

README FILE FOR DAILY GLOBAL HISTORICAL CLIMATOLOGY NETWORK (GHCN-DAILY) Version 3.31

----- How to cite:

Note that the GHCN-Daily dataset itself now has a DOI (Digital Object Identifier) so it may be relevant to cite both the methods/overview journal article as well as the specific version of the dataset used.

The journal article describing GHCN-Daily is:

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology, 29, 897-910, doi:10.1175/JTECH-D-11-00103.1.

To acknowledge the specific version of the dataset used, please cite:

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeill, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. [indicate subset used following decimal, e.g. Version 3.12].

NOAA National Climatic Data Center. <http://doi.org/10.7289/V5D21VHZ> [access date].

I. DOWNLOAD QUICK START

Start by downloading "ghcnd-stations.txt," which has metadata for all stations.

Then download one of the following TAR files:

- "ghcnd-all.tar.gz" if you want all of GHCN-Daily, OR
- "ghcnd-gsn.tar.gz" if you only want the GCOS Surface Network (GSN), OR
- "ghcnd-hcn.tar.gz" if you only want the U.S. Historical Climatology Network (U.S. HCN).

Then uncompress and untar the contents of the tar file, e.g., by using the following Linux command:

```
tar xzvf ghcnd_XXX.tar.gz
```

Where "XXX" stands for "all", "hcn", or "gsn" as applicable. The files will be extracted into a subdirectory under the directory where the command is issued.

ALTERNATIVELY, if you only need data for one station:

- Find the station's name in "ghcnd-stations.txt" and note its station identification code (e.g., PHOENIX AP (Airport) is "USW00023183"); and
 - Download the data file (i.e., ".dly" file) that corresponds to this code (e.g., "USW00023183.dly" has the data for PHOENIX AP).
- Note that the ".dly" file is located in the "all" subdirectory.

----- II. CONTENTS OF <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily>

all:	Directory with ".dly" files for all of GHCN-Daily
gsn:	Directory with ".dly" files for the GCOS Surface Network (GSN)
hcn:	Directory with ".dly" files for U.S. HCN
by_year:	Directory with GHCN Daily files parsed into yearly subsets with observation times where available. See the /by_year/readme.txt and /by_year/ghcn-daily-by_year-format.rtf files for further details
by_station:	Directory of GHCN daily station data in period of record comma separate (csv) files. See readme-by_station.txt for additional details
grid:	Directory with the GHCN-Daily gridded dataset known as HadGHCND

papers: Directory with pdf versions of journal articles relevant to the GHCN-Daily dataset

figures: Directory containing figures that summarize the inventory and processing of GHCN-Daily station records

superghcnd: Directory containing a comma delimited format of GHCN-Daily with station metadata integrated into the data. Two files are provided each day. The superghcnd_full file contains all GHCN-Daily data in comma delimited format and the superghcnd_diff file which contains the differences in the dataset between two successive update runs. See the readme.txt file in the superghcnd directory for more details.

ghcnd-all.tar.gz: TAR file of the GZIP-compressed files in the "all" directory

ghcnd-gsn.tar.gz: TAR file of the GZIP-compressed "gsn" directory

ghcnd-hcn.tar.gz: TAR file of the GZIP-compressed "hcn" directory

ghcnd-countries.txt: List of country codes (FIPS) and names

ghcnd-inventory.txt: File listing the periods of record for each station and element

ghcnd-stations.txt: List of stations and their metadata (e.g., coordinates)

ghcnd-states.txt: List of U.S. state and Canadian Province codes used in ghcnd-stations.txt

ghcnd-version.txt: File that specifies the current version of GHCN Daily

mingle-list.txt: File that provides a list of each source and source identifiers associated with each GHCN-Daily station.

readme.txt: This file

status.txt: Notes on the current status of GHCN-Daily

III. FORMAT OF DATA FILES (".dly" FILES)

Each ".dly" file contains data for one station. The name of the file corresponds to a station's identification code. For example, "USC00026481.dly" contains the data for the station with the identification code USC00026481).

Each record in a file contains one month of daily data. The variables on each line include the following:

Variable	Columns	Type
ID	1-11	Character
YEAR	12-15	Integer
MONTH	16-17	Integer
ELEMENT	18-21	Character
VALUE1	22-26	Integer
MFLAG1	27-27	Character
QFLAG1	28-28	Character
SFLAG1	29-29	Character
VALUE2	30-34	Integer
MFLAG2	35-35	Character
QFLAG2	36-36	Character
SFLAG2	37-37	Character
.	.	.
.	.	.
.	.	.
VALUE31	262-266	Integer
MFLAG31	267-267	Character
QFLAG31	268-268	Character
SFLAG31	269-269	Character

These variables have the following definitions:

ID is the station identification code. Please see "ghcnd-stations.txt"

for a complete list of stations and their metadata.
 YEAR is the year of the record.
 MONTH is the month of the record.
 ELEMENT is the element type. There are five core elements as well as a number of addition elements.

The five core elements are:

PRCP = Precipitation (tenths of mm)
 SNOW = Snowfall (mm)
 SNWD = Snow depth (mm)
 TMAX = Maximum temperature (tenths of degrees C)
 TMIN = Minimum temperature (tenths of degrees C)

The other elements are:

ACMC = Average cloudiness midnight to midnight from 30-second ceilometer data (percent)
 ACMH = Average cloudiness midnight to midnight from manual observations (percent)
 ACSC = Average cloudiness sunrise to sunset from 30-second ceilometer data (percent)
 ACSH = Average cloudiness sunrise to sunset from manual observations (percent)
 ADPT = Average Dew Point Temperature for the day (tenths of degrees C)
 ASLP = Average Sea Level Pressure for the day (hPa * 10)
 ASTP = Average Station Level Pressure for the day (hPa * 10)
 AWBT = Average Wet Bulb Temperature for the day (tenths of degrees C)
 AWDR = Average daily wind direction (degrees)
 AWND = Average daily wind speed (tenths of meters per second)
 DAEV = Number of days included in the multiday evaporation total (MDEV)
 DAPR = Number of days included in the multiday precipitation total (MDPR)
 DASF = Number of days included in the multiday snowfall total (MDSF)
 DATN = Number of days included in the multiday minimum temperature (MDTN)
 DATX = Number of days included in the multiday maximum temperature (MDTX)
 DAWM = Number of days included in the multiday wind movement (MDWM)
 DWPR = Number of days with non-zero precipitation included in multiday precipitation total (MDPR)
 EVAP = Evaporation of water from evaporation pan (tenths of mm)
 FMTH = Time of fastest mile or fastest 1-minute wind (hours and minutes, i.e., HHMM)
 FRGB = Base of frozen ground layer (cm)
 FRGT = Top of frozen ground layer (cm)
 FRTH = Thickness of frozen ground layer (cm)
 GAHT = Difference between river and gauge height (cm)
 MDEV = Multiday evaporation total (tenths of mm; use with DAEV)
 MDPR = Multiday precipitation total (tenths of mm; use with DAPR and DWPR, if available)
 MDSF = Multiday snowfall total
 MDTN = Multiday minimum temperature (tenths of degrees C; use with DATN)
 MDTX = Multiday maximum temperature (tenths of degrees C; use with DATX)
 MDWM = Multiday wind movement (km)
 MNPN = Daily minimum temperature of water in an evaporation pan (tenths of degrees C)
 MXPN = Daily maximum temperature of water in an evaporation pan (tenths of degrees C)
 PGTM = Peak gust time (hours and minutes, i.e., HHMM)
 PSUN = Daily percent of possible sunshine (percent)
 RHAV = Average relative humidity for the day (percent)

RHMN = Minimum relative humidity for the day (percent)
 RHMX = Maximum relative humidity for the day (percent)
 SN*# = Minimum soil temperature (tenths of degrees C)
 where * corresponds to a code
 for ground cover and # corresponds to a code for soil
 depth.

Ground cover codes include the following:

0 = unknown
 1 = grass
 2 = fallow
 3 = bare ground
 4 = brome grass
 5 = sod
 6 = straw mulch
 7 = grass muck
 8 = bare muck

Depth codes include the following:

1 = 5 cm
 2 = 10 cm
 3 = 20 cm
 4 = 50 cm
 5 = 100 cm
 6 = 150 cm
 7 = 180 cm

SX*# = Maximum soil temperature (tenths of degrees C)
 where * corresponds to a code for ground cover
 and # corresponds to a code for soil depth.
 See SN*# for ground cover and depth codes.

TAXN = Average daily temperature computed as
 $(TMAX+TMIN)/2.0$ (tenths of degrees C)

TAVG = Average daily temperature (tenths of degrees C)
 [Note that TAVG from source 'S' corresponds
 to an average of hourly readings for the period
 ending at 2400 UTC rather than local midnight or other
 Local Standard Time according to a specific
 Met Service's protocol]
 [For sources other than 'S' TAVG is computed in a
 variety of ways including
 traditional fixed hours of the day whereas TAXN
 is solely computed as $(TMAX+TMIN)/2.0$]

THIC = Thickness of ice on water (tenths of mm)

TOBS = Temperature at the time of observation (tenths of degrees C)

TSUN = Daily total sunshine (minutes)

WDF1 = Direction of fastest 1-minute wind (degrees)

WDF2 = Direction of fastest 2-minute wind (degrees)

WDF5 = Direction of fastest 5-second wind (degrees)

WDFG = Direction of peak wind gust (degrees)

WDFI = Direction of highest instantaneous wind (degrees)

WDFM = Fastest mile wind direction (degrees)

WDMV = 24-hour wind movement (km)

WESD = Water equivalent of snow on the ground (tenths of mm)

WESF = Water equivalent of snowfall (tenths of mm)

WSF1 = Fastest 1-minute wind speed (tenths of meters per second)

WSF2 = Fastest 2-minute wind speed (tenths of meters per second)

WSF5 = Fastest 5-second wind speed (tenths of meters per second)

WSFG = Peak gust wind speed (tenths of meters per second)

WSFI = Highest instantaneous wind speed (tenths of meters per second)

WSFM = Fastest mile wind speed (tenths of meters per second)

WT** = Weather Type where ** has one of the following values:

01 = Fog, ice fog, or freezing fog (may include heavy fog)
 02 = Heavy fog or heaving freezing fog (not always
 distinguished from fog)
 03 = Thunder
 04 = Ice pellets, sleet, snow pellets, or small hail
 05 = Hail (may include small hail)

06 = Glaze or rime
 07 = Dust, volcanic ash, blowing dust, blowing sand, or blowing obstruction
 08 = Smoke or haze
 09 = Blowing or drifting snow
 10 = Tornado, waterspout, or funnel cloud
 11 = High or damaging winds
 12 = Blowing spray
 13 = Mist
 14 = Drizzle
 15 = Freezing drizzle
 16 = Rain (may include freezing rain, drizzle, and freezing drizzle)
 17 = Freezing rain
 18 = Snow, snow pellets, snow grains, or ice crystals
 19 = Unknown source of precipitation
 21 = Ground fog
 22 = Ice fog or freezing fog

WV** = Weather in the Vicinity where ** has one of the following values:

01 = Fog, ice fog, or freezing fog (may include heavy fog)
 03 = Thunder
 07 = Ash, dust, sand, or other blowing obstruction
 18 = Snow or ice crystals
 20 = Rain or snow shower

VALUE1 is the value on the first day of the month (missing = -9999).

MFLAG1 is the measurement flag for the first day of the month. There are ten possible values:

Blank = no measurement information applicable
 B = precipitation total formed from two 12-hour totals
 D = precipitation total formed from four six-hour totals
 H = represents highest or lowest hourly temperature (TMAX or TMIN) or the average of hourly values (TAVG)
 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)

QFLAG1 is the quality flag for the first day of the month. There are fourteen possible values:

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 megaconsistency 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

SFLAG1 is the source flag for the first day of the month. There are thirty possible values (including blank, upper and lower case letters):

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

C = Environment Canada

D = Short time delay US National Weather Service CF6 daily summaries provided by the High Plains Regional Climate Center

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)

f = Data provided courtesy of the Fiji Met Service

m = Data from the Mexican National Water Commission (Comision National del Agua -- CONAGUA)

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 = NCEI Reference Network Database (Climate Reference Network and Regional Climate Reference Network)

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).

s = China Meteorological Administration/National Meteorological Information Center/ Climatic Data Center (<http://cdc.cma.gov.cn>)

T = SNOwpack TELEmetry (SNOTEL) data obtained from the U.S. Department of Agriculture's Natural Resources Conservation Service

U = Remote Automatic Weather Station (RAWS) data obtained from the Western Regional Climate Center

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

When data are available for the same time from more than one source, the highest priority source is chosen according to the following priority order (from highest to lowest):

Z,R,D,0,6,C,X,W,K,7,F,B,M,f,m,r,E,z,u,b,s,a,G,Q,I,A,N,T,U,H,S

VALUE2 is the value on the second day of the month

MFLAG2 is the measurement flag for the second day of the month.

QFLAG2 is the quality flag for the second day of the month.

SFLAG2 is the source flag for the second day of the month.

... and so on through the 31st day of the month. Note: If the month has less than 31 days, then the remaining variables are set to missing (e.g., for April, VALUE31 = -9999, MFLAG31 = blank, QFLAG31 = blank, SFLAG31 = blank).

IV. FORMAT OF "ghcnd-stations.txt"

Variable	Columns	Type
ID	1-11	Character
LATITUDE	13-20	Real
LONGITUDE	22-30	Real
ELEVATION	32-37	Real
STATE	39-40	Character
NAME	42-71	Character
GSN FLAG	73-75	Character
HCN/CRN FLAG	77-79	Character
WMO ID	81-85	Character

These variables have the following definitions:

ID is the station identification code. Note that the first two characters denote the FIPS country code, the third character is a network code that identifies the station numbering system used, and the remaining eight characters contain the actual station ID.

See "ghcnd-countries.txt" for a complete list of country codes.

See "ghcnd-states.txt" for a list of state/province/territory codes.

The network code has the following five values:

- 0 = unspecified (station identified by up to eight alphanumeric characters)
- 1 = Community Collaborative Rain, Hail, and Snow (CoCoRaHS) based identification number. To ensure consistency with with GHCN Daily, all numbers in the original CoCoRaHS IDs have been left-filled to make them all four digits long. In addition, the characters "-" and "_" have been removed to ensure that the IDs do not exceed 11 characters when preceded by "US1". For example, the CoCoRaHS ID "AZ-MR-156" becomes "US1AZMR0156" in GHCN-Daily
- C = U.S. Cooperative Network identification number (last six characters of the GHCN-Daily ID)
- E = Identification number used in the ECA&D non-blended dataset
- M = World Meteorological Organization ID (last five characters of the GHCN-Daily ID)
- N = Identification number used in data supplied by a National Meteorological or Hydrological Center
- P = "Pre-Coop" (an internal identifier assigned by NCEI for station records collected prior to the establishment of the U.S. Weather Bureau and their management of the U.S. Cooperative (Coop) Observer Program
- R = U.S. Interagency Remote Automatic Weather Station (RAWS) identifier
- S = U.S. Natural Resources Conservation Service SNOwpack TELeMtry (SNOTEL) station identifier
- W = WBAN identification number (last five characters of the GHCN-Daily ID)

LATITUDE is latitude of the station (in decimal degrees).

LONGITUDE is the longitude of the station (in decimal degrees).

ELEVATION is the elevation of the station (in meters, missing = -999.9).

STATE is the U.S. postal code for the state (for U.S. stations only).

NAME is the name of the station.

GSN FLAG is a flag that indicates whether the station is part of the GCOS Surface Network (GSN). The flag is assigned by cross-referencing the number in the WMOID field with the official list of GSN stations. There are two possible values:

Blank = non-GSN station or WMO Station number not available
GSN = GSN station

HCN/
CRN FLAG is a flag that indicates whether the station is part of the U.S. Historical Climatology Network (HCN) or U.S. Climate Reference Network (CRN). There are three possible values:

Blank = Not a member of the U.S. Historical Climatology
or U.S. Climate Reference Networks
HCN = U.S. Historical Climatology Network station
CRN = U.S. Climate Reference Network or U.S. Regional Climate
Network Station

WMO ID is the World Meteorological Organization (WMO) number for the station. If the station has no WMO number (or one has not yet been matched to this station), then the field is blank.

V. FORMAT OF "ghcnd-countries.txt"

Variable	Columns	Type
CODE	1-2	Character
NAME	4-64	Character

These variables have the following definitions:

CODE is the FIPS country code of the country where the station is located (from FIPS Publication 10-4 at www.cia.gov/cia/publications/factbook/appendix/appendix-d.html).

NAME is the name of the country.

VI. FORMAT OF "ghcnd-states.txt"

Variable	Columns	Type
CODE	1-2	Character
NAME	4-50	Character

These variables have the following definitions:

CODE is the POSTAL code of the U.S. state/territory or Canadian province where the station is located

NAME is the name of the state, territory or province.

VII. FORMAT OF "ghcnd-inventory.txt"

Variable	Columns	Type
ID	1-11	Character
LATITUDE	13-20	Real
LONGITUDE	22-30	Real
ELEMENT	32-35	Character
FIRSTYEAR	37-40	Integer
LASTYEAR	42-45	Integer

These variables have the following definitions:

ID is the station identification code. Please see "ghcnd-stations.txt" for a complete list of stations and their metadata.

LATITUDE is the latitude of the station (in decimal degrees).

LONGITUDE is the longitude of the station (in decimal degrees).

ELEMENT is the element type. See section III for a definition of elements.

FIRSTYEAR is the first year of unflagged data for the given element.

LASTYEAR is the last year of unflagged data for the given element.

VIII. FORMAT OF "mingle-list.txt"

Variable	Column(s)	Type
GHCN-Daily Identifier (ID)	1-11	Character
No of sources assoc. with the station (N)	13-14	Integer
Source Code of lowest ranking source	16	Character
Source Identifier of lowest ranking source	18-28	Character
Source Code of next lowest ranking source	30	Character
Source Identifier of next lowest ranking source	32-42	

.
.
.

Up to Nth (highest ranking) source code and source identifier'
[See section III for a definition of source codes]

IX. REFERENCES

The journal article describing GHCN-Daily is:

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily Database. Journal of Atmospheric and Oceanic Technology, 29, 897-910, doi:10.1175/JTECH-D-11-00103.1.

To acknowledge the specific version of the dataset used, please cite:

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. [indicate subset used following decimal, e.g. Version 3.12].

NOAA National Climatic Data Center. <http://doi.org/10.7289/V5D21VHZ> [access date].

Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. of Climatol., 22, 1441-1453.

Data and metadata available at <http://eca.knmi.nl>

For additional information, please send an e-mail to ncdc.ghcnd@noaa.gov.