Introduction to R-INLA

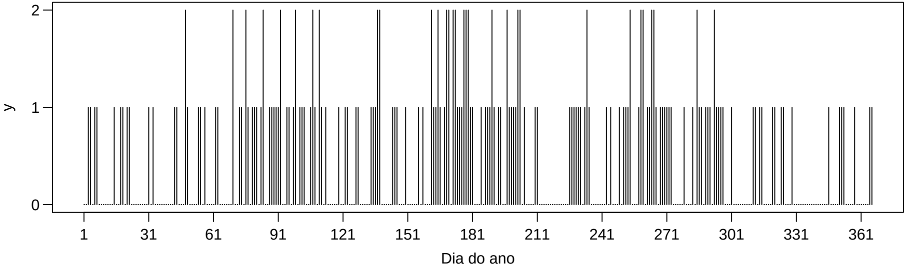
Elias T Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

May, 22 2018

Sumário

Example 1: Rain each day over two years by each day of the year



Problem: model the probability of rain each day of the year

Model

- y_i assume values 0, 1 or 2, for $i = 1, ..., n$
  - assuming conditional independence, thus

$$y_i | p_i \sim \text{Binomial}(2, p_i)$$

  - link function (logit)

$$p_i = 1/(1 + \exp(-x_i))$$

- **x**: smoothing over time
  - ***R**andon **W**alk* - RW of first order: `rw1`
  - Gaussian distribuiton for the sucessive differences (**R** esparse)

$$\Delta_i = x_i - x_{i-1} \sim N(0, \tau^{-1})$$

$$\log(\pi(\mathbf{x}|\tau)) \propto -\frac{\tau}{2}[(x_1 - x_n)^2 + \sum_{i=2}^{n}(x_i - x_{i-1})^2] = -\frac{\tau}{2}\mathbf{x}'\mathbf{R}\mathbf{x},$$

  - **x** is a *Gaussian Markov Random Field* - GMRF, Rue and Held (2005)
- τ: local precision parameter

Model, INLA

- Model

$$\begin{aligned} y_i|x_i &\sim \text{Binomial}(2, p_i) & \rightarrow \text{likelihood} \\ \mathbf{x}|\tau &\sim N(\mathbf{0}, (\tau\mathbf{R})^{-}) & \rightarrow \text{latent field, GMRF} \\ \tau &\sim p(\tau) & \rightarrow \text{prior distribution} \end{aligned}$$

- INLA: Acronym of ***I**ntegraged* ***N**ested* ***L**aplace* ***A**proximation*, Rue, Martino, and Chopin (2009)
  - First ***L**aplace* ***A**pproximation* - LA to compute

$$\pi(\tau|\mathbf{y})$$

  - Second (nested to the first) LA to compute

$$\pi(x_i|\mathbf{y})$$

- If the likelihood is Gaussian, the “approximation” is exact

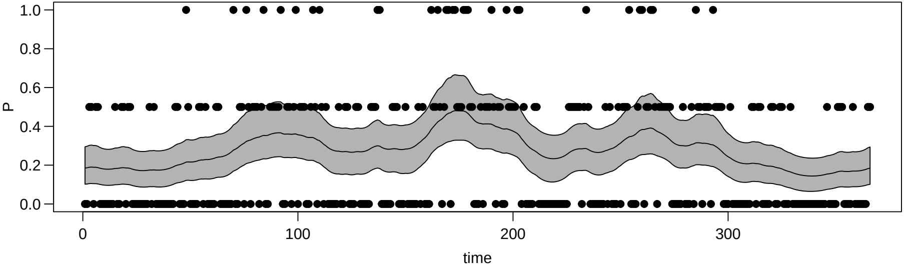
Example of a time series with rw1

```r
head(Tokyo, 5)
```

```
##   y n time
## 1 0 2    1
## 2 0 2    2
## 3 1 2    3
## 4 1 2    4
## 5 0 2    5
```

```r
model <- y ~ f(time, model='rw1', cyclic=TRUE)
result <- inla(model, family='binomial',
               data=Tokyo, Ntrials=n,
               control.compute=list(cpo=TRUE))
```

Result for the time series



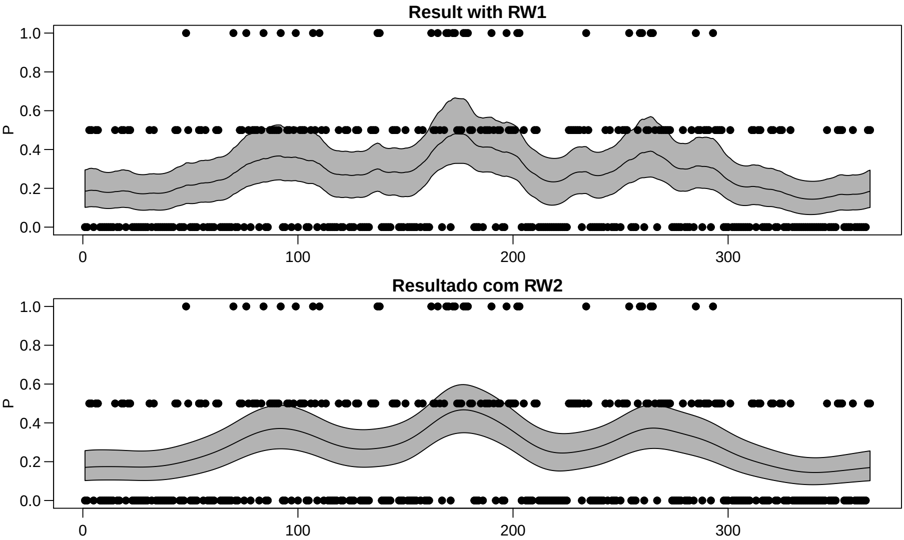
Smoothing more

Gaussian distribution for the second order differences (`rw2`)

$$\Delta_i^2 = x_i - 2x_{i-1} + x_{i-2} \sim N(0, \tau^{-1})$$

```
model2 <- y ~ f(time, model='rw2', cyclic=TRUE)
result2 <- inla(model2, family='binomial',
                data=Tokyo, Ntrials=n,
                control.compute=list(cpo=TRUE))
```

Both results for the time series



Goodness-of-fit measures

- ***C**onditional **P**redictive **O**rdinate* - CPO:

$$P(y_i^{\text{obs}}|\mathbf{y}_{-i})$$

$\mathbf{y}_{-i}$ is the $\mathbf{y}$ vector without the y_i element.

```
c(-sum(log(result$cpo$cpo)),
  -sum(log(result2$cpo$cpo)))
```

```
## [1] 312.6621 314.3368
```

- ***P**robability **I**ntegral **T**ransform* - PIT:

$$P(Y_i \leq y_i^{\text{obs}}|\mathbf{y}_{-i}).$$

- Useful to detect lack of fit or outliers

Sumário

Possibilities of models

- Generalized (mixed) models
- Generalized additive (mixed) models
- Survival models
- Dynamic models
- Stochastic volatitily models
- Smoothing spline
- Semiparametric regression
- Disease mapping
- Model based geostatistics*
- Log-Gaussian Cox processes
- Space-time models
- Semiparametric regression with spatial (space-time) varying coefficients
- +++

Possibilities of models

- Generalized (mixed) models
- Generalized additive (mixed) models
- Survival models
- Dynamic models
- Stochastic volatitily models
- Smoothing spline
- Semiparametric regression
- Disease mapping
- Model based geostatistics*
- Log-Gaussian Cox processes
- Space-time models
- Semiparametric regression with spatial (space-time) varying coefficients
- +++

$\rightarrow$ GLMM, GAM, GAMM, ... are different names for the same thing

Flexibility and responsability

- In the rainfall ocurrence example, suppose we add other terms (covariate, seazonal effect)
- Flexibility must come with responsability
- PC-prior ***P**enalized **C**omplexity* prior, Simpson et al. (2017)

Sumário

Implementation

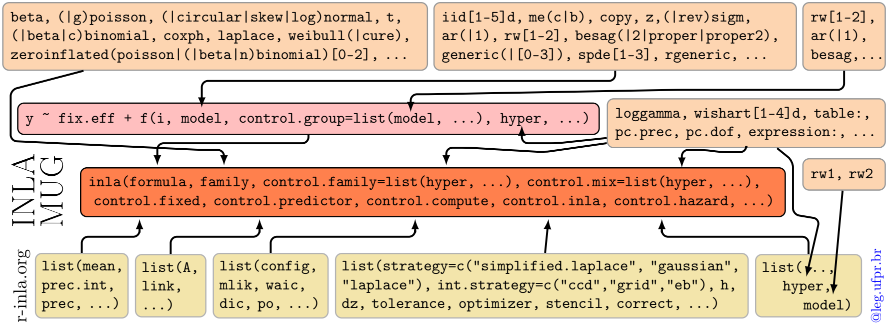
The code

- gmrflib, inlaprog, fmesher
  - ~6000 lines in Fortran
  - +100000 lines in C++
- external libraries: muparser, taucs (sparce matrices), PARDISO (experimental)
- ~60000 lines of code+help in R (R-INLA)

The work

- +15 years Håvard Rue (~80% of the code)
- +10 years Finn Lindgren (~20% of the code)
- source code in the bitbucket (open source), web page: r-inla.org

Main function in R-INLA

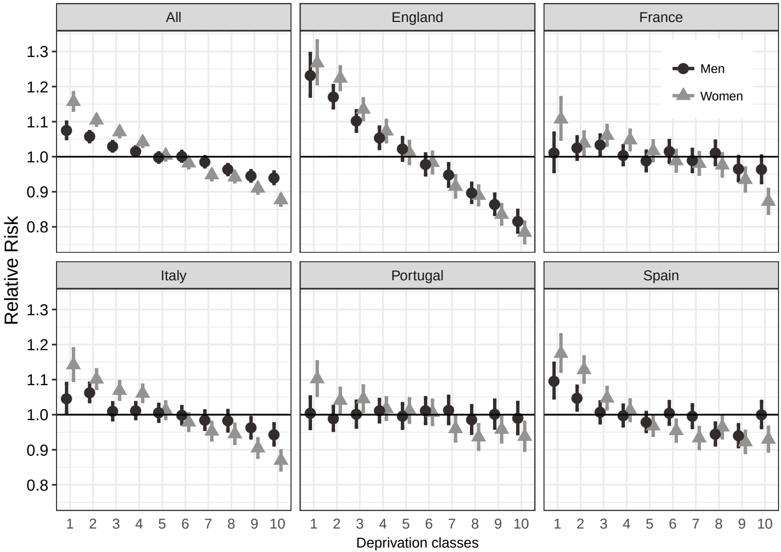


Sumário

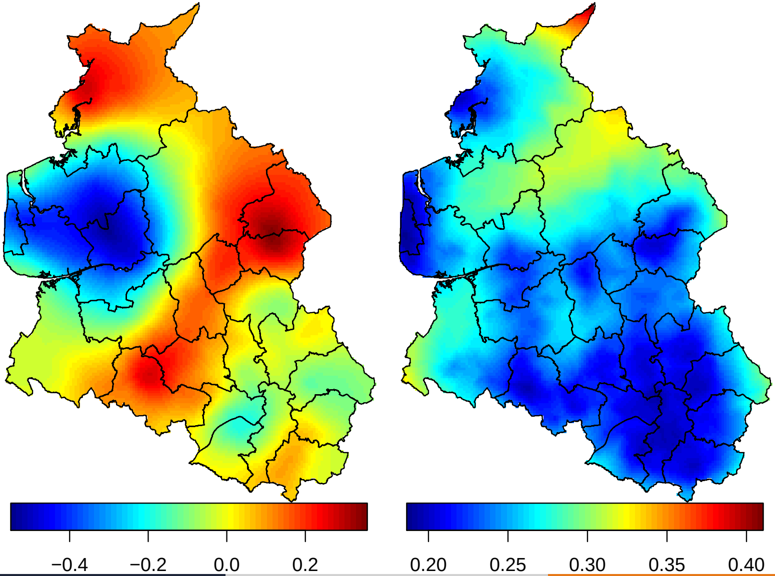
Some applications cited in Rue et al. (2017)

Recent examples of applications using the R-INLA package for statistical analysis include disease mapping (Schrödle & Held 2011a , b ; Ugarte et al. 2014 , 2016 ; Papoila et al. 2014 ; Goicoa et al. 2016 ; Riebler et al. 2016); age-period-cohort models (Riebler & Held 2016); a study of the evolution of the Ebola virus (Santermans et al. 2016); the relationships between access to housing, health, and well-being in cities (Kandt et al. 2016); the prevalence and correlates of intimate partner violence against men in Africa (Tsiko 2016); a search for evidence of gene expression heterosis (Niemi et al. 2015); analysis of traffic pollution and hospital admissions in London (Halonen et al. 2016); early transcriptome changes in maize primary root tissues in response to moderate water deficit conditions by RNA sequencing (Opitz et al. 2016); performance of inbred and hybrid genotypes in plant breeding and genetics (Lithio & Nettleton 2015); a study of Norwegian emergency wards (Goth et al. 2014); effects of measurement errors (Muff et al. 2015 , Muff & Keller 2015 , Kröger et al. 2016); network meta-analysis (Sauter & Held 2015); time-series analysis of genotyped human campylobacteriosis cases from the Manawatu region of New Zealand (Friedrich et al. 2016); modeling of parrotfish habitats (NC Roos et al. 2015); Bayesian outbreak detection (Salmon et al. 2015); long-term trends in the number of Monarch butterflies (Crewe & Mccracken 2015); long-term effects on hospital admission and mortality of road traffic noise (Halonen et al. 2015); spatio-temporal dynamics of brain tumors (Iulian et al. 2015); ovarian cancer mortality (García-Pérez et al. 2015); the effect of preferential sampling on phylodynamic inference (Karcher et al. 2016); analysis of the impact of climate change on abundance trends in central Europe (Bowler et al. 2015); investigation of drinking patterns in US counties from 2002 to 2012 (Dwyer-Lindgren et al. 2015); resistance and resilience of terrestrial birds in drying climates (Selwood et al. 2015); cluster analysis of population amyotrophic lateral sclerosis risk (Rooney et al. 2015); malaria infection in Africa (Noor et al. 2014); effects of fragmentation on infectious disease dynamics (Jousimo et al. 2014); soil-transmitted helminth infection in sub-Saharan Africa (Karagiannis-Voules et al. 2015); analysis of the effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015 (Bhatt et al. 2015); adaptive prior weighting in generalized regression (Held & Sauter 2016); analysis of hand, foot, and mouth disease surveillance data in China (Bauer et al. 2016); estimation of the biomass of anchovies in the coast of Perú (Quiroz et al. 2015); and many others.

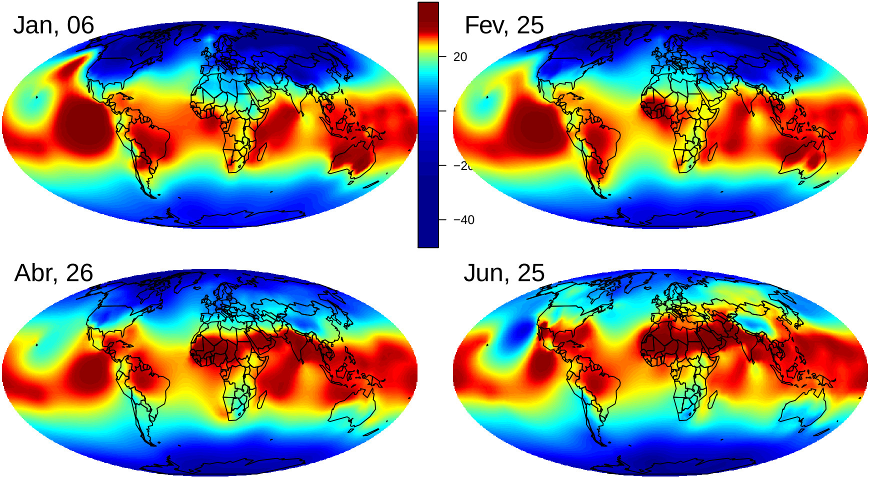
Deprivation effect, Ribeiro et al. (2018)



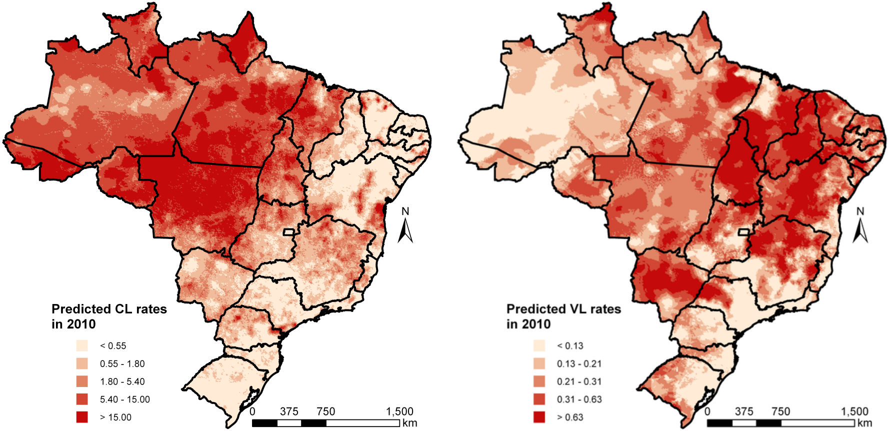
Survival: frailty map



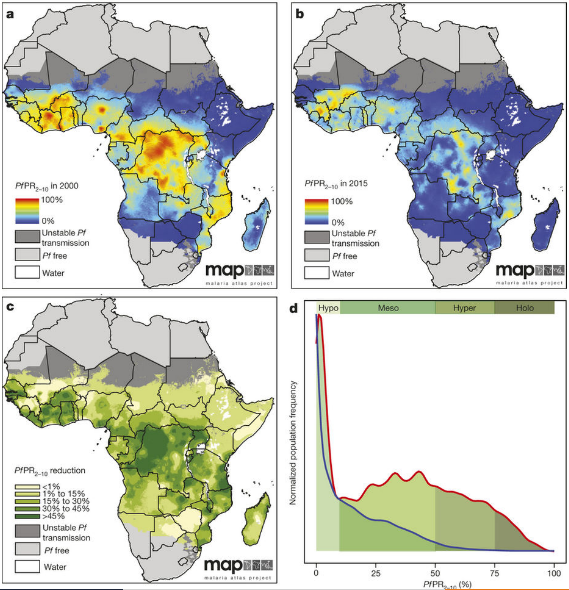
Non-separable space-time modeling in the globe



Leishmaniasis in Brazil, Karagiannis-Voules et al. (2013)



Malaria in Africa, Gething (2015)



Sumário

References

Gething, S. Bhatt AND D. J. Weiss AND E. Cameron AND D. Bisanzio AND B. Mappin AND U. Dalrymple AND K. E. Battle AND C. L. Moyes AND A. Henry AND P. A. Eckhoff AND E. A. Wenger AND O. Briët AND M. A. Penny AND T. A. Smith AND A. Bennett AND J. Yukich AND T. P. Eisele AND J. T. Griffin AND C. A. Fergus AND M. Lynch AND F. Lindgren AND J. M. Cohen AND C. L. J. Murray AND D. L. Smith AND S. I. Hay AND R. E. Cibulskis AND P. W. 2015. "The Effect of Malaria Control on Plasmodium Falciparum in Africa Between 2000 and 2015." *Nature*, no. 526 (October): 207–11.

Karagiannis-Voules, D-A, R. G. C. Scholte, L. H. Guimarães, J. Utzinger, and P. Vounatsou. 2013. "Bayesian Geostatistical Modeling of Leishmaniasis Incidence in Brazil." *PLOS Neglected Tropical Diseases*, no. 5.

Ribeiro, A. I., E. T. Krainski, M. S. Carvalho, G. Launoy, C. Pornet, and M. F. de Pina. 2018. "Does Community Deprivation Determine Longevity After the Age of 75? A Cross-National Analysis." *International Journal of Public Health*, 1–11.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics & Applied Probability. Boca Raton: Chapman; Hall.

Rue, H., S. Martino, and N. Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with Discussion)." *Journal of the Royal Statistical Society, Series B* 71 (2): 319–92.

Rue, H., A. I. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. 2017. "Bayesian Computing with Inla: A Review." *Annual Review of Statistics and Its Application* 4: 395–421.

Simpson, D. P., H. Rue, T. G. Martins, A. Riebler, and S. H. Sørbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1): 1–28.