Check for updates

Received: 23 July 2021 | Accepted: 23 August 2021

DOI: 10.1002/env.2705

RESEARCH ARTICLE

WILEY

Estimation of the spatial weighting matrix for regular lattice data—An adaptive lasso approach with cross-sectional resampling

Miryam S. Merk[1] | **Philipp Otto**[2]

[1]Chairs of Statistics and Econometrics, University of Göttingen, Göttingen, Germany

[2]Institute of Cartography and Geoinformatics, Leibniz University Hannover, Hannover, Germany

Correspondence
Miryam S. Merk, Chairs of Statistics and Econometrics, University of Göttingen, Humboldtallee 3, 37073 Göttingen, Germany.
Email: miryamsarah.merk@uni-goettingen.de

Abstract

Spatial autoregressive models typically rely on the assumption that the spatial dependence structure is known in advance and is represented by a deterministic spatial weights matrix, although it is unknown in most empirical applications. Thus, we investigate the estimation of sparse spatial dependence structures for regular lattice data. In particular, an adaptive least absolute shrinkage and selection operator (lasso) is used to select and estimate the individual nonzero connections of the spatial weights matrix. To recover the spatial dependence structure, we propose cross-sectional resampling, assuming that the random process is exchangeable. The estimation procedure is based on a two-step approach to circumvent simultaneity issues that typically arise from endogenous spatial autoregressive dependencies. The two-step adaptive lasso approach with cross-sectional resampling is verified using Monte Carlo simulations. Eventually, we apply the procedure to model nitrogen dioxide (NO_2) concentrations and show that estimating the spatial dependence structure contrary to using prespecified weights matrices improves the prediction accuracy considerably.

KEYWORDS

adaptive lasso, regular lattice data, spatial weights matrix

1 | INTRODUCTION

Modeling dependencies in spatial or spatio-temporal data theoretically requires accounting for $n^2 - n$ potential interactions between the n spatial units of a sample, which raises two important challenges. First, the number of unknown connections grows quadratically with the sample size, which can become computationally demanding especially in the context of big data. Second, the identification suffers from an incidental parameter problem because the number of unknown parameters exceeds the sample size n.

In geostatistics, it is therefore typically assumed that the spatial dependence can be described by a spatial covariance function, which is independent of the spatial location. This covariance function usually depends on additional parameters but in a much lower dimension than the number of observations. Several choices of this covariance function and its estimation have been discussed for both stationary (e.g., Cressie & Huang, 1999; Gneiting, 2002; Ma, 2003) and nonstationary spatial dependence (e.g., Nott & Dunsmuir, 2002; Nychka et al., 2002; Paciorek & Schervish, 2006; Risser & Calder, 2015;

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Environmetrics* published by John Wiley & Sons, Ltd.

Sampson & Guttorp, 1992). Under certain assumptions (e.g., as long as the misspecified covariance function is compatible with the true covariance function), the asympotic kriging performance is not affected if the covariance is misspecified or estimated (e.g., Putter & Young, 2001; Stein, 1988). Nevertheless, the true underlying spatial dependence is typically unknown (e.g., Reich & Fuentes, 2012). Moreover, modeling spatial data with Gaussian Markov random fields assuming a certain covariance function is computationally demanding, especially in the case of large data sets (e.g., Simpson et al., 2012).

Similarly, in spatial econometrics, the spatial dependence is described by the multiple of an unknown scalar and a prespecified weighing matrix to reduce dimensionality. The focus of this article is on the latter case and spatial autoregressive models in particular. For these classical econometric approaches, the unknown scalar parameter reflects the strength of the spatial dependence. The prespecified spatial weights matrix is often based on distance measures between the individual locations, which involves geographic proximity (cf., Anselin, 1988; Cliff & Ord, 1973; Ertur & Koch, 2007) or other types of adjacency, such as social (see, e.g., Cohen & Tita, 1999) or economic (see, e.g., Besner, 2002; Bodson & Peeters, 1975) characteristics. The most common specifications involve q-nearest neighbors, binary contiguity, and inverse distance matrices. Alternatively, geographically weighted regression models can capture spatial dependence by coefficients that are allowed to vary across space assuming a certain weight structure (cf., Fotheringham et al., 1997; Harris et al., 2010; Leung et al., 2000). Geniaux and Martinetti (2018) combined both approaches, that is, geographically weighted regression and spatial autoregressive models. Of course, the results of such models are only valid if the weights matrix has been correctly specified. However, in most applications the true underlying dependence structure is unknown (see also Gibbons & Overman, 2012). Hence, a priori selections of the spatial weights matrix have raised considerable criticism, especially because different weighting matrices can lead to very different and, therefore, contradicting estimation results (cf., Mizruchi & Neuman, 2008; Smith, 2009). In particular, the results must be interpreted conditioned on the assumed weighting scheme (cf., Debarsy et al., 2012).

Consequently, a variety of alternative approaches have been suggested that involve parametric or semi-parametric estimation procedures (see, e.g., Pinkse et al., 2002) or matrix selection from a set of candidate schemes based on goodness-of-fit criteria, such as maximized log-likelihood values (cf., Stakhovych & Bijmolt, 2009). Alternatively, Bhattacharjee and Jensen-Butler (2013) proposed estimating the weights from an estimated spatial autocovariance matrix. This approach uses panel data under the constraint of symmetry and a finite sample size of n. Hence, the number of unknown spatial connections or weights is reduced by half. In addition to symmetry, other identifying assumptions have been proposed to estimate the individual spatial weights. Basak et al. (2018) suggested estimating a triangular spatial weighting matrix from the covariance matrix under the assumption of causal ordering in acyclic networks. Moreover, the least absolute shrinkage and selection operator (lasso) originally suggested by Tibshirani (1996) has been used to estimate sparse spatial weighting matrices. Ahrens and Bhattacharjee (2015) suggested identifying an approximately sparse spatial weights matrix using a two-stage lasso estimation procedure. Lam and Souza (2019) proposed estimating the spatial weights matrix from a linear combination of different specifications adjusted using a potentially sparse matrix. Moreover, Otto and Steinert (2018) discussed an adaptive lasso procedure to estimate the spatial dependence and an unknown number of possible structural breaks simultaneously. Wheeler (2009) introduced a lasso estimation procedure for variable selection in geographically weighted regression models. In the geostatistical context, Krock et al. (2020) proposed to apply a graphical lasso approach instead of implying a certain dependence structure by a suitable covariance function. This is particularly useful for nonstationary spatial data when independent observations of the same spatial field are present. Recently, this approach has been extended to multivariate data (cf., Krock et al., 2021).

However, most of these approaches have in common that they require spatio-temporal data, where the spatial dependence is not varying over time, or repeated (independent) measurements of the spatial field. To the best of our knowledge, little attention has been paid to purely spatial models where the number of spatial links exceeds the sample size and where no dependence structure must be assumed in advance. Zhu et al. (2010) proposed penalized maximum likelihood (ML) estimators to select covariates and a neighborhood structure for lattice data in the context of conditional and simultaneous spatial error models. More precisely, the unknown spatial weights matrix is represented by a linear combination of individual matrices reflecting the neighborhood sets of different orders. Bhattacharjee et al. (2012) proposed estimating the spatial weighting matrix in hedonic pricing models under the structural assumption of symmetry. Merk and Otto (2021) investigated the estimation of a triangular spatial weighting matrix in the context of directional spatial autoregressive dependencies.

Contrary to these approaches, we suggest estimating all relevant spatial connections for purely spatial regular lattice data based on cross-sectional resampling. Thus, instead of estimating the spatial dependence structure over an entire (possibly high-dimensional) dataset, smaller exchangeable subsamples or blocks are repeatedly drawn from the spatial field to estimate the spatial dependence. The spatial weights and regressors are selected and estimated by the adaptive

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

lasso approach proposed by Zou (2006) as an extension of the original lasso that enables individual shrinking. To avoid endogeneity issues arising due to the dependent variable simultaneously serving as an explanatory variable, we propose a two-step approach incorporating instrumental variables (IVs) in the first step.

The rest of this article is organized as follows. Section 2 describes the theoretical framework, including the spatial autoregressive regression model, regularity assumptions, and the two-step adaptive lasso estimation procedure. The results obtained from the Monte Carlo simulations on the performance of the spatial weights estimates are reported in Section 3. Furthermore, we apply the two-step adaptive lasso approach to model nitrogen dioxide (NO_2) concentrations. Finally, the last section presents the conclusion.

2 | THEORETICAL MODEL

Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ be a univariate spatial stochastic process at known locations $\boldsymbol{s}$ in the set D, which is a subset of the d-dimensional real numbers $\mathbb{R}^d$, where $\boldsymbol{s}$ may vary discretely or continuously over D. Spatial data can be generally classified into three categories, namely, spatial point patterns, geostatistical (or continuous) processes, and lattice data (Cressie, 1993). For the latter case, the process is observed on regular or irregular grids in the two-dimensional space. In this article, the focus is on regular lattices that cover all kinds of images, like remotely sensed or microscopic imagery. Hence, D is a subset of $\mathbb{Z}^2$ and the individual locations $\{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$ represent a discrete collection of n equally sized and shaped grid cells. Due to the natural ordering and predictable relative placement, regular lattice data are an obvious choice for modeling neighborhood relations or dependence structures. Typical specifications of this spatial dependence structure involve a fixed number q of nearest neighbors, like the four first-order neighbors to the north, south, east, and west. However, the true underlying structure of the data-generating process is usually unknown (cf., Gibbons & Overman, 2012). In particular, there are up to $n^2 - n$ potential connections between the n locations of a sample. We consider these links as unknown parameters along with the degree of dependence at each true link.

While the spatial autoregressive dependencies between the individual responses denoted by the vector $\boldsymbol{Y} = (Y(\boldsymbol{s}_i))_{i=1,\dots,n}$ are commonly characterized by prespecified spatial weighting matrices, we explicitly assume that $\mathbf{W} = (w_{ij})_{i,j=1,\dots,n}$ is unknown. In other words, the ith row of $\mathbf{W}$ reflects how the corresponding observation on the dependent variable $Y(\boldsymbol{s}_i)$ is influenced by observations in all other regions. For example, $w_{ij} > 0$ indicates that $Y(\boldsymbol{s}_i)$ depends positively on $Y(\boldsymbol{s}_j)$, while $w_{ij} = 0$ corresponds to the (conditionally) independent case. Thus, the model equation of a spatially autoregressive response variable can be specified as

$$\boldsymbol{Y} = \mathbf{W}\boldsymbol{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

or as the reduced form

$$\boldsymbol{Y} = (\mathbf{I} - \mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}), \tag{2}$$

where $\mathbf{X}$ is an $n \times k$ matrix comprising k explanatory variables, $\mathbf{I}$ is the identity matrix of dimension n, and $\boldsymbol{\epsilon}$ is an n-dimensional vector of independently and identically distributed residuals with zero mean and positive finite variance σ^2. Moreover, let $\boldsymbol{y} = (y(\boldsymbol{s}_i))_{i=1,\dots,n}$ be the observed vector of $\boldsymbol{Y}$. In contrast to classical models, we do not incorporate a spatial autoregressive coefficient ρ because $\mathbf{W}$ and ρ are not separately identified if the spatial weights matrix is unknown (Bhattacharjee & Jensen-Butler, 2013; Gibbons & Overman, 2012). In the following sections, we will explain the estimation method in detail. To give the reader a structured overview of this multi-step approach, the entire procedure is shown schematically in Figure 1.

2.1 | Cross-sectional resampling

In contrast to spatio-temporal data, where the spatial dependence structure may be recovered by repeated observations over time, we propose using cross-sectional resampling for purely spatial processes. That is, we randomly draw r locations $\boldsymbol{s}_{*_1}, \dots, \boldsymbol{s}_{*_r}$ without replacement and identify the respective m nearest locations for each sampled location. Hence, there are r sampled blocks of size $m + 1$ consisting of $\boldsymbol{s}_*$ and its m nearest locations. Let $\tilde{\boldsymbol{Y}}$ denote the corresponding vector of sampled responses $(Y(\boldsymbol{s}_{*_1}), \dots, Y(\boldsymbol{s}_{*_r}))'$. To estimate the spatial weights from this sample, we assume that the data have the following two properties.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

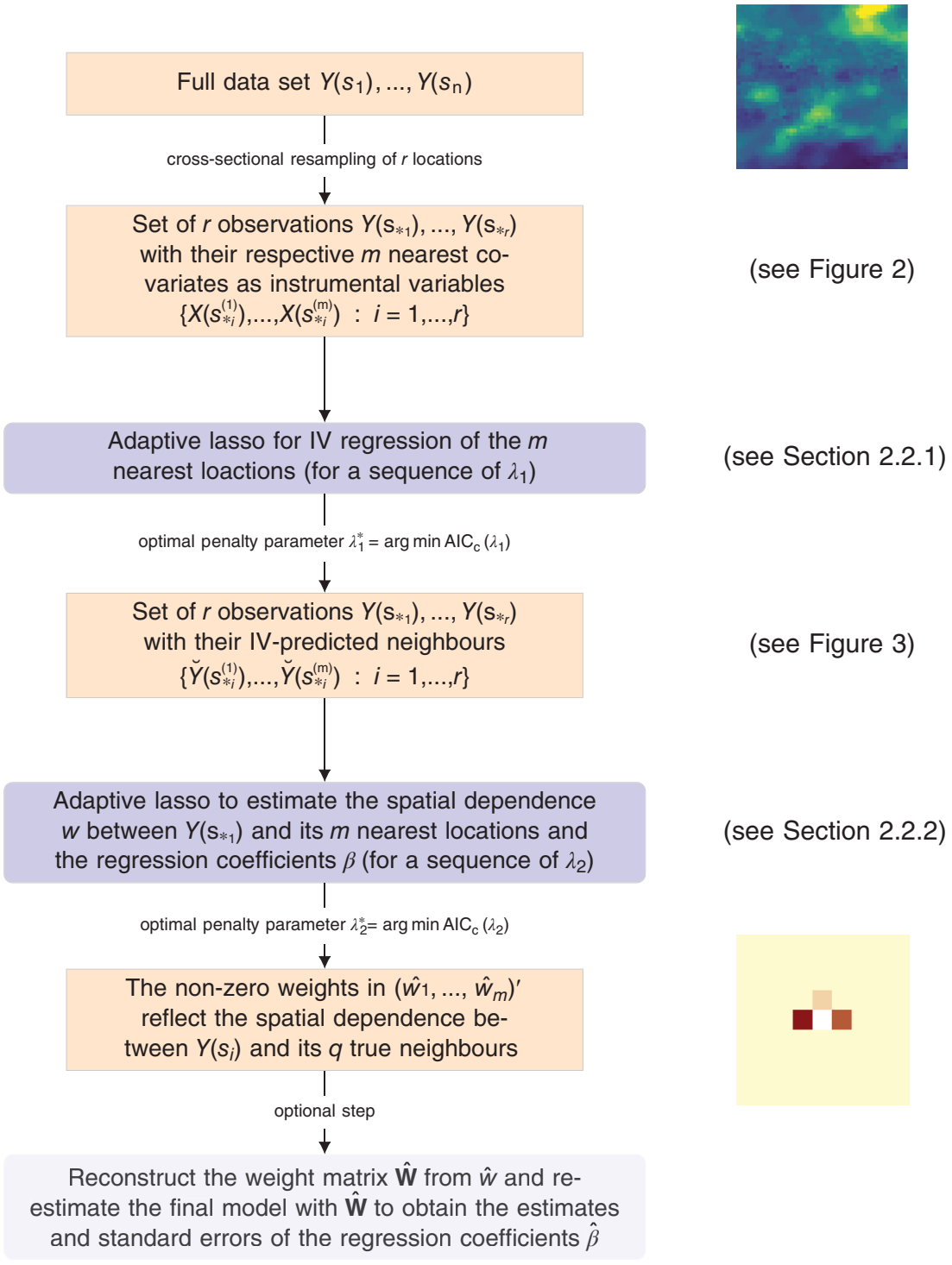


FIGURE 1 Schematic representation of the estimation procedure. The example data set on the right-hand side is taken from Table 4

First, the set of the q true (but unknown) links of the ith location N_i^* is a subset of the m nearest locations of $\boldsymbol{s}_i$, which is denoted by N_i. Accordingly, not all locations in N_i actually have an impact on $\boldsymbol{s}_i$, but only the q true links, which constitute the nonzero elements in the spatial weights matrix $\mathbf{W}$. Hence, the proportion of zero entries in $\mathbf{W}$ is always greater than $mn/(n^2 - n)$. This assumption implicitly draws on Tobler's first law of geography (Tobler, 1970) that nearby observations are more alike than distant ones.

Second, the conditional distributions of the response variables given the observations of their m nearest locations, that is, $f_{Y(\boldsymbol{s}_i)|\{Y(\boldsymbol{s}_i^{(j)}):j=1,\dots,m\}}$ with $\boldsymbol{s}_i^{(j)}$ being the jth nearest location of $\boldsymbol{s}_i$, is the same for all locations. This implies

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that the spatial dependence structure is constant across space and the process is exchangeable across these blocks. Note that this assumption is not restrictive because it is assumed for each stationary geostatistical process or spatial econometric models with uniform weighting structures. Of course, the conditional distribution varies for the locations along the edges of the random field, where the m nearest locations are not completely identified. In consequence, these bordering locations are not sampled but they may be considered as part of the neighborhood subsets of other locations.

For illustration, consider the following example depicted in Figure 2 with $n = 10 \times 10$ locations and two randomly sampled locations. For both of these locations, the $q = 3$ neighbors to the north, north-east, and east are contained in the respective $m = 8$ nearest locations. The observed dependence is the same for all locations, except for the ones along the border.

As a consequence of the above assumptions, the spatial weighting matrix is composed of a unique parameter vector $\boldsymbol{w} = (w_1, \ldots, w_m)'$ that appears in each row of $\mathbf{W}$ (except for the bordering locations) reflecting how each location is affected by its m nearest locations. The choice of m thereby determines the number of potential neighbors considered or equivalently, sets a distance-cutoff after which the spatial dependence is supposed to be zero. Large values of m ensure that all true neighbors are covered and simultaneously induce sparsity of the spatial weights vector $\boldsymbol{w}$. More precisely, the set of potential neighbors can be chosen permissively large (beyond the expected distance-cutoff) due to the lasso's

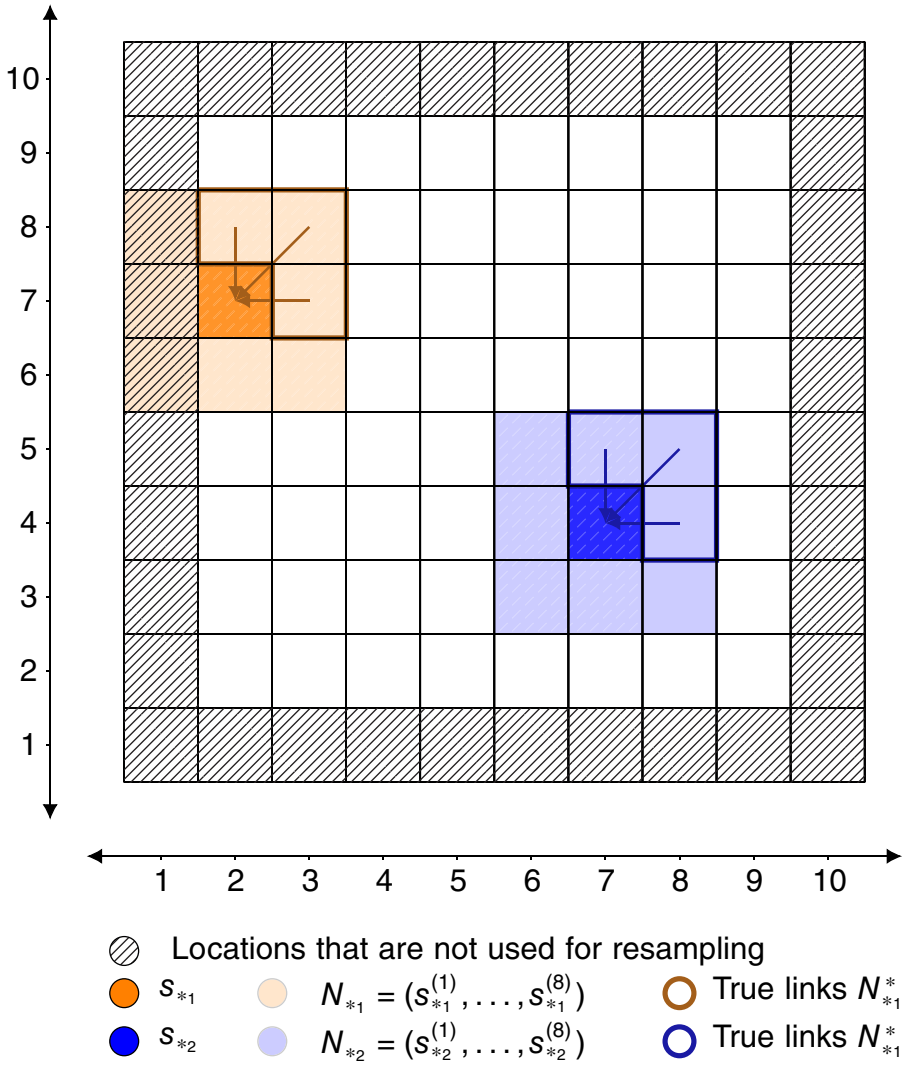


FIGURE 2 Cross-sectional resampling (first step). Two sample locations $\boldsymbol{s}_{*_1}$ and $\boldsymbol{s}_{*_2}$ are highlighted in orange and blue, respectively, including their eight nearest locations representing the sets N_i. The true links, N_i^*, are drawn by arrows. The setting has an anisotropic dependence from the north-east

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

property to select the true model from a large set of potential covariates, which is explained in more detail in the following sections.

2.2 | Two-step lasso estimator

Classical least squares estimators are inconsistent in the context of endogenous spatial dependencies. Consistent procedures, such as the ML approach (cf., Anselin, 1988; Lee, 2004; Ord, 1975) or generalized method of moments (cf., Kelejian & Prucha, 1999), circumvent simultaneity issues arising due to interdependence among neighboring locations. Alternatively, Kelejian and Prucha (1998) proposed a two-stage least squares estimation procedure, which is computationally less demanding than the ML approach. More precisely, the right-hand side endogenous variables are replaced by predictions of the response variable obtained by an IV regression. In addition, Ahrens and Bhattacharjee (2015) adapted a two-step estimation procedure for spatio-temporal data to estimate spatial weights, where the number of instruments can be larger than the sample size, assuming approximate sparsity.

Our estimation procedure is based on a two-step adaptive lasso approach that allows recovering the spatial dependence structure. The first step involves finding suitable instruments to predict the endogenous spatially lagged variables. In the second step, the response variables are replaced with their predicted values to circumvent simultaneity issues arising from simultaneous spatial autoregressive dependencies, and all parameters of the full model are estimated. Below, we will be more specific about these two steps.

2.2.1 | First step: IV regression

In the first step, we estimate $\tilde{\boldsymbol{Y}}$ using l instruments by a linear model, that is,

$$\tilde{\boldsymbol{Y}} = \tilde{\mathbf{Z}}\boldsymbol{\theta} + \boldsymbol{\nu}, \tag{3}$$

where $\tilde{\mathbf{Z}}$ is the corresponding $r \times l$ matrix of instruments associated with the r sampled locations, and $\boldsymbol{\theta}$ is the corresponding l-dimensional vector of coefficients with $l \geq 1$. Moreover, $\boldsymbol{\nu}$ denotes the r-dimensional vector of the IV random errors. The fitted values of this regression are used to instrument the spatially lagged variables (i.e., the m nearest locations) in the second step.

Of course, the instruments must fulfill certain requirements. It is assumed that the instruments are independent of the regression residuals but correlated with the response variable or endogenous variables in general.

In contrast to the proposal by Kelejian and Prucha (1998) that the instruments should be composed of exogenous regressors $\mathbf{X}$ and first and higher-order spatial lags ($\mathbf{WX}, \mathbf{W}^2\mathbf{X}, \dots$) using prespecified spatial weights, the spatial weights matrix is unknown in our case. For that reason, we take all exogenous regressors at the sampled locations and at their m nearest locations as IVs. Consequently, there are $l = k + mk$ instruments with $m \ll n$. In particular, $\tilde{\mathbf{Z}} = (\tilde{\mathbf{X}}, \overset{\approx}{\mathbf{X}})$, where $\tilde{\mathbf{X}} = (\boldsymbol{X}(\boldsymbol{s}_{*1}), \dots, \boldsymbol{X}(\boldsymbol{s}_{*r}))'$ corresponds to exogenous variables observed at the sampled locations and $\overset{\approx}{\mathbf{X}}$ comprises the exogenous variables at the m nearest locations, that is,

$$\overset{\approx}{\mathbf{X}} = \begin{pmatrix} \boldsymbol{X}(\boldsymbol{s}_{*_1}^{(1)}) & \dots & \boldsymbol{X}(\boldsymbol{s}_{*_1}^{(m)}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{X}(\boldsymbol{s}_{*_r}^{(1)}) & \dots & \boldsymbol{X}(\boldsymbol{s}_{*_r}^{(m)}) \end{pmatrix} \tag{4}$$

with $\boldsymbol{X}(\boldsymbol{s}_{*}^{(1)}), \dots, \boldsymbol{X}(\boldsymbol{s}_{*}^{(m)})$ being the exogenous variables at the m nearest locations of $\boldsymbol{s}_{*}$. For simplicity, we assume that the size of the subset of potential neighbors m is the same in the first and second step.

Finally, the first-step adaptive lasso estimator is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\| \tilde{\boldsymbol{Y}} - \tilde{\mathbf{Z}}\boldsymbol{\theta} \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\psi}_1 \circ \boldsymbol{\theta} \right\|_1 . \tag{5}$$

The objective function consists of two parts, where the first term minimizes the residual sum of squares between observations on the response variables and corresponding instruments. The second term imposes a penalty on the

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

absolute size of the coefficients, where λ_1 denotes the tuning parameter and $\boldsymbol{\psi}_1$ corresponds to the adaptive weights. Both λ_1 and $\boldsymbol{\psi}_1$ are described in more detail in the following subsection.

Equivalent to the choice of the m nearest locations, the set of instruments can be permissively large, since the adaptive lasso approach is selection consistent and shrinks irrelevant parameters toward zero (see Zou, 2006). This motivates the application of the lasso estimation in the first step.

Using this model, we can predict the vector $\boldsymbol{Y}$ (except for the locations at the edges of the random field). Eventually, these predicted values denoted by $\check{Y}(\boldsymbol{s}_i)$ are used to instrument the spatially lagged variables in the second step to avoid endogeneity.

2.2.2 | Second step: Full model

Let $\check{\mathbf{Y}}$ denote the $r \times m$ matrix of predicted values obtained from the first-step estimation. That is,

$$\check{\mathbf{Y}} = \begin{pmatrix} \check{Y}(\boldsymbol{s}_{*_1}^{(1)}) & \dots & \check{Y}(\boldsymbol{s}_{*_1}^{(m)}) \\ \vdots & \ddots & \vdots \\ \check{Y}(\boldsymbol{s}_{*_r}^{(1)}) & \dots & \check{Y}(\boldsymbol{s}_{*_r}^{(m)}) \end{pmatrix} \tag{6}$$

with $\check{Y}(\boldsymbol{s}_*^{(1)}), \dots, \check{Y}(\boldsymbol{s}_*^{(m)})$ being the IV-predictions of the response variable at the m nearest locations of $\boldsymbol{s}_*$.

In the second step, we estimate the full model given by (1) but replacing the endogenous variable on the right-hand side with $\check{\mathbf{Y}}$. To be precise, the second-step adaptive lasso estimator is given by

$$\begin{aligned} (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{w}})' = \arg\min_{(\boldsymbol{\beta},\boldsymbol{w})'} \left\| \tilde{\boldsymbol{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta} - \check{\mathbf{Y}}\boldsymbol{w} \right\|_2^2 + \lambda_2 \left(\left\| \boldsymbol{\psi}_{2,\beta} \circ \boldsymbol{\beta} \right\|_1 + \left\| \boldsymbol{\psi}_{2,w} \circ \boldsymbol{w} \right\|_1 \right), \\ \text{s.t. } \boldsymbol{w} \geq 0 \text{ and } \|\boldsymbol{w}\|_1 < 1, \end{aligned} \tag{7}$$

where the first term minimizes the residual sum of squares between observations on the response variable and exogenous variables, which consist of explanatory variables $\tilde{\mathbf{X}}$ and first-step predictions of the endogenous variable $\check{\mathbf{Y}}$.

The constraint $\|\boldsymbol{w}\|_1 < 1$ ensures that $||\hat{\mathbf{W}}||_\infty < 1$ and, thus, $(\mathbf{I} - \mathbf{W})$ is invertible and the model is well-defined and stationary. The locations are only sampled from the center (excluding the edges of the random field), such that all entries of $\check{\mathbf{Y}}$ are known. The blocks or sets of the m nearest locations of the sampled locations can potentially overlap. Biscio and Waagepetersen (2019) showed the consistency of subsampling-based statistics having an additive structure if the number of blocks is going to infinity regardless of whether the blocks are overlapping or disjointed.

In Figure 3, we illustrate the locations and values that are used in the second step (i.e., the observation $Y(\boldsymbol{s}_{*_i})$ and the IV-predicted values $\check{Y}(\boldsymbol{s}_{*_i}^{(1)}), \dots, \check{Y}(\boldsymbol{s}_{*_i}^{(m)})$ for $i \in \{1, 2\}$ sample locations).

The second term of the first and second-step estimation corresponds to the lasso penalty term that penalizes the absolute size of the coefficients in $\boldsymbol{\theta}$ and $(\boldsymbol{\beta}, \boldsymbol{w})'$, respectively. The nonnegative regularization or tuning parameters λ_1 and λ_2 are obtained based on information criteria. In particular, we focus on the corrected Akaike information criterion (AIC_c) suggested by Ninomiya and Kawano (2016) for lasso regression. However, because the large and small elements in the regression coefficients are equally driven to zero, the regular lasso yields excessively penalized large coefficients. Thus, the ℓ_1 penalty additionally consists of individual weights such that the regression coefficients may be penalized individually (cf., Zou, 2006). The weights of the first and second steps are given by the l-dimensional vector $\boldsymbol{\psi}_1 = 1/|\hat{\boldsymbol{\theta}}_0|$, the l-dimensional vector $\boldsymbol{\psi}_{2,\beta} = 1/|\hat{\boldsymbol{\beta}}_0|$ and the m-dimensional vector $\boldsymbol{\psi}_{2,w} = 1/|\hat{\boldsymbol{w}}_0|$. The individual vectors $\hat{\boldsymbol{\theta}}_0$, $\hat{\boldsymbol{\beta}}_0$, and $\hat{\boldsymbol{w}}_0$ represent prior parameter estimates that are obtained by a ridge regression. The ridge regression exploits the proposed two-step estimation procedure in order to avoid endogeneity issues due to the simultaneity. Hence, the sampled observations are regressed on their respective spatially lagged exogenous variables in the first step and on predictions of the spatially lagged dependent variables in the second step, respectively.

Finally, recall that m (i.e., the number of potential links) can be chosen much larger than q (i.e., the number of true links), because the adaptive lasso shrinks irrelevant links to zero. It just has to be ensured that there are sufficiently many observations in the center, because increasing values of m lead to broader regions at the borders of the spatial field that have to be left out.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

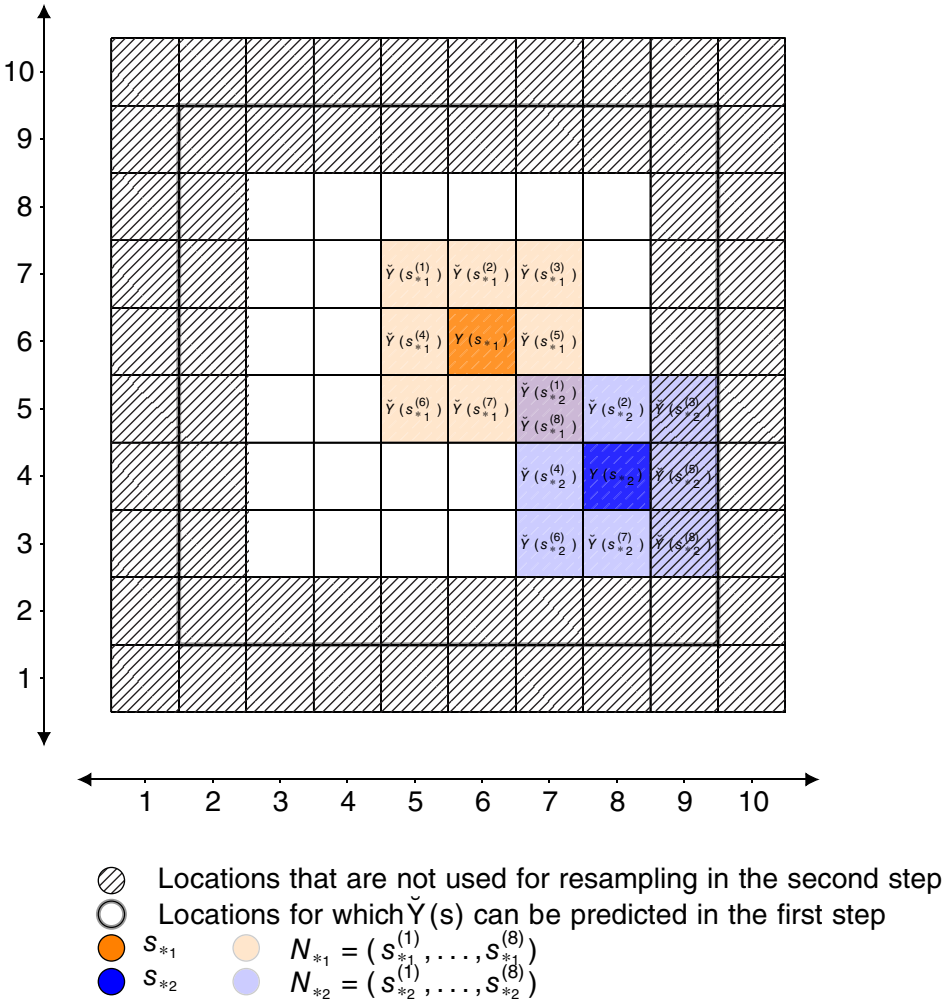


FIGURE 3 Cross-sectional resampling (second step). Two sample locations $\boldsymbol{s}_{*_1}$ and $\boldsymbol{s}_{*_2}$ are highlighted in orange and blue, respectively, including their eight nearest locations representing the sets N_i. The locations for which $\check{Y}(\boldsymbol{s})$ could be predicted by the instrumental variables in the first step coincide with all locations that are used for resampling in the first step (Figure 2)

Moreover, it is important to mention that the computational complexity does not depend on the sample size n but on the number of sampled locations r. Thus, it is potentially scalable and useful for big geospatial data (if r is chosen smaller than n).

3 | SIMULATION STUDY

In the following subsections, we analyze whether the spatial dependence structure can be recovered using the aforementioned procedure. More precisely, we critically examine the performance of the estimators with respect to (a) the number of cross-sectional resampling replications r, (b) the number of potential neighbors m, and (c) the number of the true links q, where the ratio of true to potential neighbors reflects the extent of the sparsity of the spatial weights vector. Moreover, we analyze the performance of the approach for two distinct types of spatial dependence, namely, isotropic and anisotropic processes.

3.1 | Considered settings

For both specifications of the spatial dependence, we consider a 30×30 regular lattice with $n = 900$ grid cells. Regarding the spatial structure of the locations, two different weighting schemes are employed. The first specification corresponds

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to an isotropic spatial process (Case A) commonly known as the Queen's matrix, where each location is equally affected by its $q = 8$ nearest neighbors with which it either shares a common edge or a vertex, namely,

$$w_{ij} = \begin{cases} 1/8 \cdot c, & \text{if } j \in N_i^*(q = 8) \wedge i \neq j, \\ 0, & \text{else.} \end{cases} \tag{8}$$

In contrast, the second specification corresponds to an anisotropic spatial process (Case B) with each location being equally affected by its two neighbors to the east and south-east, namely,

$$w_{ij} = \begin{cases} 1/2 \cdot c, & \text{if } j \in N_i^*(q = 2) \wedge i \neq j, \\ 0, & \text{else.} \end{cases} \tag{9}$$

Moreover, each row of $\mathbf{W}$ sums to a constant c, reflecting the strength of the spatial dependence, which is consequently constant over all locations. Different degrees of positive spatial autoregressive dependence are applied, namely, $c \in \{0.5, 0.7, 0.9\}$. For simplicity and without loss of generality, we consider $k = 1$ exogenous regressor. We perform 1000 Monte Carlo replications simulating the process

$$\boldsymbol{Y} = (\mathbf{I} - \mathbf{W})^{-1}(\boldsymbol{X}\beta + \boldsymbol{\epsilon}), \tag{10}$$

where both $\boldsymbol{X} = (X_i)_{i=1,\dots,n}$ and $\boldsymbol{\epsilon}$ are drawn from the Gaussian distribution (i.e., $X_i \sim \mathcal{N}(0, 1)$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2 = 1)$), and $\beta = 1$.

For estimating the spatial dependence structure, different sizes m are considered that ensure the sparsity of the spatial weights vector (i.e., $q < 0.5m$), namely, $m \in \{24, 48\}$ for Case A and $m \in \{8, 24, 48\}$ for Case B. In addition, three different replication levels r are employed to assess whether the estimates improve with the number of observations. The minimum number is $r_{min} = 30$, the maximum number $r_{max} = (\sqrt{n} - \sqrt{(m+1)} + 1)^2$ includes all observations of the square lattice (except for the edges) and the medium number is $r_{med} = \lfloor (r_{min} + r_{max})/2 \rfloor$.

3.2 | Recovery of the dependence structure

Figure 4 illustrates how often each connection is identified as being nonzero for the isotropic setting (Case A). In general, the proportion of correctly identified neighbors or nonzero spatial connections increases with the number of replications r. The four true neighbors that are horizontally or vertically located tend to be selected more often than the four diagonal neighbors. This might be due to spatial spillover effects that occur because the eight first-order neighbors of $\boldsymbol{s}_i$ are themselves influenced by their eight nearest neighbors, the second-order neighbors of $\boldsymbol{s}_i$, and so on. Thus, the true neighbors also transmit spillover effects from higher-order neighbors that decrease strictly monotonically with increasing order. For example, if $m = 24$, the horizontal and vertical neighbors possess the same number of second-order neighbors as the diagonal connections (i.e., eight), but their spillovers from the third-order neighbors are higher.

Figure 5 depicts the recovery rates for the anisotropic setting (Case B). As in the previous specification, the number of correctly identified neighbors increases with the number of replications. However, in contrast to the first specification (isotropic setting, Case A), the recovery frequencies are much higher for the correctly identified $q = 2$ nonzero connections. For instance, if $r \in \{r_{med}, r_{max}\}$, both true neighbors are identified across all Monte Carlo iterations. Thus, the identification of the true neighbors improves with the degree of the sparsity in $\boldsymbol{w}$ for fixed numbers of potential neighbors m.

As in Case A, the identification also suffers from spatial spillover effects, especially in the case of strong spatial dependence (i.e., $c \in \{0.7, 0.9\}$). In particular, the zero connections that are located in the east and south-east direction of the true connections are more frequently falsely selected than other zero elements if $m = 24$. Hence, the second-order neighbors are erroneously assumed to be nonzero connections because they affect the first-order neighbors. The effect from these second-order neighbors thereby increases with the strength of spatial dependence c due to the lasso's property to select irrelevant predictors that are highly correlated with the true predictors (eg., Zhao & Yu., 2006). Increasing the number of nearest locations (i.e., $m = 48$) mitigates the influence from second-order neighbors but slightly raises the frequency of falsely selected third-order neighbors. However, the frequency of falsely selected higher-order neighbors decreases with

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

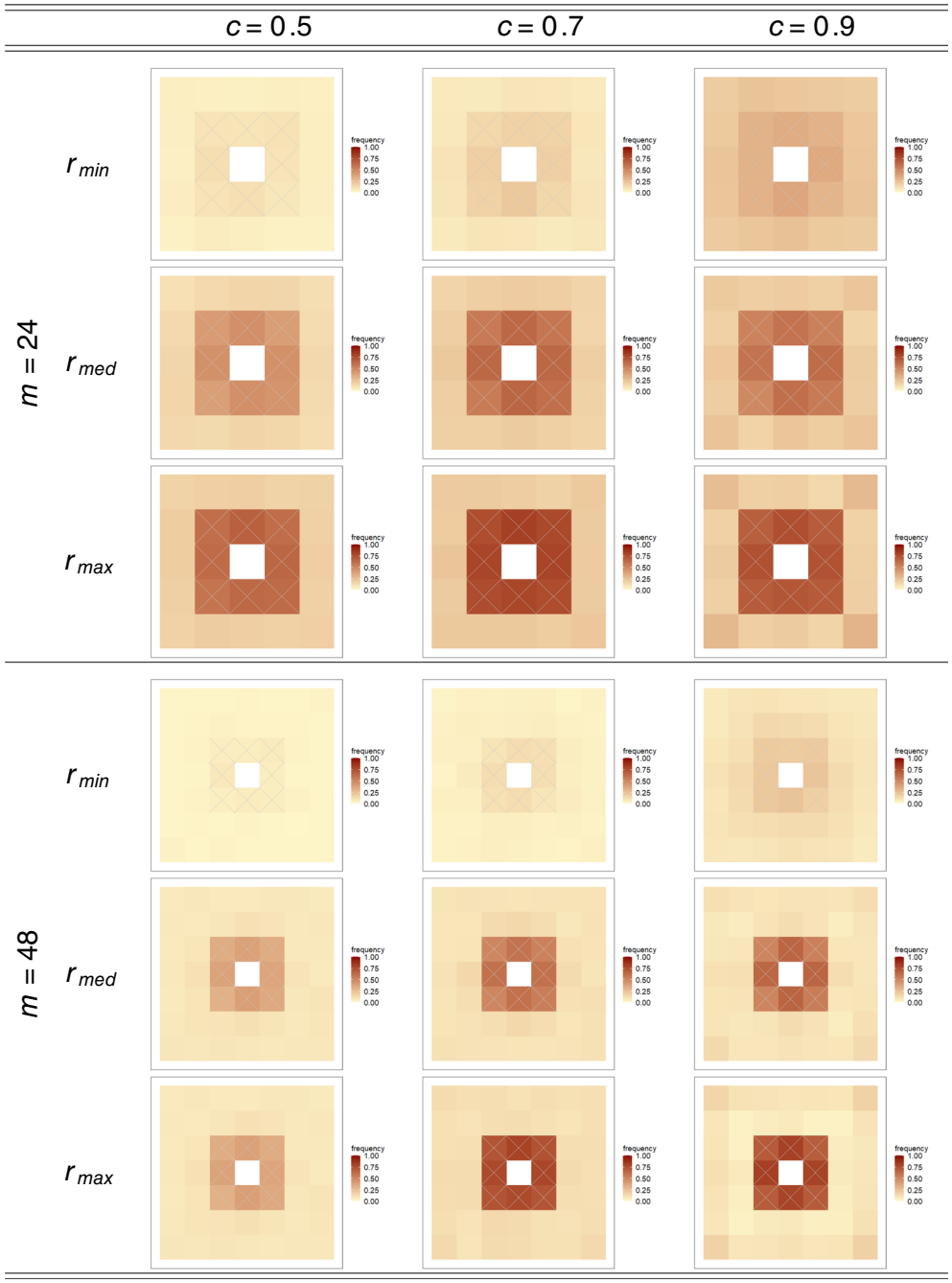


FIGURE 4 Recovery frequency of isotropic $q = 8$ spatial dependence structure (Case A). Columns correspond to the strength of the spatial dependence $c \in \{0.5, 0.7, 0.9\}$. Rows correspond to the number of replications $r \in \{r_{min}, r_{med}, r_{max}\}$ for $m = 24$ and $m = 48$, respectively. The true neighbors are marked by a cross

increasing order. Moreover, the spillover effects are dispersed over several grid cells. Thus, expanding the set of the m nearest locations contributes to the reduction of spatial spillovers and the selection of false neighbors. Alternatively, the number of nearest locations can be adjusted to exclude the second- and higher-order neighbors, for example, by choosing a small neighborhood size of $m = 8$.

In addition to the individual recovery frequencies, Table 1 reports the overall specificity Π_0 and sensitivity Π_1 statistics in addition to the frequency of minimized AIC_c. Specificity refers to the average percentage of correctly identified zero elements, while sensitivity indicates the average percentage of correctly identified nonzero weights.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

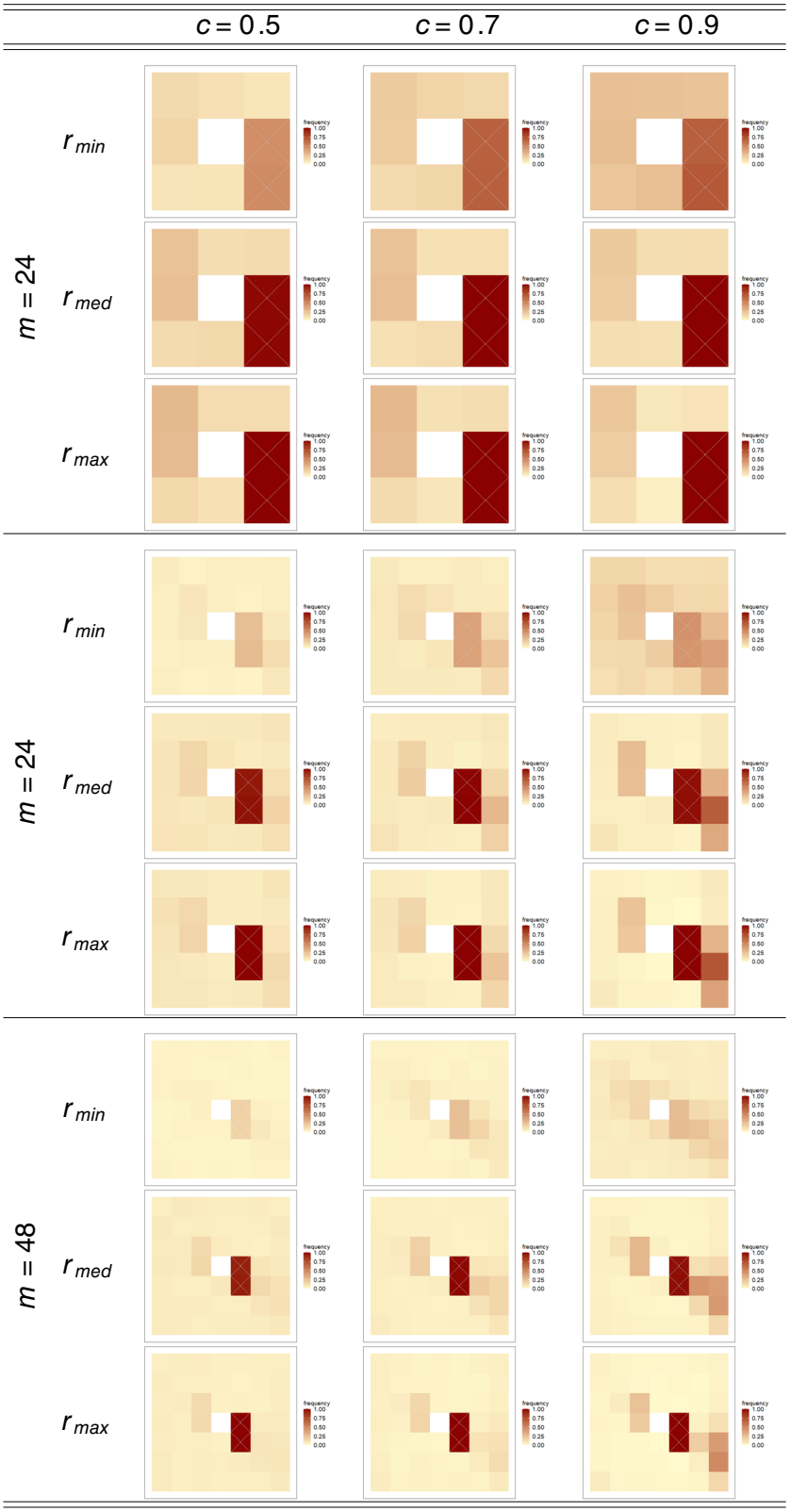


FIGURE 5 Recovery frequency of anisotropic $q = 2$ spatial dependence structure (Case B). Columns correspond to the strength of the spatial dependence $c \in \{0.5, 0.7, 0.9\}$. Rows correspond to the number of replications $r \in \{r_{min}, r_{med}, r_{max}\}$ for $m = 8$, $m = 24$, and $m = 48$, respectively. The true neighbors are marked by a cross

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Overall specificity and sensitivity of the estimated spatial weights depending on the type of weighting scheme, the true number of positive weights q, the strength of the spatial dependence c, the number of cross-sectional resampling replications r, and the number of potential neighbors m

			$c = 0.5$			$c = 0.7$			$c = 0.9$		
q	m		r_{min}	r_{med}	r_{max}	r_{min}	r_{med}	r_{max}	r_{min}	r_{med}	r_{max}
Isotropic setting (Case A)											
8	24	AIC_c	0.724	0.763	0.763	0.662	0.699	0.673	0.701	0.583	0.476
	48	AIC_c	0.276	0.237	0.237	0.338	0.301	0.327	0.299	0.417	0.524
	24	Π_0	0.953	0.859	0.818	0.918	0.817	0.795	0.783	0.801	0.794
	48	Π_0	0.973	0.916	0.887	0.958	0.894	0.882	0.899	0.903	0.914
	24	Π_1	0.101	0.438	0.627	0.174	0.609	0.780	0.324	0.564	0.713
	48	Π_1	0.059	0.358	0.551	0.112	0.559	0.755	0.211	0.590	0.750
Anisotropic setting (Case B)											
2	8	AIC_c	0.651	0.666	0.611	0.737	0.605	0.522	0.733	0.692	0.701
	24	AIC_c	0.196	0.185	0.259	0.168	0.193	0.216	0.170	0.167	0.120
	48	AIC_c	0.153	0.149	0.130	0.095	0.202	0.262	0.097	0.141	0.179
	8	Π_0	0.886	0.829	0.822	0.834	0.840	0.832	0.754	0.851	0.872
	24	Π_0	0.945	0.905	0.912	0.913	0.905	0.919	0.823	0.872	0.889
	48	Π_0	0.968	0.932	0.942	0.954	0.937	0.951	0.915	0.919	0.942
	8	Π_1	0.479	0.989	1	0.674	1	1	0.696	1	1
	24	Π_1	0.261	0.963	0.999	0.383	0.995	1	0.444	0.976	0.999
	48	Π_1	0.168	0.936	0.997	0.251	0.988	1	0.265	0.975	0.999

Note: The percentage of minimized AIC_c can be used to select m.

Across all model specifications, the sensitivity Π_1 increases with the number of cross-sectional resampling replications. Hence, the recovery of the true spatial weights improves with the number of sampled observations. In the case of anisotropic dependence (Case B) with $c = 0.5$ and $m = 48$, the sensitivity rises from 16.8% to 99.7%. Regardless of the model specification, the proportion of correctly identified zero elements Π_0 is less sensitive with respect to r and maintains a level of 75%–97%. Regarding the choice of m, smaller neighborhood sizes tend to coincide with higher sensitivity values. Conversely, the percentage of correctly identified zero elements Π_0 decreases with decreasing m. Hence, recovering the true neighbors benefits from smaller sets of potential candidates, while the identification of zero connections improves with the extent of the sparsity in $\boldsymbol{w}$. Concerning the neighborhood setting, higher individual recovery frequencies are also reflected in the overall sensitivity values, which are much higher in the case of anisotropic dependencies (Case B). In particular, for $r = r_{max}$, the average percentage of correctly identified nonzero connections is approximately 100%. In contrast, for Case A with $q = 8$ true neighbors, the average percentage of correctly identified nonzero weights does not exceed 78%, implying that not all of the eight neighbors are identified.

In accordance with the sensitivity values, lower neighborhood sizes m also tend to yield lower AIC_c values. Based on the AIC_c and sensitivity values, the neighborhood size should therefore be chosen as small as possible and as large as necessary to cover the true neighbors even if it comes at the expense of increasing the proportion of falsely selected zero weights. Bühlmann and Van De Geer (2011) pointed out that false-positive rates (i.e., falsely selected variables) can typically not be avoided due to the lasso's property to select too many variables. This property impairs if predictor variables are highly correlated leading to the selection of higher-order neighbors in the case of spatial autoregressive dependencies. Moreover, high values of Π_0 do not necessarily indicate good model performance because setting all weights to zero yields a value of $\Pi_0 = 1$.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.3 | Estimation accuracy

To evaluate the performance of the parameter estimates, the mean absolute errors (MAE) of the regression coefficient β and the spatial weights vector $\boldsymbol{w}$ are reported in Table 2. Moreover, the bias of β is displayed and the bias of the spatial weights is evaluated separately for the zero weights $\boldsymbol{w}_0$ and nonzero weights $\boldsymbol{w}_1$. The estimation accuracy improves with the number of replications r in terms of lower MAE and less biased regression coefficients and nonzero spatial weights. Similar to the specificity, the bias of the zero weights is less sensitive to changes in m. Since $\Pi_0 < 1$ across all model specifications (i.e., false-positive rates can hardly be avoided), the average bias of the zero weights is greater than zero, while the nonzero weights are downward biased. Concerning the choice of the neighborhood size, the bias of the nonzero connections tends to decay with decreasing m. In contrast, the impairment of the specificity increases the bias of the zero weights as m decreases. Due to the sparsity of $\boldsymbol{w}$, the overall MAE is predominantely determined by the $m - q$ zero connections instead of the q nonzero ones. Therefore, similar trends occur in $\text{MAE}_{\boldsymbol{w}}$ as in the bias of the zero weights, that is, the MAE of the spatial weights improves with m.

To summarize, the number of cross-sectional resampling replications is essential for the identification of the spatial weights structure and the performance of the corresponding estimates. In particular, the average percentage of correctly identified nonzero weights and the recovery frequency of each individual neighbor increase. At the same time, the estimation accuracy improves with the number of replications. The number of correctly identified zero connections (specificity) is less affected by changes in r and is smaller than 100% across all model specifications. These false-positive rates result from spatial spillover effects that lead to false selections of higher-order neighbors and the lasso's tendency to select too many variables. For higher subset sizes (i.e., $m = 48$), the individual neighbors are more difficult to determine and the estimated nonzero spatial weights are more biased. Regarding the identifying assumption of the sparsity of the spatial weights matrix, which is reflected by the number of true neighbors q, sparse matrices with few nonzero weights are better identified than dependence structures that are more connected. Finally, even though the model performance generally improved with increasing r, the medium number of replications already achieves satisfactory results. Especially in the presence of few true neighbors (Case B), the nonzero spatial weights are identified in the majority of the cases (93.6%–100%). By employing cross-sectional resampling or by considering a subsample instead of the entire spatial domain, we can therefore decrease the computational burden, which is particularly essential in the context of big geospatial data.

4 | EMPIRICAL APPLICATION

In this section, we apply the proposed approach to model the spatial dependence of NO_2 emissions. We consider satellite-derived monthly concentrations measured in billion molecules per mm^2 from October 2004 for three different 50×50 grids ($n = 2500$) with a grid-cell resolution of 0.1° across North America. NO_2 primarily arises from combustion processes and traffic and is therefore particularly concentrated around urban areas. For this reason, population size is included as an exogenous predictor and IV in addition to temperature and elevation (see, e.g., Hart et al., 2009; Skene et al., 2010 for the effect of different covariates on NO_2 emissions). Both NO_2 concentrations and the exogenous regressors are retrieved from NASA Earth Observations (NEO).

To investigate the spatial neighborhood structure, different neighborhood sizes $m \in \{8, 24, 48, 80, 120\}$ are considered in the first and second steps, respectively. The estimation is based on 90% of the center observations using the maximum number of considered neighborhood sizes $m = 120$ as a reference. In consequence, $r = 0.9 \cdot (\sqrt{n} - \sqrt{120 + 1} + 1)^2 = 1440$ observations are randomly sampled in the first step and $r = 810$ in the second step. The remaining 10% of observations are used to evaluate the out-of-sample fit of the estimation procedure.

To assess whether the estimation of the spatial dependence structure improves the predictive accuracy, we compare the model fit of the estimated spatial weights matrix with the two most commonly used contiguity-based deterministic specifications, namely the Queen's and Rook's matrices. According to the Queen's specifications, each grid cell is equally affected by its $q = 8$ neighbors with which it either shares a common edge or vertex. In contrast, the Rook's matrix is based on the assumption that the $q = 4$ true neighbors only correspond to the adjacent locations to the north, south, east, and west.

The model fit is evaluated based on the corrected AIC, in-sample, and out-of sample RMSE between observed and predicted NO_2 concentrations. Table 3 reports these model performance measures in addition to the NO_2 concentrations

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 The mean absolute error (MAE) and average bias of the regression coefficient β and the spatial weights $\boldsymbol{w}$ depending on the type of weighting scheme, the number of true positive weights q, the strength of the spatial dependence c, the number of cross-sectional resampling replications r, and the number of potential neighbors m

			$c = 0.5$			$c = 0.7$			$c = 0.9$		
q	m		r_{min}	r_{med}	r_{max}	r_{min}	r_{med}	r_{max}	r_{min}	r_{med}	r_{max}
Isotropic setting (Case A)											
8	24	MAE_β	0.2201	0.0644	0.0467	0.2274	0.0705	0.0485	0.3263	0.0940	0.0560
		MAE_w	0.0292	0.0256	0.0221	0.0398	0.0318	0.0255	0.0519	0.0430	0.0361
		$Bias_\beta$	−0.1259	0.0116	0.0122	−0.0884	0.0176	0.0128	−0.0521	0.0157	0.0162
		$Bias_{w_0}$	0.0088	0.0108	0.0105	0.0138	0.0146	0.0125	0.0270	0.0205	0.0180
		$Bias_{w_1}$	−0.0427	−0.0208	−0.0131	−0.0544	−0.0203	−0.0140	−0.0660	−0.0349	−0.0287
8	48	MAE_β	0.2467	0.0642	0.0470	0.2484	0.0706	0.0487	0.3137	0.0942	0.0546
		MAE_w	0.0153	0.0143	0.0129	0.0208	0.0176	0.0141	0.0288	0.0210	0.0165
		$Bias_\beta$	−0.1809	0.0074	0.0126	−0.1382	0.0156	0.0132	−0.0620	0.0169	0.0163
		$Bias_{w_0}$	0.0050	0.0059	0.0059	0.0070	0.0074	0.0062	0.0132	0.0079	0.0058
		$Bias_{w_1}$	−0.0512	−0.0294	−0.0199	−0.0668	−0.0261	−0.0159	−0.0808	−0.0335	−0.0231
Anisotropic setting (Case B)											
2	8	MAE_β	0.2007	0.0640	0.0462	0.2369	0.0718	0.0491	0.3428	0.1014	0.0580
		MAE_w	0.0610	0.0267	0.0191	0.0749	0.0263	0.0189	0.1045	0.0309	0.0188
		$Bias_\beta$	−0.0912	−0.0155	−0.0109	−0.1024	−0.0245	−0.0174	−0.1547	−0.0427	−0.0279
		$Bias_{w_0}$	0.0226	0.0138	0.0103	0.0322	0.0127	0.0096	0.0499	0.0123	0.0072
		$Bias_{w_1}$	−0.1252	−0.0251	−0.0152	−0.1506	−0.0274	−0.0163	−0.2262	−0.0450	−0.0261
2	24	MAE_β	0.2443	0.0652	0.0471	0.2772	0.0736	0.0507	0.3566	0.1091	0.0661
		MAE_w	0.0268	0.0133	0.0089	0.0373	0.0143	0.0089	0.0532	0.0256	0.0171
		$Bias_\beta$	−0.1729	−0.0201	−0.0119	−0.1919	−0.0327	−0.0196	−0.2073	−0.0679	−0.0444
		$Bias_{w_0}$	0.0102	0.0073	0.0049	0.0158	0.0076	0.0048	0.0246	0.0139	0.0094
		$Bias_{w_1}$	−0.1893	−0.0530	−0.0293	−0.2568	−0.0628	−0.0358	−0.3642	−0.1419	−0.0960
2	48	MAE_β	0.2869	0.0669	0.0478	0.3270	0.0761	0.0518	0.3913	0.1166	0.0715
		MAE_w	0.0152	0.0083	0.0053	0.0202	0.0083	0.0048	0.0289	0.0140	0.0079
		$Bias_\beta$	−0.2380	−0.0263	−0.0144	−0.2667	−0.0395	−0.0232	−0.2839	−0.0864	−0.0542
		$Bias_{w_0}$	0.0060	0.0047	0.0030	0.0080	0.0045	0.0026	0.0127	0.0074	0.0041
		$Bias_{w_1}$	−0.2140	−0.0680	−0.0358	−0.2920	−0.0746	−0.0359	−0.3984	−0.1555	−0.0854

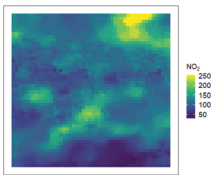
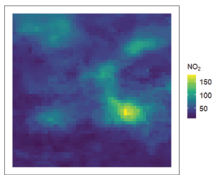
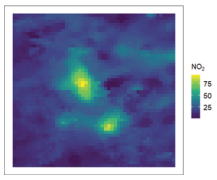
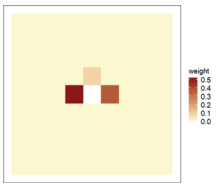
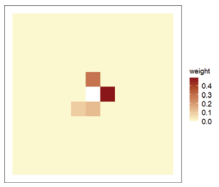
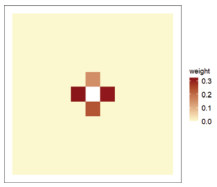
of the three selected regions and the respective estimated spatial weights. Regarding the spatial weights vector, we select the neighborhood size m yielding the lowest AIC_c value, namely, $m = 80$ in the first example and $m = 120$ in the second and third examples.

All three estimated spatial dependence structures are irregular and asymmetrical and therefore do not correspond to either the Queen's or Rook's matrices. Estimating the spatial weights structure thereby leads to improved model performances in terms of lower AIC_c values and better out-of-sample and in-sample fits across all three regions. More precisely, the out-of-sample predictions improve by around 10%–41% in contrast to employing prespecified Queen's or Rook's matrices.

Concerning the effect of exogenous regressors, the competing model specifications lead to different results because the regression coefficients depend on the choice of both the spatial weighting scheme and the spatial extent. Table 4 reports the estimated coefficients obtained from re-estimating the full model based on all $n = 2500$ observations. More precisely, we construct the spatial weighting matrix $\hat{\mathbf{W}}$ from the spatial weights vector $\hat{\boldsymbol{w}}$ and re-estimate the model using a

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 NO_2 observations (first row), estimated weights (second row), corrected AIC, in-sample RMSE, and out-of-sample RMSE based on estimated weights $\hat{\boldsymbol{w}}$, the Queen's matrix and Rook's matrix for three different regions (columnwise)

	37.5°–42.5° north			**32°–37° north**			**38°–43° north**		
	82°–87° west			**82.5°–87.5° west**			**109°–114° west**		
NO_2 observations									
Estimated weights									
	m	$\hat{\boldsymbol{w}}$	**Queen**	**Rook**	$\hat{\boldsymbol{w}}$	**Queen**	**Rook**	$\hat{\boldsymbol{w}}$	**Queen**	**Rook**
AIC_c	120	1549.3	1550.4	1547.5	**1546.7**	1616.9	1614.2	**1559.3**	1623.6	1616.8
	80	**1538.7**	–	–	1547.9	–	–	1568.4	–	–
	48	1545.1	–	–	1555.3	–	–	1559.6	–	–
	24	1584.8	–	–	1612.9	–	–	1599.8	–	–
	8	1553.8	–	–	1589.4	–	–	1646.2	–	–
In-sample RMSE	120	0.2596	0.2526	0.2456	**0.2385**	0.3821	0.3777	**0.2781**	0.3927	0.3819
	80	**0.2277**	–	–	0.2465	–	–	0.2976	–	–
	48	0.2445	–	–	0.2691	–	–	0.2831	–	–
	24	0.3373	–	–	0.3788	–	–	0.3499	–	–
	8	0.2746	–	–	0.3455	–	–	0.4326	–	–
Out-of-sample RMSE	120	0.2607	0.2491	0.2549	0.2362	0.4043	0.4035	0.2623	0.3699	0.3761
	80	0.2244	–	–	0.2604	–	–	0.2860	–	–
	48	0.2507	–	–	0.2857	–	–	0.2589	–	–
	24	0.3726	–	–	0.3331	–	–	0.3160	–	–
	8	0.2848	–	–	0.3326	–	–	0.4265	–	–

Note: Bold values correspond to minimum AIC_c values and lowest in-sample RMSE.

regular two-stage least squares approach to compute standard errors and determine the significance of the estimated model coefficients.

Most noticeably, the effect of the population size is positive and significant across all three regions if the weights are estimated. In contrast, the association between NO_2 emissions and elevation level is negative but only significant in the second example. These results are in line with other findings that NO_2 emissions increase with the population, which serves as a proxy of anthropogenic sources (e.g., road traffic), while higher elevation levels contribute in reducing the pollution (cf., Hart et al., 2009). The coefficient sign of temperature is not constant but insignificant in all three selected regions.

5 | CONCLUSION

In contrast to the classical spatial autoregressive models that require prior specifications of a spatial weights matrix, this article investigates the estimation of the spatial dependence structure for purely spatial regular lattice data.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Estimated coefficients of the exogenous regressors for the three spatial regions depending on the weighting scheme

	37.5°–42.5° north 82°–87° west	32°–37° north 82.5°–87.5° west	38°–43° north 109°–114° west
Estimated weights $\hat{W}$			
Elevation	−0.0025	−0.0138***	−0.0080
	(0.0047)	(0.0048)	(0.0054)
Temperature	0.0017	−0.0079	0.0003
	(0.0053)	(0.0052)	(0.0057)
Population	0.0158**	0.0363***	0.0132***
	(0.0063)	(0.0071)	(0.0043)
Queen's matrix			
Elevation	−0.0075***	−0.0088**	−0.0113*
	(0.0045)	(0.0045)	(0.0066)
Temperature	0.0098*	0.0044	−0.0028
	(0.0048)	(0.0048)	(0.0069)
Population	0.0226**	0.0152***	0.0147***
	(0.0054)	(0.0058)	(0.0054)
Rook's matrix			
Elevation	−0.0026	−0.0061	−0.0086
	(0.0041)	(0.0040)	(0.0058)
Temperature	0.0033	0.0014	−0.0028
	(0.0045)	(0.0043)	(0.0063)
Population	0.0112**	0.0170***	0.0077
	(0.0052)	(0.0056)	(0.0051)

Note: Standard errors are in parentheses. Significant at ***1% level, **5% level, *10% level.

A two-step adaptive lasso approach is suggested to estimate the individual spatial links, assuming the sparsity of the spatial weights and the exchangeability of the random process. Apart from the sparsity of the spatial weights, no structural assumptions such as symmetry or a triangular form are imposed, allowing all types of irregular dependence structures to be estimated. Moreover, we propose using cross-sectional resampling to recover the spatial dependence structure. Thus, in contrast to other estimation methods like ML or Gaussian Markov random fields, we avoid computationally demanding matrix operations by considering a smaller subsample of the spatial domain and by substituting the spatial weights matrix with a stacked, unique weighting vector that captures the spatial dependence structure.

The Monte Carlo results demonstrate that the estimators appear to be consistent with an increasing number of cross-sectional resampling replications. More precisely, the average percentage of correctly identified spatial connections increases, and the spatial weight estimates are most accurate for the maximum number of replications. However, even a medium number of replications already ensures the recovery of the true neighbors in the majority of the cases while maintaining low false-positive rates. This strengthens our proposal to reduce the computational burden by considering a subsample instead of the entire spatial domain, which is particularly relevant in the context of big geospatial data. Regarding the selection of the set of nearest locations, contradictory influences exist on the model performance. On one hand, a low number of nearest locations carries the risk of not covering all relevant neighbors. Moreover, spatial spillover effects are more concentrated on very few false neighbors if the set of potential neighbors covers the higher-order neighbors. On the other hand, true neighbors are better identified and the nonzero weights are less biased. In addition, more observations are available for cross-sectional resampling, which is particularly relevant for small data sets, and the computational

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

effort is less demanding. In consequence, lower neighborhood sizes are favored in terms of corrected information criteria. Finally, sparse dependence structures with very few true neighbors are better identified than spatial weights matrices that are more connected.

For the empirical application, we investigate the spatial dependence structure of the NO_2 data. We find that the prediction accuracy in terms of the in-sample and out-of-sample RMSE can be improved considerably by estimating the spatial weights in contrast to employing deterministic specifications that are based on prior selections of the q nearest neighbors.

Finally, our approach is suitable for representing the spatial dependence structure for regular lattice data provided that the spatial dependence is constant across space. Hence, structural breaks or heterogeneous spatial dependencies are not accounted for and should be investigated in future research. Moreover, additional assumptions on the ordering of the nearest locations are required if other types of spatial data, such as irregular lattice or geostatistical processes, are of interest. In particular, different distance lags and directions between point-referenced locations or polygon centroids could be considered to define the neighbors corresponding to each weight.

REFERENCES

Ahrens, A., & Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, *3*, 128–155.

Anselin, L. (1988). *Spatial econometrics: Methods and models* (Vol. *1*). Kluwer Academic Publishers.

Basak, G. K., Bhattacharjee, A., & Das, S. (2018). Causal ordering and inference on acyclic networks. *Empirical Economics*, *55*, 213–232.

Besner, C. (2002). A spatial autoregressive specification with a comparable sales weighting scheme. *Journal of Real Estate Research*, *24*, 193–212.

Bhattacharjee, A., Castro, E., & Marques, J. (2012). Spatial interactions in hedonic pricing models: The urban housing market of Aveiro, Portugal. *Spatial Economic Analysis*, *7*, 133–167.

Bhattacharjee, A., & Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics*, *43*, 617–634.

Biscio, C. A. N., & Waagepetersen, R. (2019). A general central limit theorem and a subsampling variance estimator for α-mixing point processes. *Scandinavian Journal of Statistics*, *46*, 1168–1190.

Bodson, P., & Peeters, D. (1975). Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour-demand function. *Environment and Planning A*, *7*, 455–472.

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media.

Cliff, A., & Ord, K. (1973). *Spatial autocorrelation*. Pion.

Cohen, J., & Tita, G. (1999). Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes. *Journal of Quantitative Criminology*, *15*, 451–493.

Cressie, N. (1993). *Statistics for spatial data*. Wiley.

Cressie, N., & Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, *94*, 1330–1339.

Debarsy, N., Ertur, C., & LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology*, *9*, 158–171.

Ertur, C., & Koch, W. (2007). Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, *22*, 1033–1062.

Fotheringham, A. S., Charlton, M., & Brunsdon, C. (1997). *Measuring spatial variations in relationships with geographically weighted regression*. In Fischer, M. M., Getis, A., *Recent developments in spatial analysis* (pp. 60–82). Springer.

Geniaux, G., & Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, *72*, 74–85.

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, *52*, 172–191.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, *97*, 590–600.

Harris, P., Fotheringham, A., Crespo, R., & Charlton, M. (2010). The use of geographically weighted regression for spatial prediction: An evaluation of models using simulated data sets. *Mathematical Geosciences*, *42*, 657–680.

Hart, J. E., Yanosky, J. D., Puett, R. C., Ryan, L., Dockery, D. W., Smith, T. J., Garshick, E., & Laden, F. (2009). Spatial modeling of pm10 and no2 in the continental United States, 1985–2000. *Environmental Health Perspectives*, *117*, 1690–1696.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*, 99–121.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509–533.

Krock, M., Kleiber, W., & Becker, S. (2021). Nonstationary modeling with sparsity for spatial data via the basis graphical lasso. *Journal of Computational and Graphical Statistics*, *30*, 375–389.

Krock, M., Kleiber, W., Hammerling, D., & Becker, S. (2021). *Modeling massive multivariate spatial data with the basis graphical lasso* (arXiv preprint arXiv:2101.02404).

Lam, C., & Souza, P. C. (2019). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, *38*, 693–710.

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*, 1899–1925.

Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, *32*, 9–32.

Ma, C. (2003). Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, *86*, 97–107.

Merk, M. S., & Otto, P. (2021). Directional spatial autoregressive dependence in the conditional first- and second-order moments. *Spatial Statistics*, *41*, 100490.

Mizruchi, M. S., & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, *30*, 190–200.

Ninomiya, Y., & Kawano, S. (2016). Aic for the lasso in generalized linear models. *Electronic Journal of Statistics*, *10*, 2537–2560.

Nott, D. J., & Dunsmuir, W. T. (2002). Estimation of nonstationary spatial covariance structure. *Biometrika*, *89*, 819–829.

Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, *2*, 315–331.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120–126.

Otto, P., & Steinert, R. 2018. *Estimation of the spatial weighting matrix for spatiotemporal data under the presence of structural breaks* (arXiv preprint arXiv:1810.06940).

Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*, 483–506.

Pinkse, J., Slade, M. E., & Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, *70*, 1111–1153.

Putter, H., & Young, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli*, *7*, 421–438.

Reich, B. J., & Fuentes, M. (2012). Nonparametric Bayesian models for a spatial covariance. *Statistical Methodology*, *9*, 265–274.

Risser, M. D., & Calder, C. A. (2015). Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, *26*, 284–297.

Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, *87*, 108–119.

Simpson, D., Lindgren, F., & Rue, H. (2012). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, *23*, 65–74.

Skene, K. J., Gent, J. F., McKay, L. A., Belanger, K., Leaderer, B. P., & Holford, T. R. (2010). Modeling effects of traffic and landscape characteristics on ambient nitrogen dioxide levels in Connecticut. *Atmospheric Environment*, *44*, 5156–5164.

Smith, T. E. (2009). Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis*, *41*, 307–332.

Stakhovych, S., & Bijmolt, T. H. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, *88*, 389–408.

Stein, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, *16*, 55–63.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*, 267–288.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Wheeler, D. C. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A*, *41*, 722–742.

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Zhu, J., Huang, H. C., & Reyes, P. E. (2010). On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*, 389–402.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429.

How to cite this article: Merk, M. S., & Otto, P. (2022). Estimation of the spatial weighting matrix for regular lattice data—An adaptive lasso approach with cross-sectional resampling. *Environmetrics*, *33*(1), e2705. https://doi.org/10.1002/env.2705

1099095x, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.2705 by University Of Glasgow, Wiley Online Library on [03/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License