J1.3 RELIABLE STATISTICAL INFERENCE FOR WEATHER AND CLIMATE

Alexander Gluhovsky* and Ernest Agee
Purdue University, West Lafayette, Indiana

1. INTRODUCTION

This study demonstrates the value of inferring statistics of meteorological and climatological time series by using computer intensive subsampling method, which allows one to avoid time series analysis anchored in parametric linear models with imposed perceived physical assumptions.

As motivating examples, consider time series of Palmer Drought Index (PDI) for Arizona, Division 6 from NCDC (Figure 1), and of the vertical velocity of wind (*W*) recorded under Project LESS (Lake-Effect Snow Studies) in the winter of 1983-84 (Figure 2).

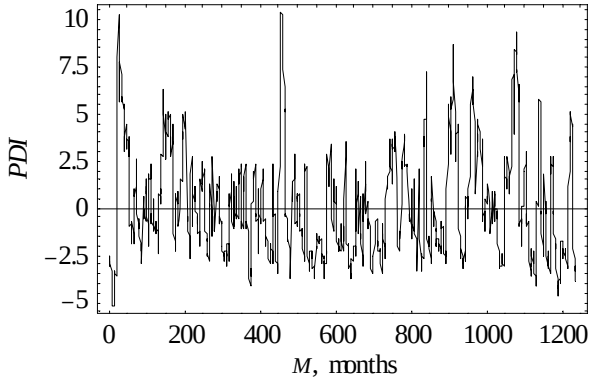

Fig.1. Palmer Drought Index (*PDI*) for years 1904-2006.

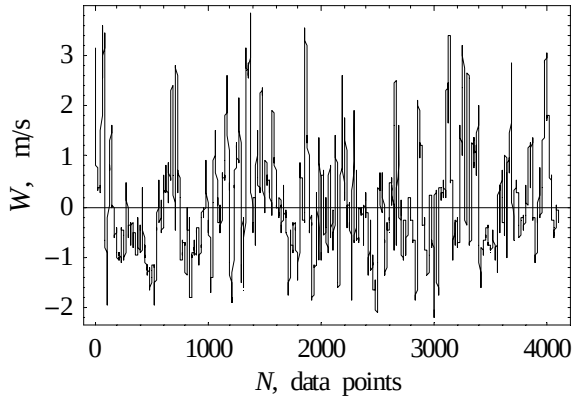

Fig. 2. Record of vertical velocity (*W*) measurements during Project LESS (see Gluhovsky and Agee 1994). *N* is the number of data points collected at a frequency of 20 data points per second.

* *Corresponding author address:* Alexander Gluhovsky, Purdue University, Dept. of Earth & Atmos. Sciences, West Lafayette, IN 47907; e-mail: aglu@purdue.edu

Figure 2 shows a segment of 4096 data points (corresponding to about 14.3 km) from the record of *W* taken at 50 m above Lake Michigan with 70 ms^{-1} flight speed and 20 Hz sampling rate. Both time series have large sample skewness (0.92 and 0.84, respectively) indicating possible nonlinearity. In statistics, confidence intervals (CIs) are used to decide how much importance is reasonable to attach to such numbers, our "best guesses", that by themselves are not guaranteed to be close to the real time series parameter (e.g., skewness). For example, positive skewness of the time series (implying that it is not normal) would be reasonably confirmed CIs for the skewness containing only positive numbers.

Since the data generating mechanism is usually unknown, the common practice is to assume a *linear* parametric model for it (thus assuming a *normal* time series), then estimate the model from the observed record, and compute CIs for parameters of the underlying time series based on the estimated model. Using Monte Carlo simulations with a model nonlinear time series, we demonstrate below that nonlinearities in the *real* data generating mechanism may render useless the inference (90% CIs for the variance of the time series) based on estimated linear parametric models, while modern computer intensive subsampling method (Politis et al. 1999) permit obtaining reliable inference (CIs for the variance and skewness in this study) without making questionable assumptions about the data generating mechanism.

A 90% CI is the range of numbers containing an unknown parameter with *coverage probability* 0.90. This implies that if instead of one time series record commonly available in practice, an enormous number of such records of equal lengths is obtainable, and from each record a CI is computed, then 90% of the resulting CIs will contain the parameter. Such coverage probability (often referred to as *nominal* or *target* coverage probability, e.g., Davison and Hinkley 1997) is attained only if all assumptions underlying the method for the CI construction are met. This is typically not the case in geosciences, so that the *actual* coverage probability may differ (sometimes considerably) from the target level. Intervals with confidence levels other than 90% (e.g., 95% or 99%) are often used in various applications (the higher confidence level the wider the interval).

2. CONFIDENCE INTERVALS FOR VARIANCE: ESTIMATED LINEAR MODELS VS. SUBSAMPLING

To get an idea of how much in error one can possibly be when computing CIs for parameters of observed time series from estimated linear models, we subjected this commonly accepted procedure to Monte Carlo simulations with a model time series. Monte Carlo simulations permit finding the actual coverage probability of such CI by using its probabilistic interpretation given above.

They also point toward a viable alternative, the computer-intensive subsampling method (Politis et al. 1999). In subsampling, independent realizations from a Monte Carlo simulation are replaced by blocks of consecutive observations from a single available record. All blocks are of the same length (the *block size*) sufficient to retain the dependence structure of the time series. One block of size b is underscored in the record below containing n observations of time series Y_t (and, therefore, $n-b+1$ blocks):

$$\left\{Y_1, \ldots, \ Y_{i-1}, \ \underline{Y_i, \ Y_{i+1}, \ldots, \ Y_{i+b-1}}, \ Y_{i+b} \ , \ldots, Y_n\right\}. \tag{4}$$

Commonly a linear model, an autoregressive moving average (ARMA) model, is fitted to the available record, and CIs are computed from the estimated model. Assume that the data are generated by the simplest such model, a first order autoregressive process (AR(1)),

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

where $0 < \phi < 1$ is a constant and ε_t is white noise (a sequence of uncorrelated random variables with zero mean and variance σ_ε^2). AR(1) is widely used in the studies of climate as a default model for correlated time series (e.g., Katz and Skaggs 1981, von Storch and Zwiers 1999, Percival et al. 2004).

If the data generating mechanism is known to be model (1), then a 90% CIs for the variance of X_t is given by

$$\hat{\sigma}_X^2 \pm 1.645 \sigma_X^2 \sqrt{\frac{2}{n} \cdot \frac{1+\phi^2}{1-\phi^2}}, \tag{2}$$

where sample variance $\hat{\sigma}_X^2$, an estimate of the "true" variance of X_t, $\sigma_X^2 = \sigma_\varepsilon^2/(1-\phi^2)$, is computed from data. When σ_X^2 in Eq. (2) is unknown (which is usually the case), it must be estimated from data (commonly unknown parameters, ϕ and σ_ε^2, are estimated). Eq. (2) follows from the fact that $\hat{\sigma}_X^2$ is asymptotically normal with mean σ_X^2 and standard error $\sigma_X^2 \sqrt{\frac{2}{n} \cdot \frac{1+\phi^2}{1-\phi^2}}$ (e.g., Priestly 1981, Brockwell and Davis 1991). For brevity the CI defined by Eq. (2) will be denoted as CI (2).

Our Monte Carlo simulations were conducted by generating 1000 records of a model nonlinear time series, fitting to each record a *linear* model, and computing from this model the 90% CI for the variance of the data generating time series. Finally, from the resulting set of 1000 CIs, the actual coverage probability was determined as the fraction of those among them that contain the "true" variance (*known* from the data generating model employed in the experiment).

First, realizations of length $n = 1024$ were generated from model (1) with $\phi = .67$ and Gaussian white noise with zero mean and variance $\sigma_\varepsilon^2 = 1-\phi^2 \approx 0.55$ (which makes $\sigma_X^2 = 1$). At the chosen value of ϕ, about 1000 data points from model (1) (and of model (3) below) allow the same accuracy in the estimation of variance as 400 independent normal observations (see, e.g., Priestly 1981). In practice, when only one record is available, determination of the optimal block size in subsampling (see below) requires the record length to be a power of 2 ($1024 = 2^{10}$).

Pretending that, as in reality, the data generating mechanism is unknown, an AR(1) model was fitted to each such realization and the goodness of fit of the model was confirmed by commonly employed diagnostic checking procedures (residual analysis, portmanteau test; see, e.g., Brockwell and Davis 1991). Not surprisingly, the coverage probability of CI (2) in this case was about its nominal value, 0.90, since the data generating time series was AR(1).

Next, model (1) was altered with a nonlinear component, so that the data were generated from the model considered earlier by Lenschow et al. (1994),

$$Y_t = X_t + a(X_t^2 - 1), \tag{3}$$

where X_t is the same as in Eq. (1) and a is a constant ($a = 0$ corresponds to model (1)). Linear models may match the first two moments (mean and variance) of observed time series, but they have zero skewness, while a nonlinear model may be capable of matching all three moments. Note that at $a = 0.14$, the mean,

variance, and skewness of Y_t are, respectively, 0, $1+2a^2 \approx 1.04$, and $(6a+8a^3)/(1+2a^2)^{3/2} \approx 0.83$, i.e., close to corresponding sample characteristics (0.04, 1.11, and 0.84 of time series *W* discussed in the introduction. Thus Y_t might provide a better description for *W* than linear models.

Monte Carlo simulations were now repeated with time series generated from nonlinear model (3) for various values of a (0.05, 0.1, 0.15, 0.20, 0.25, 0.30). In each case, 1000 realizations of time series (3) were generated, and again, an AR(1) model was fitted to each realization and passed, as before, the common residual-based postfitting diagnostic checking.

For nonlinear data, however, the *actual* coverage of CI (2), shown in Figure 3 by the solid curve, turns out to be considerably less than nominal (0.90). This means that CI (2) now becomes too narrow to provide the desired 0.90 coverage (and for $a > 0.2$ misleading). We found that the widths of CIs (2) remain the same (around .22) for all values of a, while CIs that do provide the desired 0.90 coverage should be 1.5 ($a = 0.20$) and 2.1 ($a = 0.30$) times wider.

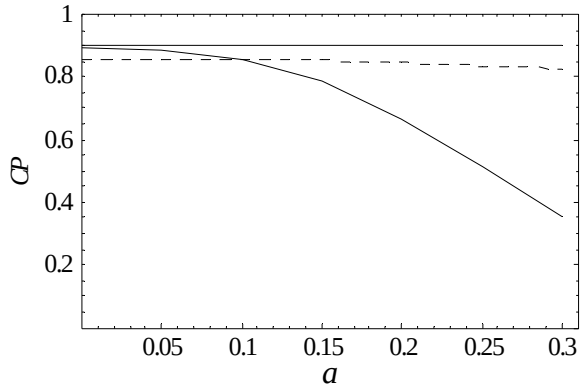

Fig. 3. Coverage probabilities (*CP*) of 90% CIs for the *variance* of time series (3) at various values of nonlinearity constant a. Solid curve corresponds to those based on estimated linear model, dashed curve – to subsampling CIs. Horizontal solid line shows 0.90 coverage. All CIs were computed from realizations of length $n = 1024$.

In contrast, when CIs (2) were replaced (in Monte Carlo simulations with the nonlinear model) by subsampling (symmetric percentile) CIs, these were expanding with increasing a, so that their coverage (dashed curve in Figure 3) remained close to the target.

It turns out that the actual coverage probability of subsampling CIs depends considerably on the block size *b*. In fact, the optimal choice of the block size is the most difficult practical problem in subsampling shared by all blocking methods. In this study, subsampling CIs were computed based on the *optimal* block size ($b = 80$ when $n = 1024$) determined through Monte Carlo simulations with model 3. In practice, when typically only one record of a time series is available, the optimal block size can be determined using a technique suggested by Gluhovsky et al. (2005), which is based on a version of the circular bootstrap (Politis and Romano 1992).

3. SUBSAMPLING CONFIDENCE INTERVALS FOR SKEWNESS

Figure 4 shows coverage probabilities of 90% subsampling CIs for the skewness of time series (3) obtained through Monte Carlo simulations analogous to those for the variance.

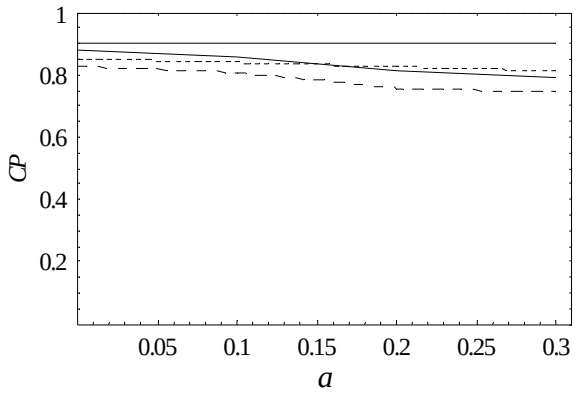

Fig. 4. Coverage probabilities (*CP*) of 90% subsampling CIs for the *skewness* of time series (3) at various values of nonlinearity constant a. Long-dashed curve corresponds to CIs computed from realizations of length $n = 1024$, short-dashed curve – same at $n = 4096$. Solid curve corresponds to to calibrated CIs at $n = 1024$. Horizontal solid line shows 0.90 coverage.

CIs based computed from realizations of length $n = 1024$ have noticeably lower coverage than their counterparts in Figure 3 (both marked by long-dashed curves). One way to improve the coverage is to increase the record length. Short-dashed curve in Figure 4 shows a better coverage due to longer records of $n = 4096$. When this is not feasible, one may use calibration, i.e., instead of the *nominal* 90% CIs providing the *actual* coverage of 0.83 at $a = 0$ and 0.74 at $a = 0.3$, employ, say, the nominal 95% CIs providing the actual coverage noticeably closer to the target (0.88 at $a = 0$, 0.79 at $a = 0.3$), as shown by the solid curve in Figure 4. In practice, calibration can be carried out using a model time series that shares certain statistical properties with the one under study (e.g., model (3) with $a = 0.14$ for the vertical velocity time series *W*).

4. SUMMARY AND CONCLUSION

This study has addressed the problem of obtaining reliable statistical inference from atmospheric and climatic time series. Two motivating examples that signify the need to depart from ubiquitous linear models were chosen, one climatological (Palmer Drought Index, *PDI*) and one meteorological (vertical velocity, *W*), whose nonzero sample skewnesses (0.92 and 0.84, respectively) indicate possible nonlinearity. In practice, a *linear* parametric model is commonly assumed for the time series under study (often a questionable assumption), then the model is estimated from the time series record, and CIs for parameters of the time series are computed based on the estimated linear model.

To investigate how nonlinearities may affect statistical inference based on *linear* models, an AR(1) (first order autoregressive) process, typically used as a default model for a correlated time series in climate studies, was altered with a *nonlinear* component. It was demonstrated that when a time series is nonlinear (which is often the case since they originate from an inherently nonlinear system), the CIs for its variance obtained from the estimated linear model are inferior and can become misleading, while those obtained through the subsampling method are valid for both the linear and nonlinear time series.

Linear models are characterized by zero skewness. We have demonstrated that subsampling can be used to estimate the skewness of nonlinear time series, although CIs for the skewness may require considerably longer records than for the variance. Meteorological observations are more likely to have adequate record lengths for nonparametric inference, while many climatological time series (such as the global annual mean surface temperature with only about 140 data points) are often too short (even for choosing the best *linear* model for the observed time series as shown by Percival et al. (2004)). On the other hand, General Circulation Models (GCMs), for example, can provide data volumes that are sufficiently large for reliable inference, which can be obtained using resampling methods.

Acknowledgments. This work was supported by National Science Foundation Grants ATM-0514674 and ATM-0541491.

5. REFERENCES

Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods.* Springer, 577 pp.

Davison, A. C. and D. V. Hinkley, 1997: *Bootstrap methods and their application*. Cambridge University Press, 582 pp.

Gluhovsky, A., and E. Agee, 1994: A definitive approach to turbulence statistical studies in planetary boundary layers. *J. Atmos. Sci.*, **51**, 1682 - 1690.

———, M. Zihlbauer, and D. N. Politis, 2005: Subsampling confidence intervals for parameters of atmospheric time series: block size choice and calibration. *J. Statist. Computation and Simulation*, **75**, 381-389.

Katz, R. W., and R. H. Skaggs, 1981: On the use of autoregressive-moving average processes to model meteorological time series. *Mon. Wea. Rev.*, **109**, 479-484.

Lenschow, D. H., J. Mann, and L. Kristensen, 1994: How long is long enough when measuring fluxes and other turbulence statistics? *J. Atmos. And Oceanic Tech.*, **11**, 661-673.

Percival, D. B., J. E. Overland, and H. O. Mofjeld, 2004: Modeling North Pacific climate time series. In *Time Series Analysis and Applications to Geophysical Systems*, D.R. Brillinger, E.A. Robinson, and F.P. Schoenberg (eds.), Vol. 139 in the series "The IMA Volumes in Mathematics and its Applications", Springer, 151–167.

Politis, D. N. and J. P. Romano, 1992: A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, eds., John Wiley, pp. 263-270.

Politis, D. N., J. P. Romano, and M. Wolf, 1999: *Subsampling*. Springer, 347 pp.

Priestley, M. B., 1981: *Spectral Analysis and Time Series.* Academic Press, 890 pp.

Von Storch, H. and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.