ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 24, NO. 3, 2007, 503–508

Prediction of Monthly Mean Surface Air Temperature in a Region of China

Jeong-Hyeong LEE*[1] and Keon-Tae SOHN[2]

[1] *Division of Management Information Science, Dong-A University, Busan* 604-714, *Korea*

[2] *Department of Statistics, Pusan National University, Busan* 609-735, *Korea*

(Received 7 June 2006; revised 29 September 2006)

ABSTRACT

In conventional time series analysis, a process is often modeled as three additive components: linear trend, seasonal effect, and random noise. In this paper, we perform an analysis of surface air temperature in a region of China using a decomposition method in time series analysis. Applications to the National Centers for Environmental Prediction/the National Center for Atmospheric Research (NCEP/NCAR) Collaborative Reanalysis data in this region of China are discussed. The main finding was that the surface air temperature trend estimated for January 1948 to February 2006 was not statistically significant at 0.5904°C (100 yr)$^{-1}$. Forecasting aspects are also considered.

Key words: Chinese region, surface air temperature, time series, decomposition method, regression

DOI: 10.1007/s00376-007-0503-1

1. Introduction

The problem of detecting and forecasting a climate change signal in the climatological record is of obvious importance in any strategy to understand regional and global change. Over the last few decades there has been growing interest in trying to quantify the magnitude of regional and global climatic trends. Trends are defined here by the mean slope of the change in amplitude of a given climatic variable with time over some prescribed period.

Since understanding climatic changes at the regional scale is one of the most important and uncertain issues within the climate change debate, the need to provide regional climate change information has increased for both impact assessment studies and policymaking (IPCC, 2001, 739–768; Lee et al., 2005). A regional climate is determined by interactions between large, regional and local scales, but the available tools have directed research toward understanding the climate system as a whole.

On the other hand, the statistical analysis of observed temperature time series has not yet provided conclusive evidence about the climatic warming effect. In fact, much has been accomplished in the detection and attribution of climatic change on a global scale. All projections of future change indicate that warming is likely to continue. This conclusion holds regardless of the computer model used or the "emission scenario"—the particular set of data describing the future emissions of greenhouse gases and aerosols—applied in the model. The advantage is that no arbitrary assumptions about the variable's statistical properties are required, the catch being that we are assuming that dynamic models are correctly reproducing these properties. However, the statistical analysis of observed temperature time series using statistical models is also important, at least to complement the physical models (Zwiers, 2002).

A summary of results from previous global and regional climate change studies is presented in Table 2 of Casey and Cornillon (2001). At the global scale, the 1960 to 1990 anomaly trends calculated using the Comprehensive Ocean-Atmosphere Data Set (COADS; Woodruff et al., 1987) are found to be 0.09°C±0.03°C and 0.10°C±0.03°C per decade for the 5° × 5° and temperature class averaging techniques, respectively. Trends for other climatic variables have also been estimated, but normally for shorter periods than for surface temperature. Wu and Straus (2004)

*E-mail: jeonglee@dau.ac.kr

have computed 55-year duration trends, for the period 1948 to 2002, for surface pressure, temperature at different heights etc. Trends have also been estimated for individual countries for a range of climatic variables, although usually only over multidecadal periods. See, for example, Jonsson and Fortuniak (1995)—surface wind directions for Sweden; Brunetti et al. (2000)—temperature trends for Italy; Kaiser (2000)—cloudiness and other trends for China; and Osborn et al. (2000)-precipitation trends for the UK. Lee et al. (2005) studied climate change detection, attribution, and prediction for the surface air temperature in the Northeast Asian region.

Our objectives are to estimate long-term warming trends and to predict surface air temperature in a region of China using a decomposition method in time series analysis. We propose an additive model consisting of a linear trend, seasonal terms and an error process with the ARMA (autoregressive moving average) model.

The paper is organized as follows. Section 2 describes the data sources and the regions analyzed. Section 3 explains the time series model, its application, and results for monthly mean surface air temperature in a region of China. Finally, some concluding remarks are provided in section 4.

2. Data and regions

The primary dataset used in this study is the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) Collaborative Reanalysis (hereafter, NNR) data obtained from the National Oceanic and Atmospheric Administration's (NOAA) Climate Diagnostics Center website (http://www.cdc.noaa.gov/). We used monthly mean surface air temperature from the NNR data for January 1948 to December 2005. The resolution of the data is $2.5° \times 2.5°$. The set of data analyzed in this paper is for monthly surface air temperature anomalies from January 1948 to February 2006 in a region of China, defined as (20°–52.5°N, 72.5°–135°E) (Fig. 1).

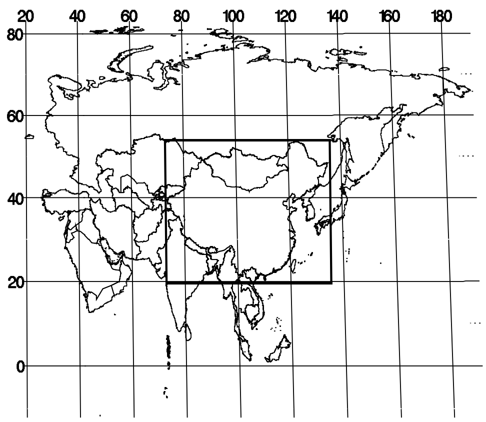

Fig. 1. Map showing the region used to compute surface air temperature anomalies.

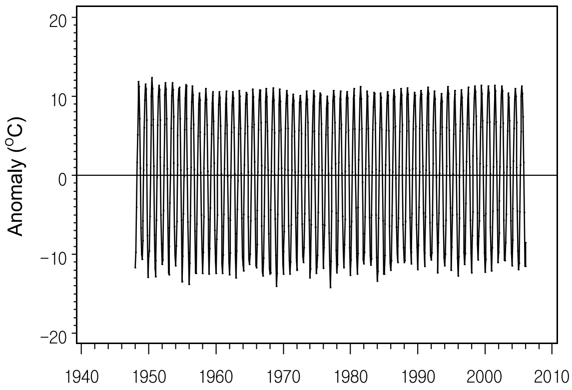

Fig. 2. Temperature anomalies covering the studied region of China.

A plot of the data set is shown in Fig. 2. We have subtracted sample means from the original dataset and thus the data plotted corresponds to this mean deleted observation.

3. Method and results

The fundamental statistical model used in this study is:

$$Y_t = T_t + S_t + I_t, \ \ t = 1, 2, \ldots, N \ ,$$

where T_t is the linear trend, S_t is the seasonal effect, and I_t is the random noise. For example, in the study of global warming, surface air temperature data consists of the long-term warming trend, seasonal variations within years, and random variation.

We consider the additive model given below for the surface air temperature data where $\{Y_t\}$ are the observed time series:

$$Y_t = c_0 + c_1 t + \sum_{i=1}^{12} r_i S_{it} + X_t \ , \ \ t = 1, 2, \ldots, N \quad (1)$$

where c_0, c_1 and r_i ($i = 1, \ldots, 12$) are parameters to be estimated, 12 is the number of months in a year,

$$S_{it} = \begin{cases} 1, & t(\text{mod } 12) = i \\ 0, & \text{otherwise} \end{cases}$$

is an index function of a month, as we are dealing with monthly surface air temperature data. Also, we assume that the stationary time series $\{X_i\}$ can be represented by an ARMA(p, q) model of the form:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q} , \quad (2)$$

where $\{e_t, t = 1, \ldots, N\}$ is a sequence of independent and identically distributed (IID) random variables with mean zero and variance σ_e^2.

We assume that the roots of the polynomial $a(B) = 1 + a_1 B + a_2 B^2 + \cdots + a_p B^p$ and $b(B) = 1 + b_1 B + b_2 B^2 + \cdots + b_q B^q$ lie outside the unit circle. We now consider the estimation of the parameters of the non-stationary model (1) the parameters of the stationary ARMA model (2). ARMA(p, q) processes are stationary, short range, correlated normal processes. White noise residuals [ARMA(0,0)] and red noise residuals [ARMA(1,0)] are the most commonly used.

We first consider the estimation of the parameters c_0 and c_1 by the ordinary least square (OLS) method. Let

$$\zeta_t = \sum_{i=1}^{12} r_i S_{it} + X_t, \quad t = 1, 2, \ldots, N .$$

Then

$$Y_t = c_0 + c_1 t + \zeta_t , \quad t = 1, 2, \ldots, N .$$

Let

$$\boldsymbol{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} , \quad \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} ,$$

$$\boldsymbol{Z} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_N \end{bmatrix} , \quad \boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N \end{bmatrix} .$$

We estimate the parameters c_0 and c_1 by the OLS method. Toyooka (1977, 1980) showed that under some conditions on the regression variables, OLS can give consistent estimates of the parameters even in the case where the errors are nonstationary and uniformly modulated. We have the OLS estimate $\hat{\boldsymbol{c}}$ of $\boldsymbol{c}$ as $\hat{\boldsymbol{c}} = (\boldsymbol{T}'_1 \boldsymbol{T}_1)^{-1} \boldsymbol{T}'_1 \boldsymbol{Y}$, where $'$ denotes the transpose of a matrix..

Having estimated the parameters c_0 and c_1 we can obtain the residuals' $\{Z_t\}$ from

$$Z_t = Y_t - \hat{c}_0 - \hat{c}_1 t, \quad t = 1, 2, \ldots, N .$$

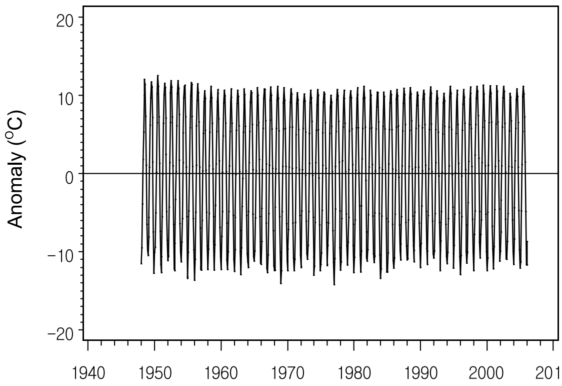

Fig. 3. Time series plot of Z_t for the studied region of China.

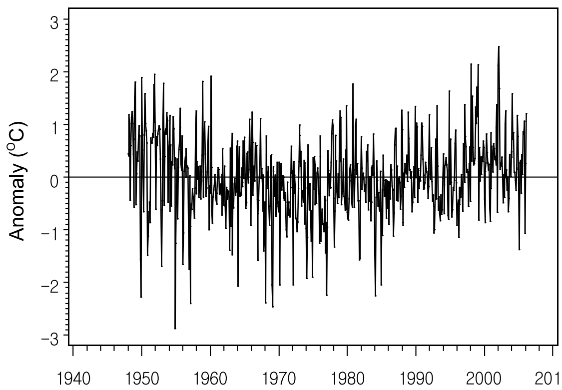

Fig. 4. Time series plot of X_t for the studied region of China.

We now estimate the r_i parameter 12 using the residual $\{Z_i\}$, given by

$$Z_i = \sum_{i=1}^{12} r_i S_{it} + \xi_y , \quad t = 1, 2, \ldots, N ,$$

where

$$\xi_t \approx X_t , \quad t = 1, 2, \ldots, N .$$

The time series plot of Z_t for the studied region of China is given in Fig. 3.

The estimation of r_i is quite simple. Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix} , \quad \boldsymbol{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{12} \end{bmatrix} , \quad \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix} ,$$

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix} ,$$

Estimates of $\{r_i\}$ can then be obtained by OLS. The least squares estimates $\{\hat{r}_i\}$ of the unknown parameters $\{r_i\}$ are given by $\hat{\boldsymbol{R}} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{Z}$.

The final stage for estimating the parameters of model (1) is to fit an ARMA model. From model (1) we have

$$Y_t - c_0 - c_1 t - \sum_{i=1}^{12} r_i S_{it} = X_t\ , \quad t = 1, 2, \ldots, N\ ,$$

and we can assume that the resulting series is stationary and satisfies an ARMA model. We used Akaike Information Criteria (AIC; Akaike, 1977) to determine the order in which to estimate the parameters of the ARMA model, and then estimated the parameters. The time series plot of X_t for the studied region of China is given in Fig. 4.

Following the method described above, we fitted models of Eq. (1) to the sets of surface air temperature data. Here, we summarize the obtained models for this set of data:

$$\begin{aligned} Y_t = & -0.172127 + 0.000492t - 11.9855S_{1t} - \\ & 9.9373S_{2t} - 5.0901S_{3t} + 0.7887S_{4t} + \\ & 5.6949S_{5t} + 9.3668S_{6t} + 10.8900S_{7t} + \\ & 10.0968S_{8t} + 6.3357S_{9t} + 0.6430S_{10t} - \\ & 5.9140S_{11t} - 10.5112S_{12t} + X_t \\ & \qquad t = 1, 2, \ldots 698, \end{aligned} \tag{3}$$

where $\{X_t\}$ is given by:

$$\begin{aligned} (1 - 0.5129B)X_t = (1 - 0.2316B)e_t\ , \\ t = 1, 2, \ldots, 698\ , \end{aligned} \tag{4}$$

and $\{e_t, t = 1, 2, \ldots, N\}$ are IID with mean zero and variance σ_e^2. Estimates of σ_e^2 and the Q statistic (Ljung and Box, 1978) are $\hat{\sigma}_e^2 = 0.4868$ and $\hat{Q}_{48} = 50.82$, which is not significant (p-value=0.2551). Thus, we conclude that the process e_t is a white noise process. As estimates of the constant c_1 are small values, we calculate the standard error. Thus, we have

$$E_s(\hat{c}_1) = 1.5400 \times 10^{-3}\ .$$

Statistically significant trends were not detected in the regional monthly anomalies temperature series.

Also, from the slope (S) of the linear trend fitted to the dataset we notice that there is an increase in the temperature, which is calculated from:

$$\text{Increase} = \text{Slope} \times \text{number of interval}\ ,$$

where is the number of months. The surface air temperature trend per decade estimated for January 1948 to February 2006 was not statistically significant at 0.5904°C (100 yr)$^{-1}$.

The trends in surface air temperature for this region of China show interesting features that appear to have relevance to global change issues. This result agrees with IPCC (2001), Zwiers (2002), and Stott and Kettleborough (2002). It is generally believed that the average global surface air temperature has increased by 0.4°C to 0.8°C since the late 19th Century (IPCC, 2001). Zwiers (2002) estimated that the global mean temperature has risen by 0.6°C±0.2°C during the past century. Stott and Kettleborough (2002) estimated that the global mean temperature in the decade 2020–2030 will be 0.3°C–1.3°C higher than it was in 1990–2000 (5%–95% likelihood range). These results are unaffected by the choice of emission scenario used to make the projection.

However, the above result is less than Liu et al. (2004) and Qian and Zhu (2001). Using surface air temperature data from 305 Chinese stations for 1955–2000, Liu et al. (2004) found that daily maximum and minimum surface air temperatures increased significantly throughout China, particularly at the higher latitudes during the past 40 years. However, the secular changes of surface air temperature for different periods and regions in China, as well as their contributions to global warming, are completely unknown. In China, the secular change of surface air temperature was found to be basically consistent with that in the Northern Hemisphere, and a positive contributor to global warming due to the larger amplitude was found (Qian and Zhu, 2001).

Our final objective is to forecast the future values of temperature in the studied region of China. A good forecast must be close to the original values. Thus, given a series $\{X_t\}$, which has zero mean and is second order stationary, a sensible criterion would be to minimize the mean square error given by:

$$M(m) = E[X_{N+m} - \tilde{X}_N(m)]^2\ ,$$

where m is the point in the future at which we want to predict our series, and $\tilde{X}_N(m)$ is the prediction of X_{N+m}. Wei (1990) demonstrated that the optimum predictor is the conditional expectation of the future values given the past observations, i.e.:

$$\hat{X}_N(m) = E(X_{N+m}|X_N, X_{N-1}, X_{N-2}, \ldots)\ , \tag{5}$$

$\hat{X}_N(m)$ is usually read as the m-step ahead of forecast of X_{N+m} at the forecast origin N. The forecast error $e_N(m)$ is given by $e_N(m) = [X_{N+m} - \hat{X}_N(m)]$.

To forecast values from a fitted model, one has to assume that its parameters are known. By replacing the parameters c_0, c_1, and for $i = 1, \ldots, 12$ in model

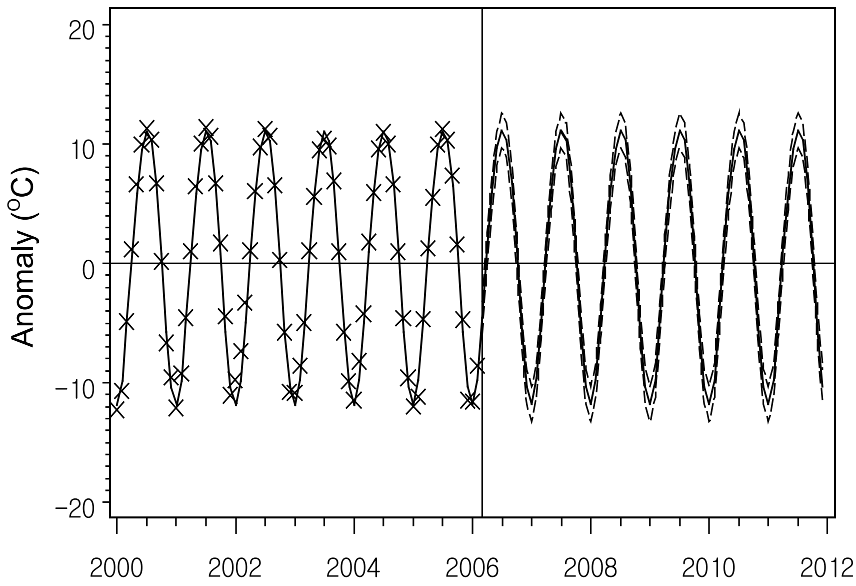

Fig. 5. Prediction for surface air temperatures (The line with crosses means actual, the solid thick line means predicted, the dashed thin line means predicted limit) in the studied region of China.

(1) by its estimates we can write

$$\hat{X}_t = Y_t - \hat{c}_0 - \hat{c}_1 t - \sum_{i=1}^{12} \hat{r}_i S_{it} ,$$

where $\{X_t\}$ is an ARMA(p, q) stationary process of the form

$$X_t + \hat{a}_1 X_{t-1} + \hat{a}_2 X_{t-2} + \cdots + \hat{a}_p X_{t-p} = e_t + \hat{b}_1 e_{t-1} + \hat{b}_2 e_{t-2} + \cdots + \hat{b}_q e_{t-q}$$

and $\{e_t, t = 1, \ldots, N\}$ is a Gaussian white noise process with mean zero and variance σ_e^2. The order of the ARMA model for each set of data has selected an ARMA(1,1) model for $\{X_t\}$. As shown above, the optimum predictor is given by Eq. (5). Thus, we have that, for ARMA(1,1), the one step ahead forecast is given by

$$X_N(1) = \hat{a}_1 X_N + \hat{b}_1[X_N - \hat{X}_{N-1}(1)] .$$

We obtain monthly forecasts for the surface air temperature data set. Each time, we predict one step ahead for the following month. We consider models (3) and (4) for the studied region of China. In Fig. 5 we plot the data of January 2000 to February 2006, and the predicted values for March 2006 to December 2011 for the set of data.

4. Concluding remarks

In this paper, the long-term warming trend and forecasting of surface air temperatures in a region of China has been analyzed by a decomposition method in time series analysis. The parameters were estimated for the sets of data and used later for prediction from March 2006 to December 2011.

The results were that statistically significant trends were not detected in the regional monthly anomalies temperature series. The surface air temperature trend per decade estimated for January 1948 to February 2006 was not statistically significant at 0.3434°C. This result agrees with Zwiers (2002) and Stott and Kettleborough (2002). Trends in surface air temperature for the studied region of China show interesting features that appear to have relevance to global change issues. However, it should be noted that there is a lot of uncertainty in this analysis because it came from a single realization of a single variable, and there existed some uncorrected problems in the NNR data, which need to be provided in a more detailed study. Thus, we will use other observations, e.g. Jones and Moberg (2003) and compare the results with those from this study. A regional climate model can be a promising way to reduce the uncertainty in the analysis of regional climate variation.

Acknowledgements. This paper was carried out using the Dong-A University Research Fund (2005).

REFERENCES

Akaike, H, 1977: On entropy maximization principle. *Application of Statistics*, P. R. Krishnaiah, Ed., North Holland, 27–41.

Brunetti, M., L. Buffoni, M. Maugeri, and T. Nanni, 2000: Trends of minimum and maximum daily temperatures in Italy from 1865 to 1996. *Theor. Appl. Climatol.*, **66**, 49–60.

Casey, K. S., and P. Cornillon, 2001: Global and regional sea surface temperature trends. *J. Climate*, **14**, 3801–3818.

IPCC, 2001: Climate Change 2001: *The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Houghton et al., Eds., Cambridge University Press, Cambridge, UK, 881pp.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223.

Jonsson, P., and K. Fortuniak, 1995: Interdecadal variations of surface wind direction in Lund, Southern Sweden 1741–1990. *International Journal of Climatology*, **15**, 447–461.

Kaiser, D. P., 2000: Decreasing cloudiness over China: An updated analysis examining additional variables. *Geophys. Res. Lett.*, **27**, 2193–2196.

Lee, J. H., B. Kim, K. T. Sohn, W. T. Kown, and S. K. Min, 2005: Climate change signal analysis for Northeast Asian surface temperature. *Adv. Atmos. Sci.*, **22**, 159–171.

Liu, B., M. Xu, M. Henderson, Y. Qi, and Y. Li, 2004: Taking China's temperature: Daily range, warming trends, and regional variations, 1955–2000. *J. Climate*, **17**, 4453–4462.

Ljung, G. M., and G. E. P. Box, 1978: The likelihood function of stationary autoregressive moving average models. *Biometrika*, **66**, 265–270.

Osborn, T. J., M. Hulme, P. D. Jones, and T. A. Basnett, 2000: Observed trends in daily intensity of United Kingdom precipitation. *International Journal of Climatology*, **20**, 347–364.

Qian, W., and Y. Zhu, 2001: Climate change in China from 1880 to 1998 and its impact on the environmental condition. *Climatic Change*, **50**, 419–444.

Stott, P. A., and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 719–723.

Toyooka, Y., 1977: A note on weak consistency of simple least squares estimators in polynomial regression with a nonstationary error process. *Keio Engineering Report*, **30**, 35–44.

Toyooka, Y., 1980: An asymptotically efficient estimation procedure in time series regression model with uniformly modulated error process. *Mathematica Japonica*, **25**, 525–532.

Wei, W., 1990: *Time Series Analysis: Univariate and Multivariate Methods*. Addison Wesley, 478pp.

Woodruff, S., R. Slutz, R. Jenne, and P. Steurer, 1987: A comprehensive ocean-atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 1239–1250.

Wu, Q., and D. M. Straus, 2004: AO, COWL, and observed climate trends. *J. Climate*, **17**, 2139–2156.

Zwiers, F. W., 2002: The 20-year forecast. *Nature*, **416**, 690–691.