

## 1 Spatial data

Space is inherent to all ecological processes, influencing dynamics such as migration, dispersal, and species interactions. A primary goal in ecology is to understand how these processes shape species distributions and dynamics across space.

Spatial data are data which have any form of geographical information attached to them. The emergence of modelling frameworks for spatial ecology has been facilitated by the rise of new technologies and data collection systems like aerial photographs, GPS tracking, satellite imagery, and biologging devices.



Figure 1: GPS tags for survey and monitoring waterbirds movement

Most of environmental variables and ecological processes of interest will vary over space to some degree and we will typically use this spatial information to help us understand the relationship between our data points. Furthermore, incorporating space into statistical modelling approaches has driven the formulation of novel ecological questions and the development of new analytical methods to address them.

The overall goal of any piece of spatial analysis is to understand the spatial patterns in our data. This could involve:

- Estimating differences in mean, variance or some other summary statistic over space.
- Predicting the value at some unobserved location.
- Identifying *hotspot* with high (or low) value compared to the rest of the region.

aims also vary with different types of spatial data...

## 1.1 Types of spatial data

Spatial patterns are omnipresent in both environmental and ecological data. In general, our environmental or ecological processes of interest can be described by three main categories of spatial data :

### Discrete space:

- Data on a spatial grid (areal data)

### Continuous space:

- Geostatistical (geo-referenced) data
- Spatial point process data

The components of the models we will cover are used to reflect spatial dependence structures in discrete and continuous space.

In point processes we measure the locations where events occur (e.g. trees in a forest, earthquakes) and the coordinates of such occurrences are our data.

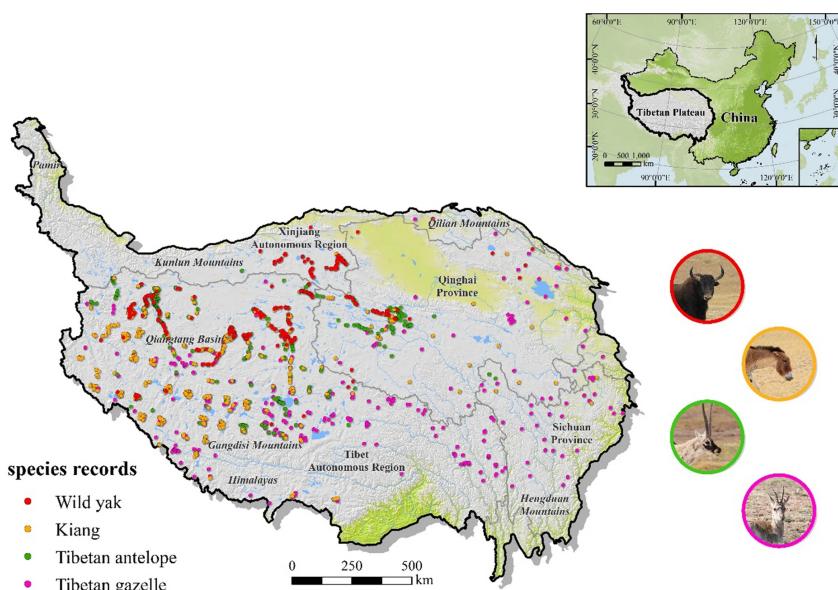


Figure 2: Occurrence records of four ungulate species in the Tibet (taken from (Liang et al. 2021)).

In geostatistical data, measurements are taken at a set of fixed locations. (e.g. air monitoring stations to quantify pollution levels, or a quadrats strategically placed to monitor wildlife).

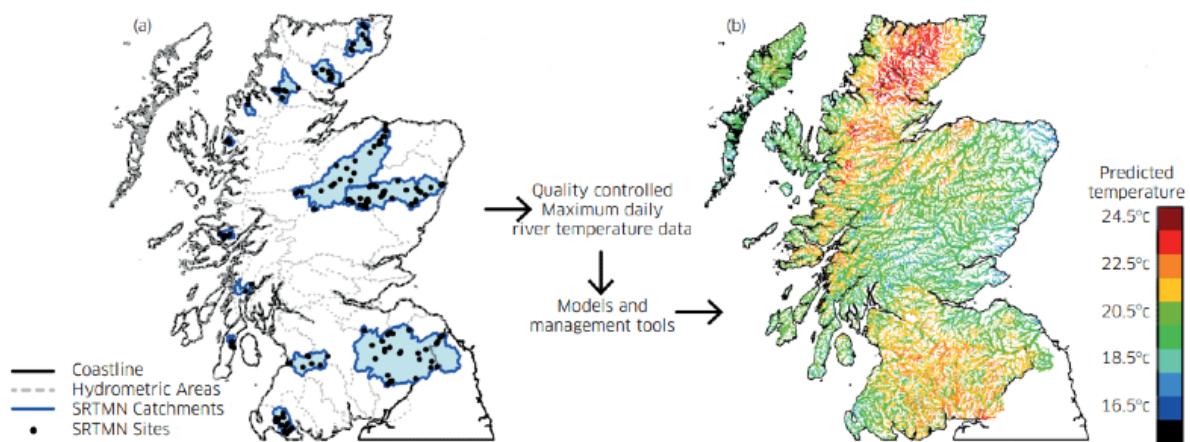


Figure 3: Scotland river temperature monitoring network

Finally, in areal data our measurements are summarised across a set of discrete, non-overlapping spatial units (e.g., delimited conservation regions, postcode areas, council regions).

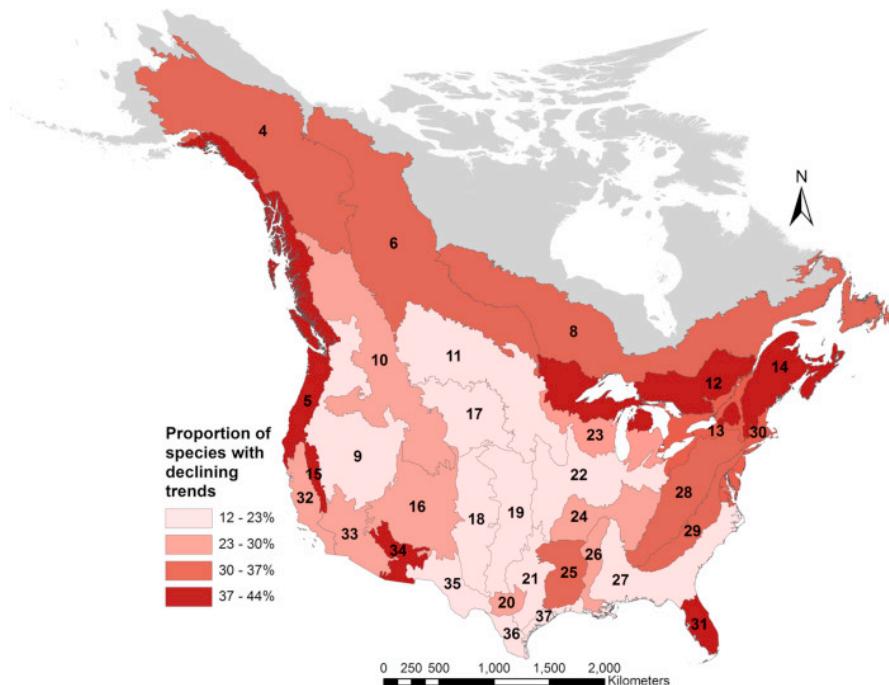


Figure 4: Map of bird conservation regions (BCRs) showing the proportion of bird species within each region showing a declining trend (Stanton et al. 2016)

## 1.2 Spatial scales

*Scales* describe the spatial and temporal dimensions at which an ecological process occurs. Many ecological patterns and processes occur at different spatial and temporal scales and thus, understanding and quantifying these scales is essential to provide an accurate interpretation of the ecological processes we aimed to study.

The spatial scale is often described by the *grain* (a.k.a. *spatial resolution*) and the *extent*. The former refers to the finest spatial unit of measurement for a given process, whereas the latter refers to the total length or area of the study. The ratio between these two measurements is known as *Scope*.

There is typically a trade-off between the grain and the extent, mainly due to practicality. For example, it can be costly to work at large extents while collecting data at fine grain sizes. But also, some of the processes occurring at these finer scales become simply noise when we look at systems at larger extents.

In spatial ecology, the scale at which an ecological process occurs can have a major impact on the interpretation of our analysis. For example, Figure 5 shows the locations of isopod burrows in the northern Negev Desert, Israel. If we look at the complete data set we can see a strong aggregation pattern. However, when focusing on two subsets with smaller extents, the observed patterns shift, ranging from random to aggregated.

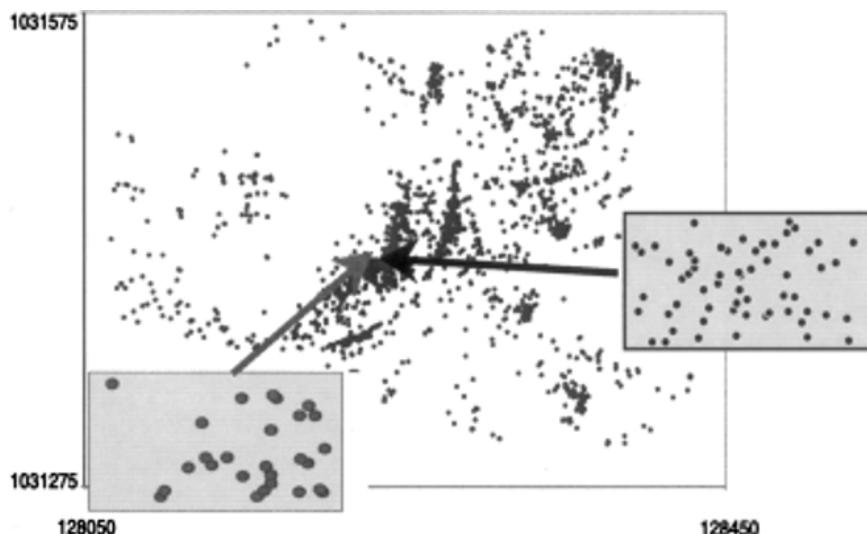


Figure 5: Isopod burrow data with  $n=2015$  individual burrows in a study area with an extent of  $75\,600\text{ m}^2$ . Inset maps show two subsets with reduced extent of  $6.4\text{ m}^2$  each.  
Source: (Dungan et al. 2002)

The spatial scale of analysis is crucial in ecological and environmental studies, as it can significantly influence the patterns we observe and the conclusions we draw. For example, changing the grain of the study (while holding the extent constant) can introduce bias and uncertainty in the observed ecological patterns. This is because aggregated data can have different properties than the sample data from which they were derived, i.e. the assumption that a relationship exists one level of aggregation does not necessarily hold at another level of aggregation, particularly in situations where data are aggregated into irregular sampling units (modifiable areal unit problem -MAUP). This is illustrated in Figure Figure 6, where two variables ( $v_1$  and  $v_2$ ) recorded on uniform point level show little correlation. As we aggregate the values on a larger grain size, we get a slight increase in the slope and  $R^2$  values. Now if we aggregate the same point data using the non-homogeneous aggregation scheme now  $v_1$  and  $v_2$  become highly correlated.

Thus, if the data we had at hand came from the non-uniform aggregation scheme, it would be wrong to assume that a correlation also exists on an individual-level data. In ecology this is known as *ecological fallacy*, which arises when incorrect inferences about individual sample units are made based on aggregated data.

For example, aggregating species abundance (counts) data at a national scale might obscure local hotspots of biodiversity observed at finer scales. Results depend on how the areas are divided, even if the total number of zones remains constant. Different arrangements of the same spatial units can lead to different conclusions. When individual-level data is grouped, the variability within units is lost, which can mask finer-scale patterns or exaggerate certain trends. As consequence, conservation strategies should consider the spatial scale for effective conservation decisions.

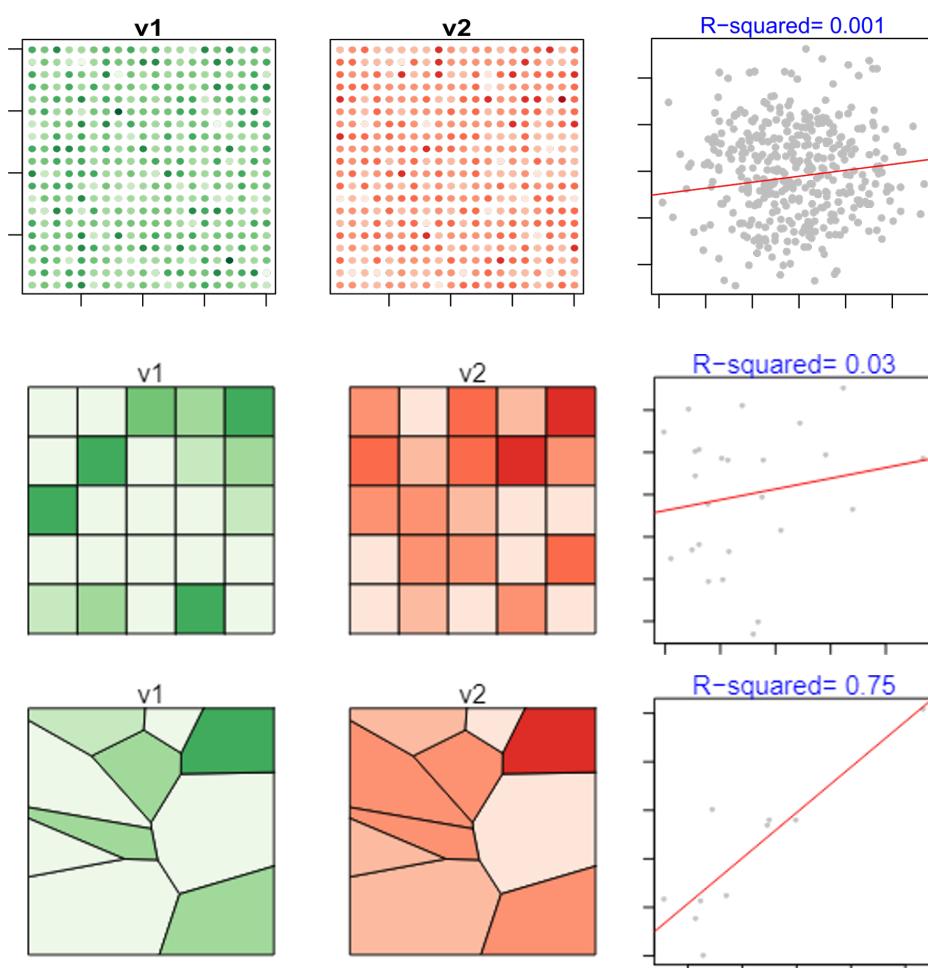


Figure 6: Plots of variables v1 and v2 for each individual in the survey and summarized using a uniform and non-uniform aggregation scheme

## 1.3 Spatial dependence

---

### Spatial modelling: why is it necessary?

Tobler's first law of geography states that:

*Everything is related to everything else, but near things are more related than distant things"*

Spatial statistics quantify spatial variance and correlation based on distance, aligning with Tobler's law—nearby measurements tend to be more correlated, while this relationship weakens with distance. Spatial dependence can reveal ecological processes like species interactions and responses to environmental gradients. However, it also poses challenges for statistical analysis, as it violates the common assumption of independent data in traditional methods.

Analyzing spatial dependence can offer insights into the biological processes shaping observed patterns, such as social behavior, resource distribution, or dispersal. While measuring spatial dependence alone may not provide definitive answers, it helps generate hypotheses and refine predictions. Additionally, spatial dependence can influence assessments of conservation threats and strategies.

Spatial dependence can arise for many reasons. It may result from **endogenous processes** inherent to the system (e.g., localized dispersal leading to organism clustering or social and grouping behaviors) or be the result of **exogenous factors** such as spatially dependent environmental gradients used by the organism of interest. Additionally, spatial dependence can stem from model *mis-specifications*, including the omission of key covariates or incorrect functional assumptions, such as failing to account for nonlinear relationships between predictors and responses.

In ecology, spatial data often come from **point samples**, where spatial relationships are captured through **x-y coordinates** (point processes) or **aerial (lattice) data**, such as counties or watersheds, where spatial dependence is modeled using a **neighborhood matrix** (spatial weights matrix). Ecological models commonly incorporate these spatial structures to account for spatial dependence in the process of interest.

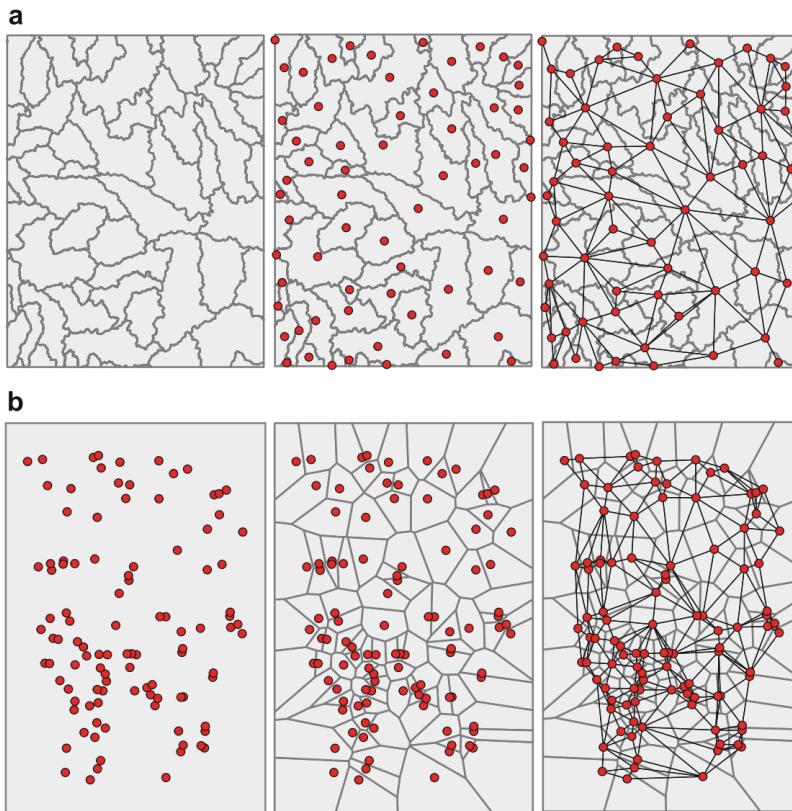


Figure 7: Examples of aerial data used in spatial modelling. Areal data can come from (a) polygon based information or can be generated (b) from point data using Voronoi tessellation. We can describe spatial dependence through the links among locations (right panel) with a spatial neighborhood (weights) matrix

Ecological models can be classified based on the spatial structure they consider. For example, *spatial filtering* methods account for spatial patterns by incorporating functions of x–y coordinates. In this approach, space is treated as a predictor variable within a regression framework—typically a flexible model such as polynomial regression or a generalized additive model (GAM), where smooth terms for x–y coordinates help capture large-scale spatial dependence within a region. Another common approach is the use of conditional autoregressive models, which are applied to areal (or lattice) data. In these models, spatial dependence is captured through a spatial neighborhood weights matrix. Before exploring these models further, let's briefly define what areal data are.

## 2 Areal Data

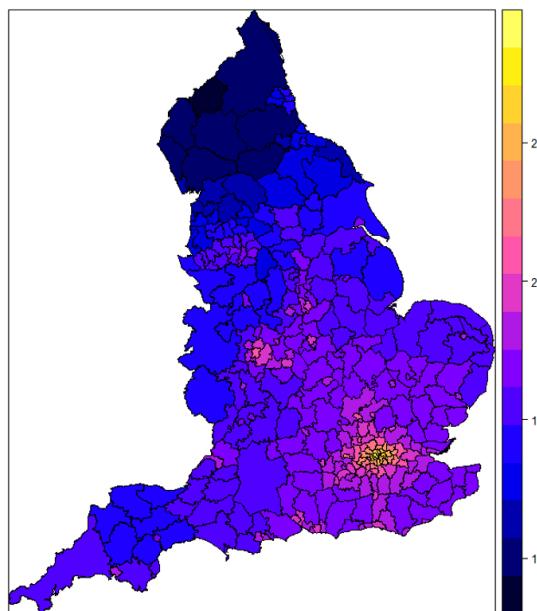
### 2.1 Overview of Areal Data

**Areal data** are data which come from well defined **geographical units** such as postcode areas, health board or pixels on a satellite image. Many public health studies use data aggregated over groups rather than data on individuals - often this is for privacy reasons, but it may also be for convenience.

As with the other types of spatial modelling, our goal is to observe and explain spatial variation in our data. Again, we are trying to understand and account for spatial dependence. Generally, we are aiming to produce a smoothed map which summarises the spatial

patterns we observe in our data. This often leads to identifying spatial extremes and/or boundaries (step changes) in the spatial surface.

Areal data are commonly used in public health settings. The image shows pollution in counties of England



Areal data is also very common in ecology where samples are taken on a lattice or become the byproduct of some sort aggregated spatial point process. For example, the map below shows the species richness (no. of different species) occur within a 10 km grid in Scotland.

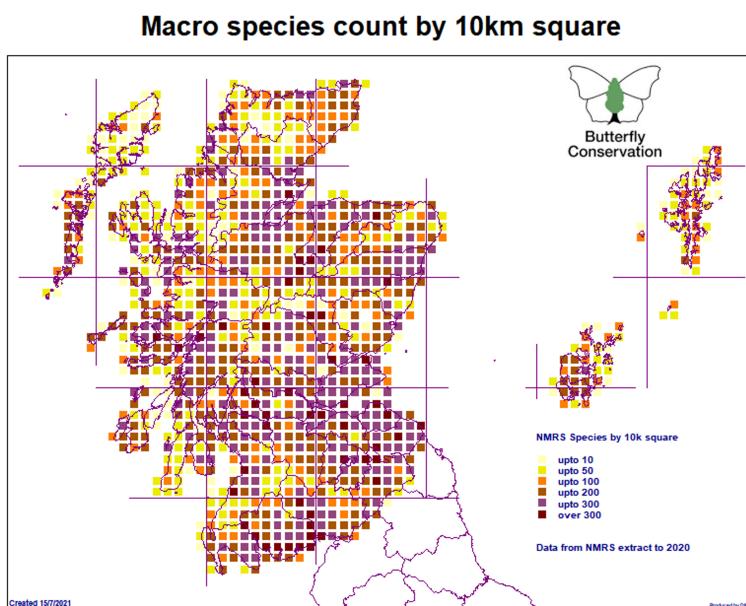


Figure 8: Species richness in Scotland in 2020 by the Butterfly conservation UKBMS

## 2.2 Definition of an Areal Process

An **areal process** (or lattice process) is a stochastic process defined on a set of regions that form a partition of our region of interest  $D$ . Let  $B_1, \dots, B_m$  be our set of  $m$  distinct regions such that:

$$\bigcup_{i=1}^m B_i = D.$$

Here we require that our regions are non-overlapping, with

$$B_i \cap B_j = \emptyset.$$

Then our areal process is simply the stochastic process

$$\{Z(B_i); i = 1, \dots, m\}.$$

### 3 Modelling spatial similarity

Imagine we have animal counts in each region. We can model them as Poisson

$$y_i \sim \text{Poisson}(e^{\eta_i})$$

How do we model the linear predictor  $\eta_i$ ?

- We could model the number of animals in each region independently

$$\eta_i = \beta_0 + \mathbf{x}'\beta + u_i \quad u_i \stackrel{iid}{\sim} N(0, \sigma_i^2)$$

- regional differences are accounted for through a *random effect*

But... what if the distribution varies across space, i.e. is structured in space? Do the covariates account for those structures? maybe some, but we cannot measure everything!

Also, note that if there's an area where the animal is rare, we'll get lots of zero counts. We would like that *Nearby regions* to have similar counts and borrow the information to improve our estimates. Thus, we could model some dependence across regions:

$$\eta_i = \beta_0 + \mathbf{x}'\beta + u_i \quad u_i \stackrel{iid}{\sim} N(0, \Sigma)$$

- Now the random effect  $u_i \sim N(0, \Sigma)$  is correlated.
- How do we do this?

#### 3.1 Sparse Matrices

---

The key lies in sparse matrices. Assume that our random effect is modelled as follows:

$$u \sim \mathcal{N}(0, \Sigma)$$

We have two options:

1. Force the *covariance* matrix  $\Sigma$  to be sparse
2. Force the *precision* matrix  $\Sigma^{-1}$  to be sparse

By forcing the *covariance* matrix  $\Sigma$  to be sparse we have that:

- $\Sigma_{ij} = \text{Cov}(u_i, u_j)$  : Covariance between  $u_i$  and  $u_j$
- $\Sigma_{ij} = 0 \rightarrow u_i$  and  $u_j$  are *independent*
- A sparse covariance matrix implies that many elements of  $\mathbf{u}$  are mutually independent.....is this desirable?

Alternatively, if we force the *precision* matrix  $\mathbf{Q} = \Sigma^{-1}$  to be sparse we might ask ourselves

- What does  $Q_{ij}$  represents?
- What does a sparse precision matrix implies?

Lets look at a simpler example.

### 3.1.1 Example: The AR(1) process

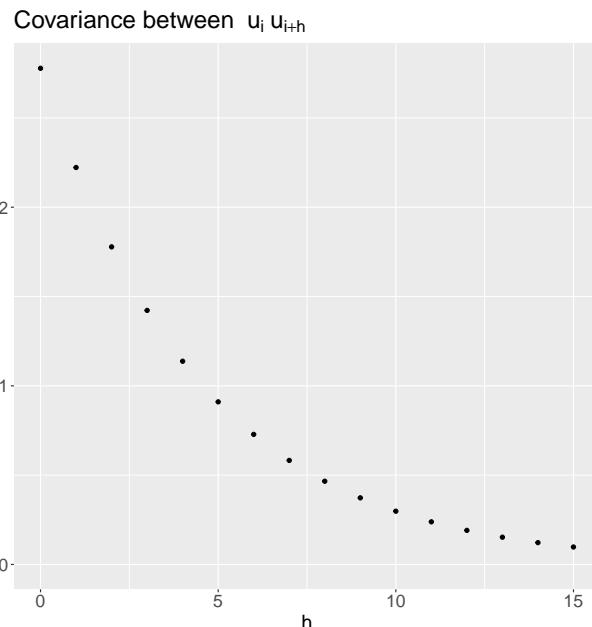
Suppose the following AR(1) process:

$$\begin{aligned}\mathbf{i = 1} : x_1 &\sim N\left(0, \frac{1}{1-\phi^2}\right) \\ \mathbf{i = 2, \dots, T} : x_i &= \phi x_{i-1} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1)\end{aligned}$$

This AR(1) is very common to model dependence in time. The *joint* distribution of  $\mathbf{x} = x_1, x_2, \dots$  is Gaussian and the **Covariance Matrix** is given by

$$\Sigma = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^N \\ \phi & 1 & \phi & \dots & \phi^{N-1} \\ \phi^2 & \phi & 1 & \dots & \phi^{N-2} \\ \dots & \dots & \dots & \dots & \dots \\ \phi^N & \phi^{N-1} & \phi^{N-2} & \dots & 1 \end{bmatrix}$$

Note that this is a *dense* matrix where all elements of the  $\mathbf{x}$  vector are *dependent*. We can visualize the covariance function for different time lags displacements  $h$  to see this:



Now lets have a look at the **Precision Matrix**

$$\mathbf{Q} = \Sigma^{-1} = \begin{bmatrix} 1 & -\phi & 0 & 0 & \dots & 0 \\ -\phi & 1 + \phi^2 & -\phi & 0 & \dots & 0 \\ 0 & -\phi & 1 - \phi^2 & -\phi & \dots & 0 \\ 0 & 0 & -\phi & 1 - \phi^2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\phi & 1 \end{bmatrix}$$

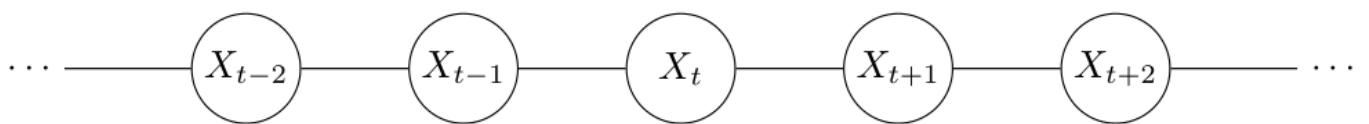
Note that:

- This is a tridiagonal matrix, it is *sparse*
- The tridiagonal form of  $\mathbf{Q}$  can be exploited for quick calculations.

What is the key property of this example that causes  $\mathbf{Q}$  to be sparse?

The key lies in the **full conditionals**

$$x_t | \mathbf{x}_{-t} \sim \mathcal{N} \left( \frac{\phi}{1 - \phi^2} (x_{t-1} + x_{t+1}), \frac{1}{1 + \phi^2} \right)$$

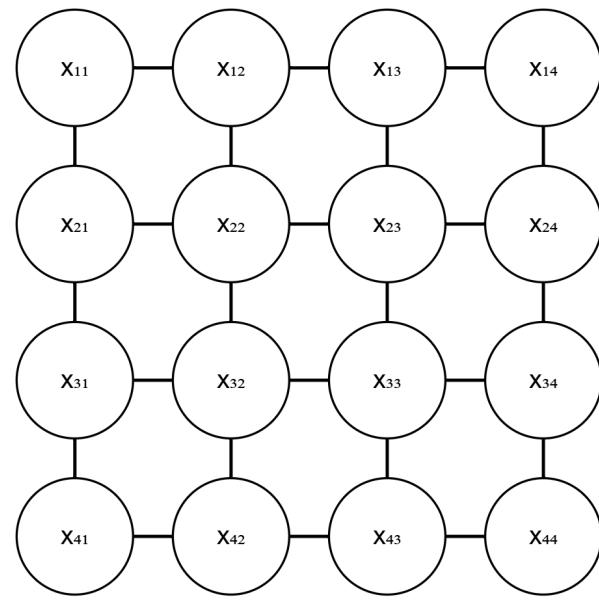


Here:

- The circles represent the values of  $x$  at individual time points
- There is a line between them if they are conditionally dependent (e.g.,  $x_t$  is conditionally independent of  $x_{t-2}$  given  $x_{t-1}$ )
- Each timepoint is only **conditionally dependent** on the two closest neighbours
- The nonzero pattern in the precision matrix is given by the neighborhood structure of the process

### 3.1.2 Markov models in space

We can extend the previous idea into space. For example, on the following first order conditional autoregressive model or a CAR(1) model every node is conditionally dependent on its four nearest neighbours



Models based on neighbourhood have a name in statistics: they are **Markovian models**.  
Markovian models are specified entirely through “neighbourhood structures”

Recall our first Model:

$$y_i \sim \text{Poisson}(e^{\eta_i}) \\ \eta_i = \beta_0 + \mathbf{x}'\beta + u_i \quad u_i \sim N(0, Q^{-1})$$

This mean  $u_i$  is independent of all the other parameters  $\mathbf{u}_{-i}$ , given the set of its neighbors.  
I.e., for any pair of elements  $(i, j)$  in  $\mathbf{u}$ ,  $u_i \perp u_j | \mathbf{u}_{-ij} \iff Q_{ij} = 0$ . In other words,  
 $Q_{ij} \neq 0$  only if  $j \in \{i, \mathcal{N}(i)\}$ .

## 3.2 Gaussian Markov random field (GMRF)

---

An informal definition of a Gauss Markov random field (GMRF) is

- a Gaussian distribution where the non-zero elements of the precision matrix are defined by a neighbourhood matrix (or graph structure)
- each region conditionally has a Gaussian distribution with
- mean equal to the average of the neighbours and
- precision proportional to the number of neighbours

A more formal definition goes as follows:

Let  $\mathbf{u}$  be a GMRF wrt  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ .

$\mathcal{V}$  Vertices:  $1, 2, \dots, n$ .  $\mathcal{E}$  Edges  $\{i, j\}$

- No edge between  $i$  and  $j$  if  $u_i \perp u_j | \mathbf{u}_{ij}$ .
- An edge between  $i$  and  $j$  if  $u_i \not\perp u_j | \mathbf{u}_{ij}$ .

*Key point: A graph defines the sparsity structure of  $Q$*

### Definition

A random variable  $\mathbf{u}$  is said to be a GMRF wrt to the graph  $\mathcal{G}$  with vertices  $\{1, 2, \dots, n\}$  and edges  $\mathcal{E}$ , with mean  $\mu$  and precision matrix  $\mathbf{Q}$  if its probability distribution is given by

$$\mathbf{u} \sim N(\mu, Q^{-1})$$

and

$$Q_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \quad \forall i \neq j$$

### 3.3 ICAR Model

An example of a GMRF is the the Besag model a.k.a. Intrinsic Conditional Autoregressive (ICAR) model. The conditional distribution for  $u_i$  is

$$u_i | \mathbf{u}_{-i}, \tau_u, \sim N \left( \frac{1}{d_i} \sum_{j \sim i} u_j, \frac{1}{d_i \tau_u} \right)$$

- $\mathbf{u}_{-i} = (u_i, \dots, u_{i-1}, u_{i+1}, \dots, u_n)^T$
- $\tau_u$  is the precision parameter (inverse variance).
- $d_i$  is the number of neighbours
- The mean of  $u_i$  is equivalent to the the mean of the effects over all neighbours, and the precision is proportional to the number of neighbours (e.g., if an area has many neighbors then its variance will be smaller)

Then, the joint distribution is given by:

$$\mathbf{u} | \tau_u \sim N \left( 0, \frac{1}{\tau_u} Q^{-1} \right),$$

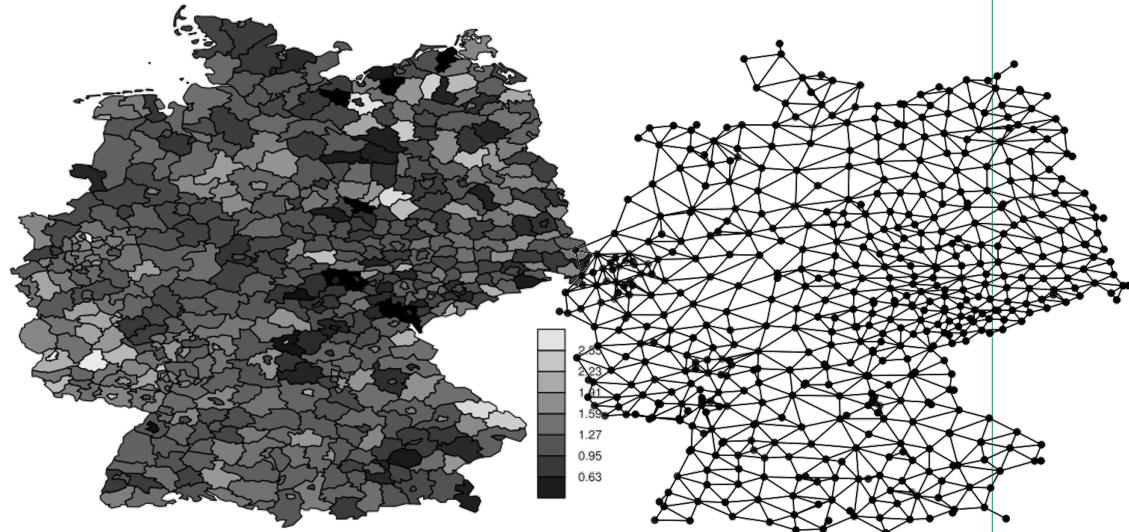
Where  $Q$  denotes the precision matrix defined as

$$Q_{i,j} = \begin{cases} d_i, & i = j \\ -1, & i \sim j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

This structure matrix directly defines the neighbourhood structure and is sparse.

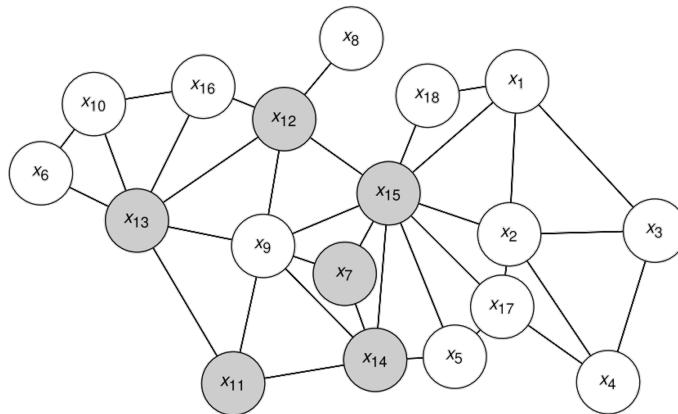
#### Example

Lets look a Larynx cancer relative risk map and how by connecting all the neighbouring areas we get a graph structure graph

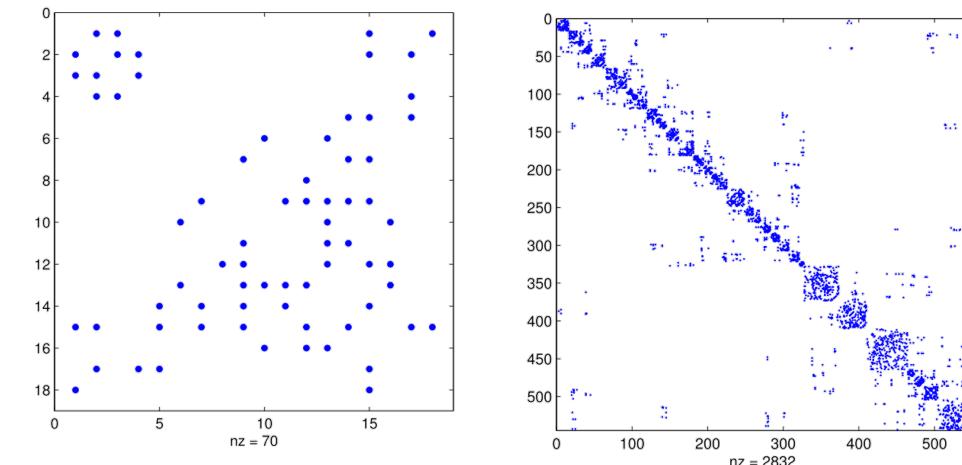


Let's focus on one small part of the graph and apply an ICAR model where each region conditionally has a Gaussian distribution with mean equal to the average of the neighbours and a precision proportional to the number of neighbour

$$x_9 \mid \mathbf{x}_{-9} \sim N \left( \frac{1}{6}(x_7 + x_{11} + x_{12} + x_{13} + x_{14} + x_{15}), \frac{1}{6\tau} \right)$$



The sub graph leads to a precision matrix with 21.6% non-zero elements while the full graph leads to a precision matrix with 0.1% non-zero elements.



The question now is how do we do choose the neighborhood structure?

### 3.4 Neighbourhood structures

Each of our regions  $B_i$  has a set of other nearby which can be considered **neighbours**. We might expect that areas have more in common with their neighbours. Therefore we can construct dependence structures based on the principle that neighbours are correlated and non-neighbours are uncorrelated. However, we need to come up with a sensible way of defining what a neighbour is in this context.

There are many different ways to define a region's neighbours. The most common ones fall into three main categories - those based on borders, and those based on distance. The methods based on **common borders** assume that regions which share a border on a map are neighbours (Figure 9). Methods based on **distance** assume that regions which are within a certain distance of each other are neighbours (Figure 10).

The common border approach is simple and easy to implement. However, it treats all borders the same, regardless of length, which can be unrealistic. Also means areas very close together are not neighbours if there is even a small gap between them.

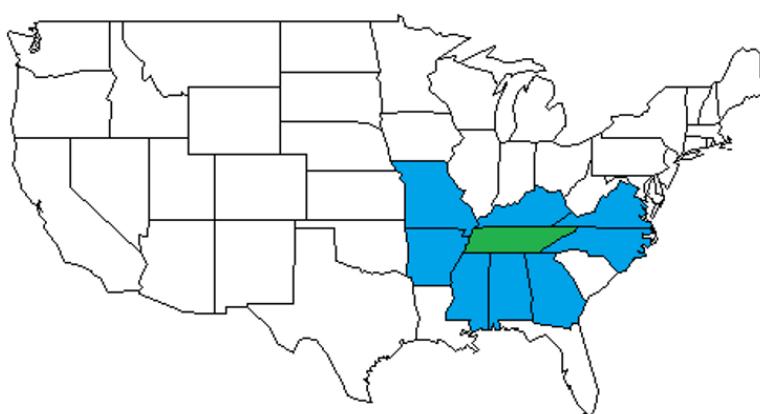


Figure 9: Illustration of border-based neighbours.

A distance-based approach may initially seem more sensible as a concept, but there are a number of challenges. What distance do you choose? How do you decide that? There is no easy answer. Where do you measure from? The value will be different depending on whether you use nearest border or central point.

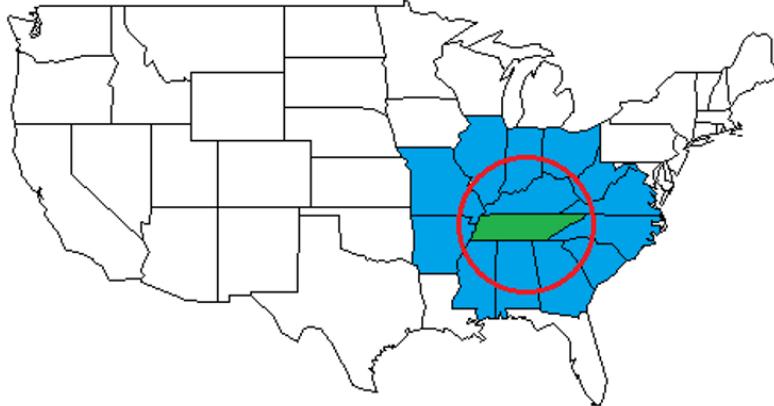


Figure 10: Illustration of distance-based neighbours.

Once we have identified a set of neighbours using our chosen method, we can use this to account for correlation. We construct a **neighbourhood matrix** (or proximity matrix), which defines how each of our  $m$  regions relate to each other. Let  $W$  denote an  $m \times m$  matrix where the  $(i, j)$ th entry,  $w_{ij}$  denotes the proximity between regions  $B_i$  and  $B_j$ . The values of this matrix can be discrete (which regions are neighbours) or continuous (how far apart are the regions).

By far the most common approach is to use a binary neighbourhood matrix,  $W$ , denoted by

$$w_{ij} = \begin{cases} 1 & \text{if areas } (B_i, B_j) \text{ are neighbours.} \\ 0 & \text{otherwise.} \end{cases}$$

Binary matrices are used for their simplicity. Fitting spatial models often requires us to invert  $W$ , and this is less computationally intensive for sparse matrices.

$$\begin{matrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{matrix}$$

Now that we have defined a measure of spatial proximity for areal data, we can use this to assess spatial dependence. Essentially, we can now ask the question of how similar a region is to its neighbours. We can consider *global* correlation, measured across the entire region, and also *local* correlation which allows for regional variation. In this course, we will focus on modelling global autocorrelation using **Moran's I**.

### 3.5 Moran's I

---

Moran's  $I$  is a measure of global spatial autocorrelation, and can be considered an extension of the Pearson correlation coefficient. For a set of data  $Z_1, \dots, Z_m$  measured on regions  $B_1, \dots, B_m$ , with neighbourhood matrix  $W$ , we can compute Moran's  $I$  as:

$$I = \frac{m}{\sum_{i=1}^m \sum_{j=1}^m w_{ij}} \frac{\sum_{i=1}^m \sum_{j=1}^m w_{ij}(Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_{i=1}^m (Z_i - \bar{Z})^2}$$

This is basically a function of differences in values between neighbouring areas.

Moran's  $I$  ranges between -1 and 1, and can be interpreted in a similar way to a standard correlation coefficient.

- $I = 1$  implies that we have **perfect spatial correlation**.
- $I = 0$  implies that we have **complete spatial randomness**.
- $I = -1$  implies that we have **perfect dispersion** (negative correlation).

Our observed  $I$  is a point estimate, and we may also wish to assess whether it is significantly different from zero. We can test for a statistically significant spatial correlation using a permutation test, with hypotheses:

$$\begin{aligned} H_0 &: \text{no spatial association } (I = 0) \\ H_1 &: \text{some spatial association } (I \neq 0) \end{aligned}$$

We carry out  $k$  random permutations of our data (reassign each data value to a random location) and compute a Moran's  $I$  for each permutation ( $I_{perm} = I_1, \dots, I_k$ ). We reject the null hypothesis if our observed Moran's  $I$  for our real data ( $I_{obs}$ ) could not have plausibly come from the distribution of  $I_{perm}$ .

The plots below show perfect dispersion, complete spatial randomness and high spatial correlation respectively.

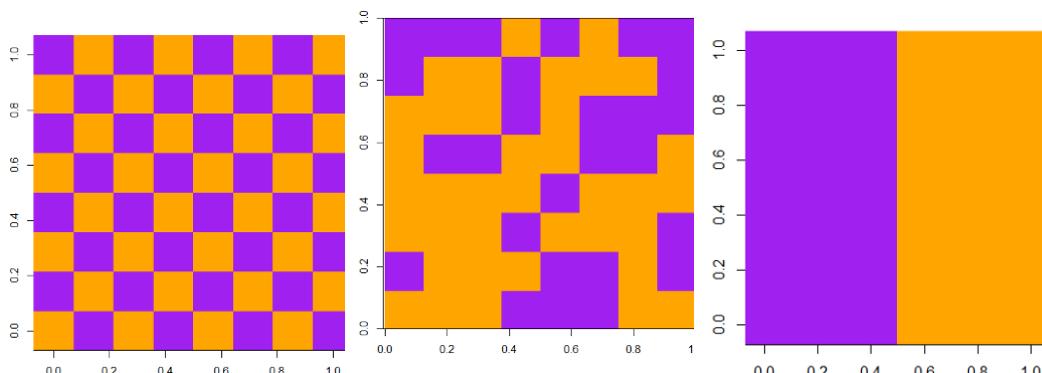


Figure 11:  $\hat{I} = 1$ ; Std.Dev 0.095; pval < 0.05

Figure 12:  $\hat{I} = 0.001$ ; Std.Dev 0.095; pval = 0.865

Figure 13:  $\hat{I} = 865$ ; Std.Dev 0.095; pval < 0.05

Illustration of Moran's  $I$  using simulated data.

## 4 Model fitting

Suppose we identified spatial autocorrelation in our data... what do we do next?

Fitting autoregressive models to non-normal data—such as presence/absence responses or counts data can be challenging due to the lack of simple distributional assumptions. For that reason, CAR models are often implemented within a Bayesian framework, which

provides a flexible approach for handling spatially structured ecological & environmental data.

- A key advantage of Bayesian hierarchical modelling is its ability to properly incorporate uncertainty and facilitate inference in complex spatial models.

## 4.1 Bayesian Inference

---

In the Bayesian framework all unknown quantities in the model are treated as random variables, and the aim is to estimate the **joint posterior distribution** of the unknown parameters  $\theta$  given the observed data  $\mathbf{y}$ .

- We obtain this distribution through Bayes' theorem:

$$\pi(\theta \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid \theta)\pi(\theta)}{\pi(\mathbf{y})}$$

- Components:

- $\pi(\mathbf{y} \mid \theta)$  is the **likelihood** of  $\mathbf{y}$  given parameters  $\theta$
- $\pi(\theta)$  is the **prior distribution** of the parameters
- $\pi(\mathbf{y})$  is the **marginal likelihood** (normalizing constant):

$$\pi(\mathbf{y}) = \int_{\Theta} \pi(\mathbf{y} \mid \theta)\pi(\theta)d\theta$$

- marginalizing  $\pi(\mathbf{y})$  means *integrating* out all the uncertainty on  $\theta$

When the posterior distribution lacks a closed-form solution, **computational methods** must be used to approximate it. To efficiently estimate spatially structured spatial models, particularly those incorporating **Conditional Autoregressive (CAR) priors**, the **Integrated Nested Laplace Approximation (INLA)** framework (Van Niekerk et al., 2023) provides a powerful alternative to traditional **Markov chain Monte Carlo (MCMC)** methods.

## 4.2 INLA

---

We can use INLA to approximate the joint posterior of our parameters of interest. INLA is a fast and accurate method to do Bayesian inference with **latent Gaussian models** and can be used with Bayesian hierarchical models where we model in different stages or levels:

- **Stage 1:** What is the distribution of the responses?
- **Stage 2:** What is the distribution of the underlying latent components?
- **Stage 3:** What are our prior beliefs about the hyperparameters?

## Stage 1: Data Generating Process

On Stage 1, we look at how our data ( $\mathbf{y}$ ) is generated from the underlying components  $\mathbf{x}$  and hyperparameters  $\theta$  in the model

### Response types:

- Gaussian (temperature, rainfall, weight)
- Count data (disease cases, species counts)
- Point patterns (tree locations)
- Binary data (yes/no responses)
- Survival data (time to event)

It is also important how data are collected!

All of this information is placed into our likelihood  $\pi(\mathbf{y}|\theta)$ . We assume that given the underlying components ( $\mathbf{x}$ ) and hyperparameter ( $\theta$ ), the data are independent on each other

$$\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i \in \mathcal{I}} \pi(y_i | \mathbf{x}_{\mathcal{I}}, \theta)$$

This implies that all the dependence structure in the data is explained on Stage 2.

## Stage 2: Dependence Structure

The underlying unobserved components  $\mathbf{x}$  are called latent components and can be:

- Fixed effects of covariates
- Unstructured random effects (individual effects, group effects)
- Structured random effects (AR(1), regional effects, ... )

These are linked to the responses in the likelihood through linear predictors.

An important feature of INLA is that it considers the latent field  $\mathbf{x}$  to be a GMRF

$$\pi(\mathbf{x}|\theta) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} \right\}$$

Thus, the precision matrix  $\mathbf{Q}$  is sparse. We can also factorize the likelihood as:  $\pi(\mathbf{y}|\mathbf{x}, \theta) = \prod_i \pi(y_i | \eta_i, \theta)$ . This means that our data are conditional independent given  $\mathbf{u}$  and  $\theta$ . Also, that each data point depends on only 1 element of the latent field: the *predictor*  $\eta_i$ , where  $\eta$  is a *linear* combination of other elements of  $\mathbf{x}$ :  $\eta = \mathbf{A}^T \mathbf{x}$ .

Lastly, we assume that the vector of hyperparameters  $\theta$  is low dimensional. See how both, the likelihood and the latent model, typically have hyperparameters that control their behavior. The hyperparameters  $\theta$  can include:

**Examples likelihood:**

- Variance of observation noise
- Dispersion parameter in the negative binomial model -
- Probability of a zero (zero-inflated models)

**Examples latent model:**

- Variance of unstructured effects
- Correlation of multivariate effects
- Range and variance of spatial effects
- Autocorrelation parameter  $\rho$

## 4.3 LGMs and the INLA strategy in short

---

These three stages constitute the basis of latent Gaussian models (LGMs).

- A LGM consists of three elements:
  1. a likelihood model
  2. a latent Gaussian field (i.e., the latent components of our model)
  3. a vector of non-Gaussian hyperparameters
- The characteristic property is that the latent part of the hierarchical model is Gaussian,  $\mathbf{x}|\theta \sim N(0, Q^{-1})$ , where the expected value is 0 and the precision matrix is  $Q$ .

**Assumption**

We are mainly interested in posterior marginals  $p(x_i|\mathbf{y})$  and  $p(\theta_j|\mathbf{y})$

**Strategy**

- We use numerical integration in a smart way
- We approximate what we do not know analytically exploiting the Gaussian structure of  $\mathbf{x}|y$ .

## 5 Case Study: Lip cancer rates in Scotland

In this example we model the number of lip cancer rates in Scotland in the years 1975–1980 at the county level in order to evaluate the presence of an association between sun exposure and lip cancer.

## 5.1 Standardized Mortality Ratios and spatial correlation

---

In epidemiology, disease risk is assessed using Standardized Mortality Ratios (SMR):

$$SMR_i = \frac{Y_i}{E_i}$$

- A value  $SMR > 1$  indicates a high risk area.
- A value  $SMR < 1$  suggests a low risk area.
- SIRs may be misleading in counties with small populations.
- model-based approaches enable to incorporate covariates and borrow information from neighboring counties to improve local estimates,

The data is available on the `SpatialEpi` R package and can be loaded and visualized as follows:

```
# Library where the data is stored
library(SpatialEpi)

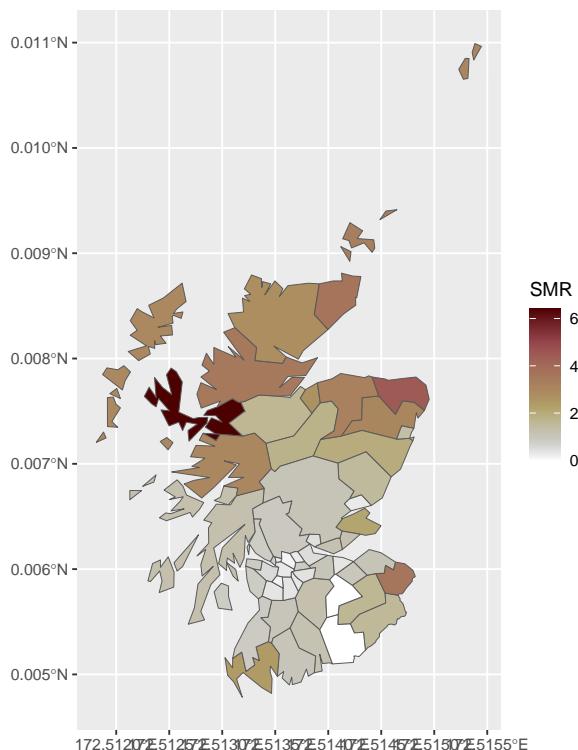
# Libraries for Data manipulation
library(tidyr)
library(dplyr)

# Libraries for producing maps
library(ggplot2)
library(sf)
library(mapview)

# Load the Data
data(scotland_sf)

# Compute the SMR and add a region index (for later modelling)
scotland_sf <- scotland_sf %>% mutate(
  SMR = cases/expected,
  region_id = 1:nrow(scotland_sf))

# Visualize the regions colored by the SMR
ggplot() + geom_sf(data=scotland_sf, aes(fill=SMR)) + scale_fill_scico(direction = -1)
```



### Do we need to model spatial dependence?

Recall that, Moran's  $I$  ranges between -1 and 1, and can be interpreted in a similar way to a standard correlation coefficient.

- $I = 1$  implies that we have **perfect spatial correlation**.
- $I = 0$  implies that we have **complete spatial randomness**.
- $I = -1$  implies that we have **perfect dispersion** (negative correlation).

Our observed  $I$  is a point estimate, and we may also wish to assess whether it is significantly different from zero. We can test for a statistically significant spatial correlation using a permutation test, with hypotheses:

$$H_0 : \text{negative or no spatial association } (I \leq 0)$$

$$H_1 : \text{positive spatial association } (I > 0)$$

We can use `moran.test()` to test this hypothesis by setting `alternative = "greater"`. To do so, we need to supply `list` containing the neighbors via the `nb2listw()` function from the `spdep` package. Lets assess now the spatial autocorrelation of the SMR in 2011:

```
# list of neighbors based on areas with contiguous boundaries
W.nb <- poly2nb(scotland_sf, queen = TRUE)

# Create proximity matrix
R <- nb2listw(W.nb, style = "B", zero.policy = TRUE)

# Global Moran's I
gmoran <- moran.test(scotland_sf$SMR, R,
                      alternative = "greater")
```

Moran I test under randomisation

```
data: scotland_sf$SMR
weights: R
n reduced by no-neighbour observations

Moran I statistic standard deviate = 5.3678, p-value = 3.984e-08
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.437119239     -0.019230769     0.007227652
```

Since we have set the alternative hypothesis to be  $I > 0$  and have a  $p$ -value  $< 0.05$ , we then reject the null hypothesis and conclude there is evidence for positive spatial autocorrelation.

## 5.2 BYM model

---

The ICAR model introduced earlier (#sec-besag), accounts only for spatially structured variability and does not include a limiting case where no spatial structure is present. Thus, we typically add an unstructured random effect  $z_i \mid \tau_z \sim N(0, \tau_z^{-1})$

- The resulting model  $v_i = u_i + z_i$  is known as the Besag-York-Mollie model (BYM)
- The structured spatial effect is controlled by  $\tau_u$  which control the degree of smoothing:
  - Higher  $\tau_u$  values lead to stronger smoothing (less spatial variability).
  - Lower  $\tau_u$  values allow for greater local variation.

This is an example of a LGM expressed as follows:

- **Stage 1:** We assume the responses are Poisson distributed:

$$\begin{aligned} y_i | \eta_i &\sim \text{Poisson}(E_i \lambda_i) \\ \log(\lambda_i) &= \eta_i = \beta_0 + \beta_1 \text{AFF} + u_i + z_i \end{aligned}$$

- **Stage 2:**  $\eta_i$  is a linear function of **four components**: an intercept, proportion of population engaged in agriculture, fishing, or forestry (AFF) effect, a spatially structured effect  $u$  and an unstructured iid random effect  $z$ :

$$\eta_i = \beta_0 + \beta_1 \text{AFF}_i + u_i + z_i$$

- **Stage 3:**  $\{\tau_z, \tau_u\}$ : Precision parameters for the random effects

The latent field is  $\mathbf{x} = (\beta_0, \beta_1, u_1, u_2, \dots, u_n, z_1, \dots)$ , the hyperparameters are  $\theta = (\tau_u, \tau_z)$ , and must be given a prior.

### The Model structure

$$\begin{aligned} y_i | \eta_t &\sim \text{Poisson}(E_i \lambda_i) \\ \log(\lambda_i) &= \eta_i = \beta_0 + \beta_1 \text{AFF}_i + u_i + v_i \end{aligned}$$

- $\beta_0$  is the intercept that represents the overall risk
- $\beta_1$  is the coefficient of the AFF covariate
- $u_i$  is a spatial structured component
- $v_i$  is a spatial unstructured component

### The neighbourhood structure

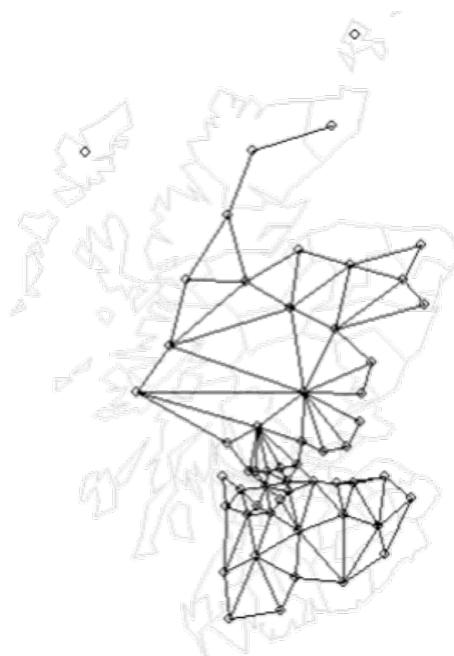
Here we generate the adjacency matrix to represent the structure of the neighbours using the `poly2nb` function from the `spdep` R package.

```
library(spdep)

W.nb <- poly2nb(scotland_sf, queen = TRUE)

plot(st_geometry(scotland_sf), border = "lightgray")

plot.nb(W.nb, st_geometry(scotland_sf), add = TRUE)
```



Then we define the proximity matrix using the `nb2mat` function and define the precision matrix  $Q$  according to Equation 1:

Recall The precision matrix  $Q$  depends on the neighboring structure and  $\tau_u$  (which will be estimated). Now, to make the precision parameters of models with different intrinsic Gaussian random field comparable we add a sum-to-zero constrain  $\sum_i^n u_i = 0$  (see `scale.model = TRUE` in the code below).

We can fit this model with `inlabru`, which is a wrapper around INLA (we will cover this more in detail on the labs). We just need to define

1. The model components (`cmp`)
2. The formula for the linear predictor
3. The observational model (via the `bru_obs` function)

# INLA Model Results

## Posterior summaries of fixed effects and hyperparameters

Parameter	Mean	2.5% Quantile	97.5% Quantile
Intercept	-0.306	-0.538	-0.069
beta_1	4.317	1.758	6.761
Precision for u_i	4.139	2.023	7.597
Precision for z_i	22,038.015	1,474.185	86,086.909

4. fit the model using the bru function

```
cmp = ~ Intercept(1) + beta_1(AFF, model = "linear") +
      u_i(region_id, model = "besag", graph = Q, scale.model = TRUE) +
      z_i(region_id, model = "iid")

formula = cases ~ Intercept + beta_1 + u_i + z_i

lik = bru_obs(formula = formula,
              family = "poisson",
              E = expected,
              data = scotland_sf)

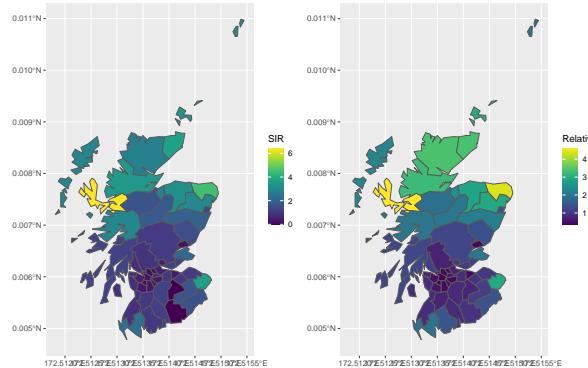
fit = bru(cmp, lik)
```

The posterior summaries for fixed effect and hyperparameters can be obtain by typing `fit$summary.fixed` and `fit$summary.hyperpar`. The following table summarises the results obtained from the model:

The model revealed a strong positive association between sun exposure and lip cancer risk. The relative risks  $\lambda_i$  can be obtained as follows:

$$\begin{aligned}\lambda_i &= \exp(\eta_i) \\ &= \exp\left(\beta_0 + \beta_1 \times \text{AFF}_i + u_i + z_i\right)\end{aligned}$$

We can then compare a map of the RR of each area against the SMR we computed previously:



We see RR values are shrunk towards 1 compared to the SMR values because each area's spatial effect is pulled toward its neighbors' average.

### 5.3 BYM2 Model & PC priors

In the original BYM, the spatially structured component must be scaled so that  $\tau_u$  produces consistent smoothness across different neighborhood structures.

Simpson *et al.* (2017) proposed a new parametrization of the BYM model that improves parameter interpretability. The BYM2 model proposed by Simpson *et al.* (2017) uses a scaled spatially structured effect  $u^*$  and unstructured random effect  $v^*$ :

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} (\sqrt{1-\phi} v^* + \sqrt{\phi} u^*).$$

Here, the precision  $\tau_b > 0$  controls the marginal variance contribution of the weighted sum  $u^*$  and  $v^*$ . The mixing parameter  $0 \leq \phi \leq 1$  measures the proportion of the marginal variance explained by the structured effect  $u^*$ . Thus, if  $\phi = 1$  the model captures only spatially structured variability, whereas when  $\phi = 0$ , it accounts solely for unstructured spatial noise.

This parametrization of the BYM2 model also allows the specification of **Penalized Complexity (PC) priors**, which provide an intuitive way to control the amount of spatial smoothing and avoid overfitting. The key idea behind PC priors is to shrink the model towards a simpler baseline (e.g., a non-spatial model) unless the data provide strong evidence for a more complex structure. To define the prior for the marginal precision  $\tau_b$  and the mixing parameter  $\phi$ , we use the probability statements  $P(1/\sqrt{\tau_b} > U) = \alpha$  and  $P(\phi < U) = \alpha$  respectively.

If we consider a marginal standard deviation of approximately 0.5 is a reasonable upper bound, we can use the rule of thumb described by Simpson *et al.* (2017) and set  $U = 0.5/0.31$  and  $\alpha = 0.01$ , this is equivalent to the probability statement  $P((1/\sqrt{\tau_b}) > 0.5/0.31) = 0.01$ . Then, the PC prior for the mixing parameter can be specified as  $P(\phi < 0.5) = 2/3$ . The later statement assumes that the unstructured random effect accounts for more of the variability than the spatially structured effect.

## 6 Summary of points

- Space is inherent to many ecological and environmental phenomena
- We can distinguish between three main **types of spatial data**:
  - Areal data - Data summarized across discrete non-overlapping regions
  - Geostatistical - Measurements of a continuous process measured at fixed location
  - Point process data - locations where events occur as realization of a continuous process
- The **spatial scale** defined by the *grain* and *extent* determine the conclusion we can draw from our analysis (MAUP and ecological fallacy).
- Spatial data are often not independent → spatial dependency can complicate statistical analysis.
- Incorporate spatial dependence through sparse precision matrices
- Spatial dependence can be incorporated through Gaussian Markov random fields (GMRF)
- GMRF conditionally dependent structures enforces sparsity into the precision matrix.
- Besag mode ( or ICAR) is a GMRF used to model spatial dependence in areal data based on a neighborhood structure.
- Different way of defining the neighborhood structure and matrices.
- CAR models can be fitted using *Bayesian methods*
- **INLA** is an efficient Bayesian efficient approximation methods for LGMs.
- **LGMs** are defined based on:
  - Model likelihood  $p(y_i | u_i \theta)$
  - Latent GMRF  $p(\mathbf{u} | \theta)$ .
  - Vector of hyper parameter  $p(\theta)$
- Use **Moran's I** to assess spatial dependence.
- Use BYM to extend ICAR model to account for *unstructured spatial variation*.
- BYM2 model provides further flexibility and interpretability
- PC priors enable intuitive construction of priors based on probability statements.

Dungan, J. L., J. N. Perry, M. R. T. Dale, P. Legendre, S. Citron-Pousty, M.-J. Fortin, A. Jakomulska, M. Miriti, and M. S. Rosenberg. 2002. "A Balanced View of Scale in Spatial Statistical Analysis." *Ecography* 25 (5): 626–40. <https://doi.org/10.1034/j.1600-0587.2002.250510.x>.

Liang, Jianchao, Zhifeng Ding, Zhigang Jiang, Xiaojun Yang, Rongbo Xiao, Paras Bikram Singh, Yiming Hu, Keji Guo, Zhixiang Zhang, and Huijian Hu. 2021. "Climate Change, Habitat Connectivity, and Conservation Gaps: A Case Study of Four Ungulate Species Endemic to the Tibetan Plateau." *Landscape Ecology* 36 (4): 1071–87. <https://doi.org/10.1007/s10980-021-01202-0>.

Stanton, Jessica C., Brice X. Semmens, Patrick C. McKann, Tom Will, and Wayne E. Thogmartin. 2016. "Flexible Risk Metrics for Identifying and Monitoring Conservation-Priority Species." *Ecological Indicators* 61 (February): 683–92. <https://doi.org/10.1016/j.ecolind.2015.10.020>.