Geostatistical Models: Model Selection and Parameter Estimation under Infill and Expanding Domain Asymptotics

Andrew A. Merton
Department of Statistics, Colorado State University
30 October 2006
Motivation

• Suppose one has collected information about a particular response variable along with several potential covariates across a finite region of study.
  – For example, one might be interested in the presence of a pollutant and its dependence on atmospheric conditions, topography, land usage, etc...

• A rich class of models for describing the relationship is \( Y(s) = X(s)'\beta + \delta(s) \), where
  – \( Y(s) \) is the response at location \( s \),
  – \( X(s) \) is a (sub)collection of the covariates,
  – \( \delta(s) \) is assumed to be a Gaussian random field.

• Common issues that arise during modeling are
  – identifying the “best” subset of covariates,
  – identifying the appropriate form for the covariance function,
  – estimation of the model parameters, and
  – making inference about the model parameter estimators.

• The work presented in this dissertation explores all of these aspects for this model class.
Scope of the dissertation

1. Model selection and parameter estimation
     • Derivation of the spatial $AIC_c$ statistic for model selection
     • Simulation studies to evaluate the performance of spatial $AIC_c$

2. Exponential class of correlation functions in one-dimension
   (a) Derive asymptotic distribution of the MLE for the range parameter under various sampling designs
   (b) Simulation studies to corroborate theoretical results

3. Simulation study of the exponential class in two-dimensions

4. Simulation study of the Matérn class in both one- and two-dimensions
Geostatistical model

• Let $Z = (Z(s_1), \ldots, Z(s_n))'$ be a partial realization of a random field $Z$, where $s_j \in D$, a fixed, finite area under study.

• We observe $\{Y(s_j)\} = \{Z(s_j) + \epsilon(s_j)\}$ where $\epsilon(s_j) \sim \text{iid}(0, \tau^2)$.

• A model for the random field at any location $s \in D$ is given by

$$Z(s) = X'(s)\beta + \delta(s),$$

where

– $X(s) = (1, X_1(s), \ldots, X_{p-1}(s))'$ is a $p$-vector of covariates observed at location $s$,
– $\beta$ is a $p$-vector of unknown coefficients, and
– $\delta(s)$ is the unobserved regression error at location $s$. 
Geostatistical model

- We assume that the error process $\delta(\cdot)$ is a stationary, isotropic Gaussian process with mean zero and covariance function

$$\text{Cov}(\delta(u), \delta(v)) = \sigma^2 \rho(d; \theta),$$

where

- $\sigma^2$ is the variance of the process,
- $d = ||u - v||$ is the Euclidean distance between locations $u$ and $v$,
- $\rho(\cdot; \theta)$ is an isotropic correlation function parameterized by the $k$-vector $\theta$. 
Parameter estimation

- Assuming no measurement error, i.e., $\tau^2 = 0$, the log-likelihood of the parameters $(\theta, \beta, \sigma^2)$ given the data, $Y$, is

$$\log L (\beta, \theta, \sigma^2; Y) \propto \frac{1}{2} \log |\sigma^2 \Gamma| + \frac{1}{2\sigma^2} (Y - X\beta)' \Gamma^{-1} (Y - X\beta),$$

where $\Gamma = [\rho(d; \theta)]$ represents the matrix of correlations between all pairs of observations.

- Parameter estimation can proceed via an iterative maximum (profile) likelihood approach.

- The resulting log profile likelihood is

$$\ell_{\text{profile}}(\theta; \hat{\beta}(\theta), \hat{\sigma}^2(\theta), Y) \propto \frac{1}{2} \log |\Gamma(\theta)| + \frac{n}{2} \log (\hat{\sigma}^2(\theta)), $$

where

$$\hat{\beta}(\theta) = (X' \Gamma^{-1}(\theta) X)^{-1} X' \Gamma^{-1}(\theta) Y,$$

$$\hat{\sigma}^2(\theta) = \frac{1}{n} (Y - X\hat{\beta}(\theta))' \Gamma^{-1}(\theta) (Y - X\hat{\beta}(\theta)).$$
Model selection: Spatial $AIC_c$

- Derivation of the AIC statistic in the spatial context, i.e., incorporate potential correlation of the residuals during model selection.

- Suppose
  1. $Z \sim f_T$, and
  2. $\{f(\cdot; \psi), \psi \in \Psi\}$ is a family of candidate probability density functions.

- The Kullback-Leibler information between $f(\cdot; \psi)$ and $f_T$ is defined by

$$I(\psi) = \int -2 \log \left( \frac{f(z; \psi)}{f_T(z)} \right) f_T(z) dz.$$ 

  – Measures the distance between $f(\cdot; \psi)$ and $f_T$.
  – Quantifies the loss of information when $f(\cdot; \psi)$ is used as the model for the data instead of $f_T$.

- The quantity

$$AIC_c = -2 \log L(\hat{\beta}, \hat{\theta}, \hat{\sigma}^2; Y) + 2(p + k + 1) \frac{n}{n - p - k - 2},$$

is an approximately unbiased estimate of the expected Kullback-Leibler information.
Motivation and roadmap for the remainder of the presentation

- The likelihood in the spatial $AIC_c$ statistic partially motivated the discussion to follow.
  - The proof requires standard asymptotic assumptions. Do these assumptions hold?
  - Maximum likelihood estimation is important of its own accord; is the normal distribution a good approximation (asymptotically) to the distribution of the MLE of the model parameters?

- Roadmap
  1. Continuous-parameter Gaussian process with exponential covariance function
     (continuous AR(1))
  2. Sampling designs for one-dimension, i.e., sampling along a transect
  3. Asymptotic limit theory distribution for estimates of the range parameter (exponential class)
     (a) Limiting results for equispaced sampling locations
     (b) Weighted least squares estimation
     (c) Maximum likelihood estimation
     (d) Empirical results corroborating theoretical results
  4. Empirical results for the Matérn class in one- and two-dimensions
Ornstein-Uhlenbeck process in one-dimension

- The continuous analogue to the discrete-time AR(1) process is defined as the solution to the stochastic differential equation

\[ dY_t = -\tau Y_t dt + \sigma \sqrt{2\tau} dB_t, \quad \tau > 0, \]

where

- \( t \) represents time (or space),
- \( \tau \) is the range parameter, and
- \( B_t \) is a standard Brownian motion.

- The solution to this equation, known as the Ornstein-Uhlenbeck process, is

\[ Y_t = e^{-\tau(t-s)} Y_s + \sigma \sqrt{2\tau} \int_s^t e^{-\tau(t-u)} dB_u \quad \text{for } s < t. \]

- The \( \{Y_{t_i}\} \) is the continuous AR(1) process that satisfies the recursions

\[ Y_{t_i} = e^{-|t_i-t_{i-1}|/\theta} Y_{t_{i-1}} + \varepsilon_{t_i}, \quad \varepsilon_{t_i} \sim \text{iid } \mathcal{N}\left(0, \sigma^2(1 - e^{-2|t_i-t_{i-1}|/\theta})\right). \]  

  - Note \( \theta = \tau^{-1} \) to coincide with the spatial literature (exponential correlation function).
Sampling along a transect

- Assume the transect is $m$ units long.
- Let $n$ be the desired sampling effort per unit length (block).
- To generate the $mn$-vector of sampling locations $t$:
  1. Define $f_T(\cdot) > 0$ a.e. on the interval $(0, 1]$, e.g., $t \sim \text{unif}(0, 1)$.
  2. Generate $m$ simple random samples from $f_T(\cdot)$ each of size $n$ such that $t_i = (i - 1)1 + (t_{i,(1)}, \ldots, t_{i,(n)})'$
  3. Define $t = \{t_1, \ldots, t_m\}'$.
- Special case: set $\Delta t \equiv n^{-1}$, i.e., regular, fixed intervals (equispaced sampling locations).
- One can generalize this procedure so that the number of sampling locations per block is random, i.e., $n_i \sim f_N(\cdot)$ and $t \sim f_T(\cdot; n_i)$.
  - One can construct “clustered” sampling designs in this manner.
Examples of sampling along a transect

- Suppose that the “true” response curve is as illustrated ($\theta = 1$).
Sampling at regular, fixed intervals

- Generate the sampling locations using a regular, fixed interval with $n = 2$, i.e., $\Delta t \equiv 1/2$. 

---

**Ph.D. Defense Seminar**

**October 2006**
Sampling at regular, fixed intervals

- At each sampling location record the value of the response.
Sampling at regular, fixed intervals

- The observed response curve for $mn = 16$ equispaced sampling locations ($m = 8$, $n = 2$).
Sampling at random intervals

- Generate $n = 2$ sampling locations per block where $f(t) \sim \text{unif}(0, 1)$. 

![Sampling locations along transect](image)
Sampling at random intervals

- At each sampling location record the value of the response.

Observed response at each sampling location
Sampling at random intervals

- The observed response curve for $mn = 16$ uniformly spaced sampling locations.
Asymptotic distribution of the MLE for $\theta$ for equispaced sampling

- If sampling at regular, fixed intervals, i.e., $\Delta t \equiv n^{-1}$, then the maximum of the profile likelihood function, conditioned on $Y_{t_0} = 0$, can be written in closed form.

$$\hat{\phi}_n = \frac{\sum_{i=1}^{mn} Y_{t_i} Y_{t_{i-1}}}{\sum_{i=1}^{mn} Y_{t_{i-1}}^2}$$

where $\phi_n = \exp\{-1/(\theta n)\}$.

- **Theorem 1**
  Let $Y$ be an $mn$-vector of observations sampled from the continuous AR(1) process as defined by (1). Define $\hat{\theta} = -(n \log(\hat{\phi}_n))^{-1}$. Then,

1. for fixed $n$ and $m \to \infty$,

$$\sqrt{m} (\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N} \left(0, \theta_0^4 n (e^{2/(\theta_0 n)} - 1)\right),$$

2. for $n \to \infty$ as $m \to \infty$,

$$\sqrt{m} (\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N} \left(0, 2\theta_0^3\right),$$

where $\theta_0$ is the “true” value of the range parameter.
Weighted least squares approach

- For notational convenience define the “spacings” between subsequent sampling locations by

\[
\delta_{i,j} = \begin{cases} 
  t_{i,(j)} - t_{i,(j-1)} & \text{for } j > 1, \\
  t_{i,(1)} - t_{i-1,(n)} & \text{for } i > 1, j = 1, \\
  \infty & \text{for } i = 1, j = 1, \text{ (i.e., no } Y_{t_0}). 
\end{cases}
\]

- The \(\delta_i = (\delta_{i,1}, \ldots, \delta_{i,n})'\) are identically distributed for all \(i \geq 2\).
- Define \(\delta\) as a randomly selected spacing from block \(i \geq 2\) such that \(\delta = \delta_{i,j}\) with probability \(1/n\).
- Define \(E[f(\delta)] = n^{-1} \sum_{j=1}^{n} E[f(\delta_j)]\).

- For a WLS estimator of \(\theta\), define the objective function as

\[
S(\phi) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( Y_{t_{i,j}} - \phi^{\delta_{i,j}} Y_{t_{i,j-1}} \right)^2 / \delta_{i,j} \text{ where } \phi = \exp\{-1/\theta\}.
\]

- The form of the objective function is a simplification of the likelihood function.
- Note that subsequent locations at close proximity are more heavily weighted.
- Techniques used to derive the asymptotic distribution of the WLS estimator are recycled for the derivation of asymptotic distribution for the MLE.
Asymptotic distribution of a WLS estimator for $\theta$ for $t \sim f_T(\cdot)$

- **Theorem 2**
  Let $\delta$ be the $mn$-vector of spacings between subsequent sampling locations and let $\delta_i$ correspond to a randomly selected spacing from block $i \geq 2$ as described above. Let $Y$ be the $mn$-vector of observations sampled from the continuous AR(1) process as defined by (1).
  Define $\hat{\theta} = - (\log(\hat{\phi}))^{-1}$ where $\hat{\phi} = \arg \min_{\phi \in (0,1)} S(\phi)$. Then,

1. for fixed $n$ and $m \to \infty$,
   \[
   \sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{E[\phi_0^2(1 - \phi_0^2)]}{(E[\phi_0^2 \delta])^2}\right)
   \]
   where $\phi_0 = \exp\{-1/\theta_0\}$, and

2. for $n \to \infty$ as $m \to \infty$,
   \[
   \sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, 2\theta_0^3\right).
   \]

- We recover the result for equispaced sampling locations when $\delta_{i,j} \equiv n^{-1}$.

- Theorem 2 suggests that the “optimal” sampling design is to use a regular, fixed interval between sampling locations.
Asymptotic variance for the WLS estimate as a function of infill

- Illustration of the asymptotic variance for the WLS estimate for each sampling method considered as a function of sampling effort per block ($\theta = 1$).

- Note that the regular sampling pattern uniformly has the smallest expected variance with respect to $n$.
Maximum likelihood approach

- Brockwell and Davis (1996) show that the profile likelihood function can be expressed in terms of the one-step prediction errors, \( Y_{tk} - \hat{Y}_{tk} \), and their variances \( \nu_{tk-1} \) \((k = (i - 1)n + j)\). Hence,

1. \( \det \Gamma = \prod_{k=1}^{mn} \nu_{tk-1} \)
2. \( Y' \Gamma^{-1} Y = \sum_{k=1}^{mn} (Y_{tk} - \hat{Y}_{tk})^2 / \nu_{tk-1} \),

where \( \hat{Y}_{tk} = \phi^{\delta_{tk}} Y_{tk-1} \) and \( \nu_{tk-1} = 1 - \phi^{2\delta_{tk}} \)

- The resulting reduced likelihood function is

\[
\ell(\phi) = \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{j=1}^{n} \log(1 - \phi^{2\delta_{i,j}}) \right] \\
+ \log \left( \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{j=1}^{n} \left( Y_{ti,j} - \phi^{\delta_{i,j}} Y_{ti,j-1} \right)^2 / (1 - \phi^{2\delta_{i,j}}) \right] \right)
\]

Minimizing (2) with respect to \( \phi \) is equivalent to maximizing the likelihood function with respect to \( \theta \).
Asymptotic distribution of the MLE for $\theta$ for $t \sim f_T(\cdot)$

- **Theorem 3**

Let $\delta$ be the $mn$-vector of spacings between subsequent sampling locations and let $\delta$ correspond to a randomly selected spacing from block $i \geq 2$ as described above. Let $Y$ be the $mn$-vector of observations sampled from the continuous AR(1) process as defined by (1).

Define $\hat{\theta} = -(\log(\hat{\phi}))^{-1}$ where $\hat{\phi}$ minimizes (2). Then,

1. for fixed $n$ and $m \to \infty$,

$$\sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\theta_0^4}{n}(2\text{Var}[\delta \phi_0^{2\delta}/(1 - \phi_0^{2\delta})] + \text{E}[\delta^2 \phi_0^{2\delta}/(1 - \phi_0^{2\delta})])^{-1}\right),$$

where $\phi_0 = \exp\{-1/\theta_0\}$, and

2. for $n \to \infty$ as $m \to \infty$,

$$\sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, 2\theta_0^3).$$
Asymptotic variance for the MLE as a function of infill

- Illustration of the asymptotic variance for the MLE for each sampling method considered as a function of sampling effort per block ($\theta = 1$).

- Note that the trajectories for each sampling design are nearly identical.
Outline of proof for Theorem 3 for fixed $n$

1. Reparameterize and scale the reduced likelihood.
   (a) Let $\phi = \phi_0 + \gamma / \sqrt{m}$.
   (b) Scale the reduced likelihood function such that
       \[ \ell^*_m(\gamma) = m \left( \ell(\gamma / \sqrt{m}) - \ell(0) \right). \]

2. Show convergence in distribution of $\ell^*_m(\gamma)$ to $\ell^*(\gamma)$ for each $\gamma$ as $m \to \infty$.
   (a) Expand $\ell(\gamma / \sqrt{m})$ about zero using Taylor’s theorem.
   (b) Collect like terms and define new rv’s as functions of $\delta$, $Y$, and $\varepsilon$.
   (c) Determine the limiting distributions of the new rv’s as $m \to \infty$ where
       - $\delta$ is generated independently of $Y$,
       - the distribution of $Y$ is conditional on $\delta$, and
       - the $\delta_i$’s are strictly stationary and 1-dependent for all $i \geq 2$.

3. Extend pointwise convergence to “uniform” convergence on compact sets.

4. Minimize $\ell^*(\gamma)$ with respect to $\gamma = \sqrt{m}(\phi - \phi_0)$. 
Exponential simulation results in one-dimension

- Finite sample variance comparison with the asymptotic variance as a function of infill.
Exponential simulation results in one-dimension

- Surface plots of the mean square error (MSE) for four different sampling methods ($\theta = 2$).
Matérn class of autocorrelation functions

- The Matérn function has two parameters:
  - A range parameter $\theta_1$ which describes the behavior of the function at moderate to large distances.
  - A smoothness parameter $\theta_2$ which describes the behavior of the function at near-zero distances.

$$
\rho(d; \theta) = \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left( \frac{2d\sqrt{\theta_2}}{\theta_1} \right)^{\theta_2} K_{\theta_2} \left( \frac{2d\sqrt{\theta_2}}{\theta_1} \right), \quad \theta_1 > 0, \theta_2 > 0
$$

- The exponential class is a subset of the Matérn class: set $\theta_2 = 1/2$.
- In recent years the Matérn class has gained favor due to its “flexibility.”
Estimated correlation functions for Matérn(2,1)

- Estimated correlation functions for each of four sampling designs with \((m, n) = (32, 8)\).
  - Each of the \(r = 100\) realizations is denoted by a blue line.
  - The “true” correlation function is denoted by the black line.
Define the mean integrated square error (MISE) as

\[
MISE = \frac{1}{r} \sum_{i=1}^{r} \int_{0}^{\infty} \left( \rho(t; \hat{\theta}_i) - \rho(t; \theta_0) \right)^2 dt.
\]
Mean integrated squared error for the 2D Matérn(2,1) simulations

- MISE surface plots for the two-dimensional case
Concluding remarks

- For the exponential class in one-dimension
  - the finite sample variances for the MLEs are well approximated by the asymptotic variance,
  - the asymptotic variances converge at similar rates for each sampling design, and
  - the “optimal” sampling design is to define a regular, fixed interval between sampling locations.

- For the Matérn class the empirical results suggest that coverage of the domain is of primary concern but some sampling locations at close proximity is desirable.

- Future research includes
  - incorporating a nugget effect into the continuous AR(1) process, i.e., allow $\tau^2 > 0$, and
  - decrease computational time of the MLE for the Matérn class by restricting the smoothness parameter such that $\theta_2 = i + 1/2$ for $i = \{0, 1, \ldots\}$. 
Acknowledgements

- The work reported here was supported by STAR Research Assistance Agreements CR-829095 awarded by the U.S. Environmental Protection Agency to the Space-Time Aquatic Research Modeling and Analysis Program (STARMAP) at Colorado State University.

- I’d like to express my deepest gratitude to my thesis advisors, Jennifer Hoeting and Richard Davis. Their guidance and patience has been paramount in the completion of my degree and I sincerely hope to carry forward their “nuggets” of wisdom in my career.

- A special thank you to my committee members Scott Urquhart and Paul Doherty.

- Finally, and most especially, a whole-hearted thank you to my wife Melanie who’s support has been unwavering through all of my successes and failures over the past few years...