UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Examination in:	STK4150 — Environmental and spatial statistics - FASIT
Day of examination:	Monday 10. june 2013.
Examination hours:	09.00 – 13.00.
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.

(a) The variogram, defined as

$$2\gamma(h) = \mathrm{Var}[Z(\boldsymbol{s}) - Z(\boldsymbol{s} + \boldsymbol{h})]$$

measures the spatial dependence as a function of displacement. For the given data, there is a large discontinuity at zero, indicating that there is a strong nugget effect (randomness with no spatial structure). Further, the increase in time indicates spatial structure. Note however that the curve do not stabilize, indicating that there are long-range dependence structures (although the uncertainty for large values is large).

(b) None of the spatial models gave any improvement over the non-spatial model.

(c) CAR models have sparse precision matrices, making them efficient to work with computationally. Further, a complex multivariate distribution is broken down to specification of many univariate distributions.

(Continued on page 2.)

(d) In this case an improvement of 28.54 in the log-likelihood was achieved. Further, there are no extra parameters in this model, implying that we get a large improvement also in AIC value. Therefore we can conclude that this model is better.

(e) The CAR model is not proper. That is, if we define $\tilde{\delta}_i = \delta_i + c$, $\{\tilde{\delta}_i\}$ will have the same distribution. This can cause identifiability problem, but the sum-constraint will solve this problem.

Problem 2.

(a) A process is temporal stationary if $E[Y(s,t)] = \mu(s)$ and $\text{cov}[Y(\boldsymbol{s};t), Y(\boldsymbol{x},r))] = C(\boldsymbol{s},\boldsymbol{x},t-r)$. Stationarity typically simplifies modelling and can improve estimation of parameters. However, many processes in real life are not stationary, and then making such assumptions can lead to wrong answers.

(b) From

$$y_t(s) = \beta_1 y_{t-\Delta_t}(s) + \beta_2[y_{t-\Delta_t}(s-\Delta_s) + y_{t-\Delta_t}(s+\Delta_s)] + \delta_t(s)$$

We have

$$\begin{aligned}
y_t(s) =&\beta_1[\beta_1 y_{t-2\Delta_t}(s) + \beta_2[y_{t-2\Delta_t}(s-\Delta_s) + y_{t-2\Delta_t}(s+\Delta_s)] + \delta_{t-\Delta_t}(s)]+\\
&\beta_2[\beta_1 y_{t-2\Delta_t}(s-\Delta_s) + \beta_2[y_{t-2\Delta_t}(s-2\Delta_s) + y_{t-2\Delta_t}(s)] + \delta_{t-\Delta_t}(s-\Delta_s)]+\\
&\beta_2[\beta_1 y_{t-2\Delta_t}(s+\Delta_s) + \beta_2[y_{t-2\Delta_t}(s) + y_{t-2\Delta_t}(s+2\Delta_s)] + \delta_{t-\Delta_s}(s+\Delta_s)]+\\
&\delta_t(s)\\
=&[\beta_1^2 + 2\beta_2^2]y_{t-2\Delta_t}(s) + 2\beta_1\beta_2[y_{t-2\Delta_t}(s-\Delta_s) + y_{t-2\Delta_t}(s+\Delta_s)]+\\
&\beta_2^2[y_{t-2\Delta_t}(s-2\Delta_s) + y_{t-2\Delta_t}(s+2\Delta_s)]+\\
&\beta_1\delta_{t-\Delta_t}(s) + \beta_2[\delta_{t-\Delta_t}(s-\Delta_s) + \delta_{t-\Delta_s}(s+\Delta_s)] + \delta_t(s)
\end{aligned}$$

giving a much more complex structure. In particular, with larger time-steps we would expect $y_t(s)$ to depend on more previous points and also that the noise structure becomes more complex.

(c) For a separable covariance function we must have that

$$\text{Cov}[y_t(s), y_{t-k}(v)] = \text{Cov}[y_t(s), y_t(v)] \times C^t(k)$$

which in particular implies that

$$\text{Cov}[y_t(s), y_{t-2}(v)] = \tfrac{C^t(2)}{C^t(1)}\text{Cov}[y_t(s), y_{t-1}(v)]$$

(Continued on page 3.)

From the model and the results in (b) we have that

$$\text{Cov}[y_t(s), y_{t-1}(v)] = 0 \qquad \text{for } |s-v| = 2\Delta_s$$

while

$$\text{Cov}[y_t(s), y_{t-2}(v)] \neq 0 \qquad \text{for } |s-v| = 2\Delta_s$$

which shows that it can not be separable.

(d) We call these hierarchical dynamical spatio-temporal models.

It is very difficult to model multivariate models for count data directly, while doing it indirectly through a latent Gaussian process is much simpler.

(e) The main problem now is that we for prediction are interested in

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}) =& \frac{p(\boldsymbol{Y};\boldsymbol{\theta})p(\boldsymbol{Z}|\boldsymbol{Y};\boldsymbol{\theta})}{p(\boldsymbol{Z};\boldsymbol{\theta})} \\ =& \frac{p(\boldsymbol{Y};\boldsymbol{\theta})p(\boldsymbol{Z}|\boldsymbol{Y};\boldsymbol{\theta})}{\int_{\mathbf{Y}'} p(\boldsymbol{Y}';\boldsymbol{\theta})p(\boldsymbol{Z}|\boldsymbol{Y}';\boldsymbol{\theta})d\boldsymbol{Y}'} \end{aligned}$$

where the integral in the denominator will be very difficult to evaluate. This denominator also corresponds to the likelihood function, making inference on parameters equally difficult.

In the course we have used the INLA method/package for doing calculations within such models. An alternative (which the results in the textbook is based on) is Markov Chain Monte Carlo methods. The latter will typically be much slower but is also more general.

END