UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

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

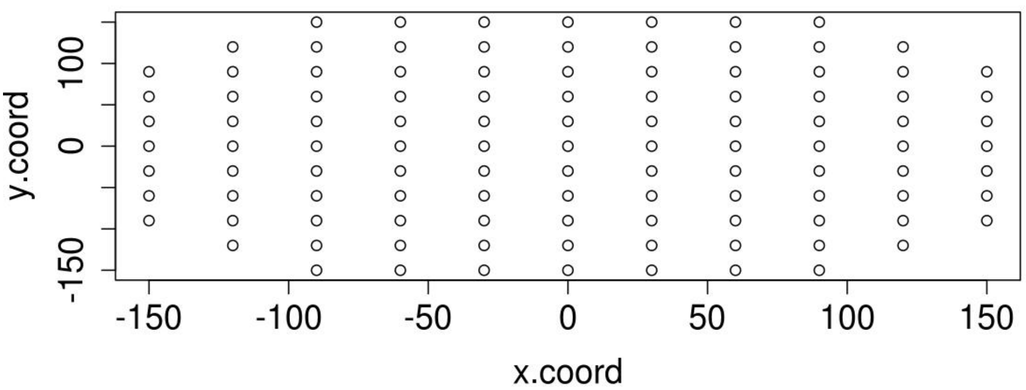
This examination set consists of 3 pages.

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.

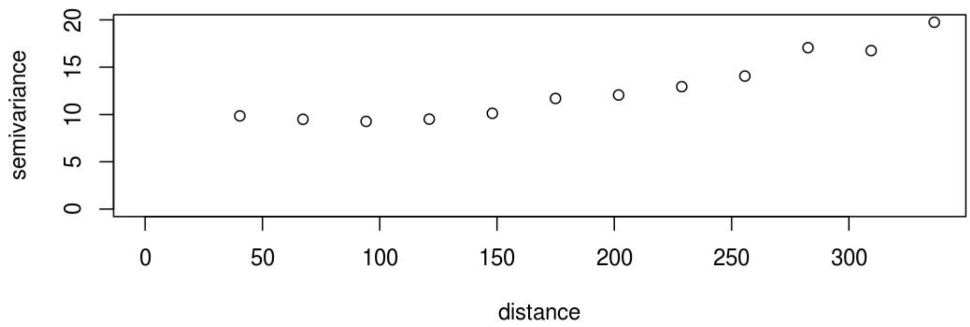
In this exercise we will consider a pure spatial data set on soil physical and chemical data collected on a field in the Weissenstaedter Becken, Germany. We will concentrate on the `moisture` content which is measured in Kg/Kg*100%. The plot below displays the spatial sites, showing that the data are collected on a regular grid



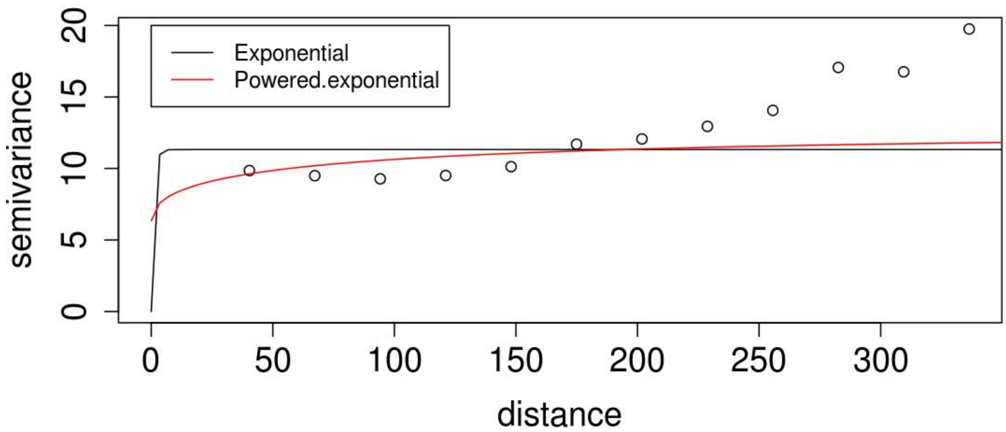
(Continued on page 2.)

We will consider the use of spatial statistics for this dataset.

(a) Below is a (semi-)variogram of the data shown (assuming an isotropic and stationary structure). Discuss the use of variograms in general and comment on the particular shape of the variogram in this case.



(b) In the plot below, an exponential and a powered exponential covariance function were fitted to the data. The log-likelihood values for these covariance functions were -287.2 and -286.9, repsectively. In comparisson, a non-spatial model gave a likelihood value equal to -286.9. Based on these results, do you find that any of these fits are satisfactory?



(c) An alternative to geostatistical models are CAR models. Discuss the benefits in using CAR models in general.

(d) A possible CAR based model is

$Z_i = \mu + \delta_i + \varepsilon_i$

(Continued on page 3.)

where $\{\delta_i\}$ follows a CAR structure (with a sum-constraint):

$$[\delta_i|\delta_j, j \neq i] = N(\frac{1}{N_i}\sum_{j\sim i}\delta_j, \sigma_\delta^2/N_i)$$

where N_i is the number of neighbors of site i and $\sum_{j\sim i}$ means the sum over all sites j that are neighbors to site i. Further, $\{\varepsilon_i\}$ are iid and $N(0,\sigma_\varepsilon^2)$.

A fit to this model using the four nearest grid points as neighbors gave $\hat{\sigma}_\delta^2 = 2.751$ and $\hat{\sigma}_\varepsilon^2 = 6.712$. Further, the loglikelihood value was -258.36 in this case.

Discuss these results and compare with previous results.

(e) In the model for $\{\delta_i\}$ a sum-constraint was imposed when doing the analysis. Why is this a reasonable restriction to impose?

Problem 2.

Consider a discritized version of a spatiotemporal process $\{y_t(s)\}$ where s is a spatial point (in one dimension) and t is a temporal point:

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

where $\{\delta_t(s)\}$ is a process that is independent in time but possibly correlated in space. We will also assume that β_1, β_2 are chosen such that the process is stationary in time.

(a) What do we mean by stationarity in time? Discuss reasons for assuming/not assuming such a property.

(b) Write $y_t(s)$ as a function of $\{y_{t-2\Delta_t}(v)\}$. Based on this, discuss why discretization with a small Δ_t is preferable. What are the disadvantages with having a small Δ_t?

(c) Is this model separable? Justify your answer.

(d) Assume now that we have observations $Z_t(s)$ that, conditional on the Y-process are independent and Poisson distributed with expectations $\exp(Y_t(s))$.

What do we call these types of models. What are the benefits in this kind of modelling compared to a direct multivariate model on the observed Z's?

(e) Discuss the computational problems related to models of the type suggested in (*d*) and discuss briefly possible computational tools for performing inference.

END