Simulation Study of Time Series Models Generated by Underlying Dynamics

Evidence Matangi[1] and Alexander Gluhovsky[2]
[1]Department of Statistics, Purdue University, West Lafayette, IN 47907
[2]Department of Earth, Atmospheric, and Planetary Sciences and Department of Statistics, Purdue University, West Lafayette, IN 47907

Abstract
Subsampling confidence intervals for parameters of atmospheric time series with the actual coverages close to the target are obtained via simulations involving realistic approximating models (G-models).

Key Words: time series, atmospheric dynamics, nonlinear dynamical systems, resampling

1. Introduction

Time series analysis has been successfully applied in many areas of science and engineering, especially when data records met strong statistical assumptions underlying traditional methods and were long enough for the results obtained by these methods to be reliable. In atmospheric and climate studies, however, observed records are often prohibitively short with only one record typically available, and the above assumptions are rarely met (Ghil et al. 2002).

1.1 Motivating Example

Figure 1 shows a typical atmospheric record – the vertical velocity of wind in a convective boundary layer, taken 29 km across Lake Michigan, 50 m above the lake). For

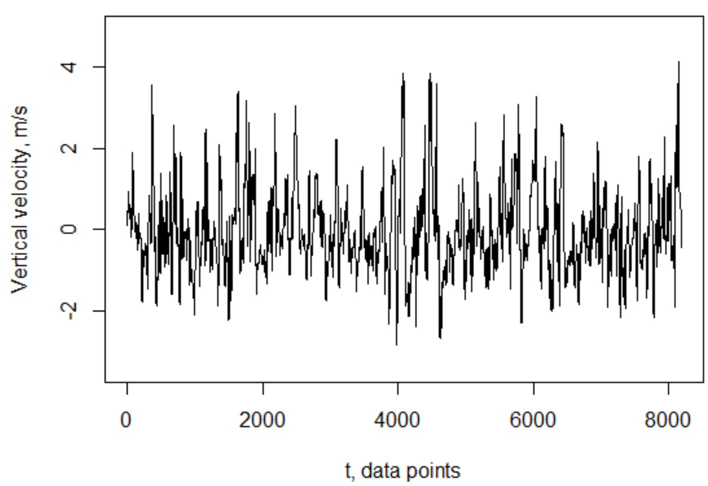

Figure 1: Record of 20-Hz vertical velocity measurements over Lake Michigan. Figure from Gluhovsky (2011).

this record, the routinely computed sample mean, variance, skewness, and kurtosis are -0.04, 1.06, 0.83, 4.10, respectively. The elevated skewness and kurtosis (from values 0 and 3 specific for a normal distribution) are attributed to the occurrence of coherent structures in turbulent flows (Ruppert-Felsot et al. 2005), but to learn how far one can trust such numbers, confidence intervals (CIs) are needed.

Their coverage probabilities (say, 0.90) are attained only if the assumptions underlying the CI construction are met, a common one being that the model generating the series is linear. Since for atmospheric time series (produced by the inherently *nonlinear* system) this is rarely the case, the *actual* coverage probability may differ from the *target* level (0.90), sometimes considerably. Moreover, CIs for the skewness cannot be based on linear models, which imply zero skewness. Thus, there is a need for nonlinear models, but finding an appropriate one among conventional models is problematic.

1.2 Subsampling Confidence Intervals and Approximating Models

Models could be avoided by employing resampling methods. In subsampling (which is especially useful for atmospheric data as it works under the weakest assumptions) the record at hand of length n is divided in $n-b+1$ subsamples, or blocks of consecutive observations, all of the same length b, that retain the dependence structure of the series (Politis et al. 1999). Subsampling yields a CI for parameter θ of the series of asymptotically correct coverage when

$$b \to \infty \text{ and } b/n \to \infty \text{ as } n \to \infty, \tag{1}$$

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

Atmospheric records, however, are typically too short to satisfy conditions (1), and so in practice, *approximating models* are needed (those sharing statistical properties with the series under study) to assess the *actual* coverage of the subsampling CI and then to fix the CI accordingly. For the latter, the *empirical* convergence rate $\tau_n = n^{\beta}$ was introduced (Gluhovsky and Nielsen 2012), where the value of exponent β was different from the theoretical one (to make for an insufficient record length and/or to avoid finding the theoretical value).

1.2.1 Approximating Model A

For a subsampling treatment of the series in Figure 1 (Gluhovsky 2011), the following model (Lenschow et al. 1994) was used,

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

referred below to as the approximating *Model A*). In Eqs. (2), Y_t is a first order autoregressive process AR(1), $Y_t = \varphi Y_{t-1} + \varepsilon_t$, $0 < \varphi < 1$ and a are constants, and ε_t is a white noise process with mean 0 and variance $\sigma_\varepsilon^2 = 1 - \varphi^2$ (so that $\sigma_Y^2 = 1$). The reason behind choosing this model was that the first four moments of X_t at $a = 0.145$ (0, 1.04, 0.84, 3.95, respectively) were close to the corresponding sample characteristics of the series in Figure 1 (0.04, 1.06, 0.83, 4.10), while setting $\varphi = 0.83$ served to fairly imitate its dependence structure as characterized by autocorrelation functions. One could then presume that Model A might be adequate for fixing subsampling CIs (but there is no guarantee that other statistical properties of the data and the model do not differ to considerably affect the intended applications).

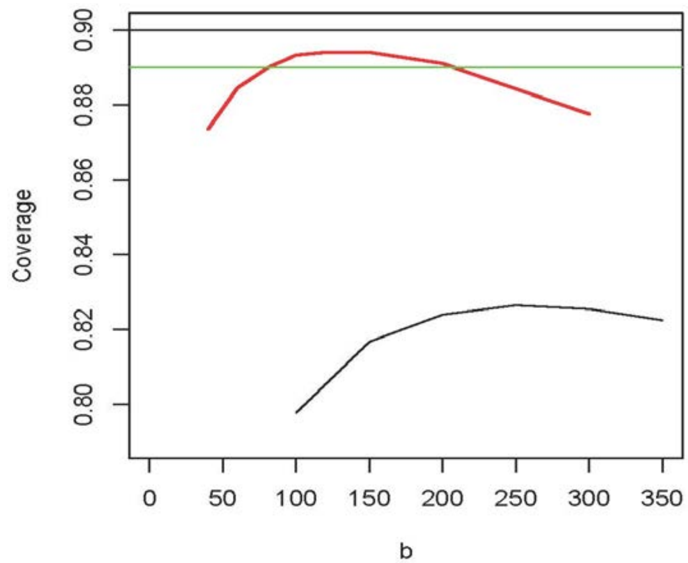

Figure 2: Actual coverage probabilities of 90% subsampling CIs with $\beta = 0.42$ for the skewness of nonlinear time series from Model A at $a = 0.145, n = 2048$. The figure is adjusted from that in Gluhovsky and Nielsen (2012).

The black curve in Figure 2 shows the actual coverage probability of subsampling CIs for the skewness of model time series (2) 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 range of block sizes, and even then the CIs undercover (the coverage is considerably 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), Gluhovsky and Nielsen (2012) suggested employing an ”empirical” rate of convergence found via MC simulations with an approximating model.

The red curve demonstrates that coverage probabilities close to the target can be achieved (within a range of block sizes) using $\beta = 0.42$. For the vertical velocity time series in Figure1, the subsampling 90% CI for the skewness with $\beta = 0.42$ is $(0.56,\ 1.10)$.

1.2.2 Approximating Model B

The efficacy of a CI depends on the record length and on how well the data generating mechanism (DGM) of the model approximates the true one. The former is given, but the DGMs of models borrowed from traditional time series analysis may be very different from those generating atmospheric data.

On the positive side, the so-called *G-models*, which were developed (Gluhovsky 2006; for the formal definition see Gluhovsky and Grady 2016) as physically sound extensions of the celebrated Lorenz (1963) model. G-models are derived from the governing equations of atmospheric dynamics and thus retain some of their physics. They were used as physically sound low-order models in problems of atmospheric dynamics (see Gluhovsky 2006, Gluhovsky and Grady 2016 and references therein) and have drawn increasing attention in various physical and mathematical studies (e.g., Bihlo and Staufer 2011, Souza and Doering 2015, Chen et al. 2017, Bianucci 2017, Majda and Qi 2018).

G-models were also suggested as alternative time series models for atmospheric dynamics (Gluhovsky 2012) since the Lorenz model flow possesses a physical ergodic invariant probability measure (Arajo et al. 2009) and satisfies the central limit theorem (Holland and Melbourne 2007, Arajo and Varandas 2012).

For the following G-model (Gluhovsky 2012, Gluhovsky and Grady 2016),

$$\begin{aligned} \dot{x}_1 &= -x_2x_3 + cx_3 - \alpha_1 x_1 + f, \\ \dot{x}_2 &= x_3x_1 - x_3 - \alpha_2 x_2, \\ \dot{x}_3 &= x_2 - cx_1 - \alpha_3 x_3, \end{aligned} \tag{3}$$

which we call *Model B*, the skewness and kurtosis of x_3 (representing the time series in Figure 1) at $c = 0.35$ proved close to those of the observed series and Model A (shown in Table 1; the results are analytical for Model A (Gluhovsky 2011) and were obtained from very long records for Model B). The autocorrelation functions of the three series are also close, which was achieved by tweaking parameter φ in Eq. (2) and the sampling rate in series generated by Eqs. (3).

Table 1. Skewness and kurtosis of the observed series and those generated by Model A and Model B

	Skewness	Kurtosis
Observed series (Figure 1)	0.83	4.1
Model A	0.84	3.9
Model B	0.81	4.2

Note that at $c = 0$, Model B is equivalent to the Lorenz (1963) model (Gluhovsky 2006). But the latter has failed to serve as an approximating model, since the skewness and kurtosis computed from its long records ($S = 0$, $K = 2.3$) were far off the sample characteristics of the observed series. This is because in addition to the Rayleigh-Bénard convection as its principal mechanism (described by the Lorenz (1963) model), the dynamics over Lake Michigan involves a hoist of others mechanisms, which in Eqs. (3) are taken care of by the single term with coefficient $c = 0.35$ (this is explained in Gluhovsky 2012, Gluhovsky and Grady 2016).

2. Simulations Results for Subsampling Confidence Intervals Enhanced via Approximating Model B

In this study, we explore how the subsampling CIs should be fixed to achieve the desired target coverage (0.90) using approximating Model B. The results of Monte Carlo simulations for the actual coverage of subsampling CIs for the skewness of the component x_3 in Eqs. (3) are shown in Figure 3. Different realizations were generated by randomly choosing the initial conditions for the runs of Eqs. (3).

In contrast to similar results for Model A shown in Figure 2, where CIs with the actual coverage of 0.90 required the empirical rate of convergence $\beta = 0.42$, for Model B it was found that $\beta = 0.65$. Accordingly, the CI for the skewness of the vertical velocity of wind with the actual coverage of 0.90 here was found to be $(0.65,\ 1.00)$ – smaller than that resulting from Model A.

Still both CIs serve the purpose of confirming that the vertical velocity skewness is positive, thus indicating nonlinearity in the series.

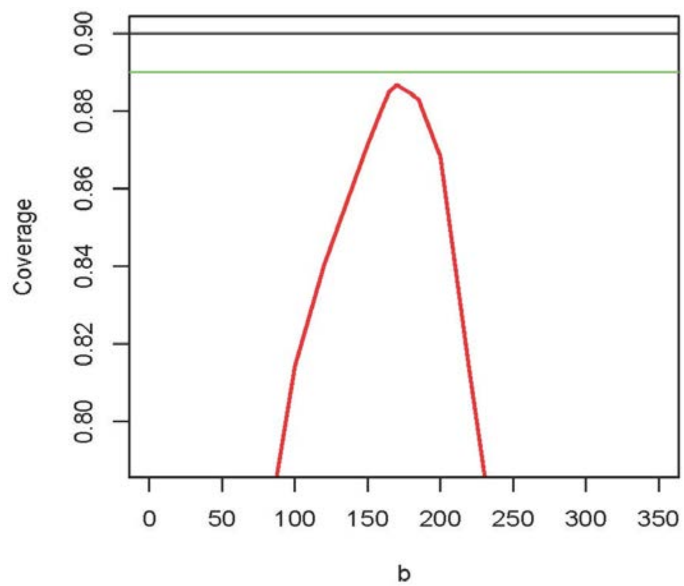

Figure 3: Actual coverage probabilities of 90% subsampling CIs with $\beta = 0.65$ for the skewness of nonlinear time series from Model B at $c = 0.35,\ n = 2048$. Horizontal green line denotes the 0.89 level.

3. Conclusions

In this paper, we have considered the construction of confidence intervals (CIs) for parameters of atmospheric time series. Their observed records are typically too short to employ resampling methods, which necessitates the use of approximating models to assess, via Monte Carlo simulations, the *actual* coverage of the CIs and the extend to which the CIs should be modified to achieve the *target* coverage.

The basic statistical characteristics of the two approximating models examined (conventional nonlinear time series Model A and a novel time series Model B derived from the underlying dynamical equations) are close to those of the observed series. However, an important advantage of Model B over Model A is that even this simple G-model shares some fundamental physics with the original system. This should help (a) to better align statistical properties of series generated by the model with those of observed series beyond the first moments and autocorrelation function, (b) to avoid a difficult task of finding an appropriate approximating model based entirely on statistical characteristics estimated with questionable accuracy, and (c) to run meaningful Monte Carlo simulations, particularly when estimators are more sensitive to properties of the DGMs.

References

Bianucci, M. (2017): Large scale emerging properties from non Hamiltonian complex systems. *Entropy*, **19**, 302.

Bihlo, A., and J. Staufer (2011): Minimal atmospheric finite-mode models preserving symmetry and generalized Hamiltonian structures. *Physica D*, **240**, 599-606.

Chen, G., N. V Kuznetsov, G. A. Leonov, and T. N. Mokaev (2017): Hidden Attractors on One Path: Glukhovsky–Dolzhansky, Lorenz, and Rabinovich Systems. *Int. J. Bifurcation and Chaos* **27**, 1750115.

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

Gluhovsky, A. (2006): Energy-conserving and Hamiltonian low-order models in geophysical fluid dynamics. *Nonlin. Proc. Geophys.*, **13**, 125-133.

Gluhovsky, A. (2011): Statistical inference from atmospheric time series: detecting trends and coherent structures. *Nonlin. Proc. Geophys*., **18**, 537-544.

Gluhovsky, A., and T. Nielsen (2012): Improving the actual coverage of subsampling confidence intervals in atmospheric time series analysis. *Nonlin. Proc. Geophys.*, **19**, 473-477, 2012.

Gluhovsky, A. (2012): Analysis of time series generated by low-order models of atmospheric dynamics. In *JSM Proceedings*, Section on Physical and Engineering Sciences. Alexandria, VA: American Statistical Association, 2244-2249.

Gluhovsky, A., and K. Grady (2016): Effective low-order models for atmospheric dynamics and time series analysis. *Chaos*, **26**, 023119, 2016.

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

Lorenz, E. N. (1963): Deterministic nonperiodic flow. *J. Atmos. Sci*., **20**, 130-141.

Majda, A. J., and D. Qi (2018): Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Review*, **60**, 491-549.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*. Springer, New York.

Ruppert-Felsot, J., O., Praud, E. Sharon, and H. L. Swinney (2005): Extraction of coherent structures in a rotating turbulent flow experiment. *Phys. Rev. E*, **72**, 016311.

Souza, A. N., Doering, C., A. (2015): Transport bounds for a truncated model of Rayleigh– Bénard convection. *Physica D*, **308**, 26-33.