View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Open Research Exeter

Spatio-temporal modelling of climate-sensitive disease risk: towards an early warning system for dengue in Brazil

Rachel Lowe[a,*], Trevor C. Bailey[a,*], David B. Stephenson[a,*], Richard J. Graham[b,*], Caio. A. S. Coelho[c,*], Marilia. Sá Carvalho[d,*], Christovam Barcellos[d,*]

[a]*School of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter, EX4 4QF, UK*

[b]*Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK*

[c]*Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Rodovia Presidente Dutra, Km 40, SP-RJ 12630-000, Cachoeira Paulista, SP, Brazil*

[d]*Oswaldo Cruz Foundation, Health Information Research Laboratory, LIS/ICICT/Fiocruz, Av. Brasil, Manguinhos, Rio de Janeiro, CEP 21045-900, Brazil*

Abstract

This paper considers the potential for using seasonal climate forecasts in developing an early warning system for dengue fever epidemics in Brazil. In the first instance, a generalised linear model (GLM) is used to select climate and other covariates which are both readily available and prove significant in

*Corresponding author. Tel: +44 (0) 1392 725280 Fax: +44 (0) 1392 217965

Email addresses: R.Lowe@exeter.ac.uk (Rachel Lowe), T.C.Bailey@exeter.ac.uk (Trevor C. Bailey), D.B.Stephenson@exeter.ac.uk (David B. Stephenson), richard.graham@metoffice.gov.uk (Richard J. Graham), caio.coelho@cptec.inpe.br (Caio. A. S. Coelho), carvalho@fiocruz.br (Marilia. Sá Carvalho), xris@cict.fiocruz.br (Christovam Barcellos)

prediction of confirmed monthly dengue cases based on data collected across the whole of Brazil for the period January 2001 to December 2008 at the microregion level (typically consisting of one large city and several smaller municipalities). The covariates explored include temperature and precipitation data on a $2.5^\circ \times 2.5^\circ$ longitude-latitude grid with time lags relevant to dengue transmission, an El Niño Southern Oscillation index and other relevant socio-economic and environmental variables. A Negative-Binomial model formulation is adopted in this model selection to allow for extra-Poisson variation (overdispersion) in the observed dengue counts caused by unknown/unobserved confounding factors and possible correlations in these effects in both time and space. Subsequently, the selected global model is refined in the context of the South East region of Brazil where dengue predominates, by reverting to a Poisson framework and explicitly modelling the overdispersion through a combination of unstructured and spatio-temporal structured random effects. The resulting spatio-temporal hierarchical model (or GLMM - generalised linear mixed model) is implemented via a Bayesian framework using Markov Chain Monte Carlo (MCMC). Dengue predictions are found to be enhanced both spatially and temporally when using the GLMM and the Bayesian framework allows posterior predictive distributions for dengue cases to be derived which can be useful for developing a dengue alert system. Using this model, we conclude that seasonal climate forecasts could have potential value in helping to predict dengue incidence months in advance of an epidemic in South East Brazil.

Keywords: dengue fever, prediction, epidemic, spatio-temporal model, seasonal climate forecasts

1. Introduction and motivation

The early identification of an epidemic of infectious disease is an important first step towards implementing effective interventions to control the disease and reducing mortality and morbidity in human populations (Kuhn et al., 2005). However, often an epidemic is under way before the authorities are notified and control measures are put in place. In this paper we assess the potential for using seasonal climate forecasts to provide early warnings of future increased and geographically specific risk of dengue fever in Brazil.

Dengue fever and its more severe form (dengue hemorrhagic fever) is one of the most important emerging tropical diseases at the beginning of the 21st century in terms of morbidity and mortality (Gubler, 2002, Guzman and Kouri, 2003). Dengue is an acute viral disease characterised by fever, headache, muscle and joint pains, rash, nausea, and vomiting, while dengue haemorrhagic fever is a potentially deadly complication that in severe cases, can cause circulatory failure. Dengue viruses are transmitted by the bite of infected *Aedes* females, in particular *Aedes aegypti*, an urban mosquito with widespread distribution in tropical cities. Field survivability of *Aedes aegypti* and patterns of dengue transmission are influenced by many factors including, but not limited to, climate which influences mosquito biology and interactions between the mosquito vector and dengue virus (Kuno, 1995; Scott et al., 2000; Sanchez et al., 2006). In many regions, epidemic dengue transmission is seasonal in response to variability in temperature and rainfall. There have been recent concerns of a worldwide spread of dengue fever because of climate changes that could favour an expansion of the transmission area.

In Brazil, the greatest incidence of cases occur from January to May when the climate is warmest and most humid (Braga and Valle, 2007). Three of the four dengue virus serotypes have spread throughout Brazil, where reported dengue cases in the last decade represent about 60% of dengue cases reported in the Americas as a whole (Nogueira et al., 2007a). Dengue epidemics impact heavily on the national health services. There is no specific treatment for dengue, but appropriate medical care frequently saves the lives of patients with the more serious dengue haemorrhagic fever. A major epidemic occurred in Brazil in 2008, with 764 040 reported cases (January to September) including 3 848 cases of hemorrhagic fever and 213 deaths [1]. In Rio de Janeiro, military field hospitals were opened during the 2008 outbreak to help to ease the pressure on emergency rooms packed with people suffering from dengue [2].

The current monitoring system in Brazil relies on observing dengue incidence in December/January to estimate epidemic potential late in the austral summer. However, this does not provide a quantitative measure or much predictive lead time. The greater the lead time available for forecasting disease risk, the greater the opportunity for effective disease risk intervention, although long term predictions often involve larger errors. Myers et al. (2000) suggested that epidemic forecasting is most useful to health services when case numbers are predicted two to six months ahead. This would allow time for the allocation of resources to interventions such as preparing health care

[1] http://portal.saude.gov.br/saude/

[2] http://news.bbc.co.uk/1/hi/world/americas/7324000.stm

services for increased numbers of dengue patients and educating populations to eliminate mosquito breeding sites i.e. by regularly emptying water that accumulates in discarded refuse, tyres and domestic water storage containers, commonplace in urban slums/favelas found in some areas of Brazil.

As seasonal climate forecasts predict seasonal or monthly average temperature and precipitation (and other variables) for the forthcoming months/season in both time and space, they could potentially be used in a national dengue early warning system (EWS) for Brazil to aid epidemic planning months in advance. EWS based on seasonal climate forecasts have been developed to predict malaria incidence, for example in Botswana (Thomson et al., 2006), but there has been limited progress in developing EWS for dengue fever. Therefore, the use of seasonal climate forecasts with lead times of one month or more within a dengue EWS is a research area in need of exploration.

Before assessing the viability of using seasonal climate forecasts in a dengue prediction model, a model driven by observed climate variables with time lags relevant to dengue transmission, issued at the same resolution as the climate forecasts, must first be evaluated. If a significant relationship is identified between observed climate and dengue in Brazil, the use of forecast climate for dengue prediction purposes could be valuable. The remainder of this paper focuses on the viability of using observed climate variables in a spatio-temporal dengue prediction model.

In Section 2 we outline some of the key processes involved in dengue transmission and describe the data used in the statistical modelling. Section 3 documents the exploratory data analysis and model selection process to find

which monthly climate variables and time lags are statistically significant for modelling dengue incidence in Brazil, using a negative-binomial generalised linear model (GLM) to allow for overdispersion. In section 4 we focus on the South East of Brazil where dengue predominates, and refine the previously selected Brazilian global model by reverting to a Poisson formulation and explicitly including spatially unstructured and spatio-temporal structured random effects via a Bayesian framework to account for unobserved/confounding factors. Section 5 then assesses the ability of the refined model to issue dengue epidemic warnings for the peak dengue season in 2008 when a serious epidemic occurred. The final section discusses future ideas for research and summarises the main findings of the paper.

2. Dengue transmission

A number of complex factors are related to dengue transmission, in particular population growth and unplanned urbanization, resulting in substandard housing, inadequate water, sewerage and waste management systems which allow mosquito reproduction. Poverty and health inequality are behind almost all of these factors (Gubler, 2002). Given favourable climatic conditions for development of the dengue-carrying mosquito, the urban environment plays a major role in determining transmission rates. Rainfall may influence the filling of containers out in the open (e.g. old tyres) which create potential breeding sites for the mosquito. More importantly, the breeding of mosquitoes depends on temperature, humidity, the mosquitoes' life expectancy, life-long fecundity, biting activity and virus incubation (Favier

et al., 2005). Several previous studies have examined the link between climate and dengue. In many tropical countries, a positive association between rainfall and dengue incidence have been documented (Li et al., 1985; Moore et al., 1978; Gould et al., 1970). However, a significant relationship was not found for other regions (Eamchan et al., 1989, Goth et al., 1987 Kuno, 1995). Some authors have found that time-lagged climate variables of up to two or three months have a statistically significant association with dengue (Li et al., 1985; Schreiber, 2001; Wu et al., 2007). Precipitation and temperature oscillations over large parts of Latin America and the Caribbean are strongly influenced by changes in Pacific sea surface temperatures (SST) as part of the El Niño Southern Oscillation (ENSO) (Glantz, 2001) and these in turn can influence vector competence and survivorship. Several studies have also used some index of ENSO to model dengue (Brunkard et al., 2008; Cazelles et al., 2005; Gagnon et al., 2001; Hales et al., 1999). Therefore the inclusion of covariates based on the urban environment, climate (e.g. temperature, precipitation, Pacific SST) and their lagged effects appear to be potentially important components of a climate informed dengue prediction model.

Dengue fever data (counts of confirmed cases per month) from January 2001 - December 2008 (96 months) were obtained at municipality level (5651 municipalities) from SINAN DATASUS - an Information System for Notifiable Diseases, established by the Brazilian Ministry of Health[3]. A network of laboratories, capable of diagnosing dengue infections, has been implemented in all states. The network is responsible for confirmation of cases to support

[3]http://dtr2004.saude.gov.br/sinanweb/novo/

epidemiological surveillance (Nogueira et al., 2007b). However, this network is not accessible to all municipalities. Dengue counts were aggregated to the microregion level (558 microregions), where a microregion typically consists of one large city and several smaller municipalities. This alleviates problems of misreporting due to variation in the availability of health services and epidemiological facilities at the municipality level. Figure 1a shows dengue counts for this period grouped into the 5 main regions of Brazil (Figure 1b) and Figure 1c shows the total dengue cases in each microregion for the period January 2001-December 2008. Dengue is most prevalent in the South East. Two major epidemics occurred in the late austral summer of 2002 and 2008, while considerably less dengue occurred in 2004 and 2005. There is very little dengue in South Brazil and the North West Amazon.

Insert Figure 1 here

National cartographic data such as altitude and biome were obtained from the Brazilian Institute for Geography and Statistics (IBGE)[4]. Census data at the microregion level such as population, percentage of urban population, and the percentage of households with a water supply provided by a network, refuse collection and at least one bathroom, was obtained from an aggregated database SIDRA maintained by IBGE. Each microregion belongs to an administrative main region (1. North, 2. North East, 3. South, 4. South East, 5. Central West) and a biome (1. Amazon Rainforest, 2. Caatinga, 3. Cerrado, 4. Atlantic Rainforest, 5. Pampa, 6. Pantanal). A spatial variable

[4]http://www.sidra.ibge.gov.br/

named zone was defined according to the 6 biomes but by also subdividing the Atlantic Rainforest biome into 3 areas (North, South East and South) according to different climatic regimes. For example, south of the Tropic of Capricorn (23.5°S) the climate is more temperate and humid, while in the North East portion of the Atlantic Rainforest the climate is relatively warmer. Therefore 8 zones are defined for which climatic, geographical and ecological conditions are homogeneous. In a modelling context, zone is treated as a categorical variable, or factor. Figure 2 shows the spatial distribution of altitude and urban population in Brazil and the location of the geographical zones (Figure 2a, b and c). Figures 2d, e and f illustrate the relationship between these covariates and standardised morbidity ratio (SMR) for the given time period where, for a microregion i, the SMR is defined as the ratio of observed (y_i) to expected (e_i) dengue cases in the time period. The expected cases e_i in each microregion are calculated as the population at risk (p_i) multiplied by the global dengue detection rate over the whole of Brazil for the time period ($e_i = p_i \times \sum y_i / \sum p_i$). Altitude has a statistically significant negative relationship with dengue SMR (as altitude increases, dengue counts decrease) and percentage of urban population had a statistically significant positive relationship, given a microregion with excess risk of dengue fever (SMR$>$ 1), as urban areas are ideal environments for mosquitoes and many people living in close proximity create a human virus reservoir.

Insert Figure 2 here

Figure 3 illustrates that dengue has a strong annual cycle which differs with geographical zone. The spatially varying dengue annual cycle is included in

the model specified in section 3, as an interaction between the categorical variables zone and month. As only part of the cycle may be attributable to climatic conditions, the inclusion of this interaction could account for other confounding variables, such as seasonal population movements, leading to differences in the annual cycle across zones.

Insert Figure 3 here

Observed gridded ($2.5° \times 2.5°$ latitude-longitude grid) monthly mean precipitation data was obtained from the Global Precipitation Climatology Project (GPCP) (Adler et al., 2003). Reanalysis monthly mean surface air temperature data was obtained from the NCAR/NCEP Reanalysis (Kalnay et al., 1996). These climatic variables are referred to as 'observed' climate for the remainder of the text. Niño 3.4 is an index used to measure the strength of El Niño and La Niña events (Barnston et al., 1997) and is defined as the departure in monthly sea surface temperature from its long-term mean averaged over the region (120°W-170°W and 5°S- 5°N). A positive (negative) index indicates El Niño (La Niña) conditions. A time series of the monthly Niño 3.4 index was obtained from NOAA Climate Prediction Center[5].

Microregion and gridded data were combined by assigning a grid point to each microregion on the basis that the microregion is contained within the grid square (see Fig. 4).

Insert Figure 4 here

[5]http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices

3. Model selection using a generalised linear model

Poisson models are widely used in the analysis of count data. However, it is well established that observed count data e.g. disease cases, often display substantial extra-Poisson variation, or overdispersion (Lawless, 1987). Overdispersion was evident in this dengue dataset. Fitting a Poisson generalised linear model (GLM) involving the full set of explanatory variables described earlier results in a residual deviance more than a hundred times larger than the residual degrees of freedom, implying that as the mean dengue count increases, the variance increases at a much greater rate. In section 4 we will consider making explicit allowance for this overdispersion within the Poisson framework via the inclusion of appropriate random effects, but for model selection purposes within this section we accommodate overdispersion implicitly by using the negative binomial distribution for the observed counts, viz:

$$f(y;\mu,\theta) = \frac{\Gamma(y+\theta)}{\Gamma(\theta)y}\frac{\mu^y\theta^\theta}{(\mu+\theta)^{y+\theta}},$$

with mean μ, scale parameter θ and variance function $V(\mu) = \mu + \mu^2/\theta$. In a GLM context the associated canonical link is $g(\mu) = \log(\mu)$.

In order to select which explanatory variables are important for modelling dengue counts in Brazil for the 96 month time period (Jan 2001 - Dec 2008), the negative binomial GLM described above was fitted using the **MASS** package (Venables and Ripley, 2002) in `R` (R Development Core Team, 2008),

starting with a maximal model based on all of the covariates described in the previous section i.e. spatial covariates related to the urban environment, altitude, the annual cycle and interactions with geographical zone, observed climate variables with associated time lags (0-3 months) and the Niño 3.4 index with time lags of up to 6 months. Exploratory analyses were then carried out using different subsets of variables to select an appropriate prediction model (e.g. examining model fit with and without climate information and with different interactions). These analyses were assisted by use of stepwise model selection algorithms based on the Akaike information criterion (AIC) which not only rewards goodness of fit, but also includes a penalty that discourages overfitting. The final most parsimonious model which emerged from the investigation is as follows:

$$
\begin{aligned}
y_{it} &\sim \text{NegBin}(\mu_{it}\,, \theta) \\
\log(\mu_{it}) &= \log(e_i) + \alpha + \sum_j \beta_j x_{jit} + \sum_j \gamma_j w_{ji} + \sum_j \delta_j z_{jit},
\end{aligned}
$$

where y_{it} is dengue count for microregion $i = 1, \ldots, 558$ and time $t = 1, \ldots, 96$, μ_{it} is the corresponding mean dengue count and θ is the scale parameter. The expected cases $e_i = p_i r$ are treated as an offset in the model based on the population p_i in microregion i and the overall average dengue rate per month r. The variables x_{jit} represent the selected climate influences: precipitation one month previous ($j = 1$), precipitation two months previous ($j = 2$), temperature one month previous ($j = 3$), temperature two months previous ($j = 4$) and Niño 3.4 six months previous ($j = 5$). The variables w_{ji}

are: altitude ($j = 1$) and percentage of urban population ($j = 2$). Finally, z_{jit} is a series of factors reflecting zone, month and interaction between zone and month.

All covariate coefficients were found to be significantly different from zero at the p =0.001 level. The estimated parameters and standard errors for the climate variables included in the final model are listed in Table 1. Precipitation and temperature with time lags of 1 and 2 months were found to be the most statistically significant and are positively related to dengue as warm, humid conditions promote mosquito development and rain water fills discarded containers outdoors to create mosquito breeding sites in the months preceding increased dengue incidence. The Niño 3.4 index is negatively related to dengue. This is because the major dengue epidemics in 2002 and 2008 in particular, were preceded by negative SST anomalies in the Niño 3.4 region. The scale parameter θ was estimated to be 0.32 with standard error 0.002, confirming a mean variance relationship considerably different from that of the Poisson (equal mean and variance), hence justifying the use of a negative binomial rather than a Poisson GLM for model selection purposes.

One important aspect of such a model to a public health decision maker is its ability to predict dengue during the peak dengue season from February-April (FMA). In Figure 5, scatter plots with fitted loess curves show the relationship between observed and predicted dengue using the GLM model for the FMA season 2001-2008 for Brazil (Fig 5a) and the South East region where dengue predominates (Fig 5b). Although the model clearly fails to capture much of the variability in dengue counts in this season, there is an overall

Table 1: Parameter estimates for climate covariates.

Observed Climate	Coefficient estimate	Standard error	Prob$> |z|$
Precipitation lag 1	0.018	0.0037	5.12×10^{-4}
Precipitation lag 2	0.022	0.0036	6.45×10^{-11}
Temperature lag 1	0.091	0.0093	2×10^{-16}
Temperature lag 2	0.161	0.0093	2×10^{-16}
Niño 3.4 lag 6	-0.204	0.0119	2×10^{-16}

positive association between observed and predicted counts at both the national and regional level. The influence of the climate variables in the model predictions is demonstrated in Figure 6a which shows the time series of total observed dengue cases for the FMA season, predicted dengue using a GLM without any climate information (dotted line) and with climate information (dashed line). The climate variables are the only source of temporal information in the model, therefore by not including them the same prediction is produced for every month/season of each year. By including climate information, some of the temporal variability is captured albeit with limited skill. Figure 6c illustrates how the GLM predicts dengue for the FMA season in 2008. In some areas, the predicted dengue level corresponds to the observed level, for example, in coastal margins of the South East region (see Fig 6b). However, low levels of dengue are overestimated in the South and the model fails to reproduce the variability in dengue cases across the Amazon. When we focus in at the region level (South East) and microregion level (Rio de Janeiro) for which dengue early warnings would be most useful, time series of

dengue for the FMA season 2001-2008 show that the climate informed GLM fails to reproduce the dengue epidemic in 2002 and the increase in dengue from 2006-2007 (Fig 7a and b).

Insert Figure 5 here

Insert Figure 6 here

Insert Figure 7 here

This GLM clearly fails to capture much of the temporal variability in dengue counts, which may be attributable to factors such as population immunity to the dominant circulating serotype or specific health interventions and vector control measures. However, information regarding these aspects of the disease system are not readily available. Therefore, the use of unstructured random effects may be valuable to allow for unobserved latent structures in the model (McCulloch and Searle, 2004), for example, to capture the impact of unknown/unobserved confounding factors, such as the introduction of a new dengue serotype in a certain area of Brazil. Also, by using a GLM independence is assumed in both time and space and neither of these assumptions may be valid. There could be strong temporal correlation effects within some areas and there could also be spatial clustering effects in neighbouring microregions. To allow for such latent effects and correlation structures, the GLM is refined in the next section by reverting to a Poisson framework but using a generalised linear mixed model (GLMM) which includes spatially unstructured and spatio-temporal structured random effects in the linear predictor. This explicitly models the extra-Poisson variation or

overdispersion previously allowed for using the negative binomial.

We focus our analysis on the South East region of Brazil (see Figure 1a) where dengue is most prevalent and there are a large number of densely populated urban centres which could benefit from a climate informed dengue EWS. This is also the region where the previously reported GLM predictions did appear to capture some of the observed spatial variability in dengue counts (see Figure 6c).

4. Development of a generalised linear mixed model

As described above, we now focus on the 160 microregions in South East Brazil and return to a Poisson model for the dengue count data to develop a GLMM that includes random effects in the linear predictor. One approach to fitting such a model is to use a Bayesian framework. Markov Chain Monte Carlo (MCMC) methods make Bayesian modelling of complex situations involving many parameters a practical feasibility (see Gilks and Spiegelhalter (1996), Brooks (1998) for more details). One further advantage of the Bayesian approach is that the associated MCMC sampling yields full posterior predictive distributions which automatically incorporate all components of variance at the different levels in the model. A full assessment of prediction uncertainty is therefore more easily obtained with Bayesian estimation than with the more traditional maximum likelihood approach.

The inclusion of random effects introduces an extra source of variability (a latent effect) into the model to capture the impact of unknown/unobserved

confounding factors. For example, serotype introduction, which can vary spatially and temporally. Unstructured random effects can help account for overdispersion in the distribution of dengue counts y_i, however, this does not allow for explicit spatial dependence between y_i. This dependence can be included by adding a spatially structured random effect. A typical choice for a spatially structured prior is a conditional intrinsic Gaussian autoregressive model (CAR) (see Besag et al., 1995);

$$\nu_i|\nu_{j\neq i} \sim \mathrm{N}\left(\frac{\sum_{j\neq i} a_{ij}\nu_j}{\sum_{j\neq i} a_{ij}}, \frac{\sigma_\nu^2}{\sum_{j\neq i} a_{ij}}\right),$$

where a_{ij} are adjacency weights for the microregions, here taken to be simple binary values: $a_{ij} = 1$ if microregion i has a common boundary with microregion j, $a_{ij} = 0$ otherwise. The hyperparameter σ_ν controls the strength of the local spatial dependence. As the CAR is improper, a 'sum to zero' constraint is applied to ν_i and it is then advisable to take a uniform flat prior for the intercept α (see model specification below).

Models to predict vector-borne disease may include an autoregressive time series component (e.g. Gomez-Elipe et al., 2007), based on the idea that the current value of the time series y_{it} can be explained as a function of past values. Accordingly, a first order autoregressive temporal effect ω_t was included in the model, where t is calendar month and ω_1 (August) is set equal to zero in the model specification to avoid identifiability problems.

Therefore the spatio-temporal GLMM adopted is given by:

$$
\begin{aligned}
y_{it} &\sim \text{Pois}(\mu_{it}) \\
\log(\mu_{it}) &= \log(e_i) + \alpha + \sum_j \beta_j x_{jit} + \sum_j \gamma_j w_{ji} + \phi_i + \nu_i + \omega_t \\
\alpha &\sim \text{U}(-\infty, +\infty) \\
\phi_i &\sim \text{N}(0, \sigma_\phi^2) \\
\nu_i &\sim \text{CAR}(\sigma_\nu^2) \\
\omega_1 &= 0, \omega_t \sim \text{N}(\omega_{t-1}, \sigma_\omega^2), \ , t = 2, \ldots, 12.
\end{aligned}
$$

Independent diffuse Gaussian priors (mean 0, precision 1×10^{-6}) were taken for β_j $(j = 1, \ldots, 5)$ and γ_j $(j = 1, 2)$, whilst independent gamma hyperpriors with equal shape and inverse scale parameter (0.01) were used for the precisions ($\tau_\phi = 1/\sigma_\phi^2$, $\tau_\nu = 1/\sigma_\nu^2$, $\tau_\omega = 1/\sigma_\omega^2$) of the priors for the spatially unstructured ϕ_i and spatially structured ν_i random effects, $(i = 1, \ldots, 160)$, and the temporally autocorrelated random effects ω_t $(t = 2, \ldots, 12)$.

This model was fit with WinBUGS[6] software, using two parallel MCMC chains, each of length 25,000 with a burn-in of 20,000 and thinning of 10 to obtain samples from $P(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\omega}, \tau_\phi, \tau_\nu, \tau_\omega | \mathbf{y})$. As mentioned earlier, the explanatory variables x_{jit} $(j = 1, \ldots, 5)$ and w_{ji} $(j = 1, 2)$ are as before, however, all covariates are now standardised to zero mean and unit variance to aid convergence. This model is fit at the region level, therefore the zone factor is omitted as there is little variation in zone type in the South East region, and any geographic differences between microregions are captured by

[6] http://www.mrc-bsu.cam.ac.uk/bugs/

the spatial random effects. Satisfactory convergence was confirmed using a range of standard criteria (Gelman et al., 2004). Posterior distributions for each parameter associated with the climate covariates in the model are given in Figure 8 (with posterior means in parentheses). The climate parameters are all significantly different from zero and the sign of the association between dengue and each climate variable is consistent with the results from the GLM fit (Table 1).

Insert Figure 8 here

A posterior predictive distribution can be obtained for each microregion by drawing random samples from a Poisson distribution with mean equal to the MCMC samples from the model fit. The mean of the posterior predictive distribution for all microregions in the South-East region were obtained for the peak dengue season FMA. In Figure 9 a scatter plot with fitted loess curve shows the relationship between observed and predicted dengue using the GLMM model for the FMA season 2001-2008. When compared to predicted values from the GLM (see Fig 5b), the loess curve has shifted towards the 45° line and more of the variability in dengue cases has been captured by the prediction model. Figure 10 illustrates the spatial distribution of observed (Fig 10a), and predicted dengue for FMA season 2008 using both the GLM (Fig 10b) and GLMM (Fig 10c). While the GLM predicted medium levels of dengue across much of the region, the GLMM captures more of the observed variation. When compared with Figure 7, Figure 11 shows how the addition of random effects to the model has improved dengue predictions for both the South East as a whole (Fig. 11a) and for the microregion Rio de Janeiro

(Fig. 11b), particularly for the 2008 epidemic.

Insert Figure 9 here

Insert Figure 10 here

Insert Figure 11 here

5. Probabilistic epidemic prediction

The specified Bayesian hierarchical model can also be used to predict the probability of dengue exceeding a pre-defined epidemic threshold in each microregion. As the GLMM here provides a posterior predictive distribution for each microregion (rather than a point estimate), the probability of exceeding an epidemic threshold can be calculated and the decision to trigger an alert can be based on the probability of exceeding the threshold being greater than a specified alert level, (e.g. probability of exceedance $\geq 90\%$). Many epidemic detection algorithms have been investigated to detect epidemics (Cullen et al., 1984; Hay et al., 2002; Teklehaimanot et al., 2004). As an example, we consider the event of dengue exceeding an epidemic threshold of the mean plus one standard deviation for each microregion in South East Brazil in FMA 2008. The epidemic threshold is based on the dengue counts in the FMA season for the previous seven years (FMA 2001-2007). We can assess the ability of the GLMM to predict 'dengue epidemics' across South East Brazil during the FMA season in 2008 using a contingency table (see Table 2). Observed dengue counts for the 3-month season were compared

with model predictions where the probability of an epidemic exceeded an alert threshold chosen to be 90%.

Table 2: Contingency table for observed dengue exceeding epidemic threshold (mean plus one standard deviation) and probability of predicted dengue exceeding alert threshold ($\geq 90\%$) for the 160 microregions.

		Observed		
		Yes	No	Total
Predicted	Yes	a=29	b=9	38
probability	No	c=19	d=103	122
$\geq 90\%$	Total	48	112	160

The contingency table provides information on the overall predictive skill of the warning system. The proportion correct, defined as the proportion of the 160 microregions for which the prediction correctly anticipated the subsequent epidemic or non-epidemic ($a + d/a + b + c + d$), was 83%. The hit rate (the proportion of epidemics that were correctly predicted $a/a + c$) was 60%. Conversely, the false alarm rate (the proportion of epidemics that were predicted but did not occur $b/b + d$) was 8%.

Figure 12 shows the posterior predictive distribution for the FMA season 2008 for the microregion Linhares, found on the coastal region of Espírito Santo, where the probability of exceeding the epidemic threshold was found to be $\geq 90\%$, based on the epidemic threshold of mean plus one standard deviation derived from the distribution of dengue for the season FMA 2001-2007. A successful epidemic alert would have been issued for this microregion using

the GLMM with the given epidemic threshold and alert level. By lowering the alert level below 90% the hit rate for the region increases but so does the false alarm rate. In practice, the choice of epidemic threshold and alert level should be selected by decision makers based on expert opinion and available resources.

Insert Figure 12 here

6. Discussion and Conclusion

The preliminary modelling results in this paper indicate that climatic covariates play a statistically significant role in the transmission of dengue fever. Although climate information alone does not account for a large proportion of the overall variation in dengue cases in Brazil, spatio-temporal climate information with the addition of spatio-temporal random effects do account for some of this variability, particularly for the 2008 peak dengue season, when a serious epidemic occurred. Therefore the inclusion of seasonal climate forecasts in a dengue EWS for Brazil is worth investigating. The next step would be to assess the predictive validity of the model when replacing 'observed' with 'hindcast' (i.e. retrospective forecasts made for a historical period in pseudo-operational mode) climate variables. 'Hindcast' precipitation, temperature and Niño3.4 data are available from forecasting systems such as the UK Met Office seasonal forecasting system (Graham et al., 2005) and the European Centre for Medium Range Forecasts (ECMWF) System 3 (Anderson et al., 2007). These systems typically produce ensemble predictions with lead times up to 6 months. By replacing 'observed' with 'hindcast' climate

variables in the above GLMM, a dengue prediction could be made 5 months ahead of the dengue season of interest. For example, to predict dengue incidence for March 2010, the model could be run in October 2009 using the observed Niño 3.4 index for September 2009 (6 month lag), and precipitation and temperature forecasts for January and February 2010 issued in October 2009. The reliability of a climate-based EWS will depend on the skill of the forecasting system or multi-model combined and calibrated system such as EUROBRISA (Coelho et al., 2006), in predicting seasonal climate conditions for the region of interest.

Previous sections have highlighted the potential for incorporating climate information into a spatio-temporal EWS for dengue in Brazil. However, before implementing such an operational system several technical issues need to be considered. For example, the definition of epidemic thresholds by public health decision makers. Thresholds should be designed to minimise false alarms and false negatives (i.e. failing to predict that an epidemic will occur) and should correspond with the epidemic response capabilities in specific locations. The spatial scale of the system affects the type of response activity that could be implemented. For example, at the microregion level interventions such as health care provisions may be possible. However, vector-control efforts may be more difficult to target. Predictive output from an EWS needs to be continuously monitored and evaluated over time and models should be refitted as new dengue/climate data becomes available. Spatial demographic data from the census (and interim projections) should also be updated when necessary. In order to issue the most reliable epidemic predictions forecast climate should be replaced with observed climate as time progresses towards

the peak epidemic season, so that updated epidemic alerts can be re-issued to public health decision makers. However, time delays in obtaining and collating real-time information for both confirmed dengue cases from SINAN and climate forecasts and observations could hinder the ability to provide warnings far enough in advance. Another important consideration is the dissemination and visualisation of early warnings of increased level of dengue risk. It is vital to train public health decision makers on how to interpret and use dengue risk forecasts, including awareness about forecast limitations to avoid misinterpretation and/or over interpretation.

Developing a climate-based EWS for dengue using climate and disease information over such short time periods remains a major challenge. During the study period, the Niño 3.4 index strongly influences the temporal signal of predicted dengue. From June 2007, a moderate La Niña event developed, which strengthened in early 2008. Was the dengue epidemic in 2008 influenced by this La Niña event or was this a coincidence? ENSO may play a role in synchronizing epidemics, however, periods between epidemics may also be a function of herd immunity from previous epidemics, and these two cycles (ENSO and herd immunity) may have coincided during the 2001-2008 study period. Further investigation is needed to understand temperature and precipitation patterns associated with warm phase and cold phase ENSO for this region in Brazil and to consider the possibility of a non-linear relationship between precipitation/temperature and dengue. The model parameterisation would benefit from the inclusion of one or more past epidemics to address these problems.

Despite this, it is hoped that this spatio-temporal dengue prediction model is a step towards the development of a useful decision making tool for the Brazilian health services. Such spatio-temporal models offer an opportunity to balance global climate variables and local responses, e.g. the influence of ENSO on dengue incidence is likely to occur unequally across the region due to particular socio-economic local conditions. Another advantage of the GLMM is the ability to address specific public health issues in terms of probabilities. This model could be extended to other regions in the world where climate-sensitive infectious diseases (e.g. cholera, malaria, leptospirosis, plague) present a burden to public health infrastructure, particularly in developing countries.

7. Acknowledgments

This work was supported by the EUROBRISA (http://eurobrisa.cptec.inpe.br/) network project (F/00 144/AT) kindly funded by the Leverhulme Trust. GPCP Precipitation data and NCEP/NCAR Reanalysis surface temperature data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/. CASC was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), processes 2005/05210-7 and 2006/02497-6. RL would like to thank Evangelina Xavier Gouveia de Oliveira (IBGE) for kindly providing socioeconomic/geographical data and Renato Vitolo and Theo Economou for technical support and advice.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, S., Arkin, P., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). Journal of Hydrometeorology 4 (6), 1147–1167.

Anderson, D., Balmaseda, M., Stockdale, T., Ferranti, L., Vitart, F., Mogensen, K., Molteni, F., Doblas-Reyes, F., Vidard, A., 2007. Development of the ECMWF seasonal forecast System 3. ECMWF Technical Memorandum, 56 pp.

Barnston, A., Chelliah, M., Goldenberg, S., 1997. Documentation of a highly ENSO-related SST region in the equatorial Pacific. Atmosphere Ocean 35, 367–383.

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian Computation and Stochastic Systems. Statistical Science 10 (1), 3–41.

Braga, I. A., Valle, D., 2007. Aedes aegypti: histórico do controle no Brasil. Epidemiologia e Serviços de Saúde 16 (2), 113–118.

Brooks, S. P., 1998. Markov chain Monte Carlo method and its application. The Statistician, 69–100.

Brunkard, J. M., Cifuentes, E., Rothenberg, S. J., 2008. Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas-Mexico border region. Salud pública de México 50, 227–234.

519 Cazelles, B., Chavez, M., McMichael, A. J., Hales, S., 2005. Nonstationary
520 Influence of El Niño on the Synchronous Dengue Epidemics in Thailand.
521 PLoS Medicine 2 (4), 313–318.

522 Coelho, C. A. S., Stephenson, D. B., Balmaseda, M., Doblas-Reyes, F. J.,
523 van Oldenborgh, G. J., 2006. Toward an integrated seasonal forecasting
524 system for South America. Journal of Climate 19 (15), 3704–3721.

525 Cullen, J. R., Chitprarop, U., Doberstyn, E. B., Sombatwattanangkul, K.,
526 1984. An epidemiological early warning system for malaria control in north-
527 ern Thailand. Bulletin of the World Health Organization 62 (1), 107–114.

528 Eamchan, P., Nisalak, A., Foy, H. M., Chareonsook, O. A., 1989. Epidemi-
529 ology and control of dengue virus infections in Thai villages in 1987. The
530 American Journal of Tropical Medicine and Hygiene 41 (1), 95–101.

531 Favier, C., Degallier, N., Dubois, M. A., 2005. Dengue epidemic modelling:
532 stakes and pitfalls. Asia Pacific Biotech News 9 (No. 22), 1191–1194.

533 Gagnon, A., Bush, A., Smoyer-Tomic, K., 2001. Dengue epidemics and the
534 El Niño Southern Oscillation. Climate Research 19 (1), 35–43.

535 Gelman, A. C., Stern, H. S., Rubin, D. B., 2004. Bayesian data analysis.
536 Texts in Statistical Science., 668 pp.

537 Gilks, W. R., Spiegelhalter, D. J., 1996. Markov chain Monte Carlo in prac-
538 tice. Chapman & Hall/CRC.

539 Glantz, M., 2001. Currents of change: impacts of El Niño and La Niña on
540 climate and society. Cambridge Univ Press.

Gomez-Elipe, A., Otero, A., Van Herp, M., Aguirre-Jaime, A., 2007. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997 – 2003. Malaria Journal 6 (1), 129–139.

Goth, K. T., Ng, S. K., Chan, Y. C., Lim, S. J., Chua, E., 1987. Epidemiological aspects of an outbreak of dengue fever/dengue haemorrhagic fever in Singapore. Southeast Asian Journal of Tropical Medicine and Public Health 18 (3), 295–302.

Gould, D., Mount, G. A., Scanlon, J. E., Ford, H. R., Sullivan, M. F., 1970. Ecology and the control of dengue vectors on an island in the Gulf of Thailand. Journal of Medical Entomology 7 (4), 499–508.

Graham, R. J., Gordon, M., McLean, P. J., Ineson, S., Huddleston, M. R., Davey, M. K., Brookshaw, A., Barnes, R. T., 2005. A Performance Comparison of Coupled and Uncoupled Versions of the Met Office Seasonal Prediction GCM. Tellus A 57 (3), 320–339.

Gubler, D. J., 2002. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends in Microbiology 10 (2), 100–103.

Guzman, M. G., Kouri, G., 2003. Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges. Journal of clinical virology: the official publication of the Pan American Society for Clinical Virology 27 (1), 1–13.

Hales, S., Weinstein, P., Souares, Y., Woodward, A., 1999. El Niño and

the dynamics of vectorborne disease transmission. Environmental Health Perspectives, 99–102.

Hay, S. I., Simba, M., Busolo, M., Noor, A. M., Guyatt, H. L., Ochola, S. A., Snow, R. W., 2002. Defining and detecting malaria epidemics in the highlands of western Kenya. Emerging Infectious Diseases 8 (6), 555–562.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al., 1996. The NCAR/NCEP 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437–471.

Kuhn, K., Campbell-Lendrum, D., Haines, A., Cox, J., Corvalán, C., Anker, M., 2005. Using climate to predict infectious disease epidemics. World Health Organization, Geneva, 55 pp.

Kuno, G., 1995. Review of the factors modulating dengue transmission. Epidemiologic Reviews 17 (2), 321–335.

Lawless, J. F., 1987. Negative binomial and mixed Poisson regression. Canadian Journal of Statistics 15 (3), 209–225.

Li, C., Lim, T., Han, L., Fang, R., 1985. Rainfall, abundance of Aedes aegypti and dengue infection in Selangor, Malaysia. The Southeast Asian journal of Tropical Medicine and Public Health 16 (4), 560–568.

McCulloch, C., Searle, S., 2004. Generalized, linear, and mixed models. Wiley-Interscience.

584 Moore, C., Cline, B., Ruiz-Tiben, E., Lee, D., Romney-Joseph, H., Rivera-
585 Correa, E., 1978. Aedes aegypti in Puerto Rico: environmental determi-
586 nants of larval abundance and relation to dengue virus transmission. The
587 American Journal of Tropical Medicine and Hygiene 27 (6), 1225–1231.

588 Myers, M. F., Rogers, D. J., Cox, J., Flahault, A., Hay, S. I., 2000. Forecast-
589 ing disease risk for increased epidemic preparedness in public health. Adv
590 Parasitol 47, 309–30.

591 Nogueira, R. M. R., Araújo, J. M. G., Schatzmayr, H. G., 2007a. Dengue
592 viruses in Brazil, 1986-2006. Revista Panamericana de Salud Pública
593 22 (5), 358–363.

594 Nogueira, R. M. R., da Araújo, J. M. G., Schatzmayr, H. G., 2007b. Aspects
595 of dengue virus infections in Brazil 1986-2007. Virus Reviews and Research
596 12, 1–17.

597 R Development Core Team, 2008. R: A Language and Environment for Sta-
598 tistical Computing. R Foundation for Statistical Computing, Vienna, Aus-
599 tria, ISBN 3-900051-07-0.
600 URL `http://www.R-project.org`

601 Sanchez, L., Vanlerberghe, V., Alfonso, L., Marquetti, M., Guzman, M.,
602 Bisset, J., van der Stuyft, P., 2006. Aedes aegypti larval indices and risk
603 for dengue epidemics. Emerging Infectious Diseases 12, 800–806.

604 Schreiber, K., 2001. An investigation of relationships between climate and
605 dengue using a water budgeting technique. International Journal of Biome-
606 teorology 45 (2), 81–89.

607 Scott, T., Amerasinghe, P., Morrison, A., Lorenz, L., Clark, G., Strickman,
608 D., Kittayapong, P., Edman, J., 2000. Longitudinal studies of Aedes ae-
609 gypti (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding
610 frequency. Journal of Medical Entomology 37 (1), 89–101.

611 Teklehaimanot, H. D., Schwatrz, J., Teklehaimanot, A., Lipsitch, M., 2004.
612 Alert threshold algorithms and malaria epidemic detection. Emerging In-
613 fectious Diseases 10 (7), 1220–1226.

614 Thomson, M., Doblas-Reyes, F., Mason, S., Hagedorn, R., Connor, S., Phin-
615 dela, T., Morse, A., Palmer, T., 2006. Malaria early warnings based on
616 seasonal climate forecasts from multi-model ensembles. Nature 439 (7076),
617 576–579.

618 Venables, W. N., Ripley, B. D., 2002. Modern Applied Statistics with S, 4th
619 Edition. Springer, New York, iSBN 0-387-95457-0.
620 URL `http://www.stats.ox.ac.uk/pub/MASS4`

621 Wu, P., Guo, H., Lung, S., Lin, C., Su, H., 2007. Weather as an effective
622 predictor for occurrence of dengue fever in Taiwan. Acta Tropica 103 (1),
623 50–57.

Figure captions

Figure 1: (a) Monthly dengue fever counts (1 000 cases) for main regions of Brazil from January 2001 to December 2008 (b) map to show main regions of Brazil (c) map of total dengue cases in each microregion (558) in Brazil for period January 2001 to December 2008.

Figure 2: Upper panel: spatial distribution of (a) altitude, (b) urban population, (c) zones in Brazil. Lower panel: scatter plot and loess curve to show relationship between dengue SMR and (d) altitude, (e) percentage of urban population, (f) boxplots to show distribution of dengue SMR in each zone. Note logarithmic y axes.

Figure 3: Annual cycle of dengue for 8 zones in Brazil, calculated for period January 2001 to December 2008.

Figure 4: Map to show centroids of microregions in Brazil (circles) and $2.5^\circ \times 2.5^\circ$ climate grid (squares). Box indicates approximate location of South East region for which GLMM is developed.

Figure 5: Scatter plot and loess curve (solid line) to show observed and predicted dengue fever, using GLM model for 3 month season FMA 2001-2008 for (a) Brazil and (b) South East region.

Figure 6: (a) time series of total observed dengue (solid line), GLM predicted dengue without climate (dashed line) and GLM predicted dengue with climate (dotted line) for FMA season 2001-2008 in Brazil, maps to show sum of (b) observed and (c) predicted dengue cases for microregions of Brazil, FMA

season 2008. Categories defined by quintiles of observed dengue for FMA season 2008.

Figure 7: Time series of total observed dengue (solid line), GLM predicted dengue without climate (dashed line) and GLM predicted dengue with climate (dotted line) for FMA season 2001-2008 for (a) South East (region level) and (b) Rio de Janeiro (microregion level).

Figure 8: Kernel density estimates for marginal posterior distributions of parameters $\beta_1, \ldots, \beta_5$ (posterior means in parentheses) associated with climate variables: (a) precipitation lag 1, (b) precipitation lag 2, (c) temperature lag 1, (d) temperature lag 2 and (e) Niño 3.4 index lag 6 for South East Brazil.

Figure 9: Scatter plot and loess curve (solid line) to show observed and predicted dengue fever using GLMM for 3 month season FMA 2001-2008 for South East Brazil.

Figure 10: Maps to show (a) observed dengue, (b) predicted dengue using GLM model and (c) predicted dengue using GLMM model for South East, FMA season 2008. Categories defined by quintiles of observed dengue for FMA season 2008.

Figure 11: Time series of total observed dengue (solid line) and predicted dengue using GLMM (dashed line) for FMA season 2001-2008 for (a) South East (region level) and (b) Rio de Janeiro (microregion level).

Figure 12: Kernel density estimate for posterior predictive distribution of dengue, FMA 2008 for Linhares (19.4°S,40.1°W), a microregion in Espírito

668 Santo. Dashed vertical line indicates epidemic threshold of mean plus one
669 standard deviation based on FMA 2001-2007. Solid vertical line indicates
670 observed dengue count in FMA 2008.

Figure 1

Click here to download high resolution image

Figure 2

Click here to download high resolution image

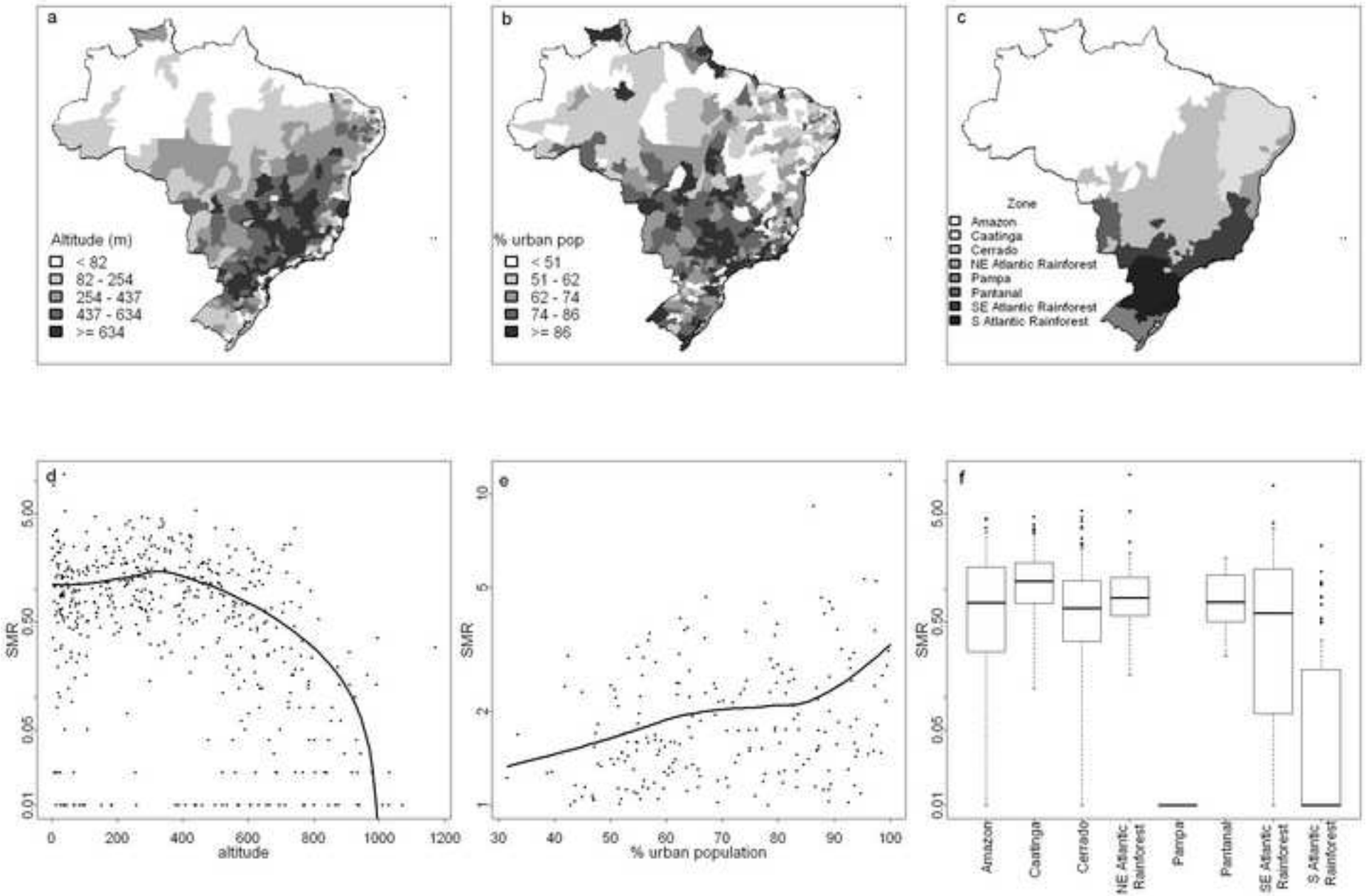


Figure 3

Click here to download high resolution image

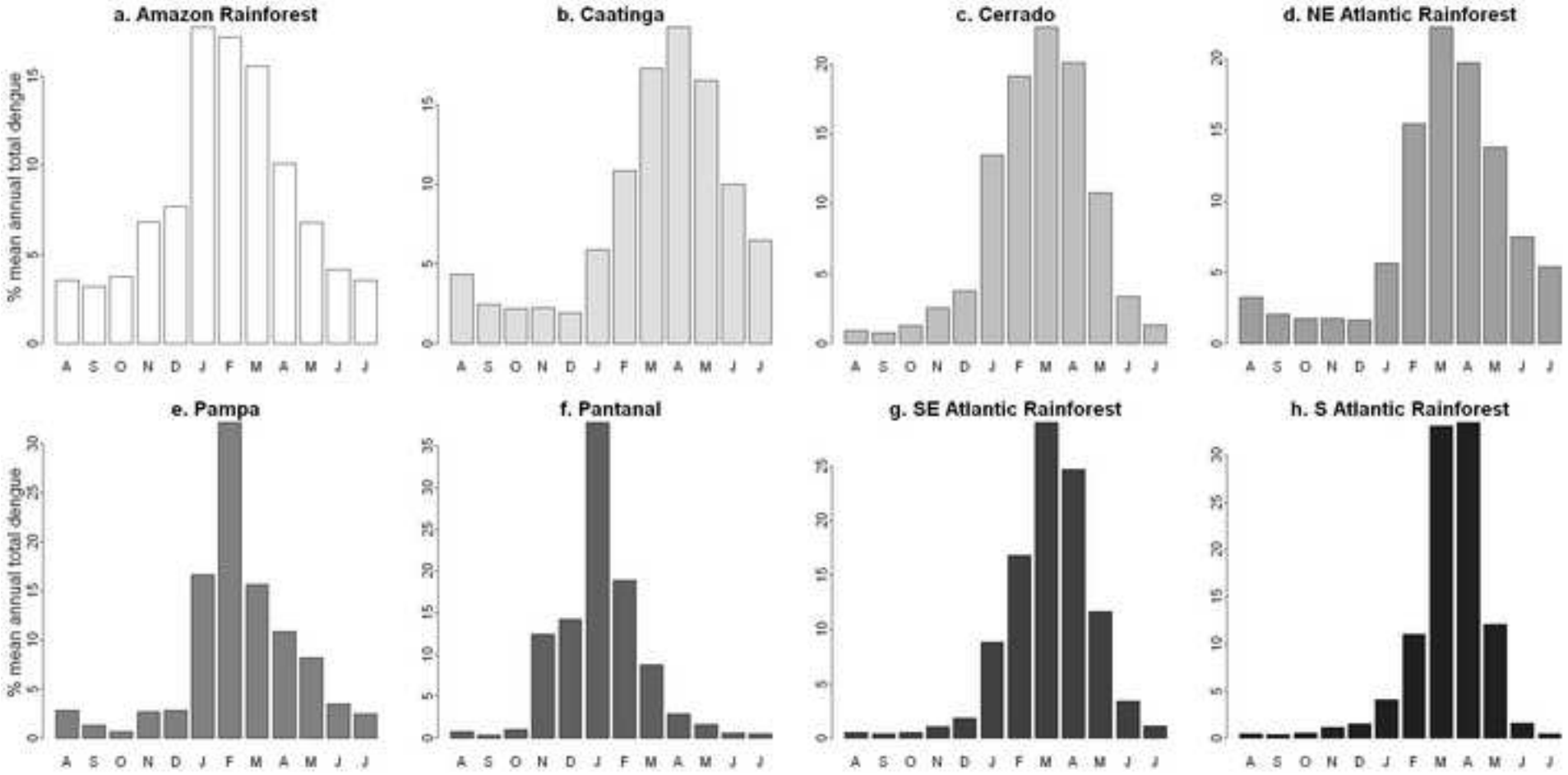


Figure 4

Click here to download high resolution image

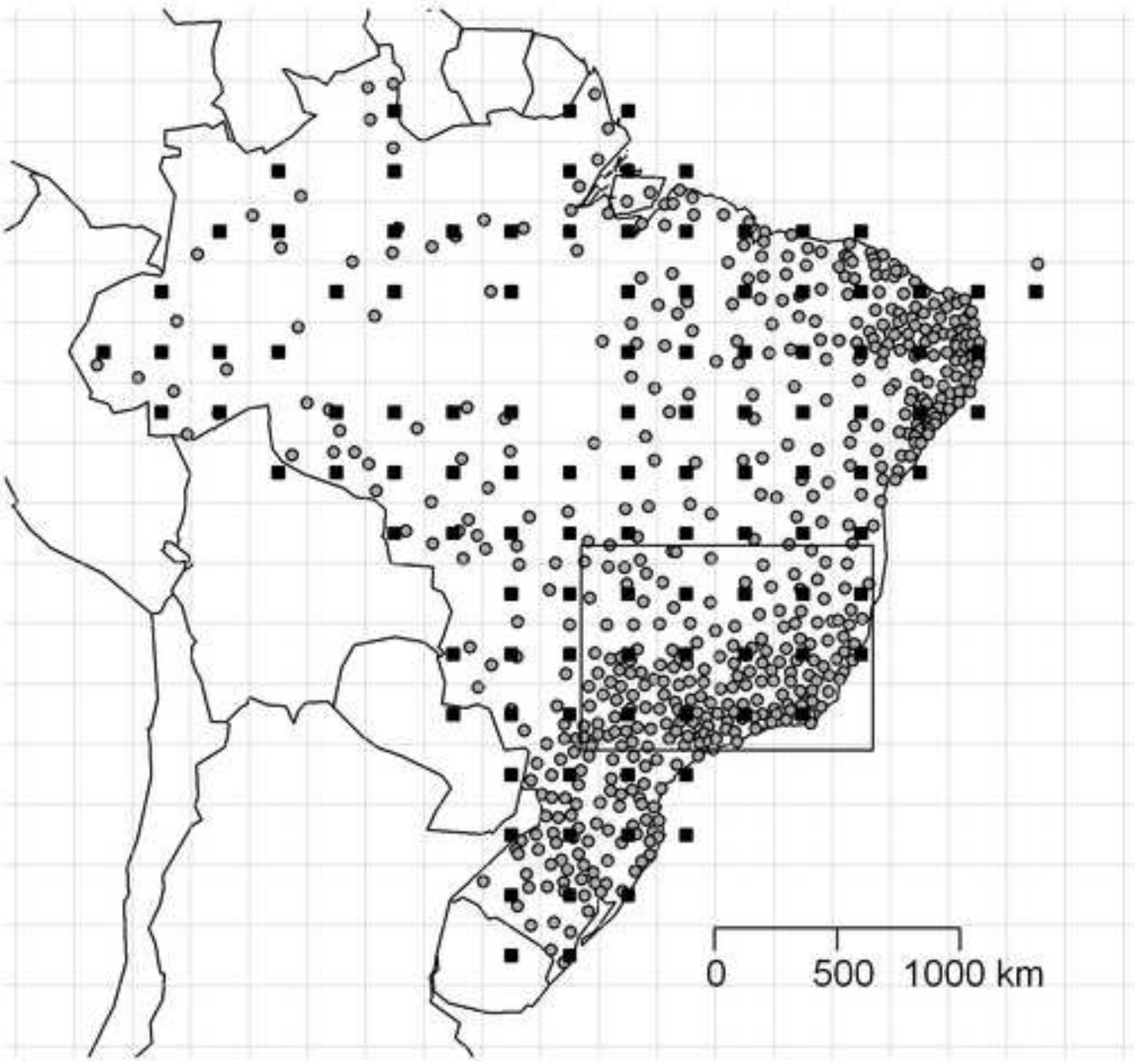


Figure 5

Click here to download high resolution image

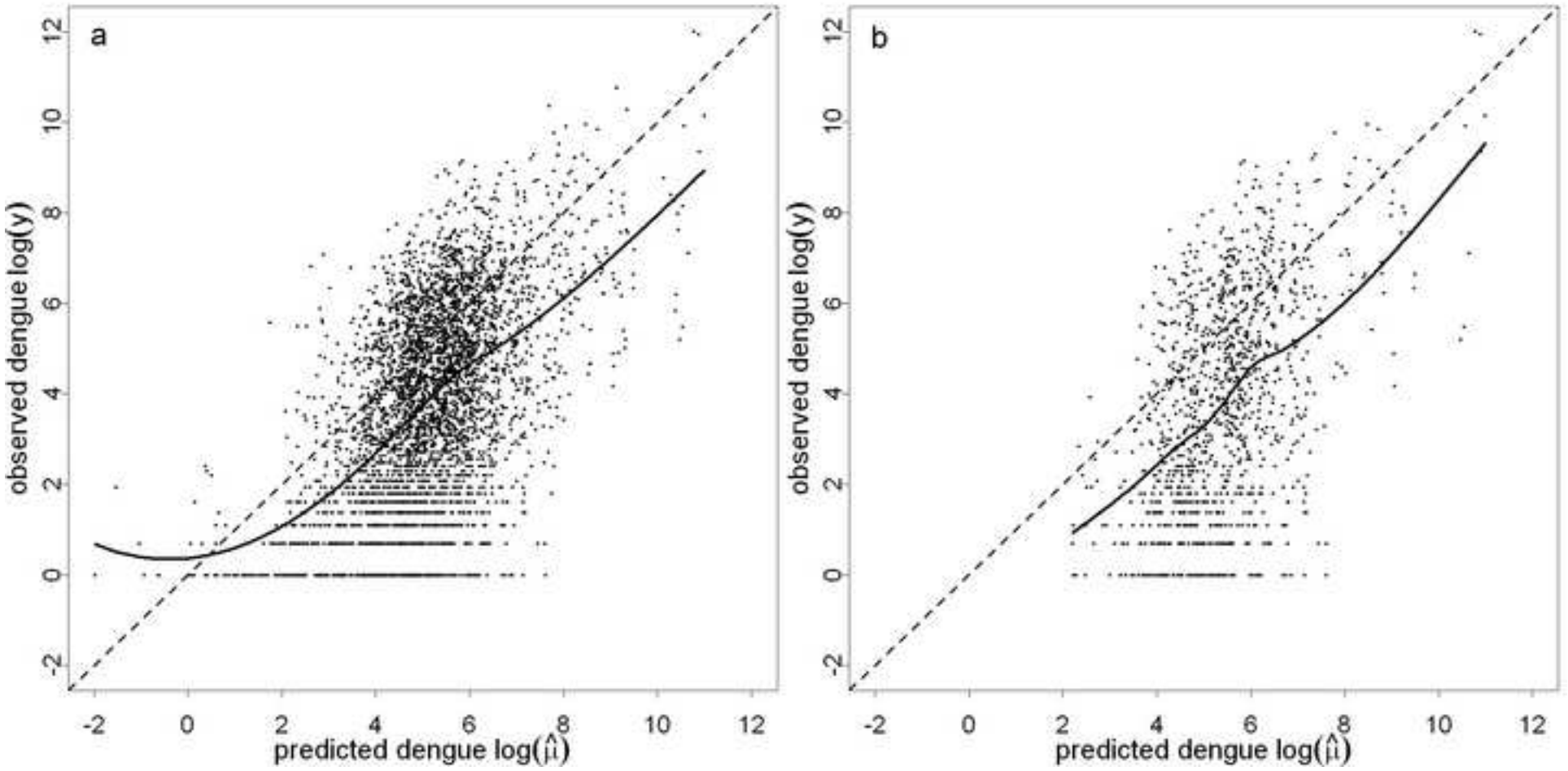


Figure 6

Click here to download high resolution image

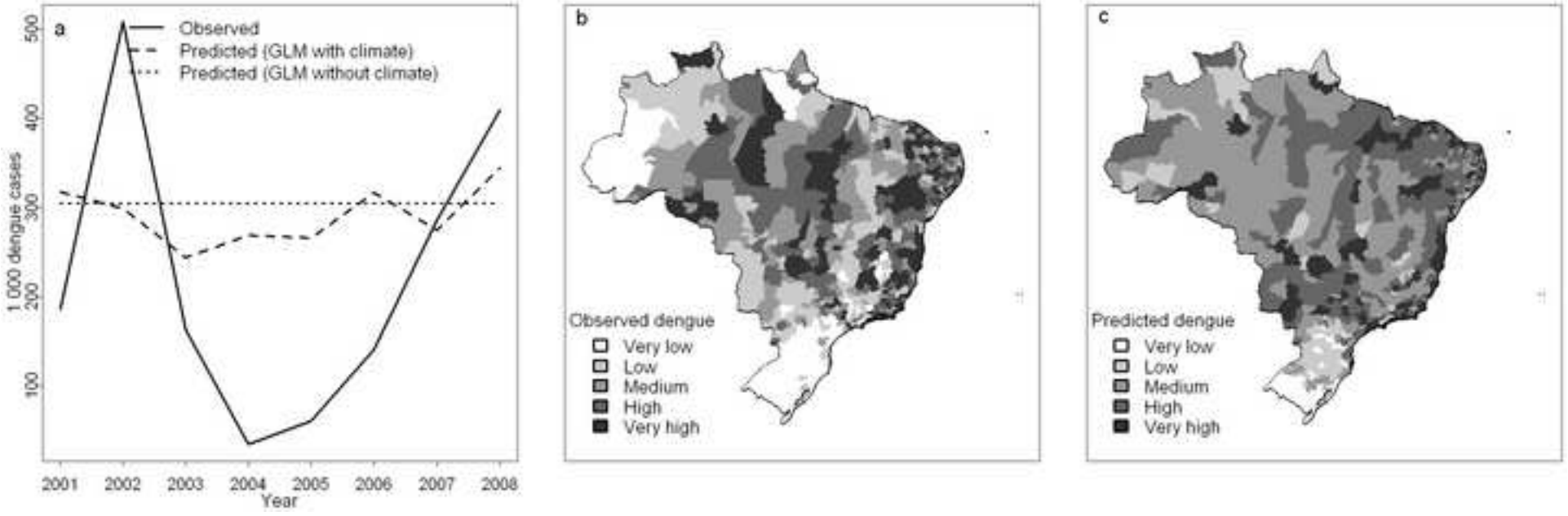


Figure 7

Click here to download high resolution image

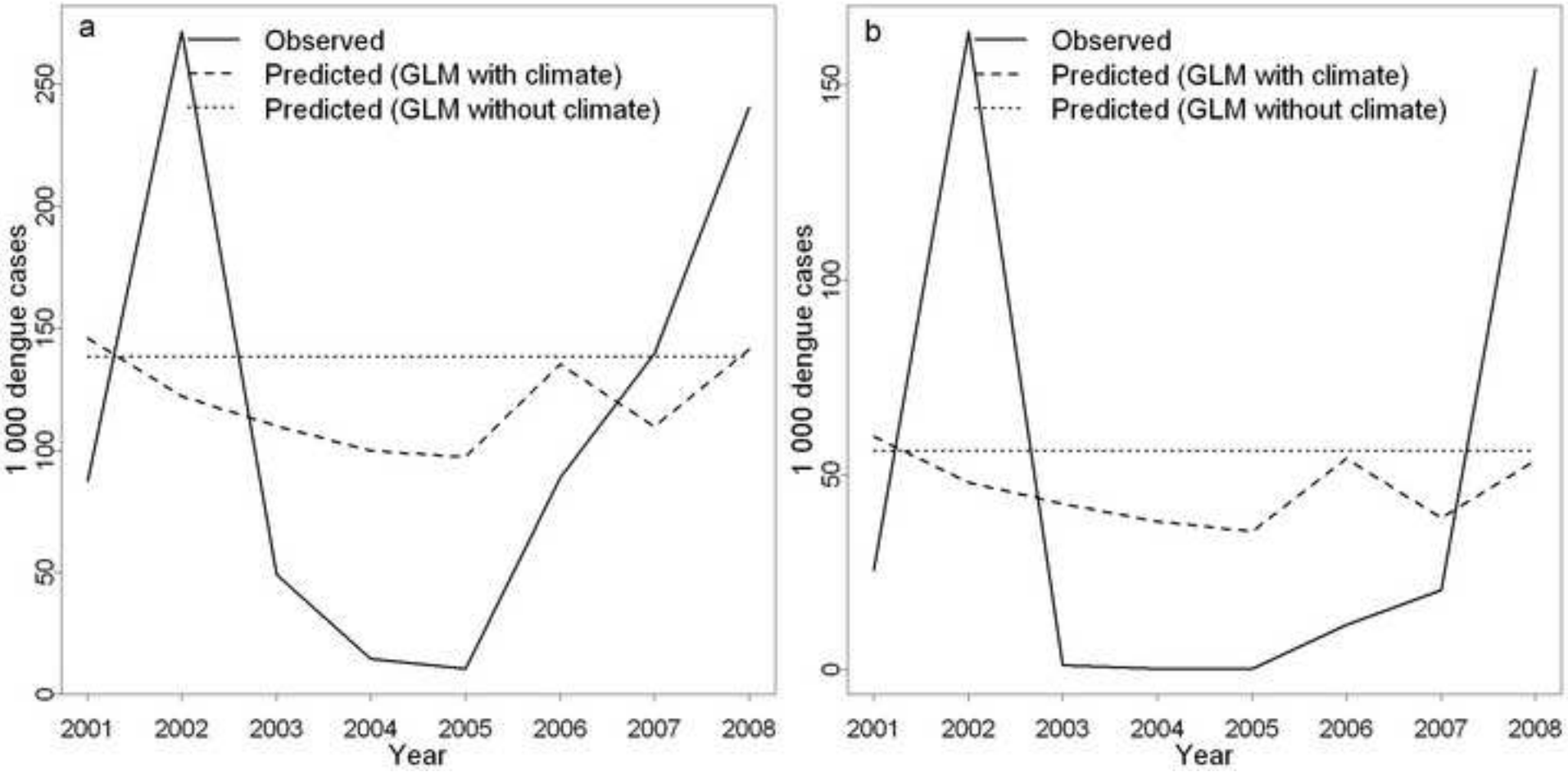


Figure 8

Click here to download high resolution image

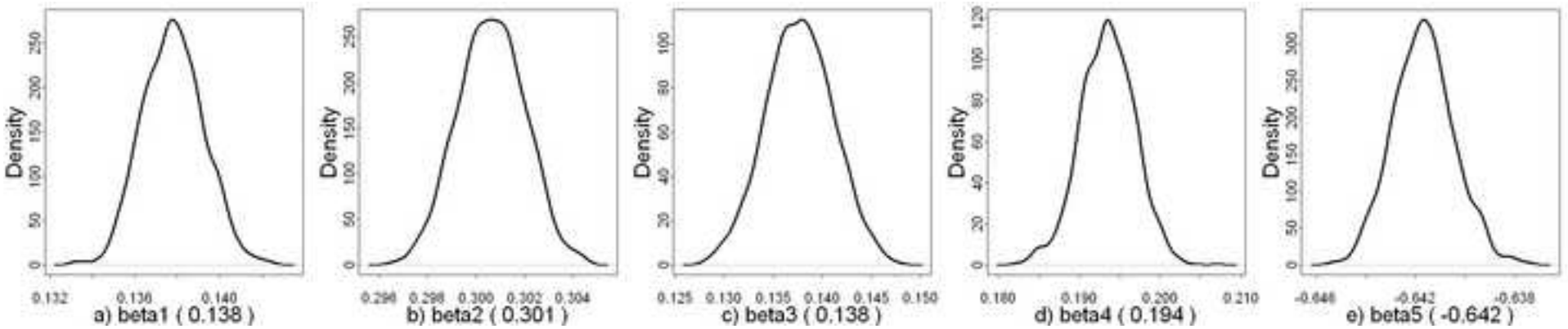


Figure 9

Click here to download high resolution image

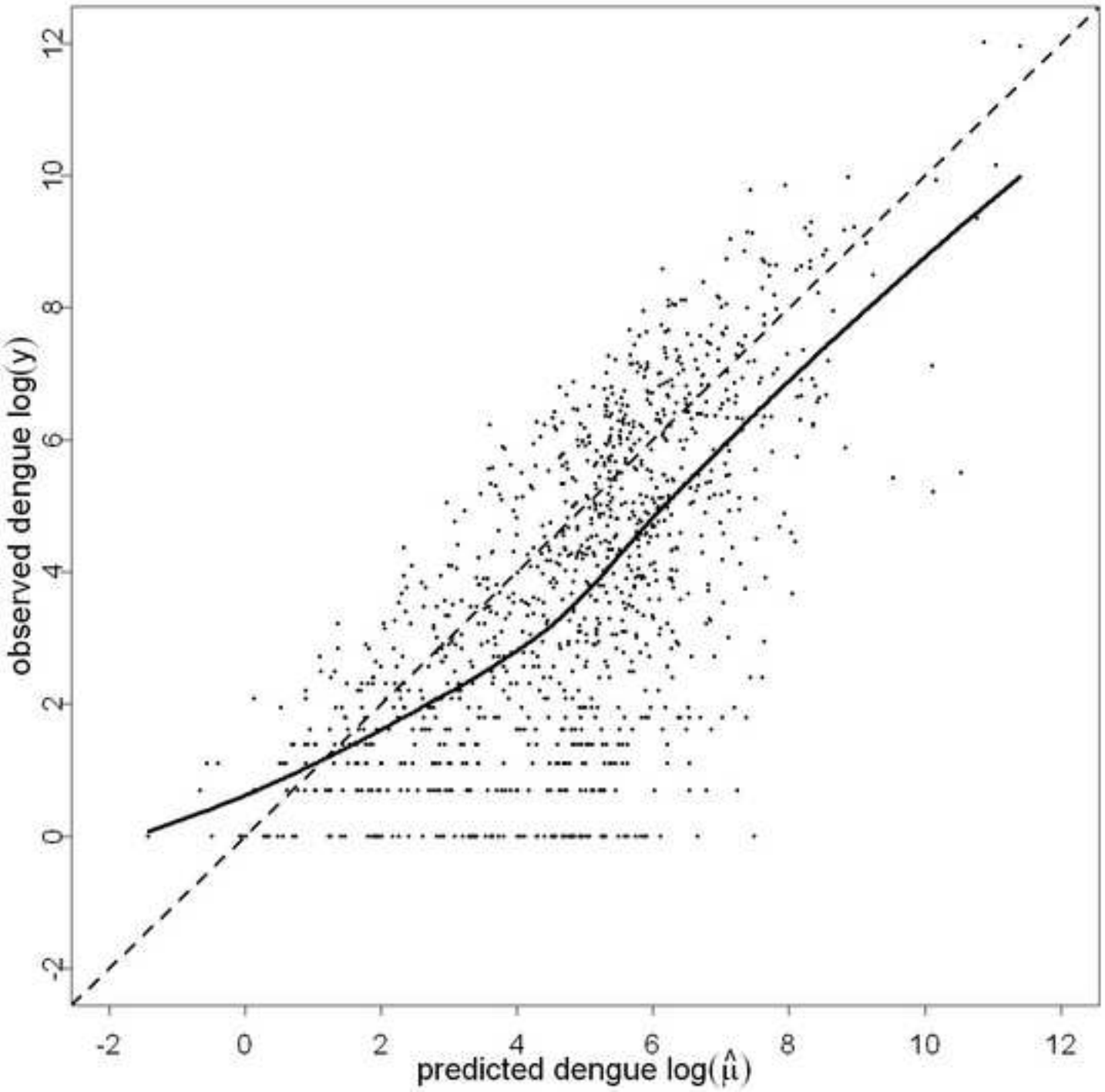


Figure 10

Click here to download high resolution image

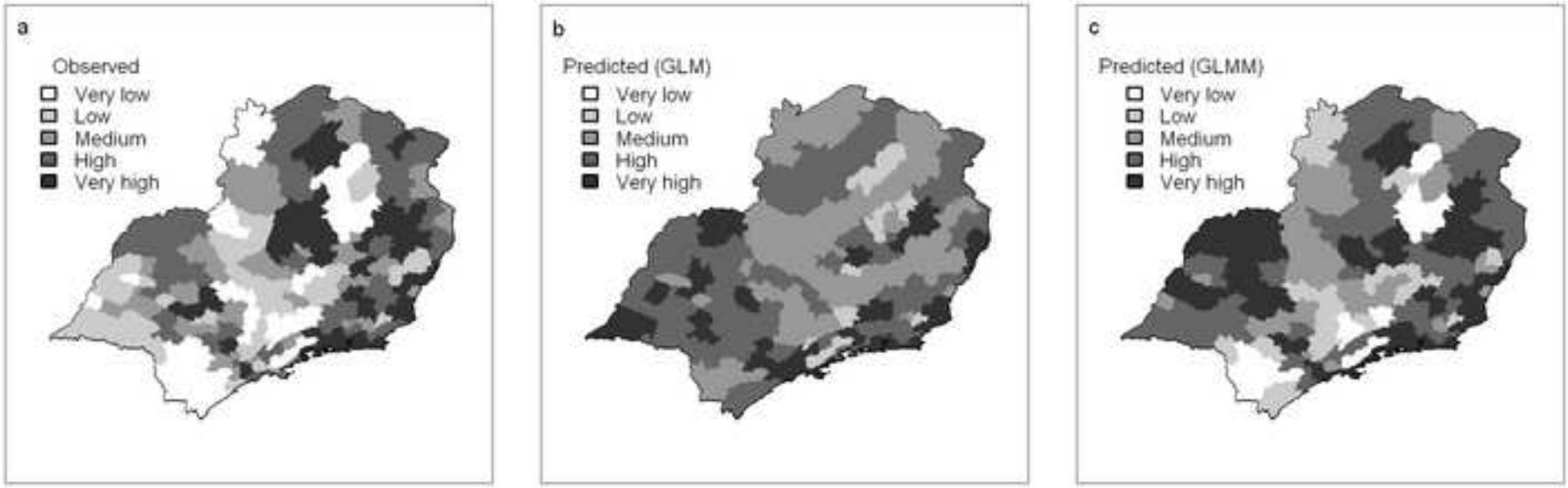


Figure 11

Click here to download high resolution image

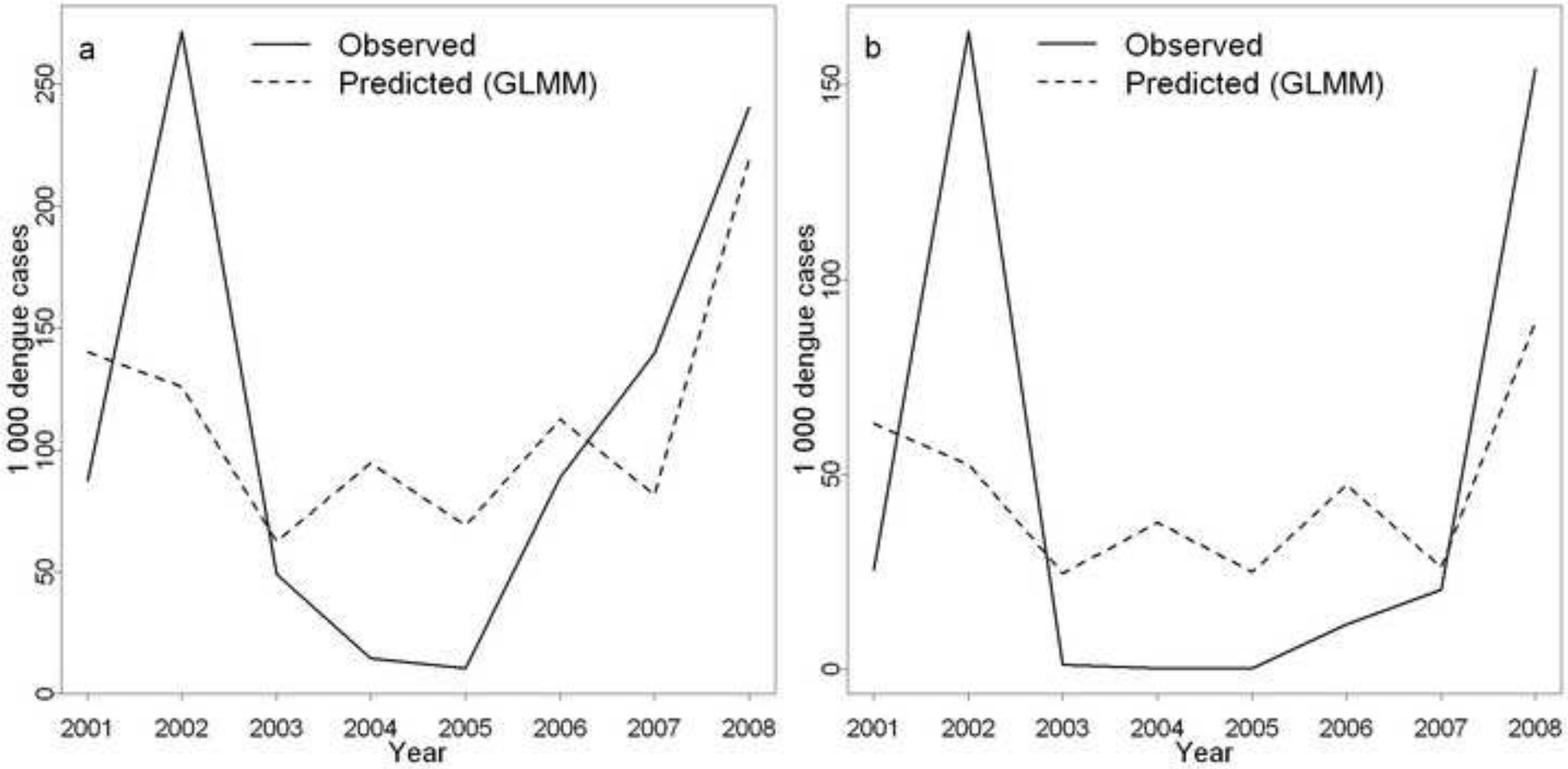
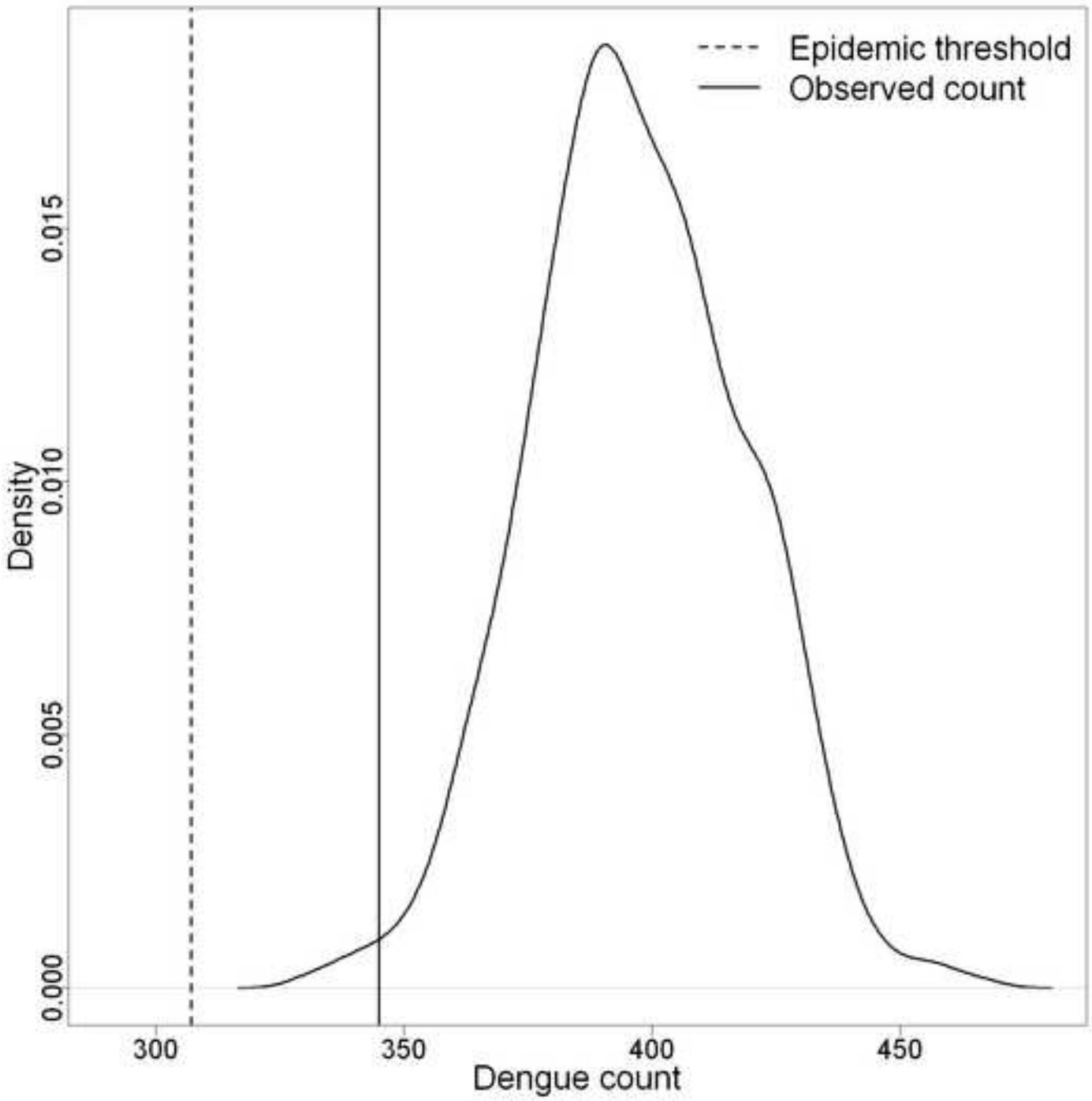


Figure 12

Click here to download high resolution image