Bayesian Spatial Prediction: BTG

Benjamin Kedem

Department of Mathematics & Inst. for Systems Research
University of Maryland, College Park

Victor De Oliveira, David Bindel, Boris and Sandra Kozintsev

GWU, February 20, 2009

Outline

1. Introduction
   - Stationary isotropic Gaussian random fields
2. Spatial Prediction
   - Ordinary Kriging
   - BTG

(•) Berger, De Oliveira, Sanso (2001). Objective Bayesian Analysis of Spatially Correlated Data. JASA, 96, 1361-1374.

(•) De Oliveira, Kedem, Short (1997). Bayesian Prediction of Transformed Gaussian Random Fields. JASA, 92, 1422-1433.

(•) De Oliveira, Ecker (2002). Bayesian Hot Spot Detection in the Presence of a Spatial Trend: Application to Total Nitrogen Concentration in the Chesapeake Bay. *Environmetrics*, 13, 85-101.

(•) Handcock, Stein (1993). A Bayesian Analysis of Kriging. *Technometrics*, 35, 403-410.

(•) Kedem, Fokianos (2002). *Regression Models for Time Series Analysis*, New York: Wiley.

(•) Kozintsev, Kedem (2000). Generation of "Similar" Images From a Given Discrete Image. *J. Comp. Graphical Stat.*, 2000, Vol. 9, No. 2, 286-302.

(•) Kozintseva (1999). Comparison of Three Methods of Spatial Prediction. M.A. Thesis, Department of Mathematics, University of Maryland, College Park.

(•) Bindel, De Oliveira, Kedem (1997). An implementation of the BTG spatial prediction model.
http://www.math.umd.edu/bnk/btg_page.html

Spatial/temporal geostatistical data often display:

- Non-Gaussian skewed sampling distributions.
- Positive continuous data.
- Heavy right tails.
- Bounded support.
- Small data sets observed irregularly (gaps).

Spatial/temporal geostatistical data often display:

- Non-Gaussian skewed sampling distributions.
- Positive continuous data.
- Heavy right tails.
- Bounded support.
- Small data sets observed irregularly (gaps).

Spatial/temporal geostatistical data often display:

- Non-Gaussian skewed sampling distributions.
- Positive continuous data.
- Heavy right tails.
- Bounded support.
- Small data sets observed irregularly (gaps).

Spatial/temporal geostatistical data often display:

- Non-Gaussian skewed sampling distributions.
- Positive continuous data.
- Heavy right tails.
- Bounded support.
- Small data sets observed irregularly (gaps).

Spatial/temporal geostatistical data often display:

- Non-Gaussian skewed sampling distributions.
- Positive continuous data.
- Heavy right tails.
- Bounded support.
- Small data sets observed irregularly (gaps).

Possible remedies:

- Bayesian Transformed Gaussian (BTG): A Bayesian approach combined with parametric families of nonlinear transformations to Gaussian data.
- BTG provides a unified framework for inference and prediction/interpolation in a wide variety of models, Gaussian and non-Gaussian.
- Will describe BTG and illustrate it using spatial and temporal data.

Possible remedies:

- Bayesian Transformed Gaussian (BTG): A Bayesian approach combined with parametric families of nonlinear transformations to Gaussian data.
- BTG provides a unified framework for inference and prediction/interpolation in a wide variety of models, Gaussian and non-Gaussian.
- Will describe BTG and illustrate it using spatial and temporal data.

Possible remedies:

- Bayesian Transformed Gaussian (BTG): A Bayesian approach combined with parametric families of nonlinear transformations to Gaussian data.
- BTG provides a unified framework for inference and prediction/interpolation in a wide variety of models, Gaussian and non-Gaussian.
- Will describe BTG and illustrate it using spatial and temporal data.

Let $\{Z(\mathbf{s})\}$, $\mathbf{s} \in D \subset R^d$, be a spatial process or a random field.

A random field $\{Z(\mathbf{s})\}$ is Gaussian if for all $\mathbf{s}_1, ..., \mathbf{s}_n \in D$, the vector $(Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))$ has a multivariate normal distribution.

$\{Z(\mathbf{s})\}$ is (second order) stationary when for $\mathbf{s}, \mathbf{s}+\mathbf{h} \in D$ we have

$$\begin{aligned}
(\bullet) \quad & \mathrm{E}(Z(\mathbf{s})) = \mu, \\
(\bullet) \quad & \mathrm{Cov}(Z(\mathbf{s}+\mathbf{h}), Z(\mathbf{s})) \equiv C(\mathbf{h}).
\end{aligned}$$

The function $C(\cdot)$ is called the covariogram or covariance function.

We shall assume that $C(\mathbf{h})$ depends only on the distance $\|\mathbf{h}\|$ between the locations $\mathbf{s} + \mathbf{h}$ and $\mathbf{s}$ but not on the direction of $\mathbf{h}$.

In this case the covariance function as well as the process are called isotropic.

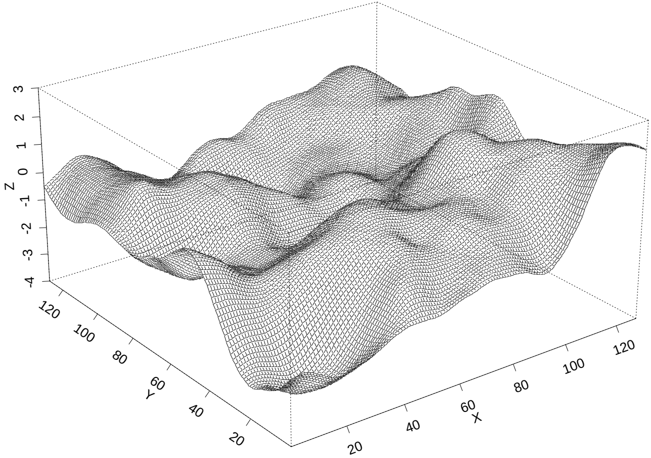
The corresponding isotropic correlation function is given by $K(l) = C(l)/C(0)$, where l is the distance between points.

Useful special case: **Matérn correlation**

$$K_{\boldsymbol{\theta}}(l) = \begin{cases} \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{l}{\theta_1}\right)^{\theta_2} \kappa_{\theta_2}\left(\frac{l}{\theta_1}\right) & \text{if } l \neq 0 \\ 1 & \text{if } l = 0 \end{cases}$$

where $\theta_1 > 0, \theta_2 > 0$, and κ_{θ_2} is a modified Bessel function of the third kind of order θ_2.

Matérn ($\theta_1 = 8, \theta_2 = 3$).

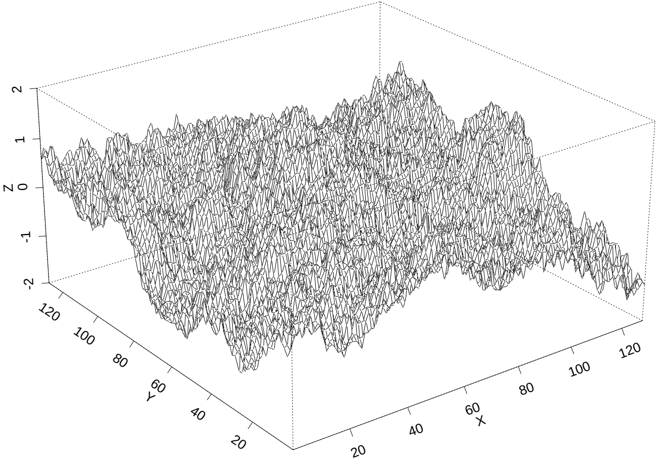


Spherical correlation:

$$K_{\boldsymbol{\theta}}(l) = \begin{cases} 1 - \frac{3}{2}\left(\frac{l}{\theta}\right) + \frac{1}{2}\left(\frac{l}{\theta}\right)^3 & \text{if } l \leq \theta \\ 0 & \text{if } l > \theta \end{cases}$$

where $\theta > 0$ controls the correlation range.

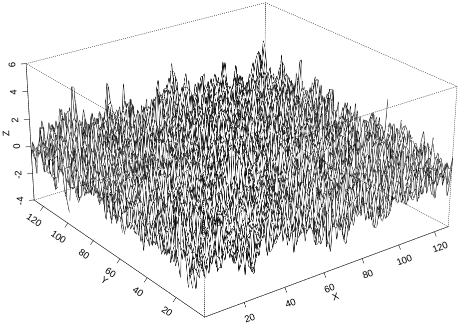
Spherical ($\theta = 120$).



Exponential correlation:

$$K_{\boldsymbol{\theta}}(l) = \exp(l^{\theta_2} \log \theta_1), \quad \theta_1 \in (0,1), \theta_2 \in (0,2]$$

Exponential ($\theta_1 = 0.5, \theta_2 = 1$).

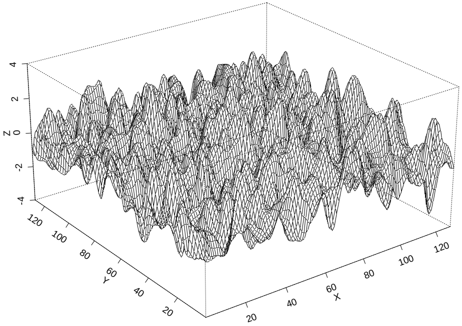


Rational quadratic:

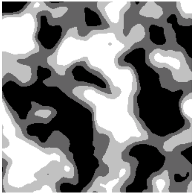
$$K_{\boldsymbol{\theta}}(l) = \left(1 + \frac{2}{\theta_1^2}\right)^{-\theta_2}$$

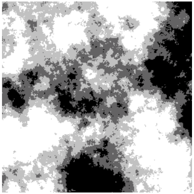
$\theta_1 > 0$, $\theta_2 > 0$.

Rational quadratic ($\theta_1 = 12, \theta_2 = 8$).

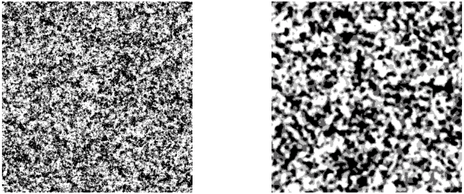


Clipped, at 3 levels, realizations from Gaussian random fields. Left: Matérn (8,3). Right: spherical (120).





Clipped, at 3 levels, realizations from Gaussian random fields.
Left: exponential (0.5,1). Right: rational quadratic (12,8).



www.math.umd.edu/-bnk/bak/generate.cgi?4
Kozintsev (1999), Kozintsev and Kedem (2000).

Outline

1. Introduction
   - Stationary isotropic Gaussian random fields
2. Spatial Prediction
   - Ordinary Kriging
   - BTG

Ordinary Kriging.

Given the data

$$\mathbf{Z} \equiv (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$$

observed at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ in D, the problem is to predict (or estimate) $Z(\mathbf{s}_0)$ at location $\mathbf{s}_0$ using the best linear unbiased predictor (BLUP) obtained by minimizing

$$\mathrm{E}\Big(Z(\mathbf{s}_0) - \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)\Big)^2 \quad \text{subject to} \quad \sum_{i=1}^{n} \lambda_i = 1$$

Define

$$
\begin{aligned}
\mathbf{1} &= (1, 1, \ldots, 1)', \quad 1 \times n \text{ vector} \\
\mathbf{c} &= \big(C(\mathbf{s}_0 - \mathbf{s}_1), \ldots, C(\mathbf{s}_0 - \mathbf{s}_n)\big)' \\
\mathbf{C} &= \big(C(\mathbf{s}_i - \mathbf{s}_j)\big), \quad i, j = 1, ..., n \\
\boldsymbol{\lambda} &= (\lambda_1, \lambda_2, \ldots, \lambda_n)'
\end{aligned}
$$

Then

$$
\hat{\boldsymbol{\lambda}} = \mathbf{C}^{-1}\left(\mathbf{c} + \frac{1 - \mathbf{1}'\mathbf{C}^{-1}\mathbf{c}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}\mathbf{1}\right).
$$

The ordinary kriging predictor is then

$$
\hat{Z}(\mathbf{s}_0) = \hat{\boldsymbol{\lambda}}'\mathbf{Z}.
$$

Define,

$$m = \frac{1 - \mathbf{1}'\mathbf{C}^{-1}\mathbf{c}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}$$

and the *kriging variance*

$$\sigma_k^2(\mathbf{s}_0) = \mathrm{E}\big(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\big)^2 = C(\mathbf{0}) - \hat{\lambda}'\mathbf{c} + m.$$

Under the Gaussian assumption,

$$\hat{Z}(\mathbf{s}_0) \pm 1.96\sigma_k(\mathbf{s}_0)$$

is a 95% prediction interval for $Z(\mathbf{s}_0)$. For non-Gaussian fields this may not hold.

Bayesian Spatial Prediction: The BTG Model

RF $\{Z(\mathbf{s}),\ \mathbf{s} \in D\}$ observed at $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$. Parametric family of monotone transformations

$$\mathcal{G} = \{g_\lambda(.) : \lambda \in \Lambda\}$$

($\star$) Assumption: $Z(.)$ can be transformed into a Gaussian random field by a member of $\mathcal{G}$.

A useful parametric family of transformations often used in applications to 'normalize' positive data is the Box-Cox (1964) family of power transformations,

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

For some unknown ‘transformation parameter’ $\lambda \in \Lambda$, $\{g_\lambda(Z(\mathbf{s})),\ \mathbf{s} \in D\}$ is a Gaussian random field with

$$E\{g_\lambda(Z(\mathbf{s}))\} = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}),$$

$$\text{cov}\{g_\lambda(Z(\mathbf{s})), g_\lambda(Z(\mathbf{u}))\} = \tau^{-1} K_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{u}),$$

Regression parameters: $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$

Covariates: $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \ldots, f_p(\mathbf{s}))$

Variance: $\tau^{-1} = \text{var}\{g_\lambda(Z(\mathbf{s}))\}$

Simplifying assumption: Isotropy,

$$K_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{u}) = K_{\boldsymbol{\theta}}(||\mathbf{s} - \mathbf{u}||), \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_q) \in \Theta \subset R^q$$

Data: $\mathbf{Z}_{obs} = (Z_{1,obs}, \ldots, Z_{n,obs})$

$$g_\lambda(Z_{i,obs}) = g_\lambda(Z(\mathbf{s}_i)) + \epsilon_i \ ; \quad i = 1, \ldots, n,$$

$\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, \frac{\xi}{\tau})$.

Parameters: $\boldsymbol{\eta} = (\boldsymbol{\beta}, \tau, \xi, \boldsymbol{\theta}, \lambda)$.

Prediction problem:

Predict $\mathbf{Z}_0 = (Z(\mathbf{s}_{01}), \ldots, Z(\mathbf{s}_{0k}))$ from the predictive density function, defined by

$$p(\mathbf{z}_o|\mathbf{z}_{obs}) = \int_{\Omega} p(\mathbf{z}_o, \eta|\mathbf{z}_{obs}) d\eta$$
$$= \int_{\Omega} p(\mathbf{z}_o|\eta, \mathbf{z}_{obs}) p(\eta|\mathbf{z}_{obs}) d\eta,$$

where $\Omega = R^p \times (0, \infty)^2 \times \Theta \times \Lambda$.

Notation: For $\mathbf{a} = (a_1, \ldots, a_n)$, we write

$$\underline{g}_\lambda(\mathbf{a}) \equiv (g_\lambda(a_1), \ldots, g_\lambda(a_n)).$$

The Likelihood:

$$L(\eta; \mathbf{z}_{obs}) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \left|\Psi_{\xi,\boldsymbol{\theta}}\right|^{-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}Q\right\} J_\lambda,$$

$$Q = \left(\underline{g}_\lambda(\mathbf{z}_{obs}) - X\beta\right)' \Psi_{\xi,\boldsymbol{\theta}}^{-1} \left(\underline{g}_\lambda(\mathbf{z}_{obs}) - X\beta\right).$$

X $n \times p$ design matrix, $X_{ij} = f_j(\mathbf{s}_i)$.
$\Psi_{\xi,\boldsymbol{\theta}} = \Sigma_{\boldsymbol{\theta}} + \xi I$, $n \times n$ matrix.
$\Sigma_{\boldsymbol{\theta};ij} = K_{\boldsymbol{\theta}}(\mathbf{s}_i, \mathbf{s}_j)$.
I identity matrix.
$J_\lambda = \prod_{i=1}^{n} |g'(z_{i,obs})|$, the Jacobian.

The Prior

Insightful arguments in Box and Cox(1964), De Oliveira, Kedem, Short (1997), as well as practical experience lead us to the prior

$$p(\boldsymbol{\eta}) \propto \frac{p(\xi)p(\boldsymbol{\theta})p(\lambda)}{\tau J_{\lambda}^{\frac{p}{n}}},$$

where $p(\xi)$, $p(\boldsymbol{\theta})$ and $p(\lambda)$ are the prior marginals of ξ, $\boldsymbol{\theta}$ and λ, respectively, which are assumed to be proper.

Unusual prior: it depends on the data through the Jacobian.

For more on prior selection see Berger, De Oliveira, Sansó (2001).

Simplifying assumption: No measurement noise ($\xi = 0$).

$$g_\lambda(Z_{i,obs}) = g_\lambda(Z(\mathbf{s}_i)), \quad i = 1, \ldots, n,$$

$$p(\boldsymbol{\beta}, \tau, \boldsymbol{\theta}, \lambda) \propto \frac{p(\boldsymbol{\theta})p(\lambda)}{\tau J_\lambda^{p/n}} \tag{1}$$

$$\boldsymbol{\eta} = (\boldsymbol{\beta}, \tau, \boldsymbol{\theta}, \lambda)'.$$

Also, write

$$\mathbf{z} = \mathbf{z}_{obs}$$

The Posterior

$$p(\boldsymbol{\eta}|\mathbf{z}) = p(\boldsymbol{\beta}, \tau, \boldsymbol{\theta}, \lambda|\mathbf{z}) = p(\boldsymbol{\beta}, \tau|\boldsymbol{\theta}, \lambda, \mathbf{z})p(\boldsymbol{\theta}, \lambda|\mathbf{z}).$$

To get the first factor:

$$(\boldsymbol{\beta}|\tau, \boldsymbol{\theta}, \lambda, \mathbf{z}) \sim \mathcal{N}_p(\hat{\beta}_{\boldsymbol{\theta},\lambda}, \frac{1}{\tau}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1})$$

$$(\tau|\boldsymbol{\theta}, \lambda, \mathbf{z}) \sim Ga(\frac{n-p}{2}, \frac{2}{\tilde{q}_{\boldsymbol{\theta},\lambda}})$$

where

$$\hat{\beta}_{\boldsymbol{\theta},\lambda} = (\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\underline{g}_{\lambda}(\mathbf{z})$$

$$\tilde{q}_{\boldsymbol{\theta},\lambda} = (\underline{g}_{\lambda}(\mathbf{z}) - \mathbf{X}\hat{\beta}_{\boldsymbol{\theta},\lambda})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\underline{g}_{\lambda}(\mathbf{z}) - \mathbf{X}\hat{\beta}_{\boldsymbol{\theta},\lambda}).$$

and we get Normal-Gamma:

$$p(\boldsymbol{\beta}, \tau|\boldsymbol{\theta}, \lambda, \mathbf{z}) = p(\boldsymbol{\beta}|\tau, \boldsymbol{\theta}, \lambda, \mathbf{z})p(\tau|\boldsymbol{\theta}, \lambda, \mathbf{z})$$

To get the second factor:

$$p(\boldsymbol{\theta}, \lambda | \mathbf{z}) \propto$$
$$|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} |\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X}|^{-1/2} \tilde{q}_{\boldsymbol{\theta},\lambda}^{-\frac{n-p}{2}} J_{\lambda}^{1-\frac{p}{n}} p(\boldsymbol{\theta}) p(\lambda)$$

In addition to the joint posterior distribution $p(\boldsymbol{\eta}|\mathbf{z})$ derived above, the predictive density $p(\mathbf{z}_o|\mathbf{z})$ also requires $p(\mathbf{z}_o|\boldsymbol{\eta}, \mathbf{z})$. We have

$$p(\mathbf{z}_o|\boldsymbol{\eta}, \mathbf{z}) = (\frac{\tau}{2\pi})^{k/2} |\mathbf{D}_{\boldsymbol{\theta}}|^{-1/2} \prod_{j=1}^{k} |g'_{\lambda}(z_{oj})|$$
$$\times \exp\left\{-\frac{\tau}{2}(\underline{g}_{\lambda}(\mathbf{z}_o) - \mathbf{M}_{\boldsymbol{\beta},\boldsymbol{\theta},\lambda})' \mathbf{D}_{\boldsymbol{\theta}}^{-1} (\underline{g}_{\lambda}(\mathbf{z}) - \mathbf{M}_{\boldsymbol{\beta},\boldsymbol{\theta},\lambda})\right\}$$

where $\mathbf{M}_{\boldsymbol{\beta},\boldsymbol{\theta},\lambda}, \mathbf{D}_{\boldsymbol{\theta}}$ are known.

We now have the integrand $p(\mathbf{z}_o|\boldsymbol{\eta}, \mathbf{z})p(\boldsymbol{\eta}|\mathbf{z})$ needed for $p(\mathbf{z}_o|\mathbf{z})$. By integrating out β and τ we obtain the simplified form of the predictive density:

$$
\begin{aligned}
p(\mathbf{z}_o|\mathbf{z}) &= \int_\Lambda \int_\Theta p(\mathbf{z}_o|\boldsymbol{\theta}, \lambda, \mathbf{z})p(\boldsymbol{\theta}, \lambda|\mathbf{z})d\boldsymbol{\theta}d\lambda \\
&= \frac{\int_\Lambda \int_\Theta p(\mathbf{z}_o|\boldsymbol{\theta}, \lambda, \mathbf{z})p(\mathbf{z}|\boldsymbol{\theta}, \lambda)p(\boldsymbol{\theta})p(\lambda)d\boldsymbol{\theta}d\lambda}{\int_\Lambda \int_\Theta p(\mathbf{z}|\boldsymbol{\theta}, \lambda)p(\boldsymbol{\theta})p(\lambda)d\boldsymbol{\theta}d\lambda}
\end{aligned}
$$

where

$$p(\mathbf{z}_o|\boldsymbol{\theta},\lambda,\mathbf{z}) = \frac{\Gamma(\frac{n-p+k}{2})\prod_{j=1}^{k}|g'_\lambda(z_{oj})|}{\Gamma(\frac{n-p}{2})\pi^{k/2}|\tilde{q}_{\boldsymbol{\theta},\lambda}\mathbf{C}_{\boldsymbol{\theta}}|^{1/2}}$$
$$\times[1+(\underline{g}_\lambda(\mathbf{z}_o)-\mathbf{m}_{\boldsymbol{\theta},\lambda})'(\tilde{q}_{\boldsymbol{\theta},\lambda}\mathbf{C}_{\boldsymbol{\theta}})^{-1}$$
$$\times(\underline{g}_\lambda(\mathbf{z}_o)-\mathbf{m}_{\boldsymbol{\theta},\lambda})]^{-\frac{n-p+k}{2}}$$

and from Bayes theorem,

$$p(\mathbf{z}|\boldsymbol{\theta},\lambda) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2}|\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}|^{-1/2}\tilde{q}_{\boldsymbol{\theta},\lambda}^{-\frac{n-p}{2}}J_\lambda^{1-\frac{p}{n}}$$

where $\mathbf{m}_{\boldsymbol{\theta},\lambda},\mathbf{C}_{\boldsymbol{\theta}}$ are known.

BTG Algorithm: Predictive Density Approximation

- **1.** Let $S = \{z_o^{(j)} : j = 1, \dots, r\}$ be the set of values obtained by discretizing the effective range of Z_0.
- **2.** Generate independently $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m$ i.i.d. $\sim p(\boldsymbol{\theta})$ and $\lambda_1, \dots, \lambda_m$ i.i.d. $\sim p(\lambda)$.
- **3.** For $z_o \in S$, the approximation to $p(z_o|\mathbf{z})$ is given by

$$\hat{p}_m(z_o|\mathbf{z}) = \frac{\sum_{i=1}^m p(z_o|\boldsymbol{\theta}_i, \lambda_i, \mathbf{z}) p(\mathbf{z}|\boldsymbol{\theta}_i, \lambda_i)}{\sum_{i=1}^m p(\mathbf{z}|\boldsymbol{\theta}_i, \lambda_i)}$$

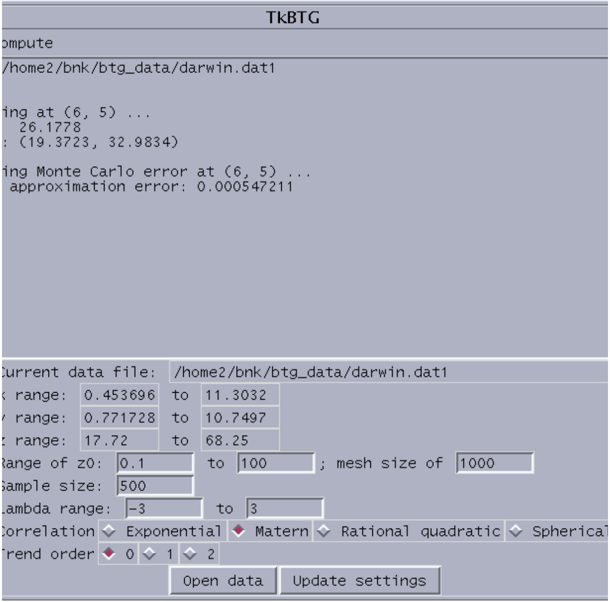
$p(z_o|\boldsymbol{\theta}, \lambda, \mathbf{z})$ and $p(\mathbf{z}|\boldsymbol{\theta}, \lambda)$ given above.

The point predictor is the median of the estimated predictive distribution:

$$(\star) \quad \hat{Z}_0 = \text{Median of } (Z_0|\mathbf{Z})$$

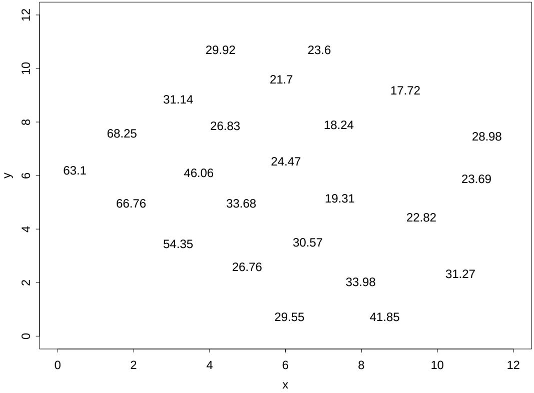
tkbtg Interface Layout

Software: **tkbtg** application. Hybrid of C++, Tcl/Tk, and FORTRAN 77 (Bindel et al (1997)).
www.math.umd.edu/-bnk/btg_page.html

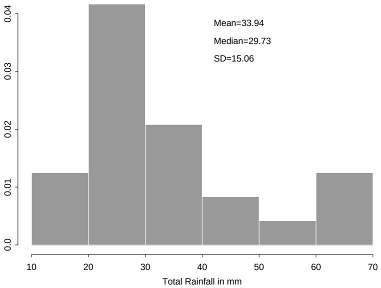

TkBTG
ompute
/home2/bnk/btg_data/darwin.dat1
ing at (6, 5) ...
26.1778
: (19.3723, 32.9834)
ing Monte Carlo error at (6, 5) ...
approximation error: 0.000547211
Current data file: /home2/bnk/btg_data/darwin.dat1
range: 0.453696 to 11.3032
range: 0.771728 to 10.7497
range: 17.72 to 68.25
Range of z0: 0.1 to 100 ; mesh size of 1000
Sample size: 500
Lambda range: -3 to 3
Correlation Exponential Matern Rational quadratic Spherical
Trend order 0 1 2
Open data
Update settings

Example: Spatial Rainfall Prediction

Rain gauge positions and weekly rainfall totals in mm, Darwin, Australia, 1991.



1. Use the Box-Cox transformation family.
2. $\lambda \sim \mathrm{Unif}(-3, 3)$.
3. $m = 500$.
4. Correlation: Matérn and exponential.
5. No covariate information. Assume constant regression: $E\{g_\lambda(Z(\mathbf{s}))\} = \beta_1$.
6. Data apparently not normal.

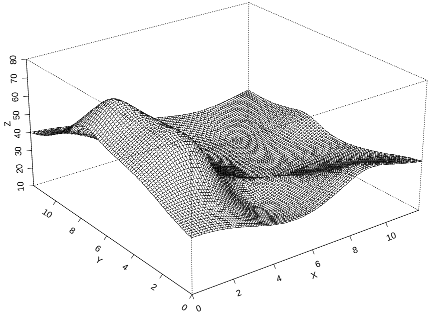


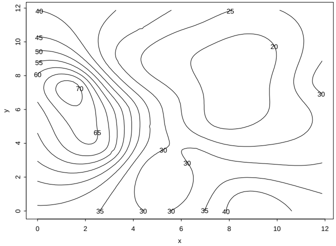
Results of Cross Validation From 23 Observations

No.	z	$\hat{z}$	95% PI
1	29.55	33.74	(10.70, 56.78)
2	41.85	34.32	(15.29, 53.35)
3	26.76	36.71	(20.93, 52.49)
4	33.98	33.39	(18.49, 48.29)
5	31.27	32.99	(9.53, 56.45)
6	54.35	39.13	(16.46, 61.79)
7	30.57	24.45	(15.40, 33.59)
8	22.82	23.09	(11.52, 34.66)
9	66.76	64.12	(28.25, 100)
10	33.68	35.16	(18.01, 52.30)
11	19.31	24.51	(15.62, 33.40)
12	23.69	26.45	(15.35, 37.54)

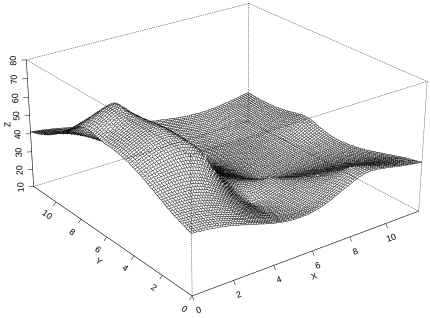
No.	z	$\hat{z}$	95% PI
13	63.10	72.07	(44.14, 100)
14	46.06	40.60	(18.58, 62.63)
15	24.47	22.32	(14.00, 30.63)
16	28.98	21.62	(13.43, 29.81)
17	68.25	46.54	(19.25, 73.84)
18	26.83	29.52	(16.57, 42.46)
19	18.24	19.00	(10.79, 27.21)
20	31.14	37.36	(20.33, 54.39)
21	21.70	22.97	(14.71, 31.22)
22	17.72	22.69	(11.64, 33.74)
23	29.92	26.56	(12.21, 40.91)
24	23.60	21.83	(10.85, 32.81)

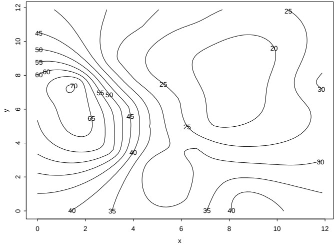
Spatial prediction and contour maps from the Darwin data using *Matérn correlation*.



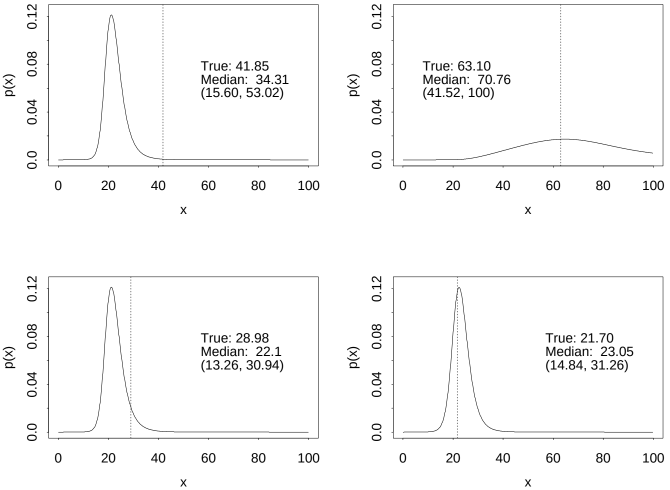


Spatial prediction and contour maps from the Darwin data using *exponential correlation*.





Predictive densities, 95% PI's, and cross-validation: Predicting a true value from the remaining 23 observations using Matérn correlation. The vertical line marks true values.



BTG vs Kriging and Trans-Gaussian kriging (Kozintseva (1999)).

- Cross validation results using artificial data on 50 $\times$ 50 grid.
- Data obtained by transforming a Gaussian (0,1) RF using inverse Box-Cox transformation.
- In Kriging and TG kriging λ, $\boldsymbol{\theta}$, were known. Not in BTG (!)
- $\lambda = 0$: Log-Normal.
  $\lambda = 1$: Normal.
  $\lambda = 0.5$: Between Normal and Log-Normal.
- In most cases BTG has more reliable but larger prediction intervals.
- BTG predicts at the original scale. TG kriging does not.

	Matérn(1,10)		
λ	0	0.5	1
KRG MSE	68397.48	7.15	0.58
TGK MSE	55260.90	7.08	0.58
BTG MSE	64134.30	7.31	0.56
KRG AvePI	2.42	2.51	2.42
TGK AvePI	291.80	8.21	2.42
BTG AvePI	330.68	10.23	2.87
KRG % out	100%	48%	6%
TGK % out	18%	8%	6%
BTG % out	12%	6%	6%

	Exponential($e^{-0.03}$, 1)		
λ	0	0.5	1
KRG MSE	12212.32	1.83	0.13
TGK MSE	11974.73	1.84	0.13
BTG MSE	12520.70	1.89	0.14
KRG AvePI	1.45	1.43	1.45
TGK AvePI	267.92	5.24	1.45
BTG AvePI	466.69	6.10	1.63
KRG % out	98%	64%	2%
TGK % out	20%	4%	2%
BTG % out	6%	2%	2%

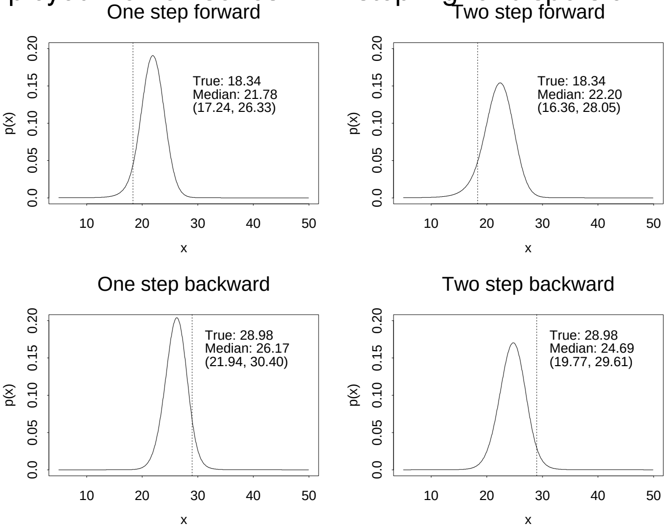
Application of BTG to Time Series Prediction.

- Short time series observed irregularly.
- Set: $\mathbf{s} = (x, y) = (t, 0)$.
- Can predict/interpolate as in state space prediction: k–step prediction forward, backward, and “in the middle”.
- Example 1: Monthly data of unemployed women 20 years of age and older, 1997–2000. Data source: Bureau of Labor Statistics. $N = 48$.
- Example 2: Monthly airline passenger data, 1949–1960. Data source: Box-Jenkins (1976). Use only $N = 36$ out of 144 observations, $t = 51, ..., 86$.

Example: Prediction of Monthly number of unemployed women $(\text{Age} \geq 20)$, 1997–2000. Data in hundreds of thousands.

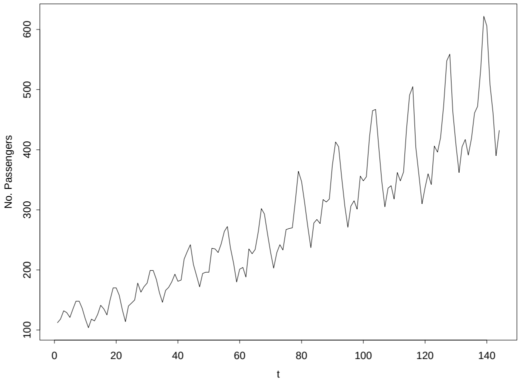
Cross validation and 95% PI's. Observations at times $t = 12, 13, 36$ are outside their 95% PI's. $N = 48 - 1 = 47$.

Forward and backward one and two step prediction in the unemployed women series. In 2-step higher dispersion.

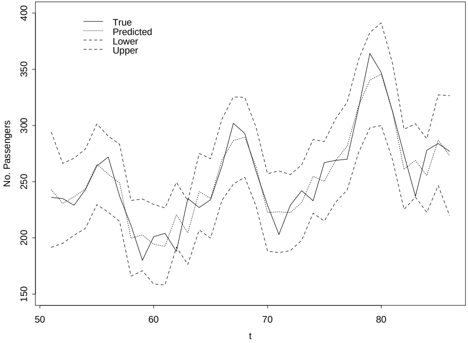


Example: Prediction of No. of Airline Passengers.

Time series of monthly international airline passengers in thousands, January 1949-December 1960. N=144. Seasonal time series.



BTG cross validation and prediction intervals for the monthly airline passengers series, $t = 51, ..., 86$, using Matérn correlation. Observations at $t = 62, 63$ are outside the PI. $N = 36 - 1 = 35$.



Application to Rainfall: Heuristic Argument (Kedem and Chiu, PNAS, 1987.)

Let X_n represent the area average rain rate over a region such that

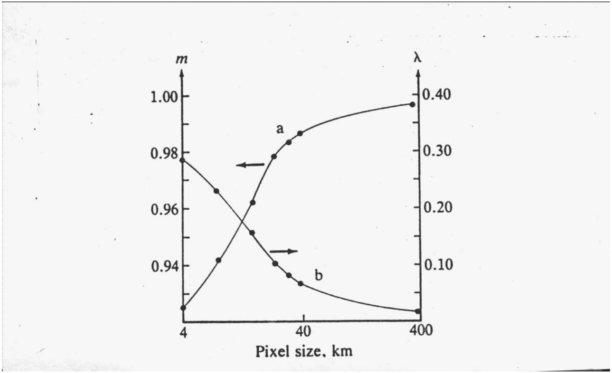
$$X_n = mX_{n-1} + \lambda + \epsilon_n, \qquad n = 1, 2, 3, \cdots ,$$

where the noise $\{\epsilon_n\}$ is a *martingale difference*. It can be argued that for

$$X_n \to \mathrm{LogNormal}, \quad n \to \infty,$$

we need $m \to 1^-$ and $\lambda \to 0^+$.

The monotone increase in $\hat{m}$ (Curve a) and the monotone decrease in $\hat{\lambda}$ (Curve b) as a function of the square root of the area. Source: Kedem and Chiu(1987).



This suggests that the lognormal distribution as a model for averages or rainfall amounts over large areas or long periods.

It is interesting to obtain the posterior $p(\lambda \mid \mathbf{z})$ of λ, the transformation parameter in the Box-Cox family, given the data, where the data are weekly rainfall totals from Darwin, Australia.

With a uniform prior for λ, the medians of $p(\lambda \mid \mathbf{z})$ in 5 different weeks are:

Week 1 median = $-$ 0.45 (to the left of 0).
Week 2 median = 0.45 (to the right of 0).
Week 3 median = 0.15 (Not far from 0).
Week 4 median = 0.20 (Not far from 0).
Week 5 median = 0.95 (Close to 1).

Weekly posterior $p(\lambda \mid \mathbf{z})$ of λ given rainfall totals from Darwin, Australia, for 5 different weeks.