NOAA NCDC dataset attributes

The attributes, or "flags", for each row of the output for data may have a flag with it. Each `datasetid` has it's own set of flags. The following are flag columns, and what they stand are. `fl_` is the beginning of each flag column name, then one or more characters to describe the flag, keeping it short to maintain a compact data frame. Some of these fields are the same across datasetids, but they may have different possible values. See below details on each dataset.

- *fl*c__ = completeness
- *fl*m__ = measurement
- *fl*d__ = day
- *fl*q__ = quality
- *fl*s__ = source
- *fl*t__ = time
- *fl*cmiss__ = consecutive missing
- *fl*miss__ = missing
- *fl*u__ units

Datasets

- ANNUAL
- GHCND
- GHCNDMS
- NORMAL_ANN
- NORMAL_DLY
- NORMAL_HLY
- NORMAL_MLY
- PRECIP_HLY
- PRECIP_15
- NEXRAD2 - not working yet
- NEXRAD3 - not working yet

Dataset: ANNUAL

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/ANNUAL_documentation.pdf

fl_m (Measurement Flag)   A - Accumulated amount. This value is a total that may include data from a previous month or months (TPCP).

B - Adjusted Total. Monthly value totals based on proportional available data across the entire month. (CLDD, HTDD)

E - An estimated monthly or annual total.

I - Monthly means or totals based on incomplete time series. 1 to 9 days are missing. (MMNT,MMXP, MMXT, MNTM, TPCP, TSNW)

M - used to indicate data element missing.

S - Precipitation for the amount is continuing to be accumulated. Total will be included in a subsequent value (TPCP).

Example: Days 1-20 had 1.35 inches of precipitation, then a period of accumulation began. The element TPCP would then be 00135S and the total accumulated amount value appears in a subsequent monthly value.

If TPCP = 0 there was no precipitation measured during the month. Flag 1 is set to "S" and the total accumulated amount appears in a subsequent monthly value.

T - Trace of precipitation, snowfall, or snow depth. The precipitation data value will = "00000". (EMXP, MXSD, TPCP, TSNW)

- - The phenomena in question occurred on several days. The date in the DAY field is the last day of occurrence.

(blank) No report

fl_q (Quality Flag) A – Accumulated amount E – Estimated value + - Value occurred on more than one day – last date of occurrence is used

fl_d (Number of Days flag) Number of days is given as 00 when all days in the month are considered in computing data value or otherwise the maximum number of consecutive days in the month considered in computing the data value.

fl_u (Units) C -Whole degree Celsius D - Whole Fahrenheit Degree Day F - Whole degree Fahrenheit HI - Hundredths of inches I - Whole inches M - Whole miles MH – Miles per hour MM – Millimeters NA – No units applicable (dimensionless) TC – Tenths of degrees Celsius TF – Tenths of degrees Fahrenheit TI – Tenths of inches TM – Tenths of millimeters 1 - Soils – degrees Fahrenheit, soil depths in inches and hundredths 2 - Soils – degrees Celsius, soil depth in whole centimeters 3 – Soils – degrees Celsius, soil, soil depth in inches and hundredths 4 – Soils – degrees Fahrenheit, soil depth in whole centimeters 5 – Soils – If the soil station closed during the current month, "5" indicates the station has closed.

Dataset: GHCND

flm_m (Measurement Flag) Blank = no measurement information applicable A = value in precipitation or snow is a multi-day total, accumulated since last measurement (used on Daily Form pdf file) B = precipitation total formed from two 12-hour totals D = precipitation total formed from four six-hour totals K = converted from knots L = temperature appears to be lagged with respect to reported hour of observation O = converted from oktas P = identified as "missing presumed zero" in DSI 3200 and 3206 T = trace of precipitation, snowfall, or snow depth W = converted from 16-point WBAN code (for wind direction)

fl_q (Quality Flag) Blank = did not fail any quality assurance check D = failed duplicate check G = failed gap check I = failed internal consistency check K = failed streak/frequent-value check L = failed check on length of multiday period M = failed mega-consistency check N = failed naught check O = failed climatological outlier check R = failed lagged range check S = failed spatial consistency check T = failed temporal consistency check W = temperature too warm for snow X = failed bounds check Z = flagged as a result of an official Datzilla investigation

fl_s (Source Flag) Blank = No source (i.e., data value missing) 0 = U.S. Cooperative Summary of the Day (NCDC DSI-3200) 6 = CDMP Cooperative Summary of the Day (NCDC DSI-3206) 7 = U.S. Cooperative Summary of the Day – Transmitted via WxCoder3 (NCDC DSI-3207) A = U.S. Automated Surface Observing System (ASOS) real-time data (since January 1, 2006) a = Australian data from the Australian Bureau of Meteorology B = U.S. ASOS data for October 2000-December 2005 (NCDC DSI-3211) b = Belarus update E = European Climate Assessment and Dataset (Klein Tank et al., 2002) F = U.S. Fort data G = Official Global Climate Observing System (GCOS) or other government-supplied data H = High Plains Regional Climate Center real-time data I = International collection (non U.S. data received through

personal contacts) K = U.S. Cooperative Summary of the Day data digitized from paper observer forms (from 2011 to present) M = Monthly METAR Extract (additional ASOS data) N = Community Collaborative Rain, Hail,and Snow (CoCoRaHS) Q = Data from several African countries that had been "quarantined", that is, withheld from public release until permission was granted from the respective meteorological services R = NCDC Reference Network Database (Climate Reference Network and Historical Climatology Network-Modernized) r = All-Russian Research Institute of Hydrometeorological Information-World Data Center S = Global Summary of the Day (NCDC DSI-9618) NOTE: "S" values are derived from hourly synoptic reports exchanged on the Global Telecommunications System (GTS). Daily values derived in this fashion may differ significantly from "true" daily data, particularly for precipitation(i.e., use with caution). u = Ukraine update W = WBAN/ASOS Summary of the Day from NCDC's Integrated Surface Data (ISD). X = U.S. First-Order Summary of the Day (NCDC DSI-3210) Z = Datzilla official additions or replacements z = Uzbekistan update

fl_t (Time of Observation)   Is the (2 digit hour, 2 digit minute) 24 hour clock time of the observation given as the local time at the station of record.

Dataset: GHCNDMS

Observation(s) is/are synonymous with elements or values, and defined in Table A below. 9's in a field (e.g.9999) indicate missing data or data that has not been received.

Flags: Missing Flag , Consecutive Missing Flag

fl_miss (Missing Flag)   Defined as total number of days observation/element is missing in that month. This can be taken as a measure of quality or completeness as the higher the number of days sampled in the month, the more representative the value is for the entire month.

fl_cmiss (Consecutive Missing Flag)   Defined as the maximum number of consecutive days in the month that an observation/element is missing.

Dataset: NORMAL_ANN

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/NORMAL_ANN_documentation.pdf

Description   The 1981-2010 Normals comprise all climate normals using the thirty year period of temperature, degree days, precipitation, snowfall, snow depth, wind, etc. Data is organized into hourly, daily, monthly, seasonal and annual. This document describes the elements and layout of the Seasonal and Annual Normals which are derived from a composite of climate records from numerous sources that were merged and then subjected to a suite of quality assurance reviews.

fl_c (Completeness Flag)   Flags accompany every Normals value and indicate the completeness of the data record used to compute each value, accounting for methodological differences for different product classes. There are six flag options described generally in Table 1 below. Due to methodological differences, the flags are applied somewhat differently between the temperature-based normals and the precipitation-based normals. For the precipitation-based and hourly normals, the following flags were assigned independently for each normals value reported based on number of years available for that individual calculation. For temperature-based normals, strong precedence is given to the monthly normals of maximum and minimum temperature or derived from the flags for these two variables.

Table 1 (CompletenessFlag/Attribute) C = complete (all 30 years used) S = standard (no more than 5 years missing and no more than 3 consecutive years missing among the sufficiently complete years) R =

representative (observed record utilized incomplete, but value was scaled or based on filled values to be representative of the full period of record) P = provisional (at least 10 years used, but not sufficiently complete to be labeled as standard or representative). Also used for parameter values on February 29 as well as for interpolated daily precipitation, snowfall, and snow depth percentiles. Q = quasi-normal (at least 2 years per month, but not sufficiently complete to be labeled as provisional or any other higher flag code. The associated value was computed using a pseudonormals approach or derived from monthly pseudonormals. Blank = the data value is reported as a special value (see section B under III. Additional Information below).

Note: Flags Q and R aren't applicable to average number of days with different precipitation, snowfall, and snow depth threshold exceedance; precipitation/snowfall/snow probabilities of occurrence. Further, Q flags are not applicable for standard deviations.

Dataset: NORMAL_DLY

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/NORMAL_DLY_documentation.pdf

Description   The 1981-2010 Normals comprise all climate normals using the thirty year period of temperature, degree days, precipitation, snowfall, snow depth, wind, etc. Data is organized into hourly, daily, monthly, seasonal and annual. This document describes the elements and layout of the Daily Normals which are derived from a composite of climate records from numerous sources that were merged and then subjected to a suite of quality assurance reviews.

fl_c (Completeness Flag)   Same as NORMAL_ANN, see the description above at Completeness Flag.

Dataset: NORMAL_HLY

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/NORMAL_HLY_documentation.pdf

Description   The 1981-2010 Normals comprise all climate normals using the thirty year period of temperature, degree days, precipitation, snowfall, snow depth, wind, etc. Data is organized into hourly, daily, monthly, seasonal and annual normals. This document describes the elements and layout of the Hourly Normals which are derived from a composite of climate records from numerous sources that were merged and then subjected to a suite of quality assurance reviews. The hourly normals provide a suite of descriptive statistics based on hourly observations at a few hundred stations from across the United States and its Pacific territories. Statistics are provided as 30- year averages, frequencies of occurrence, and percentiles for each hour and day of the year. These products are useful in examination of the diurnal change of a particular variable.

For temperature, dew point and mean sea level pressure an average hourly value as well as a 10th and 90th percentile of hourly values is given. For heating and cooling degree hours, an average hourly value is given using a 65 degree F base. Average hourly values are also given for heat index and wind chill. Cloud cover statistics include percent frequency of clear, few, scattered, broken and overcast conditions. Wind statistics include prevailing and secondary wind direction and percent frequency, average wind speed, percentage of calm winds and mean wind vector direction and magnitude.

The statistics are computed from the ISD-lite dataset for which more information can be found at http://www.ncdc.noaa.gov/oa/climate/isd/index.php?name=isd-lite. 262 stations were selected from the ISD-lite data, based on their completeness and membership in a list of what were known as "first order stations." These are typically airport locations with the needed 24 hours/day observations to make hourly normals meaningful. All stations had at least 27 of the 30 years represented.

Each hourly normal is computed on the basis of 450 possible values. This is the aggregation of the value for a particular date and time, plus and minus 7 days, over each of 30 years. If fewer than 350 valid values are present, the output is given as the special value 9999. No normals are computed for February 29, but data

for February 29 is included in the 15 day window for leap years. The original data has been shifted from Greenwich Mean Time to an end product in local standard time.

fl_c (Completeness Flag)   Same as NORMAL_ANN, see the description above at Completeness Flag.

Dataset: NORMAL_MLY

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/NORMAL_MLY_documentation.pdf

Description   The 1981-2010 Normals comprise all climate normals using the thirty year period of temperature, degree days, precipitation, snowfall, snow depth, wind, etc. Data are organized into hourly, daily, monthly, seasonal and annual. This document describes the elements and layout of the Monthly Normals which are derived from a composite of climate records from numerous sources that were merged and then subjected to a suite of quality assurance reviews.

fl_c (Completeness Flag)   Same as NORMAL_ANN, see the description above at Completeness Flag.

Dataset: PRECIP_HLY

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/PRECIP_HLY_documentation.pdf

Description   Hourly Precipitation Data (labeled Precipitation Hourly in Climate Data Online system) is a database that gives time-sequenced hourly precipitation amounts for a network of over 7000 reporting station located primarily in the United States. Data is collected from a variety of sources including National Weather Service reporting stations, volunteer cooperative observers, Federal Aviation Administration (FAA), utility companies, etc.

Concerning rain gages/data processing: Data from weighing rain gages, Fischer-Porter gages, Universal rain gages and in recent years, more modern measuring equipment in conjunction with automated recording sites, etc. have been used in this dataset over the period. Precipitation values have been checked and edited as necessary by both automated and manual methods. Because of some inconsistencies identified with the earlier data (prior to 1996), historical data were reprocessed in 1997. This rehabilitated data covered 53 million observations between 1900 and 1995. Similar quality control checks are in place that maintain consistency between the historical and operationally received data.

fl_m (Measurement Flag)   Note: This field is left blank when no flag is needed.

a: Begin accumulation. A data value of 99999 accompanies this flag. It indicates that the accumulation has begun at some time during the hour.

A: End accumulation (an amount is associated with this flag). It indicates that accumulation has ended sometime during the hour. Accumulated period indicates that the precipitation amount is correct, but only the exact beginning and ending times are known. A data value of 99999 occurring on the last day and hour of a month indicates the accumulation continues into the next month.

, (comma): Used at the beginning of a data month when an accumulation is in progress from the previous month. A data value of 99999 always accompanies this flag. This flag is used prior to 1984.

{ : Begin deleted period during the hour (inclusive). The original data were received, but were unreadable or clearly recognized as noise. A value of 99999 accompanies this flag. Primarily used since 1984. Also used in Alaska for 1976-1978.

}: End deleted period during the hour (inclusive). The original data were received, but were unreadable or clearly recognized as noise. A value of 99999 accompanies this flag. Primarily used since 1984. Also used in Alaska for 1976-1978.

[: Begin missing period during the hour (inclusive). A value of 99999 accompanies this flag.

]: End missing period during the hour (inclusive). A value of 99999 accompanies this flag. Prior to 1984 if precipitation occurred during the last hour of the missing period, the ending missing value appears with a non-zero value. Beginning in 1984, the beginning and ending hours of the missing period are recorded as "99999[" and "99999]," respectively.. A missing flag indicates that the data were not received. The flag appears on the first and last day of each month for which data were not received or not processed by NCDC.

E: Evaporation may have occurred. Data may or may not be reliable. This flag was used during the period 1984-1993.

g: Only used for day 1, hour 0100, when precipitation amount is zero.

T: Indicates a "trace" amount. Data value with this will be zero. "T" flags appear on National Weather Service data only since July 1996.

M: Missing data. No data available for this period.

fl_q (Data Quality) Z: Indicates probable amounts as a result of melting frozen precipitation. This flag may be used to identify those sites that are deficient in the manner the snow shields are employed. Used since January 1996.

R: This data value failed one of the NCDC's quality control tests.

Q: Pre 1996 usage - Indicates value failed an extreme value test (value will be present). Data are to be used with caution.

Extreme tests used are 1) value was not an accumulated amount and was higher than the one-hour statewide 100 year return period precipitation amount or 2) if they value was an accumulated amount and was higher than the 24 hour statewide extreme precipitation total.

Q: 1996 to present usage - A single erroneous value (value will be present). Rarely used since 1996.

q: An hourly value excludes one or more 15 minute periods. Lowest data resolution is 15 minutes. Used since January 1996.

Dataset: PRECIP_15

More info: http://www1.ncdc.noaa.gov/pub/data/cdo/documentation/PRECIP_15_documentation.pdf

Description 15 Minute Precipitation Data (labeled Precipitation 15 Minute in Climate Data Online system) is a database that gives time-sequenced quarter-hour precipitation amounts for a network of over 3600 reporting station located primarily in the United States. Data is collected from a variety of sources including National Weather Service reporting stations, volunteer cooperative observers, Federal Aviation Administration (FAA), utility companies, etc.

Concerning rain gages/data processing: Data from Fischer-Porter gages between May 1971 and December 1983 have been used in this dataset. Precipitation values have been checked and edited as necessary by both automated and manual methods. Data processing procedures were updated in January 1984 to produce the element structured data base files and further enhanced beginning with the January 1996 data month. Currently, interactive quality control procedures are in place that has added many checks and features and data are subjected to automated editing procedures that reduce the manual handling of the data.

fl_m (Data Measurement)   QPCP

Note: This field is left blank when no flag is needed.

a: Begin accumulation. A data value of 99999 accompanies this flag. It indicates that the accumulation has begun at some time during the 15 minute period.

A: End accumulation (an amount is associated with this flag). It indicates that accumulation has ended sometime during the 15 minute period. Accumulated period indicates that the precipitation amount is correct, but only the exact beginning and ending times are known. A data value of 99999 occurring on the last day and hour of a month indicates the accumulation continues into the next month.

, (comma): Used at the beginning of a data month when an accumulation is in progress. This flag is used prior to 1984.

{ : Begin deleted period during the hour (inclusive).

}: End deleted period during the hour (inclusive).

[: Begin missing period during the hour (inclusive).

]: End missing period during the hour (inclusive).

E: Evaporation may have occurred. Data may or may not be reliable. This flag was used during the period 1984-1993.

g: Only used on day 1 when precipitation amount is zero.

T: Indicates a "trace" amount. Data value with this will be zero. "T" flags appear on National Weather Service data only.

M: Missing data. No data available for this period.

QGAG

a: begin accumulation (indicates measurement periods overlapped) A: end accumulation [: begin missing]: end missing {: begin delete }: end delete S: gage reset

fl_q (Data Quality)   QPCP

X: Used for data prior to 1996 as part of a 1997 data rehabilitation effort. Indicates value failed an extreme value test; data are to be used with caution. Extreme tests were: 1) if the value was not an accumulated precipitation total, the value failed the one hour statewide 100 year return period precipitation and 2) if the value was an accumulated precipitation total, the value failed the 24 hour statewide extreme precipitation total.

Z: Indicates probable amounts as a result of melting frozen precipitation. This flag may be used to identify those sites that are deficient in the manner the snow shields are employed. Used since January 1996.

R: This data value failed one of NCDC's quality control tests.

Q: A single erroneous value (value will be present). Used since January 1996.

q: An hourly value excludes one or more 15 minute periods. Lowest data resolution is 15 minutes. Used since January 1996.

A: Accumulated period and amount. An accumulated period indicates that the precipitation amount is correct, but the exact beginning and ending times are only known to the extent that the precipitation occurred sometime within the accumulation period.

QGAG

Q: Questionable value. Data not used. P: Punched mechanism failure, missing punch assumed. Assumed punch value being used. V: Evaporation likely. Gage value has dropped. Data are being used.

fl_u (Units Flag)   HI indicates data values (QGAG or QPCP) are in hundredths of inches. HT indicates data values (QGAG or QPCP) are in tenths of inches.