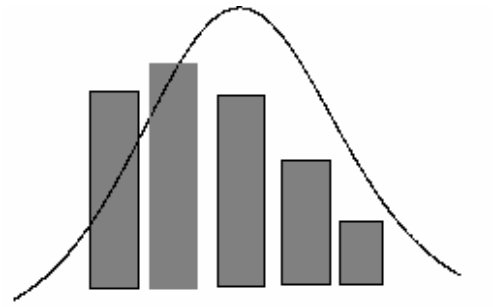
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ON TIME SERIES OF OVERDISPERSED COUNTS: MODELLING, INFERENCE, SIMULATION AND PREDICTION

Harry Pavlopoulos and Dimitris Karlis
Department of Statistics
Athens University of Economics and Business

Technical Report No 221, March 2006



DEPARTMENT OF STATISTICS

TECHNICAL REPORT

ON TIME SERIES OF OVERDISPERSED COUNTS: MODELLING, INFERENCE, SIMULATION AND PREDICTION

Harry Pavlopoulos and **Dimitris Karlis**
Department of Statistics
Athens University of Economics and Business
76 Patission Str., GR-10434 Athens, Greece

Abstract

This article is concerned with a stationary non-linear auto-regressive Markov chain on the non-negative integers, $\mathbb{N}$, known as INAR(1) model. The model's auto-regressive structure emerges from binomial thinning, a non-linear operation applied on the current state of the chain, driven additively by integer-valued i.i.d. innovations independent of the past history of the chain. In the case considered here the innovations follow a finite mixture distribution of $m \geq 1$ (independent) Poisson random variables. The stationary marginal probability distribution of the chain is self-decomposable, unimodal, and for $m > 1$ its index of dispersion (i.e. ratio of variance to mean) is greater than unity, rendering overdispersion relative to Poisson laws. The transition probability function of the model is a mixture of m Poisson-binomial laws, whence conditional maximum likelihood (CML) estimation becomes feasible by an appropriate EM-algorithm devised for this task. Furthermore, it is shown that CML estimation may be combined with estimation by the method of moments (MOM) in order to obtain a balance between descriptive and predictive abilities of the model. Integer-valued prediction is feasible via simulation of the model in a simple way. Criteria for selecting m and for assessing performance of the fitted model are discussed. As an application, the model is fitted to time series of (instantaneous) counts of pixels where spatially averaged rain rate exceeds a given threshold level. Several threshold levels are considered on (nested) sub-regions, spanning a range of spatial scales, illustrating the capabilities of the proposed model in this challenging case of application to highly overdispersed count data.

KEYWORDS: INAR Model; Dispersion Index; Poisson Mixture; Poisson-binomial Mixture; EM-Algorithm; Simulation; Integer-Valued Prediction; Threshold Rain Rate.

1 Introduction

This article is concerned with a novel version of the integer-valued auto-regressive model of first order, INAR(1) in short. The INAR(1) is actually an entire class of models, originally introduced by *McKenzie* [1985] and independently by *Al-Osh and Alzaid* [1987]. The appeal of INAR(1) stems from its strikingly simple Markovian auto-regressive structure. The model is non-linear, due to a binomial thinning operation applied at each step on the current state of the chain, and is driven additively by i.i.d. innovations independent of the past history of the chain. However, INAR(1) models do belong to the class of conditionally linear first-order auto-regressive models, CLAR(1) in short; see *Grunwald et al.* [2000]. The version considered here is driven by innovations following an m-mixture of Poisson components, rendering a marginal law of the INAR(1) model

that can account for arbitrarily large index of dispersion. In this sense, the proposed version extends the simple Poisson INAR(1) in a rather flexible manner, suitable for modelling (stationary) time series of both small and large counts with significant overdispersion relative to Poisson laws. This idea has been motivated from an interest to model time series of extremely variable and highly overdispersed count data representing spatial coverage of a region with rainfall whose intensity exceeds a given threshold level. Potentially, however, a similar interest might also pertain to spatio-temporal intermittent random fields other than rainfall intensity (e.g. of ecological, environmental and geophysical variables).

1.1 Survey of relevant literature

A substantial volume of literature is available on the probabilistic properties of INAR(1) and generalizations to INAR(p) models of higher order [*Alzaid and Al-Osh*, 1988, 1990; *Du and Li*, 1991; *Al-Osh and Aly*, 1992; *DaSilva and Oliveira*, 2004]. INAR and other types of models based on thinning, such as INMA and INARMA, are closely related to multi-type branching processes with immigration [*Dion et al.*, 1995]. They are also characterized as observation-driven models, in the sense that they formulate schemes of dependence structure directly between current and past observations. According to *Cox* [1981] observation-driven models are juxtaposed to parameter-driven models, where dependence among observations is attributed to or driven by some latent process. An overview of both classes of models is offered in *MacDonald and Zucchini* [1997], *Kedem and Fokianos* [2002], and by *McKenzie* [2003].

Statistical inference for parameters of INAR models is less developed than their probabilistic properties, motivated almost exclusively by (but also restricted to) specific cases of application on data of small counts, conducive to equidispersion or slight overdispersion. *Al-Osh and Alzaid* [1987] were concerned with estimation of the two parameters of the Poisson INAR(1) model, where innovations follow a Poisson law. Based on Monte Carlo simulations of the model they assessed the behavior of bias and mean square error (MSE) of Yule-Walker type estimators (YW), obtained by the method of moments, and of estimators obtained by methods of conditional least squares (CLS) and conditional maximum likelihood (CML), conditioning on the initial observation in the series. Recently, *Freeland and McCabe* [2005] have rigorously addressed the asymptotic properties of YW and CLS estimators of the parameters of the Poisson INAR(1) model. They derived the asymptotic covariance matrix of CLS estimators explicitly, and showed asymptotic equivalence of the distributions of CLS and YW estimators, for large samples. Estimation by CML for the Poisson INAR(1) model is treated also by *Freeland and McCabe* [2004a], along with development of methodology for assessing the model's adequacy when fitted to time series of small counts. An overview on statistical inference for INAR(1) models, from the more general standpoint of CLAR processes with discrete support, is provided by *Jung et al.* [2005].

Franke and Selingmann [1993] proved consistency and asymptotic normality of CML estimators of the vector of four parameters in the INAR(1) model with innovations following a 2-mixture of Poisson components. They referred to this model as *switching*-INAR(1), or SINAR(1) in short, and

applied it to time series of slightly overdispersed data of daily counts of epileptic seizures. Clearly, the SINAR(1) model is a special case of the INAR(1) models treated in the present article, for $m = 2$.

Thyregod et al. [1999] considered the INAR(1) and INAR(2) models with Poisson innovations, and also a *self-exciting threshold*-INAR(1) model (SETINAR in short) consisting of two Poisson INAR(1) branches. The SETINAR process alternates between its two branches at a given instant, if the sum of its two previous values (up- or down-) crosses a threshold parameter. The INAR(2) model was considered according to the formulation given by *Alzaid and Al-Osh* [1990], which implies a more complex dependence structure than that of an alternative formulation given by *Du and Li* [1991]. All three models were fitted to time series of count data from rainfall measurements at a single station. The INAR(1) model was fitted by implementing YW, CLS, and CML methods, while the INAR(2) and SETINAR models were fitted by implementing certain ramifications of the CML method.

Prediction and forecasting by INAR models is another area which has recently attracted interest and starts to pick up momentum, especially regarding the challenge of producing *integer-valued* or *"coherent"* forecasts by the fitted model. The cases addressed so far focus on INAR(1) with Poisson innovations, fitted by CML [*Freeland and McCabe*, 2004b], and on INAR(1) with Poisson, binomial, or negative binomial innovations, fitted by Bayesian methodology [*McCabe and Martin*, 2005], while *Jung and Tremayne* [2006] have produced coherent forecasts by the *Alzaid and Al-Osh* [1990] version of the INAR(2) model with Poisson innovations, fitted by the method of moments.

1.2 Outline and summary

The rest of this article is structured into five sections. Section 2 summarizes fundamental probabilistic properties of stationary INAR(1) processes, in order to facilitate understanding of the rest of the article in a self-contained manner. In the same section, the stationary marginal distribution of the INAR(1) process driven by mixed-Poisson innovations is characterized through an infinite product representation of its probability generating function (p.g.f.). This distribution is (discrete) self-decomposable, and thus unimodal. Explicit formulae for its index of dispersion, and coefficients of skewness and kurtosis can be obtained using certain moment properties, which are summarized in the Appendixes A, B, C.

Section 3 is concerned with statistical inference for the model's parameters. A certain representation of the model facilitates explicit calculation of its transition probability function as a finite mixture of m independent Poisson-binomial laws, whence an appropriate EM-algorithm is devised for CML estimation of the model's parameters. The parameter associated with binomial thinning may alternatively be estimated by the method of moments (MOM), simply as sample auto-correlation at lag-1. This option may be preferred because it offers the advantage of divorcing estimation of the thinning parameter, and subsequently the auto-correlation function (ACF) of the fitted model, from the estimation of the innovation distribution. Yet another option considered is to combine CML and MOM estimates, to the effect of regulating by desired weights the balance

of the fitted model between better descriptive qualities with regard to marginal probabilistic characteristics of the data when the thinning parameter is estimated by CML, versus better predictive qualities when the thinning parameter is estimated by MOM. This trade-off is explained on theoretical grounds, and procedures for selecting the order m of the Poisson mixture are discussed too.

Simulation of the model and issues of prediction are addressed in Section 4. Optimal prediction in the sense of minimizing mean square error (MSE) yields a linear predictor. Most importantly, integer-valued prediction is feasible via simulation of the fitted model. The two types of predictors have the same mean, but naturally the variance of the integer-valued predictor is always greater than the variance of the linear predictor. As the lead time of the prediction increases, the variance of the linear predictor vanishes and its MSE increases towards the value of the variance of the INAR(1) model, while predicted values eventually degenerate to a constant (i.e. the mean of the model). On the other hand, the asymptotic variance of the integer-valued predictor coincides with the variance of the INAR(1) model, as lead time increases. In particular, if the predicted process is indeed a (stationary) INAR(1) process, and not just modelled as INAR(1), then the whole distribution of the integer-valued predictor coincides with the stationary marginal law of the process. For these reasons, in addition to the advantage of producing integer forecasts, the integer-valued predictor is superior to the optimal predictor, the latter being linear due to the conditionally linear autoregressive (CLAR) property of the model. A certain statistic serving the purpose of quantifying the fitted model's performance is introduced and discussed in the same section.

As an application, in Section 5, the proposed model is fitted to time series of instantaneous counts of pixels where spatial averages of rain rate (SARR) exceed a given threshold level. These time series are obtained from radar scans mapping tropical rain-fields over a large region in southwestern Pacific Ocean. The count data, if scaled by the total number of pixels per scan, represent spatial coverage (of the probed region) with rainfall whose intensity exceeds the given threshold. Several threshold levels and sub-regions of different scale are considered. The obtained time series are characterized by intermittency, high overdispersion, pronounced skewness and kurtosis in all cases. All these features are captured remarkably well by the marginal distribution of the fitted model, especially so by CML versions of the fitted model. Simulations and predictions based on the fitted model yield quite adequate representations of the observed time series to which the model is fitted, more so under MOM versions. That is, the anticipated and theoretically explained trade-off between descriptive and predictive performance of the fitted model, depending on the available options for inference, is indeed verified and clearly demonstrated.

Section 6 resumes the article with additional remarks in connection with recent literature on INAR processes, remarks on potential alterations of the proposed model so as to be able to account for dependence structures more involved than first-order Markov, and remarks about prospects for utility of integer-valued time series models in prediction of spatial moments of intermittent spatio-temporal random fields.

2 Probabilistic properties of the INAR(1) model

The INAR(1) model is formally defined by the stochastic difference equation

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

where both the solution process $\{X_t\}_{t\in\mathbb{Z}}$ and the process of innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are non-negative integer-valued processes, i.e. taking values in the set of natural numbers $\mathbb{N}$. The model is specified by requiring that the innovation process $\{\varepsilon_t\}$ consists of i.i.d. random variables, and ε_t is stochastically independent of the past history $\mathcal{F}_{t-1} = \sigma(X_s; s < t)$, for every $t \in \mathbb{Z}$. The $\circ$- operation is called *binomial thinning*, and for given $\alpha \in [0, 1]$ is defined to act on a non-negative integer-valued random variable X, so that $\alpha \circ X := \sum_{i=1}^{X} Z_i$, with $\{Z_i\}_{i\geq 1}$ being an i.i.d. sequence of Bernoulli random variables, all independent of X, and each with distribution $Bin(1, \alpha)$; i.e. $\alpha = P(Z_i = 1) = 1 - P(Z_i = 0)$. Equation (1) describes a non-linear scheme of first-order autoregression, rendering $\{X_t\}$ a Markov chain on $\mathbb{N}$. The source of non-linearity of INAR(1) processes is the binomial thinning $\circ$-operator, originally introduced by *Steutel and Van Harn* [1979] in order to extend the notion of self-decomposability of distributions on the Borel line $\mathbb{R}$, to the notion of discrete self-decomposability of distributions on $\mathbb{N}$. Some elementary properties of this operation are summarized in Appendix A, and are used in calculations throughout the article.

2.1 The stationary distribution

Starting with any non-negative integer-valued random variable X_0, it follows from **A3** and **A4** that, after t iterations of (1), $X_t \overset{D}{=} \alpha^t \circ X_0 + \sum_{j=0}^{t-1} \alpha^j \circ \varepsilon_{t-j}$. Thus, if $E(X_0) < \infty$ and $\alpha < 1$, the term $\alpha^t \circ X_0$ converges in probability to zero, as $t \to \infty$. This is easily established using Markov's inequality and **A5**, whence $P\{\alpha^t \circ X_0 \geq 1\} \leq \alpha^t E(X_0)$. Therefore, the marginal probability distribution converges to that of the formal limit $\sum_{j=0}^{\infty} \alpha^j \circ \varepsilon_{t-j}$. Theorem 2.1 of *Alzaid and Al-Osh* [1990] implies that the probability generating function (p.g.f.) of this limit in law is

$$\phi(s) = \prod_{j=0}^{\infty} \psi\left(1 - \alpha^j + \alpha^j s\right), \tag{2}$$

provided that the following convergence condition holds

$$\sum_{j=1}^{\infty} \frac{P(\varepsilon_t \geq j)}{j} < \infty, \tag{3}$$

where $\psi(z) = E(z^{\varepsilon_t})$ is the p.g.f. of the marginal distribution of the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. That is, (3) is sufficient for (2), but also necessary and sufficient condition for convergence of the product on the RHS of (2) for every $s \in [0, 1]$ [*Heathcote*, 1966], so that ϕ is well defined in that domain. Moreover, Theorem 2.2 of *Alzaid and Al-Osh* [1990] implies that, if a process $\{X_t\}_{t\in\mathbb{Z}}$ started in the infinite past has reached a stationary marginal distribution, and satisfies the INAR(1) model

with binomial thinning parameter $\alpha < 1$ and innovations whose distribution satisfies (3), then the p.g.f. of the stationary marginal distribution is given by ϕ defined in (2).

These two facts, regarding solutions of (1) started in either finite or infinite past time, justify the causal integer-valued moving average of infinite order, INMA(∞), representation

$$X_t \quad \stackrel{D}{=} \quad \sum_{j=0}^{\infty} \alpha^j \circ \varepsilon_{t-j} \tag{4}$$

of stationary solutions $\{X_t\}$ of (1), which exist if and only if $\alpha < 1$. Furthermore, (4) implies that the marginal distribution of a stationary INAR(1) process is a discrete self-decomposable law on $\mathbb{N}$, uniquely determined by the law of the innovations; see also *Al-Osh and Alzaid* [1987]. Subsequently, discrete self-decomposability implies infinite divisibility and unimodality [*Steutel and Van Harn*, 1979] of the stationary marginal law. It is worthy noting that discrete stable laws form a subclass of discrete self-decomposable laws [*Van Harn*, 1978], and while Poisson laws are the only discrete stable laws with finite mean, it is also worthy to remark that a stationary INAR(1) process has a Poisson marginal law (with mean λ) if and only if the innovations follow a Poisson law too (with mean $\lambda \cdot (1 - \alpha)$).

2.2 Moments

The mean and variance of a stationary INAR(1) process $\{X_t\}_{t \in Z}$ are

$$\mu_X = E(X_t) = \frac{\mu_\varepsilon}{1-\alpha} \quad \text{and} \quad \sigma_X^2 = Var(X_t) = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2}, \tag{5}$$

where μ_ε and σ_ε^2 are respectively the mean and variance of the innovations. Formulae (5) are obtained directly from (1), exploiting properties **A5** and **A6**, independence between ε_t and $\alpha \circ X_{t-1}$, and accounting for stationarity. Consequently, the index of dispersion of the process, $ID(X) = \sigma_X^2/\mu_X$, is related to that of the innovations, $ID(\varepsilon) = \sigma_\varepsilon^2/\mu_\varepsilon$, according to the formula

$$ID(X) = \left(1 + \frac{ID(\varepsilon)}{\alpha}\right) \cdot \left(1 + \frac{1}{\alpha}\right)^{-1}, \tag{6}$$

showing that an INAR(1) process is overdispersed (i.e. $\sigma_X^2 > \mu_X$) if and only if the innovation process is overdispersed (i.e. $\sigma_\varepsilon^2 > \mu_\varepsilon$). The index of dispersion is widely used (e.g. in ecology) as a measure of clustering or repulsion, associated with overdispersion or underdispersion respectively, versus pure randomness. The latter case is associated with equidispersion, and is often represented by a Poisson law.

Every stationary INAR(1) process is positively correlated since it has positive auto-covariance function (ACVF) given by the formula [*Al-Osh and Alzaid*, 1987; *Alzaid and Al-Osh*, 1990]

$$\gamma_X(k) = Cov(X_t, X_{t-k}) = \alpha^{|k|}\sigma_X^2 \quad , \quad k \in \mathbb{Z}, \tag{7}$$

and thus the auto-correlation function (ACF) is also positive

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \alpha^{|k|} \ , \quad k \in \mathbb{Z}. \tag{8}$$

Moreover, since $\sum_{k=-\infty}^{+\infty} |\gamma_X(k)|$ is a convergent geometric series, the (non-normalized) power-spectrum of the process is absolutely continuous, and its spectral density function is obtained straight forwardly by the Fourier transform of the covariance function:

$$S_X(\omega) = \frac{\sigma_X^2}{\pi} \cdot \sum_{k=-\infty}^{+\infty} \alpha^{|k|}\ e^{-ik\omega} \ = \frac{\alpha \cdot \mu_\varepsilon + \sigma_\varepsilon^2}{\pi \cdot (1 - 2\alpha \cos\omega + \alpha^2)} \ \ , \ \omega \in [0, \pi]. \tag{9}$$

2.3 The case of mixed-Poisson innovations

Given a fixed integer $m \geq 1$, let

$$\varepsilon_t = \mathbf{Q}_t' \mathbf{\Lambda}_t = \sum_{i=1}^{m} Q_t^{(i)} \Lambda_t^{(i)}, \tag{10}$$

with $\left\{ \mathbf{Q}_t = \left(Q_t^{(1)}, \ldots, Q_t^{(m)} \right)' \right\}_{t \in \mathbb{Z}}$ and $\left\{ \mathbf{\Lambda}_t = \left(\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)} \right)' \right\}_{t \in \mathbb{Z}}$ being two independent processes of i.i.d. random vectors. Each vector $\mathbf{Q}_t$ is assumed to follow a multinomial distribution, having index 1 and weights $0 \leq p_1, \ldots, p_m \leq 1$, with $\sum_{i=1}^{m} p_i = 1$; i.e. $\mathbf{Q}_t \overset{D}{=} Multinomial\,(1,\ p_1, \ldots,\ p_m)$. Each vector $\mathbf{\Lambda}_t$ is assumed to consist of mutually independent Poisson random variables; i.e. $\Lambda_t^{(i)} \overset{D}{=} Poi\,(\lambda_i)\,,\ i = 1, 2, \ldots, m$.

In short, the probability distribution of each ε_t defined by (10) is a finite mixture of m Poisson components $\Lambda_t^{(1)}, \ldots, \Lambda_t^{(m)}$, with corresponding parameters $\lambda_1, \ldots,\ \lambda_m > 0$ and mixing weights $p_1, \ldots,\ p_m$. The probability function of these mixed-Poisson innovations is

$$P\,(\varepsilon_t = x) = \sum_{i=1}^{m} p_i e^{-\lambda_i} \frac{\lambda_i^x}{x!} \quad , \quad x = 0,\ 1,\ 2, \ldots \tag{11}$$

and the corresponding p.g.f. is given by the formula [e.g. see *Johnson et al.*, 1993 (pg. 326)]

$$\psi(z) = E\,(z^{\varepsilon_t}) = \sum_{i=1}^{m} p_i \exp\left\{\lambda_i (z-1)\right\}. \tag{12}$$

Setting $z = 1 - \alpha^j + \alpha^j s$ in (12) yields $\psi\left(1 - \alpha^j + \alpha^j s\right) = \sum_{i=1}^{m} p_i \exp\left\{\lambda_i \alpha^j (s-1)\right\}$. Thus, according to (2) the p.g.f. of the marginal law of a stationary INAR(1) process driven by these particular innovations is given by the formula

$$\phi\,(s) = \prod_{j=0}^{\infty} \left(\sum_{i=1}^{m} p_i \exp\left\{\lambda_i \alpha^j (s-1)\right\} \right). \tag{13}$$

The infinite product on the RHS of (13) does converge, so that ϕ is a well defined function on the interval $[0, 1]$. The fact that mixed-Poisson innovations indeed satisfy the convergence condition (3) is detailed with a proof in Appendix C. In addition, Appendix C provides formulae for (central and non-central) moments of mixed-Poisson laws. Appendix B demonstrates the use of such formulae in calculating central moments $\mu_r(X) = E\left[(X - EX)^r\right]$, for $r = 3, 4$, of stationary INAR(1) processes driven by innovations defined according to (10). Thereof, explicit formulae for the coefficients of

skewness $IS(X) = \mu_3(X)/\sigma_X^3$ and kurtosis $IK(X) = \mu_4(X)/\sigma_X^4$ may be obtained, along with the index of dispersion $ID(X)$ given by (6). Indeed, **B3** combined with (5) renders $IS(X)$ in terms of μ_ε, σ_ε^2, $\mu_3(\varepsilon)$, which are given in Appendix C along with a sufficient condition for positive skewness of the law of X. A more elaborate but of similar nature calculation based on **B2** and **A11** readily renders $IK(X)$ explicit too, in terms of μ_ε, σ_ε^2, μ_ε, σ_ε^2, $\mu_3(\varepsilon)$, and $\mu_4(\varepsilon)$. Notably, the coefficients of skewness and kurtosis are unit-free indexes used for characterization of the shape of a distribution, while the index of dispersion does depend on the scale of units used for measurement.

3 Statistical inference

3.1 Transition probability function and conditional likelihood

Substituting the representation (10) of mixed-Poisson innovations into (1), and accounting for the fact that $\sum_{i=1}^{m} Q_t^{(i)} = 1$, the stationary solution $\{X_t\}$ of (1) may be represented at each time $t \in \mathbb{Z}$ by the sum

$$X_t = \sum_{i=1}^{m} Q_t^{(i)} \cdot X_t^{(i)}, \tag{14}$$

where $X_t^{(i)} = \alpha \circ X_{t-1} + \Lambda_t^{(i)}$, $i = 1, 2, \ldots, m$.
The conditioned random variable $\left(X_t^{(i)} \mid X_{t-1}\right) = (\alpha \circ X_{t-1} \mid X_{t-1}) + \Lambda_t^{(i)}$ follows a Poisson-binomial law $PB\left(X_{t-1}, \alpha; \lambda_i\right)$, that is the convolution of the independent random variables $(\alpha \circ X_{t-1} \mid X_{t-1}) \sim Bin\left(X_{t-1}, \alpha\right)$ and $\Lambda_t^{(i)} \sim Poi\left(\lambda_i\right)$. Therefore, the conditioned random variable $(X_t \mid X_{t-1}) = \sum_{i=1}^{m} Q_t^{(i)} \cdot \left(X_t^{(i)} \mid X_{t-1}\right)$ follows a finite mixture law of m (independent) Poisson-binomial random variables $PB\left(X_{t-1}, \alpha; \lambda_i\right)$, with mixing weights p_i, $i = 1, 2, \ldots, m$. Consequently, the transition probability function of the Markov chain $\{X_t\}$ is identical to the probability function of this mixed Poisson-binomial law, and is given by

$$P_{\theta,m}\left(X_t = x_t \mid X_{t-1} = x_{t-1}\right) = \sum_{i=1}^{m} p_i \cdot \pi_i(x_t \mid x_{t-1}), \tag{15}$$

where

$$\pi_i(x_t \mid x_{t-1}) = \sum_{k=0}^{x_t \,\wedge\, x_{t-1}} e^{-\lambda_i} \cdot \frac{\lambda_i^k}{k!} \cdot \binom{x_{t-1}}{x_t - k} \cdot \alpha^{x_t - k} \cdot (1-\alpha)^{x_{t-1} - x_t + k} \tag{16}$$

is the probability function of $PB\left(x_{t-1}, \alpha; \lambda_i\right)$, defined for $x_t \in \mathbb{N}$, with $x_t \,\wedge\, x_{t-1} = \min(x_t, x_{t-1})$, and $\theta = (\alpha; p_1, \ldots, p_{m-1}; \lambda_1, \ldots, \lambda_m)$ is the vector of the $2m$ effective parameters of the INAR(1) model. Details on the Poisson-binomial law are given by *Shumway and Gurland* [1960].

Given a time series of observed counts $\mathbf{x} = \{x_0, x_1, x_2, ..., x_T\}$, let $L(\theta; m; \mathbf{x}) = P_{\theta,m}(X_0 = x_0, X_1 = x_1, ..., X_T = x_T)$ denote its (full) likelihood according to the INAR(1) model with parameters θ, for fixed m. Due to the Markov property this likelihood reduces to the product $L(\theta; m; \mathbf{x}) = P_{\theta,m}(X_0 = x_0) \cdot \prod_{t=1}^{T} P_{\theta,m}(X_t = x_t | X_{t-1} = x_{t-1})$, whose maximization with respect to θ is a difficult problem, especially when the dimension of θ is large. This difficulty stems from the fact that there is

no explicit expression of the marginal law, despite expression (12) of its p.g.f. as an infinite product. However, in light of the explicit expression (16) for transition probabilities, maximization of the conditional likelihood $CL(\theta; m; \mathbf{x}) = \prod_{t=1}^{T} P_{\theta,m}(X_t = x_t | X_{t-1} = x_{t-1})$ is feasible. The sense in which this is a conditional likelihood is that the initial value x_0 is treated as a fixed constant, and not as a randomly varying observation. Thus, the remaining product, $CL(\theta; m; \mathbf{x}) = L(\theta; m; \mathbf{x})/P_{\theta,m}(X_0 = x_0)$, represents (under the considered model) likelihood of the observed data among the ensemble of all possible time series of the same length $(T+1)$ and with the same initial observation x_0.

3.2 An EM-algorithm

From (15) and (16) one can see that the conditional likelihood is the likelihood of a finite mixture, namely of Poisson-binomial laws. Thus standard tools for inference in finite mixtures can be used, such as the EM algorithm, for likelihood maximization (e.g. see *Titterington et al.* [1985], *Bohning* [2000], *McLachlan and Peel* [2001]). An EM-algorithm appropriate for carrying out numerically the task of CML estimation of θ, for fixed m, is developed here. Starting with an initial estimate θ_0 in the admissible range of parameters, the algorithm comprises of iterative updating in two steps.

At the E-step, one estimates conditional expectations of latent information about the model, given the available observations and the estimate θ^{old} from the previous iteration (or θ_0 in the very first iteration). In this case the latent information lies in the innovations $\{\varepsilon_t; t = 1, ..., T\}$, which in turn are made of the non-observable variables $\left\{Q_t^{(i)}; t = 1, ..., T\right\}_{i=1}^{m}$ and $\left\{\Lambda_t^{(i)}; t = 1, ..., T\right\}_{i=1}^{m}$. Thus, at the E-step (expectation step), for each $i = 1, ..., m$ and $t = 1, \ldots, T$ the following information is computed.

E-step:

$$
\begin{aligned}
q_t^{(i)} &= E\left\{Q_t^{(i)} \mid \mathbf{x}; \theta^{old}\right\} = p_i^{old} \cdot \frac{\pi_i^{old}(x_t \mid x_{t-1})}{\sum_{j=1}^{m} p_j^{old} \cdot \pi_j^{old}(x_t \mid x_{t-1})}, \\
\ell_t^{(i)} &= E\left\{\Lambda_t^{(i)} \mid \mathbf{x}; \theta^{old}\right\} = \lambda_i^{old} \cdot \frac{\pi_i^{old}(x_t - 1 \mid x_{t-1})}{\pi_i^{old}(x_t \mid x_{t-1})}.
\end{aligned}
$$

Assuming that the data $\mathbf{x}$ are augmented with values of the latent variables as specified in the E-step (i.e. conditionally on the data and in terms of θ^{old}), then θ^{old} is updated to θ^{new} maximizing the likelihood of the augmented data. In the present case this task is straight forward, involving estimation in a standard binomial law and in a finite mixture of Poisson laws. Thus, the M-step (maximization step) yields the following updates for each $i = 1, ..., m$.

M-step:

$$a^{new} = \frac{\sum_{t=1}^{T} x_t - \sum_{t=1}^{T}\sum_{i=1}^{m} q_t^{(i)} \cdot \ell_t^{(i)}}{\sum_{t=1}^{T} x_{t-1}},$$

$$p_i^{new} = \frac{\sum_{t=1}^{T} q_t^{(i)}}{T} \quad \text{and} \quad \lambda_i^{new} = \frac{\sum_{t=1}^{T} q_t^{(i)} \cdot \ell_t^{(i)}}{\sum_{t=1}^{T} q_t^{(i)}}.$$

Iterations may be stopped as soon as some convergence criterion is satisfied, otherwise the algorithm returns to the E-step for a new iteration. All the pros and cons of the standard EM algorithm apply (e.g. see *McLachlan and Krishnan* [1997]). Although convergence to estimates in the admissible range is guaranteed, the speed of the algorithm can be slow, especially when the dimension of θ is large. In order to improve speed one may execute a few iterations that will approximate the solution closely enough, and then locate the maximum using standard numerical techniques (e.g. Newton-Raphson). Notably, few iterations are usually enough to get estimates quite close to the maximum. Another common problem in the implementation of the EM-algorithm to finite mixture models is the existence of local maxima of the (conditional) likelihood, which can trap the estimates in their vicinity. In order to overcome this problem in the application presented in Section 5, the algorithm was run for 20 different initial vectors θ_0, chosen randomly for given m, and then it was checked if more than 3 out of the 20 computed solutions reached the largest of maxima, in which case the procedure was stopped. Otherwise, if only 3 or less solutions reached the largest of maxima, the algorithm was run anew with additional initial vectors, until the largest maximum was reached at least 3 times. In each run, the convergence criterion used for stopping the algorithm is that the relative change of the (conditional) log-likelihood between two successive iterations becomes smaller than 10^{-12}, which is quite strict. Of course, the obtained solution always depends on the initial observation x_0.

Detailed accounts of the general properties of the EM-algorithm can be found in *Dempster et al.* [1977] and *McLachlan and Krishnan* [1997], while *Bohning* [2000] and *McLachlan and Peel* [2001] provide details on the EM- and other algorithms for the case of finite mixtures.

3.3 Alternative options of estimation

According to (8) the thinning parameter, α, is equal to the value of the ACF at lag-1. This very fact prompts estimation of α by the sample estimate of the auto-correlation of lag-1. This is quite a reasonable alternative to CML estimation of α, with the definite advantages that, (a) it does not depend on the distribution of innovations, thus is also independent of the choice of m, and (b) lag-1 auto-correlation of the fitted model matches exactly with lag-1 sample auto-correlation of the actual data to which the model is fitted (for any m). The only shortcoming is that this estimate may be a negative number, however not smaller than -1. In such a case, one may abort

modelling with the proposed model, or may still pursue it by setting $\alpha = 0$. In the latter case, the data are modelled as integer-valued i.i.d. noise with mixed-Poisson law. Hereafter, the estimate $\overset{\wedge}{\alpha} = \max\left\{0, \sum_{t=0}^{T-1}(x_t - \bar{x})(x_{t+1} - \bar{x}) / \sum_{t=0}^{T}(x_t - \bar{x})^2\right\}$ is referred to (for brevity) as method of moments (MOM) estimate of the thinning parameter α, where $\bar{x} = (T+1)^{-1} \cdot \left(\sum_{t=0}^{T} x_t\right)$ is the sample mean of the data. In this case the rest of the parameters can be still obtained using the CML method and the EM algorithm described in the previous section, but now keeping the value of α constant in all iterations; i.e. by setting $\alpha^{new} = \alpha_0 = \overset{\wedge}{\alpha}$ in the formulae of the previous section. The estimates obtained with this method are also referred to as MOM estimates, emphasizing that they have been obtained with α estimated by $\overset{\wedge}{\alpha}$.

To emphasize further the significance of the MOM option, it should be mentioned that the CML estimate of α by the EM-algorithm, say α_{CML}, usually under-estimates α compared to the MOM estimate $\overset{\wedge}{\alpha}$, provided that $\overset{\wedge}{\alpha} > 0$, and the deficit of α_{CML} versus $\overset{\wedge}{\alpha} > 0$ keeps increasing with m. This effect and the fact that α determines completely the ACF (8) of the INAR(1) model, suggest that it is best to not estimate α based solely on the EM-algorithm, because then the entire ACF of the fitted model can be seriously deflated, thus handicapping the performance of any reasonable predictors that one may be able to construct. Instead, if α is estimated by $\overset{\wedge}{\alpha}$, then the performance of competent predictors can be assessed more objectively, at least at lag-1 where the auto-correlation of the data and the auto-correlation of the fitted model are identical.

Two explanations of the effect of m on α_{CML} are offered here, beyond plain empiricism from computational experiments. One line of reasoning is that by increasing the order m of the mixture law of innovations one is approaching the so-called non-parametric maximum likelihood estimate (NPMLE); i.e. the mixture distribution that provides the best likelihood with respect to all probability measures [*Lindsay*, 1995]. Therefore, by increasing m the fitted model comes as close to the observed data as it can possibly be, thus optimizing distributional fit, but at the cost of ignoring and even depleting the serial dependence structure inherent in the data. Notably, several authors use the NPMLE as a semi-parametric estimate of the underlying probability distribution. A second explanation is based on equation (6), whence it is easy to check that α is a decreasing function of $ID(X)$, when $ID(\varepsilon) > 1$. This is indeed the case when the innovations follow a finite mixture of $m > 1$ Poisson laws. By increasing m in the INAR(1) model fitted to a given series of overdispersed count data, one evidently increases $ID(X)$ of the fitted model so as to match the sample index of dispersion of the data. Therefore, a decay in values of α is naturally anticipated, which propagates even further to degrade lag-1 auto-correlation and thereof the entire ACF of the fitted model.

In light of the above discussed properties of the estimators α_{CML} and $\overset{\wedge}{\alpha}$, one might be interested to consider as a third option a weighted estimator $\widetilde{\alpha} := \pi \cdot \alpha_{CML} + (1-\pi) \cdot \overset{\wedge}{\alpha}$. The role of the weight $0 \leq \pi \leq 1$ is to regulate the balance of the fitted model between excellent descriptive qualities with regard to marginal probabilistic characteristics of the data (when $\pi = 1$ and $\widetilde{\alpha} = \alpha_{CML}$) and reasonably adequate predictive qualities (when $\pi = 0$ and $\widetilde{\alpha} = \overset{\wedge}{\alpha}$). In this case, it should also be understood that the remaining parameters $(p_1, \ldots, p_m; \lambda_1, \ldots, \lambda_m)$ of the model ought to be

re-estimated by the EM-algorithm, but keeping now $\alpha^{new} = \alpha_0 = \widetilde{\alpha}$ in every iteration.

3.4 Specifying the order of the mixture

The methods of estimation presented above, and also simulation and prediction methodology to be presented in Section 4, presume that a fixed number of components in the Poisson mixture law of the innovations is given. However, none of those methods specified how this number $m \geqslant 1$ ought to be selected. The purpose of this section is to describe procedures according to which the size of m can be decided with reason.

A systematic and rather objective procedure is to fit the proposed model, to the count data of one's interest by any of the above suggested methods of inference, beginning with $m = 1$. Then, using the same method of estimation, keep fitting the model to the same data by increasing the value of m successively until the log-likelihood stops increasing (e.g. see *Lindsay* [1995], *Bohning* [2000]). The only shortcoming of this procedure is that it is not keen to parsimony, and may often lead to over-parameterized models with components which may not contribute substantially to the overall fit.

A different strategy, quite standard in mixture literature as well as in time series model selection, would be to select m according to information criteria such as Akaike Information Criterion (AIC), where the conditional log-likelihood is penalized by the number of estimated parameters, or AIC corrected (AICC) and Bayesian Information Criterion (BIC), where the penalty accounts also for the sample size T. Further criteria for selecting the order m of a mixture model are discussed in *McLachlan and Peel* [2001].

4 Simulation, prediction and performance assessment

Given a fixed m and a vector of parameters $\theta = (\alpha; p_1, \ldots, p_{m-1}; \lambda_1, \ldots, \lambda_m)$, simulation of synthetic time series from the INAR(1) model with mixed-Poisson innovations is quite straight forward. The following algorithm can be easily implemented to simulate a single stationary series of desired length N.

1. First generate a single value X_0 from a mixed-Poisson law with parameters $(p_1, \ldots, p_m; \lambda_1, \ldots, \lambda_m)$ in line with (10) and (11).

2. Then generate M values $\{X_t; t = 1, \ldots, M\}$, such that each X_t is the sum of two independent random variables, the one following a binomial law $Bin\,(X_{t-1}, \alpha)$, and the other following again a mixed-Poisson law with parameters $(p_1, \ldots, p_m; \lambda_1, \ldots, \lambda_m)$.

Note that the M iterations of step 2 follow the Markov recursion of the INAR(1) model according to (1), starting with the initial value of X_0 from step 1, and the M values generated from the mixed Poisson law with parameters $(p_1, \ldots, p_m; \lambda_1, \ldots, \lambda_m)$ in step 2 are realizations of i.i.d. innovations. Moreover, since the law of X_0 from step 1 has finite mean, $\mu_\varepsilon = \sum_{i=1}^{m} p_i \lambda_i$, the argument given in

the very beginning of Section 2.1 guarantees that after a large enough number of iterations, say $n \leq M$, the marginal law of the simulated values $\{X_t; t = n+1, \ldots, M\}$ is effectively a very good approximation of the stationary law with p.g.f. (13). Thus, considering the first n iterations of step 2 as "burn-in" period, the remaining $N = M - n$ values $\{X_t; t = n+1, \ldots, M\}$ represent the desired synthetic stationary series.

The rest of this section is concerned with prediction of time series of counts under a stationary INAR(1) model (i.e. $\alpha < 1$), driven by i.i.d. innovations with finite variance, and known parameters, either given or estimated by fitting such a model to data. As usual, the setting assumes that a regular discrete time series of non-negative integers $X_1 = x_1, X_2 = x_2, ..., X_T = x_T$ has been observed, and for some lead-time $k \geq 1$, a future value corresponding to the random variable X_{T+k} must be predicted in terms of known information consisting of the l most recent observed values $X_{T-l+1} = x_{T-l+1}, X_{T-l+2} = x_{T-l+2}, ..., X_T = x_T$, where $1 \leq l \leq T$.

4.1 Linear prediction

It is well known [e.g. see *Priestley*, 1981, Chapter 10] that the optimal predictor of X_{T+k}, given $X_{T-l+1}, ..., X_T$, is the conditional expectation

$$\hat{X}_{T+k} = E\left(X_{T+k} \quad | \quad X_{T-l+1}, X_{T-l+2}, ..., X_T\right), \tag{17}$$

in the sense that this predictor minimizes the mean square error of the prediction

$$MSE(k) = E\left\{\left(X_{T+k} - \hat{X}_{T+k}\right)^2\right\}. \tag{18}$$

After k iterations of (1), $X_{T+k} \overset{D}{=} \alpha^k \circ X_T + \sum_{j=1}^{k} \alpha^{k-j} \circ \varepsilon_{T+j}$, whence by taking conditional expectations and using basic properties of the thinning operator (Appendix A) it readily follows that

$$\hat{X}_{T+k} = \alpha^k \cdot X_T + \left(1 - \alpha^k\right) \cdot \frac{\mu_\varepsilon}{1-\alpha}, \tag{19}$$

whence one also derives that

$$E\left(\hat{X}_{T+k}\right) = \alpha^k \cdot \mu_X + \left(1 - \alpha^k\right) \cdot \frac{\mu_\varepsilon}{1-\alpha}, \tag{20}$$

$$V\left(\hat{X}_{T+k}\right) = \alpha^{2k} \cdot \sigma_X^2 \ , \tag{21}$$

$$MSE(k) = \left(1 - \alpha^{2k}\right) \cdot \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2}. \tag{22}$$

From (19) it is clear that the optimal predictor $\hat{X}_{T+k}$ is linear with respect to the most recent observation X_T. Note that if $\mu_X = \mu_\varepsilon \cdot (1-\alpha)^{-1}$, then $E\left(\hat{X}_{T+k}\right) = \mu_X$, and if $\sigma_X^2 = \left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) \cdot \left(1-\alpha^2\right)^{-1}$, then $MSE(k) + V\left(\hat{X}_{T+k}\right) = \sigma_X^2$, both of which do hold if the X-process is indeed INAR(1). In general, however, even when the X-process is merely modelled as INAR(1),

without necessarily being such, the asymptotic values of $E\left(\hat{X}_{T+k}\right)$ and $MSE(k)$ are respectively $\mu_\varepsilon \cdot (1-\alpha)^{-1}$ and $\left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) \cdot \left(1-\alpha^2\right)^{-1}$, as $k \to \infty$, while $V\left(\hat{X}_{T+k}\right)$ tends to zero, and the predictions themselves tend (in probability) to the same constant value $\mu_\varepsilon \cdot (1-\alpha)^{-1}$ as their mean.

4.2 Integer-valued prediction

The linear predictor $\hat{X}_{T+k}$ suffers two shortcomings. One is that after a few lead times it degenerates to nearly constant predictions, that is no prediction. The other is that it produces positive real-valued predictions, which are difficult to interpret meaningfully in light of the distinctively integer character of the observations. Therefore, an alternative approach is proposed and investigated here, resolving both these issues.

Given observations of the past variables $X_{T-l+1}, ..., X_T$, and taking into account the non-linear auto-regressive Markov structure of the model used for prediction, we propose prediction of a future variable X_{T+k} by the integer-valued and non-linear predictor

$$\widetilde{X}_{T+k} = \alpha^k \circ X_T + \sum_{j=1}^{k} \alpha^{k-j} \circ \widetilde{\varepsilon}_{T+j}, \tag{23}$$

formulated by k iterations of the INAR(1) model, starting from the most recent observation X_T, and driven by simulated values $\widetilde{\varepsilon}_{T+j}$ of the (non-observable) i.i.d. innovation process.

If X_T follows the marginal law of the stationary INAR(1) model, then so does the predictor $\widetilde{X}_{T+k}$ too, whence $E\left(\widetilde{X}_{T+k}\right) = \mu_X$ and $V\left(\widetilde{X}_{T+k}\right) = \sigma_X^2 > V\left(\hat{X}_{T+k}\right)$, where μ_X and σ_X^2 are further specified according to (5). In general, however, when the X-process is just modelled as INAR(1), without necessarily being such, it follows from (23) that

$$E\left(\widetilde{X}_{T+k}\right) = \alpha^k \cdot \mu_X + \left(1-\alpha^k\right) \cdot \frac{\mu_\varepsilon}{1-\alpha}, \tag{24}$$

$$V\left(\widetilde{X}_{T+k}\right) = \alpha^{2k} \cdot \sigma_X^2 + \alpha^k \cdot (1-\alpha^k) \cdot \mu_X + \frac{1-\alpha^{2k}}{1-\alpha^2} \cdot \sigma_\varepsilon^2 + \left(\frac{1-\alpha^k}{1-\alpha} - \frac{1-\alpha^{2k}}{1-\alpha^2}\right) \cdot \mu_\varepsilon. \tag{25}$$

From (20) and (24) it is clear that $E\left(\widetilde{X}_{T+k}\right) = E\left(\hat{X}_{T+k}\right)$, and thus all the comments made about $E\left(\hat{X}_{T+k}\right)$ remain valid for $E\left(\widetilde{X}_{T+k}\right)$ too. From (21) and (25) it is also clear that always $V\left(\widetilde{X}_{T+k}\right) > V\left(\hat{X}_{T+k}\right)$. Asymptotically, $V\left(\widetilde{X}_{T+k}\right)$ tends to $\left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) \cdot \left(1-\alpha^2\right)^{-1}$, as $k \to \infty$. This limit is the variance of the INAR(1) model, thus it is positive and guarantees that the integer-valued predictions will never degenerate to a constant. This same limit is also the variance of the process only if $\sigma_X^2 = \left(\alpha\mu_\varepsilon + \sigma_\varepsilon^2\right) \cdot \left(1-\alpha^2\right)^{-1}$, which happens for example when X_T follows the marginal law of the stationary INAR(1) model.

4.3 Assessment of model performance

Model performance is used here as a term that broadens the meaning of the more classical notion of goodness of fit, in the sense that the model at hand is not merely a probability distribution

whose goodness of fit to data must be assessed. Instead, it is a model for integer-valued time series, which aims to capture features of the underlying dependence structure represented by second-order characteristics (i.e. auto-correlations), along with features of variability represented by moment characteristics of the underlying marginal probability distribution.

In general, the construction of statistical criteria suitable for evaluation of model performance in time series settings is a task that may be carried out in different ways, depending on the focus of interest. This implies "relativity" among criteria of model performance, in the sense that a model considered adequate with respect to one criterion may fail seriously to fulfill another. Nevertheless, model performance as well as goodness of fit criteria are measures of proximity of information produced by the fitted model and by the observed data whose stochastic structure the fitted model is anticipated to mime. Usually such proximity is measured by some kind of collective loss of information between the actual data and the predicted or simulated data produced by the fitted model.

A definite restriction with respect to the case at hand, is that most of the existing literature on measures of model performance or goodness of fit for time series uses normality assumptions and/or asymptotic arguments for large samples. Avoiding these restrictions, *parametric bootstrap* is undertaken here adopting the approach suggested by *Tsay* [1992]. That is, to generate data from the fitted model and to check if their deviation from the observed data could have occurred if the observed series had been generated by the assumed model. This approach enables one to assess how closely the assumed model represents the stochastic evolution of the data by implementing the following test statistic

$$D = \sum_{t=2}^{T} \frac{(X_i - \hat{X}_i)^2}{s^2(\hat{X})}.$$

That is, D is the sum of normalized square residuals between the observed data X_i and their (e.g. one-step-ahead) predictions $\hat{X}_i$, in this case given by (19) or (23) for $k = 1$. The normalization $s^2(\hat{X})$ is an estimate of the variance of the predictions, as in (21) or (25) respectively, with all the involved parameters taking the same values as in the fitted model. Alternatively, $s^2(\hat{X})$ can be just the sample variance of the predictions made by (19) or (23). In the applications presented in Section 5 the normalization $s^2(\hat{X})$ is the sample variance of integer-valued predictions made by (23), for $k = 1$ and $k = 2$.

Instrumental to this approach is to obtain an empirical distribution of the test statistic D, and to check if the observed value of D belongs to an upper tail of this distribution, or not, given a desired level of significance. This test is computationally demanding, as one must fit the model to each series simulated from the fitted model, whose performance is under testing, and must also make predictions based on the simulated series. However, since good initial values are available the EM-algorithm presented in 3.2 converges fairly quickly.

Of course, the (conditional) likelihood of the fitted model offers an alternative measure of performance, which one may consider in parallel with the above proposed statistic D, as a counterpart reflecting descriptive abilities rather than predictive ones.

5 Application to rainfall

The INAR(1) model can be thought of as describing dynamics of a branching process with immigration in discrete time. In that framework, the thinned term $\alpha \circ X_t$ represents the number of survivors from the t-th generation of a population with X_t members, to the very next $(t+1)$-st generation with $X_{t+1} = \alpha \circ X_t + \varepsilon_{t+1}$ members, while the innovation term ε_{t+1} represents immigrants added (independently of X_t) to the $\alpha \circ X_t$ survivors. This viewpoint is conducive with the application of INAR(1) to spatial rainfall, presented herein.

The population of interest consists of pixels where spatially averaged rain rate (SARR) exceeds a given threshold level, in an instantaneous radar scan mapping a rain field over a fixed region. The role of survivors between two successive scans of the probed region is played by pixels that maintain SARR values greater than the given threshold in both scans. Pixels whose SARR values increased so as to up-cross and exceed the threshold level, from values lower than or equal to the threshold in the previous scan, play the role of immigrants. Moreover, since the spatial organization of rain fields is quite complex, comprising of light and moderately intense rain from stratiform clouds with wide spatial coverage, together with very intense rain from convective cells occurring very locally, and several other intermediate regimes in the evolution of a storm, it is reasonable to model immigration by considering innovations with mixed structure, such as that of mixed Poisson laws.

Motivation for pursuing this sort of application of the proposed INAR(1) model, to time series of (instantaneous) counts of pixels where SARR exceeds a threshold, is prompted from the need to predict temporal evolution of the fraction of regional coverage, τ-FRC, with rain intensity exceeding a certain "*optimal*" threshold $\tau \geq 0$. If τ-FRC can be predicted over time by a suitable time series model, then the *threshold method* (TM) can be used to recast temporal predictions of τ-FRC into temporal predictions of SARR over the entire probed region. Such information is quite valuable input for further hydrological, meteorological and climatological analysis and predictions.

When the size of pixels is sufficiently small, relative to the size of the probed region, τ-FRC can be approximated reasonably well by the ratio of the count of pixels where rain intensity exceeds τ, to the total number of pixels in a scan. However, the total number of pixels in any given scan is fixed, determined by the spatial resolution of pixels, while the count in the numerator of τ-FRC fluctuates randomly as the rain field evolves. Therefore, prediction of time series of counts of pixels where SARR exceeds a given "optimal" threshold τ, can indeed play a key role in predicting τ-FRC and subsequently SARR by implementation of TM. In this sense, the present application may be viewed as an effort aiming to parametric modelling of τ-FRC time series of a temporally evolving rain rate random field, for predictive as well as for descriptive purposes.

TM works more generally for prediction of spatial moments of non-negative random fields with sufficient intermittency between positive and zero values, and not just for prediction of spatial averages (i.e. spatial moments of first order). Thus, the idea mentioned above about recasting temporal predictions of τ-FRC to predictions of SARR, can in principle be also applied towards temporal prediction of (general) spatial moments. Detailed accounts on the statistical foundation of TM, on criteria for choosing optimal thresholds, on the performance of the method in predicting

spatial moments of rain rate, and on its valuable utility when remote sensing technology (i.e. radar, satellite) is implemented for measuring precipitation, are given in *Kedem and Pavlopoulos* [1991], *Shimizu et al.* [1993], *Shimizu and Kayano* [1994], *Mase* [1996], *Meneghini et al.* [2001].

5.1 Presentation of the data

The raw data is a regular time series of maps of radar reflectivity measurements obtained during the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE) by shipboard Doppler precipitation radar (MIT). Each scan probes a fixed oceanic region of area $240 \times 240 Km^2$ in the tropical sector of South Western Pacific Ocean (China Sea: $2^o S$, $156^o E$). Measurements of radar reflectivity echo Z (in dBZ units) from each scan, binned by spatial averaging over square pixels of area $2 \times 2\ Km^2$, have been converted to instantaneous rain rate R values (in mm/hr units) by the Z-R relationship $R = (Z/230)^{1/1.25}$. The temporal resolution between successive scans is 20 minutes, and the sample size of the regular time series considered here is $T = 184$, corresponding to a nearly three day segment (02:01 UTC, November 22, 1992 through 15:01 UTC, November 24, 1992). This is the longest regular segment (i.e. without missing scans) of Cruise 1 (20:01 UTC, November 10, 1992 through 23:41 UTC, December 9, 1992) of TOGA-COARE. In total Cruise 1 consists of 1991 scans, with a percentage of missing scans 5.19% distributed in 49 blocks of length mostly 1 or 2 (scans), and occasionally longer (the longest corresponding to 18 missing scans, amounting to a six hour gap of information). Detailed documentation of TOGA-COARE is given by *Short et al.* [1997].

Along with the region of reference, which is a square of side length 240 Km, seven square sub-regions are also considered, having the same center as the region of reference and side lengths 120 Km, 60 Km, 32 Km, 16 Km, 8 Km, 4 Km, 2 Km respectively. For each one of these centrally nested regions, time series of counts of pixels where R exceeds the threshold levels of 0 mm/hr, 1 mm/hr, 2 mm/hr, 5 mm/hr, 10 mm/hr, and 20 mm/hr were recorded from the regular time series of 184 radar scans. That is, six time series of counts for each one of the eight regions were recorded, rendering a total of 48 different series of pixel counts. The threshold of 0 mm/hr theoretically discriminates between wet and dry pixels. However, due to known deficiencies of radar technology in discerning very light rain from noise induced by plain humidity or cloud, the thresholds of 1 mm/hr or even 2 mm/hr might alternatively be considered for mapping information of spatial intermittency between wet and dry states of rainfall. The thresholds of 5 mm/hr, 10 mm/hr and 20 mm/hr are considered here only as proxies of optimal threshold levels reported in pertinent literature [*Kedem and Pavlopoulos*, 1991; *Short et al.*, 1993; *Kayano and Shimizu*, 1994].

Due to the nesting of the considered sub-regions, the count of pixels for any given threshold level is an increasing (albeit random) function of the spatial scale (or size) of the region. It is also anticipated that zero counts prevail in the smaller spatial scales and for the larger thresholds, but become more rare in regions of large spatial scales, especially for low thresholds considered there. Temporal intermittency between positive and zero values of counts, representing evolution of spatial intermittency with respect to a threshold level of precipitation intensity, is partially indicated by

the probability or sample proportion of zero counts. If this indicator is very close to its extreme bounds, 0 and 1, then the time series of counts contains only a few zeroes or is made mostly of zeroes (respectively), and in this sense its intermittency ought to be considered low.

Table 1 provides information of sample statistics (mean, index of dispersion, coefficients of skewness and kurtosis, proportion of zeroes, and lag-1 auto-correlation) from the 48 time series described above. Some series from the regions of smaller scales (4 Km and 2 Km) comprised only of zeroes, thus no statistics are reported there except the obvious values of sample mean and sample proportion of zeroes. Notably, in most cases the overdispersion is nontrivial, accompanied by pronounced (positive) skewness and kurtosis. The negative skewness of series corresponding to the threshold 0 mm/hr in scales of 32 Km and lower, results from *all* pixels being "wet" for a sufficiently large proportion of scans. However, this may be considered merely as an artifact of the rather coarse resolution by pixels of scale 2 Km relative to scales of 32 Km or lower, in combination with deficiencies of radar technology in discerning rain of very low intensity from plain noise.

Certain trends emerging from Table 1 may be summarized as follows. The sample proportion of zeroes indeed increases when the spatial scale decreases and when the threshold level increases, as anticipated earlier. The same two trends are also noted for the coefficients of skewness and kurtosis, with the exception of the 0 mm/hr threshold level, where deviations from these trends might be attributed to the aforementioned deficiencies of radar technology. Exactly the reverse type of trends is noted about the sample mean and the index of dispersion, both of which decrease when the threshold level increases and when the spatial scale decreases. Auto-correlation also presents a tendency to decay as the spatial scale decreases (for fixed threshold) and when the threshold level increases (for fixed scale), although not as consistently as noted in all the other trends.

5.2 Results from the fitted model

The performance of the proposed INAR(1) model is demonstrated in the rest of this section. The model was fitted to the time series of pixel counts presented above by the methods of estimation discussed in Section 3. For demonstrative purposes six cases have been chosen based on the criterion that there is sufficiently moderate temporal intermittency in the observed series of counts. This criterion excludes the regions of 2 Km, 4 Km, and 240 Km, since intermittency there is very little, as indicated by the proximity of sample proportion of zero counts to 1 and 0 respectively (see Table 1). Specifically, the cases of 5 mm/hr in 120 Km, 1 mm/hr and 20 mm/hr in 60 Km, 10 mm/hr in 32 Km, 0 mm/hr in 16 Km, and 2 mm/hr in 8 Km, have been chosen. These are quite representative of very similar types of behavior seen in the bulk of cases with moderate to high intermittency, and also representative of the performance of the fitted model in such cases. An exception is the case of 0 mm/hr in 16 Km, which represents cases with very low intermittency (due to only a few zeroes) and the artifact of negative skewness pointed out earlier.

Table 2 provides information about the fitted model (mean, dispersion index, skewness and kurtosis coefficients, proportion of zeroes, order m, thinning parameter α, conditional log-likelihood, D-statistic and its P-value) in these six cases, using the CML ($\pi = 1$) and MOM ($\pi = 0$) options

for estimation of parameters, when the order m is chosen according to AIC. This information offers a first impression about the model's adequacy, by simple comparison with the corresponding information in Table 1 (marked with bold characters). It is clear that all the marginal characteristics (mean, ID, IS, IK, proportion of zeroes) are recovered much more adequately by CML than by MOM, and the likelihood of the CML-fitted model is significantly greater than that of the MOM-fitted model. On the contrary, comparing the estimated values of the thinning parameter α in Table 2 with lag-1 sample auto-correlation in Table 1, it is seen that the MOM version recovers the ACF much more adequately than its CML counterpart does. Judging also by the values of the statistic D and the corresponding P-values, indicating significant model performance based on predictions, it is clear that the model fitted by the MOM option predicts more adequately than its CML counterpart.

In fact, the very transition from the better predictive ability of the MOM version ($\pi = 0$) to the better descriptive ability of the CML version ($\pi = 1$), is further detailed in Table 3. It is clearly seen there that by combining both options to fit the model, with weights determined by $0 < \pi < 1$, the log-likelihood grows while the D statistic decays, consistently as π increases from 0 to 1. The growth of the likelihood is also paralleled with the improvement of the mean of the fitted model, which may be further compared to the corresponding values given in Tables 1 and 2. It is also seen that the statistic D is partial to assessing predictive performance of the fitted model, while descriptive performance in the sense of capturing characteristics of the marginal distribution is better reflected by the conditional log-likelihood. These results advocate the possibility of even optimizing the compromise between predictive and descriptive abilities of the fitted model, by choosing the weight π appropriately. Of course that choice generally depends on the scope of the optimization, and is not pursued any further in this work. However, some interesting possibilities might be to relate π with α or m.

Table 4 shows the procedure of selecting a suitable value of m by AIC or BIC in each of the six cases presented, when the thinning parameter α is estimated strictly by the method of MOM ($\pi = 0$). The values of m that optimize the AIC criterion (marked with *) were used in fitting the model for all the values of π shown in the previous Table 3. It is worthy noting that after a certain value of m the changes in the likelihood are quite small juxtaposed to the great complexity that the additional components induce to the model and its estimation. It is also noted that D fluctuates very little in Table 4, as m increases. This stems from the fact that α (and thus the ACF) of the MOM fitted model remains fixed for all m, and thus the predictive ability of the MOM model is not very sensitive to the number of components m. However, D does grow with m if the CML or a combined MOM/CML version of the fitted model is used (not shown here). It is also seen in Table 4 that the greater values of m correspond to cases with greater overdispersion (see also Table 1). All the models were fitted implementing the EM algorithm for several different initial values in order to ensure that the global (instead of local) maxima have been obtained. Figures 1 and 2 facilitate more detailed comparison between marginal probabilistic properties of the observations and those of the INAR(1) model, fitted by the MOM ($\pi = 0$, left column) and by the CML ($\pi = 0$,

right column) options, in each of the six cases.

Figure 1 depicts tail probabilities $\{P(X > k); k = 0, 1, 2, \ldots\}$ of the empirical probability distribution of the data, along with parametric bootstrap medians $Q(0.5)$ and 95% percentile confidence intervals $[Q(0.025), Q(0.975)]$ of tail probabilities based on 1000 simulations from the fitted model. In each simulation a series of length $M = 1000$ was generated by the fitted model, from which only the tail segment of length $N = 184$ (equal to the length of the observed series) was retained. Experimentation with other values of M showed that $M = 1000$ suffices to reach stationarity in the tail segment of the desired length.

Figure 2 depicts sample moments of order $0.1 \leq k \leq 4$ (with incremental step $\Delta k = 0.1$), calculated from the observed series of pixel counts, along with parametric bootstrap medians $Q(0.5)$ and 95% percentile confidence intervals $[Q(0, 025), Q(0.975)]$ of moments based on the same 1000 simulations used for Figure 1. Parametric bootstrap percentile estimates of marginal characteristics of the fitted model are implemented as an alternative to unavailable exact formulae for tail probabilities and (general) moments of the proposed model. However, exact values of (non-central) moments of integer order $k = 1, 2, 3, 4$ have been calculated explicitly in terms of the parameters of the fitted model, and are shown in Figure 2. Note that all moments depicted in Figure 2 are represented in logarithmic scale, merely for the sake of enhancing clarity. Figures 1 and 2 reassure the already acknowledged better ability of CML versus MOM options of the fitted model to describe the empirical probability distribution of the observed time series of counts.

Figures 3-5 correspond to three of the six demonstrative cases considered above. Each depicts the corresponding time series of pixel counts (continuous line in plots (c)-(f)), along with 1-step-ahead (plots (c) and (d)) and 2-step-ahead (plots (e) and (f)) integer-valued predictions (dotted lines in plots (c)-(e)). Specifically, predictions shown in (c) and (e) are based on a single replication from the MOM fitted model, which for predictive purposes is certainly more adequate than the CML option. Predictions shown in (d) and (f) are actually parametric bootstrap medians of integer-valued predictions based on 1000 replications from the MOM fitted model, laying within 95% parametric bootstrap percentile prediction intervals shown by grey vertical lines. All predictions follow the observed data with sufficient proximity so as to represent quite adequately the overall dynamic evolution of the observed series. Notably, the observed counts of pixels lay also well inside the parametric bootstrap prediction intervals, except in a few instants corresponding to extreme variability from low to high values and vice versa. It is also worthy to point out that the skewness of the observed counts is very adequately represented by the (integer-valued) parametric bootstrap percentiles shown in plots (d) and (f) of Figures 3-5. That is, in Figures 3-4 where the data are markedly positively skewed (see Table 1), bootstrap medians $Q(0.5)$ are much closer to the lower ends $Q(0.025)$ of the 95% percentile interval $[Q(0.025), Q(0.975)]$. Yet, in Figure 5 where the data are negatively skewed (see Table 1), parametric bootstrap medians $Q(0.5)$ are more centrally located within the 95% bootstrap percentile intervals $[Q(0.025), Q(0.975)]$, and indeed closer to the upper ends $Q(0.025)$ during segments where all 64 pixels of the $16Km$ region are recorded as being "wet" (an artifact due to deficiencies pointed out earlier).

Plots (a) and (b) in each of Figures 3-5 depict sample ACF and smoothed periodogram estimates of power spectra (dotted lines), along with the MOM-fitted model's exact ACF and exact spectral density functions (continuous lines), according to formulae (8) and (9) respectively. Percentile 95% confidence limits via parametric bootstrap based on 1000 simulations of the MOM fitted model are also provided for the ACF of the fitted model. It is seen that deviations between sample and model ACF are greater at large lags, while there is generally good agreement for small lags. An equivalent representation of this behavior is seen by comparing the fitted model's exact spectra against sample spectra. There is a clear deficit of exact model spectra versus sample spectra in the band of low frequencies near the origin, which reduces appreciably and eventually diminishes in the high frequency band.

6 Concluding remarks

A general remark about integer-valued time series models, is that statistical inference for their parameters, integer-valued or "coherent" prediction, and assessment of performance are quite challenging tasks, both conceptually and computationally, perhaps more so than in linear and non-linear time series models taking values in a continuum. The additional difficulties stem from complexities inherent to non-linear dependence structure of such models, in the present case imposed via binomial thinning, as a means of retaining integer-valued realizations of the model. The degree by which the integer-valued character of the model complicates very basic concepts, even such as that of residuals upon which model adequacy might be tested, is well conveyed by *Freeland and McCabe* [2004a] in the case of Poisson INAR(1). The model proposed and studied in this article certainly perplexes all these difficulties, although it is perhaps the simplest possible way of generalizing the Poisson INAR(1) model towards modelling time series of highly overdispersed count data.

It is worthy to remark that implementation of parametric bootstrap medians for construction of integer-valued predictions, and of other percentiles for construction of integer-valued 95% prediction intervals (see again plots (d) and (f) in Figures 3-5), is an idea similar to that of "coherent forecasting" introduced by *Freeland and McCabe* [2004b]. Coherent forecasting is a method producing integer-valued predictions of minimum mean absolute error (MAE), conditionally on past observations; see also *McCabe and Martin* [2005] and *Jung and Tremayne* [2006]. Although the approach taken herein towards integer-valued predictions is not "coherent forecasting" per se, parametric bootstrap medians of k-step-ahead integer-valued predictions produced according to (23) from the MOM fitted model, for $k = 1, 2$ are indeed integer-valued too, and do minimize MAE conditionally on past values of the fitted model (instead of conditionally on past observations).

The herein studied version of INAR(1) has met exceptionally well the marginal statistical properties of the highly overdispersed pixel count time series to which it was fitted by CML. However, the same model fitted to these data, even by the MOM option so as to optimize predictive performance, seems to fall short of capturing fully the complicated dependence structure of the observed process. This is evident by the earlier noted deviations of the MOM fitted model's ACF and spectrum from their sample estimates. These deviations might be interpreted in various ways, pointing perhaps

to non-exponential ACF (e.g. ACF might be thought of as a linear combination of exponentially decaying functions of lag, or even as power-law decays), and possibly even to non-Markovian dependence structure. This prompts for consideration of alternative versions of the proposed INAR(1) model. Different possibilities in that direction might be to introduce dependence among innovations, to make mixed Poisson innovations dependent on past observations, and even to randomize the binomial thinning parameter α so as to account for possibly varying "survival" probabilities among pixels. A valuable tool in exploring and deciding such alterations might be the so called "information matrix" test (IM), if appropriately ramified so as to serve the case of INAR(1) with mixed Poisson innovations. A version of the IM test and its three components for INAR(1) with simple Poisson innovations is detailed in *Freeland and McCabe* [2004a].

Another interesting feature of the pixel count time series modelled here is a set of power-law type multi-scaling properties of sample moments, with respect to the regional scale (for fixed threshold) and with respect to the threshold level (for fixed scale). These properties might be interpreted as indications of multifractal spatial distribution of the temporally evolving rain rate field over the probed region. Although a detailed presentation of these properties is beyond the scope of the present article, it is worthy to mention that they are very well reproduced by the parametric bootstrap estimates of moments from the proposed INAR(1) model fitted to the corresponding series by CML. Similar multi-scaling properties reported in research literature about spatial moments of rain rate [e.g. see *Gupta and Waymire*, 1990, 1993], may in light of the TM be viewed as consequences of multi-scaling of τ-FRC, which in practice is essentially conveyed by multi-scaling behavior of counts of pixels where SARR exceeds τ. Therefore, parametric modelling of time series of counts of pixels where SARR exceeds a given (optimal) threshold τ is further justified by the need of stochastic mechanisms able to reproduce such multi-scaling properties.

On an even more general note, the earlier discussed idea of recasting (instantaneous) predictions of *tau*-FRC, based on pixel counts, to (instantaneous) predictions of spatial moments via TM, can in principle be also applied for prediction of (instantaneous) moments of the *spatial cumulative distribution function* (SCDF) of a temporally evolving non-negative valued random field. That is, for a spatio-temporal random field with sufficient intermittency between zero and positive values, so that TM performs well for a broad band of thresholds, and under appropriate assumptions of temporal stationarity and spatial homogeneity, so that the corresponding SCDF remains stationary. In such a setting, predictions made via a parametric stationary model for time series of pixel counts could in principle be recast to temporal predictions of moments of the SCDF of the probed random field. Non-parametric methodology for prediction of the SCDF random functional (at fixed time), based on sub-sampling schemes, along with definitions of SCDF, its moments and quantiles, and an application to predicting SCDF of an ecological index for maple tree foliage, is given in *Lahiri et al.*[1999].

APPENDIX A: Properties of Binomial Thinning

In the following list of properties X and Y denote non-negative integer-valued and stochastically independent random variables, and the symbol $\stackrel{D}{=}$ denotes equality of probability distributions.

A1. The characteristic function of $\alpha \circ X$, is $\Phi_{\alpha\circ X}(u) = E\left(e^{iu\cdot(\alpha\circ X)}\right) = E\left\{\left(1-\alpha+\alpha e^{iu}\right)^X\right\}$, for $u \in \mathbb{R}$. For a proof, first condition on X, account for the fact that the conditioned random variable $(\alpha \circ X|X)$ follows the binomial distribution $Bin(X, \alpha)$, and then take expectation of the conditioned quantity. Direct consequences of **A1** are the following properties.

A2. $0 \circ X \stackrel{D}{=} 0 \quad \& \quad 1 \circ X \stackrel{D}{=} X$.

A3. $\alpha_1 \circ (\alpha_2 \circ X) \stackrel{D}{=} (\alpha_1\alpha_2) \circ X$, for every $\alpha_1, \alpha_2 \in [0, 1]$.

A4. $\alpha \circ (X + Y) \stackrel{D}{=} (\alpha\circ X) + (\alpha\circ Y)$, for every $\alpha \in [0, 1]$.

Non-central moments $\mu'_r(\alpha \circ X) = E\left[(\alpha \circ X)^r\right]$ of $\alpha \circ X$ are obtained directly from **A1**, provided that $\mu'_r(X) = E\left(X^r\right) < \infty$, for $r = 1, 2, 3, 4$ respectively:

A5. $\mu'_1(\alpha \circ X) = \alpha\mu'_1(X)$,

A6. $\mu'_2(\alpha \circ X) = \alpha^2\mu'_2(X) + \alpha(1-\alpha)\mu'_1(X)$,

A7. $\mu'_3(\alpha \circ X) = \alpha^3\mu'_3(X) + 3\alpha^2(1-\alpha)\mu'_2(X) + \alpha(1-3\alpha+2\alpha^2)\mu'_1(X)$,

A8. $\mu'_4(\alpha \circ X) = \alpha^4\mu'_4(X) + 6\alpha^3(1-\alpha)\mu'_3(X) + \alpha^2(1-\alpha)(7-11\alpha)\mu'_2(X) + \alpha(1-\alpha)(6\alpha^2 - 6\alpha + 1)\mu'_1(X)$.

Central moments $\mu_r(\alpha\circ X) = E\left[(\alpha \circ X - E(\alpha \circ X))^r\right]$, for $r = 1, 2, 3, 4$, are obtained by elementary algebraic calculations, exploiting the previous formulae for non-central moments:

A9. $\mu_2(\alpha \circ X) = Var\left(\alpha \circ X\right) = \alpha^2 Var\left(X\right) + \alpha(1-\alpha)E(X)$,

A10. $\mu_3(\alpha \circ X) = \alpha^3\mu_3(X) + 3\alpha^2(1-\alpha)Var(X) + \alpha(1-3\alpha+2\alpha^2)E(X)$,

A11. $\mu_4(\alpha \circ X) = \alpha^4\mu_4(X) + 6\alpha^3(1-\alpha)E(X^3) + \alpha^2(1-\alpha)(7-11\alpha)E(X^2) + \alpha(1-\alpha)(6\alpha^2 - 6\alpha + 1)E(X)$
$-12\alpha^3(1-\alpha)E(X)E(X^2) - 4\alpha^2(1-3\alpha+2\alpha^2)[E(X)]^2 + 6\alpha^3(1-\alpha)[E(X)]^3$.

APPENDIX B: Central Moments of Stationary INAR(1) Processes

While the central moments of first and second order are given by (5), the central moments of third and fourth order of a stationary INAR process can also be derived directly from (1). First recall that central moments of third and fourth order of $X + Y$ when X and Y are stochastically independent, are given by $\mu_3(X + Y) = \mu_3(X) + \mu_3(Y)$ and $\mu_4(X + Y) = \mu_4(X) + \mu_4(Y) + 6Var(X)Var(Y)$. Applying these formulae to the RHS of (1), and accounting for stationarity so that $\mu_3(X_{t+1}) = \mu_3(X_t)$ and $\mu_4(X_{t+1}) = \mu_4(X_t)$, one obtains:

B1. $(1-\alpha^3)\mu_3(X) = \mu_3(\varepsilon) + \mu_3(\alpha \circ X)$,

B2. $(1-\alpha^4)\mu_4(X) = \mu_4(\varepsilon) + \mu_4(\alpha \circ X) + 6\sigma_\varepsilon^2\left[\alpha^2\sigma_X^2 + \alpha(1-\alpha)\mu_X\right]$,

whereupon a further calculation is possible by employing **A10** and **A11**, respectively. For example, substituting $\mu_3(\alpha \circ X)$ according to **A10** in the RHS of **B1**, and then substituting $E(X) = \mu_X$ and $Var(X) = \sigma_X^2$ according to (5), after some algebraic calculations one obtains

B3. $\mu_3(X) = \left[\frac{3\alpha^3}{1+\alpha} + \alpha(1-2\alpha)\right]\mu_\varepsilon + \left[\frac{3\alpha^2}{1+\alpha}\right]\sigma_\varepsilon^2 + \mu_3(\varepsilon)$.

- **Remark 1:** When the innovations follow a Poisson Law $Poi(\lambda)$, then $\mu_\varepsilon = \sigma_\varepsilon^2 = \mu_3(\varepsilon) = \lambda$, and $\mu_3(X)$ in **B3** simplifies to the third central moment of the Poisson law $Poi(\lambda/(1-\alpha))$, as expected.

- **Remark 2:** If the law of the innovations is positively skewed (i.e. $\mu_3(\varepsilon) > 0$), then the stationary INAR(1) process shall be positively skewed too, since the quantities in brackets in **B3** are all positive.

Non-central moments can be derived via standard formulae relating central moments and non-central moments (e.g. see *Kendall et al.* [1994], Chapter 3).

APPENDIX C: Properties of Mixed Poisson Laws

Verification of convergence condition (3).

For given integer $j \geq 1$, $P(\varepsilon_t \geq j) = \sum_{x=j}^{\infty} P(\varepsilon_t = x) = \sum_{x=j}^{\infty} \left(\sum_{i=1}^{m} p_i e^{-\lambda_i} \frac{\lambda_i^x}{x!} \right) = \sum_{i=1}^{m} p_i e^{-\lambda_i} \left(\sum_{x=j}^{\infty} \frac{\lambda_i^x}{x!} \right)$, and by rewriting the last inner sum as $\sum_{k=0}^{\infty} \frac{\lambda_i^{j+k}}{(j+k)!} = \frac{\lambda_i^j}{(j-1)!} \cdot \sum_{k=0}^{\infty} \frac{\lambda_i^k}{j(j+1)\cdots(j+k)}$, and then dividing by j, we obtain

$$\frac{P(\varepsilon_t \geq j)}{j} = \sum_{i=1}^{m} p_i e^{-\lambda_i} \frac{\lambda_i^j}{j!} \left(\sum_{k=0}^{\infty} \frac{\lambda_i^k}{j(j+1)\cdots(j+k)} \right) \leq \sum_{i=1}^{m} p_i e^{-\lambda_i} \frac{\lambda_i^j}{j!} \left(\sum_{k=0}^{\infty} \frac{\lambda_i^k}{k!} \right) = \sum_{i=1}^{m} p_i \frac{\lambda_i^j}{j!}.$$

Thus,

$$\sum_{j=1}^{\infty} \frac{P(\varepsilon_t \geq j)}{j} \leq \sum_{j=1}^{\infty} \left(\sum_{i=1}^{m} p_i \frac{\lambda_i^j}{j!} \right) = \sum_{i=1}^{m} p_i \left(\sum_{j=1}^{\infty} \frac{\lambda_i^j}{j!} \right) = \sum_{i=1}^{m} p_i \left(e^{\lambda_i} - 1 \right) < \infty.$$

Formulae for moments.

Let Λ be a discrete random variable assuming only m positive values, $\lambda_1, \ldots, \lambda_m$ with probabilities $p_1, \ldots, p_(m)$, respectively. Then the non-central moments $\mu_r'(\varepsilon) = E[\varepsilon_t^r]$ of order $r = 1, 2, 3, 4$ for the mixture of the m Poisson laws $Poi(\lambda_i)$, $i = 1, \ldots, m$ are given by the formulae:

C1. $E(\varepsilon) = \mu_\varepsilon = E(\Lambda)$,

C2. $E(\varepsilon^2) = E(\Lambda^2) + E(\Lambda)$,

C3. $E(\varepsilon^3) = E(\Lambda^3) + 3E(\Lambda^2) + E(\Lambda)$,

C4. $E(\varepsilon^4) = E(\Lambda^4) + 6E(\Lambda^3) + 7E(\Lambda^2) + E(\Lambda)$,

The central moments $\mu_r(\varepsilon) = E[(\varepsilon_t - E(\varepsilon_t))^r]$, for $r = 2, 3, 4$ are given by the formulae:

C5. $\mu_2(\varepsilon) = Var(\varepsilon) = \sigma_\varepsilon^2 = E(\varepsilon^2) - [E(\varepsilon)]^2 = E(\Lambda^2) + E(\Lambda) - [E(\Lambda)]^2 = E(\Lambda) + Var(\Lambda)$,

C6. $\mu_3(\varepsilon) = \mu_3(\Lambda) + 3Var(\Lambda) + E(\Lambda)$,

C7. $\mu_4(\varepsilon) = \mu_4(\Lambda) + 4Var(\Lambda) + 3E(\Lambda^2) + E(\Lambda) + 6(E(\Lambda^3) + [E(\Lambda)]^3 - 2E(\Lambda^2)E(\Lambda))$.

C1-C4 can be calculated directly from (12) by the derivatives $\mu_r'(\varepsilon) = [d^r \psi(e^u)/du^r]_{u=0}$. However, formulae **C1-C7** remain valid even for general mixtures (i.e. not necessarily finite mixtures) of Poisson laws (e.g. see *Johnson et al.* [1993]).

Regarding **Remark 2** in **Appendix B**, it is noted that when $\mu_3(\Lambda) + 3Var(\Lambda) + E(\Lambda) > 0$ the innovations are positively skewed, and then also the corresponding stationary INAR(1) distribution

is positively skewed too. Also note that **C1-C2** yield $ID(\varepsilon) = 1 + Var(\Lambda)/E(\Lambda) > 1$. Thus, according to (6), any stationary INAR(1) process driven by innovations with a finite mixture distribution of Poisson laws is overdispersed relative to Poisson laws.

References

[1] Al-Osh, M.A. and E.A.A. Aly (1992): First-order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics - Theory and Methods,* **21**, 2483-2492.

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

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

[4] Alzaid, A.A. and M.A. Al-Osh (1990): An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, **27**, 314-324.

[5] Bohning, D. (2000), *Computer assisted analysis of mixtures & applications in meta- analysis, disease mapping & others*, CRC Press, NY

[6] Cox, D. (1981): Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics*, **8**, 93-115.

[7] DaSilva, M.E. and V.L. Oliveira (2004): Difference equations for the higher order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, **25**(3), 317-333.

[8] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B*, **39**, 1-38

[9] Dion, J.P., G. Gauthier and A. Latour (1995): Branching processes with immigration and integer-valued time series. *Serdica*, **21**, 123-136.

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

[11] Franke, J. and T. Selingmann (1993): Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series Analysis*, Subba Rao, T. (ed.), Chapman & Hall, pp. 310-330.

[12] Freeland, R.K. and B.P.M. McCabe (2004a): Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis*, **25**(5), 701-722.

[13] Freeland, R.K. and B.P.M. McCabe (2004b): Forecasting discrete valued low count time series. *International Journal of Forecasting*, **20**, 427-434.

[14] Freeland, R.K. and B.P.M. McCabe (2005): Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics and Probability Letters*, **73**(2), 147-153.

[15] Gupta, V. K. and E.C. Waymire (1990): Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, **95**, 1999-2009.

[16] Gupta, V.K. and E.C. Waymire (1993): A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology*, **32**, 251-267.

[17] Grunwald, G.K., R.J. Hyndman, L. Tedesco and R.L. Tweedie (2000): Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics*, **42**, 479-495.

[18] Heathcote, C.R. (1966): A branching process allowing immigration. *Journal of the Royal Statistical Society, B*, **28**, 213-217.

[19] Johnson, N.L., S. Kotz and A.W. Kemp (1993): *Univariate Discrete Distributions, 2nd ed.*, John Wiley & Sons.

[20] Jung, R.C., G. Ronning and A.R. Tremayne (2005): Estimation in conditional first order autoregression with discrete support. *Statistical Papers*, **46**, 195-224.

[21] Jung, R.C. and A.R. Tremayne (2006): Coherent forecasting in integer time series models. *International Journal of Forecasting* (in press).

[22] Kayano, K. and K. Shimizu (1994): Optimal thresholds for a mixture of lognormal distributions as the continuous part of the mixed distribution. *Journal of Applied Meteorology*, **33** (12), 1543-1550.

[23] Kedem, B. and K. Fokianos (2002): *Regression Models for Time Series Analysis*, John Wiley & Sons.

[24] Kedem, B. and H. Pavlopoulos (1991): On the threshold method for rainfall estimation: Choosing the optimal threshold level. *Journal of the American Statistical Association*, **86**, 626-633.

[25] Kendall, M., Stuart, A. and Ord, J.K. (1994): *Kendall's advanced theory of statistics, Vol I,* 6th edition, Arnold Publishing, London

[26] Lahiri, S.N., M.S. Kaiser, N. Cressie and N.J. Hsu (1999): Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, **94**, 86-97.

[27] Lindsay, B.G. (1995). Mixture Models: Theory, Geometry and Applications. *NSF-CBMS Regional Conference Series in Probability and Statistics* ***5***, Institute of Mathematical Statistics and American Statistical Association.

[28] MacDonald, I.L. and W. Zucchini (1997): *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman & Hall.

[29] Mase, S. (1996): The threshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics*, **48** (2), 201-213.

[30] McCabe, B.P.M. and G.M. Martin (2005): Bayesian predictions of low count time series. *International Journal of Forecasting*, **21**, 315-330.

[31] McKenzie, E. (1985): Some simple models for discrete variate time series. *Water Resources Bulletin*, **21**, 645-650.

[32] McKenzie, E. (2003): Discrete variate time series. In *Handbook of Statistics, Vol. 21*, Shanbhag, D. and Rao,C. (eds.), Elsevier Science.

[33] McLachlan, G. and Krishnan, N. (1997): *The EM Algorithm and Extensions*. Wiley, NY.

[34] McLachlan, G. and Peel, N. (2001): *Finite mixture models*. Wiley NY.

[35] Meneghini, R., J.A. Jones, T. Iguch, K. Okamoto and J. Kwiatkowski (2001): Statistical methods of estimating average rainfall over large space-time scales using data from the TRMM precipitation radar. *Journal of Applied Meteorology*, **40** (3), 568-585.

[36] Priestley, M.B. (1981): *Spectral Analysis and Time Series*, Academic Press.

[37] Shimizu, K. and Kayano, K. (1994): A further study of the estimation of area rain rate moments by the threshold method. *Journal of the Meteorological Society of Japan*, **72**(6), 833-839.

[38] Shimizu, K., Short, D.A. and Kedem, B. (1993): Single- and double-threshold methods for estimating the variance of area rain rate. *Journal of the Meteorological Society of Japan*, **71**(6), 673-683.

[39] Short, D.A., P.A.Kucera, B.S. Ferrier, J.C. Gerlach, S.A. Rutledge and O.W. Thiele (1997): Shipboard radar rainfall patterns within the TOGA/COARE IFA. *Bulletin of the American Meteorological Society*, **78**, 2817-2836.

[40] Short, D.A., K. Shimizu and B. Kedem (1993): Optimal thresholds for the estimation of area rain-rate moments by the threshold method. *Journal of Applied Meteorology*, **32** (2), 182-192.

[41] Shumway, R. and J. Gurland (1960): Fitting the Poisson-binomial distribution. *Biometrics*, **16**, 522-534.

[42] Steutel, W. and K. Van Harn (1979): Discrete analogues of self-decomposability and stability. *Annals of Probability*, **7**(5), 893-899.

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

[44] Titterington, M., Smith, A.F.M, and Makov, U. (1985): *Statistical Analysis of finite mixture distributions*, Wiley NY.

[45] Tsay, R.S. (1992): Model checking via parametric bootstraps in time series analysis. *Applied Statistics*, 41, 1-15.

[46] Van Harn, K. (1978): *Classifying Infinitely Divisible Distributions by Functional Equations.* Amsterdam: Mathematisch Centrum.

threshold	mean	ID	IS	IK	$P(X=0)$	acf	mean	ID	IS	IK	$P(X=0)$	acf
			Scale = 240 *Km*						Scale = 120 *Km*			
0 *mm/hr*	3903.67	974.84	0.232	1.905	0.000	0.823	1266.17	324.62	0.217	1.992	0.000	0.807
1 *mm/hr*	810.36	706.90	1.414	4.333	0.000	0.810	149.40	255.80	1.862	5.620	0.060	0.847
2 *mm/hr*	543.41	547.05	1.526	4.602	0.000	0.791	102.51	182.10	1.769	5.114	0.103	0.844
5 *mm/hr*	253.28	246.55	1.351	3.763	0.005	0.756	**53.72**	**112.59**	**1.973**	**5.856**	**0.201**	**0.823**
10 *mm/hr*	114.16	129.28	1.689	5.029	0.011	0.758	28.03	72.04	2.183	6.718	0.255	0.819
20 *mm/hr*	37.85	59.73	1.941	6.021	0.092	0.745	10.12	35.38	2.797	10.838	0.413	0.782
			Scale = 60 *Km*						Scale = 32 *Km*			
0 *mm/hr*	383.95	93.40	0.477	2.167	0.000	0.823	119.13	25.94	-0.053	1.989	0.000	0.824
1 *mm/hr*	**25.75**	**63.24**	**2.064**	**6.871**	**0.391**	**0.760**	3.40	17.36	3.387	17.068	0.685	0.686
2 *mm/hr*	17.51	47.96	2.351	8.688	0.457	0.727	2.33	15.95	3.975	21.904	0.734	0.670
5 *mm/hr*	9.74	39.16	3.110	13.583	0.527	0.687	1.27	11.59	4.285	24.800	0.837	0.632
10 *mm/hr*	5.69	29.36	3.347	15.536	0.625	0.660	**0.71**	**8.93**	**5.167**	**35.373**	**0.853**	**0.589**
20 *mm/hr*	**2.35**	**17.63**	**3.822**	**18.527**	**0.739**	**0.583**	0.29	5.67	5.515	37.086	0.924	0.662
			Scale = 16 *Km*						Scale = 8 *Km*			
0 *mm/hr*	**39.207**	**10.019**	**-0.316**	**1.812**	**0.033**	**0.762**	10.277	3.511	-0.558	1.727	0.109	0.694
1 *mm/hr*	0.946	9.855	4.507	28.257	0.864	0.536	0.185	3.540	5.448	36.841	0.929	0.231
2 *mm/hr*	0.658	9.385	5.489	39.578	0.891	0.521	**0.098**	**2.918**	**6.930**	**55.948**	**0.951**	**0.272**
5 *mm/hr*	0.359	7.439	6.912	62.410	0.918	0.189	0.043	3.224	9.179	89.142	0.984	-0.014
10 *mm/hr*	0.196	7.344	9.201	99.346	0.940	0.186	0.022	2.492	11.790	147.417	0.989	-0.009
20 *mm/hr*	0.087	5.568	9.367	98.008	0.978	0.345	0.005	1.000	13.344	180.032	0.995	-0.005
			Scale = 4 *Km*						Scale = 2 *Km*			
0 *mm/hr*	2.690	1.100	-0.728	1.706	0.245	0.620	0.696	0.306	-0.843	1.704	0.304	0.561
1 *mm/hr*	0.033	1.643	7.465	60.050	0.978	-0.020	0	-	-	-	1	-
2 *mm/hr*	0.011	2.000	13.344	180.032	0.995	-0.005	0	-	-	-	1	-
5 *mm/hr*	0.011	2.000	13.344	180.032	0.995	-0.005	0	-	-	-	1	-
10 *mm/hr*	0.005	1.000	13.344	180.032	0.995	-0.005	0	-	-	-	1	-
20 *mm/hr*	0	-	-	-	1	-	0	-	-	-	1	-

Table 1: Sample statistics for all 48 series of pixel counts. No statistics are reported for series with only zero values. Boldface characters mark the six representative cases chosen for demonstration of performance by the fitted model.

	scale (Km)	120	60	60	32	16	8
	threshold (mm/hr)	5	1	20	10	0	2
CML	mean	54.026	25.891	2.376	0.730	39.206	0.098
MOM		90.906	41.792	3.464	1.055	45.837	0.114
CML	ID	110.441	62.757	14.963	7.261	6.643	2.778
MOM		36.031	28.085	9.625	5.802	5.757	2.545
CML	IS	1.99	2.08	3.72	5.63	-0.13	7.02
MOM		0.82	0.92	2.38	3.51	0.22	6.36
CML	IK	5.97	6.92	17.52	48.06	2.51	61.86
MOM		8.61	7.18	14.12	31.6	4.76	54.4
CML	$P(X=0)$	0.168	0.388	0.663	0.798	0	0.95
MOM		<0.0001	0.002	0.385	0.666	<0.0001	0.937
CML	m	9	9	3	4	4	2
MOM		7	7	3	4	3	2
CML	α	0.00398	6E-06	0.112	0.2499	0.376	0.129
MOM		0.82	0.76	0.58	0.59	0.76	0.27
CML	Log-L	-841.51	-646.99	-249.78	-129.9	-732.33	-44.617
MOM		-2269.2	-1189.2	-317.38	-140.98	-918.91	-45.477
CML	D	11412951	6888093	12808	2241.29	730.835	10352.3
MOM		84.469	119.468	346.058	321.681	133.678	2283.193
CML	P-value	0	0	0	0	0	0
MOM		0.07	0.3	0.25	0.53	0.09	0.58

Table 2: Information on marginal characteristics and measures of performance for the CML and MOM fitted model.

	scale = 120 Km ($m = 7$) threshold = 5 mm/hr			scale = 60 Km ($m = 7$) threshold= 1 mm/hr			scale = 60 Km ($m = 3$) threshold= 20 mm/hr		
π	D	Log-L	mean	D	Log-L	mean	D	Log-L	mean
0 (MOM)	84.469	-2269.168	90.906	119.468	-1189.212	41.792	346.058	-317.376	3.464
0.1	105.789	-1833.274	73.213	142.711	-1035.532	35.695	410.828	-304.069	3.219
0.2	141.693	-1556.156	65.479	176.719	-935.932	32.274	498.900	-292.798	3.034
0.3	202.580	-1373.076	61.418	226.967	-846.721	30.228	621.493	-283.188	2.884
0.4	309.102	-1225.659	58.928	302.684	-795.957	28.887	797.284	-275.032	2.761
0.5	506.135	-1140.181	56.975	420.391	-746.946	28.111	1059.048	-268.182	2.662
0.6	904.999	-1046.806	56.028	611.123	-725.790	27.314	1468.290	-262.316	2.593
0.7	1837.491	-983.138	55.247	939.621	-707.709	26.839	2150.615	-257.187	2.529
0.8	4660.094	-929.664	54.707	1555.822	-670.400	26.522	3396.122	-253.192	2.465
0.9	19940.453	-887.964	54.351	2871.620	-661.300	26.187	5999.228	-250.668	2.413
1 (CML)	4201804	-857.056	54.028	6359.663	-659.549	26.034	12808.441	-249.783	2.376

	scale = 32 Km ($m = 4$) threshold = 10 mm/hr			scale = 16 Km ($m = 3$) threshold = 0 mm/hr			scale = 8 Km ($m = 2$) threshold = 2 mm/hr		
π	D	Log-L	mean	D	Log-L	mean	D	Log-L	mean
0 (MOM)	321.681	-140.976	1.055	133.678	-918.912	45.837	2283.193	-45.477	0.114
0.1	363.814	-138.328	0.982	146.954	-871.850	43.636	2543.711	-45.331	0.112
0.2	416.285	-136.197	0.922	163.631	-838.222	42.197	2852.752	-45.197	0.110
0.3	482.369	-134.509	0.873	184.546	-814.725	41.195	3222.958	-45.075	0.109
0.4	566.694	-133.198	0.835	211.181	-794.143	41.170	3671.345	-44.964	0.107
0.5	675.955	-132.195	0.806	244.211	-775.821	40.641	4221.251	-44.866	0.105
0.6	820.111	-131.426	0.784	286.088	-761.917	40.218	4905.358	-44.783	0.104
0.7	1014.472	-130.825	0.769	339.710	-751.521	39.909	5770.571	-44.714	0.102
0.8	1283.507	-130.353	0.756	409.142	-743.906	39.668	6886.108	-44.662	0.101
0.9	1668.203	-130.025	0.743	500.294	-739.035	39.432	8357.540	-44.629	0.100
1 (CML)	2241.288	-129.902	0.730	622.062	-737.359	39.207	10352.280	-44.617	0.098

Table 3: Log-likelihood and D statistics measuring performance of the fitted model for several values of the weight $0 \leq \pi \leq 1$ used between MOM ($\pi = 0$) and CML ($\pi = 1$) versions of the fitted INAR(1). The number of Poisson components in the mixture law of innovations, m, is chosen by optimization of the AIC as shown in Table 4, using MOM ($\pi = 0$). The mean of the fitted model is also reported.

	scale = 120 Km, threshold = 5 mm/hr				scale = 60 Km, threshold = 1 mm/hr			
m	Log-L	AIC	BIC	D	Log-L	AIC	BIC	D
1	-4730.399	9462.798	9463.063	84.017	-3082.490	6166.980	6167.245	118.994
2	-3020.503	6047.006	6047.800	84.500	-1596.888	3199.776	3200.570	119.317
3	-2591.699	5193.398	5194.722	84.260	-1276.478	2562.956	2564.280	119.405
4	-2335.357	4684.714	4686.568	84.366	-1232.281	2478.562	2480.416	119.454
5	-2289.661	4597.322	4599.705	84.487	-1203.214	2424.428	2426.811	119.439
6	-2277.539	4577.078	4579.991	84.456	-1192.989	2407.978	2410.891	119.453
7	-2269.168	4564.336*	4567.779**	84.469	-1189.212	2404.424*	2407.867 **	119.468
8	-2269.010	4568.020	4571.992	84.467	-1187.754	2405.508	2409.480	119.453
9	-2268.497	4570.994	4575.496	84.467	-1186.882	2407.764	2412.266	119.459
	scale = 60 Km, threshold = 20 mm/hr				scale = 32 Km, threshold = 10 mm/hr			
m	Log-L	AIC	BIC	D	Log-L	AIC	BIC	D
1	-687.762	1377.524	1377.789	344.831	-273.165	548.330	548.595	321.591
2	-367.229	740.458	741.253	345.691	-150.770	307.539	308.334	321.577
3	-317.376	644.752*	646.076**	346.058	-143.041	296.082	297.406**	321. 684
4	-316.703	647.407	649.261	346.050	-140.976	295.953*	297.807	321.681
5	-316.699	651.397	653.781	346.050	-140.976	299.953	302.336	321.681
	scale = 16 Km, threshold = 0 mm/hr				scale = 8 Km, threshold = 2 mm/hr			
m	Log-L	AIC	BIC	D	Log-L	AIC	BIC	D
1	-1377.723	2757.446	2757.711	134.842	-64.058	130.117	130.382	2282.450
2	-982.298	1970.595	1971.390	134.207	-45.477	96.954*	97.748**	2283.193
3	-918.912	1847.824*	1849.148**	133.678	-45.440	100.879	102.203	2283.241
4	-917.595	1849.189	1851.043	133.760	-	-	-	-

Table 4: Selection of the number of Poisson components in the mixture law of innovations, m, by optimization of AIC and BIC, using MOM ($\pi = 0$). Log-likelihood and D statistics, reflecting performance of the fitted INAR(1) model, are also reported. Optimal values of m are indicated by * under AIC and by ** under BIC.

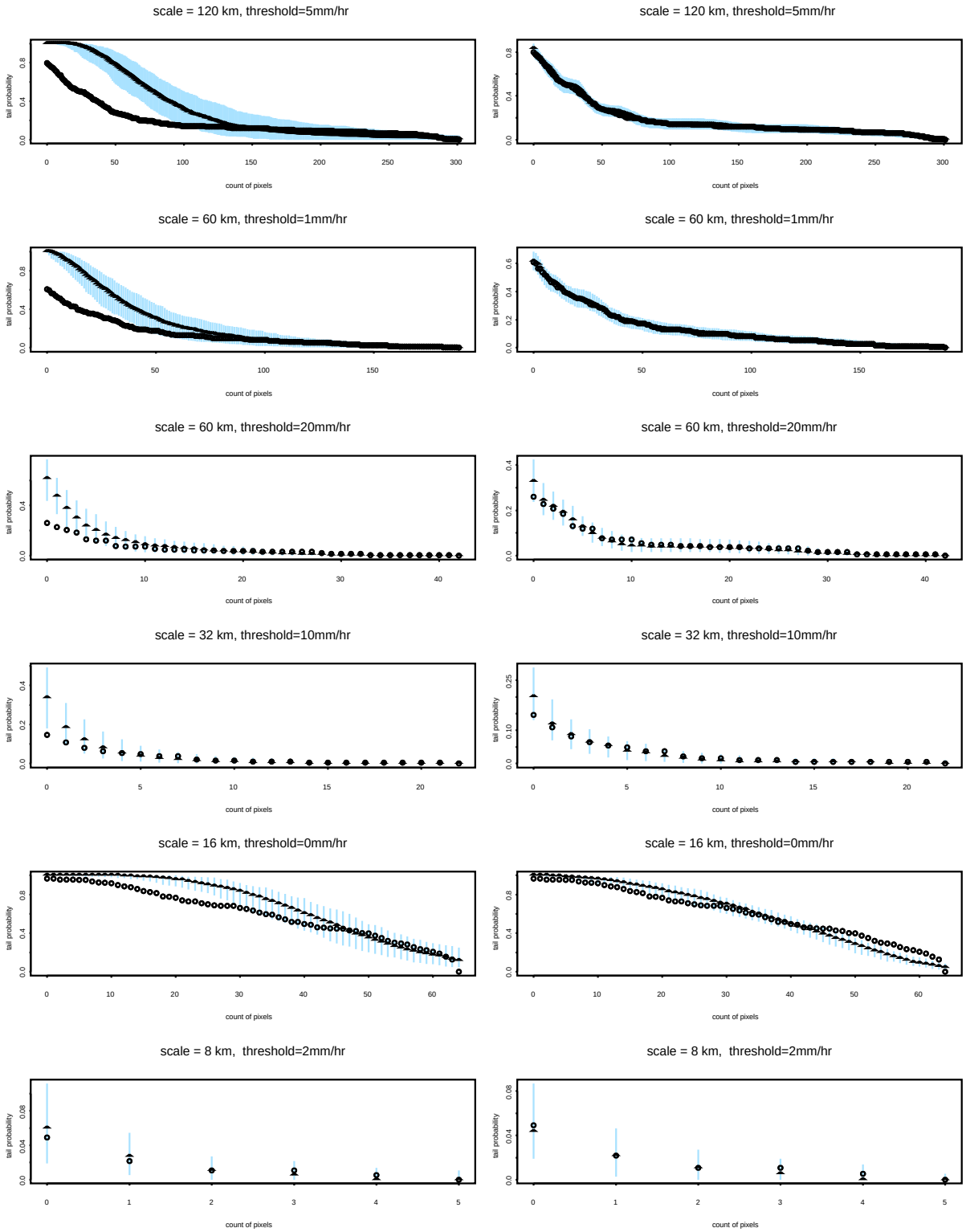


Figure 1: Plots of tail probabilities $P(X > k)$, for integer $k \geq 0$. Empirical values from the observed series of pixel counts are depicted with ∘. Parametric bootstrap medians of tail probabilities based on 1000 replications from the fitted INAR(1) model are depicted with ▲. Using parametric bootstrap percentiles, based on the same 1000 replications, 95% confidence intervals are also computed and depicted by grey vertical line segments. The left column of plots corresponds to the MOM version of the fitted model, and the right column to the CML version.

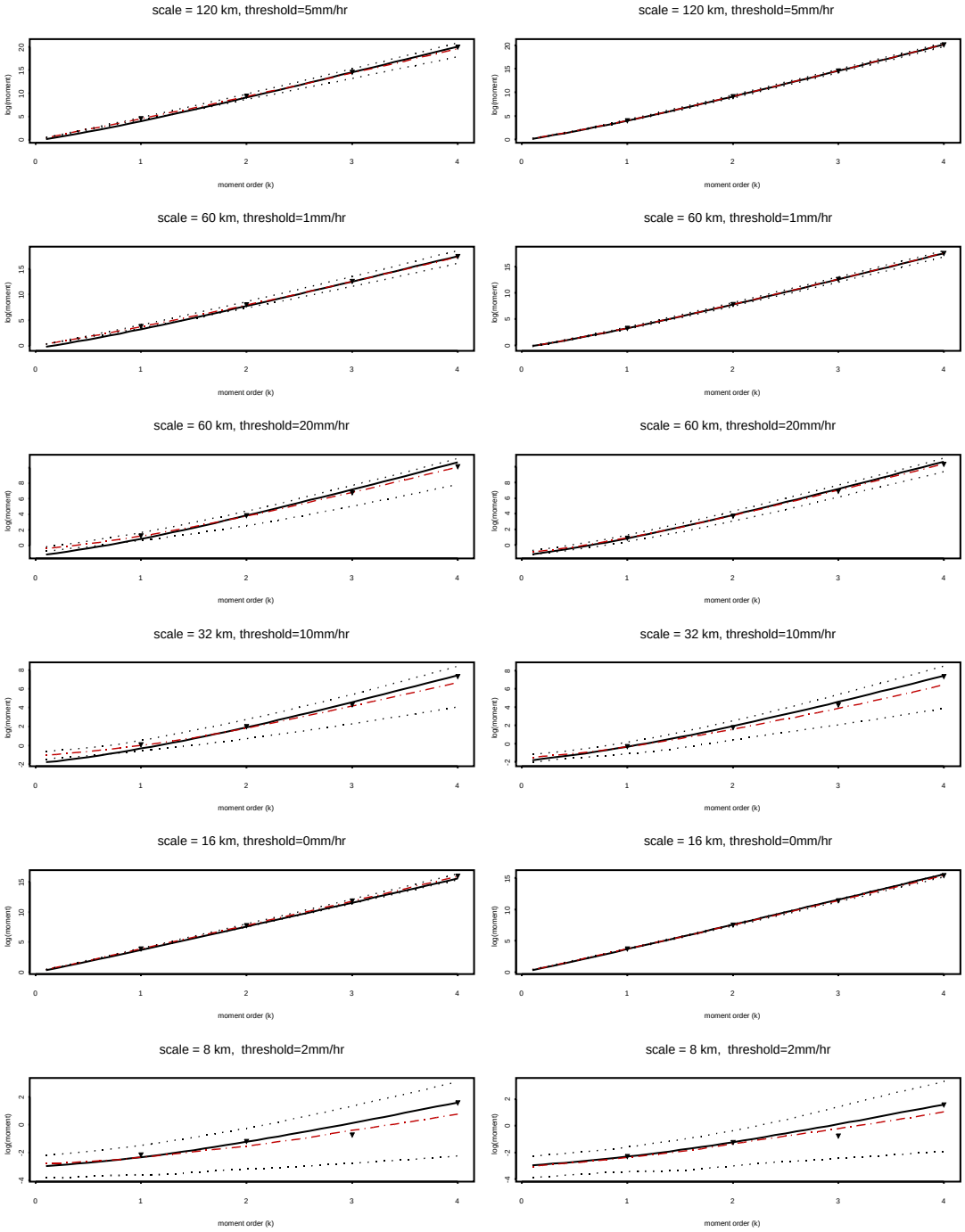


Figure 2: Plots of (non-central) log-moments, $log\{\mu'_k(X)\} = log\{E(X^k)\}$, for equi-spaced values of $0.1 \leq k \leq 4$ with step $\Delta k = 0.1$. The *continuous line* depicts sample moments from the observed series of pixel counts. Parametric bootstrap medians of moments based on 1000 replications from the fitted INAR(1) model are depicted by the *dashed line*. Using parametric bootstrap percentiles, based on the same 1000 replications, 95% confidence bounds are also computed and depicted by the two *dotted lines*. The four ▼ in each plot depict exact values of moments of integer order $k = 1, 2, 3, 4$ for the fitted model. The left column of plots corresponds to the MOM version of the fitted model, and the right column to the CML version.

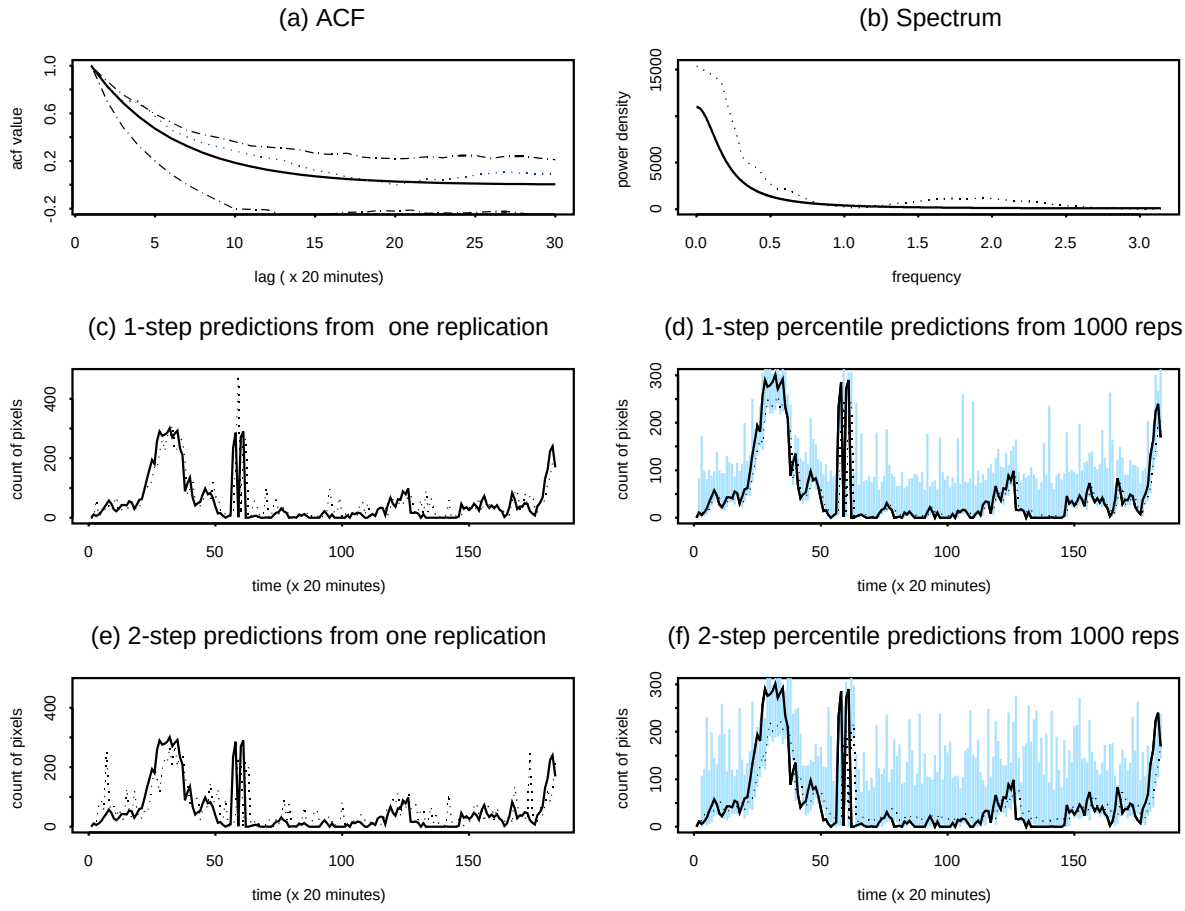


Figure 3: Case of threshold = 5 *mm/hr* in the region with scale = 120 *Km*. The corresponding time series of pixel counts is depicted by the continuous line in each of (c)-(f). Sample ACF and a smoothed periodogram estimate of the spectrum are depicted by *dotted lines* in (a) and (b). Exact ACF and exact spectral density function of the MOM-fitted model (see formulae (8) and (9)) are depicted by *continuous lines* in (a) and (b) respectively. The *dashed lines* in (a) depict 95% confidence bounds for the ACF, using parametric bootstrap percentiles based on 1000 replications of the MOM-fitted model. The *dotted lines* in (c) and (d) depict 1-step-ahead integer-valued predictions made by the MOM-fitted model. Predictions shown in (c) are based on a single replication, while predictions shown in (d) are parametric bootstrap medians of 1-step predictions from 1000 replications of the MOM-fitted model. The grey vertical line segments in (c) are 95% prediction intervals constructed from parametric bootstrap percentiles of the 1000 1-step predictions. The *dotted lines* and the grey intervals in (e) and (f) depict similar information as in (c) and (d), but for 2-step-ahead predictions based on a single and on 1000 replications of the MOM-fitted model.

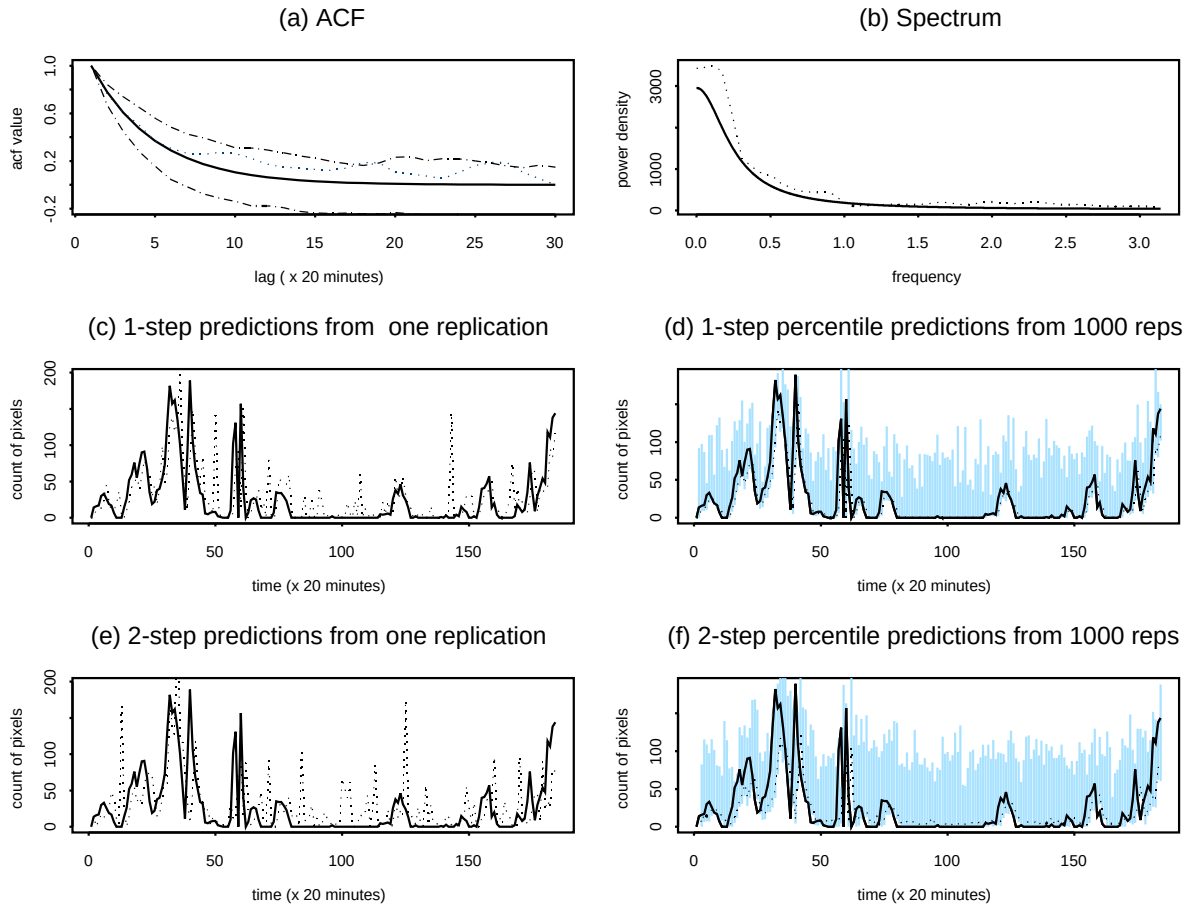


Figure 4: Case of threshold = 1 *mm/hr* in the region with scale = 60 *Km*. The corresponding time series of pixel counts is depicted by the continuous line in each of (c)-(f). Sample ACF and a smoothed periodogram estimate of the spectrum are depicted by *dotted lines* in (a) and (b). Exact ACF and exact spectral density function of the MOM-fitted model (see formulae (8) and (9)) are depicted by *continuous lines* in (a) and (b) respectively. The *dashed lines* in (a) depict 95% confidence bounds for the ACF, using parametric bootstrap percentiles based on 1000 replications of the MOM-fitted model. The *dotted lines* in (c) and (d) depict 1-step-ahead integer-valued predictions made by the MOM-fitted model. Predictions shown in (c) are based on a single replication, while predictions shown in (d) are parametric bootstrap medians of 1-step predictions from 1000 replications of the MOM-fitted model. The grey vertical line segments in (c) are 95% prediction intervals constructed from parametric bootstrap percentiles of the 1000 1-step predictions. The *dotted lines* and the grey intervals in (e) and (f) depict similar information as in (c) and (d), but for 2-step-ahead predictions based on a single and on 1000 replications of the MOM-fitted model.

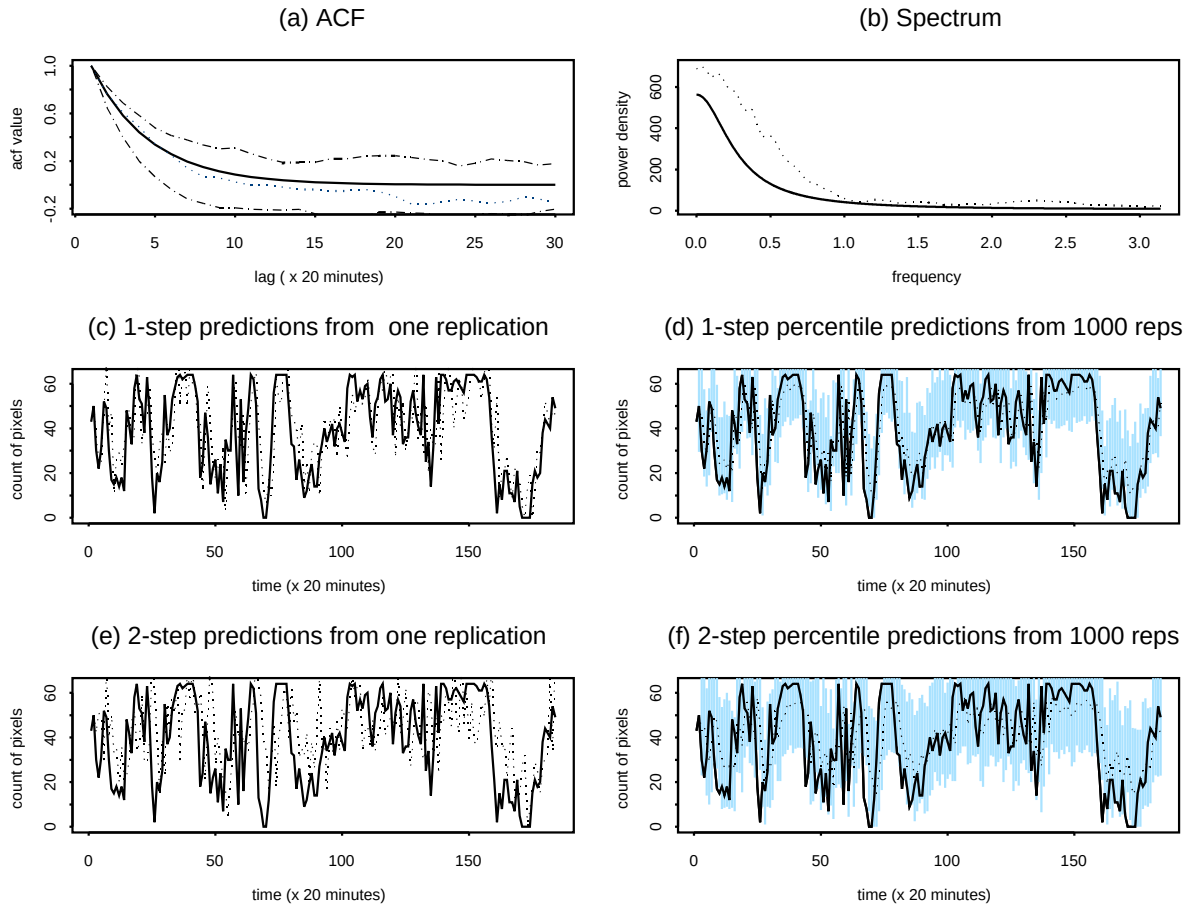


Figure 5: Case of threshold = 0 *mm/hr* in the region with scale = 16 *Km*. The corresponding time series of pixel counts is depicted by the continuous line in each of (c)-(f). Sample ACF and a smoothed periodogram estimate of the spectrum are depicted by *dotted lines* in (a) and (b). Exact ACF and exact spectral density function of the MOM-fitted model (see formulae (8) and (9)) are depicted by *continuous lines* in (a) and (b) respectively. The *dashed lines* in (a) depict 95% confidence bounds for the ACF, using parametric bootstrap percentiles based on 1000 replications of the MOM-fitted model. The *dotted lines* in (c) and (d) depict 1-step-ahead integer-valued predictions made by the MOM-fitted model. Predictions shown in (c) are based on a single replication, while predictions shown in (d) are parametric bootstrap medians of 1-step predictions from 1000 replications of the MOM-fitted model. The grey vertical line segments in (c) are 95% prediction intervals constructed from parametric bootstrap percentiles of the 1000 1-step predictions. The *dotted lines* and the grey intervals in (e) and (f) depict similar information as in (c) and (d), but for 2-step-ahead predictions based on a single and on 1000 replications of the MOM-fitted model.