Spatio-Temporal Precipitation Generator Based on Latent Gaussian fields

Sara Martino

Sintef Energy AS

Vannes
20th May 2016

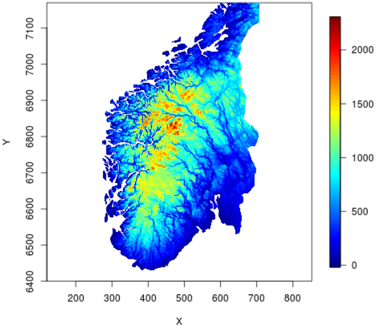
Table of contents

Motivation - Why a rain simulator for Norway

- Hydropower is a very important economic factor in Norway
- Long time production planning models need inflow (precipitation) data as input
- Hystorical data are available only at gauges, have many holes and are often unreliable
- The whole energy system in Norway in interconnected, therefore the hydropower companies would like to have a model for the whole area
- There is a strong wish from the industry to have a model that is "climate driven"

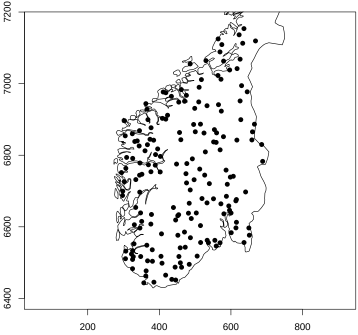
Challanges

- Large area
- Complex terrain
- Orographic precipitation
- Complex Spatial structures
- "Messy" data set



Data

- 177 Stations
- Time range: 1979-2011
- Covariates from ERA Interim reanalysis (averaged over the whole domain): Temperature, Geopotential, RH, Wind.



Model Definition

We assume there is one latent Gaussian field

$$W(s,t) \sim \mathcal{N}(\mu(s,t), \Sigma(t))$$

We define the rain field $Y(s,t)$ as follows:

$$Y(s,t) = \begin{cases} 0, & \text{if } W(s,t) \leq 0 \\ F^{-1}(s,t) \cdot \Phi^{+}(W(s;t)), & \text{if } W(s,t) > 0 \end{cases}$$

That is: we assume that for negative values of W we are in dry state, for positive values of W we transform the gaussian into a "suitable" marginal distribution.

Model Parameters

- Parameters of the marginal distribution F
- Mean and Covariance function of the latent field

Marginal Model for Intensity

We assume the intensity at each station to be distributed as a Gamma RV with mean $\alpha(s,t)$ and scale $\lambda(s)$.
We model the mean as

$$\log(\alpha(s,t)) = X^T \beta_\alpha(s)$$

The parameters $\beta_\alpha(s)$ are estimated at each station.
The covariates in X^T include some "climate" variables from ERA Interim.

Mean of W(s,t)

The mean of the latent field is linked to the Probability of wet state:

$$P(Y > 0) = P(W > 0) = \Phi(\mu)$$

If we consider the occurrence process (0,1) this is the definition of a probit model. So we can introduce covariates in the mean μ as:

$$\Phi(\mu(s,t)) = X^T \beta_\mu(s)$$

We estimate the β parameters at each station.
The covariates in X^T include seasonal components, dependence on the day before, and some "climate" variables from ERA Interim

Covariance Function

- We only observe a transformation of the latent Gaussian field
- In the wet state we can transform the observed rain into Gaussian values

$$\hat{w} = \Phi_{\mu}^{-1} \cdot F(y)$$

- We use the transformed $\hat{w}$ to estimate the covariance function

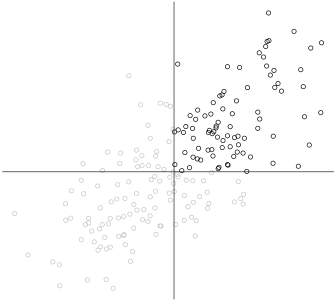
- Covariance Model

$$C(\mathbf{h}, A(t)) = \exp(-|\mathbf{d}|/A(t))$$

where

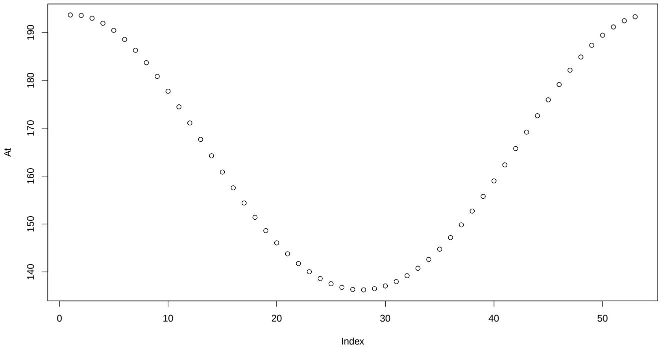
$$A(t) = a_0 + a_1 \cos(2\pi t/12) + a_2 \sin(2\pi t/12)$$

- Minimize squared distance between theoretical and empirical covariate

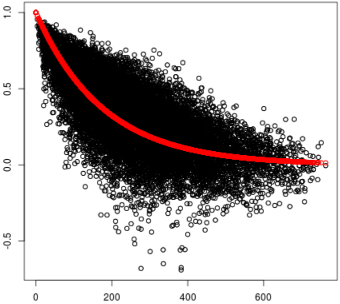


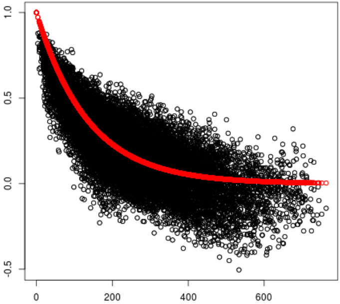
Estimated Covariance Function

Estimated range parameter



Estimated Covariance Function





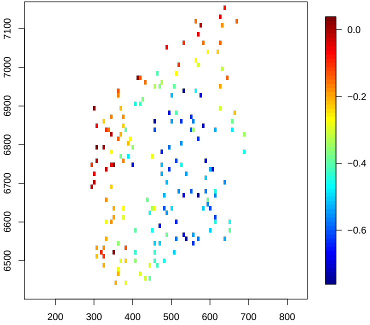
Smoothing parameters over space

- We estimated the β parameters at the station locations but we would like to simulate continuously
- Fit a spatial model

$$\hat{\beta}(s) = \xi_0 + \xi X_\beta + u(s) + \epsilon(s)$$

where:

- $\epsilon(s)$ is a nugget effect
- (u) is a Matern spatial effect represented through the SPDE approach
- ξ are possible parameters
- X are possible covariates (altitude, latitude, longitude)



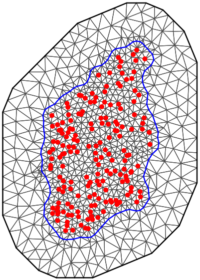
To model the smooth spatial effect we use the SPDE approach introduced by Lindgrenn et. al (2011)

- A matern field has full precition matrix, costly to simulate from on large grids.
- The aim of the SPDE approach is to is to find a GMRF, with local neighbourhood and sparse precision matrix Q, that best represents the Matern field

Spatial Effect $\mathbf{u}$ - The SPDE approach

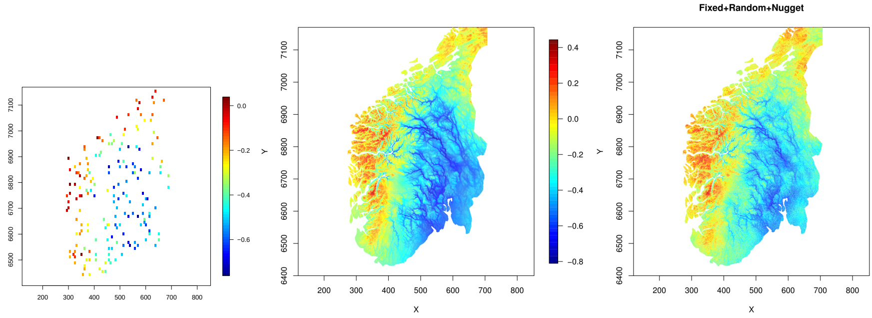
- This results in expressing the Matern field as a linear combination of basis functions defined on a triangulation of the domain $\mathcal{D}$ using n vertices.
- So the Matern field u is expressed as $u \sim \mathcal{N}(0, Q)$ with n-dimensional precision matrix Q.
- The matrix Q from the SPDE representation and is computed using Eq.(10) of Lindgren et al. (2011).
- The SPDE representation defines an explicit mapping between the parameters of the GF covariance function (range and marginal variance) to the elements of the precision matrix Q of teh GMFR

- Q is a sparse precision matrix
- Fast to simulate
- Inference through INLA
- Easy implenetation through R-INLA pachage
- Possible to introduce non-stationarity

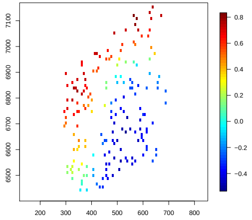


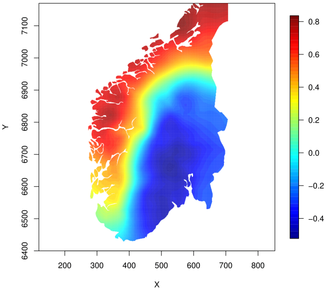
Parameter Simulation

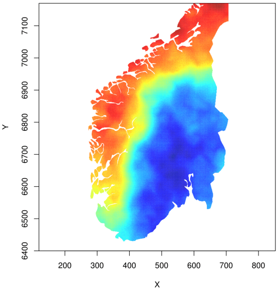
We simulate the β parameters from their posterior distribution: Example1



Example2


7100
7000
6900
6800
6700
6600
6500
200
300
400
500
600
700
800
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Y
X
7100
7000
6900
6800
6700
6600
6500
6400
200
300
400
500
600
700
800
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Fixed+Random+Nugget
Y
X
7100
7000
6900
6800
6700
6600
6500
6400
200
300
400
500
600
700
800

Simulation Algorithm

- Simulate all β parameters from the corresponding Gaussian processes
- For each dat t
  1. Simulate a realisation from the latent Gaussian field Z
  2. Set the simulated rain to 0 in all locations where $Z < 0$
  3. Compute the rain amount in the locations where $Z > 0$ using the parameters of the Gamma distribution

Model Validation - Marginals

Marginal occurrence process:

- Probabilty of dry state
- Transition Probabilities (from dry to wet etc)
- Dry Spells

Model Validation

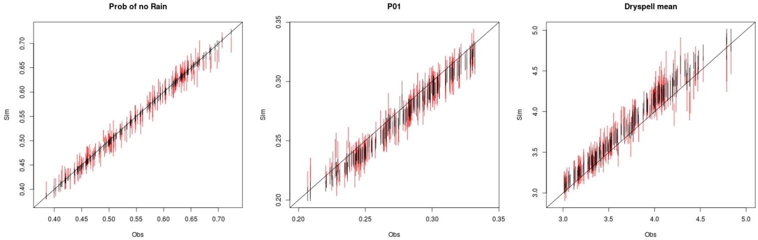


Figure: Probability of no rain, transition probability from dry to wet, mean length of dry spell

Model Validation - Marginals

Marginal occurrence process:

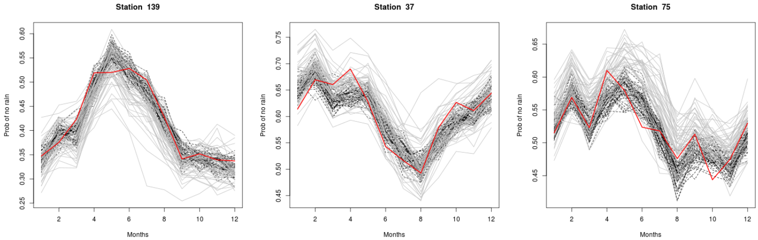


Figure: Probability of no rain for 3 stations, 95% confidence intervals. Grey withour resampling β dotted resampling β

Model Validation - Margianals

Marginal Intensity process - Daily Statistics:

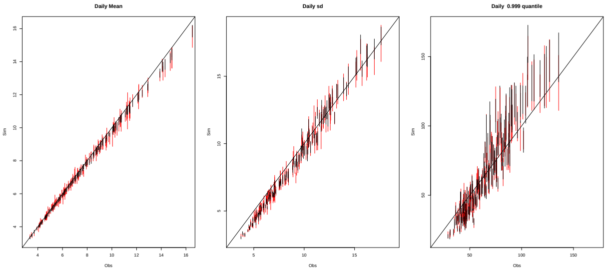
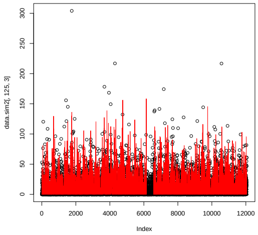


Figure: Daily intentinsity mean, standard deviation and .999 quantile

Model Validation - Marginals

Marginal Intensity process - Extreme Values:

- Unrealistic high simulated values
- This is a problem in winter and in the winter season (where the mean of the gamma distribution tends to be high)
- Need to consider other density tha gamma



Number of sumulataneously dry station

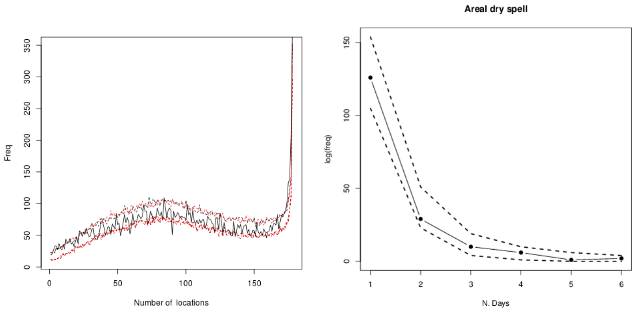
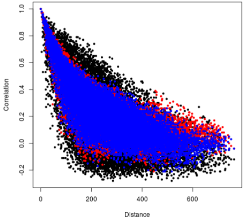
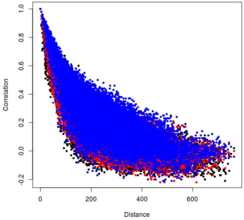


Figure: Number of contemporary dry stations and areal dry spell

Model Validation - Space

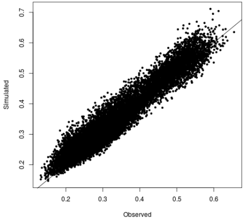
Correlation of wet stations in January and July

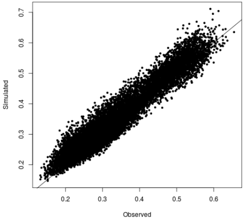




Model Validation - Space

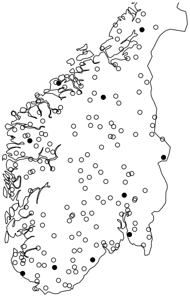
Probability of being contemporary dry in January and July



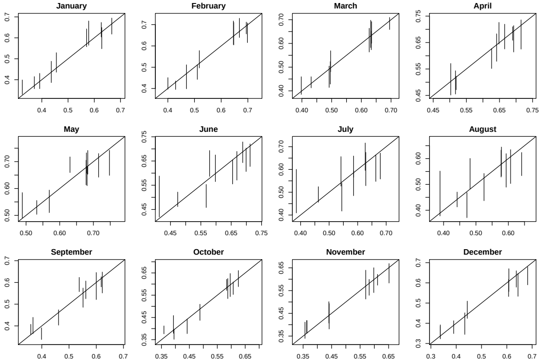


Crossvalidation

- Remove 10 random stations from the dataset
- Re-estimate all parameters
- Sample 20 time series



Crossvalidation - Probability of Dry State



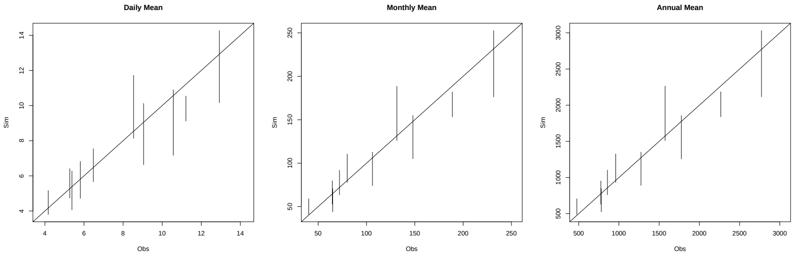


Figure: Daily, Monthly and Annual mean

Simulated Annual Map

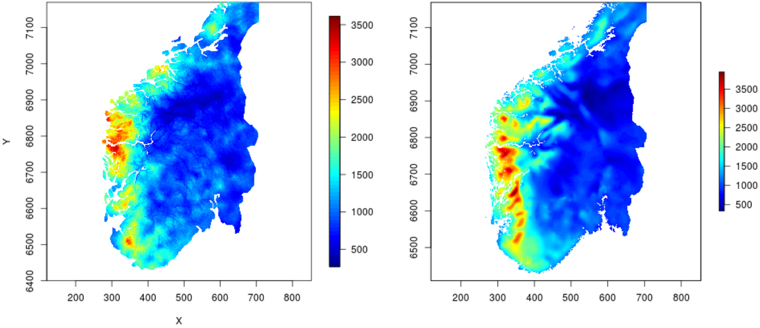


Figure: Simulated Annual Map (1980) Annual Map from Met.no (1979)