Nonlinear Processes in Geophysics

Brief communication
"Improving the actual coverage of subsampling confidence intervals in atmospheric time series analysis"

A. Gluhovsky[1] and T. Nielsen[2]

[1]Department of Earth, Atmospheric, and Planetary Sciences and Department of Statistics, Purdue University, West Lafayette, IN 47907, USA
[2]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA

Correspondence to: A. Gluhovsky (aglu@purdue.edu)

Abstract. In atmospheric time series analysis, where only one record is typically available, subsampling (which works under the weakest assumptions among resampling methods), is especially useful. In particular, it yields large-sample confidence intervals of *asymptotically* correct coverage probability. Atmospheric records, however, are often not long enough, causing a substandard coverage of subsampling confidence intervals. In the paper, the subsampling methodology is extended to become more applicable in such practically important cases.

1 Introduction

Observed and modeled data are often collected in the form of time series, but there are several predicaments for employing traditional time series analysis in atmospheric sciences. The primary challenges are: (a) the records are frequently prohibitively short, while just a few records (typically one) are available, and (b) conventional statistical methods are "based on certain probabilistic assumptions about the nature of the physical process that generates the time series of interest. Such mathematical assumptions are rarely, if ever, met in practice" (Ghil et al., 2002). One common assumption is that observations are normally distributed. Yet in reality distributions are often not normal, such as those for the velocity field in a turbulent flow (e.g., Lesieur, 2008), and new advances in statistics have made it clear that even slight departures from normality can be a source of concern (e.g., Wilcox, 2003). Another questionable assumption is that of

linear models for the observed time series, such as autoregressive moving average (ARMA) models, whereas the real data-generating mechanism (DGM) is inherently nonlinear, so that estimation commonly based on fitted linear models may be misleading (e.g., Gluhovsky and Agee, 2007).

These issues will be addressed below by considering a problem central to obtaining reliable statistical inference from limited data sets, namely, the construction of confidence intervals (CIs) for a parameter, θ, of the unknown distribution of a stationary time series from its finite record, $X_1, ..., X_n$.

As a typical example, consider a record in Fig. 1 from Gluhovsky (2011) of the vertical velocity of wind in a convective boundary layer during an outbreak of a polar air mass over the Great Lakes region. The record consists of 8192 data points over about 29 km across Lake Michigan, 50 m above the lake, and it has passed a test for stationarity from Gluhovsky and Agee (1994). The sample mean, variance, skewness, and kurtosis of the vertical velocity computed from this record are −0.04, 1.06, 0.83, and 4.10, respectively. The elevated skewness and kurtosis (the corresponding population parameters characterizing a linear time series are 0 and 3) may indicate nonlinearities in the underlying data-generating mechanism (DGM), but these sample characteristics are just point estimates (our "best guesses") of the true values of the parameters. Therefore, to learn how far one can trust these numbers, CIs are employed.

Here is the problem with CIs for parameters of atmospheric time series, which are produced by the inherently nonlinear system. A 90 % CI is the range of numbers that traps the unknown parameter with probability 0.90 called the
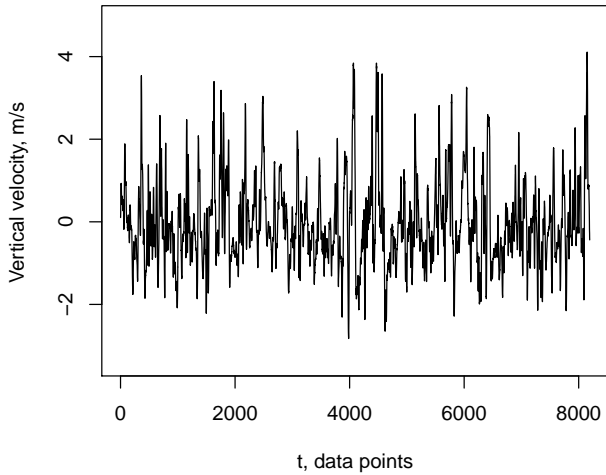
Fig. 1. Record of 20-Hz aircraft vertical velocity measurements over Lake Michigan. Figure from Gluhovsky (2011).

coverage probability. Also referred to as the *nominal* or *target* coverage probability (e.g., Davison and Hinkley, 1997), it is attained only if the assumptions underlying the method for the CI construction are met. Since for atmospheric time series this is rarely the case, the *actual* coverage probability may differ from the target level (sometimes considerably). For example, when the DGM is linear, CIs for the mean or the variance of the time series may be found analytically, but the common practice of computing CIs from fitted linear models may result in erroneous CIs when the real DGM is, in fact, nonlinear (Gluhovsky and Agee, 2007). Moreover, CIs for the skewness cannot be based on linear models that imply zero skewness. Unlike standard methods of time series analysis, resampling techniques provide asymptotically valid statistical inference from a single record without making questionable/unverifyable assumptions about the DGM. Among these methods, subsampling (Politis and Romano, 1994; Politis et al., 1999) works under the weakest assumptions, which makes it particularly applicable for atmospheric data.

When the DGM (the model) is known, CIs can be constructed using Monte Carlo (MC) simulations: one generates many records, $X_1, ..., X_n$, from the model, computes from each record a point estimate, $\hat{\theta}$, of parameter θ (skewness, for example), and an estimate, $\hat{Q}_{0.90}$, of the 0.90 quantile of the distribution of $|\theta - \hat{\theta}|$ (i.e., about 90 % of $|\theta - \hat{\theta}|$ values are below $\hat{Q}_{0.90}$, and about 10 % are above). Then a *symmetric* 90 % CI for θ is simply

$$(\hat{\theta} - \hat{Q}_{0.90}, \ \hat{\theta} + \hat{Q}_{0.90}). \tag{1}$$

Such exercise motivates the construction of subsampling CIs below. Similarly, by estimating $q_{0.05}$ and $q_{0.95}$, the 0.05 and 0.95 quantiles of the distribution of $\hat{\theta} - \theta$, one obtains an *equal-tailed* 90 % CI for θ as $(\hat{\theta} - \hat{q}_{0.95}, \ \hat{\theta} - \hat{q}_{0.05})$.

In real-life situations with unknown DGM and a single record available, $X_1, ..., X_n$, subsampling comes to the rescue by replacing independent, computer-generated MC realizations by *subsamples*, or blocks of consecutive observations from the record,

$$\underbrace{X_1, ..., X_b}_{b}, \ ..., \ \underbrace{X_i, ..., X_{i+b-1}}_{b}, \ ..., \ \underbrace{X_{n-b+1}, ..., X_n}_{b}, \tag{2}$$

which retain the dependence structure of the time series (Politis et al., 1999). Underscored above are the first, intermediate, and the last block, all of the same length b (the *block size*) in the record containing n observations and, therefore, $n - b + 1$ blocks. Using quantiles estimated from subsamples in place of independent MC records, the subsampling method yields large-sample CIs of asymptotically correct coverage, when

$$b \to \infty \quad \text{and} \quad b/n \to 0 \quad \text{as} \quad n \to \infty, \tag{3}$$

assuming the existence of a nondegenerate asymptotic distribution for $\tau_n(T_n - \theta)$ at some *known* rate τ_n (Politis et al., 1999). Typically, $\tau_n = n^\beta$, $\beta \in (0, 1]$, and $\beta = 0.5$ when estimator T_n is the sample mean, sample variance, etc.

Very often, however, real records (like that in Fig. 1) are not long enough to satisfy asymptotic conditions (3), which causes two problems for practical applications. First, the actual coverage probability of subsampling CIs greatly depends on the choice of block size b, being unacceptable beyond a relatively narrow range of b values, and second, the actual coverage may differ from the target even within this range. The first problem has been handled by another resampling technique developed (Gluhovsky et al., 2005) for the optimal choice of b based on the same single available record, so that one first determines the optimal block size using this technique, then runs subsampling with the optimal b to construct the CI. To deal with the second problem, the so-called *calibration* was suggested (Politis et al., 1999) and used ever since: a CI with a desired confidence level is obtained by running subsampling for the CI with the higher level. The latter may be determined via MC simulations with an approximating model for the time series at hand. For example, calibration was used in Gluhovsky (2011) to achieve the desired 0.90 target level for the subsampling CI for the skewness constructed from the record in Fig. 1.

In this study, another approach is explored for achieving the target coverage of subsampling CIs in a practically important case of shorter records. As mentioned above, the subsampling method requires the knowledge of the rate of convergence τ_n. In this paper, the asymptotic (known or unknown) rate of convergence is replaced by that found via simulations with an approximating model. This eliminates the need for calibration with even better overall results. For the sake of consistency and comparisons with calibration, both the time series and the approximating model are kept the same as in Gluhovsky (2011).

Table 1. Parameters of the model time series X_t (Eq. 4) distribution vs. sample characteristics of the observed series W_t in Fig. 1.

	X_t	X_t at $a = 0.145$	W_t
Mean	$M = 0$	0	-0.04
Variance	$V = 1 + 2a^2$	≈ 1.04	1.06
Skewness	$S = \frac{6a + 8a^3}{V^{3/2}}$	≈ 0.84	0.83
Kurtosis	$K = \frac{3 + 60a^2 + 60a^4}{V^2}$	≈ 3.95	4.10

2 The approximating model

When records are long enough, subsampling does not require that any model, linear or nonlinear, be fitted to the data, and it works in complex dependent data situations under the weakest assumptions among other computer-intensive techniques. The role of an approximating model that shares statistical properties with the time series under study is twofold. It becomes necessary for the assessment of the *actual* coverage of subsampling CIs via MC simulations, thus offering the opportunity denied in practice by the single observed record. From the model, one can generate numerous records, compute from each one the subsampling CI, and estimate its coverage probability by counting the fraction of times the known parameter value, θ, was within the CI. The other important role for an approximating model is to assist with the selection of the confidence level in calibration or instead with the selection of the *empirical* rate of convergence in the technique suggested in this paper.

Consider the following model (Lenschow et al., 1994),

$$X_t = Y_t + a(Y_t^2 - 1), \tag{4}$$

where Y_t is a first-order autoregressive process (AR(1)),

$$Y_t = \phi Y_{t-1} + \epsilon_t, \tag{5}$$

$0 < \phi < 1$ and a are constants, and ϵ_t is a white-noise process (a sequence of uncorrelated random variables with mean 0 and variance σ_ϵ^2). AR(1) with a Gaussian white noise is widely employed in studies of climate as a default model for correlated time series. When the white noise in model (5) is not Gaussian, the model may exhibit nonlinear behavior and is referred to as an *implicit* nonlinear model (Fan and Yao, 2003), as opposed to an *explicit* nonlinear model (4), where AR(1) is altered with a nonlinear component.

In simulations, $\sigma_\epsilon^2 = 1 - \phi^2$ so that $\sigma_Y^2 = 1$, the records contain 2048 data points, and $\phi = 0.67$, which permits to imitate the dependence structure of the vertical velocity time series in Fig. 1 as characterized by autocorrelation functions. At $a = 0.145$, the mean, variance, skewness, and kurtosis of X_t (in model 4) are close to the corresponding sample characteristics of the series (see Table 1). Thus, model (4) might provide a better description for that series than linear models, which inherently have zero skewness.
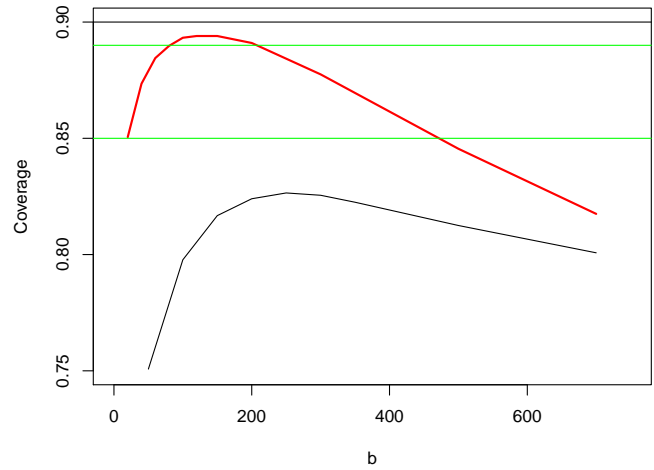
Fig. 2. Actual coverage probabilities of 90 % subsampling CIs with $\beta = 0.50$ (black) and $\beta = 0.42$ (red) for the skewness of nonlinear time series (4) at $a = 0.145$ and $n = 2048$. Horizontal green lines denote 0.85 and 0.89 levels.

3 Results of the simulation study

MC simulations with model (4) were carried out to explore how the actual coverage of subsampling CIs is affected by the introduction of the *empirical* convergence rate $\tau_n = n^\beta$, i.e., the value of exponent β now differs from the theoretical one (in an attempt to make for an insufficient record length and/or to avoid finding the theoretical value). Subsampling CIs for the skewness were computed according to Eq. (1), where $\hat\theta$ was the sample skewness computed from the whole record, while the quantiles were estimated from subsamples. The latter requires decreasing the length of resulting CIs by factor of $(n/b)^\beta$ (Politis et al., 1999).

The black curve in Fig. 2 shows the actual coverage probability of subsampling CIs for the skewness of model time series (4) at $a = 0.145$, $\beta = 0.50$ for various block sizes. One can see that because of the relatively short record, the CIs are indeed useful only within a relatively narrow range of block sizes, and even then the CIs undercover (the coverage is below the target of 0.90). Estimating the skewness does require long records, and a simple way to improve the coverage is to increase the record length. When this is not feasible (which is typically the case), we suggest employing an "empirical" rate of convergence found via MC simulations with approximating model (4). The resulting red curve demonstrates that coverage probabilities close to the target can be achieved using $\beta = 0.42$ within a range of block sizes (where the curve is above, say, 0.89).

For the vertical velocity time series in Fig. 1, the subsampling 90 % CI for the skewness with $\beta = 0.42$ is (0.56, 1.10), which compares favorably with the wider subsampling 90 % CI, (0.41, 1.24), obtained in Gluhovsky (2011) using calibration. Still both serve the purpose of confirming that the

vertical velocity skewness is positive, thus indicating nonlinearity in the series.

4 Conclusions

A new technique, an extension of subsampling methodology by introducing the empirical convergence rate, was suggested in this paper for the practically important case of short records, when subsampling CIs fail to achieve the target coverage probability. Its effectiveness was demonstrated for a model time series and then applied to the observed series of the vertical velocity of wind.

The construction of a subsampling CI for an unknown time series parameter requires the knowledge of τ_n, the convergence rate of its estimator. Although the latter can also be estimated using the subsampling methodology (Bertail et al., 1999), this causes additional sampling variability in the CIs, and achieving the target coverage in case of short records still remains problematic (see Eq. 3). Finding the empirical rate, however, removes the need for the rate estimation and simultaneously improves the coverage.

As an alternative to calibration, which has been used so far to improve the coverage of subsampling CIs, the new technique has several advantages. Again, it does not require determining the rate of convergence of the estimator (still needed for calibrated CIs); it employs the same quantile estimate (say, $\hat{Q}_{0.90}$, see Eq. 1) as the basic subsampling, whereas calibration requires higher quantiles, which are estimated less accurately; finally, new CIs are shorter. For example, the subsampling 90 % CI using $\beta = 0.42$ is (0.56, 1.10) for the skewness of vertical velocity time series in Fig. 1, which is smaller than the subsampling 90 % CI, (0.41, 1.24), obtained in Gluhovsky (2011) via calibration with $\hat{Q}_{0.96}$. In both papers, symmetric CIs were employed. Although almost all published work on resampling CIs has focused on equal-tailed intervals, symmetric CIs are often shorter and have better coverage accuracy (Hall, 1988).

To achieve the target coverage, both calibration and the new technique require an approximating model to determine (via MC simulations) the empirical confidence level or empirical rate, respectively (such as model 4 for the vertical velocity time series). In general, however, the problem of selecting an approximating model is difficult, since linear models are inappropriate whereas the multitude of nonlinear models is overwhelming.

A new possibility may result from recent progress in the work that extends statistical analysis to chaotic deterministic dynamical systems, exemplified by the celebrated Lorenz (1963) system (e.g., Collet and Eckmann, 2006; Araújo and Pacifico, 2010; Holland et al., 2012). One way to deal with formidable difficulties posed by the (deterministic) governing equations of atmospheric dynamics is to approximate them with finite systems of ordinary differential equations, the so-called *low-order models* (LOMs). LOMs in the form of coupled classical mechanical systems known as the Volterra gyrostats (gyrostatic LOMs) proved particularly useful, the simplest gyrostat being equivalent to the Lorenz system (Gluhovsky, 2006). Gyrostatic LOMs provide a bridge between the Lorenz model and the original governing equations, whose fundamental properties they inherit, thus presenting a viable alternative to standard time series models when these are ill-suited for the atmospheric data.

Edited by: W. Hsieh
Reviewed by: R. V. Donner and one anonymous referee

References

Araújo, V. and Pacifico, M. J: Three-Dimensional Flows, Springer, 2010.

Bertail, P., Politis, D. N., and Romano, J. P.: On subsampling estimators with unknown rate of convergence, J. Amer. Statist. Assoc., 94, 569–579, 1999.

Collet, P. and Eckmann, J.-P.: Concepts and Results in Chaotic Dynamics: A Short Course, Springer, 2006.

Davison, A. and Hinkley, D.: Bootstrap Methods and Their Application, Cambridge University Press, 1997.

Fan, J. and Yao, Q.: Nonlinear Time Series, Springer, 2003.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K, Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A, Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climate time series, Rev. Geophys., 40, 1–41, 2002.

Gluhovsky, A.: Energy-conserving and Hamiltonian low-order models in geophysical fluid dynamics, Nonlin. Processes Geophys., 13, 125–133, doi:10.5194/npg-13-125-2006, 2006.

Gluhovsky, A.: Statistical inference from atmospheric time series: detecting trends and coherent structures, Nonlin. Processes Geophys., 18, 537–544, doi:10.5194/npg-18-537-2011, 2011.

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

Gluhovsky, A. and Agee, E. M.: On the analysis of atmospheric and climatic time series, J. Appl. Meteorol. Climatol., 46, 1125–1129, 2007.

Gluhovsky, A., Zihlbauer, M., and Politis, D. N.: Subsampling confidence intervals for parameters of atmospheric time series: block size choice and calibration, J. Stat. Comput. Simul., 75, 381–389, 2005.

Hall, P.: On symmetric bootstrap confidence intervals, J. Royal Stat. Soc. B, 50, 35–45, 1988.

Holland, M. P., Vitolo, R., Rabassa, P., Sterk, A. E., and Broer, H. W.: Extreme value laws in dynamical systems under physical observables, Physica D, 241, 497–513, 2012.

Lenschow, D., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics?, J. Atmos. Oceanic Technol., 11, 661–673, 1994.

Lesieur, M.: Turbulence in Fluids, Springer, 2008.

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130–141, 1963.

Politis, D. N. and Romano, J. P.: Large sample confidence regions based on subsamples under minimal assumptions, Ann. Stat., 22, 2031–2050, 1994.

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

Wilcox, R. R.: Applying Contemporary Statistical Techniques, Academic Press, 2003.