...you might like to give a talk about how priors are useful for modelling spatial data but we certainly would not hold you to that

Håvard Rue
King Abdullah University of Science and Technology
Saudi Arabia

December 4, 2017

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Plan

Priors

- Background on priors
- Penalised complexity priors

Examples

- The easier ones
- Area models (more)
- Gaussian fields (less)

Joint and ongoing work with many, including...

Daniel Simpson

Haakon Bakka

Anna Sterrantino

Andrea Riebler

Geir-A Fuglstad

Finn Lindgren

Massimo Ventrucci

Sigrunn Sørbye

and others

Our background: R-INLA (www.r-inla.org)

- INLA do Bayesian inference on Latent Gaussian models
- Accurate, fast, scale well wrt size, great spatial models support, quite general with an easy R-interface (www.r-inla.org).
- Build models adding model component

$$\eta = \boldsymbol{X}\beta + f_1(\ldots; \boldsymbol{\theta}_1) + f_2(\ldots; \boldsymbol{\theta}_2) + \cdots$$

  for Gaussians $\{f_i(\cdot)\}$ conditioned on some hyperparameters θ
- Likelihood(s) have hyper-parameters as well
- Of course, the model include **prior** specification for θ, which is the topic of this talk

Our background: R-INLA (www.r-inla.org)

- INLA do Bayesian inference on Latent Gaussian models
- Accurate, fast, scale well wrt size, great spatial models support, quite general with an easy R-interface (www.r-inla.org).
- Build models adding model component

$$\eta = \boldsymbol{X}\beta + f_1(\ldots; \boldsymbol{\theta}_1) + f_2(\ldots; \boldsymbol{\theta}_2) + \cdots$$

  for Gaussians $\{f_i(\cdot)\}$ conditioned on some hyperparameters θ
- Likelihood(s) have hyper-parameters as well
- Of course, the model include **prior** specification for θ, which is the topic of this talk

Our background: R-INLA (www.r-inla.org)

- INLA do Bayesian inference on Latent Gaussian models
- Accurate, fast, scale well wrt size, great spatial models support, quite general with an easy R-interface (www.r-inla.org).
- Build models adding model component

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(\ldots; \boldsymbol{\theta}_1) + f_2(\ldots; \boldsymbol{\theta}_2) + \cdots$$

  for Gaussians $\{f_i(\cdot)\}$ conditioned on some hyperparameters θ
- Likelihood(s) have hyper-parameters as well
- Of course, the model include **prior** specification for θ, which is the topic of this talk

Our background: R-INLA (www.r-inla.org)

- INLA do Bayesian inference on Latent Gaussian models
- Accurate, fast, scale well wrt size, great spatial models support, quite general with an easy R-interface (www.r-inla.org).
- Build models adding model component

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(\ldots; \boldsymbol{\theta}_1) + f_2(\ldots; \boldsymbol{\theta}_2) + \cdots$$

  for Gaussians $\{f_i(\cdot)\}$ conditioned on some hyperparameters θ
- Likelihood(s) have hyper-parameters as well
- Of course, the model include **prior** specification for θ, which is the topic of this talk

Our background: R-INLA (www.r-inla.org)

- INLA do Bayesian inference on Latent Gaussian models
- Accurate, fast, scale well wrt size, great spatial models support, quite general with an easy R-interface (www.r-inla.org).
- Build models adding model component

$$\eta = \boldsymbol{X}\boldsymbol{\beta} + f_1(...;\boldsymbol{\theta}_1) + f_2(...;\boldsymbol{\theta}_2) + \cdots$$

  for Gaussians $\{f_i(\cdot)\}$ conditioned on some hyperparameters θ
- Likelihood(s) have hyper-parameters as well
- Of course, the model include **prior** specification for θ, which is the topic of this talk

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

What do **we** know about priors for θ in this framework?

I cannot say what **you** know, here is my story...

- Not much. And I am not proud of it!
- I knew reference priors, which I, except in simple cases, cannot compute, and I do not want to use. Conjugate priors does not apply here, and is more "math, not priors".
- I could dig up similar studies/models/examples, and copy and refer to their prior choice. (Risk averse)
- I ran into problems when a student presented his/her hierarchical model and ask about advice for how to set priors for the f.ex 5-dimensional hyperparameter θ; I did not believed my own advises.
- I do not think that I am *that* unique

Why is the additive model situation different?

Classical:

- I want to estimate the precision from data $\boldsymbol{y}$, without any context
- In this case I just want to get it right!

Additive model:

- From data $\boldsymbol{y}$ I add an additional iid random effect

```
formula = y ~ ... + f(idx, model="iid")
```

  with the "hope" it is not there.
- In this case I have a preference for "no random effect" doing inference

Why is the additive model situation different?

Classical:

- I want to estimate the precision from data $\boldsymbol{y}$, without any context
- In this case I just want to get it right!

Additive model:

- From data $\boldsymbol{y}$ I add an additional iid random effect

```
formula = y ~ ... + f(idx, model="iid")
```

  with the "hope" it is not there.
- In this case I have a preference for "no random effect" doing inference

Why is the additive model situation different?

Classical:

- I want to estimate the precision from data $\boldsymbol{y}$, without any context
- In this case I just want to get it right!

Additive model:

- From data $\boldsymbol{y}$ I add an additional iid random effect

```
formula = y ~ ... + f(idx, model="iid")
```

  with the "hope" it is not there.
- In this case I have a preference for "no random effect" doing inference

Why is the additive model situation different?

Classical:

- I want to estimate the precision from data $\boldsymbol{y}$, without any context
- In this case I just want to get it right!

Additive model:

- From data $\boldsymbol{y}$ I add an additional iid random effect

```
formula = y ~ ... + f(idx, model="iid")
```

  with the "hope" it is not there.
- In this case I have a preference for "no random effect" doing inference

How to proceed from here?

- How to think about priors in hierarchical models?
- Is it possible to *understand/have good intuition about* them?

How to proceed from here?

- How to think about priors in hierarchical models?
- Is it possible to *understand/have good intuition about* them?

Parameters!

- I have an issue with parameters.
- σ, σ^2, τ, ρ, p, ...
- I want to understand their *impact* on something I understand, not their *numerical values*!
- Invariance

Parameters!

- I have an issue with parameters.
- σ, σ^2, τ, ρ, p, ...
- I want to understand their *impact* on something I understand, not their *numerical values*!
- Invariance

Parameters!

- I have an issue with parameters.
- σ, σ^2, τ, ρ, p, ...
- I want to understand their *impact* on something I understand, not their *numerical values*!
- Invariance

Parameters!

- I have an issue with parameters.
- σ, σ^2, τ, ρ, p, ...
- I want to understand their *impact* on something I understand, not their *numerical values*!
- Invariance

KISS (Keep it simple, stupid!)

- ...most systems work best if they are kept simple rather than made complicated
- ...there is no value in a solution being "clever" but in one being easily understandable

KISS (Keep it simple, stupid!)

- ...most systems work best if they are kept simple rather than made complicated
- ...there is no value in a solution being "clever" but in one being easily understandable

KISS (Keep it simple, stupid!)

- ...most systems work best if they are kept simple rather than made complicated
- ...there is no value in a solution being “clever” but in one being easily understandable

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the “prior”-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** “optimal” or “unique” in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

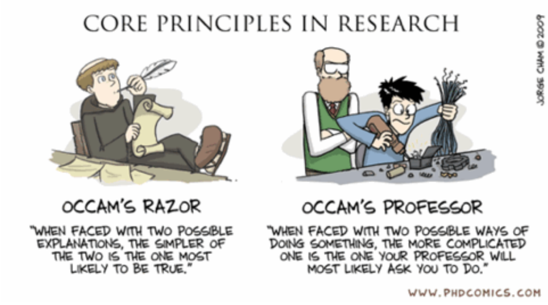
Our take on the "prior"-problem

which is

- a principled and practical approach to constructing priors
- KISS-friendly
- a unified way to think about priors
- useful
- is widely applicable
- is transparent
- invariant for reparameterisations
- something I can understand
- better than not knowing what to do

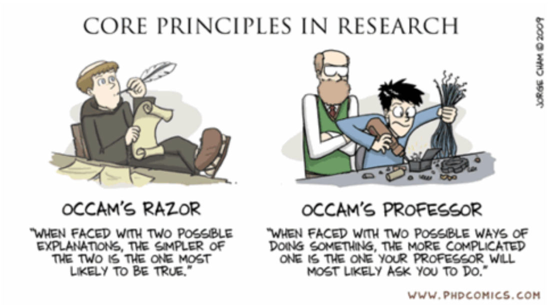
It is **not** "optimal" or "unique" in any sense. If you prefer something else, please do...

Principle I: Occam's razor



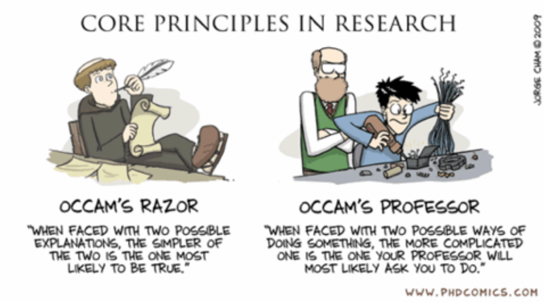
- Prefer simplicity over complexity. *Simplicity* defines the **base model**
- $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$, base model $\tau = \infty$
- Student-t, base model Gaussian
- Spline model, base model linear/constant effect
- AR(1), base model $\rho = 0$ or $\rho = 1^{-}$

Principle I: Occam's razor



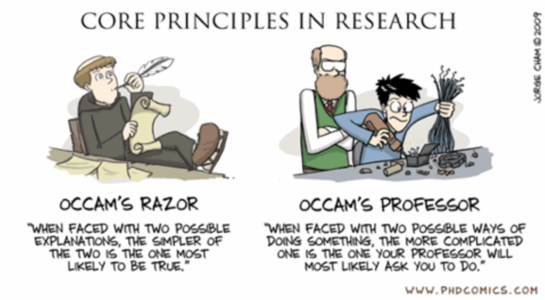
- Prefer simplicity over complexity. *Simplicity* defines the **base model**
- $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$, base model $\tau = \infty$
- Student-t, base model Gaussian
- Spline model, base model linear/constant effect
- AR(1), base model $\rho = 0$ or $\rho = 1^{-}$

Principle I: Occam's razor



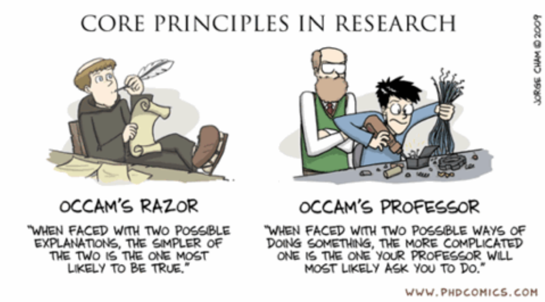
- Prefer simplicity over complexity. *Simplicity* defines the **base model**
- $x \sim \mathcal{N}(\mathbf{0}, \tau \mathbf{I})$, base model $\tau = \infty$
- Student-t, base model Gaussian
- Spline model, base model linear/constant effect
- AR(1), base model $\rho = 0$ or $\rho = 1^-$

Principle I: Occam's razor



- Prefer simplicity over complexity. *Simplicity* defines the **base model**
- $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$, base model $\tau = \infty$
- Student-t, base model Gaussian
- Spline model, base model linear/constant effect
- AR(1), base model $\rho = 0$ or $\rho = 1^-$

Principle I: Occam's razor



- Prefer simplicity over complexity. *Simplicity* defines the **base model**
- $x \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{I})$, base model $\tau = \infty$
- Student-t, base model Gaussian
- Spline model, base model linear/constant effect
- AR(1), base model $\rho = 0$ or $\rho = 1^-$

Principle I: Occam's razor

Consider the more complex model

$$\pi(x|\xi), \qquad \xi \geq 0$$

with base model $\pi(x|\xi = 0)$.

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause **overfitting/force complexity** if, loosely speaking,

$$\pi_\xi(\xi = 0) = 0$$

Principle I: Occam's razor

Consider the more complex model

$$\pi(x|\xi), \qquad \xi \geq 0$$

with base model $\pi(x|\xi = 0)$.

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause **overfitting/force complexity** if, loosely speaking,

$$\pi_\xi(\xi = 0) = 0$$

Principle I: Occam's razor

Consider the more complex model

$$\pi(x|\xi), \qquad \xi \geq 0$$

with base model $\pi(x|\xi = 0)$.

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause **overfitting/force complexity** if, loosely speaking,

$$\pi_\xi(\xi = 0) = 0$$

Principle I: Occam's razor

Consider the more complex model

$$\pi(x|\xi), \qquad \xi \geq 0$$

with base model $\pi(x|\xi = 0)$.

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause **overfitting/force complexity** if, loosely speaking,

$$\pi_\xi(\xi = 0) = 0$$

Principle I: Occam's razor

Consider the more complex model

$$\pi(x|\xi), \qquad \xi \geq 0$$

with base model $\pi(x|\xi = 0)$.

- The prior for $\xi \geq 0$ should penalise the complexity introduced by ξ
- The prior should be decaying with increasing measure by the complexity (the mode should be at the base model)

A prior will cause **overfitting/force complexity** if, loosely speaking,

$$\pi_\xi(\xi = 0) = 0$$

Principle II: Measure of complexity

Use Kullback-Leibler discrepancy to measure the increased complexity introduced by $\xi > 0$,

$$\mathrm{KLD}(f \| g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

for flexible model f and base model g.

Gives a measure of the information lost when the base model is used to approximate the more flexible models

Principle II: Measure of complexity

Use Kullback-Leibler discrepancy to measure the increased complexity introduced by $\xi > 0$,

$$\mathrm{KLD}(f \| g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

for flexible model f and base model g.

Gives a measure of the information lost when the base model is used to approximate the more flexible models

Principle III: Constant rate penalisation

Define

$$d(\xi) = \sqrt{2\,\text{KLD}(\xi)}$$

as the (uni-directional) “distance” from flexible-model to the base model. Need the square-root to get the scale right.

Constant rate penalisation:

$$\pi(d) = \lambda \exp\left(-\lambda d\right), \qquad \lambda > 0$$

with mode at $d = 0$

Invariance: OK

Principle III: Constant rate penalisation

Define

$$d(\xi) = \sqrt{2\,\mathrm{KLD}(\xi)}$$

as the (uni-directional) "distance" from flexible-model to the base model. Need the square-root to get the scale right.

Constant rate penalisation:

$$\pi(d) = \lambda \exp(-\lambda d), \qquad \lambda > 0$$

with mode at $d = 0$

Invariance: OK

Principle III: Constant rate penalisation

Define

$$d(\xi) = \sqrt{2\,\mathrm{KLD}(\xi)}$$

as the (uni-directional) “distance” from flexible-model to the base model. Need the square-root to get the scale right.

Constant rate penalisation:

$$\pi(d) = \lambda \exp\left(-\lambda d\right), \qquad \lambda > 0$$

with mode at $d = 0$

Invariance: OK

Principle IV: User-defined scaling

The rate λ is determined from knowledge of the *scale* or some interpretable property or impact, $Q(\xi)$ of ξ:

$$\Pr(Q(\xi) > U) = \alpha$$

- Problem dependent: must be!!!
- Can make the prior more informative or weakly informative this way

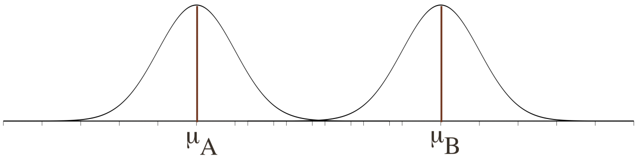
Principle IV: User-defined scaling

The rate λ is determined from knowledge of the *scale* or some interpretable property or impact, $Q(\xi)$ of ξ:

$$\Pr(Q(\xi) > U) = \alpha$$

- Problem dependent: must be!!!
- Can make the prior more informative or weakly informative this way

Example



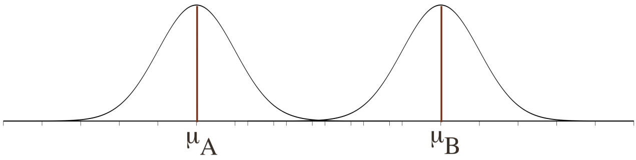
- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu, 1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example



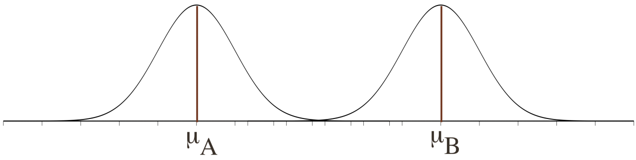
- Base model $\mathcal{N}(0, 1)$
- Flexible model $\mathcal{N}(\mu, 1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example



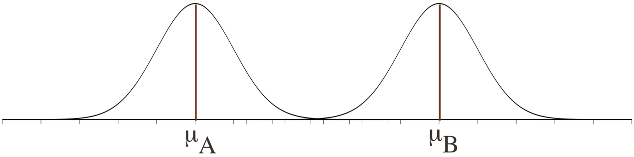
- Base model $\mathcal{N}(0, 1)$
- Flexible model $\mathcal{N}(\mu, 1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example



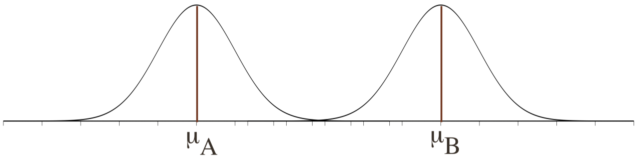
- Base model $\mathcal{N}(0,1)$
- Flexible model $\mathcal{N}(\mu,1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

Example



- Base model $\mathcal{N}(0, 1)$
- Flexible model $\mathcal{N}(\mu, 1)$, $\mu > 0$.
- KLD is $\mu^2/2$ and $d(\mu) = \mu$.
- PC prior:

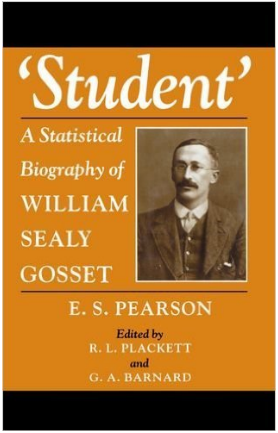
$$\pi(\mu) = \lambda \exp(-\lambda\mu)$$

- Can determine λ from a question like

$$\text{Prob}(\mu > u) = \alpha$$

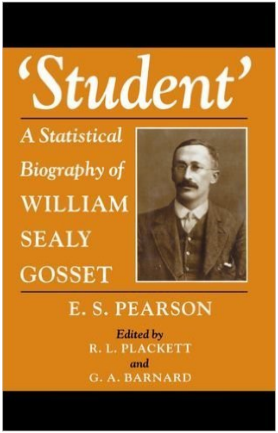
Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.



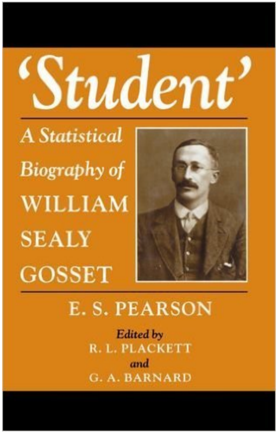
Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.



Example: Student-t with unit variance

- Degrees of freedom (dof) parameter $\nu > 2$.
- This is a difficult case: It is hard to intuitively construct any reasonable prior for ν at all.
- It is hard to even think of dof.



A useful but negative result

Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

A useful but negative result

Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

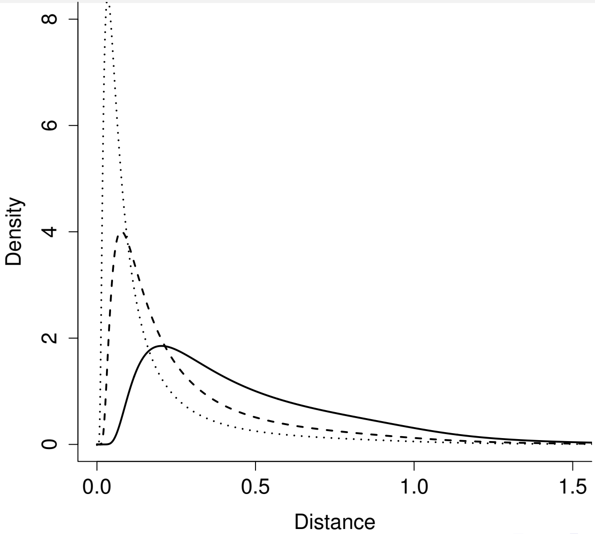
- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

A useful but negative result

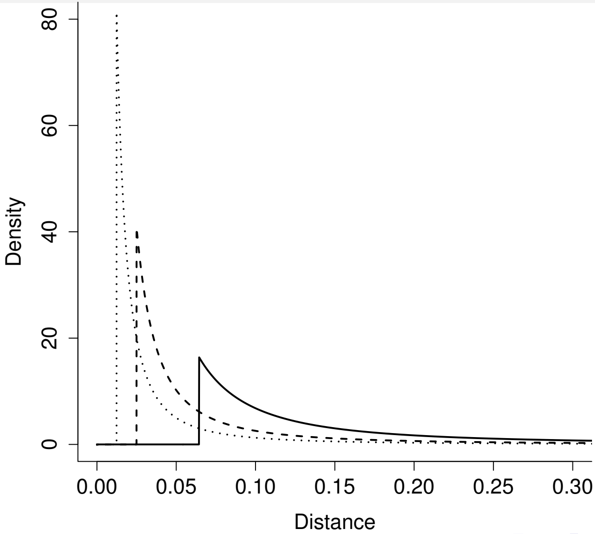
Result Let $\pi_\nu(\nu)$ be a prior for $\nu > 2$ where $E(\nu) < \infty$, then $\pi_d(0) = 0$ and the prior overfits

- Priors with finite expectation *defines* the flexible model to be *different* from the base model!!!
- Why? A finite expectation bounds the tail behaviour as $\nu \to \infty$

The exp-prior with mean 5, 10, 20, converted to a prior for the distance



The uniform prior with upper= 20, 50, 100, converted to a prior for the distance



The precision of a Gaussian

PC prior for the precision κ when $\kappa = \infty$ defines the base model

- "random effects"/iid-model
- The smoothing parameter in spline models
- etc...

Result Let $\pi_\kappa(\kappa)$ be a prior for $\kappa > 0$ where $E(\kappa) < \infty$, then $\pi_d(0) = 0$ and the prior overfits.

The precision of a Gaussian

PC prior for the precision κ when $\kappa = \infty$ defines the base model

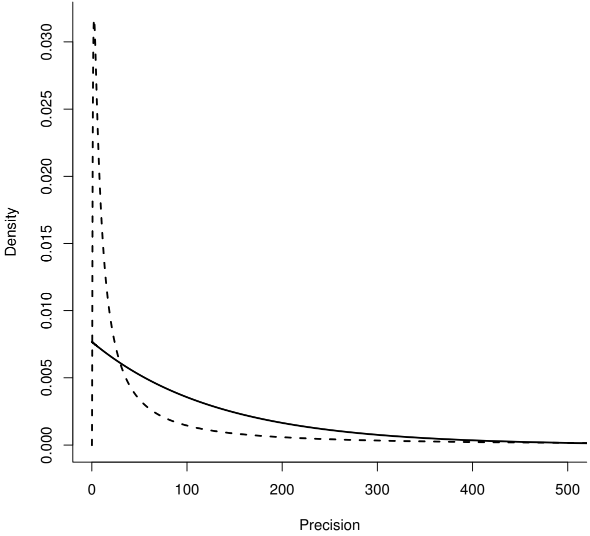
- "random effects"/iid-model
- The smoothing parameter in spline models
- etc...

Result Let $\pi_\kappa(\kappa)$ be a prior for $\kappa > 0$ where $E(\kappa) < \infty$, then $\pi_d(0) = 0$ and the prior overfits.

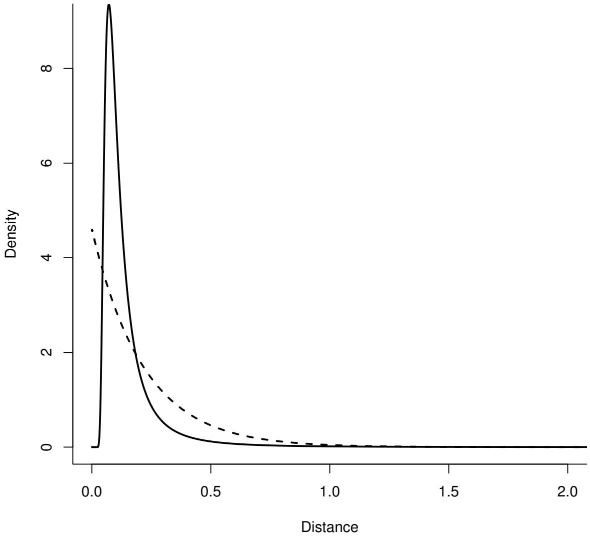
The precision case (II)

$$\pi(\sigma) = \lambda \exp(-\lambda\sigma)$$

Comparison with a similar Gamma-prior



Comparison with a similar Gamma-prior



Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are (almost) orthogonal in interpretation

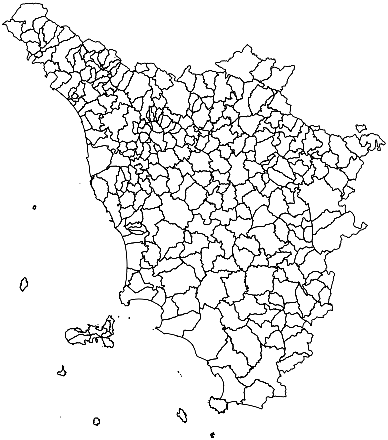
Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are (almost) orthogonal in interpretation

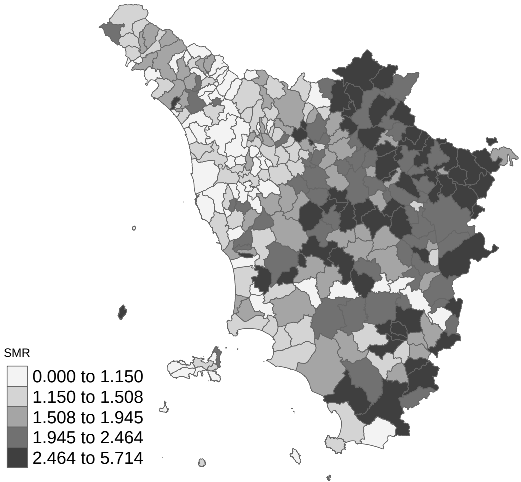
Student-t case revisited

- PC prior for the dof ν
- PC prior for precision κ
- This is OK as the parameters are (almost) orthogonal in interpretation

Area models

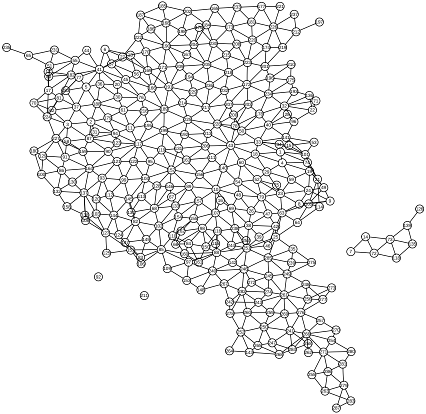


Area models


SMR
0.000 to 1.150
1.150 to 1.508
1.508 to 1.945
1.945 to 2.464
2.464 to 5.714

The "CAR"-model

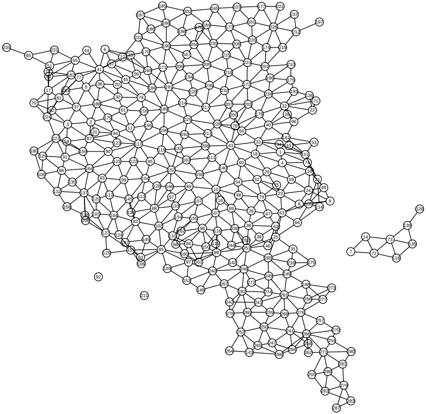
- Instrinsic GMRF model
- $E(x_i | \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a "sum-to-zero constraint"
- **The issue here, is the prior for κ**



Many call this "the CAR model", never really understood why. This is why it is named `model="besag"` in `R-INLA`

The "CAR"-model

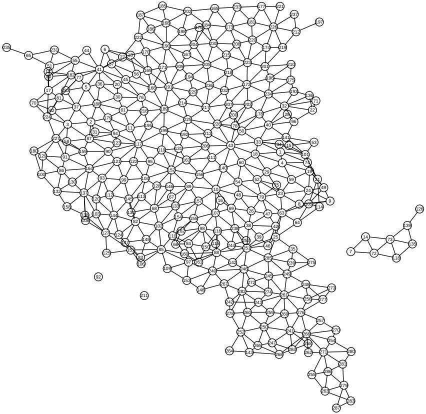
- Instrinsic GMRF model
- $E(x_i | \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a "sum-to-zero constraint"
- **The issue here, is the prior for κ**



Many call this "the CAR model", never really understood why. This is why it is named `model="besag"` in `R-INLA`

The "CAR"-model

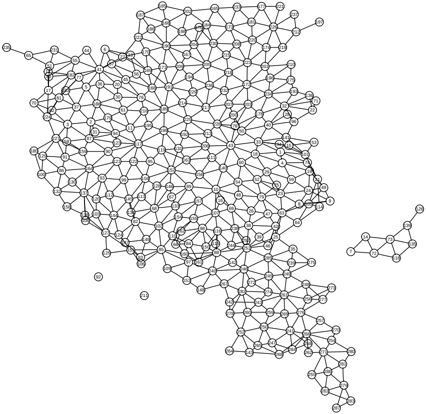
- Instrinsic GMRF model
- $\text{E}(x_i | \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a "sum-to-zero constraint"
- **The issue here, is the prior for κ**



Many call this "the CAR model", never really understood why. This is why it is named `model="besag"` in `R-INLA`

The “CAR”-model

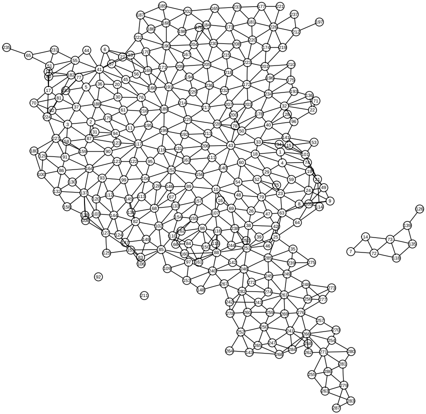
- Instrinsic GMRF model
- $\mathrm{E}(x_i | \ldots) = \overline{x_{\partial i}}$
- $\mathrm{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a “sum-to-zero constraint”
- The issue here, is the prior for κ



Many call this “the CAR model”, never really understood why. This is why it is named `model="besag"` in R-INLA

The “CAR”-model

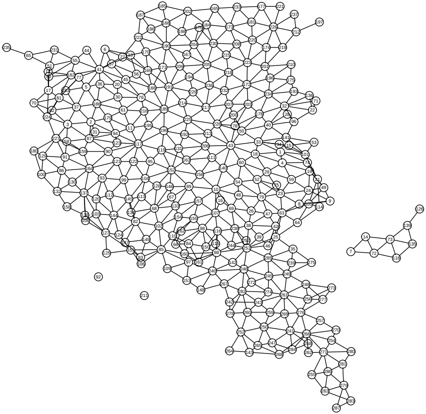
- Instrinsic GMRF model
- $\text{E}(x_i \mid \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i \mid \ldots) = 1/(\kappa|\partial_i|)$
- Usually a “sum-to-zero constraint”
- **The issue here, is the prior for κ**



Many call this “the CAR model”, never really understood why. This is why it is named `model="besag"` in `R-INLA`

The "CAR"-model

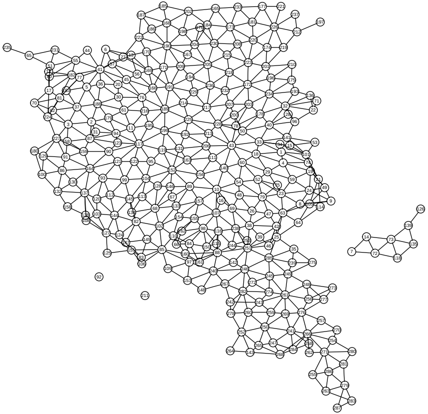
- Instrinsic GMRF model
- $\text{E}(x_i | \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a "sum-to-zero constraint"
- **The issue here, is the prior for κ**



Many call this "the CAR model", never really understood why. This is why it is named `model="besag"` in `R-INLA`

The "CAR"-model

- Instrinsic GMRF model
- $\text{E}(x_i | \ldots) = \overline{x_{\partial i}}$
- $\text{Var}(x_i | \ldots) = 1/(\kappa |\partial_i|)$
- Usually a "sum-to-zero constraint"
- **The issue here, is the prior for κ**

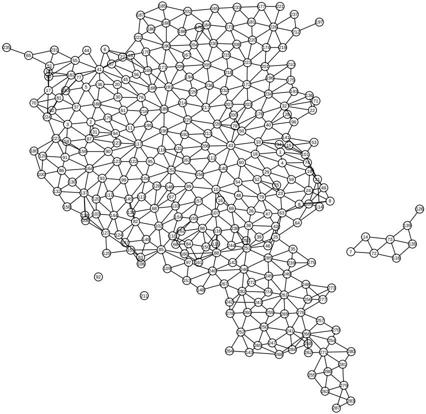


Many call this "the CAR model", never really understood why. This is why it is named `model="besag"` in `R-INLA`

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?

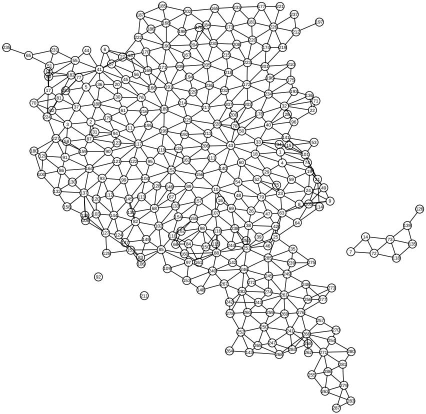


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?

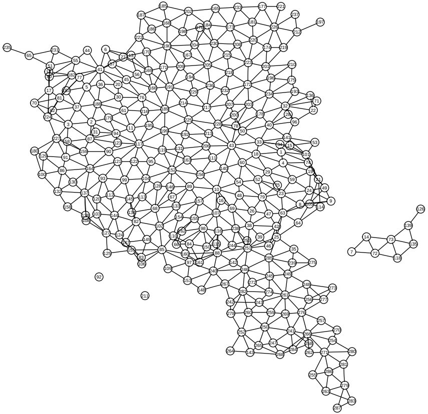


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with ‘sum-to-zero’ constraint?
- What to do with effect of the graph itself?

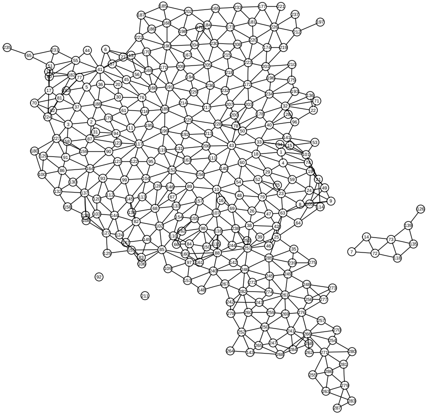


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?

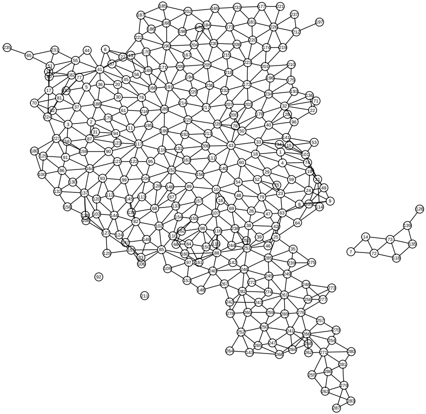


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?

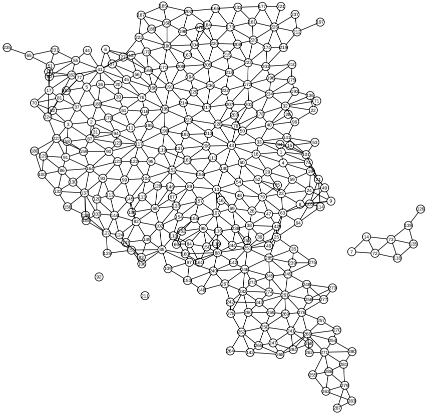


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?

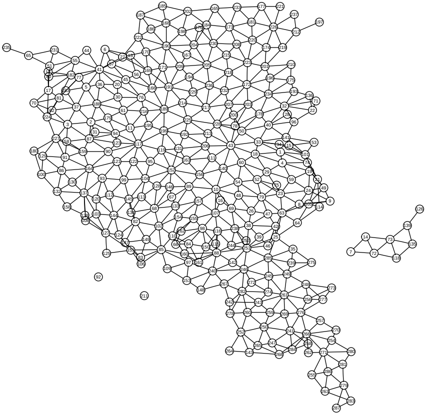


Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Prior for κ

There are several issues!

- What to do with singeltons?
- What to do with islands?
- What to do with 'sum-to-zero' constraint?
- What to do with effect of the graph itself?



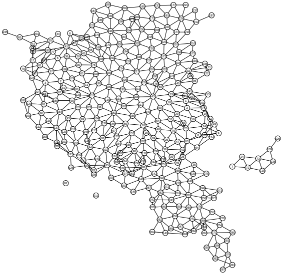
Without getting these issues right and internally coherent, there is no point of thinking of a prior for κ. Same issues, but more serious, with RW1/RW2 models.

Scaling (I)

- Assume a connected graph
- κ controls the deviation from the null-space
- The geometric mean of the marginal variances are

```
## largest area
> gmean(diag(INLA::inla.ginv(Q)))
[1] 0.4987539796
## island
> gmean(diag(INLA::inla.ginv(Q)))
[1] 0.3910586057
```

when $\kappa = 1$

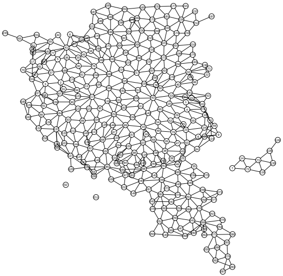


Scaling (I)

- Assume a connected graph
- κ controls the deviation from the null-space
- The geometric mean of the marginal variances are

```
## largest area
> gmean(diag(INLA::inla.ginv(Q)))
[1] 0.4987539796
## island
> gmean(diag(INLA::inla.ginv(Q)))
[1] 0.3910586057
```

when $\kappa = 1$

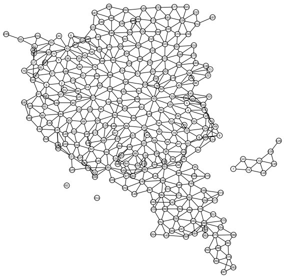


Scaling (I)

- Assume a connected graph
- κ controls the deviation from the null-space
- The geometric mean of the marginal variances are

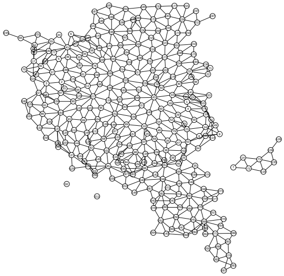
```
## largest area
> gmean(diag(INLA:::inla.ginv(Q)))
[1] 0.4987539796
## island
> gmean(diag(INLA:::inla.ginv(Q)))
[1] 0.3910586057
```

when $\kappa = 1$



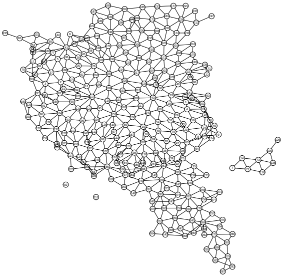
Scaling (II)

- Scale each connected component to have unit gmean
- Singletons given iid std normal distribution
- One sum-to-zero constraint for each subgraph > 1 (variants)
- Then κ is has clear interpretation as the *marginal* precision, controlling the deviance from the null-space



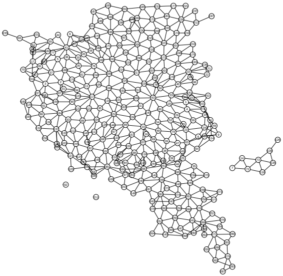
Scaling (II)

- Scale each connected component to have unit gmean
- Singletons given iid std normal distribution
- One sum-to-zero constraint for each subgraph > 1 (variants)
- Then κ is has clear interpretation as the *marginal* precision, controlling the deviance from the null-space



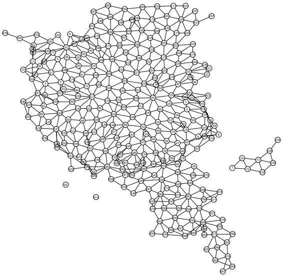
Scaling (II)

- Scale each connected component to have unit gmean
- Singletons given iid std normal distribution
- One sum-to-zero constraint for each subgraph > 1 (variants)
- Then κ is has clear interpretation as the *marginal* precision, controlling the deviance from the null-space



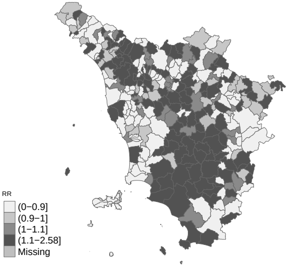
Scaling (II)

- Scale each connected component to have unit gmean
- Singletons given iid std normal distribution
- One sum-to-zero constraint for each subgraph > 1 (variants)
- Then κ is has clear interpretation as the *marginal* precision, controlling the deviance from the null-space



Scaling (III)

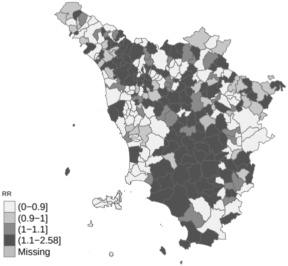
- PC-prior for κ
- Feasible max range $(-1, 1)$
- Gives PC-prior params: $(U, \alpha) = (1, 0.01)$




```
formula = obs35 ~ 1+f(id, model='besag',
                      scale.model=TRUE, graph='Toscana.graph',
                      hyper = list(prec = list(
                                   prior = "pc.prec",
                                   param = c(1, 0.01))))
res = inla(formula, family="poisson", E=exp35, data=counties@data)
```

Scaling (III)

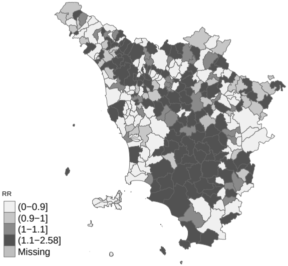
- PC-prior for κ
- Feasible max range $(-1, 1)$
- Gives PC-prior params: $(U, \alpha) = (1, 0.01)$




```
formula = obs35 ~ 1+f(id, model='besag',
                      scale.model=TRUE, graph='Toscana.graph',
                      hyper = list(prec = list(
                                   prior = "pc.prec",
                                   param = c(1, 0.01))))
res = inla(formula, family="poisson", E=exp35, data=counties@data)
```

Scaling (III)

- PC-prior for κ
- Feasible max range $(-1, 1)$
- Gives PC-prior params: $(U, \alpha) = (1, 0.01)$




```
formula = obs35 ~ 1+f(id, model='besag',
                      scale.model=TRUE, graph='Toscana.graph',
                      hyper = list(prec = list(
                                   prior = "pc.prec",
                                   param = c(1, 0.01))))
res = inla(formula, family="poisson", E=exp35, data=counties@data)
```

Towards the BYM-model

- Additional random effect: structured and unstructured
- Here, there is a lot of “confusion” in the literature
- We also need to distribute variance between the structured and unstructured part

Towards the BYM-model

- Additional random effect: structured and unstructured
- Here, there is a lot of "confusion" in the literature
- We also need to distribute variance between the structured and unstructured part

Towards the BYM-model

- Additional random effect: structured and unstructured
- Here, there is a lot of "confusion" in the literature
- We also need to distribute variance between the structured and unstructured part

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I + \phi R\right)$$

- Just the wrong way to doit!
  - Convex combination of precision matrices???
  - What is ϕ without scaling???
  - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I + \phi R\right)$$

- Just the wrong way to doit!
  - Convex combination of precision matrices???
  - What is ϕ without scaling???
  - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I+\phi R\right)$$

- Just the wrong way to doit!
  - Convex combination of precision matrices???
  - What is ϕ without scaling???
  - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I+\phi R\right)$$

- Just the wrong way to doit!
    - Convex combination of precision matrices???
    - What is ϕ without scaling???
    - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I+\phi R\right)$$

- Just the wrong way to doit!
  - Convex combination of precision matrices???
  - What is ϕ without scaling???
  - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I + \phi R\right)$$

- Just the wrong way to doit!
    - Convex combination of precision matrices???
    - What is ϕ without scaling???
    - No meaningful interpretation

The Leroux model

- Unstructured precision matrix I
- Structured precision matrix R
- Leroux model

$$\tau\left((1-\phi)I+\phi R\right)$$

- Just the wrong way to doit!
    - Convex combination of precision matrices???
    - What is ϕ without scaling???
    - No meaningful interpretation

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov(sum)} = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\mathrm{Cov}(\mathrm{sum}) = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov(sum)} = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov(sum)} = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov(sum)} = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov(sum)} = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

The BYM model: alternative paramerisation

- The basic idea is to have a convex combination of two limiting cases
  - unstructured
  - structured (scaled!)
- This means

$$\text{Cov}(\text{sum}) = \frac{1}{\tau}\left((1-\phi)I + \phi R^{-}\right)$$

  which is **not** the Leroux model
- We have priors
  - PC-prior for τ to shrink the sum to zero
  - PC-prior for ϕ to shrink to the simpler model $\phi = 0$, and is graph dependent

Computation of the prior for ϕ

- Depends on the graph or R^{-1}
- Computations need to make use of the sparse structure of R
- Let $z = x + y$, where x and y are indep normal
- Then z is the marginal from the joint distribution of (x, z)

$$x \sim \mathcal{N}(0, \ldots) \quad \text{and} \quad z|x \sim \mathcal{N}(x, \ldots)$$

  and (x, z) is Markov
- And we need to use that

$$(I + A^{-1})^{-1} = A(A + I)^{-1}$$

Computation of the prior for ϕ

- Depends on the graph or R^{-1}
- Computations need to make use of the sparse structure of R
- Let $z = x + y$, where x and y are indep normal
- Then z is the marginal from the joint distribution of (x, z)

$$x \sim \mathcal{N}(0, \ldots) \quad \text{and} \quad z|x \sim \mathcal{N}(x, \ldots)$$

and (x, z) is Markov

- And we need to use that

$$(I + A^{-1})^{-1} = A(A + I)^{-1}$$

Computation of the prior for ϕ

- Depends on the graph or R^{-1}
- Computations need to make use of the sparse structure of R
- Let $z = x + y$, where x and y are indep normal
- Then z is the marginal from the joint distribution of (x, z)

$$x \sim \mathcal{N}(0, \ldots) \quad \text{and} \quad z|x \sim \mathcal{N}(x, \ldots)$$

  and (x, z) is Markov
- And we need to use that

$$(I + A^{-1})^{-1} = A(A + I)^{-1}$$

Computation of the prior for ϕ

- Depends on the graph or R^{-1}
- Computations need to make use of the sparse structure of R
- Let $z = x + y$, where x and y are indep normal
- Then z is the marginal from the joint distribution of (x, z)

$$x \sim \mathcal{N}(0, \ldots) \quad \text{and} \quad z|x \sim \mathcal{N}(x, \ldots)$$

  and (x, z) is Markov
- And we need to use that

$$(I + A^{-1})^{-1} = A(A + I)^{-1}$$

Computation of the prior for ϕ

- Depends on the graph or R^{-1}
- Computations need to make use of the sparse structure of R
- Let $z = x + y$, where x and y are indep normal
- Then z is the marginal from the joint distribution of (x, z)

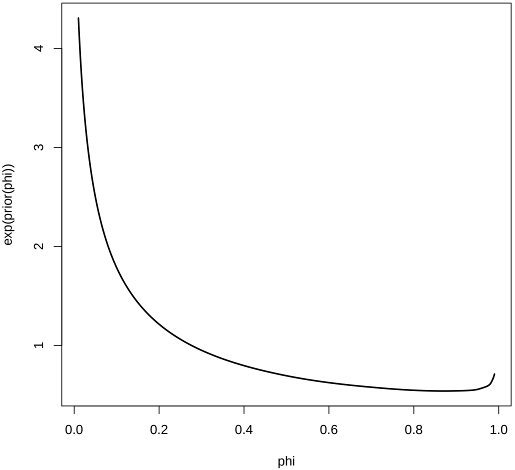
$$x \sim \mathcal{N}(0, \dots) \quad \text{and} \quad z|x \sim \mathcal{N}(x, \dots)$$

  and (x, z) is Markov
- And we need to use that

$$(I + A^{-1})^{-1} = A(A + I)^{-1}$$

Prior for ϕ

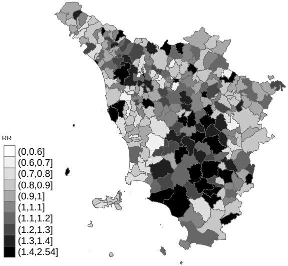
```
Q = INLA:::inla.pc.bym.Q("Toscana.graph")
prior = INLA:::inla.pc.bym.phi(Q,
        u = 0.5, alpha = 0.5)
phi = seq(0.01, 0.99, len=1000)
plot(phi, exp(prior(phi)), lwd=2, type="l")
```



Application

Proper Poisson quantile regression...

```
formula = obs35 ~ 1+f(id,
                model='bym',
                scale.model=TRUE,
                graph='Toscana.graph')
res = inla(formula, family="poisson",
          E=exp35, data=counties@data,
          control.family = list(
               control.link = list(
                    model = "quantile",
                    quantile = 0.9)))
```



Gaussian fields

- Gaussian field in $\mathbb{R}^d$ (d$\leq$ 3) with a Matérn covariance function with fixed smoothness ν.
- PC-prior for range r and variance σ^2, with base model $\sigma^2 = 0$ and $r = \infty$ (a constant).
- Joint PC-prior is (dim$=$ 2)

$$(1/r, \sigma) \sim \text{Exp}(\lambda_r) \times \text{Exp}(\lambda_\sigma),$$

with $\text{E}(r) = \infty$ to prevent overfitting.

Gaussian fields

- Gaussian field in $\mathbb{R}^d$ (d$\leq$ 3) with a Matérn covariance function with fixed smoothness ν.
- PC-prior for range r and variance σ^2, with base model $\sigma^2 = 0$ and $r = \infty$ (a constant).
- Joint PC-prior is (dim$=$ 2)

$$(1/r, \sigma) \sim \text{Exp}(\lambda_r) \times \text{Exp}(\lambda_\sigma),$$

with $\text{E}(r) = \infty$ to prevent overfitting.

Gaussian fields

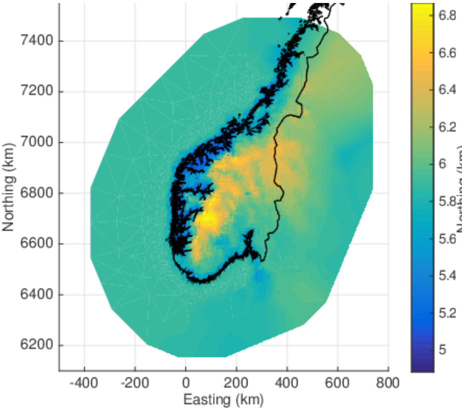
- Gaussian field in $\mathbb{R}^d$ (d$\leq$ 3) with a Matérn covariance function with fixed smoothness ν.
- PC-prior for range r and variance σ^2, with base model $\sigma^2 = 0$ and $r = \infty$ (a constant).
- Joint PC-prior is (dim$=$ 2)

$$(1/r, \sigma) \sim \text{Exp}(\lambda_r) \times \text{Exp}(\lambda_\sigma),$$

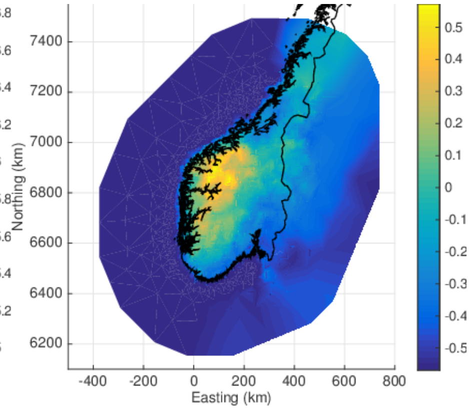
with $\text{E}(r) = \infty$ to prevent overfitting.

Non-stationary Gaussian fields

$$\pi(\theta) \;=\; \underbrace{\pi(\theta_{\text{stationary}})}_{\text{PC-prior for range \& stdev}} \;\times\; \underbrace{\pi(\theta_{\text{non-stationary}} \mid \theta_{\text{stationary}})}_{\text{shrinkage towards stationarity}}$$



Log(range)



Log(stdev)

Non-separable space-time model

Based on Finn's ideas

$$
\begin{aligned}
(\gamma_t \frac{\partial}{\partial t} - \Delta)^{\alpha_t} z(s,t) &= \gamma_s^{-1/2} \mathcal{E}(s,t) \\
(1 - \gamma_{\mathcal{E}} \Delta)^{\alpha_{\mathcal{E}}/2} \mathcal{E}(s, \delta t) &= \mathcal{W}_{\mathcal{E}}(s, \delta t)
\end{aligned}
$$

written up in the forthcoming PhD-thesis of Elias Krainski.

We need to understand the parameters in this model, which we **can** map into

- marginal variance
- spatial range
- temporal range

Non-separable space-time model

Based on Finn's ideas

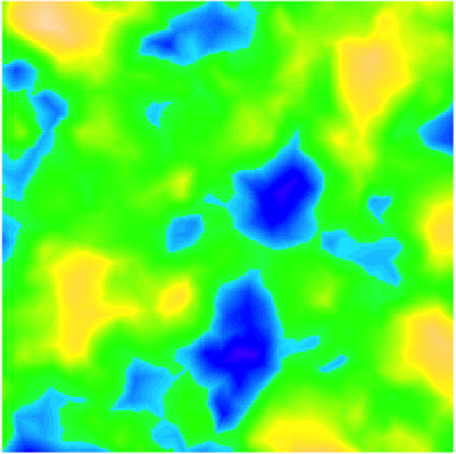
$$
\begin{aligned}
(\gamma_t \frac{\partial}{\partial t} - \Delta)^{\alpha_t} z(s,t) &= \gamma_s^{-1/2} \mathcal{E}(s,t) \\
(1 - \gamma_{\mathcal{E}} \Delta)^{\alpha_{\mathcal{E}}/2} \mathcal{E}(s, \delta t) &= \mathcal{W}_{\mathcal{E}}(s, \delta t)
\end{aligned}
$$

written up in the forthcoming PhD-thesis of Elias Krainski.

We need to understand the parameters in this model, which we **can** map into

- marginal variance
- spatial range
- temporal range

New nonsep space-time model



Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

Discussion: Priors

- Not easy
- Makes a difference
- Need to '*calibrate*' priors based on intuitive model properties
- PC-priors is a principled constructive approach to construct priors, and seems very promising
- Easy and natural interpretation, as a well defined shrinkage to a base-model: KISS!
- Still work in progress

References

- D. P. Simpson, H. Rue, A. Riebler, T. G. Martins and S. H. Sørbye (2014 & 2015 & 2016 & 2017) *Penalising model component complexity: A principled, practical approach to constructing priors (with discussion)* Statistical Science
- A. Riebler and S. H. Sørbye and D. Simpson and H. Rue, *An intuitive Bayesian spatial model for disease mapping that accounts for scaling*, 2016, Statistical Methods in Medical Research
- M. Ventrucci and H. Rue, *Penalised complexity priors for degrees of freedom in Bayesian P-splines*, 2016, Statistical Modelling
- S. H. Sørbye and H. Rue. *Penalised complexity priors for stationary autoregressive processes.* arXiv preprint arXiv:1608.08941, to appear in *Journal of Time Series Analysis*, 2017.
- G. A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. *Constructing priors that penalize the complexity of Gaussian random fields.* arXiv:1503.00256, 2016, in revision.
- and others...