National Climatic Data Center

DATA DOCUMENTATION

FOR

DATASET 9290a
Global Synoptic Climatology Network.
A. Canada

February 28, 2005

Version 1.0

National Climatic Data Center
151 Patton Ave.
Asheville, NC 28801-5001 USA

Table of Contents

1. Abstract

These files are a compilation of in situ hourly meteorological observations for Canada initially obtained within the framework of joint efforts to create Global Synoptic Climatology Network among the Meteorological Service of Canada (Downsview, Ontario and Vancouver, British Columbia), Research Institute for Hydrometeorological Information, of the Russian State Committee for Hydrometeorology, Obninsk, Russian Federation, and NOAA National Climatic Data Center. The data were then pre-processed according to a standard quality control routine and generation of derived variables (e.g., some humidity characteristics were derived from others). Currently, there are approximately 170 active stations, which can be operationally updated from the entire list of 768 locations, and another 350 stations are updatable with a delay. The maximum period of the data span is from January 1, 1953 to February 21, 2005.

Variables in the data set include sea level and station pressure, surface air temperature, water vapor pressure, relative humidity, wind speed and direction, several characteristics of cloudiness, and present weather. Each data file (extension .dat) is accompanied with a companion flags file (extension .flg). Thus, each data value is characterized by four flags: data measurement flag, quality control flag, confidence level/status flag, and data source flag. A lot of the original data are also available from the web site of Meteorological Service of Canada (MSC):
http://www.climate.weatheroffice.ec.gc.ca/climateData/canada_e.html.
Archive contains three types of data files:

1. **STN_ID.dat** (768 files, 1 per station)
2. **STN_ID.flg** (768 files, 1 per station), and
3. **Canadian_hly_stn.list.txt** (1 file with essential station metadata such as station identifier, latitude, longitude, elevation, date of the first and the last records, and station name and province).

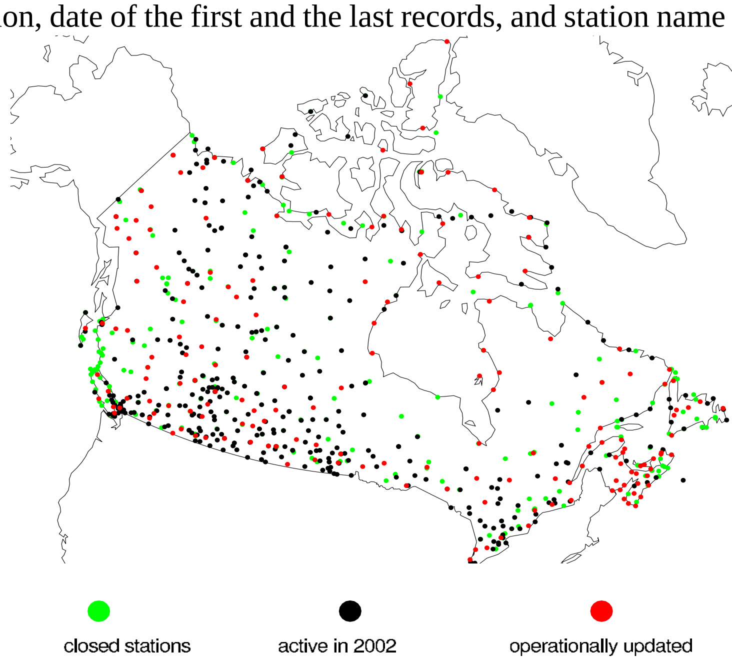

Figure 1. Map of 768 synoptic stations of Canada with the data available in this archive. Closed stations are marked in green and active and operationally updated stations in red.

2. Element Names and Definitions

2.1. File "**Canadian_hly_stn.list.txt**" has 768 lines**.** Each line contains information about one station and can be read with the help of the following FORTRAN statement:

```
      REAL lat, lon, elev
      INTEGER time_beg, time_end
      CHARACTER*11 STN_ID
      CHARACTER *30 stn_name
      CHARACTER *3 prov
       READ(*,1) STN_id,lat,lon,elev,
     *   time_beg,time_end,stn_name,prov
1     format(a11,6x,f5.2,1x,f7.2, 1x,f4.0,2(1x,i10),1x,a30,a2,a3)
```

Where

STN_ID is the station ID composed from "403" – characters that identify Canada in the coding system accepted in the Global Daily Climatology Network (DSI-9101, Appendix A), one reserved character "0" and 7-character alpha-numeric code for station accepted by the MSC. In this 7-character alpha-numeric coding system the first digit is associated with specific Canadian territory and/or province: 1 - for British Columbia; 2 - for Yukon and North-Western Territories (currently for Nunavut also); 3 - for Alberta, 4 - for Saskatchewan, 5 - for Manitoba; 6 - for Ontario, 7 - for Quebec, and 8 - for Maritime Provinces (these were further separated by the second digit; 81 - stands for New Brunswick, 82 - for Nova Scotia; 83 - for Prince Edward Island, 84 - for Newfoundland Island, and 85 - for the continental part of Newfoundland Province).

lat is the station latitude (degrees, N). It is provided with an accuracy to hundredths of a degree. Range from 41.95 to 82.52.

lon is the station longitude (degrees, W is negative). It is provided with an accuracy to hundredths of a degree. Range from -140.87 to -52.75.

elev is the station elevation in meters. Range from 0 to 2543.

time_beg and **time_end** provide year, month, day, and hour with the first and the last data record respectively available for the station in the archive. Note that these dates may not indicate the starting/ending points of hourly observations at the station. Many digital records start in January 1953 which was the beginning of the digital archive but not the actual start of the observations at a given location. Also, we stopped the near-real time update of the archive on February 21, 2005 and the large-scale update at the end of December 2002. However this does not mean termination of observations at the stations but simply the mechanical termination of infilling the current version of the data set.

stn_name provides the name of the station used in the national Canadian Catalogue and

prov a two to three letter abbreviation used for the province and/or territory.

2. 2. Each file `STN_ID`.dat contains the entire data available for STN_ID station.

These files were created using the following FORTRAN write statement:

```
      CHARACTER*1 c37
      INTEGER US_weather_codes(24),cldtyp(4), cloud_layer_elev(4),wea(24)
      INTEGER year, month, day ,hour, minute, slp, pres, temp, vp, rh, wndspd, wndir,
      op_cc, t_cc, locc, num_cld_typ, num_wx, opaque_cc, total_cc, locc
       .........
         WRITE (*,80) year, month, day, hour, minute, slp, pres, temp, vp, rh, wndspd,
      wndir, op_cc, t_cc, c37, locc, num_cld_typ, num_wx,
        * (cldtyp(i), i=1,num_cld_typ),(wea(j),j=1,num_wx),
        * (US_weather_codes(k),k=1,nwea),(cloud_layer_elev(j),j=1, num_cld_typ)
80    FORMAT (i4.4, 4i2.2, 1x, 2i5, 2i4, i3, 2i4, 2i3, 1x, a1 ,i3, i2, i3, 45i3.2)
```

The same record is described below in detail:

Variable	Position in the record	Meaning
Year	1 - 4	Year of the observation
month	5 - 6	Month of the observation
day	7 - 8	Day of the observation
hour	9 - 10	Hour of the observation
minute	11- 12	Minute of the observation [in Canada, there is a region with a non-standard time zone shifted by 30 minutes]
slp	14 - 18	Sea Level Pressure [tenths of mB]
pres	19 - 23	Station Pressure [tenths of mB]
temp	24- 27	Dry Bulb Temperature [tenths of C]
vp	28- 31	Vapor Pressure [tenths of mB]
rh	32- 34	Relative Humidity [percent]
wndspd	35- 38	Wind Speed [tenths of m/sec]
wndir	40- 42	Wind Direction [degrees from N clockwise]
op_cc	43- 45	Opaque cloud cover [percent]
t_cc	46- 48	Total cloud cover [percent]
c37	50- 50	Character code indicating which cloud element was used for cloud information[1]
locc	51- 53	Low cloud cover [percent]
num_cld_typ	55- 55	Number of cloud types to follow (cloud types, **cld_typ**, are coded according to the WMO coding rules,

[1] This code is used for the U.S. cloud cover data stream and has the following meanings:
blank = TSKC but CLC and/or CLM information is available
1= TSCE from octas (TSKC is not available)
2 = CLCx
3 = ALCx
4= CLMx
5 = ALMx
6 = C2C3 is used to calculate low cloudiness
7 = summation of layer information is used to calculate low cloudiness
8 = TSKC only is available (no ALC,ALM,CLC,CLM information)
9 = total cloudiness is calculated by summation of layer information
(see documentation for TD3240 for description of abbreviations). So far, for Canada, **c37** = ' ' or '8'.

```
                                             cf., DSI-3505)
num_wx                   57-  58          Number of weather codes to follow
                                           (present weather types, wea, are
                                          coded according to WMO coding rules)

Variable part of the record:
cld_typ               two digit numbers    Cloud types
                                             [cld_typ(i), i=1,num_cld_typ]
wea                   two digit numbers    Weather codes [wea(j), j=1, num_wx]

US_weather_codes      two digit numbers    See footnote[2].
cloud_layer_elev      two digit numbers    Cloud heights (in 100s of meters)
                                            are presented if possible after the
                                            last group of data.
```

Note about missing values. For positive values "–1" is the missing value code, for temperature, 999 in the field is used for the missing value code. See item 9.3 for missing information in **num_wx** and **num_cld_typ .** In practice, there were rare situations when the variable part of record was long. However, the restrictions used in array sizes for **cld_typ** and **wea** as well as in the FORMAT statements for the variable length portion of the records was never exceeded.

2.3. Each file STN_ID.flg contains the entire flag data available for STN_ID station.
This file was generated concurrently with the **STN_ID.dat** file and contains the set of four alpha-numeric flags for each variable in the data set:
slp_flag, pres_flag, temp_flag, vp_flag, rh_flag, wndspd_flag, wndir_flag,
op_cc_flag, t_cc_flag, c37_flag, locc_flag, (cldtyp_flag(i), i=1,num_cld_typ),
(wea_flag(j), j=1,num_wx), (US_weather_codes_flag(k), k=1,num_wx),
(cloud_layer_elev_flag(l), l=1,num_cld_typ).
Each variable in the file type **STN_ID.dat** containing the meteorological information is accompanied with four-character flag information.

These files were created using the following write statement:

```
20      FORMAT(i4.4, 4I2.2, 1X, 11A4, I2, I3,45A4)
        WRITE(*,20) year, month, day, hour, minute, slp_flag, pres_flag,
   +    temp_flag, vp_flag, rh_flag, wndspd_flag, wndir_flag,
   +    op_cc_flag, t_cc_flag, c37_flag, locc_flag,
   +    num_cld_typ, num_wx, (cldtyp_flag(i), i=1,num_cld_typ),
   +    (wea_flag(j), j=1,num_wx), (US_weather_codes_flag(k),
   +    k=1,num_wx), (cloud_layer_elev_flag(l), l=1,num_cld_typ)
```

Each variable has a CHARACTER*4 description representing a string of one-character flags: [flag 1, flag 2, flag 3, flag 4] are associated with:

- Data measurement (flag 1),

[2] For the United States and Canada it was impossible to accurately report the original weather codes used in TD3280 and by Canadian Meteorological Service. Thus the original US weather codes, [**US_weather_codes(i),i=1,num_wx]** (i.e., a conversion of the Canadian weather codes to the **U.S. weather codes** that are compatible) followed the record. Use documentation for Surface Airways Hourly TD3280 for the US weather codes (**DSI 3280**).

- Quality control (flag 2),
- Confidence level/status (flag 3), and
- Data source (flag 4).

The individual flags can have the following values:

Flag 1: Data measurement flag
blank = measured value
E = estimated value
D = derived value
U = suspect
R = Dew point (used to estimate water vapor pressure) and/or relative humidity, originally calculated with respect to ice have been recomputed with respect to water
G = overcast skies with gaps

Flag 2: Quality control flag
blank = unknown
N = normal QC by provider of data
1 = data were subjected to GSHN version 1.0 QC procedures (replaces "N" but does not replace the "H" flag)
H = instrumental homogeneity adjusted.

Flag 3:Confidence level/status flag
0 = observed data has passed all original system checks
blank = unknown
B = value failed QC checks
C = scale corrected
D = derived value
E = edited value passed all original checks
H = homologous value, rigorously tested
I = interpolated value, not verified
M = missing value
N = not tested but within observed climatological boundaries
Q = questionable (actually wrong)
R = record-breaking value
S = Suspect value (outside climatological boundaries, not verified)
T = tested value, manually checked but not perfectly homologous
U = value suspect
These flags are a result of GSCN version 1.0 (and next steps) quality control routines and replace "blank" and "N" flags when needed.
X - exceeds known world extreme or impossible value
O - Outlier **.GE**. 6 bi-weight standard deviations from the bi-weighted mean values[3]

[3] See [Lanzante 1996] and DSI-9101 for definitions and implementation of QC technique that use bi-weighted

5 - Outlier **.GE.** 5 bi-weight standard deviations from the bi-weighted mean values
4 - Outlier **.GE**. 4 bi-weight standard deviations from the bi-weighted mean values
3 - Outlier **.GE**. 3 bi-weight standard deviations from the bi-weighted mean values
K - Value occurs 10 or more days in a row (0.0 cloudiness characteristics excluded)

Flag 4: Data source flag
blank = unknown
G = GTS
N = national meteorological service
P = preliminary national data (e.g., operational update)

3. Start Date: 19530101

4. Stop Date: 20050221

5. Coverage:

a. Southernmost Latitude: 41.95 N
b. Northernmost Latitude: 82.52 N
c. Westernmost Longitude: 140.87 W
d. Easternmost Longitude: 52.75 W

6. How to Order Data:

Contact NCDC's Climate Services about costs of obtaining this dataset.

Phone 828-271-4800
Fax 828-271-4876
e-mail NCDC.Orders@noaa.gov

7. Archiving Data Center:

Name: National Climatic Data Center/NCDC
Address: Federal Building
151 Patton Ave.
Asheville, NC 28801-5001

Phone: 828-271-4800
Fax 828-271-4876
e-mail NCDC.Orders@noaa.gov

---

statistics.

8. Technical Contact:

Name: Pavel Ya. Groisman
Address: National Climatic Data Center
Federal Building
151 Patton Ave.
Asheville, NC 28801-5001

Voice Telephone: 828-271-4347
Facsimile Telephone: 828-271-4328
Electronic Mail Address: Pasha.Groisman@noaa.gov

9. Known Uncorrected Problems

9.1. The data represent a transformation of original Canadian observational data at synoptic stations into a format accepted for GSCN. Some conversions (such as weather code conversions and cloud type information) were imposed to make the data set comparable with similar information worldwide (by the use of the WMO codes) and/or across North America (by the use of the cloud code information similar to that used in the United States).
9.2. Historically, the humidity information was reported by different instruments that do not coincide with variables used in this data set (water vapor pressure and relative humidity). The Magnus formula was used for conversion. This explains the numerous occurrences of the "D" flag signifying that those humidity data are derived from other humidity elements.
9.3. It is impossible to distinguish the situations when no weather was reported versus the occasions when no data were available; thus 0 reported for **num_wx** cannot be depended upon to indicate clear weather. The same is true for **num_cld_typ** and the cloud type information for Canada.
9.4. Several changes in observational practice have occurred throughout the network that could affect homogeneity and/or data availability of cloudiness and humidity information. For example, automation effectively eliminated cloud type information. The replacement of the humidity sensor (dewcell sensor instead of psychrometer in the early 1970s) affected homogeneity (a stepwise decrease in the records) of winter season measurements of relative humidity throughout the network in the north of the country (van Wijngaarten and Vincent 2004).

10. Quality Statement

The general philosophy used in the quality control of these data is to create a data set which will cause "no harm" to the summary statistics. For example, if daily data for a station contains a temperature value of 100° C, the summary statistics will be contaminated by the large outlier. On the other hand, if all values, say > 40 ° C are flagged as unacceptable values, that can be equally bad. The idea is to try to find some middle ground. In general one can take the time to look at individual values which are questionable when the data set has a temporal resolution of a day or greater. However, this luxury is not available when dealing with hourly data. During the QC development phase, the flag parameters must be carefully adjusted so that the QC programs can run in a completely objective mode. The other part of the philosophy is that the QC process never changes data, only flags. Therefore, if a user wishes to use every bit of data regardless of possible outliers, the flags can be ignored.

When possible, depending on the element to be checked, there are two classes of inspections; a set of within-station and another of among-station checks.

A) Within station

1) Range checks based on a single near-record value. For example, the highest sea-level pressure ever recorded was1083.8mb (12/31/68) in Siberia. Therefore this value could be used for the extreme-value flag parameter for this element.
2) Range checks based on the mean and standard deviation derived from monthly statistics from the station itself. A slight trim is used on each end of the sorted vector to eliminate possible outliers. The trim insures that there will always be a certain portion of all values that will undergo further scrutiny.
3) The data which are chosen to undergo further scrutiny are also subjected to a second-difference test which is an extremely simple but yet powerful means to detect “spikes” in a time series caused by an outlier. The suspect values are then passed to the next series of tests for further checking.

B) Among stations

1) A file containing each station ID accompanied by up to 30 nearby station IDs is created along with the distance between the target station and each neighbor.
2) Data that were flagged during the within station checks are placed in an array with 4 hrs of data before and after the suspect value. The other lines are the same but for nearby neighbors. The station number and the distance from the station to be checked follow. There can be 30 nearby neighbors, though the list is seldom large and is additionally restricted to some maximum distance (generally, 200 km).
3) A matrix of absolute differences from the "target" 4hrs before and after is produced. The consistency check consists of a series of comparisons between various rows and columns. If consistency is found within the ranges of the threshold parameters (C1-C3) and the target, the suspect value passes the QC test and is not flagged. The threshold parameters are adjusted for each meteorological element individually. An example of the rules follows.
    a) If any absolute difference on line 1 < C1, then pass.
    b) If any absolute difference on neighbor’s same hour (i.e. middle column) < C2 then pass.
    c) If any value NOT on line 1 > than target, then pass.
    d) If any absolute difference off target row and column < C3 then pass.

on a particular system. NOAA makes no warranty, expressed or implied, nor does the fact of distribution constitute such a warranty.

11. Essential Companion Datasets

Data description and additional information contained in these data sets can be helpful when working with this archive:

- TD-3505 Integrated Surface Hourly Database(ISH); Data Set 3505, DSI-3505 available at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3505.pdf
- TD-9956 DATASAV3 Global Surface Hourly Data; Data Set 9956, DSI-9956 available at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td9956.pdf
- TD-3280 (Data Set 3280, DSI-3280, Surface Airways Hourly; available at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3280.pdf
- TD-9101 (Data set 9101, DSI-9101, Global Daily Climatology Network available at http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td9101.pdf (see also http://www.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html)

See also the MSC web-site http://www.climate.weatheroffice.ec.gc.ca/climateData/canada_e.html.

12. References

a. Lanzante, J.R., 1996: Resistant, robust and nonparametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197-1226.

b. Milewska, E.J. 2005: Trends in cloud types and diurnal changes in cloudiness in Canada 1953-2003. *Annual AMS Meeting*, San Diego, California, 19th Conf. on Hydrology. 9-13 January 2005. P3.4 (CD ROM).

c. Sun, B. and P.Ya. Groisman, 2004: Variations in low cloud cover over the United States during the second half of the 20th century. *J. Climate*, **17**, 1883-1888.

d. Sun, B., Groisman, P.Ya., and I. I. Mokhov, 2001: Recent Changes in Cloud Type Frequency and Inferred Increases in Convection over the United States and the Former USSR. *J. Climate*, **14**, 1864-1880.

e. Sun, B., P. Ya. Groisman, R. S. Bradley, and F. T. Keimig, 2000: Temporal Changes in the Observed Relationship between Cloud Cover and Surface Air Temperature, *J. Climate*, **13**, 4341-4357.

f. van Wijngaarden, W. A. and L. A. Vincent, 2004: Trends in Relative Humidity in Canada from 1953-2003. *Annual AMS Meeting*, Seattle, Washington, 11-15 January, 2004 JP2.3 (CD ROM).