

**STATISTICAL EXAMINATION OF CLIMATOLOGICAL DATA
RELEVANT TO GLOBAL TEMPERATURE VARIATION**

Final Report

H.L. Gray, Richard F. Gunst, and Wayne A. Woodward
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

July 1998

PREPARED FOR THE U.S. DEPARTMENT OF ENERGY
UNDER GRANT NUMBER DE-FG03-93ER61645

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

**Portions of this document may be illegible
electronic image products. Images are
produced from the best available original
document.**

SUMMARY OF RESULTS

Through DOE support, advanced statistical procedures have been developed to analyze and model climatological data. Our efforts have been presented at technical conferences, and our work has resulted in publications in scholarly journals and conference proceedings. In the following we give a summary of our results obtained under this contract.

1. *Testing the Global Temperature Series for Trend*

An important problem that we have investigated is that of determining whether the upward trending behavior in the temperature anomaly series is a "permanent" trend that should be forecast to continue, or whether it is simply a random (short-term) trend that would not necessarily be forecast to continue. Woodward and Gray (*Journal of Climate*, 1993) used the autoregressive-integrated-moving average (ARIMA(p,d,q)) model to model the temperature anomaly series, X_t . They fit ARIMA models to the temperature series with fixed mean (i.e. no permanent trend) and a single unit root. Bloomfield and Nychka (*Climatic Change*, 1992) and Bloomfield (*Climatic Change*, 1992), among others, used a trend model with stationary, autocorrelated errors.

Techniques exist for testing for trend with correlated errors, and Bloomfield and Nychka (1992) also derived a new test for this purpose. They applied this test to the temperature data sets and found a significant "deterministic" trend in the data. However, even in the tests for trend that adjust for the correlated residuals, there is a tendency for the significance levels to be inflated. Woodward and Gray (*Journal of Climate*, 1993) investigated this phenomenon as it relates to temperature data and showed that the new Bloomfield (1992) test also suffers from this tendency. For example, one would expect that when using such tests for trend with an $\alpha = 0.05$ level of significance, then only about 5% of realizations from stationary models for which there is autocorrelation present but for which there is no permanent trend would be detected as having a significant trend. In fact, Woodward and Gray showed that using maximum likelihood estimation as well as the Bloomfield and Nychka test, as many as 25-50% or more of such realizations can incorrectly be detected as having a significant

trend when being tested at the nominal $\alpha = 0.05$ level. Thus, the finding of a significant trend using these tests (as was done by Bloomfield and Nychka) certainly must be viewed with caution.

The work by Woodward and Gray (1993) on this topic was published in the *Journal of Climate*. A reprint of this paper is included in Appendix B.

Thus, we in essence have two models for the data: the fixed mean ARIMA model and the "line + correlated noise" or "trend" model. If the ARIMA model is correct, then realizations should be expected to have random trends, which may be causing the current tests for deterministic trend to detect a significant deterministic trend in a high percentage of the realizations. The two model types will often produce realizations which have similar characteristics yet they are fundamentally different. The fixed mean ARIMA process with a single unit root does not and should not predict a trend to continue, whereas the line + noise model with non-zero slope does and should predict the trend to continue.

The above considerations motivated two questions that were extensively investigated under funding from this grant: (a) *Given a time series realization of the type typically seen in temperature data, can an intelligent decision be made concerning whether the data would best be modeled as a "trend + correlated noise" model or a fixed-mean ARIMA model?* and (b) *Can tests for trend with correlated residuals be obtained that have the correct significance level?* We briefly discuss these questions below.

(a) *Distinguishing between realizations from fixed mean ARIMA models and "line + correlated noise" models.*

Under this grant we investigated the use of the parametric bootstrap combined with statistical classification to determine whether we can correctly classify realizations as coming from the correct model type. The ultimate goal is to determine whether the temperature data show characteristics most like realizations from the "line + correlated noise" model or from fixed mean ARIMA model. The procedure involves first modeling a temperature series using each model. Then, several independent realizations are then generated from each model. Features which are potentially useful for discriminating

between realizations from the two models are calculated for each of the simulated realizations. The resulting two sets of features (one set from each model) are treated as training samples from the populations of features for the two models. The classifying features are also calculated on the original time series realization and the resulting "observation" is classified as being from a fixed-mean ARIMA model or a "line + correlated noise" model using traditional statistical classification procedures. In our simulations, we consider cases in which the trend is linear and cases in which it is quadratic. Simulation results show that one can do a good job of distinguishing between the two model types with the the classification performance improving as the number of observations increases.

Application of this procedure to several sets of temperature anomaly series typically resulted in the selection of a fixed-mean ARIMA as being the more appropriate model for these temperature anomaly series. The implication of these results is that, based strictly on an examination of these series and their statistical behavior, they appear to support the position that the current observed trend may abate in the future. Simulation results also show that the bootstrap procedure presented here has the ability to distinguish between the two type of models, especially for series of length 150 or 200 years. Consequently, the results shown here suggest that if only 30 or 40 more years of reliable temperature data were available, a more definitive conclusion could be made concerning whether the trend should be forecast to continue.

This bootstrap-classification technique is described in detail in Woodward and Gray (1995) which was published in the *Journal of Climate*. A reprint of this paper is included in Appendix B.

(b) *Development of tests for trend that have appropriate significance levels when the errors are correlated.*

As mentioned above, trend tests in the literature that are intended to adjust for correlation in the residuals tend to have the problem that the observed significance levels are higher (and sometimes dramatically so) than the nominal levels. For this reason, a finding of a significant trend using these tests cannot be trusted as truly providing "significant" evidence of a trend. We have developed a test for trend in

which the control over the significance level is maintained. Part of the problem with earlier tests is that the estimated standard error of the slope estimate tends to be underestimated in realizations lengths of small to even moderately large. We successfully use the bootstrap to provide a data-based estimate of this variability and provide appropriate critical values for an α -level test.

In contrast to the findings of Bloomfield and Nychka (*Climatic Change*, 1992) and Bloomfield (*Climatic Change*, 1992), application of our test to the Hansen and Lebedeff data resulted in a finding that the slope was not significantly different from zero. Thus, based on the data alone, there is not significant evidence that the apparent trend in the temperature series is significant, i.e that it should be predicted to continue.

These findings are fully described by Woodward, Bottone, and Gray (1997) in a paper published in the *Journal of Agricultural, Biological, and Environmental Statistics*. This is a journal published by the American Statistical Association.

2. *The Effect of Initial Conditions on Autoregressive Models for Global Temperature Data*

A common approach to studying the impact of greenhouse gases on climate is via computer simulation using general circulation models (GCMs) which integrate current understanding of the various factors influencing climate through a comprehensive mathematical model. Tsonis (*Eos*, 1991) suggests modeling global temperature data by fitting an autoregressive model to the data and then using such a model to generate additional realizations with modified starting values. He assumes that the sensitivity of GCMs to initial conditions is similar to that of the autoregressive model, and he then reasons that if slightly changing the initial values in simulated realizations from the autoregressive model results in very dissimilar realizations, this provides evidence that the GCMs themselves are too sensitive to initial conditions to trust predictions made from simulations.

Gray and Woodward (*Eos*, 1992) show that Tsonis is incorrect when he concludes that predictions from a stationary autoregressive model fitted to temperature data are sensitive to the starting values; on the contrary, the opposite is true. It is well

known that for any stationary autoregressive process, if the sample size is large relative to the order of the process, the initial values will have little effect on forecasts. We show that different realizations with the same starting values from the autoregressive model fit by Tsonis typically show little similarity beyond 30 years and literally no resemblance (in the sense of correlation) beyond 50 years. The "different" realizations indicated here have different and independently derived noise sequences. The dissimilarity Tsonis observed is due to the correlation structure and noise sequence, not the starting values. As another demonstration of the situation, we generated two samples with different starting values but with the same noise sequence and correlation structure. This corresponds to the same set of random conditions driving the weather, but from different starting values. Simulations showed that the two realizations will typically merge within 30 years, again illustrating their insensitivity to the starting values.

The paper by Gray and Woodward (1992) was published in *EOS*. A copy of this article is included in Appendix B.

3. Long Memory Modeling of the CO₂ Data

A characteristic of many climate-related data sets is the fact that the correlation structure between observations may persist over long intervals of time. When this is the case, it may be true that neither the deterministic signal+noise model nor an ARIMA model may be appropriate. A flexible model, that has been developed for purposes of modeling long-memory data, is the Gegenbauer ARMA(GARMA) long memory model introduced by Gray, Zhang, and Woodward (*Journal of Time Series Analysis*, 1989). This model is an extension of the ARMA model which allows for long term dependency in stationary models, i.e. allows for stationary models with slowly damping autocorrelations. The k -factor GARMA model developed by Woodward, Cheng, and Gray (*Journal of Time Series Analysis*, 1998) allows for long-memory behavior at several frequencies. This model seems to be applicable for data related to climate change, and we have investigated its use on the monthly atmospheric CO₂ measurements. A 2-factor GARMA model was fit to the CO₂ data with the stationary, long-memory behavior occurring at frequencies of 12 and 6 months. In addition to a stationary autoregressive component, the model has the factor $(1 - B)^2$, indicating that

the model is nonstationary and that the increasing trend will be predicted to continue. As an alternative model for the carbon dioxide data, we fit a nonstationary AR(17) using standard model-fitting techniques. Mean square errors of the forecasts from the two models were compared at six different origins, and it was found that the AR model had better forecasts in the short term, but that the 2-factor GARMA model performs better for long-term forecasts.

In the paper by Woodward, Cheng, and Gray (1998), we develop the properties of the k -factor GARMA model as well as provide a useful long-memory analysis of the CO₂ data. This paper is in press and will appear this year in the *Journal of Times Series Analysis*. A copy of the galley proofs is included in Appendix B.

4. *Spatial Modeling of Global Temperature Data*

Three major advances were made in the promotion of scientifically valid spatial modeling of global temperature data. All three advances contributed to improving the scientific validity of area estimates of mean climatological variates, including temperature.

(a) *Estimating Global Mean Temperature*

At the outset of this work it was clear that proper account was not being made of spatial correlations in the estimation of global mean temperature. Ordinary averages are not optimal estimates because, while unbiased, they do not have the smallest variances when data are correlated. Among the more serious impacts of not properly accommodating spatial correlations is that measures of uncertainty, notably standard errors, are not correct. In particular, standard errors calculated from the usual formulas that are appropriate for independent data are too small if data are spatially correlated.

In the paper "Defining and Estimating Global Mean Temperature Anomalies" (Gunst, Basu, and Brunell, *Journal of Climate*, 1993), mean global temperature was formally defined. An explicit definition of mean global temperature had not prior to that time appeared in the climatological literature in papers that dealt with estimating

global mean temperature. The definition of the mean was the normalized (by the area of the region) integral of point-location temperature anomalies. Based on this definition, the authors demonstrated very clearly that the ordinary average is not optimal when data are spatially correlated. Moreover, some of the most frequently used alternatives, notably distance weighted averages, were similarly shown to be inefficient and generally biased if temperature anomalies were not constant throughout the region for which the estimates are calculated.

Issues regarding the gridding of data and data reuse were also investigated in this paper. Gridding was shown to be effective if ordinary averages or distance weighted averages are used to estimate regional means. Neither of these estimators are optimal, but gridding ameliorates the effects of irregular spatial distribution of station locations. Station reuse was examined by calculating regional averages using stations within concentric, overlapping circles. The more the overlap, the greater the bias in estimating regional means. The problem with reuse is not reuse itself. It is that no account of the spatial correlations is being made either with individual calculations in a single circle of stations or in the repeated use of station in overlapping circles.

(b) Estimating Spatial Correlations

Early use of spatial modeling in the climatological literature was referred to as optimal spatial averaging. It has also been referred to as best linear unbiased prediction (BLUP) and kriging. Central to all of these equivalent methods for the optimal estimation of regional and global means is the estimation of an appropriate correlation function that properly accounts for spatial correlations among station locations. The fitting of correlation functions to correlations calculated from pairs of station time series dominated the climatological literature during the term of this contract.

The paper "Estimating Spatial Correlations from Spatial-Temporal Meteorological Data" (Gunst, *Journal of Climate*, 1995) comprehensively demonstrated that correlations calculated from pairs of time series did not correctly account for spatial correlations. In fact, such correlations more indicated temporal correlation than spatial correlation. The proper method for accommodating spatial correlations was the estimation of structure functions (semivariograms), long-known but little used functions that could be estimated from all station data at a fixed time point.

Structure functions can be calculated isotropically or anisotropically. Since they are based on calculations from spatial locations at a fixed time point, they are not temporally biased. When averaged over several time points, they provide a very stable, usually smooth, function which can be fit to suitably chosen theoretical semivariogram models. The parameters of the fitted models characterize key aspects of the spatial correlations such as the range over which spatial correlations are nonzero. Finally, the fitted structure functions provide the basis for proper standard error calculations.

(c) Influential Spatial Data

Insidious in that they are often overlooked with the large data bases used for climatological modeling, influential spatial data can have a strong effect on the fitting of structure functions. The literature on structure function modeling made no mention of this problem and its importance was not discovered until careful examination of fitted models for temperature anomalies was undertaken. Highly irregular patterns, unlike the smooth monotonic trends expected, were found. Some of the patterns seemingly led to scientifically insupportable conclusions such as the apparent strengthening of spatial correlations as distance between stations increased. These findings were ultimately shown to be due to the presence of one or a very small number of spatially influential data values.

The paper "The Effects of Influential Observations on Sample Semivariograms," was initially printed as a technical report and ultimately published (Basu, Gunst, Guertal, and Hartfield, *Journal of Agricultural, Biological, and Environmental Statistics*, 1997). In this paper influential data are shown to produce sudden spikes, linear trends, mound-shaped excitation crests, and shifts in structure function plots. When the influential data are removed, the plots invariably revert to the smooth monotonic trends that are expected. The paper presents examples of each of these aberrations. Mathematical foundations for the appearance of these aberrations show that they are due to differences in locations of the influential data. Finally, graphical and quantitative methods for identifying spatially influential data are provided. A companion paper "Robust Semivariogram Estimation in the Presence of Influential Spatial Data Values" (Gunst and Hartfield, *Modeling Longitudinal and Spatially Correlated Data*, 1996) compares several robust structure function estimation methods.

APPENDIX A

BIBLIOGRAPHY AND OTHER SCHOLARLY ACTIVITIES

Listed below are the reports that were published under the sponsorship of this grant. Also included are other activities such as presentations at technical conferences.

1. *Reports Which Appeared in the Scientific Literature*

The following are manuscripts that were written under the support of this grant.

- (1) Basu, S., Gunst, R.F., Guertal, E.A., and Hartfield, M.I. (1997). "The Effects of Influential Observations on Sample Semivariograms," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 490-512.
- (2) Gray, H.L. and Woodward, W.A. (1992) "The Effect of Initial Conditions on Autoregressive Models for Global Temperature Data," *EOS* 73, 267-268.
- (3) Gray, H.L., Gunst, R.F. and Woodward, W.A. "Statistical Examination of Climatological Data Relevant to Global Temperature Variation," *Proceedings of the 5th International Meeting on Statistical Climatology*, 197-201.
- (4) Gunst, R.F. (1991). "The Scientific Controversy Over Global Warming," *STATS: The Magazine for Students of Statistics*, 3-8.
- (5) Gunst, R.F. (1994), "Estimating Global Mean Temperature from Instrument Data," *Proceedings of the Colloquium on Applications of Statistics to Modeling the Earth's Climate Systems*, Technical Report No. 409, National Center For Atmospheric Research, 56-60.
- (6) Gunst, R.F. (1995). "Estimating Spatial Correlations from Spatial-Temporal Meteorological Data," *Journal of Climate*, 8, 2454-2470.

- (7) Gunst, R.F., Basu, S., and Brunell, R. (1993). "Defining and Estimating Global Mean Temperature Anomalies," *Journal of Climate* 6, 1368-1374.
- (8) Gunst, R.F., Basu, S. and Woodward, W.A. (1993). "Estimating Spatial Correlations," *Proceedings of the 18th Annual Climate Diagnostics Workshop*, Boulder, Colorado, November 1-5, 1993.
- (9) Gunst, R.F. and Hartfield, M.I. (1997), "Robust Semivariogram Estimation in the Presence of Influential Spatial Data Values," *Modeling Longitudinal and Spatially Correlated Data*, (T.G. Gregoire, et al., eds.), Springer Verlag, 265-274.
- (10) Woodward, W.A., Cheng, Q.C., and Gray, H.L. (1998), "A k -Factor GARMA Long-Memory Model," *Journal of Time Series Analysis*, 18 (*in press*).
- (11) Woodward, Wayne A. and Gray, Henry L. (1992) "Testing for Trend in the Global Temperature Data," *DOE Research Summary*, No. 16, April 1992.
- (12) Woodward, W.A. and Gray, H.L. (1993), "Global Warming and the Problem of Testing for Trend in Time Series Data," *Journal of Climate* 6, 953-962.
- (13) Woodward, W.A. and Gray, H.L. (1994), "Selecting a Model for Detecting the Presence of a Trend," *Journal of Climate*, 8, 1929-1937.
- (14) Woodward, W.A. and Gray, H.L. (1993), "Distinguishing Between Deterministic and Random Trends in Time Series Data," *Computer Science and Statistics: Proceedings of the 25th Symposium on the Interface*.
- (15) Woodward, W.A., Bottone, S., and Gray, H.L. (1997), "Improved Tests for Trend in Time Series Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 403-416.

2. Professional Presentations

The following are presentations given by Southern Methodist University faculty, supported under the DOE grant.

Richard F. Gunst

1. "Statistical Issues in the Global Warming Controversy"
 - (i) Statistical Science Department Seminar, Southern Methodist University,
 - (ii) Division of Mathematics Seminar, Computer Science and Statistics, University of Texas at San Antonio, December 6, 1990.
 - (iii) Carrollton-FB ISD, LEAP students Conference Day, SMU Pre-College Programs, March 30, 1992.
2. "The Scientific Controversy over Global Warming,"
 - (i) Mathematics and Science Student Awards Banquet, May 14, 1991
 - (ii) SMU Parents Weekend, Fall 1991
3. "Defining and Estimating Global Temperature Change,"
 - (i) Department of Mathematics, Trinity University, November 14, 1991
 - (ii) SMU Chapter of Sigma Xi, November 21, 1991
4. "Estimating Annual Mean Global Temperature Change," Annual Joint Statistical Meetings, August 9-13, 1992. (with S. Basu and R. Brunnel)
5. "Statistical Challenges in the Estimation of Global Mean Temperature," Statistics Seminar Series, Texas A&M University, November 19, 1992.
6. "Assessing the Statistical Debate Over Global Warming," Mathematics Club, Sacred Heart University, April 28, 1993.

7. "Spatial Statistical Modeling of Global Temperature Data"
 - (i) Connecticut Chapter of the American Statistical Association, April 28, 1993
 - (ii) Center for Statistical Science, University of Texas (Austin),
February 17, 1994
 - (iii) Department of Statistics, Rice University, March 28, 1994
8. "Estimating Spatial Correlations," poster session at the 18th Annual Climate Diagnostics Workshop, Boulder, Colorado, November 1-5, 1993.
9. "Estimating Global Mean Temperature from Instrument Data," NSF/NCAR Conference on Applications of Statistics to Modeling the Earth's Climate System," National Center for Atmospheric Research, Boulder, CO, July 11, 1994.
10. "Estimating Spatial Correlations from Spatial-Temporal Data," Annual Joint Statistical Meetings, Toronto, August 15-18, 1994.
11. Invited Speaker at the National Science Foundation Workshop on Industry/Academic Collaboration, July 9-12, 1995, California Polytechnic University.
12. "Spatial Modeling on a Sphere," Annual Joint Statistical Meetings, Orlando, FL, August 13-17, 1995 (with Z. Yucel)
13. "The Influence of Abberant Spatial Data Values on Semivariogram Modeling"
 - (i) Statistics Department, University of Iowa, March 28, 1996
 - (ii) Statistics Department, Iowa State University, March 29, 1996
 - (iii) presentation at Conference on Applied Statistics in Agriculture, Kansas State University, April 29, 1996
14. "An Introduction to Spatial Modeling," Southern Regional Conference on Statistics, DeGray Lodge, AR, June 9-13, 1996.

H.L. Gray

1. "Warming Trend: Fact or Fiction?"
 - (i) Physics Department Seminar, Southern Methodist University
 - (ii) Freshman Seminar, Southern Methodist University
2. Served as a panelist at Texas Tech University in February 1991 in a panel discussion on global warming.

Wayne Woodward

1. "Global Warming and the Problem of Testing for Trend in Time Series Data,"
 - (i) Stephen F. Austin University, Mathematics Department Seminar
 - (ii) Southern Methodist University, Statistical Science Seminar
 - (iii) presentation before two undergraduate mathematics classes at Baylor University.
2. "Statistical Examination of Climatological Data Relevant to Global Temperature Variation," invited presentation at the 5th International Meeting on Statistical Climatology (5IMSC) held at the University of Toronto, June 22-26, 1992. (with Gray and Gunst)
3. "Distinguishing between Deterministic and Random Trends in Time Series Data," invited presentation before the 25th Symposium on the Interface between Computer Science and Statistics, San Diego, California, April 14-17, 1993.
4. "Global Warming: A Look at the Data," presentation at Parent's Weekend, Southern Methodist University, October 8, 1993.
5. "Global Warming and Trend Testing," Statistics Roundtable, University of Texas Southwestern Medical School, April, 1994.

6. "Selecting a Model for Detecting the Presence of a Trend - with Applications to the Global Temperature Anomaly Series," Environmetrics Conference, Burlington, Canada, August 10-13, 1994.

Paul Whitney

1. "Statistical Issues in Climatology: Looking for the Greenhouse Effect," Texas Academy of Science, Stephen F. Austin University, March 1, 1991.
2. "Temperature Trends During the Past Century," Texas Conference of Statisticians, Dallas, Texas, March 19, 1991.

Christopher Hill

1. "Global Climate Change and Pleistocene Pluvials in Saharan North Africa," Geological Society of America Meetings, 1991.

Harold Hietala

1. "An Analysis of Temperature Variation During the Late Holocene," American Quaternary Association meetings, 1992.

Sabyasachi Basu

1. "Influence Diagnostics for Spatial Modeling," presentation at the Indian Statistical Institute, January 1994.

APPENDIX B

Included in this appendix are copies of papers supported under this grant that have appeared in refereed scientific journals.

*Reprints removed for
separate processing*