

## **Protocol S1: The *Plasmodium falciparum* Parasite Rate Malariometric Survey Database**

The Malaria Atlas Project (MAP) *Plasmodium falciparum* parasite rate (*PfPR*) database is the largest and most contemporary global compilation of malariometric surveys [1]. *PfPR* data are updated through a continuous process of data search and acquisition and as of the 31 July 2008 the database approached 9,000 distinct community surveys, across 78 of the 87 *P. falciparum* malaria endemic countries (*PfMEC*). This document describes the important aspects of the data assembly process and provides a description of key features of the *PfPR* data used for this 2007 iteration of the global *P. falciparum* endemicity map.

### **S1.1 Summary of Data Search and Data Abstraction Procedures**

The MAP *PfPR* database has strict inclusion criteria that have been described in detail elsewhere [1]. In brief, only post-1984 *PfPR* data were included as they were most representative of contemporary malaria endemicity. The statistical precision of malariometric surveys increases with the number of people sampled; a minimum sample size of 50 was used as the threshold for inclusion [2,3]. The surveys were also required to represent random or whole community samples of the population. Repeat surveys at the same location were recorded. The assumption of independence was required between longitudinal observations and this was supported by requiring a minimum six month observation interval between surveys, after which time the presence or absence of infections in re-sampled individuals was considered to be unrelated to their infection status in the first survey [4,5]. For studies where the survey design indicated that it had been implemented to sample seasonal variation in transmission this six month rule was relaxed. Significant search efforts were then directed to find formal (published scientific papers) and informal (“grey” literature and unpublished data) sources of information on malaria surveys that fulfilled these primary inclusion criteria.

Extensive data searches were undertaken using online reference catalogues of the formal literature, mainly PubMed (<http://www.ncbi.nlm.nih.gov/sites/entrez>), ISI Web of Knowledge (<http://wok.mimas.ac.uk>) and Scopus (<http://www.scopus.com>), supplemented with additional regional archives [1]. Five main searches were undertaken in: March 2006, October 2006, May 2007, March 2008 and June 2008. In addition, several monthly and weekly malaria research summary compendia (Malaria World newsletters (<http://www.malaria-world.com/>), RBM alerts (<http://www.rollbackmalaria.org>) and the Environmental Health at USAID malaria bulletins (<http://www.ehproject.org/>)) were checked regularly for relevant publications.

The MAP *PfPR* database has a simple design that captures the main aspects of malariometric surveys [1] relevant to the modelling described in this paper. The abstraction of the data from primary sources has been conducted by the same individuals (CAG, RWS and SIH) throughout the two years of data assembly. On inspection, many information sources were

incomplete and required follow-up with the authors for missing and additional information (for example, disaggregating data to the village level, helping find the geographic location of the surveyed communities and/or providing survey dates). In order to expedite this process, further engage the malaria research and control community and request access to additional data, a contacts database was created from these initial searches. This was augmented from various sources and used to send a request for *PfPR* data in June 2006 [1]. These communications with malaria specialists have proven to be an extremely effective method of data provision and information retrieval and will be pursued actively for the duration of the MAP [1]. All contributors of data are gratefully acknowledged online (<http://www.map.ox.ac.uk/acknowledgements.html>) and arranged alphabetically by surname under the country name for which the data were provided.

One particularly challenging aspect of this data assembly was the geo-positioning of each record, for which very detailed guidelines have been developed [1]. These guidelines allowed surveys to be categorised according to the area from which they were sampled as: points ( $\leq 10$  km<sup>2</sup>), wide-areas ( $>10$  and  $\leq 25$  km<sup>2</sup>), small polygons ( $>25$  and  $\leq 100$  km<sup>2</sup>) and large polygons ( $>100$  km<sup>2</sup>). Records that were likely to be geo-positioned less precisely were marked as either a “good” (inaccuracy  $<5$  kilometres) or a “rough” guess (inaccuracy  $>5$  kilometres).

The entire database was checked for data consistency with a series of simple queries that generated reports on the number of records by data field, so that inconsistencies could be easily flagged and checked. In addition, before the data exclusion rules were applied, each row of the database was checked for missing or inconsistent information by CAG, RWS and SIH.

## **S1.2. Data Exclusion Rules**

By the 31 July 2008, 8,950 temporally independent community *P. falciparum* parasite prevalence surveys were identified globally from 78 of the 87 *PfMEC* globally [6]. Those *PfMECs* not represented in the database were Bangladesh, Belize, Bhutan, Djibouti, Dominican Republic, Guyana, Iran, Kyrgyzstan and Panama. In addition, five countries endemic only for *P. vivax* also contributed zero prevalence *P. falciparum* estimates for the modelling: Costa Rica, Iraq, Mexico, Morocco and Turkey.

Several exclusions were applied to these data in order to optimise their spatial resolution and geo-positional fidelity in the final modelling dataset (see Figure S1.2 and Section S1.5). First, 213 large and 134 small polygons were excluded because these records represented areas larger than the 5x5 km spatial resolution grid output of the model (see main text). Second, 349 surveys that could not be geo-located were excluded as this is a pre-requisite for spatial analysis. Amongst the remaining 8,254 geo-located point or wide-area records, the accuracy of the geographic coordinates for 27 was classified as a “rough” guess. These surveys were excluded because of the likely uncertainty in the estimate of their location exceeded that of the

5x5 km spatial resolution of the model output. Finally, longitudinal surveys that could not be temporally disaggregated (n=153) and those for which the month of survey was unavailable (n=83) were also excluded as this information is needed for the space-time geostatistical procedures adopted. These exclusions left 7,991 point or wide-area, geo-located parasite prevalence records for further analysis.

### S1.3. Age-Standardisation

*PfPR* data are reported in a diversity of age ranges and 311 different age groups were found in the *PfPR* database. Since population measures of malaria prevalence are age-dependent [7,8,9,10,11,12], it was necessary to standardise the *PfPR* survey estimates to a single, representative age group for comparison. All surveys were standardized to the 2 (2.00) to 10 (9.99) year age group (hereafter *PfPR*<sub>2-10</sub>) using catalytic conversion models first adapted to malaria by Pull and Grab [13] and described in detail elsewhere [14].

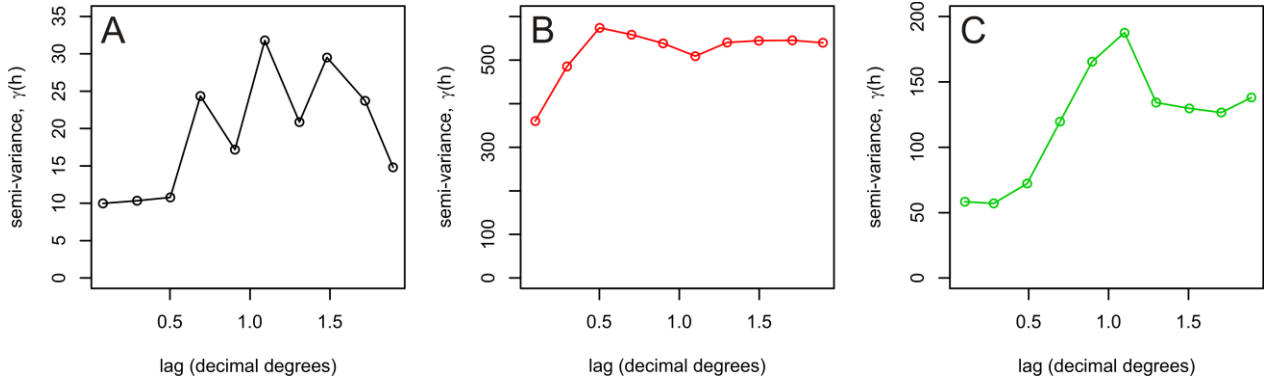
### S1.4. Semi-Variograms of *PfPR*<sub>2-10</sub> Data by Region

Sample semi-variograms were computed independently for each of the three global regions (America; Africa, Yemen and Saudi Arabia (Africa+); and Central and South East Asia (CSE Asia)) to assess the extent to which the pattern of spatial heterogeneity in observed *PfPR*<sub>2-10</sub> values differed between these regions (Figure 1, main text). Sample semi-variograms are a widely used diagnostic tool that allows simple visualisations of the spatial autocorrelation structure exhibited by spatial data [15]. More formally, they represent the effect of separation distance (termed lag,  $h$ ) on the expected difference between pairs of observed values (quantified using semi-variance,  $\gamma$ ) [16]. For these analyses, the sample semi-variogram was defined as:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (p(\mathbf{x}_i) - p(\mathbf{x}_i + h))^2 \quad (\text{S1.1})$$

where  $p(\mathbf{x}_i)$  represents a value of *PfPR*<sub>2-10</sub> observed at location  $\mathbf{x}_i$  and  $p(\mathbf{x}_i + h)$  represents a second observation at a distance  $h$  from the first. By discretizing lags into a series of bins of width  $b$ , such that each value of  $h$  actually represents a distance interval  $h \pm 1/2b$ , semi-variances  $\gamma(h)$  are computed as the mean semi-variance amongst the set of  $n(h)$  pairs of observations separated by distances within that interval. Comparison of the sample semi-variograms for each global region revealed marked differences in spatial structure (Figure S1.1). The semi-variograms for Africa+ and CSE Asia both indicate spatial autocorrelation in *PfPR*<sub>2-10</sub> values up to lags of around one decimal degree, although the much larger value of the y-axis

intercept for Africa+ indicates a larger contribution of aspatial variation in that region. The semi-variogram for America displays consistently small semi-variances indicating much more uniform  $PfPR_{2-10}$  endemicity than exhibited in Africa+ and CSE Asia, with only weak spatial autocorrelation. In conjunction with prior biogeographical, entomological and epidemiological knowledge [17,18], these distinct semi-variograms provided a rationale for the decision to stratify the modelling into the three major global regions.



**Figure S1.1.** Sample semi-variograms of  $PfPR_{2-10}$  for A America (black line), B Africa+ (red line), and C the CSE Asia (green line) regions. A decimal degree is equal to approximately 111 km at the equator.

### S1.5. Geostatistical Filter for Detection of Extreme Outliers

Section S1.2 describes the rigorous series of exclusion rules that were applied to candidate  $PfPR$  survey data before inclusion in the modelling dataset used in this study. Subsequent to the age-standardisation (Section S1.3) and as a final data quality check, the data were scanned for extreme statistical outliers: values that were deemed statistically implausible for the population of surveys sampled. A non-spatial approach to identifying  $PfPR_{2-10}$  values as statistical outliers might involve the use of an arbitrary threshold of deviance such as  $n$  standard deviations from the mean of a global or regional population. Such an approach would neglect the fact that both the mean and variance of  $PfPR_{2-10}$  may vary spatially. It is intuitive that the probability of an unusually large  $PfPR_{2-10}$  value being a genuine “outlier” is larger if it is in a neighbourhood of generally much smaller values and/or the neighbourhood is generally uniform. Conversely, if that same point lies within a neighbourhood with either similarly large values, or in which values have a tendency to vary greatly, then the probability that it is an outlier is less. To incorporate these heuristic considerations a spatial filtering algorithm was developed which implemented the following procedure for each data location  $\mathbf{x}_i$  in turn.

A large regional neighbourhood was specified around each data location,  $\mathbf{x}_i$ , defined as the set of  $n_i$  data  $\{p(\mathbf{x}_j); j = 1, \dots, n_i\}$  located within 20 decimal degrees of  $\mathbf{x}_i$ , and excluding the

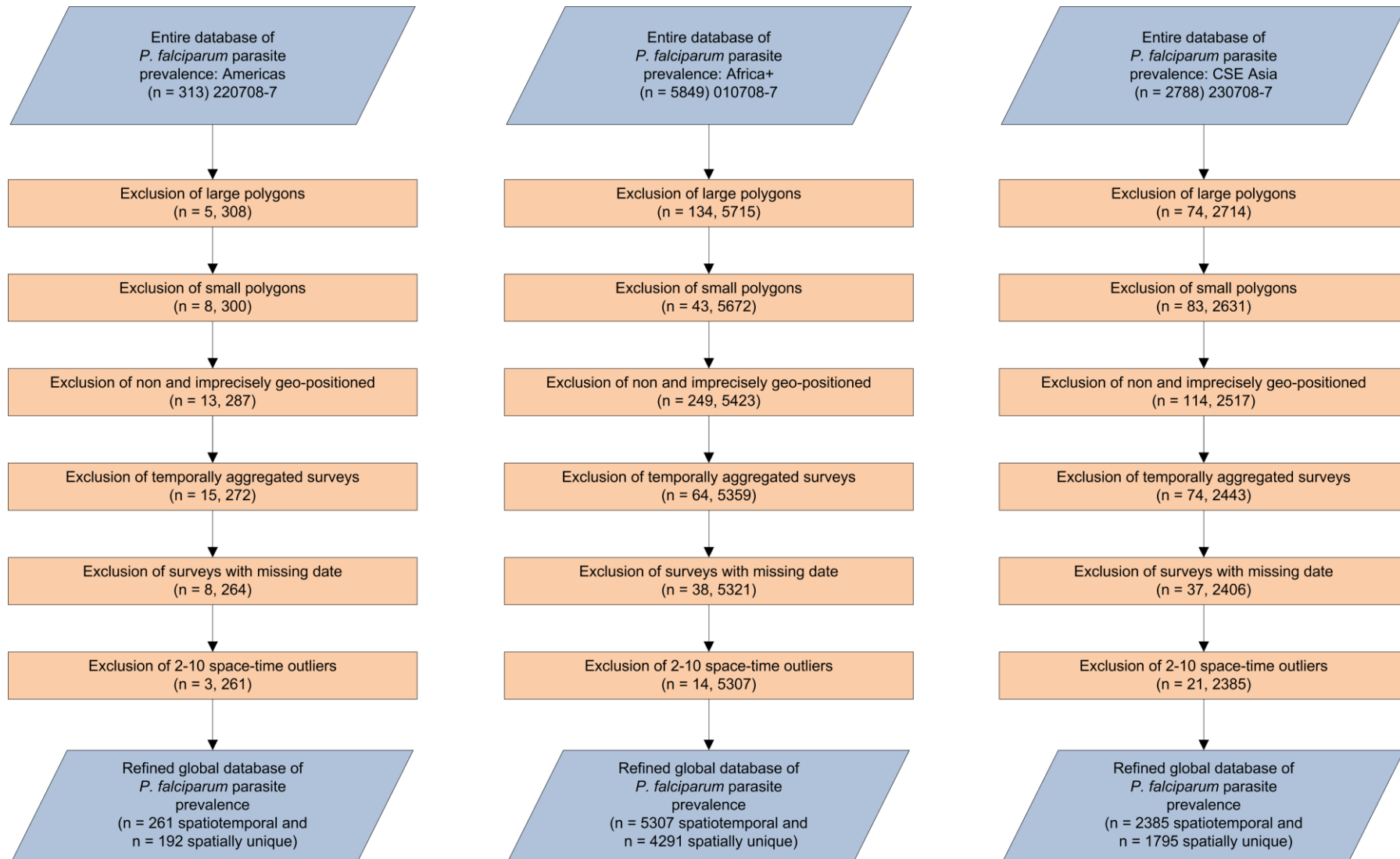
datum  $p(\mathbf{x}_i)$  itself. For each of the  $(n_i^2 - n_i)/2$  unique pairs of data  $\{(p(\mathbf{x}_j), p(\mathbf{x}_{j'}))\}$  within the regional neighbourhood the semi-variance  $\gamma_{j,j'} = 0.5(p(\mathbf{x}_j) - p(\mathbf{x}_{j'}))^2$  was computed along with the separation distance  $h_{j,j'}$  between each datum in the pair. Taken together, this set of  $(n_i^2 - n_i)/2$  pair-wise values of  $\gamma_{j,j'}$  and  $h_{j,j'}$ , calculated without reference to the datum  $p(\mathbf{x}_i)$ , captured the spatial autocorrelation structure within the regional neighbourhood of location  $\mathbf{x}_i$ . Semi-variances  $\gamma_{i,k}$  were then calculated between the datum  $p(\mathbf{x}_i)$  and its 20 nearest neighbours  $\{p(\mathbf{x}_k); k = 1, \dots, 20\}$ . Each of these 20 semi-variance values  $\gamma_{i,k}$  were then compared to regional semi-variance values  $\gamma_{j,j'}$  computed for pairs at similar separation distances, defined as within 0.1 decimal degrees (dd). This required the definition of the subset  $\gamma_{j,j'}^k = \{\gamma_{j,j'} : |h_{j,j'} - h_{i,k}| < 0.1\text{dd}\}$  for each of the semi-variance values  $\gamma_{i,k}$ . To avoid comparing data from substantially different time periods, both the local and regional semi-variance values were derived using only surveys carried out within a five year window centred on the survey at  $\mathbf{x}_i$ . The quantile position  $q_k$  of each semi-variance value  $\gamma_{i,k}$  within the corresponding subset  $\gamma_{j,j'}^k$  was recorded and the mean of the 20 quantile positions  $q_i = \frac{1}{20} \sum q_k$  was calculated.

The value  $q_i$  provides, for every datum  $p(\mathbf{x}_i)$ , an index of the extent to which the observed  $PfPR_{2-10}$  value differs from its neighbours relative to the extent expected given the distances to those neighbours, as defined by the regional semi-variances. Data for which the value  $q_i$  exceeded 0.95 were classed as outliers, and in this way a total of 38 points (America = 3, Africa+ = 14, CSE Asia = 21) were identified and removed. Details of these surveys will be made available on request. Following the implementation of this last data exclusion procedure, the final data set used for further modelling consisted of 7,953 data (America = 261, Africa+ = 5,307, CSE Asia = 2,385) of which (America = 192, Africa+ = 4,291, CSE Asia = 1,795) represented unique survey locations. The sequence of data inclusion rules and the number of data removed at each stage are summarised in Figure S1.2 and Table S1.1.

**Table S1.1.** Summary of data exclusions by region.

	America	Africa+	CSE Asia	Total
<b>Countries with <i>PfPR</i> survey data†</b>	<b>14</b>	<b>48</b>	<b>21</b>	<b>83</b>
<b>Total identified</b>	<b>313</b>	<b>5,849</b>	<b>2,788</b>	<b>8,950</b>
Large polygons	5	134	74	213
Small polygons	8	43	83	134
Unable to geo-locate	13	245	91	349
Imprecise geo-location	0	4	23	27
Exclusion of temporally aggregated surveys	15	64	74	153
Exclusion of surveys with missing month	8	38	37	83
Geostatistical space-time outliers	3	14	21	38
<b>Total surveys</b>	<b>261</b>	<b>5,307</b>	<b>2,385</b>	<b>7,953</b>
<b>Spatially unique surveys</b>	<b>192</b>	<b>4,291</b>	<b>1,795</b>	<b>6,278</b>

†Those countries in which *Plasmodium falciparum* parasite rate (*PfPR*) survey data were found are listed alphabetically by region: America (Bolivia, Brazil, Colombia, Costa Rica, Ecuador, French Guiana, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Peru, Suriname, Venezuela); Africa+ (Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Equatorial Guinea, Eritrea, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mauritania, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, São Tomé and Príncipe, Saudi Arabia, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Togo, Uganda, United Republic of Tanzania, Yemen, Zambia, Zimbabwe) and CSE Asia (Afghanistan, Cambodia, China, India, Indonesia, Iraq, Lao People's Democratic Republic, Malaysia, Myanmar, Nepal, Pakistan, Papua New Guinea, Philippines, Solomon Islands, Sri Lanka, Tajikistan, Thailand, Timor-Leste, Turkey, Vanuatu, Viet Nam).



**Figure S1.2:** Sequence of data exclusion rules for the formulation of a refined global *P. falciparum* parasite prevalence database for modelling. For each region the number excluded,  $n$ , and the number remaining are shown at each stage of exclusion.

## S1.6. Malariometric Survey Data Summary and Descriptive Statistics

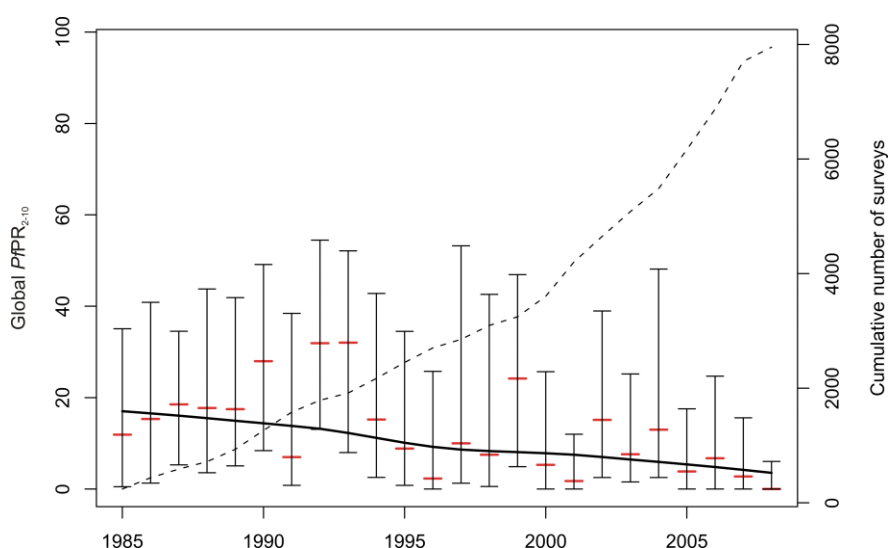
This section presents summary descriptions of some of the most relevant aspects of the *PfPR* database. Table S1.2 summarises these descriptions by region.

### Data and Geographic Coordinate Sources

The most productive source of information for the database, both primary *PfPR* data and geographic coordinates required for mapping, was direct communication with malaria specialists. Although peer-reviewed data sources were an important contributor in all regions, these represented between a sixth and a third of all the primary sources of *PfPR* data, depending on the region (Table S1.2). In terms of geographic coordinates, personal communication, including GPS, was by far the most important source for mapping survey data across the three regions, especially in the Africa+ and CSE Asia regions (Table S1.2).

### Time Periods Observed

Table S1.2 shows the frequency of *PfPR*<sub>2-10</sub> records by four time periods since 1985. In all three regions, particularly in Africa+ and CSE Asia, the frequency of surveys in the database conducted after the year 2000 increases. This is also illustrated by the increase in the gradient of the cumulative number of surveys by date (Figure S1.3). Furthermore a simple plot of the median *PfPR*<sub>2-10</sub> by year for the duration of the surveys (Figure S1.3) shows a clear secular movement of decreasing *PfPR*<sub>2-10</sub>. This has important implications for the analyses (See Protocol S3 and main text), affirming the need for a full space-time geostatistical model.



**Figure S1.3.** Median (red horizontal bars) and IQR (black horizontal bars) *PfPR*<sub>2-10</sub> per year with smooth fit line (continuous thick black) generated by a loess smoother. Also shown is the cumulative number of data available through time (dashed line).

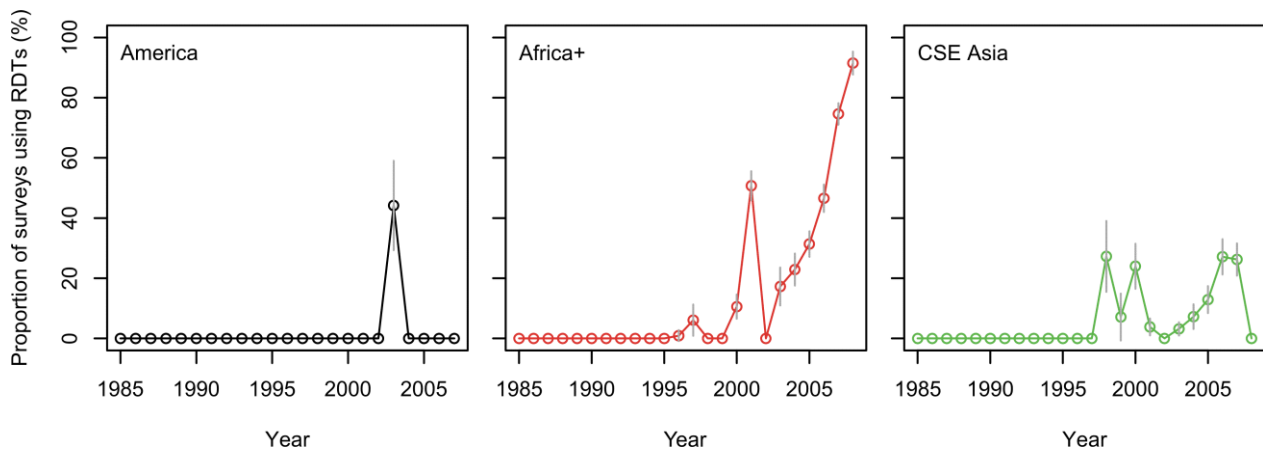


## Age Ranges Archived

The large number of age ranges observed in malariometric surveys required the standardisation to a single age group of 2-10 years (see Section 1.3). Table S1.2 presents the frequency of records by four upper age limits, given the difficulty in summarising all of the 311 age groups. In Africa+, the great majority of records corresponded to an age group of children of up to 15 years. In contrast, in America and CSE Asia, most of the sampling included adults.

## Diagnostic Methods

Microscopy was the dominant method of *P. falciparum* malaria diagnosis in the majority of surveys in the MAP database. Rapid diagnostic tests (RDTs), however, are increasingly used in large malariometric surveys, due to their relative ease of use and requirement of less training [19,20]. This was particularly evident in Africa+, where national malaria prevalence surveys using RDTs are increasingly common [21,22] (Figure S1.4).



**Figure S1.4.** Percentage of surveys using RDTs rather than microscopy by year for each region. Vertical bars are 95% binomial confidence intervals.

## Numbers Examined in Surveys

In Africa+, most sample sizes were between 50 and 100, with a median of 95, whereas in America and CSE Asia, the majority of sample sizes were between 101 and 500, with medians of 131 and 137, respectively (Table S1.2). The model (see Protocol S3 and main text) requires from each survey both the sample size and the number of individuals that are *P. falciparum* positive. Where *PfPR* was known, but the number of positive people was not stated explicitly, the latter was imputed using the assumed or known sample size and the *PfPR*. Authors sometimes also stated, for instance, that “approximately 200 individuals were sampled” and a *PfPR* for the community was reported. In these cases the approximate value for the number examined was accepted.

Despite requiring the number of blood samples examined to exceed 50, exceptions were made if this number was not specified but was strongly implied as being higher. For example, an author might state that all households of a community were sampled and the community is known to be considerably larger than 50 households. If no additional information was found, then the sample size was assumed conservatively as 50, thus maximizing the potential sampling error with their inclusion. This was the procedure for 147 records in the final database, mostly in Africa+ (Table S1.2).

**Table S1.2.** Summary figures of the most important aspects of the *PfPR*<sub>2-10</sub> data by region. Summaries are after all exclusions shown in Table S1.1.

	<b>America</b> <b>(n = 261)</b>	<b>Africa+</b> <b>(n = 5,307)</b>	<b>CSE Asia</b> <b>(n = 2,385)</b>	<b>Total</b> <b>(n = 7,953)</b>
<b>Primary source of <i>PfPR</i> data</b>				
Peer reviewed sources	102	888	492	1,482
Personal communication	113	2,735	1,497	4,345
Reports <sub>†</sub>	46	1,684	396	2,126
<b>Source of spatial coordinates</b>				
Personal communication	65	722	737	1,524
Encarta	88	1,396	230	1,714
GPS (through pers. comm.)	31	1,362	552	1,945
Combination	54	664	441	1,159
Other digital gazetteers	14	1,112	177	1,303
Paper source	8	14	11	33
Map	1	37	237	275
<b>Date range<sub>‡</sub></b>				
1985-1989	48	718	170	936
1990-1994	30	956	249	1,235
1995-1999	84	691	302	1,077
2000-2008	99	2,942	1,664	4,705

	America (n = 261)	Africa+ (n = 5,307)	CSE Asia (n = 2,385)	Total (n = 7,953)
<b>Upper age sampled</b>				
<=10	5	1,697	190	1,892
>10 and <=15	5	1,511	129	1,645
>15 and <=20	0	410	124	534
>20	251	1,689	1,942	3,882
<b>Total unique age ranges</b>	56	226	114	311‡
<b>Method</b>				
Microscopy	242	4,011	2,155	6,408
RDT	19	1,296	230	1,545
<b>Denominator</b>				
No denominator	0	138	9	147
50-100	104	2,818	771	3,693
101-500	124	2,078	1,404	3,606
>500	33	273	201	507
Median (IQR)	131 (77-259)	95 (56-164)	137 (88-216)	103 (63-190)

‡Ministry of Health reports, theses and other unpublished sources. †The end year of the survey was used to assign the date range. §These figures do not sum as some age ranges are repeated across regions.

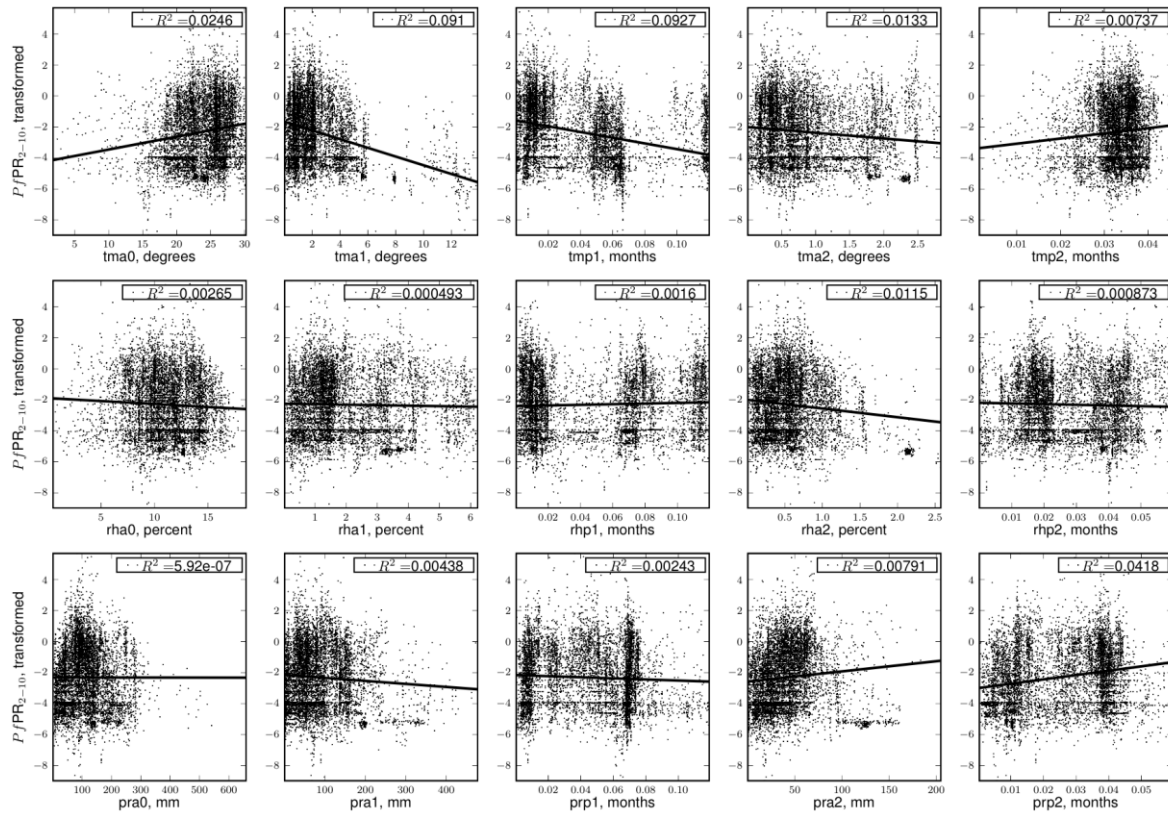
## S1.7. Relationships with Environmental Covariates

### Global Synoptic Climate Data

Global gridded climate data at 10 minute (18×18 km) spatial resolution, created through the interpolation of meteorological station data from 1961-1990 to form synoptic monthly images, were obtained [23]. From these data, the images representing monthly mean temperature, precipitation and relative humidity were extracted. These monthly images were then subject to temporal Fourier processing [24], and for each of the three climatic variables, the Fourier

variables representing the mean (amplitude 0), amplitude of the peak of the annual and biannual cycles (amplitudes 1 and 2), and timing of the peak of the annual and biannual cycles (phases 1 and 2), were obtained for analysis.

In each plot of the panel an environmental covariate from New *et al.* [23] is plotted against  $\text{logit}((PfPR_{2-10})/(N+2))$  (Figure S1.5). The argument of the logit function is the posterior expectation of the underlying  $PfPR$  in the 2-10 age group using a uniform prior. The logit function maps this value from the interval [0,1] to the real line.  $PfPR_{2-10}$  is estimated using the method of Smith *et al.* [14] rather than the Bayesian method described in Protocol S3. Each relationship is fit with the linear model and the variance explained marked in the top right of each plot. The wide scatter in the plots is evident, with the variance explained always below 10%.

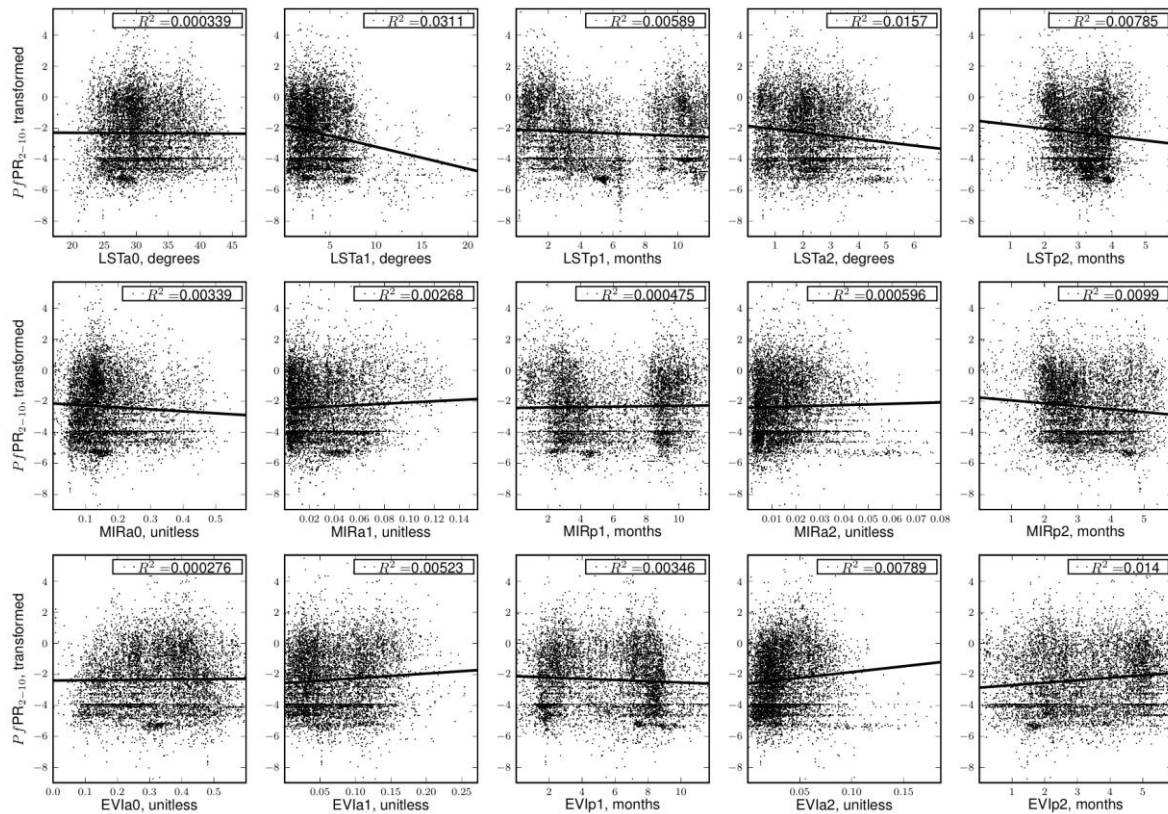


**Figure S1.5.** A panel of plots showing environmental covariates from New *et al.* [23] against  $\text{logit}((PfPR_{2-10})/(N+2))$ . On the x-axes, the first two letters indicate the environmental covariate: tm = temperature, rh = relative humidity, pr = precipitation. The second two letters indicate the temporal Fourier variable: a0 = mean, a1 = amplitude of the annual cycle, a2 = amplitude of the biannual cycle, p1 = phase of the annual cycle, p2 = phase of the biannual cycle. Temperature values are in degrees Celsius.

## Global MODIS Satellite Sensor Data

Global MODerate Resolution Imaging Spectrometer (MODIS) imagery from the Terra satellite [25] were acquired for the period 2001-2005 at 1×1 km spatial resolution. These were processed to create environmental variables that have proven valuable in the mapping of vector-borne disease distributions [24]: land surface temperature (LST), middle infrared reflectance (MIR) and enhanced vegetation index (EVI), an improved index over the widely used normalized difference vegetation index (NDVI). The 2001-2005 sequences of these variables were then subject to temporal Fourier processing, as described in Scharlemann *et al.* [26], and for each of the three environmental variables, the Fourier variables representing the mean (amplitude 0), amplitude of the peak of the annual and biannual cycles (amplitudes 1 and 2), and timing of the peak of the annual and biannual cycles (phases 1 and 2), were obtained for analysis.

In each plot of the panel an environmental covariate from Scharlemann *et al.* [26] is plotted against  $\text{logit}((P/PR_{2-10})/(N + 2))$  (Figure S1.6). The argument of the logit function is computed as in Figure S1.5. The wide scatter in the plots is again evident, with the variance explained always below 5%.



**Figure S1.6.** A panel of plots showing environmental covariates from Scharlemann *et al.* [26] against  $\text{logit}((P/PR_{2-10})/(N + 2))$ . On the x-axes, the first three letters indicate the environmental covariate: LST = land surface temperature, MIR = middle infrared reflectance, EVI = enhanced vegetation index. The following two letters indicate the temporal Fourier variable: a0 = mean, a1

= amplitude of the annual cycle,  $a_2$  = amplitude of the biannual cycle,  $p_1$  = phase of the annual cycle,  $p_2$  = phase of the biannual cycle. LST values are in degrees Celsius.

## References

1. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, et al. (2007) Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malar J* 6: 17.
2. Gregory RD, Blackburn TM (1991) Parasite prevalence and host sample size. *Parasitol Today* 7: 316-318.
3. Jovani R, Tella JL (2006) Parasite prevalence and sample size: misconceptions and solutions. *Trends Parasitol* 22: 214-218.
4. Sama W, Owusu-Agyei S, Felger I, Vounatsou P, Smith T (2005) An immigration-death model to estimate the duration of malaria infection when detectability of the parasite is imperfect. *Stat Med* 24: 3269-3288.
5. Smith DL, McKenzie FE, Snow RW, Hay SI (2007) Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol* 5: e42.
6. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med* 5: e38.
7. Baird JK (1995) Host age as a determinant of naturally acquired-immunity to *Plasmodium falciparum*. *Parasitol Today* 11: 105-111.
8. Baird JK, Jones TR, Danudirgo EW, Annis BA, Bangs MJ, et al. (1991) Age-dependent acquired protection against *Plasmodium falciparum* in people having two years exposure to hyperendemic malaria. *Am J Trop Med Hyg* 45: 65-76.
9. Baird JK, Purnomo, Basri H, Bangs MJ, Andersen EM, et al. (1993) Age-specific prevalence of *Plasmodium falciparum* among six populations with limited histories of exposure to endemic malaria. *Am J Trop Med Hyg* 49: 707-719.
10. Baird JK (1995) Host age as a determinant of naturally acquired immunity to *Plasmodium falciparum*. *Parasitol Today* 11: 105-111.
11. Baird JK, Jones TR, Danudirgo EW, Annis BA, Bangs MJ, et al. (1991) Age-dependent acquired protection against *Plasmodium falciparum* in people having two years exposure to hyperendemic malaria. *Am J Trop Med Hyg* 45: 65-76.
12. Baird JK, Purnomo, Basri H, Bangs MJ, Andersen EM, et al. (1993) Age-specific prevalence of *Plasmodium falciparum* among six populations with limited histories of exposure to endemic malaria. *Am J Trop Med Hyg* 49: 707-719.
13. Pull JH, Grab B (1974) Simple epidemiological model for evaluating malaria inoculation rate and risk of infection in infants. *Bull World Health Organ* 51: 507-516.

14. Smith DL, Guerra CA, Snow RW, Hay SI (2007) Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malar J* 6: 131.
15. Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics; Bickel P, Diggle P, Fienberg S, Gather U, Olkin I et al., editors. New York: Springer. 228 p.
16. Goovaerts P (1997) Geostatistics for natural resource evaluation. New York, U.S.A.: Oxford University Press. 483 p.
17. Macdonald G (1957) Local features of malaria. The epidemiology and control of malaria. London: Oxford University Press. pp. 63-99.
18. Mouchet J, Carnevale P, Coosemans M, Julvez J, Manguin S, et al. (2004) Paludisme et grandes régions biogéographiques. Biodiversité du paludisme dans le monde. Montrouge, France: John Libbey Eurotext.
19. Moody A (2002) Rapid diagnostic tests for malaria parasites. *Clin Microbiol Rev* 15: 66-78.
20. Murray CK, Gasser RA, Jr., Magill AJ, Miller RS (2008) Update on rapid diagnostic testing for malaria. *Clin Microbiol Rev* 21: 97-110.
21. Kleinschmidt I, Sharp B, Benavente LE, Schwabe C, Torrez M, et al. (2006) Reduction in infection with *Plasmodium falciparum* one year after the introduction of malaria control interventions on Bioko Island, Equatorial Guinea. *Am J Trop Med Hyg* 74: 972-978.
22. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, et al. (2008) Spatial prediction of *Plasmodium falciparum* prevalence in Somalia. *Malar J* 7: 159.
23. New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21: 1-25.
24. Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ (2006) Global environmental data for mapping infectious disease distribution. *Adv Parasitol* 62: 37-77.
25. Tatem AJ, Goetz SJ, Hay SI (2004) Terra and Aqua: new data for epidemiology and public health. *Int J Appl Earth Obs Geoinformation* 6: 33-46.
26. Scharlemann JPW, Benz D, Hay SI, Purse BV, Tatem AJ, et al. (2008) Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS One* 3: e1408.

## Protocol S2. Demographic Databases and Procedures

A brief description of the steps taken to define urban/peri-urban/rural status for the environments in which age-standardized *Plasmodium falciparum* parasite rate surveys,  $PfPR_{2-10}$ , (see Protocol S1) were undertaken and the impact of such status on prevalence is presented here. Full details on the exploratory analyses and experiments undertaken are provided elsewhere [1].

### S2.1 Parasite Rate Survey Urban/Peri-Urban/Rural Classification Rules

In the age-standardized  $PfPR_{2-10}$  database (up to and including Protocol S1.3), 37.8% of surveys include an author-defined urban or rural assignment (617 urban, 2400 rural). For Africa, Yemen and Saudi Arabia (hereafter Africa+), significantly lower  $PfPR_{2-10}$  values were observed in author-defined urban areas compared to rural areas when spatially (within 100 km) and temporally (within five years) proximate survey pairs of urban and rural  $PfPR_{2-10}$  were compared [1]. No significant differences were observed for urban-rural survey pairs elsewhere in the *P. falciparum* malaria endemic world. The ability to use this information in further analyses required defining the urban-rural status for all surveys and exploring the significance of this relationship for the entire  $PfPR_{2-10}$  database.

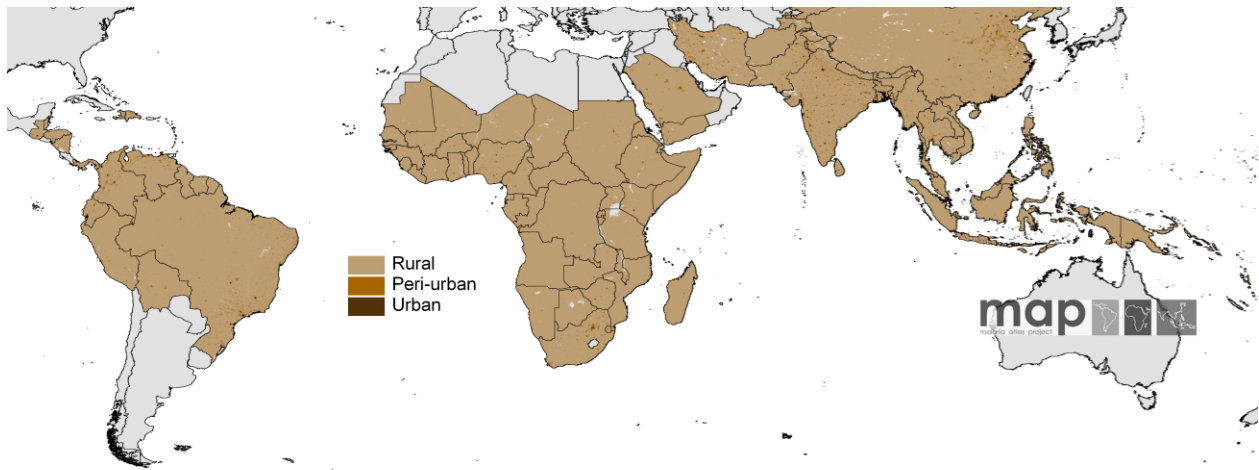
Author definitions were assumed to be the most epidemiologically pertinent definitions of urban and rural status for these surveys. These definitions were used to test which of several global urban extent maps available in the public-domain matched these author-defined definitions most accurately. The best surface would then provide the basis for an uniform urban-rural assignment of all  $PfPR_{2-10}$  surveys.

The results showed that the urban extent mask created as part of the Global Rural Urban Mapping Project (GRUMP) [2] mapped most accurately the author-defined urban/rural designations of  $PfPR_{2-10}$  surveys [1]. These were tested using confusion matrix-based accuracy measures, including percentage correct, user's and producer's accuracies and the kappa coefficient of agreement [3,4]. The GRUMP urban extent mask (GRUMP-UE) mapped 88.73% of author-defined designations correctly, with a kappa value of 0.664.

The GRUMP-UE surface is created principally using night-time lights satellite imagery, supplemented with data derived from tactical pilotage charts and known settlement points [2,5,6]. Through using satellite night-time lights as the basis for mapping urban areas, GRUMP-UE has been shown to overestimate urban extent due to the "overflow" effects seen in such imagery [7], resulting in the inclusion of less intensely-urban "peri-urban" areas. An approach was therefore required to identify the intensely urban areas within the urban extents defined by GRUMP-UE. Previous work has shown that identifying areas where population density is greater than or equal to 1000 people per km<sup>2</sup> maps a malaria-relevant divide between urban and peri-urban malaria transmission in Africa [8]. This definition was used and implemented with the Gridded Population of the World version 3 (GPW3) [2] population density database for 2000, projected to 2007 using standard techniques (see Section S2.4).

The GPW3 human population density surface is derived from the most recent available national census and other demographic data, resolved at the highest possible administrative boundary level, and areal weighted [9] to a 5×5 km spatial resolution grid. Those areas where GPW3 population densities were greater than or equal to 1000 people per km<sup>2</sup>, and within urban extents defined by GRUMP-UE, were mapped as "urban". Those areas within GRUMP-UE extents with GPW3 population densities lower than 1000 people per km<sup>2</sup> were mapped as "peri-urban". The final urban and peri-urban extents are shown in Figure S.2.1.





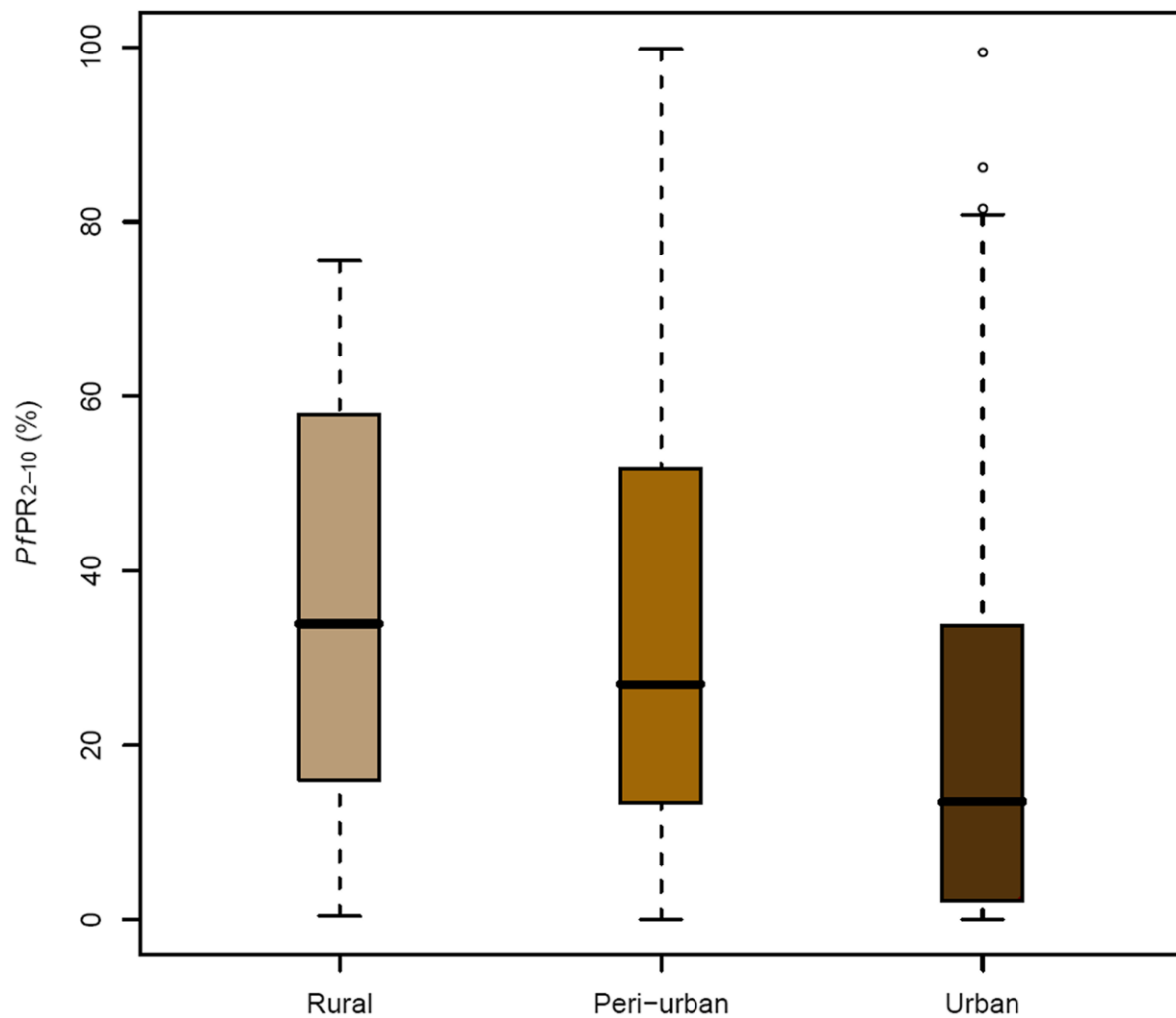
**Figure S.2.1.** Urban, peri-urban and rural extents for all malaria endemic countries. Full details of their derivation are provided in the text

## S2.2 Urban/Peri-Urban/Rural Status and Prevalence

The  $PfPR_{2-10}$  points were overlaid on the new urban classified map and then re-assigned as urban ( $n = 334$ ), peri-urban ( $n = 541$ ) or rural ( $n = 7052$ ). This changed the author-defined assignment of “rural” to peri-urban in 126 cases and urban in 21 cases, while changing the author-defined assignment of “urban” to rural for 193 surveys and peri-urban for 196. A reanalysis of the urban/peri-urban/rural status and  $PfPR_{2-10}$  was then conducted on the complete  $PfPR_{2-10}$  database.

All sets of surveys mapped as urban, peri-urban and rural within 100 km and ten years of each other were identified, and their  $PfPR_{2-10}$  values plotted. Where multiple surveys existed for the same location, the survey closest in time to the counterpart urban/peri-urban/rural survey was used. No sets of such points existed for the America region, while just six were available for analyses in CSE Asia (Figure S2.2).

Tests on neighbouring “urban-peri-urban”, “urban-rural” and “rural-peri-urban” sets of  $PfPR_{2-10}$  surveys consistently showed transmission differences in Africa+ (significant at the  $p < 0.001$  level using the Friedman’s ranking test [10]) (Figure S2.2). No significant differences were found for the CSE Asia region. These urban, peri-urban and rural classifications were then made available to the model based geostatistical approaches described in Protocol S3.



**Figure S2.2.** A boxplot showing the differences in average  $PfPR_{2-10}$  between GRUMP-UE/GPW3-defined rural, peri-urban and urban areas using 145 sets of spatially and temporally proximate malariometric survey pairs for Africa+ [1]. The central bar is the median (rural  $PfPR_{2-10}$  = 33.9%, peri-urban  $PfPR_{2-10}$  = 26.9%, and urban  $PfPR_{2-10}$  = 13.5%), the box defines the positions of the first and third quartiles of the distribution, while the whiskers extend to include all points no more than 1.5 times the inter-quartile range defined by the boxes. Points beyond this range are marked as outliers (the circles).

### S2.3 GRUMP alpha Human Population Surface

The GRUMP spatial population database provides gridded population counts and population density estimates for the years 1990, 1995 and 2000, both adjusted and unadjusted to the United Nations' national population estimates. The data and methods used to construct GRUMP are described in detail elsewhere [2,11]. In brief, the initial procedures are identical to those of GPW3 except that the areal weighting [9] is implemented on a 1×1 km spatial resolution grid. The urban extent mask, GRUMP-UE (see section S2.1), is then used to adjust population numbers within each extent to match estimated totals for each settlement in question. Finally, the individual national population surfaces are projected to the years 1990, 1995 and 2000 and adjusted to match the national population totals estimated by the United Nations' Population

Division [12]. The clearly documented methods, the substantially larger number of administrative units used to create the database, and tests showing its higher accuracy over other products [9,13], have led to the adoption of GRUMP products within the Malaria Atlas Project (MAP, <http://www.map.ox.ac.uk>).

## S2.4 Population and Area at Risk Derivation

GRUMP UN-adjusted population counts were obtained for the year 2000. This population surface was projected to 2007 by applying national, medium variant, inter-censal growth rates by country [12]. An estimate of population in 2007 was produced using the following equation;  $P_{2007} = P_x e^{rt}$  where  $P_{2007}$  is the required 2007 population within a pixel,  $P_x$  is the population within the same pixel at year  $x$ ,  $t$  is the number of years between year  $x$  and 2007, and  $r$  is the average growth rate [14]. This resulted in a population count surface of 1×1 km spatial resolution for 2007. These population counts were then stratified nationally by age group using UN defined age structures [12] to obtain under 5 years, 5-14 years and 15+ years population count surfaces.

By overlaying the 5×5 km predicted endemicity classes on each of the 2007 population count surfaces, populations at risk in each class were estimated globally. The quantification of areas within each endemicity class was undertaken by first reprojecting the predicted endemicity class map from geographic to Mollweide equal area projection in ArcGIS 9.2 (ESRI, 2006). The areas covered by each class were then calculated in km<sup>2</sup>.

## References

1. Tatem AJ, Guerra CA, Kabaria CW, Noor AM, Hay SI (2008) Human population, urban settlement patterns and their impact *Plasmodium falciparum* malaria endemicity. *Malar J* 7: 218.
2. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, et al. (2006) Determining global population distribution: methods, applications and data. *Adv Parasitol* 62: 119-156.
3. Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 37: 35-46.
4. Congalton RG, Green K (1993) A practical look at the sources of confusion in error matrix generation. *Photogramm Eng Remote Sens* 59: 641-644.
5. Elvidge CD, Baugh KE, Dietz JB, Bland T, Sutton PC, et al. (1999) Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sens Environ* 68: 77-88.
6. Elvidge CD, Hobson VR, Nelson IL, Safran JM, Tuttle BT, et al. (2003) Overview of DMSP OLS and scope of applications. In: Mesev V, editor. *Remotely Sensed Cities*. London: Taylor and Francis. pp. 281-333.
7. Tatem AJ, Noor AM, Hay SI (2005) Assessing the accuracy of satellite derived global and national urban maps in Kenya. *Remote Sens Environ* 96: 87-97.
8. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW (2005) Urbanization, malaria transmission and disease burden in Africa. *Nat Rev Microbiol* 3: 81-90.
9. Hay SI, Noor AM, Nelson A, Tatem AJ (2005) The accuracy of human population maps for public health application. *Trop Med Int Health* 10: 1073-1086.
10. Friedman M (1940) A comparison of alternative tests of significance for the problem of  $m$  rankings. *Ann Math Stat* 11: 86-92.
11. CIESIN/IFPRI/WB/CIAT (2007) Global Rural Urban Mapping Project (GRUMP) alpha: Gridded Population of the World, version 2, with urban reallocation (GPW-UR). Available at <http://sedac.ciesin.columbia.edu/gpw>. Palisades, New York, USA: Center for

International Earth Science Information Network, Columbia University / International Food Policy Research Institute / The World Bank / and Centro Internacional de Agricultura Tropical.

12. U.N.P.D. (2006) World population prospects: the 2006 revision population database. <http://esa.un.org/unpp/>. New York: United Nations Population Division (U.N.D.P).
13. Tatem AJ, Noor AM, von Hagen C, Di Gregorio A, Hay SI (2007) High resolution population maps for low income nations: combining land cover and census in East Africa. PLoS One 2: e1298.
14. Deichmann U (1996) A review of spatial population database design and modelling. Santa Barbara, California, U.S.A.: National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara (UCSB).

## **Protocol S3. Model-based geostatistical procedures**

### **S3.1 Overview of the Statistical Model**

#### **Bayesian Inference**

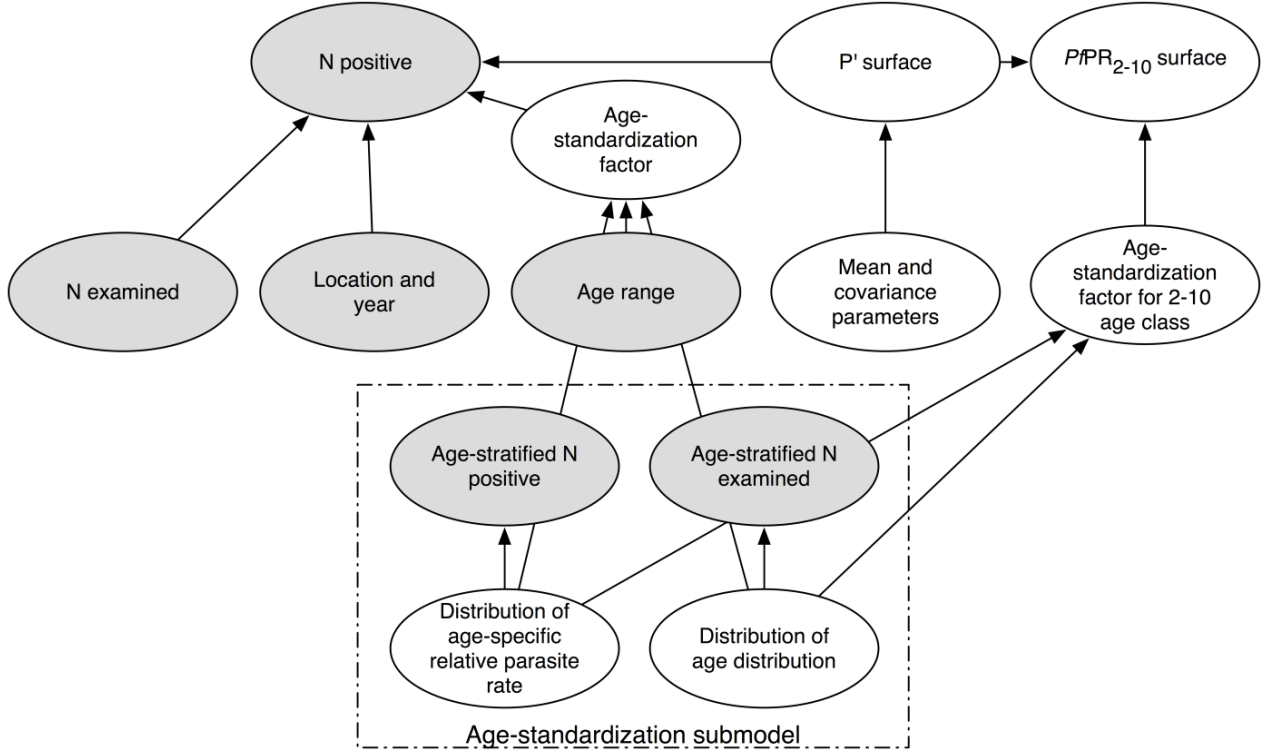
In Bayesian model-based geostatistics, conversion of geo-referenced data to a continuous surface begins with specification of a probability model. Such models are usually collections of interconnected conditional probability statements, for example, “given the number of individuals participating in a sample and the population-wide *Plasmodium falciparum* parasite rate (*PfPR*), the number of positive individuals in that sample is distributed binomially.” These statements, including “prior” probability distributions for basic model parameters, should represent the modeller’s understanding of the relationship between the inferential targets (the *PfPR* surface) and the data (the sample outcomes). The most broadly applicable class of methods for fitting Bayesian probability models is Markov chain Monte Carlo, or MCMC [1,2].

#### **Model Overview**

The probability model used in the current study assumed that individuals participating in each sample (*PfPR* survey) were *P. falciparum* positive with a probability that was the product of two components: (i) a continuous function of the time and location of the survey and (ii) a factor that depended on the age range of individuals included in the survey. The continuous function of time and location was modelled as a Gaussian process or Gaussian random field [3]. The distributions of the age-standardization factors were modelled using a Bayesian version of the procedure described by Smith *et al.* [4], and inferred based on detailed age-stratified information reported by a small number of surveys specified in that study.

This section first presents a schematic graphical representation of the full model (Figure S3.1). Such representations are helpful for visualizing complicated probability models. The main probability model is then presented formally. Section S3.2 gives a discussion of the specifications of some prior distributions; S3.3 discusses the age-standardization sub-model;

S3.4 describes the Markov chain Monte Carlo algorithm used to fit the model; and S3.5 describes the conversion of the posterior predictive distribution to imagery.



**Figure S3.1** A schematic of the probability model, expressed as a directed acyclic graph. Ovals represent variables in the model. Grey ovals represent variables that have been observed. Arrows indicate conditional distributions written down in the model. For example, the distribution of the number positive in a parasite rate survey is specified based on the corresponding age-standardization factor, the underlying  $PfPR$  surface and the time and location of the survey. It is possible to fit the age-standardization sub-model separately with minimal inconsistency. The maps presented in this study are summaries of the posterior of the upper right-hand node, the  $PfPR_{2-10}$  surface.

### Formal Presentation of Model

Each of the  $N_i$  individuals in sample  $i$  was assumed *P. falciparum* positive with probability  $\tilde{k}_i P'(x_i, t_i)$ , so the number positive  $N_i^+$  was distributed binomially:

$$N_i^+ | N_i, P'(x_i, t_i) \stackrel{\text{ind}}{\sim} \text{Bin}(N_i, \tilde{k}_i P'(x_i, t_i)) \quad (\text{S3.1})$$

The coefficient  $P'(x_i, t_i)$  was modelled as a Gaussian process. The factor  $\tilde{k}_i$  converted  $P'(x_i, t_i)$  to the probability that individuals within the age range reported for study  $i$  were *P. falciparum* positive, and that the infection was detected, thereby accounting for the influence of age on the

probability of detection [4]. The age-standardization factor  $\tilde{k}_i$  in each population was assumed drawn independently from a distribution  $D_{\tilde{k}}$  whose parameters were the lower  $A_{L,i}$  and upper  $A_{U,i}$  ages reported in study  $i$ :

$$\tilde{k}_i | A_{U,i}, A_{L,i} \stackrel{\text{ind}}{\sim} D_{\tilde{k}}(A_{U,i}, A_{L,i}) \quad (\text{S3.2})$$

The form of  $D_{\tilde{k}}$  is described in section S3.3.

$PfPR_{2-10}$  is the *P. falciparum* parasite rate for individuals between ages 2 (2.00) and 10 (9.99). Its value at an arbitrary location  $x$  and time  $t$  is the product of  $P'(x, t)$  and another age-standardization factor,  $k_{2-10}$ , distributed as  $D_k(2, 10)$ :

$$\begin{aligned} PR_{2-10}(x, t) &= P'(x, t) k_{2-10}(x, t) \\ k_{2-10}(x, t) &\stackrel{\text{ind}}{\sim} D_k(2, 10) \end{aligned} \quad (\text{S3.3})$$

The factor  $k_{2-10}$  converted  $P'(x, t)$  to the probability that individuals between ages 2 and 10 at location  $x$  are *P. falciparum* positive. The age-standardization factor  $\tilde{k}$  of a survey is the product of the age-standardization factor  $k$  associated with the same place, time and age range and the sensitivity of the survey.

The coefficient  $P'(x, t)$  at arbitrary location  $x$  and time  $t$  was modelled as the inverse-logit function applied to a random field  $f$  evaluated at  $(x, t)$ , plus an unstructured (random) component  $\epsilon(x, t)$ .

$$P'(x, t) = \text{logit}^{-1}(f(x, t) + \epsilon(x, t)) \quad (\text{S3.4})$$

The components  $\epsilon(x, t)$  were assumed independent and identically distributed for each location  $x$  and time  $t$  and a standard diffuse but proper prior with expectation 0.25 was assigned to their variance  $V$ .

$$\epsilon(x, t) | V \stackrel{\text{iid}}{\sim} N(0, V) \quad (\text{S3.5})$$

$$1/V \sim \text{Gamma}(.001, .004) \quad (\text{S3.6})$$

The random field  $f$  was modelled as a Gaussian process whose mean was a linear function of time  $\mu = \beta_x + \beta_t t + \beta_u 1_u(x) + \beta_p 1_p(x)$  and whose covariance was  $C$ :

$$\begin{aligned}
& f(x, t) | \beta_x, \beta_t, \beta_u, \beta_p, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, v \\
& \sim \text{GP}(\beta_x + \beta_t t + \beta_u 1_u(x) + \beta_p 1_p(x), C)
\end{aligned} \tag{S3.7}$$

The variables  $1_u(x)$  and  $1_p(x)$  indicated whether  $x$  was urban or peri-urban, respectively. The covariance of the field was modelled using a version of the spatiotemporal covariance function recently recommended by Stein [5] (equation 12):

$$\begin{aligned}
C(x_i, t_i; x_j, t_j) &= \tau^2 \gamma(0) \frac{(\Delta x)^{\gamma(\Delta t)} K_{\gamma(\Delta t)}(\Delta x)}{2^{\gamma(\Delta t)-1} \Gamma(\gamma(\Delta t)+1)}, \\
\gamma(\Delta t) &= \frac{1}{2\rho+2(1-\rho) \left[ (1-v)e^{-|\Delta t|/\phi_t} + v \cos(2\pi\Delta t) \right]}, \\
\Delta t &= |t_i - t_j|
\end{aligned} \tag{S3.8}$$

$K_\gamma$  is the modified Bessel function of the second kind of order  $\gamma$ , and  $\Gamma$  is the gamma function [6,7].

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  $x_j$ .  $\theta$  was computed as if latitude and longitude were Euclidean coordinates (on a cylindrical projection):

$$\Delta x = 2\sqrt{\gamma(\Delta t)} \frac{D_{GC}(x_i, x_j) \sqrt{1 - \psi^2 \cos^2(\theta(x_i, x_j) - \lambda)}}{\phi_x} \tag{S3.9}$$

Computing distance in this way allows for anisotropy.

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

$$\rho + (1 - \rho) \left[ (1 - v)e^{-|\Delta t|/\phi_t} + v \cos(2\pi\Delta t) \right] \tag{S3.10}$$

As temporal separation increases, the covariance approaches a limiting sinusoid  $\tau^2[\rho + (1 - \rho)v \cos(2\pi\Delta t)]$  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 [5].

The square root of the partial sill  $\tau$  and the spatial range parameter  $\phi_x$  were assigned skew-normal priors:

$$\log \tau | \mu_\tau, V_\tau, \alpha_\tau \sim \text{Skew-Normal}(\mu_\tau, V_\tau, \alpha_\tau) \quad (\text{S3.11})$$

$$\log \phi_x | \mu_\phi, V_\phi, \alpha_\phi \sim \text{Skew-Normal}(\mu_\phi, V_\phi, \alpha_\phi) \quad (\text{S3.12})$$

and their specification is described further in section S3.2.

The standard “one-over- $x$ ” prior for the temporal scale parameter  $\phi_t$  resulted in collapse to zero, a common artefact when data do not contain strong information. A relatively vague but proper prior, which has an expectation of ten years, was used instead.

$$\phi_t \sim \text{Exponential}(0, .1) \quad (\text{S3.13})$$

A uniform prior was assigned to the direction of anisotropy parameter  $\lambda$  and to the square of the “eccentricity” parameter  $\psi$ , which controls the amount of anisotropy,

$$\lambda \sim \text{Uniform}(0, \pi) \quad (\text{S3.14})$$

$$\psi^2 \sim \text{Uniform}(0, 1) \quad (\text{S3.15})$$

a uniform prior was assigned to the limiting autocorrelation in the temporal direction,

$$\rho \sim \text{Uniform}(0, 1) \quad (\text{S3.16})$$

and a standard prior was assigned to the components of the mean:

$$p(\beta_x, \beta_t, \beta_u, \beta_p) \propto 1 \quad (\text{S3.17})$$

### S3.2 Prior Specification

Although standard priors such as the improper “flat” prior [1] were assigned to most of the basic model parameters, subjective skew-normal priors [8] were specified for the range and partial sill parameters  $\tau$  and  $\phi_x$ , because MCMC mixing time appeared to be bottlenecked by correlations involving these parameters and because their effects on the  $PfPR_{2-10}$  surface were relatively easy to visualize. A program written in the R language [9] (available on request) was prepared to help those authors with a more extensive experience of malaria geography to visualize the parameters’ effects. The required inputs were values for  $\tau$  and  $\phi_x$ , as well as a handful of hypothetical surveys (latitude, longitude, number examined and number positive). The outputs were realizations from, and quantile surfaces for, the approximate posterior predictive



distribution of  $PfPR_{2-10}$  over a small geographic area. The authors (PWG, SIH) who provided the priors preferred to express their opinions as triples (lower 95% credible interval, mode, upper 95% credible interval), which were converted to skew-normal priors [8] on the log scale. A table of these triples, with the corresponding parameters of the log-skew-normal distribution, is shown below. The units of  $\phi_x$  are radians on the Earth's surface.

	lower 95%	mode	upper 95%	$\mu$	$V$	$\alpha$
Africa+, $\phi_x$	1	2	6	0.0535	0.559	3.21
America, $\phi_x$	0.1	1	4	0.607	1.24	-1.17
CSE Asia, $\phi_x$	1	2	6	0.0535	0.559	3.21
Africa+, $\tau$	0.0157	0.0784	0.392	-2.54	0.704	-0.0150
America, $\tau$	0.0157	0.0784	0.470	-2.58	0.741	0.0498
CSE Asia, $\tau$	0.00784	0.0470	0.157	-2.97	0.571	-0.143

**Table S3.1** The prior modes and credible intervals specified for the spatial range  $\phi_x$  and the spatial partial sill  $\tau$  in the three regions, and the corresponding parameters of the log-scale skew-normal prior [8].

### S3.3 Age-Standardization

#### The Age-Standardization Model

This section explains in more detail the age-standardization model that underpins the distributions  $D_k$  and  $D_{\tilde{k}}$  of the age-standardization factors  $k$  and  $\tilde{k}$ . Following Smith *et al.* [4], the functional form of Pull and Grab [10] was used to model the probability that an individual of age  $A$  is *P. falciparum* positive:

$$P(A) = P' [1 - e^{-bA}] , \quad (\text{S3.18})$$

which is increasing. In addition, the age dependent probability of detection of a *P. falciparum* infection (the sensitivity) was modelled as:

$$F(A; \alpha, s, c) = \begin{cases} 1, & A < \alpha \\ 1 - s [1 - e^{-c(A-\alpha)}] , & A \geq \alpha \end{cases} , \quad (\text{S3.19})$$

which is decreasing for  $A \geq \alpha$ . The resulting model for the probability that a *P. falciparum* infection would be detected in an individual drawn from the population within age range  $[A_L, A_U)$  is:

$$P' \tilde{k} = \frac{\sum_{A=A_L}^{A_U-1} P(A; b, P') F(A, \alpha, s, c) S(A)}{\sum_{A=A_L}^{A_U-1} S(A)}, \quad (\text{S3.20})$$

where  $S$  is the age distribution of the study participants. Some allowance for population-to-population variation was introduced into the age-standardization procedure of Smith *et al.* [4]. The age distribution  $S$  was assumed to be drawn from some probability distribution  $p(S)$ . Similarly,  $\alpha$ ,  $s$ ,  $c$  and  $b$  were assumed to be drawn from a probability distribution  $p(\alpha, s, c, b)$ . Further, the ages of participants in study  $i$  were assumed to be drawn uniformly from the population's age distribution within the published age limits  $A_{L,i}$  and  $A_{U,i}$ . The distributions  $p(S)$  and  $p(\alpha, s, c, b)$ , if they are known, can be converted into the probability distributions  $D_k(2, 10)$  and  $D_{\tilde{k}}(A_{L,i}, A_{U,i})$  required in the previous section.

### Adaptation of the Smith *et al.* Model

The sub-model for age-standardization used in this study is presented formally below. In the following,  $i$  indexes populations (within the training set) and  $j$  indexes age bins. Each individual in population  $i$  and age class  $j$  was assumed *P. falciparum* positive with probability given by  $p_i(A_{L,j}, A_{U,j})$ , where  $A_{L,j}$  and  $A_{U,j}$  bound age class  $j$ :

$$N_{i,j}^+ | N_{i,j}, p_i(A_{L,j}, A_{U,j}) \stackrel{\text{ind}}{\sim} \text{Bin}(N_{i,j}, p_i(A_{L,j}, A_{U,j})) \quad (\text{S3.21})$$

The probability that an individual in population  $i$  and age class  $j$  was *P. falciparum* positive was modelled as a function of the local detection-probability parameters  $\alpha_i$ ,  $s_i$  and  $c_i$ , and epidemiological parameters  $P'_i$  and  $b_i$ , as well as the age distribution  $\tilde{S}_i$  of study participants using the Smith *et al.* model for age dependence:

$$\begin{aligned} p_i(A_{L,j}, A_{U,j}) | P'_i, b_i, \alpha_i, s_i, c_i, \tilde{S}_i &= P'_i \tilde{k}_{ij} \\ &= \frac{\sum_{A=A_{L,j}}^{A_{U,j}} P(A; b_i, P'_i) F(A, \alpha_i, s_i, c_i) \tilde{S}_i(A)}{\sum_{A=A_{L,j}}^{A_{U,j}} \tilde{S}_i(A)} \end{aligned} \quad (\text{S3.22})$$

The age distribution  $\tilde{S}_i$  of study participants was assumed to have been generated by randomly drawing study participants from the population-wide age distribution  $S_i$ :

$$\tilde{S}_i | S_i \stackrel{\text{ind}}{\sim} \text{Multinomial}(N_i, S_i) \quad (\text{S3.23})$$

The age distribution  $S_i$  in each population was assumed drawn from a common Dirichlet distribution [4]. Dirichlet random variables are discrete probability distributions: positive vectors that sum to one. The distribution was “centered” on a typical age distribution  $S_0$ :

$$S_i | S_0, \nu \stackrel{\text{iid}}{\sim} \text{Dirichlet}(\nu S_0) \quad (\text{S3.24})$$

The parameter  $\nu$  controlled the extent to which  $S_i$  deviates from  $S_0$  such that large values of  $\nu$  resulted in small deviations. A Dirichlet prior was assigned to the typical age distribution  $S_0$ :

$$S_0 | \theta_0 \sim \text{Dirichlet}(\theta_0) \quad (\text{S3.25})$$

The epidemiological and detection-probability parameters, except  $P'$ , in each population were assumed drawn from a common distribution.  $P'$  in each population was modelled as independent.

$$\{\log(\alpha_i), 1/\log(c_i), 1/\log(b_i), \text{logit}(s_i)\} | \mu_A, C_A \stackrel{\text{iid}}{\sim} N(\mu_A, C_A) \quad (\text{S3.26})$$

$$P'_i \stackrel{\text{iid}}{\sim} \text{Uniform}(0, 1) \quad (\text{S3.27})$$

This assumption is inconsistent with the main spatial model, but it allows the age-standardization model to be fitted separately. Separating a single large model into several smaller models is usually computationally advantageous when MCMC is used for fitting.

A standard prior was used for the mean of the transformed epidemiological and detection-probability parameters' distribution:

$$p(\mu_A) \propto 1 \quad (\text{S3.28})$$

Following Gelman *et al.* [1], the covariance matrix  $C_A$  was factored into the vector of marginal standard deviations  $\sigma$  and the correlation matrix  $R$ . A standard prior was used for the marginal standard deviations  $\sigma$  and for the off-diagonal elements of the upper triangle of  $R$ , subject to the constraint that  $R$  must be positive definite:

$$C_A = \text{diag}(\sigma) R \text{diag}(\sigma) \quad (\text{S3.29})$$

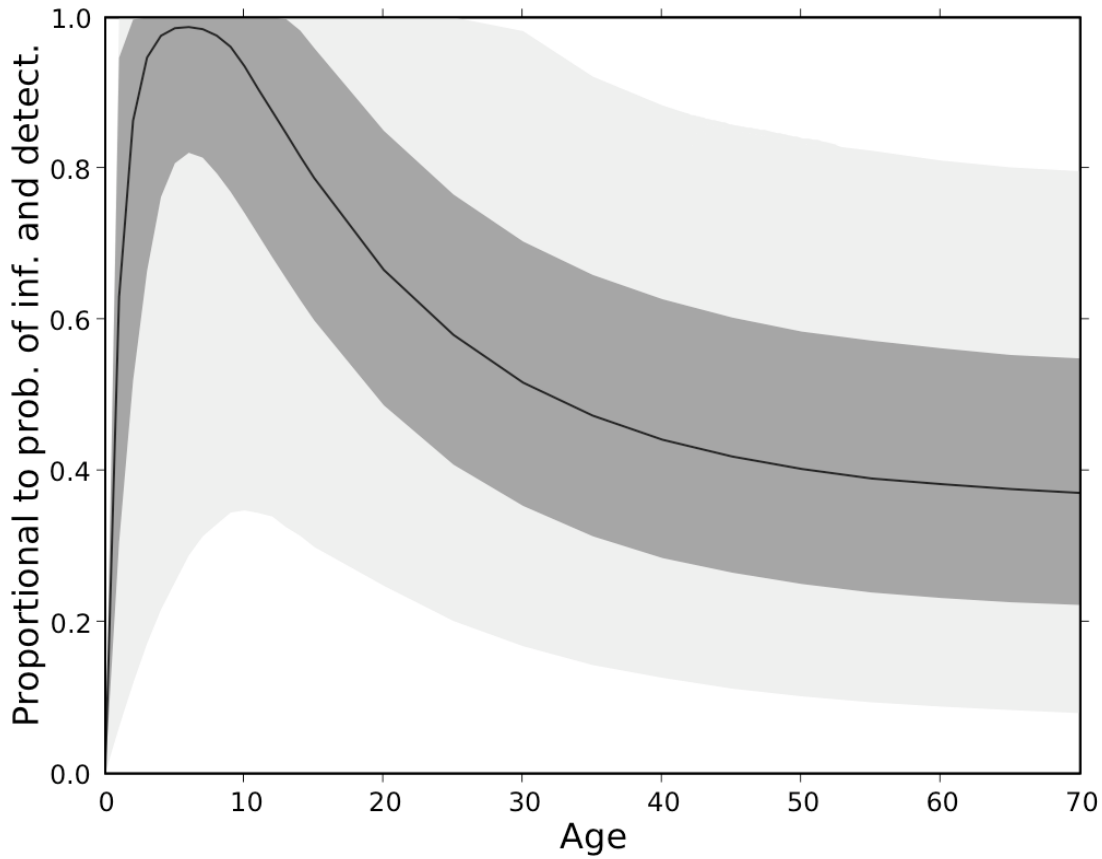
$$p(\sigma) \propto \prod_{i=1}^4 1/\sigma_i \quad (\text{S3.30})$$

$$p(R) \propto 1_R \text{ symmetric positive definite} \quad (\text{S3.31})$$

A standard prior was also assigned to  $\nu$ , the parameter controlling the concentration of the distribution of the  $S_i$ 's.

$$p(\nu) \propto 1/\nu \quad (\text{S3.32})$$

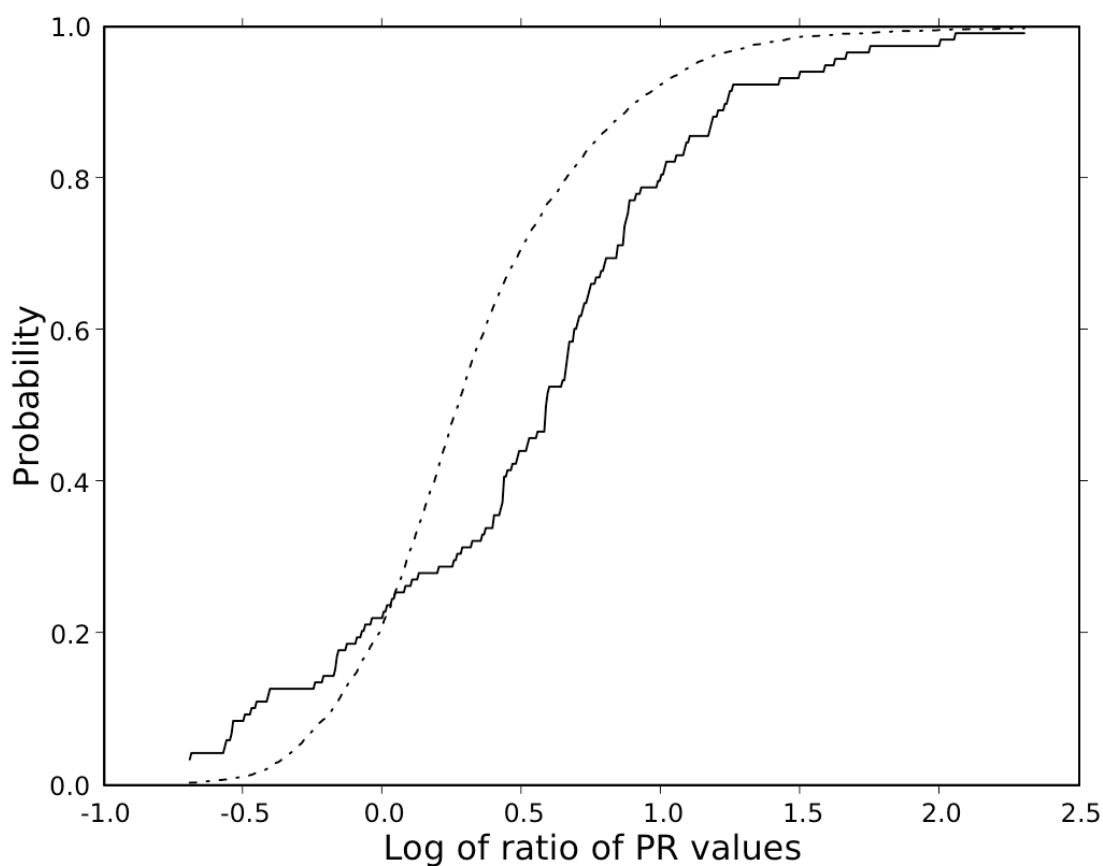
The posterior predictive distribution of  $P(x, x)/P'$  (the predictive distribution of  $PfPR$  in a single age bin given that  $P'$  is equal to 1) is shown in the figure S3.2.



**Figure S3.2** The posterior predictive distribution of the relative probability of *P. falciparum* infection and detection (assuming  $P' = 1$ ) as a function of age. The heavy black line shows the median, and the dark and light grey regions show the centred 50% and 90% credible intervals, respectively.

Validation of the Age-standardization Model

Smith *et al.* validated their methods using a testing set consisting of 121 *PfPR* surveys, each of which considered two distinct age ranges [4]. To assess the performance of the Bayesian version of their method presented here, the ratios of the paired measurements were considered. These ratios were independent of  $P'$ , a necessary feature for predictive validation since no predictive distribution for  $P'$  had been inferred at that stage. For each pair  $((A_{L,1,i}, A_{U,1,i}), (A_{L,2,i}, A_{U,2,i}))$  in the testing set, a predictive distribution for  $PfPR(A_{L,1,i}, A_{U,1,i})/PfPR(A_{L,2,i}, A_{U,2,i})$  was computed. Averaging these distributions gave a predictive distribution for  $PfPR(A_{L,1}, A_{U,1})/PfPR(A_{L,2}, A_{U,2})$  for any survey drawn at random from the testing set. Figure S3.3 shows the empirical CDF from the testing set and the CDF of the posterior predictive distribution.



**Figure S3.3** A comparison of the empirical (solid) and predicted (broken) CDFs of the log of the ratio of *PfPR* values for paired observations drawn from the testing set. The observed CDF increases more slowly than the predicted CDF, indicating that there was more variation in the testing set than predicted.

The observed CDF increases more slowly than the predicted CDF, demonstrating that there was relatively more probability mass in the tails of the empirical distribution than predicted by the

model. Future research will be directed toward obtaining better predictions. The data in the training set did not exhibit extra-binomial variation, so it may be appropriate to use more flexible distributions, such as Dirichlet process mixtures, for the parameters  $\{\log(\alpha_i), 1/\log(c_i), 1/\log(b_i), \text{logit}(s_i)\}$  and the age distributions  $S_i$ .

### S3.4 Implementation Details

Implementation of the MBG procedure was divided into two computational tasks: (i) an inference stage in which MCMC was used to generate samples from the posterior distribution of the parameter set and of the space-time random field at the data locations; and (ii) a prediction stage in which samples were generated from the posterior distribution of  $P\text{PR}_{2-10}$  at each prediction location on a 5×5 km grid within the limits of stable transmission. Each stage is explained in more detail below.

#### Markov chain Monte Carlo Algorithm

Both the main geostatistical model and the age-standardization sub-model were fitted using the MCMC algorithm [1,2]. The algorithm was implemented in the Python [11] and FORTRAN programming languages using the open-source Bayesian statistics package PyMC [12] and the numerical packages SciPy and NumPy [13]. The code (which contains working values for all tuning parameters) is available on request, as are the dynamic traces of all unobserved parameters.

#### MCMC Algorithm for the Main Spatial Model

The parameter  $V$  and the evaluation of  $f$  at the sampling locations and times were updated using Gibbs steps [1]. The evaluation of the uncorrelated process  $\epsilon$  was updated one point at a time using random-walk Metropolis steps [1]. The model parameters  $\beta_x, \beta_t, \beta_u, \beta_p, \tau, \phi_x, \phi_t, \lambda, \psi$  and  $\rho$  were updated jointly using the method of Haario, Saksman and Tamminen [14].

Within the MCMC loop, the age-standardization factors  $\tilde{k}_i$  were not imputed explicitly. We were not interested in their particular values, and marginalizing out "nuisance parameters" ahead of time usually improves the mixing of MCMC algorithms. Before the MCMC loop began, the marginal likelihood:

$$\int \text{Bin}(N_i^+; N_i, k_i P'(x_i, t_i)) D_{\tilde{k}}(\tilde{k}_i; A_{U,i}, A_{L,i}) d\tilde{k}_i \quad (\text{S3.33})$$

was approximated using standard Monte Carlo integration for several values of  $P'(x_i, t_i)$ . That is, values for the model parameters  $\alpha_i, b_i, c_i$  and  $s_i$  and the age distribution  $S_i$  were drawn from their

posterior predictive distributions, then expression (S3.1) was evaluated to obtain  $k_i$ , then the binomial probability was evaluated for several values of  $P'(x_i, t_i)$ . The probabilities resulting from many such draws were averaged. Inside the MCMC loop, the marginal likelihood function for arbitrary values of  $P'(x_i, t_i)$  was evaluated by interpolation.

### MCMC Algorithm for the Age Correction Model

The age distribution parameters  $S_i$ ,  $S_0$  and  $\nu$  are independent of the relative *PfPR* parameters  $P'_i$ ,  $\alpha_i$ ,  $c_i$ ,  $b_i$ ,  $s_i$ ,  $\mu_A$ ,  $\sigma$  and  $R$  given the data, so these two groups of parameters were inferred using separate MCMC algorithms.

In the MCMC for the age distribution parameters, the survey populations' age distributions  $S_i$  were updated using Gibbs steps [1]. The concentration parameter  $\nu$  was updated using random-walk Metropolis steps [1]. The typical age distribution  $S_0$  was represented as a normalized sequence of gamma random variables [15], and these variables were updated one at a time using random-walk Metropolis steps [1].

In the MCMC for the relative *PfPR* parameters, the distributional parameters  $\mu_A$ ,  $\sigma$  and  $R$  were updated jointly using the method of Haario, Saksman and Tamminen [14]. The parameters  $P'_i$ ,  $\alpha_i$ ,  $c_i$ ,  $b_i$  and  $s_i$  were updated jointly for each population  $i$  using the same method.

### Spatiotemporal Prediction and Map Generation

The output of the MCMC stage consisted of  $\{\theta_{(l)}; l = 1, \dots, m\}$  samples from the posterior of the parameter set  $\theta = \{\beta_x, \beta_t, \beta_u, \beta_p, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, k, V\}$  and a corresponding  $\{f(x_i, t_i)_{(l)}; l = 1, \dots, m\}$  samples from the posterior of the space-time random field at each of the  $n$  data locations  $\{(x_i, t_i); i = 1, \dots, n\}$ . For every  $l'$ th sample, the conditional distribution of the 2007 annual mean of the space-time random field was predicted at each prediction location  $x_j$  on the nodes of a regular 5x5 km grid within the spatial limits of stable *P. falciparum* transmission [16]. The distribution of the 2007 annual mean  $f(x_j)_{(l)}$  for prediction location  $x_j$  was modelled as the joint multivariate normal distribution of the 12 predicted monthly values  $\{t = 2007_{Jan}, \dots, 2007_{Dec}\}$  for that year specified by a 12 element mean vector  $\hat{\mathbf{y}}(x_j)_{(l)}$  and 12 x 12 variance-covariance matrix  $\hat{\sigma}^2(x_j)_{(l)}$ :

$$f(x_j)_{(l)} \sim MVN(\hat{\mathbf{y}}(x_j)_{(l)}, \hat{\sigma}^2(x_j)_{(l)}) \quad (\text{S3.34})$$

The mean vector  $\hat{\mathbf{y}}(x_j)_{(l)}$  was computed using:

$$\hat{\mathbf{y}}(x_j)_{(l)} = \boldsymbol{\mu}_{P(l)} + \mathbf{C}_{DP(l)}^T \cdot \mathbf{C}_{DD(l)}^{-1} \cdot (\mathbf{p}(x, t) - \boldsymbol{\mu}_{D(l)}) \quad (\text{S3.35})$$

where  $\mu_P$  and  $\mu_D$  were the predicted mean of the random field at each of the 12 prediction times  $\{t = 2007_{Jan}, \dots, 2007_{Dec}\}$  at spatial location  $x_j$  and at each of the  $n$  data locations respectively,  $C_{DP}$  and  $C_{DD}$  were the data-to-prediction and data-to-data covariance matrices respectively, and  $p(x, t)$  was the vector of  $n$  data values. The  $12 \times 12$  variance-covariance matrix  $\hat{\sigma}^2(x_j)_{(l)}$  was computed using:

$$\hat{\sigma}^2(x_j)_{(l)} = C_{PP(l)} - C_{DP(l)}^T \cdot C_{DD(l)}^{-1} \cdot C_{DP(l)} \quad (S3.36)$$

The value of the  $l'th$  sample of  $V$ , the variance of the unstructured component  $\epsilon(x, t)$ , was then added to the diagonal of the matrix  $\hat{\sigma}^2(x_j)_{(l)}$  and 1000 draws were made randomly from the distribution specified in equation S3.34. These draws represented samples from the posterior distribution of  $f(x_j)$  and were subject to an inverse logit transform and then multiplied by the  $l'th$  sample of the age-standardization parameter  $k_{2-10(l)}$  to form the  $l'th$  sample from the posterior distribution of the predicted mean annual 2007  $PfPR_{2-10}$  endemicity surface at location  $x_j$ :

$$P'_{2-10}(x_j)_{(l)} = \text{logit}^{-1} \left( f(x_j)_{(l)} + \epsilon_{(l)} \right) k_{2-10(l)} \quad (S3.37)$$

This procedure was repeated for every  $l'th$  sample to form the set  $\{P'_{2-10}(x_j)_{(l)}; l = 1, \dots, m\}$  of  $m$  samples for each prediction location. The point estimate of  $PfPR_{2-10}$  endemicity at each location was defined as the mean of this set, whilst the probability of membership to each class was computed as the proportion of these samples falling within each class definition:  $PfPR_{2-10} \leq 5\%$ ;  $PfPR_{2-10} > 5\% - < 40\%$ ;  $PfPR_{2-10} \geq 40\%$ .

#### Monte Carlo Standard Errors

Monte Carlo standard errors for the expectations of the scalar parameters and the class membership probabilities of selected points were estimated using the procedure recommended by Flegal *et al.* [17]. For the class membership probabilities, it was first verified by direct simulation that the Monte Carlo standard error of the estimated class membership probabilities given each sample from the posterior was negligible. That is, we set  $l$  large enough that the estimated class membership probability was, for all intents and purposes, a deterministic function of the model parameters and the value of  $f$  at the data locations. This observation allowed us to apply the method of Flegal *et al.* [17] to the class membership probabilities directly. The largest half-widths amongst several representative prediction points were found to be approximately 0.3%.



### S3.4 Overview of Map Generation

A deterministic model outputting a single predicted value of  $PfPR_{2-10}$  for each pixel would lead to a single predicted map. The output of the MBG model for each pixel, however, was not a single prediction but a large set of possible values representing the predicted posterior distribution of  $PfPR_{2-10}$  and, together, provide a complete model of our uncertainty. Examples of these posterior distributions are shown in Figure S3.4 for three pixels selected from nations in each modelled region (Brazil (America), Côte d'Ivoire (Africa+) and Papua New Guinea (CSE Asia)). The information contained in each pixel's posterior distribution was summarized in different ways to make three different global maps. First, the mean of each posterior distribution was calculated which became our "point estimate" of  $PfPR_{2-10}$  (3%, 68% and 22% respectively for the three example sites in Figure S3.4) and these values for each pixel generated the global map shown in Figure 3 of the main text. We also calculated the probability of membership to each of the three endemicity classes which was found by calculating the relative proportion of the posterior distribution falling within each class. These class probabilities were used to generate three further maps. The map shown in Figure 4 of the main text displays, for each pixel, which of the three classes had the largest class probability and was therefore considered the 'most likely' endemicity class for that pixel ( $\leq 5\%$ ,  $\geq 40\%$  and  $>5\% - <40\%$  respectively for the three example sites). The map shown in Figure 5 of the main text displays the class probability for that most likely class (0.89, 0.80 and 0.4 respectively for the three example sites). This latter map can be interpreted as a summary of the modelled uncertainty in assigning class memberships. Values close to one indicated a high degree of certainty in class assignment whilst values close to one-third indicated a high degree of uncertainty; that is even the "most likely class" was only marginally more likely than the other classes). Figure S4.1A-C show the probability of membership to each class individually, regardless of which was calculated as most likely. The map shown in Figure S4.2 displays the standard deviation of each posterior distribution, which can be interpreted as a further indicator of uncertainty. Where predictions have a high uncertainty the posterior distribution will be dispersed across a wide range of possible  $PfPR_{2-10}$  values and will therefore have a larger standard deviation than a more certain, less dispersed posterior distribution.

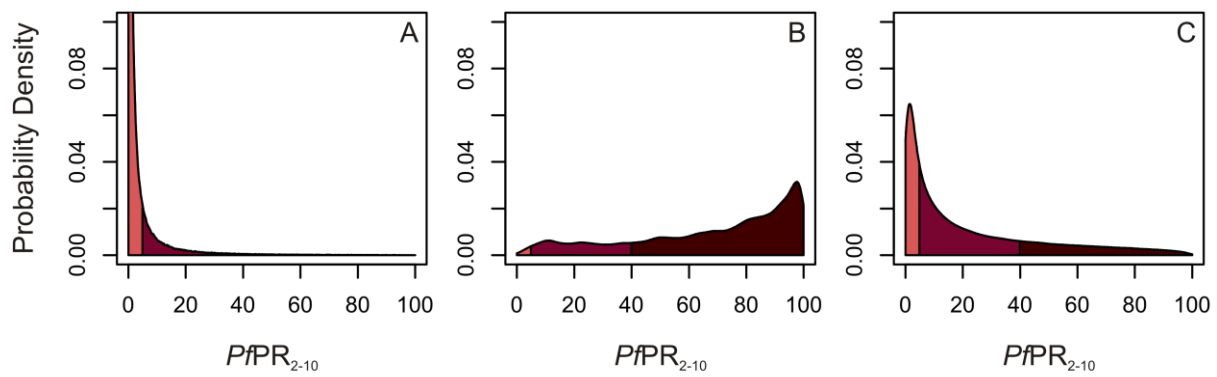


Figure S3.4. The posterior distribution predicted from the MBG model, displayed as a probability density across the range of possible  $PPR_{2-10}$ , for selected 5×5 km pixels in A Brazil (note probability density off scale), B Côte d'Ivoire and C Papua New Guinea. The equivalent posterior distributions from each pixel on a 5×5 km global grid were used to generate a series of maps, as described in the text. The three coloured regions on each plot highlight the section of each distribution corresponding to the three endemicity classes ( $\leq 5\%$ ,  $\geq 40\%$  and  $>5\% - <40\%$ ), the area of which is equivalent to the probability of membership to that class.

## References

1. Gelman A, Carlin JB, Stern HS (2003) Bayesian Data Analysis. Texts in Statistical Science. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC. 696 p.
2. Gilks WR, Spiegelhalter DJ (1999) Markov Chain Monte Carlo in Practice. Interdisciplinary statistics. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC.
3. Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Monographs on Statistics and Applied Probability 101. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC.
4. Smith DL, Guerra CA, Snow RW, Hay SI (2007) Standardizing estimates of the *Plasmodium falciparum* parasite rate. Malar J 6: 131.
5. Stein ML (2005) Space-time covariance functions. J Am Stat Assoc 100: 310-321.
6. Antosiewicz HA (1964) 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.
7. Davis GM (1964) 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.
8. Azzalini A (1985) A class of distributions which includes the normal ones. Scan J Stat 12: 171-178.

9. R Development Core Team (2008) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, URL: <http://www.R-project.org>.
10. Pull JH, Grab B (1974) Simple epidemiological model for evaluating malaria inoculation rate and risk of infection in infants. Bull World Health Organ 51: 507-516.
11. van Rossum G (2008) Python Programming Language - Official Website. Website: URL <http://www.python.org>.
12. Fonnesbeck C, Huard D, Patil AP (2008 ) PyMC 2.0 User's Guide: installation and tutorial. URL: <http://www.trichech.us/pymc>.
13. Oliphant TE (2007) Python for scientific computing. Comput Sci Eng 9: 10-20.
14. Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. Bernoulli 7: 223-242.
15. Hogg RV, Craig AT (2005) Introduction to Mathematical Statistics. Upper Saddle River, New Jersey, U.S.A: Prentice Hall Inc. 564 p.
16. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. PLoS Med 5: e38.
17. Flegal JM, Haran M, Jones GL (2007) Markov Chain Monte Carlo: can we trust the third significant figure? ArXiv.org preprint: arXiv:math/0703746v4.

## **Protocol S4. Model Validation Procedures**

Assessing the plausibility of the model output is essential for reliable interpretation of the mapped output. Many different measures of map uncertainty are available. Three different aspects of the performance of the predictive model were assessed using a range of validation statistics as described in the main text. This section describes in more detail the procedures used to define a validation set, obtain validation data, and compute a series of summary validation statistics and plots. This supplement also provides information on additional types of uncertainty quantification and provides some discussion on how they should be interpreted.

### **S4.1 Creation of the Validation Sets**

Validation statistics obtained via prediction of a validation set are representative of model performance only if the validation set itself is a representative sample of the prediction space. Visual examination of the  $PfPR_{2-10}$  point data used in this study revealed clear evidence of spatial clustering (Figure 1, main text). As such, a simple random sample drawn from this set would be similarly clustered and not spatially representative of the predicted  $PfPR_{2-10}$  surface as

a whole. To generate a spatially representative validation set, the full set of 7,953 data locations was stratified into the three global regions as defined in the main text (America, Africa+ and CSE Asia) and a spatially declustered sampling procedure was implemented within each region. Thiessen polygons were defined around each data location  $x_i$  within each region. A Thiessen polygon defines the area closest to each data point in Euclidian space relative to surrounding points. Each datum was then assigned a weight  $w_i$  defined as  $w_i = \sqrt{a_i}$  where  $a_i$  is the area of the Thiessen polygon surrounding the data location,  $x_i$ . A sample of size  $n$  was drawn without replacement from the regional set where each datum had a probability of selection proportional to its weight  $w_i$ . Those surveys located outside the stable limits of transmission were excluded from selection. The sample size  $n$  was defined as the larger of  $n = 30$  or 10% of the total data within each region, rounded to the nearest integer value, where 30 was chosen as a rule-of-thumb value for an acceptable minimum sample size.

Hold-out sets of size  $n = 30$ ,  $n = 531$ , and  $n = 239$  were defined in this way for the America, Africa+, and CSE Asia regions respectively. The model was then run in full for each region independently using the corresponding thinned sets of  $n = 231$ ,  $n = 4,776$ , and  $n = 2,146$  data to predict  $PfPR_{2-10}$  at the validation locations. In contrast to the main model run in which  $PfPR_{2-10}$  was predicted as an annual mean for 2007, the validation run predicted values for the time corresponding to the mid-point of each validation survey to enable fairer comparisons of the observed and predicted  $PfPR_{2-10}$  values.

## S4.2 Procedures for Testing Model Performance

Predictive performance of the model was tested using three different approaches: the ability of the model to (i) predict the correct endemicity class at unsampled locations; (ii) predict point-values of  $PfPR_{2-10}$  at unsampled locations; and (iii) provide realistic measure of uncertainty for each prediction.

### Predicting Endemicity Class

The accuracy of predicting class membership was determined in terms of sensitivity and specificity using the area under curve (AUC) of a receiver-operating characteristic (ROC) curve [1,2,3,4]. These statistics provide a summary of model performance across a range of sensitivity and specificity values, with values of AUC=1 representing a perfect model and AUC=0.5 a model with no discriminatory power. ROC plots and AUC statistics were computed for each of the three endemicity classes  $PfPR_{2-10} \leq 5\%$ ,  $PfPR_{2-10} > 5\% - < 40\%$ , and  $PfPR_{2-10} \geq 40\%$ . Benchmarks for interpreting AUC values are inherently arbitrary [5], but values exceeding 0.7 are commonly recognised as representing fair to good discrimination, and values exceeding 0.9 as representing excellent discrimination. Simpler statistics relating to the prediction of endemicity

class were also computed: the percentage of points assigned to the correct endemicity class, the percentage of points incorrectly assigned to a non-adjacent class, that is, points in the  $PfPR_{2-10} \leq 5\%$  class assigned as  $PfPR_{2-10} \geq 40\%$  or vice versa, as well as a full  $3 \times 3$  class contingency table.

#### Predicting Point Values of $PfPR_{2-10}$

The validation procedure generated  $n = 800$  point estimates of  $PfPR_{2-10}$ , where point estimates were calculated using the mean of each predicted posterior distribution. This set of point estimates  $(p^*(\mathbf{x}_i); i = 1, \dots, n)$  (where the asterisk denotes a prediction) was then compared to the corresponding set of known  $PfPR_{2-10}$  values  $(p(\mathbf{x}_i); i = 1, \dots, n)$  at the validation locations. The ability of the model to predict point-values of  $PfPR_{2-10}$  at unsampled locations was quantified using three simple summary statistics: the correlation coefficient between the predicted and actual set, the mean prediction error (ME) defined as:

$$ME = \frac{1}{n} \sum_{i=1}^n (p^*(\mathbf{x}_i) - p(\mathbf{x}_i)), \quad (S4.1)$$

and the mean absolute prediction error (MAE) defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |p^*(\mathbf{x}_i) - p(\mathbf{x}_i)| \quad (S4.2)$$

The correlation coefficient provides a straightforward measure of linear association between the data and prediction sets, the ME provides a measure of the bias of the predictor (the overall tendency to over or under predict  $PfPR_{2-10}$  values), and the MAE provides a measure of the mean accuracy of individual predictions (the average magnitude of difference between each actual and predicted value). ME and MAE values were presented as both absolute values and as a proportion of the mean  $PfPR_{2-10}$  in each region as calculated from the validation set. A scatter plot was also generated as a visualisation of the correspondence between point estimates of  $PfPR_{2-10}$  and the corresponding known values.

A sample semi-variogram was calculated from standardised model residuals to assess the presence of residual spatial autocorrelation unexplained by the model output. Standardised Pearson [6] residuals  $r_i$  were defined for each validation location as:

$$r_i = \frac{N_i^+ - N_i p_i^*}{\sqrt{N_i p_i^* (1 - p_i^*)}} \quad (\text{S4.3})$$

where  $N_i$  is the number of individuals surveyed in survey  $i$ ,  $N_i^+$  is the age-standardized number of *P. falciparum* positive responses in that survey and  $p_i^*$  is the corresponding point-prediction of  $PfPR_{2-10}$ . This standardisation follows established procedures [7,8] and rescales the raw model residuals to account for their variance characteristics as proportion values. Following the procedure outlined by Diggle and Ribeiro [9] this sample semi-variogram was compared to a Monte Carlo envelope computed from 99 random permutations of the same residual set. This envelope represents the range of semi-variograms that could be expected by chance in the absence of any spatial structure. Where the semi-variogram of interest lies entirely within this envelope, it can be considered to display no significant spatial structure.

### Providing Realistic Measures of Uncertainty for Each Prediction

Posterior distributions arising from Bayesian models provide an estimate of the relative probability of a particular outcome and can be used to characterize uncertainty of prediction [10]. Our model generated a posterior distribution for each unsampled location and a procedure [11,12,13] was implemented to test how well the validation set of 800 posterior distributions captured the true uncertainty in our model output. Each such distribution provided a prediction of the probability with which  $PfPR_{2-10}$  exceeded any given value in the range 0-100%. The distribution of each validation survey  $i$  was discretized into 101 probability thresholds, that is, the set  $\{p_{i(j)}^*; j = 1, \dots, 101\}$  of  $PfPR_{2-10}$  values that were predicted to exceed the known  $PfPR_{2-10}$  value  $p_i$  with a probability of  $P_j = 0.00, 0.01, \dots, 1.00$ . Consider any one of these probability thresholds. The value  $p_{i(95)}^*$ , for example, denotes the value of  $PfPR_{2-10}$  at location  $i$  that was predicted to have a probability of  $P = 0.95$  of exceeding the known value  $p_i$ . The value  $p_{i(95)}^*$  could be defined at each of the 800 validation locations and, at each, compared to the corresponding known value  $p_i$ . If the value of  $p_{i(95)}^*$  was predicted correctly at all 800 validation locations, then 95% of these values would be expected to exceed the corresponding known  $PfPR_{2-10}$  value  $p_i$  whilst 5% would be smaller. This rationale can be extended such that, for any given probability threshold  $P_j$ , the proportion  $P_j^*$  of the 800 values  $\{p_{i(j)}^*; i = 1, \dots, n\}$  that exceed their corresponding known value  $\{p_{i(j)}; i = 1, \dots, n\}$  should equal the probability threshold  $P_j$ . The proportion  $P_j^*$  is also termed the coverage probability [11]. In a perfect model of uncertainty, values of the coverage probabilities  $\{p_{i(j)}^*; j = 1, \dots, 101\}$  would correspond exactly to the probability thresholds  $\{P_j; j = 1, \dots, 101\}$  and the P-P plot comparing  $P_j$  with  $P_j^*$  would show values lying perfectly along the 1:1 line.

### S4.3 Additional Results

A contingency matrix was generated (Table S4.1) showing, for each region, the numbers of validation points that were classified into each of the three endemicity classes in relation to their known class. For all regions, the majority of points were correctly classified as shown by the shaded cells along the diagonal of each matrix. Classification errors were proportionally smaller for the highest and lowest endemicity class. Globally, 16% (50/308) of validation points from the  $\leq 5\%$  class were wrongly classified as being in the 5-40% class and 0.6% (2/308) in the  $\geq 40\%$  class. Four percent (7/172) of validation points from the  $\geq 40\%$  class were wrongly classified as being in the  $\leq 5\%$  class and 22% (37/172) in the 5-40% class. Errors were larger for the middle 5-40% class in which globally 25% (81/320) were wrongly classified in the  $\leq 5\%$  class and 18% (57/320) in the  $\geq 40\%$  class.

	America			Africa+			CSE Asia		
	$\leq 5\%$	5-40%	$\geq 40\%$	$\leq 5\%$	5-40%	$\geq 40\%$	$\leq 5\%$	5-40%	$\geq 40\%$
$\leq 5\%$	24	2	0	143	30	1	89	18	1
5-40%	4	0	0	44	109	50	33	73	7
$\geq 40\%$	0	0	0	2	29	123	5	8	5

Table S4.1. Contingency table for the America, Africa+ and CSE Asia regions comparing the observed (rows) endemicity classes of the validation surveys with those class memberships predicted (columns) by the model. If the prediction was perfect all numbers would appear in the grey highlighted cells.

Further maps illustrating the uncertainty of the predictions are also provided. In Figure S4.1 the predicted probability of  $PfPR_{2-10}$  falling in each endemicity class is shown. Note the use of a probability scale from zero to one differs from the one third (0.333) to one (1.000) scale used in Figure 5A of the main text. This difference arises because the latter plot displays only probabilities of membership to the most likely class which, by definition must have a probability in excess of one third. In Figure S4.2, the standard deviation of each per-pixel posterior distribution in units of  $PfPR_{2-10}$  is provided as an additional index of relative uncertainty.

Figure S4.3 presents a combined visualisation of the  $PfPR_{2-10}$  endemicity class assignments (presented in Figure 4 of the main text) and the certainty of those class assignments (presented in Figure 5A of the main text). Each of the three classes has a different primary colour associated and the shading for each pixel is derived as a composite of these three colours in proportion to the probability of membership to each class. Pure shades therefore represent

highly certain class assignments (the membership probability to one class was very high and to the other two very low), whilst mixed shades represent less certain assignments (membership probability was shared more evenly across two or three classes).

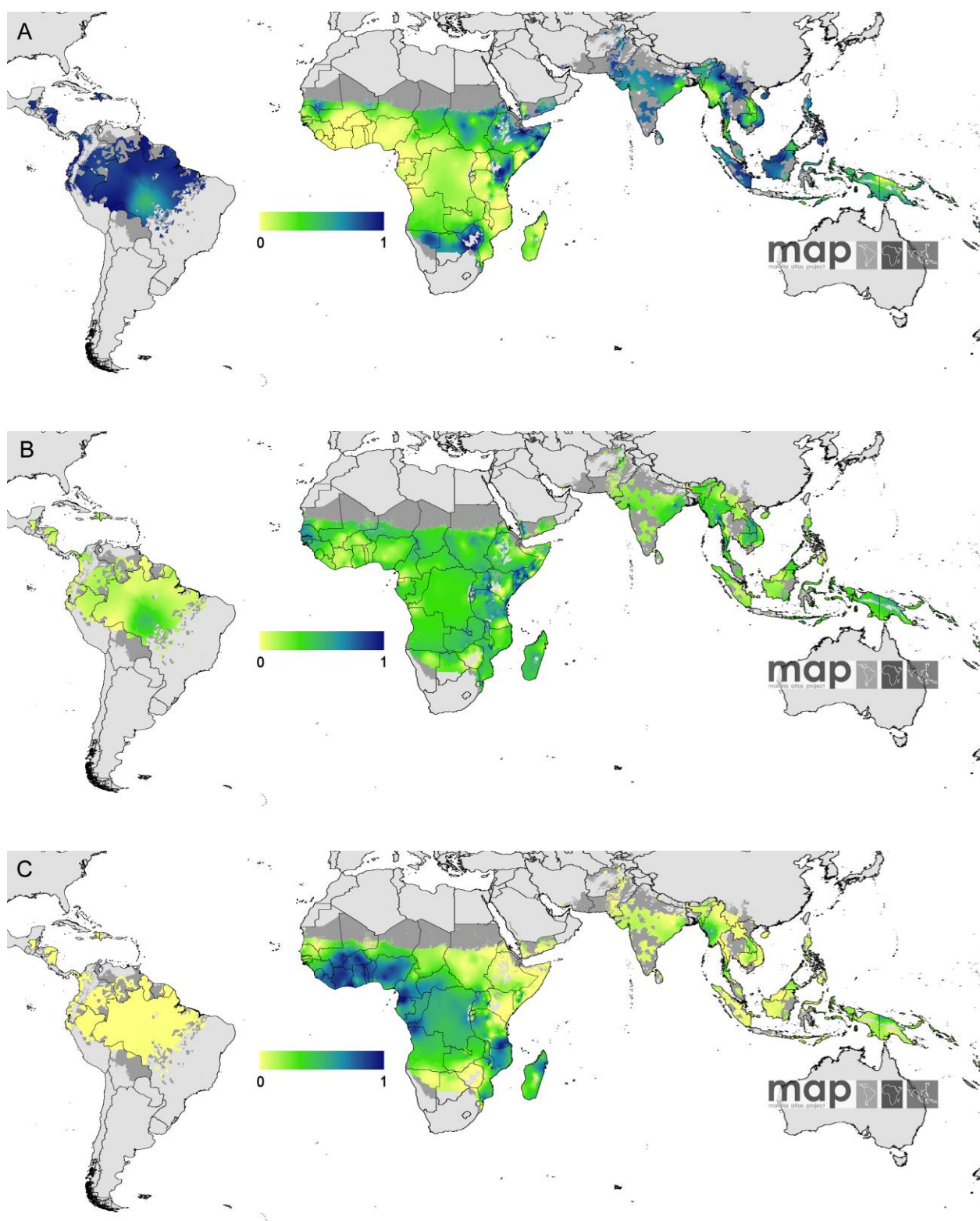


Figure S4.1. The predicted probability of  $PfPR_{2-10}$  falling in each endemicity class: A  $PfPR_{2-10} \leq 5\%$ ; B  $PfPR_{2-10} > 5\% - < 40\%$ ; C  $PfPR_{2-10} \geq 40\%$  within the stable limits of *P. falciparum*



transmission. Yellow values indicate a small probability of class membership, and blue a large probability. Dark grey areas indicate the limits of unstable risk ( $PfAPI < 0.1$  per thousand people pa) and pale grey no risk ( $PfAPI = 0$  per thousand people pa) [14,15,16].

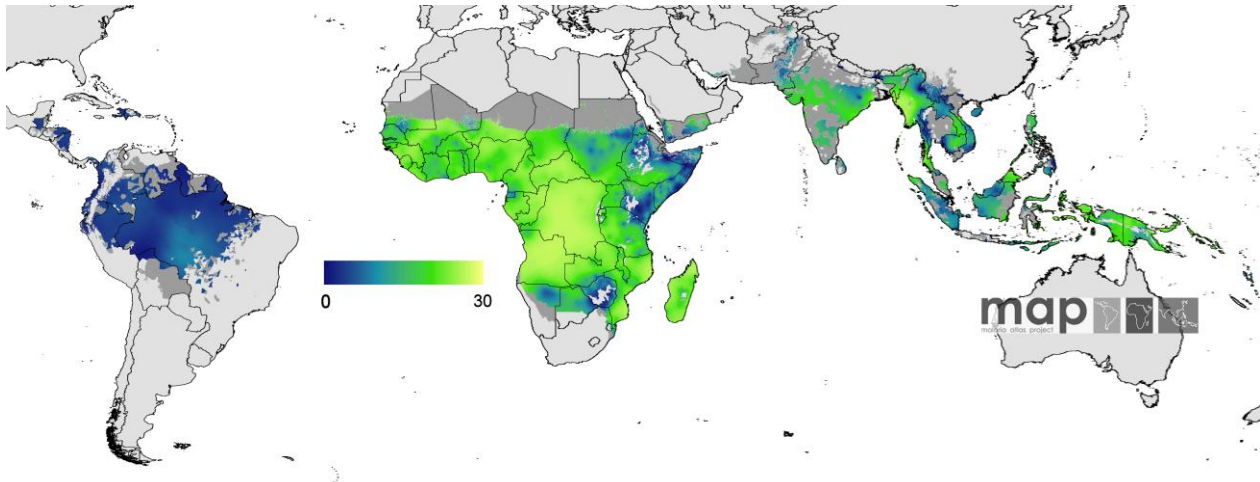


Figure S4.2. The standard deviation (in units of  $PfPR_{2-10}$ , 0-100%) of each per-pixel posterior distribution within the stable limits of *P. falciparum* transmission. These values act as an index of relative uncertainty. Where predictions were made with large uncertainty, the resulting posterior distributions were dispersed across a wide range of possible values of  $PfPR_{2-10}$  so the standard deviation of these values was relatively high. Conversely, where uncertainty was small, posterior distributions were well defined across a narrow range of values resulting in low standard deviations. Dark grey areas indicate the limits of unstable risk ( $PfAPI < 0.1$  per thousand people pa) and pale grey no risk ( $PfAPI = 0$  per thousand people pa) [14,15,16].

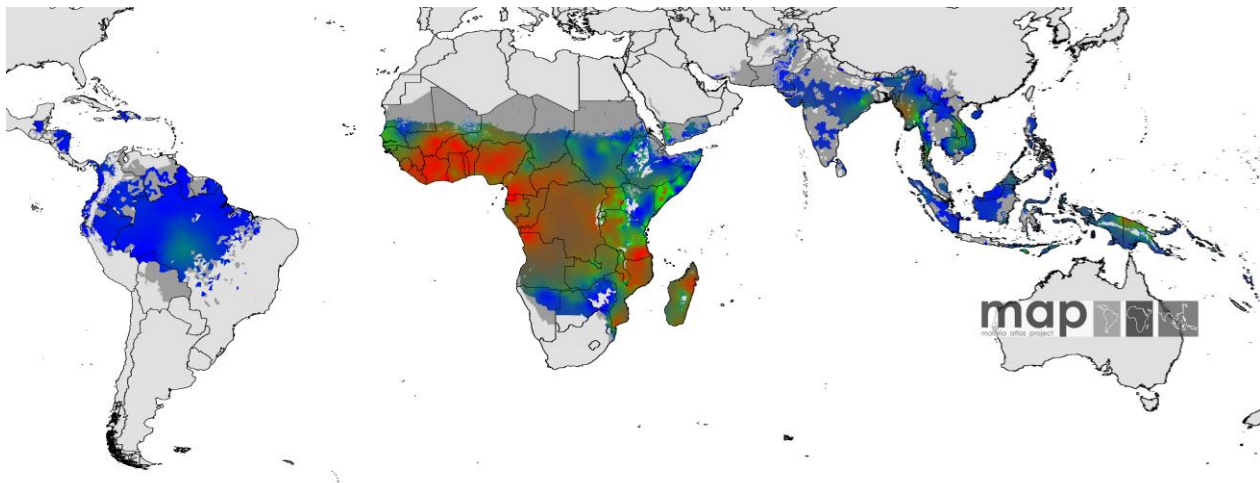


Figure S4.3. Three-colour composite showing combined probability of membership to three classes of  $PfPR_{2-10}$  endemicity categorized as low risk  $PfPR_{2-10} \leq 5\%$  (pure blue); intermediate risk  $PfPR_{2-10} > 5\% - < 40\%$  (pure green) and high risk  $PfPR_{2-10} \geq 40\%$  (pure red). The shading for each pixel is derived as a composite of the three pure class colours in proportion to the probability of membership to each class. Thus, pure shades indicate a highly certain assignment to a single class whilst mixed intermediate shades indicate an uncertain assignment with

membership probability shared across two or three classes. Dark grey areas indicate the limits of unstable risk ( $PfAPI < 0.1$  per thousand people pa) and pale grey no risk ( $PfAPI = 0$  per thousand people pa) [14,15,16].

## References

1. Brooker S, Hay SI, Bundy DA (2002) Tools from ecology: useful for evaluating infection risk models? *Trends Parasitol* 18: 70-74.
2. Clements ACA, Lwambo NJS, Blair L, Nyandindi U, Kaatano G, et al. (2006) Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop Med Int Health* 11: 490-503.
3. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Lett* 27: 861-874.
4. Metz CE (1978) Basic principles of ROC analysis. *Semin Nucl Med* 8: 283-298.
5. Conraths FJ, Schares G (2006) Validation of molecular-diagnostic techniques in the parasitological laboratory. *Vet Parasitol* 136: 91-98.
6. McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Boca Raton, Florida: Chapman and Hall / CRC Press. 540 p.
7. Clements ACA, Moyeed R, Brooker S (2006) Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* 133: 711-719.
8. Diggle P, Moyeed R, Rowlingson B, Thomson M (2002) Childhood malaria in The Gambia: a case-study in model-based geostatistics. *J Roy Stat Soc C-App* 51: 493-506.
9. Diggle PJ, Ribeiro PJ (2007) *Model-based geostatistics*; Bickel P, Diggle P, Fienberg S, Gather U, Olkin I et al., editors. New York: Springer. 228 p.
10. Congdon P (2003) *Applied Bayesian Modelling*. Chichester: John Wiley and Sons Ltd. 457 p.
11. Gething PW, Noor AM, Gikandi PW, Hay SI, Nixon MS, et al. (2008) Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. *Geogr Anal* 40: 167-188.
12. Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. *Geoderma* 103: 3-26.
13. Moyeed RA, Papritz A (2002) An empirical comparison of kriging methods for nonlinear spatial point prediction. *Math Geol* 34: 365-386.
14. Hay SI, Smith DL, Snow RW (2008) Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infect Dis* 8: 369-378.
15. Guerra CA (2007) Mapping the contemporary global distribution limits of malaria using empirical data and expert opinion [Doctoral thesis]. Oxford: University of Oxford. 258 p.
16. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, et al. (2008) The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med* 5: e38.

