Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts.

Appendix

Model formulation

To model monthly dengue cases (January 2000- December 2013) in the 553 microregions of Brazil, a generalised linear mixed model (GLMM) was formulated,[1,2] where for each spatial area, or microregion, (s=1,..., 553) and monthly time step, (t=1,..., 168), the count of dengue cases, y_{st}, follows a negative binomial distribution with an unknown scale parameter, κ, and mean, μ_{st} . A negative binomial model formulation was adopted to allow for extra-Poisson variation (overdispersion) in the observed dengue counts caused by unknown confounding factors and possible correlations in both time and space. The expected number of cases, e_{st}, was included as an offset (based upon the population of microregion (s) at time (t) multiplied by the global dengue rate for the whole dataset).

Following Lowe et al.[2] the climate variables considered were three-month average temperature and precipitation anomalies (departures from the long-term average),[3,4] over the three months preceding the dengue month of interest. This is equivalent to a two month lag when considering the mid-point of the three month average. As our model is designed as an early warning system, this aligns with the fact that temperature and precipitation data are, in practice, obtained from seasonal climate forecasting systems, which are typically issued as seasonal (e.g. March-April-May) forecasts (http://eurobrisa.cptec.inpe.br). Important non-climatic explanatory variables included in the model were altitude and population density. Altitude was found to have a significant negative association with the dengue relative risk, whereas population density was positively associated.

As a surrogate for spatio-temporal confounding factors (e.g. increased mosquito populations or the re-introduction of a dengue serotype), lagged dengue incidence was considered. The log ratio of observed to expected dengue cases (dengue relative risk) four months ahead for each microregion was included in the model. For example, a dengue prediction for June is based on the dengue risk reported in the previous February. This lag was chosen as a compromise between the longest lag plausible to provide predictive skill and the shortest lag possible to allow enough time to provide an early warning (in line with the lead-time gained from using seasonal climate forecasts). As the inclusion of an autoregressive term causes the first four observations in each microregion to be lost, we fitted the model to the dataset for the period May 2000 - December 2013 (164 months).

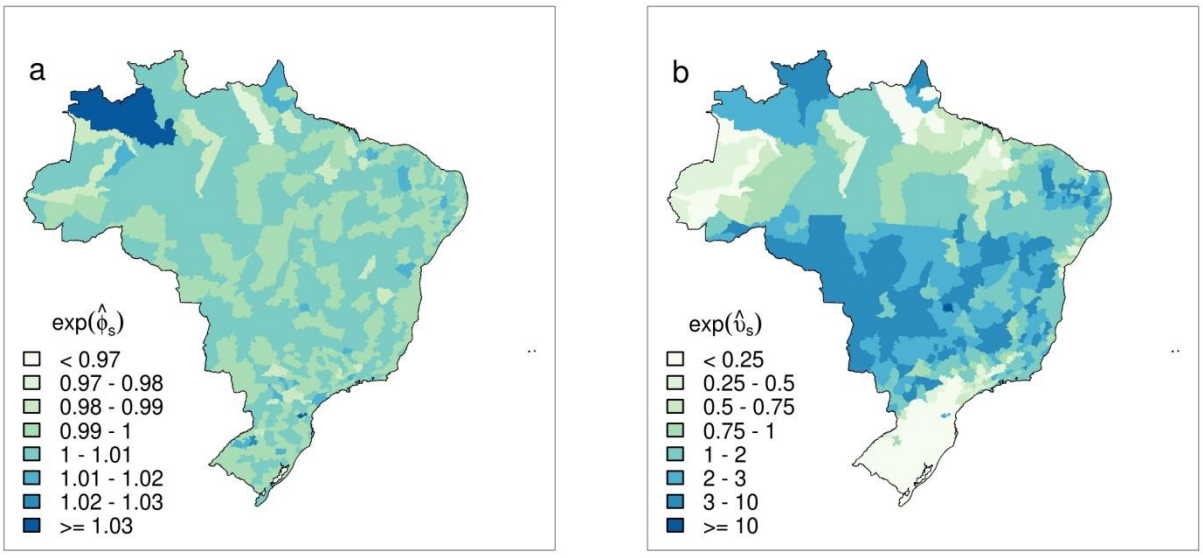

Figure S1: Multiplicative contribution of spatial random effects to dengue relative risk.
Spatial distribution of posterior mean (a) unstructured and (b) structured random effects.

We included unstructured random effects in the hierarchical model framework to help account for unknown or unobserved confounding factors in the disease system, such as population immunity, quality of healthcare services and local health interventions. Such random effects introduce an extra source of variability (a latent effect) into the model, which can assist in modelling overdispersion, in addition to the single-scale parameter in the negative binomial model. To allow for correlated heterogeneity between microregions or clustering, structured random effects were included. We imposed a spatial dependency structure by assuming a conditional intrinsic Gaussian autoregressive (CAR) model prior distribution[5] for the spatial effects, which takes the neighbourhood structure of the area under consideration into account (Fig. S1).

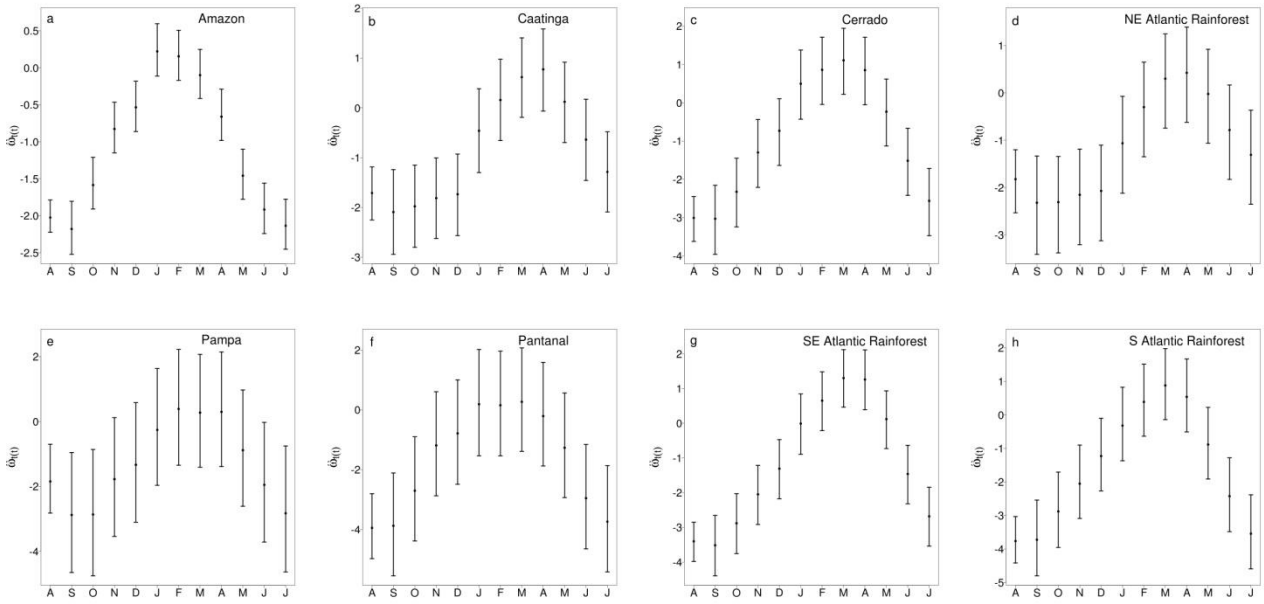

Figure S2: Posterior distributions of zone-specific autocorrelated annual cycle effects.
Posterior distribution of the autocorrelated annual cycle (from August to July) for zones: (a) Amazon, (b) Caatinga, (c) Cerrado, (d) North East Atlantic Rainforest (e) Pampa, (f) Pantanal, (g) South East Atlantic Rainforest and (h) South Atlantic Rainforest.

Brazil can be divided into six distinct ecological biomes: Amazon Rainforest, Caatinga, Cerrado, Atlantic Rainforest, Pampa, and Pantanal. A spatial variable named zone was defined according to the six biomes but with the Atlantic Rainforest biome additionally subdivided into three areas (North East, 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 and drier. Therefore, eight zones are defined for which climatic, geographical and ecological conditions are approximately homogeneous. In the model framework, zone is treated as an eight level factor with the Amazon (zone 1) set as the reference level. Dengue has a marked annual cycle in Brazil that varies between the different zones. To allow for this, a first-order autoregressive month effect[6] was interacted with zone, with August (month 1) set as the reference level (Fig. S2).

The final model comprised a combination of lagged climate variables and dengue risk, non-climate covariates and spatially and temporally structured and unstructured random effects. We formulated the model as a Bayesian GLMM as follows:

$$y_{st} \sim \text{NegBin}(\mu_{st}, \kappa)$$
$$\log(\mu_{st}) = \log(e_{st}) + \alpha + \alpha_{s'(s)} + \Sigma\beta_j x_{jst} + \Sigma\gamma_j w_{jst} + \delta z_{st} + \varphi_s + \upsilon_s + \omega_{t'(t)} + \omega_{t'(t),s'(s)}\,.$$

A Negative Binomial as opposed to a Poisson distribution was chosen for the dengue case counts to allow for observed global overdispersion in these counts (i.e. variance of counts being substantially in excess of their mean values in all months and all microregions). The spatially unstructured random effects in the model (see below) then additionally allow for local spatial variations in this overdispersion. The variables x_{jst}, represent the selected climate influences: precipitation (j=1) and temperature (j=2) anomalies (observed climate datasets used to estimate parameters). The variables w_{jst} are altitude (j=1) and population density (j=2). Variable z_{st} is the log ratio of observed to expected dengue cases 4 months previously, $z_{st}=\log(y_{st-4}/e_{st-4})$. Spatial random effects are composed of unstructured φ_s and structured

components υ_s . The spatially unstructured random effects, φ_s, are assigned independent diffuse Gaussian exchangeable prior distributions. The structured random effects, υ_s, are assigned a CAR model prior (see Lowe et al.[2] for details). We included a first-order autoregressive month effect $\omega_{t'(t),s'(s)}$ for each zone (t'(t)=2,...,12 , s'(s)=2,...,8), with month 1 (August) and zone 1 (Amazon) aliased to the model intercept, α, and subsequent months following a random walk or first difference prior in which each effect is derived from the immediately preceding effect. Note, the fixed effect for month t'(t)=1 and zone s'(s)=2,...,8, is denoted as $\alpha_{s'(s)}$. The random month effect for the Amazon s'(s)=1 and month t'(t)=2,...,12, is denoted as $\omega_{t'(t)}$.

We used independent diffuse Gaussian priors (mean 0, precision 1 x 10^{-6}) for the fixed effects ($\alpha_{s'(s)}$, β_j , γ_j , δ) and weakly informative independent gamma hyperpriors (shape parameter 0.5, inverse scale parameter 0.0005) for the scale parameter κ and for the precisions for the spatial and temporal random effects. These choices are similar to those used in previous modelling of dengue in Brazil reported in Lowe et al.[2] and represent relatively standard choices for a spatio-temporal GLMM such as that used in this study. Sensitivity of results to the choice of prior and hyperprior parameters was assessed by comparing runs with a suitably chosen range of both less and more informative choices and revealed that final models results were robust provided an appropriately long burn-in was allowed for in each case. The final choices represented a compromise between not wishing to be too informative whilst at the same time not wanting to be inefficient in the length of Markov Chain Monte Carlo (MCMC) runs required to achieve convergence. We fitted the Bayesian model via MCMC sampling using R in conjunction with the WinBUGS software[7] and the R2WinBUGS package.[8] We generated two parallel MCMC chains, each of length 25,000 with a burn-in of 20,000 and thinning of 10 to obtain 1000 samples from the joint posterior distribution.

Producing the forecast

Posterior predictive distributions[9] of June 2014 dengue incidence were obtained for each microregion. Samples of parameter estimates from the model fitted to May 2000 – December 2013 (e.g. Fig. S3) were combined with new data for dengue relative risk for February 2014 and seasonal forecasts of precipitation and temperature anomalies[10] for March-April-May 2014 (and other spatial covariates), to calculate the posterior distribution of the mean dengue count, $\boldsymbol{\mu}_{st}$. Subsequently, the posterior predictive distribution of dengue counts, $\mathbf{y}_{st}$, for each microregion was estimated by drawing random values from a negative binomial distribution with mean corresponding to the elements of $\boldsymbol{\mu}_{st}$ and scale parameter corresponding to the elements of $\boldsymbol{\kappa}$, estimated from the model.

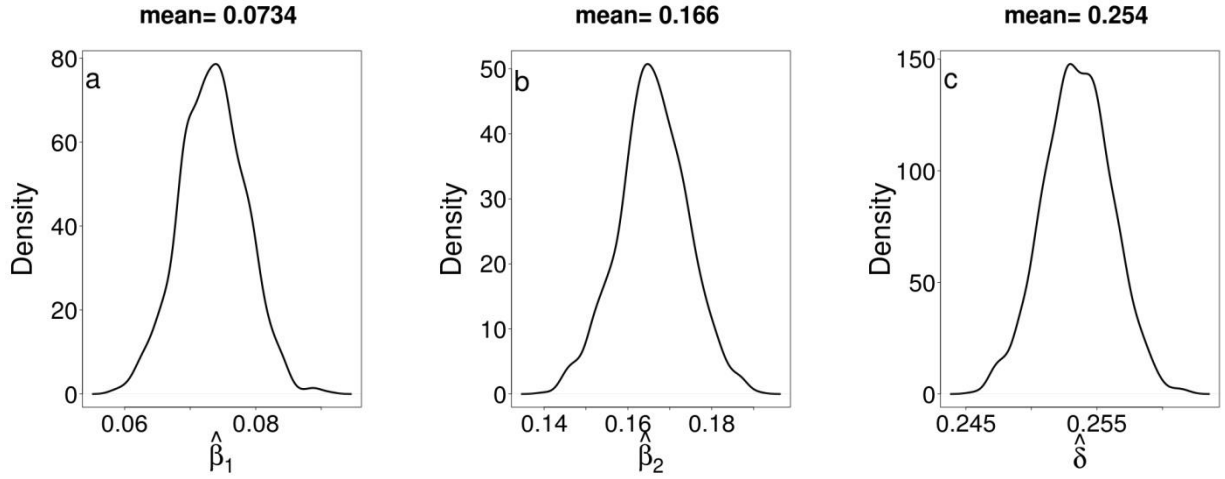

Figure S3: Kernel density estimates for dengue model drivers.

Kernel density estimates of marginal posterior distributions for the parameters associated with the variables driving the spatio-temporal dengue predictions: (a) precipitation and (b) temperature anomalies (averaged over previous 3 months) and (c) dengue risk four months earlier. Estimates from the model fitted to data 2000-2013.

Communicating ternary probabilistic forecasts

To present the forecast to decision makers, we calculated the probability of dengue incidence rates (DIR) falling into pre-defined categories. The Brazilian Ministry of Health is interested in areas where DIR < 100, indicating low risk; 100 < DIR < 300, indicating medium risk; and DIR > 300, indicating high risk. We used a novel visualisation technique[11] to produce a map in which the forecast at each microregion is expressed as a colour determined by a combination of three probabilities (see Fig. 1 in Article). We considered a ternary forecast as a point in a triangle of barycentric coordinates. This allows a unique colour to be assigned to each forecast from a continuum of colours

defined on the triangle. Colour saturation increases with information gain relative to the benchmark (reference) forecast (see Fig. S4). In this case, the benchmark distribution of June dengue incidence rates 2000-2013 was p(low=68%, medium=16%, high=16%) (Fig. S5). Posterior predictive distributions of DIR for individual microregions (i.e. the twelve microregions where the stadiums are located) were simulated, to show the probability distribution between low, medium and high risk categories for June 2014 (see Fig. S6 and table in Article).

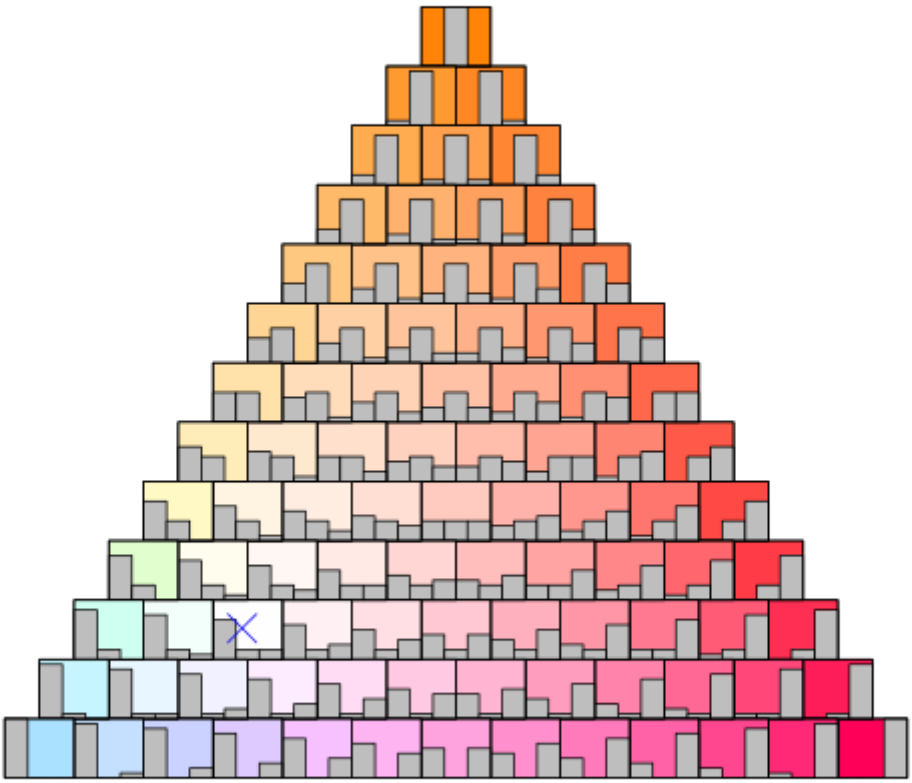

Figure S4: Palette of colours assigned to ternary forecasts.
Triangle with palette of colours assigned to ternary forecasts of low, medium and high dengue risk. The triangle is divided in 91 squares. For each square three bars are plotted indicating the probabilities for low (first bar), medium (second bar) and high (third bar) dengue risk. For example, for the central square in the middle of the triangle the three bars have the same height indicating equal (33.33%) probabilities for low, medium and high dengue risk. The square at the bottom left corner of the triangle corresponds to 100% probability of low risk and 0% probability for both medium and high risk. The square at the bottom right corner of the triangle corresponds to 100% probability of high risk and 0% probability for both medium and low risk. The square at the top corner of the triangle corresponds to 100% probability of medium risk and 0% probability for both low and high risk. The long-term average distribution of June dengue incidence rates (i.e. probability of low risk equal to 68%, probability of medium risk equal to 16%, and probability of high risk equal to 16%), indicated by the blue cross, is assigned the colour white. The blue colour in the triangle indicates that low risk is most likely. The pink colour in the triangle indicates that high risk is most likely. The orange colour in the triangle indicates that medium risk is most likely. The more saturated the colour, the more certain the model that a particular category will be observed.

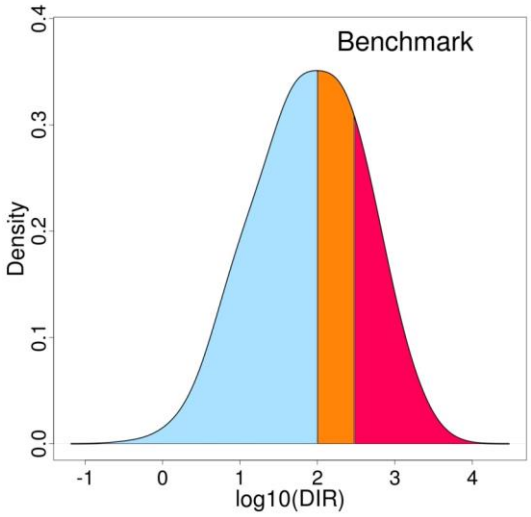

Figure S5: Benchmark probability distribution in Brazil, June 2000-2013.
The historic (long-term) average distribution of dengue incidence rates (DIR) in Brazil. The distribution is divided into low risk (DIR<100; 68% - blue), medium risk (100<DIR<300; 16% - yellow) and high risk (DIR>300; 16% - pink).

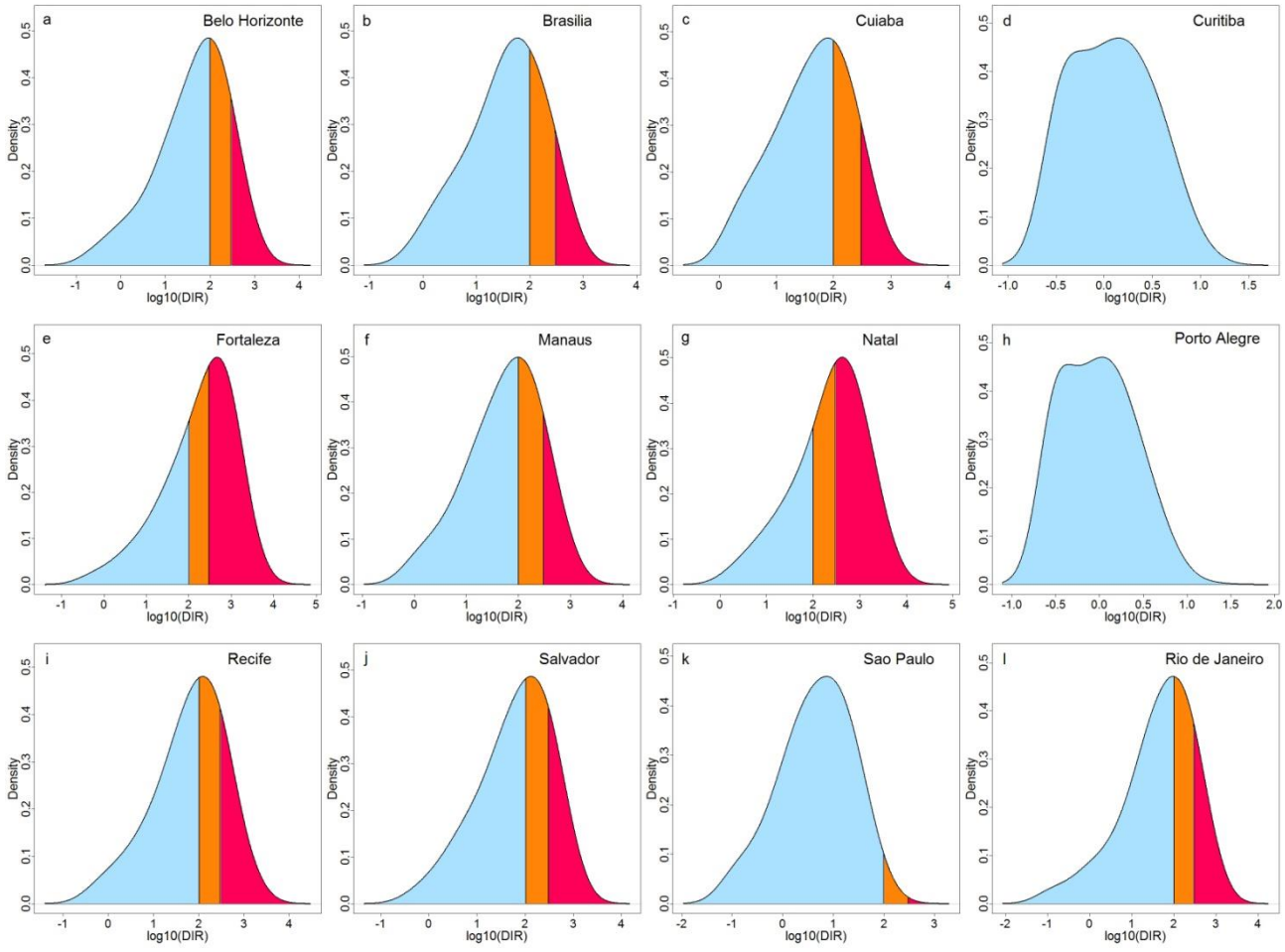

Figure S6: Predictive distributions for June 2014 for microregions where stadiums are located.
Posterior predictive distributions for June 2014 showing the probability of low risk (blue), medium risk (yellow) and high risk (pink) for June 2014, in the microregions where World Cup stadiums are located: (a) Belo Horizonte, (b) Brasília, (c) Cuiabá, (d) Curitiba, (e) Fortaleza, (f) Manaus, (g) Natal, (h) Porto Alegre, (i) Recife, (j) Salvador, (k) São Paulo and (l) Rio de Janeiro (see table in Article for category summary).

Model evaluation

To determine how well the probability forecasting system predicted the category that the observations fell into for June 2000-2013, the rank probability score (RPS), was calculated for each microregion. RPS measures the sum of the squared differences in cumulative probability space for a multi-category probabilistic forecast.[12]

$$RPS = E\left[\frac{1}{K}\sum_{k=1}^{K}(p_k - o_k)^2\right],$$

where K is number of categories, p_k is the predicted probability in category k and o_k is a value equal to 1 or 0 depending on whether the event occurred in category k, or not. The RPS penalises forecasts less severely when their probabilities are close to the true outcome and more severely when their probabilities are further from the actual outcome.

For each microregion, the rank probability skill score (RPSS) was obtained by estimating the RPS using the 14 year sample mean of probability forecasts and observations (see Fig. 2, Table 1).The RPSS essentially measures the improvement of the multi-category probabilistic forecast relative to benchmark (long-term distribution). The RPSS takes the value 1 for a perfect forecast and 0 for the benchmark (reference) forecast.

$$RPSS = 1 - \frac{RPS}{RPS_{benchmark}}.$$

Relative (or Receiver) Operating Characteristic (ROC) graphs were used in this study to evaluate the forecasting system.[12] The ROC is a graph of the hit rate (HR) against the false alarm rate (FAR) (or sensitivity against 1-specificity) for different decision thresholds. The location of the whole curve in the unit square is determined by the intrinsic discrimination capacity of the forecasting system and the location of specific points on a curve is fixed by the probability decision threshold at which the system is operating. As the probability decision threshold varies from high to low (moving from left to right) HR and FAR vary together to trace out the ROC curve. Perfect discrimination is represented by the point (0, 1) where HR= 100% and FAR= 0%. The diagonal HR=FAR represents zero skill, i.e. the forecasting system performs no better than random guessing. Trigger thresholds were defined as the point on the curve closest to the point of perfect discrimination. The area under the modelled ROC curve, abbreviated AUC,[13] is a widely used ROC-based measure of skill. AUC characterises the quality of a forecast system by describing the system's ability to anticipate correctly the occurrence or non-occurrence of pre-defined events.[14] The possible range of AUC is [0,1]. Zero skill is indicated by AUC=0.5, that is, the area under the diagonal HR=FAR. For perfect skill, AUC=1. To test the null hypothesis that the area under the ROC curve is 0.5, i.e. the forecast has no skill, a p-value can be calculated using a Mann-Whitney U-test. [14] The ROC curves in Figure 5 a and b (see Article) have AUC=0.86 (p-value < 0.0001) and AUC=0.84 (p-value < 0.0001), respectively. This indicates that the forecasting system performs significantly better than randomly guessing.

References

1 Lowe, R. Bailey TC, Stephenson DB, et al. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput. Geosci* 2011; **37**: 371–381.

2 Lowe R, Bailey TC, Stephenson DB, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Stat Med* 2013; **32**, 864–883.

3 Fan Y, Van den Dool H. A global monthly land surface air temperature analysis for 1948–present. *J Geophys Res* 2008; **113**: D1.

4 Adler RF, Huffman GJ, Chang A, et al. The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J Hydrometeorol* 2003; **4**: 1147–1167.

5 Besag J, Green P, Higdon D, Mengersen K. Bayesian computation and stochastic systems. *Stat. Sci* 1995; **10**: 3–41.

6 Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, Boca Raton, FL, 1996.

7 Lunn DJ, Thomas A, Best N, Spiegelhalter DJ. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000; **10**: 325–337.

8 Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw* 2005; **12**: 1–16.

9 Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sinica* 1996; **6**: 733–759.

10 Coelho CAS, Stephenson DB, Balmaseda M, Doblas-Reyes FJ, van Oldenborgh GJ. Toward an integrated seasonal forecasting system for South America. *J. Climate* 2006; **19**: 3704–3721.

11 Jupp TE, Lowe R, Coelho CAS, Stephenson DB. On the visualization, verification and recalibration of ternary probabilistic forecasts. *Phil Trans R Soc A* 2012; **370**: 1100–1120.

12 Jolliffe IT, Stephenson DB. Forecast Verification: A Practitioner's Guide in Atmospheric Science. Second Edition. Wiley, Chichester, 2011.

13 Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett* 2006; **27**: 861–874.

14 Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Quart J Roy Meteor Soc* 2002; **128**: 2145–2166.