J2.1 SURFACE DATA INTEGRATION AT NOAA's NATIONAL CLIMATIC DATA CENTER: DATA FORMAT, PROCESSING, QC, AND PRODUCT GENERATION

Stephen A. Del Greco *, Neal Lott *, Kathy Hawkins, Rich Baldwin,
Dee Dee Anders, Ron Ray, Dan Dellinger
NOAA National Climatic Data Center, Asheville, North Carolina

Pete Jones, Fred Smith
TMC Technologies, Fairmont, West Virginia

ABSTRACT

The National Oceanic & Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) acquires, quality controls, archives and provides dissemination services for global meteorological and climatological data. These data are processed with automated Quality Assurance/Control (QA/QC) software, with some of the software being network-specific. For example, the NOAA Automated Surface Observing System (ASOS) Network, National Cooperative Observers Network (COOP), and Climate Reference Network (CRN) are processed using systems unique to their network. Additionally, some of the data undergo additional interactive QC, which involves "visual/manual" inspection of the data.

While similar, QA/QC rules and algorithms for like parameters from different observing networks are sometimes not standardized. NCDC's goal is to integrate surface data into a standard format, and process the data through standardized QA/QC algorithms and procedures. To that end, NCDC has developed a new integrated surface database, called Integrated Surface Data (ISD). To date, numerous historical datasets have been integrated into ISD, with others to follow. Also, NCDC is designing a new QA/QC system - Integrated Surface Data Processing System (ISDPS), as an end-to-end system for processing in-situ data, where QA/QC is standardized, network independent and based on reporting frequency (hourly, daily, etc.). Together, ISDPS and ISD integrate the QA/QC algorithms into a unified system, use ISD as the input/output format, and provide integrated online products to customers.

This paper briefly describes the ISDPS/ISD methodologies, data format, QC/validation techniques, and the various products available to customers.

1. INTRODUCTION

The development of ISD (previously called ISH—Integrated Surface Hourly) and ISDPS has been an iterative process. This includes the development of the integrated format, collection of datasets to include in the initial ISD database, development of a data model to use in a relational database for customer servicing, quality control of the historical ISD, development of the end-to-end ISDPS process, and development of online products.

2. HISTORICAL DATABASE AND BACKGROUND

The National Climatic Data Center (NCDC), in conjunction with Federal Climate Complex (FCC) partners (US Air Force and Navy), developed the global ISD database to address a pressing need for an integrated global database of surface climatological data. The database of approximately 20,000 stations includes data from as early as 1901 (many stations beginning in 1948-1973 timeframe), is operationally updated with the latest data, and is now being used by numerous customers in many varied applications. This effort was made possible by funding from the Environmental Services Data and Information Management (ESDIM) office, the Office of Global Programs (OGP), and extensive contributions from member agencies in the FCC. The development of ISD Version 1 and Version 2 is now complete:

1) ISD Version 1 -- the "database build" phase produced ISD Version 1 by integrating various data sources into one set of data. The new database collects NCDC U.S. hourly data, U.S. Navy hourly data, NCDC U.S. hourly precipitation data, and Air Force global hourly and synoptic data, into one global database. These data sources included over 100 original "tapedecks" (as they were called many years ago) and formats, each having already been quality-controlled to various degrees. The building of the database involved extensive research, data format conversions, time-of-observation conversions, and development of extensive metadata to drive the processing and merging. This included the complex handling of input data stored in three different station-numbering/ID systems.

2) ISD Version 2 -- two phases of quality control produced ISD Version 2. Phase one involved the correction of errors identified after the "database build" phase (e.g., due to input data file problems). Phase two involved the research, development, and programming of algorithms to correct random and systematic errors in the data, to improve the overall quality of the database;

* *Corresponding authors address:* Stephen Del Greco, Neal Lott, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; e-mail: Stephen.A.Delgreco@noaa.gov, Neal.Lott@noaa.gov.

and the data processing of the full period of record archive through these quality control algorithms. To briefly describe the two phases of quality control:

Phase 1 –
a) Validation of the data merging process through element value comparisons, such as temperature.
b) A complete inventory of all input and output data, to ensure no data loss.
c) Thorough checking of test data and archive data, to fully test the software before full database processing began.
d) Random checks of the final output database (ISD).

Phase 2 –
a) Extremes / validation checks, to ensure no obviously erroneous values are present in the data.
b) Temporal continuity checks, to look for "spikes" in continuous elements such as temperature and pressure.
c) Consistency checks of one element vs. another within a given data record/observation, such as temperature vs. present weather (e.g., no snow at 10 degrees C).

3. DATA ACCESS AND PRODUCTS

ISD is fully accessible online through:
a) the NOAA National Data Center (NNDC) Climate Data Online (CDO) system: http://cdo.ncdc.noaa.gov (figure 1);
b) NCDC's Geographic Information System (GIS) interface (figure 2): click on "*search by map*" on NCDC's homepage: http://www.ncdc.noaa.gov; and
c) for U.S. stations, through our Local Climatological Data (LCD) system (figure 5): http://cdo.ncdc.noaa.gov/ulcd/ULCD

3.1 *CDO System and GIS Interface*

CDO and the GIS interface (a map-based interface for data in CDO) provide users with the ability to easily select data by country, state, station, and time series; for output into an ASCII text file or a web form. Output formats include a simplified space-delimited format with the key climatic elements, and a more advanced format which allows the user to select just the elements desired—e.g., temperature and dew point.

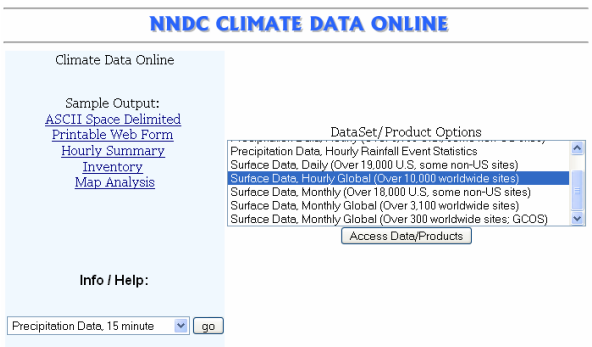

Figure 1. CDO Interface.

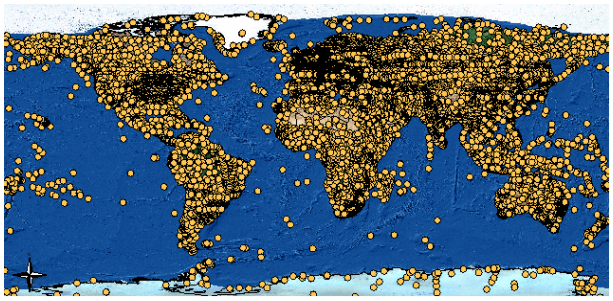

Figure 2. GIS Interface Sample Map (ISD global data).

As a new feature of the GIS interface, selected data types can be graphed for user-selected stations and periods of interest. This capability will continue to be expanded for additional data types, etc.

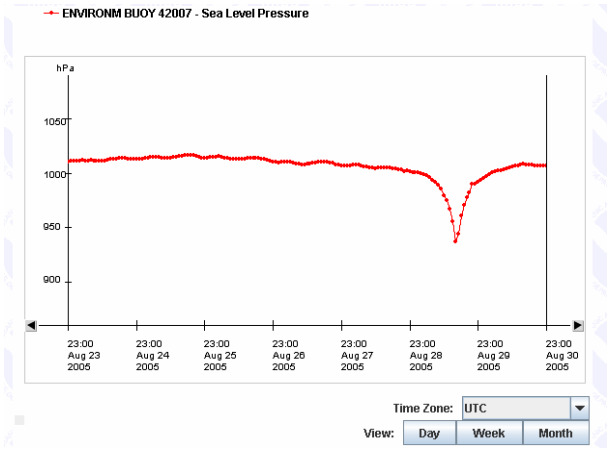

Figure 3. Graph of Sea Level Pressure During Passage of Hurricane Katrina.

Also, an ISD summary system (figure 4) is now online within the CDO system, and through our GIS interface. Fourteen different summaries can be generated, such as:

- ceiling-visibility,
- dew point statistics,
- temperature statistics,
- ceiling-visibility-wind conditions (different categories for flight criteria),
- relative humidity,
- sky cover,
- sea level and station pressure, and
- wind speed/direction.

The summaries are available as "pre-generated" 5 and 10-year summaries, or as "on-the-fly" summaries for user-selected periods.

NNDC CLIMATE DATA ONLINE

Surface Data, Hourly Global: Summaries(DS3505)

ISH Summary

POR 01/01/1995 - 12/31/2004

Temperature Summary for 87750099999/BAHIA BLANCA AERO

1995/01/01 00:00 to 2004/12/31 23:59

HOUR (UTC)		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
0	mean	23.6	21.8	19.7	14.6	11.5	8.4	7.7	9.5	11.3	14.8	18.0	21.3	15.2
	stdv	4.0	4.3	4.2	3.6	3.7	3.8	3.9	4.1	4.0	4.2	4.4	4.8	4.1
	#obs	294.0	268.0	292.0	276.0	295.0	286.0	306.0	299.0	287.0	301.0	285.0	259.0	3448.0
1	mean	22.1	20.5	18.6	13.8	10.9	7.7	7.1	8.7	10.4	13.8	16.7	20.1	14.2
	stdv	4.0	4.2	4.1	3.7	3.8	4.0	4.0	4.1	4.1	4.2	4.1	4.5	4.1
	#obs	274.0	258.0	284.0	267.0	285.0	264.0	286.0	289.0	279.0	268.0	265.0	252.0	3271.0
2	mean	21.0	19.5	17.8	13.2	10.5	7.3	6.6	8.1	9.7	13.2	15.9	18.9	13.5
	stdv	3.8	4.1	4.1	3.8	3.9	4.1	4.1	4.3	4.1	4.3	4.1	4.3	4.1
	#obs	273.0	260.0	285.0	269.0	278.0	277.0	287.0	284.0	272.0	272.0	269.0	254.0	3280.0
3	mean	20.3	18.8	17.1	12.7	10.0	7.0	6.4	7.8	9.3	12.5	15.3	18.2	13.0
	stdv	3.8	4.0	4.2	3.9	4.0	4.3	4.2	4.4	4.2	4.1	4.2	4.4	4.1
	#obs	299.0	274.0	304.0	292.0	306.0	291.0	306.0	300.0	289.0	298.0	287.0	274.0	3520.0
4	mean	19.7	18.2	16.7	12.4	9.7	6.8	6.2	7.5	8.9	12.1	14.8	17.3	12.5
	stdv	3.8	4.1	4.3	4.0	4.2	4.3	4.5	4.5	4.3	4.0	4.2	4.3	4.2
	#obs	278.0	253.0	278.0	271.0	277.0	268.0	280.0	280.0	257.0	255.0	248.0	247.0	3192.0

Figure 4. ISD Summary Sample.

3.2 *LCD System*

The LCD system provides users with a standard LCD form (figure 5) for U.S. stations, displaying hourly, daily, and monthly surface data parameters. The LCD product has been in existence in some form since 1949 and is very popular with many customers. The new system provides access to over 1500 stations' data after completion of all automated quality control, with a one-day lag time (e.g., January 15 data available on January 16). Approximately 270 "first-order" National Weather Service stations' data undergo additional quality control, and these data are then "fed" into the LCD system as the final data for those stations. In this way, users can be assured that the LCD forms they receive through this system are always the "latest and greatest" available at that time.

4. STATUS AND TARGETS -- ISDPS PROCESS

NOAA's ASOS network data are now being operationally ingested into ISD format and processed through ISDPS on a daily basis. The LCD products mentioned above are by-products of these processed data. The Climate Reference Network data are being converted to ISD, and plans are to process these data using ISDPS in 2006.

Projections for the Cooperative Observer Network are to convert to ISD in FY06 and to start processing data through ISDPS in FY07. Also, NCDC will pursue other in situ data sources (e.g., U.S. mesonets, other global/foreign datasets) for inclusion in the ISD/ISDPS process and historical database.

5. BENEFITS OF INTEGRATION

In conclusion, the bullets below list some of the benefits obtained by integrating NOAA's in situ datasets:

- Reduction of subjectivity and inconsistencies among data sets that span multiple observing networks and platforms
- Standardized QA/QC based on reporting time resolution (e.g., QC methodology for hourly temperature data independent of network)
- Standardized products are more easily developed and consistent
- Collective experience and expertise leads to a better product
- Software is modular for ease of modification
- Conformance of data to documentation, in one format for all users.

U.S. Department of Commerce
National Oceanic & Atmospheric Administration

PRELIMINARY LOCAL
CLIMATOLOGICAL DATA
HOURLY OBSERVATIONS TABLE
W B HARTSFLD/ATLA INTL APT (ATL)
ATLANTA, GA
(08/2005)

National Climatic Data Center
Federal Building
151 Patton Avenue
Asheville, North Carolina 28801

Elevation: 971 ft. above sea level
Latitude: 33.640
Longitude: -84.427
Data Version:
This form contains Quality Controlled data.

Date	Time	Station Type	Sky Conditions	Visibility	Weather Type	Dry Bulb Temp (F)	Dry Bulb Temp (C)	Wet Bulb Temp (F)	Wet Bulb Temp (C)	Dew Point Temp (F)	Dew Point Temp (C)	Rel Humd %	Wind Speed (MPH)	Wind Dir	Wind Gusts (MPH)	Station Pressure	Press Tend	Sea Level Pressure	Report Type	Precip. Total
01	0011	11	SCT008 BKN039 OVC050	10.00		73	23.0	72	22.4	72	22.0	97	7	120		29.02		M	SP	
01	0051	11	FEW010 OVC050	10.00		73	22.8	71	21.6	70	21.1	90	5	140		29.02	0	30.09	AA	
01	0151	11	FEW010 OVC080	10.00		73	22.8	70	21.3	69	20.6	87	9	110		29.01		30.08	AA	
01	0211	11	SCT009 BKN080 OVC090	10.00		73	23.0	71	21.6	70	21.0	90	9	080		29.01		M	SP	
01	0219	11	BKN009 BKN065 OVC090	10.00		73	23.0	71	21.6	70	21.0	90	10	070		29.00		M	SP	
01	0251	11	OVC009	10.00		73	22.8	71	21.6	70	21.1	90	9	080		29.01		30.07	AA	
01	0301	11	FEW007 OVC013	10.00		73	23.0	71	21.6	70	21.0	90	10	070		29.01		M	SP	
01	0329	11	FEW010 OVC016	10.00		73	23.0	71	21.6	70	21.0	90	7	090		29.00		M	SP	
01	0351	11	BKN020 OVC110	9.00		73	22.8	70	21.3	69	20.6	87	8	090		29.00	6	30.07	AA	

Figure 5. LCD Product Sample.

6. REFERENCES

Lott, Neal. Data Documentation for Federal Climate Complex Integrated Surface Data. [Asheville, N.C.]: National Climatic Data Center, 2005.

Lott, Neal. The Quality Control of the Integrated Surface Hourly Database. 84th American Meteorological Society Annual Meeting, 2004, Seattle, WA.

Lott, Neal, Rich Baldwin, and Pete Jones. NCDC Technical Report 2001-01, The FCC Integrated Surface Hourly Database, A New Resource of Global Climate Data. [Asheville, N.C.]: National Climatic Data Center, 2001.

Lott, Neal and Dee Dee Anders. NNDC Climate Data Online for Use in Research, Applications, and Education. Twelfth Conference on Applied Climatology. Pages 36-39. American Meteorological Society, May 8-11, 2000, Asheville, NC.