Spatio-temporal modelling of vector-borne disease: a case study of dengue in Brazil

Rachel Lowe

International Centre for Theoretical Physics

(rlowe@ictp.it)

Summer School on Climate Impacts Modelling for Developing Countries: Water, Agriculture and Health, ICTP, Italy, 12 September 2011

Trevor Bailey, David Stephenson, Tim Jupp (UoE), Richard Graham (Met Office), Caio Coelho (CPTEC), Marilia Sá Carvalho, Christovam Barcellos (FIOCRUZ), Antonio Miguel Vieira Monteiro (INPE)

Objectives

Demonstrate how statistical methodologies for spatio-temporal data can be applied to model climate-sensitive disease risk.

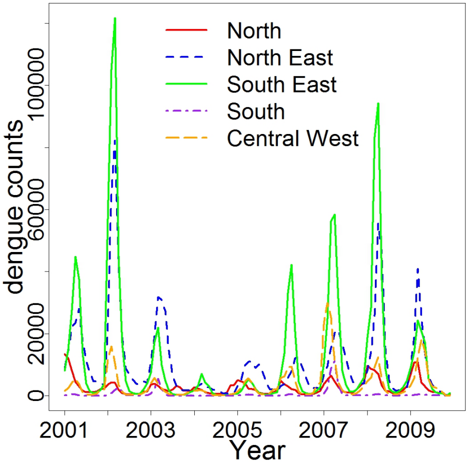
- Analyze and visualize spatio-temporal data and model results.
- Evaluate predictive validity of probabilistic forecasts.

Dengue in Brazil

- Dengue transmitted by *Aedes aegypti* mosquitoes
- Severe joint and muscle pain (rarely fatal) — ‘Break-bone fever’
- More than 3 million cases in Brazil 2001-2009
- 2008 epidemic: 787,726 cases, 448 deaths
- Seasonal pattern: increases in Jan-May when climate warmer/humid

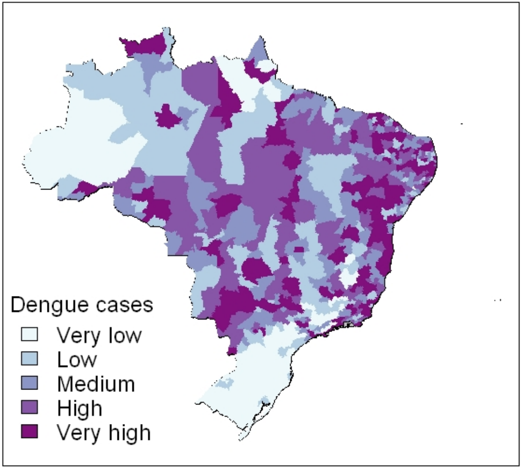
(Center for Disease Control Public Health Image Library and BBC)

Temporal variability in dengue in Brazil



Monthly dengue counts for main regions of Brazil 2001-2009

Spatial variability in dengue in Brazil



Total dengue cases in microregions (558) 2001-2009

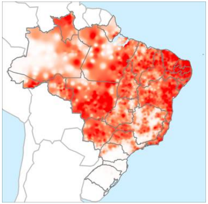
Dengue transmission

- Human drivers, e.g.
  - population growth/urbanisation/poverty (substandard housing)
  - abundance of water-storage containers
- Environmental drivers, e.g.
  - Rainfall
    (filling of containers)
  - Temperature/humidity
    (mosquito development)

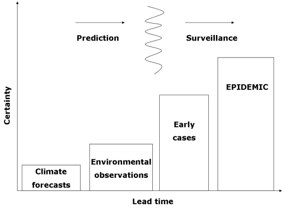
Environmental Health Perspectives, 2008

Research Aim

- Develop a modelling framework to provide spatio-temporal probabilistic forecasts of dengue risk.
    - To what extent can spatio-temporal variations in dengue risk be accounted for by climate variations?
    - Which observed and unobserved non-climatic confounding factors should be incorporated?



- Is climate information useful in a dengue Early Warning System (EWS) for Brazil?
    - How well can the developed model predict future and geographically specific dengue epidemics?
    - How does this compare with current ‘surveillance and response’ approach in Brazil (observe early dengue cases Dec/Jan then estimate epidemic potential for late austral summer)?
    - How can early warnings of dengue epidemics based on climate information be effectively communicated to public health decision makers?



TODOS CONTRA A DENGUE

Disease and Demographic Data

Disease data SINAN-DATASUS

- Monthly dengue cases Jan 2001 - Dec 2009
- Spatial unit: microregion

Census/cartographic data SIDRA-IBGE

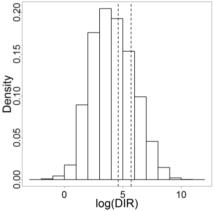
- % urban population
- Altitude
- Administrative region
- Zone or Biome (e.g. Atlantic/Amazon Rainforest)

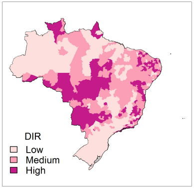
Overall dataset: 108 months, 558 locations

DIR$= \frac{y_{st}}{p_{st}} \times 100,000$

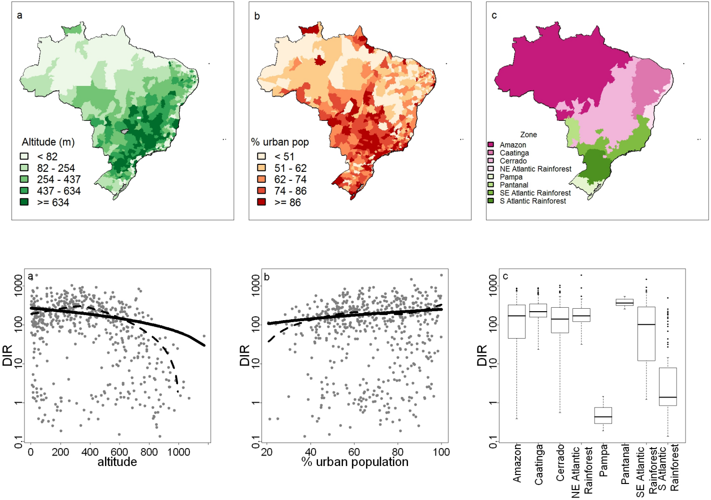
Low: DIR < 100
Med: $100 <$ DIR < 300

High: DIR > 300



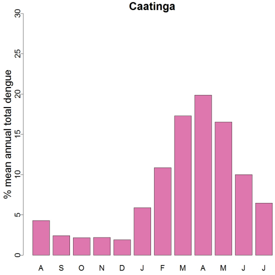


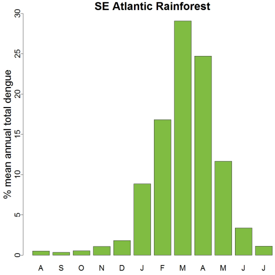
Dengue in relation to altitude, urban population and zone

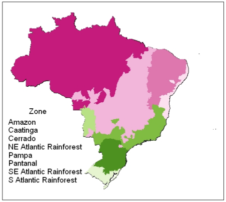


(a) Altitude, (b) % urban population, (c) geographic zone

Geographically specific annual cycle

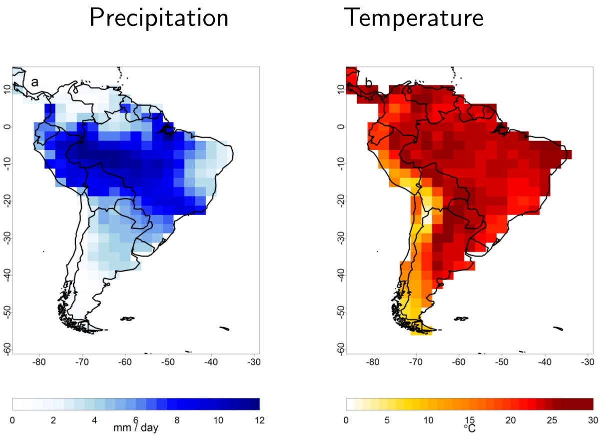






Gridded climate data (2.5° × 2.5°)

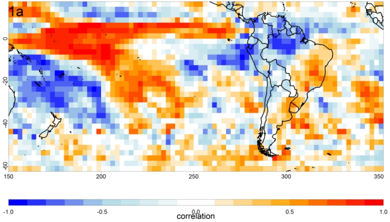
- Average precipitation rate (GPCP)
- Reanalysis average temperature (NCEP/NCAR)



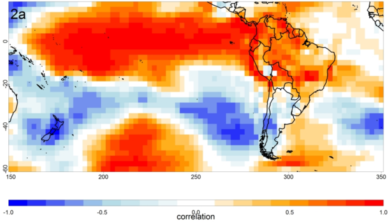
Dec-Feb climatology (2000-9)

El Niño Southern Oscillation

Precipitation

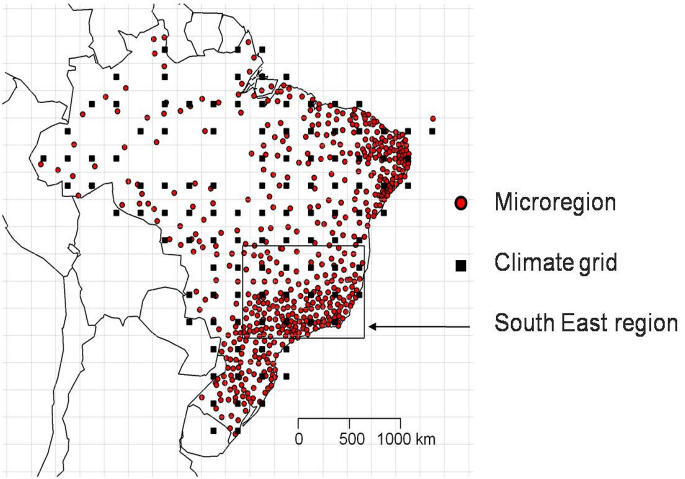


Temperature

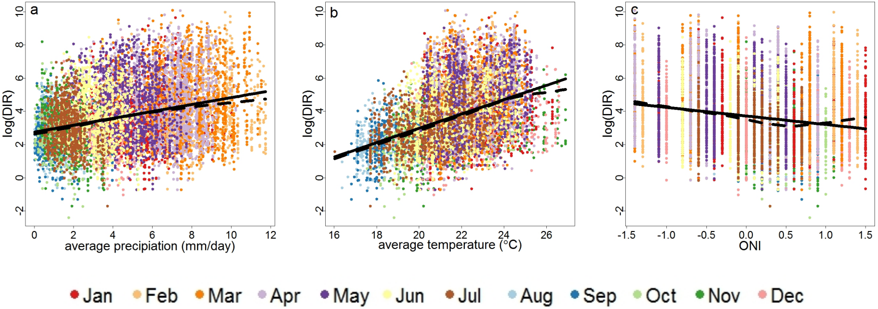


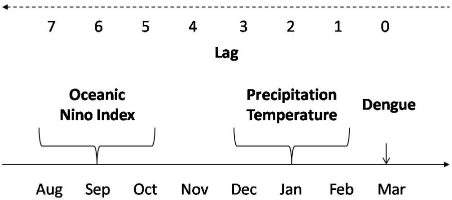
Correlation Oceanic Niño Index (ONI) vs Dec-Feb precipitation & temperature

Microregions and climate grid



Time lag considerations





Model framework: Generalised Linear Model (GLM)

$$
\begin{aligned}
y_{st} &\sim \text{NegBin}(\mu_{st}, \kappa) \\
\log \mu_{st} &= \underbrace{\log e_{st}}_{offset} + \alpha + \underbrace{\delta_{1t'(t)} + \delta_{2s'(s)} + \delta_{3s'(s)t'(t)}}_{factors} + \underbrace{\sum_j \gamma_j w_{jst}}_{non-climate} \\
&\quad + \underbrace{\sum_j \beta_j x_{jst} + \sum_j \beta_{js'(s)} x_{jst}}_{climate}
\end{aligned}
$$

y_{st} dengue count for microregion $s = 1, \ldots, 558$ and time $t = 1, \ldots, 108$

μ_{st} mean dengue count

κ scale parameter

$e_{st} = p_{st}\pi$, p_{st} population in microregion s and time t, π overall average dengue rate

x_{jst} precipitation, temperature, ONI

w_{jst} altitude and population density

$\delta_{1t'(t)}$ calendar month, $t'(t) = 1, \ldots, 12$ (categorical variable)

$\delta_{2s'(s)}$ zone $s'(s) = 1, \ldots, 8$ (categorical variable)

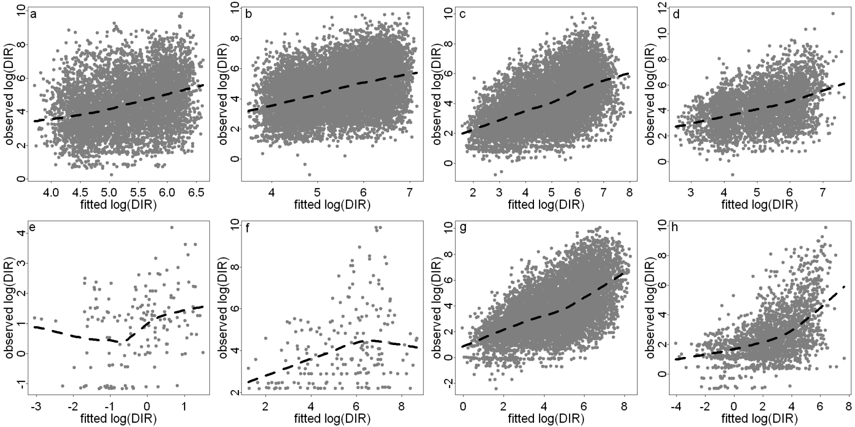
$\delta_{3s'(s)t'(t)}$ interaction between calendar month and zone

Model selection

Comparison of models with increasing complexity

Model	Deviance	R^2_D	p	$n-p$	AIC	BIC
Null model	63007	0	2	60262	404321	404330
Climate model	61882	0.21	5	60259	389550	389586
Non-climate model	61495	0.33	99	60165	380425	381308
Combined model	60520	0.39	123	60141	374515	375614

Selected results - GLM



Observed and model fit DIR (a) Amazon Rainforest, (b) Caatinga, (c) Cerrado, (d) NE Atlantic Rainforest, (e) Pampa, (f) Pantanal, (g) SE Atlantic Rainforest and (h) S Atlantic Rainforest

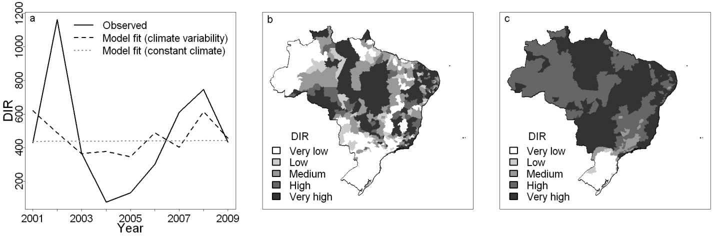
Selected results - GLM

Parameter estimates (standard error) for climate covariates in Brazilian zones

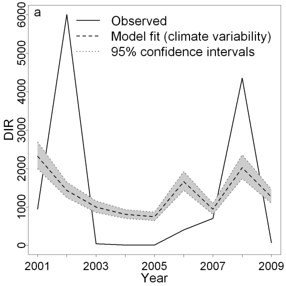
Zone	precipitation	temperature	ONI
Amazon Rainforest	-0.005 (0.007)	**-0.217** (0.019)	**-0.157** (0.034)
Caatinga	**-0.070** (0.009)	-0.02 (0.029)	-0.018 (0.054)
Cerrado	**0.068** (0.01)	**0.135** (0.028)	**-0.408** (0.055)
North East Atlantic Rainforest	**0.196** (0.02)	**0.089** (0.039)	**-0.223** (0.065)
Pampa	-0.003 (0.07)	**0.347** (0.12)	**-0.357** (0.174)
Pantanal	**0.437** (0.112)	**0.384** (0.126)	**-1.345** (0.187)
South East Atlantic Rainforest	**0.041** (0.014)	**0.466** (0.029)	**-0.611** (0.055)
South Atlantic Rainforest	**0.337** (0.019)	**0.85** (0.031)	-0.096 (0.064)

Estimates in bold face are significant at the 0.05 level.

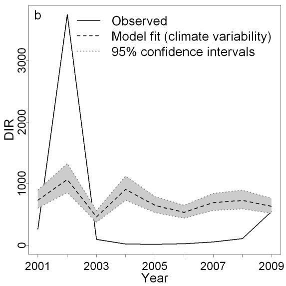
Selected results - GLM



Rio de Janeiro



Salvador da Bahia



How to account for unexplained variance?

- GLM fails to capture spatio-temporal dengue variability.
  - population immunity to circulating serotype.
  - health interventions/vector control measures.
- Problem: lack of data to model disease system.
- Solution:
  - Early cases - surrogate for unobserved and unmeasured spatio-temporal confounding factors.
  - Hierarchical model - add extra level uncertainty using random effects.

Accounting for unknown confounding factors

Early cases

- Idea: current incidence can be partly explained by past values
- Problem: short time lag, not feasible for advance warning of an impending epidemic
- Compromise: dengue risk three month previous $z_{st} = \log(\frac{y_{st-3}}{e_{st-3}})$
- Represent increased mosquito populations/circulation new serotype?

Random effects

- Unobserved latent structures
- Overdispersion
- Temporal correlation
- Spatial clustering

Selected Generalised Linear Mixed Model framework

$$
\begin{aligned}
y_{st} | \phi_s, \nu_s, \omega_{t'(t)} &\sim \mathrm{NegBin}(\mu_{st}\,, \kappa); \qquad s = 1, \ldots, 558;\ t = 1, \ldots, 108 \\
\log \mu_{st} &= \underbrace{\log e_{st}}_{\text{offset}} + \alpha + \underbrace{\delta_{1t'(t)} + \delta_{2s'(s)} + \delta_{3s'(s)t'(t)}}_{\text{month+zone factors}} \\
&\quad + \underbrace{\gamma_1 w_{1st} + \gamma_2 w_{2s}}_{\text{non-climate vars: pop dens+altitude}} \\
&\quad + \underbrace{\beta_{1s'(s)} x_{1,s,t-2} + \beta_{2s'(s)} x_{2,s,t-2} + \beta_{3s'(s)} x_{3,t-6}}_{\text{climate vars: precip+temp+ONI}} \\
&\quad + \underbrace{\delta z_{st}}_{\text{early cases}} + \underbrace{\phi_s + \nu_s}_{\text{spatial random effects}} + \underbrace{\omega_{t'(t)}}_{\text{monthly random effects}}
\end{aligned}
$$

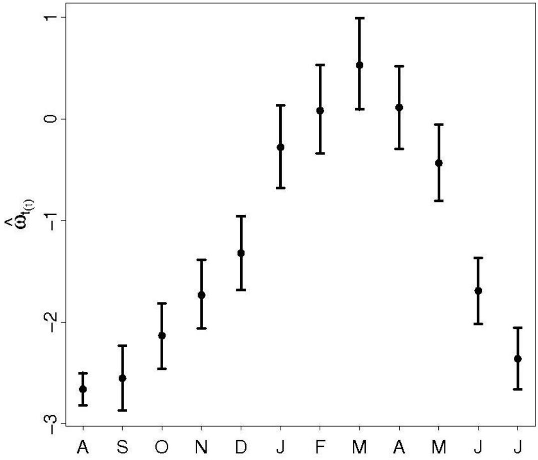
$$
\begin{aligned}
&\phi_s \sim \mathrm{N}(0, \sigma^2_\phi) \\
&(\nu_s) \sim \mathrm{CAR}(\sigma^2_\nu) \\
&\omega_1 = 0, \quad \omega_{t'(t)} \sim \mathrm{N}(\omega_{t'(t)-1}, \sigma^2_\omega); \qquad t'(t) = 2, \ldots, 12 \\
&\sigma^2_\lambda \sim \mathrm{Ga}(0.5, 0.0005)\,, \lambda = (\phi, \nu, \omega)\,, \kappa \sim \mathrm{Ga}(0.5, 0.0005).
\end{aligned}
$$

Selected GLMM model framework

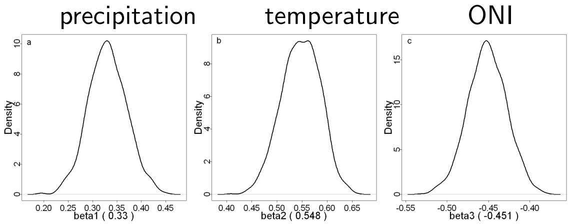
- Climate signal is weak but statistically significant.
- Precipitation and temperature averaged over preceding 3 month period, 2 month lag with dengue.
- ONI lagged 6 months with dengue, 4 months with climate variables.
- Early cases lagged 3 months, slight improvement to spatio-temporal variation.
- Random effects are important:
  - Unobserved confounding factors (population immunity to circulating serotype, health interventions/vector control measures)
  - Overdispersion
  - Temporal correlation and spatial clustering

South East Brazil: peak dengue season February-April (FMA)

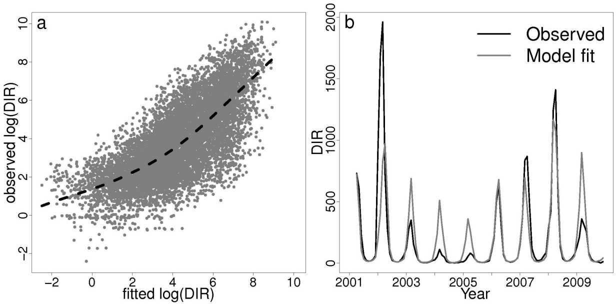
Auto-correlated annual cycle



Selected results - GLMM, SE Brazil



Climate coefficient posteriors

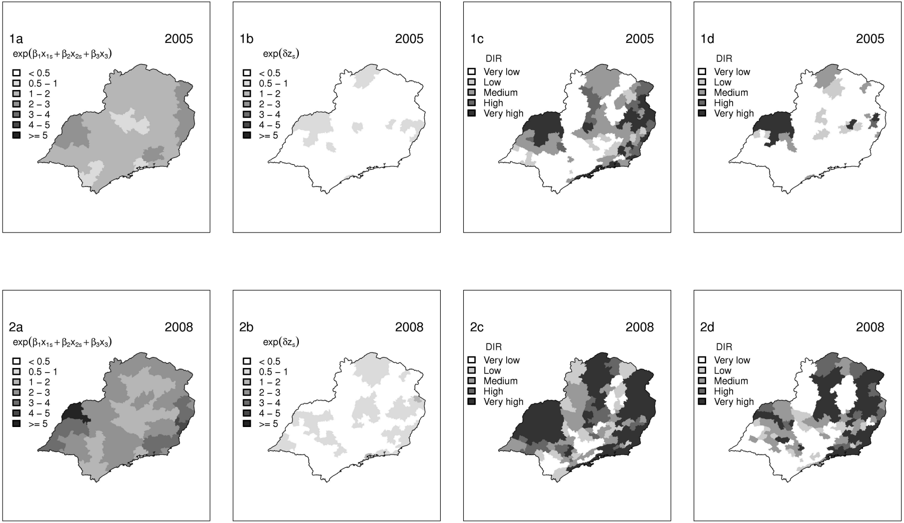


Observed vs model fit, 2001-2009

Multiplicative decomposition dengue risk, FMA season

Non-epidemic year (2005) and epidemic year (2008)

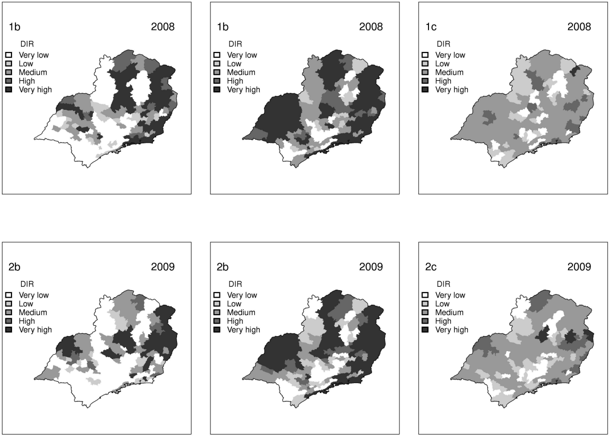
Climate | Previous cases | Model | Observed



Comparison of GLMM and current surveillance model

Current surveillance model (CSM):

$$
\begin{aligned}
y_{st} &\sim \mathrm{NegBin}(\mu_{st}, \kappa) \\
\log \mu_{st} &= \log e_{st} + \alpha + \delta z_{st}
\end{aligned}
$$

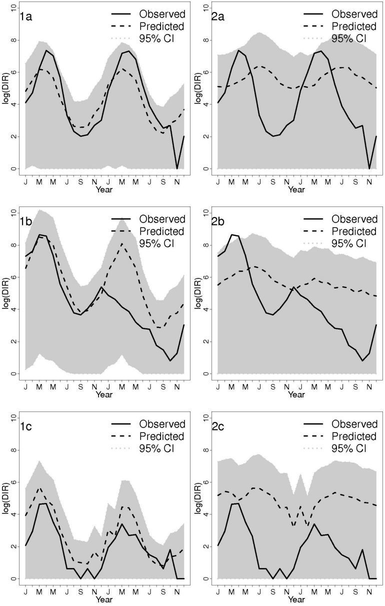


Posterior predictions selected microregions 2008-2009

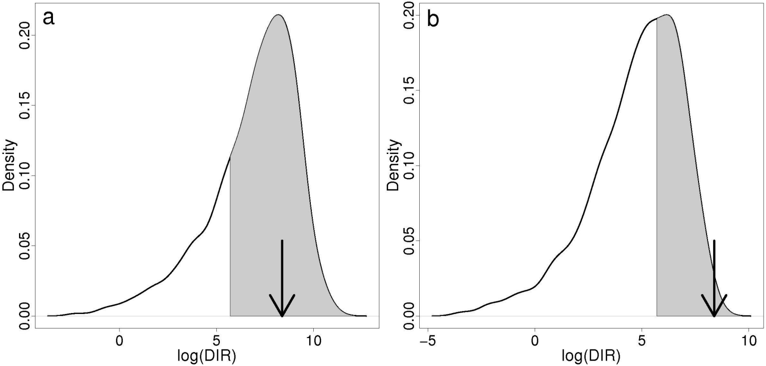
Belo Horizonte

Rio de Janeiro

São Jose dos Campos



Posterior prediction FMA 2008 epidemic, Rio de Janeiro



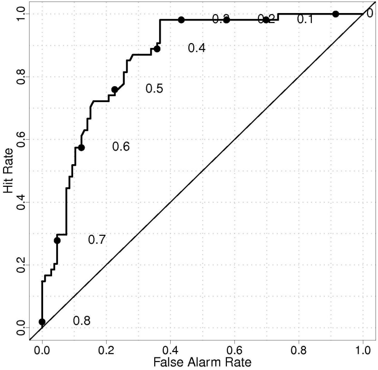
GLMM ($p(\text{DIR}) > 300 = 0.75$)   CSM ($p(\text{DIR}) > 300 = 0.37$)

- GLMM improvement to current practice
- Inclusion of climate information and observed and unobserved confounding factors improves model performance

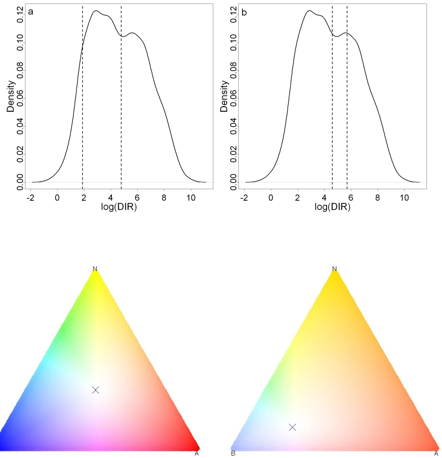
ROC analysis epidemic threshold: 300 per 100,000

Posterior predictive results in 160 microregions for cases exceeding 300 per 100,000 at probability decision thresholds ($40\%, 50\%, 60\%$)

Threshold	Hit	False Alarm	Miss	Correct Rejection	PC	HR	FAR
60%	31	13	23	93	76%	57%	12%
50%	44	27	10	79	77%	81%	25%
40%	49	36	5	70	74%	91%	34%

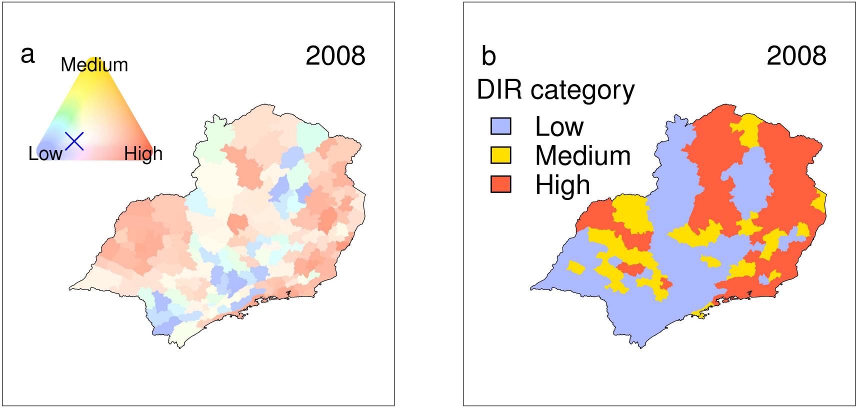


Defining and visualising epidemic risk

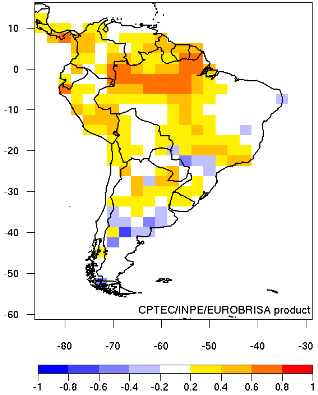


Symmetric (tercile) and non-symmetric (100 and 300 cases per 100,000) category boundaries of the observed distribution of DIR, FMA 2001-2007, SE Brazil

Visualising probabilistic forecasts 2008 epidemic

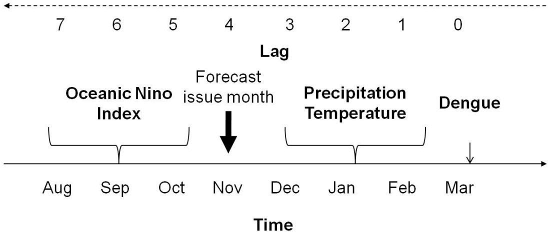


EURO-BRazilian **I**nitiative for improving **S**outh **A**merican seasonal climate forecasts `http://eurobrisa.cptec.inpe.br/`



Correlation between forecast and observed precipitation anomaly using the integrated EUROBRISA forecasting system for the period 1981-2005. Forecasts issued in November, valid for DJF season

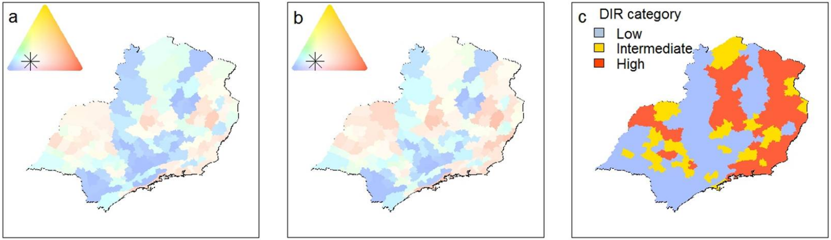
Extending prediction lead-time with forecast climate



EUROBRISA forecast

GPCP observed

Observed categories



Conclusions & Future work

Conclusions

- Climate accounts for some variation in dengue risk
- Important to account for confounding factors
- Potential for use of climate information in Brazil dengue EWS

References

- Lowe, R., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C. A. S., Carvalho, M. S., Barcellos, C., 2011. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. Computers Geosciences 37, 371-381.
- Jupp, T.E., Lowe, R., Coelho, C. A. S., Stephenson, D. B., 2011. On the visualisation, verification and recalibration of ternary probabilistic forecasts. Phil. Trans. R. Soc. A, [in press].

Future work

- Test model framework more fully in other locations
- Extent to which climate forecasts extend predictive lead time
- Addition of serotype information
- Addition of health intervention/prevention information
- Representative movement of human hosts
- Incorporate better understanding of disease transmission process

Recommended Reading

Ebi, K.L. (2009). Malaria Early Warning Systems. *In: Biometeorology for Adaptation to Climate Variability and Change*, 49–74.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters* **27**(8), 861–874.

Lawson, A.B. (2008). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman & Hall/CRC, Boca Raton, Florida, USA, 344pp.

Mason, S. J. and Graham, N. E. (2002). Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society* **128**(584), 2145–2166.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, Second Edition*. Chapman Hall, London, UK, 536pp.

Teixeira, M.G., Costa, M.C.N., Barreto, F. and Barreto, M.L. (2009). Dengue: twenty-five years since reemergence in Brazil. *Cadernos de Saúde Pública* **25**, 7–18.

Thank you!