CADB v2 Daily Summary Output Documentation

Edited: Melissa Ou
Date last edited: 6/4/2020

This document describes the output files from the climate assessment database (CADB) daily station summary observation software that was rewritten and released in 2019.

Background and changes

The original software to produce the Climate Assessment Database (CADBv1) was written many years ago in Fortran 77. This software has been rewritten in Python with improved methodology and documentation (CADBv2). Additionally, work has been done upstream to provide more extensive data to users. The goal of the rewrite was to produce a more robust version of output with well understood techniques that would be easily maintainable. Below are some changes between CADBv1 and CADBv2:

General Changes

- In many of the CADBv1 files, metar stations were identified as '99' followed by 3 characters. CADBv2 will now use 4 characters after '99'. Therefore the metar stations will change from 5 to 6 characters in total (e.g. '99CHO' will be '99KCHO'). This allows potentially more stations because the upstream data is actually reported with 4 characters, with the first character denoting the region. In CADBv1, we were using only U.S. metars. Additionally, there are stations in the CONUS and Alaska that have the same 3 last characters with a different char at the beginning. CADBv1 only picked the CONUS version of the 3-char station if it was a duplicateFor ex. There is a KABR and PABR, with KABR being the CONUS station and PABR being the Alaska station. CADBv1 only chose the CONUS station, precluding the Alaska PABR station from being output. CADBv1 would now include, e.g. 99KABR and 99PABR.
  **To map the CADBv1 stations to the new CADBv2, for CADBv1 stations - include a 'K' to the metar stations (ones with characters instead of integer) in front of the 3 characters, ONLY for stations that were used in CADBv1. CADBv2 now includes stations that begin with other letters besides 'K', whereas previously CADBv1 only produced data for stations beginning with 'K'.
- There is a new column with station call letters, if there is one. If not, a '-99999' is listed for a station that does not have a station call.

- More stations are being included in the report. The number of stations may vary day to day depending on the reports that come in.
- Many stations have higher resolution of lat/lon information (up to 4 decimal places).
- CADBv2 files are in CSV format now, rather than fixed width.
- File names have some changes - daily files are named *daily_summary_$YYYY$mm$dd.csv*.

Public web page precipitation and temperature tables:

- CADBv1 used to only include U.S. stations, and now includes global.
- Tables used to be formatted as fixed width and are now formatted as CSV, which is easier to import and use in most applications.
- Include many more variables of data than temperature and precipitation. Variables are listed in the 'File name and formatting' section below.

Internal NOAA file Changes

- CADBv1 produced a 'wx' and 'qz' file for internal NOAA users - the qz file mainly contained only temperature and precipitation, whereas the wx file contained more variables. These 'qz' files will no longer be produced, in lieu of a more expanded variable file.

The CADB should have updated output files created by 8am local eastern time (New York), daily.

File descriptions

File name and formatting

daily_summary_$YYYY$mm$dd_v2.csv - ASCII file is comma-delimited. Below contains more information about the content and format of the data.

Header - First row - contains column names.

Delimiter- Comma delimited with header of column names (CSV).

Missing values - A missing value for all variables is set to -99999

Column info:
Each of the columns in the file are listed in order below (right to left cols). Below info is formatted as:

Name of col - *(unit, # decimals) description | Any related format info or notes*

stn_id - (NA, NA) Station ID | Represented by 6 characters, prepending Metar station IDs (without a numeric only synoptic ID) with '99' (e.g. KCHO -> 99KCHO). For metar stations that have both a station number and station call, the station number will be used instead of the call. This is based off of the station reference list file. To find if there is an associated station call with station number, you can use the station ref file to pair them. Use the latest cpc_station_library.txt on CPC FTP: ftp://ftp.cpc.ncep.noaa.gov/cadb_v2/library/
stn_call - (NA, NA) Station call letters, if applicable.
city - (NA, NA) City name. Does not contain commas, apostrophes, or spaces. May contain other symbols.
state - (NA, NA) United States state abbreviation. Represented by 2 characters. For non-U.S. locations this field value is denoted as a missing value.
country - (NA, NA) Country name. Does not contain commas, apostrophes, or spaces. May contain other symbols.
date - (date, NA) Valid date | Formatted as YYYYmmdd.
lat - (deg , 4 decimal places) Latitude of station | Values range from -90.0000 to 90.0000.
lon - (deg, 4 decimal places) Longitude of station | Values range from -180.0000 to 180.0000
elev - (meters, ones) Elevation of station.
tmax - (Deg C, tenths) Maximum temperature.
tmin - (Deg C, tenths) Minimum temperature.
report_p - (mm, tenths) Reported precipitation total.
final_p - (mm, tenths) Final estimated total precipitation based on combining estimates from reported precip and weather codes. | Combination involves a weighted process to calculate a final summary value.
p_flag - (int, ones) Precipitation flag denoting the source/quality of the final precip estimation value. This may also be interpreted as a precip quality flag. | Definition of flag values are described below.
num_6hr_p - (int, ones) Number of 6hr report precip observations associated with a 3 hour time step. Time steps represent 24 hours, relative to the precip bounding period of each station. The 3 hour value is obtained by taking the maximum 6hr report precip value at the 3 hour time and the 2 hour prior. The max value would be 8. This number is prior to performing QC associated with assessing duplicate/overlapping report precip. Typically used as a diagnosis value.
wxchars - (NA, NA) 8 Weather character string associated with the weather code at the same times represented in the wind speed summary. | The default value is '////////' and is replaced by other characters if applicable per the weather codes in the report.
trace - (int, ones) Flag denoting 0 for no trace precipitation (default), or 1 for trace precipitation.
vp - (mb, tenths) Vapor pressure. | Values are based on air temperature and dewpoint values neighboring the expected heat peak time.

vp_def - (mb , tenths) Vapor pressure deficit | This is the saturated vapor pressure minus the actual vapor pressure.
slp_6, slp_12, slp_18, slp_0 - (mb, tenths) Sea level pressure at 4 different times. | The summary times are at specific date/times including 6Z, 12Z, and 18Z of the valid date, and 0Z of the day after the valid date.
max_rh - (percent, ones) Maximum relative humidity in the 24-hour precipitation report period.
min_rh - (percent, ones) Minimum relative humidity in the 24-hour precipitation report period.
at - (Deg C, tenths) Apparent temperature (heat index). | This is the maximum apparent heat temperature value during the 24-hour precipitation bounded period.
wc - (Deg C, tenths) Wind chill. | This is the minimum of either the wind chill or the minimum temperature values over the 24-hour precipitation bounded period.
wspd_3, wspd_6, wspd_9, wspd_12, wspd_15, wspd_18, wspd_21, wspd_24 - (kts, tenths) Wind speeds at 8 different times. | Summary times are different for each station. The hours in the field name (except for 0) represent the number of hours after the beginning of the 24-hour precipitation report period the summary time represents. Each station has a unique 24-hour precip bounding period based on the location.

Precip flag values

These represent values in the column ‘p_flag’.
-9 - Station precip not finished being processed yet. All stations were initialized with this.
For zero and positive flags: The lowest values indicate the most trusted/reliable type of precip estimate, assuming report precip is more reliable than weather codes and positive data is more info than zero or missings.
0 - Set final estimate to a 24hr report precip value. The 24hr precip value was a positive value. Determined from report_precip_estimate()

*Final est is the max 24hr precip value from 3hrs prior to end bound time to end bound time.

The below flags are set AFTER assessing 3hrly QC-ed/summarized report precip and wxcode estimate values by calling final_precip_estimate() below (order is from most to least trusted):
1 - There were no missing report precip values.

*Final est is sum of report precip.

3 - There was at least one missing report precip value (and one or more zero or positive report precip values) and a valid positive wxcode est was used to add to the report precip.

*Final est is combined value of report and wxcode precip.

5 - There was at least one report precip value missing (and one or more zero or positive report precip values) but no positive wxcode est deemed valid to add to report precip (there was however at least one positive wxcode value originally). The quality may be similar to level 3, since there may not have been any precip during the missing time, hence no expected positive wxcode est value.

*Final est is only the sum of report precip.

7 - All report precip values were missing but there was at least one valid positive wxcode est value included.

* Final est only includes a wxcode est.

9 - All report precip values were missing and there were no valid positive wxcode est values.

*Final est is a missing.

Resources

The latest updated CPC station reference list is available to the public:
ftp://ftp.cpc.ncep.noaa.gov/cadb_v2/cpc_station_library.txt