Quality Control and Data Estimation for Fire Weather Datasets for Fire Planning and Analysis (FPA)

Program for Climate, Ecosystem, and Fire Applications (CEFA)
Western Regional Climate Center (WRCC)
Desert Research Institute (DRI)

Version: December 2005 (vDec05)

CURRENT DATA VERSION vDec05 ARE UNVALIDATED FOR SOME FIELDS - SEE NOTE BELOW

June 12, 2007

CEFA and WRCC are processing weather datasets based on a station list provided by FPA. The purpose of this processing is to perform quality data checking on the original observations, and provide for data estimation where needed to generate a complete dataset. The North American Regional Reanalysis gridded dataset is currently being utilized for the estimation process and is briefly described below. The quality control (QC) algorithms are provided below.

Hourly Remote Automated Weather Station (RAWS) data for this project was acquired from the WRCC RAWS historical database. This data is received real-time at WRCC via the Automated Sorting, Conversion and Distribution System (ASCADS). CEFA applies a series of algorithms to determine through coarse quality control methods those data that are missing, questionable, and/or impossible to occur. Any data that was not determined to be a reasonable observation was flagged for replacement by an estimated value.

Criteria causing an original observation to be flagged and removed include:

Air temperature
- Observation exceeded historical maximum/minimum values for that state
- Observation stayed constant over past 24 hours
- 3 consecutive hours stayed constant at a value < 5°F

Relative humidity
- Observation was either < 0 or > 100
- Observation stayed constant over past 24 hours

Wind speed
- Observation was < 0 mph or > 115 mph
- Observation was constant over past 12 hours
- Observation was constant

Wind direction
- Observation was < 0 or > 360
- Observation was constant over past 8 hours

Precipitation

- Observation was < 0
- Observation exceed historical monthly total for that state
- Observation exceeded 2 inches

Estimations of flagged data were derived from multiple linear regression analyses of each station, 8 time periods in the day, and for whether or not there was a good observation both the previous hour and the same hour yesterday. Predictor variables included modeled output variables from the National Center for Atmospheric Research (NCAR) / National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR) dataset (http://wwwt.emc.ncep.noaa.gov/mmb/rreanl), daily climatologies for the 3-hourly time periods, and persistence (both from the last hour, and yesterday at the same hour). Details of this process will be provided in a project report.

PRODUCTS AND DELIVERABLES

Due to a number of weather stations not having an assigned 6-digit station ID, pre-assigned 4-character codes from the WRCC are implemented for all station ID's for this product. This 4-character code in the station ID field replaces the 6-digit WIMS ID in the *.fwx (daily) and *.fw9 (hourly) files. The new 4-character code is also used as part of the data file names so station catalogues used in software such as PCHA should also use the same 4-character code for the station ID. Data files include:

1. Daily *.fwx data - traditional 1972 WIMS format. The only modification from the original format specifications is the use of the 4-character code in lieu of the 6-digit ID
2. Hourly *.fw9 data - traditional 1998 WIMS format. The only modification from the original format specifications is the use of the 4-character code in lieu of the 6-digit ID
3. Hourly *.dat data - this is a comma delimited ASCII/text file that includes both the raw data and the flag for each value indicating whether the value is the original observation (1) or an estimation (2)

3a. The columns of the .dat files represent the following information:

Column	Definition
1	NWS ID
2	Year
3	Day of Year
4	Month
5	Day
6	Hour (LST)
7	Precipitation Amount
8	Temperature (F)
9	Relative Humidity
0	Wind Speed (mph)
11	Wind direction
12	State of the Weather
13	Precipitation QC flag
14	Temperature QC flag
15	Relative Humidity QC flag
16	Wind Speed QC flag
17	Wind Direction QC flag

4. RequestID-ESTIMATES.txt - this file list the percentages of each variable that is estimated and observed
5. RequestID-METADATA.txt - this file lists the metadata for each station of the data request including the:
   - 4-character WRCC code (filename code, station ID)
   - 6-digit NWS ID (if available)
   - Station name
   - Latitude
   - Longitude
   - Elevation
   - Period of record

VALIDATION

A validation procedure was performed on a subset of stations that were spread throughout the United States. This procedure applied the estimation methods and algorithms on values with known observations. The estimations were then correlated to the observations to quantify how close the estimations were to the known observations. The general results were as follows:

Variable	Correlation Coefficient	Average Difference
Temperature	0.97+	2F – 4F
Relative Humidity	0.94+	4% - 7%
Wind Speed	0.7 – 0.8	1 mph – 3 mph
Wind Direction	0.5 – 0.8	30 deg – 70 deg
Precipitation Amount	0.1 – 0.6	0.0 – 0.03 in

State of the Weather (SOW) has not been validated as of the release of this version (vOct05).

SOFTWARE

Note that due to the change of the station IDs from the traditional 6-digit NWS IDs to the 4-character WRCC codes both within the data files and within the filenames, these data files may not work on all versions of PCHA or FireFamily Plus. These files have only been known to work on the following:

PCHA version 1.2.31 Patch 1f or later
FireFamily Plus version 3.0.5 or later

VERSIONS

vFeb05 - First complete release of the hourly RAWS data for the US. This version was based upon equations from the Global Reanalysis dataset. Equations were developed for two seasons (May-Oct, Nov-Apr), 4 time periods in the data (00-05UTC, 06-11UTC, 12-17UTC, 18-23UTC). This version showed problems of consecutive estimated values becoming increasingly erroneous and had to be converted to missing. Precipitation and winds had relatively low validation results.

vOct05 - The first version that included the NARR data set within the multiple linear equation algorithms. This version also was the first to employ two separate sets of equations depending upon the existence of observations for the persistence values. This modification eliminated the problem of consecutive estimated values becoming increasing erroneous. Precipitation and winds continue to have relatively low validation results.

vDec05 - An error was discovered in the climatology values used for computing estimations. This error did not affect all values, only a small subset of cases. All data was re-run with these errors corrected.

CONTACTS

Beth L. Hall (CEFA)
(775) 674-7174
Beth.Hall@dri.edu
2215 Raggio Pkwy
Reno, NV 89512

Greg McCurdy (WRCC)
(775) 674-7165
Greg.McCurdy@dri.edu
2215 Raggio Pkwy
Reno, NV 89512

Howard Roose (FPA)
(208) 947-3781
Howard_Roose@nifc.blm.gov
1387 South Vinnell Way
Boise, ID 8309-1657