

Article

Not peer-reviewed version

---

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

---

[Carlo Grillenzoni](#)\*

Posted Date: 13 August 2024

doi: 10.20944/preprints202408.0947.v1

Keywords: Contiguity matrices; Consistent estimation; Spatial autoregression; Spatial data; Spatial forecasting.



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

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

Carlo Grillenzoni 

Università IUAV di Venezia, St Croce, n. 1957; 30135 Venezia, Italy

**Abstract:** Spatial auto-regressive (SAR) models are widely used in geosciences for spatial analyses; 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 not consistent in the presence of multilateral (sparse) matrices. Instead, the unilateral specification of SAR models provides triangular weight matrices which allow good statistical properties to LS and enable the implementation of sequential prediction functions. In this paper we show the better performance in out-of-sample forecasting of unilateral SAR and LS with respect to multilateral and maximum likelihood (ML) methods. This conclusion is supported by extensive numerical simulations and applications to real geological data, both on regular lattice and irregular polygonal form.

**Keywords:** Contiguity matrices; Consistent estimation; Spatial autoregression; Spatial data; Spatial forecasting

## 1. Introduction

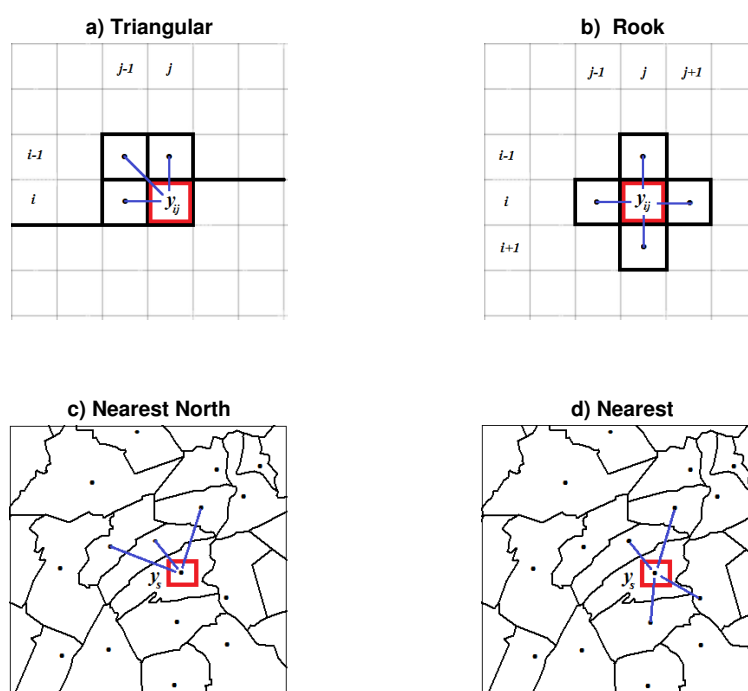
Spatial data are nowadays present in many social and natural phenomena. They are mainly available in two formats: on regular lattices (as digital images and remotely-sensed data), and irregular polygons and sparse points (as in zonings and geological surveys). These data are representable on maps by geographic information systems (GIS), and statistical models may be developed for practical purposes; such as filtering, prediction and decision. In digital images the main goals are denoising and sharpening (e.g. Grillenzoni, 2008) to support image segmentation and classification; in polygonal data the aims are structural analyzes, evaluation and out-of-sample prediction (see Barbiero and Grillenzoni, 2022).

One of the main features of spatial data is the auto-correlation (ACR), which arises from the interaction between the spatial units (pixels in images, and centroids in polygons). The general rule is that the minor is the distance of units, the greater is the size of ACR. This dependence must be properly represented in spatial regression models in order to satisfy the basic assumption of independent residuals. This, in turn, is necessary for having unbiased and efficient estimates of the parameters and their standard errors; that is, for performing statistical inference and prediction. It is also possible to show that correlating two independent, but strongly auto-correlated series, then spurious dependence between them may be found (Granger, 2012); this effect also occurs in spatial data.

On the other hand, ACR is useful for forecasting the character under study in areas where it is not measured, especially when exogenous covariates are not available. It follows that representing the ACR in spatial systems is one of the major concerns of regression modelling. 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 the 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 *causal* ordering. As an example, in lattice data one has the *triangular* and *rook* schemes in Figure 1a,b; in the first, the dynamic is unidirectional since starting from the upper-left corner it follows the sequence of writing a text (see Tjøstheim, 1983). Instead, in Figure 1b the dynamic is multi-directional as each cell involves parts of the text that are not yet written

(Bustos et al., 2009). Figure 1c,b show the analogous situations in polygonal data, where the position of each unit is defined by its center (geometric or institutional). 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 for SAR models are important. Multidirection linkages violate the condition of independence between regressors and residuals, therefore make least squares (LS) estimates biased and inconsistent. To solve this problem, many alternative estimators have been proposed, such as maximum likelihood (ML, Anselin, 2003; LeSage and Pace, 2009), generalized method of moment (GM, Kelejian and Prucha, 1999), two-stage least squares (Lee, 2003) and recently indirect inference (Bao et al., 2020). These methods use iterative algorithms of optimization, which may not converge, and matrices of spatial contiguity which are hindered by the system/sample dimension. However, Lee (2002) showed that under suitable conditions of contiguity, the ordinary LS estimator is consistent even for multilateral models.

In this paper we focus on unilateral SAR for both lattice and polygonal data and we show their ability to preserve the optimal properties of LS estimates even with respect to the efficient ML and GM methods. This feature is important because the LS method is linear and can manage dataset of large dimensions, as it may avoid the direct use of contiguity matrices (i.e. the square arrays that connect all spatial units). Further, unilateral models applied to the data sorted in the same spatial direction as the models, have a recursive structure which enable effective forecasting functions. This feature confirms the close connection between parametric identifiability and model predictability.

Spatial prediction of missing data (inside the perimeter of the observed area) has been treated by LeSage and Pace (2004, 2008), Kelejian and Prucha (2007) and Goulard et al. (2017) for multilateral models, and Mojiri et al. (2021) in the lattice case. They show that the naive predictor based on the *reduced form* of SAR models must be corrected with the best linear unbiased predictor (BLUP) of classical statistics. However, the recursive forecasts of unilateral models do not need the BLUP correction and outperform the others. We show evidence of 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. USGS, 2017) and spatial diffusion of water isotopes (e.g. Moreiras Reynaga et al., 2021).

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; two Appendices provide computational details.

## 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  $Y = \{y_{ij}\}$ , 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  $Y$  with its own values, to obtain interpolates and forecasts:  $\hat{y}_{n+h, m+k}$ . To accomplish this task, as in time series analysis (Fingleton, 2009), the SAR modeling puts each cell in relation to the contiguous ones, 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 allows to define the causality between stochastic processes (e.g. Herrera et al., 2012). 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  $Y$  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 Fig. 1a (see Tjøstheim, 1983, Bustos et al., 2009). Although causality is naturally related to the dynamic of events and aspects of human learning, digital filters of denosing and sharpening may follow non-causal models, such as the *rook* scheme in Fig. 1b (e.g. Grillenzoni, 2008). However, the unidirectional approach remains the favorite solution as allows properties of sequentiality and prediction to the models.

Under linear and unilateral constraints, the triangular SAR( $p$ ) model becomes

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

with independent normal (IN) residuals. In statistics, the model (1) is usually estimated with the maximum likelihood (ML); this requires the vectorization of the data matrix  $\mathbf{y} = \text{vec}_j(Y)$ , of length  $N = n \times m$  and the inversion of auto-covariance matrices  $\mathbf{\Gamma}_y$  of size  $N \times N$ , see Basu and Reinsel (1993, Sect. 4) and Awang and Shitan (2006). 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

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

the LS estimator becomes

$$\hat{\boldsymbol{\phi}}_N = \left( \sum_{i=p+1}^n \sum_{j=p+1}^m \mathbf{x}_{ij} \mathbf{x}_{ij}' \right)^{-1} \sum_{i=p+1}^n \sum_{j=p+1}^m \mathbf{x}_{ij} y_{ij} \quad (2a)$$

$$= \boldsymbol{\phi} + \left( \sum_{i=p+1}^n \sum_{j=p+1}^m \mathbf{x}_{ij} \mathbf{x}_{ij}' \right)^{-1} \sum_{i=p+1}^n \sum_{j=p+1}^m \mathbf{x}_{ij} e_{ij} \quad (2b)$$

At computational level, the double summation in Eq. (2) can be implemented in recursive form as follows

$$\begin{aligned} R_{i,m} &= R_{i-1,m} + P_{im}, & P_{ij} &= P_{i,j-1} + x_{ij} x'_{ij}, & P_0 &= \mathbf{O}_{(p^2-1)}, \\ s_{i,m} &= s_{i-1,m} + q_{im}, & q_{ij} &= q_{i,j-1} + x_{ij} y_{ij}, & q_0 &= \mathbf{0}_{(p^2-1)}, \\ \hat{\phi}_N &= R_{nm}^{-1} s_{nm}, & \hat{\Sigma}_N &= R_{nm}^{-1} \hat{\sigma}_e^2, \end{aligned}$$

with  $N=(n-p)(m-p)$ . Unlike the ML approach, this algorithm only involves the inversion of the matrix  $R$  of size  $(p^2-1)$ , for any values of  $n, m$ . As regards the statistical properties, the LS estimator (2) is unbiased and consistent under the assumption of unilateral dynamic of the model (1) because  $E(x_{ij}e_{ij}) = \mathbf{0}$ .

**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 Basu and Reinsel (1993, p.634) to the lagged term  $y_{i,j-1}$ , one obtains the moving average (MA) representation

$$y_{i,j-1} = e_{i,j-1} + \sum_{h=1}^i \sum_{k=1}^{j-1} \sum_{l=0}^{\max(i,j-1)} \psi_{hkl} e_{i-h,j-1-k-l}, \quad (3)$$

from which  $E(y_{i,j-1}e_{ij})=0$ . Similarly, in Eq. (2b) it follows that  $E(x_{ij}e_{ij})=\mathbf{0}$ , because all entries of  $x_{ij} = [y_{i,j-1}, y_{i-1,j}, y_{i-1,j-1}]'$  do not depend on  $e_{ij}$ , and the LS estimator (2a) is unbiased. This result does not apply to multilateral SAR models, because the decomposition (3) would involve  $e_{i+h,j+k}$ ; for the rook-type scheme in Fig. 1b the ML estimator is necessary. Finally, as in time series, the consistency of LS does not need the stability of the model (1); rather, it improves with signal-to-noise ratio  $\sigma_y^2/\sigma_e^2$  (e.g. Bhattacharyya et al., 1997).

**SAR( $p,q,r$ ).** Empirical models often have a *subset* structure, i.e. have missing lagged regressors 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,  $q$ =number of spatial units and  $r$ =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  $\alpha$ , defined as

$$y_{ij} = \alpha + \phi (y_{i-1,j} + y_{i,j-1} + y_{i-1,j-1})/3 + e_{ij}, \quad y_{i0} = y_{0j} = 0, \quad (4a)$$

$$y_{ij} = \alpha + \phi (y_{i,j-1} + y_{i,j+1} + y_{i-1,j} + y_{i+1,j})/4 + e_{ij}, \quad y_{i,m+1} = y_{n+1,j} = 0, \quad (4b)$$

see Fig. 1a,b. These models can also be written as  $y_{ij} = \alpha + \phi \bar{y}_{ij} + e_{ij}$ , where  $\bar{y}$  is an average of lagged local terms.

The models (4) are first-order both in the spatial lag and the number of coefficients; their condition of stability is given by  $|\phi| < 1$ , as a Corollary of the Proposition 1 of Basu and Reinsel (1993). The rook model (4b) has a multilateral structure, with problems of LS estimation and forecasting; however, it may represent circular dynamics and is a natural competitor of the triangular (4a). This admits an MA representation similar to Eq. (3) with weights

$$\psi_{hkl} = \frac{(h+k+l)!}{h! k! l!} \left(\frac{\phi}{3}\right)^{(h+k+l)},$$

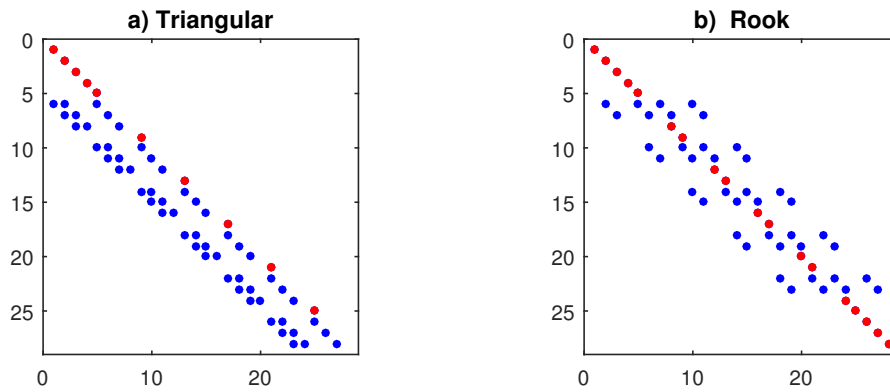
see Bustos et al. (2009). If  $e_{ij} \sim \text{IN}(0, \sigma_e^2)$ , it follows that  $E(y_{11}) = \alpha$ ; the definition of the border values is crucial in simulating the rook model, as it involves missing values. For this reason, we assume the general starting condition  $Y_0 = \alpha$  which is the simplest out-of-sample forecast.

### 2.1. The Vector Form

Previous representations are termed *raster* in GIS softwares; when the lattice dimension is moderate (e.g.  $n \times m < 1000$ ), it may be useful to pass at the *vector* form. Basu and Reinsel (1993), Awang and Shitan (2006) and Mojiri et al. (2021) consider complex forms of vectorization for ML estimation. We follow the spatial econometric approach which vectorizes the data-matrix by columns and build the contiguity matrix of all cells, see LeSage and Pace (2009). Thus let  $\mathbf{y} = \text{vec}_j(\mathbf{Y})$  the  $N = n \times m$  data vector; the first-order vector model with drift is

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

where  $\mathbf{1}_N$  is a unit vector of length  $N$  and  $\mathbf{W}$  is an  $N \times N$  contiguity matrix. Its structure depends on the spatial dynamics and the marginal values: the models (4a,b) have matrices  $\mathbf{W}$  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 Appendix A); the distinctive feature of the unilateral model (4a) is that  $\mathbf{W}$  is lower *triangular*.



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

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

$$y_l = \alpha + \phi \mathbf{w}_l' \mathbf{y} + e_l, \quad l = 1, 2, \dots, N, \quad (6)$$

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

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

$$\begin{aligned} (\mathbf{I}_N - \phi \mathbf{W}) \mathbf{y} &= \alpha \mathbf{1}_N + \mathbf{e}, \\ \mathbf{y} &= (\mathbf{I}_N - \phi \mathbf{W})^{-1} (\alpha \mathbf{1}_N + \mathbf{e}), \end{aligned} \quad (7)$$

this provides an *automatic* way to generate SAR data, independent of  $\mathbf{y}_0$ . Differences with the AR method (6) arise from the choice of the border values and the contiguity matrices; however, in models

with  $W$  triangular and  $\text{diag}(W)=0$  the algorithms (6) and (7) are equivalent. In this case, the impulse-response matrix  $(I_N - \phi W)$  is always invertible and triangular, and leads to the MA decomposition

$$\begin{aligned} M &= (I_N - \phi W)^{-1} = I_N + \sum_{k=1}^{\infty} \phi^k W^k, \\ y &= \frac{\alpha}{(1-\phi)} \mathbf{1}_N + \sum_{k=1}^{\infty} \phi^k W^k e + e, \end{aligned} \quad (8)$$

which converges when  $|\phi| < 1$ . The expression of the constant term follows from the row-normalization, where  $W \mathbf{1}_N = \mathbf{1}_N$  and also  $W^k \mathbf{1}_N = \mathbf{1}_N$ . Further, each element of  $y$  becomes a linear combination of  $e$ , and in unidirectional models, they are function of one-quadrant errors only, providing the MA decomposition

$$y_{ij} = \frac{\alpha}{(1-\phi)} + \sum_{h=1}^{i-1} \sum_{k=1}^{j-1} \psi_{hk} e_{i-h,j-k} + e_{ij},$$

where the weights  $\psi_{hk} \rightarrow 0$  are complex functions of  $\phi^k W^k$ .

Finally, the representation (5) enables to define the LS estimator of parameters  $\delta = [\alpha, \phi]'$  of the models (4) using the entire sample  $Z = [\mathbf{1}_N, W y]$  as

$$\hat{\delta}_N = (Z Z')^{-1} Z y, \quad (9a)$$

$$\hat{\Sigma}_N = (Z Z')^{-1} \hat{\sigma}_e^2, \quad (9b)$$

this improves the estimator (2) based on  $N - p$  observations. The matrix  $\hat{\Sigma}_N$  is the dispersion and if  $W$  is triangular, as in Fig. 2a, then LS (9) is consistent; its analysis will be developed in Sect. 3.

*Spatial Forecasting.* Prediction is one of the central goals of SAR modeling; typical examples are restoring parts of remote sensing data which are hidden by clouds or extrapolating the image out of its perimeter. However, existing literature on lattice data has mostly concerned with *in-sample* filtering and interpolation, focusing on techniques of image sharpening (Grillenzone, 2008), robust denosing (Bustos et al., 2009) and trend estimation (Mojiri et al., 2021), also in conjunction with non-parametric smoothing.

This section aims to forecast data segments which are external of the estimation perimeter  $n \times m$ ; e.g. on the right hand side as  $\{y_{i,m+k}\}$  or, in vector form  $y_{N+l}$  on units placed beyond  $N$ . By, defining  $h, k, l = 1, 2 \dots H, K, L$  the prediction indexes, the forecast function depends on the SAR representation used, see Kelejian and Prucha (2007). Here, we have the lattice form (4), the vector AR (6) and the reduced MA (7); for the triangular model (4a), the 3 predictors are

$$\begin{aligned} \hat{y}_{n+h,m+k} &= E(y_{n+h,m+k} | y_{n-i+1,m-j+1}), \quad h, k = 1, 2 \dots H, K, \\ \hat{y}_{n+h,m+k} &= \alpha + \phi (\hat{y}_{n+h,m+k-1} + \hat{y}_{n+h-1,m+k-1} + \hat{y}_{n+h-1,m+k}) / 3, \end{aligned} \quad (10a)$$

$$\hat{y}_{N+l} = \alpha + \phi w'_{N+l} \hat{y}, \quad \hat{y}' = [y_{11}, y_{21} \dots y_{nm}, \hat{y}_{N+1} \dots \hat{y}_{N+l-1} \dots 0], \quad (10b)$$

$$\hat{y} = (I_{N+L} - \phi W_{N+L})^{-1} \alpha \mathbf{1}_{N+L}, \quad l = 1, 2 \dots L, \quad (10c)$$

where  $w'_{N+l}$  is the  $(N+l)$ -th row of the augmented matrix  $W_{N+L}$  and the vector  $\hat{y}$  in (10b) is updated with the forecasts. The predictor (10c) is nearly automatic, but in absence of exogenous variables provides constant forecasts; on the other hand, the functions (10a,b) need border values to start. These can be set as  $\hat{y}_{1,m+k}^{(0)} = \hat{\alpha}$  for all  $k$ , and may be improved by  $r$ -iterating the forecasts with the nearby values as  $\hat{y}_{1,m+k}^{(r+1)} = \hat{y}_{2,m+k-1}^{(r)}$ , etc.. This approach can be extensively applied to the rook model (4b), to provide initial values also of its forward elements  $y_{ij+1}$ .

The variance of predictors is useful for building confidence intervals of the forecasts (10) and testing hypotheses; as in time series models, the process variance  $\sigma_y^2 = \text{var}(y_{ij})$  represent an upper bound. Applying the formula of Mojiri et al. (2018, p.187) to the triangular model (4a), one obtains

$$\sigma_y^2 = \sigma_e^2 / \left\{ 1 - 2\frac{\phi^2}{9} \left[ \frac{3}{2} + \rho(1,1) + \rho(0,1) + \rho(1,0) \right] \right\} \leq \sigma_e^2 / (1 - \phi^2),$$

where  $|\rho(\cdot, \cdot)| < 1$  are autocorrelation coefficients. The last formula is the variance of an AR(1) process; hence, for the model (4a), the approximate variance of the predictor (10a) may be  $\sigma_e^2 [1 - \phi^{2(h+k)}] / (1 - \phi^2)$ . A more general result can be obtained for the representations (5) and (10c), namely

$$V(\hat{y}) = \left( I_{N+L} - \phi W_{N+L} \right)^{-1} \left( I_{N+L} - \phi W'_{N+L} \right)^{-1} \sigma_e^2,$$

whose diagonal elements increase with the index  $l$ .

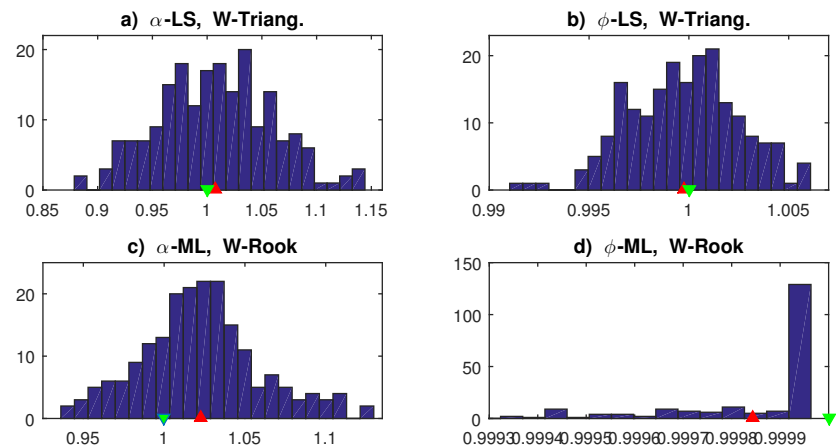
## 2.2. Simulations and Applications

We develop simulation experiments to test the performance of LS (9) and ML (19) estimators (which are based on the same matrices  $W$ ), applied to unilateral and multilateral SAR models. We consider the models (4a,b) 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.. Owing to the forward terms  $y_{i+1,j}, y_{i,j+1}$ , the generation of the rook process (4b) is not *fully* possible with the AR equation (6) and needs the MA reduced form (7), with the matrix in Fig. 2b, without the diagonal entries. Instead, for the unilateral model (4a), the generating methods (6) and (7) are equivalent.

200 replications on a  $32 \times 32$  lattice ( $N=1024$  cells) are generated 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  $\text{MSE}(\hat{\phi})/\phi_0^2$ ) and the  $p$ -values of the Jarque and Bera (1987) Normality test. Since the relative statistics of  $\hat{\alpha}, \hat{\phi}$  do not depend on the parameter size, the overall performance of LS and ML methods can be compared by summing the index values.

The results are displayed in Table 1 and Figure 3; the main findings are as follows: LS (9) is uniformly better than ML (19) when the matrix  $W$  is triangular. Its efficiency improves as  $\phi_0 \rightarrow 1$ , as a consequence of the *super-consistency* property of LS estimates in AR models (e.g. Grillenzoni, 2004). Its performance slightly worsens when using LS (2), due to the smaller sample size, which excludes border values. The performance of LS estimates in multilateral SAR models is biased for  $\phi < 1$ , but still benefits of the super-consistency. It is interesting to note that, unlike time series, the Normal distribution of LS estimates of the AR parameter  $\hat{\phi}_N$  holds even in the presence of the *unit-root*  $\phi_0 = 1$  (see Figure 3b).

As regard the ML method (19), we use the implementation of LeSage and Pace (2009), which is computationally demanding and time consuming. 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 negative, meaning that ML is mainly designed for sparse weight matrices. We will resume this topic in Sect. 3 for non-lattice data; as a conclusion, we can state that LS is suitable for unilateral models, whereas ML must be used in multilateral SAR.



**Figure 3.** Distribution of LS and ML estimates in 200 replications of the SAR models (4a,b) with  $\alpha_0=1$ ,  $\phi_0=1$  (green), and mean values of  $\hat{\alpha}$ ,  $\hat{\phi}$  (red): Panels (a,b) LS (9) of model (4a); Panels (c,d) ML (19) of model (4b).

**Table 1.** Performance of LS (9) and ML (19) estimators applied to triangular and rook SAR models (4a,b) 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 enhances the best methods (LS or ML) over model classes (Triang. and Rook).

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

We also carry out two empirical applications to test the performance of various SAR models in out-of-sample forecasting of spatial data. In comparing predictions, one cannot proceed by simulations as in the experiments of Tab. 1, because simulated data depend on the ground models, which will influence the final results. Hence, we fit SAR models to data generated from external sources.

*Random Surface.* The first application considers rough random surfaces (Sheffield, 2022); these techniques are used in physics to study electromagnetic, fluid and plasma phenomena (see Sjodin, 2017), and are applied in engineering (e.g. 3D printers). We follow an approach based on trigonometric series, where a matrix of random numbers is convolved with a Gaussian filter to achieve spatial

autocorrelation. Thus, let the fast Fourier transform (FFT) of the bivariate random sequence  $u_{ij}$  on a  $m \times n$  regular lattice be

$$F(u_{ij}) = \sum_{h=0}^{n-1} \sum_{k=0}^{m-1} u_{hk} e^{-i2\pi hi/n} e^{-i2\pi kj/m}, \quad u_{ij} \sim \text{IN}(0, \sigma_u^2),$$

where  $i$  is the imaginary unit. Let  $R_{ij}(\delta) = e^{-0.5(i^2+j^2)/\delta^2}$  an exponential ACR function with  $\delta > 0$ ; then the model surface generator becomes

$$y_{ij} = \alpha + \Re\left\{F^{-1}\left[R_{ij}(\delta) F(u_{ij})\right]\right\} + e_{ij}, \quad e_{ij} \sim \text{IN}(0, \sigma_e^2), \tag{11}$$

where  $\Re(\cdot)$  is the real part,  $F^{-1}(\cdot)$  is the inverse FFT and  $e_{ij}$  is a white noise.

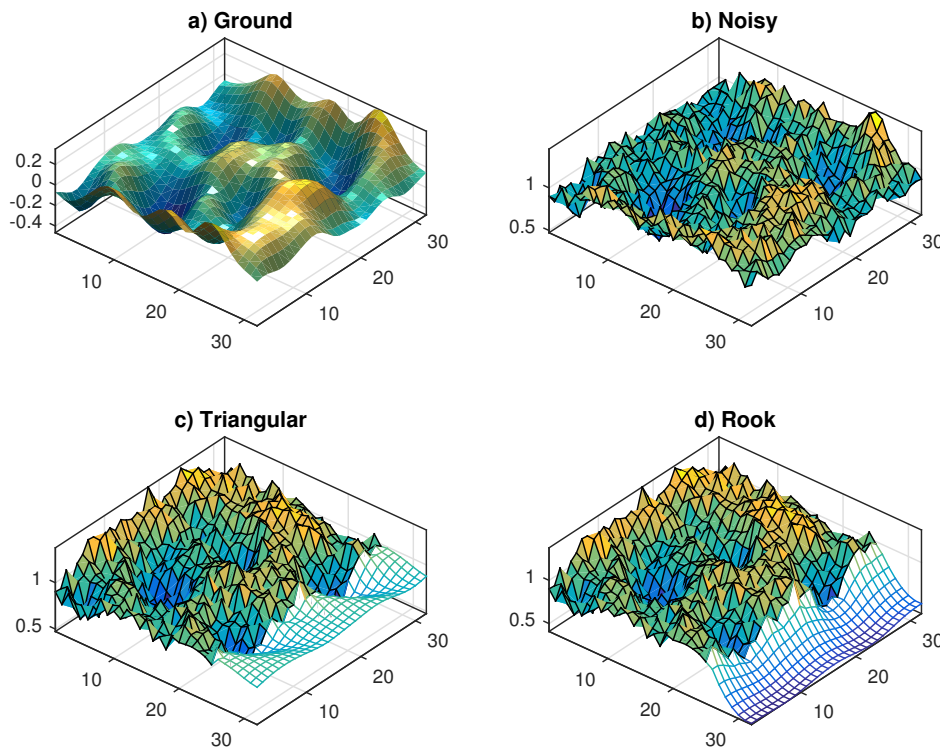
Despite the randomness of  $u_{hk}$ , the rough surface  $g_{ij} = \Re(\cdot)$  in Eq. (11) is still smooth, hence we add an external noise  $e_{ij}$  to increase its realism; the resulting series  $y_{ij}$  resemble real terrain models of geostatistics (see Figs. 4a,b). In the experiment we set  $n=m=32$  (which yields  $N=1024$ ),  $e, u$  independent Gaussian with  $\sigma_u=1$ ,  $\sigma_e=0.067$ , and  $\alpha=1$ ,  $\delta=2.5$  provide mild spatial ACR. The goal is to forecast the last  $K=7$  columns (about 22%) of the surface  $Y$  with the algorithms (10). Table 2 reports the in-sample parameter estimates (with LS and ML methods) of the models (4a,b) and the mean absolute percentage forecast errors statistics

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

A triangular SAR(1,3,3) model (with 3 distinct AR parameters) is also fitted to the data, but it does not improve the forecasts of the constrained (4a). The rook (4b) 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. The performance of the vector predictor (10b) is similar to that of the lattice method (10a) and is reported in Appendix B, together with that of the function (10c).

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

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



**Figure 4.** Panels (a,b): Simulation of the random surface (11) of size  $n=m=32$ . Panels (c,d): Out-of-sample forecasts of  $Y(:,26:32)$  with models (4a,b) and predictor (10a).

*Remote Sensing.* The second application regards a real case study; we consider high-resolution elevation data obtained with aerial laser scanners. The sample array  $Y$  comes from USGS (2017) and has  $340 \times 455$  pixels (see Fig. 5); this would yield a vector SAR model with 154,700 rows. Further, the image extrapolation may require the inclusion of spatial trend components (e.g. Grillenzoni, 2004), such as bivariate polynomials  $g_d(i, j)$  of degree  $d \geq 1$ . Given the implied computational issues, we only perform analysis with the lattice implementation (2); in the case of a triangular model with quadratic trend, the forecasting function becomes

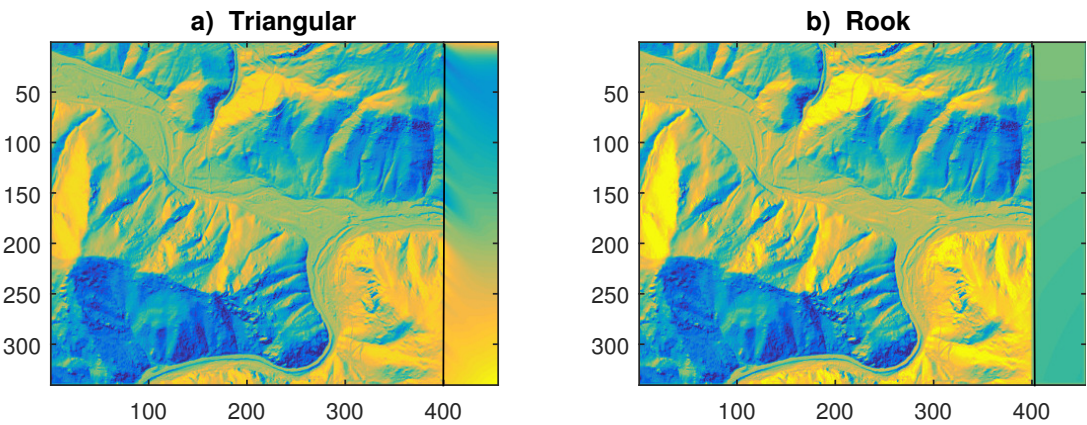
$$\begin{aligned} \hat{y}_{i+h,j+k} &= \alpha + g_2(i+h, j+k) + \phi (\hat{y}_{i+h-1,j+k-1} + \hat{y}_{i+h-1,j+k} + \hat{y}_{i+h,j+k-1})/3, \\ g_2(i+h, j+k) &= \beta_1(i+h) + \beta_2(j+k) + \beta_3(i+h)^2 + \beta_4(j+k)^2 + \beta_5(i+h)(j+k), \end{aligned}$$

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

In the application to the USGS image, 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  $R^2=0.98$ ), the rook model has a disappointing forecasting performance both in terms of MAPE statistic and visual appearance (Figure 5b). The reason is partly due to the (biased) estimates of its parameters, with  $\hat{\alpha} < 0, \hat{\phi} > 1$ ; however, adjusting their values reduces the MAPE statistics but do not change the (uniform) path of the predicted surface.

**Table 3.** Estimates of SAR models with bivariate polynomials applied to the USGS (2017) image:  $Y(:,1:400)$  in-sample data,  $Y(:,401:455)$  out-of-sample.

| Model                | $\hat{\alpha}$ | $\hat{\phi}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\sigma}_e^2$ | MAPE         |
|----------------------|----------------|--------------|-----------------|-----------------|-----------------|-----------------|-----------------|--------------------|--------------|
| Triang.<br>(4a)      | 8.50           | 0.961        | -0.024          | -0.011          | 11e-5           | 43e-7           | -88e-7          | 338.4              | <b>0.848</b> |
| $T$ -stat.,<br>$R^2$ | 25.56          | 104.9        | -10.58          | -5.47           | 22.55           | 0.735           | -2.07           | 0.913              | .            |
| Rook<br>(4b)         | -6.12          | 1.026        | 0.007           | 0.016           | -58e-6          | 17e-6           | -17e-6          | 88.15              | 1.362        |
| $T$ -stat.,<br>$R^2$ | -35.82         | 214.0        | 6.29            | 15.87           | -22.80          | 5.50            | -7.82           | 0.977              | .            |



**Figure 5.** Application of SAR models (4a,b) with quadratic trend to USGS (2027) data: out-of-sample forecasts of the image portion  $Y(:,401:455)$ .

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 missing pixels) or may be replaced by nonparametric smoothers. In any event, the role of the AR components remains fundamental, as the coefficient  $\hat{\phi}$  is the most significant in both models of Table , with statistic  $T_{\hat{\phi}} > 100$ .

**Table 4.** Results of the simulation experiments on the SAR systems (13)-(15) with spatial lags  $q=1-3$  and single  $\phi$ . 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<br>Estimator | North NN      |        |               | Simple NN |               |               | $\phi$    | lags |
|----------------|---------------|--------|---------------|-----------|---------------|---------------|-----------|------|
|                | LS            | ML     | GM            | LS        | ML            | GM            |           |      |
| Ave.<br>Rbias  | <b>0.0205</b> | 0.1338 | 0.0235        | 0.1363    | 0.0287        | <b>0.0216</b> | 0.65      | 1    |
| Ave.<br>RRMSE  | <b>0.1336</b> | 0.2402 | 0.1831        | 0.1887    | <b>0.1343</b> | 0.1581        |           |      |
| Ave.<br>Ntest  | 0.3218        | 0.2983 | 0.1675        | 0.1561    | 0.3401        | 0.3337        |           |      |
| Ave.<br>Rbias  | <b>0.0077</b> | 0.1208 | 0.0098        | 0.1232    | 0.0400        | <b>0.0075</b> | 0.65      | 1    |
| Ave.<br>RRMSE  | <b>0.1396</b> | 0.2445 | 0.1924        | 0.1873    | <b>0.1430</b> | 0.1609        | (Variable |      |
| Ave.<br>Ntest  | 0.3274        | 0.3554 | 0.2291        | 0.1637    | 0.2499        | 0.1626        | grid)     |      |
| Ave.<br>Rbias  | 0.0340        | 0.1330 | <b>0.0217</b> | 0.1324    | <b>0.0259</b> | 0.0376        | 0.65      | 3    |
| Ave.<br>RRMSE  | <b>0.1586</b> | 0.2416 | 0.2365        | 0.1998    | <b>0.1390</b> | 0.2035        |           |      |
| Ave.<br>Ntest  | 0.2442        | 0.2612 | 0.0936        | 0.1811    | 0.2559        | 0.0960        |           |      |
| Ave.<br>Rbias  | 0.0298        | 0.1289 | <b>0.0228</b> | 0.1487    | <b>0.0208</b> | 0.0343        | 0.65      | 3    |
| Ave.<br>RRMSE  | <b>0.1518</b> | 0.2360 | 0.2228        | 0.2088    | <b>0.1326</b> | 0.1963        | (IDW      |      |
| Ave.<br>Ntest  | 0.2530        | 0.2561 | 0.1673        | 0.1805    | 0.2900        | 0.1736        | weights)  |      |
| Ave.<br>Rbias  | 0.0312        | 0.0980 | <b>0.0248</b> | 0.0581    | <b>0.0109</b> | 0.0164        | 0.95      | 3    |
| Ave.<br>RRMSE  | <b>0.1475</b> | 0.2514 | 0.3190        | 0.1208    | <b>0.1086</b> | 0.1648        |           |      |
| Ave.<br>Ntest  | 0.1516        | 0.1678 | 0.0010        | 0.2611    | 0.3033        | 0.1601        |           |      |
| Ave.<br>Rbias  | <b>0.0663</b> | 0.3405 | .             | 0.3333    | <b>0.0704</b> | .             | 0.65      | 3    |
| Ave.<br>RRMSE  | <b>0.2849</b> | 0.4659 | .             | 0.4141    | <b>0.2764</b> | .             | (Durbin   |      |
| Ave.<br>Ntest  | 0.2584        | 0.3891 | .             | 0.1286    | 0.1313        | .             | model)    |      |

3. Point and Polygonal Data

When the data are irregularly distributed in space (either as point processes or polygonal areas), notations and representations have 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  $\mathbf{x}'_s = [i_s, j_s]$  (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

$$y_s = \alpha + \phi y_{s-1} + \beta' x_s + e_s, \quad s = 1, 2 \dots N, \tag{13a}$$

$$\mathbf{y} = \alpha \mathbf{1}_N + \phi \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_N), \tag{13b}$$

$$= (\mathbf{I}_N - \phi \mathbf{W})^{-1} (\alpha \mathbf{1}_N + \mathbf{X} \boldsymbol{\beta} + \mathbf{e}), \tag{13c}$$

where  $s-1$  refers to the unit which is *closest* to the  $s$ -th term, and the vector of regressors  $\mathbf{x}'_s = [i_s, j_s, \dots]$  may include exogenous covariates.

Unlike the previous section, the adjacency matrix  $\mathbf{W}$  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) direction may be followed as in the lattice case. However, polygonal data have not a lexicographic order; hence, the unilateral constraint may simply be defined along the north direction. In this setting, the observations  $\{y_s, \mathbf{x}_s\}$  are ordered 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 by  $(i_s)$  or simply  $(s)$ , and each equation of the model (13a) can be written in sequential form, as a time series model

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

where  $y_{(s-1)}$  is the north NN of  $y_{(s)}$ . The property  $E[y_{(s-1)}e_{(s)}] = 0$  as in time series is the basis for unbiased LS estimation and forecasting. However, it does not hold in the simple NN linkage, even when data are ordered along the latitude.

As an illustration of NN matrices, we simulate  $N=30$  random points with  $[i_s, j_s] \sim \text{IU}(0, 1)^2$  and consider  $q=3$  contiguous terms, see Figure f6. Under the north ordering of data, the entries of matrices are concentrated around the null main diagonal  $\text{diag}(\mathbf{W})=0$ , and with the unilateral constraint the array is lower triangular. In order to obtain non-sparse matrices, LeSage and Pace (2004) ordered data according to the sum  $(i_s + j_s)$ , but just used the simple NN rule. The matrices  $\mathbf{W}_k$  in Figure f6 yield SAR(3) models, but if they are combined as  $\bar{\mathbf{W}} = 3^{-1} \sum_{k=1}^3 \mathbf{W}_k$  they provide constrained SAR(3,1). The analogous of the lattice model (4a) then becomes

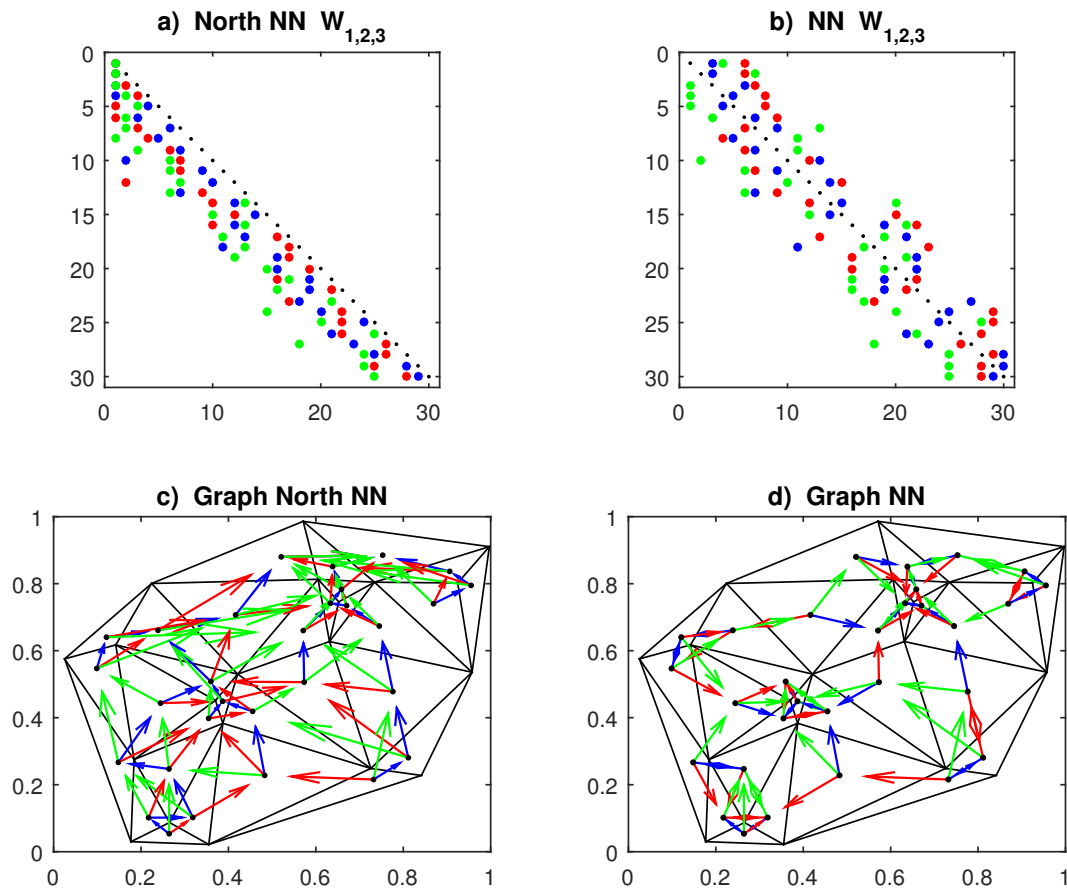
$$y_{(s)} = \alpha + \phi \left( q^{-1} \sum_{k=1}^q y_{(s-k)} \right) + \beta' \mathbf{x}_{(s)} + e_{(s)}. \quad (15)$$

The econometric literature often discusses models with autocorrelated errors, such as  $\mathbf{u} = (\mathbf{I}_N - \theta \mathbf{W})^{-1} \mathbf{e}$ , e.g. Kelejian and Prucha (2007). Apart from estimation difficulties, it should be noted that they are encompassed by unconstrained SAR(2); indeed, the combined system  $(\mathbf{I}_N - \theta \mathbf{W})(\mathbf{I}_N - \phi \mathbf{W})\mathbf{y} = (\mathbf{I}_N - \theta \mathbf{W})(\alpha + \mathbf{X}\beta) + \mathbf{e}$  involves just 2 spatial lags. In place of  $\mathbf{u}$ , it is better to insert into the model (13b) a lagged term on the exogenous part  $+\mathbf{W}\mathbf{X}\gamma$ , that corresponds to  $\gamma = \theta\beta$  in the model with  $\mathbf{u}$  errors. This is the Durbin model (LeSage and Pace, 2009), which is useful when  $\mathbf{x}_{(s)}$  are autocorrelated, i.e.  $(i_s, j_s)$  have a non-random pattern.

When the matrix  $\mathbf{W}$  is lower triangular, the transfer function  $(\mathbf{I}_N - \phi \mathbf{W})$  is always nonsingular and its inverse admits a convergent power expansion

$$\mathbf{y} = \frac{\alpha}{(1-\phi)} \mathbf{1}_N + \mathbf{X}\beta + \sum_{k=1}^{\infty} \phi^k \mathbf{W}^k \mathbf{X}\beta + \sum_{k=1}^{\infty} \phi^k \mathbf{W}^k \mathbf{e} + \mathbf{e}.$$

This expression shows that the process  $y_{(s)}$ , in the location  $s$ -th, depends on the values of the inputs  $\mathbf{x}_{(r)}$  and shocks  $e_{(r)}$  in the north locations ( $r < s$ ), and with weights  $\phi^k \mathbf{W}^k$  which decline with the distance. To appreciate this diffusion effect, note that when  $\mathbf{W}$  is strictly lower triangular, then also  $\mathbf{W}^2$  is lower triangular but with nonzero entries shifted away from the main diagonal. Hence,  $\mathbf{W}^2$  reflects the second-order contiguity (the neighbors of the neighbors), and so forth for  $\mathbf{W}^k \rightarrow \mathbf{O}$  the null matrix. The same feature, however, does *not* hold in non-triangular matrices, such as those based on the simple NN.



**Figure 6.** Contiguity matrices and planar graphs of  $N=30$  random centroids on the unit square  $[0,1]$ . Panels: (a,c) North nearest neighbors (NN); (b,d) Multidirection NN. Matrices:  $W_1$  first NN (blue);  $W_2$  second NN (red);  $W_3$  third NN (green).

The unilateral model (14) corresponds to a time-series AR with *unequally* spaced observations. This feature is more clear in the presence of multiple lagged terms  $y_{(s-k)}$ ; the suitable treatment is to weight the observations by the inverse Euclidean distance  $1/d_{(s,s-k)}$ . The analog of the constrained model (15), with north NN linkage and with row normalization, is then given by

$$y_{(s)} = \alpha + \phi \left( \sum_{k=1}^q w_{(sk)} y_{(s-k)} \right) + \beta' x_{(s)} + e_{(s)}, \quad (16)$$

$$w_{(sk)} = d_{(s,s-k)}^{-1} / \sum_{k=1}^q d_{(s,s-k)}^{-1}, \quad d_{(s,s-k)} = [(i_{(s)} - i_{(s-k)})^2 + (j_{(s)} - j_{(s-k)})^2]^{1/2},$$

this also admits a vector representation as (13b) with triangular matrix  $W$ , having  $q$  decreasing coefficients per row.

Alternatively, since the inverse distances decay too fast, one can use the exponential weights  $w_{(sk)} \propto \lambda^{d_{(s,s-k)}}$ , with  $0 < \lambda < 1$ . In vector form, one has to define the diagonal matrices  $\Lambda_k = \lambda^{D_k}$  where  $D_k = \text{diag}[0 \dots 0, d_{(k,1)}, d_{(k+1,2)} \dots d_{(N,N-k)}]$  contain the  $k$ -th north NN distances of the centroids. These arrays must be multiplied by the  $k$ -lag contiguity matrices  $W_k$ , then summed as  $W = \sum_k \Lambda_k W_k$  and finally normalized by rows. The resulting model is nonlinear in the parameters and requires iterative algorithms; however, the value of  $\lambda$  may be selected a-priori in the range  $[.9, 1)$ , then the estimators compute the suitable value of  $\phi$ .

### 3.1. Estimation Methods

Writing the models (14)-(16) in vector form, as  $y_{(s)} = \delta' z_{(s)} + e_{(s)}$ , with  $\delta = [\alpha, \phi, \beta]'$  and  $z_{(s)} = [1, \bar{y}_{(s-1)}, x_{(s)}]'$ , where  $\bar{y}_{(s-1)}$  is the average of  $q$ -north NN terms, then the LS estimator of the parameters  $\delta$  is given by

$$\begin{aligned}\hat{\delta}_N &= \left( \sum_{s=q+1}^N z_{(s)} z_{(s)}' \right)^{-1} \sum_{s=q+1}^N z_{(s)} y_{(s)}, \\ &= \delta + \left( \sum_{s=q+1}^N z_{(s)} z_{(s)}' \right)^{-1} \sum_{s=q+1}^N z_{(s)} e_{(s)},\end{aligned}\quad (17)$$

its unbiasedness follows from  $E[z_{(s)} e_{(s)}] = \mathbf{0}$  by the sequential structure of the models, in particular  $E[e_{(s)} \bar{y}_{(s-1)}] = 0$  in the unilateral model.

To see what happens in detail, consider the LS estimator in matrix form

$$\hat{\delta}_N = (Z'Z)^{-1}Z'y = \delta + (Z'Z)^{-1}Z'e \quad (18)$$

with regressors  $Z = [1_N, Wy, X]$ , the crucial term is  $Wy$ . From Eq. (13c)

$$\begin{aligned}E(Z'e) &= E[(Wy)'e] = E\left\{[(I_N - \phi W)^{-1}(\alpha 1_N + X\beta + e)]'W'e\right\} \\ &= E[e'(I_N - \phi W')^{-1}W'e] = E(e'H'e), \quad (\text{say})\end{aligned}$$

this term vanishes only if the trace (tr) of  $H' = (I_N - \phi W')^{-1}W'$  is 0, because

$$E(y'W'e) = E[\text{tr}(e'H'e)] = \text{tr}[HE(ee')] = \text{tr}(H)\sigma_e^2 I_N.$$

Now,  $\text{tr}(H) = 0$  in general holds if  $W$  is strictly lower triangular, because  $H$  is itself triangular with null diagonal.

As regards the convergence of LS estimator, multiplying Eq. (17) by  $1/N$ , under regularity conditions (stability and existence of second-order moments) one has that  $\text{Plim}_{N \rightarrow \infty}(Z'Z/N) = \Sigma_z$ , the covariance matrix of  $z_{(s)}$ . Whereas

$$\text{Plim}_{N \rightarrow \infty}(Z'e/N) = E[z_{(s)} e_{(s)}] = \text{tr}(H)\sigma_e^2,$$

which is 0 if  $W$  is triangular. Finally, under regularity conditions ( $\Sigma_z$  positive definite,  $|\phi| < 1$  and  $x_s$  stationary) one has

$$\sqrt{N}(\hat{\delta}_N - \delta) \xrightarrow[N \rightarrow \infty]{D} N(0, \Sigma_z^{-1}\sigma_e^2),$$

where the convergence is in distribution (D), see Yao and Brockwell (2006).

For multilateral SAR models ( $W$  non-triangular), the LS is generally biased and inconsistent, and alternative estimators must be sought. The ML method may be suitable if the sample size  $N$  is moderate and the observations have Gaussian distribution. This estimator maximizes the log-likelihood function

$$\begin{aligned}\ell(\alpha, \phi, \beta, \sigma) &\propto -N \log(\sigma) + \log[\det(I_N - \phi W)] - (e'e)/2\sigma_e^2, \\ e(\alpha, \phi, \beta) &= (I_N - \phi W)y - (\alpha 1_N + X\beta),\end{aligned}\quad (19)$$

which requires  $\det(\cdot) > 0$ , i.e. the stability condition  $\phi \in [0, 1)$  (LeSage and Pace, 2009, p.63). This requirement is not necessary in the LS method; moreover, numerical issues may arise in the optimization of Eq. (19), when computing  $\log \det(\cdot) = \sum_{i=1}^N \log(1 - \phi \lambda_i)$ , with the eigenvalues  $\lambda_i$  of  $W$ . This techniques cannot be applied to triangular matrices, because  $\lambda_i=0$  for all  $i$ .

Another method which is suitable for parameter estimates in multilateral  $\mathbf{W}$  is that of generalized moments (GM, Kelejian and Prucha, 1999); it arises by applying the instrumental variables (IV) method to the LS estimator (17). The first step is the definition of valid IVs: on the basis of the predictor decomposition  $\hat{\mathbf{y}} = \sum_k \phi^k \mathbf{W}^k (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha})$ , they may be given by  $\mathbf{M} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X} \dots]$ . This provides the projection matrix  $\mathbf{P} = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$  and the regressors  $\mathbf{V} = \mathbf{PZ}$ , which lead to the IV estimator  $\tilde{\boldsymbol{\delta}}_N = (\mathbf{V}'\mathbf{Z})^{-1}\mathbf{V}'\mathbf{y}$ . This solution is also of GM-type, as minimizes the functional  $Q_N(\boldsymbol{\delta}) = \mathbf{m}'\mathbf{S}^{-1}\mathbf{m}$ , with the empirical moments  $\mathbf{m} = \mathbf{M}'\mathbf{e}/N$  and  $\mathbf{S} = \mathbf{M}'\mathbf{e}\mathbf{e}'\mathbf{M}/N$ . In the following, we check the effectiveness of ML and GM estimators with simulation experiments.

### 3.2. Forecasting Functions

In forecasting, we have  $L$  out-of-sample units with value  $\mathbf{y}_L = \mathbf{0}$ , whose coordinates and regressors  $\mathbf{x}'_l = [1, i_l, j_l, x_{1l} \dots x_{ml}]$  are known for all  $l = 1, 2 \dots L$ . 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  $\mathbf{X}_{N+L}$ . 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:  $\mathbf{W}_{N+L} = \text{diag}[\mathbf{W}_N, \mathbf{W}_L]$ , where  $\mathbf{W}_L$  is overlapped to  $\mathbf{W}_N$  for the contiguity of  $L$  units with the in-sample ones. In the other cases, it has a more complex structure, with triangular sub-matrices under the unilateral constraint

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

In any case, the estimation of parameters  $\boldsymbol{\delta}$  is performed with in-sample data  $\mathbf{y}_N$  and the matrix  $\mathbf{W}_{NN}$ ; the other blocks will be used in forecasting.

The forecasting function of  $\mathbf{y}_L$  depends on the SAR representation used for  $\mathbf{y}$ ; Kelejian and Prucha (2007) also considered models with autocorrelated errors  $\mathbf{u}$ . In the reduced (MA) form (13c), the in-sample (fitted values) and out-of-sample forecasts are jointly computed as

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

where the contiguity matrix may have  $q$ -entries per row (the spatial lags) and may use non-uniform (IDW) weights. The matrix solution (20) is nearly automatic, but in absence of exogenous variables it provides constant forecasts.

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

$$\begin{aligned} \hat{\mathbf{y}}_L &= \mathbf{E}(\mathbf{y}_L | \mathbf{y}_N, \mathbf{X}_{N+L}, \mathbf{W}_{N+L}), \\ \hat{\mathbf{y}}_{N+l} &= \alpha + \phi \mathbf{w}'_{N+l} \hat{\mathbf{y}}_{N+l-1} + \mathbf{x}'_{N+l} \boldsymbol{\beta}, \quad l = 1, 2 \dots L, \\ \hat{\mathbf{y}}_{N+l-1} &= [y_1, y_2, \dots, y_N, \hat{y}_{N+1} \dots \hat{y}_{N+l-1}, 0, \dots, 0]', \end{aligned} \quad (21)$$

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

LeSage and Pace (2004) and Goulard et al. (2017) 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:

$$\begin{aligned} E(\mathbf{y}_L | \mathbf{y}_N) &= \boldsymbol{\mu}_L + \boldsymbol{\Sigma}_{LN} \boldsymbol{\Sigma}_{NN}^{-1} (\mathbf{y}_N - \boldsymbol{\mu}_N), \\ \hat{\tilde{\mathbf{y}}}_L &= \tilde{\mathbf{y}}_L + \boldsymbol{\Sigma}_{LN} \boldsymbol{\Sigma}_{NN}^{-1} (\mathbf{y}_N - \tilde{\mathbf{y}}_N), \end{aligned} \quad (22)$$

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

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

The solution (22)-(23) requires a *double* matrix inversion: first  $(\mathbf{I}_{N+L} - \phi \mathbf{W}_{N+L})^{-1}$  and next  $\boldsymbol{\Sigma}_{NN}^{-1}$ ; in the presence of large  $N$ , this may introduce approximation errors. To reduce numerical instability, one may partition the simpler array  $\boldsymbol{\Omega}_{N+L} = \boldsymbol{\Sigma}_{N+L}^{-1}$  and exploit the algebraic relationship  $\boldsymbol{\Sigma}_{LN} \boldsymbol{\Sigma}_{NN}^{-1} = -\boldsymbol{\Omega}_{LL}^{-1} \boldsymbol{\Omega}_{LN}$  (see LeSage and Pace, 2008). The latter involves a single matrix inversion  $\boldsymbol{\Omega}_{LL}^{-1}$ , of a submatrix of lower size  $L \ll N$ . Finally, the conditional variance  $V(\mathbf{y}_L | \mathbf{y}_N)$  of the BLUP estimator may provide the dispersion of the forecasts (22)

$$V(\hat{\tilde{\mathbf{y}}}_L) = \boldsymbol{\Sigma}_{LL} - \boldsymbol{\Sigma}_{LN} \boldsymbol{\Sigma}_{NN}^{-1} \boldsymbol{\Sigma}_{NL},$$

this matrix requires the estimation of the variance  $\sigma_e^2$ , which can be carried out with the in-sample residuals  $\hat{\mathbf{e}}_N = (\mathbf{y}_N - \hat{\mathbf{y}}_N)$  of Eqs. (20).

As a final remark, notice that all formulae displayed until now can be extended to unconstrained SAR( $q$ ) models, by replacing the IR matrix with one of order  $q$ -th. For the predictors (20),(21) this yields the following dispersion

$$V(\tilde{\mathbf{y}}) = \left( \mathbf{I}_{N+L} - \sum_{k=1}^q \phi_k \mathbf{W}_k \right)^{-1} \left[ \left( \mathbf{I}_{N+L} - \sum_{k=1}^q \phi_k \mathbf{W}_k \right)^{-1} \right]' \sigma_e^2,$$

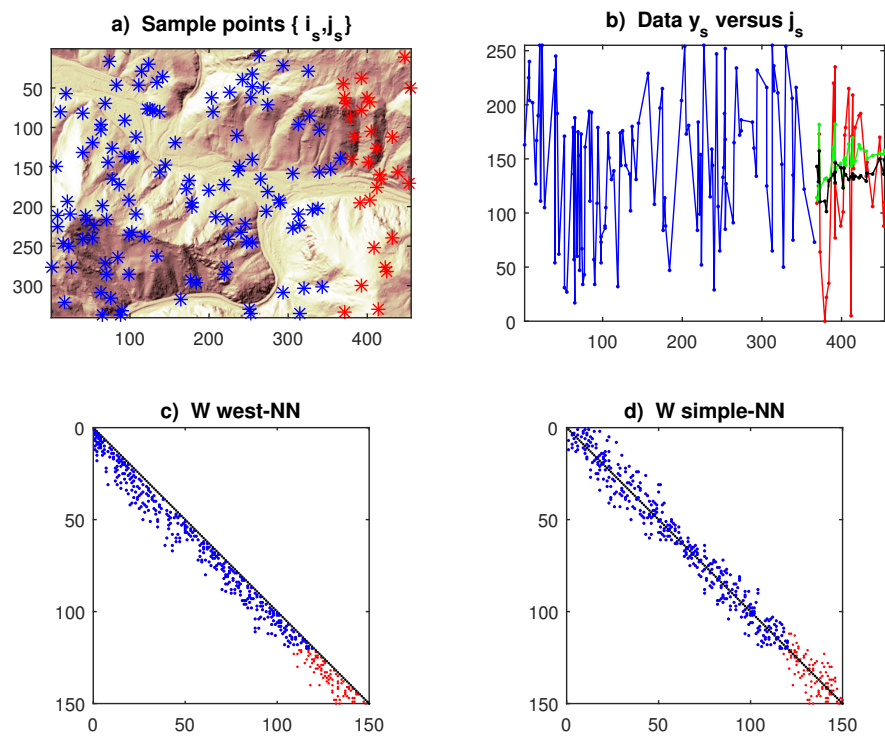
where the  $k$ -th lag contiguity matrices  $\mathbf{W}_k$  have size  $N + L$ .

### 3.3. Simulations and Applications

To compare the various estimators, we perform simulation experiments of SAR models (13)-(16) on a random grid of  $N=150$  points, uniformly distributed in the unit square:  $[i_s, j_s] \sim \text{IU}(0, 1)^2$ , as in Figure f6c. We generate realizations of the process  $\{y_s\}$  with independent inputs  $x_s \sim \text{IU}(0, 2)$  Uniform and  $e_s \sim \text{IN}(0, 1)$  Normal; parameters  $\alpha = 1, \phi = 0.65, 0.95, \beta = -1$ ; contiguity matrices  $\mathbf{W}$  with  $q=1, 3$  lags, 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 Normality test were computed for the coefficients  $\hat{\alpha}, \hat{\phi}, \hat{\beta}$  and then averaged. Note that relativization (e.g.  $\text{bias}(\hat{\phi}_N) / |\phi_0|$ ) allows to sum the individual statistics so as to directly compare the various methods.

The results are reported in Table 4, where "variable grid" means that at each replication the centroids change and "variable weights" indicates that IDW are used in the matrices  $\mathbf{W}$ . The Durbin model includes a lagged term on the exogenous part, as  $\beta_1 x + \beta_2 \mathbf{W}x$ , with coefficients  $[\beta_1, \beta_2] = [1, -.65]$ . The software used for ML and GM estimates is the Matlab package of LeSage and Pace (2009); it fits the Durbin model only with the ML method. Main conclusions from Table 4 are that ML and GM methods do not improve the LS estimates in the case of unidirectional (triangular)  $\mathbf{W}$  matrices. However, they significantly outperform LS in the multidirectional (e.g. simple NN) case; in particular, the ML method is the best. However, these results are not homogeneous as regards unbiasedness and efficiency of estimates, where the GM estimator may locally have smaller bias.

*Forecasting.* Unlike estimation, prediction evaluation of SAR methods cannot be based on pre-specified models and contiguity matrices, but it must refer to data generated from autonomous sources. As in the lattice case, we considered the random surfaces (11), sampled at random points  $i_s \sim \text{IU}[1, n]$ ,  $j_s \sim \text{IU}[1, m]$ , with  $s = 1, 2 \dots N$ ; however, the results are too favourable to unilateral SAR models. We then turn to the real USGS (2007) data in Figure 4, with  $n=340$ ,  $m=455$ ; we sample  $N=150$  cells and withhold for forecasting  $L=30$  outside, see Figure 7a. The fitted model is SAR (15) with  $q=3$  and  $\mathbf{x}_s = [i_s, j_s]'$ , and forecast functions (20)-(22); the unilateral model is estimated with LS and the multilateral one with ML.



**Figure 7.** Graphical results of a single replication of the estimates in Table 5: a) Random sampling of USGS surface (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) Contiguity matrix of the multilateral model.

**Table 5.** Average Values, over 100 Replications, of Parameter Estimates (and  $T$ -statistics) and MAPE Forecast Statistics of SAR Model (15) on Random Samples of USGS (2007) Data.

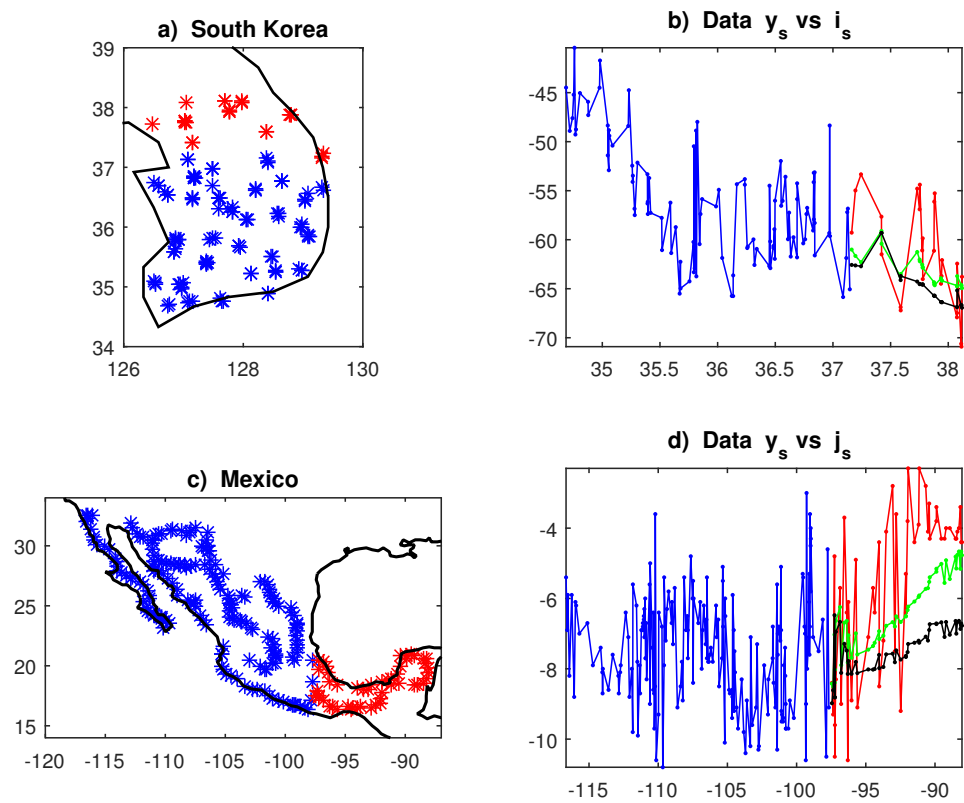
| Model      | Method                           | $\bar{\alpha}$ | $\bar{\phi}$ | $\bar{\beta}_1$ | $\bar{\beta}_2$ | $\bar{\sigma}_e$ | MAPE(20) | MAPE(21)     | MAPE(22) |
|------------|----------------------------------|----------------|--------------|-----------------|-----------------|------------------|----------|--------------|----------|
| Unilateral | LS (17)                          | 68.4           | 0.543        | 0.012           | -0.018          | 54.3             | 0.948    | <b>0.815</b> | 0.830    |
|            | $\bar{T}$ -stat.,<br>$\bar{R}^2$ | (3.0)          | (5.1)        | (0.21)          | (-0.31)         | 0.213            | .        | .            | .        |
| Multilat.  | ML (19)                          | 76.5           | 0.492        | 0.005           | -0.009          | 50.2             | 0.996    | 0.951        | 0.899    |
|            | $\bar{T}$ -stat.,<br>$\bar{R}^2$ | (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 ( $\delta^{18}\text{O}$ ) and hydrogen ( $\delta^2\text{H}$ ) in the ground water. Mapping isoscapes is useful for physical monitoring of the hydrological cycle and for archeological and forensic investigations, regarding the path of people movement. We consider the datasets of Gautam et al. (2020), recorded in South-Korea in 2010, with  $N=130$ ; and Moreiras Reynaga et al. (2021), recorded in Mexico in 2007,

with  $N=234$ . We estimate SAR model (15) with  $q=3$  for  $y_s = \delta^{18}\text{O}, \delta^2\text{H}$  and  $x_s = \text{latitude, longitude}$ , to forecast about 20% of observations; the results are in Tab. 6 and Fig. 8. They confirm the better performance of unilateral SAR models with the AR algorithm (21), especially when the data have significant autocorrelation (i.e. large  $|\phi|$ ) and marked spatial trends (significant  $\beta_{1,2}$ ). The reduction of MAPE statistics in Tab. 6 ranges from -18% in Korea to -38% in Mexico.

**Table 6.** Parameter estimates (and  $T$ -statistics) and MAPE forecast statistics of the SAR model (15) with  $q=3$  applied to the water isotope data of South Korea  $\delta^2\text{H}$  (first block) and Mexico  $\delta^{18}\text{O}$  (second block).

| Model                | Method | $\hat{\alpha}$ | $\hat{\phi}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}_e$ | MAPE(20) | MAPE(21)     | MAPE(22) |
|----------------------|--------|----------------|--------------|-----------------|-----------------|------------------|----------|--------------|----------|
| Unilateral LS (17)   |        | 123.7          | 0.337        | -0.557          | -2.51           | 4.28             | 0.069    | <b>0.065</b> | 0.065    |
| $T$ -stat.,<br>$R^2$ | "      | (1.70)         | (4.72)       | (-1.04)         | (-3.37)         | 0.478            | .        | .            | .        |
| Multilat. ML (19)    |        | 97.9           | 0.622        | -0.376          | -1.98           | 3.53             | 0.080    | 0.079        | 0.079    |
| $T$ -stat.,<br>$R^2$ | "      | (1.69)         | (9.46)       | (-0.86)         | (-3.41)         | 0.645            | .        | .            | .        |
| Unilateral LS (17)   |        | 4.30           | 0.636        | 0.089           | 0.093           | 1.28             | 0.377    | <b>0.369</b> | 0.370    |
| $T$ -stat.,<br>$R^2$ | "      | (1.69)         | (8.17)       | (3.19)          | (2.72)          | 0.336            | .        | .            | .        |
| Multilat. ML (19)    |        | -0.85          | 0.479        | 0.053           | 0.101           | 1.26             | 0.600    | 0.597        | 0.588    |
| $T$ -stat.,<br>$R^2$ | "      | (-0.36)        | (7.41)       | (1.88)          | (2.89)          | 0.351            | .        | .            | .        |



**Figure 8.** Water isotopes data and SAR forecasts (22): 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.

#### 4. Conclusions

In this paper we have compared unilateral and multilateral SAR models in forecasting spatial data of both lattice and polygonal type. 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 models can be consistently estimated with least squares, multilateral SAR require maximum likelihood or method of moments; in the latter cases, numerical complexity increase with the dimension of the contiguity matrices. Instead, LS is not sensitive to the size of the lattice or the number of spatial units involved and can manage even large scale systems, see Eqs. (9) and (17). Simulation experiments on small-medium scale systems have shown that ML and GM are suitable for multilateral models, but LS is preferable for the unilateral ones, especially in presence of unstable systems ( $|\phi| \geq 1$ ).

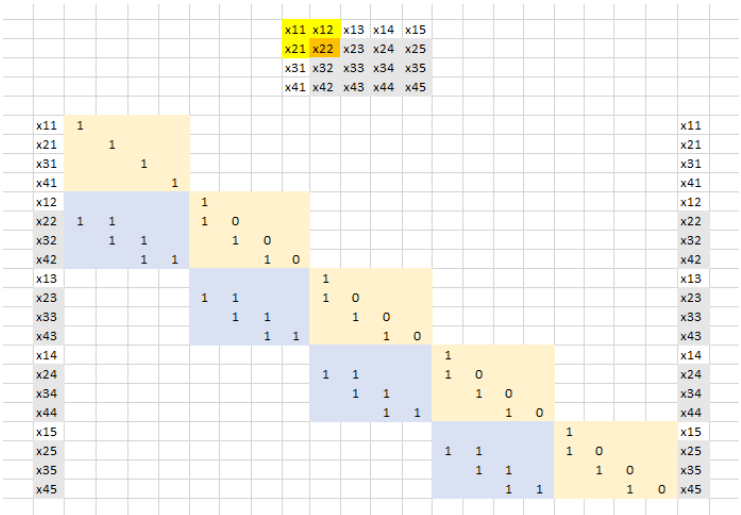
As in time series, unilateral models with data ordered in a certain direction, yield recursive (sequential) calculations, which allow simple and unconditional forecasting functions. Instead, multilateral models involve forward values which may be pre-estimated only with the reduced form (20). In the applications we have seen that out-of-sample forecasting of multilateral models benefits from the BLUP improvement (22), but the performance of the unilateral models with (21) remains better, see Tabs. 5 and 6. As regards the practical usage of SAR point forecasts, we mention the possibility to use them as alternatives to non-parametric smoothers, such as Kriging and Kernels, which are particularly sensitive to the border effects. Further, in order 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. This solution is suitable for points inside the perimeter of the investigated area, but it can also be applied to external points by using diagonal paths (e.g. NE).

These are open topics for further research; other important areas of application of triangular contiguity matrices are to space-time (ST) data; in particular, to STAR and dynamic panel-data models (see Grillenzoni, 2004 and Copiello and Grillenzoni, 2021). In the latter reference it is shown how the unidirectional specification improves the performance even of robust (M-type) estimators, designed for data with outliers. Finally, the unilateral specification is also suitable for testing the Granger causality between spatial processes (e.g. Herrera et al., 2012) because it is related to diffusion effects which follow specific directions in the space.

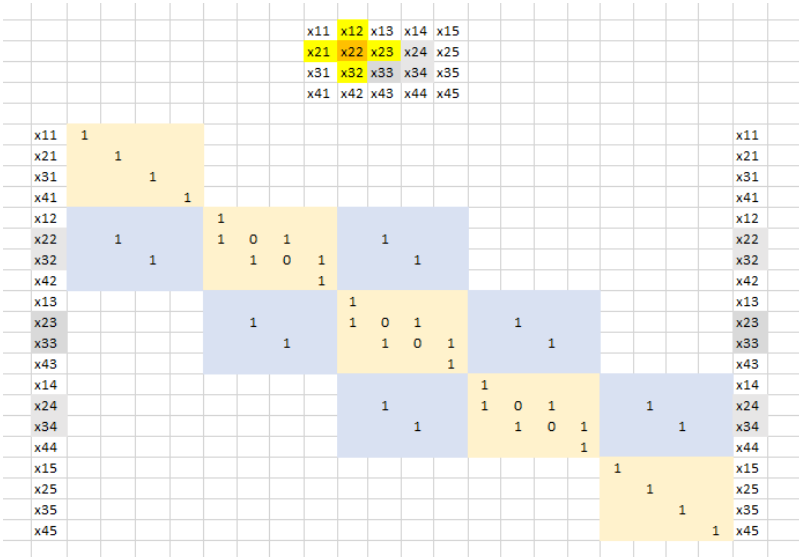
#### Appendix A. Contiguity Matrices of SAR Lattice Models

We show the derivation of the vector representation of a  $SAR(p, q, r)$  model for lattice data, where  $p$ =spatial lag,  $q$ =number of units and  $r$ =number of parameters. Consider the array  $X$  in Fig. A1 of size  $n=4$ ,  $m=5$ , and a deterministic triangular scheme  $SAR(1,3,1)$ :  $x_{ij} = (x_{i-1,j} + x_{i-1,j-1} + x_{i,j-1})$ ,  $i, j = 2 \dots 4, 5$ . Let  $x = \text{vec}_j(X)$  be the column vectorization of length  $N=20$ ; then, subject to fixed border conditions (first row and column), the vector representation becomes  $x = W_x x$ , where  $W_x$  is the contiguity matrix in Fig. A1. In general,  $W_x$  has  $m$  square blocks of size  $n$  on the main diagonal and  $m - 1$  blocks on the lower sub-diagonal, with central rows with  $q$  non-zero elements.

As regards the corresponding  $SAR(1,3,1)$  process  $y_{ij} = \alpha + \phi (y_{i-1,j} + y_{i-1,j-1} + y_{i,j-1})/3 + e_{ij}$ , with border conditions  $y_{00}=0$ , i.e.  $y_{i1} = y_{1j} = \alpha + e_{ij}$ , its vector representation becomes  $y = \alpha \mathbf{1}_N + \phi W_y \frac{1}{3} y + e$ , where the matrix  $W_y$  is obtained by equating to zero all diagonal entries of  $W_x$  in Figure A1. If  $r=q$ , the matrix  $W_y$  is decomposed into  $r$  triangular matrices which provide  $y = \alpha + \sum_{k=1}^r \phi_k W_k y + e$ . Analogously, Figure A2 shows the contiguity matrix  $W_x$  for a rook model  $SAR(1,4,1)$ ; the corresponding  $W_y$  is obtained by equating to zero all the diagonal entries.



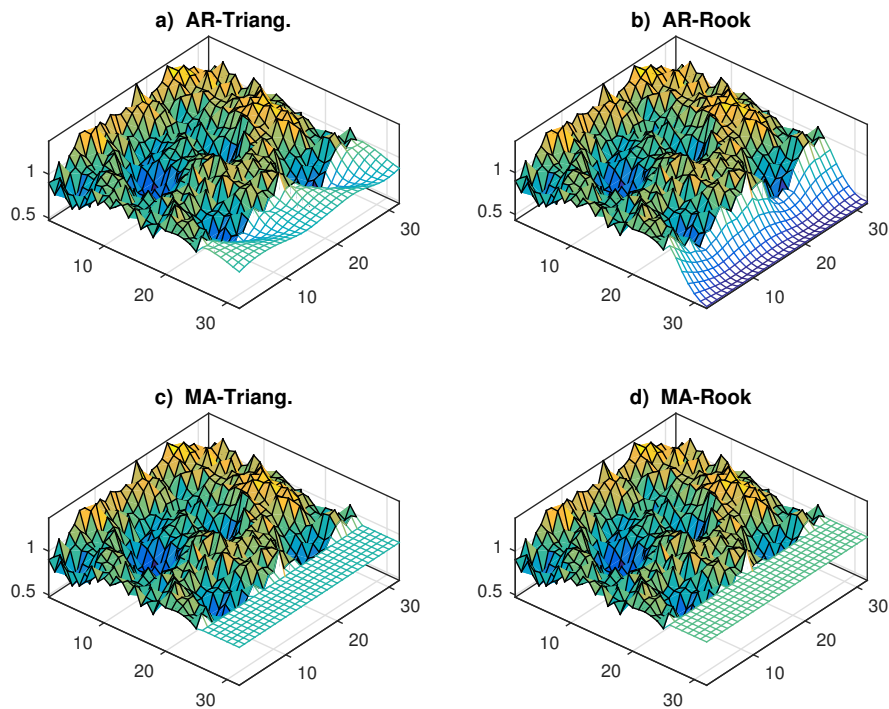
**Figure A1.** Vector representation of a triangular scheme  $x_{ij} = (x_{i-1,j} + x_{i-1,j-1} + x_{i,j-1})$  on a lattice of size 4×5, with fixed border conditions.



**Figure A2.** Vector representation of a rook model  $x_{ij} = (x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1})$  on a lattice of size 4×5, with fixed border conditions.

Appendix B. More Results on Surface Forecasting

We provide additional results of the experiment of Figure 4 and Table 2, which deals with forecasting a rough random surface (11) with models (4a,b). Figure A3 shows the forecasts of the vector algorithms (10b,c); as one can see, the results of the AR method (10b) are similar to those of the lattice function (10a), in Figure 4c,d. Instead, the forecasts of the reduced (MA) form (10c) are nearly constants in both models, due to the absence of exogenous variables, see Figure A3c,d.



**Figure A3.** Forecasts of the surface in Figure 4b provided by the models (4a,b) and the predictors (10b,c).

Table A1 provides the average values, over 10 replications of the surface (11), of the parameter estimates of the models (4a,b) and of their MAPE statistics. The results agree with those of Table 2, which regard a single replication; in particular, the forecasts of the triangular model (in lattice and vector form) are better than those of the rook scheme. The ML estimates are necessary for the multilateral model, but the unilateral one may also rely on the fast LS. In summary, the best methods, also from the graphical viewpoint, are unilateral models with the AR predictors.

**Table A1.** Average values, over 10 replications of the surface (11), of parameter estimates (and *T*-statistics) and forecast statistics of the models (4a,b) with predictors (10a-c).

| Model            | Method  | $\bar{\alpha}$ | $\bar{\phi}$ | $\bar{R}^2$ | MAPE         |
|------------------|---------|----------------|--------------|-------------|--------------|
| Triang. (10a)    | LS (2)  | 0.118 (4.87)   | 0.879 (37.0) | 0.641       | <b>0.116</b> |
| Rook (10a)       | ML (19) | 0.231 (17.3)   | 0.766 (58.2) | 0.738       | 0.183        |
| AR-Triang. (10b) | LS (9)  | 0.107 (4.78)   | 0.891 (40.6) | 0.668       | <b>0.117</b> |
| AR-Rook (10b)    | ML (19) | "              | "            | "           | 0.217        |
| MA-Triang. (10c) | LS (9)  | "              | "            | "           | 0.231        |
| MA-Rook (10c)    | ML (19) | "              | "            | "           | 0.584        |

References

Anselin, L. (2003). Spatial Econometrics; in Baltagi, B.H. (ed.), *A Companion to Theoretical Econometrics*. Blackwell Publishing, Oxford, pp 310-330.

Awang, N., Shitan, M. (2006). Estimating the parameters of the second order spatial unilateral autoregressive model. *International Journal of Statistical Sciences*, 5, 37-58.

Bao, Y., Liu, X., Yang, L. (2020). Indirect inference estimation for spatial autoregressions. *Econometrics*, 8(3), 34.

- Barbiero, T., Grillenzoni, C. (2022). Planning a novel regional methane network: Demand forecasting and economic evaluation. *Energy Conversion and Management: X*, 16, 100294.
- Basu, S., Reinsel, G.C. (1993). Properties of the spatial nilateral first-order ARMA model. *Advances in Applied Probability*, 25, 631-648.
- Bhattacharyya, B.B., Richardson, G.D., Franklin, L.A. (1997). Asymptotic inference for near unit roots in spatial autoregression. *Annals of Statistics*, 28, 173-179.
- Bustos, O., Ojeda, S., Vallejos, R. (2009). Spatial ARMA models and its applications to image filtering. *Brazilian Jour. of Probab. and Stat.*, 23(2), 141-165.
- Copiello, S., Grillenzoni, C., (2021). Robust spacetime modeling of solar photovoltaic deployment. *Energy Reports*, 7(5), 657-676.
- Fingleton, B. (2009). Spatial Autoregression. *Geograph. Anal.*, 41(4), 385-391.
- Gautam, M.K., Song, B.-Y., Shin, W.-J., Bong, Y.-S., Lee, K.-S. (2020), Datasets for spatial variation of O and H isotopes in waters and hair across South Korea. *Data in Brief*, 30, 105666.
- Goulard M., Laurent T., Thomas-Agnan C. (2017). About predictions in spatial autoregressive models: optimal and almost optimal strategies. *Spatial Economic Analysis*, 12(23), 304-325.
- Granger, C.W.J. (2012). Useful conclusions from surprising results. *J. Econom.*, 169(2), 142-146,
- Grillenzoni, C. (2004), Adaptive spatio-temporal models for satellite ecological data. *Jour. of Agric. Biol. & Environ. Statistics*, 9, 158-180.
- Grillenzoni, C. (2008), Statistics for image sharpening. *Statistica Neerlandica*, 62(2), 173-192.
- Herrera, M., Ruiz, M., Mur, J. (2012). Some examples of the use of a new test for spatial causality. *Estadística Española*, 54, 53-70.
- Jarque, C.M., Bera, A.K. (1987), A test for normality of observations and regression residuals. *International Statistical Review*, 55, 163-172.
- Kelejian, H.H., Prucha, I.R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Econom. Rev.*, 40, 509-533.
- Kelejian H.H., Prucha I.R. (2007). The relative efficiencies of various predictors in spatial econometric models containing spatial lags *Regional Science and Urban Economics*, 37, 363-374.
- Lee, L.F., (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, 18, 252-277.
- Lee, L.F., (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22, 307-335.
- LeSage, J.P., Pace R.K. (2004). Models for spatially dependent missing data. *Journal of Real Estate Finance and Economics*, 29, 233-254.
- LeSage J.P., Pace R.K. (2008). Spatial econometric models, prediction. In Shekhar, S., Xiong, H. (eds), *Encyclopedia of Geographical Information Science*, 1095-1100. Springer: New York.
- LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group.
- Mojiri, A., Waghei, Y., Nili-Sani, H.R., Mohtashami Borzadaran, H.R. (2018). The stationary regions for the parameter space of unilateral second-order spatial AR model. *Random Operators and Stochastic Equations*, 26(3), 185-191.
- Mojiri, A., Waghei, Y., Nili-Sani, H.R., Mohtashami Borzadaran, H.R. (2021). Non-stationary spatial autoregressive modeling for the prediction of lattice data. *Communications in Statistics - Simulation and Computation*, 100, 1-13.
- Moreiras Reynaga, D.K., Millaire, J.-F., Chávez Balderas, X., Román Berrelleza J.A., López Luján, L., Longstaffe, F.J. (2021). Building Mexican isoscapes: Oxygen and hydrogen isotope data of meteoric water sampled across Mexico. *Data in Brief*, 36, 107084.
- Sheffield, S. (2022). What is a random surface? Proc. Int. Cong. Math. 2022, International Mathematical Union, EMS Press.
- Sjodin B. (2017), Generate Random Surfaces in COMSOL Multiphysics<sup>(R)</sup>. Avalibale at: <https://www.comsol.com/blogs/to-generate-random-surfaces-in-comsol-multiphysics>
- Tjøstheim, D. (1983). Statistical Spatial Series Modelling II: Some further Results in Unilateral Processes. *Advances in Applied Probability*, 15, 562-684.

USGS (2017), U.S. Geological Survey, Digital Elevation Models (DEMs). Available at: <https://www.sciencebase.gov/catalog/item/59600000488d680000400000>  
Yao, Q., Brockwell, P.J. (2006), Gaussian Maximum Likelihood Estimation for ARMA Models II: Spatial Processes. *Bernoulli*, 13, 403-429.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.