MULTIVARIATE INTEGER-VALUED TIME SERIES WITH FLEXIBLE AUTOCOVARIANCES AND THEIR APPLICATION TO MAJOR HURRICANE COUNTS[1]

BY JAMES LIVSEY[*], ROBERT LUND[†,2], STEFANOS KECHAGIAS[‡]
AND VLADAS PIPIRAS[§,3]

*United States Census Bureau[*], Clemson University[†], SAS Institute[‡]
and University of North Carolina, Chapel Hill[§]*

This paper examines a bivariate count time series with some curious statistical features: Saffir–Simpson Category 3 and stronger annual hurricane counts in the North Atlantic and eastern Pacific Ocean Basins. As land and ocean temperatures on our planet warm, an intense climatological debate has arisen over whether hurricanes are becoming more numerous, or whether the strengths of the individual storms are increasing. Recent literature concludes that an increase in hurricane counts occurred in the Atlantic Basin circa 1994. This increase persisted through 2012; moreover, the 1994–2012 period was one of relative inactivity in the eastern Pacific Basin. When Atlantic activity eased in 2013, heavy activity in the eastern Pacific Basin commenced. When examined statistically, a Poisson white noise model for the annual severe hurricane counts is difficult to resoundingly reject. Yet, decadal cycles (longer term dependence) in the hurricane counts is plausible. This paper takes a statistical look at the issue, developing a stationary multivariate count time series model with Poisson marginal distributions and a flexible autocovariance structure. Our auto- and cross-correlations can be negative and have long-range dependence; features that most previous count models cannot achieve in tandem. Our model is new in the literature and is based on categorizing and super-positioning multivariate Gaussian time series. We derive the autocovariance function of the model and propose a method to estimate model parameters. In the end, we conclude that severe hurricane counts are indeed negatively correlated across the two ocean basins. Some evidence for long-range dependence is also presented; however, with only a 49-year record, this issue cannot be definitively judged without additional data.

1. Introduction. Hurricanes participate in equalizing global heat imbalances. In the Northern Hemisphere, hurricanes form in the tropics and move northward

to the Arctic, carrying equatorial heat to the pole in an attempt to equalize global surface temperatures. Geophysicists often view hurricanes, which require warm waters to form and thrive, as an Earth sweating mechanism. As surface and ocean temperatures of the Earth warm, more equatorial heat will seemingly need to be dissipated. Accordingly, many scientists believe that a warming Earth should experience increased hurricane activity, expressed by a higher frequency of hurricanes and/or stronger individual storms. Note that stronger individual storms do not necessarily mean a larger number of strong storms.

Scientific debate over increasing hurricane activity has been intense. The popular science book by Mooney (2007) narrates the scientific mudslinging and the stances taken by different "camps" on various issues, including linking hurricane changes to global warming. The debate was exacerbated by an increase in North Atlantic Basin hurricane activity circa the mid-1990s. At this time, Atlantic activity was concluded to have increased by many authors [Elsner, Jagger and Niu (2000), Elsner, Kossin and Jagger (2008), Goldenberg, Landsea and Mestas-Nunez (2001), Robbins et al. (2011)]. Some physicists [Goldenberg, Landsea and Mestas-Nunez (2001)] explained this increase as part of a natural multi-decadal cycle, whereby hurricane counts oscillate on decadal cycles. Their claim that Atlantic activity would return to normal levels was based largely on physical models; past data were not considered. Another camp, the empiricists, claimed that an era of increased hurricane activity is here to stay, is largely attributed to climate change, and is supported by the data record.

Around 2012, North Atlantic hurricane activity markedly decreased. However, at this time, activity in the eastern Pacific Basin dramatically increased. This was reflected, in particular, in the frequency of hurricanes. In 2015, the eastern Pacific Basin experienced ten severe hurricanes while the Atlantic Basin had just two. Henceforth, the term Pacific, North Pacific or Pacific Basin will be referencing the eastern North Pacific Basin. This on/off negative correlation pattern has been persistent since the mid-1960s, when reliable Atlantic and Pacific hurricane records commenced (this is the time at which satellite surveillance began). One objective of this paper is to investigate this negative dependence between the two annual basin counts. A long-range dependence cycle in the basin counts is also statistically investigated.

Forecasting annual hurricane counts is difficult. Most forecasts of the North Atlantic Basin's activity a year in advance have little predictive power. In fact, Atlantic Basin storm counts often pass Poisson white noise statistical tests, especially when only the strong storms are considered. This said, some forecasting power can be achieved with covariates such as El-Niño, North Atlantic oscillation (NAO), Northwest African rainfall, etc. at a few months lead time [Elsner and Jagger (2006), Gray (1984)]. Previous work has also used these covariates to show the different behavior of storms in basins across the globe. For example, Elsner and Kocher (2000) note a negative correlation between tropical cyclones over the North Atlantic and four other basins (western North Pacific, eastern North Pacific,

northern Indian Ocean, and Southern Hemisphere). The authors then link this phenomenon to NAO, showing that there is a significant positive correlation between NAO and the global tropical cyclone activity index used in their analysis. Confirmation of negative correlation between basins or longer memory cycles in the individual basins should aid annual storm count forecasting.

Poisson distributions are natural models for the annual severe hurricane counts due to their event-based interpretation. Indeed, many authors have used Poisson or Poisson-based models [McDonnell and Holbrook (2004), Mooley (1981), Parisi and Lund (2000), Solow (1989), Thompson and Guttorp (1986), Xiao, Kottas and Sansó (2015)] to describe hurricane counts. This said, Poisson dynamics are not perfect: some slight over-dispersion in the Pacific counts will be encountered. While Chu and Zhao (2004) and Villarini, Vecchi and Smith (2010) and others propose negative binomial marginals, which are over-dispersed, the amount of over-dispersion in our data is minimal, as Section 3 shows. As such, our work entails developing a bivariate stationary time series model with marginal Poisson distributions for the annual storm counts. Extensions to over-dispersed marginal count distributions will be addressed in our concluding discussion.

Count time series modeling is an active current area of statistical research [Davis et al. (2016)]. To describe the severe hurricane counts in both basins simultaneously, a bivariate count time series model with Poisson marginal distributions is needed—one that permits possible negative cross-correlations at lag zero between the series and nonzero correlations at decadal lags in each marginal series. Stationarity, the natural status quo model, should be posited until it can be reliably discounted—essentially, our null hypothesis is a non-changing hurricane climate. However, such a count time series model has proven difficult to devise so far. Section 4 remedies this issue.

The rest of the paper proceeds as follows. Section 2 presents a brief background on count time series models. Section 3 explores properties of the bivariate hurricane count series. The construction of the bivariate Poisson count model that allows for negative cross-correlations and long-range dependence is undertaken in Section 4. Section 5 introduces a quasi-maximum likelihood parameter estimation method for this model; its performance is investigated in a short simulation study in Section 6. Section 7 fits the proposed model to the hurricane data. Conclusions and future work are summarized in Section 8.

2. Time series background.

2.1. *Count time series*. Count time series arise in the investigation of natural phenomena such as rare disease occurrences, animal sightings, and severe weather events. This subsection reviews several stationary discrete-time models for multivariate count series.

In contrast to continuous multivariate observations, where vector autoregressive moving-average (VARMA) processes take a dominant role, no single class of count

time series models has emerged as the most flexible, parsimonious, and widely used [Davis et al. (2016), Fokianos and Kedem (2003)]. Many existing models cannot produce an arbitrary count marginal distribution with negative auto- and cross-correlations, a feature present in our hurricane counts. To handle this, a novel count time series model with positive or negative auto- and cross-correlations will be constructed in the next section. Our model also allows for long-range dependence (LRD): the slow autocorrelation decay in time exhibited in many real data sets.[4]

The most popular stationary count time series models are arguably the integer autoregressive moving-average (INARMA) models introduced in Steutel and Van Harn (1979) [see also Alzaid and Al-Osh (1990), Enciso-Mora, Neal and Subba Rao (2009), McKenzie (2003), Neal and Subba Rao (2007)]. INARMA models replace the scalar multiplication in continuous ARMA models with thinning to keep the series integer-valued. The L-dimensional first-order integer autoregressive [INAR(1)] series $\{\mathbf{Y}_t\}$, for example, obeys the recursion

$$(2.1) \qquad\qquad \mathbf{Y}_t = \boldsymbol{\alpha} \circ \mathbf{Y}_{t-1} + \mathbf{Z}_t.$$

Here, $\boldsymbol{\alpha}$ is an $L \times L$ dimensional matrix whose entries $\alpha_{i,j}$ all lie in $[0, 1]$, $\boldsymbol{\alpha} \circ \mathbf{Y}_{t-1}$ is an L-dimensional vector whose ith component is defined as $(\boldsymbol{\alpha} \circ \mathbf{Y}_{t-1})_i = \sum_{j=1}^{L} \alpha_{i,j} \circ Y_{j,t-1}$ and $\{\mathbf{Z}_t\}$ is L-dimensional independent and identically distributed (IID) count-valued noise. The symbol $\circ$ denotes thinning and operates on a nonnegative univariate integer-valued random variable Y via $p \circ Y := \sum_{i=0}^{Y} B_i$, where B_i are IID Bernoulli(p) variables. INAR series of general order and INARMA series are defined in some of the above references.

Many properties of ordinary ARMA models hold for INARMA models. For example, a unique (in mean square) causal stationary solution to (2.1) exists if and only if $\det(\mathbf{I}_L - \boldsymbol{\alpha} z)$ has no roots inside the complex unit circle $|z| \leq 1$ (equivalently, the largest eigenvalue of $\boldsymbol{\alpha}$ has an absolute magnitude less than unity), where $\mathbf{I}_k$ is the $k \times k$ identity matrix. This said, since all thinning probabilities $\alpha_{i,j}$ must lie in $[0, 1]$, one can show that an INARMA model cannot have any negative correlations [Lund and Livsey (2016)]. In this way, INARMA models are not as flexible as ARMA models.

Recently, negatively correlated count series have been investigated in the literature. Kachour and Yao (2009) achieve negative autocorrelation by rounding solutions to continuous Gaussian ARMA equations. For example, the univariate (multivariate extensions are straightforward) rounded integer autoregressive model of order p obeys

$$Y_t = \left\langle \mu + \sum_{j=1}^{p} \phi_j Y_{t-j} \right\rangle + \varepsilon_t,$$

---

[4]Many authors use the term "long memory" when referring to LRD.

where $\langle x \rangle$ rounds x to its nearest integer (round down should there be two nearest integers), μ is a location parameter, $\phi_1, \ldots, \phi_p$ are autoregressive coefficients, and $\{\varepsilon_t\}$ is count-valued IID noise. While such series can have negative autocorrelations, this method, due to the rounding, makes it difficult to produce a pre-specified marginal distribution. The ability for a user to select the marginal distribution can be important; for example, Poisson marginal distributions for the hurricane counts will be sought.

Cui and Lund (2009) use a renewal/point process based approach to devise univariate count time series models with negative autocorrelations. There, a renewal sequence is used to generate a correlated but stationary sequence of zeros and ones. IID copies of these correlated binary processes are then superpositioned akin to Blight (1989) to produce the marginal distribution sought. While renewal methods produce very flexible autocovariance structures in one dimension, they fail in two or more dimensions: in bivariate renewal processes, the item number in use at a large time t is unlikely to be the same for each component. Since different components are typically assumed independent in renewal processes, such methods will produce independent components. While Lund and Livsey (2016) discuss this issue and show how to bypass it, the fixes are unwieldy.

Other count time series methods have been devised; for example, GLARMA series [Dunsmuir (2016)], state-space approaches [Davis and Dunsmuir (2016)], finite mixtures of multivariate Poisson distributions [Karlis and Meligkotsidou (2007)], and hidden Markov techniques [MacDonald and Zucchini (2016)]. See also Barndorff-Nielsen et al. (2014) and Kerss, Leonenko and Sikorskii (2014). A recent and more detailed review of multivariate count time series models can be found in Karlis (2016). These models all have a drawback that precludes them for our use—either a fixed marginal distribution is difficult to achieve or the model cannot produce negative correlations or LRD.

2.2. *Long-range dependent models.* Univariate LRD models have attracted attention across a broad spectrum of scientific disciplines such as finance, economics, computer networks, physics, etc. In the climate sciences, the existence of long-range dependence and scaling phenomena has been intensely debated. In a celebrated work, Hasselmann (1976) advocates that climatic dynamics can often be adequately described by AR(1) processes. This view was challenged in a series of articles reviewed in Mudelsee (2013) that claim that temperatures follow a universal power law, and hence should have LRD features. Varotsos and Efstathiou (2013) examine long memory in tropical cyclone counts (not severe hurricanes); Yuan, Fu and Liu (2014) assert satisfactory performance of a fractionally integrated LRD model in describing Northern Hemisphere temperature anomalies and Pacific decadal oscillations.

Although multivariate LRD has been studied less than its scalar counterpart, it has attracted considerable attention recently. An intuitive definition of multivariate LRD extends the univariate nonsummability characterization: a multivariate

stationary series $\{\mathbf{Y}_t\}$ is said to be LRD if

$$(2.2) \qquad \sum_{h=-\infty}^{\infty} \|\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+h})\| = \infty,$$

where $\|\mathbf{A}\|$ denotes the Frobenius norm of a matrix $\mathbf{A}$. Other definitions of multivariate LRD are possible [see Kechagias and Pipiras (2015) for a detailed treatment on the subject]. The series is short-range dependent (SRD) if the autocovariances are absolutely summable in (2.2). Little has been done on LRD count series; Quoreshi (2014) and Lund, Holan and Livsey (2016) are two recent exceptions, although these are univariate works.

Vector autoregressive fractionally integrated moving average (VARFIMA) series will be used to construct our count time series model; these are bivariate extensions of the celebrated ARFIMA model class, which have been extensively studied and used in applications [Beran et al. (2013), Doukhan, Oppenheim and Taqqu (2003), Giraitis, Koul and Surgailis (2012), Palma (2007), Park and Willinger (2000), Pipiras and Taqqu (2017), Robinson (2003)]. VARFIMA models can capture both LRD and SRD. Moreover, their autocovariance function can often be expressed in a closed form that facilitates computations and statistical inference.

3. The severe hurricane data. Figure 1 depicts the annual number of major hurricanes (Saffir–Simpson Category 3 and above) recorded in the North Atlantic and North Pacific Basins since 1967. Our data commences in 1967 as problems exist in the Pacific record before this time (in pre-satellite years, storms could form over open ocean waters and not be detected). We omit 1966, the first year of satellite era, from our analysis due to the decommission of satellite ESSA-1 amidst the Pacific hurricane season. Saffir–Simpson Category 3+ storms have wind speeds of 111 mph or more at some time during the storm's lifetime. Our focus here is on counts of storms reaching this threshold but do not investigate individual storm intensity. The peak wind speed for each storm is used as a measure of the storm's severity.

Marginally, the two component series are roughly Poisson distributed (there is a slight amount of over-dispersion). Elaborating, from 1967–2015, the sample
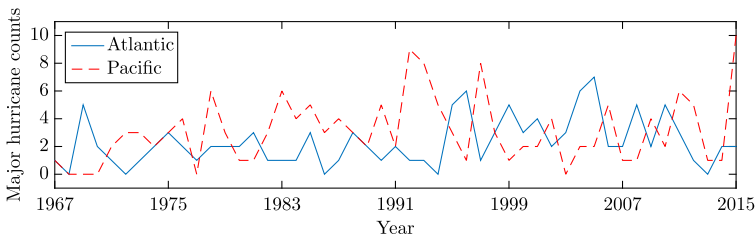

FIG. 1. *Annual number of Saffir–Simpson category* 3 *and stronger hurricanes in the North Pacific and North Atlantic Oceans.*

means and standard deviations of the annual counts are

$$\bar{y}_{\text{Atlantic}} = 2.31, \qquad \bar{y}_{\text{Pacific}} = 3.10,$$
$$s^2_{\text{Atlantic}} = 2.97, \qquad s^2_{\text{Pacific}} = 5.76.$$

The Atlantic major hurricane counts handily pass all Poisson diagnostic checks: a chi-squared goodness-of-fit test with separate bins for the counts $0, 1, 2, \ldots, 7$ and a bin for counts ≥ 8 produced a critical value of 13.00 with 7 degrees of freedom—a p-value of 0.232. For the Pacific series, the same test gives a p-value between 0.05 and 0.1, regardless of the binning choices. Most of the Pacific's Poisson departures is attributed to large counts. While other distributions allowing for over-dispersion and heavier tails are worth consideration (e.g., negative binomial, generalized Poisson), we proceed with a Poisson marginal distribution as roughly reasonable and illustrative.

Figure 2 shows the sample auto- and cross-correlation functions of the Atlantic and Pacific series (blue dashed lines). Pointwise 95% confidence bands for white
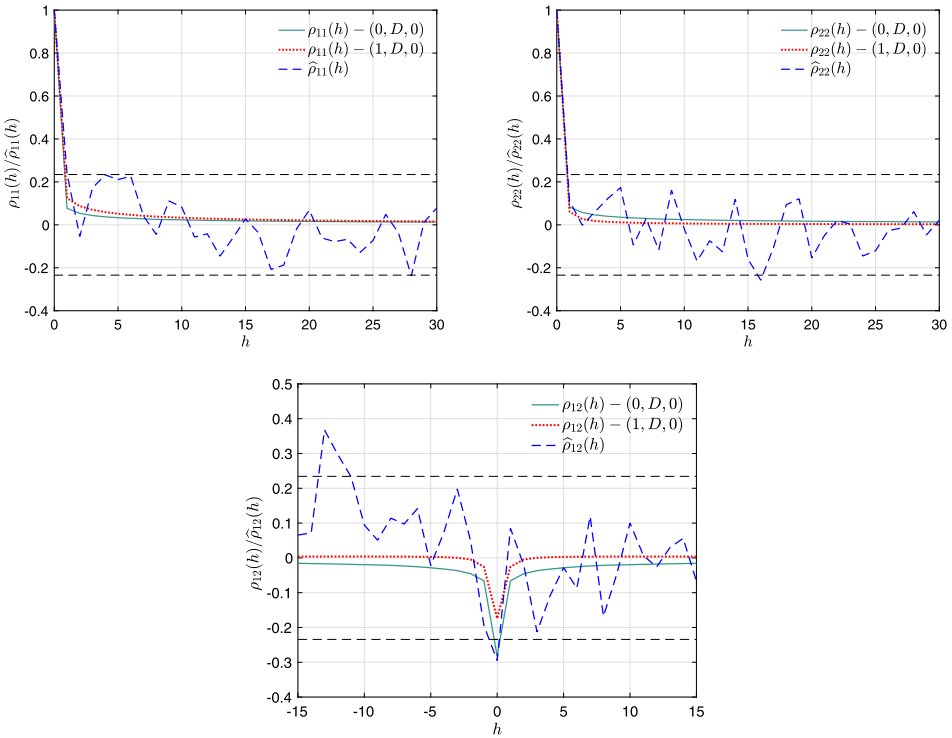

FIG. 2. *Sample (dashed lines) and theoretical (solid lines) auto-correlation functions (top plots) and cross-correlation function (bottom plot) of major hurricane counts in the Atlantic and Pacific Basins. The theoretical auto- and cross-correlation functions are computed using (4.8)–(4.9) with parameter values from Table 2. See Section 7 for more details.*

noise are included. The Atlantic counts are close to white noise; the Pacific counts less so, but still are not heavily correlated. The sample correlation between components (this is a lag zero cross-correlation) is -0.295, hinting that active North Atlantic seasons are typically accompanied by inactive North Pacific seasons (and vice versa). This value is very close to the boundary of $\pm 1.96/\sqrt{T}$, the asymptotic 95% confidence interval for an IID series. Hence, final judgement of the issue is unclear without additional analysis. In what follows, we determine this correlation to be decisively negative. In short, our model will help establish negative correlation inference by utilizing the correlated count structure of the data.

4. A multivariate Poisson count time series model.

This section constructs a multivariate count time series model with Poisson marginal distributions that allows for negative correlations and LRD. For presentation ease, we focus on the bivariate case and construct the model in four steps. We begin with a stationary bivariate Gaussian series.

Step 1: *Start with a bivariate Gaussian series.*
Let $\{\mathbf{X}_t\}_{t \in \mathbb{Z}} = \{(X_{1,t}, X_{2,t})'\}_{t \in \mathbb{Z}}$ be a bivariate, second-order stationary time series with $\mathbb{E}[\mathbf{X_t}] = \mathbf{0}$ and lag-h autocovariance matrix

$$(4.1) \qquad \mathbf{\Gamma_X}(h) = \mathbb{E}[\mathbf{X}_t \mathbf{X}'_{t+h}] = \begin{pmatrix} \gamma_{1,1}(h) & \gamma_{1,2}(h) \\ \gamma_{2,1}(h) & \gamma_{2,2}(h) \end{pmatrix}.$$

We suppose that $\mathbf{X}_t$ follows a bivariate Gaussian distribution for each fixed t, that is,

$$(4.2) \qquad \mathbf{X}_t \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where $\rho = \gamma_{1,2}(0) = \gamma_{2,1}(0)$. The unit marginal variances imply that the autocorrelation function of $\{\mathbf{X}_t\}$ satisfies

$$(4.3) \qquad \mathrm{Corr}(X_{i,t}, X_{j,t+h}) := \rho_{i,j}(h) = \gamma_{i,j}(h), \qquad i,j = 1,2, h \in \mathbb{Z}.$$

At this point, no further assumptions are placed on $\mathbf{\Gamma_X}(h)$ as $h \to \infty$; however, later in this section, a bivariate parametric model for $\{\mathbf{X}_t\}$ is posited that can capture both short- and long-range dependent dynamics.

Step 2: *Place the components of the Gaussian series into categories.*
Let $\{\mathbf{S}_t\}_{t \in \mathbb{Z}}$ be a bivariate series, whose individual components bookkeep the positive/negative signs of the components in $\{\mathbf{X}_t\}$:

$$(4.4) \qquad \mathbf{S}_t = \begin{pmatrix} S_{1,t} \\ S_{2,t} \end{pmatrix} = \begin{pmatrix} 1_{\{X_{1,t} > 0\}} \\ 1_{\{X_{2,t} > 0\}} \end{pmatrix},$$

where 1_A is the indicator of the event A. Lemma 4.1 below shows that $\{\mathbf{S}_t\}_{t \in \mathbb{Z}}$ is stationary and identifies its mean and autocovariance function $\mathbf{\Gamma_S}(h) = \mathbb{E}[\mathbf{S}_t \mathbf{S}'_{t+h}] - \mathbb{E}[\mathbf{S}_t]\mathbb{E}[\mathbf{S}_{t+h}]'$.

LEMMA 4.1. *The series $\{\mathbf{S}_t\}_{t \in \mathbb{Z}}$ is stationary with mean $\mathbb{E}[\mathbf{S}_t] = (1/2, 1/2)'$ and lag-h autocovariance matrix*

$$(4.5) \qquad \mathbf{\Gamma_S}(h) = \frac{1}{2\pi} \begin{pmatrix} \arcsin(\rho_{1,1}(h)) & \arcsin(\rho_{1,2}(h)) \\ \arcsin(\rho_{2,1}(h)) & \arcsin(\rho_{2,2}(h)) \end{pmatrix},$$

where $\rho_{i,j}(h)$, $i, j = 1, 2$, $h \in \mathbb{Z}$, are as in (4.3).

Lemma 4.1 is an immediate consequence of the univariate result given in Section III of Van Vleck and Middleton (1966). Note that $\arcsin(x) < 0$ if and only if $-1 \le x < 0$. Hence, the sign of the auto/cross-correlations of the component series $\{S_{1,t}\}$ and $\{S_{2,t}\}$ is determined by the sign of the auto/cross-correlation functions of $\{X_{1,t}\}$ and $\{X_{2,t}\}$, respectively. Therefore, $\{\mathbf{S}_t\}$ can have negative auto- and cross-correlations. Also, long memory features of $\{\mathbf{X}_t\}$ are passed on to $\{\mathbf{S}_t\}$.

COROLLARY 4.1. *If $\mathbf{\Gamma_X}(\cdot)$ satisfies* (2.2), *then so will $\mathbf{\Gamma_S}(\cdot)$.*

Corollary 4.1 follows directly from the fact $|\arcsin(x)| \ge |x|$ for $x \in [-1, 1]$, implying that $|\arcsin(\rho_{i,j}(h))| \ge |\rho_{i,j}(h)|$.

Step 3: *Superimpose IID copies of* $\{\mathbf{S}_t\}_{t \in \mathbb{Z}}$.

Let $\{\mathbf{S}_t^{(k)}\}_{k=1}^{\infty} = \{(S_{1,t}^{(k)}, S_{2,t}^{(k)})'\}_{k=1}^{\infty}$ be a sequence of IID replicates of the bivariate binary process $\{\mathbf{S}_t\}_{t \in \mathbb{Z}}$. To obtain Poisson marginal distributions, we will superimpose these binary processes as in Blight (1989) and Cui and Lund (2009). More specifically, consider the bivariate count series

$$(4.6) \qquad \mathbf{Y}_t = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{N_{1,t}} S_{1,t}^{(k)} \\ \sum_{k=1}^{N_{2,t}} S_{2,t}^{(k)} \end{pmatrix}, \qquad t \in \mathbb{Z},$$

where for each $i = 1, 2$, and $t \in \mathbb{Z}$,

$$(4.7) \qquad N_{i,t} \sim \text{Poisson}(\lambda_i),$$

for some $\lambda_i > 0$. We also assume that the processes $\{N_{1,t}\}$ and $\{N_{2,t}\}$ consist of independent variables, are mutually independent, and are also independent of the series $\{\mathbf{S}_t^{(k)}\}$, $k = 0, 1, 2, \ldots$. The components $Y_{1,t}$ and $Y_{2,t}$ are Poisson random sums of Bernoulli$(1/2)$ variables; hence, they have Poisson distributions with means $\lambda_1/2$ and $\lambda_2/2$, respectively.

In Proposition 4.1 below, $\{\mathbf{Y}_t\}$ is shown to be stationary and its mean and autocovariance function are derived. The autocovariance function involves the cumulative distribution function (CDF) of the random variable $W = M_1 - M_2$, where M_1 and M_2 are independent Poisson random variables with means λ_1 and λ_2, respectively. The random variable W follows the so-called Skellam(λ_1, λ_2) distribution [Skellam (1946)], whose CDF $F_W(\cdot; \lambda_1, \lambda_2)$ can be computed accurately and efficiently.

PROPOSITION 4.1. *The series $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$ is stationary with mean $\mathbb{E}[\mathbf{Y}_t] = (\lambda_1/2, \lambda_2/2)'$ and lag-h autocovariance matrix*

$$(4.8) \qquad \Gamma_{\mathbf{Y}}(h) = \frac{1}{2\pi} \begin{pmatrix} c_{1,1} \arcsin(\rho_{1,1}(h)) & c_{1,2} \arcsin(\rho_{1,2}(h)) \\ c_{2,1} \arcsin(\rho_{2,1}(h)) & c_{2,2} \arcsin(\rho_{2,2}(h)) \end{pmatrix},$$

where $\rho_{i,j}(h)$, $h \in \mathbb{Z}$, $i, j = 1, 2$, are as in (4.3) *and*

$$(4.9) \qquad c_{i,j} = \begin{cases} 2\lambda_i, & i = j, h = 0, \\ \lambda_i F_W(-1; \lambda_1, \lambda_2) + \lambda_j[1 - F_W(1; \lambda_1, \lambda_2)], & \text{otherwise}, \end{cases}$$

where $F_W(\cdot; \lambda_1, \lambda_2)$ is the CDF of the Skellam(λ_1, λ_2) distribution.

Corollary 4.1 applies here and shows that LRD in $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ will be inherited in $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$. Relation (4.8) will aid statistical inference, our Section 5 objective.

Step 4: *Select a parametric model for* $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$.

Given $d_1, d_2 \in (-1/2, 1/2)$, set $\mathbf{D} = \text{diag}(d_1, d_2)$ and let $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$ be the usual autoregressive and moving-average polynomials of orders p and q; namely,

$$\boldsymbol{\Phi}(z) = \mathbf{I}_2 - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p, \qquad \boldsymbol{\Theta}(z) = \mathbf{I}_2 + \boldsymbol{\Theta}_1 z + \cdots + \boldsymbol{\Theta}_q z^q,$$

where $\boldsymbol{\Phi}_i, i = 1, \ldots, p$, $\boldsymbol{\Theta}_j, j = 1, \ldots, q$, are 2×2 matrices. Let $\{\boldsymbol{\eta}_t\}_{t \in \mathbb{Z}} = \{(\eta_{1,t}, \eta_{2,t})'\}_{t \in \mathbb{Z}}$ be a bivariate Gaussian white noise series with mean $\mathbb{E}[\boldsymbol{\eta}_t] \equiv \mathbf{0}$ and covariance matrix $\mathbb{E}[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \boldsymbol{\Sigma} = (\sigma_{i,j})_{i,j=1,2}$. Recall that the univariate fractional differencing/integration operator $(I - B)^d$, where $I = B^0$ and B denote the identity and backshift operator, respectively, is defined through the Taylor series

$$(I - B)^d = \sum_{k=0}^{\infty} b_k B^k, \qquad \text{with } b_k = \frac{\Gamma(k+1)}{\Gamma(d)\Gamma(k+d)}, k = 0, 1, \ldots,$$

for any $d \in (-1/2, 1/2)$ [see, e.g., Beran et al. (2013), Pipiras and Taqqu (2017)].

Now suppose that $\{\mathbf{X}_t\}$ of Step 1 is a VARFIMA$(p, \mathbf{D}, q)$ series satisfying

$$(4.10) \qquad (I - B)^{\mathbf{D}} \boldsymbol{\Phi}(B) \mathbf{X}_t = \boldsymbol{\Theta}(B) \boldsymbol{\eta}_t,$$

where the operator $(I - B)^{\mathbf{D}}$ is understood to be

$$(I - B)^{\mathbf{D}} = \begin{pmatrix} (I - B)^{d_1} & 0 \\ 0 & (I - B)^{d_2} \end{pmatrix}.$$

The parameters d_1 and d_2 govern the decay rate of the autocovariances of $\{\mathbf{X}_t\}$ to zero.

REMARK 4.1. When $p = q = 0$, the components of the lag-h autocovariance matrix $\boldsymbol{\Gamma}_{\mathbf{X}}(h)$ of a VARFIMA$(0, \mathbf{D}, 0)$ series are

$$(4.11) \qquad \begin{aligned} \gamma_{i,j}(h) &= \sigma_{i,j} \frac{(-1)^h \Gamma(1 - d_i - d_j)}{\Gamma(1 - d_i + h)\Gamma(1 - d_j - h)} \\ &\sim \kappa_{i,j} h^{d_i + d_j - 1} \qquad \text{as } h \to \infty, \end{aligned}$$

for $i, j = 1, 2$, and for some constants $\kappa_{i,j}$.[5] Equation (4.11) illuminates the role of
the parameters d_1, d_2: if $d_1, d_2 \in (0, 1/2)$, the power-law decay in (4.11) implies
that $\{\mathbf{X}_t\}$ has LRD. When $d_1, d_2 \in (-1/2, 0)$, the left-hand side of (2.2) is finite
and the series exhibits a special type of SRD called anti-persistence. When $d_i = 0$
for $i = 1$ or 2, $\gamma_{i,i}(h) = 0$ for $h > 0$, implying that the corresponding component
series $\{X_{i,t}\}$ is white noise. Finally, the asymptotic relation in (4.11) holds for any
p, q for suitable constants $\kappa_{i,j}$.

REMARK 4.2.   When $p = 0$ and $q \geq 1$, $\mathbf{\Gamma_X}(h)$ can still be efficiently calcu-
lated [see Proposition 3.1 in Kechagias and Pipiras (2017) with $c = 0$]. When
$p = 1$, explicit expressions for $\mathbf{\Gamma_X}(h)$ are not known, but autocovariances can be
numerically computed up to any desired accuracy; however, an additional assump-
tion on the AR polynomial $\mathbf{\Phi}(B)$ is needed [all of its eigenvalues need to be pos-
itive; Sela (2010)]. Finally, computing $\mathbf{\Gamma_X}(h)$ when $p \geq 2$ is not straightforward.
As such, we focus on models with $p = 0, 1$.

REMARK 4.3.   The white noise series $\{\boldsymbol{\eta}_t\}$ must meet certain criteria for $\{\mathbf{X}_t\}$
in (4.10) to satisfy (4.2). The zero mean and Gaussian distribution of $\{\mathbf{X}_t\}$ follow
directly from the zero mean and Gaussian distribution of $\{\boldsymbol{\eta}_t\}$. However, the struc-
ture of the variance matrix in (4.2) implies that only the nondiagonal entries of $\mathbf{\Sigma}$
are free parameters. When $p = q = 0$, the variance structure is achieved by setting
$\gamma_{1,1}(0) = \gamma_{2,2}(0) = 1$, and $\gamma_{1,2}(0) = \rho$ in the first equality of (4.11), and then solv-
ing for $\sigma_{1,1}, \sigma_{2,2}$ and $\sigma_{1,2}$. The same technique can be used when $p = 1$ or $q = 1$;
unfortunately, in these cases, the solution of the linear system is not necessarily
unique, and in fact may not even be a positive definite matrix. This issue is revis-
ited at the end of Section 5. In particular, the cross-correlation ρ will be used as a
free parameter in the estimation procedure below.

5. Inference.   This section puts forth a quasi-maximum likelihood estima-
tion (QMLE) method for the model in Section 4. We consider the underlying
VARFIMA$(p, \mathbf{D}, q)$ series $\{\mathbf{X}_t\}$ satisfying (4.2) and the orders $(p, q) = (0, 0)$,
$(p, q) = (0, 1)$, and $(p, q) = (1, 0)$. Let $\boldsymbol{\xi}$ contain all model parameters; these in-
clude the long memory parameters d_1 and d_2, the Poisson means λ_1 and λ_2, the
cross-correlation ρ (instead of the parameters in $\mathbf{\Sigma}$—see Remark 4.3) and all pa-
rameters in $\mathbf{\Phi}$ and $\mathbf{\Theta}$.

The exact likelihood structure of the count series data $\{\mathbf{Y}_t\}_{t=1,\ldots,T}$, where T is
the sample size, has proven to be intractable [Davis et al. (2016), Fokianos and
Kedem (2003)]. Nonetheless, analogous to ordinary time series theory, a quasi
log-likelihood can be devised from the model's autocovariance function. Using

---

[5]For two sequences, $\{a_n\}_{n \in \mathbb{N}}$ and $\{b_n\}_{n \in \mathbb{N}}$, $a_n \sim b_n$ stands for $a_n/b_n \to 1$ as $n \to \infty$.

the multivariate Durbin–Levinson (DL) algorithm [Brockwell and Davis (2006)], this quasi log-likelihood has the form

$$(5.1) \qquad L(\boldsymbol{\xi}) \propto -\frac{1}{2} \sum_{t=1}^{T} \log |\mathbf{V}_{t-1}| - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)' \mathbf{V}_{t-1}^{-1} (\mathbf{Y}_t - \widehat{\mathbf{Y}}_t),$$

where $\widehat{\mathbf{Y}}_t := \mathbb{E}[\mathbf{Y}_t | \mathbf{1}, \mathbf{Y}_1, \ldots, \mathbf{Y}_{t-1}]$ is the best linear one-step-ahead predictor of $\mathbf{Y}_t$ from a constant and a process history ($\widehat{\mathbf{Y}}_1 = E[\mathbf{Y}_1]$) and $\mathbf{V}_{t-1} := \mathbb{E}[(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)']$ is the corresponding mean squared error. These quantities can be recursively obtained from the multivariate DL algorithm that has computational complexity $O(T^2)$; however, the form in (5.1) conveniently bypasses inversion of a $2T \times 2T$ covariance matrix. Such an inversion is needed in a brute force evaluation of the Gaussian likelihood, an approach that should only be chosen over (5.1) for small sample sizes.

The QMLE parameter estimates are

$$(5.2) \qquad \widehat{\boldsymbol{\xi}} = \underset{\boldsymbol{\xi} \in \mathcal{S}}{\arg\max} L(\boldsymbol{\xi}),$$

where the k-dimensional ($k = 5 + 4p + 4q$) parameter space $\mathcal{S}$ is defined as

$$\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^k : -1/2 < d_1, d_2 < 1/2, \lambda_1, \lambda_2 > 0, -1 < \rho < 1\}.$$

The estimates $\widehat{\boldsymbol{\xi}}$ do not have a closed form, but can be computed numerically from a quasi-Newton algorithm. This algorithm is available in the NLPQN function of SAS/IML, which is the software used in Sections 6 and 7. Moreover, using the NLPFDD function, we computed the inverse Hessian of the likelihood function which we used to obtain confidence intervals.

We conclude this section with some observations about $\mathcal{S}$. First, in view of Remark 4.1, it is important to allow the parameters d_1, d_2 to take negative values; optimization with $d_1, d_2 \in (0, 1/2)$ may yield artificial LRD in the sense that positive d_1 or d_2 are obtained due to parameter constraints and not because of the underlying LRD. Second, no constraints are imposed on the entries of $\boldsymbol{\Phi}_1$ or $\boldsymbol{\Theta}_1$ except the following: in numerical implementation of (5.2), we set $L(\boldsymbol{\xi}) = -\infty$ if $\boldsymbol{\Phi}_1$ or $\boldsymbol{\Theta}_1$ have any eigenvalues whose absolute modulus exceeds unity. This condition is equivalent to requiring that all roots of $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$ lie outside of the complex unit circle and is a standard assumption guaranteeing that a causal and invertible solution to the VARMA difference equation exists [Lütkepohl (2005)]. Candidate maximizers of $\boldsymbol{\xi}$ that violate this constraint are assigned an infinitesimally small likelihood to ensure that estimators are causal and invertible. Finally, the restrictions on $\boldsymbol{\Sigma}$ discussed in Remark 4.3 cause issues for VARFIMA$(0, \mathbf{D}, 1)$ and VARFIMA$(1, \mathbf{D}, 0)$ models, where autocovariance functions are more complex than those in the simpler VARFIMA$(0, \mathbf{D}, 0)$ model. In these cases, one can still compute the parameters $\sigma_{1,1}, \sigma_{1,2}$, and $\sigma_{2,2}$ that ensure marginal unit variances and a prescribed correlation ρ for $\{\mathbf{Y}_t\}$ by solving a linear system whose coefficients are nonlinear functions of $d_1, d_2, \rho, \boldsymbol{\Phi}$, and $\boldsymbol{\Theta}$. After experimenting with

several parameter schemes for $d_1, d_2, \rho, \mathbf{\Phi}$, and $\mathbf{\Theta}$, these systems were found to always have unique solutions. However, these solutions did not always lead to a positive definite estimate of $\mathbf{\Sigma}$. We dealt with such candidate maximizers of $L(\boldsymbol{\xi})$ by again assigning them a log-likelihood of $-\infty$.

6. Simulation study. This section fits the model of Section 4 to the simulated bivariate count data via the QMLE method of the last section. The VARFIMA orders $(0, \mathbf{D}, 0)$, $(0, \mathbf{D}, 1)$, and $(1, \mathbf{D}, 0)$ are considered. For each model, 100 series were simulated with $T = 200, 400$ and several VARFIMA parameter values. To generate the underlying Gaussian series, the fast and exact synthesis algorithm of Helgason, Pipiras and Abry (2011) is used. The steps in Section 4 are followed to generate the count series.

Table 1 shows the median bias (MB) and median absolute deviation (MAD) for the obtained estimates. The true parameter values were $d_1 = 0.3, d_2 = 0.2, \lambda_1 = 3$, $\lambda_2 = 2$ for all columns and $\rho = 0.45, -0.9, 0.5, -0.9$ for columns $(0, 0)_1, (0, 0)_2$,

TABLE 1
Median bias (top entries of each case) and median absolute deviation (bottom entries of each case) of estimated parameters for the three models. The true parameter values are $d_1 = 0.3, d_2 = 0.2, \lambda_1 = 3, \lambda_2 = 2$ *for all columns,* $\rho = 0.45, -0.9$ *for columns* $(0, 0)_1$ *and* $(0, 0)_2$, *respectively,* $\Phi_{1,1} = 0.4, \Phi_{1,2} = 0.1, \Phi_{2,1} = 0.3, \Phi_{2,2} = 0.6$ *for column* $(1, 0)$ *and* $\Theta_{1,1} = 0.1$, $\Theta_{1,2} = -0.6, \Theta_{2,1} = 0.2, \Theta_{2,2} = 0.8$ *for column* $(0, 1)$

| (p, q)	$(0, 0)_1$		$(0, 0)_2$		$(1, 0)$		$(0, 1)$	
T	200	400	200	400	200	400	200	400
d_1	−0.011	−0.010	−0.013	−0.024	−0.029	−0.039	0.059	0.069
	0.082	0.059	0.075	0.061	0.138	0.107	0.057	0.038
d_2	−0.026	0.009	−0.003	0.010	−0.019	−0.190	0.071	0.085
	0.093	0.065	0.093	0.074	0.183	0.179	0.070	0.047
λ_1	−0.046	−0.021	0.018	−0.028	−0.049	−0.065	−0.024	−0.013
	0.273	0.229	0.320	0.214	0.309	0.338	0.231	0.199
λ_2	0.029	0.023	−0.019	0.017	−0.077	−0.035	−0.023	0.010
	0.157	0.125	0.210	0.171	0.228	0.165	0.156	0.149
ρ	0.072	0.045	0.001	−0.024	−0.002	0.042	0.060	−0.011
	0.218	0.155	0.077	0.050	0.196	0.163	0.110	0.050
$\Phi_{1,1}/\Theta_{1,1}$					0.069	0.248	−0.004	−0.017
					0.243	0.267	0.021	0.023
$\Phi_{1,2}/\Theta_{1,2}$					0.017	−0.090	0.165	0.145
					0.241	0.178	0.085	0.086
$\Phi_{2,1}/\Theta_{2,1}$					−0.023	−0.071	0.130	0.133
					0.158	0.168	0.069	0.047
$\Phi_{2,2}/\Theta_{2,2}$					0.052	0.088	0.029	0.032
					0.148	0.140	0.020	0.022
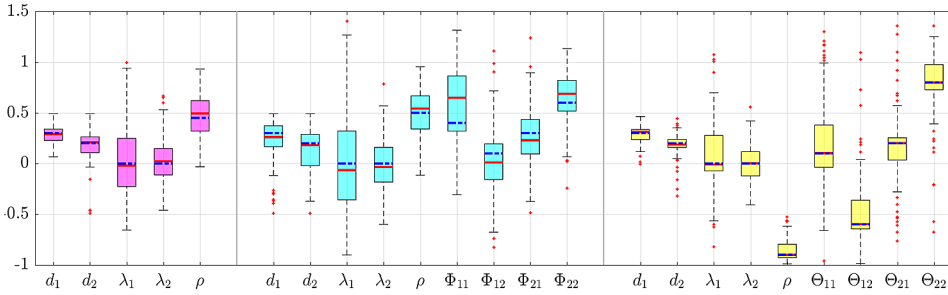
FIG. 3. *Boxplots of the estimates from columns* $(0,0)_1$ (*left box*), $(1,0)$ (*middle box*) *and* $(0,1)$ (*right box*) *of Table* 1 *for* $T = 400$. *The dashed blue lines correspond to the true parameter values, while the solid red lines are the medians.*

$(1,0)$, and $(0,1)$, respectively. Finally, for columns $(1,0)$ and $(0,1)$ the AR and MA coefficient matrices were

$$\mathbf{\Phi} = \begin{pmatrix} 0.4 & 0.1 \\ 0.3 & 0.6 \end{pmatrix} \quad \text{and} \quad \mathbf{\Theta} = \begin{pmatrix} 0.1 & -0.6 \\ 0.2 & 0.8 \end{pmatrix}$$

respectively. Overall, the QMLE method performs very well in terms of MBs in most cases, even though the sample sizes used here are considered small/medium in the LRD literature.

Most MBs and MADs decrease with increasing sample size. An exception is the MBs for some parameters in models with SRD components, especially those in the $(1,0)$ column. This is attributed to the negative definiteness issue discussed at the end of Section 5. Nevertheless, all MBs and MADs did decrease when the sample size $T = 1000$ was considered. Other parameter schemes were experimented with and produced similarly good results, but are not shown here for brevity's sake.

The boxplots in Figure 3 provide a distributional view of the parameter estimates from columns $(0,0)_1$, $(0,1)$, and $(1,0)$ of Table 1 for $T = 400$. The dashed blue lines demarcate the true parameter values, while red lines show medians. The boxplots for estimates of λ_1 and λ_2 are centered at zero by subtracting the true parameter value, providing a uniform presentation scale. Finally, when experimenting with larger sample sizes, the symmetry/outliers in these boxplots increased/decreased significantly.

7. Hurricane data. Table 2 displays parameter estimates and the corresponding AIC and BIC scores obtained by fitting the bivariate count model of Section 4 to the hurricane count series. We considered here three underlying Gaussian processes: a VARFIMA$(0, \mathbf{D}, 0)$, a VARFIMA$(1, \mathbf{D}, 0)$, and an IID series. Subscripts of unity refer to the Atlantic Basin, subscripts of two refer to the Pacific Basin, and $\Phi_{i,j}, i, j = 1, 2$ are the entries of the 2×2 AR matrix Φ_1. The IID model produced the smallest information criteria, however, for the univariate Atlantic series, where

TABLE 2
Parameter estimates and corresponding AIC and BIC scores for the hurricane series for VARFIMA
$(0, \mathbf{D}, 0)$, $(1, \mathbf{D}, 0)$, *and* $(0, \mathbf{0}, 0)$ *models*

Model	d_1	d_2	λ_1	λ_2	ρ	$\Phi_{1,1}$	$\Phi_{1,2}$	$\Phi_{2,1}$	$\Phi_{2,2}$	AIC	BIC
$(0, \mathbf{D}, 0)$	0.24	0.23	5.7	11.5	−0.96					239	249
$(1, \mathbf{D}, 0)$	−0.4	0.12	5.6	10.3	−0.70	0.88	0.26	−0.01	0.01	242	259
$(0, \mathbf{0}, 0)$	0	0	5.8	11.3	−0.97					237	242

a longer reliable record is available, LRD models are decisively preferred over the IID one by AIC. We investigate this phenomena further in Section 7.1.

The quasi-Newton algorithm converged for all models, but termination criteria differed. In the VARFIMA$(1, \mathbf{D}, 0)$ case, the maxima occurred at the boundary of the feasible region ($|\mathbf{\Sigma}| < 0$ for $\rho < -0.7$), while for the IID and VARFIMA$(0, \mathbf{D}, 0)$ models, all gradient values were $< 10^{-5}$ at the maxima indicating that the maxima occurred at an interior point of $\mathcal{S}$. The VARFIMA$(0, \mathbf{D}, 1)$ and VARFIMA$(0, \mathbf{D}, 0)$ models produced almost identical estimates, but the VARFIMA$(0, \mathbf{D}, 1)$ model had a negligible increase in log-likelihood; hence, we omit listing VARFIMA$(0, \mathbf{D}, 1)$ results in the table. For all models investigated, multiple starting points of the parameters were investigated to ensure globally optimal estimators were found in the step and search optimization algorithm. All estimated LRD parameters are between 0 and 1/2, except for the Atlantic series under the VARFIMA$(1, \mathbf{D}, 0)$ model, which is −0.4. In this case, the dependence in the series is captured by the AR coefficients $\widehat{\Phi}_{1,1}$ and $\widehat{\Phi}_{1,2}$ [this is a common phenomenon in estimation of Gaussian VARFIMA$(1, \mathbf{D}, 0)$ series with a small sample size, especially when one of the AR parameters is significantly greater than zero].

As the AIC and BIC scores are smaller for the IID and VARFIMA$(0, \mathbf{D}, 0)$ fit compared to the VARFIMA$(1, \mathbf{D}, 0)$ case, confidence intervals (CIs) for the parameters of these models will be reported. Standard errors were obtained from the usual second derivative of the quasi log-likelihood function. First, the 90% CIs for ρ are $[-1.00, -0.75]$ and $[-1.00, -0.59]$ for the IID and VARIFMA$(0, \mathbf{D}, 0)$ models, respectively. Hence, ρ is decisively negative and the negative correlation between basin counts appears real. Substituting the VARFIMA$(0, \mathbf{D}, 0)$ estimates into (4.8)–(4.9) yields a lag zero cross-correlation of −0.28, which closely matches the sample cross-correlation of −0.295. Second, 90% CIs for the LRD parameters are $[-0.095, 0.50]$ for the Atlantic Basin and $[-0.235, 0.50]$ for the Pacific Basin. As both intervals contain zero, long memory cannot be definitively declared in either basin, despite the positive estimates of the LRD parameters. Of course, wide intervals are expected with only 49 years of data; a few years of additional data may change this conclusion, especially for the Atlantic Basin, which was a close call. Finally, the 90% Poisson parameter confidence intervals from the IID and

VARFIMA$(0, \mathbf{D}, 0)$ models are $[3.90, 7.72]$ and $[3.81, 7.60]$ (Atlantic Basin) and $[7.78, 14.78]$ and $[7.21, 15.82]$ (Pacific Basin), respectively.

Recall that the mean of the ith component series is $\lambda_i/2$ for $i = 1, 2$. For feel, the auto- and cross-correlations of the fitted VARFIMA$(0, \mathbf{D}, 0)$ (green solid lines) and VARFIMA$(1, \mathbf{D}, 0)$ models (red dotted lines) are plotted together with the sample auto- and cross-correlations (blue dashed lines) in Figure 2—no radical disagreements are seen.

7.1. *Further investigation into LRD.* While long memory cannot be declared to any reasonable degree of statistical confidence, it cannot be discounted with only 49 years of observations. In this vein, we investigate here how long the observation record must be to conclude long memory with our model and we look further into long-memory dynamics of the Atlantic Basin where a longer reliable sample is available.

First, we generated bivariate LRD count series for various parameter values and checked whether confidence intervals for LRD parameters contain zero (corresponding to the SRD case and, more specifically, IID counts). The empirical proportions of confidence intervals containing zero are presented in Figure 4, for parameter values indicated in the figure titles. Note that the smallest sample sizes needed to distinguish between the IID and LRD counts are relatively small. In fact, at level 50% and for larger values of d, they are around the sample size of the data considered in this paper. But for higher certainty, larger sample sizes (than that in this paper) are required. We note that we also investigated the sample size issue through the portmanteau tests as in Percival, Overland and Mofjeld (2001), but found the needed sizes for concluding long memory to be considerably larger than those deduced from confidence intervals.

To further investigate the Atlantic Basin major hurricane counts, we fitted univariate IID and Poisson(λ)-ARFIMA$(0, d, 0)$ models to data starting in 1900. The
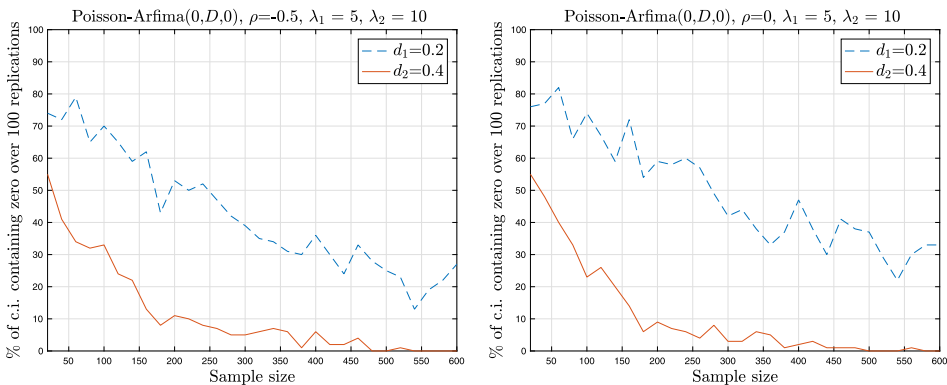

FIG. 4. *Empirical proportions of confidence intervals containing zero as functions of $\log_2 T$.*
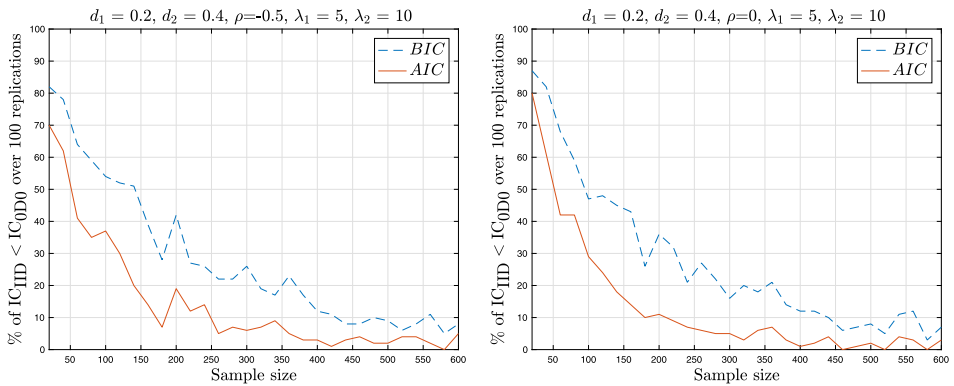
FIG. 5. *Empirical proportions over* 100 *replications where information criteria pick the IID model over the LRD model, where the latter is used in the data generating process.*

Poisson(λ)-ARFIMA$(0, d, 0)$ had smaller information criteria and produced 95% confidence intervals $[0.224, 0.497]$ and $[4.851, 6.989]$ for d and λ, respectively. Note that although aircraft surveillance in the Atlantic Basin commenced in 1944, data with good reliability on tropical cyclones striking the east coast extends back to 1900 (according to the web archive of the Atlantic Oceanographic and Meteorological Laboratory of the National Oceanic and Atmospheric Administration). For more closure on this issue, we ran a simulation study to see how often LRD is rejected by AIC/BIC when the true data generating process is in fact LRD. More specifically, we generated synthetic data using the Poisson(λ_1, λ_2)-VARFIMA$(0, \mathbf{D}, 0)$ model and then fitted two models, a misspecified IID model and the true one. Figure 5 shows empirical proportions (over 100 replications) where AIC and BIC selected the misspecified IID model over the true one for several sample sizes and parameter schemes. AIC performed well in most cases we considered, selecting the true model more than 50% of the time for sample sizes as small as 60 (red solid lines).

7.2. *Model diagnostics.* We comment here on some procedures for model diagnostics for count time series as they relate to our model and its fit to the hurricane data. More specifically, we focus on the probability integral transform (PIT) and related diagnostics plots as introduced in Czado, Gneiting and Held (2009) and further analyzed in Kolassa (2016).

These diagnostics tools are based on the one-step-ahead predictive distributions, and have been considered in the case of univariate data. Applying them here raises at least two difficulties. One is transitioning from the univariate to the bivariate case, which should be examined apart from our model. For this reason, we consider here only the univariate analogue of our model and the univariate component series, say the counts of the Atlantic hurricanes. Another issue is that our count
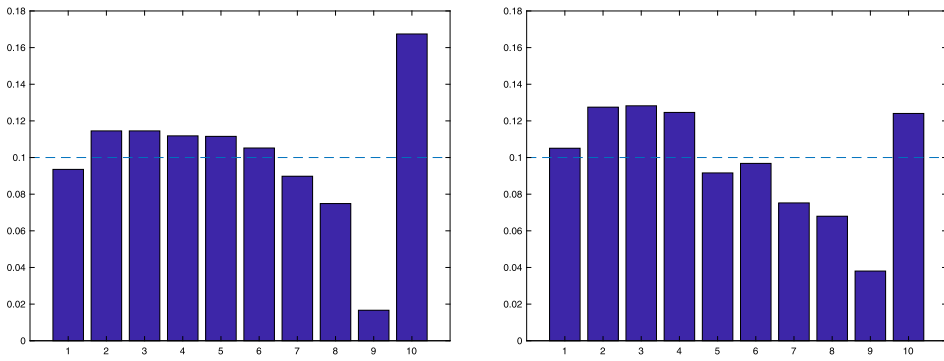
FIG. 6.   *PIT histograms for IID* (*left*) *and Poisson-ARFIMA*(0, *d*, 0) (*right*) *models fit to the Atlantic hurricane counts.*

time series model does not lend itself to analytic calculations of predictive distributions. But we have an efficient numerical algorithm to compute these predictive distributions through Monte-Carlo simulations. Describing the method here would go beyond the scope and the main theme of the paper—the details will appear elsewhere. We hence limit our discussion to several illustrative plots and some observations.

Figure 6 presents nonrandomized PIT histograms for the Atlantic hurricane counts. The left plot corresponds to the Poisson IID model and the right plot corresponds to the Poisson(λ)-ARFIMA(0, d, 0) model, with the parameter values taken as those estimated from the data. A uniform PIT histogram suggests an adequate fit. The U-shapes of the histograms point to an overdispersed fitted model, though the PIT histogram for the Poisson-ARFIMA(0, d, 0) model seems to suggest a slightly improved fit (especially focusing on the last few bins). In fact, in the IID case, we examined the PIT histograms for several other common overdispersed distributions (e.g., negative binomial), but the disagreement over the last few bins in the PIT histograms remained present.

8. Conclusions.   This paper introduced a novel stationary bivariate count time series model with Poisson marginal distributions and possible negative correlations and long-range dependence. Most count models developed to date do not allow combination of these three features. The model was used to analyze annual severe hurricane counts in the North Atlantic and Pacific Basins, series with important climatic ramifications that have been intensely scrutinized by climatologists [Mooney (2007)]. We find a definite negative correlation between the two basins. Although we are not able to decisively prove existence of long-memory dynamics, a convincing argument is presented for the Atlantic Basin and a foundation for discussion in the Pacific.

Modifications to our model are worth exploring. For example, negative binomial marginal distributions on the support set {0, 1, . . .} can be produced with our

tactics—one need only take $\{N_{1,t}\}$ and $\{N_{2,t}\}$ in (4.6) to be independent processes, each themselves composed of IID negative binomial draws. Negative binomial marginal distributions are over-dispersed and have been suggested as marginal distributions for hurricane counts in Chu and Zhao (2004) and Villarini, Vecchi and Smith (2010). However, complexities arise. Even eschewing time-correlation aspects, there is no guarantee that a bivariate random pair whose marginals are negative binomial and have a negative correlation as high as -0.3 exists. We have yet to be able to construct a case, using the above paradigm, where the correlation between the pair is anything lower than $-1/4$ (for any negative binomial parameters and background Gaussian correlation parameter, much less for parameters that give us roughly the means of our counts). Some methods of Joe (1996) will produce series with higher positive correlations than $1/4$, but negatively correlated series cannot be achieved with those methods. When we fitted this negative binomial version of the model to the count data, the estimated parameters migrated toward a degenerate bivariate Gaussian with correlation between components of -1: the fit pushed up against "model class boundaries" and was numerically unstable. Other marginal distributions are possible; these are currently being probabilistically formalized in Jia and Lund (2016).

It may be desirable to include covariates in the analysis. One simple way to do this is to allow the parameters λ_i to depend on the covariates via a log link. While the resulting series will not be technically stationary, they are natural variants of stationary series.

APPENDIX

The results stated in Section 4 are proven here.

PROOF OF PROPOSITION 4.1.   For the mean claim, $\mathbb{E}[\sum_{\ell=1}^{N_{i,t}} S_{i,t}^{(\ell)}|N_{i,t}=k] = k/2$ implies that

$$(A.1)\qquad \mathbb{E}[Y_{i,t}] = \mathbb{E}\left[\sum_{\ell=1}^{N_{i,t}} S_{i,t}^{(\ell)}\right] = \mathbb{E}\left[\mathbb{E}\left(\sum_{\ell=1}^{N_{i,t}} S_{i,t}^{(\ell)}|N_{i,t}\right)\right] = \frac{1}{2}\mathbb{E}[N_{i,t}] = \frac{\lambda_i}{2}.$$

Next, let $p_{i,j}(h) = P(N_{i,t}=n_i, N_{j,t+h}=n_j)$ and condition on $N_{i,t}$ and $N_{j,t+h}$ to get

$$\mathbb{E}[Y_{i,t}Y_{j,t+h}] = \mathbb{E}\left[\sum_{m=1}^{N_{i,t}} S_{i,t}^{(m)} \sum_{k=1}^{N_{j,t+h}} S_{j,t+h}^{(k)}\right]$$

$$(A.2)\qquad = \mathbb{E}\left[\mathbb{E}\left(\sum_{m=1}^{N_{i,t}} \sum_{k=1}^{N_{j,t+h}} S_{i,t}^{(m)} S_{j,t+h}^{(k)}|N_{i,t}, N_{j,t+h}\right)\right]$$

$$= \sum_{n_i,n_j=0}^{\infty} \mathbb{E}\left[\sum_{m=1}^{n_i} \sum_{k=1}^{n_j} S_{i,t}^{(m)} S_{j,t+h}^{(k)}\right] p_{i,j}(h).$$

Using Lemma 4.1 and the independence of $\{S_{i,t}^{(m)}\}$ and $\{S_{j,t+h}^{(k)}\}$ when $m \neq k$, we get

$$
\text{(A.3)} \qquad \mathbb{E}\big[S_{i,t}^{(m)} S_{j,t+h}^{(k)}\big] = \begin{cases} \dfrac{1}{4}, & m \neq k, \\ \dfrac{1}{4} + \dfrac{\arcsin(\rho_{ij}(h))}{2\pi}, & m = k. \end{cases}
$$

Let $\Pi_{i,j} = n_i n_j$, $M_{i,j} = \min(n_i, n_j)$, and observe that the last row of (A.2) has $M_{i,j}$ different cross products of the form $S_{i,t}^{(m)} S_{j,t+h}^{(k)}$, where $m = k$ and $\Pi_{i,j} - M_{i,j}$ cross products of the same form when $m \neq k$. Using (A.3) in (A.2) provides

$$
\begin{aligned}
\mathbb{E}[Y_{i,t} Y_{j,t+h}] &= \sum_{n_i, n_j = 0}^{\infty} \left[M_{i,j}\left(\frac{1}{4} + \frac{\arcsin(\rho_{i,j}(h))}{2\pi} \right) + \frac{(\Pi_{i,j} - M_{i,j})}{4} \right] p_{i,j}(h) \\
&= \sum_{n_i, n_j = 0}^{\infty} \left[M_{i,j} \frac{\arcsin(\rho_{ij}(h))}{2\pi} + \frac{\Pi_{ij}}{4} \right] p_{i,j}(h) \\
&= \frac{1}{2\pi} \arcsin\big(\rho_{i,j}(h)\big) \mathbb{E}\big[\min(N_{i,t}, N_{j,t+h})\big] \\
&\quad + \frac{1}{4}\mathbb{E}[N_{i,t} N_{j,t+h}].
\end{aligned}
$$

(A.4)

The expectation in the second term in the last row of (A.4) is readily calculated using the independence assumption under (4.7):

$$
\text{(A.5)} \qquad \mathbb{E}[N_{i,t} N_{j,t+h}] = \begin{cases} \lambda_i \lambda_j, & i \neq j, \\ \lambda_i + \lambda_i^2, & h = 0, i = j. \end{cases}
$$

On the other hand, as shown in Lemma A.1 below, the expectation of the first term in the last row of (A.4) is

$$
\text{(A.6)} \qquad \mathbb{E}\big[\min(N_{i,t}, N_{j,t+h})\big] = \lambda_i F_W(-1) + \lambda_j \big[1 - F_W(1)\big],
$$

where F_W is the CDF of a Skellam random variable with parameters λ_1 and λ_2.

The autocovariance in (4.8) now follows from (A.1), (A.5), (A.6), and the last row in (A.3). $\square$

LEMMA A.1. *Suppose that M_1 and M_2 are independent Poisson variables with mean $\mathbb{E}[M_i] = \lambda_i$ for $i = 1, 2$. Define $W = M_1 - M_2$ and $Y = \min(M_1, M_2)$. Then*

$$
\text{(A.7)} \qquad \mathbb{E}[Y] = \lambda_1 F_W(-1) + \lambda_2 \big[1 - F_W(1)\big],
$$

where F_W is the CDF of W.

PROOF.   Let $P_i = P(m_i; \lambda_i)$, $m_i = 0, 1, 2 \ldots$, $\lambda_i > 0$, be the probability mass functions of M_i for $i = 1, 2$. Independence of M_1 and M_2 gives

$$\mathbb{E}[Y] = \sum_{m_1, m_2 = 0}^{\infty} \min(m_1, m_2) P(m_1; \lambda_1) P(m_2; \lambda_2)$$

(A.8)

$$= \sum_{m_2=0}^{\infty} \sum_{m_1=0}^{m_2} m_1 P_1 P_2 + \sum_{m_2=0}^{\infty} \sum_{m_1=m_2+1}^{\infty} m_2 P_1 P_2.$$

Denote the first and second sums in the last row of (A.8) by κ_1 and κ_2, respectively, and note that

$$\kappa_1 = \sum_{m_2=0}^{\infty} P_2 \sum_{m_1=0}^{m_2} m_1 P_1$$

$$= \sum_{m_2=0}^{\infty} e^{-\lambda_2} \frac{\lambda_2^{m_2}}{m_2!} \sum_{m_1=0}^{m_2} m_1 e^{-\lambda_1} \frac{\lambda_1^{m_1}}{m_1!}$$

$$= \lambda_1 \sum_{m_2=0}^{\infty} e^{-\lambda_2} \frac{\lambda_2^{m_2}}{m_2!} \sum_{m_1=0}^{m_2-1} e^{-\lambda_1} \frac{\lambda_1^{m_1}}{m_1!}$$

$$= \lambda_1 \sum_{m_2=0}^{\infty} e^{-\lambda_2} \frac{\lambda_2^{m_2}}{m_2!} \left(1 - \sum_{m_1=m_2}^{\infty} e^{-\lambda_1} \frac{\lambda_1^{m_1}}{m_1!}\right)$$

(A.9)

$$= \lambda_1 \left(1 - \sum_{m_2=0}^{\infty} e^{-\lambda_2} \frac{\lambda_2^{m_2}}{m_2!} \sum_{m_1=m_2}^{\infty} e^{-\lambda_1} \frac{\lambda_1^{m_1}}{m_1!}\right)$$

$$= \lambda_1 \left(1 - \sum_{m_1, m_2=0}^{\infty} 1_{\{m_1 \geq m_2\}} P_1 P_2\right)$$

$$= \lambda_1 [1 - P(M_1 \geq M_2)]$$

$$= \lambda_1 F_W(-1).$$

Similar arguments give $\kappa_2 = \lambda_2 [1 - F_W(1)]$, thus proving (A.7).   □

Acknowledgments.   The authors would like to thank the Editor and referee for insightful comments which improved the paper considerably.

REFERENCES

ALZAID, A. A. and AL-OSH, M. (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* **27** 314–324. MR1052303
BARNDORFF-NIELSEN, O. E., LUNDE, A., SHEPHARD, N. and VERAART, A. E. D. (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scand. J. Stat.* **41** 693–724. MR3249424

BERAN, J., FENG, Y., GHOSH, S. and KULIK, R. (2013). *Long-Memory Processes*. Springer, Heidelberg.

BLIGHT, P. A. (1989). Time series formed from the superposition of discrete renewal processes. *J. Appl. Probab.* **26** 189–195. MR0981265

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series*: *Theory and Methods*. Springer, New York. Reprint of the second (1991) edition. MR2839251

CHU, P. S. and ZHAO, Z. (2004). Bayesian change-point analysis of tropical cyclone activity: The Central North Pacific case. *J. Climate* **17** 4893–4901.

CUI, Y. and LUND, R. (2009). A new look at time series of counts. *Biometrika* **96** 781–792.

CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261.

DAVIS, R. A. and DUNSMUIR, W. T. M. (2016). State space models for count time series. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. Holan, R. B. Lund and N. Ravishanker, eds.) CRC Press, Boca Raton, FL.

DAVIS, R. A., HOLAN, S. H., LUND, R. B. and RAVISHANKER, N., eds. (2016). *Handbook of Discrete-Valued Time Series*. CRC Press, Boca Raton.

DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston, MA.

DUNSMUIR, W. T. M. (2016). Generalized linear autoregressive moving average models. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. Holan, R. B. Lund and R. N. Ravishanker, eds.) CRC Press, Boca Raton, FL.

ELSNER, J. and JAGGER, T. H. (2006). Prediction models for annual US hurricane counts. *J. Climate* **19** 2935–2952.

ELSNER, J., JAGGER, T. and NIU, X. F. (2000). Changes in the rates of North Atlantic major hurricane activity during the 20th century. *Geophys. Res. Lett.* **27** 1743–1746.

ELSNER, J. B. and KOCHER, B. (2000). Global tropical cyclone activity: A link to the North Atlantic Oscillation. *Geophys. Res. Lett.* **27** 129–132.

ELSNER, J., KOSSIN, J. P. and JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455** 92–95.

ENCISO-MORA, V., NEAL, P. and SUBBA RAO, T. (2009). Efficient order selection for integer-valued ARMA processes. *J. Time Series Anal.* **30** 1–18.

FOKIANOS, K. and KEDEM, B. (2003). Regression theory for categorical time series. *Statist. Sci.* **18** 357–376. MR2061915

GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

GOLDENBERG, S. B., LANDSEA, C. W. and MESTAS-NUNEZ, A. M. (2001). The recent increase in Atlantic hurricane activity: Causes and implications. *Science* **293** 474–479.

GRAY, W. M. (1984). Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb quasi-biennial oscillation influences. *Mon. Weather Rev.* **112** 1649–1668.

HASSELMANN, K. (1976). Stochastic climate models part I. Theory. *Tellus* **28** 473–485.

HELGASON, H., PIPIRAS, V. and ABRY, P. (2011). Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding. *Signal Process.* **91** 1123–1133.

JIA, Y. and LUND, R. B. (2016). Superpositioned stationary count time series. *J. Appl. Probab.*. To appear.

JOE, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* **33** 664–677. MR1401464

KACHOUR, M. and YAO, J. F. (2009). First-order rounded integer-valued autoregressive (RINAR(1)) process. *J. Time Series Anal.* **30** 417–448. MR2536061

KARLIS, D. (2016). Models for multivariate count time series. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. Holan, R. B. Lund and N. Ravishanker, eds.) 407–424. CRC Press, Boca Raton, FL.

KARLIS, D. and MELIGKOTSIDOU, L. (2007). Finite mixtures of multivariate Poisson distributions with application. *J. Statist. Plann. Inference* **137** 1942–1960.

KECHAGIAS, S. and PIPIRAS, V. (2015). Definitions and representations of multivariate long-range dependent time series. *J. Time Series Anal.* **36** 1–25. MR3300202

KECHAGIAS, S. and PIPIRAS, V. (2017). Identification, estimation and applications of a bivariate long-range dependent time series model with general phase. Preprint.

KERSS, A., LEONENKO, N. and SIKORSKII, A. (2014). Fractional Skellam processes with applications to finance. *Fract. Calc. Appl. Anal.* **17** 532–551.

KOLASSA, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *Int. J. Forecast.* **32** 788–803.

LUND, R. B., HOLAN, S. H. and LIVSEY, J. (2016). Long memory discrete-valued time series. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. Holan, R. B. Lund and N. Ravishanker, eds.) 447–458. CRC Press, Boca Raton, FL.

LUND, R. B. and LIVSEY, J. (2016). Renewal-based count time series. In *Handbook of Discrete-Valued Time Series* (R. A. Davis, S. Holan, R. B. Lund and N. Ravishanker, eds.) 101. CRC Press, Boca Raton.

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368

MACDONALD, I. L. and ZUCCHINI, W. (2016). Hidden Markov models for discrete-valued time series. In *Handbook of Discrete-Valued Time Series* (R. Davis, S. Holan, R. B. Lund and N. Ravishanker, eds.) CRC Press, Boca Raton, FL.

MCDONNELL, K. A. and HOLBROOK, N. J. (2004). A Poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. *Weather Forecast.* **19** 440–455.

MCKENZIE, E. (2003). Discrete variate time series. In *Stochastic Processes: Modelling and Simulation. Handbook of Statistics* **21** 573–606. North-Holland, Amsterdam.

MOOLEY, D. A. (1981). Applicability of the Poisson probability model to severe cyclonic storms striking the cost around the bay of bengal. *Sankyha* **43** 187–197.

MOONEY, C. C. (2007). *Storm World. Hurricanes, Politics, and the Battle over Global Warming*. Harcourt, New York.

MUDELSEE, M. (2013). *Climate Time Series Analysis*. Springer, Berlin.

NEAL, P. and SUBBA RAO, T. (2007). MCMC for integer-valued ARMA processes. *J. Time Series Anal.* **28** 92–110. MR2332852

PALMA, W. (2007). *Long-Memory Time Series*. Wiley, New Jersey.

PARISI, F. and LUND, R. B. (2000). Seasonality and return periods of landfalling Atlantic basin hurricanes. *Aust. N. Z. J. Stat.* **42** 271–282.

PARK, K. and WILLINGER, W. (2000). *Self-Similar Network Traffic and Performance Evaluation*. Wiley Online Library.

PERCIVAL, D. B., OVERLAND, J. E. and MOFJELD, H. O. (2001). Interpretation of North Pacific variability as a short-and long-memory process. *J. Climate* **14** 4545–4559.

PIPIRAS, V. and TAQQU, M. S. (2017). *Long-Range Dependence and Self-Similarity* **45**. Cambridge Univ. Press, Cambridge.

QUORESHI, A. M. M. S. (2014). A long-memory integer-valued time series model, INARFIMA, for financial application. *Quant. Finance* **14** 2225–2235.

ROBBINS, M. W., LUND, R. B., GALLAGHER, C. M. and LU, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *J. Amer. Statist. Assoc.* **106** 89–99. MR2816704

ROBINSON, P. M. (2003). Long-memory time series. In *Time Series with Long Memory. Adv. Texts Econometrics* 4–32. Oxford Univ. Press, Oxford. MR2077520

SELA, R. J. (2010). Three essays in econometrics: Multivariate long memory time series and applying regression trees to longitudinal data. Ph.D. thesis, New York Univ.

SKELLAM, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. R. Stat. Soc.* **109** 296.

SOLOW, A. (1989). Statistical modeling of storm counts. *J. Climate* **2** 131–136.

STEUTEL, F. W. and VAN HARN, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7** 893–899.

THOMPSON, M. L. and GUTTORP, P. (1986). A probability model for severe cyclonic storms striking the cost around the bay of bengal. *Mon. Weather Rev.* **114** 2267–2271.

VAN VLECK, J. H. and MIDDLETON, D. (1966). The spectrum of clipped noise. *Proc. IEEE* **54** 2–19.

VAROTSOS, C. and EFSTATHIOU, M. N. (2013). Is there any long-term memory effect in the tropical cyclones? *Theor. Appl. Climatol.* **114** 643–650.

VILLARINI, G., VECCHI, G. A. and SMITH, J. A. (2010). Modeling the dependence of tropical storm counts in the North Atlantic basin on climate indices. *Mon. Weather Rev.* **138** 2681–2705.

XIAO, S., KOTTAS, A. and SANSÓ, B. (2015). Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *Ann. Appl. Stat.* **9** 353–382. MR3341119

YUAN, N., FU, Z. and LIU, S. (2014). Extracting climate memory using fractional integrated statistical model: A new perspective on climate prediction. *Sci. Rep.* **4**.

J. LIVSEY
CENTER FOR STATISTICAL RESEARCH
  AND METHODOLOGY
UNITED STATES CENSUS BUREAU
WASHINGTON, DC 20233
USA
E-MAIL: james.a.livsey@census.gov

R. LUND
DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
CLEMSON, SOUTH CAROLINA 29634-0975
USA
E-MAIL: lund@clemson.edu

S. KECHAGIAS
ADVANCED ANALYTICS AND RESEARCH DIVISION
SAS INSTITUTE
100 SAS CAMPUS DR.
CARY, NORTH CAROLINA 27513
USA
E-MAIL: stefanos.kechagias@sas.com

V. PIPIRAS
DEPARTMENT OF STATISTICS
  AND OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
  AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514
USA
E-MAIL: pipiras@email.unc.edu