Journal of Crop and Weed, 12(2):96-101(2016)

Application of INAR model on the pest population dynamics in Agriculture

H. S. ROY, R. K. PAUL, L. M. BHAR AND P. ARYA

ICAR-Indian Agricultural Statistics Research Institute,
New Delhi, India

Received: 17-06-2016; Revised: 25-06-2016; Accepted: 29-06-2016

ABSTRACT

In real world phenomenon, some situations often occur where observations are not continuous. Such situations in time series may be number of road accidents, number of patients hospitalized, number of phone calls, the number of products sold, etc. In agriculture, it may be number of pests emerged in a crop season, number of crops infected by a certain type of disease. In these situations, usual autoregressive (AR) model where errors follow Gaussian distribution is not applicable. Hence there is a need to develop models for count observations. The frequently used model for count observation is Poisson regression. In this paper, Integer valued autoregressive (INAR) model is applied for prediction of count observations. Here two real data sets are applied where prediction has been done on pest populations. Finally INAR model is compared with Poisson regression and it is seen that INAR model gives better forecast than the Poisson regression model.

***Keywords*:** Autoregressive model, forecast, Gaussian distribution, integer valued autoregressive model, Poisson regression, time series.

In real world condition we may face different situations where observations are not continuous. In time series framework, we may often found correlated count observations. Hence there is a need to develop models and describe the time series mathematically for such type of observations. Such time series may appear in many real life situations such as number of accidents, number of patients hospitalized, number of phone calls, number of customers arrived or departed, crime victimization, number of products sold, *etc*. Development of models and statistical inferences for this kind of integer valued time series observations is divided into two parts- Poisson regression models and Integer valued autoregressive moving average (ARMA) models.

Arya *et al*. (2015) applied ARIMAX model forpredicting pest population using weather variables. When realization of a process contains very large integers, then we can approximate it by continuous standard autoregressive process, but in practice it does not happen. So, there is a need to use integer valued models. To achieve it Markov chains and its properties (Cox and Miller, 1965) were applied, but the developed models were over parameterized and have limitation in terms of their correlation structure. Then Jacob and Lewis (1978) introduced the discrete autoregressive moving average (DARMA) models, by applying ARMA models. After these, various approaches were made by several authors for fitting these kinds of models. Finally, some results with good forecast values were obtained using Integer Valued Autoregressive models (INAR). Mainly, the INAR models were based on thinning operators. Binomial thinning operator and its generalization were first introduced by Stuetel and Van Harn (1979). These thinning operations were widely implemented for developing INAR models.

McKenzie (1985) developed Binomial autoregressive model for binomial count observations and the structure of model is well-interpretable. For stationarysequence of count observations, Al-Osh and Alzaid (1987) first introduced integer-valued random variables for lag-one that is known as INAR (1) process or Poisson INAR(1) process. They showed that this model is most suitable for count observations. They also showed that the distributional properties and correlation structure of the model are similar to the continuous valued autogressive or AR(1) process. Different estimation procedures such as maximum likelihood estimation (MLE), conditional least squares (CLS) and Yule-Walker (YW) method were also described.Alzaid and Al-Osh (1990) extended INAR(1) process up to the p^{th} order which is useful for modelling discrete-time dependent counting process. They showed the difference with the Gaussian AR (p) process in terms of correlation, Markovian property and regression.

Considering the autocorrelation and discrete nature of the data into account, Thyregod *et al*. (1999) developed a method for modelling the dynamics of rain sampled by a tipping bucket rain gauge. Different models with varying number of lags were presented and the estimation procedure of the models was also described. Karale and Sharma (2014) investigated probability models for explaining population dynamics of major insectpests under rice-potato-okra cropping system.

Bu and McCabe (2008) developed estimation and model selection procedure for a series of integer valued autoregressive models for any number of lags. Estimation was done by performing the likelihood methods. They have introduced a new method for model selection based on residuals. The forecasts were obtained

*Email:*himadriiasri@gmail.com

by treating the model as a Markov Chain and error estimate was done along with their confidence intervals. Pavlopoulos and Karlis (2008) presented another type of INAR(1) model, which discussed about the non-linear structure of auto-regressive Markov chain on total time length of the series, where error structure follows a finite mixture distribution of Poisson laws. EM-algorithm was used to make inference by maximization of a conditional likelihood. Parametric bootstrap approach for integer-valued prediction (IP) method was developed and certain test statistic based on the predictions were discussed for assessing performance of the fitted model.

Enciso-Mora *et al*. (2009) developed INAR processes which are perfectly suited for modelling count data including the explanatory variables into the model. An efficient MCMC algorithm was constructed to analyze the model and incorporates both explanatory variables and order selection.

For count data analysis, integer valued time-series occurs in many practical situations, especially when low frequency discrete data appears but their continuous approximations are not suitable (Freeland and McCabe, 2003). This motivated Neal and Rao (2005) to construct integer valued ARMA (INARMA) processes.

In this paper we have introduced the INAR (1) model. Here our prime importance is to predict pest population dynamics in agriculture. Here two real data sets were analyzed by INAR(1) model and compared with Poisson regression. It is seen that for studying discrete observation in time series framework INAR model is most suitable.

MATERIALS AND METHODS

Suppose $\{X_t\}$ $t = 0, 1, 2, \ldots, n$ be a count-data time series with a ûnite range $\{0, \ldots, n\}$ of counts, where $n \in N = \{1, 2, \ldots\}$ is known and the series has serial dependence similar to the Gaussian autoregressive (AR) process. If the marginal distribution follows binomial distribution, *i.e.*, B(n, p) where $p \in (0; 1)$ is called Binomial AR (1) model, was firstproposed by McKenzie (1985). Particular case of the binomial AR(1) model for describing binomial counts with a first-order autoregressive serially correlated structure was described by Weib and Kim (2013). Asymptotic distribution of the conditional least-squares estimators of the parameters of the binomial AR(1) model were also discussed.

INAR (1) model

Integer valued autoregressive model for first order, INAR (1) was first introduced by McKenzie (1985, 1988) and Al-Osh and Alzaid (1987) independently for modelling and forecasting the sequences of dependent counting process.

If $\{X_t\}$ follows a discrete non-negative integer valued stochastic process then the INAR (1) model is given by the following stochastic difference equation

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

where $\alpha > 0$, thinning operation is represented by '$*$' symbol and $\{\varepsilon_t\}$ is a sequence of non-negative integer-valued *i.i.d* random variables with mean μ_e and variance σ_e^2.

The INAR (1) model can be interpreted as follows:

$$\underbrace{X_t}_{\textit{Population at time t}} = \underbrace{\alpha * X_{t-1}}_{\textit{Survivors of time t-1}} + \underbrace{\varepsilon_t}_{\textit{Immigration}}$$

By using thinning operation, X_t would be interpreted as the number of new individuals. So INAR (1) process can be compared with Galton-Watson process with immigration. McKenzie (1988) compared INAR (1) process with queueing system. So by using these interpretations many authors have used INAR (1) models in many applications. As an example X_t describe the number of customers, ε_t describes new customers and the customers who have been lost at the end of the last period is given by $X_{t-1} - \alpha * X_{t-1}$.

Using the properties of binomial thinning operator the mean and variance of the stationary INAR (1) model for $\{X_t\}$ is given by

$$\mu_X = E(X_t) = \frac{\mu_\varepsilon}{1-\alpha} \quad \text{and} \quad \sigma_X^2 = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2}$$

Pavlopoulos and Karlis (2008) proposed the equation of linear prediction of INAR model for k number of lags, which is given by

$$\hat{X}_{t+k} = \alpha^k . X_t + \left(1-\alpha^k\right)\frac{\mu_\varepsilon}{1-\alpha}$$

The above equation is the linear prediction of the most recent observation X_t. The mean and variance are given by

$$E\left(\hat{X}_{t+k}\right) = \alpha^k . \mu + \left(1-\alpha^k\right)\frac{\mu_\varepsilon}{1-\alpha} \text{ and } Var\left(\hat{X}_{t+k}\right) = \alpha^{2k} . \sigma^2$$

ILLUSTRATION

In this section we analyze an application of the introduced INAR model to predict the pest population in agriculture. The data sets were collected on jassid and whitefly population on Bt cotton crop during the period from 2008 to 2012 at different meteorological weeksfrom the Guntur district of Andhra Pradesh from farmer's field. The data was analyzed by using R software. The variable under study is taken as the average number of pest on 3 leaves (randomly selected). The data consists of 110 meteorological weeks.

According to the season of pest occurrence, 22 meteorological weeks (32nd to 52nd week of a year and 1st week of the successive year) are considered in each year. In case of Jassid population, the maximum number of pest was recorded as 5 which occurred at the 39th week of 2010-11. In case of Whitefly, the maximum number of pest was recorded as 13 which occurred at 45th week of year 2010-11.It was found that the variability (*i.e.*, coefficient of variation) of Jassid population (63%) is much lower than that of the Whitefly population (125%). In case of Jassid population, mean (1.52) is greater than the variance (0.93) but in case of Whitefly population, the variance (9.38) is greater than the mean (2.45). Descriptive statistics of the data sets are given in table 1. The graphical representation of time series plot, autocorrelation function and partial autocorrelation function are shown in figure 1 and 2. A perusal of the Fig. 1 indicates that Jassid population shows nearly uniform upward and downward growth pattern over each year. Fig. 2 indicates that whitefly population does not show uniform growth pattern over each year. Fig.1 and 2 also show positive autocorrelation over time for both pest population and it also reveals that the data are stationary in nature.

Table 1: Descriptive statistics of data sets

Pest	Min.	Max.	Mode	Median	Mean	Variance	CV
Jassid	0	5	1	1	1.52	0.93	0.63
Whitefly	0	13	0	1	2.45	9.38	1.25

RESULTS AND DISCUSSION

The INAR (1) model is fitted with observations of the two data sets separately and using Yule-Walker method the estimated parameters are found to be $\lambda = 0.32$, $\alpha = 0.80$ for Jassid population and $\lambda = 0.98$, $\alpha = 0.62$ for Whitefly population. Prediction of pest population for both data sets is done using INAR (1) model and then using Poisson regression model. Table 2 shows comparative study of Poisson regression along with INAR (1) model. The meteorological weeks 49, 50, 51, 52 and 1 are considered for model validation for both the pests. For Jassid at 49th week the actual observation is 2 but INAR (1) prediction is 1.90 and Poisson regression prediction is 1.27. Similarly for 50th week actual number of observation is 2, INAR (1) prediction is 1.90 and Poisson regression prediction is 1.14. For 51st week, actual observation is 3 and INAR (1) and Poisson regression prediction is 2.79 and 0.90 respectively. Similar observations were in case of Whitefly. From the comparative study, we can say that INAR model performs better than Poisson regression model. A perusal of the Figure 3 indicates that the INAR (1) model fits the data quite satisfactorily. Root mean square error (RMSE) and Akaike Information Criteria (AIC) values are calculated for comparison of INAR (1) and Poisson regression model and these values are shown in table 3. The AIC values of INAR model are 237.62 and 420.81 for Jassid and Whitefly population respectively. The AIC values of Poisson regression model are 302.95 and 490.68 for Jassid and Whitefly population respectively. Similarly the RMSE values of INAR (1) model are 0.16 and 0.46 for Jassid and Whitefly population respectively. The RMSE Values of Poisson regression are 0.98 and 1.31 for Jassid and Whitefly population respectively. For both the data sets RMSE values and AIC values of INAR (1) model is less than the Poisson regression model.

Table 2: Comparison of actual observations versus predicted observations by INAR (1) and Poisson regression

Pests	Comparison	Meteorological weeks of 2013				
		49th	50th	51st	52nd	1st
Jassid	Actual observation	2	2	3	2	1
	INAR(1) prediction	1.90	1.90	2.70	1.91	1.12
	Poisson reg. prediction	1.27	1.14	0.90	1.12	1.07
Whitefly	Actual observation	3	2	3	2	2
	INAR(1) prediction	3.48	2.2	2.85	2.85	2.23
	Poisson reg. prediction	2.42	4.84	5.11	3.57	4.66

Table 3: RMSE and AIC values of INAR (1) model and Possion regression model

Pest		RMSE	AIC
Jassid	INAR(1)	0.16	237.62
	Poisson reg.	0.98	302.95
Whitefly	INAR(1)	0.46	420.81
	Poisson reg.	1.31	490.68

For count data analysis, the most frequently used two models are INAR model and Poisson regression model. In the present investigation, INAR (1) model was introduced. This paper also discusses the predictability of INAR (1) model in real data sets.Analysis was done on two real data sets separately. A comparative study of Poisson regression along with INAR (1) model has been carried out in order to predict the pest populations. Finally the investigation reveals that for predicting the discrete time series observations INAR (1) model is most suitable.

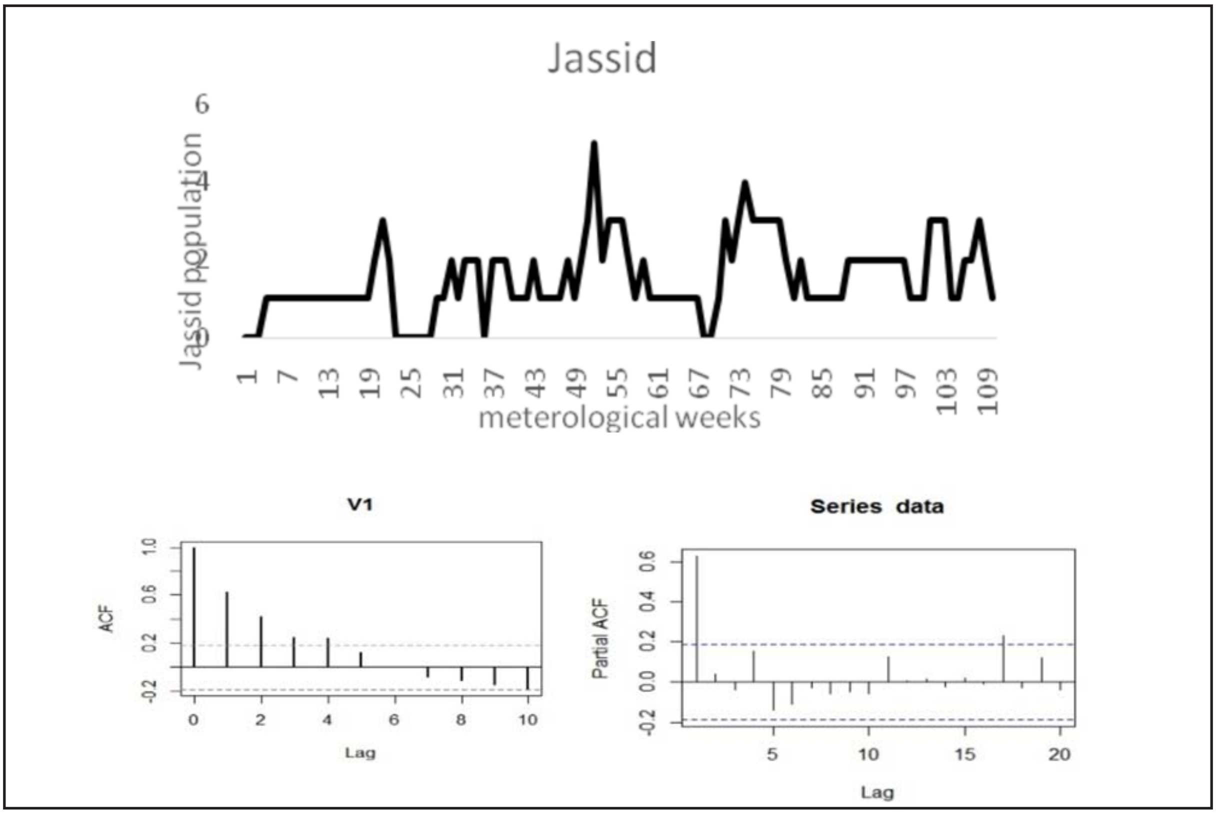


Fig.1: Time series plot, sample autocorrelation function and partial autocorrelation function of jassid population

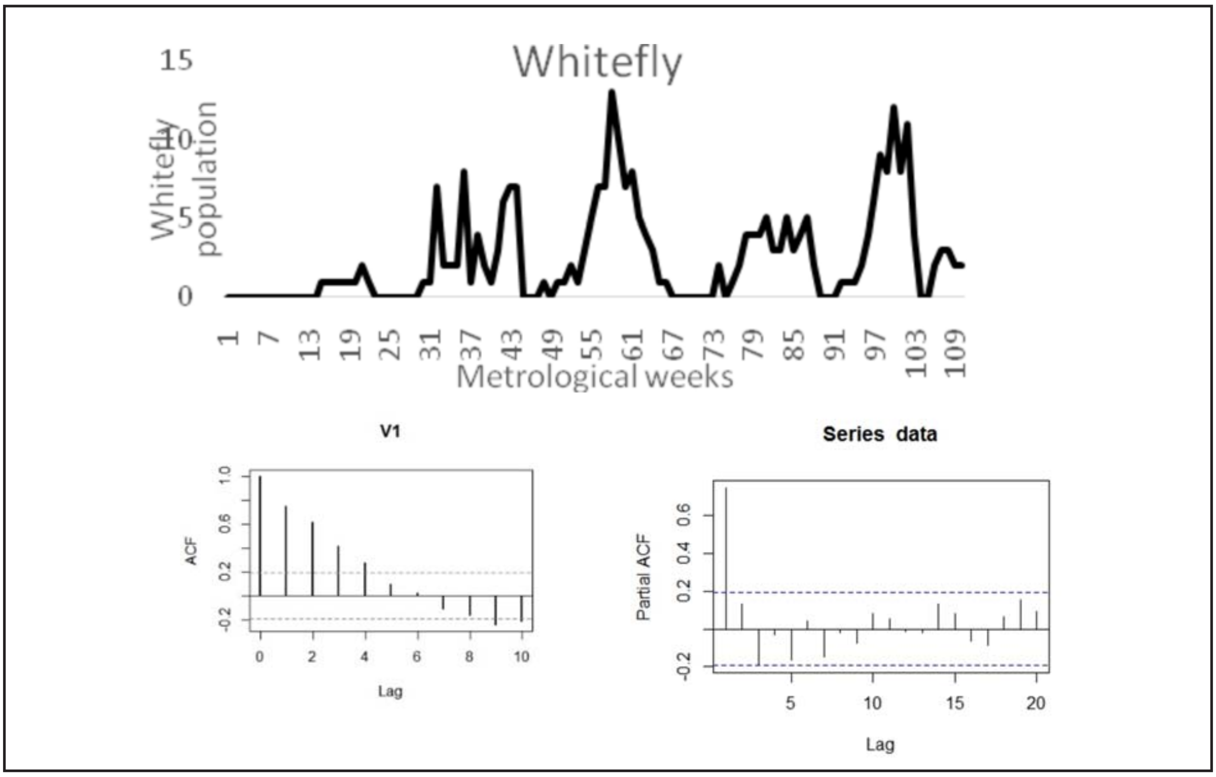
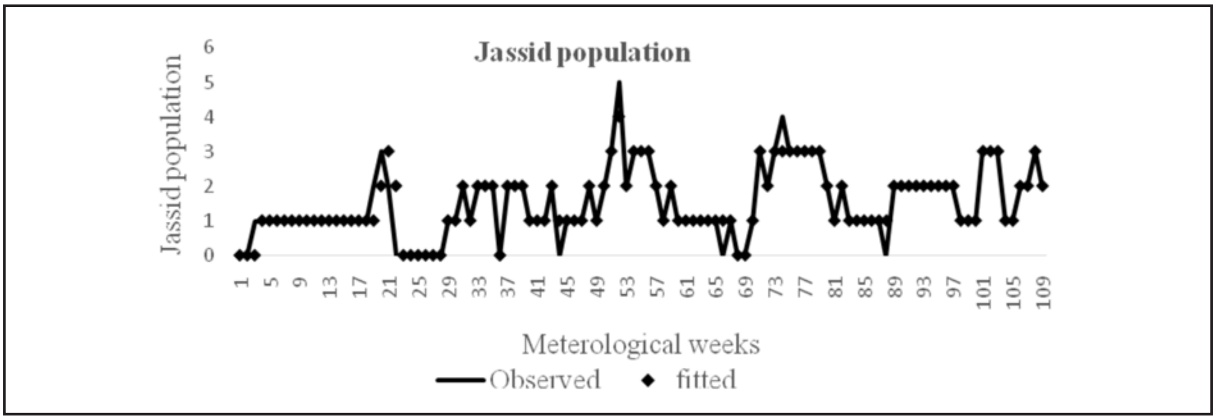


Fig. 2: Time series plot, sample autocorrelation function and partial autocorrelation function of whitefly population



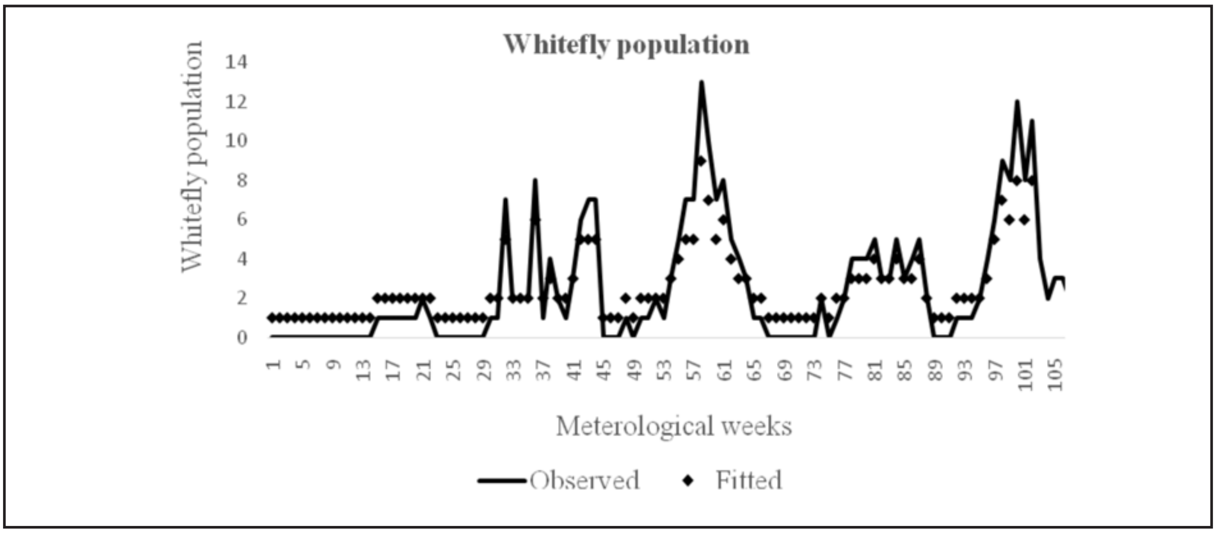


Fig. 3: Observed and fitted values of Jassid and Whitefly populations by using INAR(1) model

REFERENCES

Al-Osh, M. A. and Alzaid, A. A. 1987. First order integer-valued autoregressive INAR(1) process. *J. Time Series Anal.*, **8**: 261-75.

Alzaid, A. A. and Al-Osh, M. A. 1990. An integer-valued *p*th-order autoregressive structure (INAR(*p*)) process. *J. App. Prob.*, **27:** 314–24.

Arya, P., Paul, R. K., Kumar, A., Singh, K. N., Sivaramne, N. and Chaudhary, P. 2015. Predicting pest population using weather variables: an ARIMAX time series framework. *Int. J. Agric. Stat. Sci.*, **11**(2): 381-86.

Brännäs, K., Hellström, J. and Nordström, J. 2002. A new approach to modelling and forecasting monthly guest nights in hotels, *Int. J. Forecasting*, **18**: 19-30.

Bu, R. and McCabe, B. 2008. Model selection, estimation and forecasting in INAR(*p*) models: a likelihood-based Markov chain approach, *Int. J. Forecasting*,**24**: 151–62.

Cox, D.R. and Miller, H.D. 1965. *The Theory of Stochastic Processes*, Chapman and Hall/ CRC, Washington D. C.

Du, J. G. and Y, Li. 1991. The integer-valued autoregressive (INAR(p)) model. *J. Time Series Anal.*, **12**: 129-42.

Enciso-Mora, V., Neal, P. and Rao T. S. 2009. Integer Valued AR processes with explanatory variables, *Sankhya : The Indian J. Stat.*, **71**(B) : 248-63.

Jacobs, P. A. and Lewis, P. A. W. 1978. Discrete time series generated by mixtures I: correlation and runs properties. *J. Royal Stat. Soc.*, **40**(B)**:** 94-105.

Karale, M.and Sharma,H. L. 2014.On probability models for describing population dynamics of major insectpests under rice-potato-okra cropping system, *J. Crop Weed*, **10**:451-56.

McKenzie, E. 1985. Some simple models for discrete variate time series, *Water Resources Bull.*, **21**: 645–50.

McKenzie, E. 1988. Some ARMA models for dependent sequences of Poisson counts. *Adv. App. Prob.*,**20**: 822-35.

Neal, P. and Rao, T. S. 2005. *MCMC for Integer Valued ARMA Processes*, Research Report, **10**, Probability and Statistics Group School of Mathematics, The University of Manchester.

Pavlopoulos, H. and Karlis, D. 2008. INAR(1) modeling of overdispersed count series with an environmental application, *Environmetrics*, **19**: 369–93.

Steutel F. W. and van Harn K. 1979. Discrete Analogues of Self-Decomposability and Stability, *The Annals of Prob.*, **7**: 893-99.

Thyregod, P., Carstensen, J., Madsen, H. and Arnbjerg-Nielsen, K. 1999. Integer valued autoregressive models for tipping bucket rainfall measurements, *Environmetrics*, **10**: 395-411.

Weib, C. H. and Kim, H. 2013. Binomial AR(1) processes: moments, cumulants, and estimation, *Statistics*, **47**: 494-510.