





# Space-time integer-valued ARMA modelling for time series of counts

Ana Martins<sup>1</sup> , Manuel G. Scotto<sup>2</sup> ,  
Christian H. Weiß<sup>3</sup> , and Sónia Gouveia<sup>1,4</sup> 

<sup>1</sup>*Institute of Electronics and Informatics Engineering of Aveiro (IEETA) and  
Department of Electronics, Telecommunications and Informatics (DETI),  
University of Aveiro, Aveiro, Portugal  
e-mail: [a.r.martins@ua.pt](mailto:a.r.martins@ua.pt); [sonia.gouveia@ua.pt](mailto:sonia.gouveia@ua.pt)*

<sup>2</sup>*Center for Computational and Stochastic Mathematics (CEMAT) and  
Department of Mathematics, IST, University of Lisbon, Lisbon, Portugal  
e-mail: [manuel.scotto@tecnico.ulisboa.pt](mailto:manuel.scotto@tecnico.ulisboa.pt)*

<sup>3</sup>*Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg,  
Germany  
e-mail: [weissc@hsu-hh.de](mailto:weissc@hsu-hh.de)*

<sup>4</sup>*Intelligent Systems Associate Laboratory (LASI), Portugal*

**Abstract:** This paper introduces a new class of space-time integer-valued ARMA models referred to as STINARMA. This class arises as the natural space-time extension of the INARMA models and, simultaneously, as the integer-valued counterpart of the conventional STARMA models. In this work, the moving average subclass  $\text{STINMA}(q_{m_1, \dots, m_q})$  is studied in detail. Particular attention is given to the derivation of first- and second-order moments, including space-time autocorrelations. Due to its large potential use in real-data applications, the Poisson  $\text{STINMA}(1_1)$  process is analyzed in further detail. Estimation methods are also addressed and their performance is demonstrated through a simulation study and by analysing the daily number of hospital admissions observed over time in three Portuguese locations.

**MSC2020 subject classifications:** Primary 62M10, 62H11, 62H12; secondary 62P12, 68T09.

**Keywords and phrases:** Space-time series of counts, STINARMA models, autoregressive moving-average processes, binomial thinning operator, Poisson distribution.

Received October 2022.

## 1. Introduction

The statistical characterization of the temporal dependence structure of count time series is a topic of active research nowadays. To this end, several families of models have been proposed in the literature being the equation-based models one of the most popular. This family includes the thinning-operator-based models [40] such as the integer-valued ARMA (INARMA), in which the multiplication of the ordinary ARMA recursion is replaced by the binomial thinning operator (BTO hereafter). While univariate INARMA models have been

intensively discussed in the literature [e.g. 11, 12, 54], the development of its multivariate extensions progresses slowly. Multivariate INAR (MINAR) models based upon either independent BTOs or generalized thinning operators were first introduced in the 1990s by Franke and Subba Rao [17] and Latour [24], respectively. Several further contributions can be found in the literature including periodic MINAR [39], MINAR with seasonally varying autocorrelation parameters [9], non-stationary bivariate INAR [23], bivariate binomial INAR [41] and full bivariate INAR [30, 31]. By contrast, however, multivariate integer-valued moving averages (MINMA) models emerged considerably later with the proposal of a bivariate INMA (BINMA) model [37] and its generalization to higher dimensions [38]. More recent papers on BINMA models include [42], [49] and [47]. It is worth to mention that the family of full multivariate INARMA models is hardly ever addressed in the literature with some exceptions e.g. Sunecher, Mamode Khan and Jowaheer [48], Sunecher [46] and McKenzie [28].

As far as parameter estimation is concerned, conditional least squares (CLS) and conditional maximum likelihood (CML) have been conducted for MINAR models [17, 24] while CLS, generalized method of moments (GMM) and generalized quasi-likelihood (GQL) have been established for BINMA models [37, 38, 47, 42, 49].

Since most contributions to multivariate BTO-based models have been introduced during the last few years, it is not surprising that both purely spatial and space-time INARMA models are still quite underdeveloped and limited to a few applications with autoregressive approaches [e.g. 20, 19, 1, 22, 50]. Therefore, this work intends to contribute to this direction by introducing a novel space-time extension of the full INARMA model, hereafter referred to as STINARMA class. This new class constitutes the integer BTO-based counterpart of the space-time ARMA (STARMA) class first introduced by Pfeifer and Deutsch [35], in which the spatial component is expressed by a  $\mathbf{W}^{(\ell)}$  matrix providing weighted information on the spatial neighbors of order  $\ell$ . Following the same line of reasoning, the novel class is designated to as  $\text{STINARMA}(p_{f_1, \dots, f_p}, q_{m_1, \dots, m_q})$  highlighting that different spatial lags  $f_1, \dots, f_p$  and  $m_1, \dots, m_q$  are allowed for each temporal lag of the AR and MA component of the class; see Section 2 for details. This paper further explores the stationary  $\text{STINMA}(q_{m_1, \dots, m_q})$  subclass of models (a particular case of  $p = 0$  in the STINARMA class). Specifically, the second-order model structure is derived in Section 3 along with the space-time autocovariance and autocorrelation functions. Moreover, Section 4 presents exhaustive research on the Poisson  $\text{STINMA}(1_1)$  model, including the derivation of its moments and estimation methods based on the method of moments (MM), CLS and CML, being their performance compared through a simulation study; see Section 5. In Section 6, the Poisson  $\text{STINMA}(1_1)$  model is applied to a real-world data example concerning a space-time process on the daily number of hospital admissions in Portugal. Finally, Section 7 is devoted to the conclusions and perspectives for future research in the STINARMA class.

## 2. The STINARMA class for spatio-temporal counts

The STINARMA class is proposed as an extension of the univariate INARMA class to the space dimension. Both classes are based on the BTO defined as

$$\psi \circ X := \sum_{i=1}^X u_i, \quad X > 0 \quad (1)$$

and 0 otherwise [44]. In this formulation,  $X$  is a discrete random variable with range  $\mathbb{N}_0$  and  $u_i, i = 1, \dots, X$  are independent and identically distributed (i.i.d.) Bernoulli-distributed random variables with probability of success  $\psi \in (0, 1)$ , being independent of  $X$ . The boundary cases are defined as  $0 \circ X = 0$  and  $1 \circ X = X$ . The general INARMA( $p, q$ ) model [13] makes use of the BTO and is defined by the recursive scheme

$$Y_t = \sum_{i=1}^p \alpha_i \circ Y_{t-i} + \sum_{j=1}^q \beta_j \circ \varepsilon_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}, \quad (2)$$

where  $\varepsilon_t \in \mathbb{N}_0$  are i.i.d. count innovations with  $\mu_\varepsilon := E(\varepsilon_t) < \infty$  and  $\sigma_\varepsilon^2 := V(\varepsilon_t) < \infty$ , being independent of  $Y_u$  for  $u < t$ . Within this setting, the BTOs are assumed to be performed independently of each other and of  $\varepsilon_t$ , for each  $t$ . These operations make use of the model parameters  $\alpha_i, \beta_j \in [0, 1)$  defined for  $i = 1, \dots, p$  and  $j = 1, \dots, q$  such that  $\alpha_p \neq 0$  and  $\beta_q \neq 0$ , where  $p$  and  $q$  denote, respectively, the temporal AR and MA orders of the INARMA model.

From the INARMA representation in (2), the novel STINARMA class is defined as follows. Set  $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{S,t})^\top \in \mathbb{N}_0^S$  as the  $(S \times 1)$ -vector of a multivariate count process at locations  $s = 1, \dots, S$  and time  $t \in \mathbb{Z}$ . ( $\mathbf{Y}_t$ ) it is said to be a STINARMA process if it admits the representation

$$\mathbf{Y}_t = \sum_{i=1}^p \sum_{\ell=0}^{f_i} \alpha_{i\ell} \mathbf{W}^{(\ell)} \circ \mathbf{Y}_{t-i} + \sum_{j=1}^q \sum_{\ell=0}^{m_j} \beta_{j\ell} \mathbf{W}^{(\ell)} \circ \varepsilon_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (3)$$

where  $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{S,t})^\top \in \mathbb{N}_0^S$  are i.i.d. with  $\varepsilon_{s,t}$  being the innovation at location  $s = 1, \dots, S$  and time  $t \in \mathbb{Z}$ . The STINARMA model introduces the spatial dependency through the  $(S \times S)$ -matrix of weights  $\mathbf{W}^{(\ell)}$  for the spatial lags  $\ell = 0, \dots, \max\{f_1, \dots, f_p, m_1, \dots, m_q\} < S$ , where  $f_i$  and  $m_j$  are the spatial orders for the  $i^{\text{th}}$  autoregressive and the  $j^{\text{th}}$  moving-average term, respectively. The model in (3) is hereafter referred to as STINARMA( $p_{f_1, \dots, f_p}, q_{m_1, \dots, m_q}$ ). In analogy to the STARMA and INARMA classes, the STINARMA can be reduced to either a purely autoregressive (STINAR) or moving-average (STINMA) class by setting  $q = 0$  or  $p = 0$ , respectively.

The STINARMA representation in (3) highlights that  $\mathbf{Y}_t$  is expressed as a sum of time-lagged BTOs that introduces the temporal dependence to the multivariate process. Moreover, the spatial dependence is expressed through the multiplication of the parameters  $\alpha_{i\ell}, \beta_{j\ell} \in [0, 1)$  by the weight matrix  $\mathbf{W}^{(\ell)}$ ,

which is endowed into the BTO that takes a matrix formulation [17]. The BTO between a matrix  $\Psi$  with (fixed) entries in  $[0, 1)$  and a random vector  $\mathbf{X}$  is defined as

$$\Psi \circ \mathbf{X} := \begin{pmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1S} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{S1} & \psi_{S2} & \cdots & \psi_{SS} \end{pmatrix} \circ \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_S \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^S \psi_{1n} \circ X_n \\ \sum_{n=1}^S \psi_{2n} \circ X_n \\ \vdots \\ \sum_{n=1}^S \psi_{Sn} \circ X_n \end{pmatrix}, \quad (4)$$

where  $\psi_{sn} \circ X_n$ , for a given  $n$  and  $s$  is performed according to (1). Useful properties of such matricial operators can be found in Franke and Subba Rao [17] and [24]. In the default version of the STINARMA model, all BTOs in (3) and (4) are performed independently of each other. However, this assumption is easily modified if required by the application context. In Section 4, for example, a modified STINMA(1<sub>1</sub>) model is considered, imposing the same observed value for  $\psi_{in} \circ X_n$  and  $\psi_{jn} \circ X_n$  such that  $\psi_{in} = \psi_{jn}$ ,  $i \neq j$  in a matricial BTO (4) performed at a given time  $t$ . The interpretation is that if the spatial contribution of location  $n$  to its neighbors  $i$  and  $j$  is the same, then the result of these two BTOs should be the same for both neighbors at a time instant  $t$ . Therefore, the BTOs conveyed in this restricted STINARMA formulation of (3) are performed independently in time but not necessarily independently for different spatial locations at a given time  $t$ .

The innovation process  $\boldsymbol{\varepsilon}_t$  in (3) is assumed to be i.i.d. in time with  $\boldsymbol{\mu}_\varepsilon := E(\boldsymbol{\varepsilon}_t) = (\mu_{\varepsilon_1}, \dots, \mu_{\varepsilon_S})^\top$  where  $\mu_{\varepsilon_s} := E(\varepsilon_{s,t}) < \infty$  and  $\sigma_{\varepsilon_s}^2 := V(\varepsilon_{s,t}) < \infty$ , for each spatial component  $s = 1, \dots, S$ . Furthermore, it is assumed that  $\boldsymbol{\varepsilon}_t$  follows a (trivial) discrete multivariate distribution, i.e. with mutually independent components. Such assumption, however, might be relaxed if a more sophisticated cross-dependence structure is required for  $\mathbf{Y}_t$ ; see Section 7 for details. Under the mutual independence assumption, it holds that

$$Cov[\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+h}^\top] = \begin{cases} \mathbf{G}, & h = 0, \\ \mathbf{O}_S, & h \neq 0, \end{cases} \quad (5)$$

where  $\mathbf{G} = \text{diag}(\sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_S}^2)$  with zero off-diagonal entries, and  $\mathbf{O}_S$  denotes the  $(S \times S)$ -zero matrix. To sum up, the multivariate process  $\boldsymbol{\varepsilon}_t$  is assumed to be i.i.d. in time and independent in space, but not necessarily identically distributed in space.

A key component of the STINARMA model is the set of neighbor matrices  $\mathbf{W}^{(\ell)}$ ,  $\ell < S$ , which reflects the hierarchical ordering of the spatial neighbors, namely the first-order neighbors ( $\ell = 1$ ) are those closest to each other, second-order neighbors ( $\ell = 2$ ) are farther apart, and so on. For a given  $\ell$ , the entries

of  $\mathbf{W}^{(\ell)}$  are such that

$$w_{sn}^{(\ell)} := \begin{cases} \kappa_{sn}^{(\ell)}, & \text{if } s \text{ and } n \text{ are } \ell^{\text{th}}\text{-order neighbors,} \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where  $\kappa_{sn}^{(\ell)}$  quantifies the influence of location  $n$  in  $s$  provided that these are  $\ell^{\text{th}}$ -order neighbors. Note that locations  $s$  and  $n$  cannot be simultaneously  $\ell^{\text{th}}$ -order and  $k^{\text{th}}$ -order neighbors with  $\ell \neq k$ , i.e. at most one of  $\{w_{sn}^{(0)}, w_{sn}^{(1)}, \dots\}$  is non-zero. From (6) it follows that  $\mathbf{W}^{(0)}$  is a diagonal matrix since each location is its (only) zero-order neighbor; the normalized  $\mathbf{W}^{(0)} = \mathbf{I}_S$  (identity matrix) is considered. Also,  $w_{ss}^{(\ell)} = 0$  for  $\ell \geq 1$  since each location is not a neighbor of order  $\ell \geq 1$  of itself, which implies that  $\text{tr}(\mathbf{W}^{(\ell)}) = 0$ ,  $\ell \geq 1$ . Finally, it is important to remark that the matrix  $\mathbf{W}^{(\ell)}$  is not necessarily symmetric for  $\ell > 0$ .

Dealing with BTO implies that all entries of the matrices  $\alpha_{i\ell}\mathbf{W}^{(\ell)}$  and  $\beta_{j\ell}\mathbf{W}^{(\ell)}$  must range between 0 and 1. Since  $\alpha_{i\ell}, \beta_{i\ell} \in [0, 1)$ , it suffices to assume that  $w_{sn}^{(\ell)} \in [0, 1]$  to comply with such BTO restriction. This holds for row-normalized  $\mathbf{W}^{(\ell)}$  matrices (i.e. such that  $\sum_{s=1}^S w_{sn}^{(\ell)} = 1$  for each  $n$  and  $\ell$ ), by defining the spatial weights as in (6) with  $\kappa_{sn}^{(\ell)} = 1/m_s^{(\ell)}$ , where  $m_s^{(\ell)} \leq S$  denotes the number of  $\ell^{\text{th}}$ -order neighbors of location  $s$  [35]. As an example, Figure 1(a) shows a  $3 \times 3$  grid layout with a reference location in the centre and eight surrounding neighbor locations. According to the previous definition each first-order neighbor  $n$  of the reference location  $s$  has  $\kappa_{sn}^{(1)} = 1/4$ .

**Remark 2.1.** Other approaches for expressing the relation between  $\ell^{\text{th}}$ -order neighbors are possible, e.g. by transforming values  $\nu_{sn}^{(\ell)} \in \mathbb{R}_0^+$  into  $\kappa_{sn}^{(\ell)} = f(\nu_{sn}^{(\ell)}) \in$

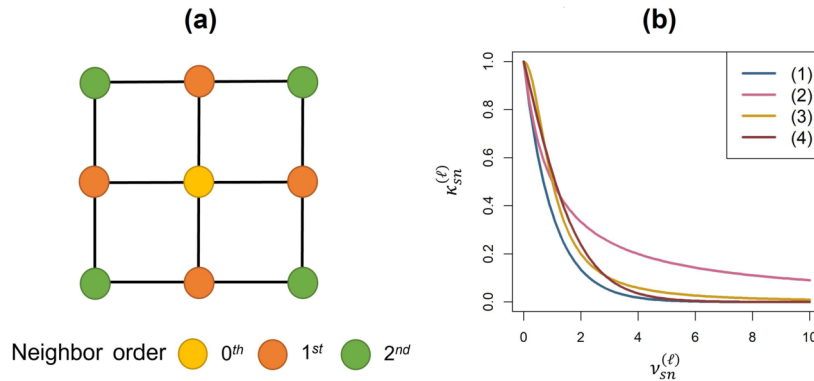


FIG 1. (a) Schematic representation of a  $3 \times 3$  grid with an anchor location (yellow) and its eight surrounding neighbor locations of different orders. (b) Examples of continuous transformations  $\kappa_{sn}^{(\ell)} = f(\nu_{sn}^{(\ell)})$  of spatial weights  $\nu_{sn}^{(\ell)} \in \mathbb{R}_0^+$  into  $\kappa_{sn}^{(\ell)} \in [0, 1]$ , namely (1)  $\kappa_{sn}^{(\ell)} = \exp(-\nu_{sn}^{(\ell)})$ , (2)  $\kappa_{sn}^{(\ell)} = 2/(\exp(\nu_{sn}^{(\ell)}) + 1)$ , (3)  $\kappa_{sn}^{(\ell)} = 1/(\exp(\nu_{sn}^{(\ell)}) + 1)$  and (4)  $\kappa_{sn}^{(\ell)} = 1/(\exp(\nu_{sn}^{(\ell)})^2 + 1)$ .

$[0, 1]$  with subsequent row-normalization. For instance, one may consider all locations as first-order neighbors and take  $\nu_{sn}^{(1)}$  as the Euclidean distance between locations  $s$  and  $n$  [45]. Figure 1(b) displays examples of transformations  $f$  that comply with the  $[0, 1]$ -restriction and follow an inverse relation between distance and weights (higher distance associated with smaller weights). These functions satisfy  $f(0) = 1$  and  $\lim_{x \rightarrow \infty} f(x) = 0$ , but exhibit different decay ratios that lead to different weighting schemes for the neighbor locations.

The new STINARMA class thus constitutes the BTO-based integer counterpart of the STARMA class introduced by [35]. The STARMA recursion, which refers to a multivariate real-valued process, does not differ considerably from that in (3) except for the fact that the BTO is replaced by the conventional multiplication, the STARMA parameters  $\alpha_{i\ell}$  and  $\beta_{j\ell}$  (and their scalar products with  $\mathbf{W}^{(\ell)}$ ) are not restricted to the  $[0, 1]$  interval, and the  $\varepsilon_t$  components are normally distributed with mean  $\mathbf{0}_S$  (zero vector) and diagonal cross-covariance matrix with variance  $\sigma_{\varepsilon_1}^2 = \dots = \sigma_{\varepsilon_S}^2 = \sigma^2$ . Furthermore, the connections of the new STINARMA model (3) to the INARMA framework are obvious: setting the spatial orders  $f_i = m_j = 0$  for all  $i, j$  leads to a multivariate INARMA( $p, q$ ) model, while setting the number of locations to  $S = 1$  leads to the univariate INARMA( $p, q$ ) formulation in (2). The novel STINARMA class also includes the so-called GSTAR(1; 1) model of Huda, Mukhaiyar and Pasaribu [22] as a special case ( $p = 1, f_1 = 1$ , and  $q = 0$ ), but it differs from the SINAR(1, 1) model of Ghodsi, Shitan and Bakouch [20] and the multiple PoINAR(1) proposal of Aldor-Noiman et al. [1] with correlated innovations.

Finally, the conditions for the STINARMA model to become a stationary process are closely linked to those of the STARMA process [35] through the results conveyed in Latour [24] regarding the MGINAR( $p$ ) model. Noticing that at most one of  $\{w_{sn}^{(0)}, w_{sn}^{(1)}, \dots\}$  is non-zero, the STINARMA recursion (3) can be rewritten as

$$\mathbf{Y}_t = \sum_{i=1}^p \underbrace{\left( \sum_{\ell=0}^{f_i} \alpha_{i\ell} \mathbf{W}^{(\ell)} \right)}_{=: \mathbf{A}_i} \circ \mathbf{Y}_{t-i} + \sum_{j=1}^q \underbrace{\left( \sum_{\ell=0}^{m_j} \beta_{j\ell} \mathbf{W}^{(\ell)} \right)}_{=: \mathbf{B}_j} \circ \varepsilon_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (7)$$

which shows that the STINARMA class can be understood as a special type of an  $S$ -variate INARMA model. This relation can be used to establish the existence of a unique stationary solution. Needless to say that the STINMA model is stationary by construction, so the focus should be put on the autoregressive part. According to (7), any STINAR( $p_{f_1, \dots, f_p}$ ) model constitutes a MGINAR( $p$ ) model in the sense of Latour [24], for which a stationary solution exists if the roots of  $\det(\mathbf{I}_S - \mathbf{A}_1 z - \dots - \mathbf{A}_p z^p)$  are outside the unit circle. Thus, the existence of a stationary STINAR( $p_{f_1, \dots, f_p}$ ) process is ensured if

$$\det \left( \mathbf{I}_S - \sum_{i=1}^p \sum_{\ell=0}^{f_i} \alpha_{i\ell} \mathbf{W}^{(\ell)} z^i \right) \neq 0 \quad \text{for } |z| \leq 1, \quad (8)$$

which is equivalent to the stationarity condition of the continuous STAR model [35, p. 37].

### 3. The general STINMA( $q_{m_1, \dots, m_q}$ ) process

The STINMA( $q_{m_1, \dots, m_q}$ ) process satisfies the recursion

$$\mathbf{Y}_t = \sum_{j=1}^q \sum_{\ell=0}^{m_j} \beta_{j\ell} \mathbf{W}^{(\ell)} \circ \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{9}$$

where  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{S,t})^\top$  is an i.i.d. multivariate process with finite  $E(\varepsilon_{s,t}) = \mu_{\varepsilon_s}$  and  $V(\varepsilon_{s,t}) = \sigma_{\varepsilon_s}^2$  for each  $s = 1, \dots, S$ . Furthermore, the matrix-BTO is defined according to equation (4) with all  $\circ$ -operations being independent. The previous recursion can be written as

$$\begin{bmatrix} Y_{1,t} \\ \vdots \\ Y_{S,t} \end{bmatrix} = \sum_{j=1}^q \sum_{\ell=0}^{m_j} \beta_{j\ell} \begin{bmatrix} w_{11}^{(\ell)} & \dots & w_{1S}^{(\ell)} \\ \vdots & \ddots & \vdots \\ w_{S1}^{(\ell)} & \dots & w_{SS}^{(\ell)} \end{bmatrix} \circ \begin{bmatrix} \varepsilon_{1,t-j} \\ \vdots \\ \varepsilon_{S,t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{S,t} \end{bmatrix}, \tag{10}$$

which highlights that the spatial component  $Y_{s,t}$  is modelled as the sum of thinning operations involving  $\boldsymbol{\varepsilon}_{t-j}$ , either through its  $s^{\text{th}}$  component or the components associated with the  $\ell^{\text{th}}$ -order neighbors of location  $s$ . As  $\mathbf{W}^{(0)} = \mathbf{I}_S$  by definition, it is clear that an  $S$ -variate MINMA( $q$ ) driven by space-time independent innovations arises by setting  $m_1 = \dots = m_q = 0$ . In this case,  $Y_{s,t}$  is formulated as the sum of thinning operations with  $\varepsilon_{s,t}$  alone and its time lagged versions up to order  $q$ , which resumes (10) to a system of independent equations. A workaround to use the MINMA( $q$ ) model in the context of spatial data analysis is to consider component-wise correlated innovations (see e.g. [42] for the bivariate case). However, this differs from the new STINMA( $q_{m_1, \dots, m_q}$ ) proposal where the spatial component is explicitly considered in the model's definition through the  $\mathbf{W}^{(\ell)}$  matrices, and the innovation process is assumed to be temporally and spatially independent.

#### 3.1. First- and second-order moments

Basic properties of the BTO can be used to show that the component-wise mean and variance of the STINMA( $q_{m_1, \dots, m_q}$ ) process are given by

$$E[Y_{s,t}] = \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S \left( \beta_{j\ell} w_{sn}^{(\ell)} \mu_{\varepsilon_n} \right) + \mu_{\varepsilon_s} \tag{11}$$

$$V[Y_{s,t}] = \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S \left( \beta_{j\ell}^2 \left( w_{sn}^{(\ell)} \right)^2 \sigma_{\varepsilon_n}^2 + \beta_{j\ell} w_{sn}^{(\ell)} \left( 1 - \beta_{j\ell} w_{sn}^{(\ell)} \right) \mu_{\varepsilon_n} \right) + \sigma_{\varepsilon_s}^2 \tag{12}$$

for a given location  $s = 1, \dots, S$ . The complete proofs are presented in Appendix A. The mean and variance of the  $S$ -variate MINMA( $q$ ) process are obtained for the particular case  $m_1 = \dots = m_q = 0$  [52]. Compared to the conventional STMA process with  $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{\mu}_\varepsilon, \mathbf{G})$ , the STINMA exhibits the same mean but the variance includes the additional terms  $\beta_{j\ell} w_{sn}^{(\ell)} (1 - \beta_{j\ell} w_{sn}^{(\ell)}) \boldsymbol{\mu}_{\varepsilon_n}$  stemming from replacing the multiplication by the  $\circ$ -operation. Finally, the mean and variance of the STMA process with  $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_S, \mathbf{G})$  are included in the above equations by setting  $\boldsymbol{\mu}_{\varepsilon_s} = \mathbf{0}$  [35] which certainly is not well-defined in the count-data case.

In the matricial formulation, the linearity of the expected value and the  $E[\psi \circ \mathbf{X}] = \psi E[\mathbf{X}]$  property [17] can be used to write the expected value of the STINMA( $q_{m_1, \dots, m_q}$ ) as

$$E[\mathbf{Y}_t] = \sum_{j=1}^q \sum_{\ell=0}^{m_j} \beta_{j\ell} \mathbf{W}^{(\ell)} \boldsymbol{\mu}_\varepsilon + \boldsymbol{\mu}_\varepsilon. \tag{13}$$

Furthermore, the autocovariance function  $\boldsymbol{\Gamma}^*(h)$  at time lag  $h \geq 0$  is given in Theorem 3.1, with complete proof in Appendix B. It is noteworthy that the result in Theorem 3.1 stands for the assumption that all BTOs are performed independently, which implies that calculations involving expected values of cross products are based on the matrix-BTO properties of Franke and Subba Rao [17]. Noticing that  $V[\psi \circ \mathbf{X}] \neq V[\psi \mathbf{X}]$ , the main diagonal of  $\boldsymbol{\Gamma}^*(0)$  (i.e. the component-wise variances of the process) differs for the STINMA and STMA processes. As the BTOs are independent, the variance terms only appear at the  $\boldsymbol{\Gamma}^*(0)$  diagonal, and the STINMA process exhibits off-diagonal covariances in  $\boldsymbol{\Gamma}^*(h)$  for  $h \geq 0$  equal to those of the STMA process [35].

**Theorem 3.1.** *Let  $(\mathbf{Y}_t)$  be the STINMA( $q_{m_1, \dots, m_q}$ ) process*

$$\mathbf{Y}_t = \sum_{j=1}^q \mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{14}$$

where  $\mathbf{B}_j := \sum_{\ell=0}^{m_j} \beta_{j\ell} \mathbf{W}^{(\ell)}$  and  $\boldsymbol{\varepsilon}_t$  is a multivariate i.i.d. process with mean  $\boldsymbol{\mu}_\varepsilon := E[\boldsymbol{\varepsilon}_t] < \infty$  and  $Cov(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+h}^\top)$  defined as in (5). Then, the autocovariance  $\boldsymbol{\Gamma}^*(h) = Cov(\mathbf{Y}_t, \mathbf{Y}_{t+h}^\top)$ ,  $h \geq 0$  is given by

$$\boldsymbol{\Gamma}^*(h) = \sum_{j=1}^{q-h} \mathbf{B}_j \mathbf{G} \mathbf{B}_{j+h}^\top + \mathbf{G} \mathbf{B}_h^\top, \quad 0 < h < q + 1 \tag{15}$$

and

$$\boldsymbol{\Gamma}^*(0) = \sum_{j=1}^q \left( \mathbf{G} (\mathbf{I}_S + \mathbf{B}_j \mathbf{B}_j^\top) + \sum_{\ell=0}^{m_j} \text{diag}(\mathbf{C}^{(\ell)} \boldsymbol{\mu}_\varepsilon) \right), \tag{16}$$

where  $\mathbf{C}^{(\ell)}$  is the  $(S \times S)$ -matrix with entries  $c_{ij}^{(\ell)} = \beta_{j\ell} w_{ij}^{(\ell)} (1 - \beta_{j\ell} w_{ij}^{(\ell)})$ ,  $i, j = 1, \dots, S$  and  $\mathbf{I}_S$  is the  $(S \times S)$  identity matrix. Moreover,  $\boldsymbol{\Gamma}^*(h) = \mathbf{O}_S$ ,  $h \geq q + 1$ .



The space-time autocovariance between the weighted  $r^{\text{th}}$ - and  $k^{\text{th}}$ -order neighbors at time lag  $h \geq 0$  is based on  $\mathbf{\Gamma}^*(h)$  and defined as

$$\gamma_{rk}(h) = \text{tr} \left( \frac{\mathbf{W}^{(k)\top} \mathbf{W}^{(r)} \mathbf{\Gamma}^*(h)}{S} \right), \quad (17)$$

where  $\text{tr}(\cdot)$  represents the trace of the squared matrix [26, 35]. Note that  $\gamma_{rk}(h)$  should not to be confused with the entries  $\gamma_{ij}^*(h)$  of the temporal autocovariance function  $\mathbf{\Gamma}^*(h)$ . Given that  $\mathbf{\Gamma}^*(h) = \mathbf{\Gamma}^*(-h)$ , the space-time autocovariance exhibits the  $\gamma_{rk}(h) = \gamma_{kr}(-h)$  symmetry [34]. Furthermore, we consider the space-time autocorrelation function defined by

$$\rho_{rk}(h) = \frac{\gamma_{rk}(h)}{[\gamma_{rr}(0)\gamma_{kk}(0)]^{1/2}}, \quad (18)$$

which is obtained by normalizing the space-time autocovariance with the variance parcels [35]. The space-time autocovariance  $\gamma_{rk}(h)$  of the STINMA( $q_{m_1, \dots, m_q}$ ) process is calculated by replacing  $\mathbf{\Gamma}^*(h)$  in (17) with the adequate expressions given in Theorem 3.1. The computation of  $\gamma_{rk}(h)$  is straightforward as it solely requires the evaluation of the trace of a product of matrices. Alternatively, general expressions for  $\gamma_{rk}(h)$  can be derived as weighted sums of  $\mathbf{\Gamma}^*(h)$  by taking into account that  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$  for any matrices  $\mathbf{A}$  and  $\mathbf{B}$  of the same dimension. For instance, the calculation of  $\gamma_{00}(h)$  for  $0 \leq h < q + 1$  requires the evaluation of the  $\mathbf{\Gamma}^*(h)$  trace and since  $\mathbf{W}^{(0)} = \mathbf{W}^{(0)\top} = \mathbf{I}_S$ , it follows that

$$\gamma_{00}(h) = v_{00}(\mathbf{G}\mathbf{B}_h^\top) + \sum_{j=1}^{q-h} \left[ v_{00}(\mathbf{B}_j \mathbf{G} \mathbf{B}_{j+h}^\top) + \mathbf{1}_{(h=0)} \sum_{\ell=0}^{m_j} v_{00} \left( \text{diag}(\mathbf{C}^{(\ell)} \boldsymbol{\mu}_\varepsilon) \right) \right], \quad (19)$$

where  $\mathbf{B}_0 = \beta_{00} \mathbf{W}^{(0)} = \mathbf{I}_S$ ,  $v_{rk}(\mathbf{X}) = \text{tr}(\mathbf{W}^{(k)\top} \mathbf{W}^{(r)} \mathbf{X})/S$  and  $\mathbf{1}_{(h=0)}$  is the indicator function. Similar expressions can be obtained for other evaluations of  $\gamma_{rk}(h)$ , however noting that  $\gamma_{rk}(h) = 0$ ,  $h \geq q + 1$  since  $\mathbf{\Gamma}^*(h) = \mathbf{O}_S$ .

Table 1 provides the first- and second-order moments of the STINMA(1<sub>1</sub>) process

$$\mathbf{Y}_t = \beta_{10} \circ \boldsymbol{\varepsilon}_{t-1} + \beta_{11} \mathbf{W}^{(1)} \circ \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (20)$$

by considering all locations as first-order neighbors [45]. The theoretical results presented in Table 1 remain valid for any weighting scheme  $\mathbf{W}^{(1)}$ , including one-dimensional line sites and two-dimensional grid systems [35]. Moreover, the component-wise expected value and variance of the STINMA(1<sub>1</sub>) arise as particular cases of equations (11) and (12) for  $q = m_1 = 1$ , respectively. Similarly, the temporal autocovariance is a particular case of the result in Theorem 3.1 for  $q = m_1 = 1$ . Regarding the space-time autocovariance function  $\gamma_{rk}(h)$ , the expressions for  $h = 0$  and even  $r + k$  exhibit the additional terms related to the

TABLE 1  
First- and second-order moments of the STINMA(1<sub>1</sub>) model.

Moment	Expression
$E[Y_{s,t}]$	$\mu_{\varepsilon_s}(1 + \beta_{10}) + \beta_{11} \sum_{n=1}^S w_{sn}^{(1)} \mu_{\varepsilon_n}$
$V[Y_{s,t}]$	$\sigma_{\varepsilon_s}^2(1 + \beta_{10}^2) + \beta_{10}(1 - \beta_{10})\mu_{\varepsilon_s} + \beta_{11}^2 \sum_{n=1}^S (w_{sn}^{(1)})^2 \sigma_{\varepsilon_n}^2 + \sum_{n=1}^S \beta_{11} w_{sn}^{(1)}(1 - \beta_{11} w_{sn}^{(1)})\mu_{\varepsilon_j}$
$\mathbf{\Gamma}^*(0)$	$\mathbf{G}(\mathbf{I}_S + \beta_{10}^2 + \beta_{10}\beta_{11}\mathbf{W}^{(1)\top} + \beta_{10}\beta_{11}\mathbf{W}^{(1)} + \beta_{11}^2\mathbf{W}^{(1)}\mathbf{W}^{(1)\top}) + \text{diag}(\mathbf{C}^{(0)}\boldsymbol{\mu}_\varepsilon) + \text{diag}(\mathbf{C}^{(1)}\boldsymbol{\mu}_\varepsilon)$
$\mathbf{\Gamma}^*(1)$	$\mathbf{G}(\beta_{10} + \beta_{11}\mathbf{W}^{(1)\top})$
$\mathbf{\Gamma}^*(h)$	$0, h \geq 2$
$\gamma_{00}(0)$	$(1 + \beta_{10}^2)v_{00}(\mathbf{G}) + \beta_{11}^2 v_{00}(\mathbf{W}^{(1)}\mathbf{G}\mathbf{W}^{(1)\top}) + v_{00}(\text{diag}(\mathbf{C}^{(0)}\boldsymbol{\mu}_\varepsilon)) + v_{00}(\text{diag}(\mathbf{C}^{(1)}\boldsymbol{\mu}_\varepsilon))$
$\gamma_{10}(0)$	$\beta_{10}\beta_{11}(v_{10}(\mathbf{W}^{(1)\top}\mathbf{G}) + v_{10}(\mathbf{W}^{(1)}\mathbf{G}))$
$\gamma_{11}(0)$	$(1 + \beta_{10}^2)v_{11}(\mathbf{G}) + \beta_{11}^2 v_{11}(\mathbf{W}^{(1)}\mathbf{G}\mathbf{W}^{(1)\top}) + v_{11}(\text{diag}(\mathbf{C}^{(0)}\boldsymbol{\mu}_\varepsilon)) + v_{11}(\text{diag}(\mathbf{C}^{(1)}\boldsymbol{\mu}_\varepsilon))$
$\gamma_{20}(0)$	$(1 + \beta_{10}^2)v_{20}(\mathbf{G}) + \beta_{11}^2 v_{20}(\mathbf{W}^{(1)}\mathbf{G}\mathbf{W}^{(1)\top}) + v_{20}(\text{diag}(\mathbf{C}^{(0)}\boldsymbol{\mu}_\varepsilon)) + v_{20}(\text{diag}(\mathbf{C}^{(1)}\boldsymbol{\mu}_\varepsilon))$
$\gamma_{21}(0)$	$\beta_{10}\beta_{11}[v_{21}(\mathbf{G}\mathbf{W}^{(1)\top}) + v_{21}(\mathbf{W}^{(1)}\mathbf{G})]$
$\gamma_{22}(0)$	$(1 + \beta_{10}^2)v_{22}(\mathbf{G}) + \beta_{11}^2 v_{22}(\mathbf{W}^{(1)}\mathbf{G}\mathbf{W}^{(1)\top}) + v_{22}(\text{diag}(\mathbf{C}^{(0)}\boldsymbol{\mu}_\varepsilon)) + v_{22}(\text{diag}(\mathbf{C}^{(1)}\boldsymbol{\mu}_\varepsilon))$
$\gamma_{00}(1)$	$\beta_{10}v_{00}(\mathbf{G})$
$\gamma_{10}(1)$	$\beta_{11}v_{10}(\mathbf{G}\mathbf{W}^{(1)\top})$
$\gamma_{11}(1)$	$\beta_{10}v_{11}(\mathbf{G})$
$\gamma_{21}(1)$	$\beta_{11}v_{21}(\mathbf{G}\mathbf{W}^{(1)\top})$
$\gamma_{22}(1)$	$\beta_{10}v_{22}(\mathbf{G})$
$\gamma_{r0}(1)$	$0, r \geq 2$
$\gamma_{rk}(h)$	$0, h \geq 2$

**Remark:**  $v_{rk}(\mathbf{X}) = \text{tr}(\mathbf{W}^{(k)\top} \mathbf{W}^{(r)} \mathbf{X})/S$ ,  $\mathbf{G} = \text{diag}(\sigma_{\varepsilon_1}^2, \dots, \sigma_{\varepsilon_S}^2)$ , and  $\mathbf{C}^{(\ell)}$  is the  $(S \times S)$  matrix with entries  $c_{sn}^{(\ell)} = \beta_{j\ell} w_{sn}^{(\ell)}(1 - \beta_{j\ell} w_{sn}^{(\ell)})$ ,  $s, n = 1, \dots, S$ .

BTO variance. By contrast,  $\gamma_{rk}(1)$  assume simpler expressions that rely either on  $\beta_{10}$  or  $\beta_{11}$ . Finally,  $\gamma_{r0}(1) = 0$  for  $r \geq 2$ , as  $\text{tr}(\mathbf{W}^{(r)}\mathbf{G}\mathbf{W}^{(1)\top}) = 0$ . Note that the results in Table 1 are in accordance with those derived for the conventional STMA(1<sub>1</sub>) with  $\mathbf{G} = \mathbf{I}_S\sigma^2$  [35] except for the autocovariances at lag 0, which exhibit the extra terms due to the BTO variances.

#### 4. The Poisson STINMA(1<sub>1</sub>) model

Since the Poisson distribution plays a central role in integer-valued time series modelling, this section presents further results for the STINMA(1<sub>1</sub>) process in (20) driven by Poisson distributed innovations. This case considers  $S = 3$  locations, linearly placed next to each other as illustrated in Fig. 2.

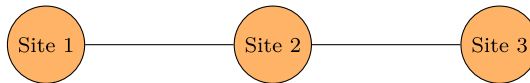


FIG 2. Spatial distribution of 3 linear locations showing (1,2) and (2,3) as two pairs of first-order neighbors.

Following the definition of spatial weights in equation (6) with  $\kappa_{sn}^{(\ell)} = 1/m_s^{(\ell)}$ , the first-order neighbor matrix is given by

$$\mathbf{W}^{(1)} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix} \quad (21)$$

and the STINMA(1<sub>1</sub>) process becomes

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \circ \varepsilon_{1,t-1} + \beta_{11} \circ \varepsilon_{2,t-1} + \varepsilon_{1,t} \\ \beta_{10} \circ \varepsilon_{2,t-1} + \frac{1}{2}\beta_{11} \circ (\varepsilon_{1,t-1} + \varepsilon_{3,t-1}) + \varepsilon_{2,t} \\ \beta_{10} \circ \varepsilon_{3,t-1} + \beta_{11} \circ \varepsilon_{2,t-1} + \varepsilon_{3,t} \end{bmatrix}. \quad (22)$$

With the extra assumption that the innovation process is such that  $\varepsilon_{s,t} \sim Poi(\lambda_s)$ ,  $s = 1, 2, 3$  and the variance-covariance matrix is  $\mathbf{G} = \text{diag}(\boldsymbol{\lambda})$  with  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^\top$ , the process  $\mathbf{Y}_t$  in (22) is addressed as Poisson STINMA(1<sub>1</sub>).

#### 4.1. Moments and other distributional properties

Theorem 4.1 states that the joint distribution of the Poisson STINMA(1<sub>1</sub>) process with  $S = 3$  is a trivariate Poisson distribution; see Appendix C for details. It is shown that, regardless of the innovations' distribution the probability generating function (p.g.f.) of any STINMA(1<sub>1</sub>) process can be written as the product of six p.g.f. functions, associated with the marginal p.g.f.s of the trivariate innovation process and the parameters related to the STINMA recursion. Furthermore, the result in Theorem 4.1 stands independently of the weighting scheme in  $\mathbf{W}^{(1)}$  as long as the BTO restrictions are satisfied. Moreover, it is shown that the p.g.f. of the Poisson STINMA(1<sub>1</sub>) process can also arise by applying the reduction method with seven independent univariate Poisson r.v.'s [25]. This work considers the notation  $\mathbf{X} = (X_1, X_2, X_3)^\top \sim TPoi(a_1, a_2, a_3, a_{12}, a_{13}, a_{23}, a_{123})$  as a trivariate Poisson random vector with  $2^3 - 1 = 7$  nonnegative parameters whenever

$$X_r = Z_r + \sum_{\substack{k=1 \\ r < k}}^3 Z_{rk} + Z_{123}, \quad r, k = 1, 2, 3 \quad \text{and} \quad r < k, \quad (23)$$

being  $Z_r$ ,  $Z_{rk}$ ,  $Z_{123}$  independent univariate Poisson r.v.'s with parameters  $a_r$ ,  $a_{rk}$ ,  $a_{123}$ , respectively. For simplicity, the notation for the multivariate distribution of vector  $\mathbf{X} = (X_1, X_2, X_3)^\top \sim TPoi(a_1, a_2, a_3) := TPoi(a_1, a_2, a_3, 0, 0, 0, 0)$  is used whenever  $a_{12} = a_{13} = a_{23} = a_{123} = 0$ , i.e., a trivariate Poisson distribution with mutually independent (uncorrelated) components.

**Theorem 4.1.** *Let  $\mathbf{Y}_t$  be the STINMA(1<sub>1</sub>) process in (22) and let the innovations  $\varepsilon_t \sim TPoi(\lambda_1, \lambda_2, \lambda_3)$ . It follows that  $\mathbf{Y}_t \sim TPoi(a_1, a_2, a_3, a_{12}, a_{13}, a_{23}, a_{123})$  where*

$$a_1 = \lambda_1(1 + \beta_{10} - \frac{1}{2}\beta_{10}\beta_{11}) + \lambda_2(1 - \beta_{10} - \beta_{11} + \beta_{10}\beta_{11})\beta_{11},$$

$$\begin{aligned}
 a_2 &= \lambda_1(\frac{1}{2} - \beta_{10})\beta_{11} + \lambda_2(1 + \beta_{10} - 2\beta_{11}\beta_{10} + \beta_{10}\beta_{11}^2) + \lambda_3(\frac{1}{2}\beta_{11} - \frac{1}{2}\beta_{10}\beta_{11}), \\
 a_3 &= \lambda_3(1 + \beta_{10} - \frac{1}{2}\beta_{10}\beta_{11}) + \lambda_2(1 - \beta_{10}\beta_{11} + \beta_{10}\beta_{11})\beta_{11}, \\
 a_{12} &= \lambda_1\frac{1}{2}\beta_{10}\beta_{11} + \lambda_2(1 - \beta_{11})\beta_{10}\beta_{11}, \\
 a_{13} &= \lambda_2(1 - \beta_{10})\beta_{11}^2, \\
 a_{23} &= \lambda_3\frac{1}{2}\beta_{10}\beta_{11} + \lambda_2(1 - \beta_{11})\beta_{10}\beta_{11}, \\
 a_{123} &= \lambda_2\beta_{10}\beta_{11}^2.
 \end{aligned}$$

The proof of Theorem 4.1 is postponed to the Appendix C considering a general  $\mathbf{W}^{(1)}$  matrix with  $S = 3$ . Clearly, Theorem 4.1 can be generalized for any dimension  $S > 3$ , by following the proof for  $S = 3$  and acknowledging that the results stand true by changing the upper limits in sums and products of parcels along the proof.

It follows from the trivariate Poisson distribution of the Poisson STINMA(1<sub>1</sub>) process that

$$E[\mathbf{Y}_t] = V[\mathbf{Y}_t] = \boldsymbol{\lambda} + \beta_{10}\boldsymbol{\lambda} + \beta_{11}\mathbf{W}^{(1)}\boldsymbol{\lambda}. \tag{24}$$

Regarding the autocovariance of the Poisson STINMA(1<sub>1</sub>) process,  $\boldsymbol{\Gamma}^*(h) = \mathbf{O}_S$ ,  $h \geq 2$ . Also, it follows from the result in Theorem 3.1 that

$$\boldsymbol{\Gamma}^*(0) = \begin{bmatrix} (1 + \beta_{10})\lambda_1 + \beta_{11}\lambda_2 & \beta_{10}\beta_{11}(\frac{1}{2}\lambda_1 + \lambda_2) & \beta_{11}^2\lambda_2 \\ \beta_{10}\beta_{11}(\frac{1}{2}\lambda_1 + \lambda_2) & (1 + \beta_{10})\lambda_2 + \frac{1}{2}\beta_{11}(\lambda_1 + \lambda_3) & \beta_{10}\beta_{11}(\lambda_2 + \frac{1}{2}\lambda_3) \\ \beta_{11}^2\lambda_2 & \beta_{10}\beta_{11}(\lambda_2 + \frac{1}{2}\lambda_3) & (1 + \beta_{10})\lambda_3 + \beta_{11}\lambda_2 \end{bmatrix} \tag{25}$$

and

$$\boldsymbol{\Gamma}^*(1) = \begin{bmatrix} \beta_{10}\lambda_1 & \frac{1}{2}\beta_{11}\lambda_1 & 0 \\ \beta_{11}\lambda_2 & \beta_{10}\lambda_2 & \beta_{11}\lambda_2 \\ 0 & \frac{1}{2}\beta_{11}\lambda_3 & \beta_{10}\lambda_3 \end{bmatrix}. \tag{26}$$

It is easy to check that the  $\boldsymbol{\Gamma}^*(0)$  of the Poisson STINMA(1<sub>1</sub>) is equivalent to that of the conventional STMA with  $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_S, \mathbf{G})$  and  $\mathbf{W}^{(1)}$  as in equation (21) [36]. Furthermore, it is important to highlight that  $\boldsymbol{\Gamma}^*(1)$  is not necessarily symmetric and that  $\gamma_{13}^*(1) = \gamma_{31}^*(1) = 0$  as locations 1 and 3 are not first-order neighbors. Finally, the space-time autocovariance and autocorrelation functions follow directly by plugging  $\boldsymbol{\Gamma}^*(h)$  into equations (17) and (18), respectively, and constitute a particular case of the expressions presented in Table 1.

A closer look at the STINMA(1<sub>1</sub>) process in (22) highlights that  $\beta_{11} \circ \varepsilon_{2,t-1}$  is performed twice for each time  $t$ . With the BTO independence assumption, both evaluations are not necessarily equal, for a given  $t$ , even if location 2 equally contributes for locations 1 and 3. The alternative assumption that both evaluations should be equal for each time  $t$ , leads to the reparametrization of

the model (22) into

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \circ \varepsilon_{1,t-1} + r_t + \varepsilon_{1,t} \\ \beta_{10} \circ \varepsilon_{2,t-1} + \frac{1}{2}\beta_{11} \circ (\varepsilon_{1,t-1} + \varepsilon_{3,t-1}) + \varepsilon_{2,t} \\ \beta_{10} \circ \varepsilon_{3,t-1} + r_t + \varepsilon_{3,t} \end{bmatrix}, \quad (27)$$

where  $r_t := \beta_{11} \circ \varepsilon_{2,t-1}$  represents the common factor in both equations. Under this assumption, the BTOs are still performed independently in time, similarly to the INARMA model, but not independently in space.

The reparameterized STINMA(1<sub>1</sub>) process in (27) also follows a trivariate Poisson distribution, being  $\mathbf{Y}_t \sim TPoi(a_1, a_2, a_3, a_{12}, a_{13}, a_{23}, a_{123})$  with

$$\begin{aligned} a_1 &= \lambda_1(1 + \beta_{10} - \frac{1}{2}\beta_{10}\beta_{11}), \\ a_2 &= \lambda_2(1 + \beta_{10} - \beta_{10}\beta_{11}) + (\lambda_1 + \lambda_3)(\frac{1}{2}\beta_{11}) - \frac{1}{2}\beta_{10}\beta_{11}, \\ a_3 &= \lambda_3(1 + \beta_{10} - \frac{1}{2}\beta_{10}\beta_{11}), \\ a_{12} &= \lambda_1\frac{1}{2}\beta_{10}\beta_{11}, \quad a_{13} = \lambda_2(\beta_{11} - \beta_{10}\beta_{11}), \quad a_{23} = \lambda_3\frac{1}{2}\beta_{10}\beta_{11}, \\ a_{123} &= \lambda_2\beta_{10}\beta_{11}, \end{aligned} \quad (28)$$

as clarified by Remark C.1. Note that the set of parameters of the joint distribution for the STINMA(1<sub>1</sub>) processes in (22) and (27) differ. In particular, it is clear that the parameter  $a_2$  in the (27) formulation (see equation (28)) contemplates a joint contribution with the term  $(\lambda_1 + \lambda_3)$ , while for the (22) formulation,  $\lambda_1$  and  $\lambda_3$  have different terms/contributions (see Theorem 4.1).

The expected value and variance of the reparametrized Poisson STINMA(1<sub>1</sub>) process are still those given in (24). The autocovariance  $\mathbf{\Gamma}^*(h)$  of the reparameterized process, however, does not follow from the result in Theorem 3.1 as this is derived with matrix-BTO properties assuming independent  $\circ$ -operations [17]. Indeed, the autocovariance of the reparameterized model can be obtained by the arguments presented in the proof in Appendix B, along with the result in Lemma 4.2 for the calculation of the expected value of cross products.

**Lemma 4.2.** *Let  $\mathbf{A}$  be a matrix with fixed entries in  $[0, 1)$  and  $\mathbf{X}$  be a trivariate count random variable. Assume that all the counting series in  $\mathbf{A} \circ \mathbf{X}$  are independent, except for the  $\circ$ -operations involving the same constant and r.v., for which the realisation of the underlying Bernoulli sequence coincide. Then, it follows that*

$$E[(\mathbf{A} \circ \mathbf{X})(\mathbf{A} \circ \mathbf{X})^\top] = \mathbf{A}E[\mathbf{X}\mathbf{X}^\top]\mathbf{A}^\top + \sum_{r=1}^3 \mathbf{Q}^{(r)}E[\mathbf{X}]e_r^\top, \quad (29)$$

where  $\mathbf{Q}^{(r)}$ ,  $r = 1, 2, 3$ , is the  $(3 \times 3)$ -matrix with entries  $q_{ij}^{(r)} = \mathbb{1}_{(a_{ij}=a_{rj})}a_{ij}(1 - a_{ij})$ ,  $i, j = 1, 2, 3$ , and  $e_r$  denotes the  $r^{\text{th}}$  unit vector.

Notice that the matrix-BTO property introduced in (29) generalizes the corresponding result of [17], admitting off-diagonal variance parcels. The proof of

the Lemma can be found in Appendix D. Nevertheless, it is clear that this result can be easily generalized to a higher dimension by considering that  $\mathbf{X}$  is a multivariate vector and, by replacing the summation  $r = 1, 2, 3$  for  $r = 1, \dots, S$ .

The assumption of a common factor in the STINMA(1<sub>1</sub>) process changes (solely) the covariance matrix  $\mathbf{\Gamma}^*(0)$  for the independent case (see (25)). Specifically, the covariance between  $Y_{1,t}$  and  $Y_{3,t}$  is now given by  $\gamma_{13}^*(0) = Cov(r_t, r_t) = V[r_t] = V[\beta_{11} \circ \varepsilon_{2,t-1}] = V[\beta_{11}\varepsilon_{2,t-1}] + E[\beta_{11}(1 - \beta_{11})\varepsilon_{2,t-1}] = \beta_{11}\lambda_2$  instead of  $\gamma_{13}^*(0) = Cov(\beta_{11} \circ \varepsilon_{2,t-1}, \beta_{11} \circ \varepsilon_{2,t-1}) = \beta_{11}^2\lambda_2$ , which is larger if considering the common factor  $r_t = \beta_{11} \circ \varepsilon_{2,t-1}$ . This modification in  $\mathbf{\Gamma}^*(0)$  has an inherent repercussion in the space-time  $\gamma_{11}(0)$ , while  $\gamma_{00}(0)$  and  $\gamma_{10}(0)$  remain unchanged. Therefore, space-time autocorrelations involving the calculation of  $\gamma_{11}(0)$ , such as  $\rho_{10}(1)$ , will also be modified. Table 2 presents the space-time autocovariance  $\gamma_{rk}(h)$  for the Poisson STINMA(1<sub>1</sub>) in (27), evaluated for several  $r, k$ , and  $h$  values. Due to the equidispersion property of the Poisson distribution, there are no terms in  $\beta_{10}^2$  nor in  $\beta_{11}^2$  appearing in the Poisson-case  $\gamma_{rk}(h)$ , by contrast to that of the general model displayed in Table 1. Moreover, note that  $\gamma_{rk}(0)$  depends on both  $\beta_{10}$  and  $\beta_{11}$ , while  $\gamma_{rk}(1)$  relies either on  $\beta_{10}$  or  $\beta_{11}$ .

TABLE 2  
Space-time autocovariance  $\gamma_{rk}(h)$  of the Poisson STINMA(1<sub>1</sub>) process in (27), evaluated for several  $\{r, k, h\}$  values.

$\gamma_{rk}(h)$	Expression
$\gamma_{00}(0)$	$(1 + \beta_{10})(\frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)) + \beta_{11}(\frac{1}{6}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2)$
$\gamma_{10}(0)$	$\beta_{10}\beta_{11}(\lambda_2 + \frac{1}{4}(\lambda_1 + \lambda_3))$
$\gamma_{11}(0)$	$(1 + \beta_{10})(\frac{1}{12}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2) + \beta_{11}(\frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3))$
$\gamma_{20}(0)$	$\beta_{11}(\frac{2}{3}\lambda_2)$
$\gamma_{21}(0)$	$\beta_{10}\beta_{11}(\frac{1}{6}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2)$
$\gamma_{22}(0)$	$(1 + \beta_{10})(\frac{1}{3}(\lambda_1 + \lambda_3)) + \beta_{11}(\frac{2}{3}\lambda_2)$
$\gamma_{00}(1)$	$\beta_{10}(\frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3))$
$\gamma_{10}(1)$	$\beta_{11}(\frac{1}{12}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2)$
$\gamma_{11}(1)$	$\beta_{10}(\frac{1}{12}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2)$
$\gamma_{21}(1)$	$\beta_{11}(\frac{1}{6}(\lambda_1 + \lambda_3))$
$\gamma_{22}(1)$	$\beta_{10}(\frac{1}{3}(\lambda_1 + \lambda_3))$
$\gamma_{r0}(1)$	$0, r \geq 2$

It is straightforward that  $\rho_{00}(0) = \rho_{11}(0) = 1$ , and  $\rho_{10}(0)$  exhibits crossed and squared terms in  $\beta_{10}$  and  $\beta_{11}$ , which arguments against the use of equations based on  $\rho_{rk}(0)$  for parameter estimation. On the other hand,  $\rho_{rk}(1)$  exhibits dependency either on  $\beta_{10}$  or  $\beta_{11}$ , e.g.

$$\rho_{00} := \rho_{00}(1) = \frac{\beta_{10}a}{(1 + \beta_{10})a + \beta_{11}b}, \tag{30}$$

$$\rho_{10} := \rho_{10}(1) = \frac{\beta_{11}c}{[[ (1 + \beta_{10})a + \beta_{11}b ] [ (1 + \beta_{10})c + \beta_{11}a ] ]^{\frac{1}{2}}}, \tag{31}$$

where  $a = \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)$ ,  $b = \frac{1}{6}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2$  and  $c = \frac{1}{12}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2$ . Figure 3 illustrates that  $\rho_{00}$  and  $\rho_{10}$  show a monotonic and inverse variation with changes in  $\beta_{10}$  and  $\beta_{11}$ . In addition,  $\rho_{00}$  in (a) increases for increasing  $\beta_{10}$  and for decreasing  $\beta_{11}$  while  $\rho_{10}$  in (b) increases for decreasing  $\beta_{10}$  and increasing  $\beta_{11}$ . Figure 3 also highlights that  $\rho_{00}$  exhibits larger variations with changes in  $\beta_{10}$  rather than in  $\beta_{11}$ , whereas  $\rho_{10}$  exhibits larger variations for changes in  $\beta_{11}$ . Furthermore, the Poisson parameters have a negligible effect on  $\rho_{00}$  although having a greater impact on  $\rho_{10}$ , in particular for increasing values of  $\beta_{11}$  for which the surfaces exhibit a larger variability for a given combination of  $\beta_{10}$  and  $\beta_{11}$ .

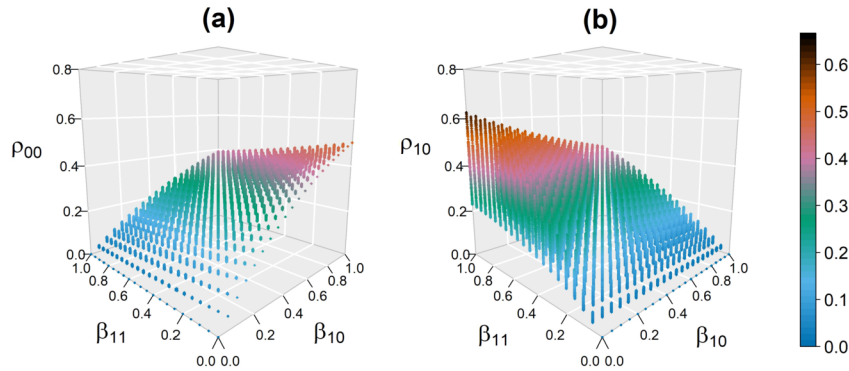


FIG 3. Theoretical  $\rho_{00}$  (a) and  $\rho_{10}$  (b) for the Poisson STINMA(1<sub>1</sub>) process in (27) as a function of  $\beta_{10}$  and  $\beta_{11}$ , and evaluated for several  $\{\lambda_1, \lambda_2, \lambda_3\} \in (0, 10)$ . The color bar is set for the range  $0 \leq \rho_{10} < 2/3$ .

It is clear that both  $\rho_{00}$  and  $\rho_{10}$  are bounded functions with range smaller than the  $(-1, 1)$  interval. Moreover,  $0 \leq \rho_{00} < 0.5$  is obtained by setting  $\beta_{10} = 0$  or  $(\beta_{10} \rightarrow 1, \beta_{11} = 0)$  in (30), which is corroborated in Figure 3(a). Furthermore,  $0 \leq \rho_{10} < \frac{2}{3}$ , where the lower bound is determined for  $\beta_{11} = 0$  in (31) and corroborated in Figure 3(b). The upper bound is less trivial due to its dependencies on parameters  $(\lambda_1, \lambda_2, \lambda_3, \beta_{10}$  and  $\beta_{11})$ . However,  $\rho_{10}$  exhibits a monotonic behaviour with  $\lambda_1, \lambda_2$ , and  $\lambda_3$ , namely increasing for larger  $\lambda_2$  and decreasing for larger  $\lambda_1 + \lambda_3$ , regardless of  $\beta_{10}$  and  $\beta_{11}$ . As  $\rho_{10}$  is maximal if  $\beta_{10} = 0$  and  $\beta_{11} \rightarrow 1$ , and by studying the limits for the Poisson parameters, one can conclude that  $\rho_{10}$  is maximal (and tends to  $2/3$ ) for  $\lambda_2 \rightarrow +\infty$  and  $\lambda_1 + \lambda_3 \rightarrow 0$ .

The  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$  setting in equations (30) and (31) leads to  $a = b = \lambda$  and  $c = \frac{5}{6}\lambda$  and, consequently,  $\rho_{00}$  and  $\rho_{10}$  do not depend on the Poisson parameter  $\lambda$ . Figure 4(a,b) displays the shape of the theoretical  $\rho_{00}$  and  $\rho_{10}$  as a function of  $\beta_{10}$  and  $\beta_{11}$ , again showing the inverse relation of these functions with respect to  $\beta_{10}$  and  $\beta_{11}$  changes. Also, the comparison between Figures 4(a,b) and 4(c,d) shows that the theoretical and empirical surfaces, obtained from the sample autocorrelation of simulated STINMA(1<sub>1</sub>) paths, exhibit good agreement with each other.

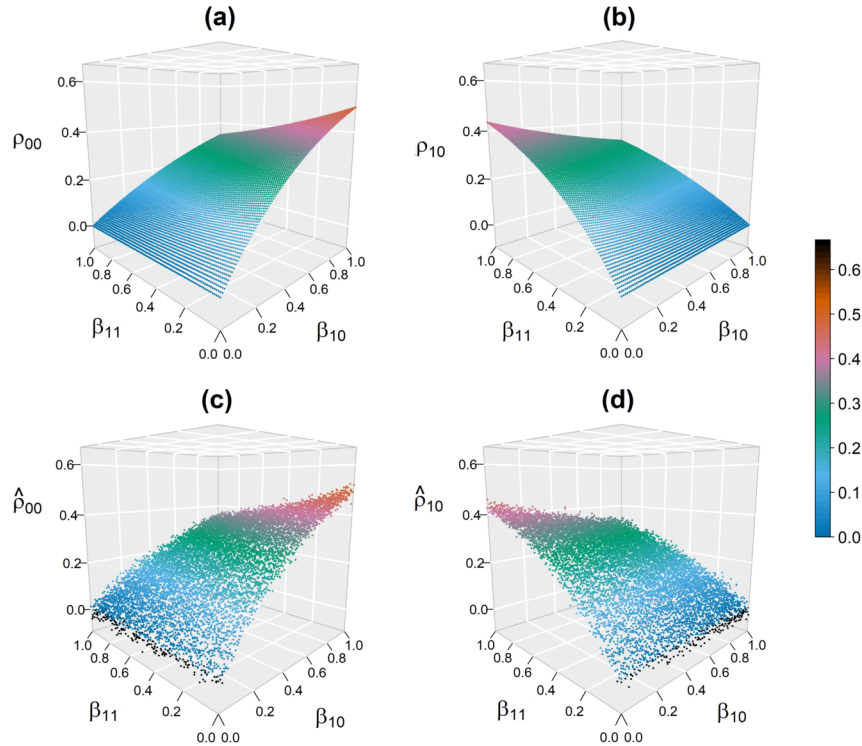


FIG 4. Theoretical (a,b) and empirical (c,d)  $\rho_{00}$  and  $\rho_{10}$  for  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ , as a function of  $\beta_{10}$  and  $\beta_{11}$ . The empirical values are based on the sample autocorrelation of a STINMA(1<sub>1</sub>) path generated for each  $(\beta_{10}, \beta_{11})$  combination. The black points identify the paths with empirical autocorrelation outside the theoretical range (approx. 8% of the paths).

Given the boundedness and the monotonic behaviour of  $\rho_{00}$  and  $\rho_{10}$  in (30) and (31), respectively, it is possible to construct an analytical region containing the  $(\rho_{00}, \rho_{10})$  values for all admissible Poisson STINMA(1<sub>1</sub>) processes (i.e. those for which  $\beta_{10}, \beta_{11} \in [0, 1)$  and  $\lambda_i > 0, i = 1, 2, 3$ ). The delimitation of such region is obtained by varying the  $\beta$ - and  $\lambda$ -parameters within the admissible intervals in the following equation

$$\rho_{10} = \frac{\beta_{11}c}{(\beta_{10}a)^{\frac{1}{2}} \left( [(1 + \beta_{10})c + \beta_{11}a] \right)^{\frac{1}{2}}} (\rho_{00})^{\frac{1}{2}}, \tag{32}$$

where  $a = \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)$  and  $c = \frac{1}{12}(\lambda_1 + \lambda_3) + \frac{2}{3}\lambda_2$ , and  $\rho_{00}$  is evaluated according to (30).

As illustrated in Figure 5, different settings for  $\lambda$  generate admissible regions with different areas but similar shapes, delimited by a trapezoid-like contour with 4 anchor points obtained from the 4 combinations among the limiting values for the  $\beta$ 's. The black contour (circles) in Figure 5 delimits the admissible region with the largest area possible, which is obtained for  $\lambda_2 \rightarrow +\infty$  and  $\lambda_1 + \lambda_3 \rightarrow 0$ .



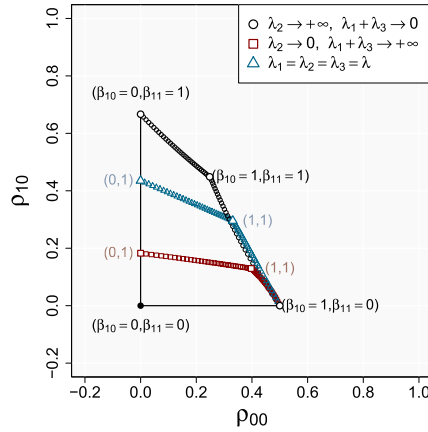


FIG 5. Theoretical region containing the  $(\rho_{00}, \rho_{10})$  values for all admissible Poisson STINMA(1<sub>1</sub>) processes (i.e. those with  $\beta_{10}$  and  $\beta_{11} \in [0, 1]$  and  $\lambda_i > 0$ ,  $i = 1, 2, 3$ ), for  $\lambda_2 \rightarrow +\infty$  and  $\lambda_1 + \lambda_3 \rightarrow 0$  (black circles),  $\lambda_2 \rightarrow 0$  and  $\lambda_1 + \lambda_3 \rightarrow +\infty$  (red squares) and  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$  (blue triangles). The corresponding  $(\beta_{10}, \beta_{11})$  coordinates are presented for each anchor point of the admissible regions.

For this case, the above equation simplifies to

$$\rho_{10} = \frac{\beta_{11} \frac{2}{3}}{(\beta_{10} \frac{1}{3})^{\frac{1}{2}} \left( [(1 + \beta_{10}) \frac{2}{3} + \beta_{11} \frac{1}{3}] \right)^{\frac{1}{2}}} (\rho_{00})^{\frac{1}{2}}, \tag{33}$$

which allows to evaluate the vertices and edges of this contour. For example, the  $(\rho_{00}, \rho_{10})$  vertices are  $\{(0, 0), (0.5, 0), (0.25, \sqrt{20}/10), (0, 2/3)\}$  for the  $(\beta_{10}, \beta_{11})$  limits in  $\{(0, 0), (1, 0), (1, 1), (0, 1)\}$ .

Figure 5 also shows a red contour (squares) that delimits the admissible region with the smallest area, constructed similarly by setting  $\lambda_2 \rightarrow 0$  and  $\lambda_1 + \lambda_3 \rightarrow +\infty$ . Other  $\lambda$  values lead to admissible regions with limits in-between the previous cases, e.g. the blue contour (triangles) constructed for the  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$  setting. Overall, Figure 5 highlights that the admissible region for  $(\rho_{00}, \rho_{10})$  of the Poisson STINMA(1<sub>1</sub>) is smaller than the possible full  $[-1, 1]$  range, only includes positive values, and its size depends on the values in  $\lambda$ .

#### 4.2. Parameter estimation of STINMA

This section discusses parameter estimation strategies for the Poisson STINMA(1<sub>1</sub>) model in (27) with  $\varepsilon_t \sim TPoi(\lambda_1, \lambda_2, \lambda_3)$ , based on a trivariate series of  $T$  observations of the process. Estimation strategies for  $\theta := (\lambda^\top, \beta_{10}, \beta_{11})$ , with  $\lambda := (\lambda_1, \lambda_2, \lambda_3)^\top$ , consider the Method of Moments (MM), Conditional Least Squares (CLS) and Conditional Maximum Likelihood (CML). Furthermore, inference based on parametric bootstrap is described to provide a strategy to compare the aforementioned estimation approaches.

All computations were accomplished through the statistical software R, version 4.1.2 [51]. The MM estimation was based on the Newton–Raphson numerical procedure (package *rootSolve*, [43]). The CLM was efficiently implemented using sparse matrices (*Matrix*, [4]), compiled C++ code (*Rcpp* [14] and *RcppArmadillo* [15]), and parallelized optimization procedures (*optimParallel* [18]). The packages *R.utils* [5], *HelpersMG* [21], *doParallel* [10] were additionally used to aid estimation. Data manipulation and visualisation was carried out with *ggplot2* [56], *patchwork* and *plot3D* [33], *sf* [29], *dplyr* [57], *ggsn* [3] and *ggforce* [32].

4.2.1. Method of moments (MM)

The MM requires the construction of a system with  $S + 2$  independent equations to estimate the  $S$  innovational parameters  $\boldsymbol{\lambda}$  in addition to  $(\beta_{10}, \beta_{11})$ , to be solved numerically via Newton-Raphson algorithm. The first  $S$  equations arise from the relation between the first-order moments of  $\mathbf{Y}_t$  and  $\boldsymbol{\varepsilon}_t$ , in equation (24), where  $E(\mathbf{Y}_t)$  is replaced by its empirical counterpart  $\bar{\mathbf{Y}} = (\bar{Y}_1, \bar{Y}_2, \bar{Y}_3)$ . Note that, for the  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$  case, the (univariate) equation (24) is also used to estimate  $\lambda$ . The MM-estimation of  $(\beta_{10}, \beta_{11})$  should be based on sample space-time autocorrelations, avoiding second-order neighbors and  $\beta_{10}\beta_{11}$  terms to simplify the estimation procedure. As presented in Table 2,  $\beta_{10}$  estimation should consider either  $\gamma_{00}(1)$  or  $\gamma_{11}(1)$ , whereas  $\beta_{11}$  estimation should be based on  $\gamma_{10}(1)$ . Therefore,  $(\bar{\mathbf{Y}}, \hat{\rho}_{00}, \hat{\rho}_{10})$  and  $(\bar{\mathbf{Y}}, \hat{\rho}_{11}, \hat{\rho}_{10})$  will be considered as two competing estimation alternatives, where  $\hat{\rho}_{00}$ ,  $\hat{\rho}_{10}$  and  $\hat{\rho}_{11}$  are the sample estimators for  $\rho_{00}$  and  $\rho_{10}$  given in (30) and (31), respectively, and

$$\rho_{11} := \rho_{11}(1) = \frac{\beta_{10}c}{(1 + \beta_{10})c + \beta_{11}a}, \tag{34}$$

with  $a$  and  $c$  as previously defined for  $\rho_{00}$  and  $\rho_{10}$ . This approach makes use of the following estimators for the mean and the (temporal) covariance

$$\bar{\mathbf{Y}} = \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_t, \quad \text{and} \quad \hat{\boldsymbol{\Gamma}}^*(h) = \frac{1}{T} \sum_{t=1}^{T-h} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t+h} - \bar{\mathbf{Y}})^\top.$$

The proposed MM-estimation for the Poisson STINMA(1<sub>1</sub>) model shares similarities with that of other models. Namely, the sample space-time autocorrelations  $(\hat{\rho}_{00}, \hat{\rho}_{10})$  are used in the MM-estimation of  $(\beta_{10}, \beta_{11})$  for the conventional STMA(1<sub>1</sub>) model [35, 36] while the MM-estimation of  $(\lambda, \beta_{10})$  in the univariate INMA(1) process  $Y_t = \beta_{10} \circ \varepsilon_{t-1} + \varepsilon_t$  with  $\varepsilon_t \sim Poi(\lambda)$ , considers the sample mean  $\bar{Y}$  and the (temporal) autocorrelation at lag 1 [2].

4.2.2. Conditional least squares (CLS)

The CLS approach is based on the minimization of the sum of squared deviations

$$S(\boldsymbol{\lambda}) = \sum_{t=q+1}^T \mathbf{e}_t^\top \mathbf{e}_t = \sum_{t=2}^T (\boldsymbol{\varepsilon}_t - \boldsymbol{\lambda})^\top (\boldsymbol{\varepsilon}_t - \boldsymbol{\lambda}), \tag{35}$$

where  $\mathbf{e}_t = \boldsymbol{\varepsilon}_t - \boldsymbol{\lambda}$  is the error function [7, 8, 38]. Straightforwardly, the roots of the partial derivatives of  $S(\boldsymbol{\lambda})$  lead to the estimator

$$\hat{\boldsymbol{\lambda}} = \frac{1}{T-1} \sum_{t=2}^T \boldsymbol{\varepsilon}_t. \tag{36}$$

The  $\beta_{10}$  and  $\beta_{11}$  estimates are obtained by solving two out of the three equations

$$E[Y_{1,t}|\boldsymbol{\varepsilon}_{t-1}] = \lambda_1 + \beta_{10}\varepsilon_{1,t-1} + \beta_{11}\varepsilon_{2,t-1}, \tag{37}$$

$$E[Y_{2,t}|\boldsymbol{\varepsilon}_{t-1}] = \lambda_2 + \beta_{10}\varepsilon_{2,t-1} + \frac{1}{2}\beta_{11}(\varepsilon_{1,t-1} + \varepsilon_{3,t-1}), \tag{38}$$

$$E[Y_{3,t}|\boldsymbol{\varepsilon}_{t-1}] = \lambda_3 + \beta_{10}\varepsilon_{3,t-1} + \beta_{11}\varepsilon_{2,t-1}, \tag{39}$$

which correspond to the component-wise conditional mean of the STINMA(1<sub>1</sub>) process. The possible combinations of equations (37)–(39) lead to three different estimators  $(\hat{\boldsymbol{\lambda}}, \bar{Y}_i, \bar{Y}_j)$ ,  $i, j = 1, 2, 3$ ,  $i \neq j$ , which are evaluated as opposing strategies. The unobserved innovations are recursively computed from  $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \beta_{10} \circ \boldsymbol{\varepsilon}_{t-1} - \beta_{11} \mathbf{W}^{(1)} \circ \boldsymbol{\varepsilon}_{t-1}$ , where  $\mathbf{y}_t$  are the observed count time series,  $\boldsymbol{\varepsilon}_0 = (c_1, c_2, c_3)^\top \in \mathbb{N}_0^3$  is constant and the MM-estimates for  $\beta_{10}$  and  $\beta_{11}$  are used as initial estimates. For the  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$  case, it is clear that equations (37)–(39) are redundant and do not allow the estimation of  $\beta_{10}$  and  $\beta_{11}$ . In such case, an additional equation based on a second-order moment that leads to admissible solutions (e.g.  $\rho_{00}$ ) should be used.

#### 4.2.3. Conditional maximum likelihood (CML)

The CML approach proposed for the Poisson STINMA(1<sub>1</sub>) is based on the ML estimation procedure for Hidden-Markov models [58] and constitutes a multivariate extension of the CML strategy developed for the INARMA(1, 1) process [55]. Instead of obtaining the  $\boldsymbol{\theta}$  estimate as the maximizer of

$$L(\boldsymbol{\theta}) = P(\mathbf{Y}_T = \mathbf{y}_T, \mathbf{Y}_{T-1} = \mathbf{y}_{T-1}, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1), \tag{40}$$

the likelihood function is decomposed into the following sum of parcels

$$L(\boldsymbol{\theta}) = \sum_{k_1, \ell_1}^{M_1} \sum_{k_2, \ell_2}^{M_2} \sum_{k_3, \ell_3}^{M_3} b_{k_1, k_2, k_3, \ell_1, \ell_2, \ell_3}(t) = \sum_{\mathbf{k}, \boldsymbol{\ell}}^M b_{\mathbf{k}\boldsymbol{\ell}}(t), \tag{41}$$

where, for a given  $\mathbf{k} = (k_1, k_2, k_3)$  and  $\boldsymbol{\ell} = (\ell_1, \ell_2, \ell_3)$ , each parcel is

$$b_{\mathbf{k}\boldsymbol{\ell}}(t) = P(\boldsymbol{\varepsilon}_t = \mathbf{k}, \boldsymbol{\varepsilon}_{t-1} = \boldsymbol{\ell}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1). \tag{42}$$

Given that  $\boldsymbol{\varepsilon}_t \leq_{st} \mathbf{Y}_t$  for all  $t$  (the operator “ $\leq_{st}$ ” is defined via  $\mathbf{X} \leq_{st} \mathbf{Z}$  iff  $P(\mathbf{X} > \mathbf{x}) \leq P(\mathbf{Z} > \mathbf{x})$ ,  $\mathbf{x} \in \mathbb{N}_0^3$ ) then the upper limits of the summations in

(41) are finite and can be set as  $\mathbf{M} = (M_1, M_2, M_3)$  where  $M_i = \max_t \{y_{i,t}\}$  for  $i = 1, 2, 3$ . Moreover,  $b_{\mathbf{k}\ell}(t)$  can be obtained recursively from

$$\begin{aligned}
 b_{\mathbf{k}\ell}(t+1) &= P(\varepsilon_{t+1} = \mathbf{k}) \times \left[ \sum_{j_2=0}^{J_{k_1 k_3}} \binom{\ell_2}{j_2} \beta_{11}^{j_2} (1 - \beta_{11})^{\ell_2 - j_2} \binom{\ell_1}{y_{1,t+1} - k_1 - j_2} \right. \\
 &\times \beta_{10}^{y_{1,t+1} - k_1 - j_2} (1 - \beta_{10})^{\ell_1 - y_{1,t+1} + k_1 + j_2} \binom{\ell_3}{y_{3,t+1} - k_3 - j_2} \beta_{10}^{y_{3,t+1} - k_3 - j_2} \\
 &\times (1 - \beta_{10})^{\ell_3 - y_{3,t+1} + k_3 + j_2} \left. \right] \times \left[ \sum_{j_2=0}^{y_{2,t} - k_2} \binom{\ell_2}{j_2} \beta_{10}^{j_2} (1 - \beta_{10})^{\ell_2 - j_2} \binom{\ell_1 + \ell_3}{y_{2,t} - k_2 - j_2} \right. \\
 &\times \left. \left( \frac{1}{2} \beta_{11} \right)^{y_{2,t} - k_2 - j_2} \left( 1 - \frac{1}{2} \beta_{11} \right)^{\ell_1 + \ell_3 - y_{2,t} + k_2 + j_2} \right] \times \sum_{i=\mathbf{0}_3}^{\mathbf{y}_{t-1}} b_{\ell i}(t), \tag{43}
 \end{aligned}$$

where  $J_{k_1 k_3} := \min_t \{y_{1,t+1} - k_1, y_{3,t+1} - k_3\}$ . Appendix E presents the detailed derivation of the recursion in (43). This equation can be rewritten in the general form  $\mathbf{a}_t = \mathbf{D}\mathbf{Q}_t\mathbf{a}_{t-1}$ , where  $\mathbf{D} = \text{diag}(P(\varepsilon = \mathbf{0}_3), \dots, P(\varepsilon = \mathbf{M}))$  conveys the innovations' joint probabilities for any combination in  $\mathbf{k}$ , and  $\mathbf{Q}_t$  represents the matrix that stores in each entry the product of convolutions for a given combination in  $\mathbf{k}$  and  $\ell$ . Finally, both  $\mathbf{D}$  and  $\mathbf{Q}_t$  have dimension  $[(M_1 + 1) \times \dots \times (M_3 + 1)]^2$ . Moreover,  $\mathbf{a}_t$  represents  $b_{\mathbf{k}\ell}(t+1)$  and  $\mathbf{a}_{t-1}$  provides the recursive information on  $b_{\mathbf{k}\ell}(t)$  through the summation of  $b_{\ell i}(t)$  terms, what should not be confused with  $b_{\mathbf{k}\ell}(t)$  itself. Both  $\mathbf{a}_t$  and  $\mathbf{a}_{t-1}$  are vectors with dimension  $(M_1 + 1) \times \dots \times (M_3 + 1)$ . In resume, the computation of the log-likelihood function is outlined in the following steps (see algorithm below). Firstly, the recursive scheme is initialized with a  $\lambda$  estimate, either obtained from MM if  $\hat{\lambda}_i > 0$ ,  $i = 1, 2, 3$ , or simply the non-informative  $\lambda$  estimate  $\hat{\mathbf{y}}$ . Secondly, the  $\mathbf{a}_t$  computation applies a scaling scheme to avoid numerical underflow [58]. Thirdly, the log-likelihood function is evaluated and subsequently maximized via L-BFGS-B box-constrained optimization, where the solution is restricted to be admissible.

1. Initialization:
  - (a) Compute initial estimate  $\hat{\theta}_0$ .
  - (b) Initial evaluation:  $\mathbf{a}_1 = P(\varepsilon_1 = \ell | \varepsilon_1 \leq \mathbf{Y}_1)$ ,  $w_1 = \mathbf{1}^\top \mathbf{a}_1$ ,  $\phi_1 = \mathbf{a}_1 / w_1$ .
2. For  $t = 2, \dots, T$ :
  - (a) Computation:  $\mathbf{u}_t = \mathbf{D}\mathbf{Q}_t\phi_{t-1}$ .
  - (b) Scaling:  $\frac{w_t}{w_{t-1}} = \mathbf{1}^\top \mathbf{u}_t$ ,  $\phi_t = \mathbf{u}_t / \left(\frac{w_t}{w_{t-1}}\right)$ .
3. Log-likelihood evaluation:  $\ell(\theta) = \ln w_T = \ln w_1 + \sum_{t=2}^T \ln \frac{w_t}{w_{t-1}}$ .

Given the large similarities with the CML algorithm used in INARMA(1,1) identification [55] it is crucial to stress that the above presented algorithm has

a multivariate dimension  $S > 1$ , for which the spatial component is embedded into the  $\mathbf{D}$  and  $\mathbf{Q}_t$  matrices as well as the  $\mathbf{a}_t$  vector, through the combinations in  $\mathbf{k}$  and  $\ell$ . Not surprisingly, the multivariate algorithm is computationally more demanding than the univariate counterpart, with its complexity increasing not only with higher  $S$  but also with higher  $M_i$  values, for  $i = 1, \dots, S$ .

#### 4.2.4. Statistical inference

Inference based on one sample requires the evaluation of a standard error. This is available for the CML from the variance-covariance matrix of  $\hat{\boldsymbol{\theta}}$  estimated from the numerical Hessian, even if reported that the numerical Hessian may be unreliable when CML estimates are too close to the boundaries of the admissible region [58]. Parametric bootstrap is one strategy to deal with this shortcoming [16], with the additional advantage of allowing the comparison between different estimation approaches. This strategy evaluates the standard errors as the standard deviation of the estimates obtained for a bootstrap sample, composed of  $B$  paths (of length  $T$ ) simulated from a Poisson STINMA(1<sub>1</sub>) data-generating process (DGP). Given a trivariate time series of length  $T$ , the parametric bootstrap is resumed in the following steps.

1. Compute  $\hat{\boldsymbol{\theta}}$  from the time series.
2. Set the Poisson STINMA(1<sub>1</sub>) with parameters  $\hat{\boldsymbol{\theta}}$  as the DGP.
3. For  $b = 1, \dots, B = 1000$ :
  - (a) Simulate the  $b^{\text{th}}$  bootstrap path from the DGP.
  - (b) Compute  $\hat{\boldsymbol{\theta}}_b^*$  from the  $b^{\text{th}}$  bootstrap path.

The set of bootstrap estimates  $\hat{\boldsymbol{\theta}}_b^*$ ,  $b = 1, \dots, B$ , for which the estimation method converges is then used to carry out inference. Confidence intervals (CI) based on the normal distribution can be computed if the bootstrap distribution of  $\hat{\boldsymbol{\theta}}^*$  is roughly normal (i.e. the asymptotic distribution as  $T \rightarrow \infty$ ). In this case, the standard errors are evaluated from the square root of  $\text{diag}(\mathbf{G}_{\hat{\boldsymbol{\theta}}})$ , where  $\mathbf{G}_{\hat{\boldsymbol{\theta}}}$  is the (empirical) variance-covariance matrix obtained from the bootstrap estimates,

$$\mathbf{G}_{\hat{\boldsymbol{\theta}}} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\boldsymbol{\theta}}_b^* - \bar{\boldsymbol{\theta}}^*)^\top (\hat{\boldsymbol{\theta}}_b^* - \bar{\boldsymbol{\theta}}^*) \quad \text{with} \quad \bar{\boldsymbol{\theta}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\boldsymbol{\theta}}_b^*. \quad (44)$$

For smaller values of  $T$ , confidence intervals based on percentiles (percentile intervals, PI) may be preferable as the normal approximation might be poor [16]. Here, both CIs and PIs were evaluated at 95% confidence.

## 5. Simulation study

The MM, CLS, and CML approaches were implemented in order to estimate the STINMA(1<sub>1</sub>) parameters given one instance of  $S = 3$  simultaneous time

series of length  $T$ . Briefly, six estimation procedures were compared (2 MM, 3 CLS and 1 CML). The MM approach considered two competing strategies based on either  $(\bar{Y}, \hat{\rho}_{00}, \hat{\rho}_{10})$  or  $(\bar{Y}, \hat{\rho}_{11}, \hat{\rho}_{10})$  denoted as “MM  $\hat{\rho}_{00}$ ” and “MM  $\hat{\rho}_{11}$ ”, respectively. The three CLS alternatives are based on  $(\hat{\varepsilon}, \bar{Y}_i, \bar{Y}_j)$  with combinations of  $i, j \in \{1, 2, 3\}$  and  $i \neq j$ , and denoted as “CLS  $\bar{Y}_i\bar{Y}_j$ ”. The simulation study uses 1000 replicates of the chosen DGP with  $S = 3$  and different lengths  $T \in \{50, 100, 250, 500, 1000\}$ . Motivated by the real data application in Section 6, the data used in the simulation study was generated from the Poisson STINMA(1<sub>1</sub>) process in (27) with parameters  $\theta = (\lambda^\top, \beta_{10}, \beta_{11})$  where  $\lambda = (2, 3, 4)^\top$ ,  $\beta_{10} = 0.3$  and  $\beta_{11} = 0.5$ .

Figure 6 presents the distributions of  $\hat{\theta}$  for the comparison between the different estimation strategies concerning bias and variability. For  $T = 1000$ , the CML approach exhibits the lowest sample bias and smallest variability among the estimates. Both CML and MM distributions are fairly symmetric around the true value of each parameter. Among the MM strategies, the variability of the estimates is smaller for “MM  $\hat{\rho}_{00}$ ” than for “MM  $\hat{\rho}_{11}$ ”. Contrarily to the MM and CML approaches, CLS shows a visible sample bias where, typically,  $\lambda$  and  $\beta_{10}$  are overestimated whereas  $\beta_{11}$  is underestimated. The three CLS approaches exhibit similar bias and variability for  $\lambda$ , which was expected as the corresponding estimator is the same. However, the approaches differ with respect to the estimation of  $\beta_{10}$  and  $\beta_{11}$ . The “CLS  $\bar{Y}_1\bar{Y}_2$ ” shows lower bias and higher variability than the “CLS  $\bar{Y}_1\bar{Y}_3$ ”, while “CLS  $\bar{Y}_2\bar{Y}_3$ ” is the CLS approach with largest bias and variability, also providing  $\beta_{10}$  and  $\beta_{11}$  estimates  $\notin [0, 1)$  in 49 out of 1000 paths.

Figure 7 displays the MM and CML comparison for  $T \leq 1000$ . As expected, the variability in the estimation increases as  $T$  decreases with the CML exhibit-

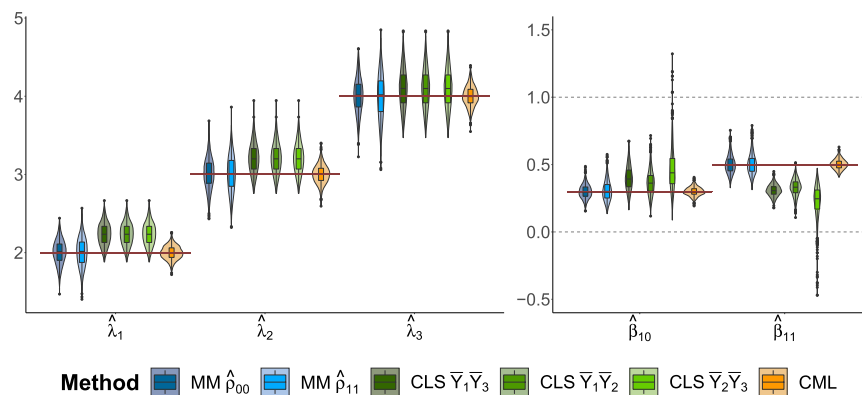


FIG 6. Violin plots of the estimates for the Poisson STINMA(1<sub>1</sub>) parameters  $\theta = (2, 3, 4, 0.3, 0.5)$ , considering MM, CLS and CML strategies (1000 paths with  $T = 1000$ ). The  $\theta$  values are displayed in the red lines and the  $[0, 1)$  interval is delimited with the black dotted lines. Violin plots use a Gaussian kernel with one unit of bandwidth. Boxplots are presented in the background.

ing the lowest variability across the different sample sizes. Also, the MM procedure fails to provide an estimate in 49/18/2 paths of length  $T = 50/100/250$  (e.g. due to non-positive values of the space-time autocovariance function  $\gamma_{rk}(h)$  or singularity of the Jacobian matrix) in addition to 100/39/2 paths, where the estimates were outside the admissible interval.

Furthermore, the MM bias increases considerably when  $T$  decreases while CML distributions remain fairly symmetric even for small sample sizes. Having in mind that, in general, the CML procedure is computationally more demanding than the MM and requires an initial guess to trigger the algorithm, the MM is a fair estimation strategy for large  $T$ . For lower values of  $T$ , the CML approach clearly exhibits lower sample bias and lower variability thus making it the preferable estimation strategy. The results from this study suggest that the combined use of CML with the initial estimate provided by MM is an adequate choice for all real-data applications as it reduced the computation time by approx. 5% when compared to the CML combined with the non-informative initial guess  $\hat{\theta}_0 = (\bar{y}, 0.5, 0.5)$ .

The CLS approach was also evaluated in a version of the procedure with known innovations, accomplished by inserting the innovations used in the simulation of the DGP paths into the estimation procedure, see Figure 7 for the resulting distributions. In comparison to the CLS with estimated innovations (Figure 6,  $T = 1000$ ), it is clear that the CLS approach with known innovations has better performance. In particular, the bias and variance are notably decreased at levels lower than those obtained for the MM approach. Hence, the

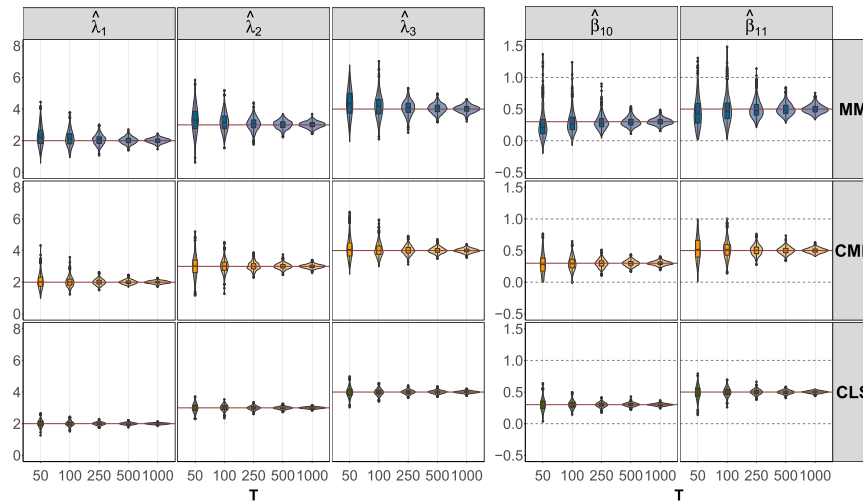


FIG 7. Violin plots of the estimates for the Poisson STINMA(1<sub>1</sub>) parameters  $\theta = (2, 3, 4, 0.3, 0.5)$ , considering the “MM  $\hat{\rho}_{00}$ ” and the CML estimation strategies as well as the “CLS  $\bar{Y}_1\bar{Y}_3$ ” with known innovations (1000 paths and several values for  $T$ ). Similar representation as in Figure 6.

CLS is a suitable estimation approach in many applications when data on the unobserved process is available, e.g. series of counts with the number of accesses to a web server within a time interval, given the total number of the different IP and the number of new IP addresses, the later constituting the time series of observed innovations [53].

In order to evaluate the full potential of the CML approach regarding inference in the Poisson STINMA(1<sub>1</sub>), the standard error (SE) evaluated via the Hessian matrix, for a single path, was compared to the SE obtained from the 1000 paths simulated via parametric bootstrap. As expected, Figure 8 shows that the median Hessian-SE decreases as  $T$  increases, in a fairly similar trend to that observed for the bootstrap-SE obtained for CML and MM from the 1000 paths (represented in orange and blue, respectively). Moreover, the interquartile range of the Hessian-SE includes the CML bootstrap-SE (orange line), supporting that the CML inference via Hessian and bootstrap are concordant for most of the Poisson STINMA(1<sub>1</sub>) paths and even for small values of  $T$ . Finally, Figure 8 also points out that the Hessian-SE is lower than the MM bootstrap-SE (blue line) for more than 95% of the paths, for each sample size  $T$ . The better properties of the CML estimation approach, however, come at a computational cost as one run of the CML algorithm (efficiently implemented via parallelisation and C++ coding) with a non-informative initial estimate is yet more demanding than 1001 runs (one to obtain the point estimate plus 1000 to carry out the statistical inference) of the MM algorithm. For instance, for  $T = 1000$ , 100 runs of MM take about  $0.002 \pm 0.001$  seconds, while the iterative parallelized procedure for CML optimization needs  $11.973 \pm 4.372$  minutes even for  $T = 50$ . Thus, when considering larger spatial systems, it is reasonable to consider MM estimation. For instance, if we take  $S = 18$  (the number of Portugal mainland districts) and  $S = 278$  (the number of Portuguese municipalities), then MM is a good choice to provide point estimates of a STINMA(1<sub>1</sub>) taking about  $0.058 \pm 0.02$  seconds ( $S = 18$ ) and  $3.76 \pm 0.06$  minutes ( $S = 278$ ) in a common personal computer, for 100 simulated paths of length  $T = 1000$ , where all locations are considered first-order neighbours. This suggests that MM is a good choice even for large systems since point estimates are obtained in a reasonable amount of time. For inference, parallelization procedures are used to speed-up computa-

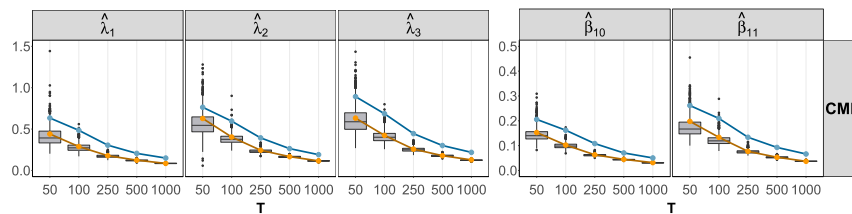


FIG 8. Boxplots of the standard error (SE) for the Poisson STINMA(1<sub>1</sub>) with parameters  $\theta = (2, 3, 4, 0.4, 0.5)$  computed for CML via the Hessian matrix. The orange and blue dots/lines show, respectively, the SE of the CML and the MM estimates for the 1000 paths.



tion. Note that in these larger systems, all locations are considered first-order (spatial) neighbors [45], for ease of establishing the spatial structure, since this task becomes increasingly challenging as  $S$  increases.

## 6. Real data application

The Poisson STINMA( $1_1$ ) model is finally applied to real data, corresponding to the daily number of hospital admissions in three Portuguese locations (Figure 9). It is important to stress that the Poisson STINMA( $1_1$ ) modelling is used merely for illustrative purposes and, as such, it is not expected that this process is the true data-generating mechanism of the time series. Nevertheless, it will be demonstrated that there is a significant spatial and temporal dependence for this experimental setting. For each location, the integer-valued time series was constructed by counting the daily episodes of respiratory system diseases' (ICD-9 codes 460–519 and ICD-10 codes J00–J99) from patients with residence within a 20 km radius of the centre of the location. The spatial locations refer to existing air quality monitoring stations, to keep the hospital anonymity and to set the data ready for a future analysis of the counts in the light of environmental correlates. We refer to [27] for further details on the data.

Figure 9 displays the map of Portugal highlighting the Algarve region and the three geographical locations, linearly placed next to each other. The count time series at each location are shown for the period between 17 February 2012 and 26 May 2012 ( $T = 100$ ). The data clearly exhibit low counts ranging within the  $[0,14]$  interval, and the three time series have the same median of 4 hospital admissions per day. The empirical space-time autocorrelations at lag 1 are  $\hat{\rho}_{00} = 0.200$  and  $\hat{\rho}_{10} = 0.203$ , respectively, which suggest that the Poisson STINMA( $1_1$ ) is a potential good candidate for modeling purposes, since  $(\hat{\rho}_{00}, \hat{\rho}_{10})$  fall within the admissible region in Figure 5.

Table 3 presents the MM and CML estimates for the parameters of the Poisson STINMA( $1_1$ ) process fitted to the experimental data. Both MM and CML approaches indicate that all parameters are statistically significant at 95% confidence level. For a given parameter, the CIs are wider for MM than for CML and overlap. Regarding the point estimates,  $\hat{\lambda}_1 \approx \hat{\lambda}_3 < \hat{\lambda}_2$  for both approaches. Finally,  $\hat{\lambda}$  are smaller for the MM approach, which is compensated by larger estimates for the temporal parameter  $\beta_{10}$  and the spatial parameter  $\beta_{11}$ .

The performance of the STINMA( $1_1$ ) model was compared to that of the MINMA(1) model, which is obtained from (9) by setting  $m_1 = 0$ . This implies that the STINMA( $1_1$ ) process extends the MINMA(1) by additionally including  $\beta_{11}$ , i.e. the spatio-temporal parameter that reflects the dependence between the process in a given location and the past history of its neighbor locations. As expected, the MINMA thinning parameter was found to be significant at a 5% level, with estimates  $\hat{\beta}_{10} = 0.250 [0.086; 0.413]$  (for MM) and  $\hat{\beta}_{10} = 0.194 [0.192; 0.196]$  (for CML). The STINMA and MINMA modelling approaches were compared through the maximum log-likelihood values and likelihood ratio test (LRT), to assess whether the MINMA model is preferable against

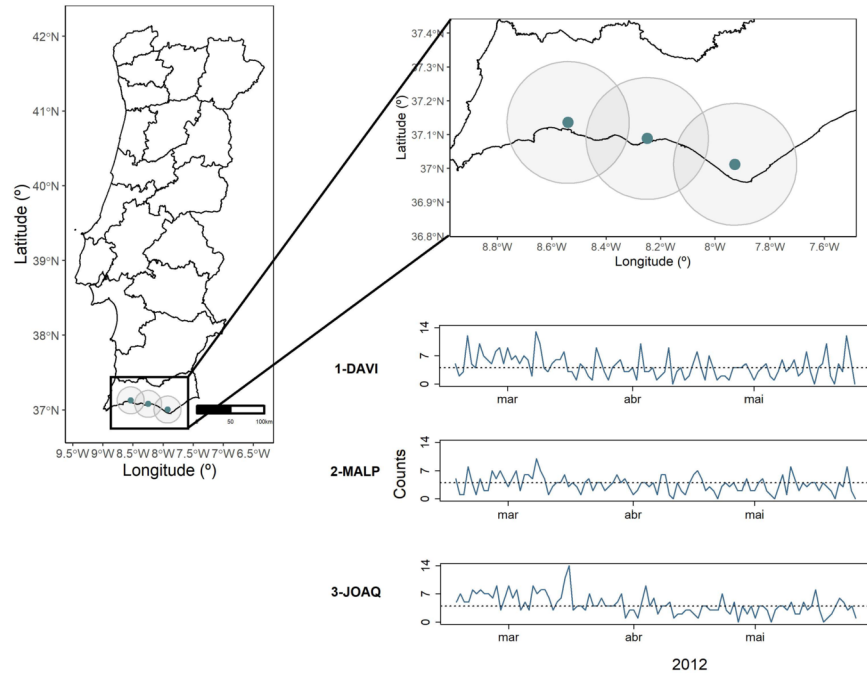


FIG 9. Portugal map highlighting the Algarve region. The data consists of time series with the daily number of hospital admissions from the residents within the 20 km radius around three locations (DAVI, MALP, JOAQ). The medians are shown with the black dotted lines. Further details of the experimental setting in [27].

the fully specified STINMA, as well as via goodness-of-fit criterion (i.e. AIC, BIC and AICc) to assess the balance between performance and complexity of the model. The MINMA exhibits a smaller log-likelihood value than that of the STINMA model ( $-678.13$  vs.  $-675.07$ ) thus favouring the STINMA approach. The LRT returned a p-value = 0.013, which lead to the conclusion that the hypothesis of the MINMA being preferable against the STINMA model is rejected (at 5% level). Additionally, the goodness-of-fit measures consistently favoured the STINMA model (AIC: 1360.14 vs. 1364.27; AICc: 1360.78 vs. 1364.69; BIC: 1373.17 vs. 1374.69). Furthermore, in the presence of spatial data, the STINMA emerges as the natural modelling approach because it incorporates a parameter that facilitates spatial interpretability.

### 7. Conclusions and future research

This work extensively analyzed the characteristics of the STINMA model, which constitutes a special case of the new STINARMA class of models. In comparison to the continuous STMA model, the second-order moments of the general STINMA( $q_{m_1, \dots, m_q}$ ) exhibit extra additive terms that steam from independent

TABLE 3

Estimate, standard error (SE), 95% confidence and percentile intervals (CI and PI) for MM and CML. SE is computed via parametric bootstrap for MM and via Hessian matrix for CML.

Parameter	MM				CML		
	Estimate	SE	95% CI	95% PI	Estimate	SE	95% CI
$\lambda_1$	2.680	0.389	[1.856, 3.504]	[2.103, 3.558]	3.439	0.292	[2.867, 4.011]
$\lambda_2$	1.730	0.396	[0.890, 2.570]	[1.140, 2.656]	2.603	0.294	[2.027, 3.179]
$\lambda_3$	2.673	0.390	[1.838, 3.508]	[2.075, 3.584]	3.415	0.295	[2.837, 3.993]
$\beta_{10}$	0.351	0.136	[0.070, 0.631]	[0.116, 0.635]	0.164	0.067	[0.033, 0.295]
$\beta_{11}$	0.468	0.172	[0.120, 0.816]	[0.127, 0.801]	0.159	0.076	[0.010, 0.308]

BTOs executed at lag zero. The Poisson STINMA(1<sub>1</sub>) model was explored for  $S = 3$  locations (with results easily generalized for  $S > 3$ ), and its second-order moments coincide with those of the STMA(1<sub>1</sub>).

The case considering non-independent spatial BTOs was further addressed. This situation leads to off-diagonal variance parcels in the Poisson STINMA(1<sub>1</sub>) autocovariance and modifications on the space-time autocovariance. Strategies for parameter estimation (MM, CLS and CML) were derived for the Poisson STINMA(1<sub>1</sub>) model under the non-independent BTO paradigm. The simulation study shows that the CML achieves the optimal performance (lowest bias and variance) among all strategies, even for small sample sizes. For larger sizes, MM also presents reduced bias and variability but, for small sizes, fails to converge or returns non-admissible estimates for some cases. CML also has the advantage of allowing inference through the Hessian matrix, unlike MM, which requires e.g. a bootstrap procedure for inference purposes (which implies several runs of the MM procedure). Balancing all pros and cons, the CML strategy has preferable properties (bias and variance) at the expense of a higher computational burden, while MM constitutes a faster and fair estimation strategy for large samples. Finally, in real applications with known innovations, CLS is also a good option for the estimation of the Poisson STINMA(1<sub>1</sub>) parameters.

There are several topics of future research that can bring more flexibility to the STINARMA class of models. Firstly, more flexibility could be added to the model by allowing that the temporal and spatial parameters to vary per location, similarly to what has been done for continuous models [6]. Secondly, one may consider an innovations' distribution allowing for overdispersion, e.g. negative-binomial. A third relevant topic is the inclusion of time-dependent covariates either through time-dependent innovations or probability parameters, introduced via an appropriate link function. Additional cross-dependence can be added to the model e.g. through the use of a non-independent innovations process or by using a copula-based approach. Lastly, non-stationary modelling approaches are also an interesting topic for future research since the stationarity assumption may be too restrictive in some applied settings.

**Appendix A: Expected value and variance of the general STINMA process**

The proofs for the component-wise expected value and variance of the general STINMA( $q_{m_1, \dots, m_q}$ ) process are set by rewriting the system of equations (10) such that, for a given location  $s$ , the model equation can be written as

$$\begin{aligned}
 Y_{s,t} &= \sum_{j=1}^q \left[ \sum_{\ell=0}^{m_j} \sum_{n=1}^S (\beta_{j\ell} w_{sn}^{(\ell)}) \circ \varepsilon_{n,t-j} \right] + \varepsilon_{s,t} \\
 &= \sum_{j=1}^q \left[ \sum_{n=1}^S \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \circ \varepsilon_{n,t-j} \right] + \varepsilon_{s,t}. \tag{45}
 \end{aligned}$$

*Proof. (Expected Value – Component-wise)* By applying the linearity of the expected value and the well-known property  $E[\psi \circ X] = E[\psi X]$  to equation (45), the expected value is obtained as follows

$$\begin{aligned}
 E[Y_{s,t}] &= E \left[ \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S (\beta_{j\ell} w_{sn}^{(\ell)}) \circ \varepsilon_{n,t-j} + \varepsilon_{s,t} \right] \\
 &= \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S E \left[ \beta_{j\ell} w_{sn}^{(\ell)} \circ \varepsilon_{n,t-j} \right] + E[\varepsilon_{s,t}] \\
 &= \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S E \left[ \beta_{j\ell} w_{sn}^{(\ell)} \varepsilon_{n,t-j} \right] + E[\varepsilon_{s,t}] \\
 &= \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S \left( \beta_{j\ell} w_{sn}^{(\ell)} \mu_{\varepsilon_n} \right) + \mu_{\varepsilon_s}. \tag{46}
 \end{aligned}$$

□

*Proof. (Variance – Component-wise)* The component-wise variance for a given location  $s$  is based on the result  $V[(\psi \circ X)] = V[\psi X] + E[\psi(1 - \psi)X]$  and the independence between innovations.

$$\begin{aligned}
 V[Y_{s,t}] &= V \left[ \sum_{j=1}^q \left[ \sum_{n=1}^S \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \circ \varepsilon_{n,t-j} \right] + \varepsilon_{s,t} \right] \\
 &\stackrel{\text{ind.}}{=} \sum_{j=1}^q V \left[ \sum_{n=1}^S \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \circ \varepsilon_{n,t-j} \right] + V[\varepsilon_{s,t}] \\
 &= (\text{independence in time}) \\
 &\stackrel{\text{ind.}}{=} \sum_{j=1}^q \sum_{n=1}^S V \left[ \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \circ \varepsilon_{n,t-j} \right] + V[\varepsilon_{s,t}] \\
 &= (\text{independence in space})
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=1}^q \sum_{n=1}^S \left( V \left[ \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \varepsilon_{n,t-j} \right] + E \left[ \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \times \right. \right. \\
 &\quad \left. \left. \times \left( 1 - \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \varepsilon_{n,t-j} \right] \right) + V[\varepsilon_{s,t}] \\
 &= \sum_{j=1}^q \sum_{n=1}^S \left[ \left( \sum_{\ell=0}^{m_j} \beta_{j\ell}^2 (w_{sn}^{(\ell)})^2 + \sum_{\substack{\ell=0 \\ k \neq \ell}}^{m_j} \sum_{k=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \beta_{jk} w_{sn}^{(k)} \right) V[\varepsilon_{n,t-j}] + \right. \\
 &\quad \left. + \left( \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) \left( 1 - \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} \right) E[\varepsilon_{n,t-j}] \right] + V[\varepsilon_{s,t}].
 \end{aligned}$$

However, taking into account that  $\mathbf{W}^{(0)} = \mathbf{I}_S$ ,  $\text{diag}(\mathbf{W}^{(\ell)}) = 0, \forall \ell \geq 1$  and the fact that for a given  $s, n$  at most one of  $\{w_{sn}^{(0)}, w_{sn}^{(1)}, \dots\}$  is non-zero, one can write more simply,

$$\begin{aligned}
 V[Y_{s,t}] &= \sum_{j=1}^q \sum_{n=1}^S \left( \sum_{\ell=0}^{m_j} \beta_{j\ell}^2 (w_{sn}^{(\ell)})^2 \sigma_{\varepsilon_n}^2 + \sum_{\ell=0}^{m_j} \beta_{j\ell} w_{sn}^{(\ell)} (1 - \beta_{j\ell} w_{sn}^{(\ell)}) \mu_{\varepsilon_n} \right) + \sigma_{\varepsilon_s}^2 \\
 &= \sum_{j=1}^q \sum_{\ell=0}^{m_j} \sum_{n=1}^S \left( \beta_{j\ell}^2 (w_{sn}^{(\ell)})^2 \sigma_{\varepsilon_n}^2 + \beta_{j\ell} w_{sn}^{(\ell)} (1 - \beta_{j\ell} w_{sn}^{(\ell)}) \mu_{\varepsilon_n} \right) + \sigma_{\varepsilon_s}^2. \quad (47)
 \end{aligned}$$

□

**Appendix B: Temporal covariance of the general STINMA process**

Theorem 3.1 specifies the temporal covariance function  $\Gamma^*(h)$  for the general STINMA( $q_{m_1, \dots, m_q}$ ) process. The proof is constructed by steps. First,  $\Gamma^*(0)$  is obtained for the STINMA( $1_{m_1}$ ) and extended to the STINMA( $q_{m_1, \dots, m_q}$ ). Then,  $\Gamma^*(h)$  is obtained by induction.

*Proof.* Let  $h = 0$ . For simplicity of the derivation of results we start by proving the result of equation (16) for a STINMA( $1_{m_1}$ )

$$\mathbf{Y}_t = \sum_{\ell=0}^{m_1} \beta_{1\ell} \mathbf{W}^{(\ell)} \circ \varepsilon_{t-1} + \varepsilon_t, \quad (48)$$

which can be rewritten using the “trick” (7),

$$\mathbf{Y}_t = \underbrace{\sum_{\ell=0}^{m_1} \beta_{1\ell} \mathbf{W}^{(\ell)}}_{=: \mathbf{B}_1} \circ \varepsilon_{t-1} + \varepsilon_t. \quad (49)$$

Using the properties of the covariance operator,

$$\Gamma^*(0) = \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_t^\top) = \text{Cov}(\mathbf{B}_1 \circ \varepsilon_{t-1} + \varepsilon_t, (\mathbf{B}_1 \circ \varepsilon_{t-1} + \varepsilon_t)^\top)$$

$$\begin{aligned}
 &= Cov\left(\mathbf{B}_1 \circ \boldsymbol{\varepsilon}_{t-1}, (\mathbf{B}_1 \circ \boldsymbol{\varepsilon}_{t-1})^\top\right) + \\
 &\quad + \underbrace{Cov\left(\mathbf{B}_1 \circ \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_t^\top\right)}_{=\mathbf{O}_S} + \underbrace{Cov\left(\boldsymbol{\varepsilon}_t, (\mathbf{B}_1 \circ \boldsymbol{\varepsilon}_{t-1})^\top\right)}_{=\mathbf{O}_S} + \underbrace{Cov(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t^\top)}_{=\mathbf{G} \text{ eq. (5)}}.
 \end{aligned}$$

The computation of the covariance of the first parcel, requires the property  $E[(\mathbf{A} \circ \mathbf{X})(\mathbf{A} \circ \mathbf{X})^\top] = \mathbf{A}E[\mathbf{X}\mathbf{X}^\top]\mathbf{A}^\top + \text{diag}(CE[\mathbf{X}])$  where  $\text{diag}(CE[\mathbf{X}])$  is a diagonal matrix and  $\mathbf{C}$  is the squared matrix ( $S \times S$ ) with entries  $c_{ij} = a_{ij}(1 - a_{ij})$ ,  $i, j = 1, \dots, S$  [17]. This implies, in terms of covariance, that

$$Cov\left(\mathbf{A} \circ \mathbf{X}, (\mathbf{A} \circ \mathbf{X})^\top\right) = \mathbf{A}Cov(\mathbf{X}, \mathbf{X}^\top)\mathbf{A}^\top + \text{diag}(CE[\mathbf{X}]). \quad (50)$$

Therefore, from (50) the general expression can be written as

$$\boldsymbol{\Gamma}^*(0) = \mathbf{G}(\mathbf{I}_S + \mathbf{B}_1\mathbf{B}_1^\top) + \sum_{\ell=0}^{m_1} \text{diag}(\mathbf{C}^{(\ell)}E[\boldsymbol{\varepsilon}_t]). \quad (51)$$

Note that  $\mathbf{C}^{(\ell)}$  has entries  $c_{ij}^{(\ell)} := a_{ij}^{(\ell)}(1 - a_{ij}^{(\ell)})$ , for  $\ell = 1, \dots, m_1$ . Now, for a general STINMA( $q, m_1, \dots, m_q$ ) it is straightforward to obtain the variance-covariance matrix, since  $\boldsymbol{\Gamma}^*(0)$  will be of the same form as equation (51) with an additional summation up to the order  $q$  of the process, i.e.

$$\boldsymbol{\Gamma}^*(0) = \sum_{j=1}^q \left( \mathbf{G}(\mathbf{I}_S + \mathbf{B}_j\mathbf{B}_j^\top) + \sum_{\ell=0}^{m_j} \text{diag}(\mathbf{C}^{(\ell)}E[\boldsymbol{\varepsilon}_t]) \right). \quad (52)$$

Now, let  $h = 1$ . Then,

$$\begin{aligned}
 \boldsymbol{\Gamma}^*(1) &= Cov(\mathbf{Y}_t, \mathbf{Y}_{t+1}^\top) = Cov\left(\sum_{j=1}^q \mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \sum_{j=1}^q (\mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j+1} + \boldsymbol{\varepsilon}_{t+1})^\top\right) \\
 &= Cov\left(\sum_{j=1}^q \mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j} + \boldsymbol{\varepsilon}_t, \sum_{i=0}^{q-1} (\mathbf{B}_i \circ \boldsymbol{\varepsilon}_{t-i+1} + \boldsymbol{\varepsilon}_{t+1})^\top\right) \\
 &= \sum_{j=1}^q \sum_{i=0}^{q-1} \underbrace{Cov\left(\mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j}, (\mathbf{B}_i \circ \boldsymbol{\varepsilon}_{t-i+1})^\top\right)}_{\neq 0, \text{ when } i=j+1} + \sum_{j=1}^q \underbrace{Cov\left(\mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j}, \boldsymbol{\varepsilon}_{t+1}^\top\right)}_{\mathbf{O}_S} \\
 &\quad + \sum_{i=0}^{q-1} \underbrace{Cov\left(\boldsymbol{\varepsilon}_t, (\mathbf{B}_i \circ \boldsymbol{\varepsilon}_{t-i+1})^\top\right)}_{\neq 0, \text{ when } i=1} + \underbrace{Cov(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+1}^\top)}_{\mathbf{O}_S} \\
 &= \sum_{j=1}^{q-1} Cov\left(\mathbf{B}_j \circ \boldsymbol{\varepsilon}_{t-j}, (\mathbf{B}_{j+1} \circ \boldsymbol{\varepsilon}_{t-j})^\top\right) + Cov\left(\boldsymbol{\varepsilon}_t, (\mathbf{B}_1 \circ \boldsymbol{\varepsilon}_t)^\top\right) \\
 &= \sum_{j=1}^{q-1} \mathbf{B}_j \mathbf{G} \mathbf{B}_{j+1}^\top + \mathbf{G} \mathbf{B}_1^\top. \quad (53)
 \end{aligned}$$

By induction, for  $0 < h < q + 1$ , it can be shown that

$$\mathbf{\Gamma}^*(h) = \sum_{j=1}^{q-h} \mathbf{B}_j \mathbf{G} \mathbf{B}_{j+h}^\top + \mathbf{G} \mathbf{B}_h^\top. \quad (54)$$

For  $h = q$ , it is straightforward that  $\mathbf{\Gamma}(q) = \mathbf{G} \mathbf{B}_q^\top$ . Finally, by definition  $\mathbf{B}_h = 0$ ,  $h \geq q + 1$  and  $\mathbf{\Gamma}^*(h) = \mathbf{O}_S$ ,  $h \geq q + 1$ .  $\square$

### Appendix C: Joint distribution of the Poisson STINMA(1<sub>1</sub>) process

Theorem 4.1 specifies the joint distribution of the Poisson STINMA(1<sub>1</sub>) process under the independent thinnings assumption. The proof is based on the derivation of the p.g.f. of the Poisson STINMA(1<sub>1</sub>) process  $\mathbf{Y}_t = \beta_{10} \circ \boldsymbol{\varepsilon}_{t-1} + \beta_{11} \mathbf{W}^{(1)} \circ \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t$ , with

$$\mathbf{W}^{(1)} = \begin{bmatrix} 0 & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & 0 & w_{23}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & 0 \end{bmatrix}. \quad (55)$$

*Proof.* For the sake of simplicity we define  $\mathbf{B} = \beta_{10} \mathbf{I}_3 + \beta_{11} \mathbf{W}^{(1)}$  with entries  $b_{ij}$ ,  $i, j = 1, 2, 3$ , and rewrite the STINMA model as  $\mathbf{Y}_t = \mathbf{B} \circ \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t$ . By definition, the p.g.f. of  $\mathbf{Y}_t$  is

$$G_{\mathbf{Y}}(s_1, s_2, s_3) = E \left[ s_1^{Y_{1,t}} s_2^{Y_{2,t}} s_3^{Y_{3,t}} \right] = E \left[ \prod_{r=1}^3 s_r^{\varepsilon_{r,t} + \sum_{j=1}^3 b_{rj} \circ \varepsilon_{r,t-1}} \right]$$

and by independence in time it follows that

$$= \prod_{r=1}^3 E \left[ s_r^{\varepsilon_{r,t}} \right] \times \prod_{r=1}^3 E \left[ \prod_{j=1}^3 s_j^{b_{rj} \circ \varepsilon_{r,t-1}} \right], \quad (56)$$

where  $E \left[ s_r^{\varepsilon_{r,t}} \right] = G_{\varepsilon_r}(s_r)$  is the p.g.f. of the univariate innovation process  $\varepsilon_{r,t}$ . Each mixed expected value takes the generic form  $E \left[ s_1^{\alpha_1 \circ X} s_2^{\alpha_2 \circ X} s_3^{\alpha_3 \circ X} \right]$  with  $\alpha_i \circ X = \sum_{j=1}^X u_{ij}$  and  $u_{ij} \sim B(1, \alpha_i)$  being independent r.v.'s. Taking into account that the thinning operations are independent,

$$\begin{aligned} E \left[ s_1^{\alpha_1 \circ X} s_2^{\alpha_2 \circ X} s_3^{\alpha_3 \circ X} \right] &= E \left[ E \left[ s_1^{\alpha_1 \circ X} s_2^{\alpha_2 \circ X} s_3^{\alpha_3 \circ X} \mid X \right] \right] \\ &\stackrel{\text{ind.}}{=} E \left[ E \left[ s_1^{\alpha_1 \circ X} \mid X \right] E \left[ s_2^{\alpha_2 \circ X} \mid X \right] E \left[ s_3^{\alpha_3 \circ X} \mid X \right] \right], \quad (57) \end{aligned}$$

and, by replacing  $E \left[ s_i^{\alpha_i \circ X} \mid X \right] = (1 - \alpha_i + \alpha_i s_i)^X \forall i = 1, 2, 3$  results in

$$E \left[ \prod_{i=1}^3 (1 - \alpha_i + \alpha_i s_i)^X \right] = G_X \left( \prod_{i=1}^3 (1 - \alpha_i + \alpha_i s_i) \right). \quad (58)$$

Furthermore,

$$\prod_{i=1}^3 (1 - \alpha_i + \alpha_i s_i) = 1 + \sum_{i=1}^3 \left( 1 - \sum_{\substack{j=1 \\ i \neq j}}^3 \alpha_j + \sum_{\substack{k=1 \\ k \neq i \neq j}}^3 \alpha_j \alpha_k \right) \alpha_i (s_i - 1) + \sum_{\substack{i=1 \\ i < j}}^3 \sum_{j=1}^3 \left( \alpha_j - \sum_{\substack{k=1 \\ k \neq i \neq j}}^3 \alpha_j \alpha_k \right) \alpha_i (s_i s_j - 1) + \alpha_1 \alpha_2 \alpha_3 (s_1 s_2 s_3 - 1). \tag{59}$$

To construct the joint p.g.f. of the STINMA(1<sub>1</sub>) process with  $S = 3$ , consider  $\prod_{i=1}^3 (1 - \alpha_i + \alpha_i s_i)$  as a function  $\eta(\alpha_1, \alpha_2, \alpha_3)$ . Using (59),

$$G_{\mathbf{Y}}(s_1, s_2, s_3) = \prod_{r=1}^3 G_{\varepsilon_r}(s_r) \times \prod_{j=1}^3 G_{\varepsilon_r}(\eta(b_{1r}, b_{2r}, b_{3r})),$$

which constitutes the product of six p.g.f. functions. Since  $\varepsilon_{r,t}$  are marginally Poisson distributed with parameter  $\lambda_r$ , we get

$$G_{\mathbf{Y}}(s_1, s_2, s_3) = \exp \left\{ \sum_{r=1}^3 \left[ \lambda_r + \left( \sum_{j=1}^3 \lambda_j \left( 1 - \sum_{\substack{i=1 \\ i \neq r}}^3 b_{ij} + \prod_{\substack{k=1 \\ k \neq r}}^3 b_{kj} \right) b_{rj} \right) (s_r - 1) \right] + \sum_{r=1}^3 \sum_{\substack{k=1 \\ k > r}}^3 \left[ \left( \sum_{j=1}^3 \lambda_j \left( b_{kj} - \prod_{\substack{i=1 \\ i \neq r}}^3 b_{ij} \right) b_{rj} \right) (s_r s_k - 1) \right] + \sum_{j=1}^3 \lambda_j b_{1j} b_{2j} b_{3j} (s_1 s_2 s_3 - 1) \right\}, \tag{60}$$

where  $Y_r, r = 1, 2, 3$ , is uniquely defined as

$$Y_r = Z_r + \sum_{\substack{k=1 \\ k \neq r}}^3 Z_{rk} + Z_{123}, \tag{61}$$

with the  $Z_r, Z_{rk}, Z_{123}$  being independent Poisson r.v.'s with parameters  $a_r, a_{rk}, a_{123}$ , respectively. This multivariate p.g.f. can be written as equation (4.1) of [25], by defining the parameters of the seven independent Poisson r.v.'s as

$$a_r = \lambda_r + \left( \sum_{j=1}^3 \lambda_j \left( 1 - \sum_{\substack{i=1 \\ i \neq r}}^3 b_{ij} + \prod_{\substack{k=1 \\ k \neq r}}^3 b_{kj} \right) b_{rj} \right) s_r, \\ a_{rk} = \left( \sum_{j=1}^3 \lambda_j \left( b_{kj} - \prod_{\substack{i=1 \\ i \neq r}}^3 b_{ij} \right) b_{rj} \right) s_r s_k, k > r,$$



$$a_{123} = \sum_{j=1}^3 \lambda_j b_{1j} b_{2j} b_{3j} s_1 s_2 s_3,$$

and by letting  $A_3$  be the sum of all terms multiplied by  $-1$ . Replacing (55) with the first-order neighbor matrix of model (22) and applying (60), the parameters of the seven Poisson distributions are obtained. This concludes the proof.  $\square$

**Remark C.1.** *The result of the previous theorem still holds for the reparameterized model in (27), where  $\beta_{11} \circ \varepsilon_{2,t-1}$  in location 1 and 3 returns the same observed value at time  $t$ . This becomes clear by following the above proof and rewriting (57) for location 2 as*

$$E[s_1^{\alpha_1 \circ X} s_2^{\alpha_2 \circ X} s_3^{\alpha_3 \circ X}] = E[s_1^{\alpha_1 \circ X} s_2^{\alpha_2 \circ X} s_3^{\alpha_1 \circ X}] \stackrel{ind.}{=} E\left[E[(s_1 s_3)^{\alpha_1 \circ X} | X] E[s_2^{\alpha_2 \circ X} | X]\right],$$

where the new joint terms are  $E[(1 - \alpha_1 + \alpha_1 s_1 s_3)^X (1 - \alpha_2 + \alpha_2 s_2)^X] = G_X((1 - \alpha_1 + \alpha_1 s_1 s_3)(1 - \alpha_2 + \alpha_2 s_2))$ . The remaining calculations are performed in an equal manner leading to a trivariate Poisson distribution but with different parameters than those of (22); see Theorem 4.1 and (28).

**Appendix D: Lemma 4.2**

The proof of Lemma 4.2 follows closely the arguments in the proof of Lemma 1 in [17].

*Proof.* For each component  $k, r = 1, \dots, 3$ , it follows that

$$E[(\mathbf{A} \circ \mathbf{X})(\mathbf{A} \circ \mathbf{X})^\top]_{k,r} = \sum_{s,n=1}^3 E[(a_{ks} \circ X_s)(a_{rn} \circ X_n)].$$

If  $a_{ks} \neq a_{rn}$  or  $s \neq n$ , then  $a_{ks} \circ X_s$  and  $a_{rn} \circ X_n$  are conditionally independent binomial random variables given  $X_s$  and  $X_n$ , respectively, and

$$E[(a_{ks} \circ X_s)(a_{rn} \circ X_n)] = a_{ks} a_{rn} E[X_s X_n].$$

However, the case  $a_{ks} = a_{rn}$  and  $s = n$ , leads to

$$E[(a_{ks} \circ X_s)(a_{rn} \circ X_n)] = E[(a_{ks} \circ X_s)^2] = a_{ks}^2 E[X_s^2] + a_{ks}(1 - a_{ks})E[X_s], \quad (62)$$

which includes the diagonal operations, i.e. those with  $k = r$  and  $s = n$  (which implies that  $a_{ks} = a_{rn}$ , as well as the off-diagonal operations with  $a_{ks} = a_{rn}$  and  $s = n$ . Therefore, with  $\mathbb{1}_{(a_{ks}=a_{rs})} = 1$  if  $a_{ks} = a_{rs}$  and 0 otherwise, then

$$E[(\mathbf{A} \circ \mathbf{X})(\mathbf{A} \circ \mathbf{X})^\top]_{k,r} = \sum_{s,n=1}^3 a_{ks} a_{rn} E[X_s X_n] + \sum_{s=1}^3 \mathbb{1}_{(a_{ks}=a_{rs})} a_{ks}(1 - a_{ks})E[X_s]. \quad (63)$$

Finally, it should be noted that

$$\sum_{r=1}^3 \mathbf{Q}^{(r)} E[\mathbf{X}] \mathbf{e}_r^\top = \sum_{r=1}^3 \begin{pmatrix} \vdots \\ \sum_{s=1}^3 \mathbb{1}_{(a_{ks}=a_{rs})} a_{ks} (1 - a_{ks}) E[X_s] \\ \vdots \end{pmatrix} \mathbf{e}_r^\top,$$

represents a matrix in which the  $(k, r)^{\text{th}}$  equals to  $\mathbb{1}_{(a_{ks}=a_{rs})} a_{ks} (1 - a_{ks}) E[X_s]$ , by defining  $\mathbf{Q}^{(r)}$  properly and by setting  $\mathbf{e}_r$  as the  $r^{\text{th}}$  unit vector. This concludes the proof.  $\square$

**Appendix E: CML recursive probabilities**

This appendix presents the proof for the derivation of the recursive equation (43).

*Proof.* Replacing  $t$  by  $t + 1$  in (42) ( $t \geq 2$ ) leads to

$$b_{\mathbf{k}\ell}(t + 1) = P(\boldsymbol{\varepsilon}_{t+1} = \mathbf{k}, \boldsymbol{\varepsilon}_t = \ell, \mathbf{Y}_{t+1} = \mathbf{y}_{t+1}, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1). \quad (64)$$

The recursive equation in  $t$  is constructed by introducing  $\boldsymbol{\varepsilon}_{t-1}$  in the above probability which will bring an explicit expression on  $b_{\cdot}(t)$  to the equation. Define  $\mathbf{i} = (i_1, i_2, i_3)$ . From the law of total probability it follows that

$$\begin{aligned} b_{\mathbf{k}\ell}(t + 1) &= \\ &= \sum_{\mathbf{i}=\mathbf{0}_3}^{\mathbf{y}_{t-1}} P(\boldsymbol{\varepsilon}_{t+1} = \mathbf{k}, \boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_{t+1} = \mathbf{y}_{t+1}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) \\ &= \sum_{\mathbf{i}=\mathbf{0}_3}^{\mathbf{y}_{t-1}} P(\mathbf{Y}_{t+1} = \mathbf{y}_{t+1}, \boldsymbol{\varepsilon}_{t+1} = \mathbf{k} | \boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_1 = \mathbf{y}_1) \times \\ &\quad \times P(\boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) \\ &= \sum_{\mathbf{i}=\mathbf{0}_3}^{\mathbf{y}_{t-1}} P(\mathbf{Y}_{t+1} = \mathbf{y}_{t+1} | \boldsymbol{\varepsilon}_{t+1} = \mathbf{k}, \boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_1 = \mathbf{y}_1) \times \\ &\quad \times P(\boldsymbol{\varepsilon}_{t+1} = \mathbf{k} | \boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2, \mathbf{Y}_1 = \mathbf{y}_1) \times \\ &\quad \times P(\boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1). \end{aligned}$$

Note that for any STINMA(1<sub>1</sub>) model driven by i.i.d. innovations which are independent of  $\mathbf{Y}_u$  for  $u < t$ , it follows that

$$\begin{aligned} b_{\mathbf{k}\ell}(t + 1) &= P(\boldsymbol{\varepsilon}_{t+1} = \mathbf{k}) P(\mathbf{Y}_{t+1} = \mathbf{y}_{t+1} | \boldsymbol{\varepsilon}_{t+1} = \mathbf{k}, \boldsymbol{\varepsilon}_t = \ell) \times \\ &\quad \times \sum_{\mathbf{i}=\mathbf{0}_3}^{\mathbf{y}_{t-1}} P(\boldsymbol{\varepsilon}_t = \ell, \boldsymbol{\varepsilon}_{t-1} = \mathbf{i}, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1), \quad (65) \end{aligned}$$

and by (64) the last parcel is simply

$$\sum_{i=\mathbf{0}_3}^{\mathbf{y}_{t-1}} P(\varepsilon_t = \ell, \varepsilon_{t-1} = i, \mathbf{Y}_t = \mathbf{y}_t, \dots, \mathbf{Y}_2 = \mathbf{y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) = \sum_{i=\mathbf{0}_3}^{\mathbf{y}_{t-1}} b_{\ell i}(t), \quad (66)$$

which constitutes the sum of  $b_{\ell i}$  parcels evaluated at time  $t$  that will be subsequently used to quantify  $b_{\mathbf{k}\ell}$  at time  $t+1$  in a recursive manner. Moreover, for the particular STINMA(1<sub>1</sub>) model defined in (27),

$$\begin{aligned} & P(\mathbf{Y}_{t+1} = \mathbf{y}_{t+1} | \varepsilon_{t+1} = \mathbf{k}, \varepsilon_t = \ell) \\ &= P\left(\beta_{10} \circ \ell_1 + \beta_{11} \circ \ell_2 + k_1 = y_{1,t+1}, \beta_{10} \circ \ell_2 + \frac{1}{2}\beta_{11} \circ (\ell_1 + \ell_3) + k_2 = y_{2,t+1}, \right. \\ &\quad \left. \beta_{10} \circ \ell_3 + \beta_{11} \circ \ell_2 + k_3 = y_{3,t+1}\right) \\ &= P\left(\beta_{10} \circ \ell_1 + \beta_{11} \circ \ell_2 = y_{1,t+1} - k_1, \beta_{10} \circ \ell_3 + \beta_{11} \circ \ell_2 = y_{3,t+1} - k_3\right) \times \\ &\quad \times P\left(\beta_{10} \circ \ell_2 + \frac{1}{2}\beta_{11} \circ (\ell_1 + \ell_3) = y_{2,t+1} - k_2\right) = p_{13} \times p_2, \end{aligned}$$

where

$$\begin{aligned} p_{13} &= \sum_{j_2=0}^{J_{k_1 k_3}} P(\beta_{11} \circ \ell_2 = j_2) P(\beta_{10} \circ \ell_1 = y_{1,t+1} - k_1 - j_2) \times \\ &\quad \times P(\beta_{10} \circ \ell_3 = y_{3,t+1} - k_3 - j_2) \\ p_2 &= \sum_{j_2=0}^{y_{2,t+1}-k_2} P(\beta_{10} \circ \ell_2 = j_2) P\left(\frac{1}{2}\beta_{11} \circ (\ell_1 + \ell_3) = y_{2,t+1} - k_2 - j_2\right) \end{aligned}$$

with  $J_{k_1 k_3} = \min_t\{y_{1,t+1}-k_1, y_{3,t+1}-k_3\}$  due to the  $P(\beta_{11} \circ \ell_2 \leq y_{1,t+1}-k_1) = 1$  and  $P(\beta_{11} \circ \ell_2 \leq y_{3,t+1}-k_3) = 1$ , restrictions that arise from the parcels evaluating probabilities of sums of r.v.'s. Recalling that  $\beta \circ \ell \sim B(\ell, \beta)$  with  $\ell \in \mathbb{N}_0$  then the above parcels are of the type  $P(\beta \circ \ell = x) = \binom{\ell}{x} \beta^x (1-\beta)^{\ell-x}$ . Altogether, it follows that

$$b_{\mathbf{k}\ell}(t+1) = P(\varepsilon_{t+1} = \mathbf{k}) p_{13} p_2 \sum_{i=\mathbf{0}_3}^{\mathbf{y}_{t-1}} b_{\ell i}(t), \quad (67)$$

resulting in (43), which concludes the proof.  $\square$

**Remark E.1.** Note that the assumption that all thinning operations are independent, such as in the STINMA(1<sub>1</sub>) model (22), the above parcel  $p_{13}$  is replaced by the following product of probabilities,

$$\begin{aligned} p_{13} &= P(\beta_{10} \circ \ell_1 + \beta_{11} \circ \ell_2 + k_1 = y_{1,t+1}) \times P(\beta_{10} \circ \ell_3 + \beta_{11} \circ \ell_2 + k_3 = y_{3,t+1}) \\ &= p_1 \times p_3 \end{aligned}$$

with

$$\begin{aligned}
 p_1 &= \sum_{j_1=0}^{y_{1,t+1}-k_1} P(\beta_{10} \circ \ell_1 = j_1) P(\beta_{11} \circ \ell_2 = y_{1,t+1} - k_1 - j_1), \\
 p_3 &= \sum_{j_3=0}^{y_{3,t+1}-k_3} P(\beta_{30} \circ \ell_3 = j_3) P(\beta_{31} \circ \ell_2 = y_{3,t+1} - k_3 - j_3), \quad (68)
 \end{aligned}$$

and consequently  $b_{\mathbf{k}\ell}(t+1) = P(\varepsilon_{t+1} = \mathbf{k}) p_1 p_2 p_3 \sum_{i=0_3}^{y_{t-1}} b_{\ell i}(t)$ .

### Acknowledgments

The authors thank the Associate Editor and the three reviewers for their valuable comments on an earlier draft of this manuscript. The authors thank Administração Central do Sistema de Saúde for providing hospital admissions data. The authors are also grateful to Dr. Philipp Wittenberg (HSU Hamburg) for the helpful advice on implementing the parameter estimation in R.

### Funding

This work was partially funded by the Foundation for Science and Technology, FCT (<https://www.fct.pt/>), Portugal through national (MEC) and European structural (FEDER) funds, in the scope of the research projects IEETA/UA (UIDB/00127/2020, [www.ieeta.pt](http://www.ieeta.pt)) and CEMAT/IST/UL (UIDB/04621/2020, <http://cemat.ist.utl.pt>). This work also benefited from funding of the National Network for Advanced Computing, RNCA (<https://rnca.fccn.pt/>), which is part of the Scientific Computing Unit (FCCN) of the FCT, under the project CPCA/A1/449427/2021. AM acknowledges a PhD grant from the FCT (SFRH/BD/143973/2019) funded by the Portuguese state budget, through the Ministry for Science, Technology and Higher Education and, by the European Social Fund within the Framework of PORTUGAL2020, namely through Programa Operacional Capital Humano (PO CH) and Programa Operacional Regional do Centro (Centro 2020).

### References

- [1] ALDOR-NOIMAN, S., BROWN, L. D., FOX, E. B. and STINE, R. A. (2016). Spatio-Temporal Low Count Processes with Application to Violent Crime Events. *Statistica Sinica* **26** 1587–1610. <https://doi.org/10.5705/ss.2014.217t>. MR3586230
- [2] ALEKSANDROV, B. and WEISS, C. H. (2020). Parameter Estimation and Diagnostic Tests for INMA(1) Processes. *TEST* **29** 196–232. MR4063388
- [3] SANTOS BAQUERO, O. (2019). ggsn: North Symbols and Scale Bars for Maps Created with ‘ggplot2’ or ‘ggmap’ R package version 0.5.0.

- [4] BATES, D. and MAECHLER, M. (2021). Matrix: Sparse and Dense Matrix Classes and Methods R package version 1.4-0.
- [5] BENGTTSSON, H. (2021). R.utils: Various Programming Utilities R package version 2.11.0.
- [6] BOROvkOVA, S., LOPUHAA, H. and NURANI, B. (2002). Generalized STAR Model with Experimental Weights. In *Proceedings of the 17th International Workshop on Statistical Modelling* 139–147.
- [7] BRÄNNÄS, K. and HALL, A. (2001). Estimation in Integer-Valued Moving Average Models. *Applied Stochastic Models in Business and Industry* **17** 277–291. [MR1864091](#)
- [8] BRÄNNÄS, K. and QUORESHI, S. (2010). Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks. *Applied Financial Economics* **20** 1429–1440.
- [9] BUTEIKIS, A. and LEIPUS, R. (2020). An Integer-Valued Autoregressive Process for Seasonality. *Journal of Statistical Computation and Simulation* **90** 391–411. [MR4044916](#)
- [10] MICROSOFT CORPORATION and WESTON, S. (2020). doParallel: Foreach Parallel Adaptor for the ‘parallel’ Package R package version 1.0.16.
- [11] DAVIS, R. A., HOLAN, S. H., LUND, R. and RAVISHANKER, N. (2016). *Handbook of Discrete-Valued Time Series*. CRC Press. [MR3642975](#)
- [12] DAVIS, R. A., FOKIANOS, K., HOLAN, S. H., JOE, H., LIVSEY, J., LUND, R., PIPIRAS, V. and RAVISHANKER, N. (2021). Count Time Series: A Methodological Review. *Journal of the American Statistical Association* **116** 1533–1547. [MR4309291](#)
- [13] DION, J. P., GAUTHIER, G. and LATOUR, A. (1995). Branching Processes with Immigration and Integer-Valued Time Series. *Serdica Mathematical Journal* **21** 123–136. [MR1338812](#)
- [14] EDELBUETTEL, D. and BALAMUTA, J. J. (2018). Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician* **72** 28–36. <https://doi.org/10.1080/00031305.2017.1375990>. [MR3790565](#)
- [15] EDELBUETTEL, D. and SANDERSON, C. (2014). RcppArmadillo: Accelerating R with High-Performance C++ Linear Algebra. *Computational Statistics and Data Analysis* **71** 1054–1063. [MR3132026](#)
- [16] EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press. [MR1270903](#)
- [17] FRANKE, J. and SUBBA RAO, T. (1993). Multivariate First-Order Integer-Valued Autoregressions Technical Report, University of Kaiserslautern.
- [18] GERBER, F. and FURRER, R. (2019). optimParallel: An R Package Providing a Parallel Version of the L-BFGS-B Optimization Method. *The R Journal* **11** 352–358. <https://doi.org/10.32614/RJ-2019-030>
- [19] GHODSI, A. (2015). Conditional Maximum Likelihood Estimation of the First-Order Spatial Integer-Valued Autoregressive (SINAR(1,1)) Model. *Journal of the Iranian Statistical Society* **14** 15–36. <https://doi.org/10.7508/jirss.2015.02.002>. [MR3518270](#)
- [20] GHODSI, A., SHITAN, M. and BAKOUCH, H. S. (2012). A First-Order Spatial Integer-Valued Autoregressive SINAR (1, 1) Model. *Communications*

- in Statistics-Theory and Methods* **41** 2773–2787. [MR2946646](#)
- [21] GIRONDOT, M. (2021). HelpersMG: Tools for Environmental Analyses, Ecotoxicology and Various R Functions R package version 4.8.
- [22] HUDA, N., MUKHAIYAR, U. and PASARIBU, U. (2021). The Approximation of GSTAR Model for Discrete Cases through INAR Model. In *Journal of Physics: Conference Series* **1722** 012100. IOP Publishing.
- [23] JOWAHEER, V., SUNECHER, Y. and MAMODE KHAN, N. (2016). A Non-Stationary BINAR(1) Process with Negative Binomial Innovations for Modeling the Number of Goals in the First and Second half: The Case of Arsenal Football Club. *Communications in Statistics: Case Studies, Data Analysis and Applications* **2** 21–33. [MR3523164](#)
- [24] LATOUR, A. (1997). The Multivariate GINAR (p) Process. *Advances in Applied Probability* **29** 228–248. [MR1432938](#)
- [25] MAHAMUNULU, D. (1967). A Note on Regression in the Multivariate Poisson Distribution. *Journal of the American Statistical Association* **62** 251–258. [MR0216613](#)
- [26] MARTIN, R. L. and OEPPEN, J. (1975). The Identification of Regional Forecasting Models using Space: Time Correlation Functions. *Transactions of the Institute of British Geographers* 95–118.
- [27] MARTINS, A., SCOTTO, M. G., DEUS, R., MONTEIRO, A. and GOUVEIA, S. (2021). Association between Respiratory Hospital Admissions and Air quality in Portugal: A Count Time Series Approach. *PLOS One* **16** e0253455.
- [28] MCKENZIE, E. (1988). Some ARMA Models for Dependent Sequences of Poisson Counts. *Advances in Applied Probability* **20** 822–835. [MR0968000](#)
- [29] PEBESMA, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* **10** 439–446. <https://doi.org/10.32614/RJ-2018-009>
- [30] PEDELI, X. and KARLIS, D. (2011). A Bivariate INAR(1) Process with Application. *Statistical modelling* **11** 325–349. [MR2906704](#)
- [31] PEDELI, X. and KARLIS, D. (2013). Some Properties of Multivariate INAR(1) Processes. *Computational Statistics & Data Analysis* **67** 213–225. [MR3079598](#)
- [32] PEDERSEN, T. L. (2019). ggforce: Accelerating ‘ggplot2’ R package version 0.3.1.
- [33] PEDERSEN, T. L. (2020). patchwork: The Composer of Plots R package version 1.1.1.
- [34] PFEIFER, P. E. (1979). Spatial-Dynamic Modeling, PhD thesis, Georgia Institute of Technology.
- [35] PFEIFER, P. E. and DEUTSCH, S. J. (1980a). A Three-Stage Iterative Procedure for Space-Time Modeling. *Technometrics* **22** 35–47.
- [36] PFEIFER, P. E. and DEUTSCH, S. J. (1980b). Identification and Interpretation of First Order Space-Time ARMA Models. *Technometrics* **22** 397–408.
- [37] QUORESHI, A. S. (2006). Bivariate Time Series Modeling of Financial Count Data. *Communications in Statistics-Theory and Methods* **35**

- 1343–1358. [MR2328481](#)
- [38] QUORESHI, A. S. (2008). A Vector Integer-Valued Moving Average Model for High Frequency Financial Count Data. *Economics Letters* **101** 258–261.
- [39] SANTOS, C., PEREIRA, I. and SCOTTO, M. G. (2021). On the Theory of Periodic Multivariate INAR Processes. *Statistical Papers* **62** 1291–1348. [MR4262195](#)
- [40] SCOTTO, M. G., WEISS, C. H. and GOUVEIA, S. (2015). Thinning-Based Models in the Analysis of Integer-Valued Time Series: A Review. *Statistical Modelling* **15** 590–618. [MR3441230](#)
- [41] SCOTTO, M. G., WEISS, C. H., SILVA, M. E. and PEREIRA, I. (2014). Bivariate Binomial Autoregressive Models. *Journal of Multivariate Analysis* **125** 233–251. [MR3163841](#)
- [42] SILVA, I., SILVA, M. E. and TORRES, C. (2020). Inference for Bivariate Integer-Valued Moving Average Models based on Binomial Thinning Operation. *Journal of Applied Statistics* **47** 2546–2564. [MR4149569](#)
- [43] SOETAERT, K. (2009). rootSolve: Nonlinear Root Finding, Equilibrium and Steady-State Analysis of Ordinary Differential Equations R package 1.6.
- [44] STEUTEL, F. W. and VAN HARN, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability* 893–899. [MR0542141](#)
- [45] SUBBA RAO, T. and COSTA ANTUNES, A. M. (2004). Spatio-Temporal Modelling of Temperature Time Series: A Comparative Study. In *Time series analysis and applications to geophysical systems* 123–150. Springer. [MR2111482](#)
- [46] SUNECHER, Y. (2021). Application of the BINARMA(1, 1) Model with NB Innovations on Accident Data. In *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* 1–4. IEEE.
- [47] SUNECHER, Y., MAMODE KHAN, N. and JOWAHEER, V. (2017). Estimating the Parameters of a BINMA Poisson Model for a Non-Stationary Bivariate Time Series. *Communications in Statistics-Simulation and Computation* **46** 6803–6827. [MR3764940](#)
- [48] SUNECHER, Y., MAMODE KHAN, N. and JOWAHEER, V. (2019). The Non-Stationary BINARMA(1, 1) Process with Poisson Innovations: An Application on Accident Data. *International Journal of Mathematical and Computational Sciences* **13** 193–196.
- [49] SUNECHER, Y., MAMODE KHAN, N. and JOWAHEER, V. (2020). BINMA(1) Model with COM-Poisson Innovations: Estimation and Application. *Communications in Statistics-Simulation and Computation* **49** 1631–1652. [MR4120951](#)
- [50] TABANDEH, A. and GHODSI, A. (2022). First-Order Spatial Random Coefficient Non-Negative Integer-Valued Autoregressive (SRCINAR(1,1)) model. *Communications in Statistics – Simulation and Computation* **0** 1–13. <https://doi.org/10.1080/03610918.2022.2083164>
- [51] R CORE TEAM (2021). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
- [52] WEISS, C. H. (2008). Serial Dependence and Regression of Poisson

- INARMA Models. *Journal of Statistical Planning and Inference* **138** 2975–2990. [MR2442225](#)
- [53] WEISS, C. H. (2012). Fully Observed INAR (1) Processes. *Journal of Applied Statistics* **39** 581–598. [MR2880436](#)
- [54] WEISS, C. H. (2021). Stationary Count Time Series Models. *Wiley Interdisciplinary Reviews: Computational Statistics* **13** e1502. [MR4186769](#)
- [55] WEISS, C. H., FELD, M. H.-J., MAMODE KHAN, N. and SUNECHER, Y. (2019). INARMA Modeling of Count Time Series. *Stats* **2** 284–320.
- [56] WICKHAM, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [57] WICKHAM, H., FRANÇOIS, R., HENRY, L. and MÜLLER, K. (2020). dplyr: A Grammar of Data Manipulation R package version 1.0.2.
- [58] ZUCCHINI, W. and MACDONALD, I. L. (2009). *Hidden Markov Models for Time Series: an Introduction using R*. Chapman and Hall/CRC. [MR2523850](#)