Daily stochastic precipitation generator based on Gaussian field: a comparison of two estimation approaches

Komlan A. KPOGO-NUWOKLO

Météo-France, DIROP/PI, Toulouse

Institut für Meteorologie, Freie Universität Berlin, Berlin

Rennes, November, 28th to 30th, 2018

Motivation

- SWGen are increasingly used as a complementary approach for precipitation simulation:
  - low costs and rapid simulations
  - uncertainty assessment
- Precipitation occurrences and intensities depend on atmospheric conditions

Motivation

- SWGen are increasingly used as a complementary approach for precipitation simulation:
  - low costs and rapid simulations
  - uncertainty assessment
- Precipitation occurrences and intensities depend on atmospheric conditions
- Gaussian field based model (occurrence+intensity) estimated from rain gauge network + atmospheric variables (used as covariables)
- Direct estimates of the model parameters are available only at locations with observed data

Motivation

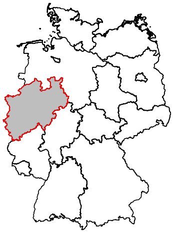
- SWGen are increasingly used as a complementary approach for precipitation simulation:
  - low costs and rapid simulations
  - uncertainty assessment
- Precipitation occurrences and intensities depend on atmospheric conditions
- Gaussian field based model (occurrence+intensity) estimated from rain gauge network + atmospheric variables (used as covariables)
- Direct estimates of the model parameters are available only at locations with observed data
- To generate a complete precipitation field across the entire simulation domain one could:
  - include the geographical coordinates as covariates in the model (TSA)
  - estimate the parameters at the location of stations and spatially interpolate them (Kriging)
- **Goal:** compare the two approaches using a model based on Gaussian field

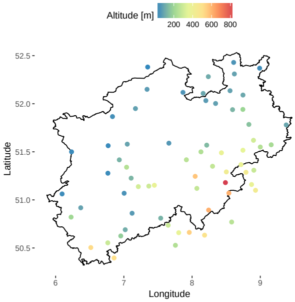
Outline

- Data
- Precipitation generator based on Gaussian field
- Parameters estimation
- Results
- Conclusion

Data

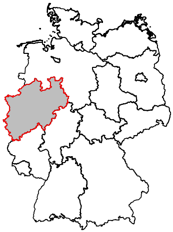
- Daily precipitation from 1980 to 2015 at 66 stations (North Rhine-Westphalia)
    - altitude varies from 28 to 840 m

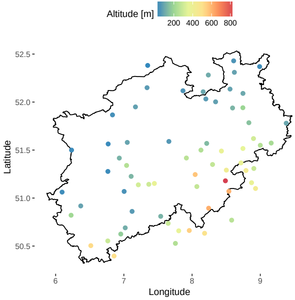




Data

- Daily precipitation from 1980 to 2015 at 66 stations (North Rhine-Westphalia)
  - altitude varies from 28 to 840 m





- Topography data from the Shuttle Radar Topography Mission (NASA)

- Covariates from ERA-Interim reanalysis
  - near surface variables: wind, dewpoint temperature
  - atmospheric variables (500hPa): geopotential, specific humidity, relative humidity, vertical velocity

Model

- Use a Gaussian field to model the occurrence
- Transform the same Gaussian field to obtain nonzero amounts
- Assume the precipitation field $Y(\mathbf{s}, t)$ depends on the latent Gaussian variable $Z(\mathbf{s}, t)$:

$$Y(\mathbf{s}, t) = \begin{cases} \psi\left(Z(\mathbf{s}, t)\right), & \text{if } Z(\mathbf{s}, t) > 0 \Longrightarrow \text{wet} \\ 0, & \text{if } Z(\mathbf{s}, t) \leq 0 \Longrightarrow \text{dry} \end{cases}$$

Model

- Use a Gaussian field to model the occurrence
- Transform the same Gaussian field to obtain nonzero amounts
- Assume the precipitation field $Y(\mathbf{s}, t)$ depends on the latent Gaussian variable $Z(\mathbf{s}, t)$:

$$Y(\mathbf{s}, t) = \begin{cases} \psi\left(Z(\mathbf{s}, t)\right), & \text{if } Z(\mathbf{s}, t) > 0 \Longrightarrow \text{wet} \\ 0, & \text{if } Z(\mathbf{s}, t) \leq 0 \Longrightarrow \text{dry} \end{cases}$$

- Model components:
  - Anamorphosis ψ
  - Mean function $\mu = \mathbb{E}\left[Z(\mathbf{s}, t)\right]$
  - Covariance structure $\text{Cov}\left(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\right)$

1- Anamorphosis

$$\psi(.) = G_{\mathbf{s},t}^{-1} \circ \Phi_\mu(.)$$

$G_{\mathbf{s},t} \hookrightarrow \text{Gamma}\left(\alpha(\mathbf{s},t)\,\gamma(\mathbf{s},t)\right)$
Φ_μ is a cdf of a censored normal random variable with mean μ and variance unity.

- if a day is wet i.e. $Z(\mathbf{s},t) > 0$,

$$Y(\mathbf{s},t) = G_{\mathbf{s},t}^{-1} \circ \Phi_\mu\left(Z(\mathbf{s},t)\right)$$

1- Anamorphosis

$$\psi(.) = G_{\mathbf{s},t}^{-1} \circ \Phi_{\mu}(.)$$

$G_{\mathbf{s},t} \hookrightarrow \text{Gamma}\left(\alpha(\mathbf{s},t)\,\gamma(\mathbf{s},t)\right)$
Φ_{μ} is a cdf of a censored normal random variable with mean μ and variance unity.

- if a day is wet i.e. $Z(\mathbf{s},t) > 0$,

$$Y(\mathbf{s},t) = G_{\mathbf{s},t}^{-1} \circ \Phi_{\mu}\left(Z(\mathbf{s},t)\right)$$

- Include non-stationnarity and effect of external covariates in the gamma distribution

$$\log(\alpha(\mathbf{s},t)) = \beta_{\alpha}(\mathbf{s})X(\mathbf{s},t)$$

$$\log(\gamma(\mathbf{s},t)) = \beta_{\gamma}(\mathbf{s})X(\mathbf{s},t)$$

- Covariates are:

seasons, wind, dewpoint temperature, mean sea level pressure, potential vorticity .

2- Mean function

- Link mean function of Z and the precipitation data Y through the probability of wet days:

$$p(\mathbf{s}, t) = P\left(Y(\mathbf{s}, t) > 0\right) = P\left(Z(\mathbf{s}, t) > 0\right) = \Phi\left(\mu(\mathbf{s}, t)\right)$$

- Then,

$$\hat{\mu}(\mathbf{s}, t) = \Phi^{-1}\left(\hat{p}_{\mathbf{s},t}\right)$$

2- Mean function

- Link mean function of Z and the precipitation data Y through the probability of wet days:

$$p(\mathbf{s}, t) = P\left(Y(\mathbf{s}, t) > 0\right) = P\left(Z(\mathbf{s}, t) > 0\right) = \Phi\left(\mu(\mathbf{s}, t)\right)$$

- Then,

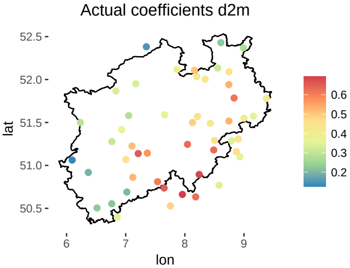
$$\hat{\mu}(\mathbf{s}, t) = \Phi^{-1}\left(\hat{p}_{\mathbf{s}, t}\right)$$

- Estimation of $p(\mathbf{s}, t)$ trough logistic regression (probit model)

$$\hat{p}(\mathbf{s}, t) = \Phi\left(\beta_p'(\mathbf{s}) X_p(\mathbf{s}, t)\right)$$

- Covariates $(X_p(\mathbf{s}, t))$ are:
seasons, wind, dewpoint temperature, geopotential, specific humidity, relative humidity, vertical velocity.

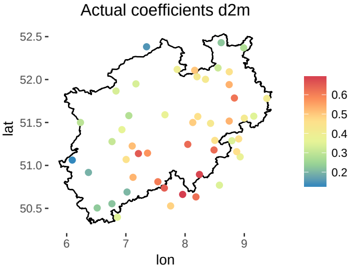
Parameters estimation



β estimation for $d2m$

- β are estimated at the station locations but we would also like to simulate over the all domain

Parameters estimation



β estimation for $d2m$

- β are estimated at the station locations but we would also like to simulate over the all domain
- Two approaches are used
  - **Kriging approach:** Estimation of β at the station locations and smoothing over space using kriging (Universal kriging)
  - **Trend Surface Analysis (TSA) approach:** Use of a polynomial expansion of geographical coordinates as covariates (Legendre polynomials)
- We aim to compare the two approaches

Kriging approach: universal kriging

- $\beta(s_1), \ldots, \beta(s_n)$ with β such as

$$E\left(\beta(s)\right) = az(s)$$

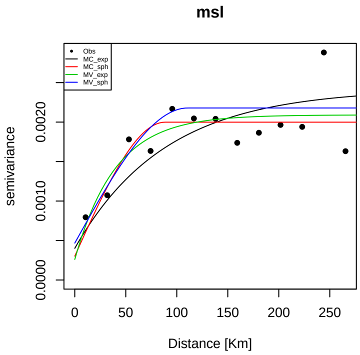
with $z(s)$ being the altitude of station s and a the regression coeficient.

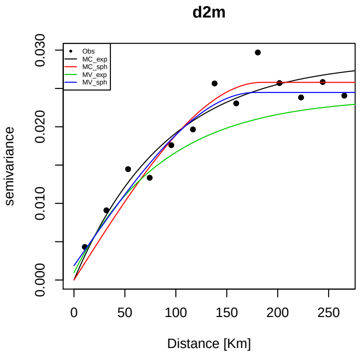
$$var\left(\beta(s) - \beta(s+h)\right) = 2\gamma(h).$$

We are looking for

$$\beta^*(s_0) = \sum \lambda_i \beta(s_i)$$

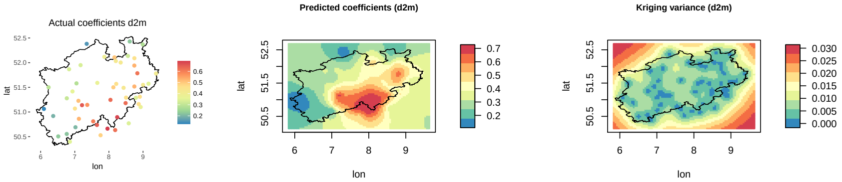
- Variogram



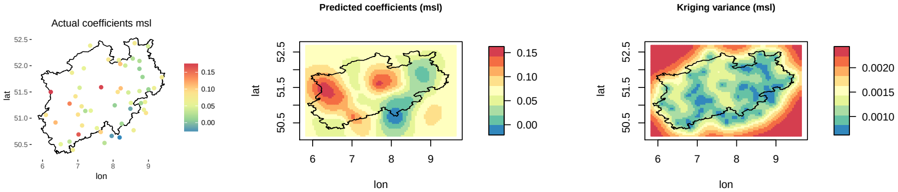


Kriging approach: universal kriging

- Parameter associated to d2m



- Parameter associated to msl



TSA: Legendre polynomials

- To capture spatial variations, we use a series expansion using Legendre polynomials for longitude x, latitude y and altitude z

- Example for the mean function

$$\hat{\mu}(s,t) = \mu_{climat} + \mu_{season} + \mu_{geo} + \mu_{int}$$

with

$$\mu_{climat} = \sum_{c=1}^{C} \beta_c X_c$$

X_c reprents a large scale atmospheric variable

$$\mu_{season} = \beta_0 + \sum_{i=1}^{I} \left[\beta_{i,sin} \sin\left(\frac{2i\pi}{365.25} t \right) + \beta_{i,cos} \cos\left(\frac{2i\pi}{365.25} t \right) \right]$$

$$\mu_{geo} = \sum_{j=1}^{J} \beta_{j,P} P_j(x) + \sum_{k=1}^{K} \beta_{k,P} P_k(y) + \sum_{l=1}^{L} \beta_{l,P} P_l(z)$$

where $P_j(.)$ is the j^{ieme} Legendre polynomial used for x. μ_{int} is the interaction terms

3- Covariance structure

- Stationary, isotropic covariance function

$$\text{Cov}\left(Z(s_i), Z(s_j)\right) = C\left(s_i, s_j\right) = C\left(\parallel h \parallel\right)$$

with $h = s_i - s_j$

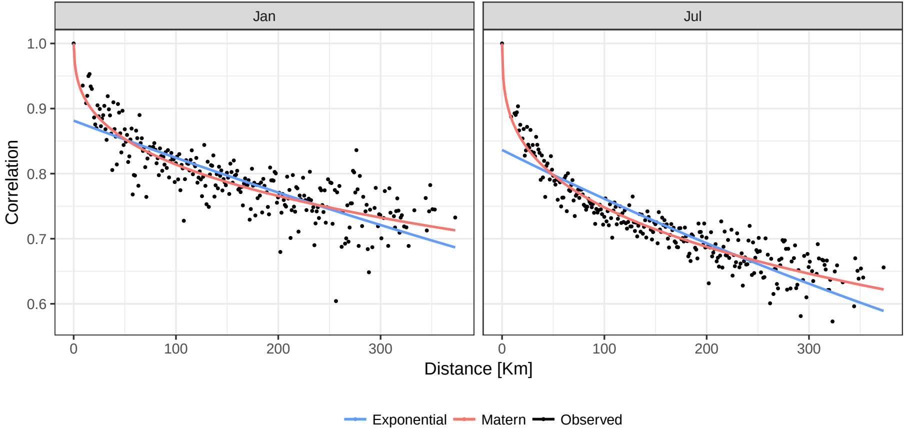
- Estimation of empirical covariance
- Fit a parametric covariance function
  - Exponential

$$C(h) = \exp\left(-\frac{\parallel h \parallel}{\rho}\right)$$

  - Matérn

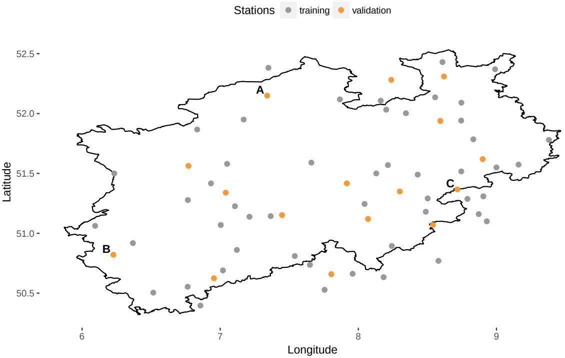
$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu} \parallel h \parallel}{\rho}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu} \parallel h \parallel}{\rho}\right)$$

3- Covariance structure

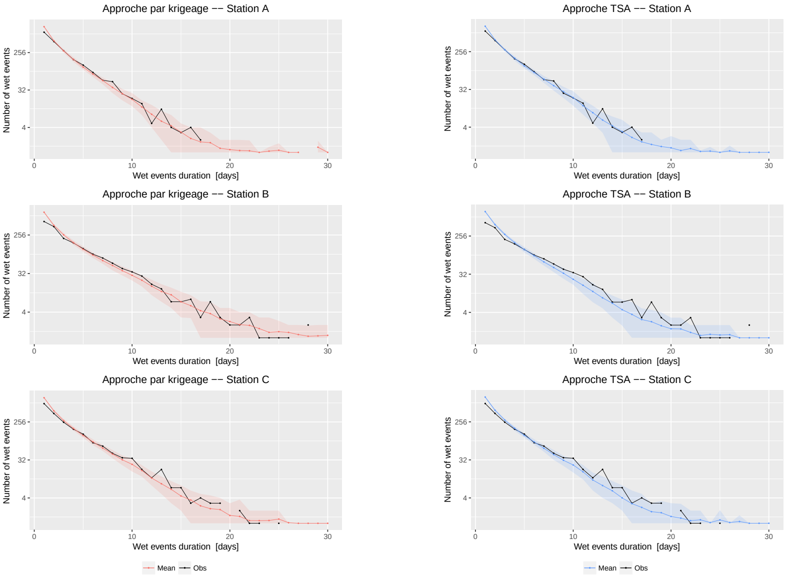


Calibration and validation stations

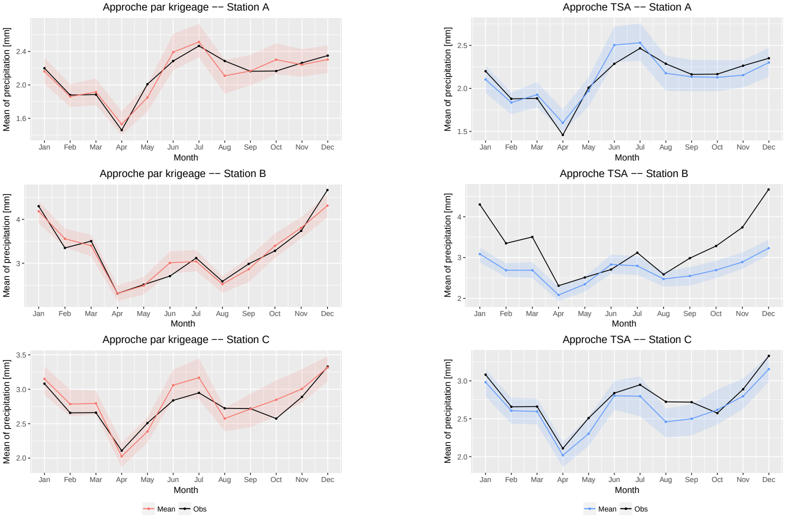
- 50 stations are used for calibration and 16 for validation
- Simulation of 100 time series from 1980 to 2015 for each approach (Kringing and TSA)



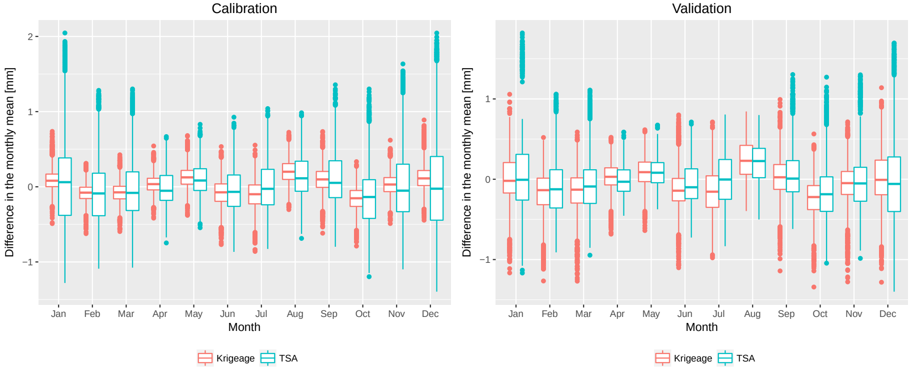
Wet spell length



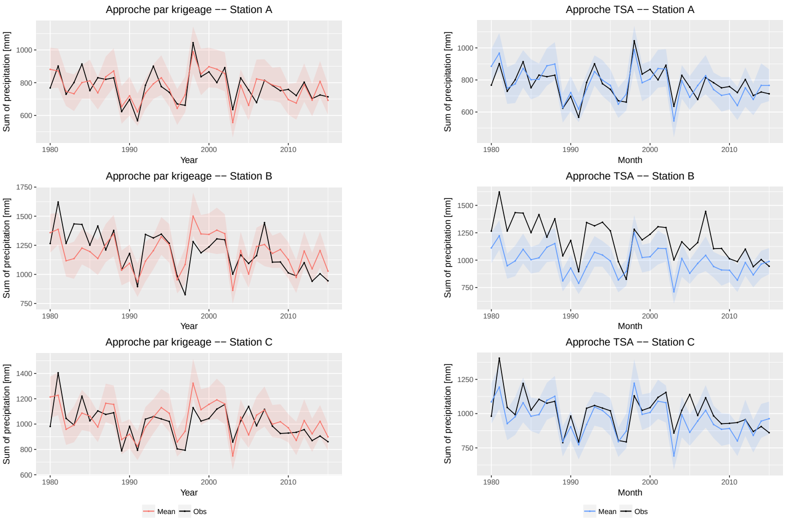
Seasonal mean



Seasonal mean error



Annual precipitation sum



Variogram winter

Conclusion

- Two approaches of parameters estimation for Gaussain field based precipitation generator are compared
- Wet days frequency and spell length are globally well reproduced by two approaches
- TSA tends to underestimate precipitation intensities for some points.

Thank you!