A practical introduction to Latent Gaussian Models with INLA

Elias T. Krainski

August 2024

Introduction

What is INLA?

- **The short answer:**

INLA - Integrated Nested Laplace Approximations

- fast method to do Bayesian inference with latent Gaussian models
- `INLA` is an `R`-package that implements this method with a flexible and simple interface.
- http://www.r-inla.org

A much longer answer:

- INLA (classic)
  - Rue, Martino, and Chopin (2009) **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society: Series B*. 319–392
  - Or, first, read Rue et al. (2017) **Bayesian Computing with INLA: A Review.** *Annual Review of Statistics and Its Application* 4, 395–421.

A much longer answer:

- INLA (classic)
  - Rue, Martino, and Chopin (2009) **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society: Series B*. 319–392
  - Or, first, read Rue et al. (2017) **Bayesian Computing with INLA: A Review.** *Annual Review of Statistics and Its Application* 4, 395–421.
- The new “INLA” (compact)
  - Niekerk and Rue (2024) **Low-rank Variational Bayes correction to the Laplace method.** *Journal of Machine Learning Research*
  - Or, first, read van Niekerk et al. (2023) **A new avenue for Bayesian inference with INLA.** *CSDA*

A much longer answer:

- INLA (classic)
  - Rue, Martino, and Chopin (2009) **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society: Series B*. 319–392
  - Or, first, read Rue et al. (2017) **Bayesian Computing with INLA: A Review.** *Annual Review of Statistics and Its Application* 4, 395–421.
- The new "INLA" (compact)
  - Niekerk and Rue (2024) **Low-rank Variational Bayes correction to the Laplace method.** *Journal of Machine Learning Research*
  - Or, first, read van Niekerk et al. (2023) **A new avenue for Bayesian inference with INLA.** *CSDA*
- There are some books around . . .

Why should you use `R-INLA`?

- What type of problems can we solve?
- What type of models can we use?
- When can we use it?

Why should you use `R-INLA`?

- What type of problems can we solve?
- What type of models can we use?
- When can we use it?

To have proper answers, we need to start at the very beginning

Basic statistical model structure

- Observations of a phenomena may follow the model

$$
\begin{aligned}
\mathbf{y} &\sim \pi(\mu, \theta_1) \\
h(\mu) &= \mathrm{F}\beta + \mathrm{Zb}
\end{aligned}
$$

Basic statistical model structure

- Observations of a phenomena may follow the model

$$\begin{aligned} \mathbf{y} &\sim \pi(\mu, \theta_1) \\ h(\mu) &= \mathrm{F}\beta + \mathrm{Z}\mathrm{b} \end{aligned}$$

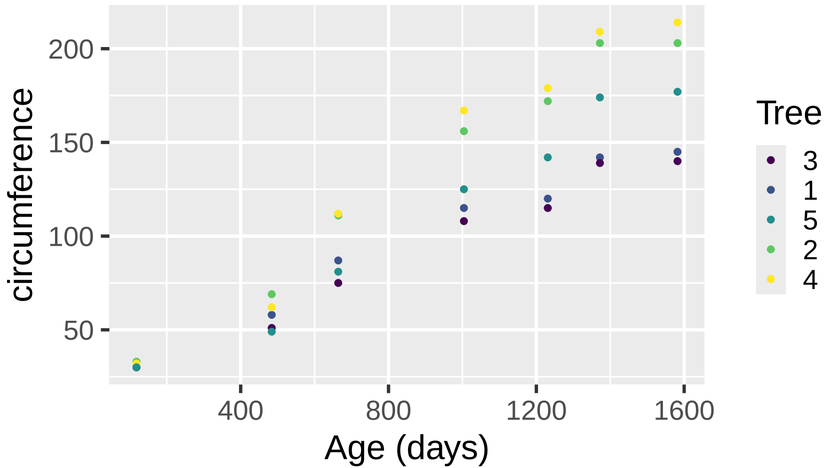
- **F** includes the design matrix, factors, explanatory variables, covariates, independent variables, etc.
- β is a vector of **unknown** *constants*
- **Z** random effects design matrix
- **b** is a vector of **unknown** *constants*

Orange data

```
##   Tree age circumference
## 1    1 118            30
## 2    1 484            58
## 3    1 664            87
```

```
##    Tree  age circumference
## 33    5 1231           142
## 34    5 1372           174
## 35    5 1582           177
```

Orange data (visualize)



Orange: model 1

- Circumference increases as age increases
- Simple linear model

$$\text{circumference} = \beta_0 + \beta_1 \text{Age} + \text{error}$$

Orange: model 1

- Circumference increases as age increases
- Simple linear model

$$\text{circumference} = \beta_0 + \beta_1 \text{Age} + \text{error}$$

- Outcome (circunference): $\mathbf{y} = (\mathrm{y}_1\ , \ldots, \mathrm{y}_n)$
- Covariate (age): $\mathbf{w} = (\mathrm{w}_1\ , \ldots, \mathrm{w}_n)$
- Simple linear model

$$\mathrm{y}_i = \beta_0 + \beta_1 \mathrm{w}_i + \mathrm{u}_i\ , \quad \mathrm{u}_i \sim N(0, \sigma^2)$$

$$\mathsf{E}(\mathrm{y}_i) = \mu_i = \beta_0 + \beta_1 \mathrm{w}_i, \quad \mathrm{y}_i \sim N(\mu_i, \sigma^2)$$

$i = 1, \ldots, n$, $\sigma^2 = \mathrm{e}^{-\theta_1}$

On the common linear model

- Observation model $\mathbf{y} \mid \underbrace{\beta_0, \beta_1}_{\mathbf{x}}, \underbrace{\tau}_{\theta}$:
  - Encodes information about observed data
- Latent model $\mathbf{x}$: The unobserved process
- Hyperprior for θ

On the common linear model

- Observation model $\mathbf{y} \mid \underbrace{\beta_0, \beta_1}_{\mathbf{x}}, \underbrace{\tau}_{\theta}$:
  - Encodes information about observed data
- Latent model $\mathbf{x}$: The unobserved process
- Hyperprior for θ
- From this we can compute the posterior distribution

$$\pi(\mathbf{x}, \theta \mid \mathbf{y}) \propto \pi(\mathbf{y} \mid \mathbf{x}, \theta)\pi(\mathbf{x})\pi(\theta)$$

  and then the corresponding posterior marginal distributions.
  - each model parameter has its own posterior marginal distribution, which is the distribution after accounting for the other parameters

Fitting using INLA

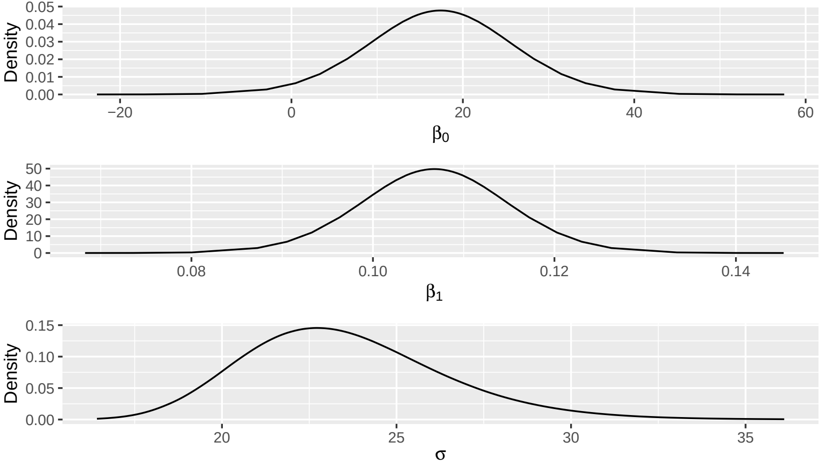
```
m1 <- inla(circumference ~ age, data=Orange,
           control.compute = list(cpo = TRUE))
m1$summary.fixed
```

```
##                mean      sd 0.025quant 0.5quant 0.975quan
## (Intercept) 17.400 8.57832     0.4808   17.400      34.31
## age          0.107 0.00823     0.0905    0.107       0.12
```

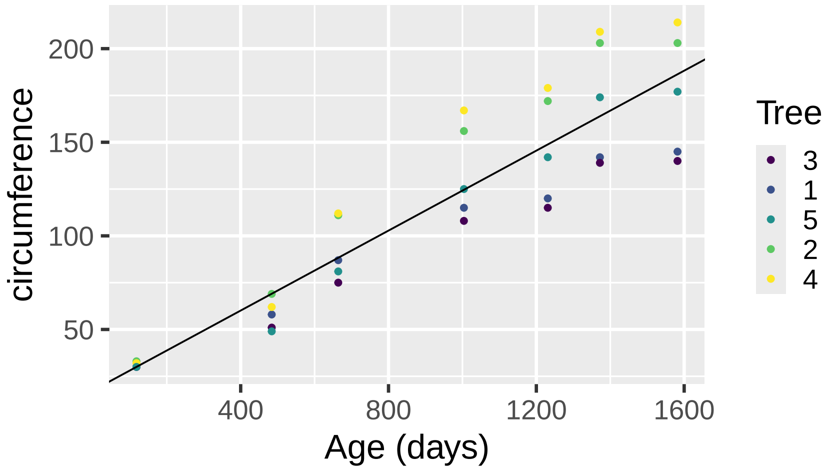
```
m1$summary.hyperpar[1,]
```

```
##                                          mean      sd
## Precision for the Gaussian observations 0.00188 0.00045
##                                         0.975quant     mo
## Precision for the Gaussian observations    0.00286 0.001
```

Posterior marginals



Model 1 fit



200
150
100
50
circumference
400
800
1200
1600
Age (days)
Tree
3
1
5
2
4

Goodness-of-fit measures

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

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

  $\mathbf{y}_{-i}$ is the $\mathbf{y}$ vector without the y_i element
  - useful for model comparison

Goodness-of-fit measures

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

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

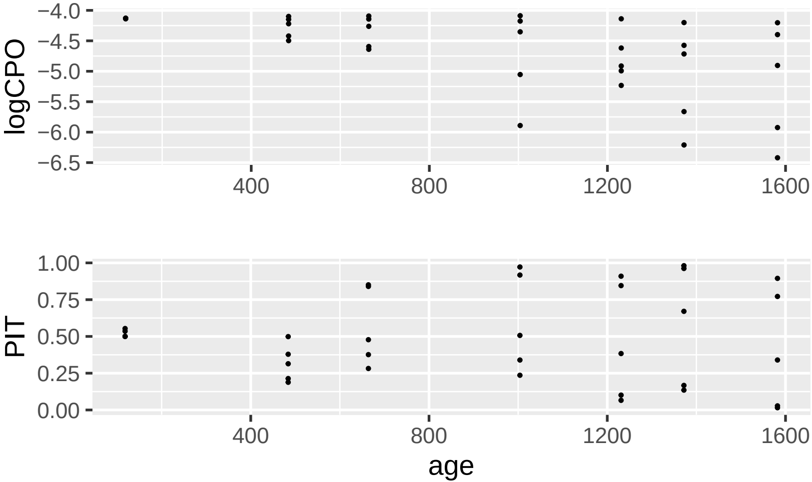
  $\mathbf{y}_{-i}$ is the $\mathbf{y}$ vector without the y_i element
  - useful for model comparison

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

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

  - useful to detect lack of fit or outliers

Orange: model 1 check



Extending the basic model

Orange example: effect for each tree

- **model 2** the increase in circumference with age is different for each tree

$$\text{circumference} = \beta_0 + \beta_{tree}\text{Age} + \text{error}$$

- β_0 and β_j, j for each tree, are unknown
- Now we have: $\mathbf{y} \mid \underbrace{\beta_0, \beta_1, \ldots, \beta_5}_{\mathbf{x}}, \underbrace{\log(1/\sigma^2)}_{\theta}$

About the coefficients

- Usually it is assumed that β_j is
  - *a random sample from a population*
    - because each tree is sampled from a population of trees

About the coefficients

- Usually it is assumed that β_j is
  - *a random sample from a population*
    - because each tree is sampled from a population of trees
- It is **very common** to consider $\beta_j \sim N(0, \tau_\beta^{-1})$
  - even non Bayesian does this

About the coefficients

- Usually it is assumed that β_j is
  - *a random sample from a population*
    - because each tree is sampled from a population of trees
- It is **very common** to consider $\beta_j \sim N(0, \tau_\beta^{-1})$
  - even non Bayesian does this
- Being Bayesian:
  - It is **also common** to consider $\beta_0 \sim N(m_0, \tau_0^{-1})$, m_0 and τ_0 fixed
  - $\beta = \{\beta_0, \beta_1, \ldots, \beta_5\}$ is a Gaussian with precision

$$\begin{bmatrix} \tau_0 & & & & & \\ & \tau_\beta & & & & \\ & & \tau_\beta & & & \\ & & & \tau_\beta & & \\ & & & & \tau_\beta & \\ & & & & & \tau_\beta \end{bmatrix}$$

A small point to think about

- From a Bayesian point of view fixed effects and random effects are all the same (non-observable and unknown)
- Fixed effects are also random
- They only differ in the prior we put on them

Orange example, model 2

```
f2 <- circumference ~ 1 + f(Tree, age, model='iid')
m2 <- inla(f2, data=Orange, control.compute=list(cpo=TRUE))
m2$summary.fixed
```

```
##                 mean    sd 0.025quant 0.5quant 0.975quant
## (Intercept) 18.14 3.635      11.04    18.12      25.37 1
```

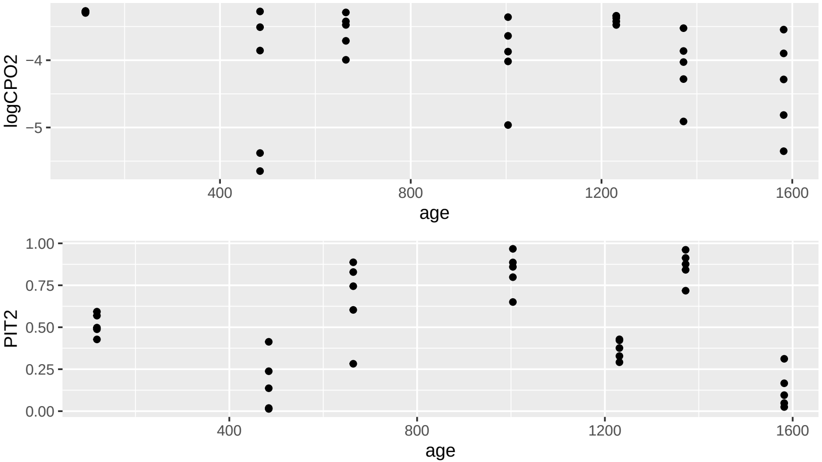
```
m2$summary.random$Tree
```

```
##   ID    mean       sd 0.025quant 0.5quant 0.975quant
## 1  3 0.08188 0.004757    0.07244  0.08191    0.09119 0.0
## 2  1 0.08669 0.004757    0.07724  0.08671    0.09600 0.0
## 3  5 0.10289 0.004759    0.09344  0.10291    0.11220 0.1
## 4  2 0.12640 0.004761    0.11694  0.12643    0.13572 0.1
## 5  4 0.13198 0.004761    0.12252  0.13201    0.14130 0.1
```

```
c(m1=-sum(log(m1$cpo$cpo)), m2=-sum(log(m2$cpo$cpo)))
```

```
##    m1    m2
## 162.5 135.3
```

Orange example, model 2 check



Hierarchical models

Hierarchical models

- Level 1, **Likelihood**: Conditional model for the outcome, $\mathbf{y}$

$$p(\mathbf{y}|\mathbf{x}, \theta, \mathbf{F}, \mathbf{Z}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}, \theta_{\mathbf{1}}, \mathbf{F}_i, \mathbf{Z}_i) \qquad \text{(conditional independence)}$$

- Level 2, **Latent**: prior for $\mathbf{x} = \{\beta, \mathbf{b}\}$

$$p(\mathbf{x}|\theta) = p(\mathbf{x}|\theta_2)$$

- Level 3, **Hyper-parameter**: prior for $\theta = \{\theta_1, \theta_2\}$

$$p(\theta)$$

Latent Gaussian models and INLA

Latent Gaussian models

- Assume a Gaussian distribution for the
  - regression coefficients
  - smoothed effects (basis functions, coefficients)
  - random effects

Latent Gaussian models

- Assume a Gaussian distribution for the
  - regression coefficients
  - smoothed effects (basis functions, coefficients)
  - random effects
- Write the linear predictor as $\eta = \mathbf{Ax}$
  - $\mathbf{A} = [\mathbf{F}, \mathbf{Z}]$
  - $\mathbf{x} = [\beta^T, \mathbf{b}^T]^T$
  - $\mathbf{x}$ is non-observable (latent), assumed Gaussian

Latent Gaussian models

- Assume a Gaussian distribution for the
  - regression coefficients
  - smoothed effects (basis functions, coefficients)
  - random effects
- Write the linear predictor as $\eta = \mathbf{A}\mathbf{x}$
  - $\mathbf{A} = [\mathbf{F}, \mathbf{Z}]$
  - $\mathbf{x} = [\beta^T, \mathbf{b}^T]^T$
  - $\mathbf{x}$ is non-observable (latent), assumed Gaussian
- Latent Gaussian Model - **LGM**
  - Basically, if you have Gaussian distribution for each of the unknowns in the linear predictor you have a LGM

INLA overview

- **Step 1**: Approach $p(\theta|\mathbf{y}) \approx \tilde{p}(\theta|\mathbf{y})$, Rue and Martino (2007)
- Laplace approximation at its mode $\tilde{\theta}$, $\tilde{p}(\tilde{\theta}|y)$

$$
\begin{aligned}
p(\theta|\mathbf{y}) &= \frac{p(\theta)\int p(\mathbf{y}|\mathbf{x},\theta)\partial\mathbf{x}}{p(\mathbf{y})} \\
&\propto p(\theta)\int p(\mathbf{y}|\mathbf{x},\theta)\partial\mathbf{x} \\
&= \frac{p(\theta)(\mathbf{y}|\mathbf{x},\theta)}{p(\mathbf{x}|\mathbf{y},\theta)} \\
&\approx \frac{p(\theta)(\mathbf{y}|\mathbf{x},\theta)}{\tilde{p}_G(\mathbf{x}|\mathbf{y},\theta)}\Big|\Big|_{\mathbf{x}=\tilde{\mathbf{x}}}
\end{aligned}
$$

- When $p(\mathbf{y}|...)$ is Gaussian, there is no approximation here

INLA overview (cont.)

Step 2: Marginals for each $\mathbf{x}_i$

- Classic (INLA): Rue, Martino, and Chopin (2009)
  - Laplace approximation (nested, simplified or full) for $p(x_i|y,\theta)$

$$p(x_i|\mathbf{y},\theta) \approx \frac{p(\theta)p(\mathbf{x}|\theta)p(\mathbf{y}|\mathbf{x},\theta)}{p_{GG}(\mathbf{x}_{-i}|\mathbf{y},\theta)}$$

INLA overview (cont.)

Step 2: Marginals for each $\mathbf{x}_i$

- Classic (INLA): Rue, Martino, and Chopin (2009)
  - Laplace approximation (nested, simplified or full) for $p(x_i|y,\theta)$

$$p(x_i|\mathbf{y},\theta) \approx \frac{p(\theta)p(\mathbf{x}|\theta)p(\mathbf{y}|\mathbf{x},\theta)}{p_{GG}(\mathbf{x}_{-i}|\mathbf{y},\theta)}$$

- Compact (LA + VBC): Niekerk and Rue (2024)
  - select some elements of $\mathbf{x}$: $\mathbf{x}_I$
  - apply a (low rank, implicit) mean correction for $\mathbf{x}$

INLA overview (cont.)

Step 2: Marginals for each $\mathbf{x}_i$

- Classic (INLA): Rue, Martino, and Chopin (2009)
  - Laplace approximation (nested, simplified or full) for $p(x_i|y,\theta)$

$$p(x_i|\mathbf{y},\theta) \approx \frac{p(\theta)p(\mathbf{x}|\theta)p(\mathbf{y}|\mathbf{x},\theta)}{p_{GG}(\mathbf{x}_{-i}|\mathbf{y},\theta)}$$

- Compact (LA + VBC): Niekerk and Rue (2024)
  - select some elements of $\mathbf{x}$: $\mathbf{x}_I$
  - apply a (low rank, implicit) mean correction for $\mathbf{x}$

Step 3: Numerical integration over θ

- Select a **good** set of values for θ (around $\tilde{\theta}$)
  - *eb*: just the mode (empirical Bayes)
  - *grid*: grid around the mode
  - *ccd*: central composite design
- Compute $p(x_i|y)$ and $p(\theta_j|y)$

INLA overview (cont.)

Step 2: Marginals for each $\mathbf{x}_i$

- Classic (INLA): Rue, Martino, and Chopin (2009)
  - Laplace approximation (nested, simplified or full) for $p(x_i|y,\theta)$

$$p(x_i|\mathbf{y},\theta) \approx \frac{p(\theta)p(\mathbf{x}|\theta)p(\mathbf{y}|\mathbf{x},\theta)}{p_{GG}(\mathbf{x}_{-i}|\mathbf{y},\theta)}$$

- Compact (LA + VBC): Niekerk and Rue (2024)
  - select some elements of $\mathbf{x}$: $\mathbf{x}_I$
  - apply a (low rank, implicit) mean correction for $\mathbf{x}$

Step 3: Numerical integration over θ

- Select a **good** set of values for θ (around $\tilde{\theta}$)
  - *eb*: just the mode (empirical Bayes)
  - *grid*: grid around the mode
  - *ccd*: central composite design
- Compute $p(x_i|y)$ and $p(\theta_j|y)$

IF $p(\mathbf{y}|...)$ is Gaussian: **no approximations** in steps 1 and 2

Several models under this framework

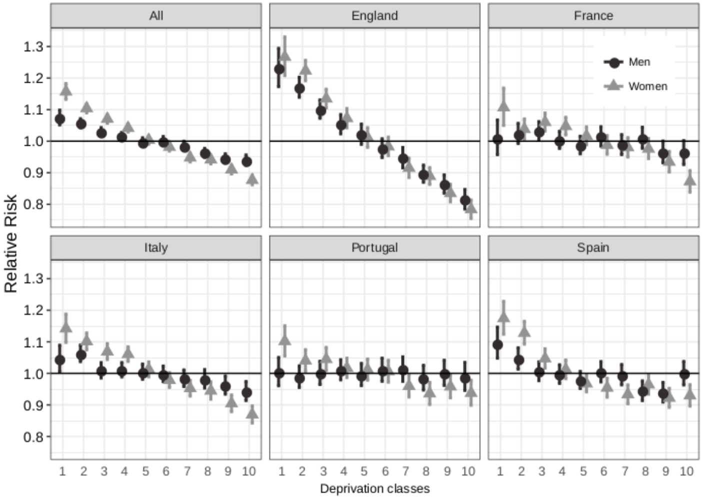
- Generalized (mixed) models
- Generalized additive (mixed) models
- Survival models
- Dynamic models
- Stochastic volatility models
- Smoothing spline
- Semi-parametric regression
- Disease mapping
- Model based geostatistics
- Log-Gaussian Cox processes
- Space-time models
- Semi-parametric regression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

Applications

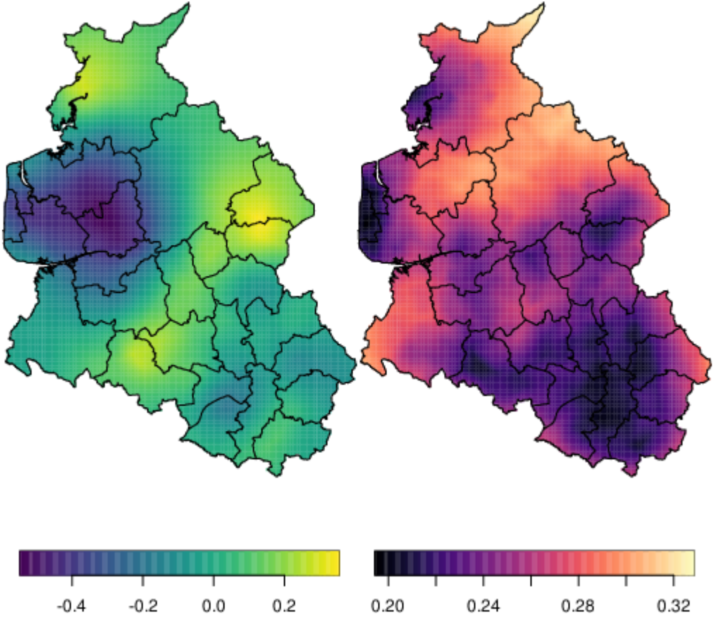
Some applications cited in Rue et al. (2017)

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



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

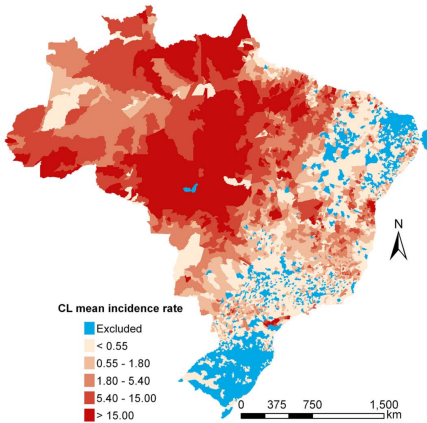


Figure 1. Raw incidence rates. Raw incidence rates (per 10,000) averaged over a 10-year period (2001–2010) for cutaneous leishmaniasis (left) and visceral leishmaniasis (right). Municipalities colored in blue, were excluded from analysis due to missing data.

Malaria in Africa, Gething (2015)

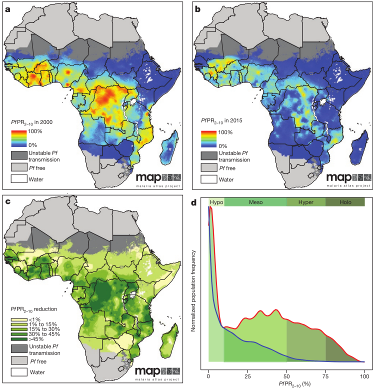
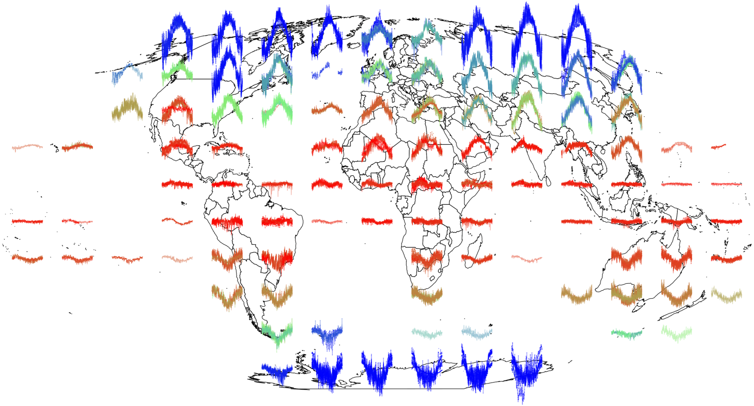


Figure 1 | Changes in infection prevalence 2000–2015. **a**, *Pf*PR$_{2-10}$ for the year 2000 predicted at 5×5 km resolution. **b**, *Pf*PR$_{2-10}$ for the year 2015 predicted at 5×5 km resolution. **c**, Absolute reduction in *Pf*PR$_{2-10}$ from 2000 to 2015. **d**, Smoothed density plot showing the relative distribution of endemic populations by *Pf*PR$_{2-10}$ in the years 2000 (red line) and 2015 (blue line). The frequencies on the vertical axis have been scaled to make the densities visually comparable. The classical endemicity categories are shown for reference in green shades. Results shown in all panels are derived from a Bayesian geostatistical model fitted to $n = 27{,}573$ *Pf*PR survey points; $n = 24{,}868$ ITN survey points; $n = 96$ national survey reports of ACT coverage; $n = 688$ country-year reports on ITN, ACT and IRS distribution by national programs; and $n = 20$ environmental and socioeconomic covariate grids. Maps in **a**–**c** are available from the Malaria Atlas Project (http://www.map.ox.ac.uk/) under the Creative Commons Attribution 3.0 Unported License.

Non-separable space-time modeling in the globe



Flexibility must come with responsability

- PC-prior ***P****enalized* ***C****omplexity* prior
  - Simpson et al. (2016)
  - Fuglstad et al. (2020)

References

Fuglstad, Geir-Arne, Ingeborg Gullikstad Hem, Alexander Knight, Håvard Rue, and Andrea Riebler. 2020. "Intuitive Joint Priors for Variance Parameters." *Bayeian Analysis* 15 (4): 1109–37.

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.

Niekerk, Janet van, and Haavard Rue. 2024. "Low-Rank Variational Bayes Correction to the Laplace Method." *Journal of Machine Learning Research* 25 (62): 1–25. http://jmlr.org/papers/v25/21-1405.html.

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 S. Martino. 2007. "Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Field Models." *Journal of Statistical Planning and Inference* 137: 3711–3192.

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., J. B. Illian, F. Lindren, S. H Sørbye, and H. Rue. 2016. "Going Off Grid: Computationally Efficient Inference for Log-Gaussian Cox Processes." *Biometrika* 103 (1): 49–70.

van Niekerk, J., E. T. Krainski, D. Rustand, and H. Rue. 2023. "A New Avenue for Bayesian Inference with INLA." https://doi.org/10.1016/j.csda.2023.107692.