Random environment INAR models of higher order

Authors: Aleksandar S. Nastić
– Department of Mathematics, University of Niš, Serbia (anastic78@gmail.com)

Petra N. Laketa
– Department of Mathematics, University of Niš, Serbia (petra.laketa@pmf.edu.rs)

Miroslav M. Ristić
– Department of Mathematics, University of Niš, Serbia (miristic72@gmail.com)

Abstract:

- Two different random environment INAR models of higher order, RrNGINARmax(p) and RrNGINAR$_1$(p), are introduced. Both of them are of variable order, which is not a random variable. Each step order is defined using the random environment process. Properties of the defined models are analyzed in parallel. The conditional expectation and variance are calculated. The strongly consistent Yule-Walker estimators are defined and new modified estimators are given. Models are applied to the simulated and the real-life data and the results show benefits of the introduced models.

Key-Words:

- *Random environment; INAR(p); RrNGINAR; negative binomial thinning; geometric marginals.*

AMS Subject Classification:

- 62M10.

1. INTRODUCTION

The integer-valued autoregressive (INAR) processes are introduced by McKenzie ([9]) and Al-Osh and Alzaid ([2]). They were the subject of research of many scientists, so there are a lot of different models which all intend to better describe the data obtained from some natural processes. The most of them are stationary, since this property gives some simplifications. Some of the models which investigate different thinning operators are given in [3], [8], [17, 18] and [14]. Models with various marginal distributions can be found in [10], [1], [5] and [6]. Weiß ([16]) and Nastić and Ristić ([11]) considered mixed processes. The first combined INAR(p) process is introduced in [16] and the combined process which is important for this paper is CGINAR(p) from [12]. It is combined in the sense that in every step recursive formula for an element of the process has one of the p possible forms (which match with formulas for INAR process) with some given probabilities.

We can say that stationary processes are rigid, because some of their properties are conserved in time. However, the real data are not often like that. One of the models which improves this weakness is RrNGINAR(1), defined in [13]. This is achieved by letting elements of the process to have varying distribution. Namely, quantitative properties observed from the nature depend on the environment. Since these values are represented by the elements of the process, it is natural to expect mentioned distribution to depend on the environment, too. It is supposed that environment conditions can be divided into r different types, which are called states, and each state is associated with a fixed distribution, so element of the process has the distribution of its state.

The main idea of this article is to make CGINAR(p) process more flexible, using the idea from RrNGINAR(1) process. Therefore, the aim is to construct a CGINAR(p) process with random states. So, in the second section of this article two different ways of constructing such a process which overcome problems that occur in classical (the most intuitive) way of defining this kind of model are discussed. Its correlation structure is analyzed in the third section. The fourth section is about Yule-Walker (YW) estimators of the parameters of the defined models. The quality of YW estimators is examined on the simulated data in Section 5. In the final section the introduced processes are applied to the real data and the results are compared for different models.

2. MODELS DEFINITIONS AND PROPERTIES

As mentioned earlier, our aim in this paper is to introduce the combined RrNGINAR process, where RrNGINAR process of order one is introduced in [13]. An attempt to construct this kind of combined process in the classical way, as

it was done in [16], [12] and [15] will bring some difficulties, so new approaches will be used. In this section we will define two processes, which overcome this problem. Discussion about some of their properties will be given.

Let $E_r = \{1, 2, \ldots, r\}$ be the set of all possible states, where $r \in \mathbb{N}$ and let $\{z_n\}$, $n \in \mathbb{N}_0$, be a realization of an r states random environment process $\{Z_n\}$ (we use Definition of the r random environment process given in [13]). For $i, j \in E_r$, let $\{\varepsilon_n(i,j)\}$, $n \in \mathbb{N}$, be sequences of independent identically distributed (i.i.d.) random variables. We will use notation $X_n(z_n)$ for an element of the new process, where z_n (which represents realized value of the random environment process in the moment $n \geq 0$) determines the distribution of that element. Let $\alpha*$ be the negative binomial thinning operator, for $\alpha \in (0,1)$, with a counting sequence $\{U_i,\, i \geq 1\}$ of i.i.d. random variables with probability mass function (pmf) given by

$$P(U_i = u) = \frac{\alpha^u}{(1+\alpha)^{u+1}}\,,\ u = 0, 1, 2, \ldots$$

As it was noted, it would be natural to introduce the combined random environment NGINAR process of order p in the following (classical) way

$$X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & \text{w.p. } \phi_1, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2, \\ \qquad\qquad \vdots & \qquad\quad \vdots \\ \alpha * X_{n-p}(z_{n-p}) + \varepsilon_n(z_{n-p}, z_n), & \text{w.p. } \phi_p, \end{cases} \tag{2.1}$$

for arbitrary $n \geq p$ and fixed $p \in \mathbb{N}$, where $\phi_i \geq 0$, $i \in \{1, 2, \ldots, p\}$, $\sum_{i=1}^{p} \phi_i = 1$, where distribution of $X_n(z_n)$ is given by

$$P(X_n(z_n) = x) = \frac{\mu_{z_n}^x}{(1+\mu_{z_n})^{x+1}}\,,\ x = 0, 1, 2, \ldots, \quad n = 0, 1, 2, \ldots,$$

$\mu_{z_n} \in \{\mu_1, \mu_2, \ldots, \mu_r\}$ is the parameter determined by the value z_n, $\mu_i > 0,\ i \in \{1, 2, \ldots, r\}$ and where the next conditions are satisfied

(*A*1) $\{Z_n\}$, $\{\varepsilon_n(1,1)\}$, $\{\varepsilon_n(1,2)\}$, $\ldots$, $\{\varepsilon_n(r,r)\}$, are mutually independent for all $n \geq 1$,

(*A*2) $X_n(l)$ is independent of Z_m and $\varepsilon_m(i,j)$, for $0 \leq n < m$ and any $i, j, l \in E_r$.

If we try to derive the distribution of $\varepsilon_n(i,j)$, $i, j \in E_r$, using procedure similar to the one of CGINAR(p) process, it wouldn't be so easy. Actually, it is not necessary that z_{n-j} for all $j = 1, 2, \ldots, p$ are the same, so, consequently, $\varepsilon_n(z_{n-j}, z_n)$ do not have to be identically distributed for all $j = 1, 2, \ldots, p$. This leads to a complex expression for the distribution of $\varepsilon_n(i,j)$, where i and j are arbitrary values from E_r.

The first method for avoiding this problem is to define $X_n(z_n)$ using (2.1), but substituting p with p_n, where p_n is the maximal number less or equal to

the given value p ($p \in \mathbb{N}$ is a fixed number, not depending on n), which satisfies $z_{n-1} = \cdots = z_{n-p_n}$. Then $\varepsilon_n(z_{n-j}, z_n)$ for $j = 1, 2, \ldots, p_n$ are the same, and obviously all have the same distribution. Let's define this more precisely.

Definition 2.1. Let z_n be the realization of the random environment process $\{Z_n\}$ in the moment $n \geq 0$. We say that $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$ is an INAR process with r-states random environment guided geometric marginals based on the negative binomial thinning operator of maximal order p (RrNGINARmax(p)), $p \in \mathbb{N}$, if the random variable $X_n(z_n)$ is defined as

$$X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & \text{w.p. } \phi_1^{(p_n)}, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2^{(p_n)}, \\ \vdots & \vdots \\ \alpha * X_{n-p_n}(z_{n-p_n}) + \varepsilon_n(z_{n-p_n}, z_n), & \text{w.p. } \phi_{p_n}^{(p_n)}, \end{cases} \tag{2.2}$$

for $n \geq 1$, where

$$p_n = \begin{cases} p, & p_n^* \geq p, \\ p_n^*, & p_n^* < p, \end{cases}$$

$p_n^* = \max\{i \in \{1, 2, \ldots, n\} : z_{n-1} = z_{n-2} = \cdots = z_{n-i}\}$ and the following conditions are satisfied:

1. $\phi_i^{(p_n)} \geq 0$, $i \in \{1, 2, \ldots, p_n\}$, $\sum_{i=1}^{p_n} \phi_i^{(p_n)} = 1$,

2. $\alpha \in (0, 1)$ and the counting sequence $\{U_i\}_{i\in\mathbb{N}}$ of the negative binomial thinning operator $\alpha*$ has pmf $P(U_i = u) = \frac{\alpha^u}{(1+\alpha)^{u+1}}$, $u \in \{0, 1, 2, \ldots\}$,

3. $P(X_n(z_n) = x) = \frac{\mu_{z_n}^x}{(1+\mu_{z_n})^{x+1}}$, $x \in \{0, 1, 2, \ldots\}$, where $\mu_{z_n} \in \{\mu_1, \mu_2, \ldots, \mu_r\}$, $\mu_i > 0$, $i \in \{1, 2, \ldots, r\}$ and $r \in \mathbb{N}$ is the number of states of the random environment process $\{Z_n\}$,

4. for fixed $i, j \in E_r = \{1, 2, .., r\}$, $\{\varepsilon_n(i, j)\}_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables,

5. $\{Z_n\}, \{\varepsilon_n(1, 1)\}, \{\varepsilon_n(1, 2)\}, \ldots, \{\varepsilon_n(r, r)\}$ are mutually independent sequences of random variables,

6. $X_n(l)$ is independent of Z_m and $\varepsilon_m(i, j)$, for $0 \leq n < m$ and any $i, j, l \in E_r$.

We want to emphasize that this model contains p different sets of the probability parameters $\Psi_i = \{\phi_1^{(i)}, \phi_2^{(i)}, \ldots, \phi_i^{(i)}\}$, for $i \in \{1, 2, \ldots, p\}$. Set Ψ_i has i elements, so the total number of the probability parameters is $1 + 2 + \cdots + p = \frac{p(p+1)}{2}$. For each i there is a condition $\sum_{j=1}^{i} \phi_j^{(i)} = 1$, so there are $\frac{p(p+1)}{2} - p = \frac{p(p-1)}{2}$ unknown probability parameters. Specially, for $i = 1$, we have $\phi_1^{(1)} = 1$.

Remark 2.1. Important feature of the introduced process is a variable order. Actually, $\{X_n(z_n)\}$ is defined like a process of order p_n, where p_n is not fixed and depends on n. But, p_n is not a random variable due to the fact that it could be calculated for given $\{z_n\}$, using its building mechanism given in Definition 2.1 and the fact that our process is defined for the realized random environment process $\{z_n\}$. Once p_n reaches p, process takes shape of the model of fixed order, p, and this lasts as long as the state does not change. When it changes ($z_n \neq z_{n-1}$), then order (p_{n+1}) becomes equal 1. The order further continues to grow until the state changes again or until it reaches p. Therefore, we consider process which is mostly of order p, but it has some transitional periods of variable and ascending order, which begin when the state changes and end when the process reaches order p, or when state changes again.

This is similar to the idea of the Variable-Order Markov (VOM) model, which was investigated in [7]. As it is known a random variable in the Markov chain model depends on a fixed number of previous conditioning elements. However, in VOM models the number of conditioning random variables (which is called the context) depends on the specific observed realization and may vary over time.

Now, we will describe one more combined random environment NGINAR process. It is similar to the previous one, but differs during the transitional period (for $p_n^* < p$), where the process of variable order is replaced with the process of order one.

Definition 2.2. Let z_n be the realization of the random environment process $\{Z_n\}$ in the moment $n \geq 0$. We say that $\{X_n(z_n)\}_{n\in\mathbb{N}_0}$ is an INAR process with r-states random environment guided geometric marginals based on the negative binomial thinning operator of order p (RrNGINAR$_1(p)$) if random variable $X_n(z_n)$ is defined as

$$X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & \text{w.p. } \phi_1^{(p_n)}, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & \text{w.p. } \phi_2^{(p_n)}, \\ \qquad\qquad \vdots & \qquad\quad \vdots \\ \alpha * X_{n-p_n}(z_{n-p_n}) + \varepsilon_n(z_{n-p_n}, z_n), & \text{w.p. } \phi_{p_n}^{(p_n)}, \end{cases} \tag{2.3}$$

for $n \geq 1$, where

$$p_n = \begin{cases} p, & p_n^* \geq p, \\ 1, & p_n^* < p, \end{cases}$$

$p_n^* = \max\{i \in \{1, 2, \ldots, n\} : z_{n-1} = z_{n-2} = \cdots = z_{n-i}\}$ and conditions $1-6$ from Definition 2.1 are satisfied.

For $p_n = 1$ we have only one probability parameter $\phi_1^{(1)}$ and from the condition $\sum_{i=1}^{p_n} \phi_i^{(p_n)} = 1$ it follows $\phi_1^{(1)} = 1$, so only $\phi_1^{(p)}, \ldots, \phi_p^{(p)}$ are unknown, but

related via one equation. Therefore, it is sufficient to determine $p-1$ probability parameters.

Remark 2.2. For the process given by Definition 2.2, p_n is not a random variable as well. Now, p_n takes one of the two possible values. Every time when state changes ($z_n \neq z_{n-1}$), order (p_{n+1}) becomes 1 and it remains the same until there is a series of enough (p) previous elements corresponding to the same state. So, we can divide this process into the series which can be represented as parts of the processes of order 1 and order p. In the series of order 1 state can be changed, so this series has the same form as RrNGINAR(1) process. For the series of order p it is necessary to stay in the same state, so they are whole in the one state and have the same form as CGINAR(p) process. There are as much different (by their marginals) CGINAR(p) processes as much different states we have.

Our next step is derivation of the distribution of the random variable $\varepsilon_n(i,j)$.

Theorem 2.1. *Let $\{X_n(z_n)\}$ be the RrNGINARmax(p) time series process or the RrNGINAR$_1$(p) process, and let $\mu_1 > 0$, $\mu_2 > 0$, ..., $\mu_r > 0$. If $0 \leq \alpha \leq \min\left\{\frac{\mu_l}{1+\mu_k}, k,l \in E_r\right\}$, then if $z_n = j$ and $z_{n-1} = i$, for $i,j \in E_r$, the distribution of the random variable $\varepsilon_n(i,j)$ can be written as a mixture of two geometric distributions*

$$\varepsilon_n(i,j) \stackrel{d}{=} \begin{cases} Geom\left(\frac{\mu_j}{1+\mu_j}\right), & \text{w.p. } 1-\frac{\alpha\mu_i}{\mu_j-\alpha}, \\ Geom\left(\frac{\alpha}{1+\alpha}\right), & \text{w.p. } \frac{\alpha\mu_i}{\mu_j-\alpha}, \end{cases} \tag{2.4}$$

for $n \geq 1$.

Proof: Consider the probability generating function (pgf) of a random variable $X_n(z_n)$ in the case when $\{X_n(z_n)\}$ is the RrNGINARmax(p) process. Due to the properties of the pgf and the definition of the negative binomial thinning operator, it holds

$$\begin{aligned} \Phi_{X_n(z_n)}(s) &= \sum_{l=1}^{p_n} \phi_l^{(p_n)} E(s^{\alpha * X_{n-l}(z_{n-l})}) E(s^{\varepsilon_n(z_{n-l},z_n)}) \\ &= \sum_{l=1}^{p_n} \phi_l^{(p_n)} \Phi_{X_{n-l}(z_{n-l})}(\Phi_U(s)) \Phi_{\varepsilon_n(z_{n-l},z_n)}(s). \end{aligned}$$

We used notation Φ_U for Φ_{U_m}, $m \geq 1$, since U_m have all the same distribution. Because $z_{n-1} = z_{n-2} = \cdots = z_{n-p_n} = i$, it holds $\Phi_{\varepsilon_n(z_{n-1},z_n)}(s) = \Phi_{\varepsilon_n(z_{n-2},z_n)}(s) = \cdots = \Phi_{\varepsilon_n(z_{n-p_n},z_n)}(s) = \Phi_{\varepsilon_n(i,j)}(s)$ and $\Phi_{X_{n-1}(z_{n-1})}(s) = \Phi_{X_{n-2}(z_{n-2})}(s) = \cdots = \Phi_{X_{n-p_n}(z_{n-p_n})}(s) = \Phi_{X_{n-1}(i)}(s)$, so

$$\Phi_{X_n(j)}(s) = \sum_{l=1}^{p_n} \phi_l^{(p_n)} \Phi_{X_{n-1}(i)}(\Phi_U(s)) \Phi_{\varepsilon_n(i,j)}(s) = \Phi_{X_{n-1}(i)}(\Phi_U(s)) \Phi_{\varepsilon_n(i,j)}(s).$$

The last equation is equivalent to

$$\frac{1}{1+\mu_j-\mu_j s} = \Phi_{\varepsilon_n(i,j)}(s)\frac{1}{1+\mu_i-\frac{\mu_i}{1+\alpha-\alpha s}},$$

because $X_n(j)$ has $Geom\left(\frac{\mu_j}{1+\mu_j}\right)$ distribution, $X_{n-1}(i)$ has $Geom\left(\frac{\mu_i}{1+\mu_i}\right)$ distribution and U_m has $Geom\left(\frac{\alpha}{1+\alpha}\right)$ distribution. Calculation of $\Phi_{\varepsilon_n(i,j)}(s)$ gives

$$\Phi_{\varepsilon_n(i,j)}(s) = \frac{\alpha\mu_i}{\mu_j-\alpha}\cdot\frac{1}{1+\alpha-\alpha s} + \left(1-\frac{\alpha\mu_i}{\mu_j-\alpha}\right)\cdot\frac{1}{1+\mu_j-\mu_j s},$$

which implies (2.4).

Suppose now that $\{X_n(z_n)\}$ is the $\mathrm{RrNGINAR}_1(p)$ process. Let fix $n\in\mathbb{N}$. If $p_n^*\geq p$, then $X_n(z_n)$ is generated in the same way as $\mathrm{RrNGINARmax}(p)$ process and it holds $p_n=p$. So, applying the same procedure as before, substituting p_n with p we get (2.4). If the previous condition doesn't hold, $X_n(z_n)$ has the form like in the RrNGINAR(1) process, for which we know that $\varepsilon_n(i,j)$ has the required distribution. □ □

Now, we derive conditional expectation and variance of the introduced processes.

Theorem 2.2. *Let $\{X_n(z_n)\}$ be RrNGINARmax(p) or $RrNGINAR_1(p)$ time series process, and let $\mu_1>0$, $\mu_2>0$, …, $\mu_r>0$. If $0\leq\alpha\leq\min\left\{\frac{\mu_l}{1+\mu_k},\ k,l\in E_r\right\}$, $z_{n+1}=j$ and $z_n=i$, for $i,j\in E_r$, then the conditional expectation and the conditional variance of this process are given by*

$$E(X_{n+1}|H_n) = \mu_j-\alpha\mu_i+\alpha\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}X_{n+1-l},$$

$$\begin{aligned}Var(X_{n+1}|H_n) \;=\; & \mu_j(\mu_j+1)-\alpha\mu_i(1+2\alpha+\alpha\mu_i)+\alpha(1+\alpha)\sum_{l=1}^{p_{n+k}}\phi_l^{(p_{n+1})}X_{n+1-l}\\ & +\alpha^2\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}X_{n+1-l}^2-\alpha^2\left(\sum_{l=1}^{p_{n+k}}\phi_l^{(p_{n+1})}X_{n+1-l}\right)^2,\end{aligned}$$

where $H_n=\sigma(X_n,X_{n-1},\ldots,X_{n-p_{n+1}})$ represents the σ-field generated by $\{X_n,$ $X_{n-1},\ldots,X_{n-p_{n+1}}\}$.

Proof: For the simplicity of notation, we will use X_n instead of $X_n(z_n)$, for $n\geq 0$ and ε_n instead of $\varepsilon_n(z_{n-1},z_n)$, for $n\geq 1$. From the definition of the

negative binomial thinning and the properties of the conditional expectation, the conditional probability generating function is

$$\begin{aligned}\Phi_{X_{n+1}|H_n}(s) \equiv E\left(s^{X_{n+1}}|H_n\right) &= \Phi_{\varepsilon_{n+1}}(s)\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}E\left(s^{\alpha * X_{n+1-l}}|H_n\right)\\ &= \Phi_{\varepsilon_{n+1}}(s)\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}\Phi_U^{X_{n+1-l}}(s),\end{aligned}$$

where $\Phi_U(s)=\frac{1}{1+\alpha-\alpha s}$. It holds

$$E(X_{n+1}|H_n)=\Phi'_{X_{n+1}|H_n}(1)$$

and

$$Var(X_{n+1}|H_n)=\Phi''_{X_{n+1}|H_n}(1)+\Phi'_{X_{n+1}|H_n}(1)-(\Phi'_{X_{n+1}|H_n}(1))^2.$$

Derivating function $\Phi_{X_{n+1}|H_n}(s)$ with respect to s and using results

$$\Phi_U(1)=1,\quad \Phi'_U(1)=\alpha,\quad \Phi''_U(1)=2\alpha^2$$

and

$$\Phi_{\varepsilon_{n+1}}(1)=1,\quad \Phi'_{\varepsilon_{n+1}}(1)=\mu_j-\alpha\mu_i,\quad \Phi''_{\varepsilon_{n+1}}(1)=2\mu_j^2-2\alpha\mu_i(\mu_j+\alpha)$$

gives

$$\Phi'_{X_{n+1}|H_n}(1)=\mu_j-\alpha\mu_i+\alpha\sum_{l=1}^{p_{n+k}}\phi_l^{(p_{n+1})}X_{n+1-l}$$

and

$$\begin{aligned}\Phi''_{X_{n+1}|H_n}(1) &= 2\mu_j^2-2\alpha\mu_i(\mu_j+\alpha)+\alpha^2\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}X_{n+1-l}^2\\ &\quad +\alpha(2\mu_j-2\alpha\mu_i+\alpha)\sum_{l=1}^{p_{n+1}}\phi_l^{(p_{n+1})}X_{n+1-l}.\end{aligned}$$

The requested formulas directly follow from here. □ □

The conditional expectation and the conditional variance of higher order can be calculated using following recurrent relations:

$$E(X_{n+k}|H_n)=\mu_j-\alpha\mu_i+\alpha\left[\sum_{l=1}^{k-1}\phi_l^{(p_{n+1})}E(X_{n+k-l}|H_n)+\sum_{l=k}^{p_{n+k}}\phi_l^{(p_{n+1})}X_{n+k-l}\right]$$

for $2\le k\le p_{n+k}$,

$$E(X_{n+k}|H_n)=\mu_j-\alpha\mu_i+\alpha\sum_{l=1}^{p_{n+k}}\phi_l^{(p_{n+1})}E(X_{n+k-l}|H_n),\quad k>p_{n+k},$$

$$
\begin{aligned}
Var(X_{n+k}|H_n) \;=\; & \mu_j(\mu_j+1) - \alpha\mu_i(1+2\alpha+\alpha\mu_i) \\
& +\alpha^2 \sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} Var(X_{n+k-l}|H_n) \\
& +\alpha(1+\alpha)\left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) + \sum_{l=k}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l}\right) \\
& +\alpha^2\left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} \left[E(X_{n+k-l}|H_n)\right]^2 + \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l}^2\right) \\
& -\alpha^2\left(\sum_{l=1}^{k-1} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n) + \sum_{l=k}^{p_{n+k}} \phi_l^{(p_{n+1})} X_{n+k-l}\right)^2
\end{aligned}
$$

for $2 \le k \le p_{n+k}$,

$$
\begin{aligned}
Var(X_{n+k}|H_n) \;=\; & \mu_j(\mu_j+1) - \alpha\mu_i(1+2\alpha+\alpha\mu_i) \\
& +\alpha^2 \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} Var(X_{n+k-l}|H_n) \\
& +\alpha(1+\alpha)\left(\sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n)\right) \\
& +\alpha^2 \sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} \left[E(X_{n+k-l}|H_n)\right]^2 \\
& -\alpha^2\left(\sum_{l=1}^{p_{n+k}} \phi_l^{(p_{n+1})} E(X_{n+k-l}|H_n)\right)^2
\end{aligned}
$$

for $k > p_{n+k}$.

3. CORRELATION STRUCTURE

Let's now investigate the correlation structure of the defined processes. First, we will examine the RrNGINARmax(p) process. From (2.2), it follows

$$
\begin{aligned}
Cov(X_n, X_{n-1}) \;&=\; \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} Cov(X_{n-i}, X_{n-1}), \\
Cov(X_n, X_{n-2}) \;&=\; \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} Cov(X_{n-i}, X_{n-2}), \\
&\;\;\vdots \\
Cov(X_n, X_{n-p_n}) \;&=\; \alpha \sum_{i=1}^{p_n} \phi_i^{(p_n)} Cov(X_{n-i}, X_{n-p_n}),
\end{aligned}
$$

where we have used X_n instead of $X_n(z_n)$ for the simplicity of notation. Denote $Cov(X_n, X_{n-h})$ as $\gamma_n^{(h)}$, for $h \geq 0$. This system can be represented in the matrix form

$$\begin{bmatrix} \gamma_n^{(1)} \\ \gamma_n^{(2)} \\ \vdots \\ \gamma_n^{(p_n)} \end{bmatrix} = \begin{bmatrix} \gamma_{n-1}^{(0)} & \gamma_{n-1}^{(1)} & \cdots & \gamma_{n-1}^{(p_{n-1})} \\ \gamma_{n-1}^{(1)} & \gamma_{n-2}^{(0)} & \cdots & \gamma_{n-2}^{(p_{n-2})} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1}^{(p_n-1)} & \gamma_{n-2}^{(p_n-2)} & \cdots & \gamma_{n-p_n}^{(0)} \end{bmatrix} \cdot \begin{bmatrix} \theta_1^{(p_n)} \\ \theta_2^{(p_n)} \\ \vdots \\ \theta_{p_n}^{(p_n)} \end{bmatrix}, \tag{3.1}$$

where $\theta_i^{(p_n)} = \alpha\phi_i^{(p_n)}$, for $i \in \{1, 2, \ldots, p_n\}$. More simple form is of course obtained by matrix notation, i.e.

$$\boldsymbol{\gamma}_n = \boldsymbol{\Gamma}_n \cdot \boldsymbol{\theta}_n,$$

where we denoted the corresponding vectors with $\boldsymbol{\gamma_n}$ and $\boldsymbol{\theta}_n$, and $\boldsymbol{\Gamma}_n$ is the covariance matrix of the vector $(X_{n-p_n}, X_{n-p_n-1}, ..., X_{n-1})'$.

In accordance with the definition of p_n random variables $X_{n-1}, X_{n-2}, \ldots, X_{n-p_n}$ have the same distribution, so it holds

$$Cov(X_{n-1}, X_{n-1}) = Cov(X_{n-2}, X_{n-2}) = \cdots = Cov(X_{n-p_n}, X_{n-p_n}) = \sigma^2_{X_{n-1}}.$$

Now, it is possible to divide covariance matrix $\boldsymbol{\Gamma}_n$ with $\sigma^2_{X_{n-1}}$ and the result is the correlation matrix

$$\mathbf{R}^{(n-1)}_{p_n \times p_n} = \begin{bmatrix} 1 & \rho_{n-1}^{(1)} & \cdots & \rho_{n-1}^{(p_{n-1})} \\ \rho_{n-1}^{(1)} & 1 & \cdots & \rho_{n-2}^{(p_{n-2})} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1}^{(p_n-1)} & \rho_{n-2}^{(p_n-2)} & \cdots & 1 \end{bmatrix}. \tag{3.2}$$

However, dividing the left side of the equation (3.1) with $\sigma^2_{X_{n-1}}$ will not give the vector of the correlations, because $z_n \neq z_{n-1}$ in general. Actually, the equation which is satisfied by the correlation matrix is

$$\begin{bmatrix} 1 & \rho_{n-1}^{(1)} & \cdots & \rho_{n-1}^{(p_n-1)} \\ \rho_{n-1}^{(1)} & 1 & \cdots & \rho_{n-2}^{(p_n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1}^{(p_n-1)} & \rho_{n-2}^{(p_n-2)} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \theta_1^{(p_n)} \\ \theta_2^{(p_n)} \\ \vdots \\ \theta_{p_n}^{(p_n)} \end{bmatrix} = \frac{\sigma_{X_n}}{\sigma_{X_{n-1}}} \begin{bmatrix} \rho_n^{(1)} \\ \rho_n^{(2)} \\ \vdots \\ \rho_n^{(p_n)} \end{bmatrix}. \tag{3.3}$$

Remark 3.1. In the special case when $z_n = z_{n-1}$, we have that $\sigma_{X_n} = \sigma_{X_{n-1}}$, so the equation for the correlation matrix takes the same form as the equation for the covariance matrix. It is important to notice that the subsample $X_{n-1}, X_{n-2}, \ldots, X_{n-p_n}$ of the RrNGINARmax(p) process cannot be seen as a subsample of a stationary process in general. Really, it is possible to be $p_i \neq p_j$, for $i, j \in \{n-1, n-2, \ldots, n-p_n\}$ (for example, if $z_{n-p_n-1} \neq z_{n-p_n}$, then $p_{n-p_n+1} = 1, p_{n-p_n+2} = 2, \ldots, p_{n-1} = p_n - 1$), so we deal with a process of

variable order, which does not have to be stationary. However, if $p_{n-p_n} = p$ and $z_{n-p_n} = z_{n-p_n-1}$ then elements of the subsample $X_{n-1}, X_{n-2}, \ldots, X_{n-p_n}$ all have the same distribution and are defined based on the p previous elements, so it is possible to consider this subsample as a subsample of CGINAR(p) process, which is stationary. Really, from the definition of p_m it holds $z_{m-1} = z_{m-2} = \ldots = z_{m-p_m}$ for arbitrary m. For $m = n$ we have $z_{n-1} = z_{n-2} = \ldots = z_{n-p_n}$ and for $m = n - p_n$ it holds $z_{n-p_n-1} = z_{n-p_n-2} = \ldots = z_{n-p_n-p}$, where we used relation $p_{n-p_n} = p$. Combining these results with equation $z_{n-p_n} = z_{n-p_n-1}$ gives $z_{n-p_n-\max p} = \ldots = z_{n-p_n-1} = z_{n-p_n} = \ldots = z_{n-1}$ and consequently, $p_{n-p_n+1} = p_{n-p_n+2} = \ldots = p_n = p$. If $z_n = z_{n-1}$, additionally, then the same conclusion holds for the subsample $X_n, X_{n-1}, \ldots, X_{n-p_n}$.

Now, let's consider the RrNGINAR$_1(p)$ process. It can be partitioned into samples of CGINAR(p) or RrNGINAR(1) processes. So, correlation structure is determined by the correlation structure of the mentioned processes. For $p_n = 1$ we have

$$\gamma_n^{(1)} = \alpha \gamma_{n-1}^{(0)},$$

and for $p_n = p$ it holds

$$\begin{bmatrix} \gamma_n^{(1)} \\ \gamma_n^{(2)} \\ \vdots \\ \gamma_n^{(p)} \end{bmatrix} = \begin{bmatrix} \gamma_{n-1}^{(0)} & \gamma_{n-1}^{(1)} & \cdots & \gamma_{n-1}^{(p-1)} \\ \gamma_{n-1}^{(1)} & \gamma_{n-2}^{(0)} & \cdots & \gamma_{n-2}^{(p-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1}^{(p-1)} & \gamma_{n-2}^{(p-2)} & \cdots & \gamma_{n-p}^{(0)} \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix},$$

where $\theta_i = \alpha\phi_i$, for $i \in \{1, 2, \ldots, p\}$. These equations can be represented by (3.1), substituting p_n with 1 and p. It also holds (3.2) and (3.3). RrNGINARmax(p) cannot have series parts with two or more successive elements of order one in the same state. However, for RrNGINAR$_1(p)$ process the maximal length of such a series is p. The Theorem 3, from [13], holds for n and k which satisfy $z_n = \cdots = z_{n-k}$ and $p_n = 1$. Therefore, based on this theorem, the maximal value that k can take is p.

Remark 3.2. For RrNGINAR$_1(p)$ process, subsample $X_{n-1}, X_{n-2}, \ldots, X_{n-p_n}$ can always be viewed as a sample of a stationary process. The case $p_n = 1$ is trivial since it gives subsample of only one element. If $p_n = p$ and $p_{n-p_n} = p$, then directly from the definition of this process it follows that $X_{n-1}, X_{n-2}, \ldots, X_{n-p_n}$ are all in the same state and of the same order. If $z_n = z_{n-1}$ then the same holds for $X_n, X_{n-1}, \ldots, X_{n-p_n}$.

4. YULE-WALKER ESTIMATION OF THE PARAMETERS

In the proof of Theorem 5 from [13] stationarity of the processes attached to the maximal subsamples provided the strong consistency of the estimators. If

we want here to prove the same, it would be useful to define estimators only on the part of the process to which we can attach a stationary process.

In accordance with Remark 3.1 let $p_n = p_{n-p_n} = p$, $z_{n-p_n} = z_{n-p_n-1}$ and $z_n = z_{n-1} = k \in E_r$ for RrNGINARmax(p) process. Then $p_i = p$ and $z_i = k$, for all $i \in \{n-p_n, n-p_n+1, \ldots, n\}$. Because of stationarity it is possible to write $\gamma_{|i-j|}$ instead of $\gamma_j^{(i)}$, for $i \in \{0, 1, \ldots, p_n\}, j \in \{n-p_n, n-p_n+1, \ldots, n\}$ without loss of generality. Then, we introduce $\gamma_{|i-j|}^{(k)}$ as a more informative and adequate notation, where index k indicates the random state. Applying this to the system (3.1), using analogously $\theta_{p,i}$ instead of $\theta_i^{(p_n)}$ we obtain

$$\begin{bmatrix} \gamma_1^{(k)} \\ \gamma_2^{(k)} \\ \vdots \\ \gamma_p^{(k)} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(k)} & \gamma_1^{(k)} & \cdots & \gamma_{p-1}^{(k)} \\ \gamma_1^{(k)} & \gamma_0^{(k)} & \cdots & \gamma_{p-2}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1}^{(k)} & \gamma_{p-2}^{(k)} & \cdots & \gamma_0^{(k)} \end{bmatrix} \cdot \begin{bmatrix} \theta_{p,1} \\ \theta_{p,2} \\ \vdots \\ \theta_{p,p} \end{bmatrix}. \tag{4.1}$$

Notice that p_n equals p, as it is assumed above. We estimate μ_k and $\gamma_h^{(k)}$, $h \in \{0, 1, 2, \ldots, p-1\}$ as in [13], but only based on the part of a sample. Precisely, estimators for state $k \in \{1, 2, \ldots, r\}$ are based on the sets $V_{0,p}^{(k)} = \{i \in \{1, 2, \ldots, N\} | z_i = k,\ p_i = p\}$ and $V_{h,p}^{(k)} = \left\{ i \in V_{0,p}^{(k)} | i + h \in V_{0,p}^{(k)} \right\}$, for $h \geq 1$, where N is the size of the sample, and are given by

$$\widehat{\mu}_k = \frac{1}{n_{0,p}^{(k)}} \sum_{i \in V_{0,p}^{(k)}} X_i(k), \quad \widehat{\gamma}_{h,p}^{(k)} = \frac{1}{n_{h,p}^{(k)}} \sum_{i \in V_{h,p}^{(k)}} (X_{i+h}(k) - \widehat{\mu}_k)(X_i(k) - \widehat{\mu}_k), \tag{4.2}$$

for $h \geq 0$, $k \in \{1, 2, \ldots, r\}$ and where $n_{h,p}^{(k)} = \left| V_{h,p}^{(k)} \right|$, for $h \geq 0$.

Substituting the theoretical moments in (4.1) with the empirical ones and then expressing the vector of the unknown parameters, we get

$$\begin{bmatrix} \widehat{\theta}_{p,1}^{(k)} \\ \widehat{\theta}_{p,2}^{(k)} \\ \vdots \\ \widehat{\theta}_{p,p}^{(k)} \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_{0,p}^{(k)} & \widehat{\gamma}_{1,p}^{(k)} & \cdots & \widehat{\gamma}_{p-1,p}^{(k)} \\ \widehat{\gamma}_{1,p}^{(k)} & \widehat{\gamma}_{0,p}^{(k)} & \cdots & \widehat{\gamma}_{p-2,p}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1,p}^{(k)} & \widehat{\gamma}_{p-2,p}^{(k)} & \cdots & \widehat{\gamma}_{0,p}^{(k)} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \widehat{\gamma}_{1,p}^{k)} \\ \widehat{\gamma}_{2,p}^{(k)} \\ \vdots \\ \widehat{\gamma}_{p,p}^{(k)} \end{bmatrix}.$$

Now, it is possible to estimate α and $\phi_i^{(p)}$. First we obtain, respectively

$$\widehat{\alpha}^{(k)} = \sum_{i=1}^{p} \widehat{\theta}_{p,i}^{(k)}, \qquad \widehat{\phi}_{p,i}^{(k)} = \frac{\widehat{\theta}_{p,i}^{(k)}}{\widehat{\alpha}^{(k)}}, \qquad i \in \{1, 2, \ldots, p\},\ k \in \{1, 2, \ldots, r\},$$

where (k), as above, indicates that the estimators are based on the subsample with state equal k. At last, taking into account all states and their frequencies of occurrence, the final Yule-Walker estimators are

$$\widehat{\alpha}^{YW} = \sum_{k=1}^{r} \frac{n^{(k)}}{N} \widehat{\alpha}^{(k)}, \qquad \widehat{\phi}_{p,i}^{YW} = \sum_{k=1}^{r} \frac{n^{(k)}}{N} \widehat{\phi}_{p,i}^{(k)}, \qquad \widehat{\mu}_k^{YW} = \widehat{\mu}_k. \tag{4.3}$$

Now, we move our attention to RrNGINAR$_1(p)$ process. Let $z_n = z_{n-1}$. Estimators of the process covariances are based on the maximal union (in the sense of number of elements) of the samples which can be treated as the samples of stationary processes. Estimators are given by

$$\widehat{\mu}_{k,j} = \frac{1}{n_{0,j}^{(k)}} \sum_{i \in V_{0,j}^{(k)}} X_i(k), \quad \widehat{\gamma}_{h,j}^{(k)} = \frac{1}{n_{h,j}^{(k)}} \sum_{i \in V_{h,j}^{(k)}} (X_{i+h}(k) - \widehat{\mu}_{k,j})(X_i(k) - \widehat{\mu}_{k,j}), \tag{4.4}$$

where $h \geq 0$, $k \in \{1, 2, \ldots, r\}$, $j \in \{1, p\}$, and they are based on the sets $V_{0,1}^{(k)} = \{i \in \{1, 2, \ldots, N\} |\ z_i = k,\ p_i = 1\}$, $V_{h,1}^{(k)} = \left\{ i \in V_{0,1}^{(k)} | i + h \in V_{0,1}^{(k)} \right\}$, $n_{h,1}^{(k)} = \left| V_{h,1}^{(k)} \right|$, for $h \geq 1$ and $V_{h,p}^{(k)}$ and $n_{h,p}^{(k)}$, for $h \geq 0$ are defined as before. Let $\alpha_j^{(k)}$ represents the autocorrelation parameter corresponding to the process subsamples of state k and order j, where $j \in \{1, p\}$.

Similarly as for RrNGINARmax(p) process we get

$$\widehat{\alpha}_1^{(k)} = \frac{\widehat{\gamma}_{1,1}^{(k)}}{\widehat{\gamma}_{0,1}^{(k)}}$$

and

$$\begin{bmatrix} \widehat{\theta}_1^{(k)} \\ \widehat{\theta}_2^{(k)} \\ \vdots \\ \widehat{\theta}_p^{(k)} \end{bmatrix} = \begin{bmatrix} \widehat{\gamma}_{0,p}^{(k)} & \widehat{\gamma}_{1,p}^{(k)} & \cdots & \widehat{\gamma}_{p-1,p}^{(k)} \\ \widehat{\gamma}_{1,p}^{(k)} & \widehat{\gamma}_{0,p}^{(k)} & \cdots & \widehat{\gamma}_{p-2,p}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1,p}^{(k)} & \widehat{\gamma}_{p-2,p}^{(k)} & \cdots & \widehat{\gamma}_{0,p}^{(k)} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \widehat{\gamma}_{1,p}^{k)} \\ \widehat{\gamma}_{2,p}^{(k)} \\ \vdots \\ \widehat{\gamma}_{p,p}^{(k)} \end{bmatrix}. \tag{4.5}$$

From (4.5) we obtain the estimators as

$$\widehat{\alpha}_p^{(k)} = \sum_{i=1}^{p} \widehat{\theta}_i^{(k)}, \qquad \widehat{\phi}_i^{(k)} = \frac{\widehat{\theta}_i^{(k)}}{\widehat{\alpha}_p^{(k)}}, \qquad i \in \{1, 2, \ldots, p\},\ k \in \{1, 2, \ldots, r\}.$$

Now, we get

$$\widehat{\alpha}^{(k)} = \frac{n_1^{(k)} \widehat{\alpha}_1^{(k)} + n_p^{(k)} \widehat{\alpha}_p^{(k)}}{n_1^{(k)} + n_p^{(k)}}, \quad \widehat{\mu}_k = \frac{n_1^{(k)} \widehat{\mu}_{k,1} + n_p^{(k)} \widehat{\mu}_{k,p}}{n_1^{(k)} + n_p^{(k)}},$$

and finally, using preceding results for all states, we obtain YW estimators as

$$\widehat{\alpha}^{YW} = \sum_{k=1}^{r} \frac{n_1^{(k)} + n_p^{(k)}}{N} \widehat{\alpha}^{(k)}, \qquad \widehat{\phi}_i^{YW} = \sum_{k=1}^{r} \frac{n_p^{(k)}}{N} \widehat{\phi}_i^{(k)}, \qquad \widehat{\mu}_k^{YW} = \widehat{\mu}_k. \tag{4.6}$$

Theorem 4.1. *Estimators given by* (4.2) *and* (4.4) *are strongly consistent.*

Proof: The general idea is to divide the process subsample, indexed, i.e. determined by $V_{0,p}^{(k)}$, into maximal subsamples and then use the proof of Theorem 5 from [13]. It is easy to notice that this theorem can be expanded so that applies for $h > 1$. Really, if in the expression for γ_1 we replace $i + 1$ by $i + h$, it becomes γ_h. Further procedure is the same. If $X_i, X_{i+1}, \ldots, X_j$ is a subsample such that $\{i, i + 1, \ldots, j\} \subseteq V_{0,p}^{(k)}$, we say that it is maximal if $z_i = z_{i+1} = \cdots = z_j = k$, $z_{j+1} \neq k$, $p_i = p$ and $p_{i-1} \neq p$. Based on Remark 3.1, $X_i, X_{i+1}, \ldots, X_j$ represents a sample of CGINAR(p) process, so it is stationary. The rest of the proof is the same as the proof of Theorem 5. The same procedure is applied to the RrNGINAR$_1(p)$ process. □ □

Since the quotient of linear combinations of the strongly consistent statistics is also strongly consistent, we have the following corollary.

Corollary 4.1. *Estimators given in* (4.3) *and* (4.6) *are strongly consistent.*

If we want to estimate the probabilities $\phi_i^{(p_n)}$, where $1 < p_n < p$ and $1 \leq i \leq p_n$, given in (2.2) for the definition of RrNGINARmax(p) process, preceding approach cannot be used. The problem is that elements X_n, for $1 < p_n < p$ are isolated in the sense that the both of their neighbors have different order, so it is impossible to form the subsample containing X_n (to define the estimators) with two or more successive elements of the same order.

This problem is worked out by defining new modified YW estimators which have less restrictive conditions in using the corresponding subsamples of the process. These modified estimators are obtained from the strongly consistent YW estimators, discussed above, by substituting their corresponding sets $V_{h,p}^{(k)}$, for $h \geq 1$, with $V_{0,p}^{(k)}$. In other words, if the corresponding sets of the modified YW estimators are denoted by $\tilde{V}_{h,p}^{(k)}$, then $\tilde{V}_{h,p}^{(k)} = V_{0,p}^{(k)}$, for $h \geq 1$. Note that $\tilde{V}_{h,p}^{(k)} \supseteq V_{h,p}^{(k)}$. However, because of these modifications, we cannot claim the modified YW strong consistence, but their goodness may be verified in the application on the simulated process values. In this regard, the results obtained in the next section show that the corresponding estimates gradually converge towards parameter values when the size of the sample increases.

5. SIMULATIONS

In this section we investigate the correctness of the modified Yule-Walker estimators. For this purpose we have simulated realizations of the processes RrNGINARmax(p) and RrNGINAR$_1(p)$, and estimated unknown parameters in both cases. There are 100 replicates, each of size 10000. Both of the processes

are considered in parallel. We choose the parameters $\alpha, p, r, \boldsymbol{\mu}, \mathbf{p}_{mat}$ and $\boldsymbol{\phi}$. The random environment process transition probability matrix is noted by $\mathbf{p}_{mat}$, and $\boldsymbol{\mu}$ is a vector of means. In the case of RrNGINARmax(p) process, the p_nth row, $p_n \in \{2, \ldots, p\}$, of the matrix $\boldsymbol{\phi}$ contains (up to the p_nth column) probabilities $\phi_i^{(p_n)}$, $i \in \{1, 2, \ldots, p_n\}$, from (2.2). In the case of RrNGINAR$_1$(p) process, the last row represents probabilities in (2.3). Matrix $\mathbf{p}_{mat}$ controls changing of the states, where diagonal elements represents the probabilities of staying in the same state. When its diagonal values are high, it is expected for the random environment process to stay in the same state more often than to change the state. This is a preferable situation since it makes the sets $V_{h,p}^{(k)}$, for $h \geq 1$, to be bigger and in this way relatively less different from $\tilde{V}_{h,p}^{(k)}$. Consequently, modified estimators became approximately equal (at least being close) to the strongly consistent YW estimators.

The simulation of the random environment process $\{Z_n\}$ represents the first step in the simulation of the defined processes. After the generation of the observed values $\{z_n\}$, we can easily evaluate the process of orders $\{p_n\}$ for both processes by using their definitions. Finally, we can simulate the values of both defined processes by using the observed values $\{z_n\}$ and $\{p_n\}$, and definitions of the defined processes.

We considered six different cases. In each case we obtained the modified YW estimators of the unknown parameters for both of the processes. All the results are given in the appropriate tables. Comparison of the results is based on the relative errors, since values of the parameters are different.

1) In the first case vector of means is $\boldsymbol{\mu} = (1, 2)$. For these values maximal value for α is $1/3$ and we chose $\alpha = 0.3$. The random state process transition probability matrix we used is $\mathbf{p}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$. Diagonal elements are equal 0.8, so, based on the discussion above, good estimates are expected. Matrix of probabilities is $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}$.

2) Here we investigate what happens when α, from case 1), reduces to $\alpha = 0.15$, where $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$. Since lower value of α contributes to less correlation, it is natural to expect that estimates for $\boldsymbol{\phi}$ are worse than in the case 1 and this is confirmed by the results. Estimates for α and $\boldsymbol{\mu}$ are almost the same, but slightly better.

3) This case differs from the first by the probabilities of changing state. We used $\boldsymbol{p}_{mat} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$. Now, there is equal probability to stay in the same state as it is to change it. As we discussed earlier, this is not favorably, so worse estimates for $\boldsymbol{\phi}$ are expected and this conclusion is confirmed by the results. However, our sample is big enough, so difference is very small.

Estimates for α and $\boldsymbol{\mu}$ are again slightly better then the estimates of case 1).

4) Vector of means is now $\boldsymbol{\mu} = (4, 5)$ and probabilities on the diagonal of the transition matrix of random states are 0.7, i.e. $\boldsymbol{p}_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$. Estimates are almost the same as in the first case. Thereby, RrNGINARmax(p) process provides better results, while the RrNGINAR$_1$(p) process gives worse results than in the first case.

These preceding four cases refer to the processes based on the environment process with two random states. The corresponding results are presented in Table 1.

5) Here we consider what happens when, in case 1), maximal order p increases to 3. The estimates of α are slightly better, but the estimates of $\boldsymbol{\phi}$ are worse, because the probabilities for the order 2 (in case of RrNGINARmax(p) process) are estimated using very small sample. However, they are significantly improved when sample size increases to 10000. These results are presented in Table 2.

6) In the last case we have simulated process with three possible states ($r = 3$). Parameters, as well as the results, are given in Table 3. The greater number of states contributes to the smaller probability of staying in the same state, so for the small sample sizes, estimates are not so good, but increasing the size of the simulated sample gives much better results. Estimates for $\phi_1^{(2)}$ and $\phi_2^{(2)}$ are not so good, which is reasonable because elements x_n of the simulated sample, such that $p_n = 2$, have neighbors with order different from 2.

In the cases 1), 5) and 6) the RrNGINAR$_1$(p) process provides better estimates, while in cases 2) and 4) the RrNGINARmax(p) process is better choice. In the third case they are almost equally good. It is important to notice that in each case results are better when the size of the sample increases.

6. APPLICATION

Quality of the processes introduced in this paper will be investigated by comparing the results obtained in the application of the various models to the same data. For this purpose, here we use two data sets of counts. Since the processes introduced in this article are not stationary, we expect them to perform well on the data chosen in [13]. So, in the first case, we choose this time series, which was created by counting drug offenses per month registered in the 27th police car beat in Pittsburg from January 1990 to December 2001. It has a

Table 1: Estimates for $p = 2$, $r = 2$

n_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{2,1}^{YW}$	$\widehat{\phi}_{2,2}^{YW}$	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{2,1}^{YW}$	$\widehat{\phi}_{2,2}^{YW}$
1) True values $\boldsymbol{\mu} = (1, 2)$, $\alpha = 0.3$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$								
500	0.9924	1.9782	0.3226	0.6389	0.3611	0.318	0.624	0.376
SE	0.1200	0.2097	0.1474	0.2009	0.2009	0.1477	0.2338	0.2338
1000	0.9912	2.0064	0.3185	0.6198	0.3802	0.3188	0.6084	0.3916
SE	0.0872	0.1487	0.1071	0.1371	0.1371	0.1066	0.1282	0.1282
5000	0.9953	1.9952	0.3048	0.6074	0.3926	0.2984	0.6089	0.3911
SE	0.0375	0.0539	0.0525	0.0578	0.0578	0.0466	0.057	0.057
10000	0.9978	1.9999	0.3049	0.6038	0.3962	0.2993	0.5971	0.4029
SE	0.0288	0.0407	0.0388	0.0381	0.0381	0.0318	0.0409	0.0409
2) True values $\boldsymbol{\mu} = (1, 2)$, $\alpha = 0.15$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$								
500	0.9988	2.0047	0.1568	0.688	0.312	0.1781	2.2577	-1.2577
SE	0.0939	0.1965	0.13	0.8714	0.8714	0.1438	15.5205	15.5205
1000	0.9933	2.0136	0.1532	0.4913	0.5087	0.1636	0.4655	0.5345
SE	0.0745	0.1312	0.0983	1.2527	1.2527	0.1038	0.7767	0.7767
5000	1.0011	1.9999	0.1526	0.5068	0.4932	0.1547	0.4921	0.5079
SE	0.0349	0.0562	0.0442	0.0984	0.0984	0.0459	0.0992	0.0992
10000	1.0031	1.9995	0.1543	0.5008	0.4992	0.1523	0.4951	0.5049
SE	0.0252	0.0368	0.0284	0.0649	0.0649	0.0309	0.0647	0.0647
3) True values $\boldsymbol{\mu} = (1, 2)$, $\alpha = 0.3$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$								
500	1.0044	1.9879	0.3201	0.5478	0.4522	0.3279	0.5864	0.4136
SE	0.1082	0.2065	0.1306	0.9084	0.9084	0.1501	1.2909	1.2909
1000	1.0069	1.9916	0.3132	0.7188	0.2812	0.3095	0.6959	0.3041
SE	0.0783	0.1413	0.0888	0.7523	0.7523	0.0955	0.6136	0.6136
5000	0.9995	1.9948	0.3047	0.5965	0.4035	0.3024	0.5946	0.4054
SE	0.0293	0.0618	0.043	0.0892	0.0892	0.0415	0.0926	0.0926
10000	0.9997	1.9922	0.3032	0.5847	0.4153	0.3018	0.5991	0.4009
SE	0.0197	0.0430	0.0283	0.0624	0.0624	0.0274	0.0657	0.0657
4) True values $\boldsymbol{\mu} = (4, 5)$, $\alpha = 0.5$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 \\ 0.6 & 0.4 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$								
500	4.0397	4.9501	0.5506	0.3952	0.6048	0.5466	0.4286	0.5714
SE	0.4132	0.4852	0.1469	0.1352	0.1352	0.1489	0.2029	0.2029
1000	4.0295	4.9785	0.5259	0.4237	0.5763	0.526	0.4309	0.5691
SE	0.3016	0.3590	0.105	0.0819	0.7523	0.1162	0.0945	0.0945
5000	4.0046	5.0008	0.5037	0.4163	0.5837	0.514	0.411	0.589
SE	0.1286	0.1601	0.0522	0.0408	0.0408	0.0492	0.0396	0.0396
10000	3.9947	4.9978	0.4982	0.4132	0.5868	0.5054	0.4135	0.5865
SE	0.0900	0.1173	0.0366	0.0278	0.0278	0.0354	0.0279	0.0279

length of 144 realizations and is downloaded from a website Forecasting Principles (http://www.forecastingprinciples.com). The plots of the given series and its

Table 2: Estimates for $p = 3$, $r = 2$

n_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{2,1}^{YW}$	$\widehat{\phi}_{2,2}^{YW}$	$\widehat{\phi}_{3,1}^{YW}$	$\widehat{\phi}_{3,2}^{YW}$	$\widehat{\phi}_{3,3}^{YW}$
5) True values $\boldsymbol{\mu} = (1, 2)$, $\alpha = 0.3$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$								
500	1.0167	1.9948	0.3185	2.1388	-1.1388	0.4742	0.2559	0.2699
SE	0.1181	0.2203	0.1656	8.6452	8.6452	0.9294	0.4239	0.6798
1000	1.0076	1.9915	0.2954	0.6158	0.3842	0.5373	0.2636	0.199
SE	0.0813	0.1452	0.1052	1.0074	1.0074	0.2578	0.2562	0.2199
5000	0.9994	2.0023	0.2956	0.6182	0.3818	0.4983	0.2893	0.2124
SE	0.0416	0.0717	0.0435	0.1311	0.1311	0.0659	0.0651	0.0686
10000	0.9978	1.9983	0.2956	0.6115	0.3885	0.5001	0.2964	0.2035
SE	0.0254	0.0509	0.0345	0.0935	0.0935	0.0437	0.0413	0.0489

	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{3,1}^{YW}$	$\widehat{\phi}_{3,2}^{YW}$	$\widehat{\phi}_{3,3}^{YW}$
500	0.2949	0.4969	0.3206	0.1825
SE	0.109	1.136	0.572	0.6893
1000	0.3052	0.4992	0.2923	0.2085
SE	0.0853	0.1857	0.1793	0.1791
5000	0.2966	0.491	0.3073	0.2017
SE	0.0378	0.0809	0.0661	0.0735
10000	0.2998	0.4916	0.3121	0.1963
SE	0.0241	0.0522	0.0496	0.0482

autocorrelation and partial autocorrelation function are given in Figures 1, 3 and 5. Here, we might have used two different approaches for choosing the model order p. The first one, which is more intuitive, is based on choosing p as the number of first p significant values of the partial autocorrelation functions observed from the diagram (in this case Figure 5). The other approach is defining p as the value from $\{1, 2, \ldots, q\}$ for which the smallest RMS value is obtained (RMS is the quality criterion explained later in this paragraph), where q is some reasonably large integer value. However, to make things easier to follow, we have decided to use the compromise of these two approaches. Namely, we choose the maximal considered model order p as the larger of the two numbers obtained by the first (intuitive) approach used for both data sets, increased by one. Since, these values for both data sets are 2 (for the 27th police car station) and 3 (for the second data set considered later in this section), we choose 4 as a maximal order of the INAR models considered for both of the observed counting time series. Therefore, INAR(p) models, for $p \in \{1, 2, 3, 4\}$, might be the reasonable choice. Considering the referent models of order 1, we chose INAR(1) model with Poisson marginals (PoINAR(1)) given in [2], quasi-binomial INAR(1) model with generalized Poisson marginals (GPQINAR(1)) from [5], geometric INAR(1) model (GINAR(1)) introduced in [4], new geometric INAR(1) (NGINAR(1)) defined in [14], negative binomial INAR(1) (NBINAR(1)) introduced in [19, 20], iterated INAR(1) model (NBIINAR(1)) with negative binomial marginals given in [1] and random coefficient INAR(1) model with negative binomial marginals (NBRCINAR(1)) constructed in [18]. Since our models, which quality we want to verify, are com-

Table 3: Estimates for $p = 3,\ r = 3$

n_1	$\widehat{\mu}_1^{YW}$	$\widehat{\mu}_2^{YW}$	$\widehat{\mu}_3^{YW}$	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{2,1}^{YW}$	$\widehat{\phi}_{2,2}^{YW}$	$\widehat{\phi}_{3,1}^{YW}$	$\widehat{\phi}_{3,2}^{YW}$	$\widehat{\phi}_{3,3}^{YW}$
6) True values $\boldsymbol{\mu} = (1, 1.5, 2)$, $\alpha = 0.3$, $\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}$, $\mathbf{p}_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}$									
500	1.0068	1.4991	1.9849	0.3228	4.9236	-3.9236	-2.3274	5.6962	-2.3688
SE	0.1617	0.1734	0.2385	0.1181	25.8252	25.8252	26.8363	53.7332	26.9268
1000	1.0084	1.5031	1.9742	0.3156	0.6707	0.3293	0.5876	0.2652	0.1472
SE	0.1130	0.1321	0.1704	0.0946	1.447	1.447	0.4793	0.3186	0.3543
5000	0.9976	1.5018	1.9981	0.3092	0.6306	0.3694	0.4833	0.302	0.2147
SE	0.0465	0.0544	0.0679	0.0435	0.1296	0.1296	0.0884	0.098	0.0854
10000	0.9970	1.5024	1.9983	0.3058	0.5994	0.4006	0.4743	0.3106	0.2151
SE	0.0329	0.0328	0.0510	0.0297	0.0796	0.0796	0.0492	0.0538	0.054

	$\widehat{\alpha}^{YW}$	$\widehat{\phi}_{3,1}^{YW}$	$\widehat{\phi}_{3,2}^{YW}$	$\widehat{\phi}_{3,3}^{YW}$
500	0.3205	0.4554	0.4348	0.1098
SE	0.1042	0.7894	0.8933	1.2224
1000	0.3096	0.5513	0.333	0.1157
SE	0.0804	0.3914	0.4626	0.4641
5000	0.2957	0.5002	0.3123	0.1875
SE	0.0322	0.0951	0.0986	0.0918
10000	0.2955	0.4876	0.3158	0.1967
SE	0.0232	0.0705	0.0659	0.0639

binations of the RrNGINAR(1) process from [13] and CGINAR(p) process from [12] in some way, it is natural to include them in consideration, too. For this purpose we used R2NGINAR(1), R3NGINAR(1), CGINAR(2), CGINAR(3) and CGINAR(4) models. Another process of higher order which is included in this section, because of the completeness of the comparison, is PoINAR(p) ([16]), precisely, PoINAR(2), PoINAR(3) and PoINAR(4). The root mean squares (RMS) of differences between the observations and predicted values (using maximum likelihood estimation) are calculated and all the results are given in Table 4. The lower values of RMS indicate the better and more appropriate models. The values of the maximum likelihood parameter estimates and RMS statistics are also calculated in case of application of our processes RrNGINARmax(p) and RrNGINAR$_1$(p) of appropriate orders 2, 3 and 4, taking into account the cases with two or three possible random states. (see Tables 5 and 6).

The same procedure of comparison of our processes to all the INAR models used above is also conducted in the second case of counting time series, i.e. on the drugs offenses counting data which were registered in the 58th police car beat in Pittsburg. The corresponding results are given by Figures 2, 4 and 6 and Tables 7, 8 and 9.

In both cases of the observed offenses data the realization of the random environment process $\{z_n\}$ is determined in the same way as in [13], by clustering

the data. Each cluster is assigned to a state. Then the corresponding sequence $\{p_n\}$ is calculated using the definitions of the models. The plots of the clusterings (Figures 7, 8, 9 and 10) show big difference between the data recorded by the police stations in the way of the environment state changing. For the data from the 27th police station probability of staying in the same state is much higher than for the observations from the 58th station. Analysis of the results lead us to a conclusion that in both cases of the selected data, the models introduced in this paper are better then the others which we applied. Namely, in the case of the 27th car beat drug offenses, R3NGINARmax(4) is the most appropriate model, while in the case of the data recorded by the 58th police station, R3NGINAR$_1$(3) process shows the best performance. It is interesting to note that the optimal order 3 obtained for the 58th police station data is in accordance with the value obtained by the graphical intuitive (the first one) approach for choosing order p. However, in the case of the 27th police station this is not the case (where graphical approach gives $p = 2$), which justify the usage of our compromise approach for choosing order of the model.

Successful performance of our Random Environment INAR models in both of the cases show that they might be very appropriate for the processes which quite often change their marginal distribution (58th car beat data counting), as well as for the processes which are, on the other hand, much more passive (27th police car beat data), i.e. which only rarely shift from one set of environment circumstances to another. Increase of the number of random states contributes to the improving of the results for both models in each case of the observed data. However, optimal order depends on the data. Thus, for the 27th station the more appropriate Random Environment INAR models are mostly the ones with the higher order, while for the 58th police station we have the opposite situation.

REFERENCES

[1] Al-Osh, M.A. and Aly, E.E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals, *Commun. Statist. Theory Meth.*, **21**, 2483–2492.

[2] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *J. Time Ser. Anal.*, **8**, 261–275.

[3] Aly, E.E.A.A. and Bouzar, N. (1994). On some integer-valued autoregressive moving average models, *Journal of Multivariate Analysis*, **50**, 132–151.

[4] Alzaid, A.A. and Al-Osh, M.A. (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties, *Statistica Neerlandica*, **42**, 53–61.

[5] Alzaid, A.A. and Al-Osh, M.A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions, *Ann. Inst. Statist. Math.*, **45**, 223–232.

[6] Bakouch, H.S. and Ristić, M.M. (2010). Zero truncated Poisson integer valued AR(1) model, *Metrika*, **72**, 2, 265–280.

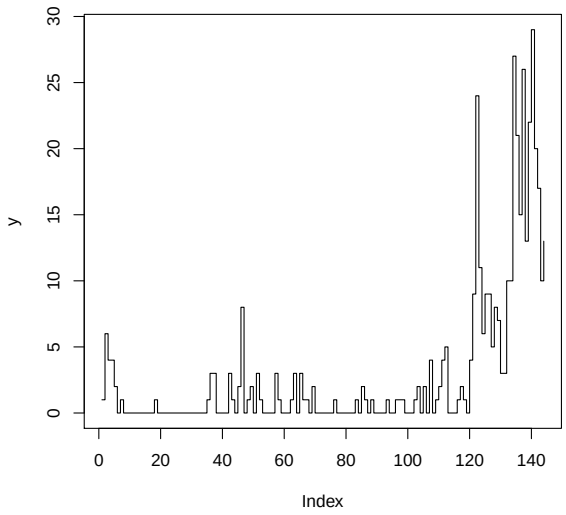

Figure 1: Drugs data from the 27th police station.

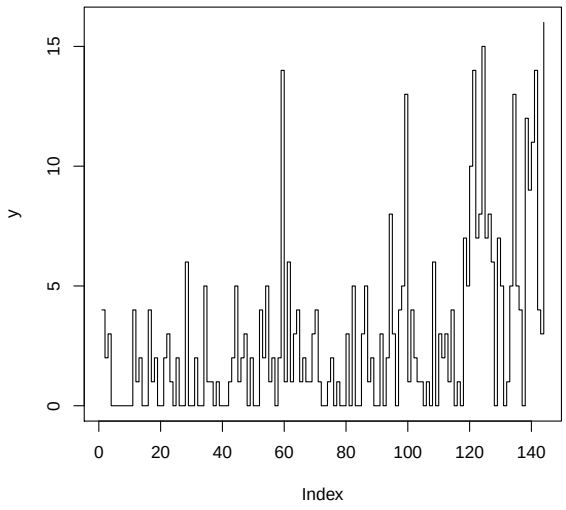

Figure 2: Drugs data from the 58th police station.

[7] Begleiter, R.; El-Yaniv, R. and Yona, G. (2004). On prediction using variable order Markov models, *Journal of Artificial Intelligence Research*, **22**, 385–421.

[8] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process, *J. Time Ser. Anal.*, **19**, 439–455.

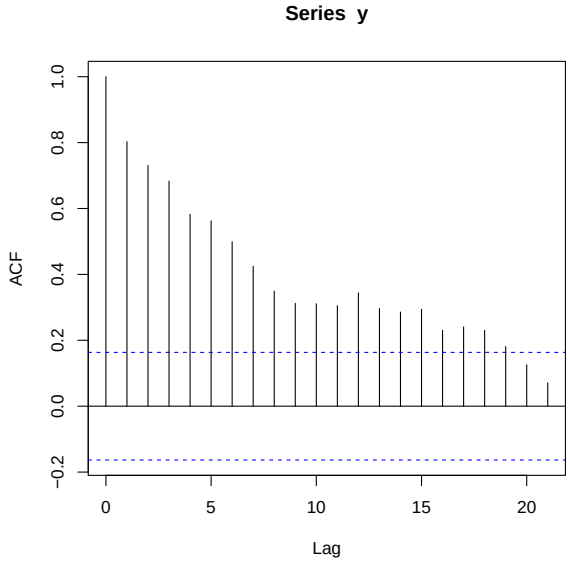

Figure 3: ACF for the data from the 27th police station.

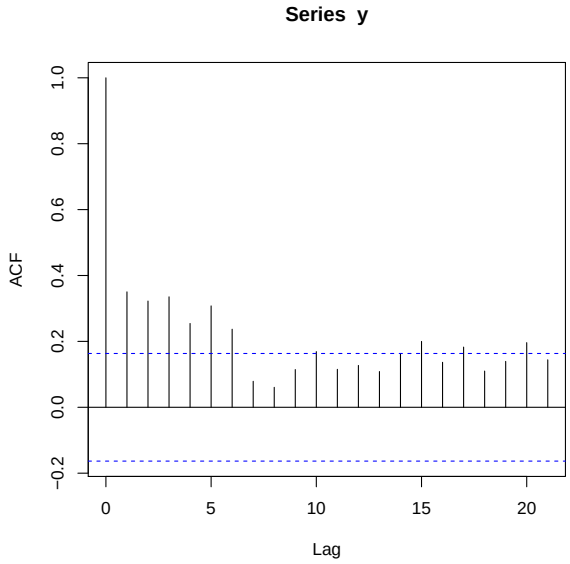

Figure 4: ACF for the data from the 58th police station.

[9] McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resour. Bull.*, **21**, 645–650.

[10] McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric distributions, *Adv. Appl. Prob.*, **18**, 679–705.

[11] Nastić, A.S. and Ristić, M.M. (2012). Some geometric mixed integer-valued

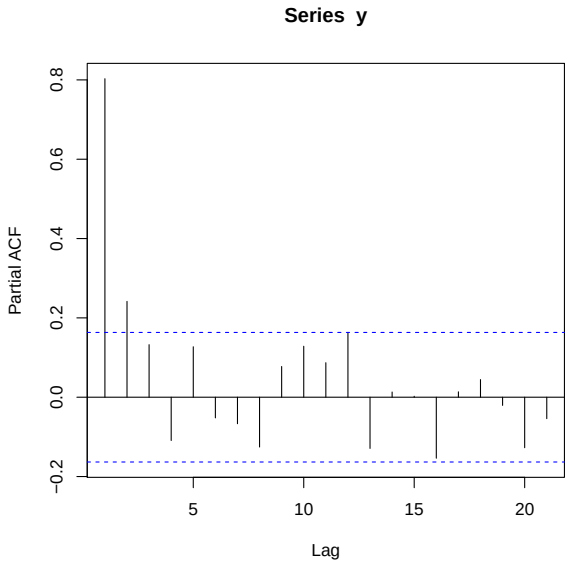

Figure 5: PACF for the data from the 27th police station.

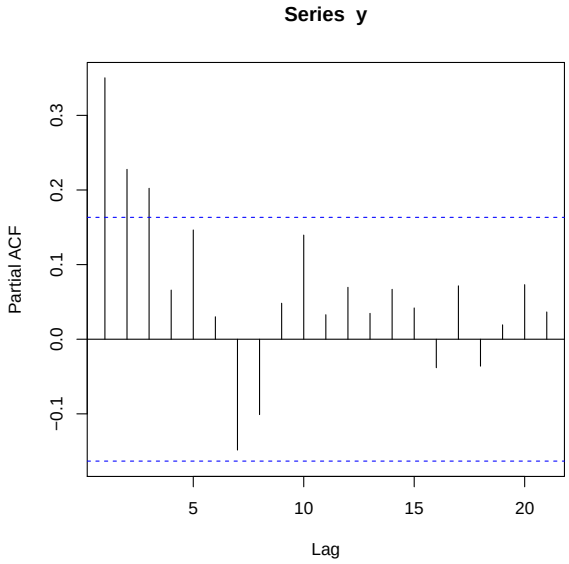

Figure 6: PACF for the data from the 58th police station.

autoregressive (INAR) models, *Statist. Probab. Lett.*, **82**, 805–811.

[12] Nastić, A.S.; Ristić, M.M. and Bakouch, H.S. (2012). A combined geometric INAR(p) model based on negative binomial thinning, *Mathematical and Computer Modelling*, **55**, 1665–1672.

[13] Nastić, A.S.; Laketa, P.N. and Ristić, M.M. (2016). Random environment

Table 4: ML parameter estimates and RMS for different models for the data from the 27th police station.

Model	MLE	RMS
PoINAR(1)	$\widehat{\lambda} = 1.237$	
	$\widehat{\alpha} = 0.5948$	3.6613
GPQINAR(1)	$\widehat{\lambda} = 0.5505$	
	$\widehat{\theta} = 0.6108$	
	$\widehat{\rho} = 0.392$	4.3398
GINAR(1)	$\widehat{q} = 0.7596$	
	$\widehat{\alpha} = 0.4809$	3.9456
NGINAR(1)	$\widehat{\mu} = 3.3014$	
	$\widehat{\alpha} = 0.7308$	3.4595
NBINAR(1)	$\widehat{q} = 0.2173$	
	$\widehat{\theta} = 0.834$	
	$\widehat{\alpha} = 0.4563$	4.0185
NBIINAR(1)	$\widehat{n} = 0.323$	
	$\widehat{p} = 0.5335$	
	$\widehat{\rho} = 0.8107$	3.4211
NBRCINAR(1)	$\widehat{n} = 0.5435$	
	$\widehat{p} = 0.1854$	
	$\widehat{\rho} = 0.46$	4.0232
R2NGINAR(1)	$\widehat{\mu}_1 = 1.1085$	
	$\widehat{\mu}_2 = 12.9138$	
	$\widehat{\alpha} = 0.052$	3.1090
R3NGINAR(1)	$\widehat{\mu}_1 = 9.5906$	
	$\widehat{\mu}_2 = 0.821$	
	$\widehat{\mu}_3 = 23.249$	
	$\widehat{\alpha} = 0.028$	1.6628
CGINAR(2)	$\widehat{\mu} = 3.2042$	
	$\widehat{\alpha} = 0.7473$	3.3801
CGINAR(3)	$\widehat{\mu} = 3.1122$	
	$\widehat{\alpha} = 0.745$	3.3749
CGINAR(4)	$\widehat{\mu} = 3.124$	
	$\widehat{\alpha} = 0.7464$	3.398
PoINAR(2)	$\widehat{\lambda} = 0.9903$	
	$\widehat{\alpha} = 0.687$	3.4822
PoINAR(3)	$\widehat{\lambda} = 0.8616$	
	$\widehat{\alpha} = 0.7449$	3.4253
PoINAR(4)	$\widehat{\lambda} = 0.84$	
	$\widehat{\alpha} = 0.7682$	3.4803

integer-valued autoregressive process, *J. Time Ser. Anal.*, **37**, 267–287.

[14] Ristić, M.M.; Bakouch, H.S. and Nastić, A.S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, *J. Stat. Plann.*

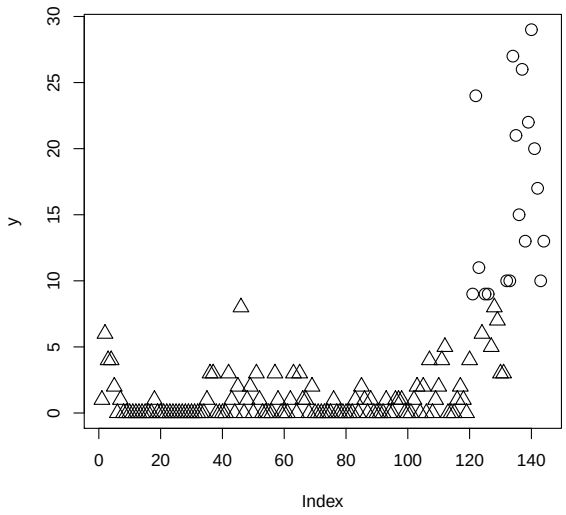

Figure 7: Clusters for two states for the 27th police station.

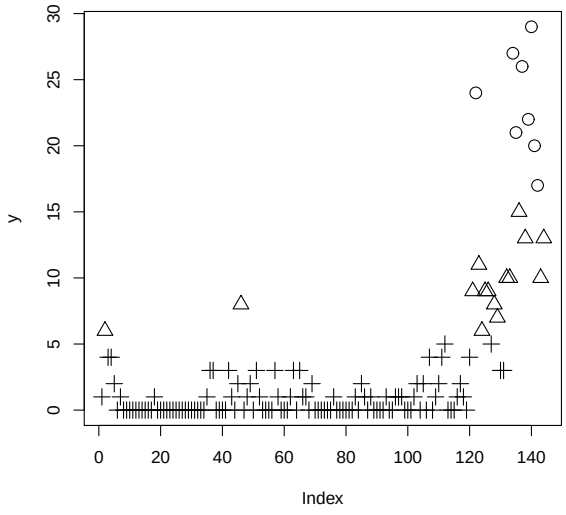

Figure 8: Clusters for three states for the 27th police station.

Inf., **139**, 2218–2226.

[15] Ristić, M.M. and Nastić, A.S. (2012). A mixed INAR(p) model, *J. Time Ser. Anal.*, **33**, 903–915.

[16] Weiss, C.H. (2008). The combined INAR(p) models for time series of counts, *Statist. Probab. Lett.*, **72**, 1817–1822.

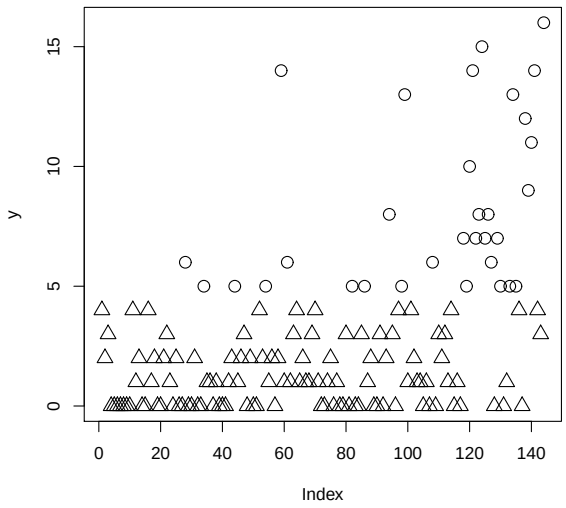

Figure 9: Clusters for two states for the 58th police station.

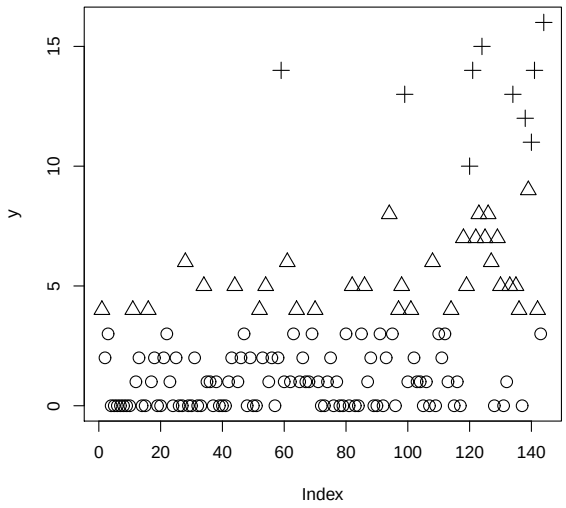

Figure 10: Clusters for three states for the 58th police station.

[17] Zheng, H.; Basawa, I.V. and Datta, S. (2006). Inference for pth-order random coefficient integer-valued autoregressive processes, *J. Time Ser. Anal.*, **27**, 411–440.

[18] Zheng, H.; Basawa, I.V. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes, *J. Stat. Plann. Inf.*, **137**, 212–229.

Table 5: ML parameter estimates and RMS for R2NGINARmax(p) and R3NGINARmax(p) process for the data from the 27th police station.

r	2	2	2
p	2	3	4
$\widehat{\alpha}$	0.0368	0.0141	0.0556
$\widehat{\boldsymbol{\mu}}$	(12.2357, 0.6728)	(15.8556, 0.6226)	(19.8312, 1.3864)
$\widehat{\boldsymbol{\phi}}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.4601 & 0.5310 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0.0000 & 0.0000 \\ 0.3395 & 0.3670 & 0.2935 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.3310 & 0.3315 & 0.3375 & 0.0000 \\ 0.2384 & 0.2157 & 0.2493 & 0.2966 \end{bmatrix}$
RMS	3.2154	2.9089	2.8400
r	3	3	3
p	2	3	4
$\widehat{\alpha}$	0.0296	0.0124	0.0241
$\widehat{\boldsymbol{\mu}}$	(23.2410, 9.0586, 0.7170)	(23.2500, 9.0570, 0.4382)	(23.2500, 9.0585, 0.7009)
$\widehat{\boldsymbol{\phi}}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0432 & 0.9568 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.3391 & 0.3226 & 0.3382 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.3300 & 0.3300 & 0.3401 & 0.0000 \\ 0.2477 & 0.1993 & 0.2383 & 0.3146 \end{bmatrix}$
RMS	1.6528	1.6532	1.6175

Table 6: ML parameter estimates and RMS for R2NGINAR$_1$(p) and R3NGINAR$_1$(p) process for the data from the 27th police station.

r	2	2	2
p	2	3	4
$\widehat{\alpha}$	0.0368	0.0276	0.0226
$\widehat{\boldsymbol{\mu}}$	(12.2357, 0.6728)	(13.1349, 0.6075)	(13.6046, 0.5570)
$\widehat{\boldsymbol{\phi}}$	(0.4601, 0.5400)	(0.3958, 0.4267, 0.1775)	(0.3399, 0.3956, 0.1292, 0.1353)
RMS	3.2154	3.1079	3.0595
r	3	3	3
p	2	3	4
$\widehat{\alpha}$	0.0296	0.0090	0.0093
$\widehat{\boldsymbol{\mu}}$	(23.2450, 9.0586, 0.7170)	(22.8019, 8.0055, 0.2286)	(22.7873, 7.9381, 0.2277)
$\widehat{\boldsymbol{\phi}}$	(0.0432, 0.9568)	(0.3828, 0.4293, 0.1878)	(0.3617, 0.4204, 0.1618, 0.0560)
RMS	1.6498	1.7446	1.7271

[19] Zhu, R. and Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning, *J. Time Ser. Anal.*, **27**, 727–738.

[20] Zhu, R. and Joe, H. (2010). Negative binomial time series models based on expectation thinning operators, *J. Stat. Plann. Inf.*, **140**, 1874–1888.

Table 7: ML parameter estimates and RMS for different models for the data from the 58th police station..

Model	MLE	RMS
PoINAR(1)	$\widehat{\lambda} = 2.2349$	
	$\widehat{\alpha} = 0.2189$	3.4011
GPQINAR(1)	$\widehat{\lambda} = 1.0578$	
	$\widehat{\theta} = 0.541$	
	$\widehat{\rho} = 0.17$	3.4624
GINAR(1)	$\widehat{q} = 0.7449$	
	$\widehat{\alpha} = 0.1342$	3.4629
NGINAR(1)	$\widehat{\mu} = 2.9157$	
	$\widehat{\alpha} = 0.1734$	3.4315
NBINAR(1)	$\widehat{q} = 0.2188$	
	$\widehat{\theta} = 0.8033$	
	$\widehat{\alpha} = 0.1155$	3.4789
NBIINAR(1)	$\widehat{n} = 1$	
	$\widehat{p} = 0.5$	
	$\widehat{\rho} = 0.5$	3.4184
NBRCINAR(1)	$\widehat{n} = 0.8442$	
	$\widehat{p} = 0.2327$	
	$\widehat{\rho} = 0.1827$	3.4247
R2NGINAR(1)	$\widehat{\mu}_1 = 1.5485$	
	$\widehat{\mu}_2 = 9.1053$	
	$\widehat{\alpha} = 0.0521$	2.0096
R3NGINAR(1)	$\widehat{\mu}_1 = 0.8719$	
	$\widehat{\mu}_2 = 6.0089$	
	$\widehat{\mu}_3 = 14.3936$	
	$\widehat{\alpha} = 0.3012$	1.3361
CGINAR(2)	$\widehat{\mu} = 2.9524$	
	$\widehat{\alpha} = 0.3232$	3.3403
CGINAR(3)	$\widehat{\mu} = 2.9326$	
	$\widehat{\alpha} = 0.395$	3.2912
CGINAR(4)	$\widehat{\mu} = 2.9517$	
	$\widehat{\alpha} = 0.4241$	3.2769
PoINAR(2)	$\widehat{\lambda} = 1.8966$	
	$\widehat{\alpha} = 0.3534$	3.3152
PoINAR(3)	$\widehat{\lambda} = 1.5812$	
	$\widehat{\alpha} = 0.4779$	3.2438
PoINAR(4)	$\widehat{\lambda} = 1.4452$	
	$\widehat{\alpha} = 0.5521$	3.2236

Table 8: ML parameter estimates and RMS for R2NGINARmax(p) and R3NGINARmax(p) process for the data from the 58th police station.

r	2	2	2
p	2	3	4
$\widehat{\alpha}$	0.1328	0.1322	0.1317
$\widehat{\boldsymbol{\mu}}$	(8.3430, 1.2408)	(8.3422, 1.2347)	(8.3420, 1.2307)
$\widehat{\boldsymbol{\phi}}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0967 & 0.9033 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0511 & 0.9489 & 0.0000 \\ 0.3366 & 0.3414 & 0.3220 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0574 & 0.9426 & 0.0000 & 0.0000 \\ 0.3290 & 0.3340 & 0.3371 & 0.0000 \\ 0.2690 & 0.2194 & 0.2450 & 0.2667 \end{bmatrix}$
RMS	2.0759	2.0785	2.0821
r	3	3	3
p	2	3	4
$\widehat{\alpha}$	0.0488	0.0488	0.0488
$\widehat{\boldsymbol{\mu}}$	(4.6800, 12.8180, 0.6746)	(4.6798, 12.8181, 0.6747)	(4.6798, 12.8181, 0.6748)
$\widehat{\boldsymbol{\phi}}$	$\begin{bmatrix} 1.0000 & 0.0000 \\ 0.0483 & 0.9517 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0154 & 0.9846 & 0.0000 \\ 0.3388 & 0.3363 & 0.3250 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0070 & 0.9930 & 0.0000 & 0.0000 \\ 0.3341 & 0.3295 & 0.3364 & 0.0000 \\ 0.2557 & 0.2094 & 0.2404 & 0.2945 \end{bmatrix}$
RMS	1.1813	1.1764	1.1795

Table 9: ML parameter estimates and RMS for R2NGINAR$_1$(p) and R3NGINAR$_1$(p) process for the data from the 58th police station.

r	2	2	2
p	2	3	4
$\widehat{\alpha}$	0.1328	0.1321	0.1321
$\widehat{\boldsymbol{\mu}}$	(8.3430, 1.2408)	(8.3425, 1.2345)	(8.3421, 1.2340)
$\widehat{\boldsymbol{\phi}}$	(0.0967, 0.9033)	(0.3363, 0.4347, 0.2290)	(0.3283, 0.3309, 0.2725, 0.0683)
RMS	2.0536	2.0732	2.0784
r	3	3	3
p	2	3	4
$\widehat{\alpha}$	0.0488	0.0491	0.0491
$\widehat{\boldsymbol{\mu}}$	(4.6800, 12.8180, 0.6746)	(4.6801, 12.8181, 0.6782)	(4.6804, 12.8181, 0.6789)
$\widehat{\boldsymbol{\phi}}$	(0.0483, 0.9517)	(0.3193, 0.4168, 0.2640)	(0.2732, 0.2719, 0.25489, 0.2001)
RMS	1.1889	1.1760	1.1793