forecasting

MDPI

Article

Forecasting Lattice and Point Spatial Data: Comparison of Unilateral and Multilateral SAR Models

Carlo Grillenzoni

Dipartimento Di Culture Del Progetto, Universitá IUAV di Venezia, St Croce, n. 1957, 30135 Venezia, Italy; carlog@iuav.it

Abstract: Spatial auto-regressive (SAR) models are widely used in geosciences for data analysis; their main feature is the presence of weight (W) matrices, which define the neighboring relationships between the spatial units. The statistical properties of parameter and forecast estimates strongly depend on the structure of such matrices. The least squares (LS) method is the most flexible and can estimate systems of large dimensions; however, it is biased in the presence of multilateral (sparse) matrices. Instead, the unilateral specification of SAR models provides triangular weight matrices that allow consistent LS estimates and sequential prediction functions. These two properties are strictly related and depend on the linear and recursive nature of the system. In this paper, we show the better performance in out-of-sample forecasting of unilateral SAR (estimated with LS), compared to multilateral SAR (estimated with maximum likelihood, ML). This conclusion is supported by numerical simulations and applications to real geological data, both on regular lattices and irregularly distributed points.

Keywords: contiguity matrices; consistent estimation; spatial autoregression; spatial data; spatial forecasting

Citation: Grillenzoni, C. Forecasting Lattice and Point Spatial Data: Comparison of Unilateral and Multilateral SAR Models. *Forecasting* **2024**, *6*, 700–717. https://doi.org/10.3390/forecast6030036

Academic Editor: Sonia Leva

Received: 24 June 2024
Revised: 16 August 2024
Accepted: 19 August 2024
Published: 23 August 2024

1. Introduction

Spatial data are nowadays present in many social and natural phenomena. They are mainly available in two formats: on regular lattices (as digital images and remotely-sensed data), and irregular polygons and sparse points (as in territorial zoning and environmental surveys); see [1] for a detailed description. Spatial data are representable within geographic information systems (GIS) and statistical models may be built for practical purposes, such as filtering and prediction. In digital images, the main goals are denoising and sharpening (e.g., [2]), to perform image segmentation and classification. In polygonal data, the aims are structural analyses and spatial forecasting (e.g., [3]) to support planning decisions.

One of the main features of spatial data is auto-correlation (ACR), which arises from the interaction between random variables located ad different spatial units (pixels in images, and centroids in polygons). The general rule is that the minor is the distance of units, and the greater is the size of the ACR. This dependence must be represented in regression models in order to satisfy the basic assumption of uncorrelated residuals. This, in turn, is necessary to have unbiased and efficient estimates of the parameters and their standard errors; that is, for performing statistical inference and prediction without bias.

On the other hand, ACR is useful for forecasting the character under study in areas where it is not measured, especially when exogenous (X) covariates are not available. It follows that representing the ACR in spatial systems is one of the major concerns of statistical modeling. As in time series analysis, it may be accomplished by introducing into the equations suitable "lagged" terms, i.e., the value of the dependent variable in nearby units. This leads to the spatial auto-regressive (SAR) models, which resemble the AR schemes of time series and dynamical systems. However, while in time there is only a single direction (from past to present), in the plane there are almost infinite directions.

The specification of the models is twofold: unilateral or multilateral, depending on whether the linkage between the nearby units reflects or not a sequential ordering. As an example, in lattice data, one has the triangular and rook schemes in Figure 1a,b [1]; in the first, the dynamic is unidirectional since starting from the upper-left corner it follows the sequence of writing a text [4]. Instead, in Figure 1b the dynamic is multi-directional as each cell involves parts of the text that are not yet written [2]. Figure 1b,c show the analogous situations in polygonal data, where the position of each unit is defined by its center (geometric or political). In Figure 1d the link is multi-directional with four nearest neighbors (NN), whereas in Figure 1c it is in the north direction only.

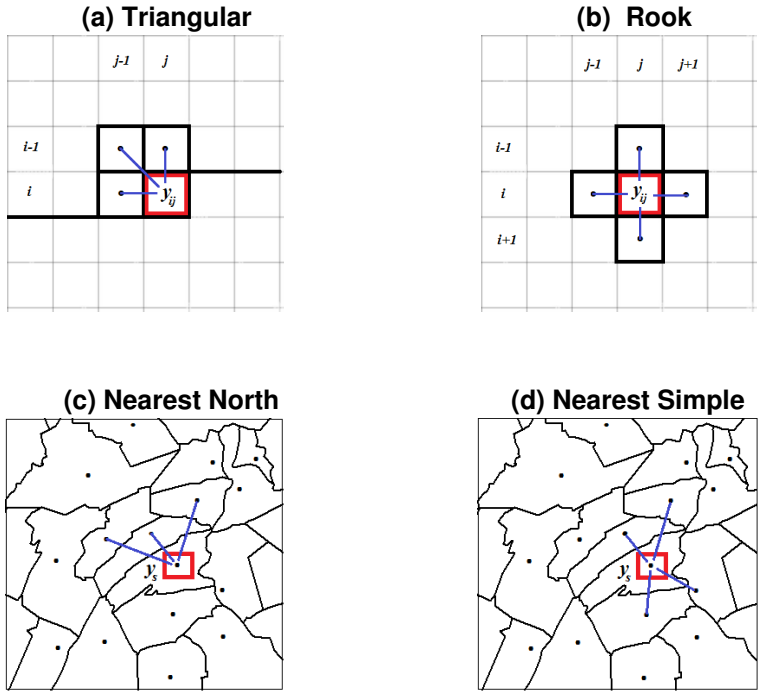

Figure 1. Examples of lattice (**a**,**b**) and polygonal (**c**,**d**) data, with unilateral (**a**,**c**) and multilateral (**b**,**d**) spatial dependence schemes, with respect to their nearest neighbors.

The statistical consequences of the two specifications of SAR models are important. Multidirection linkages violate the condition of independence between regressors and residuals, therefore, making least squares (LS) estimates biased and inconsistent. To solve this problem, many alternative estimators have been proposed, such as maximum likelihood (ML [1,5,6]), generalized method of moments (GM [7]), two-stage least squares [8] and, recently, indirect inference (II [9]). However, these methods are computationally demanding and use iterative algorithms of maximization which may not converge; involve the matrices of spatial contiguity, that must be inverted; require a condition of spatial stationarity (AR parameter $|\phi| < 1$) that may not be fulfilled.

There are also theoretical works that analyze the LS method in multilateral SAR: [10] showed that when the distance among random variables goes to infinity as the spatial dimension increases or the weight matrix converges to zero, then LS is consistent. Ref. [11] investigate the sensitivity of various LS estimators of the AR coefficient using Taylor expansions and find that it is moderate for moderate ACR. Ref. [12] under the null hypothesis, develop refined tests for first-order ACR based on Edgeworth expansion of the LS distribution. Ref. [13] developed an II estimator that implicitly corrects the bias of LS with a mechanism that involves data simulations from a related model; they find that finite sample performance is similar to that of ML. These works show the potential use of the LS method even in SAR models with Toepliz-type contiguity matrices, which are used in social and economic studies.

In this paper, we focus on unidirectional SAR for lattice and polygonal data and we show their ability to preserve the optimal properties of LS estimates even compared to the efficient ML and GM methods. This feature is important because the LS method is linear and can manage datasets of large dimensions, as it may avoid the direct use of the contiguity matrices. Further, unilateral SAR models have a recursive structure, which enables the application of the chain rule of forecasting also outside the observation area. Instead, multilateral SAR models have prediction functions that involve "forward" values and require iterations.

The spatial prediction of missing data (inside the perimeter of the observed area) has been treated by [14–17] for multilateral SAR models, and [18] in the lattice case. They show that the naive predictor based on the reduced form of SAR systems must be corrected with the best linear unbiased predictor (BLUP) of classical statistics. However, the recursive forecasts of unilateral SAR models do not need this BLUP correction and outperform the others. We show these results with Monte Carlo experiments on synthetic data and out-of-sample forecasting on real data of geosciences, such as digital elevation models (e.g., [19]) and the spatial diffusion of water isotopes (e.g., [20]).

The paper is organized as follows: Section 2 deals with lattice models, it reviews the conditions of identification and consistency and evaluates their forecasts on random surfaces and digital images; Section 3 deals with SAR models for polygonal and sparse point data, it compares the forecasts of unilateral and multilateral models with various algorithms.

2. Regular Lattice Data

Regular lattice data are mostly present in remote sensing and digital images. These data are in the form of rectangular arrays of the type $\boldsymbol{Y} = \{y_{ij}\}$, with $i = 1, 2 \dots n$, $j = 1, 2 \dots m$ the indexes of position, which may be transformed into latitude and longitude. The values y_{ij} are usually autocorrelated (e.g., clustered) and one of the main goals is to filter the array $\boldsymbol{Y}$ with its own values, to obtain interpolates and forecasts: $\hat{y}_{n+h,m+k}$. To accomplish this task, SAR modeling puts each cell in relation to the contiguous ones [1,21], such as $y_{ij} = g(y_{i\pm h,j\pm k}) + e_{ij}$, where $h, k = 1, 2 \dots p$ are spatial lags, p is the order of dependence and $\{e_{ij}\}$ is an unpredictable sequence.

In time series, the unidirectional (past-present) ordering of data is also called causal and found the analysis of causality between stochastic processes. In lattice data, the causal ordering can be established by following the lexicographic way of reading/writing a text, i.e., processing the cells of $\boldsymbol{Y}$ starting from the upper-left corner. Unilateral dynamics which satisfy this feature are the row-wise $y_{i,j-k}$, the half-plane $y_{i-h,j\pm k}$, $h > 0$ and the one-quadrant (or triangular) with $y_{i-h,j-k}$, $(h + k) > 0$ in Figure 1a [2,4]. Although causality is naturally related to the dynamic of events and aspects of human learning, digital filters of denoising and sharpening may follow non-causal models, such as the rook scheme in Figure 1b (e.g., [22]). However, the unidirectional approach remains the favorite solution, because it enjoys properties of sequentiality.

Under linear and unilateral constraints, the triangular SAR(p) model for y_{ij} is defined as follows

$$y_{ij} = \sum_{h=0}^{p} \sum_{\substack{k=0 \\ (h+k)>0}}^{p} \phi_{hk}\, y_{i-h,j-k} + e_{ij}, \qquad e_{ij} \sim \mathrm{IN}(0, \sigma_e^2), \tag{1}$$

with independent normal (IN) residuals. In statistics, the system (1) is usually estimated with the maximum likelihood (ML); this requires the vectorization of the data matrix $\boldsymbol{y} = \mathrm{vec}_j(\boldsymbol{Y})$, of length $N = n \times m$ and the inversion of auto-covariance matrices $\boldsymbol{\Gamma}_{\boldsymbol{y}}$ of size $N \times N$, see [3,23]. This solution is statistically efficient but computationally expensive and can be implemented only for moderate dimensions of the lattice field.

Instead, the LS method can be applied even for large values of n, m; rewriting the model (1) in regression form as

$$
\begin{aligned}
y_{ij} &= \boldsymbol{\phi}' \boldsymbol{x}_{ij} + e_{ij}, \\
\boldsymbol{x}_{ij} &= \left[y_{i,j-1} \dots y_{i-h,j-k} \dots y_{i-p,j-p} \right]', \\
\boldsymbol{\phi} &= \left[\phi_{01} \dots \phi_{hk} \dots \phi_{pp} \right]',
\end{aligned}
$$

the LS estimator becomes

$$
\begin{aligned}
\widehat{\boldsymbol{\phi}}_N &= \left(\sum_{i=p+1}^{n} \sum_{j=p+1}^{m} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}' \right)^{-1} \sum_{i=p+1}^{n} \sum_{j=p+1}^{m} \boldsymbol{x}_{ij} y_{ij}, \\
&= \boldsymbol{\phi} + \left(\sum_{i=p+1}^{n} \sum_{j=p+1}^{m} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}' \right)^{-1} \sum_{i=p+1}^{n} \sum_{j=p+1}^{m} \boldsymbol{x}_{ij} e_{ij}.
\end{aligned} \tag{2}
$$

The second equation follows from (2) by inserting $y = \boldsymbol{\phi}' \boldsymbol{x} + e$ and is useful for statistical analysis. The linearity of the algorithm (2) can be appreciated with respect to both $y_{ij} \Big| \boldsymbol{x}_{ij}$ and the recursive calculation of the sums, see [24]. Unlike the ML approach, it only involves the inversion of a matrix of size $(p+1)^2 - 1$, for any value of n, m. The LS estimator is unbiased and consistent for the unilateral model (1) because $\mathrm{E}(\boldsymbol{x}_{ij} e_{ij}) = \mathbf{0}$.

SAR(1). As an example, let $p = 1$, then the model (1) is $y_{ij} = (\phi_1 y_{i,j-1} + \phi_2 y_{i-1,j} + \phi_3 y_{i-1,j-1}) + e_{ij}$. Applying the formula of [25] to the term $y_{i,j-1}$, one obtains the moving average (MA) representation

$$
y_{i,j-1} = e_{i,j-1} + \sum_{\substack{h=1 \\ (h+l)<i}}^{i} \sum_{\substack{k=1 \\ (k+l)<j}}^{j-1} \sum_{l=0}^{\max(i,j-1)} \psi_{hkl} \, e_{i-h-l,j-1-k-l},
$$

from which $\mathrm{E}(y_{i,j-1} e_{ij}) = 0$. Similarly, $\mathrm{E}(\boldsymbol{x}_{ij} e_{ij}) = \mathbf{0}$ because all entries of $\boldsymbol{x}_{ij} = [y_{i,j-1}, y_{i-1,j}, y_{i-1,j-1}]'$ do not depend on e_{ij}; hence, the LS estimator (2) is unbiased. This result does not apply to multilateral SAR models, because the MA decomposition would involve $e_{i+h,j+k}$; for the rook scheme in Figure 1b the ML estimator is necessary. Finally, as in a time series, the consistency of LS does not need the stability of the model (1), e.g., $|\phi_{hk}/3| < 1$; rather, it improves with the signal-to-noise ratio σ_y^2 / σ_e^2 (e.g., [26]).

SAR(p,q,r). Empirical models often have a subset structure, i.e., have missing lagged terms or they aggregate $y_{i-k,j-h}$ under a common parameter. The proper definition of such models is SAR(p, q, r), where p = maximum lag of regressors, q = the number of spatial units and r = the number of parameters. Two parsimonious models that will be extensively applied in the paper are the triangular SAR(1,3,1) and the rook SAR(1,4,1) with drift α, defined as

$$
y_{ij} = \alpha + \phi \left(y_{i-1,j} + y_{i,j-1} + y_{i-1,j-1} \right) / 3 + e_{ij}, \qquad y_{i0} = y_{0j} = 0, \tag{3}
$$

$$
y_{ij} = \alpha + \phi \left(y_{i,j-1} + y_{i,j+1} + y_{i-1,j} + y_{i+1,j} \right) / 4 + e_{ij}, \qquad y_{i,m+1} = y_{n+1,j} = 0, \tag{4}
$$

see Figure 1a,b and [1]. These models have almost the same number of regressors and can be written as $y_{ij} = \alpha + \phi \, \overline{y}_{ij} + e_{ij}$, where $\overline{y}$ is a local average. The condition of stability in (3) is $|\phi/3| < 1$, because it admits a MA decomposition with weights $\psi_{hkl} \propto (\phi/3)^{(h+k+l)}$ which converge to 0, see [2]. By analogy, the model (4) is stable if $|\phi/4| < 1$, and this is a necessary condition for the convergence of its ML estimator (see Section 2.1).

2.1. The Vector Form

Previous representations are termed raster in GIS software. When the lattice dimension is moderate (e.g., $n \times m < 1000$), it may be useful to pass at the vector form. Refs. [18,23,25] consider complex forms of vectorization for ML estimation. We follow the spatial econometric approach which vectorizes the data matrix by columns and builds the contiguity

matrix of all cells, see [6]. Thus, let $\boldsymbol{y} = \text{vec}_j(\boldsymbol{Y})$ of size $N = n \times m$; the first-order vector model with drift becomes

$$\boldsymbol{y} = \alpha \mathbf{1}_N + \phi\, \boldsymbol{W}\boldsymbol{y} + \boldsymbol{e}, \qquad \boldsymbol{e} \sim \text{N}(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_N), \tag{5}$$

where $\mathbf{1}_N$ is a unit vector of length N and $\boldsymbol{W}$ is an $N \times N$ contiguity matrix. Its structure depends on the order p and the marginal values: the models (3), (4) have matrices $\boldsymbol{W}$ with sub-diagonals < 1 and sparse 1 on the main diagonal, in correspondence of the border values $y_{i1}, y_{1j}, y_{im}, y_{nj}$. For $n = 4$, $m = 7$ one has the arrays in Figure 2, see [24]; the distinctive feature of unilateral models is that $\boldsymbol{W}$ is always triangular, this provides recursivity to the system.

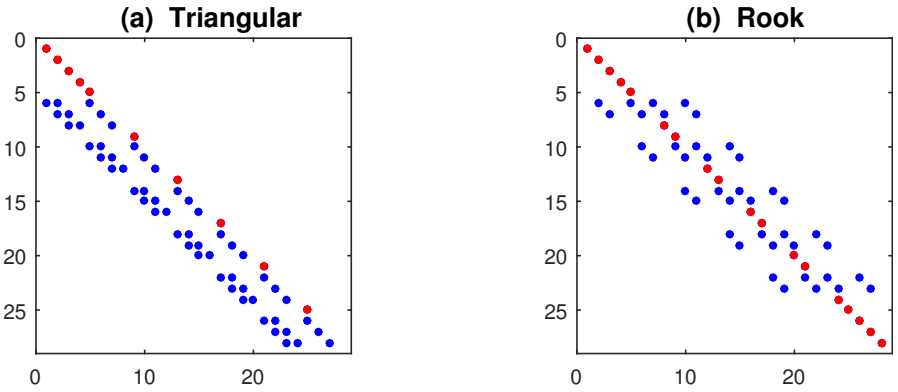

Figure 2. Contiguity matrices $\boldsymbol{W}$ of the models (3), (4) on a lattice of size $n = 4$, $m = 7$. Red dots have value 1 and correspond to the border values; while the blue have value $1/3$ in Panel (**a**) and $1/4$ in Panel (**b**); see [24].

The arrays in Figure 2 are obtained by solving the static system $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{y}$, where m determines the number of blocks of size $n \times n$, with matrices $\boldsymbol{I}_n$ at the corners. However, when $\boldsymbol{W}$ is inserted in the dynamic model (5), its diagonal elements $w_{ll} = 1$ must be replaced by 0, obtaining $\boldsymbol{W}_0$. Each sub-diagonal corresponds to a specific regressor $y_{i-h,j-k}$ of models (3), (4); thus, the array in Figure 2a can be decomposed as $\boldsymbol{W}_0 = (\boldsymbol{W}_1 + \boldsymbol{W}_2 + \boldsymbol{W}_3)$, providing an SAR(3) model. In simulations, the data-generation of $\{y_l\}$ proceeds by rows, starting from an initial $\boldsymbol{y} = \boldsymbol{y}_0$, as

$$y_l = \alpha + \phi\, \boldsymbol{w}_l' \boldsymbol{y} + e_l, \qquad l = 1, 2 \ldots N,$$

where $\boldsymbol{w}_l'$ is the l-th row of $\boldsymbol{W}_0$. Unilateral processes are insensitive to $\boldsymbol{y}_0$, since $\boldsymbol{W}_0$ triangular fills $\boldsymbol{y}$ recursively; instead, the model (4), with the forward terms $y_{i,j+1}, y_{i+1,j}$, needs a non-null initial vector, e.g., $\boldsymbol{y}_0 = \alpha \mathbf{1}_N$.

Using matrix algebra, one may obtain the reduced (MA) form of the system (5)

$$\begin{aligned} \left(\boldsymbol{I}_N - \phi \boldsymbol{W}\right) \boldsymbol{y} &= \alpha \mathbf{1}_N + \boldsymbol{e}, \\ \boldsymbol{y} &= \left(\boldsymbol{I}_N - \phi \boldsymbol{W}\right)^{-1} \left(\alpha \mathbf{1}_N + \boldsymbol{e}\right), \end{aligned} \tag{6}$$

this provides an automatic way to generate SAR data, independent of the initial/border condition $\boldsymbol{y}_0$. The reduced form (6) is fundamental for ML estimation because, under Gaussianity, the log-likelihood function takes the form

$$\begin{aligned} \ell(\alpha, \phi, \sigma) &\propto -N \log(\sigma) + \log\left[\det\left(\boldsymbol{I}_N - \phi \boldsymbol{W}\right)\right] - \left(\boldsymbol{e}'\boldsymbol{e}\right)/2\sigma_e^2, \\ \boldsymbol{e}(\alpha, \phi) &= \left(\boldsymbol{I}_N - \phi \boldsymbol{W}\right)\boldsymbol{y} - \alpha \mathbf{1}_N, \end{aligned} \tag{7}$$

The maximization of Equation (7) proceeds iteratively and requires $\det(\cdot) > 0$, i.e., the stability condition $|\phi| < 1$, to converge, see [6]. The ML method is compulsory for the

rook model (4), but may also be applied to the triangular (3); although LS is preferable (see Table 1).

Table 1. Performance of LS (8) and ML (7) estimators applied to triangular and rook SAR models (3), (4) with $n = m = 32$: R-bias = relative bias, R-RMSE = relative root mean squared error, N-test = p-values of the normality test. The bold character indicates the best methods (LS or ML) over model classes (Triang and Rook).

Method	**LS**	**Triang.**	**LS**	**Rook**	**ML**	**Triang.**	**ML**	**Rook**	$\boldsymbol{\alpha_0 = 1}$
Stat.	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\phi}}$	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\phi}}$	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\phi}}$	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\phi}}$	$\boldsymbol{\phi_0}$
R-bias	**0.015**	**0.013**	0.279	0.325	0.295	0.324	**0.014**	**0.014**	
R-RMSE	**0.078**	**0.078**	0.289	0.334	0.302	0.329	**0.062**	**0.066**	0.5
N-test	>0.5	0.415	0.456	0.267	0.319	0.065	0.221	0.138	
R-bias	**0.021**	**0.007**	0.330	0.134	0.372	0.148	**0.023**	**0.008**	
R-RMSE	**0.088**	**0.032**	0.339	0.136	0.384	0.153	**0.074**	**0.028**	0.75
N-test	>0.5	>0.5	0.413	0.135	>0.5	>0.5	0.192	0.052	
R-bias	**0.007**	**0.0003**	**0.004**	**0.0001**	0.049	0.003	0.023	0.0002	
R-RMSE	**0.053**	**0.003**	0.054	0.0004	0.067	0.004	**0.043**	**0.0002**	1
N-test	>0.5	>0.5	0.412	0.119	0.126	0.001	0.092	0.001	
R-bias	**0.0057**	**0.0001**	**0.003**	**0.0000**	0.510	0.024	2.270	0.023	
R-RMSE	**0.047**	**0.002**	**0.035**	**0.0001**	0.512	0.024	2.286	0.024	1.025
N-test	>0.5	>0.5	>0.5	0.001	>0.5	0.001	0.450	0.001	

In the vector representation (5), the LS estimator of parameters $\boldsymbol{\delta} = [\alpha, \phi]'$ of the models (3), (4) may use the entire sample $\boldsymbol{Z} = \left[\mathbf{1}_N, \boldsymbol{W}\boldsymbol{y}\right]$ as

$$\begin{aligned} \widehat{\boldsymbol{\delta}}_N &= \left(\boldsymbol{Z}\,\boldsymbol{Z}'\right)^{-1}\boldsymbol{Z}\,\boldsymbol{y}, \\ \widehat{\boldsymbol{\Sigma}}_N &= \left(\boldsymbol{Z}\,\boldsymbol{Z}'\right)^{-1}\hat{\sigma}_e^2, \end{aligned} \tag{8}$$

improving the estimator (2) based on $N - p(n+m)$ observations. The matrix $\widehat{\boldsymbol{\Sigma}}_N$ provides the dispersion of estimates and if $\boldsymbol{W}$ is triangular the method is consistent; a formal analysis is in [24].

Spatial Forecasting. Prediction is one of the central goals of SAR modeling; typical examples with lattice data are restoring parts of remote-sensing images hidden by clouds or extrapolating their value outside the observed range. However, existing literature has mostly concerned with in-sample filtering and interpolation, focusing on techniques of image sharpening [22], robust denoising [2] and trend estimation [18], also in conjunction with non-parametric smoothing.

This paper aims to forecast data that are external to the measured perimeter $n \times m$, e.g., on the right-hand side as $\{y_{i,m+k}\}$ with $i, k > 0$ or, in vector form $\boldsymbol{y}_{N+l}$ on units placed beyond N. By defining $h, k, l = 1, 2 \ldots H, K, L$ the forward indexes, the forecast function depends on the SAR representations, see [15]. Here, we have the lattice form (3), (4), the vector AR (5) and the reduced MA (6); for the triangular model (3), the 3 predictors are

$$\hat{y}_{n+h,m+k} = \mathrm{E}\left(y_{n+h,m+k} \middle| y_{n-i+1,m-j+1}\right), \qquad h, k = 1, 2 \ldots H, K,$$

$$\hat{y}_{n+h,m+k} = \alpha + \phi\left(\hat{y}_{n+h,m+k-1} + \hat{y}_{n+h-1,m+k-1} + \hat{y}_{n+h-1,m+k}\right)/3, \tag{9}$$

$$\hat{y}_{N+l} = \alpha + \phi\,\boldsymbol{w}'_{N+l}\,\widehat{\boldsymbol{y}}, \qquad \widehat{\boldsymbol{y}}' = \left[y_{11}, y_{21} \ldots y_{nm}, \hat{y}_{N+1} \ldots \hat{y}_{N+l-1} \ldots 0\right], \tag{10}$$

$$\widehat{\boldsymbol{y}} = \left(\boldsymbol{I}_{N+L} - \phi\,\boldsymbol{W}_{N+L}\right)^{-1}\alpha\mathbf{1}_{N+L}, \qquad l = 1, 2 \ldots L, \tag{11}$$

where $\boldsymbol{w}'_{N+l}$ is the $(N+l)$-th row of the augmented matrix $\boldsymbol{W}_{N+L}$ and the vector $\widehat{\boldsymbol{y}}$ in (10) is updated with the forecasts. The predictor (11) is nearly automatic, but in the absence of

exogenous variables provides constant forecasts; on the other hand, the functions (9), (10) need border values to start. These can be set as $\hat{y}^{(0)}_{1,m+k} = \hat{\alpha}$ for all k, and may be improved by r-iterating the forecasts with the nearby values as $\hat{y}^{(r+1)}_{1,m+k} = \hat{y}^{(r)}_{2,m+k-1}$, etc. This approach can extensively be applied to the rook model (4), with regard to its forward elements $y_{i,j+1}, y_{i+1,j}$; however, the convergence of $\hat{y}^{(r)}_{n+h,m+k}$, as r increases, is not guaranteed and may give biased forecasts.

The variance of predictors is useful for building confidence intervals and testing hypotheses. A general expression can be obtained for the representations (6) and (11), such as

$$\mathrm{V}(\widehat{\boldsymbol{y}}) = \left(\boldsymbol{I}_{N+L} - \phi\, \boldsymbol{W}_{N+L}\right)^{-1} \left(\boldsymbol{I}_{N+L} - \phi\, \boldsymbol{W}'_{N+L}\right)^{-1} \sigma_e^2,$$

Finally, in order to compare the models' performance, a portion of the observed data, e.g., $y_{i,m-k}$, $k = 1 \dots K$ is excluded from the parameter estimation; then the forecasts of each model are computed $\hat{y}_{i,m-K+k}$ and are compared with the mean absolute percentage error (MAPE) statistic

$$\mathrm{MAPE} = \frac{1}{nK} \sum_{i=1}^{n} \sum_{j=m-K+1}^{m} \left| \frac{\hat{y}_{ij} - y_{ij}}{y_{ij}} \right|. \tag{12}$$

2.2. Simulations and Applications

We develop simulation experiments to test the performance of LS (8) and ML (7) estimators, applied to unilateral and multilateral SAR models. We consider the models (3), (4) with parameters $\alpha_0 = 1$, $\phi_0 = 0.5, 0.75, 1.0, 1.025$ and border values $(\alpha + e_{ij})$, i.e., $y_{00} = 0$, etc. Data are generated with the reduced form (6), with matrices as in Figure 2 with zero diagonal; 200 replications on a 32×32 lattice (N=1024 cells) are obtained with Normal disturbances and $\sigma_e^2 = 1$. Performance statistics are the relative bias $|\hat{\phi} - \phi_0|/\phi_0$, the relative root mean squared errors (the square root of $\mathrm{MSE}(\hat{\phi})/\phi_0^2$) and the p-values of the Normality test of [27]. Since the relative statistics of $\hat{\alpha}, \hat{\phi}$ do not depend on the parameter size, their value is averaged to provide a single indicator of performance.

The results are displayed in Table 1 and Figure 3; the main findings are as follows: LS (8) is uniformly better than ML (7) when the matrix $\boldsymbol{W}$ is triangular. Its efficiency improves as $\phi_0 \to 1$, as a consequence of the super-consistency property of LS estimates in SAR models (e.g., [26]). LS estimates in multilateral SAR models are biased for $\phi < 1$ but still benefit from the super-consistency. Further, unlike time series, the Normal distribution of LS estimates holds even in the presence of the unit root $\phi_0 = 1$ (see Figure 3b).

Regarding the ML method (7), we used the Matlab implementation of [6], which is computationally demanding. However, it significantly improves the estimates of multilateral models, providing good levels of unbiasedness and efficiency when $\phi_0 \ll 1$. On the other hand, it requires conditions of stationarity and its performance in triangular models is disappointing, meaning that ML is mainly designed for sparse-weight matrices. In conclusion, we can state that LS is suitable for unilateral models, whereas ML must be used in multilateral SAR.

We also carry out two applications to test the performance of SAR models in out-of-sample forecasting of lattice data. In comparing predictions, one cannot proceed as in Table 1, because data generated by a certain SAR model will only be better predicted by that model. Hence, we consider data from external sources.

Random Surface. The first application considers rough random surfaces [28]; these models are used in physics to study electromagnetic, fluid and plasma phenomena. We follow an approach based on a fast Fourier transform, where a matrix of random numbers u_{ij} is convolved with a Gaussian filter to achieve a certain spatial autocorrelation. The resulting surfaces $g(u_{ij})$ are random but smooth and one may add further noise to make them similar to the real data of geostatistics $y_{ij} = g(u_{ij}) + e_{ij}$.

Figure 4a,b show an example of $g(u_{ij})$ and y_{ij} obtained with two Gaussian noises on a 32×32 lattice; the computational details are in [24]. The goal is to forecast the last $K = 7$

columns (about 22%) of the surface $\boldsymbol{Y}$ with the algorithms (9)–(11). Table 2 reports the in-sample parameter estimates (with LS and ML methods) of the models (3), (4) and their MAPE statistics (12).

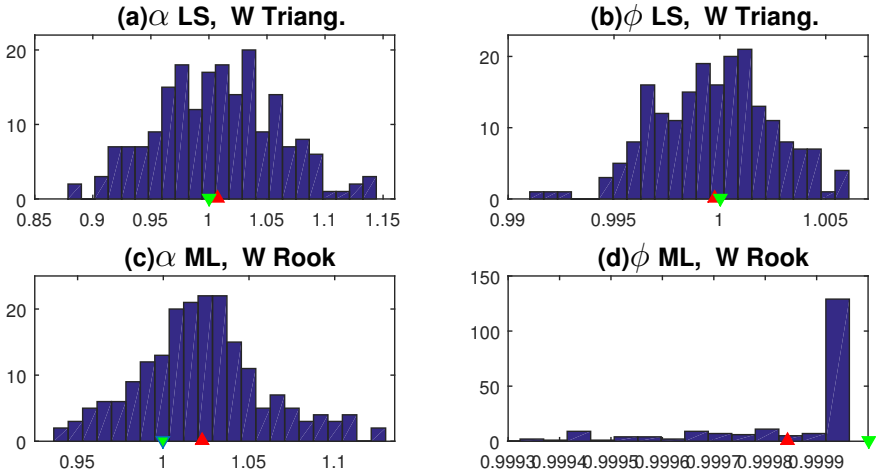

Figure 3. Distribution of LS and ML estimates in 200 replications of the SAR models (3), (4) with α_0=1, ϕ_0=1 (green), and mean values of $\hat{\alpha}, \hat{\phi}$ (red): Panels (**a**,**b**) LS (8) of model (3); Panels (**c**,**d**) ML (7) of model (4).

A triangular SAR(1,3,3) (with three AR parameters) is also fitted to the data, but it does not improve the forecasts of the constrained (3). However, the best performance is provided by (3) with ML estimates. The rook (4) performs poorly with the (biased) LS estimates, whereas it significantly improves with the ML ones; anyway, the path of the predicted surface in Figure 4d remains disappointing. Ref. [24] provides further results; on 10 replications of the experiment, the mean value of MAPE of (3)-LS is − 37% smaller than that of (4)-ML.

Table 2. Parameter estimates (with T-statistics in parentheses) of SAR models (3), (4) on the surface $\boldsymbol{Y}$(:,1:25) in Figure 4b, and forecast statistics on the data $\boldsymbol{Y}$(:,26:32).

Model	Estim.	$\hat{\alpha}$	$\hat{\boldsymbol{\phi}}_1$	$\hat{\boldsymbol{\phi}}_2$	$\hat{\boldsymbol{\phi}}_3$	R^2	MAPE
Triang. (3)	LS	0.096 (4.45)	0.893 (40.3)	.	.	0.686	0.143
Triang. (3)	ML	0.247 (15.2)	0.739 (45.1)	.	.	0.665	**0.125**
SAR(1,3,3)	LS	0.094 (4.42)	0.448 (13.3)	0.097 (2.69)	0.352 (10.1)	0.701	0.140
Rook (4)	LS	−0.008 (−0.39)	1.008 (50.6)	.	.	0.788	0.864
Rook (4)	ML	0.198 (16.3)	0.795 (64.9)	.	.	0.753	0.224

Remote Sensing. The second application regards a real case study; we consider high-resolution elevation data obtained with aerial laser scanners. The sample array $\boldsymbol{Y}$ comes from USGS [19] and has 340 × 455 pixels (see Figure 5); this would yield a vector SAR model with 154,700 rows. Further, the terrain morphology may require the inclusion of spatial trend components (e.g., [18]), such as bivariate polynomials $g_d(i,j)$ of degree $d \geq 1$. Given the numerical issues implied by computing $\boldsymbol{W}$, we only perform analyses with the LS estimator (2); for a model (3) with a quadratic trend, the forecasting function (9) becomes

$$\hat{y}_{i+h,j+k} = \alpha + g_2(i+h, j+k) + \phi\left(\hat{y}_{i+h-1,j+k-1} + \hat{y}_{i+h-1,j+k} + \hat{y}_{i+h,j+k-1}\right)/3,$$
$$g_2(i+h, j+k) = \beta_1(i+h) + \beta_2(j+k) + \beta_3\,(i+h)^2 + \beta_4(j+k)^2 + \beta_5\,(i+h)(j+k),$$

with $h, k = 1, 2 \ldots H, K$, which can be sequentially managed. The AR part of the rook model involves forward values in the lower-right side; however, these may be provided by the polynomial itself: $\tilde{y}_{i+h+1,j+k+1} = \hat{g}_2(i+h+1, j+k+1)$.

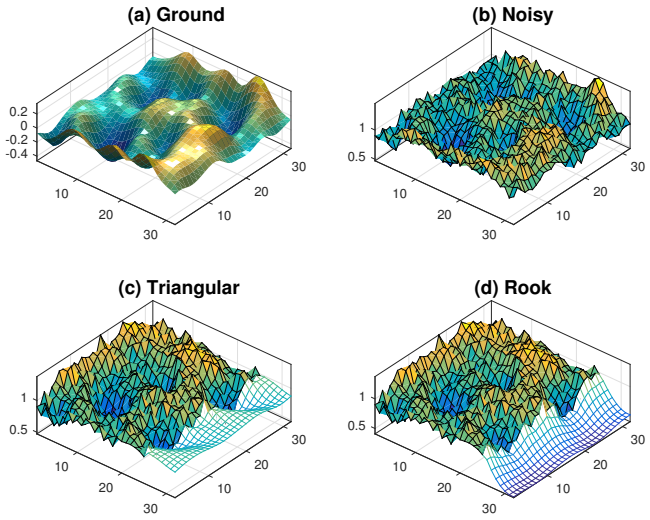

Figure 4. Panels (**a**,**b**): Simulation of a rough random surface $\boldsymbol{Y}$ of size $n = m = 32$. Panels (**c**,**d**): Out-of-sample forecasts of $\boldsymbol{Y}$(:,26:32) with models (3), (4) and predictors (9), (10).

In the prediction analysis, we left the right portion of 55 columns (about 12% of the array) as out-of-sample data to forecast. LS estimates of the parameters are reported in Table 3 and graphical results are displayed in Figure 5. Despite the better in-sample fitting (with $R^2 = 0.977$), the rook model has a disappointing performance both in terms of MAPE statistic and visual appearance (see Figure 5b). The reason is partly due to LS estimates of its parameters, with $\hat{\alpha} < 0, \hat{\phi} > 1$; however, adjusting their values reduces the MAPE but does not change the (uniform) path of the predicted surface.

Table 3. LS estimates of SAR models (3), (4) with bivariate polynomials (T-statistics are in parentheses) applied to USGS data [19]: $\boldsymbol{Y}$(:,1:400) in-sample, $\boldsymbol{Y}$(:,401:455) forecasts.

Model	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\sigma}^2_e, R^2$	**MAPE**
Triang. (3)	8.50	0.961	-0.024	-0.011	11×10^{-5}	43×10^{-7}	-88×10^{-7}	338.4	**0.848**
	(25.6)	(105)	(-10.6)	(-5.5)	(22.5)	(0.73)	(-2.1)	0.913	.
Rook (4)	-6.12	1.026	0.007	0.016	-58×10^{-6}	17×10^{-6}	-17×10^{-6}	88.15	1.362
	(-35.8)	(214)	(6.3)	(15.9)	(-22.8)	(5.5)	(-7.8)	0.977	.

In real-life applications, where small and inner portions of images must be restored, the spatial polynomials $g(i, j)$ may be fitted locally (around the missing pixels) or may be replaced by nonparametric smoothers (e.g., [18]). In any event, the role of the AR component remains fundamental, as the coefficient $\hat{\phi}$ is the most significant in both models of Table 3, with a statistic $T_{\hat{\phi}} > 100$.

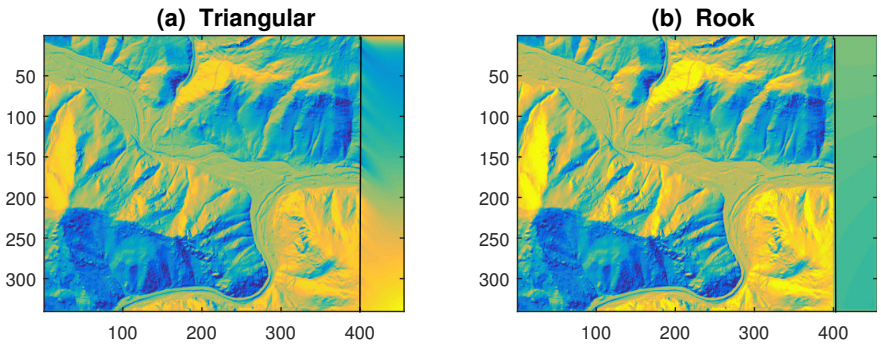

Figure 5. Application of SAR models (3), (4) with quadratic trend to [19] data: Out-of-sample forecasts of the image portion $\boldsymbol{Y}$(:,401:455), displayed in pseudocolor.

3. Point and Polygonal Data

When the data are irregularly distributed in space (either as point processes or polygonal areas), the representation has to change. The spatial process is now defined as $\{y_s\},\ s = 1, 2, \dots N$; where the index s also applies to the planar coordinates $\boldsymbol{x}'_s = [\, i_s, j_s \,]$ (latitude and longitude) of the polygon centers (geometric or political), see Figure 1c,d. These variables are irregularly distributed in the plane but have a fixed (non-stochastic) nature; moreover, they may influence the level of the process y_s, according to spatial trends. Thus, the SAR(1) model becomes

$$\begin{aligned} y_s &= \alpha + \phi\, y_{s,1} + \boldsymbol{\beta}' \boldsymbol{x}_s + e_s, \qquad s = 1, 2 \dots N, \\ \boldsymbol{y} &= \alpha \mathbf{1}_N + \phi \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \qquad \boldsymbol{e} \sim \mathrm{N}\big(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_N\big), \\ &= \big(\boldsymbol{I}_N - \phi \boldsymbol{W}\big)^{-1} \big(\alpha \mathbf{1}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}\big), \end{aligned} \tag{13}$$

where $y_{s,1}$ belongs to the unit which is closest to the s-th term, and the vector of regressors $\boldsymbol{x}'_s = [i_s, j_s]$ may include other covariates.

Unlike the previous section, the adjacency matrix $\boldsymbol{W}$ has an irregular structure, which depends on the rule of contiguity. The most common rule is to put each observation y_s in relation to its nearest neighbor (NN) term y_r, according to the Euclidean distance: $\min_r\{d_{sr} = [(i_s - i_r)^2 + (j_s - j_r)^2]^{1/2}\}$. Furthermore, under the unilateral constraint, the north/west (NW) $\rightarrow$ south/east (SE) direction may be followed as in the lattice case. However, polygonal and point data have not lexicographic order; hence, the unilateral constraint may simply be defined along the N-S direction. In this setting, it is useful to order the observations $\{y_s, \boldsymbol{x}_s\}$ according to the shortest distance from the northern border, i.e., according to the inverse of the latitude $1/i_s$. Such ordering may be denoted as (i_s) or simply (s), and the model (13) can be written in a sequential form as

$$y_{(s)} = \alpha + \phi\, y_{(s,-1)} + \boldsymbol{\beta}' \boldsymbol{x}_{(s)} + e_{(s)}, \qquad y_{(0,-1)} = 0, \tag{14}$$

where $y_{(s,-1)}$ is the north NN of $y_{(s)}$, while the simple NN may be denoted as $y_{(s,1)}$. As in the lattice case, the property $\mathrm{E}[y_{(s,-1)} e_{(s)}] = 0$ is the basis for unbiased LS estimation and forecasting.

As an illustration of NN matrices, we simulate $N = 30$ random points with $[i_s, j_s] \sim \mathrm{IU}(0,1)^2$ and consider $q = 3$ contiguous terms, see Figure 6. Under the north-south ordering of data, the entries of matrices are concentrated around the (null) main diagonal, and with the unilateral (north) constraint the array is lower triangular. Similarly, to obtain concentrated matrices, ref. [16] ordered the data according to the sum $(i_s + j_s)$, but they used the simple NN rule. The matrices $\boldsymbol{W}_k$ in Figure 6 provide SAR(3) models but averaged as $\overline{\boldsymbol{W}} = \frac{1}{3}\sum_{k=1}^{3} \boldsymbol{W}_k$, they yield constrained SAR(3,1). The analogous model to the lattice model (3) is then

$$y_{(s)} = \alpha + \phi \left(q^{-1} \sum_{k=1}^{q} y_{(s,-k)} \right) + \boldsymbol{\beta}' \boldsymbol{x}_{(s)} + e_{(s)}, \tag{15}$$

where $y_{(s,-k)}$ is the k-north NN of $y_{(s)}$. Instead, in the multilateral model of Figure 1d one has $q = 4$ and the simple NN terms $y_{(s,k)}$.

Unlike regular lattices, polygonal and point data are not equidistant; therefore, it is useful to consider spatially weighted averages in the model (15). A simple approach is based the inverse distance weighting (IDW)

$$\begin{aligned} y_{(s)} &= \alpha + \phi \left(\sum_{k=1}^{q} w_{(sk)}\, y_{(s,k)} \right) + \boldsymbol{\beta}' \boldsymbol{x}_{(s)} + e_{(s)}, \\ w_{(sk)} &= d_{(s,k)}^{-1} \Big/ \sum_{k=1}^{q} d_{(s,k)}^{-1}, \qquad d_{(s,k)} = \left[\big(i_{(s)} - i_{(s,k)}\big)^2 + \big(j_{(s)} - j_{(s,k)}\big)^2 \right]^{1/2}, \end{aligned} \tag{16}$$

where $[i_{(s,k)}, j_{(s,k)}]$ is the k-NN of the point $[i_{(s)}, j_{(s)}]$. Alternatively, since the inverse distances decay too fast, one can use the exponential weights $w_{(sk)} \propto \lambda^{d_{(s,k)}}$, with coefficient $0 < \lambda < 1$. The resulting model is nonlinear and requires complex estimators; however, the value of λ may be selected a priori and its effect may be evaluated on the LS-ML estimates of ϕ.

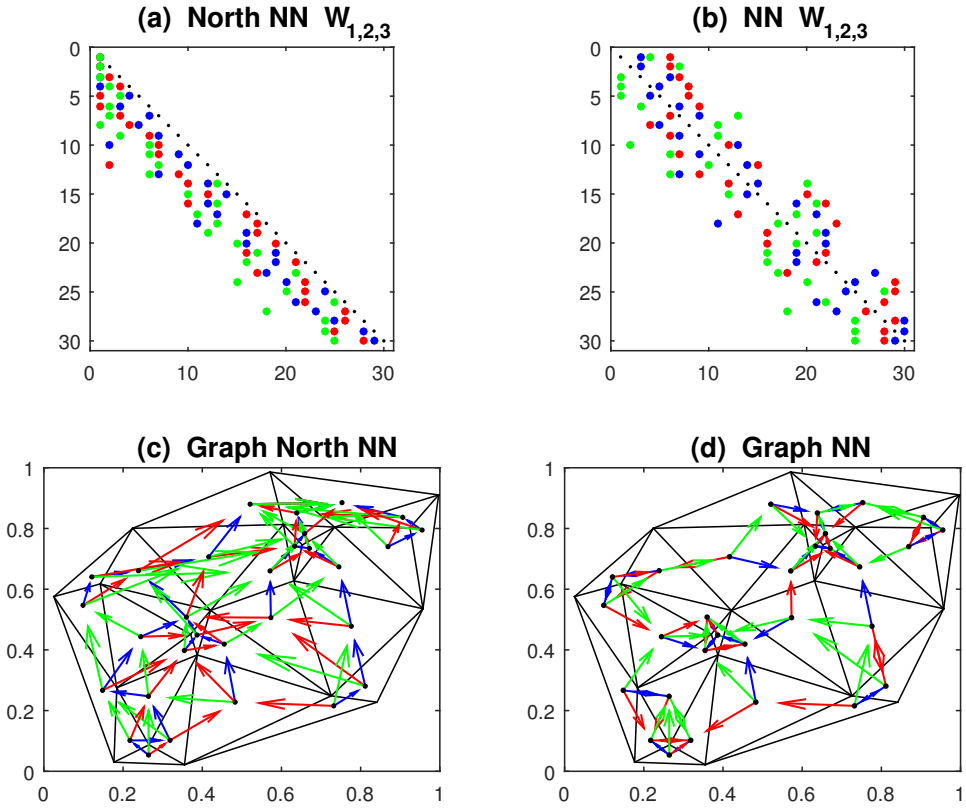

Figure 6. Contiguity matrices and planar graphs of $N = 30$ random centroids on the unit square [0, 1], based on nearest neighbors (NN). Panels: (**a**,**c**) North NN; (**b**,**d**) Simple NN. Matrices: $\boldsymbol{W}_1$ first NN (blue); $\boldsymbol{W}_2$ second NN (red); $\boldsymbol{W}_3$ third NN (green).

Models (15) and (16) admit a vector representation (13), with $\boldsymbol{W}$ triangular or sparse as in Figure 6. One may also include a lagged term on the exogenous part

$$\boldsymbol{y} = \alpha \mathbf{1}_N + \phi \boldsymbol{W}_{\mathrm{y}} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}_{\mathrm{x}} \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{e}, \tag{17}$$

this is the Durbin model [6], which is useful when $\boldsymbol{x}_{(s)}$ are autocorrelated, i.e., the points (i_s, j_s) have a non-random pattern. The contiguity matrices in Equation (17) may be different, e.g., $\boldsymbol{W}_{\mathrm{y}}$ is preferably triangular, while $\boldsymbol{W}_{\mathrm{x}}$ may be sparse, without affecting the consistency of LS estimates.

3.1. Estimation and Forecasting

The LS estimator of the models (13)–(17) can be obtained from Equation (2) by writing the models in regression form, as $y_{(s)} = \delta' z_{(s)} + e_{(s)}$, with $\delta = \left[\alpha, \phi, \boldsymbol{\beta}\right]'$ and $z_{(s)} = \left[1, \bar{y}_{(s)}, \boldsymbol{x}_{(s)}\right]'$, where $\bar{y}_{(s)}$ is an average of q NN terms. In the matrix form (8), the LS estimator of δ is given by

$$\begin{aligned} \widehat{\delta}_N &= \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{y}, \\ \boldsymbol{Z} &= \left[\mathbf{1}_N, \boldsymbol{W}\boldsymbol{y}, \boldsymbol{X}\right], \end{aligned} \tag{18}$$

which can be applied to the Durbin model (17) by including in $\boldsymbol{Z}$ the term $\boldsymbol{WX}$. In the case of triangular $\boldsymbol{W}$, the estimator (18) is unbiased because $\mathrm{E}\left[\bar{y}_{(s)}e_{(s)}\right] \propto \sum_k \mathrm{E}[y_{(s,-k)}e_{(s)}] = 0$ by the recursive computation of residuals. In general, under stationarity one can also prove the general result

$$\sqrt{N}\left(\widehat{\delta}_N - \delta\right) \quad \xrightarrow[N\to\infty]{\mathrm{D}} \quad \mathrm{N}\left[\mathbf{0}, \mathrm{E}\left(z_s z_s'\right)^{-1}\sigma_e^2\right], \tag{19}$$

see [24], where the convergence is in distribution (D).

For non-triangular $\boldsymbol{W}$, the LS is generally biased and one must use ML or generalized moments (GM, [7]). The GM solution arises by applying the instrumental variables (IV) method to the LS estimator (18); by using the matrix of instruments $\boldsymbol{M} = \left[\boldsymbol{X}, \boldsymbol{WX}, \boldsymbol{W}^2\boldsymbol{X} \ldots \boldsymbol{W}^k\boldsymbol{X}\right]$, the projection matrix $\boldsymbol{P} = \boldsymbol{M}\left(\boldsymbol{M}'\boldsymbol{M}\right)^{-1}\boldsymbol{M}'$ and the modified regressors $\boldsymbol{V} = \boldsymbol{PZ}$, the GM estimator is given by $\widetilde{\delta}_N = \left(\boldsymbol{V}'\boldsymbol{Z}\right)^{-1}\boldsymbol{V}'\boldsymbol{y}$. In the numerical section, we compare ML and GM estimators with simulation experiments.

In forecasting, we have L out-of-sample units with value $\boldsymbol{y}_L = \mathbf{0}$, whose coordinates and regressors $\boldsymbol{x}_l' = \left[1, i_l, j_l, x_{1l} \ldots x_{ml}\right]$ are known for all $l = 1, 2 \ldots L$. Their locations (i_l, j_l) may be inside or outside the observed region; in both cases, they are placed at the end of the data matrix $\boldsymbol{X}_{N+L}$. If the data are ordered in a certain direction (e.g., north-south with $i_{(s)} \geq i_{(s+1)}$), and all L units are outside the observed region, and in the same direction, then the weight matrix is nearly block-diagonal: $\boldsymbol{W}_{N+L} = \mathrm{diag}\left[\boldsymbol{W}_N, \boldsymbol{W}_L\right]$, where $\boldsymbol{W}_L$ is overlapped to $\boldsymbol{W}_N$ for the contiguity of L target units with the observed ones. In the other cases, it has a more complex structure, with triangular sub-matrices under the unilateral constraint

$$\boldsymbol{W}_{N+L} = \begin{bmatrix} \boldsymbol{W}_{NN} & \boldsymbol{O}_{NL} \\ \boldsymbol{W}_{LN} & \boldsymbol{W}_{LL} \end{bmatrix}$$

The forecasting function of $\boldsymbol{y}_L$ depends on the SAR representation used for $\boldsymbol{y}$; in the reduced (MA) form of (13), the fitted values and the forecasts are jointly computed as

$$\begin{aligned} \widetilde{\boldsymbol{y}}_L &= \mathrm{E}\left(\boldsymbol{y}_L \middle| \boldsymbol{X}_{N+L}, \boldsymbol{W}_{N+L}\right), \\ \widetilde{\boldsymbol{y}}_{N+L} &= \left(\boldsymbol{I}_{N+L} - \phi\boldsymbol{W}_{N+L}\right)^{-1}\left(\alpha\mathbf{1}_{N+L} + \boldsymbol{X}_{N+L}\boldsymbol{\beta}\right), \\ &= \left[\widetilde{\boldsymbol{y}}_N', \widetilde{\boldsymbol{y}}_L'\right]', \end{aligned} \tag{20}$$

where the joint matrix $\boldsymbol{W}_{N+L}$ has q-entries per row and may use non-uniform (IDW) weights. The solution (20) is nearly automatic, but in the absence of exogenous variables, it provides constant forecasts.

The second predictor comes from the structural (AR) representation (13); it is not automatic and must be managed sequentially by rows as

$$\begin{aligned} \widehat{\boldsymbol{y}}_L &= \mathrm{E}\left(\boldsymbol{y}_L \middle| \boldsymbol{y}_N, \boldsymbol{X}_{N+L}, \boldsymbol{W}_{N+L}\right), \\ \widehat{y}_{N+l} &= \alpha + \phi\, \boldsymbol{w}_{N+l}'\, \widehat{\boldsymbol{y}}_{N+l-1} + \boldsymbol{x}_{N+l}'\, \boldsymbol{\beta}, \qquad l = 1, 2 \ldots L, \\ \widehat{\boldsymbol{y}}_{N+l-1} &= \left[y_1, y_2, \ldots, y_N, \widehat{y}_{N+1} \ldots \widehat{y}_{N+l-1}, 0, \ldots, 0\right]', \end{aligned} \tag{21}$$

where $\boldsymbol{w}_l', \boldsymbol{x}_l'$ are the l-th rows of the matrices $\boldsymbol{W}_{N+L}, \boldsymbol{X}_{N+L}$. Note, that for non-triangular $\boldsymbol{W}$, Equation (21) involves missing values in the running vector $\widehat{\boldsymbol{y}}_{N+l-1}$. If these values are provided by Equation (20), then the vector of observations and forecasts at the l-th step becomes $\widehat{\boldsymbol{y}}_{N+l-1} = \left[y_1 \ldots y_N, \widehat{y}_{N+1} \ldots \widehat{y}_{N+l-1}, \widetilde{y}_{N+l} \ldots \widetilde{y}_{N+L}\right]'$ and in the end $l = L$, and will only contain the improved forecasts (21).

Refs. [14,17] also discussed solutions based on the best linear unbiased predictor (BLUP) of Goldberger. This approach arises from the conditional mean and variance of Gaussian random vectors:

$$\begin{aligned} \mathrm{E}(\boldsymbol{y}_L|\boldsymbol{y}_N) &= \boldsymbol{\mu}_L + \boldsymbol{\Sigma}_{LN}\boldsymbol{\Sigma}_{NN}^{-1}(\boldsymbol{y}_N - \boldsymbol{\mu}_N), \\ \widehat{\widetilde{\boldsymbol{y}}}_L &= \widetilde{\boldsymbol{y}}_L + \boldsymbol{\Sigma}_{LN}\boldsymbol{\Sigma}_{NN}^{-1}(\boldsymbol{y}_N - \widetilde{\boldsymbol{y}}_N), \end{aligned} \tag{22}$$

where $\widetilde{\boldsymbol{y}}_L$ is a subvector of the predictor in Equation (20) and $\boldsymbol{\Sigma}_{LN}, \boldsymbol{\Sigma}_{NN}$ come from the partition of the joint covariance matrix of $\boldsymbol{y}_{N+L}$, given by

$$\boldsymbol{\Sigma}_{N+L} = \left(\boldsymbol{I}_{N+L} - \phi\boldsymbol{W}_{N+L}\right)^{-1}\left(\boldsymbol{I}_{N+L} - \phi\boldsymbol{W}'_{N+L}\right)^{-1}\sigma_e^2 = \begin{bmatrix} \boldsymbol{\Sigma}_{NN} & \boldsymbol{\Sigma}_{NL} \\ \boldsymbol{\Sigma}_{LN} & \boldsymbol{\Sigma}_{LL} \end{bmatrix}. \tag{23}$$

The conditional variance $\mathrm{V}(\boldsymbol{y}_L|\boldsymbol{y}_N)$ of the partition of Gaussian vectors also provides the dispersion of the forecasts (22), such as

$$\mathrm{V}(\widehat{\widetilde{\boldsymbol{y}}}_L) = \boldsymbol{\Sigma}_{LL} - \boldsymbol{\Sigma}_{LN}\boldsymbol{\Sigma}_{NN}^{-1}\boldsymbol{\Sigma}_{NL},$$

this matrix requires the error variance σ_e^2, which can be estimated with the in-sample residuals $\widehat{\boldsymbol{e}}_N = (\boldsymbol{y}_N - \widehat{\boldsymbol{y}}_N)$ of Equations (20).

3.2. Simulations and Applications

To compare the various estimators, we perform simulation experiments on SAR models (13)–(16) defined on a random grid of N=150 points, uniformly distributed in the unit square: $[i_s, j_s] \sim \mathrm{IU}(0,1)^2$, as in Figure 6c. We generate the process $\{y_s\}$ with an exogenous input $x_s \sim \mathrm{IU}(0,2)$, Normal residuals $e_s \sim \mathrm{IN}(0,1)$, parameters $\alpha = 1$, $\phi = 0.65, 0.95, \beta = -1$, contiguity matrices $\boldsymbol{W}$ with q = 1, 3 lags and with north NN and simple NN links. Subsequently, M = 500 replications are fitted with LS, ML, GM estimators and relative biases, relative RMSE and p-values of the Normality test were computed for the coefficients $\hat{\alpha}, \hat{\phi}, \hat{\beta}$ and then averaged. Note, that relativization (e.g., $\mathrm{bias}(\hat{\phi}_N)/|\phi_0|$) allows to combine the statistics of three parameters to have a single indicator of performance.

The results are reported in Table 4, where "variable grid" means that at each replication the centroids change, and "IDW weights" refers to the model (16). The Durbin model includes a lagged term on the exogenous part, as $\beta_1\boldsymbol{x} + \beta_2\boldsymbol{W}\boldsymbol{x}$, with coefficients $[\beta_1, \beta_2] = [1, -0.65]$. The software used for ML and GM estimates is the Matlab package of [6], but it does not fit the Durbin model with the GM method. The main conclusions from Table 4 are that ML and GM methods do not improve the LS estimates in the case of unilateral (triangular) $\boldsymbol{W}$ matrices. However, they significantly outperform LS in the multidirectional (simple NN) case; in particular, the ML method is the best in terms of MSE. However, these results are not homogeneous regarding unbiasedness and efficiency, and the GM estimator may locally have a smaller bias.

Forecasting. As in the lattice case, when comparing the predictions of SAR models one must consider data generated from autonomous sources. An idea is to sample the surfaces in Figure 4 at random points $i_s \sim \mathrm{IU}[1,n]$, $j_s \sim \mathrm{IU}[1,m]$, and define y_s as the terrain height; however, the forecasting results were too favorable to the unilateral SAR model. We then use the real USGS image in Figure 5, with n = 340, m = 455; we sampled N = 150 pixels and withheld for forecasting L = 30 on the right side, see Figure 7a. We fitted the constrained SAR (15) with q = 3, $\boldsymbol{x}_s = [i_s, j_s]'$, $\boldsymbol{W}$ triangular and sparse, and forecast functions (20)–(22). The unilateral (west NN) is estimated with LS and the multilateral (simple NN) with ML.

The experiment is replicated M = 100 times (25 in each direction NS, SN, WE, EW, obtained by simply rotating the image $\boldsymbol{Y}$) and the average estimates are reported in Table 5. Both ML and LS estimates detect non-significant spatial coordinates (β_1, β_2), which means that the terrain level y_s does not have a (monotone) spatial trend. As in the lattice case, the fitting statistics σ_e, R^2 of the multilateral model are the best, but they are not confirmed out-of-sample. The predictor (21) moderately outperforms the naive (20), with a mild superiority (about 10% on average) of the unilateral model. Notice that the BLUP solution

(22) mainly improves the multilateral model, whereas slightly worsens the unilateral one. Figure 7b shows the results of the predictor (21) in a single replication.

Table 4. Results of the simulation experiments on the SAR models (13)–(16) with spatial lags q = 1–3 and single ϕ. The statistics are average values (over $\hat{\alpha}, \hat{\phi}, \hat{\beta}$) of relative biases, relative RMSE and p-value of N-test. Bold characters indicate the best results.

W	North NN			Simple NN				
Estimator	**LS**	**ML**	**GM**	**LS**	**ML**	**GM**	ϕ	**Lags**
Ave. Rbias	**0.0205**	0.1338	0.0235	0.1363	0.0287	**0.0216**	0.65	1
Ave. RRMSE	**0.1336**	0.2402	0.1831	0.1887	**0.1343**	0.1581		
Ave. Ntest	0.3218	0.2983	0.1675	0.1561	0.3401	0.3337		
Ave. Rbias	**0.0077**	0.1208	0.0098	0.1232	0.0400	**0.0075**	0.65	1
Ave. RRMSE	**0.1396**	0.2445	0.1924	0.1873	**0.1430**	0.1609	(Variable	
Ave. Ntest	0.3274	0.3554	0.2291	0.1637	0.2499	0.1626	grid)	
Ave. Rbias	0.0340	0.1330	**0.0217**	0.1324	**0.0259**	0.0376	0.65	3
Ave. RRMSE	**0.1586**	0.2416	0.2365	0.1998	**0.1390**	0.2035		
Ave. Ntest	0.2442	0.2612	0.0936	0.1811	0.2559	0.0960		
Ave. Rbias	0.0298	0.1289	**0.0228**	0.1487	**0.0208**	0.0343	0.65	3
Ave. RRMSE	**0.1518**	0.2360	0.2228	0.2088	**0.1326**	0.1963	(IDW	
Ave. Ntest	0.2530	0.2561	0.1673	0.1805	0.2900	0.1736	weights)	
Ave. Rbias	0.0312	0.0980	**0.0248**	0.0581	**0.0109**	0.0164	0.95	3
Ave. RRMSE	**0.1475**	0.2514	0.3190	0.1208	**0.1086**	0.1648		
Ave. Ntest	0.1516	0.1678	0.0010	0.2611	0.3033	0.1601		
Ave. Rbias	**0.0663**	0.3405	.	0.3333	**0.0704**	.	0.65	3
Ave. RRMSE	**0.2849**	0.4659	.	0.4141	**0.2764**	.	(Durbin	
Ave. Ntest	0.2584	0.3891	.	0.1286	0.1313	.	model)	

Table 5. Average values, over 100 replications, of parameter estimates (T-statistics are in parentheses), and MAPE forecast statistics of the SAR model (15) on random samples of pixels of the USGS image. Bold character indicates the best results.

Model	**Estim.**	$\bar{\alpha}$	$\bar{\phi}$	$\bar{\beta}_1$	$\bar{\beta}_2$	$\bar{\sigma}_e, \bar{R}^2$	**MAPE (20)**	**MAPE (21)**	**MAPE (22)**
Unilateral	LS (18)	68.4	0.543	0.012	−0.018	54.3	0.948	**0.815**	0.830
(west NN)		(3.0)	(5.1)	(0.21)	(−0.31)	0.213	.	.	.
Multilateral	ML (7)	76.5	0.492	0.005	−0.009	50.2	0.996	0.951	0.899
(simple NN)		(4.2)	(6.3)	(0.06)	(−0.19)	0.328	.	.	.

As a final application, we consider original point data concerning the measurement of stable isotopes of oxygen (δ^{18}O) and hydrogen (δ^{2}H) in the groundwater. Mapping isoscapes is useful for physical monitoring of the hydrological cycle and for anthropological and forensic investigations, e.g., regarding the path of people's movement. We consider the datasets of [29], recorded in South Korea in 2010, with N = 130 and [20], recorded in Mexico in 2007, with N = 234. We estimate the SAR model (15) with q = 3 for $y_s = \delta^{18}$O, δ^{2}H and $\boldsymbol{x}_s$ = latitude, longitude, to forecast about 20% of observations withheld from estimations. The unilateral constraints of the $\boldsymbol{W}$ matrix are south NN for Korea and west NN for Mexico; the results are in Table 6 and Figure 8. They confirm the better performance of unilateral SAR models with the forecast Function (21), especially when the data have significant autocorrelation (i.e., large $|\phi|$) and marked spatial trends (significant $\beta_{1,2}$). The reduction in the MAPE statistics in Table 6 ranges from −18% in Korea, to −38% in Mexico.

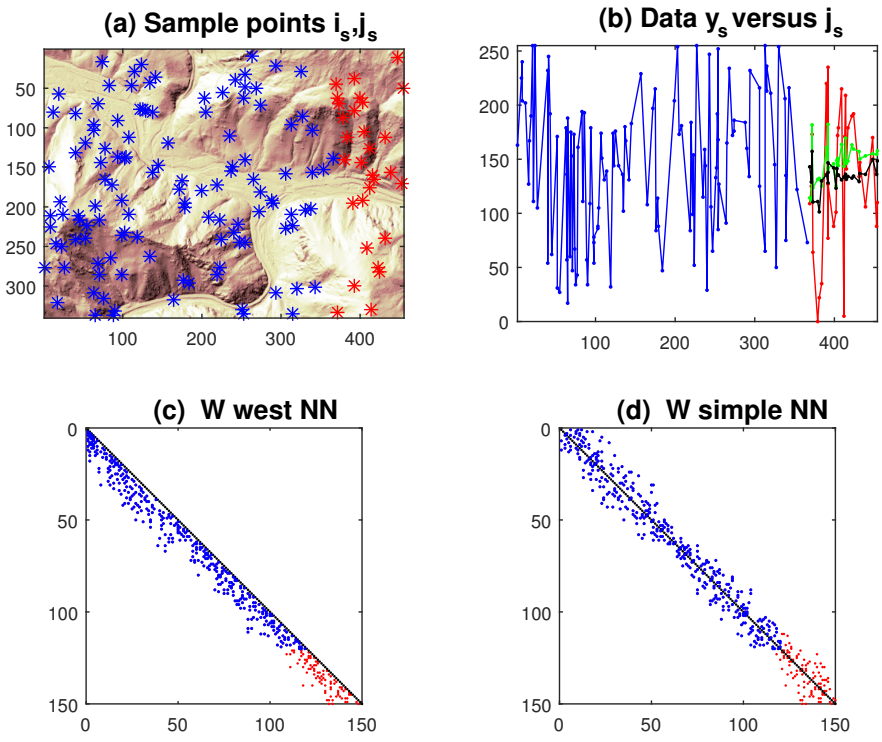

Figure 7. Graphical results of a single replication of the estimates in Table 5: (**a**) Random sampling of USGS surface (the points to forecast are in red); (**b**) Longitudinal display of data and forecasts (21) (unilateral = green, multilateral = black); (**c**) Contiguity matrix of the unilateral model with $q = 3$; (**d**) W matrix of the multilateral model (simple NN).

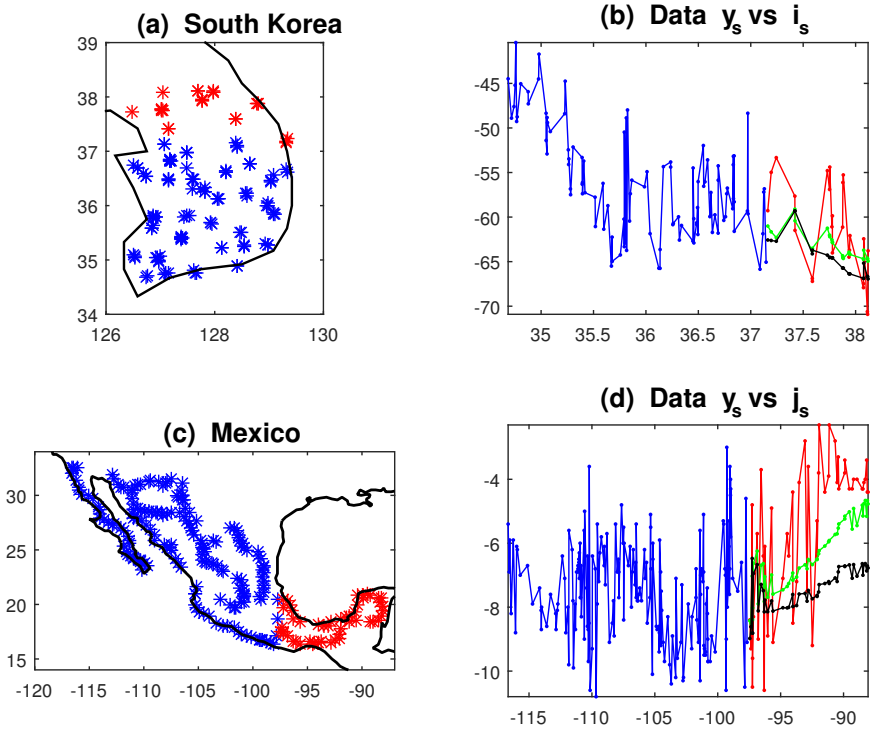

Figure 8. Water isotope data and SAR forecasts (22): Blue = in-sample data, Red = out-of-sample data, Green = forecasts with W triangular; Black = forecasts with W multilateral. Panels: (**a**) South-Korea sample locations; (**b**) Latitudinal view of Hydrogen isotope; (**c**) Mexico sample locations; (**d**) Longitudinal view of Oxygen isotope.

Table 6. Parameter estimates (with T-statistics in parentheses) and MAPE forecast statistics of the SAR model (15) with $q = 3$ applied to the water isotope data of South Korea δ^2H (first block) and Mexico δ^{18}O (second block). Bold character indicates the best results.

Model		$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_e, R^2$	MAPE (20)	MAPE (21)	MAPE (22)
Unilateral	LS (18)	123.7	0.337	−0.557	−2.51	4.28	0.069	**0.065**	0.065
(south NN)		(1.70)	(4.72)	(−1.04)	(−3.37)	0.478	.	.	.
Multilateral	ML (7)	97.9	0.622	−0.376	−1.98	3.53	0.080	0.079	0.079
(simple NN)		(1.69)	(9.46)	(−0.86)	(−3.41)	0.645	.	.	.
Unilateral	LS (18)	4.30	0.636	0.089	0.093	1.28	0.377	**0.369**	0.370
(west NN)		(1.69)	(8.17)	(3.19)	(2.72)	0.336	.	.	.
Multilateral	ML (7)	−0.85	0.479	0.053	0.101	1.26	0.600	0.597	0.588
(simple NN)		(−0.36)	(7.41)	(1.88)	(2.89)	0.351	.	.	.

4. Conclusions

In this paper, we have compared unilateral and multilateral SAR models in forecasting spatial data of both lattice and point types. SAR systems are natural extensions of classical autoregressions, their difference is in the treatment of the spatial dependence: while unilateral models choose a single direction in space, as in time series, multilateral models consider multiple directions. These approaches lead to different contiguity matrices: triangular and sparse, usually computed with the nearest-neighbor approach. While triangular SAR can be consistently estimated with least squares, multilateral SARs require maximum likelihood or method of moments; in the latter cases, numerical complexity increases with the dimension of the contiguity matrices. Instead, LS is not sensitive to the size of the lattice or the number of spatial units and can manage even large-scale systems; see the estimator (2). Simulation experiments on small-medium scale systems have shown that ML and GM are suitable for multilateral models, but LS is preferable for unilateral ones, especially in the presence of unstable roots ($|\phi| \geq 1$).

These features can be summarized in Table 7, which are vertically dependent; in fact, the concepts of triangularity, sequentiality, consistency of LS, recursive calculation, chain rule of forecasting, etc., are closely intertwined in the unilateral case.

Table 7. Summary of the main features of SAR models.

Item	Unilateral	Multilateral
Matrix $\boldsymbol{W}_y$	Triangular	Sparse
System	Sequential	Simultaneous
Estimator	LS	ML, GM
Sample size	Large	Moderate
AR root	$\|\phi\| \geq 1$	$\|\phi\| \ll 1$
Computation	Recursive	Iterative
Application	Forecasting	Analysis

For structural analyses, multilateral models are preferable because the interaction between spatial units does not occur in a single direction. However, structural analyses are mainly concerned with the dependence between the variables $y_s, \boldsymbol{x}_s$; now, in the Durbin model (17), the matrix $\boldsymbol{W}_x$ may be sparse (for detecting simultaneous relationships), without affecting the properties of unilateral SAR. The role of $\boldsymbol{W}_y$ in structural models is nearly ancillary: it has to remove the ACR of residuals e_s, to have consistent estimates of the standard errors of $\hat{\boldsymbol{\beta}}$. It is unlikely that a specific direction, i.e., a particular $\boldsymbol{W}_y$ triangular, may hinder this goal.

Things are different in forecasting, unilateral models with ordered data enable recursive calculations, which allow linear predictions (the chain rule of forecasting). Instead, multilateral models involve forward values which may be pre-estimated only with the (inefficient) reduced form (20) and require iterations. In the applications we have seen

that these models benefit from the BLUP improvement (22), but the performance of the unilateral models with the AR predictor (21) remains better, (see Tables 5 and 6). Regarding the practical usage of SAR point forecasts, we mention the possibility of using them as alternatives to non-parametric smoothers (such as Kriging and Kernels), which are unstable outside the observation perimeter. Further, to overcome the limits of unidirectionality, one can estimate unilateral models in various directions (e.g., NS, SN, WE, EW), and then combine their forecasts with weights proportional to their fitting statistics (e.g., R^2 or $1/\hat{\sigma}_e^2$). This solution is suitable for points within the investigated area, but can also be applied to external points, by using diagonal paths (e.g., NW-SE).

Funding: This research has not received fundings.

Data Availability Statement: The data are published in the references [19,20,29].

Conflicts of Interest: The author declares no conflict of interest.

References

1. Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 192–236. [CrossRef]
2. Bustos, O.; Ojeda, S.; Vallejos, R. Spatial ARMA models and its applications to image filtering. *Braz. J. Probab. Stat.* **2009**, *23*, 141–165. [CrossRef]
3. Barbiero, T.; Grillenzoni, C. Planning a novel regional methane network: Demand forecasting and economic evaluation. *Energy Convers. Manag. X* **2022**, *16*, 100294. [CrossRef]
4. Tjøstheim, D. Statistical Spatial Series Modelling II: Some further Results in Unilateral Processes. *Adv. Appl. Probab.* **1983**, *15*, 562–684. [CrossRef]
5. Anselin, L. Spatial Econometrics; In *A Companion to Theoretical Econometrics*; Baltagi, B.H., Ed.; Blackwell Publishing: Oxford, UK, 2003; pp. 310–330.
6. LeSage, J.P.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2009.
7. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Econom. Rev.* **1999**, *40*, 509–533. [CrossRef]
8. Lee, L.F. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econom. Rev.* **2003**, *22*, 307–335. [CrossRef]
9. Bao, Y.; Liu, X.; Yang, L. Indirect inference estimation for spatial autoregressions. *Econometrics* **2020**, *8*, 34. [CrossRef]
10. Lee, L.F. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econom. Theory* **2002**, *18*, 252–277. [CrossRef]
11. Liu, S.; Polasek, W.; Sellner, R. Sensitivity analysis of SAR estimators: a numerical approximation. *J. Stat. Comput. Simul.* **2012**, *82*, 325–342. [CrossRef]
12. Robinson, P.M.; Rossi, F. Refined tests for spatial correlation. *Econom. Theory* **2015**, *31*, 1249–1280. [CrossRef]
13. Kyriacou, M.; Phillips, P.C.B.; Rossi, F. Indirect inference in spatial autoregression. *Econom. J.* **2017**, *20*, 168–189. [CrossRef]
14. Goulard, M.; Laurent, T.; Thomas-Agnan, C. About predictions in spatial autoregressive models: optimal and almost optimal strategies. *Spat. Econ. Anal.* **2017**, *12*, 304–325. [CrossRef]
15. Kelejian, H.H., Prucha, I.R. The efficiencies of various predictors in spatial econometric models containing lags. *Reg. Sci. Urban Econ.* **2007**, *37*, 363–374. [CrossRef]
16. LeSage, J.P.; Pace R.K. Models for spatially dependent missing data. *J. Real Estate Financ. Econ.* **2004**, *29*, 233–254. [CrossRef]
17. LeSage J.P.; Pace R.K. Spatial econometric models, prediction. In *Encyclopedia of Geographic Information Science*; Shekhar, S., Xiong, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1095–1100.
18. Mojiri, A.; Waghei, Y.; Nili-Sani, H.R.; Mohtashami Borzadaran, H.R. Non-stationary spatial autoregressive modeling for the prediction of lattice data. *Commun. Stat.-Simul. Comput.* **2023**, *52*, 5714–5726 . [CrossRef]
19. USGS (2017), U.S. Geological Survey, Digital Elevation Models (DEMs). Avalilable online: https://www.sciencebase.gov/catalog/item/543e6b86e4b0fd76af69cf4c (accessed on 1 January 2023).
20. Moreiras, Reynaga, D.K.; Millaire, J.-F.; Chávez, Balderas, X.; Romxaxn, Berrelleza, J.A.; López, Lujxaxn, L.; Longstaffe, F.J. Building Mexican isoscapes: Oxygen and hydrogen isotope data of meteoric water. *Data Brief* **2021**, *36*, 107084. [CrossRef] [PubMed]
21. Fingleton, B. Spatial Autoregression. *Geograph. Anal.* **2009**, *41*, 385–391. [CrossRef]
22. Grillenzoni, C. Statistics for image sharpening. *Stat. Neerl.* **2008**, *62*, 173–192. [CrossRef]
23. Awang, N.; Shitan, M. Estimating the parameters of the second order spatial unilateral autoregressive model. *Int. J. Stat. Sci.* **2006**, *5*, 37–58.
24. Grillenzoni, C. Forecasting lattice and point spatial data: Comparison of unilateral and multilateral SAR models. *Preprints* **2024**, *1*, 2024080947. [CrossRef]
25. Basu, S.; Reinsel, G.C. Properties of the spatial nilateral first-order ARMA model. *Adv. Appl. Probab.* **1993**, *25*, 631–648. [CrossRef]

26. Bhattacharyya, B.B.; Richardson, G.D.; Franklin, L.A. Asymptotic inference for near unit roots in spatial autoregression. *Ann. Stat.* **1997**, *28*, 173–179. [CrossRef]
27. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **1987**, *55*, 163–172. [CrossRef]
28. Sheffield, S. What is a random surface? In Proceedings of the International Congress of Mathematicians 2022, Virtual , 6–14 July 2022; International Mathematical Union, EMS Press: Philadelphia, PA, USA, 2022.
29. Gautam, M.K.; Song, B.-Y.; Shin, W.-J.; Bong, Y.-S.; Lee, K.-S. Datasets for spatial variation of O and H isotopes in waters and hair across South Korea. *Data Brief* **2020**, *30*, 105666. [CrossRef] [PubMed]