`GeoModels` Tutorial: simulation, estimation and prediction of positive spatial data using Weibull random fields

Moreno Bevilacqua
Christian Caamaño

Introduction

In this tutorial we show how to analyze geo-referenced spatial data with positive support using Weibull random fields (RFs) (Bevilacqua et al., 2018) with the `R` package `GeoModels` (Bevilacqua and Morales-Oñate (2018)). The Weibull distribution is a flexible parametric model for positive data allowing both right and left skewness.

We first load the R libraries needed for the analysis and set the name of the model in the `GeoModels` package:

```
rm(list=ls())
require(GeoModels)
require(fields)
require(hypergeo)
set.seed(24)
model="Weibull"  # model name in the  GeoModels package
```

Simulation of Weibull random fields

The definition of a Weibull RF starts by considering a 'parent' Gaussian RF $Z := \{Z(\boldsymbol{s}), \boldsymbol{s} \in S\}$, where $\boldsymbol{s}$ represents a location in the domain S. In this tutorial, we assume $S = [0,1]^2 \subseteq \mathbb{R}^2$ and that Z is stationary with zero mean, unit variance and correlation function $\rho(\boldsymbol{h}) := \mathrm{cor}(Z(\boldsymbol{s}+\boldsymbol{h}), Z(\boldsymbol{s}))$.

Given Z_1, Z_2, two independent copies of Z, a RF $U = \{U(\boldsymbol{s}), \boldsymbol{s} \in S\}$ with marginal distribution $Weibull(\kappa, \nu(\kappa))$ can be derived by the transformation

$$U(\boldsymbol{s}) = \nu(\kappa)\left(\frac{1}{2}\sum_{k=1}^{2} Z_k(\boldsymbol{s})^2\right)^{1/\kappa}, \tag{1}$$

where $\nu(\kappa) = \Gamma^{-1}(1+1/\kappa)$, $\kappa > 0$ is a shape parameter and $\Gamma(\cdot)$ is the gamma function. Under this specific parametrization, $\mathbb{E}(U(\boldsymbol{s})) = 1$ $\mathrm{var}(U(\boldsymbol{s})) = (\Gamma(1+2/\kappa)\,\nu^2(\kappa) - 1)$ and the correlation function is given by:

$$\rho_U(\boldsymbol{h}) = \frac{\nu^{-2}(\kappa)}{[\Gamma(1+2/\kappa) - \nu^{-2}(\kappa)]}\left[{}_2F_1\left(-1/\kappa, -1/\kappa; 1; \rho^2(\boldsymbol{h})\right) - 1\right]. \tag{2}$$

Here ${}_2F_1(a, b; c; x)$ is the Gaussian hypergeometric function (Abramowitz and Stegun (1970)). In the `GeoModels` package it is computed using the function `hypergeo` of the `hypergeo` package (Hankin (2016)).

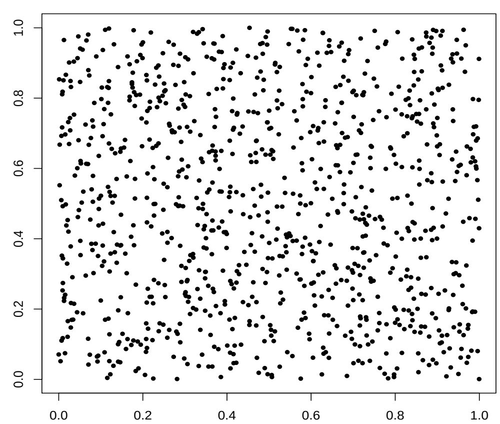
Then a non stationary version can be defined trough as:

$$W(\boldsymbol{s}) = \mu(\boldsymbol{s})U(\boldsymbol{s}), \qquad \mu(\boldsymbol{s}) > 0. \tag{3}$$

In this case $\mathbb{E}(W(\boldsymbol{s})) = \mu(\boldsymbol{s})$, $\text{var}(W(\boldsymbol{s})) = \mu(\boldsymbol{s})^2(\Gamma\left(1+2/\kappa\right)\nu^2(\kappa)-1)$ and a spatial regression model can be obtained by assuming that $\mu(\boldsymbol{s}) = e^{X(\boldsymbol{s})^T\boldsymbol{\beta}}$ where $X(\boldsymbol{s})$ is a k-dimensional vector of covariates and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^T$ is a k-dimensional vector of (unknown) parameters.

Thus, in order to obtain a realization from a Weibull RF we need to specify a regression mean parameters, a shape parameter and a parametric correlation model for $\rho(\boldsymbol{h})$. We first set the spatial coordinates

```
N=1000  # number of location sites
x = runif(N, 0, 1)
y = runif(N, 0, 1)
coords=cbind(x,y)  # spatial coordinates
plot(coords,pch=20,xlab="",ylab="")
```



Then we fix $k = 2$ and we build the matrix covariates and fix the regression mean parameters

```
X=cbind(rep(1,N),runif(N))      # matrix covariates
mean = -0.3; mean1=0.5          # regression parameters
```

where `mean` and `mean1` are respectively β_1 and β_2.

The names of the marginal parameters associated with the Weibull model can be obtained with the function NuisParam (note that the option num_betas is the number of regression parameters involved):

```
NuisParam(model,num_betas=2)
[1] "mean"   "mean1"   "nugget"   "shape"
```

For the correlation function we assume a special case of the isotropic Generalized Wendland class (Bevilacqua et al. (2019)) i.e the Askey model.

$$\rho(\boldsymbol{h};\alpha,\delta) := \begin{cases} (1-||\boldsymbol{h}||/\alpha)^{\delta} & ||\boldsymbol{h}|| < \alpha \\ 0 & \text{otherwise} \end{cases}.$$

Using asymptotic arguments Bevilacqua et al. (2019) show that this correlation model has the same features of the exponential correlation model. Additionally it is compactly supported an interesting feature from computational point of view. We set the Askey model and the associated parameters. Note that the function CorrParam returns the names of the parameters associated for a given correlattion model.

```
corrmodel = "Wend0"    ## correlation model
CorrParam(corrmodel)   ## names of correlation model parameter
[1] "power2" "scale"
scale = 0.2
power2 = 4
```

Here the scale parameter corresponds to α, the compact support of the correlation model. Finally, we set the shape parameter of the Weibull RF and the nugget parameter

```
shape=2    # shape of the  weibull RF
nugget=0   # nugget parameter
```

We are now ready to simulate a Weibull random field using the function GeoSim:

```
param=list(mean=mean,mean1=mean1, nugget=nugget, scale=scale ,power2=power2 ,shape=
set.seed(312)
data = GeoSim(coordx=coords , corrmodel=corrmodel , model=model , param=param ,
X=X)$data
```

The simulation is performed using Cholesky decomposition for the two Gaussian RFs involved (see Equation (1)).

Estimation of Weibull random fields

The density of the bivariate random vector $U(\boldsymbol{s}_i), U(\boldsymbol{s}_j)$ is given by (Bevilacqua et al. (2018)).

$$f_U(u_i,u_j) \;=\; \frac{\kappa^2 (u_i u_j)^{\kappa-1}}{\nu^{2\kappa}(\kappa)(1-\rho_{ij}^2)} \exp\left[-\frac{u_i^\kappa + u_j^\kappa}{\nu^\kappa(\kappa)(1-\rho_{ij}^2)}\right] I_0\left(\frac{2|\rho|(u_i u_j)^{\kappa/2}}{\nu^\kappa(\kappa)(1-\rho_{ij}^2)}\right). \tag{4}$$

where $I_\alpha(x)$ denotes the modified Bessel function of the first kind of order α and the bivariate densities of W can be derived from (4) as

$$f_W(w_i,w_j) \;=\; (\mu_i\mu_j)^{-1} f_U(w_i/\mu_i, w_j/\mu_j). \tag{5}$$

Given $w(\mathbf{s}_1),\ldots,w(\mathbf{s}_n)$, $i=1,\ldots,n$ observations, then, the pairwise likelihood function is defined as:

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{N}\sum_{j\neq i}^{N} log(f_W(w_i,w_j))c_{ij}$$

. where $\boldsymbol{\theta} = (\beta_0,\beta_1,\kappa,\alpha,\delta)^T$, $w_i = w(\mathbf{s}_i)$ for notation convenience

An efficient way to specify the (non symmetric) weights from computational and efficient viewpoint is based on neighborhoods:

$$c_{ij}(k) = \begin{cases} 1 & \mathbf{s}_i \in N_k(\mathbf{s}_j) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $N_k(\mathbf{s}_l)$ is the set of the neighbors of order $k = 1,2,\ldots$ of the point $\mathbf{s}_l$.

The pairwise likelihood estimator $\hat{\boldsymbol{\theta}}_{pl}$ is obtained maximizing (5) with respect to $\boldsymbol{\theta}$. In the `GeoModels` package we can choose the fixed parameters and the parameters that must be estimated. Pairwise likelihood estimation is performed with the function `GeoFit`:

```
start=list(mean=mean,mean1=mean1,scale=scale,shape=shape)
fixed=list(,nugget=nugget,power2=power2)
# Maximum composite-likelihood fitting of the Weibull random field:
fit = GeoFit(data=data,coordx=coords, corrmodel=corrmodel,model=model,
    X=X, optimizer="BFGS",start=start,fixed=fixed,neighb=3)
```

The object `fit` include informations about the pairwise likelihood estimation

```
fit


#################################################################
```

```
Maximum  Composite-Likelihood Fitting of Weibull Random Fields
Setting: Marginal Composite-Likelihood
Model: Weibull
Type of the likelihood objects: Pairwise
Covariance model: Wend0
Optimizer: BFGS
Number of spatial coordinates: 1000
Number of dependent temporal realisations: 1
Type of the random field: univariate
Number of estimated parameters: 4
Type of convergence: Successful
Maximum log-Composite-Likelihood value: -3693.89
Estimated parameters:
   mean     mean1     scale     shape
-0.2418    0.3988    0.2267    1.9590
##################################################################
```

Note that the option `neighb=3` set the compact support of the weight function (6) i.e. $k = 3$.

Checking model assumptions

Given the estimation of the mean $\widehat{\mu(\boldsymbol{s})} = e^{X_1(\boldsymbol{s})\hat{\beta}_1 + X_2(\boldsymbol{s})\hat{\beta}_2}$, the estimated residuals

$$\widehat{u(\boldsymbol{s}_i)} = w(\boldsymbol{s}_i)/\widehat{\mu(\boldsymbol{s}_i)} \qquad i = 1, \ldots, N \tag{7}$$

can be viewed as a realization of U a stationary RF with marginal distribution $Weibull(\kappa, \nu(\kappa))$ with unit mean and correlation function given by (2).

The estimated residuals can be computed using the `GeoResiduals` function:

```
res=GeoResiduals(fit)  # computing residuals
```

Then the agreement of the marginal distribution assumption on the residuals with the theoretical model can be graphically checked with the `GeoQQ` function (Figure 1 left part):

```
GeoQQ(res)
```

The covariance model assumption can be checked comparing the empirical and the estimated semivariogram using the `GeoVariogram` and `GeoCovariogram` functions (Figure 1 right part). In particular the function `GeoVariogram` compute the empirical semivariogram:

```
### checking model  residuals assumptions: covariance model
vario = GeoVariogram(data=res$data,
coordx=coords,maxdist=0.3) # empirical variogram
GeoCovariogram(res,show.vario=TRUE, vario=vario,pch=20)
```

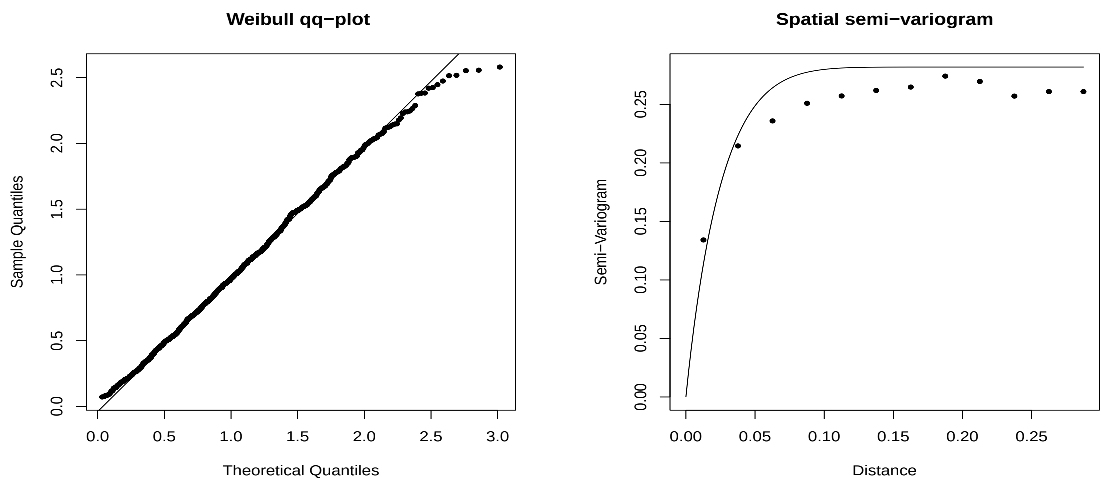


Figure 1: Left: QQ-plot for the Weibull model residuals. Right: empirical vs estimated semi-variogram function for the residuals

Prediction of Weibull random fields

The optimal linear prediction of Weibull RF at a location $\boldsymbol{s}_0$ is given by (Bevilacqua et al. (2018)):

$$\widehat{W(\boldsymbol{s}_0)} = \widehat{\mu(\boldsymbol{s}_0)}\left(1 + \sum_{i=1}^{N} \lambda_i[\widehat{U(\boldsymbol{s}_i)} - 1]\right) \tag{8}$$

where the vector of weights $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)'$ is given by $\boldsymbol{\lambda} = R^{-1}\boldsymbol{c}$.

Here $\boldsymbol{c} = (cor(U(\boldsymbol{s}_0), U(\boldsymbol{s}_1)), \dots, cor(U(\boldsymbol{s}_0), U(\boldsymbol{s}_n)))'$ and $R = [\mathrm{cor}(U(\boldsymbol{s}_i), U(\boldsymbol{s}_j)]_{i,j=1}^{N}$ is the (estimated) correlation matrix associated to (2).

We first set the spatial locations to predict and the associated covariates. In this example, we choose a regular fine grid in order to construct a prediction map.

```
# locations to predict  and associated covariates
xx=seq(0,1,0.013)
loc_to_pred=as.matrix(expand.grid(xx,xx))
```

```
Nloc=nrow(loc_to_pred)
Xloc=cbind(rep(1,Nloc),runif(Nloc))
```

Then the optimal linear prediction (8), using the estimated parameters, can be performed using the `GeoKrig` function:

```
pr=GeoKrig(fit, loc=loc_to_pred, Xloc=Xloc,mse=TRUE,sparse=TRUE)
```

Note that the option `sparse=TRUE` allows to exploit specific algorithms for sparse matrices implemented in the `spam` package (Gerber et al. (2017)) when computing the inverse of the variance covariance matrix. (Furrer and Sain (2010)).

We can compare the map of simulated data with the kriging prediction (and associated mean square error) with the following code (see Figure 2):

```
colour = rainbow (100)
par(mfrow=c(1,3))
quilt.plot(coords, data,col=colour,main="Data")
quilt.plot(loc_to_pred, pr$pred,col=colour,xlab="",ylab="",main="Kriging")
quilt.plot(loc_to_pred, pr$mse,col=colour,xlab="",ylab="",main="MSE")
```

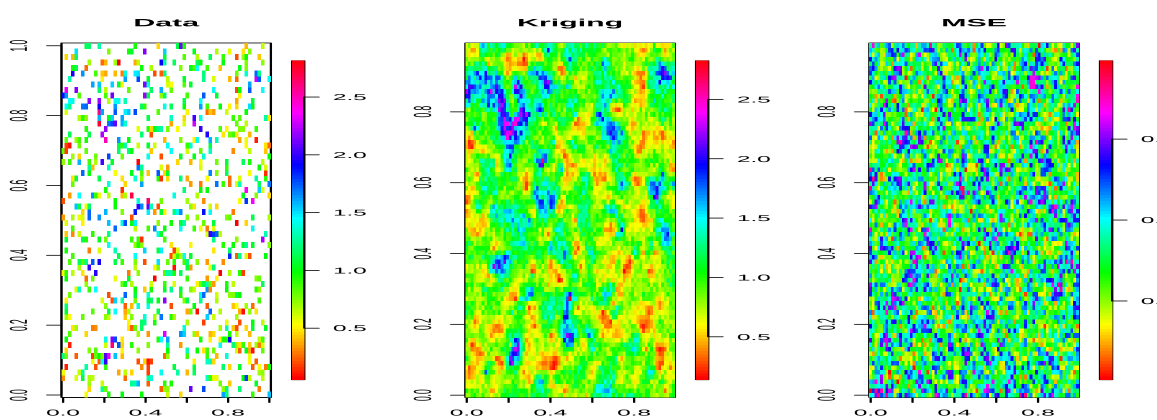


Figure 2: From left to right: colored map of observed data, kriging prediction and associated mean squared error

References

Abramowitz, M. and I. A. Stegun (1970). *Handbook of Mathematical Functions*. New York: Dover.

Bevilacqua, M., C. Caamano, and C. Gaetan (2018). On modelling positive continuous data with spatio-temporal dependence. *ArXiv e-prints*.

Bevilacqua, M., T. Faouzi, R. Furrer, and E. Porcu (2019). Estimation and prediction using generalized Wendland functions under fixed domain asymptotics. *The Annals of Statistics 47*, 828–856.

Bevilacqua, M. and V. Morales-Oñate (2018). *GeoModels: A Package for Geostatistical Gaussian and non Gaussian Data Analysis*. R package version 1.0.3-4.

Furrer, R. and S. R. Sain (2010). spam: a sparse matrix R package with emphasis on mcmc methods for Gaussian Markov random rields. *Journal of Statistical Software 36*, 1–25.

Gerber, F., K. Moesinger, and R. Furrer (2017). Extending R packages to support 64-bit compiled code: An illustration with spam64 and GIMMS NDVI3g data. *Computer & Geoscience 104*, 109–119.

Hankin, R. K. S. (2016). *hypergeo: The Gauss Hypergeometric Function*. R package version 1.2-13.