Statistical Data Analysis Assignment
Climate Variation

Jonathan Newnham

June 2, 2010

1 Introduction

In this assignment we investigate daily weather observations from Alice Springs, Woomera and Charleville from 1950–2006. We attempt to correlate the increase in CO_2 in the atmosphere with a change in the effective Stefan's constant for the earth by fitting the difference between the maximum and minimum temperatures overnight.

2 Scripts and Analysis Tools

Matlab is used to perform the analysis. The scripts I used are available at http://www.ph.unimelb.edu.au/~jnnewn/EM/. Most of the analysis is performed by the script in asgt.m, which calls other functions as necessary. This script is divided into cells so that parts of it can be run separately.

3 Initial Analysis

3.1 Import

To read the data in from the provided text files, I used the ezread function. The script import_all.m imports each data file. It calls the function in import_data.m, which uses ezread and hoursdarkness. Hoursdarkness is a function to calculate how many hours of darkness there will be on a given date at a given latitude.

3.2 Data presence

To see which years have reasonably complete data, a script reports how many valid rows of data there are in each year. This script is in the file valid_years.m. Running this file shows that reasonably complete data begins in the year 1952.

3.3 Some Quick Histograms

For this year, we calculate sample means and standard deviations using the following formula:

$$\hat{\mu} = \frac{\sum_i x_i}{N} \qquad \text{(sample mean)}$$

$$s = \sqrt{\frac{\sum_i (x_i - \mu)^2}{N-1}} \qquad \text{(sample standard deviation)}$$

where we have used $N - 1$ instead of N as the usual Bessel's correction to reduce the bias of the standard deviation estimator.

The first cell (lines 3–26 of asgt.m) perform this calculation using the function plotHist. PlotHist uses the matlab functions mean and std to calculate the above quantities. The results are given in Figures 1–3. The quantities and their sample means and standard deviations are given in Table 1.

The histograms look reasonably Gaussian, with the exception of Cloudiness which is heavily biased toward 0 for Alice Springs and decidedly non-Gaussian for the other data sets.

3.4 Correlations

The sample correlation coefficient between two data sets X and Y is defined as

$$r_{xy} \equiv \frac{\sum_i (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{(n-1)\, s_x s_y}.$$

It is a good indicator of a linear dependence between X and Y. This calculation is implemented by the matlab function corrcoef.

The second cell (lines 35–66) of asgt.m calculates Table 2 and plots Figures 4–6. This code uses the function plotCorr.m which performs the calculation and plots the graph.

Very strong correlations are observed between maximum and minimum temperature. The correlations are slightly stronger when the max to min temperature change is considered (i.e. maximum - next day minimum, or dT2).

Quantity	Alice Springs	Charleville	Woomera
Minimum temperature	12.853 ± 7.281	12.812 ± 7.423	11.552 ± 5.462
Maximum temperature	27.831 ± 7.830	27.207 ± 7.427	24.812 ± 7.547
Max temp - (min temp for the same day) (dT1)	14.978 ± 4.620	14.395 ± 4.186	13.259 ± 4.035
Max temp - (min temp for next day) (dT2)	14.966 ± 3.841	14.408 ± 3.449	13.273 ± 3.638
Average overnight cloudiness (C)	2.533 ± 2.527	2.977 ± 2.403	3.558 ± 2.375
Average overnight relative humidity (H)	33.204 ± 17.429	42.553 ± 18.235	45.562 ± 16.775

Table 1: Sample Means and Standard Deviations for 1952.

The correlation between the change in temperature and the length of the night is negative, as would be expected – with less time to cool (longer night), the change in temperature will be smaller. Strong negative correlations are also observed between the overnight cloudiness and humidity. This is logical because the increased water vapour in the atmosphere acts as a blanket and reduces the temperature change. Cloudiness and humidity are also strongly correlated.

Interestingly, the correlations for Woomera are significantly different to the other data sets. This indicates something unusual about that data set, such as different data collection conditions.

3.5 Monte Carlo methods

In order to determine if there is a statistically significant trend in the temperature data, we use Monte Carlo resampling from the assumed Gaussian distributions (see Figures 1–3) of the minimum temperature, the maximum temperature and dT2. This allows us to determine the variance of the experimentally measured mean temperatures[1] for each year and hence determine a variance for the computed fit.

The code to perform this procedure is in the file MonteCarlo.m. Cell 3 in asgt.m (lines 68-110) sets up the data and plots and saves the results. The fit is performed by the file york_fit.m.

Figures 7–9 show the results. For most minimum and maximum temperatures, there is a statistically significant increase.

Studying these figures, we see that changing the size of the interval (using 1- or 5-year blocks) has little effect on the error in the gradient, as one would expect (it is effectively the same statistical analysis). There is usually a small effect on the actual computed value of the gradient, but the two values (for 1- or 5-year blocks) are well within each others' error bounds.

We also see that using 5-year blocks usually results in smaller standard devations for each point (as we are Monte-Carlo sampling five times as many points, with a correspondingly smaller variation in the mean. This does not hold true for dT2, indicating a wider variance for this computed value.

[1] which is different to (considerably smaller than) the variance of the data itself, of course

4 Model theory

In this section we develop a simple theory of how the temperature changes from the maximum during the day to the minimum at night. We begin with the Stefan-Boltzmann Law:

$$P = \sigma T^4 \tag{4.1}$$

where P is the power radiated by a black body, $\sigma \sim 5.67 \times 10^{-8}\mathrm{Wm}^{-2}\mathrm{K}^{-4}$ is the Stefan-Boltzmann constant, and T is the temperature of the black body.

To allow for the emissivity of the surface, we consider a modified σ,

$$\sigma' = \varepsilon\sigma$$

where ε is the emissivity of the earth. Our model for σ' is

$$\sigma' = 0.5\,(\mathrm{S0} + \mathrm{S1}\times\mathrm{C} + \mathrm{S2}\times\mathrm{H})\,\sigma$$

which allows for first-order effects of cloudiness and humidity on the emissivity. The 0.5 appears because this will ensure that S0 is approximately 1, as the surface of the earth only cools for half the time (the rest of the time its temperature is increasing as a result of being in sunlight).

We assume that the basic equation that governs how the atmosphere cools at night is

$$Q = C_H \Delta T \tag{4.2}$$

Quantity	Alice Springs	Charleville	Woomera
Minimum and maximum temp on the same day	0.8155	0.8411	0.8553
Maximum temp and minimum temp the next day	0.8734	0.8915	0.8928
dT1 and the number of hours of darkness	-0.0726	-0.1010	**-0.4788**
dT2 and the number of hours of darkness	-0.0741	-0.1095	**-0.5253**
dT2 and C	-0.5862	-0.6252	**-0.2904**
dT2 and H	-0.5342	-0.5283	-0.6161
H and C	0.4275	0.4451	0.3131

Table 2: The normalised correlation coefficient between the various data sets for 1952.

where Q is the amount of heat radiated,

$$C_H = (1 + C1 \times H + C2 \times C)\, 8 \times 10^5 \mathrm{JK}^{-1}$$

is the heat capacity of air (taking humidity and cloudiness into account), and ΔT is the change in temperature. We approximate Q by

$$\begin{aligned} Q &= \int_0^N P(t)\, dt \\ &= \int_0^N \sigma' T(t)^4\, dt \end{aligned}$$

where N is the number of seconds between T_{max} and T_{min} (we will use the length of the night for N).

Approximating

$$T(t) = T_{\mathrm{max}} - \frac{\Delta T\, t}{N}$$

(i.e. a linear drop in temperature) and using the binomial expansion (with T in Kelvin, $\Delta T \ll T$), we work through to

$$Q = \sigma' N T_{\mathrm{max}}^4 \left(1 - 2\frac{\Delta T}{T_{\mathrm{max}}}\right).$$

Inserting this expression into (4.2) yields

$$C_H \Delta T = \sigma' N T_{\mathrm{max}}^4 \left(1 - 2\frac{\Delta T}{T_{\mathrm{max}}}\right)$$

and now solving for ΔT:

$$\begin{aligned} C_H \Delta T &= \sigma' N T_{\mathrm{max}}^4 - 2\sigma' N \Delta T T_{\mathrm{max}}^3 \\ \sigma' N T_{\mathrm{max}}^4 &= \Delta T \left(C_H + 2\sigma' N T_{\mathrm{max}}^3\right) \\ \Delta T &= \frac{\sigma' N T_{\mathrm{max}}^4}{\left(C_H + 2\sigma' N T_{\mathrm{max}}^3\right)} \\ T_{\mathrm{max}} - T_{\mathrm{min}} &= \frac{\sigma' N T_{\mathrm{max}}^4}{\left(C_H + 2\sigma' N T_{\mathrm{max}}^3\right)} \\ T_{\mathrm{min}} &= T_{\mathrm{max}} - \frac{\sigma' N T_{\mathrm{max}}^4}{\left(C_H + 2\sigma' N T_{\mathrm{max}}^3\right)}. \end{aligned} \tag{4.3}$$

Equation (4.3) is the model that we will fit the data to, by finding the best values of S0, S1, S2, C1 and C2.

5 Fitting to the model

The code in Cell 4 (lines 112-218 of asgt.m) fits various-duration blocks of data to this model, and then fits a trendline through the resulting sequence of S0 values to determine if there is a statistically significant trend in the measured radiative forcing of the earth during 1950–2006. The results are given in Figures 11–14. Some discussion of these results is given in Section 6.

6 Results and Analysis

At a temperature of 300K, Equation (4.1) gives the power emission from the Earth's surface as about 459Wm^{-2}. From Figure 15, we see that the amount by which CO_2 prevents emission through the atmosphere is about 1.5Wm^{-2}. Figure 17 shows results from the ModTran modelling package. This indicates that a change from 310–385ppm (Figure 15) will result in a reduction of about 0.9Wm^{-2} to the radiative forcing component. (The values from ModTran are lower because it simulates conditions at an altitude of

70km, at which height the atmosphere is considerably cooler.)

As a result, to be sensitive to the effect of CO_2 levels changing, the reduction in S0 will be about $0.9/459 = 0.2\%$. This corresponds to an expected gradient of S0 with respect to years of about

$$\begin{aligned} g_{S0} &= -\frac{0.2\%}{(2006-1950)} \times \underbrace{1.5}_{\text{initial value for S0}} \\ &= -5 \times 10^{-5} \text{ per year} \end{aligned}$$

This is quite a small result. The observed gradients for the various data sets are presented in Table 3.

From these results we can see that we need about an order of magnitude improvement in the uncertainty to be able to observe the effect of carbon dioxide on the temperature.

Also, it is important to note that other trends in the data are apparent. Our model does not take these into account and they swamp the effect we are looking for. For example, the best fit to the Alice Springs data indicates a significant increase in S0 (i.e. the region now cools *faster* at night than it did in 1950. This could be due to many things, including changing weather patterns, a change in the colour of the surface (becoming darker and hence a better emitter), increased vegetation or housing development changing the emissivity of the surface.

In order to reduce the error margin, it could be important to measure the temperature on both sides of the CO_2 barrier, i.e. at ground level and at 70km. This would be expensive. Several measurements of the temperature at the same time for each day would also help fit to the model better.

7 Conclusion

Our analysis is inconclusive as there were too many unknown errors which swamped the signal we were looking for. Clear trends in the emissivity of the earth for different locations were observed, in different directions for different locations, indicating other effects our model did not take into account.

Data Set	g_{S0} ($\times 10^{-4}$ per year)	uncertainty (2s.d.)
Alice Springs	30	6
Charleville	0	9
Woomera	-16	10
Bourke Post Office	-20	16

Table 3: . S0 gradients. The expected value corresponding to CO_2 levels increasing is about 0.5 to 1. These figures come from Figures 11–14.

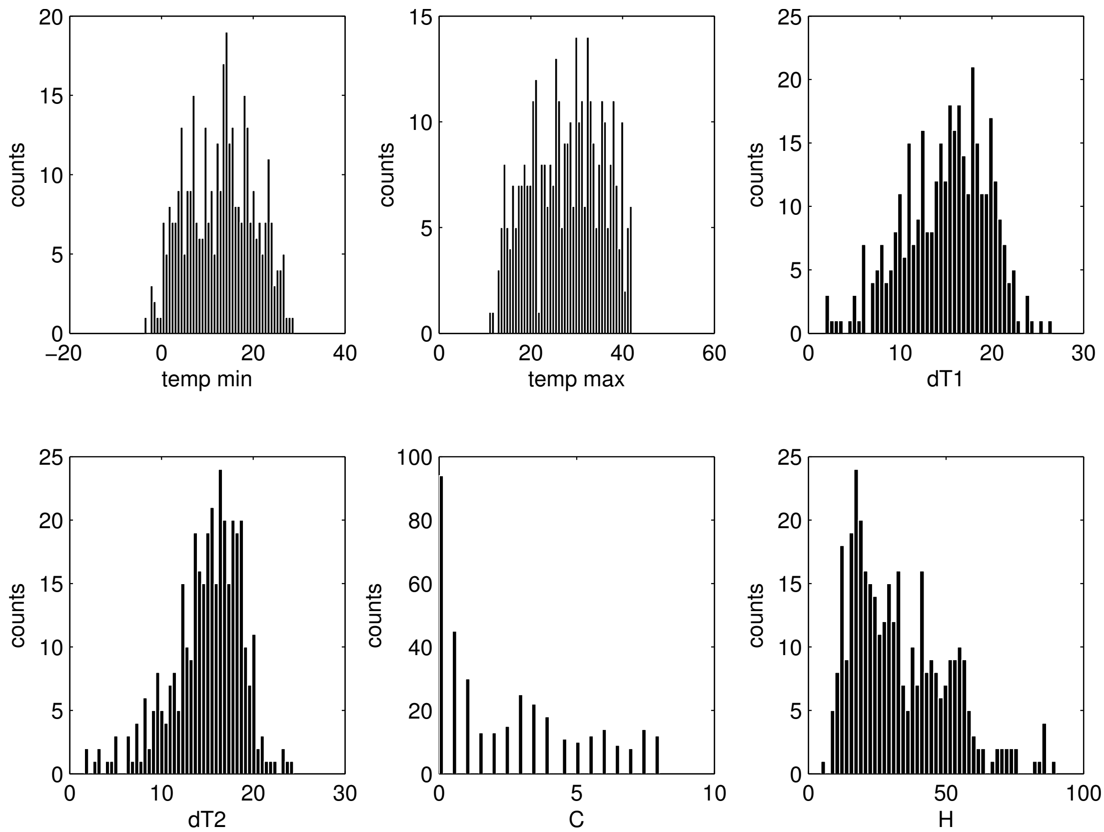


Figure 1: Histograms for Alice Springs data set for 1952.

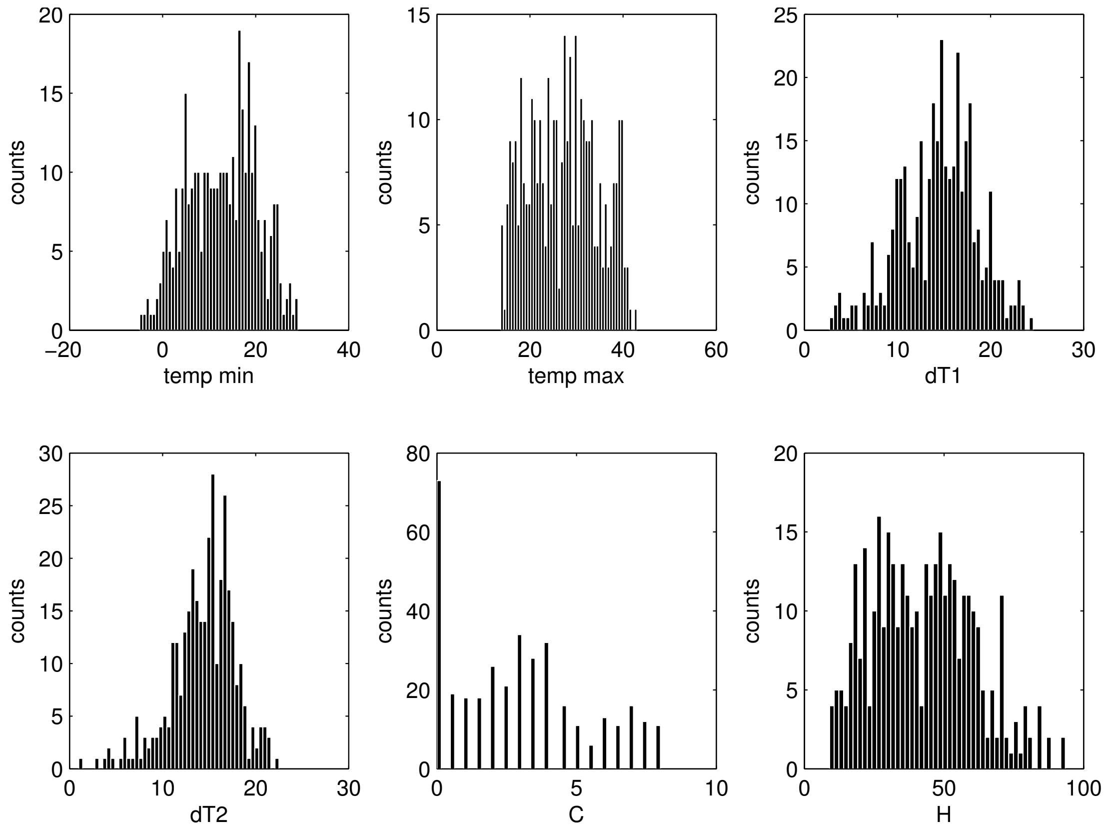


Figure 2: Histograms for Charleville data set for 1952.

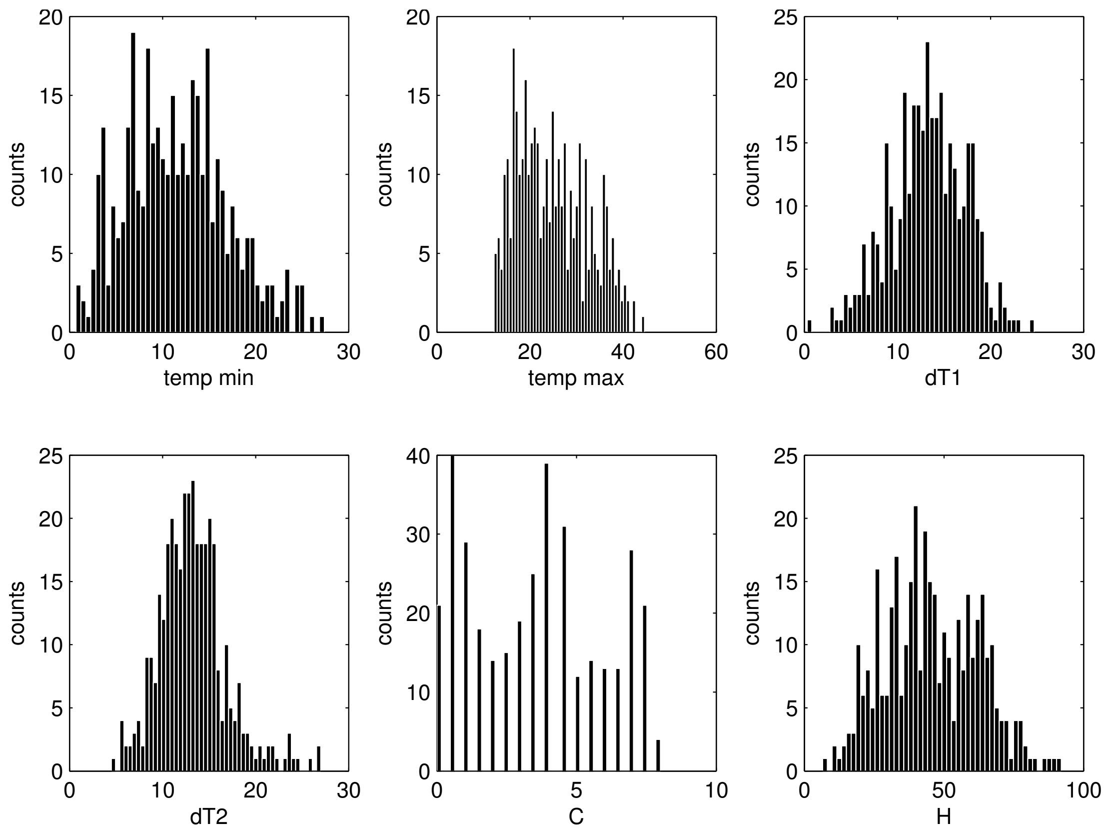


Figure 3: Histograms for Woomera data set for 1952.

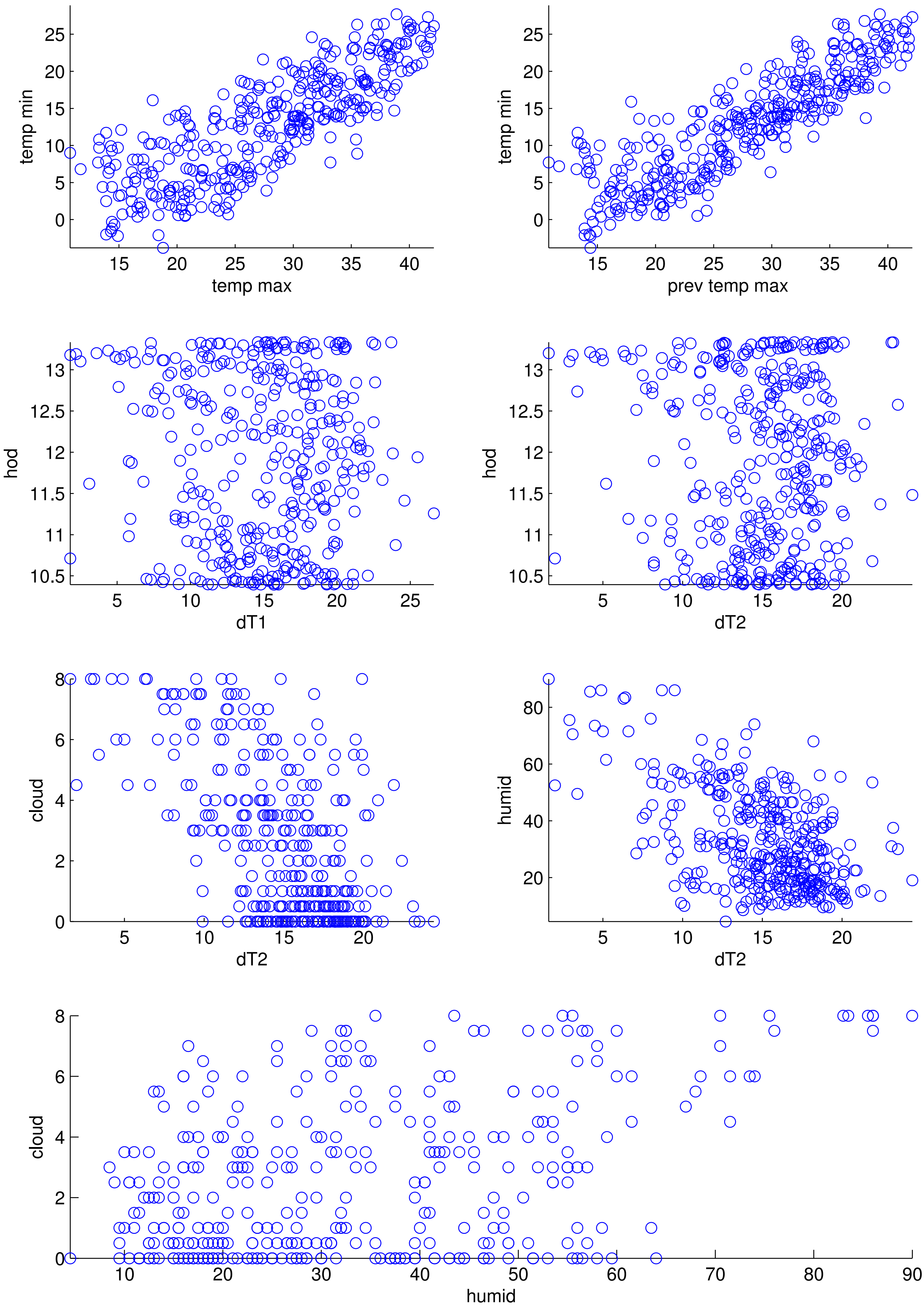


Figure 4: Various comparisons for Alice Springs 1952 data.

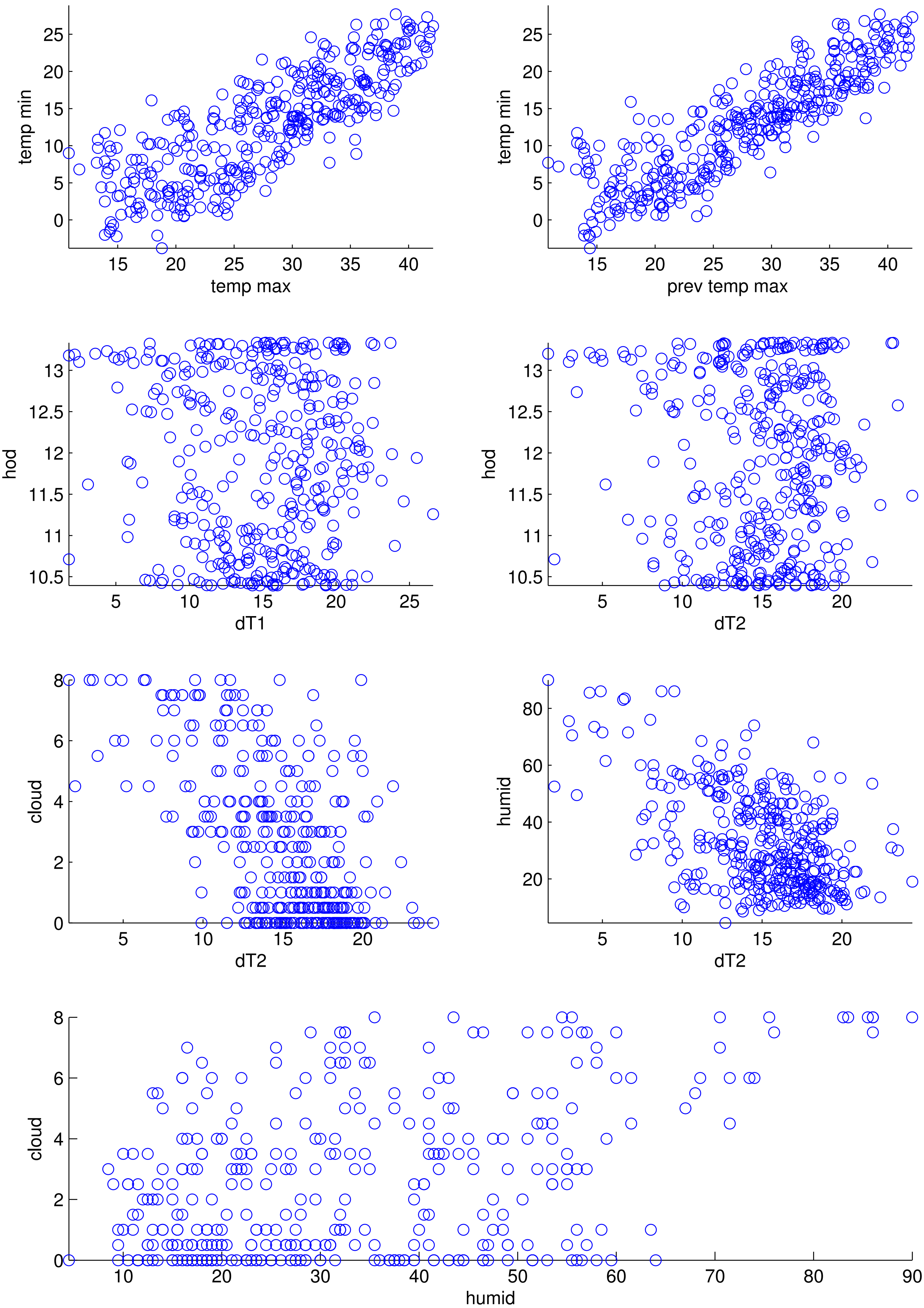


Figure 5: Various comparisons for Charleville 1952 data.

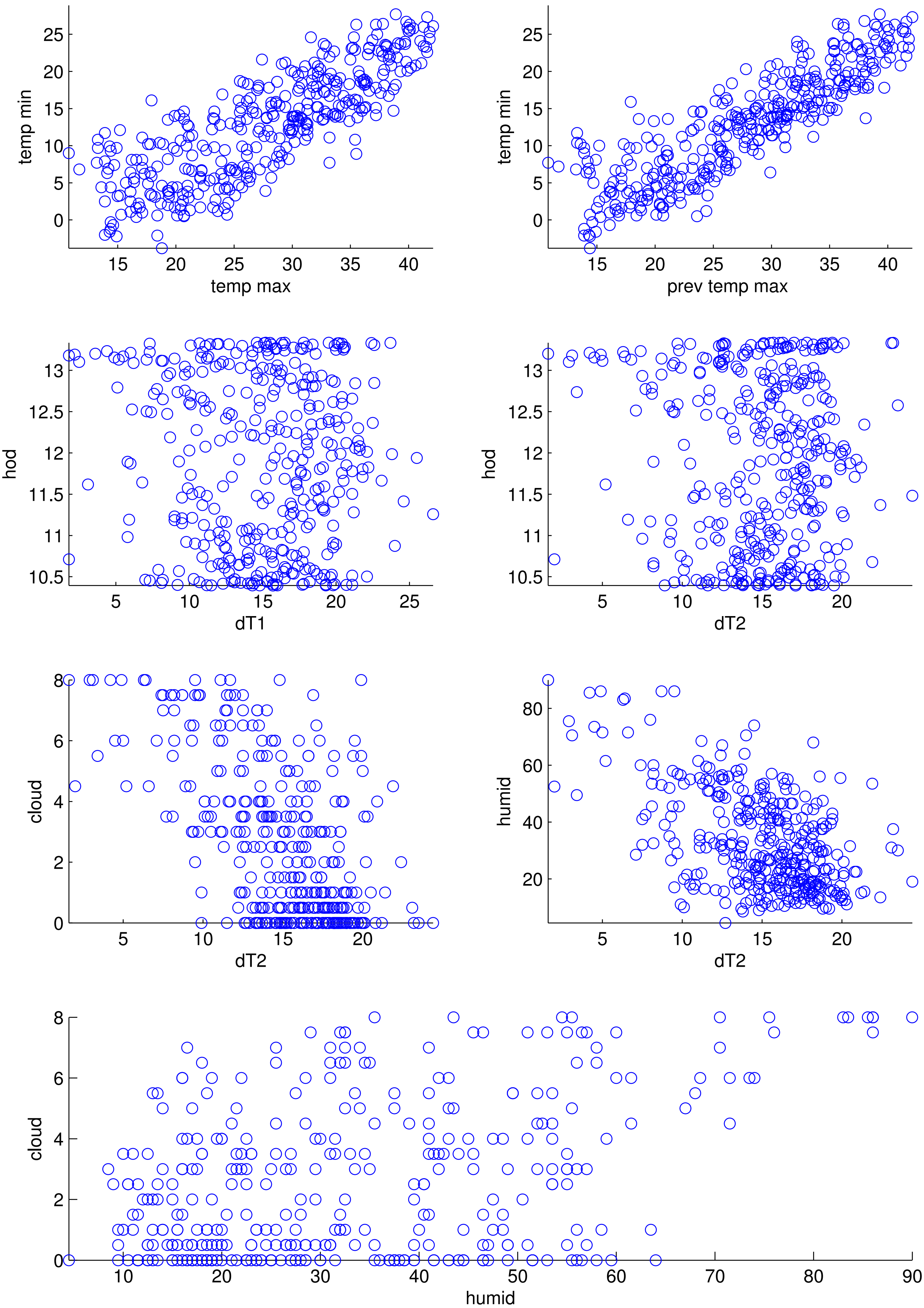


Figure 6: Various comparisons for Woomera 1952 data.

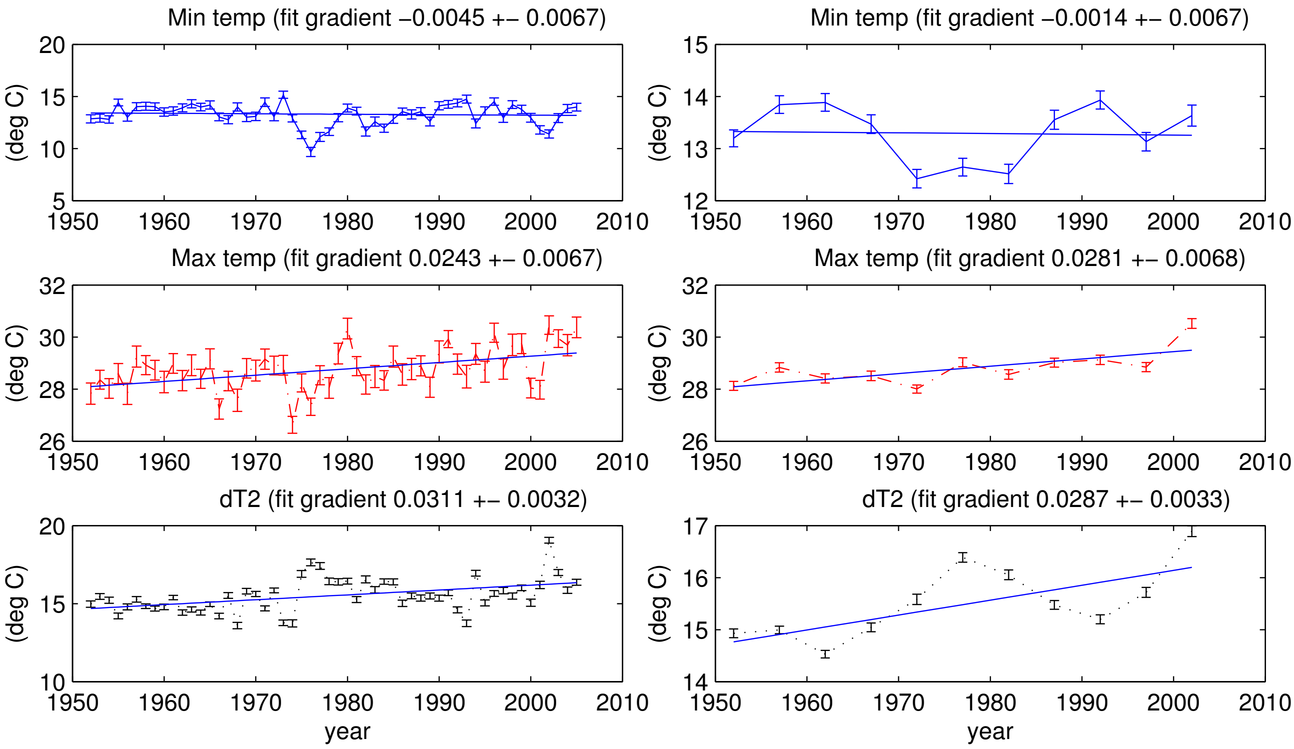


Figure 7: One- and five-year trends for Alice data. The confidence interval displayed for the gradient is two standard deviations, indicating statistically significant increases for `max temp` and `dT2`.

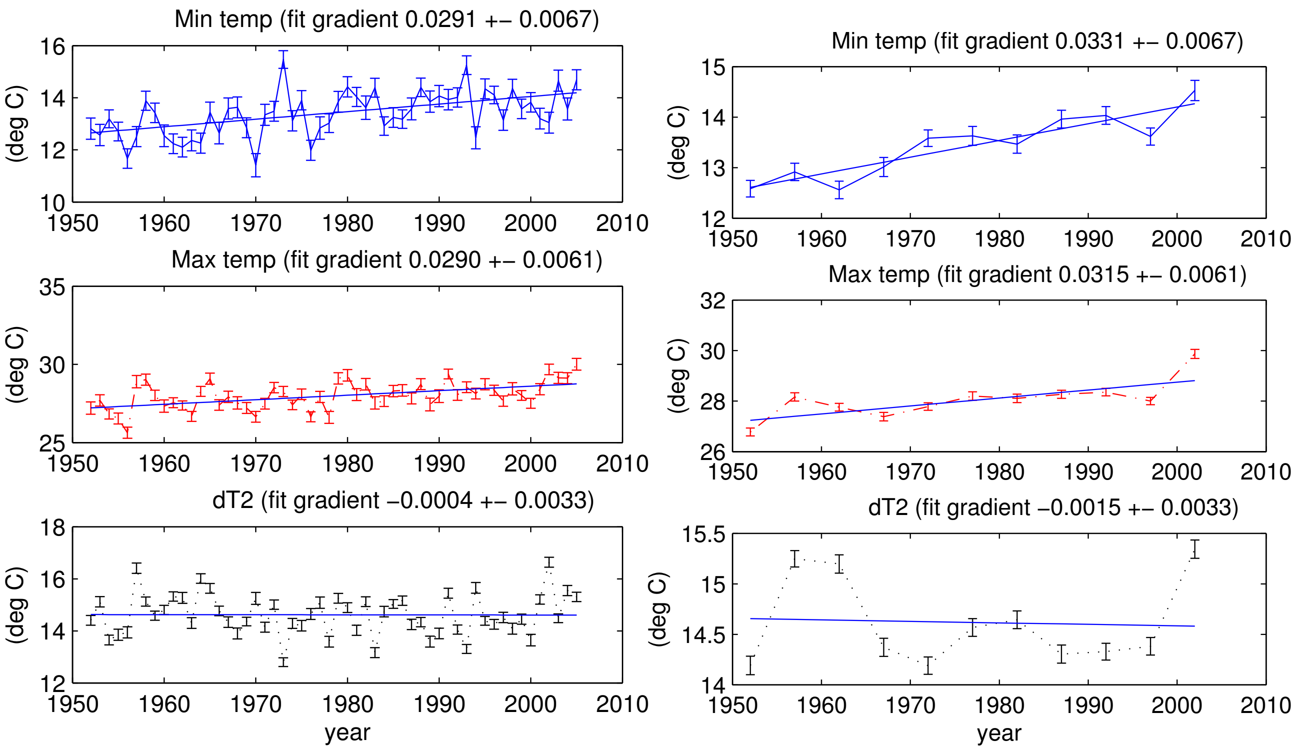


Figure 8: One- and five-year trends for Charleville data.

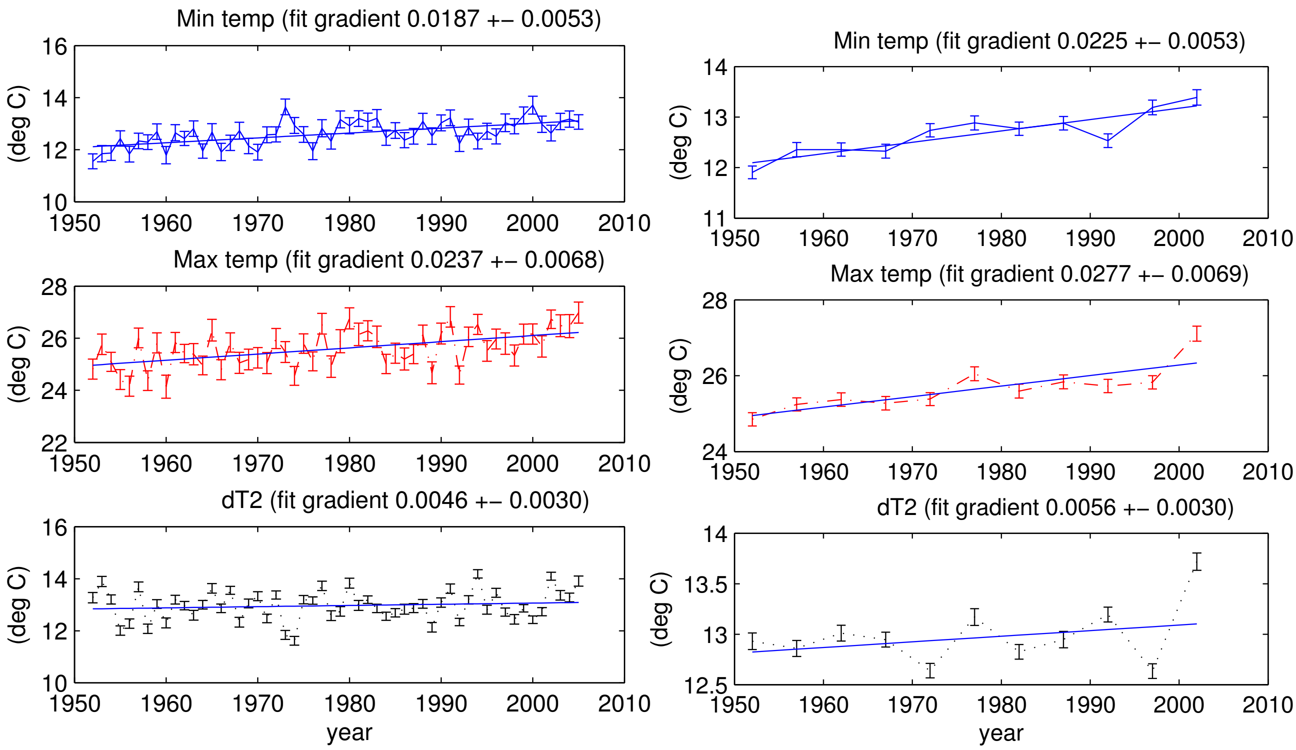


Figure 9: One- and five-year trends for Woomera data.

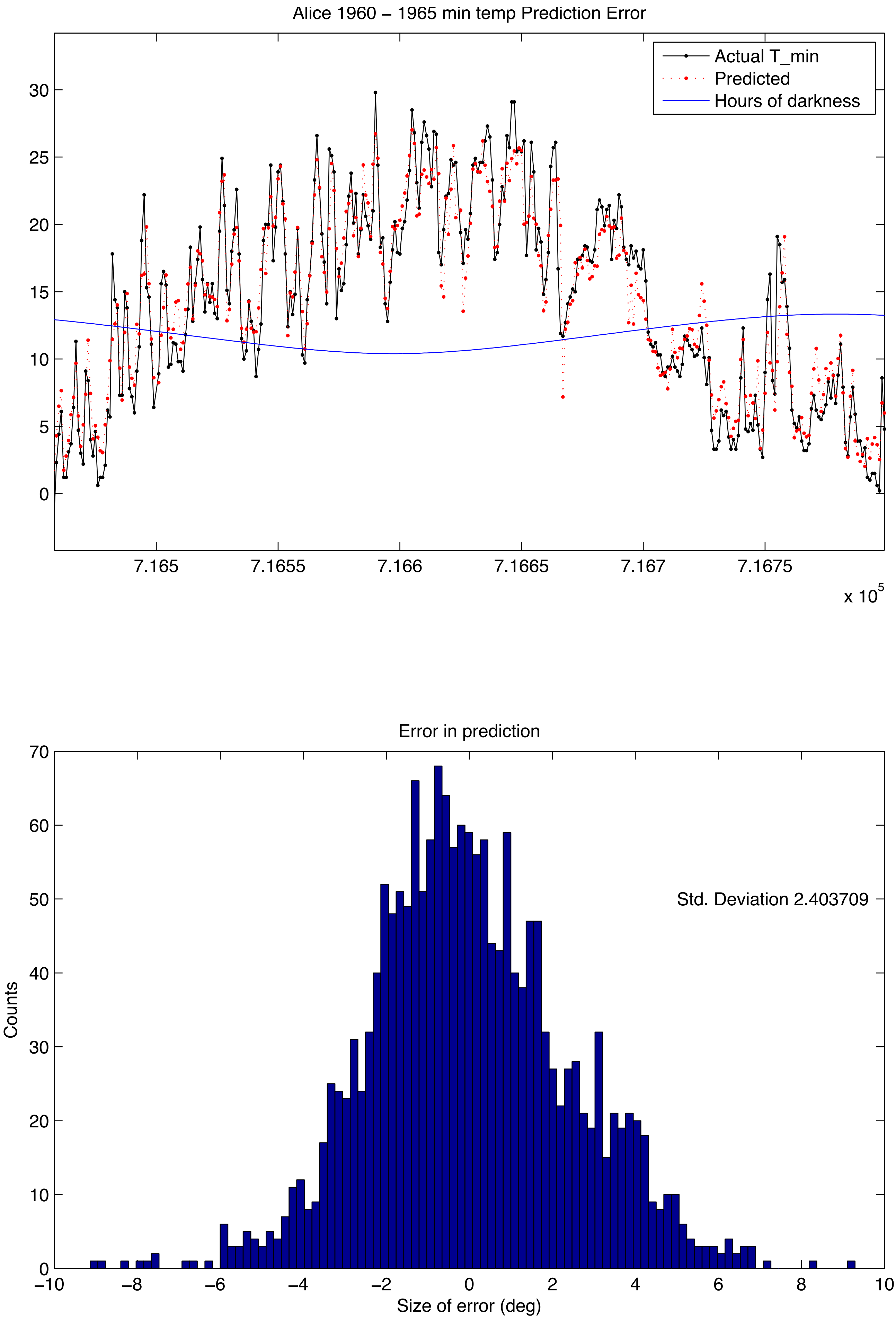


Figure 10: An example of the fit to the model.

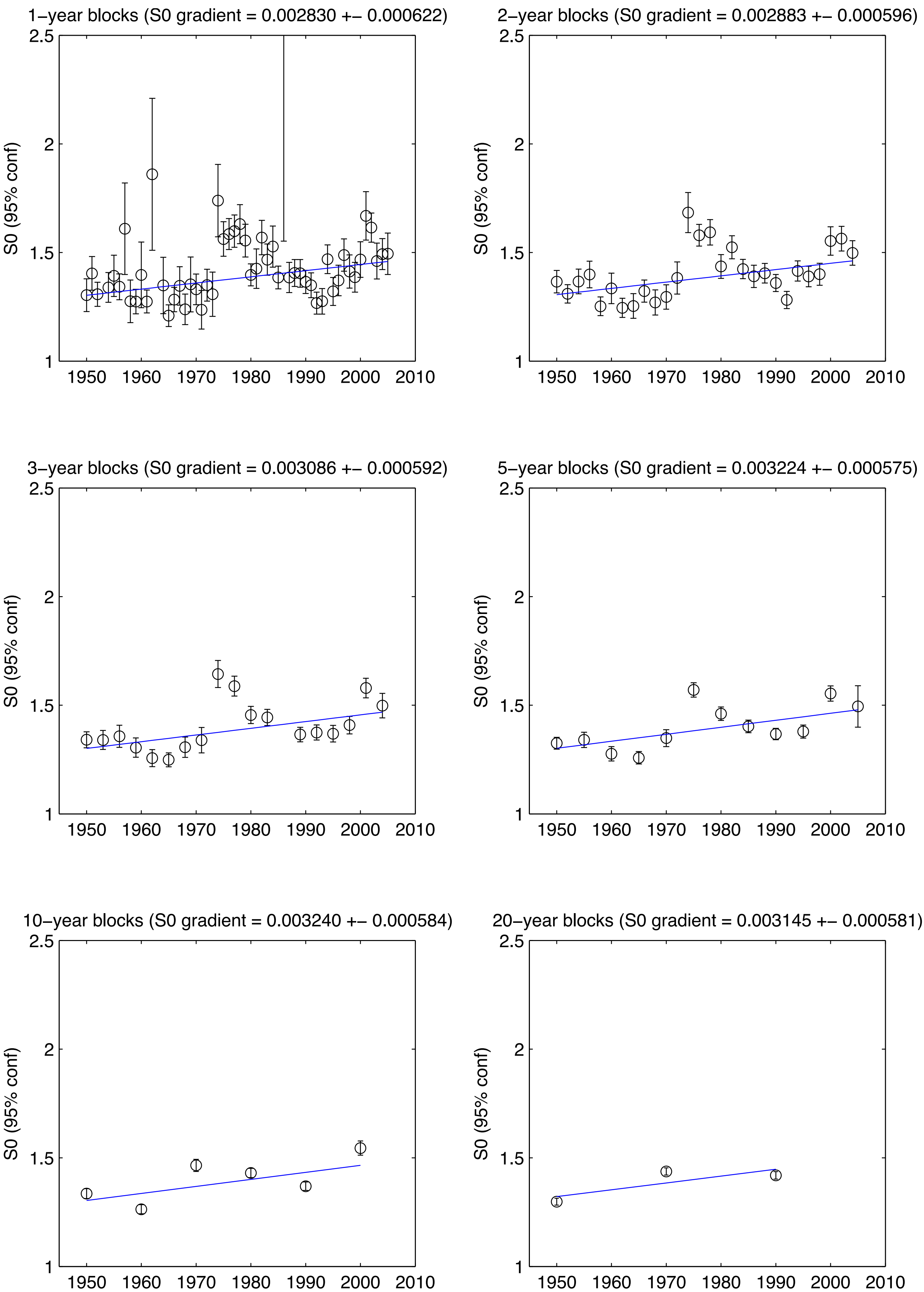


Figure 11: Variation in S0 for Alice data set. Each graph shows the fit for variously-sized blocks of years: 1,2,3,5,10 and 20. Error bars and quoted errors are 2 standard deviations.

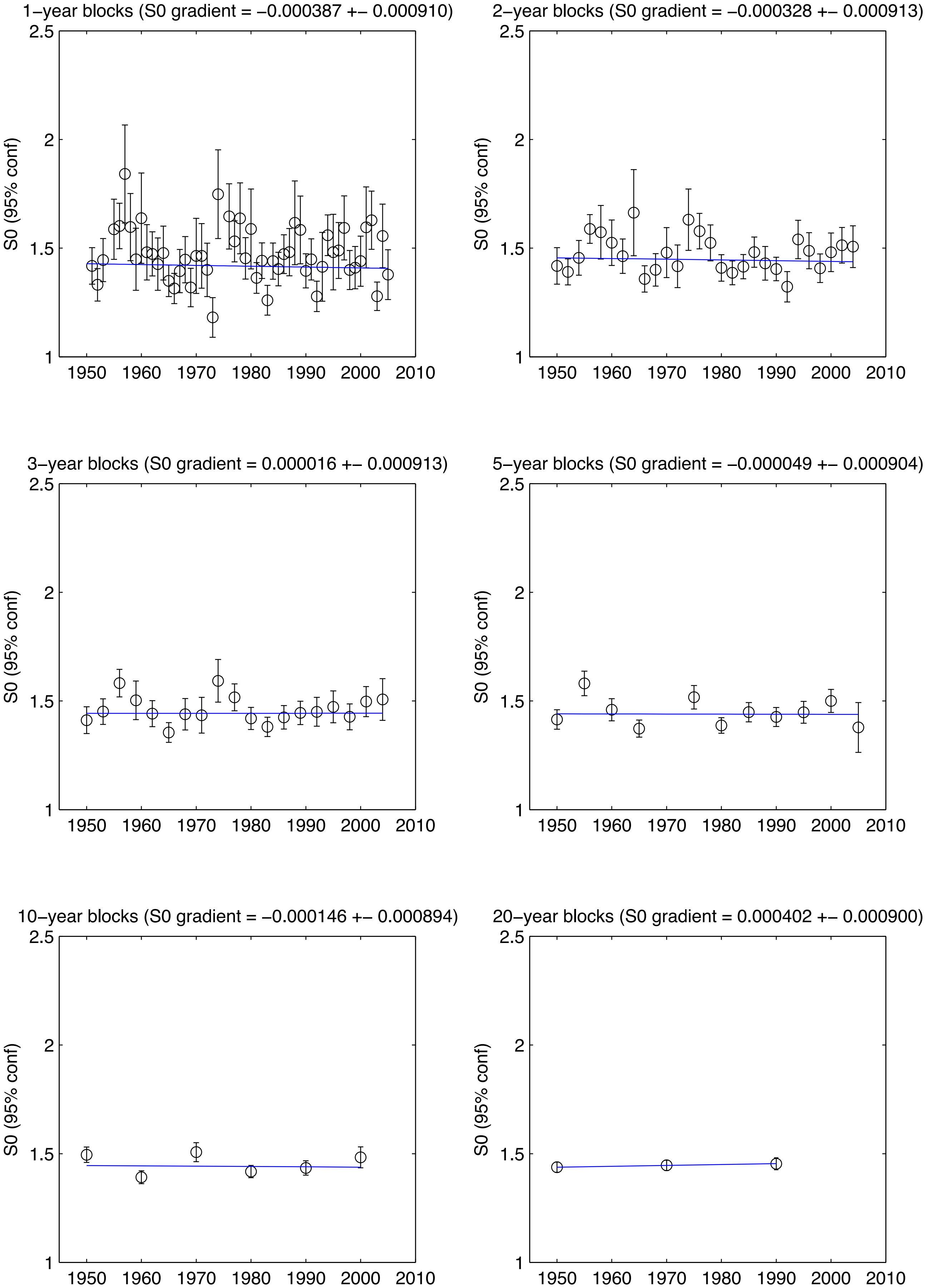


Figure 12: Variation in S0 for Charleville data set.

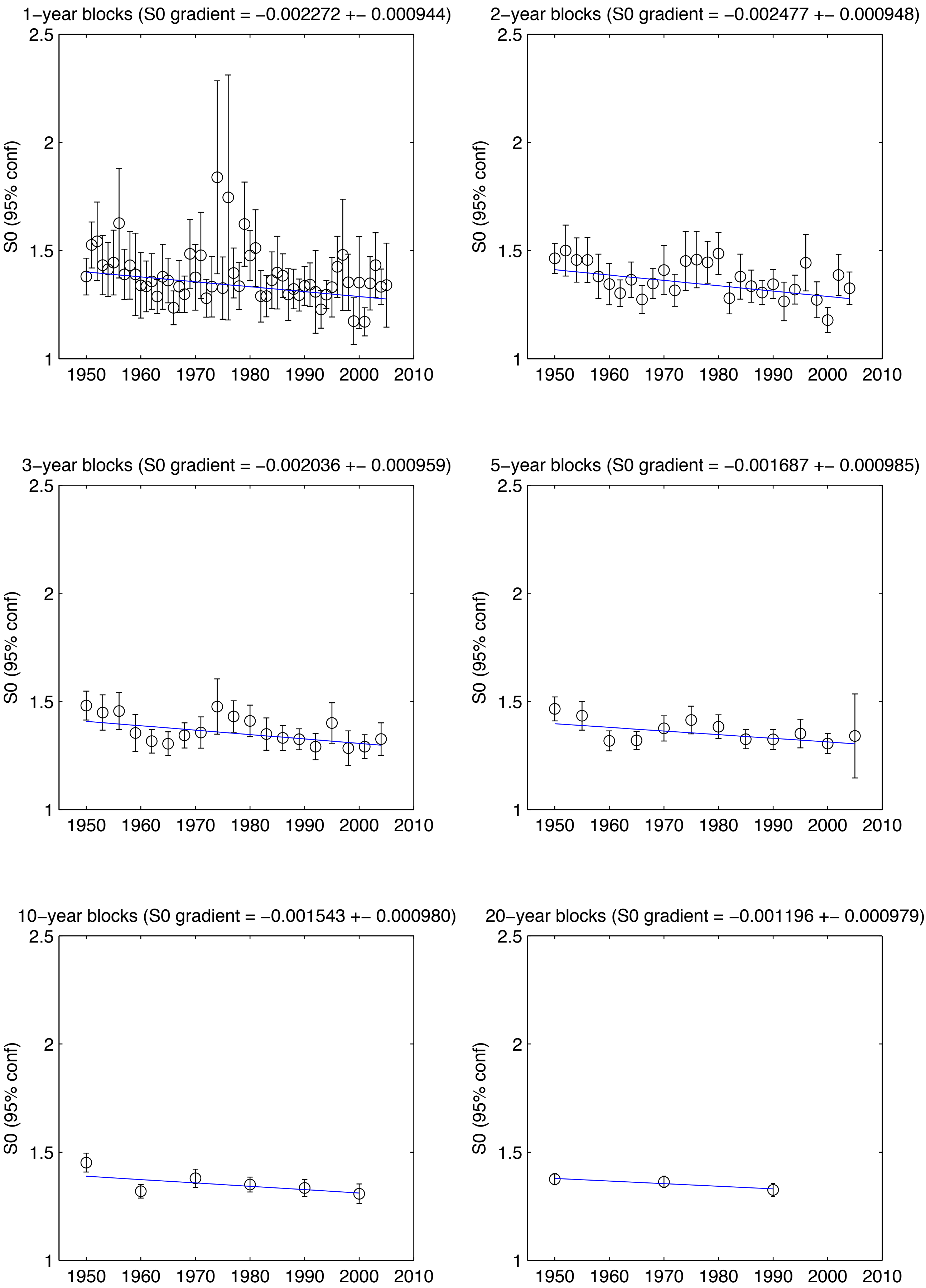


Figure 13: Variation in S0 for Woomera data set.

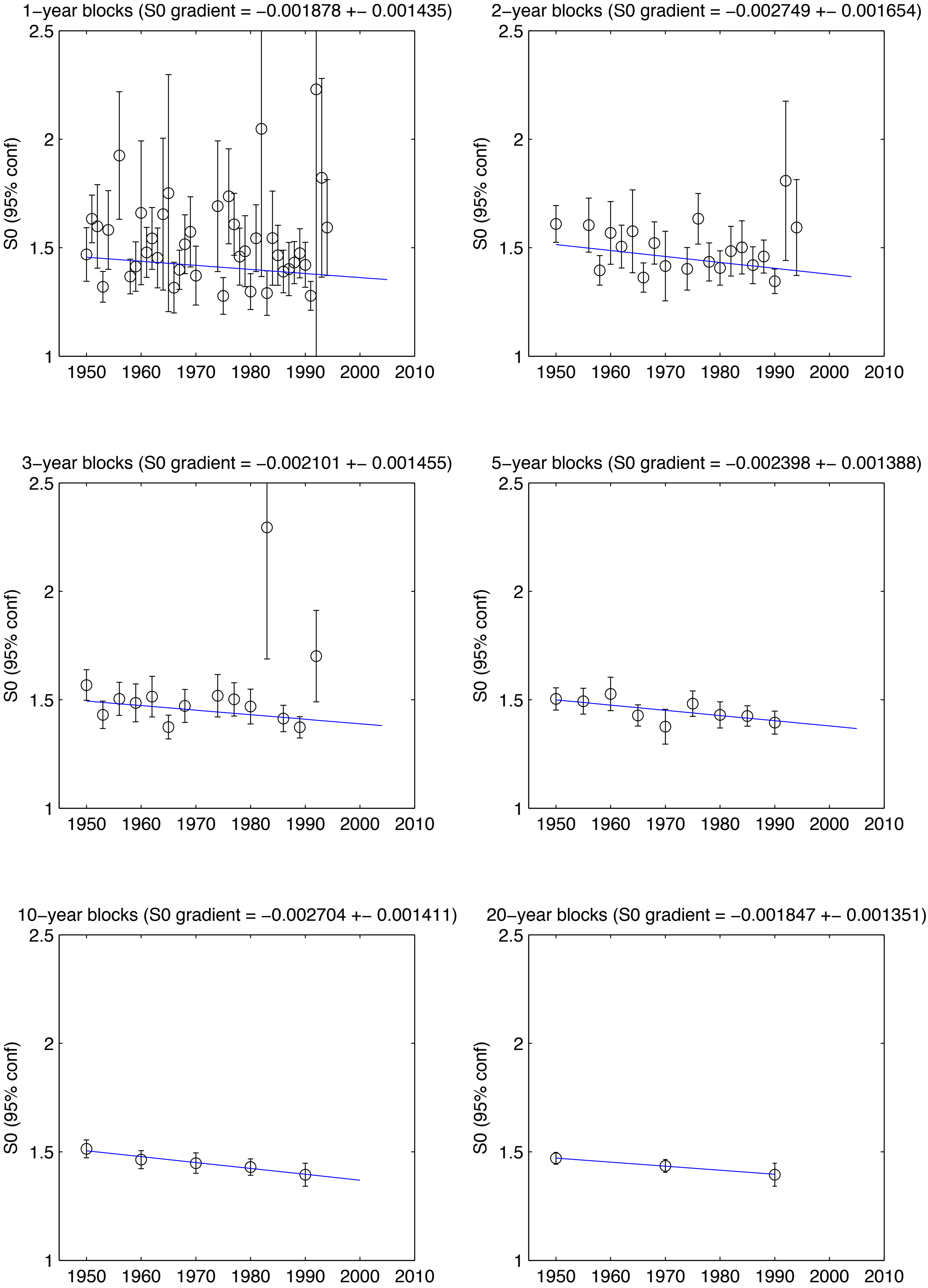


Figure 14: Variation in S0 for Bourke Post Office data set.

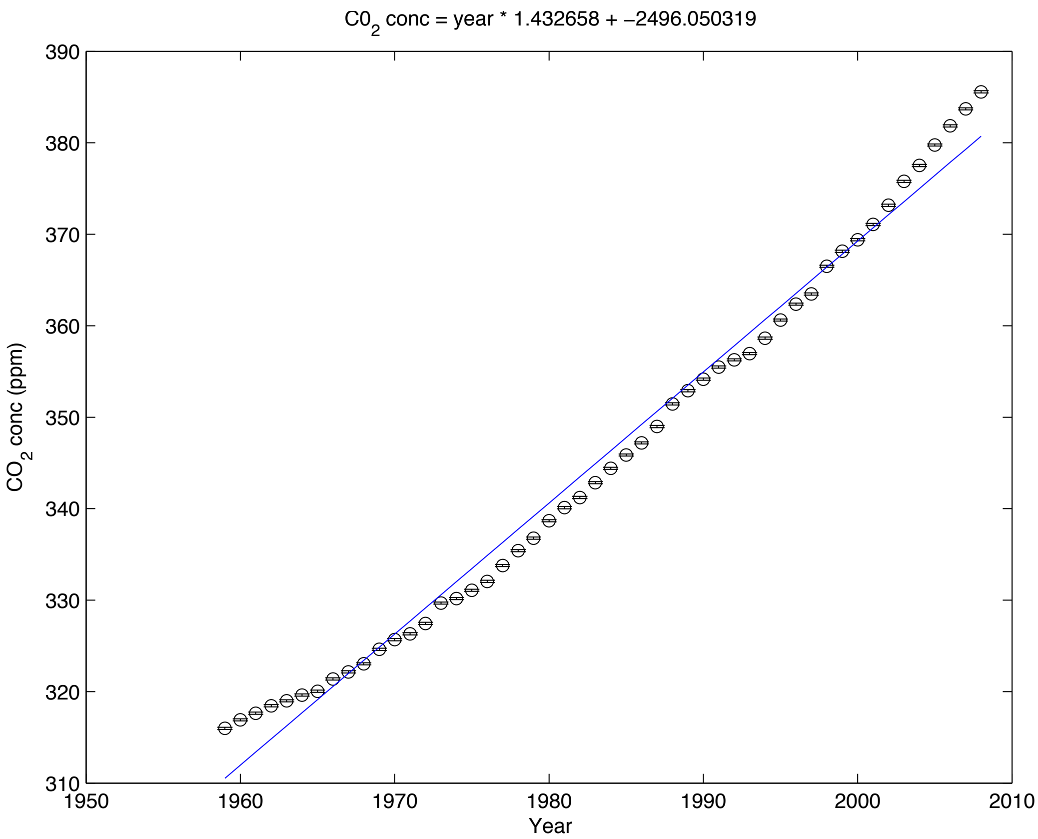


Figure 15: Atmospheric CO_2 concentration over time.

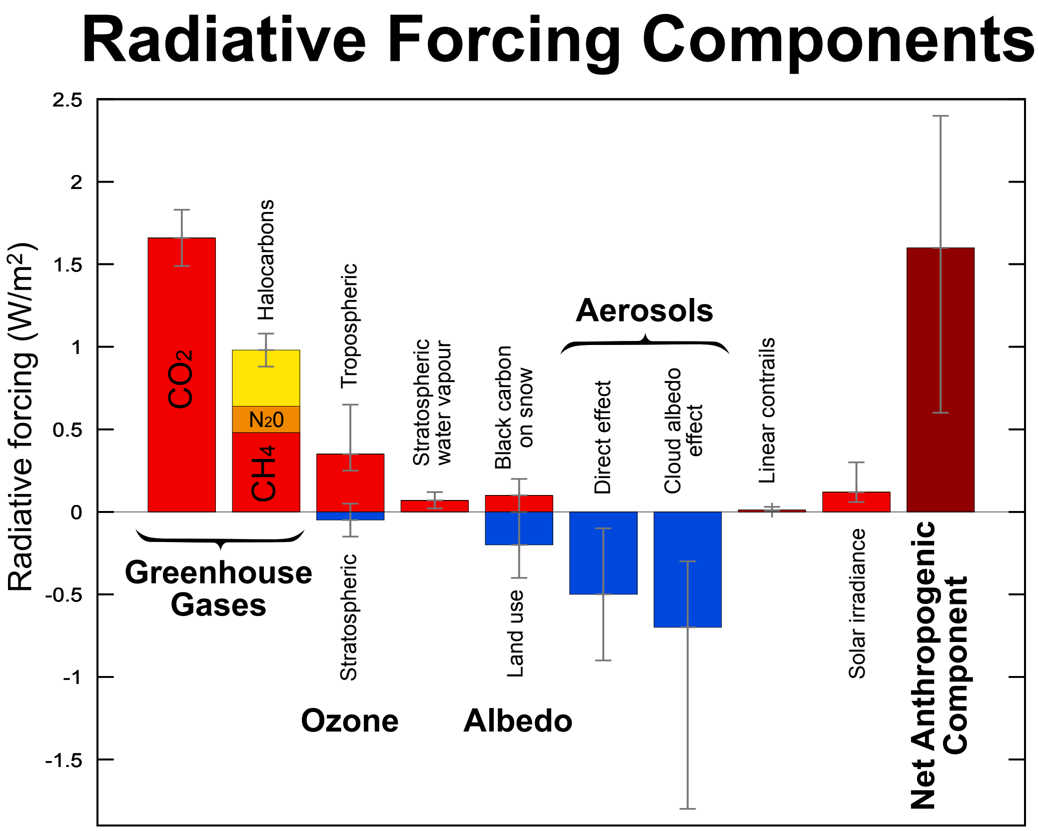
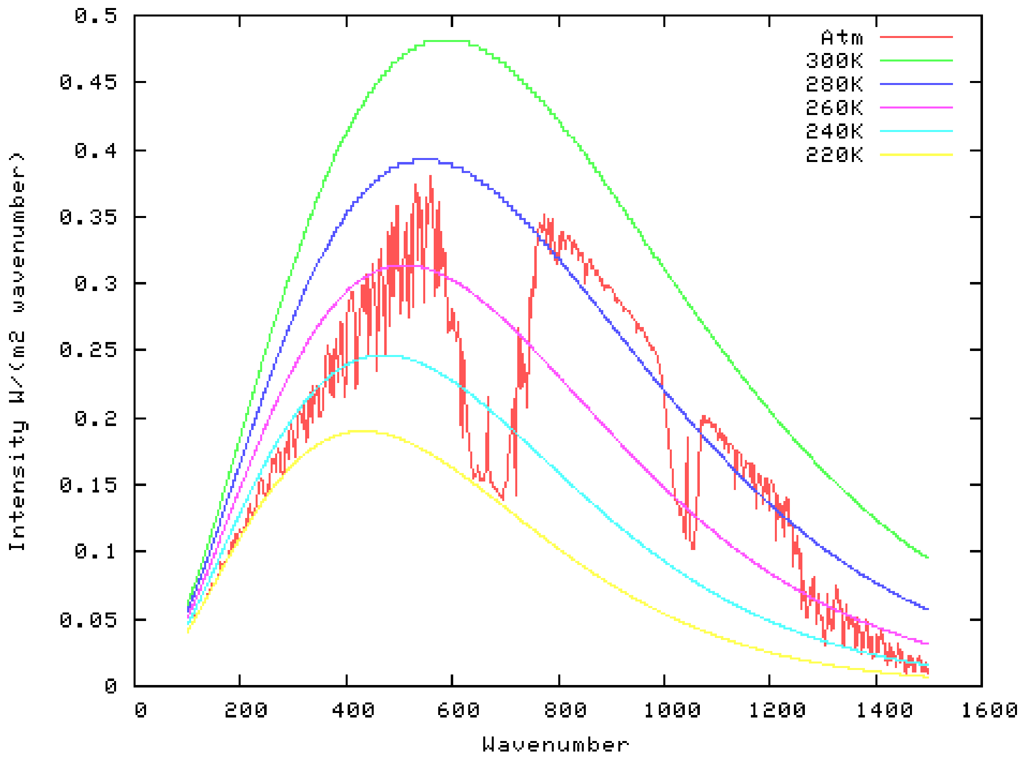
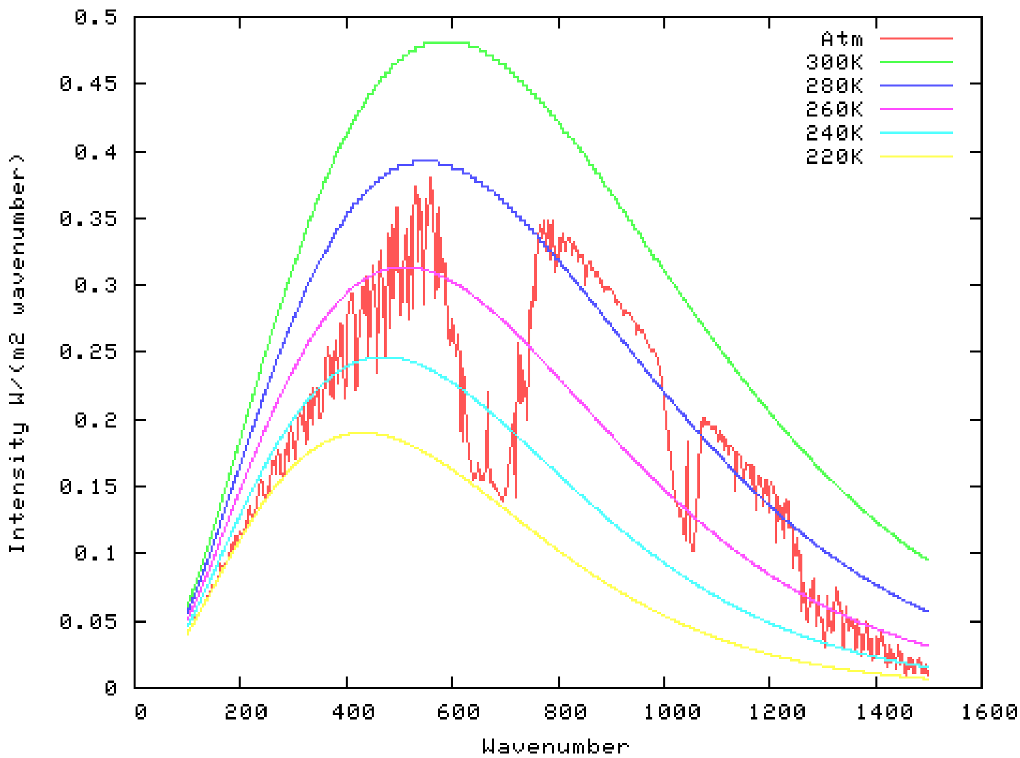


Figure 16: Climate Data, from IPCC.



(a) Radiative forcing curve for 310ppm CO_2. Here, $I = 259.6 \mathrm{Wm}^{-2}$.



(b) Radiative forcing curve for 385ppm CO_2. Here, $I = 258.7 \mathrm{Wm}^{-2}$.

Figure 17: Change in Radiative Forcing, from http://geoflop.uchicago.edu/forecast/docs/Projects/modtran.html .

8 Indices

List of Figures

List of Tables