Statistical Science
2003, Vol. 18, No. 4, 436–456

Spatial Statistics in the Presence of Location Error with an Application to Remote Sensing of the Environment

Noel Cressie and John Kornak

Abstract. Techniques for the analysis of spatial data have, to date, tended to ignore any effect caused by error in specifying the spatial locations at which measurements are recorded. This paper reviews the methods for adjusting spatial inference in the presence of data-location error, particularly for data that have a continuous spatial index (geostatistical data). New kriging equations are developed and evaluated based on a simulation experiment. They are also applied to remote-sensing data from the Total Ozone Mapping Spectrometer instrument on the *Nimbus-7* satellite, where the location error is caused by assignment of the data to their nearest grid-cell centers. The remote-sensing data measure total column ozone (TCO), which is important for protecting the Earth's surface from ultraviolet and other radiation.

Key words and phrases: Attribute error, CP model, FP model, errors-in-variables, geographic information systems, geostatistics, GIS, GPS, kriging.

1. INTRODUCTION

The field of spatial (and spatiotemporal) statistics is a fertile area for innovations in data analysis and statistical inference. Data are considered spatial if they contain locational information, where interest can lie in the locations themselves or in "attributes" associated with those locations. Although much attention has been given to explaining the sources of "signal" and "error" for an observation (i.e., attribute) as a function of its location, there has been relatively little effort given by spatial statisticians to considering the effect of spatial uncertainty associated with that location. From navigation using the stars, to the use of global positioning systems (GPS) to trace a 911 call on a cell phone, location error has been a source of variability for anyone wishing to "go from A to B" (and back again). Location error is ubiquitous in seismology, as scientists attempt to predict both locations and magnitudes of earthquakes (e.g., Veneziano and Van Dyck, 1987). Also, ship captains have reason to be concerned with location error that arises from two sources: Kielland and Tubman (1994) attempt to reduce the risk of ships running aground by accounting for uncertainty in both a navigator's current position and the ship's navigational maps. A computed two-sigma error ellipse is used to describe the sea vessel's positional uncertainty (e.g., associated with a GPS). Additionally, kriging is performed on data determined from sparsely sampled depth soundings, from which depth contours are generated. Ad hoc tolerance regions are put around each contour to account for location error, and navigation is then performed based on a worst-case analysis.

Understanding locational variability, making scientific inferences in its presence and controlling for it in design settings, is the realm of statistical science. In this paper, we consider the effect of location error in spatial statistics, but mainly in the subfield known as geostatistics (Matheron, 1963). Here the spatial index varies continuously over a given spatial domain, and geostatistical models for attributes delineate global trend effects, local medium-scale spatial ef-

Noel Cressie is Professor, Department of Statistics, The Ohio State University, Columbus, Ohio 43210-1247 (e-mail: ncressie@stat.ohio-state.edu). John Kornak is Assistant Professor at the Magnetic Resonance Unit, University of California, San Francisco, Veterans Affairs Medical Center, San Francisco, California 94121-1545 (e-mail: kornak@itsa.ucsf.edu).

fects, a microscale effect and a measurement-error effect (e.g., Cressie, 1993, page 112; Diggle, Tawn and Moyeed, 1998). Mapping in environmental science is ubiquitous and this is something geostatistics does very well. Furthermore, with the rise in high-level statistical programming languages, geostatistical tools can now be found in standard statistical software packages such as S-PLUS, SAS and in contributed libraries for R.

The environmental problem considered in this paper is mapping the ozone layer that serves as a protective shield against harmful radiation from outer space. Depletion of the ozone layer results in increased ultraviolet radiation, which causes cell damage in living organisms leading to increased skin-cancer rates and reduced rates of growth and reproduction (e.g., Cocchi and Trivisano, 2002). The principal data source is the Total Ozone Mapping Spectrometer on a remote-sensing platform. The variable of interest is total column ozone (TCO), measured in Dobson units (London, 1985). In this context, location error is often purposively imposed by data analysts struggling with the massiveness of the data, who for convenience move the data to nearest nodes of a regular grid.

We acknowledge here some of the models found in the field of morphometrics, where the data usually consist of the (relative) *locations* of landmarks (e.g., Bookstein, 1986) and where one is concerned with making inference on an object's "size" and "shape" defined by these landmark locations. We contrast this with the theme of our paper, which is concerned with making inferences pertaining to *attributes* at these locations. An obvious model of interest for us is the following simple one described by Bookstein (1986): landmarks $L_1, L_2, \ldots, L_K$ are distributed about their centroids $W_1, W_2, \ldots, W_K$, according to a Gaussian normal distribution, with covariance matrix equal to the identity matrix multiplied by a variance parameter; that is, the landmarks have a coordinate-wise-independent (circular) Gaussian distribution about their centroids. This is a special case of the location-error model given by (1) below.

Broadly speaking, the models of spatial statistics can be categorized into geostatistical models, lattice models and spatial point processes (Cressie, 1993), where the relative locations in space between data points are used to determine an overall spatial-dependence structure crucial for statistical inferences. Therefore, errors in location may transfer to errors in specifying and estimating the spatial-dependence structure. Specifically, in geostatistics, location error affects the covariates at a location, as well as the spatial lag between locations. This in turn affects the theoretical and empirical spatial covariance function (or variogram) and has a potentially serious effect on trend-parameter estimation and kriging. The rest of this section is essentially review in nature: in Section 1.1 we present known results for location error in geostatistical models, and Section 1.2 shows that there is an analogy between errors-in-variables models and geostatistical location-error models. Section 1.3 reviews location-error research in lattice models and spatial point processes.

1.1 Location Error in Geostatistics

Geographical information systems (GIS) provide a framework within which extremely large spatial databases can be stored, retrieved, manipulated and mapped with ever-increasing speed. The datasets can contain vast quantities of information that create the need for data-analysis and statistical-inference techniques that are efficient (in terms of computational speed, accuracy and precision). Geostatisticians have used location information in these datasets to model trend and correlation between attribute values (e.g., generalized-least-squares parameter estimation, variogram estimation, kriging and so forth); see Cressie (1993, Part I). However, based on our review of the geostatistics literature, we conclude that uncertainty in the location information is generally ignored.

There have been sporadic treatments of location error in geostatistics, most of it simply noting its consequences without incorporating it into a full spatial-statistical-inference engine. Matérn (1960, page 56) derives the covariance between two observations whose locations are perturbed by location error. However, he does not derive the variance of any individual observation, something that is essential for inference (trend-parameter estimation, kriging). Chiles (1976) considers the consequences of location error for kriging when the data have a simple linear trend (but no measurement error). However, he does not derive the variances of the observations, and he does not seem to be aware that the covariance between the prediction location and a datum has only one source of location error. Atkinson (1997) derives some second-moment properties of a geostatistical model in the presence of location error. Chiles and Delfiner (1999) assume a process with constant mean and derive the variogram between any two observations whose locations are perturbed by location error; they do not do kriging. Finally, Gabrosek and Cressie (2002) consider a geostatistical model with measurement error, location er-

ror and a trend given by a linear combination of spatial covariates. They give kriging equations that are correct for constant mean but lack a component of variation for the more general trend term. In what follows in this paper, we give general results for the spatial variances and covariances, and we derive a kriging methodology for geostatistics in the presence of location error.

Let $\mathbf{s}$ be a generic spatial index that ranges continuously over the spatial domain $D \subset \mathbb{R}^d$, where $|D| > 0$. Then a spatial process of the attribute Z is denoted as $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$. In Gabrosek and Cressie (2002) two major location-error models were defined in the geostatistical setting (i.e., where the spatial index varies continuously).

1.1.1 *The coordinate-positioning model.* This kind of location error occurs when one attempts to take observations (i.e., attributes) on a spatial process at a predetermined set of locations, given by their coordinates. The model was referred to as the "design model" in Gabrosek and Cressie (2002), to emphasize that the locations of samples are chosen by design. We find the current nomenclature more descriptive.

Consider a set of n intended sites $S = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, perturbed by location error to sites $R = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ at which the observations are actually taken. The actual sites R are typically unknown to us. The source of location error may be imprecise positioning instruments, positional coordinate rounding or human error (e.g., in map reading). In the coordinate-positioning (CP) model location error occurs because of an inability to go exactly to (and hence sample from) the intended sites S. Then one possible location-error model for the distribution of actual (but unknown) site locations R, given the intended site locations S, is

$$R|S \sim (S, \Sigma_g), \tag{1}$$

where Σ_g is the covariance matrix of the (unknown) sites R, distributed about the (known) mean S. [The distribution in (1) is momentarily left unspecified, but two obvious choices are the d-dimensional Gaussian distribution and the uniform distribution on a d-dimensional ball.]

An example of this kind of location error might occur when sampling contaminated soil in an environmental characterization study. Suppose that a set of test-site locations is specified to be on a regular grid. However, due to imprecise positioning of the equipment, sampling actually occurs at coordinates that are displaced from the regular grid, and these displacements are unknown.

It was mentioned earlier that location error is sometimes introduced purposively. This can happen when datasets are extremely large. For example, in satellite remote sensing, data locations are usually reassigned to nearest nodes of a (regular) grid; the U.S. space agency NASA uses regular latitude–longitude grids. To date, these imposed errors are generally ignored and, when analyzed, the data are assumed to have been recorded at the grid-node location to which they were reassigned. A little thought reveals that the CP model is sufficiently flexible to handle this situation: in this instance, the actual data locations make up R, and the grid nodes make up S. [It is not relevant that R is known, since this information is ignored.] Although in this case the location error is clearly not a random process, it can be handled quite easily this way. In essence, we describe the purposive uncertainty in the true locations via a probabilistic framework, so that although information is lost due to reassigning the data locations, at least the resulting location error is accounted for. In Section 5, we shall use the CP model to analyze total column ozone data from the *Nimbus-7* satellite.

1.1.2 *The feature-positioning model.* This form of location error occurs when we have a set of easily recognizable "features" at which to take measurements. However, we determine the location of these features with some error. This form of data collection was referred to as *resource sampling* in Gabrosek (1999), because the feature at which a measurement is taken usually represents a resource of some sort. For example, the feature might be a tree whose logging potential is of interest. The tree is fixed in space, but we are unable to obtain its exact location because we do not have infinitely accurate positioning instruments. Here we have the set of true feature locations $A = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$, which are unknown, and a corresponding set of observed locations $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$, that are observed with error (e.g., using a global positioning system). In this case, we could model

$$B|A \sim (A, \Gamma_g). \tag{2}$$

The feature-positioning (FP) model could also be considered as a component of a hierarchical marked-point-process model, where the feature positions, modeled by a spatial point process, are then observed with location error. The "marks" of the marked point processes are what we have called attributes, and we may be interested in the effect of location error on spatial inference for the mark process.

Although at first glance the coordinate-positioning model (1) and the feature-positioning model (2) appear similar, there is a very important distinction between them. In the first model the locations of the intended sites are known, and the location-error distribution is (generally) centered around those *known* locations. However, in the second model, the location-error distribution is centered around the *unknown* locations of the features. This is further illustrated in the next subsection, where we note differences between the CP and FP models by viewing them as analogous to errors-in-variables models.

1.2 Errors-in-Variables Models

There is a strong correspondence between location-error models in geostatistics and errors-in-variables models (e.g., Fuller, 1987; Fazekas, Baran, Kukush and Lauridsen, 1999). Consider a set of covariates $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and a corresponding set of response variables $\{z_1, \ldots, z_n\}$, linked through the model

$$z_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \delta_i, \tag{3}$$

where $f(\mathbf{x}_i, \boldsymbol{\beta})$ is the (possibly nonlinear) mean function, and δ_i is a zero-mean error term independent of $\mathbf{x}_i$. A special case of (3) is a model described by Berkson (1950):

$$z_i = \alpha + \beta x_i + \delta_i,$$

where

$$w_i = x_i + p_i.$$

Now, one possibility is to specify $\{x_i\}$ as target values (controls) at which to observe $\{z_i\}$, but because of errors $\{p_i\}$ one observes z_i at $w_i = x_i + p_i$, where w_i is *not observed*. The other possibility is to consider a collection of *observed* $\{w_i\}$, where again $w_i = x_i + p_i$, but instead one observes z_i at x_i. The former model is often called Berkson's model (see, e.g., Burr, 1988), and it is analogous to our CP model for location error. The latter model is analogous to our FP model. Clearly, if an inappropriate model is used, the resulting inference is biased.

1.3 Location Error in Other Spatial Models

Geographers and users of raster-based GIS have often assumed that the spatial domain is a fixed regular grid of pixels. Then, for investigating the quality of output maps, the statistical-image-analysis approach of Geman and Geman (1984) has been adapted to quantify how location error propagates through to the output maps (e.g., Goodchild, 1989; Haining and Arbia, 1993; Arbia, Griffith and Haining, 1998).

Output maps often result from overlay operations that combine two or more source maps at potentially different scales of resolution. The unobservable true source map (T) is perturbed to give a location-error-corrupted output map (Z). Denote T_{ij} as the true attribute value at pixel (i, j) and Z_{ij} as the corresponding location-error-corrupted output value. Let (g, h) be the location shift of pixel (i, j), so that its location-error-corrupted position is $(i + g, j + h)$, where $g, h = 0, \pm 1, \pm 2, \ldots$. Arbia, Griffith and Haining (1998) use the corruption model

$$Z_{ij} = \sum_g \sum_h w_{ij}(g, h) T_{i+g, j+h} + u_{ij}.$$

In the equation above, $\{w_{ij}(g, h) : g, h = 0, \pm 1, \pm 2, \ldots\}$ satisfies

$$\sum_g \sum_h w_{ij}(g, h) = 1 \quad \text{and} \quad w_{ij}(g, h) \geq 0 \quad \text{for all } (i, j),$$

and can be thought of as a location-error mixing distribution; the last term u_{ij} is the attribute error. We do not pursue this model further, since it is not appropriate for location error expressed as a *random* displacement.

For Markov-random-field (MRF) lattice models, the structure of the spatial dependence remains unchanged provided the neighborhoods (those sites that are conditionally spatially dependent) of the MRF are unaffected by location error. [However, the strength of the spatial dependence may change if the dependence is defined in terms of actual distances; see, e.g., Cressie (1993), equation (7.6.20).] The location-error problem is an important one in the social sciences, where error is introduced (e.g., data are aggregated to zip codes) to maintain confidentiality of subjects.

There is a literature on location error in point processes that dates back to at least the 1960s; see, for example, the review given in Cox and Isham (1980, pages 104–106). An important theorem of Dobrushin (1963), whose proof was corrected and extended to d dimensions by Stone (1968), states that under suitable regularity conditions a translated point process will converge to the homogeneous Poisson process, as the translation (i.e., location-error) variance tends to infinity. Veneziano and Van Dyck (1987) consider the effect of location error on the relationship between observed attributes ("marks" in the point process literature) at points, applied to earthquake epicenter locations and their corresponding magnitudes.

Diggle (1993, pages 97–99) provides some results with respect to estimation of parameters of the true point process before perturbation, from only the point patterns observed after perturbation. Assumptions of (i) a point process that is stationary, isotropic and exhibits clustering and (ii) a radially symmetric location error (to preserve overall isotropy) are used. The first-order intensity,

$$\lambda \equiv \lim_{|d\mathbf{s}|\to 0}\left\{\frac{E[N(d\mathbf{s})]}{|d\mathbf{s}|}\right\},$$

where $N(\cdot)$ denotes the number of points in the region A, is unaffected by the perturbations, since on average no events are added or deleted. Second-moment properties of stationary and isotropic point processes are often characterized by the K-function $K(h)$, defined by

$$\lambda K(h) \equiv E[\text{no. extra events within distance } h \text{ of an arbitrary event}].$$

Diggle's general conclusion is that if one uses a standard K-function estimator based on the perturbed point pattern, then it will be an underestimate of the K-function of the nonperturbed point process.

For the remainder of this paper, we concentrate on the *CP model in geostatistics*. In Section 2, we consider the effect of location error on the spatial trend and covariance function, and we discuss their estimation. This is illustrated via an artificial-data example, which is used again in Section 3. Section 3 is devoted to optimal spatial prediction using linear predictors (kriging) and the extensions necessary to adjust for the location error. In Section 4, we give the results of a simulation experiment that compares kriging adjusting for location error (KALE) with kriging ignoring location error (KILE). Section 5 provides an illustration of KALE on remotely sensed total column ozone data. Finally, Section 6 contains discussion and conclusions.

2. CP LOCATION-ERROR MODEL IN GEOSTATISTICS

We focus here on introducing the CP location-error model into geostatistical methodology, extending and correcting the earlier work of Gabrosek and Cressie (2002). Consider a spatial dataset $\mathbf{Z} \equiv \{Z(\mathbf{s}_i) : \mathbf{s}_i \in D \subset \mathbb{R}^d; i = 1, \ldots, n\}$, which are observations from a random process $Z(\cdot)$ with continuous spatial index over $D \subset \mathbb{R}^d$. In all the examples we consider in this paper, $d = 2$, although when deriving the geostatistical methodology we retain the generality of working in $\mathbb{R}^d$.

The data $\mathbf{Z}$ correspond to a single partial realization of $Z(\cdot)$ and, to make inference on the noiseless version of $Z(\cdot)$ and its parameters, further assumptions need to be made. Specifically, we assume that $Z(\cdot)$ has linear trend in spatial covariates and second-order stationary errors:

$$\begin{aligned}
&\text{(i)} && \operatorname{var}(Z(\mathbf{s})) < \infty && \text{for all } \mathbf{s} \in D;\\
&\text{(ii)} && E(Z(\mathbf{s})) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} && \text{for all } \mathbf{s} \in D;\\
&\text{(iii)} && \operatorname{cov}\big(Z(\mathbf{s}_1), Z(\mathbf{s}_2)\big) && \\
& && \quad = C(\mathbf{s}_2 - \mathbf{s}_1) && \text{for all } \mathbf{s}_1, \mathbf{s}_2 \in D,
\end{aligned} \tag{4}$$

where $C(\cdot)$ is a positive-definite function on $\mathbb{R}^d$. These assumptions imply the following: (i) the first two moments of the process are defined; (ii) the mean function is linear in a $(q \times 1)$ vector of unknown parameters $\boldsymbol{\beta}$; and (iii) the covariance between any two points depends only on their relative spatial locations.

The noiseless version of $Z(\cdot) = Y(\cdot) + \varepsilon(\cdot)$ is the random process $Y(\cdot)$, where $\varepsilon(\cdot)$ is an independent white-noise process of measurement errors with variance c_{ME}. Now $Z(\cdot)$ has stationary covariance function if and only if $Y(\cdot)$ does; we denote the latter as $C_Y(\cdot)$.

In the CP model an intended site $\mathbf{s}$ is perturbed by location error to an actual (but unobserved) site $\mathbf{r} = \mathbf{s} + \mathbf{p}(\mathbf{s})$. In this paper the location errors $\{\mathbf{p}(\mathbf{s})\}$ are assumed to be independent continuous random vectors each with probability density function g that is invariant over $\mathbf{s}$, $\mathbf{s} \in D$. Then the process $Z(\cdot)$ is sampled at $\mathbf{r}$ but mistakenly located at $\mathbf{s}$, yielding the process $Z_g(\cdot)$.

We attempt to follow the notation for geostatistical models in Cressie (1993) as closely as possible, and consequently the CP location-error model becomes

$$\begin{aligned}
Z_g(\mathbf{s}) &\equiv Y(\mathbf{r}) + \varepsilon(\mathbf{r}) = \mathbf{x}(\mathbf{r})'\boldsymbol{\beta} + \nu(\mathbf{r}) + \varepsilon(\mathbf{r}),\\
&\mathbf{r} = \mathbf{s} + \mathbf{p}(\mathbf{s}), \quad \mathbf{s} \in D,
\end{aligned} \tag{5}$$

where $Y(\cdot)$ is the noiseless version of $Z(\cdot)$; the location-error vector $\mathbf{p}(\mathbf{s})$ has probability density function $g(\cdot)$ in $\mathbb{R}^d$; $\mathbf{x}(\cdot)'\boldsymbol{\beta}$ is the deterministic trend; and $\varepsilon(\cdot)$ is the (attribute) measurement error with measurement-error variance, $\operatorname{var}(\varepsilon(\mathbf{s})) \equiv c_{\mathrm{ME}}$. The spatial error $\nu(\cdot)$ can be broken down into two additive, zero-mean components. That is, $\nu(\cdot) = W(\cdot) + \eta(\cdot)$, where $W(\cdot)$ is the smooth small-scale variation, and $\eta(\cdot)$ is sometimes known as the *nugget effect* (Matheron, 1963), but we shall call it the microscale variation (Cressie, 1993, page 128). The microscale variation has a highly localized covariance

structure (generally at higher resolution than the spatial sampling frequency, and hence indistinguishable from white noise) with $\operatorname{var}(\eta(\mathbf{s})) = c_{\mathrm{MS}}$. Consequently, $C_Y(\mathbf{h})$ has a discontinuity of c_{MS} at $\mathbf{h} = \mathbf{0}$.

We make the point here that $\eta(\cdot)$ is inherently different from $\varepsilon(\cdot)$ in its genesis, even though they are both modeled as white noise (Cressie, 1993, Section 3.1). Specifically, $\eta(\cdot)$ is a component of $\nu(\cdot)$, which is in turn a component of the "true" process of interest $Y(\cdot)$, whereas $\varepsilon(\cdot)$ is a noise process that we want to filter out. However, without multiple observations per location on $Z(\cdot)$, $\eta(\cdot)$ and $\varepsilon(\cdot)$ are indistinguishable, since $C(\mathbf{0}) - \lim_{\mathbf{h}\to\mathbf{0}} C(\mathbf{h}) = c_{\mathrm{MS}} + c_{\mathrm{ME}}$. Henceforth, we shall assume that the measurement-error variance c_{ME} is known, as it usually is from independent instrument-calibration experiments.

The process $Z_g(\cdot)$ defined by (5) should be compared to the location-error-free process

$$\begin{aligned} Z(\mathbf{s}) &\equiv Y(\mathbf{s}) + \varepsilon(\mathbf{s}) \\ &= \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \nu(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad \mathbf{s} \in D, \end{aligned} \tag{6}$$

whose properties are summarized by (4). Notice that $Z_g(\mathbf{s}) = Z(\mathbf{s} + \mathbf{p}(\mathbf{s}))$, $\mathbf{s} \in D$, and hence in the presence of location error we shall denote the spatial dataset as

$$\mathbf{Z}_g \equiv \big(Z_g(\mathbf{s}_1), \ldots, Z_g(\mathbf{s}_n)\big)', \tag{7}$$

associated with intended sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$.

In addition to our earlier independence and invariance assumptions about the components of $\mathbf{p}(\cdot)$, we make the further assumption that $\mathbf{p}(\cdot)$ is independent of $\varepsilon(\cdot)$ and of $Y(\cdot)$. That is, location error, measurement error and spatial-process error are mutually independent. In most cases, this is a reasonable assumption.

Suppose that we formally define the location-error component of variation in the attribute as $\xi_g(\mathbf{s}) \equiv Z_g(\mathbf{s}) - Z(\mathbf{s})$. That is, $Z_g(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s}) + \xi_g(\mathbf{s})$, $\mathbf{s} \in D$. Then, for $\mathbf{p}(\mathbf{s}) \not\equiv \mathbf{0}$,

$$\begin{aligned} \operatorname{var}\big(\xi_g(\mathbf{s})\big) &= \operatorname{var}\big(Z_g(\mathbf{s}) - Z(\mathbf{s})\big) \\ &= E\big[\operatorname{var}\big(Z(\mathbf{s}+\mathbf{p}(\mathbf{s})) - Z(\mathbf{s})|\mathbf{p}(\mathbf{s})\big)\big] \\ &\quad + \operatorname{var}\big[E\big(Z(\mathbf{s}+\mathbf{p}(\mathbf{s})) - Z(\mathbf{s})|\mathbf{p}(\mathbf{s})\big)\big] \\ &= E\big[\operatorname{var}\big(Y(\mathbf{s}+\mathbf{p}(\mathbf{s})) - Y(\mathbf{s})\big)\big] + 2c_{\mathrm{ME}} \\ &\quad + \operatorname{var}\big(\boldsymbol{\beta}'\mathbf{x}(\mathbf{s}+\mathbf{p}(\mathbf{s}))\big) \\ &= 2\int \big(C_Y(\mathbf{0}) - C_Y(\mathbf{u})\big)g(\mathbf{u})\,d\mathbf{u} \\ &\quad + 2c_{\mathrm{ME}} + \boldsymbol{\beta}' M_g(\mathbf{s})\boldsymbol{\beta}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} M_g(\mathbf{s}) &\equiv \int \mathbf{x}(\mathbf{s}+\mathbf{u})\mathbf{x}(\mathbf{s}+\mathbf{u})'g(\mathbf{u})\,d\mathbf{u} \\ &\quad - \mathbf{x}_g(\mathbf{s})\mathbf{x}_g(\mathbf{s})', \\ \mathbf{x}_g(\mathbf{s}) &\equiv \int \mathbf{x}(\mathbf{s}+\mathbf{u})g(\mathbf{u})\,d\mathbf{u} \end{aligned}$$

and

$$C_Y(\mathbf{h}) \equiv C(\mathbf{h}) - c_{\mathrm{ME}} I(\mathbf{h} = \mathbf{0}), \quad \mathbf{h} \in \mathbb{R}^d,$$

is the covariance function of $Y(\cdot)$ with $I(\cdot)$ denoting the indicator function. The three components of variation in (8) are, respectively, due to spatial dependence, measurement error and trend. Notice that, when $\mathbf{p}(\mathbf{s}) \equiv \mathbf{0}$, $\operatorname{var}(\xi_g(\mathbf{s})) = 0$.

However, it would not be correct to interpret $\xi_g(\cdot)$ as an independent component of attribute variability. First, it is easy to see that $E(\xi_g(\mathbf{s})) = (\int \mathbf{x}(\mathbf{s}+\mathbf{u})g(\mathbf{u})\,d\mathbf{u} - x(\mathbf{s}))'\boldsymbol{\beta}$, which is not zero in general. Second, a derivation similar to (8) yields $\operatorname{cov}(Y(\mathbf{s}), \xi_g(\mathbf{s})) = -\int (C_Y(\mathbf{0}) - C_Y(\mathbf{u}))g(\mathbf{u})\,d\mathbf{u}$, which is always nonpositive (Atkinson, 1997).

2.1 Adjusting the Moments to Account for Location Error

As can be seen from (4), the mean and covariance functions play a major role in understanding the spatial process $Z(\cdot)$. In the presence of location error, it is the first two moments of $Z_g(\cdot)$ that are of interest. From them we gain insight into the hidden process $Y(\cdot)$. To estimate $\boldsymbol{\beta}$ and to predict $Y(\mathbf{s}_0)$, $\mathbf{s}_0 \in D$, in the presence of location error, we use first and second moments adjusted for location error (see Sections 2.3 and 3).

To find the mean of the process $Z_g(\cdot)$, one integrates the mean of the process $Z(\cdot)$ over all possible locations, weighted by the density g of the location error; see Gabrosek and Cressie (2002). Since $\varepsilon(\cdot)$ has zero mean, we obtain the CP-model-adjusted mean,

$$\begin{aligned} \mu_g(\mathbf{s}) &\equiv E(Z_g(\mathbf{s})) \\ &= \left(\int \mathbf{x}(\mathbf{s}+\mathbf{u})g(\mathbf{u})\,d\mathbf{u}\right)'\boldsymbol{\beta} \\ &\equiv \mathbf{x}_g(\mathbf{s})'\boldsymbol{\beta}. \end{aligned} \tag{9}$$

More generally, $\mu_g(\mathbf{s}) = \int \mu(\mathbf{s}+\mathbf{u})g(\mathbf{u})\,d\mathbf{u}$, where $\mu(\cdot) \equiv E(Z(\cdot))$. Notice that, when $\mu(\cdot)$ is linear in $\boldsymbol{\beta}$, so too is $\mu_g(\cdot)$; from (9), we see that it is simply the covariates that are adjusted.

The covariance for $Z_g(\cdot)$ can also be adjusted, but it is somewhat more complicated than adjusting

the mean. Note that the spatial lag in the stationary covariance function becomes a random vector in the presence of location error. For $\mathbf{h} \neq \mathbf{0}$ the adjusted covariance is

$$C_g(\mathbf{h}) \equiv \operatorname{cov}\big(Z_g(\mathbf{s}), Z_g(\mathbf{s}+\mathbf{h})\big) = \int\!\!\int C_Y(\mathbf{h}+\mathbf{v}-\mathbf{u}) g(\mathbf{u}) g(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v}, \tag{10}$$

which is stationary. Now, if $C_Y(\mathbf{h})$ depends on $\mathbf{h}$ through $h \equiv \|\mathbf{h}\|$ (isotropic) and $g(\mathbf{u})$ depends on $\mathbf{u}$ only through $\|\mathbf{u}\|$, it is straightforward to show that $C_g(\mathbf{h})$ depends on $\mathbf{h}$ only through h, $\mathbf{h} \neq \mathbf{0}$.

In contrast to the adjusted covariances, the adjusted variance function is no longer necessarily stationary. It is

$$\begin{aligned} v_g(\mathbf{s}) &\equiv \operatorname{var}(Z_g(\mathbf{s})) \\ &= E\big[\operatorname{var}\big(Y\big(\mathbf{s}+\mathbf{p}(\mathbf{s})\big)|\mathbf{p}(\mathbf{s})\big)\big] \\ &\quad + c_{\mathrm{ME}} + \operatorname{var}\big(\boldsymbol{\beta}'\mathbf{x}\big(\mathbf{s}+\mathbf{p}(\mathbf{s})\big)\big). \end{aligned}$$

That is,

$$v_g(\mathbf{s}) = C_Y(\mathbf{0}) + c_{\mathrm{ME}} + \boldsymbol{\beta}' M_g(\mathbf{s}) \boldsymbol{\beta}. \tag{11}$$

Only in the constant-mean case, where $\mathbf{x}(\cdot) \equiv 1$, do we have $M_g(\cdot) \equiv 0$, and hence $v_g(\mathbf{s}) = C_Y(\mathbf{0}) + c_{\mathrm{ME}} = C(\mathbf{0})$. Consequently, $K_g(\mathbf{s},\mathbf{t}) \equiv \operatorname{cov}(Z_g(\mathbf{s}), Z_g(\mathbf{t}))$ is given by

$$K_g(\mathbf{s},\mathbf{t}) = \begin{cases} v_g(\mathbf{s}), & \mathbf{s}=\mathbf{t}, \\ C_g(\mathbf{t}-\mathbf{s}), & \mathbf{s}\neq\mathbf{t}. \end{cases}$$

In general, $K_g(\mathbf{s},\mathbf{t})$ is not a function of $\mathbf{t}-\mathbf{s}$ (covariance stationarity), except when the mean $\mu(\cdot)$ is constant. Notice that $K_g(\cdot,\cdot)$ satisfies the requirement of positive-definiteness on $\mathbb{R}^d \times \mathbb{R}^d$ because $C_Y(\cdot)$ is positive-definite on $\mathbb{R}^d$ and each $M_g(\cdot)$ is a nonnegative-definite matrix.

As pointed out in Gabrosek and Cressie (2002), the integrals in (9)–(11) are analytically intractable for most mean and covariance models, even when combined with simple location-error models [e.g., $g(\cdot)$ is Gaussian or uniform]. In this paper, we concentrate on a Monte-Carlo-integration approach for evaluation of (9)–(11), which opens up a wide range of location-error–distribution/mean-and-covariance-function combinations. The only requirements for our approach are that we can generate realizations from $g(\cdot)$, and that we can invert the matrix $\operatorname{var}(\mathbf{Z}_g)$ for parameter estimation and for kriging. The first requirement is generally easy to meet, and the second requirement should be no more difficult in general than for the cases where we have analytical solutions.

We now present the Monte-Carlo-integration algorithm for evaluation of (9)–(11). Because it is assumed that location error is a continuous random vector with density $g(\cdot)$, the discontinuity in $C_Y(\mathbf{h})$ at $\mathbf{h}=\mathbf{0}$ disappears after integration. Hence (10) can be written as $C_g(\mathbf{h}) = \int\!\!\int C^+(\mathbf{h}+\mathbf{v}-\mathbf{u}) g(\mathbf{u}) g(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v}$, $\mathbf{h} \neq \mathbf{0}$, where $C^+(\mathbf{h}) \equiv C_Y(\mathbf{h}) - c_{\mathrm{MS}} I(\mathbf{h}=\mathbf{0})$, which is continuous at $\mathbf{h}=\mathbf{0}$.

The Monte-Carlo-integration algorithm to evaluate the first and second moments adjusted for location error is as follows:

1. Simulate the set of points $\{\mathbf{u}_i : i = 1,\ldots,N\}$ in $\mathbb{R}^d$, each independently and identically distributed from $g(\cdot)$.
2. Approximate $\mathbf{x}_g(\mathbf{s})$ by $\frac{1}{N}\sum_{i=1}^N \mathbf{x}(\mathbf{s}+\mathbf{u}_i)$.
3. Approximate $C_g(\mathbf{h})$ by $\frac{1}{N^2}\sum_i \sum_j C^+(\mathbf{h}+\mathbf{u}_j-\mathbf{u}_i)$, $\mathbf{h}\neq\mathbf{0}$.
4. Approximate $\int \mathbf{x}(\mathbf{s}+\mathbf{u})\mathbf{x}(\mathbf{s}+\mathbf{u})' g(\mathbf{u})\, d\mathbf{u}$ by $\frac{1}{N}\sum_{i=1}^N \mathbf{x}(\mathbf{s}+\mathbf{u}_i)\mathbf{x}(\mathbf{s}+\mathbf{u}_i)'$, and hence obtain $M_g(\mathbf{s})$; then $v_g(\mathbf{s})$ is obtained from (11).

Clearly, the approximation of the covariance function for $Z_g(\cdot)$ is positive-definite, since $C^+(\cdot)$ is positive-definite and each $M_g(\cdot)$ is a nonnegative-definite matrix. Choice of N depends on which approximation, 2, 3 or 4 has the biggest coefficient of variation; N should be chosen large enough so that the maximum coefficient of variation is at most 2.5%, ensuring a (rounded) relative accuracy to the first decimal place.

Figure 1 illustrates how, in the constant-mean case, the covariance function for $Z_g(\cdot)$ changes as the dispersion in the location-error density increases. In the figure we have chosen (12) for $C_Y(\cdot)$, (13) for $g(\cdot)$ and location-error-dispersion parameter $\psi = 0.05, 0.15, 0.25$. [Recall that, in the constant-mean case, $v_g(\mathbf{s}) = C(\mathbf{0})$, independent of $\mathbf{s}$.]

2.2 An Artificial Dataset

We illustrate the Monte-Carlo-integration approach on an artificial dataset, obtained through simulation. Data generation and analysis of this and other datasets were performed in the statistical programming language R.

We simulated 100 locations $\{\mathbf{s}_1,\ldots,\mathbf{s}_{100}\}$ uniformly on the unit square. Then the 100 intended locations are perturbed with location error to the locations $\{\mathbf{r}_1,\ldots,\mathbf{r}_{100}\}$, which results in a total of 200 spatial locations. The Gaussian process $Y(\cdot)$ is simulated at these 200 locations, as well as at 50×50 locations comprising a fine-scale grid on the unit square. The

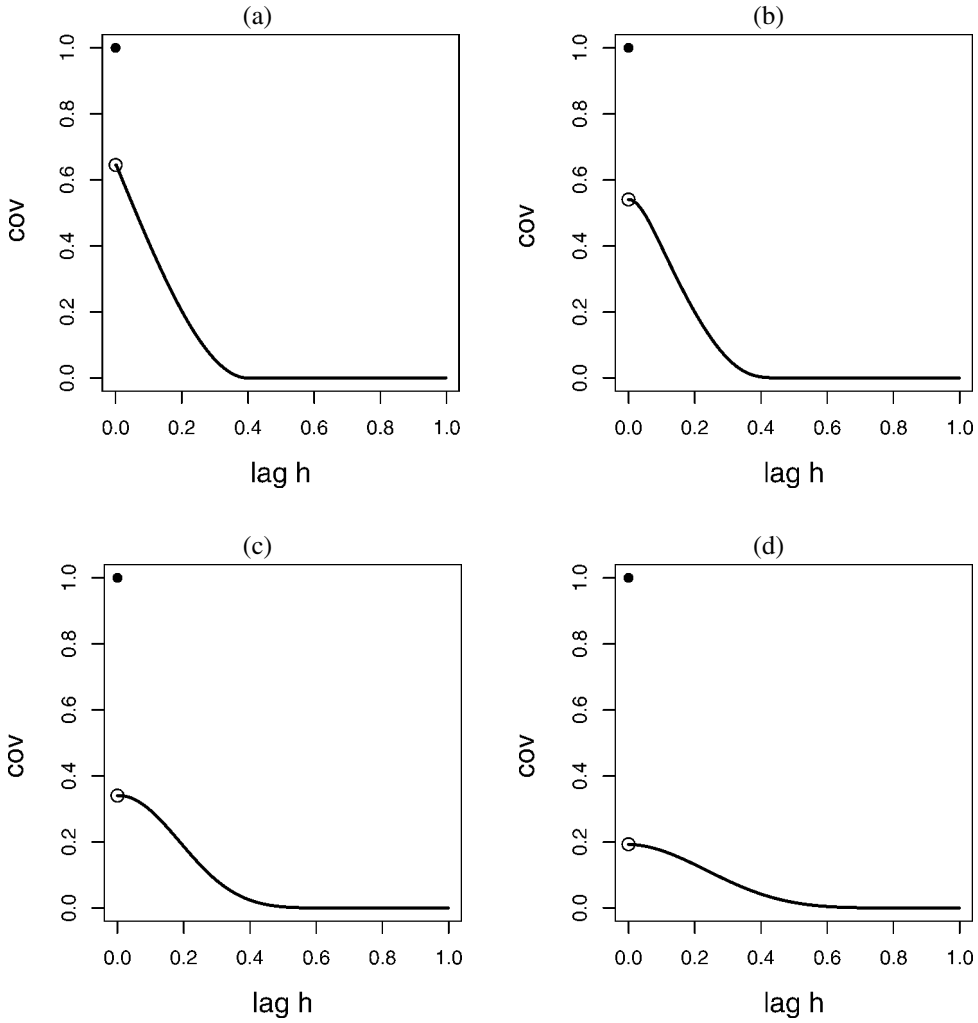

FIG. 1. *Panel* (a) *shows the spherical covariance function* (12) *in the absence of location error* [*i.e.*, $\mathbf{p}(\mathbf{s}) \equiv \mathbf{0}$]. *The other three panels show the covariance function when the process* $Z(\cdot)$ *has been perturbed by location error whose distribution is given by* (13), *namely uniform on the circle centered at* $\mathbf{0} \in \mathbb{R}^2$ *with radius* ψ: (b) $\psi = 0.05$; (c) $\psi = 0.15$; (d) $\psi = 0.25$. *In all panels,* $\boldsymbol{\theta}$ *in* (12) *is given by* $(\tau^2, c_{\mathrm{MS}}, \phi) = (0.65, 0.05, 0.4)$, *and* $c_{\mathrm{ME}} = 0.3$.

2,500 Y-values on the fine-scale grid are used later for assessing the quality of prediction of $Y(\cdot)$ from the artificial data $\mathbf{Z}_g$, defined below.

For the purpose of illustration we choose the simplest case, where we simulate from a $Y(\cdot)$ that has constant mean $\mu = 0$, although when we analyze the dataset, μ is estimated. Further, $C_Y(\cdot)$ is chosen to be an isotropic spherical covariance function given by

$$C_Y(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} \tau^2 + c_{\mathrm{MS}}, & \mathbf{h} = \mathbf{0}, \\ \tau^2\big[1 - 1.5(\|\mathbf{h}\|/\phi) \\ \quad + 0.5(\|\mathbf{h}\|/\phi)^3\big], & 0 < \|\mathbf{h}\| < \phi, \\ 0, & \|\mathbf{h}\| \geq \phi, \end{cases} \tag{12}$$

where $\boldsymbol{\theta} \equiv (\tau^2, c_{\mathrm{MS}}, \phi)'$; $C_Y(\cdot)$ is valid in $\mathbb{R}^1$, $\mathbb{R}^2$ and $\mathbb{R}^3$. For the artificial dataset $\boldsymbol{\theta} = (0.65, 0.05, 0.4)'$, although when we analyze the dataset, $\boldsymbol{\theta}$ is estimated.

To obtain $Z(\cdot)$, we add to $Y(\cdot)$ the measurement-error process $\varepsilon(\cdot)$, a zero-mean Gaussian white-noise process with variance $c_{\mathrm{ME}} = 0.3$. In general, $\mathrm{var}(Z(\mathbf{s})) = \tau^2 + c_{\mathrm{MS}} + c_{\mathrm{ME}}$; for the artificial dataset, $\mathrm{var}(Z(\mathbf{s})) = 1$. Hence, the choice of $c_{\mathrm{ME}} = 0.3$ represents a moderate measurement error component of 30% of the attribute variance. Note that while it is the intended locations $\{\mathbf{s}_i\}$ that we use in our analysis (based on the CP location-error model), the data $\mathbf{Z}_g \equiv (Y(\mathbf{r}_1) + \varepsilon(\mathbf{r}_1), \ldots, Y(\mathbf{r}_{100}) + \varepsilon(\mathbf{r}_{100}))'$ are obtained from actual locations $\{\mathbf{r}_i\}$, as defined by (5).

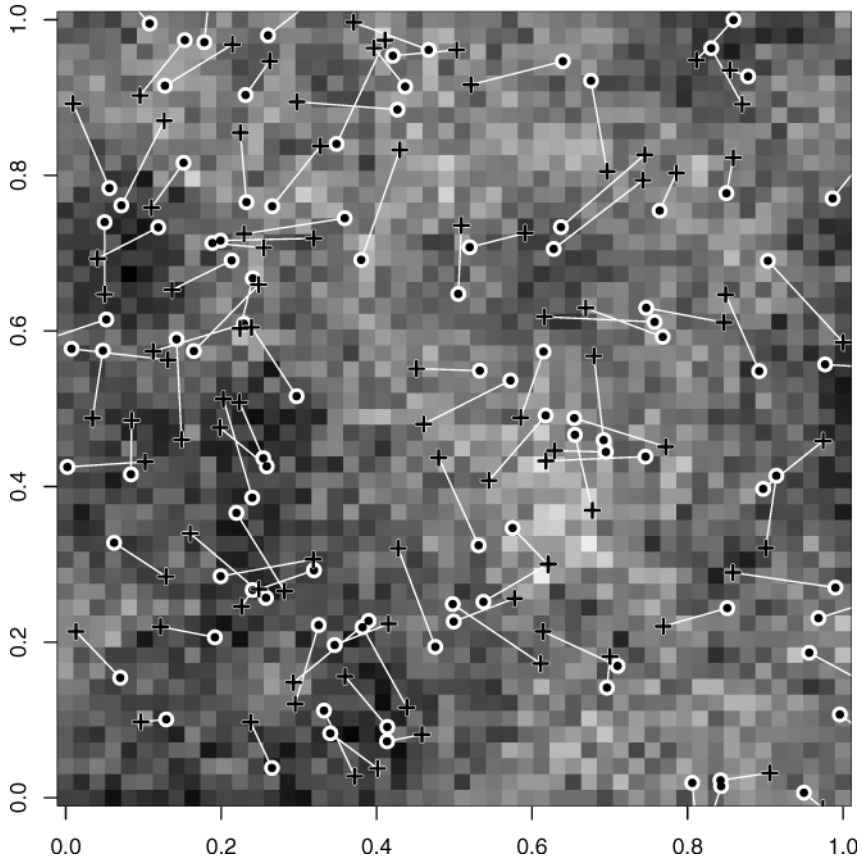

FIG. 2. *Gray-scale image of the artificially generated process $Z(\cdot)$ evaluated at the nodes of a 50×50 grid on $[0,1]^2$. Dots correspond to the intended sites S and pluses correspond to the actual sites R. The lines show the correspondence between intended and actual sites.*

The location-error simulation that yields the perturbations is obtained by sampling independently from a density g defined by the uniform distribution on a disk with center $\mathbf{0}$ and radius $\psi = 0.15$. For general $\psi > 0$ the density is

$$g(\mathbf{u}) = \begin{cases} (\pi\psi^2)^{-1}, & \|\mathbf{u}\| \le \psi, \\ 0, & \|\mathbf{u}\| > \psi. \end{cases} \tag{13}$$

Specifically, let $\{\mathbf{p}_1, \ldots, \mathbf{p}_{100}\}$ be a random sample from (13) with dispersion parameter $\psi = 0.15$. Then the intended locations $\{\mathbf{s}_i\}$ are perturbed to the actual locations $\{\mathbf{r}_i\}$, where $\mathbf{r}_i = \mathbf{s}_i + \mathbf{p}_i$, $i = 1, \ldots, 100$. Figure 2 shows $\{\mathbf{s}_i\}$ and $\{\mathbf{r}_i\}$ superimposed on a discretized gray-scale image of $Z(\cdot) = Y(\cdot) + \varepsilon(\cdot)$.

In the next section, we discuss the effect location error can have on estimation of the parameters contained in the first two moments of the (hidden) spatial process $Y(\cdot)$.

2.3 Parameter Estimation

We have assumed that the spatial process $Y(\cdot)$ is Gaussian, and hence its mean and covariance functions fully specify the distribution. Parameters in the mean function are denoted as $\boldsymbol{\beta}$ and are assumed to appear linearly:

$$E(Y(\mathbf{s})) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}, \quad \boldsymbol{\beta} \in \mathbb{R}^q, \tag{14}$$

where $\mathbf{x}(\mathbf{s})$ is a $(q \times 1)$ vector of covariates at location $\mathbf{s}$. Parameters in the (stationary) covariance function are potentially more problematic, since the only requirement is that they yield a positive-definite covariance function. We write

$$\operatorname{cov}\big(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})\big) = C_Y(\mathbf{h}; \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta. \tag{15}$$

In most geostatistical problems, neither $\boldsymbol{\beta}$ nor $\boldsymbol{\theta}$ is known *a priori*. Provided one can write down the likelihood of $(\boldsymbol{\beta}', \boldsymbol{\theta}')$ and maximize it, the resulting maximum-likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be used to obtain "plug-in" estimates of the mean and covariance functions, which are in turn "plugged into" the kriging equations given in Section 3. Alternatively, one could estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on minimizing generalized and weighted least squares criteria, respectively, and "plugging in" as above (Cressie, 1993, Chapter 2).

In either case, this leads to kriging predictors that are approximately unbiased but kriging standard errors that tend to be smaller than they should be (e.g., Cressie and Zimmerman, 1992). We shall adopt this classical geostatistical strategy here.

Assuming that the original (location-error-free) data are Gaussian, we could use maximum pseudolikelihood estimation (e.g., Carroll and Ruppert, 1988, page 71) to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. To see that the process $Z_g(\cdot)$ is no longer Gaussian in general, and hence that there is a need for a pseudolikelihood, we note that the marginal cumulative distribution function for the process $Z_g(\cdot)$ takes the form

$$\begin{aligned}\Pr\bigl(Z_g(\mathbf{s}) \le z\bigr) &= \int \Pr\bigl(Y(\mathbf{s}+\mathbf{u}) + \varepsilon(\mathbf{s}+\mathbf{u}) \le z\bigr) g(\mathbf{u})\, d\mathbf{u} \\ &= \int\int_{-\infty}^{z} (2\pi C(\mathbf{0}))^{-1/2} \exp\left\{\frac{-(t - \mathbf{x}(\mathbf{s}+\mathbf{u})'\boldsymbol{\beta})^2}{2C(\mathbf{0})}\right\} \\ &\qquad \cdot g(\mathbf{u})\, dt\, d\mathbf{u},\end{aligned}$$

which is not Gaussian unless $\mathbf{x}(\cdot) \equiv 1$ and $g(\cdot)$ is of a conjugate form. This can easily be seen by choosing, for example, the location error to take only two possible distinct values; when $x(\cdot) \not\equiv 1$, the result is a bimodal distribution. In our case, the location error has a density, so that the joint distribution of $\mathbf{Z}_g$ is a continuous mixture of Gaussian densities. Also, although it should be possible to obtain the maximum likelihood estimator using a Monte Carlo algorithm such as simulated annealing (e.g., Geman and Geman, 1984), this will be computationally expensive and negate some of the advantages of the geostatistical approach.

The pseudolikelihood approach we shall use does not depend on Gaussian assumptions, but it does make use of the functional form of a Gaussian density and the mean–variance–covariance formulas given by (9)–(11). According to Carroll and Ruppert (1988, page 71), "Pseudo-likelihood estimates of $\boldsymbol{\theta}$ are based on pretending that the regression parameter $\boldsymbol{\beta}$ is known and equal to the current estimate ... and then estimating $\boldsymbol{\theta}$ by maximum likelihood assuming normality." There are versions of this that one could iterate, using the current estimate of $\boldsymbol{\theta}$ to improve (generalized-least-squares) estimation of $\boldsymbol{\beta}$, which is then used to obtain a new estimate of $\boldsymbol{\theta}$, and so forth. In the spirit of maximum pseudolikelihood, we could simply substitute $E(\mathbf{Z}_g) = X_g\boldsymbol{\beta}$, where $X_g \equiv (\mathbf{x}_g(\mathbf{s}_1), \ldots, \mathbf{x}_g(\mathbf{s}_n))'$, and $\operatorname{var}(\mathbf{Z}_g) \equiv \Sigma_g$, which depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (Section 2.1), into the Gaussian density obtained from a Gau$(E(\mathbf{Z}_g), \operatorname{var}(\mathbf{Z}_g))$ distribution. Define

$$\begin{aligned}&\ell_g(\boldsymbol{\beta}, \boldsymbol{\theta}) \\ &\equiv (2\pi)^{-n/2} |\Sigma_g|^{-1/2} \\ &\qquad \cdot \exp\bigl\{-\tfrac{1}{2}(\mathbf{Z}_g - X_g\boldsymbol{\beta})'\Sigma_g^{-1}(\mathbf{Z}_g - X_g\boldsymbol{\beta})\bigr\}.\end{aligned} \tag{16}$$

The resulting estimates are given by

$$(\widehat{\boldsymbol{\beta}}_g', \widehat{\boldsymbol{\theta}}_g') = \arg \sup_{(\boldsymbol{\beta}', \boldsymbol{\theta}')} \ell_g(\boldsymbol{\beta}, \boldsymbol{\theta}). \tag{17}$$

Notice that $\widehat{\boldsymbol{\theta}}_g = \arg\sup_{\boldsymbol{\theta}} \ell_g(\widehat{\boldsymbol{\beta}}_g(\boldsymbol{\theta}), \boldsymbol{\theta})$, which gives the connection with the more usual (iterated) maximum pseudolikelihood estimates defined directly below.

In the case of the artificial dataset, recall that the mean function of $Y(\cdot)$ is constant, equal to μ. Then this maximum pseudolikelihood estimator for μ and $\boldsymbol{\theta}$ is based on the n-variate Gaussian density with $(n \times 1)$ mean vector $(\mu, \ldots, \mu)'$, and $(n \times n)$ variance–covariance matrix Σ_g obtained from variance function

$$v_g(\mathbf{s}) = C_Y(\mathbf{0}; \boldsymbol{\theta}) + c_{\mathrm{ME}}, \quad \mathbf{s} \in D,$$

and covariance function

$$C_g(\mathbf{h}) = \int\int C_Y(\mathbf{h} + \mathbf{v} - \mathbf{u}; \boldsymbol{\theta}) g(\mathbf{u}) g(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v};$$

here $C_Y(\cdot; \boldsymbol{\theta})$ is given by (12) and $g(\cdot)$ is given by (13) with $\psi = 0.15$.

We compare this to the more usual iterated pseudolikelihood procedure, where first the mean-function parameter $\boldsymbol{\beta}$ is estimated by ordinary least squares, yielding $\boldsymbol{\beta}_g^0 \equiv (X_g'X_g)^{-1}X_g'\mathbf{Z}_g$. Then

$$\boldsymbol{\theta}_g^0 \equiv \arg\sup_{\boldsymbol{\theta}} \ell_g(\boldsymbol{\beta}_g^0, \boldsymbol{\theta})$$

is an estimator of $\boldsymbol{\theta}$ based on the pseudolikelihood. Substituting $\boldsymbol{\beta}_g^0$ and $\boldsymbol{\theta}_g^0$ into Σ_g to obtain $\widetilde{\Sigma}_g$, we compute a generalized-least-squares estimator, $\widetilde{\boldsymbol{\beta}}_g \equiv (X_g'\widetilde{\Sigma}_g^{-1}X_g)^{-1}X_g'\widetilde{\Sigma}_g^{-1}\mathbf{Z}_g$. Finally, $\widetilde{\boldsymbol{\theta}}_g$ is obtained analogously to $\boldsymbol{\theta}_g^0$ but with $\boldsymbol{\beta}_g^0$ replaced by $\widetilde{\boldsymbol{\beta}}_g$ in the formula displayed immediately above.

For the artificial dataset described in Section 2.2 the mean function is constant, equal to μ; Table 1 compares the maximum pseudolikelihood estimates $\widehat{\mu}_g$ and $\widehat{\boldsymbol{\theta}}_g$ [given by (17)] with iterated versions $\widetilde{\mu}_g$ and $\widetilde{\boldsymbol{\theta}}_g$, for this dataset. Although we have no standard errors with which to gauge the differences, a small-scale sensitivity study indicated that there is little to choose between the two estimation procedures, and they did not appear to be biased over simulations of like datasets using the same $\boldsymbol{\theta} = (0.65, 0.05, 0.4)'$ and $\psi = 0.15$. In all that is to follow we use the maximum pseudolikelihood estimates given by (17), which we refer to as PALE (parameter estimation adjusting for

TABLE 1
Parameter estimates for the artificial dataset: maximum pseudolikelihood estimates $\widehat{\mu}_g$ and $\widehat{\boldsymbol{\theta}}_g$, and iterated maximum pseudolikelihood estimates $\widetilde{\mu}_g$ and $\widetilde{\boldsymbol{\theta}}_g$

Parameter	$\boldsymbol{\mu}, \boldsymbol{\theta}$	$\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{g}}$	$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{g}}, \widetilde{\boldsymbol{\theta}}_{\boldsymbol{g}}$
μ	0	−0.3041	−0.3040
τ^2	0.65	0.7011	0.7007
c_{MS}	0.05	0.1007	0.1012
ϕ	0.4	0.2813	0.2813

location error) estimates. This is to be contrasted with PILE (parameter estimation ignoring location error) estimates, which are given by

$$(\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\theta}}') = \arg \sup_{(\boldsymbol{\beta}', \boldsymbol{\theta}')} \big[(2\pi)^{-n/2}|\Sigma|^{-1/2} \cdot \exp\{(-\tfrac{1}{2})(\mathbf{Z}_g - X\boldsymbol{\beta})' \cdot \Sigma^{-1}(\mathbf{Z}_g - X\boldsymbol{\beta})\}\big]. \tag{18}$$

Importantly, in (18), $X = (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$ and $\Sigma = (C(\mathbf{s}_i - \mathbf{s}_j))$, where recall that $C(\mathbf{h}) = C_Y(\mathbf{h}; \boldsymbol{\theta}) + c_{\mathrm{ME}} I(\mathbf{h} = \mathbf{0})$, which are both obtained by ignoring the presence of location error.

Note that maximum pseudolikelihood estimation (and any other type of parameter estimation for that matter) is only able to estimate $(c_{\mathrm{MS}} + c_{\mathrm{ME}})$, but it cannot disentangle c_{MS} and c_{ME} unless either the measurement-error variance (c_{ME}) or the microscale variance (c_{MS}) is known. Estimation of measurement error can be obtained via repeated measurements at exactly the same location, and estimation of the microscale variance requires measurements between sites within very small distances of each other. It is more likely that we have external information on c_{ME}, such as instrument-precision specifications from the instrument manufacturer, and hence, by subtraction, c_{MS} [a parameter of the hidden process $Y(\cdot)$] can be estimated. In analyzing the artificial dataset, we have used the knowledge that $c_{\mathrm{ME}} = 0.3$, the true value, and then estimated $\boldsymbol{\theta} = (\tau^2, c_{\mathrm{MS}}, \phi)'$; see Table 1.

We performed another sensitivity study, this time to assess the variability of the PALE estimators $\widehat{\mu}_g$ and $\widehat{\boldsymbol{\theta}}_g$, and the PILE estimators $\widehat{\mu}$ and $\widehat{\boldsymbol{\theta}}$, of μ and $\boldsymbol{\theta}$. The study consisted of generating 50 datasets in the same manner as the zero-trend simulations given in Section 4, and obtaining PALE and PILE estimates of μ and $\boldsymbol{\theta}$ for each dataset. Figure 3 shows histograms

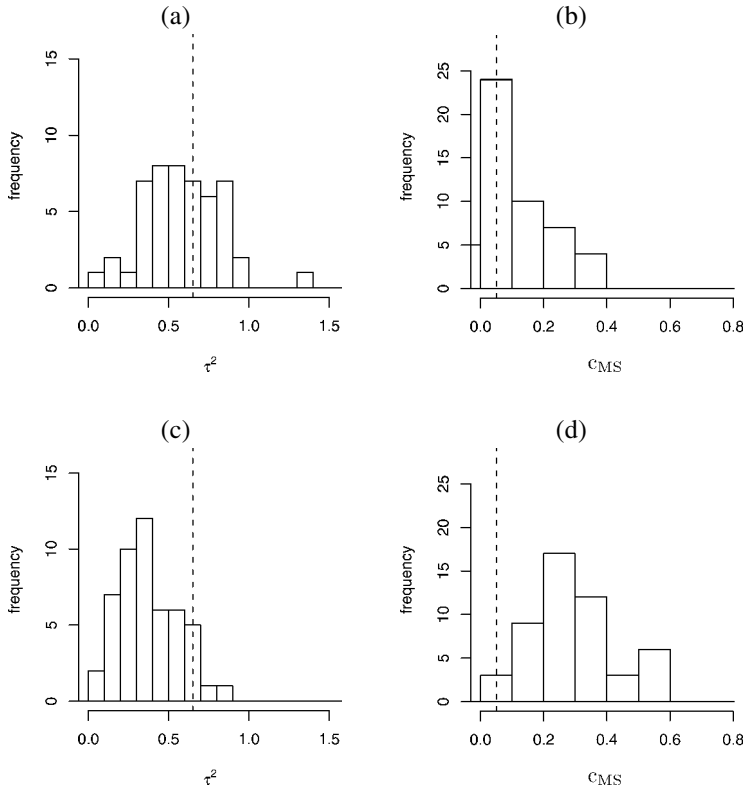

FIG. 3. *Histograms of estimates of τ^2 and c_{MS}, for PALE and PILE, with the true value indicated by a vertical dashed line*: (a) *PALE for τ^2*; (b) *PALE for c_{MS}*; (c) *PILE for τ^2*; (d) *PILE for c_{MS}*.

of the PALE and PILE estimates for τ^2 and c_{MS}. The PALE estimates do not appear to be biased, whereas the PILE estimates appear to underestimate τ^2 and overestimate c_{MS}. Since PILE (inappropriately) ignores the presence of location error, the data $\mathbf{Z}_g$ will indicate a spatial covariance function with τ^2 smaller than 0.65 [see Figure 1(c)] and consequently a c_{MS} that is larger than 0.05. The estimates for the other two parameters, μ and ϕ, were similar for both PALE and PILE, in that their histograms (not shown) did not indicate any bias and showed similar levels of variability. As the location-error dispersion parameter ψ increases, we expect to see more variability in the estimates, since the data $\mathbf{Z}_g$ have a spatial covariance that is obtained from more integration; see how Figure 1(a) is integrated to become Figures 1(b)–(d), for $\psi = 0.05, 0.15, 0.25$, respectively.

3. KRIGING ADJUSTING FOR LOCATION ERROR

Kriging in geostatistics is nothing more than best linear prediction in statistical linear-model theory, where the variances and covariances are derived from the observed and hidden spatial processes (Cressie, 1993, Chapter 3). We now give a brief description of kriging, without location error, in order to provide comparison for the subsequent derivation of kriging adjusting for location error.

3.1 Optimal Spatial Prediction—No Location Error

The geostatistical model upon which the usual kriging equations are based is

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{19}$$

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \nu(\mathbf{s}), \tag{20}$$

where $Y(\cdot)$ is the (hidden) process of scientific interest that we wish to predict from a finite number of observations $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$ on the (noisy) process $Z(\cdot)$. Let $\mathbf{s}_0 \in \mathbb{R}^d$ be a generic location. Then the simple-kriging equations given below provide the best linear predictor

$$p(\mathbf{Z}; \mathbf{s_0}) \equiv \boldsymbol{\ell}'\mathbf{Z} + k \tag{21}$$

of $Y(\mathbf{s}_0)$. Cressie (1993, page 173) has shown that the coefficients in (21) are given by

$$\boldsymbol{\ell}' = \mathbf{c}(\mathbf{s}_0)'\Sigma^{-1} \quad \text{and} \quad k = \big(\mathbf{x}(\mathbf{s}_0) - X'\boldsymbol{\ell}\big)'\boldsymbol{\beta}, \tag{22}$$

where $\Sigma \equiv \mathrm{var}(\mathbf{Z})$ is an $(n \times n)$ variance matrix, $\mathbf{c}(\mathbf{s}_0) \equiv \mathrm{cov}(Y(\mathbf{s}_0), \mathbf{Z})$ is an $(n \times 1)$ vector, $\mathbf{x}(\mathbf{s}_0)$ is the $(q \times 1)$ vector of covariates at $\mathbf{s}_0$ and $X \equiv (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$ is the $(n \times q)$ matrix of covariates at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. That is, the best linear predictor (21), (22) depends on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ constitutes the variance–covariance parameters.

To obtain a predictor that is a function only of the data, it is natural to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, substitute them into $\mathbf{c}(\mathbf{s}_0)$ and Σ [yielding $\widehat{\mathbf{c}}(\mathbf{s}_0)$ and $\widehat{\Sigma}$] and then substitute the estimates into (21) and (22). Choice of the generalized-least-squares estimate for $\boldsymbol{\beta}$,

$$\widehat{\boldsymbol{\beta}} = (X'\widehat{\Sigma}^{-1}X)^{-1}X'\widehat{\Sigma}^{-1}\mathbf{Z}, \tag{23}$$

yields the universal-kriging predictor (Cressie, 1993, page 173). An estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained in any of a number of ways, including Gaussian maximum likelihood (Cressie, 1993, Section 2.6). The mean squared prediction error (MSPE) of (21) for known $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is

$$\begin{aligned}\mathrm{MSPE}(\mathbf{s}_0) &\equiv E\big(Y(\mathbf{s}_0) - \boldsymbol{\ell}'\mathbf{Z} - k\big)^2 \\ &= C_Y(\mathbf{0}) - \mathbf{c}(\mathbf{s}_0)'\Sigma^{-1}\mathbf{c}(\mathbf{s}_0).\end{aligned} \tag{24}$$

The MSPE (24) depends on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and again it is natural to substitute the estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in (24). Note that the resulting estimated $\mathrm{MSPE}(\mathbf{s}_0)$ will be an underestimate due to unaccounted-for uncertainty inherited from parameter estimation (Harville, 1985; Cressie and Zimmerman, 1992).

The kriging equations depend on Σ and $\mathbf{c}(\mathbf{s}_0)$, whose entries are functions of spatial lags; in the presence of location error these spatial lags are random. The next section gives the appropriate adjustments for kriging when location error is present.

3.2 Optimal Spatial Prediction—Location Error Present

Assume the CP model for location error, and recall the notation established in Section 2. We have observations $\mathbf{Z}_g$ on the process $Z_g(\mathbf{s}) \equiv Z(\mathbf{s} + \mathbf{p}(\mathbf{s}))$, $\mathbf{s} \in D$, from which we wish to predict $Y(\mathbf{s}_0)$. Using the class of linear predictors,

$$p_g(\mathbf{Z}_g; \mathbf{s}_0) \equiv \boldsymbol{\ell}_g'\mathbf{Z}_g + k_g, \tag{25}$$

the same algebra as above yields the optimal coefficients as

$$\begin{aligned}\boldsymbol{\ell}_g' &= \mathbf{c}_g(\mathbf{s}_0)'\Sigma_g^{-1} \quad \text{and} \\ k_g &= \big(\mathbf{x}(\mathbf{s}_0) - X_g'\boldsymbol{\ell}_g\big)'\boldsymbol{\beta},\end{aligned} \tag{26}$$

where $\Sigma_g \equiv \mathrm{var}(\mathbf{Z}_g)$, $\mathbf{c}_g(\mathbf{s}_0) \equiv \mathrm{cov}(Y(\mathbf{s}_0), \mathbf{Z}_g)$, $E(Y(\mathbf{s}_0)) = \mathbf{x}(\mathbf{s}_0)'\boldsymbol{\beta}$, $E(\mathbf{Z}_g) = X_g\boldsymbol{\beta}$, and recall that

$X_g = (\mathbf{x}_g(\mathbf{s}_1), \ldots, \mathbf{x}_g(\mathbf{s}_n))'$. From Section 2.1 and similar calculations we have

$$\operatorname{var}(Z_g(\mathbf{s}_i)) = C_Y(\mathbf{0}; \boldsymbol{\theta}) + c_{\mathrm{ME}} + \boldsymbol{\beta}' M_g(\mathbf{s}) \boldsymbol{\beta},$$

$$\operatorname{cov}\big(Z_g(\mathbf{s}_i), Z_g(\mathbf{s}_j)\big) = \iint C_Y(\mathbf{s}_j - \mathbf{s}_i + \mathbf{v} - \mathbf{u}; \boldsymbol{\theta}) \cdot g(\mathbf{u}) g(\mathbf{v})\, d\mathbf{u}\, d\mathbf{v},$$

$$\operatorname{cov}\big(Y(\mathbf{s}_0), Z_g(\mathbf{s}_i)\big) = \int C_Y(\mathbf{s}_i - \mathbf{s}_0 + \mathbf{u}; \boldsymbol{\theta}) g(\mathbf{u})\, d\mathbf{u},$$

$$\mathbf{x}_g(\mathbf{s}_i) = \int \mathbf{x}(\mathbf{s}_i + \mathbf{u}) g(\mathbf{u})\, d\mathbf{u}; \quad i = 1, \ldots, n,\ j = 1, \ldots, n.$$

We refer to the optimal predictor (25) and (26) as kriging adjusting for location error (KALE). Notice that KALE is unbiased; that is, $E(p_g(\mathbf{Z}_g; \mathbf{s}_0)) = E(Y(\mathbf{s}_0))$.

As in Section 3.1, (25) and (26) depend on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is not as straightforward as in the case where there is no location error, because $\boldsymbol{\beta}$ appears in both the mean function, $\mu_g(\mathbf{s}) \equiv E(Z_g(\mathbf{s}))$, and in the variance function, $v_g(\mathbf{s}) \equiv \operatorname{var}(Z_g(\mathbf{s}))$. In particular, there is no universal-kriging predictor for $Y(\mathbf{s}_0)$, and Gabrosek and Cressie's (2002) universal-kriging equations are not correct. However, their ordinary kriging equations *are* correct, as we now demonstrate: if $E(Z(\mathbf{s})) = \mu$, then from (9), $\mu_g(\mathbf{s}) \equiv \mu$, and from (11), $v_g(\mathbf{s}) \equiv C_Y(\mathbf{0}; \boldsymbol{\theta}) + c_{\mathrm{ME}}$, which is independent of μ.

The KALE predictor (25) and (26), and the KALE MSPE,

$$\begin{aligned} \mathrm{MSPE}_g(\mathbf{s}_0) &\equiv E\big(Y(\mathbf{s}_0) - \boldsymbol{\ell}_g' \mathbf{Z}_g - k_g\big)^2 \\ &= C_Y(\mathbf{0}) - \mathbf{c}_g(\mathbf{s}_0)' \Sigma_g^{-1} \mathbf{c}_g(\mathbf{s}_0), \end{aligned} \tag{27}$$

both depend on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Section 2.3 discusses how to obtain pseudolikelihood-based estimates $\widehat{\boldsymbol{\beta}}_g$ and $\widehat{\boldsymbol{\theta}}_g$ of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, which we substitute into (25)–(27) to obtain quantities that are functions only of the data.

3.3 Kriging the Artificial Dataset

We wish to compute the optimal spatial predictor for the artificial dataset described in Section 2.2. Because in this example we actually know the true (hidden) process $Y(\cdot)$, there is an opportunity to get some idea of how well KALE performs.

Recall that Figure 2 shows the intended locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_{100}\}$ as dots superimposed on $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, where D is discretized onto S, a 50×50 grid. The dataset $\mathbf{Z}_g$ is actually observed at $\{\mathbf{r}_1, \ldots, \mathbf{r}_{100}\}$, shown in Figure 2 as pluses, but for KALE these locations are assumed unknown. From the data $\mathbf{Z}_g$ and the formulas for KALE given in Section 3.2, we obtain $\{p_g(\mathbf{Z}_g; \mathbf{s}) : \mathbf{s} \in D\}$, which is compared to the true process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ in Figure 4. In practice, the PALE estimates $\widehat{\mu}_g$ and $\widehat{\boldsymbol{\theta}}_g$ given by (17) are substituted into (25) and (26) to yield the KALE map.

From Figure 4, we see the smoothing effect that is typical of kriging. The true process $Y(\cdot)$, shown in Figure 4(a), is much more variable than the KALE map $p_g(\mathbf{Z}_g; \cdot)$, shown in Figure 4(b). Furthermore, by inspecting similar plots for ψ other than 0.15, we conclude that the larger the location-error dispersion, the smoother is KALE.

To ascertain the impact of variation in parameter estimates on KALE, we looked at the squared prediction error of KALE as a function of prediction location, for various choices of parameters. Five vectors (one consisting of the true parameter values) were chosen to mimic sample variability of the parameter estimates for the artificial dataset. The average (here, an average

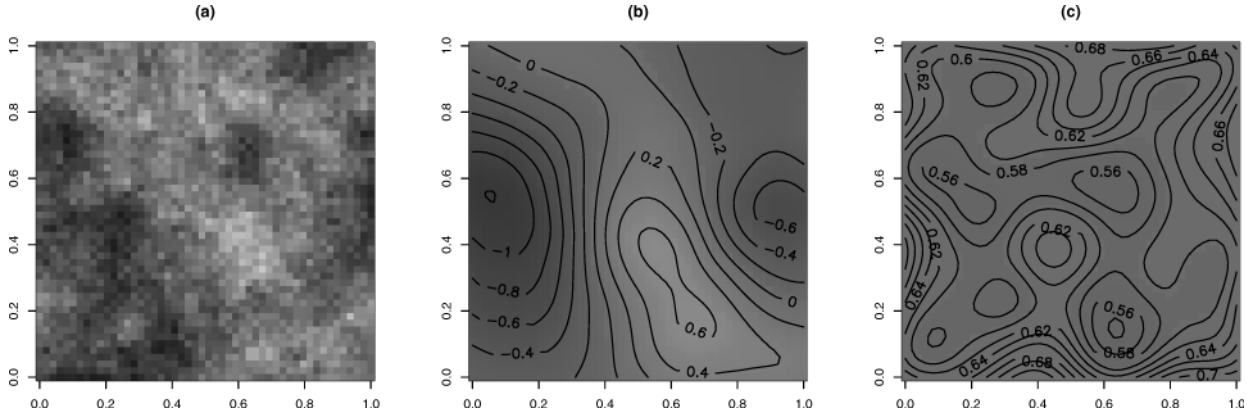

FIG. 4. (a) *Hidden process* $Y(\cdot)$, *from which artificial data* $\mathbf{Z}_g$ *were generated*; (b) *KALE map given by* $p_g(\mathbf{Z}_g; \cdot)$; (c) $\{\mathrm{MSPE}_g(\cdot)\}^{1/2}$ *given by* (24). *In* (b) *and* (c) $\widehat{\boldsymbol{\beta}}_g$ *and* $\widehat{\boldsymbol{\theta}}_g$ *are substituted for unknown* $\boldsymbol{\beta}$ *and* $\boldsymbol{\theta}$.

over $\mathbf{s}_0 \in S$) squared prediction error based on the vector of true parameter values was 0.3478, which is to be compared with 0.3387 based on the vector of estimates obtained from PALE; the ratio is 1.020. The remaining four vectors of parameters yielded ratios of 1.013, 0.960, 1.028 and 0.995. The choice of these four parameter vectors, based on histograms of parameter estimates like those seen in Figure 3, ensured wide coverage of the parameter space. That the ratios hardly deviate from 1.000 indicates rather weak dependence of KALE on the parameters substituted into (25)–(27). This lack of sensitivity will be exploited in the simulation experiment described in Section 4.

4. SPATIAL-PREDICTION PROPERTIES OF KALE—A SIMULATION EXPERIMENT

The methodology of kriging adjusting for location error, as presented in Section 3, is split into two steps. Like most geostatistical methods for mapping, the data $\mathbf{Z}_g$ are used twice. First, the optimal predictor, from the class of all predictors linear in $\mathbf{Z}_g$, is derived; see Section 3.2. The coefficients of the optimal linear unbiased predictor, given by (26), and the MSPE, given by (27), depend on unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The second step reuses the data for estimation of any unknown parameters, which are then substituted into (25)–(27). The result is the spatial-prediction methodology that we call KALE.

We also consider a methodology that we call KILE (kriging ignoring location error). That is, we treat $\mathbf{Z}_g$ as if it were observed with no location error: in the first step of KILE, the predictor is $p(\mathbf{Z}_g; \mathbf{s}_0)$ given by (21) and (22). Notice that $E(p(\mathbf{Z}_g; (\mathbf{s}_0))) = \mathbf{x}(\mathbf{s}_0)'\boldsymbol{\beta} + \mathbf{c}(\mathbf{s}_0)'\Sigma^{-1}(X_g - X)\boldsymbol{\beta}$, and $E(Y(\mathbf{s}_0)) = \mathbf{x}(\mathbf{s}_0)'\boldsymbol{\beta}$. Hence, KILE yields a predictor that is generally biased; define $\mathrm{BIAS}(\mathbf{s}_0) \equiv \mathbf{c}(\mathbf{s}_0)'\Sigma^{-1}(X_g - X)\boldsymbol{\beta}$. If one were to ignore location error to compute the MSPE using (24), it would be an underestimate; the true MSPE, which we denote as $\mathrm{MSPE}(\mathbf{s}_0)$, is given by

$$\begin{aligned} &C_Y(\mathbf{0}) - 2\mathbf{c}_g(\mathbf{s}_0)'\Sigma^{-1}\mathbf{c}(\mathbf{s}_0) \\ &\quad + \mathbf{c}(\mathbf{s}_0)'\Sigma^{-1}\Sigma_g\Sigma^{-1}\mathbf{c}(\mathbf{s}_0) \\ &\quad + \big(\mathbf{c}(\mathbf{s}_0)'\Sigma^{-1}(X_g - X)\boldsymbol{\beta}\big)^2; \end{aligned} \tag{28}$$

the last term is a squared bias component that depends on $\boldsymbol{\beta}$. The KILE predictor $p(\mathbf{Z}_g; \mathbf{s}_0)$, $\mathrm{BIAS}(\mathbf{s}_0)$ and $\mathrm{MSPE}(\mathbf{s}_0)$ all depend on unknown parameters, so that in the second step of KILE these parameters are estimated, again ignoring location error [e.g., PILE estimates given by (18)]; then the estimates are substituted into (21), (22) and (28).

We have conducted a simulation experiment that enables us to assess the effect of no, small, medium and large location-error dispersion on the bias and mean squared prediction error of KALE and KILE; see Section 4.1. The results of the experiment show the improvement of KALE over KILE as a function of the location-error dispersion; see Section 4.2.

4.1 Design of the Simulation Experiment

Recall the description of the artificial dataset given in Section 2.2. The simulation experiment is very similar, in that the process $Z_g(\cdot)$ was simulated on $[0, 1]^2$ in the same way except, as part of our design, we considered a mean function with linear trend as well as a mean function with no trend (i.e., constant mean). That is, we first simulated $Z(\cdot)$ with two choices for $\mu(s^x, s^y) \equiv E(Z(s^x, s^y))$:

$$\begin{aligned} &\mu(s^x, s^y) = 0 \quad \text{or} \\ &\mu(s^x, s^y) = -0.125 + 0.25s^x, \end{aligned} \tag{29}$$

where $\mathbf{s} \equiv (s^x, s^y) \in [-0.25, 1.25]^2$, a slightly larger region than the unit square to account for location errors. The covariance function of the hidden process $Y(\cdot)$ is given by (12), with $\boldsymbol{\theta} = (\tau^2, c_{\mathrm{MS}}, \phi)' = (0.65, 0.05, 0.4)'$, and the measurement-error variance is $c_{\mathrm{ME}} = 0.3$.

Recall that the process $Z_g(\cdot)$ is defined by

$$Z_g(\mathbf{s}) \equiv Z\big(\mathbf{s} + p(\mathbf{s})\big), \quad \mathbf{s} \in D,$$

where $\{\mathbf{p}(\mathbf{s})\}$ are i.i.d. with density g that is uniform on a disk centered at $\mathbf{0}$ and of radius ψ; see (13). We consider four levels of location error: the first is where $\mathbf{p}(\mathbf{s}) \equiv \mathbf{0}$, which we write as $\psi = 0$; the other three are $\psi = 0.05$, 0.15 and 0.25, corresponding to small, medium and large location-error dispersion, respectively. Some control was imposed on the way the different levels were simulated. The intended sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_{100}\}$ were chosen by sampling uniformly from the unit square $[0, 1]^2$; they were then fixed throughout the experiment. In the case of no location error, $\mathbf{Z}_g = \mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_{100}))'$. In the case of small location-error dispersion, $\{\mathbf{p}(\mathbf{s}_1), \ldots, \mathbf{p}(\mathbf{s}_{100})\}$ were randomly sampled from $g(\cdot)$ with $\psi = 0.05$. That is, for small location error, we obtained the simulated data, $\mathbf{Z}_g \equiv (Z(\mathbf{r}_1), \ldots, Z(\mathbf{r}_{100}))'$, where $\mathbf{r}_i = \mathbf{s}_i + \mathbf{p}(\mathbf{s}_i)$, $i = 1, \ldots, 100$. Then, for medium location error, we obtained $\mathbf{Z}_g$ from actual sites $\mathbf{r}_i = \mathbf{s}_i + 3\mathbf{p}(\mathbf{s}_i)$, $i = 1, \ldots, 100$; and for large location error, we obtained $\mathbf{Z}_g$ from actual sites $\mathbf{r}_i = \mathbf{s}_i + 5\mathbf{p}(\mathbf{s}_i)$, $i = 1, \ldots, 100$.

The sources of randomness in the experiment were from the generation of new processes $Y(\cdot)$, $\varepsilon(\cdot)$ and $\mathbf{p}(\cdot)$ for each of L simulations. While the KALE and KILE procedures were quite time-consuming, we found that parameter estimation of $\boldsymbol{\theta}$ was prohibitive for L large. Consequently, for the simulation experiment we chose to estimate $\boldsymbol{\beta}$ (using generalized least squares) and fix $\boldsymbol{\theta}$ at its true value. [In the study described in Section 3.3, we verified a lack of sensitivity to the choice of parameters.] All the responses from the simulation experiment relate to the performance of KALE and KILE for *spatial prediction* of the hidden process $Y(\cdot)$.

KALE and KILE were evaluated on S, which recall is a 50×50 grid in D. At each location $\mathbf{s}_0 \in S$ we know the true value $Y(\mathbf{s}_0)$, which we compare to the KALE predictor $p_g(\mathbf{Z}_g; \mathbf{s}_0)$ given by (25) and (26), and the KILE predictor $p(\mathbf{Z}_g; \mathbf{s}_0)$ given by (21) and (22), as follows. Let $\widehat{Y}^{(\ell)}(\mathbf{s}_0)$ denote a generic predictor of $Y^{(\ell)}(\mathbf{s}_0)$, for simulation $\ell = 1, \ldots, L$. Then we compute the empirical prediction bias,

$$\text{EBIAS}(\mathbf{s}_0) \equiv \frac{1}{L}\sum_{\ell=1}^{L}\big(\widehat{Y}^{(\ell)}(\mathbf{s}_0) - Y^{(\ell)}(\mathbf{s}_0)\big), \quad \mathbf{s}_0 \in S,$$

and the empirical mean squared prediction error,

$$\text{EMSPE}(\mathbf{s}_0) \equiv \frac{1}{L}\sum_{\ell=1}^{L}\big(\widehat{Y}^{(\ell)}(\mathbf{s}_0) - Y^{(\ell)}(\mathbf{s}_0)\big)^2, \quad \mathbf{s}_0 \in S,$$

and use them to compare KALE to KILE. In our simulation study, L was chosen to be 5,000.

4.2 Results from the Simulation Experiment

We discuss principally the no-trend case and compare it to the linear-trend case later. After looking carefully at the simulation results for EBIAS$(\cdot)$, we concluded that KALE and KILE exhibited no bias, which agrees with the theory given in Section 3.2 and at the beginning of this section.

A comparison of KALE's theoretical MSPE given by (27) and its EMSPE defined in Section 4.1 showed good agreement. Similarly good agreement was shown between KILE's theoretical MSPE given by (28) and its EMSPE. In all that is to follow, we compare KALE to KILE via their EMSPE's obtained from the simulation experiment.

In Figure 5 we give for $\psi = 0.15$ a gray-scale image of

$$\log\left\{\frac{\text{EMSPE}(\mathbf{s}_0) \text{ for KILE}}{\text{EMSPE}(\mathbf{s}_0) \text{ for KALE}}\right\},$$

which can be interpreted as a log relative efficiency.

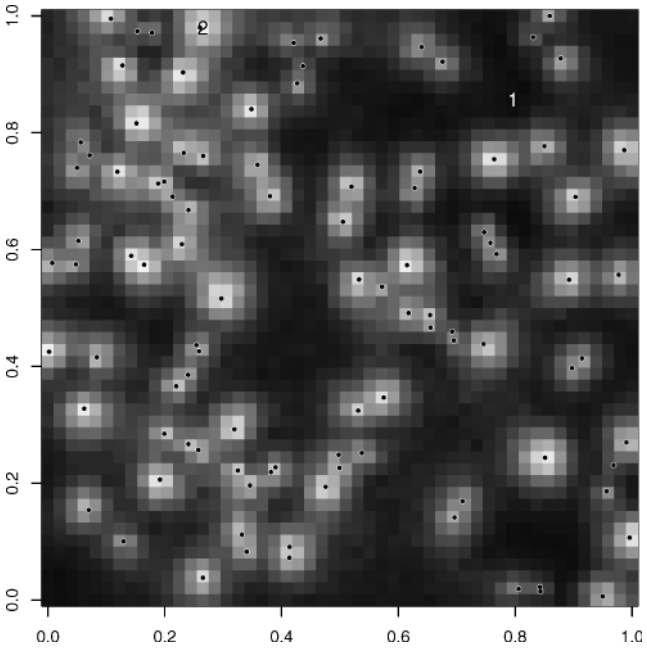

FIG. 5. *Gray-scale image of* log *relative efficiency of KALE with respect to KILE for the simulation experiment. White/light-grey represents high efficiency above* 100%; *black/dark-gray represents efficiencies around* 100%; *values on the gray scale are* log{(*empirical MSPE for KILE*)/(*empirical MSPE for KALE*)}. *Superimposed on the image are the intended sites and two other sites* (1 *and* 2) *used in the text for efficiency comparisons.*

Superimposed on the image are the intended sites $\mathbf{s}_1, \ldots, \mathbf{s}_{100}$ (fixed throughout the experiment). It is apparent from the figure that proximity of the prediction point $\mathbf{s}_0$ to an intended site, as well as that site's degree of isolation, are important factors that help explain high (log) relative efficiencies. Consequently, Figure 6 shows plots of log relative efficiency as a function of distance to nearest intended site, for (a) $\psi = 0.05$, (b) $\psi = 0.15$ and (c) $\psi = 0.25$.

For a given location-error dispersion ψ, it is clear from Figure 6 that the efficiency of KALE relative to KILE increases as distance to the nearest intended site decreases; also notice that KALE is typically as efficient or more efficient than KILE. The effect of increasing ψ is quite remarkable; for $\psi = 0.15$ and 0.25, all relative efficiencies are greater than 100%, with some efficiencies more than 120%. To illustrate the potential of KALE, we chose two prediction locations that give quite different results. Point 1 in Figure 5 is far from an intended site and had relative efficiencies of 100.0% ($\psi = 0.05$), 100.4% ($\psi = 0.15$) and 103.5% ($\psi = 0.25$), indicating that for this point there was little benefit in adjusting for location error. In contrast, point 2 in Figure 5 is very close to an isolated

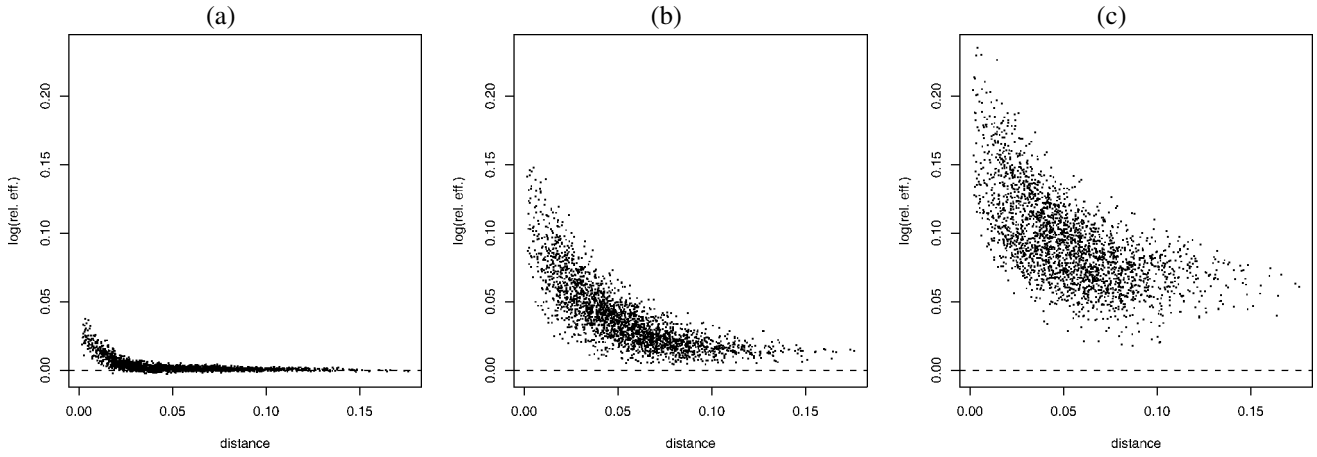

FIG. 6. *Panels* (a)–(c) *show plots of* log *relative efficiency* (*see Figure* 5) *versus distance to nearest intended site*: (a) $\psi = 0.05$; (b) $\psi = 0.15$; (c) $\psi = 0.25$.

intended site and had relative efficiencies of 102.6% ($\psi = 0.05$), 115.9% ($\psi = 0.15$) and 121.6% ($\psi = 0.25$). These types of efficiency gains are substantial and worth the trouble of adjusting for location error [i.e., using KALE to predict $Y(\mathbf{s}_0)$].

A global comparison of KALE and KILE is given in Table 2. There we see an average over S of the empirical and theoretical MSPEs for KALE and KILE. The theoretical values are very close to the empirical values, but always a little smaller since they do not account for parameter estimation. The ratio of the average of EMSPE$(\cdot)$ for KILE to the same for KALE is 100.2% ($\psi = 0.05$), 104.0% ($\psi = 0.15$) and 109.8% ($\psi = 0.25$).

Maps of EBIAS$(\cdot)$ and EMSPE$(\cdot)$, along with global comparisons analogous to those given in Table 2, were studied for the linear-trend case. We do not present them here, but rather summarize the effect on KALE and KILE of adding a linear trend term. In the linear-trend case, we observe from the maps of EBIAS$(\cdot)$ that both KALE's and KILE's bias are very small [compared to EMSPE$(\cdot)$] but increasing in the opposite direction to the trend in $Y(\cdot)$. That there is bias is not unexpected since $\boldsymbol{\beta}$ is not assumed known in the simulation experiment; rather, it is estimated.

TABLE 2
Entries are ave$\{\text{EMSPE}(\boldsymbol{s}_0) : \boldsymbol{s}_0 \in S\}$ *for KALE and KILE in the no-trend case*; *average theoretical MSPEs are shown in parentheses*

ψ	KALE	KILE
0	0.2905 (0.2888)	0.2905 (0.2888)
0.05	0.3029 (0.3011)	0.3037 (0.3021)
0.15	0.3750 (0.3719)	0.3900 (0.3879)
0.25	0.4740 (0.4692)	0.5204 (0.5164)

From looking at the maps of EMSPE$(\cdot)$, there were imperceptable differences between the no-trend and linear-trend cases. However, global comparisons like those given in Table 2 showed an increase of 1–2% in ave$\{\text{EMSPE}(\mathbf{s}_0) : \mathbf{s}_0 \in S\}$ for the linear-trend case over the no-trend case (for both KALE and KILE). For more general trend surfaces, we expect that a comparison of trend versus no-trend will lead to bigger differences for KILE and KALE, that is, higher efficiencies for KALE relative to KILE. This is because the last term in KILE's theoretical MSPE given by (28) depends on $\boldsymbol{\beta}$ and could be substantial.

5. REMOTELY SENSED TOTAL COLUMN OZONE

Ozone is important for protecting the Earth's surface from ultraviolet and other radiation in the 290–400 nm wavelength. An excess of this type of radiation can damage DNA and cellular proteins involved in biochemical processes for growth and reproduction. Total column ozone is the amount of ozone in a column starting from a point on the Earth's surface and integrating through all levels of the atmosphere; it is measured in Dobson units. To illustrate a geostatistical analysis in the presence of location error, we use data from the Total Ozone Mapping Spectrometer (TOMS) on the *Nimbus-7* satellite, obtained from the NASA Goddard Distributed Active Archive Center (http://daac.gsfc.nasa.gov/).

The *Nimbus-7* satellite is polar-orbiting and covers the entire globe in a 24-hour period. We chose a

region of interest D over ocean off the coast of Chile [Figure 7(a)] because of obvious trend in the data. The actual locations $\{\mathbf{r}_i\}$ of the TOMS measurements come from satellite swaths recorded on October 1, 1988. They are shown as pluses in Figure 7(b), and they are in the range $-87.5°$ to $-81.25°$ longitude (lon) and $-33°$ to $-26°$ latitude (lat). Data analysts often superimpose a regular lon–lat grid on the Earth's surface and purposively move the spatially irregular data to the nearest center of the lon-lat rectangles. This introduces a location error that we shall adjust for in the geostatistical analysis to follow. The centers, which play the role of intended sites $\{\mathbf{s}_i\}$, are shown as dots in Figure 7(b).

After standard exploratory spatial data analysis, we posed the following geostatistical model:

$$Z(s^x, s^y) = \beta_1 + \beta_2 s^x + \beta_3 s^y + \nu(s^x, s^y) + \varepsilon(s^x, s^y), \tag{30}$$

where s^x denotes degrees longitude and s^y denotes degrees latitude; $\nu(\cdot)$ is the stochastic component of the hidden process, $Y(\cdot) = \mathbf{x}(\cdot)'\boldsymbol{\beta} + \nu(\cdot)$, where $\mathbf{x}(\mathbf{s}) = (1, s^x, s^y)'$; and $\varepsilon(\cdot)$ is the measurement error. We posed the spherical model (12) for $\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \mathrm{cov}(\nu(\mathbf{s}), \nu(\mathbf{s}+\mathbf{h}))$. Because of the curvature of the Earth's surface, great-arc distances were used for computing $\|\mathbf{h}\|$.

To quantify the different sources of variation, we first obtained an estimate of c_{ME}. We made the reasonable assumption that when no location error has been imposed on the data the spatial process is sufficiently smooth to assume that $c_{\mathrm{MS}} = 0$. Therefore, we can use the actual locations to obtain a maximum likelihood estimate of $c_{\mathrm{MS}} + c_{\mathrm{ME}} = c_{\mathrm{ME}}$, and we subsequently assumed this estimate, $\widehat{c}_{\mathrm{ME}} = 2.65$, to be the true value of c_{ME} when performing PALE estimation and KALE prediction. It would obviously be preferable to use an independent estimate of c_{ME} obtained from the instrument manufacturer's quality-assurance experiments; however, such an estimate was not available to us.

Estimates on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ proved to be somewhat sensitive to their starting values in the optimization routine (the Nelder-Mead algorithm using the optim routine within R). This is due to the lack of identifiability between the trend surface $\mathbf{x}(\cdot)'\boldsymbol{\beta}$ and the second-order stationary process $\nu(\cdot)$. The problem is potentially present in all geostatistical analyses, as has been pointed out by Cressie (1993, page 25). We make the recommendation that, when inferences are to be made on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, a range of starting values should be tried for the optimization procedure that maximizes the pseudolikelihood. However, when prediction is the main objective, a careful choice of starting values may

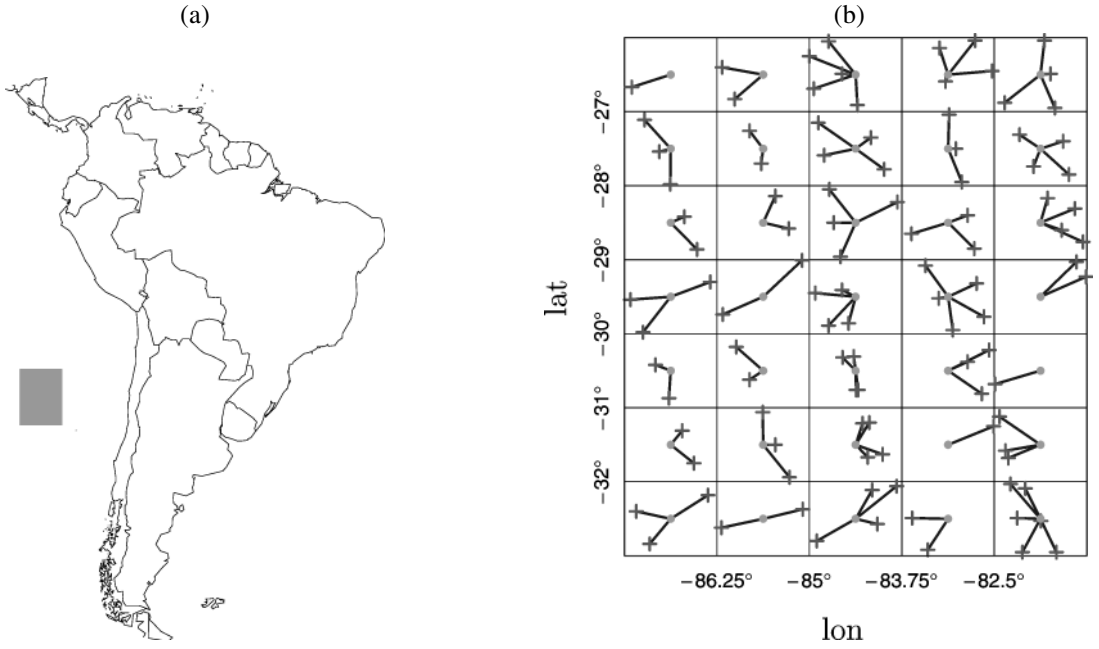

FIG. 7. (a) *Gray rectangle indicates region of interest for studying KALE of total column ozone (TCO)*; (b) *actual and intended sites of TCO data in lon–lat space. Dots correspond to the intended sites (centers of lon–lat rectangles); pluses correspond to the actual sites of TCO data on October 1, 1988; a datum is recorded at a plus but is purposively moved to its nearest dot.*

be less important; see Section 2.3, where it is seen that KALE is not particularly sensitive to the choice of parameters. Here, we used ordinary-least-squares estimation to obtain a starting value for $\boldsymbol{\beta}$. We then obtained the residual mean square and used it as an estimate of the total variance. The starting value for τ^2 was 75% of the total variance, since $\hat{c}_{\mathrm{ME}}$ was found to be 25% of the estimated total variance. The starting value for ϕ was obtained by taking the median of all distances between all pairs of intended sites in the region of interest. Maximization of (16) yielded PALE estimates $\widehat{\boldsymbol{\beta}}_g = (134.96, -0.50, -4.42)'$ and $\widehat{\boldsymbol{\theta}}_g \equiv (\widehat{\tau}_g, \widehat{c}_{\mathrm{MS},g}, \widehat{\phi}_g)' = (10.97, 4.56, 361.55)'$, where the range-parameter estimate $\widehat{\phi}_g$ is measured in kilometers.

The data vector we use for prediction is given by $\mathbf{Z}_g \equiv (Z(\mathbf{r}_1), \ldots, Z(\mathbf{r}_n))'$, but we assume that the data are observed at the centers $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ of the lon–lat rectangles that contain them. Recall that the actual locations of the data are not available to the data analyst; only the assigned grid centers (intended locations) are used.

Notice that the satellite tracks are evident from the pluses in Figure 7(b). Although the process of data assignment is clearly not random, the uncertainty induced by not knowing the actual locations where the data were recorded can be modeled by the CP location-error model (Section 2). Figure 8 displays the purposively imposed location errors, $\mathbf{p}_i \equiv \mathbf{r}_i - \mathbf{s}_i$, $i = 1, \ldots, n$, that were incurred within D. The random scatter of errors in the rectangle and the approximately uniform marginal histograms in the lon and lat directions, imply that the realizations $\{\mathbf{p}_i : i = 1, \ldots, n\}$ can be approximated as coming from the uniform distribution on the $1.25^\circ \times 1^\circ$ rectangle centered at $(0, 0)$.

From (25) and (26), the KALE predictor is given by

$$p_g(\mathbf{Z}_g; \mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \mathbf{c}_g(\mathbf{s})'\Sigma_g^{-1}(\mathbf{Z}_g - X_g\boldsymbol{\beta}), \quad \mathbf{s} \in D, \tag{31}$$

where the PALE estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are substituted

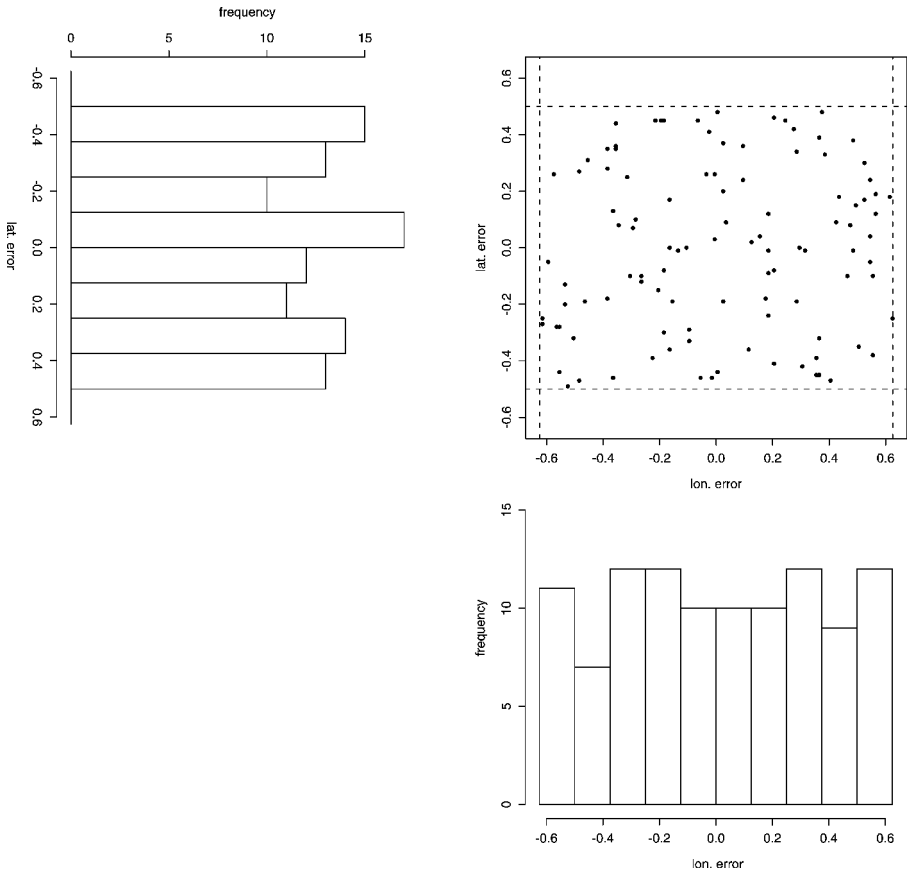

FIG. 8. *Two-dimensional dot plot showing the* 105 *location errors imposed by moving the actual sites* (*from the satellite swaths*) *on October 1, 1988, to the intended sites* (*nearest centers of the* $1.25^\circ \times 1^\circ$ *lon–lat rectangles*). *The density of dots appears uniform*; *this is reinforced by the marginal histograms shown in the figure.*

into the right-hand side of (31). Figure 9 illustrates the results of KALE; to obtain a comparable gray-scale for the maps, we subtracted an overall average of the trend (307.3 Dobson units) from both sides of (31). It is clear from Figure 9 that, while the trend $\mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$ is important, the residual nonlinear features of the map are effectively captured by

$$\widehat{\nu}(\mathbf{s}) \equiv \mathbf{c}_g(\mathbf{s})'\Sigma_g^{-1}(\mathbf{Z}_g - X_g\widehat{\boldsymbol{\beta}}_g).$$

The map of $\{(\mathrm{MSPE}_g(\mathbf{s}))^{1/2} : \mathbf{s} \in D\}$ shown in Figure 9 is an essential part of KALE.

6. DISCUSSION AND CONCLUSIONS

Location error can have a considerable effect on attribute prediction (kriging). Generally speaking, adjusting inferences for location error is more efficient than ignoring it. Examples include purposive relocation of massive, irregularly spaced data $\mathbf{Z}$ to become data $\mathbf{Z}_g$ on regularly spaced grid nodes in remote sensing, and reporting people's locations only at the zip-code level in confidential surveys. We believe that (spatial) inference in the presence of location error is an important problem, although it tends to be ignored by most spatial analysts.

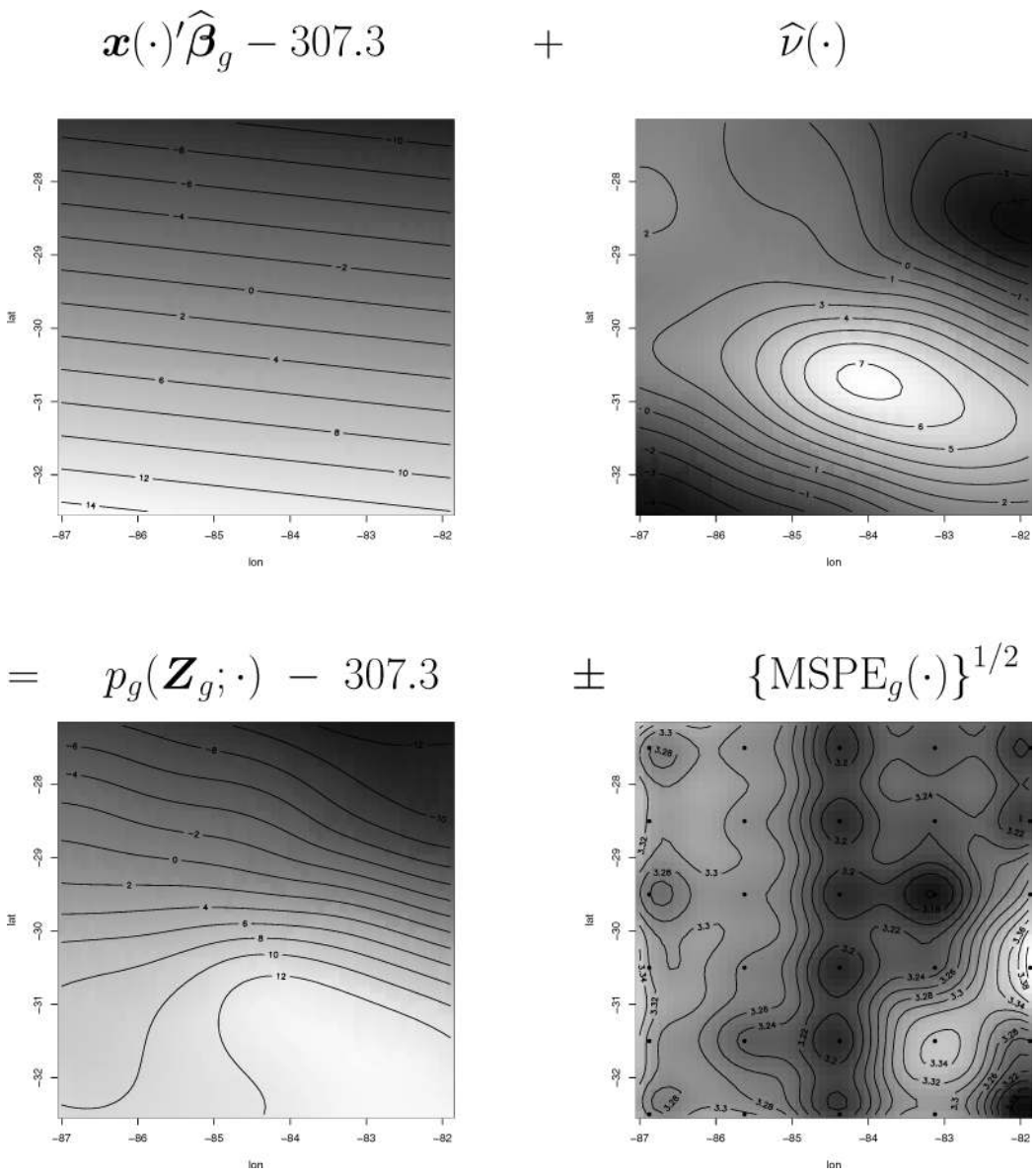

FIG. 9. *The recentered trend map plus the predicted small-scale variation,* $\widehat{\nu}(\cdot) = \mathbf{c}_g(\cdot)'\Sigma_g^{-1}(\mathbf{Z}_g - X_g\widehat{\boldsymbol{\beta}})$, *is equal to the recentered KALE map. Also shown is* $\{\mathrm{MSPE}_g(\cdot)\}^{1/2}$ *given by* (24). *Lighter shading corresponds to larger values.*

In this paper, we have used artificial data, a simulation experiment and an analysis of remotely sensed TCO data to show that adjusting for location error takes a little extra effort, but that it is definitely worth it. Location error plays a different role to measurement error, however some characteristics are similar. In each case we assume that experiments extraneous to the spatial statistical analysis have been conducted that determine, respectively, the measurement-error variance c_{ME} and the location-error density $g(\cdot)$. That is, in our spatial analyses estimation is confined to trend and covariance parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, all pertaining to the process $Y(\cdot)$. It is possible that the disk radius ψ in $g(\cdot)$ could be estimated via maximum pseudolikelihood, but in many cases it would be known in advance. Moreover, in remote-sensing applications, such as given in Section 5, we can sometimes rehabilitate subset of the actual data locations and use the differences between actual and intended locations to examine the form of $g(\cdot)$. We did this for the TCO data in Section 5; see Figure 8, where the realizations from $g(\cdot)$ appear to be uniform on the rectangles of the latitude–longitude grid.

It is clear that the building blocks for location-error adjustment involve integrals with respect to the location-error density $g(\cdot)$. In the more general case of a mixture of discrete and continuous location-error distributions, the integral would be of the Riemann–Stieltjes type. In particular, if $\mathbf{p}(\mathbf{s}) = \mathbf{0}$ with positive probability p_0, then $C_g(\mathbf{h})$ has a jump equal to $p_0 c_{\mathrm{MS}}$ at $\mathbf{h} = \mathbf{0}$, and care must be taken to account for this in the Monte Carlo integration given in Section 2.1.

There are a number of open problems still to be solved. The presence of the trend parameters $\boldsymbol{\beta}$ on the diagonal of the variance–covariance matrix $\mathrm{var}(\mathbf{Z}_g)$ indicates that more efficient estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ may be available through quasilikelihoods or generalized estimating equations (e.g., Diggle, Heagerty, Liang and Zeger, 2002); this is a topic of future research. Alternatively, one could use the ad hoc approach that is often seen in practical geostatistics: based on an ordinary-least-squares estimate of $\boldsymbol{\beta}$ $[= (X_g' X_g)^{-1} X_g' \mathbf{Z}_g]$, detrend the data $\mathbf{Z}_g$, compute an empirical covariance function and fit $\boldsymbol{\theta}$ via, for example, nonlinear weighted least squares [using only nonzero lags, since $\boldsymbol{\beta}$ is not present in $\mathrm{cov}(Z_g(\mathbf{s}), Z_g(\mathbf{s}+\mathbf{h}))$ when $\mathbf{h} \neq \mathbf{0}$]. This latter approach requires assumptions of second-order stationarity of $Y(\cdot)$ and of independent and identically distributed location errors, both of which are made in this paper. However, KALE is quite general in that it simply requires first- and second-order moments associated with the data $\mathbf{Z}_g$ and the hidden process $Y(\cdot)$. Clearly, (9)–(11) and all other components of KALE can be generalized to accommodate a nonstationary $\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{u}))$ and a noninvariant, nonindependent location-error process $\mathbf{p}(\cdot)$. In this case, estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ via PALE would be (pseudo- or quasi-)likelihood based.

Substituting an estimate of unknown parameters into the kriging equations generally has a minimal effect on the predictor, but it can lead to an underestimate of the predictor's mean squared prediction error. Ideally, we should carry out a fully Bayesian analysis of spatial prediction in the presence of location error. This would necessitate, *inter alia*, putting a prior on $g(\cdot)$, or equivalently on its parameters. Inference based on Markov chain Monte Carlo is quite complicated in this case, and will be the subject of a future paper.

In conclusion, we have shown that substantial gains in efficiency can often be made by adjusting kriging in the presence of location error. It can be seen from the kriging equations that the trend parameter $\boldsymbol{\beta}$ plays an important role; we expect the efficiency gains to be even greater for trends more complicated than the simple linear ones considered in this paper.

ACKNOWLEDGMENTS

Research supported by U.S. Environmental Protection Agency Grant R827257-01-0 and Office of Naval Research Grant N00014-02-1-0052. The authors thank John Gabrosek for helpful discussions, Konstantin Krivoruchko for his guidance to early references on location error, Gardar Johannesson for assistance with the TCO data and the referees for helpful comments.

REFERENCES

ARBIA, G., GRIFFITH, D. and HAINING, R. (1998). Error propagation modelling in raster GIS: Overlay operations. *Internat. J. Geographical Information Sci.* **12** 145–167.

ATKINSON, P. M. (1997). Simulating locational error in field-based measurements of reflectance. In *geoENV I—Geostatistics for Environmental Applications* (A. Soares, J. Gómez-Hernández and R. Froidevaux, eds.) 297–338. Kluwer, Dordrecht.

BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45** 164–180.

BOOKSTEIN, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.* **1** 181–242.

BURR, D. (1988). On errors-in-variables in binary regression—Berkson case. *J. Amer. Statist. Assoc.* **83** 739–743.

CARROLL, R. and RUPPERT, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, London.

CHILES, J. P. (1976). How to adapt kriging to non-classical problems: Three case studies. In *Advanced Geostatistics in the Mining Industry* (M. Guarascio, M. David and C. Huijbregts, eds.) 69–89. Reidel, Dordrecht.

CHILES, J. P. and DELFINER, P. (1999). *Geostatistics*: *Modeling Spatial Uncertainty*. Wiley, New York.

COCCHI, D. and TRIVISANO, C. (2002). Ozone. In *Encyclopedia of Environmetrics* **3** 1518–1523. Wiley, New York.

COX, D. R. and ISHAM, V. (1980). *Point Processes*. Chapman and Hall, London.

CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. ed. Wiley, New York.

CRESSIE, N. and ZIMMERMAN, D. L. (1992). On the stability of the geostatistical method. *Math. Geol.* **24** 45–59.

DIGGLE, P. J. (1993). Point process modelling in environmental epidemiology. In *Statistics for the Environment* (V. Barnett and K. F. Turkman, eds.) 89–110. Wiley, New York.

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K. Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford Univ. Press.

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics (with discussion). *Appl. Statist.* **47** 299–350.

DOBRUSHIN, R. L. (1963). On Poisson laws for distributions of particles in space. *Ukrain. Mat. Zh.* **8** 127–134. (In Russian.)

FAZEKAS, I., BARAN, S., KUKUSH, A. and LAURIDSEN, J. (1999). Asymptotic properties in space and time of an estimator in nonlinear functional errors-in-variables models. *Random Oper. Stochastic Equations* **7** 389–412.

FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York.

GABROSEK, J. (1999). The effect of locational uncertainty in geostatistics. Ph.D. dissertation, Dept. Statistics, Iowa State Univ., Ames.

GABROSEK, J. and CRESSIE, N. (2002). The effect on attribute prediction of location uncertainty in spatial data. *Geographical Analysis* **34** 262–285.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721–741.

GOODCHILD, M. (1989). Modelling error in objects and fields. In *The Accuracy of Spatial Databases* (M. Goodchild and S. Gopal, eds.) 107–113. Taylor and Francis, London.

HAINING, R. and ARBIA, G. (1993). Error propogation through map operations. *Technometrics* **35** 293–305.

HARVILLE, D. A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132–138.

KIELLAND, P. and TUBMAN, T. (1994). On estimating map model errors and GPS position errors: Applying more science to the art of navigation. *Internat. Hydrographic Rev.* **71** 47–67.

LONDON, J. (1985). The observed distribution of atmospheric ozone and its variations. In *Ozone in the Free Atmosphere* (R. C. Whitten and S. S. Prasad, eds.) 11–80. Van Nostrand Reinhold, New York.

MATÉRN, B. (1960). *Spatial Variation*. Springer, Berlin.

MATHERON, G. (1963). Principles of geostatistics. *Economic Geology* **58** 1246–1266.

STONE, C. (1968). On a theorem by Dobrushin. *Ann. Math. Statist.* **39** 1391–1401.

VENEZIANO, D. and VAN DYCK, J. (1987). Statistical analysis of earthquake catalogs for seismic hazard. In *Stochastic Approaches in Earthquake Engineering* (Y. K. Lin and R. Minai, eds.) 385–427. Springer, New York.