UNIVERSITETET I OSLO

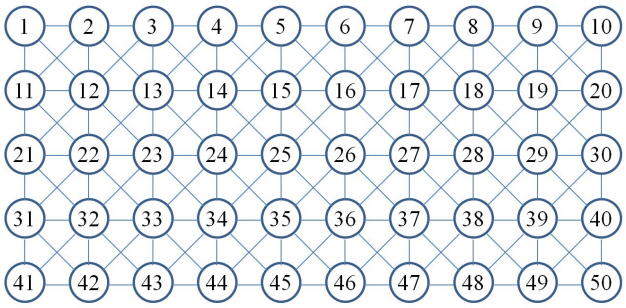
Det matematisk-naturvitenskapelige fakultet

Examination in:	STK9150 — Environmental and spatial statistics
Day of examination:	Tuesday May 30st 2017.
Examination hours:	09.00 – 13.00.
This examination set consists of 6 pages.	
Appendices:	None
Permitted aids:	Approved calculator, Cressie and Wikle: Statistics for Spatio-temporal data

Make sure that your copy of the examination set is complete before you start solving the problems.

Problem 1.

Consider a lattice model of 50 nodes, and let the neighborhood system be defined using the undirected graph in the figure below.



(a) List the neighbours of node 25.
In the setting of lattice models what is meant by a clique?
Make a drawing of all the cliques which contain node 25. It is sufficient to draw the unique shapes of the cliques, but include all orientations.

Assume that the lattice model is defined through a Gibbs distribution, i.e.

(Continued on page 2.)

$$p(\boldsymbol{Y} = \boldsymbol{y}) = C \cdot \exp\{Q(\boldsymbol{y})\}$$

with C beeing a normalizing constant, $p(\boldsymbol{Y} = \boldsymbol{0}) > 0$, and

$$Q(\boldsymbol{y}) = \sum_{c \in \mathcal{C}} y_{i_1} y_{i_1} \cdots y_{i_p} G_c(\boldsymbol{y}_c)$$

$$= \sum_{i=1}^{n} y_i G_i(y_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} y_i y_j G_{ij}(y_i, y_j) + \cdots + y_1 y_2, \ldots y_n G_{12\ldots n}(y_1, y_2, \ldots, y_n)$$

where $\mathcal{C}$ is the set of all subsets in the lattice, $c = (i_1, i_2, \ldots, i_p)$ is the counter for these subsets and, $\boldsymbol{y}_c = (y_{i_1}, y_{i_2}, \ldots, y_{i_p})$, holds the values of the lattice model in the subset c.

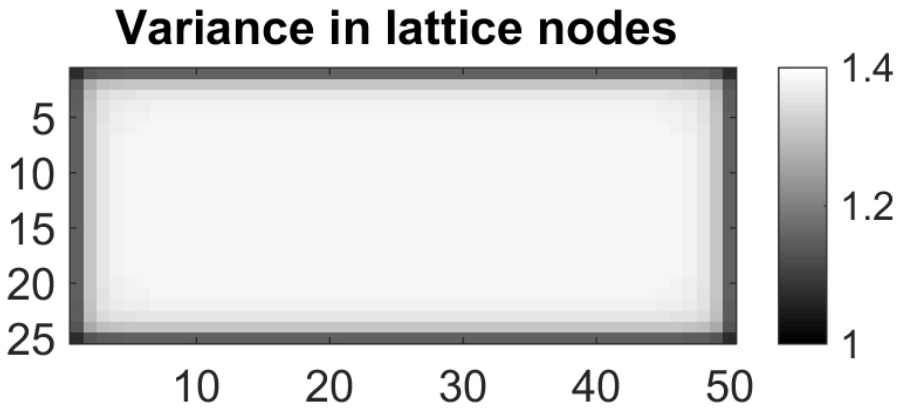
(b) What does Hammersley - Clifford Theorem state about the relation between $c = (i_1, i_2, \ldots, i_p)$ and $G_c(\boldsymbol{y}_c)$?
Assume further that the model is a Gaussian CAR model. Which additional simplification does this imply with respect to $G_c(\boldsymbol{y}_c)$?

The same neighborhood system as above is used on a 50 by 25 lattice, for this neighborhood system one possible CAR model is defined by assigning equal weight to the neighbors, i.e. by letting the precision matrix have the form $\boldsymbol{Q} = \boldsymbol{I} - \phi \boldsymbol{A}$ where

$$A_{i,j} = \begin{cases} 1 \text{ if } i \text{ is neighbor to } j \\ 0 \text{ otherwise} \end{cases}$$

(c) Is the matrix $\boldsymbol{A}$ symmetric? Give a reason for your answer.
The largest and smallest eigenvalues of $\boldsymbol{A}$ are respectively $\lambda_{\max} = 7.9449$ and $\lambda_{\min} = -3.9741$. What is the maximum and minimum value of ϕ that gives a valid CAR model?
The figure below shows the variance of the model in each lattice point for a particular choice of ϕ. What kind of problem does this figure indicate?
Is it a big or small problem in this case?
Comment in general on how this type of problem can be mitigated.

(Continued on page 3.)



Problem 2.

Consider a spatial process $\{Y(\boldsymbol{s})\}$ specified through

$$Y(\boldsymbol{s}) = \beta_0 + \beta_1 x_1(\boldsymbol{s}) + \beta_2 x_2(\boldsymbol{s}) + \delta(\boldsymbol{s})$$

where $\{\delta(\boldsymbol{s})\}$ is a zero-mean Gaussian spatial process with covariance function $C_\delta(\boldsymbol{s}, \boldsymbol{v}) = C_\delta(||\boldsymbol{s} - \boldsymbol{v}||)$.
Assume further that we have observations

$$Z(\boldsymbol{s}_i) = \boldsymbol{Y}(\boldsymbol{s}_i) \quad i = 1, ..., n$$

which are collected in a region of 20km by 5km. Estimates of parameters obtained by using the R function `gls` are displayed Table 1, with σ_{β_j} being the estimated standard deviation for the estimator of β_j. Four different models for the covariates are considered, all models assume that the correlation function of $\delta(\boldsymbol{s})$ is exponential.

Table 1

Parameter	Model 1	Model 2	Model 3	Model 4
β_0	5.78	3.92	5.19	4.08
σ_{β_0}	7.20	2.65	4.77	2.66
β_1	-	-1.46	-	-1.20
σ_{β_1}	-	-0.42	-	0.50
β_2	-	-	0.80	0.36
σ_{β_2}	-	-	0.54	0.44
R ([km])	7.01	2.19	4.31	2.18
σ	14.01	9.02	11.58	9.06
AIC	222.3	232.5	235.87	247.39
logLik	-108.1	-112.2	-113.94	-118.69

(Continued on page 4.)

The exponential correlation function is defined by

$$\mathrm{Cov}(\delta(\boldsymbol{s}), \delta(\boldsymbol{s}+\boldsymbol{h})) = \sigma^2 \exp\left(-\|\boldsymbol{h}\|/R\right).$$

(a) Comment on the results for the four models, what is positive and what is problematic?
Which model would you select? Give a reason for your choice.
Does the set of models give you any more insight about the data than what you would get from just the selected model?

(b) A spatial model can be defined using the estimated parameters for the selected model above, this model can then be used to make spatial predictions. What would you call this approach?
An alternative would be to assign prior distributions to all the parameters and apply a Bayesian approach. What are the strengths and weaknesses of the two different methods?
Describe briefly two different approaches that can be used for inference in the Bayesian model.

Define the random vector $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n), Y(\boldsymbol{s}_{n+1}), \ldots, Y(\boldsymbol{s}_{n+m}))$, where the initial n locations are the ones where the process has been observed, and the m locations to follow are new locations. Let the distribution of $\boldsymbol{Y}$ be defined according to the two blocks $\boldsymbol{Y}_1 = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))$ and $\boldsymbol{Y}_2 = (Y(\boldsymbol{s}_{n+1}), Y(\boldsymbol{s}_{n+2}), \ldots, Y(\boldsymbol{s}_{n+m}))$, such that :

$$p(\boldsymbol{Y}) = N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right).$$

Let $\boldsymbol{Y}^s = (\boldsymbol{Y}^s_1, \boldsymbol{Y}^s_2)$ be a random sample from $p(\boldsymbol{Y})$, and define

$$\boldsymbol{Y}^s_{2|1} = \boldsymbol{Y}^s_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{b} - \boldsymbol{Y}^s_1)$$

(c) Derive the mean and covariance matrix of $\boldsymbol{Y}^s_{2|1}$.
How can this be used to sample form the posterior distribution of $p(\boldsymbol{Y}_2|\boldsymbol{Y}_1 = \boldsymbol{y}_1)$?

Problem 3.

We will in this case consider a spatio-temporal process $\{Y_t(\boldsymbol{s})\}$ discrete in time and defined on a fixed set of points in space,$(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m)$. where $\boldsymbol{Y}_t = (Y_t(\boldsymbol{s}_1), Y_t(\boldsymbol{s}_2), \ldots, Y_t(\boldsymbol{s}_n))$. The model is defined through an vector AR process.

$\boldsymbol{Y}_1 \sim N(\boldsymbol{\mu}_Y, \boldsymbol{C}_Y)$

(Continued on page 5.)

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{M}\boldsymbol{Y}_{t-1} + \boldsymbol{\delta}_t$$

where $\boldsymbol{\delta}_t$ is a zero mean Gaussian spatial process with covariance $\boldsymbol{C}_\delta$, and independent of $\boldsymbol{Y}_s$ for $s < t$.

(a) Which restrictions/relations are required for $\boldsymbol{\mu}_Y, \boldsymbol{C}_Y, \boldsymbol{\mu}, \boldsymbol{M}$ and $\boldsymbol{C}_\delta$ in order for $\{Y_t(\boldsymbol{s})\}$ to be second order stationary in time?
Under the distributional assumptions made, is the process time stationary in the strict sense if these relations hold?

(b) Set up an expression for the full joint distribution of $\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_T\}$, use this expression to show that $p(\boldsymbol{Y}_t|\boldsymbol{Y}_u, u \neq t) = p(\boldsymbol{Y}_t|\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t+1})$, and derive the conditional distribution.
Hint: For the latter part derive a quadratic form for the logarithm of the probability distribution and identify the parameters of a Gaussian distribution.

(c) Under the assumption that the process is stationary in time there is also a reverse recursion

$$\boldsymbol{Y}_{t-1} = \tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{M}}\boldsymbol{Y}_t + \tilde{\boldsymbol{\delta}}_t$$

Set up the joint distribution of Y_{t-1} and Y_t, and use this to derive the expressions for $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{M}}$ and $\boldsymbol{C}_{\tilde{\delta}_t}$.
Argue why $\tilde{\boldsymbol{\delta}}_t$ is independent of Y_s for $s >= t$.

In the setting of the Kalman filter, observations are made at each time step $t = 1, 2, \ldots, T$ according to a time varying incidence matrix $\boldsymbol{H}_t$.

$$\boldsymbol{Z}_t = \boldsymbol{H}_t\boldsymbol{Y}_t + \boldsymbol{\epsilon}_t$$

where all observation errors are independent, have mean zero and a Gaussian distribution. We now want to derive an expression for the Kalman smoother of $\boldsymbol{Y}_t$. Use the notation

$$\boldsymbol{Y}_t|\boldsymbol{Z}_{1:u} \sim N(\boldsymbol{Y}_{t|u}, \boldsymbol{P}_{t|u}),$$

and

$$\boldsymbol{Y}_t|\boldsymbol{Z}_{u:T} \sim N(\tilde{\boldsymbol{Y}}_{t|u}, \tilde{\boldsymbol{P}}_{t|u}).$$

(Continued on page 6.)

(d) What is the difference between a Kalman filter and a Kalman smoother?
If you start a Kalman filter from the initial time and run it forward, which parameters $\boldsymbol{Y}_{t|u}$, $\boldsymbol{P}_{t|u}$ are derived in this process?
If you start a Kalman filter at the final time and use the inverse transitions from (c) and standard updating procedure, which $\tilde{\boldsymbol{Y}}_{t|u}$, $\tilde{\boldsymbol{P}}_{t|u}$ do you compute then? Write your answers in terms of pairs of t and u.

(e) Use the results from (b) and (d) and the likelihood corresponding to $\boldsymbol{Z}_t$ to derive an expression for the Kalman smoother $p(\boldsymbol{Y}_t|\boldsymbol{Z}_{1:T})$. Introduce intermediate vectors and matrices to simplify the computations. You may assume that $P(\boldsymbol{Y}_1, \boldsymbol{Y}_T) \approx P(\boldsymbol{Y}_1)P(\boldsymbol{Y}_T)$.
When is this assumption reasonable?

END