Optimal Alarm Systems for Count Processes

Magda Monteiro
Escola Superior de Tecnologia e Gestão de Águeda
Research Unit 'Matemática e Aplicações'
Universidade de Aveiro
Portugal

Isabel Pereira and Manuel G. Scotto*
Departamento de Matemática
Research Unit 'Matemática e Aplicações'
Universidade de Aveiro
Campo Universitário de Santiago
3810-193 Aveiro, Portugal

Abstract

In many phenomena described by stochastic processes the implementation of an alarm system becomes fundamental to predict the occurrence of future events. In this work we develop an alarm system to predict whether a count process will upcross a certain level and give an alarm whenever the upcrossing level is predicted. We consider count models with parameters being functions of covariates of interest and varying on time. The paper presents classical and Bayesian methodology for producing optimal alarm systems. Both methodologies are illustrated and their performance compared through a simulation study. The work finishes with an empirical application to a set of data concerning the number of sunspot on the surface of the sun.

Keywords: Count processes; Optimal alarm systems; Autoregressive processes

*Manuel G. Scotto, Departamento de Matemática, Universidade de Aveiro. Campo Universitário de Santiago, 3810-193 Aveiro, Portugal. Email: mscotto@ua.pt.

1 Introduction

Currently, a major theme in the analysis of a large variety of random phenomena consists in detecting and warning the occurrence of a catastrophe or some other event connected with an alarm mechanism. Examples range from the prediction of increases in mean sea level and flood frequencies due to global warming arising from increased greenhouse gas concentrations to the assessment of the health impact of air pollution. One way of doing so is by using a *naive* alarm system, based on the predictor $\hat{X}_{t+h} = E\left[X_{t+h}|X_s, -\infty < s \leq t\right]$, for $h > 0$, where an alarm is given every time the predictor exceeds some level. This alarm system, however, is far from being optimal because it does not have a good performance on the ability to detect the events, locate them accurately in time and give as few false alarms as possible. Addressing this issue Lindgren (1975a and b, 1980, 1985) and de Maré (1980) set the principles of optimal prediction of level crossings; later Svensson et al. (1996) applied Lindgren and de Maré's results on the development of optimal alarm systems to predict high water levels in the Baltic. A major drawback of the alarm system introduced by Lindgren and de Maré is that it ignores the sampling variation of the model parameters. In order to overcome this limitation Amaral Turkman and Turkman (1990) suggested a Bayesian approach and particular calculations were carried out for an autoregressive model of order one, although no attempt was made to solve the difficult computational problems involved. More recently, Antunes et al. (2003) extended the results given in Amaral Turkman and Turkman (1990) autoregressive models of order p and show how the alarm characteristics can numerically be obtained.

It is worth to mention that all references given in the previous paragraph deal with the case of continuous-valued processes. A related interesting problem, which has not been addressed yet, is to develop an alarm system for series of counts which are represented through integer-valued autoregressive models. This paper aims at giving a contribution towards this direction.

The analysis of count processes has become an important area of research in the last two decades partially because its wide applicability to experimental biology (Zhou and

Basawa, 2005), social science (McCabe and Martin, 2005), international tourism demand (Nordström 1996, Garcia-Ferrer and Queralt, 1997, Brännäs et al., 2002, and Brännäs and Nordström, 2006), queueing systems (Ahn et al., 2000) and economy (Quoreshi, 2006). We refer to McKenzie (2003) for an overview of the early work in this area. Among the most successful integer-valued time series models proposed in the literature we mention the INteger-valued AutoRegressive model of order p (INAR(p)) and the INteger-valued Moving Average model of order q (INMA(q)). The former was first introduced by McKenzie (e.g., 1985) and Al-Osh and Alzaid (1987) for the case $p = 1$. Empirical relevant extensions have been suggested by Brännäs (1995, explanatory variables), Blundell et al. (2002, panel data), Brännäs and Hellström (2001, extended dependence structure), and more recently by Silva et al. (2005, replicated data). Extensions and generalizations were proposed by Du and Li (1991) and Latour (1988). The INMA(q) model was proposed by Al-Osh and Alzaid (1988) and subsequently studied by Brännäs and Hall (2001). Related models were introduced by Aly and Bouzar (1994, 2005), Zhu and Joe (2003) and more recently by Neal and Subba Rao (2007). Extensions for random coefficients integer-valued autoregressive models have been proposed by Zheng et al. (2006, 2007) who investigate basic probabilistic and statistical properties of these models. Zheng and co-workers illustrate their performance in the analysis of epileptic seizure counts (e.g., Latour, 1988) and in the analysis of the monthly number cases of poliomyelitis in the US for the period 1970-1983.

Potential applications of optimal alarm systems for count processes can be found in the study of short-term effects of environmental factors, such as pollutants (ozone, nitrogen dioxide, etc) and climate variables (pressure, temperature, relative humidity, etc) on mortality (daily or monthly number of deaths). Much of the early work in this subject is based on the use of generalized linear models and generalized additive models using nonparametric techniques. Examples can be found in the study of the relationship between mortality and air pollution (Katsouyanni et al., 2002), hospital admissions and air pollution (Touloumi et al., 2004), atmospheric pressure with mortality (Campbell et al., 2001, Braga et al., 2001), and infectious gastrointestinal illness related to drinking water (Schwartz et al., 1997); see Koop and Tale (2004) for further references. All the above referred references, however, are not directly applicable to predict in advance *future* upcrossings (i.e., a large number of deaths). Is in this context that the implementation of an alarm system reveals to be useful. Similar questions occur when modelling daily or monthly guest nights in hotels (Brännäs et al., 2002 and Brännäs and Nordström, 2006) incorporating the effect of economic variables such as the income level of the country of guest's origin, prices and exchange rates. Again the models proposed by Brännäs and co-workers fail in predicting the probability of a catastrophe such as, for example, the accommodation demand exceeding the capacity of the hotels.

For completeness and reader's convenience background description of basic theoretical concepts related with event prediction are given below. We follow closely Antunes et al. (2003). Their ideas will be extensively used throughout this paper.

Let $\mathbf{X} = (X_t)_{t\in\mathbb{N}}$ be a count process with parameter space $\Theta \subset \mathbb{R}^k$, for some $k \in \mathbb{N}$. The time sequel $\{1, 2, \ldots, t-1, t, t+1, \ldots\}$ is divided in three sections $\{1, 2, \ldots, t-q\}$, $\{t-q+1, \ldots, t\}$ and $\{t+1, \ldots\}$ i.e., the past, present and future such that for some $q > 0$ the sets $D_t = \{X_1, X_2, \ldots, X_{t-q}\}$, $\mathbf{X_2} = \{X_{t-q+1}, \ldots, X_t\}$ and $\mathbf{X_3} = \{X_{t+1}, \ldots\}$ represent respectively the informative experiment, the present experiment and the future experiment at time t. Any event of interest, say $C_{t,j}$, in the σ-field generated by $\mathbf{X_3}$ is defined as a catastrophe. Throughout this work a catastrophe will be considered as the upcrossing event $C_{t,j} = \{X_{t+j-1} \leq u < X_{t+j}\}$, for some $j \in \mathbb{N}$. Moreover, any event $A_{t,j}$ in the σ-field generated by $\mathbf{X_2}$, predictor of $C_{t,j}$, will be alarm region. It is said that an alarm is given at time t, for the catastrophe $C_{t,j}$, if the observed value of $\mathbf{X_2}$ belongs to the alarm region. In addition, the alarm is said to be correct if the event $A_{t,j}$ is followed by the event $C_{t,j}$. Conversly, a false alarm is defined as the occurrence of $A_{t,j}$ without $C_{t,j}$. If an alarm is given when the catastrophe occurs, it is said that the catastrophe is detected. Furthermore the alarm region $A_{t,j}$ is said to have size $\alpha_{t,j}$ if $\alpha_{t,j} = P(A_{t,j}|D_t)$. The alarm region is optimal of size $\alpha_{t,j}$ if

$$P(A_{t,j}|C_{t,j}, D_t) = \sup_{B\in\sigma_{\mathbf{X_2}}} P(B|C_{t,j}, D_t), \tag{1}$$

with $P(B|D_t) = \alpha_{t,j}$. Note that this alarm region also is the supreme, among all events in $\sigma_{\mathbf{X_2}}$, of the probability of correct alarm; see Lemma 1 of Antunes et al. (2003) for details.

Definition 1.1. *An optimal alarm system of size* $\{\alpha_{t,j}\}$ *is a family of alarm regions* $\{A_{t,j}\}$ *in time satisfying* (1).

Lemma 1.1 below is a slight modification of Lemma 2 in Antunes et al. (2003) for count processes.

Lemma 1.1. *Let* $p(\boldsymbol{x}_2|D_t)$ *and* $p(\boldsymbol{x}_2|C_{t,j}, D_t)$ *be the predictive probabilities of* $\boldsymbol{X}_2$ *and* $\boldsymbol{X}_2|C_{t,j}$*, respectively. The alarm system* $\{A_{t,j}\}$ *defined by*

$$A_{t,j} = \{\boldsymbol{x}_2 \in \mathbb{N}^q : \frac{P(C_{t,j}|\boldsymbol{x}_2, D_t)}{P(C_{t,j}|D_t)} \geq k_{t,j}\},$$

for a fixed $k_{t,j} : P(\boldsymbol{X}_2 \in A_{t,j}|D_t) = \alpha_{t,j}$*, is optimal of size* $\alpha_{t,j}$*.*

This lemma ensures that the alarm region defined above renders the highest detection probability. Moreover to enhance the fact that the optimal alarm system depends on the choice of $k_{t,j}$, it is important to stress that in view of the fact that $P(C_{t,j}|D_t)$ does not depend on $\mathbf{x}_2$, the alarm region can be rewritten in the form

$$A_{t,j} = \{\mathbf{x}_2 \in \mathbb{N}^q : P(C_{t,j}|\mathbf{x}_2, D_t) \geq k\}, \tag{2}$$

where $k = k_{t,j}P(C_{t,j}|D_t)$ is chosen in some optimal way to accommodate conditions over the following operating characteristics of the alarm system:

Definition 1.2 [Operating characteristics]

1. *Size of the alarm:* $P(A_{t,j}|D_t)$*;*
2. *Probability of correct alarm:* $P(C_{t,j}|A_{t,j}, D_t)$*;*
3. *Probability of detecting the event:* $P(A_{t,j}|C_{t,j}, D_t)$*;*
4. *Probability of false alarm:* $P(\bar{C}_{t,j}|A_{t,j}, D_t)$*;*
5. *Probability of undetected the event:* $P(\bar{A}_{t,j}|C_{t,j}, D_t)$*.*

The choice of k must be such that maximizes the probabilities of correct alarm and detection and simultaneously minimizes the probabilities of non detection and false alarm. In view of the fact that simultaneous maximization of the probability of correct alarm and the probability of detection it is not possible in general, a compromise must be reached between those operating characteristics. Svensson et al. (1996), for example, suggest to choose the value of k that corresponds to the equality of the above referred probabilities. In this work a different approach will be adopted. Details are given in Section 4.

The rest of the paper is organized as follows: in Section 2 an optimal alarm system for a Doubly Stochastic INteger-valued AutoRegressive (DSINAR) process of order one is developed. Expressions for the probability of the alarm size, correct alarm and the probability of detecting the catastrophe are given. Parameter estimation from both classical and Bayesian approaches is covered in Section 3. In Section 4, the results are illustrated through a simulation study. Section 5 gives and empirical example of a set of data concerning the number of sunspot on the surface of the sun. Finally, some concluding remarks are given in Section 6.

In this paper we want to highlight the following issues:

1. whereas for the continuous-valued models considered by Svensson et al. (1996) the percentages of correct alarm, when considering as the upcrossing level an extremal event, is nearly 50% in the discrete case this percentage falls to 30% at least for the models under consideration. Our results are not directly comparable with those obtained by Antunes et al. (2003) since the authors have considered as upcrossings non-extremal level events;

2. there is no unifying criterion to choose an optimal alarm system. The ones considered by Antunes et al. (2003) and Svensson et al. (1996) do not directly apply in the present setting mainly due to the discrete nature of the data. Thus a different approach has to be adopted;

3. the event prediction considered in this work allow us *on-line prediction* in the sense that, the parameter estimates of the model, the alarm regions and the operating

characteristics are updated at each time point.

2 Optimal Alarm Systems for DSINAR(1) Processes

In this section an optimal alarm system for a Doubly Stochastic INteger-valued AutoRegressive processes of order one (hereafter DSINAR(1)) is developed. A DSINAR(1) model is defined by the recursive equation

$$X_t = \alpha_t \circ X_{t-1} + Z_t, \qquad t = 1, 2, \ldots, \tag{3}$$

where the *thinning* operator $\circ$ is defined as

$$\alpha_t \circ X_{t-1} \overset{d}{=} \sum_{i=1}^{X_{t-1}} U_{i,t}(\alpha_t),$$

being $(U_{i,t}(\alpha_t))$, for $i = 1, 2, \ldots$, an i.i.d. sequence of Bernoulli random variables with success probability $P(U_{i,t}(\alpha_t) = 1) = \alpha_t$. Furthermore $(Z_t)_{t\in\mathbf{N}}$ constitutes an i.i.d. sequence of Poisson-distributed random variables with mean λ, which are assumed to be independent of X_{t-1}, α_t and $\alpha_t \circ X_{t-1}$. Note that the operator $\circ$ incorporates the discrete nature of the variates and acts as the analogue of the standard multiplication used in the continuous-valued processes. We further assume that $X_0 = x_0$, is observed. For α_t the conventional specification

$$\alpha_t = \frac{e^{\mathbf{Y}_{t-s}\boldsymbol{\omega}}}{1 + e^{\mathbf{Y}_{t-s}\boldsymbol{\omega}}}, \; s \geq 0,$$

is adopted with $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t}, \ldots, Y_{l,t})$ being a vector of covariates of interest, and $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_l)^T$ the corresponding unknown vector of parameters. It is worth to mention that although for the large majority of processes involving series of counts the assumption of $s > 0$ is tenable, the motivation for considering the case $s = 0$ comes from the fact that in the analysis of international tourism demand and other related processes this assumption is more appropriate in terms of modelling purposes; see Brännäs et al. (2002) and Brännäs and Nordström (2004, 2006).

For the DSINAR(1) model the probabilistic computation of the operating characteristics is as follows: the probability of catastrophe conditional on D_t and $\mathbf{X_2}$, i.e., $P\left(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right)$ and the probability of catastrophe conditional on D_t, $P(C_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})$ (hereafter P_1

and P_2 respectively) with $\mathsf{S}_{\mathbf{Y}}$ defined as the σ-field generated by $(\mathbf{Y}_t)_{t\in\mathbb{N}}$, can be obtain through the calculation of the set of conditional probabilities $P\left(x_{t+h}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right)$, with $h \in \mathbb{N}$. In order to obtain $P\left(x_{t+h}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right)$ we need the following result.

Proposition 2.1. *For the DSINAR(1) model defined in* (3) *it follows that*

$$X_{t+h}|\mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega} \overset{d}{=} \left(\prod_{i=0}^{h-1} \alpha_{t+h-i}\right) \circ X_t + \left[Z_{t+h} + \sum_{j=1}^{h-1} \left(\prod_{m=0}^{j-1} \alpha_{t+h-m}\right) \circ Z_{t+h-j}\right], \tag{4}$$

with the convention $\sum_{j=1}^{0} = 0$.

Proof. See Appendix A. □

The expression on the right-hand side of (4) is the sum of two independent components, one depending only on X_t and the other depending only on the innovations. Having in mind the properties of the thinning operator it is easy to check that the 2nd term in brackets is Poisson-distributed with parameter $\lambda^* = \lambda \sum_{j=0}^{h-1} \beta_{t+h,j}$ being

$$\beta_{t,s} = \begin{cases} \prod_{m=0}^{s-1} \alpha_{t-m} & s \geq 1 \\ 1 & s = 0 \end{cases}.$$

Moreover, the first term on the right-hand side of (4) conditioned on $X_t = x_t$ is Binomial-distributed with parameters x_t and $\beta_{t+h,h}$. Hence the probability function of X_{t+h} conditional on X_t and $\mathsf{S}_{\mathbf{Y}}$ can be written as

$$P(x_{t+h}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \;=\; e^{-\lambda^*} \sum_{i=0}^{\min(x_t, x_{t+h})} \frac{(\lambda^*)^{x_{t+h}-i}}{(x_{t+h}-i)!} C_i^{x_t} \left(\beta_{t+h,h}\right)^i \left(1-\beta_{t+h,h}\right)^{x_t - i}, \tag{5}$$

providing

$$\begin{aligned}
&P\left(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right) = \\
&\quad = P(X_{t+j-1} \leq u < X_{t+j}|x_1, \ldots, x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \\
&\quad = \sum_{x_{t+j-1}=0}^{u} P\left(X_{t+j-1} = x_{t+j-1}, X_{t+j} > u|\mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right) \\
&\quad = \sum_{x_{t+j-1}=0}^{u} \left[P\left(X_{t+j-1} = x_{t+j-1}|\mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right) \times P\left(X_{t+j} > u|x_{t+j-1}, \mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right)\right] \\
&\quad = \sum_{x_{t+j-1} \leq u} p(x_{t+j-1}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \times \left(1 - \sum_{x_{t+j} \leq u} p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})\right).
\end{aligned} \tag{6}$$

Finally, P_2 can be calculated through the expression

$$
\begin{aligned}
&P(C_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) = \\
&= P(X_{t+j-1} \leq u < X_{t+j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \\
&= \sum_{x_{t+j-1}=0}^{u} P(X_{t+j-1} = x_{t+j-1}, X_{t+j} > u|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \\
&= \sum_{x_{t+j-1}=0}^{u} [P(X_{t+j-1} = x_{t+j-1}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \times P(X_{t+j} > u|x_{t+j-1}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})] \\
&= \sum_{x_{t+j-1} \leq u} p(x_{t+j-1}|x_{t-q}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \times \left(1 - \sum_{x_{t+j} \leq u} p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})\right). \qquad (7)
\end{aligned}
$$

Since $\mathbf{X}$ is Markovian the probabilities $p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})$ and $p(x_{t+j-1}|x_1, \ldots, x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})$ are calculated through (5), after making the necessary adaptations. With these preliminaries out of the way, the operating characteristics can now be easily calculated.

1. **Alarm size**

$$
\begin{aligned}
P(A_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) &= \sum_{\mathbf{x_2} \in A_{t,j}} P(\mathbf{X_2} = \mathbf{x_2}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \\
&= \sum_{\mathbf{x_2} \in A_{t,j}} \prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega});
\end{aligned}
$$

2. **Probability of correct alarm**

$$
\begin{aligned}
&P(C_{t,j}|A_{t,j}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) = \\
&= \frac{P(C_{t,j} \cap A_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}{P(A_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})} \\
&= \frac{\sum_{\mathbf{x_2} \in A_{t,j}} P(\mathbf{X_2} = \mathbf{x_2}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) P(C_{t,j}|\mathbf{X_2} = \mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}{P(\mathbf{X_2} = \mathbf{x_2}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})} \\
&= \frac{\sum_{\mathbf{x_2} \in A_{t,j}} \prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) P(C_{t,j}|\mathbf{X_2} = \mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}{\sum_{\mathbf{x_2} \in A_{t,j}} \prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}.
\end{aligned}
$$

The expression of the alarm size along with the expression in (6) allow us to rewrite the probability of correct alarm as

$$
\begin{aligned}
&P(C_{t,j}|A_{t,j}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) = \\
&= \sum_{\mathbf{x_2}\in A_{t,j}} \left[\prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \sum_{x_{t+j-1}=0}^{u} p\left(x_{t+j-1}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right) \times \right. \\
&\times \left. \left(1 - \sum_{x_{t+j}=0}^{u} p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})\right)\right] \left[\sum_{\mathbf{x_2}\in A_{t,j}} \prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \right]^{-1};
\end{aligned}
$$

3. **Probability of detecting the catastrophe**

$$
\begin{aligned}
&P(A_{t,j}|C_{t,j}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) = \\
&= \frac{P(C_{t,j} \cap A_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}{P(C_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})} \\
&= \frac{\sum_{\mathbf{x_2}\in A_{t,j}} P(\mathbf{X_2} = \mathbf{x_2}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) P(C_{t,j}|\mathbf{X_2} = \mathbf{x_2}, D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})}{P(C_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})} \\
&= \sum_{\mathbf{x_2}\in A_{t,j}} \left[\prod_{i=0}^{q-1} p(x_{t-i}|x_{t-i-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \sum_{x_{t+j-1}=0}^{u} p\left(x_{t+j-1}|x_t, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}\right) \times \right. \\
&\times \left. \left(1 - \sum_{x_{t+j}=0}^{u} p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})\right)\right] \times \\
&\times \left[\sum_{x_{t+j-1}\le u} p(x_{t+j-1}|x_{t-q}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega}) \left(1 - \sum_{x_{t+j}\le u} p(x_{t+j}|x_{t+j-1}, \mathsf{S}_{\mathbf{Y}}, \lambda, \boldsymbol{\omega})\right)\right]^{-1}.
\end{aligned}
$$

3 Estimation methods

In this section we consider the estimation of the operating characteristics. From the classical framework an estimative method (plug-in) is used to estimate these probabilities based on the the well-known conditional maximum likelihood (CML) method. The CML estimates of the unknown parameters $\boldsymbol{\theta} = (\lambda, \boldsymbol{\omega})$ are obtained maximizing the conditional log-likelihood function with respect to $\boldsymbol{\theta}$, recurring to the iterative Newton-Raphson method. The starting values needed to initialize the algorithm are the conditional least squared estimates. *Moreover, the values of Y_{t+j} are calculated through the optimal linear j-period ahead forecast given by $\hat{Y}_{t+j} = E(Y_{t+j}|\mathbf{y})$, being $\mathbf{y} = (y_t, y_{t-1}, \ldots)$.*

From the Bayesian perspective a prior distribution for the vector of parameters $\boldsymbol{\theta}$ is needed. This distribution is intended to represent beliefs about parameter values, prior to the availability of data. We consider that $\lambda \sim Gamma(a, b)$, $a, b > 0$ and $\omega_i \sim N(\mu_i, \tau_i^{-1})$ with $\mu_i \in \mathbb{R}$ and $\tau_i > 0$, $i = 1, \ldots, l$. Assuming independence between all the parameters involved the prior distribution of $\boldsymbol{\theta}$, say $h(\boldsymbol{\theta})$, is proportional to

$$h(\boldsymbol{\theta}) \propto \frac{1}{\prod_{i=1}^{l} \sqrt{2\pi\tau_i^{-1}}} \lambda^{a-1} \exp\left\{ -b\lambda - \frac{1}{2} \sum_{i=1}^{l} \tau_i (\omega_i - \mu_i)^2 \right\}. \tag{8}$$

Note that this distribution is vague when the hyperparameters tend to zero. Moreover, the distribution of D_t conditioned on x_0 is the convolution of the Binomial and the Poisson distributions taking the form

$$f_{D_t}(d_t|x_0, \boldsymbol{\theta}, \mathsf{S}_\mathbf{Y}) = \prod_{n=1}^{t-q} e^{-\lambda} \sum_{i=0}^{\min(x_{n-1}, x_n)} \frac{\lambda^{x_n - i}}{(x_n - i)!} C_i^{x_{n-1}} \alpha_n^i (1 - \alpha_n)^{x_{n-1} - i}. \tag{9}$$

Conjugating (8) and (9) it follows that the posterior distribution is proportional to

$$\begin{aligned} h(\boldsymbol{\theta}|d_t, \mathsf{S}_\mathbf{Y}) \quad \propto \quad & \frac{1}{\prod_{i=1}^{l} \sqrt{2\pi\tau_i^{-1}}} \lambda^{a-1} \exp\left\{ -b\lambda - \frac{1}{2} \sum_{i=1}^{l} \tau_i (\omega_i - \mu_i)^2 - (t-q)\lambda \right\} \\ & \times \prod_{n=1}^{t-q} \sum_{i=0}^{\min(x_{n-1}, x_n)} \frac{\lambda^{x_n - i}}{(x_n - i)!} C_i^{x_{n-1}} \alpha_n^i (1 - \alpha_n)^{x_{n-1} - i}. \end{aligned} \tag{10}$$

The probability P_1 is given by

$$P(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}) = \int_\Gamma \int P(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}, \boldsymbol{\theta}) h(\boldsymbol{\theta}|D_t, \mathsf{S}_\mathbf{Y}) \, d\boldsymbol{\omega} d\lambda, \tag{11}$$

with $\Gamma = \{(\omega_1, \omega_2, \ldots, \omega_l) \in (-\infty, \infty)^l\}$ whereas the probability of P_2 takes the form

$$P(C_{t,j}|D_t, \mathsf{S}_\mathbf{Y}) = \int_\Gamma \int P(C_{t,j}|D_t, \mathsf{S}_\mathbf{Y}, \boldsymbol{\theta}) h(\boldsymbol{\theta}|D_t, \mathsf{S}_\mathbf{Y}) \, d\boldsymbol{\omega} d\lambda, \tag{12}$$

where $P(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}, \boldsymbol{\theta})$ and $P(C_{t,j}|D_t, \mathsf{S}_\mathbf{Y}, \boldsymbol{\theta})$ are calculated through (6) and (7) respectively. *In this case, the predictive values Y_{t+j} are estimated by calculating the mean of the corresponding predictive distribution via the composition method suggested by Tanner*

(1996).

It is worth to mention that the complexity of expressions (11) and (12) do not permit their analytical calculation, even in the simplest case $j = 1$. Regarding expression (11) since by definition $P\left(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}\right) = E_{\boldsymbol{\theta}|D_t=d_t,\mathsf{S}_\mathbf{Y}}\left[P\left(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}, \boldsymbol{\theta}\right)\right]$ it is easy to obtain its respective Monte Carlo approximation through the expression

$$\hat{P}\left(C_{t,j}|\mathbf{x_2}, D_t, \mathsf{S}_\mathbf{Y}\right) = \frac{1}{m}\sum_{i=1}^{m} P\left(C_{t,j}|\mathbf{x_2}, D_t, \boldsymbol{\theta}_i, \mathsf{S}_\mathbf{Y}\right), \tag{13}$$

where the observations $\boldsymbol{\theta}_i = (\lambda_i, \boldsymbol{\omega}_i)$, $i = 1, \ldots, m$ form a sample of the posterior distribution $h(\boldsymbol{\theta}|D_t, \mathsf{S}_\mathbf{Y})$. In view of the fact that this sample can not be generate directly from the posterior distribution, we use the Gibbs methodology with Metropolis step, available in the program WINBUGS, to sample from (10). A similar procedure it is applied to estimate the probability P_2 and the operating characteristics.

4 Simulation study

In this section we present a simulation study to illustrate the performance of the alarm system using data sets generated from the DSINAR(1) model in (3) with one covariate. Moreover, we assume that $q = 1$. The simulation study contemplates six different combinations of (λ, ω_1) namely $\lambda = 2, 3, 4$ and $\omega_1 = 0.2, 0.3$. For the covariate Y_t the continuous-valued first-order autoregressive model

$$Y_t - 3 = 0.6(Y_{t-1} - 3) + \epsilon_t,\ t = 1, 2, \ldots,$$

with $\epsilon_t \sim N(0, 1)$, is adopted. 200 samples of size 250 are generated for each combination of (λ, ω_1). The analysis of the alarm system is carried out at $t = 200$, i.e., $\mathbf{X_2} = \{x_{200}\}$. The event of interest is the two step ahead catastrophe given by the upcrossing level u at time $t + 2$, i.e., $C_{200,2} = \{X_{201} \leq u < X_{202}\}$. A two-step ahead catastrophe was used to diminish the computational effort, but it is possible to construct a j-step ahead catastrophe with $j > 2$. The choice of u is carried out in four stages: (a) for each one of the 200 samples obtained for a fixed combination of (λ, ω_1) the corresponding probability

P_2 in (7) considering the true values for the parameters, say $P_2^{(i)}$, given by

$$P_2^{(i)} = \sum_{x_{t+j-1}^{(i)} \leq u} p(x_{t+j-1}^{(i)} | x_{t-1}^{(i)}, \mathsf{S}_Y) \left(1 - \sum_{x_{t+j}^{(i)} \leq u} p(x_{t+j}^{(i)} | x_{t+j-1}^{(i)}, \mathsf{S}_Y) \right),$$

with $i = 1, \ldots, 200$ is calculated, for a fixed value of $u \in [10, 25]$; (b) for the set of probabilities $(P_2^{(1)}, \ldots, P_2^{(200)})$ calculate the corresponding sample mean $(P_2^{(1)} + \cdots + P_2^{(200)})/200$ (c) repeat steps (a) and (b) for different values of u ranging from 10 to 25; (d) choose as the appropriate value of u the one corresponding to the sample mean closer to 0.1. The choice of this value is justified by the fact that we are interested in relatively rare events. The previous procedure lead us to select $u = 8$ for $(\lambda, \omega_1) = (2, 0.2)$, $u = 9$ for $(\lambda, \omega_1) = (2, 0.3)$, $u = 11$ for $(\lambda, \omega_1) = (3, 0.2)$, $u = 13$ for $(\lambda, \omega_1) = (3, 0.3)$, $u = 12$ for $(\lambda, \omega_1) = (4, 0.2)$, and $u = 17$ for $(\lambda, \omega_1) = (4, 0.3)$.

For each one of the 200 samples obtained for a fixed combination of (λ, ω_1) an optimal alarm region is generated through expression (2) for values of k ranging from P_2 to $P_2 + 0.1$. For each optimal alarm region the corresponding operating characteristics are calculated. This procedure is repeated for the classical (using the true values of the parameters and their maximum likelihood estimates) and the Bayesian approach. In the Bayesian setting a sample of length 35000, including a burn-in period of 15000 observations, of the posterior distribution is generated. Furthermore, only every twentieth iteration is stored in order to obtain an, approximately, independent and identically distributed sample.

As previously mentioned, the choice of k plays a key role in order to obtain the best collection of operating characteristics. Besides Svensson's criterion already mentioned in section 2 another procedures to deal with this problem include (a) to choose the value of k such that the alarm size approximately equals the probability P_2; and (b) to choose the value of k that verifies $P(A_{t,2}|D_t) \approx 2P(C_{t,2}|D_t)$; see Antunes (2002) for details. When dealing with count processes, however, these criteria can not be directly applied due to the discrete nature of the data. Giving heed to this problem we consider two different criteria for the selection of k:

Criterion 1: (B_1) $0.5 < \frac{P(A_{t,2}|D_t)}{P(C_{t,2}|D_t)} \leq 1.5$; ($B_2$) $1.5 < \frac{P(A_{t,2}|D_t)}{P(C_{t,2}|D_t)} \leq 2.5$; ($B_3$) $2.5 < \frac{P(A_{t,2}|D_t)}{P(C_{t,2}|D_t)} \leq 3.5$;

For criterion B_j with $j = 1, 2, 3$ the optimal value $k = k^*$ turn out to be

$$k^* = \min_{k: j-1/2 < \frac{P(A_{t,2}|D_t)}{P(C_{t,2}|D_t)} \leq j+1/2} \left| \frac{P(A_{t,2}|D_t)}{P(C_{t,2}|D_t)} - j \right|.$$

Since for all values of k considered the probability of correct alarm is too small we introduce the alternative Criterion 2 in which the value of k is such that the probability of detection is approximately 0.5.

Criterion 2: (B_4) $P(A_{t,2}|C_{t,2}, D_t) \approx 0.5$.

Table 1 below shows the operating characteristics of the alarm system obtained by considering the classical approach replacing the true values for their maximum likelihood estimates and the Bayesian approach for $\lambda = 2, 3, 4$ and $\omega_1 = 0.2, 0.3$, considering only the values of k satisfying criterion B_2.

(Table 1 about here)

Table 2 represents the results for $\lambda = 2$. For $\omega_1 = 0.2$ the ratio alarms/catastrophes is in general above the interval defined by the respective criterion whereas for $\omega_1 = 0.3$ this ratio is below the considered interval. The number of false alarms is very high, for both values of ω_1, being nearly 80%. For $\omega_1 = 0.2$ the criterion B_3 has the highest percentages of detection regardless the approach considered whereas for $\omega_1 = 0.3$ the criteria B_4 rends the best collection of operating characteristics.

(Table 2 about here)

In Table 3 for $\lambda = 3$ and $\omega_1 = 0.2$ the ratio alarms/catastrophes falls into the respective interval regardless the criteria and approach adopted. For the case $\lambda = 3$ and $\omega_1 = 0.3$, however, the ratio alarms/catastrophes does not fall into the respective interval for the criterion B_3; and criterion B_2 when considering the true parameters. In terms of false alarms the results in Table 3 are very similar to the ones obtained in Table 2 being the percentages around 80%. Arguably, for $\lambda = 3$ it seems that criterion B_4 rends the best collection of operating characteristics.

(Table 3 about here)

The results in Table 4 show that the ratio alarms/catastrophes is in general below the interval defined by the respective criterion. Note that for $\lambda = 4$ and $\omega_1 = 0.2$ the most balanced criterion seems to be B_2. The percentages of false alarms are below 70% and the percentages of detection are between 30% and 44.4%. For $\lambda = 4$ and $\omega_1 = 0.3$, however, the criterion that seems to rend the best collection of operating characteristics is B_3. The percentages of false alarms are below 70% and the percentage of detecting the catastrophe for the classical approach (considering the true values of the parameters) is 30% whereas for the other two approaches this percentage is equal to 44.4%.

(Table 4 about here)

Finally, in order to illustrate how the online prediction performs in practice, we consider a single realization of the DSINAR(1) process with parameters $\lambda = 4$ and $\omega_1 = 0.3$. The event to predict is $C_{t,2} = \{X_{t+1} \leq 17 < X_{t+2}\}$, $t = 199, \ldots, 205$. First of all, we start by analysing the alarm region and the operating characteristics at the time $t = 200$. Hence for this process, the alarm regions for values of k ranging from 0.151 to 0.175, considering both Bayesian and maximum likelihood estimtes, are calculated. The plot containing the alarm regions obtained by using maximum likelihood estimates is given in Figure 1. The alarm regions obtained by adopting the Bayesian perspective are identical.

(Figure 1 about here)

Table 5 contains the corresponding operating characteristics obtained by considering both Bayesian and maximum likelihood estimates.

(Table 5 about here)

In Figure 2 the probability of alarm region, considering maximum likelihood estimates, against k is plotted. In the figure are also represented the criteria B_1, B_2 and B_3. The optimal alarm system and the associated operating characteristics for criterion B_1 are obtained with $k = 0.175$, for criterion B_2 the value of k is 0.172 and for criterion B_3 k is 0.164.

(Figure 2 about here)

Table 6 summarizes the operating characteristics of the alarm system for $t = 199, 200, \ldots, 205$ and the corresponding alarm region obtained by criteria B_3.

(Table 6 about here)

5 Example

In this section we model a data set containing the number of sunspot groups (Zurich classification) available in the annual *Solar Observations Bulletins* published by the Catania Astrophysical Observatory. These reports track the birth, the growth, and the decay of spot groups, as solar rotation carries them across the face of the sun.

Sunspots appear as dark spots on the surface of the sun. Temperatures in the dark centers of sunspots drop to about 3700K (compared to 5700K for the surrounding photosphere). They typically last for several days, although very large ones may live for several weeks. Sunspots are magnetic regions on the sun with magnetic field strengths thousands of times stronger than the Earth's magnetic field. Sunspots usually come in groups with two sets of spots. One set will have positive or north magnetic field while the other set will have negative or south magnetic field. The field is strongest in the darker parts of the sunspots - the umbra. The field is weaker and more horizontal in the lighter part - the penumbra.

The sunspot series of day-by-day observations is collected for the period December 1, 1998, to 30 April, 1999. As covariate we consider the corrected total area in millionths of solar hemisphere. The first problem to be dealt with is the lack of some data in the records. The method to fill gaps in a time series depends basically on their duration. In the time series under analysis (sunspot of the surface on the sun and the corrected total area in millionths of solar hemisphere) small gaps have been filled by a direct ordinary interpolation between the neighboring observations. Note that this method, when applied to the sunspot data set, lacks of data *coherency* in the sense that the interpolated values have to be restricted to the set of the integers. To overcome this difficulty the interpolated values were adjusted to the nearest integer. Both series are exhibited in Figure 3.

(Figure 3 about here)

The autocorrelation function and the partial autocorrelation function of the series of corrected total area in millionths of solar hemisphere is displayed in Figure 4. Included as dotted lines on each plot are approximate 95% data confidence limits.

(Figure 4 about here)

Both functions are consistent with the data being generated by an AR(1) and it seems reasonable, therefore, to proceed with the estimation of the parameters and the events of interest. In order to fit the DSINAR(1) model we only use the observations between December 1, 1998 to February 12, 1999 for modelling purposes whereas the rest of the observations (i.e., from February 13 to April 30 of 1999) are used to illustrate the performance of the online prediction procedure to predict the extremal upcrossing event $C_{t,2} = \{X_{t+1} \leq 4 \leq X_{t+2}\}$ with $t = 75, 76, \ldots, 149$. Note that the probability of catastrophe given the past D_t is $P(C_{t,j}|D_t, \mathsf{S}_{\mathbf{Y}}) \approx 0.09$. *In order to calculate the optimal alarm region and its corresponding operating characteristics at every fixed instant t, the values of Y_{t+j}, $j = 0, 1, 2$, are obtained through the optimal linear j-period ahead forecast given by $\hat{Y}_{t+j} = \hat{\mu}_{t-1} + \hat{\phi}^j_{t-1}(Y_{t-1} - \hat{\mu}_{t-1})$, being $(\hat{\mu}_{t-1}, \hat{\phi}_{t-1})$ the CML estimates of $\mu = E(Y_t)$ and ϕ the parameter of the autoregressive model of order one, respectively, for the classical approach. From the Bayesian point of view the predictive values Y_{t+j} are estimated by calculating the mean of the corresponding predictive distribution.* The results for the online alarm system of the DSINAR(1) model considering criterion B_4 are presented in Table 7.

(Table 7 about here)

In February the alarm system correctly predicts the catastrophe whereas in April the probability of detecting a catastrophe is 33% (3 catastrophes being 1 detected). The number of false alarms is rather high although this depends on our strict definition of correct alarm. This number will be smaller if we accept alarms that are one step early and do not count the alarms we get while we are still in the catastrophe state. Hence, it can be discussed whether these kinds of false alarms in a practical meaning should be considered false.

It is worth to mention that other studies have found similar result regarding the number of false alarms. For example, Svensson and Holst (1998) in the analysis of high water levels at the Danish coast in the Baltic sea report a rate of false alarms nearly 95%; see also Svensson and Holst (1997) for further details.

6 Conclusions

This paper has presented an optimal alarm system for processes described by integer-valued autoregressive processes with parameters being functions of covariates of interest and varying on time. The optimal alarm technique leads to optimal event predictors in the sense that they give the least number of false alarms for a predetermined alarm size. As stressed throughout the paper the number of false alarms is rather high both in the simulation study and in the working example. A possibility to lower the number of false alarms is to include in the model additional external information or to consider a time-varying catastrophe level. This remains a topic of future research.

Appendix A

Proof of Proposition 2.1:

the proof follows by induction and relies on the properties of the thinning operator (see Silva and Oliveira, 2004). If $h = 1$ then $X_{t+1} \stackrel{d}{=} \alpha_{t+1} \circ X_t + Z_{t+1}$. If the result is true for $h = p - 1$, $(p > 1)$, then for $h = p$

$$\begin{aligned}
X_{t+p}|\mathsf{S}_{\mathbf{Y}},\lambda,\boldsymbol{\omega} &\stackrel{d}{=} \alpha_{t+p}\circ X_{t+p-1}+Z_{t+p}\\
&\stackrel{d}{=} \alpha_{t+p}\circ\left[\left(\prod_{i=0}^{p-2}\alpha_{t+p-1-i}\right)\circ X_t+Z_{t+p-1}+\sum_{j=1}^{p-2}\left(\prod_{m=0}^{j-1}\alpha_{t+p-1-m}\right)\circ Z_{t+p-1-j}\right]\\
&+ Z_{t+p}\\
&\stackrel{d}{=} \left(\alpha_{t+p}\prod_{i=0}^{p-2}\alpha_{t+p-1-i}\right)\circ X_t+\alpha_{t+p}\circ Z_{t+p-1}+\sum_{j=1}^{p-2}\left(\alpha_{t+p}\prod_{m=0}^{j-1}\alpha_{t+p-1-m}\right)\circ Z_{t+p-1-j}\\
&+ Z_{t+p}\\
&\stackrel{d}{=} \left(\prod_{i=0}^{p-1}\alpha_{t+p-i}\right)\circ X_t+\alpha_{t+p}\circ Z_{t+p-1}+\sum_{j=1}^{p-2}\left(\prod_{m=0}^{j}\alpha_{t+p-m}\right)\circ Z_{t+p-1-j}+Z_{t+p}\\
&\stackrel{d}{=} \left(\prod_{i=0}^{p-1}\alpha_{t+p-i}\right)\circ X_t+\alpha_{t+p}\circ Z_{t+p-1}+\sum_{j=2}^{p-1}\left(\prod_{m=0}^{j-1}\alpha_{t+p-m}\right)\circ Z_{t+p-j}+Z_{t+p}\\
&\stackrel{d}{=} \left(\prod_{i=0}^{p-1}\alpha_{t+p-1-i}\right)\circ X_t+Z_{t+p}+\sum_{j=1}^{p-1}\left(\prod_{m=0}^{j-1}\alpha_{t+p-m}\right)\circ Z_{t+p-j}.
\end{aligned}$$

References

[1] Ahn, S., Gyemin, L. and Jongwoo, J. (2000). Analysis of the M/D/1-type queue based on an integer-valued autoregressive process. Oper. Res. Lett. 27:235–241.

[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] Al-Osh, M. and Alzaid, A. (1988). Integer-valued moving average (INMA) process. Stat. Pap. 29:281–300.

[4] Aly, E.-E. and Bouzar, N. (1994). Explicit stationary distributions for some Galton-Watson processes with immigration. Comm. Statist. Stochastic Models 10:499–517.

[5] Aly, E.-E. and Bouzar, N. (2005). Stationary solutions for integer-valued autoregressive processes. Int. J. Math. Math. Sci. 1:1–18.

[6] Amaral-Turkman, M. A. and Turkman, K. F. (1990). Optimal alarm systems for autoregressive processes; a Bayesian approach. Comput. Statist. Data Anal. 10:307–314.

[7] Antunes, M. (2002). Some problems in non-linear prediction. Unpublished doctoral dissertation. University of Lisbon.

[8] Antunes, M., Amaral-Turkman, M. A. and Turkman, K. F. (2003). A Bayesian approach to event prediction intervals. J. Time Ser. Anal. 24:631–646.

[9] Blundell, R., Griffith, R. and Windmeijer, F. (2002). Individual effects and dynamics in count data models. J. Econometrics 108:113–131.

[10] Braga, A. L., Zanobetti, A., Schwartz, J. (2001). The time course of weather-related deaths. Epidemiology 12:662-667.

[11] Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. Umeå Economic Studies 381.

[12] Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models. Appl. Stochastic Models Bus. Ind. 17:277–291.

[13] Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression. Econometric Rev. 20:425–443.

[14] Brännäs, K. Hellström, and J. Nordström, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels. Int. J. Forecast. 18:9–30.

[15] Brännäs, K. and Nordström, J. (2004). An integer-valued time series model for hotels that accounts for constrained capacity. Studies in Nonlinear Dynamics and Econometrics 8:4, article 6.

[16] Brännäs, K. and Nordström, J. (2006). Tourist accomodation effects of festivals. Tourism Economics 12:291–302.

[17] Campbell, M. J., Julious, S. A., Peterson, C. K. and Tobias, A. (2001). Atmospheric pressure and sudden infant death syndrome in Cook County, Chicago. Paediatr Perinat Epidemiol. 15:287-289.

[18] de Maré, J. (1980). Optimal prediction of catastrophes with application to Gaussian process. Ann. Probab. 8:841–850.

[19] Du, J.-G. and Li, Y. (1991). The integer valued autoregressive (INAR(p)) model. J. Time Ser. Anal. 12:129–142.

[20] Garcia-Ferrer, A. and Queralt, R. A. (1997). A note on forecasting international tourism deman in Spain. Int. J. Forecast. 13:539–549.

[21] Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopolis Y. and Rossi, G. (2002). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA 2 project. Epidemiology 12:521-531.

[22] Koop, G. and Tale, L. (2004). Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? J. Environ. Econ. Manage. 47:30-54.

[23] Latour, A. (1988). Existence and stochastic structure of a non-negative intger-valued autoregressive processes. J. Time Ser. Anal. 4:439–455.

[24] Lindgren, G. (1975a). Prediction for a random point time. Ann. Probab. 3:412–423.

[25] Lindgren, G. (1975b). Prediction of catastrophes and high level crossings. Bull. Internat. Statist. Inst. 46:225–240.

[26] Lindgren, G. (1980). Model processes in non-linear prediction, with application to detection and alarm. Ann. Probab. 8:775–792.

[27] Lindgren, G. (1985). Optimal prediction of level crossings in Gaussian processes and sequences. Ann. Probab. 13:804–824.

[28] McCabe, B. P. M. and Martin, G. M. (2005). Bayesian prediction of low count time series. Int. J. Forecast. 21:315–330.

[29] McKenzie, E. (1985). Some simple models for discrete variate time series. Water Res. Bull. 21:645–650.

[30] McKenzie, E. (2003). Discrete variate time series. In: Shanbhag, D. N. Rao, C. R. ed., Handbook of Statistics. Elsevier Science, 573–606.

[31] Neal, P. and Subba Rao, T. (2007). MCMC for integer-valued ARMA processes. J. Time Ser. Anal. 28:92–110.

[32] Nordström, J. (1996). Tourism satellite account for Sweden 1992-93. Tourism Economics 2:13–42.

[33] Quoreshi, A. M. M. S. (2006). Bivariate time series modelling of financial count data. Comm. Statist. Theory Methods 35:1343–1358.

[34] Schwartz, J., Levin, R. and Hodge, K. (1997). Drinking water turbidity and paediatric hospital use for gastrointestinal illness in Philadelphia. Epidemiology 8:615-620.

[35] Silva, M. E. and Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. J. Time Ser. Anal. 25:317–333.

[36] Silva, I., Silva, M. E., Pereira, I. and Silva, N. (2005). Replicated INAR(1) processes. Methodol. Comput. Appl. Probab. 7:517–542.

[37] Svensson, A., Lindquist, R. and Lindgren, G. (1996). Optimal prediction of catastrophes in autoregressive moving average processes. J. Time Ser. Anal. 17:511–531.

[38] Svensson, A. and Holst, J. (1997). Optimal prediction of level-crossing in piecewise linear Gaussian processes with changing catastrophe level. Tech. rep. 1997:14, Department of Mathematical Statistics, Lund Institute of Technology.

[39] Svensson, A. and Holst, J. (1998). Prediction of the high water levels in the Baltic. Istatistik, Journal of the Turkish Statistical Association 1:39–46.

[40] Tanner, M. A. (1996). Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions. 3rd edition. New-York: Springer-Verlag.

[41] Touloumi, G., Atkinson, R., Le Tertre, A., Samoli, E., Schwartz, J., Schlinder, C., Vonk, J. M., Rossi, G., Saez. M., Rabszenko, D. and Katsouyanni K. (2004). Analysis of health outcome time series data in epidemiological studies. Environmetrics 15:101-117.

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

[43] Zheng, H. T., Basawa, I. V. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. J. Statist. Plann. Inference 173:212–229.

[44] Zhou, J. and Basawa, I. V. (2005). Least-squared estimation for bifurcation autoregressive processes. Statist. Probab. Lett. 74:77-88.

[45] Zhu, R. and Joe, H. (2003). A new type of discrete self-decomposability and its applications to continuous-time Markov processes for modeling count data time series. Stoch. Models 19:235–254.

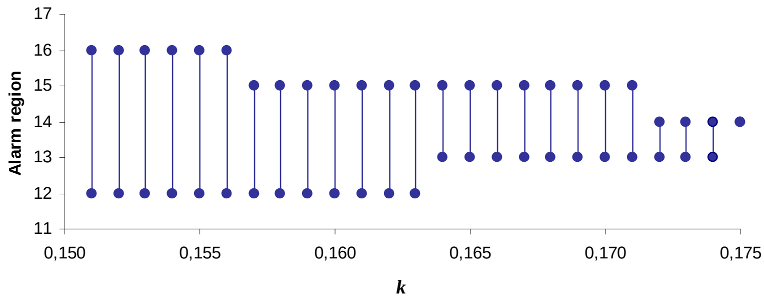


Figure 1: Alarm regions for values of $k \in [0.151, 0.175]$ considering maximum likelihood estimates at time point $t = 200$.

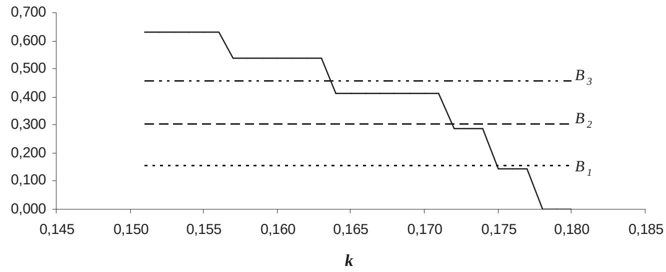
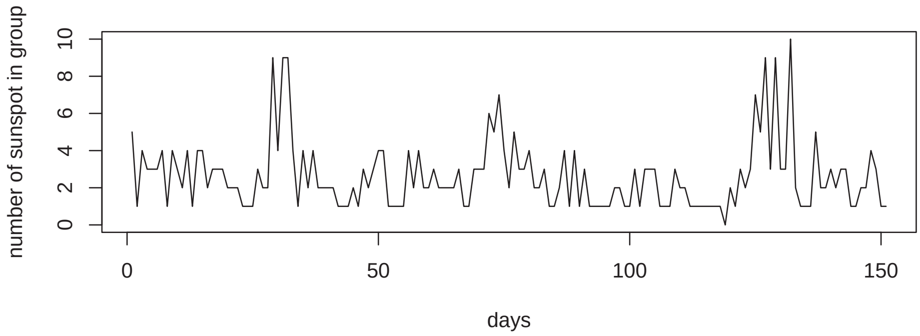


Figure 2: Plot of the probability of alarm region versus criteria B_1, B_2 and B_3 at time $t = 200$ and $j = 2$ considering maximum likelihood estimates.



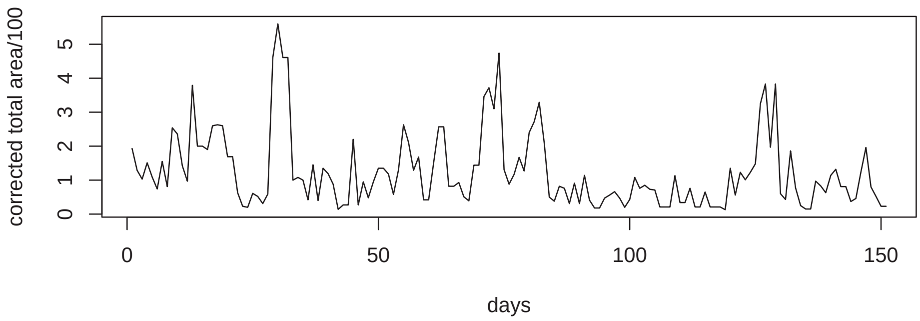


Figure 3: Time series plots for the number of sunspot in groups and the corrected total area in millionths of solar hemisphere/100 for the period December 1, 1998 to 30 Abril, 1999.

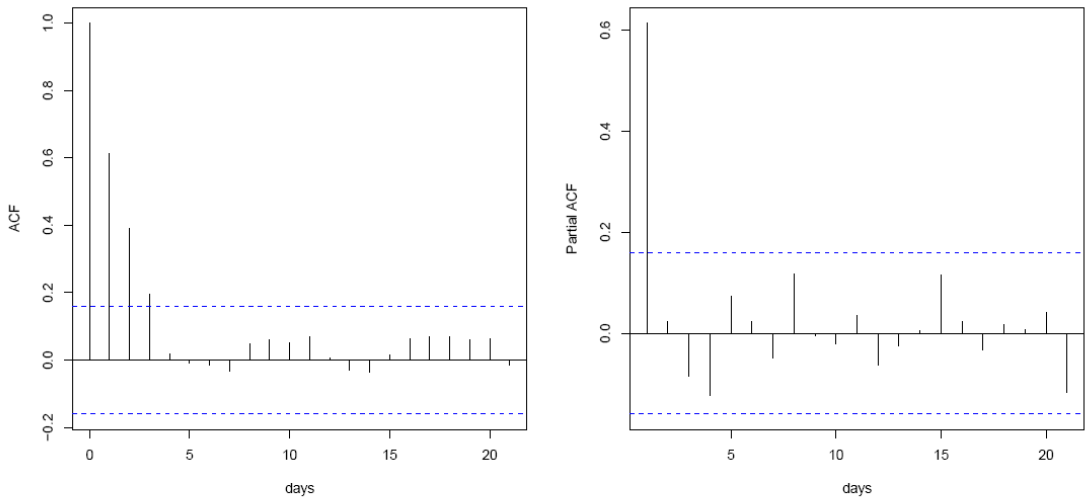


Figure 4: Autocorrelation function and partial autocorrelation of the series corrected total area in millionths of solar hemisphere.

Parameters	x_{t-1}	M.L. Estimates k	$P(C_{t,2}\|D)$	$P(A_{t,2,k}\|D)$	$P(C_{t,2}\|A_{t,2,k})$	$P(A_{t,2,k}\|C_{t,2})$	Bayessian approach k	$P(C_{t,2}\|D)$	$P(A_{t,2,k}\|D)$	$P(C_{t,2}\|A_{t,2,k})$	$P(A_{t,2,k}\|C_{t,2})$
$(\lambda,\omega_1)=(2,0.2)$ $u=8$	2	0.078	0.051	0.109	0.094	0.200	0.091	0.052	0.106	0.094	0.194
	4	0.110	0.082	0.140	0.116	0.198	0.110	0.083	0.140	0.116	0.198
	7	0.109	0.093	0.207	0.111	0.245	0.108	0.093	0.206	0.111	0.245
	9	0.113	0.095	0.172	0.117	0.212	0.113	0.094	0.171	0.117	0.212
	10	0.110	0.096	0.144	0.112	0.179	0.109	0.089	0.143	0.112	0.178
	14	0.089	0.040	0.065	0.102	0.166	0.088	0.040	0.071	0.102	0.169
$(\lambda,\omega_1)=(2,0.3)$ $u=9$	2	0.108	0.071	0.137	0.149	0.286	0.109	0.072	0.139	0.149	0.287
	4	0.122	0.080	0.168	0.128	0,.70	0.122	0.080	0.168	0.129	0.270
	8	0.134	0.114	0.211	0.140	0.259	0.135	0.113	0.211	0.148	0.261
	9	0.137	0.107	0.194	0.144	0.260	0.136	0.106	0.193	0.143	0.259
	13	0.109	0.049	0.112	0.131	0.296	0.109	0.049	0.114	0.130	0.296
	14	0.100	0.035	0.054	0.141	0.220	0.099	0.035	0.058	0.141	0.222
$(\lambda,\omega_1)=(3,0.2)$ $u=11$	3	0.092	0.048	0.076	0.120	0.189	0.092	0.048	0.077	0.120	0.189
	5	0.079	0.055	0.134	0.089	0.217	0.088	0.056	0.128	0.090	0.206
	10	0.118	0.103	0.168	0.120	0.196	0.119	0.103	0.169	0.120	0.197
	12	0.109	0.090	0.159	0.113	0.201	0.109	0.089	0.158	0.113	0.201
	16	0.108	0.059	0.136	0.118	0.274	0.108	0.057	0.137	0.118	0.273
	17	0.095	0.048	0.095	0.111	0.218	0.098	0.047	0.094	0.111	0.211
$(\lambda,\omega_1)=(3,0.3)$ $u=13$	6	0.118	0.079	0.183	0.122	0.282	0.118	0.081	0.183	0.124	0.282
	9	0.124	0.107	0.245	0.129	0.297	0.124	0.107	0.244	0.129	0.298
	12	0.158	0.125	0.253	0.165	0.336	0.158	0.124	0.253	0.165	0.335
	14	0.129	0.103	0.247	0.134	0.322	0.129	0.101	0.246	0.133	0.322
	18	0.116	0.065	0.157	0.128	0.312	0.117	0.063	0.157	0.129	0.311
	22	0.070	0.013	0.022	0.110	0.184	0.069	0.014	0.025	0.108	0.185
$(\lambda,\omega_1)=(4,0.2)$ $u=12$	4	0.099	0.075	0.179	0.101	0.240	0.098	0.075	0.178	0.101	0.240
	5	0.089	0.067	0.112	0.092	0.155	0.090	0.068	0.113	0.093	0.155
	11	0.116	0.102	0.250	0.118	0.291	0.116	0.102	0.250	0.119	0.291
	12	0,137	0.117	0.261	0.141	0.316	0.138	0.116	0.206	0.141	0.250
	18	0.115	0.073	0.158	0.123	0.267	0.115	0.072	0.158	0.122	0.265
	19	0.117	0.071	0.113	0.137	0.219	0.116	0.070	0.113	0.137	0.217
$(\lambda,\omega_1)=(4,0.3)$ $u=17$	6	0,102	0.057	0.130	0.120	0.273	0.101	0.057	0.130	0.120	0.272
	8	0.139	0.109	0.215	0.153	0.303	0.139	0.108	0.214	0.152	0.302
	16	0.144	0.117	0.245	0.154	0.321	0.144	0.117	0.243	0.155	0.321
	18	0.138	0.102	0.165	0.145	0.235	0.138	0.101	0.164	0.145	0.234
	22	0.127	0.059	0.141	0.145	0.345	0.127	0.058	0.140	0.144	0.343
	23	0.094	0.034	0.065	0.129	0.248	0.093	0.033	0.067	0.128	0.248

Table 1: Operating characteristics at time point $t = 200$.

Approach	Criterion	$\omega_1 = 0.2$				$\omega_1 = 0.3$			
		Alarms		Catastrophes		Alarms		Catastrophes	
		False	Total	Detected	Total	False	Total	Detected	Total
True parameters	B_1	1	1	0	4	11	11	0	6
	B_2	32 (0.84)	38	6 (0.35)	17	14 (0.88)	16	2 (0.13)	16
	B_3	22(0.85)	26	4(0.67)	6	19 (0.83)	23	4 (0.33)	12
	B_4	81 (0.9)	90	9 (0.50)	18	57 (0.84)	68	11 (0.58)	19
M.L. Estimates	B_1	4	4	0	5	10	10	0	3
	B_2	28 (0.88)	32	4 (0.31)	13	19 (0.90)	21	2 (0.12)	17
	B_3	16 (0.84)	19	3 (0.6)	5	21 (0.84)	25	4 (0.31)	13
	B_4	83 (0.90)	92	9 (0.50)	18	108 (0.89)	121	13 (0.68)	19
Bayesian	B_1	4	4	0	6	10	10	0	3
	B_2	28 (0.88)	32	4 (0.20)	20	19 (0.90)	21	2(0.11)	18
	B_3	18 (0.82)	22	4 (0.67)	6	21 (0.84)	25	4 (0.31)	13
	B_4	82 (0.90)	91	9 (0.50)	18	62 (0.85)	73	11 (0.58)	19

Table 2: Results for $\lambda = 2$ at time point $t = 200$. Percentages in parenthesis.

Approach	Criterion	$\omega_1 = 0.2$				$\omega_1 = 0.3$			
		Alarms		Catastrophes		Alarms		Catastrophes	
		False	Total	Detected	Total	False	Total	Detected	Total
True parameters	B_1	11 (0.85)	13	2 (0.20)	10	23 (0.852)	27	4 (0.14)	28
	B_2	23 (0.92)	25	2 (0.15)	13	15 (0.94)	16	1 (0.33)	3
	B_3	45 (0.80)	56	11(0.52)	21	37 (0.80)	46	9 (0.35)	26
	B_4	74 (0.85)	87	13 (0.54)	24	65 (0.84)	77	12 (0.43)	28
M.L. Estimates	B_1	15 (0.88)	17	2 (0.17)	12	18 (0.82)	22	4 (0.20)	20
	B_2	25 (0.89)	28	3 (0.21)	14	22 (0.88)	25	3 (0.23)	13
	B_3	37 (0.84)	44	7 (0.47)	15	36 (0.80)	45	9 (0.38)	24
	B_4	78 (0.85)	92	14 (0.58)	24	66 (0.85)	78	12 (0.43)	28
Bayesian	B_1	14 (0.88)	16	2 (0.20)	10	19 (0.83)	23	4 (0.20)	20
	B_2	25 (0.89)	28	3 (0.21)	14	21 (0.88)	24	3 (0.27)	11
	B_3	40 (0.85)	47	7 (0.50)	14	36 (0.81)	44	8 (0.35)	23
	B_4	78 (0.85)	92	14 (0.58)	24	65 (0.84)	77	12 (0.43)	28

Table 3: Results for $\lambda = 3$ at time point $t = 200$. Percentages in parenthesis.

Approach	Criterion	$\omega_1 = 0.2$				$\omega_1 = 0.3$			
		Alarms		Catastrophes		Alarms		Catastrophes	
		False	Total	Detected	Total	False	Total	Detected	Total
True parameters	B_1	0	0	0	27	13 (0.81)	16	3 (0.10)	30
	B_2	25 (0.76)	33	8 (0.30)	27	23 (0.76)	30	7 (0.23)	30
	B_3	0	0	0	27	21 (0.70)	30	9 (0.30)	30
	B_4	76 (0.81)	93	17 (0.63)	27	46 (0.75)	61	15 (0.50)	30
M.L. Estimates	B_1	3	3	0	6	8 (0.80)	10	2 (0.07)	28
	B_2	14 (0.66)	21	7 (0.36)	19	18 (0.90)	20	2 (0.10)	20
	B_3	21 (0.80)	26	5 (0.33)	15	23 (0.65)	35	12 (0.44)	27
	B_4	71 (0.80)	88	17 (0.63)	27	50 (0.75)	66	16 (0.53)	30
Bayesian	B_1	4	4	0	6	9 (0.81)	11	2 (0.09)	27
	B_2	14 (0.63)	22	8 (0.44)	18	19 (0.90)	21	2 (0.28)	7
	B_3	21 (0.75)	28	7 (0.50)	14	23 (0.67)	35	12 (0.44)	27
	B_4	72 (0.81)	89	17 (0.63)	27	49 (0.75)	65	16 (0.53)	30

Table 4: Results for $\lambda = 4$ at time point $t = 200$. Percentages in parenthesis.

Approach	k	$P(C_{t,2}\|D)$	$P(A_{t,2,k}\|D)$	$P(C_{t,2}\|A_{t,2,k})$	$P(A_{t,2,k}\|C_{t,2})$
M.L.Estimates	0.151	0.151	0.633	0.170	0.712
	0.157	0.151	0.537	0.172	0.612
	0,164	0.151	0.410	0.175	0.476
	0.172	0.151	0.286	0.176	0.334
	0.175	0.151	0.142	0.178	0.168
Bayesian	0.151	0.150	0.633	0.169	0.714
	0.157	0.150	0.538	0.172	0.615
	0.164	0.150	0.411	0.175	0.478
	0.172	0.150	0.287	0.176	0.336
	0.175	0.150	0.143	0.178	0.169

Table 5: Operating characteristics for the DSINAR(1) process with $\lambda = 4$ and $\omega_1 = 0.3$ at time point $t = 200$.

Approach	t	x_{t-1}	k	$P(C_{t,2}\|D)$	$P(A_{t,2,k}\|D)$	$P(C_{t,2}\|A_{t,2,k})$	$P(A_{t,2,k}\|C_{t,2})$	Alarm Region
True parameters	199	13	0.192	0.182	0.487	0.207	0.554	{14,15,16}
	200	12	0.138	0.120	0.347	0.149	0.432	{14, 15, 16}
	201	15	0.134	0.119	0.310	0.140	0.365	{15, 16}
	202	18	0.163	0.149	0.438	0.185	0.545	{14, 15, 16, 17}
	203	14	0.119	0.101	0.305	0.122	0.368	{14, 15}
	204	15	0.103	0.092	0.297	0.107	0.344	{16, 17}
	205	13	0.127	0.106	0.297	0.145	0.403	{15, 16, 17, 18}
M.L. Estimates	199	13	0.198	0.197	0.585	0.224	0.666	{12, 13, 14, 15, 16}
	200	12	0.164	0.151	0.410	0.175	0.476	{13, 14, 15}
	201	15	0.151	0.140	0.449	0.165	0.530	{13, 14, 15, 16}
	202	18	0.171	0.159	0.459	0.203	0.587	{12, 13, 14, 15, 16, 17}
	203	14	0.140	0.128	0.395	0.150	0.465	{13, 14, 15}
	204	15	0.131	0.119	0.385	0.136	0.441	{14, 15, 16}
	205	13	0.156	0.148	0.442	0.171	0.51	{14, 15, 16, 17}
Bayesian	199	13	0.197	0.195	0.587	0.223	0.669	{12, 13, 14, 15, 16}
	200	12	0.164	0.150	0.411	0.175	0.478	{13, 14, 15}
	201	15	0.151	0.139	0.448	0.165	0.531	{13, 14, 15, 16}
	202	18	0.170	0.157	0.459	0.202	0.587	{12, 13, 14, 15, 16, 17}
	203	14	0.139	0.127	0.396	0.150	0.466	{13, 14, 15}
	204	15	0.131	0.119	0.385	0.137	0.442	{14, 15, 16}
	205	13	0.157	0.148	0.442	0.171	0.512	{14, 15, 16, 17}

Table 6: Operating characteristics at different time points.

Month	Approach	Alarms		Catastrophes	
		False	Total	Detected	Total
February	M.L. Estimates	8 (0.88)	9	1 (1.00)	1
	Bayesian	8 (0.88)	9	1 (1.00)	1
March	M.L. Estimates	8 (1.00)	8	0	0
	Bayesian	8 (1.00)	8	0	0
April	M.L. Estimates	13 (0.92)	14	1 (0.33)	3
	Bayesian	13 (0.92)	14	1 (0.33)	3

Table 7: Results of the alarm system with $u = 4$ Percentages in parenthesis.