MDPI

Article

Model-Based and Model-Free point prediction algorithms for locally stationary random fields

Srinjoy Das[1], Yiwen Zhang [2] and Dimitris N. Politis [3]

[1] School of Mathematical and Data Sciences, West Virginia University, Morgantown, WV, 26506 USA
[2] MIT Sloan School of Management, Cambridge, MA 02142 USA
[3] Department of Mathematics and Halicioglu Data Science Institute, University of California—San Diego, La Jolla, CA 92093, USA
* Correspondence: srinjoy.das@mail.wvu.edu

Abstract: The Model-Free Prediction Principle has been successfully applied to general regression problems, as well as problems involving stationary and locally stationary time series. In this paper we demonstrate how Model-Free Prediction can be applied to handle random fields that are only locally stationary such as pixel values over an image or satellite data observed on an ocean surface, i.e., they can be assumed to be stationary only across a limited part over their entire region of definition. We construct novel one-step-ahead Model-Based and Model-Free point predictors and compare their performance using synthetic data as well as images from the CIFAR-10 dataset. In the latter case we demonstrate that our best Model-Free point prediction results outperform those obtained using Model-Based prediction.

Keywords: Kernel smoothing, linear predictor, random fields, nonstationary series, point prediction.

1. Introduction

Consider a real-valued random field dataset $\{Y_{\underline{t}}, \underline{t} \in Z^2\}$ defined over a 2-D index-set D e.g. pixel values over an image or satellite data observed on an ocean surface. It may be unrealistic to assume that the stochastic structure of such a random field $Y_{\underline{t}}$ has stayed invariant over the entire region of definition D hence, we cannot assume that $\{Y_{\underline{t}}\}$ is stationary. Therefore it is more realistic to assume a slowly-changing stochastic structure, i.e., a *locally stationary model*. The theory of locally stationary time series and parametric and nonparametric methods for their estimation have been covered extensively in the literature including references [1], [2], [3], [4], [5], [6], [7], [8]. In [9] we propose Model-Based and Model-Free algorithms for point prediction and prediction intervals of locally stationary time series and demonstrate their applications for both synthetic and real-life datasets. Our work in this paper extends this framework to point prediction over locally stationary random fields with applications involving synthetic and real-life image data.

In the context of random fields two principal modeling approaches are usually followed in order to perform estimation. In the first case for fields of study such as econometrics and ecology where the sampling points can be irregular the random field data $\{Y_{\underline{s}}, \underline{s} \in S\}$ is defined over a continuous subset S of R^d. Modeling strategies for such non-uniformly spaced spatial data have been discussed in [10], [11]. Kernel estimation for locally stationary random fields defined over such irregularly spaced locations has been proposed in [12] and [13] while autoregressive estimation of similarly defined locally stationary random fields has been proposed in [14]. In the other style of modeling the random field $\{Y_{\underline{t}}, \underline{t} \in S\}$ is defined over a regularly spaced grid $S \subset Z^d$. Examples of such applications arise in fields of study such as image processing and radiography. Two dimensional (2D) autoregressive models for such random fields covering various regions of support have been proposed in [15]. Autoregressive estimation of such data regularly spaced on a lattice has also been discussed in [16]. Applications of 2D autoregressive models for analysis and synthesis

Citation: Lastname, F.; Lastname, F.; Lastname, F. Model-Based and Model-Free point prediction algorithms for locally stationary random fields. *Appl. Sci.* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

of textural images are shown in [17]. Local Linear based nonparametric estimators of such random fields and their theoretical properties have been discussed in [18], [19]. In this paper we assume a locally stationary model for random fields $Y_{\underline{t}} \in R$ defined over $\underline{t} \in S$ where $S \subset Z^d$, $d = 2$. Given data $Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n}$, our objective is to perform point prediction for a *future* unobserved data point $Y_{\underline{t}_{n+1}}$. Here $\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_n, \underline{t}_{n+1} \in Z^2$ denote the coordinates of the random field over the 2-D index set D and the notion of a *future* datapoint over a coordinate of a random field for purposes of predictive inference over $\underline{t} \in Z^2$ is defined in Section 2.

The usual approach for dealing with locally stationary series is to assume that the data can be decomposed as the sum of three components:

$$\mu(\underline{t}) + S_{\underline{t}} + W_{\underline{t}}$$

where $\mu(\underline{t})$ is a deterministic trend function, $S_{\underline{t}}$ is a seasonal (periodic) series, and $\{W_{\underline{t}}\}$ is (strictly) stationary with mean zero. This type of decomposition has been proposed for time series [20] and can also be used for decomposition of locally stationary random field data. The seasonal (periodic) component, be it random or deterministic, can be easily estimated and removed and having done that, the 'classical' decomposition simplifies to the following model with additive trend, i.e.,

$$Y_{\underline{t}} = \mu(\underline{t}) + W_{\underline{t}} \tag{1}$$

which can be generalized to accommodate a coordinate-changing variance as well, i.e.,

$$Y_{\underline{t}} = \mu(\underline{t}) + \sigma(\underline{t}) W_{\underline{t}}. \tag{2}$$

In both above models, the series $\{W_{\underline{t}}\}$ is assumed to be (strictly) stationary, weakly dependent, e.g. strong mixing, and satisfying $E(W_{\underline{t}}) = 0$; in model (2), it is also assumed that $var(W_{\underline{t}}) = 1$. The deterministic functions $\mu(\cdot)$ and $\sigma(\cdot)$ are unknown and can be assumed to belong to a class of functions that is either finite-dimensional (parametric) or not (nonparametric). In this paper we focus on the nonparametric case and assume that $\mu(\cdot)$ and $\sigma(\cdot)$ have some degree of smoothness i.e. change smoothly (and slowly) with $\underline{t}$.

Models (1) and (2) can be used to capture the first two moments of the locally stationary random field; however it may be the case that the skewness and/or kurtosis of $Y_{\underline{t}}$ changes with $\underline{t}$. In addition it may also be the case that the correlation $\text{Corr}(Y_{\underline{t}_j}, Y_{\underline{t}_{j+1}})$ changes smoothly (and slowly) with $\underline{t}_j \in Z^2$. To address this more general case we propose a methodology for point prediction of locally stationary random fields that does not rely on simple additive models such as (1) and (2). This is accomplished by using the Model-Free Prediction Principle of [21], [22]. The key towards Model-Free inference is to be able to construct an invertible transformation $H_n : \underline{Y}_{\underline{t}_n} \mapsto \underline{\epsilon}_n$ where $\underline{Y}_{\underline{t}_n} = (Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n})$ denotes the random field data under consideration and $\underline{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)'$ is a random vector with i.i.d. components.

The rest of the paper is arranged as follows. In Section 2 we set up the framework for defining causality of random fields in order to enable us perform point prediction. In Section 3 we visit the problem of Model-Based inference and develop a point prediction methodology for locally stationary random fields. In Section 4 we construct the framework for point prediction of locally stationary random fields using Model-Free inference. In Section 5 we describe how cross validation be used to determine the optimal bandwidths for both Model-Based and Model-Free inference. Finally in Section 6 using finite sample experiments we compare the two novel approaches namely Model-Based of Section 3 and Model-Free of Section 4 using synthetic and real life data.

2. Causality of Random Fields

Given the random field observations $Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_n}$ our goal is to perform predictive inference for the "next" unknown datapoint $Y_{\underline{t}_{n+1}}$. In this context a definition of causality

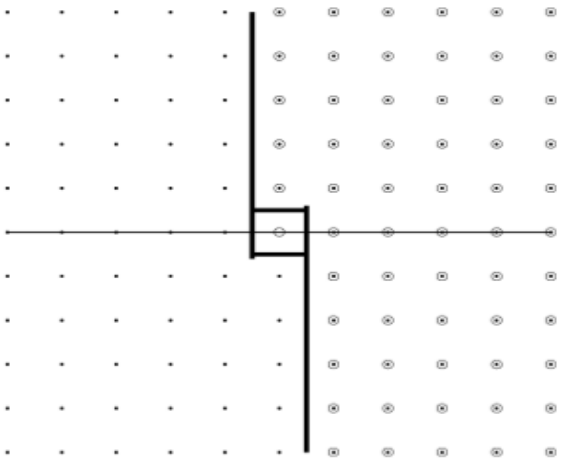

Figure 1. Non Symmetric Half-Plane

is necessary to specify the random field coordinate $\underline{t}_{n+1}$ where predictive inference will be performed. For this purpose we adopt the framework proposed in [15] and consider random fields discussed in this paper to be defined over a subset of the non symmetric half-plane (NSHP) denoted as H_∞. Figure 1 shows an NSHP centered at (0, 0). The NSHP can also be centered at any other point $\underline{t}$ as follows:

$$NSHP(\underline{t}) = \underline{t} + \underline{s} \quad \forall \underline{s} \in NSHP(0,0) \tag{3}$$

Such nonsymmetric half-planes have been used previously for specifying causal 2-D AR models [15]. In such cases a causal 2-D AR model with $H_p \subset H_\infty$ can be defined as below in equation (4) where the set H_p is termed as the region of support (ROS) of the 2-D AR model. Here $H_p = \{(j,k) \mid j = 1,2,\ldots,p \text{ and } k = 0,\pm 1,\ldots,\pm p\} \cup \{(0,k) \mid k = 1,2,\ldots,p\}$ and v_{t_1,t_2} is a 2-D white noise process with mean 0 and variance $\sigma^2 > 0$.

$$Y_{t_1,t_2} = \sum_{(j,k)\in H_p} \beta_{j,k} Y_{t_1-j,t_2-k} + v_{t_1,t_2} \tag{4}$$

Based on [23] a 2-D AR process with ROS S is causal if there exists a subset C of Z^2 satisfying the following conditions:

- The set C consists of 2 rays emanating from the origin and the points lie between the rays
- The angle between the 2 rays is strictly less than 180 degrees
- $S \subset C$

In this case since $H_p \subset H_\infty$ satisfies these conditions the 2-D AR process denoted by (4) is causal. We use this framework to describe a causal random field defined over the NSHP and perform predictive inference on the same. Given this our setup for point prediction of random fields is described as below.

Consider random field data $\{Y_{\underline{t}},\ \underline{t} \in E\}$ where E can be any finite subset of Z^2 for e.g. $E_{\underline{n}} = \{\underline{t} \in Z^2 \text{ with } \underline{n} = (n_1, n_2)\}$. Our goal is predictive inference at $\underline{t} = (t_1, t_2)$ where $0 < t_1 < n_1$ & $0 < t_2 < n_2$,. This "future" value Y_{t_1,t_2} is determined using data defined over the region as shown in Figure 2:

$$E_{\underline{t},\underline{n}} = NSHP(\underline{t}) \cap E_{\underline{n}}$$

Both Model-Based and Model-Free causal inference for Y_{t_1,t_2} are performed using the data specified over this region $E_{\underline{t},\underline{n}}$. We consider predictive inference at $Y_{\underline{t}} = Y_{t_1,t_2}$ given the data $(Y_{\underline{s}} \mid \underline{s} \prec \underline{t}\ \&\ \underline{s} \in E_{\underline{t},\underline{n}})$ where the symbol $\prec$ denotes lexicographical ordering on the region of support of the random field i.e. $(a_k, b_k) \prec (a_{k+1}, b_{k+1})$ if and

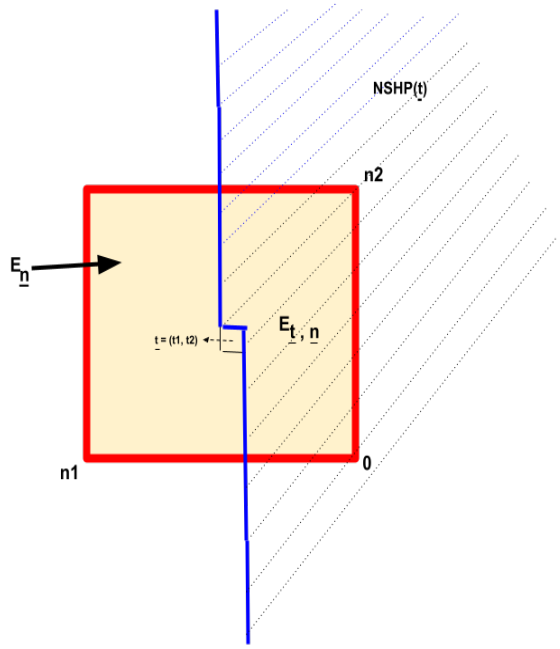

Figure 2. Prediction point for NSHP. In this drawing NSHP($\underline{t}$) denotes the non symmetric half-plane centered at $\underline{t} = (t_1, t_2)$ covering the hashed area. $E_{\underline{n}}$ denotes the finite subset of Z^2 marked by the red boundary. The intersection of the two gives $E_{\underline{t},\underline{n}}$. Point prediction is performed at $\underline{t} = (t_1, t_2)$.

only if either $a_k < a_{k+1}$ or $(a_k = a_{k+1}$ and $b_k < b_{k+1})$ [15]. In the subsequent discussion the lexicographically ordered "past" data $Y_{\underline{s}}$ will be denoted as $Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n}$ and point prediction will be performed at $Y_{\underline{t}} = Y_{\underline{t}_{n+1}}$.

3. Model-Based Point Prediction

We adopt the time changing mean and variance model as given by Equation (2). The L_2–optimal predictor of $Y_{\underline{t}_{n+1}}$ given the data $Y_{\underline{s}} = \underline{Y}_{\underline{t}_n} = (Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_n})'$ is the conditional expectation $E(Y_{\underline{t}_{n+1}}|\underline{Y}_{\underline{t}_n})$. Using model (2) and assuming that $\underline{W}_{\underline{t}_n}$ is weakly dependent it can be shown that [9]:

$$E(Y_{\underline{t}_{n+1}}|\underline{Y}_{\underline{t}_n}) = \mu(\underline{t}_{n+1}) + \sigma(\underline{t}_{n+1})E(W_{\underline{t}_{n+1}}|\underline{W}_{\underline{t}_n}). \tag{5}$$

From the above equation we can see that for Model-Based point prediction we need to estimate the conditional expectation $E(W_{\underline{t}_{n+1}}|\underline{W}_{\underline{t}_n})$ as well as the coordinate changing trend and variance i.e. $\mu(\underline{t}_{n+1})$ and $\sigma(\underline{t}_{n+1})$.

Estimating the conditional expectation: This is done by fitting a (causal) AR(p, q) model to the data $\underline{W}_{\underline{t}_n} = (W_{\underline{t}_1}, \ldots, W_{\underline{t}_n})$ with p, q chosen by minimizing AIC, BIC or a related criterion as described in [15]. Using this framework involves estimating the coefficients of the following 2-D AR model defined over a ROS H_p as described in Section 2:

$$W_{r,s} = \sum_{(j,k)\in H_p} \beta_{j,k} W_{r-j,s-k} + v_{r,s} \tag{6}$$

Here $v_{r,s}$ is a 2-D white noise process i.e., an uncorrelated sequence, with mean 0 and variance $\sigma^2 > 0$. For our point prediction problem setting this implies that

$$\bar{E}(W_{\underline{t}_{n+1}}|\underline{W}_{\underline{t}_n}) = \sum_{(j,k)\in H_p} \beta_{j,k} W_{t_{n_1}-j,t_{n_2}-k} \tag{7}$$

Estimating the trend and variance: This can be performed by using kernel smoothing [24–26] using 2D kernels i.e. Nadaraya-Watson (NW) estimation. In addition since predicting $Y_{\underline{t}_{n+1}}$ is essentially a boundary problem it is also possible to use local linear fitting which has been reported to have smaller bias than kernel smoothing for such estimation problems [26–28]. For time series problems $\{Y_t,\ t \in Z\}$ local linear nonparametric estimation can approximate the trend locally by a straight line whereas for the case of random fields $\{Y_{\underline{t}},\ \underline{t} \in Z^2\}$ discussed in this paper local linear estimation can be used to approximate the trend locally with a plane.

In order to estimate $E(W_{\underline{t}_{n+1}}|\underline{W}_{\underline{t}_n})$ the stationary data $W_{\underline{t}_1}, \ldots, W_{\underline{t}_n}$ needs to be estimated. In this case $W_{\underline{t}}$ has to be calculated in a one-sided manner for all points including those at the center of the dataset else the obtained values $\hat{W}_{\underline{t}_1}, \ldots, \hat{W}_{\underline{t}_n}$ will not be stationary which leads to incorrect estimation of the conditional expectation of $W_{\underline{t}_{n+1}}$. The one-sided kernel smoothed and local linear estimators used for calculating $\hat{W}_{\underline{t}_1}, \ldots, \hat{W}_{\underline{t}_n}$ can be defined in two ways as shown in the equations below on NW–Regular, NW–Predictive, LL–Regular and LL–Predictive fitting. Here the bandwidth parameter b is assumed to satisfy

$$b \to \infty \text{ as } n \to \infty \text{ but } b/n \to 0 \tag{8}$$

We will assume throughout that $K(\cdot)$ is a nonnegative, symmetric 2-D Gaussian kernel function for which the diagonal values are set to the bandwidth b and the off-diagonal terms are set to 0. Random field data is denoted as $Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_k}, \ldots Y_{\underline{t}_n}$.

1. **NW–Regular fitting:** Let $\underline{t}_k \in [\underline{t}_1, \underline{t}_n]$, and define

$$\begin{aligned} \hat{\mu}(\underline{t}_k) &= \sum_{i=1}^{k} Y_{\underline{t}_i}\, \hat{K}\left(\frac{\underline{t}_k - \underline{t}_i}{b}\right) \quad \text{and} \\ \hat{M}(\underline{t}_k) &= \sum_{i=1}^{k} Y_{\underline{t}_i}^2\, \hat{K}(\frac{\underline{t}_k - \underline{t}_i}{b}) \end{aligned} \tag{9}$$

where

$$\begin{aligned} \hat{\sigma}(\underline{t}_k) &= \sqrt{\hat{M}_{\underline{t}_k} - \hat{\mu}(\underline{t}_k)^2} \quad \text{and} \\ \hat{K}\left(\frac{\underline{t}_k - \underline{t}_i}{b}\right) &= \frac{K(\frac{\underline{t}_k - \underline{t}_i}{b})}{\sum_{j=1}^{k} K(\frac{\underline{t}_k - \underline{t}_j}{b})}. \end{aligned} \tag{10}$$

Using $\hat{\mu}(\underline{t}_k)$ and $\hat{\sigma}(\underline{t}_k)$ we can now define the *fitted* residuals by

$$\hat{W}_{\underline{t}_k} = \frac{Y_{\underline{t}_k} - \hat{\mu}(\underline{t}_k)}{\hat{\sigma}(\underline{t}_k)} \text{ for } \underline{t}_k = \underline{t}_1, \ldots, \underline{t}_n. \tag{11}$$

2. **NW–Predictive fitting:**

$$\tilde{\mu}(\underline{t}_k) = \sum_{i=1}^{k-1} Y_{\underline{t}_i} \tilde{K}\left(\frac{\underline{t}_k - \underline{t}_i}{b}\right) \text{ and}$$
$$\tilde{M}(\underline{t}_k) = \sum_{i=1}^{k-1} Y_{\underline{t}_i}^2 \tilde{K}(\frac{\underline{t}_k - \underline{t}_i}{b}) \tag{12}$$

where

$$\tilde{\sigma}(\underline{t}_k) = \sqrt{\tilde{M}_{\underline{t}_k} - \tilde{\mu}(\underline{t}_k)^2} \text{ and}$$
$$\tilde{K}\left(\frac{\underline{t}_k - \underline{t}_i}{b}\right) = \frac{K(\frac{\underline{t}_k - \underline{t}_i}{b})}{\sum_{j=1}^{k-1} K(\frac{\underline{t}_k - \underline{t}_j}{b})}. \tag{13}$$

Using $\tilde{\mu}(\underline{t}_k)$ and $\tilde{\sigma}(\underline{t}_k)$ we can now define the *predictive* residuals by

$$\tilde{W}_{\underline{t}_k} = \frac{Y_{\underline{t}_k} - \tilde{\mu}(\underline{t}_k)}{\tilde{\sigma}(\underline{t}_k)} \text{ for } \underline{t}_k = \underline{t}_1, \ldots, \underline{t}_n. \tag{14}$$

Similarly, the one-sided local linear (LL) fitting estimators of $\mu(\underline{t}_k)$ and $\sigma(\underline{t}_k)$ can be defined in two ways.

1. **LL–Regular fitting:** Let $\underline{t}_k \in [\underline{t}_1, \underline{t}_n]$, and define

$$\hat{\mu}(\underline{t}_k) = \frac{\sum_{j=1}^{k} w_j Y_{\underline{t}_j}}{\sum_{j=1}^{k} w_j + n^{-2}} \text{ and}$$
$$\hat{M}(\underline{t}_k) = \frac{\sum_{j=1}^{k} w_j Y_{\underline{t}_j}^2}{\sum_{j=1}^{k} w_j + n^{-2}} \tag{15}$$

Denoting

$$\underline{a} = (a_1, a_2) = (\underline{t}_j - \underline{t}_k) \tag{16}$$

$$s_{t1,1} = \sum_{j=1}^{k} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1 \tag{17}$$

$$s_{t2,1} = \sum_{j=1}^{k} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_2 \tag{18}$$

$$s_{t1,2} = \sum_{j=1}^{k} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1^2 \tag{19}$$

$$s_{t2,2} = \sum_{j=1}^{k} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_2^2 \tag{20}$$

$$s_{t1,t2} = \sum_{j=1}^{k} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1 a_2 \tag{21}$$

$$
\begin{aligned}
w_j &= K(\frac{\underline{t}_j - \underline{t}_k}{b}) \\
&\Big\{ s_{t1,2}s_{t2,2} - s^2_{t1,t2} \\
&-a_1(s_{t1,1}s_{t2,2} - s_{t2,1}s_{t1,t2}) \\
&+a_2(s_{t1,1}s_{t1,t2} - s_{t1,2}s_{t2,1}) \Big\}
\end{aligned}
\tag{22}
$$

The term n^{-2} in eq. (15) is just to ensure the denominator is not zero; see [29]. Eq. (10) then yields $\hat{\sigma}(\underline{t}_k)$, and eq. (11) yields $\hat{W}_{\underline{t}_k}$.

2. **LL–Predictive fitting:**

$$
\begin{aligned}
\tilde{\mu}(\underline{t}_k) &= \frac{\sum_{j=1}^{k-1} w_j Y_{\underline{t}_j}}{\sum_{j=1}^{k-1} w_j + n^{-2}} \text{ and} \\
\tilde{M}(\underline{t}_k) &= \frac{\sum_{j=1}^{k-1} w_j Y^2_{\underline{t}_j}}{\sum_{j=1}^{k-1} w_j + n^{-2}}
\end{aligned}
\tag{23}
$$

where

$$\underline{a} = (a_1, a_2) = (\underline{t}_j - \underline{t}_k) \tag{24}$$

$$s_{t1,1} = \sum_{j=1}^{k-1} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1 \tag{25}$$

$$s_{t2,1} = \sum_{j=1}^{k-1} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_2 \tag{26}$$

$$s_{t1,2} = \sum_{j=1}^{k-1} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1^2 \tag{27}$$

$$s_{t2,2} = \sum_{j=1}^{k-1} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_2^2 \tag{28}$$

$$s_{t1,t2} = \sum_{j=1}^{k-1} K\left(\frac{\underline{t}_j - \underline{t}_k}{b}\right) a_1 a_2 \tag{29}$$

$$
\begin{aligned}
w_j &= K(\frac{\underline{t}_j - \underline{t}_k}{b}) \\
&\Big\{ s_{t1,2}s_{t2,2} - s^2_{t1,t2} \\
&-a_1(s_{t1,1}s_{t2,2} - s_{t2,1}s_{t1,t2}) \\
&+a_2(s_{t1,1}s_{t1,t2} - s_{t1,2}s_{t2,1}) \Big\}
\end{aligned}
\tag{30}
$$

Eq. (13) then yields $\tilde{\sigma}(\underline{t}_k)$, and eq. (14) yields $\tilde{W}_{\underline{t}_k}$.

Using one of the above four methods (NW vs. LL, regular vs. predictive) gives estimates of the quantities needed to compute the L_2–optimal predictor (5). The bandwidth b in all 4 algorithms can be determined by cross-validation as described in Section 5.

4. Model-Free Point Prediction

For the Model-based case Equation (2) accounts for spatially-changing mean and variance of $Y_{\underline{t}}$. More generally however it may happen that the random field $\{Y_{\underline{t}}$ for $\underline{t} \in Z^2\}$ has a nonstationarity in its third (or higher moment), and/or in some other feature of its mth marginal distribution. This is addressed by the Model-Free Prediction Principle of Politis (2013, 2015).

The key towards Model-Free inference is to be able to construct an invertible transformation $H_n : \underline{Y}_{\underline{t}_n} \mapsto \underline{\epsilon}_n$ where $\underline{Y}_{\underline{t}_n} = (Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n})$ denotes the random field data under consideration and $\underline{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)'$ is a random vector with i.i.d. components. In order to do this in our context, let some $m \geq 1$, and denote by $\mathcal{L}(Y_{\underline{t}_k}, Y_{\underline{t}_{k-1}}, \ldots, Y_{\underline{t}_{k-m+1}})$ the mth marginal of $Y_{\underline{t}_k}$ i.e. the joint probability law of the vector $(Y_{\underline{t}_k}, Y_{\underline{t}_{k-1}}, \ldots, Y_{\underline{t}_{k-m+1}})'$. We assume that $\mathcal{L}(Y_{\underline{t}_k}, Y_{\underline{t}_{k-1}}, \ldots, Y_{\underline{t}_{k-m+1}})$ changes smoothly (and slowly) with $\underline{t}_k$ in order to use nonparametric smoothing for estimation. In this case $\{Y_{\underline{t}_k}, \underline{t}_k \in Z^2\}$ can be defined over a 2-D index-set D and the set $(Y_{\underline{t}_k}, Y_{\underline{t}_{k-1}}, \ldots, Y_{\underline{t}_{k-m+1}})$ can be considered to be lexicographically ordered as discussed previously in Section 2.

Similar to the framework proposed in [9] in order to ensure both the smoothness and data-based consistent estimation of $\mathcal{L}(Y_{\underline{t}_k}, Y_{\underline{t}_{k-1}}, \ldots, Y_{\underline{t}_{k-m+1}})$ we assume that, for all $\underline{t}_k$,

$$Y_{\underline{t}_k} = \mathbf{f}_{\underline{t}_k}(W_{\underline{t}_k}, W_{\underline{t}_{k-1}}, \ldots, W_{\underline{t}_{k-m+1}}) \tag{31}$$

for some function $\mathbf{f}_{\underline{t}_k}(w)$ that is smooth in both arguments $\underline{t}_k$ and w, and some strictly stationary and weakly dependent, univariate series $W_{\underline{t}_k}$ where without loss of generality, it is assumed that $W_{\underline{t}_k}$ is a Gaussian series. Here model (2) is a special case of Eq. (31) with $m = 1$, and the function $\mathbf{f}_{\underline{t}_k}(w)$ being affine/linear in w. Therefore, for comparison with the Model-Based case of Eq. (2), in this section we focus on the case $m = 1$. For reference Model-Free estimators for point prediction and prediction intervals in the case of locally stationary time series for $m = 1$ have been discussed in [9]. Below we describe the steps necessary to construct the invertible transformation H_n required to perform Model-Free point prediction for locally stationary random fields for the case $m = 1$.

Step 1: Transformation to uniform samples

With $m = 1$ let $D_{\underline{t}}(y) = P\{Y_{\underline{t}} \leq y\}$ denote the first marginal distribution of the random field $\{Y_{\underline{t}}\}$. Applying the probability integral transform we have

$$U_{\underline{t}} = D_{\underline{t}}(Y_{\underline{t}}) \text{ for } \underline{t} = \underline{t}_1, \ldots, \underline{t}_n \tag{32}$$

Here $U_{\underline{t}_1}, \ldots, U_{\underline{t}_n}$ are random variables having distribution Uniform $(0, 1)$. In this case it is assumed that $D_{\underline{t}}(y)$ is (absolutely) continuous in y for all $\underline{t}$. Therefore we can use either local constant or local linear fitting to estimate it.

Using **local constant fitting** a smooth estimator can be defined as:

$$\bar{D}_{\underline{t}_k}(y) = \sum_{i=1}^{T} \Lambda\left(\frac{y - Y_{\underline{t}_i}}{h_0}\right) \tilde{K}\left(\frac{\underline{t}_k - \underline{t}_i}{b}\right) \tag{33}$$

where $\tilde{K}(\frac{\underline{t}_k - \underline{t}_i}{b}) = K(\frac{\underline{t}_k - \underline{t}_i}{b}) / \sum_{j=1}^{T} K(\frac{\underline{t}_k - \underline{t}_j}{b})$, $\Lambda(y)$ is a smooth distribution function which is strictly increasing with density $\lambda(y) > 0$ i.e. $\Lambda(y) = \int_{-\infty}^{y} \lambda(s)ds$ and h_0 is a secondary bandwidth. Furthermore, as in Section 3, we can let $T = k$ or $T = k - 1$ leading to a **fitted vs. predictive** way to estimate $D_{\underline{t}_k}(y)$ using $\bar{D}_{\underline{t}_k}(y)$. Similar to the Model-Based case we will assume throughout that $K(\cdot)$ is a nonnegative, symmetric 2-D Gaussian kernel function

for which the diagonal values are set to the bandwidth b and the off-diagonal terms are set to 0. Note that the kernel estimator (33) is one-sided for the same reasons discussed before in Section 3. Cross-validation is used to determine the bandwidths h_0 and b ; details are described in Section 5.

Since point prediction is performed on the boundary of the random field one can also consider **local linear estimation** as an alternative to the local constant based smoothing approach. $\bar{D}_{\underline{t}_k}(y)$ as defined in eq. (33) is the Nadaraya-Watson smoother of the variables $v_1, \ldots, v_n$ where $v_i = \Lambda(\frac{y-Y_{\underline{t}_i}}{h_0})$. It is possible to define $\bar{D}^{LL}_{\underline{t}_k}(y)$ which is the local linear estimator of $D_{\underline{t}_k}(y)$ based on the smoothed variables $\Lambda(\frac{y-Y_{\underline{t}_i}}{h_0})$. This estimator is expected to have smaller bias than $\bar{D}_{\underline{t}_k}(y)$. However, there is no guarantee that this will be a proper distribution function as a function of y, i.e., being nondecreasing in y with a left limit of 0 and a right limit of 1 as discussed in [26]. A proposed solution put forward by Hansen [30] involves a straightforward adjustment to the local linear estimator of a conditional distribution function that maintains its favorable asymptotic properties. The local linear version of $\bar{D}_{\underline{t}_k}(y)$ adjusted via Hansen's (2004) proposal is given as follows:

$$\bar{D}^{LLH}_{\underline{t}_k}(y) = \frac{\sum_{i=1}^{T} w_i^{\diamond} \Lambda(\frac{y-Y_{\underline{t}_i}}{h_0})}{\sum_{i=1}^{T} w_i^{\diamond}}. \tag{34}$$

The weights $w_i^{\diamond}$ are derived from weights w_i described in equations (22) and (30) for the fitted and predictive cases where:

$$w_i^{\diamond} = \begin{cases} 0 & \text{when } w_i < 0 \\ w_i & \text{when } w_i \geq 0 \end{cases} \tag{35}$$

As with eq. (33), we can let $T = k$ or $T = k-1$ in the above, leading to a fitted vs. predictive local linear estimators of $D_{\underline{t}_k}(y)$ using $\bar{D}^{LLH}_{\underline{t}_k}(y)$.

One problem with the local linear estimator described above is that it replaces negative weights by zeros, and then renormalizes the nonzero weights. However if estimation is performed on the boundary (as in the case with one-step ahead prediction of random fields), negative weights are crucially needed in order to ensure the extrapolation takes place with minimal bias. To address this problem we modify the original, possibly non-monotonic local linear distribution estimator $\bar{D}^{LL}_{\underline{t}_k}(y)$ to construct a monotonic version denoted by $\bar{D}^{LLM}_{\underline{t}_k}(y)$. The **Monotone Local Linear Distribution Estimator** $\bar{D}^{LLM}_{\underline{t}_k}(y)$ can be constructed by Algorithm 1 as given below[31].

Algorithm 1. Monotone Local Linear Distribution Estimation

1. *Recall that the derivative of* $\bar{D}^{LL}_{\underline{t}_k}(y)$ *with respect to* y *is given by*

$$\bar{d}^{LL}_{\underline{t}_k}(y) = \frac{\frac{1}{h_0}\sum_{j=1}^{T} w_j \lambda(\frac{y-Y_{\underline{t}_j}}{h_0})}{\sum_{j=1}^{n} w_j}$$

   where $\lambda(y)$ *is the derivative of* $\Lambda(y)$ *and the weights* w_j *can be derived based on equations (22) and (30) for the fitted and predictive cases.*
2. *Define a nonnegative version of* $\bar{d}^{LL}_{\underline{t}_k}(y)$ *as* $\bar{d}^{LL+}_{\underline{t}_k}(y) = \max(\bar{d}^{LL}_{\underline{t}_k}(y), 0)$.
3. *To make the above a proper density function, renormalize it to area one, i.e., let*

$$\bar{d}^{LLM}_{\underline{t}_k}(y) = \frac{\bar{d}^{LL+}_{\underline{t}_k}(y)}{\int_{-\infty}^{\infty} \bar{d}^{LL+}_{\underline{t}_k}(s)ds}. \tag{36}$$

4. *Finally, define* $\bar{D}^{LLM}_{\underline{t}_k}(y) = \int_{-\infty}^{y} \bar{d}^{LLM}_{\underline{t}_k}(s)ds$.

The above modification of the local linear estimator allows one to maintain monotonicity while retaining the negative weights that are helpful in problems which involve estimation at the boundary. As with eq. (33), we can let $T = k$ or $T = k-1$ in the above, leading to a **fitted vs. predictive** local linear estimators of $D_{\underline{t}_k}(y)$ that are monotone.

Step 2: Transformation to iid normal samples

Starting from the original random field data $Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_n}$ by using either the local constant, the local linear or the monotone local linear distribution estimator in Step 1 it is possible to obtain samples $U_{\underline{t}_1}, \ldots, U_{\underline{t}_n}$ having distribution Uniform $(0,1)$. However these samples are dependent and therefore additional steps are necessary to convert them to i.i.d. samples as required for Model-Free inference. This is done as described below.

Let Φ denote the cumulative distribution function (cdf) of the standard normal distribution. Therefore we have:

$$Z_{\underline{t}} = \Phi^{-1}(U_{\underline{t}}) \text{ for } \underline{t} = \underline{t}_1, \ldots, \underline{t}_n; \tag{37}$$

Here $Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}$ are correlated standard normal random variables. Now let Γ_n denote the $n \times n$ covariance matrix of the random vector $\underline{Z}_{\underline{t}_n} = (Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n})'$. Consider the Cholesky decomposition $\Gamma_n = C_n C_n'$ where C_n is (lower) triangular, and construct the *whitening* transformation:

$$\underline{\epsilon}_n = C_n^{-1} \underline{Z}_{\underline{t}_n}. \tag{38}$$

It then follows that the entries of $\underline{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)'$ are uncorrelated standard normal. Assuming that the random variables $Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}$ are *jointly* normal it can then be inferred that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0,1)$. Joint normality can be established by assuming a generative model of the random field as given by Equation (31); for a more detailed discussion refer to [9].

To implement the whitening transformation (38), it is necessary to estimate Γ_n, i.e., the $n \times n$ covariance matrix of the random vector $\underline{Z}_{\underline{t}_n} = (Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n})'$ where the $Z_{\underline{t}}$ are the normal random variables defined in eq. (37). The problem involves positive definite estimation of Γ_n based on the sample $Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}$. This estimate is based on the sample autocovariance which is defined for a 2D second-order stationary random field $\{y_{r,s} | r = 1, 2, \ldots, R, \;\; s = 1, 2, \ldots, S\}$ as follows [15]:

$$\begin{aligned} \check{\gamma}(j,k) &= \check{\gamma}(-j,-k) \\ &= \frac{1}{(R-j)(S-k)} \sum_{s=1}^{R-j} \sum_{t=1}^{S-k} \{y_{r+j,s+k} - \bar{y}\}\{y_{r,s} - \bar{y}\} \end{aligned} \tag{39}$$

$$\begin{aligned} \check{\gamma}(j,-k) &= \check{\gamma}(-j,k) \\ &= \frac{1}{(R-j)(S-k)} \sum_{s=1}^{R-j} \sum_{t=k+1}^{S} \{y_{r+j,s-k} - \bar{y}\}\{y_{r,s} - \bar{y}\} \end{aligned} \tag{40}$$

where $(j, k = 0, 1, 2, \ldots)$

Now let $\hat{\Gamma}_n^{AR}$ be the $n \times n$ covariance matrix associated with the fitted AR(p,q) model to the data $Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}$ with p, q by minimizing AIC, BIC or a related criterion as described in [15]. Let $\hat{\gamma}^{AR}_{|i-j|}$ denote the i, j element of the Toeplitz matrix $\hat{\Gamma}_n^{AR}$. Using the 2D Yule-Walker equations to fit the AR model implies that $\hat{\gamma}^{AR}_{k,l} = \check{\gamma}_{k,l}$ for $k = 0, 1, \ldots, p$ and $l = 0, 1, \ldots, q$.

For the cases where $k > p$ or $l > q$, $\hat{\gamma}^{AR}_{k,l}$ can be fitted by iterating the difference equation that characterizes the fitted 2D AR model. In the R software this procedure is automated for time series using the `ARMAacf()` function, here we extend the same approach for stationary data over random fields.

Estimating the 'uniformizing' transformation $D_{\underline{t}}(\cdot)$ and the whitening transformation based on Γ_n allows us to construct the transformation $H_n : \underline{Y}_{\underline{t}_n} \mapsto \underline{\epsilon}_n$. Here $\underline{\epsilon}_n$ is a random vector with i.i.d. components as required by the Model-Free prediction principle. Since all the steps in the transformation, i.e., eqs. (32), (37) and (38), are invertible; therefore, the composite transformation $H_n : \underline{Y}_{\underline{t}_n} \mapsto \underline{\epsilon}_n$ is also invertible. However, in order to put the Model-Free Prediction Principle to work, we also need to estimate the transformation H_{n+1} (and its inverse). To do so, we need a positive definite estimator for the matrix Γ_{n+1}; this can be accomplished by extending the covariance matrix associated with the fitted 2D AR(p,q) model to $(n+1)$ by $(n+1)$ i.e. calculate $\hat{\Gamma}^{AR}_{n+1}$.

Consider the following vectors which include the additional values $Y_{\underline{t}_{n+1}}, Z_{\underline{t}_{n+1}}$ and ϵ_{n+1} that have not yet been estimated:

- $\underline{Y}_{\underline{t}_{n+1}} = (Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_n}, Y_{\underline{t}_{n+1}})'$,
- $\underline{Z}_{\underline{t}_{n+1}} = (Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}, Z_{\underline{t}_{n+1}})'$ and
- $\underline{\epsilon}_{n+1} = (\epsilon_1, \ldots, \epsilon_n, \epsilon_{n+1})'$

We now show how to obtain the inverse transformation $H^{-1}_{n+1} : \underline{\epsilon}_{n+1} \mapsto \underline{Y}_{\underline{t}_{n+1}}$. Since $\underline{\epsilon}_n$ and $\underline{Y}_{\underline{t}_n}$ are related in a one-to-one way via transformation H_n, therefore the values $Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_n}$ are obtainable by $\underline{Y}_{\underline{t}_n} = H^{-1}_n(\epsilon_n)$. Similar to the framework proposed for locally stationary time series in [9] below we show how to create the unobserved $Y_{\underline{t}_{n+1}}$ from $\underline{\epsilon}_{n+1}$ using the following three steps.

Algorithm 2. *GENERATION OF UNOBSERVED DATAPOINT FROM FUTURE INNOVATIONS*

i. *Let*

$$\underline{Z}_{\underline{t}_{n+1}} = C_{n+1}\underline{\epsilon}_{n+1} \tag{41}$$

where C_{n+1} is the (lower) triangular Cholesky factor of (our positive definite estimate of) Γ_{n+1}. From the above, it follows that

$$Z_{\underline{t}_{n+1}} = \underline{c}_{n+1}\underline{\epsilon}_{n+1} \tag{42}$$

where $\underline{c}_{n+1} = (c_1, \ldots, c_n, c_{n+1})$ is a row vector consisting of the last row of matrix C_{n+1}.

ii. *Create the uniform random variable*

$$U_{\underline{t}_{n+1}} = \Phi(Z_{\underline{t}_{n+1}}). \tag{43}$$

iii. *Finally, define*

$$Y_{\underline{t}_{n+1}} = D^{-1}_{\underline{t}_{n+1}}(U_{\underline{t}_{n+1}}); \tag{44}$$

where in practice, the above will be based on an estimate of $D^{-1}_{\underline{t}_{n+1}}(\cdot)$.

Since $\underline{Y}_{\underline{t}_n}$ has already been created using (the first n coordinates of) $\underline{\epsilon}_{n+1}$, the above completes the construction of $\underline{Y}_{\underline{t}_{n+1}}$ based on $\underline{\epsilon}_{n+1}$, i.e., the mapping $H^{-1}_{n+1} : \underline{\epsilon}_{n+1} \mapsto \underline{Y}_{\underline{t}_{n+1}}$. By combining eq. (42), (43) and (44) we can write the formula:

$$Y_{\underline{t}_{n+1}} = D^{-1}_{\underline{t}_{n+1}}(\Phi(\underline{c}_{n+1}\underline{\epsilon}_{n+1})).$$

The term $\underline{c}_{n+1}\underline{\epsilon}_{n+1}$ can be written as $\sum_{i=1}^{n} c_i\epsilon_i + c_{n+1}\epsilon_{n+1}$; hence, the above can be compactly denoted as

$$\begin{aligned} Y_{\underline{t}_{n+1}} &= g_{n+1}(\epsilon_{n+1}) \text{ where} \\ g_{n+1}(x) &= D^{-1}_{\underline{t}_{n+1}}\left(\Phi\left(\sum_{i=1}^{n} c_i\epsilon_i + c_{n+1}x\right)\right). \end{aligned} \tag{45}$$

Eq. (45) is the predictive equation required in the Model-free Prediction Principle where $Y_{\underline{t}_{n+1}}$ is estimated conditionally on $\underline{Y}_{\underline{t}_n} = (Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n})$. The complete algorithm for constructing the Model-Free point predictors is as described below:

Algorithm 3. MODEL-FREE (MF) POINT PREDICTORS FOR $Y_{\underline{t}_{n+1}}$

1. *Construct* $U_{\underline{t}_1}, \ldots, U_{\underline{t}_n}$ *by eq. (32) with* $D_{\underline{t}_k}(\cdot)$ *estimated by either* $\bar{D}_{\underline{t}_k}(\cdot)$, $\bar{D}^{LLH}_{\underline{t}_k}(\cdot)$ *or* $\bar{D}^{LLM}_{\underline{t}_k}(\cdot)$ *where* $\underline{t}_k \in [\underline{t}_1, \underline{t}_n]$
2. *Construct* $Z_{\underline{t}_1}, \ldots, Z_{\underline{t}_n}$ *by eq. (37)*
3. *Construct* $\epsilon_1, \ldots, \epsilon_n$ *by eq. (38), and let* $\hat{F}_n$ *denote their empirical distribution.*
4. *The Model-free* L_2*–optimal point predictor of* $Y_{\underline{t}_{n+1}}$ *is then*

$$\hat{Y}_{\underline{t}_{n+1}} = \int g_{n+1}(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} g_{n+1}(\epsilon_i)$$

   where the function g_{n+1} *is defined in the predictive equation (45) with* $D_{\underline{t}_{n+1}}(\cdot)$ *being again estimated by either* $\bar{D}_{\underline{t}_{n+1}}(\cdot)$, $\bar{D}^{LLH}_{\underline{t}_{n+1}}(\cdot)$ *or* $\bar{D}^{LLM}_{\underline{t}_{n+1}}(\cdot)$
5. *The Model-free* L_1*–optimal point predictor of* $Y_{\underline{t}_{n+1}}$ *is given by the median of the set* $\{g_{n+1}(\epsilon_i)$ *for* $i = 1, \ldots, n\}$.

5. Random Fields cross-validation

To choose the bandwidth b for either Model-Based or Model-Free point prediction, we perform one-step-ahead prediction at several coordinates of the given random field data. To elaborate, consider a random field $Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_n}$ and suppose only subseries $Y_{\underline{t}_1}, Y_{\underline{t}_2}, \ldots, Y_{\underline{t}_k}$ has been observed where $k < n$. Let $\hat{Y}_{\underline{t}_{k+1}}$ denote the predicted value based on the data $Y_{\underline{t}_1}, \ldots, Y_{\underline{t}_k}$; this can be estimated by using either the Model-Based or Model-Free approaches as described in Sections 3 and 4 for some choice of b. However, since $Y_{\underline{t}_{k+1}}$ is known, the quality of the predictor can be assessed. So, for each value of b over a reasonable range, we calculate the sum of squared errors:

$$SSE(b) = \sum_{k=k_o}^{n-1} (\hat{Y}_{\underline{t}_{k+1}} - Y_{\underline{t}_{k+1}})^2 \tag{46}$$

here k_o should be big enough so that estimation is accurate, e.g., k_o can be of the order of $\sqrt{n}$. The cross-validated bandwidth choice would then be the b that minimizes $SSE(b)$. For the problem of selecting h_0 in the case of Model-Free point predictors, as in [21], our final choice is $h_0 = h^2$ where $h = b/n$. Note that an initial choice of h_0 (needed to perform cross-validation to determine the optimal bandwidth b) can be set by any plug-in rule as the effect of choosing an initial value of h_0 is minimal.

6. Model-Free vs. Model-Based Inference: empirical comparisons

The point prediction performance of the Model-Free and Model-Based predictors described above are empirically compared using simulated as well as real-life data. The Model-Based local constant and local linear methods are denoted as MB-LC and MB-LL respectively. Model-Based predictors MB-LC and MB-LL are described in Section 3. The Model-Free methods using local constant, local linear (Hansen) and local linear (Monotone) are denoted as MF-LC, MF-LLH, MF-LLM. Model-Free predictors are described in Section

4. Both fitted and predictive residuals as described in Sections 3 and 4 are used for point prediction and their performance as indicated by Mean Squared Error (MSE) is used to compare the various estimators.

Baseline comparisons: Besides the above MB and MF estimators we also provide three baselines estimators for comparison as described below:

- Model-Based estimation of $Y_{\underline{t}_{n+1}}$ involves nonparametric mean and variance estimation followed by estimating the conditional expectation $E(W_{\underline{t}_{n+1}}|\underline{W}_{\underline{t}_n})$ which involves calculating the coefficients of the 2-D AR model as given by Equation 6. In this case as a baseline we have included results for both the synthetic and real-life datasets using only local linear estimation of the mean i.e. in this case the L_2–optimal predictor of $Y_{\underline{t}_{n+1}}$ is given by:

$$\hat{Y}_{\underline{t}_{n+1}} = \mu(\underline{t}_{n+1}) \tag{47}$$

  Here $\mu(\underline{t}_{n+1})$ is calculated using local linear fitting based on Equations 23–30 (in this case regular and predictive fitting are the same as stated in Remark 2.2 in [9]). In Tables 1 and 2 this estimator is shown as **LL** .

- Model-Free estimation of $Y_{\underline{t}_{n+1}}$ involves nonparametric estimation of the first marginal distribution followed by estimating the autocovariance matrix $\hat{\Gamma}_n^{AR}$. In this case as a baseline we have included results using only local linear estimation of the uniformizing transformation as given by Equations 34–35 (Local Linear Hansen) and Algorithm 1 (Monotone Local Linear Distribution Estimation). For the same reasons as stated above regular and predictive fitting are also the same in this case. The L_2–optimal predictor of $Y_{\underline{t}_{n+1}}$ in this case is given by:

$$\hat{Y}_{\underline{t}_{n+1}} = \frac{1}{M}\sum_{i=1}^{M} D^{-1}_{\underline{t}_{n+1}}(u_i) \tag{48}$$

  Here $u_1, \ldots, u_M \sim U[0,1]$ where M is some large integer, U is the uniform distribution and $D_{\underline{t}_{n+1}}$ is estimated by using $\bar{D}^{LLH}_{\underline{t}_{n+1}}(\cdot)$ or $\bar{D}^{LLM}_{\underline{t}_{n+1}}(\cdot)$. In Tables 1 and 2 these estimators are shown as **LLH, LLM**.

The code for all algorithms used for the synthetic and real-life datasets as discussed in this paper can be found under https://github.com/srinjoyd/randomfields_pp.

6.1. Simulation: Additive model with stationary 2-D AR errors

Let a random field be generated using the 2-D AR process as below:

$$\begin{aligned} W_{\underline{t}} = W_{t_1,t_2} = 0.25W_{t_1-1,t_2-1} + 0.2W_{t_1-1,t_2+1} \\ - 0.05W_{t_1-2,t_2} + v_{t_1,t_2} \end{aligned} \tag{49}$$

Let this field be generated over the region defined by $0 \le t_1 \le n_1$ & $0 \le t_2 \le n_2$ where $n_1 = 101, n_2 = 101$. The NSHP limits are set from $(101, 101)$ to $(50, 50)$, this defines the region $E_{\underline{t},\underline{n}}$ as shown in Figure 2. The data $Y_{\underline{t}}$ is generated using the additive model in eq. (1) with trend specified as $\mu(\underline{t}) = \mu(t_1, t_2) = \sin(4\pi\frac{t_2-1}{n_2-1})$ where $0 \le t_1 \le n_1$ & $0 \le t_2 \le n_2$. Here v_{t_1,t_2} are i.i.d. $N(0, \tau^2)$ where $\tau = 0.1$. Let $t_1 = 50, t_2 = 50$ where point prediction is performed. Bandwidths for all Model-Based, Model-Free and baseline predictors are calculated using cross-validation as described in Section 5.

Results for point prediction using mean square error (MSE) over all MB and MF methods are shown in Table 1. A total of 100 realizations of the dataset were used for measuring point prediction performance. From this table it can be seen that MB-LL is the best point predictor. This is expected since the data was generated by a 2D AR model

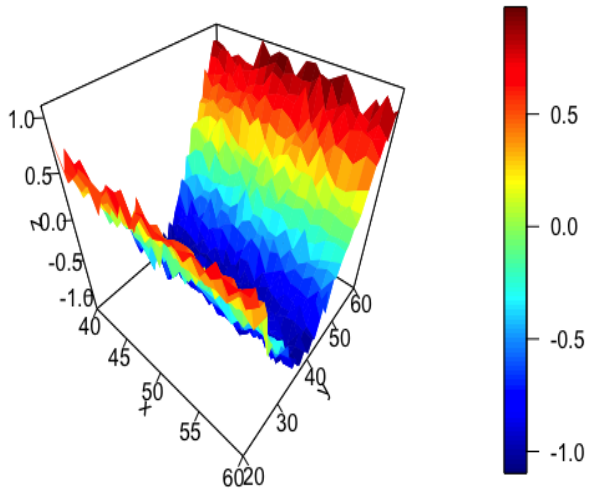

Figure 3. Linear trend for NSHP where prediction is performed (50, 50). Here the axes labeled x and y denote the coordinates of the random field and the axis labeled z denotes the corresponding value of the random field at those coordinates.

which is the same used in MB-LL prediction. In addition the estimation is performed at the boundary of the random field with a strong linear trend as shown in Figure 3 where LL regression is expected to perform the best. In addition it can be observed that MF-LLM performs the best among all MF point predictors and approaches the performance of MB-LL. This shows that monotonicity correction in the LLM distribution estimator has minimal effect on the center of the distribution that is used for point prediction.

In addition comparing the performance of MB-LL versus its corresponding baseline LL and that of MF-LLH and MF-LLM versus their corresponding baselines LLH and LLM respectively, show that the baseline estimators underperform as they do not take into account the spatial dependence present in the data either by estimating the coefficients of the 2D AR model (Equation 6 as in the Model-Based case) or by estimating the autocovariance matrix $\hat{\Gamma}_n^{AR}$ (Model-Free case).

Table 1. Point Prediction performance for 2-D AR dataset

Prediction Method	Residual Type	MSE
MB-LC	P	1.488e-02
	F	1.520e-02
MB-LL	P	1.393e-02
	F	1.400e-02
MF-LC	P	1.530e-02
	F	1.549e-02
MF-LLH	P	1.471e-02
	F	1.515e-02
MF-LLM	P	1.414e-02
	F	1.456e-02
LL	Not Applicable	1.488e-02
LLH	Not Applicable	1.651e-02
LLM	Not Applicable	1.455e-02

6.2. Real-life example: CIFAR images

The CIFAR-10 dataset [32] is used as a real-life example to compare the Model-Based and Model-Free prediction algorithms discussed before. The original CIFAR-10 dataset consists of 60000 32 by 32 color images in 10 classes, with 6000 images per class. We pick 100 images from the class "dog" where the original images have 3 RGB (red, green, blue) channels with discrete pixel values. We pick the R (red) channel of each image, and standardize these to generate a new real-valued dataset. Our final transformed dataset has 100 32 by 32 random fields. The NSHP limits are set from $(32, 32)$ to $(16, 16)$, this defines the region $E_{\underline{t},\underline{n}}$ as shown in Figure 2. Rest of the image is considered as occluded and their pixel values are not available for prediction. Sample images used for prediction are shown in Figure 4. Let $t_1 = 16, t_2 = 16$ where point prediction is performed. Bandwidths for all Model-Based, Model-Free and baseline predictors are calculated using cross-validation as described in Section 5.

Results for point prediction using mean square error (MSE) over all MB and MF methods are shown in Table 2. From this table it can be seen that MF-LLH and MF-LLM are the best point predictors. The superior performance of the Model-Free estimators as compared to their Model-Free counterparts can be attributed to the fact that the CIFAR-10 image data is not compatible with additive model as given by eq. (1). It can also be seen that unlike the synthetic 2D AR dataset the two best predictors MF-LLH and MF-LLM are closer in performance which is owing to lack of a linear trend at the point where prediction is performed.

In addition comparing the performance of MB-LL versus its corresponding baseline LL and that of MF-LLH and MF-LLM versus their corresponding baselines LLH and LLM respectively, show that the baseline estimators underperform as they do not take into account the spatial dependence present in the data either by estimating the coefficients of the 2D AR model (Equation 6 as in the Model-Based case) or by estimating the autocovariance matrix $\hat{\Gamma}_n^{AR}$ (Model-Free case).

Figure 4. Sample images from CIFAR-10 dataset with label dog (Note: Here full images are shown although only part of it is used for prediction.)

Table 2. Point Prediction performance for CIFAR-10 dataset

Prediction Method	Residual Type	MSE
MB-LC	P	1.98e-01
	F	2.20e-01
MB-LL	P	1.79e-01
	F	1.95e-01
MF-LC	P	1.79e-01
	F	2.12e-01
MF-LLH	P	1.60e-01
	F	1.89e-01
MF-LLM	P	1.64e-01
	F	1.70e-01
LL	Not Applicable	2.12e-01
LLH	Not Applicable	2.38e-01
LLM	Not Applicable	2.14e-01

7. Conclusions and Future Work

In this paper we investigate the problem of one-sided prediction over random fields that are stationary only across a limited part over their entire region of definition. For such locally stationary random fields we develop frameworks for point prediction using both a Model-Based approach which includes a coordinate changing trend and/or variance and also by using the Model-Free principle proposed by [21], [22]. We apply our algorithms to both synthetic data as well as a real-life dataset consisting of images from the CIFAR-10 dataset. In the latter case we obtain the best performance by using the Model-Free approach and thereby demonstrate the superiority of this technique versus the Model-Based case where an additive model is assumed arbitrarily for purposes of prediction. In future work we plan to investigate both Model-Based and Model-Free prediction using random fields with non-uniform spacing of data as well as consider extending our algorithms for estimating prediction intervals.

Author Contributions: Software and experiments—S.D. and Y.Z; conceptualization and writing—S.D. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: D.P. was partially supported by NSF grant DMS 19-14556. S.D. and Y.Z. did not receive any external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CIFAR-10 publicly available dataset was analyzed in this study. This data can be found in [32]. R code for the algorithms discussed in this paper can be found at https://github.com/srinjoyd/randomfields_pp.

Acknowledgments: This research was partially supported by NSF grant DMS 19-14556. The authors would like to acknowledge the Pacific Research Platform, NSF Project ACI-1541349 and Larry Smarr (PI, Calit2 at UCSD) for providing the computing infrastructure used in this project.

Conflicts of Interest: The authors declare no conflict of interest..

References

1. Priestley, M.B. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **1965**, pp. 204–237.
2. Priestley, M.B. *Non-linear and non-stationary time series analysis*; Academic Press, London, 1988.
3. Dahlhaus, R.; et al. Fitting time series models to nonstationary processes. *The Annals of Statistics* **1997**, *25*, 1–37.
4. Dahlhaus, R.; Rao, S.S. Statistical inference for time-varying ARCH processes. *The annals of Statistics* **2006**, *34*, 1075–1114.
5. Zhou, Z.; Wu, W.B. Local linear quantile estimation for nonstationary time series. *The Annals of Statistics* **2009**, *37*, 2696–2729.
6. Dahlhaus, R. Locally stationary processes. In *Handbook of statistics*; Rao, T.S.; et al., Eds.; Elsevier, 2012; Vol. 30, chapter 13, pp. 351–412.
7. Zhou, Z. Nonparametric specification for non-stationary time series regression. *Bernoulli* **2014**, *20*.
8. Kley, T.; Preuß, P.; Fryzlewicz, P. Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electronic Journal of Statistics* **2019**, *13*, 3710–3774.
9. Das, S.; Politis, D.N. Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association* **2021**, *116*, 919–934.
10. Lu, Z.; Tjøstheim, D. Nonparametric estimation of probability density functions for irregularly observed spatial data. *Journal of the American Statistical Association* **2014**, *109*, 1546–1564.
11. Fuglstad, G.A.; Simpson, D.; Lindgren, F.; Rue, H. Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics* **2015**, *14*, 505–531.
12. Kurisu, D. On nonparametric inference for spatial regression models under domain expanding and infill asymptotics. *Statistics & Probability Letters* **2019**, *154*, 108543.
13. Kurisu, D. Nonparametric regression for locally stationary random fields under stochastic sampling design. *Bernoulli* **2022**, *28*, 1250–1275.
14. Matsuda, Y.; Yajima, Y. Locally stationary spatio-temporal processes. *Japanese Journal of Statistics and Data Science* **2018**, *1*, 41–57.
15. Choi, B.; Politis, D.N. Modeling 2-D AR processes with various regions of support. *IEEE transactions on signal processing* **2007**, *55*, 1696–1707.
16. Mojiri, A.; Waghei, Y.; Nili-Sani, H.; Mohtashami Borzadaran, G.R. Non-stationary spatial autoregressive modeling for the prediction of lattice data. *Communications in Statistics-Simulation and Computation* **2021**, pp. 1–13.
17. Vaishali, D.; Ramesh, R.; Christaline, J.A. 2 D autoregressive model for texture analysis and synthesis. In Proceedings of the 2014 International Conference on Communication and Signal Processing. IEEE, 2014, pp. 1135–1139.
18. Hallin, M.; Lu, Z.; Tran, L.T. Local linear spatial regression. *Annals of Statistics* **2004**, *32*, 2469–2500.
19. El Machkouri, M.; Es-Sebaiy, K.; Ouassou, I. On local linear regression for strongly mixing random fields. *Journal of Multivariate Analysis* **2017**, *156*, 103–115.
20. Brockwell, P.J.; Davis, R.A. *Time series: theory and methods*, second ed.; Springer, New York, 1991.
21. Politis, D.N. Model-free model-fitting and predictive distributions. *Test* **2013**, *22*, 183–221.
22. Politis, D.N. *Model-Free Prediction and Regression*; Springer, New York, 2015.
23. Dudgeon, D.E.; Mersereau, R.M. *Multidimensional Digital Signal Processing Prentice-Hall Signal Processing Series*; Prentice-Hall, Englewood Cliffs, NJ, 1984.
24. Härdle, W.; Vieu, P. Kernel regression smoothing of time series. *Journal of Time Series Analysis* **1992**, *13*, 209–232.
25. Kim, T.Y.; Cox, D.D. Bandwidth selection in kernel smoothing of time series. *Journal of Time Series Analysis* **1996**, *17*, 49–63.
26. Li, Q.; Racine, J.S. *Nonparametric econometrics: theory and practice*; Princeton University Press, Princeton, 2007.
27. Fan, J.; Gijbels, I. *Local polynomial modelling and its applications: monographs on statistics and applied probability*; Vol. 66, CRC Press, Boca Raton, 1996.
28. Fan, J.; Yao, Q. *Nonlinear time series: nonparametric and parametric methods*; Springer, New York, 2007.
29. Fan, J. Local linear regression smoothers and their minimax efficiencies. *The annals of Statistics* **1993**, pp. 196–216.
30. Hansen, B.E. Nonparametric estimation of smooth conditional distributions. *Unpublished paper: Department of Economics, University of Wisconsin* **2004**.

31. Das, S.; Politis, D.N. Nonparametric estimation of the conditional distribution at regression boundary points. *The American Statistician* **2019**.
32. Krizhevsky, A.; Nair, V.; Hinton, G. Cifar-10 and cifar-100 datasets. *URl: https://www. cs. toronto. edu/kriz/cifar. html* **2009**, *6*, 1.