Exercises for STK4150/9150 - Environmental and spatial statistics

Odd Kolbjørnsen, Geir Storvik

Spring 2017

Exercise 1 (AR(1) models)
Consider the AR(1) process

$$x_1 = \mu_1 + \sigma_1 \varepsilon_1$$
$$x_i = \mu + a(x_{i-1} - \mu) + \sigma \varepsilon_i, \quad i = 2, 3, ...$$

where $\varepsilon_1, \varepsilon_2, ...$ are iid standard normal variables and $\sigma_1 = \sigma/\sqrt{1-a^2}$. We will assume $0 < a < 1$ and $\sigma > 0$.

(a). Show that if $\mu_1 = \mu$ and $\sigma_1 = \sigma/\sqrt{1-a^2}$, then

$$E(x_i) = \mu$$
$$\text{Var}(x_i) = \text{Var}(x_1)$$
$$\text{Cov}(x_i, x_{i+\tau}) = \text{Var}(x_i)a^\tau$$

for all $i > 1$.

Hint: Show it recursively.

A random vector $\mathbf{x} = (x_1, \cdots x_n)^T$ is said to have the multivariate normal distribution if it satisfies the following condition:
Every linear combination of its components

$$y = a_1 x_1 + \cdots + a_n x_n$$

is normally distributed. That is, for any constant vector $\mathbf{a} \in \mathcal{R}^n$, the random variable $\mathbf{y} = \mathbf{a}^T \mathbf{x}$ has a univariate normal distribution.

If $\boldsymbol{\mu}$ is the expectation vector of $\mathbf{X}$ and $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{x}$, we say that $\mathbf{x} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or just $\mathbf{x} \sim \text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(b). Show that $\mathbf{x} = (x_1, ..., x_n)$ is $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is a matrix with elements

$$\Sigma_{i,i} = \sigma_1^2$$
$$\Sigma_{i,i+\tau} = \sigma_1^2 a^\tau$$

Maximum likelihood estimates for (a, μ, σ^2) can be obtained by maximizing

$$L(a, \mu, \sigma^2) = f(\mathbf{x}|a, \mu, \sigma^2)$$

In general these equations become a bit complicated. We will therefore do two "tricks" to make it simpler.

The first trick is to define $\tilde{\mu} = \mu(1 - a)$ such that

$$x_i = \tilde{\mu} + ax_{i-1} + \sigma\varepsilon_i, \quad i = 2, 3, ...$$

The second trick is to consider an approximate likelihood function defined as

$$\tilde{L}(a, \mu, \sigma^2) = f(\mathbf{x}_{-1}|x_1, a, \mu, \sigma^2),$$

where $\mathbf{x}_{-1} = (x_2, ..., x_n)$. That is we assume x_1 do not contain any information about (a, μ, σ^2) (this is not true in practice, but if n is large, the information loss is negligible).

(c). Write down the log-likelihood based on $\tilde{L}$ and use this to derive a linear equation system that the ML estimates for $(a, \tilde{\mu})$ must satisfy. How can you obtain an estimate of μ from this?

Also find an explicit expression for the ML estimate for σ^2 as a function of the other parameters.

Luckily, we do not have to implement these equations because most packages have suitable routines for doing this. We will see how we can perform estimation in R using the routine *gls* within the library *nlme*.

(d). Show that

$$\mathrm{cor}(x_i, x_j) = \exp(-|i - j|/\theta)$$

for a proper choice of θ. This correlation structure is usually named the exponential correlation function

(e). At the course home page, there is a file *x_sim.dat* which contains a simulated time series from the AR(1) model with $a = 0.9, \mu = 2$ and $\sigma = 1$.

The following R code can be used for estimation of the parameters involved:

```
tim = 1:100
fit = gls(x~1,correlation=corAR1(form=~tid))
summary(fit)
```

Take a look on the help page for the *gls* and make sure you understand the call above.

Read out the parameter estimates from the output of the call and compare with the true values.

(f). Try also out the command

```
fit2 = gls(x~1,correlation=corExp(form=~tid))
```

and from exercise (d) try to make the connection between the phi parameter in `fit` and the range parameter in `fit2`.

Exercise 2

Consider the state space model

$$Y_1 \sim N(\mu_1, \sigma_1^2)$$
$$Y_t = \alpha Y_{t-1} + W_t, t = 2, 3, ... \qquad\qquad W_t \overset{ind}{\sim} (0, \sigma^2)$$
$$Z_t = Y_t + \eta_t, t = 1, 2, 3, ... \qquad\qquad \eta_t \overset{ind}{\sim} (0, \tau^2)$$

where all $\{W_t\}$ also are independent of all $\{\eta_t\}$. We will assume that $\mu_1 = 0$ and $\sigma_1^2 = \sigma^2/(1 - \alpha^2)$ in the following (corresponding to 2. order stationarity).

Assume we observe $\mathbf{Z} = (Z_1, ..., Z_T)^T$. We will in this exercise consider methods for calculating the density $f(z_1, ..., z_T; \boldsymbol{\theta})$, corresponding to the likelihood $L(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the parameters involved.

The following results from probability theory will be useful: Assume X_1 and X_2 are arbitrary random variables while X_3 may be a random variable or a vector of random variables. Then

$$E[X_1|X_2] = E[E[X_1|X_2, X_3]]$$
$$\text{var}[X_1|X_2] = \text{var}[E[X_1|X_2, X_3] + E[\text{var}[X_1|X_2, X_3]]$$
$$\text{cov}(X_1, X_2) = E[\text{cov}(X_1, X_2)|X_3] + \text{cov}[E(X_1|X_3), E(X_2|X_3)].$$

Further, if $\mathbf{X} = (X_1, X_2)$ is bivariate normally distributed with expectation vector $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$ then

$$E(X_1|X_2 = x_2) = \nu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \nu_2)$$
$$\text{var}(X_1|X_2 = x_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

This result also holds if $\mathbf{X}$ is conditioned on some other variable(s).

(a). Show that $\mathbf{Z}$ is multivariate normal distributed.

Hint: Use the results for the AR(1) model in the previous model.

(b). Find the expectation of Z_t.

Hint: Use the rule of double expectation and condition on Y_t.

3

(c). Show that

$$\text{cov}(Z_t, Z_{t+\tau}) = \text{cov}(Y_t, Y_{t+\tau}) + \tau_t^2 I(\tau = 0)$$

where $I(\cdot)$ is the indicator function.

Hint: Use the rules above with $X_3 = (Y_t, Y_{t+\tau})$.

The results above combined with that the multivariate density is given by

$$f(\mathbf{z}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{T/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\{-0.5(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu})\}$$

can be used to calculate the density for any T. A problem with this approach is however that when T is large, the matrix calculations involved can be computationally costly, and cheaper alternatives would be preferable. We will look at how calculations can be performed recursively.

For the following, define

$$\hat{y}_{t|s} = E(Y_t|Z_1, ..., Z_s) \qquad\qquad \hat{z}_t = E(Z_t|Z_1, ..., Z_{t-1})$$
$$P_{t|s} = \text{var}(Y_t|Z_1, ..., Z_s) \qquad\qquad S_t = \text{var}(Z_t|Z_1, ..., Z_{t-1})$$

where $s = 0$ correspond to conditioning on the empty set. Note that $\hat{y}_{1|0} = \mu_1$ and $P_{1|0} = \sigma_1^2$ is given directly by the model.

(d). Show that

$$f(\mathbf{z}; \boldsymbol{\theta}) = \phi(z_1; \mu_1, \sigma_1^2 + \tau^2) \prod_{t=2}^{T} \phi(z_t; \hat{z}_t, S_t)$$

where $\phi(z; \mu, \sigma^2)$ is the univariate normal density for $Z = z$ having mean μ and variance σ^2.

(e). Show that

$$\hat{z}_t = \hat{y}_{t|t-1}, \qquad\qquad S_t = P_{t|t-1} + \tau_t^2 \qquad\qquad (*)$$

(f). Show that

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} |z_1, ..., z_{t-1} \sim \text{MVN}\left(\begin{pmatrix} \hat{y}_{t|t-1} \\ \hat{z}_t \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & P_{t|t-1} \\ P_{t|t-1} & S_t \end{pmatrix}\right)$$

and use this to show that

$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + \frac{P_{t|t-1}}{S_t}(z_t - \hat{y}_{t|t-1}), \qquad\qquad P_{t|t} = P_{t|t-1} - \frac{P_{t|t-1}^2}{S_t}$$

Show that these equations are equivalent to

$$K_t = \frac{P_{t|t-1}}{S_t}$$
$$\hat{y}_{t|t} = \hat{y}_{t|t-1} + K_t(z_t - \hat{y}_{t|t-1}), \quad P_{t|t} = P_{t|t-1}(1 - K_t) \qquad\qquad (**)$$

4

(g). Show that

$$\hat{y}_{t+1|t} = \alpha \hat{y}_{t|t}, \qquad\qquad\qquad P_{t+1|t} = \alpha^2 P_{t|t} + \sigma^2 \qquad\qquad (\text{***})$$

(h). Consider now a case where observation Z_t is missing. Argue that in this case,

$$\hat{y}_{t|t} = \hat{y}_{t|t-1}, \qquad\qquad\qquad P_{t|t} = P_{t|t-1}$$

while there will be no contribution in the likelihood at this time-step.

(i). Argue why equations $(*),(**)$ and $(***)$ can be used to calculate the likelihood recursively.

Equations $(*),(**)$ and $(***)$ are actually the univariate version of the famous Kalman filter which you now have verified!

(j). At the course webpage there is an R-script, `exer2_kalman.R`, including simulation of data (using the parameter values $\mu_1 = 0, \sigma_1 = 0.1, \alpha = 0.9, \beta = 1, \sigma = 0.5, \tau = 1$) and performing Kalman filtering.

Run the Kalman filter with the true parameter values and plot the predictions $\hat{y}_{t|t}$ together with a confidence band (95% prediction intervals). You should also include the observation points in the plot.

Comment on the results. In particular, what happens with the prediction intervals when there are several missing values in a row.

(k). In the script there are also commands for calculating the log-likelihood for different values of α, keeping the other parameters fixed. Look at the plot you get and comment on the results.

(l). Finally, the script contain a call to the optimization routine `nlm` which can be used to maximize the likelihood with respect to all the parameters involved. Look at the final estimates and comment on the differences from the true values.

Run the Kalman filter again, now using the estimated parameter values. Make a plot similar to the one made in the routines and discuss differences.

Exercise 3 (Autocorrelation functions)
(a). Consider again the data set from exercise 1 which was simulated data from an AR(1) model. The autocorrelation function (ACF) can be estimated in R by the `acf` function. Perform the commands

```
x.sim = scan("x_sim.dat")
acf(x.sim)
```

The plot shows estimates of ACF for different lags (lag is here defined as the difference in time). The horizontal dashed lines shows confidence bands which can be interpreted in that estimates outside these bands are significantly different from zero.

How many significant terms are there? How do the ACF decay with increasing lag?

(b). Consider now data simulated from an MA(q) process which can be simulated (for $q = 2$) by

```
n=100;q=2;beta = c(0.7,0.3)
eps = rnorm(q+n)
x = rep(NA,q+n)
for(i in 1:n)
  x[q+i] = eps[q+i]+sum(beta*eps[q+i-1:q])
x = x[-c(1:q)]    #Removing the fist q which are just NA
acf(x)
```

where again the `acf` routine is used to plot the autocorrelation function.

Repeat the commands above several times. Try also out different values of q (with your own choice of β's). What is typically the number of lags for which the correlations are significantly different from zero? Relate this to q.

Discuss how this can can be used to identify the order of the MA process.

It is more difficult to read the order of an AR process from the ACF. A useful alternative in that case is the partial autocorrelation function (PACF) which is defined as

$$\rho_\tau = \text{cor}[x_i, x_{i+\tau}|x_{i+1}, ..., x_{i+\tau-1}]$$

(c). Show that for an AR(p) process $\rho_\tau = 0$ for $\tau > p$

(d). The PACF can also be estimated by the `acf` routine using the option `type="partial"`, i.e. try out

```
acf(x.sim,type="partial")
```

for the data set from exercise 1. How many significant lags do you get in this case? How does that fit with the true model?

Exercise 4

The data set `Hawaii` (available from the course webpage) contains abundance of a bird species at an island in Hawaii from 1956 to 2003 in addition to the amount rainfall within each year. A possible regression model in this case is

$$Y_i = \beta_0 + \beta_1 i + \beta_2 R_i + \varepsilon_i$$

where i is a year index, R_i is rainfall in year i while Y_i in this case is the square root of abundance. (in the R-language)

```
Hawaii = read.table("Hawaii.txt")
library(nlme)
M0 = gls(sqrt(Abundance)~Year+Rainfall,data=Hawaii)
acf(resid(M0))
```

where the last command plots the autocorrelation function on the residuals.

(a). Assume first that all the noise terms are independent. R-commands for performing the regression analysis is given by

```
Hawaii = read.table("Hawaii.txt")
library(nlme)
M0 = gls(sqrt(Abundance)~Year+Rainfall,data=Hawaii)
acf(resid(M0))
```

where the last command also plot the ACF. Try out these commands and plot the ACF function of the residuals. Do they look independent?

(b). An alternative analysis is to assume that the noise terms follow an AR(1) model, i.e.

$$\varepsilon_i = \phi \varepsilon_{i-1} + \eta_i$$

where now the η_i's are independent. This model can be analyzed by the R-command

```
M1 = gls(sqrt(Abundance)~Year+Rainfall,data=Hawaii,
  correlation=corAR1(form=~Year))
```

What is the estimate of ϕ? How can this be used to calculate estimates of η_i?

(c). Calculate estimates of η_i and make a plot of the ACF for these terms. Do the plot indicate that the AR(1) model is reasonable?

Exercise 5 (Estimation of variograms)
Within the field of geostatistics, it has been more common to talk about variograms rather than covariance functions. The variogram is defined as

$$2\gamma_Y(\mathbf{h}) = \text{var}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})]$$

where we indirectly have assumed that the variogram is stationary (do not depend on $\mathbf{s}$).

(a). Show that the variogram can be calculated from the covariance function.

(b). Show that the variogram is equal to zero at $\mathbf{h} = 0$. Find also $\lim_{||\mathbf{h}||\to 0} 2\gamma_Y(\mathbf{h})$ and relate this to the nugget effect.

(c). Assume the covariance function $C_Y(\mathbf{h})$ converges to zero as $||\mathbf{h}||$ increases to infinity. Find the limit value of the variogram $2\gamma_Y(\mathbf{h}$ as $||\mathbf{h}||$ increases to infinity. This limit value is usually called the (partial) sill of the variogram.

The variogram can be estimated by

$$2\widehat{\gamma}_Z^0(h) = \mathrm{ave}\{(Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2; \|\mathbf{s}_i - \mathbf{s}_j\| \in T(h)\}$$

Here $T(h)$ can be equal to $\{h\}$ in which case the variogram only can be estimated for the distances that are available in the data or it can be a tolerance region around h (e.g. $h \pm \Delta$ where Δ is small). In principle we could program this estimate by yourself but usually it is easier to use some available routine. We will here consider the routine *variogram* within the R package *gstat*.

For illustration we will use the sea surface temperature data from the book at February 98 (see fig 5.4). These data are available from the course web page and can be read by the R command

```
d = read.table("SST_feb98.txt",header=T)
```

(d). Calculate and plot the variogram through the command

```
library(gstat)
plot(variogram(temp~1,d,locations=~x+y)
```

(Note that it is the estimated semivariogram $\widehat{\gamma}_Z^0(h)$ that is plotted.)

(e). Fitting an empirical variogram to a parametric function can be performed through the command *fit.variogram* routine. Run the commands

```
vgm1 = variogram(temp~1,d,locations=~x+y)
res = fit.variogram(vgm1,vgm(psill=1,"Mat",range=300,nugget=1,kappa=0.5))
```

Here the second input to the *fit.variogram* both specifies the variogram model (Matern in this case) and starting values on the parameters involved (see the help function for the *fit.variogram*). Note that the range parameter corresponds to θ_1 and kappa corresponds to θ_2 in equation (4.6) in the book. Note also that the routine uses a parameter called partial sill (*psill*). This is defined as

$$\mathrm{psill} = \mathrm{Var}[Y(s)] - \mathrm{nugget}$$

which is the part of the total variance that is not the nugget effect. This can be used to calculate the nugget effect.

What are the estimates of the parameters in this case?

You can also plot the empirical variogram together with the estimated parametric variogram through the commands

```
plot(vgm1$dist,vgm1$gamma)
lines(variogramLine(vgm(res[2,2],"Mat",res[2,3],kappa=0.5),
    maxdist=max(vgm1$dist)))
```

Do the fitted variogram look reasonable in this case?

(f). The `fit.variogram` actually do not estimate kappa but only uses the specified initial value. Try out different values of kappa to get a better fit (it is most important that the fit is good for small values of h).

Exercise 6 (Estimation under dependence)

Assume $\mathbf{Z}$ is a random vector of length m following a multivariate normal distribution $MVN(\mu\mathbf{1}, \sigma^2\mathbf{C})$ where $\mathbf{C}$ in this case is a correlation matrix (the diagonal elements are all equal to one). Our interest will be in estimation of μ and σ^2, assuming $\mathbf{C}$ is known and invertible. We will look at two ways of obtaining the ML estimates.

(a). Show that the density for $\mathbf{Z} = \mathbf{z}$ can be written as

$$f(\mathbf{z}; \mu, \sigma^2) = \frac{1}{(2\pi)^{m/2}|\sigma^2\mathbf{C}|^{1/2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{z} - \mu\mathbf{1})^T\mathbf{C}^{-1}(\mathbf{z} - \mu\mathbf{1})\}$$

(b). Show that the maximum likelihood estimates for μ and σ^2 are given by

$$\hat{\mu} = \frac{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{z}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}$$
$$\hat{\sigma}^2 = \frac{1}{m}(\mathbf{z} - \hat{\mu}\mathbf{1})^T\mathbf{C}^{-1}(\mathbf{z} - \hat{\mu}\mathbf{1})$$

The vector $\frac{\mathbf{1}^T\mathbf{C}^{-1}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}$ can be considered as a set of weights given to each observation.

(c). Show that for independent data, we get the standard estimates.

(d). As an alternative, write $\mathbf{C}^{-1} = \mathbf{L}^T\mathbf{L}$ (such a matrix $\mathbf{L}$ always exists since $\mathbf{C}$ is symmetric). Define $\widetilde{\mathbf{Z}} = \mathbf{L}\mathbf{Z}$. Show that $\widetilde{\mathbf{Z}} \sim \text{MVN}(\mu\mathbf{L}\mathbf{1}, \sigma^2\mathbf{I})$ where $\mathbf{I}$ is the identify matrix of dimension $m \times m$. Argue why this corresponds to an ordinary linear regression setting with $\mathbf{L}\mathbf{1}$ corresponding to the design matrix.

(e). Find the ML estimates for μ and σ^2 based on $\widetilde{\mathbf{Z}}$ and show that this corresponds to the expressions above.

Consider now a concrete example where the covariances are defined through a stationary covariance function $C_Z(h) = \exp(-0.5||h||)$. Assume further you have observed points at positions $\{(0.95, 1), (1, 0.95), (1, 1.05), (1.05, 1), (10, 10)\}$.

(f). Calculate the weight vector in this case. Argue why these weights are reasonable

Exercise 7 (Wiener processes)

Variograms can be stationary even in cases where the covariance function not is stationary. One such example is the Wiener process. A Wiener process $\{W(t), t \in \mathcal{R}^+\}$ is a continuous time process characterized by the three properties

9

- $W(0) = 0$

- The function $t \to W(t)$ is almost surely continuous

- W_t has independent increments with $W(t) - W(s) \sim N(0, t - s)$

(a). Show that $\text{cov}(W(s), W(t)) = \min(s, t)$ so that the covariance function is not stationary

(b). Show that the variogram $\text{Var}[W(s + h) - W(s)] = |h|$ so that the variogram is stationary

Exercise 8 (Boreality data)

In the introductory lecture we discussed a data set giving the number of species that belong to a set of boreal species divided by the total number of species at a site (*Bor* below). As an explanatory variable we considered an index of wetness (*Wet* below). In the introductory lecture we discusses the problems with the independence assumption. The following commands reproduce the analysis made then:

```
Boreality = read.table("Boreality.txt",header=T)
Boreality$Bor<-sqrt(1000*(Boreality$nBor+1)/(Boreality$nTot))
library(nlme)
B.lm<-gls(Bor~Wet,data=Boreality)
summary(B.lm)
plot(Boreality$Wet,Boreality$Bor)
E<-resid(B.lm)
library(gstat)
mydata<-data.frame(E,Boreality$x,Boreality$y)
coordinates(mydata)<-c("Boreality.x","Boreality.y")
bubble(mydata,"E",col=c("black","grey"),
main="Residuals",xlab="X-coordinates",
ylab="Y-coordinates")
Vario1 = variogram(E ~ 1, mydata,locations=seq(0,2500,length=20))
plot(Vario1$dist,Vario1$gamma)
```

In this exercise we will consider a spatial model.

(a). Repeat the commands above and argue why a spatial model is necessary.

(b). A model with a Gaussian covariance function can be fitted by the following commands (without and with a nugget effect):

```
fit.gaus0 = gls(Bor~Wet,correlation=corGaus(form=~x+y),data=Boreality)
fit.gaus = gls(Bor~Wet,correlation=corGaus(form=~x+y,nugget=TRUE),data=Boreality)
```

(These calls may be a bit time-consuming.)

A possible model comparison criterion is the AIC criterion, defined by

$$AIC = -2 * \log.\text{likelihood} + 2 * \text{npar}$$

Here, the best model is considered to be the one with the lowest AIC value.

In R you can easily get the AIC values by the command

```
AIC(B.lm,fit.gaus0,fit.gaus)
```

Perform these commands and discuss your results.

(c). Consider now the exponential covariance function (both with and without nugget effect). Do this model give a better fit to the data? (Use the AIC criterion again.)

(d). Now calculate the fitted variogram for the best model you have found and draw this line on top of the empirical variogram you fitted above. Do the fit seem reasonable?

Hint: Look at the help page for the appropriate covariance function in order to calculate this. You will then need to convert from covariance function to variogram. (Note: I was not able to find out how to extract nugget effect and range from the fitted objects and had to insert these values manually. If you find out how to do this, tell me!)

We will now consider the problem of prediction. Assume we want to predict at the locations $\mathbf{s}_{p1} = (2197.922, 2086.046)$ with corresponding value of Wet $= -0.0252$ and $\mathbf{s}_{p2} = (1940.748, 1778.798)$ with corresponding value of Wet $= -0.02594$. A reasonable predictor is the conditional expectation using the "plug-in" rule. The plug-in rule means that we treat the estimates of the parameters as the true values (and ignore the uncertainty involved here).

Use $Z(\mathbf{s})$ for the response at site $\mathbf{s}$. Denoting $\mathbf{Z}_1 = (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_m))$ the set of observations and $\mathbf{Z}_2 = (Z(\mathbf{s}_{p1}), Z(\mathbf{s}_{p2}))$ the prediction points, we are interested in prediction of $\mathbf{Z}_2$ given $\mathbf{Z}_1$ (so we here are actually interested in predicting observations at new sites).

(e). Calculate the expectation vector $\boldsymbol{\mu}_1 = E[\mathbf{Z}_1]$ and $\boldsymbol{\mu}_2 = E[\mathbf{Z}_2]$

(f). Calculate the covariance matrix of $\mathbf{Z}_1$.

Hint: You will here need to calculate all the distances between the positions for $\mathbf{Z}_1$ and then use the parametric form of the covariance function to obtain the matrix. The R function *dist* can be useful here.

Denote this matrix by $\boldsymbol{\Sigma}_{11}$.

(g). Calculate the covariance matrix of $\mathbf{Z}_2$.

Denote this matrix by $\boldsymbol{\Sigma}_{22}$.

(h). Also calculate the $m \times 2$ matrix $\boldsymbol{\Sigma}_{12}$ which contains all the covariances between elements of $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

(i). Argue why

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T$ is the covariance matrix of $(\mathbf{Z}_1, \mathbf{Z}_2)$.

(j). Calculate the quantities

$$\begin{aligned} \widehat{\mathbf{Z}}_2 &= \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Z}_1 - \boldsymbol{\mu}_1) \\ \mathbf{V}_2 &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{aligned}$$

and argue why these are useful for performing predictions.

(k). Plot $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$ against the the distances between $\mathbf{s}_{pi}$ and the observation points (one plot for each prediction point). Discuss this plot.

Assume now that our real interest is in a latent process $\{Y(\mathbf{s})\}$ where $Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), i = 1, ..., m$ and where the ε's are independent measurement noise. Assume further that the nugget effect corresponds to the measurement noise in this case (so that the latent Y process has a covariance function that is continuous in zero). Denote $\mathbf{Y}_2 = (Y(\mathbf{s}_{p1}), Y(\mathbf{s}_{p2}))$.

(l). Argue why $\widehat{\mathbf{Z}}_2$ can be used as a prediction of $\mathbf{Y}_2$

(m). Argue also that uncertainty measures about $\mathbf{Y}_2$ can be obtained by subtracting the nugget effect from the diagonal of $\mathbf{V}_2$. Compare these with the uncertainties about $\mathbf{Z}_2$.

Exercise 9 (Gaeten & Guyon (2009) exercise 1.8)
For a stationary process $Y(\cdot)$ on $\mathcal{R}$, with covariance function $C_Y(\cdot)$, we observe $Y(0) = -1$ and $Y(1) = 1$. (Note that this correspond to a case where we do not have any measurement errors, i.e $Z(0) = Y(0), Z(1) = Y(1)$.)

(a). Show that $\hat{\mu}_{gls} = 0$ in this case.

(b). Show that ordinary kriging gives

$$\widehat{Y}(s) = \frac{C_Y(s-1) - C_Y(s)}{C_Y(0) - C_Y(1)}$$

Hint: Calculate $\mathbf{C}_Y^{-1}\mathbf{Y}$ first.

(c). One can also show that the variance of the prediction error at s is

$$\sigma^2_{Y,ok}(s) = C_Y(0)\left(1 - \frac{(C_Y(s)+C_Y(s-1))^2}{C_Y(0)^2 - C_Y(1)^2}\right) + 2\frac{C_Y(s)C_Y(s-1)}{C_Y(0)-C_Y(1)}$$

but this you do not need to show. Show however that $\sigma^2_{Y,sk}(s) = 0$ for $s = 0, 1$ and argue that this is reasonable.

(d). Draw the graphs of $s \to \widehat{Y}(s)$ and $s \to \sigma^2_{Y,sk}(s)$ for $s \in [-3, 3]$ when C_Y is the exponential covariance function with $C_Y(0) = 1$ and $\theta_1 = 1$. as well as the Gaussian covariance function with the same parameters. Comment on these graphs.

Exercise 10 (Bayesian analysis)
Assume $z_1, ..., z_m$ are iid, $z_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$. Our interest is in estimating μ through a Bayesian approach, we will for simplicity assume σ^2 known. As a prior, we will assume $\mu \sim N(\mu_0, k\sigma^2)$ where k can be seen as the relative uncertainty in μ compared to the uncertainty in one observation.

(a). Show that

$$f(\mathbf{z}; \mu) = g(\bar{z}; \mu)h(\mathbf{z})$$

where $h(\cdot)$ is a function not depending on μ. $\bar{z}$ is in this case what we call a sufficient statistic.

(b). Show that

$$p(\mu|\mathbf{z}) \propto p(\mu|\bar{z})$$

(where $\propto$ means proportional to) and that this implies that

$$p(\mu|\mathbf{z}) = p(\mu|\bar{z})$$

Argue why this means that the posterior distribution for μ only depend on $\mathbf{z}$ through the sufficient statistic $\bar{z}$.

(c). Show that

$$\begin{pmatrix} \mu \\ \bar{z} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_0 \\ \mu_0 \end{pmatrix}, \sigma^2 \begin{pmatrix} k & k \\ k & k + 1/m \end{pmatrix}\right)$$

(d). Show that

$$E[\mu|\mathbf{z}] = \mu_0 + \frac{k}{k + 1/m}(\bar{z} - \mu_0) = \frac{k^{-1}\mu_0 + m\bar{z}}{k^{-1} + m}$$

$$\text{Var}[\mu|\mathbf{z}] = \frac{\sigma^2}{k^{-1} + m}$$

13

(*e*). Assume in general we have some data $\mathbf{z}$ and our interest is in estimating $\boldsymbol{\theta}$ where $p(\boldsymbol{\theta})$ is some prior for $\boldsymbol{\theta}$. Assume further that

$$f(\mathbf{z}; \boldsymbol{\theta}) = g(\mathbf{T}; \boldsymbol{\theta})h(\mathbf{z})$$

where $h(\cdot)$ is a function not depending on $\boldsymbol{\theta}$. $\mathbf{T}$, possibly a vector, is in this case the sufficient statistic. Show that

$$p(\boldsymbol{\theta}|\mathbf{z}) = p(\boldsymbol{\theta}|\mathbf{T})$$

that is the posterior only depend on the sufficient statistic $\mathbf{T}$.

(*f*). Assume now in the normal iid example above that σ^2 also is unknown so that $\boldsymbol{\theta} = (\mu, \sigma^2)$. Show that $\mathbf{T} = (\bar{z}, \hat{\sigma}^2)$ where $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(z_i - \bar{z})^2$ is a sufficient statistic in this case.

In Bayesian analysis it is mathematically more convenient to work with the precision parameter $\tau = 1/\sigma^2$. Define now $\boldsymbol{\theta} = (\mu, \tau)$ and assume

$$p(\mu, \tau) = p(\mu|\tau)p(\tau)$$

where $\mu|\tau \sim N(\mu_0, k\tau^{-1})$ as before while τ is given a Gamma prior with parameters (a, b), i.e.

$$p(\tau) = \frac{b^a}{\Gamma(a)}\tau^{a-1}e^{-b\tau}$$

(*g*). Show that

$$p(\mu, \tau|\mathbf{z}) = p(\mu|\tau, \mathbf{z})p(\tau|\mathbf{z})$$

where $p(\mu|\tau, \mathbf{z})$ is the distribution you found in (*c*) (now expressed with τ instead of σ^2). Also show that

$$p(\tau|\mathbf{z}) \propto p(\tau)p(\mathbf{z}|\tau)$$

(*h*). Show that

$$\mathbf{z}|\tau \sim \text{MVN}\left(\mu_0\mathbf{1}, \sigma^2[\mathbf{I} + k\mathbf{1}\mathbf{1}^T]\right) = \text{MVN}\left(\mu_0\mathbf{1}, \tau^{-1}[\mathbf{I} + k\mathbf{1}\mathbf{1}^T]\right)$$

Hint: Argue first that $\mathbf{z}$ is MVN and then use rules about expectations and variances to find these.

(*i*). Convince yourself on that $[\mathbf{I} + k\mathbf{1}\mathbf{1}^T]^{-1} = [\mathbf{I} - \frac{k}{1+km}\mathbf{1}\mathbf{1}^T]$ and use this to show that

$$p(\mathbf{z}|\tau) \propto \tau^{m/2}\exp\{-0.5\tau(\mathbf{z} - \mu_0\mathbf{1})^T[\mathbf{I} - \frac{k}{1+km}\mathbf{1}\mathbf{1}^T](\mathbf{z} - \mu_0\mathbf{1})\}$$

(j). Show that

$$p(\tau|\mathbf{z}) \propto \tau^{\bar{a}-1}e^{-\bar{b}\tau}$$

for some properly chosen $\bar{a}, \bar{b}$ (which will depend on the data. Use this to argue that also the posterior of τ is a Gamma distribution.

Exercise 11 (Identifying acid-sensitive waters)

The data to be used in this exercise are a subset of data analyzed in Cruikshanks et al. [2006]. We follow Zuur et al. [2009] and only use the 2003 data, and several recordings are dropped. The original research sampled 257 rivers in Ireland. We quote Zuur et al. [2009]:

> One of the aims was to find a different tool for identifying acid-sensitive waters, which currently uses measures of pH. The problem with pH is that it is extremely variable within a catchment and depends on both flow conditions and underlying geology. As an alternative measure, the Sodium Dominance Index (SDI) is proposed as an indicator of the acid sensitivity of rivers. SDI is defined as the contribution of sodium (Na+) to the sum of the major cations. The motivation for this research is the increase in plantation forestry cover in Irish landscapes and its potential impacts on aquatic resources.

Of the 257 sites, 192 were non-forested and 65 were forested. The data are available in the *SDI2003.txt* file from the webpage and can be read into R through

```
SDI2003 = read.table("SDI2003.txt",header=TRUE)
```

A plot giving the observed sites with filled circles indicating forested sites and open circles the non-forested sites can be produced by the following commands:

```
MyPch<-vector(length=dim(SDI2003)[1])
MyPch[SDI2003$Forested==1]<-16
MyPch[SDI2003$Forested==2]<-1
plot(SDI2003$Easting,SDI2003$Northing,pch=MyPch)
```

(a). Consider first an ordinary regression model (in R terminology)

$$pH \sim 1 + SDI * log(Altitude) * factor(Forested)$$

that is an intercept and the three covariates *SDI*, *log(Altitude)*, *Forested* (the last one a factor with two levels) in addition to all pairwise interactions as well as the 3-way interaction between all three covariates.

Fit this model by the command

```
M1 = gls(pH~SDI*Forested*lAltitude,data=SDI2003,method="ML")
```

Model selection based on the AIC criterion can be performed through the command

```
stepAIC(M1)
```

which is available from the **MASS** library What is your final model?

Note: Above the option **method="ML"** is used for **gls** to use Maximum likelihood for estimation. for selection of fixed effects (covariates), this is the preferred estimation method. However, when testing for correlation structures, it is preferable to use REML which is the default method in **gls**.

(b). Inspect the spatial structure of the residuals from the **M1** model above. Is the assumption about independence reasonable? (Look at variograms in different directions in this case)

Also try to perform a permutation test for a formal testing of spatial dependence.

(c). Another way of validating the need for spatial structures is to include a covariance structure into the model. Try out different covariance functions. Use the AIC criterion for validating the need for spatial structures.

In case you get a different answer here compared to the previous exercise, discuss possible reasons for this.

(d). For the best spatial model you found, use the **stepAIC** routine to do model selection with respect to the covariates again (this part can be quite time-consuming) (note that you should use ML estimation in this part). Compare your results with the previous ones.

(e). Based on your results, make a conclusion on the importance of the various covariates.

(f). Consider now a more theoretical problem. Let

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_p x_{ij} + \varepsilon_i$$

with $\{\varepsilon_i\}$ being independent $N(0, \sigma^2)$ variables.

Assume however that covariate x_{ip} is missing and that you perform your analysis based on the simplified model

$$y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_p x_{ij} + \widetilde{\varepsilon}_i$$

Assume further that the missing covariate x_{ip} has a spatial structure with

$$\mathrm{cov}[x_{ip}, x_{i'p}] = C_p(||\mathbf{s}_i - \mathbf{s}_{i'}||)$$

where $\mathbf{s}_i$ is the spatial position of observation i.

Calculate the covariance structure of $\{\widetilde{\varepsilon}_i\}$.

Discuss the importance of this result related to the conclusions you made above.

Exercise 12 (Gaussian processes and precision matrices)

Assume $\mathbf{x} \sim \text{MVN}(\mu, \mathbf{\Sigma})$. We will assume $\mathbf{\Sigma}$ is of full rank. Assuming

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix},$$

we will make use of the results

$$E[\mathbf{x}_1|\mathbf{x}_2] = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$
$$\text{Var}[\mathbf{x}_1|\mathbf{x}_2] = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{11}$$

The precision matrix is defined as $\mathbf{Q} = \mathbf{\Sigma}^{-1}$. We will in this exercise discuss some properties of $\mathbf{Q}$. We will start with showing a general result on inverses of partitioned matrices.

(a). Assume

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}$$

where

$$\mathbf{Q}_{11} = [\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}]^{-1}$$
$$\mathbf{Q}_{12} = -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{Q}_{22}$$
$$\mathbf{Q}_{21} = -\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{Q}_{11}$$
$$\mathbf{Q}_{22} = [\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}]^{-1}$$

Show that $\mathbf{\Sigma}\mathbf{Q} = \mathbf{I}$ so that $\mathbf{Q}$ is the inverse of $\mathbf{\Sigma}$.

(b). Show that $\text{Var}[x_i|\mathbf{x}_{-i}] = Q_{ii}^{-1}$ where $\mathbf{x}_{-i}$ is the subvector of $\mathbf{x}$ excluding x_i.

(c). Show that $\text{E}[x_i|\mathbf{x}_{-i}] = \mu_i - Q_{ii}^{-1}\sum_{j \neq i} Q_{i,j}(x_j - \boldsymbol{\mu}_j)$.

(d). Explain why $Q_{i,j} = 0$ is equivalent to that x_i is conditional independent of x_j (where we condition on all other elements).

Hint: Consider the case $i = 1, j = 2$ and define $\mathbf{x}_1 = (x_1, x_2)$. Show that the conditional covariance matrix is diagonal if $Q_{1,2} = 0$.

Consider now the AR(1) process

$$x_1 \sim N(0, \sigma^2/(1 - a^2))$$
$$x_i = ax_{i-1} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2).$$

(e). Show that

$$x_1|\mathbf{x}_{-1} \sim N(ax_2, \sigma^2)$$

$$x_i|\mathbf{x}_{-i} \sim N(\tfrac{a}{1+a^2}(x_{t-1} + x_{t+1}), \tfrac{\sigma^2}{1+a^2}), \qquad\qquad i = 2, ..., n-1$$

$$x_n|\mathbf{x}_{-n} \sim N(ax_{n-1}, \sigma^2)$$

Use this to specify the structure of $\mathbf{Q}$.

(f). Specify the ratio of non-zero elements in $\mathbf{Q}$. Calculate this ratio for $n = 100, n = 1000$ and $n = 5000$.

For each of these values of n, compute the number of operations needed for calculating $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ by direct calculation and by utilization of the sparse structure of $\mathbf{Q}$.

(g). For calculations involving covariance or precision matrices, Cholesky decompositions are essential (for instance, determinants can be calculated through products of the diagonal elements in the Cholesky decomposition. Assuming Q contains the precision matrix, the time used for calculating the Cholesky decomposition in R can be found by the command

```
unix.time(chol(Q))
```

Try this out for the different values of n and record the time that is used.

(h). A sparse representation of Q and a corresponding Cholesky decomposition using sparse matrix computation can be obtained by the following commands:

```
library(Matrix)
n = dim(Q)[1]
i = c(1:n,1:(n-1),2:n)
j = c(1:n,2:n,1:(n-1))
x = c(diag(Q),diag(Q[-1,-n]),diag(Q[-n,-1]))
Q.sparse = sparseMatrix(i=i,j=j,x=x)
unix.time(chol(Q.sparse))
```

Here i and j are the indices giving the non-zero entries in the precision matrix while x are the corresponding values obtained from the diagonal and the two closes off-diagonals in Q.

Try these commands out for different values of n and compare the computing time with the ones obtained in the previous exercise.

Exercise 13
Assume $\mathbf{Y} = (Y_1, ..., Y_m)$ has a density

$$p(\mathbf{y}) \propto \exp\{-0.5[\sum_{j=1}^{m} \alpha_j(y_j - \mu_j)^2 + \sum_{j=1}^{n} \sum_{k \in N_j} \beta_{jk}(y_j - \mu_j)(y_k - \mu_k)]\}$$

where $N_j \subset \{1, ..., m\}$ is a neighborhoodset of j where we assume $k \in N_j \Rightarrow j \in N_k$.

(a). Show that $\mathbf{Y}$ is multivariate normal and find the expectation vector $\boldsymbol{\mu}$ and the precision matrix $\mathbf{Q}$.

(b). Show that $[Y_i|\mathbf{Y}_{-i}]$ only depend on $\{Y_j, j \in N_i\}$. Find the conditional distribution.

(c). Assume now that $Z_i|Y_i \sim \text{Poisson}(\exp(Y_i))$ and that conditional on $\mathbf{Y}$, all the Z_i's are independent. Show that the distribution of $\mathbf{Y}|\mathbf{Z}$ can be written on the form

$$p(\mathbf{y}|\mathbf{z}) \propto \exp\{-0.5[\sum_{j=1}^{m} g_i(y_i, z_i) + \sum_{j=1}^{n} \sum_{k \in N_j} \beta_{jk}(y_j - \mu_j)(y_k - \mu_k)]\}$$

for some appropriate functions g_i.

Use this to show that the distribution of $Y_i|\mathbf{Y}_{-i}, \mathbf{Z}$ only depend on $\mathbf{Y}$ through $\{Y_j, j \in N_i\}$.

Is the conditional distribution still Gaussian?

Exercise 14 (The geoR package)

(a). Go into the webpage `http://leg.ufpr.br/geoR/` for geoR and go through the three first sections of the illustration session. Make sure you understand the main structure of a `geodata` object.

We will now see how we can put a time-series into the geoR package. We will illustrate this through the AR(1) model

$$y_t = ay_{t-1} + \sigma\varepsilon_t$$

where $\varepsilon_t \sim N(0, 1)$ and all ε_t's are independent.

(b). Show that if $y_1 \sim N(0, \sigma^2/(1 - a^2))$, then all y_t have the same distribution.

Also calculate the semivariogram

$$2\gamma(h) = \text{Var}[Z(t + h) - Z(t)]$$

for $h = 0, 1, 2, ...$

(c). Simulate a time series of length $n = 1000$ from an AR(1) model using $\sigma^2 = 1$ and $a = 0.95$.

Put the data into a `geodata` object and use the geoR package to estimate the variogram.

Plot the (semi)variogram and superimpose the true function.

Hint: Put the second coordinate equal to 1 for all observations. Note that geoR plots by default the semivariogram.

(*d*). Repeat (*c*) several times and summarize your findings.

Exercise 15 (geoR (cont))
We will in this exercise consider further the data you simulated in exercise 14. Fitting a variogram to a parametric form can be based on maximum likelihood theory (the routine `gls` does that) or on least squares principles (which the `variofit` routine does). Assuming $\hat{\gamma}$ is a vector of variogram estimates at different distances (using for example the `variog` routine) and $\gamma(\theta)$ the corresponding parametric variogram for a set of parameter values θ, the (weighted) least squares approach is based on minimizing

$$[\hat{\gamma} - \gamma(\theta)]^T \mathbf{V}(\theta)^{-1}[\hat{\gamma} - \gamma(\theta)]$$

with respect to θ. Here $\mathbf{V}(\theta)$ is a weight function (equal to identity for ordinary least squares).

(*a*). Explore the R-routine `variofit` and use this routine to perform a least squares fit to the data assuming an exponential covariance function.

(*b*). Modify you call to `variofit` to perform weighted least squares. Use the estimates obtained by the ordinary least squares method as initial values.

Look on the help page to see how these weights are constructed.

(*c*). Explore the R-routine `likfit` and use this routine to obtain ML-estimates assuming again an exponential covariance function. Use estimates obtained by the ols or wls as initial values.

(*d*). Plot all variograms fitted in the same plot.

(*e*). Repeat the above a few times and summarize your findings.

Exercise 16 (Cliques)
Remember that cliques are subsets c of $\{1,...,m\}$ such that if $i, j \in c$ then $i \in N_i$ (and vice verse).

(*a*). Assume the AR(1) model. What are the cliques in this case?

(*b*). Consider then the AR(2) model. What are the cliques in this case?

(*c*). Consider the following graphical model
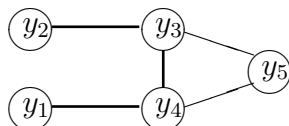
where a line between component means that they are neighbors. Specify the set of cliques in this case.

Exercise 17 (The Potts model)

Assume $Y_1, ..., Y_n$ are random variables over a lattice indexed by $i, ..., n$. Assume $Y_i \in \{1, ..., K\}$, i.e. Y_i is a categorical variable. Assume the probability for a particular configuration $\mathbf{y} = (y_1, ..., y_n)$ is given by

$$\Pr(\mathbf{Y} = \mathbf{y}) = C \exp\{\sum_{i=1}^n \alpha_{i,y_i} + \tfrac{1}{2}\beta \sum_i \sum_{j \in N_i} I(y_i = y_j)\}$$

where N_i is the set of neighbor points to i while $I(\cdot)$ is the indicator function and C is a proportionality constant. This is called the Potts model within statistical mechanics, but is widely used in other fields such as image analysis.

(a). Write down a formula for C. How many terms needs to be added if $n = 100^2$ and $K = 4$?

(b). Consider the case where $\beta = 0$, what kind of interpretation can you have on $\alpha_{i,k}, i = 1, ..., n, k = 1, ..., K$?

(c). Assuming now $\beta > 0$, what kind of interpretation do you have on this parameter?

(d). Show that

$$\Pr(Y_i = k | Y_j = y_j, j \neq i) = \frac{\exp\{\alpha_{i,k} + \beta \sum_{j \in N_i} I(y_j = k)\}}{\sum_l \exp\{\alpha_{i,l} + \beta \sum_{j \in N_i} I(y_j = l)\}}$$

(e). Assume now we have observations $Z_1, ..., Z_n$ that conditional on $Y_1, ..., Y_n$ are independent and that $[Z_i | Y_i = k] = f_k(z_i)$, that is the observations follow different distributions depending on which group Y_i belongs to. Show that

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}) = C \exp\{\sum_{i=1}^n \bar\alpha_{i,y_i} + \tfrac{1}{2}\beta \sum_i \sum_{j \in N_i} I(y_i = y_j)\}$$

for some properly chosen $\bar\alpha_{i,y_i}$

Exercise 18 (CAR models)

Consider the CAR model

$$[Y_i | Y_j, j \neq i] = N(\sum_{j \in \mathcal{N}_i} c_{ij} y_j, \sigma_i^2)$$

We will consider the cases where $\mathbf{C} = \phi \mathbf{H}$ and $\mathbf{M} = \tau^2 \mathbf{\Delta}$ with $\mathbf{\Delta}$ being diagonal (so that $\sigma_i^2 = \tau^2 \Delta_{ii}$).

Remember the requirements needed:

(∗) $\Delta_{ii}^{-1} h_{ij} = \Delta_{jj}^{-1} h_{ji}$

(∗∗) $\tilde{\lambda}_{(1)}^{-1} < \phi < \tilde{\lambda}_{(n)}^{-1}$ where $\tilde{\lambda}_{(1)} \leq \tilde{\lambda}_{(2)} \leq \cdots \leq \tilde{\lambda}_{(n)}$ are the ordered eigenvalues of $\mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{1/2}$.

Define $\mathbf{W}$ such that $w_{ij} = I(i \text{ and } j \text{ are neighbors})$, that is the adjacency matrix.

We will consider processes on a regular lattice, so that $i \in \{(u,v), 1 \leq u \leq n_1, 1 \leq v \leq n_2\}$ where we will use $n_1 = n_2 = 10$. Also assume $\mathcal{N}_{u,v} = \{(u-1,v),(u+1,v),(u,v-1),(u,v+1)\}$ (with some corrections on the borders). The adjacency model can then be constructed by the commands

```
n1=n2=10;n = n1*n2
W = matrix(0,nrow=n,ncol=n)
#Neighbors in horizontal direction
for(j in 1:(n2))
for(i in 1:(n1-1))
  {
    W[(j-1)*n1+i,(j-1)*n1+i+1] = 1
    W[(j-1)*n1+i+1,(j-1)*n1+i] = 1
  }
#Neighbors in vertical direction
for(j in 1:(n2-1))
for(i in 1:(n1))
  {
    W[(j-1)*n1+i,j*n1+i] = 1
    W[j*n1+i,(j-1)*n1+i] = 1
  }
N = rowSums(W)
```

(The last command gives the number of neighbors to each site.)

(a). Show that $\mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{1/2} = \boldsymbol{\Delta}^{-1/2}\mathbf{H}\boldsymbol{\Delta}^{1/2}$

(b). Assume we are interested in $\text{cor}[Y_1, Y_2 | Y_k, k \neq 1, 2]$. Define $\mathbf{Q}_{11}$ to be the block of $\mathbf{Q}$ corresponding to (Y_1, Y_2).

How can $\text{cor}[Y_1, Y_2 | Y_k, k \neq 1, 2]$ be calculated from $\mathbf{Q}_{11}$?

Also express the correlation as a function of elements in $\boldsymbol{\Delta}$ and $\mathbf{H}$.

(c). Consider the case $\boldsymbol{\Delta} = \mathbf{I}, \mathbf{H} = \mathbf{W}$, which is called the Homogeneous CAR (HCAR) model.

   (i) Show that requirement (∗) is satisfied in this case.

   (ii) Calculate $\text{cor}[Y_1, Y_2 | Y_k, k \neq 1, 2]$. What restrictions does this give on ϕ.

   (iii) Using the eigenvalue requirements, what requirements are needed on ϕ in this case?

22

(iv) For $\phi = 0.2$ and $\tau = 1$, calculate the covariance matrix matrix and look in particular on the variances. Make a plot of the variances over the $n_1 \times n_2$ lattice. Comment on the results.

(d). Consider now the case $\boldsymbol{\Delta} = \mathrm{diag}\{|N(\mathbf{s}_i)|^{-1}\}$, $h_{ij} = w_{ij}|N(\mathbf{s}_i)|^{-1}$ which is called the Weighted CAR (WCAR) model.

(i) Show that requirement $(*)$ is satisfied in this case.

(ii) Calculate $\mathrm{cor}[Y_1, Y_2|Y_k, k \neq 1, 2]$. What restrictions does this give on ϕ.

(iii) Using the eigenvalue requirements, what requirements are needed on ϕ in this case?

(iv) For $\phi = 0.8$ and $\tau = 1$, calculate the covariance matrix matrix and look in particular on the variances. Make a plot of the variances over the $n_1 \times n_2$ lattice. Comment on the results.

(e). Consider now the ACAR model $\boldsymbol{\Delta} = \mathrm{diag}\{|N(\mathbf{s}_i)|^{-1}\}$, $h_{ij} = w_{ij}|N(\mathbf{s}_i)|^{-1/2}|N(\mathbf{s}_j)|^{1/2}$

(i) Show that requirement $(*)$ is satisfied in this case.

(ii) Calculate $\mathrm{cor}[Y_1, Y_2|Y_k, k \neq 1, 2]$. What restrictions does this give on ϕ.

(iii) Using the eigenvalue requirements, what requirements are needed on ϕ in this case?

(iv) For $\phi = 0.2$ and $\tau = 1$, calculate the covariance matrix matrix and look in particular on the variances. Make a plot of the variances over the $n_1 \times n_2$ lattice. Comment on the results.

Exercise 19 (Computation by simulation)

For hierarchical models we are interested in conditional distributions $p(\boldsymbol{\theta}|\mathbf{z})$ where $\mathbf{Z}$ denotes the set of observations while $\boldsymbol{\theta}$ is some unknown quantity of interest (which could be either be a latent process or, in a Bayesian setting, unknown parameters, or both).

In many situations the distribution $p(\boldsymbol{\theta}|\mathbf{z})$ is not numerically tractable. Assume however that we are able to simulate from $p(\boldsymbol{\theta}|\mathbf{z})$ and let $\{\boldsymbol{\theta}^{(s)}, s = 1, ..., S\}$ be a set of such simulations.

(a). Assume our interest is in $E[g(\boldsymbol{\theta})|\mathbf{Z} = \mathbf{z}]$. Suggest an estimator for this quantity based on the simulations.

(b). Assume the simulations are obtained independently. Calculate the variance of the estimator in (a) and comment on its dependence on S.

(c). Consider a hierarchical model where $q \sim \mathrm{Beta}(\alpha, \beta)$ and where $Z|q \sim \mathrm{Binom}(n, q)$. Our interest will be in $g(q) = q/(1-q)$ which we call the odds.

Note: The Beta distribution is given by $p(q; \alpha, \beta) \propto q^{\alpha-1}(1-q)^{\beta-1}$.

(i) Show that $q|Z = z$ also follows a Beta distribution with parameters updated.

(*ii*) Assume now that $\alpha = \beta = 1$ and that $n = 20, z = 5$. Use simulations to estimate $E[g(q)|Y = y]$. How can you assess the variance of this estimate?

Hint: The command `rbeta(S,a,b)` can be used to simulate from the Beta distribution.

(*iii*) How can you use the simulations to approximate the density $p(q|z)$?

Exercise 20 (INLA and the Besag model)

The *besag* model specified in the *INLA* package is defined as

$$[\delta_i|\delta_j, j \neq i] = N(|\mathcal{N}(\mathbf{s}_i)|^{-1} \sum_{j \in \mathcal{N}(\mathbf{s}_i)} \delta_j, \tau^2 |\mathcal{N}(\mathbf{s}_i)|^{-1})$$

(*a*). Show that this corresponds to a CAR model where $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{H})$ with $\mathbf{M} = \tau^2 \mathbf{\Delta}, \mathbf{\Delta} = \text{Diag}\{|\mathcal{N}(\mathbf{s}_i)|^{-1}\}$ and $H_{ij} = a_{ij}|\mathcal{N}(\mathbf{s}_i)|^{-1}$ where $a_{ij} = I(j \in \mathcal{N}(\mathbf{s}_i))$.

Show that this is a special case of the WCAR model with $\phi = 1$.

Hint: Use the results from exercise 12.

(*b*). Show that $\mathbf{1}^T(\mathbf{I} - \mathbf{H}) = \mathbf{0}$ and use this to explain why $\mathbf{I} - \mathbf{H}$ has an eigenvalue equal to 0. Discuss the implication of this.

(*c*). Related to the problem in the previous exercise, define $\xi_i = \delta_i + c$ where c is a constant. Show that $\mathbf{\xi}$ has exactly the same conditional distributions as $\mathbf{\delta}$.

One can actually show that these properties of the *besag* model makes it improper in the sense that it does not integrate to one. This has not restricted people from using the model because in most cases we are interested in the distribution of $\mathbf{\delta}$ conditional on data, and such models typically turn out to be proper.

(*d*). Assume in this part that $|\mathcal{N}(\mathbf{s}_i)|$ are equal for all i so that we can write $\mathbf{M} = \tilde{\tau}^2 \mathbf{I}$. Assume further that we have observations

$$Z_i = \delta_i + \varepsilon_i, \quad i = 1, ..., m$$

where the ε_i's are iid and $N(0, \sigma^2)$.

Show that $[\mathbf{\delta}|\mathbf{Z}]$ becomes a multivariate Gaussian distribution with a precision matrix that is of the form $\mathbf{Q}_{post} \propto \mathbf{M}^{-1}(\mathbf{I} - \phi\mathbf{H})$ where ϕ is now a parameter with $0 < \phi < 1$.

Using the script from exercise 18, verify numerically that $0 < \phi < 1$ is sufficient for $\mathbf{Q}_{post}$ to be positive definite.

Using improper models for inference within a model is by this possible. If however one considers model selection and use e.g. marginal likelihoods, problems might arise. In such cases it is better to consider proper models for the latent processes.

In the following, we will consider the WCAR model specified by $\mathbf{Q} = \tau^{-2}\mathbf{\Delta}^{-1}(\mathbf{I} - \phi\mathbf{H})$ where $0 < \phi < 1$ and where $\mathbf{\Delta}$ and $\mathbf{H}$ are defined as in (*a*).

(e). Show that in this case

$$[\delta_i|\delta_j, j \neq i] = N\left(\frac{\phi}{|\mathcal{N}(\mathbf{s}_i)|}\sum_{j \in \mathcal{N}(\mathbf{s}_i)}\delta_j, \tau^2|\mathcal{N}(\mathbf{s}_i)|^{-1}\right)$$

(f). Define $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\delta}$. Show that the precision matrix $\widetilde{\mathbf{Q}}$ for $\widetilde{\boldsymbol{\delta}}$ is equal to $\tau^{-2}(\mathbf{I} - \phi\widetilde{\mathbf{H}})$ where $\widetilde{\mathbf{H}}$ is a known matrix.

(g). Consider now a model $\mathbf{Z} = \mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}$. Show that this can be rewritten in a similar way as the spatial rates model considered in the lecture notes.

(h). Consider now the doctor's prescription data. Use the above analysis to perform inference using the WCAR model.

(You will here need to modify the code involving the spatial rates model appropriately.)

Compare your results both with the independence model and the spatial rates model with respect to

- Estimates of $\boldsymbol{\beta}$
- The marginal likelihood
- The DIC

Summarize your findings.

Exercise 21 (Markov Chain Monte Carlo)

In complex models, direct simulation from $p(\mathbf{y}|\mathbf{z})$ may be impossible. Markov chain Monte Carlo (MCMC) algorithms are algorithms constructed such that $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \mathbf{y}^{(3)}, ...$ are generated in sequence such that

- $\mathbf{y}^{(s)}$ is generated only depending on $\mathbf{y}^{(s-1)}$ (which makes it a Markov chain).

- As s increases, the distribution of $\mathbf{y}^{(s)}$ converges towards $p(\mathbf{y}|\mathbf{z})$

Under some additional requirements, we also have that $S^{-1}\sum_{s=1}^{S}g(\mathbf{y}^{(s)})$ converges towards $E[g(\mathbf{Y})|\mathbf{Z} = \mathbf{z}]$.

There are many possible MCMC algorithms. Perhaps the most simple one is the Gibbs sampler. Assume $\mathbf{y} = (y_1, ..., y_n)$. The algorithm is then defined by

- For $s = 1, 2, 3, ...$

    - simulate $y_1^s \sim p(y_1|y_2^{s-1}, ..., y_n^{s-1})$
    - simulate $y_2^s \sim p(y_2|y_1^s, y_3^{s-1}, ..., y_n^{s-1})$
    - $\vdots$

– simulate $y_n^s \sim p(y_n | y_1^s, y_2^s, ..., y_{n-1}^s)$

Note that we only need univariate simulations and that only conditional distributions are required.

Consider a simple model where

$$[y_i | \mu, \tau] = N(\mu, \tau^{-1}), \quad i = 1, ..., n$$
$$[\mu, \tau] = [\mu][\tau]$$

where

$$[\mu] = N(\mu_0, \tau_0^{-1})$$
$$[\tau] = \text{Gamma}(a, b)$$

(a). Show that

$$[\mu | \mathbf{y}, \tau] = N(\bar{\mu}, \bar{\tau}^{-1})$$
$$[\tau | \mathbf{y}, \mu] = \text{Gamma}(\bar{a}, \bar{b})$$

for appropriatly chosen $\bar{\mu}, \bar{\tau}$ and $\bar{a}, \bar{b}$.

Show in particular that the conditional distributions only depend on $\mathbf{y}$ through $\bar{y}$ and $\sum_{i=1}^{n}(y_i - \bar{y})^2$.

(b). Use the above results to implement a Gibbs sampler for simulating from $p(\mu, \tau | \mathbf{y})$.

For $\mu_0 = 0, \tau_0 = 0.01$ and $a = b = 0.05$ with $n = 5, \bar{y} = 15.4$ and $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 17.2$, run the Gibbs sampler. Use $\mu = 0, \tau = 1$ as starting values for the simulations.

Plot the simulated values of μ and τ as functions of iteration numbers. Do the simulations seem to stabilize fast?

Hint: The command `rgamma(1,shape=a,rate=b)` can be used to simulate from the Gamma(a, b) distribution.

(c). Use the simulations to produce approximations of $E[\mu | \mathbf{y}]$ and $E[\tau | \mathbf{y}]$ as well as 95% credibility intervals for both parameters.

Also use the simulations to estimate the correlation between μ and τ conditional on the data $\mathbf{y}$.

Exercise 22 (Markov random fields and blurring)
Consider an Ising random field $Y : \{Y_u; u \in \mathcal{L}_\mathcal{D}\}$ where $\mathcal{L}_\mathcal{D}$ is a regular lattice over $\mathcal{D} \subset \mathcal{R}^2$, and $Y_u \in \{0, 1\}$. The clique system consists of all closest neighbors of lattice nodes and the interaction parameter is β. The Gibbs formulation of random field is:

$$\Pr(\mathbf{Y} = \mathbf{y}) = \text{const} \times \exp\{-\beta \sum_{v \sim u} y_u y_v\}$$

where $v \sim u$ means that v and u are closest neighbors. In the following only lattice nodes in the interior of $\mathcal{L}_\mathcal{D}$ need to be considered - hence border problems can be ignored.

(a). Specify the Markov formulation for the random field and specify in particular the associated neighborhood system $\{\mathcal{N}(u); u \in \mathcal{L}_\mathcal{D}\}$.

Observations are available in all lattice nodes $\mathbf{Z} = \{Z_u; u \in \mathcal{L}_\mathcal{D}\}$ and the observation model contains *blurring*:

$$[Z_u|\mathbf{Y} = \mathbf{y}] = \frac{1}{5} \sum_{t \in \nu(u)} y_t + e_u$$

with blurring-domain $\nu(u)$ being the lattice node u and the four closest lattice nodes. If $u = (i, j)$ the blurring-domain is $\{(i, j), (i-1, j), (i, j-1), (i+1, j), (i, j+1)\}$, ie five lattice nodes. The observation errors e_u are spatially independent.

(b). Why do you think the observation model is called "blurring"?

(c). Develop the expression for the Markov formulation for the posterior model, $[\mathbf{Y}|\mathbf{z}]$, and specify in particular the associated neighborhood system.

Also specify all clique configurations in this case.

Exercise 23 (Potts models and computation)

Direct calculations from the Potts models is usually intractable, and Monte Carlo methods are usually applied instead.

Simulation from a Potts model can be performed by a Markov Chain Monte Carlo algorithm which is an iterative algorithm sampling $\mathbf{Y}^{s+1}$ conditional on $\mathbf{Y}^s$ in such a way that $\mathbf{Y}^s$ is approximately distributed according to the Potts model when s is large. In particular, a Gibbs sampler algorithm, corresponding to sampling Y_i^{s+1} from the conditional distribution derived above, can easily be implemented. A script for simulation from this model is given at the course web page as ***potts_sim.R***.

(a). Go through the script and make sure you understand how the algorithm works.

(Note in particular the trick with including an extra line of points around the grid-points which makes it easier to calculate the number of neighbors at the borders.)

(b). Run the script. Try also out different values of β and comment on differences.

Note: These simulations are somewhat time-consuming.

Markov Chain Monte Carlo methods should be used with care, since they might fail to converge (actually having very slow convergence). A usual way to check this is to look at trace plots of $g(\mathbf{X}^s)$ for some function(s) $g(\cdot)$ and see if it both has stabilized and is varying reasonable. Many choices of $g(\cdot)$ are possible. Some possibilities for the Potts model are

$$g_k(\mathbf{y}) = \sum_{i=1}^n I(y_i = k), k = 1, ..., K$$

and

$$g_{K+1}(\mathbf{y}) = \sum_i \sum_{j \in N_i} I(y_i = y_j)$$

(*c*). Use the given script to plot these functions and try to evaluate whether it seems like things have converged.

(*d*). In the script there are also commands for simulating data conditional on the last simulated configuration. Here it is assumed that

$$Z_i | \mathbf{Y} = \mathbf{y} \sim N(\mu_{y_i}, \sigma)$$

Try to modify the simulation part such that the simulations now are conditional on the observations **Z**. Perform simulations and plot the images you obtain in this case. Remark on differences you get in this case compared to the unconditional simulation.

Exercise 24 (Spatial interpolation using geoR)
As part of monitoring the pollution in the Barents region, a whole range of chemical elements were analyzed in the Kola region. Parts of these data can be downloaded from the course project page (the **kola.Rdata** file). A short description of the data is given by the **env_geochemical_atlas.pdf** file on the same page. The data can be loaded into R with the command

```
load("kola.Rdata")
```

which gives you an object **kola.geoR** of class **geodata**. A plot of the geographical locations (after you have loaded the geoR library) for the data and some specific plots for the Nickel measurements can be obtained by the command

```
plot(kola.geoR,data=kola.geoR$data[,"Ni"])
```

and similarly for the other measurements (**colnames(kola.geoR$data)** gives a list of the available measurements). We will in this exercise analyze these data using the geoR package. Note that it might be a good idea to transform the x- and y- coordinates to be in the range $[0, 1]$ in order to make some of the routines in geoR work properly. I.e. perform the commands

```
kola.geoR$coords = coords2coords(kola.geoR$coords,xlim=c(0,1),ylim=c(0,1))
kola.geoR$borders = coords2coords(kola.geoR$borders,xlim=c(0,1),ylim=c(0,1))
```

We will in the following concentrate on the Nickel component.

(*a*). From the histogram of the Nickel components, the data does not look very Gaussian. Make an alternative plot using the log-transform of the response instead. Does this look better from a Gaussian distribution perspective?

(*b*). Fit an ordinary linear regression model with the logarithm of Nickel as response and the geographical coordinates and elevation (given in **kola.geoR$covariate**) as explanatory variables. Explore the spatial correlation of the residuals (through estimates of variograms) and discuss why the inferences about the regression parameters obtained should be used with care.

Defining $Z(\mathbf{s})$ to be the observed log-value of Nickel at site $\mathbf{s}$, assume now a model

$$Z(\mathbf{s}) = \mathbf{x}(s)^T \boldsymbol{\beta} + \delta(\mathbf{s})$$

where $\mathbf{x}(s)$ are covariates at site $\mathbf{s}$ and $\{\delta(\mathbf{s})\}$ is a zero-mean Gaussian process. We will be interested in selecting the best model both with respect to the expectation structure (selection of covariates) and the best model for the covariance structure. Usually one wants most of the structure to be modeled through the expectation structure (easiest to predict). A good strategy is this therefore to start with choosing a model for the covariance structure using as many covariates as possible and thereafter perform model selection using the chosen model for the covariance structure.

(c). A model including the coordinates and elevation in the expectation structure and using a Matern covariance structure can be fitted by the following command:

```
ml1 = likfit(kola.geoR,data=log(kola.geoR$data[,"Ni"]),
  trend=trend.spatial(~coords+ELEV,kola.geoR),
  ini.cov.pars=c(0.5,0.5),fix.kappa=FALSE)
```

(Note: This call can be a bit time-consuming.) Further, the AIC value for this model can be obtained by

```
AIC(ml1)
```

By including an option *fix.nugget=TRUE* in the call to *likfit*, the nugget effect will be fixed (to the default value zero).

Try out different models for the covariance structure (including the Matern, exponential, Gaussian and spherical) both with and without nugget effects. Based on the AIC values, which model performs the best?

Summarize your findings.

(d). Try out a Bayesian approach for fitting the model you found in the previous exercise. Use this to predict the Nickel content on a grid within the area of interest.

Hint: You need to deal with that the covariates involved are only available in the observation sites. Imputing these missing values in some way might be a possibility.

(e). Assume our interest is in high values of Nickel. Estimate the size of the region for which the Nickel value is above 100. Include an uncertainty measure to your estimate.

(f). Discuss strengths and weaknesses with the points you have been through.

Exercise 25

(a). Assume C_1 and C_2 are two covariance functions for processes within $\mathcal{R}^d$ that are both non-negative definite. Show that the sum of these covariance functions also is non-negative definite

Consider now the spatio-temporal model (Eq (6.8) in the textbook)

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \beta(s) + \gamma(t) + \kappa(\mathbf{s}, t) + \delta(\mathbf{s}, t)$$

where

$$
\begin{aligned}
\text{cov}(\beta(\mathbf{s}), \beta(\mathbf{x})) &= C_\beta(\mathbf{s}, \mathbf{x}) \\
\text{cov}(\gamma(t), \gamma(r)) &= C_\gamma(t, r) \\
\text{cov}(\kappa(\mathbf{s}, t), \kappa(\mathbf{x}, r)) &= C_\kappa(\mathbf{s}, \mathbf{x}, t, r) \\
\text{cov}(\delta(\mathbf{s}, t), \delta(\mathbf{x}, r)) &= \sigma_\delta^2 I(\mathbf{x} = \mathbf{s}, t = r)
\end{aligned}
$$

(b). Express the covariance function $C_Y(\cdot; \cdot)$ for $\{Y(\mathbf{s}, t)\}$ through the individual covariance functions above.

(c). Show that if each of the individual covariance functions are non-negative definite, then also $C_Y(\cdot; \cdot)$ is non-negative definite.

Exercise 26

Consider a time-discrete spatio-temporal process $\{Y_t(\mathbf{s})\}$ where

$$
\begin{aligned}
Y_1(\mathbf{s}) &= \frac{1}{\sqrt{1 - a^2}} \delta_1(\mathbf{s}) \\
Y_t(\mathbf{s}) &= a Y_{t-1}(\mathbf{s}) + \delta_t(\mathbf{s}) \qquad\qquad t = 2, 3, ...
\end{aligned}
$$

where $\{\delta_t(\mathbf{s})\}$ is a zero-mean spatially correlated but independent in time. Assume further that

$$\text{Cov}[\delta_t(\mathbf{s}), \delta_t(\mathbf{x})] = C_\delta(\mathbf{s}, \mathbf{x})$$

for all t.

(a). Show that

$$\text{cov}[Y_t(\mathbf{s}), Y_t(\mathbf{x})] = \frac{1}{1 - a^2} C_\delta(\mathbf{s}, \mathbf{x})$$

for all t.

(b). Show that

$$\text{cov}[Y_t(\mathbf{s}), Y_{t-1}(\mathbf{x})] = a\,\text{cov}[Y_{t-1}(\mathbf{s}), Y_{t-1}(\mathbf{x})]$$

and use this to show that the process is separable.

(c). Show that the process is 2. order stationary in time.

What requirements are needed for the process to also be 2. order stationary in space?

Exercise 27 (Separable covariance functions)

Assume $C^{(s)}(\mathbf{h})$ is a purely (stationary) spatial covariance function in $\mathcal{R}^d$ and $C^{(t)}(\tau)$ is a purely (stationary) temporal covariance function. We will in this exercise show that

$$C(\mathbf{h}, \tau) = C^{(s)}(\mathbf{h}) \times C^{(t)}(\tau)$$

is a valid (stationary) covariance function in space-time.

(a). Recall Bochner's Theorem (Theorem 4.1) in the book and use this to show that $C(\mathbf{h}, \tau)$ is a valid covariance function if and only if it can represented by

$$C(\mathbf{h}, \tau) = \int_{\boldsymbol{\omega} \in \mathcal{R}^d} \int_{\nu \in \mathcal{R}} \cos(\boldsymbol{\omega}^T h + \nu\tau) f(\boldsymbol{\omega}, \nu) d\boldsymbol{\omega} d\nu$$

where $f(\boldsymbol{\omega}, \nu) \geq 0$ is symmetric about $(\mathbf{0}, 0)$

(b). Using that $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$ show that $C(\mathbf{h}, \tau)$ is a valid covariance function if and only if it can represented by

$$C(\mathbf{h}, \tau) = \int_{\boldsymbol{\omega} \in \mathcal{R}^d} \int_{\nu \in \mathcal{R}} \cos(\boldsymbol{\omega}^T h)\cos(\nu\tau) f(\boldsymbol{\omega}, \nu) d\boldsymbol{\omega} d\nu \qquad (*)$$

Hint: Use the symmetry properties of $f(\cdot, \cdot)$ to show that the sinus-parts integrates to zero.

(c). Show that $C(\mathbf{h}, \tau) = C^{(s)}(\mathbf{h}) \times C^{(t)}(\tau)$ indeed can be represented by (*) and that this proves that it is a valid covariance function in space-time.

Exercise 28 (Separable covariance functions and kriging)

Consider a zero-mean time-discrete spatio-temporal process $\{Y_t(\mathbf{s})\}$ with a separable covariance function given by

$$C_Y((\mathbf{s}, t), (\mathbf{x}, t + \tau)) = C_Y^{(s)}(\mathbf{s}, \mathbf{x}) C_Y^{(t)}(\tau)$$

where we assume $C_Y^{(t)}(0) = 1$.

Assume further that we have the observations $\{Y(\mathbf{s}_j, v), j = 1, ..., m, v = 1, ..., t\}$ and we want to predict $Y(\mathbf{s}_0, t + 1)$.

We will further define

$$\mathbf{Y}_v^T = (Y(\mathbf{s}_1, v), ..., Y(\mathbf{s}_m, v))$$
$$\mathbf{Y}^T = (\mathbf{Y}_1^T, ..., \mathbf{Y}_t^T)$$

(a). Show that

$$\operatorname{cov}[Y(s_i, v), Y(s_j, w)] = \operatorname{cov}[Y(s_i, v), Y(s_j, v)] C^{(t)}(w - v).$$

(b). Define $\mathbf{C}_Y$ to be the $mt \times mt$ covariance matrix for the observed $\mathbf{Y}$. Show that we can write

$$\mathbf{C}_Y = \mathbf{C}_{Y,t} \otimes \mathbf{C}_{Y,s}$$

where $\mathbf{C}_{Y,s}$ is an $m \times m$ matrix equal to the covariance matrix for $\{Y(\mathbf{s}_j, v), j = 1, ..., m\}$ for a fixed v while $\mathbf{C}_{Y,t}$ is a $t \times t$ matrix equal to the correlation matrix for $\{Y(\mathbf{s}, v), v = 1, ..., t\}$ for a fixed $\mathbf{s}$.

Here $\otimes$ denotes the Kronecker product which means that each element of $\mathbf{C}_{Y,t}$ is multiplied by $\mathbf{C}_{Y,s}$.

(c). Show that $\mathbf{C}_Y$ becomes non-negative definite if both $\mathbf{C}_{Y,t}$ and $\mathbf{C}_{Y,s}$ are non-negative definite.

(d). Why do we need to use the correlation matrix in time and not the covariance matrix?

(e). Show that

$$\mathbf{C}_{Y,t}^{-1} \otimes \mathbf{C}_{Y,s}^{-1}$$

is the inverse of $\mathbf{C}_Y$.

(f). Show that $\mathbf{c}_0 = \mathrm{cov}[Y(\mathbf{s}_0, t+1), \mathbf{Y}]$ is given by

$$\mathbf{c}_0 = \mathbf{c}_{0,t} \otimes \mathbf{c}_{0,s}$$

where $\mathbf{c}_{0,s} = \mathrm{cov}[Y(\mathbf{s}_0, t), \mathbf{Y}_t]$ and $\mathbf{c}_{0,t} = (C_Y^{(1)}(t), C_Y^{(1)}(t-1), ..., C_Y^{(1)}(1))$.

(g). Show that

$$
\begin{aligned}
E[Y(\mathbf{s}_o, t+1)] &= \mathbf{c}_0^T \mathbf{C}_Y^{-1} \mathbf{Y} \\
&= [\mathbf{c}_{0,t}^T \mathbf{C}_{Y,t}^{-1}] \otimes [\mathbf{c}_{0,s}^T \mathbf{C}_{Y,s}^{-1}] \mathbf{Y} \\
&= \sum_{v=1}^{t} b_v [\mathbf{c}_{0,s}^T \mathbf{C}_{Y,s}^{-1}] \mathbf{Y}_v
\end{aligned}
$$

where b_v is the vth element of $\mathbf{c}_{0,t}^T \mathbf{C}_{Y,t}^{-1}$.

(h). Consider the special case (exercise 26) where

$$C_Y^{(t)}(\tau) = a^{|\tau|}$$

Show that

$$
\mathbf{C}_{Y,t} = \begin{pmatrix}
1 & a & a^2 & a^3 & \cdots & a^{t-2} & a^{t-1} \\
a & 1 & a & a^2 & \cdots & a^{t-3} & a^{t-2} \\
a^2 & a & 1 & a & \cdots & a^{t-4} & a^{t-3} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
a^{t-2} & a^{t-3} & a^{t-4} & a^{t-5} & \cdots & 1 & a \\
a^{t-1} & a^{t-2} & a^{t-3} & a^{t-4} & \cdots & a & 1
\end{pmatrix}
$$

and

$$\mathbf{C}_{Y,t}^{-1} = \frac{1}{1-a^2} \begin{pmatrix} 1 & -a & 0 & 0 & \cdots & 0 & 0 \\ -a & 1+a^2 & -a & 0 & \cdots & 0 & 0 \\ 0 & -a & 1+a^2 & -a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+a^2 & -a \\ 0 & 0 & 0 & 0 & \cdots & -a & 1 \end{pmatrix}$$

(i). Show that b_v as defined in (g) is equal to zero for all v except for $v = t$. Based on the model in exercise 26, argue why this is reasonable.

Exercise 29 (Discretization of SPDE)
Consider an extension of the diffusion-injection model given by

$$\frac{\partial Y(s;t)}{\partial t} = \beta \frac{\partial^2 Y(s;t)}{\partial s^2} + \gamma \frac{\partial Y(s;t)}{\partial s} - \alpha Y(s;t) + \delta(s;t)$$

(a). Using a similar discretization as in the lecture notes, including the approximation

$$\frac{\partial Y(s;t)}{\partial s} \approx \frac{Y(s+\Delta_s;t) - Y(s;t)}{\Delta_s},$$

write a discrete approximation of the model relating $Y(s, t + \Delta_t)$ to $\{Y(\cdot, t)\}$.

(b). Now consider an alternative approximation

$$\frac{\partial Y(s;t)}{\partial s} \approx \frac{1}{2}\frac{Y(s+\Delta_s;t) - Y(s;t)}{\Delta_s} + \frac{1}{2}\frac{Y(s;t) - Y(s-\Delta_s;t)}{\Delta_s}$$

and again write a discrete approximation of the model relating $Y(s, t + \Delta_t)$ to $\{Y(\cdot, t)\}$.

(c). Discuss the differences between these two approximations.

Exercise 30 (Vector autoregressive models as CAR models)
Consider a process $\{Y_t(s)\}$ defined on a 1-dimensional grid of integer locations $s \in \{1, ..., m\}$ and over a discrete time set $t \in \{1, ..., T\}$. Define $\mathbf{Y}_t = (Y_t(1), ..., Y_t(m))'$. Assume that

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{ind}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$$

and further that $\mathbf{Y}_0 = \mathbf{0}$.

(a). Assume first that $\mathbf{M} = \mathbf{I}$ and $\boldsymbol{\Sigma}_\eta = \sigma_\eta^2 \mathbf{I}$. Find $[Y_t(s)|Y_v(x), (v, x) \neq (t, s)]$ and use this to describe the process as a CAR model with corresponding neighborhood system.

(b). Now extend the model to assume $\mathbf{Q}_\eta = \mathbf{\Sigma}_\eta^{-1}$ has a sparse structure where the (i,j) element of $\mathbf{Q}_\eta$ is zero if $|i-j| > 1$. Again describe the process as a CAR model with corresponding neighborhood system.

(c). Extend the model further to assume

$$M_{ij} \neq 0 \quad \text{if } |j-i| \leq 1$$
$$M_{ij} = 0 \quad \text{if } |j-i| > 1$$

Again describe the process as a CAR model with corresponding neighborhood system.

Exercise 31

Define $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), ..., Y_t(\mathbf{s}_n))^T$. Consider a model

$$\mathbf{Y}_t = a\mathbf{Y}_{t-1} + \boldsymbol{\delta}_t$$

where $\boldsymbol{\delta}_t = (\delta_t(\mathbf{s}_1), ..., \delta_t(\mathbf{s}_n))^T \sim N(\mathbf{0}, \mathbf{\Sigma}_\delta)$ and independent in time.

(a). Assuming $\mathbf{\Sigma}_\delta$ given and that $\mathbf{\Sigma}_Y = \text{Var}[\mathbf{Y}_t]$ for all t. Find $\mathbf{\Sigma}_Y$ in this case.

(b). Show that the model above is a special case of

$$\mathbf{Y}_t = \mathbf{M}\mathbf{Y}_{t-1} + \boldsymbol{\delta}_t \qquad (*)$$

and that $\mathbf{\Sigma}_Y$ is the solution of the equation

$$\text{vec}(\mathbf{\Sigma}_Y) = \{\mathbf{I} - \mathbf{M} \otimes \mathbf{M}\}^{-1}\text{vec}(\mathbf{\Sigma}_\delta) \qquad (**)$$

(c). Consider now a model

$$Y_t(\mathbf{s}_i) = a_i\mathbf{Y}_{t-1}(\mathbf{s}_i) + \boldsymbol{\delta}_t(\mathbf{s}_i)$$

Find $\mathbf{\Sigma}_Y$ in this case.

(d). Also in this extended model, show that it is a special case of $(*)$ and that the $\mathbf{\Sigma}_Y$ you found is a solution of $(**)$.

Hint: Consider the case $n = 2$ first.

(e). Consider now a general model of the form $(*)$. In the case of $n = 2$, show that a time-stationary covariance matrix must satisfy $(**)$.

Exercise 32 (Spatio-temporal modelling using INLA)

We will in this exercise (due to computational burden) consider a subset of the sea surface temperature data, using only the first 5 time points and a 30×30 grid. The data is available at the course home-page in the file *SST_sub.dat* where each column correspond to a time point. Each row corresponds to a spatial point where first the 30 point in longitude direction is given for the first latitude level, then for the second latitude level and so on. It can be read into R by

```
SST = matrix(scan("SST_sub.dat"),ncol=5,byrow=T)
nx=30;ny=30;nT=dim(SST)[2]
```

For the spatial part we will consider a CAR model, and for that you also need the file *SST.graph* which give you the spatial graph. It can be read into R by

```
library(INLA)
g = inla.read.graph("SST.graph")
```

This exercise contain a lot of R commands. They are directly available at the script *INLA_SST.R*. Note that some of the commands will be time-consuming.

(a). Consider first a simple spatial model ignoring temporal dependence:

$$Y(\mathbf{s},t) = \mu + \beta_0(\mathbf{s}) + \beta_1(\mathbf{s}) + \varepsilon(\mathbf{s},t)$$

where $\{\beta_0(\mathbf{s})\}$ is a spatial process with a CAR structure, $\{\beta_1(\mathbf{s})\}$ is a spatial process with iid components and $\varepsilon(\mathbf{s},t)$ are iid additional noise terms. This model can be analysed by the following commands

```
Y = as.vector(as.matrix(SST))    #Vector of responses
SPnode1 = rep(1:(nx*ny),nT)        #Index for spatial nodes
SPnode2 = rep(1:(nx*ny),nT)        #Copy of index for spatial nodes
d = data.frame(Y=Y,SPnode1=SPnode1,SPnode2=SPnode2)
g = inla.read.graph("SST.graph")
formula1 = Y~f(SPnode1,model="iid")+
    f(SPnode2,model="besag",graph=g)
mod1 = inla(formula1,data=d)
```

Try out these commands. How large are the variances for the different components involved?

(b). An extension of the previous model is to include temporal components:

$$Y(\mathbf{s},t) = \mu + \beta_0(\mathbf{s}) + \beta_1(\mathbf{s}) + \gamma_0(t) + \gamma_1(t) + \varepsilon(\mathbf{s},t)$$

where now $\{\gamma_0(t)\}$ is a random walk process (a one dimensional CAR model) while $\{\gamma_0(t)\}$ is an iid process. This model can be analysed by the following additional commands

```
d$Tnode1 = rep(1:nT,each=nx*ny)   #Index for temporal nodes
d$Tnode2 = rep(1:nT,each=nx*ny)   #Copy of index for temporal nodes
formula2 = Y~f(SPnode1,model="iid")+
            f(SPnode2,model="besag",graph=g)+
            f(Tnode1,model="iid")+f(Tnode2,model="rw1")
mod2 = inla(formula2,data=d)
```

Again look at the sizes of the variances of the different components and comment.

Finally, we will look at a model where an interaction term between space and time is included:

$$Y(\mathbf{s}, t) = \mu + \beta_0(\mathbf{s}) + \beta_1(\mathbf{s}) + \gamma_0(t) + \gamma_1(t) + \delta(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

We will assume a separable covariance structure for $\{\delta(\mathbf{s}, t)\}$. Assuming $\mathbf{Q}^{(s)}$ is the precision matrix for $\{\beta_0(\mathbf{s})\}$ and $\mathbf{Q}^{(t)}$ is the precision matrix for $\{\gamma_0(t)\}$, we will assume that the precision matrix for $\{\delta(\mathbf{s}, t)\}$ is $\mathbf{Q} = \mathbf{Q}^{(t)} \otimes \mathbf{Q}^{(s)}$ (actually we are here talking about the covariance matrices for the points that we have observations, not for the whole process).

Now, for the models we have assumed for the pure spatial and pure temporal components above, we have $\mathbf{Q}^{(s)} = \tau^{(s)} \mathbf{R}^{(s)}$ and $\mathbf{Q}^{(t)} = \tau^{(t)} \mathbf{R}^{(t)}$ and $\mathbf{R}^{(s)}$ and $\mathbf{R}^{(t)}$ are known matrices.

(c). Show that $\mathbf{Q} = \tau \mathbf{R}$ where $\mathbf{R}$ is a known matrix.

The spatial matrix $\mathbf{R}^{(s)}$ can be calculated by the commands

```
g = inla.read.graph("SST.graph")
Rspat = diag(g$nnbs)
for(i in 1:g$n)
  Rspat[i,g$nbs[[i]]] = -1
```

while $\mathbf{R}^{(t)}$ can be calculated directly by

```
Rtemp = diag(c(1,rep(2,nT-2),1))
Rtemp[1,2] = -1
for(i in 2:(nT-1))
  Rtemp[i,c(i-1,i+1)] = -1
Rtemp[nT,nT-1] = -1
```

(d). What is the neighborhood structure corresponding to $\mathbf{R}^{(t)}$?

Finally we can analyse the full model by the commands

```
Rspat.temp = kronecker(Rtemp,Rspat)
d$STnode = 1:(nx*ny*nT)
formula3 = Y~f(SPnode1,model="iid")+
           f(SPnode2,model="besag",graph.file="SST.graph")+
           f(Tnode1,model="iid")+f(Tnode2,model="rw1")+
           f(STnode,model="generic0",Cmatrix=Rspat.temp,constr=TRUE)
mod3 = inla(formula3,data=d)
```

(e). Perform the commands above. Again look at the sizes of the variances and comment on the results.

Exercise 33 (INLA and missing data)
We will again consider the data in the previous exercise, but now we will see how INLA can handle missing data.

We can randomly insert missing data in the respond variable by the following commands

```
n = dim(d)[1]
nmiss = 450
ind = sample(1:n,niss)
dmiss = d
dmiss$Y[ind] = NA
```

(a). Try out the first model from exercise 32 using *dmiss* as input data. Compare the results with the ones obtained by using the full data set

(b). Increase the number of missing data to 2250 and repeat the previous point. Again compare your results.

(c). Also look at the computation times used when missing data are included. Try to explain the increase in computation time.

Exercise 34 (Ireland wind data)
Haslett and Raftery (1989) analysed a dataset

Exercise 35 (Air quality data)
The file *Piemonte_data_byday.csv* contain data on PM_{10} (particulate matter with an aerodynamic diameter of less than 10 μm) daily observed over a period from 01/10/05 to 31/03/06 at 24 stations in the North-Italian region Piemonte. This is a dataset taken from Cameletti et al. [2011] which we will analyse using spatio-temporal models later. Here we will concentrate on models where the spatial correlation is ignored. In order to make the response approximately Gaussian distributed, we will consider $\log(PM_{10})$.

8 covariates are available: daily mean wind speed (*WS*, m/s), daily maximum mixing height (*HMIX*, m), daily precipitation (*P*, mm), daily mean temperature (*TEMP*, $°K$), daily emissions (*EMI*, g/s), altitude (*A*, m) and spatial geographic coordinates (*UTMX* and *UTMY*, in km). In addition we will make use of a time index (*Time*, a numerical value going from 1 to 182).

The dataset can be read into R by the commands

```
d <-read.table("Piemonte_data_byday.csv",header=TRUE,sep=",")
d$Time = rep(1:182,each=24)
d$logPM10 = log(d$PM10)
```

(a). Consider first a simple regression model

$$y_t(s_i) = \mu + \mathbf{x}_t(s_i)\boldsymbol{\beta} + \varepsilon_t(s_i)$$

where $\mathbf{x}_t(s_i)$ is the vector of covariates at time t and station i, *Time* excluded.

Fit such a model using ordinarly least squares and use different tools for checking whether the residuals are uncorrelated.

Hint: Possible commands are

```
fit = lm(logPM10~A+UTMX+UTMY+WS+TEMP+HMIX+PREC+EMI,data=d)
acf(resid(fit)[d$Station==1],na.action=na.omit)
```

(b). Include now *Time* as a covariate. Look at the acf curve for the first station in this case. Does it look like some of the autocorrelation is removed?

(c). Include now *Time* as a categorical covariate by the command

```
d$Time = as.factor(d$Time)
options(contrasts=c("contr.sum","contr.sum"))
fit3=lm(logPM10~A+UTMX+UTMY+WS+TEMP+HMIX+PREC+EMI+Time,data=d)
```

Again look at the autocorrelation function. How do the autocorrelationfunction look in this case?

(d). The model above can be written as

$$y_t(s_i) = \mu + \mathbf{x}_t(s_i)\boldsymbol{\beta} + \alpha_t + \varepsilon_t(s_i)$$

where α_t are parameters, one for each time point. The command

```
options(contrasts=c("contr.sum","contr.sum"))
```

above make a constraint $\sum_t \alpha_t = 0$ in order to make all the parameters identifiable.

Find the estimates of α_t for all t (note that you have to calculate the last one using the constraint above) and plot this as a function of time. Also look at the autocorrelation function of the estimates.

(e). An alternative to consider the α_t's to be fixed parameters, is to consider them as random variables. A particular choice is to assume $\alpha_t \overset{iid}{\sim} N(0, \sigma_\alpha^2)$. Fit such a model with INLA using the commands

```
library(INLA)
formula = logPM10~A+UTMX+UTMY+WS+TEMP+HMIX+PREC+EMI+f(Time,model="iid")
fit4=inla(formula,data=d,control.inla=list(int.strategy="eb"))
```

The "estimates" of the α's are in this case available in the object

```
fit4$summary.random$Time[,2]
```

Why do I write "estimates" in this case?

Compare these estimates with the one obtained by considering the α's as fixed parameters. In particular, look at the square sum of the two types of estimates. Why do you think the square sum of the estimates based on the α's being treated as random are smaller?

(f). For the model where the α_t's are treated as random, calculate the covariance structure for $\{Y_t(\mathbf{s}_i)\}$.

(g). Now consider an extension where the α_t's are considered to follow an AR(1) process. Such a model can be fitted into INLA through the commands

```
formula2 = logPM10~A+UTMX+UTMY+WS+TEMP+HMIX+PREC+EMI+f(Time,model="ar1")
fit5=inla(formula2,data=d,control.inla=list(int.strategy="eb"))
```

What is the estimate of the autoregressive parameter in this case?

Compare the "estimates" of the α_t's for this model with the ones obtained using the iid assumption.

(h). Calculate the covariance structure for $\{Y_t(\mathbf{s}_i)\}$ for the extended model.

Exercise 36 (Air quality data (cont))
Consider again the data from exercise 35. We will now extend the model by taking spatial dependence into account as well. In that case we need to read in the spatial coordinates for the data, which can be done with the following commands (including plotting the spatial points as well as the border of the region:

```
coordinates <-read.table("coordinates.csv",header=TRUE,sep=",")
borders = read.table("Cameletti/Piemonte_borders.csv",header=TRUE,sep=",")
plot(borders, lwd=3,type="l")
points(coordinates$UTMX, coordinFFates$UTMY, pch=20, cex=2, col=2)
```

In order to avoid some numerical problems, we will also reduce the dataset to the first 50 days:

```
d = d[d$Time<=50,]
```

Consider first a model of the form

$$y_t(s) = \mu + \mathbf{x}_t(s)\boldsymbol{\beta} + \alpha_t + \delta(\mathbf{s}) + \varepsilon_t(\mathbf{s})$$

where now $\{\delta(s_i)\}$ is a zero-mean Gaussian spatial process. Further, we will in the beginning only use A as a covariate.

(a). Assume first that $\{\delta(\mathbf{s})\}$ has an independence structure. Such a model can be fitted by INLA with the commands

```
ind = rep(1:nstations,ndays)
formula = logPM10 ~  A + f(Time,model="ar1") +f(ind,model="iid")
res = inla(formula,data=d,control.inla=list(int.strategy="eb"))
```

Fit this model and extract the predictions of the δ's.

Use different methods to explore whether there are any spatial structure in the δ's.

(Hint: You can use the variogram approach or using Moran's I, both discussed in Chapter 4. For the latter you can look at the STI-calculations for the SST data, noting that Moran's I is a spatial version of STI.)

(b). Assume now that $\{\delta(\mathbf{s})\}$ has an exponential covariance structure, that is

$$\text{Cov}[\delta(s), \delta(s+\mathbf{h})] = \sigma_\beta^2 \exp(-||\mathbf{h}||/\theta)$$

INLA does not have this covariance function available directly. Assume however that θ is known. Show that the precision matrix for $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), ..., \delta(\mathbf{s}_n))$ can be written as $\tau \mathbf{C}$ where $\mathbf{C}$ is a known matrix. This situation is covered by the *generic0* model in INLA. Such a model can be fitted by the commands

```
dis = as.matrix(dist(coordinates[,c("UTMX","UTMY")],upper=TRUE,diag=TRUE))
theta = 7
Sigma.delta = exp(-dis/theta)
C0 = solve(Sigma.delta)
formula2=logPM10~A+f(Time,model="ar1")+f(ind,model="generic0",Cmatrix=C0)
res2 = inla(formula2,data=d,control.inla=list(int.strategy="eb"))
```

for $\theta = 7$. Perform these commands as well.

A possible measure for comparing models is the marginal (log) likelihood which is given by `res2$mlik` for the second model. Which model gives the best fit?

Also try out different θ. Use the best one in the following.

(c). A further extension of the model is to assume

$$y_t(s) = \mu + \mathbf{x}_t(s)\boldsymbol{\beta} + \delta_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$$

where now $\{\delta_t(\mathbf{s})\}$ is a spatio-temporal process following the dynamic model

$$\delta_t(\mathbf{s}) = \rho\delta_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s})$$

Such a model can be fitted in INLA using the commands

```
formula3 = logPM10 ~  A + f(Time,model="ar1") +
  f(ind,model="generic0",Cmatrix=C0,group=Time,control.group=list(model="ar1"))
res3 = inla(formula3,data=d,control.inla=list(int.strategy="eb"))
res3$mlik
```

40

Compare this model with the ones you have obtained earlier.

(d). Now repeat the points above including the covariates *WS,HMIX,P,TEMP,EMI*.

Why is it not necessary to include *UTMX,UTMY*?

Why does it seem like the spatial structure becomes less important now?

(e). Summarize your findings.

Note: In Cameletti et al. [2011] an alternative approach utilizing the computational advantages of CAR models is used. We will however not consider that further here.

Exercise 37 (A hierarchical spatio-temporal model for air quality data)
In this exercise we will look at a dataset of particulate matter concentration in the North-Italian region Piemonte. In particular, we will be interested in PM_{10} (particulate matter with an aerodynamic diameter of less than 10 μm) dayly observed over a period from 01/10/05 to 31/03/06. 8 covariates are available: daily mean wind speed (*WS*, m/s), daily maximum mixing height (*HMIX*, m), daily precipitation (*P*, mm), daily mean temperature (*TEMP*, $°K$), daily emissions (*EMI*, g/s), altitude (*A*, m) and spatial geographic coordinates (*UTMX* and *UTMY*, in km).

(a). First read the data in by the commands

```
d <-read.table("Piemonte_data_byday.csv",header=TRUE,sep=",")
d$logPM10 = log(d$PM10)
coordinates <-read.table("coordinates.csv",header=TRUE,sep=",")
borders = read.table("Piemonte_borders.csv",header=TRUE,sep=",")
```

and plot the spatial points by the commands

```
plot(borders, lwd=3,type="l")
points(coordinates$UTMX, coordinates$UTMY, pch=20, cex=2, col=2)
```

(b). Consider first a simple regression model where

$$y_t(s) = \mathbf{x}_t(s)\boldsymbol{\beta} + \varepsilon_t(s)$$

where $y_t(s)$ is $\log(PM_{10})$, $\{\varepsilon_t(s)\}$ are iid noise variables following a $N(0,\sigma^2)$ distribution while $\mathbf{x}_t(s)$ is the vector of covariates.

Fit such a model using ordinarly least squares and use different tools for checking whether the residuals are uncorrelated.

(c). Consider now an extended model

$$y_t(s) = \mathbf{x}_t(s)\boldsymbol{\beta} + \xi_t(s) + \varepsilon_t(s)$$
$$\xi_t(s) = a\xi_y(s-1) + \omega_t(s)$$

where $\mathbf{x}_t(s)$ is a p-dimensional vector of covariates and where

$$\text{Cov}[\omega_t(\mathbf{s}_i), \omega_{t'}(\mathbf{s}_j)] = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C(||\mathbf{s}_i - \mathbf{s}_j||) & \text{if } t = t' \end{cases}$$

where $C(\cdot)$ is a known function.

Calculate $C_\xi(\mathbf{H}, \tau) = \text{Cov}[\xi_t(\mathbf{s}), \xi_{t+\tau}(\mathbf{s} + \mathbf{h})$ and show that the covariance function is separable.

(d). Propose a reasonable estimator for a based on the residuals from (b) and calculate the estimate. Use this value for a in the following.

(e). Show that for given a, the precision matrix $\mathbf{Q}$ for $\{\xi_t(\mathbf{s}_i), i = 1, ..., m, t = 1, , , T\}$ is given by $\tau \times \mathbf{Q}_0$ where $\mathbf{Q}$ is a known matrix.

(f). Assume know $C(\mathbf{h}) = \exp(-||h||/\theta)$.

Exercise 38

In this exercise we will consider the `meuse` dataset which gives locations and topsoil heavy metal concentrations, along with a number of soil and landscape variablesat the observation locations, collected in a flood plain of the river Meuse, near the village of Stein (NL). Heavy metal concentrations are from composite samples of an area of approximately 15 m x 15 m.

Consider a model

$$Z_i = \mu + \alpha_i + \varepsilon_i, i = 1, ..., m,$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_I)$ is multivariate normal with expectation vector equal to zero and with a covariance matrix given by $\sigma_\alpha^2 \mathbf{C}$ while $varepsilon_i$ is a vector of iid random variables normally distributed with expectation zero and variance σ_ε^2.

(a). Estimate the variogram for the data and use this to argue why a model including correlations between the α_i's is reasonable to consider.

(b). Fit a Matern variogram to the data using $\theta_2 = \kappa = 0.5$. Do the fit look reasonable?

Exercise 39 (Monte Carlo ML and geoRglm)
(Part of this exercise is a copy of what we went through in the lecture March 5.)

Assume the model

$$Z(\mathbf{s}) = \mathbf{X}_\mathbf{s}^T \boldsymbol{\beta} + \eta(\mathbf{s}) \tag{1}$$
$$\widehat{Z}(\mathbf{s}_i) \sim g(\cdot|Z(\mathbf{s}_i)), \quad i = 1, ..., n \tag{2}$$

where $\{\eta(\mathbf{s})\}$ is a Gaussian process with mean zero, constant variance σ^2 and correlation function $c(||\mathbf{h}||; \boldsymbol{\theta}) = \text{Corr}[\eta(\mathbf{s}), \eta(\mathbf{s} + \mathbf{h}); \theta]$. We further assume that all $\widehat{Z}(\mathbf{s})$ are independent given the $Z(\mathbf{s})'s$. Our aim is to estimate the parameters $\boldsymbol{\psi} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$ based on observations $\widehat{\mathbf{Z}} = (\widehat{Z}(\mathbf{s}_1), ..., \widehat{Z}(\mathbf{s}_n))^T$.

Define $\mathbf{Z} = (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))^T$, $f_1(\mathbf{Z}; \boldsymbol{\psi})$ to be the distribution for $\mathbf{Z}$ and $f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z})$ to be the conditional distribution for $\widehat{\mathbf{Z}}$ given $\mathbf{Z}$. Note that the observation distribution is assumed to be completely known (i.e. no unknown parameters are involved). We furthermore will use $f_2(\widehat{\mathbf{Z}}; \boldsymbol{\psi})$ to be the marginal distribution for $\widehat{\mathbf{Z}}$.

(a). Show that the likelihood function $L(\boldsymbol{\psi})$ is given by

$$L(\boldsymbol{\psi}) = f_2(\widehat{\mathbf{Z}}; \boldsymbol{\psi}) = \int_{\mathbf{Z}} f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z}) f_1(\mathbf{Z}) d\mathbf{Z}$$

(b). Define $\tilde{f}(\mathbf{Z}, \widehat{\mathbf{Z}}) = \tilde{f}_1(\mathbf{Z}) f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z}))$ where $\tilde{f}_1(\mathbf{Z})$ is an arbitrary fixed density for $\mathbf{Z}$. Show that

$$\begin{aligned}
L(\boldsymbol{\psi}) &= \int_{\mathbf{Z}} f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z}) f_1(\mathbf{Z}; \boldsymbol{\psi}) d\mathbf{Z} \\
&= \int_{\mathbf{Z}} f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z}) \frac{f_1(\mathbf{Z}; \boldsymbol{\psi})}{\tilde{f}_1(\mathbf{Z})} \tilde{f}_1(\mathbf{Z}) d\mathbf{Z} \\
&= \int_{\mathbf{Z}} \tilde{f}_{1|2}(\mathbf{Z}|\widehat{\mathbf{Z}}) \tilde{f}_2(\widehat{\mathbf{Z}}) \frac{f_1(\mathbf{Z}; \boldsymbol{\psi})}{\tilde{f}_1(\mathbf{Z})} d\mathbf{Z} \\
&= \tilde{f}_2(\widehat{\mathbf{Z}}) \int_{\mathbf{Z}} \tilde{f}_{1|2}(\mathbf{Z}|\widehat{\mathbf{Z}}) \frac{f_1(\mathbf{Z}; \boldsymbol{\psi})}{\tilde{f}_1(\mathbf{Z})} d\mathbf{Z}
\end{aligned}$$

where

$$\tilde{f}_2(\widehat{\mathbf{Z}}) = \int_{\mathbf{Z}} \tilde{f}_1(\mathbf{Z}) f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z}) d\mathbf{Z}$$

and

$$\tilde{f}_{1|2}(\mathbf{Z}|\widehat{\mathbf{Z}}) = \frac{\tilde{f}_1(\mathbf{Z}) f_{2|1}(\widehat{\mathbf{Z}}|\mathbf{Z})}{\tilde{f}_2(\widehat{\mathbf{Z}})}$$

Give an interpretation of $\tilde{f}_2(\widehat{\mathbf{Z}})$ and $\tilde{f}_{1|2}(\mathbf{Z}|\widehat{\mathbf{Z}})$ and use this to argue how to use Monte Carlo approximations in order to optimize $L(\boldsymbol{\psi})$.

(c). One possible choice is to use $\tilde{f}_1(\mathbf{Z}) = f_1(\mathbf{Z}|\boldsymbol{\psi}_0)$, i.e. the model with a fixed set of parameters. Assume now that we want to compare different models through the AIC criterion

$$AIC = -2 \log L(\hat{\boldsymbol{\psi}}_{ML}) + 2|\boldsymbol{\psi}|$$

where $\hat{\psi}_{ML}$ is the maximum likelihood estimate under the given model while $|\psi|$ is the number of parameters involved. Argue why such a choice of $\tilde{f}_1(\mathbf{Z})$ within each model to be considered may be problematic and that a common $\tilde{f}_1(\mathbf{Z})$ for all models to be considered is preferable.

(d). For the rongelap data considered in the lecture, repeat the analysis using a Poisson distribution for the data with the intensity given by $\kappa(\mathbf{s}) = \exp(Z(\mathbf{s}))$ or equivalently $Z(\mathbf{s}) = \log \kappa(\mathbf{s})$.

Christensen(2004) "Monte Carlo Maximum Likelihood in Model-Based Geostatistics" considered more general models given by the Box-Cox transformation

$$Z(\mathbf{s}) = \begin{cases} (\kappa(\mathbf{s})^\lambda - 1)/\lambda & \text{if } \lambda > 0 \\ \log \kappa(\mathbf{s}) & \text{if } \lambda = 0 \end{cases}$$

and argued that $\lambda = 1$ gives a better model for the rongelap data.

(e). In order for comparison of models, the call to *prepare.likfit.glsm* need the extra option *use.intensity = TRUE*, i.e. the call should now be

```
mcmcobj <- prepare.likfit.glsm(MCmle.input.fixed, use.intensity = TRUE)
```

which then can be followed by a call to *likfit.glsm* given by

```
lik.expon.boxcox.1 <- likfit.glsm(mcmcobj, ini.phi=10, fix.nugget.rel=TRUE,
                                  lambda=1,cov.model="exponential")
```

Perform these commands and use the AIC criterion to compare the two models.

(f). An alternative to specify a value on λ is to estimate it. Look at the *help* page for the *likfit.glsm* routine and now estimate λ as well. Compare the model you now get with the previous ones. Which one do you prefer?

Exercise 40 (The Potts model)
Assume $X_1, ..., X_n$ are random variables over a lattice indexed by $i, ..., n$. Assume $X_i \in \{1, ..., K\}$, i.e. X_i is a categorical variable. Assume the probability for a particular configuration $\mathbf{x} = (x_1, ..., x_n)$ is given by

$$\Pr(\mathbf{X} = \mathbf{x}) = C \exp\{\sum_{i=1}^n \alpha_{i,x_i} + \tfrac{1}{2}\beta \sum_i \sum_{j \in N_i} I(x_i = x_j)\}$$

where N_i is the set of neighbor points to i while $I(\cdot)$ is the indicator function and C is a proportionality constant. This is called the Potts model within statistical mechanics, but is widely used in other fields such as image analysis.

(a). Write down a formula for C. How many terms needs to be added if $n = 100^2$ and $K = 4$?

(b). Consider the case where $\beta = 0$, what kind of interpretation can you have on $\alpha_{i,k}, i = 1, ..., n, k = 1, ..., K$?

(c). Assuming now $\beta > 0$, what kind of interpretation do you have on this parameter?

(d). Show that

$$\Pr(x_i = k | x_j, j \neq i) = \frac{\exp\{\alpha_{i,k} + \beta \sum_{j \in N_i} I(x_j = k)\}}{\sum_l \exp\{\alpha_{i,l} + \beta \sum_{j \in N_i} I(x_j = l)\}}$$

Direct calculations from the Potts models is usually intractable, and Monte Carlo methods are usually applied instead.

Simulation from a Potts model can be performed by a Markov Chain Monte Carlo algorithm which is an iterative algorithm sampling $\mathbf{X}^{s+1}$ conditional on $\mathbf{X}^s$ in such a way that $\mathbf{X}^s$ is approximately distributed according to the Potts model when s is large. In particular, a Gibbs sampler algorithm, corresponding to sampling X_i^{s+1} from the conditional distribution derived above, can easily be implemented. We will not go into the details of this here, but rather use a library in R which performs the necessary calculations for us. This library is a bit limited though, in that is assumes

$$\beta_{i,j} = \beta \text{ and } \alpha_{i,k} = \alpha_k.$$

We can obtain a sample from this special Potts model by the following R commands:

```
library(potts)
K = 4
nrow = 100
ncol = 100
alpha = rep(0,K)
beta = 1.0
theta = c(alpha, beta)
x = matrix(1, nrow = nrow, ncol = ncol)
foo = packPotts(x, k)
out = potts(foo, theta, nbatch = 10)
image(out$final)
```

Here **alpha** is a vector of length K (the number of categories) while **beta** is a scalar.

Markov Chain Monte Carlo methods should be used with care, since they might fail to converge. A usual way to check this is to look at trace plots of $g(\mathbf{X}^s)$ for some function(s) $g(\cdot)$ and see if it both has stabilized and is varying reasonable. Many choices of $g(\cdot)$ are

possible in the **potts** routine, see the help function on how to do this. By default, the functions

$$g_k(\mathbf{x}) = \sum_{i=1}^{n} I(x_i = k), k = 1, ..., K$$

and

$$g_{K+1}(\mathbf{x}) = \sum_{i} \sum_{j \in N_i} I(x_i = x_j)$$

are considered and the latter can be plotted by the command

```
plot.ts(out$batch[,K+1])
```

(e). Try out the commands above for different values of β. Try in particular $\beta = 0$ and some values above 2. Look particularly at the g_{K+1} function and explore if convergence seems to be reached for the different values of β.

Assume now that $\mathbf{X} = (X_1, ..., X_n)$ is not directly observed, but that it is indirectly observed through some variables $Y_1, ..., Y_n$ who are conditionally (on $\mathbf{X}$) independent and where we assume that

$$Y_i | X_i = k \sim f_k(\cdot)$$

with f_k some known distribution.

(f). Show that

$$\Pr(\mathbf{x}|\mathbf{y}) = \widetilde{C} \exp\{\sum_{i=1}^{n} \tilde{\alpha}_{i,x_i} + \tfrac{1}{2}\beta \sum_{i} \sum_{j \in N_i} I(x_i = x_j)\}$$

for some appropriate choices of $\tilde{\alpha}_{i,k}$ and where $\widetilde{C}$ is a new proportionality constant.

Even if we assume $\alpha_{i,k} = \alpha_k$, will we now be able to use the **potts** library?

Exercise 41 (Potts model (cont))
Assume again the Potts model

$$\Pr(\mathbf{X} = \mathbf{x}) = C \exp\{\sum_{i=1}^{n} \alpha_{i,x_i} + \tfrac{1}{2}\beta \sum_{i} \sum_{j \in N_i} I(x_i = x_j)\}$$

where $X_i \in \{1, ..., K\}$ and $n = n_1 \times n_2$ corresponding to a $n_1 \times n_2$ grid of points (so i is actually a two-dimensional index). The Gibbs sampler, which is a special case of a Markov chain Monte Carlo algorithm, is an iterative process for which $\mathbf{X}^{s+1}$ at iteration $s + 1$ is generated by the following updates:

- For $i = 1, ..., n$ simulate X_i^{s+1} from the conditional distribution of X_i given that for $j \neq i$

$$X_j = \begin{cases} X_j^{s+1} & \text{if } j < i \\ X_j^s & \text{if } j > i. \end{cases}$$

Here some ordering of the variables are needed, but this can be arbitrary. Note that the conditional distributions are easy to sample from. For s large, $\mathbf{X}^s$ will approximately follow the Potts model. On the course web page (under "exercises") an R-function which performs such simulations is given. This routine runs **burnin+nMCMC** iterations where the first **burnin** $\mathbf{X}^s$'s are discarded while the last **nMCMC** $\mathbf{X}^s$'s are stored.

(a). Have a look on the code and see if you can understand the essentials.

(b). Simulate $\mathbf{X}$ for $n_1 = n_2 = 20, K = 3, \beta = 0.9$ and $\alpha_{i,k} = 0$ and with the default values for **burnin** and **nMCMC**.

Pick out the values of $g(\mathbf{X}^s)$ for $g(\mathbf{X}) = \sum_i I(X_i = k)$. Make a time-series plot of the values and also a table of the realizations.

Do similarly for $g(\mathbf{X}^s) = \sum_{i \sim j} I(X_i = X_j)$. Compare the values you get with what you would expect if $\beta = 0$.

In the following we will use $\mathbf{X} = \mathbf{X}^{burnin+nMCMC}$, i.e. the last simulated value of $\mathbf{X}$.
Assume now that $Y_1, ..., Y_n$ are conditionally (on $\mathbf{X}$) independent and

$$Y_i | X_i \sim N(\mu_k, \sigma^2)$$

(c). Simulate the Y_i's conditional on the X_i's with $\mu_k = k$ and $\sigma = 0.5$.

Our aim now will be to reconstruct $\mathbf{X}$ based on the Y_i's.

(d). One possible prediction of X_i is

$$\widehat{X}_i^{naive} = \text{argmax}_k \, \text{Pr}(X_i = k | Y_i).$$

Assuming $\text{Pr}(X_i = k) = K^{-1}$, show that

$$\text{argmax}_k \, \text{Pr}(X_i = k | Y_i) = \text{argmax}_k f_k(Y_i)$$

Calculate $\widehat{X}_i^{naive}$ for all i and also calculate the error rate for this predictor,

$$\textbf{err}^{naive} = n^{-1} \sum_i I(\widehat{X}_i^{naive} = X_i).$$

An alternative to the naive predictor above is to put

$$\widehat{X}^{mode} = \text{argmax}_{\mathbf{x}} \Pr(\mathbf{X} = \mathbf{x}|\mathbf{Y}).$$

Such a mode is in practice difficult to find, but an approximation can be obtained by the Iterated Conditional Mode (ICM) algorithm (Besag, 1986, On the statistical analysis of dirty pictures). This is an iterative algorithm where at each iteration

$$\widehat{X}_i^{s+1} = \text{argmax}_k \Pr(X_i = k|Y_i, X_j = X_j^s, j \neq i).$$

and the routine is continued until no more changes are made.

(e). Discuss the strengths and the weaknesses for the ICM algorithm.

(f). Also an ICM function is available at the webpage. Try this out and compare the error rate now with the one obtained from the naive predictor.

Yet another predictor can be obtained by

$$\widehat{X}_i^{marg.mode} = \text{argmax}_{x_i} \Pr(X_i = x_i|\mathbf{Y}).$$

(g). Explain how the marginal distribution $\Pr(X_i = x_i|\mathbf{Y})$ can be estimated through the Gibbs sampler routine. Perform the necessary simulations and calculate $\widehat{X}_i^{marg.mode}$ for each i. Again calculate the error rate and compare with the previous results.

(h). Plot the images predicted using the different methods and compare.

Exercise 42 (Estimation using sparse matrices)
(a). Simulate an AR(1) model using $n = 1000, \mu = 0, a = 0.95$ and $\sigma = 1$.

(b). Write a routine for calculating the likelihood using the spare matrix formulation from the previous exercise.

(c). Find maximum likelihood estimates by numerical optimization of the likelihood function. Compare your results with that obtained from the routine **arima** in R.

Exercise 43 (Estimation in Gaussian random fields)
In this exercise we will try to reproduce some of the results obtained in Section 4.3 of Smith: Environmental statistics. The data are available from the course web page (the *yield* data) which can be read into R by

```
yield = matrix(scan("yield.dat"),ncol=25)
```

Assume we have a Gaussian random field model with

$$X_{i,j}|X_{i',j'}, (i', j') \neq (i, j) \sim N(\alpha + \beta_1(X_{i-1,j} + X_{i+1,j}) + \beta_2(X_{i,j-1} + X_{i,j+1}), \sigma^2)$$

Define $\mathbf{X}_1 = \{X_{i,j} : |i - j| \text{ even }\}$ and $\mathbf{X}_2 = \{X_{i,j} : |i - j| \text{ odd }\}$.

(*a*). Argue that the coding methods based on maximizing $p(\mathbf{X}_1|\mathbf{X}_2)$ with respect to the unknown parameters can be performed by ordinary linear regression.

(*b*). Make a matrix containing $\mathbf{X}_1$ as the first column and the corresponding elements $X_{i-1,j} + X_{i+1,j}$ and $X_{i,j-1} + X_{i,j+1}$ as the second and third columns and use this matrix to estimate $beta_1, \beta_2$ in addition to α and σ^2. Compare the results with those obtained by Smith.

Note that you probably will not get exactly the same results, due to how edge-effects are handled. Here you can treat the covariates as missing when (i, j) is on the border.

Also perform the analysis when switching $\mathbf{X}_1$ and $\mathbf{X}_2$.

(*c*). Argue why also estimated based on the by the pseudolikelihood can be performed by ordinary linear regression. Do this and compare the results with the ones obtained by the coding method. Discuss approaches for obtaining standard errors in this case (you do not need to try this out).

Exercise 44 (Extreme values of ozone)
In this exercise we will look at extreme values for Ozone. One the course web-page, hourly ozone measurements from Birkenes over the period 1990-2008 is available and can be read into R with the command

```
ozoneB = read.table("ozoneBirkenes.dat",header=T)
```

Yearly maxima, which will be of our main interest here can be calculated by the command

```
M = as.vector(by(ozoneB$Val,ozoneB$year,max,na.rm=T))
```

where `M` becomes a vector containing the yearly maxima.

(*a*). A common assumption is that yearly maxima are iid variables. Given that is actually is a time-series over years, auto-correlations are possible.

Convince yourself on that independence indeed is plausible in this case by comparing an AR(1) model with an AR(0) model using the AIC criterion.

Hint: Use the **`arima(M,c(p,0,0))`** command for fitting an AR model of order `p`.

(*b*). The R library **`evd`** contains a routine **`fgev`** which can be used for fitting the GEV distribution to data.

Perform such a fitting on the ozone data, both estimating the shape parameter ξ and putting the shape parameter to zero (the latter can be obtained by including the option **`shape=0`** to the **`fgev`** call.

The output from **`fgev`** include a deviance measure which is equal to $-2 * \text{loglik}$ (up to an additive constant). Use the AIC criterion in order to decide whether $\xi \neq 0$ is reasonable or not.

(c). For your fitted model, specify the value x_N such that

$$1 - H(x_N; \hat{\mu}, \hat{\psi}, \hat{\xi}) = \frac{1}{N}.$$

Construct standard errors and confidence intervals for x_N using parametric boot-strapping. (Hint: Simulation from a GEV model can be performed by the **rgev** routine).

Discuss the meaning of x_N.

Exercise 45 (Extreme values of ozone and the Poisson-GPD model)
Continuing on the data from the previous exercise, we will now consider fitting models to exceedances above $u = 120$ using the Poisson-GPD model. Within the R-library **evd** there is a routine **fpot** that performs the fitting for us.

(a). Take a look on the **fpot** help page in order to understand the main issues. In particular, look at the **cmax** option and the corresponding routine **clusters**.

(b). For the ozone data set, play around with the **fpot** routine. Also plot the results and try to understand the outputs.

Exercise 46 (Spatio-temporal processes)
(a). Assume

$$Z(t, \mathbf{x}) = S(\mathbf{x}) + T(t) + \varepsilon(t, \mathbf{x})$$

where the tree terms on the right hand side are assumed independent. Further, all $\varepsilon(t, \mathbf{x})$ are independent in time with expectation zero and covariance function $C_\varepsilon(\mathbf{v})$. $\{S(\mathbf{x})\}$ is a pure spatial process, stationary with covariance function $C_S(\mathbf{v})$ while $\{T(t)\}$ is an AR(1) model given by

$$T(t) = aT(t - 1) + \eta(t)$$

with initial conditions such that the process becomes stationary and $\eta(t)$ are iid, zero mean and constant variance σ_T^2.

Calculate the covariance $\text{cov}[Z(t, \mathbf{x}), Z(t + s, \mathbf{x} + \mathbf{v})]$ and show that it is stationary both in space and time.

(b). Consider now an alternative model

$$Z(t, \mathbf{x}) = S(\mathbf{x}) + k[Z(t - 1, \mathbf{x}) - S(\mathbf{x})] + c\varepsilon(t, \mathbf{x})$$

with the individual components having the same properties as above.

Express $\text{cov}[Z(t, \mathbf{x}), Z(t, \mathbf{x} + \mathbf{v})]$ by $\text{cov}[Z(t - 1, \mathbf{x}), Z(t - 1, \mathbf{x} + \mathbf{v})]$. Assuming stationarity in time, i.e. $\text{cov}[Z(t, \mathbf{x}), Z(t, \mathbf{x} + \mathbf{v})] = C(0, \mathbf{v})$, derive an expression for $C(0, \mathbf{v})$.

Also derive $C(s, \mathbf{v}) = \text{cov}[Z(t, \mathbf{x}), Z(t + s, \mathbf{x} + \mathbf{v})]$ for general s (start with $s = 1$ and do the calculations recursively).

Can you find values of k and c such that you end up with the model given in (a)?

(c). Now assume $\mathbf{x}$ is defined on a regular grid and consider the model

$$Z(t, \mathbf{x}_i) = S(\mathbf{x}_i) + k_1[Z(t-1, \mathbf{x}_i) - S(\mathbf{x}_i)] + k_2 \sum_{j \sim i}[Z(t-1, \mathbf{x}_j) - S(\mathbf{x}_j)] + c\varepsilon(t, \mathbf{x}_i)$$

where $j \sim i$ means that j is one of the four closest neighbors of i.

Assuming $E[Z(0, \mathbf{x}_i) = \mu$ for all $\mathbf{x}_i$, what is the expectation of $E[Z(t, \mathbf{x}_i)$ for a general t?

Assuming $\mathsf{cov}[Z(0, \mathbf{x}_i), \mathbf{Z}(0, \mathbf{x}_j)] = C_0(||\mathbf{x}_j - \mathbf{x}_i||)$, try to derive an expression for $\mathsf{var}[Z(1, \mathbf{x}_i), Z(1, \mathbf{x}_j)]$. Is the variance at time 1 also constant in space?

(d). Consider now a continuous version of the previous model:

$$Z(t, \mathbf{x}) = S(x) + \int_{\mathbf{v}}[Z(t-1, \mathbf{x} + \mathbf{v}) - S(\mathbf{x} + \mathbf{v})]k(\mathbf{v})d\mathbf{v} + \varepsilon(t, \mathbf{x}_i)$$

where $k(\cdot)$ is a kernel function assumed to be integrable.

Assume $E[Z(0, \mathbf{x}) = \mu$ for all $\mathbf{x}$ and $\mathsf{cov}[Z(0, \mathbf{x}), \mathbf{Z}(0, \mathbf{x} + \mathbf{v})] = C_0(||\mathbf{v}||)$.

What requirements are needed on $k(\cdot)$ in order for $E[Z(t, \mathbf{x}) = \mu$ for all t?

Derive an expression for $\mathsf{var}[Z(1, x)]$ as a function of C_S, C_0 and $\sigma^2 = \mathsf{var}[\varepsilon(t, \mathbf{x}_i)]$.

Hint: It might be easier to consider $\widetilde{Z}(t, \mathbf{x}) = Z(t, \mathbf{x}) - S(\mathbf{x})$.

Discuss verbally what kind of requirements are needed in order for the variance not to increase in time.

Exercise 47 (Doctor-Prescription Amounts per Consultation)
Following the description in sec 4.2.6 in the text book, assume the hierarchical model

$$Z(\mathbf{s}_i)|Y(\mathbf{s}_i), \sigma_\varepsilon^2 \sim \text{ind.Gau}(Y(\mathbf{s}_i), \sigma_\varepsilon^2)$$
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is multivariate Gaussian. At the course web-page there is a script for doing analysis of these data using the "Besag" model in INLA, that is

$$\delta_i|\delta_{-i} \sim N(\frac{1}{n_i} \sum_{j \in \mathcal{N}_i} \delta_j, \frac{\sigma^2}{n_i})$$

where $n_i = |\mathcal{N}_i|$ is the number of neighbors of i.

(a). Try out both Bayesian and empirical Bayesian estimation using INLA. Compare the estimates and comment on differences.

In the textbook an alternative model is used, taking into account that each observation are based on different amounts of data. This model is not directly available in INLA, but we will see that we can make proper transformations in order to obtain the model anyway. We will first consider this in a general setting. Assume therefore the CAR model

$$\boldsymbol{\delta} \sim \text{MVN}(\mathbf{0}, [\mathbf{I} - \mathbf{C}]^{-1}\mathbf{M})$$

where $\mathbf{C} = \phi\mathbf{H}$ and $\mathbf{M} = \sigma^2\boldsymbol{\Delta}$ and where $\mathbf{H}$ is a known sparse matrix and $\boldsymbol{\Delta}$ is a known diagonal matrix.

(b). Show that the precision matrix for $\mathbf{Y}$ is equal to $\frac{1}{\sigma^2}\boldsymbol{\Delta}^{-1}[\mathbf{I} - \phi\mathbf{H}]$.

(c). Define $\widetilde{\mathbf{Y}} = \boldsymbol{\Delta}^{-1/2}\mathbf{Y}$. Show that

$$E[\widetilde{\mathbf{Y}}] = \widetilde{\mathbf{X}}\boldsymbol{\beta}$$

where $\widetilde{\mathbf{X}}$ is specified through $\mathbf{X}$ and $\boldsymbol{\Delta}$.

(d). Find the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ for $\widetilde{\mathbf{Y}}$ and use this to show that the corresponding precision matrix $\widetilde{\mathbf{Q}}$ is of the form

$$\widetilde{\mathbf{Q}} = \frac{1}{\sigma^2}[\mathbf{I} - \phi\widetilde{\mathbf{H}}]$$

where $\widetilde{\mathbf{H}}$ is a sparse matrix with the same non-zero elements as $\mathbf{H}$.

(e). Assume first that $\phi = 0$ and also $\sigma_\varepsilon^2 = 0$.

INLA has a model called *generic0* which assumes the precision matrix $\mathbf{Q} \propto \mathbf{C}$. Show that this is the case for $\phi = 0, \sigma_\varepsilon^2 = 0$. Use the commands in the script (the ones resulting in *res0*) to fit this model. Compare the estimates of the parameters with those obtained from the Besag model.

(f). Using the transformed variables $\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}$, show that this correspond to the δ_i's being iid. Use the commands resulting in *res0.2* to fit this alternative. Again compare with previous results.

(g). Now consider the spatial case but still with $\sigma_\varepsilon^2 = 0$. INLA has a model called *generic0* which assumes the precision matrix $\mathbf{Q} \propto \mathbf{I} - \phi\mathbf{C}$. Use the commands resulting in *res2* to fit this model. Again compare with the previous results.

(h). Finally consider the general case. Use the commands resulting in *res3* to fit this model. Again compare with the previous results.

Exercise 48 (Separable covariance functions)
Assume a separable covariance function

$$\text{cov}[Y(\mathbf{s};t), Y(\mathbf{x};r)] = C(\mathbf{s},t,\mathbf{x},r) = C^{(s)}(\mathbf{s},\mathbf{x}) \cdot C^{(t)}(t,r)$$

We will show that if $C^{(s)}(\mathbf{s}, \mathbf{x})$ is a legal covariance function in space and $C^{(t)}(t, r)$ is a legal covariance function in time, then their product become a legal covariance function in space-time.

(a). Assume first $\mathbf{Y} = \{Y(\mathbf{s}_i, t_j), i = 1, ..., m, j = 1, ..., n\}$. Show that the covariance matrix $\boldsymbol{\Sigma}_Y$ of $\mathbf{Y}$ can be written as $\boldsymbol{\Sigma}^{(s)} \otimes \boldsymbol{\Sigma}^{(t)}$ where $\boldsymbol{\Sigma}^{(s)}$ is defined through $C^{(s)}(\mathbf{s}, \mathbf{x})$ and $\boldsymbol{\Sigma}^{(t)}$ is defined through $C^{(t)}(t, r)$. Here $\otimes$ is the kroenecker product, i.e.

$$
\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \cdots & A_{2n}\mathbf{B} \\ \vdots & & & \\ A_{n1}\mathbf{B} & A_{n2}\mathbf{B} & \cdots & A_{nn}\mathbf{B} \end{pmatrix}
$$

(b). Show that $\boldsymbol{\Sigma}_Y$ becomes non-negative definite.

(c). Use this to show that the covariance matrix for any collection of variables $\{Y(\mathbf{s}_k, t_k), k = 1, , , ., N\}$ becomes non-negative definite.

Exercise 49 (MRF's for "geostatistical" data)
This exercise is now a part of the project for v13

The file *Piemonte_data_byday.csv* contain data on PM_{10} (particulate matter with an aerodynamic diameter of less than 10 μm) daily observed over a period from 01/10/05 to 31/03/06 at 42 stations in the North-Italian region Piemonte.

This is a dataset taken from Cameletti et al. [2011] which we will analyse using spatio-temporal models later. Here we will only use data from one day (day 110) and concentrating on the spatial structure In order to make the response approximately Gaussian distributed, we will consider $\log(PM_{10})$.

The dataset can be read into R by the commands

```
d <-read.table("Piemonte_data_byday.csv",header=TRUE,sep=",")
d$Time = rep(1:182,each=24)
d1 = d[d$Time==110,]
d1$logPM10 = log(d1$PM10)
```

d1 now also contain several covariates, but we will ignore these in this exercise. We also need the spatial coordinates which are available by

```
coordinates <-read.table("Cameletti/coordinates.csv",header=TRUE,sep=",")
borders = read.table("Cameletti/Piemonte_borders.csv",header=TRUE,sep=",")
```

It might also be a good idea to remove those NA's:

```
coord = coord[!is.na(d1$PM10),]
d1 = d1[!is.na(d1$PM10),]
```

(a). Do different types of exploratory analyses to evaluate the spatial structure of the data

(b). Fit an exponential covariance function to the data. Do the fit look reasonable? There is also a second dataset available, which can be used for evaluation.

References

M. Cameletti, F. Lindgren, DP Simpson, and H. Rue. Spatio-temporal modelling of particulate matter concentration through the spde approach. AStA Adv Stat Anal, Submitted, 2011.

Robert Cruikshanks, Rasmus Lauridsen, Alison Harrison, Mark GJ Hartl, Mary Kelly-Quinn, Paul S Giller, and John OHalloran. Evaluation of the use of the sodium dominance index as a potential measure of acid sensitivity. WATERAC Final Report. EPA and COFORD, Ireland, 2006.

Alain F Zuur, Elena N Ieno, Neil Walker, Anatoly A Saveliev, and Graham M Smith. Mixed effects models and extensions in ecology with R. Springer, 2009.