Day 1 – Spatial & Spatio-temporal Modelling

Olatunji Johnson

Introduction

About me

- Completed a Bachelor's degree in Statistics at FUTA
- Master's in Mathamatical Sciences at AIMS-Tanzania
- PhD in Statistics and Epidemiology at University of Lancaster
- Postdoc at University of Manchester
- Lecturer in Statistics at the University of Manchester

Overview of the 3 days

- Spatial and spatio-temporal analysis (different likelihoods)
- Joint modelling of multiple malaria processes
- Non-stationary spatial processes
- Hybrid machine learning + geostatistical models

Linear Regression

- Goal: Model a continuous response variable as a linear function of predictors.
- Model $Y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$
- Key Assumption
  - Linearity
  - Independence
  - Homoscedasticity (constant variance)
  - Normality of errors

Generalized Linear Models (GLMs)

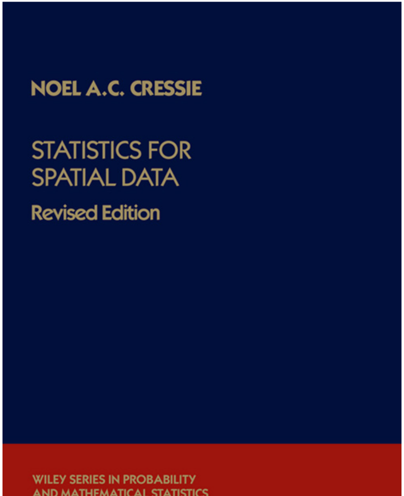
- Extension of linear models to handle non-normal response distributions.
- Three components:
  - Random component: Distribution from the exponential family (e.g., Binomial, Poisson).
  - Systematic component: Linear predictor $\eta = X\beta$.
  - Link function: Relates $E(Y)$ to η.
- Examples:
  - Logistic regression for binary outcomes – `logit link function`
  - Poisson regression for count data – `log link function`

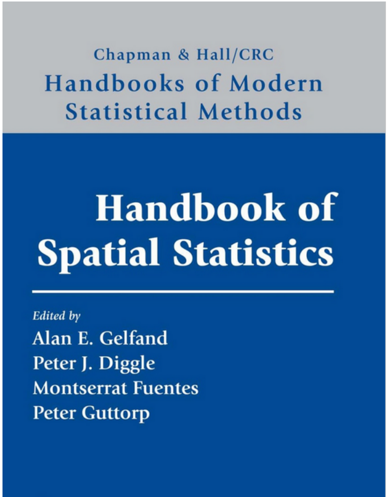
Why Spatial Statistics?

- **Spatial Dependence:**
  Observations collected at nearby locations are often more similar than those farther apart.
- **Ignoring Spatial Structure:**
  - Leads to biased parameter estimates.
  - Underestimates uncertainty.
  - Misses important spatial patterns.
- **Applications:**
  - Disease mapping
  - Environmental monitoring
  - Agricultural field trials
- **Goal:**
  Account for spatial correlation to improve prediction and inference.

Spatial Analysis

Spatial Statistics





Classification of spatial statistics

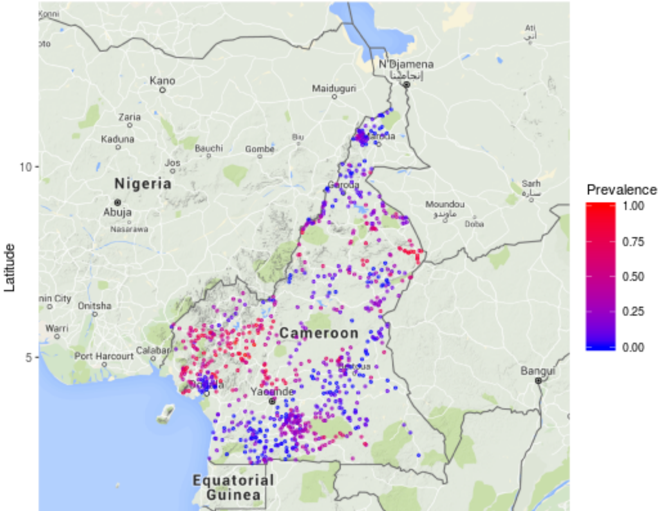
Cressie's book classifies spatial statistics according to **data format**:

1. Geostatistical data
2. Lattice data
3. Point patterns

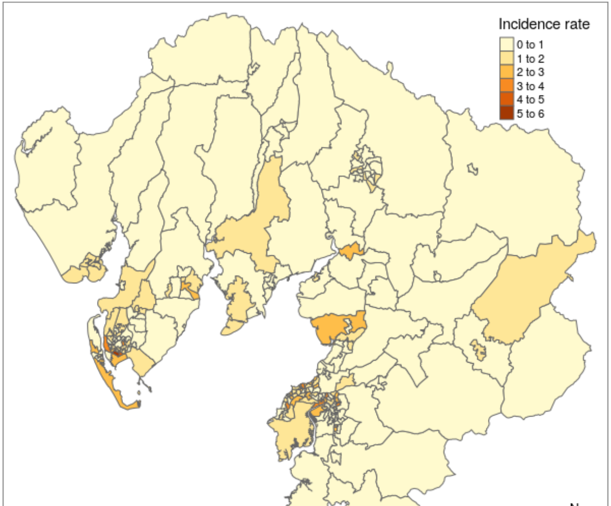
Gelfand's book classifies spatial statistics according to **spatial variation**:

1. Discrete spatial variation: Gaussian Markov Random Field (GMRF)
2. Continuous spatial variation: Gaussian Random Field (GMRF)

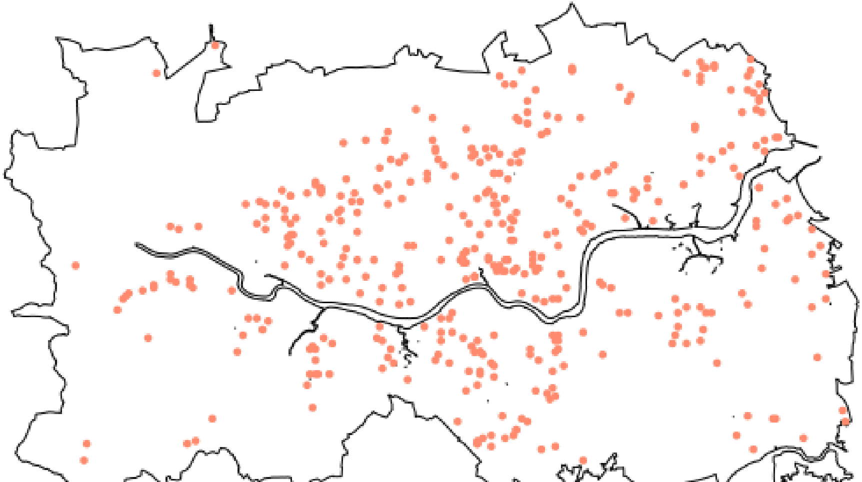
Geostatistical data: River blindness in Cameroon



Lattice Data: COPD emergency admission



Point pattern: Primary biliary cirrhosis data



Model-based Geostatistics

Modelling Geostatistical Data – Model-based Geostatistics

- The term **Model-based Geostatistics (MBG)** was coined by Peter Diggle in 1998 (Diggle, Tawn, and Moyeed 1998; Diggle and Giorgi 2019).
- MBG applies general principles of statistical modelling and inference to the analysis of geostatistical data.
- It emphasises the use of likelihood-based inference.
- **and the use of latent spatial process (Gaussian or stochastic process)**

Standard MBG model

$$Y_i = \beta_0 + \underbrace{\beta_1 d_1(x_i) + \beta_2 d_2(x_i)}_{\text{explained}} + \underbrace{S(x_i) + Z_j}_{\text{unexplained}},$$

where

- β_0, β_1, and β_2 are regression coefficients;
- $d_1(x)$ and $d_2(x)$ denote covariates/predictors at location x;
- $S(x)$ is the stochastic spatial process;
- Z_j represents measurement error.

Modelling $S(x)$ and Z_j

- $Z_i \sim N(0, \tau^2)$
- We assume that $S(x)$ is a zero-mean **stationary and isotropic Gaussian process**. i.e $S \sim \text{MVN}(0, \Sigma)$
- The (i, j)th entry of Σ is

$$\Sigma_{ij} = \text{Cov}\left(S(x_i), S(x_j)\right) = \sigma^2 \rho(u)$$

- $u = \|x_i - x_j\|$ is the Euclidean distance between locations x_i and x_j
- **How do we choose parametric correlation function $\rho(\cdot)$?**

The Matérn correlation function

$$\rho(u;\phi,\kappa) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{u}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{u}{\phi}\right),$$

- $K_\kappa(\cdot)$: modified Bessel function of order κ
- **Interpretation**
  - κ determines the smoothness: $\kappa > r \Rightarrow S(x)$ is r times differentiable
  - ϕ determines the scale of spatial correlation
- **Special cases**
  - $\kappa = 0.5 \Rightarrow \rho(u) = \exp\{-u/\phi\}$
  - $\kappa \to \infty \Rightarrow \rho(u) = \exp\{-(u/\phi)^2\}$
- Often sufficient to choose $\kappa \in \{0.5, 1.5, 2.5\}$