Chapter 4

The Making of the Hockey Stick

The scientists tell us the 1990s were the hottest decade of the entire millennium.

—President Bill Clinton, State of the Union Address (January 27, 2000)

Although scientific revolutions in how we see the world do occur, the bulk of our scientific understanding comes from the cumulative impact of numerous incremental studies that together paint an increasingly coherent picture of how nature works. The hockey stick was no different. To employ a mixed sports metaphor, the hockey stick did not suddenly appear out of left field. Rather, it arose as a logical consequence of decades of work by paleoclimate researchers that led to increasingly rich networks of climate proxy data and the introduction of new ways to use such data to reconstruct past climates. My colleagues and I were the beneficiaries of this substantial body of past work.

The Reconstruction Zone

A half-century ago, when the British climatologist Hubert Lamb set out to trace temperature trends over the past millennium, he obviously wasn't attempting to address the issue of anthropogenic climate change. It wasn't considered an issue at the time. Reconstructing past climate change was interesting in its own right, however. It was an engaging piece of puzzle-solving, the sort of thing that excites scientists. How could one deduce, from the sparse and imperfect clues available, how climate varied in the past? Could one perhaps even explain some key historical events, such as the rise or collapse of various civilizations in terms of climate changes? Big picture, gee-whiz stuff.

These sorts of questions, rather than the threat of human-caused climate change, drove dozens of paleoclimate researchers around the world to attempt to piece together, over several decades, the riddle of how Earth's climate had varied over time. By the mid-1990s, it was becoming possible to use proxy records such as tree rings and ice cores to build year-by-year chronologies of climate change at many locations around the globe, reaching back centuries and in some cases millennia. These proxy records could be used to address, for example, the question of how cold it really was during the "year without a summer" of 1816 that followed the explosive Tambora volcanic eruption of April 1815. How did that eruption influence rainfall and atmospheric circulation patterns around the world? Was there a relationship between the 1791–1792 El Niño and the deadly failure of the Indian monsoon during that period, which led to drought, famine, and the death of millions?

Thanks to decades of curiosity-driven work by paleoclimatologists, extensive global networks of long-term proxy data were available for analysis by the early 1990s. Of course, by that time, climate change itself and what it might portend for the future were far more prominent issues than they had been in Lamb's day. By featuring the Lamb curve in the First Assessment Report in 1990, the IPCC had suddenly made proxy reconstructions of past temperature policy-relevant. The Lamb curve, as we've seen, implicitly raised the issue of whether modern warming really was unusual; it seemed at the time to resolve that proposition in the negative. Like any controversial and high-profile scientific finding, the curve became a straw man for other scientists to either confirm, refine, or reject. And indeed, other scientists would soon approach the issue using more up-to-date data and more rigorous approaches.

The first truly quantitative reconstruction of past temperature changes at the hemispheric or global scale was attempted by Ray Bradley and Phil Jones in 1993.[1] Their approach was perhaps primitive by today's standards, but their contribution to our understanding was significant. The two researchers assembled a set of about two dozen proxy records representing temperature variations in distinct regions of the Northern Hemisphere (largely during the summer season), supplemented with the few long historical temperature records available. They formed a composite of the proxy records representing the average over all regions and then scaled the composite to match the scale of the modern instrumental temperature record[2] to produce an estimate of Northern

Hemisphere average temperature back in time. The Bradley and Jones reconstruction stretched back to A.D. 1500 and was adopted as the new standard in the 1995 IPCC Second Assessment Report, replacing—to the chagrin of climate change contrarians—the considerably less quantitative or reliable Lamb curve of the 1990 report.

The Bradley and Jones reconstruction did not encompass the medieval period, but it did, for the first time, characterize the extent of Northern Hemisphere average cooling during the period known as the Little Ice Age. Despite the existence of greater cooling in some regions (e.g., Europe) at certain times (e.g., the seventeenth century), the temperature changes recorded in the various proxy records were not synchronous and, in some cases, were even of opposite sign. As a result, the average cooling over the entire Northern Hemisphere at the height of the Little Ice Age was modest—less than 1°C cooler than the late twentieth century, nearly a factor of two smaller than what Lamb had originally estimated for central England.

Numerous efforts were made in the ensuing years to advance the science further. In 1995, Bradley collaborated with solar physicists Judith Lean and Juerg Beer to investigate the relative roles that both natural and human factors might have played in the long-term temperature changes documented by the Bradley and Jones reconstruction.[3] The group built on earlier studies by scientists such as John Eddy,[4] who correlated estimated changes in solar output (derived from historical sunspot measurements and radiocarbon data)[5] with Lamb's estimates of past temperature changes in an attempt to deduce the impact on temperatures of an apparent lowering of solar output during the peak of the European Little Ice Age in the latter half of the seventeenth century. During this period—termed the "Maunder minimum," after British astronomer Edward W. Maunder (1851–1928) who had first studied this anomalous period—no sunspots were observed at all, implying a substantial reduction in solar output at the time. Eddy concluded that there was a statistical connection between the Maunder minimum and coincident changes in climate, including not only cooling temperatures in Europe, but also shifting drought patterns in North America. With the data available at the time, he was unable to provide a meaningful quantitative estimate of the effect.

Bradley and his collaborators revisited the problem with aid of the more extensive quantitative estimates that were now available. They confirmed that the dip in solar output during the Maunder minimum appeared responsible for a cooling of a bit less than 0.5°C during the Little Ice Age. A modest increase in solar output might also have played a role in the early twentieth-century warming, they concluded, but solar impacts could not explain most of the warming of recent decades. Paleoclimatologist Jonathan Overpeck and collaborators reached similar conclusions in a subsequent 1997 study of Arctic temperature trends.[6]

These more recent studies still suffered some limitations. The temperature reconstructions only resolved decadal timescale variations, not individual years, so nothing could be said about, say, the "year without a summer" of 1816 or the apparent mother of all El Niño events in 1791–1792. Moreover, these studies reconstructed only a single time series representative of hemispheric mean temperatures and thus could not establish precisely which regions were warm and which cold in a given year; in other words, they didn't produce a spatial pattern of relative temperature. Finally, these studies didn't provide any margin of error, or "error bars," to indicate the uncertainty in the estimated temperature changes given the imperfect and uncertain nature of the proxy records. While the recent decades were nominally the warmest in these reconstructions, without knowing how much uncertainty there was in the estimates, it was impossible to rule out, with any degree of confidence, that past temperatures might have been as warm as today's.

It was at this point that I stepped onto the paleoclimate reconstruction scene. After defending my Ph.D. dissertation in the spring of 1996, I was funded on a Department of Energy postdoctoral fellowship[7] to continue my paleoclimate work with Ray Bradley at the University of Massachusetts (U. Mass) in Amherst, the town I grew up in. I would also be working closely with Ray's colleague Malcolm Hughes from the University of Arizona, a specialist in the use of tree ring data.

After having left for college more than a decade earlier, I was quite literally returning home. I moved into an upstairs apartment in the house I'd grown up in, just down the street from the center of town. I now constantly ran into old friends and acquaintances, including some of my old high school and elementary school teachers. My father had recently retired as a math professor from U. Mass, and we only narrowly missed being colleagues. Indeed, some of my U. Mass colleagues were parents of kids I'd grown up with. It was all a bit odd. But Amherst was a nice place to live as a "thirty-something," and I couldn't have been happier with my newfound academic home in the U. Mass Department of Geosciences. It was a lively, friendly department. And it also happened to back up on the U. Mass faculty tavern, a favorite gathering place on Friday afternoons and evenings.

The Game Begins

My postdoctoral research was aimed at developing and applying a new statistical approach to the problem of proxy climate reconstruction. Seeking to improve upon previous efforts, I wanted to reconstruct surface temperatures not just for individual decades, but also for individual years. Moreover, I was interested in reconstructing the underlying spatial patterns of temperature variation, not simply average trends over large regions like the Northern Hemisphere or the Arctic. Reconstructing these spatial patterns would not only tell us where it was warm or cold in any particular year, but also would give us insight into the workings of the climate system. It was these patterns that could tell us about the long-term role of the El Niño phenomenon, for example, or the wiggles in the track of the jet stream from year to year called the North Atlantic Oscillation (NAO). How were these patterns influenced by external factors such as volcanic eruptions and changes in solar output? It was questions such as these, rather than the effects of humanity on climate, that I was seeking to address.

There was already a rich history of using sophisticated statistical methods in certain subfields of paleoclimatology, such as dendroclimatology, the study of tree ring data to infer past climate change. Over the previous few decades, researchers such as Hal Fritts, the godfather of dendroclimatology, Ed Cook of Columbia University's Lamont Doherty Earth Observatory (LDEO), and Keith Briffa of the University of East Anglia's Climatic Research Unit had developed various approaches to reconstructing patterns of temperature, rainfall, surface pressure, and drought from tree ring data. In dendroclimatology, one is often lucky enough to be dealing with a set of proxy data with similar attributes, such as a network of precipitation-sensitive tree ring records from, say, the western United States. And there were well-established methods for relating the patterns in the tree ring data to the patterns of climate.

A climate reconstruction could thus be formed by establishing a statistical relationship between the two datasets over their common years of overlap (typically the twentieth century), called the "calibration" or "training" period, which could then be projected back in time for periods long ago when instrumental climate records were not available.[8]

The circumstances of the problem my collaborators and I were attacking were somewhat different. Unlike the typical situation encountered in past work, we were making use of diverse proxy data; our dataset included regional networks of tree ring data, but it also included data from ice cores, corals, and lake sediments, and a smattering of historical climate records. It wasn't appropriate to lump all of the proxy records together. We had far more tree ring data than other types of proxy records, yet the tree ring data represented only a restricted region of the globe, the midlatitude continents. The smaller amount of data drawn from corals, ice cores, and lake sediments represented the other key regions: the oceans, the tropics, and the poles. Allowing the sheer amount of tree ring data to overwhelm the less abundant information from other proxy records would weight our results primarily toward the midlatitude continents, whereas we were seeking to reconstruct patterns over the entire surface of the globe, land and ocean, from equator to pole.

We used a statistical procedure called "principal component analysis" (PCA) to get around the problem. PCA can be used to represent a very large dataset (be it modern thermometer-based temperature measurements distributed over the globe or a set of tree ring records reflective of temperatures distributed across a region such as North America) in terms of a small number of patterns in space and time that describe the most variation in the data. Each pattern can be described by a combination of its temporal history (the principal component or PC time series) and its spatial variations (the empirical orthogonal function or EOF).

This widely used statistical tool would become a major bone of contention among hockey stick critics, so it is worth a special effort to understand how and why it is used. To that end, let us consider a simple synthetic example (see figure 4.1 on page 46) where only two regions make up the world: the hemisphere west of the Greenwich meridian (the west) and the hemisphere east of it (the east). Temperatures, furthermore, are always uniform within each of the two regions. In our example, surface temperatures overall warm in the course of a century (top row), but they fluctuate differently for the west (panel a) and east (panel b). These fluctuations can be measured relative to the average temperature over the century as a whole—a number that defines the zero of the temperature scale (*y*-axis) and, thus, determines the vertical centering of the data. In the west, temperatures start out cold, about 1 degree below the long-term average. There is substantial warming in the century's first decades, bringing temperatures nearly 0.5 degrees above the long-term average by midcentury. That above-average warming is offset by slight cooling during subsequent decades, bringing temperatures back down roughly to their long-term average by century's end. In the east, we see the converse of this pattern. Slight cooling occurs during the first half-century, when temperatures spend much of the time below their long-term average. But that cooling is more than overcome by a substantial warming during the second half, with temperatures ending up nearly 1 degree above the long-term average by the end of the century.

The variations in surface temperature over the globe as a whole can be fully described mathematically in terms of just two spatiotemporal patterns of variation in the data. One of these two patterns (bottom row) is characterized by a linear increase in temperature over the hundred-year period of just under 1°C in amplitude. In this characterization, the warming is uniform across both west (panel e) and east (panel f) and is described by a PC series that is a simple upward ramp (for purposes of this example, you could think of this pattern as global warming).[9] The other pattern (middle row) is a roof-shaped temperature variation of 1°C amplitude that plays out oppositely in the west (panel c) and east (panel d): The west warms during the first half-century and cools during the later half-century. The east does just the opposite, yielding what looks like an inverted roof. This second pattern in the temperature record can be described, for the west hemisphere, by a PC series that increases toward the middle of the century and decreases thereafter, with precisely the reverse evolution for the east.[10] This pattern could be, say, the imprint on the global temperature record of a single cycle of a multidecadal climate oscillation like the Atlantic Multidecadal Oscillation (AMO) described in chapter 3, which acts to redistribute heat from one part of the globe to another (e.g., from west to east), but doesn't change the average temperature of the globe. Adding together the two patterns yields the total pattern of temperature variation (i.e., top row; adding c and e, gives a, while adding d and f gives b).

We can rank the two contributing patterns by what fraction of the total variation in the overall data (temperature, in this case) they explain. These rankings will depend, however, on what baseline we choose for the data, that is, how we center the data with respect to the *y*-axis. It is conventional to center the data about their long-term average, as we have done in the above example. By this convention, the relative temperature departures average to zero over the full hundred-year dataset. With the data centered that way, the second of the two patterns (c and d) of warming/cooling (in the west) and cooling/warming (in the east) explains 60 percent of the variation and thus constitutes PC#1. The global warming pattern (e and f) in our example turns out to be the second most prominent pattern, explaining 40 percent of the variation in the data. It is PC#2.

In more realistic examples, there are generally a greater number of significant patterns of variation in the data. Moreover, rarely can they be extracted as cleanly as in this example, since there is typically some degree of contamination by noise, be it random disorganized temperature variations or biases and errors in the data. This example nonetheless demonstrates how PCA can efficiently describe the few leading patterns of variation in a larger dataset, a step that is essential if one is interested, as with paleoclimate reconstructions, in establishing robust relationships between patterns in potentially quite large and noisy datasets. Applying PCA in that case can help sort out the climate signals (the key, most robust patterns of variation in the datasets) from the noise in which they are immersed.

Using PCA, we were able to represent the main information in each of the various regional networks of tree ring data (e.g., North America, Eurasia) in terms of a small number of representative PC series. With the tree ring data that dominate the midlatitude continents[11] represented by a few leading patterns, the handful of coral series from the tropics and ice core series from the Arctic would have commensurate representation in the dataset. As far as each proxy's opportunity to help tease out the past patterns of climate variation was concerned, it was now a fair fight.

Our statistical method established a relationship between the proxy data (which extend several centuries back in time) and the modern surface temperature record during the period of the twentieth century where both datasets overlap. This procedure required finding a way to

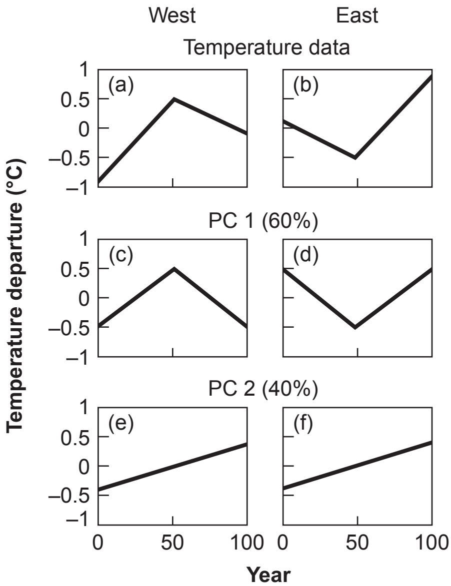

Figure 4.1: PCA Example: Spatial and Temporal Variations in Temperature Data This example shows how PCA can be used to characterize efficiently a global temperature dataset (top) in terms of a small number of patterns in space and time. In this case, just two patterns (middle and bottom) characterize the data. The horizontal axis is time in years, and the vertical axis depicts the relative departure of temperature in degrees from the average over the baseline, 0 (in this case, the average over the full hundred-year period).

represent efficiently the information from the modern instrumental record. Spanning a century, the instrumental surface temperature dataset contained more than a thousand months of data at more than a thousand locations around the globe.[12] But only a handful of patterns could potentially be captured by our noisy proxy dataset. The temperature data thus also needed to be represented in terms of a modest number of its most prominent underlying patterns. We recognized that the same tool we had used to simplify the representation of dense tree ring data networks—PCA—could again prove useful. Using PCA, we could represent the key information in the instrumental temperature dataset with just a dozen or fewer distinct patterns. The leading temperature pattern related to the overall warming of the globe, while subsequent patterns related to phenomena such as El Niño, the North Atlantic Oscillation, and the Atlantic Multidecadal Oscillation.[13]

With the aid of our statistical method, we simply needed to figure out the combination of these dozen or so temperature patterns that most closely matched the behavior of the proxy series we had during the twentieth-century period of overlap. Once those relationships had been established, they could be tested through a process known as "validation" or "verification." In common scientific parlance, these terms describe the process of independent confirmation of a previous finding. In the present context, unlike say, performing a new experiment that backs up previously reported results, one is instead seeking to demonstrate that a statistical model for a phenomenon can successfully predict independent data that were not used in establishing the model in the first place. In other words, it is a way of demonstrating that the statistical relationship is real, and not just a fluke of statistics. Rather than using the full available instrumental record for the calibration of the proxy data, one leaves some subinterval of the instrumental record aside for testing purposes. The proxy data are then calibrated over the shortened calibration interval, and the resulting statistical model is used to predict variations in climate over the remaining subinterval that was not used in the calibration process. Since no information from that part of the instrumental record was used to calibrate the proxy data, the extent to which the climate variations predicted by the statistical model match the climate changes that were actually observed over that time interval provides a true test of the reliability of the climate reconstruction.

This process of validation or verification is essential in establishing the credibility of a proxy-based climate reconstruction. There is, however, an important compromise that must be struck: The shorter the calibration period, the less robust the resulting statistical model, but the shorter the validation period, the less reliable the independent validation test.[14] The information from the calibration and validation process can also be used to estimate statistical uncertainties in the reconstructions, yielding the important margins of error (variously referred to as "error bars" or "confidence intervals") that characterize the envelope of uncertainty surrounding the climate reconstructions.

The instrumental data become increasingly sparse as one goes back in time before the twentieth century; indeed, only a handful of records are available as far back as the early nineteenth century. On the other hand, proxy data are available only up through the time they were collected. Many of the key proxy records were obtained during the 1970s or early 1980s, and proxy data become increasingly sparse after that period. We consequently chose the data-rich interval of 1902 to 1980 for our calibration period, leaving aside the earlier nearly half-century of sparser instrumental data for the validation tests.[15] These tests established that the statistical reconstructions were skillful—that is, meaningful in a statistically rigorous sense[16]—as far back as A.D. 1400, but not earlier than that.

Onto the World Stage

When we initially wrote up our results for publication, we focused on what we felt was most scientifically interesting, for example, that we recovered an unusual pattern for the 1816 "year without a summer" that indicated a very cold Eurasia and lower than average temperatures in North America (observations that are independently confirmed by historical accounts), but a warmer than usual Middle East and Labrador (who knew?). Or that we had independently affirmed anecdotal accounts that there was a whopper of an El Niño event in 1791—a year that, according to our reconstruction, also happened to be a comparative scorcher for Europe and a large part of North America. Then we did the least scientifically interesting thing one could possibly do with these rich spatial patterns: We averaged them to obtain a single number for each year: the Northern Hemisphere average temperature.

That single aspect of our work got all the attention. In truth, it did also represent an advance. Compared to previous Northern Hemisphere temperature reconstructions, ours went further back, it resolved individual years rather than just decades, and it had error bars associated with it. Thus, we were able to draw the specific conclusion that, even taking into account the margin for error, three recent years—1990, 1995, and 1997—all appeared warmer than any other in the past six centuries.

Our study, which has come to be known as "MBH98" for the authors—Mann, Bradley, Hughes—was published in *Nature* on April

22, 1998—Earth Day.[17] I was caught completely off-guard by the amount of media attention the article received. Generally, one is lucky to get a nibble or two from the local media in response to press releases on a published scientific paper. This time was different. No sooner had the press releases gone out (one from U. Mass, another from *Nature*, and a third from the National Science Foundation) than the phone calls began coming in nonstop. Our study was written up in the *New York Times*, *USA Today*, *Boston Globe*, and a host of other major U.S. newspapers. Articles soon appeared in *Time* magazine and *U.S. News and World Report*. We even made it into *Rolling Stone* (though not the cover). I was asked in one afternoon to do television interviews with CNN, CBS, and NBC. In the CBS interview, John Roberts put it to me bluntly: "So does this prove humans are responsible for global warming?" He repeated the question at least three times during the interview, clearly not having gotten the money quote he was fishing for. I wouldn't take the bait. I repeated that our results were "highly suggestive" of that conclusion, but I wouldn't go further than that. I well knew that establishing that recent warming is anomalous in a long-term context alone did not establish that human factors were responsible for it. Any conclusion about *causality* required the use of climate models to estimate the relative contributions of the various factors, including human increases in greenhouse gas concentrations, hypothesized to be responsible for the observed changes.

There are several reasons that our paper might have received more than the usual expected attention. The globe had experienced record-breaking warmth that winter. The first three months of 1998 were the warmest on record (in western Massachusetts, where I was living, it barely felt like we'd had a winter). Temperatures had likely been spiked, to some extent, by a fluke of nature, a particularly large El Niño event. It was in part due to this fluke that the 1998 record for the hottest year in the instrumental record had still not unambiguously been broken by 2011 (one group has 2005 and 2010—in a statistical tie—narrowly beating it out, but another group has 1998 still holding the title).[18] In any case, 1998 was as of that date the warmest year in the instrumental record; with the advent of our study, "warmest on record" meant not just in 150 years but in at least 600!

That our paper coincidentally happened to be published on Earth Day no doubt gave journalists an extra news hook to cover the study.

Some commentators attached an almost diabolical significance to the timing, as if *Nature* was somehow conspiring with the world's environmental activists. The truth is much less interesting; the publication date at *Nature* is determined by the date of a paper's final acceptance and placement in the journal's publication queue.

Extending the Handle of the Stick

The original MBH98 hockey stick had a comparatively short "handle,"[19] extending back six centuries, and did not reach back far enough to establish whether a medieval warm period actually existed, let alone when it might have ended or begun. Some went so far as to attack us for only going back six centuries, claiming we had intentionally stopped short to avoid running into the putative medieval warm period.[20] We were guided, of course, only by what the validation tests had objectively indicated: that with the data we had, we couldn't go back any further and still obtain meaningful estimates. Why would anyone impugn our integrity so brazenly and make an allegation so false? Boy, would I later be in for a lesson.

Shortly after our study was published, Phil Jones and collaborators published another attempt to trace the Northern Hemisphere's mean temperature. They used the same composite approach as had Bradley and Jones in 1993, but this time they extended the estimate back over the entire past millennium.[21] We were a bit skeptical. Our own statistical tests, after all, had told us that it was not possible to obtain a meaningful reconstruction further back than six centuries, using more or less the same proxy data.[22] Nonetheless, my colleagues and I decided it was worth taking a closer look.

We decided to examine more closely the two dozen proxy records we had that extended back six centuries or more, which included those used by Jones and colleagues, and several others. I performed a series of so-called sensitivity tests, in which various proxy records are removed or—to use standard statistical terminology, "censored"—from the network, and the sensitivity of the results to those records is gauged by noting how much of an effect their removal has on the result. I titled the computer directory "censored" accordingly—a choice I would later regret.

The tests revealed that not all of the records were playing an equal role in our reconstructions. Certain proxy data appeared to be of critical importance in establishing the reliability of the reconstruction—in particular, one set of tree ring records spanning the boreal tree line of North America published by dendroclimatologists Gordon Jacoby and Rosanne D'Arrigo.[23] These records didn't extend any further back than A.D. 1400, however, and that's when we could no longer achieve a reliable reconstruction. This result actually made perfect sense since the pattern of twentieth-century warming in the instrumental record showed particularly strong warming in western and northern North America—precisely where these tree ring data were located. Moreover, a previous analysis of long climate model simulations had suggested that western North America was a "sweet spot" for estimating the average temperature of the Northern Hemisphere.[24]

We had other proxy data from this region, and they extended considerably farther back in time, in many cases more than a thousand years. There was a group of tree ring records from high-elevation sites in the western United States (primarily California, Nevada, and Arizona) that should in principle have reflected temperature variations in the region. Yet, despite the availability of these data, our validation tests were telling us that we couldn't go further back than six centuries. Intrigued by this apparent inconsistency, I undertook a closer comparison of the two different North American datasets.

Something rather remarkable emerged when the two datasets were laid on top of each other. The two tracked each other almost perfectly from the beginning of their period of overlap (A.D. 1400) until the early nineteenth century. Then the two series began to diverge, with the western U.S. series showing an almost exponential increase in tree ring growth relative to D'Arrigo and Jacoby's boreal tree line series. Then, at the beginning of the twentieth century, the two series began tracking each other again. As it turns out, the likely reason for this enigmatic observation had already been discussed in the scientific literature.

It was known that high-elevation trees are often not limited in growth simply by climate conditions such as growing season warmth, but also by carbon dioxide levels. Trees need carbon dioxide for photosynthesis, but immersed in the thinner atmosphere of high elevations they may be somewhat starved of this resource. In a 1993 article, Graybill and Idso had shown that these very trees might be expected to exhibit a positive growth response to increasing atmospheric carbon dioxide levels.[25] This so-called CO_2 fertilization mechanism could explain the divergence between the growth rates of the high-elevation western U.S. trees and the low-elevation boreal tree line stands: The timing of the divergence was almost perfectly correlated with the exponential rise of atmospheric carbon dioxide since the early nineteenth century associated with the Industrial Revolution. The disappearance of the divergence in the twentieth century was consistent with warmth once again returning as the key factor controlling the growth of the high-elevation trees in the presence of adequate carbon dioxide.

By correcting for that carbon dioxide effect through comparing otherwise similar trends in low- and high-elevation temperature-sensitive North American trees,[26] it seemed we might now be able to make use of the far-longer-term western U.S. data. Indeed, when we used the corrected version of the western U.S. tree ring data in our analysis, our validation tests gave us the green light; we could indeed now meaningfully reconstruct Northern Hemisphere average temperatures over the entire past millennium.

Despite this success, we were rather guarded in our conclusions, recognizing that our ability to obtain a reliable millennial Northern Hemisphere temperature reconstruction relied heavily on those western U.S. tree ring data as well as our somewhat *ad hoc* correction for potential CO_2 fertilization effects. The abstract of our article, entitled "Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations," stated: "We focus not just on the reconstructions, but the *uncertainties therein*, and *important caveats*," and "Though *expanded uncertainties* prevent decisive conclusions for the period prior to A.D. 1400, our results *suggest* that the latter 20th century is anomalous in the context of at least the past millennium" (emphasis added). The expanded uncertainties were a result of additional cross-checks that indicated that the reconstruction was less effective at capturing century-scale and longer variations in earlier centuries, which led us to increase the error bars relative to MBH98. Even with these larger uncertainties taken into consideration, the reconstruction indicated that the 1990s were likely the warmest decade and 1998 likely the warmest year of the millennium.[27]

Our new paper[28] (henceforth "MBH99"), published just under a year after our original *Nature* article, once again garnered a fair amount of media attention, including a major spread in the Tuesday science section of the *New York Times*. But perhaps because it represented a cautious and incremental development relative to our earlier work and appeared in the lower-profile journal *Geophysical Research Letters* (and wasn't published on Earth Day), it didn't make quite the same splash, at least at the time. The article, however, would soon gain the attention of climate change contrarians, who perceived that our findings undermined one of their primary arguments against human-caused global warming: that a period of warmth comparable to that of the present day had existed in the relatively recent past, prior to any appreciable increase in greenhouse gas concentrations.

The Third Assessment Report

In order to understand the rise to prominence of the hockey stick, it is necessary to delve into the history behind the IPCC Third Assessment Report published in 2001. It was in that publication that our work truly entered onto the world stage. There is a pyramid-like hierarchy of IPCC report authorship. For any given chapter, there are dozens of contributing authors, fewer than a dozen lead authors, and, among the latter, two convening authors who have primary responsibility for the chapter. While I had been nominated to be an author of some sort prior to the publication of our work on paleoclimate reconstructions, I was still relatively fresh out of graduate school, and I was surprised when I learned in late 1998 that I had been selected as a lead author for the new report—a choice presumably related to the scientific attention our recent work had received.

My task as a lead author was to work with the numerous contributing authors in assessing the state of knowledge regarding evidence from the paleoclimate record, which—as in the previous IPCC reports—was part of the "observations" chapter. The first job was to solicit input from the leading experts in the field, including Keith Briffa and Phil Jones of the University of East Anglia's Climate Research Unit, who had performed their own proxy reconstructions of past climate, and Henry Pollack, a researcher at the University of Michigan who had derived an independent assessment of past temperatures from boreholes.[29] Boreholes are deep holes in the ground generated for purposes of geophysical exploration. Under the right circumstances (when there is minimal intrusion of fluid flow, the properties of the bedrock are reasonably uniform, changes in seasonal snow cover are minimal, etc.), the penetration of heat from the surface down into the upper layer of Earth can reasonably be assumed to follow a simple diffusion process. The physics of diffusion dictates that the vertical temperature profile down the hole can then be used to provide an estimate of how temperatures at the surface changed back in time.

To be credible as a true assessment of prevailing scientific understanding, IPCC reports must accurately reflect the diverse views within the scientific community on any particular issue—no simple task when it comes to contentious disciplines such as climate change. In addition, since the IPCC report is an assessment report, not simply a literature review, it is necessary that the IPCC evaluate the collective work of the scientific community in such a way that it is clear to readers on what points scientists agree, and where there is still active debate.

There were several rounds of discussion among the lead authors regarding each of the various drafts of our chapter. The lead authors included scientists with a wide range of views such as John Christy, a scientist with a somewhat contrarian outlook on climate change. In the end, whatever was said in the chapter would have to meet with John's approval. Moreover, whatever would be concluded in our chapter would have to withstand the scrutiny of the rigorous IPCC review process.[30]

My primary responsibilities involved the assessment of the paleoclimate records of past centuries.[31] It would be essential in this work to reflect the diverse range of views in the recent scientific literature and to present the range of estimates that had been published. Any key conclusions arising from the chapter could not rely on one study or one group's findings. They had to reflect a consensus of recent studies, if indeed that existed. The MBH99 hockey stick was shown in a plot by itself for two important reasons: (1) It was the only reconstruction done at the level of individual years rather than decadal or longer-term averages, and (2) it came with error bars, which the other reconstructions didn't. Thus, unlike other studies, it spoke to whether recent years, such as 1998, stood out as unusual against the backdrop of the longer-term reconstruction and its uncertainties.

An additional plot in the chapter compared three different independent or largely independent quantitative reconstructions of Northern Hemisphere temperatures over the past millennium. One was an estimate by Keith Briffa and collaborators based on tree ring density (as opposed to ring width) measurements that extended back six centuries,[32] while the other two—each extending back the full millennium—were the Jones et al. multi-proxy estimate discussed earlier in the chapter and the MBH99 hockey stick. Each showed recent warming to be anomalous in the context of the longer-term temperature history of the record. Another reconstruction of Northern Hemisphere temperatures over the past millennium by Crowley and Lowery was published too late to be included in the comparison figure, but its conclusions, similar to those of the other three reconstructions, were summarized in the IPCC chapter.[33]

Yet another figure in the chapter depicted the Pollack et al. estimate of past temperature change from borehole records, an estimate that was entirely independent of the other proxy estimates. Though the borehole temperature reconstructions came with their own caveats, they provided independent evidence that recent warming was unusual in at least a five-hundred-year time frame—as far back as reliable borehole temperature estimates went. A separate plot in the chapter showing glacier mass balance records spanning the past five hundred years indicated that the melting of mountain glaciers worldwide too was unprecedented over at least this time frame.

Collectively, the data clearly indicated that the modern warming was unprecedented in a long time frame. But only two estimates (the Jones estimate and the MBH99 estimate) gave a quantitative depiction of hemispheric-scale temperature changes for the full millennium, and only one (the MBH99) had error bars attached to it. After much discussion among all the lead authors, a consensus was reached on a tentative conclusion. The word *likely*, the group decided, would be attached to the conclusion that recent warming for the Northern Hemisphere on the whole was anomalous in a millennial context. In the parlance of the IPCC, this careful phrasing indicated confidence of about 67 percent, that is, a two-out-of-three chance that the conclusion was correct. That is a far cry from the 90 percent threshold required for "confident" inferences—what in IPCC parlance was referred to as "very likely." Like MBH99 itself, the IPCC Third Assessment Report was extremely cautious in the level of confidence it attached to its conclusion that recent warming was anomalous in a millennial context.

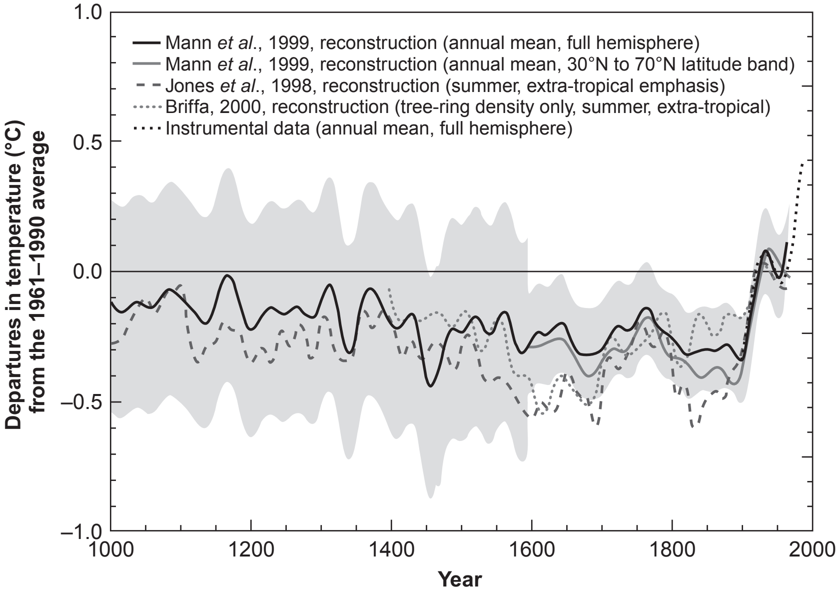

Figure 4.2: Competing Estimates of Millennial Temperature Trends
The graph compares different reconstructions of Northern Hemisphere temperature trends over the past thousand years as shown in the 2001 IPCC *Third Assessment Report*. Shown for comparison is the instrumental record from the mid-nineteenth to late twentieth centuries. The data have been smoothed to emphasize multidecadal and longer timescale variations.

The Implications

Though we didn't quite realize it at the time, the gauntlet had been laid down with the initial publication of the extended MBH99 millennial hockey stick and, especially, its subsequent prominence in the 2001 IPCC report. We had taken on a sacred cow of climate change contrarianism: the medieval warm period (MWP). Our reconstruction did not "eliminate the MWP," as our detractors liked to claim.[34] It did in fact include a period of relative warmth during the medieval era of the eleventh to fourteenth centuries. While medieval conditions were relatively warm, however, the modern tip of the blade—as can be seen in the graphic featured in the prologue—was warmer than the peak reconstructed medieval warmth.

One pillar of the contrarian case against human-caused climate change was that the mere existence of a warmer period centuries earlier somehow disproved any human influence on modern warmth. In reality, this was not true. Scientists had known for some time, for example, that there were periods in the deep geological past during which temperatures were warmer than today, such as in the mid-Cretaceous period 100 million years ago. The reason was atmospheric CO_2 concentrations that were several times higher than today owing to the slow geological processes that modify atmospheric composition on very long timescales. It was indeed possible that other natural factors, be they changes in solar output or volcanic activity, could have led to conditions that were as warm as today.

Whether conditions in past centuries might have been warmer than today, then, would not have a scientific bearing on the case for the reality of human-caused climate change. That case, as we've seen, rests on multiple independent lines of evidence: the basic physics and chemistry of the greenhouse effect, for example; the relationship between greenhouse gas concentrations and temperatures over geological time; and a pattern of observed climate change that can only be explained by climate models when human-produced greenhouse gases are included in the calculations. While the presence of a medieval warm period warmer than today would not negate the reality of modern human-caused climate change, evidence of its absence nonetheless would take away an important (if misguided) talking point of contrarians—something of which they were well aware.

Our finding that recent warming was anomalous in a long-term (now, apparently, millennial) context was suggestive of the possibility that human activity was implicated in the warming. I was always very careful not to claim that our work could firmly establish a human role in the warming. To draw such a conclusion based on our work alone would necessarily buy into the classic logical fallacy of "correlation without causation." We had established correlation—the anomalous warming that we documented coincided with the human-caused ramp-up in greenhouse gas concentrations—but we hadn't established causality.

A little more than a year after we had published our millennial hockey stick reconstruction, paleoclimatologist Thomas Crowley of Texas A&M University (and coauthor of the Crowley and Lowery reconstruction discussed earlier) published findings based on the use of a theoretical climate model simulation designed to investigate causes of past temperature change.[35] Crowley subjected the model to estimated changes in natural factors over the past thousand years using indirect measures of changes in solar output and explosive volcanic activity, information on both of which can be recovered from atmospheric deposits in polar ice cores. These simulations revealed that the natural factors could explain the extent of medieval warmth in our reconstruction; in the model, this warmth arose from a relative lack of cooling volcanic eruptions combined with relatively high levels of solar output. The natural factors could also explain the cooler conditions of the ensuing Little Ice Age, which resulted from relatively low levels of solar output and more frequent explosive volcanic eruptions. Fed the natural factors only, the model could not, however, reproduce the abrupt twentieth-century warming. In fact, the model predicted that the climate should have cooled in recent decades, rather than warmed, if only natural factors had been at play. It was only when Crowley added the modern human influences—increasing greenhouse gas concentrations primarily from fossil fuel burning and the regional cooling effect of industrial sulfate aerosols emissions—to the model simulation that it was able to track the hockey stick all the way through to the present. The conclusion was clear: Natural factors could explain the temperature changes of the past millennium through the dawn of the Industrial Revolution, but only human influences could explain the unusual recent warming. Causality was at least tentatively established now.