

Stochastic modeling of regime shifts

Franco Biondi^{1,*}, Tomasz J. Kozubowski², Anna K. Panorska²

¹Department of Geography, and ²Department of Mathematics, University of Nevada, Reno, Nevada 89557, USA

ABSTRACT: Probabilistic methods for modeling the distribution of regimes and their shifts over time are developed by drawing on statistical decision and limit theory of random sums. Multi-annual episodes of opposite sign are graphically and numerically represented by their duration, magnitude, and intensity. Duration is defined as the number of consecutive years above or below a reference line, magnitude is the sum of time series values for any given duration, and intensity is the ratio between magnitude and duration. Assuming that a regime shift can occur every year, independently of prior years, the waiting times for the regime shift (or regime duration) are naturally modeled by a geometric distribution. Because magnitude can be expressed as a random sum of N random variables (where N is duration), its probability distribution is mathematically derived and can be statistically tested. Here we analyze a reconstructed time series of the Pacific Decadal Oscillation (PDO), explicitly describe the geometric, exponential, and Laplace probability distributions for regime duration and magnitude, and estimate parameters from the data obtaining a reasonably good fit. This stochastic approach to modeling duration and magnitude of multi-annual events enables the computation of probabilities of climatic episodes, and it provides a rigorous solution to deciding whether 2 regimes are significantly different from one another.

KEY WORDS: Climatological probabilities · Environmental change · Proxy records · Pacific Decadal Oscillation · Random sums

—Resale or republication not permitted without written consent of the publisher—

1. INTRODUCTION

Visual identification of particularly relevant episodes in a climatological time series is a quite common and often useful practice in scientific research and literature. For instance, the plot of an environmental parameter (temperature, moisture, wildfire frequency, tree-ring indices, pollen percentages, etc.) over time is often used as a starting point to describe which historical periods were most remarkable (warmer, cooler, wetter, drier, with more fires, with fewer fires, etc.). Such classification of multi-annual events is usually subjective, and our efforts began as an attempt at ranking climatic episodes. We then proceeded to answer questions related to the probability of observing episodes of a given size and/or of one episode being 'more remarkable' than another. Although we study temporal processes, time series analysis methods do not provide

solutions to our inquiries, as they help with other important issues, for example, deciding whether a process has changed significantly over time in terms of its mean or covariance structure (e.g. de Jong & Penzer 1998, Harvey 1990, and references therein). Our methodology does not detect change in a time series, because the change is explicitly identified. We *define* regimes (or episodes) as time periods being continuously above or below a reference line. In other words, we consider the time series process as fluctuating around a constant level, and the covariance structure of the process is irrelevant to our model. Our problem and approach are similar to those used in hydrology to summarize drought properties (see Dracup et al. 1980, Sedeghipour & Dracup 1985), as discussed in Section 2 below.

The method we propose is an application of the limits of random sums that is capable of yielding probabil-

*Email: fbiondi@unr.edu

ity statements on regimes and regime shifts even for heavy-tailed processes, e.g. processes with finite mean but without finite variance. In fact, random sums have been successfully applied to modeling financial asset returns, which exhibit high volatility (Mittnik & Rachev 1993, Rachev & SenGupta 1993, Kozubowski & Rachev 1994, Kozubowski & Panorska 1999, Kozubowski & Podgórski 2001). Other extensive applications of random sums are in reliability and queuing theories (e.g. Kalashnikov 1997). We begin by quantifying multi-annual time series events in terms of 3 random variables (parameters). First, *duration* is the number of years the series remains continuously above (or below) its reference line. Second, *magnitude* is the sum of all series values for a given duration; hence it is equivalent to the area under (or above) the curve. Third, *intensity* is the ratio between magnitude and duration; therefore it is equal to the average magnitude. Computation of the 3 parameters for every multi-annual event is straightforward, and it allows for objective ranking of the events themselves. For instance, the event with the longest duration or the largest magnitude (absolute or positive or negative) or the highest intensity can be easily identified (Fig. 1). At the same time, comparisons between events can be carried out using any 1 (or a combination) of the 3 numerical parameters. A useful classification scheme is based on ranking events according to duration, then to magnitude, and then to intensity. The sum of the 3 ranks assigned to an event provides a numerical score, and the entire classification scheme offers a quantitative solution to the need for identifying the 'strongest', 'greatest', or 'most remarkable' periods, which would otherwise be selected differently by different observers. An example of the procedure is given below, and summarized

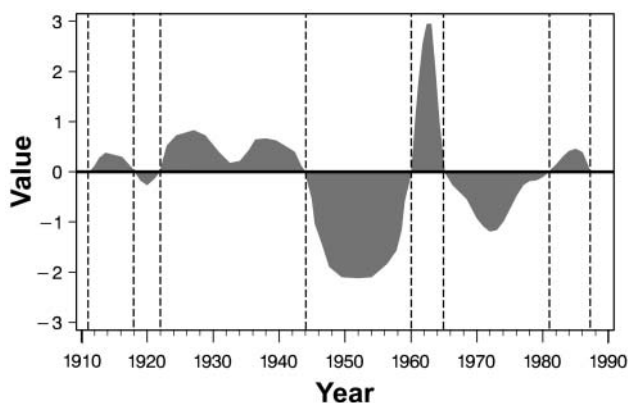


Fig. 1. Time-series plot showing the duration (dashed lines) and magnitude (solid areas) of multi-annual events in a fictitious annual record: 1922-1944 has the longest duration, 1944-1960 the largest magnitude, and 1960-1965 the highest intensity

in Table 1. It should be noted that our procedure extends runs analysis, which has been widely applied in water resources research to describe drought properties. Dracup et al. (1980) provide a clear description of how the truncation level (what we call a reference line) is chosen. They also explain in detail the meaning and computation of drought duration, severity (what we call magnitude), and magnitude (what we call intensity).¹

2. STOCHASTIC MODEL

While it is useful to quantify the relative importance of multi-annual events in a time series, a stochastic framework was required to answer questions about the probability of occurrence and statistical significance of differences between regimes. In particular, our objectives are to provide answers to the following questions: (1) What is the probability of an episode being larger than a given amount? (2) Is the difference between 2 episodes statistically significant? (3) Do positive episodes behave differently from negative ones? Because our questions are probabilistic, they can only be answered using a distributional model for the duration and magnitude of episodes. As previously mentioned, we note that common time series methods (including cusum-like tests) do not provide such models. We thus answered our questions by means of a comprehensive and 'natural' (as being derived naturally from the definition of regime duration, magnitude, and intensity) stochastic model of the distribution of episodes and regime shifts observed through time.

First, assuming that at each time step (e.g. a year) the regime can shift independently of prior time steps with a time-invariant probability, p , the waiting time for the shift (i.e. the event duration) is naturally modeled by a geometric random variable, N , with the following probability function:

$$P(N=x) = p(1-p)^{x-1} \quad x=1,2,\dots$$

Magnitude can then be expressed as a random sum of N random variables $\sum_{i=1}^N X_i$, where N is the duration of an event and X_i are the series values. It follows from the limit theory for random sums (Brown 1990, Kalashnikov 1997, Kozubowski & Panorska 1998) that if the number of terms in the summation, N , follows a geometric distribution with mean $1/p$, then under certain conditions when the number of terms increases ($p \rightarrow 0$),

¹Our terminology applies to any type of negative or positive episodes, whereas the term 'severity' entails a negative connotation that is appropriate for drought, but not necessarily for other phenomena

the distribution of the appropriately scaled random sum converges to an exponential law with the following probability density function:

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \quad x > 0$$

where μ is the expected value. For example, if the summands X_i are independent of N and satisfy the weak law of large numbers (WLLN), i.e.

$$\frac{\sum_{i=1}^n X_i}{n} \Rightarrow \mu \quad (\text{in probability}) \quad \text{for } n \rightarrow \infty$$

then the scaled random sum $p(\sum_{i=1}^N X_i)$ converges to an exponential random variable with mean μ as $p \rightarrow 0$. Consequently, for small values of p , the sum $\sum_{i=1}^N X_i$ can be approximated by an exponential distribution, as is the case when X_i are *iid* variables with finite mean. It should be noted, however, that we are only making the assumption that the WLLN holds, which is weaker than assuming that the summands are *iid* random variables. In Appendix 1, we provide a formal proof of the above convergence result and a discussion of assumptions under which the exponential approximation holds. To summarize, the exponential distribution is the natural model for regime magnitude. A caveat exists for any limit-based approximation technique on the rate of convergence that provides an estimate of the distance between the approximated quantity and the approximation (limit). In our case, this translates into the following question: How small does p have to be for the exponential approximation of the random sum to be reasonable? Like any other rate of convergence problem, this is a difficult mathematical question, and there are few results and only for special cases (Brown 1990, Kalashnikov 1997; see Appendix 1 for additional details). In practice, it is recommended to employ goodness-of-fit statistics for the exponential match to the distribution of regime magnitude. Furthermore, magnitude differences can then be statistically described and thus rigorously tested for significance because the difference between 2 independent, exponential random variables has a symmetric or asymmetric Laplace distribution (Kotz et al. 2001) depending on whether the means of the exponential distributions are the same or different. The probability density function of the Laplace distribution is:

$$f(x; \sigma, k) = \frac{1}{\sigma} \frac{k}{1+k^2} \begin{cases} e^{-\frac{k}{\sigma}|x|} & \text{for } x \geq 0 \\ e^{\frac{-1}{k\sigma}|x|} & \text{for } x < 0 \end{cases} \quad \sigma > 0, k > 0$$

where σ and k are scale and skewness parameters, respectively. When $k = 1$, we have the classical symmetric Laplace distribution with mean 0 and variance $2\sigma^2$. The likelihood ratio test for:

$$H_0: k = 1 \text{ versus } H_1: k \neq 1$$

rejects the null hypothesis if the value of the statistic S given by:

$$S(x_1, \dots, x_n) = n \left(2 - \frac{\left(\sqrt{\sum_{j=1}^n \max\{x_j, 0\}} + \sqrt{\sum_{j=1}^n \max\{-x_j, 0\}} \right)^2}{\sum_{j=1}^n \max\{x_j, 0\} + \sum_{j=1}^n \max\{-x_j, 0\}} \right)$$

is greater than some constant c appropriately chosen to obtain the desired level of significance. To compute p -values or critical regions for the test when the sample size n is large, we use the fact that the limiting distribution of S under the null hypothesis is χ_1^2 , and for small n we use its exact distribution, as shown in Kozubowski & Panorska (2002). To date we have focused on duration and magnitude because intensity, as defined within our framework, converges to a constant random variable.

While related to drought studies (Dracup et al. 1980, Sedeghipour & Dracup 1985, Touchan et al. 1999), our models provide a considerable improvement and expansion of those techniques whose main objective is the estimation of a drought frequency curve, usually obtained by runs analysis combined with numerical (Monte Carlo) simulation. Such an approach does not provide a probabilistic comparison between episodes (for instance, droughts) of different duration, because it cannot answer distributional questions on the magnitude of a drought. An explicit call for stochastic modeling of drought magnitude is present in Sedeghipour & Dracup (1985), because by modeling magnitude one can compare episodes of different duration. To compare droughts of different duration, Sedeghipour & Dracup (1985) proposed using ‘standardized droughts’ but they conceded that exceedance probabilities obtained from their method could be improved by assigning a probability distribution to the drought record—a distribution they had not yet discovered. Our methodology provides a unified, probabilistic model for both duration and magnitude of multi-annual events that are either negative (such as dry spells) or positive (such as wet spells).

3. APPLICATION TO THE PACIFIC DECADAL OSCILLATION

The ranking system and stochastic framework described above are applied here to a reconstructed time series of the Pacific Decadal Oscillation (PDO; Fig. 2). The PDO time series consists of a total of 48 multi-annual events, 24 above and 24 below the median reference line (Table 1). The total event score, computed

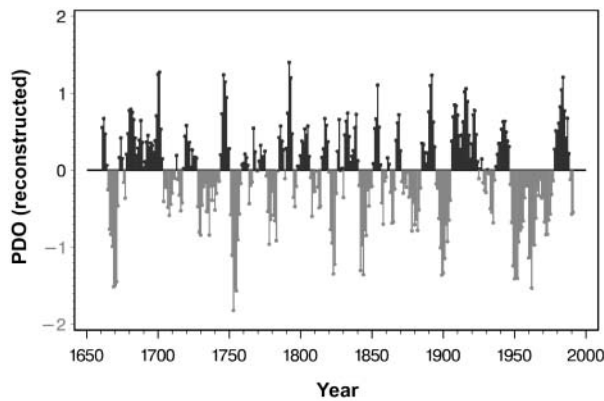


Fig. 2. Annual values of the reconstructed Pacific Decadal Oscillation (PDO) index from 1661 to 1991 (Biondi et al. 2001) plotted as deviations from the overall median (0.138). Positive (warm) and negative (cold) episodes were numerically ranked and statistically analyzed

as the sum of ranks (in ascending order) based on duration, absolute magnitude, and absolute intensity, shows that the most remarkable event was 1947–1977, followed by 1895–1905 and 1906–1924. The 1678–1703 event, which could have been subjectively considered as the greatest one, is in fourth position. The 3 strongest episodes occurred in the 20th century, confirming previous results showing that the 1900s were anomalous compared to the previous $2\frac{1}{2}$ centuries (Biondi et al. 2001).

To compute episode probabilities and assess statistical significance of differences between PDO events using our stochastic model, we first tested the goodness-of-fit of the model distributions to the PDO data. The fit of geometric distributions to PDO durations was tested separately for positive and negative events using the chi-square goodness-of-fit test (D'Agostino & Stephens 1986). The test could not reject the null hypothesis that the distributions of durations are in fact geometric (both p-values > 0.2; Fig. 3). This result allowed us to compute climatological probabilities of event duration. For instance, the longest event in the PDO series of Fig. 2 spanned 31 yr (1947–1977), and the probability of exceeding such a duration is 0.0076. In other words, the chance that a PDO event should last longer than 31 yr is very small (below 1%). Such information is useful in forecasting or computing the likelihood of a regime shift.

We then turned our attention to the event magnitude, and again we started by testing how well the exponential distribution represents the magnitude of positive and negative events. We assume that the number of positive events is equal to the number of negative events, as is the case for the PDO (Table 1). Their magnitudes can be formally described as 2

series, 1 for positive events (M_1^+, \dots, M_n^+) and 1 for negative events (M_1^-, \dots, M_n^-), with n equal to 24 in the PDO time series. An assumption implicitly included in our approach is that the positive and negative magnitudes can be treated as sets of independent observations. The p-values (≥ 0.1) of the autocorrelation and partial autocorrelation functions (Box &

Table 1. Multi-annual Pacific Decadal Oscillation (PDO) events. The 3 strongest episodes (given in **bold**) occurred during the 20th century

Start	End	Duration	Magnitude	Intensity	Score
1661	1664	4	1.76	0.44	71
1665	1672	8	-7.75	-0.97	24
1673	1675	3	0.74	0.25	108
1676	1677	2	-0.52	-0.26	115
1678	1703	26	12.37	0.48	19
1704	1712	9	-2.86	-0.32	62
1713	1713	1	0.19	0.19	127
1714	1717	4	-1.39	-0.35	83
1718	1727	10	2.50	0.25	65
1728	1743	16	-6.76	-0.42	30
1744	1750	7	4.81	0.69	38
1751	1758	8	-8.24	-1.03	22
1759	1763	5	0.59	0.12	104
1764	1766	3	-0.79	-0.26	102
1767	1769	3	0.78	0.26	104
1770	1770	1	-0.01	-0.01	137
1771	1775	5	1.03	0.21	95
1776	1784	9	-4.65	-0.52	38
1785	1788	4	1.64	0.41	77
1789	1789	1	-0.11	-0.11	132
1790	1794	5	4.09	0.82	41
1795	1797	3	-1.03	-0.34	94
1798	1806	9	2.72	0.30	65
1807	1814	8	-2.42	-0.30	69
1815	1819	5	2.06	0.41	65
1820	1825	6	-4.61	-0.77	39
1826	1829	4	0.93	0.23	100
1830	1830	1	-0.35	-0.35	110
1831	1840	10	4.20	0.42	42
1841	1850	10	-6.59	-0.66	25
1851	1856	6	3.03	0.50	51
1857	1860	4	-1.35	-0.34	86
1861	1862	2	0.24	0.12	124
1863	1866	4	-1.89	-0.47	69
1867	1870	4	1.97	0.49	65
1871	1884	14	-5.60	-0.40	38
1885	1894	10	5.22	0.52	30
1895	1905	11	-8.55	-0.78	14
1906	1924	19	11.23	0.59	16
1925	1925	1	-0.11	-0.11	134
1926	1927	2	0.16	0.08	130
1928	1930	3	-0.71	-0.24	110
1931	1931	1	0.01	0.01	139
1932	1936	5	-1.90	-0.38	72
1937	1946	10	3.92	0.39	49
1947	1977	31	-21.89	-0.71	8
1978	1988	11	7.03	0.64	22
1989	1991	3	-1.24	-0.41	86

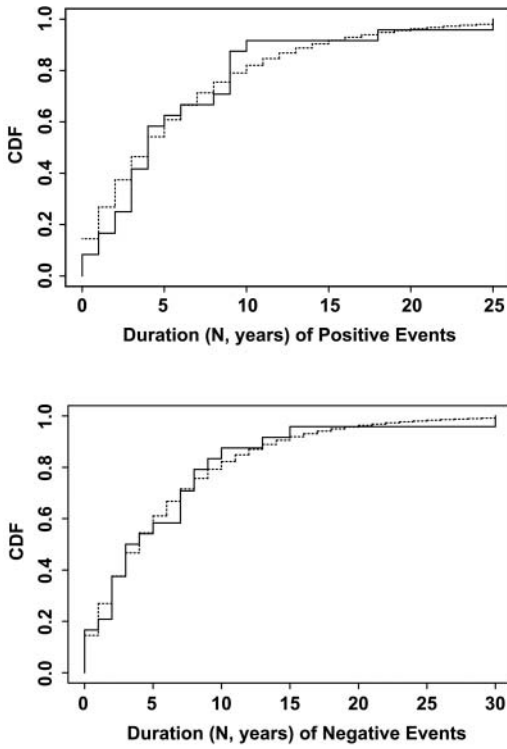


Fig. 3. Empirical (solid line) and hypothesized (dashed line) geometric cumulative density function (CDF) of duration for positive and negative PDO episodes

Jenkins 1976), and the minimum value of the Akaike information criterion (Akaike 1974), support the independence assumption. The goodness-of-fit between the magnitude series M_1^+, \dots, M_n^+ or M_1^-, \dots, M_n^- and the exponential distribution (Fig. 4) was tested using the Kolmogorov-Smirnov test (D’Agostino & Stephens 1986). The test results could not reject the null hypothesis that both samples follow exponential distributions (both p-values > 0.3). Computing climatological probabilities was deferred until we tested the positive and negative magnitude series for significant differences in their distributions, which was the last step in fitting our model to the PDO series. To decide whether the magnitude series follow a common or different exponential distribution is equivalent to testing their Laplace distributed difference for symmetry. The test for symmetry described in the previous section assumes that we have 2 independent exponentially distributed samples. We checked that assumption using the Pearson, Spearman, and Kendall correlation coefficients (Conover 1980), which were not significant (p-values > 0.1). The lack of dependence between positive and negative magnitudes was also supported by their scatterplot (Fig. 5). The symmetry test yielded an S statistic for the PDO magnitudes of 0.52 with p-value of 0.48. Because the difference

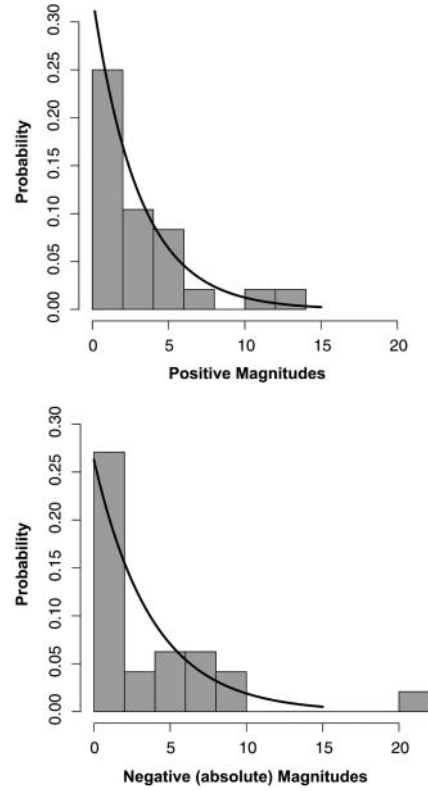


Fig. 4. Probability histogram and fitted exponential distribution for positive and negative magnitudes of PDO episodes

between PDO magnitudes can be modeled as a symmetric Laplace distribution (Fig. 6), the positive and (absolute value of) negative magnitudes are modeled by a common exponential distribution. An estimate of the common mean magnitude is given by the average of the mean positive (3.05) and negative (3.81 in absolute value) magnitudes, equal to 3.43. As for duration, the distributional models for magnitudes

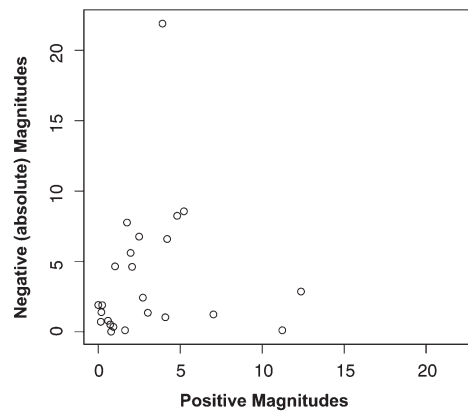


Fig. 5. Scatterplot of negative vs positive magnitude of PDO episodes, ordered according to time

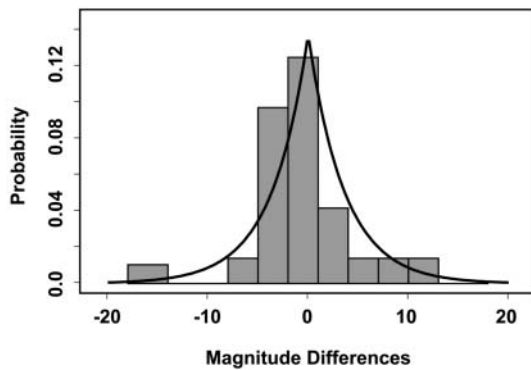


Fig. 6. Probability histogram and fitted symmetric Laplace distribution for the difference between positive and negative magnitudes of PDO episodes

enable estimation of climatological probabilities related to the magnitude of PDO events. For instance, the 1947–1977 event had the greatest magnitude (21.89 in absolute value), and the chance of exceeding such a value is minuscule (0.17%). Finally, we could estimate climatological probabilities of any size difference between 2 multi-annual PDO events. For example, the chance that the difference between the magnitudes of 2 events exceeds the observed difference between the (absolute) magnitudes of the 1947–1977 and 1906–1924 events is rather small (2.2%).

An issue that arises at this point is whether the results of our analysis are sensitive to the choice of reference line. Indeed, our approach shares a known shortcoming of other runs analyses of temporal data, i.e. a single observation crossing the reference line can turn a long episode into 3 shorter episodes. Although we do not offer a remedy, we investigated the following question: How different would our models be if we used a mean or zero reference line? We considered both cases and obtained the same models, with very similar parameters, as those detailed above for the median reference line. The reason for this empirical consistency is perhaps more fundamental than chance alone. Choosing a reference line, albeit somewhat arbitrary, is nevertheless dictated by the nature of the process and the scientific questions that need to be answered. Therefore, suitable reference lines are likely to be close to one another, as was the case for the PDO data (either 0 or the mean 0.083 or the median 0.138). Since the statistical methods we employed are fairly robust, our stochastic analysis yielded consistent results for similar reference lines.

It should also be emphasized that error affects any observation and/or reconstruction of an environmental process. However, our focus is on stochastic mod-

eling of what was actually observed/reconstructed. We cannot know whether the true process consists of a lot of short-lived episodes and frequent regime changes instead of a few long episodes tainted by noise. Our model is applicable to (and can be tested on) both scenarios, as long as the regime duration is compatible with the limit theory of random sums. If the investigator so desires, s/he has the option to smooth the data before applying the models presented here. Any smoothing or other manipulation of the data will reflect the belief of the investigator and may be dictated by the needs of the particular question being addressed. We decided to minimize data pre-processing, and let the models speak for themselves.

4. CONCLUSION

We have described an innovative and practical approach to the analysis of regime shifts based on a minimum set of parameters that fully describe multi-annual events. By applying statistical methods for random sums and their limits to duration and magnitude of climatic episodes, we provide a rigorous approach to the estimation of climatological probabilities, such as the likelihood of a regime exceeding any given duration. These models enable us to quantify the probability of occurrence and statistical significance of multi-annual events, and they help us decide whether 2 episodes are significantly different from one another. Testing whether the distributional properties of magnitude and duration for a time series are different from those of another time series can also be used to test for significant differences between time series of the same parameter, such as different PDO time series, either instrumental (Mantua et al. 1997) or reconstructed. The stochastic framework described here can be applied to any time series, including heavy-tailed processes, that has a reference line and can be expressed as a number of events of known duration and magnitude. Future extensions of this work could include the relaxation of independence between N and X_i or the removal of the assumption that the WLLN holds at all, which happens when X_i have an infinite mean. Further work could also extend the model to the analysis of 2 or (possibly) more time series at once.

Acknowledgements. F.B. was supported by the Paleoclimate Program of the National Science Foundation, under Grant Nos. ATM-0096316 and CAREER/ATM-0132631. We thank 3 anonymous reviewers for their helpful comments and suggestions.

Appendix 1. Mathematical supplement

Since the proof of convergence under the assumption of the weak law of large numbers (WLLN) is not easily available in the literature, we provide it below. Our results are more general than those mentioned without proof by Brown (1990) for the special case of geometric convolution.

Lemma. Let X_1, X_2, \dots be a sequence of random variables satisfying a WLLN,

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{d} \mu \text{ as } n \rightarrow \infty \quad (\text{A1})$$

(where \xrightarrow{d} stands for convergence in distribution). Assume that (N_n) is a sequence of integer-valued random variables independent of the X_i 's and such that as $n \rightarrow \infty$

$$N_n \xrightarrow{p} \infty \text{ (in probability) and } \frac{N_n}{n} \xrightarrow{d} Z \quad (\text{A2})$$

where Z is a non-negative random variable. Then,

$$\frac{1}{n} \sum_{i=1}^{N_n} X_i \xrightarrow{d} \mu Z \text{ as } n \rightarrow \infty \quad (\text{A3})$$

Proof. First, we show that

$$\frac{1}{N_n} \sum_{i=1}^{N_n} X_i \xrightarrow{p} \mu \text{ (in probability)} \quad (\text{A4})$$

that is for each $\varepsilon > 0$ and each $\eta > 0$ there exists some integer k such that

$$P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon \right) < \eta \text{ whenever } n > k \quad (\text{A5})$$

Let $\varepsilon > 0$ and $\eta > 0$. By Eq. (A1), there exists an integer N_0 such that

$$P_r \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) < \frac{\eta}{2} \text{ for all } n > N_0 \quad (\text{A6})$$

Since $N_n \xrightarrow{p} \infty$, for the above N_0 there exists an integer N_1 such that for each $n > N_1$ we will have

$$P_r(N_n \leq N_0) < \frac{\eta}{2} \quad (\text{A7})$$

Let now $k = \max\{N_0, N_1\}$. Then, for $n > k$, we will have

$$\begin{aligned} P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon \right) &= P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon, N_n \leq N_0 \right) \\ &\quad + P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon, N_n > N_0 \right) \\ &= I + II \end{aligned} \quad (\text{A8})$$

Note that by Eq. (A7) we have

$$I = P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon, N_n \leq N_0 \right) \leq P_r(N_n \leq N_0) < \frac{\eta}{2} \quad (\text{A9})$$

On the other hand, utilizing the independence of the N_n 's and the X_i 's, we obtain

$$\begin{aligned} II &= P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon, N_n > N_0 \right) \\ &= \sum_{j=N_0+1}^{\infty} P_r \left(\left| \frac{1}{N_n} \sum_{i=1}^{N_n} X_i - \mu \right| > \varepsilon, N_n = j \right) \\ &= \sum_{j=N_0+1}^{\infty} P_r \left(\left| \frac{1}{j} \sum_{i=1}^j X_i - \mu \right| > \varepsilon, N_n = j \right) \\ &= \sum_{j=N_0+1}^{\infty} P_r \left(\left| \frac{1}{j} \sum_{i=1}^j X_i - \mu \right| > \varepsilon \right) P_r(N_n = j) < \sum_{j=N_0+1}^{\infty} \frac{\eta}{2} P_r(N_n = j) \leq \frac{\eta}{2} \end{aligned} \quad (\text{A10})$$

where the above inequalities hold because of Eq. (A6). This proves Eq. (A5). Next, by virtue of Eqs. (A2) & (A4), as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{i=1}^{N_n} X_i = \frac{N_n}{n} \frac{1}{N_n} \sum_{i=1}^{N_n} X_i \xrightarrow{p} \mu Z \text{ (in probability)} \quad (\text{A11})$$

and the result has been proven.

Note that the above result holds when the X_i 's are *iid* variables with finite mean μ , although it does not require that the X_i 's be *iid*. For instance, the result holds if the X_i 's form a stationary sequence with autocovariance function converging to zero, since such sequence satisfies the WLLN (e.g. Brockwell & Davis 1991). Let us also note that the condition of independence of the N_n 's and the X_i 's can be relaxed for geometric sums if the strong law of large numbers holds (Brown 1990).

With regard to the rate of convergence, available theorems use multiple ways of measuring closeness of 2 distributions, and results are often difficult to use in practice. Following Brown (1990), let X_i 's be positive *iid* variables with finite mean μ_1 and finite second moment $EX_i^2 = \mu_2$, and let N be a geometric random variable with mean $1/p$ and independent of the X_i 's. Then, the error bound can be stated in terms of the distance between the cumulative distribution functions of the random sum $Y = \sum_{i=1}^N X_i$ and its limiting exponential distribution Z , as follows:

$$|P(Y \leq t) - P(Z \leq t)| \leq (p/q) \max \left\{ 1 + \frac{\mu_2}{2\mu_1^2} \frac{1}{q}, \frac{\mu_2}{\mu_1^2} - 1 \right\}$$

where $q = 1 - p$.

LITERATURE CITED

- Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automatic Control AC-19:716–723
- Biondi F, Gershunov A, Cayan DR (2001) North Pacific decadal climate variability since AD 1661. J Clim 14:5–10
- Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control, revised edn. Holden-Day, Oakland
- Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer-Verlag, New York

- Brown DM (1990) Error bounds for exponential approximations of geometric convolutions. Ann Prob 18:1388–1402
- Conover WJ (1980) Practical nonparametric statistics, 2nd edn. John Wiley, New York
- D'Agostino RB, Stephens MA (1986) Goodness-of-fit techniques. Marcel Dekker, New York
- de Jong P, Penzer J (1998) Diagnosing shocks in time series. J Am Stat Assoc 93:796–806
- Dracup JA, Lee KS, Paulson EG (1980) On the definition of droughts. Water Resour Res 16:297–302

- Harvey AC (1990) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge
- Kalashnikov V (1997) Geometric sums: bounds for rare events with applications. Kluwer Academic Publishers, Dordrecht
- Kotz S, Kozubowski TJ, Podgórski K (2001) The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance. Birkhäuser, Boston
- Kozubowski TJ, Panorska AK (1998) Weak limits for multivariate random sums. *J Multivar Anal* 67:398–413
- Kozubowski TJ, Panorska AK (1999) Multivariate geometric stable distributions in financial applications. *Math Comput Model* 29:83–92
- Kozubowski TJ, Panorska AK (2002) Testing for symmetry under a skew Laplace model. *J Statist Plann Inference* (in press)
- Kozubowski TJ, Podgórski K (2001) Asymmetric Laplace laws and modeling financial data. *Math Comput Model* 34: 1003–1021
- Kozubowski TJ, Rachev ST (1994) The theory of geometric stable distributions and its use in modeling financial data. *Eur J Operational Res* 74:310–324
- Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull Am Meteorol Soc* 78:1069–1079
- Mittnik S, Rachev ST (1993) Modeling asset returns with alternative stable distributions. *Econ Rev* 12:261–330
- Rachev ST, SenGupta A (1993) Laplace-Weibull mixtures for modeling price changes. *Management Sci* 39:1029–1038
- Sedeghipour J, Dracup JA (1985) Regional frequency analysis of multiyear droughts. *Water Resour Bull* 21:481–486
- Touchan R, Meko DM, Hughes MK (1999) A 396 year reconstruction of precipitation in southern Jordan. *J Am Water Resour Assoc* 35:49–59

*Editorial responsibility: Andrew Comrie,
Tucson, Arizona, USA*

*Submitted: January 23, 2002; Accepted: September 22, 2002
Proofs received from author(s): November 28, 2002*