Received: 3 March 2025 | Revised: 3 March 2025 | Accepted: 17 March 2025

DOI: 10.1111/sjos.12784

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

A unifying class of compound Poisson integer-valued ARMA and GARCH models

Johannes Bracher[1,2] | **Barbora Němcová**[1,3]

[1]Institute of Statistics, Karlsruhe Institute of Technology, Germany

[2]Computational Statistics Group, Heidelberg Institute for Theoretical Studies, Germany

[3]HIDSS4Health - Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany

Correspondence
Johannes Bracher, Institute of Statistics, Karlsruhe Institute of Technology, Blücherstraße 17, 76185 Karlsruhe, Germany.
Email: johannes.bracher@kit.edu

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 512483310; Helmholtz Association

Abstract

INAR (integer-valued autoregressive) and INGARCH (integer-valued GARCH) models are among the most commonly employed approaches for count time series modeling, but have been studied in largely distinct strands of literature. In this paper, a new class of generalized integer-valued ARMA (GINARMA) models is introduced which unifies a large number of compound Poisson INAR and INGARCH processes. Its stochastic properties, including stationarity and geometric ergodicity, are studied. Particular attention is given to a generalization of the INAR(p) model which parallels the extension of the INARCH(p) to the INGARCH(p, q) model. For inference, we consider maximum likelihood, Gaussian quasi-likelihood, and moment-based approaches, along with likelihood ratio tests to distinguish between selected instances of our class. Models from the proposed class have a natural interpretation as stochastic epidemic processes, which throughout the article is used to illustrate our arguments. In a case study, different instances, including both established and newly introduced models, are applied to weekly case numbers of measles and mumps in Bavaria, Germany.

KEYWORDS

branching process, count time series, forward algorithm, geometric ergodicity, integer-valued ARMA

1 | INTRODUCTION

Count time series arise in many contexts from hydrology (McKenzie, 1985) to criminology and traffic studies (Scotto et al., 2015). Numerous modeling approaches for such data exist, including, for example, hidden Markov (Zucchini & MacDonald, 2009), generalized linear ARMA (Benjamin et al., 2003) and latent Gaussian models (Jia et al., 2023). This diversity led a recent review to conclude that "the field developed without a unifying theory" (Davis et al., 2021). In the present paper, we aim to provide an overarching framework for two particularly influential model classes, namely the INAR (integer-valued autoregressive) and INGARCH (integer-valued GARCH) classes. These have been highlighted as "probably the most widely used approaches for stationary count time series" (Weiß, 2021). While INAR models employ thinning operations and resemble branching processes (Dion et al., 1995), INGARCH models take their starting point in generalized linear regression. Despite some known links between the two (Lu, 2021; Weiß, 2015), they have been treated in largely distinct strands of literature. Recently, various model classes have been suggested which combine elements of INGARCH and INAR models (Aknouche & Scotto, 2024; Weiß & Zhu, 2024), but these always contain at most one of the two as a special case. Our contributions to bridging this gap are the following:

- Building on generalized INAR models (Latour, 1998), we define a broad model class comprising many well-known INAR and INGARCH processes. As it is inspired by a state-space representation of the continuous-valued ARMA model and shares some of its properties, we refer to it as the *generalized INARMA* class.
- As an important special case, we study a generalization of the INAR(p) model which parallels the extension of the INARCH(p) to the INGARCH(p, q).
- Properties of the new class are studied with a focus on compound Poisson (CP) formulations, and various inference schemes are proposed.

The practical benefit of the generalized INARMA class is that it provides a unified and modular framework covering many established models. These are characterized by their lag orders and a set of distributional assumptions. As different applications may suggest other types of distributions, this can guide model choice. In many cases, relevant models can even be nested into suitable sub-classes of the GINARMA framework, meaning that hypothesis tests can be conceived to select appropriate specifications.

To strengthen the intuition for the "mechanics" of the new class, we propose an interpretation as a stochastic epidemic process. Indeed, both INAR (Cardinal et al., 1999; Pedeli et al., 2015) and INGARCH (Bracher & Held, 2021; Ferland et al., 2006) models are commonly applied to infectious disease counts, though often without discussion of the implied assumptions on disease spread (see Bauer & Wakefield, 2018 for an exception). Building on these arguments, we will conclude with a case study on measles and mumps in the German state of Bavaria. Similarly to Kucharski et al. (2014), we will estimate local effective reproductive numbers and the fraction of imported cases.

The article is structured as follows. In Section 2, we provide some background on CP-INAR and INGARCH models. In Section 3, we introduce our general model class, before turning to an extension of the INAR class in Section 4. Sections 5 and 6 are concerned with inference and simulation studies, respectively. In Section 7, the real-data application is presented before Section 8 concludes with a discussion.

2 | PRELIMINARIES

We start by reviewing relevant fundamentals of (generalized) INAR and INGARCH models.

2.1 | Poisson (G)INAR(1) and INGARCH(1,1) models

The generalized INAR(1) model (Latour, 1998), or GINAR(1), is defined as $\{X_t, t \in \mathbb{Z}\}$ with

$$X_t = \alpha \bullet X_{t-1} + \varepsilon_t \tag{1}$$

and $\alpha > 0$. The *innovations* $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) count random variables with mean $\nu > 0$ and variance $\sigma_\nu^2 < \infty$, while $\bullet$ is the *generalized thinning operator*. With $N \in \mathbb{N}_0$ and $\alpha > 0$ it is defined as $\alpha \bullet N = 0$ if $N = 0$ and

$$\alpha \bullet N = \sum_{i=1}^{N} Z_i \tag{2}$$

otherwise. The *counting series* $Z_1, \ldots, Z_N$ consists of i.i.d. draws from an integer-valued distribution with mean $\alpha > 0$ and variance σ_α^2. Borrowing terminology from branching process theory, we will refer to this distribution as the *offspring distribution*. All thinnings in Equation (1) are performed independently of each other, the innovations $\{\varepsilon_t\}$ and the past of the process $\{X_t\}$, an assumption we will make throughout the paper unless relaxed explicitly.

Model (1) can be read as an adaptation of the classic Gaussian AR(1) process with multiplication replaced by generalized thinning. As GINAR(1) models are first-order conditionally linear autoregressive models (CLAR; Grunwald et al., 2000), they preserve many stochastic properties of their continuous counterpart. Two particularly influential instances of the class are the Poisson INAR(1) (Al-Osh & Alzaid, 1987; McKenzie, 1985) and INARCH(1) (Ferland et al., 2006; Fokianos et al., 2009) models (the naming of the latter being somewhat controversial, see Remark 4.1.2 in Weiß, 2018). While in both models we assume $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{Pois}(\nu)$, the offspring distributions differ. In the INAR(1), given by

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \tag{3}$$

binomial thinning $\circ$ is used, which results from $Z_i \overset{\text{i.i.d.}}{\sim} \text{Bern}(\alpha)$ (Steutel & van Harn, 1979). For the INARCH(1), Poisson thinning $\star$ with $Z_i \overset{\text{i.i.d.}}{\sim} \text{Pois}(\alpha)$ is assumed instead.

> *Remark* 1. GINAR(1) models can be thought of as simple epidemic processes (Cardinal et al., 1999). Each of X_t infectives present in a population at time t causes on average α new infectives ("offspring") at $t+1$ and then recovers. Outside sources cause infections at a rate ν. As will be seen in Section 3.2, similar interpretations apply to extended models.

While the GINAR(1) representation of the Poisson INARCH(1) is well-known (e.g., Weiß, 2018, p. 56), the model is more commonly defined in terms of a conditional generalized linear regression model (GLM). It then becomes $\{X_t, t \in \mathbb{N}\}$ with

$$X_t | X_{t-1}, \ldots, X_0, \lambda_0 \sim \text{Pois}(\lambda_t) \tag{4}$$

$$\lambda_t = \nu + \alpha X_{t-1} \tag{5}$$

and fixed starting values λ_0, X_0. This formulation is attractive as it can be extended to the Poisson INGARCH(1, 1) model (Ferland et al., 2006), where (5) becomes

$$\lambda_t = \nu + \alpha X_{t-1} + \beta \lambda_{t-1} \tag{6}$$

with $0 \leq \beta$. The *feedback term* $\beta\lambda_{t-1}$ here leads to an ARMA(1, 1) autocorrelation function.

2.2 | Compound Poisson distributions

To handle overdispersion in a flexible way, Poisson (G)INAR and INGARCH models are commonly extended using compound Poisson (CP) distributions. A random variable Y is said to follow a CP distribution (Feller, 1968, Chapter 3) if it can be written as a randomly stopped sum $Y = \sum_{i=1}^{N} Z_i$, where N is Poisson distributed and $Z_1, \ldots, Z_N \overset{\text{i.i.d.}}{\sim} G(\theta)$ independently of N. We assume throughout that the *cluster distribution* $G(\theta)$ has a single parameter θ and support $\{1, \ldots, r\}$, where r is the *order* of the CP distribution and $r = \infty$ is allowed. For simplicity we identify θ with the mean of G and in analogy to (2) use the shorthand

$$\theta * N = \sum_{i=1}^{N} Z_i, \quad Z_1, \ldots, Z_N \overset{\text{i.i.d.}}{\sim} G(\theta). \tag{7}$$

We assume $0 < \theta < \infty$ and denote the variance of $G(\theta)$ by $\sigma_\theta^2 < \infty$. Setting $N \sim \text{Pois}(\mu/\theta)$, we obtain a CP distribution with mean $\mathbb{E}(Y) = \mu$ and variance $\text{Var}(Y) = \mu \times (\sigma_\theta^2/\theta + \theta)$.

> *Remark* 2. We use the term "*cluster distribution*" as in disease modeling and ecology, CP distributions are often applied to phenomena that occur in clusters. The number N of clusters is then Poissonian, while the number of units per cluster follows $G(\theta)$. Note that other authors like Weiß (2018) use "*compounding distribution*" instead.

Two popular CP distributions are the Hermite and negative binomial. A random variable Y is Hermite distributed if it can be written as $Y = A_1 + 2A_2$ where independently $A_1 \sim \text{Pois}(\lambda_1)$, $A_2 \sim \text{Pois}(\lambda_2)$. In slight variation of Gupta and Jain (1974) we parameterize the distribution by its mean $\mu = \lambda_1 + 2\lambda_2$ and a dispersion parameter $\psi = 2\lambda_2/(\lambda_1 + 2\lambda_2) \in [0, 1]$, implying $\text{Var}(Y) = (1 + \psi)\mu$. The probability mass function is then

$$\Pr(Y = y) = \exp\left[\mu\left(-1 + \frac{\psi}{2}\right)\right]\mu^y(1-\psi)^y \sum_{j=0}^{[y/2]} \frac{\psi^j}{2^j \mu^j (1-\psi)^{2j}(y-2j)!j!}, y = 0, 1, 2, \ldots$$

where $[y/2]$ is the integer part of $y/2$. The Hermite is a CP distribution of order $r = 2$, see Supporting Information A.1.1.

For the negative binomial distribution, we likewise use a parameterization via its mean μ and a dispersion parameter $\psi > 0$, given by

$$\Pr(Y = y) = \frac{\Gamma(1/\psi + y)}{y!\Gamma(1/\psi)} \left(\frac{1/\psi}{1/\psi + \mu}\right)^{1/\psi} \left(\frac{\mu}{1/\psi + \mu}\right)^y,$$

while $\text{Var}(Y) = (1 + \psi\mu)\mu$. The negative binomial distribution is a CP distribution with a logarithmic cluster distribution (Weiß, 2018) and hence $r = \infty$, see Supporting Information A.2.1.

2.3 | CP-(G)INAR(p) and INGARCH(p, q) models

In the (generalized) INAR framework, the extension of the model (1) to the CP case (Schweer & Weiß, 2014) is straightforward as only the innovation distribution is replaced by a CP with mean ν and variance σ_ν^2. Hermite and negative binomial innovations have been considered, for example, by Fernández- Fontelo et al. (2017) and Pedeli et al. (2015). A higher-order GINAR(p) model is obtained by setting (Dion et al., 1995)

$$X_t = \sum_{i=1}^{p} \alpha_i \bullet X_{t-i} + \varepsilon_t. \tag{8}$$

Slightly generalizing Latour (1998), we allow for dependent offspring $(\alpha_1 \bullet X_t, \ldots, \alpha_p \bullet X_t)$, see Definition 1 in the next section for details. This is because we will extend the INAR(p) by Alzaid and Al-Osh (1990), where

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \varepsilon_t \tag{9}$$

with $\alpha_1, \ldots, \alpha_p \geq 0, 0 < \sum_{i=1}^{p} \alpha_i < 1$ is combined with multinomial thinnings,

$$(\alpha_1 \circ X_t, \ldots, \alpha_p \circ X_t) \sim \text{Mult}(X_t, \alpha_1, \ldots, \alpha_p). \tag{10}$$

We note that an equally well-known INAR(p) model with independent thinning operations has been proposed by Du and Li (1991), but it is less fruitful within our framework.

The GINAR(p) model can be generalized further to the GINARMA$_{\text{DGL}}(p, q)$, given by

$$X_t = \sum_{i=1}^{p} \alpha_i \bullet X_{t-i} + \sum_{j=1}^{q} \delta_j \bullet \varepsilon_{t-j} + \varepsilon_t, \tag{11}$$

where all thinnings $\bullet$ are performed independently of each other. We here add the initials of its authors – Dion, Gauthier and Latour – to the notation to distinguish this GINARMA model from our suggestion presented later on.

To extend (6) to a CP-INGARCH(p, q) model we adopt notation from Weiß et al. (2017). The model is then defined as $\{X_t, t \in \mathbb{N}\}$ with

$$N_t \mid X_{t-1}, \ldots, X_{1-p}, \lambda_0, \ldots, \lambda_{1-q} \sim \text{Pois}(\lambda_t/\theta) \tag{12}$$

$$X_t = \sum_{i=1}^{N_t} Z_{t,i} \quad \text{where} \quad Z_{t,1}, \ldots, Z_{t,N_t} \overset{\text{i.i.d.}}{\sim} G(\theta) \tag{13}$$

$$\lambda_t = \nu + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}. \tag{14}$$

Here, $\lambda_{1-q}, \ldots, \lambda_0 \geq 0$ and $X_{1-p}, \ldots, X_0 \in \mathbb{N}_0$ are again fixed and we assume $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0, \sum_{i=1}^{p} \alpha_i > 0$. Given the past, X_t then follows a CP distribution with mean λ_t and variance $\lambda_t \times (\sigma_\theta^2/\theta + \theta)$. This is more restrictive than in Gonçalves et al. (2015a) where θ_t is a function of λ_t. It nonetheless contains, for example, the negative binomial (Weiß, 2018; Xu et al., 2012),

generalized Poisson (Xu et al., 2012; Zhu, 2012) and Neyman Type A (Gonçalves et al., 2015b) INGARCH models. For the Hermite and negative binomial cases we provide details in Supporting Information A.1.2 and A.2.2. We note that despite its name, the INGARCH(p, q) has an ARMA$(\max[p, q], p)$-like correlation structure (Ferland et al., 2006, section 2.7); nonetheless it behaves quite differently than the GINARMA$_{\text{DGL}}(p, q)$, see Section 4.2.1.

3 | A NEW GINARMA(p, q) MODEL

3.1 | Model definition

We now propose an alternative GINARMA extension of the model (8), which as we shall see subsumes all INAR and INGARCH models from Section 2. Rather than directly replacing the multiplications in the Gaussian ARMA(p, q) model

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j} + \nu + \varepsilon_t, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{N}(0, \sigma_\varepsilon^2) \tag{15}$$

by thinnings as in Equation (11), we start by re-formulating it as follows. Setting

$$\kappa_i = \frac{\alpha_i + \delta_i}{1 + \sum_{j=1}^{q} \delta_j}, \quad \beta_j = -\delta_j, \quad \tau = \frac{\nu}{1 + \sum_{j=1}^{q} \delta_j},$$

with $\delta_j = 0$ for $j > q$, an equivalent of (15) is (see Supporting Information F.1)

$$X_t = \left(1 - \sum_{j=1}^{q} \beta_j\right) \times E_t + \tau + \varepsilon_t. \tag{16}$$

Here, $\{E_t\}$ is an auxiliary process defined as

$$E_t = \sum_{j=1}^{q} \beta_j \times E_{t-j} + \sum_{i=1}^{p} \kappa_i \times X_{t-i}. \tag{17}$$

This is a state-space representation of the ARMA(p, q), see also Remark 3. We are not aware of any previous use of this exact form, but similar ones can be found, for example, in de Jong and Penzer (2004). Based on the above, we define our model as follows. Note that we add a compounding step that serves to accommodate CP-INGARCH models in the class.

Definition 1. The *GINARMA*(p, q) model is a stochastic process $\{X_t, t \in \mathbb{N}\}$ with

$$X_t = \theta * \left[\left(1 - \sum_{j=1}^{q} \beta_j\right) \circ E_t + \varepsilon_t\right] \tag{18}$$

$$E_t = \sum_{j=1}^{q} \beta_j \circ E_{t-j} + \sum_{i=1}^{p} \kappa_i \bullet X_{t-i} \tag{19}$$

and $\kappa_i, \beta_j \geq 0; 0 < \sum_{i=1}^{p} \kappa_i; 0 \leq \sum_{j=1}^{q} \beta_j < 1$. Specifically, the following is assumed.

i. The sequence $\{\varepsilon_t\}$ consists of i.i.d. realizations from an integer-valued innovation distribution with mean $0 < \tau < \infty$ and variance $\sigma_\tau^2 < \infty$.
ii. The offspring $(\kappa_1 \bullet X_t, \ldots, \kappa_p \bullet X_t)$ result from a generalized thinning operation $\bullet$ and can be dependent. Specifically, we assume $(\kappa_1 \bullet X_t, \ldots, \kappa_p \bullet X_t) = \mathbf{Z}_{t,1} + \cdots + \mathbf{Z}_{t,X_t}$, where, independently of X_t, the vectors $\mathbf{Z}_{t,j}$ are i.i.d. on $\mathbb{N}_0^p$ with $\mathbb{E}(\mathbf{Z}_{t,j}) = (\kappa_1, \ldots, \kappa_p)$ and finite variances $\sigma_{\kappa_1}^2, \ldots, \sigma_{\kappa_p}^2$. To avoid dealing with, for example, purely even-valued offspring distributions we assume $\text{Prob}(\kappa_i \bullet 1 = 1) > 0$ if $\kappa_i > 0$.
iii. The thinnings of E_t are coupled via

$$\left[\beta_1 \circ E_t, \ldots, \beta_q \circ E_t, \left(1 - \sum_{j=1}^{q} \beta_j\right) \circ E_t\right] \Big| E_t \sim \text{Mult}\left(E_t; \beta_1, \ldots, \beta_q, 1 - \sum_{j=1}^{q} \beta_j\right). \quad (20)$$

iv. As in Equation (7), $\theta *$ denotes a compounding step with a clustering distribution G. Its mean and variance are $0 < \theta < \infty$ and $\sigma_\theta^2 < \infty$, respectively.
v. Apart from the dependencies between different thinnings of the same X_t or E_t introduced in (ii) and (iii), all thinnings and compoundings are performed independently of each other, the past of the process and the innovation sequence $\{\varepsilon_t\}$.
vi. Unless stated otherwise, the values $X_{1-p}, \ldots, X_0$, $E_{1-q}, \ldots, E_0$ are fixed. In some instances, we will initialize them with the respective stationary distributions instead.

GINARMA models are thus characterized by the orders p and q along with an innovation, an offspring, and a compounding distribution. As detailed in Section 3.3, many INGARCH models can be obtained via Poisson offspring (setting $\bullet$ to Poisson thinning $\star$). Bernoulli offspring, on the other hand, yield an extension of the INAR class; see Section 4. We note that assumption (iii) may seem arbitrary at first sight, but is central to obtaining appealing stochastic properties. Notably, it ensures that the model reduces to the GINAR(p) if the compounding step $\theta *$ is omitted and $\beta_1 = \cdots = \beta_q = 0$.

Remark 3. From Equation (18), it is clear that given E_t, X_t is independent of $E_1, \ldots, E_{t-1}$ and $X_1, \ldots, X_{t-1}$. We are thus dealing with a (generalized) state-space model (Brockwell & Davis, 2016, Chapter 9.8; Weiß, 2018, Remark 5.2.1), with $\{E_t\}$ the state variable. However, using terminology from these references, the model in its general form is neither purely *parameter-driven* nor *observation-driven*. While $\{X_t\}$ has no Markov property unless $q = 0$, $\{E_t\}$ is a $\max(p, q)$-th order Markov process, as is $\{(E_t, X_t)\}$; see Corollaries 1 and 3.

3.2 | Interpretation as a stochastic epidemic process

Formulation (18) to (20) can be interpreted as a discrete-time model for the spread of an infectious disease, which provides a useful language and intuition for the following. We first illustrate this for $p = q = 1$ and omitting the compounding step, that is, we consider

$$X_t = (1 - \beta) \circ E_t + \varepsilon_t; \quad E_t = \beta \circ E_{t-1} + \kappa \bullet X_{t-1}. \quad (21)$$

A graphical display of the following interpretation is provided in the top panel of Figure 1.

1. X_t is the number of infectious individuals at time t. These stay infectious for one time period and independently cause new infections with mean κ and variance σ_κ^2.
2. Individuals newly infected at time t do not necessarily become infectious already at $t+1$. Instead, they enter into an "*exposed pool*" E_{t+1}. Following common disease modeling terminology, *exposed* individuals are already infected, but not yet infectious.
3. At each time t, each of the E_t exposed individuals can either remain in the exposed pool (with probability β) or advance to infectiousness (with probability $1-\beta$).
4. An exposed individual from E_t advancing to infectiousness becomes part of X_t.
5. At each time t, ε_t individuals get infectious due to external sources.

The effective reproductive number, that is, the mean number of new infections caused by one infected is $R_e = \kappa$. The latent period, defined as the number of time points an infected spends in the exposed pool, is geometrically distributed with mean $1/(1-\beta)$. The same holds for the generation time, that is, the time between the start of infectiousness of one individual and that of a second individual infected by the first.

When allowing $p, q > 1$, X_t can be seen as the number of individuals becoming *newly* infectious at time t. These are contagious over p time steps, with $(\kappa_1, \ldots, \kappa_p)$ the infectivity profile. In the exposed pool, individuals can "move forward" up to q time periods at once, leading to more complex latent period distributions; see Figure 1, middle panel.

When adding the compounding step $\theta *$ from Equation (18) to (21), E_t and ε_t can be thought of as clusters of exposed individuals, each containing a $G(\theta)$-distributed number of members (see Remark 2). All members of a cluster turn infectious simultaneously. The effective reproductive number then becomes $R_e = \kappa\theta$; see Figure 1, bottom panel.

The above mechanisms resemble classic epidemic models like the SEIR (susceptible-exposed-infectious-removed), with the difference that immunity due to infection is ignored. See Bauer and Wakefield (2018) for a similar argument on the INARCH(1) and Bjørnstad et al. (2002) for a related model accounting for immunity. Ignoring immunity due to infection is appropriate, for example, for vaccine-preventable diseases with high, but not complete vaccination levels in a population (De Serres et al., 2000). In this situation, only minor outbreaks seeded by imported cases occur, which do not meaningfully reduce the number of remaining susceptibles. We will return to such a setting in our case study in Section 7.

3.3 | GINARMA forms of GINAR and INGARCH models

We now describe how the various models laid out in Section 2.3 fit into the GINARMA framework. As one would expect, the GINAR(p) is just a GINARMA($p, 0$) model, further simplified by the omission of the compounding step $\theta *$. As in this case (brushing over the abuse of notation) we have $1 - \sum_{j=1}^{0} \beta_j = 1$, the model (18) and (19) collapses to

$$X_t = E_t + \varepsilon_t, \quad E_t = \sum_{i=1}^{p} \kappa_i \bullet X_{t-i}. \tag{22}$$

The term E_t hence becomes a simple placeholder in the GINAR(p) structure.

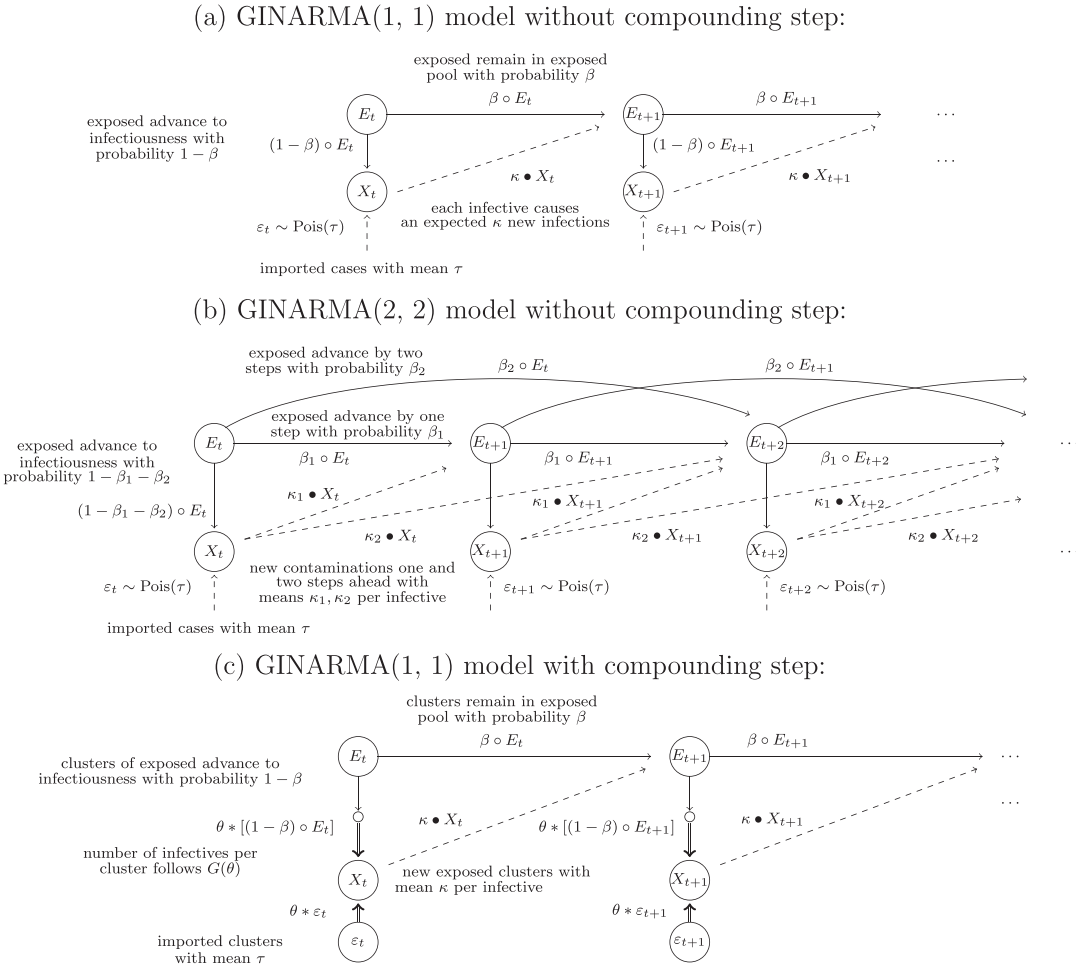

FIGURE 1 Interpretation of GINARMA models as stochastic epidemic processes. (a) GINARMA(1, 1) without a compounding step, see Equation (21). (b) GINARMA(2, 2) without a compounding step. (c) Full GINARMA(1, 1) model including compounding step. Solid lines represent multinomial thinning; dashed lines are generalized thinning or innovations, and double lines are thinning with a cluster distribution $G(\theta)$. In the bottom panel, the small circles represent the intermediate step $(1-\beta) \circ E_t$, which is then subject to compounding.

The link to INGARCH-type models is somewhat more intricate (see Supporting Information F.4 for the derivations). We develop the correspondence step by step, starting with the Poisson INGARCH(1, 1) model (6). If $\beta < 1$ it can be represented as

$$X_t = (1-\beta) \circ E_t + \varepsilon_t, \quad E_t = \beta \circ E_{t-1} + \kappa \star X_{t-1}, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{Pois}(\tau) \tag{23}$$

with $\tau = \nu/(1-\beta)$ and $\kappa = \alpha/(1-\beta)$. The two thinnings of E_t are coupled as in Equation (20) while $E_0 \sim \text{Pois}(\eta), \eta = (\lambda_0 - \tau)/(1-\beta)$. Note that while the equivalence even holds if $\alpha + \beta \geq 1$, that is, if the process is not stationary, we will not deal with this case in the following as the stochastic properties and inference would get considerably more involved. A technical condition (that also applies in the following) is $\lambda_0 \geq \tau$ so that $\eta \geq 0$. This, however, is natural as $\lambda_t \geq \tau$ also

holds for all $t \geq 1$ if $\lambda_0 \geq \tau$. The model structure corresponds to the top panel in Figure 1. We note that am construction of the Poisson INGARCH(1, 1) using a "cascade of thinning operations" was also introduced in Ferland et al. (2006), but is considerably more involved than our representation.

Remark 4. In epidemic modeling, Poisson offspring as in Equation (23) are widely used (Farrington & Grant, 1999; Kucharski et al., 2014). They arise, for example, when approximating a Reed–Frost chain binomial model for a large population (Bauer & Wakefield, 2018).

Formulation (23) can be extended to the Poisson INGARCH(p, q) case by setting

$$X_t = \left(1 - \sum_{j=1}^{q} \beta_j\right) \circ E_t + \varepsilon_t, \qquad E_t = \sum_{j=1}^{q} \beta_j \circ E_{t-j} + \sum_{i=1}^{p} \kappa_i \star X_{t-i} \tag{24}$$

with $\tau = \nu/(1 - \sum_{j=1}^{q} \beta_j)$, $\kappa_i = \alpha_i/(1 - \sum_{j=1}^{q} \beta_j)$. For initialization we need to set $E_m \overset{\text{ind.}}{\sim} \text{Pois}(\eta_m)$ with $\eta_m = (\lambda_m - \tau)/(1 - \sum_{j=1}^{q} \beta_j)$, again assuming $\lambda_m \geq \tau, m = 1 - q, \ldots, 0$ and $\sum_{j=1}^{q} \beta_j < 1$. This corresponds to the structure shown in the middle row of Figure 1.

A CP-INGARCH(1, 1) model as in Equations (12) to (14) is obtained by extending Equation (23) to

$$X_t = \theta * [(1 - \beta) \circ E_t + \varepsilon_t], \tag{25}$$

where $*$ denotes thinning with the clustering distribution G as in Equation (7). Here, we need to set $\tau = (\nu/\theta)/(1 - \beta), \kappa = (\alpha/\theta)/(1 - \beta)$ and $\eta = (\lambda_0/\theta - \tau)/(1 - \beta)$, assuming $\lambda_0/\theta \geq \tau$. This extension corresponds to the bottom panel of Figure 1.

3.4 | Stochastic properties

For the case $p = q = 1$, various properties of the GINARMA model can be obtained by noting that the process $\{E_t\}$ is a Galton-Watson branching process with immigration.

Lemma 1. *The process* $\{E_t\}$ *from Equation* (19) *with* $p = q = 1$ *can be expressed as*

$$E_t = \sum_{k=1}^{E_{t-1}} B_{t-1,k} + \varepsilon_t^*. \tag{26}$$

Here we set $\varepsilon_t^* = \kappa \bullet (\theta * \varepsilon_{t-1})$ *and independently for each* $k = 1, \ldots, E_{t-1}$

$$B_{t-1,k} = \begin{cases} 1 & \text{with probability } \beta \\ \kappa \bullet (\theta * 1) & \text{with probability } 1 - \beta. \end{cases} \tag{27}$$

Corollary 1. *While the process* $\{X_t\}$ *lacks a Markov property, both* $\{E_t\}$ *and the joint process* $\{(X_t, E_t)\}$ *are first-order Markov processes.*

Corollary 2. *If* $\kappa\theta < 1$, *the Markov chain* $\{E_t\}$ *is aperiodic and irreducible.*

As we typically initialize our processes at fixed values E_0, X_0, they are not actually stationary. However, under mild conditions, they have limiting distributions that are independent of the

initialization and are also obviously stationary distributions. Following Pakes (1971), we will refer to these distributions and associated properties as *limiting-stationary*.

Proposition 1. *For $p = q = 1$, the processes $\{E_t\}$ and $\{X_t\}$ from Equations (18) and (19) have unique limiting-stationary distributions if $\kappa\theta < 1$. The limiting-stationary moments are finite up to order r if the same is true for the moments of $\{\epsilon_t\}$, $\kappa \bullet 1$ and $\theta * 1$.*

Lemma 2. *Given they exist, the limiting-stationary means and variances of $\{E_t\}$ and $\{X_t\}$ in a GINARMA(1, 1) process are*

$$\mu_E = \frac{\kappa\tau\theta}{1-\beta-(1-\beta)\kappa\theta}, \quad \mu_X = \frac{\tau\theta}{1-\kappa\theta},$$
$$\sigma_E^2 = \frac{\sigma_\kappa^2\theta\tau + (\sigma_\theta^2\tau + \sigma_\tau^2\theta^2)\times\kappa^2 + \mu_E\times(1-\beta)\times\{\beta(1-\kappa\theta)^2 + \sigma_\kappa^2\theta + \sigma_\theta^2\kappa^2\}}{1-\{\beta+(1-\beta)\kappa\theta\}^2},$$
$$\sigma_X^2 = (1-\beta)\mu_E\sigma_\theta^2 + \theta^2(1-\beta)\{\beta\mu_E + (1-\beta)\sigma_E^2\} + \tau\sigma_\theta^2 + \sigma_\tau^2\theta^2.$$

The autocovariance functions of $\{E_t\}$ and $\{X_t\}$ are of AR(1) and ARMA(1, 1) type, as

$$\gamma_E(d) = \{\beta+(1-\beta)\kappa\theta\}^d \times \sigma_E^2$$
$$\gamma_X(d) = \{\beta+(1-\beta)\kappa\theta\}^{d-1}\times(1-\beta)\theta\times\{\theta\beta(1-\beta)(\sigma_E^2-\mu_E) + \kappa\sigma_X^2\}.$$

Combining arguments from Pakes (1971) and Meitz and Saikkonen (2008), it can be shown that $\{E_t\}$ is geometrically ergodic under mild conditions, which translates to $\{X_t\}$.

Proposition 2. *For $p = q = 1$, the joint process $\{(X_t, E_t)\}$ from Equations (18) and (19) is geometrically ergodic if $\kappa\theta < 1$ and $\sigma_\tau^2, \sigma_\kappa^2, \sigma_\theta^2 < \infty$. If the initial value E_0 is generated from its stationary distribution, the process is moreover β-mixing with geometrically decaying coefficients.*

Proposition (2) notably implies geometric ergodicity of the process $\{X_t\}$ in CP-INGARCH(1, 1) models, a topic that has received much attention (e.g., Fokianos et al., 2009; Gonçalves et al., 2015a; Davis et al., 2021 and references therein). Typically, the employed arguments are more sophisticated than what we use, the difficulty being that the state space of $\lambda_{t+1}|\lambda_t$ in Equation (6) depends on λ_t. We circumvent this via a fully discrete display involving $\{E_t\}$ rather than $\{\lambda_t\}$. A natural consequence of this approach, however, is that our arguments do not imply ergodicity of the joint process $\{\lambda_t, X_t\}$, as addressed, for example, by Fokianos et al. (2009).

For the general case $p, q > 1$ we can only make a few statements on the overarching model from Definition 1. As already discussed in Remark 3, both $\{E_t\}$ and $\{E_t\}$ are $\max(p, q)$-th order Markov processes. It is easily shown that for $\sum_{i=1}^p \kappa_i < 1$ a limiting-stationary mean exists and is

$$\mu_X = \frac{\tau\theta}{1-\theta\times\sum_{i=1}^p\kappa_i}, \quad \mu_E = \mu_X \times \frac{\sum_{i=1}^p\kappa_i}{1-\sum_{j=1}^q\beta_j}.$$

Stationary variances and covariances can be derived for specific instances of the class, but we defer these to Section 4.3.

4 | EXTENDING THE INAR CLASS

4.1 | A new INARMA(p, q) model

Despite the parallels between the Poisson INAR(1) and INARCH(1) models seen in Section 2.1, it is not obvious how the INGARCH(p, q) recursion (14) could be transposed to the INAR case. In the thinning-based representation (24), however, we can simply swap all Poisson thinnings for multinomial thinnings. This leads to a new extension of the INAR(p) model (9) and (10), which as we shall see has attractive stochastic properties. Omitting the compounding step $\theta *$ from Definition (1), we define our INARMA(p, q) process as

$$X_t = \left(1 - \sum_{j=1}^{q} \beta_j\right) \circ E_t + \varepsilon_t \tag{28}$$

$$E_t = \sum_{j=1}^{q} \beta_j \circ E_{t-j} + \sum_{i=1}^{p} \kappa_i \circ X_{t-i}. \tag{29}$$

In addition to the constraints from Definition 1, we assume that $\sum_{i=1}^{p} \kappa_i < 1$ and set

$$(\kappa_1 \circ X_t, \ldots, \kappa_p \circ X_t) \mid X_t \sim \text{Mult}\big(X_t; \kappa_1, \ldots, \kappa_p\big), \tag{30}$$

thus paralleling Equation (10). The innovations $\{\varepsilon_t\}$, thinnings of E_t and the initialization with $X_{1-p}, \ldots, X_0, E_{1-q}, \ldots, E_0$ are handled as in Definition 1.

Remark 5. In terms of the interpretation from Section 3.2, an infected can cause at most one new infection in the INAR/INARMA model. This case is occasionally studied in theory, but corresponds to an unusual practical setting. As argued by Farrington and Grant (1999), the disease would need to be of very modest infectivity (low κ_i), or infectives would need to be isolated systematically after a first event of onward transmission.

4.2 | Properties of the INARMA(1, 1) model

4.2.1 | General innovation distributions

Many statements from Section 3.4 simplify considerably for the INARMA(1, 1) model.

Lemma 3. *As we assume $\kappa < 1$, the limiting stationary mean, variance, and autocorrelation function of an INARMA*(1, 1) *process $\{X_t\}$ are given by*

$$\mu_X = \frac{\tau}{1-\kappa} \tag{31}$$

$$\sigma_X^2 = \frac{\kappa(1+\beta)}{1+\xi} \times \frac{\tau}{1-\kappa} + \left(1 - \frac{\kappa(1+\beta)}{1+\xi}\right) \times \frac{\sigma_\tau^2}{1-\kappa} \tag{32}$$

$$\gamma_X(d) = (1-\beta)\kappa\xi^{d-1} \times \left(1 + \frac{\kappa\beta(\sigma_\tau^2 - \tau)}{(1+\beta)\{(1-\kappa)\sigma_\tau^2 + \kappa\tau\} + (1-\beta)\kappa\sigma_\tau^2}\right) \times \sigma_X^2. \tag{33}$$

Here, we use the shorthand

$$\xi = \gamma_X(2)/\gamma_X(1) = \beta + (1-\beta)\kappa. \tag{34}$$

If relative to a Poisson, the innovation distribution is overdispersed ($\sigma_\tau^2 > \tau$) or underdispersed ($\sigma_\tau^2 < \tau$), respectively, the same thus holds for the marginal distribution of $\{X_t\}$. Moreover, for overdispersed (underdispersed) innovations, the autocorrelations will be stronger (weaker) than for identical τ, β, κ and equidispersed innovations. It is easily shown that $\gamma_X(2) \geq \gamma_X(1)^2$ always holds in Equation (33), with equality for $\beta = 0$. Like the INGARCH(1, 1), our INARMA(1, 1) thus has a "longer memory" than an INAR(1), and there is no instance with an MA(1) structure. These aspects differ from the $\text{INARMA}_{\text{DGL}}(1,1)$, that is, model (11) with binomial thinning, which generally displays marginal overdispersion even for equidispersed innovations, implies $\gamma_X(2) \geq \gamma_X(1)^2$ and contains the INMA(1) as a special case.

Remark 6. In the INARMA(1, 1), $\{E_t\}$ is an INAR(1) process, while $\{X_t\}$ is an INAR(∞) with geometrically decaying autoregressive parameters; see Supporting Information Remark S1. Alternatively, the INARMA(1, 1) has an MA(∞) representation akin to arguments from Pei et al. (2024), see Supporting Information Remark S2.

4.2.2 | Poisson innovations

The Poisson INARMA(1, 1) model

$$X_t = (1-\beta)\circ E_t + \varepsilon_t, \quad E_t = \beta \circ E_{t-1} + \kappa \circ X_{t-1}, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{Pois}(\tau) \tag{35}$$

has been discussed in Bracher (2019), but for completeness some results are repeated and extended.

Lemma 4. *If* $\{X_t\}$ *is a Poisson INARMA*(1,1) *process with* $E_0 \sim \text{Pois}[\kappa\tau/(1-\xi)]$, *then the process is strictly stationary with Poisson marginals. Expressions* (31)–(33) *simplify to*

$$\mu_X = \sigma_X^2 = \frac{\tau}{1-\kappa}; \quad \rho_X(d) = (1-\beta)\kappa\xi^{d-1}. \tag{36}$$

Moreover, for $t \in \mathbb{N}$ *and* $d > 0$ *it holds that*

$$(X_t, X_{t+d}) \sim \text{BPois}\{\rho_X(d)\mu_X, [1-\rho_X(d)]\mu_X, [1-\rho_X(d)]\mu_X\}, \tag{37}$$

with BPois the bivariate Poisson distribution as defined in Johnson et al. (1997).

Further particularities of the Poisson INARMA(1, 1) process include that it is time-reversible and closed to binomial thinning. The latter means that if $\{X_t\}$ is a Poisson INARMA(1, 1), then so is $\{\tilde{X}_t\}$ with $\tilde{X}_t = \pi \circ X_t$ independently for each t; see Bracher (2019) and Supporting Information Remark S3.

Remark 7. In terms of the thinning-based formulation (23), the limiting-stationary second-order properties of the Poisson INGARCH(1, 1) model are $\mu_X = \tau/(1-\kappa)$ and

$$\sigma_X^2 = \left(1 + \frac{(1-\beta)^2\kappa^2}{1-\xi^2}\right) \times \mu_X, \quad \rho_X(d) = \left(1 + \frac{\kappa\beta(1-\beta)}{1-\xi^2+\kappa^2(1-\beta)^2}\right) \times (1-\beta)\kappa\xi^{d-1}.$$

Poisson rather than binomial offspring thus lead to higher dispersion and stronger autocorrelations than in the Poisson INARMA(1, 1), but the ACFs are proportional.

4.2.3 | Compound Poisson innovations

As the class of CP distributions is closed to binomial thinning and summation, the marginal distributions of INAR(1) models with CP innovations are CP (Schweer & Weiß, 2014). Remark 6 thus implies that in the CP-INARMA(1, 1) process, $\{E_t\}$ has CP marginals, and it is easy to show that X_t inherits this property. The order of the CP distribution, too, will be inherited (Weiß & Puig, 2015), meaning that Hermite innovations lead to Hermite marginals; see Supporting Information Remark S4.

A structural difference between CP-INARMA and CP-INGARCH models is that the latter features a compounding step acting on the Poisson innovations and offspring, see Equation (25). In CP-INARMA models, on the other hand, no compounding step is used, but the innovation distribution becomes a CP. Overdispersion thus enters purely via innovations.

4.3 | Properties of the INARMA(p, q) model

For $p, q > 1$, the stationary moments can still be derived, but take somewhat more involved forms. We start by establishing expressions for the Poisson case, which we can subsequently use as building blocks for the more general case.

Proposition 3. *The limiting-stationary mean of a Poisson INARMA(p, q) process $\{X_t\}$ is*

$$\mu_X = \sigma_X^2 = \frac{\tau}{1 - \sum_{i=1}^{p} \kappa_i} \tag{38}$$

while the autocorrelation function can be computed recursively via

$$\rho(0) = 1 \quad \text{and} \quad \rho_X(d) = \left(1 - \sum_{j=1}^{q} \beta_j\right) \times \left(\sum_{i=1}^{d} \rho_X(d-i) \times s_i\right),$$
$$\text{where} \quad s_i = \sum_{k=1}^{\min\{i,p\}} \kappa_k \pi_{i-k} \quad \text{for } i = 1, 2, \ldots$$
$$\pi_0 = 1, \quad \pi_k = \sum_{l=1}^{\min\{k,q\}} \beta_l \pi_{k-l} \quad \text{for } k = 1, 2, \ldots \tag{39}$$

Moreover, the bivariate Poisson property (37) from Lemma 4 still holds.

Corollary 3. *In the Poisson INARMA(p, q) model, the process $\{E_t\}$ is a Poisson INAR$(\max\{p, q\})$ process as defined in Equations (9) and (10); see Lemma S1 in the Supporting Information for details.*

Corollary 4. *In the INARMA(p, q) model with general innovation distribution, the stationary mean is still given by expression (38). The stationary variance is given by*

$$\sigma_X^2 = \mu_X + (\sigma_\tau^2 - \tau) \times \sum_{i=0}^{\infty} \rho_X^{\text{Po}}(i)^2,$$

where τ and σ_τ^2 are the innovation mean and variance, while ρ_X^{Po} is the autocorrelation function of the Poisson INARMA(p, q) model from Equation (39). The autocovariance function can be expressed as

$$\gamma_X(d) = \mu_X \times \rho_X^{\text{Po}}(d) + (\sigma_\tau^2 - \tau) \times \left[\sum_{i=0}^{\infty} \rho_X^{\text{Po}}(i) \times \rho_X^{\text{Po}}(i+d) \right].$$

As for $p = q = 1$ (Lemma 2), overdispersed innovations with $\sigma_\tau^2 > \tau$ are thus a necessary and sufficient condition for marginal overdispersion $\sigma_X^2 > \mu_X$. Moreover, it can be shown that with the same parameters τ, β, κ, overdispersed innovations again lead to stronger autocorrelations compared to the Poisson case.

Remark 8. Following arguments from Alzaid and Al-Osh (1990), it can be shown that the ACF of the INARMA(p, q) model with general innovation distribution bears resemblance with, but is not identical to that of a Gaussian ARMA$(\max\{p, q\}, \max\{p, q\})$ model; see Supporting Information Remark S5.

Remark 9. The Poisson INGARCH(p, q) model, which as described in Section 2.1 is a GINARMA(p, q) model with Poisson offspring, likewise has closed-form stationary moments (Weiß, 2018, section 4.1). Unlike for $p = q = 1$ (Remark 7), however, we are not aware of any obvious parallels between the expressions for INARMA(p, q) and INGARCH(p, q).

5 | INFERENCE

5.1 | Maximum likelihood inference in the case $p = q = 1$

For GINARMA models with Poisson innovation and offspring distributions, INGARCH-type representations as in Section 3.3 exist, making likelihood evaluation straightforward. Whenever no such display is available, as is the case for all INARMA models, computing the likelihood gets tedious. This is because unless $q = 0$, no useful Markov properties can be exploited (see Remark 3). In the following, we present an algorithm that can be applied to GINARMA(1, 1) models with arbitrary innovation and offspring distributions. For simplicity, we omit the compounding step from Equation (18), but accommodating it would be straightforward. The procedure represents an adaptation of the forward algorithm (Zucchini & MacDonald, 2009) and resembles an existing approach for the INARMA$_{\text{DGL}}(1, 1)$ model (Weiß et al., 2019). To facilitate notation we introduce the shorthand $A_t = (1 - \beta) \circ E_t$ which enables us to write

$$X_t = \underbrace{A_t}_{(1-\beta)\circ E_t} + \varepsilon_t, \qquad E_t = \underbrace{E_{t-1} - A_{t-1}}_{\beta \circ E_{t-1}} + \kappa \bullet X_{t-1}.$$

The substitution $E_{t-1} - A_{t-1} = \beta \circ E_{t-1}$ results from point (iii) in Definition 1 which implies

$$(\beta \circ E_{t-1}, (1 - \beta) \circ E_{t-1}) \,\big|\, E_{t-1} \sim \text{Mult}\left[E_{t-1}; \beta, (1 - \beta)\right].$$

Now denote the sequence of observed values by $\{x_1, \ldots, x_T\}$, with T the length of the time series. As a first step, a sufficiently large support $\mathcal{E} = \{0, 1, \ldots, M\}$ needs to be chosen for $\{X_t\}$ and

$\{E_t\}$. In practice, we set M to the maximum of $1.2 \times \max(x_1,\ldots,x_t)$ and the 0.999 quantiles of the stationary distributions of E_t and X_t under the respective parameters. As $E_t \geq A_t$, $\mathcal{E}$ implies a support $\mathcal{E}^* = \{(0,0),(1,0),(1,1),(2,0),\ldots,(M,M-1),(M,M)\}$ for the tuple (E_t, A_t). Moreover we introduce the following shorthands: $\Pr(Y = y \mid X_{<t})$ is the probability that $Y = y$ provided that $X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2},\ldots,X_1 = x_1$; for $t = 1$ this corresponds to the marginal distribution of Y under some suitable initialization. $\Pr(Y = y \mid X_{\leq t})$ is defined analogously, but $X_t = x_t$ is also included in the condition.

Algorithm 1. *The algorithm is initialized by setting* $E_1 \sim \text{Pois}(\eta)$*, with* η *treated like an extra parameter. Then the following steps are iterated for* $t = 1,\ldots,T$.

1. *For each tuple* $(e_t, a_t) \in \mathcal{E}^*$ *compute*

$$\Pr(E_t = e_t, A_t = a_t \mid X_{<t}) = \Pr(A_t = a_t \mid E_t = e_t) \times \Pr(E_t = e_t \mid X_{<t}).$$

2. *Compute and store*

$$\begin{aligned}\Pr(X_t = x_t \mid X_{<t}) &= \sum_{(e_t,a_t)\in\mathcal{E}^*} \Pr(X_t = x_t \mid E_t = e_t, A_t = a_t) \\ &\quad \times \Pr(E_t = e_t, A_t = a_t \mid X_{<t}) \\ &= \sum_{(e_t,a_t)\in\mathcal{E}^*} \Pr(\varepsilon_t = x_t - a_t) \times \Pr(E_t = e_t, A_t = a_t \mid X_{<t}).\end{aligned}$$

3. *For* $(e_t, a_t) \in \mathcal{E}^*$ *compute*

$$\begin{aligned}\Pr(E_t = e_t, A_t = a_t \mid X_{\leq t}) &= \frac{\Pr(E_t = e_t, A_t = a_t, X_t = x_t \mid X_{<t})}{\Pr(X_t = x_t \mid X_{<t})} \\ &= \frac{\Pr(E_t = e_t, A_t = a_t \mid X_{<t}) \times \Pr(\varepsilon_t = x_t - a_t)}{\Pr(X_t = x_t \mid X_{<t})}.\end{aligned}$$

4. *For* $l_t \in \mathcal{E}$ *compute*

$$\Pr(E_t - A_t = l_t \mid X_{\leq t}) = \sum_{(e_t,a_t)\in\mathcal{E}^* : e_t - a_t = l_t} \Pr(E_t = e_t, A_t = a_t \mid X_{\leq t}).$$

5. *For* $e_{t+1} \in \mathcal{E}$ *compute*

$$\begin{aligned}\Pr(E_{t+1} = e_{t+1} \mid X_{\leq t}) &= \sum_{l_t\in\mathcal{E}} \Pr(E_{t+1} = e_{t+1} \mid E_t - A_t = l_t, X_{\leq t}) \\ &\quad \times \Pr(E_t - A_t = l_t \mid X_{\leq t}) \\ &= \sum_{l_t\in\mathcal{E}} \Pr(\kappa \circ x_t = e_{t+1} - l_t) \times \Pr(E_t - A_t = l_t \mid X_{\leq t}).\end{aligned}$$

The values stored in Step 2 of each iteration serve to evaluate the (conditional) likelihood of the observed time series as

$$\Pr(X_1 = x_1,\ldots,X_t = x_T) = \Pr(X_1 = x_1) \times \prod_{t=2}^{T} \Pr(X_t = x_t \mid X_{<t}).$$

Maximization of the log-likelihood is done using the Nelder–Mead method as implemented in the `R` function `optim`. Whenever available we use moment estimators (see next section) to initialize the optimization. All parameters are handled on suitable transformed scales allowing for unconstrained optimization. For parameters constrained to the unit interval, we use logit transformations, for parameters that can take any positive value we use the natural logarithm. Standard errors are estimated via the inverse observed Fisher information (obtained by numerical differentiation) with subsequent application of the delta method. Fitted values and Pearson residuals can be obtained using the probabilities $\Pr(X_t = x_t \mid X_{<t}), x_t \in \mathcal{E}$. We note that conceptually the algorithm could be extended to models with $p, q > 1$, but the size of the matrices and arrays needed to keep track of all hidden states would quickly become too demanding in practice.

The likelihood function can be evaluated with arbitrary precision, but is not available in closed form. Establishing consistency or asymptotic normality of the estimators is thus not straightforward (a typical proof strategy relying on threefold continuous differentiability of the log-likelihood function, Fokianos et al., 2009). Indeed, results on the asymptotics of maximum likelihood estimators seem to be lacking even for INAR models with general innovation distributions.

5.2 | Moment-based estimation

The suggested likelihood evaluation method can get slow even for moderately high count values. As a computationally fast alternative in the case $p = q = 1$ we consider moment-based estimators, see Supporting Information C.1. For Poisson innovations, consistency and asymptotic normality of the estimators can be established. For general innovation distributions, they do not have a closed form, but can be evaluated by solving a cubic equation numerically.

For $p, q > 1$, the moment-based approach lacks stability and it becomes challenging to handle cases where no exact solution exists. A potential solution is the generalized method of moments (GMM) as described for a related setting by Pei et al. (2024). In GMM, a larger number of moment conditions can be used, and cases, where the empirical moments do not match those resulting from any combination of parameters exactly, are handled more naturally. Relevant moment conditions might be obtained along the lines of Lu (2021). Due to its technical complexity, however, we defer this approach to future work.

5.3 | Gaussian quasi-likelihood inference

Lastly, we consider a Gaussian quasi-likelihood approach, which extends more easily to $p, q > 1$ than the previously discussed options. It is similar to approaches previously explored, for example, by Cui and Lund (2009) and Armillotta and Gorgi (2024). To evaluate the quasi-likelihood of a (generalized) INARMA(p, q) model with a given set of parameter values, we match it to a Gaussian AR(p^*) model with identical stationary mean, variance and autocovariance up to lag p^* (in practice we use $p^* = 2 \times$ (number of fitted parameters + 1)). The quasi-likelihood for the parameters in question is then given by the (conditional) likelihood of said Gaussian AR(p^*) model (note that we lose $p^* - 1$ likelihood contributions due to the use of p^* lags). Optimization of the quasi-likelihood is done numerically, and standard errors can be obtained via the inverse of the quasi-Fisher information matrix.

5.4 | Likelihood ratio testing of nested models

A benefit of an overarching class like GINARMA is that the nesting of different instances can be used for testing purposes. We address this aspect in two examples centered around the Poisson INARMA(1, 1) from Equation (35). Firstly, we address the pair of hypotheses

$$\mathrm{H}_0 : \beta = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \beta > 0, \tag{40}$$

that is, test whether the Poisson INAR(1) is sufficient or needs to be extended to the INARMA(1, 1) version. To this end, we apply a likelihood ratio test statistic given by

$$\Lambda = -2 \log \left[\frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)} \right] \tag{41}$$

where $\theta = (\tau, \beta, \kappa)$ is the parameter vector, Θ is the unconstrained parameter space and Θ_0 is the parameter space under H_0, that is, under the constraint $\beta = 0$. As the null hypothesis is on the boundary of the parameter space, the asymptotic distribution of Λ is not a $\chi^2(1)$ distribution as general theory would suggest. While the lack of closed-form likelihood function prevents us from establishing a formal result on the distribution of Λ, we can follow the heuristic argument from Molenberghs and Verbeke (2007) which suggests an equal mixture of the $\chi^2(1)$ distribution and a point mass at zero.

Secondly, we may be interested in choosing between the Poisson INARMA(1, 1) and INGARCH(1, 1) models. To this end we nest both into an *extended* Poisson INARMA(1, 1) model, which borrowing ideas from Weiß (2015) we define as

$$X_t = (1 - \beta) \circ E_t + \varepsilon_t, \qquad E_t = \beta \circ E_{t-1} + \kappa \circledast_\zeta X_{t-1}, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{Pois}(\tau). \tag{42}$$

Here, the thinning operator $\circledast_\zeta$ is defined as

$$\alpha \circledast_\zeta N = [\alpha \times (1 - \zeta)] \circ N + (\alpha\zeta) \star N. \tag{43}$$

For $\zeta = 0$, the operator $\circledast_\zeta$ reduces to binomial thinning, leading to a regular INARMA model. For $\zeta = 1$ we obtain Poisson thinning and thus an INGARCH model. We focus on testing the appropriateness of the Poisson INARMA(1, 1), that is, we consider

$$\mathrm{H}_0 : \zeta = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \zeta > 0. \tag{44}$$

The likelihood ratio test statistic is defined in analogy to (41) and can be assumed to asymptotically follow the same mixture distribution.

Our simulation studies in Section 6.3 indicate that the assumed mixture distribution works well for the testing problem (40). For problem (44), however, the resulting test is rather conservative. This parallels evidence from Vu and Maller (1996) who find a similar bias towards the binomial rather than Poisson distribution in an i.i.d testing setting. Unfortunately, their suggestion to obtain the reference distribution by simulation is not computationally feasible in our case. We therefore suggest a heuristic approach, employing a computationally efficient simulation step to adjust the proposed mixture distribution in a parametric manner; see details in Supporting Information Section C.2.

6 | SIMULATION STUDY

6.1 | Parameter estimation in the INARMA(1, 1) model

To assess our estimators in the case $p = q = 1$ we specify three simulation scenarios:

1. Scenario 1: $\tau = 1, \beta = 0.5, \kappa = 0.5$. In the Poisson case this implies $\mu_X = \sigma_X^2 = 2, \rho_X(d) = 0.25 \times 0.75^{d-1}$. In the negative binomial / Hermite cases we set $\psi = 0.5$, resulting in $\mu_X = 2; \sigma_X^2 = 2.57; \rho_X(d) = 0.26 \times 0.75^{d-1}$.
2. Scenario 2: $\tau = 1, \beta = 0.2, \kappa = 0.6$. In the Poisson case this implies $\mu_X = \sigma_X^2 = 2.5; \rho_X(d) = 0.48 \times 0.68^{d-1}$. In the negative binomial / Hermite cases we set $\psi = 0.7$, resulting in $\mu_X = 2.5; \sigma_X^2 = 3.5; \rho_X(d) = 0.5 \times 0.68^{d-1}$.
3. Scenario 3: $\tau = 1, \beta = 0.1, \kappa = 0.8$. In the Poisson case this implies $\mu_X = \sigma_X^2 = 5; \rho_X(d) = 0.72 \times 0.82^{d-1}$. In the negative binomial / Hermite cases we set $\psi = 0.9$, resulting in $\mu_X = 5; \sigma_X^2 = 7.32; \rho_X(d) = 0.74 \times 0.82^{d-1}$.

As we chose $\tau = 1$ in all settings we get the same second-order properties for the same values of ψ in the Hermite and the negative binomial versions. Note however that this is not generally the case. We simulated 1000 time series for each scenario and different lengths of time series $T \in \{250, 500, 1000\}$. The results for maximum-likelihood estimation are presented in Table 1. Results for moment-based and Gaussian quasi-likelihood estimators are available in Supporting Information Tables S1 and S2, respectively.

All three fitting procedures yield approximately unbiased estimates for τ, β, and κ, with some small-sample biases. The dispersion parameters ψ are subject to some biases, and their estimation can become unstable if $\hat{\tau}$ is small. The maximum-likelihood estimators have smaller standard errors than the two other methods. Both in the maximum likelihood and the Gaussian quasi-likelihood schemes, the estimated standard errors are mostly in good agreement with the observed standard errors, but in some instances biased downwards.

6.2 | Parameter estimation in the INARMA(2, 1) model

As stated in Section 5.3, the Gaussian quasi-likelihood approach extends to higher-order models. We illustrate this with two additional scenarios on the INARMA(2, 1) model:

1. Scenario 4: $\tau = 2, \beta = 0.15, \kappa_1 = 0.2, \kappa_2 = 0.6$. In the Poisson case, this implies $\mu_X = \sigma_X^2 = 10$, while the first five values of the autocorrelation function are $0.17, 0.56, 0.27, 0.37, 0.26$. In the negative binomial and Hermite cases we set $\psi = 0.7$.
2. Scenario 5: $\tau = 1.5, \beta = 0.35, \kappa_1 = 0.45, \kappa_2 = 0.25$. In the Poisson case, this implies $\mu_X = \sigma_X^2 = 5$, while the first five values of the autocorrelation function are $0.29, 0.35, 0.27, 0.23, 0.19$. In the negative binomial and Hermite cases we set $\psi = 0.5$.

Table 2 summarizes the results. For length $T = 250$, some noteworthy biases occur; see, for example, the estimates $\hat{\tau}$ and $\hat{\kappa}_1$ in the Hermite and negative binomial cases. These, however, decrease for longer time series and become negligible for $T = 2000$. The average estimated standard errors are overall well-aligned with the observed standard errors.

TABLE 1 Simulation results for maximum likelihood estimators in the Poisson, Hermite, and negative binomial settings, Scenarios 1–3, $T \in \{250, 500, 1000\}$, 1000 runs.

Poisson

	τ				ψ				β				κ			
T	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean of $\widehat{\text{se}}$	se	Mean	True	Mean	se	Mean of $\widehat{\text{se}}$
250	1.000	1.044	0.266	0.245	—	—	—	—	0.500	0.449	0.167	0.130	0.500	0.477	0.132	0.111
500		1.025	0.186	0.179		—	—	—		0.476	0.113	0.098		0.486	0.091	0.084
1000		1.015	0.131	0.128		—	—	—		0.486	0.073	0.070		0.493	0.064	0.061
250	1.000	1.008	0.186	0.185	—	—	—	—	0.200	0.190	0.091	0.089	0.600	0.596	0.072	0.071
500		1.011	0.132	0.132		—	—	—		0.196	0.066	0.065		0.595	0.052	0.050
1000		1.005	0.092	0.092		—	—	—		0.198	0.048	0.046		0.598	0.035	0.035
250	1.000	1.009	0.182	0.184	—	—	—	—	0.100	0.097	0.039	0.040	0.800	0.797	0.036	0.035
500		1.007	0.123	0.128		—	—	—		0.099	0.028	0.028		0.798	0.024	0.025
1000		1.010	0.089	0.090		—	—	—		0.098	0.020	0.020		0.798	0.017	0.018

Hermite

	τ				ψ				β				κ			
T	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$
250	1.000	1.045	0.267	0.217	0.500	0.483	0.225	0.170	0.500	0.465	0.171	0.123	0.500	0.477	0.127	0.094
500		1.017	0.187	0.157		0.494	0.171	0.132		0.480	0.112	0.092		0.491	0.092	0.071
1000		1.022	0.129	0.122		0.488	0.113	0.105		0.491	0.075	0.069		0.489	0.062	0.057
250	1.000	1.032	0.189	0.178	0.700	0.656	0.195	0.160	0.200	0.191	0.092	0.086	0.600	0.586	0.069	0.062
500		1.019	0.125	0.123		0.679	0.128	0.117		0.193	0.063	0.063		0.591	0.046	0.044
1000		1.009	0.085	0.085		0.690	0.088	0.083		0.196	0.044	0.044		0.596	0.031	0.031
250	1.000	1.051	0.201	0.198	0.900	0.814	0.224	0.154	0.100	0.095	0.041	0.040	0.800	0.789	0.036	0.033
500		1.019	0.127	0.127		0.863	0.145	0.110		0.098	0.027	0.028		0.796	0.023	0.022
1000		1.008	0.087	0.086		0.884	0.099	0.081		0.101	0.020	0.020		0.798	0.015	0.016

TABLE 1 (Continued)

Negative binomial																
	τ				ψ				β				κ			
T	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$
250	1.000	1.017	0.255	0.208	0.500	0.599	0.668	0.549	0.500	0.474	0.168	0.122	0.500	0.491	0.121	0.093
500		1.013	0.173	0.159		0.532	0.269	0.136		0.485	0.109	0.093		0.494	0.084	0.073
1000		1.007	0.118	0.116		0.515	0.186	0.092		0.491	0.072	0.068		0.497	0.058	0.055
250	1.000	1.019	0.177	0.179	0.700	0.713	0.379	0.309	0.200	0.191	0.091	0.088	0.600	0.590	0.067	0.064
500		1.012	0.122	0.125		0.708	0.258	0.192		0.196	0.063	0.064		0.595	0.047	0.045
1000		1.005	0.090	0.087		0.706	0.181	0.127		0.199	0.044	0.044		0.598	0.033	0.032
250	1.000	1.024	0.190	0.192	0.900	0.911	0.469	0.501	0.100	0.099	0.043	0.041	0.800	0.795	0.035	0.034
500		1.019	0.128	0.131		0.895	0.312	0.299		0.097	0.029	0.029		0.796	0.024	0.024
1000		1.007	0.092	0.090		0.901	0.221	0.204		0.099	0.021	0.020		0.798	0.017	0.016

Note: In 0.3% of runs, estimates and/or standard errors $\widehat{\text{se}}$ could not be evaluated due to numerical problems.

TABLE 2 Simulation results for the Gaussian quasi-likelihood estimators for Poisson, Hermite, and negative binomial INARMA(2, 1) models, Scenarios 4 and 5, $T \in \{250, 500, 1000, 2000\}$, 1000 runs.

Poisson INARMA(2, 1)

	τ				ψ				β				κ_1				κ_2			
T	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean of $\widehat{\text{se}}$	se	Mean	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$
250	2.000	2.1	0.619	0.626	—	—	—	—	0.150	0.149	0.090	0.092	0.200	0.187	0.117	0.117	0.600	0.602	0.088	0.085
500		2.1	0.450	0.445		—	—	—		0.153	0.067	0.068		0.181	0.086	0.086		0.608	0.061	0.061
1000		2.057	0.315	0.306		—	—	—		0.150	0.048	0.048		0.191	0.063	0.061		0.603	0.045	0.042
2000		2.016	0.211	0.212		—	—	—		0.151	0.033	0.034		0.197	0.043	0.043		0.601	0.031	0.030
250	1.500	1.589	0.499	0.467	—	—	—	—	0.350	0.329	0.157	0.152	0.450	0.439	0.155	0.150	0.250	0.241	0.108	0.104
500		1.54	0.324	0.318		—	—	—		0.340	0.112	0.109		0.444	0.105	0.105		0.247	0.077	0.074
1000		1.532	0.228	0.221		—	—	—		0.344	0.079	0.076		0.447	0.071	0.073		0.247	0.051	0.051
2000		1.514	0.156	0.154		—	—	—		0.348	0.053	0.053		0.448	0.050	0.051		0.249	0.036	0.036

Hermite INARMA(2, 1)

	τ				ψ				β				κ_1				κ_2			
T	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$	True	Mean	se	Mean of $\widehat{\text{se}}$
250	2.000	2.35	0.729	0.727	0.700	0.524	0.359	0.388	0.150	0.154	0.091	0.092	0.200	0.170	0.105	0.106	0.600	0.593	0.084	0.083
500		2.183	0.503	0.518		0.602	0.290	0.362		0.150	0.065	0.068		0.189	0.075	0.077		0.593	0.055	0.058
1000		2.107	0.335	0.367		0.626	0.228	0.288		0.152	0.048	0.048		0.190	0.054	0.055		0.599	0.042	0.041
2000		2.056	0.252	0.285		0.667	0.184	0.258		0.151	0.033	0.034		0.195	0.038	0.039		0.599	0.028	0.030
250	1.500	1.685	0.498	0.498	0.500	0.416	0.309	0.358	0.350	0.332	0.156	0.156	0.450	0.424	0.144	0.144	0.250	0.240	0.111	0.102
500		1.597	0.351	0.346		0.450	0.244	0.293		0.340	0.115	0.111		0.436	0.104	0.103		0.245	0.077	0.075
1000		1.55	0.245	0.240		0.478	0.194	0.221		0.345	0.083	0.078		0.445	0.072	0.072		0.245	0.053	0.053
2000		1.523	0.164	0.168		0.488	0.141	0.159		0.347	0.056	0.054		0.446	0.052	0.050		0.249	0.039	0.037

TABLE 2 (Continued)

Negative binomial INARMA(2, 1)

	τ				ψ				β				κ_1				κ_2			
T	True	Mean	se	Mean of $\widehat{se}$	True	Mean	se	Mean of $\widehat{se}$	True	Mean of $\widehat{se}$	se	Mean	True	Mean	se	Mean of $\widehat{se}$	True	Mean	se	Mean of $\widehat{se}$
250	2.000	2.318	0.766	0.737	0.700	0.669	0.607	0.651	0.150	0.153	0.092	0.094	0.200	0.176	0.100	0.097	0.600	0.592	0.077	0.081
500		2.158	0.521	0.498		0.704	0.470	0.466		0.148	0.066	0.068		0.188	0.069	0.070		0.596	0.055	0.056
1000		2.105	0.364	0.347		0.677	0.308	0.309		0.151	0.050	0.048		0.191	0.051	0.049		0.598	0.040	0.039
2000		2.052	0.247	0.241		0.693	0.250	0.225		0.151	0.034	0.034		0.195	0.035	0.034		0.600	0.028	0.027
250	1.500	1.645	0.485	0.491	0.500	0.516	0.494	0.543	0.350	0.328	0.163	0.153	0.450	0.436	0.147	0.145	0.250	0.236	0.116	0.105
500		1.579	0.344	0.337		0.507	0.331	0.356		0.334	0.117	0.112		0.440	0.105	0.101		0.244	0.080	0.075
1000		1.531	0.232	0.234		0.512	0.227	0.247		0.344	0.083	0.079		0.448	0.075	0.072		0.245	0.057	0.054
2000		1.517	0.169	0.165		0.507	0.160	0.171		0.348	0.057	0.055		0.450	0.052	0.050		0.247	0.038	0.037

Note: In 1.6% of runs, estimates and/or standard errors $\widehat{se}$ could not be evaluated due to numerical problems.

We note that we chose settings with slightly higher average counts here than in Scenarios 1–3. In exploratory analyses, we found that in very low-count settings like Scenario 1, higher-order lags were often not well identified.

6.3 | Likelihood ratio testing

Lastly, we present a simulation study on the proposed likelihood ratio tests. For the testing problem (40), that is, Poisson INAR(1) vs. Poisson INARMA(1, 1), we consider again Scenarios 1–3. In each of them, we generate 1000 time series of lengths $T = 250, 500, 1000$ to assess the power of the test. We then generate the same amount of data with β set to zero to assess the size. Table 3 summarizes the results for significance level $\alpha = 0.05$ (see Supporting Information Table S3 for $\alpha = 0.10$). The size of the tests is close to the nominal levels, and the power is overall high. Somewhat surprisingly, it is highest in Scenario 3, even though the true value of β is only 0.1. However, this is plausible in light of the results from Table 1, where the ratio of $\beta/\text{se}(\hat{\beta})$ is largest for Scenario 3. This is likely because the average counts are highest in Scenario 3, leading to less sparse and, hence, more informative data.

For the second testing problem (44), that is, Poisson INARMA(1, 1) vs. the extended version (42), we use Scenarios 1–3 to generate data under the null ($\zeta = 0$) and assess the size of the test. To evaluate the power, we generate data using $\zeta = 0.5$ and $\zeta = 1$, the latter corresponding to a Poisson INGARCH(1, 1). Results for significance level $\alpha = 0.05$ are shown in Table 4 (see Supporting Information Table S4 for $\alpha = 0.1$). When using the asymptotic distribution, that is, an equal mixture of a $\chi^2(1)$ and a point mass at zero, the test is quite conservative (columns labeled "Asym."). The simulation-based correction of the asymptotic distribution yields sizes closer to the nominal level (columns "Corr."), but in Scenario 1 the test remains conservative. In Scenarios 2 and 3 the test is well-behaved after the correction. The power is rather modest in Scenario 1, but

TABLE 3 Simulation results for the testing problem Poisson INAR(1) vs. Poisson INARMA(1, 1) as given in Equation (40).

		Rejection rates by true value of β			
Scenario	T	β	**Size**	β	**Power**
1	250	0.00	4.70	0.50	80.00
1	500		5.00		95.40
1	1000		5.00		99.80
2	250	0.00	4.40	0.20	70.00
2	500		4.30		90.50
2	1000		4.60		99.10
3	250	0.00	4.20	0.10	84.10
3	500		4.10		98.50
3	1000		5.20		100.00

Note: The null hypothesis corresponds to $\beta = 0$. Results are presented for Scenarios 1–3 with $T \in \{250, 500, 1000\}$ and based on 1000 runs. The size and power are reported as percentage values for significance level $\alpha = 5.0\%$. Results for $\alpha = 10.0\%$ are provided in Supporting Information Table S3.

TABLE 4 Simulation results for the testing problem (44), that is, Poisson INARMA(1, 1) vs. the extension (42). The null hypothesis corresponds to $\zeta = 0$.

		Rejection rates by true value of ζ								
		Size			Power					
Scenario	T	ζ	**Asym.**	**Corr.**	ζ	**Asym.**	**Corr.**	ζ	**Asym.**	**Corr.**
1	250	0.00	2.50	3.50	0.50	16.10	20.70	1.00	31.80	36.70
1	500		3.60	4.50		34.53	39.64		50.85	56.06
1	1000		2.70	3.40		60.50	65.30		74.80	77.10
2	250	0.00	3.20	4.50	0.50	71.20	75.90	1.00	83.70	85.40
2	500		3.80	4.90		93.70	94.60		95.00	95.20
2	1000		3.50	4.40		99.70	99.80		99.90	99.90
3	250	0.00	2.00	4.50	0.50	99.90	99.90	1.00	97.70	97.80
3	500		3.30	5.50		100.00	100.00		99.40	99.40
3	1000		3.60	5.50		100.00	100.00		99.90	99.90

Note: Results are presented for Scenarios 1–3 with $T \in \{250, 500, 1000\}$ and based on 1000 runs. The size and power are reported as percentage values for significance level $\alpha = 5.0\%$. We provide results based on the asymptotic distribution (columns "Asym.") and the corrected asymptotic distribution (columns "Corr."). Results for $\alpha = 10\%$ are provided in Supporting Information Table S4.

considerably higher in Scenarios 2 and 3. Similarly to parameter estimation, we thus observe that the test is more well-behaved when average counts are higher.

7 | APPLICATION: CHILDHOOD DISEASES IN BAVARIA

Lastly, we apply various instances of the introduced model class to two time series of infectious disease counts. We consider weekly numbers of reported measles and mumps cases in the German state of Bavaria, 2014–2019. Measles and mumps are vaccine-preventable childhood diseases and have become rare in Western Europe. While both diseases exhibit seasonal patterns in the absence of vaccination, during the considered period they only occurred sporadically. They thus match the setting described in Section 3.2 well, and are indeed commonly modeled using subcritical branching processes (Chiew et al., 2014; De Serres et al., 2000). The data, available from Robert Koch Institute (https://survstat.rki.de), are displayed in Figure 2. Both series exhibit slowly decaying autocorrelation functions and some degree of overdispersion.

Table 5 summarizes the fits of the INARCH(1), INGARCH(1, 1), INAR(1) and INARMA(1, 1) models. Each of them was applied in the Poisson, Hermite, and negative binomial versions. To make the results more easily comparable across the different models, we present them in terms of the epidemiological interpretation from Section 3.2; results for the original parameterizations are shown in Supporting Information Table S5. The mean generation times obtained from the INGARCH and INARMA models are in good agreement with commonly used estimates from the literature (slightly below 2 weeks for measles; 18 days for mumps; Bjørnstad et al., 2002; Vink et al., 2014). The estimated reproductive numbers here are around 0.7 and 0.6 for measles and mumps, respectively. While no comparable estimates for Germany exist, these values seem plausible in light of estimates from Australia, a country with somewhat higher vaccination coverage ($R_e = 0.47$ to 0.65 for measles depending on the exact method for 2009–2011; Chiew et al., 2014).

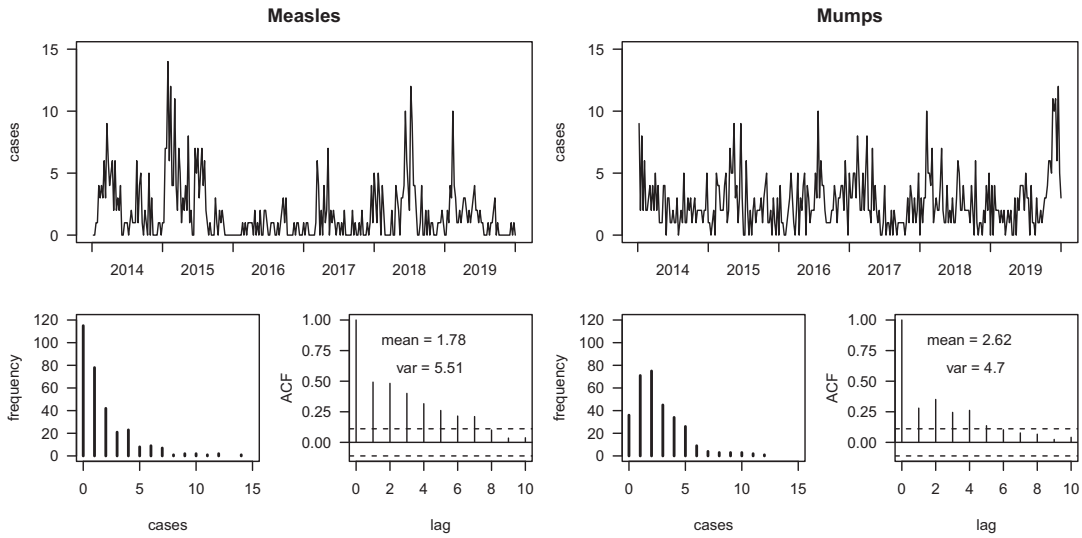

FIGURE 2 Top: Weekly counts of reported measles and mumps cases in Bavaria, 2014–2019. Bottom: Marginal distributions and autocorrelation functions.

TABLE 5 Model fits for measles and mumps. For better comparability, estimates are presented in terms of the epidemiological interpretation of the models.

	Measles					Mumps				
Model	**IC**	**R_e**	**GT**	**CS**	**AIC**	**IC**	**R_e**	**GT**	**CS**	**AIC**
Poisson INARCH	0.83	0.54	1*	1*	1159.13	1.93	0.26	1*	1*	1274.26
Hermite INARCH	0.85	0.52	1*	1.37	1082.40	1.95	0.25	1*	1.24	1249.33
NegBin INARCH	0.88	0.51	1*	1.51	1055.04	1.98	0.24	1*	1.24	1244.75
Poisson INGARCH	0.46	0.74	2.08	1*	1096.91	1.04	0.60	2.98	1*	1238.27
Hermite INGARCH	0.51	0.72	2.01	1.31	1046.09	1.07	0.58	2.98	1.18	1224.43
NegBin INGARCH	0.55	0.67	2.03	1.41	1028.23	1.11	0.57	2.96	1.19	1222.86
Poisson INAR	1.17	0.34	1*	1*	1232.94	2.12	0.18	1*	1*	1283.22
Hermite INAR	1.18	0.34	1*	1*	1122.68	2.07	0.20	1*	1*	1252.77
NegBin INAR	1.17	0.34	1*	1*	1068.77	2.08	0.20	1*	1*	1245.57
Poisson INARMA	0.72	0.60	2.00	1*	1166.26	1.38	0.47	2.50	1*	1257.34
Hermite INARMA	0.81	0.55	1.86	1*	1094.07	1.30	0.50	2.51	1*	1235.48
NegBin INARMA	0.81	0.53	1.81	1*	1046.65	1.41	0.46	2.46	1*	1231.73

Note: IC $= \tau\theta$ is the mean weekly number of imported cases; $R_e = \kappa\theta$ is the effective reproductive number; GT $= 1/(1-\beta)$ is the mean generation time in weeks; CS $= \theta$ is the mean cluster size. The definitions of IC, R_e, GT, CS refer to the general formulation (18) and (19) with $p = q = 1$. Entries with asterisks* are not estimated but implied by the model definition.

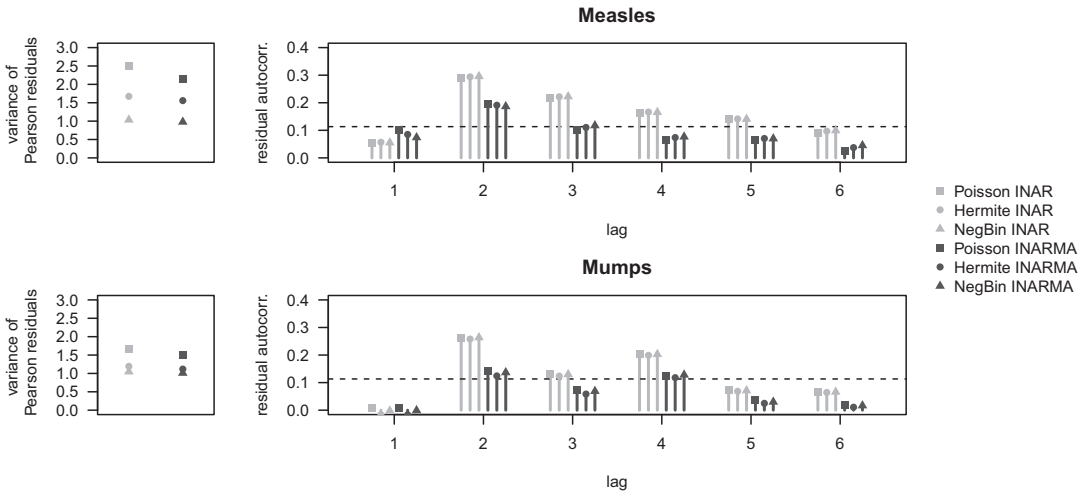

FIGURE 3 Analysis of Pearson residuals of INAR and INARMA models. Left: Variances, which should be close to 1. Right: Autocorrelations, which should be small. The dashed line shows $2/\sqrt{T}$, that is, the 97.5% quantile for the empirical ACF of white noise.

In terms of the Akaike information criterion (AIC), both a more flexible autocorrelation structure (i.e., INGARCH or INARMA) and accounting for overdispersion considerably improve model fits. The INGARCH approach, where overdispersion enters both via the innovation and offspring distributions, leads to better results than the INARMA, where the offspring distribution is always Bernoulli. This is biologically plausible as the occurrence of multiple secondary infections is commonly observed in outbreak investigations (see Nishiura et al., 2017 for measles). The same qualitative results are obtained when using the Bayesian information criterion instead, see Supporting Information Table S5.

Figure 3 shows an analysis of the Pearson residuals of the INAR and INARMA models. The Pearson residuals are too dispersed for the Poisson version (variance exceeding 1); the negative binomial version can remedy this, while the Hermite model only partly does so. The INAR models show pronounced residual autocorrelation at lags 2 and 3, which is largely remedied by the INARMA versions. For the INARCH and INGARCH models, the picture is similar, see Supporting Information Figure S2. Graphical representations of the fits are provided in Supporting Information Figures S3 and S4.

Despite the good agreement with literature estimates, we emphasize that aggregate-level analyses like the above should not be overinterpreted; see discussion in the next Section.

8 | DISCUSSION

In this paper, we introduced an overarching class of count time series models, which includes many popular CP-INAR and INGARCH models. Each of them is characterized by an innovation, an offspring, and a clustering distribution. We gave particular attention to a new INARMA(p, q) model which mirrors the INGARCH(p, q) formulation. Numerous other instances could be examined, for example, models based on other thinning operators (Joe, 1996; Scotto et al., 2015) or compounding distributions (e.g., generalized Poisson models; Zhu, 2012). Other potential avenues are the inclusion of covariates and multivariate extensions.

We note that our class only comprises linear CP-INGARCH models with a time-constant clustering distribution. For instance, the negative binomial INGARCH model by Zhu (2011), which features a parameter θ_t that depends on λ_t, is not contained; nor are log-linear models (Fokianos et al., 2009) or other non-linear variations.

As noted before, the introduced INARMA(1, 1) class only allows for ACFs which from lag 2 onwards decay *more slowly* than in the corresponding INAR(1) model. For other INARMA models suggested in the literature (Dion et al., 1995), the converse is true. It would be desirable to construct a model able to accommodate both patterns.

Concerning the real-data application, several caveats are needed. First, our aggregate analysis glosses over population heterogeneities, ignoring, for example, that non-vaccination may be clustered in certain groups. Given the sparse data, we pragmatically assumed constant R_e within and across seasons. We consider this acceptable for the pre-COVID-19 period, but it would certainly not hold for the years since. Routine surveillance data are moreover subject to many biases, including reporting delays and underreporting, which can distort estimates of R_e (Bracher & Held, 2021). These aspects can moreover vary over time, for example, due to changes in healthcare-seeking or testing practices. In practice, branching process models are usually not fitted to surveillance counts alone, but also data, for example, on the type of infection (imported/domestic), and estimates based on different data types are compared to assess robustness (Chiew et al., 2014). This will yield more reliable estimates than we provide in our example.

ACKNOWLEDGMENTS

We would like to thank Mirko Armillotta, Konstantinos Fokianos, Melanie Schienle, and Christian Weiß for discussions on earlier versions of the paper. Both authors were supported by the German Research Foundation (DFG), project 512483310. Barbora Němcová was moreover supported by the Helmholtz Association under the joint research school *HIDSS4Health – Helmholtz Information and Data Science School for Health*. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

ORCID

Johannes Bracher https://orcid.org/0000-0002-3777-1410
Barbora Němcová https://orcid.org/0009-0004-7565-4145

REFERENCES

Aknouche, A., & Scotto, M. G. (2024). A multiplicative thinning-based integer-valued GARCH model. *Journal of Time Series Analysis*, *45*(1), 4–26.

Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*(3), 261–275.

Alzaid, A. A., & Al-Osh, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, *27*(2), 314–324.

Armillotta, M., & Gorgi, P. (2024). Pseudo-variance quasi-maximum likelihood estimation of semi-parametric time series models. *Journal of Econometrics*, *246*(1), 105894.

Bauer, C., & Wakefield, J. (2018). Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in China. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *67*(5), 1379–1398.

Benjamin, M. A., Rigby, R. A., & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, *98*(461), 214–223.

Bjørnstad, O. N., Finkenstädt, B. F., & Grenfell, B. T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, *72*(2), 169–184.

Bracher, J. (2019). *A new INARMA(1, 1) model with Poisson marginals*. In *Stochastic models, statistics and their applications* (pp. 323–333). Springer.

Bracher, J., & Held, L. (2021). A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts. *Biometrics*, *77*(4), 1202–1214.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.

Cardinal, M., Roy, R., & Lambert, J. (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine*, *18*(15), 2025–2039.

Chiew, M., Gidding, H. F., Dey, A., Wood, J., Martin, N., Davis, S., & McIntyre, P. (2014). Estimating the measles effective reproduction number in Australia from routine notification data. *Bulletin of the World Health Organization*, *92*(3), 171–177.

Cui, Y., & Lund, R. (2009). A new look at time series of counts. *Biometrika*, *96*(4), 781–792.

Davis, R. A., Fokianos, K., Holan, S. H., Joe, H., Livsey, J., Lund, R., Pipiras, V., & Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, *116*(535), 1533–1547.

de Jong, P., & Penzer, J. (2004). The ARMA model in state space form. *Statistics & Probability Letters*, *70*(1), 119–125.

De Serres, G., Gay, N. J., & Farrington, C. P. (2000). Epidemiology of transmissible diseases after elimination. *American Journal of Epidemiology*, *151*(11), 1039–1048.

Dion, J., Gauthier, G., & Latour, A. (1995). Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal*, *21*, 123–136.

Du, J. G., & Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, *12*(2), 129–142.

Farrington, C. P., & Grant, A. D. (1999). The distribution of time to extinction in subcritical branching processes: applications to outbreaks of infectious disease. *Journal of Applied Probability*, *36*(3), 771–779.

Feller, W. (1968). *An introduction to probability theory & its applications* (Vol. *1*). Wiley.

Ferland, R., Latour, A., & Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis*, *27*(6), 923–942.

Fernández-Fontelo, A., Fontdecaba, S., Alba, A., & Puig, P. (2017). Integer-valued AR processes with Hermite innovations and time-varying parameters: An application to bovine fallen stock surveillance at a local scale. *Statistical Modelling*, *17*(3), 172–195.

Fokianos, K., Rahbek, A., & Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, *104*(488), 1430–1439.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015a). Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis*, *36*(4), 503–527.

Gonçalves, E., Mendes-Lopes, N., & Silva, F. (2015b). A new approach to integer-valued time series modeling: The Neyman type-A INGARCH model. *Lithuanian Mathematical Journal*, *55*(2), 231–242.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., & Tweedie, R. L. (2000). Theory & methods: Non-Gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, *42*(4), 479–495.

Gupta, R. P., & Jain, G. C. (1974). A generalized Hermite distribution and its properties. *SIAM Journal of Applied Mathematics*, *27*(2), 359–363.

Jia, Y., Kechagias, S., Livsey, J., Lund, R., & Pipiras, V. (2023). Latent Gaussian count time series. *Journal of the American Statistical Association*, *118*(541), 596–606.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, *33*(3), 664–677.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1997). *Discrete multivariate distributions*. Wiley.

Kucharski, A., Mills, H., Pinsent, A., Fraser, C., Van Kerkhove, M., Donnelly, C. A., & Riley, S. (2014). Distinguishing between reservoir exposure and human-to-human transmission for emerging pathogens using case onset data. *PLOS Currents: Outbreaks*, *6*. https://pubmed.ncbi.nlm.nih.gov/24619563/

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, *19*(4), 439–455.

Lu, Y. (2021). The predictive distributions of thinning-based count processes. *Scandinavian Journal of Statistics*, *48*(1), 42–67.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Journal of the American Water Resources Association*, *21*(4), 645–650.
Meitz, M., & Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory*, *24*(5), 1291–1320.
Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *American Statistician*, *61*(1), 22–27.
Nishiura, H., Mizumoto, K., & Asai, Y. (2017). Assessing the transmission dynamics of measles in Japan, 2016. *Epidemics*, *20*, 67–72.
Pakes, A. G. (1971). Branching processes with immigration. *Journal of Applied Probability*, *8*(1), 32–42.
Pedeli, X., Davison, A. C., & Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association*, *511*, 1229–1238.
Pei, J., Lu, Y., & Zhu, F. (2024). Mixed causal-noncausal count process. *TEST*, 1–36.
Schweer, S., & Weiß, C. H. (2014). Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis*, *77*, 267–284.
Scotto, M. G., Weiß, C. H., & Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, *15*(6), 590–618.
Steutel, F. W., & van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, *7*(5), 893–899.
Vink, M. A., Bootsma, M. C. J., & Wallinga, J. (2014). Serial intervals of respiratory infectious diseases: A systematic review and analysis. *American Journal of Epidemiology*, *180*(9), 865–875.
Vu, H. T. V., & Maller, R. A. (1996). The likelihood ratio test for Poisson versus binomial distributions. *Journal of the American Statistical Association*, *91*(434), 818–824.
Weiß, C. H. (2015). A Poisson INAR(1) model with serially dependent innovations. *Metrika*, *78*(7), 829–851.
Weiß, C. H. (2018). *An introduction to discrete-valued time series*. Wiley.
Weiß, C. H. (2021). Stationary count time series models. *WIREs Computational Statistics*, *13*(1), e1502.
Weiß, C. H., Feld, M. H. J., Mamode Khan, N., & Sunecher, Y. (2019). INARMA modeling of count time series. *Stats*, *2*(2), 284–320.
Weiß, C. H., Gonçalves, E., & Lopes, N. M. (2017). Testing the compounding structure of the CP-INARCH model. *Metrika*, *80*(5), 571–603.
Weiß, C. H., & Puig, P. (2015). *The marginal distribution of compound Poisson INAR(1) processes*. In *Stochastic models, statistics and their applications* (pp. 351–359). Springer.
Weiß, C. H., & Zhu, F. (2024). Conditional-mean multiplicative operator models for count time series. *Computational Statistics & Data Analysis*, *191*, 107885.
Xu, H. Y., Xie, M., Goh, T. N., & Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Computational Statistics & Data Analysis*, *56*(12), 4229–4242.
Zhu, F. (2011). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, *32*(1), 54–67.
Zhu, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, *389*(1), 58–71.
Zucchini, W., & MacDonald, I. (2009). *Hidden markov models for time series*. Chapman and Hall/CRC.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article. An R package implementing the presented estimation method is available at https://github.com/jbracher/rinarma. Data and code to reproduce all results are available at https://github.com/jbracher/ginarma.

How to cite this article: Bracher, J., & Němcová, B. (2025). A unifying class of compound Poisson integer-valued ARMA and GARCH models. *Scandinavian Journal of Statistics*, 1–30. https://doi.org/10.1111/sjos.12784