UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Examination in:	STK9150 — Environmental and spatial statistics
Day of examination:	Wednesday June 3th 2015.
Examination hours:	14.30 – 18.00.
This examination set consists of 4 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 spatial process $\{Y(\boldsymbol{s})\}$ specified through

$$Y(\boldsymbol{s}) = \beta_0 + \beta_1 x(\boldsymbol{s}) + \delta(\boldsymbol{s})$$

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

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

where the ε_i's are independent and also independent of the $\{Y(\boldsymbol{s})\}$ process. Connected to the observations the covariates $\{x(\boldsymbol{s}_i), i = 1, ..., n\}$ are also observed.

(a) Find the expectations and covariance functions for the $\{Y(\boldsymbol{s})\}$ and $\{Z(\boldsymbol{s})\}$ processes.

What kind of processes are these?

Assume $C_\delta(h)$ is continuous in $h = 0$. What are the nugget effects for the two processes?

(Continued on page 2.)

(b) Assume now $C_\delta(h) = C_\delta^0(h) + \sigma_1^2 I(h = 0)$ where $C_\delta^0(h)$ is continuous in zero, what are the nugget effects in this case?

What kind of inferential problems can occur if we assume $\sigma_1^2 > 0$?

In the following we will assume $C_\delta(h)$ is continuous in $h = 0$.

(c) Assume that $\beta_0, \beta_1, \sigma_\varepsilon^2, x(\boldsymbol{s})$ and $C_\delta(h)$ are known. Write down a formula for the optimal predictor for $Y(\boldsymbol{s}_0)$ and $Z(\boldsymbol{s}_0)$ expressed trough the quantities known (you do not need to derive these formulae).

In what sense is it optimal?

Also write down prediction errors.

(d) What is the lowest value possible for the prediction error for $Z(\boldsymbol{s}_0)$ both with and without the restriction that $\boldsymbol{s}_0 \neq \boldsymbol{s}_i$ for all $i = 1, ..., n$?

Hint: Consider the case $n = 1$ first.

Consider now the case where $x(\boldsymbol{s})$ are known in the observation points $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ but unknown in $\boldsymbol{s}_0$. Assume that $\{x(\boldsymbol{s})\}$ also can be modeled as a zero-mean Gaussian process with covariance function $C_x(\boldsymbol{s}, \boldsymbol{v}) = C_x(||\boldsymbol{s}-\boldsymbol{v}||)$. Further, the $\{x(\boldsymbol{s})\}$ process is assumed to be independent of $\{\delta(\boldsymbol{s})\}$ and the observation errors.

(e) Find the expectations and covariance functions for the Y- and Z-processes in this case. Describe (without going into much details) how prediction of $Y(\boldsymbol{s}_0)$ and $Z(\boldsymbol{s}_0)$ can be performed based on the Z-observations only.

(f) Is there any benefit in utilizing the x-observations in this case? Describe how this can be done.

Problem 2.

We will in this case consider a spatio-temporal process $\{Y_t(\boldsymbol{s})\}$ discrete in time and defined on a regular lattice in space.
Assume that

$$Y_t(\boldsymbol{s}) = \beta_0 + \delta(\boldsymbol{s}) + \gamma_t + \kappa_t(\boldsymbol{s}) \qquad (*)$$

where $\{\delta(\boldsymbol{s})\}$ is a zero-mean Gaussian spatial process defined through a Markov random field of the *Besag* type, that is

$$[\delta(\boldsymbol{s})|\delta(\boldsymbol{v}), \boldsymbol{v} \neq \boldsymbol{s}] = N(\tfrac{1}{n_{\boldsymbol{s}}} \sum_{\boldsymbol{v} \in \mathcal{N}_{\boldsymbol{s}}} \delta(\boldsymbol{v}), \tfrac{\sigma_\delta^2}{n_{\boldsymbol{s}}}),$$

(Continued on page 3.)

$\{\gamma_t\}$ is a Gaussian random walk process,

$$\begin{aligned}\gamma_0 =&\eta_0\\ \gamma_t =&\gamma_{t-1}+\eta_t, \quad t=1,2,...\end{aligned}$$

where $\eta_t \overset{iid}{\sim} N(0,\sigma_\eta^2)$ and finally $\{\kappa_t(\boldsymbol{s})\}$ is a zero-mean Gaussian spatio-temporal process where

$$\begin{aligned}\kappa_0(\boldsymbol{s}) =&\nu_0(\boldsymbol{s})\\ \kappa_t(\boldsymbol{s}) =&\kappa_{t-1}(\boldsymbol{s})+\nu_t(\boldsymbol{s}), \quad t=1,2,...\end{aligned}$$

where $\{\nu_t(\boldsymbol{s})\}$ is for each t a Gaussian Markov random field with a model similar to $\{\delta(\boldsymbol{s})\}$, but replacing σ_δ with σ_ν, and independent in time.
These three random processes are further independent of each other.

(a) Consider first the $\{\kappa_t(\boldsymbol{s})\}$ process. Is it stationary in time? Is the associated covariance function separable?

(b) Derive the covariance function for $\{Y_t(\boldsymbol{s})\}$ as a function of the covariance functions for the different components involved.

Argue why this covariance function is not separable.

Is the process stationary in time?

Consider now observations $\{Z_t(\boldsymbol{s})\}$, connected to the $\{Y_t(\boldsymbol{s})\}$ process through

$$Z_t(\boldsymbol{s}) = Y_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}), \quad \varepsilon_t(\boldsymbol{s}) \overset{iid}{\sim} N(0,\sigma_\varepsilon^2).$$

This observation model together with model (*) was fit to the sea surface temperature data from the textbook (only the first 10 time-points and using an empirical Bayes approach). The following estimates were obtained:

Table 1

Parameter	Estimate	2.5% quantile	97.5% quantile
β_0	0.0056	0.0038	0.0073
σ_ε	0.0024	0.0022	0.0029
σ_δ	0.0075	0.0039	0.0283
σ_η	0.0079	0.0040	0.0303
σ_ν	0.1903	0.1887	0.1923

(c) Based on the estimates above, discuss the importance of the different components in the model. In particular, consider whether spatial and/or temporal dependence seems to be important.

(Continued on page 4.)

Empirical orthogonal functions (EOF) are popular quantities that can be used to explain some of the variations involved. Consider now an extension of model (*) to

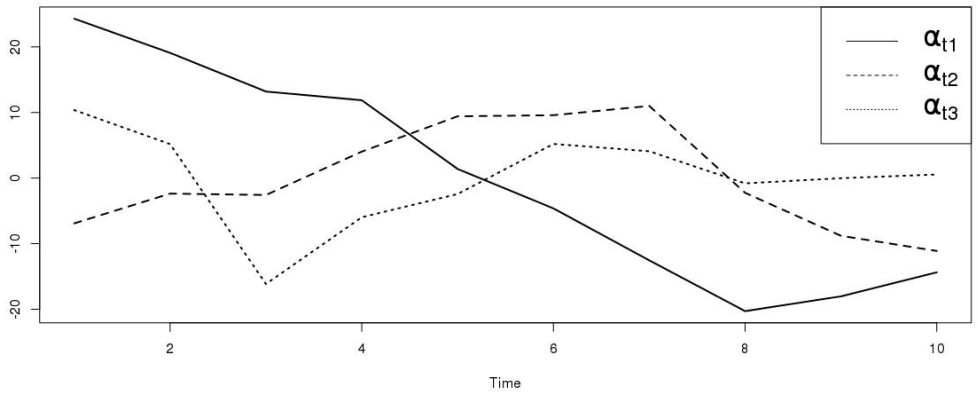
$$Y_t(\boldsymbol{s}) = \beta_0 + \sum_{k=1}^{3} \phi_k(\boldsymbol{s})\alpha_{t,k} + \delta(\boldsymbol{s}) + \gamma_t + \kappa_t(\boldsymbol{s}) \qquad (**)$$

where $\phi_k(\boldsymbol{s})$ is the kth EOF at spatial location $\boldsymbol{s}$. These EOFs are calculated based on all the 399 time-points.
A fit to this model gave

Table 2

Parameter	Estimate	2.5% quantile	97.5% quantile
β_0	0.0056	-0.2024	0.2135
σ_ε	0.0023	0.0020	0.0028
σ_η	0.0072	0.0038	0.0279
σ_γ	0.0064	0.0038	0.0155
σ_ν	0.1728	0.1715	0.1734

while the $\alpha_{t,k}$ estimates are displayed on the plot below



(d) Explain what EOFs are and why it may be useful to use these as part of the model. Why is it reasonable that $\alpha_{t,1}$ is most variable?

(e) Compare the results in Tables 1 and 2. How does the introduction of the EOFs in the model influence the parameter estimates?

Given that the log-likelihood for the first model is 4581.7 while for the second model it is 6840.2, which model would you prefer?

END