American Journal of Epidemiology

Vol. 163, No. 3
DOI: 10.1093/aje/kwj026
Advance Access publication December 15, 2005

Practice of Epidemiology

Malaria Mapping Using Transmission Models: Application to Survey Data from Mali

A. Gemperli[1], P. Vounatsou[1], N. Sogoba[2], and T. Smith[1]

[1] Biostatistics and Basic Epidemiology Group, Department of Public Health and Epidemiology, Swiss Tropical Institute, Basel, Switzerland.
[2] Epidemiology and GIS Unit, Department of Medical Entomology and Vector Ecology, Malaria Research and Training Center, Faculty of Medicine, Pharmacy and Odonto-Stomatology, University of Bamako, Bamako, Mali.

Received for publication July 22, 2003; accepted for publication September 1, 2005.

Geographic mapping of the distribution of malaria is complicated by the limitations of the available data. The most widely available data are from prevalence surveys, but these surveys are generally carried out at arbitrary locations and include nonstandardized and overlapping age groups. To achieve comparability between different surveys, the authors propose the use of transmission models, particularly the Garki model, to convert heterogeneous age prevalence data to a common scale of estimated entomological inoculation rates, vectorial capacity, or force of infection. They apply this approach to the analysis of survey data from Mali, collected in 1965–1998, extracted from the Mapping Malaria Risk in Africa database. They use Bayesian geostatistical models to produce smooth maps of estimates of the entomological inoculation rates obtained from the Garki model, allowing for the effect of environmental covariates. Again using the Garki model, they convert kriged entomological inoculation rates values to age-specific malaria prevalence. The approach makes more efficient use of the available data than do previous malaria mapping methods, and it produces highly plausible maps of malaria distribution.

disease transmission; kriging; malaria; Markov chain Monte Carlo

Abbreviations: EIR, entomological inoculation rate; MARA, Mapping Malaria Risk in Africa; NDVI, Normalized Difference Vegetation Index.

Reliable maps of the prevalence or transmission intensity of malaria are urgently needed, especially in endemic areas of sub-Saharan Africa. Such maps are fundamental for estimating the scale of the problem, and hence the resources needed to combat malaria. They provide benchmarks for assessing the progress of control and indicate which geographic areas should be prioritized.

Malariological measures that might be mapped include categories of endemicity (e.g., unstable, mesoendemic, holoendemic); vector-based factors (vector densities, vectorial capacity, entomological inoculation rate (EIR)); incidence of disease; or the force of infection. However, although malaria endemicity can vary widely over only short distances, most of these measures have been studied in only a few widely separated localities, and, in general, the measurements available from distinct sites differ. The most widely available malariological measures are point prevalence data, assessed by microscopy. Malaria prevalence at unsampled locations can be estimated by incorporating information from environmental covariates (1). The precision of such estimates can be further improved by using spatial smoothing or geostatistical methods (2–5).

Spatial statistical models have already made substantial contributions to the modeling of malaria risk (2–7), and Bayesian geostatistical methods have demonstrated their value for this application in the work of Diggle et al. (5)

Correspondence to Dr. Armin Gemperli, Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205 (e-mail: agemperl@jhsph.edu).

(also refer to Thomson et al. (8)) for mapping childhood malaria risk in The Gambia and Gemperli et al. (9) for relating infant mortality to malaria risk. Spatial statistical models have also been used to produce malaria maps of the whole of West Africa (3) and specifically of Mali (2). All of these analyses directly modeled the prevalence data without taking into account age dependence of the malaria risk.

Malaria prevalence data are usually reported by age group but with different age groupings used in different series of surveys. Direct mapping of age-prevalence data therefore involves choosing a target age group (with some flexibility in the choice of age-category boundaries) and discarding data for other age groups and for sites for which data are not available for the target age group.

We propose to replace this subjective and inefficient procedure with a mathematical model to convert a set of heterogeneous malariological indices into a common scale for mapping purposes. Mathematical models, such as that from the Garki project (10), can be used to predict the associations of different measures of malaria transmission and endemicity with the shape of the age-prevalence relation. Statistical fitting of the Garki model can therefore be used to estimate any malariological parameter predicted by the model as a function of whatever community-based malariological data are available for a site. We recently used this approach to obtain an interval estimate of the EIR on the island of Príncipe from the age-prevalence curve for *Plasmodium falciparum* malaria (11).

In the present study, we applied this approach to an assemblage of age-prevalence data from Mali. The data were extracted from the Mapping Malaria Risk in Africa (MARA) database (12), the most comprehensive database on malaria in Africa, containing survey data obtained since the early 1960s. Using the Garki model, we translated the raw prevalence data from each MARA survey into an (interval) estimate of the EIR. We then followed Bayesian geostatistical methods to generate smooth maps of the EIR, allowing for the effects of environmental covariates.

We used estimates from the fitted model to produce smooth EIR maps for Mali via Bayesian kriging. The predicted EIRs were compared with observed EIR values obtained from spray catches from 186 surveys in Mali. Again using the Garki model, we also converted the kriged EIR values to estimates of malaria prevalence in children less than 5 years of age and in children 2–10 years of age and produced maps of these parameters.

TABLE 1. Age range of participants in the Mapping Malaria Risk in Africa surveys conducted in Mali, 1965–1998

Age categories (years)	No. of surveys	No. of blood slides	No. positive	Malaria prevalence (%)
2–9	52	2,787	1,842	66.1
5–9 and 10–14	15	3,176	1,786	56.2
0–44	12	2,488	230	9.2
0–1 and 2–4 and 5–9	12	8,842	4,084	46.2
0–1 and 2–4 and 5–9 and 10–15	11	2,722	1,528	56.1
1–2 and 3–5 and 6–10	10	8,284	3,883	46.9
0–1 and 2–4 and 5–9 and 10–14	9	3,616	912	25.2
5–9	9	1,435	468	32.6
0–12	6	360	160	44.4
1–15	4	2,715	736	27.1
6–14	4	923	108	11.7
0–15	3	800	481	60.1
2–15 and 16–70	3	582	207	35.6
0–1 and 2–9	2	129	66	51.2
0–5 and 6–10	2	279	121	43.4
1–9	2	712	93	13.1
2–9 and 10	2	346	215	62.1
8–14 and 15–19 and 20–29 and 30–39 and 40–49 and 50–59	2	2,023	1,063	52.5
0–1 and 2–4 and 5–9 and 10–14 and 15–19	1	110	72	65.5
1–4 and 5–9 and 10–14 and 15–24 and 25–34 and 35–44 and 45–54 and 55–64	1	476	308	64.7
2–9 and 10–60	1	251	124	49.4
6–9	1	300	77	25.7

MATERIALS AND METHODS

Data sources

Data on malaria prevalence were extracted from the MARA/ARMA database (12). To our knowledge, this is the most comprehensive database compiled by an international collaboration initiated to provide a database and an atlas of malaria in Africa by collating both published and unpublished results of malariological surveys since 1965. We selected data from 164 surveys carried out in 147 locations in Mali between 1965 and 1998 covering various age-group ranges (table 1). EIRs were estimated by fitting the Garki model to the malaria survey data. A geostatistical model was fitted to EIR estimates to produce smooth maps of malaria transmission. The maps were adjusted for environmental and climatic covariates. In our analysis, we considered the same covariates used by Kleinschmidt et al. (2) to produce smooth maps of malaria prevalence fitted directly to MARA prevalence data—that is, the average maximum temperature from March to May, the length of the rainy season defined as the number of months with more than 60 mm of rainfall, the distance from the nearest water source, and the Normalized Difference Vegetation Index (NDVI).

Data on temperature and length of the rainy season were obtained from the Topographic and Climate Database for Africa, version 1.1, by Hutchinson et al. (13). The database includes spatial estimates of monthly values averaged over years for the whole continent of Africa at a resolution of 0.05 degrees of longitude and latitude. The base data are collected from diverse research agencies and contain measurements between 1920 and 1980 averaged for at least 5

years. Daily maximum temperature is recorded at 1,499 stations and rainfall at 6,051 stations in Africa. The predictions are created by using thin-plate splines (14), where the standard errors are reported to lie below 0.5°C for the temperature and between 5 and 15 percent for the rainfall data.

The NDVI values were extracted from the NOAA/NASA Pathfinder AVHRR Land Project database (15), which records daily observed emitted and reflected radiations in different channels of the electromagnetic spectrum, sent by a satellite on a spatial resolution of 8 km. To reduce distortion because of clouds and atmospheric contaminants, we used as a composite measure the maximum value over 10 days because clouds reduce the reported NDVI value. The NDVI is derived from the reflectance rate of two (visual and near-infrared) channels. It has been shown to be highly correlated with vegetation parameters (16) and is used as a proxy for vegetation. In contrast to the other predictors used, the NDVI was able to express temporary variability.

Data on the observed EIR were extracted from the largest EIR database in Mali that compiles historic surveys conducted by the Malaria Research and Training Center of the University of Mali in Bamako. The data used for this study were obtained via spray catches from 186 surveys carried out at 14 distinct locations. They were geocoded and used to validate the model-based predicted EIR values.

TABLE 2. Estimates of coefficients and covariance parameters in the regression model for the natural logarithm of the annual entomological inoculation rate*

Variable	Median	5% Quantile	95% Quantile
Intercept	−13.46	−32.85	8.17
Distance to water			
4–40 km	−1.44	−2.18	−0.71
>40 km	−1.07	−2.57	0.42
Duration of the rainy season	0.31	−0.37	1.00
Temperature†	0.34	−0.20	0.82
NDVI‡	9.73	−2.14	21.16
σ^2‡	3.341	2.733	4.153
ρ‡	118.520	26.165	340.886
τ^2‡	2.743	0.883	6.868

* The annual entomological inoculation rate was derived by fitting the Garki model to the malaria prevalence data from the Mapping Malaria Risk in Africa surveys conducted in Mali, 1965–1998.

† Average maximum temperature from March to May in degrees Celsius.

‡ NDVI, Normalized Difference Vegetation Index; σ^2, variance of the spatial process; ρ, range parameter of the spatial exponential process (in kilometers); τ^2, residual variance.

Statistical analysis

Fitting the Garki model. We used the mathematical model of the Garki project to convert the observed prevalence at each location to estimated EIR values. The model comprises a set of linked difference equations describing transitions among seven categories of host distinguished by their infection and immunologic status. This model can be used to predict the age-specific prevalence of *P. falciparum* in humans as a function of transmission measures, including the vectorial capacity, the EIR, or the force of infection. To estimate the EIR from community parasitological survey data, the equilibrium age-prevalence curves for the Garki model were estimated for different values of the EIR, using a golden section search routine to locate the maximum likelihood estimate (17). Asymptotic 95 percent confidence estimates were obtained by numerical estimation of the Fisher's information. Further details of the models used are reported in the Appendix.

Spatial modeling of EIR. We assumed that the logarithmic-transformed EIR estimates, Y_j, for location j were normally distributed, having a mean that is a function of the covariates $\boldsymbol{X}_j$. We modeled the spatial dependency by assuming that the covariance of the EIR values at two locations, say, i and j, decreases with their distance d_{ij}; that is, $\Sigma_{ij} = \mathrm{Cov}(Y_i, Y_j) = \sigma^2 \exp(-d_{ij}/\rho)$, where σ^2 is the spatial variance and ρ is a parameter describing the degree of correlation decay. In addition, the variance of EIR at each location i was specified by $\Sigma_{ii} = \mathrm{Var}(Y_i) = \tau^2 + \sigma^2$, where τ^2 models the remaining nonspatial variation in EIR that is not explained by the covariates. Under the assumptions of second-order stationarity, the covariance matrix Σ determines a well-known exploratory tool in geostatistics, the variogram. The τ^2 corresponds to the nugget parameter, the σ^2 estimates the partial sill, and the ρ is related to the range, that is, the minimum distance at which the spatial correlation is less than 5 percent, which is 3ρ.

We chose Bayesian methods in model fit and prediction (kriging) because they allow estimation of the precision of model parameters and kriged EIR values without depending on asymptotic inference, for which several competing definitions exist (18, p. 350). We estimated the model parameters by using Markov chain Monte Carlo methods. Further details of this modeling approach are given in the Appendix. The analysis was implemented by using software written by the authors in Fortran 95 (Compaq Visual Fortran, version 6.6; Hewlett-Packard Company, Palo Alto, California) using IMSL numerical libraries (Visual Numerics, Inc., Houston, Texas).

Model validation. The model-based, predicted EIR estimates were validated against field-based EIR data by constructing a confidence interval for the difference between the two values (19). This method was also applied to compare model-based prevalence with observed prevalence data. The prevalence data were validated by using the same data as those used to fit the model.

RESULTS

Parameter estimates are summarized in table 2. The only environmental covariate significantly related to transmission intensity was distance from the water, indicating high transmission in the areas within 4 km of the water source. Duration of the rainy season, NDVI, and temperature were not statistically significantly related to EIR. Estimates of ρ suggest a strong spatial correlation reflected in the high

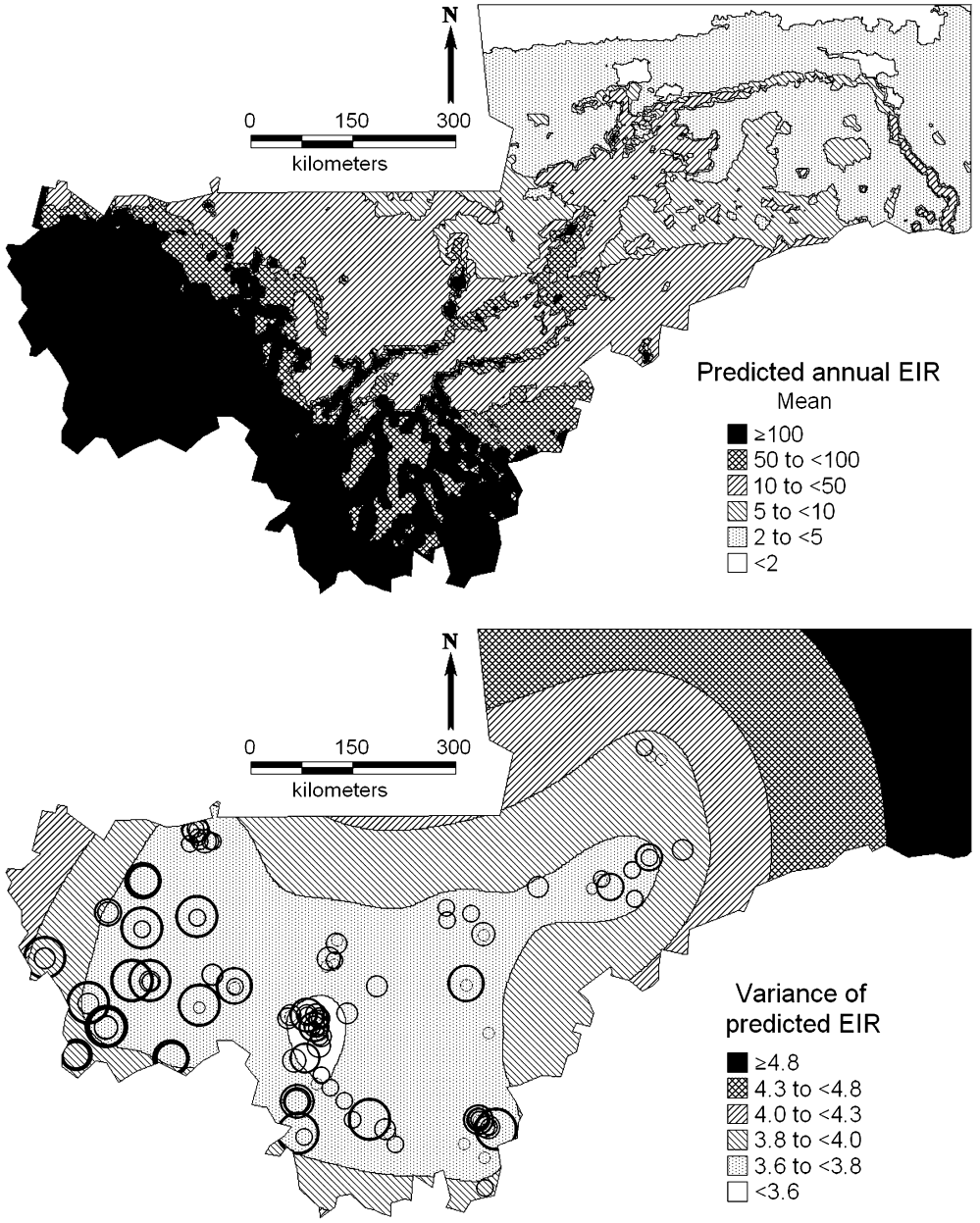

FIGURE 1. Spatial prediction of the annual entomological inoculation rate (EIR) for malaria in Mali. Bottom: survey locations are indicated by circles whose diameters are proportional to the natural logarithm of the estimated annual EIR.

median range distance—the minimum distance between two points whose correlation is below 5 percent—of 356 km (90 percent confidence interval: 78 km, 1,023 km). Locations 50 km apart have a median correlation of 65 percent. In the survey data, 92.5 percent of all distances between pairs of locations are within the distance of the median range of 356 km.

In addition, the EIR estimates show high variability estimated by τ^2. The map of predicted EIR estimates is shown in figure 1. It depicts a clear north-south and east-west pattern of transmission, ranging from disease-free regions in the Sahara Desert to high-prevalence areas in the southern and western parts of Mali. With this map, it is possible to predict the high transmission areas along the Niger River and in the Niger Delta, which brings large water masses to otherwise low-endemic areas. It can also be used to identify distinct foci of high EIR around water sources. Estimates of the prediction error are shown in the bottom map of figure 1. The small prediction error in the regions around Bamako, Nioro, and Mopti reflects the high density of surveys carried

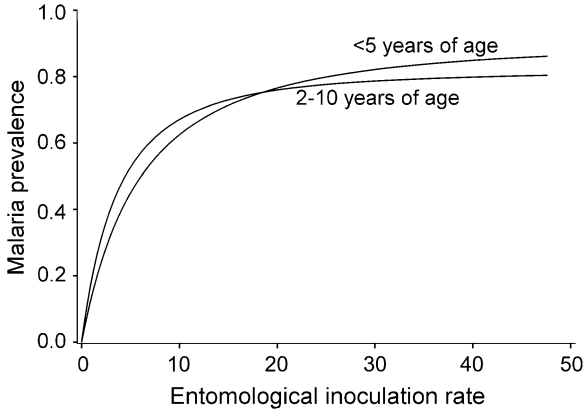

FIGURE 2. Relation between the prevalence of malaria and the entomological inoculation rate in Mali, estimated by using the Garki model for two age categories of children (neglecting the effects of seasonality).

out in those regions. In contrast, predictions in the northern part of the country are not reliable because of the sparse malaria surveys conducted in the Sahara Desert.

The mean difference between the field data and the model-based predicted EIR values was 7.59 (95 percent confidence interval: −139.41, 154.58), with EIRs from field data being larger. Comparison of the observed prevalence data with the model-based prevalence estimates gave a mean difference of −0.03 (95 percent confidence interval: −0.50, 0.45) for those less than 5 years of age and −0.03 (95 percent confidence interval: −0.42, 0.36) for those aged 2–10 years.

Figure 2 displays the relation between malaria prevalence and transmission intensity estimated by fitting the Garki model to the malaria survey data. This model enabled us to estimate this relation according to age. We chose two groups: those younger than 5 years of age and those 2–10 years of age. The figure shows that, at high levels of transmission, children younger than 5 years of age tend to be at higher risk than older children. The opposite is observed for areas of low transmission. Maps of malaria prevalence for these two age groups of children are shown in figure 3. Data for these maps were calculated by converting the EIR values to malaria prevalence using the EIR-prevalence relation shown in figure 2. The middle level of transmission in the districts of Koulikoro and Segou in central-western Mali contributes to higher malaria risk in the group less than 5 years of age than in the group 2–10 years of age. This reversion of risk order is due to the change point in the prevalence-EIR relation for the chosen age categories at that level of transmission.

DISCUSSION

An advantage of the transmission-model-based approach is that it is age adjusted and makes use of all the survey data available (we did not have to discard any surveys because of inappropriate age groups). In principle, this approach can exploit whatever population-based malariological data are available.

The transmission-model-based approach also makes it possible to tailor the outputs of the mapping exercise to the specific needs of users, who may be interested in specific age groups of hosts or predictions of malariological indices that are rarely measured in the field, as illustrated by our maps of malaria transmission intensity in Mali as well as of age-specific prevalence.

In contrast to earlier geostatistical malaria models, we fitted the Bayesian spatial model to the log-transformed EIR values, a continuous outcome with an approximately Normal distribution, estimated by fitting the Garki model to the available age-prevalence data. The Bayesian approach allows flexible model fitting and estimation and mapping of the prediction error. The method also enabled us to generate maps of the prediction error, demonstrating which geographic areas need further field investigation if the maps are to be uniformly reliable.

The maps we produced broadly correspond to the known distribution of malaria in Mali and, in particular, indicate high transmission of malaria in the areas around the main rivers, the Niger and Senegal. However, a specific difficulty arises in modeling the relation between malaria and distance to water bodies in West Africa. The presence of water bodies and flooding in low-rainfall zones undeniably leads to malaria transmission in areas that would otherwise be malaria free; in general, mosquito numbers are highest near water, especially areas prone to flooding. However, recent studies in Niono in Mali (20) found that the highest malaria risk can be several kilometers away from the main anopheline breeding sites. This finding may reflect a greater tendency of people exposed to very high mosquito densities to adopt protective measures, together with the lower average age of mosquitoes close to sites of emergence. In Mali, the center of the inland delta of the Niger is not considered a zone of highest risk (21), and, in the model of Kleinschmidt et al. (2), the river system did not appear to strongly influence malaria distribution. Risk in locations within 4 km of the nearest water body was estimated to be lower than in locations 4–40 km from water. In the map of malaria across West Africa (3), broad zones of lower malaria risk close to rivers were estimated.

Most malaria surveys include people in areas of several square kilometers, so surveys close to water bodies may include some people from the riverbank and others from several kilometers away. Therefore, it is not obvious what relation with distance to water to expect. The exact relation between proximity to rivers and malaria appears to be very sensitive to which data points are included and to the details of the model, especially when there are very few data points in the critical areas of the river floodplains. It may also be that the lack of adjustment for age in the earlier models biased some of the covariate effects. Because we were able to include data from 164 surveys rather than just the 101 analyzed by Kleinschmidt et al. (2), we have some confidence that the present model provides an improved estimate of the broad geographic pattern of malaria, although possibly not of local variation within this pattern.

We chose to use the Garki model to estimate EIR from age-prevalence curves because, when it was originally developed, it was designed to accurately reproduce this

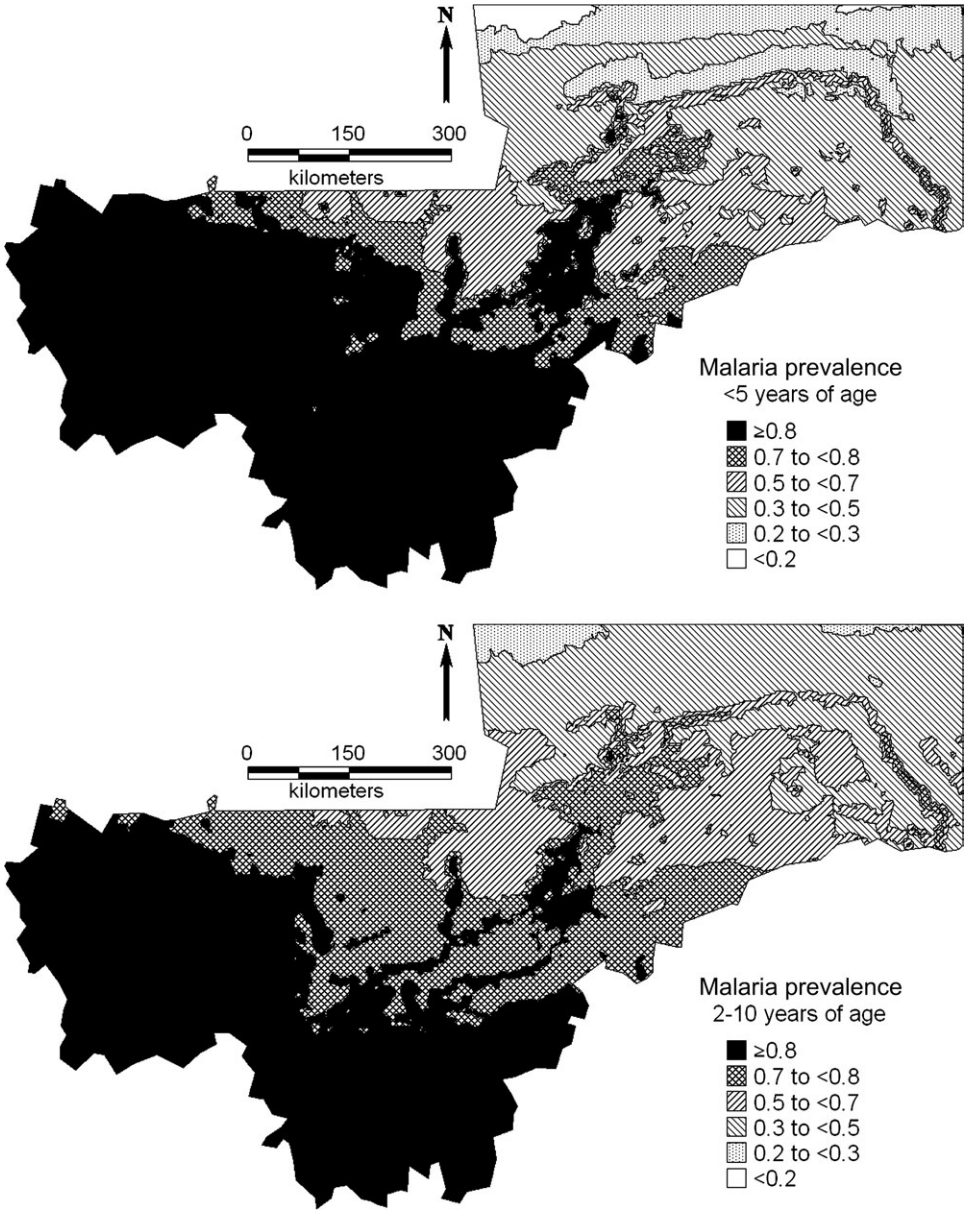

FIGURE 3. Spatial prediction of the age-specific prevalence of malaria among children in Mali, derived after transforming the predicted entomological inoculation rate.

relation in field data from the savannah zone of Nigeria (22). The Nigerian field site was in many ways similar to southern Mali; hence, the model is likely to be most accurate for the range of conditions seen in our study. However, flexibility in the outputs comes at the price of making many approximations. Like all mathematical models, the Garki model is a simplification. Some elements of it could probably be improved by using recent insights from molecular epidemiology studies and advances in statistical computation. Comparison of the model predictions with the entomological data shows that this approach has potential, although the validation data are sparse and therefore the confidence intervals for the mean difference are very wide. The original fitting of the Garki model to data was rather limited, and better malaria transmission models are clearly needed.

The main simplification inherent in our application to Malian data is that, following other exercises in empirical mapping of malaria (2), we ignored the seasonal patterns, although both acquisition of the data and transmission of malaria itself were seasonal. Seasonality in transmission is an important consideration in interpreting the EIR map (figure 1), because, when many inoculations occur over a

short period of time, the proportion resulting in erythrocytic infections is reduced (23, 24). Clustering of inoculations in the transmission season thus means that the average force of infection is lower than would result from the same number of entomological inoculations spread over the whole year. The Garki model does capture this phenomenon, but only if a seasonal input of vectorial capacity is assumed. Since, in the present analysis, we assumed a constant vectorial capacity for each location, the true EIR values in Mali must be higher than those we estimated.

The prevalence maps (figure 3) are less affected by seasonality because the transformation back to a scale of prevalence corrects the bias introduced by assuming uniform yearly transmission. Moreover, prevalence is known to show much less seasonality than does EIR (22, 24, 25). To compare malaria risk at the regional level, and in zones where the degree of seasonality varies considerably, it will be essential to correctly allow for seasonality when estimating transmission parameters from age-prevalence data. In further developing our model-based approach to malaria mapping, we propose to use maps of seasonality in transmission as an input to the modeling procedure to correct for the biases in EIR estimates.

ACKNOWLEDGMENTS

The work of the first author was supported by Swiss National Science Foundation grant 3200-057165.99.

The authors thank the NOAA/NASA Pathfinder AVHRR Land Project (University of Maryland, College Park, Maryland) and the Distributed Active Archive Center (Code 902.2) at the Goddard Space Flight Center, Greenbelt, Maryland, for producing and distributing these data, respectively.

This work is a product of the MARA collaboration. The authors also acknowledge the contributions of the many field, laboratory, and office workers who carried out the surveys and compiled the malariological database.

Conflict of interest: none declared.

REFERENCES

1. Hay SI, Rogers DJ, Toomer JF, et al. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, Internet access and review. Trans R Soc Trop Med Hyg 2000;94:113–27.
2. Kleinschmidt I, Bagayoko M, Clarke GPY, et al. A spatial statistical approach to malaria mapping. Int J Epidemiol 2000; 29:355–61.
3. Kleinschmidt I, Omumbo J, Briët O, et al. An empirical malaria distribution map for West Africa. Trop Med Int Health 2001;6:779–86.
4. Kleinschmidt I, Sharp BL, Clarke GPY, et al. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. Am J Epidemiol 2001;153:1213–21.
5. Diggle PJ, Moyeed RA, Rowlinson B, et al. Childhood malaria in The Gambia: a case-study in model-based geostatistics. Appl Stat 2002;51:493–506.
6. Ribeiro JMC, Seulu F, Abose T, et al. Temporal and spatial distribution of anopheline mosquitos in an Ethiopian village: implications for malaria control strategies. Bull World Health Organ 1996;74:299–305.
7. Thomas CJ, Lindsay SW. Local-scale variation in malaria infection amongst rural Gambian children estimated by satellite remote sensing. Trans R Soc Trop Med Hyg 2000;94: 159–63.
8. Thomson MC, Connor SJ, D'Alessandro U, et al. Predicting malaria infection in Gambian children from satellite data and bednet use surveys: the importance of spatial correlation in the interpretation of results. Am J Trop Med Hyg 1999;61:2–8.
9. Gemperli A, Vounatsou P, Kleinschmidt I, et al. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. Am J Epidemiol 2004;159:64–72.
10. Dietz K, Molineaux L, Thomas A. A malaria model tested in the African savannah. Bull World Health Organ 1974;50: 347–57.
11. Hagmann R, Charlwood JD, Gil V, et al. Malaria and its possible control on the island of Príncipe. Malar J 2003;2:15.
12. MARA/ARMA. Towards an atlas of malaria risk in Africa. First technical report of the MARA/ARMA collaboration. Durban, South Africa, 1998. (http://www.arma.org.za/tech_report_eng.htm).
13. Hutchinson MF, Nix HA, McMahon JP, et al. Topographic and climate database for Africa. (CD-ROM). The Australian National University, Canberra, ACT 0200, Australia, 1995. (http://cres.anu.edu.au/outputs/africa.php).
14. Hutchinson MF. The application of thin plate splines to continent-wide data assimilation. In: Jasper JD, ed. Data assimilation systems. Melbourne, Australia: Bureau of Meteorology, 1991:104–13. (BMRC res. rep. no. 27).
15. Agbu PA, James ME. NOAA/NASA Pathfinder AVHRR Land Data Set User's Manual. Goddard Distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt, MD. 1994. (http://disc.sci.gsfc.nasa.gov/landbio/AVHRR_GD.pdf).
16. Justice CO, Townshend JRG, Holben BN, et al. Analysis of the phenology of global vegetation using meteorological satellite data. Int J Remote Sens 1985;6:1271–318.
17. Press WH, Flannery BP, Teukolsky SA, et al. Numerical recipes in C: the art of scientific computing. Cambridge, MA: Cambridge University Press, 1988.
18. Cressie NAC. Statistics for spatial data. New York, NY: Wiley, 1993.
19. Bland MJ, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
20. Dolo G, Briët OJT, Dao A, et al. Rice cultivation and malaria transmission in the irrigated Sahel of Mali, West Africa. (In French). Cahiers Agricultures 2000;9:425.
21. Doumbo O. Epidemiology of malaria in Mali, study of chloroquinoresistance, testing control strategies using permethrine impregnated bed-nets and its relation to the treatment of accesses of fever. (In French). Doctoral thesis. Montpellier, France, University of Montpellier, 1992.
22. Molineaux L, Gramiccia G. The Garki project: research on the epidemiology and control of malaria in the Sudan savanna of West Africa. Geneva, Switzerland: World Health Organization, 1980.
23. Beier JC, Oster CN, Onyango FK, et al. *Plasmodium falciparum* incidence relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in western Kenya. Am J Trop Med Hyg 1994;50:529–36.

24. Charlwood JD, Smith T, Lyimo E, et al. Incidence of *Plasmodium falciparum* infection in infants in relation to exposure to sporozoite-infected anophelines. Am J Trop Med Hyg 1998;59:243–51.
25. Smith T, Charlwood JD, Kihonda J, et al. Absence of seasonal variation in malaria parasitaemia in an area of intense seasonal transmission. Acta Trop 1993;54:55–72.
26. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85:398–409.

APPENDIX

Garki model

The mathematical model from the Garki project is used to predict the age-specific prevalence of *P. falciparum* in humans as a function of the vectorial capacity, C. It comprises a set of linked difference equations describing transitions among seven categories of host distinguished by their infection and immunological status. Two compartments, comprising proportions x_1 and x_3 of the population, account for the uninfected individuals; two are for those with prepatent infections (x_2 and x_4); and the remaining three, comprising proportions y_1, y_2, and y_3, account for those with blood-stage infections.

The model consists of an algorithm for predicting the proportion of the human population of each age in each of these compartments and is defined by a set of difference equations (equations 1–7) that specify the change in each of these proportions from one time point to the next (i.e., $\Delta x_1 = x_1(t+1) - x_1(t)$).

$$\Delta x_1 = \delta + y_2 R_1(h) - (h+\delta)x_1 \quad (1)$$

$$\Delta x_2 = hx_1 - (1-\delta)^N + h(t-N)x_1(t-N) - \delta x_2 \quad (2)$$

$$\Delta x_3 = y_3 R_2(h) - (h+\delta)x_3 \quad (3)$$

$$\Delta x_4 = hx_3 - (1-\delta)^N + h(t-N)x_3(t-N) - \delta x_4 \quad (4)$$

$$\Delta y_1 = (1-\delta)^N + h(t-N)x_1(t-N) - (\alpha_1+\delta)y_1 \quad (5)$$

$$\Delta y_2 = \alpha_1 y_1 - (a_2 + R_1(h) + \delta)y_2 \quad (6)$$

$$\Delta y_3 = \alpha_2 y_2 - (1-\delta)^N + h(t-N)x_3(t-N) - (R_2(h)+\delta)y_3 \quad (7)$$

The meanings of the additional symbols are given in appendix table 1. For simplicity, the time points to which the proportions and the force of infection refer are indicated in these seven equations only when they differ from t.

To complete the model, h, the force of infection, must be specified as a function of the vectorial capacity C. From the definition of C, it follows that each mosquito bite on an infective individual will result in C new inoculations, n days later (where n is the duration of sporogony). Since a proportion $y_1(t)$ of the population is infective, the entomological inoculation rate E is $E(t) = C(t-n)\, y_1(t-n)$.

APPENDIX TABLE 1. Quantities in the Garki model

Symbol	Meaning	Default value
δ	Human birth and death rates	36.5/100 years
α_1	Rate at which nonimmunes move into the noninfective category	0.002/day
α_2	Rate at which nonimmunes recovering from infection move into the immune category	0.00019/day
h	Force of infection (rate of infection of susceptibles)	To be estimated
N	Duration of prepatent period	15 days
r_1	Recovery rate for individual clones (nonimmune)	0.0023/day
r_2	Recovery rate for individual clones (immune)	$10r_1$
$R_1(h)$	Recovery rate from infection in nonimmunes y_2 (as a function of h)	To be estimated
$R_2(h)$	Recovery rate from infection in immunes y_3 (as a function of h)	To be estimated
g	Maximum value of force of infection	0.097/5 days
q_1	Detectability of parasites in infectives (y_1)	1
q_2	Detectability of parasites in nonimmunes (y_2)	1
q_3	Detectability of parasites in immunes (y_3)	0.7

To ensure that the model reproduces the observed saturation in the force of infection as E increases, $h(t)$ is assumed to be related to $E(t)$ via the equation $h(t) = g(1 - \exp(-E(t)))$, where g then represents the upper limit of the force of infection and is hence a parameter measuring host susceptibility. The problems of superinfection and acquired immunity are addressed by specifying the recovery rates, R_1, and R_2, as functions of h using the relation $R = h/(\exp(h/r) - 1)$, where r is the recovery rate for single clone infections. Nonimmunes are assumed to recover at rate R_1, calculated from this equation by setting $r = r_1$. Immunes recover at rate R_2, calculated by setting $r = r_2$, where $r_2 > r_1$. In addition to this difference, acquisition of immunity prevents transmission from the human host to the mosquito.

To examine the fit of the model to real parasite prevalence data, three detectability parameters, q_1, q_2, and q_3, are required to allow for imperfect detection of parasitemia in each of the three infected classes (appendix table 1). The overall observed prevalence is then $z(t) = q_1 y_1(t) + q_2 y_2(t) + q_3 y_3(t)$. Predicting the age-specific parasite rate for any given vector $C(t)$ then involves a two-stage process. Initially, the model (equations 1–7) is simulated by starting with arbitrary values of x_1 to x_4 and y_1 to y_3 until equilibrium is reached. The input, $C(t)$, may be constant or may vary cyclically. Following the original implementation, we use time intervals of 5 days. It is therefore natural to consider the

input to follow a 365-day (73 time intervals) repeating cyclical pattern. Convergence is judged to have been achieved when each of x_1 to x_4 and y_1 to y_3 is equal to the value attained 73 time units previously. The life history of a cohort of individuals born into the nonimmune susceptible category is then simulated by running the model with x_1 initialized to be 1, δ set to 0, and $C(t)$ set to the equilibrium values. To incorporate effects of the season of birth, a series of such cohorts are simulated with birth dates spread uniformly throughout the year.

Geostatistical model

Let Y_j be the logarithmic transformation of the EIR estimates, at location $j = 1, \ldots, n$. We assume that Y_j are normally distributed, having a mean that is a function of the covariates $\boldsymbol{X}_j$ at j. We model the spatial dependency by assuming that the covariance of the EIR values at two locations, say, i and j, decreases with their distance d_{ij}, that is, $\Sigma_{ij} = \mathrm{Cov}(Y_i,Y_j) = \sigma^2 R_{ij}(\rho)$ with $R_{ij} = \exp(-d_{ij}/\rho)$, where σ^2 is the spatial variance and ρ is a parameter describing the degree of correlation decay. In addition, the variance of EIR at each location i is specified by $\Sigma_{ii} = \mathrm{Var}(Y_i) = \tau^2 + \sigma^2$, where τ^2 models the remaining nonspatial variation in EIR that is not explained by the covariates. The likelihood function of the EIR is multivariate Normal, that is, $\boldsymbol{Y} \sim MVN(\boldsymbol{X}^T\beta, \Sigma)$, where $\Sigma = \tau^2 I_n + \sigma^2 R(\rho)$ and I_n is the unity matrix of dimension n.

To complete Bayesian formulation of the model, we specify prior distributions for the model parameters β, ρ, σ^2, and τ^2. In particular, we adopt independent Normal noninformative priors for the regression coefficients β_k, $\beta_k \sim N(0, 10^6)$; inverse Gamma priors for the variance parameters $\sigma^2 \sim IG(a_1, b_1)$ and $\tau^2 \sim IG(a_2, b_2)$; and a Gamma prior for the ρ parameter, where $\rho \sim Ga(a_3, b_3)$ with hyper-priors $a_1 = a_2 = 2.01$, $b_1 = b_2 = 1.01$, and $a_3 = b_3 = 0.01$. Following the Bayesian paradigm, the full posterior distribution takes the form

$$[\beta, \rho, \sigma^2, \tau^2 \mid \boldsymbol{Y}] \propto \det(\tau^2 I_n + \sigma^2 R(\rho))^{-1/2} \times \exp\left(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}^T\beta)^T(\tau^2 I_n + \sigma^2 R(\rho))^{-1}(\boldsymbol{Y} - \boldsymbol{X}^T\beta)\right) \times [\beta, \rho, \sigma^2, \tau^2].$$

We estimate the parameters of the model by using Markov chain Monte Carlo and, in particular, Gibbs sampling (26). Implementation of the Gibbs sampler requires simulating from the conditional posterior distributions of all parameters.

The full conditional posterior distribution of β is a Normal distribution, and it is straightforward to simulate from. The conditional posterior distributions of σ^2, τ^2, and ρ have nonstandard forms. We sampled from these distributions by using a random-walk Metropolis algorithm having a Gaussian proposal density with mean equal to the estimate from the previous iteration and variance derived from the inverse second derivative of the log-posterior. To estimate the unobserved logarithm of EIR at a set of new locations $s_{01}, s_{02}, \ldots, s_{0l}$, we use Bayesian kriging. Let $\boldsymbol{Y}_0 = (Y(s_{01}), Y(s_{02}), \ldots, Y(s_{0l}))$ denote the values to predict. Then, the predictive distribution

$$P(\boldsymbol{Y}_0 \mid \boldsymbol{Y}) = \int P(\boldsymbol{Y}_0 \mid \boldsymbol{Y}, \beta, \sigma^2, \tau^2, \rho) \times P(\beta, \sigma^2, \tau^2, \rho \mid \boldsymbol{Y})\, d\beta d\tau^2 d\sigma^2 d\rho \qquad (8)$$

is numerically approximated by the average $1/r \times \sum_{k=1}^{r} P(\boldsymbol{Y}_0 \mid \boldsymbol{Y}, \beta^{(k)}, \sigma^{2(k)}, \tau^{2(k)}$. $\beta^{(k)}, \sigma^{2(k)}, \tau^{2(k)}$ and $\rho^{(k)}$ are samples drawn from the posterior $P(\beta, \sigma^2, \tau^2, \rho \mid \boldsymbol{Y})$ and $P(\boldsymbol{Y}_0 \mid \boldsymbol{Y}, \beta, \sigma^2, \tau^2, \rho) = \mathcal{N}(\boldsymbol{X}_0^T\beta + \Sigma_{01}\Sigma_{11}^{-1}(\boldsymbol{Y} - \boldsymbol{X}^T\beta), \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10})$, when $\Sigma_{01} = \Sigma_{10}^T = \mathrm{Cov}(\boldsymbol{Y}_0, \boldsymbol{Y}), \Sigma_{11} = \mathrm{Var}(\boldsymbol{Y})$ and $\Sigma_{00} = \mathrm{Var}(\boldsymbol{Y}_0)$.