Non-stationary Gaussian models with physical barriers

Haakon Bakka;
in collaboration with Jarno Vanhatalo, Janine Illian, Daniel Simpson and Håvard Rue

King Abdullah University of Science and Technology

Nov 8, 2018

Avignon, France. Ca 40 min.

Summary

The coastline problem

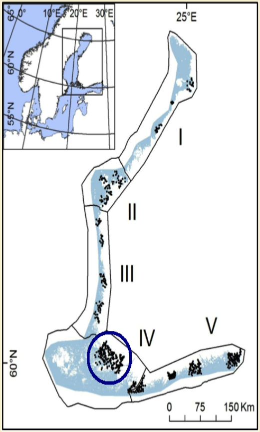
- Data near coastline
- Boundary effects
- Stability and practical usefulness
- Some solutions and why they do not work

Our solution

- The Barrier model (component)
- Practical presentation
- Theoretical presentation

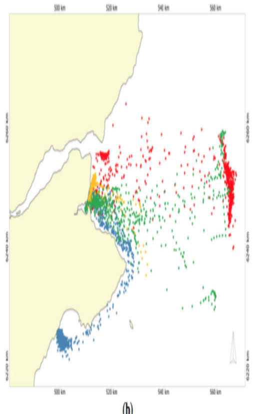
Our motivation to work on this

Jarno Vanhatalo



Esther Jones

Grey seal with telemetry tag (photo credit: SMRU).



Oceans and Fisheries Canada

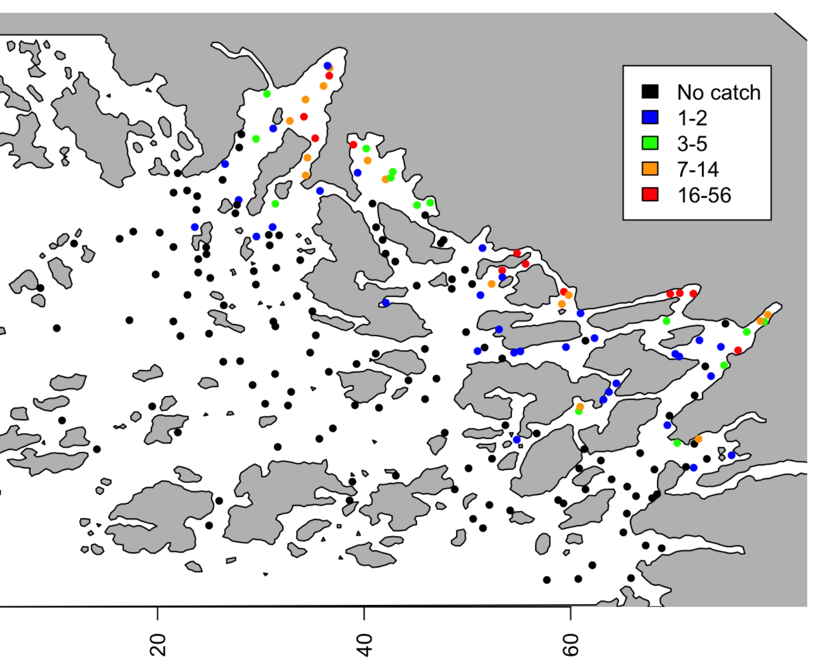
Statistical Ecology, Species Distribution Modeling

Construct spatial abundance maps
$\rightarrow$ Predictive surface

Habitat preference
$\rightarrow$ Linear or spline (nonparametric) covariate effects

Spatial structure not explained by covariates
$\rightarrow$ Spatial random effect

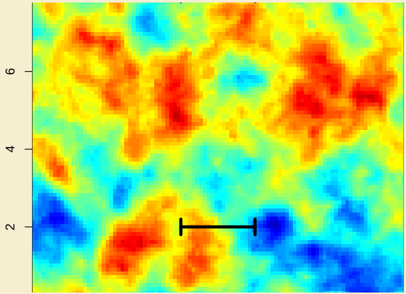
Main example



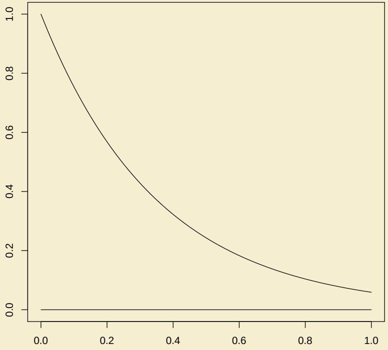
Spatial Modeling With Matérn Fields

Our starting point

Simulated field



Correlation by distance



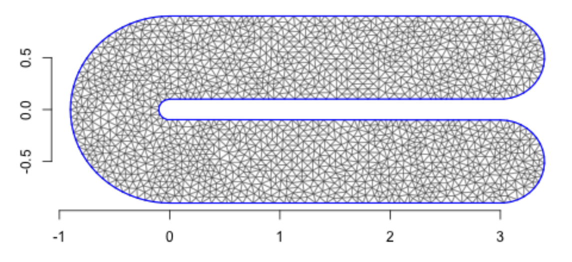
$u(s)$ hyperparameters σ, r (range)

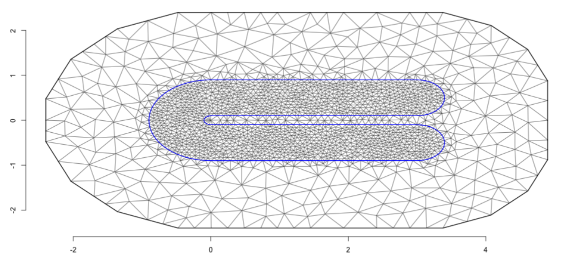
$$\vec{u} \sim \mathcal{N}(0, Q^{-1})$$

Hierarchical model: random effect is part of a GAM/GLMM, or an LGM.

Moving the boundary away

Any boundary condition

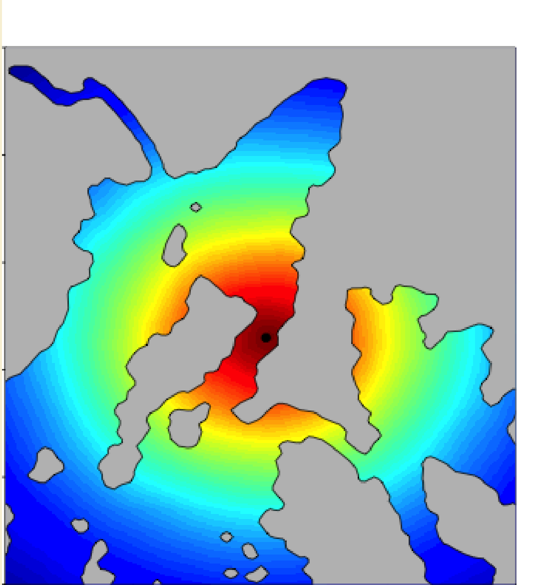




How to deal with the coastline?

But can't we just...? (1)

- Create a mesh/grid over the entire area, including land, and put a spatial model on this?



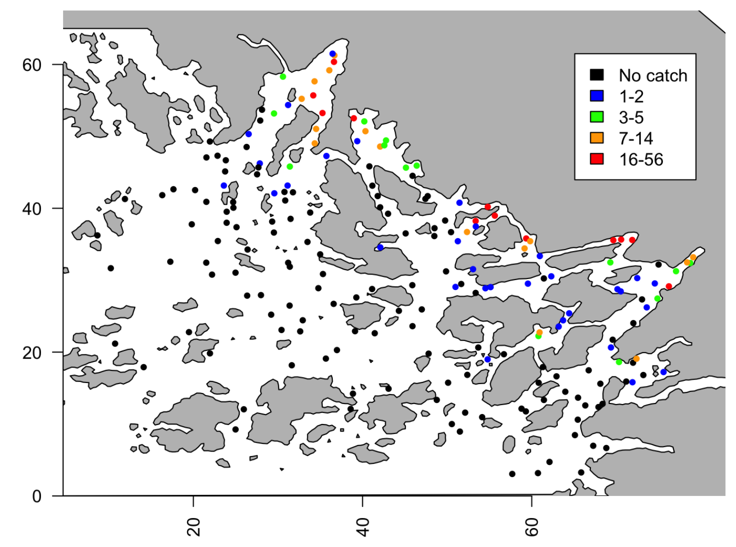
But can't we just...? (2)

- Set a known value at the boundary, like 0?

But can't we just...? (2)

- Set a known value at the boundary, like 0?

This dominates the data. Is there "constant" fish along the coast?



But can't we just...? (3)

- Model in water and use Neumann boundary conditions?

But can't we just...? (3)

- Model in water and use Neumann boundary conditions?

No.

But can't we just...? (4)

- Model the unknown values along the boundary with splines?

But can't we just...? (4)

- Model the unknown values along the boundary with splines?

200 datapoints and 200 splines!
Change the spline model as you deform boundary, how?

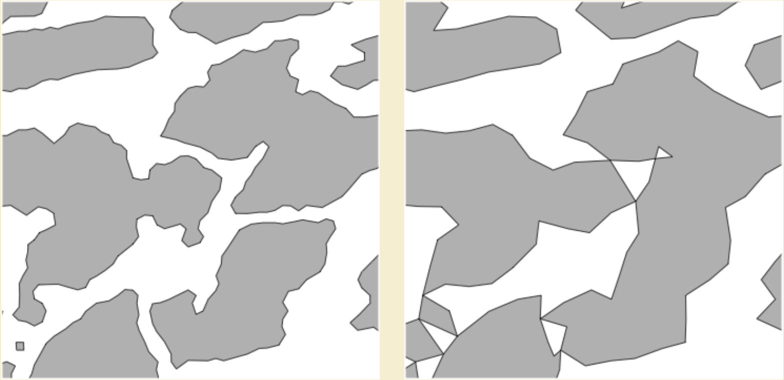
But can't we just...? (5)

- Redefine distance to be "distance around land" and construct covariance here? (e.g. as several)

But can't we just...? (5)

- Redefine distance to be "distance around land" and construct covariance here? (e.g. as several)

Coastlines not always well defined. This approach is not stable.



What more do you ask?

(of a good solution)

Property 2 Computational cost $\sim$ simple model

Property 3 Easy to use

Property 5 No new unrealistic assumptions

What more do you ask?

(of a good solution)

Property 2 Computational cost $\sim$ simple model

Property 3 Easy to use

Property 5 No new unrealistic assumptions

Something I cannot answer well:

- When can you ignore the coastline problem?

What is “easy to use”?

When using hierarchical models / GAMs, e.g. in R-INLA

Observation likelihood $\pi(y_i|\eta_i)$, e.g. poisson.
Linear predictor

$$\eta_i = \beta_0 + \beta_1 x_{1,i} + ... + u(s_i)$$

```
>> formula1 = y ~ 1+x1 + ... + f(s, model = spatial.model)

>> formula2 = y ~ 1+x1 + ... + f(s, model = new.model)
```

Replace the spatial random effect with a new random effect, without adding *any* other complexity.

The Barrier model (component)

Practical presentation

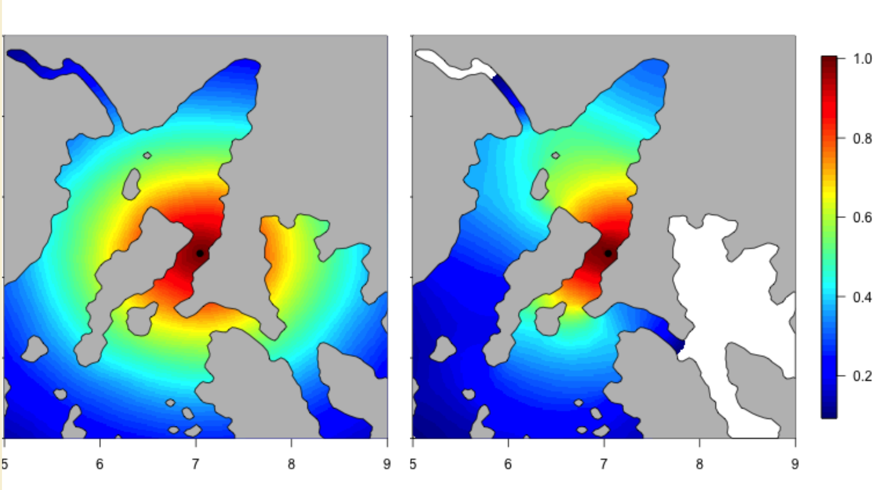
- Needed for applications, experts, communication
- Priors and conditional priors

Theoretical presentation

- Needed for deep understanding, implementation, generalisations
- Mathematical ideas

Practical presentation

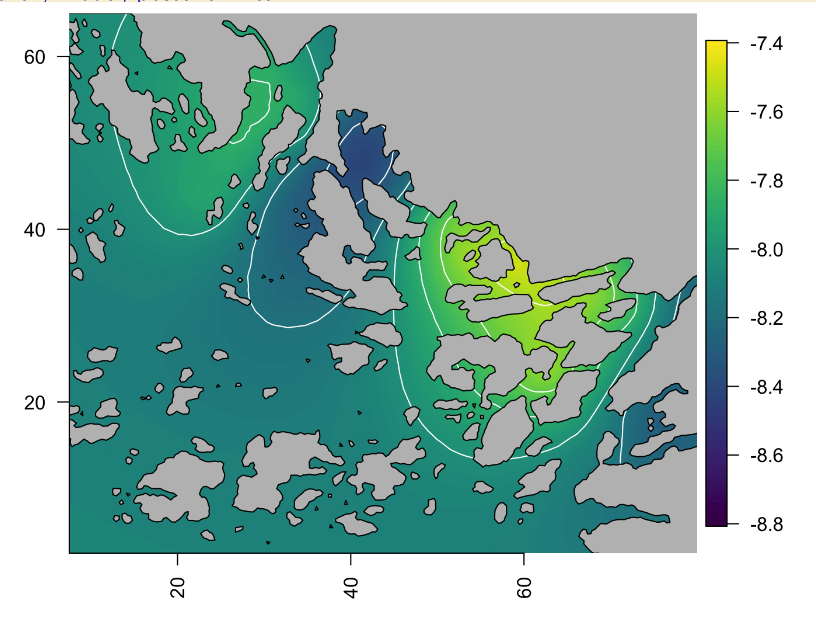
A priori: Fish do not swim over land



Correlation plots.

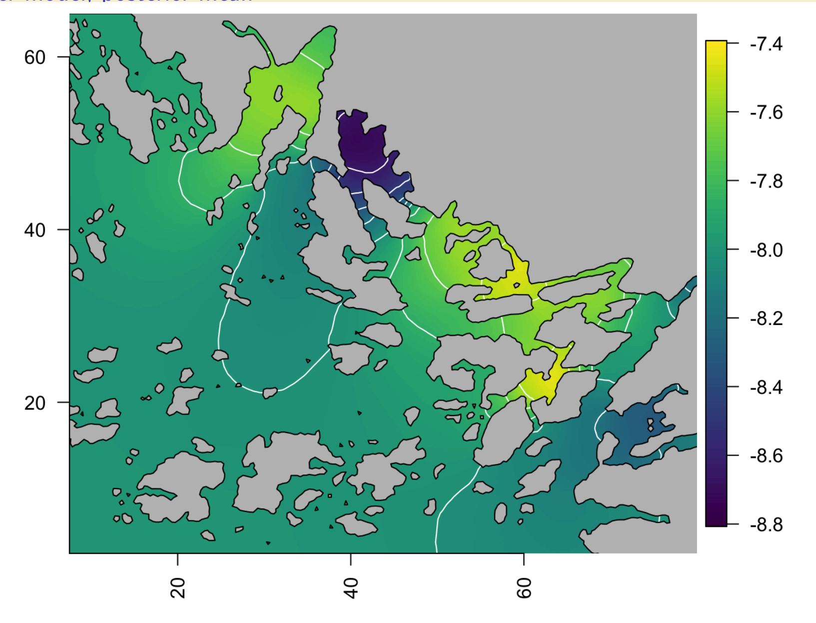
Back to the Fish larvae

Stationary model, posterior mean



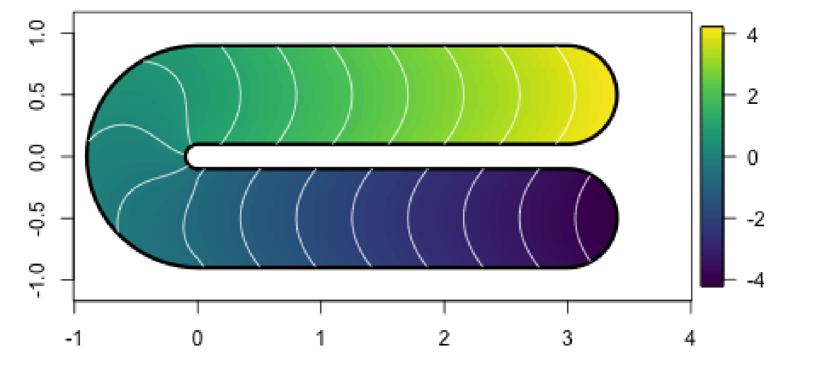
Back to the Fish larvae

Barrier model, posterior mean



Practical presentation

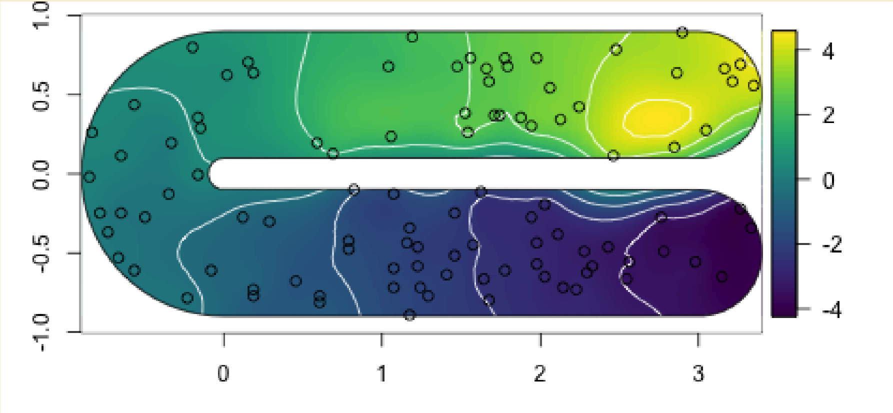
Surface reconstruction



Unknown true surface. Wood et al. (2008).

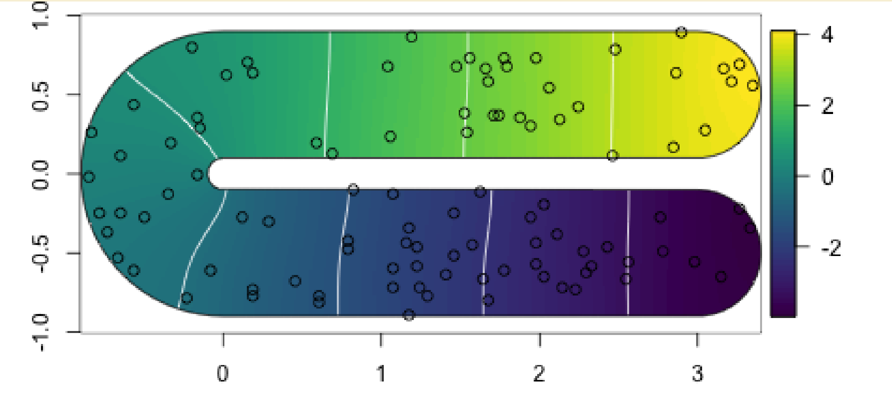
Practical presentation

Stationary model



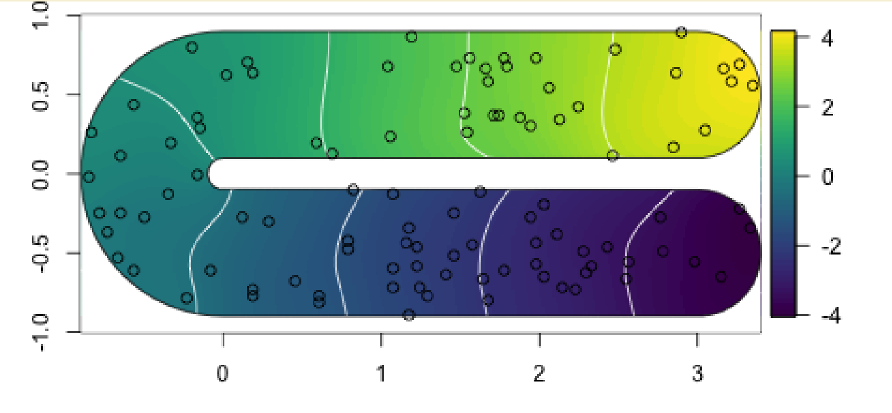
Practical presentation

Neumann model



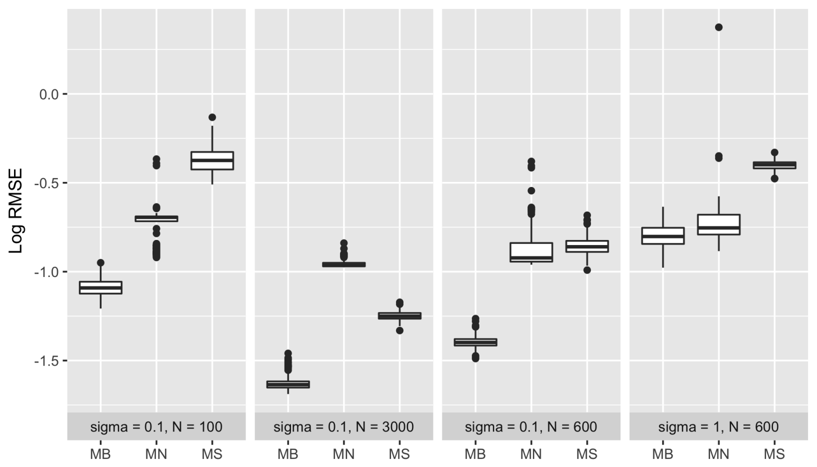
Practical presentation

Barrier model



Practical asymptotics

What model is best with a ton of data?



Would the Barrier problem disappear if we had enough data?

Before we get to the theory

More information:

`https://haakonbakka.bitbucket.io/btopic107.html`

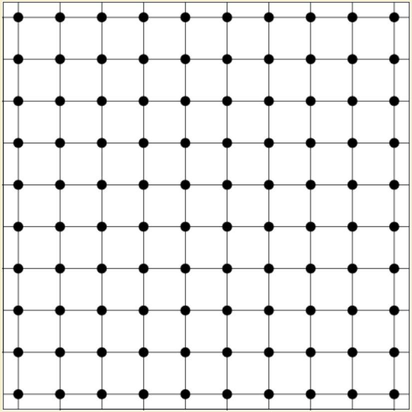
Arxiv: "Bakka"
- Non-stationary Gaussian models with physical barriers
(also the review and the tutorial)

Theoretical approach

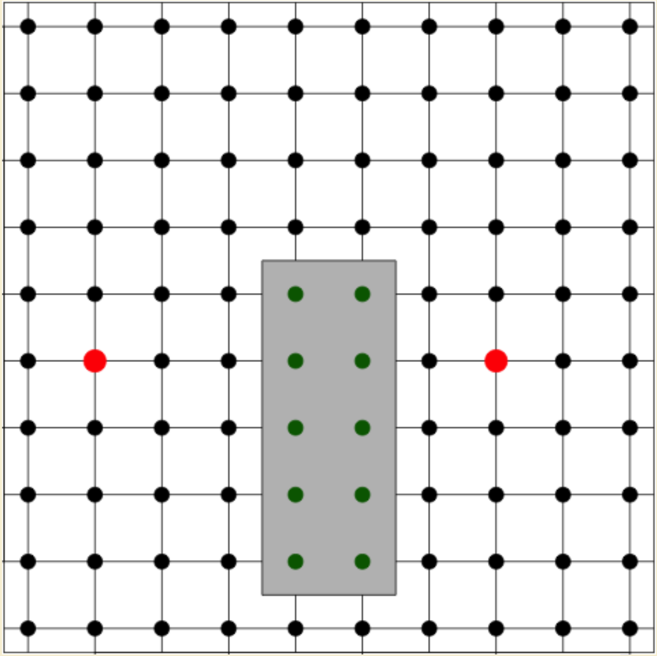
Matérn SAR formulation

Simultaneous Autoregressive

$$U_{i,j} - k\left(U_{i-1,j} + U_{i,j-1} + U_{i+1,j} + U_{i,j+1}\right) = z_{i,j},$$



A new SAR



A new SAR

“At” one point i, j

$$(1-k_2)U_{0,0} - k\left(U_{-1,0} + U_{0,-1} + (1-k_3)U_{1,0} + U_{0,1}\right) = z_{0,0}$$

Decreases dependency in positive x direction.

Disclaimer

Disclaimer

I have no idea how to make this work

Disclaimer

I have no idea how to make this work, if I have to use the SAR (or CAR) specification directly.

Stationary Matérn model again

$$u(s) - \nabla \cdot \frac{r^2}{8} \nabla u(s) = r \sqrt{\frac{\pi}{2}} \sigma_u \mathcal{W}(s) \tag{1}$$

Stationary Matérn model again

$$u(s) - \nabla \cdot \frac{r^2}{8} \nabla u(s) = r\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(s) \tag{1}$$

Why this is better ("easier") than CAR/SAR formulation is explained in "Spatial modelling with R-INLA: A review" (2018), and the references therein. How to understand/solve this equation is explained in the tutorial "How to solve...".

The Barrier model again

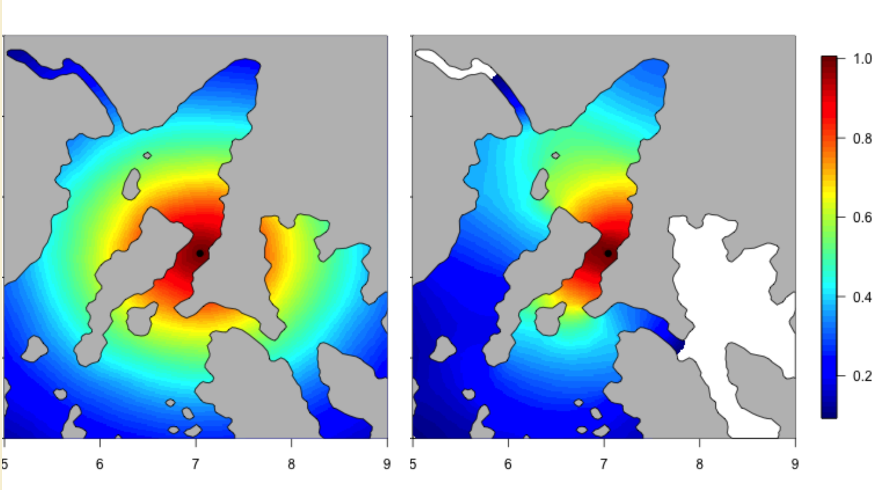
$$
\begin{aligned}
u(s) - \nabla \cdot \frac{r^2}{8} \nabla u(s) &= r\sqrt{\frac{\pi}{2}} \sigma_u \mathcal{W}(s), \text{ for } s \in \Omega_w \\
u(s) - \nabla \cdot \frac{r_b^2}{8} \nabla u(s) &= r_b\sqrt{\frac{\pi}{2}} \sigma_u \mathcal{W}(s), \text{ for } s \in \Omega_b \qquad (2)
\end{aligned}
$$

Now what?

Discretize this differential equation, similarly to Lindgren et al. (2011).

Summary

A priori: Fish do not swim over land



Fast, easy to use, robust and automatic.

The End

Thank you for listening!

Please contact me if you are interested: `bakka@r-inla.org`

Special thanks to: Håvard Rue, Jarno Vanhatalo, Janine Illian, Daniel Simpson, Elias Krainski, Finn Lindgren, Simon Wood, Rosa Crujeiras, and Esther Jones.