Received: 27 May 2022 | Revised: 25 April 2023 | Accepted: 2 May 2023

DOI: 10.1002/env.2817

INVITED ARTICLE

WILEY

Long memory conditional random fields on regular lattices

Angela Ferretti[1] | **L. Ippoliti**[1] | **P. Valentini**[1] | **R. J. Bhansali**[2]

[1]Department of Economics, University G. d'Annunzio, Pescara, Italy

[2]Division of Statistics and Probability, Department of Mathematical Sciences, University of Liverpool, Liverpool, UK

Correspondence
L. Ippoliti, Department of Economics, University G. d'Annunzio, Viale Pindaro 42, 65127 Pescara, Italy.
Email: luigi.ippoliti@unich.it

Abstract

This paper draws its motivation from applications in geophysics, agricultural, and environmental sciences where empirical evidence of slow decay of correlations have been found for data observed on a regular lattice. Spatial ARFIMA models represent a widely used class of spatial models for analyzing such data. Here, we consider their generalization to conditional autoregressive fractional integrated moving average (CARFIMA) models, a larger class of long memory models which allows a wider range of correlation behavior. For this class we provide detailed descriptions of important representative models, make the necessary comparison with some other existing models, and discuss some important inferential and computational issues on estimation, simulation and long memory process approximation. Results from model fit comparison and predictive performance of CARFIMA models are also discussed through a statistical analysis of satellite land surface temperature data.

KEYWORDS

CARFIMA models, conditional autoregression, Gauss-Markov random fields, long memory random fields, spatial processes

1 | INTRODUCTION

Long memory (LM) processes are known to play a significant role in many scientific disciplines. Historically, these processes were first introduced by Kolmogorov (1941) to model scaling behavior in turbulence. In parallel, Wiener considered extension of the Wiener process to more general diffusion processes, including fractional Brownian motion. Subsequently, the subject was brought to the attention of a wider audience by the pioneering works of Mandelbrot and colleagues who defined the notion of fractional Gaussian noise (Mandelbrot & van Ness, 1968) as a model for the LM observed by Hurst in the Nile river flood records. A similar path-breaking role can also be attributed to Granger for studies in economics (Granger, 1980).

Since these early contributions, different approaches have been introduced to model LM in temporal and spatial data. In the temporal case, fractional time series models (see, e.g., Beran et al., 2013) and regression models (see, Geweke & Porter-Hudak, 1983) have become increasingly popular with applications in many areas, as for example, astronomy (Newcomb, 1985), hydrology (Hurst, 1951), econometrics (Baillie et al., 1996; Porter-Hudak, 1990), biology (Mandelbrot & Wallis, 1969), meteorology (Haslett & Raftery, 1989) and environmental sciences (Burrough, 1981; Sørbye & Rue, 2017), and telecommunication (Beran et al., 2013; Doukhan et al., 2003). Further applications and theory in time series are also reviewed in Cox (1984), Beran (1992), and Samorodnitsky (2006).

The simplest processes with LM are stationary processes with correlations decaying hyperbolically. In particular, their autocorrelation function, decaying as a power law in the lag variable, sums to infinity for large lag values. Because of this slow decay, LM processes are also referred to as long range dependent (LRD) processes. Two well-known classes of stationary time series processes with slowly decaying correlations are the fractional ARIMA(p, d, q) model, introduced by Granger and Joyeux (1980) and Hosking (1981), and the fractional exponential—FExp(s)—model of Beran (1993). With the inclusion of the fractional (LM) parameter d, these models are a natural generalization of standard ARIMA models defined in Box and Jenkins (1970) and of the Exponential model of Bloomfield (1973). The development of a theory of statistical inference for these models has become a very active field of research in the past years. In particular, the models became fashionable through time and favored also the development of new approaches for modeling spatial data with LM.

The notion of spatial LM dates back to the works of Smith (1938) and Whittle (1956, 1962) on agricultural data. In the geostatistical framework, Matheron (1973) also developed the mathematical theory for intrinsic and self-similar processes; however, the practical significance of these topics have not been fully recognized until Mandelbrot's pioneering work (see, comments in Beran et al., 2013). Further proposals of spatial models with long-range dependence have been formulated in Atmospheric sciences (see, e.g., Anh et al., 2000; Beran et al., 2009; Gneiting, 2000; Kashyap & Lapsa, 1984; Kohli, 2016Lin et al., 2007;), Oceanography (Percival et al., 2008), ground water flow (Guo et al., 2009), statistical physics (Wang, 2009), and image analysis (Anh & Lunney, 1995). More theoretical details can be found also in Leonenko (1999), Gneiting and Schlather (2004), Lavancier (2006), and Mateu et al. (2007).

Several papers on spatial LM have appeared in the last decade. However, as highlighted by Robinson (2020), the topic has not been developed as systematically or comprehensively as LM time series and, to the best of our knowledge, there are no comprehensive surveys available in the spatial framework. Some distinctive issues arising in the standard spatial statistics, as for example those regarding inference on second-order properties of stationary random fields, have been studied far more under short memory than LM. Furthermore, we have also found that some of the theoretical results achieved for LM random fields appeared in less known journals or dissertations and, accordingly, a complete literature is not readily available or properly quoted in scholarly publications.

Our work thus provides an overview of LM random fields for both interested researchers who want to enter this developing field, and experts who would like to compare some of the models presented in the literature. In particular, our paper makes contributions summarized as follows:

- **New class of LM models**. By exploiting the stochastic properties of conditional spatial models (Cressie, 1993; Ippoliti et al., 2013), we define the new class of conditional autoregressive fractional integrated moving average (CARFIMA) models which includes some existing LM models (e.g., spatial ARFIMA) as special cases.
- **Comprehensive review**. By introducing the CARFIMA models, we make the necessary comparison with some important representative models existing in the literature, and discuss some inferential and computational issues both on estimation and LM process approximation.
- **Future directions**. We discuss some theoretical aspects of LM random fields, analyze the limitations of existing methods, and suggest possible future research directions in terms of modeling strategies, parameter estimation and spatial interpolation.

The remainder of the manuscript is organized as follows. Section 2 provides an introduction to conditional random fields, while Section 3 provides details on LM random fields. In Section 4, we introduce and discuss the class of Gaussian conditional LM models through different examples. Model parameter estimation is then discussed in Section 5, and simulation results are presented in Section 6 to compare different estimation methods. In Section 7, we analyze land surface temperatures, addressing computational issues in an interpolation problem. Section 8 briefly discusses approximating a conditional LM process with a Conditional Autoregressive (CAR) model or Gaussian Markov random field (GMRF). Finally, Section 9 concludes the paper with a discussion and suggestions for future research.

2 | CONDITIONAL SPATIAL MODELS FOR A REGULAR LATTICE

Suppose that $X = \{X_{\boldsymbol{t}}, \boldsymbol{t} \in \mathbb{Z}^k, k \geq 1\}$ is a weakly stationary real-valued random field with zero mean, covariance function, $R_x(\boldsymbol{u}) = \mathrm{Cov}(X_{\boldsymbol{t}}, X_{\boldsymbol{t}+\boldsymbol{u}})$, and correlation function, $r_x(\boldsymbol{u}) = R_x(\boldsymbol{u})/\sigma_x^2$, where $\sigma_x^2 = R_x(\boldsymbol{0})$. The spectral density function of X

is defined as follows

$$f_x(\lambda) = (2\pi)^{-k} \sum_{\boldsymbol{u}\in\mathbb{Z}^k} R_x(\boldsymbol{u})\, e^{-i\boldsymbol{u}'\lambda}, \quad \lambda \in T^k, \quad T^k = [-\pi, \pi]^k.$$

We consider the situation in which the data, $\boldsymbol{x} = \{X_{\boldsymbol{t}}, \boldsymbol{t} \in \mathcal{L}\}$, are observed on a finite lattice $\mathcal{L}$ such that $\boldsymbol{x}$ may be regarded as a partial realization of X. This means that for each $\boldsymbol{t} \in \mathcal{L}$, $x_{\boldsymbol{t}}$ is an observed value of $X_{\boldsymbol{t}}$.

The question of constructing suitable finite parameter stochastic models for the observed data has attracted much attention lately and several different models have been introduced—see, Guyon (1995) and Cressie (1993). As discussed in Ippoliti et al. (2013), the existing models may be summarized under the following three broad headings: (1) Simultaneous, (2) Unilateral, and (3) Conditional.

The simultaneous models were proposed by Whittle (1954) and express $X_{\boldsymbol{t}}$ for each $\boldsymbol{t}$ as a linear combination of its neighbors plus a random white noise error, $\epsilon_{\boldsymbol{t}}$. Although an aim is to generalize the standard autoregressive models for time series ($k = 1$) to a random field, as pointed out in a pioneering paper by Besag (1974), this class of models suffers from severe logical difficulties; for example, the error $\epsilon_{\boldsymbol{t}}$ is serially correlated with all observations.

The unilateral models (see Martin, 1996; Tjostheim, 1978, 1983), depend on the choice of site ordering. Although they enjoy some useful properties, see Yao and Brockwell (2006), and references therein, they can appear contrived.

The conditional models, have the advantage of being justified in the classical linear least-squares interpolation theory for random fields (Rozanov, 1967). Below we give a brief account of the conditional approach, especially the CAR and CAR moving average models, CARMA. Note that every Simultaneous autoregressive, SAR, model is also a CAR but not vice versa, and for $k \geq 2$, the class of CAR models is larger than that of the SAR models. Similarly, although every Unilateral autoregressive, UAR, model is a CAR, the converse is not true and there are CAR's which do not admit a UAR representation. Hence the class of CAR models is larger than that of UAR models.

Suppose that the covariance function $R_x(\boldsymbol{u})$ is absolutely summable and the spectral density function is "non-vanishing"

$$\sum_{\boldsymbol{u}\in\mathbb{Z}^k} |R_x(\boldsymbol{u})| < \infty, \quad f_x(\lambda) \neq \boldsymbol{0}, \quad \lambda \in T^k. \tag{1}$$

Then, see Guyon (1995), $f_x^{-1}(\lambda) = 1/f_x(\lambda)$ exists for all $\lambda \in T^k$ and admits a Fourier series expansion. Let

$$f_{ix}(\lambda) = (2\pi)^{-2k} f_x^{-1}(\lambda),$$

denote the inverse spectral density function (Yuan & Subba Rao, 1993). Then, $f_{ix}(\lambda)$ admits a Fourier expansion as follows

$$f_{ix}(\lambda) = (2\pi)^{-k} \sum_{\boldsymbol{u}\in\mathbb{Z}^k} R_{ix}(\boldsymbol{u})\, e^{-i\boldsymbol{u}'\lambda}, \quad \lambda \in T^k. \tag{2}$$

Here,

$$R_{ix}(\boldsymbol{u}) = (2\pi)^{-2k} \int_{T^k} e^{i\boldsymbol{u}'\lambda} f_{ix}(\lambda)\, d\lambda, \tag{3}$$

defines the inverse covariance function. The corresponding inverse correlations are defined by $r_{ix}(\boldsymbol{u}) = R_{ix}(\boldsymbol{u})/R_{ix}(\boldsymbol{0})$. Suppose that for a fixed $\boldsymbol{t} \in \mathbb{Z}^k$ $X_{\boldsymbol{t}}$ is unknown. As in Rozanov (1967),

$$\hat{X}_{\boldsymbol{t}} = -\sum_{\boldsymbol{j}\in\mathbb{Z}^k\setminus\{\boldsymbol{0}\}} r_{ix}(\boldsymbol{j}) X_{\boldsymbol{t}-\boldsymbol{j}}, \tag{4}$$

defines the linear least squares interpolator of $X_{\boldsymbol{t}}$ for all $\boldsymbol{t}$, and

$$\eta_{\boldsymbol{t}} = \sum_{\boldsymbol{j}\in\mathbb{Z}^k} r_{ix}(\boldsymbol{j}) X_{\boldsymbol{t}-\boldsymbol{j}}, \tag{5}$$

the corresponding interpolation error process. Note that under the additional assumption that X is Gaussian,

$$\hat{X}_{\boldsymbol{t}} = \mathrm{E}\left(X_{\boldsymbol{t}} | X_{\boldsymbol{t}-\boldsymbol{j}}, \quad \forall\, \boldsymbol{j} \in \mathbb{Z}^k \ \setminus\ \{\boldsymbol{0}\}\right),$$

defines the conditional expectation of $X_{\boldsymbol{t}}$, given the knowledge of all its neighbors. We also have $\mathrm{E}(\eta_{\boldsymbol{t}}\, X_{\boldsymbol{t}+\boldsymbol{j}}) = 0$ for all $\boldsymbol{j} \neq \boldsymbol{0}$, $E(\eta_{\boldsymbol{t}} X_{\boldsymbol{t}}) = \mathrm{E}(\eta_{\boldsymbol{t}}^2) = \sigma_\eta^2 = 1/R_{ix}(\boldsymbol{0})$ and $\mathrm{E}(\eta_{\boldsymbol{t}}\, \eta_{\boldsymbol{t}+\boldsymbol{j}}) = \sigma_\eta^2\, r_{ix}(\boldsymbol{j})$. We may also write

$$X_{\boldsymbol{t}} = \hat{X}_{\boldsymbol{t}} + \eta_{\boldsymbol{t}}, \tag{6}$$

where $\hat{X}_{\boldsymbol{t}}$ is uncorrelated with $\eta_{\boldsymbol{t}}$.

Note that Equation (5) provides a two-sided CAR representation of $X_{\boldsymbol{t}}$ with $\eta_{\boldsymbol{t}}$ as the error term. This representation may be inverted to give a two-sided conditional moving average (CMA) representation for $X_{\boldsymbol{t}}$ in terms of the interpolation error process, $\eta_{\boldsymbol{t}}$, as follows

$$X_{\boldsymbol{t}} = \sum_{\boldsymbol{u} \in \mathbb{Z}^k} r_x(\boldsymbol{u}) \eta_{\boldsymbol{t}-\boldsymbol{u}}. \tag{7}$$

The representations (5) and (7) parallel the one-sided autoregressive and moving average representations for a time series with the innovation process $\epsilon_{\boldsymbol{t}}$ as random error. However, whereas in time series, the problem is that of prediction of an unobserved future value from a knowledge of the past, in space the problem is that of explaining the observation at a given site in terms of its neighbors. These two representations also bring out a fundamental difference between analyses of time series and spatial data. A main purpose of fitting a model in time series is to reduce an observed correlated process to a purely random white noise error term. By contrast, the corresponding aim in the conditional approach to constructing a model for spatial data is to transform a correlated observed process to an error process, $\eta_{\boldsymbol{t}}$, which, while correlated itself, is uncorrelated with all the neighbors $X_{\boldsymbol{t}-\boldsymbol{j}},\ \boldsymbol{j} \neq \boldsymbol{0}$ of $X_{\boldsymbol{t}}$, except $X_{\boldsymbol{t}}$ itself.

We also have

$$f_x(\boldsymbol{\lambda}) = \frac{\sigma_\eta^2}{(2\pi)^k \ \sum_{\boldsymbol{u} \in \mathbb{Z}^k} r_{ix}(\boldsymbol{u})\ e^{-i\boldsymbol{u}'\boldsymbol{\lambda}}}, \tag{8}$$

$$f_\eta(\boldsymbol{\lambda}) = \frac{\sigma_\eta^2}{(2\pi)^k} \sum_{\boldsymbol{u} \in \mathbb{Z}^k} r_{ix}(\boldsymbol{u})\ e^{-i\boldsymbol{u}'\boldsymbol{\lambda}}. \tag{9}$$

The following variance decomposition readily follows from Equation (6)

$$\mathrm{Var}(X_{\boldsymbol{t}}) = \mathrm{Var}\left(\hat{X}_{\boldsymbol{t}}\right) + \mathrm{Var}(\eta_{\boldsymbol{t}}) = \{R_x(\boldsymbol{0}) R_{ix}(\boldsymbol{0})\} + R_{ix}(\boldsymbol{0})^{-1}.$$

The first term to the right of this equation gives the variance of $X_{\boldsymbol{t}}$ that may be explained from a knowledge of its neighbors and the second term the unexplained variance. As in standard regression analysis, the ratio

$$\mathcal{F} = \mathrm{Var}\left(\hat{X}_{\boldsymbol{t}}\right) / \mathrm{Var}(X_{\boldsymbol{t}}) = 1 - \{R_x(\boldsymbol{0}) R_{ix}(\boldsymbol{0})\}^{-1}, \tag{10}$$

provides an R^2 measure of the interpolability of the process and it can be used as an index of linear determinism to gain an understanding of how much of the variability of $X_{\boldsymbol{t}}$ can be explained from a knowledge of all its neighbors, $X_{\boldsymbol{t}-\boldsymbol{u}}, \boldsymbol{u} \neq \boldsymbol{0}$.

Equations (8) and (9) show that appropriate conditional finite parameter models for X may be constructed by in turn specifying suitable models for $f_x(\lambda)$. As is well known (Guyon, 1995), the CAR models are specified by requiring that the inverse correlations, $r_{ix}(\boldsymbol{u})$, vanish outside a finite neighborhood, S_p, of $X_{\boldsymbol{t}}$. The order of the neighborhood set is denoted by p, and is defined sequentially by the maximum distance between the origin and a point in S_p. Thus, for example, the first-order neighbors of a site are those four sites which are adjacent to it; and the second order neighbors are these plus the four diagonally adjacent sites.

The CAR process admits a linear representation as follows

$$X_{\boldsymbol{t}} + \sum_{\boldsymbol{j} \in S_p} \alpha_{\boldsymbol{j}} X_{\boldsymbol{t}-\boldsymbol{j}} = \eta_{\boldsymbol{t}},$$

in which $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}} = r_{ix}(\boldsymbol{j})$, $\eta_{\boldsymbol{t}}$ is the interpolation error process and

$$\mathrm{Var}(\eta_{\boldsymbol{t}}) = \sigma_\eta^2 = \{R_{ix}(\boldsymbol{0})\}^{-1}.$$

Its spectral density function may then be written as (Besag, 1974)

$$f_x(\lambda) = \frac{\sigma_\eta^2}{(2\pi)^k \left\{1 + 2\sum_{\boldsymbol{j}\in S_p} \alpha_{\boldsymbol{j}} \cos(\boldsymbol{j}'\lambda)\right\}}.$$

This CAR model has the important Markov property that even when a knowledge of all the neighbors of $X_{\boldsymbol{t}}$ is available, only those in a finite set of neighbors, S_p, are relevant for its interpolation, and in view of this property, it is also said to define a Gauss Markov random field, GMRF. Since its introduction by Besag (1974), the class of CAR models has witnessed much development (see, Rue & Held, 2005) and recognized to provide a flexible class of conditional models for spatial data. It is, however, also known to suffer from the following three limitations.

Limitation 1. As noted by Besag and Kooperberg (1995), a CAR does not generate significant correlations, except at the boundary of its parameter space. Consequently, the index $\mathcal{F}$ typically takes fairly small values (see, Bhansali & Ippoliti, 2005) and a CAR may fail to explain a substantial proportion of the variability of an observed dataset.

Limitation 2. Although the correlations of a CAR decay at an exponential rate, they can on occasions decrease very slowly (Besag, 1981).

Limitation 3. A CAR does not apply to situations in which the conditional dependence of $X_{\boldsymbol{t}}$ on its neighbors is not merely confined to a finite set, but it extends to a possibly infinite set of sites in such a way that the degree of this dependence decreases as a function of the distance to the neighbor. Although, as further discussed below, a GMRF may still be used in such situations for approximating the underlying conditional dependence structure, such a course of action need not necessarily be warranted or applicable in all situations; for example, it may not be parsimonious in its use of parameters, or, see Rue and Tjelmeland (2002), Fontanella et al. (2008), and Song et al. (2008), specialist techniques may need to be invoked.

The class of conditional CARMA models was introduced by Ippoliti et al. (2013) in response to these three limitations. It is defined by the following equation

$$\sum_{\boldsymbol{j}\in S_p\cup\boldsymbol{0}} \alpha_{\boldsymbol{j}} X_{\boldsymbol{t}-\boldsymbol{j}} = v_0 \sum_{\boldsymbol{j}\in S_q\cup\boldsymbol{0}} \beta_{\boldsymbol{j}} \eta_{\boldsymbol{t}-\boldsymbol{j}},$$

in which $\eta_{\boldsymbol{t}}$ is the interpolation error process of $X_{\boldsymbol{t}}$, $v_0 = 1/\sum_{\boldsymbol{j}\in S_q\cup\boldsymbol{0}} \beta_{\boldsymbol{j}} r_{ix}(\boldsymbol{j})$ is a proportionality constant, the $\alpha_{\boldsymbol{j}}$ and $\beta_{\boldsymbol{j}}$ are real coefficients such that if

$$A(\boldsymbol{z}) = \sum_{\boldsymbol{j}\in S_p\cup(\boldsymbol{0})} \alpha_{\boldsymbol{j}}\, \boldsymbol{z}^{\boldsymbol{j}}, \quad B(\boldsymbol{z}) = \sum_{\boldsymbol{j}\in S_q\cup(\boldsymbol{0})} \beta_{\boldsymbol{j}}\, \boldsymbol{z}^{\boldsymbol{j}}, \quad \boldsymbol{z}\in C^k, \quad \alpha_{\boldsymbol{0}} = \beta_{\boldsymbol{0}} = 1,$$

are two symmetric finite Laurent series (Ahlfors, 2021) defined on the k-dimensional complex plane, C^k, then $A(\boldsymbol{z})$ and $B(\boldsymbol{z})$ have no common factors, $A(\boldsymbol{z}) = A(\boldsymbol{z}^{-\boldsymbol{1}})$, $B(\boldsymbol{z}) = B(\boldsymbol{z}^{-\boldsymbol{1}})$, and for $|\boldsymbol{z}| = 1$, $A(\boldsymbol{z}) \neq \boldsymbol{0}$, $B(\boldsymbol{z}) \neq \boldsymbol{0}$. Note that the conditions on $A(\boldsymbol{z})$ and $B(\boldsymbol{z})$ imply that $\alpha_{\boldsymbol{j}} = \alpha_{-\boldsymbol{j}}$ and $\beta_{\boldsymbol{j}} = \beta_{-\boldsymbol{j}}$ for all $\boldsymbol{j} \neq \boldsymbol{0}$.

The spectral density function of $X_{\boldsymbol{t}}$ is given by

$$f_x(\lambda) = \frac{\sigma_\eta^2\, v_0}{(2\pi)^k} \frac{B(e^{-i\lambda})}{A(e^{-i\lambda})}, \tag{11}$$

and it may be viewed as proportional to a ratio of two CAR spectral densities functions. The correlations and inverse correlations satisfy the Yule-Walker equations

$$\sum_{\boldsymbol{j}\in\mathbb{Z}^k} \alpha_{\boldsymbol{j}} r_x(\boldsymbol{u}-\boldsymbol{j}) = 0, \quad \text{all } \boldsymbol{u} \notin S_p, \tag{12}$$

$$\sum_{\boldsymbol{j}\in\mathbb{Z}^k} \beta_{\boldsymbol{j}} r_{ix}(\boldsymbol{u}-\boldsymbol{j}) = 0, \quad \text{all } \boldsymbol{u} \notin S_q. \tag{13}$$

The CAR is a special case with $S_q = \emptyset$. However, if $S_q \neq \emptyset$, Limitation 3 with the CAR can be overcome as the $r_x(\boldsymbol{u})$ do not vanish in a finite neighborhood. Also, numerical results given in Ippoliti et al. (2013) demonstrate that for $S_q \neq \emptyset$ a CARMA can generate non trivial correlations and Limitations 1 and 2 of CAR may be mitigated.

The CMA model is also a special case with $S_p = \emptyset$. Now, $r_x(\boldsymbol{u}) = 0$, for all $\boldsymbol{u} \notin S_q$.

Ippoliti et al. (2013) show that a CARMA can arise under certain operations on a CAR. Thus, the correlations of a process obtained by summing two CARs admit the same structure as that of a CARMA. Similarly, smoothing or sampling a CAR gives rise to a CARMA correlation structure.

It may be noticed that CAR, CMA, and CARMA models postulate that the spectral density of $X_{\boldsymbol{t}}$ is a rational function of $\boldsymbol{z} = e^{-i\lambda}$. An alternative approach is suggested by Bloomfield (1973), who introduced the class of exponential models for a time series. We note also that an example of an exponential spectral density also occurs in the work of Baxter (1962). More recently, this class of models has attracted attention in the literature of LM time series; see for example, Beran (1993) and Moulines and Soulier (1999), among others. An important property of this class of models is that the logarithm of the spectral density admits a finite Fourier expansion,

$$\log f_x(\lambda) = \sum_{s=-m}^{m} c_s e^{-i\lambda},$$

where m denote the order of the model and the c_s are called cepstral correlations, see Bogert et al. (1963). For spatial data, Solo (1986) has considered a two-dimensional analogue of this specification and pointed out its advantages for modeling certain properties of spatial data such as separability and isotropy. For more recent interest in this class of models for spatial data see Kohli (2016).

More generally, the k-dimensional stationary random field, X, satisfying conditions given in Equation (1), follows a Conditional Exponential, CEXP, model of order $\boldsymbol{m} \geq 0$, if the logarithm of its spectral density function, $\log f_x(\lambda)$, admits a finite Fourier series expansion as follows

$$\log f_x(\lambda) = \sum_{\boldsymbol{s}=-\boldsymbol{m}}^{\boldsymbol{m}} c_{\boldsymbol{s}}\, e^{-i\boldsymbol{s}'\lambda}, \quad \text{for all } \lambda \in T^k. \tag{14}$$

It readily follows that, for each fixed $\boldsymbol{u} \in Z^k$, X_t admits a CAR representation, (5) and a CMA representation (7) in which the coefficients $r_{ix}(\boldsymbol{u})$ and $r_x(\boldsymbol{u})$ are some functions of the cepstral correlations, $c_{\boldsymbol{s}}$. However, neither admits the Yule-Walker equations (12) and (13) satisfied by a process following a CARMA model. In this sense, the class of CEXP models provides an alternative to the class of CARMA models for spatial data.

It may be observed that the conditional models described above are parametric in the sense that the spectral density function, $f_x(\lambda)$, of each model is a function of a finite number of parameters. The latter will be unknown for the observed dataset, $\boldsymbol{x}$, and further statistical inferences will be based on estimated parameters. An alternative "non-parametric" approach has attracted some attention lately. In this approach, the observed data are assumed to be a realization of a process admitting the CAR representation (5), and a GMRF of a suitable finite order is fitted for approximating the underlying stochastic structure of the observed process. This approach draws on the success of an autoregressive model fitting approach for time series, see Parzen (1969) and Berk (1974), but now applied to lattice spatial processes. A driving idea is that the risk of bias due to truncating this expansion at a suitable finite value could be off-set by the risk of an increase in the variance due to fitting a model of an undue higher order. A variety of different approaches have been suggested in the literature for determining a suitable GMRF to use. Thus, Besag and Kooperberg (1995) adopt the Kullback–Leibler information for assessing the extent to which a GMRF may approximate an arbitrary random

field. Rue and Tjelmeland (2002) show, however, that while this approach provides a good approximation to the correlations within the neighborhood, S_p, used in estimating the GMRF, it is poor at approximating correlations outside of this neighborhood. The last-referenced authors suggest the use of an alternative matched correlation (MC)approach for this purpose. By contrast, Cressie and Verzelen (2008) adopt a prediction point of view for finding an optimal GMRF. They suggest doing this by finding that GMRF which gives the best prediction of an individual component from a knowledge of other components. A conditional-mean least squares criterion is introduced for this purpose and its performance relative to the MCs approach is also investigated. Verzelen (2010) has introduced a model selection criterion based on this last approach, but in which a penalty term is added. This author gives a lower bound on the penalty term such that, for an observed dataset, a best trade-off is obtained between the aforementioned increased bias due to choosing a too small GMRF and increased variance due to choosing a too large GMRF. Little is currently known, however, about the behavior of this approach with an observed dataset. In addition, there has also been interest, see Lindgren et al. (2011), in approximating a continuous Matern random field by a GMRF. It should, nevertheless, be emphasized that since the spectral density, $f_x(\lambda)$, of a GMRF is continuous and bounded for all $\lambda \in T^k$, the stochastic structure of a GMRF differs from that of conditional LM models described in Section 4. By contrast, at an empirical level, the question of which approach to adopt requires further investigation and we consider this question in Section 8. It is also pertinent to note that Dutta and Mondal (2021) consider the usefulness of a lattice approximation for continuous fractional Gaussian fields.

3 | LM RANDOM FIELDS

An important mathematical property of the CARMA, CEXP and associated parametric models introduced in Section 2 is that their respective spectral density functions, $f_x(\lambda)$, belong to the class of real analytic functions. This result, in turn, implies that the correlations, $r_x(\boldsymbol{u})$, and inverse correlations, $r_{ix}(\boldsymbol{u})$, of each of this class of models converge to 0 at an exponential rate as $\boldsymbol{u} \to \infty$, meaning that there exist ψ_1 and ψ_2 with $0 < \psi_1, \psi_2 < 1$, and bounded constants M_1 and M_2 such that, for all $\boldsymbol{u}$,

$$|r_x(\boldsymbol{u})| \leq M_1 \psi_1^{\tilde{\boldsymbol{u}}}, \quad |r_{ix}(\boldsymbol{u})| \leq M_2 \psi_2^{\overline{\boldsymbol{u}}}, \tag{15}$$

where $\tilde{\boldsymbol{u}} = \sum_i |u_i|$ denote the L_1 norm of vector $\boldsymbol{u}$. It is now standard practice, see Beran et al. (2013), to refer to such processes as possessing short memory, and to distinguish them from stationary processes whose correlations decay at a subexponential rate, for example, polynomial or logarithmic. The latter class of processes, by contrast, are said to possess LM. Their distinguishing feature is that the correlations decay so slowly that their covariances are not summable, that is, it is possible to find an $\boldsymbol{m}^k \subset \mathbb{Z}^k$ such that

$$\sum_{\boldsymbol{u} \in \boldsymbol{m}^k} |R_x(\boldsymbol{u})| \to \infty, \quad \text{as} \quad \boldsymbol{m}^k \to \infty,$$

and, consequently, the first part of condition (1) fails to hold. Lavancier (2006) and Beran et al. (2013) have considered the formal mathematical question of how to make precise this last stated property. These authors, in particular, define a LM random field as a stationary process whose covariances, $R_x(\boldsymbol{u})$, decay at a polynomial rate as follows

$$R_x(\mathbf{u}) \sim \|\mathbf{u}\|^{-\tau} \mathcal{W}(\|\mathbf{u}\|)\, b\left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\right), \quad 0 < \tau < k. \tag{16}$$

Here, $\|\mathbf{u}\|$ denote the Euclidean norm of $\mathbf{u}$, $b(\mathbf{u}/\|\mathbf{u}\|)$ is a continuous function on the unit sphere in $\mathbb{Z}^k$ and $\mathcal{W}(\|\mathbf{u}\|)$ is slowly varying at infinity. It is common to assume that $\mathcal{W}(\cdot)$ is also bounded, and it can be replaced by a constant, C, say. This need not necessarily be the case, however, and $\mathcal{W}(\cdot)$ could, for example, be a logarithmic function and in which case the correlations decay at a mixture of polynomial and logarithmic rates. In the boundary case, in which, $\tau = 0$ and $\mathcal{W}(\cdot)$ is a logarithmic function, the correlations decay at a logarithmic case. An example of a random field where this last property holds is given in Bouchaud et al. (2014).

As for a time series, an alternative approach to defining a LM random field is based on the behavior of its spectral density function at the origin. In this approach, a random field, X, is said to possess LM, if its spectral density function,

$f_x(\lambda)$, is continuous everywhere except at the origin, where it admits a pole of the following form

$$f_x(\lambda) \sim \|\lambda\|^{\tau-k}\mathcal{W}\left(\frac{1}{\|\lambda\|}\right) b\left(\frac{\lambda}{\|\lambda\|}\right), \quad 0 < \tau < k, \quad \lambda \to \mathbf{0}. \tag{17}$$

These two definitions are known to be equivalent when $k = 1$, see, for example, Zygmund (2003). If, $k = 2$, $\mathcal{W}(\cdot)$ is bounded and $\tau = 2 - 2d$, $d \in (0, 1)$, then it readily follows that condition (16) specifies that as $\boldsymbol{u} \to \infty$, the covariances, $R_x(\boldsymbol{u})$ decrease to $\mathbf{0}$ at a polynomial rate

$$R_x(\boldsymbol{u}) \sim C\|\mathbf{u}\|^{2d-2}, \quad \text{as} \quad \boldsymbol{u} \to \infty$$

where C is a bounded constant. Also, now condition (17) specifies that $f_x(\lambda)$ diverges to infinity at the origin as follows

$$f_x(\lambda) \sim C\|\lambda\|^{-2d}, \quad \text{as} \quad \lambda \to \mathbf{0}.$$

Doukhan et al. (1996) have extended the classical Zygmund result to bivariate Fourier series by establishing that these two conditions are equivalent. For further discussion on the equivalence between the behavior of the spectral density at the origin and the asymptotic decay of the autocovariance function, see also Lighthill (1962), Beran (1994), Stein (1999), and Samorodnitsky (2006).

As a typical example of data admitting a long-memory behavior, Figure 1 shows land surface temperatures collected during daytime by the Terra instrument onboard the MODIS (Moderate Resolution Imaging Spectroradiometer) satellite on August 4, 2016 (Level-3 data). The data are on (256×256) regular lattice with 65,536 observations and represent a portion of bigger dataset analyzed in Heaton et al. (2019). The variation of the land surface temperature is one of the most important components of climatic and environmental changes, affecting organisms and ecosystems from local to global scales (Peng et al., 2018). The land surface temperature controls the effective radiating temperature of the Earth's surface. Its derivation is fundamentally influenced by the extreme heterogeneity of most natural land surface, such as atmospheric temperature and humidity variations, clouds and large aerosol particles. However, the recent advancements in remote sensing have made the acquisition of land surface temperatures data at high spatial resolution feasible.

To check the behavior of the spectral density function, Figure 2a represents the logarithm of the periodogram $I(\lambda_1, \lambda_2)$, here used as a sample version of $f_x(\lambda_1, \lambda_2)$, versus the log-frequencies. The plot shows that the points are scattered around a straight line and the negative slope suggests that the spectral density has a pole at the origin of the order $\|\lambda\|^{-1.35}$, giving a rough estimate of d equal to 0.68. However, this plot implicitly assumes that the LM parameter is the same in all directions while the data show a clear anisotropic structure. To this purpose, Figure 2b,c shows the values of the periodograms in log-log-coordinates obtained from the row and column series. In this case, the negative slopes suggest that an estimate

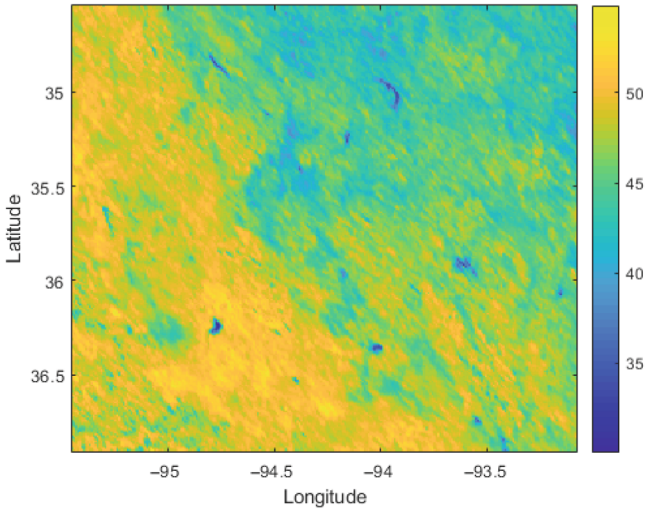

FIGURE 1 Land surface temperatures. The satellite data are observed over a (256×256) regular lattice.

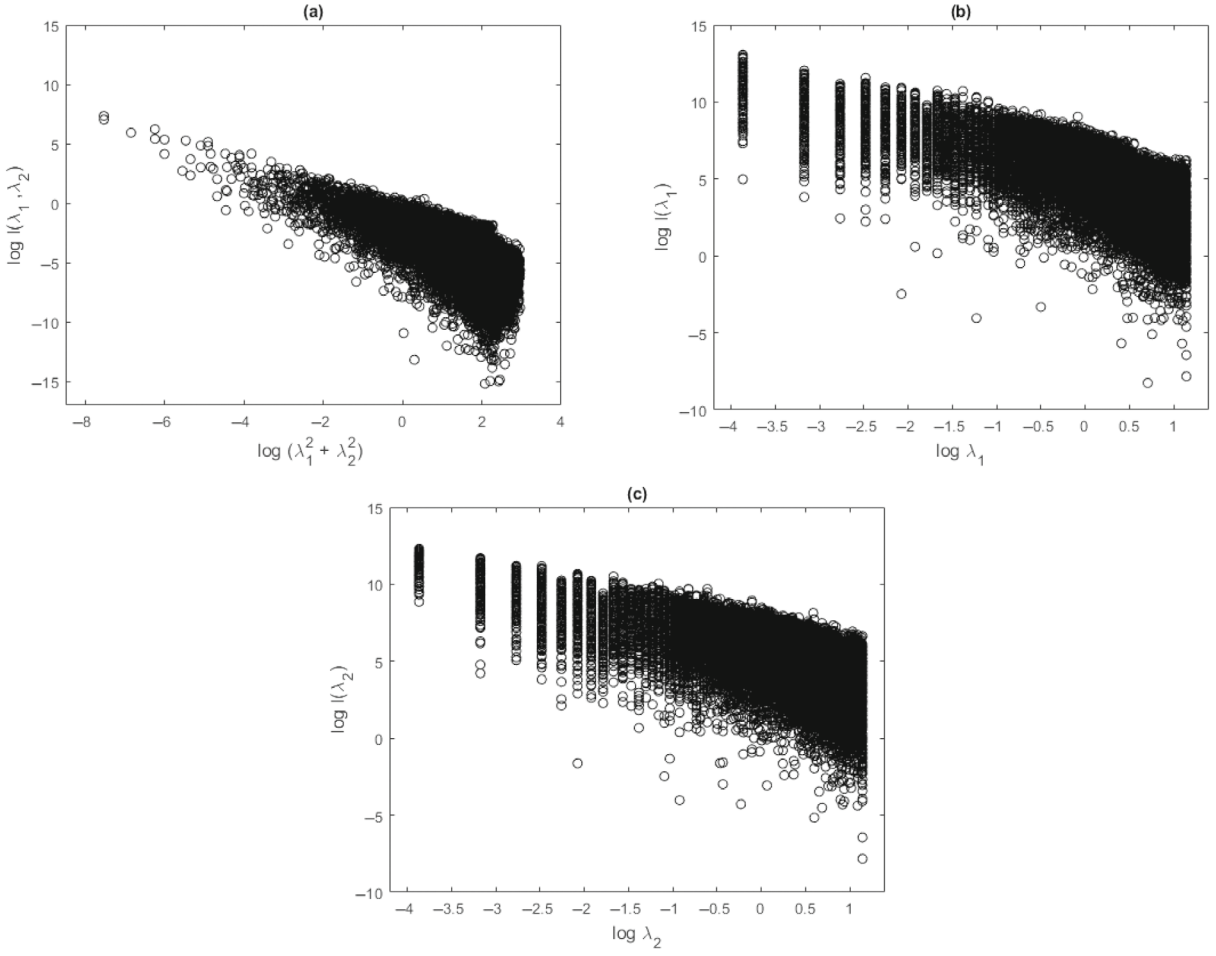

FIGURE 2 Log-periodogram versus the log-frequencies. Figure (a) refers to the isotropic case while figures (b) and (c) explore the existence of possible different long memory dependence in the North-South and East-West directions.

of d in the North-South and East-West directions are, respectively, equal to $\hat{d}_1 = 0.90$ and $\hat{d}_2 = 0.76$. In all cases the LM parameters appear quite large for the raw data suggesting a slower convergence of the correlation coefficients in Equation (7)—that is, the estimated LM parameters provide a measure of persistence or memory (Beran et al., 2013).

There exists a substantial literature on LM time series. The existing approaches to modeling LM for spatial data mainly follow a unilateral approach and are based on the time series literature. In Section 4 below, we consider how to construct Conditional LM, CLM, models for spatial data on a regular lattice. For simplicity, a univariate Gaussian process, indexed in $\boldsymbol{t} = (s, t) \in \mathbb{Z}^2$, will be considered henceforth.

4 | CLM MODELS

The inverse correlations, $r_{ix}(\boldsymbol{u})$ are readily seen to play an important role in the conditional approach to constructing stochastic models for X. It will be useful, therefore, to introduce the class of conditional LM models for a random field by first defining the generating function, $\Phi(\boldsymbol{z})$, of inverse correlations as follows

$$\Phi(\boldsymbol{z}) = \sum_{\boldsymbol{u} \in \mathbb{Z}^k} r_{ix}(\boldsymbol{u})\, \boldsymbol{z}^{\boldsymbol{u}}.$$

We assume below that $\Phi(\boldsymbol{z})$, $\boldsymbol{z} \in C^2$ may be factorized as follows

$$\Phi(z_1, z_2) = \Phi_S(z_1, z_2)\, \Phi_L(z_1, z_2), \tag{18}$$

where $\Phi_S(z_1, z_2) = \sum_{(u,v)\in\mathbb{Z}^2} \psi_{u,v}\, z_1^u z_2^v$ is a real analytic function such that, $\psi_{0,0} = 1$, $\Phi_S(z_1, z_2) > 0$ for $|z_1| = |z_2| = 1$. It readily follows that the coefficients, $\psi_{u,v}$, decrease to 0 at an exponential rate, that is, $|\psi_{u,v}| \leq \nu^{|u|+|v|}$ with $0 < \nu < 1$. Possible examples for $\Phi_S(z_1, z_2)$ are:

(a) $\Phi_S(z_1, z_2)$ is a polynomial as in GMRFs;
(b) $\Phi_S(z_1, z_2)$ is a rational function of finite symmetric Laurent series; and
(c) $\Phi_S(z_1, z_2)$ is an exponential function.

Furthermore, we also assume that $\Phi_L(z_1, z_2) = \sum_{(u,v)\in\mathbb{Z}^2} \theta_{u,v}\, z_1^u z_2^v$, with $\theta_{0,0} = 1$, and $\Phi_L(z_1, z_2) = 0$ for $|z_1| = 1$ and $|z_2| = 1$. Below, we consider different specifications for $\Phi(z_1, z_2)$ and show that these give rise to different types of conditional LM models on a regular lattice. To this purpose, let L be a shift operator on an index, such that $L_1 X_{s,t} = X_{s-1,t}$ and $L_2 X_{s,t} = X_{s,t-1}$. In addition, let $\overline{L} = (\overline{L}_1 + \overline{L}_2)$, $\overline{L}_1 = (L_1 + L_1^{-1})$ and $\overline{L}_2 = (L_2 + L_2^{-1})$.

Example 1. Conditional isotropic Fractionally Integrated (FI) noise process.

Koul et al. (2016) consider an isotropic FI random field, $Z_{s,t}$, defined as follows

$$\left(1 - \frac{1}{4}\overline{L}\right)^{\tau} Z_{s,t} = \epsilon_{s,t}, \quad 0 < \tau < 1/2, \tag{19}$$

where $\epsilon_{s,t}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ_ϵ^2, and τ denotes the order of fractional integration.

These authors show in the cited reference that the left-hand side of equation (19) may be written as

$$\sum_{(u,v)\in\mathbb{Z}^2} a_{u,v} Z_{s-u,t-v}, \tag{20}$$

where

$$a_{u,v} = \sum_{j=0}^{\infty} \psi_j(d) p_j(u, v), \tag{21}$$

$\psi_j(d) = \Gamma(j - d)/\Gamma(-d)\Gamma(j + 1)$, and $p_j(u, v)$ are j-step transition probabilities of the symmetric nearest-neighbor random walk $\{W_k,\ k = 0, 1, \dots\}$ on $\mathbb{Z}^2$, with transition probabilities $P(W_1 = (u, v)|W_0 = (0, 0)) = 1/4,\ |u| + |v| = 1$. In addition, the representation given in (20) above may be inverted to write $Z_{s,t}$ as a moving average

$$Z_{s,t} = \sum_{(u,v)\in\mathbb{Z}^2} b_{u,v}\ \epsilon_{s-u,t-v},$$

where

$$b_{u,v} = \sum_{j=0}^{\infty} \theta_j(-d) p_j(u, v), \tag{22}$$

and $\sum_{(u,v)\in\mathbb{Z}^2} b_{u,v}^2 < \infty$. An implication of the fact that the $b_{u,v}$ are square summable, is that the random field defined in equation (19) is stationary with mean 0, and spectral density function

$$f_z(\lambda_1, \lambda_2) = \frac{\sigma_\epsilon^2}{(2\pi)^2}\ 2^{2\tau}\{(1 - \cos(\lambda_1)) + (1 - \cos(\lambda_2))\}^{-2\tau}, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2. \tag{23}$$

It also readily follows from this Equation (23), that as $(\lambda_1^2 + \lambda_2^2) \to 0$,

$$f_z(\lambda_1, \lambda_2) \sim C\ (\lambda_1^2 + \lambda_2^2)^{-2\tau},$$

and the random field defined by Equation (19) admits LM. In addition, proposition 5.1 of Koul et al. (2016) shows also that as $\left(u^2+v^2\right)\to\infty$,

$$b_{u,v}\sim C\,(u^2+v^2)^{-1+\tau},\quad 0<\tau<1/2.$$

It may be noticed that the process defined in equation (19), and considered by Koul et al. (2016), provides a simultaneous representation of an isotropic random field admitting LM. We show below that it is also possible to adopt a conditional approach to define LM for this class of processes.

Suppose that $X_{s,t}$ is a completely symmetric GMRF of order 1

$$\left(1-\phi\overline{L}\right)\,X_{s,t}=\eta_{s,t},\tag{24}$$

where $\eta_{s,t}$ is the interpolation error. If $|\phi|<1/4$, $X_{s,t}$ is stationary, but for $\phi=1/4$, it is nonstationary. In the latter case, however, $X_{s,t}$, defines (see Besag & Kooperberg, 1995) an intrinsic CAR of order 1 and intrinsic order 0. We show below that even when $\phi=1/4$, it is possible to define a stationary LM process as follows.

Consider the equation

$$\left(1-\frac{1}{4}\overline{L}\right)^{d}\,X_{s,t}=\eta_{s,t},\quad 0<d<1,\tag{25}$$

where d denotes the LM parameter. The moving average and autoregressive representations of $X_{s,t}$ are as follows

$$\sum_{(u,v)\in\mathbb{Z}^2}a_{u,v}X_{s-u,t-v}=\eta_{s,t},\tag{26}$$

$$X_{s,t}=\sum_{(u,v)\in\mathbb{Z}^2}b_{u,v}\,\eta_{s-u,t-v},\tag{27}$$

where the $a_{u,v}$ and the $b_{u,v}$ are as defined above in Equations (21) and (22). A comparison of Equations (26) and (27) with (5) and (7) shows that the $b_{u,v}$ now define the correlations of $X_{s,t}$ and the $a_{u,v}$ define the corresponding inverse correlations. The spectral density function of $X_{s,t}$ is given by

$$\begin{aligned}f_x(\lambda_1,\lambda_2)&=\frac{\sigma_\eta^2}{(2\pi)^2}\left\{\sin^2\left(\frac{\lambda_1}{2}\right)+\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-d}\\&\sim C\,(\lambda_1^2+\lambda_2^2)^{-d},\quad\text{for }(\lambda_1,\lambda_2)\to 0.\end{aligned}\tag{28}$$

In a personal communication, D. Surgailis has confirmed that the decay rate for the $b_{u,v}$ given in Equation (22) above continues to hold for all $\tau\in(0,1)$. Moreover, the $a_{u,v}$ satisfy a similar decay rate, as follows

$$a_{u,v}\sim C\,(u^2+v^2)^{-1-\tau},\quad 0<\tau<1.$$

We note also that despite their apparent similarity, the models given in Equations (19) and (25) are radically different. Thus, the conditional model in (25) does not admit a simultaneous representation. This follows by recognizing that its spectral density function, f_x, say, for brevity, does not admit a factorization as $f_x=|Q|^{-2}$, where Q denotes a Laurent polynomial corresponding to the generating function of the coefficients of a simultaneous autoregressive process, and that this last condition is necessary for a stationary random field to admit a simultaneous representation. By contrast, the simultaneous model, (19), does admit a conditional representation as follows

$$\Phi_L(L_1,L_2)\,Z_{s,t}=\eta_{s,t},$$

where $\Phi_L(L_1,L_2)=\left(1-\phi_{1,0}\overline{L}-\phi_{2,0}(L_1^2+L_1^{-2}+L_2^2+L_2^{-2})-\phi_{1,1}\overline{L}_1\overline{L}_2\right)^{\tau}$, and with $\phi_{1,0}=2m\phi=2/5$, $\phi_{2,0}=-m\phi^2=-1/20$, $\phi_{1,1}=-2m\phi^2=-1/10$, $m=(\phi_{0,0})^{-1}=(1+4\phi^2)^{-1}=4/5$ and $\sigma_\eta^2=\sigma_\epsilon^2 m^{-d}$.

It is also relevant to note that if $\tau = 1$, $\Phi_L(L_1, L_2)$ gives the polynomial of a completely symmetric intrinsic CAR of order 1 and neighborhood order, $p = 3$. The latter process may thus be interpreted as the limit of a completely symmetric SAR with parameter $\phi = 1/4$. As shown in Rue and Held (2005), this intrinsic CAR process gives the parametrization of the thin-plate spline. □

Example 1 specifies that the LM parameter, d, is the same in all directions. In some applications, however, this specification may be too restrictive and where may be plausible to specify differing LM parameters in the "horizontal" and in the "vertical" dimensions. An example is given in a study of ground water flow and contaminant transport by Guo et al. (2009) who discuss random fields with a stronger memory parameter in the direction of the flow. In addition, Beran et al. (2009) consider the possibility of fitting a nonisotropic LM dependence. Example 2 below demonstrates how this aim may be achieved by specifying a pair of LM parameter, d_1 and d_2, instead of a just a single memory parameter, d, as is done in Example 1.

Example 2. *Conditional FI noise with separable fractional difference parameters.*

Consider a separable GMRF of order 2 of the following form

$$\left(1 - \phi_{1,0}\overline{L}_1\right)\left(1 - \phi_{0,1}\overline{L}_2\right) X_{s,t} = \eta_{s,t}. \tag{29}$$

As $X_{s,t}$ is separable, we may write the spectral density function of $X_{s,t}$ as

$$f_x(\lambda_1, \lambda_2) = \frac{\sigma_\eta^2}{4\pi^2} f_1(\lambda_1)\, f_2(\lambda_2),$$

where

$$f_1(\lambda_1) = (1 - 2\phi_{10}\cos(\lambda_1))^{-1} \quad \text{and} \quad f_2(\lambda_2) = (1 - 2\phi_{01}\cos(\lambda_2))^{-1}.$$

Hence, it readily follows that the process defined by Equation (29) is stationary if $|\phi_{01}| < 1/2$ and $|\phi_{10}| < 1/2$. If, however, $\phi_{10} = 1/2$ and $0 < \phi_{01} < 1/2$, then $f_1(\lambda_1)$ tends to infinity as λ_1 tends to 0 but $f_2(\lambda_2)$ remains bounded. Thus, now $f_x(\lambda_1, \lambda_2)$ is unbounded at the origin but only along the axis defined by λ_1. Conversely, if $\phi_{01} = 1/2$ and $0 < \phi_{10} < 1/2$, $f_2(\lambda_2)$ tends to infinity as λ_2 tends to 0 while $f_1(\lambda_1)$ remains bounded, and now $f_x(\lambda_1, \lambda_2)$ approaches infinity at the origin, but only along the axis defined by λ_2. In the situation when $\phi_{01} = \phi_{10} = 1/2$, $f_x(\lambda_1, \lambda_2)$ tends to infinity, as (λ_1, λ_2) tends to $(0, 0)$. We show below that in this last situation, it is still possible to define a stationary process admitting LM.

Suppose that $X_{s,t}$ follows a separable LM model given below

$$\left(1 - \frac{1}{2}\overline{L}_1\right)^{d_1}\left(1 - \frac{1}{2}\overline{L}_2\right)^{d_2} X_{s,t} = \eta_{s,t}, \quad 0 < d_1, \quad d_2 < 1/2, \tag{30}$$

where η_{st} denotes the interpolation error process of $X_{s,t}$ and d_1 and d_2 are a pair of memory parameters which may take different values along the horizontal and vertical axes. The spectral density function of $X_{s,t}$ is as follows

$$f_x(\lambda_1, \lambda_2) = \frac{\sigma_\eta^2}{4\pi^2}\left\{2\sin^2\left(\frac{\lambda_1}{2}\right)\right\}^{-d_1}\left\{2\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-d_2}. \tag{31}$$

We show next that the conditional process, $X_{s,t}$, defined in Equation (30), admits a representation as follows

$$(1 - L_1)^{\tau_1}(1 - L_2)^{\tau_2} Z_{s,t} = \epsilon_{s,t}. \tag{32}$$

This last process is also called a separable fractional noise process and it has previously attracted much attention in the literature: see Boissy et al. (2005), Shitan (2008), Ghodsi and Shitan (2009), Beran et al. (2009), and Guo et al. (2009), among others. Its spectral density function is as follows:

$$f_z(\lambda_1, \lambda_2) = \frac{\sigma_\epsilon^2}{4\pi^2}\left\{4\sin^2\left(\frac{\lambda_1}{2}\right)\right\}^{-\tau_1}\left\{4\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-\tau_2}. \tag{33}$$

A comparison of the right-hand sides of Equations (31) and (33) shows that apart from a multiplying constant of σ_η^2 occurring in (31) and $\sigma_\epsilon^2\ 2^{-(\tau_1+\tau_2)}$ occurring in (33), the spectral shapes of these two processes are the same, provided $\tau_1 = d_1$ and $\tau_2 = d_2$; moreover, if, in addition, $\sigma_\epsilon^2 = \sigma_\eta^2\ 2^{\tau_1+\tau_2}$ the two spectra coincide. The correlations and inverse correlations, $r_x(u)$, $ri_x(u)$ and $r_z(u)$ and $ri_z(u)$ of $X_{s,t}$ and $Z_{s,t}$ are the same and may be derived from the results given in Gradshteyn and Rhyzhik (1965) and Hosking (1981). We have

$$r_x(u,v) = \frac{\Gamma(d_1+u)\Gamma(1-d_1)\Gamma(d_2+v)\Gamma(1-d_2)}{\Gamma(d_1)\Gamma(1-d_1+u)\Gamma(d_2)\Gamma(1-d_2+v)},$$

and

$$r_{ix}(u,v) = \frac{\Gamma(-d_1+u)\Gamma(1+d_1)\Gamma(-d_2+v)\Gamma(1+d_2)}{\Gamma(-d_1)\Gamma(1+d_1+u)\Gamma(-d_2)\Gamma(1+d_2+v)}.$$

Also, on using the approximation, $\Gamma(a+x)/\Gamma(a+y) \sim a^{x-y}$, for $a \to \infty$, we also have, for $(u,v) \to \infty$,

$$r_x(u,v) \sim \frac{\Gamma(1-d_1)\Gamma(1-d_2)}{\Gamma(d_1)\Gamma(d_2)}|u|^{2d_1-1}|v|^{2d_2-1},$$

and

$$r_{ix}(u,v) \sim \frac{\Gamma(1+d_1)\Gamma(1+d_2)}{\Gamma(-d_1)\Gamma(-d_2)}|u|^{-1-2d_1}|v|^{-1-2d_2}.$$

□

It may be noticed that the interpolation error process, $\eta_{s,t}$, occurs on the right of Equations (25) and (30). An implication of this specification is that the short memory components of the processes defined in Examples 1 and 2 are not being explicitly modeled and, in the factorization (18) of $\Phi(z_1,z_2)$, $\Phi_S(z_1,z_2)$ is trivially set equal to 1. It is possible to specify a non-trivial $\Phi_S(z_1,z_2)$, by replacing $\eta_{s,t}$ by a short memory process, $Y_{s,t}$. Examples 4.3 and 4.4 given below illustrate this possibility by specifying that $Y_{s,t}$ follows a CAR and CMA model, respectively.

Example 3. *Diagonally symmetric CARFIMA*$(2,d_1,d_2,0)$ *model.*
Suppose that the LM component of the observed process, $X_{s,t}$, is the same as that of the process considered on the left hand side of equation (30) but its right-hand side is replaced by a second order CAR model with diagonally symmetric structure. The model is represented as follows

$$\left(1-\frac{1}{2}\overline{L}_1\right)^{d_1}\left(1-\frac{1}{2}\overline{L}_2\right)^{d_2} X_{s,t} = Y_{s,t}, \quad 0 < d_1, \quad d_2 < 1/2, \tag{34}$$

where $Y_{s,t}$ is a diagonally symmetric CAR(2),

$$A(L_1,L_2)Y_{s,t} = \eta_{s,t},$$

with $A(L_1,L_2) = 1-\alpha_{1,0}\overline{L}_1 - \alpha_{0,1}\overline{L}_2 - \alpha_{1,1}\overline{L}_1\overline{L}_2$ and $\alpha_{1,1} = -\alpha_{0,1}\alpha_{1,0}$. The spectral density function of $X_{s,t}$ is given by

$$f_x(\lambda_1,\lambda_2) = \frac{\sigma_\eta^2\ (1-\cos(\lambda_1))^{-d_1}\ (1-\cos(\lambda_2)^{-d_2}}{4\pi^2\left(1-2\alpha_{0,1}\cos(\lambda_1)-2\alpha_{1,0}\cos(\lambda_2)-4\alpha_{1,1}\cos(\lambda_1)\cos(\lambda_2)\right)}. \tag{35}$$

An important feature of this model is that X_{st} represents a conditional form of the following LM separable unilateral AR(1) process (Martin, 1986) which is also known as the Pickard process (Pickard, 1980)

$$(1-L_1)^{\tau_1}(1-L_2)^{\tau_2}\ Z_{s,t} = W_{s,t},$$

where

$$W_{s,t} = \left(1 - \vartheta_{1,0}L_1\right)^{-1}\left(1 - \vartheta_{0,1}L_2\right)^{-1}\epsilon_{s,t}$$
$$= \left(1 - \vartheta_{1,0}L_1 - \vartheta_{0,1}L_2 - \vartheta_{1,1}L_1\ L_2\right)^{-1}\epsilon_{s,t},$$

$\vartheta_{1,1} = -\vartheta_{0,1}\vartheta_{1,0}$ and, as in Example 2, $\epsilon_{s,t}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ_ϵ^2.

It is straightforward to show that the spectral density function of $Z_{s,t}$ is given by

$$\begin{aligned} f_z(\lambda_1,\lambda_2) &= \frac{\sigma_\epsilon^2}{4\pi^2}\left\{4\sin^2\left(\frac{\lambda_1}{2}\right)\right\}^{-\tau_1}\left\{4\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-\tau_2}\left|1 - \vartheta_{1,0}\ e^{i\lambda_1}\right|^{-2}\left|1 - \vartheta_{0,1}\ e^{i\lambda_2}\right|^{-2} \\ &= \frac{\sigma_\epsilon^2}{4\pi^2}\left(1 + \vartheta_{1,0}^2\right)^{-1}\left(1 + \vartheta_{0,1}^2\right)^{-1}\left\{4\sin^2\left(\frac{\lambda_1}{2}\right)\right\}^{-d_1}\left\{4\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-d_2} \\ &\quad\times\left\{1 - 2\frac{\vartheta_{1,0}}{1+\vartheta_{1,0}^2}\cos(\lambda_1) - 2\frac{\vartheta_{0,1}}{1+\vartheta_{0,1}^2}\cos(\lambda_2) - 4\frac{\vartheta_{1,0}\vartheta_{0,1}}{\left(1+\vartheta_{1,0}^2\right)\left(1+\vartheta_{0,1}^2\right)}\cos(\lambda_1)\cos(\lambda_2)\right\}^{-1} \\ &= \frac{\sigma_\epsilon^2\ C(1-\cos(\lambda_1))^{-\tau_1}\ (1-\cos(\lambda_2)^{-\tau_2}}{4\pi^2\left(1 - 2\rho_{1,0}\cos(\lambda_1) - 2\rho_{0,1}\cos(\lambda_2) - 4\rho_{1,1}\cos(\lambda_1)\cos(\lambda_2)\right)}. \end{aligned} \tag{36}$$

where $C = 2^{-(d_1+d_2)}(1+\vartheta_{1,0}^2)^{-1}(1+\vartheta_{0,1}^2)^{-1}$, $\rho_{1,0} = \vartheta_{1,0}/(1+\vartheta_{1,0}^2)$, $\rho_{0,1} = \vartheta_{0,1}/(1+\vartheta_{0,1}^2)$ and $\rho_{1,1} = -\rho_{0,1}\rho_{1,0}$. A comparison of the right-hand sides of Equations (35) and (36) shows that the two spectra coincide if $\tau_1 = d_1$ and $\tau_2 = d_2$, $\sigma_\epsilon^2 = \sigma_\eta^2\ C^{-1}$, $\alpha_{0,1} = \rho_{0,1}$ and $\alpha_{1,0} = \rho_{1,0}$.

Parameter estimation for the Pickard process has been considered in several different studies; some relevant references for the least squares estimation and for maximum likelihood estimation are Beran et al. (2009), Sethuraman and Basawa (1995) and Boissy et al. (2005). In addition, Ghodsi and Shitan (2009) have studied the normalized periodogram of the process with the aim of paralleling the work of Hurvich and Beltrao (1993) for a time series. □

Equations (8) and (9) show that the observed process, $X_{s,t}$, and its interpolation error process, $\eta_{s,t}$, are in an inverse relationship with each other in that the spectral density function of the former is proportional to the reciprocal of the latter, and vice versa. A suitably scaled version of the interpolation error process is accordingly also called the inverse process of $X_{s,t}$, see Battaglia (1983). Moreover, the inverse process of a standard autoregressive model for a time series is also known to follow a standard moving average process (see Bhansali & Ippoliti, 2005) and this property is at the heart of a model selection procedure for moving average models suggested in Bhansali (1983). In the following example, this property is also utilized by specifying that the short memory process, $Y_{s,t}$, follows a CMA model which is also the model followed by the interpolation error process of the short memory process, $Y_{s,t}$, considered in Example 4.

Example 4. *The diagonally symmetric CARFIMA*$(0, d_1, d_2, 2)$ *model*
Suppose that $X_{s,t}$ satisfies the same LM model as that specified by Equation (34) in Example 4, but $Y_{s,t}$ follows a diagonally symmetric CMA(2) model. The model is expressed as follows

$$\left(1 - \frac{1}{2}\overline{L}_1\right)^{d_1}\left(1 - \frac{1}{2}\overline{L}_2\right)^{d_2} X_{s,t} = Y_{s,t}, \tag{37}$$

where with $B(L_1, L_2) = 1 + \beta_{1,0}\overline{L}_1 + \beta_{0,1}\overline{L}_2 + \beta_{1,1}\overline{L}_1\overline{L}_2$,

$$Y_{s,t} = B(L_1, L_2)\ \eta_{s,t}.$$

The spectral density function of $X_{s,t}$ is given by

$$f_x(\lambda_1,\lambda_2) = \frac{\sigma_\eta^2\ \left(1 + 2\beta_{0,1}\cos(\lambda_1) + 2\beta_{1,0}\cos(\lambda_2) + 4\beta_{1,1}\cos(\lambda_1)\cos(\lambda_2)\right)}{4\pi^2\ (1-\cos(\lambda_1))^{d_1}\ (1-\cos(\lambda_2)^{d_2}}. \tag{38}$$

It may be observed that, apart from a slight change of notation, the polynomial $B(L_1, L_2)$ defined above and the polynomial $A(L_1, L_2)$ occurring in Example 4 are the same. However, whereas the latter occurs in the denominator of the spectral density function of the process defined in Example, the former occurs in the numerator of Equation (38). This last result, nevertheless, accords with the inverse relationship between the class of CMA processes considered in this example and the CAR processes considered in Example 4. Let

$$Z_{s,t} = (1 - L_1)^{-\tau_1}(1 - L_2)^{-\tau_2}\ W_{s,t}, \tag{39}$$

be a unilateral MA(1) process, where

$$W_{s,t} = \left(1 + \varphi_{1,0}L_1\right)\left(1 + \varphi_{0,1}L_2\right)\epsilon_{s,t},$$

and $\epsilon_{s,t}$ is a sequence of uncorrelated random variables, each with mean 0 and variance σ_ϵ^2. The spectral density of $Z_{s,t}$ is given by

$$f_z(\lambda_1, \lambda_2) = \frac{\sigma_\epsilon^2\ C\ \left(1 + 2\rho_{1,0}\cos(\lambda_1) + 2\rho_{0,1}\cos(\lambda_2) + 4\rho_{1,1}\cos(\lambda_1)\cos(\lambda_2)\right)}{4\pi^2\ (1 - \cos(\lambda_1))^{\tau_1}\ (1 - \cos(\lambda_2)^{\tau_2}}, \tag{40}$$

where $C = 2^{-(\tau_1+\tau_2)}(1 + \vartheta_{1,0}^2)(1 + \vartheta_{0,1}^2)$, $\rho_{1,0} = \vartheta_{1,0}/(1 + \vartheta_{1,0}^2)$, $\rho_{0,1} = \vartheta_{0,1}/(1 + \vartheta_{0,1}^2)$ and $\rho_{1,1} = \rho_{0,1}\rho_{1,0}$. On comparing equations (38) and (40), it readily follows that the two spectra are identical if $\tau_1 = d_1$, and $\tau_2 = d_2$, $\sigma_\epsilon^2 = \sigma_\eta^2\ C^{-1}$, $\beta_{0,1} = \rho_{0,1}$, $\beta_{1,0} = \rho_{1,0}$ and $\beta_{1,1} = \rho_{1,1}$.

The LM property of the process $Z_{s,t}$ follow by noting that

$$Z_{s,t} = \left(1 + \varphi_{1,0}L_1\right)\left(1 + \varphi_{0,1}L_2\right)\ \xi_{s,t},$$

where

$$\xi_{s,t} = (1 - L_1)^{-d_1}(1 - L_2)^{-d_2}\ \epsilon_{s,t},$$

is a separable fractional difference noise process considered previously in Example 2. This process is known to admit LM with spectral density function given on the right-hand side of Equation (33). The covariance function, $R_z(u, v)$, of $Z_{s,t}$ is related to the corresponding covariance function, $R_\xi(u, v)$, say, of $\xi_{s,t}$, as follow

$$R_z(u) = (1 + \varphi_{1,0}^2)\ R_\xi(u) + 2\varphi_{1,0}\left(R_\xi(u - 1) + R_\xi(u + 1)\right),$$

and

$$R_z(v) = (1 + \varphi_{0,1}^2)\ R_\xi(v) + 2\varphi_{0,1}\left(R_\xi(v - 1) + R_\xi(v + 1)\right).$$

Hence, it follows that as $(u, v) \to \infty$,

$$R_z(u, v) \sim C\ u^{-2\tau_1+1}v^{-2\tau_2+1},$$

where C denotes a bounded constant.

Moreover, since $Z_{s,t}$ and $X_{s,t}$ are equivalent, the latter also admits LM. The LM property of $Z_{s,t}$ also follows more readily by observing that it is a finite linear combination of $\xi_{s,t}$. The latter, by Example 4, admits LM and hence so does $Z_{s,t}$. Moreover, since $X_{s,t}$ is equivalent to $Z_{s,t}$, it also admits LM. □

Example 2 considered earlier introduces a two-parameter LM model for specifying differing LM behavior in the "horizontal" and the "vertical" dimensions. An alternative approach for achieving this aim is suggested in Example 5 below in which a GMRF of order 1 with only a single memory parameter is specified, but differing autoregressive coefficients for the neighbors along the rows and the columns are allowed for.

Example 5. *FI CAR*(1) *model with differing row and column coefficients*

The model considered is

$$\left(1-\phi_{1,0}\overline{L}_1-\phi_{0,1}\overline{L}_2\right)^d X_{s,t}=\eta_{s,t}. \tag{41}$$

where for some $\mu\in(-1/4,1/4)$, $\phi_{1,0}=1/4+\mu$, $\phi_{0,1}=1/4-\mu$ and $\eta_{s,t}$ denotes the interpolation error process of $X_{s,t}$. Following example 2 in Koul et al. (2016), $X_{s,t}$ admits a moving average representation as follows

$$X_{s,t}=\sum_{j=0}^{\infty}\psi_j(-d)\sum_{k=0}^{j}\binom{j}{k}\phi_{01}^{k}\phi_{10}^{j-k}\,\eta_{s-k,t-j-k}.$$

Its spectral density function is given by

$$f_x(\lambda_1,\lambda_2)=\frac{\sigma_\eta^2\,2^{-d}}{(2\pi)^2}\left\{\left(\frac{1}{2}+2\mu\right)\sin^2\left(\frac{\lambda_1}{2}\right)+\left(\frac{1}{2}-2\mu\right)\sin^2\left(\frac{\lambda_2}{2}\right)\right\}^{-d}. \tag{42}$$

Hence, we have, as $(\lambda_1,\lambda_2)\to 0$

$$f_x(\lambda_1,\lambda_2)\sim C\left\{\left(\frac{1}{2}+2\mu\right)\lambda_1^2+\left(\frac{1}{2}-2\mu\right)\lambda_2^2\right\}^{-d}. \tag{43}$$

An implication of this last result is that the memory parameter for the unequal coefficient model considered here is the same as that for the completely symmetric model (25) considered in Example 1. Nevertheless, the spectral density functions of the two models differ. In particular, the multiplying factors, $\left(\frac{1}{2}+2\mu\right)$ and $\left(\frac{1}{2}-2\mu\right)$ occurring in front of the two sine-squared terms on the right of Equation (42) are not present in Equation (28). These factors enable differing weights to be given to the two sine-squared terms. Thus, for $0<\mu<1/4$, the first term to the right of (42) is given greater weight than the second term. Conversely, for $-1/4<\mu<0$, the second term is given a greater weight than the first. At an empirical level, the extent to which the presence of μ influences the LM behavior of the model considered in this example and that considered in Example 1 requires further investigation. □

As demonstrated in Examples 1–5, the conditional approach to specifying LM random fields enables a more general class of models to be deployed than has hitherto been possible by either the simultaneous or the unilateral approaches; for example, the use of fractional differencing which is used in the unilateral approach may be avoided and yet an equivalent model could be specified. Moreover, these examples also suggest the following generic approach to specifying the factor $\Phi_L(z_1,z_2)$ occurring in Equation (18):

As in Equation (11), let S_p denote a finite neighborhood of $(0,0)$ and let

$$A(z_1,z_2)=\sum_{(s,t)\in S_p\cup(0,0)}\alpha_{s,t}\,z_1^s z_2^t,\ (z_1,z_2)\in C^2,$$

denote a symmetric finite Laurent series defined on the two-dimensional complex plane, C^2, such that $A(z_1,z_2)=A(z_1^{-1},z_2^{-1})$. A necessary condition for $A(z_1,z_2)$ to define a stationary Gauss-Markov random field is that $A(z_1,z_2)>0$ on the two-dimensional unit circle, $|z_1|=|z_2|=1$. Let, $\overline{\alpha}_{s,t}$ denote the values of $\alpha_{s,t}$ for which $A(z_1,z_2)=0$, $|z_1|=|z_2|=1$ vanishes on the unit circle. We may then set, with $0<d<1$ denoting the LM parameter,

$$\Phi_L(z_1,z_2)=\left(\sum_{(s,t)\in S_p\cup(0,0)}\overline{\alpha}_{s,t}\,z_1^s z_2^t\right)^d.$$

Also, $\Phi_S(z_1,z_2)$ could be specified as a real analytic function following one of the standard conditional models discussed in Section 2.

5 | ESTIMATION OF CONDITIONAL LM RANDOM FIELDS

There has been much development on the question of how to estimate the memory parameter, d, of a fractionally-differenced stationary time series with a review of the existing methods may be found in Palma (2007) and Giraitis and Robinson (2003), among others. As discussed in these references, the existing methods may be grouped under the following three broad approaches: (1) Parametric; (2) Nonparametric; (3) Semi-parametric. In the parametric approach, the observed time series is typically assumed to be a realization of a LM process, X_t, say such that its dth fractional difference, Y_t, say, follows an ARMA or an Exponential model of known order. As discussed in the pioneering work of Fox and Taqqu (1986), Dahlhaus (1989), and Giraitis and Surgailis (1990), a maximum likelihood, or an equivalent procedure, is then used for estimating the parameters of the model specified for Y_t together with the fractional-difference parameter.

In the more realistic situation in which the order of the process is unknown, the use of Akaike's information criterion (AIC), or a related model selection criterion, may also be considered for determining the order to adopt; see, for example, Beran et al. (1998). The nonparametric approach, by contrast, only specifies the behavior of the theoretical spectral density function near the origin. The memory parameter is then estimated, see Geweke and Porter-Hudak (1983), by regressing the logarithm of the periodogram of the observed time series on the logarithm of the frequency. Early work on the semi-parametric approaches combined the parametric and nonparametric approaches by specifying only the behavior of the spectral density near the origin. A truncated form of the Whittle likelihood, which is defined only for a frequency band close to the origin, is then employed for estimating the memory parameter; see, for example, Robinson (1995a).

More recent work, by contrast, specifies that the dth fractionally-differenced time series, Y_t, admits an infinite exponential, or autoregressive representation. The memory parameter together with the parameters of the specified parametric model for Y_t, are then estimated by a regression, or a maximum likelihood, procedure, in which the finite order to adopt is determined by a model selection criterion; however, the specified model for Y_t, is assumed to be of an infinite order and the sampling properties of the estimates are studied by assuming that the fitted order approaches infinity simultaneously, but sufficiently slowly, with the length of the observed time series.

We explain below how the three approaches described above may be generalized for estimating the parameters of a conditional LM random field.

5.1 | Exact maximum likelihood

Assume that X_t follows either a CARFIMA(p, d, q) model or a CARFIMA(p, d_1, d_2, q). Also, let $\mathcal{L} \subset \mathbb{Z}^2$ be an $(n_1 \times n_2)$ lattice with n sites and let $x_{s,t}$ be the process observed at sites $s = 1, \ldots, n_1; t = 1, \ldots, n_2$. Each row of the lattice is mapped lexicographically into $(N_2 \times 1)$ vectors $\boldsymbol{x}_s = \left[x_{s,1}, x_{s,2}, \ldots, x_{s,n_2}\right)\Big]'$ that are stacked one on the top of the other to form the $(n \times 1)$ vector $\boldsymbol{x} = \left[\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_{n_1}'\right]'$. Then, the Gaussian log likelihood function of X is given by

$$\log\ L(\boldsymbol{\omega}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{R}_x| - \frac{1}{2}\mathbf{x}'\mathbf{R}_x^{-1}\mathbf{x}, \tag{44}$$

where $\mathbf{R}_x$ is the $(n \times n)$ covariance matrix of X dependent on the unknown model parameters collected in the vector $\boldsymbol{\omega}$.

Under suitable regularity conditions on $f_x(\lambda_1, \lambda_2)$—see, for example, Beran et al. (2013) and Palma (2007), the ML estimator of $\boldsymbol{\omega}$ is consistent and asymptotically Gaussian, with asymptotic variance given by the inverse of the Fisher information matrix $\mathbf{J}(\boldsymbol{\omega})$, with (i, j)th element given by $\mathbf{J}_{i,j}(\boldsymbol{\omega}) = \frac{1}{2}tr\left(\mathbf{R}_x^{-1}\mathbf{R}_x^{(i)}\mathbf{R}_x^{-1}\mathbf{R}_x^{(j)}\right)$, where $tr(\cdot)$ is the trace operator and $\mathbf{R}_x^{(i)} = \partial\mathbf{R}_x/\partial\omega_i$. Note that closed form expressions of the information matrix are in general difficult to obtain and, hence, $\mathbf{J}(\boldsymbol{\omega})$ is often calculated numerically.

A main difficulty in implementing MLE is that, in practice, the datasets may be extremely large and may have missing values. This makes likelihood inference impracticable, since exact Cholesky decompositions require $O(n^3)$ operations. One possibility to overcome the problem is to take advantage of using separable covariance functions for which the dimension of the covariance matrices to be inverted is reduced dramatically. For a separable fractional autoregressive specification of the model, the asymptotic normality of the ML estimator was proved by Sethuraman and Basawa (1995). Another possibility is to view the observed data as a partial realization from a Gaussian random field on a periodic lattice. In particular, if toroidal boundary conditions are assumed, then $\mathbf{R}_x$ is block circulant with circulant blocks and can be easily diagonalized with a two dimensional discrete Fourier transform—see, for example, Dryden et al. (2002) for inference

on CAR models, Ippoliti et al. (2013) for spatial processes with rational spectral density functions or Stroud et al. (2017) for stationary Gaussian processes observed on a large lattice with missing values. This approach has a number of computational advantages since the eigenvalues, matrix-vector multiplications and quadratic forms can be computed efficiently in $O(n \log n)$ operations by exploiting the FFT. Thus ML estimation can be carried out efficiently by maximizing the log-likelihood over the valid parameter space, that is, subject to the covariance matrix being strictly positive definite. In general, this can be ensured by checking numerically all its eigenvalues are positive.

5.2 | Whittle likelihood method

When working with processes for which an exact likelihood is hard to compute or is not analytically tractable, Whittle estimation represents another useful approach for estimating the model parameters.

Assume that both the covariance matrix, $\mathbf{R}_x$, and the spectral density function, $f_x(\lambda_1, \lambda_2)$, are functions of an unknown vector $\boldsymbol{\omega}$ of model parameters. Then, the Whittle method evaluates the likelihood by means of the following approximations

$$\log |\mathbf{R}_x| \approx \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} \log\{4\pi^2 f_x(\lambda_1, \lambda_{2,j_2})\},$$

and

$$\mathbf{x}' \mathbf{R}_x^{-1} \mathbf{x} \approx \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} \frac{I_x(\lambda_{1,j_1}, \lambda_{2,j_2})}{f_x(\lambda_{1,j_1}, \lambda_{2,j_2})},$$

where $I_x(\lambda_{1,j_1}, \lambda_{2,j_2})$ is the periodogram

$$I_x(\lambda_{1,j_1}, \lambda_{2,j_2}) = \frac{1}{4\pi^2 n} \left| \sum_{s=1}^{n_1} \sum_{t=1}^{n_2} x_{s,t} \, e^{i(s\lambda_{1,j_1} + t\lambda_{2,j_2})} \right|^2,$$

with discrete Fourier frequencies $\lambda_{1,j_1} = 2\pi j_1 / n_1$ and $\lambda_{2,j_2} = 2\pi j_2 / n_2$, and $j_k = 1, \ldots, n_r$, for $r = 1, 2$. The negative Whittle likelihood can thus be written as

$$\log L(\boldsymbol{\omega}) \approx - \left[\sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \log \left\{ 4\pi^2 f_y(\lambda_{1,j_1}, \lambda_{2,j_2}) \right\} + \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \frac{I_y(\lambda_{1,j_1}, \lambda_{2,j_2})}{f_y(\lambda_{1,j_1}, \lambda_{2,j_2})} \right]. \tag{45}$$

The properties of Whittle estimation have been studied extensively. For example, Hannan (1973) proved the consistency and asymptotic normality of the Whittle estimator for the class of linear processes and his results were subsequently extended by Rice (1979). Then, Fox and Taqqu (1986) and Giraitis and Surgailis (1990); Giraitis and Taqqu (1999) studied its properties in the presence of long-range dependence, and showed that both for LM Gaussian processes and nonlinear processes the estimator is consistent but does not necessarily have a limiting multivariate normal distribution after standardization. More recently, considering a separable fractional autoregressive process, Boissy et al. (2005) proved the consistency of the Whittle estimator in the spatial context. Theoretical results for the multidimensional periodogram can be found, for example, in Porcu et al. (2009).

A difficulty in implementing both ML and Whittle estimation is that the order (p, q) is invariably unknown, and, for an observed spatial series, there may not even be a true value of this order. The question of model choice has been considered by different authors by relying on model selection criteria, such as AIC or Bayesian Information criterion (BIC). Schmidt and Tcherning (1995) and Crato and Ray (1996) analyzed, by Monte Carlo experiments, the performance of various model selection criteria for both ARMA and ARFIMA models. Beran et al. (1998) investigated model choice criteria for stationary and nonstationary, fractional and nonfractional autoregressive processes. These authors have shown that while the AIC criterion does not provide a consistent order selection procedure for this class of processes, the BIC is consistent for such processes. On the other hand, to the best of our knowledge, there are no comparable results on the

asymptotic sampling properties of information criteria in the spatial framework. However, the use of information criteria suggests further alternative strategies for estimating LM models as described in the next section.

5.3 | CARFI method

In this section we discuss an alternative approach to the estimation of d in which the short-memory process is postulated to follow an infinite CAR process and a CARFIMA$(p, d, 0)$ model is fitted to an observed spatial series, where p is such that $p \to \infty$ simultaneously but sufficiently slowly with n. This approach extends to the spatial context the *FAR* method introduced by Bhansali and Kokoszka (2001) and Bhansali et al. (2006) in a LM time series framework.

In common with the *FAR* method, our differenced process is postulated to follow an infinite order autoregressive model such that an estimate of d is constructed by fitting a hierarchical family of CAR models with increasing order $p = p(n)$, such that $p(n) \to \infty$, $p(n)/n \to 0$, as $n \to \infty$.

In the following, it is thus assumed that $X_{s,t}$ is a realization of a two-dimensional stochastic process defined by

$$\Phi_L(L_1, L_2) X_{s,t} = Y_{s,t},$$

where Y is a stationary short-memory process satisfying

$$Y_{s,t} = \sum_{(u,v)\in\mathbb{Z}^2} \alpha_{u,v}\, Y_{s-u,t-v} + \eta_{s,t}, \tag{46}$$

with a bounded spectral density at the origin and an exponential decaying autocorrelation function. For example, if $\Phi_L(L_1, L_2) = \left(1 - \phi\tilde{L}\right)^{\tau}$, the process X has spectral density

$$f_x(\lambda_1, \lambda_2) = \frac{\sigma_\eta^2}{(2\pi)^2}\, \Phi_L(e^{-i\lambda_1}, e^{-i\lambda_2})^{-1}\, A(e^{-i\lambda_1}, e^{-i\lambda_2})^{-1},$$

with $A(e^{-i\lambda_1}, e^{-i\lambda_2}) = \sum_{(u,v)\in\mathbb{Z}^2} \alpha_{u,v} e^{-i\lambda_1 u} e^{-i\lambda_2 v}$. For estimation purposes the summation is confined to a finite set S_p such that X can be thought of as a realization of a CARFIMA$(p, \tau, 0)$—or a CARFIMA$(p, \tau_1, \tau_2, 0)$—whose parameters can be estimated by ML or Whittle likelihood and the order determined by appealing to the AIC, or BIC, criterion. Under several assumptions, and in the framework of time series, Bhansali et al. (2006) have shown that the FAR estimate of the LM parameter is asymptotically normal and that that this estimator is asymptotically superior to the nonparametric and semiparametric described below. Unfortunately, there are no theoretical results about the asymptotic properties of this estimator in the framework of random fields.

Considering the likelihood procedures described above, one may argue that the knowledge of the complete structure of the short-memory process $Y_{s,t}$ might be a rather restrictive assumption for many practical applications. Little is known on the behavior of model diagnostic criteria in the presence of spatial LM and this is clearly undesirable in applied studies where misspecifications of the short-run dynamics (i.e., the CARMA part) may result in inconsistencies in the estimate of the parameter d (or d_1, d_2) which only relates to the long-run behavior of the series and can be the only parameter of interest for many applications. These considerations thus motivate the introduction of a number of methods which seek to estimate the LM parameter(s) under few prior assumptions concerning the spectral density of $X_{s,t}$ and, in particular, without specifying a finite parameter model for the short memory process $Y_{s,t}$. The estimators considered below have been developed in the frequency domain and have led to the specification of log-periodogram regressions.

5.4 | Geweke and Porter-Hudak method

A widely used procedure in time series analysis is known as the Geweke and Porter-Hudak (GPH) method. This estimator is due to Geweke and Porter-Hudak (1983), and it is based on the observation that if in (17), the function $\mathcal{W}$ is considered as a constant C, for all λ, $\log f_x(\lambda)$ follows a linear function of $\log \lambda$ in a neighborhood of the origin, and, hence, an estimate of τ may be found by regressing log-periodogram on $\log \lambda$. As a result, the GPH estimator belongs to the so-called class of narrowband estimators because $f_x(\lambda)$ is evaluated only in the neighborhood of zero frequency. The extension of GPH to

the spatial domain was considered by Shitan (2008) and Ghodsi and Shitan (2016) for the separable fractional processes of Equations (32) and (39), and by Wang (2009) for the isotropic model of Equation (25).

Suppose, for example, that X is a separable LM unilateral ARMA model considered by Shitan (2008) and Ghodsi and Shitan (2016), with spectral density function

$$f_x(\lambda_1,\lambda_2)=\left|2\sin\left(\frac{\lambda_1}{2}\right)\right|^{-2d_1}\left|2\sin\left(\frac{\lambda_2}{2}\right)\right|^{-2d_2}f_y(\lambda_1,\lambda_2),\quad 0<d_1,\quad d_2<1/2. \tag{47}$$

Then, by taking the log of the spectral density function, it follows that

$$\begin{aligned}\log\{f_x(\lambda_1,\lambda_2)\}&=\log\{f_y(0,0)\}-d_1\log\left\{4\sin^2(\lambda_1/2)\right\}\\&\quad-d_2\log\left\{4\sin^2(\lambda_2/2)\right\}+\log\left\{f_y(\lambda_1,\lambda_2)/f_y(0,0)\right\}.\end{aligned}$$

By adding the log of the periodogram, $I_y(\lambda_{1j_1},\lambda_{2j_2})$, to both sides and rearranging gives

$$\begin{aligned}\log\left\{I_x(\lambda_{1j_1},\lambda_{2j_2})\right\}&=\log\{f_y(0,0)\}-d_1\log\left\{4\sin^2\left(\lambda_{1j_1}/2\right)\right\}-d_2\log\left\{4\sin^2\left(\lambda_{2j_2}/2\right)\right\}\\&\quad+\log\left\{f_y(\lambda_{1j_1},\lambda_{2j_2})/f_y(0,0)\right\}+\log\left\{I_x(\lambda_{1j_1},\lambda_{2j_2})/f_x(\lambda_{1j_1},\lambda_{2j_2})\right\}.\end{aligned}$$

To motivate a log-periodogram regression, Geweke and Porter-Hudak (1983) argued that $\log\left\{I_x(\lambda_{1j_1},\lambda_{2j_2})\right\}-\log\left\{f_x(\lambda_{1j_1},\lambda_{2j_2})\right\}$ turns into an iid sequence and that, following Brillinger (1974),

$$\frac{I_x(\lambda_{1j_1},\lambda_{2j_2})}{f_x(\lambda_{1j_1},\lambda_{2j_2})}\sim\chi^2_2\quad\text{as}\quad\min(n_1,n_2)\to\infty,$$

where χ^2_2 represents a chi-square distribution with two degrees of freedom for which we have

$$\mathrm{E}\left[\log\left\{\frac{I_x(\lambda_{1j_1},\lambda_{2j_2})}{f_x(\lambda_{1j_1},\lambda_{2j_2})}\right\}\right]=\psi(1),$$

where $\psi(\cdot)$ is the digamma function and $\psi(1)=0.57722$ is the Eulero's constant. Accordingly, this suggests to consider the terms

$$\epsilon_{j_1j_2}=\log\left\{\frac{I_x(\lambda_{1j_1},\lambda_{2j_2})}{f_x(\lambda_{1j_1},\lambda_{2j_2})}\right\}+\psi(1),$$

as regression errors with zero mean and constant variance $\pi^2/6$ (Brillinger, 1974). Furthermore, since for frequencies λ_{1j_1} and λ_{2j_2}, in the neighborhood of zero, $f_y(\lambda_{1j_1},\lambda_{2j_2})\sim f_y(0,0)$, the term $f_y(\lambda_{1j_1},\lambda_{2j_2})/f_y(0,0)$ becomes negligible and the log-periodogram regression can be expressed as

$$\log\left\{I_x(\lambda_{1j_1},\lambda_{2j_2})\right\}=c-d_1\log\left\{4\sin^2\left(\lambda_{1j_1}/2\right)\right\}-d_2\log\left\{4\sin^2\left(\lambda_{2j_2}/2\right)\right\}+\epsilon_{j_1j_2},$$

where $c=\log\{f_x(0,0)\}+\psi(1)$, $j_1=1,\ldots,m_1$, $j_2=1,\ldots,m_2$ and m_1 and m_2 are bandwidth hyper-parameters selecting the smallest Fourier frequencies around the origin. In addition, since $\sin(\lambda_{rj_r}/2)\sim\lambda_{rj_r}/2$, for $j_r/n_r\to 0$ and $r=1,2$, the LM parameters could be estimated by considering the following regression

$$\log\{I_x(\lambda_{1j_1},\lambda_{2j_2})\}=c-\tau_1\log\lambda_{1j_1}-\tau_2\log\lambda_{2j_2}+\epsilon_{j_1j_2}.$$

where now $\tau_1=2d_1$ and $\tau_2=2d_2$. Estimation of the LM parameter d using a nonseparable fractional differential operator is also straightforward and leads to the following log-periodogram regression

$$\log\left\{I_x(\lambda_{1j_1},\lambda_{2j_2})\right\}=c-d\log\{\sin^2(\lambda_{1j_1}/2)+\sin^2(\lambda_{2j_2}/2)\}+\epsilon_{j_1j_2}.$$

The SEs of the model parameters can be obtained by following the classical theory on model regression. The intuitive explanation is that most of the residuals in the GPH regression are approximately uncorrelated, and this turns out to be enough to obtain a result analogous to an i.i.d. situation. In principle, results from time series analysis (Beran et al., 2013) directly apply to the separable case while results for the nonseparable case can be found in Wang (2009). We note further that only the lowest κ_1 and κ_2 harmonic frequencies are involved to fit the regression in GPH. An optimal choice of κ_r as positive integers which tend to infinity slower than n_r for $r = 1, 2$ was proposed, for example, by Ghodsi and Shitan (2016) with $\kappa_r = \sqrt{n_r}$ following Geweke and Porter-Hudak (1983) in time series. In contrast, Wang (2009) chose κ_r equal to $n_r^{4/5}$ as suggested by Hurvich et al. (1998). However, there exist some difficulties in applying the GPH estimator in that, in practical applications, the method exhibits high variability and it is not easy to determine an optimal band of frequencies over which the regression is performed.

5.5 | Fractional exponential method

As opposed to the GPH, which uses a limited amount of the spectral density at low frequencies, the fractional exponential (FEXP) method finds an estimate of the LM parameters from the use of the full spectrum and it thus represents a broad-band estimator. Since all the frequencies are used in estimation, the difficult question of how to determine an optimal band of frequencies as in the GPH is no longer applicable here.

In the framework of time series, the FEXP method was first discussed by Janacek (1982) and then by Moulines and Soulier (1999) and Hurvich and Brodsky (2001). In space, this semiparametric estimator was recently extended by Kohli (2016) for random fields observed on a regular lattice. In this case, the logarithm of the spectral density of the short-memory process is postulated to possess an infinite Fourier series expansion. Hence, if X is a LM process with spectral density function, say as in Equation (47), an estimator of the LM parameters and of the Fourier coefficients, $c_{\boldsymbol{s}}$, say, is found by first truncating the infinite Fourier series at some finite value as in Equation (14), and then by applying an ordinary linear least-squares procedure for the following log-periodogram regression

$$\begin{aligned}\log\{I_x(\lambda_{1,j_1}, \lambda_{2,j_2})\} \approx c_{0,0} - d_1 \log\left\{4\sin^2(\lambda_{1,j_1}/2)\right\} - d_2 \log\left\{4\sin^2(\lambda_{2,j_2}/2)\right\} \\ + 2\sum_{r=1}^{p_1} c_{r,0}\cos(\lambda_{1,j_1} r) + 2\sum_{l=1}^{p_2} c_{0,l}\cos(\lambda_{2,j_2} l) \\ + 2\sum_{r=1}^{p_1}\sum_{l=1}^{p_2} c_{r,l}^{(+)}\cos(\lambda_{1,j_1} r + \lambda_{2,j_2} l) + 2\sum_{r=1}^{p_1}\sum_{l=1}^{p_2} c_{r,l}^{(-)}\cos(\lambda_{1,j_1} r - \lambda_{2,j_2} l) + \epsilon_{j_1,j_2},\end{aligned}$$

where $\boldsymbol{m} = (p_1, p_2)$ is the order of the model, $c_{r,l} = c_{-r,-l}$ represent the two-dimensional cepstral coefficients of the random field and, as for GPH, the term ϵ_{j_1,j_2} is assumed to be an asymptotically independent and identically distributed sequence of random errors. If $\boldsymbol{m}$ tends to infinity, then one obtains a perfect approximation, provided that the estimates $c_{\boldsymbol{s}}$ converge fast enough to the true values. The latter is guaranteed by preventing that $\boldsymbol{m}$ grows too fast, since otherwise there would be too many parameters to estimate. In general, there are $K = (2p_1p_2 + p_1 + p_2 + 1)$ cepstral coefficients to be estimated. A challenge in the FEXP method is the choice of model order $\boldsymbol{m}$. Paralleling works in time series studies, Kohli (2016) suggests the use of global and local versions of Mallow's C criterion (Hurvich, 2001) and the corrected Akaike's information criterion (AICc) of Hurvich and Tsai (1989), though no theoretical justifications are given for the use of such criteria. As for the GPH, Kohli (2016) computes the standard errors of the model parameters by following the classical theory on model regression.

It is finally worth noting that the FEXP approach bears some similarity to the CARFI approach, in that both methods approximate the short memory process by truncating either an infinite Fourier series expansion or an infinite order autoregressive model. However, the CARFI approach is likelihood-based and it may be expected to be asymptotically more efficient than the regression-based approach.

5.6 | Other methods

Two of the early methods for estimating the LM parameter are graphical and are based on the use of the *rescaled range* (R/S) statistic and/or the variance-time plot (Beran et al., 2013). A main difficulty in using the graphical methods is

that, especially for moderate values of n, they typically have large biases. In another semiparametric approach, see for example, Künsch (1987), Robinson (1995b), and Frias et al. (2008), a precise parametric form for the spectral density of the random field $X_{s,t}$ is not specified and it is assumed to be of the form given in Equation (17) but with $\mathcal{W}$ considered as a continuous function of the LM parameter in a neighborhood of the origin. Thus, this approach combines the parametric and nonparametric methods discussed above and it is often called *local Whittle*. However, this method suffers from the same difficulties as described for the nonparametric approach in that the bandwidth of frequencies must be chosen in an arbitrary fashion.

6 | MODEL PARAMETER ESTIMATION: A SIMULATION STUDY

To illustrate the efficacy of the different estimators in estimating the LM parameter, we provide a comparison within a simulation study based on realizations from a wide range of CARFIMA(p, d_1, d_2, q) models, where p and q take values over the region $0 \leq p, q \leq 2$. The aim here is to compare the overall performance of different estimators over a wide range of models, rather than point out which estimators are best for certain models. In particular, considering the general representation

$$\left(1 - \frac{1}{2}\overline{L}_1\right)^{d_1} \left(1 - \frac{1}{2}\overline{L}_2\right)^{d_2} X_{s,t} = Y_{s,t},$$

we simulate realizations from the following different generating models

I. CARFIMA$(0, d_1, d_2, 0)$ where $Y_{s,t} = \eta_{s,t}$.
II. CARFIMA$(1, d_1, d_2, 0)$ where $\left(1 - 0.2\overline{L}_1 - 0.2\overline{L}_2\right) Y_{s,t} = \eta_{s,t}$.
III. CARFIMA$(2, d_1, d_2, 0)$ where $\left(1 - 0.2\overline{L}_1 - 0.2\overline{L}_2 + 0.08\overline{L}_1\overline{L}_2\right) Y_{s,t} = \eta_{s,t}$.
IV. CARFIMA$(0, d_1, d_2, 1)$ where $Y_{s,t} = \left(1 + 0.2\overline{L}_1 + 0.2\overline{L}_2\right)\ \eta_{s,t}$.
V. CARFIMA$(1, d_1, d_2, 1)$ where $\left(1 - 0.2\overline{L}_1 - 0.2\overline{L}_2\right) Y_{s,t} = \left(1 + 0.2\overline{L}_1 + 0.15\overline{L}_2\right)\ \eta_{s,t}$.
VI. CARFIMA$(0, d_1, d_2, 2)$ where $Y_{s,t} = \left(1 + 0.2\overline{L}_1 + 0.2\overline{L}_2 + 0.08\overline{L}_1\overline{L}_2\right)\ \eta_{s,t}$.
VII. CARFIMA$(2, d_1, d_2, 2)$ where

$$\left(1 - 0.2\overline{L}_1 - 0.2\overline{L}_2 + 0.08\overline{L}_1\overline{L}_2\right) Y_{s,t} = \left(1 + 0.2\overline{L}_1 + 0.1\overline{L}_2 + 0.05\overline{L}_1\overline{L}_2\right)\ \eta_{s,t}.$$

Note that all models have moderate AR and MA coefficients so as not to favor a priori any of the model fitting methods. Furthermore, we note that each model has $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 + 3$ parameters with $\mathcal{P}_1$ from S_p, $\mathcal{P}_2$ from S_q plus σ_η^2, d_1 and d_2. For each generating model, we set $d_1 = d_2 = 0.3$ and $\sigma_\eta^2 = 1$ and the following methods of LM parameters estimation are compared:

1. **GPH** The LM parameters d_1 and d_2 were estimated by least squares estimation on a general multiple regression model as defined in Section 5.4. Different bandwidths, $(\kappa_1, \kappa_2) = (n_1^{b_1}, n_2^{b_2})$ with $b_1 = b_2 = 2/5, 3/5, 4/5$ were used to fit the regression model.
2. **CARFIMA**$(1, d_1, d_2, 1)$ The parametric method estimates the LM parameters by Gaussian maximum likelihood and assumes that the simulated random field truly follows a fixed CARFIMA$(1, d_1, d_2, 1)$ process, whether or not the generated model was of this particular form.
3. **CARFIMA**$(2, d_1, d_2, 2)$ Same as above, but a CARFIMA$(2, d_1, d_2, 2)$ was fitted.
4. **CARFIMA AIC** In this case, we consider a CARFIMA(p, d_1, d_2, q) model with $p, q = 0, 1, 2$. The model for which the LM parameters is estimated by Gaussian maximum likelihood is selected by minimizing the AIC

$$\text{AIC}(p, q) = -2 \log L(\boldsymbol{\omega}) + 2\mathcal{P}.$$

5. **CARFIMA BIC** Same as above, but the LM parameters obtained by the maximum likelihood estimation were selected by minimizing the BIC.

$$\mathrm{BIC}(p,q) = -2\log L(\boldsymbol{\omega}) + (1+\ln\ n)\mathcal{P}.$$

6. **CARFI AIC** In this approach, a CARFI(p, d_1, d_2) model with $p = 0, 1, 2, 3, 4, 5$ is considered. The model for which the LM parameters were estimated by the maximum likelihood was selected by minimizing

$$AIC(p) = -2\log L(\boldsymbol{\omega}) + 2(\mathcal{P}_1 + 3).$$

7. **CARFI BIC** Here the procedure used is the same as above, except that the AIC is replaced by the BIC.

$$\mathrm{BIC}(p) = -2\log L(\boldsymbol{\omega}) + (1+\ln n)(\mathcal{P}_1 + 3).$$

Note that the choice of estimating CARFIMA($1, d_1, d_2, 1$) and CARFIMA($2, d_1, d_2, 2$) models is considered because they coincide with two true generating processes (see models V and VII above) and, accordingly, simulation results may provide information about possible effects of model selection on the estimation of d_1 and d_2. A further reason is that they correspond to under- or over-fitting the generated process and the results may again provide some information about how misspecification of the short-memory model structure influences the estimation of LM parameters.

Focusing on bias, SE, and the root mean square error (RMSE) for the parameters, the following tables provide, for each model and method, estimation results from 100 simulations on a (128×128) lattice. In general, parametric methods

TABLE 1 Bias, SD, and root of mean square error (RMSE) of GPH estimator of d_1 and d_2 for different bandwidths $(\kappa_1, \kappa_2) = (n_1^{b_1}, n_2^{b_2})$ in 100 simulations of various Gaussian models on a (128×128) lattice.

GPH							
		Bias		**SD**		**RMSE**	
	$(\boldsymbol{b_1}, \boldsymbol{b_2})$	$\hat{\boldsymbol{d}}_1$	$\hat{\boldsymbol{d}}_2$	$\hat{\boldsymbol{d}}_1$	$\hat{\boldsymbol{d}}_2$	$\hat{\boldsymbol{d}}_1$	$\hat{\boldsymbol{d}}_2$
I: CARFIMA($0, d_1, d_2, 0$)	(2/5, 2/5)	−0.051	−0.088	0.267	0.237	0.272	0.253
	(3/5, 3/5)	0.042	0.014	0.069	0.084	0.081	0.085
	(4/5, 4/5)	0.002	−0.001	0.027	0.028	**0.027**	**0.028**
II: CARFIMA($1, d_1, d_2, 0$)	(2/5, 2/5)	0.032	−0.016	0.235	0.234	0.238	0.234
	(3/5, 3/5)	0.160	0.153	0.080	0.068	**0.179**	**0.167**
	(4/5, 4/5)	0.294	0.291	0.028	0.025	0.295	0.292
III: CARFIMA($2, d_1, d_2, 0$)	(2/5, 2/5)	−0.049	−0.086	0.267	0.237	0.271	0.252
	(3/5, 3/5)	0.061	0.033	0.069	0.084	**0.092**	**0.090**
	(4/5, 4/5)	0.177	0.173	0.027	0.028	0.179	0.175
IV: CARFIMA($0, d_1, d_2, 1$)	(2/5, 2/5)	−0.048	−0.085	0.267	0.237	0.271	0.251
	(3/5, 3/5)	0.063	0.035	0.069	0.084	**0.094**	**0.091**
	(4/5, 4/5)	0.164	0.160	0.027	0.028	0.166	0.163
V: CARFIMA($1, d_1, d_2, 1$)	(2/5, 2/5)	0.027	−0.017	0.231	0.232	0.232	0.232
	(3/5, 3/5)	0.182	0.170	0.080	0.065	**0.199**	**0.182**
	(4/5, 4/5)	0.456	0.452	0.029	0.025	0.457	0.453
VI: CARFIMA($0, d_1, d_2, 2$)	(2/5, 2/5)	−0.046	−0.083	0.267	0.237	0.271	0.251
	(3/5, 3/5)	0.073	0.045	0.069	0.084	**0.101**	**0.095**
	(4/5, 4/5)	0.185	0.181	0.027	0.028	0.187	0.184
VII: CARFIMA($2, d_1, d_2, 2$)	(2/5, 2/5)	−0.013	−0.038	0.242	0.212	0.242	0.216
	(3/5, 3/5)	0.088	0.074	0.076	0.083	**0.116**	**0.111**
	(4/5, 4/5)	0.358	0.356	0.024	0.029	0.359	0.357

Note: Bold indicate the values of the bandwidths, (κ_1, κ_2), for which the RMSE is minimized.

TABLE 2 Estimation results for $\hat{d}_1$ and $\hat{d}_2$ provided by different methods in 100 simulations of CARFIMA(p, d_1, d_2, q) models I–VII on a (128×128) lattice.

		Generated model													
		I CARFIMA $(0, d_1, d_2, 0)$		**II CARFIMA $(1, d_1, d_2, 0)$**		**III CARFIMA $(2, d_1, d_2, 0)$**		**IV CARFIMA $(0, d_1, d_2, 1)$**		**V CARFIMA $(1, d_1, d_2, 1)$**		**VI CARFIMA $(0, d_1, d_2, 2)$**		**VII CARFIMA $(2, d_1, d_2, 2)$**	
Method		$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
GPH	(a)	0.042	0.014	0.160	0.153	0.061	0.033	0.063	0.035	0.182	0.170	0.073	0.045	0.088	0.074
	(b)	0.069	0.084	0.080	0.068	0.069	0.084	0.069	0.084	0.080	0.065	0.069	0.084	0.076	0.083
	(c)	0.081	0.085	0.179	0.167	0.092	0.090	0.094	0.091	0.199	0.182	0.101	0.095	0.116	0.111
CARFIMA$(1, d_1, d_2, 1)$	(a)	−0.002	−0.006	−0.008	−0.002	0.060	0.040	−0.001	−0.006	−0.005	−0.002	−0.027	−0.031	0.025	0.017
	(b)	0.022	0.023	0.028	0.026	0.081	0.085	0.022	0.022	0.029	0.027	0.058	0.053	0.058	0.042
	(c)	0.022	0.024	0.029	0.026	0.101	0.094	0.022	0.023	0.030	0.027	0.064	0.061	0.063	0.045
CARFIMA$(2, d_1, d_2, 2)$	(a)	−0.002	−0.006	−0.007	−0.003	−0.007	−0.015	−0.001	−0.005	−0.005	−0.002	0.000	−0.004	−0.002	−0.008
	(b)	0.021	0.020	0.028	0.027	0.028	0.028	0.018	0.019	0.029	0.028	0.020	0.021	0.037	0.038
	(c)	0.021	0.021	0.030	0.027	0.029	0.032	0.018	0.020	0.030	0.028	0.020	0.021	0.037	0.039
CARFIMA AIC	(a)	0.002	−0.002	−0.006	−0.001	−0.005	−0.014	0.001	−0.003	−0.005	−0.002	−0.000	−0.003	−0.003	−0.008
	(b)	0.018	0.016	0.027	0.027	0.029	0.029	0.018	0.020	0.030	0.027	0.018	0.020	0.037	0.039
	(c)	0.018	0.016	0.028	0.027	0.029	0.032	0.018	0.020	0.030	0.027	0.018	0.021	0.037	0.039
CARFIMA BIC	(a)	0.003	0.000	−0.005	−0.002	−0.003	−0.009	0.001	−0.002	−0.005	−0.002	−0.002	−0.005	−0.002	−0.011
	(b)	0.014	0.014	0.026	0.025	0.024	0.025	0.016	0.018	0.029	0.027	0.019	0.021	0.046	0.043
	(c)	0.014	0.014	0.027	0.025	0.024	0.026	0.016	0.018	0.030	0.027	0.019	0.022	0.046	0.045
CARFI AIC	(a)	0.002	−0.002	−0.006	−0.001	−0.005	−0.014	0.108	0.098	0.089	0.086	−0.026	−0.027	−0.096	−0.095
	(b)	0.018	0.015	0.026	0.027	0.028	0.028	0.106	0.118	0.132	0.131	0.059	0.067	0.054	0.062
	(c)	0.018	0.015	0.027	0.027	0.029	0.031	0.152	0.153	0.160	0.160	0.064	0.072	0.110	0.113
CARFI BIC	(a)	0.003	0.000	−0.005	−0.002	−0.003	−0.009	0.097	0.085	0.067	0.070	−0.047	−0.052	−0.093	−0.103
	(b)	0.014	0.014	0.026	0.025	0.024	0.025	0.108	0.117	0.105	0.109	0.028	0.031	0.048	0.049
	(c)	0.014	0.014	0.027	0.025	0.024	0.026	0.145	0.144	0.125	0.130	0.055	0.060	0.104	0.114

Notes: (a) Bias, (b) Standard deviation, (c) Root of mean square error.

Abbreviations: AIC, Akaike's information criterion; BIC, Bayesian Information criterion.

appear to perform better than GPH. In all simulated models, the bias and SEs obtained by parametric methods are much smaller than the bias and SEs obtained by GPH. Table 1 shows the results of GPH for different bandwidths and, as it can be noticed, the RMSE is minimized for $(\kappa_1, \kappa_2) = \left(n_1^{3/5}, n_2^{3/5}\right)$. Hence, to facilitate the comparison of the different estimators, only the results for these bandwidths are reported henceforth.

In particular, considering Table 2, we note that the fixed order CARFIMA$(1, d_1, d_2, 1)$ and CARFIMA$(2, d_1, d_2, 2)$ perform very well for models V and VII. However, the same thing does not happen for all other generating models, especially for CARFI models, for which both the variance and the bias are much larger. This result is due to the fact that the above fixed order models coincide with the generating models V and VII, respectively, and in all other cases a problem of over- or under-fitting occurs. For example, the effect of the fitting a CARFIMA$(2, d_1, d_2, 2)$ model to any of the models I-VI is given by an increase in the SEs. For model VII, instead, fitting a fixed CARFIMA$(1, d_1, d_2, 1)$ with a smaller number of parameters determines a clearly increase of the bias. These results suggest that, when a model with too many parameters is fitted, the additional parameters attempt to model the LM component, effectively altering the generated value of d_1 and d_2. At the same time, the variance in estimating the LM parameters increases because of the excess variability introduced by the estimation of the redundant parameters. It should be noted, however, that the situation here is slightly different from when a pure short memory is being over-fitted in which case the estimation of additional parameters increases the variance but does not unduly influence the bias in estimating the nonzero parameters. On the other hand, when a model with too few parameters is fitted, the short memory component is not adequately modeled and this introduces bias in the estimation of d_1 and d_2 because the spectral density of the generated process is being approximated by the spectral density of the underparametrized model. The variance in estimating d_1 and d_2 is, however, reduced because fewer parameters are estimated. Consider now the CARFI method and its relative behavior in comparison with the fitting of CARFIMA models, with order selected by the AIC or the BIC criterion. The method of fitting a fractional CAR model is seen to provide good results for all models, but especially for models I−III, where the generated model for Y is a finite CAR.

It should be noted, however, that the CARFIMA BIC and AIC methods have a smaller mean squared error for models for which we have $q > 0$. This finding is probably not surprising because if the selected model coincides with the generated model the resulting estimates of d_1 and d_2 are expected to be asymptotically efficient. It should be emphasized, nevertheless, that the full CARFIMA models were fitted only up to order 2 and the chance of selecting an incorrect model is quite small in our simulations. As regards the question of whether the AIC or BIC criterion should be used for implementing the CARFI method suggested in this paper, the simulation results appear to favor the latter, probably because the AIC criterion is known to frequently select an "overparametrized" model resulting in a large mean square error as explained above when discussing possible effects of overfitting the simulated models.

Finally, we note that the results reported here refer to a (128×128) lattice. As expected, for smaller lattices the bias, the SEs and the RMSE are larger in magnitude for all models, though their values appear to be consistent with the conclusions drawn here.

7 | LAND SURFACE TEMPERATURE DATA

Here, we continue the statistical analysis started in Section 3 to see if LM models can be useful to fit the data. The land surface temperature image, represented in Figure 1, has pixel values ranging from 30.01 to 55.41° C, with sample mean 48.02, and sample SD 2.53. In order to consider a more realistic scenario, missing data are considered to test the performance of the fitted model in terms of interpolation. In particular, we consider two designs where the observations are missing at blocks or at random. We thus suppose that the lattice can be partitioned as $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$, where $\mathcal{L}_2$ contains $u \ll n$ grid-points for which $X_{s,t}$ is unknown, and $\mathcal{L}_1$ contains $g = n - u$ sites for which $X_{s,t}$ is known. Assuming that the sites are reordered, so that $\mathcal{L}_2 = \{g+1, g+2, \ldots, n\}$, we write

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_o \\ \mathbf{X}_m \end{bmatrix},$$

where $\mathbf{X}_o$ is the $(g \times 1)$ vector of variables observable at the g grid locations and $\mathbf{X}_m$ is the $(u \times 1)$ vector of the remaining variables to be predicted. In our case study, we consider a percentage of about 12% for the missing data, such that we work with a training set of $g = 57{,}536$ observations and a test set of $u = 8000$ pixels. Figure 3 shows the patterns of the missing data for the two designs considered.

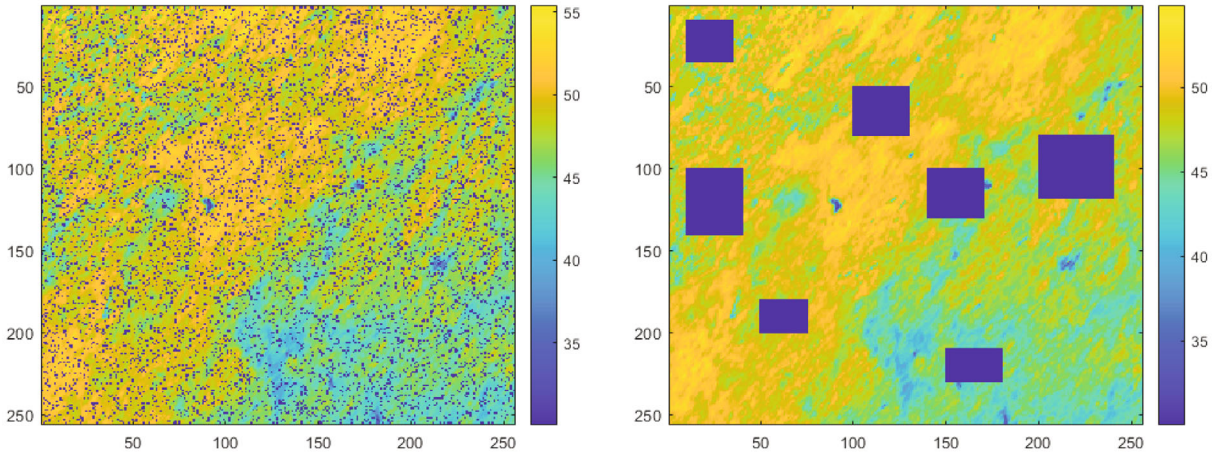

FIGURE 3 Left. Land surface temperature with data missing at random; Right. Land surface temperature with data missing at blocks. Missing data are represented as NaN in dark blue color.

To reconstruct the missing data, the linear interpolator is written in terms of the precision matrix

$$\mathbf{P}_x = \mathbf{R}_x^{-1} = \begin{bmatrix} \mathbf{P}_{oo} & \mathbf{P}_{om} \\ \mathbf{P}_{mo} & \mathbf{P}_{mm} \end{bmatrix},$$

and, as in Fontanella et al. (2008), it is represented by the following expectation

$$\tilde{\mathbf{X}} = \mathrm{E}[\mathbf{X}_m | \mathbf{X}_o = \mathbf{x}_o] = \boldsymbol{\mu}_m - \mathbf{P}_{mm}^{-1}\mathbf{P}_{mo}(\mathbf{x}_o - \boldsymbol{\mu}_o),$$

with conditional variance

$$\boldsymbol{\Xi} = \mathrm{Var}[\mathbf{X}_m | \mathbf{X}_o = \mathbf{x}_o] = \mathbf{P}_{mm}^{-1}.$$

Since in our case the ratio $u/n \approx$ is small, this interpolator is much quicker and more accurate than the usual kriging interpolator which, because of the required inversion of the $(g \times g)$ matrix $\mathbf{R}_{oo}$, appears difficult to apply in this study. For a CAR model, $\mathbf{P}_x$ is sparse and there exist a direct specification. In contrast, there is no simple direct specification of $\mathbf{P}_x$ for a CARFIMA model. However, following Equation (3), the elements of its first row can be obtained by the inverse FFT of $f_{ix}(\lambda_1, \lambda_2)$ and the full *circulant* matrix $\mathbf{P}_x$ reconstructed row-by-row since, as it is known, each row of the matrix is a right cyclic shift of the row above.

For modeling purposes, an exploratory analysis of the data suggest that the mean of the process can be represented by a parametrized linear deterministic spatial trend function. For simplicity, the index of linear determinism $\mathcal{F}$ and the empirical covariances are thus computed from the mean-corrected data at the available sites, using the Fourier transform on the full lattice with the missing values set to 0. The numbers of pairs are obtained from the two-dimensional Fourier transform of the incidence matrix of the observed sites.

Considering for example the case of missing at blocks, Figure 4 shows the periodogram of X smoothed by using the two-dimensional form of the Daniell window (Priestley, 1981) with length (5, 5). The periodogram shows a high peak at the origin with ridges along the North-South and East-West directions. The analysis of the low-lag sample correlations, represented in Table 3, also shows that these are high for neighboring sites and that they then drop away quite slowly. Furthermore, the correlations are slightly larger between columns than between rows; however, they do not suggest strong asymmetries. A similar pattern is found for the design in which the data are missing at random, though the correlations are smaller in this case—the low lag correlations are $R_x(0,1) = 0.749$, $R_x(1,0) = 0.795$, $R_x(1,1) = 0.669$, $R_x(0,2) = 0.621$ and $R_x(2,0) = 0.676$. Finally, the nonparametric estimates (Bhansali & Ippoliti, 2005) of $\mathcal{F}$ are 0.95 for the case of missing at blocks and 0.81 for the missing at random.

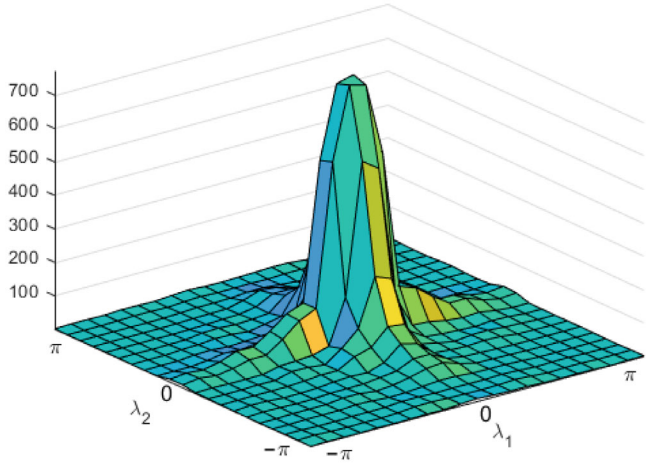

FIGURE 4 Land surface temperature with missings at blocks. Sample periodogram smoothed by using a two-dimensional Daniell window.

TABLE 3 Land surface temperature.

v									
4	0.469	0.494	0.520	0.546	0.570	0.576	0.579	0.577	0.566
3	0.502	0.533	0.564	0.594	0.626	0.639	0.646	0.637	0.609
2	0.538	0.576	0.617	0.662	0.711	0.740	0.738	0.694	0.635
1	0.577	0.626	0.684	0.762	0.855	0.883	0.808	0.710	0.632
0					1	0.903	0.771	0.676	0.612
	−4	−3	−2	−1	0	1	2	3	4
					u				

Note: Sample correlations at lags (u, v), $u = -4, \ldots, 4$, $v = 0, \ldots, 4$.

Several CARFIMA$(p, \tau, 0)$ and CARFIMA$(p, \tau_1, \tau_2, 0)$, with $0 \le p \le 2$, and several completely symmetric and homogeneous CAR(p) models, with $1 \le p \le 9$, were fitted to the data using the Whittle likelihood. As described in Section 5, this approach is not expensive computationally and, based on the periodogram computed as described above, also provides a naive solution to the problem of estimating the model parameters in the presence of missing data. Considering the design with missing at blocks, the comparison of the AIC and BIC values suggests that the following CARFIMA$(1, d, 0)$ model

$$\left(1 - \frac{2}{5}\overline{L} + \frac{1}{20}(L_1^2 + L_1^{-2} + L_2^2 + L_2^{-2}) + \frac{1}{10}\overline{L}_1\overline{L}_2\right)^d \left(1 - \alpha_{1,0}\overline{L}_1 - \alpha_{0,1}\overline{L}_2\right) Z_{s,t} = \eta_{s,t},$$

provides the best fit to the data, where, $Z_{s,t}$, is now the mean-corrected process, $\Phi_L(L_1, L_2)$ is the same fractional operator discussed in Example 1 and $\Phi_S(L_1, L_2)$ is the polynomial of a homogeneous CAR(1) process. The model has AIC = 1404.8 and BIC = 1405.2 (the value is divided by 100) with the following four estimated parameters (and standard errors): $\hat{d} = 0.434$ (0.003), $\hat{\alpha}_{0,1} = 0.355$ (0.003), $\hat{\alpha}_{1,0} = 0.092$ (0.002) and $\hat{\sigma}_\eta^2 = 0.342$(0.001). Among the completely symmetric CARs, the AIC (1416.2) and BIC (1416.9) suggest that a CAR(6) with 7-parameters is necessary. In contrast, a parsimonious first order CAR with 3-parameters is preferred among the class of homogeneous processes. In this case, however, the AIC (1495.4) and BIC (1495.8) are much larger and the short memory parameters α's, are very close to the stationary limit of the CAR. Similar comments hold also for the design in which the data are missing at random with the difference that a second order structure (i.e., $p = 2$) is necessary for the short-memory part of the LM model and for the class of homogeneous CAR models.

TABLE 4 Interpolation results from the estimated long memory models.

	Missings at blocks		Missings at random	
	$\mathbf{x}_m$	$\tilde{\mathbf{X}}_m$	$\mathbf{x}_m$	$\tilde{\mathbf{X}}_m$
Min	34.65	37.44	32.93	34.79
Max	55.41	52.43	54.19	54.38
$\mathcal{F}$	0.93	0.95	0.81	0.78
I_1	—	0.15	—	0.01
I_2	—	1.01	—	1.02
I_3	—	1.49	—	0.61
I_4	—	0.73	—	0.97

Note: For each design, $\mathbf{x}_m$ refers to the statistics computed on the true data and $\tilde{\mathbf{X}}_m$ predicted values.

To evaluate the performance of the selected model in terms of predictions we use the following statistics:

$$I_1 = \frac{1}{u}\sum_{i\in\mathcal{L}_2}\frac{Y_i - \tilde{Y}_i}{\hat{\xi}_i}, \quad I_2 = \sqrt{\frac{1}{u}\sum_{i\in\mathcal{L}_2}\frac{\left(Y_i - \tilde{Y}_i\right)^2}{\hat{\xi}_i^2}},$$
$$I_3 = \sqrt{\frac{1}{u}\sum_{i\in\mathcal{L}_2}\left(Y_i - \tilde{Y}_i\right)^2}, \quad I_4 = \mathrm{Corr}(\mathbf{Y}_m, \tilde{\mathbf{Y}}_m),$$

where $\hat{\xi}_i^2$ is the ith estimated model conditional variance, $\mathcal{L}_2$ is the set of sites containing the missing observations and u is its cardinality. When predictions are accurate, we expect that I_1 is close to 0 and that the weighted RMSE I_2 is close to 1. In addition, the RMSE I_3 should be small and I_4, being a measure of correlation between true and predicted data, should be close to 1. Interpolation results from the estimated models are shown in Table 4 which provides the values of the indices $I_1 - I_4$ and some summary statistics for the distribution of true and predicted values. In general, it seems there are some difficulties in predicting the lower temperatures. It also appears more difficult to predict the process when the data are missing at blocks. In this case, in fact, the index I_1 shows a small bias and the index I_3 is more than twice that of missing values at random. Finally, although the value of index I_4 is lower, it is still reasonably high. The values of of the $\mathcal{F}$ statistic provided by the estimated models are also reasonably close to the values estimated nonparametrically.

8 | APPROXIMATING CONDITIONAL LM PROCESSES BY GMRFS

In this section, we discuss the problem of approximating a LM Gaussian process. In particular, our investigation aims to empirically explore the extent to which an LM conditional random field can be approximated by a GMRF. This investigation is relevant when a dataset could be considered as a realization of an LM process, but a GMRF is used to fit the data without explicitly acknowledging the presence of LM. The task is also important in Bayesian spatial modeling, where a GMRF can be used to represent correlations directly estimated from observed data or consistent with prior knowledge (see, e.g., Rue & Tjelmeland, 2002).

In the following, it is thus assumed that the observed data represent a realization of a LM spatial process admitting the CAR representation of Equation (5) and that a GMRF of a suitable finite order can be fitted for approximating the underlying stochastic structure of the observed process.

One approach to approximating a Gaussian LM process is to fit a GMRF that closely approximates its correlation function. To improve the results obtained using the Kullback–Leibler divergence (Besag & Kooperberg, 1995), Rue and Tjelmeland (2002) proposed a MC approach that enhances the accuracy of the approximation beyond the neighborhood S_p. This approach aims to match the correlation structure of the LM process with that of the GMRF, resulting in a more accurate approximation. This method ensures that all sample (or theoretical) covariances of a LM process and the estimated theoretical covariances of the GMRF are as similar as possible in a weighted least-squares sense. In this case, the approximation is achieved by minimizing the function

$$\mathrm{MC}(\mathbf{R}_y, \mathbf{R}_x|\boldsymbol{\alpha}) = \sum_{(u,v)\in\mathcal{L}} \left[R_x(u,v) - R_y(u,v|\ \boldsymbol{\alpha})\right]^2 \rho(u,v), \tag{48}$$

where $\rho(u,v)$ is a user-defined weighting function of the lag-distance. A possible weighting function with $\rho(u,v)$ decreasing with $\boldsymbol{u} = (u,v)$ uses the reciprocal lag distance weights $\rho(\boldsymbol{u}) = 1/(||\boldsymbol{u}|| + 1)$.

Spectral densities and covariance functions are related through Fourier transforms. Although analyzing the variability of a process using the covariance function and spectral density may be considered equivalent, they offer different ways of studying the process, and spectral analysis may provide some advantages over analysis based on the covariance function (Stein, 1999). Hence, if we denote the spectral density function of the LM process as $f_x(\lambda_1, \lambda_2)$, and the spectral density function of the GMRF as $f_y(\lambda_1, \lambda_2)$, we can use the Chi-square distance, inspired by the MC approach and following Song et al. (2008), as a measure of discrepancy between f_x and f_y

$$\mathrm{MSDF}(f_x, f_y|\boldsymbol{\alpha}) = \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2} \left[\frac{\left(f_x(\lambda_{1,j_1}, \lambda_{2,j_2}) - f_y(\lambda_{1,j_1}, \lambda_{2,j_2}|\boldsymbol{\alpha})\right)^2}{f_x(\lambda_{1,j_1}, \lambda_{2,j_2})}\right]. \tag{49}$$

In the following, we discuss the efficacy of the GMRF approximation through a simulation study based on the MC criterion (48) and the Matched Spectral Density Function (MSDF) criterion (49). Additionally, we consider the effect of fitting a hierarchical family of GMRFs with increasing order, with p chosen using the AIC and BIC criteria. Similar to the CARFI method described in Section 5.3, we choose the value of p to obtain an adequate finite order approximation of the given LM process, rather than as an estimator of an unknown true order. The simulation is conducted by assuming that the Gaussian random field belongs to the general class of CARFIMA process whose parametrizations are reported in the following Table 5.

In particular, models M_1 and M_2 represent a fractionally differenced Gaussian noise (FGN) with two different values for the LM parameter. Models M_3 and M_4 represent two parametrizations of CARFIMA$(1, d_1)$ models with a CAR(1) structure for the short-memory component. Finally, model M_5 also represents a FGN but with a separable structure of the correlation function. Note that, apart from model M_5, all the models reported in Table 5 have a completely symmetric structure.

To evaluate the approximation of the correlation function of the first four models, we fit a sequence of completely symmetric GMRFs up to order 9 with a maximum of 10 parameters. For model M_5, instead, homogeneous CARs are fitted with $p = 1, \ldots, 5$. As discussed in Section 5, in order to derive computationally efficient algorithms, the two-dimensional FFT is used to perform the approximation task on a toroidal lattice. Note that to evaluate the MC and the MSDF criteria of Equations (48) and (49), we do not need to generate the data from the LM spatial process while this is necessary to evaluate the effect of penalizing the Whittle likelihood through the AIC and BIC criteria. Furthermore, for simplicity, we only plot the approximation provided by the MC and the MSDF criteria since, for GMRFs fitted by using the Whittle likelihood, an exact fit of the correlation function is only achieved within the neighborhood S_p (see Cressie, 1993; Dempster, 1972).

Regarding the nonseparable models M_1 and M_2, the first two panels of Figure 5 show that a good fit of the correlation functions of the LM-generating models can be achieved by fitting a CAR(9) using either the MC or the MSDF criteria. There exist small differences between the two criteria with a possible advantage of MSDF which also does not require the use of weights. The use of AIC and BIC suggest that selecting the correct generating model can be

TABLE 5 Parameter structure of the simulated long memory models.

Models	# Par	d	d_1	d_2	$\alpha_{1,0}$	$\alpha_{0,1}$	σ^2_η
M_1	2	0.35	—	—	—	—	1
M_2	2	0.95	—	—	—	—	1
M_3	4	0.3	—	—	0.18	0.18	1
M_4	4	0.7	—	—	0.18	0.18	1
M_5	3	—	0.40	0.30	—	—	1

Note: # Par denotes the number of model parameters.

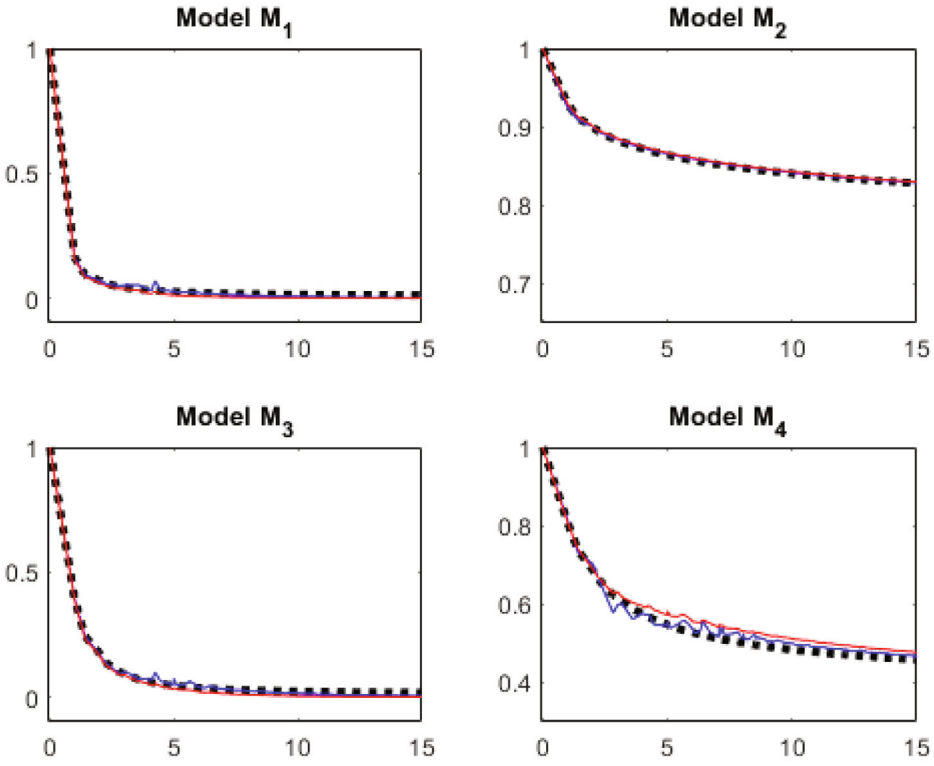

FIGURE 5 Correlation functions of the generating long memory models M_1-M_6 (dashed black curve) and of the conditional autoregressive model providing the best approximation among the models fitted with increasing order $p = 1, \ldots, 9$. The fit shown in blue correspond to the CM criterion while the fit shown in red correspond to the matched spectral density function criterion.

difficult, especially with the increasing of the LM parameter d. In fact, for $d = 0.35$, model M_1 is correctly selected by AIC with a percentage of 73% (BIC 100%) which, however, drops to 2% (BIC 22%) for model M_2, where $d = 0.95$. Results from fitting the GMRFs using the Whittle likelihood, also suggest that as d gets larger, and close to 1, both AIC and BIC favor the fit of a first order CAR. In this case, the estimated autoregressive coefficient of the CAR(1) is very close to the stationary limit of 0.25 such that the approximation is essentially provided by a completely symmetric Intrinsic CAR(1).

Regarding the models M_3 and M_4, the last two panels of Figure 5 show that, in principle, it is possible to obtain a good approximation of the correlation function by fitting higher-order GMRFs using the MC or the MSDF criteria. However, as shown for model M_4, the approximation gets much worse for larger values of d. Fitting the GMRFs using the Whittle likelihood, shows that both AIC and BIC have some difficulties to select correctly the generating LM model. This is especially true for model M_3, and for small values of d in general. For model M_3, in fact, AIC selects the correct model in the 20% of cases while BIC has only a 3% of correct classification. The model selection procedure improves substantially for BIC when d gets larger. For model M_4, in fact, the percentage raises to 96% but for AIC the percentage is still small and close to 16%. Compared with models M_1-M_2, the bias of the estimated LM parameters are larger for models M_3 and M_4 and this may be due to the presence of a positive autoregressive component which can make difficult the distinction between the short and LM components. This phenomenon is known in time series literature (Beran et al., 1998) and might explain the behavior of the information criteria in the model selection procedure. Among the class of CAR models, both AIC and BIC suggest that second-order or third-order GMRFs must be generally preferred to fit the data generated by models M_3 or M_4.

Regarding the separable model M_5, Figure 6 shows that it is difficult to obtain a good approximation of the correlation function of the LM models by using both the MC and the MSDF criteria, even fitting higher-order GMRFs. In the separable case, both the AIC and BIC criteria are able to select the generating model with a percentage of 100%. Comparing the fit of the different GMRFs using the Whittle likelihood, the AIC and BIC criteria also suggest that higher order CARs are necessary to model the generated data, with a preference of CAR(8) and CAR(9) for larger values of d_1 and d_2.

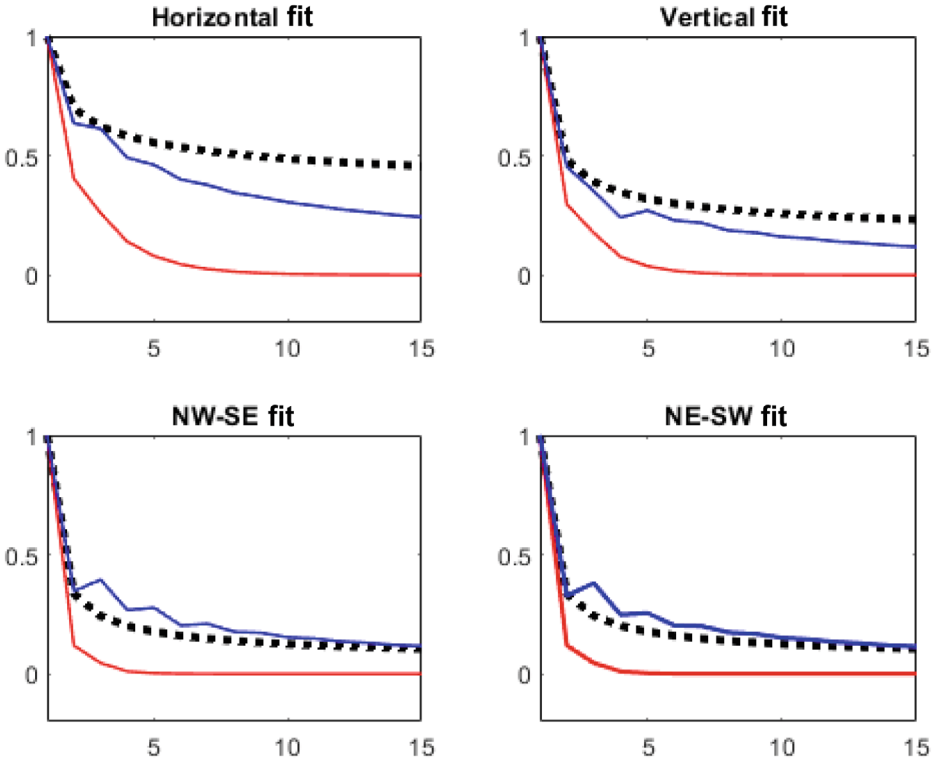

FIGURE 6 Correlation functions along different directions of the generating long memory model M_5 (dashed black curve) and of the selected homogeneous conditional autoregressive models. The fit shown in blue correspond to the CM criterion while the fit shown in red correspond to the matched spectral density function criterion.

9 | DISCUSSION AND SUGGESTIONS FOR FUTURE WORK

Much analysis of spatial data focuses on stochastic modeling. Polynomial regressions can explain some trend in the data, but residuals typically exhibit a stochastic structure. On the lattice, this stochastic component may be modeled parametrically by relying on the class of conditional ARMA models (Ippoliti et al., 2013), which also includes the Simultaneous and Unilateral modeling approaches of Whittle (1954) and Tjostheim (1978).

In some cases, however, we may need to allow for a greater degree of persistence or "memory." One way of doing this is to consider stationary spatial processes whose autocorrelations are not absolutely summable or whose spectral densities are unbounded. We can regard these processes as LM processes and for them we have introduced the class of CARFIMA models. This represents a new class of models which, as shown in Section 4, can include known LM models as special cases.

Parameter estimation for CARFIMA models was discussed in Sections 5 and 6 where, within a simulation framework, we have studied the performance of parametric and semiparametric estimators of the LM parameter. The simulations have confirmed that asymptotic results are valid for maximum likelihood estimation (MLE) under the assumption that the assumed parametric model is correct. Hence, in order to be applicable in practice, parametric approaches have to be combined with an appropriate model choice criterion and in this paper we have explored the use of simple information criteria such as AIC and BIC. Within the semiparametric framework, the use of these information criteria has also led to the proposal of the CARFI method for which we have shown that the difference between the parametric and this semiparametric approach is likely to be small.

In practical applications, the exact Gaussian MLE can be difficult to manage numerically as it requires the computation of the inverse of the covariance matrix as well as its determinant. To overcome this problem, we have assumed toroidal boundary conditions to evaluate the likelihood in $O(n \log n)$ steps. However, the correlations depend on the size of the lattice, and might be unrealistic for sites a long way apart on the planar lattice but close on the torus. This problem can be overcome by embedding the planar lattice in a bigger torus lattice but further studies are needed to fully understand *edge effect* on parameter estimation. Several other approaches have been proposed in literature of spatial statistics to overcome

this large matrix problem (see, e.g., Sun et al., 2012 for a discussion). Examples include imposing tapering the covariance matrix, using composite likelihoods, truncating the spectral representation of a random field, modeling the realizations by a latent process with reduced dimension, and approximating the random field with a Gaussian Markov random field. Section 8 was partly dedicated to the study of this latter approach but the use of other methods in the context of LM needs to be investigated. On the other hand, imposing separability on covariance functions can significantly reduce the dimension of the covariance matrices that need to be inverted, hence they can alleviate the computational problem. A further advantage is that when a separable structure is assumed for a CARFIMA model, closed form expressions for the correlations and inverse correlations are in general possible to obtain. In contrast, for nonseparable structures, there is no known analytic formula for the correlations and inverse correlations and their computation must be carried out numerically.

Compared to parametric estimation, semiparametric methods have the advantage of providing consistent estimates without the necessity of specifying a fixed functional form of the spectral density. These methods are valid for a wide range of processes and have the appealing property of being invariant to the presence of a nonzero mean or being robust to nonstandard behavior of the spectral density function implied, for example, by a pole. However, although semiparametric estimators are asymptotically unbiased, they can have a considerable finite sample bias and the main price to pay for the greater generality that these methods allow is a slower rate of convergence compared with that achieved by parametric alternatives; we note that this discussion also prompts the other important issue of bandwidth selection. One way to avoid or reduce the finite sample bias is to consider a tapered periodogram, for example, in the definition of the GPH or the local Whittle estimator. To overcome such problems, various bias and variance reduction techniques, such as tapering and pooling, have been proposed in the time series literature and this will be considered for future work also for CARFIMA models.

By considering a dataset of land surface temperatures, we have shown that CARFIMA models warrant consideration when fitting spatial models to environmental data observed on regular lattices. However, the data considered in Section 7 pose several challenging statistical issues which require further investigation.

For example, we have considered a simple solution to overcome the difficulties encountered in model parameter estimation in the presence of missing data. Estimating the LM parameter in the presence of missing data is a much less studied problem with sparse literature and the development of both parametric and semiparametric estimators is worth considering for future work. Extensions of the methods proposed in Chan and Palma (1998), which introduces a state-space representation of an ARFIMA model, or those proposed in Martin (1984) for correlated Gaussian processes could be of interest since estimates of the missing values can also be obtained if required.

In this paper we have concentrated on some properties of LM models. Topics for further work on CARFIMA models include investigating efficient methods for checking model adequacy and obtaining predicted values for large lattices.

Another problem which would require further investigation regards the computation of the SE of the model parameters. Computing them evaluating the Jacobian and the Hessian matrix is possible but it requires special attention to avoid numerical problems. To this purpose, alternative approaches, based for example on bootstrap, could be worth considering.

The MA representation of the conditional random field is particularly important because it is the usual tool for analyzing the persistence of the process and care has to be taken in partitioning the parameter space. Accordingly, another topic for future work includes investigating the presence of possible different sources of nonstationarity. Often, stationarity is not a realistic assumption, or it is at least uncertain. This makes identification of stochastic LM even more difficult, because typical LM features may be confounded with nonstationary components. Identifying and assessing possible LM components is thus essential for correct inference about the nonstationary components and one main question here is whether there is evidence for a systematic or stochastic (e.g., in the sense of random walk) trend in the data. Since visual distinction between deterministic, stochastic and spurious trends can be very difficult, a class of semiparametric fractional CAR models can be proposed that includes deterministic trends, Intrinsic CARs and stationarity with short and long range dependence. By extending in the spatial domain results reported in Beran et al. (1998), this approach could help the data analyst to decide whether the observed process contains a stationary short- or long-memory component, a spatial "difference" stationary component, and or a deterministic trend component.

Further methods for approximating CARFIMA models also need to be explored. Since the LM implies that the spectral density function has a singularity at the origin, theoretically it would not be possible to approximate this spectral density by a short memory process, which implies a continuous and bounded spectral density at the origin. However, in applications, it may still be convenient to find a GMRF which gives a good enough approximation. Exploring the extent to which this argument is valid for approximating a LM conditional random field would be relevant in all cases in which a GMRF is fitted to the data by not recognizing explicitly the presence of LM. This is would be especially important

in interpolation problems as the inverse of the precision matrix involved in the conditional expectation (i.e., the linear predictor) is sparse and would allow to work with very large lattices. Preliminary simulation results in Section 8 have shown that the goodness of the approximation depends on the complexity of the covariance structure of the LM process and that the metric used to measure the discrepancy between LM Gaussian random fields and GMRFs can be an important part of the study. The performance of other metrics based on the interpolation theory given in Section 2 could also be explored in the future. Possible examples may consider the maximization of the index of linear determinism $\mathcal{F}$ or different forms of multivariate R^2 measures. This approach has attractive properties when the goal is to predict missing values and parallels the conditional approach proposed by Cressie and Verzelen (2008). Within this framework, theoretical results in Stein (1999), and the lattice approximation approach proposed by Dutta and Mondal (2021) for approximating continuous fractional Gaussian fields, also need to be considered.

Finally, topics for further work refer to the modeling of long-range dependence both in multivariate and in space-time contexts. For the former case, extensions of multivariate CAR models on a regular lattice (Ippoliti et al., 2018) to LM appears straightforward. For spatiotemporal models, early studies were first attempted by Haslett and Raftery (1989), Frias et al. (2006, 2009). Possible extensions can be thought along the research line of the Bayesian Generalized Spatio-Temporal Structural Equation model of Ippoliti et al. (2012) and Gamerman et al. (2022), where spatial and temporal long range dependence structures can be introduced at different levels of the hierarchy and for the different processes specified within the state-space formulation. In this case, efficient parameter estimation of the LM parameter could be obtained by following, for example, Chopin et al. (2013) and Liseo et al. (2001).

ACKNOWLEDGEMENTS

The authors thank the Editor, the Associate Editor, and a referee for helpful comments and suggestions which have substantially improved the quality of the paper. The authors are also very grateful to D. Surgailis for helpful discussion on Example 4.1.

REFERENCES

Ahlfors, L. (2021). *Complex analysis: An introduction to the theory of analytic functions of one complex variable* (Third ed.). AMS Chelsea Publishing.

Anh, V. V., Lam, K. C., Leung, Y., & Tieng, Q. (2000). Multifractal analysis of Hong Kong air quality data. *Environmetrics*, *10*, 139–149.

Anh, V. V., & Lunney, K. E. (1995). Parameter estimation of random fields with long range dependence. *Mathematical and Computer Modelling*, *21*, 67–77.

Baillie, R. T., Chung, C. F., & Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics*, *11*(1), 23–40.

Battaglia, F. (1983). Inverse autocovariances and a measure of linear determinism for a stationary process. *Journal of Time Series Analysis*, *4*, 79–87.

Baxter, G. (1962). An asymptotic result for the finite predictor. *Mathematica Scandinavica*, *10*, 137–144.

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science*, *7*, 404–416.

Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika*, *80*, 817–822.

Beran, J. (1994). *Statistics for long-memory processes. Monographs on statistics and applied probability*. Chapman and Hall/CRC.

Beran, J., Bhansali, R. J., & Ocker, D. (1998). On unified model selection for stationary and non-stationary short- and long-memory autoregressive processes. *Biometrika*, *85*, 921–934.

Beran, J., Feng, Y., Ghosh, S., & Kulik, R. (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Springer.

Beran, J., Ghosh, S., & Schell, D. (2009). On least squares estimation for long-memory lattice processes. *Journal of Multivariate Analysis*, *100*, 2178–2194.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, *2*, 489–502.

Besag, J. (1981). On a system of two-dimensional recurrence equations. *Journal of the Royal Statistical Society, Series B*, *43*, 302–309.

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, *82*, 733–746.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *36*, 192–236.

Bhansali, R. J. (1983). Estimation of the order of a moving average model from autoregressive and window estimates of the inverse correlation function. *Journal of Time Series Analysis*, *4*, 137–162.

Bhansali, R. J., Giraitis, L., & Kokoszka, P. S. (2006). Estimation of the memory parameter by fitting fractionally differenced autoregressive models. *Journal of Multivariate Analysis*, *97*(10), 2101–2130.

Bhansali, R. J., & Ippoliti, L. (2005). Inverse correlations for multiple time series and gaussian random fields and measures of their linear determinism. *Journal of Mathematics and Statistics*, *4*, 287–299.

Bhansali, R. J., & Kokoszka, P. S. (2001). Estimation of the long-memory parameter: A review of recent developments and an extension. *Lecture Notes - Monograph Series*, *37*, 125–150.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, *60*, 217–226.
Bogert, B. P., Healy, M. J. R., & Tukey, J. W. (1963). *The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking*. In M. Rosenblatt (Ed.), *Time series analysis* (pp. 209–243). Academic Press.
Boissy, Y., Bhattacharyya, B. B., Li, X., & Richardson, G. D. (2005). Parameter estimates for fractional autoregressive spatial processes. *Annals of Statistics*, *33*, 2553–2567.
Bouchaud, J. P., Borghesi, C., & Jensen, P. (2014). On the emergence of an intention field for socially cohesive agents. *Journal of Statistical Mechanics: Theory and Experiment*, *3*, P03010.
Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.
Brillinger, D. R. (1974). Fourier analysis of stationary processes. *Proceedings of the IEEE*, *62*(12), 628–642.
Burrough, P. A. (1981). Fractal dimensions of landscapes and other environmental data. *Nature*, *294*, 240–242.
Chan, N. H., & Palma, W. (1998). State space modeling of long-memory processes. *The Annals of Statistics*, *26*, 719–740.
Chopin, N., Rousseau, J., & Liseo, B. (2013). Computational aspects of bayesian spectral density estimation. *Journal of Computational and Graphical Statistics*, *22*, 533–557.
Cox, D. R. (1984). *Long-range dependence: A review*. In H. A. David & H. T. David (Eds.), *Statistics: An appraisal* (pp. 55–74). Iowa State University Press.
Crato, N., & Ray, K. R. (1996). Model selection and forecasting for long-range dependent processes. *Journal of Forecasting*, *15*, 107–125.
Cressie, N. (1993). *Statistics for spatial data* (2nd ed.). Wiley.
Cressie, N., & Verzelen, N. (2008). Conditional-mean least-squares fitting of gaussian markov random fields to gaussian fields. *Computational Statistics & Data Analysis*, *52*, 2794–2807.
Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, *17*(4), 1749–1766.
Dempster, A. P. (1972). Covariance selection. *Biometrics*, *28*, 157–175.
Doukhan, P., Leon, J. R., & Soulier, P. (1996). Central and non central limit theorems for quadratic forms of a strongly dependent Gaussian field. *Brazilian Journal of Probability and Statistics*, *10*, 205–223.
Doukhan, P., Oppenheim, G., & Taqqu, M. (2003). *Theory and applications of long-range dependence*. Birkhauser.
Dryden, I., Ippoliti, L., & Romagnoli, L. (2002). Adjusted maximum likelihood and pseudo-likelihood estimation for noisy Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, *11*, 370–388.
Dutta, S., & Mondal, D. (2021). *On the usefulness of lattice approximations for fractional gaussian fields*. In C. R. Rao & A. S. R. Srinivasa Rao (Eds.), *Data science: Theory and applications, handbook of statistics* (Vol. *44*, pp. 131–154). Elsevier.
Fontanella, L., Ippoliti, L., Martin, R. J., & Trivisonno, S. (2008). Interpolation of spatial and spatio-temporal Gaussian fields using Gaussian Markov random fields. *Advances in Data Analysis and Classification*, *2*(1), 63–79.
Fox, R., & Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, *14*(2), 517–532.
Frias, M. P., Alonso, F. J., Ruiz-Medina, M. D., & Angulo, J. M. (2008). Semiparametric estimation of spatial long-range dependence. *Journal of Statistical Planning and Inference*, *138*, 1479–1495.
Frias, M. P., Ruiz-Medina, M. D., Alonso, F. J., & Angulo, J. M. (2006). Spatiotemporal generation of long-range dependence models and estimation. *Environmetrics*, *17*, 139–146.
Frias, M. P., Ruiz-Medina, M. D., Alonso, F. J., & Angulo, J. M. (2009). Spectral-marginal-based estimation of spatiotemporal long-range dependence. *Communication in Statistics - Theory and Methods*, *38*, 103–114.
Gamerman, D., Ippoliti, L., & Valentini, P. (2022). A dynamic structural equation approach to estimate the short-term effects of air pollution on human health. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *71*, 739–769. https://doi.org/10.1111/rssc.12554
Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*, 221–238.
Ghodsi, 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.
Ghodsi, 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.
Giraitis, L., & Robinson, P. M. (2003). *Parameter estimation under long-range dependence*. In P. Doukhan, G. Oppenheimer, & M. Taqqu (Eds.), *In theory and applications of long range dependence* (pp. 229–249). Birkhäurser.
Giraitis, L., & Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, *86*, 87–104.
Giraitis, L., & Taqqu, M. S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *The Annals of Statistics*, *27*, 178–203.
Gneiting, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *Journal of Applied Probability*, *37*, 1104–1109.
Gneiting, T., & Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, *46*, 269–282.
Gradshteyn, I. S., & Rhyzhik, I. M. (1965). *Tables of integrals, series and products*. Academic Press.
Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, *14*(2), 227–238.
Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series. *Journal of Time Series Analysis*, *1*, 15–30.
Guo, H., Lim, C. Y., & Meerschaert, M. (2009). Local whittle estimator for anisotropic random fields. *Journal of Multivariate Analysis*, *100*, 993–1028.

Guyon, X. (1995). *Random fields on a network: Modeling, statistics, and applications*. Springer-Verlag.
Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, *10*, 130–145.
Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics*, *38*(1), 1–50.
Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., & Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*, 398–425.
Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.
Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, *116*(1), 770–808.
Hurvich, C. M. (2001). Model selection for broadband semiparametric estimation of long memory in time series. *Journal of Time Series Analysis*, *22*(4), 679–709.
Hurvich, C. M., & Beltrao, K. I. (1993). Asymptotics for the low-frequency of the periodogram of a long-memory time series. *Journal of Time Series Analysis*, *14*, 455–472.
Hurvich, C. M., & Brodsky, J. (2001). Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis*, *22*(4), 221–249.
Hurvich, C. M., Deo, R., & Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of a long-memory time series. *Journal of Time Series Analysis*, *19*, 19–46.
Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(4), 297–307.
Ippoliti, L., Martin, R. J., & Bhansali, R. J. (2013). Rational spectral density models for lattice data. *Spatial Statististics*, *6*, 91–108.
Ippoliti, L., Martin, R. J., & Romagnoli, L. (2018). Efficient likelihood computations for some multivariate Gaussian Markov random fields. *Journal of Multivariate Analysis*, *168*(2), 185–200.
Ippoliti, L., Valentini, P., & Gamerman, D. (2012). Space-time modelling of coupled spatio-temporal environmental variables. *Journal of the Royal Statistical Society, Series C*, *61*, 175–200.
Janacek, G. J. (1982). Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis*, *3*, 177–183.
Kashyap, R. L., & Lapsa, P. M. (1984). Synthesis and estimation of random fields using long-correlation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 800–809.
Kohli, P. (2016). Fractional bivariate exponential estimator for long-range dependent random field. *Spatial Statistics*, *15*, 22–38.
Kolmogorov, A. N. (1941). *(a) Dissipation of energy in the locally isotropic turbulence, (b) the local structure of turbulence in incompressible viscous fluid for very large Reynolds number, (c) on degeneration of isotropic turbulence in an incompressible viscous liquid*. In S. K. Friedlander & L. Topper (Eds.), *Turbulence, classic papers on statistical theory* (pp. 151–161). Interscience Publishers.
Koul, H. L., Mimoto, N., & Surgailis, D. (2016). A goodness-of-fit test for marginal distribution of linear random fields with long memory. *Metrika*, *79*, 165–193.
Künsch, H. R. (1987). Statistical aspects of self-similar processes. *Bernoulli*, *1*, 67–74.
Lavancier, F. (2006). *Long memory random fields*. In P. Bertail et al. (Eds.), *Dependence in probability and statistics* (pp. 325–339). Springer.
Leonenko, N. N. (1999). *Limit theorems for random fields with singular spectrum*. Kluwer Academic Publishers.
Lighthill, M. J. (1962). *Introduction to Fourier analysis and generalised functions. Cambridge monographs on mechanics and applied mathematics*. Cambridge University Press.
Lin, G., Chen, X., & Fu, Z. (2007). Temporal-spatial diversities of long-range correlation for relative humidity over China. *Physica A: Statistical Mechanics and its Applications*, *383*, 585–594.
Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B*, *73*, 423–498.
Liseo, B., Marinucci, D., & Petrella, L. (2001). Bayesian semiparametric inference on long-range dependence. *Biometrika*, *88*, 1089–1104.
Mandelbrot, B. B., & van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review*, *10*, 422–437.
Mandelbrot, B. B., & Wallis, J. R. (1969). Computer experiments with fractional gaussian noises. *Water Resources Research*, *5*, 228–267.
Martin, R. J. (1984). Exact maximum likelihood for incomplete data from a correlated gaussian process. *Communications in Statistics, Theory and Methods*, *13*, 1275–1288.
Martin, R. J. (1986). On the design of experiments under spatial correlation. *Biometrika*, *73*, 247–277.
Martin, R. J. (1996). Some results on unilateral ARMA lattice processes. *Journal of Statistical Planning and Inference*, *50*(3), 359–411.
Mateu, J., Porcu, E., & Nicolis, O. (2007). A note on decoupling of local and global behaviours for the dagum random field. *Probabilistic Engineering Mechanics*, *22*, 320–329.
Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, *5*, 439–468.
Moulines, E., & Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, *27*, 1415–1439.
Newcomb, S. (1985). *Astronomical constants: The elements of the four inner planets and the fundamental constants of astronomy*. Government Printing Office.
Palma, W. (2007). *Long-memory time series — Theory and methods*. Wiley.
Parzen, E. (1969). *Mulitple time series modelling*. In P. R. Krishnaiah (Ed.), *Multivariate analysis II* (pp. 389–409). Academic Press.
Peng, J., Ma, J., Liu, Q., & Liu, Y. (2018). Spatial-temporal change of land surface temperature across 285 cities in China: An urban-rural contrast perspective. *Science of the Total Environment*, *635*, 487–497.

Percival, D. B., Rothrock, D. A., Thorndike, A. S., & Gneiting, T. (2008). The variance of mean sea-ice thickness: Effect of long range dependence. *Journal of Geophysical Research Oceans*, *113*, C01004. https://doi.org/10.1029/2007JC004391

Pickard, D. K. (1980). Unilateral markov fields. *Advances in Applied Probability*, *12*, 655–671.

Porcu, E., Crujeiras, R., Mateu, J., & Gonzalez-Manteiga, W. (2009). On the second order properties of the multidimensional periodogram for regularly spaced data. *Theory of Probability & Its Applications*, *53*, 349–356.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association*, *85*, 338–344.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.

Rice, J. (1979). On the estimation of the parameters of a power spectrum. *Journal of Multivariate Analysis*, *9*, 378–392.

Robinson, P. (2020). Spatial long memory. *Japanese Journal of Statistics and Data Science*, *3*, 243–256.

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.

Rozanov, Y. A. (1967). On gaussian fields with given conditional distributions. *Theory of Probability and its Applications*, *12*, 381–391.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields. Theory and Applications*. Chapman and Hall/CRC.

Rue, H., & Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, *29*, 31–49.

Samorodnitsky, G. (2006). Long range dependence. *Foundations and Trends in Stochastic Systems*, *1*, 163–257.

Schmidt, C. M., & Tcherning, R. (1995). *The identification of fractional ARIMA models. Discussion paper*. Humboldt Universitat zu Berlin.

Sethuraman, S., & Basawa, I. V. (1995). Maximum likelihood estimation for a fractionally differenced autoregressive model on a two-dimensional lattice. *Journal of Statistical Planning and Inference*, *44*(2), 219–235.

Shitan, M. (2008). Fractionally integrated separable spatial autoregressive (fissar) model and some of its properties. *Communications in Statistics - Theory and Methods*, *37*, 1266–1273.

Smith, H. F. (1938). An empirical law, describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, *28*, 1–23.

Solo, V. (1986). Modeling of two-dimensional random fields by parametric cepstrum. *IEEE Information Theory*, *32*, 743–750.

Song, H.-R., Fuentes, M., & Ghosh, S. (2008). A comparative study of gaussian geostatistical models and gaussian markov random field models. *Journal of Multivariate Analysis*, *99*, 1681–1697.

Sørbye, S. H., & Rue, H. (2017). Fractional Gaussian noise: Prior specification and model comparison. *Environmetrics*, *29*, e2457.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer Verlag.

Stroud, J. R., Stein, M. L., & Lysen, S. (2017). Bayesian and maximum likelihood estimation for gaussian processes on an incomplete lattice. *Journal of Computational and Graphical Statistics*, *26*, 108–120.

Sun, Y., Li, B., & Genton, M. G. (2012). *Geostatistics for large datasets*. In *Advances and challenges in space-time modelling of natural events* (pp. 55–77). Springer.

Tjostheim, D. (1978). Statistical spatial series modeling. *Advances in Applied Probability*, *10*, 130–154.

Tjostheim, D. (1983). Statistical spatial series modeling ii: Some further results on unilateral lattice processes. *Advances in Applied Probability*, *15*, 562–584.

Verzelen, N. (2010). Adaptive estimation of stationary Gaussian fields. *The Annals of Statistics*, *38*, 1363–1402.

Wang, L. (2009). Gaussian semiparametric estimation for long range dependent random fields. *Statistics and Probability Letters*, *79*, 2297–2306.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*, 434–449.

Whittle, P. (1956). On the variation of yield variance with plot size. *Biometrika*, *43*, 337–343.

Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin de L'Institut International de Statistique*, *39*, 105–129.

Yao, Q., & Brockwell, P. J. (2006) Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli 12*(3), 403–429. https://doi.org/10.3150/bj/1151525128

Yuan, J., & Subba Rao, T. (1993). *Spectral estimation for random fields with applications to markov modelling and texture classification*. In R. Chellappa & A. Jain (Eds.), *Markov random fields, theory and applications* (pp. 179–209). Academic Press.

Zygmund, A. (2003). *Trigonometric series* (3rd ed.). Cambridge University Press.

How to cite this article: Ferretti, A., Ippoliti, L., Valentini, P., & Bhansali, R. J. (2023). Long memory conditional random fields on regular lattices. *Environmetrics*, e2817. https://doi.org/10.1002/env.2817