Estimation of Locally Stationary Spatial Processes with Application to the American Community Survey

Daniel Weinberg* Tucker McElroy† Soumendra Lahiri‡

Abstract

The American Community Survey (ACS) multiyear estimates provide detailed economic and demographic information at a census tract level. The assumption of spatial stationarity for many variables is dubious, which motivates our formulation of local stationarity that can take into account changes in the covariance structure across census tracts. In addition, we adopt a nonparametric modeling approach that remains agnostic about specific distributional features. We present fairly general constructions in both the frequency and spatial domains, deriving an estimator for the local covariance. The properties of the local covariance estimator are explored through simulation. For our application, we utilize our estimator on the ACS data of median household income in the state of Iowa.

Key Words: American Community Survey, local stationarity, random field, stochastic design

Disclaimer

This paper is intended to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

1. Introduction

The American Community Survey (ACS) is a widely-studied survey conducted by the U. S. Census Bureau [2, 19]. Current products from the survey are 5-year Multi-year Estimates (MYEs) for all census block groups in the United States and 1-year MYEs for regions with population 65,000 or higher. However, data users may require additional data, such as MYEs for custom geographies, which amounts to a change-of-support problem. In classical geostatistics, this is commonly done with block kriging [4]. Before kriging can be performed, it is necessary to develop an estimator either for the random field's variogram or for its covariogram. Typically, weak stationarity is assumed. However, human populations are spread heterogenously; we would not expect for the properties of urban and rural regions to be similar in structure, for instance. Therefore, we develop a fresh approach to modeling ACS data that can account for spatial heterogeneity.

The use of heterogenous covariance functions in the spatial setting has been explored extensively in the literature. Sampson and Guttorp [15, 10] propose a spatial deformation approach. Here, the authors use multidimensional scaling (MDS) to transform a nonstationary process into one that is both stationary and

*US Census Bureau, Washington, DC 20233. Email: daniel.weinberg@census.gov.

†US Census Bureau, Washington, DC 20233.

‡North Carolina State University, Raleigh, NC 27695.

isotropic. These ideas are utilized in a Bayesian framework by Schmidt and O'Hagan [16]. Fuentes [8, 9] suggests dividing one's spatial domain into smaller sub-grids, fitting a different stationary process for each, and then writing the overall process as a spatially-weighted sum of these processes. Basis function models have been considered by various authors; decompositions methods include empirical orthogonal functions (EOFs) [17], Fourier series [18], wavelets [12, 7], and wavelet packets [3]. See [14] for a good review on this topic.

We adopt an approach that is nonparametric in spirit, being based upon the fundamental spectral representation of stationary processes, but adapted to account for a form of spatial non-stationarity referred to as *local stationarity*. First, recall that a zero-mean stationary time series $X_t, t \in \mathbb{Z}$, has a spectral representation [1]

$$X_t = \int_{-\pi}^{\pi} \exp(\iota\omega t) A(\omega) d\xi(\omega),$$

where $\xi(\omega)$ is an orthogonal increment process and $A(\omega)$ is a transfer function. Although Priestley [13] developed a local structure by allowing the transfer function to vary in time, no meaningful asymptotic theory is available for this formulation. To correct for this, Dahlhaus [5] considered a triangular array of stochastic processes, $X_{t,T}(t = 1, \ldots, T)$, that possesses a representation of the form

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(\iota\omega t) A_{t,T}(\omega) d\xi(\omega) \tag{1}$$

for which meaningful asymptotics can be derived.

Our work advances a nonparametric locally stationary methodology for spatial data, motivated by (1) and the above applications to the ACS. This paper is organized as follows. In Section 2, we develop the framework for local stationarity. Section 3 sets up a large class of locally stationary processes, both in the frequency and spatial domains. Our nonparametric covariance estimator is defined in Section 4. In Section 5, we explore the properties of the estimator through simulations. Our application to the American Community Survey is found in Section 6. We conclude in Section 7.

2. Framework

In this work, we study estimation of the covariance structure of a locally stationary spatial process on the continuum based on finitely many observations. Specifically, suppose that $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ for $d \geq 1$ is a spatial process observed at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ in a spatial region $R \subset \mathbb{R}^d$. We do not assume that the $Y(\cdot)$-process is stationary. The type of nonstationarity we consider requires the process to have approximate stationarity in its local spatial interactions.

As in the time series case (cf. [6]), the formulation of the local stationarity is done using a triangular array framework where the transition from the local approximate stationarity in the covariance structure is given by a smooth function. However, there are two main differences. First, unlike the time series case where the process is observed on a regular lattice, the spatial process we are interested in is on the continuum and,

therefore, the covariance function is also a function defined on the continuum. Second, the data-locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ need not be on a regular lattice.

The triangular array formulation involves process, region, and data locations depending on the sample size n: $Y(\cdot) = Y_n(\cdot)$, $R = R_n$, and $\mathbf{s}_i = \mathbf{s}_{in}$, $i = 1, \ldots, n$. The asymptotics of our statistical estimators are driven by n. We have the following working assumptions:

1. We shall suppose that the sampling regions $R \equiv R_n$ are given by

   $$R_n = \lambda_n R_0$$

   for some prototype set $R_0 \in (-1, 1)^d$, where R_0 does not depend on n and the scalar $\lambda_n \to \infty$ as $n \to \infty$. Suppose that R_0 an open set.

2. We shall suppose that the data locations are generated by the realizations of a sequence of iid random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$ through the relation:

   $$\mathbf{s}_i \equiv \mathbf{s}_{in} = \lambda_n \mathbf{X}_i, \;\; 1 \leq i \leq n, \; n \geq 1.$$

   This gives a stochastic spatial design.

3. $EY_n(\mathbf{s}) = 0$ for all $\mathbf{s} \in \mathbb{R}^d$. Further, the covariance structure of $Y_n(\cdot)$ is given by

   $$\mathrm{Cov}\Big(Y_n(\mathbf{s}), Y_n(\mathbf{s}+\mathbf{h})\Big) = C(\lambda_n^{-1}\mathbf{s}; \mathbf{h}) + O(\lambda_n^{-1}) \tag{2}$$

   for some function $C : [-1, 1]^d \times \mathbb{R}^d \to \mathbb{R}$, where $C(\mathbf{u}; \mathbf{h})$ is smooth in $\mathbf{u}$ and

   $$|C(\mathbf{u}; \mathbf{h})| \leq g(\mathbf{u})(1 + \|\mathbf{h}\|)^{-[d+b]} \tag{3}$$

   for some $g : [-1, 1]^d \to (0, \infty)$ and $b \in (0, \infty)$.

The first assumption describes R_n as an expanding template, a standard device in asymptotic statistical theory [5]. The second assumption similarly expands the data locations. The last assumption is adapted from the time series case, c.f. Eq. (73) in [6]. Note that the first argument of C corresponds to the rescaled location of $Y_n(\mathbf{s})$ in R_0, and the second argument gives the lag gap between $Y_n(\mathbf{s})$ and $Y_n(\mathbf{s}+\mathbf{h})$. The bound on C is needed for later asymptotic results, and precludes some forms of long memory, i.e., a slowly decaying autocorrelation function.

3. A class of locally stationary processes

Given our framework's three assumptions, we provide a class of stochastic processes that satisfy these criteria. Let $L^{\mathbf{x}}(\cdot)$ denote the *lag-operator* defined by

$$L^{\mathbf{x}}\Big(Z(\mathbf{s})\Big) = Z(\mathbf{s} - \mathbf{x}), \; \mathbf{x}, \mathbf{s} \in \mathbb{R}^d.$$

Note that for an integrable function $\phi : \mathbb{R}^d \to \mathbb{R}$ and a second order stationary process (SOS) $Z(\mathbf{s})$, we can define the operator $\Phi(L) \equiv \int \phi(\mathbf{x}) L^{\mathbf{x}} dx$ through the identity:

$$\int Z(\mathbf{s}-\mathbf{x})\phi(\mathbf{x})d\mathbf{x} = \int L^{\mathbf{x}}(Z(\mathbf{s}))\phi(\mathbf{x})d\mathbf{x} \equiv \Phi(L)\Big(Z(\mathbf{s})\Big).$$

It is easy to see that the operator $\Phi(L)$ applied to a SOS process $Z(\cdot)$ produces a SOS process. The construction of a locally stationary process can be effected through a time varying version of $\Phi(L)$ where the "lag-weight" or the "kernel" function $\phi(\cdot)$ depends locally on the spatial index. Specifically, let

$$\Phi_{\mathbf{t}}(L) = \int \phi_{\mathbf{t}}(\mathbf{x}) L^{\mathbf{x}} d\mathbf{x}$$

and consider the location-varying lagged process $Y(\mathbf{t}) = \Phi_{\mathbf{t}}(L)Z(\mathbf{t})$.

3.1 A Frequency Domain Formulation

Here we shall consider functions $\phi_{\mathbf{t}}(\cdot)$ such that $\Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega}) \equiv \int \phi_{\mathbf{t}}(\mathbf{x}) \exp(-\iota \boldsymbol{\omega}'\mathbf{x}) d\mathbf{x} \in L^2(\mathbb{R}^d)$ and $\log \Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega})$ is well defined. Here and in the following, we set $\iota = \sqrt{-1}$. Note that $\log \Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega})$ is a complex valued function of $\boldsymbol{\omega}$. Write

$$\log \Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega}) = g_{\mathbf{t}}(\boldsymbol{\omega}) + \iota a_{\mathbf{t}}(\boldsymbol{\omega}),\ \boldsymbol{\omega} \in \mathbb{R}^d, \tag{4}$$

with $g_{\mathbf{t}}(\boldsymbol{\omega})$ being an even function of $\boldsymbol{\omega}$ and $a_{\mathbf{t}}(\boldsymbol{\omega})$ being an odd function of $\boldsymbol{\omega}$. Note that $g_{\mathbf{t}}(\boldsymbol{\omega}) = 2^{-1} \log |\Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega})|^2$. Since $\Phi^{\diamond}(\boldsymbol{\omega}) \in L^2(\mathbb{R}^d)$, its Inverse Fourier Transform (IFT) is well defined and it follows from the above that

$$\begin{aligned}
\phi_{\mathbf{t}}(\mathbf{x}) &= \frac{1}{(2\pi)^d} \int \Phi^{\diamond}(\boldsymbol{\omega}) \exp(\iota \boldsymbol{\omega}'\mathbf{x}) d\boldsymbol{\omega} \\
&= \frac{1}{(2\pi)^d} \int \exp\Big(g_{\mathbf{t}}(\boldsymbol{\omega}) + \iota a_{\mathbf{t}}(\boldsymbol{\omega}) + \iota \boldsymbol{\omega}'\mathbf{x}\Big) d\boldsymbol{\omega}.
\end{aligned}$$

Next, consider the location-varying lagged process

$$\begin{aligned}
Y(\mathbf{t}) &= \Phi_{\mathbf{t}}(L)Z(\mathbf{t}) \\
&= \int Z(\mathbf{t}-\mathbf{x})\phi_{\mathbf{t}}(\mathbf{x})d\mathbf{x} \\
&= \int\int \exp(\iota(\mathbf{t}-\mathbf{x})\boldsymbol{\omega}) dZ^{\diamond}(\boldsymbol{\omega})\phi_{\mathbf{t}}(\mathbf{x})d\mathbf{x} \\
&= \int\int \exp(-\iota\boldsymbol{\omega}'\mathbf{x})\phi_{\mathbf{t}}(\mathbf{x})d\mathbf{x} \exp(\iota\boldsymbol{\omega}'\mathbf{t}) dZ^{\diamond}(\boldsymbol{\omega}) \\
&= \int \Phi^{\diamond}_{\mathbf{t}}(\boldsymbol{\omega}) \exp(\iota\boldsymbol{\omega}'\mathbf{t}) dZ^{\diamond}(\boldsymbol{\omega}) \\
&= \int \exp\Big(g_{\mathbf{t}}(\boldsymbol{\omega}) + \iota a_{\mathbf{t}}(\boldsymbol{\omega}) + \iota\boldsymbol{\omega}'\mathbf{x}\Big) dZ^{\diamond}(\boldsymbol{\omega}),
\end{aligned}$$

where $\{Z^{\diamond}(\boldsymbol{\omega}) : \boldsymbol{\omega} \in \mathbb{R}^d\}$ is the orthogonal increment process associated with the spectral representation of the SOS $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$. Write $F(\cdot)$ for the spectral distribution function of $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$. Then, it

follows (using results in [1])

$$\begin{aligned} & \mathrm{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) \qquad (5) \\ = & \int \exp(\iota \mathbf{h}'\boldsymbol{\omega}) \cdot \overline{\Phi_{\mathbf{t}+\mathbf{h}}(\exp(\iota\boldsymbol{\omega}))}\Phi_{\mathbf{t}}(\exp(\iota\boldsymbol{\omega}))dF(\boldsymbol{\omega}) \\ = & \int \exp(\iota \mathbf{h}'\boldsymbol{\omega}) \cdot \exp\Big(g_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) + \iota a_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) + g_{\mathbf{t}}(\boldsymbol{\omega}) - \iota a_{\mathbf{t}}(\boldsymbol{\omega})\Big)dF(\boldsymbol{\omega}) \\ = & \int \exp(\iota \mathbf{h}'\boldsymbol{\omega}) \cdot \exp([g_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) + g_{\mathbf{t}}(\boldsymbol{\omega})]) \cdot \exp(\iota[a_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) - a_{\mathbf{t}}(\boldsymbol{\omega})])dF(\boldsymbol{\omega}) \\ = & \int \exp([g_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) + g_{\mathbf{t}}(\boldsymbol{\omega})]) \cdot \cos(\mathbf{h}'\boldsymbol{\omega} + a_{\mathbf{t}+\mathbf{h}}(\boldsymbol{\omega}) - a_{\mathbf{t}}(\boldsymbol{\omega}))dF(\boldsymbol{\omega}). \qquad (6) \end{aligned}$$

Now, we posit some structural conditions on the functions $a_{\mathbf{t}}(\cdot)$ and $g_{\mathbf{t}}(\cdot)$ to specify the local stationarity component of the filtered process $Y(\mathbf{t}) \equiv Y_{\mathbf{t}}^{(\lambda)}$, where recall that $\lambda = \lambda_n$ determines the size of the sampling region R_n.

(C.1): *There exist functions $a_0(\cdot;\cdot)$, $a_1(\cdot;\cdot)$ and $g_1(\cdot;\cdot)$ such that*

$$\begin{aligned} a_{\mathbf{t}}(\boldsymbol{\omega}) \equiv a_{\mathbf{t}}^{(\lambda)}(\boldsymbol{\omega}) &= \lambda a_0(\lambda^{-1}\mathbf{t};\boldsymbol{\omega}) + a_1(\lambda^{-1}\mathbf{t};\boldsymbol{\omega}) + O(\lambda^{-1}) \qquad (7) \\ g_{\mathbf{t}}(\boldsymbol{\omega}) \equiv g_{\mathbf{t}}^{(\lambda)}(\boldsymbol{\omega}) &= g_1(\lambda^{-1}\mathbf{t};\boldsymbol{\omega}) + O(\lambda^{-1}) \qquad (8) \end{aligned}$$

where $a_0(\mathbf{u};\boldsymbol{\omega})$ is differentiable w.r.t. $\mathbf{u}$ with gradient $\nabla_{\mathbf{u}} a_0(\mathbf{u};\boldsymbol{\omega})$,

$$|a_1(\mathbf{u}_1;\boldsymbol{\omega}) - a_1(\mathbf{u}_2;\boldsymbol{\omega})| \leq \|\mathbf{u}_1 - \mathbf{u}_2\| A(\boldsymbol{\omega}) \qquad (9)$$

for some integrable functions $A(\cdot)$ and $g_1(\mathbf{u};\boldsymbol{\omega})$ is continuous in $\mathbf{u}$ uniformly over $\boldsymbol{\omega} \in \mathbb{R}^d$. Further, the order symbols are uniform in $\mathbf{t}$ and $\boldsymbol{\omega}$.

For $||\mathbf{h}||$ small, $g_{\mathbf{t}+\mathbf{h}} \approx g_{\mathbf{t}}$ and $a_{\mathbf{t}+\mathbf{h}} \approx a_{\mathbf{t}} + \lambda^{-1}\mathbf{h}'\nabla_{\mathbf{u}} a_{\mathbf{t}}$ so that

$$\mathrm{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) \approx \int \exp(\iota \mathbf{h}'\boldsymbol{\omega}) \cdot \exp\left(2g_{\mathbf{t}}(\boldsymbol{\omega}) + \iota\lambda^{-1}\mathbf{h}'\nabla_{\mathbf{u}} a_{\mathbf{t}}\right) dF(\boldsymbol{\omega})$$

Under (C.1), and assuming λ is large enough that only the higher order terms contribute, it follows that

$$\begin{aligned} & \mathrm{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) \\ \approx & \int \exp\Big(2g_1(\mathbf{u};\boldsymbol{\omega}) + \iota[\mathbf{h}'\nabla_{\mathbf{u}} a_0(\mathbf{u};\boldsymbol{\omega}) + \mathbf{h}'\boldsymbol{\omega}]\Big)dF(\boldsymbol{\omega}) \\ \equiv & C(\mathbf{u};\mathbf{h}). \end{aligned}$$

Assuming that $g_1(\mathbf{u};\boldsymbol{\omega})$ is an even function and $\nabla_u a_0(\mathbf{u};\boldsymbol{\omega})$ is an odd function of $\boldsymbol{\omega}$, for each $\mathbf{u}$, one gets

$$C(\mathbf{u};\mathbf{h}) = \int \exp\Big(2g_1(\mathbf{u};\boldsymbol{\omega})\Big) \cos\Big(\mathbf{h}'\nabla_{\mathbf{u}} a_0(\mathbf{u};\boldsymbol{\omega}) + \mathbf{h}'\boldsymbol{\omega}\Big)dF(\boldsymbol{\omega}). \qquad (10)$$

There are conditions under which this $C(\mathbf{u};\mathbf{h})$ satisfies the third assumption of our framework.

3.2 A Spatial Domain Formulation

Recall that the locally stationary process $Y(\mathbf{t})$ is generated from a SOS process $\{Z(\mathbf{t})\}$ by applying the location-varying lag-operator $\Phi_{\mathbf{t}}(L) = \int \phi_{\mathbf{t}}(\mathbf{x}) L^{\mathbf{x}} d\mathbf{x}$, leading to

$$\begin{aligned} Y(\mathbf{t}) &= \Phi_{\mathbf{t}}(L) Z(\mathbf{t}) \equiv \int \phi_{\mathbf{t}}(\mathbf{x}) Z(\mathbf{t}-\mathbf{x}) d\mathbf{x} \\ &= \int \phi_{\mathbf{t}}(\mathbf{t}-\mathbf{x}) Z(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Note that

$$\begin{aligned} \mathrm{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) &= E\left[\left(\int \phi_{\mathbf{t}+\mathbf{h}}(\mathbf{t}+\mathbf{h}-\mathbf{x}) Z(\mathbf{x}) d\mathbf{x}\right)\left(\int \phi_{\mathbf{t}}(\mathbf{t}-\mathbf{x}) Z(\mathbf{x}) d\mathbf{x}\right)\right] \\ &= \int\int \phi_{\mathbf{t}+\mathbf{h}}(\mathbf{t}+\mathbf{h}-\mathbf{x}) \phi_{\mathbf{t}}(\mathbf{t}-\mathbf{y}) E[Z(\mathbf{x})Z(\mathbf{y})] d\mathbf{x} d\mathbf{y} \\ &= \int\int \phi_{\mathbf{t}+\mathbf{h}}(\mathbf{t}+\mathbf{h}-\mathbf{x}) \phi_{\mathbf{t}}(\mathbf{t}-\mathbf{y}) C_0(\mathbf{x}-\mathbf{y}) d\mathbf{x} d\mathbf{y}, \end{aligned}$$

where $C_0(\cdot)$ is the autocovariance function of the stationary process $Z(\cdot)$. We shall use the following conditions to describe the local stationary structure of the filtered process $Y(\mathbf{t}) \equiv Y_{\mathbf{t}}^{(\lambda)}$:

(C.2): *There exists a function* $\phi_0(\cdot;\cdot) : (-1,1)^d \times \mathbb{R}^d \to \mathbb{R}$ *such that*

$$\phi_{\mathbf{t}}(\mathbf{x}) \equiv \phi_{\mathbf{t}}^{(\lambda)}(\mathbf{x}) = \phi_0(\lambda^{-1}\mathbf{t}; \mathbf{x})$$

and $\phi_0(\mathbf{u};\mathbf{x})$ *is continuous in* $\mathbf{u}$ *for each* $\mathbf{x}$ *and* $|\phi_0(\mathbf{u};\mathbf{x})| \le A(\mathbf{x})$ *for some integrable function* $A(\cdot)$.

Then, it is easy to check that for $\|\mathbf{h}\| = o(\lambda)$,

$$\begin{aligned} \mathrm{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) &= \int\int \phi_0(\lambda^{-1}[\mathbf{t}+\mathbf{h}]; \mathbf{t}+\mathbf{h}-\mathbf{x}) \phi_0(\lambda^{-1}\mathbf{t}; \mathbf{t}-\mathbf{y}) C_0(\mathbf{x}-\mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &= \int\int \phi_0(\lambda^{-1}[\mathbf{t}+\mathbf{h}]; \mathbf{x}) \phi_0(\lambda^{-1}\mathbf{t}; \mathbf{y}) C_0(\mathbf{h}-[\mathbf{x}-\mathbf{y}]) d\mathbf{x} d\mathbf{y} \\ &\approx \int\int \phi_0(\mathbf{u};\mathbf{x}) \phi_0(\mathbf{u};\mathbf{y}) C_0(\mathbf{h}-[\mathbf{x}-\mathbf{y}]) d\mathbf{x} d\mathbf{y} \equiv C(\mathbf{u};\mathbf{h}). \end{aligned}$$

Hence, the covariance $C(\mathbf{u};\mathbf{h})$ clearly satisfies the condition (3) with suitable conditions on C_0, and the third working assumption of our framework can be thereby validated.

4. Estimation

Our goal here is to estimate the function $C(\mathbf{u};\mathbf{h})$ for $\mathbf{u} \in R_0$. In the spatial domain, one possible class of estimators is:

$$\widehat{C}_n(\mathbf{u},\mathbf{h}) = \frac{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u}) H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i)) Y(\mathbf{s}_i) Y(\mathbf{s}_j)}{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u}) H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))}, \tag{11}$$

where $K_b(\cdot) = b^{-d}K(\cdot/b)$ and $H_c(\cdot) = c^{-d}H(\cdot/c)$ for some nonnegative kernel functions (pdfs) on $\mathbb{R}^d$ vanishing outside a compact set, and where $b \equiv b_n \in (0,\infty)$, $c \equiv c_n \in (0,\infty)$ are bandwidth parameters satisfying

$$(b+c) + [\lambda(b+c)]^{-1} = o(1) \quad \text{as} \quad n \to \infty \tag{12}$$

and

$$nb^d \to \infty, \frac{nc^d}{\lambda^d} \to \infty \quad \text{as} \quad n \to \infty. \tag{13}$$

The form of (11) is similar to the kernel estimator in [11]. The main difference is that ours uses two kernel functions to obtain the local covariance at position $\mathbf{u}$ and lag $\mathbf{h}$. This ensures that we strongly weight pairs of points $(\mathbf{s}_i, \mathbf{s}_j)$ such that $\lambda^{-1}\mathbf{s}_i$ is close to $\mathbf{u}$ and $\mathbf{h}$ is close to $\mathbf{s}_j - \mathbf{s}_i$. See also Eq. (9) in [6].

Let $E_{\cdot|\mathcal{X}}$ denotes the conditional expectation given $\mathcal{X} = \sigma\langle \mathbf{X}_1, \mathbf{X}_2, \ldots \rangle$. Note that by the assumption 2,

$$\begin{aligned}
E_{\cdot|\mathcal{X}}\widehat{C}_n(\mathbf{u},\mathbf{h}) &= \frac{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))E_{\cdot|\mathcal{X}}[Y(\mathbf{s}_i)Y(\mathbf{s}_j)]}{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))} \\
&\approx \frac{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))C(\lambda^{-1}\mathbf{s}_i; \mathbf{s}_j - \mathbf{s}_i)}{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))} \\
&\approx \frac{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))C(\mathbf{u};\mathbf{h})}{\sum_{\mathbf{s}_i,\mathbf{s}_j} K_b(\lambda^{-1}\mathbf{s}_i - \mathbf{u})H_c(\mathbf{h} - (\mathbf{s}_j - \mathbf{s}_i))} \\
&= C(\mathbf{u};\mathbf{h}),
\end{aligned}$$

which indicates that the estimator is asymptotically unbiased.

5. Simulations

Here, we will use a direct approach where we start with the functions $g_{\mathbf{t}}(\cdot)$ and $a_{\mathbf{t}}(\cdot)$ satisfying the conditions of Section 3.1, and apply it to a Gaussian WN process $Z(\cdot)$. For instance, let $g_{\mathbf{t}}(\boldsymbol{\omega}) = \sum_{j=0}^{N} g_{j,\mathbf{t}}\|\boldsymbol{\omega}\|^j$ with $g_{N,\mathbf{t}} < 0$, and choose $a_{\mathbf{t}}(\boldsymbol{\omega})$ to be an odd-polynomial function in $\boldsymbol{\omega}$ with coefficients depending on $\mathbf{t}$. We consider the case where $g = g_1 = \frac{1}{2}||\boldsymbol{\omega}||^2$ and $a_0(\mathbf{u};\boldsymbol{\omega}) = \frac{1}{2}\mathbf{u}'B(\boldsymbol{\omega})\mathbf{u}, a_1 \equiv 0$, where $B(\boldsymbol{\omega}) = diag(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$. Note that $\nabla_{\mathbf{u}} a = B(\boldsymbol{\omega})\mathbf{u}$. Using (6), we can compute exactly

$$\begin{aligned}
&\text{Cov}(Y(\mathbf{t}+\mathbf{h}), Y(\mathbf{t})) \\
&= \int \exp(-||\boldsymbol{\omega}||^2)\exp\left\{\iota\boldsymbol{\omega}'\left[\mathbf{h} + \frac{\mathbf{h}*\mathbf{t}}{\lambda} + \frac{\mathbf{h}*\mathbf{h}}{2\lambda}\right]\right\} d\boldsymbol{\omega} \qquad (14) \\
&= \pi\exp\left(-\frac{1}{4}\left|\left|\mathbf{h} + \mathbf{h}*\mathbf{u} + \frac{\mathbf{h}*\mathbf{h}}{2\lambda}\right|\right|^2\right)
\end{aligned}$$

where $*$ is the Hadamard product.

We can also compute with (10) that

$$\begin{aligned} & C(\mathbf{u};\mathbf{h}) \\ = & \int \exp(-||\boldsymbol{\omega}||^2)\cos\left\{\mathbf{h}'\left[B(\boldsymbol{\omega})\mathbf{u}+\boldsymbol{\omega}\right]\right\}d\boldsymbol{\omega} \\ = & \int \exp(-||\boldsymbol{\omega}||^2)\exp\left\{\iota\boldsymbol{\omega}'\left[\mathbf{h}+\mathbf{h}*\mathbf{u}\right]\right\}d\boldsymbol{\omega} \\ = & \pi\exp\left(-\frac{1}{4}\left|\left|\mathbf{h}+\mathbf{h}*\mathbf{u}\right|\right|^2\right). \end{aligned}$$

To ensure a resonable point density for estimation purposes, we set $n = 1000$ and $\lambda = n^{1/3} = 10$. Let $R_0 = (-1,1)^2$ and X_i be i.i.d. uniform on R_n. For simplicity, let $K(\mathbf{x}) = H(\mathbf{x}) = (2/\pi)(1 - ||\mathbf{x}||^2)\mathbb{1}_{||\mathbf{x}||<1}$. To choose the bandwidths, we fix $\mathbf{u}$ and $\mathbf{h}$ and vary the parameters, empirically choosing the ones that give the lowest MSE over 5000 simulations. In this way, we determine that $b = c = 0.4$ are effective values.

We perform simulations to obtain bias and variance properties of the estimator. One realization of our random field is shown in Figure 1. We fix the value of $\mathbf{u}$, and compute the approximate bias and variance over 1000 simulations, as $\mathbf{h}$ varies. The results for $\mathbf{u} = (0,0)$ and $\mathbf{u} = (0,0.5)$ are found in Tables 1 and 2, respectively.

6. Application to American Community Survey (ACS) Data

We apply our methods to the 5-year estimates for median household income in Iowa from 2011-2015 at the level of census tract. Our main goal here is to model; interpolation (kriging) will be considered in future work. We begin by assigning each areal measurement to the centroid of the corresponding region. After plotting the log median incomes, we observe that there are many outliers, due to the presence of small tracts, such as colleges, whose characteristics differ significantly from the neighboring values. As a result, we smooth the data by replacing each value with the median of the eight nearest neighbors; see Figure 2. We obtain the residuals $e(\cdot)$ by subtracting the smoothed values from the original values.

To illustrate the necessity of the locally stationary approach, we consider the isotropic empirical semivariogram (c.f. (3.7) in [20]) defined by

$$\hat{\gamma}(||\mathbf{h}||) = \frac{1}{2\cdot\#(N(||\mathbf{h}||))}\sum_{N(||\mathbf{h}||)}(e(\mathbf{s}_i)-e(\mathbf{s}_j))^2$$

where

$$N(||\mathbf{h}||) = \left\{(\mathbf{s}_i,\mathbf{s}_j) : \left|\,||\mathbf{s}_i-\mathbf{s}_j|| - ||\mathbf{h}||\,\right| < c\right\}.$$

We divide Iowa into four quadrants, computing the semivariogram utilizing only those values within each quadrant. The results for $\mathbf{u} = (0,0)$ are found in Figure 3. If the field is stationary, we would expect the semivariogram to be similar for all quadrants, as it is an estimator for

$$\frac{1}{2}\text{Var}[e(\mathbf{s})-e(\mathbf{t})] = \gamma(||\mathbf{s}-\mathbf{t}||).$$

The semivariograms computed in quadrants I and III are quite different from the ones computed in quadrants II and IV, demonstrating the need for developing a model in which the covariance can vary by position.

We next utilize our nonparametric estimator on the ACS data. Here, $n = 825$. We fix $\mathbf{u}$ and vary $\mathbf{h}$ to produce a surface plot of the local covariance at the origin. Results for $\mathbf{u} = (0, 0)$ and $\mathbf{u} = (0, 1)$ are found in Figure 4. These surfaces are similar in shape, which is consistent with the locally stationary framework. We note the anisotropy in both surface plots, with the covariance falling more rapidly in the direction of the $\mathbf{h}_1$-axis. Since $\mathbf{h}_1$ is a rescaled and shifted longitude value, while $\mathbf{h}_2$ corresponds to a latitude, this implies a relatively high north-south correlation. In addition, the estimation becomes unstable for large $||\mathbf{h}||$. Furthermore, the covariances values in Fig. 4a are slightly higher than in Fig. 4b, about 0.4% larger at their respective maxima.

7. Conclusions and Future Directions

We defined a large class of locally stationary processes. We have given a frequency domain formulation of these processes. We derived a nonparametric estimator for the local covariance, investigating its bias and variance properties for a particular choice of covariance function. We next demonstrated that a position-varying covariance structure is needed in the case of median income data from the ACS. The estimated local covariance is anisotropic, falling more slowly in the north-south direction than in the east-west direction.

Our future tasks include deriving asymptotic properties of the bias and variance of the local covariance estimator, determining optimal values of bandwidth parameters through cross-validation, and developing methodology for performing predictions on custom geographies. In principle the basic kriging formulas, which are based on minimal MSE prediction among the class of linear predictors, can be applied with our estimated covariances inserted, but the invertibility of the covariance matrix estimate needs to be established. A final consideration is the extension of our framework to areal data.

References

[1] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Science & Business Media, 2013.

[2] U.S. Census Bureau. Design and methodology, American Community Survey. Technical Report 73, 2006.

[3] A. Cardinali and G. P. Nason. Locally stationary wavelet packet processes: Basis selection and model fitting. *Journal of Time Series Analysis*, 38(2):151–174, 2017.

[4] N. Cressie. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992.

[5] R. Dahlhaus. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens conference on applied probability and time series analysis*, volume 2, pages 145–159, 1996.

[6] R. Dahlhaus. Locally stationary processes. *Handbook of statistics*, 30:351–412, 2012.

[7] I. A. Eckley, G. P. Nason, and R. L. Treloar. Locally stationary wavelet fields with application to the modelling and analysis of image texture. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(4):595–616, 2010.

[8] M. Fuentes. A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12(5):469–483, 2001.

[9] M. Fuentes. Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1):197–210, 2002.

[10] P. Guttorp and P. D. Sampson. 20 methods for estimating heterogeneous spatial covariance functions with environmental applications. *Handbook of statistics*, 12:661–689, 1994.

[11] P. Hall and P. Patil. Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, 99(3):399–424, 1994.

[12] D. Nychka, C. Wikle, and J. A. Royle. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331, 2002.

[13] M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237, 1965.

[14] P. D. Sampson. Constructions for nonstationary spatial processes. *Handbook of spatial statistics*, pages 119–130, 2010.

[15] P. D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

[16] A. M. Schmidt and A. O'Hagan. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):743–758, 2003.

[17] T. M. Smith, R. W. Reynolds, R. E. Livezey, and D. C. Stokes. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *Journal of Climate*, 9(6):1403–1420, 1996.

[18] J. Stephenson, C. Holmes, K. Gallagher, and A. Pintore. A statistical technique for modelling non-stationary spatial processes. *Geostatistics Banff 2004*, 14:125, 2008.

[19] N. K. Torrieri. America is changing, and so is the census: The American Community Survey. *The American Statistician*, 61(1):16–21, 2007.

[20] D. L. Zimmerman and M. Stein. Classical geostatistical methods. *Handbook of spatial statistics*, pages 29–44, 2010.

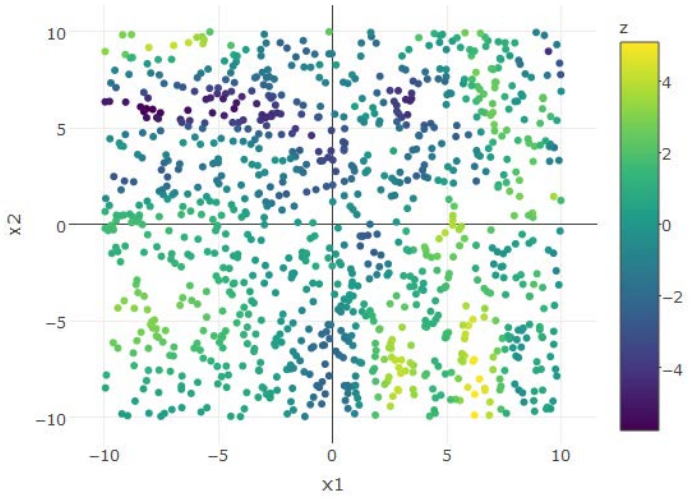


Figure 1: A realization of the random field with covariance structure given by (14).

Table 1: Biases (Variances) for $\mathbf{u} = (0, 0)$.

h	$h_2 = 0$	0.5	1	1.5	2
$h_1 = 0$	$-0.39(0.55)$	$-0.33(0.53)$	$-0.15(0.48)$	$0.08(0.42)$	$0.34(0.37)$
0.5	$-0.34(0.52)$	$-0.30(0.50)$	$-0.16(0.45)$	$0.07(0.40)$	$0.31(0.35)$
1	$-0.19(0.46)$	$-0.19(0.44)$	$-0.10(0.40)$	$0.10(0.35)$	$0.28(0.32)$
1.5	$0.00(0.39)$	$0.00(0.37)$	$0.06(0.34)$	$0.16(0.31)$	$0.26(0.29)$
2	$0.25(0.34)$	$0.24(0.33)$	$0.23(0.30)$	$0.25(0.28)$	$0.26(0.27)$

Table 2: Biases (Variances) for $\mathbf{u} = (0, 0.5)$.

h	$h_2 = 0$	0.5	1	1.5	2
$h_1 = 0$	$-0.43(0.49)$	$-0.19(0.46)$	$0.30(0.39)$	$0.58(0.33)$	$0.60(0.29)$
0.5	$-0.38(0.46)$	$-0.16(0.43)$	$0.26(0.37)$	$0.55(0.32)$	$0.56(0.29)$
1	$-0.23(0.39)$	$-0.09(0.37)$	$0.22(0.32)$	$0.47(0.29)$	$0.48(0.28)$
1.5	$-0.03(0.32)$	$0.06(0.31)$	$0.27(0.28)$	$0.41(0.26)$	$0.38(0.26)$
2	$0.24(0.28)$	$0.27(0.26)$	$0.34(0.25)$	$0.36(0.24)$	$0.30(0.24)$

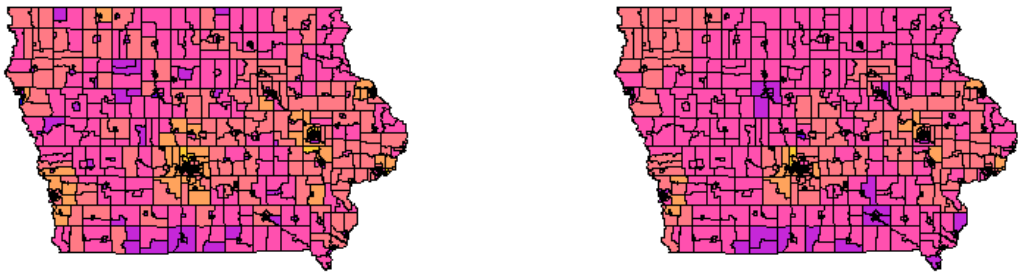


Figure 2: Log median income by census tract for raw data (left) and smoothed data (right).

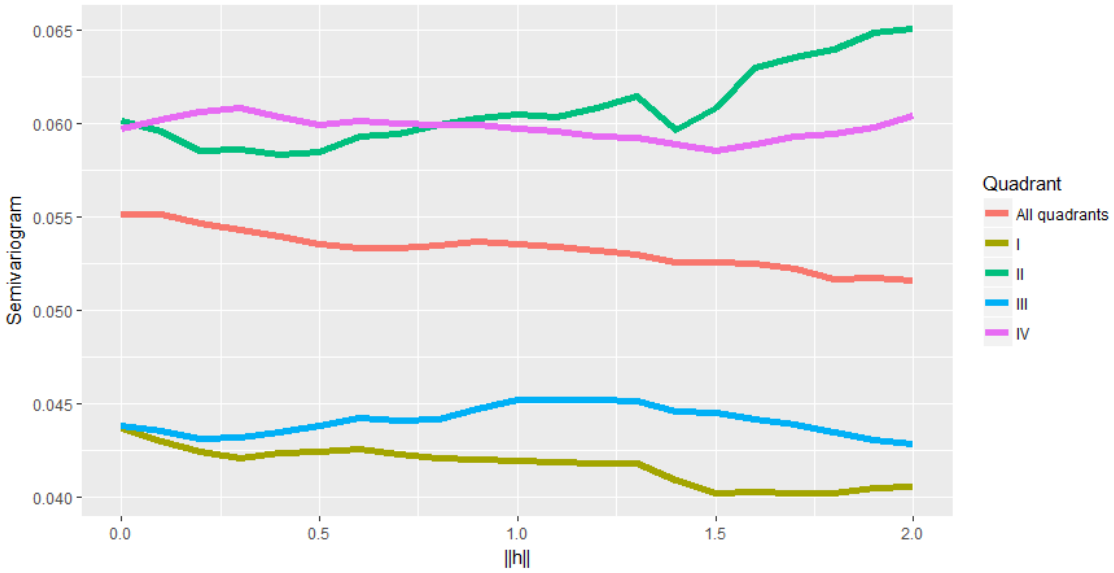
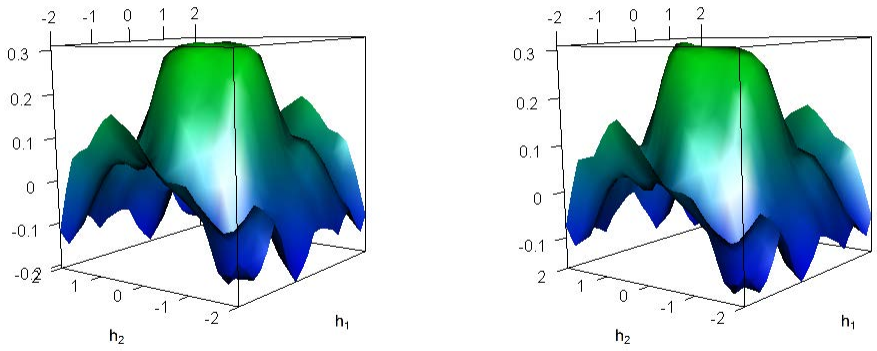


Figure 3: Empirical semivariograms calculated using values within a particular quadrant of the state of Iowa.



(a) $\mathbf{u} = (0, 0)$ (b) $\mathbf{u} = (0, 1)$

Figure 4: Estimated covariance $(\times 10^4)$ for $\mathbf{u} = (0, 0)$ and $\mathbf{u} = (0, 1)$ as $\mathbf{h}$ varies.