NEAR REAL-TIME WEATHER DATA ARCHIVE

Nicholas Long
National Renewable Energy Laboratory, Golden, CO

ABSTRACT

The National Weather Service (NWS) collects weather information from 4,000 stations around the world. The National Renewable Energy Laboratory (NREL) collectes, parses, and stores this weather data in a local database to give researchers easy access to the data. NREL's parsing program simplies the data format for users. The data are available to anyone via the Internet and e-mail.

The purpose of this paper is to describe the availability, operation, and limitations of the available data. Currently, the database only contains temperature, wind, and cloud cover information for the locations. The data are available in Comma Separated Values (CSV) and International Weather for Energy Calculations (IWEC); however, the data may not be continuous.

INTRODUCTION

Weather information is collected from NWS, parsed, and stored into a local database at NREL.

The accuracy and continuity of the data is dependent on many procedures operating seamlessly. The majority of data collected are the dry- and wet-bulb temperatures, barometric pressure, and wind conditions. In some cases, more information is available and is included in the database. This paper will document how the data are parsed, how the data are filled, the limitations of the data, and how to access the database.

DATA SOURCE

The NWS reports the data in the Meteorological Aviation Report (METAR) format. These data are taken by the NWS once an hour during different times within the hour for each station. The observations that are taken within the hour are considered to be of the same cycle.

PARSING DATA

The METAR format is extremely inconvenient to read especially for trending historical data; therefore, a parsing program was created to simplify the data.

The weather data is parsed using a local program that allows for the gigabytes of data to be processed into an accessible indexed database format. Although there are many consistent errors with the format of the weather data files, the data parser finds and compensates for them.

Many stations have multiple readings around the hour. The program finds the closest entry to the hour and uses those values. However, if new information (such as cloud cover) is received in a reading and the data point is not already stored, then the new data point is stored as if it occurred for that hour. The overwriting of old data will not occur unless the new time stamp is closer to the hour of interest or the value was not previously stored.

FILLING MISSING DATA

The weather database does not have a contiguous collection of data. There are basically two different failure areas; the first is on the local server side where the connection to the Internet was interrupted either by a network failure or a computer failure that prevents the weather data from begin downloaded. The other failure area is on the NWS side, where the station was not accessed to collect the data. Also, heavy traffic to the NWS ftp site has prevented several data sets from being retrieved.

If missing data are present, the data can be exported with interpolated values for the holes. The only data that are filled are the wet- and dry- bulb temperatures. The filled values are not stored in the database; they are calculated every time a data request is received.

If the user requests filled data, there are two different routines that operate to smooth the data, while a third routine simply breaks the file and skips the missing data. The first method of filling occurs when the missing period is less then 6 hours, in which case the data are simply filled linearly. Second, if there are more than 6 hours and less then 48 hours of missing data, then the data are filled by taking the trend of the previous day that is valid. The trend is offset by the values that surround the missing data to smooth the filled data. Finally, if the missing data covers more than 48 hours, then the program creates a new file where the data starts again.

The filling data routine does not scan for values outside of a given range, for instance, if the temperature suddenly dropped 50ºC, then there must have been an

SimBuild 2004, IBPSA-USA National Conference Boulder, CO, August 4-6, 2004.

1

error with the reading of the station or parsing of the data file. In this case, the user must manually identify these outliers and correct them.

During the filling, if the wet-bulb temperature is larger than the dry-bulb temperature, then the program will force the wet bulb to equal the dry bulb.

ACCESSING DATA

The preferred way to request data is to access the weather request web page currently located at http://www.eere.energy.gov/buildings/energy_tools/cfm/weather_request.cfm. The process is simple and outlined on the web page. It is possible to enter the 4-character ICAO (International Civil Aviation Orgranization) station ID directly or to search for a valid ID by selecting the "search" option on the home page. If the user requests data that is outside of the available time stored in the database, then the program will only report the data that falls within the bounds. All the time periods displayed on the web page are in Greewich Mean Time (GMT). Upon completion of the request, data will be sent within an hour.

If no data are received within eight hours, the process should be repeated. Certain e-mail clients do not allow for large attachments to be sent during peak hours. In this case, the data may take longer to receive. Also, some e-mail clients do not allow for large attachments at all. In this case, the requester should break up the dates to receive smaller files, then piece the files together manually. It is possible that information may not be found for the station entered, and in this case, an empty email will be returned to the user.

RECEIVED DATA FORMAT

The output of the weather data is available in two formats: CSV and IWEC.

Both formats search for a station listing that details the parameters of the station such as longitude, latitude, and time offset for GMT. The program attempts to convert the time formats in the files to local standard time. If no station information was found or the time offset was not present, then a line in the output file will warn the user that the time stamps are not in local standard time. All IWEC files need to be in local standard time before using them in an energy simulator. Conversely, if the time offset values are found, then the time stamps will be in local standard time. To decipher the format of the IWEC file refer to the IWEC Manual available from www.eere.energy.gov. The format of the CSV files is listed below.

CSV

The CSV files have several columns of information and different units, which are defined in Table 1. The definitions for the weather observations are located in Table 2.

Table 1. Column and Units of CSV Files

Column	Units/Notes
Local Date and Time	MM/DD/YYYY HH:MM[1]
Greenwich Mean Time	MM/DD/YYYY HH:MM
Dry-Bulb Temperature	Tenths of Degree Celsius
Wet-Bulb Temperature	Tenths of Degree Celsius
Wind Direction	Degrees (0º = North)
Wind Speed	Tenths of Meter Per Second
Wind Gusts	Tenths of Meter Per Second
Wind Variability	Range in Degrees
Altimeter	Hundreds of Pascals
Visibility	Tenths of Kilometers
Weather Description	MI, BC, PR, TS, BL, SH, DR, FZ
Weather Precipitation	DZ, RA, SN, SG, IC, PL, GR, GS, UP
Weather Obscuration	BR, FG, FU, VA, SA, HZ, PY, DU
Weather Other	SQ, SS, DS, PO, FC
Cloud Cover Level 1	Identifier + Height in Hundreds of Feet
Cloud Cover Level 2	Identifier + Height in Hundreds of Feet
Cloud Cover Level 3	Identifier + Height in Hundreds of Feet
Sea Level Pressure	Hundreds of Pascals

Table 2. Weather Observation Definitions

Weather Observation	Description	Weather Observation	Description
BC	Patches	MI	Shallow
BL	Blowing	PL	Ice Pellets
BR	Mist	PO	Sand Whirls
DR	Low Drifting	PR	Partial
DS	Dust storm	PY	Spray
DU	Widespread Dust	RA	Rain
DZ	Drizzle	SA	Sand
FC	Funnel Cloud	SG	Snow Grains
FG	Fog	SH	Showers
FU	Smoke	SN	Snow
FZ	Freezing	SQ	Squalls
GR	Hail	SS	Sandstorm
GS	Small Hail	TS	Thunderstorm
HZ	Haze	UP	Unknown
IC	Ice Crystals	VA	Volcanic Ash

More information on the cloud cover and weather descriptors can be found in the *Surface Weather Observations and Reports Handbook.*

DATA LIMITATIONS

There are several limitations to the data including missing data and no solar insolation. The biggest limitation is missing data. To prevent missing data, many procedures must operate seamlessly. The NWS ftp site must be active, the local data collection

---

[1] If the program does not find a valid GMT offset for the station, then this column will be the same as GMT.

SimBuild 2004, IBPSA-USA National Conference Boulder, CO, August 4-6, 2004.

2

computer must be on and stable, and local Internet services must be available. If any of these variables are unavailable, then missing data will occur. Also, several international sites download data at different intervals other than hourly.

The data received from the NWS is only as good as the data reported. The parser and database do not determine if the data was off by a certain percent because of a failing sensor or unusual conditions during the time of the reading.

The data also lacks solar insolation values. The data does contain sky cover information, see Table 3, that could potentially be translated to solar insolation values, but this database does not perform these calculations.

Table 3. Sky Cover Information

Sky Condition	Description
VV	Vertical Visibility
SKC or CLR	Clear
FEW	Few
SCT	Scattered
BKN	Broken
OVC	Overcast

CONCLUSIONS

The weather data contained in the database is limited, but it can provide a decent starting point in determining weather trends for a local area. The database fills data up to two days using linear interpolation or trending the previous days using offsets. The data are available in IWEC format or CSV format.

ACKNOWLEDGMENTS

REFERENCES

International Weather for Energy Calculations (IWEC Weather Files) User's Manual (2001); American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE).

Surface Weather Observations and Reports (1995); FCM-H1-1995; Washington, D.C.: U.S. Department of Commerce/National Oceanic and Atmospheric Administration.

SimBuild 2004, IBPSA-USA National Conference Boulder, CO, August 4-6, 2004.

3