

Data-Based Models for Global Temperature Variations

Bert W. Rust and Barend J. Thijsse

Reprinted from

CSC'07

Proceedings of the 2007 International Conference on
Scientific Computing

Editors: Hamid R. Arabnia, Jack Y. Yang, and Mary Qu Yang

CSREA Press

2007

ISBN: 1-60132-037-X

Pages 10-16

Proceedings of a WORLDCOMP'07 Conference held in Las Vegas, Nevada

June 25-28, 2007

Data-Based Models for Global Temperature Variations

Bert W. Rust

Math. and Comp. Sci. Division
Nat. Inst. of Standards and Tech.
Gaithersburg, MD, USA

Barend J. Thijsse

Department of Materials Sci. and Eng.
Delft University of Technology
Delft, Netherlands

Abstract *This paper presents two data-based models for the measured time series of global annual average tropospheric temperatures. One model is a smoothing spline fit designed to give an optimal separation of signal from noise. The other combines an optimal spline fit to the measured record of CO₂ concentration in the atmosphere with a previously reported 70 year cycle in the temperatures. It assumes a simple linear relation between changes in temperature and changes in the CO₂ concentration. When the cycle is added to the model, its fit to the temperature data is very similar to the optimal spline fit. The differences between the two fits are smaller in magnitude than the residuals for either one of them.*

Keywords: regression splines, global warming

1 Introduction

The observed increase in global average temperatures over the last 150 years is widely believed to be a consequence of the rising concentration of the greenhouse gas carbon dioxide in the atmosphere. Good measurements for both the atmospheric CO₂ concentrations and the temperature variations since 1856 are now readily available on the Web. We will examine the greenhouse hypothesis by using optimal spline fits to both records to construct a simple mathematical model relating them.

2 The spline2 algorithm

The optimal spline fits were obtained from the *spline2* algorithm [1, 2, 3] which uses a regression spline with the number and locations of the knots chosen to optimize the separation of the data record into signal and noise components. Since it does not assume anything about the mathematical form of

the underlying trend, *spline2* must base its judgement completely on the noise, i.e., on the residuals of the fit. It does so by considering a range of possible autocorrelation lengths ξ . For each ξ , *spline2* successively fits a large number of trial splines in the least squares sense to the data. Many of these trial splines have their knot locations optimized on the basis of the preceding fit. It then applies the Durbin-Watson test [4, 5, 6] to the weighted residuals of these fits. From all trial functions that pass this test, *spline2* selects the simplest one, i.e., the spline with the fewest number of knots. For this spline it computes a number α which measures the discrepancy between the autocorrelation function of the residuals and an assumed negative exponential autocorrelation function. Finally, the ξ value with the smallest α is selected and the corresponding best spline is taken to be the overall optimum spline.

3 The CO₂ concentrations

The measured atmospheric CO₂ concentrations are plotted in Figure 1. The South Pole data, which start in 1958, were direct measurements from the atmosphere made by Keeling and Whorf [7]. The ice core measurements by Etheridge, et. al. [8] and Neftel, et. al. [9] were derived from air bubbles trapped at different depths in the Antarctic ice sheet. They contain larger measuring errors than do the atmospheric measurements, but the good agreement with the atmospheric values in the interval of overlap (1958 to 1978) shows that the two kinds of data are consistent. All of these data are readily available from the web site <http://cdiac.ornl.gov/> which is maintained by the Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory.

The *spline2* fit shown in Figure 1 is a cubic regression spline using 6 knots located at times 1765.0, 1888.8, 1938.1, 1963.5, 1978.51, and 2004.5. This curve will be used in the following as the func-

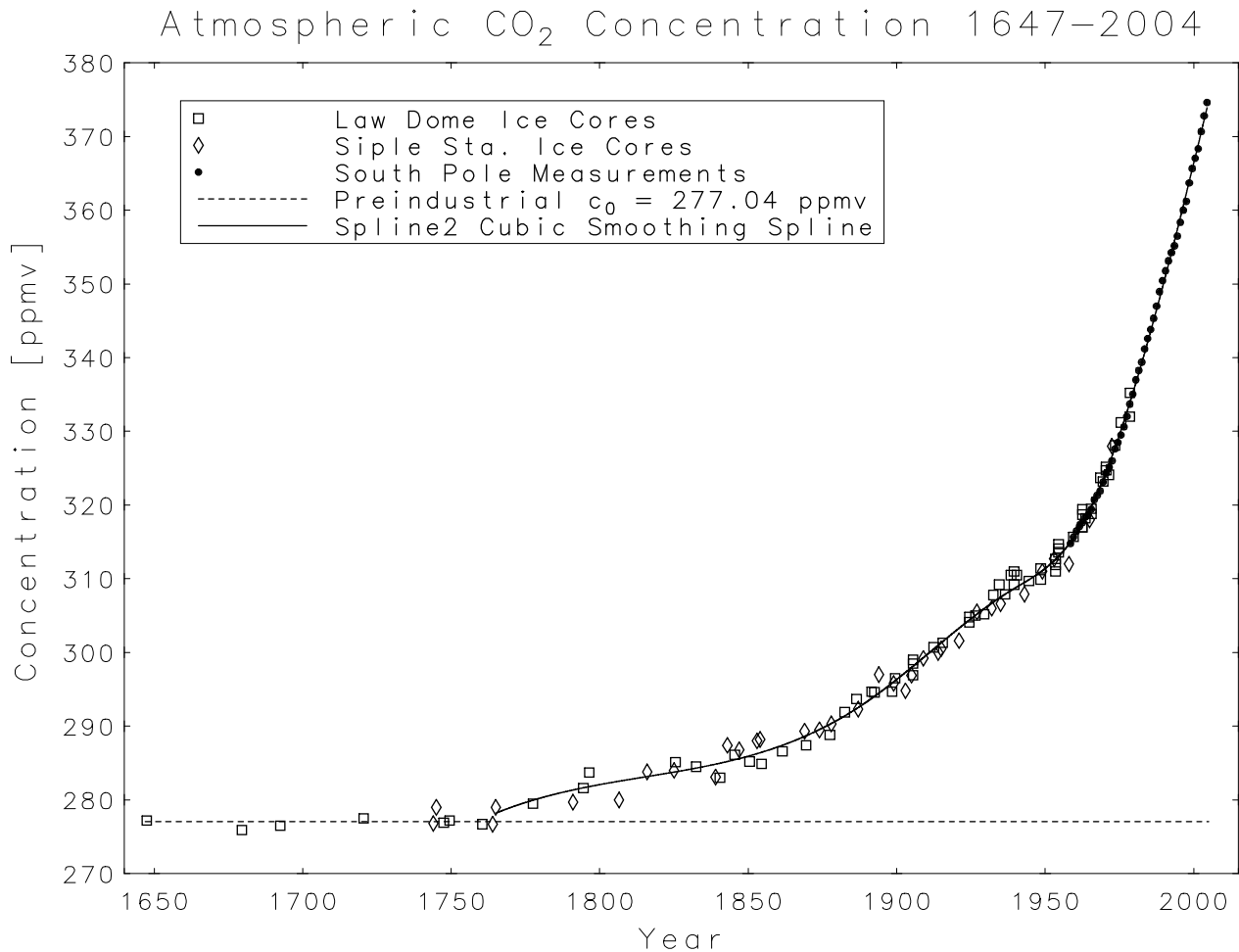


Figure 1: Annual average atmospheric CO₂ concentrations in parts per million by volume [ppmv]. The solid curve is the *spline2* fit for years 1765 to 2004. The measurements can be found at:

Law Dome <http://cdiac.ornl.gov/ftp/trends/co2/lawdome.combined.dat>

Siple Sta. <http://cdiac.ornl.gov/ftp/trends/co2/siple2.013>

South Pole <http://cdiac.ornl.gov/ftp/trends/co2/sposio.co2>

tional representation $c(t)$ of the measured concentrations. The horizontal dashed line in the plot gives the average value for the years 1647 to 1764, i.e., $c_0 = 277.04$ ppmv. This value will be used as an estimate of the preindustrial concentration.

4 The temperature anomalies

The measured time series of annual global average temperature anomalies [10, 11] is plotted in Figure 2. The *temperature anomaly*, for any given year, can be defined by

$$\begin{bmatrix} \text{Temp.} \\ \text{Anomaly} \\ \text{for Year} \end{bmatrix} \equiv \begin{bmatrix} \text{Average} \\ \text{Temp.} \\ \text{for Year} \end{bmatrix} - \begin{bmatrix} \text{Reference} \\ \text{Temp.} \end{bmatrix}.$$

The quantity of interest in global warming studies is temperature change, so it does not matter where the zero point is chosen. Absolute temperatures can be recovered from the anomalies by adding the reference temperature. For the anomalies plotted here the average temperature for the years 1961 to 1990 was used as the reference temperature.

The dashed curve in the Figure is the optimal *spline2* fit to the data. It used 4 knots at times 1856.5, 1905.83, 1955.17, and 2004.5. The sum of squared residuals for the fit was

$$\text{SSR} = \sum_{i=1}^{149} [T_{\text{obs}}(t_i) - T_{\text{spline}}(t_i)]^2 = 1.4374,$$

and the corrected total sum of squares for the data

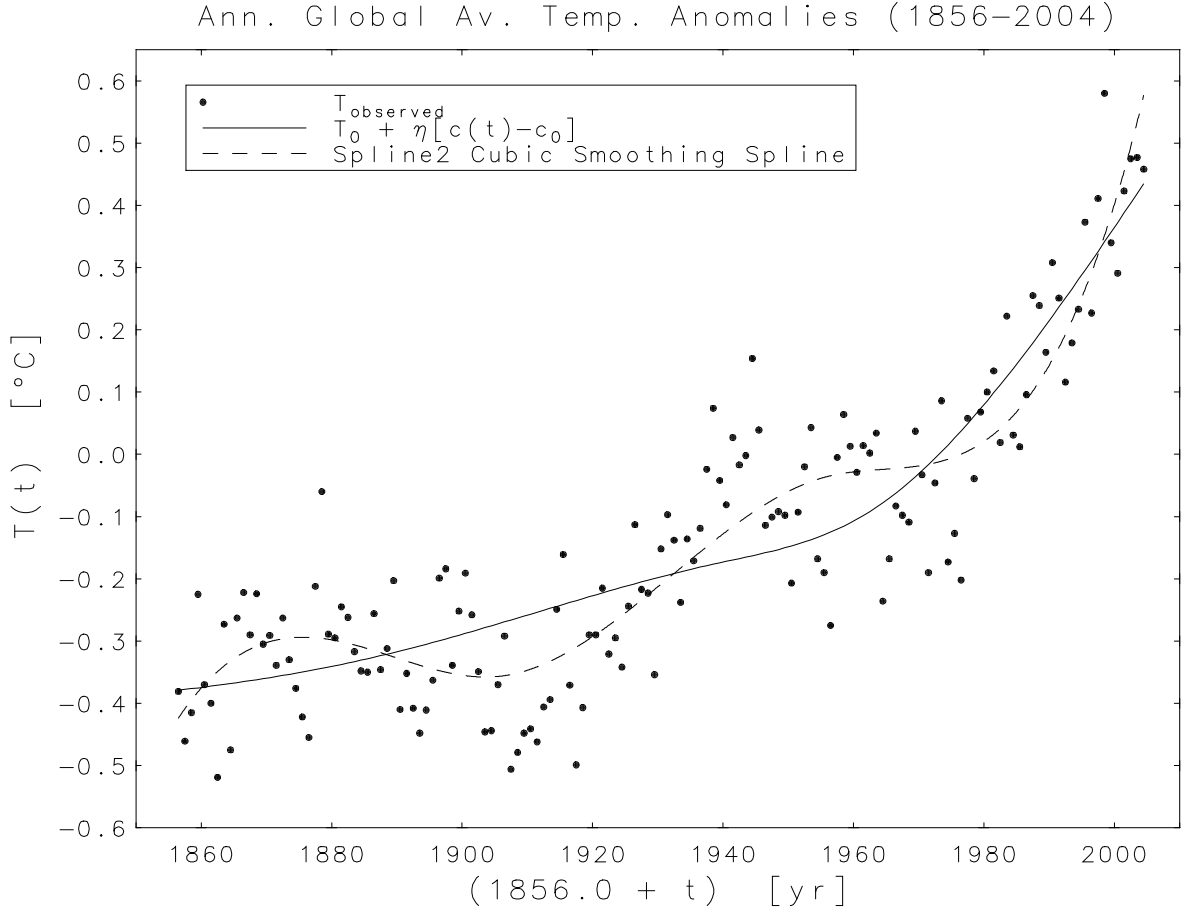


Figure 2: Annual global average temperature anomalies compiled by the Climatic Research Unit at <http://www.cru.uea.ac.uk/ftpdata/tavegl2v.dat>. The dashed curve is the *spline2* fit and the solid curve is the fit of the model (2) using the *spline2* fit from Figure 1 for $c(t)$.

was

$$\text{CTSS} = \sum_{i=1}^{149} [T_{\text{obs}}(t_i) - T_{\text{average}}]^2 = 8.5563,$$

so the coefficient of determination is

$$R^2 = 1 - \frac{\text{SSR}}{\text{CTSS}} = 0.8320.$$

Thus the fit explains 83 % of the variance in the data.

The solid curve in Figure 2 is the fit of a model based on the atmospheric CO_2 concentrations. The time scale was chosen so that $t = 0$ at epoch 1856.0. Let $T(t)$ be the temperature anomaly and $c(t)$ the atmospheric CO_2 concentration at time t . Assume that changes in the former are linearly proportional to changes in the latter, so

$$\frac{dT}{dt} = \eta \frac{dc}{dt}, \quad (1)$$

where η is the constant of proportionality. Integrating this equation gives

$$T(t) = T_0 + \eta[c(t) - c_0], \quad (2)$$

where c_0 is the preindustrial CO_2 concentration and T_0 is the corresponding temperature anomaly. Using $c_0 = 277.04$ ppmv, the least squares fit of this model gives the parameter estimates and standard uncertainties

$$\begin{aligned} \hat{T}_0 &= (-0.469 \pm .017) [^\circ\text{C}], \\ \hat{\eta} &= ([9.33 \pm .41] \times 10^{-3}) [^\circ\text{C}/\text{ppmv}], \end{aligned}$$

with

$$\text{SSR} = 1.8893, \quad \text{and} \quad R^2 = 0.7792.$$

These last two values are not as good as the corresponding ones for the spline fit, but still the model explains 78 % of the variance in the data and it

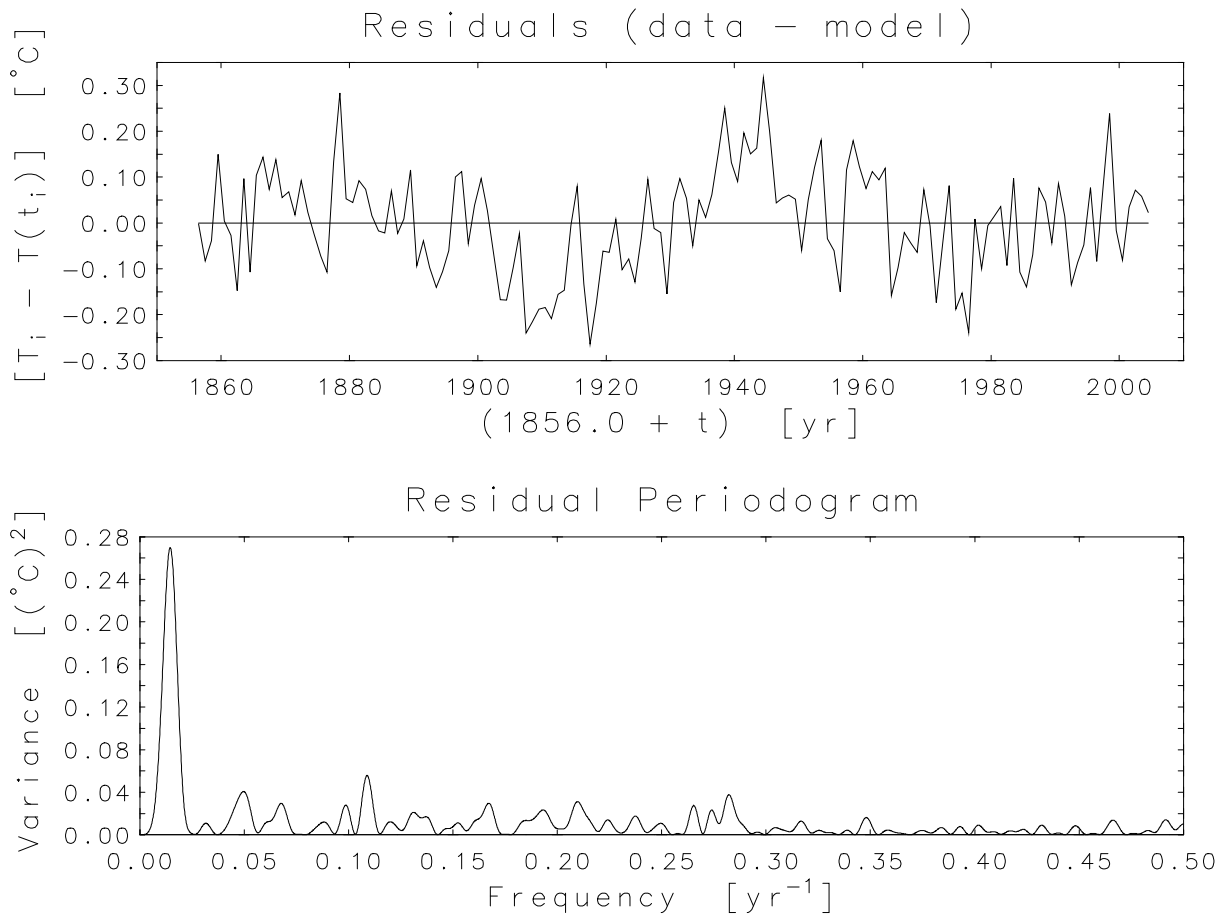


Figure 3: Upper Panel: The residuals for the solid curve fit in Figure 2.
Lower Panel: Fourier power spectrum of the above residuals.

tracks the data quite well. The spline fit undulates around the model fit in a quasi-periodic manner and a close inspection shows that the data do the same.

5 The temperature cycle

The residuals for the fit are plotted in the top panel of Figure 3, and the power (variance) spectrum for those residuals is plotted in the bottom panel. The dominant peak, which is centered at frequency 0.01453 yr^{-1} (period = 68.82 yr) was shown by both Fisher's test and the cumulative periodogram test to be statistically significant at the 95 % level. This is not surprising because this cycle was previously reported by Schlesinger and Ramankutty [12] who used a completely different technique to isolate it. They suggested that it "arises from predictable internal variability of the ocean-atmosphere system." Its presence in the data suggests a model

of the form

$$T(t) = T_0 + \eta[c(t) - c_0] + A \sin \left[\frac{2\pi}{\tau}(t + \phi) \right], \quad (3)$$

with free parameters T_0 , η , A , τ , and ϕ . The fit of this model is plotted as the solid curve in Figure 4. The parameter estimates and standard uncertainties are

$$\begin{aligned} \hat{T}_0 &= (-0.507 \pm .016) [^\circ\text{C}], \\ \hat{\eta} &= (0.01039 \pm .00042) [^\circ\text{C/ppmv}], \\ \hat{A} &= (0.099 \pm .012) [^\circ\text{C}], \\ \hat{\tau} &= (71.5 \pm 2.2) [\text{yr}], \\ \hat{\phi} &= (-1.0 \pm 1.4) [\text{yr}], \end{aligned}$$

and the corresponding estimate for the angular frequency of the cycle is

$$\hat{\omega} \equiv \frac{2\pi}{\hat{\tau}} = 0.0879 [\text{rad/yr}].$$

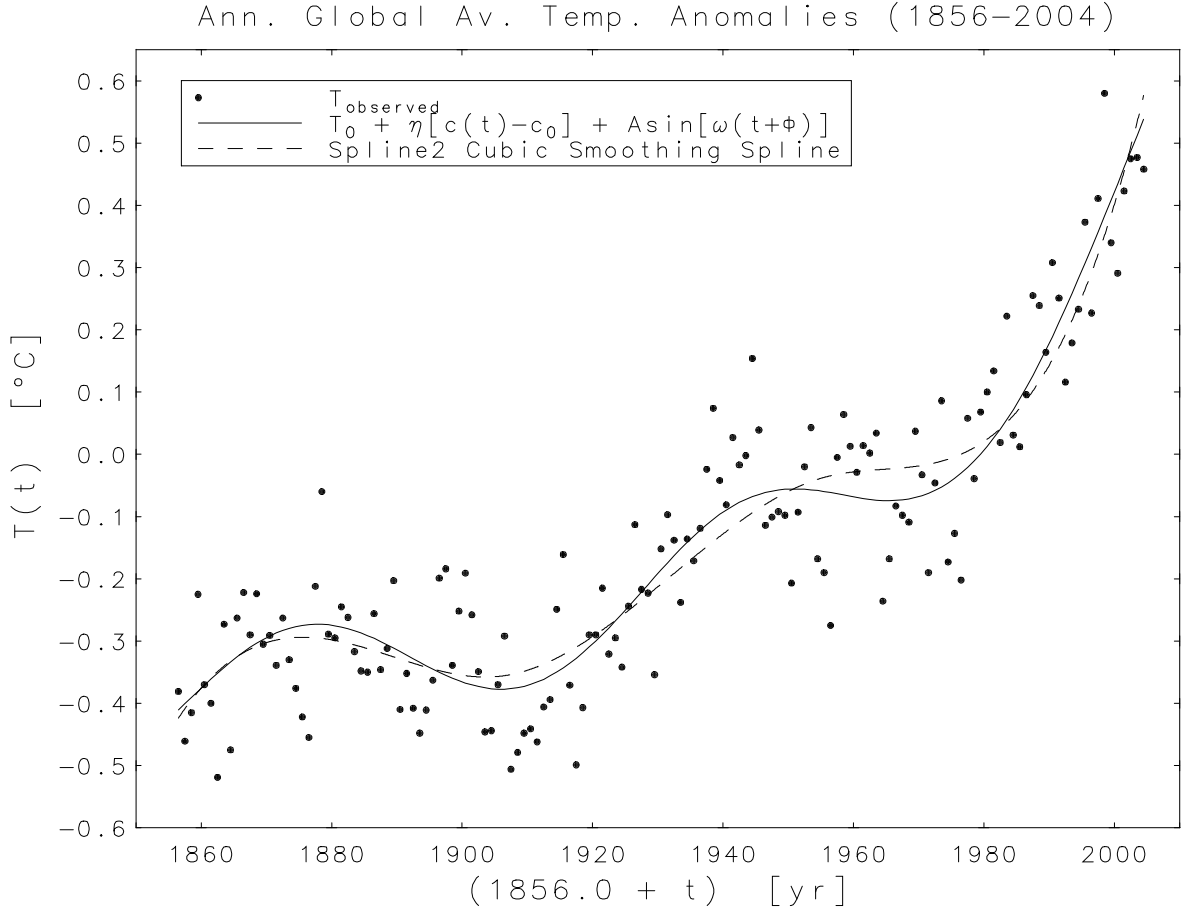


Figure 4: Two data-based fits to the temperature anomaly time series. The dashed curve is the *spline2* fit and the solid curve is the fit of the model (3) using the *spline2* fit from Figure 1 for $c(t)$.

The sum of squared residuals and coefficient of determination for the fit were

$$\text{SSR} = 1.2674 \quad \text{and} \quad R^2 = 0.8519 ,$$

so it explains $\approx 85\%$ of the variance in the data. This is a little better than the $\approx 83\%$ explained by the *spline2* fit which is also plotted in the Figure for comparison. The agreement between the two fits is actually quite good when one considers their very different origins. The magnitudes of the differences between the two are small in comparison to the magnitudes of the residuals for either of them. Those differences are plotted in Figure 5 together with the residuals for the model fit. A plot of the residuals for the spline fit would look very similar to those for the model. It seems fair to say that the differences between the two models are smaller than the noise in the data. This is a powerful argument for the correctness of both approaches to isolating the signal.

6 Conclusions

The results from fitting the model (3) suggest that the temperature anomaly record can be decomposed into three variance components

$$T_{obs} \equiv \text{Baseline} + \text{Sinusoid} + \text{Noise} ,$$

where

$$\text{Baseline} \equiv \hat{T}_0 + \hat{\eta} [c(t) - c_0] ,$$

$$\text{Sinusoid} \equiv \hat{A} \sin [\hat{\omega}(t + \hat{\phi})] , \quad \text{and}$$

$$\text{Noise} \equiv \text{residuals for the fit} .$$

These three variance components are plotted in Figure 6. An approximate analysis of variance shows that the Baseline, Sinusoid, and Noise account for $\approx 77\%$, $\approx 8\%$, and $\approx 15\%$ of the variance, respectively. The Baseline suggests that:

1. the atmosphere has warmed by $\approx 0.9^\circ\text{C}$ since 1856,

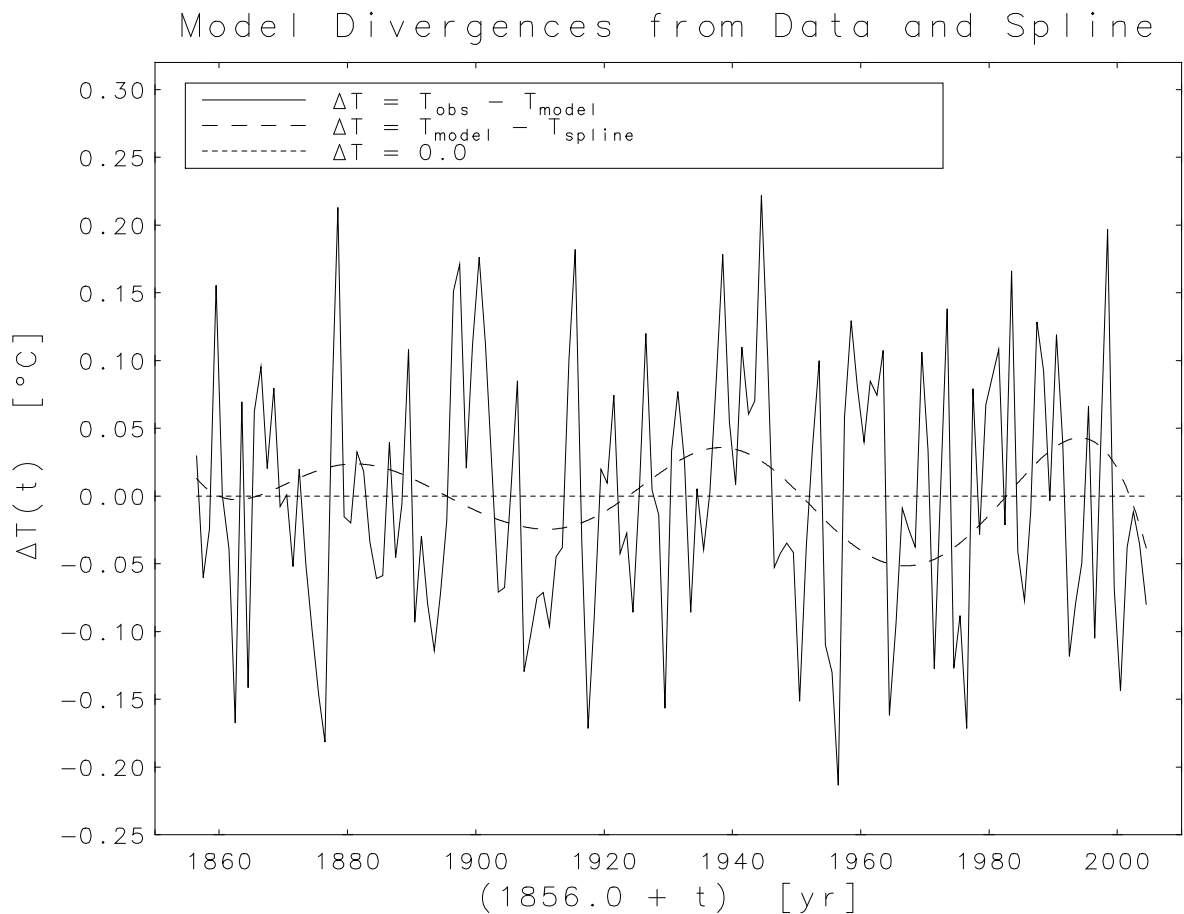


Figure 5: The residuals for the solid curve fit in Figure 4 and the difference between the solid curve and dashed curved fits. The figure would look much the same if the residuals for the dashed curve fit were plotted instead.

2. there is a linear relationship between the warming and the increase in atmospheric CO_2 in the years 1856 to 2004, and
3. the warming is accelerating.

The first half of the warming occurred in the years 1856 to 1975 and the second half in the years 1976 to 2004.

References

- [1] B. J. Thijsse, M. A. Hollanders, and J. Hendrikse, "A practical algorithm for least-squares approximation of data containing noise," *Computers in Physics*, vol. 12 (1998) pp. 393-399.
- [2] B. J. Thijsse, *Spline2 V6.0 Tutorial and User Manual – Unix version*, Delft University of Technology (Feb. 1, 2002) <http://www.3me.tudelft.nl/live/pagina.jsp?id=e919f062-a34d-4f20-8cd2-ae5dea65f942&lang=en>.
- [3] B. J. Thijsse and B. W. Rust, "Freestyle data fitting and global temperatures," *Computing in Science & Engineering*, in press (2007).
- [4] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression:I," *Biometrika*, vol. 37 (1950) pp. 409-428.
- [5] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression:II," *Biometrika*, vol. 38 (1951) pp. 159-177.
- [6] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression:III," *Biometrika*, vol. 58 (1971) pp. 1-19.
- [7] C. D. Keeling and T. D. Whorf, "Atmospheric CO_2 records from sites in the SIO air sampling network," in *Trends: A Compendium of Data*

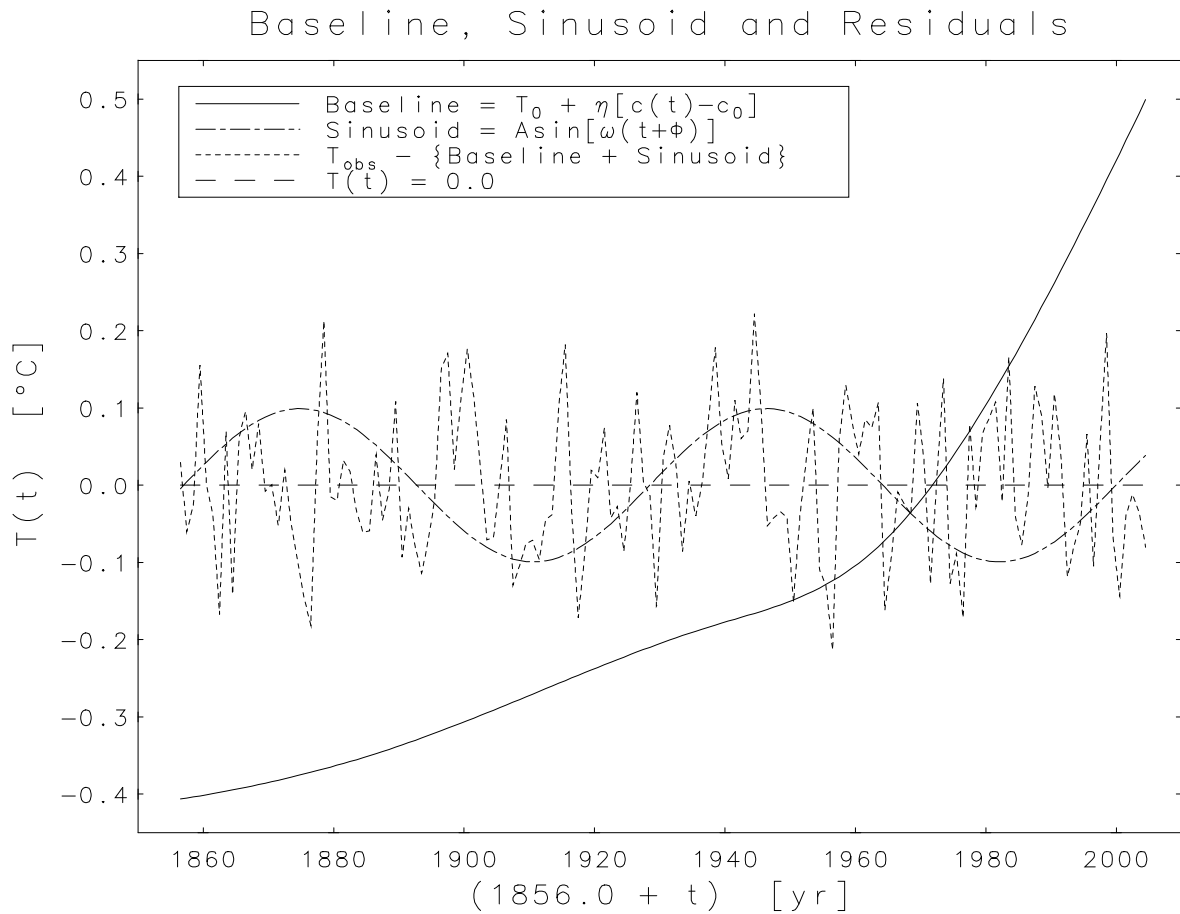


Figure 6: The components of the variance in the temperature anomaly time series.

- on *Global Change*, (Carbon Dioxide Information Analysis Center, Oak Ridge, 2005).
- [8] D. M. Etheridge, L. P. Steele, R. L. Langenfelds, R. J. Francey, J. M. Barnola, and V. I. Morgan, "Historical CO₂ records from the Law Dome DE08, DE08-2, and DSS ice cores," in *Trends: A Compendium of Data on Global Change*, (Carbon Dioxide Information Analysis Center, Oak Ridge, 2005).
- [9] A. Neftel, H. Friedli, E. Moor, H. Löttscher, H. Oeschger, U. Siegenthaler, and B. Stauffer, "Historical CO₂ record from the Siple Station ice core," in *Trends: A Compendium of Data on Global Change*, (Carbon Dioxide Information Analysis Center, Oak Ridge, 2005).
- [10] Jones, P. D., New, M., Parker, D. E., Martin, S., and Rigor, I. G. "Surface air temperature and its changes over the past 150 years," *Reviews of Geophysics*, vol. 37 (1999) pp. 173-199.
- [11] Jones, P. D. and Moberg, A. "Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001," *Journal of Climate*, vol. 16 (2003) pp. 206-223. <http://www.cru.uea.ac.uk/ftpdata/tavegl2v.dat>
- [12] Schlesinger, M. E. and Ramankutty, N. "An oscillation in the global climate system of period 65-70 years," *Nature*, vol. 367 (1994) pp. 723-726.