

Model-based Geostatistics

Hannes Kazianka*

Vienna University of Technology

Jürgen Pilz†

Alpen-Adria University of Klagenfurt

Keywords and Phrases: Spatial Statistics, Geostatistics, Bayesian Kriging, Transformed-Gaussian Kriging, Correlation Function, Geometric Anisotropy, Random Field, Generalized Linear Geostatistical Model

1 Stochastic models for spatial data

Diggle and Ribeiro (2007) and Mase (2011) describe geostatistics as a branch of spatial statistics that deals with statistical methods for the analysis of spatially referenced data with the following properties. Firstly, values Y_i , $i = 1, \dots, n$, are observed at a discrete set of sampling locations \mathbf{x}_i within some spatial region $\mathcal{S} \subset \mathbb{R}^d$, $d \geq 1$. Secondly, each observed value Y_i is either a measurement of, or is statistically related to, the value of an underlying continuous spatial phenomenon, $Z(\mathbf{x})$, at the corresponding sampling location \mathbf{x}_i . The term model-based geostatistics refers to geostatistical methods that rely on a stochastic model. The observed phenomenon is viewed as a realization of a stochastic process with continuous spatial index, a so-called random field.

Such a random field $Z(\mathbf{x})$ is fully determined by specifying all multivariate distributions, i.e. $P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_n) \leq z_n)$ for arbitrary $n \in \mathbb{N}$ and $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{S}$. Since a full characterization of a random field is usually impossible, the mean function $m(\mathbf{x}) = E(Z(\mathbf{x}))$ and the covariance function $K(\mathbf{x}_i, \mathbf{x}_j) = Cov(Z(\mathbf{x}_i), Z(\mathbf{x}_j))$ play a prominent role. Thereby, $m(\mathbf{x})$ represents the trend while $K(\mathbf{x}_i, \mathbf{x}_j)$ defines the dependence structure of the random field. It is typical that the assumption of weak (second-order) isotropy is made about the random field, i.e. its mean function is constant and its covariance function $K(\mathbf{x}_1, \mathbf{x}_2)$ depends on \mathbf{x}_1 and \mathbf{x}_2 only through $h = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$, where $\|\cdot\|_2$ denotes the Euclidean distance. In this case K is called an isotropic autocovariance function. The covariance function is directly related to smoothness properties of the random field such as mean square continuity and differentia-

*Vienna University of Technology, Wiedner Hauptstrasse 8 /105-1, 1040 Vienna, Austria. Email: hannes.kazianka@fam.tuwien.ac.at

†Alpen-Adria University of Klagenfurt, Universitaetsstrasse 65-67, 9020 Klagenfurt, Austria. Email: juergen.pilz@uni-klu.ac.at

bility. A widely used parametric family of isotropic autocovariance functions is the Matèrn family

$$K_{\sigma^2, \boldsymbol{\theta}}(h) = \sigma^2 \left((1 - \vartheta_2) + \frac{\vartheta_2}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{2\kappa^{\frac{1}{2}} h}{\vartheta_1} \right)^\kappa \mathcal{K}_\kappa \left(\frac{2\kappa^{\frac{1}{2}} h}{\vartheta_1} \right) \right),$$

where \mathcal{K}_κ denotes the modified Bessel function of order $\kappa > 0$, $\vartheta_1 > 0$ is called the “range parameter” controlling how fast the covariance decays as the distance h gets large, $\vartheta_2 \in [0, 1]$ is called the “nugget parameter” and is the proportion of the variance attributed to measurement error, σ^2 controls the variance and $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2, \kappa)$ denotes the vector of correlation parameters. The parameter κ controls the smoothness of the corresponding process. A thorough mathematical introduction to the theory of random fields is given in Stein (1999) and Yaglom (1987).

The most important geostatistical model is the linear Gaussian model

$$Y_i = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta} + Z(\mathbf{x}_i), i = 1, \dots, n, \quad (1)$$

where $Z(\mathbf{x})$ is an isotropic zero-mean Gaussian random field with autocovariance function $K_{\sigma^2, \boldsymbol{\theta}}$, \mathbf{f} is a vector of location-dependent explanatory variables and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the vector of regression parameters. The likelihood function for the linear Gaussian model is

$$p(\mathbf{Y} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} |\sigma^2 \boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{F}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_\theta^{-1} (\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}) \right\},$$

where $\boldsymbol{\Sigma}_\theta$ denotes the correlation matrix, $\mathbf{F} = (\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_n))^T$ is the design matrix and $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ is the vector of observations. The maximum likelihood estimates for $\boldsymbol{\beta}$ and σ^2 in the linear Gaussian model are

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^T \boldsymbol{\Sigma}_\theta^{-1} \mathbf{F})^{-1} \mathbf{F}^T \boldsymbol{\Sigma}_\theta^{-1} \mathbf{Y}, \quad (2)$$

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Z} - \mathbf{F}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_\theta^{-1} (\mathbf{Z} - \mathbf{F}\hat{\boldsymbol{\beta}}). \quad (3)$$

Plugging these estimates into the log-likelihood, we arrive at the so-called profiled log-likelihood, which just contains the parameters $\boldsymbol{\theta}$

$$\log p(\mathbf{Y} | \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \boldsymbol{\theta}) = -\frac{n}{2} (\log(2\pi) + 1) - \frac{1}{2} \log |\boldsymbol{\Sigma}_\theta| - \frac{n}{2} \log(\hat{\sigma}^2).$$

To obtain $\hat{\boldsymbol{\theta}}$ we have to maximize the latter equation for $\boldsymbol{\theta}$ numerically. Note that this maximization problem is a lot simpler than the maximization of the complete likelihood where $\boldsymbol{\beta}$ and σ^2 are additional unknowns, especially when p is large. Spatial prediction, which is often the goal in geostatistics, is performed based on the estimated parameters. The plug-in predictive distribution for the value of the random field at an unobserved location \mathbf{x}_0 is Gaussian

$$Y_0 | \mathbf{Y}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}} \sim \mathcal{N} \left(\mathbf{k}^T \mathbf{K}^{-1} \mathbf{Y} + \mathbf{s}^T \hat{\boldsymbol{\beta}}, \hat{\sigma}^2 - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} + \hat{\sigma}^2 \mathbf{s}^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{s} \right), \quad (4)$$

where $\mathbf{K} = \hat{\sigma}^2 \boldsymbol{\Sigma}_{\hat{\theta}}$, $\mathbf{s} = \mathbf{f}(\mathbf{x}_0) - \mathbf{F}^T \mathbf{K}^{-1} \mathbf{k}$, $\mathbf{k} = \text{Cov}(\mathbf{Z}, Z(\mathbf{x}_0))$, $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))^T$.

Weak isotropy is a rather strong assumption and environmental processes are typically not direction independent but show an anisotropic behavior. A popular extension to isotropic random fields is to consider random fields that become isotropic after a linear transformation of the coordinates (Schabenberger and Gotway (2005)). This special variant of anisotropy is called geometric anisotropy. Let $Z_1(\mathbf{x})$ be an isotropic random field on \mathbb{R}^d with autocovariance function K_1 and mean μ . For the random field $Z(\mathbf{x}) = Z_1(\mathbf{T}\mathbf{x})$, where $\mathbf{T} \in \mathbb{R}^{d \times d}$, we get that $E(Z(\mathbf{x})) = \mu$ and the corresponding autocovariance function is $\text{Cov}(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) = K_1(\|\mathbf{T}(\mathbf{x}_1 - \mathbf{x}_2)\|_2)$. When correcting for geometric anisotropy we need to revert the coordinate transformation. $Z(\mathbf{T}^{-1}\mathbf{x})$ has the same mean as $Z(\mathbf{x})$ but isotropic autocovariance function K_1 . When correcting for stretching and rotation of the coordinates we have

$$\mathbf{T}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}.$$

Here, λ and φ are called the anisotropy ratio and anisotropy angle, respectively. All the models that we consider in this chapter can be extended to account for geometric anisotropy by introducing these two parameters.

2 Bayesian kriging

The first steps towards Bayesian modeling and prediction in geostatistics were made by Kitanidis (1986) and Omre (1987) who developed a Bayesian version of universal kriging. One of the advantages of the Bayesian approach, besides its ability to deal with the uncertainty about the model parameters, is the possibility to work with only a few measurements. A disadvantage of Bayesian kriging over the non-Bayesian methodology is certainly the increased computational complexity. Assume a Gaussian random field model in the form of the form Eq. (1) with known covariance matrix \mathbf{K} but unknown parameter vector $\boldsymbol{\beta}$. From Bayesian analysis we know that it is natural to assume a prior of the form $\boldsymbol{\beta} \sim \text{N}(\mathbf{m}_b, \sigma^2 \mathbf{V}_b)$ for $\boldsymbol{\beta}$, where \mathbf{V}_b is a positive semidefinite matrix. It can be shown that the posterior distribution for $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} | \mathbf{Z} \sim \text{N}(\tilde{\boldsymbol{\beta}}, \sigma^2 \mathbf{V}_{\tilde{\boldsymbol{\beta}}}),$$

where $\tilde{\boldsymbol{\beta}} = \mathbf{V}_{\tilde{\boldsymbol{\beta}}} (\sigma^2 \mathbf{F}^T \mathbf{K}^{-1} \mathbf{Z} + \mathbf{V}_b^{-1} \mathbf{m}_b)$ and $\mathbf{V}_{\tilde{\boldsymbol{\beta}}} = (\sigma^2 \mathbf{F}^T \mathbf{K}^{-1} \mathbf{F} + \mathbf{V}_b^{-1})^{-1}$. The predictive distribution of $Z(\mathbf{x}_0)$ is also Gaussian and given by

$$Z(\mathbf{x}_0) | \mathbf{Z} \sim \text{N}(\mathbf{k}^T \mathbf{K}^{-1} \mathbf{Z} + \mathbf{s}^T \tilde{\boldsymbol{\beta}}, \sigma^2 - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} + \sigma^2 \mathbf{s}^T \mathbf{V}_{\tilde{\boldsymbol{\beta}}} \mathbf{s}),$$

where \mathbf{F} , \mathbf{s} and \mathbf{k} are defined as in Section 1. From the above representation of the Bayesian kriging predictor it becomes clear that Bayesian kriging bridges the gap between simple and universal kriging. We get simple kriging

in case of complete knowledge of the trend, which corresponds to $\mathbf{V}_b = \mathbf{0}$, whereas we get the universal kriging predictor if we have no knowledge of $\boldsymbol{\beta}$ ($\mathbf{V}_b^{-1} = \mathbf{0}$ in the sense that the smallest eigenvalue of \mathbf{V}_b converges to infinity). Interestingly, the Bayesian universal kriging predictor has a smaller or equal variance than the classical universal kriging predictor (see Eq. (4)) since $(\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F} + \sigma^{-2} \mathbf{V}_b^{-1})^{-1} \preceq (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1}$, where \preceq denotes the Loewner partial ordering.

Bayesian universal kriging is not fully Bayesian because \mathbf{K} is assumed known. Diggle and Ribeiro (2007) summarize the results for a fully Bayesian analysis of Gaussian random field models of the form Eq. (1), where $K_{\sigma^2, \theta} = \sigma^2 \Sigma_{\vartheta_1}$ and ϑ_1 is the range parameter of an isotropic autocorrelation function model.

3 Transformed Gaussian kriging

Probably the simplest way to extend the Gaussian random field model is to assume that a differentiable transformation of the original random field, $Z_1(\mathbf{x}) = g(Z(\mathbf{x}))$, is Gaussian. The mean of the transformed field is unknown and parameterized by $\boldsymbol{\beta}$, $E(Z_1(\mathbf{x})) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$. If we assume that the transformation function g and the covariance function K of $Y(\mathbf{x})$ are known, the optimal predictor for $Z(\mathbf{x}_0)$ can be derived using the results from Section 1. However, in practice neither K nor g is known and we have to estimate them from the data.

A family of one-parameter transformation functions g_λ that is widely used in statistics is the so-called Box-Cox family

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(z), & \lambda = 0. \end{cases}$$

The Box-Cox transformation is valid for positive-valued random fields and is able to model moderately skewed, unimodal data.

The likelihood of the data \mathbf{Y} in the transformed Gaussian model can be written as

$$p(\mathbf{Y} | \boldsymbol{\Theta}) = J_\lambda(\mathbf{Y}) (2\pi)^{-\frac{n}{2}} |\sigma^2 \boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma^2} (g_\lambda(\mathbf{Y}) - \mathbf{F}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_\theta^{-1} (g_\lambda(\mathbf{Y}) - \mathbf{F}\boldsymbol{\beta}) \right],$$

where, $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \lambda)$, $J_\lambda(\mathbf{Y})$ is the determinant of the Jacobian of the transformation, $g_\lambda(\mathbf{Y}) = (g_\lambda(Y_1), \dots, g_\lambda(Y_n))$ and λ is the transformation parameter. De Oliveira et al. (1997) point out that the interpretation of $\boldsymbol{\beta}$ changes with the value of λ , and the same is true for the covariance parameters σ^2 and $\boldsymbol{\theta}$, to a lesser extent though. To estimate the parameters λ and $\boldsymbol{\theta}$, we make use of the profile likelihood approach that we have already encountered in Section 1. For fixed values of λ and $\boldsymbol{\theta}$, the maximum likelihood estimates for $\boldsymbol{\beta}$ and σ^2 are given by Eqs. (2)-(3) with \mathbf{Y} replaced by $g_\lambda(\mathbf{Y})$. Again, the estimates for λ and $\boldsymbol{\theta}$ cannot be written in closed form and must be found numerically by plugging $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ in the likelihood for numerical maximization.

The estimated parameters $\hat{\Theta}$ are subsequently used for spatial prediction. To perform a plug-in prediction we make use of the conditional distribution of the Gaussian variable $Y_0 | \mathbf{Y}, \hat{\Theta}$ and back-transform it to the original scale by g_λ^{-1} . A Bayesian approach to spatial prediction in the transformed Gaussian model is proposed in De Oliveira et al. (1997).

The copula-based geostatistical model (Kazianka and Pilz (2010)) also works with transformations of the marginal distributions of the random field and is a generalization of transformed Gaussian kriging. In this approach all multivariate distributions of the random field are described by a copula (Genest (2011)) and a family of univariate marginal distributions. Due to the additional flexibility introduced by the choice of the copula and of the marginal distribution, these models are able to deal with extreme observations and multi-modal data.

4 Generalized linear geostatistical models

Generalized linear models (McCullagh and Nelder (1989)) provide a unifying framework for regression modeling of both continuous and discrete data. Diggle and Ribeiro (2007) extend the classical generalized linear model to what they call the generalized linear geostatistical model (GLGM). The responses Y_i , $i = 1, \dots, n$, corresponding to location \mathbf{x}_i are assumed to follow a family of univariate distributions indexed by their expectation, μ_i , and to be conditionally independent given $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$. The μ_i are specified through

$$h(\mu_i) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta} + Z(\mathbf{x}_i),$$

where $Z(\mathbf{x})$ is a Gaussian random field with autocovariance function K_θ and h is a pre-defined link function. The two most frequently applied GLGMs are the Poisson log-linear model, where Y_i is assumed to follow a Poisson distribution and the link function is the logarithm, and the binomial logistic-linear model, where Y_i is assumed to follow a Binomial distribution, $Bin(n_i, p(\mathbf{x}_i))$, with $\mu_i = n_i p(\mathbf{x}_i)$ and $h(\mu_i) = \log(p(\mathbf{x}_i) / (1 - p(\mathbf{x}_i)))$. These models are suitable for representing spatially referenced count data and binomial trials.

Since maximum likelihood estimation of the parameters is difficult, a Markov chain Monte Carlo (Robert and Casella (2004)) approach is proposed to sample from the posteriors of the model parameters as well as from the predictive distributions at unobserved locations \mathbf{x}_0 . The algorithm proceeds by sampling from $\mathbf{Z} | \mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\theta}$, from $\boldsymbol{\theta} | \mathbf{Z}$ and from $\boldsymbol{\beta} | \mathbf{Z}, \mathbf{Y}$ with the help of Metropolis-Hastings updates. At iteration $t + 1$ and actual sample $(\mathbf{Z}^t, \boldsymbol{\theta}^t, \boldsymbol{\beta}^t, Z^t(\mathbf{x}_0))$, perform the following steps:

- (a). Update \mathbf{Z} . For $i = 1, \dots, n$, sample a new proposal $Z'(\mathbf{x}_i)$ from the conditional Gaussian distribution $p(Z(\mathbf{x}_i) | \boldsymbol{\theta}^t, \mathbf{Z}_{-i}^t)$, where \mathbf{Z}_{-i}^t denotes $\mathbf{Z}^t = (Z^t(\mathbf{x}_1), \dots, Z^t(\mathbf{x}_n))$ with its i th element removed. Accept $Z'(\mathbf{x}_i)$ with probability $r = \min \left\{ 1, \frac{p(Y_i | \boldsymbol{\beta}^t, Z'(\mathbf{x}_i))}{p(Y_i | \boldsymbol{\beta}^t, Z^t(\mathbf{x}_i))} \right\}$.

- (b). Update $\boldsymbol{\theta}$. Sample a new proposal $\boldsymbol{\theta}'$ from a proposal distribution $J(\boldsymbol{\theta} | \boldsymbol{\theta}^t)$.
 Accept the new proposal with probability $r = \min \left\{ 1, \frac{p(\mathbf{Z}^{t+1} | \boldsymbol{\theta}') J(\boldsymbol{\theta}^t | \boldsymbol{\theta}')}{p(\mathbf{Z}^{t+1} | \boldsymbol{\theta}^t) J(\boldsymbol{\theta}' | \boldsymbol{\theta}^t)} \right\}$.
- (c). Update $\boldsymbol{\beta}$. Sample a new proposal $\boldsymbol{\beta}'$ from a proposal distribution $J(\boldsymbol{\beta} | \boldsymbol{\beta}^t)$.
 Accept the new proposal with probability $r = \min \left\{ 1, \frac{\prod_{i=1}^n p(Y_i | Z^{t+1}(\mathbf{x}_i), \boldsymbol{\beta}') J(\boldsymbol{\beta}^t | \boldsymbol{\beta}')}{\prod_{i=1}^n p(Y_i | Z^{t+1}(\mathbf{x}_i), \boldsymbol{\beta}^t) J(\boldsymbol{\beta}' | \boldsymbol{\beta}^t)} \right\}$
- (d). Draw a sample $Z^{t+1}(\mathbf{x}_0)$ from the conditional Gaussian distribution $Z(\mathbf{x}_0) | \mathbf{Z}^{t+1}, \boldsymbol{\theta}^{t+1}$.

If point predictions for $Z(\mathbf{x}_0)$ are needed, the Monte Carlo approximation to the expected value of $Z(\mathbf{x}_0) | \mathbf{Y}$ can be used, i.e. $E(Z(\mathbf{x}_0) | \mathbf{Y}) \approx \frac{1}{M} \sum_{t=1}^M Z^t(\mathbf{x}_0)$, where M is the number of simulations.

5 References

1. De Oliveira, V., Kedem, B. & Short, D. (1997) Bayesian prediction of transformed Gaussian fields. *Journal of the American Statistical Association* **92**, 1422–1433.
2. Diggle, P. & Ribeiro, P. (2007) *Model-based geostatistics*. New York: Springer
3. Genest, C. (2011) Copulas in inference. In: Lovric, M. *International Encyclopedia of Statistical Science*. Heidelberg: Springer Science+Business Media, LLC.
4. Kazianka, H. & Pilz, J. (2010) Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment* **24**, 661-673.
5. Kitanidis, P. (1986) Parameter uncertainty in estimation of spatial function: Bayesian analysis. *Water Resources Research* **22**, 499-507.
6. Mase, S. (2011) Geostatistics and kriging predictors. In: Lovric, M. *International Encyclopedia of Statistical Science*. Heidelberg: Springer Science+Business Media, LLC.
7. McCullagh, P. & Nelder, J. (1989) *Generalized linear models*. Boca Raton: Chapman & Hall/CRC.

8. Omre, H. (1987) Bayesian kriging - merging observations and qualified guesses in kriging. *Mathematical Geology* **19**, 25-39.
9. Robert, C. & Casella, G. (2004) *Monte Carlo statistical methods*. New York: Springer.
10. Schabenberger, O. & Gotway, C. (2005) *Statistical methods for spatial data analysis*. Boca Raton: Chapman & Hall/CRC.
11. Stein, M. (1999) *Interpolation of spatial data*. New York: Springer.
12. Yaglom, A. (1987) *Correlation theory of stationary and related random functions*. New York: Springer.

Reprinted with permission from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science. Heidelberg: Springer Science+Business Media, LCC