BATTLING AMBIGUITY:
A PUBLIC GUIDE TO UNDERSTANDING THE SCIENCE BEHIND
THE GREAT ‘HOCKEY STICK’ DEBATE

A thesis submitted to the
Kent State University Honors College
in partial fulfillment of the requirements
for University Honors

by
Stephanie Pike

May, 2012
# TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................. vii  
ACKNOWLEDGEMENTS ....................................................................................................... viii  

CHAPTER  

INTRODUCTION ...................................................................................................................... 1  

I. RELYING ON SCIENCE ...................................................................................................... 11  
   The Scientific Method .............................................................................................................. 12  
   Peer Review .......................................................................................................................... 14  
   Science in the Public Mind .................................................................................................... 15  

II. MANN AND HIS HOCKEY STICK ...................................................................................... 19  
   How the Hockey Stick Was Born .......................................................................................... 20  
   What is Climate .................................................................................................................... 21  
   Tree Ring Data ..................................................................................................................... 22  
   Ice Core Data ....................................................................................................................... 24  
   Coral Band Data .................................................................................................................. 26  
   Beyond Data Collection ...................................................................................................... 28  
   Principal Component Analysis (PCA) .................................................................................. 28  
   Monte Carlo Analysis .......................................................................................................... 29  
   Where Things Fall Apart .................................................................................................... 30  
   MBH99: The Hockey Stick Paper ....................................................................................... 31
LIST OF FIGURES

FIGURE 1.  THE HOCKEY STICK GRAPH..................................................4
FIGURE 2.  TREE RING GROWTH SERIES..............................................23
FIGURE 3.  ICE CORE STRUCTURE...........................................................25
FIGURE 4.  CORAL BANDS.................................................................27
FIGURE 5.  CORRECTED TEMPERATURE RECONSTRUCTION.................34
FIGURE 6.  PC1 TESTED ON RED NOISE..............................................40
FIGURE 7.  HOCKEY STICK INDEX......................................................42
FIGURE 8.  THE CRU/HADLEY INSTRUMENTAL DATA SET.................62
# TABLE OF CONTENTS

LIST OF FIGURES .............................................................................................................. vii

ACKNOWLEDGEMENTS ..................................................................................................... viii

CHAPTER

INTRODUCTION .................................................................................................................. 1

I. RELYING ON SCIENCE ................................................................................................. 11

The Scientific Method ........................................................................................................ 12

Peer Review ....................................................................................................................... 14

Science in the Public Mind ............................................................................................... 15

II. MANN AND HIS HOCKEY STICK .................................................................................. 19

How the Hockey Stick Was Born ....................................................................................... 20

What is Climate .................................................................................................................. 21

Tree Ring Data ................................................................................................................... 22

Ice Core Data .................................................................................................................... 24

Coral Band Data ............................................................................................................... 26

Beyond Data Collection ................................................................................................... 28

Principal Component Analysis (PCA) .............................................................................. 28

Monte Carlo Analysis ...................................................................................................... 29

Where Things Fall Apart .................................................................................................. 30

MBH99: The Hockey Stick Paper ..................................................................................... 31
III. MCINTYRE AND MCKITRICK: THE BATTLE RAGES

What is Noise? 
White and Red Noise
Autoregressive Models
Eigenvalues and Eigenvectors
Where We Left Off
Bitterness

IV. WEGMAN’S FOLLY

V. CLIMATEGATE: EMAIL AND CONTROVERSY

VI. ARGUMENT FOR PROGRESS SAKE

The Great Debate
Where the Hockey Stick is Now
Common Ground
Ambiguity, Argument, and Aesthetic

WORKS CITED

APPENDIX

1. THE MBH 1999 PAPER
2. MCINTYRE AND MCKITRICK (2005)
3. MANN’S RESPONSE TO BARTON’S LETTER
LIST OF FIGURES

FIGURE 1. THE HOCKEY STICK GRAPH ......................................................... 4
FIGURE 2. TREE RING GROWTH SERIES .................................................. 23
FIGURE 3. ICE CORE STRUCTURE ................................................................. 25
FIGURE 4. CORAL BANDS ........................................................................ 27
FIGURE 5. CORRECTED TEMPERATURE RECONSTRUCTION ................... 34
FIGURE 6. PC1 TESTED ON RED NOISE ...................................................... 40
FIGURE 7. HOCKEY STICK INDEX ................................................................. 42
FIGURE 8. THE CRU/HADLEY INSTRUMENTAL DATA SET .................... 62
ACKNOWLEDGEMENTS

I would like to extend my utmost appreciation for my advisor, Dr. Sara Newman, for encouraging me throughout the process and never giving up on me no matter how many times I gave up. I would also like to thank each of my committee members, Dan Ross, Dr. Paul Sampson, and Dr. Natasha Levinson, for their time spent both reading and listening to me talk about this controversial topic.

Special thanks to my mother, for sending me articles on the topic, for coaching me through the roughest patches, and for making me a writer. To my father, thanks for trying to explain weather and climate, for arguing with me to get me to see another side, and for making me a scientist. Thanks to my sister and brothers for your encouragement and input--it meant the a lot to me.

I am grateful to Eric for pressing me to sit down and write when I needed to and, later, stopping me when it got to be too much. Thanks to Erika, Jennifer, Lesly, Angeli, Renee, and Kate, for politely nodding your heads when I talked about my thesis.

And, finally, thank you to my niece, Lilia, and nephews, Daltin and Lathan, for helping me remember to keep wondering about the world and keep hoping for the future.
been adequately and correctly represented and criticized by other scientists? And finally, how, if at all, has it fanned the flames of this touchy subject?

Climate is a scientific topic deserving of scientific discussion. That being said, you ought to know what is to be considered a valid scientific argument. I'm talking, yes, about the scientific method. Simply put, to have a valid argument, a scientist will hypothesize about something, let's say the temperature of a certain place over a specific period of time. He or she will then record temperatures for a specific length of time then write a paper and send it in to a journal publication—for example *Science* or *Nature*. Before it gets published, various other scientists

![The Hockey Stick Graph](image)

*Figure 1. The Hockey Stick Graph. Shown as it was featured in the 2001 IPCC Third Scientific Assessment Report*
read it to assess the validity of the argument. If none can falsify it or find no major inconsistencies, the paper gets published and then is used as the basis for other scientific hypotheses (Jarraud, Steiner, et al. 95). So, these peer reviewed articles are what constitute the scientific fact. However, it is important to note that these articles are subject to interpretation and can be misinterpreted as easily as the smile in Da Vinci’s Mona Lisa. It is in these misinterpretations that we find the controversy.

In the case of Mann’s hockey stick graph, there are many opposing and supporting responses. Picking and choosing a legitimate opposing argument seems daunting when a simple Google search brings up thousands of varying opinions on the issue. I used Stephen McIntyre and Ross McKitrick’s 2003 and 2005 papers, namely Hockey sticks, principal components, and spurious significance. This particular paper, published in the same scientific journal, Geophysical Research Letters, criticizes Mann et al.’s techniques and statistical analyses claiming that they isolated certain data to make their final results seem more drastic than they really are—leading to that hockey stick shape. This statistician-economist duo claim that Mann’s 1999 work holds little scientific merit in interpreting climatic data through misuse of statistical methodology, either with intentional or non-intentional means (McIntyre, and McKitrick 1).

For the purpose of this thesis, I assess this work of McIntyre and McKitrick to determine how they justified their claims. Do their results hold the same scientific merit as Mann’s? What have the effects been in the scientific
community and on the debate itself? Has their work held up? And finally, how has their work impacted the public’s view on climate change?

Continuing onward, I also analyze Edward Wegman’s 2006 *Ad hoc committee report on the ‘hockey stick’ global climate change reconstruction* wherein Wegman worked as a self-proclaimed independent evaluator of McIntyre and McKitrick’s arguments. Wegman served on the Committee on Applied and Theoretical Statistics for the US National Academy of Sciences and as such was trusted with this task. The report itself came to similar, if not the same, conclusions as the McIntyre and McKitrick paper by questioning Mann et al.’s statistical methodology in the construction of the hockey stick. It is important to note, however, that the report itself has recently been subject to claims of extreme plagiarism and erroneous citations--see chapter four. Despite this circumstance, I found it important to include it to illustrate the rifts that exist between the paleoclimatologists and statisticians as well as show just how sticky and muddled the debate has become especially in the popular mind.

To further illustrate the controversy and where we are at today with it, I look into the Climatic Research Unit’s (CRU) 2009 email leak, which has since been dubbed the Climategate scandal. The emails leaked from the CRU were of a somewhat questionable nature leading many to believe that the scientists who are pro-anthropogenic climate change are trying to hide something to prove their point. In 2010, the CRU was acquitted of these claims through a series of
investigations. However, there are many who still believe in the scandalous nature of the emails.

Climategate is such a recent occurrence that it is difficult to find hard evidence or documents to support either side of the issue. Most of the information and rebuttal has taken place via the internet on various blogging sites, which is hardly a place for proper scientific debate. For this reason, I have taken as much of an objective approach as is possible to assess the credibility of the arguments. Are the emails as menacing as they seem or have statements been taken so far out of context that they only seem scandalous?

This thesis serves to elucidate some of the ambiguity surrounding the debate on anthropogenic global climate change. I wanted to help put the foundational ideas behind the debate into a clearer perspective for all of us--the nonscientific community. After all is said and done I ask the questions: is there a common ground? Can there even be a common ground? Is it possible for all of us and the debate’s constituents to meet somewhere in the middle? I believe it may very well be possible but only once the arguments are understood and communicated in a more mature, directed manner. I believe that through this, our arguing, it may be possible to obtain a more clear conclusion. In the words of entomologist E.O. Wilson, “I may be wrong, you may be wrong. We both may be partly right” (Wilson 4).

Perhaps this debate will go on for centuries more--I hope not as this issue appears far too pressing to stagnate that long. Maybe coming to common ground
in which we know what is going on and how dire it is will allow us to finally work together to do what needs to get done.

In the following chapter, I discuss the scientific method. To understand how this debate has progressed, it is important to understand how information is processed to produce a valid scientific argument. It is also important to determine where things can go wrong and become misunderstood or miscommunicated--there are a variety of filters that ideas go through before arriving in the hands of the general public. The next chapter serves as the first step in understanding the underlying nature of scientific debate.

The chapter that follows discusses the specific methodology behind building the hockey stick graph. I discuss the specific data Mann collected and used. I then take a look at what major statistical methods were used to manipulate the raw data. Because the science and mathematics behind the graph are complicated, I have made every effort to clarify and simplify the meaning behind the hockey stick.

I then discuss McIntyre and McKitrick's research and the points at which they diverge from Mann's conclusions. In this chapter, I explain more in-depth information about paleoclimatological statistics and methodology in an increasing effort to clarify why the two sides still disagree. We come to find that it is in the minor details where issues arise.

Analysis of the Wegman Report follows the McIntyre-McKitrick chapter. As stated before, this flawed report is important to take a look at because it shows
how convoluted the debate has become. Wegman also makes startling points about the rifts between two different professional scientific groups.

Later, I explain the Climategate scandal. This event has played an important role in keeping the rifts between the scientists and statisticians well alive. And, although legally acquitted of guilt, the face of the Climatic Research Unit has been tainted making it seem like the scientists themselves no longer believe in true scientific method. The effects have been dizzying.

Finally, I end with the questions with which I began: will there ever be a common ground? Or, will this debate continue in its war-wreaking path? Will anything ever be solved? Are we doomed to perpetually argue the topic in a never-ending feedback loop? Should we take the risk to keep arguing? Ultimately, would it be more beneficial to be proactive as opposed to reactive? Would it not be smart to take necessary precautions, in the pro-anthropogenic climate change scientists are correct?

The idea of global climate change presents us with a seemingly dire situation. Environmental patterns will shift leading to more intense weather-related catastrophes--stronger hurricanes, stronger rains, longer droughts, severe heat waves, and countless more. The consequences could be immense if we do not deal with them, if we can still deal with them. We could be facing ecological disaster and economical meltdown. On the other hand, we may be just fine. It may be just an idea, blown out of proportion. The planet’s climates are in a constant, fluctuating state change. Nothing in the universe stagnates and our
planet is not an exception. Who are we to narcissistically believe that we have this much power over the environment? And finally, as many have so aptly put, if scientists can hardly predict the weather two weeks from now, how can they predict other weather patterns 50 years from now?
CHAPTER I
RELYING ON SCIENCE

My organic chemistry professor began the first day of class with the words “science does not prove anything.” In other words, essential nature of science is the ideas of process and progress. Nothing is solid. The world is constantly changing and everything, every idea progressively builds upon itself or another for years. This notion of perpetual improvement inhibits ideas from becoming permanent or stale. Despite the existence of a few various laws of nature, I still wonder if the law of gravity will still mean the same thing in the far, distant future.

It is important to note that “certainties are rare in science… you just never know” (Hoggan 16). Science is rooted in distrust, disbelief, and discontent. It is an ever-changing, ever-evolving liquid, so to speak, too fluid for exactitude but viscous enough to keep the materials all flowing together. Theoretically science, in its purest and unfiltered form, can be used to predict the future, rather future occurrences. Many errors arise when we try to communicate the observations and patterns. In order for us to understand and talk of the ideas we need to distill them through a series of filters—through the scientist, through their writing and speaking, then the editorial process, and finally through the mass media to communicate to us, the general, nonscientific public. Evolutionary biologist, Stephen Jay Gould said, “Science is rooted in creative interpretation” (Gould 74).
And, it is these creative interpretations that the seeds of uncertainty begin to sprout and grow.

Some uncertainty lies in the scientist. Human error. It happens. A lot, I’d say. Analogous to the processing of pure sugar cane, we lose quite a bit when we, as scientists, process pure science—no matter how much we try not to. Since “nature” is this major web of interconnecting aspects, we would be painfully overwhelmed if we were to take into account each miniscule individual factor involved in a pattern. The human brain can only handle so much at once before it shuts down and nothing gets accomplished. It is necessary then that we carefully pick what to specifically focus on. Herein lies the issue. Science is meant to be an objective enterprise but when we try to arrange and communicate the details, it becomes subjective. We, as scientists and citizen rhetors, decide what is important and what is not. We assign the values and the weight that various factors have over one another.

The Scientific Method

As mentioned in the introduction, there is a method to the madness of science. Nearly every student learns or is taught it in virtually every science-related class from elementary school onward. The scientific method can be described from something as simple as question, hypothesis, experiment, and conclusion. It can also be explained as something that is often so complex, tedious and/or specific that it could only be described using a weaving series of
interconnected flow-chart diagrams. To complicate the process even more, when strained through the professional peer-review process the final product is something that is incredibly sticky. In the great scheme of things, when filtered and funneled in a labyrinth of processes there seems to be much more madness to the method than originally, traditionally thought.

Nonetheless, I can try to describe it in a simple manner. The scientific method begins with an observation. From this observation, a scientist identifies a pattern. The scientist then speculates on how this pattern came about. This speculation is put forth as a falsifiable statement we call a hypothesis. The key is that it is falsifiable. For example, one might observe that when a sea vessel gets far enough out, it disappears—seemingly falling off the earth’s edge. One might then go on to say that the Earth is flat because of this phenomenon. Now, this can be falsified by creating an experiment to test the statement’s validity, or rather falsify it. The scientist could board a vessel and sail to what they believe to be the edge of the earth. If no edge is found, he keeps sailing and the hypothesis is falsified.

When the hypothesis is falsified, the good scientist will go back and rework the hypothesis to let’s say, the earth is round because you can sail straight in any direction and always arrive back to where you began. The hypothesis is then tested again and again, the hypothesis each time being revised, until it can no longer be falsified. It is an arduous process and quite possible for one to get caught going around and around, dizzily, in an unforgiving
loop. However, when the hypothesis fails to be falsified, the scientist rejoices, records the findings and the hypothesis becomes a theory.

In an effort to publicly share these cool, new findings, the scientist then writes a scholarly paper explaining this now polished theory addressing how it was formed and how it can be used to predict future occurrences of the previously observed pattern. Thus begins the peer-review process.

Peer Review

There are many facets to the peer review process, which make it seem much more complicated than it really is. The scientist first submits the paper to a publisher much like a writer submits a novel except the publisher involves a community of other scientific professionals who, in turn, read and comment on it. As the ideas percolate downward through the hierarchy of readers and editors, more and more of the original findings are edited and molded into what would look good and gratifying to the scientific community. Editors of the prestigious scientific journals must judge whether the scientist’s work and/or the scientist is valid enough and is of enough importance for publication. This process is based on the following questions: Who is a valid scientist? Who has the expertise? Who has proper credentials? Can a scientist outside his or her area of expertise have equal sway?

Ultimately, the editors and readers authorize whether or not the scientist's work is valid. Keep in mind that editors may have their own agenda for their
particular publication—whether driven by a specific subject matter or a personal bias. If the premise of the paper does not fit into the correct spectrum of publication, the scientist will find it extremely difficult to get it published.

If a scientific paper is not published, the scientist can either submit elsewhere, rewrite or begin a new project. Generally, theories are not accepted into the public domain until they have been published in these scientific journals in this manner. When published, though, these theories are accepted into the scientific community until they fail to accurately predict the future in the way they were designed to. This failure involves a competition among scientists—who conduct their own research on similar topics based on the scientific method and send it off to be reviewed and published through the peer-review process.

Put simply the peer review process works in this way, an article first goes through the editor who determines whether the article is relevant and whether it is of sound quality. It is then pushed through several scientists in a similar, related field—thus peer review. They decide whether the paper fits the standard for their particular publication (Sandewell).¹

Science in the Public Mind

When looking into anything controversial, one has to be careful about who is read and how their works are read. One must have a critical and skeptical eye

¹ The peer review process is shifting in detail although not in spirit. Because the internet has introduced a new era of technological publication in which the cost to print and distribute has hugely, if not completely, vanished. This could very well lead to increases in published articles correlated to a decline in intensive peer review.
in everything read. Everyone has an agenda—whether it be grant-based or out of a need for self-fulfillment. So, it is important to include many varying perspectives in attempt to understand the whole picture.

An issue that arises in the world of scientific exploration is the idea of expectation. For scientists, “Expectation is a powerful guide to action” (Gould 65). Scientists pour their blood and sweat into discovering the true nature of the universe around us. They spend many tedious hours researching and collecting data that they hope will depict and support their ideas. This is not only a physically, emotionally draining task but also a financially costly task. No one wants to spend years of their life to come up short, inconclusive, or worse yet, wrong. It is not conducive to our fast-paced, grant-driven scientific goals.

With that in mind, it is entirely possible for scientists to develop tunnel vision where they see only what they want to see with respect to their hypothesis. Gould discusses, in *The Mismeasure of Man*, how scientists once compared brain sizes to prove that the white man is smarter than men of color and women in general. This, we know of course, is not accurate. Scientists have long realized that head sizes are proportional to body size. And so, in essence, all humans have virtually the same physical properties, and brain size has little to do with intelligence and more to do with inherited genetic traits. However, scientists would go to great length to show how certain physical characteristics of the white man’s brain enabled him to be smarter than the black man (Gould 83-84). These accepted theories simply show that when you focus so intently and expect your
results to be a certain way, you can get what you expect to get. If you expect all women to be less intelligent than you, you can find ways to show reasons how all women are less intelligent than you

Most of us do not subscribe to scientific journals in which the articles are reviewed by official experts. We like to get our information in bite size snippets from newspapers. For one, it’s easier for us to wrap our heads around. When the non-scientist reads Mann’s 1999 paper, it is very difficult to understand and superficially convoluted unless you have a background in statistical methods in paleoclimatology. And, most of us do not have the necessary expertise to maneuver the technical jargon that seems commonplace in that particular sector of the scientific community. Is this a failure on the part of the scientific method? Are the conclusions of the scientific method meant to be so isolated from the public realm and understanding?

In my opinion, it is a possible fault of the scientific process that, in general, the ideas are presented in too complicated a manner for many to comprehend. It may be necessary for us to have knowledgeable experts who can objectively interpret what the scientists say and then communicate it in a way we can better understand while at the same time discouraging any possible misinterpretation. Without these experts to explain the nuances to us, we are left to fend for ourselves and better understanding becomes increasingly more difficult.

---

2 See Appendix 1 for full article
In the next chapter, I take an in-depth approach to Mann’s papers, namely the 1999 paper. Keeping chapter one in mind, I discuss in particular the method that the team used to develop the hockey stick graph. The key to their method lies in which data they collected and from where they collected it as well as the way they analyzed their data. I then discuss the impact of the hockey stick—where we see it used and whether or not it is still a piece of cited scientific data.
CHAPTER TWO

MANN AND HIS HOCKEY STICK

I began this thesis in search of the quintessential climate change article. For years, I’ve been bombarded with massive onslaughts of graphs and information regarding this topic. There was so much information. Little of it affected me. Even amidst heated classroom discussions, I was not moved by any particular slice of evidence. This lasted until a movie, *The Simpsons Movie*, parodied that hugely publicized hockey stick graph. I had seen this graph before. It was a key feature in Al Gore’s *An Inconvenient Truth*. From here, I had a jumping point. What was the hockey stick and what did it represent? How did it come about? And, what did it mean for us, the human race?

The hockey stick. Oh, the famed hockey stick. This widely used work of scientific “art” was published in 1999 by the journal, *Geophysical Research Letters*. Geophysicist Michael Mann, along with climatologists Raymond Bradley and Malcolm Hughes, developed this graph by measuring climatic fluctuations over the preceding 1000 years. They used evidence from tree rings, ice cores and a range of other physical proxies (Mann et al. 759). The graph itself depicts an anomaly in late 20th century temperatures in the northern hemisphere, which dramatically increase from the average, essentially creating the hockey stick shape. From that shape, Mann and his cohort concluded that the 1990s was the warmest decade, at the time, and 1998 was the warmest year of the last
millennium. They also concluded that “the 20th century warming counters a millennial-scale cooling trend which is consistent with long-term astronomical forcing” (Mann et al. 759). Basically, they mean that the warming in the 20th century does not fit into what we would expect to be normal, environmental variation. And, the team suggests that there has to be an outlying cause for this warming, which they believe is a direct result of humanity’s carbon-emitting activities.

After the publication of Mann and his team’s research and conclusions, many efforts to confirm or disprove the graph ensue. The backlash was fierce. But, before I cover the debate in subsequent chapters, let me first cover the all-important graph completely. It is hugely important to first understand the particular pieces of hard science because therein we are able to find the exact places where the debate seems to sprout.

**How the Hockey Stick Was Born**

To understand the arguments in support of and against this graph, we need to understand the techniques employed by Mann and his team. Yes, they used tree rings, ice cores, corals, etc, but why and to what persuasive purpose? Why did they chose these particular environmental facets? What about these climate indicators tells us about the climate hundreds of years ago? The following section acts simply as a user’s guide to acquiring basic understanding of the methods used in this particular article.
What is Climate?

When we think climate, typically, we think weather. Although slightly oversimplified, this works. Weather represents current conditions—think, oh, it is snowing outside right now or it will be sunny tomorrow. Climate can be considered a location’s average weather. Think Seattle—rainy. Or Chicago—windy. Climate is a combination of all the measurable facets of weather—namely including temperature, precipitation, wind speeds and directions, humidity, etc (Gardiner).

Advances in technology over the past century or so have allowed us to easily measure climate data. In this case, think thermometers, barometers, and that whole entourage of other neat scientific equipment. Devices like these have simplified calculating climatic conditions and since their invention we have been able to keep arduous track of climate patterns. Please note that, in retrospect, this complex equipment has only been around for a short time frame—about 100 to 150 years (Gardiner). Keep in mind too that, as the technology advances, we continuously become better fitted to more accurately and precisely record information. Because technology had not existed long before we had the means to measure temperature and wind speed, comparing current conditions to those in the distant past, let’s say, around the 1400s proves extremely problematic. To do this, though, scientists use a variety of indicators. These indicators, called proxies, include data from tree rings, ice cores, and corals. The following sections
explain how each of these can be used to develop a network of data to numerically estimate long-past climates.

Tree Ring Data

Most of us have seen the rings of a tree. Chances are, if you’ve seen a stump, you’ve seen the series of concentric rings. These rings separate years of growth as every year during the growing season, there is a new ring of growth between the older wood and the bark ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). These rings are typically banded into lighter and darker colors. Each light band and dark band represents the seasonal effects in a single year. Figure 2 provides an example of how tree ages can be determined by the various banding in the tree rings.

The width of these tree bands is affected by a wide variety of factors which include “the tree species, tree age, stored carbohydrates…nutrients in the soil, and climatic factors including sunlight, precipitation, temperature, wind speed, humidity, and namely the carbon dioxide availability in the atmosphere” (Wegman et al. 13). Width is a great indicator in determining “good” or “bad” growing seasons depending on the tree species ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). A “good” growing season correlates to thicker bands and gives a clear indication of how much carbon dioxide was in the atmosphere. Remember, carbon dioxide is one of the primary
drivers of photosynthesis. Photosynthesis, as we have learned in high school biology class, is what makes plants grow.

Another good climate indicator is tree ring density. As in the case of tree ring widths, when the density is larger, it usually indicates that the environmental conditions were more favorable and the growing season was good--there was an abundance of sunlight and carbon dioxide.

In determining ancient climates from tree rings, Mann's team used a large portion of samples from bristlecone pine trees. These trees are some of the oldest, living trees on the planet and, even when they do die, they can remain
incredibly intact over hundreds to thousands of years ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). The team relied heavily on tree ring data from these particular trees because their age allowed them to gather information about what the climate might have been like 1000 years ago.

**Ice Core Data**

Picture snow floating gracefully toward land. As the snow falls it intermixes with the surrounding isotopes in the air—for example carbon dioxide or ozone. As this snow hits the ground, it traps these bubbles of atmospheric gas. This snow layer is soon covered in a new layer and then another and another and so on. As the layers beneath the newly fallen snow compress, they slowly turn to firn, a type of slushy snow, and then later to ice (Wegman et al. 14). Now, in warmer years, the snow melts and the layers thin out.

Scientists carefully drill into the these huge ice sheets, namely glaciers, to obtain their specimens. They must carefully preserve these ice cores to prevent contamination as any exposure to outlying substances, like gases, can derail their analyses ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). Ice cores are fragile and must be immediately cold-stored to preserve structure. Even a slightly melted ice core can disrupt the results and findings.
Each band in the ice cores provides information about past temperatures, rainfall, carbon dioxide levels and other gases in the atmosphere, and solar variability and volcanic activity (Riebeek). Even as the ice becomes more compressed, the layers are still preserved. Figure 3 shows a three depths of ice cores maintain their annual bands. Even as more sand and silt intermix with the ice as depth increases, scientists can still make inferences about what the ancient climates had been like through physical and chemical analysis of these ice bands.

![Figure 3. Ice Core Structure. Banded ice cores differ in physical structure as, over time, they are compressed by the overlying layers (Riebeek).](image)

Ice cores are predominantly taken from places like Antarctica, Greenland or high atop mountains where glaciers and hard snow packs are formed ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). Mann's team used several ice cores samples from Quelccaya in Peru, which is one of the largest glaciated regions in the tropics. They also used ice cores from Greenland. The ice cores used only provide information from only 488
to 553 years ago; compared to the data from tree rings, ice cores are not as useful in reconstructing data over the preceding millennium (Mann et al. 759). This is likely why the team relied more on tree ring data in their principal component analysis, which I will talk about later in this chapter.

**Coral Band Data**

I will only discuss this briefly as corals provide the least amount of weight in the final analyses by the Mann team. Corals have a symbiotic relationship with various species of surrounding algae. The longevity of these algae are affected by the depth of the water, currents, water clarity and composition (Wegman et al. 15). Since the algae provide the coral with a source of carbohydrates, their growth is integral to coral growth (Wegman et al. 15).

When the endosymbiotic algae die off, the coral polyps are soon to follow suit. They leave behind these layers of calcium carbonate. Upon these deposits, new symbiotic algae and polyps form and grow until eventually they, themselves, die. This process is repeated essentially creating a series of layered coral bands ("Global Change Project: Paleontological Research Institution and its Museum of the Earth"). This coral banding is depicted in Figure 4. Like the tree rings, you can determine what the growing season was like by the bands. In corals, clearer and more distinct bands depict a more hardy growing season (Cobb et al. 1). In periods of increased environmental and climatic stress, the algae will expend
less energy on growth and more on survival. This can be clearly seen by bands of slower and more blurry growth—as seen in Figure 4.

![Figure 4. Coral Bands. Coral is a decent indicator of the environmental and climatic conditions were like as the polyps died off (Shatto).](image)

**Beyond data collection**

You should now be familiar with the types of mediums the team used to obtain their raw data for reconstruction.¹ But where do we go from here? We have a massive pile of information—what do we do with it? At this point to further

---

¹ The data for each of these proxies is available online via the World Data Center-A for Paleoclimatology at [http://www.ngdc.noaa.gov/paleo/paleo.html](http://www.ngdc.noaa.gov/paleo/paleo.html).
our understanding, we need to familiarize ourselves with what they did with this raw data to develop the hockey stick. The next section explains the methods used to analyze the climate data. From here, things become a little more complicated.

**Principal Component Analysis (PCA)**

This is by far the most crucial, complicated and controversial aspect of Mann's article. The aim of PCA is to reduce “the dimension of a high dimensional set while preserving most of the information in those data” (Wegman et al. 15). Think about it. We have this huge mass of seemingly unconnected data. Tree rings and ice cores are vastly different from each other. The basic goal of PCA is to shrink down this large unconnected data set into something more manageable and then to identify any new underlying variables. Instead of tediously including all of the data, only the most common, correlated and definitive patterns within the data sets are used to describe a certain percentage of variation over the whole record (Schmidt, and Amman).

The outcome of this process is a set of uncorrelated values or patterns, called principal components (PCs). These PCs are meant to describe the entire data set and each one describes it in a slightly different way. Typically, only the first few principal components hold much mathematical weight and account for the most variance.
Please note that principal components, in themselves, do not tell us anything significant about the physics of the situation itself. They are not meant to work in isolation and it is only when they have been combined and redistributed over a time series, essentially plotted onto a line graph, that they mean anything (Schmidt, and Amman).

**Monte Carlo Analysis**

In Mann's methodology, the team used 'Monte Carlo' analysis to construct their graph. This analysis is done in concordance with PCA and is key to obtaining that hockey stick shape (McIntyre, and McKitrick 1). Using Monte Carlo analysis means that the team relied a certain type of random sampling in their statistical analysis. Picking data at random, helps prevents scientists from inadvertently choosing particular data and maintaining that all-important, objective air.

Both the PCs and the results from the Monte Carlo analysis can be plotted against each other--the PCs represent results from real data whereas the Monte Carlo analysis results from the random data (Schmidt, and Amman). These are both plotted against a noise line, described in the next chapter, to test just how well the first PC describes the actual data.

Keep in mind that this is incredibly oversimplified and rather than explaining the complicated reasoning behind this set of statistical methodology, I
will leave it at that. Without years of studying statistics, it is virtually impossible to grasp for our non-statistical driven minds.

**Where Things Fall Apart**

So we know now what principal components are, where they come from and what they do—so far so good. From where, then, does the controversy arise? The answer is painfully transparent. The issue lies partially in determining what constitutes the principal components and how the data is centered within the PCA—“i.e. what time interval is used to define the ‘zero’ baseline for the data series” (“RealClimate: Climate science from climate scientists”). The Mann team centered their PCs over the mean period between 1902 and 1980. It is not centered over the entire period—from 1400 AD onward. Why would they not center the data over the entire time period? This centering oddity presents an issue I discuss in the following chapter.

The ones who do the analyses are, essentially, the ones who decide what information holds the most weight and what information should be used to describe the entire dataset. Assigning mathematical weight, however, is a scientifically subjective task much like making a very educated guess. Your results rest entirely upon which educated guess you make and what holds the most value. With a few minor tweaks, perhaps centering your data over a different time period, may cause your results may be entirely different.
Another issue is the statistical method you choose to use. There is are quite a few techniques you can use to manipulate your data. Scientists may choose or develop the best mathematical way to translate the data. This translation issue also leads to controversy simply because there are so many ways to approach data manipulations. One scientist's idea of the best method may differ immensely from another's idea.

**MBH99: The Hockey Stick Paper**

In Mann’s methodology, there was a total of 112 proxies used to indicate climate. Of these, only twelve give information that predates AD 1400. Within these twelve proxy indicators lie the three main principal components. Remember that the first few PCs hold the majority of the weight and account for the most variance. These three PCs are based off of tree ring widths in North America, namely tree cores of bristlecone pines (Mann et al. 759). These North American tree rings hold the most mathematical weight as they provide more information over a longer timeframe. From here, the methods become blurry and difficult to understand but please bare with me. Mann needed to calibrate the data. This involves choosing a mean for which to center the data in an effort to not only standardize the data but also to use it as a basis for comparison. Mann calibrated the 1902-1980 period to have a mean of zero. He did this because he wanted to know if the tree ring data was significantly different from the
information provided by the twentieth-century’s temperature record (Schmidt, and Amman).

We know now how the hockey stick team obtained their data from tree rings, ice cores, and coral reefs--these proxies give us information about what the climate was like hundreds to thousands of years ago. The team then used this data, giving more weight to the information provided by bristlecone pine tree rings, and plotted them over a time series. This product is a graph and it, in this case, is in the shape of a hockey stick. This particular shape indicates that temperatures in the late-twentieth century do not fit into the random climate cycling of the previous 1000 years. They have increased. They have increased dramatically. By comparison, they are anomalous. The team attributes the anomaly to human, carbon dioxide-producing activity. After all, the twentieth century marks that big, industrial boom corresponds to higher carbon dioxide emissions.

In the next chapter, I look at the main scientific naysayers, Stephen McIntyre and Ross McKitrick. These statisticians have crusaded now for over a decade to show how flawed they found Mann’s research and inferences to be. Time and time again they reiterate that they fully disagree. They do not agree with the way the hockey stick was constructed. They do not agree with the idea of the inferences made about our, human connection to climate change. The following chapter discusses the methods they used to argue against the stick’s validity and then explains their main contentions with Mann methodology.
CHAPTER THREE

MCINTYRE AND MCKITRICK: THE BATTLE RAGES

Stephen McIntyre and Ross McKitrick are a Canadian statistician-economist duo who have fought Mann tooth and nail for over a decade. Well, maybe not fought, physically speaking, but they have heartily disagreed and spent a great deal of energy trying to disprove the MBH98/99 papers.

In 2003, years after Mann released the hockey stick graph, McIntyre and McKitrick (MM) published their paper *Corrections to the Mann et. al. (1998) Proxy Data Base and Northern Hemispheric Average Temperature Series* in journal, *Energy & Environment*. The two made the claim that the hockey stick was “primarily an artefact of poor data handling, obsolete data and incorrect calculation of principal components” (McIntyre, and McKitrick 751). They go on to list the numerous, specific problems they found with the MBH98/99 papers--namely errors in the statistical methodology, which include truncation errors along with a disregard for the “Medieval Warming Period” and “Little Ice Age.”

Moreover, in a later 2005 paper, the two made the claim that Mann and his team had “carried out an unusual data transformation which strongly affects the resulting PCs” (McIntyre, and McKitrick 1). They developed their own temperature reconstruction graph, changing a particular mathematical convention related to PCA, and compared it against that of Mann’s hockey stick. Their result is shown in Figure 5.
This temperature reconstruction shows that McIntyre and McKitrick, using similar methodology, obtained similar results to that of Mann and his hockey stick--except for a divergence in the 15th century which can be said to represent the “Medieval Warming Period.” This divergence is a result of altering the PCs and final analysis. More specifically, their reconstruction can be duplicated simply by removing PC4, which is also known as the ‘St Anne River’ Northern treeline series (Schmidt, and Amman).

In their 2005 paper, *Hockey sticks, principal components, and spurious significance*, McIntyre and McKitrick stated that, “their [Mann et al.’s] method, when tested on persistent red noise, nearly always produces a hockey stick

![Corrected temperature reconstruction. McIntyre and McKitrick used Mann’s methods to develop their own reconstruction (2003).](image-url)
shaped first principal component (PC1) and overstates the first eigenvalue” (McIntyre, and McKitrick 1). Now, the majority most of us are not quite certain what this means. We know the problem lies in something related to Mann’s statistical methodology. But, we do not know specifically where the issues lie. For this, we need to have a better grasp on the specific statistical methodology which involves understanding the key mathematical terms that are used by McIntyre, McKitrick and Mann. After all, we cannot even begin to understand what they are trying to get at if we have no background to their statistical jargon. The following section is a continuation of the guide I began in the previous chapter, taking on a more mathematical perspective to better explain McIntyre and McKitrick’s arguments.

**What is Noise?**

Noise is defined to be “unknown external factors or fluctuations in signals” (Wegman et al. 15). This definition means that noise represents all the unpredictable facets of what you are looking at. With regards to climatology, noise is that little hole in your ability to predict what the climate will be like, let’s say, in a given season. It is a sum of all the little minor factors that get in the way of us accomplishing our scientific goal—prediction. We would call it chaos. Scientists call it noise. This noise hugely limits our ability to predict what the climate will be like (Strauss 1). With respect to paleoclimatology, noise is all the

---

¹ See appendix 2 for full article
little factors that affect our proxy data. In other words, this noise refers to the unpredictable and random variations in growth of tree rings, ice cores, and coral bands (Wegman et al. 16).

Noise is hugely important in that it allows us to account for what we do not quite know yet or are not exactly certain of. That sounds strange, yes. How can we account for what we do not know? Think about it this way: I know that there will be a new episode of *The Simpsons* on Sunday night. However, I am not sure if the Cleveland Browns game will go into overtime and interrupt the show. If I assume that this football game will carry over, because it sometimes does, then I can account for this by not even bothering to try watching the new *Simpsons*. Or, perhaps I can simply delay the time between switching channels.

The Browns game would be considered noise in this case and accounting for it helps me decide what I want to do with my data, or time. I may not know exactly how long it will carry over or if it will even carry over at all but by taking it into account I can play it safe and not have to be subjected to watching the game.

Noise can be classified in terms of colors--namely red and white in the case of climatology. These different colors, or labels, tell us something specific about the particular unpredictable facet. In defining them, we are able to assign them specific mathematical weights which can later be used in calculation--remember, it helps to account for what we do not know.
**White and Red Noise**

Applied to tree rings, ice cores, and coral, white noise represents an even density that is spread equally across the entire spectrum (Wegman et al. 16). Think of that sound on the television when you have no signal and there’s that even, crunchy sound—that’s white noise. The sound, you know, is uneven yet even. You cannot set a particular piece apart from the rest—a sort of pulsating evenness. Now take this thought and place over the mathematical spectrum. White noise is pulsating but averages out to the roughly the same thing—equal energy per frequency band which gives us an average baseline of zero. In the mathematical context, integrating white noise will produce red noise (Wegman et al. 16).

Red noise depends on “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” (Wegman et al. 16). In other words, a good year of growth will beneficially influence the following year of growth. Time is not neatly divided, rather a continuous progression. For this, consecutive years are not necessarily self-isolated from each other. The preceding year has a huge influence over the following year in terms of growth. You have to account for this influence that one has over the other.

When you add all the noise together and assign each appropriate values, you must then send them through a model to graphically show how they might impact your raw data. In Mann’s methodology, they used the previously
described red noise and modeled it through an autoregressive model. They also used the white noise and modeled it with a moving average model.

**Autoregressive Models**

An autoregressive model (AR) uses specific formulas, based on previous outputs in attempt to predict an output of the system ("Curricular Linux Environment at Rice (CLEAR)"). Essentially, the data you obtain from these models is based solely on what the previous outputs data of the model. Models that are based off of your inputs are called moving average models (MA) and can be used in conjunction with AR models to create autoregressive-moving-average models (ARMA) ("Curricular Linux Environment at Rice (CLEAR)"). Mathematically speaking, the AR model has only poles and the MA model, only zeros.

Now, you may be wondering what this all means. Most people do. It is incredibly complicated. To put it in the simplest terms, scientists use the red noise for the AR model, and use the white noise for the MA model--note that the average white noise is equal to zero, as that is what makes it white noise. The two models are combined together (ARMA) and the results are then plotted over a time series which can then used for its predictive power. This model can then be superimposed over the raw data to see how they compare.

Think back to the previous chapter where the PCs and results from the Monte Carlo analysis were plotted against a noise line--the goal being to validate
the PCs. The ARMA model is the technique they used to compare the aforementioned data.

**Eigenvalues and Eigenvectors**

Eigenvalues and eigenvectors are used when doing linear algebra. They are simply outputs of matrix calculations. An eigenvalue is a singular output of the matrix calculations whereas the eigenvector is essentially a set of eigenvalues.

Both the eigenvalues and eigenvectors are solved for in the PCA process. As such, the eigenvector associated with the highest eigenvalue follows direction of the first PC. The eigenvector with the second highest eigenvalue follows the second PC--so on and so forth (Boersma, and Weenink).

**What We Left Off**

At this point, you may have forgotten the point of all this. Think back to what McIntyre and McKitrick were stating before about Mann’s methodology always creating a hockey stick shape. What they were trying to get at was that Mann overemphasized the first eigenvalue, or PC1. When they superimposed this data over the red noise, they obtained the upward trend. With Mann’s methodology and algorithm, this overemphasis will always select for the upward trend, that hockey stick shape, no matter what data you input into the statistical
program. The basis for this argument is illustrated in Figure 6. McIntyre and McKitrick used twelve PC1s samples generated by a red noise model.

Figure 6. PC1 tested on red noise. McIntyre and McKitrick pointed out that PC1 always leads to the hockey stick shape (2005).

In hand with the methodological disagreement, the two found problems the proxy data itself--namely with the bristlecone pine tree data which dominate the PC1 (McIntyre, and McKitrick 69). Bristlecone pines tend to grow in the same regions and environments. More importantly, it has been noted that there is an “undoubted growth pulse for a small network of bristlecone pines in the Western USA in the late 19th and 20th centuries, and their ring width “chronologies” thus have a hockey stick shape” (McIntyre, and McKitrick 4). In other words, growth in
the bristlecone pine trees in the 1900s does not correlate to the instrumental temperature record of the same timeframe. For this, bristlecone pine trees are not a reliable temperature proxy (McIntyre, and McKitrick 69).

The strange phenomenon and lack of data correlation is one of the main points of contention for McIntyre and McKitrick. The two infer that Mann's reconstruction in the 19th and 20th centuries does not reflect temperature, rather, it is more associated with this growth pulse in bristlecone pine trees (McIntyre, and McKitrick 4). This, they believe, is another reason we may see that steep incline in the 20th century.

Another key issue, the two noted, was in the way Mann calibrated his data. Remember, Mann calibrated his data to have a mean over the period between 1902 and 1980. McIntyre and McKitrick disagreed with Mann's reasoning for calibrating the data this way and claimed that this "data transformation results in the PC algorithm mining the data for hockey stick patterns" (McIntyre, and McKitrick 1). What they meant is that Mann's statistical techniques are intentionally aimed at obtaining that hockey stick shape.

The histogram in Figure 7 compares how Mann centered his data against how data is conventionally centered in normal statistical applications. As you can see, the two methods are mathematical opposites. The two contend that the hockey stick's sharp, upward inflection, primarily begins at the start of the 1902-1980 calibration period (McIntyre, and McKitrick 2). Hence, the hockey stick shape is selected for.
At the very root, McIntyre and McKitrick identified the main problems to be related to Mann’s PCA analysis. Moreover, the constantly repeated theme seems to be that Mann’s methodology always selects for the hockey stick trend. For one, the PC1 is overemphasized. This is likely due to the PC1 being dominated by an unreliable temperature proxy indicator, bristlecone pine trees, which exhibit growth trends that do not correlate with temperature data. Their 20th century growth pulse pushes the trend upward even when tested on a red noise model. Additionally, the calibration period is questionable. Calibrating the data in this way causes the PC algorithm to, once again, select for the hockey stick shape.
Bitterness

What we have seen thus far is good. All the crucial elements for proper scientific process are evident. At this point, the scientists are doing what they are supposed to be doing--research, writing, and checking each other’s work for errors. From here on out, however, we begin to get a sense of the distaste brewing between both sides.

In a later 2005 paper, published in *Energy & Environment*, the McIntyre and McKitrick argue that Mann’s published methodology was not the same as the one used to obtain the results (McIntyre and McKitrick 69). The pair also delve into the hardships they endured while trying to obtain information from Mann and other climatologists in attempt to replicate their conclusions. Lastly, they discuss limitations and criticize faults they found in the peer review process.

Indeed, McIntyre and McKitrick had a difficult time obtaining the complete information they needed to fully reproduce Mann’s hockey stick graph. Whether or not this was because the information was unpublished, unavailable, or whatever, the two experienced a lack of cooperation from the paleoclimatological community. Keep in mind that, over time, the rifts between these groups of scientists has grown immensely and there is quite a bit of bitterness wafting back and forth between the two. Albeit on petty grounds, even professionals may refuse to lend a hand to those working to discredit them.

Contrary to the difficulty encountered by McIntyre and McKitrick, other scientists have been able replicate Mann’s data (Mann 4). Mann claims that all
the data necessary is available online or upon request. In response to a 2005 confrontational letter from Senator Joe Barton, which demanded full data documentation as well a list of all who fund Mann’s research, Mann attached a detailed list of various accessible websites where the data is located. There is something awry in McIntyre and McKitrick’s inability to obtain the “necessary” data.

McIntyre and McKitrick felt that the peer review process had let them down. In their mind, they could not figure out how these peer-reviewed journals did not discover the same statistical and methodological errors that they themselves did in Mann’s research (McIntyre, and McKitrick 94). A key point to note, however, is that Energy & Environment, the journal in which McIntyre and McKitrick were published numerous times, is not a peer reviewed journal. In fact, it has been known to have a certain bias against the idea of anthropogenic climate change (Mann 7). How can McIntyre and McKitrick contend that MBH99 was improperly peer reviewed when they themselves were not even peer reviewed?

These trivial arguments have stubbornly persisted throughout the past decade. Neither side is budging. The growing gap between the climate scientists and the statisticians leads us to question whether or not we will ever reach a clear consensus on the topic. If not a consensus, then, at the very least some

---

2 See Mann’s letter to Senator Joe Barton in appendix 3
overlap. If neither side is willing to give a little, though, we cannot expect progress to happen anytime soon.

The next chapter takes a look at the 2006 Wegman Report. This report evaluated the claims of both McIntyre and McKitrick with respect to Mann and his team. This report was intended to scientifically and independently assess the criticisms put forth by McIntyre and McKitrick about the hockey stick graph. Although the report had huge flaws of its own, it offers an interesting glimpse into the various ways members of the scientific community have approached and reacted to the debate.
CHAPTER FOUR
WEGMAN’S FOLLY

We have now been through a painstaking process of understanding some of the hard science behind the debate. We have looked at the first-hand arguments of the yea-sayers and those of the nay-sayers. Now, we will gaze into the broader spectrum, a more familiar place where the facts have become more distorted and convoluted. We need to take look at the reactions of other scientists in attempt to see how far the arms of the debate have stretched and where we are at with it now. There is no better place to start than with Edward Wegman.

Edward Wegman was a well respected statistician in the scientific world. He serve as chairman of the Committee on Applied and Theoretical Statistics for the National Academy of Sciences as well as president of the International Association of Statistical Computing. In 2006, Wegman worked with an ad hoc committee to investigate and verify McIntyre and McKitrick’s claims against the MBH99/98 papers. The report, Ad Hoc Committee Report on the ‘Hockey Stick’ Global Climate Reconstruction or simply the Wegman Report, claimed to be an independent, pro bono work in which no compensation or financial interest was a goal in the outcome (Wegman et al. 2).

Wegman came to several conclusions after assessing Mann’s research and analysis. These conclusions are almost identical to those of McIntyre and
McKitrick. Wegman disagreed with Mann’s calibration and other statistical procedures. Wegman found it difficult to access to the complete raw data and computer code necessary to replicate Mann’s data. And finally, Wegman goes on to list issues in the peer review process (Wegman et al. 3).

Wegman found the same statistical issues as McIntyre and McKitrick, which essentially validated their arguments. He agreed that the data had been incorrectly centered and that calibration point was invalid and not statistically sound (Wegman et al. 4). He also agreed that the data from bristlecone pine trees are not a reliable proxy, for the same reason, and then goes on to pontificate on why the hockey stick mistakenly leaves out the “Medieval Warming Period” and “Little Ice Age” (Wegman et al. 49).

The Wegman Report also discussed problems with obtaining Mann’s original data (Wegman et al. 4). Initially, Wegman could not reproduce the results because Mann supposedly did not release all of the data and corresponding computer codes—I say supposedly because Mann claims that all his data needed for reconstruction are available online.

Wegman received assistance from Senator Joe Barton who, as mentioned in the previous chapter, wrote a letter inquiring Mann about his research. Interestingly enough, Wegman claims that Mann’s response to Barton’s letter also had a confrontational tone (Wegman et al. 7). Regardless, Mann made his data available in his response to Senator Barton. How could it still be that Wegman could not reproduce Mann’s results?
Scholarly papers are archived as per scientific record however, supplementary material is not required to be a part of this archive. For this reason, it is often not documented in a clear and organized manner (Wegman et al. 51). In Mann’s case, the supplementary data is considered to be his intellectual property. Legally, he is not obligated to release any of his supplementary data or computer code (Mann 6). Superficially, this circumstance impedes the scientific process. However, Mann explains that his computer programs are “irrelevant to whether our results can be reproduced. And whether I make my computer programs publicly available or not is a decision that is mine alone to make” (Mann 5). If all the data needed to make the results replicable are available, Mann has no obligation to loose his intellectual property.

There are likely many reasons why Mann did not want to release his complete data and computer code directly, which probably include the fact that the naysayers of global warming have repeatedly requested this information (Mann 6). After a decade of being hounded by these people who want to prove you wrong, any human would start being a little territorial over his or her work. Then again, making this data only available to people who are your side is like preaching to a choir--it only increases the rifts.

Another problem, Wegman thought, was that Mann and his team did not work as independently as most scientists are generally expected to do. He made the claim that they had worked too deeply within their own social network (Freese). This is not surprising. Think about it--how many people are going to go
into the paleoclimatological field or something related? Not a whole lot. For this reason, the research you do is likely going to be in concert or based on people you may personally know if you have established yourself in this particular community. The problem with this occurs when scientists work too closely with other like-minded scientists. This is too much of the same thing. Think of the magnifying glass effect where some details are clear and focused while the others are left out, blurred because you were so fixated on the magnifying glass. There can be a lot lost or unaccounted without a new perspective.

Wegman feels the paleoclimatological community is too tight knit. The peer review was done by peers with too close a connection to Mann and his team (Freese). The statistical community, seemingly, was left out and not incorporated enough even though the scientists relied so heavily on statistical methodology (Wegman et al. 4). To an extent, Wegman is spot on with this claim. In the MBH99 paper, the team cites their own previous research and co-authored papers in seven of the sixteen references. None of the authors or co-authors cited by MBH99 were, in fact, statisticians. And, a majority, if not all, were some type of climatologist, physicist, or dendrologist.

Contrarily, most scientists do this. An invertebrate zoologist will be peer reviewed by other invertebrate zoologists--after all, they are the experts on the subject. The problem, in Wegman’s mind, was that Mann’s results depended solely on statistical processes. There is no way of knowing exactly what the temperature and climate was like 1000 years ago. Mann had to use statistics to
analyze his proxies but he did not substantially obtain input from professional statisticians outside the paleoclimatological community, who could very well have come up with better, alternative methods.

There was a flaw in Wegman’s logic here, and it became apparent at a 2006 hearing on the report. Wegman had accused Mann, his team and the paleoclimatological community of being too closely knit and not having included professional statisticians. On the other side, members of Wegman’s team were not paleoclimatologists, they were statisticians. Lo and behold, Wegman and his team were guilty of the same thing. He claimed that this does not mean that he did not speak with paleoclimatologists on the subject (Freese). He himself had asked a group of people he knew, climatologists, to do the peer review of the report. He is guilty of the same thing. He, himself, had stayed in his comfortable, cohort bubble to compose his report. Now, this does not mean that Mann is off the hook. New perspective could prove beneficial. Even so, it is difficult to criticize someone for something you yourself have done--this is hypocrisy and is simply petty in the realm of scientific debate.

Before I go any further discussing what was contained in Wegman’s report, I must mention a major issue that came to light in 2008. Just two years after the report’s publication, Edward Wegman was accused of plagiarizing large portions of his report. This shocked the public and science world together. Plagiarism is unspeakable for a man of his position--a crime in the scientific world. Not only was he accused of plagiarism, but he was accused of plagiarizing
from articles by none other than Raymond Bradley himself, remember a co-author in the MBH99 paper. As if this was not a big enough shock, Wegman was also accused of including erroneous citations as well as manipulated phrases from his sources (Vergano). In other words, he took information out of context or, simply, just made it up to cover his tracks.

In 2010, the accusations against Wegman were confirmed. Computer scientist John Mashey, author of the DeepClimate blog, had found that 35 pages of the 91 page report had been directly plagiarized from outside sources including not only Bradley’s work but also information off of internet sites like Wikipedia (“HuffPost College: The Internet Newspaper: News Blogs Video Community”). Mashey wrote a lengthy 250-page report on Wegman’s scientific misconduct. We, the public, were left not knowing what to think.

Naturally, Wegman denied the allegations. He claimed that the plagiarism, if indeed there was any, was committed by students who had aided in the work and therefore unknowingly published the flawed report (Vergano). Regardless of awareness, Wegman ought to have at least proofread the final copy. Plagiarism is directly below lying and is hugely eschewed by the scientific community—the consequences will be immense. Wegman’s reputation is tainted. He will likely lose his professional position and never publish a report again.

This misconduct not only presents a problem for Wegman’s arguments but it also leads to wonder where the peer reviewers were when they read this. How could such evident plagiarism pass under so many ever watchful eyes? Why did
it take four years for it to become apparent? And, do his results, though plagiarized, hold any weight whatsoever?

Wegman plagiarized mass sections from his report. I might venture to say that because he himself did not have a thorough understanding of the climatological perspective. Maybe he tried to reword the information from others, albeit poorly reworded, in an effort to make it seem like he understood. Who knows? Regardless, is his statistical argument valid? I would venture to say his contentions matched those of McIntyre and McKitrick. However, Wegman’s report will likely never be used as a reliable source. It may be safe to say his argument is invalid, at least, in the scientific realm.

As for the work of McIntyre and McKitrick, Wegman and his team were able to reproduce their results (Wegman et al. 48). Wegman also makes a comment about McIntyre and McKitrick saying that “they were attempting to draw attention to the deficiencies of the MBH98-type methodologies and were not trying to do paleoclimatic temperature reconstructions” (Wegman et al. 48). As statisticians, they were testing the method not the temperature reconstructions themselves, at least in Wegman’s mind. In McIntyre and McKitrick’s mind, they were trying to show that there is no statistically significant information about how modern climates can be compared to those of the 15th century using the hockey stick graph (McIntyre, and McKitrick 6). The goal was to disprove, not improve.

One very compelling point that Wegman makes, regardless of the plagiarism, is that the debate over the hockey stick is mainly taking place via
internet--on competing web blogs, namely Realclimate, run by the “climate scientists,” and Climateaudit, run by Steve McIntyre, and both are powered by Wordpress (Wegman et al. 49). To have this debate over such a medium is completely inappropriate. As Wegman so aptly puts it “the blogs give credence to the fact that these global warming issues are[sic] have migrated from the realm of rational scientific discourse,” and “unfortunately, the factions involved have become highly and passionately polarized” (Wegman et al. 49). Arguing via blog is simply unprofessional and incredibly juvenile and is contributing to the rapidly growing gap. This is the bad face of the scientific community. If we are unable have a discussion as adults how can we expect to get passed where we currently stand? Sitting on the computerized sidelines typing “foul” during every play by the opposing team will not get us anywhere.

Wegman brought forth some very interesting points about the lack of collaboration and communication between two differing groups of professionals. If we can take anything from his flawed report, we can take it that the current system is not working as well as it should. The lack of collaboration is preventing us from obtaining a full picture of what is occurring in our atmosphere. The lack of proper communication is making it difficult, if not impossible, to even discuss the topic at hand. There can be little or no scientific progress if these issues are not addressed in a timely and more professional manner. If we do not fix the barriers, we may forever be stuck immobile in a never-ending, unforgiving loop, essentially taking us nowhere.
In the next chapter, the seriousness of the issues heightens. In 2009, emails were leaked from one of the leading climate change research facilities, the Climatic Research Unit (CRU). These emails appear to be scandalous at first glance and have been depicted in such a way that make it seem like climate scientists are trying to hide something and are abusing the scientific process by manipulating their own research to prove their points and reach their intellectual and monetary goals. This scandal has occurred so recently that it is difficult to make much sense of it all, so my major goal here is to simply show a reason why we the public are still so uncertain of what to think about this topic.
CHAPTER FIVE

CLIMATEGATE: EMAIL AND CONTROVERSY

On a cold day in November 2009, just a month before the international climate summit in Copenhagen, hackers made their way into the Climatic Research Unit’s (CRU) servers. The CRU houses many of the climatologists and scientists who work closely with the International Panel on Climate Change (IPCC) which releases integrative reports concerning the causes and effects of anthropogenic climate change. This breach caused thousands of emails and data from the CRU to leak into the public spectrum. You can find these emails on several online databases. However, and typically it seems, many contributors to these websites have made it a point to “cherry pick” certain emails to defend their climate change perspective (“Carbon Brief: The Blog”). To combat this, you must take care to keep what you read within its proper context.

If you were to sift through the huge mass of leaked CRU emails, you would find many strange things. There are a couple emails in which CRU director, Phil Jones, tells his cohort, including Michael Mann, to delete any emails that pertain to “AR4”--AR4 refers to the fourth IPCC report which was published in 2007 (Biello).¹ Statements like this make people wary. However, Mann claims

¹ The IPCC reports, as many of us may know, are the main documents used to influence international policies on dealing with climate change. Mann’s hockey stick graph is featured in the 2001 Third Assessment Report.
that no one acted on this request (Biello). Still, why would Jones ask them to delete information regarding AR4 or the IPCC? The loudest speculations have been that the CRU is trying to hide something. These opinions originate mostly from climate change skeptics. Others feel that Jones knew the information was leaked and did not want the information in the hands of people who were going to misuse it, manipulate it, or possibly plagiarize it before the new IPCC report is published—the new IPCC report is set release in September 2013. Whatever the case may be, it seems a sketchy email and has led many to think the scientists at the CRU are up to something, hence the catchy name *Climategate*.

In sifting through the assemblage of emails, you might come across an email from Peter Thorne, an environmental science professor at the University of East Anglia who works with the CRU, written to Phil Jones. Thorne writes about problems he had encountered with someone else’s research. In the emails, they had been discussing an upcoming meeting on the AR4 and Thorne was voicing some of his concerns about recent research flaws that he wanted to discuss at that meeting. Thorne makes the startling point that, “The science is being manipulated to put a political spin on it which for all our sakes might not be too clever in the long run” (Thorne). This is an interesting point. But, without the proper context, this can be interpreted in many ways. Some might think that the fact of the matter is that the CRU is trying to squeeze themselves into a political box in which they cannot fit simply to appease their funding politicians.
Others might say it could also indicate that the issues are literally being manipulated for political or personal gain and, as such, the science behind the issues is being manipulated to fit an agenda. This means that the science may become distorted and warped, as it is communicated, leading to a cascading effect where the initial meaning and true nature of it is lost. In the long run, that would be bad for everyone. Nevertheless, we don’t know for certain what exactly Thorne meant. Even with the other emails there to provide some context, there is simply not enough there to know what he meant. This is important to keep in mind when looking into the Climategate emails or any person’s interpretation of the emails therein.

There is one frequently mentioned and seemingly popular email in which Phil Jones writes, “I’ve just completed Mike’s Nature trick of adding in real temps to each series for the last 20 years (i.e. from 1981 onwards) and from 1961 for Keith’s to hide the decline” (“RealClimate: Climate science from climate scientists”). This statement has been one of the most widely used pieces of evidence to support the contention that climate scientists are trying to cover up something.

The Realclimate blog, run by purveyors of climate science, responded to this particular concern by explaining it within its actual context. They explained that when scientists use the word “trick,” it is often in reference to a good technique used to deal with an issue (“RealClimate: Climate science from climate scientists”). Often, this involves some sort of mathematical technique or simply
moving the data around. They then go on to explain that, in “hiding the decline,”
they are simply referring to a divergence issue from Ketih Briffa’s tree ring proxy
data. Briffa noticed that after 1960, the temperature data does not correlate with
the tree ring data and, for this, the climate scientists have agreed not to use the
tree ring data, admitting that they need additional research to why this is the way
it is ("RealClimate: Climate science from climate scientists").

In November 2011, almost two years after the initial leak, the CRU servers
were hacked once again and more emails were leaked into the public spectrum.
This second leak has received much less media coverage. This is probably
because the supposed scandal associated with the first leak has already been
exonerated by several various organizations who investigated the allegations of
scientific misconduct (Donovan).

We can put any spin on these emails that we like. We can pick apart the
emails however we like. In the end, though, this will not change the fact that we
may never know the underlying nature of the emails. An important thing to
remember is that these are personal emails. They are not meant to be
professional. It is possible that there is no underlying nature at all--these are just
email communications from one scientist to another talking about something
relevant to their shared, scientific world.

Scientists are human beings like the rest of us. They are entirely capable
of using “inside” jokes as well as being spiteful against people they don’t like--
Steve McIntyre. They can discuss work and disagree with each other about
certain aspects of research. They can talk about parts of their research they have doubts about--discussion may help them figure out better ways to approach their doubts or change experimental mechanisms. It is their personal business and we serve to gain very little from digging for concrete meaning in a pile of these emails.
CHAPTER SIX
ARGUMENT FOR PROGRESS SAKE

We have gone seemingly deep into our search for the quintessential truth and have not gotten very far. We have gained a basic understanding of the underlying hard science. We know the whats and hows involved in the scientific methodology. We know the places where subjective flaws came about. But, what do we do with that now? I return to the beginning back to the initial question. Is there common ground? Let’s take a look back at what we have examined thus far.

The Great Debate

Mann found that with a certain degree of certainty that the 1990s was anomalous to the past millennium. He did this by using the information provided by tree rings, ice cores, and coral bands. Using statistical methodology, Mann plotted his information along a time series and obtained the hockey stick.

McIntyre and McKitrick claimed that Mann’s statistical methodology was flawed. They did disagree with the overemphasis of some proxy data, namely bristlecone pines trees which were clustered in roughly the same area and do not serve as good temperature indicators. However, their issues were mainly on the statistical side—the biggest being the way Mann centered his data using the 1902-1980 calibration period.
Edward Wegman confirmed the claims made by McIntyre and McKitrick but was found guilty of plagiarism. Disregarding his scientific crimes, Wegman made excellent points about the debate itself. There are not enough statisticians in the climatological world and there are not enough climatologists in the statistical world. Each is an entity in and of itself. There needs to be more overlap. There needs to be more give and take. More importantly, both sides arguing blog versus blog is unacceptable. It is time to act like real scientists.

The Climategate scandal, or lack thereof, gives the debate an even more childish air. With a magnifying glass, you can find something wrong with what anyone says. You can also find something right with what anyone says. Yes, the climatologists insulted their opponents. Yes, they discussed problems they encountered with their research and problems they had with other scientist's research. The emails were not written for public viewing. There were not meant for anyone besides the intended recipient, so it is no wonder we do not understand and are suspicious of them--they're emails.

Where the Hockey Stick is Now

The MBH hockey stick of 1999 is most certainly not the last word on the matter (Amman, and Casper). While the debate has raged on between Mann, McIntyre and McKitrick, other climate scientists have developed more techniques and methods for a better temperature reconstruction. These temperature reconstructions, seen in Figure 8, have some minor differences but the
conclusion and inferences still remain the same. These scientists still believe that the temperatures in 20th century do not follow the natural cycles, rather, they are anomalous and likely caused by human, carbon dioxide emitting activities.

Figure 8. The CRU/Hadley Instrumental data set. This temperature reconstruction includes data from a variety of sources, each depicted by a different line color.

The figure above is similar to the hockey stick, shown here as the blue line, except that these reconstructions include the long sought after “Medieval Warming Period” as well as the “Little Ice Age.” Both of these constitute one of the major complaints logged by McIntyre and McKitrick. The contributors for this graph, as described, are constituents or researchers in the CRU. Mann, himself, is a co-contributor of two other lines--the orange and the yellow. Finally, one of
the major conclusions drawn by this graph is that 1998 is no longer the hottest year on record--it is now 2010.

The rebuttal has been much the same with regards to these graphs. Flaws are found and accentuated. Many of these flaws include similar issues in statistical methodology and unreliable indicators. These graphs too will likely be taken back to the drawing board someday and reworked until one day, everything just fits--nothing will ever fit exactly right in this day and age but the key is that every time these theories are reworked, they get a little stronger.

**Common Ground**

McIntyre and McKitrick admit to having a couple overlap points with Mann. The first is that they agree on the effect differing assumptions have over the temperature reconstruction (McIntyre, and McKitrick 5). By specifying the assumptions, the way the data is centered and the PCs used, both groups obtain similar reconstructions. The second is that bristlecone pine trees are not a reliable temperature proxy.

The common ground may not encompass much, but it is a start. Both groups appear to be incredibly stubborn and tend to interact with each other with condescending and confrontational tones. To even have this much common ground where they can agree on something little is maybe just enough to get the ball rolling.
Ambiguity, Argument, and Aesthetic

I am still surrounded by some ambiguity and indecision as to which side to attach to. And I, like many, am not the only one who feels this way. In my own mind, both sides make some interesting cases. But they, the scientists and statisticians, are arguing in such a way that we, the public, are left in the dust, clueless as to what to think or where to go from here. Science may be all about disproving this or disproving that but this does not mean there should be no give and take somewhere along the way just to give us a jumping point.

This may be the way the cookie crumbles unless we become professional climatologists or statisticians. It is up to us, the public, to make up our minds on the subject. We may never fully grasp what is going on or how conclusions are reached. The most we can hope for is a little help along the way--maybe one day a consensus. This will likely never happen. There will always be that back and forth. There will always be nay-sayers fighting with yea-sayers and vice versa. But, maybe that's the point. Maybe it is all about having somebody disagree with you. If we all agreed, scientists would be out of a job. After all, disproving is essential in keeping scientific progress striding forward. Is it perfect? No. And, it never will be. But, that is the beauty of it. The very essence of science is the argument. We humans are strange, strange beings.
BIBLIOGRAPHY


Shatto, Rahilla. Coral and Climate: Like tree rings, coral layers reveal climate change. 1998. Photograph. The Ocean and Global Climate


APPENDIX

1. THE MBH 1999 PAPER

Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations

Michael E. Mann and Raymond S. Bradley
Department of Geosciences, University of Massachusetts, Amherst Massachusetts
Malcolm K. Hughes
Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona

Abstract. Building on recent studies, we attempt hemispheric temperature reconstructions with proxy data networks for the past millennium. We focus not just on the reconstructions, but the uncertainties therein, and important caveats. Though expanded uncertainties prevent decisive conclusions for the period prior to AD 1400, our results suggest that the later 20th century is anomalous in the context of at least the past millennium. The 1990s was the warmest decade, and 1998 (or warmest year, at modestly high levels of confidence. The 20th century warming counters a millennium-scale cooling trend which is consistent with long-term astronomical forcing.

Introduction

Estimates of climate variability during past centuries must rely upon indirect “proxy” indicators. Natural archives that record past climate variations Trends over several centuries are evident in the recession of glaciers [Covacevich and Swanson, 1984], and in the sub-surface information from boreholes [Pelletier et al., 1990]. Annual climate estimates, however, require proxies such as tree rings, layered sediments, ice cores, and corals (combined with any available instrumental or historical records), which record seasonal/annual variations. Studies based on such “multiproxy” data networks [e.g., Bradley and Jones, 1993; Hughes and Diaz, 1994; Mann et al., 1995] have allowed the 20th century climate to be placed in a longer-term perspective, thus allowing for improved estimates of the influence of climate forcings [Lean et al., 1995; Crowley and Kim, 1996; Overpeck et al., 1997], and validation of the low-frequency behavior exhibited by climate models [e.g., Jones et al., 1996].

Recently, Mann et al. [1999, hereinafter “MBH98”] reconstructed yearly global surface temperature patterns back in time through the calibration of multiproxy networks against the modern temperature record. Skillful reconstruction of Northern Hemisphere mean annual surface temperature (“NH”) was possible back to AD 1800, as the pattern of surface temperature most readily calibrated by the available multiproxy network corresponds largely to synchronous large-scale temperature variation. It has been speculated that temperatures were warmer even further back, ~1000 years ago, a period described by Lamb [1966] as the Medieval Warm Epoch (though Lamb, examining evidence mostly from western Europe, never suggested this was a global phenomenon). We here apply the methodology detailed by MBH98 to the sparser proxy data network available prior to AD 1400, to critically revisit this issue, extending NH reconstructions as far back as is currently feasible. We also re-estimate earlier estimates of uncertainties in the NH series.

Data and Method

The multiproxy data network and instrumental temperature data used to calibrate it are discussed in detail by MBH98 (see supplementary information therein). Before AD 1400, only 12 indicators (of the more than 100 described by MBH98) are available. This includes the first 3 principal components (PC) of the (26) dendroclimate series available back to AD 1000 in the International Tree Ring Data Bank (ITRDB) mail from North America. The 12 indicators (54 counting two proxy index cases) are summarized in Table 1.

The calibration procedure (see MBH98) involves the assumptions (1) that a linear relationship exists between proxy climate indicators and and some combination of large-scale temperature patterns, and (2) that patterns of surface tem-

Table 1. 12 Proxy Indicators Available Back to AD 1000. Descriptions (“SERIES” see MBH98 for details regarding data and reference), location (“LOC”) region or lat/long coordinates, start year (“YEAR”), AD, and type (“TYPE”) of series is indicated. These data (and the NH series discussed in the text) are available over the Internet through the World Data Center-A for Paleoclimatology (http://www.ngdc.noaa.gov/paleo/paleo.html).

<table>
<thead>
<tr>
<th>SERIES</th>
<th>LOC</th>
<th>YEAR</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITEDH (PC #1)</td>
<td>N. Amer</td>
<td>1600</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>ITEDH (PC #2)</td>
<td>N. Amer</td>
<td>1600</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>ITEDH (PC #3)</td>
<td>N. Amer</td>
<td>1600</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>FFVHVH</td>
<td>658 21E</td>
<td>190</td>
<td>T. Ring density</td>
</tr>
<tr>
<td>Polar Urals</td>
<td>679 63E</td>
<td>190</td>
<td>T. Ring density</td>
</tr>
<tr>
<td>Tasia</td>
<td>435 18W</td>
<td>190</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>N. Patagonia</td>
<td>358 95W</td>
<td>190</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>Basler</td>
<td>144 7E</td>
<td>188</td>
<td>T. Ring wide</td>
</tr>
<tr>
<td>Greenland stacked core</td>
<td>775 66W</td>
<td>185</td>
<td>ice core O2</td>
</tr>
<tr>
<td>Quebec (2)</td>
<td>145 71W</td>
<td>88</td>
<td>ice core O2</td>
</tr>
<tr>
<td>Quebec (2)</td>
<td>145 71W</td>
<td>88</td>
<td>ice core O2</td>
</tr>
</tbody>
</table>

Copyright 1999 by the American Geophysical Union.


759

69
temperature in the past can be suitably described in terms of some linear combination of the dominant present-day surface temperature patterns. MBH98 performed extensive cross-validation experiments to verify the reliability of the reconstruction using global temperature data from 1854-1901 withheld from (1902-1980) calibration, and, further back, by the small number of instrumental temperature series available back through the mid-18th century.

In using the sparser dataset available over the entire millennium (Table 1), only a relatively small number of indicators are available in regions (e.g., western North America) where the primary pattern of hemispheric mean temperature variation has significant amplitude (see Fig. 2 in MBH98), and where regional variations appear to be closely tied to global-scale temperature variations in model-based experiments [Bradley, 1996]. These few indicators thus take on a particularly important role (in fact, as discussed below, one such indicator - PC #1 of the ITRDB data-is found to be essential), in contrast with the post AD 1400 reconstructions of MBH98 for which indicators are available in several key regions [e.g., the North American northern treeline ("NT") dendroclimatic chronologies of Jacoby and D'Arrigo, 1989].

Due to the leverage of ITRDB PC #1 in the millennial reconstruction, any non-climatic influence must first be removed before it can meaningfully be used in the reconstructions. Spurious increases in variance back in time associated with decreasing sample sizes [see e.g. Jones et al. 1998] are not an issue with this series, owing to the high degree of replication in the underlying chronologies back to AD 1000. A number of the highest elevation chronologies in the western U.S. do appear, however, to have exhibited long-term growth increases that are more dramatic than can be explained by instrumental temperature trends in these regions. Graybill and Kelsey [1993] suggest that such high-elevation, CO2-limited trees, in moisture-stressed environments, should exhibit a growth response to increasing CO2 levels. Though ITRDB PC #1 shows significant loadings among many of the 28 constituent series, the largest are indeed found on high-elevation western U.S. sites. The ITRDB PC#1 is shown along with the non-climatic NT series, during their 1400-1980 period of overlap (Figure 1)

The low-frequency coherence of the ITRDB PC#1 series and composite NT series during the initial four centuries of overlap [1400-1800] is fairly remarkable, considering that the two series record variations in entirely different environments and regions. In the late 10th century, the series diverge. As there is no a priori reason to expect the CO2 effect discussed above to apply to the NT series, and, furthermore, that series has been verified through cross-comparison with a variety of proxy series in nearby regions [Overpeck et al. 1997], it is plausible that the divergence of the two series is related to a CO2 influence on the ITRDB PC#1 series. The residual is indeed coherent with rising atmospheric CO2 (Figure 1b) until it levels off in the 20th century, which we speculate may represent a saturation effect whereby a new limiting factor is established at high CO2 levels. For our purposes, however, it suffices that we consider the residual to be non-climatic in nature, and consider the ITRDB PC #1 series "corrected" by removing it from the residual, forcing it to align with the NT series at low frequencies throughout their mutual interval of overlap. This correction is independently justified by the fact that temperatures averaged over the NT region and western U.S. region dominating ITRDB PC#1 exhibit very similar low-frequency trends this century (not shown).

Verification and Consistency Checks

The calibration/verification statistics for reconstructions based on the 12 indicators available back to AD 1800, are, as expected, somewhat degraded relative to those for the post AD 1400 period. The calibration and verification resolved variance (30% and 34% respectively) are consistent with each other, but lower than for reconstructions back to AD 1400 (42% and 51% respectively-see MBH98). Results further back than a millennium, based on even sparser data (see Table 1) are yet further degraded. With only a single eigenvector of the instrumental temperature data (#1- see Figure 2 in MBH98) skillfully resolved by the network available back to AD 1000, the total spatial variance calibrated is far more modest than that for the NH areas (= 5% in calibration and verification). Thus, the NH series, but not the spatial details, are most meaningful in the millennial reconstructions.

Further consistency checks are required. The most basic involves checking the potential resolvability of long-term variations by the underlying data used. An indicator of climate variability should exhibit, at a minimum, the red noise

![Figure 1](image-url)
Figure 2. Spectrum of NH series calibration residuals from 1902-1980 for post-AD 1820 (solid) and AD 1000 (dotted) reconstructions (scaled by their mean white noise levels). Median and 90%, 95%, and 99% significance levels (dashed lines) are shown.

The climate itself is known to exhibit [see Mann and Lees, 1996 and references therein]. A significant deficit of power relative to the median red noise level thus indicates a possible bias of true climatic variance, with a deficit of zero frequency power indicative of less trend than expected from noise alone, and the likelihood that the longest ("secular") timescales under investigation are not adequately resolved. Only 5 of the indicators (including the ITRDB PC #1, Polar Urals, Fenno-Scandia, and both Greenland series) are observed to have at least median red noise power at zero frequency for the pre-calibration (AD 1000-1991) period. It is furthermore found that only one of these series—PC #1 of the ITRDB data—exhibits a significant correlation with the time history of the dominant temperature pattern of the 1502-1980 calibration period. Positive calibration/variance scores for the NH series cannot be obtained if this indicator is removed from the network of 12 (in contrast with post-AD 1400 reconstructions for which a variety of indicators are available which correlate against the instrumental record). Though, as discussed earlier, ITRDB Project PC #1 represents a vital region for resolving hemispheric temperature trends, the assumption that this relationship holds up over time nonetheless demands cautious application. Clearly, a more widespread network of quality millennial proxy climate indicators will be required for more confident inferences.

A further consistency check involves examining the calibration residuals. In Figure 2 we show the power spectrum of the residuals of the NH calibration from 1902-1980 for both the calibrations based on all indicators in the network available back to 1820 (see MBH98), and the calibrations based on the 12 indicators available back to AD 1000. Not only (as indicated earlier) is the calibrated variance lower for the millennial reconstruction, but there is evidence of possible bias. While the residuals for the post-AD 1820 reconstructions are consistent with white noise (at no frequency does the spectrum of the residuals breach the 95% significance level for white noise—this holds in fact back to AD 1600), a roughly five-fold increase in unresolved variance is observed at secular frequencies (>95% significant) for the millennial reconstruction. In contrast to MBH98 where uncertainties were self-consistently estimated based on the assumption of Gaussian residuals, we here take account of the spectrum of unresolved variance, separately treating unresolved components of variance in the secular (longer than the 50-year calibration interval in this case) and higher-frequency bands. To be conservative, we take into account the slight, though statistically insignificant inflation of unresolved secular variance for the post-AD 1600 reconstructions. This procedure yields composite uncertainties that are moderately larger than those estimated by MBH98, though none of the primary conclusions therein are altered.

Temperature Reconstruction

The reconstructed NH series and estimated uncertainties are shown in Figure 3, along with its associated power spectrum. The substantial secular spectral peak is highly sig-

Figure 3. Millennial temperature reconstruction. (a) NH reconstruction (solid) and raw data (dotted) from AD 1000-1998. Smoothed version of NH series (thick solid), linear trend from AD 1000-1850 (dot-dashed) and two standard error limits (shaded) are also shown. (b) Power spectrum of the NH series based on full (AD 1000-1980) and pre-calibration (AD 1000-1901) intervals. Robustly estimated median and 90%, 95%, and 99% significance levels relative to red noise are shown [see Mann and Lees, 1996].
significant relative to red noise, associated with a long-term cooling trend in the NH series prior to industrialization ($\Delta T = -0.02\degree$C/century). This cooling is possibly related to astronomical forcing, which is thought to have driven long-term temperatures downward since the mid-Holocene at a rate within the range of $-0.01$ to $-0.04\degree$C/century [see Berger, 1988]. In addition, significant century-scale variability may be associated with solar irradiance variations [see Lean et al., 1995; Mih08], and a robust spectral peak centered at 50-70 year period seems to correspond to a multidisciplinary climate signal discussed by Mann et al. [1995]. The 20th century (1900-1988) (anomaly of $T = 0.07\degree$C relative to the 1902-1980 calibration period mean) is nominally the warmest of the millennium (11-12th: $-0.09$, 10th: $0.09$, 14th: $0.07$, 18th: $0.14$, 16th: $-0.14$, 17th: $-0.18$, 18th: $-0.14$, 19th: $-0.21$). Expanded uncertainties in centennial means prior to AD 1600, and warmer conditions during the earlier centuries of the millennium, however, preclude a definitive statement prior to AD 1400-the 11th and 12th centuries are within a (centennial) standard error of the 20th century. The late 11th, late 12th, and late 14th centuries rival those in 20th century temperature levels (see Figure 3a). Our reconstruction thus supports the notion of relatively warm hemispheric conditions earlier in the millennium, while cooling following the 14th century could be viewed as the initial onset of the Little Ice Age severe late. Considerable spatial variability is evident however [see Hughes and Diaz, 1994] and, as in Lanzhub’s [1965] original concept of a Medieval Warm Epoch, there are episodes of cooler as well as warmer conditions punctuating this period. Even the warmer intervals in our reconstruction pale, however, in comparison with modern (mid-to late 20th century) temperatures. For the NH series, both the past year (1998) and past decade (1989-1998) are well documented as the warmest in the 20th century instrumental record. Furthermore, the past decade ($T = 0.45\degree$C) is nearly two (decadal) standard errors warmer than the next warmest decade prior to the 20th century (1366-1375 $T = 0.11$), and 1998 ($T = 0.78\degree$C) more than two standard errors warmer than the next warmest year (1249 with an anomaly $T = 0.27\degree$C). 1253 and 1366 with $T = 0.25\degree$C are the only other two years approaching typical modern warmth), supporting the conclusion that both the past decade and past year are likely the warmest for the Northern Hemisphere this millennium. The recent warming is especially striking if viewed as defying a long-term cooling trend associated with astronomical forcing.

Conclusions

Although NH reconstructions prior to about AD 1400 exhibit expanded uncertainties, several important conclusions are still possible. While warmth early in the millennium approaches 20th century levels, the late 20th century still appears anomalous: the 1990s are likely the warmest decade, and 1998 the warmest year, in at least a millennium. More widespread high-resolution data which can resolve millennium-scale variability are needed before more confident conclusions can be reached with regard to the spatial and temporal details of climate change in the past millennium and beyond.

Acknowledgments. We thank P.D. Jones and two anonymous reviewers for their comments. We gratefully acknowledge the numerous researchers who have contributed to the ITRDB. This research was supported by grants from the NSF (ATM-9626833) and DOE. M.E.M. acknowledges support through the Alexander Hollaender Distinguished Postdoctoral Fellowship Program (DOE).

References

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

M.E. Mann, Department of Geosciences, Morrill Science Building, University of Massachusetts, Amherst MA 01003-5820

(Revised October 14, 1998; revised January 21, 1999; accepted January 27, 1999.)
Hockey sticks, principal components, and spurious significance

Stephen McIntyre
Northwest Exploration Co., Ltd., Toronto, Ontario, Canada

Ross McKitrick
Department of Economics, University of Guelph, Guelph, Ontario, Canada

Received 14 October 2004; revised 22 December 2004; accepted 17 January 2005; published 12 February 2005.

[1] The “hockey stick” shaped temperature reconstruction of Mann et al. (1998, 1999) has been widely applied. However it has not been previously noted in print that, prior to their principal components (PCs) analysis on tree ring networks, they carried out an unusual data transformation which strongly affects the resulting PCs. Their method, when tested on persistent red noise, nearly always produces a hockey stick shaped first principal component (PCI) and overshakes the first eigenvalue. In the controversial 15th century period, the MBH98 method effectively selects only one species (bristlecone pine) into the critical North American PCI, making it implausible to describe it as the “dominant pattern of variance”. Through Monte Carlo analysis, we show that MBH98 benchmarks for significance of the Reduction of Error (RE) statistic are substantially under-stated and, using a range of cross-validation statistics, we show that the MBH98 15th century reconstruction lacks statistical significance. 


1. Introduction

[2] The term “hockey stick” is often used to describe the shape of the Northern Hemisphere (NH) temperature index introduced in Mann et al. [1998] (hereinafter referred to as MBH98). For convenience, we define the “hockey stick index” of a series as the difference between the mean of the closing sub-segments (here 1902–1980) and the mean of the entire series (typically 1400–1980 in this discussion) in units of the long-term standard deviation (σ), and a “hockey stick shaped” series is defined as one having a hockey stick index of at least 1 σ. Such series may be either upside-up (i.e., the “blade” trends upwards) or upside-down. Our focus here is on the 1400–1450 step (“AD1400 step”) of MBH98, because of controversy over early 15th century temperature reconstructions [McIntyre and McKitrick, 2003; M. E. Mann et al., Note on paper by McIntyre and McKitrick in Energy and Environment, unpublished manuscript, 2003]. Our particular interest in the performance of the Reduction of Error (RE) statistic arises out of that controversy. We also focus on the North American tree ring network (“NOAEMER”), because the first principal component (“PCI”) of this network has been identified as essential for controversial periods of the MBH98 temperature reconstruction [Mann et al., 1999, unpublished manuscript, 2003]. MBH98 has recently been criticized on other grounds in von Storch et al. [2004].

[3] MBH98 used principal components (PCs) to reduce the dimensionality of tree ring networks and stated that they used “conventional” PC analysis. A conventional PC algorithm centers the data by subtracting the column means of the underlying series. For the AD1400 step highlighted here, this would be the full 1400–1980 interval. Instead, MBH98 Fortran code (ftp://holocene.evsc.virginia.edu/pub/MBH98/TREERINGNOAMER/pca-noamer) contains an unusual data transformation prior to PC calculation that has never been reported in print. Each tree ring series was transformed by subtracting the 1902–1980 mean, 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 1902–1986 period. The PCs were then computed using singular value decomposition on the transformed data. (The effects reported here would have been partly mitigated if PCs had been calculated using the covariance or correlation matrix.) This previously unreported transformation was recently acknowledged in the Supplementary Information to a Corrigendum to MBH98 [Mann et al., 2004], where they asserted that it has no effect on the results, a claim we refute herein.

[4] PCs can be strongly affected by linear transformations of the raw data. Under the MBH98 method, for those series in which the 1902–1980 mean is close to the 1400–1980 mean, subtraction of the 1902–1980 mean has little impact on weightings for the PCI. But if the 1902–1980 mean is different than the 1400–1980 mean (i.e., a hockey stick shape), the transformation translates the “shaft” off a zero mean; the magnitude of the residuals along the shaft is increased, and the series variance, which grows with the square of each residual, gets inflated. Since PC algorithms choose weights that maximize variance, the method reallocates variance so that hockey stick shaped series get overweighted. In effect, the MBH98 data transformation results in the PC algorithm mining the data for hockey stick patterns.

[5] In a network of persistent red noise, there will be some series that randomly “trend” up or down during the ending sub-segment of the series (as well as other sub-segments). In the next section, we discuss a Monte Carlo experiment to show that these spurious “trends” in a closing segment are sufficient for the MBH98 method, when applied to a network of red noise, to yield hockey stick PCIs, even though the underlying data gener-
Figure 1. Simulated and MBH98 Hockey Stick Shaped Series. Top: Sample PC1 from Monte Carlo simulation using the procedure described in text applying MBH98 data transformation to persistent trendless red noise; Bottom: MBH98 Northern Hemisphere temperature index re-construction.

"The simulations nearly always yielded PCs with a hockey stick shape, some of which bore a quite remarkable similarity to the actual MBH98 temperature reconstruction — as shown by the example in Figure 1. A sharp inflection was regularly observed at the start of the 1902–1980 "calibration period". Figure 2 shows histograms of the hockey stick index of the simulated PCs. Without the MBH98 transformation (top panel), a 1σ hockey stick occurs in the PC1 only 15.3% of the time (1.5σ ~ 0.1%). Using the MBH98 transformation (bottom panel), a 1σ hockey stick occurs over 99% of the time, (1.5σ ~ 73%; 1.75σ ~ 21% and 2σ ~ 0.2%)."

2. Monte Carlo Simulations of Hockey Sticks on Trendless Persistent Series

"We generated the red noise network for Monte Carlo simulations as follows. We downloaded and collated the NOAMER tree ring site chronologies used by MBH98 from M. Mann's FTP site and selected the 70 sites used in the AD 1400 step. We calculated autocorrelation functions for all 70 series for the 1400–1980 period. For each simulation, we applied the algorithm hosking.rtx from the waveslim package version 1.3 downloaded from www.cran.r-project.org/doc/packages/waveslim.pdf [Gencay et al., 2001], which applied a method due to Hosking [1984] to simulate trendless red noise based on the complete auto-correlation function. All simulations and other calculations were done in R version 1.9 downloaded from www.R-project.org [R Development Core Team, 2003]. Computer scripts used to generate simulations, figures and statistics, together with a sample of 100 simulated "hockey sticks" and other supplementary information, are provided in the auxiliary material. We carried out 10,000 simulations, in each case obtaining 70 stationary series of length 581 (corresponding to the 1400–1980 period). By the very nature of the simulation, there were no 20th century trends, other than spurious "trends" from persistence. We applied the MBH98 data transformation to each series in the network: the 1902–1980 mean was subtracted, then the series was divided by the 1902–1980 standard deviation, then by the 1902–1980 detrended standard deviation. We carried out a singular value decomposition on the 70 transformed series (following MBH98) and saved the PC1 from each calculation."

"The hockey sticks were upside-down about half the time and upside-down half the time, but the 1902–1980 mean is almost never within one σ of the 1400–1980 mean under the MBH98 method. PC series have no inherent orientation and, since the MBH98 methodology uses proxies (including the NOAMER PC1) in a regression calculation, the fit of the regression is indifferent to whether the hockey stick is upside-up or upside-down. In the latter case, the slope coefficient is negative. In fact, the North American PC1 of Mann et al. [1999] is an upside-down hockey stick, as shown at ftp://ftp.ngdc.noaa.gov/paleo/contributions_by_author/mann1999/proxies/90b2name-pc1.dat."

"The loadings on the first eigenvalues were inflated by the MBH98 method. Without the transformation, the median fraction of explained variance of the PC1 was only 4.1% (99th percentile=5.5%). Under the MBH98 transformation, the median fraction of explained variance from PC1 was 13% (99th percentile=23%), often making the PC1 appear"
collected by Donald Graybill [Graybill and Ido, 1993] (see Table 1). The weights in the MBH98 PC1 have a nearly linear relationship to the hockey stick index. The most heavily weighted site in the MBH98 PC1, Sheep Mountain, is a bristlecone pine site with the most pronounced hockey stick shape (1.6 σ in the network; it receives over 390 inches of snow each winter). The weights of other heavily weighted sites, Mayberry Slough, whose hockey stick index is near 0.

[12] Under the MBH98 data transformation, the distinctive contribution of the bristlecone pines is in the PC1, which has a spuriously high explained variance coefficient of 38% (without the transformation = 18%). Without the data transformation, the distinctive contribution of the bristlecones only appears in the PC4, which accounts for less than 8% of the total explained variance.

[13] This substantially reduced share of explained variance, together with the fact that species other than bristlecone/foxtail pines are effectively omitted from the MBH98 PC1, argues strongly against interpreting it as the “dominant component of variance” in the North American network (M. E. Mann et al., Reply to “Global-scale temperature patterns and climate forcings over the past six centuries: A comment” by S. McIntyre and R. North American Network.

We now show the effect of the MBH98 algorithm on the actual NOAMER network in the controversial AD 1400 step.

[14] Without the data transformation the PC1 is very similar to the unweighted mean of all the series and, as shown in the bottom panel of Figure 3, does not have a hockey stick shape. However, under the MBH98 algorithm, the PC1 has a marked hockey stick shape, as shown in the top panel of Figure 3. The MBH98 method creates a PC1 which is dominated by bristlecone pines and closely related foxtail pines. (Fossil pine needles are located in an adjacent mountain range, interbred with bristlecone pines and are included here with bristlecone pines collectively). Out of 70 sites in the network, 93% of the variance in the MBH98 PC1 is accounted for by only 15 bristlecone and foxtail pine sites.

Table 1. 15 Highly Weighted Sites in MBH98 PC1

<table>
<thead>
<tr>
<th>ID Code</th>
<th>Name</th>
<th>Species</th>
<th>Elevation (m)</th>
<th>Author</th>
<th>Graybill and Ido [1993]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ac510</td>
<td>San Francisco Pia</td>
<td>PIAA</td>
<td>3535</td>
<td>D.A. Graybill</td>
<td>10</td>
</tr>
<tr>
<td>ca227</td>
<td>Flower Lake</td>
<td>PIAA</td>
<td>3291</td>
<td>D.A. Graybill</td>
<td>13</td>
</tr>
<tr>
<td>ca299</td>
<td>Timber Gap Upper</td>
<td>PIAA</td>
<td>3261</td>
<td>D.A. Graybill</td>
<td>14</td>
</tr>
<tr>
<td>ca310</td>
<td>Cleaver Peak</td>
<td>PIAA</td>
<td>3505</td>
<td>D.A. Graybill</td>
<td>12</td>
</tr>
<tr>
<td>ca333</td>
<td>Campito Mountain</td>
<td>PILO</td>
<td>3400</td>
<td>D.A. Graybill</td>
<td>5</td>
</tr>
<tr>
<td>ca344</td>
<td>Sheep Mountain</td>
<td>PILO</td>
<td>3475</td>
<td>D.A. Graybill</td>
<td>11</td>
</tr>
<tr>
<td>ca522</td>
<td>Mount Goliath</td>
<td>PIAA</td>
<td>3535</td>
<td>D.A. Graybill</td>
<td>2</td>
</tr>
<tr>
<td>ca523</td>
<td>Windy Ridge</td>
<td>PIAA</td>
<td>3570</td>
<td>D.A. Graybill</td>
<td>4</td>
</tr>
<tr>
<td>ca524</td>
<td>Almagra Mountain</td>
<td>PIAA</td>
<td>3536</td>
<td>D.A. Graybill</td>
<td>1</td>
</tr>
<tr>
<td>ca525</td>
<td>Hermit Lake</td>
<td>PIAA</td>
<td>3660</td>
<td>D.A. Graybill</td>
<td>3</td>
</tr>
<tr>
<td>ca510</td>
<td>Charleston Peak</td>
<td>PILO</td>
<td>3425</td>
<td>D.A. Graybill</td>
<td>6</td>
</tr>
<tr>
<td>ca512</td>
<td>Pearl Peak</td>
<td>PILO</td>
<td>3170</td>
<td>D.A. Graybill</td>
<td>9</td>
</tr>
<tr>
<td>ca513</td>
<td>Mount Washington</td>
<td>PILO</td>
<td>3415</td>
<td>D.A. Graybill</td>
<td>8</td>
</tr>
<tr>
<td>ca514</td>
<td>Spitzbar Mountain</td>
<td>PILO</td>
<td>3110</td>
<td>D.A. Graybill</td>
<td>6</td>
</tr>
<tr>
<td>ca516</td>
<td>Hill 10642</td>
<td>PILO</td>
<td>3050</td>
<td>D.A. Graybill</td>
<td>5</td>
</tr>
</tbody>
</table>

*15 high-altitude bristlecone (PILO, PIAA) and foxtail (PIBA) sites dominating MBH98 PC1, constituting 13 of 14 sites listed in Table 1 of Graybill and Ido [1993].
and hence no exact or asymptotic tables of significance levels [Cook et al., 1994]. MBH98 attempted to benchmark the significance level for the RE statistic using Monte Carlo simulations based on AR1 red noise with a lag coefficient of 0.2, yielding a 99% significance level of 0.0. However their simulation under-estimates the actual persistence of tree ring proxies and ignores the effect of the MBH98 data transformation in over-weighting hockey stick shaped series.

[15] In order to obtain more accurate significance benchmarks, we regressed each of the 10,000 simulated PC1s against the MBH98 northern hemisphere temperature series (the “sparse” subset used by MBH98 for verification ftp://ftp.ngdc.noaa.gov/paleo/paleocan/by_contributor/mann1998/nhem-sparse.dat) in the 1901–1980 calibration period – a procedure which more closely emulates actual MBH98 methods. Since the simulated PC1s are red noise series containing no information about the climate, they can be used to establish lower limits for the significance levels which the actual proxy data must exceed to indicate reconstructive skill. Since MBH98 used 22 indicators in their AD1400 step calculation, whereas the Monte Carlo simulation used only the simulated NOAMER PC1, the actual RE significance level would be higher than the benchmark calculated here, which is only a lower limit, making the arguments herein conservative.

[16] For each regression, we calculated the temperature “reconstruction” from the simulated PC1 in the verification period (1854–1901), and used the “reconstruction” to calculate the RE, R², CE, Sign Test and Product Mean Test. From this data, we determined the 99% significance levels in the verification period as shown in Table 2. The pattern of verification statistics was quite consistent: a high RE statistic, a very low CE statistic and a low R² statistic, relative to white or weakly red noise values.

[17] According to our calculations, the lower-limit critical value for 99% RE significance is 0.59 (5% – 0.54), values much higher than the 99% critical value of 0.0 reported by MBH98. The reported RE value for the AD1400 step of the MBH98 reconstruction was 0.51 (90th percentile under our RE distribution). Mann et al. have not archived supporting calculations for the AD1400 step. Accordingly, we emulated the AD1400 step of MBH98 using their data, obtaining the verification period statistics shown in Table 3. We were only able to obtain an RE statistic of 0.46 (80th percentile under our RE distribution) and an R² statistic of 0.02 (statistically insignificant). Other verification statistics also lack statistical significance and the high RE-low R² pattern is obviously similar to the patterns from comparably treated red noise.

5. Discussion and Conclusions

[18] PC analyses are sensitive to linear transformations of data, even if such transformations only appear to be

<table>
<thead>
<tr>
<th>Table 2. Statistical Significance Levels*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verification Statistic</td>
</tr>
<tr>
<td>RE (0)</td>
</tr>
<tr>
<td>R²</td>
</tr>
<tr>
<td>CE</td>
</tr>
<tr>
<td>Sign Test</td>
</tr>
<tr>
<td>Product Mean Test</td>
</tr>
</tbody>
</table>

*99% benchmarks from simulations described in text and as reported by MBH98.

Table 3. Verification Period Statistics for AD1400 Step of MBH98 Reconstruction

<table>
<thead>
<tr>
<th>AD1400 Step</th>
<th>RE (0)</th>
<th>R²</th>
<th>CE</th>
<th>Sign Test</th>
<th>Product Mean Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emulation</td>
<td>0.46</td>
<td>0.02</td>
<td>−0.26</td>
<td>22</td>
<td>1.54</td>
</tr>
<tr>
<td>MBH98</td>
<td>0.51</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*From emulation and as reported by MBH98.

“Standardizations”. Here we have shown, in the case of MBH98, that a “standardization” step (that the authors did not even consider sufficiently important to disclose at the time of their study) significantly affected the resulting PC series. Indeed, the effect of the RE transformation is so strong that a hockey-stick shaped PC1 is nearly always generated from (trendless) red noise with the persistence properties of the North American tree ring network. This result is disquieting, given that the NOAMER PC1 has been reported to be essential to the shape of the MBH98 Northern Hemisphere temperature reconstruction.

[20] For evaluation of statistical skill in paleoclimatic studies, the Reduction of Error (RE) statistic is widely used, but lacks a theoretical distribution. Practitioners use Monte Carlo models to establish significance benchmarks. Here we have shown that the benchmarks can be dramatically affected by the Monte Carlo model itself and that the 99% significance level reported by the Monte Carlo model more accurately represents actual MBH98 procedures 0.59, as compared to the level of 0.0 reported in the original study. More generally, this example shows that changes in methodology will generally require new Monte Carlo modeling, that benchmarks carried forward from one methodology cannot necessarily be applied to a new methodology – even if the method changes may appear slight, and that great caution is required prior to concluding statistical significance based on RE statistics.

[21] An obvious guard against spurious RE significance is to examine other cross-validation statistics, such as the R² and CE statistics, as recommended, for example, by Cook et al. [1994]. While there are limitations to the R² statistic, the analysis of statistical “skill” of Murphy [1988] presupposes that the R² statistic exceeds the skill statistic and cases where the RE statistic exceeds the R² statistic are of particular concern [Cook et al., 1994]. In the case of MBH98, unfortunately, neither the R² and other cross-validation statistics nor the underlying construction step have ever been reported for the controversial 15th century period. Our calculations have indicated that they are statistically insignificant. Timely reporting of these statistics (in the original article) might have led to an earlier consideration of the discrepancy between the apparently high RE value and the low values of other statistics, and thus enabled earlier identification of the underlying data transformation resulting in this problem.

[22] Acknowledgment. No funding was sought or received for this work.

References


S. McIntyre, Northwest Exploration Co., Ltd., 512-120 Adelaide St. West, Toronto, Ontario, Canada M5H 1T1. (stephen.mcintyre@utoronto.ca)

R. McKitrick, Department of Economics, University of Guelph, Guelph, Ontario, Canada N1G 2W1.
July 15, 2005

Via Federal Express

Joe Barton, Chairman
House Committee on Energy and Commerce
Ed Whitfield, Chairman
Subcommittee on Oversight and Investigations
2125 Rayburn House Office Building
Washington, D.C. 20515

Dear Chairman Barton and Chairman Whitfield:

This letter responds to your letter of June 23, 2005, which seeks information on issues relating to my research on the historical record of temperatures and climate change. Your letter lays out a number of “concerns” about the research my colleagues and I have conducted about global warming. Your letter also inquires about the role I played in the preparation of the United Nation’s Intergovernmental Panel on Climate Change Third Assessment Report (the so-called “TAR”).

I will address each of your questions in turn. Before doing so, however, let me state that my research findings, which support the conclusion that the earth’s surface is warming, and that recent warming is due in large part to human influences, are consistent with the overwhelming scientific consensus on climate change. My research has been subject to intensive peer-review. Other scientists have replicated all facets of my research and have found it accurate and reliable. The specific conclusion published by my colleagues and me that late 20th century Northern Hemisphere warmth is anomalous in the context of at least the past millennium is common to many studies. Based on multiple supporting studies, the TAR came to a similar conclusion. The TAR did not rely solely on the work of my colleagues and me in reaching this conclusion. Recent work since the TAR has provided further support for this conclusion, which is now common to more than a dozen independent studies published in the peer-reviewed scientific literature. (I have provided for reference a comprehensive review by Jones and Mann in the journal “Reviews of Geophysics” of the American Geophysical Union (AGU.) The criticisms your letter cites have been soundly rejected by the scientific community.

1 This response is submitted without waiving any objection I might have to the Committee’s jurisdiction over the subject matter of this inquiry.
The most serious contention in your letter — namely, that my work has not been subject to replication because I have failed to make available the underlying research data — is incorrect. Your letter notes that the National Research Council’s “gold standard” for scientific research is the ability of other scientists to replicate first-generation research, and I fully agree. My colleagues and I follow the National Research Council’s guidance with regards to the disclosure of research data, and all of our data and methodologies have been fully disclosed and are available to anyone with a computer and an internet connection. As a result of our willingness to share our research with others, an independent team of scientists has used the research data my colleagues and I have made public to replicate our research and confirm the reliability of our findings. See Wahl, E.R., Ammann, C.M., Robustness of the Mann, Bradley, Hughes Reconstruction of Surface Temperatures: Examination of Criticisms Based on the Nature and Processing of Proxy Climate Evidence. Climate Change (2005) (forthcoming) and associated website: http://www.cgd.ucar.edu/ccr/ammann/millennial/MBH_reevaluation.html.

Let me now turn to your specific questions, which ask that I provide the following information:

Q1: Your letter first asks that I furnish the Committee my curriculum vitae, along with 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 those studies.”

A: This material is attached.

Q2: Your letter next asks that I “[l]ist all financial support” I have received to support my research.

A: See attachment.

Q3: Your letter requests that I provide, for all “work involving federal grants or funding support under which you were a recipient of funding or a principal investigator,” “all agreements relating to those underlying grants or funding, including, but not limited to, any provisions, adjustments, or exceptions made in the agreement relating to the sharing of research results.”

A: These requests are not directed to the appropriate person. The Committee should contact the University of Massachusetts and University of Virginia offices of grant administration for these materials. With respect to the UMass NSF research funds (which supported the 1998 Nature article), it should furthermore be noted that I was not the Principal Investigator for this NSF project, and I am not, nor have I ever, been in possession of any official paperwork related to this grant.

Q4: Your next question asks for “the location of all data archives relating to each published study for which I was an author or co-author” and whether such data would be sufficient to permit other researchers to replicate the work.
A: The data, descriptions of methods, and results related to my research — more than sufficient to permit other researchers to replicate the research — have been extensively archived (in many cases, in several archives) on public websites, and data links within the websites. The website addresses appear in the margin. 9

Q5: This question begins by stating that, “[a]ccording to The Wall Street Journal, you have declined to release the exact computer code you used to generate your results.” The question then poses a series of questions: “(a) Is that correct? (b) What policy on
sharing research and methods do you follow? (c) What is the source of that policy? (d) Provide this exact computer code used to generate your results.”

The question presumes that in order to replicate scientific research, a second researcher has to have access to exactly the same computer program (or “code”) as the initial researcher. This premise is false. The key to replicability is unfettered access to all of the underlying data and methodologies used by the first researcher. My data and methodological information, and that of my colleagues, are available to anyone who wants them. As noted above, other scientists have reproduced our results based on publicly available information.

It also bears emphasis that my computer program is a private piece of intellectual property, as the National Science Foundation and its lawyers recognized. The National Science Foundation—the government agency that establishes policy in this area—has confirmed that my colleagues and I have met every requirement of transparency and openness in our research. My research is all based on data sets regarding the Earth’s

3 All of the proxy data (tree-rings, coral, ice cores, and historical documents) used in Mann et al. (1998) has been available since May 2000 on this public website: ftp://holocene.evsc.virginia.edu/pub/MBH98. The methodology used by my colleagues and me is described in detail in the initial publication, and further expanded upon in July 2004 on Nature’s supplementary website (http://www.nature.com/nature/journal/v430/n6995/suppinfo/nature02478.html) and on our own website, ftp://holocene.evsc.virginia.edu/pub/MANNETAL98. Moreover, independently-derived source codes for implementing our algorithm, and all required input data, have been posted on the website of the National Center for Atmospheric Research. See http://www.ucar.edu/ccr/amusmann/milennium/CODES_MBH.html. For these reasons, charges that our work is not subject to replication are unfounded. The initial description of the work was sufficient to permit researchers to independently produce the key algorithms. See, e.g., Zorita, E., F. Gonzalez-Rouco, and S. Legutke, Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G Coupled Climate Model, J. Climate, 16, 1378-1390 (2003); Von Storch, H., E. Zorita, J.M. Jones, Y. Dimitriev, F. Gonzalez-Rouco, F., and S.F.B. Tett, Reconstructing Past Climate from Noisy Data, Science, 306, 679-682 (2004). Not only have we replicated our results with a different methodology (Rutherford, S., Mann, M.E., Osborn, T.J., Bradley, R.S., Briggs, K.R., Hughes, M.K., Jones, P.D., Proxy-based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Methodology, Predictor Network, Target Season and Target Domain, Journal of Climate (2005) (to appear in July issue), but an independent group has replicated our original methods and results (See Wahl, E.R. and Ammann, C.M., Robustness of the Mann, Bradley, Hughes Reconstruction of Surface Temperatures: Examination of Criticisms Based on the Nature and Processing of Proxy Climate Evidence, Climatic Change (2005) (forthcoming)).
climate that are freely and widely available to all researchers. Whether I make available my computer programs is irrelevant to whether our results can be reproduced. And whether I make my computer programs publicly available or not is a decision that is mine alone to make. Since other scientists have used the methods we described and the data we archived to replicate our results, the issue of whether my computer program is available has no bearing whatsoever on the veracity of our results. The question you posed — whether I have fully satisfied established scientific standards for data-sharing — has been fully considered by the National Science Foundation. As your letter notes, two Canadian researchers, Steve McIntyre and Ross McKitrick, contacted NSF to inquire whether I had complied with National Science Foundation requirements. The National Science Foundation twice informed them that I have, in fact, complied with all applicable transparency and openness standards and that, under long-standing Foundation policy, the computer codes referred to by The Wall Street Journal are considered the intellectual property of researchers and are not subject to disclosure.4

4 For the sake of completeness, let me quote in its entirety the email message sent by Dr. David J. Varando, Director, Palaeoclimate Program, Division of Atmospheric Sciences, National Science Foundation to Mr. Steve McIntyre (copied to me), on December 17, 2003, in response to a previous email that McIntyre had sent to Dr. Varando (copied to me):

Dear Mr. McIntyre,

I apologize if my last electronic message was not clear but let me clarify the US NSF’s view in this current message. Dr. Mann and his other US colleagues are under no obligation to provide you with any additional data beyond the extensive data sets they have already made available. He is not required to provide you with computer programs, codes, etc. His research is published in the peer-reviewed literature which has passed muster with the editors of those journals and other scientists who have reviewed his manuscripts. You are free to your analysis of climate data and he is free to his. The passing of time and evolving new knowledge about Earth’s climate will eventually tell the full story of changing climate. I would expect that you would respect the views of the US NSF on the issue of data access and intellectual property for US investigators as articulated by me to you in my last message under the advisement of the US NSF’s Office of General Counsel.

Respectfully,
David J. Varando
Director, Palaeoclimate Program
Division of Atmospheric Sciences
National Science Foundation
4201 Wilson Blvd.
Arlington, VA 22203

Even more recently, the National Science Foundation confirmed its view that my computer codes are my intellectual property. A recent issue of the Chronicle of Higher Education states: “According to David Stormer, of the Congressional-
With this background in mind, let me now respond to your specific inquiries:

A (Q5A): I have made available all of the research data that I am required to under United States policy as set by the National Science Foundation. In accordance with the rules promulgated by the Foundation and supported by the Foundation’s General Counsel, I maintain the right to decline to release any computer codes, which are my intellectual property.

A (Q5B): The policy regarding sharing research and methods I and my colleagues follow is to disclose any information that might be useful to other researchers, including the data, description of methodology, and so forth, that would enable a competent scientist to replicate our work. This proof, of course, is that other scientists have in fact succeeded in replicating our work. And, as noted above, our policies are fully in keeping with those established by the National Science Foundation.

A (Q5C): The source of these policies is the National Science Foundation.

A (Q5D): My computer program is a piece of private, intellectual property, as the National Science Foundation and its lawyers recognize. It is a bedrock principle of American law that the government may not take private property “without [a] public use,” and “without just compensation.”

That notwithstanding, the program used to generate the original Mann et al. 1998 temperature reconstructions is posted at this website: ftp://holoclima.ee.ucc.virginia.edu/pub/MANNETAL98/ (see “METHODS” subdirectory)

Q6: The Committee next asks that, “Regarding study data and related information that is not publicly archived, what requests have you and your co-authors received for data relating to climate change studies, what was your response, and why?”

A: I can of course only speak for myself, but I do not believe that there is any “study data” used in my published work that is not publicly archived. Having said that, I do respond diligently to any requests from scientific colleagues for data or methodological details relating to our research.

affairs office at the National Science Foundation, Mr. McIntyre contacted the foundation last year to ask for Mr. Mann’s computer code. Mr. Stomper said the agency had told Mr. McIntyre that the code was the intellectual property of Mr. Mann . . . .” Richard Monastersky, Congressman Demands Complete Records on Climate Research by 3 Scientists Who Support of Global Warning, Chronicle of Higher Education (July 1, 2005), available at: http://chronicle.com/temp/email.php?id=dopjhw74wvqr3k9tae5avlof/vb2yu.
Q7: This question poses a number of questions based on an article published by McIntyre and McKitrick in *Energy & Environment*. The question states that these authors "report a number of errors and omissions in Mann et al. 1998 and how these may affect the underlying conclusions of the work." The question goes on to list a number of topics that I should address in a "narrative explanation."

A: I want to begin by emphasizing that nothing in McIntyre and McKitrick's article undermines the conclusion of my research. My colleagues and I stand foursquare behind our work. So does the scientific community.

The various claims of McIntyre and McKitrick — including the ones repeated in your question — have been exhaustively examined by two different groups of climate researchers who have found their objections to be unfounded. See also National Center for Atmospheric Research, Media Advisory: The Hockey Stick Controversy New Analysis Reproduces Graph of Late 20th Century Temperature Rise (May 11, 2005) (available at: http://www.ucar.edu/news/releases/2005/annanu.htm). Moreover, it is my understanding that several other groups of climate researchers have examined McIntyre and McKitrick's criticisms and also have found their criticisms lacking in merit. On the other hand, I know of no independent scientific group that has found any of McIntyre and McKitrick’s claims to be valid.

Nor is that surprising. *Energy & Environment* is not a peer reviewed scientific journal; it is a journal primarily devoted to policy rather than science that appears to engage in, at most, haphazard review of its articles. And neither McIntyre nor McKitrick is a trained climate scientist. According to the biographical data on their websites, Mr. McIntyre is a mining industry executive with no formal training in any discipline related to climate research and Mr. McKitrick is an economist with no scientific training, hardly credentials that lend force to their academic arguments. See http://www.nongulph.ca/~smckitri/cv.html and http://www.nongulph.ca/~smckitri/research/vtevbioc.doc.

Adding to the problem, the editor of *Energy & Environment*, Ms. Sonja Boehm-Christiansen, has candidly acknowledged that the publication has a clear editorial bias. In the September 5, 2003 issue of the *Chronicle of Higher Education*, Ms. Boehm-Christiansen is quoted as describing the editorial policy of *Energy & Environment* in this way: "I'm following my political agenda — a bit, anyway. *** But isn't that the right of

an editor?" As to "peer review," Ms. Boehmert-Christiansen has acknowledged in an email to Dr. Tim Osborn of the Climatic Research Unit at the University of East Anglia (U.K.), that in her rush to get the McIntyre and McKitrick piece into print for political reasons Energy & Environment dispensed with what scientists consider peer review ("I was rushing you to get this paper out for policy impact reasons, e.g. publication well before COP9"). As Ms. Boehmert-Christiansen added, the "paper was amended until the very last moment. There was a trade off in favour of policy." McIntyre and McKitrick's work has been discredited by ample peer-reviewed, scientific work.

Nonetheless, let me try to respond to the Committee's specific questions.

A(7A,7B): The Committee inquires about the sensitivity of the results of the Mann et al. 1998 study to the inclusion or omission of certain North American tree-ring data ("the bristlecone pine series" and "archived Gaspé tree ring data" referred to in the Committee's letter). For a complete scientific response, you should consult the article my co-authors and I published back in 1999 addressing precisely these issues: Mann, M.E., Bradley, R.S., and Hughes, M.K., Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, Geophysical Research Letters, 26, 759-762 (1999).

The issues raised by the Committee involve a 100 year sub-interval of our reconstruction from AD 1400-1500. As my co-authors and I explained in our 1999 article cited above, given the proxy data available at that time, certain key tree-ring data (including the series mentioned above) were essential, if the reconstructed temperature record during early centuries were to have any climatologic "skill" (that is, any validity or meaningfulness). These conclusions were of course reached through analyses in which these key datasets were excluded, and the results tested for statistical validity. Our conclusions have been confirmed by Wahl and Ammann (see above). These researchers have demonstrated that the reconstructions produced by McIntyre and McKitrick result from ignoring these key data, and fail the accepted, basic tests for statistical validity. Moreover, Wahl and Ammann demonstrate that the climatologically improbable results obtained by McIntyre and McKitrick, which would suggest that the Northern Hemisphere was unusually warm during the 15th century (the middle of the so-called "Little Ice Age"), are statistically meaningless, and an artifact of both their exclusion of key proxy data (as discussed above) and the use of a flawed implementation of the Mann et al. 1998 method. See http://www.cgd.ucar.edu/cas/annan/mannmillennial/CODES_MBH.html (chart at the bottom of the page).

Since 1999 new proxy data have become available and new methodologies developed for using them. Studies using these data and methodologies have confirmed the primary conclusion of our work (e.g. Mann et al. 1998 and Mann et al. 1999) that the most recent decades were likely the warmest of the past 1,000 years for the Northern Hemisphere on the average. The most recent such study (published in Nature in fact extends this conclusion to at least the past 2,000 years. Moberg, A., D.M. Soneckin, K. Holmgren, N.M. Dusenko, and W. Karlén, Highly Variable Northern Hemisphere

A(7C): The Committee inquires about the calculation of the R2 statistic for temperature reconstruction, especially for the 15th Century proxy calculations. In order to answer this question it is important to clarify that I assume that what is meant by the “R2” statistic is the squared Pearson dot-moment correlation, or $r^2$ (i.e., the square of the simple linear correlation coefficient between two time series) over the 1856-1901 “verification” interval for our reconstruction. My colleagues and I did not rely on this statistic in our assessment of “skill” (i.e., the reliability of a statistical model, based on the ability of a statistical model to match data not used in constructing the model) because, in our view, and in the view of other reputable scientists in the field, it is not an adequate measure of “skill.” The statistic used by Mann et al. 1998, the reduction of error, or “RE” statistic, is generally favored by scientists in the field. See, e.g., Luterbacher, J.D., et al., European Seasonal and Annual Temperature Variability, Trends and Extremes Since 1500, Science 303, 1499-1503 (2004).

RE is the preferred measure of statistical skill because it takes into account not only whether a reconstruction is “correlated” with the actual test data, but also whether it can closely reproduce the mean and standard deviation of the test data. If a reconstruction cannot do that, it cannot be considered statistically valid (i.e., useful or meaningful). The linear correlation coefficient ($r$) is not a sufficient diagnostic of skill, precisely because it cannot measure the ability of a reconstruction to capture changes that occur in either the standard deviation or mean of the series outside the calibration interval. This is well known. See Wilks, D.S., Statistical Methods in Atmospheric Science, chap. 7 (Academic Press 1995); Cook, et al., Spatial Regression Methods in Dendroclimatology: A Review and Comparison of Two Techniques, International Journal of Climatology, 14, 379-402 (1994). The highest possible attainable value of $r^2$ (i.e., $r^2 = 1$) may result even from a reconstruction that has no statistical skill at all. See, e.g., Rutherford, et al., Proxy-based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Methodology, Predictor Network, Target Season and Target Domain, Journal of Climate (2005) (in press, to appear in July issue)(available at: http://holocene.esr.virginia.edu/pub/mann/RutherfordClimate-inpress05.pdf). For all of these reasons, we, and other researchers in our field, employ RE and not $r^2$ as the primary measure of reconstructive skill.

As noted above, in contrast to the work of Mann et al. 1998, the results of the McIntyre and McKitrick analyses fail verification tests using the accepted metric RE. This is a key finding of the Wahl and Ammann study cited above. This means that the reconstructions McIntyre and McKitrick produced are statistically inferior to the simplest possible statistical reconstruction: one that simply assigns the mean over the calibration period to all previous reconstructed values. It is for these reasons that Wahl and Ammann have concluded that McIntyre and McKitrick’s results are “without statistical and climatological merit.”
A(7D): The Committee asks: “[w]hat validation statistics did you calculate for the reconstruction prior to 1820, and what were the results?” Our validation statistics were described in detail in a table provided in the supplementary information on Nature’s website accompanying our original nature article, Mann, M.E., Bradley, R.S., Hughes, M.K., Global-Scale Temperature Patterns and Climate Forcing Over the Past Six Centuries, Nature, 392, 779-787 (1998). These statistics remain on Nature’s website (see http://www.nature.com/nature/journal/v392/n6678/supinfo/392779a0.html) and on our own website. See ftp.holocene.ewc.virginia.edu/pub/Mannetal98.

A(7E): The Committee asks how I “choose particular proxies and proxy series.” Again, this information is furnished in detail in both our original 1998 article in Nature, and expanded upon in a follow-up article published in 2000. See Mann et al., Global Temperature Patterns in Past Centuries: An Interactive Presentation, Earth Interactions 4-4, 1-29 (2000), specifically this link therein: http://www.ncdc.noaa.gov/paleo/ea/ea_nodendro.htm.

As our 1998 study and the additional information mentioned above make clear, we made use of all long-term, annually-resolved proxy indicators available to us in the public domain or through colleagues at the time the research was initiated (1996-1997) that met requirements for suitable length, age, model reliability, and in the case of tree ring series, replication, inter-correlation and metadata as described above.

Q8: This question asks me to “[e]xplain in detail” my work “for and on behalf of the Intergovernmental Panel on Climate Change,” including my “role in the Third Assessment Report” (referred to as “TAR”), and a host of information as to how TAR was prepared and how the authors of TAR verified the soundness of the data that formed the basis for the conclusions set forth in TAR.

A: As is set forth on my curriculum vitae, I was one of ten lead authors of chapter 2 of TAR, and I served as a contributing author for chapters 7, 8, and 12 of the report. Given the breadth of the project, there were two layers of editorial review that oversaw the work of the lead authors for each chapter, so the chapter reflected a consensus scientific view, not merely the views of any single author. The TAR had 672 scientist reviewers. In the United States, anyone who wanted to review the draft was allowed access to them to provide a review. I am not myself familiar with any scientific document that has been more comprehensively reviewed than the TAR.

Information concerning the “dates of key meetings,” the steps taken by “reviewers, and lead authors to ensure the data . . . were sound and accurate,” and the “identity of people who wrote and reviewed” portions of TAR should be obtained directly from the Intergovernmental Panel on Climate Change (“IPCC”). As I am sure you can appreciate, I am not an agent of the IPCC and I am not empowered to speak on IPCC’s behalf on these matters. Nor have I been authorized by the IPCC to make public information that the IPCC itself has not chosen to make publicly available. If the Committee is interested in pursuing these matters, I would urge that the Committee
contact Sir John Houghton, the head of the Working Group, at the Hadley Centre in England.

For the Committee’s convenience, I have sent along with this letter copies of key scientific articles referred to in this letter. Please let me know if you have questions.

Respectfully submitted,

Michael E. Mann, Ph.D.
Associate Professor and
Director of Earth System Science Center
Department of Meteorology
The Pennsylvania State University

---

6 I do not formally assume this position until August 22, 2005. I currently serve as Assistant Professor, Department of Environmental Sciences, University of Virginia, Charlottesville.