AD HOC COMMITTEE REPORT ON THE 'HOCKEY STICK' GLOBAL CLIMATE RECONSTRUCTION[2]

EXECUTIVE SUMMARY

The Chairman of the Committee on Energy and Commerce as well as the Chairman of the Subcommittee on Oversight and Investigations have been interested in an independent verification of the critiques of Mann et al. (1998, 1999) [MBH98, MBH99] by McIntyre and McKitrick (2003, 2005a, 2005b) [MM03, MM05a, MM05b] as well as the related implications in the assessment. The conclusions from MBH98, MBH99 were featured in the Intergovernmental Panel on Climate Change report entitled *Climate Change 2001*[3]*: The Scientific Basis*. This report concerns the rise in global temperatures, specifically during the 1990s. The MBH98 and MBH99 papers are focused on paleoclimate temperature reconstruction and conclusions therein focus on what appear to be a rapid rise in global temperature during the 1990s when compared with temperatures of the previous millennium. These conclusions generated a highly polarized debate over the policy implications of MBH98, MBH99 for the nature of global climate change, and whether or not anthropogenic actions are the source. This committee, composed of Edward J. Wegman (George Mason University), David W. Scott (Rice University), and Yasmin H. Said (The Johns Hopkins University), has reviewed the work of both articles, as well as a network of journal articles that are related either by authors or subject matter, and has come to several conclusions and recommendations. This Ad Hoc Committee has worked pro bono, has received no compensation, and has no financial interest in the outcome of the report.

Global Climate Reconstruction Methodology

MBH98, MBH99 use several indicators to measure global climate change. Primarily, these include historical records, tree rings, ice cores, and coral reefs. The width and density of tree rings vary with climatic conditions (sunlight, precipitation, temperature, humidity, and carbon dioxide and nitrogen oxides availability), soil conditions, tree species, tree age, and stored carbohydrates in the trees. However, tree ring density is useful in paleoclimatic temperature reconstructions because in mature trees, tree rings vary approximately linearly with age. The width and density of tree rings are dependent on many confounding factors, making it difficult to isolate the climatic temperature signal. It is usually the case that width and density of tree rings are monitored in conjunction in order to more accurately use them as climate proxies.

---

[2] This report was authored by Edward J. Wegman, George Mason University, David W. Scott, Rice University, and Yasmin H. Said, The Johns Hopkins University. We would also like to acknowledge the contributions of John T. Rigsby, III, Naval Surface Warfare Center, and Denise M. Reeves, MITRE Corporation.

[3] The IPCC report *Climate Change 2001: Third Assessment Report* consists of four sub-reports: 1) *Climate Change 2001: The Scientific Basis*, 2) *Climate Change 2001: Impacts, Adaptation, and Vulnerabilities*, 3) *Climate Change 2001: Mitigation*, and 4) *Climate Change 2001: Synthesis Report.*

Ice cores are the accumulation of snow and ice over many years that have recrystallized and have trapped air bubbles from previous time periods. The composition of these ice cores, especially the presence of hydrogen and oxygen isotopes, provides a picture of the climate at the time. Because isotopes of water vapor exhibit a lower vapor pressure, when the temperature falls, the heavier water molecules will condense faster than the normal water molecules. The relative concentrations of the heavier isotopes in the condensate indicate the temperature of condensation at the time, allowing for ice cores to be used in global temperature reconstruction. In addition to the isotope concentration, the air bubbles trapped in the ice cores allow for measurement of the atmospheric concentrations of trace gases, including greenhouse gases carbon dioxide, methane, and nitrous oxide. The air bubbles may also contain traces of aerosols, which are produced in great concentrations during volcanic eruptions.

Coral is similar to trees in that the growth and density of the coral is dependent upon temperature. X-rays of coral cross sections show the relative density and growth over time. High density layers of coral are produced during years of high ocean surface temperatures. Hence, corals can be calibrated to estimate sea surface temperatures.

Principal Component Analysis and the CFR and CPS Methodology

Principal component analysis is a method often used for reducing multidimensional datasets to lower dimensions for analysis. In this context, dimensions refer to the number of distinct variables. The time series proxy data involved are transformed into their principal components, where the first principal component is intended to explain most of the variation present in the variables. Each subsequent principal component explains less and less of the variation. In the methodology of MBH98/99, the first principal component is used in the temperature reconstruction, and also has the highest explained variance. This method is intended for dimension reduction. In most datasets, the first principal component should be the least smooth (because of the higher variance). However, in MBH98, MBH99, the proxy data are incorrectly centered, which inflates the variance of certain proxies and selectively chooses those decentered proxies as the temperature reconstruction.

There are several time series models that exist for the purpose of modeling series with dependence, including autoregressive, moving averages, autoregressive moving average models, and long memory processes. MBH98 and MBH99 focus on simple signal plus superimposed noise models for paleoclimate temperature reconstruction. Because of complex feedback mechanisms involved in climate dynamics, it is unlikely that the temperature records and the data derived from the proxies can be adequately modeled with a simple temperature signal with superimposed noise. We believe that there has not been a serious investigation to model the underlying process structures nor to model the present instrumented temperature record with sophisticated process models.

Two principal methods for temperature reconstructions have been used; CFR[4] (climate field construction) and CPS (climate-plus-scale). The CFR is essentially a principal component analysis and the CPS is a simple averaging of climate proxies, which are then scaled to actual temperature records. The controversy of Mann's methods lies in that the proxies are centered on the mean of the period 1902-1995, rather than on the whole time period. This mean is, thus, actually decentered low, which will cause it to exhibit a larger variance, giving it preference for being selected as the first principal component. The net effect of this decentering using the proxy data in MBH98 and MBH99 is to produce a "hockey stick" shape. Centering the mean is a critical factor in using the principal component methodology properly. It is not clear that Mann and associates realized the error in their methodology at the time of publication. Because of the lack of full documentation of their data and computer code, we have not been able to reproduce their research. We did, however, successfully recapture similar results to those of MM. This recreation supports the critique of the MBH98 methods, as the offset of the mean value creates an artificially large deviation from the desired mean value of zero.

Findings

In general, we found MBH98 and MBH99 to be somewhat obscure and incomplete and the criticisms of MM03/05a/05b to be valid and compelling. We also comment that they were attempting to draw attention to the discrepancies in MBH98 and MBH99, and not to do paleoclimatic temperature reconstruction. Normally, one would try to select a calibration dataset that is representative of the entire dataset. The 1902-1995 data is not fully appropriate for calibration and leads to a misuse in principal component analysis. However, the reasons for setting 1902-1995 as the calibration point presented in the narrative of MBH98 sounds reasonable, and the error may be easily overlooked by someone not trained in statistical methodology. We note that there is no evidence that Dr. Mann or any of the other authors in paleoclimatology studies have had significant interactions with mainstream statisticians.

In our further exploration of the social network of authorships in temperature reconstruction, we found that at least 43 authors have direct ties to Dr. Mann by virtue of coauthored papers with him. Our findings from this analysis suggest that authors in the area of paleoclimate studies are closely connected and thus 'independent studies' may not be as independent as they might appear on the surface. This committee does not believe that web logs are an appropriate forum for the scientific debate on this issue.

It is important to note the isolation of the paleoclimate community; even though they rely heavily on statistical methods they do not seem to be interacting with the statistical community. Additionally, we judge that the sharing of research materials, data and results was haphazardly and grudgingly done. In this case we judge that there was too much reliance on peer review, which was not necessarily independent. Moreover, the work has been sufficiently politicized that this community can hardly reassess their public positions without losing credibility. Overall, our committee believes that Mann's

[4] The CFR methodology is essentially the methodology used in the MBH98/99 papers, but the terminology was not used until later.

assessments that the decade of the 1990s was the hottest decade of the millennium and that 1998 was the hottest year of the millennium cannot be supported by his analysis.

Recommendations

Recommendation 1. Especially when massive amounts of public monies and human lives are at stake, academic work should have a more intense level of scrutiny and review. It is especially the case that authors of policy-related documents like the IPCC report, *Climate Change 2001: The Scientific Basis*, should not be the same people as those that constructed the academic papers.

Recommendation 2. We believe that federally funded research agencies should develop a more comprehensive and concise policy on disclosure. All of us writing this report have been federally funded. Our experience with funding agencies has been that they do not in general articulate clear guidelines to the investigators as to what must be disclosed. Federally funded work including code should be made available to other researchers upon reasonable request, especially if the intellectual property has no commercial value. Some consideration should be granted to data collectors to have exclusive use of their data for one or two years, prior to publication. But data collected under federal support should be made publicly available. (As federal agencies such as NASA do routinely.)

Recommendation 3. With clinical trials for drugs and devices to be approved for human use by the FDA, review and consultation with statisticians is expected. Indeed, it is standard practice to include statisticians in the application-for-approval process. We judge this to be a good policy when public health and also when substantial amounts of monies are involved, for example, when there are major policy decisions to be made based on statistical assessments. In such cases, evaluation by statisticians should be standard practice. This evaluation phase should be a mandatory part of all grant applications and funded accordingly.

Recommendation 4. Emphasis should be placed on the Federal funding of research related to fundamental understanding of the mechanisms of climate change. Funding should focus on interdisciplinary teams and avoid narrowly focused discipline research.

1. INTRODUCTION

Global warming is an issue that has gathered much public, legislative, national and international attention. Uncertainty about magnitude and consequences of global warming has caused considerable friction among governments and among their citizens. The Intergovernmental Panel on Climate Change (IPCC) report (Intergovernmental Panel on Climate Change, 2001) entitled, *Climate Change 2001: Third Assessment Report*, featured alarming statistics concerning the rapid rise in global temperatures during the decade of the 1990s and suggested that this rapid rise was due principally to anthropogenicly generated greenhouse gas emissions, specifically carbon dioxide. This document was taken to be a strong justification for the Kyoto Accord. Featured prominently in the IPCC report was the work of Dr. Michael Mann, Dr. Raymond Bradley, and Dr. Malcolm Hughes (Mann et al., 1998, 1999) [MBH98, MBH99]. These papers featured temperature reconstructions going back as far as 1000 years. The methodology found in Mann et al. employed a statistical technique known as principal components analysis (PCA). Challenges to the way in which PCA was used have arisen from McIntyre and McKitrick (2003, 2005a, 2005b) [MM03, MM05a, MM05b]. The challenges are based on rather subtle mathematical nuances. The discussion and evaluation of the use of PCA to some extent has degenerated in to the battle of competing web blogs: http://www.climateaudit.org, http://www.climate2003.org, and http://www.realclimate.org.

The Chairman of the House Committee on Energy and Commerce along with Chairman of the Subcommittee of Oversight and Investigations have been interested in discovering whether or not the criticisms of Mann et al. are valid and if so, what are the implications. To this end, Committee staff asked for advice as to the validity of the complaints of McIntyre and McKitrick [MM] and related implications. Dr. Wegman formed an ad hoc Committee (Drs. Edward J. Wegman – George Mason University, David W. Scott – Rice University, and Yasmin H. Said – The Johns Hopkins University). The Committee was organized with our own initiative as a pro bono committee.

We have attempted to address several broad issues. We have sought to reproduce the results of MM in order to determine whether their criticisms are valid and have merit. We will also comment on whether issues raised by those criticisms discussed in McIntyre and McKitrick (2005a, 2005b) raise broader questions concerning the assessment of Mann et al. (1998, 1999) in peer review and the IPCC and whether such science assessments involving work of a statistical nature require some type of strengthening to provide reliable guidance for policy makers.

Prior to the work of our committee and independently of our committee, Chairman Barton and Chairman Whitfield wrote letters to Drs. Michael Mann, Raymond Bradley, and Malcolm Hughes as well as to the Intergovernmental Panel on Climate Change and the National Science Foundation. All three of the authors responded, but as lead author Dr. Mann's responses were most extensive. Dr. Mann's responses had something of a confrontational tone. No member of our Committee participated in the design or structure of the questions to Dr. Mann. However, based on his responses and the extensive literature we have reviewed, we will also attempt to address some of our findings explicitly to issues raised by the questions to Dr. Mann and his responses. The specific questions of Chairman Barton and Chairman Whitfield are listed below.

Requests to Drs. Mann, Bradley and Hughes.

Provide:

1. Your *curriculum vitae*, including, but not limited to, a list of all studies relating to climate change research for which you were an author or co-author and the source of funding for these studies.

2. List of all financial support you have received related to your research, including, but not limited to, all private, state, and federal assistance, grants, contracts (including subgrants or subcontracts), or other financial awards or honoraria.

3. Regarding all such work involving federal grants or funding support under which you were a recipient of funding or principal investigator, provide all agreements relating to the underlying grants or funding, including, but not limited to, any provisions, adjustments, or exceptions made in the agreements relating to the dissemination and sharing of research results.

4. Provide the location of all data archives relating to each published study for which you were an author or co-author and indicate: (a) whether this information contains all the specific data you used and calculations you performed, including such supporting documentation as computer source code, validation information, and other ancillary information, necessary for full evaluation and application of the data, particularly for another party to replicate your research results; (b) when this information was available to researchers; (c) where and when you first identified the location of this information; (d) what modifications, if any, you have made to this information since publication of the respective study; and (e) if necessary information is not fully available, provide a detailed narrative description of the steps somebody must take to acquire the necessary information to replicate your study results or assess the quality of the proxy data you used.

5. According to the *Wall Street Journal*, you have declined to release the exact computer code you used to generate your results. (a) Is this correct? (b) What policy on sharing research and methods do you follow? (c) What is the source of this policy? (d) Provide this exact computer code used to generate your results.

6. Regarding study data and related information that is not publicly archived, what requests have you or your co-authors received for data relating to the climate change studies, what was your response, and why?

7. The authors McIntyre and McKitrick (*Energy & Environment*, Vol. 16, No. 1, 2005) report a number of errors and omissions in Mann *et al.*, 1998. Provide a

detailed narrative explanation of these alleged errors and how these may affect the underlying conclusions of the work, including, but not limited to answers to the following questions:

a. Did you run calculations without the bristlecone pine series referenced in the article and, if so, what was the result?
b. Did you or your co-authors calculate temperature reconstructions using the referenced "archived Gaspe tree ring data," and what were the results?
c. Did you calculate the R2 statistic for the temperature reconstruction, particularly for the 15th Century proxy record calculations and what were the results?
d. What validation statistics did you calculate for the reconstruction prior to 1820, and what were the results?
e. How did you choose particular proxies and proxy series?

8. Explain in detail your work for and on behalf of the Intergovernmental Panel on Climate Change, including, but not limited to: (a) your role in the Third Assessment Report [TAR]; (b) the process for review of studies and other information, including the dates of key meetings, upon which you worked during the TAR writing and review process; (c) the steps taken by you, reviewers, and lead authors to ensure the data underlying the studies forming a basis for key findings of the report were sound and accurate; (d) requests you received for revisions to your written contributions; and (e) the identity of the people who wrote and reviewed the historical temperature-record portions of the report, particularly Section 2.3, "Is the Recent Warming Unusual?"

2. BACKGROUND

2.1 Background on Paleoclimate Temperature Reconstruction

Paleoclimatology focuses on climate, principally temperature, prior to the era when instrumentation was available to measure climate artifacts. Many natural phenomena are climate dependent and, where records are still available, these phenomena may be used as proxies to extract a temperature signal. Of course the proxy signals are extremely noisy and thus temperature reconstruction becomes more problematic as one attempts reconstructions further back in time. Climate is not solely a matter of (global) temperature, although concerns with the effects of global warming focus primarily on temperature reconstructions. As just suggested, temperature reconstruction is based on proxy signals contained in the historical records of climate dependent natural phenomena. Table 1 based on Bradley (1999) illustrates the wide variety of these natural phenomena that may be used as proxies. Some proxies measure very low frequency (slowly varying) climatic variables and thus are not useful for measuring average annual temperature changes. Table 2 found in Bradley (1999), which was reproduced from Bradley and Eddy (1991) summarizes a variety of proxies and also indicates their minimum sampling interval as well as the range of years for which they could reasonably be used for temperature reconstruction. The high frequency proxies that could be used on an annual basis include tree rings, ice cores, and corals. In addition to serving as temperature proxies, these measurements are proxies for other climatic variables including, for example, precipitation, chemical composition of air and water, and solar activity.

1. **Glaciological (ice cores)**
    a. Geochemistry (ions and isotopes of oxygen and hydrogen)
    b. Gas content in air bubbles
    c. Trace element and microparticle concentrations
    d. Physical properties (e.g., ice fabric)
2. **Geological**
    a. Marine (ocean sediment cores)
        i. Biogenic sediments
            1. oxygen isotopic composition
            2. faunal and floral abundance
            3. morphological variations
            4. alkenones (from diatoms)
        ii. Inorganic sediments
            1. terrestrial (aeolian) dust and ice-rafted debris
            2. clay mineralogy
    b. Terrestrial
        i. Glacial deposits and features of glacial erosion
        ii. Preglacial features
        iii. Shorelines (Eustatic and glacio-eustatic features)
        iv. Aeolian deposits (loess and sand dunes)
        v. Lacustrine sediments and erosionial features (shorelines)
        vi. Pedological features
        vii. Speleothrems (age and stable isotope composition)
3. **Biological**
    a. Tree rings (width, density, stable isotope composition)
    b. Pollen (type, relative abundance, an/or absolute concentration)
    c. Plant macrofossils (age and distribution)
    d. Insects (assemblage characteristics)
    e. Corals (geochemistry)
    f. Diatoms, ostracods, and other biota in lake sediments (assemblages, abundance, and/or geochemistry)
    g. Modern population distribution (refugia and relict populations in plants and animals)
4. **Historical**
    a. Written records of environmental indicators (parameteorlogical phenomena)
    b. Phonological records

Table 1: Principal Sources of Proxy Data for Paleoclimatic Reconstructions

After Bradley (1999)

Archive	Minimum Sampling Interval	Temporal Range (order:yr)	Potential Information Derived
Historical records	day/hr	$\sim 10^3$	T, P, B, V, M, L, S
Tree rings	yr/season	$\sim 10^4$	T, P, B, V, M, S, C_A
Lake sediments	yr to 20 yr	$\sim 10^4$-10^6	T, B, M, P, V, C_W
Corals	yr	$\sim 10^4$	C_W, L, T, P
Ice cores	yr	$\sim 5 \times 10^4$	T, P, C_A, B, V, M, S
Pollen	20 yr	$\sim 10^5$	T, P, B
Speleothems	100 yr	$\sim 5 \times 10^5$	C_W, T, P
Paleosols	100 yr	$\sim 10^6$	T, P, B
Loess	100 yr	$\sim 10^6$	P, B, M
Geomorphic features	100 yr	$\sim 10^6$	T, P, V, L, P
Marine sediments	500 yr	$\sim 10^7$	T, C_W, B, M, L, P

Table 2: Characteristics of Natural Archives

T = temperature
P = precipitation, humidity, water balance
C = chemical composition of air (C_A) or water (C_W)
B = information on biomass and vegetation patterns
V = volcanic eruptions
M = geomagnetic field variations
L = sea level
S = solar activity

After Bradley and Eddy (1991)

Tree Rings – A cross section of a temperate forest tree shows variation of lighter and darker bands that are usually continuous around the circumference of the tree. These bands are the so-called tree rings and are due to seasonal effects. Each tree ring is composed of large thin-walled cells called early wood and smaller more densely packed thick walled cells called late wood. The average width of a tree ring is a function of many variables including the tree species, tree age, stored carbohydrates in the tree, nutrients in the soil, and climatic factors including sunlight, precipitation, temperature, wind speed, humidity, and even carbon dioxide availability in the atmosphere. Obviously there are many confounding factors so the problem is to extract the temperature signal and to distinguish the temperature signal from the noise caused by the many confounding factors. Temperature information is usually derived from interannual variations in the ring width as well as interannual and intra-annual density variations. Density variations are valuable in paleoclimatic temperature reconstructions because they have a relatively simple growth function that, in mature trees, is approximately linear with age. The density variations have been shown empirically to contain a strong climatic temperature signal. Two values of density are measured within each growth ring: minimum density representing early wood and maximum density representing late wood. Maximum density values are strongly correlated with April to August mean temperatures in trees across the boreal forest from Alaska to Labrador, Schweingruber et al., (1993). Both tree ring width and density data are used in combination to extract the maximal climatic temperature signal.

Climate signal is strongest in trees that are under stress. Trees growing in sites where climate does not limit growth tend to produce rings that are uniform. Trees that are growing close to their extreme ecological range are greatly influenced by climate. Climate variations strongly influence annual growth increments. Two types of stress are commonly recognized, moisture stress and temperature stress. Trees growing in semiarid regions are limited by water availability and thus variations in ring width reflect this climatic moisture signal. Trees growing near to their ecological limits either in terms of latitude or altitude show growth limitations imposed by temperature and thus ring width variations in such trees contain a relatively strong temperature signal. However, the biological processes are extremely complex so that very different combinations of climatic conditions may cause similar ring width increments. Tree growth and carbohydrate production by a tree in one year will precondition the tree for strong growth in the subsequent year so that there is a strong autocorrelation in the ring width time series. Photosynthetic processes are accelerated with the increased availability of carbon dioxide in the atmosphere and, hence, it is conjectured that ring growth would also be correlated with atmospheric carbon dioxide; see Graybill and Idso (1993). In addition, oxides of nitrogen are formed in internal combustion engines that can be deposited as nitrates also contributing to fertilization of plant materials. It is clear that while there are temperature signals in the tree rings, the temperature signals are confounded with many other factors including fertilization effects due to use of fossil fuels.

Wider rings are frequently produced during the early life of a tree. Thus the tree rings frequently contain a low frequency signal that is unrelated to climate or, at least, confounded with climatic effects such as temperature. In order to use tree rings as a temperature signal successfully, this low frequency component must be removed. This is typically done by a nonlinear parametric trend fit using a polynomial or modified exponential curve. Because the early history of tree rings confounds climatic signal with low frequency specimen specific signal, tree rings are not usually effective for accurately determining low frequency, longer-term effects. Once there is reasonable confidence that the tree ring signal reflects a temperature signal, and then a calibration is performed using the derived tree ring data and instrumented temperature data. The assumption in this inference is that when tree ring structure observed during the instrumented period that is similar to tree ring structure observed in the past, both will have correspondingly similar temperature profiles. As pointed out earlier, many different sets of climatic conditions can and do yield similar tree ring profiles. Thus tree ring proxy data alone is not sufficient to determine past climate variables. See Bradley (1999) for a discussion of the fitting and calibration process for dendritic-based temperature reconstruction.

Ice Cores – The accumulated past snowfall in the polar caps and ice sheets provide a very useful record for paleoclimate reconstruction. We shall refer to ice cores in this section even though strictly speaking there is a combination of snow and ice. Somewhat compressed old snow is called a firn. The transition from snow to firn to ice occurs as the weight of overlying material causes the snow crystals to compress, deform and recrystalize in more compact form. As firn is buried beneath subsequent snowfalls, density is increased as air spaces are compressed due to mechanical packing as well as plastic deformation. Interconnected air passages may then be sealed and appear as individual air bubbles. At this point the firn becomes ice. Paleoclimatic information derived from ice cores is obtained from four principal mechanisms: 1) analysis of stable isotopes of water and atmospheric oxygen; 2) analysis of other gases in the air bubbles in the ice; 3) analysis of dissolved and particulate matter in the firn and ice; and 4) analysis of other physical properties such as thickness of the firn and ice.

The mechanism by which stable isotopes of oxygen and hydrogen carry a temperature signal is as follows. An oxygen atom can exist in three stable forms with atomic weights of 16, 17 or 18. Oxygen with an atomic weight of 16 makes up 99.76% of all oxygen atoms. Similarly, hydrogen can exist in two stable forms with atomic weights of one or two, the latter being called deuterium. Hydrogen with atomic weight of one comprises 99.984% of all hydrogen. Thus water molecules can exist in several heavy forms when compared with normal water, which is made up of two atomic-weight-1 hydrogen atoms and one atomic-weight-16 oxygen atom. The vapor pressure of normal water is higher than the heavier forms of water with evaporation resulting in a vapor that is relatively speaking poor in the heavier forms of water. Conversely, the remaining water will be enriched in water containing the heavier isotopes. When condensation occurs, the lower vapor pressure of water containing the heavier isotopes will cause that water to condense more rapidly than normal water. The greater the fall in temperature, the more condensation will occur; hence, the water vapor will exhibit less heavy isotope concentration when compared to the original (sea) water. Thus the relative isotope concentrations in the condensate will be a direct indicator of the temperature at which condensation occurred.

In addition to the relative heavy/light isotope ratios, the trapped bubbles in ice cores provide a record of atmospheric concentrations of trace gases including greenhouse gases such as carbon dioxide, methane and nitrous oxide. In addition the ice cores contain record of aerosols and dust content resulting from volcanic eruptions and other changes in particulate content in the atmosphere. The relative atmospheric concentrations of greenhouse gases as well as aerosol and particulate content coupled with other climate information gives insight into both the importance of these as drivers of temperature as well as how these drivers might couple in either a positive or negative feedback sense.

Corals – The term "coral" refers to the biological order *Scleractinia*, which have hard calcium-based skeletons supporting softer tissues. An important subgroup for paleoclimate studies is the reef-building corals in which the coral polyp lives symbiotically with single-celled algae. These algae produce carbohydrates by means of photosynthesis and are affected by water depth, water turbidity, and cloudiness. Much of the carbohydrates diffuse away from the algae providing food to the coral polyp, which in turn provide a protective environment for the algae. Reef-building corals are strongly affected by temperature and, as temperature drops, the rate of calcification drops with lower temperature potentially presaging the death of the colony. Coral growth rates vary over a year and can be sectioned and x-rayed to reveal high- and low-density bands. High density layers are produced during times of higher sea surface temperatures. Thus not unlike tree rings, data on corals also can be calibrated to estimate (sea) surface temperatures.

2.2 Background on Principal Components

Principal Components

Principal Component Analysis (PCA) is a method for reducing the dimension of a high dimensional data set while preserving most of the information in those data. Dimension is here taken to mean the number of distinct variables (proxies). In the context of paleoclimatology, the proxy variables are the high dimensional data set consisting of several time series that are intended to carry the temperature signal. The proxy data set in general will have a large number of interrelated or correlated variables. Principal component analysis tries to reduce the dimensionality of this data set while also trying to explain the variation present as much as possible. To achieve this, the original set of variables is transformed into a new set of variables, called the principal components (PC) that are uncorrelated and arranged in the order of decreasing "explained variance." It is hoped that the first several PCs explain most of the variation that was present in the many original variables. The idea is that if most of the variation is explained by the first several principal components, then the remaining principal components may be ignored for all practical purposes and the dimension of the data set is effectively reduced.

Noise, White and Red

In a statistical signal processing sense, noise is defined to be unknown external factors or fluctuations in signals. Noise is typically modeled as what are called random innovations, random meaning not predictable. Signals of interest may be actual audio signals, signals in the electromagnetic spectrum, or a known function to which simple statistical noise has been added. In the paleoclimatology context, noise is the unpredictable portion of the climate signal caused by random variations in the factors related to tree ring growth, ice core development, or coral growth. Noise is often thought of in terms of periodic variation. There are many types of noise with varying frequencies, each classified by a color. The color names for these different types of sounds are derived from an analogy between the spectrum of frequencies and the equivalent spectrum of light wave frequencies. That is, if the pattern of 'blue noise' were translated into light waves, the resulting light would be blue.

White noise, has equal power density across the entire frequency spectrum, that is, it has constant energy at all frequencies. When this is graphically represented, white noise has a flat power spectral density. In a practical example, white noise is what is used to refer to that steady, even soothing sound produced when tuning in to an unused radio or TV frequency. White noise has an equal amount of energy *per frequency band* in contrast to pink noise, which has an equal amount of energy per octave. Pink noise has a frequency spectrum that is flat in logarithmic space. The power density of pink noise, compared with white noise, decreases by 3 dB (decibels) per octave. It is said that pink noise is the most soothing sound to the human ear. Pink noise has the same frequency distribution as falling rain.

Red noise is similar to pink noise, but it has relatively more energy at lower frequencies than pink noise. Red noise has a power density that decreases 6 dB per octave as the frequency increases. Of course, red noise was named after a connection with red light, which is on the low end of the visible light spectrum. Mathematically speaking, integrating white noise produces red noise. Red noise in the paleoclimatology context comes from the fact that tree rings have correlation from year to year, that is, if a tree grows well in a given year, it will store carbohydrates and will tend to have a good year of growth the following year as well. Red noise in the paleoclimatology context is modeled by a first-order autoregressive model.

Autoregressive, Moving Average and ARMA Models

Autoregressive, moving averages, and ARMA models are statistical time series models. An autoregressive model of order p means that the present value of the time series depends only on the p most recent past values of the time series. The dependence is taken to be linear. If $p = 1$, then we say that the process is a first order autoregressive process as indicated for the red noise model above. A moving average process of order q is formed by taking a weighted average of q uncorrelated white noise terms, that is, zero mean constant variance terms. The moving average means that the next value of the moving average process drops off the oldest term from the average and adds a new term. Autoregressive moving average (ARMA) models, sometimes called Box-Jenkins models, are also used to model time series data. These models are used for understanding and predicting future values in the series. There are two parts to the ARMA model, an autoregressive (AR) part and a moving average (MA) part.

Gaussian Noise and Long Memory Processes

Although we have not specified the probability distribution of the random innovations, it is often the case that a normal or Gaussian probability distribution is appropriate to model noise or what we have called random innovations. The basic paleoclimatology model is taken to be a trend with superimposed white or red noise usually with a Gaussian distribution. The Gaussian distribution assumption is a convenient one mathematically. Random (or stochastic) processes whose autocorrelation function, decaying as a power law, sums to infinity are known as long range correlations or long range dependent processes. Because the decay is slow, as opposed to exponential decay, these processes are said to have long memory. Applications exhibiting long-range dependence include Ethernet traffic, financial time series, geophysical time series such as variation in temperature, and amplitude and frequency variation in EEG signals. Fractional Brownian motion is a self-similar Gaussian process with long memory. The Box-Jenkins ARMA models described in the previous section are all short-term memory processes.

In reality, temperature records and hence data derived from proxies are not modeled accurately by a trend with superimposed noise that is either red or white. There are complex feedback mechanisms and nonlinear effects that almost certainly cannot be modeled in any detail by a simple trend plus noise. These underlying process structures appear to have not been seriously investigated in the paleoclimate temperature reconstruction literature. Cohn and Lin (2005) make the case that much of natural time series, in their case hydrological time series, might be modeled more accurately by a long memory process. Long memory processes are stationary processes, but the corresponding time series often make extended sojourns away from the stationary mean value and, hence, mimic trends such as the perceived hockey stick phenomena.

One type of such long memory processes is a process driven by fractional Gaussian noise (fractional Brownian motion). An object with self-similarity is exactly or approximately similar to a part of itself. For example, many coastlines in the real world are self-similar since parts of them show the same properties at many scales. Self-similarity is a common property of many fractals, as is the case with fractional Brownian motion. A serious effort to model even the present instrumented temperature record with sophisticated process models does not appear to have taken place.

2.3 Background on Social Networks

Networks, Relations and Structure

A social network is a mathematical structure made of nodes, which are generally taken to represent individuals or organizations. A network graph illustrates how the nodes are connected. Social network analysis (also called *network theory*) has emerged as a key technique and a topic of study in modern sociology, anthropology, social psychology and organizational theory. Research has demonstrated that social networks, operating on many levels, from families up to the level of nations, play a critical role in determining the way problems are solved, organizations are run, and the degree to which individuals succeed in achieving their goals. The shape of the social network helps determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections (weak ties) to individuals outside the main network. More "open" networks, with many weak ties and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties. In other words, a group of friends who only do things with each other already share the same knowledge and opportunities. Yet a group of individuals with connections to other social worlds is likely to have access to a wider range of information. It is better for individual success to have connections to a variety of networks rather than many connections within a single network. Similarly, individuals can exercise influence or act as brokers within their social networks by bridging two networks that are not directly linked (called filling social holes).

Networks operate anywhere that energy and information are exchanged: between neurons and cells, computers and people, genes and proteins, atoms and atoms, and people and people. Social theories are built on more than just metaphors. Social network analysis assumes that interpersonal ties matter, whether they exist among individuals, organizations or countries. Interpersonal connections matter because they are conduits for the transmission of information, goods, behavior and attitudes. Ties and connections form networks, which can be analyzed. The main goal of social network analysis is the detection and interpretation of patterns of social ties among people, nations, or organizations involved in social relationships.

There are several key concepts at the heart of network analysis. We outline these concepts next and then define a social network.

Actor: Social network analysis is concerned with understanding the linkages among social entities and the implications of these linkages. The social entities are referred to as actors. Actors do not necessarily have the desire or the ability to act. Most social network applications consider a collection of actors that are all of the same type. These are known as one-mode networks.

Relational Tie: Social ties link actors to one another. The range and type of social ties can be quite extensive. A tie establishes a linkage between a pair of actors. Examples of ties include the evaluation of one person by another (such as expressed friendship, liking, respect), transfer of material resources (such as business transactions, lending or borrowing things), association or affiliation (such as jointly attending the same social event or belonging to the same social club), behavioral interaction (talking together, sending messages), movement between places or statues (migration, social or physical mobility), physical connection (a road, river, bridge connecting two points), formal relations such as authority and biological relationships such as kinship or descent.

Dyad: A linkage or relationship establishes a tie at the most basic level between a pair of actors. The tie is an inherent property of the pair. Many kinds of network analysis are concerned with understanding ties among pairs and are based on the dyad as the unit of analysis.

Triad: The analysis of a subset of three actors (a triad) and the possible ties among them is motivated and informed by balance theory. Balance theory asks whether or not a triad is transitive or balanced. A transitive triad is characterized by transitive relations such as if actor *i likes* actor *j*, and actor *j likes* actor *k*, then actor *i* also *likes* actor *k*. A balanced triad means that if actors *i* and *j* like each other, then *i* and *j* should have similar evaluations of a third actor, whereas if they dislike each other then they are expected to differ in their evaluations.

Subgroup: Dyads are pairs of actors and associated ties, triads are triples of actors and associated ties. We can define a subgroup of actors as any subset among actors with associated ties. Locating and studying these subgroups using specific criteria is one of the primary objectives of social network analysis.

Group: Network analysis is not only concerned with collections of dyads, triads, or subgroups. Social network analysis has the ability to model the relationships among systems of actors. A group is a collection of actors on which ties are measured.

Relation: The collection of ties of a specific kind among members of a group is called a relation, for example, the set of friendships among pairs of children in a classroom or the set of formal diplomatic ties maintained by pairs of nations in the world. A relation refers to the collection of ties of a given kind measured on pairs of actors from a specified actor set.

Social Network: We are now in a position to define a social network. A social network consists of a finite set or sets of actors and the relation or relations defined on them. The presence of relational information is a significant feature of a social network.

Computational Facets of Social Network Analysis

The main goal of social network analysis is the detection and interpretation of patterns of social ties among actors. Social network analysis may be viewed as a broadening or generalization of standard data analytic techniques and applied statistics that focus on observational units and their characteristics. Complex network data sets may contain information about the characteristics of the actors (such as the gender of people in a group or the GNP of nations of the world) as well as structural variables. Network problems naturally give rise to graphs. The structural and compositional variables necessary for social network analysis often result in complicated data sets that must be modeled with sophisticated graph theoretic, algebraic and statistical methods. The underlying mathematical frameworks used to build social network models are called graphs. A graph is a discrete structure consisting of vertices (nodes) and edges (links), where the vertices correspond to the objects, and the edges to the relations of the structure to be modeled.

A network consists of a graph and additional information on the vertices or lines of the graphs. Names of people or businesses or countries represent additional information on vertices. Line values are numbers for arcs and edges that indicate the strength of relationships between actors. This flexible definition allows a wide variety of empirical phenomena to be modeled as networks.

Properties of vertices are used to find and interpret patterns of ties in a network. Social networks are often complicated and may be large. Partitions are used to reduce a network so that different facets can be studied.

- Partitions – A partition of a network is a classification or clustering of the vertices in the network so that each vertex is assigned to exactly one class or cluster. Partitions may specify some property that depends on attributes of the vertices. Partitions divide the vertices of a network into a number of mutually exclusive subsets. That is, a partition splits a network into parts. We can produce a local view defined by a selected class of vertices that consists of all of the structural ties between nodes in the selected class of vertices. Partitions are also sometimes called blocks or blockmodels. These are essentially a way to cluster actors together in groups that behave in a similar way.

- Allegiance – Allegiance measures the support that an actor provides for the structure of his block. An actor supports his block by having internal block edges. A measure of this is the total number of edges that an actor has internal to his block. An actor supports his block by not having external edges from the block to other actors or blocks. A measure of this is the total number of possible external edges minus the total number of existing external edges. The allegiance for a block is a weighted sum of a measure of internal allegiance and a measure of external allegiance. The overall allegiance for a social network is the sum of the allegiances for the individual blocks. If the overall allegiance is positive then a good partition was made. The partitioning continues recursively until a new partition no longer contributes to a positive allegiance.

- Global View – We may want a global view of a network that allows us to study relationships among classes.

- Cohesion – Solidarity, shared norms, identity, collective behavior, and social cohesion are considered to emerge from social relations. The first concern of social analysis is to investigate who is related and who is not. The general hypothesis assumes that people who match on social characteristics will interact more often and people who interact regularly will foster a common attitude or identity. Social networks usually contain dense pockets of people who stick together. They are called cohesive subgroups and usually more than interaction joins the people involved. People who interact intensively are likely to consider themselves as a social group. This phenomenon is known as homophily: "birds of a feather flock together". There are several techniques that detect cohesive subgroups in social networks. All of these techniques are based on the ways in which the vertices are interconnected. These techniques are used to investigate whether a cohesive group represents an emergent or established social group. Social cohesion is used to describe structural concepts of density and connectedness. Density refers to the number of links between vertices. A network is strongly connected if it contains paths between all of its vertices and is weakly connected when semi-paths connect all of its vertices. Connected networks and networks with high average degree are thought to be more cohesive. There are several techniques to detect cohesive subgroups based on density and connectedness.

- Affiliations – Membership in an organization or participation in an event is a source of social ties. An affiliation is a relationship between people and an organization. Affiliations are often institutional or structural and tend to be less personal as they result from private choices to a lesser degree than sentiments and friendship.

- Brokerage – Social relations can be considered to be channels that transport information, services, or goods between people or organizations. From a bird's eye view, social structure helps to explain how information, goods or even attitudes and behavior diffuses within a social system. Network analysis reveals social structure and helps to trace the routes that goods and information may follow. Some social structures permit rapid diffusion of information, whereas others contain sections that are difficult to reach. We can also focus on the position of specific people or organizations within the network. In general, being well connected is advantageous. Contacts are necessary to have access to information and help. The number and intensity of a person's ties are called his or her sociability or social capital. Social capital is known to correlate positively to age and education in Western societies. Some people occupy central or strategic positions within the system of channels and are crucial for the transmission process. Some positions may exert pressure on their occupants, but they also yield power and profit. The direction of ties is not very important in social network structures that capture the exchange of information.

- Centrality – This is one of the oldest concepts in network analysis. Most social networks contain people or organizations that are central. Because of their position, they have better access to information, and better opportunity to spread information. This is known as the ego-centered-approach to centrality. The network is centralized from socio-centered perspective. The notion of centrality refers to the positions of individual vertices within the network, while centralization is used to characterize an entire network. A network is highly centralized if there is a clear boundary between the center and the periphery. In a highly centralized network, information spreads easily, but the center is indispensable for the transmission of information.

There are several ways to measure the centrality of vertices and the centralization of networks. The concepts of vertex centrality and network centralization are best understood by considering undirected communication networks. If social relations are channels that transmit information between people, central people are those people who have access to information circulating in the network or who may control the circulation of information.

The accessibility of information is linked to the concept of distance. If you are closer to the other people in the network, the paths that information has to follow to reach you are shorter, so it is easier for you to acquire information. If we take into account direct neighbors only, the number of neighbors (the degree of a vertex in a simple undirected network) is a simple measure of centrality. If we also want to consider other indirect contacts, we use closeness centrality, which measures our distance to all other vertices in the network. The closeness centrality of a vertex is higher if the total distance to all other vertices is shorter.

The importance of a vertex to the circulation of information is captured by the concept of betweenness centrality. From this perspective, a person is central if he or she is a link in more information chains between other people in the network. High betweenness centrality indicates that a person is an important intermediary in the communication network. Information chains are represented by geodesics and the betweenness centrality of a vertex is simply the proportion of geodesics between other pairs of vertices that include the vertex. The centralization of a network is higher if it contains very central vertices as well as very peripheral vertices.

3. LITERATURE REVIEW OF GLOBAL CLIMATE CHANGE RESEARCH

Michael Mann's Dissertation and Related Work

In his 1998 dissertation, Michael Mann used instrumental data and multi-proxy datasets to observe climate variability over the past few centuries. He also used a simplified coupled ocean-atmosphere model to describe mechanisms that may contribute to the climate variability. In his dissertation, Dr. Mann described a 70 to 100 year oscillation in the climate signal formed by the proxy and instrumental data. He notes that this century-scale variation in the climate involves a combination of meridional overturning (the circulation of cold water in the ocean) and gyre-scale circulation.

After being awarded his doctorate, Dr. Mann, together with his colleagues Dr. Bradley and Dr. Hughes, continued multi-proxy reconstruction research with his 1998 paper, *Global-Scale Temperature Patterns and Climate Forcing over the Past Six Centuries*, [MBH98]. In this paper, he attempts to use PCA to find an eigenbasis (a new coordinate system where the axes are the eigenvectors, or principal components that represent the significant relationships in the data) for the multi-proxy data series for the period 1610-1995. He also uses a multivariate regression method to observe possible forcing agents, or contributors to warming. Dr. Mann uses linear relationships between possible forcing agents (greenhouse gases, solar irradiance and volcanic aerosols) and climate in previous studies by R.S. Bradley and T.J. Crowley as a basis for regression analysis. He reports that the results are a large spike in greenhouse gas forcing in the 20th century. Additionally, he notes that 1995 and 1997 were likely the hottest years since 1400 AD within a 99.7% level of certainty.

In 1999, Dr. Mann and colleagues supplemented MBH98 by a new paper, *Northern Hemisphere Temperatures during the Past Millennium: Inferences, Uncertainties, and Limitations* MBH99. In this work they used similar methods to reconstruct temperatures further back to the beginning of the millennium. Although uncertainties are magnified with each previous year of reconstruction, their results suggested that 20th century warming counters a millennial-scale cooling trend and that the 1990s was likely the hottest decade in the millennium, with moderate certainty.

McIntyre and McKitrick

After MBH99, Stephen McIntyre and Ross McKitrick [MM03] published their critique of the 1998 paper, citing calculation errors, unjustified truncation or extrapolation of source data, obsolete data, geographical location errors and incorrect calculation of principal components. They also claimed that using the MBH98 methodology and the Northern Hemisphere average temperature index for the period 1400-1980 shows that temperatures in the 15th century exceeded those of the late 20th century. In particular, they claim that MBH98's incorrect usage of PCA alone resulted in the well-known "hockey stick" shape.

In a 2004 corrigendum, Mann et al. replied to these criticisms, contending that McIntyre and McKitrick's finding resulted from an elimination of over 70% of the 15th century proxy data used by MBH98. They also assert that MM03's results fail independent cross-validation tests. In subsequent response papers, MM05a and MM05b noted that the data eliminated in their earlier critique of MBH98 was restricted to two proxy series, the Gaspe cedar ring-width series, and the first principal component from the North American tree ring network in MBH98. In the case of the first principal component, McIntyre and McKitrick stated that the mean was not centered correctly over the period of analysis, 1400-1980. Instead of subtracting the mean of each data series between the years 1400 and 1980, they subtracted the mean of 1902-1980. McIntyre and McKitrick state that this decentering of the mean causes the explained variance of certain proxies to be inflated, namely the proxy series that causes the hockey stick shape. Subsequently, that particular proxy series is chosen as the principal component, indicating it is the most significant correlation in the data. With regard to the Gaspe cedar tree ring series, McIntyre and McKitrick state that Mann 1998 did not use archived data, but rather made an extrapolation in which they misrepresented the start date of the series. They also state that this extrapolation depresses the early 15th century results. Lastly, they note that the underlying dataset up to 1421 is based only on one tree, and only on two trees up to 1447.

Mann and Rutherford's 2005 paper in turn responded to these new criticisms, stating that McIntyre and McKitrick's misunderstanding of their methodology and use of an incorrect version of the proxy dataset is the source of the discrepancy in their results. They argue that the Mann et al. 1998 implementation calculates PC series of proxy networks over progressively longer intervals, which allows for the use of the maximum amount of data. For example, if there were 50 proxy series, and only 10 date back to AD 1400, then calculating one set of PC would eliminate 40 of the 50 series available back to AD 1600. By using two different intervals, 1400-1980 and 1600-1980 in this example, all proxy series can be utilized. Mann et al. contend that this misunderstanding is what led to the elimination of data prior to 1500 AD and is also what gives rise to the warmth in the 15th century of McIntyre and McKitrick's version of MBH98.

To address the extrapolation critique, Mann et al. terminated the 1971 calibration period in which they filled in missing proxy values in the multi-proxy PC network between 1971 and 1980. They also approached the reconstruction using a different method, regularized expectation maximization (REGEM), and yielded the same results. They then conclude that their reconstruction is robust and reproducible, based on their use of an independent Climate Field Reconstruction method (the REGEM method) and their use of individual proxies instead of the multi-proxy PC representation used in Mann et al.1998.

Other Notable Works

While Mann et al. have focused much of their work on high frequency proxies, or those proxies that provide data on climate variability on a decadal or even yearly scale, Jan Esper and colleagues have investigated the effect of using low-frequency proxies that preserve data on climate variability on a centennial scale in their paper *Low-Frequency Signals in Long Tree-Ring Chronologies for Reconstructing Past Temperature Variability*. Esper et al. contend that preserving multi-centennial variability in tree-ring records is critical in comparing the temperatures of the Medieval Warming Period

(MWP) and those of the 20th century. By carefully selecting tree-ring chronologies from fourteen sites in the Northern Hemisphere (NH) extratropics, Esper et al. produced a reconstruction that preserved the multi-centennial variation, as well as supported the large-scale occurrence of the MWP over the NH extratropics. Using the regional curve standardization (RCS) method for their chronologies, Esper et al. found that there were significant multi-centennial differences between Mann et al. 1998 and their reconstruction. These differences may be explained by the fact that Mann et al.'s analysis includes data from the tropical and subtropical Northern Hemisphere whereas Esper's analysis includes only the extra tropic region.

In their 2005 paper *Highly Variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data*, Moberg et al. further studied the use of low-resolution proxy data in order to preserve multi-centennial variability in climate reconstructions. Specifically, they focused on lake and ocean sediment cores, which tend to have a lower time resolution, but provide climate information at multi-centennial timescales that may not be captured by tree-ring data. Moberg et al. reconstructed Northern Hemisphere temperatures for the past 2,000 years by combining low-resolution proxies with tree-ring data. Their reconstruction shows a larger multi-centennial variability than most previous multi-proxy reconstructions. Furthermore, their reconstruction depicted high temperatures that are comparable to the 20th century temperatures in the period AD 1000-1100. Their results also suggest a natural trend in multi-centennial variability that is likely to continue.

Following in this same trend, Hans von Storch et al., in their 2004 paper *Reconstructing Past Climate from Noisy Data*, used a coupled atmosphere-ocean model simulation of the past 1000 years to test the skill of past empirical reconstructions, specifically those of Mann et al. 1998, 1999. They found that while previous millennium based multi-proxy records find small amplitude variations followed by a clear warming trend in the past two centuries, the centennial variability of the Northern Hemisphere temperature is underestimated by these regression based methods. Their results also suggest that actual variability may have been at least twice as large as the variability predicted in these past studies. The authors surmise that this conclusion probably applies to most regression-based methods of analysis and that other methods that estimate past temperatures with physical (instead of statistical) methods or regression methods that address retention of low-frequency variability in proxies may be free from this critique.

Another 2005 paper *Are Multiproxy Climate Reconstructions Robust?* by Gerd Burger and Ulrich Cubasch questions whether these methods are statistically significant enough to be able to make robust conclusions. Burger and Ulrich describe sixty-four climate reconstructions, based on regression of temperature fields on multi-proxy networks, which are each distinguished by at least one of six standard criteria of this method. By combining these criteria Burger and Ulrich define numerous variants on millennial histories. No one criterion can account for the number of variations and no particular variant is more valid than another. Even the variant with the best reduction of error statistic is the furthest variation from the climate history of Mann et al. 1998. Burger and

Cubasch conclude that the regression model is not valid when applied in an extrapolative manner, as in climate reconstruction.

In a 2006 paper, *The Spatial Extent of 20th Century Warmth in the Context of the Past 1200 Years*, Timothy Osborn and Keith Briffa examine the most prominent anomalies in proxy records since AD 1200. They state that the most significant anomaly is the geographical extent of warmth in the middle to late 20th century. They also detail anomalies during AD 890 to 1170 and AD 1580 to 1850 as being consistent with the concepts of a Medieval Warming Period (MWP) and Little Ice Age (LIA), respectively. However, they found that when comparing these anomalies with instrumental temperatures of the 20th century, the spatial or geographical extent of this warm anomaly is far greater than that of the MWP or LIA. Their study consisted of fourteen regional temperature-related proxy records. Since it is not possible to conduct a direct comparison between proxy records and instrumental temperatures, the proxy data analysis was conducted with each series normalized over the 1856 to 1995 period, or the period where proxy and instrumental data overlap. Relative to a decadal time scale, Osborn and Briffa found supporting evidence for the MWP and LIA, but their geographical reach appeared restricted since these anomalies were sensitive to specific proxy records.

Analysis

While the work of Michael Mann and colleagues presents what appears to be compelling evidence of global temperature change, the criticisms of McIntyre and McKitrick, as well as those of other authors mentioned are indeed valid. Because the error and uncertainty involved in climate reconstructions is magnified with each preceding year, the ability to make certain conclusions about the climate at the beginning of the millennium is not very robust. This is even less robust considering the inability to actually calculate an accurate uncertainty for these reconstructions. Additionally, the work of Esper, Von Storch and Moberg make valid arguments for the inclusion of low-frequency proxies as well as the inability of PCA to effectively measure variations on a multi-centennial scale. This pitfall of PCA is further complicated by its tendency for misuse during the calibration process, specifically the decentering of the mean that McIntyre and McKitrick mention.

The papers of Mann et al. in themselves are written in a confusing manner, making it difficult for the reader to discern the actual methodology and what uncertainty is actually associated with these reconstructions. Vague terms such as “moderate certainty” (Mann et al. 1999) give no guidance to the reader as to how such conclusions should be weighed. While the works do have supplementary websites, they rely heavily on the reader’s ability to piece together the work and methodology from raw data. This is especially unsettling when the findings of these works are said to have global impact, yet only a small population could truly understand them. Thus, it is no surprise that Mann et al. claim a misunderstanding of their work by McIntyre and McKitrick.

In their works, Mann et al. describe the possible causes of global climate change in terms of atmospheric forcings, such as anthropogenic, volcanic, or solar forcings. Another questionable aspect of these works is that linear relationships are assumed in all forcing-climate relationships. This is a significantly simplified model for something as complex as the earth's climate, which most likely has complicated nonlinear cyclical processes on a multi-centennial scale that we do not yet understand. Mann et al. also infer that since there is a partial positive correlation between global mean temperatures in the 20th century and CO_2 concentration, greenhouse-gas forcing is the dominant external forcing of the climate system. Osborn and Briffa make a similar statement, where they casually note that evidence for warming also occurs at a period where CO_2 concentrations are high. A common phrase among statisticians is *correlation does not imply causation*. The variables affecting earth's climate and atmosphere are most likely to be numerous and confounding. Making conclusive statements without specific findings with regard to atmospheric forcings suggests a lack of scientific rigor and possibly an agenda.

It is also interesting to note that Mann's dissertation focused on 70 to 100 year climate signal variability, yet his future work does not have a similar component. His subsequent papers focus heavily on tree ring measurements, which provide data on a decadal or yearly scale. In later work, he also makes no mention of the ocean circulation variables, which he describes in his thesis as being integral to the variation in climate. If this type of forcing is a natural variable, it makes the conclusions about atmospheric forcings seem incomplete.

The work initiated by Mann and his colleagues is still in its infancy, and as such further study, the use of wider proxy networks and the development of more sophisticated climate models will all be necessary future steps in propagating this research. It is not expected or likely that after preliminary research, definitive conclusions can be made about the earth's climate over the past millennium.

4. RECONSTRUCTIONS AND EXPLORATION OF PRINCIPAL COMPONENT METHODOLOGIES

Mann et al. (2005) identify two major methods of climate reconstruction, which they describe respectively as climate field reconstruction (CFR) methods[5] and what they describe as *simple* climate-plus-scale (CPS) methods. CFR methods are claimed to "assimilate proxy records into a reconstruction of underlying patterns of past climate change" and among papers identified as using these methods are MBH 98, Evans et al. (2002), Luterbacher et al. (2002), Rutherford et al. (2005) and Zhang et al. (2004). In contrast CPS methods are said to "composite a number of proxy series and scales the resulting composite against a target (e.g. Northern Hemisphere temperature) instrumental series." Examples of papers using the CPS methods include Jones et al. (1998), Crowley and Lowery (2000), Briffa et al. (2001), Esper et al. (2002), Mann and Jones (2003) and Crowley et al. (2003). Although the language describing both of these methods seems somewhat obscure, it would appear that CFR methods are just principal component methods as describe earlier and in the appendix and that CPS methods are just simple averaging of climate proxies and then scaling them to actual temperature records.

The key issue in dispute is the CFR methodology as used in MBH98 and MBH99. The description of the work in MBH98 is both somewhat obscure and as others have noted incomplete. The essence of the discussion is as follows. Principal component methods are normally structured so that each of the data time series (proxy data series) are centered on their respective means and appropriately scaled. The first principal component attempts to discover the composite series that explains the maximum amount of variance. The second principal component is another composite series that is uncorrelated with the first and that seeks to explain as much of the remaining variance as possible. The third, fourth and so on follow in a similar way. In MBH98/99 the authors make a simple seemingly innocuous and somewhat obscure calibration assumption. Because the instrumental temperature records are only available for a limited window, they use instrumental temperature data from 1902-1995 to calibrate the proxy data set. This would seem reasonable except for the fact that temperatures were rising during this period. So that centering on this period has the effect of making the mean value for any proxy series exhibiting the same increasing trend to be decentered low. Because the proxy series exhibiting the rising trend are decentered, their calculated variance will be larger than their normal variance when calculated based on centered data, and hence they will tend to be selected preferentially as the first principal component. (In fact the effect of this can clearly be seen RPC no. 1 in Figure 5 in MBH98.). Thus, in effect, any proxy series that exhibits a rising trend in the calibration period will be preferentially added to the first principal component.

The centering of the proxy series is a critical factor in using principal components methodology properly. It is not clear that Dr. Mann and his associates even realized that

[5]CFR methods are essentially the methodology that was used in MBH98 and MBH99. However, the methods in MBH98 and MBH99 were not formally called CFR methods, the climate field reconstruction phrase being coined only later.

their methodology was faulty at the time of writing the MBH paper. The net effect of the decentering is to preferentially choose the so-called hockey stick shapes. While this error would have been less critical had the paper been overlooked like many academic papers are, the fact that their paper fit some policy agendas has greatly enhanced their paper's visibility. Specifically, global warming and its potentially negative consequences have been central concerns of both governments and individuals. The 'hockey stick' reconstruction of temperature graphic dramatically illustrated the global warming issue and was adopted by the IPCC and many governments as the poster graphic. The graphics' prominence together with the fact that it is based on incorrect use of PCA puts Dr. Mann and his co-authors in a difficult face-saving position. We have been to Michael Mann's University of Virginia website and downloaded the materials there. Unfortunately, we did not find adequate material to reproduce the MBH98 materials.

We have been able to reproduce the results of McIntyre and McKitrick (2005b). While at first the McIntyre code was specific to the file structure of his computer, with his assistance we were able to run the code on our own machines and reproduce and extend some of his results. In Figure 4.1, the top panel displays PC1 simulated using the MBH98 methodology from stationary trendless red noise. The bottom panel displays the MBH98 Northern Hemisphere temperature index reconstruction.

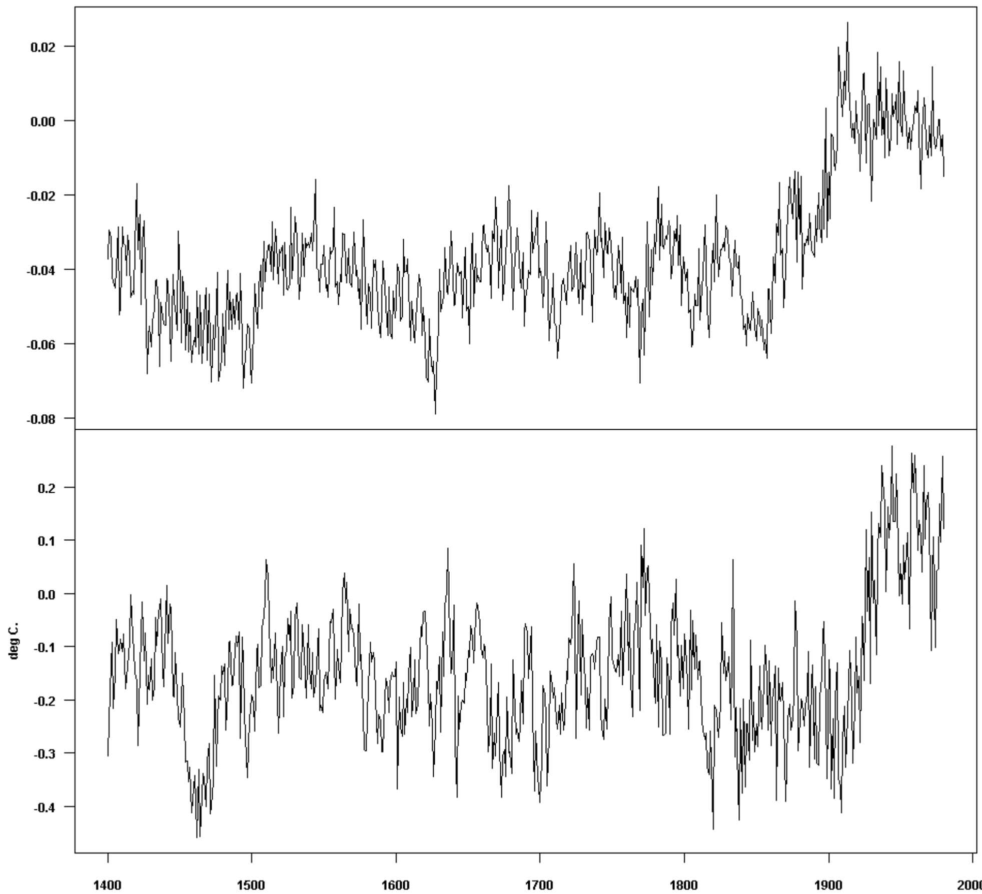

Figure 4.1: Reproduced version of Figure 1 in McIntyre and McKitrick (2005b). Top panel is PC1 simulated using MBH 98 methodology from stationary trendless red noise. Bottom panel is the MBH98 Northern Hemisphere temperature index reconstruction.

Discussion: The similarity in shapes is obvious. As mentioned earlier, red noise exhibits a correlation structure, which, although it is a stationary process, to will depart from the zero mean for minor sojourns. However, the top panel clearly exhibits the hockey stick behavior induced by the MBH98 methodology.

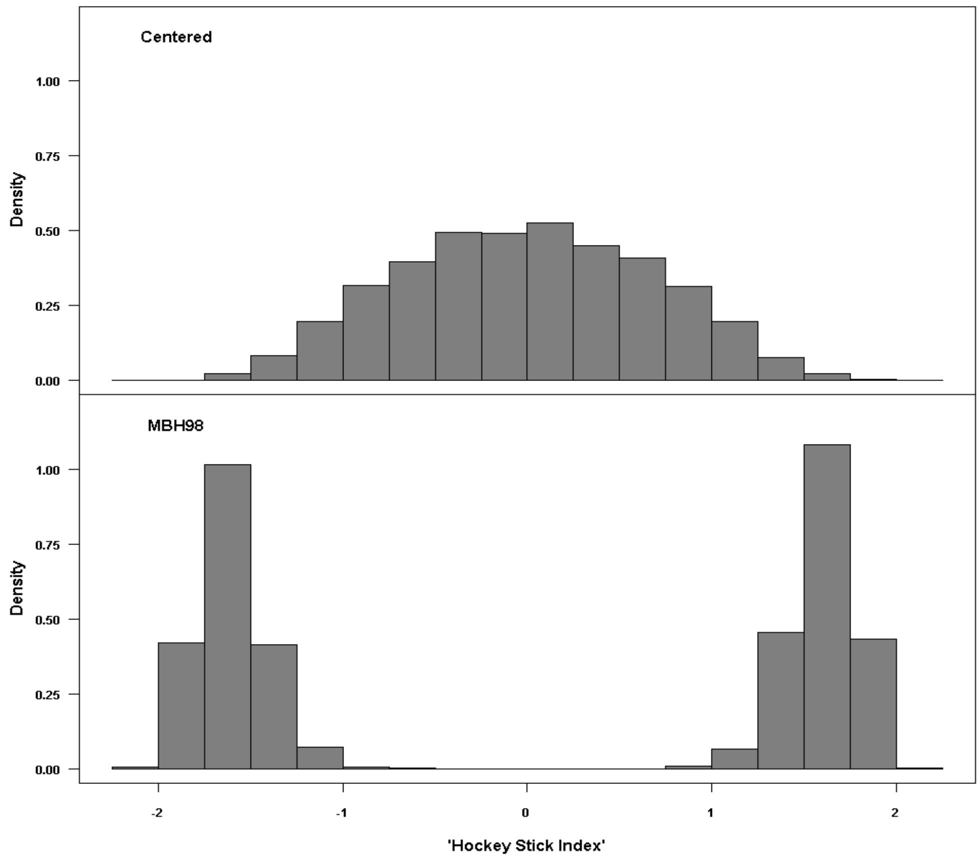

Figure 4.2: This is our recomputation of the Figure 2 in McIntyre and McKitrick (2005b). The lower panel is what MM05b call the 'Hockey Stick Index' for PC1s. For 10,000 simulated PC1s, the histogram shows the distribution of the difference between the 1902-1980 mean and the 1400-1980 mean divided by the 1400-1980 standard deviation using the MBH98 data transformation. The top histogram is based on the centered PCA computation.

Discussion: Figure 4.2 is our recomputation of the Figure 2 in McIntyre and McKitrick (2005b). The lower panel is what MM05b call the 'Hockey Stick Index' for PC1s. For 10,000 simulated PC1s, the histogram shows the distribution of the difference between the 1902-1980 mean and the 1400-1980 mean divided by the 1400-1980 standard deviation using the MBH98 data transformation. The top histogram is based on the centered PCA computation. Although our result is not identical to Figure 2 in MM05b, it reproduces the essential features of MM05b. In particular, the MBH98 methodology (and follow-on studies that use the MBH98 methodology) show a marked preference for 'hockey stick" shapes. The negative values between –2 and –1 indicate the 1902-1980 mean is lower hence the blade of the hockey stick is turned down, while the positive values between 1 and 2 in the bottom panel indicate the hockey stick blade is turned up.

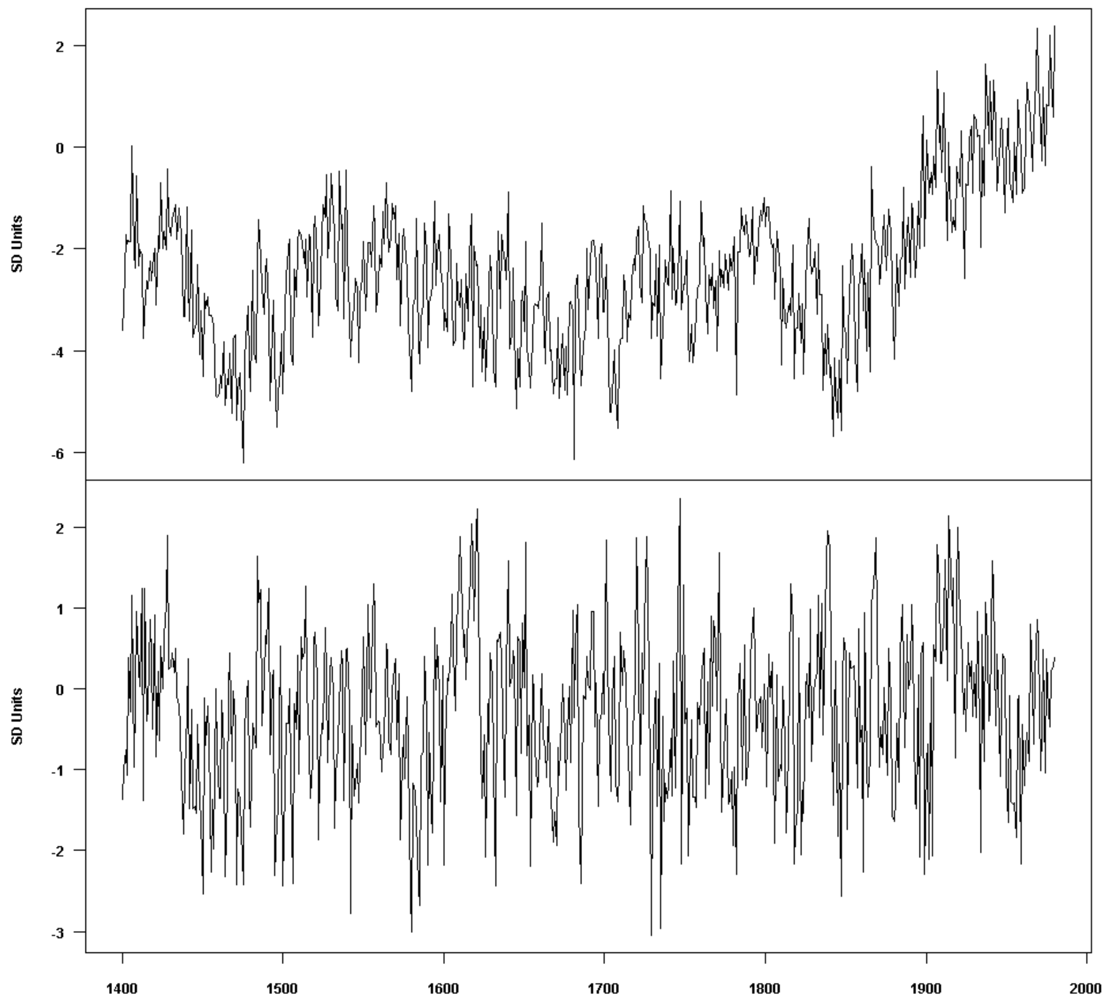

Figure 4.3: This is a recomputation of Figure 3 from MM05b. The North American Tree Network PC1 is a proxy featured prominently in MBH98. It is a PCA reconstruction of a series of tree ring proxies using the MBH98 methodology. The upper panel is the PCA reconstruction using the MBH98 data transformation. The lower panel is the reconstruction using the centered PCA methodology.

Discussion: In addition to the hockey stick shape of the upper panel it is worth noting that the lower panel exhibits considerably more variability. As mentioned in earlier discussions, PCA seeks to identify the largest contributor to the variance. It is not inherently a smoothing mechanism. The MBH98 offset of the mean value creates an 'artificially large deviation' from the desired mean value of zero.

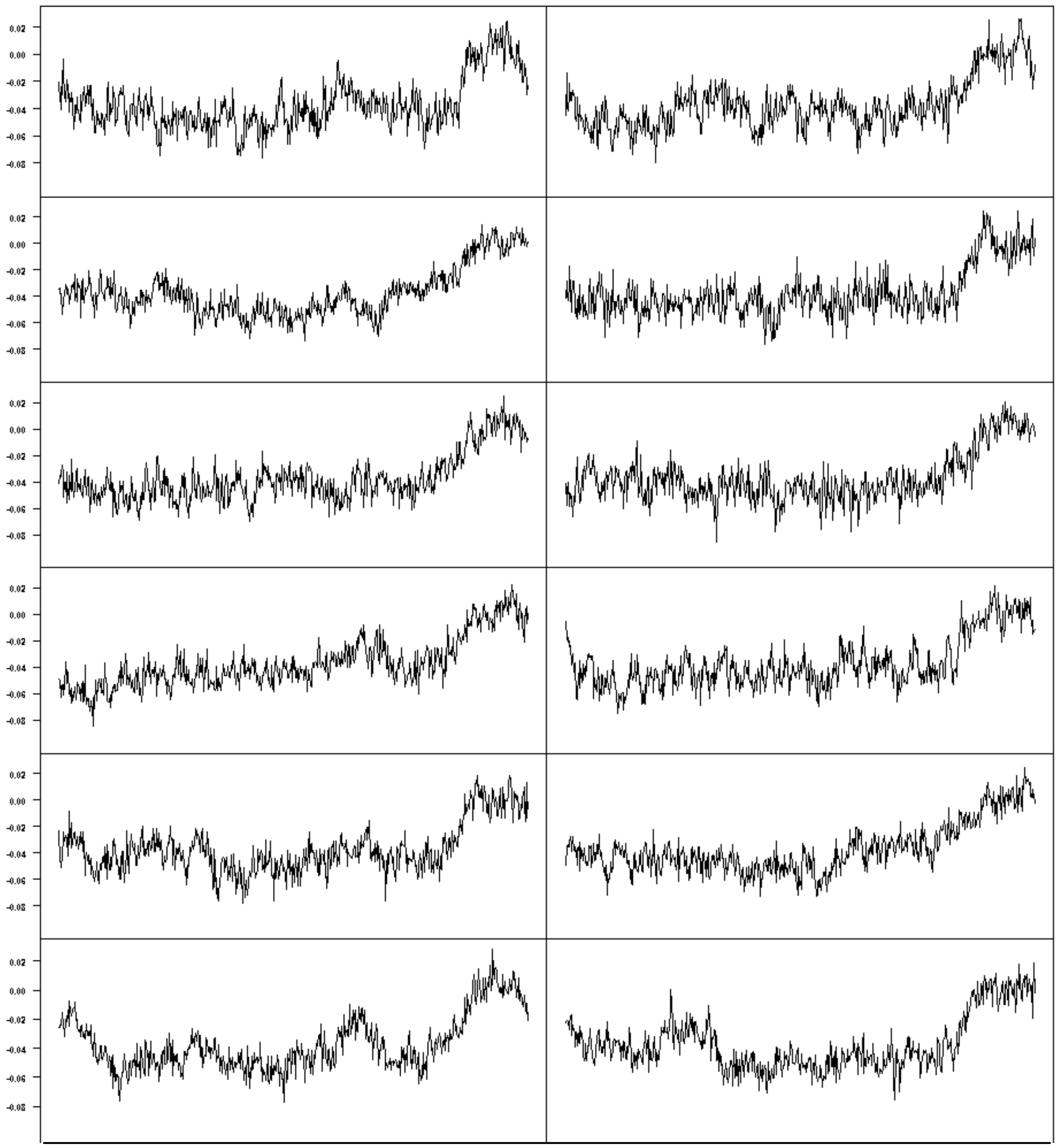

Figure 4.4: One of the most compelling illustrations that McIntyre and McKitrick have produced is created by feeding red noise [AR(1) with parameter = 0.2] into the MBH algorithm. The AR(1) process is a stationary process meaning that it should not exhibit any long-term trend. The MBH98 algorithm found 'hockey stick' trend in each of the independent replications.

Discussion: Because the red noise time series have a correlation of 0.2, some of these time series will turn upwards [or downwards] during the 'calibration' period[6] and the MBH98 methodology will selectively emphasize these upturning [or downturning] time series.

[6] 1902-1980

Figure 4.5: Here we have digitized the temperature profile as presented in the IPCC Assessment Report 1990. The early period between 1100 to about 1400 of above average temperatures is known as the Medieval Warm Period and the period from about 1500 to 1900 is known as the Little Ice Age.

Discussion: In Figure 4.5, we have digitized the temperature profile as presented in the IPCC Assessment Report 1990. The early period between 1100 to about 1400 of above average temperatures is known as the Medieval Warm Period and the period from about 1500 to 1900 is known as the Little Ice Age. The 1990 report was not predicated on a global warming scenario. It is clear that at least in 1990, the Medieval Warm Period was thought to have temperatures considerably warmer than the present era.

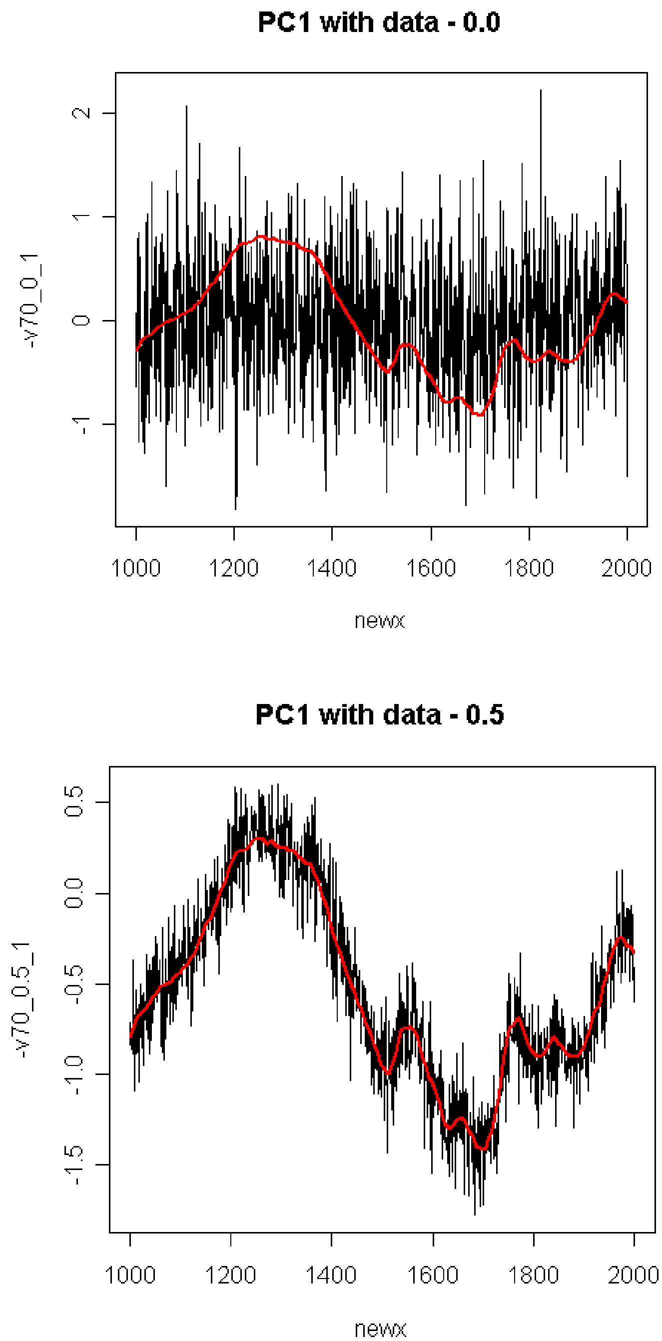

Figure 4.6: We created a pseudo-proxy scenario by creating 69 white noise process simulations, i.e. 69 white noise proxies and adding the profile in Figure 4.5 as the 70th proxy. All 70 pseudo-proxies were standardized to have a variance of one. In the top panel we applied the properly centered version of PCA. The black curve is the PC1. Offsetting the IPCC 1990 profile from zero mean by .5, and using it as the 70th proxy, we applied the MBH98 (CFR) algorithm to the proxies to obtain a 'reconstruction' of the Figure 4.5 profile in the bottom panel.

Discussion: Although there has been much discussion about the 'hockey stick' shape, the point of Figure 4.6 is that the offset procedure used in MBH98 (CFR) algorithm will reproduce any desired shape depending on what shape exists in the proxies. Recall in Figure 4.6 only one proxy of 70 has the shape, but the offset version of the PCA clearly picks out the shape.

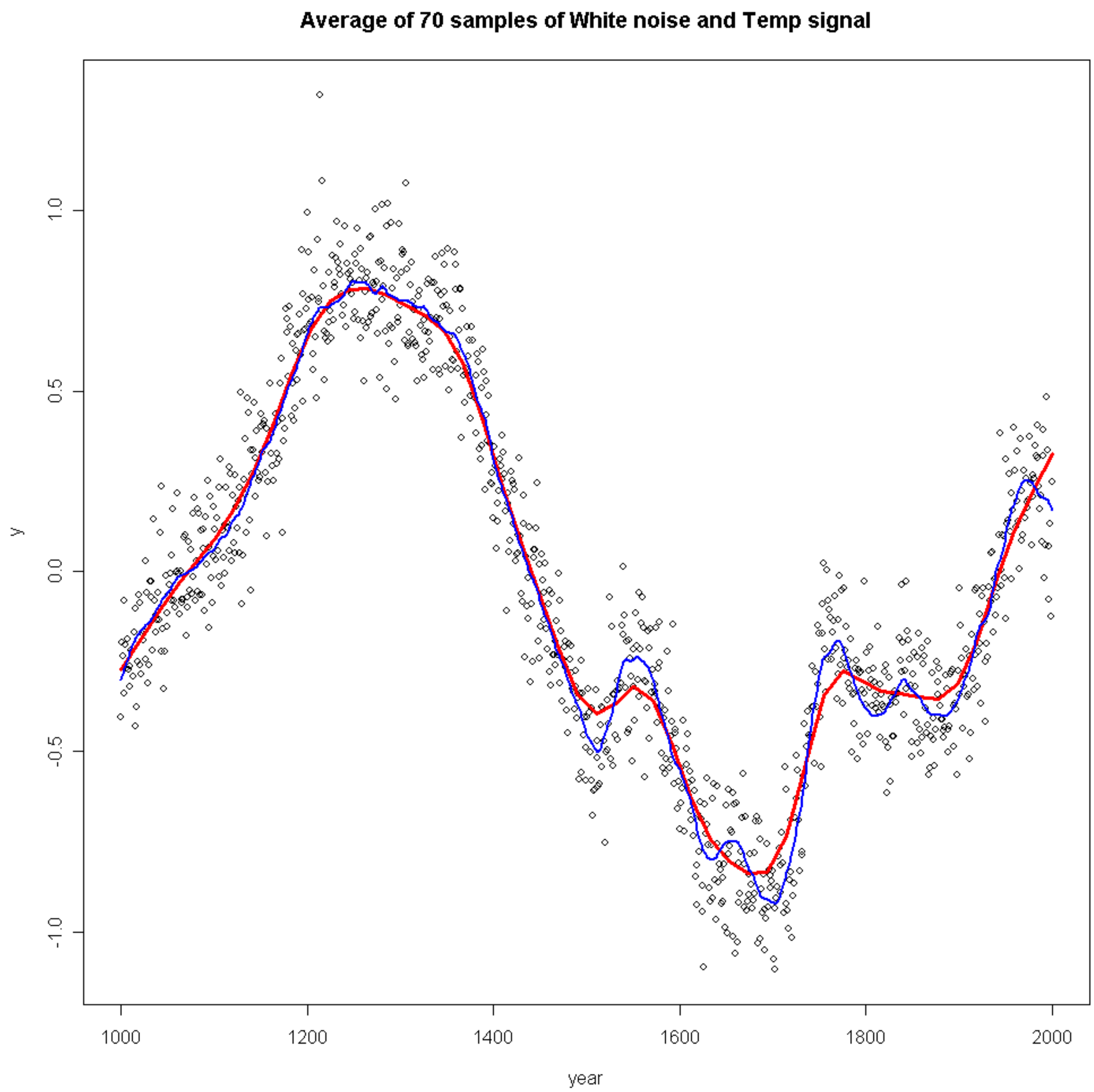

Figure 4.7: In this illustration, we created a different pseudo-proxy scenario by taking 70 copies of the Figure 4.5 profile and adding independent white noise processes to it. We then applied the CPS methodology to these pseudo-proxies to obtain another 'reconstruction' of the Figure 4.5 profile.

Discussion: In Figure 4.7, the blue curve is the original profile, the black dots are the result of applying the CPS methodology, i.e. a simple average, and the red curve is the result of applying a smoothing procedure known in the statistical literature as LOESS.

The point being made with Figures 4.6 and 4.7 is that if there are hockey sticks in the underlying data and if they are decentered, then the CFR methodology will selectively emphasize them. Similarly, if there are ‘hockey sticks’ in the data series and the remainder of the data are uncorrelated noise, then the CPS method will also emphasize the ‘hockey stick’ shape. However, if the data contain other shapes and these methods are applied to data containing these other shapes, then these methods will selectively pick out those shapes. In Figure 4.6, by decentering the 1990 profile, we inflate its effective variance so that PCA will preferentially pick it as the first principal component. In Figure 4.7, the independent white noise will be incoherent[7] and thus tend to cancel out while the ‘signal’ is the same in every proxy and will thus tend to be additive. The point here is that if each (or even a large percentage) of the proxies is selected with the hockey stick shape, then the incoherent noise will cancel and the coherent ‘hockey stick’ shape will emerge. Thus even discussions of ‘independent replications’ of the hockey stick results by different methods may not be what they superficially appear to be.

Remark: Technically speaking, the MBH98 algorithm is not principal components. Principal components are obtained theoretically through an eigenanalysis of the covariance matrix (which uses the centered data). Now there is an equivalent and numerically preferred method of obtaining the principal components by finding the so-called singular-value decomposition (SVD) of the original data matrix. The PC1 is the first of the right singular vectors in the SVD. However, this is only the case if the data matrix columns have been centered. Since the MBH98 algorithm does not center the data matrix, the SVD is actually returning a different vector than PC1. One may investigate this vector but it is incorrect to identify it as the principal component. Appendix A provides the mathematical details of this remark.

[7] Incoherent is a technical term meaning that the time series have no common form and that they are statistically independent. If each time series contains the same or similar form (shape), then this is referred to as a coherent part of the time series.

5. SOCIAL NETWORK ANALYSIS OF AUTHORSHIPS IN TEMPERATURE RECONSTRUCTIONS

One of the interesting questions associated with the 'hockey stick controversy' are the relationships among the authors and consequently how confident one can be in the peer review process. In particular, if there is a tight relationship among the authors and there are not a large number of individuals engaged in a particular topic area, then one may suspect that the peer review process does not fully vet papers before they are published. Indeed, a common practice among associate editors for scholarly journals is to look in the list of references for a submitted paper to see who else is writing in a given area and thus who might legitimately be called on to provide knowledgeable peer review. Of course, if a given discipline area is small and the authors in the area are tightly coupled, then this process is likely to turn up very sympathetic referees. These referees may have co-authored other papers with a given author. They may believe they know that author's other writings well enough that errors can continue to propagate and indeed be reinforced.

In order to answer such questions about the relationships among authors in the area of temperature reconstructions, we developed two datasets. The first specifically focusing on Dr. Mann was developed by first considering all of his co-authors and then examining the abstracts produced by the co-authors. We focus on Dr. Mann because he is the lead author of MBH98/99 and because he is extremely influential in this area as can be seen by his high degree of centrality. Drs. Bradley and Hughes also appear in the social network, but do not exhibit the centrality that Dr. Mann exhibits. We used the Engineering Compendex database, which is available on the web, to develop the abstract database of his coauthors. Based on the expanded database we examined the co-authors of his co-authors. This first database is Dr. Mann centric with the idea of understanding the relationships among his reasonably close associates. This first database consisted of 43 individuals all of whom have close ties to Dr. Mann. The second database was developed by looking for abstracts in the Engineering Compendex that treated aspects of temperature reconstruction. This second more expanded database contained more authors. In our analysis, we considered only the top 50 and top 75 in terms of numbers of papers published. There were more authors who wrote only one paper in the area and are thus not consistent researchers in the area. We report here the analysis with the top 75 authors, i.e. the 75 most frequently published authors.

Figures 5.1 through 5.4 deal with the first dataset of the closest associates of Dr. Mann. Figures 5.5 through 5.7 deal with the 75 most frequently published authors.

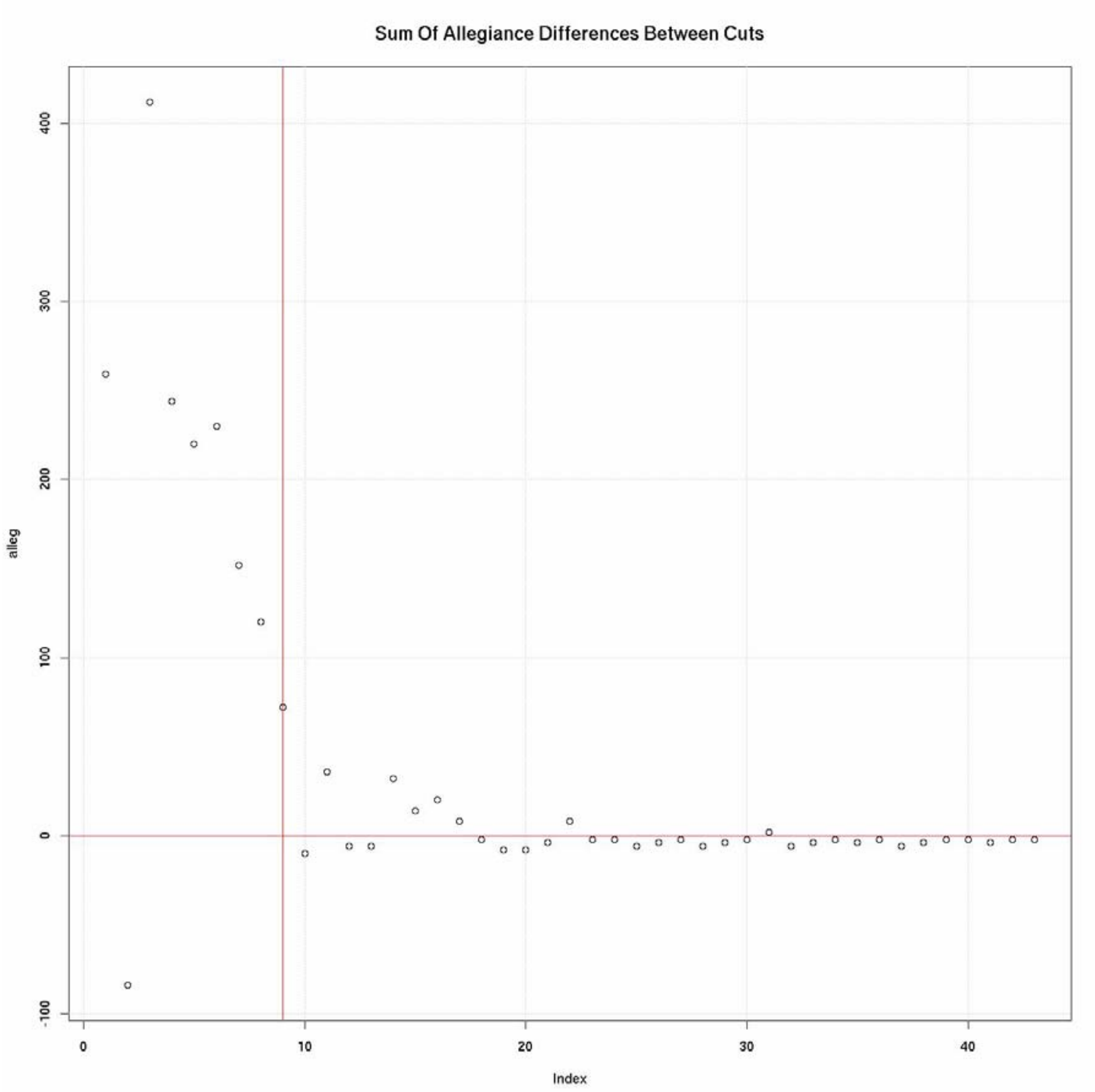

Figure 5.1: The allegiance plot for the Mann co-authors

Discussion: As mentioned in Section 2.3, allegiance is a concept/method useful for determining the clusters and cluster sizes for a social network. In some sense this determines second-order substructure within a social network. Recall that the allegiance for a block or cluster depends on having a large number of internal connections within the block and relatively few connections external to the block. The allegiance criteria are applied recursively until the allegiance criteria drop essentially to zero. This happens after nine blocks or clusters. Michael Mann is treated as a separate block because he has connections with every one of the other 42 researchers. Thus the second partition has an allegiance that temporarily drops below zero.

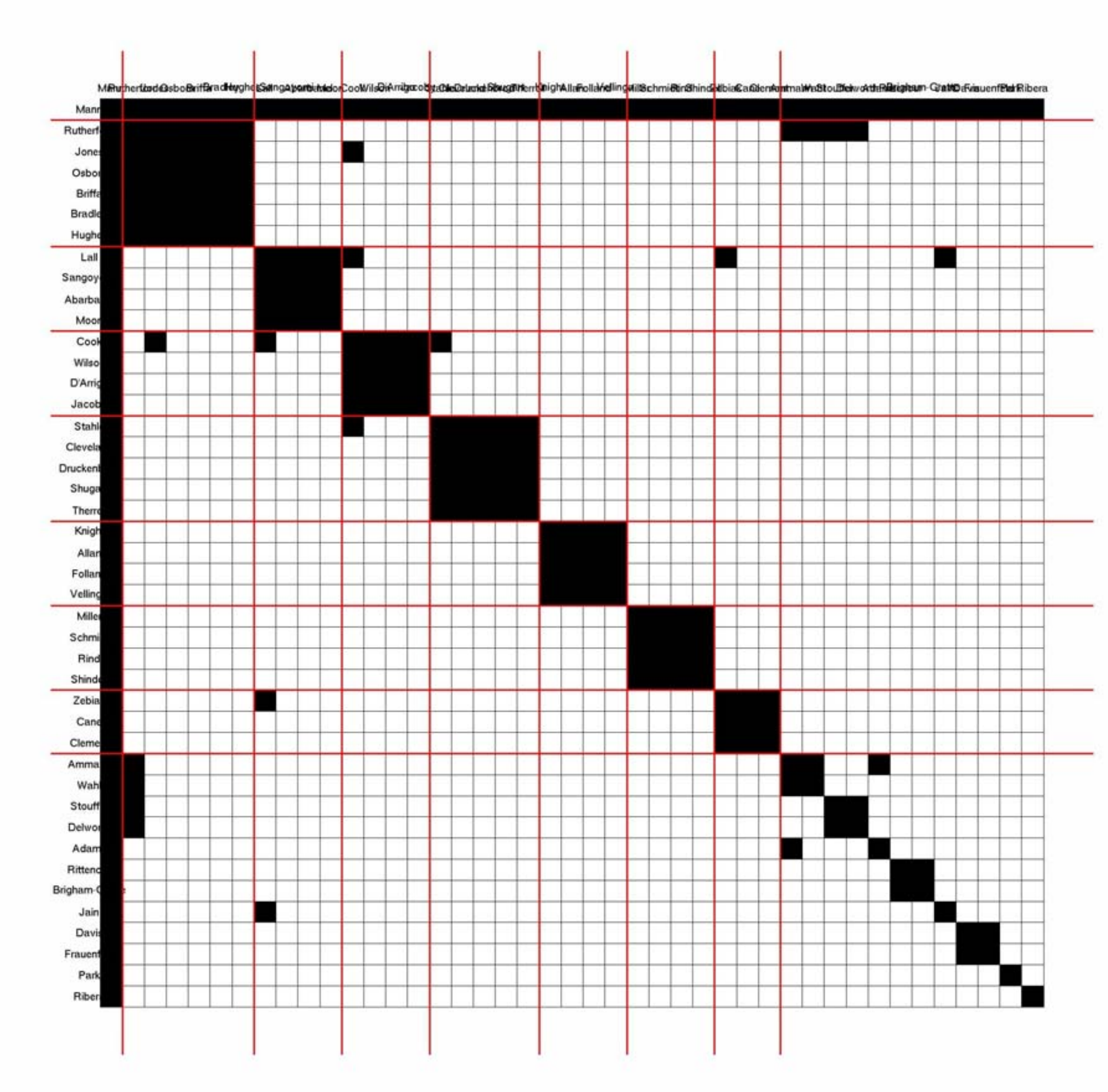

Figure 5.2: This is a matrix indicating the nine blocks of the 43 investigators. The black squares indicate that there is a co-author relationship.

Discussion: The block (cluster) structure is very clear. Michael Mann is a co-author with every one of the other 42. The black squares on the diagonal indicate that the investigators work closely within their group, but not so extensively outside of their group. The occasional off diagonal boxes indicate that some investigators have joint papers with investigators outside of their immediate group. The order of the authors on the vertical and horizontal axes is the same. Unfortunately, there is overprinting on the horizontal so that individual authors are not readable. However, it is immediately clear that the Mann, Rutherford, Jones, Osborn, Briffa, Bradley and Hughes form a clique, each interacting with all of the others. A clique is a fully connected subgraph, meaning everyone in the clique interacts with every one else in the clique.

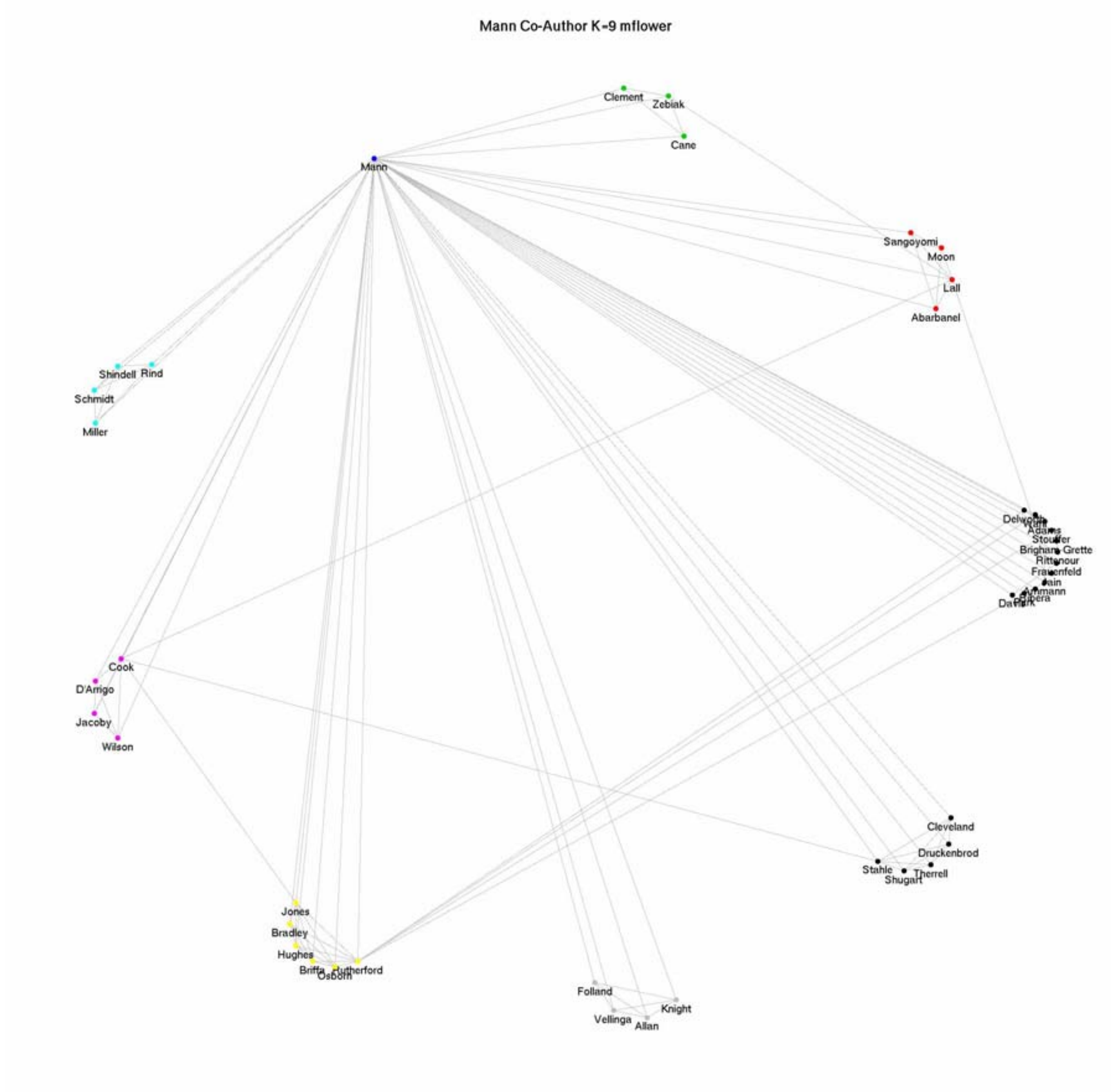

Figure 5.3: The classic social network view of the Mann co-authors. Each block or subcluster is represented along an arc.

Discussion: As mentioned before, Michael Mann is his own group since he is a co-author with each of the other 42. The cliques are very clear in this layout. In addition to the Mann-Rutherford-Jones-Osborn-Briffa-Bradley-Hughes clique there are several others that are readily apparent. They are Rind-Shindell-Schmidt-Miller, Cook-D'Arrigo-Jacoby-Wilson, Folland-Vellinga-Allan-Knight, Stahle-Shugart-Therrell-Druckenbrod-Cleveland, Sangoyomi-Moon-Lall-Abarbanel, and Clement-Zebiak-Cane. The last cluster is somewhat of the miscellaneous cluster of people who had published with Michael Mann, but not much if at all with each other.

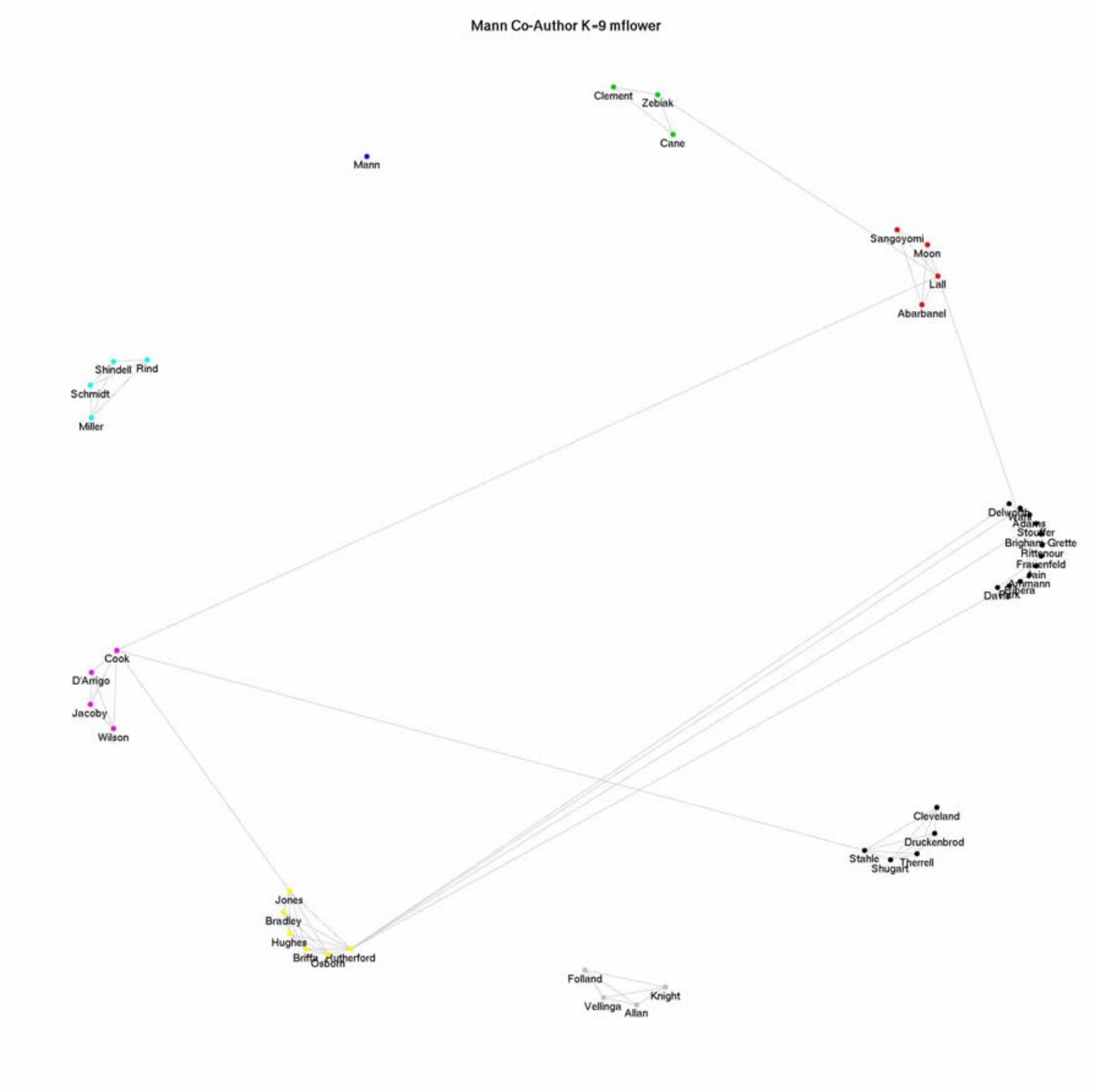

Figure 5.4: The Figure 5.3 with the linkages to Dr. Michael Mann removed.

Discussion: The dominance of Dr. Mann in terms of centrality is very clear comparing Figure 5.3 with Figure 5.4. Other authors of the remaining 42 that have some degree of centrality are Cook, Rutherford, and Lall.

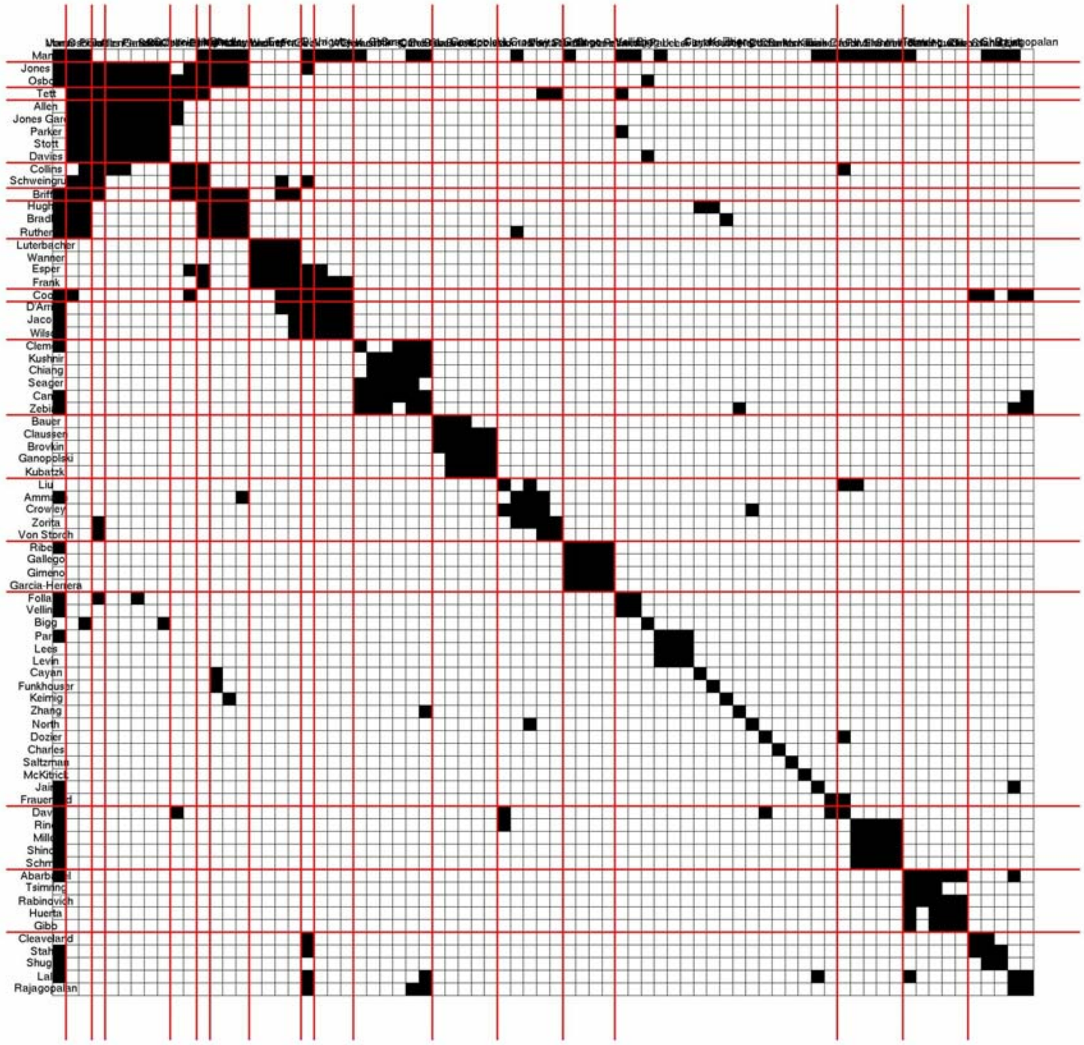

Figure 5.5: Listed here are the 75 most frequently published authors in the area of climate reconstruction. Here there are 18 blocks.

Discussion: The structure among these most frequently published authors is somewhat less than the earlier group of 43. However, the block diagonal structure still remains strong so that we do find clusters interacting with each other. The Mann-Briffa-Hughes-Bradley-Rutherford clique still is readily apparent from just the allegiance computation perspective without any forced joining.

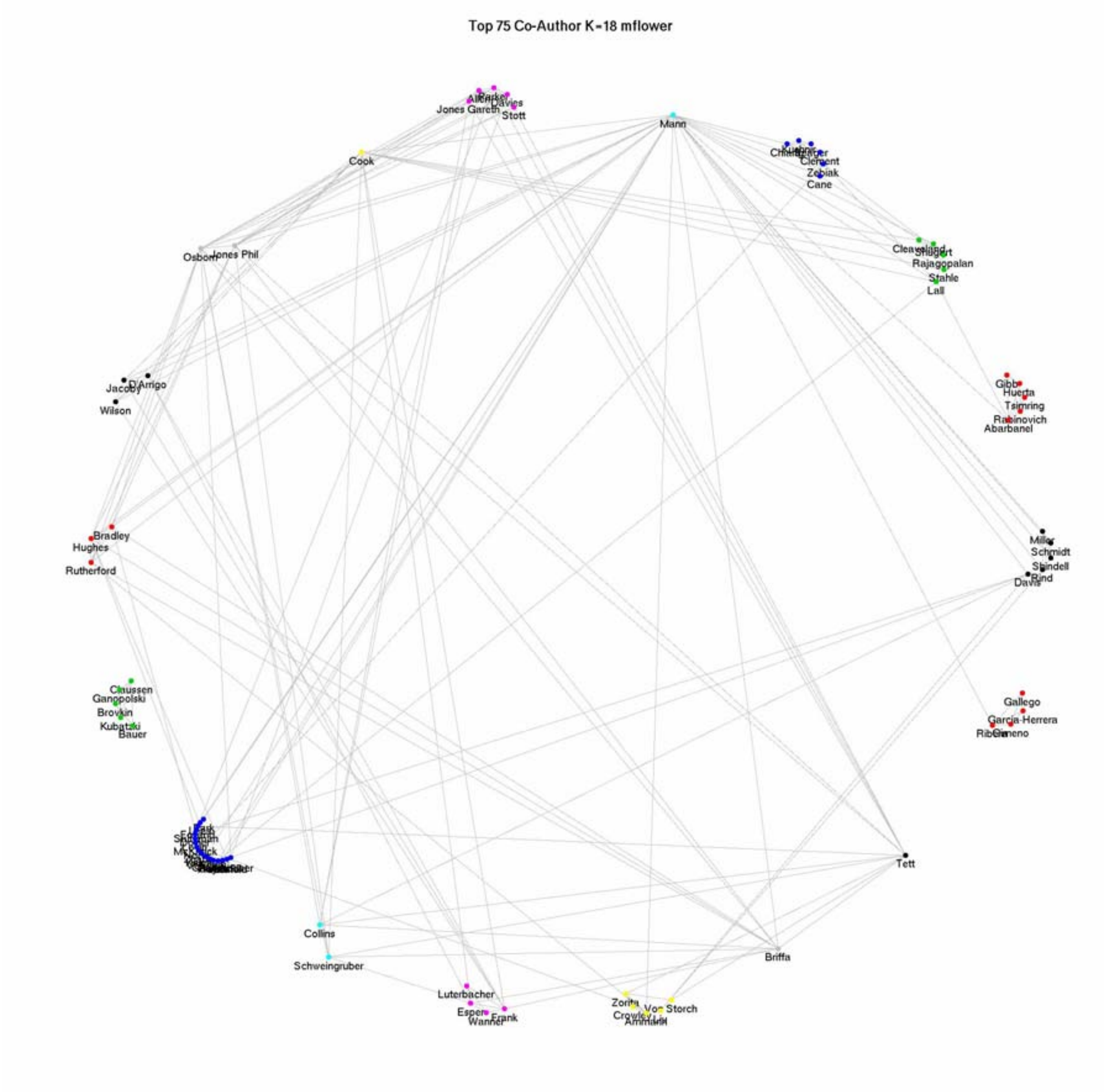

Figure 5.6: The social network of the 75 most frequently published authors in the area of climate reconstruction.

Discussion: There are some interesting features. Although Michael Mann remains an author with high centrality, Tett, Briffa and Cook emerge as belonging to their own cluster and they also exhibit high centrality. Schweingruber and Collins also appear to have relatively high centrality. One interesting observation is that although Tett is fairly central, he has no direct linkage to Mann. Similarly the Gareth Jones-Allen-Parker-Davies-Stott clique also has no direct linkage to Mann. There are two Joneses Gareth Jones is not the same person as the person previously labeled as Jones.

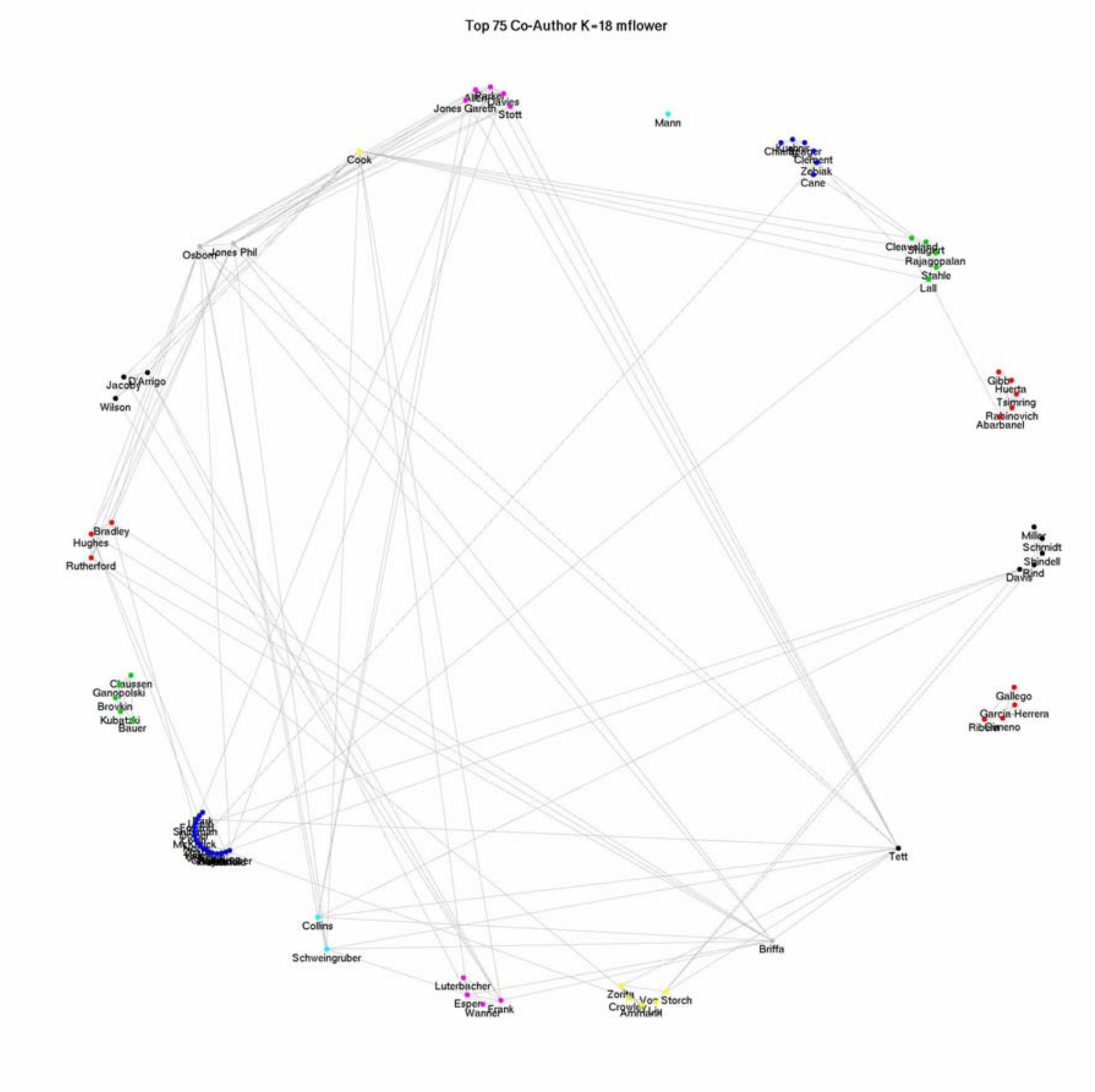

Figure 5.7: Figure 5.6 with linkages to Michael Mann removed.

Discussion: Several other interesting details emerge. The clique Claussen-Ganopolski-Brovkin-Kubatzki-Bauer is completely isolated from the other researchers in the area. Similarly, the Ribe-Gimeno-Garcia-Herrera-Gallego clique and the Arbarbanel-Rabinovich-Tsimning-Huerta-Gibb clique are nearly isolated with only linkages to Mann in the first case and linkages to Mann and Lall in the second case.

Figure 5.8: Relationships of major temperature reconstruction papers and the proxies that they used.

Discussion: The social network analysis of authors' relations suggests that the "independent reconstructions" are not as independent as one might guess. Indeed, the matrix outlined in Figure 5.8 illustrates the proxies that are used more than one time in twelve major temperature reconstruction papers. The black boxes indicate that the proxy was used in a given paper. It is clear that many of the proxies are re-used in most of the papers. It is not surprising that the papers would obtain similar results and so cannot really claim to be independent verifications. As a graphical comparison of a number of the reconstructions, see Figure 5.9 below taken from D'Arrigo et al. (2006).

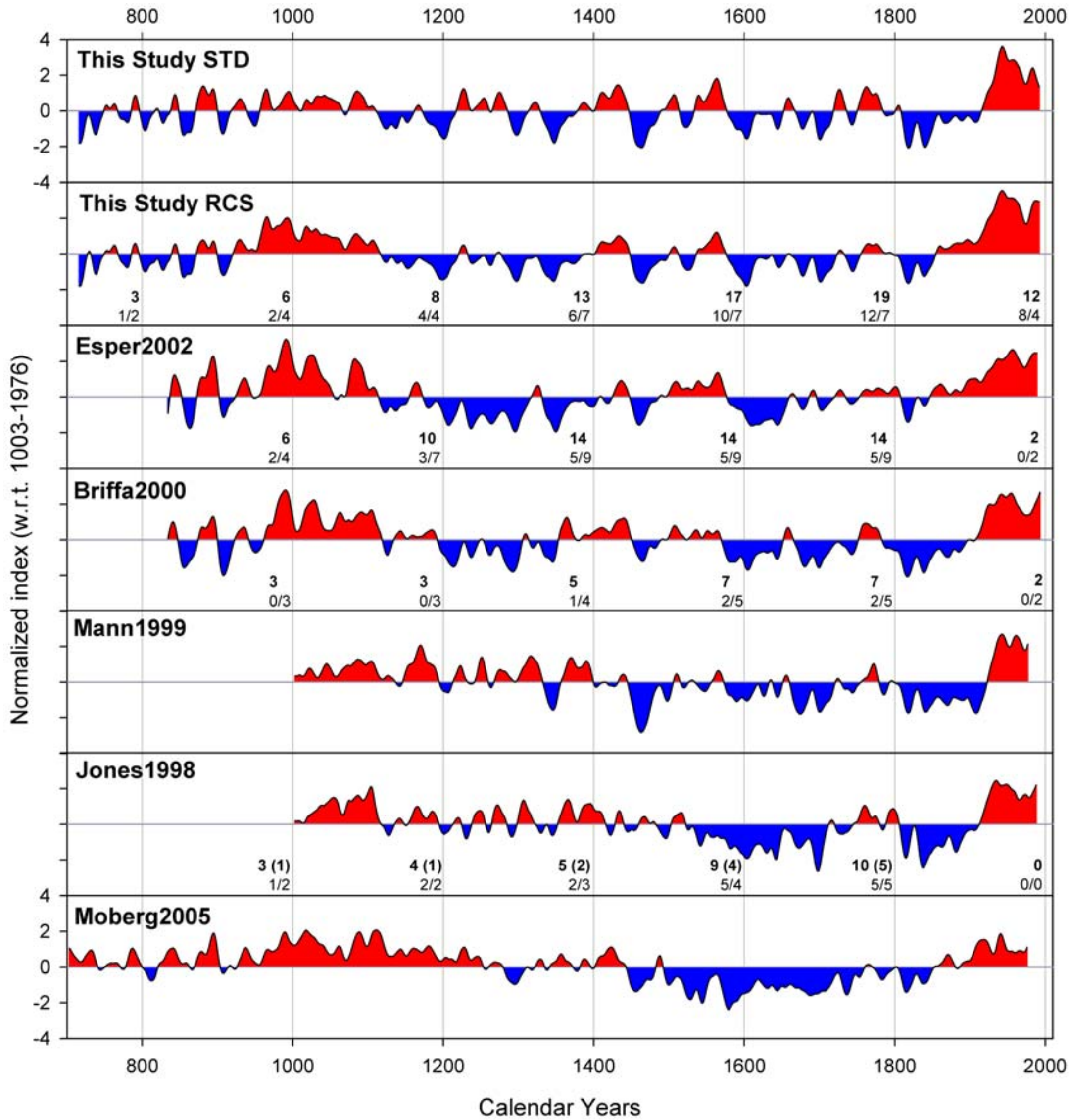

Figure 5.9. A comparison of several different reconstructions. From D'Arrigo et al. (2006)

Discussion: There are variations in the temperature reconstruction indicating the fundamental uncertainty in the reconstruction process. Essentially all agree that there was a medieval warm period centered on AD 1000 and a little ice age from at least 1600 to 1850. There is consensus in these reconstructions that the global average temperature has risen over the last 400 years. However, what must be added is that temperatures were below average in AD 1600. Both Esper et al. (2002) and Moberg et al. (2005) indicate that current global temperatures are not warmer that the medieval warm period.

6. FINDINGS

Some of the issues have been addressed in the text of the description of the methodologies in the earlier part of our discussion. For completeness and clarity we will summarize again here.

1. In general we found the writing of MBH98 somewhat obscure and incomplete. The fact that MBH98 issued a further clarification in the form of a corrigendum published in *Nature* (Mann et al. 2004) suggests that these authors made errors and incomplete disclosures in the original version of the paper. This also suggests that the refereeing process was not as thorough as it could have been.

2. In general, we find the criticisms by MM03, MM05a and MM05b to be valid and their arguments to be compelling. We were able to reproduce their results and offer both theoretical explanations (Appendix A) and simulations to verify that their observations were correct. We comment that they were attempting to draw attention to the deficiencies of the MBH98-type methodologies and were not trying to do paleoclimatic temperature reconstructions[8].

3. As we mentioned earlier, it is common for data analysis and pattern recognition experts to use a training sample for calibration. Normally one would seek to have the training data to be representative of what one might expect from the entire dataset. Because the temperature profile in the 1902-1995 is not similar, because of increasing trend, to the millennium temperature profile, it is not fully appropriate for the calibration and, in fact, leads to the misuse of the principal components analysis. However, the narrative in MBH98 on the surface sounds entirely reasonable on this calibration point, and could easily be missed by someone who is not extensively trained in statistical methodology. Dr. Mann has close ties to both Yale University and Pennsylvania State University. We note in passing that both Yale University and Pennsylvania State University have Departments of Statistics with excellent reputations[9]. Even though their work has a very significant statistical component, based on their literature citations, there is no evidence that Dr. Mann or any of the other authors in paleoclimatology studies have significant interactions with mainstream statisticians.

4. In response to the letter from Chairman Barton and Chairman Whitfield, Dr. Mann did release several websites with extensive materials, including data and code. The material is not organized or documented in such a way that makes it practical for an outsider to replicate the MBH98/99 results. For example, the directory and file structure Dr. Mann used are embedded in the code. It would

---

[8] MM05a was critiqued by Wahl and Ammann (2006) and the Wahl et al. (2006) based on the lack of statistical skill of their paleoclimate temperature reconstruction. Thus these critiques of the MM05a and MM05b work are not to the point.

[9] The Penn State and Yale Departments were ranked 19 and 20 respectively in the National Research Council publication, *Research-Doctorate Programs in the United States*, NRC (1995).

take extensive restructuring of the code to make it compatible with a local machine. Moreover, the cryptic nature of some of the MBH98/99 narratives means that outsiders would have to make guesses at the precise nature of the procedures being used.

5. As mentioned in our introduction, much of the discussion on the 'hockey stick' issue has taken place on competing web blogs. Our committee believes that web blogs are not an appropriate way to conduct science and thus the blogs give credence to the fact that these global warming issues are have migrated from the realm of rational scientific discourse. Unfortunately, the factions involved have become highly and passionately polarized.

6. Generally speaking, the paleoclimatology community has not recognized the validity of the MM05 papers and has tended dismiss their results as being developed by biased amateurs. The paleoclimatology community seems to be tightly coupled as indicated by our social network analysis, has rallied around the MBH98/99 position, and has issued an extensive series of alternative assessments most of which appear to support the conclusions of MBH98/99.

7. Our committee believes that the assessments that the decade of the 1990s was the hottest decade in a millennium and that 1998 was the hottest year in a millennium cannot be supported by the MBH98/99 analysis. As mentioned earlier in our background section, tree ring proxies are typically calibrated to remove low frequency variations. The cycle of Medieval Warm Period and Little Ice Age that was widely recognized in 1990 has disappeared from the MBH98/99 analyses, thus making possible the hottest decade/hottest year claim. However, the methodology of MBH98/99 suppresses this low frequency information. The paucity of data in the more remote past makes the hottest-in-a-millennium claims essentially unverifiable.

8. Although we have not addressed the Bristlecone Pines issue extensively in this report except as one element of the proxy data, there is one point worth mentioning. Graybill and Idso (1993) specifically sought to show that Bristlecone Pines were CO_2 fertilized. Bondi et al. (1999) suggest [Bristlecones] "are not a reliable temperature proxy for the last 150 years as it shows an increasing trend in about 1850 that has been attributed to atmospheric CO_2 fertilization." It is not surprising therefore that this important proxy in MBH98/99 yields a temperature curve that is highly correlated with atmospheric CO_2. We also note that IPCC 1996 stated that "the possible confounding effects of carbon dioxide fertilization need to be taken into account when calibrating tree ring data against climate variations." In addition, as use of fossil fuels has risen, so does the release of oxides of nitrogen into the atmosphere, some of which are deposited as nitrates, that are fertilizer for biota. Thus tree ring growth would be correlated with the deposition of nitrates, which, in turn, would be correlated with carbon dioxide release. There are clearly confounding factors for using tree rings as temperature signals.

9. Based on discussion in Mann et al. (2005) and Dr. Mann's response to the letters from the Chairman Barton and Chairman Whitfield, there seems to be at least some confusion on the meaning of R^2. R^2 is usually called the *coefficient of determination* and in standard analysis of variance; it is computed as 1 – (SSE/SST). SSE is the sum of squared errors due to lack of fit (of the regression or paleoclimate reconstruction) while SST is the total sum of squares about the mean. If the fit is perfect the SSE would be zero and R^2 would be one. Conversely, if the fit of the reconstruction is no better than taking the mean value, then SSE/SST is one and the R^2 is 0. On the other hand, the Pearson product moment correlation, r, measures association rather than lack of fit. In the case of simple linear regression, $R^2 = r^2$. However, in the climate reconstruction scenario, they are not the same thing. In fact, what is called β in MBH98 is very close what we have called R^2.

10. We note here that we are statisticians/mathematicians who were asked to comment on the correctness of the methodology found in MBH98/99. In this report we have focused on answering this question and not on whether or not the global climate is changing. We have discussed paleoclimatology only to the extent that it was necessary to make our discussion of the statistical issues clear. The instrumented temperature record makes it clear that global temperatures have risen since 1850 CE. How this present era compares to previous epochs is not clear because the uncertainties in the proxies. However, it is clear that average global temperature increases are not the real focus. It is the temperature increases at the poles that matter and average global or Northern Hemisphere increases do not address the issue. We note that according to experts at NASA's JPL, the average ocean height is increasing by approximately 1 millimeter per year, half of which is due to melting of polar ice and the other half due to thermal expansion. The latter fact implies that the oceans are absorbing tremendous amounts of heat, which is much more alarming because of the coupling of ocean circulation to the atmosphere. (See Wunsch 2002, 2006).

7. CONCLUSIONS AND RECOMMENDATIONS

Conclusion 1. The politicization of academic scholarly work leads to confusing public debates. Scholarly papers published in peer reviewed journals are considered the archival record of research. There is usually no requirement to archive supplemental material such as code and data. Consequently, the supplementary material for academic work is often poorly documented and archived and is not sufficiently robust to withstand intense public debate. In the present example there was too much reliance on peer review, which seemed not to be sufficiently independent.

Recommendation 1. Especially when massive amounts of public monies and human lives are at stake, academic work should have a more intense level of scrutiny and review. It is especially the case that authors of policy-related documents like the IPCC report, *Climate Change 2001: The Scientific Basis*, should not be the same people as those that constructed the academic papers.

Conclusion 2. Sharing of research materials, data, and results is haphazard and often grudgingly done. We were especially struck by Dr. Mann's insistence that the code he developed was his intellectual property and that he could legally hold it personally without disclosing it to peers. When code and data are not shared and methodology is not fully disclosed, peers do not have the ability to replicate the work and thus independent verification is impossible.

Recommendation 2. We believe that federally funded research agencies should develop a more comprehensive and concise policy on disclosure. All of us writing this report have been federally funded. Our experience with funding agencies has been that they do not in general articulate clear guidelines to the investigators as to what must be disclosed. Federally funded work including code should be made available to other researchers upon reasonable request, especially if the intellectual property has no commercial value. Some consideration should be granted to data collectors to have exclusive use of their data for one or two years, prior to publication. But data collected under federal support should be made publicly available. (As federal agencies such as NASA do routinely.)

Conclusion 3. As statisticians, we were struck by the isolation of communities such as the paleoclimate community that rely heavily on statistical methods, yet do not seem to be interacting with the mainstream statistical community. The public policy implications of this debate are financially staggering and yet apparently no independent statistical expertise was sought or used.

Recommendation 3. With clinical trials for drugs and devices to be approved for human use by the FDA, review and consultation with statisticians is expected. Indeed, it is standard practice to include statisticians in the application-for-approval process. We judge this to be a good policy when public health and also when substantial amounts of monies are involved, for example, when there are major policy decisions to be made based on statistical assessments. In such cases, evaluation by statisticians should be standard practice. This evaluation phase should be a mandatory part of all grant applications and funded accordingly.

Conclusion 4. While the paleoclimate reconstruction has gathered much publicity because it reinforces a policy agenda, it does not provide insight and understanding of the physical mechanisms of climate change except to the extent that tree ring, ice cores and such give physical evidence such as the prevalence of green-house gases. What is needed is deeper understanding of the physical mechanisms of climate change.

Recommendation 4. Emphasis should be placed on the Federal funding of research related to fundamental understanding of the mechanisms of climate change. Funding should focus on interdisciplinary teams and avoid narrowly focused discipline research.

BIBLIOGRAPHY

Academic Papers and Books

Anderson, Richard G. (2006) "Replicability, real-time data, and the science of economic research: FRED, ALFRED, and VDC," *Federal Reserve Bank of St. Louis Review*, 81-93.

Anderson, Richard G., Greene, William H., McCullough, Bruce D., and Vinod, H. D. (2005) "The role of data and program code archives in the future of economic research," Federal Reserve Bank of St. Louis, Working Paper 2005-014B.

Anderson, Theodore L., Charlson, Robert J., Schwartz, Stephen E., Knutti, Reto, Boucher, Olivier, Rodhe, Henning, and Heintzenberg, Jost (2003) "Climate forcing by aerosols -- a hazy picture," *Science*, 300(5622), 1103 - 1104.

Biondi, F., Perkins, D. L., Cayan, D. R. and Hughes, M. K. (1999) "July temperature during the second millennium reconstructed from Idaho tree rings," *Geophysical Research Letters*, 26(10), 1445-1448.

Bradley, R. S. (1999) *Paleoclimatology: Reconstructing Climates of the Quarternary, 2nd Edition*, San Diego: Academic Press.

Bradley, R. S. and Eddy, J. A. (1991) "Records of past global changes," in *Global Changes of the Past* (R. S. Bradley, ed.) Boulder: University Corporation for Atmospheric Research, 5-9.

Bradley, R. S. and Jones, P. D. (1993) "'Little Ice Age' summer temperature variations: Their nature and relevance to recent global warming trends," *Holocene*, 3, 367-376

Bradley, Raymond S., Hughes, M. K., and Diaz, H. F. (2003) "Climate change: Climate in medieval time," *Science*, DOI: 10.1126/science.1090372, 404-405.

Bradley, R. S., Briffa, K. R., Cole, J. E., Hughes, M. K., and Osborn, T. J. (2003) "The climate of the last millennium," in *Paleoclimate, Global Change and the Future* (Eds. K. D. Alverson, R. S. Bradley and Thomas F. Pedersen), 105-141, New York: Springer-Verlag

Briffa, K. R., Osborn, T. J., Schweingruber, F. H. (2004) "Large-scale temperature inferences from tree rings: A review," *The Global and Planetary Change*, 40, 11-26.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A. (2001) "Low-frequency temperature variations from a northern tree-ring density network," *Journal of Geophysical Research*, 106, 2929-2941.

Bürger, Gerd and Cubasch, Ulrich (2005) "Are multiproxy climate reconstructions robust?" *Geophysical Research Letters*, 32, L23711, doi:10.1029/2005GL024155.

Bürger, G., Fast, I. and Cubasch, U. (2006) "Climate reconstruction by regression – 32 variations on a theme," *Tellus*, 58A 227-235.

Center for Science and Public Policy (2005) "Climate change and the insurance industry: A critical look at the Cere report - Availability and affordability of insurance under climate change, a growing challenge for the U.S." The Center for Science and Public Policy, October 28, 2005.

Cohn, Timothy A. and Lins, Harry F. (2005) "Nature's style: naturally trendy," *Geophysical Research Letters*, 32, L23402, doi: 10.1029/2005GL024476, 2005

Cronin, T. M. (1999) *Principles of Paleoclimatology*, New York: Columbia University Press

Crowley, Thomas J., (2000) "Causes of climate change over the past 1000 years," S*cience*, 289, 270-277.

Crowley, Thomas J. (2002) "Cycles, cycles everywhere," *Science*, 295, 1473-1474.

Crowley, T. J. and North, G. R. (1991) *Paleoclimatology*, New York: Oxford University Press.

Crowley, T. J. and Lowery, T. S. (2000) ""How warm was the Medieval Warm Period? A comment on 'Man-made versus natural climate change," *Ambio*, 29, 51-54.

Crowley, T. J., Baum, S. K., Kim, K.-Y., Hegerl, G. C., and Hyde, W. T. (2003) "Modeling ocean heat content changes during the last millennium," *Geophysical Research Letters*, 30, 1932.

D'Arrigo, R., Wilson, R. ,and Jacoby, G. (2006) "On the long-term context for late twentieth century warming," *J. Geophys. Res*. 111, D03103, doi:10.1029/2005JD006352.

Esper, J., Cook, E. R., Schweingruber, F. H. (2002) "Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability," *Science*, 295, 2250-2253.

Esper, Jan, Wilson, Robert J. S., Frank, David C., Moberg, Anders, Wanner, Heinz, and Luterbacher, Jürg (2005) "Climate: past ranges and future changes," *Quaternary Science Reviews*, 24, 2164-2166.

Evans, D. L., Freeland, H. J., Hays, J. D., Imbrie, J. ans Shackleton, N. J. (1976) "Variations in the Earth's Orbit: Pacemaker of the ice ages?" *Science*, 198, 528-530.

Evans, M. N., Kaplan, A., Cane, M. A. (2002) “Pacific sea surface temperature field reconstruction from coral $\delta^{18}O$ data using reduced space objective analysis,” *Paleoceanography*, 17, 1007.

Graybill, D.A., and Idso, S.B. (1993) “Detecting the aerial fertilization effect of atmospheric CO2 enrichment in tree-ring chronologies,” *Global Biogeochemical Cycles*, 7, 81–95.

Huang, Shaopeng (2005) “Climate blog could score with newer hockey stick,” *Nature* 800 (24 February 2005); doi: 10.1038/433800b.

Huang, Shaopeng and Pollack, Henry N. (1997) “Late Quaternary temperature changes seen in world-wide continental heat flow measurements,” *Geophysical Research Letters*, 24(15), 1947-1950,

Huang, Shaopeng, Pollack, Henry N., and Shen, Po-Yu (2000) “Temperature trends over the past five centuries reconstructed from borehole temperatures,” *Nature*, 403 (17 February 2000), 403, 756-758.

Huybers, H. (2005) “Comment on ‘Hockey sticks, principal components, and spurious significance’ by McIntyre and McKitrick,” *Geophysical Research Letters.*

Jones, P. D., Briffa, K. R., Barnett, T. P. and Tett, S. F. B. (1998) “High resolution paleoclimatic records for the last millennium: Integration, interpretation, and comparison with General Circulation Model control run temperatures,” *Holocene*, 8, 455-471.

Jones, P. D. and Mann, M. E. (2004) “Climate over the past millennia,” *Reviews of Geophysics*, 42, RG2002.

Lindzen, Richard (2005) “Understanding common climate claims,” to appear *Proceedings of the 2005 Erice Meeting of the World Federation of Scientists on Global Emergencies.*

Luterbacher, J. and Coauthors (2002) “Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500,” *Climate Dynamics*, 18, 545-561.

Mann, Michael E. (1998) *A Study of Ocean-Atmosphere Interaction and Low-Frequency Variability of the Climate* System, a dissertation presented to the faculty of the graduate school of Yale University.

Mann, Michael E., Bradley, Raymond E., and Hughes, Malcolm K. (1998) “Global-scale temperature patterns and climate forcing over the past six centuries,” *Nature*, 392, 779-787.

Mann, Michael E., Bradley, Raymond S., and Hughes, Malcolm K. (1999) "Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations," *Geophysical Research Letters*, 26(6), 759-762.

Mann, Michael E., Bradley, Raymond E., and Hughes, Malcolm K. (2004) "Global-scale temperature patterns and climate forcing over the past six centuries - corrigendum," *Nature*, 430, 105.

Mann, M. E. and Emanuel, K. A. (2006) "Atlantic hurricane trend linked to climate change," *EOS*, 87(24), 233-244.

Mann, Michael E., Gille, Ed, Bradley, Raymond S., Hughes, Malcolm K., Overpeck, Jonathon, Keimig, Frank T. and Gross, Wendy (2000) "Global temperature patterns in past centuries: An interactive presentation," *Earth Interactions*, 7, Paper No. 4, 1-29.

Mann, M. E. and Jones, P. D. (2003) "Global surface temperature over the past two millennia," *Geophysical Research Letters*, 30, 1820.

Mann, Michael E., Rutherford, Scott, Wahl, Eugene, and Ammann, Caspar (2005) "Testing the fidelity of methods used in proxy-based reconstructions of past climate," *Journal of Climate*, 18, 4097-4107.

McIntyre, Stephen and McKitrick, Ross (2003) "Corrections to the Mann et al. (1998) proxy data base and Northern hemispheric average temperature series," *Energy and Environment*, 14, 751-771.

McIntyre, Stephen and McKitrick, Ross (2005a) "The M&M critique of MBH98 Northern hemisphere climate index: Update and implications," *Energy and Environment*, 16(1), 69-100.

McIntyre, Stephen and McKitrick, Ross (2005b) "Hockey sticks, principal components, and spurious significance," *Geophysical Research Letters*, 32, L03710, doi: 10.1029/2004GL021750

Moberg, A., Sonechkin, D.M., Holmgren, K., Datsenko, N.M., Karlen, W., Lauritzen, S.E. (2005) "Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data," *Nature*, 433(7026):613-617.

Osborn, Timothy J. and Briffa, Keith R. (2006) "The spatial extent of the 20th-century warmth in the context of the past 1200 years," *Science*, 311, 841-844.

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., and Jones, P. D. (2005) "Proxy-based Northern Hemisphere surface reconstructions: Sensitivity to method, predictor network, target season, and target domain," *Journal of Climate*, 18, 2308-2329.

Schweingruber, F. H., Briffa, K. R., and Nogler, (1993) “A tree-ring densiometric transect from Alaska to Labrador,” *International Journal of Biometeorology*, 37, 151-169.

Tennekes, H. (1991) “Karl Popper and the accountability of numerical weather forecasting,” in *New Developments in Predictability*, ECMWF Workshop Proceedings, ECMWF, Shinfield Park, Reading, U.K.

Valentine, Tom (1987) “Magnetics may hold key to ozone layer problems,” *Magnets*, 2(1) 18-26.

von Storch, Hans and Zorita, Eduardo (2005) “Comment on ‘Hockey sticks, principal components, and spurious significance, by S. McIntyre and R. McKitrick,” *Geophysics Research Letters*, 32, L20701, doi:10.1029/2005GL022753, 2005.

von Storch, H. E., Zorita, E., Jones, J., Dimitriev, Y., González-Rouco, F., and Tett, S. (2004), “Reconstructing past climate from noisy data,” *Science*, 306, 679–682.

von Storch, H. E., Zorita, E., Jones, J. M., Gonzalez-Rouco, F. and Tett, S. F. B. (2006) “Response to Comment on ‘Reconstructing past climate from noisy data’,” *Science*, 312, 529c.

Wahl, E. R. and Ammann, C. M. (2006) “Robustness of the Mann, Bradley, Hughes reconstruction of the Northern hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence,” *Climatic Change*, in press.

Wahl, E. R., Ritson, D. M., and Ammann, C. M. (2006) “Comment on ‘Reconstructing past climate from noisy data’,” *Science*, 312, 529b.

Wunsch, C. (2002) “Ocean observations and the climate forecast problem,” in *Meteorology at the Millennium,* (R. P. Pearce, ed.) London: Academic Press, 233

Wunsch, C. (2006) “Abrupt climate change: An alternative view,” *Quartenary Research,* 65, 191-203.

Zhang, Z., Mann, M.E., and Cook, E. R. (2004) “Alternate methods of proxy-based climate field reconstruction: Application to summer drought over conterminous United States back to AD 1700 from tree-ring data,” *Holocene*, 14, 502-516.

Other Literature Including Articles in the Popular Press

Colligan, Douglas (1973) “Brace yourself for another ice age,” *Science Digest*, 73(2), 57-61.

Crok, Marcel (2005) “Proof that mankind causes climate change is refuted: Kyoto protocol based on flawed statistics,” *Natuurwetenschap & Techniek*, February 2005.

Gwynne, Peter (1975) “The cooling world,” *Newsweek*, April 28, 1975.

Intergovernmental Panel on Climate Change (2001) *Climate Change 2001: Third Assessment Report*, IPCC (Especially Chapter 2: “Observed climate variability and change,” Folland C. K. and Karl, T. R., coordinating lead authors).

Kerr, Richard A. (2006) “Yes, it’s been getting warmer in here since the CO_2 began to rise,” *Science*, 312, 1854.

Legates, David (2005) “Where’s the data?: holding science to prospectus standards would stop climate researchers from launching misrepresentations like the ‘Hockey Stick’,” *Financial Post*, September 20, 2005.

Lindzen, Richard (2001) “Scientist’ report doesn’t support the Kyoto Treaty,” *Wall Street Journal*, June 11, 2001.

Michaels, Patrick J. and Douglass, David H. (2004) “Global-warming sciences meltdown,” *Washington Times*, Aug 16, 2004, page A17.

Muller, Richard (2004) “Global warming bombshell,” *MIT Technology Review*, http://www.technologyreview.com/BizTech/wtr_13830,296,p1.htm

NRC (1995) *Research-Doctorate Programs in the United States*, National Academy of Sciences, Washington, DC.

NOAA (2005) *SAP-1.1 Prospectus for Temperature Trends in the Lower Atmosphere: Understanding and Reconciling the Differences,* Climate Change and Assessment Program Synthesis and Assessment Product 1.1.

Sullivan, Walter (1975a) “Scientists ask why world climate is changing: Major cooling may be ahead,” *The New York Times*, p. 92, May 21, 1975.

Sullivan, Walter (1975b) “Climatic changes by aerosols in atmosphere feared,” *The New York Times*, pg. 1, September 14, 1975.

Zidek, James V. (2006) “Editorial: (post-normal) statistical science,” *Journal of the Royal Statistical Society (A)*, 169, pt. 1, 1-4.

Presentations

Mann, Michael E. (2006) “Climate Over the Past 1-2 Millennia,” presentation at the National Academy of Science, March 3, 2006

McIntyre, Stephen (2005) "Low-Frequency Ranges in Multiproxy Climate Reconstructions," AGU December 2005 Meeting

McIntyre, Stephen and McKitrick, Ross (2006) "Surface Temperature Reconstructions for the Past 1,000-2,000 Years," presentation at the National Academy of Science, March 3, 2006

McIntrye, Stephen and McKitrick, Ross (2005) "The Hockey Stick Debate: Lessons in Disclosure and Due Diligence," September 7, 2005

McKitrick, Ross (2005) "What is the 'Hockey Stick' debate about?" APEC Study Group, Australia, April 4, 2005.

von Storch, Hans and Zorita, Eduardo (2006) "Reconstruction of historical temperatures during the last millennium - analysis of methods - significance of anthropogenic climate change," presentation at the National Academy of Science, March 3, 2006

Websites

http://www.climateaudit.org (McIntyre and McKitrick website)

http://www.climate2003.org (McIntyre website)

http://www.realclimate.org (Mann website)

http://www.meteo.psu.edu/~mann/Mann/index.html

http://unfccc.int/meetings/cop_11/items/3394.php

APPENDIX

Appendix A. Mathematical Underpinnings of PCA

A.1 Eigenvectors and Eigenvalues

Consider a general square $d \times d$ matrix $\boldsymbol{M}$ and a vector $\boldsymbol{v}$. The vector $\boldsymbol{v}$ is said to be an eigenvector of the matrix $\boldsymbol{M}$ if $\boldsymbol{M}\boldsymbol{v} = s\boldsymbol{v}$ where s is a scalar. The way to think about this situation geometrically is that a matrix such as $\boldsymbol{M}$ could have three effects. It could rotate a vector, it could mirror image a vector, and it could rescale a vector. In general it cannot translate a vector. If it rotates a vector or mirror images a vector into a scaled version of itself, then the vector is an eigenvector of the matrix.

As a simple example, consider $\boldsymbol{M} = \begin{pmatrix} 8 & 10 \\ 3 & -5 \end{pmatrix}$ and let $\boldsymbol{v} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$. Then $\boldsymbol{M}\boldsymbol{v} = \begin{pmatrix} 8 & 10 \\ 3 & -5 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 50 \\ 10 \end{pmatrix} = 10\begin{pmatrix} 5 \\ 1 \end{pmatrix} = 10\boldsymbol{v}$. Thus $\begin{pmatrix} 5 \\ 1 \end{pmatrix}$ is an eigenvector of $\boldsymbol{M}$ and the scaler s is said to be the eigenvalue corresponding to the eigenvector.

Eigenvectors can only be found for square matrices and not all matrices may have eigenvectors. If a $d \times d$ symmetric matrix does have an eigenvector, it will have d of them. The d eigenvectors will be orthogonal (perpendicular) to each other. It is also clear that if we scale a matrix by a scalar amount, say c, then $(c\boldsymbol{M})\boldsymbol{v} = c(\boldsymbol{M}\boldsymbol{v}) = c(s\boldsymbol{v}) = (cs)\boldsymbol{v}$. Thus if we scale the matrix by an amount c, we scale the corresponding eigenvalue by the same amount. It is convenient to scale the largest eigenvector to length 1. In this case the rescaled eigenvectors become an orthonormal basis for the d-dimensional vector space. In the example above the length of the eigenvector is $\sqrt{5^2 + 1^2} = \sqrt{26}$. Thus $\begin{pmatrix} \frac{5}{\sqrt{26}} \\ \frac{1}{\sqrt{26}} \end{pmatrix}$ is the unit eigenvector whose length is 1.

A.2 Principal Components

The orthogonality of eigenvectors is the attractive feature that we would like to use for multivariate data analysis. We may start out with a series of data vectors, say $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d$, which may, in fact, be highly correlated. We would like to create another set of data vectors that are orthogonal. We assume that the vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d$, are column vectors of length n. We may form $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_d$, the sample means of each column vector. Then $\boldsymbol{x}_1 - \overline{x}_1, \boldsymbol{x}_2 - \overline{x}_2, \ldots, \boldsymbol{x}_d - \overline{x}_d$ are the column vectors of centered data, which can be organized into an $n \times d$ data matrix $\boldsymbol{X}$. Then $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is the $d \times d$ covariance matrix corresponding to the data matrix $\boldsymbol{X}$. Here the $\boldsymbol{X}^{\mathrm{T}}$ indicates $\boldsymbol{X}$ transpose. Because $\widehat{\boldsymbol{\Sigma}}$ is a square matrix, the eigenvectors associated with $\widehat{\boldsymbol{\Sigma}}$ are $\boldsymbol{v}_{x1}, \boldsymbol{v}_{x2}, \ldots, \boldsymbol{v}_{xd}$. Let the corresponding eigenvalues be $\alpha_{x1} \geq \alpha_{x2} \geq \cdots \geq \alpha_{xd}$ so that the eigenvalues are arranged in decreasing order of magnitude. The geometric interpretation of the first eigenvector is that it points in the direction that has the largest variance. The second eigenvector points in a direction orthogonal to the first eigenvector and in a direction that has the second largest variance, and so on for the remaining eigenvectors. The eigenvectors are each $d \times 1$ column vectors. The fact that these vectors are unit vectors (i.e. have magnitude 1) is captured by $\boldsymbol{v}_{xi}^{\mathrm{T}}\boldsymbol{v}_{xi} = 1, i = 1, 2, \ldots, d$ and that they are orthogonal is captured by $\boldsymbol{v}_{xi}^{\mathrm{T}}\boldsymbol{v}_{xj} = 0,\ i \neq j,\ i, j = 1, 2, \ldots, d.$ If $\boldsymbol{v}_{xj} = (v_{xj1}, v_{xj2}, \ldots, v_{xjd})^{\mathrm{T}}$ is the column vector for the jth eigenvector, then

$$y_{kj} = \sum_{i=1}^{d}(x_{ki} - \overline{x}_k)v_{xji}$$

is the kth entry of the jth principal component. In matrix form $\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{v}_j$ is the jth principal component. In the notation of Mann et al. (1998), $\boldsymbol{y_1}$ is PC_1 and this corresponds to the first reconstruction. Because $\boldsymbol{X}$ is a $n \times d$ matrix and $\boldsymbol{v}_j$ is a $d \times 1$ column vector, $\boldsymbol{y}_j$ is a $n \times 1$ column vector and in the case of Mann et al. (1998), it will represent the jth principal component time series. If we assemble the eigenvectors into a $d \times d$ matrix $\boldsymbol{V} = (\boldsymbol{v}_{x1}, \boldsymbol{v}_{x2}, \ldots, \boldsymbol{v}_{xd})$, then $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{V}$ would be the matrix of principal components. Principal component methodology is typically used for dimension reduction counting on the fact that for some j_0, $\alpha_j \approx 0$ for $j \geq j_0$. Thus for $j \geq j_0$ the principal components may be treated as negligible and thus ignored. In the temperature reconstruction model, the PC_1 (first principal component) is being used to reconstruct a time series that is capturing the most variability in the data.

A.3 Numerical Methods and Bias Illustration

With large datasets, a preferred numerical method for finding the eigenvalues is to use a singular value decomposition of the data matrix $\boldsymbol{X}$. The singular value decomposition (SVD) is given by

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}},$$

where $\boldsymbol{D}$ is a diagonal matrix of d singular values and the matrices $\boldsymbol{U}$ $(n \times d)$ and $\boldsymbol{V}$ $(d \times d)$ satisfy $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}_d$ and $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}_d$. In this case

$$\widehat{\boldsymbol{\Sigma}} = \tfrac{1}{n}(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}) = \tfrac{1}{n}\boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}} = \tfrac{1}{n}\boldsymbol{V}\boldsymbol{D}^2\boldsymbol{V}^{\mathrm{T}}.$$

The eigenvectors of $\widehat{\boldsymbol{\Sigma}}$ are the right singular vectors of $\boldsymbol{X}$ given by $\boldsymbol{V}$.

In Mann et al. (1998), the study period is partitioned into a reconstruction period 1400-1995 and a training period 1902-1980 in which all the proxy variables are available. The data matrix is centered using the training data rather than the overall means. Because the training period has higher temperatures, this biases the overall data lower for the period 1400-1995, thus inflating the variance. In this case the right singular vectors, $\boldsymbol{V}$, are no longer the eigenvectors.

In general,

$$\widehat{\boldsymbol{\Sigma}} = \tfrac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \ - \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^{\mathrm{T}}$$

where $\overline{\boldsymbol{x}}$ is the $p \times 1$ vector of variable means in $\boldsymbol{X}$ after partial centering. For this $\boldsymbol{X}$ we let the column vectors of the right singular matrix $\boldsymbol{V}$ be $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d$. Because these form an orthonormal basis, we can find constants, β_i, that satisfy

$$\overline{\boldsymbol{x}} = \sum_{i=1}^{d} \beta_i \boldsymbol{v}_i \ \text{ so that } \ \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^{\mathrm{T}} = \sum_{i=1}^{d} \sum_{j=1}^{d} \beta_i \beta_j \boldsymbol{v}_i \boldsymbol{v}_j^{\mathrm{T}}$$

To see how close the singular values, $\boldsymbol{v}_k$, are to the eigenvectors of $\widehat{\boldsymbol{\Sigma}}$ let us substitute

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}_k &= \left(\tfrac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \ - \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^{\mathrm{T}}\right)\boldsymbol{v}_k \\
&= \tfrac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_k - \sum_{i=1}^{d} \sum_{j=1}^{d} \beta_i \beta_j \boldsymbol{v}_i \boldsymbol{v}_j^{\mathrm{T}} \boldsymbol{v}_k \\
&= \tfrac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_k - \sum_{i=1}^{d} \beta_i \beta_k \boldsymbol{v}_i \\
&= \tfrac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}_k - \beta_k \overline{\boldsymbol{x}}.
\end{aligned}$$

Unless $\overline{\boldsymbol{x}}$ happens to be an eigenvector so that $\boldsymbol{v}_k = \overline{\boldsymbol{x}}$ for some k, the singular vectors are not the eigenvector directions. They are biased toward the unsubtracted mean vector $\overline{\boldsymbol{x}}$. Note that if one data vector (proxy variable) $\boldsymbol{x}_k$ happens to have a much larger variance than the others, then the principal components method will attempt to fit that variable causing the eigenvector to align with variable. Similarly the largest component of the mean vector will be the contribution from variable $\boldsymbol{x}_k$ so that the eigenvector and the mean vector will be such that the directions of $\boldsymbol{v}_k$ and $\overline{\boldsymbol{x}}$ tend to align. If $\boldsymbol{v}_k = \overline{\boldsymbol{x}}$, then the associate eigenvalue for $\frac{1}{n}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is biased upwards by the amount $\overline{\boldsymbol{x}}^{\mathrm{T}}\overline{\boldsymbol{x}}$. This may be substantial. Thus the first eigenvector of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ will be biased towards $\overline{\boldsymbol{x}}$ and its importance overstated by $\overline{\boldsymbol{x}}^{\mathrm{T}}\overline{\boldsymbol{x}}$.

APPENDIX B. Request from the House Committee on Science/Chairman Boehlert

In addition to the effort undertaken by the House Committee on Energy and Commerce, the House Committee on Science also sought to clarify these issues. To this end the House Committee on Science undertook to charge the National Research Council (NRC) of the National Academy of Science (NAS).

House Committee on Science Charge to NAS

The group should, in a clear and concise report issued in a relatively short period of time, answer the following questions:

1. What is the current scientific consensus on the temperature record of the last 1,000 to 2,000 years? What are the main areas of uncertainty and how significant are they?

2. What is the current scientific consensus on the conclusions reached by Drs. Mann, Bradley and Hughes? What are the principal scientific criticisms of their work and how significant are they? Has the information needed to replicate their work been available? Have other scientists been able to replicate their work?

3. How central is the debate over the paleoclimate temperature record to the overall scientific consensus on global climate change (as reflected in previous reports from the Academy)? How central is the work of Drs. Mann, Bradley, and Hughes to the consensus on the temperature record?

NAS/NRC Internal Translation of the Charge to NAS/Atmospheric Science Board

The committee will describe and assess the state of scientific efforts to reconstruct surface temperature records for the Earth over approximately the past 2,000 years. The committee will summarize current scientific information on the temperature record for the past two millennia, describe the main areas of uncertainty and how significant they are, describe the principal methodologies used and any problems with these approaches, and explain how central the debate over the paleoclimate temperature record is to the state of scientific knowledge on global climate change. As part of this effort, the committee will address tasks such as:

1. Describe the proxy records that have been used to estimate surface temperatures for the pre-instrumental period (e.g., tree rings, sediment cores, isotopes in water and ice, biological indicators, indicators from coral formations, geological boreholes, and historical accounts) and evaluate their limitations.

2. Discuss how proxy data can be used to reconstruct surface temperature over different geographical regions and time periods.

3. Assess the various methods employed to combine multiple proxy data to develop large-scale surface temperature reconstructions, the major assumptions associated with each approach, and the uncertainties associated with these methodologies.

4. Comment on the overall accuracy and precision of such reconstructions, relevant data quality and access issues, and future research challenges.

Note: statement of task revised on 03/30/2006. The National Academies is sponsoring this study. The approximate starting date for the project is 01/19/2006. The committee will issue a Final Report in approximately four months.

Note: Although the House Committee on Science initiated the Academy study, the Academy decided not to address the specific questions of the House Committee on Science and decided to focus of the Academy study away from the specific questions and address broader issues.

We attempt here to give answers to the House Committee on Science questions.

- What is the current scientific consensus on the temperature record of the last 1,000 to 2,000 years?

  Ans: There is strong evidence from the instrumented temperature record that temperatures are rising since 1850 and that global warming is a fact. How accurate the reconstructions over the past millennium are is a matter of debate and we do not believe there is a consensus on this issue.

- What are the main areas of uncertainty and how significant are they?

  Ans: The proxy data have many factors encoded in them, one of which is temperature. However, the temperature proxy is confounded with many other factors that have not been teased out including carbon dioxide fertilization effects. The high level of variability in the proxy data as well as the lack of low frequency effects make the reconstructions more problematic than the advocates of these methods would have us believe. In addition the lack of a really substantial stationary, instrumented temperature record handicaps the calibration.

- What is the current scientific consensus on the conclusions reached by Drs. Mann, Bradley and Hughes?

  Ans: Based on the literature we have reviewed, there is no overarching consensus on MBH98/99. As analyzed in our social network, there is a tightly knit group of individuals who passionately believe in their thesis. However, our perception is that this group has a self-reinforcing feedback mechanism and, moreover, the work has been sufficiently politicized that they can hardly reassess their public positions without losing credibility.

- What are the principal scientific criticisms of their work and how significant are they?

  Ans: Our perception is that principal components (statistical) analysis was used incorrectly and, based on this, unsupportable inferences were drawn about the current magnitude of global warming relative to the historical past. We hasten to repeat that the Earth is getting warmer. What does not appear to be true is that the process mechanism is as well understood as some scholars would have us believe. In addition the use of some proxies does not appear to be as carefully managed as we might like.

- Has the information needed to replicate their work been available?

  Ans: In our opinion, the answer is no. As mentioned earlier, there were gaps in MBH98.

- Have other scientists been able to replicate their work?

  Ans: von Storch, Zorita and González-Rouco in their presentation to the NAS Committee report that Bürger et al. (2006) have reproduced the MBH98 results. We have not verified this independently.

- How central is the debate over the paleoclimate temperature record to the overall scientific consensus on global climate change (as reflected in previous reports from the Academy)?

  Ans: In a real sense the paleoclimate results of MBH98/99 are essentially irrelevant to the consensus on climate change. The instrumented temperature record since 1850 clearly indicates an increase in temperature. Whether this is unprecedented in the last millennium seems less clear and to what extent the natural planetary processes can mitigate the excess green-house gas release is unknown. What is more important in our view is real insight into and understanding of the processes of global warming.

- How central is the work of Drs. Mann, Bradley, and Hughes to the consensus on the temperature record?

  Ans: MBH98/99 has been politicized by the IPCC and other public forums and has generated an unfortunate level of consensus in the public and political sectors and has been accepted to a large extent as truth. Within the scholarly community and in certain conservative sectors of the popular press, there is at least some level of skepticism.

APPENDIX C. Summaries of Important Papers Discussed In the Report

Summary of Michael E. Mann's Ph.D. Dissertation, *A Study of Ocean-Atmosphere Interaction and Low-Frequency Variability of the Climate System* (1998)

Dr. Mann attempts to clarify low frequency climate variations so that the effects of other factors (anthropogenic forcing, etc.) on climate can be defined. Dr. Mann's method includes a simplified theoretical model to provide a description of the effects of ocean circulation on climate as well as application of multivariate statistical methodology in the reconstruction of oscillatory low-frequency signals, using proxy and instrumental data. While Dr. Mann states that there is scant robust evidence for periodic climate signals other than the yearly seasonal variations, he notes there does seem to be an oscillatory character in many climatic processes. Dr. Mann attributes this "quasi-oscillatory" climate signal to linear and non-linear feedback mechanisms. Additionally, these signals are marked by defined, finite, decadal to centennial scale variations. Dr. Mann concludes, from the investigation of proxy data, that the most likely source of the climate variations is the coupled ocean-atmosphere processes.

In his analysis, Dr. Mann determines that the traditional approach of climate signal detection, univariate spectral analysis, underestimates multi-decadal variations. He opts for a multivariate method, principal component analysis combined with spectral analysis, for climate signal detection. However, this method also presents with problems, as the distinct principal components present different climate signals and varying decomposition of the associated noise. Dr. Mann attributes this to two consecutive statistical operations with confounding optimization properties. Dr. Mann goes on to examine several different methods of principal component analysis that mitigate these negative effects, but eventually settles on multitaper frequency-domain singular value decomposition, or "MTM-SVD."

MTM-SVD isolates statistically significant oscillations that are correlated with the independent time series. This method is useful in describing spatially-correlated oscillatory signals that have arbitrary spatial relationships in amplitude and phase. Subsequently, this approach can detect standing and traveling patterns in a spatial-temporal dataset as well as account for periodic or aperiodic oscillatory patterns. Dr. Mann contends that this method allows for an accurate climate reconstruction of spatiotemporal patterns immersed in noise.

Using these methods, Dr. Mann found a long-term global warming trend and anomalous atmospheric circulation patterns. These patterns show similarity to a modeled response of climate to increased greenhouse gases. Additionally, Dr. Mann found significant internal 50-100 year oscillations with similar features occurring over several centuries. Similar oscillatory signals have been attributed to variability in the thermohaline circulation and coupled ocean-atmosphere processes in other model simulations. Dr. Mann also found a distinct 10-15 year oscillation in the instrumental data. This evidence of several inter-annual climate signals makes the interpretation of data more complicated in terms of a simple linear dynamical mechanism.

Next, Dr. Mann proposes an alternative method for modeling the ocean circulation variations with respect to climate. He takes the equations governing ocean circulation and subdivides them into two sections. Dr. Mann studies these two dominant modes of circulation, the gyre and mean meridional overturning circulations, separately as well as dynamically coupled. He notes that this type of modeling is only relevant when the non-uniform effects for each section are properly taken into account. After deriving these two components, Dr. Mann introduces some approximations and simplifications to allow for their coupling. From this, Dr. Mann creates what he believes to be a reasonably faithful description of large-scale ocean circulation, temperature and salinity fields of the world's oceans. However, he notes one caveat, that this model is very sensitive to dynamics created when the gyre circulation is not taken into account. When these gyre-scale processes are absent from the model, a 200-300 year mode of ocean variability is clearly defined, taken by Dr. Mann to be the linear mode of the variability in the meridional overturning circulation. Dr. Mann estimates the effects of the ocean circulation on the atmosphere by parameterizing the modeled response of the atmosphere to sea surface temperature variations. When gyre-scale processes are accounted for, a 70-to-100 year instability is present. Dr. Mann interprets this variation as an oceanic delayed oscillator mechanism caused by changes in the meridional overturning, which subsequently causes changes in the near-surface salinity and heat advection. These effects in turn dampen the meridional overturning circulation before it can become large-scale. Dr. Mann contends that the results of the model study underscore possible interactions between these two major circulation processes and the nature of decadal to century scale variability.

Dr. Mann contends that his work shows strong evidence for the existence of 50-100 year scale oscillations centered in the North Atlantic, persistent over several centuries. This is suggestive (but not conclusive) of damped oscillations in the climate system. Dr. Mann also contends that an atmospheric response to both of these major circulation processes is identified, corresponding closely to sea level pressure variations. Lastly, Dr. Mann defends the robustness of this simplified model, stating that the observed climate variability is consistent with many complex climatic mechanisms not included in this study. However, with more long-term proxy data and more large-scale climate reconstructions, the application of the signal detection methods described here will provide further insight into the nature of these decadal to century scale climate signals.

Summary of *Global-scale temperature Patterns and Climate Forcing Over the Past Six Centuries* by Michael Mann, Raymond Bradley, and Malcolm Hughes (1998)

In an attempt to understand long-term global climate variation, Mann et al. use a widely distributed set of proxy and instrumental climate indicators to reconstruct global temperature patterns over the past 500 years. Using this data, they attempt to estimate the relationship between global temperature changes and variations in volcanic aerosols, solar irradiance, and greenhouse gas concentrations.

The data consisted of a multiproxy network. In this case proxy is a time series constructed using data from various sources, such as tree ring measurements, ice cores, ice melts, and historical records. Overall the network includes 112 proxies, and each series has been formatted into annual mean anomalies relative to the reference period used for this data, 1902-1980. Certain tree-ring datasets have been represented by a small number of leading principal components. The dendroclimatic data has also been carefully reviewed to ensure standardization and sizeable segment lengths. Although the data network covers large portions of the globe, there is only enough reliable information to conduct a spatial analysis of the Northern Hemisphere.

Because of the heterogeneity of the information available, Mann et al. calibrated the datasets by first decomposing the 20th century instrumental data into its dominant patterns of variability using principal component analysis, and subsequently calibrating the individual climate proxy indicators against the time histories of these distinct patterns during their mutual interval of overlap. Included in this calibration approach are three assumptions: 1) the indicators in our network are linearly related to one or more of the instrumental training patterns, 2) a relatively sparse, but widely distributed sampling of long proxy and instrumental records may measure the small number of degrees of freedom in climate patterns at interannual and longer timescales, and 3) patterns of variability captured by the multiproxy network have analogs in the patterns they find in the shorter instrumental data. In their principal component analysis (PCA), Mann et al. isolated a small number of dominant patterns of variability, otherwise labeled 'empirical eigenvectors'. Each of these patterns or eigenvectors has a characteristic spatial pattern and a pattern evolving over time (also referred to as the 'principal component'). These eigenvectors are ranked according to the percentage of variance they describe. The first five eigenvectors describe 93% of the total variance. Each of the indicators in this study was calibrated using these five eigenvectors.

The temperature reconstructions derived using all indicators and the most optimal eigenvector subsets show long term trends including pronounced cold periods during the mid-seventeenth and mid-nineteenth centuries and warmer intervals during the mid-sixteenth and late eighteenth centuries. Based on their methods, almost all of the years before the twentieth century exhibit temperatures well below the twentieth century mean temperature. Taking into account the uncertainties in their reconstruction, they find that the years 1990, 1995, and 1997 each show anomalies that are greater than any other year back to 1400, with roughly a 99.7% level of certainty.

Other general circulation and energy-balance model experiments, including some statistical comparisons of twentieth century global temperatures with forcing series, suggest that although both solar and greenhouse gas forcings play some role in explaining twentieth century climate trends, greenhouse gases appear to play an increasingly dominant role during this century. Additionally, it is hoped that as larger numbers of high quality proxy reconstructions become available it may be possible to create a more globally representative multiproxy data network for further study.

Summary of *Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations* Mann et al. (1999)

Estimates of climate variability during past centuries rely upon indirect “proxy” indicators – natural archives that record past climate variations: tree rings, sediments, ice cores and corals. MBH98 used these temperature proxies to reconstruct yearly global surface temperature patterns back to CE 1400. In this article, Mann et al. attempt to reconstruct global surface temperature patterns prior to 1400 because it is surmised that temperatures were warmer even before the period reconstructed in MBH98. However, in order to recreate these temperature patterns, the same methodology employed in MBH98 was applied to an even sparser proxy data network available prior to CE 1400. Only 12 viable indicators are available for this time period. Because only a small number of indicators are available in regions where the primary pattern of hemispheric mean temperature variation has significant amplitude, these indicators have a particularly important role. Just as in MBH98, the calibration procedure for these 12 indicators invokes two assumptions: first, that a linear relationship exists between proxy climate indicators and some combination of large-scale temperature patterns and second, that patterns of surface temperature in the past can be suitably described in terms of some linear combinations of the dominant present-day surface temperature patterns. The calibration/verification statistics for reconstructions based on the 12 indicators are somewhat degraded compared to those for the post CE 1400 period. The explained variance in the MBH98 data (post-1400 AD) was between 42% and 51%, whereas the explained variance among these 12 indicators is between 34% and 39%. Furthermore, the first principal component of the ITRDB (International Tree Ring Data Bank) data in this analysis is the only one of these series that exhibits a significant correlation with time history of the dominant temperature pattern of the 1902-1980-calibration period. If this indicator is omitted, positive calibration/variance scores cannot be obtained for the Northern Hemisphere (NH) series. Thus, ITRDB PC1 is the most meaningful component in resolving hemispheric temperature trends. The assumption that this relationship is consistent with time requires closer study and as such a more widespread network of proxy indicators will be required for more confident inferences.

The reconstructed NH series indicates a cooling period prior to industrialization, possibly driven by astronomical forcing[1], which is thought to have driven long term temperatures down since the mid-Holocene period. In addition, significant long-term climate variability may be associated with solar irradiance variations. Our reconstruction supports the notion that warmer hemispheric conditions took place early in the millennium followed by a long period of cooling beginning in the 14^{th} century, which can be viewed as the initial onset of the Little Ice Age. However, even the warmer intervals in our reconstruction pale in comparison with modern (mid-to-late 20^{th} century) temperatures. The data still upholds the conclusion that the 1990s was likely the hottest decade and that 1998 was likely the hottest year of the millennium. However, without more widespread high-resolution data, further conclusions cannot be drawn in regard to the spatial and temporal details of climate change in the past millennium and beyond.

---

[1] Astronomical forcing refers to the effect on climate of changes in the tilt and the shape of the orbit of the Earth.

Summary of *Global Temperature Patterns in Past Centuries: An Interactive Presentation* by Michael Mann, Ed Gille, Raymond Bradley et al. (2000)

This paper expands on earlier work by the same authors (*Global-scale temperature Patterns and Climate Forcing Over the Past Six Centuries* by Mann et al.), which used multiproxy networks of indirect climate variability indicators, such as tree-ring measurements, ice cores, coral growth, etc., to reconstruct climate variability and temperatures over the past centuries. In subsequent papers, Mann et al. expanded the reconstruction to Northern Hemisphere (NH) temperature variations over the past millennium, to examine ENSO-scale patterns of climate variability during past centuries, to compare observed patterns of climate variability in the Atlantic, and to assess the relationship between global patterns of climate variation and particular regional patterns. Most recently, Mann et al. have made available for the first time seasonally resolved versions of global temperature surface patterns in an interactive format, allowing users to select specific spatial regions or time periods of interest. Details of the data and methods involved are discussed in Mann et al. (1998) and (1999).

The data consisted of a multiproxy network. In this case proxy is a time series constructed using data from various sources, such as tree ring measurements, ice cores, ice melts, and historical records. Overall the network includes 112 proxies, and each series has been formatted into annual mean anomalies relative to the reference period used for this data, 1902-1980. Certain tree-ring datasets have been represented by a small number of leading principal components. The dendroclimatic data has also been carefully reviewed to ensure standardization and sizeable segment lengths. Although the data network covers large portions of the globe, there is only enough reliable information to conduct a spatial analysis of the Northern Hemisphere.

Because of the heterogeneity of the information available, Mann et al. calibrated the datasets by first decomposing the 20th century instrumental data into its dominant patterns of variability using principal component analysis, and subsequently calibrating the individual climate proxy indicators against the time histories of these distinct patterns during their mutual interval of overlap. Included in this calibration approach are three assumptions: 1) the indicators in our network are linearly related to one or more of the instrumental training patterns, 2) a relatively sparse but widely distributed sampling of long proxy and instrumental records may measure the small number of degrees of freedom in climate patterns at interannual and longer timescales., and 3) patterns of variability captured by the multiproxy network have analogues in the patterns we resolve in the shorter instrumental data. In their principal component analysis (PCA), Mann et al. isolated a small number of dominant patterns of variability, otherwise labeled 'empirical eigenvectors'. Each of these patterns or eigenvectors has a characteristic spatial pattern and a pattern evolving over time (also referred to as the 'principal component'). These eigenvectors are ranked according to the percentage of variance they describe. The first five eigenvectors describe 93% of the total variance. Each of the indicators in this study was calibrated using these five eigenvectors.

The most recent temperature reconstructions indicate that 1998 (as opposed to 1990, 1995 and 1997 as previously proposed in Mann et al. 1998, 1999) was most likely the warmest year of at least the past millennium. There are also distinct temperature trends for the Northern and Southern hemispheres. While both hemispheres have similar trends, the coldness of the 19th century appears to be somewhat more pronounced in the Northern hemisphere. Additionally, evidence suggests that the post-1850 warming was more dramatic at higher latitudes relative to lower latitudes due to larger positive feedbacks at high latitudes. The annual mean temperature trends at higher latitudes are seen to be greater than the hemispheric trends themselves. In contrast, the tropical band shows less change than the entire Northern Hemisphere series.

Mann et al. also provide yearly global temperature maps for annual mean, boreal cold season, and warm season for the reconstructed temperature fields from 1730 to 1980, the raw temperature data from 1902-1993 (used for calibration) and the sparse raw "verification" data from 1854 to 1901 (used for cross-validation). Users can investigate spatial patterns and time histories of this global temperature data at http://www.ngdc.noaa.gov/cgi-bin/paleo/mannplot2.pl.

The statistical relationship between variations in the NH mean temperature and estimates of the histories of solar, greenhouse gas and volcanic forcings suggest that while the natural forcings play a role, only greenhouse gas forcing alone can explain the unusual warmth of the past few decades. Mann et al. also examined the sensitivity surrounding these forcings and found that when physically reasonable lags are incorporated into the attribution analysis there is evidence of even greater statistical relationships with particular forcings. At the physical lag of one year, the relationship between temperature variations and volcanic forcing is slightly more consistent. At the physical lag of 10-15 years the relationship between greenhouse gas increases and increasing temperatures is considerably more significant, while the relationship with solar irradiation is less significant. Thus, there is significant evidence that recent anthropogenic activities are contributing to the recent warming.

It is clear that the primary limitations of large-scale proxy-based reconstruction in past centuries, both temporally and spatially, reside in the increasingly sparse nature of available proxy networks available to provide reliable, past climate information. Arduous efforts are needed to extend such networks in space and time to the point where significant improvements will be possible in order to gain a more empirical understanding of climate variations during the past millennium.

Summary of *Ocean Observations and the Climate Forecast Problem* by Carl Wunsch (2002)

Due to the recent importance of studying climate change, it has become apparent that there are significant problems in observing the ocean and its climate. Much of the problem is technical, but there is also the matter of culture and misapprehension. Many in the field of meteorology continue to have an antiquated and misleading perception of the ocean circulation. In his article, Wunsch outlines the reasons for many of the problems in observing the ocean.

Since the opacity of the ocean has made it difficult to observe until recent technological innovations, and the cost of supporting oceanographic ships is prohibitive, time series of oceanic variables were almost nonexistent. The only variables relatively easy to measure and interpret were properties such as temperature, salinity and oxygen. Since these properties are particularly stable, the static picture of ocean circulation became the predominant view. However, with the advent of modern electronics, obtaining time series of oceanographic data became easier. After years of literature and data on the subject, it became clear that the ocean is actually quite turbulent under the surface and that few, if any, elements of ocean circulation are truly steady.

There exists a large-scale oceanic circulation that appears to be steady over decades, but is thought to be slowly changing everywhere in ways not yet known. However, the current understanding of how oceanic circulation will affect the climate is actually very narrow. Additionally, the problem is further compounded by the fact that models have become so sophisticated and interesting, it is tempting to assume they must be skillful. Most papers written on the subject of oceanographic models give little or no guidance to the reader as to the actual expected skill of the model. This type of research begs the question, is it really plausible that a 4° or 1° ocean model can be integrated with skill for 1000 years? The magnitude of the error in these models is enormous when integrated over such a long time period. The evidence for the skillfulness of similar models is scant.

The assumption that the oceanic system is much simpler than it actually is leads to a corruption of the entire literature. Readers of paleoclimate papers will notice that extraordinarily complicated and far-reaching changes in the climate system are often reduced to simple assertions about how the "global conveyor" changed. One might also be suspicious of "concrete" evidence of atmospheric modeling because atmospheric modeling must be equally if not more difficult than modeling the ocean. In order to begin to make any kind of model, years of observation with oceanographic satellites are needed. Most of these satellites are not currently regarded as operational. Of primary concern is to insure that everyone understands the problem and to recognize the great influence past assumptions exercise over future necessity.

Summary of *Corrections to the Mann et al. (1998) Proxy Database and Northern Hemispheric Average Temperature Series* by Stephen McIntyre and Ross McKitrick (MM03) (2003)

In their paper, *Corrections to the Mann et al. (1998) Proxy Database and Northern Hemispheric Average Temperature Series,* (hereafter referred to as MM03), McIntyre and McKitrick assess the methodology and results of the widely referenced Mann, Bradley, and Hughes paper, *Global Scale Temperature Patterns and Climate Forcing Over the Past Six Centuries* (hereafter referred to as MBH98). In MBH98 the authors attempted to reconstruct a temperature history of the Northern Hemisphere for the period 1400-1980. Their result was a "hockey stick" shaped graph, from which they concluded that the temperatures of the late 20^{th} century were unprecedented and that 1990-2000 was likely the hottest decade in the millennium, and 1998 was likely the hottest year in the millennium. These findings have been prominent in the discussion on global climate change and in subsequent policy discussions. MM03 attempts to recreate the research in MBH98 in order to prove or disprove their findings.

In the course of the research reproduction, McIntyre and McKitrick found errors in the statistical methodology of MBH98. Primarily, MM03 found that the creation of the proxy database itself held serious errors. In this context proxy denotes one of the 112 physical measurements used that can serve as an indicator of climatic conditions, including temperature. Examples of proxies include tree measurements, ice cores, and coral calcification rates. The time series created from these measurements form the basis of the MBH98 study.

MM03 claimed the following errors in the MBH98 proxy database:

1. unjustified truncation of three time series
2. copying 1980 values from one series onto another
3. displacement of 18 series to one year earlier than apparently intended
4. Statistically unjustified extrapolations or interpolations to cover missing entries in 19 series
5. geographical mislocations and missing identifiers of location
6. inconsistent use of seasonal temperature data where annual data is available
7. obsolete data in at least 24 series, some of which may have been obsolete at the time of the MBH98 study
8. listing of unused proxies
9. incorrect calculation on all 28 tree ring principal components.

Having accounted for the major errors, MM03 reconstructed the temperature history. Using the MBH98 methodology, they were able to accurately reproduce the "hockey stick" shaped graph in the MBH98 findings. Still using the same basic methodology, MM03 prepared the data with improved quality control, including using the most recent data and collating it correctly. The result was a northern hemisphere temperature reconstruction that takes on a different shape in which the temperature index peaks at around 1450 AD, near the earliest measured point on the graph. MM03 concluded that the errors in MBH98 make the data unreliable and obsolete such that it does not support their end conclusions.

Summary of *Global Surface Temperature over the Past Two Millennia* by Michael Mann and Philip Jones (2003)

Mann and Jones present their reconstructions of Northern and Southern Hemisphere mean surface temperature over the past two millennia based on high-resolution (annually or decadally scaled) proxies. For the Northern Hemisphere, they use previous temperature reconstructions from eight distinct regions based on 23 individual proxy records and for the Southern Hemisphere, they use temperature reconstructions from five distinct regions. Composites were performed separately for each hemisphere, based on available regional temperature records. Each regional temperature record was standardized by removal of the long-term mean and division by the standard deviation, after decadal smoothing. The composites were weighted combinations of the standardized proxy series, weighted by size of region and estimated reliability of the climate signal in the proxy. Proxy records exhibiting negative or approximately zero local correlations were eliminated from the study. Each composite was also standardized to have the same mean and decadal standard deviation as the target instrumental series over the period of common overlap.

The Northern Hemisphere reconstruction was observed to be largely insensitive to elimination of shorter proxy records or to the weighting of the proxy series, suggesting a significant robustness. The reconstruction is consistent with previous Mann reconstructions, in that the warmth in the late 20^{th} century is unprecedented. Larger uncertainties in the Southern Hemisphere reconstruction preclude a similar conclusion for this. Increased quality and quantity of Southern Hemisphere proxy records are needed to decrease the current uncertainties surrounding the reconstruction and definitively make conclusions about the climate variability.

Summary of *Reconstructing Past Climate from Noisy Data* by Hans von Storch et al. (2004)

While attempting to measure anthropogenic effects on the earth's climate, it is necessary to create a reconstruction of past climate variations. Most studies have identified varying warm values in the 11th and 12th centuries followed by secular cooling periods in the mid-16th, 17th and early 19th centuries. These cooler intervals were followed by warming that is still experienced today. The amplitude of these preindustrial variations is debated, although the most notable study on the subject and the most quoted, Mann et al. 1998 (MBH98), as well as the Intergovernmental Panel on Climate Change (IPCC), report that these variations were of small amplitude. However, recent studies have suggested that centennial variations may have been larger than previously thought. This study uses a coupled atmosphere-ocean model simulation of the past millennia as a surrogate climate to test the reconstruction method of MBH98.

Using this model as a virtual world to determine the skill of regression-based reconstruction models like MBH98, von Storch et al. found that the model is reasonably skilled at reproducing short-term variations but substantial underestimation occurs in the long-term estimations. On an inter-annual scale, the reconstruction has a calibration reduction-of-error statistic of .7 for perfect pseudo-proxies and .3 for pseudo-proxies with a higher degree of noise. However, only 20% of the 100-year variability is recovered when the noise level is approximately 50%. Similar results were obtained using the third Hadley Centre coupled model (HadCM3), indicating the results are not dependent on the model used.

Von Storch et al. also tested a number of other hypotheses. They found that including more instrumental data in the proxies does not improve results, expanding the proxy set in sparse areas improved results marginally, and that expanding the range of temperature variability present in the pseudo-proxies greatly improves the results. Additionally, von Storch et al. questioned the validity of linear regression models in general in estimating climate. Using pseudo-proxies to estimate local temperatures which were then spatially averaged to derive a Northern Hemisphere temperature, they found similar problems that occur in MBH98: underestimation of low-frequency variability for a given amount of noise. The authors conclude that climate simulations of the past millennium are burdened by model limitations and uncertainties in external forcing and therefore the output must be considered with care. Additionally, the linear regression methods as used in MBH98, suffer from marked losses of centennial and multidecadal variations.

Summary of *The M&M Critique of the MBH98 Northern Hemisphere Climate Index: Update and Implications* by Stephen McIntyre and Ross McKitrick (MM05a) (2005a)

In an extension of their 2003 paper (*Corrections to the Mann et. Al. (1998) Proxy Database and Northern Hemispheric Average Temperature Series)*, McIntyre and McKitrick further detail their critique of Mann et. al. (1998) and respond to its subsequent update Mann et. al. (1999). In response to McIntyre and McKitrick (2003), Mann et. al. published new information regarding their original research that MM03 attempted to replicate. While the new information did not include the source code used to generate the original results, it did include an extensive archive of data and supplementary information on the methods at the University of Virginia FTP site.

In their article, M&M indicate that the individual data series (proxies) used to reconstruct the temperature index are important, and that errors within these series do not get washed out in a multi-proxy study. Specifically, MM05a found that the differences in MBH98 and MM03 can be almost fully reconciled through the variations in handling of two distinct series, the Gaspe "northern treeline" series and the first principal component (PC1) from the North American proxy roster (NOAMER). In MBH98, the first four years of both of these series were extrapolated. The extrapolation has the effect of depressing early 15th century results, and was not disclosed by Mann et al. until a later paper, Mann et al. (2004). The underlying dataset that was subject to extrapolation also fails to meet the data quality standards described by Mann et al. elsewhere in the paper.

In the MBH98 methodology, they used a principal component analysis, which they reported to be conventional or centered. However, in further disclosure of information on the UVA FTP site, it has been determined that the principal component analysis was not actually centered. In fact the mean used in their calculations is the 1902-1980 mean, but it was applied to the period 1400-1980. The effect of de-centering the mean is a persistent "hockey stick" shaped PC1, even when layered with persistent red noise. It follows from this shape that the climate of the late 20th century was unprecedented. Because the original code is in FORTRAN, which takes much more programming to run statistical processes than modern software such as R, it is very possible that this is due to a programming error, although Mann et al. have not admitted to any such error.

In the MBH98 de-centered principal component calculation, a group of twenty primarily bristlecone pine sites govern the first principal component. Fourteen of these chronologies account for over 93% variance in the PC1 and 38% of the total variance. The effect is that it omits the influence of the other 56 proxies in the network. In a centered version of the data, the influence of the bristlecone pine drops to the fourth principal component, where it accounts for 8% of the total variance. The MM03 results are obtained if the first two NOAMER principal components are used. The MBH98 results can be obtained if the NOAMER network is expanded to five principal components. Subsequently, their conclusion about the climate of the late 20th century is contingent upon including low-order principal components that only account for 8% of the variance of one proxy roster. Furthermore, the MM03 results occur even in a decentered PC calculation, regardless of the presence of PC4, if the bristlecone pine sites are excluded.

In the Gaspe "northern treeline" series, MM05a found that the MBH98 results occur under three conditions: 1) the series must be used as an individual proxy; 2) the series must contain the portion of the series that relies only on one or two trees for data; and 3) it must contain the ad-hoc extrapolation of the first four years of the chronology. Under all other conditions, including using an archived version of the series without extrapolation, MM03 type results occur.

MM05a also addresses the MBH98 claims of robustness in their findings. The sensitivity of the 15th century results to slight variations in the data and method of two individual series show a fundamental instability of the results that flatly contradicts the language used in MBH98 and in Mann et al. (2000) where it states "…whether we use all data, exclude tree rings, or base a reconstruction only on tree rings, has no significant effect on the form of the reconstruction for the period in question…" Additionally, MM05a notes much of the specialist literature raises questions about these indicators and at the least these questions should be resolved before using these two series as temperature proxies, much less as uniquely accurate stenographs of the world's temperature history.

In response to MM03, Mann et al. wrote several critiques that appeared in *Nature* magazine as letters and as separate articles. The Mann et al. (2004) paper argued that the MM03 use of centered principal components calculations amounted to an "effective omission" of the 70 sites of the North American network. However, the methodology used omits only one of the 22 series. A calculation like this should be robust enough that it is relatively insensitive to the removal of one series. Also, "effective omission" is more descriptive of the MBH98 de-centering method, which uses 14 bristlecone sites to account for over 99% of explained variance.

In another response, Mann et al. claim that the PC series are linear combinations of the proxies and as such cannot produce a trend that is not already present in the underlying data. However, the effect of de-centering the mean in PC analysis is that it preferentially selects series with hockey-stick shapes and it is this over weighting that yields a pattern not representative of the underlying data. Additionally, Mann et al. responded to the MM03 critique of the bristlecone pine, which pointed out that the bristlecone pine had no established linear response to temperature and as such was not a reliable temperature indicator. Mann et al. responded by stating that their indicators were linearly related to one or more instrumental training patterns, not local temperatures. Thus, the use of the bristlecone pine series as a temperature indicator may not be valid.

The authors of MM05 concluded that the various errors and adverse calculations that were not disclosed exhibit the limitations of the peer review process. They also note the limited due diligence of paleoclimate journal peer review and that it would have been prudent to have checked the MBH98 data and methods against original data before accepting the findings as the main endorsement of the Intergovernmental Panel on Climate Change.

Summary of *Hockey sticks, principal components, and spurious significance* by Stephen McIntyre and Ross McKitrick (2005b)

In their critique of *Global-scale temperature Patterns and Climate Forcing Over the Past Six Centuries* (MBH98) by Mann et al., McIntyre and McKitrick (M&M) note several errors in the methodology and subsequent conclusions made by Mann et al. First, M&M discuss the incorrect usage of principal component analysis (PCA) in MBH98. A conventional PC algorithm centers the data by subtracting the column means of the underlying series. For the 1400 to 1450 data series, the FORTRAN code contains an unusual data transformation prior to the PC calculation, which was never reported in print. Each tree ring series was transformed by subtracting the 1902-1980 mean and then dividing by the 1902-1980 standard deviation and dividing again by the standard deviation of the residuals from fitting a linear trend in the 1902-1980 period. For PCA, if the 1902-1980 mean is close to the 1400-1980 mean, then there will be very little impact from this linear transformation. However, if the means differ, then the explained series variance is inflated. Since PCA gives more weight to series that have more explained variance, the effect is preference for the 'hockey stick' shape seen in Mann et al.. This 'hockey stick' shape supports the conclusions that climatic conditions in the late twentieth century are anomalies.

M&M also ran a Monte Carlo Simulation on 70 of the stationary proxy data series. When applying the linear transformation described above that was found in MBH98, nearly every simulation yielded first principal components (PC1) with a 'hockey stick' shape. Without this transformation, the 'hockey stick' shape appeared in the PC1 only 15.3% of the time. Additionally, the MBH98 method creates a PC1 that is dominated by bristlecone pine and foxtail pine tree ring series (both closely related species). Out of the 70 sites in the network, 93% of the variance in the MBH98 PC1 is accounted for by only 15 bristlecone and foxtail pine sites, all with data collected by one man, Donald Graybill. Without the transformation, these sites have an explained variance of less than 8%. The substantially reduced share of explained variance coupled with the omission of virtually every species other than bristlecone and foxtail pine, argues strongly against interpreting it as the dominant component of variance in the North American network. There is also evidence present in other articles calling the reliability of bristlecone pines as an effective temperature proxy into question.

M&M also evaluated the MBH98 usage of the Reduction of Error statistic in place of the more reliable and widely used Monte Carlo Model to establish significant benchmarks. By using the Monte Carlo Model, M&M found that a more accurate significance level for the MBH98 procedures is .59, as opposed to the level of 0.0 reported in the original study. A guard against spurious RE significance is to examine other statistics, such as the R^2 and CE statistics. However, MBH98 did not report any additional statistics for the controversial 15^{th} century period. The M&M calculations indicate that these values for the 15^{th} century section of the temperature reconstruction are not significant, thereby refuting the conclusions made by MBH98.

Summary of *Highly Variable Northern Hemisphere Temperatures Reconstructed from Low- and High-Resolution Proxy Data* by Anders Moberg et al. (2005)

In their study, Moberg et al. reconstruct a climate history for the past 2,000 years using low resolution proxies (proxies that provide climate information at multi-centennial timescales, such as ocean sediment cores) and high resolution proxies (proxies that provide climate information on a decadal scale, such as tree rings). Due to the high profile of high-resolution proxies in reconstructions, mostly from Mann et al. 1998, views have been expressed that only tree ring and other high resolution data are useful for quantitative large scale temperature reconstructions. However, tree ring data has a well documented unreliability in reproducing multi-centennial temperature variability. By using low-resolution data for multi-centennial information combined with high-resolution data for decadal information, the most unreliable timescales for each proxy can be avoided.

The dataset used for this study was limited since proxies were required that dated back 2,000 years. Seven tree-ring series and eleven low-resolution proxy series were used. To obtain a reconstruction covering the complete range of timescales Moberg et al. created a wavelet transform to ensure tree-ring records contribute only to timescales less than 80 years and all low-resolution proxies contribute only to longer timescales. To calibrate the reconstruction, its mean value and variance were adjusted to agree with the instrumental record of Northern Hemisphere annual mean temperatures in the overlapping period 1856-1979.

The reconstruction indicates two warm peaks around A.D. 1000 and 1100 and pronounced cold periods in the 16^{th} and 17^{th} centuries. The peaks in medieval times are comparable to those of the 20^{th} century, although warmth seen in post-1990 seems to be unprecedented. Reconstructions of the temporal evolution of warming variables (volcanic aerosols, solar irradiance and greenhouse gases) have been used to drive simple energy balance climate models as well as fully coupled atmosphere-ocean general circulation models. Moberg et al. note that the Northern Hemispheric temperature series obtained from such an experiment with the coupled model ECHO-G bears a strong qualitative similarity to their reconstruction. This supports the case of a pronounced hemispheric low-frequency temperature variability resulting from the climate's response to natural changes in radioactive forcing.

There are notable differences in the Moberg et al. reconstruction and that of Mann et al. 1998. While there is a large amount of data in common between the two reconstructions, Mann et al. combined tree-ring data with decadally resolved proxies without any separate treatment at different timescales. Additionally, this study's dataset contains centennially resolved data from the oceans while Mann et al. used only annually or decadally resolved data from continents or locations near the coast. Mann et al. also used a different calibration method (regression versus variance scaling as in this study).

Further study in the process of weighting different timescales and spatial representation of the data should be conducted to see which method most accurately depicts past climate variability. This study finds no evidence for any earlier periods in the past two millennia with warmer conditions than the post-1990 period. However, natural multi-centennial climate variability, especially as a response to solar irradiance, may be larger than previously thought. This does not imply that global warming has been caused by natural factors alone, but that there is a need to improve scenarios for future climate change by also including forced natural variability.

Summary of *Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate* by Michael Mann, Scott Rutherford, Eugene Wahl and Caspar Ammann (2005)

In this article Mann et al. examine two prominent methods in historical climate reconstruction, Climate Field Reconstruction (CFR) and the Composite-Plus Scale (CPS). The former combines several different proxy[2] records in order to reconstruct underlying patterns of past climate change. The latter combines many different proxy series (such as tree ring series, ice core series, etc.) and scales the resulting composite against a target series (i.e. the Northern Hemisphere) that is measured instrumentally. In order to assess both methods, Mann et al. used climate simulation to create a known climate record. They then layered the model with the typical noise associated with real-world uncertainties found in actual proxies. Thus, Mann et al. created pseudo proxies that they could use to test the two methods of climate reconstruction. They constructed three distinct networks of pseudo proxies, each with attributes similar to actual proxy networks used in past CFR and CPS studies.

Following the standard CPS procedure, each pseudo proxy was smoothed by decade and standardized. The weighted composite of these proxies was then scaled to have the same mean and standard deviation as the actual Northern Hemisphere series. Using different levels of the signal-to-noise ratio (SNR) (relative amplitudes of noise variance), Mann et al. evaluated the effectiveness of each method. In CPS experiments, the results most closely resembled those obtained from actual proxies for SNR=1.0. The lower the SNR level (.25 and .5 were also measured), the lower the skill of reconstruction. Additionally, when SNR=1.0, the CPS method was found to be relatively insensitive to the length of the calibration interval. Mann et al. found that in general, CPS or regression based methods employing a short calibration period are likely to underestimate long-term variability.

Mann et al.'s implementation of the CFR approach makes use of the regularized expectation maximization (RegEM), which is similar to Principal Component Analysis (PCA), but it employs estimates of data covariances in iterations. Mann et al. tested three types of this method: the straight application of RegEM, a "hybrid frequency-domain calibration" approach and a stepwise version of RegEM. All three of these methods yielded similar results in the study. Similar to CPS, Mann et al. found that when SNR=1.0, this method yielded a similar resolved variance and it was relatively insensitive to the calibration period. However, this method yielded a moderately more skillful reconstruction with a long calibration period. Additionally, the CFR method appears to systematically underestimate the amplitude of the larger volcanic cooling events, most likely because of the small number of volcanic events present in the calibration interval.

In general, Mann et al. found no evidence that real-world proxy-based temperature reconstructions are likely to suffer from any systematic underestimate of low-frequency variability. Their findings also suggest that both of these standard methods, CPS and

[2] In this case, proxy refers to a time series of indicators such as tree rings, ice cores, and coral.

CFR, are likely to provide a faithful estimate of actual long-term hemispheric temperature histories, within estimated uncertainties.

Summary of *Low-Frequency Signals in Long Tree-Ring Chronologies for Reconstructing Past Temperature Variability* by Jan Esper, Edward Cook, and Fritz Schweingruber (2005)

In this article Esper et al. address the debate revolving around the reliability of tree-ring records as substantial basis of temperature reconstruction before the 17th century. The authors' present analysis of centuries-long ring-width trends in 1205 radial tree-ring series from 14 high-elevation and middle-to-high latitude sites distributed over a large part of the Northern Hemisphere extratropics. Esper et al. looked at growth trends in tree ring proxies by analyzing individual raw ring-width measurements using Regional Curve Standardization (RCS) methods. Successful use of the RCS method usually requires a large number of ring-width series because the method of detrending is not based on any explicit curve fitting to the individual series, but rather over series of a similar region. However, the series are further broken down into two groups, those that age linearly and those with age trends that are non-linear.

In each of these groups, the smoothed regional curves were estimated from the averaged biological age-aligned data. The resulting tree ring indices were then averaged into linear and non-linear mean value functions to produce two nearly independent tree-ring chronologies covering the years 800-1990. Each of these chronologies showed evidence of above average temperatures during the Medieval Warming Period (900-1300), below average temperatures during the Little Ice Age (1200-1850), and large-scale warming after 1850, consistent with instrumental temperature records. Overall, these results demonstrate that properly selected and processed tree-ring records can preserve such long time-scale climate variability.

Additionally, using RCS methods, climate variability of the Medieval Warming Period (MWP) can be reconstructed, and it approaches the magnitude of 20th-century warming in the Northern Hemisphere up to 1990. Consistent with other analyses of the MWP, it appears to be more temporally variable than the warming trend of the past century. Analysis also supports that the warmest period of the MWP may have begun in the early 900s, with the warmest interval being from 950 to 1045 AD. This finding suggests that past comparisons of the MWP with the 20th-century warming may not have included all of the MWP, especially its warmest period.

Summary of *Are Multiproxy Climate Reconstructions Robust?* By Gerd Bürger and Ulrich Cubasch (2005)

Bürger and Cubasch review the robustness of multiproxy climate reconstructions, especially with regard to Mann et al. 1998 (MBH98), a prominent and widely discussed paper on the topic. The MBH98 reconstruction applies an inverse regression between a set of multiproxies on one hand and the dominant temperature principal components (PCs) on the other. The sparse availability of proxies prior to 1450 is accounted for by estimating the regression for seven successive time periods. Bürger and Cubasch skip this last step in their approximation of the MBH98 settings. In this study, they use the proxies available for the period 1400-1450 (which includes 18 tree ring and ice core proxies). Using the 1902-1980 as the calibration period, an empirical model is fitted and applied to the full proxy record. Before estimating the regression model, the proxies undergo a PC transformation, a measure against collinearity, which can inflate the model error.

Bürger and Cubasch found 64 variants of reconstructed millennial Northern Hemisphere temperatures. The spread about the MBH98 is immense, especially around the years 1450, 1650, and 1850. There is no evidence that one variation should be chosen over the others and even the variant with the best reduction of error statistic (79%) is the variant that most strongly deviates from MBH98. When the setting was moved to AD 1600 instead of 1400, the spread is still quite large in the early part of the reconstruction, even though more proxies are available.

Fundamental to all proxy inferences is the assumption that the variations in the proxy are related to the temperature and somewhat uniformly. However, the results of this study do not give such a relationship, at least not one that is robust. Bürger and Cubasch could not find one criterion solely responsible for the spread of variants, but it is possible that a significant source of uncertainty could be the scale mismatch between the full millennial and the calibrating proxy variations. In that case, the regression model leaves its general domain of validity and is applied in an extrapolative manner. The further the estimated regression laws are extrapolated the less robust the method is. This error growth is especially critical for parameter-insensitive, multi-proxy climate field reconstructions of the MBH98 type. In order to salvage such a method, there must be a mathematical derivation of the model error and more sophisticated regularization schemes that can minimize the error.

Summary of *Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season and Target Domain* by Rutherford et al. (2005)

Rutherford et al. discuss the necessity of climate reconstruction with multi-proxy networks as empirical evidence in documenting past climate variability. In this case, they note the advantage of using high-resolution proxies (annually or seasonally resolved proxies, such as tree rings, corals and ice cores) because they overlap with instrumental data of the past century, allowing analysis of their climate signal and reliability. These proxies have been used to reconstruct spatial climate fields which not only provide a climate variability record but which also retain information of the mechanisms or forcing underlying the variability. Annually resolved proxy networks have also been used to directly reconstruct indices of climate variability, but these methods are somewhat flawed in that they assume a direct relationship between the recorded proxy variables and temperature and precipitation but large-scale climate influences may change over time. Rutherford et al. focus specifically on recent constructions of this type of Northern Hemisphere temperatures and the reasons for the differences between reconstructions.

There are four identifiable factors that largely contribute to differences in reconstructions. Those are 1) using proxies as calibrators for surface temperature patterns, 2) the difference in character of the proxy networks used, 3) the target season of reconstruction, and 4) the target region of reconstruction. The intent of this study is to provide an assessment of the relative impacts of these four factors.

To measure the sensitivity of the proxy network selected, three networks were used: the multiproxy dataset used by Mann et al., the MXD data used by Briffa et al., and a combination of these datasets for the third network. To perform the reconstruction on these three networks a RegEM approach of climate field reconstruction was used. The RegEM method is an iterative method for estimating missing data through the estimation of means and covariances from an incomplete data field. The calibration interval for this approach was the time interval that includes overlap of proxy and instrumental data. Rutherford et al. made two modifications to the RegEM approach. First, they applied the method in a stepwise fashion, performing the reconstruction one step at a time using all available climate information. Second, they separated the datasets into low and high frequency datasets to create two independent reconstructions, which were then combined at the conclusion of the experiment to create a complete reconstruction. In the findings, Rutherford et al. stated that using a 20 year boundary for the frequency calibration gave superior results in almost all cases while the stepwise modification of the RegEM method did not produce any different results. Additionally, since the combined network showed only marginal improvement over the other two, it is likely that these reconstructions are relatively insensitive to the proxy network used.

To measure the sensitivity of the target season and region on reconstruction, Rutherford et al. performed an array of RegEM climate field reconstructions based on various seasons and regions. These reconstructions were compared with several previous reconstructions based on common predictor datasets. Rutherford et al. found that the optimal results for the MXD data were produced for the period in which the cold season *ends* while the optimal results for the multiproxy principal component reconstruction (Mann and coworkers reconstruction) were produced for the period in which the cold season *begins*. Additionally, the MXD network was found to outperform the combined network in the warm season. In terms of region, Rutherford et al. found differences in the target region lead to significant variability in the hemispheric mean estimates. While they found that reconstructions are sensitive to changes in season and region, Rutherford et al. maintained that the unprecedented temperatures in the late 20^{th} century that are seen in many reconstructions are supported with respect to all of the factors considered in this study.

Summary of *Abrupt Climate Change: An Alternative View* by Carl Wunsch (2006)

A Dansgaard-Oeschger (D-O) event is a rapid climate fluctuation, taking place at the end of the Ice Age. Twenty-three such events have been identified between 110,000 and 23,000 years before present. A widely held view of abrupt climate change during the last glacial period is that these D-O events are at least hemispheric, if not global and caused by changes in ocean circulation. It has been hypothesized that there may be abrupt climate change similar to a D-O event because of ongoing global warming and its oceanic affects. Underlying the major conclusions about D-O events and abrupt climate change there are several assumptions, including (1) the ^{18}Oxygen variations appearing in ice cores are viable as a proxy, (2) climate fluctuations in Greenland reflect those on a hemispheric or global basis, (3) the cause of D-O events can be traced to major changes of the North Atlantic meridional overturning circulation and perhaps failure of the Gulf Stream, and (4) apparent detection of a D-O event at a remote location in a proxy implies local climatic importance. In this article Wunsch reexamines these assumptions in order to assess their relevance, specifically focusing on (2) and (3).

In terms of using ^{18}Oxygen in the Greenland ice cores as a climate proxy, Wunsch found that although is was relatively accurate for central Greenland, when aligned with other locations a visual similarity would appear on the spectral graph, but that there was actually little statistical correlation; this occurred when comparing time periods of less than 900 years. While this does not disprove the hypothesis of a large impact of the D-O events, it cannot be used to support this assumption. There are three possible explanations for the disappearance of covariance for these periods less than 900 years. First, although both records have wide variability, it is primarily regional in character and there is no simple relationship between them. Second, the age-model (the calibration of age versus depth in the core) error has a larger influence on the short period variations than the long period ones. Third, different physical processes dominate the proxies at high frequency in the two separate locations, but they have roughly similar low spectral moments. Any of these factors could affect the lack of covariance between geographical locations. Subsequently, the assumption that there exist large-scale hemispheric correlations with the D-O events is neither proven nor disproven.

The heat flux associated with meridional overturning (the sinking and spreading of cold water and dispersion of heat) of the ocean has the most direct impact on the atmosphere in terms of oceanic circulation patterns. The contribution of the oceanic Northern Hemisphere to this pole-ward circulation falls very rapidly as heat is transferred to the atmosphere. At the 40th latitude North, the oceanic contribution is less than 25% of the atmospheric contribution. Hypothetically, if warming continues, and the Northern Atlantic is injected with fresh water from glacial melting, the meridional overturning circulation would be dramatically reduced, resulting in a D-O-like event. However, models attempting to construct this theoretical climate change have not been successful, mostly in that they have not taken into account the overlying wind field response to this event. Since much of the temperature flux of the North Atlantic is carried in the Gulf Stream, scenarios requiring wind shifts sufficient to shut it down are likely a physical impossibility because of the need to conserve angular momentum in the atmosphere.

Coupled models that have been claimed to show an atmospheric response to oceanic flux shifts are so simplified and lack adequate resolution that they cannot be skillfully integrated over the time periods required to describe true climatic time scales. Again, these models are only indicators of processes that *can* be operating but with no evidence that they dominate.

While the abrupt climate changes in Greenland may not have occurred in other parts of the globe, there still is the question of why it occurred in Greenland. One apparent observation is that the D-O events ceased in the Holocene and have been remarkably placid since. As such, the operative mechanism causing the D-O events must have also disappeared. The answer is the disappearance of the Laurentide and Fennoscandian ice sheets. Two enormous mountain ranges of high albedo (reflection factor) were removed. In a study by Jackson (2000), he noted that small, regional changes in the ice sheet elevations had a large effect on the atmospheric stationary wave patterns. As a standing wave, the wind encountering the ice sheets had more than one equilibrium state. Major local climate change could appear with a slight shift in the wave pattern of the wind system. While the model for this hypothesis is rough, other studies have indicated great influence of the ice sheets on atmospheric scales as well. The body of these theories suggests that the most important and sensitive determinant of oceanic circulation is wind, and not the temperature flux. Similarly, the widely accepted view that D-O events were of global impact and may occur as a result of recent warming is based on four assumptions, which in turn are based on ambiguous data and a high degree of uncertainty. As such, to make conclusions about such events would be imprudent without first addressing the uncertainties in the age-model as well a cautious reinterpretation of proxy signals.

Summary of *The Spatial Extent of 20th Century Warmth in the Context of the Past 1200 Years* by Timothy Osborn and Keith Briffa (2006)

In this article Osborn and Briffa review past work on proxy-based climate reconstruction in an attempt to assess if the claim that the late 20th century was the warmest period during the past millennium is supported. Whether or not this claim is supported depends on the comparison of recent instrumental temperature records with the earlier proxy-based temperature reconstructions. This comparison is only valid if it takes an account of the uncertainties associated with interpreting a specific reconstruction as an estimate of the actual temperature. Some of the reviewed studies do not provide a reconstruction of the entire millennia and some do not estimate the uncertainty in an appropriate manner assessing the significance of late 20th century warmth. Osborn and Briffa focus on three studies that meet the criteria of a formal quantitative comparison of late 20th century temperatures against reconstructed temperatures for the past millennia. These studies are Mann et al. 1999, Mann and Jones 2003, and Jones, Osborn and Briffa 2001. While all of these studies supported the claim of unprecedented temperatures in the 20th century and published uncertainties associated with proxy reconstructions above the 95% uncertainty range, the Intergovernmental Panel on Climate Change (IPCC) concluded that this claim could only be made with a 66 to 90% confidence because of un-quantifiable error that may arise from the proxies in the dataset. Osborn and Briffa conduct their own analysis of the proxy data of Mann et al. 1999 by smoothing the data and simply counting the fraction of records with values that exceed one or two standard deviations from the mean. The differences between pairs of these fractional exceedance time series (or the fraction of records at least one standard deviation above the mean minus the fraction of records with at least one standard deviation below the mean) were also analyzed. The highest positive deviations occur in the late 20th century, even far exceeding those of the mid-20th century.

The instrumental temperature results show a close correspondence with the proxy records, especially for the early 20th century increase and variations during 1930-1975. Additionally, the multi-decadal intervals support the concepts of the medieval warming period and Little Ice Age period. However, the dates of onset are vague and the analysis geographically restricted. The most conclusive finding is that the 20th century is the most anomalous interval in the entire period of analysis, including significant positive extremes in the proxy records.