

# A practical introduction to Latent Gaussian Models with INLA

Elias T. Krainski

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

**Laboratório de Estatística e Geoinformação**, August 2019



# What is INLA?

- **The short answer:**

- INLA: fast method to do Bayesian inference for latent Gaussian models
- INLA: R-package that implements this method

# What is INLA?

- **The short answer:**

- INLA: fast method to do Bayesian inference for latent Gaussian models
- INLA: R-package that implements this method

- **A much longer answer:**

- Is in the paper H. 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

# What is INLA?

- **The short answer:**
  - INLA: fast method to do Bayesian inference for latent Gaussian models
  - INLA: R-package that implements this method
- **A much longer answer:**
  - Is in the paper H. 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
  - Easier reading: H. Rue et al. (2017) **Bayesian Computing with INLA: A Review.** *Annual Review of Statistics and Its Application* 4, 395–421.

# What is INLA?

- **The short answer:**
  - INLA: fast method to do Bayesian inference for latent Gaussian models
  - INLA: R-package that implements this method
- **A much longer answer:**
  - Is in the paper H. 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
  - Easier reading: H. Rue et al. (2017) **Bayesian Computing with INLA: A Review.** *Annual Review of Statistics and Its Application* 4, 395–421.
  - some books around, see [r-inla.org](http://r-inla.org)
  - a video in YouTube

# Why use INLA?

- Because it can handle a large class of models
- Because it is faster
- “If we’re going to be wrong, we might as well be wrong quickly.” - D. Simpson

# Informations? <http://www.r-inla.org>

## The R-INLA project

Search this site

The home of the  
R-INLA project

A. The SPDE book

Bayesian Models

Books

Contact us, stay  
updated, get help or  
report an error

Discussion forum

Download

▼ Examples and  
tutorials

Case studies and  
code from papers

N-Mixture models

Small tutorials by

Haakon Bakka

Tutorials

Volume I

Volume II

FAQ

Help

Internal use

▼ Models

Latent models,  
likelihoods and  
priors.

Tools to manipulate  
models and  
likelihoods

▼ News

"A nonita INLA"

## Bayesian computing with INLA !

This site provides documentation to the [R-INLA package](#) which solves a large class of statistical models using the [INLA](#) approach.

### Recent posts to the discussion group

Google Group

NOVO TEMA

C

Ajuda

Pesquisa

R-inla discussion group

Compartilhado publicamente

30 de muitos temas

Sobre

Best,

H

Welcome to this discussion group about r-inla. Please ask your questions here in case you think they will be useful for others, otherwise send them to [help@r-inla.org](mailto:help@r-inla.org). You are of course free to comment on questions from others as well.

### Recent announcements

#### Recent Announcements

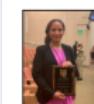
[Code for the report "Improving Bayesian Local Spatial Models in Large Data Sets"](#) Code for the above arxiv-report by A.Lenzi et al, is made available here.

Posted 22 Jul 2019, 05:10 by Havard Rue



[Statsref article about INLA](#) Sara and Andrea have submitted a contribution (as far as I know) to Wiley Statsref about INLA. Worth reading. H

Posted 7 Jul 2019, 04:15 by Havard Rue



[Congratulations!!!](#) Centers for Disease Control and Prevention (Atlanta, USA) awarded the paper A BAYESIAN SPATIAL AND TEMPORAL MODELING APPROACH TO MAPPING GEOGRAPHIC VARIATION IN MORTALITY RATES FOR SUBNATIONAL AREAS WITH R-INLA ...

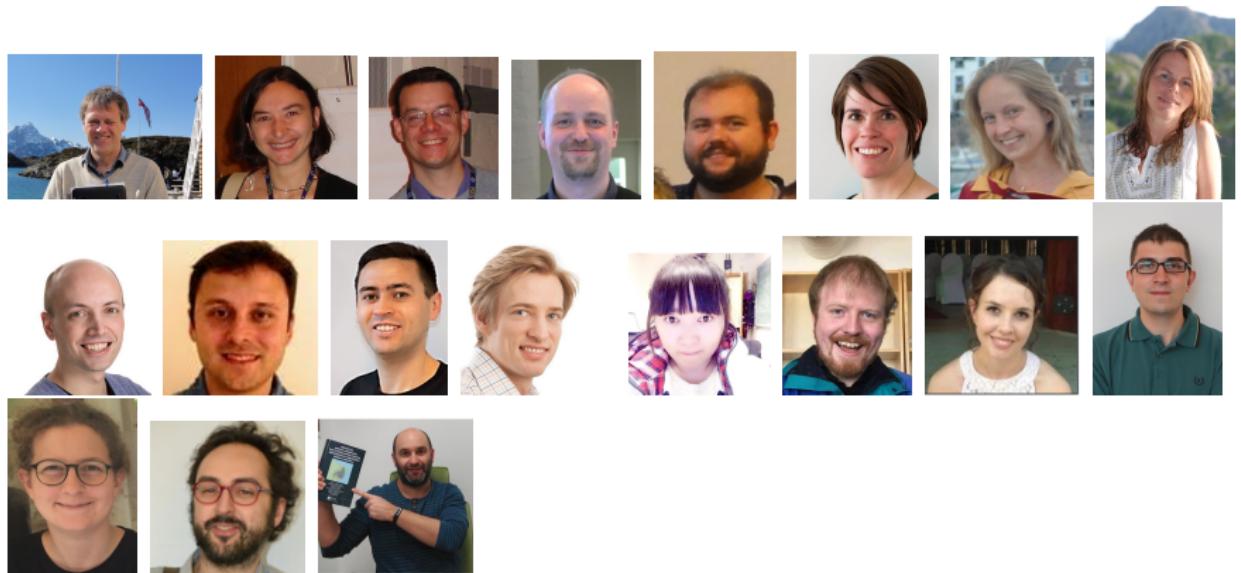
Posted 26 Jun 2019, 13:04 by Havard Rue

[Missing links](#) There may be some missing links when the web server was moved

# Who is involved in working with INLA



.. (certainly not all the) people working around INLA



## So... 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

# So... 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

- **The core**

- We have questions
- We observe/collect some data.
- We want answers

# So... 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

- **The core**

- We have questions
- We observe/collect some data.
- We want answers

- **How do we find answers?**

- We need to make choices:
  - Bayesian or frequentist?
  - How do we model the data?
  - How do we compute the answer?

- **These questions are *not* independent.**

# Sumário

1 The basic model idea

2 Extending the basic model

3 Hierarchical models

4 INLA

5 References

# Basic statistical model structure

- Observations of a phenomena may follow the model

$$\mathbf{y} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{e}$$

- $\mathbf{y}$  is the observation
- $\mu(\mathbf{F}, \boldsymbol{\beta})$  is the explanation
  - if it is a linear model, then

$$\mu(\mathbf{F}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 \mathbf{F}_{i,1} + \dots + \beta_p \mathbf{F}_{i,p}$$

- $\mathbf{e}$  is the unexplained part

# Basic statistical model structure

- Observations of a phenomena may follow the model

$$\mathbf{y} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{e}$$

- $\mathbf{y}$  is the observation
- $\mu(\mathbf{F}, \boldsymbol{\beta})$  is the explanation
  - if it is a linear model, then

$$\mu(\mathbf{F}_i, \boldsymbol{\beta}) = \beta_0 + \beta_1 \mathbf{F}_{i,1} + \dots + \beta_p \mathbf{F}_{i,p}$$

- $\mathbf{e}$  is the unexplained part
- The “explanation part” may not be the “truth”
  - choose  $\mu(.,.)$  that reduces  $\mathbf{e}$
  - there may be some options for  $\mu(.,.)$
  - $\mu(.,.)$  is “a vision of the world”

# Basic statistical model structure contd.

- Observations of a phenomena may follow the model

$$\mathbf{y} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{e}$$

- $\mathbf{y}$  is the observation
- $\mu(\mathbf{F}, \boldsymbol{\beta})$  is the explanation
- $\mathbf{e}$  is the unexplained part

# Basic statistical model structure contd.

- Observations of a phenomena may follow the model

$$\mathbf{y} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{e}$$

- $\mathbf{y}$  is the observation
  - $\mu(\mathbf{F}, \boldsymbol{\beta})$  is the explanation
  - $\mathbf{e}$  is the unexplained part
- 
- Statistics at this point (more to come):
    - $\mathbf{e}$  follows a probability distribution
    - $\mu(.,.)$  may be a simplification
    - *all the models are wrong, but some are useful*

# The statistical modeling problem

- Propose  $\mu(F, \beta)$  that
  - sets  $e$  as completely random
    - i.e. no other information available to explain  $e$

# The statistical modeling problem

- Propose  $\mu(F, \beta)$  that
  - sets  $e$  as completely random
    - i.e. no other information available to explain  $e$
- For a given  $\mu(F, \beta)$ ,  $\beta$  is unknown
  - estimate  $\beta$

# The statistical modeling problem

- Propose  $\mu(F, \beta)$  that
  - sets  $e$  as completely random
    - i.e. no other information available to explain  $e$
- For a given  $\mu(F, \beta)$ ,  $\beta$  is unknown
  - estimate  $\beta$
- Account for uncertainty

# The linear predictor

Suppose  $\mu(F, \beta)$  is a linear function of  $\beta$  on  $F$ , the linear predictor is

$$\mu(F_i, \beta) = \beta_0 + \beta_1 F_{i,1} + \dots + \beta_p F_{i,p}$$

- this can be written as  $E(\mathbf{y}|F, \beta)$ , where we model the expected value of  $\mathbf{y}$  conditional on  $F$  and  $\beta$

# The linear predictor

Suppose  $\mu(F, \beta)$  is a linear function of  $\beta$  on  $F$ , the linear predictor is

$$\mu(F_i, \beta) = \beta_0 + \beta_1 F_{i,1} + \dots + \beta_p F_{i,p}$$

- this can be written as  $E(\mathbf{y}|F, \beta)$ , where we model the expected value of  $\mathbf{y}$  conditional on  $F$  and  $\beta$
- $\mathbf{F}$  includes the design matrix, factors, explanatory variables, covariates, independent variables, etc.
  - usually it is assumed to be fixed

# The linear predictor

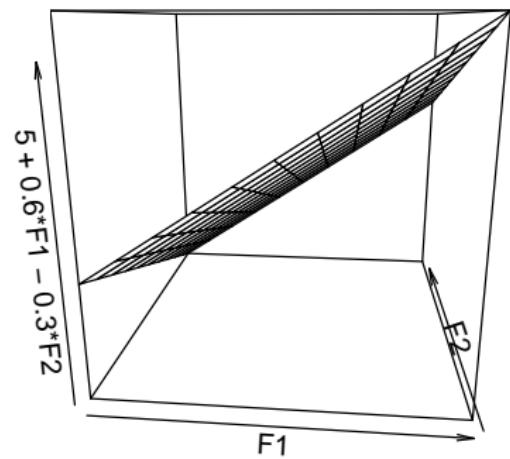
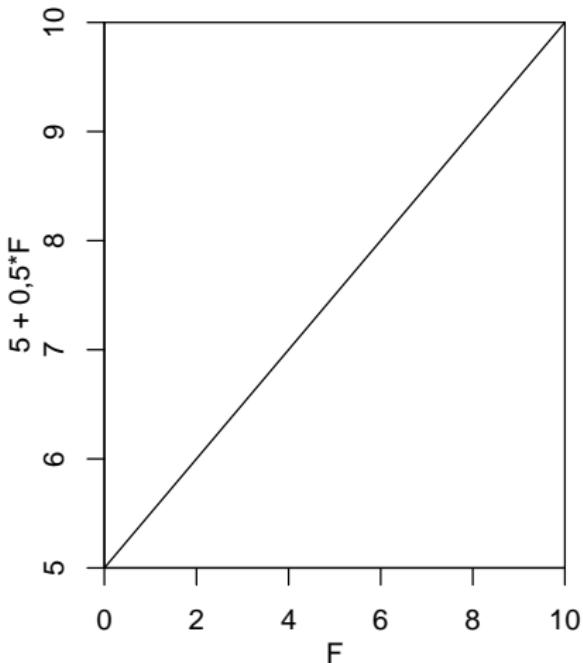
Suppose  $\mu(F, \beta)$  is a linear function of  $\beta$  on  $F$ , the linear predictor is

$$\mu(F_i, \beta) = \beta_0 + \beta_1 F_{i,1} + \dots + \beta_p F_{i,p}$$

- this can be written as  $E(\mathbf{y}|F, \beta)$ , where we model the expected value of  $\mathbf{y}$  conditional on  $F$  and  $\beta$
- $\mathbf{F}$  includes the design matrix, factors, explanatory variables, covariates, independent variables, etc.
  - usually it is assumed to be fixed
- $\beta$  is a vector of unknown *constants*
  - regression coefficients (measure the effect of the covariates)
  - usually are the parameters of main interest

# About the coefficients

- the effect of  $F_j$  is constant ( $\beta_j$ ) among the range of  $F_j$  values
- It is a hyper-plane on the  $p$  dimensional space

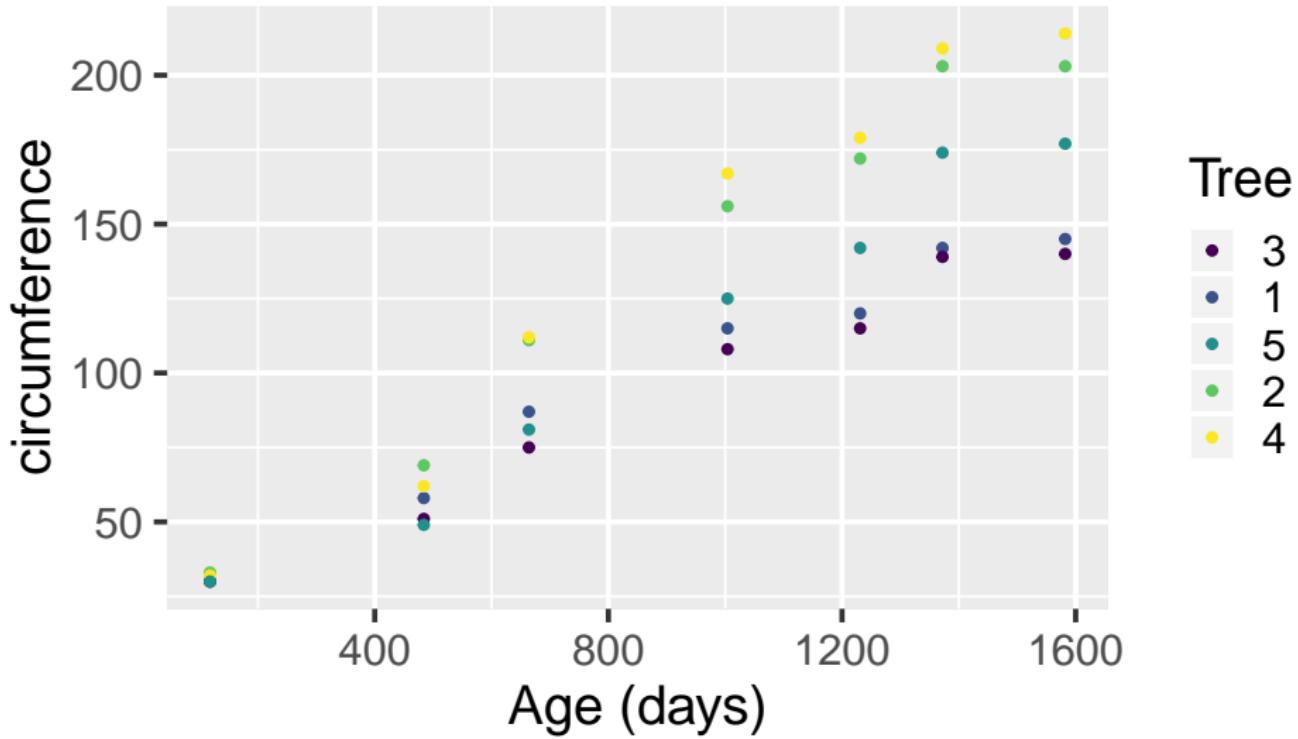


# Orange data

```
##      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

- **model 1:** circumference increases as age increases

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

No accounting for tree.

# Orange: model 1

- **model 1:** circumference increases as age increases

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

No accounting for tree.

- Outcome (circunference):  $\mathbf{y} = (y_1, \dots, y_n)$
- Covariate (age):  $\mathbf{F} = (F_1, \dots, F_n)$

$$\mathbb{E}(y_i) = \beta_0 + \beta_1 F_i, \quad \text{Var}(y_i) = \tau^{-1}, \quad i = 1, \dots, n$$

# On the common linear model

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

# On the common linear model

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

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

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

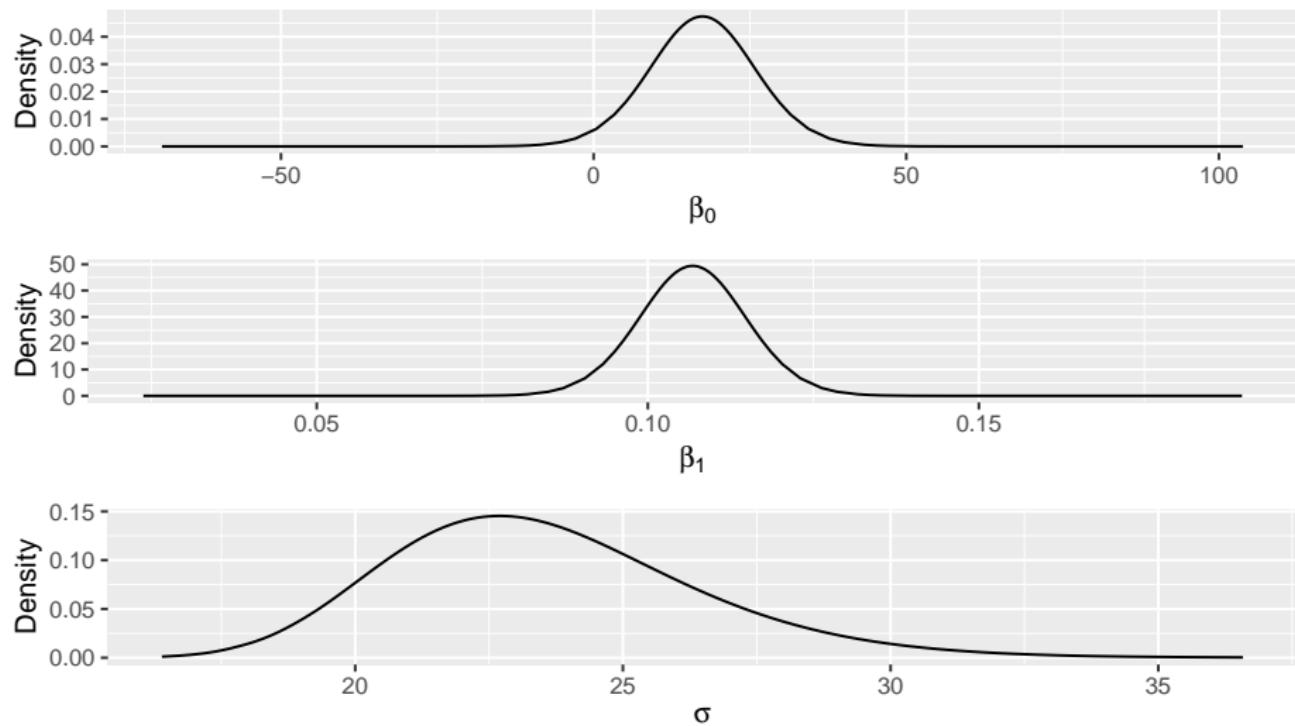
```
m1 <- inla(circumference ~ age, data=Orange,  
            control.compute = list(cpo = TRUE))  
m1$summary.fixed
```

```
##               mean        sd 0.025quant 0.5quant 0.975quant  
## (Intercept) 17.400 8.59090      0.4363   17.399   34.350 1  
## age         0.107 0.00825      0.0905    0.107    0.123
```

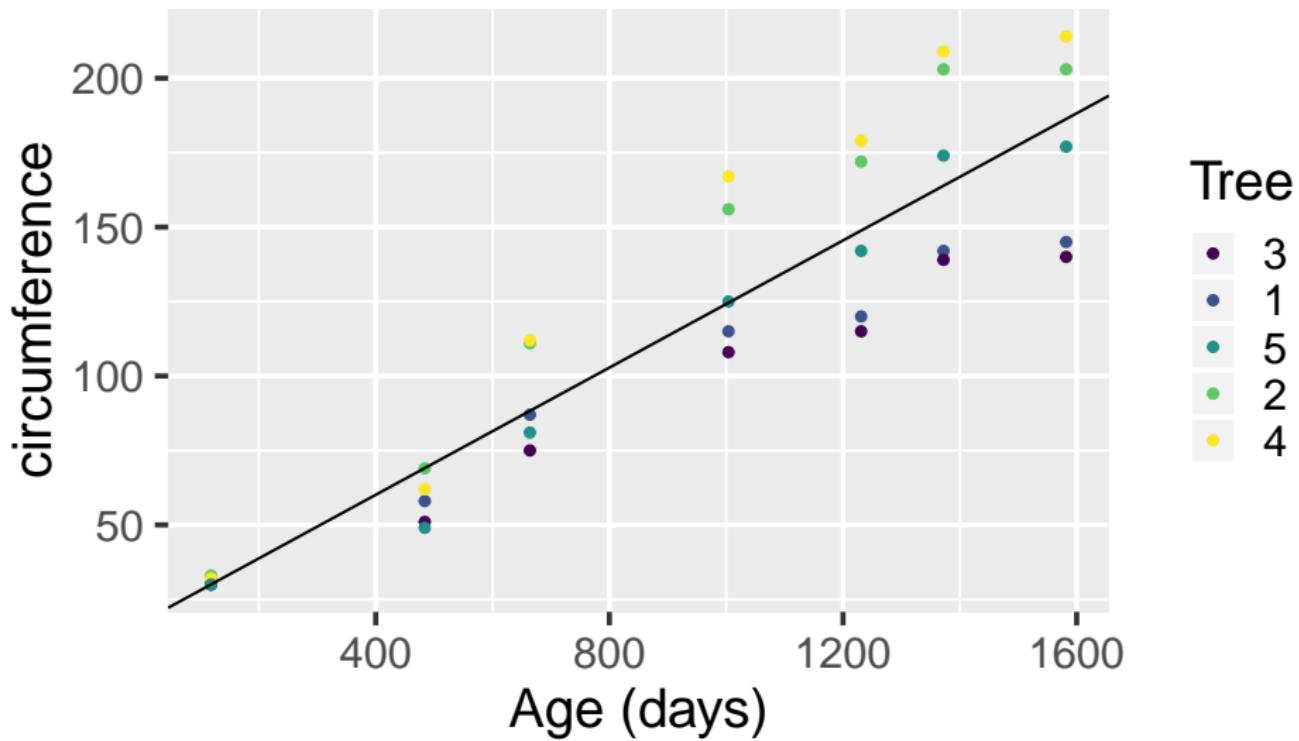
```
m1$summary.hyperpar[1,]
```

```
##               mean        sd 0.  
## Precision for the Gaussian observations 0.00188 0.000449  
##                                         0.5quant 0.975quant  
## Precision for the Gaussian observations 0.00185 0.00286
```

# Posterior marginals



## Model 1 fit



# Goodness-of-fit measures

- **Conditional Predictive Ordinate** - CPO:

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

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

- useful for model comparison

# Goodness-of-fit measures

- **Conditional Predictive Ordinate** - CPO:

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

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

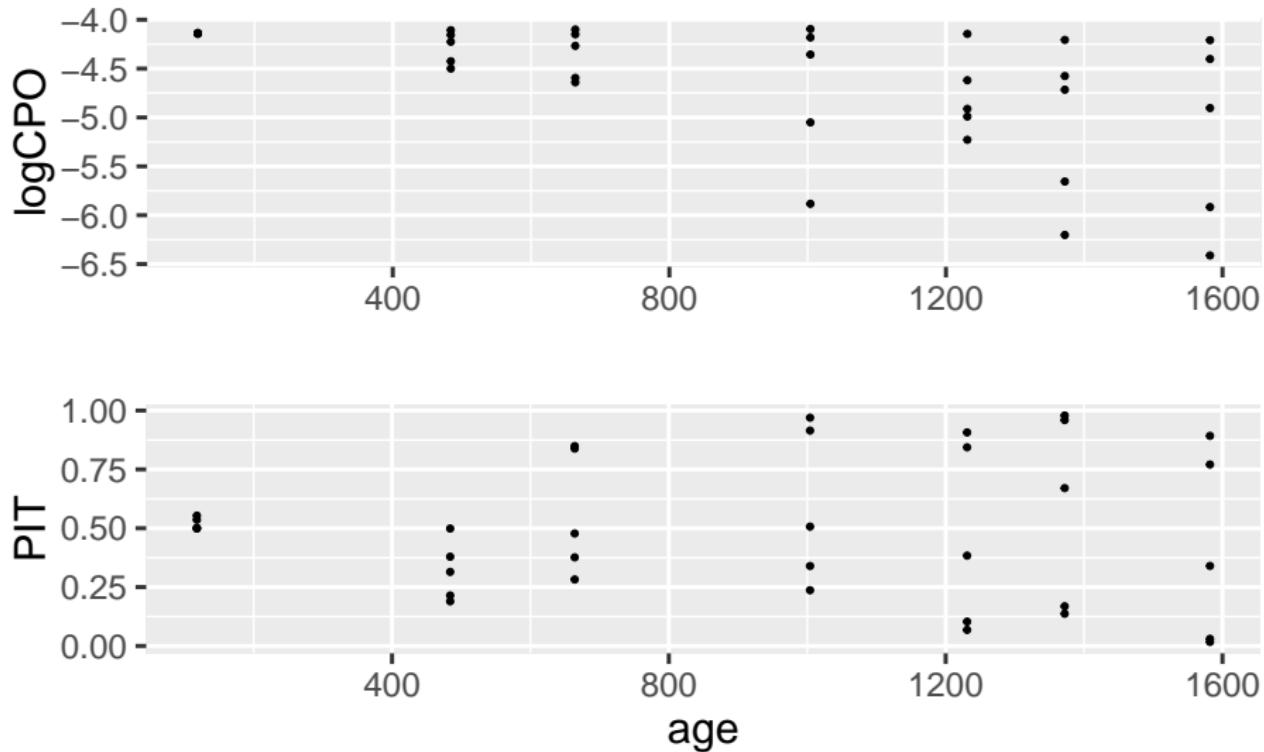
- useful for model comparison

- **Probability Integral Transform** - PIT:

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

- useful to detect lack of fit or outliers

## Orange: model 1 check



# Sumário

- 1 The basic model idea
- 2 Extending the basic model
- 3 Hierarchical models
- 4 INLA
- 5 References

## Orange example: effect for each tree

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

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

- $\beta_0$  and  $\beta_j$ ,  $j$  for each tree, are unknown
- Now we have:  $\mathbf{y} | \underbrace{\beta_0, \beta_1, \dots, \beta_5}_{\mathbf{x}}, \underbrace{\tau}_{\theta}$

# About the coefficients

- Usually it is assumed that  $\beta_j$  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  $\beta_j$  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  $\beta_j$  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  $\tau_0$  fixed
  - $\beta = \{\beta_0, \beta_1, \dots, \beta_5\}$  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 (unobservable and unknown)
- Fixed effects are also random
- They only differ in the prior we put on them

## Orange example, model 2

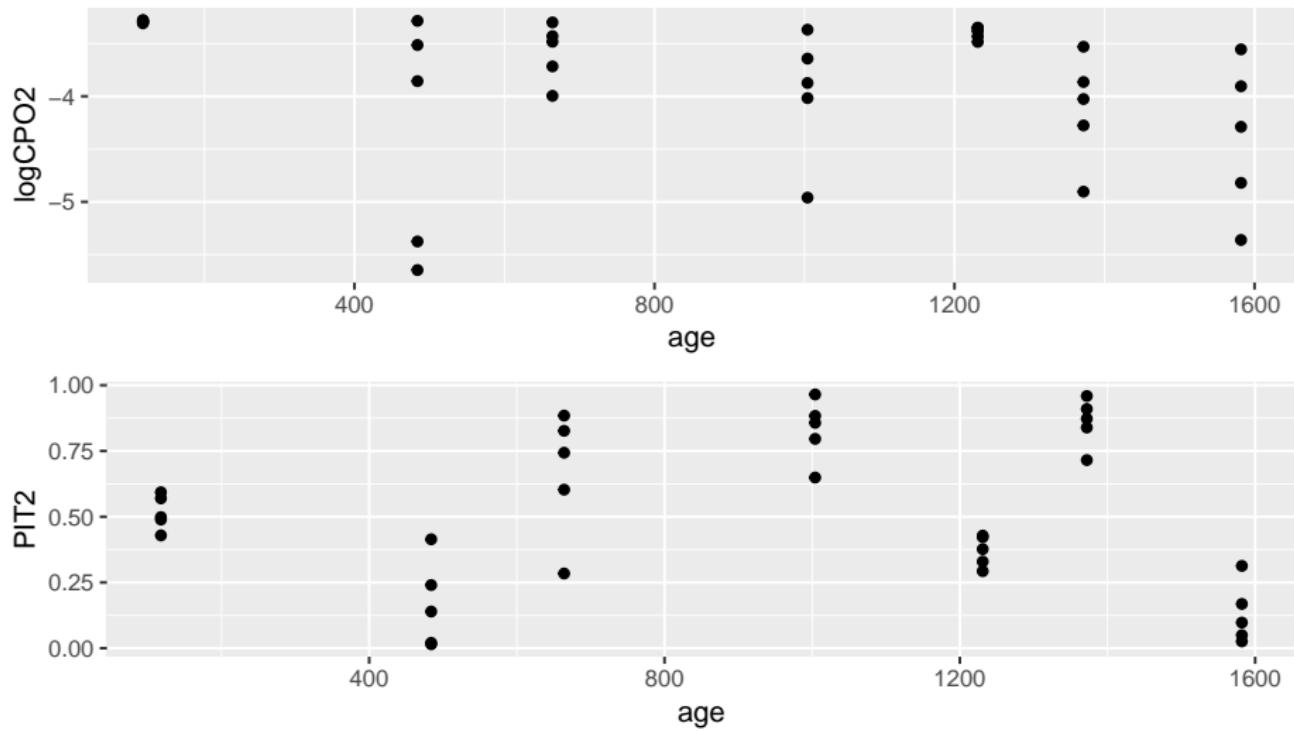
```
f2 <- circumference ~ 1 + f(Tree, age, model='iid')
m2 <- inla(f2, data=Orange, control.compute=list(cpo=TRUE))
m2$summary.fixed

##           mean      sd 0.025quant 0.5quant 0.975quant mod
## (Intercept) 18.11 3.674        10.91    18.08    25.43 18.0

m2$summary.random$Tree

##   ID      mean      sd 0.025quant 0.5quant 0.975quant mod
## 1  3 0.08192 0.004807  0.07234  0.08194  0.09134 0.0819
## 2  1 0.08672 0.004808  0.07715  0.08675  0.09615 0.0867
## 3  5 0.10293 0.004809  0.09335  0.10295  0.11235 0.1030
## 4  2 0.12644 0.004812  0.11685  0.12647  0.13587 0.1265
## 5  4 0.13202 0.004812  0.12243  0.13205  0.14145 0.1321
```

# Orange example, model 2 check



# Extending the model framework

- So far the basic (linear) model

$$\begin{aligned}\mathbf{y} &= \mathbf{F}\boldsymbol{\beta} + \mathbf{e} \\ &= \boldsymbol{\eta} + \mathbf{e}\end{aligned}$$

does not solves all the problems

# Extending the model framework

- So far the basic (linear) model

$$\begin{aligned}\mathbf{y} &= \mathbf{F}\boldsymbol{\beta} + \mathbf{e} \\ &= \boldsymbol{\eta} + \mathbf{e}\end{aligned}$$

does not solves all the problems

- random effects

$$\boldsymbol{\eta} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{Z}\mathbf{b}$$

- **Exercice:** define  $\mathbf{F}$  in model 2 for the orange data

# Extending the model framework

- So far the basic (linear) model

$$\begin{aligned}\mathbf{y} &= \mathbf{F}\boldsymbol{\beta} + \mathbf{e} \\ &= \boldsymbol{\eta} + \mathbf{e}\end{aligned}$$

does not solve all the problems

- random effects

$$\boldsymbol{\eta} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{Z}\mathbf{b}$$

- Exercise:** define  $\mathbf{F}$  in model 2 for the orange data
- non-linear effects

- work more on  $\eta = \mu(F, \beta)$ . Example:  $\eta_i = \alpha e^{\beta F_i}$

# Extending the model framework

- So far the basic (linear) model

$$\begin{aligned}\mathbf{y} &= \mathbf{F}\boldsymbol{\beta} + \mathbf{e} \\ &= \boldsymbol{\eta} + \mathbf{e}\end{aligned}$$

does not solves all the problems

- random effects

$$\boldsymbol{\eta} = \mu(\mathbf{F}, \boldsymbol{\beta}) + \mathbf{Z}\mathbf{b}$$

- **Exercice:** define  $\mathbf{F}$  in model 2 for the orange data
- non-linear effects
  - work more on  $\boldsymbol{\eta} = \mu(F, \boldsymbol{\beta})$ . Example:  $\eta_i = \alpha e^{\beta F_i}$
- non-Gaussian outcomes
  - $p(\mathbf{y}|...)$  may be non-Gaussian

# Salmonella example

Breslow (1984) analyses some mutagenicity assay data (shown below) on salmonella in which three plates have been processed at each dose  $i$  of quinoline and the number of revertant colonies of TA98 Salmonella measured. A certain dose-response curve is suggested by theory.

dose of quinoline ( $\mu\text{g}$ per plate)						
0	10	33	100	333	1000	
15	16	16	27	33	20	
21	18	26	41	38	27	
29	21	33	69	41	42	

Figure 1: Salmonella data

## Salmonella model

This is assumed to be a random effects Poisson model allowing for over-dispersion. Let  $x_i$  be the dose on the plates  $i = 1, 2$  and  $3$ . Then we assume

$$y_{ij} \sim \text{Poisson}(m_{ij})$$

$$\log(m_{ij}) = a + b \log(x_i + 10) + g x_i + l_{ij}$$

$$l_{ij} \sim \text{Normal}(0, t)$$

$a, b, g, t$  are given independent ``noninformative" priors. The appropriate

Figure 2: Salmonella model

# Salmonella model fit

```
data(Salm)
head(Salm)
```

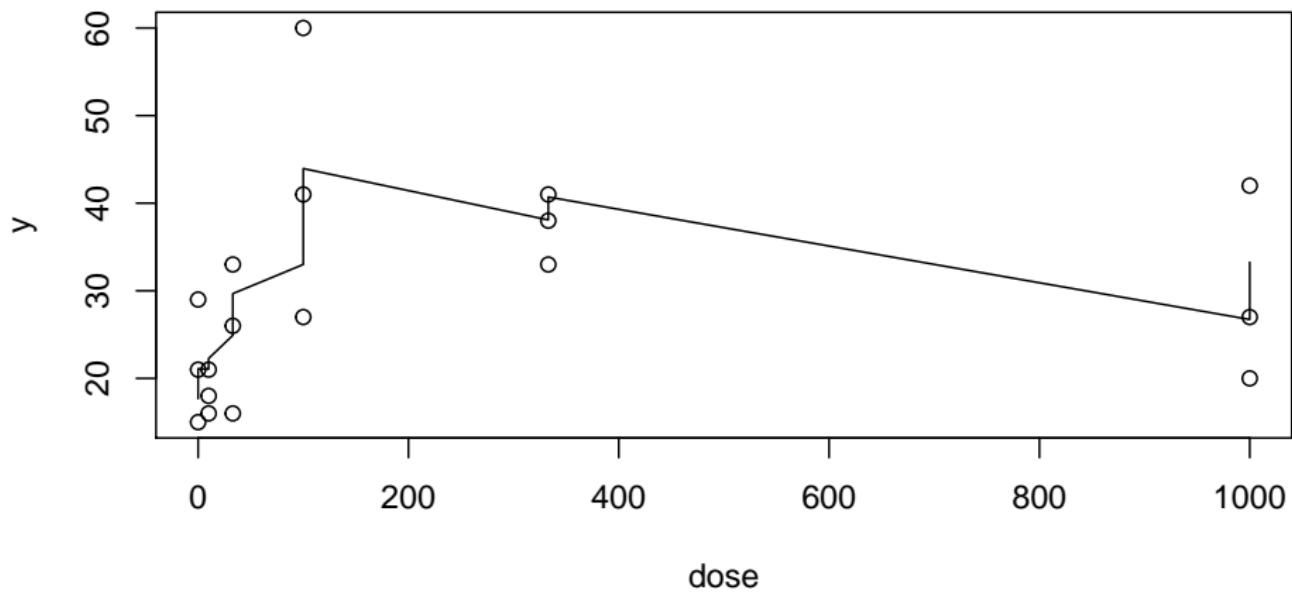
```
##      y dose rand
## 1 15    0    1
## 2 21    0    2
## 3 29    0    3
## 4 16   10    4
## 5 18   10    5
## 6 21   10    6
```

```
salmm <- inla(y ~ log(dose+10) + dose + f(rand, model='iid'),
                 family='Poisson', data=Salm,
                 control.compute=list(cpo=TRUE))
```

# Salmonella model results

```
##               mean     sd 0.025quant 0.5quant 0.975quant mod
## (Intercept) 2.17 0.28      1.61      2.17      2.72 2.17
## log(dose + 10) 0.32 0.07      0.16      0.32      0.46 0.32
## dose        0.00 0.00      0.00      0.00      0.00 0.00
## 
##               mean     sd 0.025quant 0.5quant 0.975quant
## Precision for rand 8205 15938                  9          71      5553
```

# Salmonella model fit result



# Epilepsia example

Breslow and Clayton (1993) analyse data initially provided by Thall and Vail (1990) concerning seizure counts in a randomised trial of anti-convulsant therapy in epilepsy. The table below shows the successive seizure counts for 59 patients. Covariates are treatment (0,1), 8-week baseline seizure counts, and age in years. The structure of this data is shown below

Patient	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	Trt	Base	Age
1	5	3	3	3	0	11	31
2	3	5	3	3	0	11	30
3	2	4	0	5	0	6	25
4	4	4	1	4	0	8	36
...							
8	40	20	21	12	0	52	42
9	5	6	6	5	0	12	37
...							
59	1	4	3	2	1	12	37

# Epilepsia example, model

We consider model *III* of Breslow and Clayton (1993), in which Base is transformed to  $\log(\text{Base}/4)$  and Age to  $\log(\text{Age})$ , and a Treatment by  $\log(\text{Base}/4)$  interaction is included. Also present are random effects for both individual subjects  $b_{1j}$  and also subject by visit random effects  $b_{jk}$  to model extra-Poisson variability within subjects.  $V_4$  is an indicator variable for the 4th visit.

$$y_{jk} \sim \text{Poisson}(m_{jk})$$

$$\log m_{jk} = a_0 + a_{\text{Base}} \log(\text{Base}_j / 4) + a_{\text{Trt}} \text{Trt}_j + a_{\text{BT}} \text{Trt}_j \log(\text{Base}_j / 4) + a_{\text{Age}} \text{Age}_j + a_{V4} V_4 + b_{1j} + b_{jk}$$

$$b_{1j} \sim \text{Normal}(0, t_{b1})$$

$$b_{jk} \sim \text{Normal}(0, t_b)$$

Coefficients and precisions are given independent "noninformative" priors.

# Epilepsia data

```
data(Epil)  
head(Epil)
```

```
##   y Trt Base Age V4 rand Ind  
## 1 5   0   11  31  0    1    1  
## 2 3   0   11  31  0    2    1  
## 3 3   0   11  31  0    3    1  
## 4 3   0   11  31  1    4    1  
## 5 3   0   11  30  0    5    2  
## 6 5   0   11  30  0    6    2
```

- See Epil-example.R

# Sumário

- 1 The basic model idea
- 2 Extending the basic model
- 3 Hierarchical models
- 4 INLA
- 5 References

# Hierarchical models, level 1

- **Likelihood**, the conditional model for the outcome,  $\mathbf{y}$

$$\mathbf{y}|\mathbf{x}, \theta_1 \sim \pi(\mathbf{y}|\mathbf{x}, \theta_1) = \prod_{i=1}^n \pi(y_i|x_i, \theta_1)$$

(conditional independence)

# Hierarchical models, level 1

- **Likelihood**, the conditional model for the outcome,  $\mathbf{y}$

$$\mathbf{y}|\mathbf{x}, \theta_1 \sim \pi(\mathbf{y}|\mathbf{x}, \theta_1) = \prod_{i=1}^n \pi(y_i|x_i, \theta_1)$$

(conditional independence)

- $\mathbf{x}$ , see H. Rue et al. (2017) for an example
  - $x_i$ , for  $i = 1, \dots, n$  is the linear predictor
  - $x_j$ ,  $j > n$  includes fixed and random effects

# Hierarchical models, level 1

- **Likelihood**, the conditional model for the outcome,  $\mathbf{y}$

$$\mathbf{y}|\mathbf{x}, \theta_1 \sim \pi(\mathbf{y}|\mathbf{x}, \theta_1) = \prod_{i=1}^n \pi(y_i|x_i, \theta_1)$$

(conditional independence)

- $\mathbf{x}$ , see H. Rue et al. (2017) for an example
  - $x_i$ , for  $i = 1, \dots, n$  is the linear predictor
  - $x_j$ ,  $j > n$  includes fixed and random effects
- $\theta_1$ : likelihood extra parameter
  - example: variance (dispersion), zero inflation

# Likelihood:

The likelihood,  $\pi(\mathbf{y}|\mathbf{x}, \theta)$  depends on

- the kind of response
  - binary, counts, continuous, censored

# Likelihood:

The likelihood,  $\pi(\mathbf{y}|\mathbf{x}, \theta)$  depends on

- the kind of response
  - binary, counts, continuous, censored
- how is it collected
  - usually each individual has only one observation
  - possible for more than one
  - *Unusual example:* point process (point pattern) where we only have the locations of a set of events Can you explain in the course what you mean by having only the locations of a set of events?

# Hierarchical model, level 2

- the model for the random effect
  - not observable, latent
  - assumed to have a probability distribution
    - usually Gaussian → INLA

$$\mathbf{x}|\theta_2 \sim \pi(\mathbf{x}|\theta_2) = N(\mathbf{0}, \mathbf{Q}(\theta_2)^{-1})$$

# Hierarchical model, level 2

- the model for the random effect
  - not observable, latent
  - assumed to have a probability distribution
    - usually Gaussian → INLA

$$\mathbf{x}|\theta_2 \sim \pi(\mathbf{x}|\theta_2) = N(\mathbf{0}, \mathbf{Q}(\theta_2)^{-1})$$

- this distribution has its own parameters,  $\theta_2$ , the hyper-parameters

# Random effect distribution:

- The random effect distribution,  $\pi(\mathbf{x}|\mathbf{Q}(\theta))$  is
  - Non-observable (thus latent)
  - if Gaussian  $\rightarrow$  latent Gaussian
  - Markovian  $\rightarrow \mathbf{Q}(.)$  sparse  $\rightarrow$  computational benefits

# Random effect distribution:

- The random effect distribution,  $\pi(\mathbf{x}|\mathbf{Q}(\theta))$  is
  - Non-observable (thus latent)
  - if Gaussian  $\rightarrow$  latent Gaussian
  - Markovian  $\rightarrow \mathbf{Q}(\cdot)$  sparse  $\rightarrow$  computational benefits
- It represents
  - covariate effects (coefficients or smoothed effects)
  - random effects (individuals, temporal, spatial)
    - unstructured or structured

# Random effect distribution:

- The random effect distribution,  $\pi(\mathbf{x}|\mathbf{Q}(\theta))$  is
  - Non-observable (thus latent)
  - if Gaussian  $\rightarrow$  latent Gaussian
  - Markovian  $\rightarrow \mathbf{Q}(.)$  sparse  $\rightarrow$  computational benefits
- It represents
  - covariate effects (coefficients or smoothed effects)
  - random effects (individuals, temporal, spatial)
    - unstructured or structured
- It can be
  - unstructured (independent, non-correlated individuals)
  - structured (dependent, correlated, similar neighbour effects)
  - more than one structure (or level) combined

# Hierarchical model, level 3

- if Bayesian
  - assumed a distribution for the hyper-parameters

# Hierarchical model, level 3

- if Bayesian
  - assumed a distribution for the hyper-parameters
- have  $\theta = \{\theta_1, \theta_2\}$

$$\theta \sim \pi(\theta)$$

# Prior distribution for the hyper-parameters $\theta$ : $\pi(\theta)$

- likelihood examples
  - precision parameter
    - Normal, gamma, beta, binomial negative
  - zero inflation probability
- random effect examples
  - random effect precision parameter
  - correlation parameter
  - range parameter

# Hierarchical Model Summary

What are the

- ① distribution of the responses?
- ② distribution of the underlying unobserved (latent) components?

# Hierarchical Model Summary

What are the

- ① distribution of the responses?
- ② distribution of the underlying unobserved (latent) components?

... *if Bayesian*

- ③ prior beliefs about the parameters (distribution) on the hyper-parameters in the model?

# Latent Gaussian models

- Assume a Gaussian distribution for the
  - regression coefficients
  - smoothed effects
  - random effects

# Latent Gaussian models

- Assume a Gaussian distribution for the
  - regression coefficients
  - smoothed effects
  - random effects
- Latent Gaussian Model - **LGM**
  - Basically, if you have Gaussian distribution for each of the unknowns in the linear predictor you have a LGM

# Sumário

- 1 The basic model idea
- 2 Extending the basic model
- 3 Hierarchical models
- 4 INLA
- 5 References

# What is INLA?

- Integrated Nested Laplace Approximations
- Short answer: **fast** method for Bayesian inference on **LGM**
- More details: see H. Rue, Martino, and Chopin (2009)
  - Recommended to start with the review in H. Rue et al. (2017)

# INLA overview

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$

# INLA overview

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$
- Step 1: approach  $p(\theta|y) \approx \tilde{p}(\theta|y)$ 
  - Laplace approximation at its mode  $\tilde{\theta}$ ,  $\tilde{p}(\tilde{\theta}|y)$
  - select a **good** set of values for  $\theta$  around  $\tilde{\theta}$ 
    - *eb*: just the mode (empirical Bayes)
    - *grid*: grid around the mode
    - *ccd*: central composite design

# INLA overview contd

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$

## INLA overview contd

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$
- Step 2: approach  $p(x_i|y, \theta) \approx \tilde{p}(x_i|y, \theta)$ 
  - for a set of values of  $\theta$
  - Gaussian, adaptive, simplified Laplace or (full) Laplace approximation

# INLA overview contd

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$
- Step 2: approach  $p(x_i|y, \theta) \approx \tilde{p}(x_i|y, \theta)$ 
  - for a set of values of  $\theta$
  - Gaussian, adaptive, simplified Laplace or (full) Laplace approximation
- Step 3: approach  $p(x_i|y)$  and  $p(\theta_j|y)$ 
  - numerical integration over  $\theta$

# INLA overview contd

- Integrated Nested Laplace Approximations for  $p(\theta_j|y)$  and  $p(x_i|y)$
- Step 2: approach  $p(x_i|y, \theta) \approx \tilde{p}(x_i|y, \theta)$ 
  - for a set of values of  $\theta$
  - Gaussian, adaptive, simplified Laplace or (full) Laplace approximation
- Step 3: approach  $p(x_i|y)$  and  $p(\theta_j|y)$ 
  - numerical integration over  $\theta$
- **IF  $p(y|...)$  is Gaussian, there are 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
- +++

# 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
- +++

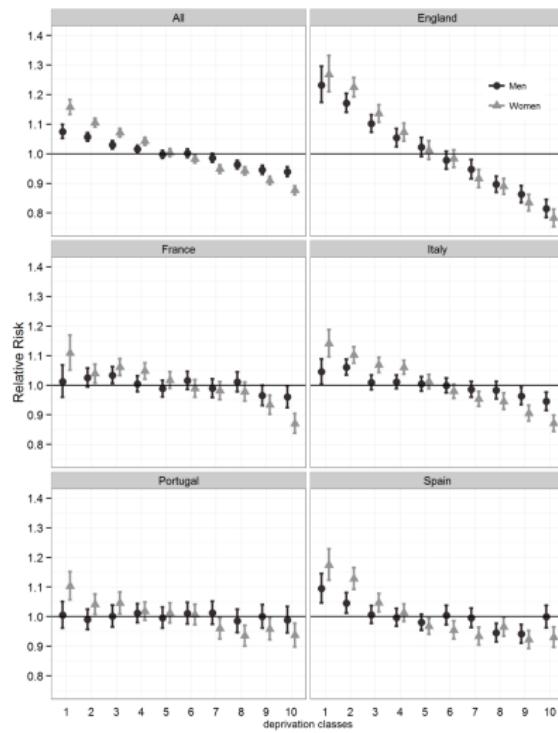
→ GLMM, GAM, GAMM, ... different names for a similar thing

# Some applications cited in H. Rue et al. (2017)

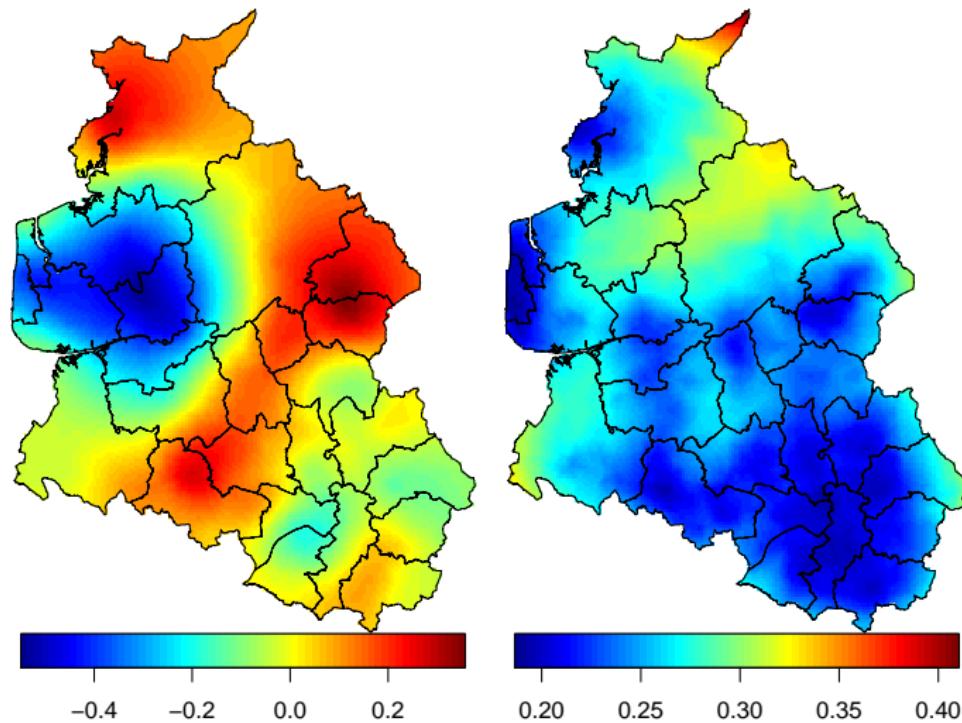
Recent examples of applications using the R-INLA package for statistical analysis include disease mapping (

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 ( Julian 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.

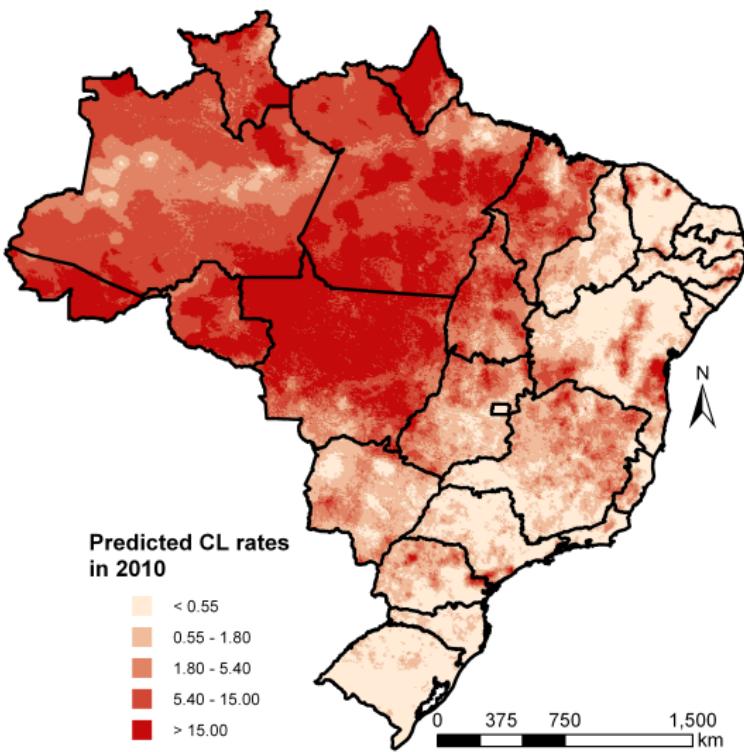
# Deprivation effect, Ribeiro et al. (2018)



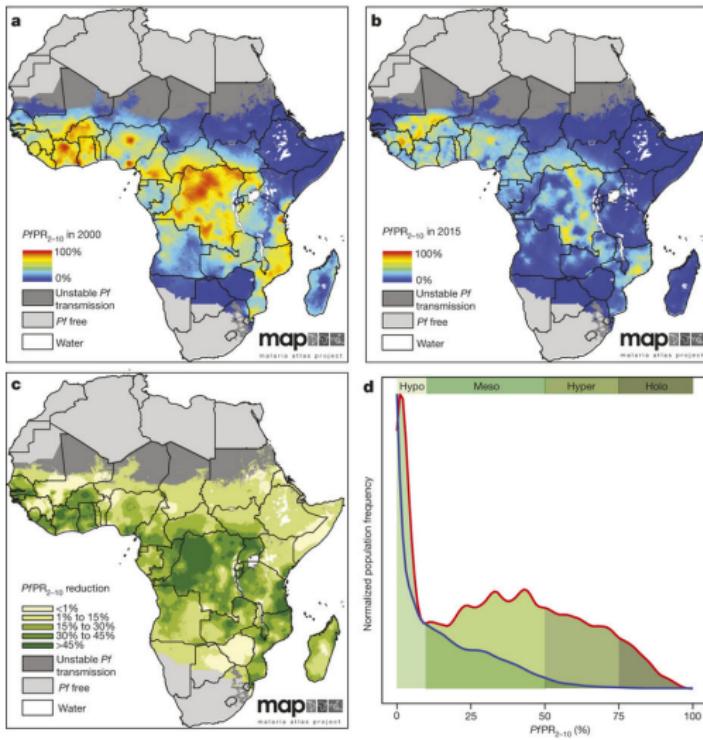
# Survival: frailty map



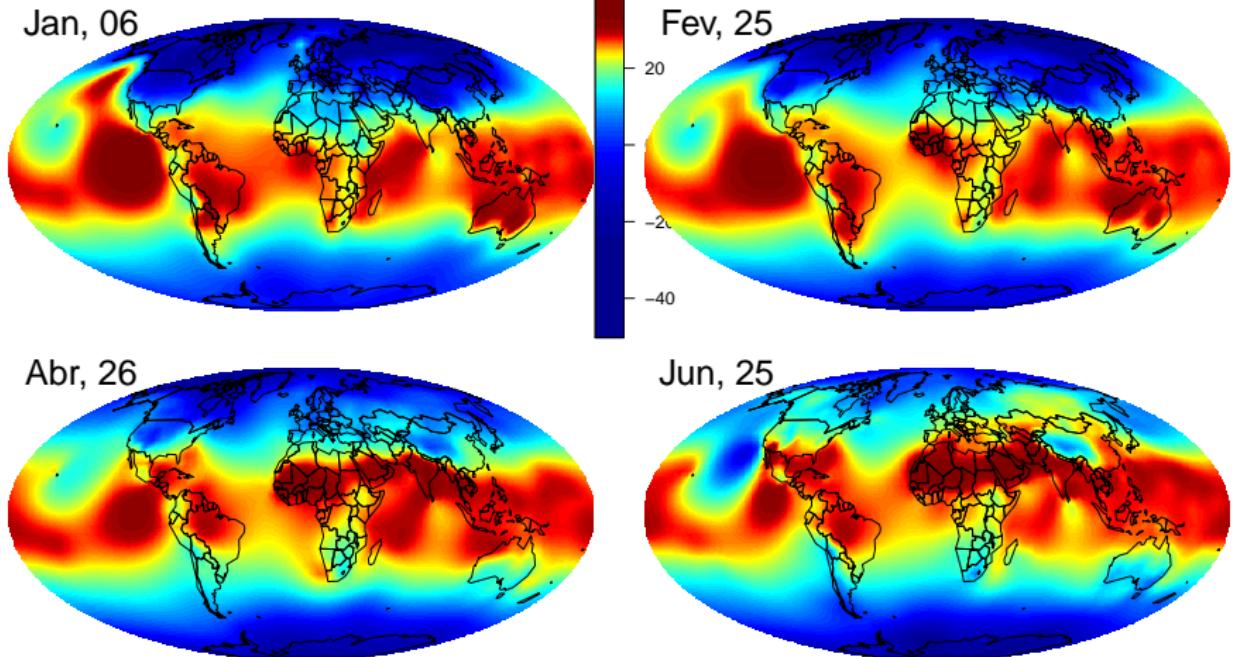
# Leishmaniasis in Brazil, Karagiannis-Voules et al. (2013)



# Malaria in Africa, Gething (2015)



# Non-separable space-time modeling in the globe



# Flexibility must come with responsibility

- PC-prior *Penalized Complexity* prior
  - Simpson et al. (2017)
  - Fuglstad et al. (2019)

# Sumário

- 1 The basic model idea
- 2 Extending the basic model
- 3 Hierarchical models
- 4 INLA
- 5 References

# References

- Fuglstad, Geir-Arne, Ingeborg Gullikstad Hem, Alexander Knight, Håvard Rue, and Andrea Riebler. 2019. "Intuitive Principle-Based Priors for Attributing Variance in Additive Model Structures," February, Submitted.
- 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. Guimaraes, 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. Pernet, 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., 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.