# NCDC Air Temperature (Degrees C)

Examples:

20022416,26631,701040,12/31/2010 23:00:00,FM-15,4,68.883,-166.133,-28.0,1

20022416,26631,701040,12/31/2010 23:00:00,SAO,4,68.883,-166.133,-27.8,1

0022376,26616,701330,04/18/1973 21:00:00,SAO,5,99999,999999,-18.9,5

20022376,26616,701330,04/18/1973 21:00:00,SAOSP,4,66.883,-162.6,-18.8,1

20022376,26616,701330,04/02/1976 18:00:00,SAO,2,99999,999999,-15.0,5

20022376,26616,701330,04/02/1976 18:00:00,SY-SA,1,66.867,-162.633,-14.9,1

20022376,26616,701330,04/13/1997 20:54:00,FM-15,3,66.867,-162.633,-8.3,5

20022376,26616,701330,04/13/1997 20:54:00,FM-16,4,66.867,-162.633,-9.0,1

20022376,26616,701330,01/03/1999 00:53:00,FM-15,3,66.883,-162.6,-15.6,5

20022376,26616,701330,01/03/1999 00:53:00,FM-16,4,66.87,-162.63,-16.0,1

**Quality Code:**

5: Passed all quality control checks, data originate from an NCDC data source

1: Passed all quality control checks

4: Passed gross limits check, data originate from an NCDC data source

0: Passed gross limits check

C: Temperature and dew point received from Automated Weather Observing

Systems (AWOS) are reported in whole

degrees Celsius.  Automated QC flags these values, but they are accepted

as valid.

9: missing

6: Suspect, data originate from an NCDC data source

2: Suspect

7: Erroneous, data originate from an NCDC data source

3: Erroneous

**Duplicate Timestamps Situations**

The data values are the same, but have different data quality codes.

One data value is missing, the other data value is reported.

The two data values are different, with the same data quality code.

The two data values are different, with different data quality codes

**Processing:**

Remove duplicate lines from master file.

awk '!x[$0]++' filename

Create station list:

cut -d, -f1,1 02\_dewpoint\_unique.txt | sort -u > stations.txt

Split file into separate text files for each station.

for ids in `cat stations.txt`

do

grep $ids 02\_dewpoint\_unique.txt > $ids.txt

done

Order of selection, done using data quality codes, if timestamps are duplicates (this is done with remove\_dup2.awk): 5,1,4,0,9,C,6,2,7,3

grep -v -E '^([^,]\*,){5}[12]' $ids.txt | sort -t, -k4.7,4.10 -k4.1,4.2

-k4.4,4.5 -k4.12,4.19 | awk -f remove\_dup2.awk > ${ids}\_sorted\_unique.txt

If there is a duplicate timestamp:

If the timestamps are the same, select the row with the quality control code of 5.

If one of the data values is missing, select the one with a data value.

If the data values are different, select the data value that contains the quality code of 5. If the data quality codes are the same, it will select the last data value that was read in.

The data values are different at the same timestamp, depending on the data source and the report code (the type of geophysical surface observation). Some of the reports include the SPECI values and others do not.

# Wind Direction and Wind Speed

Wind direction contained a wind-observation type code, but only the codes 9 (missing) and N (normal) were found within the file.

# Ceiling

Did not load ceiling determination code or the CAVOK code.

Visibility not loaded.

# Precipitation

Only loaded those values that were collected hourly. 24 stations had precipitation data.

SNOW ACCUMULATION (17\_AL): There were no hourly snow accumulation records (17\_AL).

RELATIVE HUMIDITY (30\_CH): There were no hourly relative humidity records. Only one station available and it was every 5 minutes.

HOURLY TEMP: Did not load average air temp , min air temp and max air temp that failed all checks (The code that indicates ISD's evaluation of the quality status of the hourly temperature average.

) I noticed that the average air temps that failed the ISD’s quality status check had a code that was assigned to it that was not a ‘9’ or a ‘0’. They had codes of ‘2’ or ‘4’ for the CRN Temp\_avg\_flag quality code. I used ‘999.9’ as the missing marker for Avg Temp, since I did a check on all AVG\_TEMP\_QC codes for ‘9’ and they were all ‘999.9’. I believe ‘999.9’ was generated by the Python script.

54\_KB: Contained no hourly air temperatures. It was all averaged out over an hour.

GHCN Processing:

For GHCN, if the quality code indicated failure, I deleted the data value and set the QualifierID to ‘17’, for missing data value.

Remove duplicate lines from master file.

awk '!x[$0]++' filename

Create station list:

cut -d, -f1,1 02\_dewpoint\_unique.txt | sort -u > stations.txt

Split file into separate text files for each station.

for ids in `cat stations.txt`

do

grep $ids 02\_dewpoint\_unique.txt > $ids.txt

done

To create station list:

for stations in `cat stations.txt`;

do

grep ^$stations, station\_all.txt >> list.txt;

done

station\_all.txt is a list of all NCDC stations that are in IARCOD. To add the correct description to the station, use:

sed -E -e 's/,[^,]\*$/&\_GHCN\_SNWD/' list.txt > station\_list.txt

SNOTEL:

Inserted missing marker ’-99.9’ where values were missing and a date reported.

# Snow Course:

* Inserted missing marker ‘-99.9’ where values were missing and a date reported.
* If E/ST was used, I replaced the date with either MM/01, if the card was 1 or MM/15, if the card was 2. I did not replace an estimated date, if the card was 3. However, it appeared that all cards were of type 1 with dates that needed to be replaced.

# USGS\_GTN-P:

AK104\_hourly\_ave.txt: Removed duplicate timestamp ‘3/15/11 10:00’. All data values were duplicates

# USGS\_BLM:

Converted files to AKST, so that we wouldn’t lose an hour of data.

For Precip, there are quite a few values that look totally wrong. I didn’t load Precip Cumulative.

# Ivotuk/WERC

Ivotuk Moss Met

1999: Insert missing markers ‘6999’

* in date range 7/7/99 9:00 to 7/7/99 12:00
* date ‘7/8/99 11:00’
* 8/15/99 16:00

2003: Deleted 1st occurrence of timestamp ‘2/28/2003 12:00’.

2005: Deleted 2nd occurrence of timestamp range ‘6/19/2005 00:00’ to ‘6/20/2005 19:00’

Ivotuk Moss Radiation:

2000:

* Deleted duplicate timestamp ‘3/21/00 10:00’ and ‘9/23/00 8:00’.
* Deleted 2nd occurrence of date range ‘12/31/2000 1:00’ to ‘12/31/2000 23:00’.

2002: Replaced ‘#N/A’ with missing marker ‘6999’

2004: Deleted 2nd occurrence of duplicate date range ‘4/30/2004 04:00’ to 5/1/2004 11:00’

2005: Deleted 2nd duplicate date range ‘6/19/2005 00:00’ to ‘6/20/2005 29:00’

Ivotuk Moss Snow:

2004: Inserted missing marker ‘6999’ in date range ‘10/26/2004 16:00’ to ‘1/1/2005 0:00’.

2005:

* Inserted missing marker ‘6999’ for timestamp ‘7/29/2005 13:00’
* Deleted 2nd occurrence of date range ‘4/30/2004 4:00’ to ‘5/1/2004 11:00’
* Deleted 2nd occurrence of date range ‘6/19/2005 00:00’ to 6/20/2005 19:00’

Ivotuk Moss Soils:

2002: Replaced #NUM! with missing marker ‘6999’ 39 times.

2006: Deleted duplicate timestamp ‘6/20/2005 18:00’.

Ivotuk Shrub Met

2000: Deleted duplicate row for timestamp ‘3/21/00 13:00’

2001: Removed 2nd duplicate date range ‘9/10/01 8PM’ to ‘9/22/01 4PM’

2005: Removed 2nd duplicate date range ‘6/19/2005 0:00’ to ‘6/20/2005 20:00’

Ivotuk Shrub Rad:

2000: Deleted duplicate timestamp ‘3/21/2000 13:00’

Ivotuk Shrub Snow:

2000: Deleted duplicate timestamp ‘3/21/2000 13:00’

Ivotuk Shrub Soil:

2004: Deleted duplicate timestamp ‘11/4/04 12:00’

# TOOLIK LTER

Reference this URL for hourly data definitions: <http://ecosystems.mbl.edu/arc/meta_template.php?FileName=./weather/tl/88dltlh.html>

Metadata3.csv:

* Deleted 2nd duplicate timestamp for ‘6/7/1988 0:00’ (aka 6/6/1988 24:00). Note, this had differing values.
* Deleted 2nd duplicate timestamps for ‘6/7/99 3:00’ to ‘6/7/99 3:00’.
* Deleted first duplicate for timestamp ‘1/1/2001 0:00’ (aka 12/31/2000 24:00)
* Deleted 2nd duplicate timestamps for ‘6/5/06 3:00’ to ‘6/5/06 15:00’