The Definition of the Standard WMO Climate Normal

The Key to Deriving Alternative Climate Normals

BY ANTHONY ARGUEZ AND RUSSELL S. VOSE

The World Meteorological Organization (WMO) and its predecessor, the International Meteorological Organization (IMO), have been coordinating the publication of global climate normals at the monthly scale for about 75 years. Member nations of the IMO/WMO were first mandated to compute climate normals for their respective countries for the 1901–30 period, and are required to update these climate normals every 30 years, resulting in the 1931–60 normals and the 1961–90 normals. Since 1956, the WMO has recommended that each member country recompute their 30-year climate normals every 10 years. Although some member countries do not update their climate normals every decade, for ease of comprehension we hereafter refer to the recommended decadally updated 30-year average as the standard WMO climate normal.

Given substantial evidence (e.g., Solomon et al. 2007; Milly et al. 2008) indicating that the stationarity of climate statistics can no longer be (and never should have been) taken for granted, the justification for using a 30-yr normal for describing current and future climate conditions has increasingly been called into question (e.g., the 2007 *Journal of Applied Meteorology and Climatology* article by Livezey et al., hereafter referred to as L07). The key problem is that climate normals are calculated retrospectively, but are often utilized prospectively. Specifically, climate normals are calculated using data from a recent 30-yr period, but one of their primary utilities is to provide stakeholders and decision makers with a metric of future climate conditions that can be taken into account in long-term planning considerations. The utilization of climate normals in this manner adheres to the well-known maxim, "The best predictor of future behavior is past behavior." Implicit in this link between the calculation and the utilization of climate normals is the notion of stationarity. Weak stationarity assumes that the expectation (i.e., the mean value) of a variable is time invariant, and that second-moment statistics are a function of lag only. Significant trends in a time series (as opposed to natural fluctuations about a mean state) violate the weak stationarity assumption. In turn, if stationarity is violated, a retrospective 30-yr average becomes considerably less useful as an indicator of current and future climate conditions.

As discussed by WMO (2007), climate normals are not only used as predictors of future climate conditions, but are also used to provide a reference value for the computation of climate anomalies. For placing current climate conditions in a historical perspective (i.e., real-time climate monitoring), there are compelling statistical reasons to use climate normals that are rarely updated—if at all—so that the meaning of a particular anomaly value will be consistent across time. This is true whether there are significant trends in climate time series or not. Similarly, for stationary climate time series, there would be little reason to update climate normals because, by definition, a stationary climate's mean does not change in time. The 30-yr climate normal under the stationarity assumption could be interpreted as the true background state, offset by decadal and longer-term tendencies, and further tweaked by interannual variability (e.g., ENSO-related variations) as well as random and systematic errors. Thus, for stationary time series, the standard WMO climate normal is a reasonable

AFFILIATIONS: ARGUEZ AND VOSE—NOAA's National Climatic Data Center, Asheville, North Carolina
CORRESPONDING AUTHOR: Anthony Arguez, NOAA/NCDC Room 506, 151 Patton Avenue, Asheville, NC 28801
E-mail: Anthony.Arguez@noaa.gov

DOI: 10.1175/2010BAMS2955.1

metric with respect to both of its primary utilizations. Conversely, if climate conditions are deemed to be nonstationary, the standard WMO normal still retains its utility for placing current conditions in a historical context, but the predictive value is compromised.

Climate scientists have been concerned with the definition of climate normals since before the WMO mandate was put in place, with renewed interest in the last 30 years due in large part to observed climate change. To address the shortcomings of traditional climate normals in a changing climate, L07 and others have been advocating for the development of alternative normal products that are better indicators of current and future climate conditions. We contend that the most straightforward approach for creating alternative climate normals is to alter the generalized definition of the standard WMO climate normal. Arguably, every possible alternative climate normal that can be devised is the result of altering one or more of five fundamental attributes. Below, we describe these five attributes, formulate a generalized equation for WMO-type climate normals, and briefly consider a few ways to alter the standard WMO definition to arrive at alternative climate normals.

FIVE KEY ATTRIBUTES OF THE STANDARD WMO CLIMATE NORMAL.

Although climate normals are simply 30-yr averages, the computation of climate normals is a nontrivial, multifaceted process. The WMO provides member nations with considerable leeway on the methodology employed in computing climate normals, such as quality control, the handling of missing data values, etc. Here, we ignore these methodological specifics and restrict ourselves to the statistical definition of the standard WMO climate normal (i.e., the metric of "typical" climate conditions). There are five important attributes of the normals metric:

- it is a temporal average;
- the average is unweighted;
- the averaging period is 30 consecutive years;
- it is a causal filter (using past and current values only); and
- it is updated once per decade.

Considering these five attributes, the generalized equation form for this class of average-based normals metrics is as follows:

$$y\left(t_0+k\Delta t\right)=\sum_{i=t_0+k\Delta t-N+1}^{t_0+k\Delta t} w\left(i\right)x\left(i\right) \tag{1}$$

Here, y is the climate normal, x is the observed annual time series, w is a weighting function, k is an integer, Δt is the update frequency, t_0 is a reference year, and N is the number of years averaged. For the standard WMO climate normal, $N = 30$, w is set to a constant value of 1/30, $\Delta t = 10$ years, and t_0 is a multiple of 10 years. Substituting, the standard WMO climate normal metric is defined as follows:

$$y\left(t_0+10k\right)=\frac{1}{30}\sum_{i=t_0+10k-29}^{t_0+10k} x\left(i\right) \tag{2}$$

For the case of the 1971–2000 climate normals (setting $k = 0$, presuming $t_0 = 2{,}000$), Eq. (2) reduces further to an even more familiar form as follows:

$$y\left(2000\right)=\frac{1}{30}\sum_{i=1971}^{2000} x\left(i\right) \tag{3}$$

Alternative normals products can be created by changing one or more of the five attributes listed above. In the remainder of this section, we provide additional details for each of the five attributes, and briefly describe how the attributes can be modified to arrive at alternative normals.

Temporal average. The defining characteristic of traditional climate normals is that they are based on averages. The average, or mean, is ubiquitous in weather and climate applications as an indication of central tendency. Specifically, climate normals are temporal averages, and can be considered running averages of sorts, although they are only updated once per decade. In time series filtering theory, a running average is a very simple low-pass filter, which means it smoothes out high-frequency variations (e.g., year-to-year to interannual fluctuations such as those associated with the El Niño–Southern Oscillation) to highlight a background state. Assuming stationarity, the rationale is that these higher-frequency fluctuations are superimposed on the mean background state; this background state is precisely what the WMO climate normal metric attempts to quantify.

There is no natural law mandating that "typical" weather conditions be represented as an averaged value. The median is a viable alternative that also provides a measure of central tendency. Further, a strong trend in a climate time series renders a temporal average an unsuitable choice for describing a background climate state. A temporal average essentially undermines the predictability inherent with a trend, since it involves simply taking the arithmetic mean of 30 values without regard to their temporal ordering, effectively smoothing out relative outliers in the first and second halves of the time series.

Truly time-dependent normals exist that do not rely on averaging. For example, L07 shows that a simple regression line can be considered a time-dependent normal. The point in time through which the regression line passes is the normal value for that year. Specifically, L07 proposes a Hinge Fit regression consisting of a constant value through 1975 and a linear fit thereafter. Similarly, the relatively new technique known as Empirical Mode Decomposition (EMD) has been used to define a normals metric. The lowest-order residual time series resulting from EMD analysis of climate time series is purported to represent a climate normal function. Both of these methods may be particularly useful for defining "normal" conditions for time series that exhibit large trends (either positive or negative).

Unweighted. The WMO climate normal is an unweighted average. Every single year in the averaging period imparts the same influence on the normal value. Therefore, the first year of the period has the same influence as the last year. Similarly, the first half of the period exerts the same influence as the second half. As an example, consider the 1971–2000 normals. The 1971–85 subperiod has the same impact as the 1986–2000 subperiod, whereas the individual contributions of the 1971 value and the 2000 value are equivalent. For a climate series that exhibits neither a significant trend nor positive serial autocorrelation, there is little incentive to use a weighted average. However, observations do indicate that significant trends in temperature, for example, exist over many parts of the world. Therefore, it is conceivably advantageous to provide greater weight to more recent data and limit the influence of the earliest values. This could be imposed via the function w in (1). Presumably, w would take the form of a monotonically increasing function (i.e., each successive year would be assigned a greater weight than the previous year). The weights could be determined based on theoretical techniques developed for filtering near endpoints, such as those described by Mann (2004, 2008) and Arguez et al. (2008). Alternatively, empirically determined weights could be utilized based on individual time series characteristics, analogous to the empirical weight exercise employed by Arguez et al. (2008).

Thirty years. Arguably the most intuitive and practical alternative to a 30-yr normal is to average over a different number of years (N). Basing climate normals on 30-yr averages has been standard practice for almost a century now, since the IMO first mandated that member countries provide climate normals for their respective countries. Interestingly, elementary statistics texts often state that a sample size of 30 is the "rule of thumb" threshold for which reliable estimates can be determined.

Considering climate change (e.g., the warming that has occurred over much of the U.S. since the 1970s), one would expect a shorter time interval average would be more representative of the current state of the climate, at the time of reporting, than a 30-yr average. Changing the value of N in (1) results in a simple alternative normal. Technically, this can also be accomplished by fixing N to a large value and removing unwanted years by setting the corresponding values of w to zero, essentially imposing a filtering window. However, we include both parameters N and w to highlight the distinctions between weighted averages and unweighted N-yr averages.

An abundance of anecdotal evidence suggests that the U.S. energy industry, particularly with respect to load forecasting by utilities and rate setting by state agencies, is moving to shorter-term averages for determining "normal" weather (McMenamin 2008; J. Sanderson 2007, personal communication; C. Marple 2007, personal communication; A. Heinen 2007, personal communication; T. Hennessey 2008, personal communication). It is not uncommon for industry representatives to utilize 10-, 15-, and/or 20-yr normals, although the number of years to average over (N) is sometimes determined somewhat arbitrarily and/or a posteriori.

In a 1996 *Journal of Climate* article, Huang et al. developed a method for computing normals based on an "optimal" averaging period (N). These so-called Optimal Climate Normals (OCN) are based on the predictive skill of normals for a 1-yr lead time. Citing practical reasons for choosing fixed averaging periods for the entire United States, their analysis determined that the optimal averaging period is 10 years for temperature normals and 15 years for precipitation normals over the United States. More recently, L07 argued that the N values for computing OCN should be computed separately for each of a station's annually sampled time series. It is easily shown that for stations exhibiting near-zero trends, the N value determined by the OCN technique is typically greater than 30 years. This is because, for a seemingly stationary time series, the best estimate results when the largest possible sample is included in the average. For time series with very large trends—regardless of sign—the OCN technique as described in L07 can result in N values much smaller than 30 (in practice as low as 5 years) for U.S. monthly temperatures.

Causal filter. Time-series filtering is used to extract salient time scales from time series, often to "smooth out" high-frequency variations. A causal filter is a filter in which the output value—the filtered value—is a function of past and/or present values only. The implication is that the current filtered value was "caused" by the previously recorded conditions. The standard WMO climate normal is essentially computed as a causal filter, since it is calculated retrospectively. This is inferred from (1) because the index of y is identical to the upper summation limit, meaning that the normal value is a function of past and present values only.

This stands in sharp contrast to acausal filters, which depend on "future" values. Acausal filtering, such as using conventional running means, typically results in filtered values that depict the midpoint of the filtering range. Thus, acausal filters are often referred to as *centered* filters. For example, a 5-month running average of August–December 2010 temperature values represents a smoothed value for October 2010. Consequently, the filtered value for October cannot be computed until data for December are available.

Following this alternate convention, it is reasonable to regard the 1971–2000 climate normals as indications of typical climate conditions for 1985/1986, which is the midpoint of the averaging range. The next recommended installment of WMO climate normals (covering 1981–2010) will be released no sooner than 2011. Until this product release, the "current" climate normals will be, arguably, up to ~25 years out-of-date. However, note that even when a new product is released every decade, the *centering* aspect of filter theory implies that standard WMO climate normals will always be *at least* 15 years out-of-date.

There are several ways to alter the normals metric definition such that the output value is indicative of the time of computation, rather than indicative of the middle of the averaging range. One indirect option was discussed earlier: using filter weights, determined either empirically or theoretically, to allow more recent observations to exert more influence on the average. However, a truly centered, acausal solution requires extrapolation, inevitably injecting some degree of prediction error. Predicting future values can either be accomplished via statistical methods (such as autoregressive models) or via downscaled climate model projections. A 30-yr average centered on today could be computed from the most recent 15 years of observations, along with the forecast for the next 15 years. In work commissioned by the U.K. energy industry, the Met Office Hadley Centre has used an analogous approach to update the climatological temperature baselines used in energy demand planning. A dynamical decadal prediction system was used to "extend" observed historical temperature records into the future. The long-term temperature average centered on the current year, or any year in the forthcoming decade, was then calculated using a mix of observed and predicted temperatures (personal communication, Richard Graham).

Decadal updates. The WMO mandates member countries to compute 30-yr normals once every 30 years (1901–30, 1931–60, 1961–90, 1991–2020, etc.), but recommends that member countries create decadal updates as well. Presuming stationarity, the true mean background state (μ) would not fluctuate from one decade to another (or from one 30-yr period to another), yet differences between decadal updates would mostly highlight long-term variability (and shorter-term variability to a lesser extent) superimposed on a constant background state. Conversely, if we presume a trend exists in the data record, then decadal updates become essential for monitoring such a trend's effects on what is considered "normal." In fact, a prominent trend would warrant that updates be initiated as frequently as possible. The obvious alternative to a decadally updated climate normal is to update the 30-yr average annually—setting Δt equal to 1 yr in (1)—as recommended in L07. Simple calculations using monthly mean temperature data demonstrate that for station-month time series exhibiting strong relative trends, annually updated climate normals can outperform decadally updated normals over 90% of the time as the decadal average becomes more out-of-date during the intervening decade between calculations of standard WMO climate normals. This effect is magnified for member nations that only compute normals every 30 years.

CONCLUSIONS. The standard WMO climate normal is a useful, albeit imperfect, metric. Indeed, no metric can be perfect by definition. Climate change, and in particular significant nonzero trends in climate time series, renders the standard WMO climate normal less useful. For use as a reference period average for computing climate anomalies, climate normals retain their usefulness despite climate change, although updating the reference period can lead to dramatic changes in the anomaly values (and their interpretations). Climate monitoring centers should proceed with caution if and when base periods are changed for computing real-time anomalies. If we accept that climate conditions are indeed nonstationary, then for the purposes of providing more accurate depictions of current and future climate conditions, climate normals should be 1) updated as frequently as possible (i.e., annually); and/or 2) computed in an alternative manner. Alternative approaches include choosing $N \neq 30$, computing climate normals as an acausal filter, using a weighted average, and/or redefining "normal" as some quantity other than an average.

Note that we have focused on the definition of the climate-normals metric, which is a statistical construct. While the statistical definition is universal, the real-world applicability of a particular alternative is not. For example, it is highly likely that the best alternative for monthly temperature normals will differ for monthly precipitation normals; consider the possibility of defining "normal" as a 15-yr average for the former and a 40-yr median for the latter. Further, varying underlying time series characteristics, such as trend and residual autocorrelation (L07), result in seasonal and regional disparities in the performance of particular alternative techniques. These issues need to be considered in any evaluation of alternative techniques.

Clearly, the standard WMO climate normal is not ideal in an era of observed climate change. Future work should be undertaken to identify a thorough list of alternative climate normals, conduct an evaluation of all viable techniques, and recommend and provide specific alternative normals products to stakeholders and decision makers. It is our contention that accurate depictions of current and future climate conditions necessitate the development of alternative climate normal products.

ACKNOWLEDGMENTS. The authors kindly thank R. Graham, T. Peterson, P. Thorne, and T. Karl for fruitful discussions on this topic, as well as M. Kruk and M. Squires for reviewing earlier versions of this manuscript. The authors also appreciate the instructive comments of several anonymous reviewers. Partial support for this work was provided by the Office of Biological and Environmental Research, U.S. Department of Energy (Grant number DE-AI02-96ER62276).

FOR FURTHER READING

Angel, J. R., W. R. Easterling, and S. W. Kirtsch, 1993: Towards defining appropriate averaging periods for climate normals. *Clim. Bulletin,* **27,** 29–44.

Arguez, A., P. Yu, and J. J. O'Brien, 2008: A new method for time series filtering near endpoints. *J. Atmos. Oceanic Technol.,* **25,** 534–546.

Hamming, R. W., 1989: *Digital Filters.* 3rd ed., Prentice Hall, 284 pp.

Huang, J., H. M. van den Dool, and A. G. Barnston, 1996: Long-lead seasonal temperature prediction using optimal climate normals. *J. Climate,* **9,** 809–817.

Huang, N. E., and S. S. P. Shen, Eds., 2005: *Hilbert–Huang Transform and Its Applications.* World Scientific, 360 pp.

Kunkel, K. E., and A. Court, 1990: Climatic means and normals—A statement of the American Association of State Climatologists (AASC). *Bull. Amer. Meteor. Soc.,* **71,** 201–204.

Lamb, P. J., and S. A. Changnon Jr., 1981: On the "best" temperature and precipitation normals: The Illinois situation. *J. Appl. Met.,* **20,** 1383–1390.

Livezey, R. E., K. Y. Vinnikov, M. M. Timofeyeva, R. Tinker, and H. M. van den Dool, 2007: Estimation and extrapolation of climate normals and climatic trends. *J. Appl. Met. Climatol.,* **46,** 1759–1776.

Mann, M. E., 2004: On smoothing potentially nonstationary climate time series. *Geophys. Res. Lett.,* **31,** L07214, doi:10.1029/2004GL019569.

——, 2008: Smoothing of climate time series revisited. *Geophys. Res. Lett.,* **35,** L16708, doi:10.1029/2008GL034716.

Marvin, C. F., 1923: Concerning normals, secular trends and climatic changes. *Mon. Wea. Rev.,* **51,** 383–390.

McMenamin, J. S., Weather Normalization Benchmarking Study, Itron, Inc., cited 2008: Weather normalization. [Available online at www.aeic.org/load_research/docs/1_WeatherNorm_McMenamin.ppt.]

Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate,* **22,** 1700–1717.
——, ——, and R. S. Vose, 2009: The U.S. Historical Climatology Network Monthly Temperature Data, Version 2. *Bull. Amer. Meteor. Soc.,* **90,** 993–1007.
Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science,* **319,** 573–574.
Sabin, T. E., and M. D. Shulman, 1985: A statistical evaluation of the efficiency of the climatic normal as a predictor. *Intl. J. Climatol.,* **5,** 63–77.
Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Avery, M. Tignor, and H. L. Miller, Eds., 2007: *Climate Change 2007: The Physical Science Basis.* Cambridge University Press, 996 pp.
World Meteorological Organization, 1989: Calculation of Monthly and Annual 30-Year Standard Normals. WCDP-No. 10, WMO-TD/No. 341, World Meteorological Organization.
——, 2007: The Role of Climatological Normals in a Changing Climate. WCDMP-No. 61, WMO-TD/No. 1377, World Meteorological Organization.

The Integrated Surface Database

Recent Developments and Partnerships

BY ADAM SMITH, NEAL LOTT, AND RUSS VOSE

Hourly surface-based meteorological observations are the most-used, most-requested type of climatological data, but historically they have been scattered across multiple repositories worldwide in a variety of disparate formats. This greatly complicated the life of the end user and significantly increased the cost of data usage. To address this problem, in 1998 NOAA's National Climatic Data Center (NCDC) initiated the Integrated Surface Database (ISD) project. The goal of the project was to merge numerous surface hourly datasets into a common format and data model, thus providing a single collection of global hourly data for the user that was continuously updated and available. Additional benefits of integration include the reduction of subjectivity and inconsistencies among datasets that span multiple observing networks and platforms; standardized quality control (QC) based on reporting time resolution (e.g., a QC methodology for hourly temperature data independent of network); and products that are more easily developed and improved by collective experience and expertise.

The outcome of this effort is a dataset containing data from more than 100 original data sources that collectively archived hundreds of meteorological variables. The primary data sources include the Automated Surface Observing System (ASOS), Automated Weather Observing System (AWOS), Synoptic, Airways, METAR, Coastal Marine (CMAN), Buoy, and various others, from both military and civilian stations including both automated and manual observations. "Summary of day" parameters such as maximum/minimum temperature, 24-h precipitation, and snow depth are also included in ISD, to the extent that they are reported in the hourly data sources. Also, for ASOS sites, the daily summaries transmitted by each station are now being ingested into ISD. Some of the most common meteorological parameters include wind speed and direction, wind gust, temperature, dew point, cloud data, sea level pressure, altimeter setting, station pressure, present weather, visibility, precipitation amounts for various time periods, and snow depth. Total data

AFFILIATIONS: SMITH, LOTT, AND VOSE—NOAA's National Climatic Data Center, Asheville, North Carolina
CORRESPONDING AUTHOR: Adam Smith, NOAA's National Climatic Data Center, Federal Building, 151 Patton Avenue, Asheville, NC 28801-5001
E-mail: Adam.Smith@noaa.gov

DOI:10.1175/2011BAMS3015.1

volume (uncompressed) is approximately 500 GB.

ISD contains over 2 billion surface weather observations from more than 20,000 stations worldwide included in the archive (1900–present). Figure 1 shows the spatial distribution of reporting ISD stations in 1925, 1950, 1975, and 2000. Since 1950, spatial coverage has been quite reasonable over North America, Europe, Australia, and parts of Asia, with noteworthy gaps in Africa and South America until the early 1970s, when the Global Telecommunications System came into existence. At present there are more than 11,000 active stations that are updated daily in the database (i.e., near real-time data that are ingested each day). Figure 2 depicts the approximate number of stations per year, which generally increase through time. One notable exception is the decline in reporting stations during the late 1960s through early 1970s due to the transition from keying of data to digital transmission/receipt of data. Some stations have more than 50 years of continuous reporting during the latter half of the time period; however, many stations have breaks in the period of record (e.g., 40 years of data may be spread over a 70-year period).

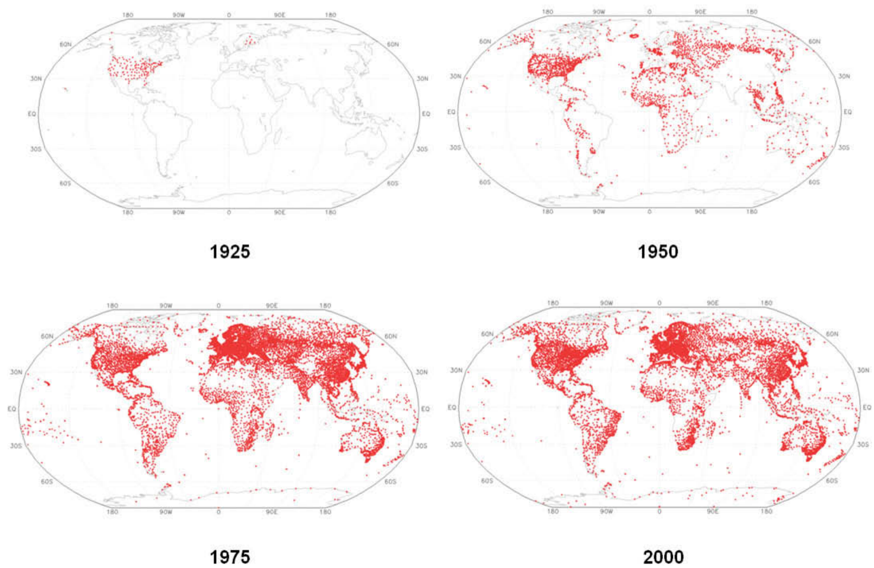

Fig. 1. The red dots represent the distribution of stations that contribute to the ISD data collection. Fewer stations were reporting in the early twentieth century. Since 1950, station coverage has been reasonable over North America, Europe, Australia, and parts of Asia, with noteworthy gaps in Africa and South America until the early 1970s, when the Global Telecommunications System came into existence.

ISD Version 1 was released in 2001, with Version 2 (additional quality control applied) following in 2003. Thereafter, continued incremental improvements have been implemented in automated quality control software, along with additional partnerships to further enhance the temporal and spatial coverage of the data. Current ISD partnerships include:

- the Federal Climate Complex (FCC) USAF Fourteenth Weather Squadron and the U.S. Navy Fleet Numerical Meteorological and Oceanographic Command Detachment (FNMOC Det), which provide historical data along with current data streams of global hourly, synoptic, and military station data (note: the FCC in Asheville, North Carolina, consists of NCDC and its DoD partners);
- NOAA's National Weather Service (NWS), the Federal Aviation Administration (FAA), and NOAA's Climate Reference Network (CRN), which provide data streams into ISD on a daily basis;
- the Climate Data Modernization Program (CDMP), which provides for publications and forms as far back as the 1800s, such as U.S. data prior to 1948. These are scanned, digitized, and integrated into ISD, and include data processing at the Northeast Regional Climate Center (NERCC); and
- the National Center for Atmospheric Research (NCAR), which provides numerous datasets of global and national origin.

The remainder of the paper is structured as follows: section 2 provides a brief overview of the QC system; section 3 provides examples of ISD usage in research and industry; section 4 discusses recent progress and future plans for ISD; and lastly, section 5 provides information for accessing a wide variety of ISD data applications, products, and services.

QUALITY CONTROL. It is important to note that a number of datasets included in ISD already have internal quality control procedures such as the Climate Reference Network (CRN), Regional U.S. Historical

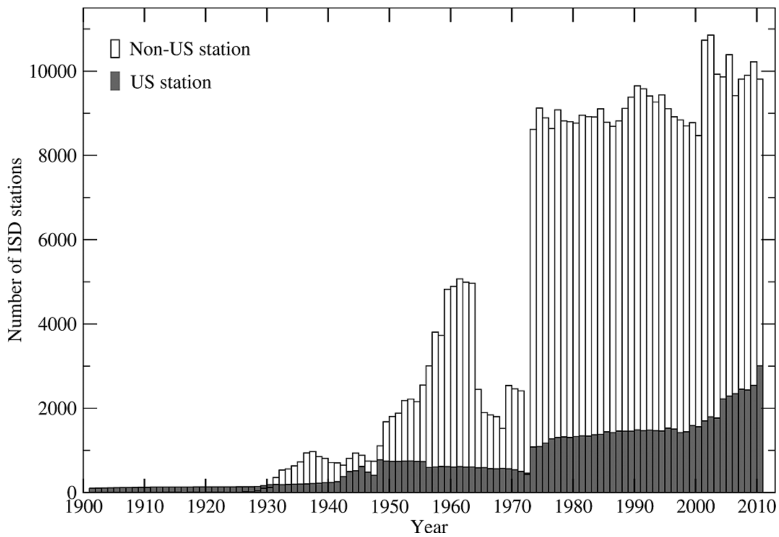

Fig. 2. There are more than 11,000 stations actively reporting hourly data and updated daily in the ISD. A notable decline occurs for both U.S. and non-U.S. ISD stations during the late 1960s–early 1970s due to the transition from keying of data to digital transmission/receipt of data causing interruption in reporting.

Climatology Network, ASOS/AWOS, CDMP, U.S. Air Force global hourly data, and U.S. hourly precipitation data. However, the ISD provides integration of many disparate datasets and additional QC checks to better facilitate data access.

Since 2003, there have been continued incremental improvements in automated QC software. ISD contains 54 quality control (QC) algorithms, which serve to process each data observation through a series of validity checks, extreme value checks, internal (within observation) consistency checks, and external (versus another observation for the same station) continuity checks. This QC is conservative in that it was designed to eliminate obvious errors in the data, minimize overflagging of data, and ensure to the greatest extent possible that valid values were not removed or flagged as erroneous. However, this does not include any spatial quality control (e.g., buddy checks with nearby stations). Such checks are employed at the source dataset level in some cases and provide an opportunity to further improve ISD in the future.

Though all data observation parameters are quality controlled as briefly described above, the parameters validated most extensively are wind data, temperature and dew point data, pressure data, cloud data, visibility and present weather data, precipitation amounts, and snowfall and snow depth. Each day, the ISD is updated with new global hourly data, and the QC process is applied to each day's data. Therefore, the full period of record, including the latest day's data, have been through a consistent QC process, which is a key aspect for spatially variable, research-quality data.

EXAMPLES OF ISD USAGE IN RESEARCH AND INDUSTRY.

A number of peer-reviewed research studies have employed historical data records from ISD. For instance, Willett et al. (2007) derived a homogenized gridded dataset of surface humidity from ISD to examine changes in surface-specific humidity over the late twentieth century. Camalier et al. (2007) used ISD data to model the effects of meteorology on ozone in 39 urban areas in the eastern United States. Zou (2009) applied ISD data in a comparative evaluation of the accuracy levels of exposure risk estimate models. Brown and DeGaetano (2009) employed data from 10 stations in the conterminous United States to develop a method to detect inhomogeneities in historical hourly dew point data. Compo et al. (2011) utilized ISD as one of the primary data sources to develop a gridded global pressure reanalysis for the twentieth century.

Innovative usage of ISD data is also occurring in the private business/industry sector. The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) uses ISD as input for the data summaries/tables in its *Handbook of Fundamentals*, which provides climatic design information for 5,564 global locations, with more than 500 parameters in each table. The climatic design information is commonly used for design, sizing, distribution, installation, and marketing of heating, ventilating, air conditioning, and dehumidification equipment, as well as for other energy-related processes in residential, agricultural, commercial, and industrial applications. These summaries include various values of dry-bulb, wet-bulb, and dew-point temperature; monthly degree-days to various bases; clear sky solar irradiance; and wind speed with direction at various frequencies of occurrence (e.g., 0.4%). A subset of the elements most often used for stations representing major urban centers is also presented in the handbook. Prior to using ISD, the tables only included several hundred locations in the United States and Canada.

Other examples of ISD data usage include

- engineering design such as ice loads for towers, cables, wires, etc.;
- wind loads for buildings, etc.;
- drainage/runoff extremes (pipes, culverts);
- aircraft operations: crosswinds (runway design), instrument landing systems, etc.;
- ship routing and oil rig placement;
- global reanalyses for climate trends assessment, etc.;
- HAZMAT operations and studies: oil spills, toxic release, etc.;
- weather risk management industry: estimates of risk and verification;
- insurance investigations and verification;
- court cases and criminal investigations;
- aircraft accident investigations;
- wind energy studies: wind farms, United States and overseas; and
- commercial innovation and design: typical and extreme conditions for a new market.

RECENT PROGRESS/FUTURE PLANS.

Efforts are well underway to integrate additional data sources into ISD, which will provide additional U.S. data prior to 1950 and some data prior to 1900. Plans are also in place to gradually integrate hourly datasets provided by various countries to increase data coverage and periods of record for some areas. Most recently, data sets from Brazil, Australia, Greenland, and Mexico were converted to ISD format, quality controlled, merged, and integrated into ISD. This effort, in addition to the CDMP data preservation effort, has provided an additional 54 million surface observations covering a period of more than 100 years for integration into the ISD.

Future plans for ISD include integrating additional datasets and data sources, enhancing the metadata, developing additional applications and products based on customer requirements, further refining and developing new QC techniques, and incorporating additional operational data streams and data partners. We also plan to better consolidate the station ID numbers over time, so that to the greatest extent possible, a single station location will have a single station ID for its full period of record. Historically, this has been an issue with many data sources.

In the future, a high priority will be continuing to reduce global climate data gaps in both space and time, especially in the Southern Hemisphere where gaps are large. It is particularly important to assist other countries in the world where climate data can be rescued (e.g., CDMP), vetted through appropriate QC checks, and integrated into ISD. Reanalysis of existing data is also a priority but would require considerable resources to accomplish. We welcome readers' comments and input as this long-term effort continues to improve the availability of global climatological data for years to come.

DATA ACCESS AND PRODUCTS.

Additional detail regarding ISD data applications, usage, links to related products and services, and references, is available at www.ncdc.noaa.gov/oa/climate/isd/index.php.

Examples of products and services include the following:

a) ISD-Lite, with the goal of making ISD less complex for general research and scientific purposes, is a

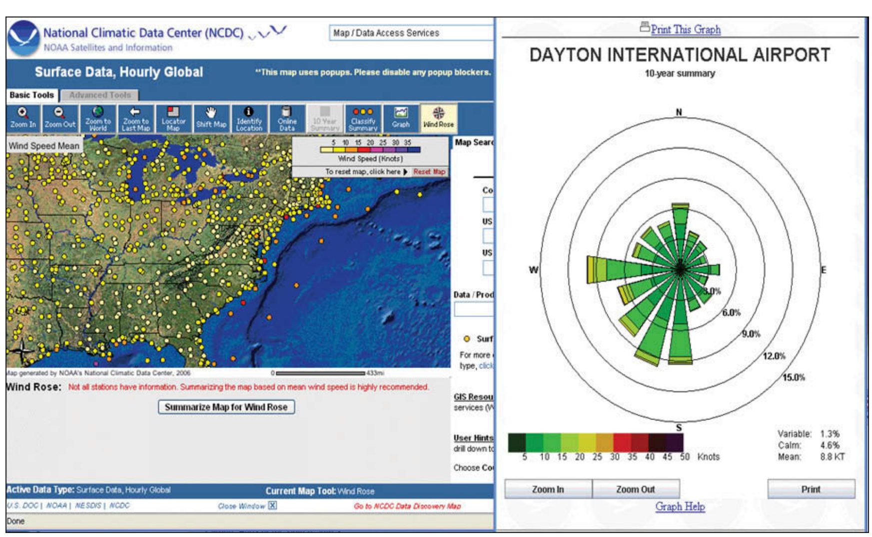

FIG. 3. Dynamic GIS maps and station-based wind rose diagrams are examples of available hourly and summary global data products.

subset of the full ISD containing 1 value per hour for the 8 most popular surface parameters. Data volume is approximately 10% of the full ISD data set. (See ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite.)

b) The Climate Data Online (CDO) Web system (http://cdo.ncdc.noaa.gov) provides ASCII text output and printable Web forms for numerous datasets, including ISD. GIS interface with ISD global map/numerous search parameters (Fig. 3): http://gis.ncdc.noaa.gov.
c) For U.S. stations—the Quality Controlled Local Climatological Data (LCD) product: http://cdo.ncdc.noaa.gov/qclcd/QCLCD?prior=N.
d) For global stations updated daily—Global Surface Summary of the Day (GSOD): http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html.
e) ISD summaries provide climatological summaries in tabular and graphical form, for various parameters such as temperature, dew point, wind speed/direction, cloud ceiling vs. visibility, sea level pressure, and various others: http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=SUMMARIES.

These products and services are also accessible via the NOAA Climate Services Portal at www.climate.gov.

ACKNOWLEDGMENTS. A large number of people have contributed to this overall effort since its inception in the late 1990s. In addition to the database development, this comprises product development, online services, and operational support for the data. The cast of contributors includes (in addition to the authors): NCDC employees (including federal and contract) Rich Baldwin, Dee Dee Anders, Tom Whitehurst, Pete Jones, Fred Smith, Richard Smith, Glen Reid, Brian May, Scott Chapal, Mark Lackey, Vickie Wright, Jon Burroughs, Xungang Yin, Imke Durre, Byron Gleason, David Wuertz, Jay Lawrimore, Steve Del Greco, Dan Dellinger, Ron Ray, Alan Hall, Mike Urzen, Tom Ross, Mark Seiderman, Vincent Stanton, Rod Truesdell; NCAR employees Steve Worley and Joey Comeaux; the Northeast Regional Climate Center's Arthur DeGaetano and Keith Eggleston; and numerous employees with the USAF 14th Weather Squadron (especially George Moody, Jon Whiteside, and Robert Davy) and U.S. Navy FNMOC (Brian Wallace and Joe Covert—both retired). Without their excellent efforts, this database and its online services would not exist.

FOR FURTHER READING

Brown, P. J., and A. T. DeGaetano, 2009: A method to detect inhomogeneities in historical dewpoint temperature series. *J. App. Meteor. Climatol.*, **48,** 2362–2376.

Camalier, L., W. Cox, and P. Dolwick, 2007: The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmos. Environ.*, **41,** 7127–7137.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137,** 1–28, doi:10.1002/qj.776.

Del Greco, S. A., N. Lott, R. Ray, D. Dellinger, P. Jones, and F. Smith, 2007: Surface data processing and integration at NOAA's National Climatic Data Center. *Preprints, 87th AMS Annual Meeting,* San Antonio, TX.

Lott, N., 2004: The quality control of the integrated surface hourly database. *Preprints, 84th AMS Annual Meeting,* Seattle, WA.

——, R. Baldwin, and P. Jones, 2001: The FCC Integrated Surface Hourly Database, A New Resource of Global Climate Data. NCDC Technical Report 2001-01. National Climatic Data Center, 42 pp.

——, S. A. Del Greco, T. Ross, and R. Vose, 2008: The integrated surface database: Partnerships and progress. *Preprints, 88th AMS Annual Meeting,* New Orleans, LA.

Phillips, C. S., 1985: An objective method for minimizing non-precipitation effects in precipitation data from punched paper tape. *Preprints, 65th AMS Annual Meeting,* Boston, MA.

Willett, K. M., N. P. Gillett, P. D. Jones, and P. W. Thorne, 2007: Attribution of observed surface humidity changes to human influence. *Nature,* **449,** 710–712, doi:10.1038/nature06207.

Zou, Bin, 2009: How should environmental exposure risk be assessed? A comparison of four methods for exposure assessment of air pollutions. *Environ. Monit. Assess.,* doi:10.1007/s10661–009-0992–8.