Additional File 2: Bayesian model-based geostatistical modelling procedures

2.1 Model overview

The underlying value of *Plasmodium falciparum* parasite rate in the 2-10 year age cohort, $Pf\mathrm{PR}_{2-10}(x_i)$, at each location x_i for the year 2009 was modelled as a transformation $g(\cdot)$ of a spatiotemporally structured field superimposed with additional random variation $\epsilon(x_i)$ [1]. The count of individuals tested as *P. falciparum* positive N_i^+ from the total sample of N_i in each survey was modelled as a conditionally independent binomial variate given the unobserved underlying age-standardized $Pf\mathrm{PR}_{2-10}$ value [2]. The spatiotemporal component was represented by a stationary Gaussian process $f(x_i, t_i)$ with mean μ and covariance C. The unstructured component $\epsilon(x_i)$ was represented as Gaussian with zero mean and variance V. Both the inference and prediction stages were coded using Python (PyMC version 2.0) [3]

Mean definition

The mean component μ was modelled as a linear function of time, t, and whether the prediction location x was: urban (denoted by the indicator variable $1_{ur_1}(x)$) rather than rural; of maximum temperatures <25°C ($1_{tm_1}(x)$) or >30°C ($1_{tm_2}(x)$) compared to between 25°C-30°C; of zero ($1_{pr_1}(x)$) or 1-3 ($1_{pr_2}(x)$) sets of three continuous months of precipitation >60 mm in average year compared to >3 sets; at distance to water bodies of >12 km ($1_{wb_1}(x)$) rather than ≤12 km; and of Enhanced Vegetation Index (EVI) of ≤0.3 ($1_{ev_1}(x)$) rather than >0.3. The mean component was therefore defined by nine parameters:

$$\begin{aligned} mu =& \beta_x + \beta_t t + \beta_{ur_1} 1_{ur_1}(x) + \beta_{tm_1} 1_{tm_1}(x) + \beta_{tm_2} 1_{tm_2}(x)) \\ &+ \beta_{pr_1} 1_{pr_1}(x) + \beta_{pr_2} 1_{pr_2}(x) + \beta_{wb_1} 1_{wb_1}(x) + \beta_{ev_1} 1_{ev_1}(x) \end{aligned}$$

where β_x denotes the intercept. Each survey was referenced temporally using the mid-point (in decimal years) between the recorded start and end months.

Covariance definition

Covariance between spatial and temporal locations was modelled using a spatially anisotropic space-time covariance function C with a periodic component (wavelength=12 months) added to the time-marginal covariance model to capture seasonality [4]:

$$C(x_i, t_i; x_j, t_j) = \tau^2 \gamma(0) \frac{(\Delta x)^{\gamma(\Delta t)} K_{\gamma(\Delta t)}(\Delta x)}{2^{\gamma(\Delta t)-1}\Gamma(\gamma(\Delta t)+1)},$$

$$\gamma(\Delta t) = \frac{1}{2\rho + 2(1-\rho)\left[(1-\upsilon)e^{-|\Delta t|/\phi_t} + \upsilon\cos(2\pi\Delta t)\right]},$$

$$\Delta t = |t_i - t_j|$$

where K_γ is the modified Bessel function of the second kind of order γ, and Γ is the gamma function [5, 6].

To allow for spatial anisotropy and to accommodate the effect of the curvature of the earth on point-to-point separations, spatial distance between a pair of points x_i and x_j was computed as great-circle distance $D_{GC}(x_i, x_j)$ multiplied by a factor that depends on the angle of inclination $\theta(x_i, x_j)$ of the vector pointing from x_i to x_j. θ was computed as if latitude and longitude were Euclidean coordinates (on a cylindrical projection):

$$\Delta x = 2\sqrt{\gamma(\Delta t)} \frac{D_{GC}(x_i, x_j)\sqrt{1 - \psi^2 \cos^2(\theta(x_i, x_j) - \lambda)}}{\phi_x}$$

When $\Delta x = 0$ (that is, for points at the same location but different times), the covariance function reduces to

$$\rho + (1-\rho)\left[(1-\upsilon)e^{-|\Delta t|/\phi_t} + \upsilon\cos(2\pi\Delta t)\right]$$

As temporal separation increases, the covariance approaches a limiting sinusoid $\tau^2[\rho + (1-\rho)\upsilon\cos(2\pi\Delta t)]$ rather than zero. When $\Delta t = 0$, on the other hand (for points at different locations but the same time), it reduces to a standard exponential form with range parameter $\phi_x\sqrt{2}$. Unlike standard sum-product models, this covariance function does not have problematic ridges along its axes [4].

2.2 Table of prior specifications

All the covariate coefficients had zero mean, standard deviation 100 normal priors. The covariate values were transformed by: (value - mean) / standard deviation.

Parameter	Prior
V	$1/V \sim \text{Gamma}(.001, .004)$
τ	$\log \tau \vert \mu_\tau, V_\tau, \alpha_\tau \sim \text{Skew-Normal}(.0535, .559, 3.21)$
ϕ_x	$\log \phi_x \vert \mu_\phi, V_\phi, \alpha_\phi \sim \text{Skew-Normal}(-2.54, .704, -.015)$
ϕ_t	$\phi_t \sim \text{Exponential}(.1)$
λ	$\lambda \sim \text{Uniform}(0, \pi)$
ψ	$\psi^2 \sim \text{Uniform}(0, 1)$
ρ	$\rho \sim \text{Uniform}(0, 1)$

2.3 Model implementation and output

Bayesian inference was implemented using Markov Chain Monte Carlo to generate samples from the posterior distribution of the Gaussian field $f(x_i, t_i)$ at each data location and of the unobserved parameters of the mean, covariance function and Gaussian random noise component.

Samples were generated from the mid-year 2009 mean of the posterior distribution of $f(x_i, t_i)$ at each prediction location. For each sample of the joint posterior, predictions were made using space-time conditional simulation over the 12 months of 2009 $\{t = 2009_{Jan}, ..., 2009_{Dec}\}$ [7]. These predictions were made at points on a regular 1 × 1 km spatial grid across Kenya. Model output therefore consisted of samples from the predicted posterior distribution of the 2009 annual mean *Pf*PR$_{2\text{-}10}$ at each grid location, which were used to generate point estimates (Figure 3a in the main text) (computed as the mean of each set of posterior samples) and endemicity class membership probabilities (computed as the proportion of posterior samples for each pixel falling in the various class ranges). These latter values were used to present maps of both the class with the highest posterior probability of membership (Figure 3b in the main text), and the probability associated with that assignment (Figure 5 in the main text).

References

1. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ et al.: **A world malaria map: *Plasmodium falciparum* endemicity in 2007**. *PLoS Med* 2009, **6**: e1000048.
2. Diggle PJ, Thomson MC, Christensen OF, Rowlingson B, Obsomer V, et al: **Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty**. *Ann Trop Med Parasitol* 2007, **101**: 499-509.
3. Anand Patil, David Huard and Christopher Fonnesbeck. PyMC 2.0: **Bayesian Stochastic Modelling in Python**. *J Stat Software* 2009, submitted.
4. Stein ML: **Space-time covariance functions**. *J Am Stat Assoc* 2005, **100**: 310-321.
5. Antosiewicz HA: **Bessel Functions of Integer Order. In: Abramowitz M, Stegun IA, editors. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables**. New York, U.S.A.: Dover Publications Inc. pp. 435-478, 1964.
6. Davis GM: **Gamma function and related functions. In: Abramowitz M, Stegun IA, editors. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables**. New York, U.S.A.: Dover Publications Inc. pp. 253-293, 1964.
7. Gething PW, Noor AM, Gikandi PW, Hay SI, Nixon MS, et al: **Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data**. *Geogr Anal* 2008, **40**: 167-188.