NOTES AND CORRESPONDENCE

The Standard Error of Time-Average Estimates of Climatic Means

C. E. LEITH¹*Dept. of Meteorology, Massachusetts Institute of Technology, Cambridge 02139*

6 March 1973 and 17 May 1973

ABSTRACT

Climate change studies with general circulation models require the estimation of climatic means by calculation of finite-time averages. A simple stochastic model is used to estimate the standard error of such an estimation method as a function of averaging time. Consequences for long-range forecasting are also discussed.

1. Introduction

In a recent study of possible effects of man-made thermal energy on the climate, Washington (1972) has described a fundamental statistical limitation in the use of finite-time averages taken from a global circulation model for computing climate change. This limitation has been further studied by Warshaw and Rapp (1973) in a set of experiments using a different global circulation model. The climate, by definition, is described in terms of statistics of the atmosphere such as means and variances. Any single-model calculation represents only a single realization out of an ensemble of possible realizations, but if the dynamics of the model atmosphere is ergodic or transitive (Lorenz, 1968, 1970) then the climate of the model can be estimated as is that of the atmosphere itself by averages computed for long time records of a single realization. Values are available from numerical models at such short time intervals that we may treat the records as if they were continuous. The climate itself, of course, is defined in terms of a hypothetical record of infinite duration.

Since the cost in computer time for the acquisition of a model record is proportional to its length, it is important as a practical matter to decide how the accuracy of an estimate of a climatic mean depends on the length of the record used for computing averages. It is the purpose of this note to provide for this question an approximate but simple answer that can serve as a guide in planning climate change experiments. The con-

sequences for long-range forecasting will also be discussed.

2. Effect of sample length

It is a simple and well-known result of sampling theory (Cramer, 1946) that if we draw a sample of n independent values of a random variable defined over a population with mean μ and standard deviation σ , then under usual conditions the sample average is an asymptotically normal random variable with mean μ and standard deviation $\sigma/n^{1/2}$. As a result the sample average is an unbiased estimate of the mean with a standard error $\sigma/n^{1/2}$ that decreases as the inverse of the square root of the number of independent observations in the sample.

This simple result cannot be applied directly to time averages of an atmospheric variable for these are usually constructed from sample values drawn at time intervals too short to be independent. The statistical dependence of two sample values drawn at different times is revealed, of course, by the structure of the lagged time correlation. We shall find as expected, however, that the standard error of a time average as an estimate of a climatic mean decreases asymptotically as the inverse of the square root of the record length.

The atmosphere is a complicated nonlinear dynamical system which we know to have limited predictability, and over long periods we must treat it as a random time series with certain statistical properties. To avoid the statistical complexity of the real atmosphere we shall use a simple stochastic model for atmospheric variables, namely, a first-order Markov process or red-noise time series whose lagged time correlation

¹ Present affiliation: National Center for Atmospheric Research, Boulder, Colo., which is sponsored by the National Science Foundation.

$R(\tau) = \exp(-\nu|\tau|)$ is fitted to that observed in the real atmosphere. The decay rate of predictability in the model, which is given by the constant ν , is somewhat more rapid than that of the atmosphere. In both there is an inner time scale of the order of days for correlation and deterministic predictability. For the simple red-noise model it is a straightforward matter to compute the standard deviation of a time average. This will then serve as an approximation for either the real atmosphere or any numerical model of it.

3. Analysis of a sample time series

In Fig. 1 is shown the average correlation $R(\tau)$ as a function of lag time τ in days for heights of the 200-, 500- and 850-mb surfaces as computed from five years of Northern Hemisphere data by Lorenz (1973). The correlation for the red-noise model is also shown for $\nu = 0.3 \text{ day}^{-1}$ as a straight line. The value of ν has been intentionally chosen to underestimate the correlation in order that the resulting approximation would underestimate the standard error by overestimating the independence of the sampling process.

We are faced with the following problem: At a typical location the height can be considered to be a random time series $\psi(t) = \mu + \phi(t)$ which is made up of random fluctuations $\phi(t)$ about a mean μ . The stationary random time series of anomalies $\phi(t)$ has ensemble mean $\langle \phi(t) \rangle = 0$ and lagged time covariance given by

$$\langle \phi(t)\phi(t+\tau) \rangle = \sigma^2 R(\tau),$$

where σ^2 is the variance and $R(\tau)$ the correlation at lag time τ . We now compute as an estimate of μ a time average $\bar{\psi}(t)$ for a finite time interval of duration T :

$$\begin{aligned} \bar{\psi}(t) &= \frac{1}{T} \int_{-T/2}^{T/2} \psi(t+s) ds = \mu + \frac{1}{T} \int_{-T/2}^{T/2} \phi(t+s) ds \\ &= \mu + \bar{\phi}(t). \end{aligned}$$

FIG. 1. Northern Hemisphere geopotential height correlations $R(\tau)$ as a function of lag time τ using data from Lorenz (1973). The straight line is a red-noise model fit.

FIG. 2. Relative rms error σ_T/σ as a function of nondimensional averaging time νT .

The term $\bar{\phi}(t)$ evidently represents the error in $\bar{\psi}(t)$ as an estimate of μ . We are interested, therefore, in the variance σ_T^2 of the random time series $\bar{\phi}(t)$, and this is given (see, e.g., Papoulis, 1965) by

$$\begin{aligned} \sigma_T^2 &= \langle \bar{\phi}(t)\bar{\phi}(t) \rangle = \frac{1}{T^2} \int_{-T/2}^{T/2} ds_1 \int_{-T/2}^{T/2} ds_2 \langle \phi(t+s_1)\phi(t+s_2) \rangle \\ &= \frac{2\sigma^2}{T} \int_0^T [1-(\tau/T)] R(\tau) d\tau. \end{aligned}$$

For the red-noise model we find that

$$\begin{aligned} (\sigma_T/\sigma)^2 &= (2/T) \int_0^T [1-(\tau/T)] \exp(-\nu\tau) d\tau \\ &= (2/\nu T) \{1 - (1/\nu T)[1 - \exp(-\nu T)]\}. \quad (1) \end{aligned}$$

The square root of this last expression gives the relative standard error of a time-average estimate of the mean and is plotted in Fig. 2 as a function of the dimensionless averaging time νT . We see from Fig. 2 how slowly the standard error decreases with increasing T . For example, a seasonal average with $T = 90$ days, which with $\nu = 0.3 \text{ day}^{-1}$ gives a value of $\nu T = 27$, has a standard error as an estimate of the mean which is about $\frac{1}{4}$ of the standard deviation of the unaveraged time series. Since the asymptotic value of σ_T/σ is $(2/\nu T)^{1/2}$, we may define an effective independent sample size $n = T/T_0$, where

$$T_0 = \int_{-\infty}^{\infty} R(\tau) d\tau = 2/\nu$$

is a characteristic time between effectively independent sample values. For our model choice of ν we have $T_0 = 6.67$ days.

Eq. (1) is, incidentally, the special zero lag case of a more general result on the influence of time averaging on the lagged covariance function for a red-noise stochastic model that was used by Munk (1960) in another connection.

FIG. 3. Diagram showing the influence of a shift in the mean on the probability of outcomes in various categories: A, above normal; N, near normal; B, below normal.

4. Application to climate change

The study of climate change with numerical models is concerned with detecting the influence on climate of external changes such as in sea surface temperature or in incoming solar radiation. The problem, of course, is that the signal of interest, namely, the change in climatic mean, is obscured by the noise of unpredictable fluctuations in finite-time average estimates of the mean.

The significance of any computed change will depend on the ratio of the actual signal to the unpredictable noise. What we have called noise leads to a probability distribution of time averages with a standard deviation σ_T about a climatic mean μ as is shown schematically by the solid line in Fig. 3. The vertical lines divide this probability distribution into three equal parts and serve to divide anomalies in time averages into three categories, namely, below normal (B), near normal (N) and above normal (A). If a particular external influence leads to a shift in the mean by $\Delta\mu = \frac{1}{2}\sigma_T$, i.e., we have a signal-to-noise ratio of $\frac{1}{2}$, then the resulting probability distribution of anomalies is shifted as shown by the dashed line in Fig. 3. Although the new mean has shifted enough to be in category A, the unpredictable noise still leads to an appreciable probability that we would be misled by the fact that the computed time average is in either category N or B. We are assuming here that the noise distribution is normal with an rms value independent of the climatic mean.

The general situation is illustrated in Fig. 4 which shows the probability that the computed time average will be in each of the three categories as a function of the ratio of signal to rms noise. It is clear from Fig. 4 that we need a signal-to-noise ratio of at least 1.0 before we can draw significant conclusions about climate change.

A typical climate change experiment in which we might try to detect a shift in climatic mean of $\frac{1}{8}$ of the standard deviation of day-to-day fluctuations would, according to our simple estimate, require a record length of at least one year. Since this is much longer than deterministic predictability times, there is no particular advantage in starting a pair of climate perturbation calculations from the same initial conditions. In cases where this has been done the first few weeks have represented primarily a predictability study since a few

weeks is too short a time to give much information on mean climate change.

5. Application to long-range forecasting

An important practical consequence of a successful determination of the response of a climatic mean to changing external influences is the possibility that this provides for long-range forecasting. By long-range forecasting is meant forecasting for times longer than deterministic predictability times and therefore forecasting not of the day-to-day evolution of atmospheric states but rather of relatively slow changes in atmospheric statistics or climate.

One suggested source of such slow changes is the possible existence of separate regimes of internal behavior such that the atmosphere, having entered one regime, tends to remain there for extended periods of time. The long internal time scales induced by such almost-intransitive behavior (Lorenz, 1968) give a possibility of long-range forecasting skill especially if the transitions between regimes are predictable. Convincing evidence of regimes has not yet been found, however, in the large global circulation models.

A source of slow changes in atmospheric statistics that certainly does exist is the possible slow change in an external influence such as sea surface temperature (Namias, 1972). Anomalies in sea surface temperature have been observed to last for months, and they can lead to predictable changes in atmospheric climatic means for corresponding periods if the magnitude and delay of the response have been determined.

Now we are faced again with a practical consequence of the residual noise left in finite-time averages. Consider, for example, a seasonal forecast of 90-day average values which could have, if skillful, great value for agricultural planning. Let us say that from observed sea surface temperature anomalies we can confidently predict for that period a shift in mean values $\Delta\mu = \frac{1}{2}\sigma_T \approx (\frac{1}{8})\sigma$ corresponding to the case shown in Fig. 3. Then we see from Fig. 4 that the probabilities that the true seasonal average will turn out to be in the categories A, N and B, defined earlier, are, respectively, 0.53, 0.30 and 0.17. A shift of this much from the *a priori* probabilities of $\frac{1}{3}$ for each category is about enough to be of

FIG. 4. The probability of outcomes in the three categories of Fig. 3 as a function of signal-to-noise ratio $\Delta\mu/\sigma_T$.

economic importance. We have therefore established a threshold value of $|\Delta\mu|/\sigma = \frac{1}{8}$ below which predicted values have little use for seasonal forecasts. Note that this threshold value decreases like $T^{-\frac{1}{2}}$ with increasing averaging times T and thus that the noise limitation on possible long-range forecasting skill becomes less severe the longer the range.

6. Discussion

We have been treating the atmosphere as if its statistical properties were stationary and ignoring the obvious annual cycle of the seasons. The annual cycle of mean values is so well determined that it can easily be removed first with the variance σ^2 calculated for the remaining time series. The variance, too, has an annual cycle which will lead to significance levels that differ in different seasons. It would therefore be desirable to stratify climate change studies by, say, four seasons, but this, of course, requires four times longer records for a given level of significance.

We have also only been considering the standard error of estimates arrived at by time averaging. Space averaging, of course, can further increase the effective number of independent samples and decrease the standard error of estimates if the averaging region is of a scale large compared to spatial correlation lengths of a few thousand kilometers. This can be useful for certain measures of mean climate but is not particularly relevant to the long-range forecasting problem since economically useful decisions can seldom be made for regions of a scale much larger than that of spatial correlations.

The computation times required to extract significant climatic information from detailed global circulation models do not apply to statistical models in which the mean values are directly computed. Such models, if successful, will have already averaged out the day-to-day fluctuations. Again this is of no help for long-range forecasting where the difficulty arises from the unpredictable variability of finite time averages for the real atmosphere.

A more complex but more realistic stochastic model could have been constructed by first carrying out a spectral decomposition of the height field into surface spherical harmonic components and then making a different choice of ν for each absolute wavenumber k . Our experience with predictability calculations leads to the expectation that $\nu(k)$ would be an increasing function, at least for sufficiently large values of k . The value of ν appropriate to the height field in our simple model represents some sort of weighted average of $\nu(k)$. For different variables the weighting is likely to be different, and, in particular, for a derived quantity such as a velocity component we expect the average value of ν to be greater, and the independent sampling time T_0 to be shorter, than for the height field. The averaging time required for a given level of significance would also be shorter since it is proportional to T_0 .

Acknowledgments. This research was supported by the National Science Foundation under Grant GA28724X. I wish to thank E. Lorenz for providing me before publication with the data displayed in Fig. 1 as well as for many valuable comments.

REFERENCES

- Cramer, H., 1946: *Mathematical Methods of Statistics*. Princeton University Press, p. 213.
- Lorenz, E. N., 1968: Climatic determinism. *Meteor. Monogr.*, 8, No. 30, 1-3.
- , 1970: Climate change as a mathematical problem. *J. Appl. Meteor.*, 9, 325-329.
- , 1973: On the existence of extended range predictability. *J. Appl. Meteor.*, 12, 1543-1546.
- Munk, W. H., 1960: Smoothing and persistence. *J. Meteor.*, 17, 92-93.
- Namias, J., 1972: Experiments in objectively predicting some atmospheric and oceanic variables for the winter of 1971-72. *J. Appl. Meteor.*, 11, 1164-1172.
- Papoulis, A., 1965: *Probability, Random Variables, and Stochastic Processes*. New York, McGraw-Hill, Sec. 9-8.
- Warshaw, M., and R. R. Rapp, 1973: An experiment on the sensitivity of a global circulation model. *J. Appl. Meteor.*, 12, 43-49.
- Washington, W. M., 1972: Numerical climatic-change experiments: The effect of man's production of thermal energy. *J. Appl. Meteor.*, 11, 768-772.