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 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).

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:

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

- 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 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.

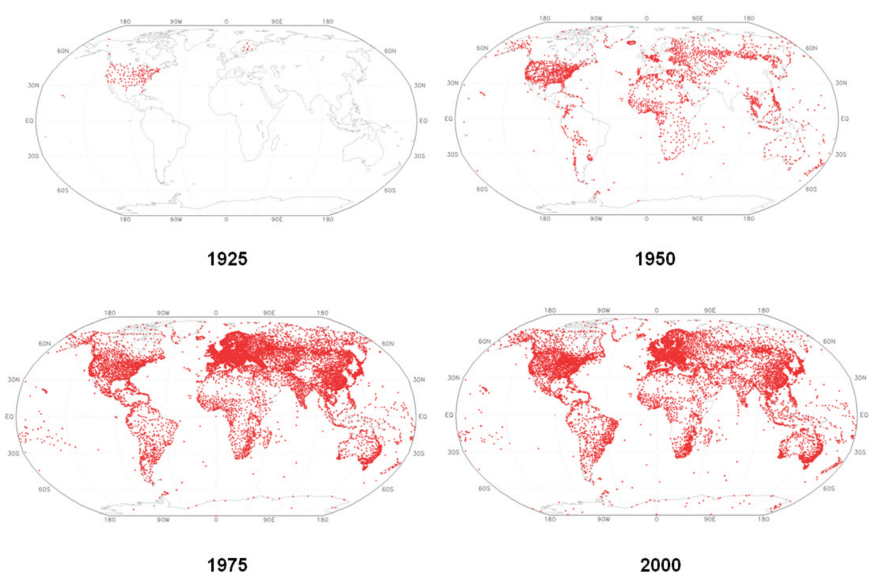

Fig. 1. The red dots represent the distribution of stations that contribute to the ISD data collection. Fewer stations were reporting in the early twentieth century. Since 1950, station coverage has been reasonable over North America, Europe, Australia, and parts of Asia, with noteworthy gaps in Africa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

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

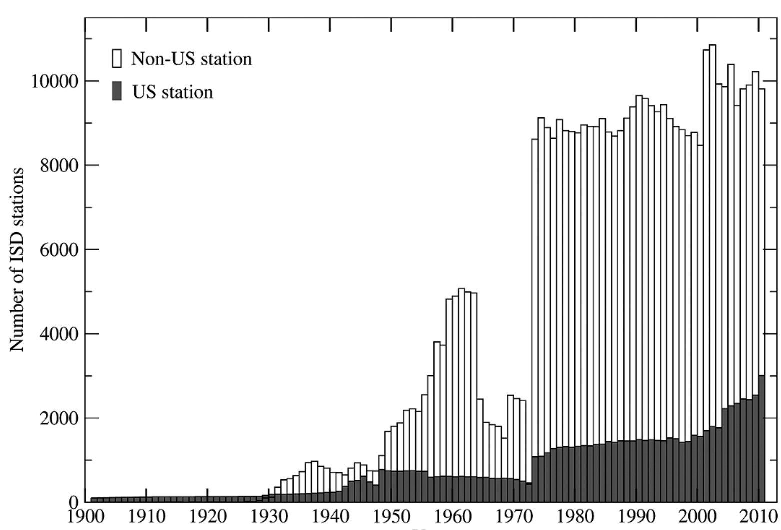

Fig. 2. There are more than 11,000 stations actively reporting hourly data and updated daily in the ISD. A notable decline occurs for both U.S. and non-U.S. ISD stations during the late 1960s–early 1970s due to the transition from keying of data to digital transmission/receipt of data causing interruption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

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.

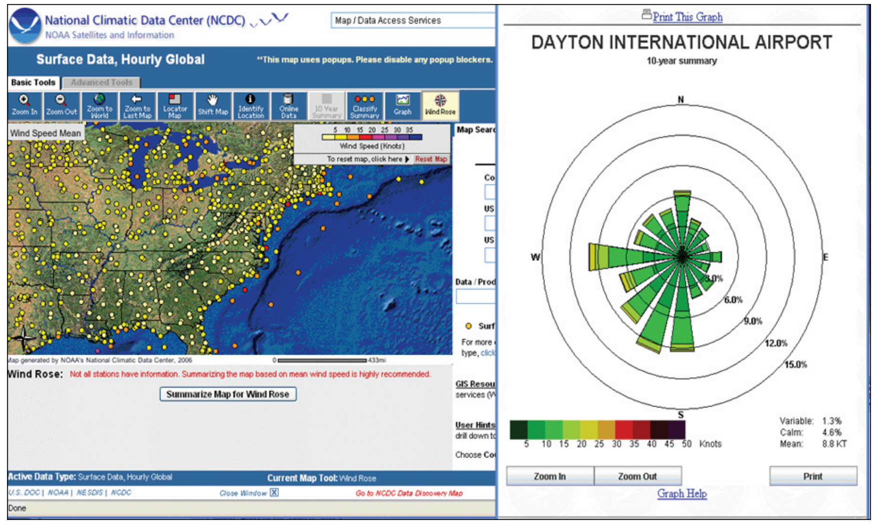

Fig. 3. Dynamic GIS maps and station-based wind rose diagrams are examples of available hourly and summary global data products.

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 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.