

Bayesian Modeling of Non-Stationary, Univariate, Spatial Data for the Earth Sciences

Chapter 24 of
Section C, Computer Programs
Book 7, Automated Data Processing and Computations

Techniques and Methods 7–C24

Bayesian Modeling of Non-Stationary, Univariate, Spatial Data for the Earth Sciences

By Karl J. Ellefsen and Bradley S. Van Gosen

Chapter 24 of
Section C, Computer Programs
Book 7, Automated Data Processing and Computations

Techniques and Methods 7–C24

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Ellefsen, K.J., and Van Gosen, B.S., 2020, Bayesian modeling of non-stationary, univariate, spatial data for the Earth sciences: U.S. Geological Survey Techniques and Methods, book 7, chap. C24, 20 p., <https://doi.org/10.3133/tm7C24>.

ISSN 2328-7055 (online)

Contents

Abstract.....	1
Introduction.....	1
Method.....	1
Background.....	1
Bayesian Hierarchical Model.....	2
Data Model.....	2
Process Model.....	3
Parameter Model.....	3
Bayes' Rule.....	4
Numerical Solution.....	5
Demonstration of the Method.....	5
Discussion.....	9
Conclusions.....	9
Acknowledgments.....	9
Data, Software, and Reproducibility.....	9
References Cited.....	10
Appendix 1. Checks of Statistical Model.....	12
Appendix 2. Sensitivity Analysis.....	16
Appendix 3. Covariance Function.....	19

Figures

1. Graph showing transformed titanium concentrations along a transect in the southeastern United States.....	2
2. Example of 9 basis functions for a linear domain.....	2
3. Map of sample locations, region of interest, and domain. Along the seacoast, the boundaries of the region of interest and the domain are congruent. The inset shows the location of the domain within the United States.....	5
4. Map of the centers of the basis functions. Along the seacoast, the boundaries of the region of interest and the domain are congruent.....	6
5. Graphs showing results of the 10-fold cross-validation.....	7
6. Maps showing, <i>A</i> , the process mean, <i>B</i> , the process standard deviation, <i>C</i> , the exceedance probability, and, <i>D</i> , the 0.95 quantile.....	8

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
kilometer (km)	0.6214	mile (mi)
kilometer (km)	0.5400	mile, nautical (nmi)

Bayesian Modeling of Non-Stationary, Univariate, Spatial Data for the Earth Sciences

By Karl J. Ellefsen and Bradley S. Van Gosen

Abstract

Some Earth science data, such as geochemical measurements of element concentrations, are non-stationary—the mean and the standard deviation vary spatially. It is important to estimate the spatial variations in both statistics because such information is indicative of geological and other Earth processes. To this end, an estimation method is formulated as a Bayesian hierarchical model. The method represents the spatially varying mean and the spatially varying standard deviation with basis functions; this formulation implicitly accounts for a spatially varying covariance function. A unique advantage of this method is that it can map the mean, the standard deviation, quantiles, and exceedance probabilities. The method is demonstrated by mapping titanium concentrations, which are measured in the coastal plain of the southeastern United States. Various checks demonstrate that the model fits the data and that the estimated statistics are geologically plausible.

Introduction

This work is motivated by a dataset consisting of 3,316 titanium concentrations that were measured in samples of stream sediments from the coastal plain in the southeastern United States. The complete dataset is described later in this report; for now, consider just those concentrations that are within 10 kilometers (km) of a transect (fig. 1). The concentrations have undergone a nonlinear transformation, which is described later in this report. Both the center and the spread of the transformed concentrations vary along the transect. These variations reflect changes in geological and geochemical processes occurring along the transect (Van Gosen and Ellefsen, 2018).

The center and the spread of the transformed concentrations may be quantified by, respectively, the mean and the standard deviation. Estimating both statistics is important because both provide information about the geological and geochemical processes that caused the current spatial distribution; however, estimating these statistics presents special challenges. First, because many Earth science datasets are large, estimating both statistics requires a lot of computation. Nonetheless, methods for statistical modeling of such large datasets have

been developed and were reviewed by Wikle (2010) and Banerjee and others (2015, p. 381–411). Second, the statistics change with location (fig. 1), which is a property called non-stationarity. Methods for statistical modeling of non-stationary datasets have been developed and were reviewed by Sampson (2010) and Banerjee and others (2015, p. 63–70).

One group of methods for statistical modeling of non-stationary datasets is called process convolution models (Sampson, 2010). With such models, the mean of a measured physical quantity may be estimated over the region in which the measurements were made (Higdon, 2002; Banerjee and others 2008; Lemos and Sansó, 2009). In this report, the process convolution model is extended to estimate the mean and the standard deviation. A similar extension already has been made for a completely different method of statistical modeling that involves stochastic partial differential equations (Ingebrigtsen and others, 2014).

This report is divided into three major sections. The Method section primarily describes the statistical model (which is a type of Bayesian hierarchical model) and its numerical solution. With this model, the spatially varying mean and the spatially varying standard deviation are estimated. The Demonstration of Method section shows how the method is used to map titanium concentrations in the southeastern United States. The Discussion section presents additional details regarding the method.

Method

Background

The data are measurements of a physical property, and all measurements are within a domain that may be represented by symbol D . The domain may be a line such as a transect, an area such as a geographic region, or a volume such as a mineral deposit.

The goal of the statistical modeling is to estimate, within domain D , the distribution of a spatially varying mean and the distribution of a spatially varying standard deviation. The method to represent the spatial variation must be flexible so that it can be applied to many different domains. Furthermore, this method must be computationally efficient so that statistical

2 Bayesian Modeling of Non-Stationary, Univariate, Spatial Data for the Earth Sciences

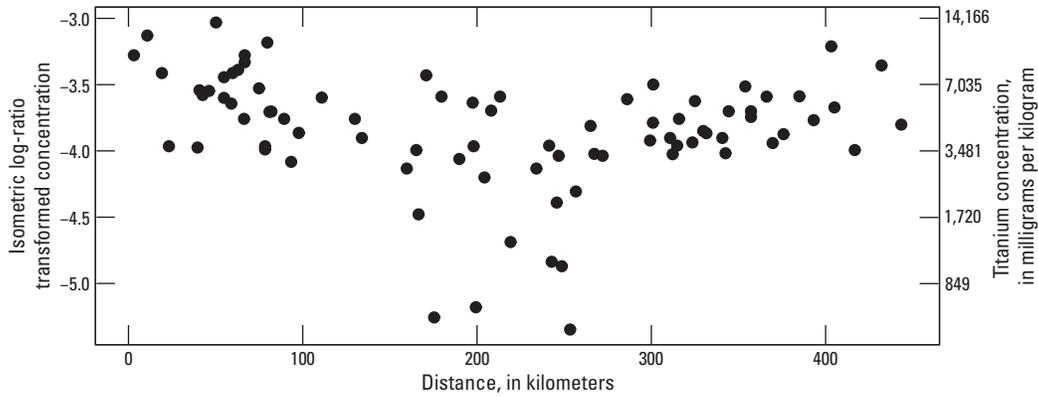


Figure 1. Graph showing transformed titanium concentrations along a transect in the southeastern United States.

modeling of large datasets is feasible. These two conditions are satisfied if the method uses smoothing kernels (Higdon, 2002). In practice, a smoothing kernel is simply a basis function, and this term is used in the rest of this report.

The key idea underlying basis functions is most easily explained for a linear domain. Within this domain, a set of overlapping functions is defined, as seen in figure 2. Each basis function is the local bi-square function (Cressie and Johannesson, 2008); this function was chosen because it is simple, and it is nonzero only in a small region around its center. (The later property can be exploited to significantly reduce the amount of computations and computer memory.) The nine functions have the same shape; they are just translated copies of one another. Appropriate weights for these functions are selected, and the weighted functions are added together to represent trends in the data.

Bayesian Hierarchical Model

Data Model

A helpful way to understand Bayesian hierarchical models for physical processes involves separating them into three submodels (Berliner, 1996; Cressie and Wikle, 2011, p. 21–23). The first is called the data model and is explained in this section. The second and third are called the process model and the parameter model respectively and are explained in subsequent sections.

Vector X consists of measurements of a physical property in domain D ; each element of X corresponds to a measurement at a different location. The dependence on location is omitted from the notation so that the notation is as easy to read as possible. The measurements are associated with errors that may be represented by a normal distribution:

$$\text{Normal}(X | Y, \text{diag}(\sigma^2)) \quad (1)$$

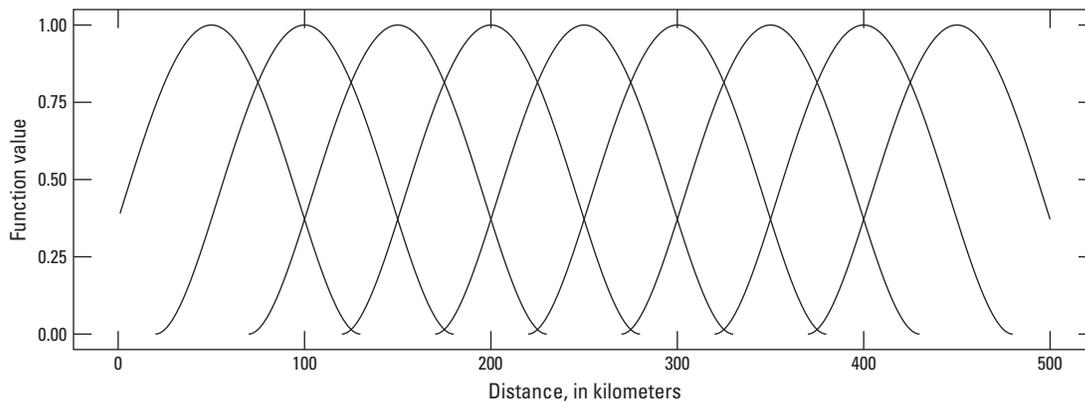


Figure 2. Graph showing an example of nine basis functions for a linear domain.

Vector Y represents the modeled physical property, and its dependence on location is omitted from the notation. Scalar σ represents the standard deviation associated with the measurement error. The standard deviation σ may be estimated, for example, from repeated measurements that are part of quality-control procedures. The notation $\text{diag}()$ indicates that its argument is converted to a diagonal matrix of the appropriate size.

Process Model

Vector Y , which represents the modeled physical property, is generated by a normal distribution:

$$\text{Normal}(Y | \mu \mathbf{1} + \mathbf{B}\phi, \text{diag}[(\rho \exp[\mathbf{B}\psi])^2]) \quad (2)$$

The mean of the distribution has two terms. In the first term, scalar μ represents the average value of the physical property in domain D . It is a constant and is estimated from the data X before the statistical modeling occurs. The symbol $\mathbf{1}$ represents a vector of ones. In the second term, the product $\mathbf{B}\phi$ represents smooth, spatially varying excursions from the average μ . Matrix \mathbf{B} consists of the values of the basis functions at the measurement locations, and vector ϕ consists of the weights for those basis functions.

The standard deviation $\rho \exp(\mathbf{B}\psi)$ has two parts. First, scalar ρ represents the average value of the standard deviation throughout domain D . Second, vector $\exp(\mathbf{B}\psi)$ represents smooth, spatially varying, positive-valued excursions from scalar ρ . The notation $\exp()$ indicates that exponentiation is applied to each element of the vector $\mathbf{B}\psi$. Vector ψ consists of the weights for the basis functions.

Parameter Model

An appropriate prior distribution for vector ϕ must satisfy two requirements. First, the excursions from the mean, which are specified by the product $\mathbf{B}\phi$, must be zero when averaged over domain D because that condition is inherent in equation 2. This requirement is satisfied approximately if the average of the elements in vector ϕ is approximately zero. Second, the weights for adjacent basis functions should be similar to one another so that the mean varies smoothly throughout domain D .

To satisfy these two requirements, vector ϕ is represented by a normal distribution:

$$\text{Normal}(\phi | \mathbf{0}, \Sigma_\phi) \quad (3)$$

The mean of the distribution is a vector of zeros $\mathbf{0}$, so the first requirement is satisfied. The covariance matrix Σ_ϕ is chosen so that weights for adjacent basis functions are similar to one another, a procedure that has been used by Lang and Brezger (2004). Thus, the second requirement is satisfied. The covariance matrix is formulated as a conditional autoregressive (CAR) model (Banerjee and others, 2015, p. 80–84). In a CAR

model, the covariance matrix is decomposed in the following manner:

$$\Sigma_\phi = (\tau_\phi^2 [D_\phi - \alpha_\phi W_\phi])^{-1} \quad (4)$$

The number of basis functions is denoted J , so the covariance matrix Σ_ϕ has dimension $J \times J$. Matrix W_ϕ , which also has dimension $J \times J$, is a proximity matrix (Banerjee and others, 2015, p. 74); it specifies the associations among weight ϕ_j and its neighbors. Matrix D_ϕ , which also has dimension $J \times J$, is diagonal; element j on the diagonal is the sum of the weights associated with weight ϕ_j . Scalar parameter α_ϕ , which is restricted to the interval $(0,1)$, ensures that the covariance matrix is nonsingular (Strang, 1988, p. 13–14). The closer that α_ϕ is to 1, the stronger the spatial association between weight ϕ_j and its neighbors. Scalar parameter τ_ϕ^2 represents precision; its reciprocal quantifies how a weight varies with respect to its neighboring weights but not the degree of spatial association among them.

A suitable prior distribution for parameter α_ϕ must satisfy two criteria. First, it must limit the range of α_ϕ to the interval $(0,1)$ (Banerjee and others, 2015, p. 82). Second, it must be adjustable so that α_ϕ has a high probability of being close to 1. Values close to 1 ensure that weight ϕ_j has a strong spatial association with its neighbors. The first criterion can be satisfied with the beta distribution:

$$\text{Beta}(\alpha_\phi | p_1, p_2) \quad (5)$$

Constants p_1 and p_2 control the shape of the distribution and must be positive. If $p_1 > p_2 > 1$, then the second criterion is satisfied. For example, suitable values might be $p_1 = 2.5$ and $p_2 = 1.2$.

A suitable prior distribution for parameter τ_ϕ^2 must satisfy two criteria. First, it must ensure that $\tau_\phi^2 > 0$ because negative precision is physically meaningless, and a precision of zero would correspond to infinite variability. Second, it must exclude small and large values because they correspond to excessively large and excessively small variances, respectively. (Small and large depend on the specific problem, so a general definition cannot be formulated.) The first criterion can be satisfied with the gamma distribution:

$$\text{Gamma}(\tau_\phi^2 | q_1, q_2) \quad (6)$$

Constants q_1 and q_2 control the shape of the distribution and must be positive. If $q_1 > 1$, then the second criterion is satisfied. For example, suitable values might be $q_1 = 2.0$ and $q_2 = 0.3$.

An appropriate distribution for vector ψ is formulated similarly. The equations are identical to equations 3, 4, 5, and 6, except that ϕ is replaced by ψ ; consequently, the equations are not repeated.

The prior distribution for standard deviation ρ must satisfy two criteria. First, it must permit only positive values for ρ . Second, it must permit ρ to have a wide range of values because usually little is known about ρ before the

statistical modeling. To satisfy these two requirements, ρ is chosen to have a truncated Cauchy distribution:

$$\text{Truncated Cauchy}(\rho | 0, s) \quad (7)$$

To understand this distribution, consider a Cauchy distribution with a center of zero and a scale parameter of s . Then, truncate the Cauchy distribution at zero, removing the negative-valued part. The resulting distribution ensures that the standard deviation has a positive value, which satisfies the first criterion. A suitable value for scale parameter s might be 3, which satisfies the second criterion.

Bayes' Rule

Bayes' rule is needed to infer the unknown parameters and can be formulated directly from equations 1 to 7. However, the numerical calculation of the posterior probability density function has severe problems because the function has regions of high curvature where the Monte Carlo sampler can be trapped. This problem can be overcome with a slight change in the formulation. To this end, equation 1 is considered from a different perspective. The measurement of a physical property X is represented by the equation $X=Y+\varepsilon$. Vector ε represents the measurement error; it has a normal distribution with mean zero and covariance matrix $\text{diag}(\sigma^2)$. Because vector Y has a normal distribution (eq. 2), the sum $Y+\varepsilon$ also has a normal distribution (Grimmett and Stirzaker, 2001, p. 114); thus, the measurements can be represented by the normal distribution:

$$\text{Normal}(X | \mu + B\phi, \text{diag}[\sigma^2 + (\rho \exp[B\psi])^2]) \quad (8)$$

With this slight change, Bayes' rule is formulated from equations 3 to 8:

$$p(\phi, \alpha_\phi, \tau_\phi^2, \psi, \alpha_\psi, \tau_\psi^2, \rho | y) \propto \prod_{i: I_i=1} \text{Normal}(X_i | \mu + B\phi, \sigma^2 + [\rho \exp(B\psi)]^2) \times$$

$$\prod_{i: I_i=0} \int_{-\infty}^{x_0} \text{Normal}(x_i | \mu + B\phi, \sigma^2 + [\rho \exp(B\psi)]^2) dx_i \times$$

$$\text{Normal}(\phi | 0, \Sigma_\phi) \times \text{Beta}(\alpha_\phi | p_1, p_2) \times \text{Gamma}(\tau_\phi^2 | q_1, q_2) \times$$

$$\text{Normal}(\psi | 0, \Sigma_\psi) \times \text{Beta}(\alpha_\psi | p_1, p_2) \times \text{Gamma}(\tau_\psi^2 | q_1, q_2) \times \text{TruncatedCauchy}(\rho | 0, s) \quad (9)$$

The expression on the left is the posterior probability density function for the model parameters. The first product on the right is that part of the likelihood function that accounts for those measurements that are not censored. The second product on the right is that part of the likelihood function that accounts for those measurements that are left censored at value x_0 . It is important to include this product because left-censored measurements are common in the Earth sciences. (There could be another product to account for measurements that are right censored. Because such measurements are rare in the Earth sciences, the product is omitted from equation 9.) These two products constitute the likelihood function. The remaining seven probability density functions on the right side are the prior probability density functions.

Numerical Solution

Equation 9 is coded in the Stan probabilistic programming language (Carpenter and others, 2017). The CAR model is coded in the manner proposed by Joseph (2016), which executes quickly and requires little computer memory because it accounts for sparsity in the matrices. Samples of the posterior probability density function are obtained using the Hamiltonian Monte Carlo method (Neal, 2011; Gelman and others, 2014, p. 300–305), which is implemented within Stan.

Convergence of the sampling is assessed by examining the parameter traces, the potential scale reduction factor, the effective number of independent simulation draws, and various other statistics (Gelman and others, 2014, p. 281–288). The effective numbers for α_ϕ , τ_ϕ^2 , α_ψ , and τ_ψ^2 usually are relatively small compared to the effective numbers for the other parameters. The reason is that these parameters appear relatively deep in the hierarchy, making them somewhat difficult to sample. These low effective numbers have no practical effect on inference.

Demonstration of the Method

To demonstrate the method, it is used to model titanium concentrations that were measured in 3,316 samples of stream sediments (Ellefsen, 2017). The samples were collected in the coastal plain of the southeastern United States, where the sediments are relatively rich in titanium-bearing minerals, namely ilmenite, rutile, and leucoxene. These minerals, as well as others, make these sediments an important economic resource.

The sample locations are scattered throughout the region of interest, except for a few small locales such as the southwestern corner of Georgia (fig. 3). The density of the samples varies a lot. It is relatively low in the northeast but relatively high in the southwest, especially near the center of Alabama. A margin was added to the region of interest because the margin improves the modeling near the boundary of the region of interest. Of course, a margin could not be added along the seacoast. The region of interest and the margin constitute the domain.

The titanium concentrations are an example of compositional data. Such data have special mathematical properties (Pawlowsky-Glahn and others, 2015, p. 8–31) that, if not accounted for, may adversely affect statistical inference (Aitchison, 2003, p. 48–63). A way to account for these properties is to transform the concentrations using the isometric log-ratio (ilr) transformation (Pawlowsky-Glahn and others, 2015, p. 36–38), which is a generalization of a logit transform. The resulting transformed concentrations do not have units.

Repeated measurements of titanium concentrations in two reference materials were made as part of the quality-control procedure for measuring element concentrations in the stream sediment samples. For each reference material, its titanium concentrations undergo the ilr transform. Analysis of these ilr-transformed concentrations indicates that their distribution may be represented by a normal distribution, which is important because this distribution represents the measurement error

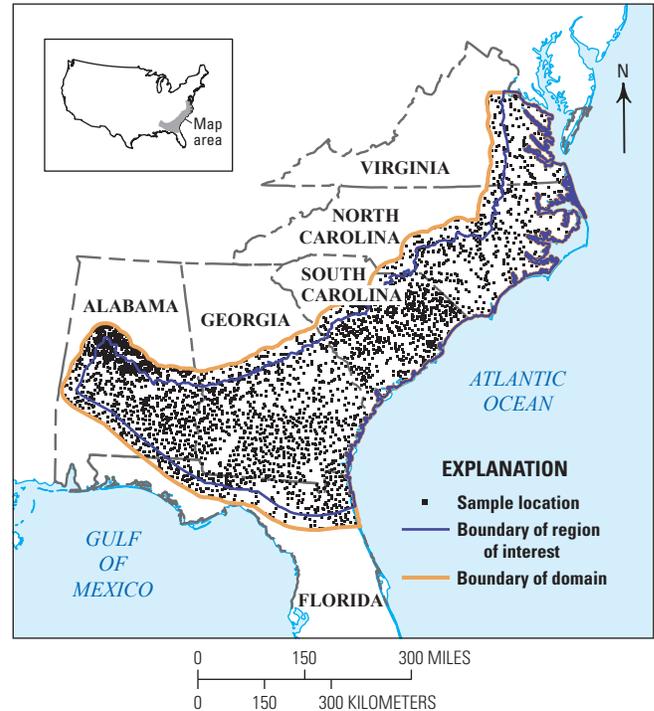


Figure 3. Map showing sample locations, region of interest, and domain in the coastal plain of the southeastern United States. Along the seacoast, the boundaries of the region of interest and the domain are congruent. The inset shows the location of the domain within the United States.

(eq. 1). The standard deviation of the distribution is 0.032 for one reference material and 0.031 for the other material; consequently, the standard deviation in the data model is chosen to be 0.032.

The basis function for a geographic region must be two-dimensional. The basis function—the local bisquared function—is nonzero only in a circular region, and the peak of the function is at the center of the circular region. For the specified domain (fig. 3), an example set of basis functions is shown in figure 4; only the locations of the basis function centers are plotted. The centers extend slightly beyond the domain. This feature improves the modeling near the domain boundary (Cressie and Kang, 2010, p. 49–63). (An exception is along part of the seashore of North Carolina where there are so few stream sediments samples that having centers outside the study area is infeasible.) The centers are chosen to be in a hexagonal pattern because this pattern requires fewer basis functions to represent the transformed concentrations than a rectangular pattern requires (Dudgeon and Mersereau, 1984, p. 44–47). Another reason is that the hexagonal pattern is more isotropic than a rectangular pattern is.

The next step in the statistical modeling is selecting the spacing between the basis functions. If the spacing is too small, then there will be many basis functions, and the statistical model will overfit the data. Conversely, if the spacing is too large,

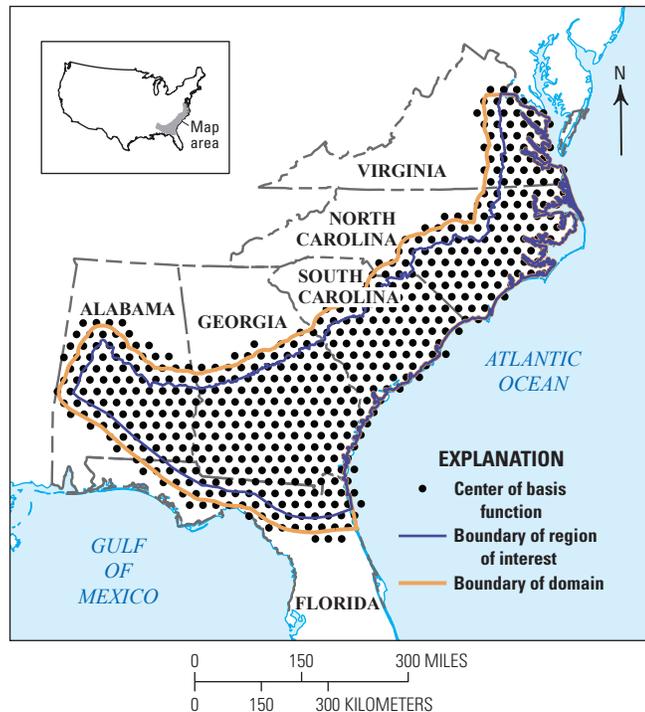


Figure 4. Map showing an example set of basis functions for the domain. Along the seacoast, the boundaries of the region of interest and the domain are congruent.

then there will be few basis functions, and the statistical model will underfit the data. For both situations, the consequences are the same: the statistical model will poorly predict the modeled physical property Y . That is, the resulting map will be inaccurate. So, selecting an appropriate spacing is important.

The spacing is selected using K -fold cross validation (Hastie and others, 2009, p. 241–245). The essential idea of this method is that a small fraction, say, 10 percent of the data are reserved, the remaining 90 percent of the data are used to estimate the statistical model, the estimated model is used to predict values for the reserved data, and finally the predicted values are compared to the actual values in the reserved data. The accuracy of the predicted values is measured by two different statistics: the deviance and the mean squared error (Gelman and others, 2014, p. 166–182; McElreath, 2016, p. 165–205). This procedure is repeated K times (namely, K fold), yielding K values of each statistic. K -fold cross validation is performed for many different spacings between the basis functions.

The results of 10-fold cross validation are shown in figure 5. The spacings between the basis functions centers range from 20 to 100 km; this range extends beyond the spatial scales of anomalies that are observed in transects. For both the deviance and the mean squared error, smaller values indicate higher predictive accuracy.

At each spacing, the 10 deviances vary a lot (fig. 5A). Consequently, it is difficult, perhaps impossible, to discern any trend in the deviances as the spacing increases from 20 to 100 km. To mitigate the effects of this variability, the 10 deviances at

each spacing are summarized by their mean and the associated standard error. The means for the deviance appear relatively small between 25 and 40 km. Similarly, the means for the mean squared error (fig. 5B) appear relatively small between 25 and 40 km. For both statistics, there is significant uncertainty in the interpreted relations because the changes between successive spacings are smaller than the standard errors. Because of this uncertainty, the cross validation is repeated using 20 folds, which reduces the variability in the statistics. The results, which are not shown, are practically the same. Consequently, the spacing between the basis functions is chosen to be 30 km, which is the very spacing shown in figure 4.

Sampling the posterior probability density function and assessing the convergence of the sampling are performed as described in the Numerical Solution section. The sampling consists of three chains of 1,500 samples each, so there are at total of 4,500 samples of the posterior probability density function. Various checks of the statistical model are presented in appendix 1.

To display the mean and the standard deviation of the process model as maps, many regularly spaced points are chosen within the domain. For this demonstration, the number is 10,000. There is nothing special about this number; other numbers could also yield suitable maps. The mean and the standard deviation of the process model must be predicted at these 10,000 points. To this end, the values of the basis functions at these prediction points are calculated and stored in a matrix B_p . For one sample of the posterior probability density function, the mean of the process model $\mu + B_p \phi$ is calculated at each of the 10,000 prediction points. Hence, one sample of the posterior probability density function yields one map of the mean of the process model. This procedure is repeated for all samples in the posterior probability density function, yielding 4,500 maps. Such a large number of maps cannot be analyzed by a human, so they are averaged, yielding a single map. The resulting map might be called the mean of the maps of the mean of the process model. Because this phrase is cumbersome, the resulting map is simply called the process mean. The procedure for the standard deviation of the process model is analogous, and the resulting map is called the process standard deviation. These maps are shown in figures 6A and B.

At each prediction point, the process mean and the process standard deviation are the parameters of a normal distribution that represents the process. This property allows alternative ways to communicate information about the process. One way involves the probability that the physical property exceeds a specified threshold; the probability of exceeding this threshold is called the exceedance probability. This exceedance probability is calculated for all prediction points, and the result is displayed as a map. For example, such a map is calculated for the titanium concentrations, using a threshold of 8,000 milligrams per kilogram (fig. 6C). Maps of exceedance probability are especially important in studies of contaminants, as researchers want to know where it is highly likely that the contaminant concentration exceeds some threshold.

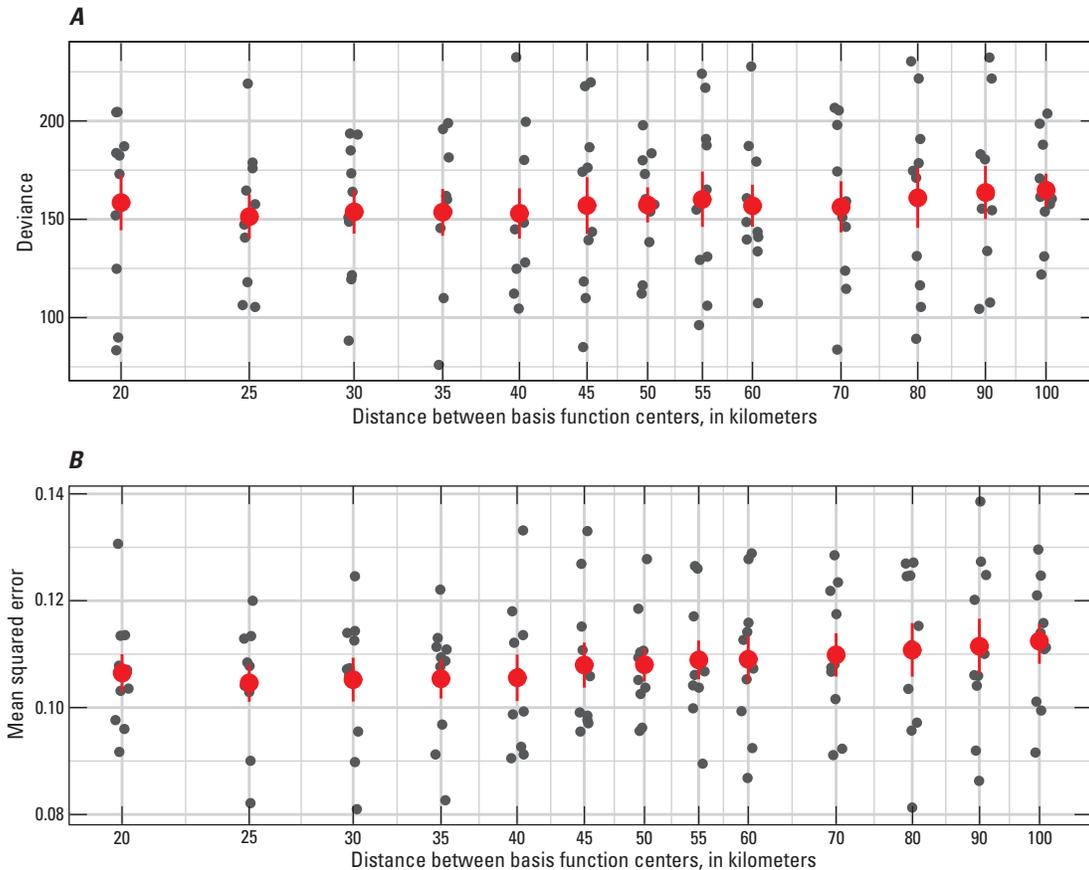


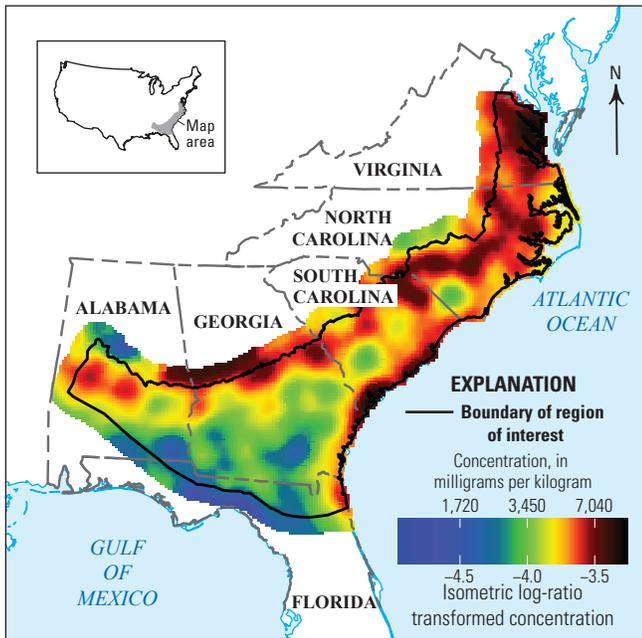
Figure 5. Graphs showing results of the 10-fold cross-validation. *A*, Results for the deviance. A gray dot represents a deviance for one cross validation test. (A small amount of random noise is added to the horizontal coordinate to reduce the number of symbols plotting atop one another.) A red dot represents the mean of the 10 deviances; a red line above (or below) a red dot represents the standard error of the mean. *B*, Results for the mean squared error. The definitions of the graph elements are analogous to those in *A*.

Another way to communicate information about the process involves the value of the physical property that is associated with a specified probability, say, 0.95. In the normal distribution that represents the process at a prediction point, this 0.95 probability is associated with a specific value of the physical property, which is called the 0.95 quantile. This 0.95 quantile is calculated for all prediction points, and the result is displayed as a map. For example, a 0.95 quantile map is calculated for the titanium concentrations (fig. 6*D*). Such maps communicate information about the right tail of the normal distribution that represents the process. Such information is especially important in studies of mineral

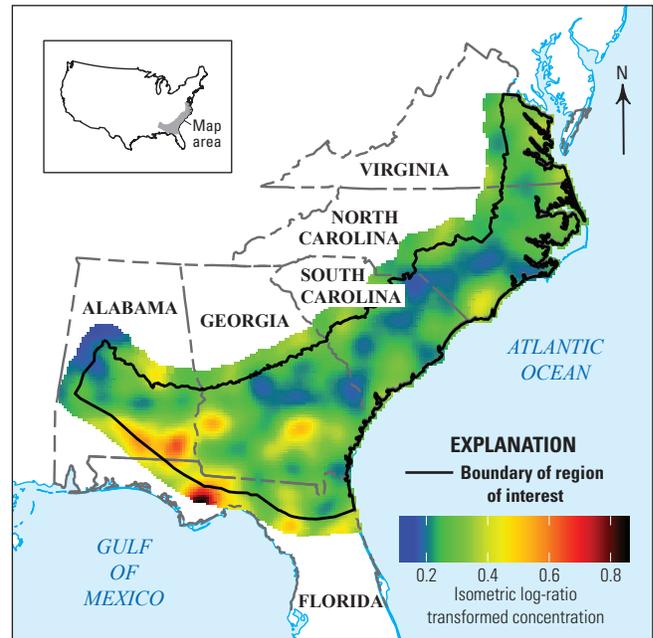
resources; economic geologists want to know where mineral resource concentrations are high.

Recall that the statistical modeling is performed on ilr-transformed concentrations. Consequently, the maps in figures 6*A*, *B*, and *D* are expressed in terms of these ilr-transformed concentrations. Because these ilr-transformed concentrations are difficult to interpret, scales for the process mean and the 0.95 quantile are re-expressed in terms of the equivalent titanium concentrations (figs. 6*A* and *D*). Unfortunately, there is no similar transformation for the process standard deviation (Pawlowsky-Glahn and others, 2015, p. 108–112).

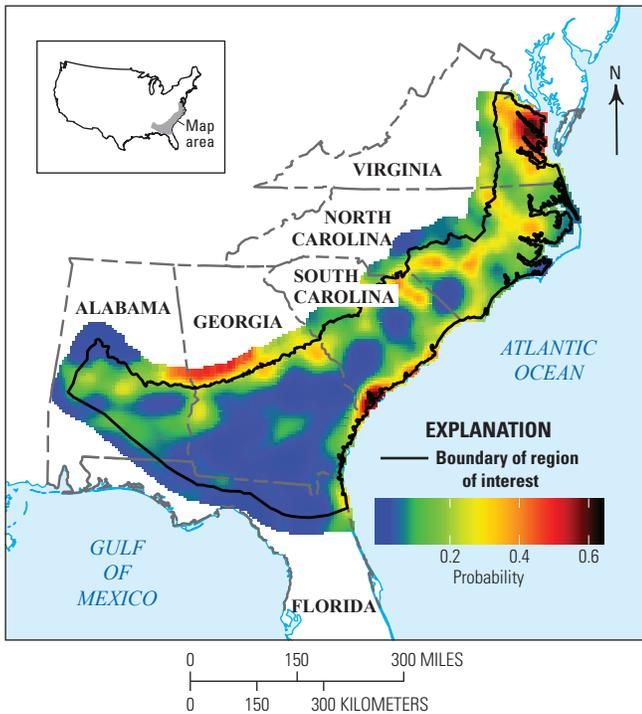
A. Process mean



B. Process standard deviation



C. Exceedance probability for 8,000 milligrams per kilogram



D. 0.95 quantile

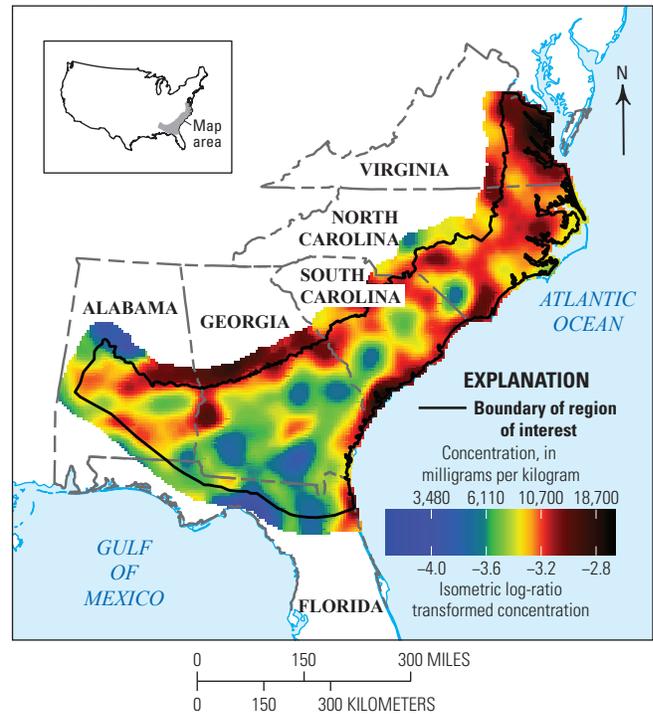


Figure 6. Maps showing, A, the process mean, B, the process standard deviation, C, the exceedance probability, and D, the 0.95 quantile.

Discussion

In the demonstration of the method, sample density varies over a wide range (fig. 3). Such variation strongly affects classical kriging techniques, so there are procedures to compensate for it (Issaks and Srivastava, 1989, p. 237–248); however, the effects are different for the method presented herein. Regions with high sample density strongly affect the weights of the basis functions that intersect those regions. Conversely, regions with low sample density weakly affect the weights of the basis functions that intersect those regions. These differences in effect are localized, and they do not affect other parts of the Bayesian hierarchical model. Consequently, there is no need to compensate for variations in sample density.

Recall that the probability density functions in the parameter model are characterized by constants $p_1, p_2, q_1, q_2,$ and s (eqs. 5, 6, and 7). These constants are chosen by the user based upon their understanding of the distributions that parameters $\alpha_\phi, \tau_\phi^2, \alpha_\psi, \tau_\psi^2,$ and ρ should have. Because these constants are chosen, not estimated from the data, the sensitivity of the model to these constants is a concern. This concern is addressed with a sensitivity analysis, which is presented in appendix 2. The analysis shows that, if reasonable values are assigned to the constants in the parameter model, then these constants have no discernible effect on the solution.

Several extant methods for modeling non-stationary data explicitly specify a spatially varying covariance function (Sampson, 2010; Banerjee and others, 2015, p. 63–70). Such a function was not specified in the formulation of the method. Nonetheless, the covariance function for the process model can be derived from the formulating equations (appendix 3). The covariance function is calculated at four different locations in the domain, and the results show that the covariance function varies spatially. Hence, although a spatially varying covariance is not explicitly specified in the statistical model, it is implicit in the model.

The basis function—the local bisquare function—is zero throughout most of the domain, so almost all elements in matrix B are zero. The number of zero-valued elements depends upon the number of measurements and the number of basis functions, so the percentage of zero-valued elements is different for every dataset. Nonetheless, the percentage frequently ranges from 97 to greater than 99. This sparsity is exploited in the computer implementation, which greatly decreases the computer time required to sample the posterior probability density function and greatly increases the amount of data that can be modeled. Regarding the latter advantage, a dataset for the conterminous U.S. with 64,368 measurements has been processed. An upper limit for the amount of data is currently unknown.

Conclusions

Non-stationary, univariate, spatial data are represented by a Bayesian hierarchical model that consists of three submodels. The first is the data model that represents the measurement procedure. The second is the process model that represents, throughout the domain, the spatial changes in the mean and the standard deviation of the data. This process model accounts for the non-stationary nature of the data, as well as a spatially varying covariance function. The third submodel is the parameter model that represents the parameters in the process model. All parameters in the Bayesian hierarchical model are estimated using the Hamiltonian Monte Carlo method.

This method is used to generate maps of the process mean, the process standard deviation, an exceedance probability, and a quantile. These maps provide Earth scientists with information about geological and other Earth processes. For example, maps of exceedance probability are especially helpful in studies of contaminants, and maps of quantiles are especially helpful in studies of mineral resources.

Acknowledgments

This work was funded by the Heavy-Mineral Sand Resources in the Southeastern U.S. project within the Mineral Resources Program of the U.S. Geological Survey. The manuscript was reviewed by M.A. Goldman and three anonymous reviewers; their suggestions improved the manuscript.

Data, Software, and Reproducibility

The titanium data that are used in this article are publicly available (Ellefsen, 2017). The numerical computations and figure generation for the main part of the report and appendix 1 are performed with software that is published in a user guide (Ellefsen and Van Gosen, 2019). This user guide presents step-by-step instructions for processing the titanium data. The numerical computations and figure generation for appendixes 2 and 3 are performed with software in file `SupplementaryMaterials.zip`, which accompanies this report. Before performing these numerical computations for appendixes 2 and 3, the scripts in the user guide must be executed so that several necessary data files are created. In summary, the publicly available data and the software can be used to reproduce the results in this article.

References Cited

- Aitchison, J., 2003, The statistical analysis of compositional data: Caldwell, N.J., The Blackburn Press, 416 p. and two supplements. [Originally published in 1986 by Chapman Hall, Inc.]
- Banerjee, S., Gelfand, A.E., Finley, A.O., and Sang, H., 2008, Gaussian predictive process models for large spatial data sets: *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, v. 70, no. 4, p. 825–848. [Also available at <https://doi.org/10.1111/j.1467-9868.2008.00663.x>.]
- Banerjee, S., Carlin, B.P., and Gelfand, A.E., 2015, Hierarchical modeling and analysis for spatial data (2d ed.): Boca Raton, CRC Press, 562 p.
- Berliner, L.M., 1996, Hierarchical Bayesian time-series models, in Hanson, K.M., and Silver, R.N., eds., *Maximum entropy and Bayesian methods*: Kluwer Academic Publishers, p. 15–22. [Also available at https://doi.org/10.1007/978-94-011-5430-7_3.]
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A., 2017, Stan—A probabilistic programming language: *Journal of Statistical Software*, v. 76, no. 1. [Also available at <https://doi.org/10.18637/jss.v076.i01>.]
- Cressie, N., and Johannesson, G., 2008, Fixed rank kriging for very large spatial data sets: *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, v. 70, no. 1, p. 209–226. [Also available at <https://doi.org/10.1111/j.1467-9868.2007.00633.x>.]
- Cressie, N., and Kang, E.L., 2010, High-resolution digital soil mapping—Kriging for very large datasets, in Viscarra Rossel, R.A., McBratney, A.B., and Minasny, B., eds., *Proximal soil sensing*: Heidelberg, Springer Science+Business Media B.V., p. 49–63.
- Cressie, N., and Wikle, C.K., 2011, *Statistics for spatio-temporal data*: Hoboken, John Wiley & Sons, Inc., 588 p.
- Dudgeon, D.E., and Mersereau, R.M., 1984, *Multidimensional signal processing*: Englewood Cliffs, N.J., Prentice-Hall, Inc., 400 p.
- Ellefsen, K.J., 2017, Titanium concentrations in stream sediments from the Atlantic Coastal Plain of the southeastern U.S. (1975-1999): U.S. Geological Survey data release, accessed September 24, 2018, at <https://doi.org/10.5066/F7J38R16>.
- Ellefsen, K.J., Goldman, M.A., and Van Gosen, B.S., 2020, User guide to the Bayesian modeling of non-stationary, univariate, spatial data using R-language package BMNUS: U.S. Geological Survey Techniques and Methods, book 7, chap. C20, 27 p., <https://doi.org/10.3133/tm7C20>.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., 2014, *Bayesian data analysis* (3d ed.): Boca Raton, CRC Press, 661 p.
- Grimmett, G.R., and Stirzaker, D.R., 2001, *Probability and random processes* (3d ed.): New York, Oxford University Press, 596 p.
- Hastie, T., Tibshirani, R., and Friedman, J., 2009, *The elements of statistical learning—Data mining, inference, and prediction*, 2nd ed.: New York, Springer Science+Business Media, 745 p.
- Higdon, D., 2002, Space and space-time modeling using process convolutions, in Anderson, C.W., Barnett, V., Chatwin, P.C., and El-Shaarawi, A.H., eds., *Quantitative methods for current environmental issues*: London, Springer, p. 37–56. [Also available at https://doi.org/10.1007/978-1-4471-0657-9_2.]
- Ingebrigtsen, R., Lindgren, F., and Steinsland, I., 2014, Spatial models with explanatory variables in the dependence structure: *Spatial Statistics*, v. 8, p. 20–38. [Also available at <https://doi.org/10.1016/j.spasta.2013.06.002>.]
- Issaks, E.H., and Srivastava, R.M., 1989, *Applied geostatistics*: New York, Oxford University Press, 561 p.
- Joseph, M., 2016, Exact sparse CAR models in Stan: Stan web page, accessed September 24, 2018, at <http://mc-stan.org/documentation/case-studies/mbjoseph-CARStan.html>.
- Lang, S., and Brezger, A., 2004, Bayesian p-splines: *Journal of Computational and Graphical Statistics*, v. 13, no. 1, p. 183–212. [Also available at <https://doi.org/10.1198/1061860043010>.]
- Lemos, R.T., and Sansó, B., 2009, A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature: *Journal of the American Statistical Association*, v. 104, no. 485, p. 5–18. [Also available at <https://doi.org/10.1198/jasa.2009.0018>.]
- McElreath, R., 2016, *Statistical rethinking—A Bayesian course with examples in R and Stan*: Boca Raton, CRC Press, 469 p.
- Neal, R., 2011, MCMC for using Hamiltonian dynamics, in Brooks, S., Gelman, A., Jones, G., and Xiao-Li, M., eds., *Handbook of Markov Chain Monte Carlo*: Boca Raton, CRC Press, p. 113–162.
- Pawlowsky-Glahn, V., Egozcue, J.J., and Tolosana-Delgado, R., 2015, *Modeling and analysis of compositional data*: Chichester, England, John Wiley and Sons, Ltd., 247 p.
- Sampson, P.D., 2010, Constructions for nonstationary spatial processes, in Gelfand, A.E., Diggle, P.J., Fuentes, M., and Guttorp, P., eds., *Handbook of spatial statistics*: Boca Raton, CRC Press, p. 119–130.

- Strang, G., 1988, *Linear algebra and its applications* (3d ed.): New York, Harcourt Brace Jovanovich College Publishers, 505 p.
- Van Gosen, B.S, and Ellefsen, K.J., 2018, Titanium mineral resources in heavy-mineral sands in the Atlantic Coastal Plain of the southeastern United States: U.S. Geological Survey Scientific Investigations Report 2018–5045, 32 p., accessed September 28, 2018, at <https://doi.org/10.3133/sir20185045>.
- Wikle, C.K., 2010, Low-rank representations for spatial processes, *in* Gelfand, A.E., Diggle, P.J., Fuentes, M., and Guttorp, P., eds., *Handbook of spatial statistics*: Boca Raton, CRC Press, p. 107–118.

Appendix 1. Checks of Statistical Model

This appendix presents five different checks of the statistical model. The first check involves analyzing the fit between the model and the data along one or more transects within the domain. This check is particularly valuable because it reveals model properties that are difficult to discern in map view. In this appendix, this analysis is demonstrated for only one transect (fig. 1.1). The data that are projected into this transect are within 10 kilometers (km) of it and are shown in figure 1.2A. (These data are also shown in figure 1.)

The distributions of the mean of the process model, the standard deviation of the process model, and the predicted measurement are calculated at the locations of the data in the transect. Then, these distributions are projected into the transect too (fig. 1.2B, C, and D). The trend defined by the medians of the mean of the process model is similar to the trend in the transformed titanium concentrations (fig. 1.2A and B). Likewise, the trend defined by the medians of the standard deviation of the process model is similar to the trend in the variability of the transformed titanium concentrations (fig. 1.2A and C). For the mean and the standard deviation of the process model, the distribution spreads at both transect



Figure 1.1. Map showing the location of the transect for the first and second checks of the statistical model.

ends are relatively large because there are less data here to estimate the model parameters (fig. 1.2B and C).

The distribution for predicted measurement i is a normal distribution with mean $\mu + B\phi_i$ and with standard deviation $(\sigma^2 + [\rho \exp(B\psi_i)]^2)^{1/2}$. This distribution accounts for the mean of the process model, the standard deviation of the process model, and the measurement error—it analogous to the prediction interval in classical regression (DeGroot and Schervish, 2002, p. 641–642). The distributions for the predicted measurements are compared to the transformed concentrations in figure 1.2D. The trend defined by the medians is similar to the trend in the transformed concentrations. The distribution spreads at both transect ends are relatively large because there are less data here to estimate the model parameters. The distribution spreads near the middle are large because the data here are highly variable. Thus, the distributions for the predicted measurements satisfactorily represent the actual data.

A plot symbol for a distribution for a predicted measurement shows the 95-percent probability interval. So, if the distributions and their associated measurements are considered collectively, approximately 95 percent of the measurements should plot within the 95-percent probability intervals. Consequently, for the 82 measurements in figure 1.2D, it is expected that 78 will plot within this interval; actually, 77 measurements do. This finding further reinforces the conclusion that the distributions for the predicted measurements satisfactorily represent the actual measurements.

The second check of the statistical model also involves the transect: the observed isometric log-ratio (ilr)-transformed concentrations are compared to seven sets of ilr-transformed concentrations that are simulated with the statistical model (fig. 1.3). The observed and the simulated transformed concentrations appear similar, with one exception. The exception occurs between about 175 and 250 km—there is large scatter in the observed ilr-transformed concentrations but only moderately large scatter in the simulated ilr-transformed concentrations. The reason for this difference is that the basis functions constrain the spatial changes in both the mean and the standard deviation of the process model; consequently, they have difficulty representing the abrupt changes at 175 and 250 km.

The third check of the statistical model involves analyzing the standardized residuals for all data. The formula for the standardized residual is derived from equation 8. For measurement X_i , the standardized residual is

$$\frac{[X_i - E(\mu + B\phi_i)]}{\sqrt{\sigma^2 + E[\rho \exp(B\psi_i)]}} \quad (1-1)$$

for which the notation $E()$ denotes the expected value. Equation 8 indicates that the standardized residuals should

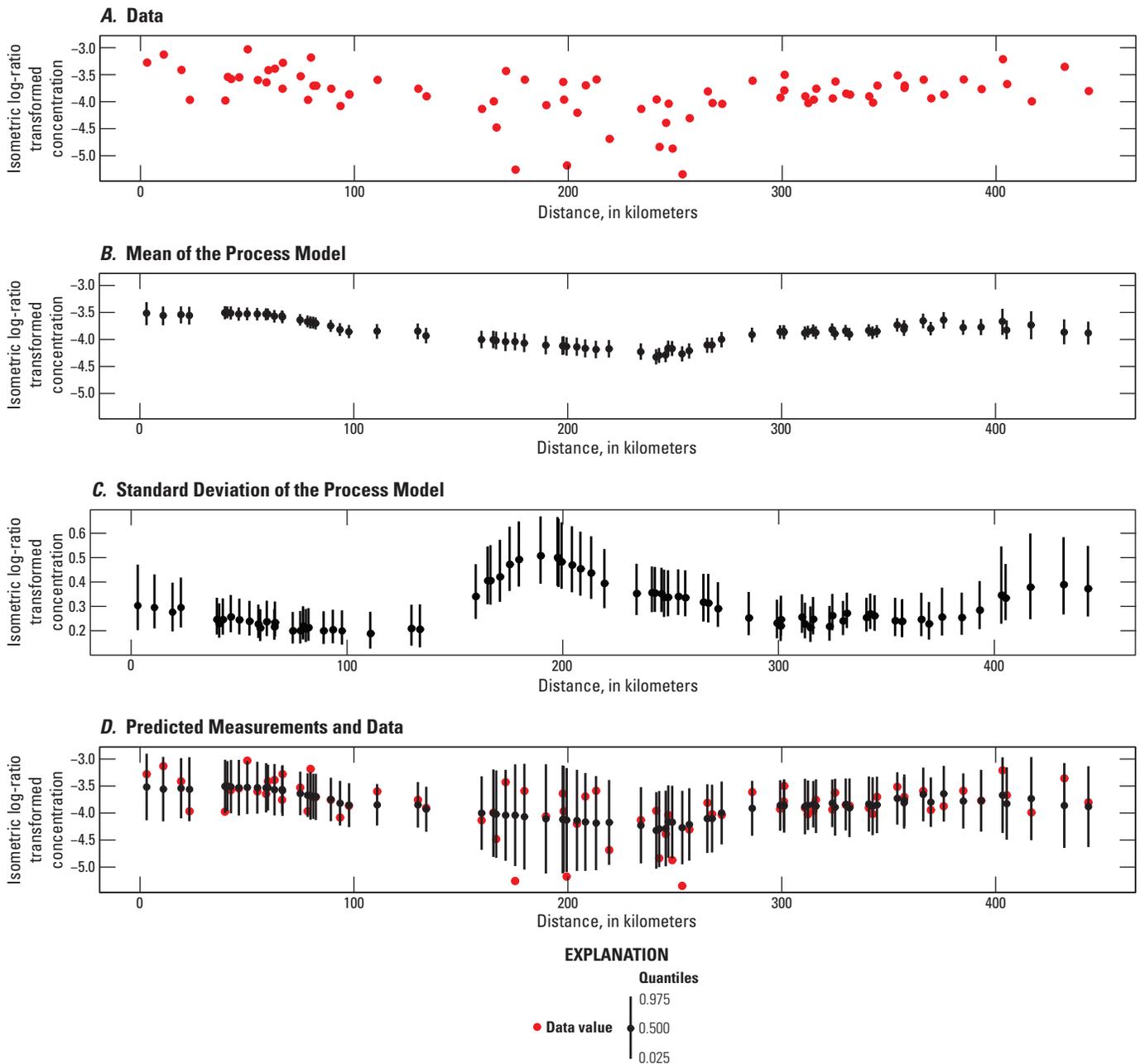


Figure 1.2. Graphs showing the model fit along the transect (fig. 1-1). *A*, isometric log-ratio transformed concentrations. *B*, distributions of the mean of the process model at the sample locations, *C*, distributions of the standard deviation of the process model at the sample locations, *D*, distributions of the predicted measurement at the sample locations, and the transformed concentrations.

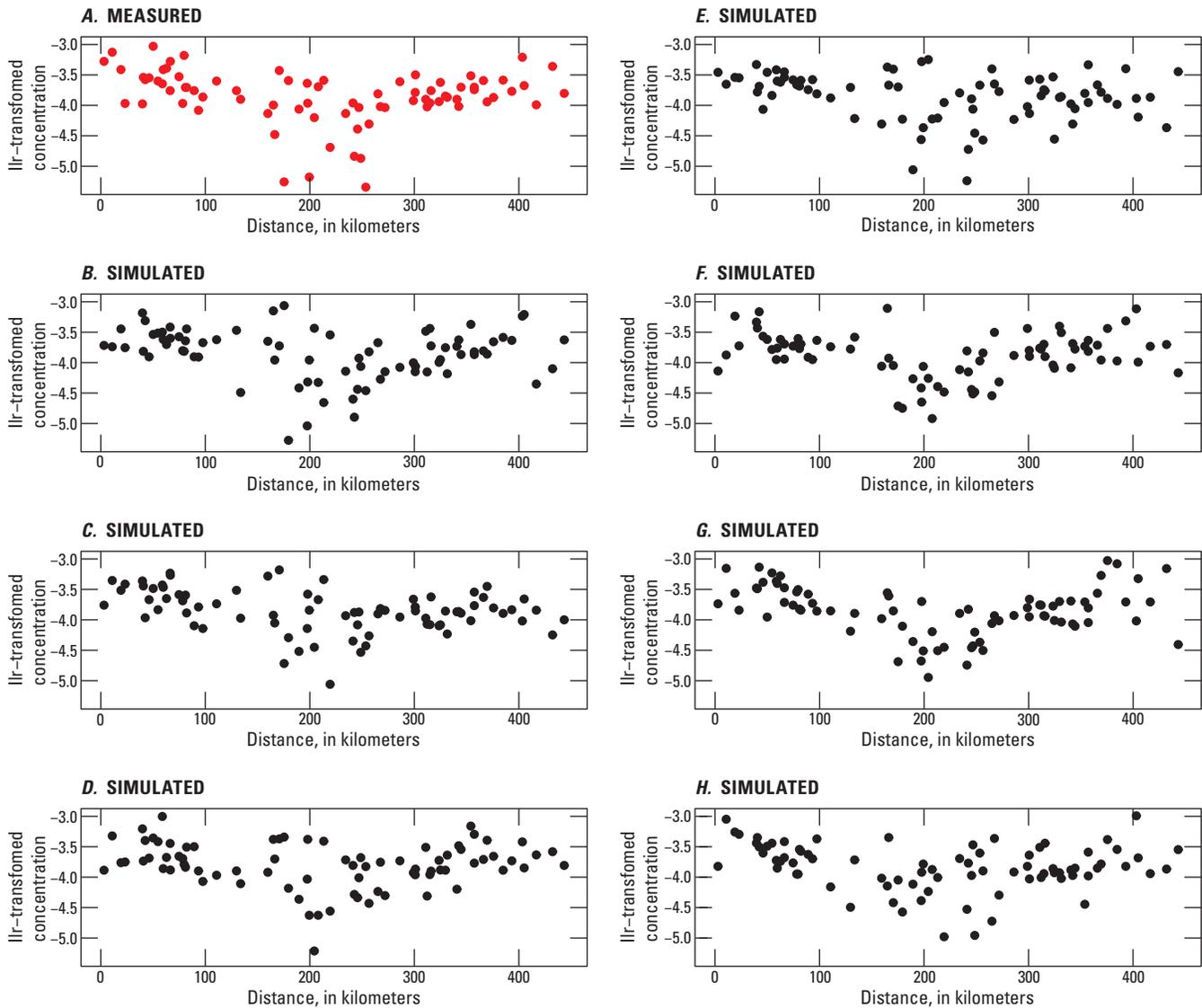


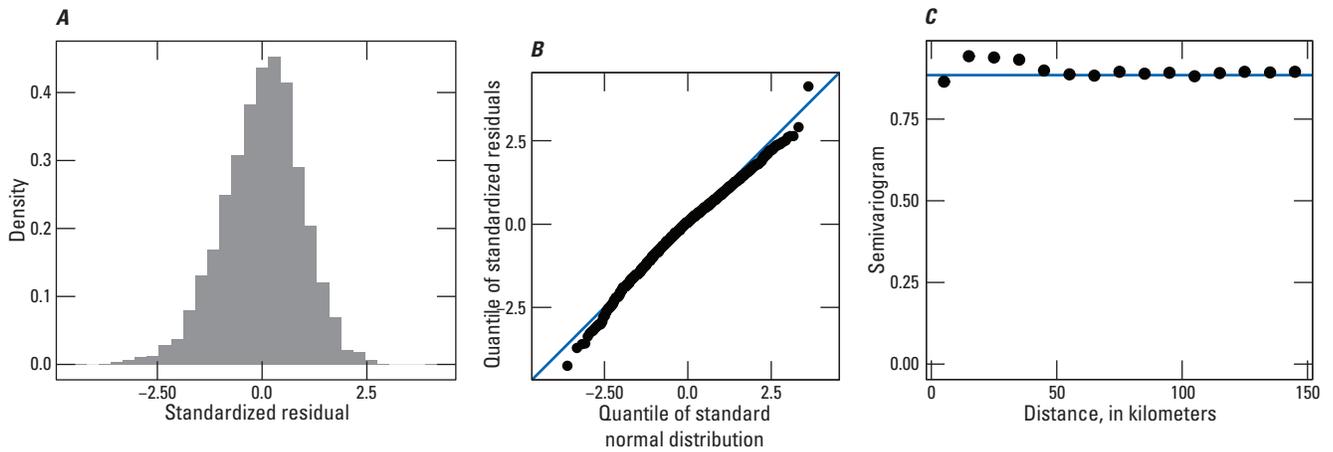
Figure 1.3. Graphs showing, *A*, observed and, *B–H*, simulated isometric log-ratio transformed concentrations along the transect (fig. 1-1).

have a standard normal distribution, which has a mean of 0, a standard deviation of 1, and a skew of 0. However, the actual distribution of the standardized residuals has a mean of -3.6×10^{-3} , a standard deviation of 0.94, and a skew of -0.34 (fig. 14*A*). The quantile-quantile plot, for which the theoretical distribution is a standard normal distribution, indicates that the actual distribution departs slightly from a standard normal distribution in its tails (fig. 14*B*). The interpretation of these findings is that the discrepancy between the standard normal distribution and the actual distribution is small. Thus, the statistical model is deemed appropriate for these data.

If the statistical model properly fits the data, then the standardized residuals should be spatially uncorrelated with one another. That is, a standardized residual for one field sample should be unrelated to the standardized residuals for nearby field samples. This property can be evaluated with a variogram. Because the standardized residuals have a variance of 0.88, the

variogram should equal 0.88 for all distances. The calculated variogram is close to this ideal for the distances specified in the variogram (fig. 14*C*). The interpretation of this finding is that the standardized residuals are uncorrelated with one another for the distances specified in the variogram. In practice, it is prudent to calculate variograms for different ranges.

For the fourth check of the statistical model, the arithmetic signs of the standardized residuals are mapped. That is, a map is generated showing the locations of those field samples for which the associated standardized residuals are positive-valued and similarly the locations of those field samples for which the associated standardized residuals are negative-valued (fig. 1-5). The map shows no large-scale clusters of positive-valued or negative-valued standardized residuals. The interpretation of this finding is that the statistical model adequately fits the data on a large-spatial scale. However, the map shows small-scale clusters. Comparison of figure 1.5 and figure 4 shows that the



EXPLANATION

In B, the blue line represents equality between the two quantile types
 In C, the blue line represents the variance of the standardized residuals

Figure 1.4. Graphs that are used to analyze the standardized residuals. A, histogram of the standardized residuals, B, quantile-quantile plot, and C, variogram of standardized residuals.

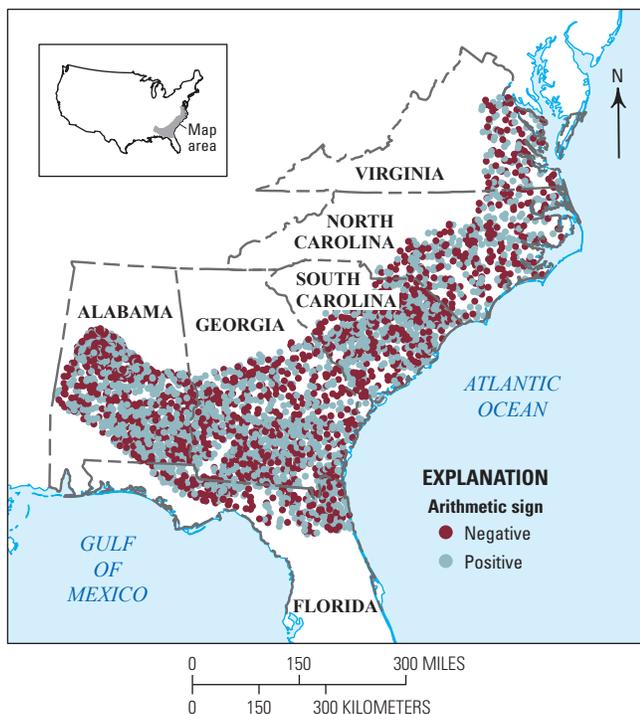


Figure 1.5. Map showing the arithmetic signs of the standardized residuals, which are plotted at the locations of the associated field samples.

sizes of the clusters roughly equal the spacing between the basis functions centers, namely 30 km. This spacing is the lower bound on any feature that cannot be resolved by the statistical model; consequently, standardized residuals that are less than 30 km apart could be spatially correlated with one another. Indeed, this phenomenon is observed in the map.

The fifth check of the statistical model, which is perhaps the most important check, involves interpretation of the maps in figure 6: The interpretation must be consistent with independent geologic information. Completely interpreting these maps is beyond the scope of this report, but it is appropriate to interpret a couple features. Along the seacoast in southeastern South Carolina, the process mean is high (fig. 6A). This anomaly can be explained by longshore currents that transported titanium-bearing minerals along the seacoast. These minerals are deposited inland during large storms such as hurricanes. In south-central Georgia, the process standard deviation is relatively high (fig. 6B), which can be explained by lateral changes in the sediments of the coastal plain. A more detailed interpretation of the maps is provided in Van Gosen and Ellefsen (2018).

Reference Cited

- DeGroot, M.H., and Schervish, M.J., 2002, Probability and statistics (3d ed.): New York, Addison Wesley, 816 p.
- Van Gosen, B.S, and Ellefsen, K.J., 2018, Titanium mineral resources in heavy-mineral sands in the Atlantic Coastal Plain of the southeastern United States: U.S. Geological Survey Scientific Investigations Report 2018–5045, 32 p., accessed September 28, 2018, at <https://doi.org/10.3133/sir20185045>.

Appendix 2. Sensitivity Analysis

In this appendix, the sensitivity of the model to changes in the constants p_1 , p_2 , q_1 , q_2 , and s is analyzed. The essential idea of this analysis is to systematically change each constant and then observe how this change affects the model parameters. To implement the analysis, the first step is to estimate the model parameters using the constant values that are specified in the Parameter Model section. To keep the sensitivity analysis tractable, it is restricted to the mean and the standard deviation of the process model, which are the most important model parameters. In addition, these two model parameters are analyzed at only four points (fig. 2.1). These four points are chosen because they are in areas with different sample densities (fig. 3) and are in different parts of the domain.

Consider point 1 (fig. 2.1). The distribution for the mean of the process model is shown in figure 2.2A in category None. (This category name indicates that none of the constants have been changed.) Likewise, the distribution for the standard deviation of the process model is shown in figure 2.2B in category None. The second step in the sensitivity analysis is to change just constant p_1 —its value is doubled. Then, the model parameters are estimated again. The distributions of the mean and standard deviation at point 1 are shown in figure 2.2A and 2.2B in category p_1 . Then, constant p_1 is returned to its original value. This step is repeated for the remaining four constants.

The six distributions for the mean of the process model (fig. 2.2A) are practically identical; thus, the changes in the constants have no practical effect on the mean of the process model at point 1. The findings for the standard deviation of the process model are the same (fig. 2.2B), and the inference is analogous. The sensitivity analyses for points 2, 3, and 4 are shown in figures 2.3, 2.4, and 2.5, respectively, and the inferences are the same.

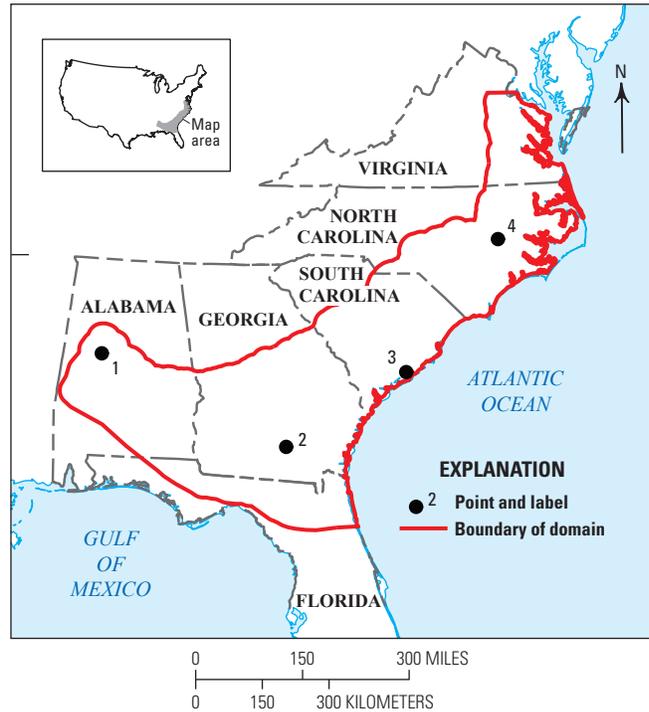


Figure 2.1. Map showing the locations of the four points in the domain, at which sensitivity analyses are performed and covariance functions are calculated (appendix 3).

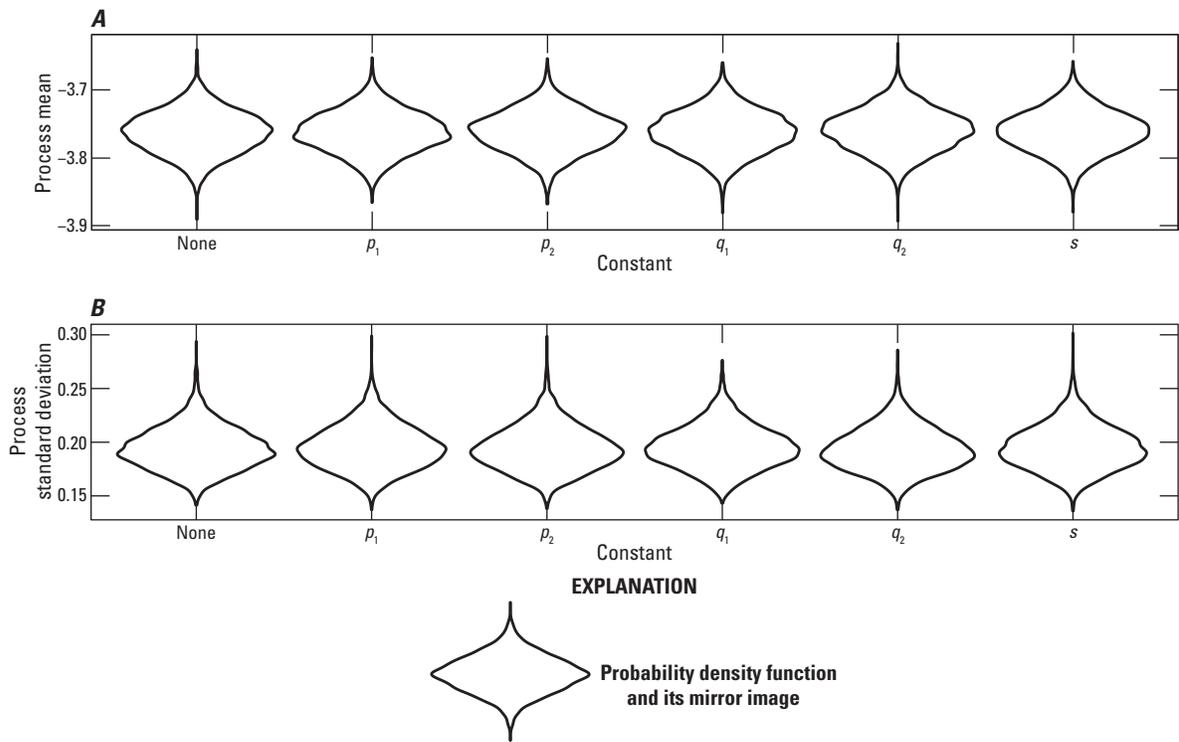


Figure 2.2. Graphs showing the sensitivity analysis for *A*, the process mean and *B*, the process standard deviation at point 1 (fig. 2.1).

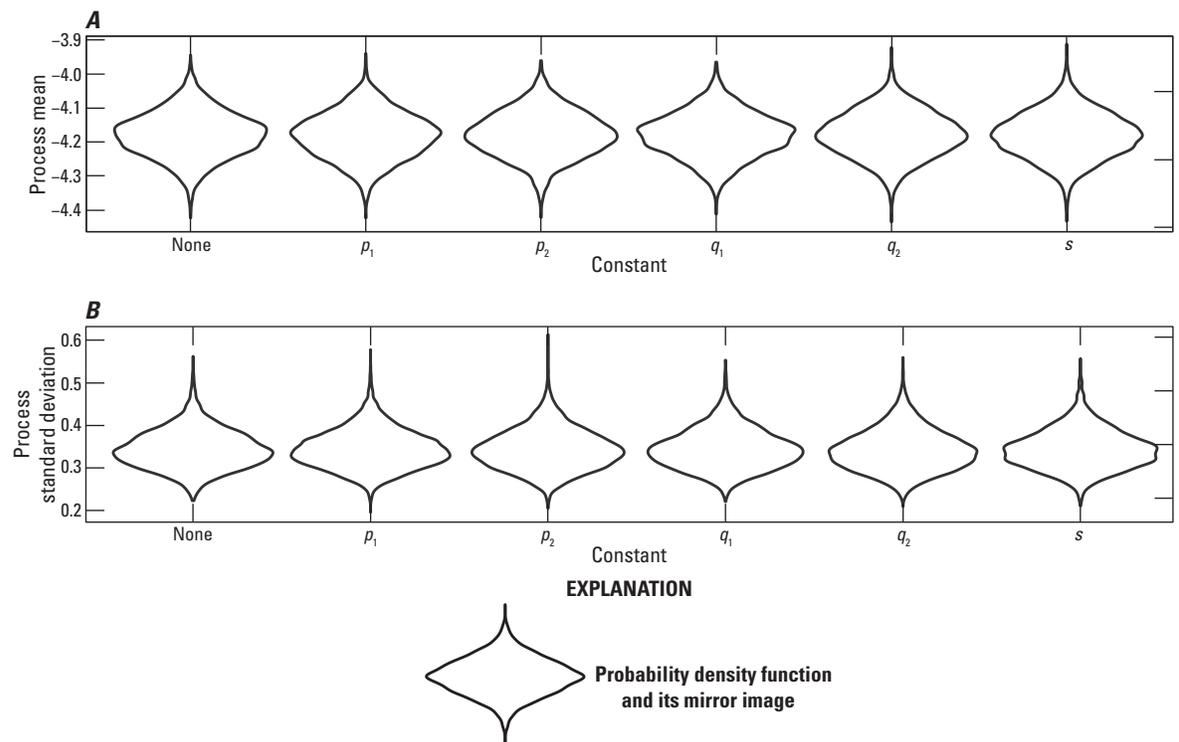


Figure 2.3. Graphs showing the sensitivity analysis for *A*, the process mean and *B*, the process standard deviation at point 2 (fig. 2.1).

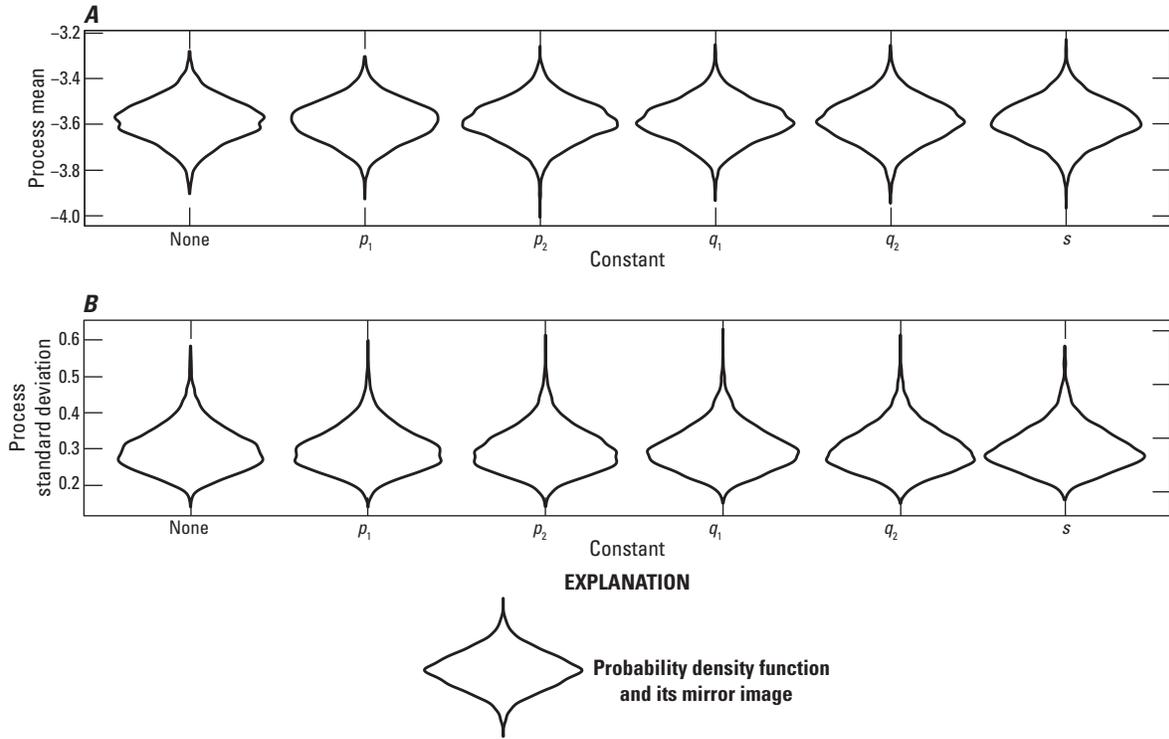


Figure 2.4. Graphs showing the sensitivity analysis for, *A*, the process mean and, *B*, the process standard deviation at point 3 (fig. 2.1).

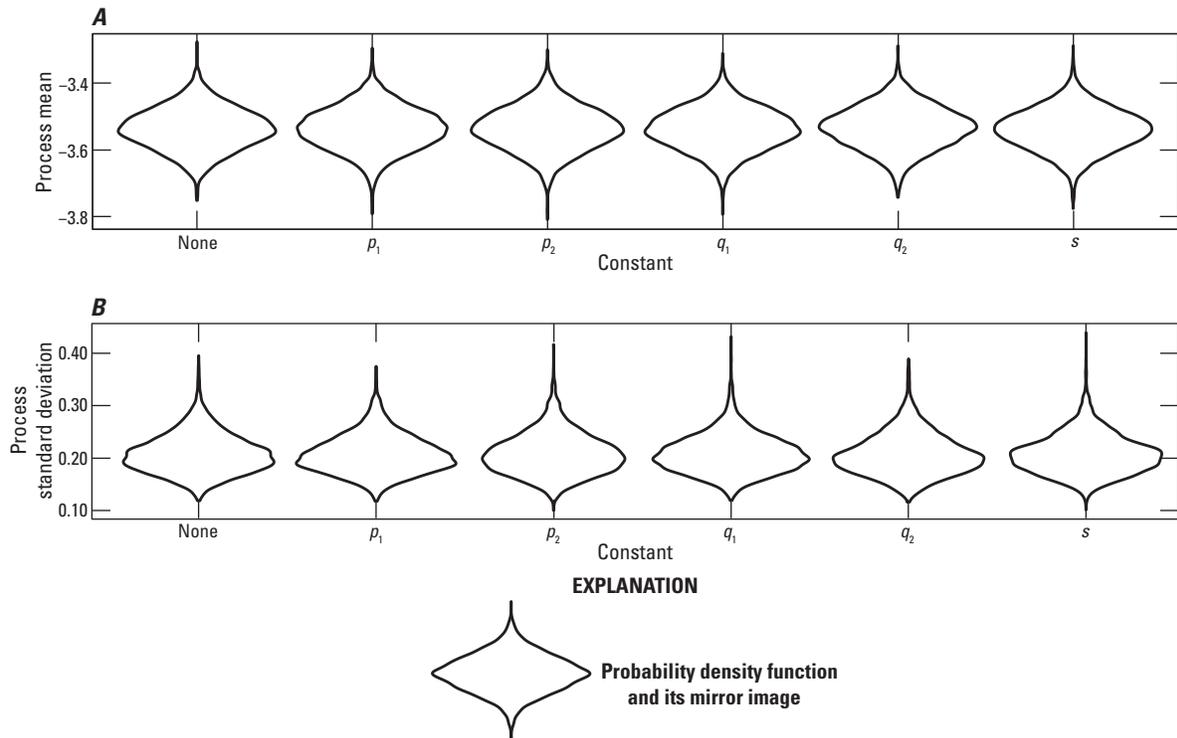


Figure 2.5. Graphs showing the sensitivity analysis for, *A*, the process mean and, *B*, the process standard deviation at point 4 (fig. 2.1)

Appendix 3. Covariance Function

In this appendix, the covariance function is derived for the process model. It is shown that the covariance function is inherited in the model, despite not being explicitly specified, and that the covariance function is spatially varying.

For the derivation, the process model (eq. 2) is re-expressed as:

$$Y = Y_1 + \zeta \quad (3.1)$$

The first term on the right side represents the mean of the process model, namely $\mu 1 + B\phi$ (see the “Process Model” section). The second term on the right side represents the variability in the process model, which is represented by a normal distribution with zero mean and a standard deviation of $\rho \exp(B\psi)$. Random variables Y_1 and ζ are assumed to be independent of one another. Consequently:

$$\text{Cov}(Y, Y) = \text{Cov}(Y_1, Y_1) + \text{Cov}(\zeta, \zeta) \quad (3.2)$$

The first term on the right side is the covariance associated with the mean of the process model; this covariance, expressed in terms of the model variables, is:

$$\text{Cov}(Y_1, Y_1) = B \text{Cov}(\phi, \phi) B^T \quad (3.3)$$

Interpreting matrix $\text{Cov}(Y_1, Y_1)$ is easier if it is decomposed:

$$\text{Cov}(Y_1, Y_1) = S \text{Cor}(Y_1, Y_1) S \quad (3.4)$$

Matrix S is square and diagonal. The elements along the diagonal are the square roots of the diagonal elements of $\text{Cov}(Y_1, Y_1)$; thus, they are standard deviations. Each standard deviation is associated with an element in vector Y_1 , and hence with one location in domain D . Each standard deviation is interpreted as the scaling applied to the associated correlation function. Using the terminology developed for variograms (Banerjee and others, 2015, p. 24–30), each standard deviation is the square root of the partial sill, which is defined as the sill minus the nugget.

Matrix $\text{Cor}(Y_1, Y_1)$ is a correlation matrix; it is square and symmetric. All elements along the diagonal of matrix $\text{Cor}(Y_1, Y_1)$ equal 1, whereas all other elements range between -1 and 1 . Element ij in the correlation matrix is interpreted as the degree of association between elements i and j in vector Y_1 . A row of the correlation matrix is related to, but not identical to, the shape of a variogram.

The second term on the right side of equation 3.2 is the covariance associated with the variability of the process model; this covariance, expressed in terms of the model variables, is:

$$\text{Cov}(\zeta, \zeta) = \text{diag}([\rho \exp(B\psi)]^2) \quad (3.5)$$

Matrix $\text{Cov}(\zeta, \zeta)$ is square and diagonal. Each diagonal element of matrix $\text{Cov}(\zeta, \zeta)$ is associated with an element in vector Y_1 , and hence with one location in domain D . The elements of matrix $\text{Cov}(\zeta, \zeta)$ are interpreted as the variance in the process at distances less than the spacing between the field samples. Using the terminology developed for variograms (Banerjee and others, 2015, p. 24–30), a diagonal element of matrix $\text{Cov}(\zeta, \zeta)$ is the nugget.

To gain intuition about the covariance function $\text{Cov}(Y, Y)$, its components (equations 3.4 and 3.5) are calculated and plotted for the example presented in the Demonstration of Method section. The diagonal elements of matrix S are plotted as a map (fig. 3.1). The map shows that these standard deviations are spatially varying. The highest values are near the edge of the domain, and relatively low values are in central Alabama and South Carolina. This pattern is related to the sampling density (fig. 3)—the higher the sampling density, the lower the standard deviation. An exception is along the coast of North Carolina where there is very low sampling density and no basis functions beyond the seacoast (fig. 4).

A correlation function, which is a row of matrix $\text{Cor}(Y_1, Y_1)$, is plotted as a map (fig. 3.2). This correlation function pertains to point 2, which is identified in figure 2.1. This correlation function has a maximum (which equals 1) at the center of the red dot, which is the location of point 2. The correlation function is approximately circularly symmetric with respect to point 2. As the radial distance from point 2 increases, the correlation decreases, becomes slightly negative (which is indicated by the gray ring surrounding the red dot), and increases to approximately 0. The maps of the correlation functions for the other three points (which also are identified in figure 2.1) are similar to that in figure 3.2—except, of course, that correlation functions are translated to the point locations; consequently, these maps are not shown.

To analyze the correlation matrix $\text{Cor}(Y_1, Y_1)$ in greater detail, the correlation functions along radial transects are plotted for each of the four points that are identified in figure 2.1 (fig. 3.3). The maximum distance along a transect is 95.1 kilometers; however, if the transect intersects the boundary of the domain, then it terminates there. The correlation functions along the four transects have approximately the same shape, but they differ in some details. For example, the function associated with point 1 is narrower than the function associated with point 2. The inference is that the correlation function $\text{Cor}(Y_1, Y_1)$ is spatially varying. The correlation functions associated with points 1, 2, and 4 are slightly asymmetric with respect to zero. (Because transect 3 terminates at -11.9 kilometers, no claim regarding asymmetry is made.) Thus, the correlation function as shown in map view (fig. 3.2) is only approximately radially symmetric.

The square root of the diagonal elements of matrix $\text{Cov}(\xi, \xi)$, which are standard deviations, are plotted as a map in figure (fig. 6B). This map shows that these standard deviations vary spatially. These standard deviations are much larger than the standard deviation that are the diagonal elements of the matrix S (fig. 3.1).

The standard deviations in matrix S , the standard deviations in matrix $\text{Cov}(\xi, \xi)$, and the correlation functions in matrix $\text{Cor}(Y_1, Y_1)$ vary spatially. That is, all components of $\text{Cov}(Y, Y)$ vary spatially, so $\text{Cov}(Y, Y)$ must vary spatially too.

Reference Cited

Banerjee, S., Carlin, B.P., and Gelfand, A.E., 2015, Hierarchical modeling and analysis for spatial data (2d ed.): Boca Raton, Fla., CRC Press, 562 p.

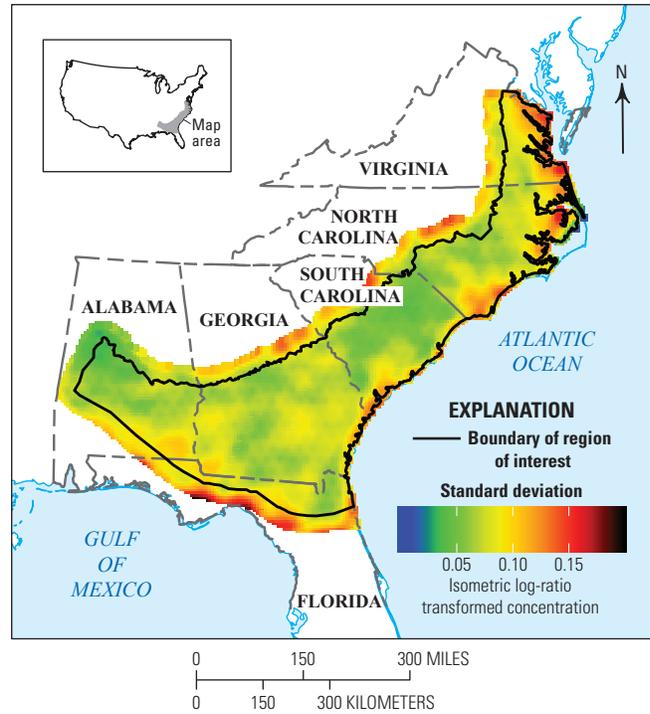


Figure 3.1. Map showing the standard deviations that are the diagonal elements of matrix S .

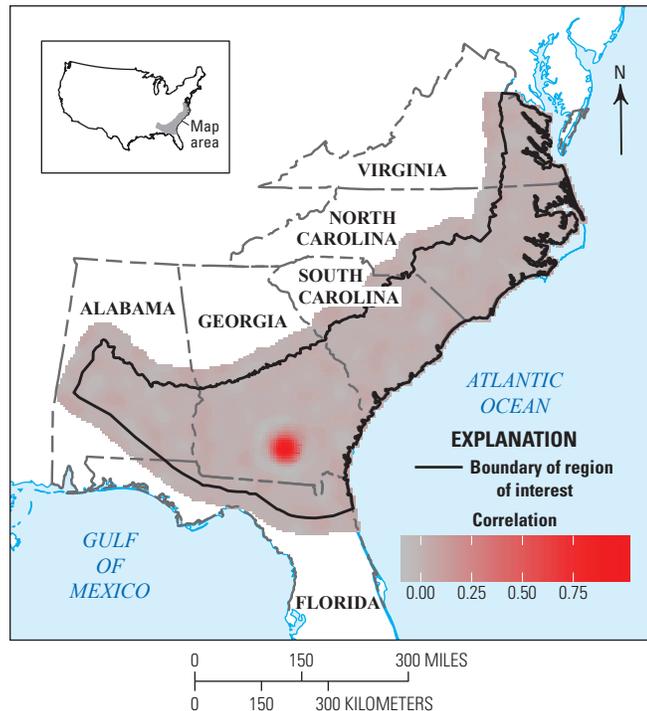


Figure 3.2. Map showing the correlation function for point 2 (which is identified in fig. 2.1).

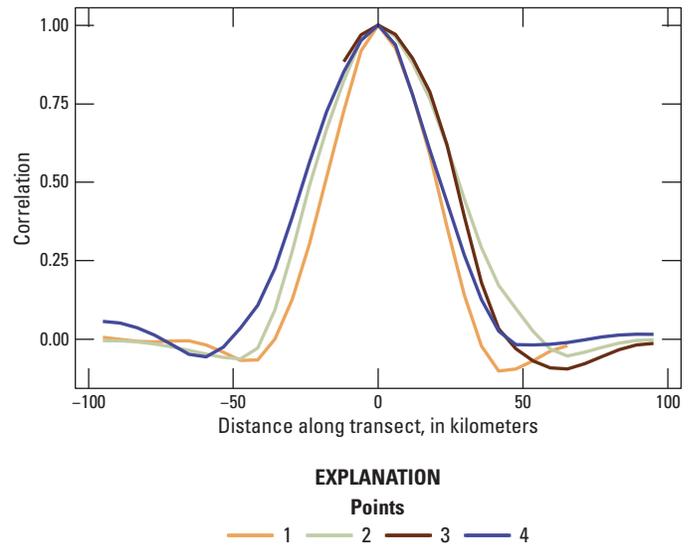


Figure 3.3. Graphs showing the correlation functions for the four points identified in figure 2.1. Each correlation function is along a transect that trends north-south. Distance along a transect is measured with respect to its associated point. Positive-valued distance is north of the point, whereas negative-valued distance is south of the reference point.

