

# Tutorial Sheet 4

## 1 Areal processes

The figure below shows coloured dissolved organic matter (CDOM) across a lake along with the corresponding Moran's I plot (Figure 1).

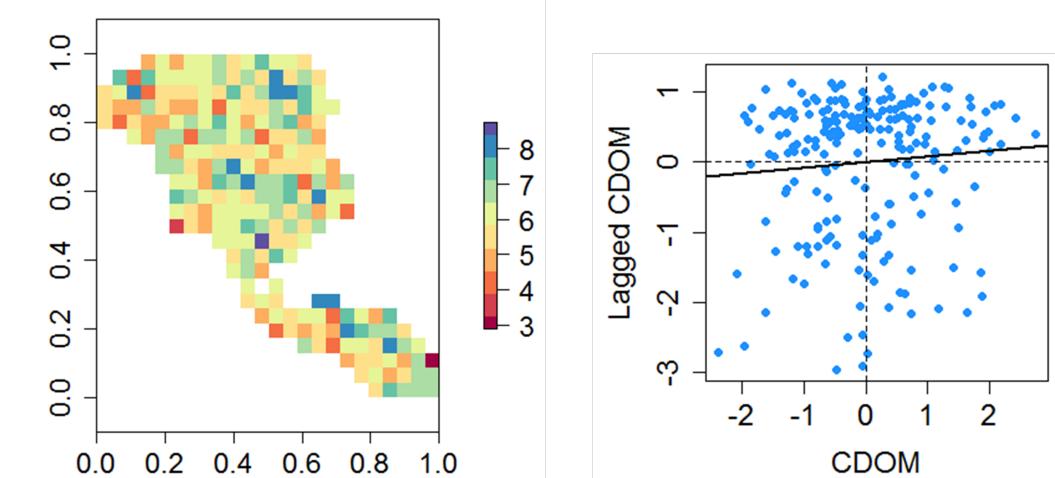


Figure 1: Lake CDOM plotted over space (left) and the corresponding Moran's I plot (right) based on Queen's distance

The estimated Global Moran's I is 0.02 with a variance of 0.02.

### **i** Interpreting Moran's I plot

A **Moran's I scatter plot** visualizes spatial autocorrelation by plotting the values of observations in different areas against their **spatially lagged values**, which represent the weighted average of neighboring values. A positive correlation in the plot suggests **spatial clustering** (similar values are located near each other), while a negative correlation indicates **spatial dispersion** (neighboring values tend to be dissimilar). The slope of the regression line through the points provides an estimate of **Moran's I**, a statistic that quantifies the degree of spatial autocorrelation.

## Task 1

Define what is meant by the term neighbourhood matrix and discuss two approaches for defining it, giving a drawback of each

## Task 2

Comment on the spatial variability in CDOM for this lake, with specific reference to the Moran's I plot

## Task 3

In your own words what does the Modifiable Areal Unit Problem (MAUP) refers to?

## Task 4

Figure 2 shows the number of lung cancer cases in Pennsylvania per county, USA, in 2002.

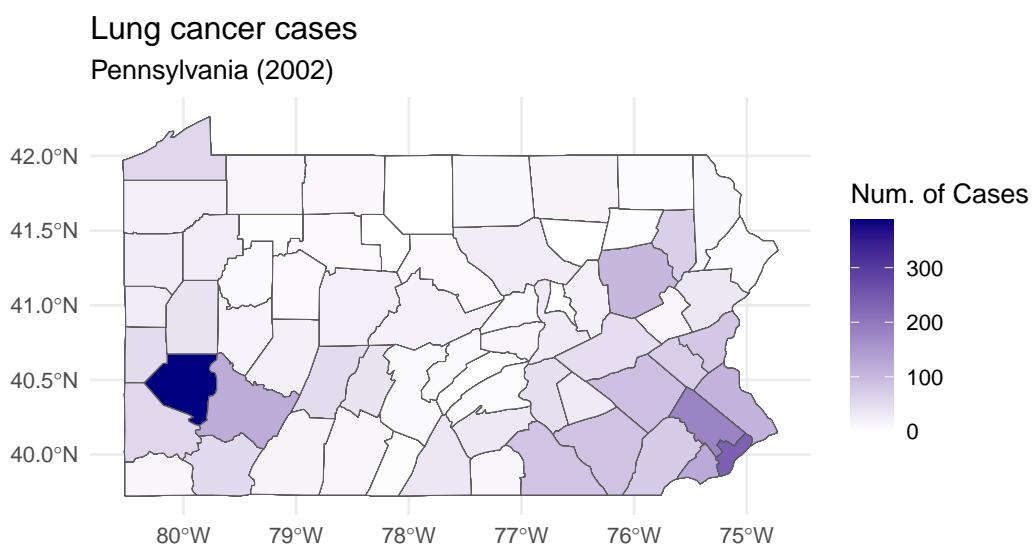


Figure 2: Number of Lung Cancer cases in Pennsylvania counties in 2002

1. Interpret the map , what does this tells you about the number of lung cancer cases?
2. In addition to the number of cases, the data set contains county-level information of the following variables:
  - Expected number of cases  $E_i$  computed as  $E_i = \sum_{j=1}^m r_j \times n_j$  ( $r_j$  is rate of disease and  $n_j$  the population in stratum  $j$ )
  - Overall county smoking rate

Propose a reasonable spatially explicit model to assess the relationship between lung cancer risk and smoking.

## 2 Geostatistics

## Task 5

Suppose we have a geostatistical process,  $\{Z(s); s \in D\}$ ,  $D \subset \mathbb{R}^2$  which is stationary with mean,  $\mu_z$  and covariance  $\text{Cov}(h)$ . Define what is meant by:

- weakly stationary
- isotropic

Then, write down an expression for the autocorrelation function  $\rho_z(h)$  in terms of the covariance function.

## Task 6

Figure 3 below shows summary plots for measurements of a geostatistical process, namely levels of Nitrogen measured at different locations across Chesapeake Bay, USA. Subjectively, comment on what each of these plots tells you about the data

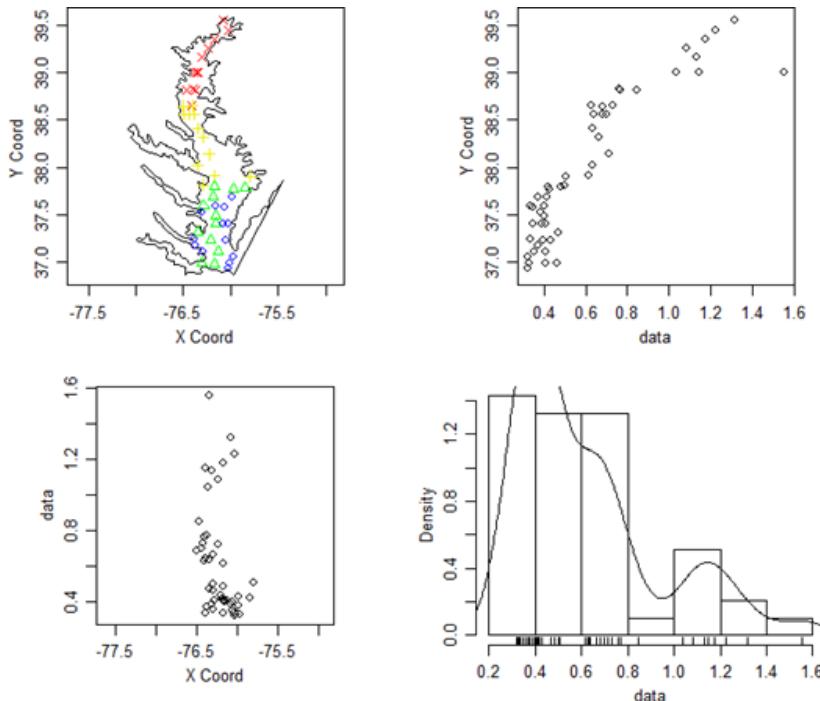


Figure 3: Summary plots for nitrogen levels across Chesapeake bay. Top left plot colour scale red to blue (high to low values).

## Task 7

Describe how you would check for spatial correlation in the data and model it if it were present.

## Task 8

An empirical semi-variogram with a MC envelope is shown below (Figure 4). Comment on the plot with regards to the presence of spatial correlation.

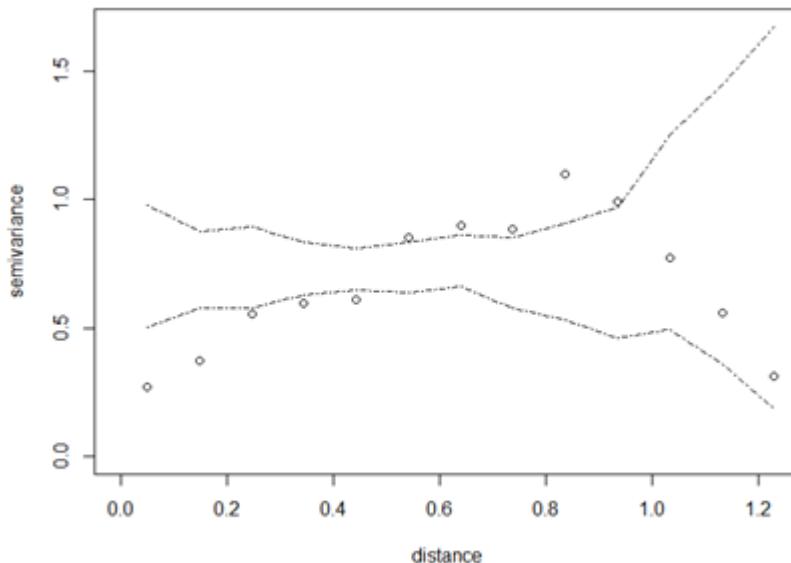


Figure 4: Empirical variogram and MC simulation envelope

## 3 Point Processes

---

### Task 9

Explain what a spatial point pattern is, define complete spatial randomness, and how it is assessed.

---

### Task 10

Discuss two limitations of using Ripley's  $K$  for assessing CSR in spatial point patterns.

---

### Task 11

Figure 5 (a, left) shows the  $K$  function for the locations of 500 trees of a particular species within an area of tropical rainforest while Figure 5 (b, right) shows the  $K$  function for the Lansing wood tree species. For each of these plots comment on the spatial pattern of the data that generated the  $K$  functions with respect to complete spatial randomness.

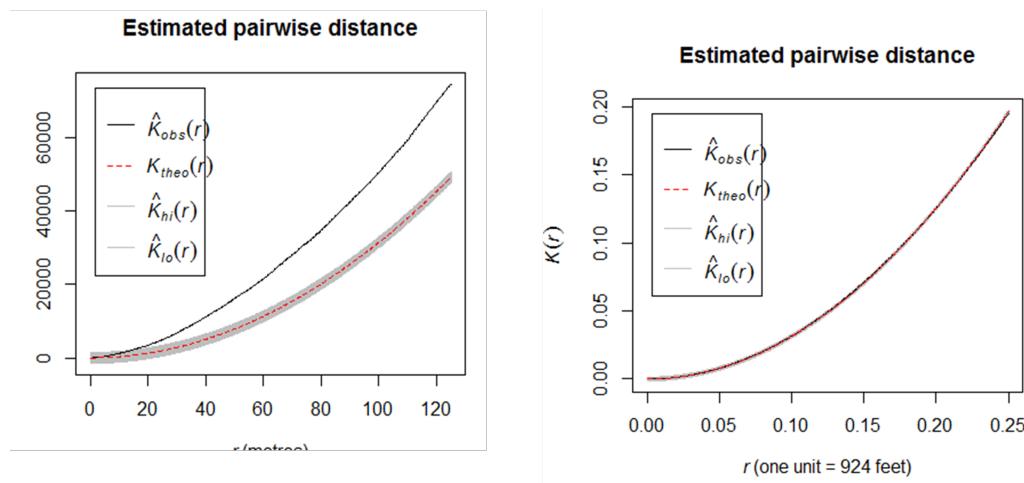


Figure 5: (a)  $K$  function for location of a particular species of tree in a tropical rainforest and (b) the  $K$  function for the distribution of trees in Lansing Wood.