Quality Control of Automatic Weather Station Data

Mark W. Seefeldt
University of Colorado - Boulder

Linda M. Keller
University of Wisconsin - Madison

Outline

- Current AWS data processing
- Weaknesses in the current data processing
- New automated quality control process
- New manual quality control process
- New AWS data features
- New AWS data products

Current AWS Data Processing

089191205.r

089191205.r - Notepad

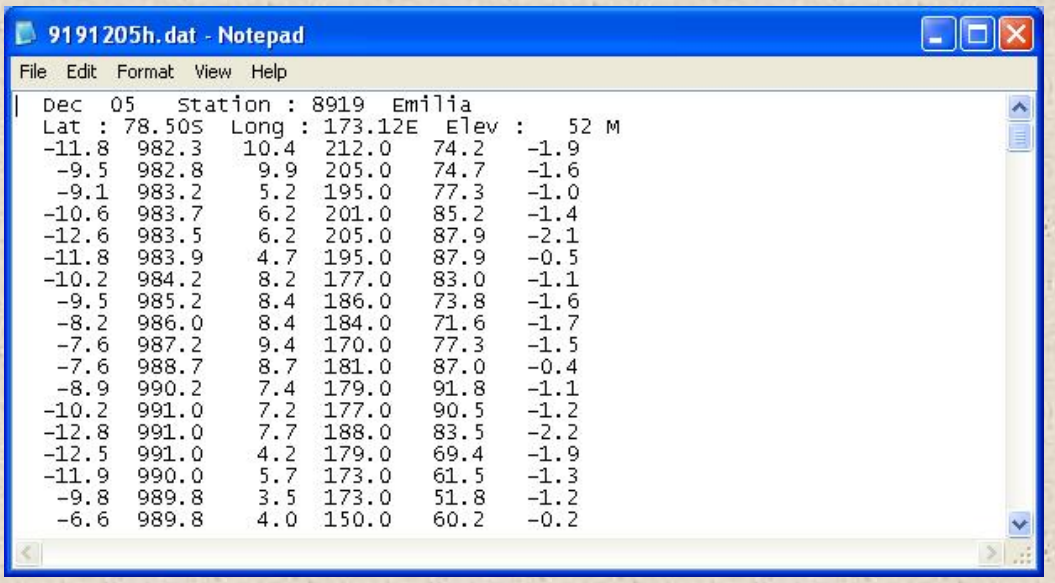
File Edit Format View Help

```
Dec  05   Station : 8919  Emilia
Lat : 78.50S  Long : 173.12E  Elev :    52 M
335    1  -11.8  982.3   10.4  212.0   74.2   -1.9
335    2  -11.2  982.5    6.9  210.0   73.8   -1.8
335    3  -11.4  982.5    8.2  208.0   78.2   -2.0
335    4  -10.8  982.6    8.4  210.0   73.4   -1.5
335    5  -10.5  982.7    8.2  212.0   73.8   -1.4
335    6  -10.4  982.7    8.2  215.0   73.4   -1.6
335    7  -10.4  982.7    8.4  207.0   73.4   -1.7
335    8  -10.2  982.7    8.2  205.0   71.6   -1.9
335    9  -10.2  982.6    8.4  208.0   70.3   -1.9
335   10  -10.0  982.7    8.4  204.0   71.2   -1.8
335   11  -10.0  982.7    8.9  215.0   72.5   -1.9
335   12   -9.6  982.8    7.9  211.0   68.1   -1.8
335   13   -9.6  982.9    6.2  204.0   69.8   -1.9
335   14   -9.5  982.9    5.7  205.0   67.6   -1.9
335   15   -9.6  982.9    7.4  201.0   71.6   -2.0
335   16   -9.6  982.8    8.2  200.0   72.9   -1.9
335   17   -9.6  982.9    6.7  204.0   74.2   -1.7
335   18   -9.5  982.9    9.2  205.0   74.2   -1.7
```

- The 10-minute observations from ARGOS are processed to eliminate extreme outliers, and to format the data into .r 10-minute files
- This is done typically within a few weeks after the end of a month
- There are no quality control procedures performed on the observations, other than the extreme outliers

Current AWS Data Processing

9191205h.dat



- 3-hourly files are created and then scanned manually by Linda Keller
- If an observation looks suspect, in relation to the surrounding observations, the 10-minute observations are checked
- A bad observation is replaced with a surrounding 10-minute observation, if one exists within 40 minutes of the hour
- The process is extremely time consuming

AWS Data Processing - Dilemma

- The current method is very time consuming
- The expanding AWS network is requiring more AWS data to be processed every year
- Due to the subjective nature, it is best for one person to perform the quality control, if possible
- The quality controlled 3-hourly data is currently five years behind
- In order to include the AWS data in climatology archives a more expedient system is needed

AWS Data Processing - Solution

- Attempts to create an automated quality control method have proven to be difficult and not refined enough

  (good points are excluded, and bad points are included)
- Through the use of IDL software a semi-automated quality control process has been developed
  - An automated procedure removes data points based on a statistical evaluation of the observations
  - A manual procedure is performed to review and correct the automated quality control procedure
- This new semi-automated quality control procedure results in a faster quality control process, and additional products

AWS Data Quality Control - Automated

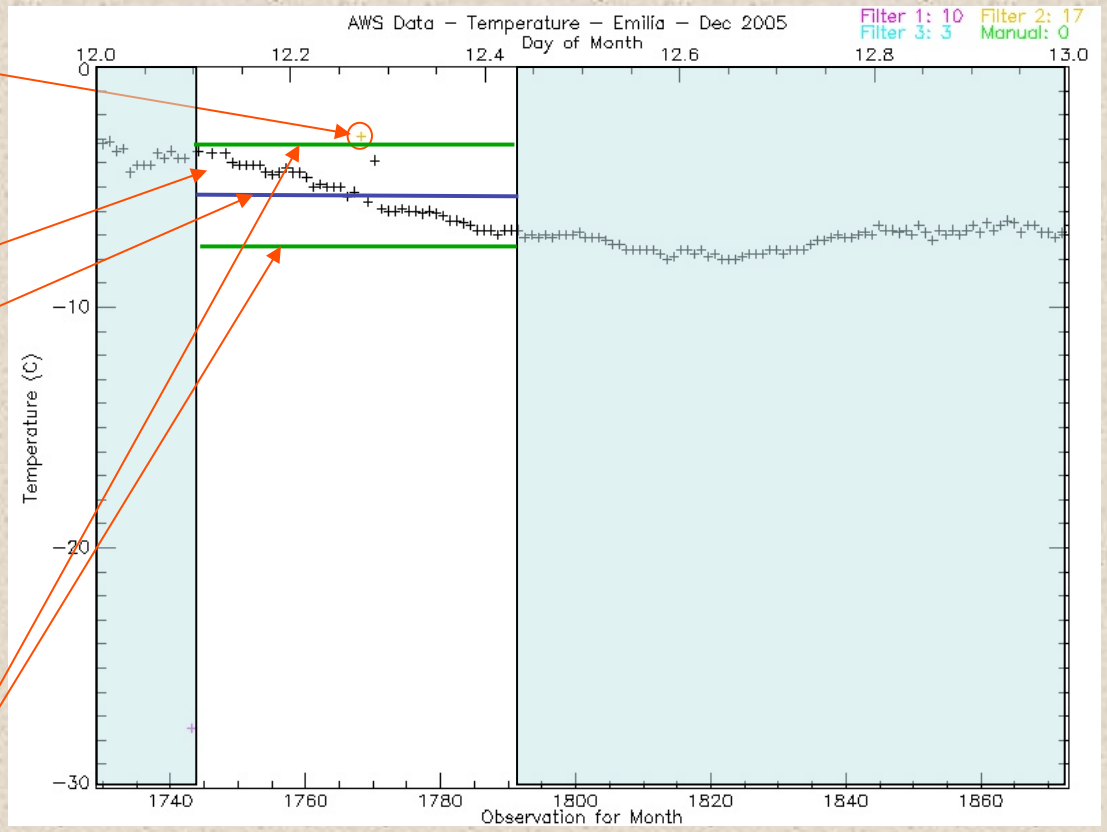
•selected observation

•a window is defined around both sides of the observation (ie. +/- 24)

•the mean of the observations within the window is calculated

•the standard deviation of the observations within the window is calculated

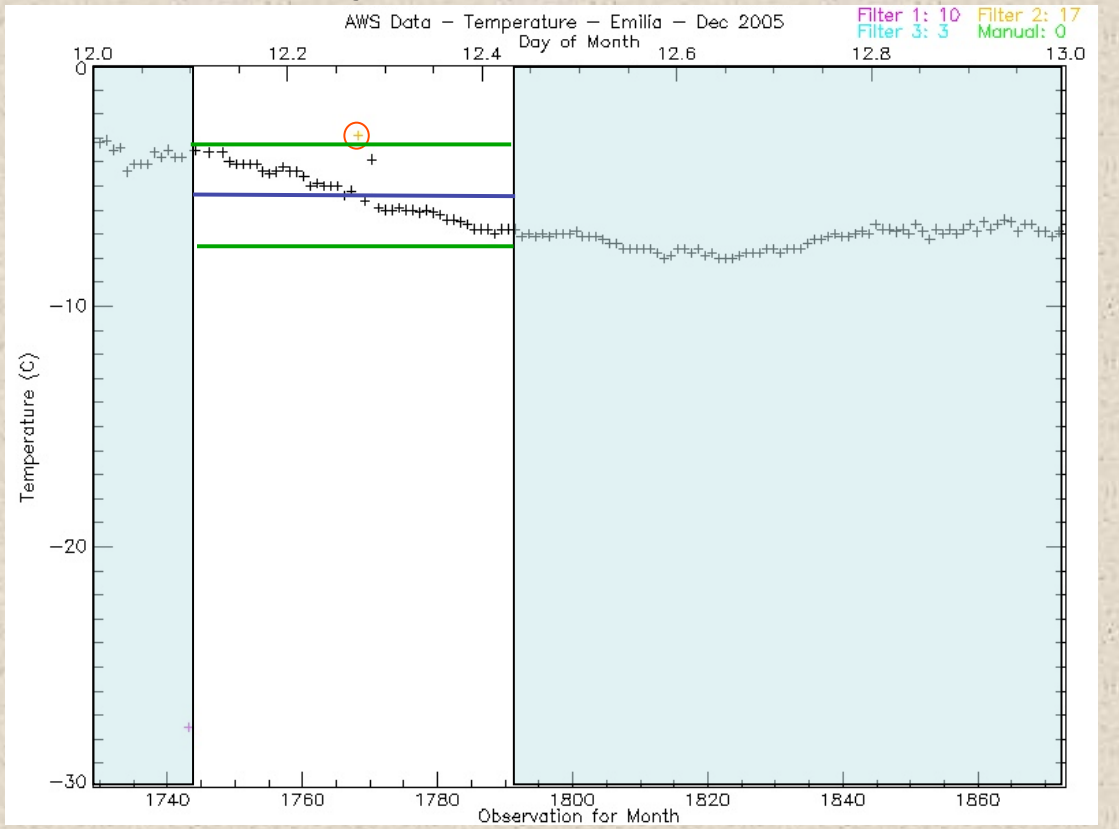
•threshold values are established based on multipliers of the standard deviation (ie. +/- 3 SD)



•if the observation is outside of the threshold values it is removed

•the entire process is repeated for every observation

AWS Data Quality Control – Automated Features



- Up to three different and progressive filters can be used
- The size of the window and the threshold multiplier can be custom specified for each sensor
- This is not able to applied to wind direction due to the more random variability and the discontinuity between 360° and 0°

AWS Data Quality Control – Automated Features

- The automated filter can be configured to remove all of the observations of a specific sensor (ie. bad wind direction at an AWS)
- A consecutive observation filter can be used to remove observations from a specific sensor which remain constant over a set value

AWS Data Quality Control – Manual

- clicking on any observation point will ‘flip’ it from good to bad, or bad to good
- clicking and dragging a box will ‘flip’ all observations in the box
- controls are provided to adjust the day and data ranges
- all sensors, and nearby observations can be viewed in data table

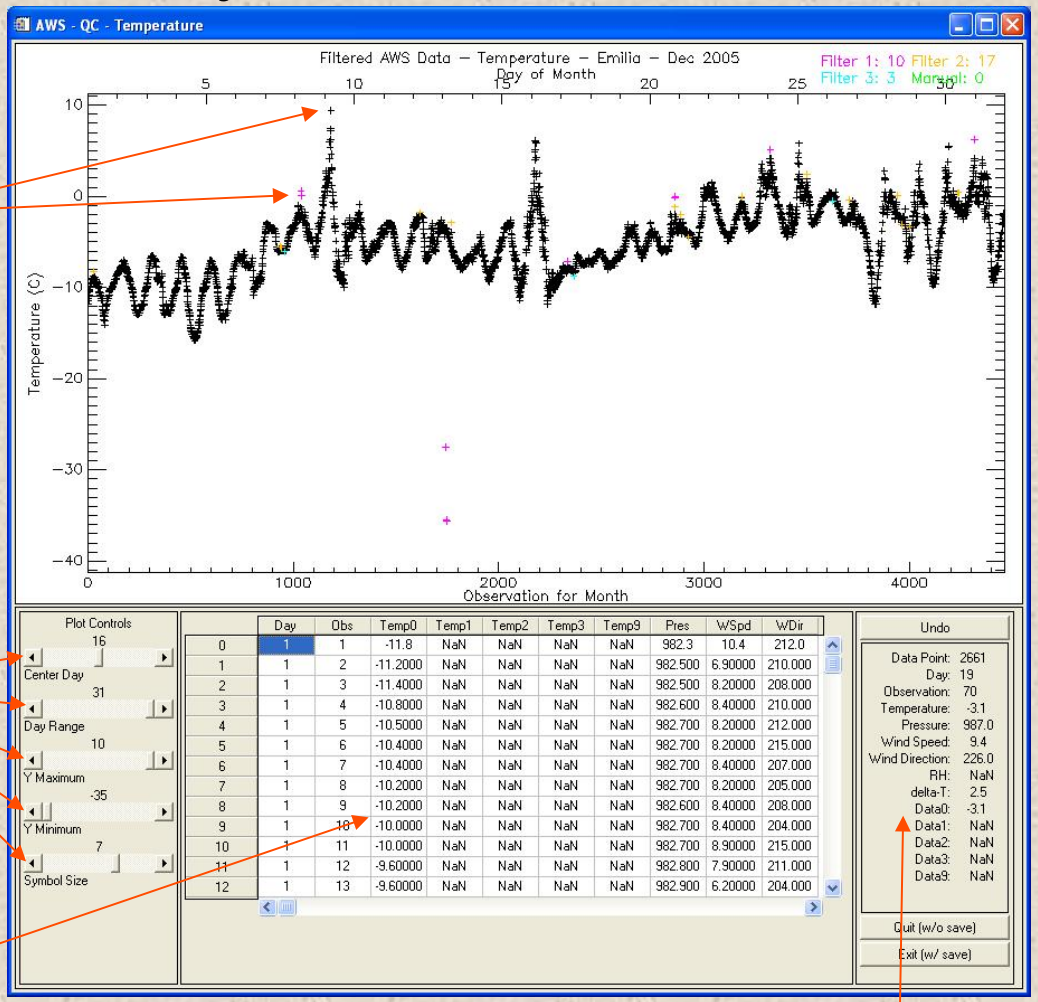
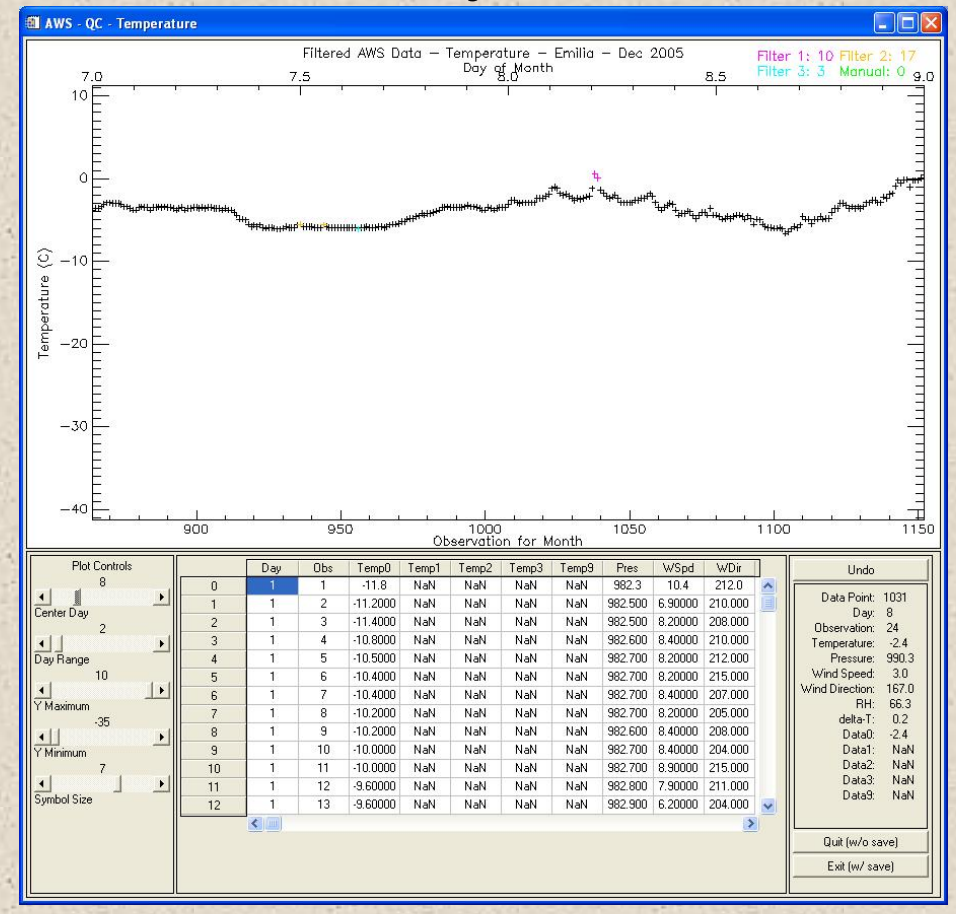


table displays observation information from all the sensors for the point at the cursor

AWS Data Quality Control – Manual



- The manual quality control process is repeated for each sensor
- At the completion of all the sensors the end products are created

AWS Data – New Features

- New data files and formats have been introduced
- AWS sites are identified by a six-letter ID in the filenames:

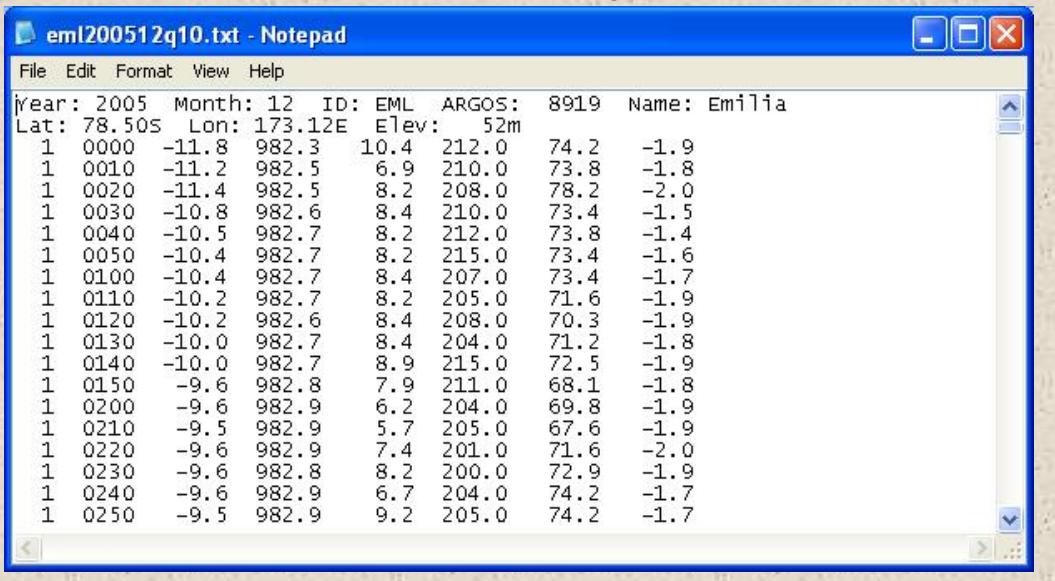
  ex. Emilia: emilia Elizabeth: elzbth Gill: gill__

  Note: The official six-letter IDs have not been determined yet.
- Three-letter IDs are included in the data files to establish a common set of IDs for use in station plots, or other data displays.

  ex. Emilia: EML Gill: GIL Carolyn: CRL
- The day is indicated as the day of month, and the observation as the time of the observation
- All data files will be given the .txt extension

New AWS Data Formats – 10-minute files (q10)

emilia200512q10.txt



- Quality controlled 10-minute data
- Includes all of the processed AWS data

New AWS Data Formats – 1-hour files (q1h)

emilia200512q1h.txt

eml200512q1h.txt - Notepad

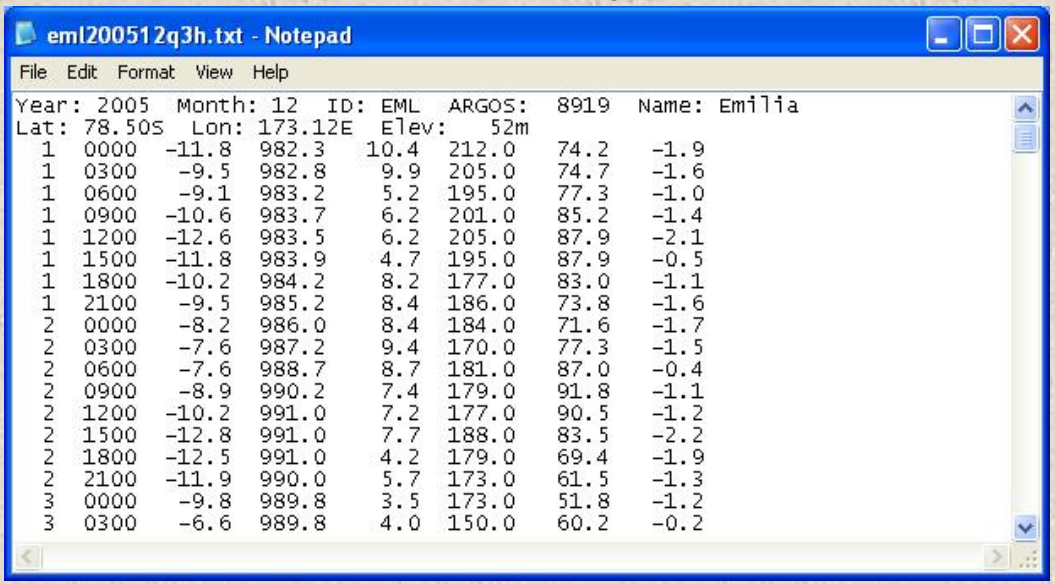
File Edit Format View Help

```
Year: 2005  Month: 12  ID: EML  ARGOS:  8919  Name: Emilia
Lat: 78.50S  Lon: 173.12E  Elev:   52m
  1  0000  -11.8  982.3   10.4  212.0   74.2   -1.9
  1  0100  -10.4  982.7    8.4  207.0   73.4   -1.7
  1  0200   -9.6  982.9    6.2  204.0   69.8   -1.9
  1  0300   -9.5  982.8    9.9  205.0   74.7   -1.6
  1  0400   -8.8  982.9    6.9  205.0   71.6   -1.1
  1  0500   -8.9  983.0    7.7  191.0   72.9   -1.2
  1  0600   -9.1  983.2    5.2  195.0   77.3   -1.0
  1  0700   -9.5  983.3    6.9  197.0   83.0   -1.0
  1  0800  -10.0  983.6    5.5  197.0   85.7   -1.1
  1  0900  -10.6  983.7    6.2  201.0   85.2   -1.4
  1  1000  -10.8  983.6    6.2  201.0   88.3   -0.9
  1  1100  -11.0  983.5    6.9  207.0   93.6   -0.7
  1  1200  -12.6  983.5    6.2  205.0   87.9   -2.1
  1  1300  -14.1  983.5    5.5  200.0   87.4   -2.0
  1  1400  -12.5  983.7    6.4  200.0   90.5   -0.3
  1  1500  -11.8  983.9    4.7  195.0   87.9   -0.5
  1  1600  -11.2  983.9    6.2  200.0   89.2   -0.9
  1  1700  -10.5  983.9    7.4  198.0   84.8   -0.7
```

- Quality controlled 1-hour data
- The observation, within 10 minutes of the hour (0, +10, -10), with the most reporting sensors (T,p,WS,WD) is selected for the hourly observation
- Any missing sensors are filled in with valid observations within 10 minutes of the hour

New AWS Data Formats – 3-hour files (q3h)

emilia200512q3h.txt


eml200512q3h.txt - Notepad

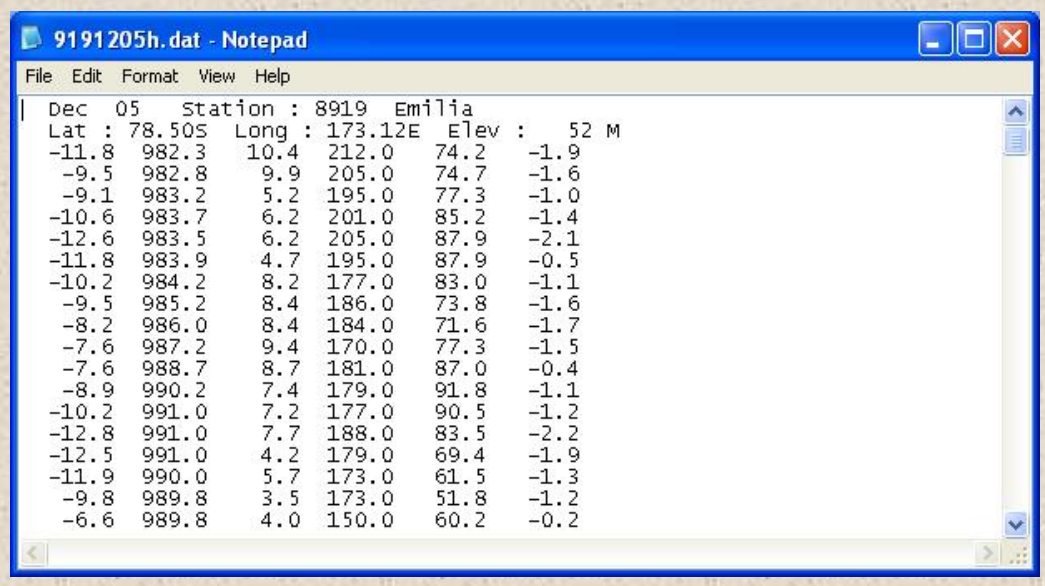
File Edit Format View Help

```
Year: 2005  Month: 12  ID: EML  ARGOS:  8919  Name: Emilia
Lat: 78.50S  Lon: 173.12E  Elev:   52m
  1  0000  -11.8  982.3  10.4  212.0   74.2   -1.9
  1  0300   -9.5  982.8   9.9  205.0   74.7   -1.6
  1  0600   -9.1  983.2   5.2  195.0   77.3   -1.0
  1  0900  -10.6  983.7   6.2  201.0   85.2   -1.4
  1  1200  -12.6  983.5   6.2  205.0   87.9   -2.1
  1  1500  -11.8  983.9   4.7  195.0   87.9   -0.5
  1  1800  -10.2  984.2   8.2  177.0   83.0   -1.1
  1  2100   -9.5  985.2   8.4  186.0   73.8   -1.6
  2  0000   -8.2  986.0   8.4  184.0   71.6   -1.7
  2  0300   -7.6  987.2   9.4  170.0   77.3   -1.5
  2  0600   -7.6  988.7   8.7  181.0   87.0   -0.4
  2  0900   -8.9  990.2   7.4  179.0   91.8   -1.1
  2  1200  -10.2  991.0   7.2  177.0   90.5   -1.2
  2  1500  -12.8  991.0   7.7  188.0   83.5   -2.2
  2  1800  -12.5  991.0   4.2  179.0   69.4   -1.9
  2  2100  -11.9  990.0   5.7  173.0   61.5   -1.3
  3  0000   -9.8  989.8   3.5  173.0   51.8   -1.2
  3  0300   -6.6  989.8   4.0  150.0   60.2   -0.2
```

- Quality controlled 3-hour data
- The observation, within 40 minutes of the hour (0,+10,-10,+20, -20,+30,-30,+40,-40), with the most reporting sensors (T,p,WS,WD) is selected for the hourly observation
- Any missing sensors are filled in with valid observations within 40 minutes of the hour

AWS Data Formats – 3-hour files (dat)

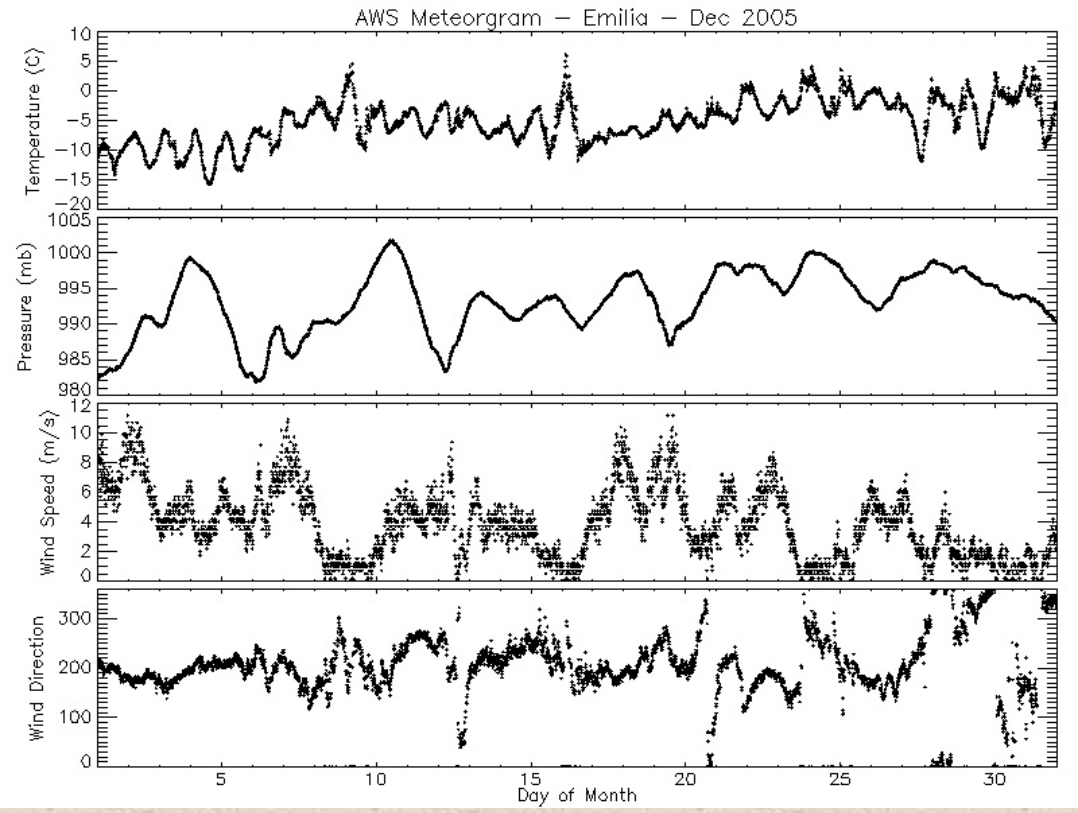
9191205h.dat



- Quality controlled 3-hour data – same data as the q3h files
- The file naming and data format is the exact same as the current 3-hourly files
- This product will be discontinued in the future and should no longer be used

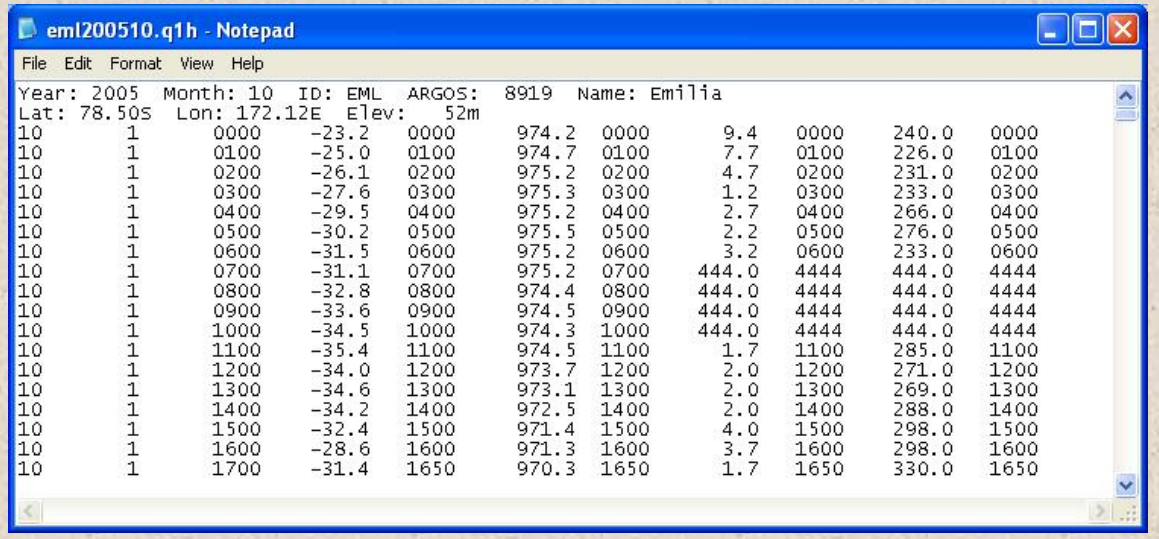
New AWS Data Product – Monthly Meteograms

emila200512met.jpg



- Plot of temperature, pressure, wind speed, and wind direction, from the quality controlled 10-minute observations, for a month

Future AWS Data Product – Detailed 1h/3h data



eml200510.q1h - Notepad

File Edit Format View Help

```
Year: 2005  Month: 10  ID: EML  ARGOS:  8919  Name: Emilia
Lat: 78.50S  Lon: 172.12E  Elev:   52m
10       1      0000    -23.2   0000     974.2  0000      9.4   0000    240.0   0000
10       1      0100    -25.0   0100     974.7  0100      7.7   0100    226.0   0100
10       1      0200    -26.1   0200     975.2  0200      4.7   0200    231.0   0200
10       1      0300    -27.6   0300     975.3  0300      1.2   0300    233.0   0300
10       1      0400    -29.5   0400     975.2  0400      2.7   0400    266.0   0400
10       1      0500    -30.2   0500     975.5  0500      2.2   0500    276.0   0500
10       1      0600    -31.5   0600     975.2  0600      3.2   0600    233.0   0600
10       1      0700    -31.1   0700     975.2  0700    444.0   4444    444.0   4444
10       1      0800    -32.8   0800     974.4  0800    444.0   4444    444.0   4444
10       1      0900    -33.6   0900     974.5  0900    444.0   4444    444.0   4444
10       1      1000    -34.5   1000     974.3  1000    444.0   4444    444.0   4444
10       1      1100    -35.4   1100     974.5  1100      1.7   1100    285.0   1100
10       1      1200    -34.0   1200     973.7  1200      2.0   1200    271.0   1200
10       1      1300    -34.6   1300     973.1  1300      2.0   1300    269.0   1300
10       1      1400    -34.2   1400     972.5  1400      2.0   1400    288.0   1400
10       1      1500    -32.4   1500     971.4  1500      4.0   1500    298.0   1500
10       1      1600    -28.6   1600     971.3  1600      3.7   1600    298.0   1600
10       1      1700    -31.4   1650     970.3  1650      1.7   1650    330.0   1650
```

- In the future, a FORTRAN program will be provided which creates an output file which includes the time of the observation from each sensor
- The input files will be the q10 file and either the q1h or q3h files

Summary

- A new semi-automated data quality control system has been establish to process AWS data
- The new system provides easy creation of new data products (10-minute, 1-hour, 3-hourly, meteograms)
- Additional changes will be implemented with the new data products (use of six-letter IDs, new filenameing, etc.)
- The new system is not quite operational, an announcement will be made when the new system and products are implemented

Questions

- Mark Seefeldt mark.seefeldt@colorado.edu
- Linda Keller lmkeller@wisc.edu