**IMIQ Data Processing**

Processing notes, scripts, source documentation

International Arctic Research Center, University of Alaska Fairbanks

**Abstract**

This document contains the data processing notes for the data that has been loaded into the IMIQ database. Amy Jacobs, Rob Cermak and Greta Burkart provided the data processing notes.

NCDC 5

GHCN 5

Initial GHCN Processing 5

Secondary GHCN Processing 7

ISH 8

Initial ISH Processing 8

Secondary ISH Processing 9

WERC Overview 12

WERC: Kuparuk River Watershed Studies 12

Coastal Plain 12

Betty Wetland Met: 12

Betty Radiation: 12

Betty Pingo Soils Wetland (Greta Myerchin's noted changes): 13

Betty Pingo Soils Wetland (New changes): 13

Greta Myerchin's changes that have been fixed on website: 13

Betty Pingo Soils Upland (Greta Myerchin’s noted changes): 13

Betty Pingo Soils Upland (New Changes) 13

Greta Myerchin’s changes that have been fixed on website: 13

Putuligayuk River Gage Discharge 13

West Dock Met 14

West Dock Radiation 14

Imnavait Watershed Sites 14

Imnavait Basin Met 14

Imnavait Basin Radiation 16

Imnavait Basin Snow 17

Imnavait Basin Soils 17

Imnavait Basin Site A Met 18

Imnavait Basin Site A Radiation 19

Imnavait Flume Discharge 19

Imnavait Flume Snow Depth 20

Upper Kuparuk Watershed Sites 20

North Headwater Met 20

Green Cabin Lake Met 20

East Headwater Met 21

Upper Headwater Met 21

West Headwater Met 21

Upper Kuparuk River Discharge 21

Upper Kuparuk Met 21

Upper Kuparuk Radiation 22

Upper Kuparuk Snow 22

Upper Kuparuk Soils 22

Western Kuparuk Sites 22

West Kuparuk Met 22

West Kuparuk Radiation 23

West Kuparuk Soils 23

Eastern Kuparuk Sites 24

Sagwon Bluffs Met 24

Sagwon Bluffs Radiation: 24

Sagwon Bluffs Snow: 25

Sagwon Bluffs Soils: 25

Lower Kuparuk Met 25

Umiat Corridor Project 25

Upper May Creek Soils 25

Itikmalakpak Soils 25

Anaktuvuk Soils 26

Bullen Hydrology Project 26

Ribdon Soils 26

Juniper Soils 26

Sag-Ivishak Soil 26

Upper Kadleroshilik Met Soils 27

Bullen Met Soils 27

Kadleroshilik Soils 27

Shaviovik Soils 27

Ivotuk Site 27

Ivotuk Moss Met 27

Ivotuk Moss Radiation 27

Ivotuk Moss Snow 28

Ivotuk Moss Soils 28

Ivotuk Shrub Met 28

Ivotuk Shrub Rad 28

Ivotuk Shrub Snow 28

Ivotuk Shrub Soil 28

North Slope Lake Chemistry 28

NSL 05\_12 WChem 29

NSL 06-01 WChem 29

NSL 06-02 WChem 29

NSL 06-03 WChem 30

Snow Course: 31

USGS\_GTN-P: 31

USGS\_BLM: 31

TOOLIK LTER 31

Steuffer, SourceID = 33 32

Trawicki, SourceID = 48 32

USGS NWIS, SourceID= 139,199 32

Lyons and Trawicki, SourceID=154 33

Whitman and Arp, SourceID=164 33

March, SourceID=179,180,181,182,183,131 34

Appendix A: GHCN-Daily, Version 2.80 Documentation 35

Appendix B: Formula for converting decimal dates 47

Appendix C: ISH Processing files 48

station\_all.txt 48

‘remove\_dup2.awk’ 52

Appendix D: GHCN-Daily scripts 55

create\_unique.sh 55

error.awk 55

remove\_dup2.awk 55

# NCDC

The Integrated Surface Hourly (ISH) documentation provided by the National Climatic Data Center can be found here, dated February 2, 2011. : [http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3505.pdf](http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td3505.pdf#The%20first%20part%20of%20the%20ISH%20documentation%20included.%20%20For%20the%20full%20document,%20please%20go%20to:%20%20http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddo)

The Global Historical Climatology Network (GHCN) Daily, Version 2.80, was used. The GHCN-Daily documentation is contained in Appendix A.

## GHCN

Rob Cermak did the initial GHCN processing. His notes are in the ‘Initial GHCN Processing’ section.

The scripts for the initial processing can be found at: <http://irdp.arsc.edu/data/>. The directory ‘daily’ is the ‘GHCN-Daily’.

### Initial GHCN Processing

#### Step 1: Finalize data inventory mapping GHCN IDs to SiteCodes.

For the GHCN, there are two files that contain station identification and station history:

(1) identification: <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt>

(2) history: <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-inventory.txt>

A snapshot of the Google Spreadsheet was taken at 2 PM AKDT on 10/12/2011. The file name was GHCN\_stations\_V3.csv. The creation of the spreadsheet can be partially completed using a script, validating WMO IDs, SiteCodes and dates with the ISH data inventory. The bulk of GHCN stations without WMO IDs had to be completed through manual identification.

#### Step 2: Extract the data from the compressed archive of all GHCN data.

A snapshot of the GHCN data was downloaded from NCDC on 08/31/2011 at 10:44 am AKDT.

Source: <http://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd_all.tar.gz>

The downloaded data file should be placed in a directory called “data”. Please create a sub-directory called “ghcnd\_all” in “data”.

Script: [extract\_from\_archive.py](http://extract_from_archive.py)

This script will extract requested data from ghcnd\_all.tar.gz and place it in data/ghcnd\_all. The files will be compressed and be of the form: GHCNID.dly.gz

#### Step 3: Generate available fields from extracted data.

Script: [gen\_fields.py](http://gen_fields.py)

#### Step 4: Extract GHCN data and associate with appropriate SiteCode.

Script: [rd\_data.py](http://rd_data.py)

The only stations that require a lookup of the GHCN\_ID by date are those that have multiple SiteCode’s assigned to them in GHCN\_stations\_V3.csv.

Data is written to data/db/PARAMETER.csv

The general format for each file is:

SiteCode,date,value,MFLAG,QFLAG,SFLAG

20028931,2004/01/02,-5.2, , ,G

Values that mentioned units in 'tenths' were adjusted.

All metadata is in the GHCN readme.txt file especially for MFLAG, QFLAG, SFLAG.

Data is packaged up using pkzip:

zip daily\_db.zip db/\*.csv

### Secondary 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. This was done, since we did not want a known bad data value to make its way to the public.

The GHCN data from Rob Cermak contained exact duplicates and all of the station data was in one file. To make it easier to load into the database, I split the data into separate station files.

* 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 (Appendix C) is a list of all NCDC stations that are in IARCOD. To add the correct descriptive name to the station, use:

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

* Run script ‘create\_unique.sh’ (Appendix D)

## ISH

Rob Cermak did the initial ISH processing. His notes are in the ‘Initial GHCN Processing’ section.

The scripts for the initial processing can be found at: <http://irdp.arsc.edu/data/>. The directory ‘hourly’ is the ‘ISH’.

### Initial ISH Processing

Purpose: Using identified NCDC stations decode data for ingestion into database.

#### Hourly (ISH)

#### Step A: get\_data.py

This script downloads the data from the NCDC website according to what is found in ISH\_stations\_to\_process.csv.

This was run on 09/29/2011 at 4:31 pm AKDT.

#### Step B: For each data element or group of elements, please repeat each group of steps. Total: 68 data elements

#### Step 1: Create a metadata template for data elements (e.g. 01\_temperature\_info.txt)

#### Step 2: Run script

./rd\_data\_01\_temperature.py

#### Step 3: Check output

Make sure there are no un-matched ID entries in output data.

grep nomatch 01\_temperature.txt

#### Step 4: Finish and verify metadata for data element

#### 01\_temperature\_info.txt

#### Step 5: Package data for website

#### zip 01\_temperature.zip 01\_temperature\_info.txt 01\_temperature.txt

Move zip file to web server directory and remove data file to conserve space.

### Secondary ISH Processing

The data that was received from Rob Cermak contained duplicate rows. This section describes how the duplicates were removed and how the individual station data files were generated and loaded into the database.

Example of data received from Rob Cermak:

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

**ISH 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**

When dealing with the duplicate timestamps, there were four different scenarios that needed be considered:

* 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 > filename\_unique.txt

* 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

*‘remove\_dup2.awk’ is in Appendix D.*

* 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 a Python script.

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

# WERC Overview

The work data was downloaded from the WERC websites. If there was a problem with a file that was downloaded, then Rob Gieck provided data/information for the Kuparuk/North Slope sites and Emily Youcha provided additional data/information for the Bullen/Umiat sites. Greta Myerchin-Tape also provided additional data/information—especially for the soils data and North Slope Lake Water Chemistry.

The WERC data was found here:

* North Slope: <http://ine.uaf.edu/werc/projects/nsl/nslakes.html>
* Kuparuk: <http://ine.uaf.edu/werc/projects/NorthSlope/northslope.html>
* Bullen: <http://ine.uaf.edu/werc/projects/bullen/bullen.html>
* Umiat: <http://ine.uaf.edu/werc/projects/umiat_corridor/stations.html>

# WERC: Kuparuk River Watershed Studies

## Coastal Plain

### Betty Wetland Met:

* 1997: Deleted repeat day, 12/13/1997, first entry. Checked WERCOD and discovered they had done the same.
* 2005: Had 6999.0 for marker. Updated 873 rows in database to NULL and QualifierID to 14 where the DataValue was ‘6999’
* 2004: Had 6999.0 for marker. Updated 196 rows in database to NULL and QualifierID to 14 for DatastreamID = 1060
* 2006: Inserted missing marker ‘6999’ for data points in the date range: 2/16/06 13:00 to 4/15/06 18:00.

### Betty Radiation:

* 1996: Removed a blank line (line number 7831).  Removed duplicate decimal date 258.63
* 1997: Duplicate 347 decimal date.  Removed first occurrence.  Note:  This decimal date had a duplicate in 1997 BM Met.
* 2004: Missing  reflected shortwave radiation on 6/26/2004 13:00 and 6/26/2004 14:00.  I loaded into IARCOD with the missing marker ‘6999'.
* 2008: Missing Upland NetRad value for entire year.  I loaded into IARCOD with the ‘7777’ marker for the Upland Net Rad.  The reason I did so was that Wetland Net Rad was also missing this value for the entire year and the '7777' marker was used.

### Betty Pingo Soils Wetland (Greta Myerchin's noted changes):

* BPW1999:  Duplicate 281 decimal date.  I removed the decimal date for 281 that contained  ‘6999’.
* BPW2006: No data points in range 2/16/2006 13:00 to 04/15/2006 18:00.  Inserted value ‘6999'.
* BPW2007:  Replaced ‘69999.0’ with ‘6999’ for SoilTemp06.

### Betty Pingo Soils Wetland (New changes):

* BPW1999:  12/31/1999 is missing timestamps from 01:00-23:00.  1/1/2000 0:00 is also missing.  I put in the missing marker ‘6999’ for these timestamps.
* BPW2006:  No data points in range 2/16/2006 13:00 to 4/15/2006 18:00.  Inserted value '6999'.
* BPW2008:  Replaced ‘6999.0’ with ‘6999’ for SoilTemp06.

### Greta Myerchin's changes that have been fixed on website:

BPW2001, BPW2002, BPW2004

### Betty Pingo Soils Upland (Greta Myerchin’s noted changes):

* BPU1994: For 15cm, 35cm, and 60cm data points, inserted ‘6999’ in the range 5/24/1994 19:00 to 5/27/1994 23:00.

### Betty Pingo Soils Upland (New Changes)

#### BPU2003:

Decimal dates 287.58 and 287.63 are missing. Inserted missing marker data point value ‘6999’.

* Decimal dates 338.17,338.21,338.25,338.29,338.33,338.38 are missing. Inserted missing marker data point value ‘6999’.
* BPU2006: Inserted missing marker ‘6999’ into data points for date range: 2/16/06 13:00 to 4/15/06 18:00.

### Greta Myerchin’s changes that have been fixed on website:

BPU2005, BPU2006

### Putuligayuk River Gage Discharge

PR1999q: Removed February 29th timestamps, since 1999 is not a leap year.

#### PR2000q:

* Duplicate date at ‘6/9/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/10/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/11/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/12/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/13/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/14/00 12:00’. Removed 2nd occurrence.
* Duplicate date at ‘6/15/00 22:00’. Removed 2nd occurrence.

### West Dock Met

* WD1997m: Deleted 2nd occurrence of decimal date range ‘325.38’ to ‘326.33’.
* WD1999m: Updated timestamps, starting at ‘11/2/99 23:00’ to ‘11/2/99 00:00’ to ‘11/2/99 11:00’. Before the timestamp was incrementing the day and not the hour.
* WD2002m: Inserted missing marker ‘6999’ for numerous data values, starting at decimal date ’18.96’
* WD2003m: Inserted missing marker ‘6999’ for numerous data values, starting at decimal date ‘205.5’
* WD2005m: Insert missing marker ‘6999’ for numerous data values, starting at decimal date ‘224.71’
* WD2007m:
  + Deleted 2nd occurrence of decimal date range ‘322.625’ to ‘328.458’
  + For decimal date ‘313.875’, updated to the correct timestamp of ‘11/9/07 21:00’

### West Dock Radiation

* WD1997r: Deleted 2nd occurrence of decimal date range ‘325.38’ to ‘326.33’.
* WD1999r: Updated timestamps, starting at ‘11/2/99 23:00’ to ‘11/2/99 00:00’ to ‘11/2/99 11:00’. Before the timestamp was incrementing the day and not the hour.
* WD2002r:
  + Inserted missing marker ‘7777’ for decimal date ‘115.42’
  + Inserted missing marker ‘7777’ for decimal date ‘251.42’
* WD2005r: Inserted missing marker ‘6999’ for decimal dates ‘183.29’ and ‘183.33’
* WD2007r:
  + Deleted 2nd occurrence of decimal date range ‘249.79’ to ‘250.33’
  + Deleted 2nd occurrence of decimal date range ‘263.42’ to ‘263.71’
  + Deleted 2nd occurrence of decimal date range ‘270.63’ to ‘270.88’
  + Deleted 2nd occurrence of decimal date range ‘273.5’ to ‘273.75’
  + Deleted 2nd occurrence of decimal date range ‘274.75’ to ‘275.75’
  + Deleted 2nd occurrence of decimal date range ‘278.63’ to ‘279.17’
  + Deleted 2nd occurrence of decimal date range ‘283.00’ to ‘283.21’
  + Deleted 2nd occurrence of decimal date range ‘283.46’ to ‘283.67’
  + Deleted 2nd occurrence of decimal date range ‘286.33’ to ‘286.50’

## Imnavait Watershed Sites

### Imnavait Basin Met

* IB1986m:
  + Used the IB1986m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Missing decimal date 286.75. Inserted missing marker value ‘6999’ for all variables.
* IB1987m: Changed column name from rainfall (mm) to 1.5m wind speed. This is because WERCOD has loaded in 1.5m wind speed values that match the precip values from IB1987m.
* IB1988m: Missing decimal date 107.46. Inserted missing marker value ‘6999’ for all variables.
* IB1989m:
  + Used the IB1989m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Relative humidity is being shown as being collected at a height of 2m in one location and 1m in another. In WERCOD, the RH sensor height is given as 1m—so this is the RH sensor height that I used.
  + What date did the 1.5m AT/RH/WS move to 1m? I used 4/21/1989 15:00.
* IB1989m fixes. These were wrong on the website, but were corrected in file from Greta M:
  + Decimal date 41.88 changed the 10m wind speed data point value from ‘6999.4’ to ‘6999’.
  + Decimal date 48.96 looks wrong. WERCOD shifted all data values to the left. However it is using the value ‘0.9’ for both 10m and 2m air temp. I put in ‘6999’ for 10m air temp.
  + For decimal date range 111.42 to 111.58 changed the 10 meter wind speed data point value to ‘6999’ from ‘6999.4’.
  + What date did the 1.5m AT/RH/WS move to 1m?
  + Decimal date 138.58 changed the 10m wind speed data point value from ‘6999.4’ to ‘6999’.
  + For decimal date range 153.54 to 153.83 changed the 10 meter wind speed data point value to ‘6999’ from 6999.4’
* IB1990m:
  + Used the IB1990m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Relative humidity is being shown as being collected at a height of 2m in one location and 1m in another. In WERCOD, the RH sensor height is given as 1m—so this is the RH sensor height that I used.
* IB1991m:
  + Used the IB1991m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Relative humidity is being shown as being collected at a height of 2m in one location and 1m in another. In WERCOD, the RH sensor height is given as 1m—so this is the RH sensor height that I used.
  + In IB1991m.dat from Greta Myerchin, the column header at the top of the file indicates that the WS order is WS1, then WS10. However, the column headers right above the data say WS10, WS1. WERCOD entered it as WS10 appearing first, so that is what I did.
  + Only two wind speeds are in IB1991m. In WERCOD, it was WS10 and WS1 that were loaded. WS3 is missing for IB1991.
* IB1992m: Relative humidity is being shown as being collected at a height of 2m in one location and 1m in another. In WERCOD, the RH sensor height is given as 1m—so this is the RH sensor height that I used.
* IB1993m:
  + Used the IB1993m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Deleted rows that came after decimal date 366.
* IB1995:
  + Used the IB1995m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick.
  + Put a comment that there is no 3mAT data between 4/21/1989 15:00 to 1/1/1995 0:00 in datastream for 3mAT in IARCOD. Precip has also started again, but there is a note that it is from Site A.
* IB1999m: Duplicate decimal dates for 150.46 Deleted 2nd occurrence, since that is what WERCOD had done.
* IB2000m: Removed character after precip at timestamp ‘9/14/2000 18:00’.
* IB2001: Used the IB2001m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick
* IB2003: Used the IB2003m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick
* IB2004m: 2mAT starts on 1/1/2000 1:00. There was no 2mAT from 1/1/2000 1:00 to 1/1/2004 0:00. No 2mAT from 1/1/2005 01:00 to 1/1/2006 0:00. Made a note in Datastreams in IARCOD.

NOTE: WERCOD has values for 3mAT, even though it was changed to 2mAT in 1989.

#### IB2006:

* + Used the IB2006m.dat file from Greta Myerchin that was sent to her on 2/22/2010 by Rob Geick. No 10m air temperature or 1m relative humidity.
  + In WERCOD precipitation was not loaded for this year. It also appears that they loaded in a different data file than I received from Greta Myerchin, since the wind speeds have three decimal places. However, there are still no 10m air temperature and 1m relative humidity in WERCOD.
* IB2007:
  + Does not appear to have been loaded into WERCOD for all variables.
  + Loading in IB2007\_old.csv until ‘9/8/2007 19:00’. After this time, will switch to IB2007\_new.csv.
* IB2009: Received copy of file from Rob Geick on 6/21/2011. Rob and Greta Myerchin have reviewed this file. Does not include 3m wind speed.

### Imnavait Basin Radiation

* IB1986r: Missing decimal date 286.75. Inserted missing marker value ‘6999’ for all variables.
* IB1999r:
  + Duplicate decimal dates for 150.46 Selected decimal date that included all values.
  + Starting at decimal date 123.17 there are 759 missing values . I did a search and replace. Replaced all missing values with ‘6999’.
* IB2000r: Missing values for decimal date 257.83. Inserted ‘7777’
* IB2003r:
  + Decimal date 118.63: Inserted ‘7777’ for net radiation.
  + Missing values starting at decimal date ‘331.08’ to ‘331.46’. Inserted ‘7777’.
* IB2002-2008: Loaded in this column order: terrestrial longwave, atmospheric longwave, reflected shortwave, incident shortwave, net radiation. Emailed Rob Geick about order on 6/28/2011.

### Imnavait Basin Snow

Precipitation in cm until 1994, where it switches to mm.

Rob Geick said not to load this data, as it was really not WERC.

### Imnavait Basin Soils

* IB1986t: Column header descriptions do not match. Used the column headers from the top of the file, which said ‘10cm Snow Temp, Surface Temp, 05cm Soil Temp and 10cm Soil Temp’. Checked Greta Myerchin’s standardized soils and this is what she had done.
* IB1989t: Decimal date ‘138.58’ was missing. Inserted ‘6999’ for all values.
* IB1991t: Column headers did not match. Went with the first version in the file, since that is what Greta Myerchin had done with standardized soils.
* IB1996t: Duplicate decimal dates 121-126. Deleted those with missing marker.
* IB1997t: Used the 1/1/1997 data value from IB1997t, not IB1996t. These files contain different values for the same date.
* IB1998t: There was no 10cm snow temp, only two surface temps. One of the surface temps contained ‘6999’ for the entire year, so I only loaded the surface temp that contained values. Greta Myerchin did the same thing with the standardized soils file.

#### IB1999t:

* + Confirmed with Rob Geick that the file from web was correct (since he sent me the IB1999t.dat file on 6/30/2011). The standardized soils file from Greta Myerchin has totally different values.
  + One of the surface temps contained ‘6999’ for the entire year, so I only loaded the surface temp that contained values.
* IB2000-2003t: One of the surface temps contained ‘6999’ for the entire year, so I only loaded the surface temp that contained values. Greta Myerchin did the same thing with the standardized soils file.

#### IB2004t:

* One of the surface temps contained ‘6999’ for the entire year, so I only loaded the surface temp that contained values. Greta Myerchin added a new column ‘Surf Temp 1cm’ for the ‘6999’ values.
* Deleted duplicate row for ‘10/27/04’.
* There is not a soils file for 2005.
* IB2006h:
* Replaced ‘#NUM!’ with ‘6999’ marker
* Did not load last column of ‘6999’ values.
* 2008: Loaded from IB2008th until decimal date ‘254.54’. After that loaded in from file IB2008t2h, starting with decimal date ‘254.58’.

#### IB2008t2h:

* + Replaced missing values with ‘6999’
  + Used headers right above data column, so did not load 15cm soil temp.

### Imnavait Basin Site A Met

* IA1985m: Inserted multiple missing decimal dates and used ‘6999’ for the missing values.
* IA1987m: Inserted missing decimal date ’75.54’ and used ‘6999’ for missing values.

#### IA1988m:

* + Inserted missing values on the midnight hour on 7/28/1988 and 8/1/1988 to 8/28/1988. Used missing value marker ‘6999’
  + Inserted missing decimal date from ‘303.5’ to ‘303.87’ and used ‘6999’ for missing values.

#### IA1989m:

* + Inserted missing marker ‘6999’ for wind speed for decimal dates in range ’42.46’ to ’42.58’.
  + Inserted missing marker ‘6999’ for wind speed for decimal dates in range ‘109.88’ to ’110.67’.
  + Inserted missing marker ‘6999’ for wind speed for decimal dates in range ‘226.46’ to ’228.54’.

#### IA1991m:

* + Inserted missing decimal date ‘122.41’ and used ‘6999’ for missing values.
  + On decimal date ‘122.458’, changed a value to ‘6999’ from ‘6999.007’ and inserted missing values at end of row.
  + Inserted missing decimal date ‘132.5’ with missing marker ‘6999’.
  + Inserted missing decimal dates ‘205.33’ to ‘206.37’ with missing marker ‘6999’.
  + Inserted missing decimal dates ‘242.58’ to ‘243.33’ with missing marker ‘6999’.
  + Inserted missing decimal dates ‘309.62’ to ‘309.66’ with missing marker ‘6999’.
  + For decimal date range ‘235.62’ to ‘236.25’. Moved data values over to the right one column. Inserted missing marker ‘6999’ into 2m relative humidity and updated 1m relative humidity to value before ‘6999’. It appears that the 2m relative humidity ‘6999’ values were appended to 1m relative humidity and messed up column order.
  + For decimal date ‘236.792’ moved values over to the left one column and inserted ‘6999’ for 2m relative humidity. Also updated 1m relative humidity to 96.

#### IA1992m:

* Inserted decimal date ‘119.83’ with missing marker ‘6999’.
* Ends at decimal date ‘244.958’.

### Imnavait Basin Site A Radiation

#### IA1985R:

* + Inserted missing decimal dates in range ‘213.75’ to ‘213.96with the missing value marker ‘6999’.
  + Inserted missing decimal dates in range ‘216 to ‘216.37’ with the missing value marker ‘6999’.
* IA1987R: Inserted missing decimal date ‘75.54’ with the missing marker ‘6999’
* IA1990R: Deleted last two columns. Appear to be terrestrial longwave and atmospheric longwave data values have been inserted twice.

### Imnavait Flume Discharge

* IH1986q: Removed duplicate decimal date ‘206.792’.

#### IH1987q:

* + Removed duplicate decimal date ‘179.25’.
* IH1990q:
  + Removed second decimal date ‘140.583’, since it appeared twice.
* IH1991q:
  + Removed second decimal date ‘129.6’, since it appeared twice.
  + Removed second decimal date ‘129.67’, since it appeared twice.
  + Removed second decimal date ‘129.71’, since it appeared twice.
* IH1993q:
  + Did not load the daily values, since they all had the missing marker value of ‘7777’.
  + Removed second decimal date ‘146.88’, since it appeared twice.
* IH1995q:
  + Removed second decimal date ‘133.75’, since it appeared twice.
  + Removed second decimal date ‘137.63’, since it appeared twice.
  + Removed second decimal date ‘143.46’, since it appeared twice.
  + Removed second decimal date ‘146.38’, since it appeared twice.
* IH1997q:
  + Removed second decimal date ‘146.71’, since it appeared twice.
  + Removed second and third decimal date ‘154.38’, since it appeared three times.
* IH1998q:
  + Removed second decimal date ‘141.38’, since it appeared twice.
* IH1999q:
  + Removed second decimal date ‘150.54’, since it appeared twice.
  + Starting at decimal date ‘219.21’ to decimal date ‘262.5’, there are duplicates for each decimal date. Removed second set of duplicate values.
* IH2001q:
  + Removed second and third decimal date ‘177.88’ rows, since it appeared three times.

General Processing notes for Imnavait Flume Discharge:

* Processed years 1985-2001 as minute data, since it was not exactly hourly data. At various times, there were data points that did not occur on the hour. I tried to process hourly, but could not get one value for the hour. Thought it best to enter it as minute and at a later date create an hourly file.
* Processed 2004 as hourly data.
* Processed 2002-2003, 2005-2008 as minute data, occurring every 15 minutes.

To process the actual discharge, used a grep command: ‘grep -E '^[^,]\*,[^,]\*,[^,]\*,[0-9]' ih2006q.csv > ih2006q\_actual.csv’ for 2005, 2006 and 2008.

### Imnavait Flume Snow Depth

ih2006-2007snow.csv: Deleted 734 rows that were from 2008 from data file. These values all contained the ‘6999’ missing value marker.

## Upper Kuparuk Watershed Sites

### North Headwater Met

* NH2003m: Decimal dates ‘235.92’ to ‘236.88’ have the incorrect Date timestamps. Should be one day later.

### Green Cabin Lake Met

* GL1997m: Removed duplicate entries for 21:00 and 22:00 hour, starting at 227.88.
* GL1998m: Removed duplicate decimal date ‘261.5’, second occurrence.
* GL1999m: Removed duplicate decimal date ‘228.96’, second occurrence.
* GL2000m: Removed duplicate decimal date ‘109.63’, second occurrence.

### East Headwater Met

* EH1997m: Changed second occurrence of decimal date ’33.95’ to ‘337.25’.
* EH1998m: Deleted decimal date ‘0’ that occurred after decimal date ‘106.42’ and after ‘107.42’,
* EH2001m: Inserted missing marker value of ‘6999’ for decimal date range ‘114.67’ to ‘119.21’.
* EH2005m: Inserted missing marker value of ‘6999’, starting at decimal date ‘9.04’ for missing values. Inserted missing marker value of ‘6999’ and ‘7777’ for decimal dates ‘137.92’ to ‘140.54’.

### Upper Headwater Met

* UH1999m:
  + Deleted duplicate decimal dates for ‘201.88’.
  + Deleted duplicate decimal dates for ‘203.21’.
  + Deleted duplicate decimal dates for ‘203.92’.
  + Deleted duplicate decimal dates for ‘204.04’
  + Deleted duplicate decimal dates for ‘204.13’
  + Deleted duplicate decimal dates for ‘204.21’
  + Deleted duplicate decimal dates for ‘238.33’
  + Deleted duplicate decimal dates for ‘239.46’
  + Deleted duplicate decimal dates for ‘239.88’
  + Deleted duplicate decimal dates for ‘240.04’
  + Deleted duplicate decimal dates for ‘204.29’
  + Deleted duplicate decimal dates for ‘240.42’
* UH2001m: Added missing marker ‘7777’ for decimal dates ‘145’ and ‘279.04’
* UH2007m: Removed extra ‘0s’ at end of the rows.

### West Headwater Met

* WH1999m: Inserted missing marker value ‘6999’ for decimal date range ‘175.63’ to ‘175.71’.

### Upper Kuparuk River Discharge

* UK1994q: Deleted duplicate decimal date ‘160.875’, that contained ‘6999’
* UK1996q:
  + Deleted duplicate decimal date ‘175.58’, second occurrence.
  + Deleted duplicate decimal dates for range ‘260’ to ‘261’
* UK2002q: Decimal date ‘216.083’: timestamp not in correct format. Changed to ‘08/04/02 2:00’ from its Excel date number format.

### Upper Kuparuk Met

* UK1995m: Deleted duplicate decimal dates ‘191.75’ and ‘191.79’ that contained ‘7777’ markers. Started loading in AT1, RH1, WS1 values at decimal date ‘191.75’.
* UK2004m: Duplicate timestamp ‘12/13/04 17:00’. Deleted second occurrence.

#### UK2005m:

* + Rob Geick sent a revised UK2005.csv file on 8/29/2011. He also said to understand that the precip data values that contained a ‘0’ when the air temperature was below freezing, should be updated to ‘7777’.
  + Inserted missing marker ‘6999’ for decimal dates ‘245.54’ to ‘245.63’ in the UK2005m\_revised file. And for decimal dates ‘245.79’, ‘245.83’

### Upper Kuparuk Radiation

* UK1997r: Inserted missing marker ‘6999’ for decimal date ‘212.54’.
* UK2001r: Inserted missing marker ‘6999’ starting at decimal date ‘116’, 3151 times.
* UK2004r: Deleted duplicate row with timestamp of ‘12/13/2004 17:00’.
* UK2005r: Inserted missing marker ‘6999’ for timestamps in range ‘6/1/2005 1:00’ to ‘6/1/2005 9:00’. Inserted missing marker ‘6999’ an additional 83 times.
* UK2006r: Inserted missing marker ‘7777’ for ‘--‘ for decimal date range: ‘193.667’ to ‘234.417’.

### Upper Kuparuk Snow

* UK2006-2007snow: Deleted 2nd occurrence of timestamp ’07/15/07 19:00’.

### Upper Kuparuk Soils

* UK1999t: Inserted missing marker ‘6999’ for decimal dates ‘217’ to ‘229’.
* UK2000t: Inserted missing marker ‘6999’ starting at decimal date ’60 to ‘176’ for missing values.
* UK2004h: Inserted missing marker ‘6999’ for date range ‘6/13/2004 19:00’ to

‘6/23/2004 19:00’.

* UK2007h: Deleted 2nd occurrence of decimal date ‘196.79’
* UK2008h: Inserted missing marker ‘6999’ 3,408 times, starting at decimal date ‘136.17’.

## Western Kuparuk Sites

### West Kuparuk Met

* WK1997m:
  + Inserted missing marker ‘6999’, starting at decimal date ‘144.17’, 350 times.
  + Changed timestamp for 2nd occurrence of decimal date ‘274.04’ to ‘274.38’ to 10/2/1997 from 10/1/1997, since it appeared that it should have the decimal date ‘275.04’ to ‘275.38’. This pattern occurred for the rest of the year, so I updated all of the remaining timestamps in the same manner.

#### WK1998m:

* + Inserted missing marker ‘6999’, starting at decimal date ‘109.5’. Missing markers were missing for multiple rows. There should have been 10 ‘6999’ missing markers per row and there were only 9.
  + Used this column order to load in variables: 1m air temp, 3m air temp, 10m air temp, 1m relative humidity, 10m relative humidity, pressure, 1m wind speed, 10m wind speed, wind direction, precipitation.
* WK1999m: Inserted missing marker ‘6999’ for 10m relative humidity for the entire year.
* WK2000m: Inserted missing marker ‘6999’, starting at decimal date ‘126.5’, 3909 times.
* WK2001m: Deleted 1st occurrence of decimal date ‘257.62’, since it contained all missing marker values ‘6999’ and was a repeat of a timestamp that had values.
* WK2004m: Inserted missing marker ‘6999’, starting at decimal date ‘248.21’ numerous times.

### West Kuparuk Radiation

There was no 2001 radiation file for West Kuparuk on the website.

Read in WK1994 and WK1995 in this column order: atmospheric longwave, terrestrial longwave, incident shortwave, reflected shortwave and net radiation.

* WK1997r: Updated the timestamps, starting at 2nd occurrence of 10/1/1997. It appears that the decimal date was incorrect for the rest of the year, causing repeat timestamps to occur.
* WK2002r:
  + Inserted missing marker ‘7777’ for decimal date ‘113.63’.
  + Inserted missing marker ‘6999’, starting at decimal date ‘117.21’, numerous times.
* WK2004r: The values all look off. I loaded them anyways, but perhaps this file should be reviewed.
* WK2008r:
  + Inserted missing marker ‘7777’ for decimal date ‘120.58’
  + Inserted missing marker ‘6999’ for decimal date ‘217.29’
  + Inserted missing marker ‘6999’ for decimal date ‘224.04’

### West Kuparuk Soils

Only loaded daily values for West Kupark, since the hourly values did not contain all of the depths.

* WK1996t:
  + Inserted missing marker ‘6999’ for decimal date range ‘272’ to ‘278’ for missing data values.
  + Inserted missing marker ‘6999’ for decimal date range ‘232’ to ‘239’ for missing data values.
  + Updated timestamp for the 2nd occurrence of timestamp ‘6/5/1997 0:00’ to ‘6/6/1997 0:00’, since 6/6/1997 contained the missing data values and not 6/5 (which had data values).
  + Updated 2nd occurrence of decimal date ‘157’ to ‘158’, since it appeared that the decimal dates where copied from the previous year and included duplicate decimal dates for ‘157’.
* WK1999t: Updated the missing marker ‘6999’ from 14 values per row to 12 values per row. The entire year is missing.

#### WK2004t:

* + Inserted missing marker ‘6999’ for 75cm Soil temp for entire year
  + Inserted missing marker ‘6999’ for decimal dates in range ‘248’ to ‘250’

## Eastern Kuparuk Sites

### Sagwon Bluffs Met

* SH1988m:
  + Duplicate date at ‘5.67’. Removed 2nd occurrence.
  + Duplicate decimal date at ‘129.54’. Removed 2nd occurrence.
* SH1994m: For decimal date ‘161.25’, inserted ‘6999’ for precipitation, since it was missing the value.
* SH1996m: Duplicate decimal date ‘91’. Changed to decimal date ‘90’, since that decimal date was missing. This row contained all ‘6999’ values.
* SH2004m:
  + Column headers do not appear to match column data. E-mailed Rob Geick about this on 7/20/ 2011. Rob sent a revised 2004 met file on 7/20/11.
  + Inserted missing marker ‘6999’ for missing wind direction values for decimal date ‘161.67’ to ‘246.38’.

#### SH2008m:

* + Inserted missing marker ‘6999’ for missing values for 794 instances.

### Sagwon Bluffs Radiation:

* SH1996r: Changed first occurrence of decimal date ‘91’ to ‘90’.

#### SH2004R:

* + Inserted missing marker ‘7777’ for decimal date ‘119.54’ for all data values.
  + Inserted missing marker ‘6999’ for decimal date ‘119.54’ for atmospheric longwave radiation.
  + Inserted missing marker ‘6999’ starting at decimal date ‘122.17’ for missing atmospheric longwave radiation data values.
  + Inserted missing marker ‘7777’ for decimal date ‘245.79’ for all data values.
  + Starting on decimal date ‘119.58’ to ‘143.50’ net radiation value is missing. Inserted missing marker value ‘6999’.
* SH2006R: Inserted missing marker ‘6999’ for atmospheric longwave on decimal date ‘117.5’.

### Sagwon Bluffs Snow:

* SW1992M: Removed decimal dates ‘275’ to ‘366’, since they appear in the 1993 file. All of these values were ‘6999’ in 1992, but had actual values in 1993.

### Sagwon Bluffs Soils:

* SH2004h: Added in missing marker ‘6999’ for missing values.
* SH2005h: Inserted ‘6999’ for ‘#NUM!’ values.
* SH2006h: Inserted ‘6999’ for missing values.
* SH2007h: Inserted ‘6999’ for missing values.
* SH2008h1: Inserted ‘6999’ for missing values.

## Lower Kuparuk Met

LK1994m:

* Decimal date ‘50’ occurred twice. Set 2nd occurrence of decimal date ‘50’ to decimal date ‘51’.
* Set 2nd occurrence of decimal date ‘52’ to ‘53’, since it was incorrect.

## Umiat Corridor Project

DUM7: Snow depth error for 2010

DUM2: 2011 on 2/14 at 4pm, error

DUM4: Loaded as ‘6999’ for 9/29/2009 at 15:00 and 16:00

DUM2: Need to load Soil

DUM3: Need to load Soil

DUS2: Need to load Soil

DUS1: Not found

DUM4: Need to load Soils

DUM5: Need to load Soils

### Upper May Creek Soils

* DUM2\_MayCreek\_Soil\_2009: Inserted missing markers ‘6999’ and ‘7777’ for decimal date ‘254.54’

### Itikmalakpak Soils

* DUM1\_Itikmalakpak\_Soil\_2009.csv:
  + Inserted missing marker ‘6999’ for ‘NAN’
* Deleted 2nd occurrence of timestamp range ‘7/31/2009 11:00’ to ‘7/31/2009 12:00’

### Anaktuvuk Soils

* DUS2\_Anaktuvuk\_Soil\_2009:
  + Inserted missing marker ‘7777’ at decimal date ‘160.5’
  + Inserted missing marker value ‘7777’ at decimal date ‘161.33’
  + Inserted missing marker value ‘7777’ for decimal date range ‘218.38’ to ‘218.46’
  + Deleted 2nd occurrence of timestamp range ‘7/7/2009 12:00’ to ‘8/6/2009 11:00’.

## Bullen Hydrology Project

DBM5: Error in 2001 met file, Errors in 2007,2008

DBM6: Snow depth error 2010 met file

DFM3: 5/16/2007 09:00-19:00 errors

DBM1: Missing WS value in 2008 met file

DBM2: Error in 2008 Soils data

DFM2: Confirm soils end date

DFM2: No Soils loaded past 2007

DBM3: Need to load Soils

### Ribdon Soils

#### DBM2\_Ribdon\_ Soil\_ 2006.csv:

* Deleted 2nd=4th occurrence of decimal date ‘239.54’.
* Deleted 2nd occurrence of decimal date range ‘246.5 to 246.58’
* DBM2\_Ribdon\_Soil\_2008.csv: Starting at decimal date ‘295.38’, inserted missing marker ‘6999’ for all values except for 135cm soil temp, 150cm soil temp, 10cm soil water and 20 cm soil water. For those four variables, I used the missing marker ‘7777’, which is what was used for the rest of the data file.

### Juniper Soils

* DBM3\_Juniper\_Soil\_ 2006.csv: Inserted missing markers ‘6999’ and ‘7777’ for decimal date range: ‘1’ to ‘184.46’.
* DBM3\_Juniper\_Soil\_ 2008.csv: Deleted 2nd occurrence of decimal date range ‘250.5’ to ‘250.83’.

### Sag-Ivishak Soil

* DBM4\_Sag-Ivishak\_Soil\_2008:
  + Inserted missing marker ‘6999’ for decimal date ‘169.38’ for value that was ‘NAN’.
  + Inserted missing marker ‘6999’ for decimal date ‘225.17’ for value that was ‘NAN’
  + Inserted missing marker ‘6999’ for decimal date ‘243.33’ for value that was ‘NAN’

### Upper Kadleroshilik Met Soils

* DBM5\_Upper\_Kad\_Soil\_2007:
  + Inserted missing markers ‘6999’ and ‘7777’ for decimal date range ‘240.63’ to ‘240.75’.
  + Deleted 2nd occurrence of decimal date ‘287’

### Bullen Met Soils

* DBM8\_Bullen\_Soil\_2008: Deleted 2nd occurrence of decimal date ‘166.63’

### Kadleroshilik Soils

* DBS1\_Kadleroshilik\_Soil\_2009: For decimal date ‘247.42’, inserted missing marker value ‘6999’ for the ‘NAN’ value

### Shaviovik Soils

* DBS2\_Shaviovik\_Soil\_2009:
  + Inserted missing marker value ‘6999’ for decimal date ‘247.63’
  + Inserted missing marker value ‘6999’ for decimal date range ‘255.04’ to ‘255.38’

UPPER HEADWATER PRECIP update datavalues

Updated datavalues from 77770 for precip for DatastreamID=4145, Upper\_Headwater\_precip, SiteID=956, VariableID=84

Updated DatastreamID= 5822, West\_Kuparuk\_(WK)\_Met\_precipitation, from ‘6669’ to NULL for the precip values (VariableID = 84) and SiteID=960

## Ivotuk Site

### Ivotuk Moss Met

* 1999: Insert missing markers ‘6999’ in date range ‘7/7/99 9:00’ to ‘7/7/99 12:00’ , and ‘7/8/99 11:00’ to ‘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’

## North Slope Lake Chemistry

No timestamps for data values. I copied them over from the original formatted data tables.

If there were no timestamps for each depth, I used the timestamp for the original formatted data table ONLY when the depths were unique.

### NSL 05\_12 WChem

#### Kuparuk Dead Arm Lake 2 - Shore, KDA2-2:

* + Missing Turbidity values from ‘12/15/05 12:46’ to ‘12/15/05 13:04’
  + Time is recorded in original formatted data tables as 12:36, 12:34, 12:31, 12:39, 12:42, etc. This is the way I entered the values also.

### NSL 06-01 WChem

#### Lake 9817 – Shore:

* + LDO data value 12.9 is assigned to LDO Depth 7, should be LDO Depth 8 at 14:48. LDO Depth 7 at 14:44 should be ‘9999’.

#### Lake 9312 - Raft B:

* + Stable RDO not entered.

#### Lake 9817 - Hole 2

* + WaterTemp (variableid=101) is entered twice. Second time is pH values. Deleted 2nd occurrence.

#### Mine Site B North - North East:

* + First data value (VariableID=101 at 3.5 depth) is given as SiteID=1 and not as SiteID=2. Updated this data value to SiteID=2.

#### Lake 9312 - Raft A:

* + LDO at depth of 5 was entered as ‘15..3’. Updated to ’15.3’

#### Mine Site B South - Center

* + Second 14 depth has a timestamp of ‘1533’ and not ‘1646’

### NSL 06-02 WChem

* Kuparuk Dead Arm Lake 2 - Center, KDA2-1: Updated timestamp for 2nd occurrence of depth 15 and depth 18 to 10:33 and 10:59 (from 10:00).
* Lake 9312 - Raft A: All timestamps are now at 12:56, since I created separate datastreams for Depths 10.5a and 10.5b.
* Lake 9817 - Hole 3: Updated duplicate depth 5 timestamp to ’12:50’ for 2nd occurrence.

#### MS13-NC-CT:

* + Update timestamp for second reading of Depth 9 to ‘2/14/06 10:17’.
  + Updated the MethodID for LDO and LDO Temp to the Hatch method id.
* MSB-SC-CT: Updated timestamp for 2nd occurrence of depth 8 to ’15:18’.

### NSL 06-03 WChem

* MSB-NC-CT: Updated duplicate depth of 13 timestamp to ’14:20’ for 2nd occurrence.

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. ‘MM’ being the month. 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 the data values were duplicates for the timestamps.

# USGS\_BLM:

Converted files to AKST.

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

# 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’

The following are notes written by Greta Burkart for the sources that she entered into the database.

# Steuffer, SourceID = 33

1. FYI: Originally this was sourceID 3 and was later moved to SourceID 33. Many items in this folder have not been updated to SourceID 33.

2. Data Values were entered from reports. There was no information about sampling dates in the 2009 and 2010 reports, so I requested dates from Sveta. Sveta sent snow survey forms. I used the dates in these snow survey forms. In some cases the data values in the form did not match up with snow survey data in the reports. In these cases, I did not add the data because I was not sure which one was correct. I sent Sveta inquiring about discrepancies. Please read this email (it is in the Source Request folder), make it more concise, and follow up with Sveta -- mainly we need to know should we go with use data values in the report or data values in the survey forms? Notice at the top of the 0033\_DataValues.xlsx file there are 15 surveys that still need dates added. These data are not in the database.

3. As you know, there are additional sources listed under sourceID 33 that are associated with WERC. With the exception of sites that are North Slope Lakes snow survey sites, it is unclear whether these sites are duplicates, are associated with a different SourceID or should be left under SourceID 33. Sveta and Rob Gieck should know what these additional sites are. I moved the North SLope Lakes snow survey sites to the North SLope Lakes SourceID.

# Trawicki, SourceID = 48

1. Data from Trawicki et al 1991 has been entered. See processing table for more information on what still needs to be done on entered data (i.e. contact Trawicki for information on sampling dates).

2. Enter additional sites and data (water level, etc) from "Lakes of the 1002 Area AK Tech Report #10 and Criteria 1002 Lakes (Bayha).

3. Add information on water rights -- this can be found in files downloaded from the FWS WRB website.

3. Jess Grunblatt is interested in entering bathymetry data from maps -- this is will be a data rescue effort. Maps can be georeferenced and you can geotag bathymetry values for each point. Then data can be added to the database when Jess is finsished.

# USGS NWIS, SourceID= 139,199

1) All available daily discharge data for all sites in the ALCC boundaries has been added. Some datasets were not available (i.e. data from Hulahula River is not online yet). Some, but not all of these datasets are stored in the 'OriginalDataSource' file. It was easiest to download them from CUAHSI HIS and add them directly to the database. Unfortunately, qualifiers on all but the most recently added datasets did not load into the database. All datasets without qualifiers should be reloaed with qualifiers. Update DataValues so that all datastreams associated with this source have "-9999" when there is no data value -- I have found some that are blank fields.

2) Contact USGS for additional datasets at a higher temporal resolution (i.e. 15 minute). Some of this data is available on a USGS archive site, but I would ask someone in the Alaska office for these datasets.

3) Some precipitation, air temperature, and water temperature datasets have been added. Add additional datasets for stations within the ALCC boundaries.

NOTE -- daily data can be downloaded using CUAHSI HIS or downloaded from USGS waterdata website. You can also use W-Get to automoatically add data to IARCOD.

# Lyons and Trawicki, SourceID=154

1. Data from reports have been added to the database. I have done a quick QAQC, but it would be good to conduct additional QAQC.

2. I have added WaterRights information to the comments section of the sites table. It would be nice to add Water Rights information in a way that it can be visualized. More information about water rights is available on the FWS WRB websites (search for FWS Trawicki WRB).

3. I added information about disturbances to the comments section of the sites table. Can you add this to the incidents table?

# Whitman and Arp, SourceID=164

1. Follow up on email request for additional data (see email and summary in SourceInfoRequest folder).

2. Request additional data from the 2011 field season.

3. Request data from Bear Creek.

Lijedahl, SourceID=169

1. Use watershed area and runoff data to calculate discharge.

2. These data have been QAQC'd and published, so change the QualityControlLevel to reflect that.

2. Contact Anna and add any new data from 2010 and 2011. These data have probably not been published yet; therefore they will need to be listed under a different Datastream.

# March, SourceID=179,180,181,182,183,131

1) See processing table in database

2) Read, interpret, revise email to Jenny March requesting additional information about datasets (see SourceCorrespondence folder)

3) Note, the file DataValuesWithOriginalSourceInfo.xlsx has the source info from Jenny's Master's thesis -- the name of these sheets has the suffix '\_old' at the end of the sheet. The reformatted DataValues are in a new sheet -- these have the suffix \_added at the end of the name. Note not all datavalues in these \_added sheets were added. There were some data values that I could not decide on a variable name for. These datavalues are to the right of the reformatted datavalues that were actually added.

# Appendix A: GHCN-Daily, Version 2.80 Documentation

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

Version 2.80

--------------------------------------------------------------------------------

--------------------------------------------------------------------------------

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 AIRPORT is "USC00026481"); and

- Download the data file (i.e., ".dly" file) that corresponds to this code

(e.g., "USC00026481.dly" has the data for PHOENIX AIRPORT).

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 information.

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

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)

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 precipiation

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)

FMTM = 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 degress 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)

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 multch

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.

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 guest 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

distinquished 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

five 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

K = converted from knots

L = temperature appears to be lagged with respect to reported

hour of observation

O = converted from oktas

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

SFLAG1 is the source flag for the first day of the month. There are

twenty one possible values (including blank):

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

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)

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

X = U.S. First-Order Summary of the Day (NCDC DSI-3210)

z = Uzbekistan update

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

GSNFLAG 73-75 Character

HCNFLAG 77-79 Character

WMOID 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)

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

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.

GSNFLAG 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

HCNFLAG is a flag that indicates whether the station is part of the U.S.

Historical Climatology Network (HCN). There are two possible values:

Blank = non-HCN station

HCN = HCN station

WMOID is the World Meteorological Organization (WMO) number for the

station. If the station has no WMO number, then the field is blank.

--------------------------------------------------------------------------------

--------------------------------------------------------------------------------

V. FORMAT OF "ghcnd-countries.txt"

------------------------------

Variable Columns Type

------------------------------

CODE 1-2 Character

NAME 4-50 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.

--------------------------------------------------------------------------------

--------------------------------------------------------------------------------

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

# Appendix B: Formula for converting decimal dates

1. Reference Cell: =INT(87\*365.255)-1
2. Formula for daily and hourly data: =DATE(YEAR($A$13+B14),MONTH($A$13+B14),DAY($A$13+B14)) + IF(MINUTE($A$13+B14)>29,TIME(HOUR($A$13+B14)+1,0,0),TIME(HOUR($A$13+B14),0,0))
3. Formula for minute data is reference cell plus decimal date.
4. To check for duplicates: =IF(COUNTIF($B$2:$B$190258,B2)>1,"DUPLICATE","UNIQUE")
5. To go by Minutes, add decimal date to Reference cell. Example: $A$3+B9

# Appendix C: ISH Processing files

## station\_all.txt

20028563,2134,AKLAVIK\_20028563

20022418,2136,AKULIK\_20022418

20022446,2137,ALASKA ISLAND\_20022446

20022369,2138,ALATNA\_20022369

20022367,2139,ALLAKAKET\_20022367

20022524,2140,AMBLER\_20022524

30002008,2141,AMBLER AIRPORT\_30002008

20022387,2142,AMBLER WEST\_20022387

20022403,2143,ANAKTUVUK AUTO\_20022403

30002785,2144,ARCTIC VILLAGE\_30002785

20022401,2145,ARCTIC VILLAGE\_20022401

20022404,2146,ATIGUN\_20022404

30003068,2147,ATQASUK EDWARD BURNELL SR MEMORIAL AP\_30003068

20022421,2148,AUFEIS\_20022421

20022420,2149,AWUNA RIVER\_20022420

20028932,2150,BABBAGE RIVER\_20028932

30014577,2151,BARROW 4 ENE\_30014577

20022477,2152,BARROW GMCC\_20022477

20022474,2153,BARROW POINT BARROW\_20022474

20022473,2154,BARROW POINT BARROW\_20022473

20022476,2155,BARROW W POST-W ROGERS AP\_20022476

20022442,2156,BARTER ISLAND BAR M\_20022442

20022441,2157,BARTER ISLAND WSO AP\_20022441

20022378,2158,BETTLES\_20022378

20022377,2159,BETTLES AIRPORT\_20022377

20022379,2160,BETTLES CAA\_20022379

20022412,2161,BETTY LAKE\_20022412

20022582,2162,BPE SOURDOUGH\_20022582

20022436,2163,CAMDEN BAY\_20022436

20022388,2164,CANYON VILLAGE\_20022388

20022419,2165,CAPE BEAUFORT LIZ A\_20022419

20022508,2166,CAPE LISBURNE\_20022508

20022416,2167,CAPE LISBURNE AFS\_20022416

20022470,2168,CAPE SIMPSON POW A\_20022470

20022402,2169,CAPE THOMPSON\_20022402

20022371,2170,CHALKYITSIK\_20022371

20022447,2171,CHALLENGE ISLAND\_20022447

20022393,2172,CHANDALAR LAKE\_20022393

20022400,2173,CHANDALAR SHELF\_20022400

20022394,2174,CHANDALAR SHELF\_20022394

30000993,2175,CHANDALAR SHELF DOT\_30000993

20022413,2176,COBBLESTONE\_20022413

20022383,2177,COLD FOOT\_20022383

20022386,2178,COLD FOOT\_20022386

20022578,2179,COLDFOOT\_20022578

20022389,2180,COLDFOOT CAMP\_20022389

20022397,2181,COLLEEN RIVER\_20022397

30000169,2182,COLVILLE VILLAGE\_30000169

20022380,2183,DAHL CREEK\_20022380

30003070,2184,DEADHORSE ALPINE AIRSTRIP\_30003070

20022444,2185,DEADHORSE AP\_20022444

20022396,2186,DIETRICH\_20022396

20022415,2187,EAGLE CREEK\_20022415

20022468,2188,EAST SIMPSON\_20022468

20022456,2189,ENDEAVOR ISLAND\_20022456

20022443,2190,FLAXMAN ISLAND POW 3\_20022443

30000726,2191,FORT YUKON\_30000726

20022365,2192,FORT YUKON\_20022365

30015570,2193,FORT YUKON #2\_30015570

20022368,2194,FORT YUKON AIRPORT\_20022368

20022366,2195,FORT YUKON RIVER\_20022366

20022430,2196,FRANKLIN BLUFF\_20022430

20022410,2197,GALBRAITH LAKE\_20022410

20022409,2198,GALBRAITH LAKE AIRPORT\_20022409

20022417,2199,HAPPY VALLEY\_20022417

20022422,2200,HAPPY VALLEY CAMP\_20022422

20028924,2201,HERSCHEL ISLAND\_20028924

20028931,2202,HERSCHEL ISLAND\_20028931

20022450,2203,HULL\_20022450

20022435,2204,HUMPHREY POINT BAR A\_20022435

20022457,2205,ICY CAPE LIZ B\_20022457

20022460,2206,IKPIKPUK\_20022460

20022438,2207,INIGOK\_20022438

20028564,2208,INUVIK\_20028564

30005214,2209,INUVIK CLIMATE, NWT\_30005214

20028566,2210,INUVIK UA\_20028566

30009635,2211,INUVIK UA, NWT\_30009635

20022411,2212,IVOTUK\_20022411

30005468,2213,IVVAVIK NAT. PARK, YT\_30005468

20022440,2214,KAD RIVER\_20022440

30000186,2215,KAKTOVIK\_30000186

20022429,2216,KAVIK RIVER\_20022429

20022496,2217,KILLIK\_20022496

20022408,2218,KILLIK\_20022408

20028930,2219,KINGS POINT\_20028930

20028562,2220,KITTIGAZUIT\_20028562

20022399,2221,KIVALINA\_20022399

20022398,2222,KIVALINA\_20022398

30000862,2223,KIVALINA AIRPORT\_30000862

20022381,2224,KOBUK\_20022381

20022479,2225,KOGRU RIVER POW B\_20022479

20022432,2226,KOLUKTAK\_20022432

20028928,2227,KOMAKUK BEACH\_20028928

20028929,2228,KOMAKUK BEACH\_20028929

30004713,2229,KOMAKUK BEACH, YT\_30004713

20022484,2230,KOTZEBUE 25 N\_20022484

20022376,2231,KOTZEBUE RALPH WEIN MEMORIAL AP\_20022376

20022488,2232,KUPARUK\_20022488

20022454,2233,KUPARUK DEV AIRSTRIP\_20022454

20022423,2234,KUYANAK\_20022423

30004748,2235,LITTLE CHICAGO, NWT\_30004748

30005463,2236,LIVERPOOL BAY, NWT\_30005463

20022467,2237,LONELY\_20022467

30005467,2238,MARGARET LAKE, YT\_30005467

20022459,2239,MCINTYRE POW C\_20022459

30000190,2240,NOATAK\_30000190

30002718,2241,NOATAK AP\_30002718

20022375,2242,NOORVIK\_20022375

30015284,2243,NOORVIK\_30015284

20022428,2244,NORA FEDERAL\_20022428

20022451,2245,NORTH INIGOK\_20022451

30000855,2246,NUIQSUT AIRPORT\_30000855

20028922,2247,OLD CROW AUT\_20028922

20022391,2248,OLD MAN\_20022391

20022461,2249,OLIKTOK POW 2\_20022461

20022465,2250,PEARD\_20022465

20022466,2251,PEARD BAY LIZ C\_20022466

20028592,2252,PELLY ISLAND\_20028592

20022437,2253,PINGO\_20022437

10500001,2254,POINT BARROW OBSERVATORY\_10500001

20022475,2255,POINT BARROW POW M\_20022475

20022405,2256,POINT HOPE\_20022405

20022406,2257,POINT HOPE AIRPORT\_20022406

20022433,2258,POINT LAY\_20022433

20022431,2259,POINT LAY LIZ 2\_20022431

20022373,2260,PROSPECT CREEK\_20022373

20022374,2261,PROSPECT CREEK\_20022374

20022452,2262,PRUDHOE BAY\_20022452

20022449,2263,PRUDHOE BAY\_20022449

20022448,2264,PRUDHOE BAY\_20022448

20022522,2265,PRUDHOE BAY\_20022522

30015294,2266,RAMPART HOUSE\_30015294

30014238,2267,RED DOG\_30014238

20022478,2268,REINDEER ISLAND\_20022478

30004866,2269,ROCK RIVER, YT\_30004866

30000749,2270,SAG RIVER D O T\_30000749

20022426,2271,SAGWON\_20022426

20022372,2272,SELAWIK\_20022372

20022370,2273,SELAWIK\_20022370

30002783,2274,SELAWIK AP\_30002783

20028925,2275,SHINGLE POINT\_20028925

30009639,2276,SHINGLE POINT A, YT\_30009639

20028919,2277,SHINGLE POINT BAR-2\_20028919

20022550,2278,SHUNGNAK\_20022550

20022382,2279,SHUNGNAK CAA\_20022382

20022463,2280,SOUTH MEADE RIVER\_20022463

20022427,2281,SUSIE 1\_20022427

10013254,2282,THETIS ISLAND HELIPAD\_10013254

20022407,2283,TIGLUKPUK\_20022407

20022414,2284,TOOLIK\_20022414

30001370,2285,TRAIL VALLEY\_30001370

30004904,2286,TRAIL VALLEY, NWT\_30004904

20028565,2287,TUKTOYAKTUK\_20028565

20028567,2288,TUKTOYAKTUK\_20028567

30005474,2289,TUKTOYAKTUK, NWT\_30005474

20022577,2290,TULAGA OIL RIG #14\_20022577

20022472,2291,TULAGEAK\_20022472

20022445,2292,TUNALIK RIVER\_20022445

20022434,2293,TUNGAK CREEK\_20022434

20022458,2294,UGNU\_20022458

20022453,2295,UGNU KUPARUK AP\_20022453

20022425,2296,UMIAT AFS\_20022425

20022424,2297,UMIAT AP\_20022424

20022518,2298,UMIAT NS\_20022518

20022385,2299,VENETIE\_20022385

20022384,2300,VENETIE\_20022384

20022464,2301,WAINWRIGHT\_20022464

30000830,2302,WAINWRIGHT AIRPORT\_30000830

20022462,2303,WAINWRIGHT LIZ 3\_20022462

20022469,2304,WALAKPA\_20022469

20022471,2305,WEST DEASE\_20022471

20022439,2306,WEST KAVIK\_20022439

20022455,2307,WEST KUPARUK\_20022455

20022392,2308,WILD LAKE\_20022392

20022395,2309,WILD LAKE 2\_20022395

20022390,2310,WISEMAN\_20022390

20022262,7958,SHAKTOOLIK\_20022262

CA002100635,7993,HERSCHEL ISLAND

CA002203900,7994,TUKTOYAKTUK

CA002200200,7995,AKLAVIK RADIOSONDE

## ‘remove\_dup2.awk’

This script varies slightly for each variable, but it pretty much the same. It varies in the missing marker.

BEGIN { FS = "," }

{

if ($4 == prev\_dt)

{

# remove airtemp "+9999"

if ("+9999" == prev\_at || "+9999" == $9)

{

if ("+9999" == prev\_at)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if (5 == prev\_qc || 5 == $10)

{

if (5 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if (1 == prev\_qc || 1 == $10)

{

if (1 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if (4 == prev\_qc || 4 == $10)

{

if (4 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if (0 == prev\_qc || 0 == $10)

{

if (0 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if ("C" == prev\_qc || "C" == $10)

{

if ("C" == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

else if (9 == prev\_qc || 9 == $10)

{

if (9 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

# suspect

else if (6 == prev\_qc || 2 == prev\_qc || 6 == $10 || 2 == $10)

{

if (6 == $10 || 2 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

# erroreous

else if (7 == prev\_qc || 3 == prev\_qc || 7 == $10 || 3 == $10)

{

if (7 == $10 || 3 == $10)

{

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

}

else

{

if ("" != prev\_line)

print prev\_line;

prev\_line = $0;

prev\_dt = $4;

prev\_at = $9;

prev\_qc = $10;

}

}

END {

print prev\_line;

}

# Appendix D: GHCN-Daily scripts

## create\_unique.sh

#!/bin/bash

for ids in `cat $1`

do

echo $ids

awk -f error.awk $ids.txt | sort -t, -k2.1,2.4 -k2.6,2.7 -k2.9,2.10 | awk -f remove\_dup2.awk > ${ids}\_sorted\_unique.txt

done

## error.awk

BEGIN { FS = "," }

{

if ("D" == $5 || "G" == $5 || "I" == $5 || "K" == $5 || "L" == $5 || "M" == $5 || "N" == $5 || "O" == $5 || "R" == $5 || "S" == $5 || "T" == $5 || "X" == $5)

{

print $1 "," $2 ",-9999," $4 ",9," $6;

}

else

{

print $0;

}

}

## remove\_dup2.awk

BEGIN { FS = "," }

{

if ($2 == prev\_dt)

{

# remove airtemp "-9999"

if ("-9999" == prev\_at || "-9999" == $3)

{

if ("-9999" == prev\_at)

{

prev\_line = $0;

prev\_dt = $2;

prev\_at = $3;

prev\_qc = $5;

}

}

else if (5 == prev\_qc || 5 == $5)

{

if (5 == $5)

{

prev\_line = $0;

prev\_dt = $2;

prev\_at = $3;

prev\_qc = $5;

}

}

else if ("W" == prev\_qc || "W" == $5)

{

if ("W" == $5)

{

prev\_line = $0;

prev\_dt = $2;

prev\_at = $3;

prev\_qc = $5;

}

}

# ERROR, set data value to -9999 and marker to 9 (missing)

else if ("9" == prev\_qc || "9" == $5)

{

if ("9" == $5)

{

prev\_line = $0;

prev\_dt = $2;

prev\_at = $3;

prev\_qc = $5;

}

}

}

else

{

if ("" != prev\_line)

print prev\_line;

prev\_line = $0;

prev\_dt = $2;

prev\_at = $3;

prev\_qc = $5;

}

}

END {

print prev\_line;

}