Separating Signal from Noise in Global Warming

Bert W. Rust

Reprinted from the CD


or

Separating Signal from Noise in Global Warming

Bert W. Rust

Mathematical and Computational Sciences Division
National Institute of Standards and Technology
100 Bureau Drive, Stop 8910
Gaithersburg, MD 20899-8910
bwr@cam.nist.gov

July 23, 2003

Abstract

One argument often used against global warming is that the global temperature record is too noisy to allow a clear determination of the signal. This paper presents two models for the signal which suggest that: (1) the warming is accelerating, (2) the warming is related to the growth in fossil fuel emissions, and (3) the warming in the last 146 years has been at least 10 times greater than the noise level. One model uses a constant rate for the acceleration and the other an exponential whose rate constant is exactly one half that of the growth in fossil fuel emissions. The two models can be viewed as best case and worst case scenarios for extrapolations into the future, but the data measured so far cannot reliably distinguish between them.

1 Introduction

Measurements by C. D. Keeling and his colleagues [7] at the Mauna Loa Observatory in Hawaii show that in the years 1959-2001 the atmospheric CO$_2$ concentration has risen from 316 to 371 parts per million by volume, an increase of 17.4%. The main source of this increase is thought to be fossil fuel emissions. Since CO$_2$ is a strong greenhouse gas, it is generally agreed that these additions to the atmosphere should produce some global warming, but there is considerable disagreement over the magnitude of the effect.

Figure 1 is a plot of annual global average tropospheric temperature anomalies (relative to the average for 1961-1990) for the years 1856-2001. These data were computed and tabulated by P. D. Jones and his colleagues [5, 6] at the Climatic Research Unit of the University of East Anglia. They can be obtained online at [http://www.cru.uea.ac.uk/cru/cru.htm](http://www.cru.uea.ac.uk/cru/cru.htm). The horizontal lines indicate the average anomaly ± one standard deviation, i.e., (−0.145 ± 0.231) °C. Before 1980, it was possible to argue that the apparent increase was too small, relative to the scatter in the data, to indicate a systematic trend, but measurements over the last two decades have removed much of the doubt about the warming.
2 Polynomial Models

Figure 2 gives a plot of the least squares fit of the straight line model
\[ T(t - t_0) = T_0 + C_1(t - t_0), \]  with \( t_0 = 1856.0, \) \hspace{1cm} (1)
where \( t \) is time (measured in years A.D.) and \( T \) is temperature anomaly (measured in \( ^\circ \text{C} \)). For all of the fits in this paper, the zero point of the time scale was shifted to epoch 1856.0, which is the beginning of the record, and each yearly average was assigned to the mid-point of its corresponding year. The parameters estimates were
\[ \hat{T}_0 = (-0.463 \pm 0.023) \ [^\circ \text{C}], \quad \hat{C}_1 = (4.36 \pm 0.28) \times 10^{-3} \ [^\circ \text{C/yr}]. \] \hspace{1cm} (2)

The residual time series, shown in Figure 3, exhibits a quasi-periodic oscillation about a concave upward base line. The last property is not obvious by inspection but is strikingly confirmed by a Fourier power (variance) spectrum. Figure 4 gives the periodogram of the residuals, truncated at frequency 0.10 yr\(^{-1}\) rather than 0.50 yr\(^{-1}\) because there were no high frequency features comparable to the two low frequency peaks. The periods corresponding to the centers of those peaks are \( \tau_0 \approx 143 \) years and \( \tau_1 \approx 62.5 \) years. Since the total length of the time series was only 146 years, the first peak is an obvious indicator that the straight line is an inadequate base line.

If the straight line, which represents warming with a constant rate \( C_1 \), will not do, then the next logical candidate for a base line is a quadratic
\[ T(t - t_0) = T_0 + C_1(t - t_0) + C_2(t - t_0)^2, \] \hspace{1cm} (3)
which represents warming with a constant acceleration \( C_2 \). The least squares estimates for this model were
\[ \hat{T}_0 = -0.311 \pm 0.031 \ [^\circ \text{C}], \quad \hat{C}_1 = (-1.88 \pm 0.97) \times 10^{-3} \ [^\circ \text{C/yr}], \] \[ \hat{C}_2 = (4.27 \pm 0.64) \times 10^{-5} \ [^\circ \text{C/yr}^2]. \] \hspace{1cm} (4)
Figure 2: Straight line fit to the global temperature anomalies.

Figure 3: Residual time series for the fit in Figure 2.
Figure 4: Truncated (in frequency) periodogram for the residuals in Figure 3.

The relatively large uncertainty in $\hat{C}_1$ suggests that a reduced quadratic model

$$T(t - t_0) = T_0 + C_2(t - t_0)^2$$

might be indicated by the data. The estimates for that fit were

$$\hat{T}_0 = -0.363 \pm 0.015 \ [^\circ C], \quad \hat{C}_2 = (3.06 \pm .16) \times 10^{-5} \ [^\circ C/yr^2],$$

and a formal F-test accepts the null hypothesis $C_1 = 0$ at the 95% level of significance [11]. Thus, the data seem to demand a monotonically increasing, accelerated warming. Figure 5 shows both fits, and Figure 6 shows the residual time series for the reduced quadratic fit.

Figure 7 gives the periodogram of the residuals in Figure 6. It is dominated by a single peak whose central frequency corresponds to a period $\tau_1 \approx 61.5$ years. Since there are $\approx 2.4$ repetitions of this cycle in the record, it almost certainly represents a real oscillation. Attempts [10, 11] to accommodate this variation by fitting higher order polynomials were ineffective for polynomials of order 3 and 4. That is, no statistically significant reduction in the sum of squared residuals was obtained, and the uncertainties in the parameter estimates were almost as large as the estimates themselves. That situation changed for a 5th order polynomial (6 free parameters) which was able to capture the variation, with statistical significance, but high order polynomial behavior is rare in nature whereas cycles are ubiquitous. And this particular cycle has previously been noted by Mitchell [9], who was using an older, cruder temperature record, and by Schlesinger and Ramankutty [16], who suggested that “the oscillation arises from predictable internal variability of the ocean-atmosphere system.” Even though the cause of the cycle is not known, it is very important because a cycle with the same period occurs in the record of fossil fuel CO$_2$ emissions which are widely thought to be driving the global warming.
Figure 5: Fits of the full quadratic model (3) and the reduced quadratic model (5) to the global temperature anomalies.

Figure 6: Residual time series for the reduced quadratic fit in Figure 5.
3 The Fossil Fuel Connection

Figure 8 is a plot of the annual global total fossil fuel emissions of CO$_2$, measured in millions of metric tons of carbon [MtC]. These data, compiled by Marland, Boden, and Andres [8], can be found at \url{http://cdiac.ornl.gov/trends/emis/em_cont.htm}. The solid curve is a nonlinear least squares fit of the model

\[ P(t - t_0) = \left\{ P_0 + P_1 \sin \left( \frac{2\pi}{\tau_1}(t - t_0 + \theta_1) \right) \right\} e^{\alpha(t - t_0)}, \]  

(7)

with free parameters $P_0$, $P_1$, $\alpha$, $\tau_1$, $\theta_1$. Although it looks complicated, it represents a simple exponential base line with a superposed sinusoidal oscillation whose amplitude is growing with the same exponential rate as the base line. The parameter estimates were

\[ \hat{P}_0 = 133.8 \pm 5.0 \text{ [MtC]}, \quad \hat{P}_1 = 25.4 \pm 1.6 \text{ [MtC]}, \]
\[ \hat{\alpha} = 0.02814 \pm 0.00034 \text{ [yr$^{-1}$]}, \quad \hat{\tau}_1 = 64.9 \pm 1.5 \text{ [yr]}, \]
\[ \hat{\theta}_1 = 26.6 \pm 2.7 \text{ [yr]}, \]

(8)

and the fit explained $100 \times R^2 = 99.73\%$ of the total variance, which is quite remarkable since the model had no provision for the temporary dislocations, quite visible in the plot, coinciding with the two World Wars, the Great Depression and the OPEC oil embargo.

Even more remarkable is the fact that the oscillation, with its sign reversed (i.e., with its phase shifted by exactly a half cycle), appears also in the global temperature record. The solid curve in Figure 9 was obtained by a linear least squares fit of

\[ T(t - t_0) = T_0 + C_2(t - t_0)^2 + A_1 \sin \left( \frac{2\pi}{64.9}(t - t_0 - 5.85) \right), \]

(9)

with free parameters $T_0$, $C_2$, $A_1$. The period of the sinusoid is exactly the same as the estimate $\hat{\tau}_1$ obtained by the nonlinear fit of (7) to the emissions data, and
Figure 8: Annual global total fossil fuel CO$_2$ emissions. The curve is the fit of the exponential/sinusoidal model (7).

Figure 9: Fits of the quadratic/sinusoidal model (9) and the exponential/sinusoidal model (10) to the global temperature anomalies.
the phase constant -5.85 has been set to make the oscillation exactly one half cycle ahead (or behind) the one obtained for that fit. Thus, maxima in the temperature oscillation correspond to minima in the emissions oscillation and vice versa. This inverse correlation has previously been noted by Rust and Kirk [15] and Rust and Crosby [14] who argued that it might represent a Gaiaen feedback by which increasing temperatures would cause reductions in fossil fuel production. That may be wishful thinking, however, and it must be admitted that the cause of the oscillation is unknown. But its occurrence in both records is strong evidence for a connection between the warming and the fossil fuel emissions.

We have seen that the temperature data seem to demand a monotonically increasing base line which we modelled with the reduced quadratic (5). But the data are fit just as well by an exponential base line whose rate constant is exactly one half of the \( \alpha \) obtained by fitting (7) to the emissions data. The dashed curve in Figure 9 was obtained by a linear least squares fit of

\[
T(t - t_0) = T_0 + C_2 \exp \left[0.01407(t - t_0)\right] + A_1 \sin \left[\frac{2\pi}{64.9} (t - t_0 - 5.85)\right],
\]

with free parameters \( T_0, C_2, A_1 \). The estimates for these parameters, together with those for the fit of (9), are given in Table 1. Those results and the two curves in the figure indicate that both models are equally acceptable, with each explaining approximately 80.75% of the variance in the record. But the two models will diverge when they are extrapolated into the future because the exponential base line will accelerate much faster than the quadratic. At the present time, they might be considered as best case and worst case scenarios for future warming.

The fact that the data are well modelled by an exponential with a rate exactly half that of the one for the fossil fuel emissions may be only a coincidence. If the constant rate 0.01407 \( \text{yr}^{-1} \) is replaced in (10) by a variable \( \beta \), and the resulting expression is fit, with free parameters \( T_0, C_2, A_1, \beta \), to the data, then the percentage of variance explained increases only to 80.94%, and the estimate for \( \beta \) is \( \hat{\beta} = (0.0165 \pm 0.0025) \text{ yr}^{-1} \). Since the rate 0.01407 lies just inside the \( \pm 1\sigma \) uncertainty interval for \( \hat{\beta} \), we can say that if an exponential base line is the correct model, then the data demand a rate approximately one half that for the fossil fuel
emissions, but it is hard to imagine a dynamical process connecting the two records that would exactly halve that rate.

The last two rows in Table 1 give the residual standard errors,

\[ \hat{\sigma}_r = \sqrt{\frac{1}{143} \sum_{i=1}^{146} (T_i - \hat{T}_i)^2}, \]  

(11)

and the net increases, over the length of the record, in the model predicted temperatures. If we take the former as a measure of the noise in the record and the latter as a measure of the total warming, then the two models predict that the total warming so far has been \( \approx 8.34 \) times greater than the noise level.

4 Completing the Model for the Signal

The fits in Figure 9 track the data well, but neither produces white noise residuals. The residuals for the quadratic model are given in Figure 10, and the periodogram and cumulative periodogram [3, Chapt. 7] of those residuals are given in Figure 11. The two diagonal lines plotted with the cumulative periodogram define the 95% Kolmogorov-Smirnov bounds for white noise. Since the cumulative periodogram lies outside these bounds at \( \approx 34\% \) of the frequencies, the white noise hypothesis is resoundingly rejected. Comparing the periodogram with the one in Figure 4 shows that fitting (9) completely removed the two dominant peaks and that the two insignificant looking bumps at frequencies \( \approx 0.05 \) and \( \approx 0.10 \) yr\(^{-1} \) in Figure 4 reappear as the two most significant peaks in Figure 11. These peaks are responsible for most of the cumulative periodogram’s excursion outside the 95% white noise band.

The detection of temperature oscillations with periods 21, 16, 6, and 5 years has been reported by Ghil and Vautard [4], but their findings were sharply challenged.
Figure 11: Periodogram and cumulative periodogram of the residuals in Figure 10. Note the large difference between the vertical scale of the periodogram here and the one in Figure 4.

by Elsner and Tsonis [2]. The debate was continued, inconclusively, by Allen, Read and Smith [1] and by Tsonis and Elsner [17]. The principle guiding the current effort was to incorporate into the models (9) and (10) just enough of the cycles indicated in Figure 11 to reduce the residuals to white noise. To do so, it was necessary to include the three largest peaks which were at frequencies 0.0487, 0.109, and 0.210 yr\(^{-1}\), i.e., at periods \(\tau_2 = 20.5\), \(\tau_3 = 9.22\), and \(\tau_4 = 4.77\) yr. This gave, for the quadratic base line, the model

\[
T(t - t_0) = T_0 + C_2(t - t_0)^2 + A_1 \sin \left[\frac{2\pi}{64.9}(t - t_0 - 5.85)\right] \\
+ A_2 \sin \left[\frac{2\pi}{20.5}(t - t_0 + \theta_2)\right] \\
+ A_3 \sin \left[\frac{2\pi}{9.22}(t - t_0 + \theta_3)\right] \\
+ A_4 \sin \left[\frac{2\pi}{4.77}(t - t_0 + \theta_4)\right],
\]

(12)

with 6 additional free parameters \(A_2, \theta_2, A_3, \theta_3, A_4, \theta_4\). A similar expression (17) can be written for the exponential base line model. Both models can be fit by linear least squares, and the estimates \(\hat{B}_j\) and \(\hat{D}_j\) can be converted into estimates of \(A_j\).

\[
A_j \sin \left[\frac{2\pi}{\tau_j}(t - t_0 + \theta_j)\right] = B_j \sin \left[\frac{2\pi(t - t_0)}{\tau_j}\right] + D_j \cos \left[\frac{2\pi(t - t_0)}{\tau_j}\right],
\]

(13)

where the \(B_j\) and \(D_j\) are equivalent (to \(A_j\) and \(\theta_j\)), new free parameters defined by

\[
B_j = A_j \cos \left(\frac{2\pi\theta_j}{\tau_j}\right), \quad D_j = A_j \sin \left(\frac{2\pi\theta_j}{\tau_j}\right), \quad j = 2, 3, 4.
\]

(14)

Since the new parameters appear linearly in the model, the fit can be done by linear least squares, and the estimates \(B_j\) and \(D_j\) can be converted into estimates of \(A_j\).
Global Warming

and $\theta_j$ by the inverse relations

$$\hat{A}_j = +\sqrt{\hat{B}_j^2 + \hat{D}_j^2}, \quad \hat{\theta} = \frac{\pi \theta_j}{2\pi} \tan^{-1} \left( \frac{\hat{D}_j}{\hat{B}_j} \right), \quad j = 2, 3, 4.$$  \hspace{1cm} (15)

More details on the use of these transformations and on converting the uncertainties in the $\hat{B}_j$ and $\hat{D}_j$ to equivalent uncertainties in the $\hat{A}_j$ and $\hat{\theta}_j$ can be found in two recent tutorial papers by Rust [12, 13].

Some of the parameter estimates and statistics for the fit of the expanded model (12) are given in Table 2, together with the estimates for the simpler model (9) for comparison. The estimated phase shifts $\hat{\theta}_2$, $\hat{\theta}_3$, $\hat{\theta}_4$ were not included because

Table 2: Parameter estimates and some diagnostics for the models (9) and (12).

<table>
<thead>
<tr>
<th>Param.</th>
<th>3-Param. Model</th>
<th>9-Param. Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_0$</td>
<td>$-0.380 \pm .013$ [°C]</td>
<td>$-0.382 \pm .012$ [°C]</td>
</tr>
<tr>
<td>$C_2$</td>
<td>$(3.27 \pm .14) \times 10^{-5}$ [°C/yr$^2$]</td>
<td>$(3.28 \pm .12) \times 10^{-5}$ [°C/yr$^2$]</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$0.105 \pm .013$ [°C]</td>
<td>$0.106 \pm .011$ [°C]</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$0.041 \pm .011$ [°C]</td>
<td></td>
</tr>
<tr>
<td>$A_3$</td>
<td>$0.042 \pm .011$ [°C]</td>
<td></td>
</tr>
<tr>
<td>$A_4$</td>
<td>$0.033 \pm .011$ [°C]</td>
<td></td>
</tr>
<tr>
<td>$SSR$</td>
<td>$1.4975$ [°C$^2$]</td>
<td>$1.1576$ [°C$^2$]</td>
</tr>
<tr>
<td>$100R^2$</td>
<td>$80.71%$</td>
<td>$85.09%$</td>
</tr>
<tr>
<td>$\dot{\sigma}_r$</td>
<td>$0.1023$ [°C]</td>
<td>$0.09192$ [°C]</td>
</tr>
<tr>
<td>$T_{146} - \hat{T}_1$</td>
<td>$0.8300$ [°C]</td>
<td>$0.9193$ [°C]</td>
</tr>
</tbody>
</table>

their uncertainties are not reliable indicators of the significance of the cycles. The estimated amplitudes $\hat{A}_2$, $\hat{A}_3$, $\hat{A}_4$ and their uncertainties do indicate the strength of the cycles, and in all three cases the amplitudes are at least 3 times larger than their corresponding uncertainties. This suggests statistical significance, and the standard F-test confirms that the 6 new free parameters do produce a significant reduction in the sum of squared residuals. Let the null hypothesis be $H_0: A_2 = A_3 = A_4 = \theta_2 = \theta_3 = \theta_4 = 0$. Then, with $m = 146$, $n = 9$, and $k = 6$ we have

$$u = \frac{(SSR)_H - (SSR)_F}{(SSR)_F} \times \frac{m - n}{k} = 6.703 > 2.165 = F_{0.95}(k, m - n),$$  \hspace{1cm} (16)

so $H_0$ is rejected.

The fit is shown in Figure 12. Although the three new oscillations are not readily discernable in the data, they are quite obvious in the fit, so they look somewhat contrived. But a good argument for their reality is the fact that including them reduces the residuals, shown in Figure 13, to normally distributed white noise.
274

B. W. Rust

Figure 12: Fit of the quadratic base line model (12) to the temperature anomalies.

Binning the residuals in a histogram and applying the $\chi^2$ test showed them to be acceptably normal, and estimating their autocorrelation function revealed no significant correlations at lags greater than 0. The periodogram and cumulative periodogram are shown in Figure 14. The cumulative periodogram lies outside the 95% Kolmogorov-Smirnov bounds at only 3.60% of the frequencies, so the residuals are an acceptable realization of normally distributed white noise.

Similar results are obtained for the exponential base line model,

$$T(t-t_0) = T_0 + C_2 \exp \left[0.01407(t-t_0)\right] + A_1 \sin \left[\frac{2\pi}{64.9}(t-t_0-5.85)\right] + \sum_{j=2}^{4} A_j \sin \left[\frac{2\pi}{\tau_j}(t-t_0+\theta_j)\right].$$

The fit, which is graphically almost indistinguishable from the one in Figure 12, explains 85.22% of the variance in the data. The cumulative periodogram of the residuals strays outside the 95% white noise bounds at only 3.23% of the frequencies. The estimated residual standard error is $\hat{\sigma}_r = 0.09152$ °C, and the estimated total warming is $\hat{T}_{146} - \hat{T}_1 = 0.9634$ °C, so for both models the total warming for the period 1856-2001 has been at least 10 times greater than the noise level in the record.

5 Extrapolating into the Future

Both models (12) and (17) reduce the temperature anomaly residuals to normally distributed white noise, so both fits are adequate representations of the signal. The two fits differ very little over the period 1856-2001, but, as can be seen in Figure 15, they diverge quickly when extrapolated into the future. The last minimum of the 64.9 year oscillation occurred at epoch 1975.4, so the period 1975-2001 corresponds to a time when the cycle was reinforcing the base line increase. This condition will continue, at an abated rate, until the cycle’s next maximum at epoch 2007.9. For
Figure 13: Residual time series for the fit in Figure 12.

Figure 14: Periodogram and cumulative periodogram of the residuals in Figure 13.
the next 32.5 years after that, the cycle will decline and partially offset the baseline increase, an effect clearly visible in the quadratic extrapolation, but less so in the exponential extrapolation. After that the exponential increases so rapidly that the oscillation is scarcely discernible.

These extrapolations assume that the dynamics causing the signals will continue unchanged into the future. Fossil fuel reserves are probably sufficient to support a continuation of the exponential growth observed in Figure 8, though doing so might require massive switches from petroleum fuels back to coal and its derivatives. The causes of the 64.9 year oscillations are unknown, but their persistence and coherence in both records for 2.25 cycles certainly suggest stability. But there remains the possibility that warming might trigger other events which would produce positive feedbacks. For example, an extensive thawing of the permafrost in the Arctic and/or the ignition of peat in the soil by hot forest fires in the tropical and temperate regions might add new transfers of CO$_2$ to the atmosphere which would not be subject to human control. Such events might have disastrous consequences, especially if the exponential baseline scenario is correct.

6 Acknowledgments

The author would like to thank Drs. Isabel Beichl, R. F. Boisvert and Anastase Nakassis for their helpful reviews and suggestions for improving this paper.

References


