11 Data Wars

Climate is essentially the history of weather, averaged over time. So one might be forgiven for assuming that climate *data* are simply weather data, averaged over time. After all, we are talking about the same primary variables (temperature, pressure, wind, etc.), measured by the same instruments at the same places, usually by the same people, transmitted through the same communication systems, under standards set by agencies of the same organizations.

Yet in the course of writing this book I was regularly met with blank stares, puzzlement, and hostile comments. "You're talking about *weather* data, not *climate* data," some interlocutors said, apparently offended on some deep level that I would dare to do such a thing. It took me a long time to understand the reason for their discomfort. An outsider, I had walked onto the central battlefield of a long, low-level war. Meteorologists are perhaps the nicest people you will ever meet; they fight quietly and politely. But the war is no less real, and the stakes are high.

Over decades, climate scientists have accumulated a long litany of frustrations with the overwhelming focus of the global observing system on forecasting. Many of them feel that operational agencies mostly ignored climatology's needs until quite recently, managing instruments and handling data in ways detrimental to the long-term climate record. As a result, the colossal quantities of weather data mostly cannot be used by climatologists, who are instead forced to rely on the much smaller "climate quality" data sets they have collected themselves. Since the 1980s, a series of high-level review commissions and WMO programs have repeatedly argued the case for more consistent data practices, better documentation and calibration of changes, and climate-relevant satellite instrumentation in the overall data system.[1] Some weather scientists now think the tables have turned. Burgeoning budgets for climate-change research, they believe, have stripped funding from other priorities, such as the forecasting of severe weather. Since hurricanes, tornadoes, and similar phenomena regularly kill people and cause extreme damage, they bitterly complain, forecasting of severe weather should receive at least as much priority as climate change.

So exactly what *is* the difference between weather data and climate data, and why does it matter so much? Does the difference *make* a difference? How does it affect knowledge of climatic history, validation of climate models, and forecasts of the climatic future? Can computer models make "weather data" into "climate data"?

The distinction stems from important differences between forecasting and climatology. These concern, especially, the following:

- the overriding *purposes* of data collection
- the major *priorities* governing data collection
- which *data sources* fit those priorities and purposes
- how to assess and control *data quality*
- the degree to which *computation* is centralized or distributed
- how each field *preserves data* from the past.

As we have seen throughout this book, observing systems and standards changed often and rapidly over time, creating temporal discontinuities and inconsistencies. These "inhomogeneities," as meteorologists call them, rendered large quantities of weather data unusable for climatological purposes. Meanwhile, until the late 1980s, when climate politics became a major public issue, climatologists' pleas for greater attention to their data needs fell mostly on deaf ears.

Yet in recent decades, with fast, high-quality 4-D data assimilation, forecasting and climatology have begun to converge. Reanalysis of global weather data is producing—for the first time—consistent, gridded data on the planetary circulation, over periods of 50 years or more, at resolutions much higher than those achieved with traditional climatological data sets. Reanalysis may never replace traditional climate data, since serious concerns remain about how assimilation models "bias" data when integrated over very long periods. Nonetheless, the weather and climate data infrastructures are now inextricably linked by the "models of data" each of these infrastructures requires in order to project the atmosphere's future and to know its past.

The next two sections sketch how scientists understand weather data and climate data. The rest of the chapter examines how the historical distinction between these two forms of data is changing, producing a proliferation of data images of the atmosphere's history.

Weather Data

When scientists talk about "weather data," they mean the information used in forecasting. For this purpose, speed and well-distributed coverage are the highest priorities. In today's World Weather Watch (WWW), synoptic data from the Global Observing System (GOS) are transmitted via the Global Telecommunication System (GTS) to the Global Data Processing and Forecast System (GDFS). GDFS models first analyze the data, filling in values for voids in the observing system, then create forecasts from the analysis. Once analyzed, the original sensor data are mostly archived in vast tape libraries. Since only a few facilities have the connectivity and the computer power to receive and analyze global sensor data in real time, the analyses produced by those facilities matter much more than the raw sensor signals used to produce them.

At this writing, the Global Observing System includes roughly 15 satellites, 100 moored buoys, 600 drifting buoys, 3000 aircraft, 7300 ships, 900 upper-air (radiosonde) stations, and 11,000 surface stations. The GOS's core remains the synoptic network, consisting of six Regional Basic Synoptic Networks (RBSNs) of land-based stations reporting in real time (ideally, eight times a day). This network structure is tightly linked to numerical modeling: "GOS requirements are dictated to a large degree by the needs of numerical [forecasting] techniques."[2] The World Meteorological Organization's goal for the RBSNs has hovered at about 4000 stations ever since the WWW's inception in 1968. However, that goal has been frustrated by the failure of Africa, Central America, South America, and parts of Asia to expand their networks. In 2004, only 2836 RBSN stations met the full reporting goal, a decline of more than 200 stations since 1988.[3] The crucial upper-air network also shows a troubled picture. Station implementation peaked in the early 1990s before dropping off again, with coverage of the southern hemisphere reaching only about 50 percent of the WWW's goal. (Because upper-air stations launch "consumables"—radiosondes and rawinsondes—they cost more than simple ground stations.) Further, more than 30 percent of all weather-station reports are never used in forecasting, having failed to reach forecasting centers within 2 hours. The number of unavailable reports reaches 65 percent in some regions.[4]

Satellites are expensive in absolute terms. Merely launching one into orbit can cost \$75 million–\$400 million. A typical mission (including the launch, the satellite, insurance, and ground monitoring over the satellite's lifetime of 3–10 years) costs between \$500 million and \$1 billion. But since a single polar-orbiting satellite can survey the entire planet twice a day, such satellites actually cost little relative to the gigantic quantities of information they provide. Sending up satellites also requires far less *political* effort than expanding the surface network, since it can be done by a single country or by a regional consortium such as the European Space Agency. For these reasons, meteorologists have persistently sought more and better satellites to make up for continuing deficiencies in GOS surface and radiosonde observations. These factors have reduced incentives to build out the surface network, even though many meteorologists would still like to see that happen.

Today satellites provide about 98 percent of the roughly 75 million data items evaluated by the European Centre for Medium-Range Weather Forecasts during every 12-hour forecast period.[5] Not all of these data are actually used by the system, however. In particular, only about 5 percent of satellite radiance data enters the analysis. Even after this filtering, however, satellite data still outnumber—by ten to one—those from all other instruments combined. Nonetheless, recent studies confirm the importance of the terrestrial network (surface plus upper-air). In the ocean-dominated southern hemisphere, where the terrestrial network is small and poorly distributed, satellites improve forecast quality dramatically. But in the extra-tropical northern hemisphere, where the terrestrial network is densest (and where most of the world's people live), satellite data provide only modest improvement over the terrestrial network alone.[6]

Throughout the NWP era, forecasters sought to improve the quality of forecasts as quickly as they could. They introduced new instruments, new mathematical techniques, and new computer models into the observation-analysis-forecast system as soon as they could be shown to work better than their predecessors. As a result, the observing system changed constantly, with new satellite radiometers, radiosondes, precipitation gauges, thermometers, and many other instruments added to the mix. Meanwhile, analysis and forecast models also evolved swiftly. The ECMWF, for example, produced its first operational forecast in 1979. That center soon formalized a continuous revision procedure, testing each new model by running it in parallel with the previous version for several months before replacing the old model with the new one. At this writing, the ECMWF model had already completed 32 revision cycles—on average, more than one revision each year. These kinds of changes in observing systems and forecast models never took place all at once or globally. Instead, national meteorological services managed their own changes and schedules. As we will see in more detail below, the frequent and rather chaotic revisions of observing systems and analysis processes created tremendous difficulties for climate scientists.

The World Meteorological Center system, developed for the World Weather Watch, centralized the computation of global weather forecasts. The weather stations that generate the original sensor data perform virtually no computation; they simply report instrument readings. Indeed, today more than 3000 of the roughly 11,000 surface stations in the Global Observing System are automated. Unstaffed, they broadcast readings of pressure, temperature, winds, and other variables by radio, Internet, satellite uplink, and other means. Under the WWW, stations in the synoptic network transmit reports to the WMCs (Moscow, Melbourne, and Washington) via the Global Telecommunication System (GTS). The Global Data Processing and Forecast System (GDFS),[7] consisting of the WMCs and a handful of specialized centers such as the ECMWF, process the incoming data with their global analysis and forecast models. These centers forward their analyzed global data and their forecasts to national meteorological centers, which may subject them to further processing in order to generate their own forecasts.

Forecasters historically put low priority on preserving raw data. Even in the era of empirical forecasting, forecasters kept vast libraries of analyzed charts, but rarely attended to the fate of the original instrument readings on which the charts were based. Occasionally the forecasting community has adopted some particular set of original sensor data as a benchmark, to test the performance of analysis and forecast models. For example, modelers and forecasting centers mined data from the 1979–80 Global Weather Experiment for well over a decade, using those data to refine analysis and forecast models. Similarly, data about particular hurricanes and other extreme events sometimes acquire the status of benchmarks, used to compare, test, and improve regional forecast models. For the most part, however, once the daily forecast cycle ends, forecasters work only with the *processed* data, i.e. the analysis. Original sensor data may or may not be stored; usually they are never used again. Sensor data that take more than 12 hours to arrive—such as a ship's weather log, data submitted by mail, or even ordinary weather station data not reported on time owing to a power failure or other problems—are never incorporated into the weather data record. For this reason, some have said that four-dimensional data assimilation is actually "three and a half dimensional." In other words, such systems can ingest data that arrive up until the cutoff point for a given analysis period, but they cannot (of course) take account of near-future data that might correct the analysis further. Reanalysis permits true four-dimensional assimilation, in which future observations as well as past observations can influence the state of the analysis at any point in time.[8]

Climate Data

The purposes, priorities, sources, and character of climate data contrast with those of weather data. The purpose of climate data is to characterize and compare patterns and trends. This requires statistics—averages, maxima, minima, etc.—rather than individual observations. And climate scientists care more about measurement quality, station stability, and the completeness and length of station records than they care about the speed of reporting. For example, whereas up to 35 percent of weather station reports are never used in forecasting because they fail to reach forecast centers within 2 hours, climatologists often can obtain much higher reporting ratios.

Climatologists use many of the same data sources as forecasters, but they also use many others. Certain kinds of data, such as precipitation measurements or paleoclimatic proxies, are crucial to climatology yet have little relevance to forecasting. Conversely, some kinds of data useful in forecasting play little or no role in climatology. For example, Doppler radar revolutionized daily precipitation forecasting, but the data it produces are of little interest to climatologists.[9]

Among the best regional data sources is the US Cooperative Observer Network, founded in 1874 under the Army Signal Service. The network comprises more than 5000 "full climatological" stations, and 7000 "B" and "C" stations that record mainly hydrological and other special-purpose data. Cooperative Observer Network stations are run by private individuals and institutions, including university research centers, reservoirs, water treatment plants, and agricultural businesses. The Coop Network provides both weather and climate data to the National Weather Service (supplementing the NWS's official staffed and automated stations) and to many other entities. Of the 5000 "A" stations, about 1200 belong to the Historical Climate Network, which includes only stations that have "provided at least 80 years of high-quality data in a stable environment."[10] The HCN thus includes more than twice as many stations as the 512 in the WMO Regional Basic Synoptic Network for North and Central America. Although at present the total number of US weather stations (including private, educational, and agricultural stations and "mesonets") dwarfs the HCN, these stations cannot provide long-term climatological data, because they do not meet the HCN's quality and stability criteria. Only HCN data meet the standards for detecting long-term regional and global climate change. The point here is that, at least until recently, only some weather-station data were useable for climatological purposes, while many climatological stations did not contribute to the weather forecast system.

When examining pre-twentieth century and paleoclimatic data, climatologists also use numerous "proxy" sources, including data on non-meteorological phenomena that depend strongly on climatic conditions. These data can provide indirect information about past weather conditions. Examples include ice cores, harvest records, tree rings, and species ranges.[11] The precision of proxy data is inherently lower than that of instrument observations, and the number of locations for most proxy data is small relative to the number of climatological stations operating today. Still, before about 1850 proxy data are all we have. Climatologists also make creative use of other non-standard sources, including diaries in which people recorded such seasonal events as when spring flowers first appeared or snow first fell.

Because all knowledge of the climate is based on information about the past, climatology has always prioritized preserving data for the long term. Climatologists prize long-term, stable, homogeneous data sets, accepting transient data sources (such as moving ships or short-lived weather stations) only in regions where no other data are available. To control data quality, climatologists may compare one data set with another for the same area, perhaps taken with different instruments (e.g. radiosonde vs. satellite). Metadata—information about station or instrument history, location, etc.—are crucial to this process.

Table 11.1 summarizes the differences between weather and climate data. Because of these differences, today's Global Climate Observing System (GCOS) bears less relevance to the issue of climate change than its parallel with the Global Observing System (for weather) would suggest. Formally established in 1992, the GCOS coordinates, links, and standardizes elements of existing observing systems to create more consistent global climate data. It is not a separate observing system; instead, it represents "the climate-focused 'system-of-systems' framework, or interface, through which all the global observing systems of WMO and its UN and non-UN system partners work together to meet the totality of national and international needs for climate observations."[12] At present the GCOS network includes about 800 surface stations selected from the roughly 10,000 in the Global Observing System, plus additional stations from other data networks (such as upper-air and ocean stations). GCOS coordination efforts

Table 11.1
The principal differences between weather data and climate data.

	Weather data	Climate data
Purpose	Dynamics: forecasting of atmospheric motion and state Forward-looking, short term (days to weeks)	Statistics: climate characterization, detection of climatic change, evaluation of climate models Dynamics: general circulation Backward-looking, long term (years to centuries)
Priorities	Speed Well-distributed coverage Forecast skill	Completeness (spatial and temporal) Stability of data record
Principal data sources	Surface synoptic network (land, ocean) Upper air (radiosonde, rawinsonde) Radar Satellites (radiances, visual images, cloud winds) 4-D data assimilation (generates additional data)	Climatological stations (some climate-only, some in synoptic network) Upper air (radiosonde, rawinsonde) Satellites: useful but problematic; period of record is short (1979–present); must be correlated with radiosonde record Ships' logs (starting in 1850s) Non-instrument sources: diaries, proxy measures (harvest dates, tree rings, ice cores, etc.), especially prior to 1850 Reanalysis (generates additional data)
Quality assessment and control	Low quality data sometimes useful in areas of poor instrument coverage Transient data sources OK Automatic consistency and conservation checks during data assimilation Data analysis detects systematic instrument errors	Long-term, stable data sources prized Transient and low-quality data sources rarely used Hand inspection and correction of individual station records for systematic biases and other errors Comparing "duplicate" data sets Comparing models and observations Metadata on station history required for quality control

Table 11.1
(continued)

	Weather data	Climate data
Computation	Centralized in Global Data Processing and Forecast System, Regional and National Meteorological Centers	Partially distributed: stations calculate monthly means Global statistics calculated by localized collectors (WWR, MCDW, Global Historical Climatology Network, individual dataset projects)
Preservation	Low priority Stored data used for model diagnosis and improvement	High priority Stored long-term data are fundamental basis of climate knowledge

will doubtless improve the global climate record in the long term, but they are still too recent to have much effect on the quality of the long-term climate record.

Meanwhile, climatologists also began to assess the homogeneity of individual weather stations' records over time. Numerous factors can reduce the stability of any time series. Instruments can drift out of calibration. Moving an instrument from a sheltered to an open location, or from the south side of a hill to the north side, can raise or lower recorded values. Over long periods, trees growing around a station can reduce recorded wind speeds. Industrial energy use, automobiles, home heating, and heat-absorbing pavement raise local temperature readings in growing cities, creating the "urban heat island" effect.

In an influential article published in 1953, J. Murray Mitchell dissected the many causes of "long-period" temperature changes in station records, dividing them into two principal types. "Apparent" changes, such as changes in thermometer location or shelters, were purely artifactual, stemming from causes unrelated to the actual temperature of the atmosphere. "Real" changes represented genuine differences in atmospheric conditions. These could be either "directly" or "indirectly" climatic, for example resulting from shifts in the general circulation (direct) or variations in solar output (indirect). But not all "real" temperature changes reflected actual climatic shifts, since some were caused by essentially local conditions (such as urban heat islands, industrial smoke, and local foliage cover) that had nothing to do with the climates of the region or the globe. Figure 11.1 details Mitchell's classification of the issues.[13]

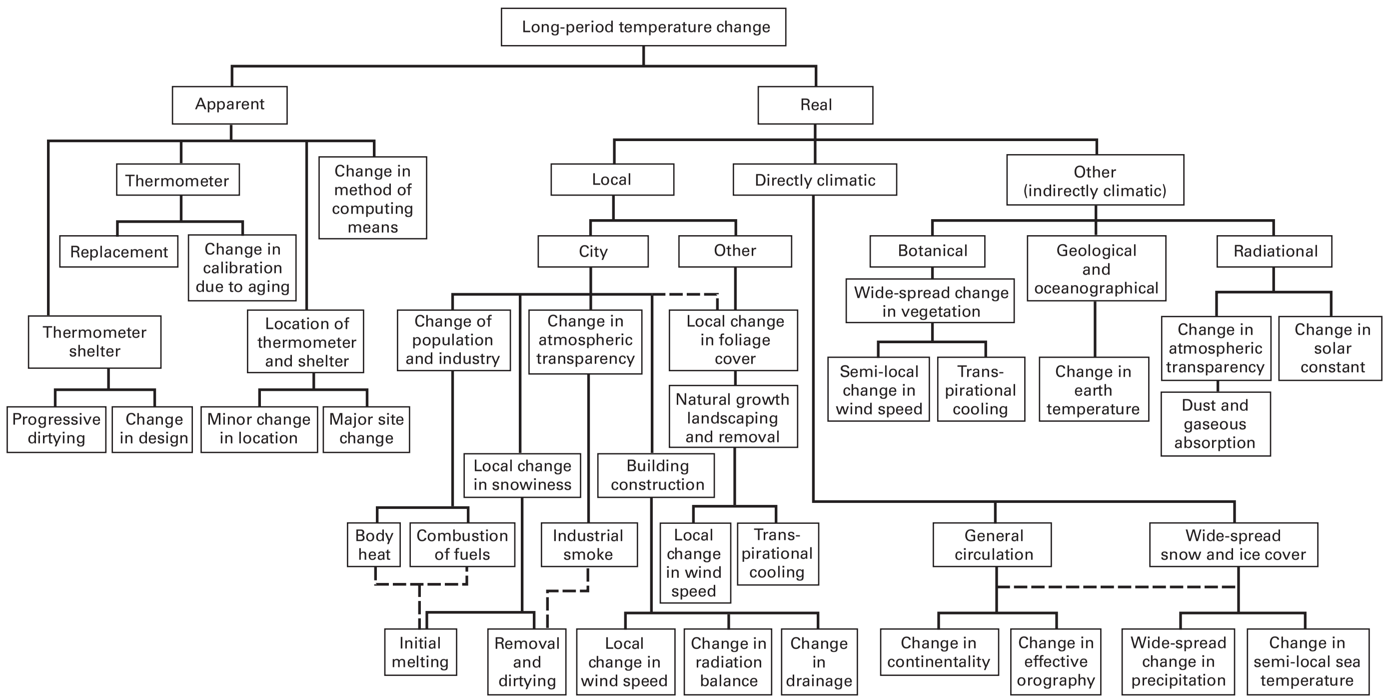


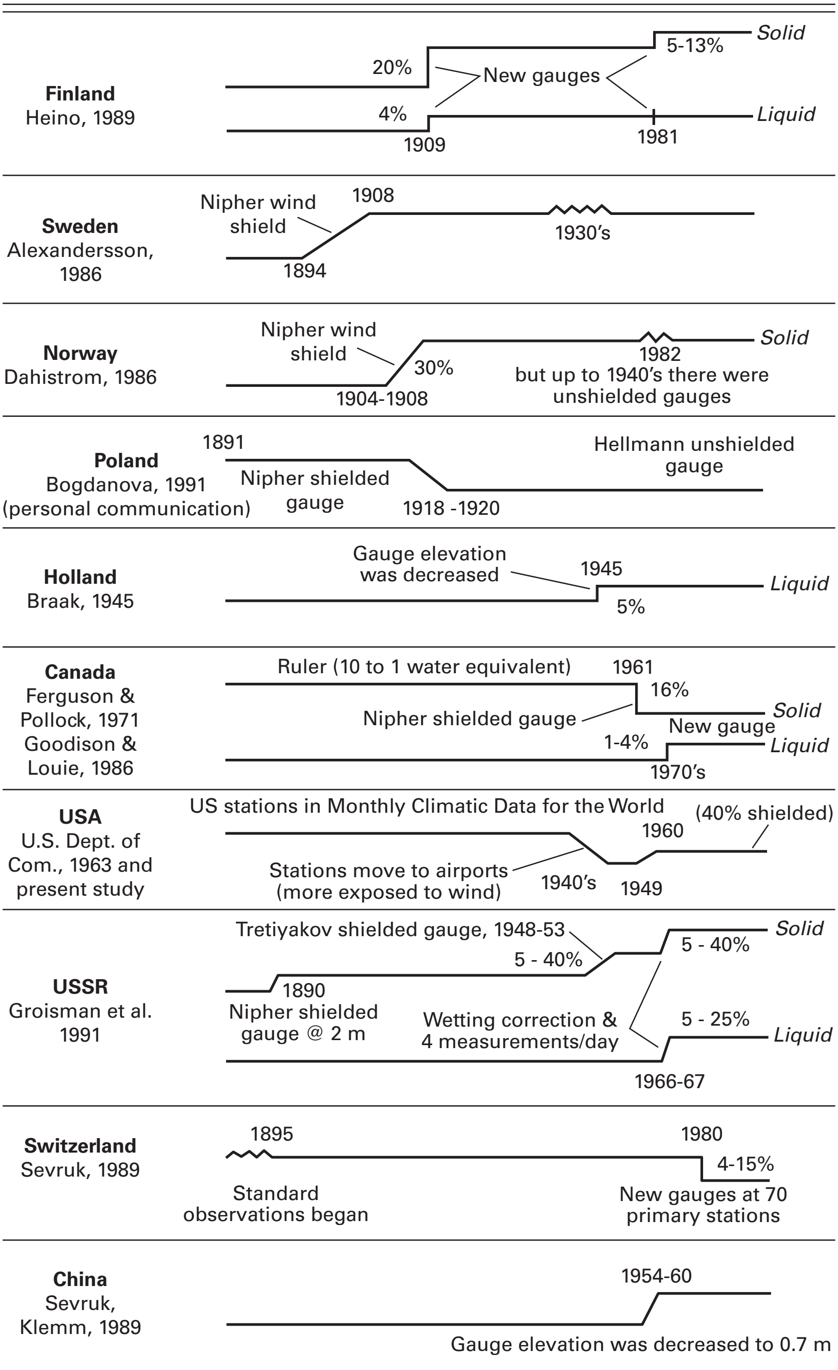
Figure 11.1
Apparent and real causes of temperature change at climatological observing stations.
Source: J. M. Mitchell, "On the Causes of Instrumentally Observed Secular Temperature Trends," *Journal of the Atmospheric Sciences* 10, no. 4 (1953): 245.

Since the majority of weather stations are located in or near population centers, urban heat islands proved especially problematic for climate data. Stations originally located miles outside a city could be engulfed within it as time went on, slowly creating an apparent climate change. Similarly, at some point virtually every station changed its observing hours, its method of calculating daily means, and/or its method of calculating monthly means. Changes in stations' locations also proved extremely common, accounting for up to 80 percent of the inhomogeneities in station records. Numerous variables related to the siting, the housing, and other local circumstances of instruments can also cause readings to differ substantially, even between sites quite close together.

Imperfect calibration of instruments to a standard is another source of error. Instruments made by different manufacturers often exhibit systematic differences. But even standardized, mass-produced instruments made by a single manufacturer can exhibit slight variations. Over time, exposure to the elements can alter individual instrument behavior. Then there is the problem of *temporal* calibration, as instrument manufacturers change their designs and sensor technologies. For example, between 1951 and the present, the Finnish Väisälä radiosondes used in Hong Kong employed at least five different temperature sensors, including two kinds of bimetal strips, two kinds of bimetal rings, and a capacitive bead. Such changes create discontinuities in the data record.[14] The occasionally dramatic discontinuities caused by adopting new instruments, changing station locations, and other issues are visible in figure 11.2, which illustrates how precipitation readings in ten countries changed over a 100-year period.

Most inhomogeneities in climate data have little political valence, but there are important exceptions. As we saw in chapter 8, during the Cold War the Soviet Union withheld some data, while the People's Republic of China withheld virtually all data. These data were not included in any Western climate data set until the mid 1980s.[15] Another kind of political issue is subtler:

> . . . consider the station Pula, . . . now managed by the Croatian Hydrometeorological Service. Its turbulent history started with the K.K. Central-Anstalt für Meteorologie und Erdmagnetismus [Austria]. From 1918 until 1930 it was managed by the Ufficio Centrale in Rome; from 1931 until 1941 it belonged to the Federal Republic of Yugoslavia. During World War II it was occupied by the Germans and after 1945 it belonged to the Socialist Republic of Yugoslavia. Since 1991 it has been part of the network of Croatia.[16]


Finland
Heino, 1989
20%
New gauges
5-13%
Solid
4%
Liquid
1909
1981
Sweden
Alexandersson, 1986
Nipher wind shield
1908
1894
1930's
Norway
Dahistrom, 1986
Nipher wind shield
30%
1904-1908
Solid
1982
but up to 1940's there were unshielded gauges
Poland
Bogdanova, 1991 (personal communication)
1891
Nipher shielded gauge
1918 -1920
Hellmann unshielded gauge
Holland
Braak, 1945
Gauge elevation was decreased
1945
5%
Liquid
Canada
Ferguson & Pollock, 1971 Goodison & Louie, 1986
Ruler (10 to 1 water equivalent)
1961
16%
Solid
Nipher shielded gauge
New gauge
1-4%
Liquid
1970's
USA
U.S. Dept. of Com., 1963 and present study
US stations in Monthly Climatic Data for the World
(40% shielded)
1960
Stations move to airports (more exposed to wind)
1940's
1949
USSR
Groisman et al. 1991
Tretiyakov shielded gauge, 1948-53
5 - 40%
5 - 40%
Solid
1890
Nipher shielded gauge @ 2 m
Wetting correction & 4 measurements/day
5 - 25%
Liquid
1966-67
Switzerland
Sevruk, 1989
1895
1980
4-15%
Standard observations began
New gauges at 70 primary stations
China
Sevruk, Klemm, 1989
1954-60
Gauge elevation was decreased to 0.7 m

These political changes may have created inhomogeneities due to changes in station management or applicable national standards—inhomogeneities that may be hard to detect if other nearby stations were simultaneously affected by the same changes.

The issues discussed above represent only a small sample of the many inhomogeneities affecting climate data. Many of these, such as recording errors and instrument placement, can be ignored because they are structurally random and thus as likely to be negative as to be positive. Therefore, given a sufficiently large number of stations, random errors of negative sign would approximately cancel errors of positive sign. Indeed, studies confirm that inhomogeneities in the *temperature* record can significantly affect trends on local and regional scales, but on hemispheric scales their effects are minimal. However, it cannot be assumed that all climate variables exhibit this self-cancellation of random errors—in precipitation, for example, the introduction of shielded gauges generally produced higher readings worldwide.[17]

Some recent work suggests that systematic errors may be more widespread than was previously believed. For example, nineteenth-century meteorologists throughout the Alps placed precipitation gauges on rooftops and thermometers in windows; later, they moved precipitation gauges to ground level and mounted thermometers inside screening devices placed in open areas (thus reducing the artifactual effects of buildings and pavement). Although stations modified their instrument placement at different times, precipitation measurements were systematically higher and temperature measurements lower after instrument placements were changed.[18] Similarly, a volunteer survey by surfacestations.org indicates a possible bias toward positive temperature errors at the majority of US climatological stations, due mainly to instrument placement near local heat sources such as air conditioner exhaust and parking lots. (However, the quality of the survey method and the possibly dubious motives of many volunteers leave this result open to question.) Still, on continental to global scales the effects of such changes on temperature trends are likely to cancel out.

A different and much more problematic issue arises with respect to satellite data. As we saw in chapter 10, most raw sensor readings from satellites require some kind of processing to convert them into meteorological

◂ **Figure 11.2**

Discontinuities in precipitation readings caused by changes in instrumentation, observing practices, and other factors, 1890s–1980s.

Source: T. R. Karl et al., "Detecting Climate Variations and Change: New Challenges for Observing and Data Management Systems," *Journal of Climate* 6 (1993), 1483.

information. This can be a complex modeling process, as in the inversion of microwave radiances, but it also can be a much simpler data-reduction process. For example, starting in 1966 the National Oceanic and Atmospheric Administration produced gridded data on snow cover from visual satellite imagery, interpreted by hand. In 1972, a higher-resolution satellite view improved the accuracy of this measurement (still interpreted by hand), resulting in an instrument-related increase in the extent of snow cover. Snow-cover readings were weekly; monthly data counted a grid cell as snow covered if covered with snow for two weeks or longer. The data-reduction process was altered in 1981; now data workers averaged the snow/no snow information from weekly charts, an approach that reduced the calculated monthly snow cover in every month except August. Figure 11.3 shows the effect of applying the newer algorithm (solid line) versus the older algorithm (dashed line) to all data from 1972 on.

Further, the National Aeronautics and Space Administration began producing snow-cover charts from microwave radiances from the Scanning Multichannel Microwave Radiometer (SMMR) in 1978; in 1987 it introduced the Special Sensor Microwave Imager, which included an additional microwave channel. These data required processing by multiple algo-

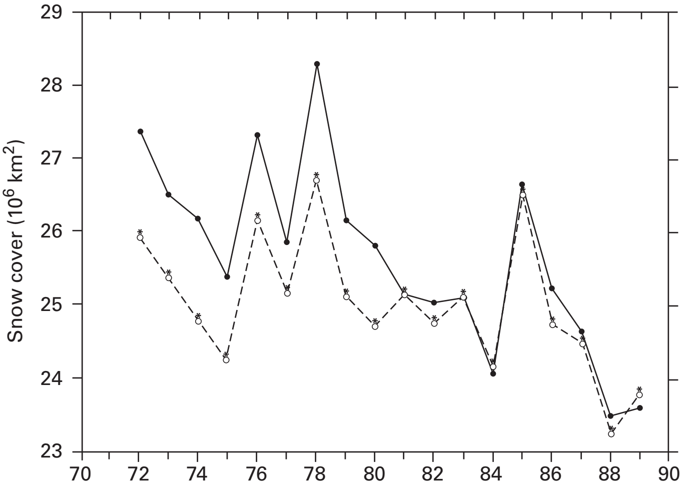


Figure 11.3

Northern hemisphere snow cover from NOAA polar orbiting satellites, processed using a consistent data-reduction algorithm (solid line), vs. the same data as processed by earlier algorithms (dashed line).

Source: Karl et al., "Detecting Climate Variations and Change," 1487.

rithms, each specific to how a particular region's land surface qualities affect the properties of snow.[19] NASA's microwave data did not agree perfectly with NOAA's visual charts; both were recognized to have virtues and defects. Thus both instrument changes and algorithm changes affected this record dramatically. On satellites, at most only two or three instruments of any given type are usually orbiting at any one time. Therefore, the self-cancellation of random instrument errors that stabilizes the global surface record cannot occur with satellites. Further, the dependence of satellite data on algorithms and modeling makes such data especially vulnerable to an increasingly common problem. Data specialists warn that

> as more and more data become dependent on processing algorithms, problems such as this are likely to grow rapidly unless special care is taken to avoid or at least document changes in processing algorithms. For example, traditional direct measurements of precipitation derived from stick measurements or weighing are being replaced by tipping devices with built-in conversion processing software. Similarly, liquid-in-glass thermometers are being replaced by electronic systems. . . . Automation will require indirect sensing of all our climate variables. Special procedures are required to archive these measurements in their proper and most basic units so that changes in either external software or internal built-in microprocessors will [enable] homogeneous reprocessing of the data when the inevitable improvements in the system occur.[20]

Fighting for the Long Term: Building Stability into Change

The assortment of issues described above only scratches the surface of the vast array of data problems scientists face in their ongoing quest to refine and extend the historical climate record. Most, if not all, of those problems stem from the divergence of purpose and focus between the operational systems responsible for weather forecasting and the related but not identical systems for monitoring climate. Historically, climate scientists have often found themselves stuck with damaged goods—leftovers from a much larger forecasting enterprise that ignored consistency and reliability in its constant drive to move forward with newer, better technology. Until climate change became a major public issue, and even afterward, this was often a losing battle.

Concerns about the stability and reliability of data led to calls—repeated over decades in various forms—to improve the climate data infrastructure.[21] The US National Research Council, the Global Climate Observing System, the UN Framework Convention on Climate Change, the Intergovernmental Panel on Climate Change, and other agencies have endorsed principles for climate monitoring similar or identical to the following:

1) *Management of network change.* Assess how and the extent to which a proposed change could influence the existing and future climatology.
2) *Parallel testing.* Operate the old system simultaneously with the replacement system.
3) *Metadata.* Fully document each observing system and its operating procedures.
4) *Data quality and continuity.* Assess data quality and homogeneity as a part of routine operation procedures.
5) *Integrated environmental assessment.* Anticipate the use of data in the development of environmental assessments.
6) *Historical significance.* Maintain operation of observing systems that have provided homogeneous datasets over a period of many decades to a century or more.
7) *Complementary data.* Give the highest priority in the design and implementation of new sites or instrumentation within an observing system to data-poor regions, poorly observed variables, regions sensitive to change, and key measurements with inadequate temporal resolution.
8) *Climate requirements.* Give network designers, operators, and instrument engineers climate monitoring requirements at the outset of network design.
9) *Continuity of purpose.* Maintain a stable, long-term commitment to these observations, and develop a clear transition plan from serving research needs to serving operational purposes.
10) *Data and metadata access.* Develop data management systems that facilitate access, use, and interpretation of data and data products by users.[22]

The ideal is clear: to integrate the weather and climate data networks, forming a genuine, robust, and enduring climate-data infrastructure. Governments and private-sector elements have begun to focus on the need for reliable predictions of seasonal and interannual climate, as well as of long-term climate change. As a result, calls have emerged for an *operational* climate forecasting system that would provide real-time climate analysis and prediction capabilities on a local or a regional scale. The US NOAA/NCEP Climate Test Bed, for example, envisions "'a Seamless Suite of Forecasts' spanning weather, intraseasonal, interannual, and multi-decadal timescales."[23]

Looking backward, climate scientists face the daunting task of refining and reconstructing the historical record. The infrastructural inversion process I have described throughout this book is one major tool for rectifying the climate data record: looking at each station's record, recovering whatever can be learned about the station's history, correcting for some kinds of changes, rejecting anomalous data points as likely errors, and so on. After taking the infrastructure apart, scientists can—sometimes, to some degree—correct the record for errors and systematic biases. From this perspective, the climate data infrastructure looks much more fluid and less stable than the dry and certain-seeming lists of numbers it provides might otherwise suggest.

Global Climate Data—Plural

We live on just one planet, with just one real climatic history. Yet when climate scientists talk about global climate data, they are talking in the plural. Just as with weather, we are multiplying data images for global climate. Hundreds, even many thousands of variant data images exist, though only a few gain authoritative status, and they have generally (but not always) converged over time.

As an example, let us explore some data sets behind the most important figure derived from global climate data: global average temperature. Usually climate scientists express this figure as a temperature anomaly time series. The "temperature anomaly" is simply the difference (positive or negative) between a given year's temperature and the average temperature of a chosen reference period; in figure 11.4, the reference period is 1961–1990. This technique allows direct comparison of different data sets without regard to their absolute temperature values, which may differ. It also permits comparing records from different types of instruments. For example, temperature anomaly trends for the lower troposphere from radiosonde data correlate well with those from surface thermometer data, but the absolute values of radiosonde and surface readings differ substantially. (Temperature varies as a function of altitude.)

Figure 11.4, from the IPCC Fourth Assessment Report (2007), compares global temperature trends from nine well-known datasets, all expressed as anomalies from a 1961–1990 average. Dozens of other global average temperature datasets (not shown here) have also been created from surface thermometer data.[24] On the decadal scale, all the trend lines show similar tendencies after 1900. Yet clearly they also disagree, sometimes strikingly. The figure reveals a maximum difference between the various trends of about 0.6°C before 1900 and 0.2°C after 1900. This may not sound like a large difference. Yet across the entire period 1840–2005, the total difference between the minimum temperature (Willett's –0.9°C) and the maximum one (Brohan et al.'s +0.45°C) is only 1.35°C. Thus the maximum disagreement among the trends is nearly half as large as the maximum total temperature change one might read from this chart. Further examination reveals other oddities. For example, from 1850 to 1900 the Willett trend line offers values well below all the others; also, its slope rises from 1875

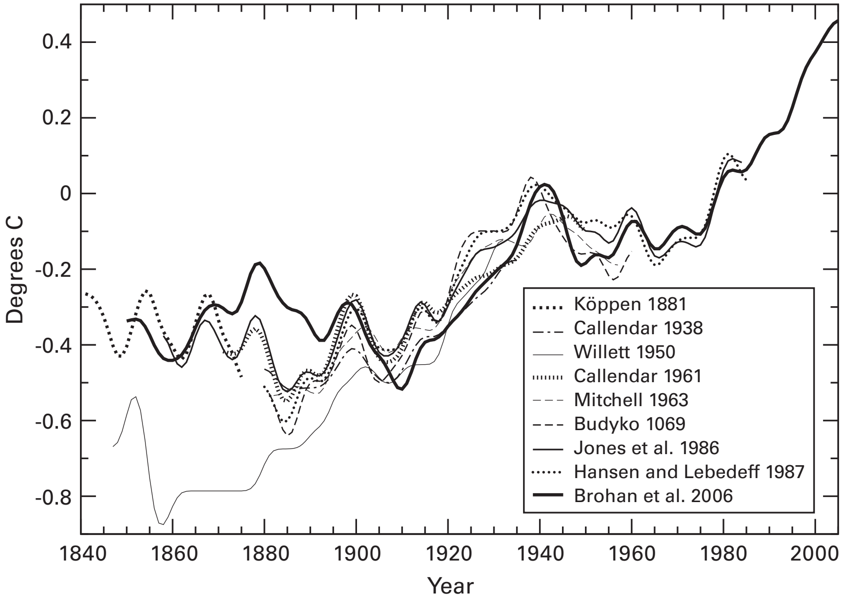


Figure 11.4

Original legend: "Published records of surface temperature change over large regions. Köppen (1881) tropics and temperate latitudes using land air temperature. Callendar (1938) global using land stations. Willett (1950) global using land stations. Callendar (1961) 60°N to 60°S using land stations. Mitchell (1963) global using land stations. Budyko (1969) Northern Hemisphere using land stations and ship reports. Jones et al. (1986a,b) global using land stations. Hansen and Lebedeff (1987) global using land stations. Brohan et al. (2006) global using land air temperature and sea surface temperature data is the longest of the currently updated global temperature time series (Section 3.2). All time series were smoothed using a 13-point filter. The Brohan et al. (2006) time series are anomalies from the 1961 to 1990 mean (°C). Each of the other time series was originally presented as anomalies from the mean temperature of a specific and differing base period. To make them comparable, the other time series have been adjusted to have the mean of their last 30 years identical to that same period in the Brohan et al. (2006) anomaly time series."

Source: *Climate Change 2007: The Physical Science Basis* (Cambridge University Press, 2007). Image courtesy of IPCC.

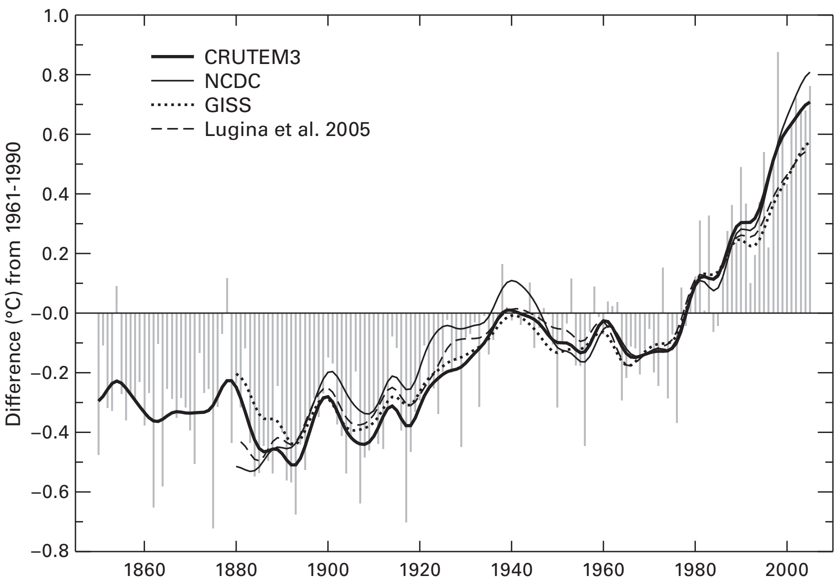


Figure 11.5

Four recent global surface temperature anomaly data sets. The thick trend line here represents the same data as the thick trend line in figure 11.4. Original legend: "Annual anomalies of global land-surface air temperature (°C), 1850 to 2005, relative to the 1961 to 1990 mean for CRUTEM3 updated from Brohan et al. (2006). The smooth curves show decadal variations. The [thick] curve from CRUTEM3 is compared with those from the National Climatic Data Center [thin], the Goddard Institute for Space Studies [dotted], and Lugina et al. [dashed]."

Source: *Climate Change 2007: The Physical Science Basis* (Cambridge University Press, 2007). Image courtesy of IPCC. Colors replaced as noted in brackets to allow non-color reproduction.

to 1895, where several other lines trend downward. Meanwhile, from 1870 to 1895 the Brohan et al. trend shows a considerably higher value than any of the others.

Recently produced global data sets converge more closely, but they still show significant differences. Figure 11.5 compares four data sets produced between 2001 and 2006. The trend lines here exhibit a maximum disagreement of around 0.3°C in 1880–1890 and 0.2°C beyond that. Yet the 1880–1890 trends disagree in both slope and sign. What is going on here? Why do we have multiple climate data sets in the first place? Why do they disagree? How can the "same" data produce different results? The classical scientific approach to these questions tries to narrow the field. Some data sets are better than others; through a process of examination and correction, you hope to approach, asymptotically, a single master data set, a definitive quantitative account of what actually happened in the atmosphere across the period of record. This approach sees data as history; it represents a fundamental premise of scientific understanding. Obviously we would like to know with greater certainty exactly how much the planet has warmed, and numerous scientists have spent their entire working lives trying to do just that. The increased convergence of the trend lines in figure 11.5 (versus those in figure 11.1) represents the fruit of this effort.

Yet just as in the case of ensemble weather forecasting, rather than reducing to a single definitive set, global climate data images have instead proliferated dramatically in the computer age. All the themes I have developed in this book played parts in this proliferation. The path of change closely mirrored that of weather forecasting, but with a delay of 10–20 years. As they proceeded, climate science shifted from the partially distributed computation characteristic of its past toward massively centralized computation in Latourian "centers of calculation."

First, in the 1970s, digitization projects made global climate data widely available for the first time by reducing data friction. Digitization made it possible to transfer formerly inconceivable volumes of data by exchanging data tapes. More recently, data sets mounted on Internet file servers reduced friction even further. Next, computers reduced the computational friction involved in applying complex mathematical techniques. Data centers, using an increasingly sophisticated array of interpolation methods, began producing *gridded* historical data sets. These centers could now apply consistent standards, and techniques could be retrospectively applied in order to *make data global* in the sense of chapter 10. As a result, scientists moved beyond simple monthly and annual averages toward a profusion of statistical methods and modeling techniques. This culminated in the 1990s with reanalysis of 40-year data sets through frozen 4-D data assimilation systems (discussed in chapter 12). Now scientists are working toward reanalyzing the entire 150-year history of instrumental observations.

To unearth the changing "memory practices" of climate science, let us first ask: What exactly *is* a "global climate data set"[25]? As we have seen, "global climate data" are never simply the total collection of relevant instrument readings. Until reanalysis, climatology always required long-term, homogeneous records (see previous section). Box 11.1 shows that, with the exception of Callendar's 1961 effort, the major global climate data sets assembled before 1965 used fewer than 200 station records. Like virtually all hemispheric and global data sets published before 1960, these relied

Box 11.1
About the Historical Surface Temperature Series in Figure 11.4

Köppen 1881 Used fewer than 100 land stations, covering only the tropics and temperate zones.[a]

Callendar 1938 Discussed in chapter 4; selected about 200 stations worldwide.

Willett 1950 The most comprehensive global calculation to date, reaching back to 1845. Seeking a geographically representative sample, Willett chose one station to represent each 10° latitude-longitude cell, picking the one with the best available long term record. By this procedure he selected 129 station records of more than 50 years, plus another 54 stations with 20–50 year records. Thus Willett managed to cover 183 of a possible 648 10° × 10° cells. Data for all but two stations were taken from *World Weather Records*. Willett's global average weighted each station record equally.[b]

Callendar 1961 Examined some 600 station records, "the majority" from World Weather Records and "a few score others from a variety of sources too numerous to mention."[c] Of these, Callendar estimated that about 74 percent were probably reliable; 8 percent showed spurious temperature increases probably related to urban locations (heat islands); and 18 percent were probably unreliable. Therefore he retained only about 450 stations for his study, notable also for promoting (again) the carbon dioxide theory of climatic change.[d]

Mitchell 1963 A student of Willett, Mitchell updated Willett's 1950 study using a nearly identical method and selection of stations. Unlike his mentor, however, Mitchell weighted stations according to the surface area of the latitude band in which they were located. (Since the surface area of latitude bands is greater near the equator and smaller near the poles, Willett's simple averaging technique gave too much weight to high-latitude stations.) For this reason, Mitchell's series[e] begins in 1882, the first year in which all latitude bands could be adequately represented. The weighting procedure accounts for the considerable difference between Mitchell's calculation and Willett's.

Budyko 1969 This trend line represents only the northern hemisphere above about 20°N. Budyko's data came from temperature-anomaly analysis maps created at the Main Geophysical Observatory in Leningrad. Although many station records used to create the maps were drawn from *WWR*, they also included Soviet records not previously available to Western climatologists.[f]

Jones et al. 1986 Produced by the Climatic Research Unit at the University of East Anglia, this global data set used 2194 stations, including hundreds of new records recovered from meteorological archives covering previously undocumented areas of the Soviet Union, the People's Republic of China,

Box 11.1
(continued)

northern Africa, and northern Europe. Jones et al. also produced a gridded version of the dataset (5° latitude by 10° longitude). Rather than use Willett's and Mitchell's method of choosing a single station to represent a grid cell, Jones et al. interpolated values for each gridpoint from a number of stations surrounding each gridpoint. They further introduced numerous corrections to existing data. This data set was later expanded to include marine temperatures. Continuously updated and corrected, the CRU gridded data set remains one of the most authoritative.[g]

Hansen and Lebedeff 1987 Based principally on W. M. L. Spangler and R. L. Jenne, *World Monthly Surface Station Climatology* (National Center for Atmospheric Research, 1984), a digitized, corrected, and updated version of *World Weather Records* and *Monthly Climatic Data for the World*. Like most other investigators, Hansen and Lebedeff introduced corrections of their own. For example, they tested the effect of urban heat islands on the global data by extracting all stations associated with population centers larger than 100,000 people[h]; this reduced the global average change from 0.7°C to 0.6°C across the period 1880–1980.

Brohan et al. 2006 The massively revised third release (CRUTEM3) of the Climatic Research Unit 1986 dataset. It includes quality-controlled monthly average temperatures for 4349 land stations. Between 1986 and 2006, numerous previously unreported stations were added, while station records were corrected or homogenized by national meteorological services and others.[i] The trend lines represent only the land-surface component of this dataset. The full dataset, HadCRUT3, combines marine and land surface data to provide complete global surface coverage (Brohan et al., "Uncertainty Estimates in Regional and Global Observed Temperature Changes").

a. W. Köppen, "Über Mehrjährige Perioden der Witterung—III. Mehrjährige Änderungen der Temperatur 1841 bis 1875 in den Tropen der Nördlichen und Südlichen Gemässigten Zone, an den Jahresmitteln Untersucht," *Zeitschrift der Österreichischen Gesellschaft für Meteorologie* (1881): 141–50.

b. H. C. Willett, "Temperature Trends of the Past Century," *Centenary Proceedings of the Royal Meteorological Society*, 1950: 195–206.

c. Callendar, "Temperature Fluctuations and Trends Over the Earth."

d. Ibid.

e. J. M. Mitchell Jr., "On the World-Wide Pattern of Secular Temperature Change," in *Changes of Climate* (1963).

f. A. Robock, "The Russian Surface Temperature Data Set," *Journal of Applied Meteorology* 21, no. 12, 1982.

Box 11.1
(continued)

g. P. D. Jones et al., "Global and Hemispheric Temperature Anomalies—Land and Marine Instrumental Records," in *Trends Online*, Carbon Dioxide Information Analysis Center, US Department of Energy, Oak Ridge National Laboratory, 2006; Jones et al., "Southern Hemisphere Surface Air Temperature Variations: 1851–1984," *Journal of Applied Meteorology* 25, no. 9 (1986): 1213–30.
h. J. Hansen and S. Lebedeff, "Global Trends of Measured Surface Air Temperature," *Journal of Geophysical Research* 92, no. D11, 1987: 13,345–72.
i. P. Brohan et al., "Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Dataset from 1850," *Journal of Geophysical Research* 111, no. D12106, 2006; P. D. Jones and A. Moberg, "Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001," *Journal of Climate* 16, no. 2, 2003: 206–23; P. D. Jones, "Hemispheric Surface Air Temperature Variations: A Reanalysis and an Update to 1993," *Journal of Climate* 7, no. 11, 1994: 1794–802.

almost exclusively on *World Weather Records*. *WWR*, first published in 1927 by the Smithsonian Institution, became the *de facto* standard data source for large-scale climatology.[26] Yet the early *WWR* collection effort was *ad hoc*, based not on a systematic and exhaustive survey but on the personal contacts of the collectors. Therefore, the *WWR* collection omitted vast swathes of data. For example, it captured only about 5 percent of all available mid-nineteenth-century data. Over half of its nineteenth-century records originated in just three countries: the United States, Russia, and India.[27]

Until quite recently, in fact, climatologists studying regional or global climate rarely used raw instrument readings at all. Instead, as we saw in chapter 5, from its earliest days climatology relied on a strategy of partially distributed computation. Each station calculated its own statistics, such as monthly means, and transmitted *only* those results—rather than the instrument readings from which they were calculated—to central collecting entities. Initially, this was the mission of the Réseau Mondial. Later, under sponsorship of the International Meteorological Organization and the World Meteorological Organization, *World Weather Records* assumed responsibility for collecting world climate data. In 1959 the US Weather Bureau took over *WWR*, eventually handing it off to the US National Oceanic and Atmospheric Administration, which also published *Monthly Climatic Data for the World*. Figure 11.6 shows typical data tables from these two publications.

ZANZIBAR, EAST AFRICA

Lat. 6° 10′ S. Long. 39° 11′ E. $H_b = 56$ ft.

TEMPERATURE IN DEGREES F.

Means of ½ (daily Max. + daily Min.)

Date	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
1891	...	...	...	...	*78.3	77.3	76.3	†75.9	77.7	78.9	...	81.1	...
1892	81.6	83.4	82.7	81.2	79.3	78.5	76.8	77.3	78.1	78.9	80.5	82.2	**80.0**
1893	81.1	82.0	81.6	79.8	*77.5	‡77.0	76.0	75.4	76.6	78.7	79.0	81.8	**78.9**
1894	82.9	83.8	82.6	81.1	78.3	76.9	76.0	77.0	78.2	79.5	79.4	81.9	**79.8**
1895	82.9	83.8	83.2	81.0	79.3	77.6	76.5	77.3	77.6	79.5	80.2	82.9	**80.2**

SURFACE

JANUARY 1998

STATION	LATITUDE	LONGITUDE	ELEVATION	DAYS OBS.	PRESSURE		TEMPERATURE		VAPOR PRESSURE		PRECIPITATION					SUNSHINE	
					MEAN STATION	MEAN SEA LEVEL	MEAN	DEPART	MEAN	DEPART	DAYS 1 MM	TOTAL	DEPART	QUINTILE	TOTAL	% OF AV.
	°′	°′	m		mb	mb	°C	°C	mb	mb		mm	mm		hr	
AUSTRIA																
11028 ST. POELTEN	4812N	01537E	282		985.0	1020.3	.7		5.5	.8	6	37	3	4	58	109
11035 WIEN/HOHE WARTE	4815N	01622E	209		994.1	1020.1	2.2		5.5	.7	5	30	-8	3	74	132
11120 INNSBRUCK-FLUGHAFEN	4716N	01121E	593		949.7	1021.9	.4		4.9		9	45		2	127	
11146 SONNBLICK	4703N	01257E	3109		689.4	2992 Z	-10.8		1.4	-.5	12	83	-175	1	148	130
11150 SALZBURG-FLUGHAFEN	4748N	01300E	450		965.9	1021.1	1.0		5.3	.6	7	31	-32	1	96	132
11155 FEUERKOGEL	4749N	01344E	1621		833.2	1461 Y	-1.8		3.1		17	94		1	135	
11212 VILLACHERALPE	4636N	01340E	2160		781.2	1490 Y	-5.1		2.6		5	24		1	152	
11231 KLAGENFURT-FLUGHAFEN	4639N	01420E	476		963.5	1022.2	-.9		4.7	.6	3	6	-30	1	128	178
11240 GRAZ-THALERHOF-FLUGHAFEN	4700N	01526E	347		977.6	1020.5	.4		5.3	.8	0	3	-28	1	134	213

Figure 11.6

Sample sections from *World Weather Records* (*WWR*, top) and *Monthly Climatic Data for the World* (bottom). The *WWR* entry includes the station's method of calculating mean monthly temperature. (See table 11.2.)

Sources: H. H. Clayton, *World Weather Records* (Smithsonian Institution, 1927); *Monthly Climatic Data for the World: January 1998* (National Oceanic and Atmospheric Administration, 1998).

Starting in 1935, to standardize and facilitate collection of climatic data, the International Meteorological Organization asked all national weather services to transmit "mean monthly values of the main climatological elements" early in the following month. To integrate this reporting into the condensed weather codes sent over the international telegraph network, the IMO (and later the WMO) developed a separate set of CLIMAT codes. These reporting requirements were later extended to ocean stations and upper-air networks. WMO technical regulations initially set the density of the requested climatological network at one station every 300 km, but this proved impractical; it was later relaxed to one station every 500 km.[28]

The distributed computation strategy reduced computational friction, since climatologists could calculate regional or global statistics from the reported monthly averages—a much smaller set of numbers than the individual daily readings. It also reduced data friction, decreasing the volume of data that had to be moved from one place to another. But the strategy also created latent uncertainty about data quality. While assembling and summarizing the incoming data, editors of *World Weather Records* and *Monthly Climatic Data for the World* could (and often did) detect gross errors, such as reversals of sign, as well as missing data. As a quality-control measure, they checked each station's reports against those of neighboring stations. However, the large distances (typically 500–1000 km) between stations made it impossible to detect small or subtle errors with this technique.

As computers and communication systems matured, climatologists began to revisit the entire strategy. In the 1960s, a handful of climate research centers began a painful process of reviewing and recalculating all available global climate data. At first they did this mainly by examining and cross-checking the statistical reports. The next steps were to seek out and include station data missed by *WWR* and other collectors, and to develop new techniques for including data records previously deemed too brief.

The Soviet data sets created in 1969 by Mikhail Budyko and his colleagues represent a special case. In addition to data from *WWR*, Budyko's group included data from Soviet and other sources that had never been incorporated into Western data sets. Whereas most others calculated averages mathematically, the Soviet climatologists deployed a graphical mapping technique. In a project lasting eight years (1960–67), the Main Geophysical Observatory in Leningrad (now St. Petersburg) prepared monthly analysis maps of hemispheric, and later global, surface temperature anomalies starting from 1881.[29] Beginning with 246 stations in 1881, by 1980 the Soviet data set included about 2000 stations. Budyko's group first corrected and "homogenized" the station records. They then plotted these records on maps, creating "hand-drawn, smooth, synoptic-type analyses" much like the subjective analyses of pre-NWP weather forecasting.[30] Finally, they overlaid a 5°×10° grid on the maps and interpolated from the map contours to the gridpoints, releasing the gridded data set on digital tape in 1980.

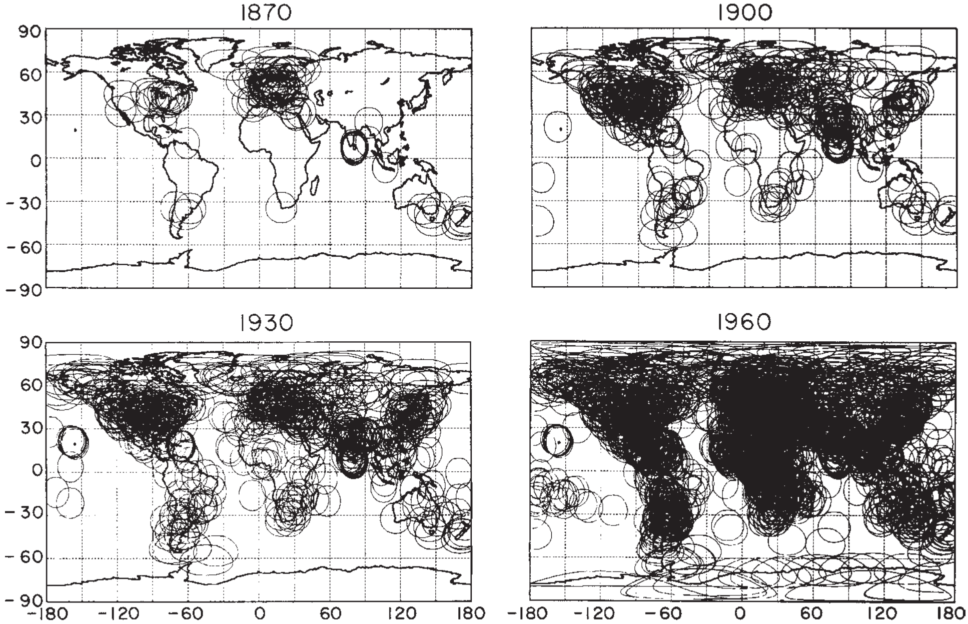
Until the mid 1980s, many climate scientists regarded this Soviet data set as the best available global source, despite a lack of clarity about exactly what corrections had been applied and how. Alan Robock and other climatologists working with the Soviet authors under a joint US-USSR environmental protection agreement concluded that the Soviet data-homogenization techniques, though mostly undocumented, were probably of high quality. Yet the group's subjective analysis techniques rendered direct comparison with other data sets problematic.[31]

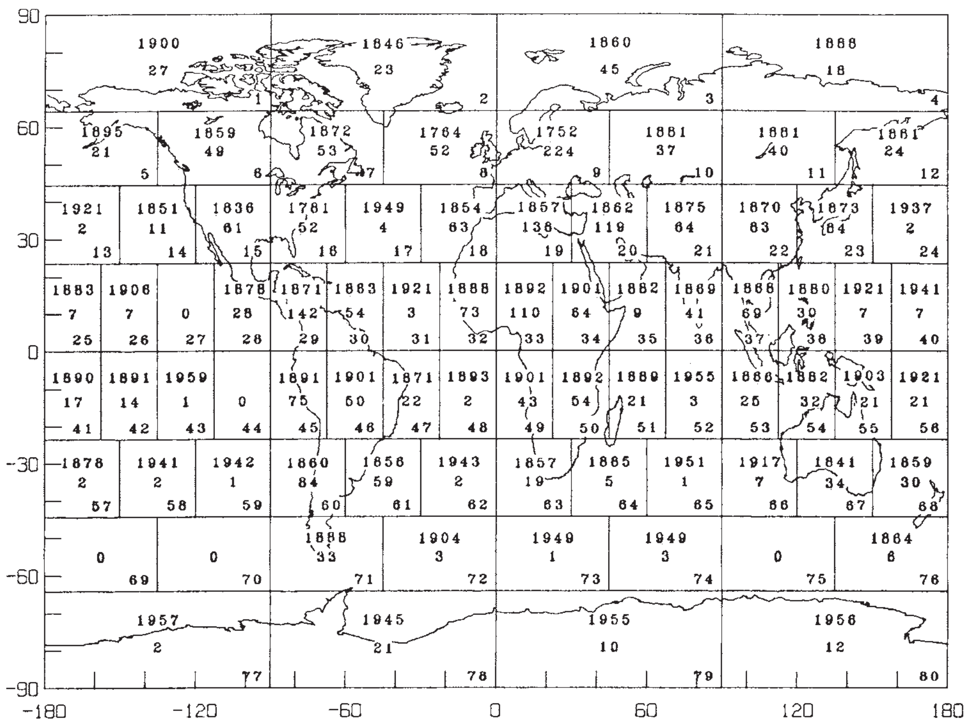
Reconstructing the Climate Record

In the mid 1970s, as climate models matured, climate data specialists began to digitize historical climate data and grid it for computer processing. These efforts naturally started with the major existing sets of climate data. The US National Climatic Data Center and National Center for Atmospheric Research digitized and combined *World Weather Records* and *Monthly Climatic Data for the World* around 1975, producing a database known as *World Monthly Surface Station Climatology* (*WMSSC*). NCAR and NCDC made this dataset available to other researchers on magnetic tapes.[32] This became the basis for numerous subsequent datasets. The middle numbers in each cell of figure 11.7 show the number of stations represented in one such dataset produced by NASA's Goddard Institute for Space Studies.[33] The surface temperature trend derived from this dataset appears as the Hansen-Lebedeff line in figure 11.4.

In the early 1980s, in an international effort based at the Climatic Research Unit of the University of East Anglia (with participation from the US Department of Energy's newly formed Carbon Dioxide Research Division), Phil Jones, Raymond Bradley, Tom Wigley, and others embarked on a systematic attempt to recover every available station record from 1850 on. They surveyed numerous meteorological data centers, libraries, and archives to find records not included in *WWR*. The team more than doubled the number of nineteenth-century northern-hemisphere station records. They also increased geographical coverage to include much of northern Asia and a number of Atlantic islands. The group then carefully interrogated the *WWR* data. Among other things, they found that the *WWR* collectors, in their zeal to create a definitive record, had introduced

Figure 11.7
Top: Evolution of coverage by surface stations in *World Monthly Surface Station Climatology* (based principally on *World Weather Records* and *Monthly Climatic Data for the World*), with coverage shown as a 1200 km radius around each station. Bottom: Surface stations included in the Goddard Institute for Space Studies version of *World Monthly Surface Station Climatology* as of 1987. Grid cells demarcate regions of equal area. Numbers in each cell represent the date on which coverage began (top), total number of stations in that region (middle), and a grid cell identifier (bottom right). Note that 12 of the 40 cells in the southern hemisphere contain 0–2 stations.
Source: J. Hansen and S. Lebedeff, "Global Trends of Measured Surface Air Temperature," *Journal of Geophysical Research* 92, no. D11 (1987), 13,346–47.


1870
1900
1930
1960
90
60
30
0
-30
-60
-90
-180 -120 -60 0 60 120 180
-180 -120 -60 0 60 120 180


90
60
30
0
-30
-60
-90
-180 -120 -60 0 60 120 180
1900 27 1
1846 23 2
1860 45 3
1888 18 4
1895 21 5
1859 49 6
1872 53 7
1764 52 8
1752 224 9
1881 37 10
1881 40 11
1881 24 12
1921 2 13
1851 11 14
1836 61 15
1781 52 16
1949 4 17
1854 63 18
1857 138 19
1862 119 20
1875 64 21
1870 83 22
1873 84 23
1937 2 24
1883 7 25
1906 7 26
0 27
1878 28 28
1871 142 29
1883 54 30
1921 3 31
1888 73 32
1892 110 33
1901 64 34
1882 9 35
1869 41 36
1868 69 37
1880 30 38
1921 7 39
1941 7 40
1890 17 41
1891 14 42
1959 1 43
0 44
1891 75 45
1901 50 46
1871 22 47
1893 2 48
1901 43 49
1892 54 50
1889 21 51
1955 3 52
1886 25 53
1882 32 54
1903 21 55
1921 21 56
1878 2 57
1941 2 58
1942 1 59
1860 84 60
1856 59 61
1943 2 62
1857 19 63
1885 5 64
1951 1 65
1917 7 66
1841 34 67
1859 30 68
0 69
0 70
1888 33 71
1904 3 72
1949 1 73
1949 3 74
0 75
1864 6 76
1957 2 77
1945 21 78
1955 10 79
1956 12 80

“adjustments” that would create complex difficulties for future investigators. With respect to the US data—“a very mixed bag”—they noted that *WWR* included “records which have had no adjustments, records adjusted for observation time change, records adjusted for location, all of the above, none of the above, or some of the above for some of the time.”[34] For example, they found that standard observing times had “varied chaotically from one country to another, and even from one station to another within the same country.” Similarly, up until 1950 the *WWR* collectors had applied a stringent standard for mean temperature at US stations: that daily means should be calculated by averaging 24 one-hourly observations. Few stations had ever collected so many readings, so the *WWR* applied an adjustment factor to bring them into line with a “true” 24-hourly mean.[35] The CRU successfully “deconvoluted” the US data by working backward from the *WWR* adjustments. Yet the *WWR* collectors documented many of their other adjustments poorly or not at all, leaving future scientists to puzzle out exactly what they had done, such as how they had compensated for station site changes.[36] The CRU also found that national meteorological services had employed a vast variety of slightly different methods to calculate mean daily temperatures. Table 11.2 shows a sample of the dozens of different methods used in the nineteenth and early twentieth centuries. Only fifteen countries maintained a consistent method from the nineteenth century onward.

Changes in observing hours and methods of calculation could have a considerable effect on temperature averages. Figure 11.8 illustrates the problem. Since most of the stations in the US Cooperative Observer Network are staffed by amateurs paid (if at all) only tiny sums for their work, the network originally accommodated their sleep schedules by setting a “climatological day” ending at sunset, typically between 5 p.m. and 8 p.m. local time, when observers record the maximum and minimum for the preceding 24 hours. Later on, the climatological day was shifted to end at 7 a.m., introducing a “time of observation bias” into the climate record, first noted by Mitchell.[37] For example, “observers who report the minimum temperature ending at 0700 local standard time can have twice as many days with temperatures below freezing under certain climate regimes than if they were to observe the 24-hour minimum at 1700 local standard time.”[38] This and other changes in standard observing hours caused an artifactual reduction in the average annual temperature of the United States of 0.16°C across the period 1931–1985 (figure 11.8). Similarly, neither the 0700 nor the 1700 climatological-day regime syncs with the midnight-to-midnight calendar day. Yet readings from both are *attributed* to the calendar

Table 11.2

Different methods used to calculate mean daily temperatures in selected countries, primarily during the nineteenth century. 00, 03, etc. refer to observing hours. Reproduced from J. P. Palutikof and C. M. Goddess, "The Design and Use of Climatological Data Banks, with Emphasis on the Preparation and Homogenization of Surface Monthly Records," *Climatic Change* 9, no. 1, 1986, 139).

Country	Methods used to calculate mean daily temperature
Egypt	½(max + min); means of 3-hourly observations, ⅛(00 + 03 + · · · 21); ¼(09 + 21 + max + min); ¼(06Z + 12Z + 18Z + min); means of 24 hourly values (exact hours unknown)
France	½(max + min); ⅓(06 + 13 + 21); ⅓(06 + 14 + 22); means of 24 hours, 1/24(01Z + 02Z + · · · 24Z); means of eight 3-hourly observations
Ghana	⅛(03 + 06 + · · · 24); ½(max + min)
Guyana	½(max + min); 1/12(00 + 02 + · · · 22); ⅓(07 + 13 + 18) local time; ½(12Z + 18Z).
Tunisia	Means of 24 hours (exact hours unknown); ½(max + min); ¼(07 + 13 + 19 + (19 + min)/2)
USSR	¼(01 + 07 + 13 + 19); ⅓(07 + 13 + 21); ¼(01 + 07 + 13 + 21); ¼(07 + 14 + 21 + 21) 105°E meridian time; means of 2–4 daily observations in 53 different combinations

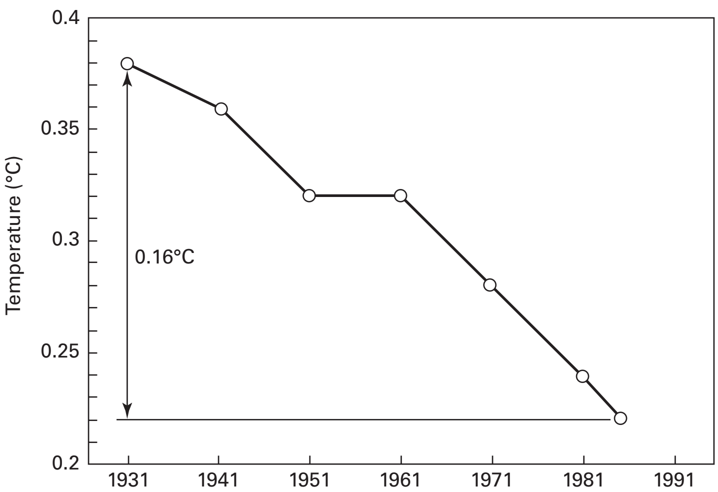


Figure 11.8

Estimated changes in US national average temperature caused by changes (at circled points) to standard observing hours.

Source: Karl et al., "Detecting Climate Variations and Change," 1486.

day on which they end. The use of climatological days can thus shift minima from the first or the last calendar day of a month into the preceding or the following month, altering the calculated monthly average.

Before digitization, most efforts to calculate global temperature had selected a few stations to represent a given area, such as a 10° latitude-longitude cell.[39] The Soviet graphical effort described above was an exception, but the lack of documentation for adjustments and interpolation techniques made it unreplicable and caused an enduring debate over its quality. After digitization, data specialists began to apply mathematical interpolation techniques similar to those used in weather forecasting to produce gridded global data sets. Here, however, the purpose of gridding was not to feed numerical models. Instead, gridding offered a principled way to integrate data from multiple stations to generate a single value for each grid cell. This, investigators hoped, would reduce the effects of minor inhomogeneities and "locally unrepresentative individual stations." Various interpolation strategies and gridding techniques could then be compared. An early CRU effort produced the gridded Northern Hemisphere data set shown in figure 11.9. The criterion used required at least six stations within 300 nautical miles of each gridpoint, each with at least ten years' data. Temperature anomalies calculated from these data correlated very closely with the Soviet data set originally produced by Budyko's group and updated by Vinnikov et al.[40]

Another project took on the 70 percent of Earth's surface area occupied by oceans. Beginning in 1981, the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, originally COADS) project prepared a digitized version of marine atmosphere and sea surface temperatures, using marine logbooks dating back to Maury's pioneering effort in 1854. Digital versions of some of these records had already been created as early as the 1890s, when the US Hydrographic Office introduced Hollerith punch cards. By the 1920s, punch-card recording had become routine in the United States and in Europe. During and after World War II, collectors in the United States and elsewhere began accumulating and combining decks of cards, recoding them into a single format and eventually transferring them to magnetic tape. By 1981, various digitized sources included more than 140 million records, about half of them duplicates. These data suffered from most of the homogeneity problems described earlier in this section, and from many other problems unique to marine data. For example, marine observers originally measured sea surface temperatures by throwing a canvas or wooden bucket over the side of a ship, hauling up some water, and inserting a thermometer. Some of these buckets were

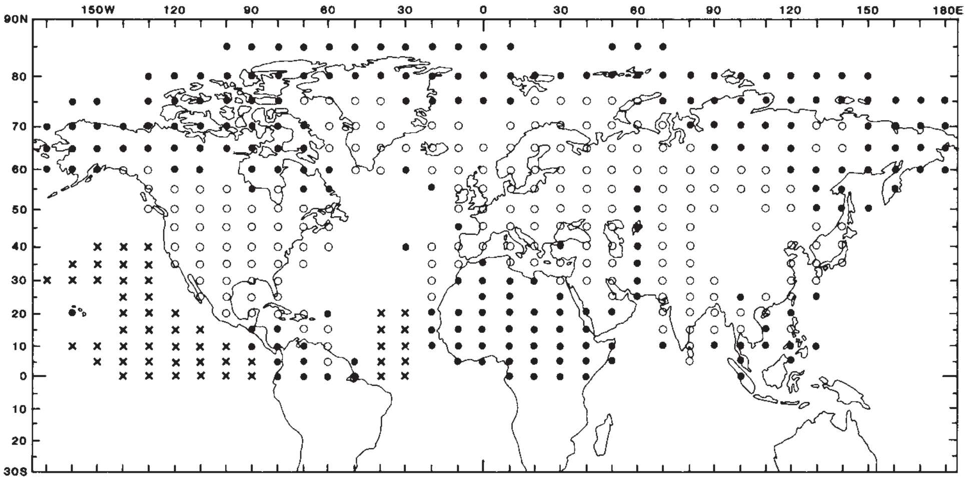


Figure 11.9

Northern hemisphere climatological data coverage, 1881–1980. To count as "covered," a gridpoint needed at least six stations within 300 nautical miles, each with at least 10 years' data. Open circles: gridpoints with data available starting in 1900. Full circles: gridpoints with data coverage starting in the 1950s. Crosses: gridpoints without sufficient data even after a "relaxed interpolation/extrapolation procedure." Full circles plus open circles represent usable gridpoints. Note the absence of data points across most of the People's Republic of China.

Source: P. D. Jones et al., "Variations in Surface Air Temperatures: Part 1. Northern Hemisphere, 1881–1980," *Monthly Weather Review* 110, no. 2 (1982), 67.

insulated; others were not. With the advent of powered vessels, many fleets switched to a different technique, recording temperatures with a sensor placed in the ship's engine-cooling-water intake. Heat from the engine can raise the detected value if the sensor is nearby, biasing these readings warmer than the sea outside the vessel. The exact depth of the engine intake also affects readings. These three methods produce systematically different results, and variations within each method create further inhomogeneities.[41] The more scientists inverted the infrastructure and recovered metadata, the more they could use algorithms to render data collected by these various methods comparable.[42]

Metadata Friction

One might imagine that by now every conceivable source of error and possible improvement would have been found, but this is certainly not the case. Take an example: In the 1990s climatologists began mining WMO Publication No. 47, *International List of Selected, Supplementary and Auxiliary Ships* (published annually starting in 1955, with numerous irregularly issued supplements). These documents described numerous features of ships and their onboard observing systems, such as what type of thermometer, anemometer, and barometer they used to measure atmospheric variables and their height above the sea surface. Publication No. 47's original purpose was to help weather forecasters identify and interpret data reported by Voluntary Observing Ships. Once a new version arrived, these operational users had no need for the old one, so they discarded it. When climate scientists began building the Comprehensive Ocean-Atmosphere Data Set, they used these manuals as metadata. But most of the year-by-year manuals and supplements had been discarded long before.

The effort involved in finding existing metadata, digitizing them, and combining them with whatever metadata you already have might be termed "metadata friction." Starting in the latter half of the 1990s, investigators gradually recovered older copies of Publication No. 47, its supplements, and other similar manuals, such as the 1963 UK *Marine Observer's Handbook* and *Lloyd's Register of Shipping*. They then digitized them and added them to existing marine metadata. Where early investigators had applied corrections on a fleet-wide basis, these metadata permitted an increasingly fine-grained application of corrections, down to the level of individual ships.[43] Recently these metadata recovery efforts have led to the detection and explanation of a large, sudden drop (–0.3°C) in sea surface temperature (SST) starting in 1945, probably related to a sudden switch in the COADS database from dominance by data from US-based vessels using engine-intake measurements (1942–1945) to data from UK-registered vessels using bucket measurements (starting in 1945).[44] This spurious drop, if confirmed, will be corrected in future versions of SST data.

Unlike the surface station reports in *WMSSC* and its predecessors, most marine records were not inherently climatological. Instead, because ships, weather buoys, and most other marine observing platforms are always moving, these logs were simply weather records, each for a particular location on a particular day. Despite the large number of individual records, outside the major shipping lanes the amount of data for any given grid cell in any given month is small, especially in the little-traveled southern hemisphere. The ICOADS project gridded these data, initially at 2°×2°. Then it calculated climatological statistics for each grid cell. In effect, ICOADS treated each cell as a single weather station, integrating all measurements from all platforms within that cell for each analysis period (monthly, yearly, etc.) to create a climatological data set. Combining data from such a large area (up to 200×200 km, depending on latitude) makes more sense at sea than it would on land, where topography and vegetation create large local and regional variations.

The aforementioned studies have brought increasing attention to sea surface temperature, rather than land air temperature, as a potential marker of global climate change:

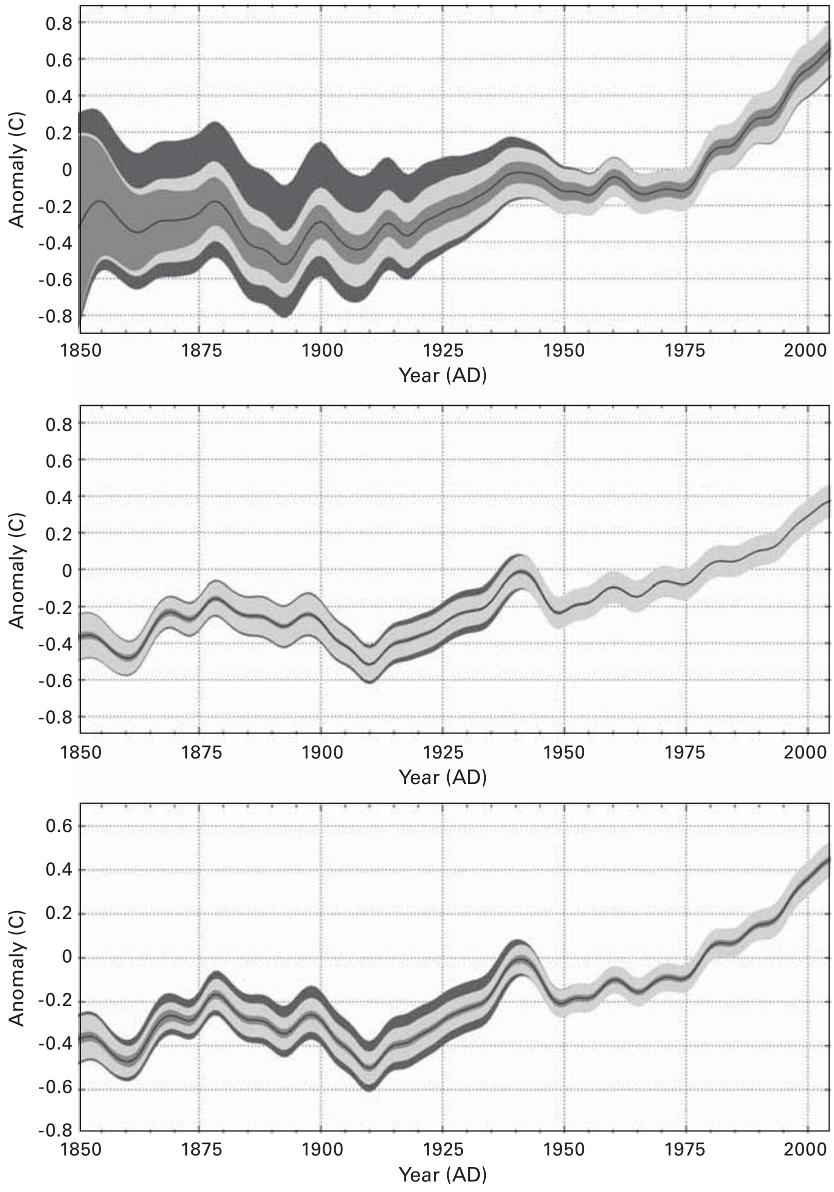
> Surface air temperature over land is much more variable than the SST. SSTs change slowly and are highly correlated in space; but the land air temperature at a given station has a lower correlation with regional and global temperatures than a point SST measurement, because land air temperature (LAT) anomalies can change rapidly in both time and space. *This means that one SST measurement is more informative about large-scale temperature averages than one LAT measurement*.[45]

Furthermore, marine air temperature near the surface generally correlates closely with sea surface temperature. Therefore, SST can serve as a reasonable surrogate for marine air temperatures.[46]

Proliferation within Convergence: Climate Data Today

By 1986, the Climatic Research Unit had combined the COADS marine data with its land surface data to produce the first comprehensive global surface temperature dataset.[47] Since then, this dataset—now in its third release, known as HadCRUT3—has been co-produced by the UK's Hadley Centre, which manages marine records, and the CRU, which handles land surface data.[48] In the decades since these pioneering efforts, more and more land and marine records have been retrieved and digitized. The third release of the CRU land surface temperature dataset CRUTEM3 contained records for 4349 land stations—more than twice the number contained in the 1986 release. While many of the additional records represent new stations added since 1986, many others were historical records omitted from *WWR*, *MCDW*, and *WMSSC* and recovered through an ongoing search.

Figure 11.10 shows temperature anomaly trends for land surface (top) and SST (middle) components of the HadCRUT3 data set. The much larger uncertainties in the land data stem from the much greater variability of land surface temperatures, as noted above. The combined whole-global trend (bottom) reduces the overall uncertainty because the SST (at 70 percent of total surface area) dominates. A notable feature of these charts is that while the best-estimate land and SST trends are essentially identical


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Anomaly (C)
1850
1875
1900
1925
1950
1975
2000
Year (AD)
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Anomaly (C)
1850
1875
1900
1925
1950
1975
2000
Year (AD)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Anomaly (C)
1850
1875
1900
1925
1950
1975
2000
Year (AD)

until about 1980, they diverge after that, with land temperatures rising faster than SST. The reasons for this remain unclear. As Brohan et al. explain, this "could be a real effect, the land warming faster than the ocean (this is an expected response to increasing greenhouse gas concentrations in the atmosphere), but it could also indicate a change in the atmospheric circulation, it could indicate an uncorrected bias in one or both data sources, or it could be a combination of these effects."[49] Hence, the work of infrastructural inversion remains unfinished, as it probably always will.

The trend lines in figure 11.10 also represent model/data symbiosis (discussed in the previous chapter) in action. The data behind these trend lines were first adjusted to reduce homogeneities, then gridded using an interpolation procedure, then adjusted for variance (differences in the number of observations available within each grid cell). Without these adjustments, the data could not be joined into a (relatively) uniform and usable whole.

What you learn from all this is that if you want global data, you have to make them. You do that by inverting the infrastructure, recovering metadata, and using models and algorithms to blend and smooth out diverse, heterogeneous data that are unevenly distributed in space and time. There are many ways to do this, and as a result there have been many versions of global data: global climate, plural, a moving target that continues to shift (albeit within a restricted range). Since the 1980s, a series of projects have vastly expanded the number and type of measurements available to do this, as well as the available tools. These began with digitized traditional climatological data from *WWR* and *MCDW*, but they have expanded to incorporate other sources such as marine weather records

◂ **Figure 11.10**
Global averages of land (top), marine (middle), and combined land-marine (bottom) atmospheric temperatures from HadCRUT3. The combined land-marine data have lower uncertainty than the land surface data alone because 70 percent of Earth's surface is ocean. From original legend: "The black [center] line is the best estimate value; the [medium gray] band gives the 95% uncertainty range caused by station, sampling, and measurement errors; the [light gray] band adds the 95% error range due to limited coverage; and the [dark gray] band adds the 95% error range due to bias errors."
Source: P. Brohan et al., "Uncertainty Estimates in Regional and Global Observed Temperature Changes: A New Dataset From 1850," *Journal of Geophysical Research* 111, no. D12106 (2006). Colors replaced as noted in brackets to allow non-color reproduction.

(COADS). The sustained effort and expertise required has had the effect of concentrating analysis in a few “centers of calculation” (in Bruno Latour’s phrase)—principally the Climatic Research Unit at the University of East Anglia, the UK Hadley Centre, NASA’s Goddard Institute for Space Studies, and the US National Climatic Data Center. These efforts represent the *ex post facto* standardization process I described at the beginning of chapter 10. They have narrowed the range of global temperature trend estimates considerably since 30 years ago, when digitization began. Yet each of these efforts continues to produce new versions of its global data. Furthermore, as we are about to see, this is not the only way scientists have found to re-create the history of climate.