Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model

Alexandre Wadoux[1] Dick Brus[2] Miguel Rico-Ramirez[3]
Gerard Heuvelink[1]

[1] Environmental sciences, Soil Geography and Landscape group
University of Wageningen, Netherlands

[2] Environmental Sciences, Soil, Water and Landuse group
Alterra, Netherlands

[3] Civil engineering, Water and Environment Management group
University of Bristol, United Kingdom

October 10, 2016

Introduction

Conventional geostatistical models assume that the property being monitored is the realisation of a second-order stationary random process

$$Z(s) = \mu + \varepsilon(s)$$

$\mu = constant$

$Cov(\varepsilon(s), \varepsilon(s+h)) = C(h)$

if $h = 0 => Cov(\varepsilon(s), \varepsilon(s)) = Var(\varepsilon(s)) = C(0)$

Conventional geostatistical models assume that the property being monitored is the realisation of a second-order stationary random process

$$Z(s) = \mu + \varepsilon(s)$$

$\mu = constant$

$Cov(\varepsilon(s), \varepsilon(s+h)) = C(h)$

if $h = 0 => Cov(\varepsilon(s), \varepsilon(s)) = Var(\varepsilon(s)) = C(0)$

But this is often an invalid assumption
$=>$ can be checked with exploratory analysis of the observed data

Expectation

Objectives...

1. Account for non-stationarity in the mean and variance of rainfall
2. Optimize the sampling locations of rain gauges for mapping rainfall over time

Introduction

Simple solutions exist for non-stationarity

In the mean

$$Z(s) = m(s) + \varepsilon(s)$$

and in the variance

$$Z(s) = m(s) + \sigma(s) \cdot \varepsilon(s)$$

- Mean rainfall at location s

$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$

Nonstationary variance model

- Mean rainfall at location s

$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$

- Multiplier for error at location s

Nonstationary variance model

- Mean rainfall at location s

$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$

- Multiplier for error at location s
- Standardized random error

In matrix notation

$$\mathbf{Z} = \mathbf{F}\beta + \underbrace{\mathbf{G}\boldsymbol{\kappa} \cdot \varepsilon}$$

$\mathbf{C} = diag\{\mathbf{G}\kappa\} \cdot \mathbf{R} \cdot diag\{\mathbf{G}\kappa\}^T$ is the variance-covariance matrix

In matrix notation

$$\mathbf{Z} = \mathbf{F}\beta + \underbrace{\mathbf{G}\boldsymbol{\kappa} \cdot \varepsilon}$$

$\mathbf{C} = diag\{\mathbf{G}\kappa\} \cdot \mathbf{R} \cdot diag\{\mathbf{G}\kappa\}^T$ is the variance-covariance matrix

Predictions at new location

$$\hat{\mathbf{z}}(\boldsymbol{s}_0) = \mathbf{f}(\boldsymbol{s}_0)^T\hat{\beta} + \mathbf{g}(\boldsymbol{s}_0)^T\hat{\boldsymbol{\kappa}} \cdot \hat{\varepsilon}(\boldsymbol{s}_0)$$

In matrix notation

$$\mathbf{Z} = \mathbf{F}\beta + \underbrace{\mathbf{G}\boldsymbol{\kappa} \cdot \varepsilon}$$

$\mathbf{C} = diag\{\mathbf{G}\kappa\} \cdot \mathbf{R} \cdot diag\{\mathbf{G}\kappa\}^T$ is the variance-covariance matrix

Predictions at new location

$$\hat{\mathbf{z}}(\boldsymbol{s}_0) = \mathbf{f}(\boldsymbol{s}_0)^T \hat{\beta} + \mathbf{g}(\boldsymbol{s}_0)^T \hat{\boldsymbol{\kappa}} \cdot \hat{\varepsilon}(\boldsymbol{s}_0)$$

Prediction error variance at new location

$$\sigma^2(\boldsymbol{s}_0) = \underbrace{\boldsymbol{c}(0) - \boldsymbol{c}_0^T \mathbf{C}^{-1} \boldsymbol{c}_0}_{\text{prediction error variance of the residuals}}$$

$$+ \underbrace{(\boldsymbol{f}_0 - \mathbf{F}^T \mathbf{C}^{-1} \boldsymbol{c}_0)^T (\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F})^{-1} \boldsymbol{f}_0 - \mathbf{F}^T \mathbf{C}^{-1} \boldsymbol{c}_0)}_{\text{error variance of the trend}}$$

Parameter estimation

With exponential correlogram,

$$r(h) = c_0 + (1 - c_0)\{exp(\frac{-3h}{a})\}$$

Parameter estimation

With exponential correlogram,

$$r(h) = c_0 + (1 - c_0)\{exp(\frac{-3h}{a})\}$$

We need to estimate $\Phi = [\kappa_i, c_0, a]$, and β_i

Parameter estimation

With exponential correlogram,

$$r(h) = c_0 + (1 - c_0)\{exp(\frac{-3h}{a})\}$$

We need to estimate $\Phi = [\kappa_i, c_0, a]$, and β_i
Independant of β_i, Restricted loglikelihood:

$$\ell(\Phi|\mathbf{z}) = Constant - \frac{1}{2}ln|\mathbf{C}| - \frac{1}{2}ln|\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}(\mathbf{I} - \mathbf{Q})\mathbf{z}$$

Parameter estimation

With exponential correlogram,

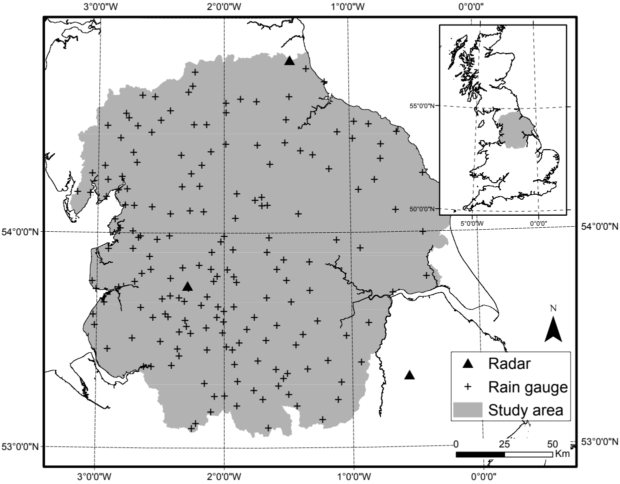
$$r(h) = c_0 + (1 - c_0)\{exp(\frac{-3h}{a})\}$$

We need to estimate $\Phi = [\kappa_i, c_0, a]$, and β_i
Independant of β_i, Restricted loglikelihood:

$$\ell(\Phi|\mathbf{z}) = Constant - \frac{1}{2} ln|\mathbf{C}| - \frac{1}{2} ln|\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}(\mathbf{I} - \mathbf{Q})\mathbf{z}$$

β_i are estimated with GLS using REML estimates of kappa, c_0 and a.

Illustration with a simple case, daily rainfall mapping with radar and rain-gauge

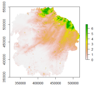


Covariates

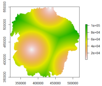
$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$

Covariates

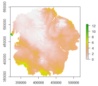
$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$



Rainfall from radar



Distance from radar

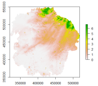


Previous predicted rainfall

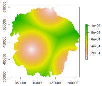
f_k

Covariates

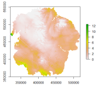
$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$



Rainfall from radar

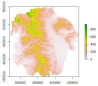


Distance from radar

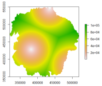


Previous predicted rainfall

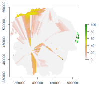
f_k +



Elevation



Distance from radar

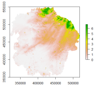


beam blockage

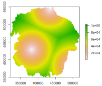
g_l ·

Covariates

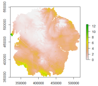
$$Z(s) = \sum_{k=0}^{K} \beta_k f_k(s) + \sum_{l=0}^{L} \kappa_l g_l(s) \cdot \varepsilon(s)$$



Rainfall from radar

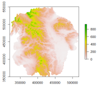


Distance from radar

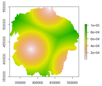


Previous predicted rainfall

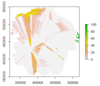
f_k +



Elevation

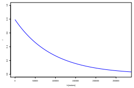


Distance from radar



beam blockage

g_l ·

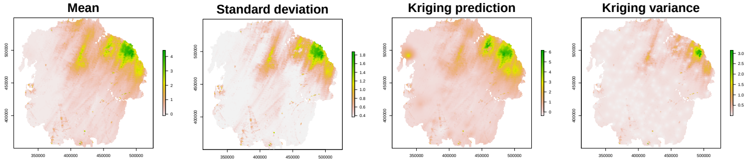


Correlogram

Model calibration

Example, February 11th, 2010...

Parameter	Estimated value	Associated to
c_0	0.0001278	*nugget*
a_1	8914	*range [meters]*
β_1	-0.02205	*intercept*
β_2	-0.1141	*radar image*
β_3	1.967e-05	*distance from radar*radar image*
β_4	0.1771	*previous estimated rainfall map*
κ_1	0.3699	*intercept*
κ_2	4.555e-11	*elevation*radar image*
κ_3	6.445e-06	*distance from radar*radar image*
κ_4	1.35e-10	*beam blockage*radar image*

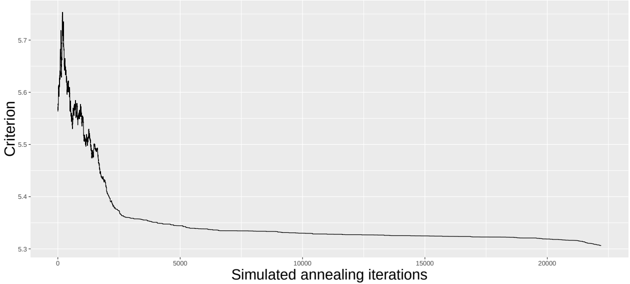


Minimizing the variance criterion by a random search called Spatial Simulated Annealing (SSA)

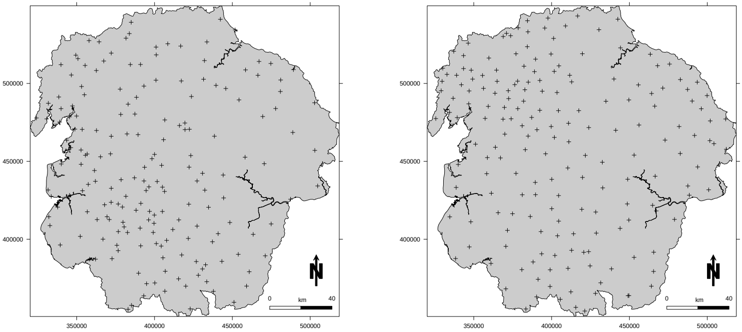
$$Criterion = \frac{1}{T}\frac{1}{|A|}\int_{t=0}^{T}\int_{s\in A} Var(Z(s,t) - \hat{Z}(s,t))dsdt \tag{1}$$

Minimizing the variance criterion by a random search called Spatial Simulated Annealing (SSA)

$$Criterion = \frac{1}{T}\frac{1}{|A|}\int_{t=0}^{T}\int_{s\in A} Var(Z(s,t) - \hat{Z}(s,t))dsdt \tag{1}$$



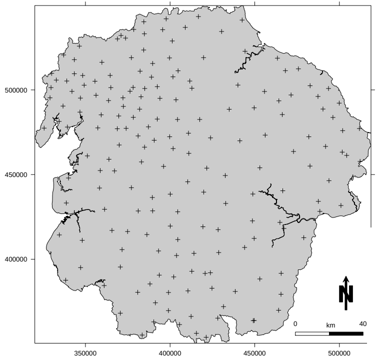
Sampling design optimisation

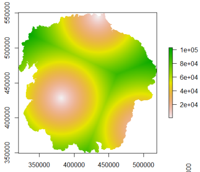


Initial

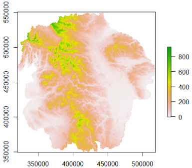
Optimized

Sampling design optimisation

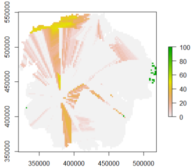




Distance from radar



Elevation



beam blockage

Final remarks

Decrease of the rainfall prediction error variance is obtained by the optimized rain-gauge network

1. It pays off to place rain-gauges at locations where the radar imagery is inaccurate
2. Uniform distribution of rain-gauge over the study area is also important

Interesting for:

- Fast radar-gauge merging accounting for the radar uncertainty (and soon the rain-gauge uncertainty too)
- The optimisation method could be applied to specific targets (flood forecasting)

Thank you for your attention

This project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.