Problem Set 5: MAE 127

due Friday, May 13, 2005

1. Temperatures in the Southern Ocean are reported to have risen in the last few decades. The data file deltaT.mat (available from the course web site and from the UCSD server) contains three variables: "deltaT" indicates the difference between temperatures measured historically from ships and temperatures measured in the 1990s; "decade" indicates the decade when shipboard observations were collected; "sigma" indicates the one standard deviation uncertainty in the temperature change estimates.

(a) Fit a constant and linear slope to the data, using a weighted least-squares fit. Estimate the uncertainty in your fit.

To do a basic least squares fit, use the following commands:

```
dref=mean(decade);
A=[ones(size(decade)) decade-dref];
Awt=A./(sigma*ones(1,2));
Twt=deltaT./sigma;

x_1a=inv(Awt'*Awt)*Awt'*Twt;
xe_1a=sqrt(diag(inv(Awt'*Awt)));

% other variables of interest
chi2_1a=sum((Awt*x_1a-Twt).^2);
T_1a=(Awt*x_1a).*sigma;

% plotting results
errorbar(decade,deltaT,sigma); hold on
plot(decade,T_1a,'g'); xlabel('time (years)'); ylabel('\Delta T (^oC)')
```

This shows a best fit of $\Delta T = -0.1002 \pm 0.0087 + (0.0027 \pm 0.0006)(t - 1965)$*. (If you didn't subtrack a reference time from the time column, then you found* $\Delta T = -5.34 \pm 1.1141 + (0.0027 \pm 0.0006)t$*.)*

(b) Using only the data from the 1950s through 1990s, repeat your fit.

To reduce the time period from part (a), we just do a minor modification to the procedure by selecting specific rows of data to consider.

```
ii=3:7;
x_1b=inv(Awt(ii,:)'*Awt(ii,:))*Awt(ii,:)'*Twt(ii);
xe_1b=sqrt(diag(inv(Awt(ii,:)'*Awt(ii,:))));


% other variables of interest
```

```
chi2_1b=sum((Awt(ii,:)*x_1b-Twt(ii)).^2)
T_1b=(Awt(ii,:)*x_1b).*sigma(ii);

% plotting results
plot(decade(ii),T_1b,'r');
```

This shows a best fit of $\Delta T = -0.1185 \pm 0.0115 + (0.0042 \pm 0.0008)(t - 1965)$. *(If you didn't subtract a reference time, then* $\Delta T = -8.3713 \pm 1.6630 + (0.0042 \pm 0.0008)t$.*)*

(c) On the basis of the results from (a) and (b), what would you predict total ocean warming relative to the 1990s to be in 2100? What are the uncertainties in your estimate?

To find the total temperature change in 2100, we simply plug the year 2100 into the fits derived in parts (a) and (b). We also have to use error propagation to estimate the uncertainties. Thus $\delta_T = \sqrt{\delta_{x_1}^2 + \delta_{x_2}^2 (t - 1965)^2}$. *This is coded:*

```
Delta_Ta = x_1a(1)+x_1a(2)*(2100-1965)
Delta_Tb = x_1b(1)+x_1b(2)*(2100-1965)
sigma_Ta=sqrt(xe_1a(1)^2 + xe_1a(2)^2*(2100-1965)^2)
sigma_Tb=sqrt(xe_1b(1)^2 + xe_1b(2)^2*(2100-1965)^2)
```

In case (a) this produces a predicted temperature change of $\Delta T_a = 0.26 \pm 0.08^\circ C$. *For case (b),* $\Delta T_b = 0.45 \pm 0.11^\circ C$.

In this particular case, subtracting the mean year from the dates has not changed the results much, but it has a big influence on the uncertainty in our prediction for ΔT *in 2100. You did not lose points if you did not subtract the mean, but you should look closely at the influence this has on your results. If you didn't subtract a reference time, then the uncertainties are* $\pm 1.63^\circ C$ *for 1a and* $\pm 2.4^\circ C$ *for 1b.*

2. Using the same data as in question 1, now consider the "goodness of fit" using the χ^2 criteria.

(a) Estimate χ^2 for the fits in (a) and (b). On the basis of the χ^2 criteria which is a better fit?

For case (a) $\chi_a^2 = 9.3$, *and we are fitting two functions to 7 data points, so* $N - M = 5$ *degrees of freedom. Thus* χ_a^2 *exceeds what we might predict, though not enormously. For case (b),* $\chi_b^2 = 3.1$, *and we are fitting two functions to 5 data points, for* $N - M = 3$. *Thus* χ_b^2 *more or less matches what we might predict for a good prediction. This tells us that given their uncertainties, the 1950 to 1990 data are more consistent with a linear trend than the 1930 to 1990 data.*

(b) Some people have suggested that temperature variability might be cyclical, with a 60 year time scale. To evaluate this, fit the full records (1930s-1990s) to a constant, linear trend, plus a cosine and sine with 60 year periodicity. On the basis of the χ^2 criteria, does adding additional parameters measurably improve your fit? Please comment on the results.

For this fit we define a new matrix A with four columns:

```
A=[ ones(size(decade)) decade-dref cos(decade*2*pi/60) sin(decade*2*pi/60)];
Awt=A./(sigma*ones(1,4));
x_2b=inv(Awt'*Awt)*Awt'*Twt;
xe_2b=sqrt(diag(inv(Awt'*Awt)));

% other variables of interest
T_2b=(Awt*x_2b).*sigma;
chi2_2b=sum((Awt*x_2b-Twt).^2)

plot(decade,T_2b,'c');
legend('data','7 decade fit','5 decade fit','60-year cyclical fit')
```

The basic fit yields $\Delta T = -0.108 \pm 0.011 + (0.0014 \pm 0.0008)(t - 1965) + (0.045 \pm 0.020)\cos(2\pi t/60) + (0.019 \pm 0.012)\sin(2\pi t/60)$. *In this case* $\chi_b^2 = 0.07$, *and we expect a fit to have a summed difference of about* $N - M = 7 - 4 = 3$. *The additional parameters improve the fit. In fact, this four-parameter fit is really better than we might expect, suggesting that we may have be fitting too many parameters to too little data.*

3. Again using the same data as in question 1, consider how well you could fit a constant, linear trend, and 500 year cycle to the data. Use a singular value decomposition to check whether the matrix inversion is likely to be stable. If you have doubts about the results, you might consider what the uncertainties in your fit imply for extrapolations for future ΔT.

In this case, we solve the same way as in question 2, but with a longer periodicity.

```
p=500;
A=[ ones(size(decade)) decade-dref cos(decade*2*pi/p) sin(decade*2*pi/p)];
Awt=A./(sigma*ones(1,4));

[U,S,V]=svd(Awt); diag(S)

x_3a=inv(Awt'*Awt)*Awt'*Twt;
xe_3a=sqrt(diag(inv(Awt'*Awt)));
[x_3a xe_3a]

chi2=sum((Awt*x_3a-Twt).^2)
T_3a=(Awt*x_3a).*sigma;
```

If we solve this way, we find that the singular values of A_{wt} *are* $[1.95 \times 10^3,\ 160.3,\ 1.9,\ 0.16]$. *Thus the ratio between the smallest and largest is* $O(10^{-4})$, *which is still mathematically tractable, and the solution will probably be OK, although the uncertainties turn out to be a little large. The solution itself is:* $\Delta T = 0.35 \pm 0.37 - (0.19 \pm 0.08)(t - 1965) + (6.25 \pm 2.80)\cos(2\pi t/200) + (14.40 \pm 5.73)\sin(2\pi t/200)$. *We find that* $\chi^2 = 0.11$ *which is small, as in problem 2b. If we predict* ΔT *for 2100, we find:*

```
Delta_T_3a = x_3a(1)+x_3a(2)*(2100-1965)+x_3a(3)*cos((2100-1965)*2*pi/500) ...
   +x_3a(4)*sin((2100-1965)*2*pi/500)
delta_Delta = sqrt(xe_3a(1)^2 + (2100-1965)^2*xe_3a(2)^2 + ...
   +xe_3a(3)^2*(cos((2100-1965)*2*pi/500))^2 ...
   +xe_3a(4)^2*(sin((2100-1965)*2*pi/500))^2)
```

which produces $\Delta T(2100) = -11.96 \pm 12.07$*. This is not a terribly meaningful result, since the answer is not statistically different from zero.*

The situation changes for the worse if we don't subtract the mean time from the linear slope term. In this case:

```
p=500;
A=[ ones(size(decade)) decade cos(decade*2*pi/p) sin(decade*2*pi/p)];
Awt=A./(sigma*ones(1,4));

[U,S,V]=svd(Awt); diag(S)

x_3a=inv(Awt'*Awt)*Awt'*Twt;
xe_3a=sqrt(diag(inv(Awt'*Awt)));
[x_3a xe_3a]

chi2=sum((Awt*x_3a-Twt).^2)
T_3a=(Awt*x_3a).*sigma;
```

The singular values in this case are $[2.5\times10^5,\ 21.8,\ 2.7,\ 0.0065]$*, so the ratio of the smallest to largest is* $O(10^{-8}$*, which is a big enough difference to imply numerical problems. The estimate of* χ^2 *is still small, around 0.11, and the fit appears very good, but the uncertainties are enormous. We find* $\Delta T(2100) = 364 \pm 155^\circ C$*, not a result I'd necessarily want to trust.*

The large difference between the smallest and largest singular values might suggest that we should use fewer singular values to derive the solution, or that we should edit the matrix A to have fewer columns.