Exam **Spatial Statistics**,
July 3rd, 2025, 13.45–16.45.

This is a closed book exam. The use of electronic devices is not allowed. Please answer all questions clearly and legibly. For each of the questions, a correct answer is worth ten points. Please make sure that your name and student identification number are on every sheet of paper that you hand in.

---

1. **a** In one sentence, state the difference between deterministic and stochastic components of a geostatistical regression model

   b Consider the Gaussian semi-variance model for rainfall data in Switzerland

$$\gamma(h) = \begin{cases} 500 + 1500\left(1 - \exp\left(-3\left(\frac{h}{60}\right)^2\right)\right) & \text{if } 0 \leq h < a \\ 2000 & \text{if } h \geq a \end{cases}$$

   i. State the nugget and the sill of this model

   ii. At what distance does spatial autocorrelation cease to exist?

   iii. What proportion of variance can be attributable to spatial autocorrelation?

   iv. What is the semi-variance at the 70 km

   v. Write the expression of its corresponding covariance

2. **a** Consider the general geostatistical model below

$$y(s_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_k(s_i) + \epsilon(s_i),$$

where β_k is the regression coefficient of the independent variable $x_k(s_i)$, and $\epsilon(s_i)$ is the spatially correlated error term. Write the expression for its corresponding spatial predictor under the assumptions of ordinary kriging.

b Consider the ordinary least-squares (1) and generalized least-squares (2) estimators of regression coefficients for spatial data

$$\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y, \quad (1)$$

$$\hat{\beta}_{GLS} = (X^\top \Sigma^{-1} X)^{-1} (X^\top \Sigma^{-1} y), \quad (2)$$

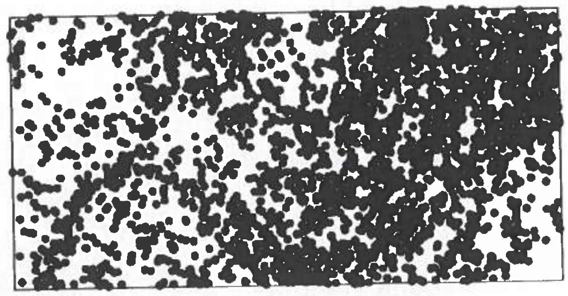
where y is the dependent variable, X is the design matrix of the independent variables, and Σ is the variance-covariance matrix.

i. From your perspective, describe the information that needs to be provided in the variance-covariance matrix Σ.

ii. Under what situation does equation (2) become equivalent to equation (1)

iii. Describe how you would estimate $\hat{\beta}_{GLS}$

3. The figure below shows the locations of 3535 *Capparis frondosa* trees on Barro Colorado island in a 1000 by 500 metre plot.



a To investigate whether the point process that generated the data is stationary, a quadrat count test is performed, which returns the following output.

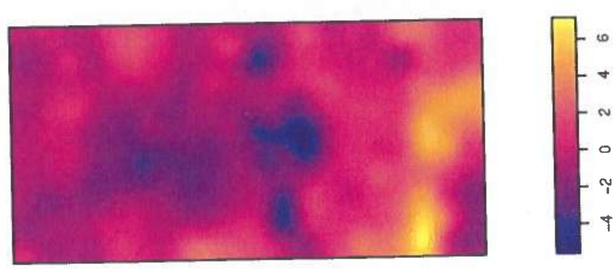
```
> Chi-squared test of CSR using quadrat counts

X2 = 1711.1, df = 199, p-value < 2.2e-16
alternative hypothesis: two.sided

Quadrats: 20 by 10 grid of tiles
```

What is your conclusion?

b A forester poses the question whether the prevalence of trees depends on the content of Cu in the soil shown in the figure below.



He fits the following Poisson model

$$\log \lambda((x, y); \theta_0, \theta_1) = \theta_0 + \theta_1 C(x, y),$$

where λ is the intensity function, $C(x, y)$ is the content of Cu at (x, y) and θ_0, θ_1 are unknown parameters. Formulate a statistical test that answers the forester's question at significance level 1%.

c To assess the fit of the model in [**b**], the forester considers the residuals, which are shown here.

How would you assess the fit.

The forester also fits an alternative fourth-order polynomial model with log intensity function

$$\log \lambda(x, y) = \theta_0 + \theta_1 x + \theta_2 y + \cdots + \theta_{13} x y^3 + \theta_{14} y^4.$$

d He would like to compare the two models using the Akaike Information Criterion (AIC). Please give the motivation behind this criterion as well as a formal definition.

e The forester obtains an AIC of 41985.27 for the model based on Cu content and 41289.83 for the alternative polynomial model. Which of these models do you prefer and why?

4. Let X be a Poisson process on the unit ball $B(0,1) = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq 1\}$ in the plane with intensity function $\lambda(x,y) = \theta(x^2+y^2)$ for some unknown parameter $\theta > 0$.

a Let N_X denote the total number of points placed by X in the unit ball. Calculate the expectation and the second-order factorial moment of N_X.

b Propose an estimator for θ.

c Is the estimator in [**b**] efficient? Motivate your answer.

Let Y be a Strauss process on the unit ball defined by the density function

$$f(\{(x_1,y_1),\ldots,(x_n,y_n)\}) \propto \prod_{i=1}^{n} \beta(x_i,y_i) \prod_{i<j} \gamma((x_i,y_i),(x_j,y_j)),$$

$(x_i,y_i) \in B(0,1)$ for $i=1,\ldots,n$, with $\beta(x,y) = \theta(x^2+y^2)$ and

$$\begin{cases} \gamma((x_i,y_i),(x_j,y_j)) &= \gamma \text{ if } (x_i-x_j)^2+(y_i-y_j)^2 < 0.01 \\ \gamma((x_i,y_i),(x_j,y_j)) &= 1 \text{ otherwise} \end{cases}$$

for $\gamma \in (0,1)$.

d Explain the idea behind maximum pseudo-likelihood estimation.

e Derive maximum pseudo-likelihood estimators for θ and γ. Compare your answer to that in [**b**].