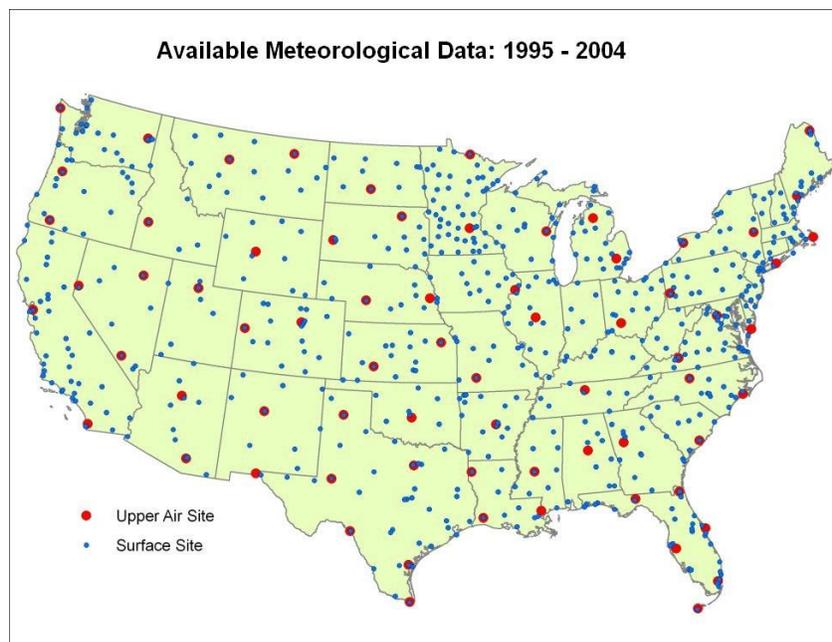


Details of the Omnibus Meteorological Database (OMD or “MetDat”)

Data Record: 1/1/1995 through 06/30/2006. The original omnibus meteorological database (OMD) was processed for the period 1995 to 2004. Subsequently, the data have been updated through 2005 and the first half of 2006. Additional months/years can be added as needed.

Data Source: The surface data were extracted from the Integrated Surface Hourly (ISH)¹ database maintained by NCDC. The upper air data was taken from the Integrated Global Radiosonde Archive (IGRA)² dataset, which is also maintained by NCDC.

Spatial/Temporal: 692 surface sites and 85 upper air (UA) sites^{3, 4} were used in creation of the OMD. There are two separate files within the OMD: one contains daily records, the other contains hourly data. Both the daily and hourly files combine the surface and UA data by pairing each surface site with its nearest UA site neighbor. This pairing was only done for surface sites within three degrees of its nearest UA neighbor. If this criteria was not met, then all UA data for that site are labeled as missing. Both the daily and hourly files have time-shifted the raw UTC data to the local standard time appropriate for the individual site.



¹ The ISH database (TD 3505) is a quality-assured global database of hourly meteorological observations. It is an aggregation of the following pre-existing datasets: all of the NCDC and Navy surface hourly data (TD3280), the NCDC hourly precipitation data (TD3240), and the Air Force Datsav3 surface hourly data (TD9956).

² The Integrated Global Radiosonde Archive (IGRA) consists of radiosonde observations at over 1500 globally distributed stations. Observations are available for standard, surface, and significant levels. Individual station records are updated daily and are freely available online at <ftp://ftp.ncdc.noaa.gov/pub/data/igra/>

³ The OMD uses all U.S. sites that had 75% or more of the expected observations from 1995-2004.

⁴ A list of the sites is available at G:/user/share/omd/Met-Database_tables.xls. The “site” tab of this spreadsheet also contains the State, latitude / longitude, elevations, UTC offset, USAFID number, and WBAN number.

Meteorological Parameters (Daily data): There are 60 meteorological parameters on the daily OMD file as shown in the table below. This table also shows the units, time duration, and variable name associated with each parameter. The full SAS dataset is approximately 250 Mbytes for each full year within the OMD.

PARAMETER_ID	NAME	UNIT_ID	DURATION_ID	PARM
1	Maximum surface temperature (°K)	C	24-hr calc	TMAX
2	Morning (7-10 a.m. LST) average temperature (°K)	C	4-hr calc	TAVGAM
3	Afternoon (1-4 p.m. LST) average temperature (°K)	C	4-hr calc	TAVGPM
4	Diurnal temperature change (°K)	C	24-hr calc	TDELTA
5	Average daily wind U vector (m/s)	m/s	24-hr calc	UAVG
6	Average daily wind V vector (m/s)	m/s	24-hr calc	VAVG
7	Average daily wind speed (m/s)	m/s	24-hr calc	WSAVG
8	Average daily wind direction (deg)	deg	24-hr calc	WDAVG
9	Average morning (7-10 a.m. LST) wind U vector (m/s)	m/s	4-hr calc	UAVGAM
10	Average morning (7-10 a.m. LST) wind V vector (m/s)	m/s	4-hr calc	VAVGAM
11	Average morning (7-10 a.m. LST) wind speed (m/s)	m/s	4-hr calc	WSAVGAM
12	Average morning (7-10 a.m. LST) wind direction (deg)	deg	4-hr calc	WDAVGAM
13	Average afternoon (1-4 p.m. LST) wind U vector (m/s)	m/s	4-hr calc	UAVGPM
14	Average afternoon (1-4 p.m. LST) wind V vector (m/s)	m/s	4-hr calc	VAVGPM
15	Average afternoon (1-4 p.m. LST) wind speed (m/s)	m/s	4-hr calc	WSAVGPM
16	Average afternoon (1-4 p.m. LST) wind direction (deg)	deg	4-hr calc	WDAVGPM
17	Average daily relative humidity (%)	%	24-hr calc	RHAVG
18	Midday (10 a.m.-4 p.m. LST) average relative humidity (%)	%	7-hr calc	RHAVGMID
19	Nighttime (8 p.m.-7 a.m. LST) average relative humidity (%)	%	12-hr calc	RHAVGNIGHT
20	Average dew point temperature (°K)	C	24-hr calc	DPAVG
21	Maximum dew point temperature (°K)	C	24-hr calc	DPMAX
22	Maximum water vapor mixing ratio (g/kg)	g/kg	24-hr calc	MRMAX
23	Average Sea level pressure (mb)	mb	24-hr calc	SLPAVG
24	Average Station pressure (mb)	mb	24-hr calc	STPAVG
25	Minimum apparent temperature (°K)	C	24-hr calc	TAMIN
26	Average apparent temperature (°K)	C	24-hr calc	TAAVG
27	Maximum apparent temperature (°K)	C	24-hr calc	TAMAX
28	Morning (7-10 a.m. LST) average fractional cloud cover (tenths)	tenths	4-hr calc	CCAVTAM
29	Afternoon (1-4 p.m. LST) average fractional cloud cover (tenths)	tenths	4-hr calc	CCAVGPM
30	Total precipitation (inches)	in	24-hr calc	PRECIP
31	Occurrence of rain (hours)	hrs	24-hr calc	RAINHRS
32	Occurrence of haze (hours)	hrs	24-hr calc	HAZEHRS
33	Occurrence of fog (hours)	hrs	24-hr calc	FOGHRS
34	Morning (~1200 UTC) Height of 500 mb surface (m)	m	1-hr obs	H500
35	Morning (~1200 UTC) Height of 700 mb surface (m)	m	1-hr obs	H700
36	Morning (~1200 UTC) Height of 850 mb surface (m)	m	1-hr obs	H850
37	Morning (~1200 UTC) Temperature at 925 mb	C	1-hr obs	T925
38	Morning (~1200 UTC) Temperature at 850 mb	C	1-hr obs	T850
39	Morning (~1200 UTC) Temperature at 700 mb	C	1-hr obs	T700
40	Morning (~1200 UTC) Temperature at 500 mb	C	1-hr obs	T500
41	Morning (~1200 UTC) Dewpoint Temperature at 850 mb	C	1-hr obs	TD850
42	Morning (~1200 UTC) 925 mb-surface temperature difference (°K)	C	1-hr calc	DT925
43	Morning (~1200 UTC) 850 mb-surface temperature difference (°K)	C	1-hr calc	DT850
44	Morning (~1200 UTC) 700 mb-surface temperature difference (°K)	C	1-hr calc	DT700
45	Maximum rate of mixing height increase (m/hr) 6am-12pm LST	m/hr	7-hr calc	MRMH
46	Maximum mixing height (m) 4am-4pm LST	m	13-hr calc	MHMAX
47	24-hr Transport Direction	deg	24-hr calc	TRANDIR
48	24-hr Transport Distance	km	24-hr calc	TRANDIS
49	X-component of the 24-hr Transport Vector	km	24-hr calc	TRANU
50	Y-component of the 24-hr Transport Vector	km	24-hr calc	TRANV
51	Z-component of the 24-hr Transport Vector	km	24-hr calc	TRANW
52	24-hr Scalar Wind Run	km	24-hr calc	WNRUN
53	24-hr Change in Morning (~1200 UTC) Temperature at 850 mb	C	1-hr calc	D24T850
54	24-hr Change in Morning (~1200 UTC) Dewpoint Temperature at 850 mb	C	1-hr calc	D24TD850
55	Deviation in Morning (~1200 UTC) Temperature (°K) at 850 mb from the 10-year monthly mean	C	1-hr calc	DEVT850
56	Deviation in Morning (~1200 UTC) Temperature (°K) at 700 mb from the 10-year monthly mean	C	1-hr calc	DEVT700
57	Deviation in Morning (~1200 UTC) Temperature (°K) at 500 mb from the 10-year monthly mean	C	1-hr calc	DEVT500
58	Deviation in Morning (~1200 UTC) Height of 850 mb surface (m) from the 10-year monthly mean	m	1-hr calc	DEVH850
59	Deviation in Morning (~1200 UTC) Height of 700 mb surface (m) from the 10-year monthly mean	m	1-hr calc	DEVH700
60	Deviation in Morning (~1200 UTC) Height of 500 mb surface (m) from the 10-year monthly mean	m	1-hr calc	DEVH500

Some of the parameters shown in the table above are extracted directly from the IGRA data (e.g., morning 500-mb temperatures). A larger segment of the daily parameters were calculated based on hourly data from the ISH (e.g., average 7-10a wind direction)⁵. The OMD also includes some “second-order” variables which are generated by processing the raw

⁵ All time-averaged data are hours-inclusive, i.e., 7-10a wind direction averages 4 hours worth of observations.

ISH/IGRA data with additional meteorological data outside the original databases. Each of the OMD parameters that are non-intuitive are explained in the next section.

TDELTA – This is the difference between daily maximum and daily minimum temperatures.

TAMIN, TAAVG, TAMAX: Each of these variations of apparent temperature used this equation⁶.

MRMAX, MRMHI: Hourly estimations of mixing height were calculated using the Holzworth method. This involves assuming air parcels originating at the surface will mix dry-adiabatically until no longer buoyant (warmer than surrounding environment). The mixing height is determined by the level at which the hourly surface temperatures (cooled with height according to the dry-adiabatic lapse rate) equal the temperature profile from the nearest UA site. This approach is a highly uncertain estimate of mixing height and the values of MRMAX and MRMHI should be used with discretion.

TRANDIR, TRANDIS, TRANU, TRANV, TRANW, WNDRUN: For each day, the HYSPLIT model is run to calculate 24-hr backward trajectories from each surface site. The trajectories will start at noon LST at a height of 300m (i.e., within the mixed layer). Generally, EDAS meteorological data are used to derive the trajectories, but for some older years NCEP Reanalysis data was used. The HYSPLIT output provides the location of an inert particle 24 hours prior based on the wind fields, and from this value a daily transport vector (u, v, w) is calculated. The model also outputs a scalar wind run distance along the trajectory.

D24T850, D24TD850: These values compare daily 1200 UTC temperatures and dew point temperatures against the same values from the day before.

DEVT_{xxx}, DEVH_{xxx}⁷: These parameters compare daily 1200 UTC temperatures and geopotential heights against the monthly mean temperature/heights for that particular site, where the monthly means are averaged over 10 years.

Each daily parameter value has a quality control code associated with its value. Missing data, or values that fail certain completeness or spatial representative checks are flagged accordingly. In general, the following parameters have the highest frequency of flagged data. Care should be exercised when using these parameters.

Occasional: AVAGAM, VAVGPM, WDAVG, RHAvg, RHAvgMID, CCAVGPM, and CCAVTAM

Moderate: WDAVGPM, VAVGPM, UAVGAM, WDAVGAM, D24T850, D24TD850, DEVT850, DEVT700, DEVT500, DEVH850, DEVH700, and DEVH500

⁶ $T_a = -1.3 + 0.92T + 2.2e$, where T_a is apparent temperature (C), T is ambient air temperature (C), and e is the RH-dependent water vapor pressure (kPa). See <http://www.ncdc.noaa.gov/oa/climate/research/heatstress> for more detail.

⁷ Calculated at 850, 700, and 500 mb.

Frequent: PRECIP, H500, H700, H850, T925, T850, T700, T500, TD850, DT925, DT850, DT700, MRMHI, and MHMAX

Meteorological Parameters (Hourly data): There are 18 meteorological parameters on the daily OMD file as shown in the table below. This table also shows the units, time duration, and variable name associated with each parameter. The full hourly SAS dataset is approximately xxx Gbytes for each full year within the OMD.

PARAMETER_ID	NAME	UNIT_ID	DURATION_ID	PARM
1	Temperature	C	1-hr obs	T
2	Dew Point	C	1-hr obs	DP
3	Wind Speed	m/s	1-hr obs	WS
4	Wind Direction	deg	1-hr obs	WD
5	Sea Level Pressure	mb	1-hr obs	SLP
6	Station Pressure	mb	1-hr obs	STP
7	Altimeter Setting	mb	1-hr obs	ALT
8	Cloud Cover	octas	1-hr obs	CC
9	1-hr Precipitation	in	1-hr obs	PCP01
10	3-hr Precipitation	in	3-hr obs	PCP03
11	6-hr Precipitation	in	6-hr obs	PCP06
12	12-hr Precipitation	in	12-hr obs	PCP12
13	24-hr Precipitation	in	24-hr obs	PCP24
14	Occurance of Liquid Precip	0=no, 1=yes	1-hr obs	RAIN
15	Occurance of Frozen Precip except Hail	0=no, 1=yes	1-hr obs	SNOW
16	Occurance of Haze or Mist	0=no, 1=yes	1-hr obs	HAZE
17	Occurance of Fog	0=no, 1=yes	1-hr obs	FOG
18	Occurance of Hail	0=no, 1=yes	1-hr obs	HAIL

The original time stamp for the hourly data was UTC. Separate post-processing must be done to convert these data to local standard time (see last page). A list of the UTC to LST adjustments on a site by site basis is available on the “met_database” spreadsheet.

Development of Additional Years of Data⁸:

There are seven basic steps to adding additional meteorological time periods to the OMD. Each of the steps is explained below. The scripts and executables to run these steps currently reside on the Linux machine “bill” in the directory /data/OMD.

Step 1a: Acquiring the ISH data⁹. The data are available at the following location: <ftp.ncdc.noaa.gov>. According to the site, this data is only available internally within NOAA, but it appears to be accessible via anonymous login. Change directory (“cd”) to /pub/data/noaa. Then “cd” to the appropriate year. All ISH data for each of the worldwide sites for that particular year is available. The file name format for the files in this directory is: *usaf-wban-year.gz*, where *usaf* is the 6-digit USAF #, *wban* is the 5-digit WBAN #, and *year* is the year listed as a 4-digit number. (For example, the 2006 Whidbey Island, WA data is in the file name 690230-24255-2006.gz.) Unfortunately, there is not a way to only download the 692 sites of

⁸ Most of the code and scripts for processing additional years of data were provided by STI.

⁹ The 1995-2004 data were extracted from CDs and DVDs provided to EPA by NCDC.

interest for the OMD, so one must get more sites than is necessary. Using wildcards in the ftp process and accessing the needed files by their first two digits, one can grab the OMD station data by the use of the following command(s) (make sure you have prompt turned off and are doing the ftp transfer in non-passive binary mode): ‘mget xx*’, where xx equals 69, 70, 72, 74, 78, and 91. Once the files have been downloaded to the local Linux computer (e.g., bill), then unzip all the individual data files with the command, ‘gunzip *’. The ISH data are variable-width ASCII files at this point. Typically, the ftp site seems to have the latest data to within approximately one week.

Step 1b: Acquiring the IGRA data. The data are available from the same ftp site (<ftp.ncdc.noaa.gov>), but in the directory: /pub/data/igra/data-xxx, where xxx equals “y2d” for data since 2004 and xxx equals “por” for earlier data. There is one file per station. All IGRA data for each of the worldwide sites for that particular year is available. The file name format for the files in this directory is: *wmo-y2d.Z*, where *wmo* is the 5-digit WMO #. Using wildcards in the ftp process and accessing the needed files by their first two digits, one can grab the OMD station data by the use of the following command: ‘mget 70* 72* 74* 91*’. Once the files have been downloaded to the local Linux computer (e.g., bill), then uncompress all the individual data files with the command, ‘uncompress *.Z’. The IGRA files are in fixed-width ascii format. Typically, the data are up-to-date to about 48 hours from present-time.

Step 2: Extracting the ISH data. The source code for this processing step is called “extract_ish”. This code¹⁰ does the following tasks:

- a) reads in the ISH data record,
- b) determines date, time, station location, and observation type,
- c) reads the mandatory and additional data sections from the variable-length ISH data,
- d) makes any necessary unit conversions, and
- e) writes the output to a fixed-width hourly text file.

Prior to running extract_ish¹¹, one must create a list of the stations from which data is to be expected. For the OMD, the file has already been created and is located in the extract_ish script directory (“OMD_station_list”). The run script and code/executable directory are listed below. As with all scripts, it is suggested that the screen output be captured into a .log file (e.g., ‘extract_ish.csh >&! 2006_extract_ish.log’.) The output data from this step are individual hourly files (in UTC), one per station in a flat fixed-width format with each line representing an hour’s worth of data.

Step 3: Processing the surface data: The source code and script for this step is called “process_sfc”. This code requires the input of the OMD site information (e.g., UTC offset) and conducts the following tasks:

- a) reads in the hourly data from extract_ish,

¹⁰ All code necessary to generate OMD data are currently on “bill” in the directory: /data/OMD/src.

¹¹ All scripts needed to generate OMD data are currently on “bill” in the directory: /data/OMD/scripts.

- b) calculates the derived hourly meteorological variables (e.g., u, v, RH, etc.),
- c) converts the UTC data to local standard time¹²,
- d) calculates daily output values (e.g., daily maximum temperature), and
- e) outputs two separate sets of output files: “dbdaily” and “dbhourly”.

Two ascii files are created for each OMD station, one hourly and one daily. The output format is fixed-width ascii with one parameter per line, per day or hour (depending on the file). Consistent values for missing data are provided at this stage of the processing and appropriate QC codes are added for missing data or calculations with insufficient data for daily averaging.

Step 4: Processing the upper air data¹³: The source code and script for this step is called “process_ua”. This code is broken up into four separate programs which must be run sequentially as each step builds upon the next.

a) process_ua1¹⁴: This step reads the raw IGRA data for each of the 85 upper air sites in the OMD and outputs a single binary file for each location.

b) process_ua2: This step creates upper air parameter files for each of the OMD surface locations. Data from the closest UA site to the surface location are used. This stage of the processing is also where mixing heights and rate of rise are estimated. If the UA to surface distance is greater than 3 degrees, then the data are QC flagged as “not spatially representative”. If the distance is between 1 and 3 degrees, then the data are labeled “suspect spatial representativeness”. There is also a representativeness check on the height of the surface and UA observations. If the values are within 100 m of each other the data is considered valid. 100-500m separation indicates “suspect” data. Greater than 500m separation leads to a flag of “not spatially representative”.

c) process_ua3: This program calculates the ten-year monthly means of temperature and geopotential height (at heights of 850, 700, and 500 mb) that daily UA values are later compared against. The OMD uses the years from 1995 to 2004 as the time period for the ten year mean, thus this code only needs to be run once. This code generates 85 individual ascii files for each UA location.

d) process_ua4: The last stage of the UA processing calculates the “second order” parameters such as deviation from monthly means and changes from yesterday’s values.

Step 5: Running HYSPLIT and calculating trajectory data: This step involves the configuration, modeling, and post-processing of HYSPLIT trajectory model data. A single script

¹² This conversion takes place for the daily output file, but not for the hourly data. See footnote 8.

¹³ Note that only 1200 UTC data is extracted into the OMD. We generally expect the upper air data at this time period to be generally representative of the entire day.

¹⁴ Depending on whether one is processing .y2d data or older data, one may need to recompile process_ua1.f as the extension is hardwired within that code. At present, I have it set up to handle .y2d files.

in the “hysplit” directory is used to conduct each of these individual steps (hystraj, hymodelt, and hysplitpost). The trajectory model is driven by the EDAS 40km meteorological dataset (80 km EDAS if years earlier than 2004 are being added). These input data are available at: <http://www.arl.noaa.gov/pub/archives/edas40>. Trajectories are generated for all OMD sites in the lower 48 States, as listed in the `tzx.csv`¹⁵ files. The trajectories are 24-hour backward trajectories starting at noon (1200 LST). The trajectories have a release height of 300m and use the isobaric vertical velocity option within HYSPLIT. All of these options can be modified by either changing the entries within the “*hysplit.csh*” script. At the end of this processing stage there are four files per day (one per time zone) containing the output trajectory information (six trajectory parameters) in ascii format. As a last step, one should concatenate all of these individual files (in the “derived” directory) into one large ascii file (e.g., “`cat hysplit_2* > hysplit.dat`”).

Step 6a: Combine and reformat daily files: This script “`daily_combine.csh`” takes the following four sets of output data and merges them into a single fixed-width ascii file with each line containing data for a date/site/parameter. The four sets of files are:

- a) the daily surface output,
- b) the ua2 output,
- c) the ua4 output, and
- d) the combined hysplit file.

The resultant combined file is processed via `reformat_daily.csh` into two separate files (`daily1` and `daily2`). “`Daily1`” has the same format as the combined file (e.g., one date/site/parameter per line), but in ascending order by site, then date, then parameter. “`Daily2`” is the more useful file in that it contains a single record of data (all parameters) for each site/date.

Step 6b: Combine and reformat hourly files: There is code and scripts called “`reformat_hourly.csh`” which would takes existing intermediate output data and merges them into a single fixed-width ascii file with each line containing data for a date/site/parameter, as done in the daily data. Currently, it is not working because of memory issues. Instead, this step is now done within SAS (see step 7).

Step 7: Create SAS datasets: There are multiple SAS codes that convert the ascii data files into SAS format datasets¹⁶:

- `daily1.sas` – creates SAS dataset from “`daily1`” file
- `daily2.sas` – creates SAS dataset from “`daily2`” file
- `rawhourly.sas` – creates “`hourly1`” SAS dataset from combined “`dbhourly`” file
- `wideout4.sas` – creates “`hourly2`” SAS dataset from “`hourly1`” SAS file, also converts time zone of data from UTC to LST.

¹⁵ Each time zone must be considered separately within HYSPLIT so that the trajectory start time is correct.

¹⁶ Currently, these SAS scripts are on bill in `/data/OMD/sascode`.