Geostatistical models – Part II

Applied Spatial Statistics

Spatial linear models

Part I introduced the standard model

$$Y_i = \mu_i + Z_i + \varepsilon_i$$

- The mean is the same as linear regression

$$\mu_i = \beta_0 + X_{i1}\beta_1 + ... + X_{ip}\beta_p$$

- There are two error terms:
  - Z_i is spatially-correlated
  - ε_i are independent across i

In Part II we will discuss models for Z_i

Spatial covariance models

- The Z_i capture spatial correlation not explained by the covariates
- Assume they are Gaussian with mean zero and variance σ^2
- Define covariance function as

$$\text{Cov}(Z_i, Z_j) = C(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2 \rho(\mathbf{s}_i, \mathbf{s}_j)$$

- The spatial correlation function ρ is large when $\mathbf{s}_i$ and $\mathbf{s}_j$ are close and small when they are far apart
- In the extreme, assume that $\rho(\mathbf{s}, \mathbf{s}) = 1$ so that Z is a continuous function of $\mathbf{s}$

Spatial covariance models

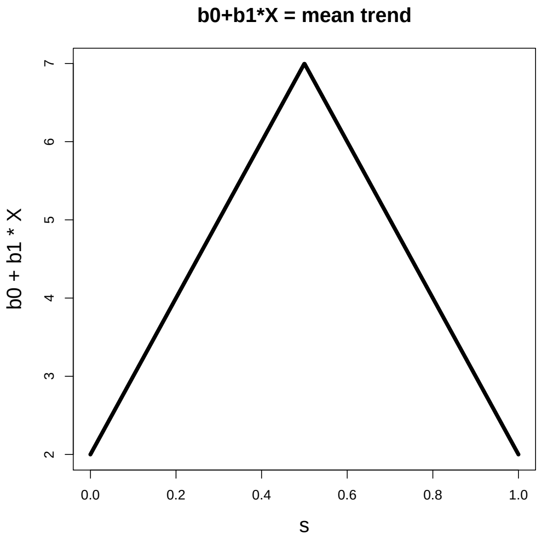
Common sources of spatial variation:

- Missing covariates that have spatial patterns
- Shared exposure to a common point sources (power plant)
- Share exposure to meteorological events
- Dispersion (e.g., pests, air, water)
- Measurement error, e.g., blurred satellite images

Fake data in one dimension (say s=longitude)

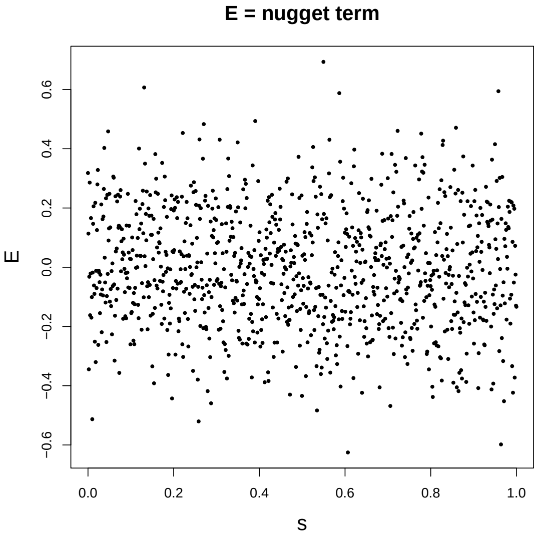
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



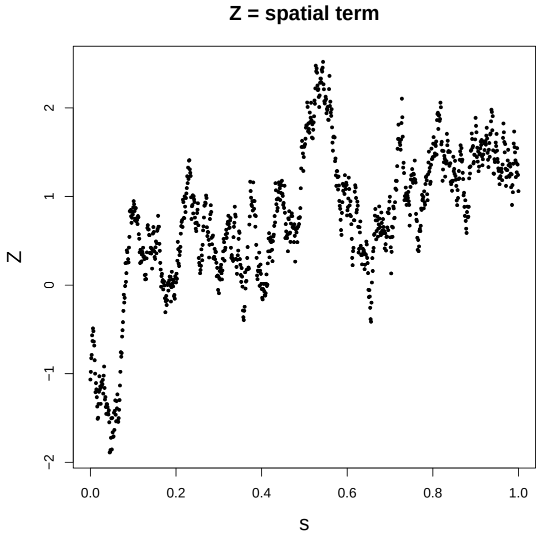
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



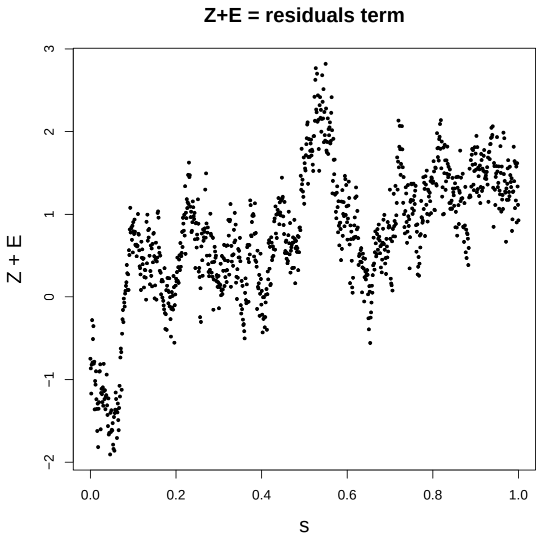
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



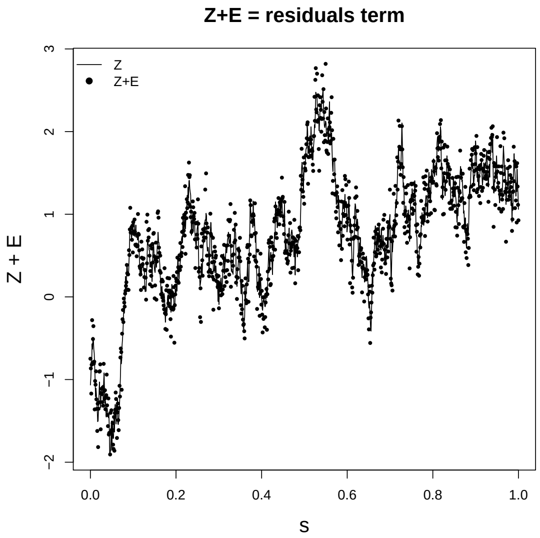
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



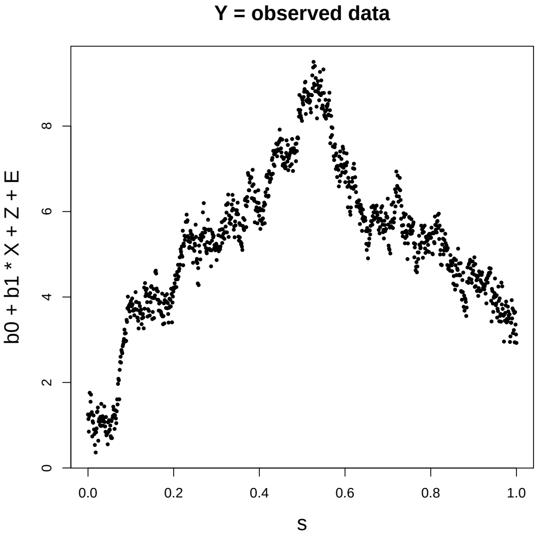
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



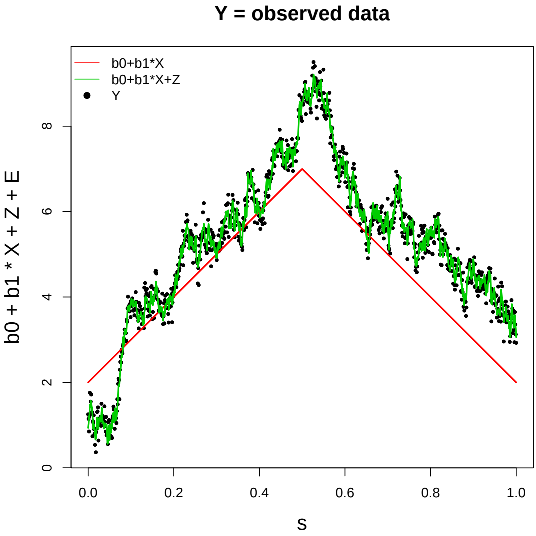
Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Fake data in one dimension (say s=longitude)



Model: $Y_i = \beta_0 + \beta_1 X_i + Z_i + E_i$

Spatial covariance models

Which factors might contribute each type of variation in the air pollution example?

- Mean trend:
- Spatial error:
- Nugget error:

Spatial covariance models - isotropy

- The spatial correlation term $\rho(\mathbf{s}_i, \mathbf{s}_j)$ is the novel term in a spatial analysis
- As illustrated from the plots above, it is challenging to estimate
- For example, how would you estimate $\rho(0.2, 0.3)$?
- We will make several simplifying assumptions (isotropic, stationary, etc)
- We will then propose models for the correlation function that satisfy these assumptions

Spatial covariance models - isotropy

- The simplest assumption is **isotropy**
- A covariance is isotropic if it depends only on the distance between locations
- Denoting the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$ as d_{ij}, then

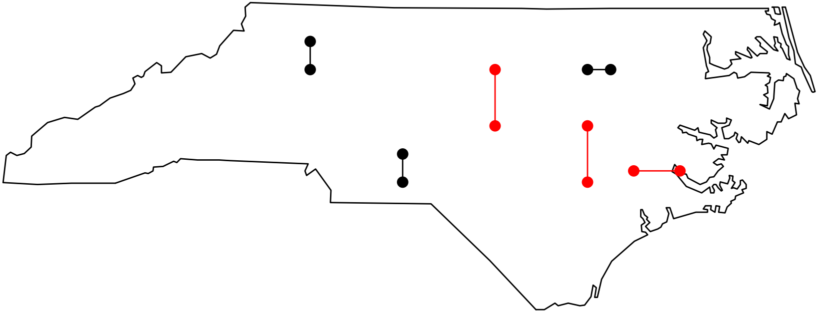
$$\text{Cor}(Z_i, Z_j) = \rho(\mathbf{s}_i, \mathbf{s}_j) = \rho(d_{ij})$$

- For example, the exponential correlation model is

$$\rho(d) = \exp(-d/\phi)$$

where ϕ determines the range of correlation

Spatial covariance models - isotropy



Pairs with the same color have the same correlation

Spatial covariance models - stationary

- Stationarity is more general than isotropy
- A covariance is stationary if it depends only on the difference $\mathbf{s}_i - \mathbf{s}_j$

$$\text{Cov}(Z_i, Z_j) = \rho(\mathbf{s}_i, \mathbf{s}_j) = \rho(\mathbf{s}_i - \mathbf{s}_j)$$

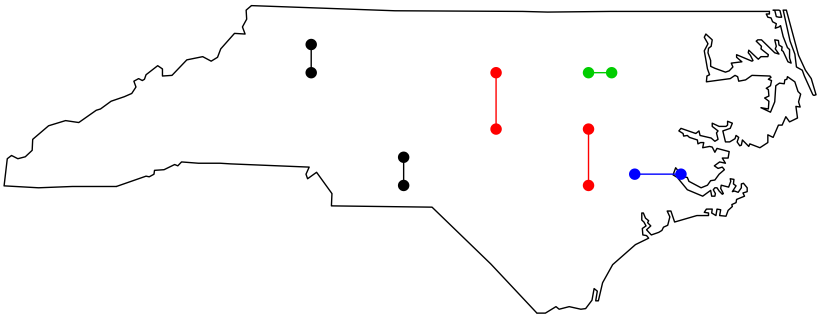
- Stationary means that the covariance is the same throughout the spatial domain
- For if two sites in the east have $\mathbf{s}_1 - \mathbf{s}_2 = \mathbf{h}$ and two sites in the west have $\mathbf{s}_3 - \mathbf{s}_4 = \mathbf{h}$, then

$$\text{Cor}(Z_1, Z_2) = \text{Cor}(Z_3, Z_4)$$

Spatial covariance models - anisotropy

- Stationary means that the covariance is the same throughout the spatial domain
- However, a stationary covariance can depend on both the distance between locations and the angle
- A covariance that is depends on the angle between location is called **anisotropic**
- For example, if the predominant wind pattern is east/west, then east/west pairs may have stronger correlation than north/south pairs

Spatial covariance models - anisotropy



Pairs with the same color have the same correlation

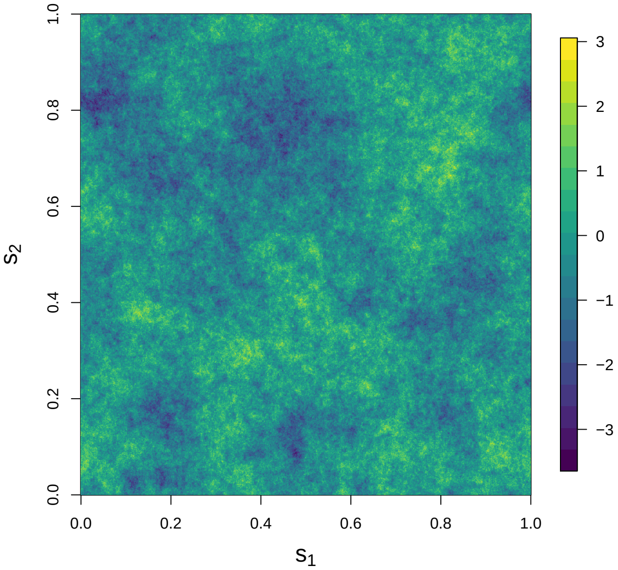
Spatial covariance models - nonstationary

If the spatial covariance can change in different parts of the spatial domain, it is **nonstationary**

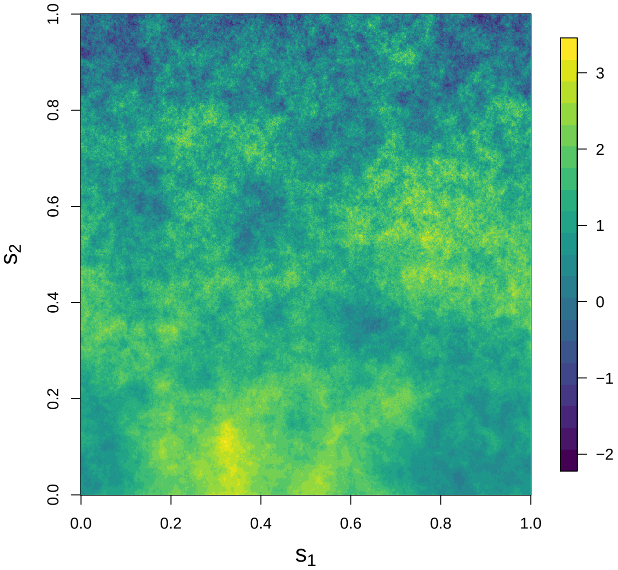
- Stronger correlation in the east than the west
- Larger variance in the north than the south
- Stronger correlation between sites at the same elevation
- Larger variance in cities than rural areas

Nonstationary covariance are much harder to estimate

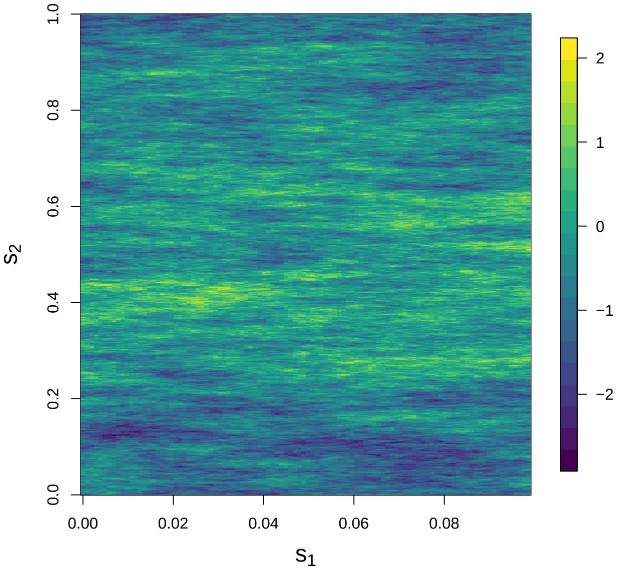
Isotropic, anisotropic or nonstationary?



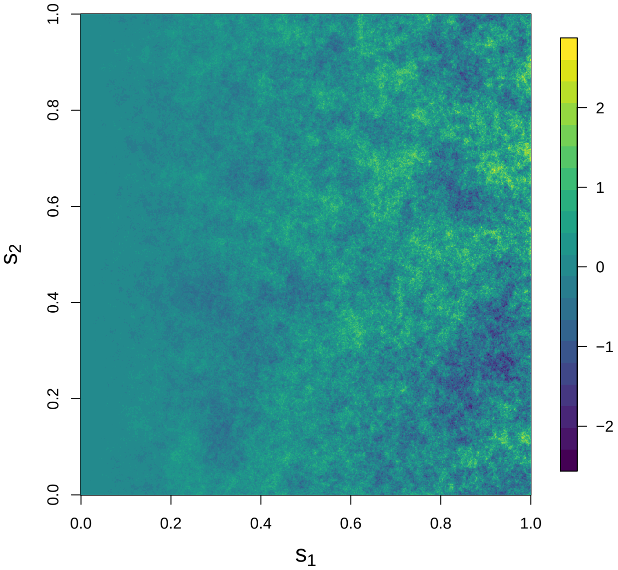
Isotropic, anisotropic or nonstationary?



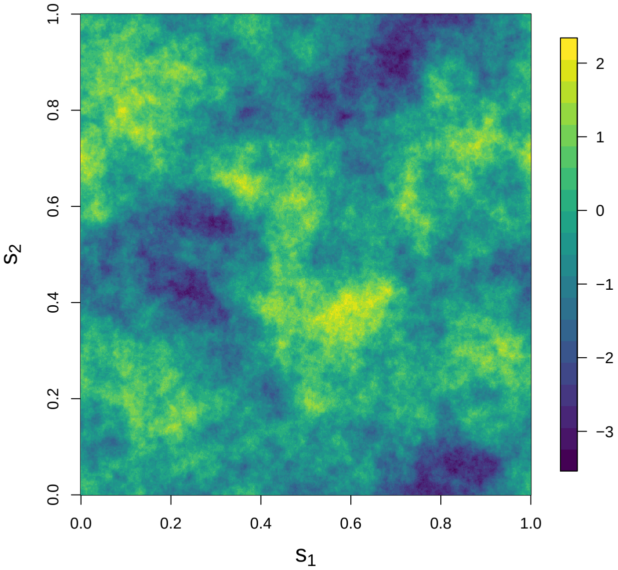
Isotropic, anisotropic or nonstationary?



Isotropic, anisotropic or nonstationary?



Isotropic, anisotropic or nonstationary?



Isotropic, anisotropic or nonstationary?

- Landsat data
- Malaria data
- Link to maps of the microbiome data