MEMORANDUM

Tetra Tech, Inc.
10306 Eaton Place, Suite 340
Fairfax, VA 22030
Telephone (703) 385-6000
Fax (703) 385-6007

Date:	October 26, 2016
From:	Khalid Alvi and Mustafa Faizullabhoy – Tetra Tech
To:	Mark Voorhees, EPA Region I
Subject:	Download and Process Weather Data for SWMM Model

This memorandum broadly discusses the various options available to the user for downloading climate data from the internet, process it, and create SWMM climate input files. Climate data can be downloaded from NOAA's Climate Data Online (CDO) website. The website provides free access to National Climate Data Center's (NCDC's) archive of global historical weather and climate data NOAA National Data Center (NNDC). The website offers a wealth of historical data at no cost to the user, allowing the user to choose from several data product options. The user can get historical climate data via a search tool, mapping tool, and various other data tools. For this example we will use the mapping tool option available as it is more visual and allows the user to explore available stations in the vicinity of the site of interest, which in our example is located at Fairhaven, MA.

STEPS FOR DOWNLOADING AND PROCESSING WEATHER DATA

1) Open a web browser and type in the web link to NCDC's CDO website
   https://www.ncdc.noaa.gov/cdo-web/

2) Click on Mapping Tool from the orange box as shown below.
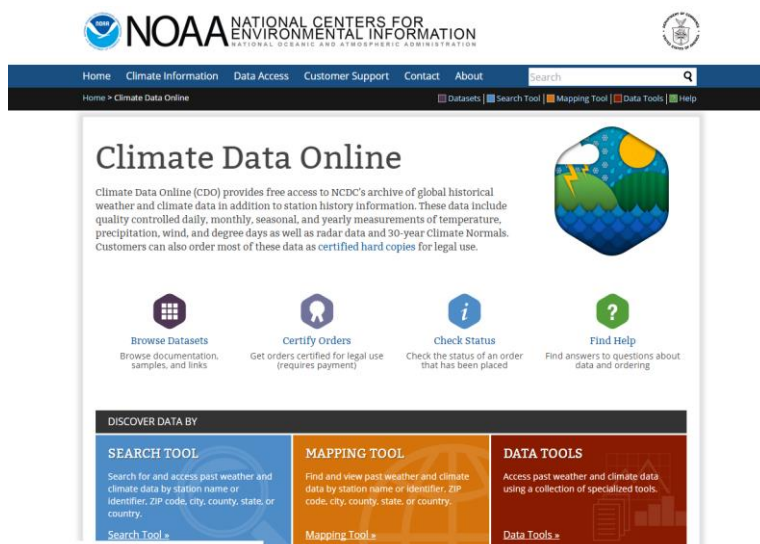
3) This will bring up a form which allows the user to select from several options.

   a. Click on the "Time-Related Maps" tab. This will bring up several time series related data options.

   b. Since we are looking for hourly time series data click on the "Hourly/Sub-Hourly" option.

4) Clicking on the "Hourly/Sub-Hourly" brings up the GIS Map Portal, with several hourly/sub-hourly data options to choose from. We are looking for hourly data so check the box for "Hourly Global" (default selection).

5) Locate the area of interest and zoom into the area of interest using the map interface. The user can either zoom and pan through the map or locate a place using the "Search for a Location" option.
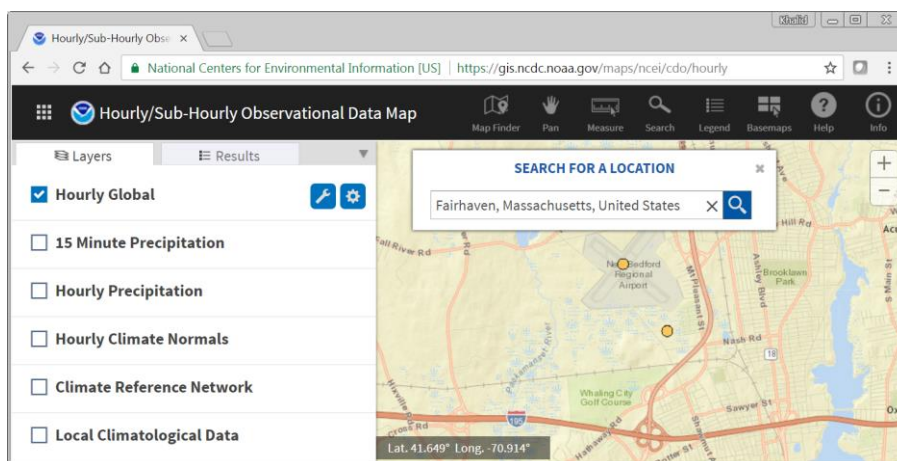

6) For our purposes the Integrated Surface Global Hourly Data (ISD) (Hourly Global), Hourly Precipitation Data (HPD), and Local Climatological Data (LCD) offer the best options. Within the US the ISD and LCD data are summaries of climatological conditions from airport and other

prominent weather stations managed by National Weather Service (NWS), Federal Aviation Administration (FAA), and Department of Defense (DOD). The HPD stations are also reported from the prominent weather stations like ISD/LCD and in addition also offer data at several other co-operative observer station locations.

In addition to hourly precipitation the ISD/LCD stations offer several other parameters of interest such as air temperature, dew point temperature, relative humidity, wind speed/direction, and atmospheric pressure. The LCD product includes hourly observations and associated remarks, and a record of hourly precipitation on a monthly basis for each year (monthly ASCII or HTML). The LCD dataset reports data on the hour and eliminates sub-hourly data. Since it reports hourly data on a monthly basis it can be a little cumbersome with no automation of the workflow. However, the ISD provides a merged ASCII dataset for a per-defined historical time period over several years. For this example it is suggested that the ISD dataset be used at it provides a merged dataset and reports both hourly precipitation and air temperature measurements. In some other locations it may be necessary to take precipitation and air temperature from different data sources based on the proximity of the stations to the site of interest.

For a complete list of parameters and description of the dataset the user is referred to the metadata of the datasets. The metadata can be accessed by clicking the ⚙ icon which shows up once the dataset is selected in the browser.
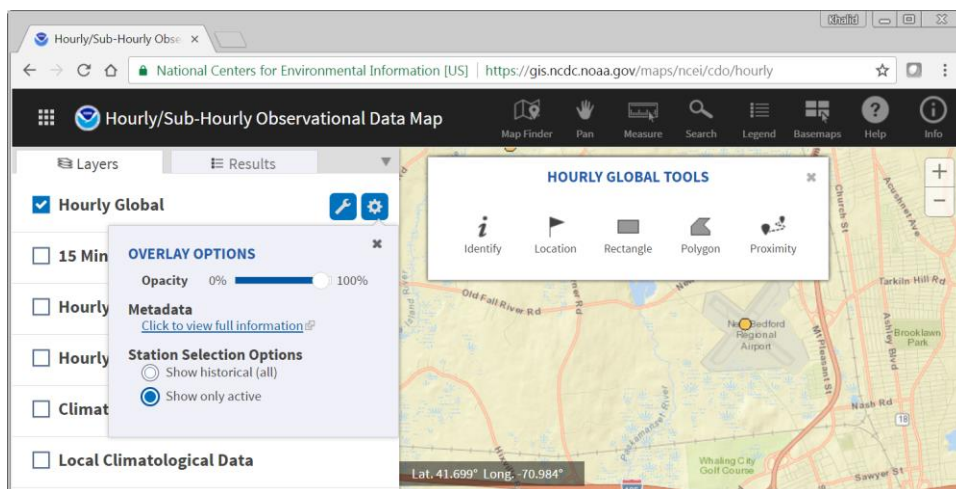
Clicking on the 🔧 icon brings up a Tools menu box which is useful for querying the data points. Clicking on the identify tool and selecting a station provides a brief descriptions of the station. Several stations can be selected using the rectangle or polygon option available on Hourly Global Tools menu box.

7) Select the active station at the New Bedford Regional Airport by using the identify tool. The resulting station description along with the period of record indicates that the New Bedford Regional Airport station has the more recent data of interest from 1/1/2006 to 10/17/2016 (current). Selecting the checkbox for the selected station on the dataset makes the "Access Data" button active (turns orange) at the bottom of the page. Click on the "Access Data" button to access options to download the data. Select Simplified as the data access option.
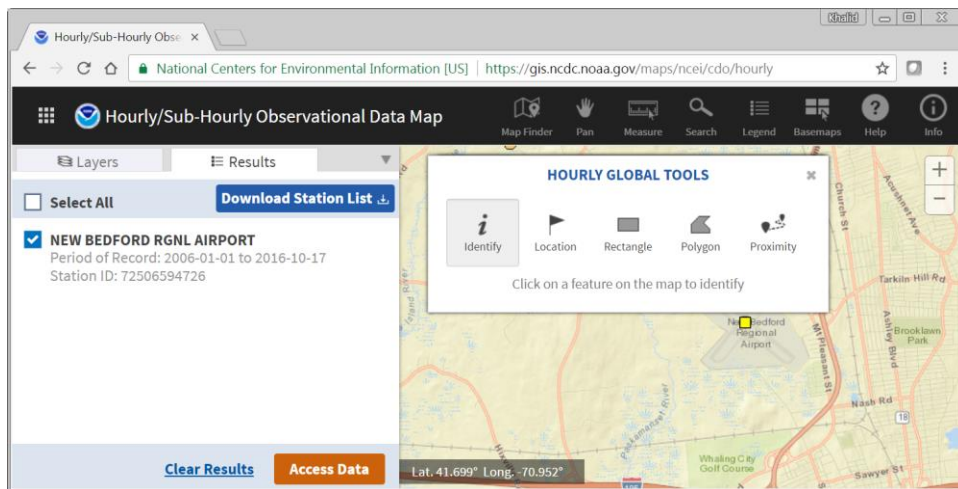
8) Next select the period of interest using the data range. Click on the radio button "Select Only Obs. On the Hour". This will eliminate several of the sub-hourly data reported. Click on "Continue" to go to the next page.

9) Check the period of record selected and click on the check boxes for Inventory Review and the check box next to "I'm not a robot". Be sure to provide a valid email address. NOAA National Data Center (NNDC) will send an email at the email address provided with links to access the data. Click on Submit Request. The amount of time it takes to get the ftp link to the data via email depends on the length of the period of record chosen and number of stations selected.

10) The email from NNDC provides links to several files including the data file for the period of interest — Data File, Web File, Stations included in your search, Data inventory for your search, and format documentation. All the files are space delimited text files, except for the Web file which is in HTML format.

NOAA Satellite and Information Service
National Environmental Satellite, Data, and Information Service (NESDIS)

National Climatic
Data Center
U.S. Department of Commerce

DOC > NOAA > NESDIS > NCEI

Land-Based Data / NNDC CDO / Product Search / Help

NNDC CLIMATE DATA ONLINE

DS3505 - Surface Data Hourly Global, Data Access

Request ID: CDO02642991

NOTICE!

From the time you submitted your request it will be several minutes (up to 24 hours for larger volumes) before your data, data inventory and station list files are available. **For this reason it is recommended that you bookmark this web page for future reference and access to your files.** You will also receive email notification when your data files are ready.

This web page and the data files / web forms listed below will be available for **7 days**, after which they will be deleted from NCEI's web server.

File Contents	Access URL	File Size Estimate (bytes)
DS3505 - Surface Data Hourly Global - Web Form	http://www.ncdc.noaa.gov/orders/isd/4240147161858dat.html	72117936
DS3505 - Surface Data Hourly Global - Data File	http://www.ncdc.noaa.gov/orders/isd/4240147161858dat.txt	18029484
DS3505 - Surface Data Hourly Global - Inventory	http://www.ncdc.noaa.gov/orders/isd/4240147161858inv.txt	1277
Station List	http://www.ncdc.noaa.gov/orders/isd/4240147161858stn.txt	449
DS3505 - Surface Data Hourly Global format documentation	http://www.ncdc.noaa.gov/cdohtml//3505doc.txt	Not available - see email message

Notes:
*File sizes shown are calculated estimates, and in certain cases may differ significantly from actual file sizes.
*Click on the ftp links above to view the files / web forms in your web browser.
*To download the files / web forms to your local computer choose **File/Save As** or **Save Link As** from your web browser.
*If your data file / web form has a **.gz** extension, the file has been compressed using UNIX compress. For information on how to uncompress the file, please access the CDO Receipt and Usage of Data page.
*For further information on data usage, see the CDO Receipt and Usage of Data page.

An excerpt of the Data File is shown below.

```
 USAF  WBAN YR--MODAHRMN DIR SPD GUS CLG SKC L M H  VSB MW MW MW MW AW AW AW AW W TEMP DEWP    SLP   ALT    STP MAX MIN PCP01 PCP06 PCP24 PCPXX SD
725065 94726 200601010053 050   6 ***  11 OVC * * *  1.3 71 10 ** ** 71 10 ** ** *   32   30 1010.1 29.83 1007.3 *** ***  0.00T***** ***** ***** **
725065 94726 200601010153 060   7 ***   4 OVC * * *  1.3 71 10 ** ** 71 10 ** ** *   32   31 1009.3 29.81 1006.6 *** ***  0.01 ***** ***** ***** **
725065 94726 200601010253 050   7 ***   4 OVC * * *  2.5 ** ** ** ** 71 10 ** ** *   33   31 1008.7 29.79 1005.9 *** ***  0.01 ***** ***** 0.02 **
725065 94726 200601010353 040   6 ***   4 OVC * * *  2.0 ** ** ** ** 71 10 ** ** *   33   31 1008.5 29.78 1005.6 *** ***  0.02 ***** ***** ***** **
725065 94726 200601010453 030   3 ***   4 OVC * * *  1.8 ** ** ** ** 71 10 ** ** *   33   31 1008.4 29.78 1005.6  36  23  0.02 ***** ***** ***** **
725065 94726 200601010459 *** *** *** *** *** * * *  **** ** ** ** ** ** ** ** ** * **** ****  ****** ***** ******  36 *** ***** ***** 0.07 ***** **
725065 94726 200601010553 020   3 ***   4 OVC * * *  4.0 ** ** ** ** 71 10 ** ** *   33   31 1008.6 29.78 1005.6  33  32  0.03  0.08 ***** ***** **
725065 94726 200601010653 350   3 ***   8 OVC * * *  2.5 ** ** ** ** 71 10 ** ** *   33   31 1010.0 29.82 1006.9 *** ***  0.02 ***** ***** ***** **
725065 94726 200601010753 360   5 ***  12 OVC * * *  3.0 ** ** ** ** 71 10 ** ** *   33   31 1010.5 29.84 1007.6 *** ***  0.02 ***** ***** ***** **
725065 94726 200601010759 360   5 ***   4 OVC * * *  2.5 ** ** ** ** 71 10 ** ** *   34   30 ****** 29.84 1007.6 *** ***  0.00T***** ***** ***** **
725065 94726 200601010853 360   6 ***   6 OVC * * *  7.0 ** ** ** ** ** ** ** ** *   33   31 1010.9 29.85 1008.0 *** ***  0.03 ***** ***** 0.08 **
725065 94726 200601010953 010   7 ***   8 OVC * * * 10.0 ** ** ** ** ** ** ** ** *   33   31 1011.6 29.87 1008.6 *** ***  0.02 ***** ***** ***** **
725065 94726 200601011053 010   5 ***  31 OVC * * * 10.0 ** ** ** ** ** ** ** ** *   33   30 1012.9 29.91 1010.0 *** ***  0.01 ***** ***** ***** **
725065 94726 200601011153 350   5 ***   5 OVC * * *  5.0 71 10 ** ** 71 10 ** ** *   32   30 1013.7 29.93 1010.7  33  32  0.01  0.12  0.21 ***** **
725065 94726 200601011253 ***   0 ***   5 OVC * * *  9.1 ** ** ** ** ** ** ** ** *   30   27 1014.6 29.96 1011.7 *** ***  0.00 ***** ***** ***** **
725065 94726 200601011353 ***   0 ***  35 OVC * * *  2.5 10 ** ** ** 10 ** ** ** *   29   27 1015.9 30.00 1013.0 *** ***  0.00T***** ***** ***** **
```

11) Pull the Data file into MS Excel for processing. The data file is space delimited and can be easily parsed using the Data→Text to Columns option in MS Excel and choosing space as the delimiter.

12) Review the format documentation which includes detailed metadata for each of the parameters in the data file. Determine the column for hourly precipitation and air temperature. An excerpt of the data format of the parameters of interest are shown below

```
Column PCP01 is 1-HOUR LIQUID RECIP REPORT IN INCHES AND HUNDREDTHS —
THAT IS, THE PRECIP FOR THE PRECEDING 1 HOUR PERIOD
Column TEMP is in Fahrenheit.
Column YR—MODAHRMN is YEAR-MONTH-DAY-HOUR-MINUTE IN GREENWICH MEAN TIME
(GMT)
```

Caution: *The date reported in the data file is given in GMT and should be converted to the local standard time*

13) The meteorological data sets sometimes contained intervals of missing data. The ISD dataset flags missing data using a series of asterisk "****". For watershed modeling, a continuous record is required to adequately represent continuous hydrology. The New Bedford data were complete with no missing data (except for a few hours) and did not require patching of missing data. Any missing data were filled using the previous hour (especially air temperature).

14) Ensure that the reported values in the data file are for each hour and that there are no missing hours. Typically NCDC reports all hourly data at the 53rd minute of the hour. The user needs to make sure this is indeed true for their dataset too and any dataset they download in the future. Extract the hour from the date and filter out any values other than the 53rd minute and delete it. The goal of this exercise is to have one value for each hour in a day for the entire period of record (in line with the requirements for the SWMM meteorological files).

15) Next compute the daily maximum and minimum value from the hourly air temperature.

16) The SWMM model requires two weather files

- an hourly precipitation file (*.dat)
- a daily min/max air temperature file (*.swm)

The data format for both the files is space delimited.

The format for the *.dat file is as follows:

```
Station ID followed by the year, month, day, hour, minute, and Precipitation (inches)
94726_uo 2006 1 1 0 0 0.
94726_uo 2006 1 1 1 0 0.02
94726_uo 2006 1 1 2 0 0.03
94726_uo 2006 1 1 3 0 0.02
94726_uo 2006 1 1 4 0 0.03
94726_uo 2006 1 1 5 0 0.01
94726_uo 2006 1 1 6 0 0.02
```

The format for the *.swm file is as follows:

```
Station ID followed by the year, month, day, Daily Maximum Air Temperature (°F), and Daily
Minimum Air Temperature (°F)
94726_uo 2006 1 1 33 29
94726_uo 2006 1 2 41 31
94726_uo 2006 1 3 36 32
94726_uo 2006 1 4 36 29.5
94726_uo 2006 1 5 41 31.5
94726_uo 2006 1 6 40 29.5
94726_uo 2006 1 7 33 23
```

RECOMMENDATION FOR PATCHING STATIONS WITH MISSING DATA

When missing data exists, the missing intervals can be estimated using weather data at nearby stations with unimpaired data. Missing precipitation can be estimated using the normal-ratio method, which estimates a missing value with a weighted average from surrounding index stations with similar precipitation patterns according to this relationship:

$$P_A = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{N_A}{N_i}P_i\right)$$

where P_A is the estimate for the impaired value at station A, n is the number of surrounding index stations with unimpaired data at the same specific point in time, N_A is the long-term average value at station A, N_i is the long term average value at nearby index station i, and P_i is the observed value at nearby index station i. For each impaired daily value at station A, n consists of only the surrounding index stations with unimpaired data; therefore, for each record, n varies from 1 to the maximum number of surrounding stations. In the case of precipitation, when no precipitation is available at the surrounding stations, zero precipitation is assumed at station A. Because normalization is the underlying principle, the method is adaptable to regions where there is large orographic variation in weather.

The first step in the process should be to compute the percent missing on an annual or preferably monthly basis for each of the stations in the modeling period of interest. Second, for each station with missing data within the period of interest, nearby index stations should be selected according to (1) shortest straight-line distance from the station, and (2) availability of unimpaired data for periods of impaired data at the station. This search for additional unimpaired stations can be conducted using the mapping interface discussed previously in the memo. A minimum of three nearby daily stations is typically adequate for patching impaired data. Patching of missing intervals can be performed at a daily or monthly time step.

For air temperature typically if a few hours in a day are missing, then patching with the previous hour value may suffice. However, for missing data over several days would require patching or filling the data using the closest station. Stations having similar seasonal average temperatures should be selected and used for patching missing air temperature data.