



Spatial long memory

Peter M. Robinson¹

Received: 4 July 2019 / Accepted: 14 October 2019
© The Author(s) 2019

Abstract

We discuss developments and future prospects for statistical modeling and inference for spatial data that have long memory. While a number of contributions have been made, the literature is relatively small and scattered, compared to the literatures on long memory time series on the one hand, and spatial data with short memory on the other. Thus, over several topics, our discussions frequently begin by surveying relevant work in these areas that might be extended in a long memory spatial setting.

Keywords Spatial data · Long memory

1 Introduction

“Long memory”, or ‘long range dependence’, or ‘strong dependence’, in time series is a topic that has been quite extensively studied in recent years. After a number of probabilistic contributions and empirical studies, serious treatment of issues of statistical inference could be said to have begun in the mid-1980s, with activity then increasing through the 1990s and this century. Predominately, this literature has focussed on observations that are regularly spaced over time, and the bulk of the theoretical development has been in terms of asymptotic statistical theory, with the number of observations, n , regarded as diverging, finite-sample theory proving mathematically intractable, even under the precise distributional assumptions that are typically not required in a large-sample treatment. Parametric, semiparametric and nonparametric models have all featured. The major characteristic feature of a long memory covariance stationary time series process is that its autocovariance function decays so slowly with increase in lag length as not to be summable, or, nearly equivalently, its spectral density diverges, typically at zero frequency; while some nonstationary processes, such as ones with a unit root can, a fortiori, be regarded as having even longer memory. By contrast, ‘short memory’ time series typically have autocovariances that are summable and

✉ Peter M. Robinson
p.m.robinson@lse.ac.uk

¹ London School of Economics, London, UK

spectral density that is more or less smooth (for example a stationary autoregressive moving average (ARMA) has exponentially decaying autocovariance and analytic spectral density); though for some relevant purposes, a short memory process is sometimes defined as merely having spectral density that is positive and finite at zero frequency.

Spatial data have long attracted the attention of statisticians, and the configurations of some such data, especially some arising in such fields as meteorology, cosmology and agriculture, can be viewed as generalizations of the typical regularly spaced time series one mentioned above. In particular, they constitute observations in $M \geq 2$ dimensions, (where $M = 1$ in the time series case), such that there are n_m observations across dimension m , for $m = 1, \dots, M$, so $n = \prod_{m=1}^M n_m$. We call this a '(rectangular) lattice', and data recorded on it belong to the class of 'random fields'. There is equal spacing across each dimension, but the spacing can vary across dimensions. In Fig. 1 for (horizontal) dimension 1, $n_1 = 15$, while for (vertical) dimension 2, $n_2 = 7$. When $n_1 = n_2$ we can speak of a 'square lattice'.

Figure 1 brings to mind regularly spaced agricultural plots on a field. In fact, 'spatial long memory' goes back at least to the agriculturally motivated paper of Smith (1938), which is also a very early reference relative to the literature on time series long memory. Consider $n = n_1 n_2$ agricultural plots on a field, as in Fig. 1. Denote the yield (of rice, say) at location $i = (i_1, i_2)$ by $x_{i_1 i_2}$ and the sample mean by $\bar{x} = n^{-1} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} x_{i_1 i_2}$. Smith (1938) assumed in effect that

$$\text{Var}(\bar{x}) \sim C n^{2d-1}, 0 \leq d < 1/2, C > 0, \quad (1)$$

and estimated the unknown d and C from the logged relation

$$\log(\text{Var}(\bar{x})) \sim \log C + (2d - 1) \log n,$$

meaning that the left side is estimated from data over several fields. Now, in the special square lattice case, $n_1 = n_2 = n^{1/2}$, the model (1) is consistent with the underlying isotropic model

$$\text{Cov}(x_{i_1 i_2}, x_{j_1 j_2}) \sim c \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}^{2d-1} = c \|i - j\|^{2d-1},$$

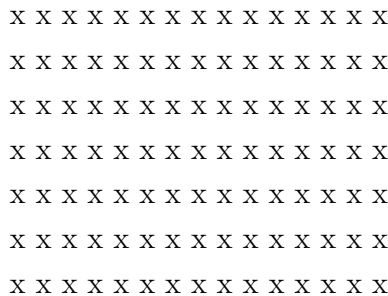


Fig. 1 Spatial lattice

for $i = (i_1, i_2)$, $j = (j_1, j_2)$, which is equivalent to the usual autocovariance function of a stationary long memory time series with differencing or memory parameter d , e.g., a fractional autoregressive integrated moving average (FARIMA). To verify (1), note that

$$\begin{aligned}\text{Var}(\bar{x}) &= n^{-2} \sum_{i_1, i_2} \sum_{j_1, j_2} \text{Cov}(x_{i_1 i_2}, x_{j_1 j_2}) \\ &\sim cn^{-2} \sum_{i_1, i_2} \sum_{j_1, j_2} \sqrt{\left((i_1 - j_1)^2 + (i_2 - j_2)^2\right)^{2d-1}} \\ &= cn^{-2} \sum_{h, k} (n^{1/2} - |k|)(n^{1/2} - |h|) \sqrt{h^2 + k^2}^{2d-1} \\ &\sim Cn^{2d-1}.\end{aligned}$$

It is interesting that Smith (1938) thought of a power law decay: this might seem natural to one coming to the subject unschooled in time series modeling which, after World War II, stressed exponential decay, as in ARMA modeling.

Since then, many papers on 'spatial long memory' have appeared, but the topic has not been developed as systematically or comprehensively as 'long memory time series'. Some distinctive issues arising in the 'spatial' case, all of which have been studied far more under short memory than long memory, are:

1. Is there isotropy or not? If not, we might model each dimension separately, or with some interaction, and possibly with a different memory parameter for each dimension.
2. Is sampling regularly or irregularly spaced? Whereas in time series regular spacing has been predominately studied, irregular spacing is perhaps more likely to be found with spatial data.
3. Should modeling be unilateral or multilateral? For time series unilateral modeling, reflecting one-sided transition from past to future, is usually natural, but this is not the case with spatial data, where, for example, the dimensions might be latitude and longitude.
4. The edge effect. In estimating lagged quantities, there is loss of information at the boundary of the observation region, which has negligible effect when $M = 1$, but increasing, and damaging, effect as M increases unless corrected for.

In this, quite personal, view of the subject, we discuss the following topics in the remaining three sections of the paper:

1. Inference on location and regression with long memory errors.
2. Inference on second-order properties of long memory stationary processes.
3. Miscellaneous topics: nonstationary processes, irregular spacing, adaptive estimation, nonparametric regression.

We do not consider 'spatial autoregressive'-type ('SAR') models (which depend on a user-chosen weight matrix of geographic or economic inverse distances); these do not fit into our framework but typically possess a kind of short memory.

2 Inference on location and regression with long memory errors

The sample mean \bar{x} is a basic statistic, whose asymptotic properties are well known under time series short and long memory. In the time series case, suppose

$$x_i = \mu + \sum_{j=0}^{\infty} \alpha_j \varepsilon_{i-j}, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \quad (2)$$

where $\{\varepsilon_j\}$ is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance, or even homoscedastic (but not necessarily identically distributed) martingale differences. The model (2) implies x_i is covariance stationary, and it can have 'short memory' or 'long memory' or 'negative memory'.

First suppose x_i has short memory, i.e., x_i has spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{Cov}(x_i, x_{i+j}) \cos(j\lambda), \quad -\pi \leq \lambda \leq \pi,$$

that satisfies

$$0 < f(0) < \infty.$$

with also $f(\lambda)$ continuous at $\lambda = 0$. Then, $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ is an asymptotically normal and efficient estimate of μ ,

$$\sqrt{n}(\bar{x} - \mu) \rightarrow_d \mathcal{N}(0, 2\pi f(0)), \quad \text{as } n \rightarrow \infty.$$

Long memory can be described by (2) with, for example,

$$\alpha_j \sim \ell(j)j^{d-1}, \quad 0 < d < 1/2, \quad (3)$$

for a slowly varying function $\ell(j)$ so $f(0) = \infty$. We can also consider the asymptotic approximation in (3) for $-1/2 < d < 0$, assuming also $\sum_{j=0}^{\infty} \alpha_j = 0$, whence $f(0) = 0$, and there is 'negative memory'. In both cases, there is interest in allowing for a nonconstant $\ell(j)$ (e.g., $\ell(j) = \log j$), and this has been done in limit theory for basic statistics, and also in (semiparametric) estimation of d . But for simplicity, we focus on constant $\ell(j)$, which covers FARIMA and fractional Brownian motion models. In this setting, \bar{x} is no longer asymptotically efficient (Adenstedt 1974) but it can still be asymptotically normal, albeit with a different rate of convergence. In particular, under suitable conditions on $\{\varepsilon_j\}$,

$$n^{1/2-d}(\bar{x} - \mu) \rightarrow_d \mathcal{N}(0, \sigma^2(d)) \quad \text{as } n \rightarrow \infty, \quad 0 < |d| < 1/2,$$

for a function $\sigma^2(d)$ of known form.

More generally consider the centered and self-normalized statistic

$$u = (\bar{x} - \mu) / \left\{ E \left(n^{-1} \sum_{i=1}^n (x_i - \mu) \right)^2 \right\}^{1/2}.$$

Ibragimov and Linnik (1971) showed that

$$u \rightarrow_d \mathcal{N}(0, 1), \text{ as } n \rightarrow \infty,$$

if only the ε_j are iid and

$$\sum_{j=0}^{\infty} \alpha_j^2 < \infty, E \left(\sum_{i=1}^n (x_i - \mu) \right)^2 \rightarrow \infty, \text{ as } n \rightarrow \infty.$$

Note that under negative memory

$$E \left(\sum_{i=1}^n (x_i - \mu) \right)^2$$

diverges more and more slowly as $d \downarrow -1/2$, indeed in the moving average (MA) model with an MA unit root, $y_t = \mu + \varepsilon_t - \varepsilon_{t-1}$, we have $\sum_{i=1}^n (x_i - \mu) = \varepsilon_n - \varepsilon_0$, and it does not diverge at all.

For spatial data, consider an M -dimensional lattice, as described in the previous section (though results can be extended to a more general, irregular region). Let i be the multiple index (i_1, \dots, i_M) with $i_j \in \mathbb{Z} = \{0, \pm 1, \dots\}$, $j = 1, \dots, M$. Consider a covariance stationary process x_t observed on the lattice $\mathbb{N} = \{i : 1 \leq i_m \leq n_m, m = 1, \dots, M\}$,

$$x_i = \sum_{j \in \mathbb{Z}^M} \alpha_{i-j} \varepsilon_j, \sum_{j \in \mathbb{Z}^M} \alpha_j^2 < \infty,$$

for ε_j iid (say) with zero mean and unit variance. For example under isotropy, with $M = 2$,

$$\alpha_j \sim \|j\|^{-\delta} = (j_1^2 + j_2^2)^{-\delta/2},$$

with memory parameter $\delta > 1$. Note that for time series, long memory has been 'explained' as arising from cross-sectional aggregation of AR(1) processes, see, Robinson (1978), Granger (1980), an interpretation extended to spatial processes by, e.g., Lavancier (2011).

For $n = \prod_{m=1}^M n_m$ and the multiple partial sum process

$$S = \sum_{i \in \mathbb{N}} x_i,$$

Specializing their results to the case $M = 2$, Lahiri and Robinson (2016) obtained

$$n^{\delta-3}S \rightarrow_d \mathcal{N}(0, \sigma^2(\delta)),$$

with

$$\delta \in (1, 2) \text{ (long memory),}$$

$$\delta \in (2, 2.5), \sum_{j \in \mathbb{Z}^2} \alpha_j = 0 \text{ (negative memory),}$$

and also

$$n^{-1/2}S \rightarrow_d \mathcal{N}(0, \sigma^2(\delta))$$

with

$$\delta > 2.5, \sum_{j \in \mathbb{Z}^2} \alpha_j = 0.$$

Such results can be extended to regression models, especially ones with deterministic regressors. One interesting case, for $M = 2$, and with y_i a response variable that is observable on the lattice is, with $\beta_0, \beta_1, \beta_2, \theta_1, \theta_2$ all unknown,

$$y_i = \beta_0 + \beta_1 t_1^{\theta_1} + \beta_2 t_2^{\theta_2} + x_i,$$

with x_i now an unobservable error process. For a more general regression model, and with general M , Robinson (2012) considered nonlinear least squares estimation of the β 's and θ 's, for short memory x_i , where $\theta_1 > -1/2$, $\theta_2 > -1/2$ is required (with negative values of the θ_i conferring a decaying trend that could be relevant for some spatial data). In this case, the estimates are asymptotically efficient (in the Gauss–Markov sense), but the β 's are estimated slightly less well than the θ 's, with rates depending on the θ 's (and slightly less well than if the θ 's were known). With long memory x_i , where regression powers will 'interact' with memory parameters, the rates will be slower, and if memory is strong enough relative to the magnitude of the θ 's we may not be able to estimate some β 's and θ 's, due to domination of the regression component by the error x_i .

3 Inference on second-order properties of long memory stationary processes

Now consider estimating either a full parametric model for the spectrum/autocovariance function of x_i , or else a 'semiparametric' one specified for only low frequencies/long lags. In the parametric case, for short memory time series, early work on asymptotics covered least squares and Yule–Walker estimates of stationary autoregressive (AR) processes. Continuous and discrete frequency Whittle, and Gaussian pseudo-maximum likelihood, estimates of ARMA and other short memory time series were well covered by Hannan (1973). He relaxed the commonly imposed iid assumption on

innovations, allowing them, and centred squared innovations, to be stationary martingale differences (expressing the natural ordering of time series data), with only finite second moments required (for estimation of dependence parameters); in fact stationarity can be relaxed, the squared innovations needing only to be uniformly integrable). His central limit theorem proof was essentially found to work under long memory with differencing parameter $d \in (0, 1/4)$ by Yajima, (1985). Using a different method of proof, Fox and Taqqu (1986) established a central limit theorem valid over the whole stationary long memory interval $d \in (0, 1/2)$.

For spatial lattice data, a vital early (short memory) reference is Whittle (1954). Noting that multilateral MA/AR representations are more natural for spatial data than the unilateral or one-sided ones normal in time series, he pointed out identifiability problems with multilateral representations, and extended the (unilateral) Wold representation for purely nondeterministic time series (cf. (1)) to spatial processes, introducing 'half-plane' representations. But a given multilateral ARMA does not necessarily have a neat half-plane representation, where AR or MA orders may be infinite; so this property may have limited practical use. Sometimes, a 'quarter plane' representation is possible.

Whittle (1954) and others (e.g., Martin 1979, Tjøstheim 1978, Kashyap and Lapsa 1984; Huang and Anh 1992; Ma 2003) considered estimation of various short memory parametric spatial models, some extending ARMA in an isotropic or separable way, e.g., for $M = 2$,

$$(1 - \alpha_1 L_1)(1 - \alpha_2 L_2)x_t = \varepsilon_t,$$

where $L_1 x_t = x_{t_1-1, t_2}$, $L_2 x_t = x_{t_1, t_2-1}$ and the ε_t are iid, a kind of first-order, unilateral AR in 2 dimensions. Other specifications, such as the Matern model, were considered by Stein (1999).

For asymptotic theory of estimates, one basic question is whether to use increasing domain (as usually in time series) or fixed-domain (infill) asymptotics (or some hybrid of the two, where the domain increases slowly while the sampling interval decays slowly). Since spatial observations can often be thought of as confined to a bounded region, infill asymptotics, as studied by Stein (1999), might seem natural. However, infill asymptotics can lead to results that are not easy to use in practice and even to inconsistency, whereas increasing domain asymptotics often provides a central limit theorem that is convenient for practical use.

Another important issue is the 'edge effect'. Consider estimating the autocovariance $\gamma_j = \text{Cov}(x_i, x_{i+j})$ of a (zero mean) covariance stationary process x_i on an M -dimensional lattice, by

$$\hat{\gamma}_j = \frac{1}{n} \sum_{i, i+j \in \mathbb{N}} x_i x_{i+j}.$$

This is of great interest because various estimates of parameters, in particular ones based on a Gaussian pseudo-likelihood or approximation thereto (and indeed nonparametric spectral estimates) are essentially functions of the $\hat{\gamma}_j$. For $M = 1$, as in time series, the summation includes $n - j$ points and so (for fixed j), $\hat{\gamma}_j$ has bias $O(n^{-1})$, which does not prevent a central limit theorem for $n^{1/2}(\hat{\gamma}_j - \gamma_j)$ from

holding. But for $M = 2$, and $\mathbb{N} = [1, n^{1/2}] \times [1, n^{1/2}]$ the bias is of exact order $n^{-1/2}$ so the limit distribution of $n^{1/2}(\hat{\gamma}_j - \gamma_j)$ has nonzero mean. And for $M = 3$ and $\mathbb{N} = [1, n^{1/3}] \times [1, n^{1/3}] \times [1, n^{1/3}]$ the bias is of exact order $n^{-1/3}$; so, no central limit theorem results, likewise for $M > 3$. Thus, the obvious extensions of time series estimates to such spatial lattice data do not work.

The following solutions, have been proposed, all in a short memory context.

1. (Guyon (1982)) Use unbiased estimates

$$\tilde{\gamma}_j = n\hat{\gamma}_j / \#\{i : i, i+j \in \mathbb{N}\}$$

of γ_j . However the $\tilde{\gamma}_j$ become unstable for large $\|j\|$ and a desirable non-negative definiteness property of the $\hat{\gamma}_j$ is lost.

2. (Dahlhaus and Kuensch (1987)) A data taper, of the type extending that used to correct for bias in a variety of time series settings, is applied to x_i .
3. (Robinson and Vidal Sanz (2006); Robinson (2007)) Trimming is employed: omitting the $\tilde{\gamma}_j$ defined above for large $\|j\|$, the first reference considering a parametric setting, the second a nonparametric one.

Because of the greater importance of long lags, i.e., large $\|j\|$, in long memory processes relative to short memory ones, the edge effect seems an even bigger issue under long memory.

Nevertheless, a number of parametric long memory stationary linear spatial models, both isotropic and separable, have been considered in the literature, see, e.g., Sethuram and Basawa (1995), Boissy et al. (2005), Ghodsi and Shitan (2009), Beran et al. (2009). A simple illustration of a separable long memory spatial parametric model is

$$(1 - L_1)^{d_1} (1 - L_2)^{d_2} x_i = \varepsilon_i, \quad 0 < d_1, d_2 < 1/4.$$

For cyclic/seasonal long memory models, with a spectral pole at one or more non-zero frequencies, see, e.g., Cisse et al. (2016).

For time series 'semiparametric' estimates of the memory parameter, only local-to-zero assumptions are made about the spectrum, and an increasing number but vanishing fraction of Fourier frequencies near the origin, depending on a user-chosen bandwidth, are used. These have also been extended to spatial lattice data. Log periodogram spatial estimates have been considered by Ghodsi and Shitan (2016), the latter extending asymptotic theory of Robinson (1995a) (though Ghodsi and Shitan (2016) assume a parametric model). Local Whittle or Gaussian semiparametric estimates have been developed by Guo et al. (2009), Durastanti et al. (2014), extending asymptotic theory of Robinson (1995b).

There is considerable scope for further study of issues of model choice (especially due to the danger of 'curse of dimensionality' in lattice models) and bandwidth choice. Basic asymptotic theory for estimation of second-order features of long memory lattice processes has been developed by Lavancier (2006, 2007, 2008), and Lavancier and Philippe (2011).

Sometimes, though observations are made only at discrete points in \mathbb{Z}^M , it is natural to think in terms of an underlying continuous model defined on \mathcal{R}^M , where \mathcal{R} denotes the real line. This topic was long ago extensively discussed in the time series case $M = 1$, especially for short memory models. The most popular covariance stationary model here has been expressed as a rational spectral density, of a continuous process $x(t)$, $t \in \mathcal{R}$,

$$g(\lambda) = \frac{1}{2\pi} \frac{|b(i\lambda)|^2}{|a(i\lambda)|^2}, \lambda \in \mathcal{R}, \quad (4)$$

where a and b are polynomials of degrees r and s , respectively such that $r > s$ (required for integrability of g , equivalently for $x(t)$ to have finite variance), and such that the zeroes of a have negative real parts. When $s = 0$, we can think of $x(t)$ as generated by a stochastic differential equation (of order r) driven by 'continuous white noise', that is an unobservable process $\zeta(t)$, $t \in \mathcal{R}$, having (non-integrable) spectral density

$$h(\lambda) = \frac{1}{2\pi}, \lambda \in \mathcal{R}, \quad (5)$$

the b factor in (4) providing further structure when $r > 1$, $s > 0$. Suppose that $x_t = x(t)$ is observable at $t \in \mathbb{Z}$ only. Then (see, e.g., Bartlett 1946, Walker 1960, Phillips 1959; Robinson 1980a) x_t has an ARMA($r, r - 1$) representation. This can be established in more than one way. Note that, if $f(\lambda)$ denotes the spectral density of x_t ,

$$f(\lambda) = \sum_{j=-\infty}^{\infty} g(\lambda + 2\pi j), \quad (6)$$

and Robinson (1980a) thence used (6) with contour integration/theorem of residues to derive the ARMA($r, r - 1$) property of x_t , also using it to derive the ARMA(r, r) property of the averaged process $X_t = \int_{t-1}^t x(u)du$, $t \in \mathbb{Z}$ (as well as analogous properties when the underlying process is skip-sampled, being defined discretely at narrow intervals $1/k$, for an integer $k \geq 2$, see also Telser (1967)). Generally, there are efficiency gains if the underlying parameters in a and b are estimated, rather than an unconstrained ARMA. But $f(\lambda)$ is naturally parameterised in terms of the zeroes of a and b , and when $r \geq 2$ its form depends on the extent of any multiplicities in a , whose number is likely unknown, even for given r . Note also that due to the aliasing property (6), a cannot be uniquely estimated from the AR coefficients of x_t when $r \geq 2$, and though the presence of b may help, a Gaussian pseudo-likelihood may have multiple peaks from which it is difficult to choose (Robinson 1980b). Generally, corresponding to (6), there are uncountably many continuous processes consistent with a given discrete one, so attempting to estimate a continuous model is highly hazardous. It should be added that though (4) is classical, the model has two parts, the frequency response function b/a and the noise ζ , and either might be specified differently, for example, the aliasing problem is avoided if ζ is band limited to Nyquist frequency, so instead of (5)

$$h(\lambda) = 0, |\lambda| > \pi, \quad (7)$$

leading to a quite different model for the discretely observed x_t , see Robinson (1977a), while Robinson (1976a, b) evaluated bias of estimates based on (7) when in fact (4) holds, again using the theorem of residues. Extensions to vector $x(t)$ were developed by Phillips (1974), Robinson (1993). The topic has been rediscovered more recently, with (4) referred to by the acronyms CAR (when $s = 0$) and CARMA (when $s > 0$).

For long memory, one can consider a continuous process such as

$$S(\lambda) = \frac{\lambda^{-2d} |b(i\lambda)|^2}{2\pi |a(i\lambda)|^2}, \lambda \in \mathcal{R},$$

with memory parameter $d \in (0, 1/2)$. Again (6) can be applied but there is now no neat analytic solution characterizing x_t and $f(\lambda)$ (see Comte and Renault 1996; Chambers 1998).

For continuously defined spatial processes that are observed on a lattice, extensions of the above are clearly possible (see, e.g. Matsuda 2019), and analogous issues discussed for the time series case apply.

4 Miscellaneous topics

4.1 Nonstationary processes

Really, a time series AR with a unit root is a 'long memory' process because it has even longer memory than a stationary long memory process, and it coincides with a fractional process with memory parameter $d = 1$. Asymptotic theory for unit root processes in an AR setting produces nonstandard limit distributions, but nesting the unit root in the fractional, rather than AR, class leads to standard asymptotics, for example a central limit theorem with $n^{1/2}$ norming and classical local power properties, for memory and other parameter estimates, even under nonstationarity, using either tapering and skipping of frequencies in discrete-frequency Whittle estimation (see Velasco and Robinson 2000), or using conditional sum-of-squares estimation (see Hualde and Robinson 2011). These ideas seem capable of extension to nonstationary fractional processes on a spatial lattice. Kuensch (1987) considered an alternative class of nonstationary spatial processes.

In a semiparametric time series setting, Velasco (1999) studied log-periodogram regression estimation under nonstationarity, using tapering and skipping of frequencies, as did Yajima and Matsuda (2019) for data on a spatial lattice of dimension $M = 2$ with an underlying continuous nonstationary isotropic process. In econometrics, study of nonstationary time series has much focussed on multivariate processes, and cointegration and fractional cointegration. Clearly, there is scope for corresponding study of multivariate spatial processes.

4.2 Irregular spacing

There is a time series literature on missing and irregularly spaced data, for example, Robinson (1977b) considered the underlying continuous model (4) with $r = 1$, $s = 0$, and showed that irregularly spaced observations have a kind of time-varying AR representation, with unknown parameters estimable by Gaussian pseudo likelihood. Irregular spacing is more likely to be an issue with spatial data. A general approach, for a parametric underlying continuous stationary spatial process, would be to construct a Gaussian pseudo-likelihood, conditional on the observation points. But conditions for asymptotic theory are messy relative to the regularly spaced case, mixing properties of the continuous process $x(t)$ with those of the observation points. Matsuda and Yajima (2009) developed parametric and nonparametric methods and theory in the frequency domain, for spatial data with an underlying model defined on \mathcal{R}^M . Another approach that could be pursued in an irregularly spaced long memory spatial setting involves averaging over the process generating the observation times as well as over $x(t)$ itself. In particular, in the time series case, Robinson (1980b) considered the model (4) with irregularly spaced observation points τ_k , $k \in \mathbb{Z}$, generated by a point process, deriving autocovariance properties that are not conditional on the observation times. For example, treating the time intervals as a renewal process, he obtained an ARMA($r, r - 1$) representation for the observed pseudo-regularly spaced process $X_k = x(\tau_k)$, $k \in \mathbb{Z}$.

Sometimes irregular spacing is due to missing from a regular lattice. Parzen (1963) proposed an amplitude modulating sequence for missing data in time series x_t , $t \in \mathbb{Z}$, studying the regularly spaced process $a_t x_t$, where a_t is a (possibly stochastic) zero–one sequence, taking the value 1 when x_t is observed, and 0 when it is missed. Dunsmuir and Robinson (1981) studied Whittle estimation of parametric models using this approach. It seems clear how it could be extended to cover spatial lattice processes with missing observations and long memory.

4.3 Adaptive estimation in semiparametric models

As is well known, under regularity conditions Gaussian pseudo-maximum likelihood estimates for parametric models are typically consistent and asymptotically normal with the same asymptotic covariance matrix irrespective of whether or not Gaussianity actually holds. But such estimates are not asymptotically efficient if the process is non-Gaussian. Considering a semiparametric model, with innovations having distribution of unknown form, it is possible to obtain asymptotically efficient ‘adaptive’ estimates, using nonparametric smoothing. For long memory time series, this was studied by Hallin and Serroukh (1998) under stationarity, and by Robinson (2005) under stationarity and nonstationarity. Such ideas are certainly extendable to spatial lattice processes.

4.4 When ordering does not matter

Unlike with time series data, there is no natural uni-dimensional ordering of spatial data. For a spatial lattice, there is typically an ordering of locations only with respect to each of the dimensions. For inference on many features, such as spatial

dependence parameters, respect for relative spatial locations is crucial. But for estimating some 'instantaneous' features, such as location and static parametric regression, and in stochastic design nonparametric regression, ordering can be disregarded and the spatial process $x_i, i \in \mathbb{N}$, mapped arbitrarily into a sequence $u_i, i = 1, \dots, n$.

We might assume, say (Robinson 2011), the nonstationary infinite MA

$$u_i = \sum_{j=1}^{\infty} a_{ij} \eta_j, \quad 1 \leq i \leq n, \quad \sup_i \sum_{j=1}^{\infty} a_{ij}^2 < \infty, \quad (8)$$

which can be checked in some spatial model settings. We have 'long memory' if

$$\frac{1}{n} \sum_{i,j=1: i \neq j} \text{Cov}(u_i, u_j) = \frac{1}{n} \sum_{i,j=1: i \neq j} \sum_{k=1}^{\infty} a_{ik} a_{jk} \rightarrow \infty, \quad \text{as } n \rightarrow \infty.$$

For stochastic design nonparametric (kernel) regression, with spatial data, Robinson (2011) gave conditions for asymptotic normality, based partly on assuming (8) for the error process and on conditions on the distance between multivariate and product univariate densities for the regressors, which can also cover long memory. Such conditions might be extended to other models and methods for spatial data.

Compliance with ethical standards

Conflict of interest The author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics*, 6, 1095–1107.
- Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Supplement to the Journal of the Royal Statistical Society*, 8, 27–41.
- Beran, J., Ghosh, S., & Schell, D. (2009). On least squares estimation for long-memory lattice processes. *Journal of Multivariate Analysis*, 100, 2178–2194.
- Boissy, Y., Bhattacharyya, B. B., Li, X., & Richardson, G. D. (2005). Parameter estimates for fractional autoregressive spatial processes. *Annals of Statistics*, 33, 2553–2567.
- Chambers, M. J. (1998). Long memory and aggregation in macroeconomic time series. *International Economic Review*, 39, 1053–1072.
- Comte, F., & Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics*, 73, 101–149.
- Cisse, P. O., Dongue, A. K., & Guegan, D. (2016). Statistical properties of the seasonal fractionally integrated separable spatial autoregressive model. *Afrika Statistika*, 11, 901–922.
- Dahlhaus, R., & Kuensch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74, 877–882.

- Dunsmuir, W., & Robinson, P. M. (1981). Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association*, 76, 560–568.
- Durastanti, C., Lan, X., & Marinucci, D. (2014). Gaussian semiparametric estimates on the unit sphere. *Bernoulli*, 20, 28–77.
- Fox, J., & Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, 14, 517–532.
- Ghods, A., & Shitan, M. (2009). Estimation of the memory parameters of the fractionally integrated separable spatial autoregressive (FISSAR(1, 1)) model: A simulation study. *Communication in Statistics-Simulation and Computation*, 6, 1256–1268.
- Ghods, A., & Shitan, M. (2016). Asymptotic properties of GPH estimators of the memory parameters of the fractionally integrated separable spatial ARMA (FISSARMA) models. *Probability and Mathematical Statistics*, 36, 247–265.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Guo, H., Lim, C. Y., & Meerschaert, M. (2009). Local Whittle estimator for anisotropic random fields. *Journal of Multivariate Analysis*, 100, 993–1028.
- Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69, 95–105.
- Hallin, M., & Serroukh, A. (1998). Adaptive estimation of the lag of a long-memory process. *Statistical Inference for Stochastic Processes*, 1, 111–129.
- Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10, 130–145.
- Hualde, J., & Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics*, 39, 3152–3181.
- Huang, D., & Anh, V. (1992). Estimation of spatial ARMA models. *Australian Journal of Statistics*, 34, 513–530.
- Ibragimov, I. A., & Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff.
- Kashyap, R. L., & Lapsa, P. M. (1984). Synthesis and estimation of random fields using long-correlation models. *IEEE Transactions on Pattern Analysis of Machine Intelligence*, 6, 800–809.
- Kuensh, H. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika*, 74, 517–524.
- Lahiri, S. N., & Robinson, P. M. (2016). Central limit theorems for long range dependent spatial linear processes. *Bernoulli*, 22, 345–375.
- Lavancier, F. (2006). Long memory random fields. In P. Bertail, et al. (Eds.), *Dependence in Probability and Statistics*. New York: Springer.
- Lavancier, F. (2007). Invariance principles for non-isotropic long memory random fields. *Statistical Inference for Stochastic Processes*, 10, 255–282.
- Lavancier, F. (2008). The V/S test of long-range dependence in random fields. *Electronic Journal of Statistics*, 2, 1373–1390.
- Lavancier, F. (2011). Aggregation of isotropic autoregressive fields. *Journal of Statistical Planning and Inference*, 141, 3862–3866.
- Lavancier, F., & Philippe, A. (2011). Some convergence results on quadratic forms for random fields and application to empirical covariances. *Probability Theory and Related Fields*, 149, 493–514.
- Ma, C. (2003). Power-law correlations and other models with long-range dependence on a lattice. *Journal of Applied Probability*, 4, 690–703.
- Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika*, 66, 209–217.
- Matsuda, Y. (2019). Bivariate CARMA random fields. Preprint.
- Matsuda, Y., & Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d . *Journal of the Royal Statistical Society Series B*, 71, 191–217.
- Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhya Series A*, 25, 383–392.
- Phillips, A. W. (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika*, 46, 67–76.
- Phillips, P. C. B. (1974). The problem of identification in finite-parameter continuous-time models. *Journal of Econometrics*, 1, 351–362.

- Robinson, P. M. (1976a). The aliasing problem in differential equation estimation. *Journal of the Royal Statistical Society Series B*, 38, 180–188.
- Robinson, P. M. (1976b). Some problems in the identification and estimation of continuous time systems from discrete time data. In R. K. Mehra & D. G. Lainiotis (Eds.), *System identification: Advances and case studies* (pp. 407–439). Cambridge: Academic Press.
- Robinson, P. M. (1977a). The construction and estimation of continuous time models and discrete approximations in econometrics. *Journal of Econometrics*, 6, 173–197.
- Robinson, P. M. (1977b). Estimation of a time series model from unequally spaced data. *Stochastic Processes and Their Applications*, 6, 9–24.
- Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5, 163–168.
- Robinson, P. M. (1980a). Efficient estimation of a rational spectral density. In Proceedings of the First European Signal Processing Conference, North-Holland, 701–704.
- Robinson, P. M. (1980b). Continuous model fitting from discrete data. In Directions in Time Series (D.R. Brillinger and G.C. Tiao, Eds.), Institute of Mathematical Statistics, 263–278.
- Robinson, P. M. (1993). Continuous-time models in econometrics: Closed and open systems, stocks and flows. In P. Phillips (Ed.), *Models Methods and Applications of Econometrics* (pp. 71–90). Basil: Blackwell.
- Robinson, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23, 1048–1072.
- Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23, 1630–1661.
- Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics*, 25, 2054–2083.
- Robinson, P. M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics*, 33, 1800–1842.
- Robinson, P. M. (2007). Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference* (issue in honour of Madan L. Puri), 137, 1024–1034.
- Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, 165, 5–19.
- Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli*, 18, 644–677.
- Robinson, P. M., & Vidal Sanz, J. (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis*, 97, 1090–1120.
- Sethuram, S., & Basawa, J. V. (1995). Maximum likelihood estimation for a fractionally differenced autoregressive model on a two-dimensional lattice. *Journal of Statistical Planning and Inference*, 44, 219–235.
- Smith, H. Fairfield. (1938). An empirical law, describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, 28, 1–23.
- Stein, M. (1999). *Interpolation of spatial data*. New York: Springer.
- Telsner, L. G. (1967). Discrete samples and moving sums in stationary stochastic processes. *Journal of the American Statistical Association*, 62, 484–499.
- Tjøstheim, D. (1978). Statistical spatial series modelling. *Advances in Applied Probability*, 10, 130–154.
- Velasco, C. (1999). Non-stationary log-periodogram regression. *Journal of Econometrics*, 91, 325–371.
- Velasco, C., & Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, 95, 1229–1243.
- Walker, A. M. (1960). Note on a generalisation of the large sample goodness of fit test for linear autoregressive schemes. *Journal of the Royal Statistical Society Series B*, 12, 102–107.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41, 434–449.
- Yajima, (1985). On estimation of long-memory time series models. *Australian Journal of Statistics*, 27, 303–320.
- Yajima & Matsuda, Y. (2019). Log-periodogram regression of two-dimensional intrinsic stationary random fields. Preprint.