NTNU

Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4250 Spatial Statistics**

Academic contact during examination: Ingelin Steinsland

Phone: 9266 3096

Examination date: June 2, 2017

Examination time (from–to): 09:00–13:00

Permitted examination support material: C
Tabeller og formler i statistikk, Akademika,
Calculator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College or HP30S
Yellow stamped A5-sheet with personal hand written notes

Language: English

Number of pages: 3

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1 Humidity start-up

After graduation you and a friend establish a start-up. The business idea is to provide a spatial map (daily) of humidity for a grid over Norway with uncertainty, and to sell this product to SeNorge. Humidity is defined as the mass of water vapor contained in a unit mass of dry air, and is hence without unit.

We assume that humidity can be modeled as a Gaussian Random field (GRF) with exponential covariance function, i.e. that the covariance between humidity at location s_1 and s_2 is given by $\sigma^2 \exp(-d(s_1, s_2)/\theta)$, where $d(s_1, s_2)$ is the Euclidean distance between s_1 and s_2 (in km), and σ^2 and θ are parameters.

Hint In this problem you might want to use that for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

a) Assume that we have two observations at location $s_1 = (0, 0)$, and at location $s_2 = (30, 0)$. We want to predict at location $s_3 = (0, 40)$. Further, for this point and in e) only, we know that the GRF has mean 3.0, $\sigma^2 = 1$ and $\theta = 10$. In this point only we assume that the observations are perfect.

Derive the predictive distribution for humidity at location s_3, and set up the expressions for the mean and variance when observed humidity at location s_1 is 3.0 and at location s_2 is 4.0. Calculate the predictive mean and variance.

In reality humidity is measured with uncertainty. We assume that the measurement error follows a Gaussian distribution with mean 0 and unknown variance σ_ϵ^2

b) Specify mathematically, including assumptions, a spatial hierarchical model for humidity. For the parameter level you only need to specify which parameters that need to be assign a prior in a full Bayesian model (you do not need to assign priors).

c) Discuss how you can make inference and predictions based on observations and the model specified in point b). Suggest at least two approaches, and give key formulas.

Networking with meteorologists reveal that the humidity measurements are done with two different instruments, type A and type B, and that these might have different measurement uncertainty. The instrument type is known for each location.

d) Extend your hierarchical model, and explain how you can do inference now (changes compared to point c)).

e) We now consider the same situation as in point a), but want to account for the measurement uncertainty.

Derive / set up the expressions for the predictive distribution for true humidity at location s_3 when measurement at location s_1 is of type A with measurement variance $\sigma_A^2 = 0.01^2$, and the measurement at location s_2 is of type B with measurement variance $\sigma_B^2 = 0.5^2$.

Further, assume that humidity at location s_1 is 3.0 and at location s_2 is 4.0.

Briefly explain for a non-specialist the differences you expect compared to the results for predictive mean and variance in a) (smaller or larger?, why?).

There are more than 100 observations of humidity in Norway, and the approach you have taken to inference is too slow for SeNorge to be interested in buying your product. To speed up the calculations a good friend of you suggest that you discretize the domain to your grid of interest, assign observations to their corresponding grid node and use a Gaussian Markov random field (GMRF) model, where each grid node has its four closest grid nodes as neighbors.

f) For the model visualized with the conditional independence graph in Figure 1:

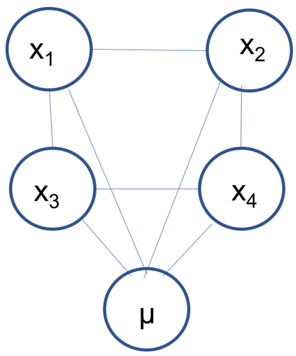

Figure 1: Conditional Independence Graph

- Specify all cliques
- Give the non-zero structure of the precision matrix (i.e. inverse of the covariance matrix)

g) Set up pseudo-code for using Laplace approximations to estimate the variances for measure type A and type B.

Explain briefly the computational benefits of using a GMRF model compared to a GRF model.

Problem 2 Cell firing of a rat

In Figure 2 the locations of a rat in a cage when a specific grid cell is fired is plotted (black dots). Consider these locations as a realization of a point process.

- Is this a realization of a random, clustered or regular point process? Argue why.
- Briefly describe one useful plot you can make (using a computer) to explore the underlying process.
- Sketch the plot you suggest for the cell firing data set, and argue for the key features in your sketch.

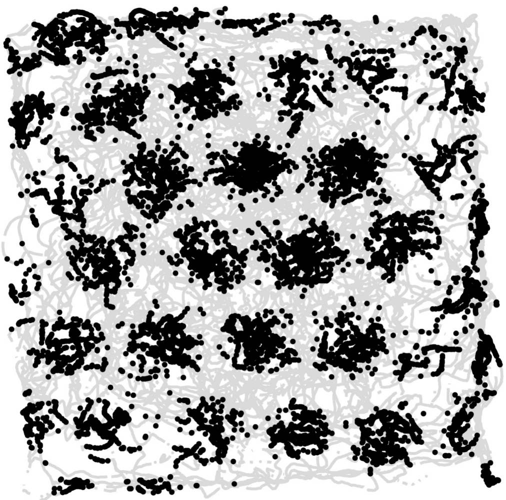

Figure 2: Cell firing

The regular hexagonal signal from a grid cell. Each firing of the cell is seen as a black dot. The path of the freely roaming rat is seen in gray. Figure from `http://www.ntnu.edu/kavli/discovering-grid-cells`