Consider the problem of using spatial data $\boldsymbol{X} = (X(s_1), \ldots, X(s_n))'$ to obtain the linear predictor $\hat{X}(s_0)$ that minimizes the error $E(\hat{X}(s_0) - X(s_0))^2$. As $\hat{X}$ is a linear predictor, we have that $\hat{X}(s_0) = \lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X}$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$. Plugging in the predictor in the error formula

$$E(\hat{X}(s_0) - X(s_0))^2 = \mathrm{var}(\boldsymbol{\lambda}'\boldsymbol{X} - X(s_0)) + (\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{\mu} - \boldsymbol{\mu}(s_0))^2$$

where $\mu(s)$ is the mean of $X(s)$.

Consider the problem of using spatial data $\boldsymbol{X} = (X(s_1), \ldots, X(s_n))'$ to obtain the linear predictor $\hat{X}(s_0)$ that minimizes the error $E(\hat{X}(s_0) - X(s_0))^2$. As $\hat{X}$ is a linear predictor, we have that $\hat{X}(s_0) = \lambda_0 + \boldsymbol{\lambda}'\boldsymbol{X}$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$. Plugging in the predictor in the error formula

$$E(\hat{X}(s_0) - X(s_0))^2 = \mathrm{var}(\boldsymbol{\lambda}'\boldsymbol{X} - X(s_0)) + (\lambda_0 + \boldsymbol{\lambda}'\boldsymbol{\mu} - \boldsymbol{\mu}(s_0))^2$$

where $\mu(s)$ is the mean of $X(s)$.

The second term has minimum value equal to 0. Letting $\sigma^2 = \mathrm{var}(X(s))$, $\boldsymbol{\sigma} = \mathrm{cov}(\boldsymbol{X}, X(s_0))$ and $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{X})$, then

$$\mathrm{var}(\boldsymbol{\lambda}'\boldsymbol{X} - X(s_0)) = \sigma^2 + \boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda} - 2\boldsymbol{\sigma}'\boldsymbol{\lambda}$$

which has a minimum at $\boldsymbol{\lambda}'\Sigma = \boldsymbol{\sigma}'$.

Thus, the optimal choices of λ_0 and $\boldsymbol{\lambda}$ are

$$\lambda_0 = \mu(s_0) - \boldsymbol{\lambda}'\boldsymbol{\mu} \text{ and } \quad \boldsymbol{\lambda} = \Sigma^{-1}\boldsymbol{\sigma}$$

So, the optimal linear predictor is

$$\hat{X}(s_0) = \mu(s_0) + \boldsymbol{\sigma}'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$$

This predictor is known in the traditional geostatistics literature as Simple Kriging.

Mean Square Prediction

More generally, consider the problem of predicting the value of a random variable T using the observed values Y. Denote the predictor as $\hat{T}$. Then the Mean Square Prediction Error is

$$MSE(\hat{T}) = E(T - \hat{T})^2$$

where the expectation is taken WRT the joint distribution of T and $\hat{T}$.

Mean Square Prediction

More generally, consider the problem of predicting the value of a random variable T using the observed values Y. Denote the predictor as $\hat{T}$. Then the Mean Square Prediction Error is

$$MSE(\hat{T}) = E(T - \hat{T})^2$$

where the expectation is taken WRT the joint distribution of T and $\hat{T}$.

Theorem: $MSE(\hat{T})$ is minimized at $\hat{T} = E(T|Y)$

Proof: $E(T-\hat{T})^2 = E_Y(E_T(T-\hat{T})^2|Y))$. The inner expectations is

$$E_T(T-\hat{T})^2|Y) = \mathrm{var}_T(T-\hat{T}|Y) + E_T^2(T-\hat{T}|Y)$$

Given Y, any function Y is constant, thus

$$E_T(T-\hat{T})^2|Y) = \mathrm{var}_T(T|Y) + (E_T(T|Y) - \hat{T})^2$$

and

$$E_Y E_T(T-\hat{T})^2|Y) = E_Y \mathrm{var}_T(T|Y) + E_Y(E_T(T|Y) - \hat{T})^2$$

which is minimized when $\hat{T} = E(T|Y)$.

Proof: $E(T-\hat{T})^2 = E_Y(E_T(T-\hat{T})^2|Y))$. The inner expectations is

$$E_T(T-\hat{T})^2|Y) = \mathrm{var}_T(T-\hat{T}|Y) + E_T^2(T-\hat{T}|Y)$$

Given Y, any function Y is constant, thus

$$E_T(T-\hat{T})^2|Y) = \mathrm{var}_T(T|Y) + (E_T(T|Y) - \hat{T})^2$$

and

$$E_Y E_T(T-\hat{T})^2|Y) = E_Y \mathrm{var}_T(T|Y) + E_Y(E_T(T|Y) - \hat{T})^2$$

which is minimized when $\hat{T} = E(T|Y)$.

From the former proof we have that the Prediction Variance is

$$E(T-\hat{T})^2 = E_Y(\mathrm{var}(T|Y))$$

Conditional Normal Distribution

Let $Y = (Y_1, Y_2)$ a joint multivariate normal vector with mean $\mu = (\mu_1, \mu_2)$, and covariance matrix

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

Then the conditional distribution of Y_1 given Y_2 is $N(\mu_{1|2}, \Sigma_{1|2})$ where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2) \quad \text{and} \quad \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Predictions for Gaussian Processes

Assume that m realizations of a Gaussian process $X(s)$, $X = (X(s_1), \ldots, X(s_m))$ have been observed with a $N(0, \tau^2)$ error. Suppose that the unobserved values of the process at locations $s_1^*, \ldots, s_p^*$ need to be predicted. Let $X^* = (X(s_1^*), \ldots, X(s_p^*))$. Then we have that $Y_1 = X$ and $Y_2 = X^*$.

Suppose the mean of the process is μ and the covariance is $\sigma^2 \rho(||s - s'||)$, then

$$EX = 1_m \mu \ , \ \ \mathrm{var}(X) = \sigma^2 R + \tau^2 I = \sigma^2 V$$

and

$$EX^* = 1_p \mu \ , \ \ \mathrm{var}(X^*) = \sigma^2 R^*$$

Here $R_{ij} = \rho(||s_i - s_j||)$ and $R_{ij}^* = \rho(||s_i^* - s_j^*||)$

The predicted values of the Gaussian process at the new locations are

$$\hat{X} = \mathbb{1}_p\mu + rV^{-1}(X - \mathbb{1}_m\mu)$$

where r is a matrix such that $r_{ij} = \rho(||s_i^* - s_j||)$.

The prediction variance is given by

$$\sigma^2(R^* - rV^{-1}r')$$

The predicted values of the Gaussian process at the new locations are

$$\hat{X} = \mathbb{1}_p \mu + r V^{-1}(X - \mathbb{1}_m \mu)$$

where r is a matrix such that $r_{ij} = \rho(||s_i^* - s_j||)$.

The prediction variance is given by

$$\sigma^2 (R^* - r V^{-1} r')$$

Notice that, if there is no observational error, $p = m$ and $s_i = s_i^*, i = 1 \ldots, m$, then $r = V$. Thus

$$\hat{X} = \mathbb{1}_p \mu + r V^{-1}(X - \mathbb{1}_m \mu) = X$$

so the prediction is an interpolation.

Prediction of Linear Functionals

To predict an averaged process like

$$T = \int_A w(s)X(s)ds$$

we use the conditional expectation

$$\hat{T} = E(T|X) = \int_A w(s)E(X(s)|X)ds = \int_A w(s)\hat{X}(s)ds$$

Under Gaussianity of X, this is the mean of a normal random variable whose variance is

$$\text{var}(T|X) = \int_A \int_A w(s)w(s')\text{cov}(X(s), X(s'))dsds'$$

var$(T|X)$ involves a double integral and so, it can be computationally too demanding to calculate in practice. Notice that the kriging predictor of T is $E(T|X) = \mu + \Sigma_{T,X}\Sigma_{X,X}^{-1}(X - \mu)$. So the pointwise predictor involves only integrals of the form

$$\int_A w(s')\mathrm{cov}(X(s), X(s'))ds'$$

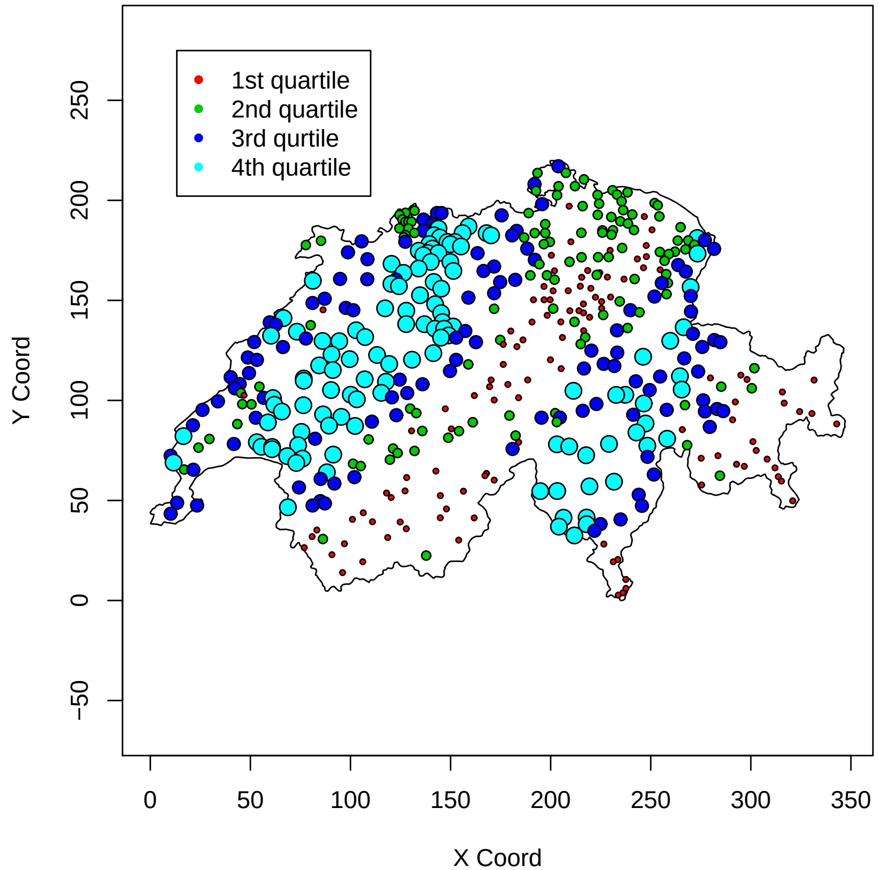
Prediction of Linear Functionals

$\text{var}(T|X)$ involves a double integral and so, it can be computationally too demanding to calculate in practice. Notice that the kriging predictor of T is $E(T|X) = \mu + \Sigma_{T,X}\Sigma_{X,X}^{-1}(X - \mu)$. So the pointwise predictor involves only integrals of the form

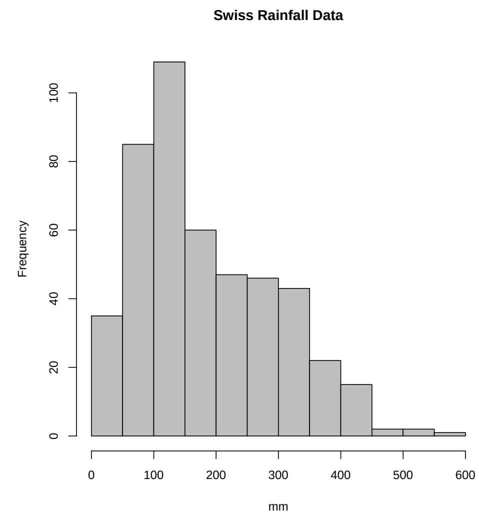
$$\int_A w(s')\text{cov}(X(s), X(s'))ds'$$

On the other hand, if T was a finite linear combination of, say, n X values, then $\text{var}(T|X) = 0$, as we can exactly determine the value of T from those of X. As the integral that defines T is a limit of finite sums, we can assume $\text{var}(T|X) \approx 0$ for large enough n.

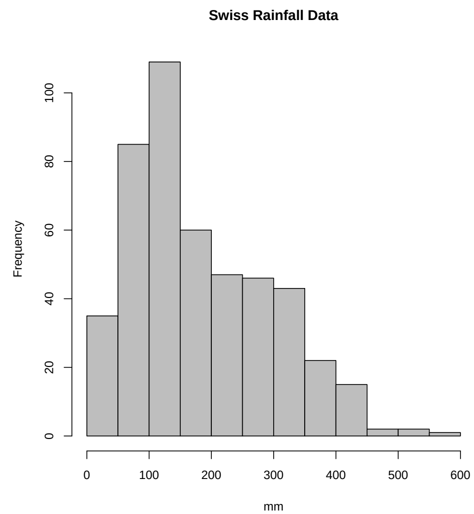
Swiss rainfall data from D&R. These data correspond to the measured rainfall on May 8 1986 at 467 locations in Switzerland. The figure was produced with `points(sic.all, borders=sic.borders ,col=2:5,pt.divide= "quartiles")`



A histogram of the Swiss rainfall data reveals some skewness that is incompatible with a Gaussian assumption.



A histogram of the Swiss rainfall data reveals some skewness that is incompatible with a Gaussian assumption.



One possibility is to use a Box-Cox transformation as

$$Y^* = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log Y & \lambda = 0 \end{cases}$$

We can perform the ML parameter estimation using `likfit` in `geoR`. Here we are using a Matern correlation function with fixed smoothness parameter 1. We are also performing a Box and Cox transformation of the observed data with fixed parameter $\lambda = .5$.

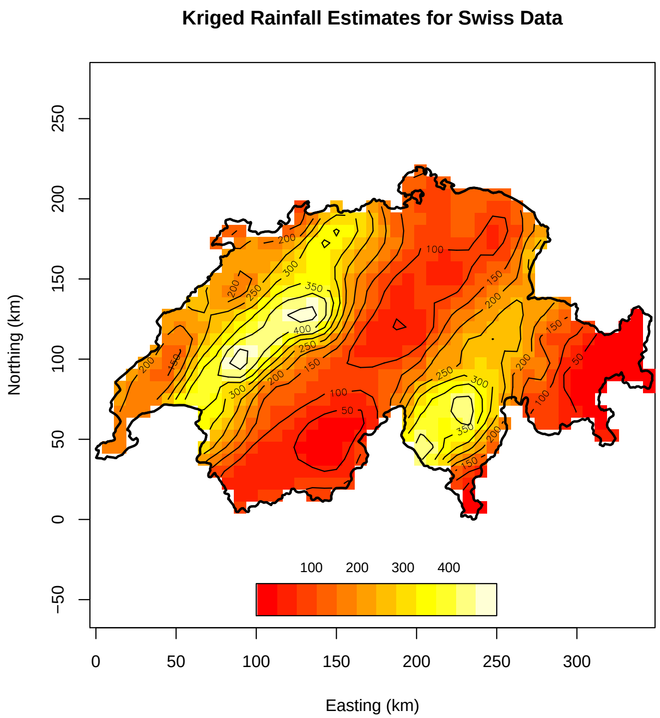
```
> ml=likfit(sic.all,ini=c(100,40),nug=10,lambda=.5,kappa=1)
> ml
likfit: estimated model parameters:
     beta     tausq   sigmasq       phi
" 20.134" "  6.921" "105.027" " 35.788"
```

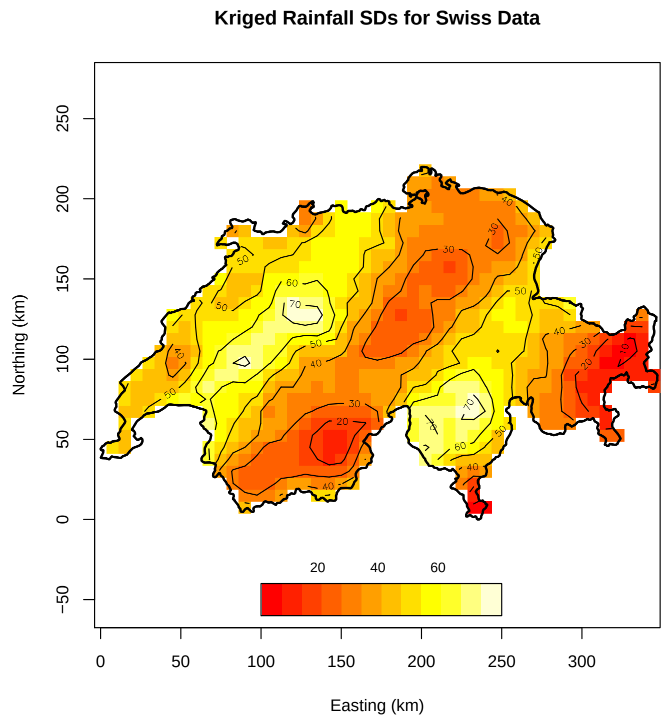
A first order trend can be obtained with

```
ml1=likfit(sic.all,trend='1st',ini=c(100,40),nug=10,
lambda=.5,kappa=1)
> ml1
likfit: estimated model parameters:
    beta0     beta1     beta2     tausq   sigmasq       phi
"24.7884" "-0.0524" " 0.0496" " 6.7465" "75.2158" "28.9214"
```

We can obtain the kriged surface using the following commands:

```
gr=pred_grid(sic.borders,by=7.5) #Create a grid
kc=krige.control(obj.model=ml) #Create krige control
pred=krige.conv(sic.all,loc=gr,borders=sic.borders,krige=kc)
#Calculate predictions
image(pred,x.leg=c(100,250),y.leg=c(-60,-40),ylab='Northing (km)',
xlab='Easting (km)')
contour(pred,add=T)
title('Kriged Rainfall Estimates for Swiss Data')
#Standard Deviations:
image(pred,x.leg=c(100,250),y.leg=c(-60,-40),val=sqrt(
pred$krige.var),ylab='Northing (km)',xlab='Easting (km)')
contour(pred,val=sqrt(pred$krige.var),add=T)
title('Kriged Rainfall SDs for Swiss Data')
```


Kriged Rainfall Estimates for Swiss Data
Northing (km)
Easting (km)
100 200 300 400


Kriged Rainfall SDs for Swiss Data
Northing (km)
Easting (km)
20 40 60