Date: Tue, 18 Jul 2006 10:28:23 -0400 (EDT)
From: Noel Cressie *<email omitted>*
To: Edward Wegman *<email omitted>*
Cc: Yasmin Said *<email omitted>*,
Noel Cressie *<email omitted>*
Subject: Reactions to your report to congress

Dear Ed:
....

1. The recommendations are solely about the support of research in the area of climate change. I would have thought that one recommendation would be to do a careful re-analysis of the data that takes into account, e.g., meas. error and temporal dependence (see your p. 3 and my attached technical comments).

2. With the properly centered data matrix X, I suggest you show PC_1 *on the same scale* as PC_1 from the improperly centered data matrix. After all your discussion in the report, a Rep. may, quite sensibly, ask for the "right" answer. Your message would be clearer if you also compared it to the "wrong" answer. Two graphs, one under the other, or superimposed, would send the appropriate message.

3. In Figure 4.4, MM showed the hockey stick to bend upwards in all their (well chosen) realizations. In fairness, you should show some realizations where it bends downwards too.

4. In Appendix A, it would help if you give different notation for the properly and improperly centered covariance matrix. Equally, x-bar is used to mean two different things in that appendix.

5. Finally, I see you did not go on to correlate (or regress) PC_1 (properly centered) on CO_2. I know you are eschewing linear relationships, but I recall that Al Gore was making the point visually. A better visual would be a scatter diagram, which would be suggestive of a possible relationship. In fact, in light of 2., you could do *two* scatter diagrams, one showing the correct PC_1 v. (lagged)CO_2, and the other showing the same with the incorrect PC_1.

With my best regards,
Noel

[Attachment follows]

7/18/06

Comments on Wegman et al. report to
House of Representatives on Paleoclimate Reconstruction

by
Noel Cressie
The Ohio State University

I concur with the technical contents of the report. The SVD and PCA based on (properly centered) X or $(1/n)X^TX$, respectively, is a data analytic tool. Of course it does not represent your call for "serious investigation to model the underlying process structures nor ... the present instrumented temperature record with sophisticated process models" (p. 3 of your report).

I would like to consider two "structures" mentioned in the report.

1. Measurement error. The data $\{\mathbf{X}(t_i)\}$ should be decomposed as,

$$\mathbf{X}(t_i) = \mathbf{Y}(t_i) + \boldsymbol{\varepsilon}(t_i)\,; \quad i = 1, \ldots, n\,,$$

where $\{\boldsymbol{\varepsilon}(t_i)\}$ represents independent measurement error, independent of the underlying process $\mathbf{Y}(\cdot)$. Generally, $\mathrm{var}(\boldsymbol{\varepsilon}(t)) = V(t)$, although independence of components of $\boldsymbol{\varepsilon}(\cdot)$ and stationarity of $V(t)$ might be assumed. Now,

$$\mathrm{var}(\mathbf{X}(t)) = \mathrm{var}(\mathbf{Y}(t)) + \mathrm{var}(\boldsymbol{\varepsilon}(t))\,,$$

and assuming stationarity over time,

$$\Sigma^X = \Sigma^Y + V\,.$$

Clearly, $(1/n)X^TX$ estimates Σ^X but we actually want to estimate Σ^Y. One approach is to de-noise the data, perhaps by Kalman smoothing.

2. Temporal dependence (red noise). Write

$$\mathbf{X}(t_i) = \mathbf{Y}(t_i) + \mathbf{W}(t_i)\,; \quad i = 1, \ldots, n\,,$$

where for illustration assume

$$\mathbf{W}(t) = \alpha\mathbf{W}(t-1) + \boldsymbol{\nu}(t)\,;$$

$0 \leq |\alpha| < 1$; $\{\boldsymbol{\nu}(t_i)\}$ are independent; and $\boldsymbol{\nu}(t)$ is independent of $\mathbf{W}(t-1)$. Then it is easy to see that

$$\Sigma^X = \Sigma^Y + (1-\alpha^2)^{-1}\Sigma^\nu \,.$$

Once again, $(1/n)X^T X$ is a biased estimate of Σ^Y, and de-noising the data to obtain an estimate of Σ^Y is more difficult. It will in general depend on the temporal dependence of the hidden process $\mathbf{Y}(\cdot)$.