Teleinformatics Engineering Department, Federal University of Ceará

Bayesian Inference of Disease Mappings with Gaussian Processes

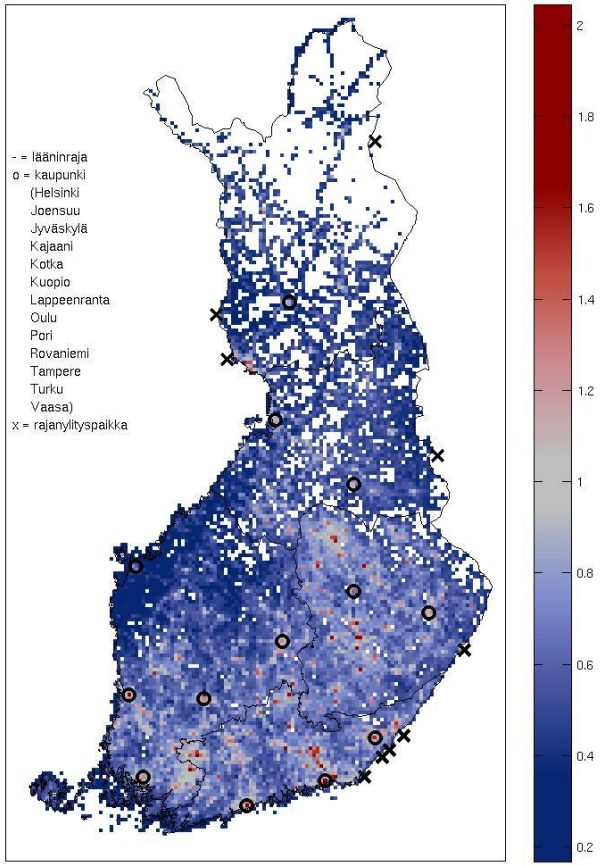
Filipe Pereira de Farias, Bachelor's student.
filipepfarias@fisica.ufc.br

April 27, 2021

Outline

Introduction

⇝ **What it is and what we can do with it?**



Definition

A **disease map** is a collection of objects (e.g. the count of occurrences in a group of individuals) in their geographical position (Lawson et al., 2001).

With disease mapping we can (Best et al., 2005):

- Identify **disease patterns** in space.
- Obtain clues as to the **disease aetiology** and highlight areas of **elevated risk**.
- Define the likelihood of a disease.

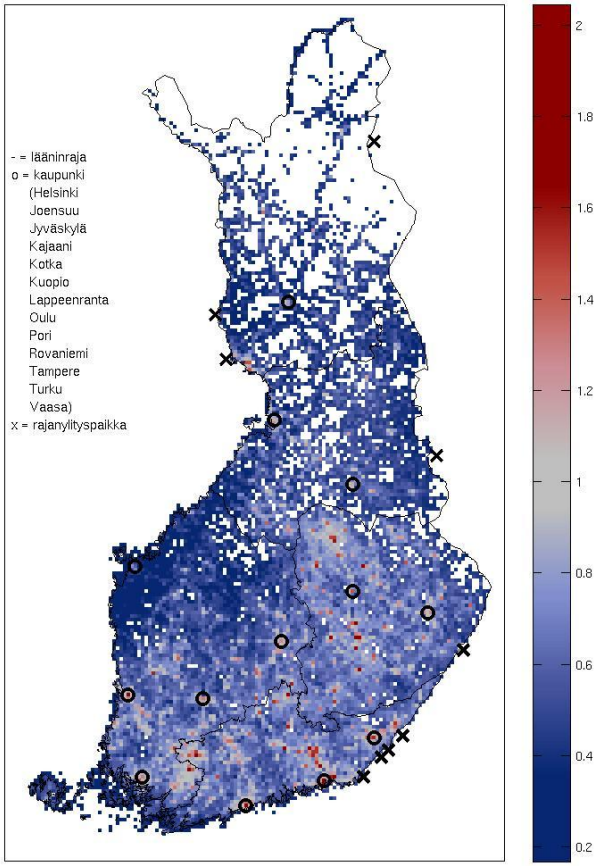
Our objetive is to **infer** the **spatial distribution** of the risk of a disease

- Given its **incidence**.

Figure by Vanhatalo, Mäkelä, et al. (2010).

Introduction
Disease Mapping

⇝ What it is and what we can do with it?



Mapping the risk of a disease allows us to identify potential **risk factors** and to **develop public policies** to mitigate their effects.

With disease mapping we can (Best et al., 2005):

- Identify **disease patterns** in space.
- Obtain clues as to the **disease aetiology** and highlight areas of **elevated risk**.
- Define the likelihood of a disease.

Our objetive is to **infer** the **spatial distribution** of the risk of a disease

- Given its **incidence**.

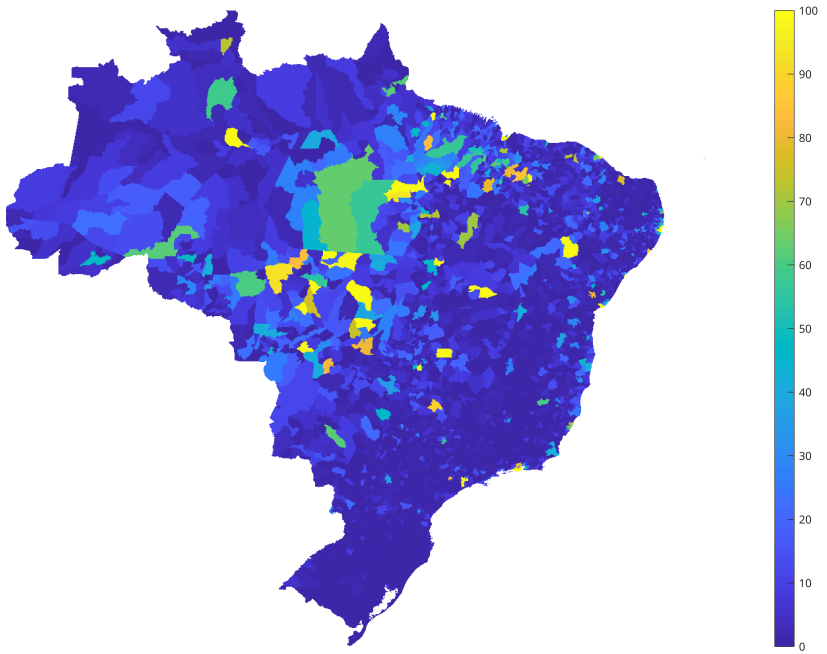
Figure by Vanhatalo, Mäkelä, et al. (2010).

Introduction
Disease Mapping

⇝ Our problem: Leprosy in Brazil

In this work, the **risk of Leprosy**, a neglected tropical disease, is studied.

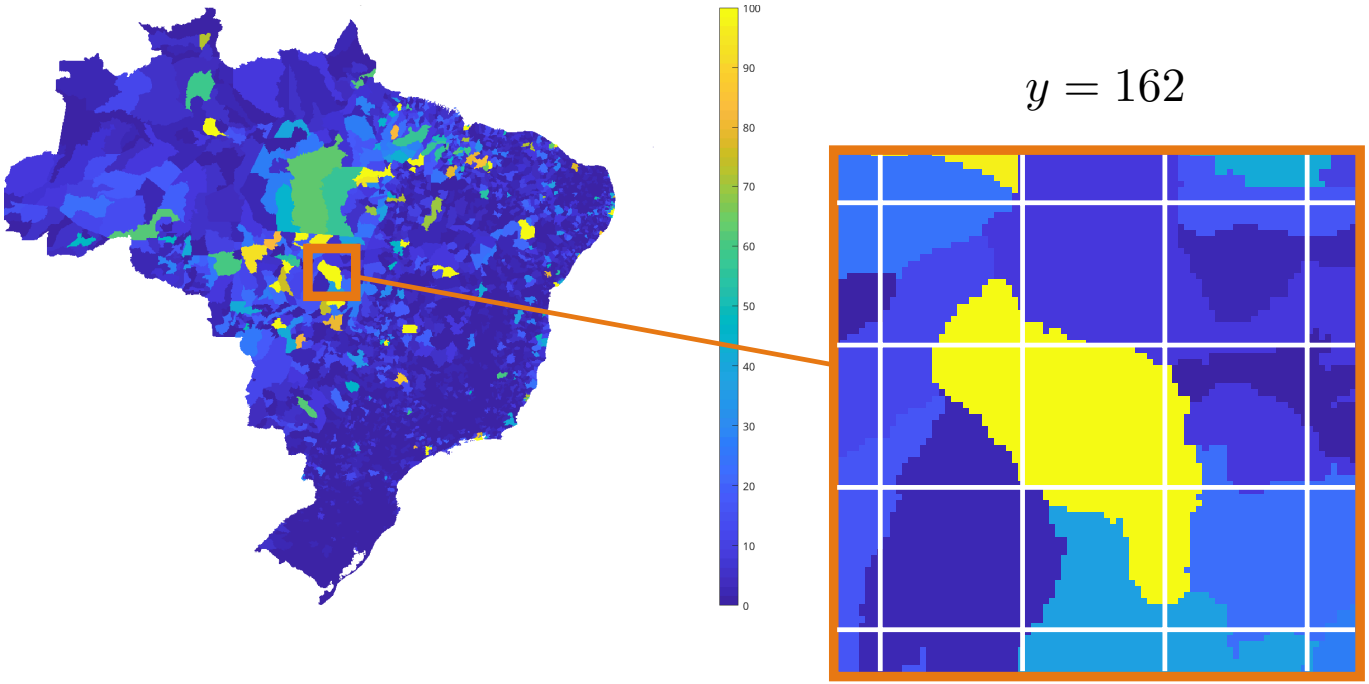
- For each municipality, the number of **notified cases** is available.
- The data are collected by the National Health System (SUS).
- We use data relative to the period 2001-2019.



⇝ Our problem: Leprosy in Brazil

On a grid over Brazil, we associate to the n-th cell the corresponding cases y_n.

- We use **Bayesian inference** to determine the risk μ_n in each cell.

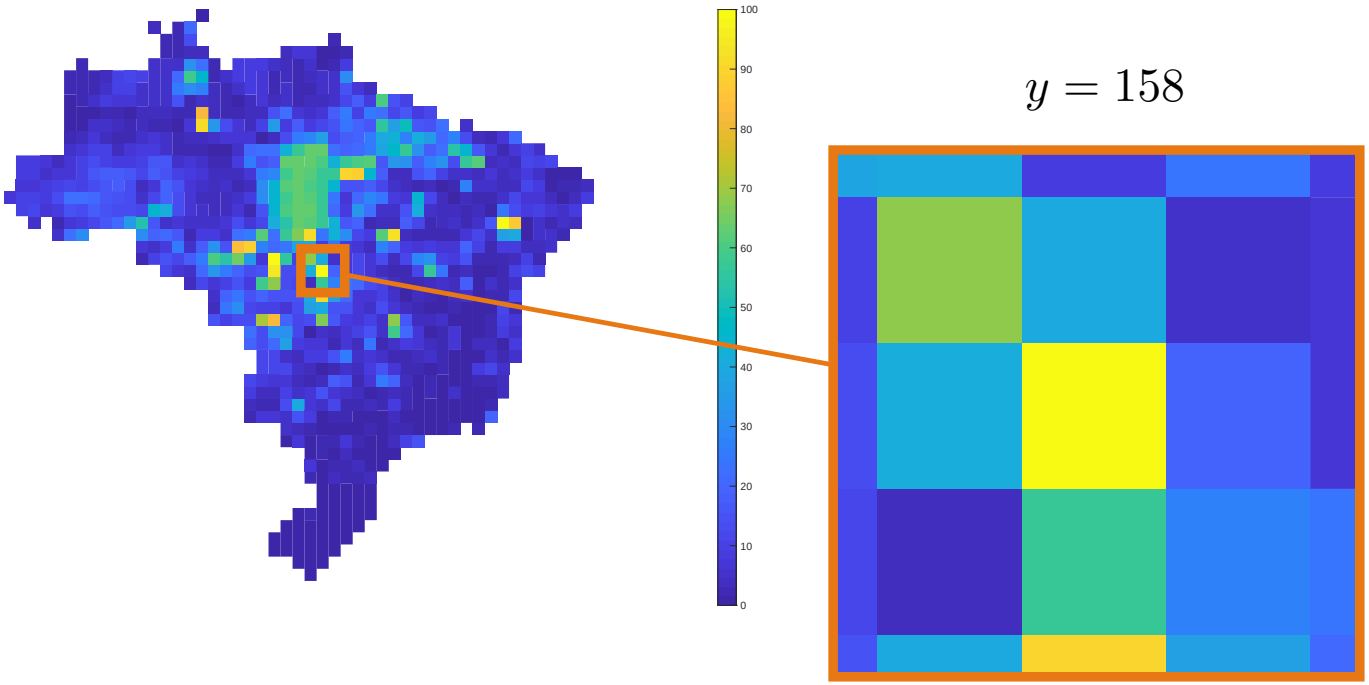


In the figure, municipality of Querência, state of Mato Grosso.

⇝ Our problem: Leprosy in Brazil

On a grid over Brazil, we associate to the n-th cell the corresponding cases y_n.

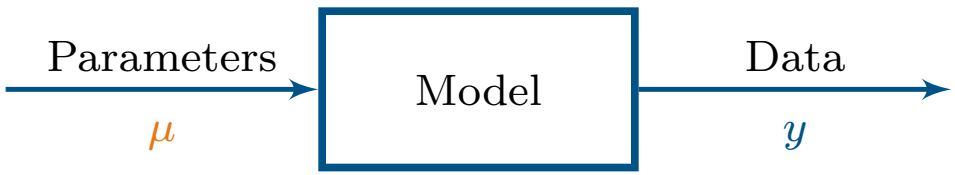
- We use **Bayesian inference** to determine the risk μ_n in each cell.



In the figure, municipality of Querência, state of Mato Grosso.

Bayesian Inference

⇝ What is inference?



- We're interested in the **process** that has generated the **observed data** y.
- We assume a **model** that depends on some **unobserved parameters**.

We assume that the **disease occurrence** y depends on the **hidden risk** μ.

- We want to estimate μ, given y and given the model.

⇝ What is *Bayesian* inference?

To model the variables in our problem (y, μ), we use **probability distributions**.

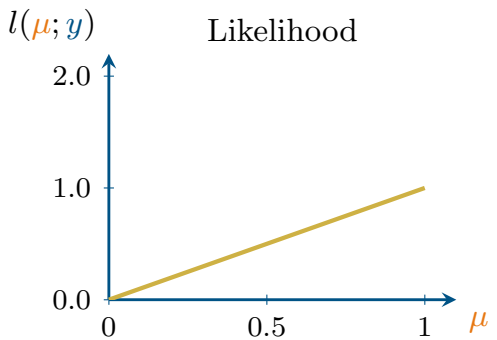
We combine the knowledge about the data (**likelihood**) with knowledge about the parameters (**prior**), to get an updated knowledge about parameters (**posterior**).

⇝ What is *Bayesian* inference?

To model the variables in our problem (y, μ), we use **probability distributions**.

We combine the knowledge about the data (**likelihood**) with knowledge about the parameters (**prior**), to get an updated knowledge about parameters (**posterior**).

$$p(y|\mu)$$



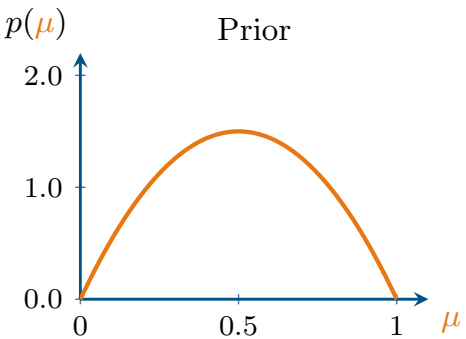
The **likelihood** l expresses how the **probability** $p(y|\mu)$ of the data varies for different values of the **parameter** μ.

⇝ What is *Bayesian* inference?

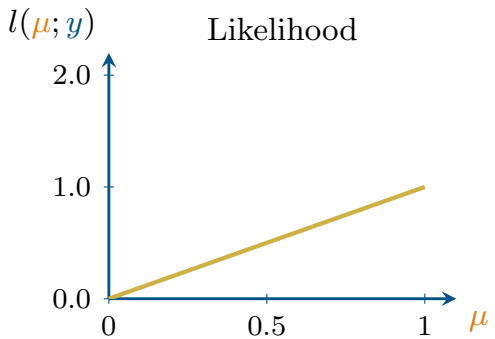
To model the variables in our problem (y, μ), we use **probability distributions**.

We combine the knowledge about the data (**likelihood**) with knowledge about the parameters (**prior**), to get an updated knowledge about parameters (**posterior**).

$$p(y|\mu)p(\mu) = p(y, \mu)$$



The **prior** expresses the **probability** $p(\mu)$ of the parameter, before observing the data y.



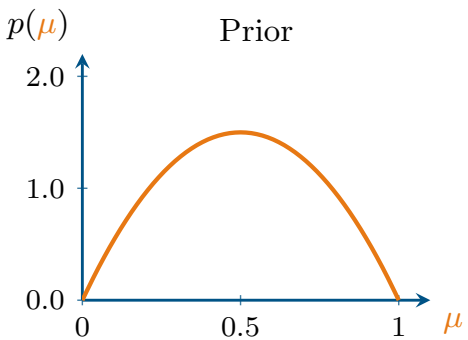
The **likelihood** l expresses how the **probability** $p(y|\mu)$ of the data varies for different values of the **parameter** μ.

⇝ What is *Bayesian* inference?

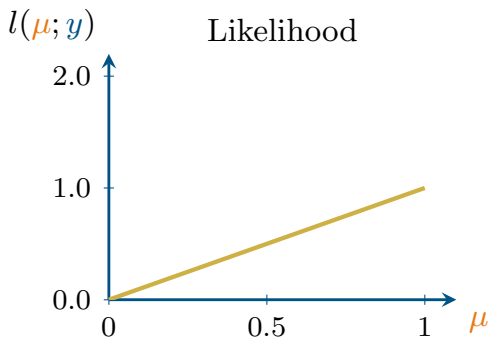
To model the variables in our problem (y, μ), we use **probability distributions**.

We combine the knowledge about the data (**likelihood**) with knowledge about the parameters (**prior**), to get an updated knowledge about parameters (**posterior**).

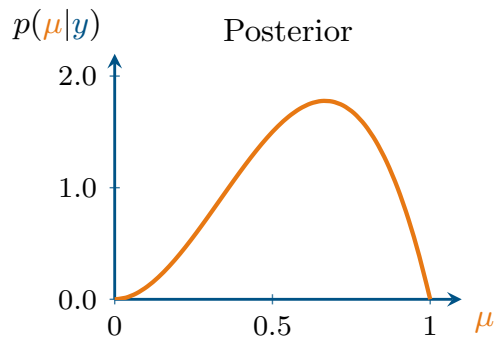
$$p(y|\mu)p(\mu) = p(y, \mu)$$



The **prior** expresses the **probability** $p(\mu)$ of the parameter, before observing the data y.



The **likelihood** l expresses how the **probability** $p(y|\mu)$ of the data varies for different values of the **parameter** μ.



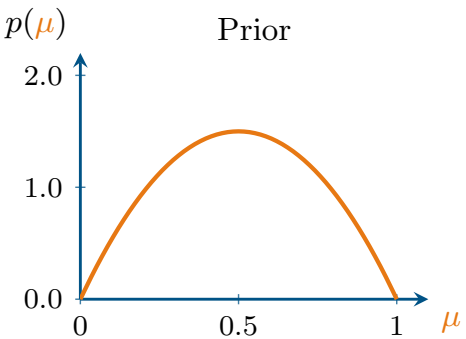
The **posterior** $p(\mu|y)$ combines these probabilities and updates the prior using the observed data.

⇝ What is *Bayesian* inference?

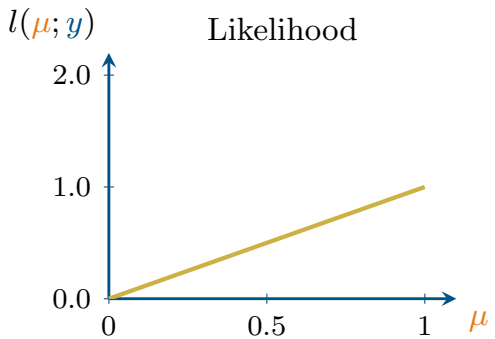
To model the variables in our problem (y, μ), we use **probability distributions**.

We combine the knowledge about the data (**likelihood**) with knowledge about the parameters (**prior**), to get an updated knowledge about parameters (**posterior**).

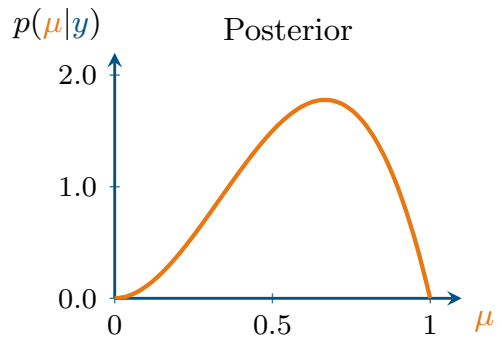
$$\frac{\overbrace{p(\mu)}^{\text{Prior}}\ \overbrace{p(y|\mu)}^{\text{Likelihood}}}{p(y)} = \overbrace{p(\mu|y)}^{\text{Posterior}} \qquad \text{(Bayes' rule)}$$



The **prior** expresses the **probability** $p(\mu)$ of the parameter, before observing the data y.



The **likelihood** l expresses how the **probability** $p(y|\mu)$ of the data varies for different values of the **parameter** μ.



The **posterior** $p(\mu|y)$ combines these probabilities and updates the prior using the observed data.

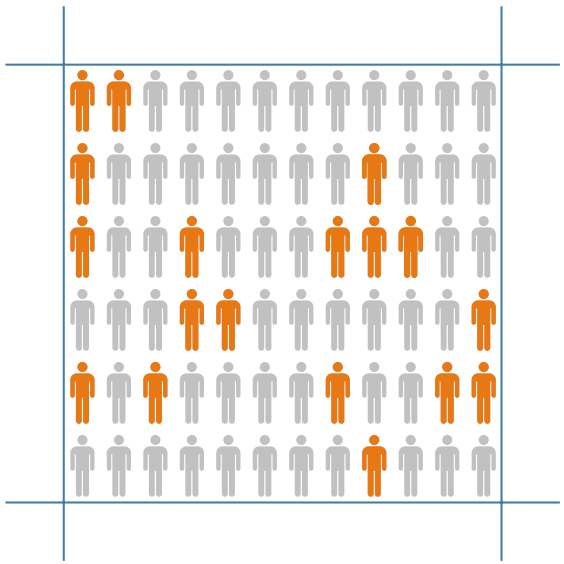
$\rightsquigarrow$ **How to model the disease cases?**

Let α be the probability of an individual (in a cell) to be positive to the disease.

- If we test m individuals (gray), then the probability of y of them being sick (orange) is

$$\binom{m}{y} \alpha^y (1-\alpha)^{m-y}. \quad (1)$$

- A **Binomial distribution**.



⇝ How to model the disease cases?

- We can rewrite equation (1) as

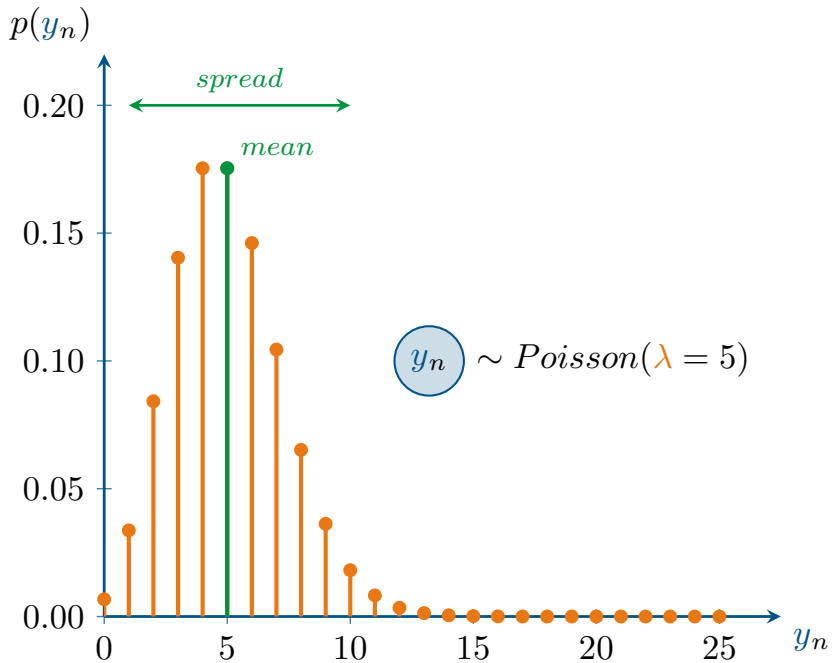
$$\frac{m!}{(m-y)!y!}\alpha^y(1-\alpha)^{m-y}$$

- For $m \to \infty$ and $\alpha \to 0$, we obtain the **Poisson distribution**.

$$\frac{\lambda^y e^{-\lambda}}{y!},$$

where $m\alpha \to \lambda$.

- The Poisson distribution is parametrised by λ.



We assume that in each cell n the number of occurrences y_n follows a Poisson distribution with parameter λ_n, which is thus the expected number of cases.

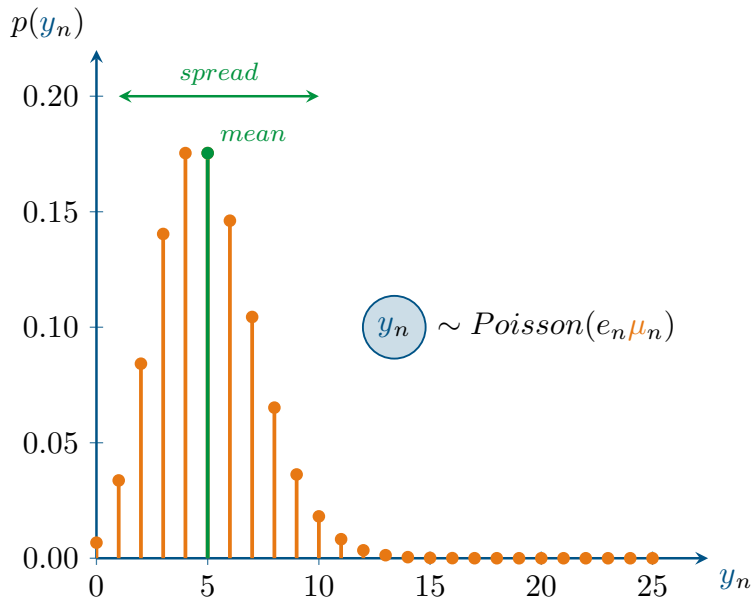
⇝ **How to count the disease cases?**

The **expected number of cases** in cell n is also given as weighted number of cases $\lambda_n = e_n \mu_n$, where the **standardised cases** e_n are weighted by the **risk** μ_n.

- The standardised cases are

$$e_n = \frac{\text{total cases}}{\text{population}} \times p_n$$

- p_n is the cell population.



Bayesian Inference
Observation Model

⇝ How to estimate the risk?

Our goal is to **infer** the risk μ_n given the observed cases y_n, for the whole country.

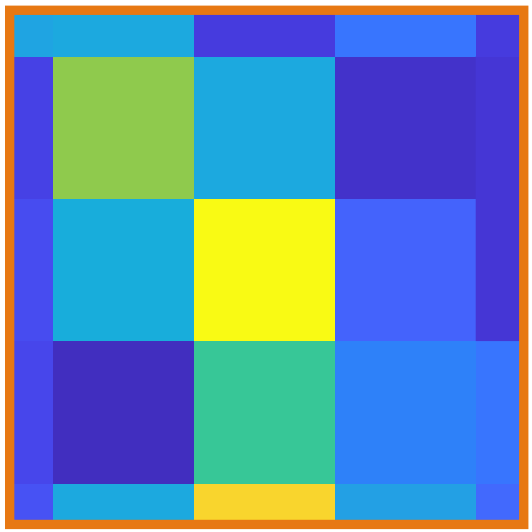
- We model each y_n as an independent Poisson variable with mean $e_n \mu_n$.
- The risks are **positive variables**, to **jointly model** them as Gaussian Process we transform them

$$f_n = \log \mu_n$$

- The resulting **Gaussian Process** is

$$f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right)$$

- Because of this model, each risk μ_n is **connected** with all the other risks, and thus they relate to each other.



Bayesian Inference
Observation Model

⇝ How to estimate the risk?

Our goal is to **infer** the risk μ_n given the observed cases y_n, for the whole country.

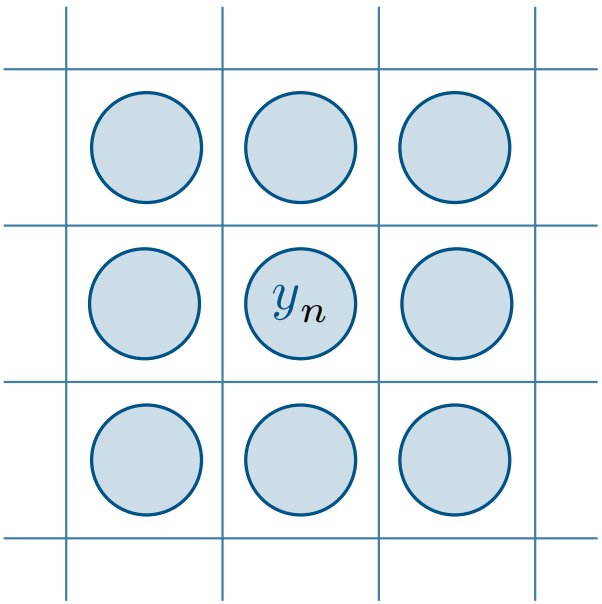
- We model each y_n as an independent Poisson variable with mean $e_n \mu_n$.
- The risks are **positive variables**, to **jointly model** them as Gaussian Process we transform them

$$f_n = \log \mu_n$$

- The resulting **Gaussian Process** is

$$f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right)$$

- Because of this model, each risk μ_n is **connected** with all the other risks, and thus they relate to each other.



Bayesian Inference
Observation Model

⇝ How to estimate the risk?

Our goal is to **infer** the risk μ_n given the observed cases y_n, for the whole country.

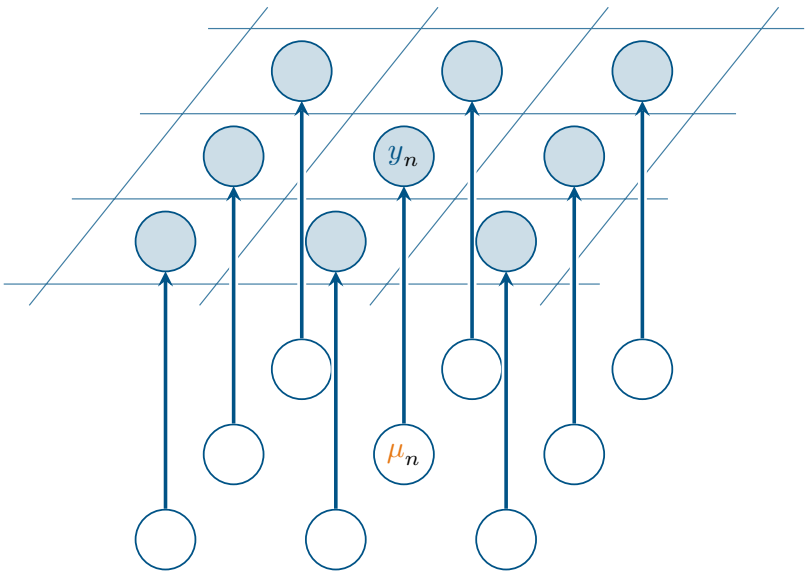
- We model each y_n as an independent Poisson variable with mean $e_n \mu_n$.
- The risks are **positive variables**, to **jointly model** them as Gaussian Process we transform them

$$f_n = \log \mu_n$$

- The resulting **Gaussian Process** is

$$f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right)$$

- Because of this model, each risk μ_n is **connected** with all the other risks, and thus they relate to each other.



Bayesian Inference
Observation Model

⇝ How to estimate the risk?

Our goal is to **infer** the risk μ_n given the observed cases y_n, for the whole country.

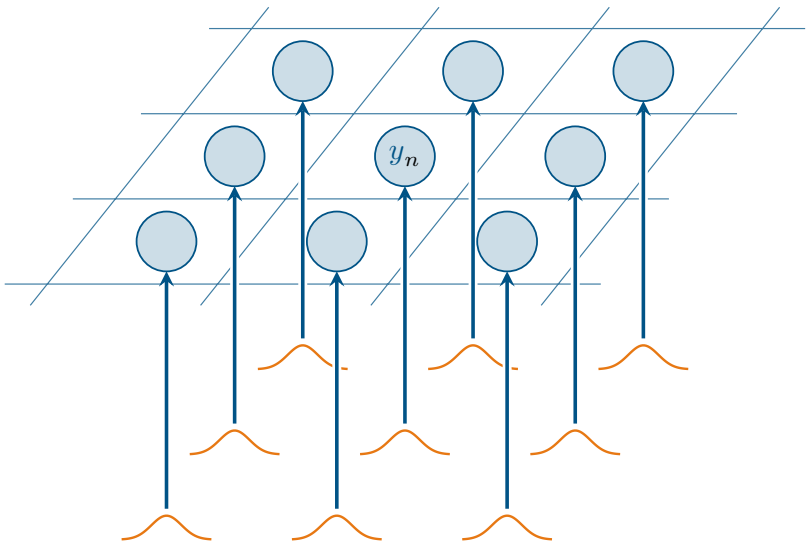
- We model each y_n as an independent Poisson variable with mean $e_n \mu_n$.
- The risks are **positive variables**, to **jointly model** them as Gaussian Process we transform them

$$f_n = \log \mu_n$$

- The resulting **Gaussian Process** is

$$f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right)$$

- Because of this model, each risk μ_n is **connected** with all the other risks, and thus they relate to each other.



Bayesian Inference
Observation Model

⇝ How to estimate the risk?

Our goal is to **infer** the risk μ_n given the observed cases y_n, for the whole country.

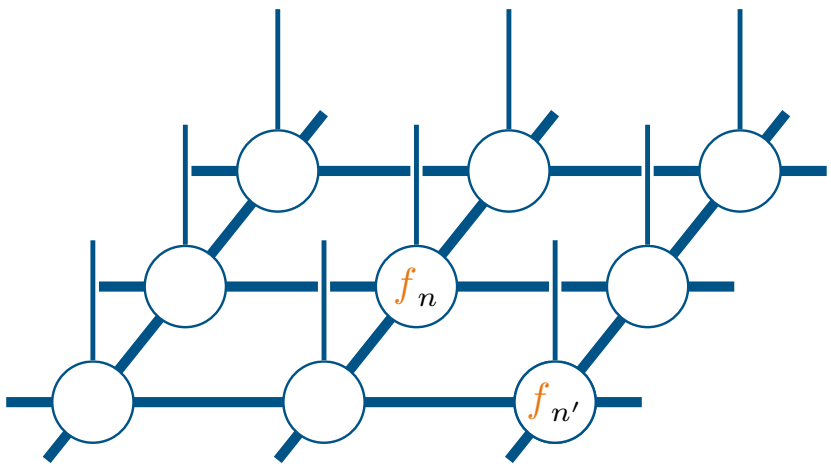
- We model each y_n as an independent Poisson variable with mean $e_n \mu_n$.
- The risks are **positive variables**, to **jointly model** them as Gaussian Process we transform them

$$f_n = \log \mu_n$$

- The resulting **Gaussian Process** is

$$f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right)$$

- Because of this model, each risk μ_n is **connected** with all the other risks, and thus they relate to each other.

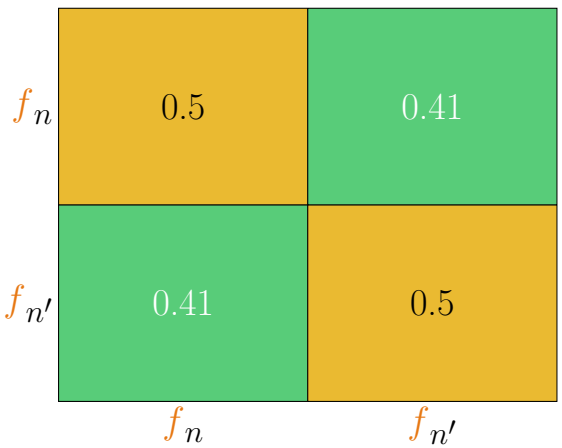


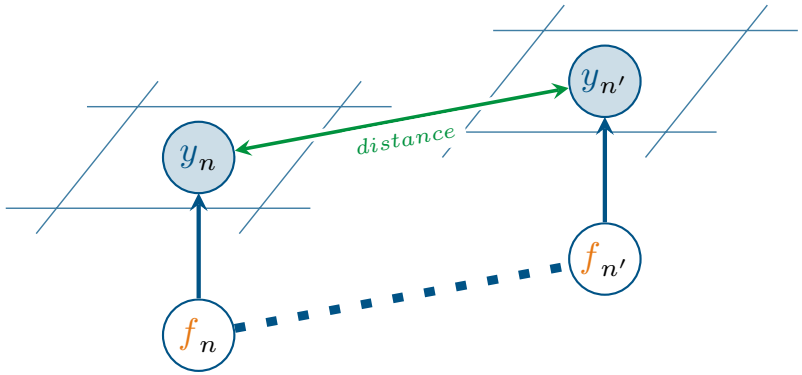
The strength of the relation between cells n and n' will depend on their distance

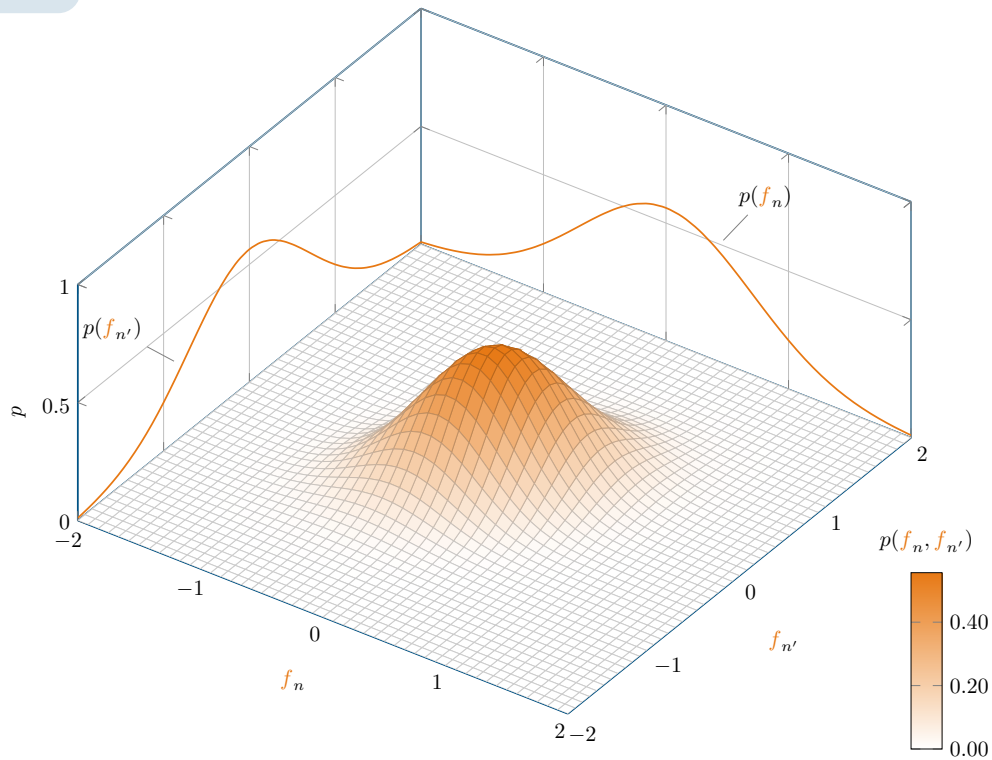
$$|\mathbf{x} - \mathbf{x}'|$$

A **covariance matrix** encodes how (log) risks f_n and $f_{n'}$ relate to each other.

When cells n and n' are **near** to each other, the covariance relating f_n and $f_{n'}$ will be large and small otherwise.





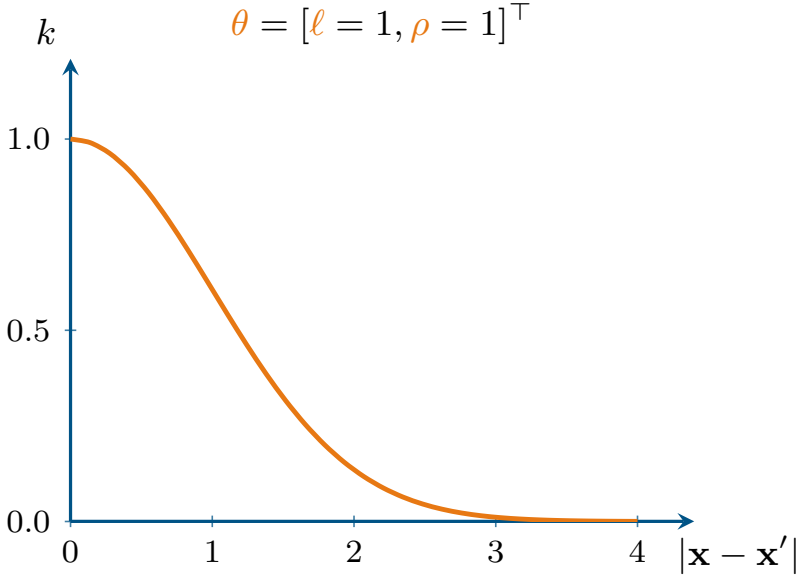


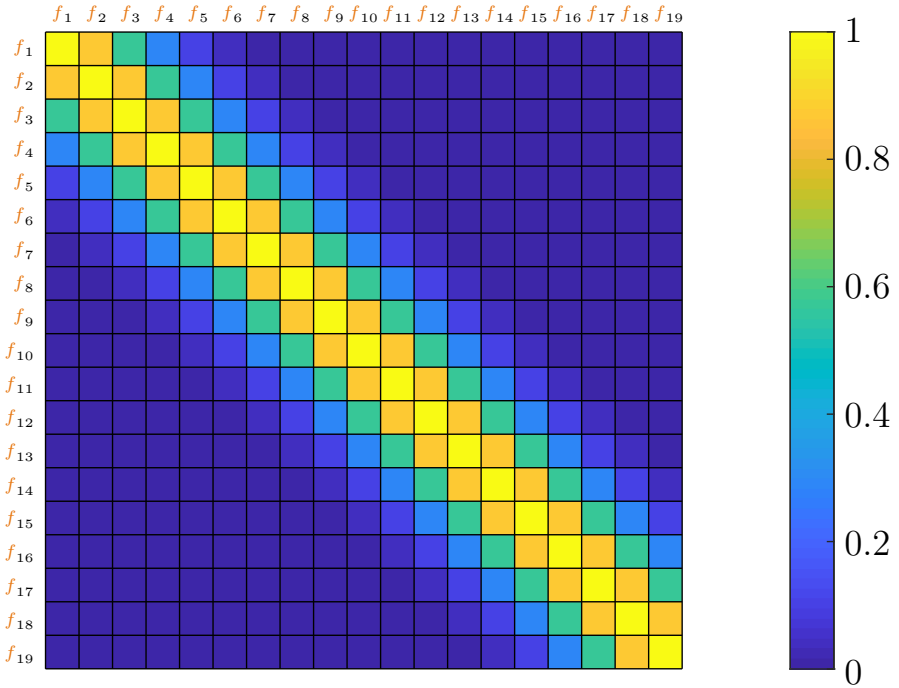
⇝ Creating covariance matrices with covariance functions...

Covariance matrices are obtained from **covariance functions** (CF)

- CFs model the relation between f_n and $f_{n'}$.

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \rho^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell^2}\right)$$



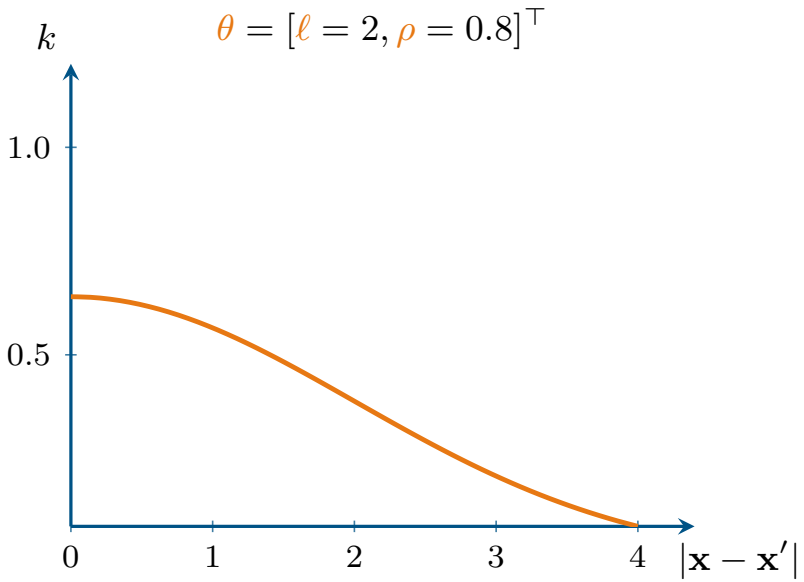


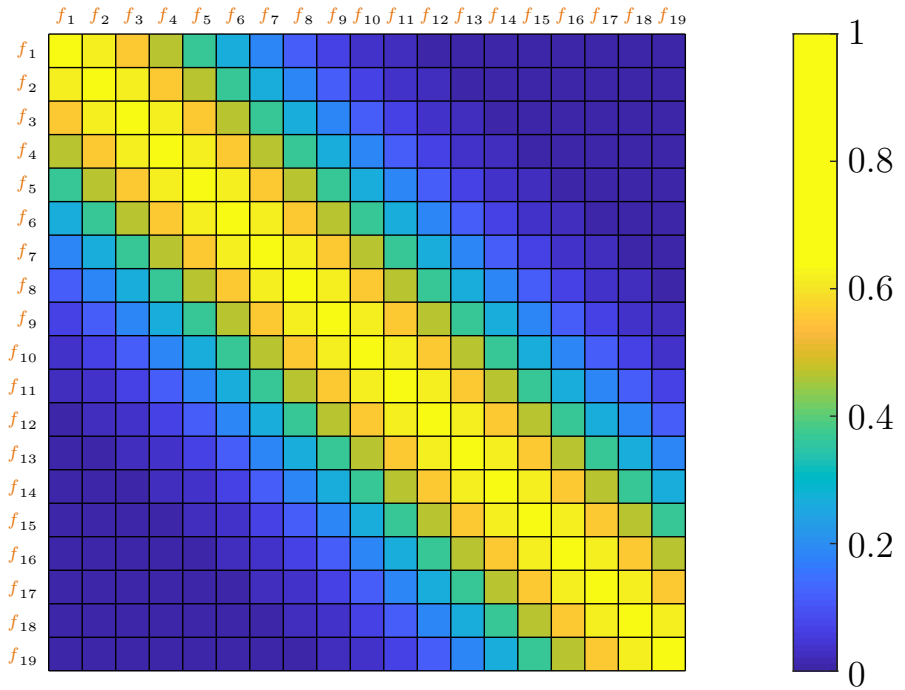
⇝ Creating covariance matrices with covariance functions...

Covariance matrices are obtained from **covariance functions** (CF)

- CFs model the relation between f_n and $f_{n'}$.

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \rho^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell^2}\right)$$



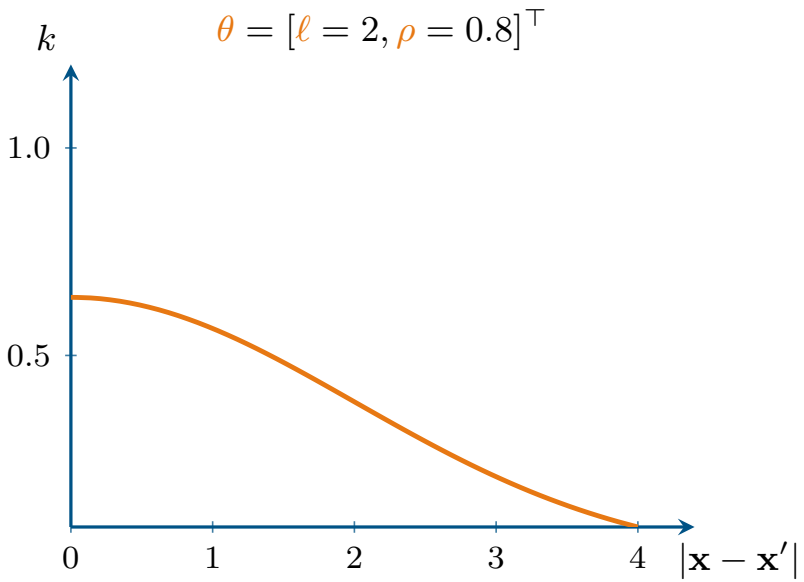


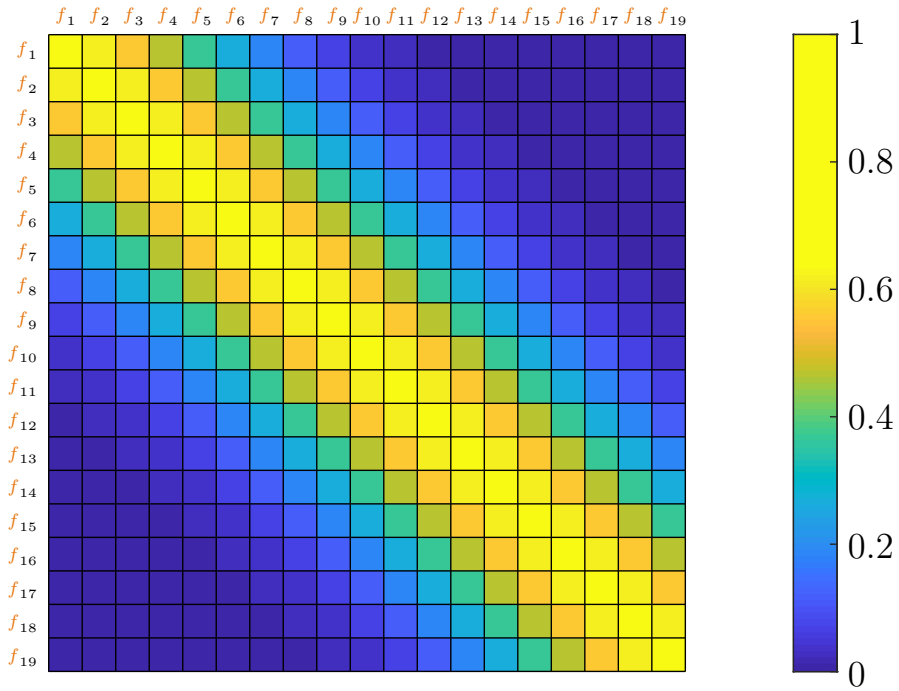
$\rightsquigarrow$ **Creating covariance matrices with covariance functions...**

$$k_{\text{SE}}(\mathbf{x}, \mathbf{x}') = \rho^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\ell^2}\right)$$

The covariance functions are also **parametrised** by a set of random variables θ.

- $\theta = [\ell, \rho]$ are parameters, the **length-scale** and **magnitude**.



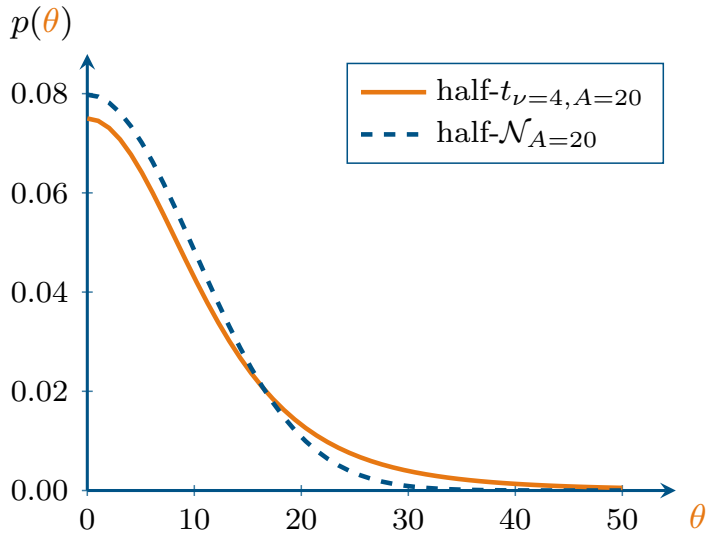


Bayesian Inference
Prior Model

$\rightsquigarrow$ **The parameters of the CF's are also random variables**

Each element of θ is modeled with a **half-Student's t distribution**, which is a **weakly informative hyperprior**

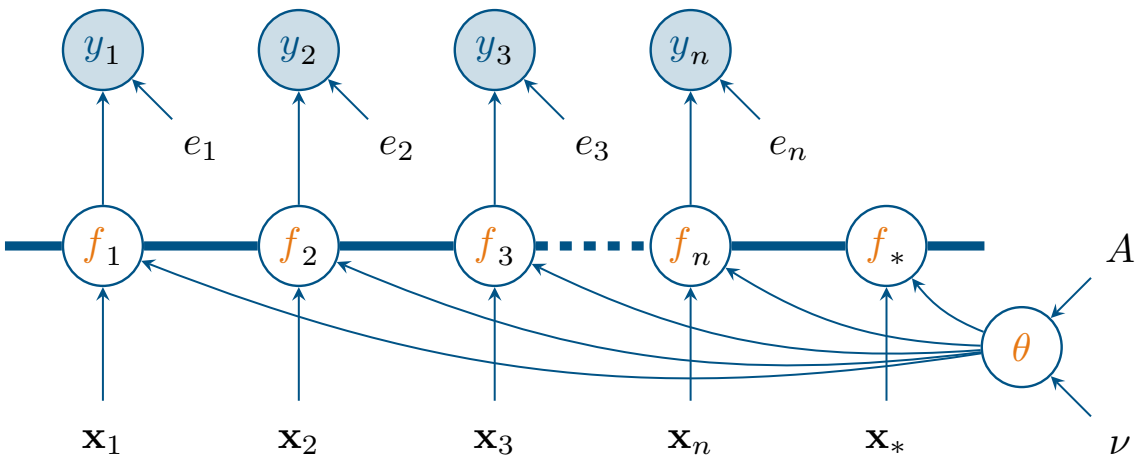
- Distributions with **large variance** are called **weakly informative**.



⇝ A graphical representation

The probabilistic model of the disease mapping problem is

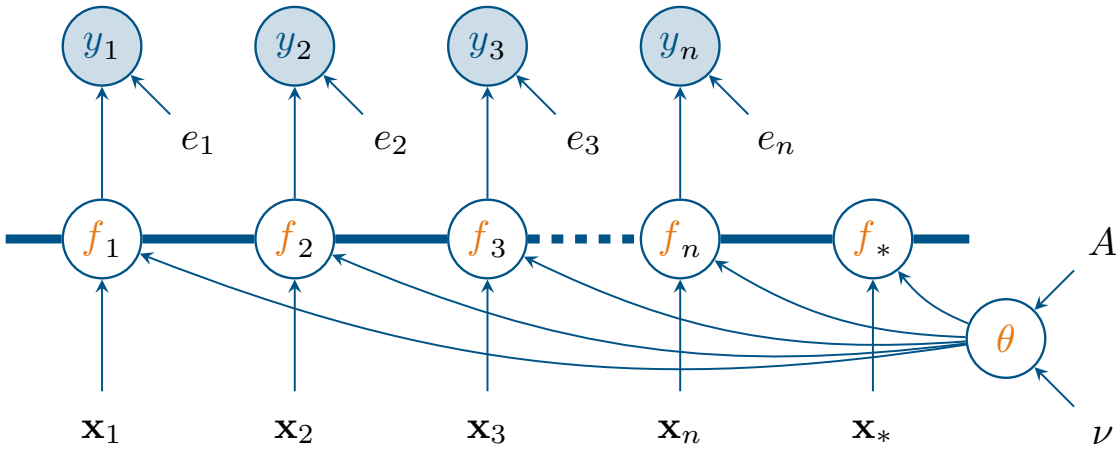
$$
\begin{cases}
y_1, \ldots, y_n | \mu_1, \ldots, \mu_n \sim \prod_{n=1}^{N} Poisson(e_n \mu_n) & \text{(Observation model)} \\
\mu_n = \exp(f_n), \quad \forall n & \text{(log-Gaussian transform)} \\
f_1, \ldots, f_n | \theta \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}' | \theta)\right) & \text{(GP prior)} \\
\theta \sim \text{half-Student's } t(A, \nu) & \text{(Hyperprior)}
\end{cases}
$$



Conducting the inference

Conducting the inference

$$p(\mathbf{f}, \theta, \mathbf{y}) = p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}, \theta)$$
$$= p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}|\theta)p(\theta)$$



We are interested in the distribution of the risk given data, the **posterior** $p(\mathbf{f} \mid \mathbf{y})$

$$p(\mathbf{f} \mid \mathbf{y}) = \int p(\mathbf{f}, \theta \mid \mathbf{y}) d\theta = \frac{1}{Z_p} \int \overbrace{p(\mathbf{y} \mid \mathbf{f})}^{\text{Poisson Obs. Model}} \underbrace{p(\mathbf{f} \mid \mathbf{X}, \theta)}_{\text{GP Prior Model}} \overbrace{p(\theta)}^{\text{half-}t\text{ Hyperprior}} d\theta,$$

where $\mathbf{X}$ are the cells' position $\mathbf{X} = [\mathbf{x}_n]_{n=1}^N$, $\mathbf{f} = [f_n]_{n=1}^N$, $\mathbf{y} = [y_n]_{n=1}^N$ and Z_p

$$Z_p = \iint p(\mathbf{y} \mid \mathbf{f})\, p(\mathbf{f} \mid \mathbf{X}, \theta)\, p(\theta) d\theta d\mathbf{f}.$$

This problem is **not solvable analytically** and we need to simplify it.

Conducting the inference
Empirical Bayes

$\rightsquigarrow$ **Pretend it is *fully Bayesian*...**

When the **marginal posterior** $p(\theta \mid \mathbf{y})$ is **smooth near its mode** $\hat{\theta}$, we can use its **evidence approximation**, also known as the empirical Bayes approximation*

$$p(\mathbf{f} \mid \mathbf{y}) \approx p(\mathbf{f} \mid \mathbf{y}, \hat{\theta}), \quad \text{where } \hat{\theta} = \arg\max_{\theta} p(\theta \mid \mathbf{y}).$$

In a **fully Bayesian** approach, we **consider** all the possible values of θ, whereas in empirical Bayes we **rank** them and choose $\hat{\theta}$, as the "**best** one" among them.

- The marginal posterior is obtained using the Bayes' rule

$$p(\theta \mid \mathbf{y}) \propto p(\theta)\, p(\mathbf{y} \mid \theta) = p(\theta) \int p(\mathbf{y} \mid \mathbf{f}) p(\mathbf{f} \mid \mathbf{X}, \theta)\, d\mathbf{f}$$

- The posterior of the risk becomes

$$p(\mathbf{f} \mid \mathbf{y}, \hat{\theta}) = \frac{p(\mathbf{y} \mid \mathbf{f})\, p(\mathbf{f} \mid \mathbf{X}, \hat{\theta})}{\int p(\mathbf{y} \mid \mathbf{f})\, p(\mathbf{f} \mid \mathbf{X}, \hat{\theta}) d\mathbf{f}}$$

*See D. J. MacKay (1992) for a more detailed discussion.

Conducting the inference
Laplace Approximation

To compute $\hat{\theta}$, we maximise $\int p(\mathbf{y} \mid \mathbf{f}) p(\mathbf{f} \mid \mathbf{X}, \theta)\, d\mathbf{f}$ after approximating it with an **unnormalised Gaussian distribution** using second order Taylor expansions*.

We approximate the integral in the computation of the posterior in a similar way.

As a result, we obtain

Approximated Posterior

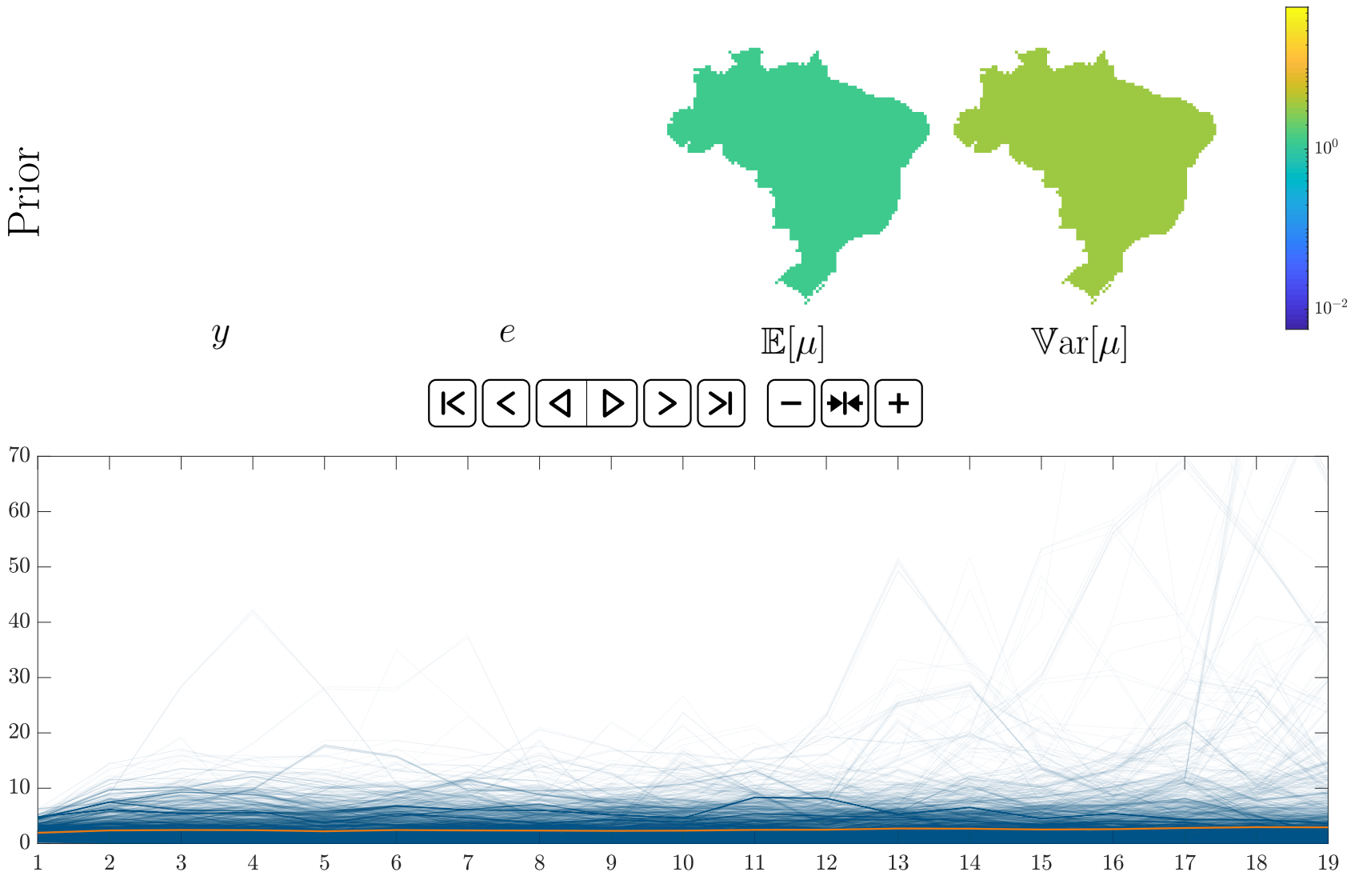
$$p(\mathbf{f} \mid \mathbf{y}, \hat{\theta}) \approx \mathcal{N}\left(\mathbf{f} \mid \hat{\mathbf{f}}, \mathbf{H}^{-1}\right)$$

where $\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} p(\mathbf{f} \mid \mathbf{y}, \hat{\theta})$, $\mathbf{H}^{-1} = -\nabla^2_{\mathbf{f}} \log p(\hat{\theta} \mid \mathbf{y}, \hat{\theta})$ and $\hat{\theta} = \arg\max_{\theta} p(\theta \mid \mathbf{y})$.

*Details in Vanhatalo, Pietiläinen, et al. (2010).

Results

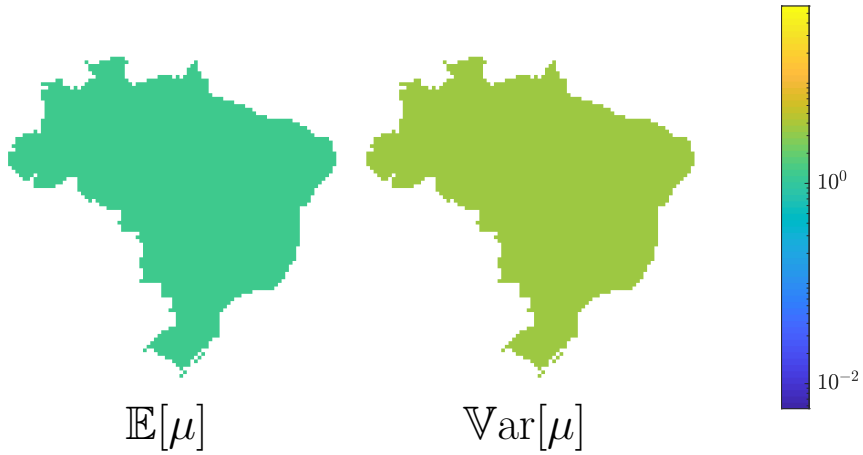
Results



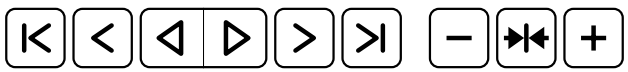
Each cell is 43×43 km

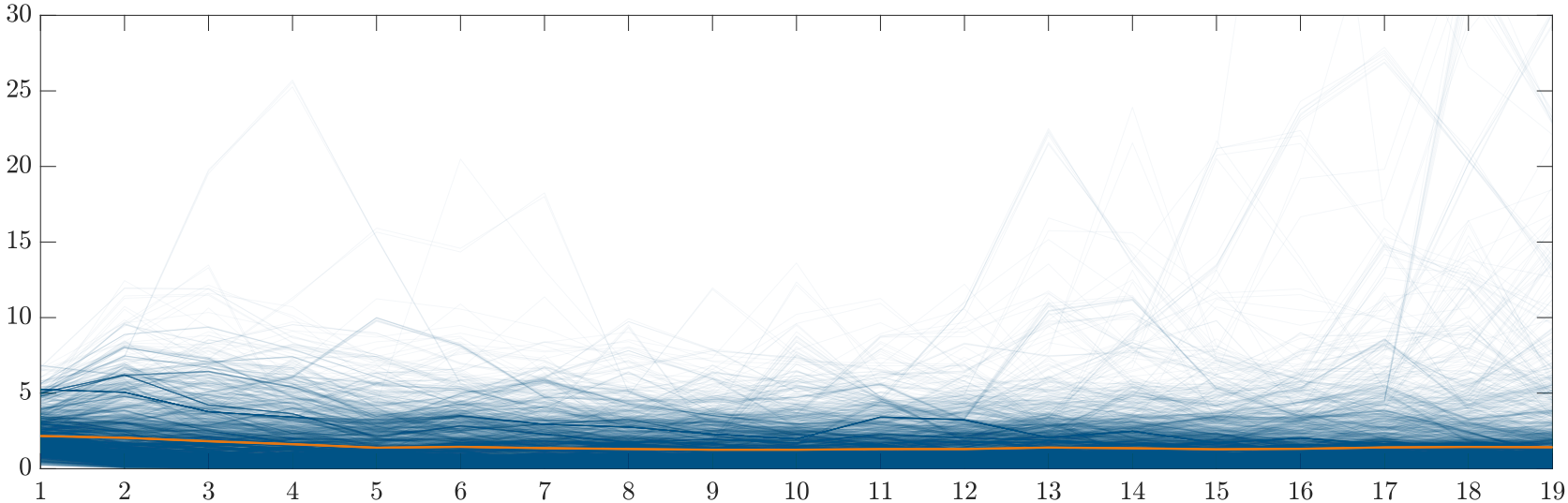
Results

Prior



y e $\mathbb{E}[\mu]$ $\mathbb{V}\mathrm{ar}[\mu]$





Each cell is 43×43 km

Bibliography I

Best, Nicky et al. (2005). "A comparison of Bayesian spatial models for disease mapping". In: *Statistical Methods in Medical Research* 14.1, pp. 35–59. DOI: 10.1191/0962280205sm388oa.

Bishop, C.M. (2016). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York. ISBN: 9781493938438.

Diggle, P.J. et al. (2019). *Model-based Geostatistics for Global Public Health: Methods and Applications*. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press. ISBN: 9781351743273.

Ferreira, A. F. (Feb. 2019). "Hanseníase em territórios das regiões norte e nordeste do Brasil: contextos epidemiológicos e operacionais de controle (*Leprosy in territories in the North and Northeast regions of Brazil: epidemiological and operational control contexts*)". M.Sc. in Public Health. Faculdade de Medicina, Universidade Federal do Ceará.

Gelman, A. et al. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN: 9781439840955.

Lawson, A.B. et al. (2001). *An Introductory Guide to Disease Mapping*. Pharmaceutical Medicine. Wiley. ISBN: 9780471860594.

MacKay, D.J.C. et al. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press. ISBN: 9780521642989.

MacKay, David JC (1992). "Bayesian interpolation". In: *Neural computation* 4.3, pp. 415–447.

Rasmussen, C.E. et al. (2006). *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning. MIT Press. ISBN: 9780262182539.

Bibliography II

Rue, Håvard et al. (2017). "Bayesian Computing with INLA: A Review". In: *Annual Review of Statistics and Its Application* 4.1, pp. 395–421. DOI: 10.1146/annurev-statistics-060116-054045.

Sousa], Delma B. [de et al. (2020). "Hot spots of leprosy in the endemic area of São Luís, Maranhão State, Northeastern Brazil". In: *Journal of Infection and Public Health* 13.2, pp. 228–234. ISSN: 1876-0341. DOI: https://doi.org/10.1016/j.jiph.2019.08.006.

Vanhatalo, Jarno (Oct. 2010). "Speeding Up the Inference in Gaussian Process Models". English. PhD thesis. Finland. ISBN: 978-952-60-3380-8.

Vanhatalo, Jarno, Pia Mäkelä, et al. (Nov. 2010). "Alkoholikuolleisuuden alueelliset erot Suomessa 2000-luvun alussa (*Regional Differences in alcohol mortality in Finland in the early 2000s*". In: *Yhteiskuntapolitiikka-YP 75 (2010) : 3*, pp. 265–273.

Vanhatalo, Jarno, Ville Pietiläinen, et al. (2010). "Approximate inference for disease mapping with sparse Gaussian processes". In: *Statistics in medicine* 29.15, pp. 1580–1607.