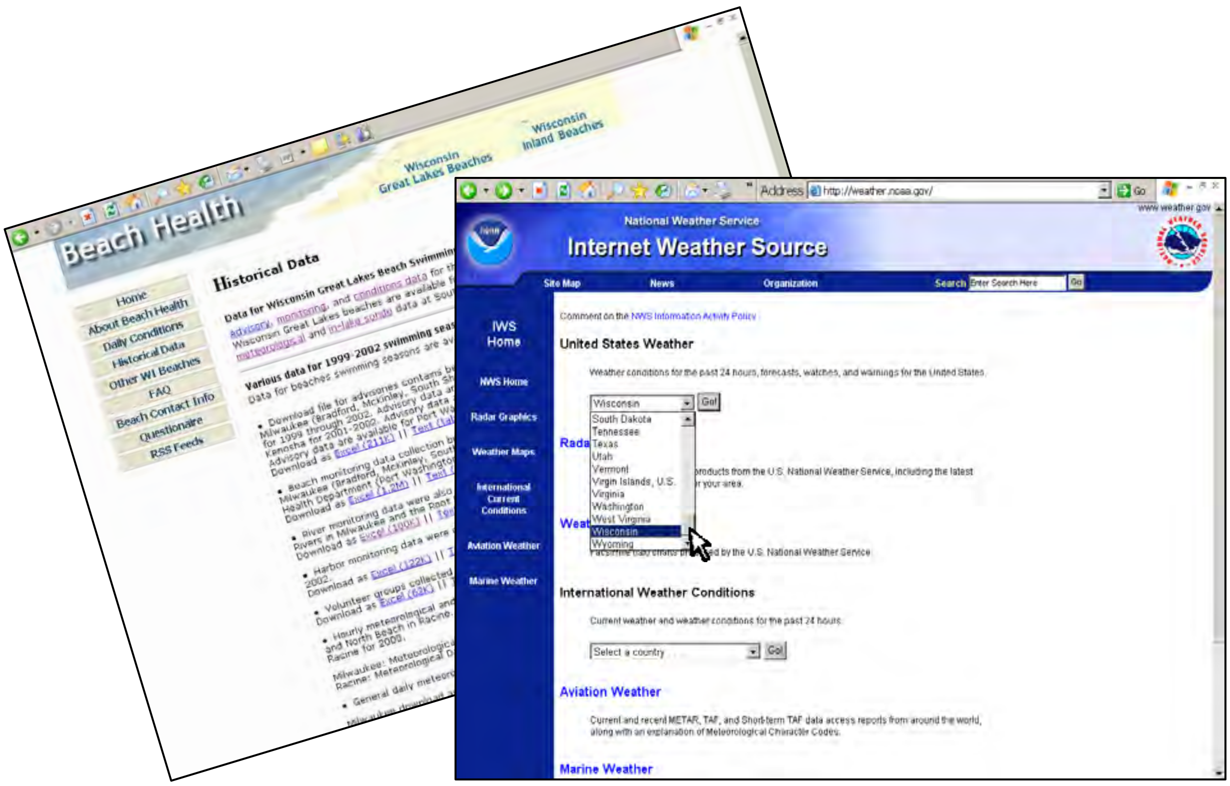
Accessing Online Data for Building and Evaluating Real-Time Models to Predict Beach Water Quality



Bureau of Science Services
Wisconsin Department of Natural Resources
P.O. Box 7921
Madison, WI 53707-7921

Miscellaneous Publication PUB-SS-1063 2009

Summary: This document lists automated data collection and reporting systems that can be accessed via the World-Wide Web and used to provide inputs to multivariate statistical models. Operated by the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS), these online systems provide near real-time data, as well as archives of historic data, on variables such as temperature, hourly rainfall, wind speed and direction, lake levels, and stream flow. Additionally, USGS operates an online archive of beach-specific monitoring and conditions data for Wisconsin. These various databases are listed in this report, along with instructions for downloading and formatting the data contained therein.

The U.S. Environmental Protection Agency supported this work, in part, with Federal Assistance Agreement No. X700E54401-0/1 (Forecasting Bacterial Exposure in Beach Settings). Points of view expressed in this report do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names and commercial products does not constitute endorsement of their use.

Accessing Online Data for Building and Evaluating Real-Time Models to Predict Beach Water Quality

Adam C. Mednick
Bureau of Science Services
Wisconsin Department of Natural Resources
P.O. Box 7921
Madison, WI 53707-7921

Adam.Mednick@Wisconsin.gov

September 2009

Contents

Note on the Currency of Web Resources

Neither the Wisconsin DNR nor the U.S. EPA maintains any of the Web resources described in this report. All of the websites, databases, mapping portals, and other Web resources included herein were active and current as of publication in 2009. By their nature, these systems are dynamic, and modifications and improvements are likely to occur over time. Modifications could include new features and functionality, changed user interfaces, and new Web addresses. Some of these systems eventually may be supplemented or replaced by newer systems not in operation at the time we prepared this publication.

Accessing Online Data for Building and Evaluating Real-Time Models to Predict Beach Water Quality

1. Introduction

In order to build a multivariate statistical model for predicting pathogen indicator concentrations at your beach – using software such as U.S. EPA's *Virtual Beach* or another statistical package – you will need regular monitoring data on your "response" variable (i.e. *E. coli* concentration), together with data on two or more "explanatory" variables measured at approximately the same time and, to the extent possible, the same location. Commonly used explanatory variables include:

Near Shore Conditions
- Wave height
- Turbidity
- Lake current speed and direction
- Water temperature
- Lake level

Weather Conditions
- Antecedent rainfall
- Wind speed and direction
- Air temperature
- Cloud cover

Onshore Conditions
- Number of bathers
- Presence of algae
- Number of gulls

Watershed Conditions
- Stream flow

A number of automated data collection/reporting systems exist that can be accessed via the World-Wide Web and used to provide inputs to your model. Operated by the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS), these systems provide near real-time data, as well as archives of historic data, on variables such as temperature, hourly rainfall, wind speed and direction, lake levels, and stream flow. Additionally, USGS operates an online archive of beach-specific monitoring and conditions data for Wisconsin. These various databases are listed in this report, along with instructions for downloading and formatting the data contained therein.

2. Data / Model Requirements

For guidance on building and evaluating models see *Procedures for Developing Models to Predict Exceedances of Recreational Water-Quality Standards at Coastal Beaches* by Donna Francy and Robert Darner (http://pubs.usgs.gov/tm/2006/tm6b5). These authors recommend assembling at least two seasons' worth of data collected four times per week to ensure that models capture seasonal as well as short-term variability in beach conditions. This level of sampling results in approximately 120 data points (i.e. days) with which to build a predictive model. The minimum number of data points required to build a satisfactory model, however, varies from beach to beach and depends on your own criteria on what constitutes a satisfactory model (e.g., the acceptable percentage of incorrect predictions of exceedances and non-exceedances). This, in turn, will depend on your purpose for building the model. If you intend to use outputs to directly inform decisions on whether or not to post swim advisories your criteria likely will be more stringent than if you plan to use the model as a screening tool for determining when to employ rapid water quality assessments.

Francy and Darner note that moderately-contaminated beaches are typically easier to model than beaches with few exceedances. Bear in mind, however, that ***all beaches are unique***. Your beach may prove easier to model with fewer data points, fewer explanatory variables, and seemingly less accurate data collection procedures than another beach with a similar frequency of contamination events. The opposite may also occur. Model validation (i.e. the comparison of model-predicted water quality to observed conditions) is the ultimate arbiter of whether or not you have chosen the right variables, collected enough data points, or have otherwise done a good job of building and refining your model using *Virtual Beach*, or whatever other statistical tool is at your disposal. For a comparative description of different metrics used to assess the predictive power of statistical models for beach water quality see the paper by Zhongfu Ge and Walter Frick[1]

3. Data Collection and Organization

Generally speaking, explanatory data collected on site are preferable to data collected off site. Where these are not available, however, data collected in the vicinity of the beach may suffice – depending on the explanatory variable in question, the distance involved, and the characteristics of your beach and its surroundings. Hourly rainfall collected at a regional airport several miles inland, for example, may prove to be as good of a predictor of *E. coli* concentrations as rainfall measured on the beach itself. Wind speed and direction, on the other hand, are more likely to differ between the beach and

[1] Zhongfu Ge and Walter Frick. 2007. Some Statistical Issues Related to Multiple Linear Regression Modeling of Beach Bacteria Concentrations. *Environmental Research* 103(3):358-364. See www.elsevier.com/locate/envres.

an inland measuring station. Ultimately, partial regression slope coefficients and associated measures of statistical significance, such as *p* values – which are computed automatically by *Virtual Beach* and most other statistical software – will tell you whether or not data collected offsite are useful predictors or not.

Standardized beach sanitary surveys (www.epa.gov/waterscience/beaches/sanitarysurvey) provide a useful means of organizing data on conditions that can influence water quality at your beach. In addition to helping you identify contamination sources for potential mitigation, these surveys provide a systematic approach for compiling data that can be used to build and update your predictive model. Conversely, statistical models can aid in source-identification efforts, by estimating the independent relationships between different contributing factors and pathogen indicator levels – as measured by partial regression slope coefficients.

4. Using Online Sources vs. Field-Collected Data

The online sources listed below include *real-time* data (i.e. data collected and reported continuously), *historic* data (i.e. archived data sets), or both. Historic data can provide a good jumping-off point for the model-building process if you have one or more seasons of regular data on pathogen indicator concentrations, but have not yet started collecting data on beach conditions through a sanitary survey, or otherwise have only recently begun that process. Alternatively, if you have already collected onsite data on a variable for which archived public data also exist, you can substitute the archived data within the model-building process and compare the results to help determine whether continued onsite collection for that particular variable is worth the effort.

When building your model you ***cannot mix-and-match data sources*** for the same variable. For example, if you build a model using historic wind direction data downloaded from NOAA, switching to your own field-collected data later-on will require that you remove from your model all data points (days) preceding the first onsite collection. Because of differences in methods and location, field-collected data and online sources technically constitute different variables. Mixing NOAA- or USGS-collected data with field-collected data is likely to lower the predictive power of your model. Transitioning from online data sources for a particular variable to field-collected data, therefore, can take one or more seasons to implement. On the other hand, transitioning from field-collected data to online sources – which typically are archived in addition to being reported in real-time – can occur immediately.

5. Accessing Online Data Sources

Use these online databases to locate, view, and download data for building and executing your model. Detailed instructions follow. In addition to these public databases, the commercial Weather Underground site (www.wunderground.com) provides access to meteorological data that can be used for model building purposes. Many of the datasets accessible through Weather Underground originate from public sources listed below. In addition, this site streams data from citizen monitors ("personal weather stations"). These data are not quality assured and the distinctions between different primary sources underlying the user interface are not always clear. As such, users should exercise caution when using this site for building and implementing predictive models. **Table 1** lists data sources by variable, provides web links, and identifies report sections that address each source.

A. *E. coli* Monitoring Data and Onshore Conditions

B. Meteorological and Nearshore Conditions

C. Watershed Conditions

Table 1. *Online Data Sources.**

Variable†	Historic Data (for Model-Building)	Real-time Data (for "Nowcasting")
Air Temperature	National Data Buoy Center www.ndbc.noaa.gov (B.6.6 - B.6.16)	National Data Buoy Center www.ndbc.noaa.gov (B.6.1 - B.6.5)
	Online Climate Data Directory – *Hourly* www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly (B.4.1 - B.4.22)	Internet Weather Source http://weather.noaa.gov (B.2.1 - B.2.5)
Lake Level	Tides & Currents http://tidesandcurrents.noaa.gov/station_retrieve.shtml (B.7.1 - B.7.16)	Tides & Currents http://tidesandcurrents.noaa.gov/station_retrieve.shtml (B.7.17 - B.7.31)
Rainfall, Hourly	Online Climate Data Directory – *Hourly* www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly (B.4.1 - B.4.22)	Internet Weather Source http://weather.noaa.gov (B.2.1 - B.2.5)
Rainfall, 24-Hour‡	Online Climate Data Directory – *Hourly* www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly (B.4.23 - B.4.25)	Internet Weather Source http://weather.noaa.gov (B.2.1 - B.2.5)
	Online Climate Data Directory – *Daily* www.ncdc.noaa.gov/oa/climate/climatedata.html#daily (B.5.1 - B.5.18)	Daily Hydrometeorological Data www.crh.noaa.gov/product_sites.php?site=MKX&product=HYD (B.3.1 - B.3.3)
Sky Conditions	Online Climate Data Directory – *Hourly* www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly (B.4.1 - B.4.22)	Internet Weather Source http://weather.noaa.gov (B.2.1 - B.2.5)

(Table Continues)

Table 1. *Online Data Sources, Continued.* *

Variable†	**Historic Data** (for Model-Building)	**Real-time Data** (for "Nowcasting")
River/Stream Discharge	Waterwatch http://water.usgs.gov/waterwatch (C.1.24 - C.1.40)	Waterwatch http://water.usgs.gov/waterwatch (C.1.1 - C.1.23)
Water Temp.	National Data Buoy Center www.ndbc.noaa.gov (B.6.6 - B.6.16)	National Data Buoy Center www.ndbc.noaa.gov (B.6.1 - B.6.5)
Wave Height	National Data Buoy Center www.ndbc.noaa.gov (B.6.6 - B.6.16)	National Data Buoy Center www.ndbc.noaa.gov (B.6.1 - B.6.5)
Wind Speed/ Direction	National Data Buoy Center www.ndbc.noaa.gov (B.6.6 - B.6.16)	National Data Buoy Center www.ndbc.noaa.gov (B.6.1 - B.6.5)
	Online Climate Data Directory – *Hourly* www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly (B.4.1 - B.4.22)	Internet Weather Source http://weather.noaa.gov (B.2.1 - B.2.5)

* Sources and web links are organized by variable and whether data are *historic* (required for building predictive models) or *real-time* (required for "nowcasting" current water quality conditions using predictive models). Report sections describing how to download and format data are identified in parentheses.

† Only explanatory variables are listed here. Data on your response variable (e.g., *E. coli*) should be collected at the beach. In Wisconsin, see www.wibeaches.us for historic monitoring data, as well as data on a site-specific explanatory variables ("beach conditions").

‡ 24-hour rainfall data are generally reported each day as of approximately 12:00 UTC (7:00 CDT, 8:00 EDT).

A. *E. coli* Monitoring Data and Onshore Conditions

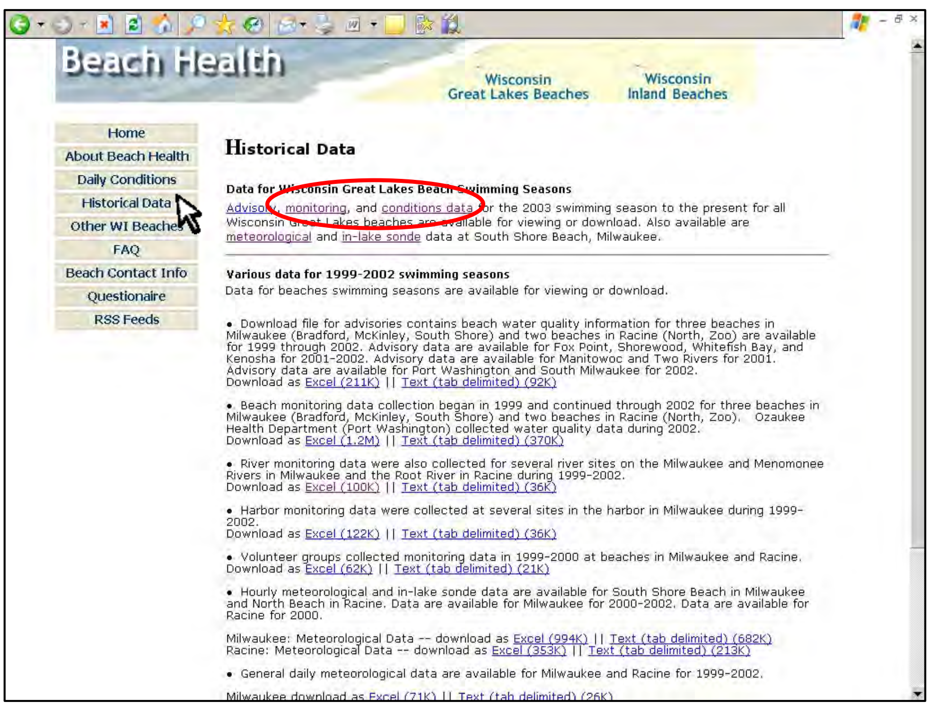
A.1. Wisconsin Beach Health

Provides access to a downloadable archive of historic data on *E. coli* monitoring results and various beach conditions collected at individual Great Lakes beaches throughout Wisconsin. Pertinent variables include *E. coli*, water temperature, number of bathers, wave height, water clarity, wind speed, wind direction, and rainfall.

Wisconsin Beach Health is the one of two sites that archive historic data on *E. coli* monitoring, as well as beach conditions in the Great Lakes. The other is the Minnesota Lake Superior Beach Monitoring Program-Data Viewer, which can be found at www.mnbeaches.org/gmap/DataViewer.html.

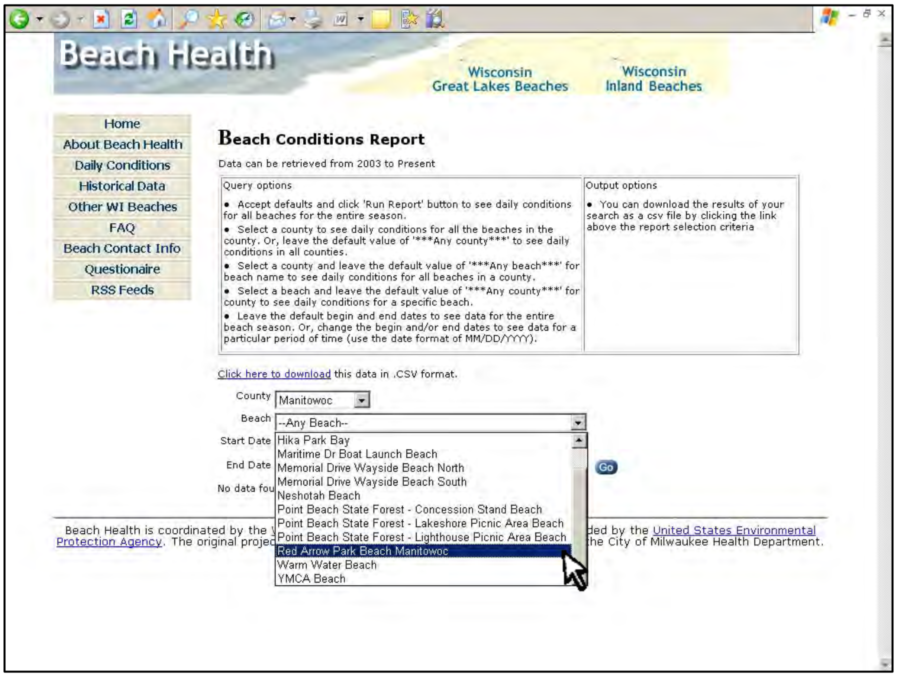
A.1.1. Open the main page (www.wibeaches.us) and click the “Historical Data” link on the left-hand navigation menu.

A.1.2. On the “Historical Data” page, click the link for either “monitoring” or “conditions data” in the first paragraph.

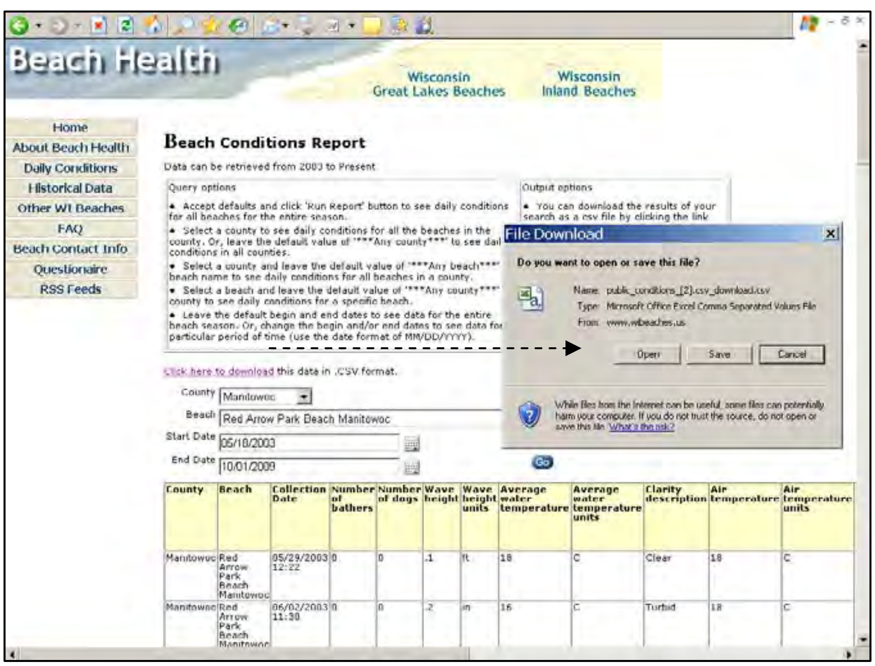


A.1.3. Depending on your selection, this will open either the “Beach Act Monitoring Report” page or the “Beach Conditions Report” page.

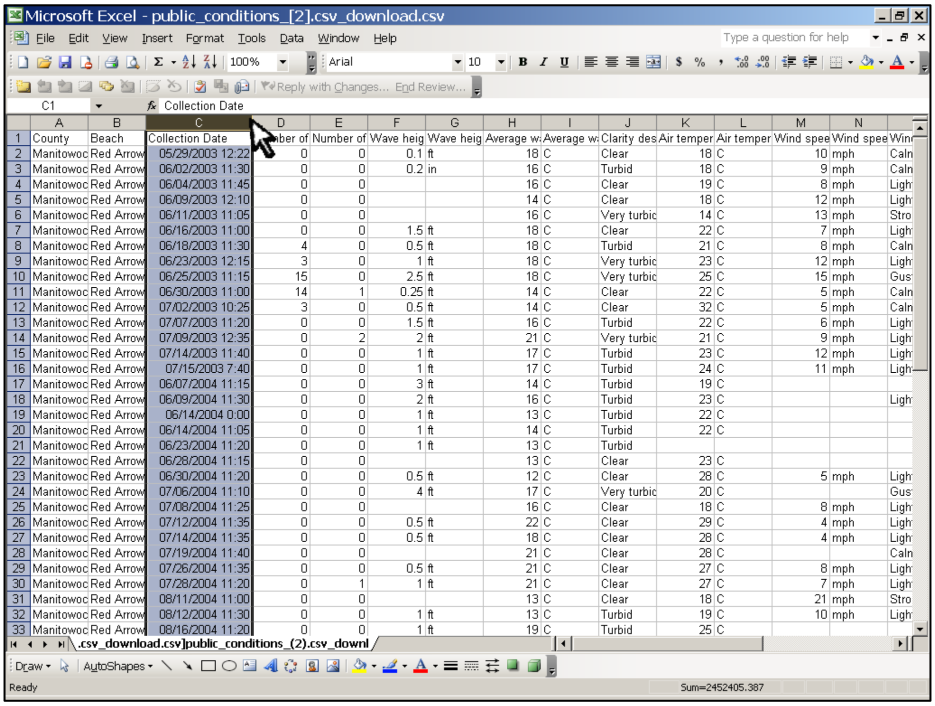
A.1.4. On either the “Beach Act Monitoring Report” or the “Beach Conditions Report” page, use the pull-down menus to select the “County”, “Beach”, “Start Date”, and “End Date” of interest, and then click the “Go” button.



A.1.5. A table will appear for online viewing. Click the “Click here to download” link to download data as comma-separated values (.csv). You can either “Open” the data directly (will launch in MS Excel) or “Save” to a local directory and folder.



A.1.6. Once data are open in MS Excel, you can expand, delete, or re-name columns as desired. In order to view the dates and times under "Collection Date", click on the right edge of the column header (C) and drag right until the values appear.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	County	Beach	Collection Date	ber of	Number of	Wave heig	Wave heig	Average w	Average w	Clarity des	Air temper	Air temper	Wind spee	Wind spee	Win
2	Manitowoc	Red Arrow	05/29/2003 12:22	0	0	0.1	ft	18	C	Clear	18	C	10	mph	Caln
3	Manitowoc	Red Arrow	06/02/2003 11:30	0	0	0.2	in	16	C	Turbid	18	C	9	mph	Caln
4	Manitowoc	Red Arrow	06/04/2003 11:45	0	0			16	C	Clear	19	C	8	mph	Ligh
5	Manitowoc	Red Arrow	06/09/2003 12:10	0	0			14	C	Clear	18	C	12	mph	Ligh
6	Manitowoc	Red Arrow	06/11/2003 11:05	0	0			16	C	Very turbic	14	C	13	mph	Stro
7	Manitowoc	Red Arrow	06/16/2003 11:00	0	0	1.5	ft	18	C	Clear	22	C	7	mph	Ligh
8	Manitowoc	Red Arrow	06/18/2003 11:30	4	0	0.5	ft	18	C	Turbid	21	C	8	mph	Caln
9	Manitowoc	Red Arrow	06/23/2003 12:15	3	0	1	ft	18	C	Very turbic	23	C	12	mph	Ligh
10	Manitowoc	Red Arrow	06/25/2003 11:15	15	0	2.5	ft	18	C	Very turbic	25	C	15	mph	Gus
11	Manitowoc	Red Arrow	06/30/2003 11:00	14	1	0.25	ft	14	C	Clear	22	C	5	mph	Caln
12	Manitowoc	Red Arrow	07/02/2003 10:25	3	0	0.5	ft	14	C	Clear	32	C	5	mph	Caln
13	Manitowoc	Red Arrow	07/07/2003 11:20	0	0	1.5	ft	16	C	Turbid	22	C	6	mph	Ligh
14	Manitowoc	Red Arrow	07/09/2003 12:35	0	2	2	ft	21	C	Very turbic	21	C	9	mph	Ligh
15	Manitowoc	Red Arrow	07/14/2003 11:40	0	0	1	ft	17	C	Turbid	23	C	12	mph	Ligh
16	Manitowoc	Red Arrow	07/15/2003 7:40	0	0	1	ft	17	C	Turbid	24	C	11	mph	Ligh
17	Manitowoc	Red Arrow	06/07/2004 11:15	0	0	3	ft	14	C	Turbid	19	C			
18	Manitowoc	Red Arrow	06/09/2004 11:30	0	0	2	ft	16	C	Turbid	23	C			Ligh
19	Manitowoc	Red Arrow	06/14/2004 0:00	0	0	1	ft	13	C	Turbid	22	C			
20	Manitowoc	Red Arrow	06/14/2004 11:05	0	0	1	ft	14	C	Turbid	22	C			
21	Manitowoc	Red Arrow	06/23/2004 11:20	0	0	1	ft	13	C	Turbid					
22	Manitowoc	Red Arrow	06/28/2004 11:15	0	0			13	C	Clear	23	C			
23	Manitowoc	Red Arrow	06/30/2004 11:20	0	0	0.5	ft	12	C	Clear	28	C	5	mph	Ligh
24	Manitowoc	Red Arrow	07/06/2004 11:10	0	0	4	ft	17	C	Very turbic	20	C			Gus
25	Manitowoc	Red Arrow	07/08/2004 11:25	0	0			16	C	Clear	18	C	8	mph	Ligh
26	Manitowoc	Red Arrow	07/12/2004 11:35	0	0	0.5	ft	22	C	Clear	29	C	4	mph	Ligh
27	Manitowoc	Red Arrow	07/14/2004 11:35	0	0	0.5	ft	18	C	Clear	28	C	4	mph	Ligh
28	Manitowoc	Red Arrow	07/19/2004 11:40	0	0			21	C	Clear	28	C			Caln
29	Manitowoc	Red Arrow	07/26/2004 11:35	0	0	0.5	ft	21	C	Clear	27	C	8	mph	Ligh
30	Manitowoc	Red Arrow	07/28/2004 11:20	0	1	1	ft	21	C	Clear	27	C	7	mph	Ligh
31	Manitowoc	Red Arrow	08/11/2004 11:00	0	0			13	C	Clear	18	C	21	mph	Stro
32	Manitowoc	Red Arrow	08/12/2004 11:30	0	0	1	ft	13	C	Turbid	19	C	10	mph	Ligh
33	Manitowoc	Red Arrow	08/16/2004 11:20	0	0	1	ft	19	C	Turbid	25	C			

For other formatting and editing procedures, see the MS Excel help menu.

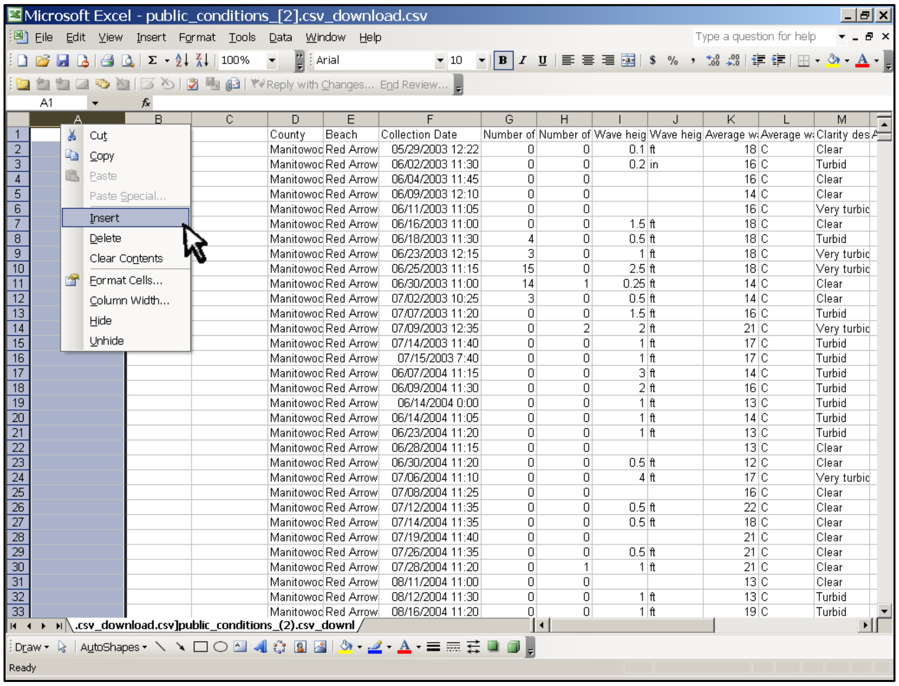
Note: If data for the same variable are reported in two or more different units (e.g., English and metric), they must be converted to common units before building your model. (It does not matter which units you choose, as long as they are consistent.) Failure to do so will reduce significantly the predictive power of your model.

Common conversions include:

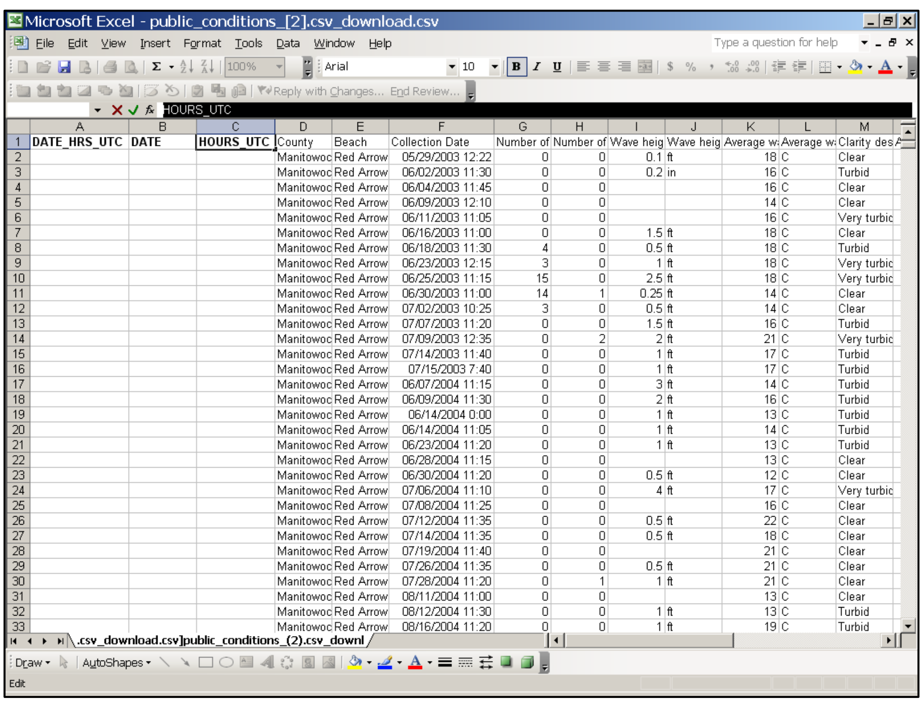
- Temperature: *Degrees F* = ([*Degrees C*] * 9/5) + 32
- Rainfall: *Inches* = *Centimeters* * 0.393700787
- Wave ht./Lake level: *Feet* = *Meters* * 0.3048
- Wind speed: *M.P.H.* (*Knots*) = *Meters-per-Second* * 0.44704

Note: In order to include them in your model, categorical (*qualitative*) variables for clarity, wind direction, and weather must be converted into separate "0/1" (absence/presence) variables. For example, clarity (which has 3 possible values: "Clear", "Turbid", and "Very Turbid") should be converted into *two* 0/1 variables: "Turbid" (0/1) and "Very Turbid" (0/1). In any such conversion, one of the possible conditions must be excluded. In this example, "Clear" is the excluded variable, with clear conditions represented by a combined value of 0 in the "Turbid" and "Very Turbid" variables.

A.1.7. Insert a new column by right-clicking on the header for column A and selecting “Insert” from the menu. Repeat this step to create a second- and a third new column.



A.1.8. To name the three new columns you created in A.1.7, first click on cell A1 (column A, row 1) and type “DATE_HRS_UTC”. Next, click on cell B1 and type “DATE”. Finally, click on cell C1 and type “HOURS_UTC”.



Note: Wisconsin Beach Health data are reported in Central Daylight Time (CDT). Other sources listed in this report use Coordinated Universal Time (UTC), also known as "Zulu Time," which is effectively the same as Greenwich Mean Time. A UTC converter is available at http://hurricanes.noaa.gov/zulu-utc.html. For the Central time zone, UTC = CDT + 5 hours.

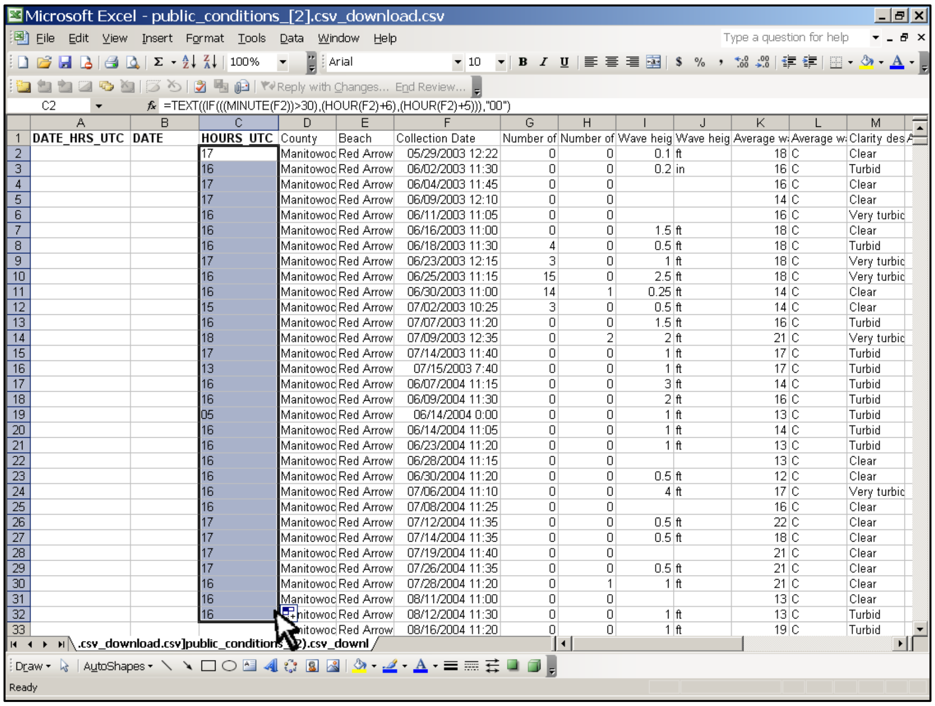
Note: As you build your input data set for *Virtual Beach* or another statistical package, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to "join" the tables into a single table of just those dates and times for which *E. coli* was measured.

Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and "joining" tables using queries in MS Access, see Section 6 of this report, "Formatting Data for Model-Building.")

A.1.9 To calculate the hour-of-day in UTC (column C: “HOURS_UTC”), click on cell C2 and type the following function:

=TEXT((IF(((MINUTE(F2))>30),(HOUR(F2)+6),(HOUR(F2)+5))),"00")

Hit “Enter.” The result should be a two-digit, rounded value for the hour-of-day that is between 5 and 6 hours ahead of Central Daylight Time (depending on the minutes listed under “Collection Date” in column F). Copy this function by clicking in the lower right corner of cell C2 and dragging to the end of the table.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DATE_HRS_UTC	DATE	HOURS_UTC	County	Beach	Collection Date	Number of	Number of	Wave heig	Wave heig	Average w	Average w	Clarity des
2			17	Manitowoc	Red Arrow	05/29/2003 12:22	0	0	0.1	ft	18	C	Clear
3			16	Manitowoc	Red Arrow	06/02/2003 11:30	0	0	0.2	in	16	C	Turbid
4			17	Manitowoc	Red Arrow	06/04/2003 11:45	0	0			16	C	Clear
5			17	Manitowoc	Red Arrow	06/09/2003 12:10	0	0			14	C	Clear
6			16	Manitowoc	Red Arrow	06/11/2003 11:05	0	0			16	C	Very turbic
7			16	Manitowoc	Red Arrow	06/16/2003 11:00	0	0	1.5	ft	18	C	Clear
8			16	Manitowoc	Red Arrow	06/18/2003 11:30	4	0	0.5	ft	18	C	Turbid
9			17	Manitowoc	Red Arrow	06/23/2003 12:15	3	0	1	ft	18	C	Very turbic
10			16	Manitowoc	Red Arrow	06/25/2003 11:15	15	0	2.5	ft	18	C	Very turbic
11			16	Manitowoc	Red Arrow	06/30/2003 11:00	14	1	0.25	ft	14	C	Clear
12			15	Manitowoc	Red Arrow	07/02/2003 10:25	3	0	0.5	ft	14	C	Clear
13			16	Manitowoc	Red Arrow	07/07/2003 11:20	0	0	1.5	ft	16	C	Turbid
14			18	Manitowoc	Red Arrow	07/09/2003 12:35	0	2	2	ft	21	C	Very turbic
15			17	Manitowoc	Red Arrow	07/14/2003 11:40	0	0	1	ft	17	C	Turbid
16			13	Manitowoc	Red Arrow	07/15/2003 7:40	0	0	1	ft	17	C	Turbid
17			16	Manitowoc	Red Arrow	06/07/2004 11:15	0	0	3	ft	14	C	Turbid
18			16	Manitowoc	Red Arrow	06/09/2004 11:30	0	0	2	ft	16	C	Turbid
19			05	Manitowoc	Red Arrow	06/14/2004 0:00	0	0	1	ft	13	C	Turbid
20			16	Manitowoc	Red Arrow	06/14/2004 11:05	0	0	1	ft	14	C	Turbid
21			16	Manitowoc	Red Arrow	06/23/2004 11:20	0	0	1	ft	13	C	Turbid
22			16	Manitowoc	Red Arrow	06/28/2004 11:15	0	0			13	C	Clear
23			16	Manitowoc	Red Arrow	06/30/2004 11:20	0	0	0.5	ft	12	C	Clear
24			16	Manitowoc	Red Arrow	07/06/2004 11:10	0	0	4	ft	17	C	Very turbic
25			16	Manitowoc	Red Arrow	07/08/2004 11:25	0	0			16	C	Clear
26			17	Manitowoc	Red Arrow	07/12/2004 11:35	0	0	0.5	ft	22	C	Clear
27			17	Manitowoc	Red Arrow	07/14/2004 11:35	0	0	0.5	ft	18	C	Clear
28			17	Manitowoc	Red Arrow	07/19/2004 11:40	0	0			21	C	Clear
29			17	Manitowoc	Red Arrow	07/26/2004 11:35	0	0	0.5	ft	21	C	Clear
30			16	Manitowoc	Red Arrow	07/28/2004 11:20	0	1	1	ft	21	C	Clear
31			16	Manitowoc	Red Arrow	08/11/2004 11:00	0	0			13	C	Clear
32			16	nitowoc	Red Arrow	08/12/2004 11:30	0	0	1	ft	13	C	Turbid
33				nitowoc	Red Arrow	08/16/2004 11:20	0	0	1	ft	19	C	Turbid

⚠ **Note:** In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Also, be sure to include period-separators between the year, month, day, and hour (e.g., 2003.05.29.12 as opposed to 2003052912). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

A.1.10. To calculate "DATE" (column B), click on cell B2 and type the following function:

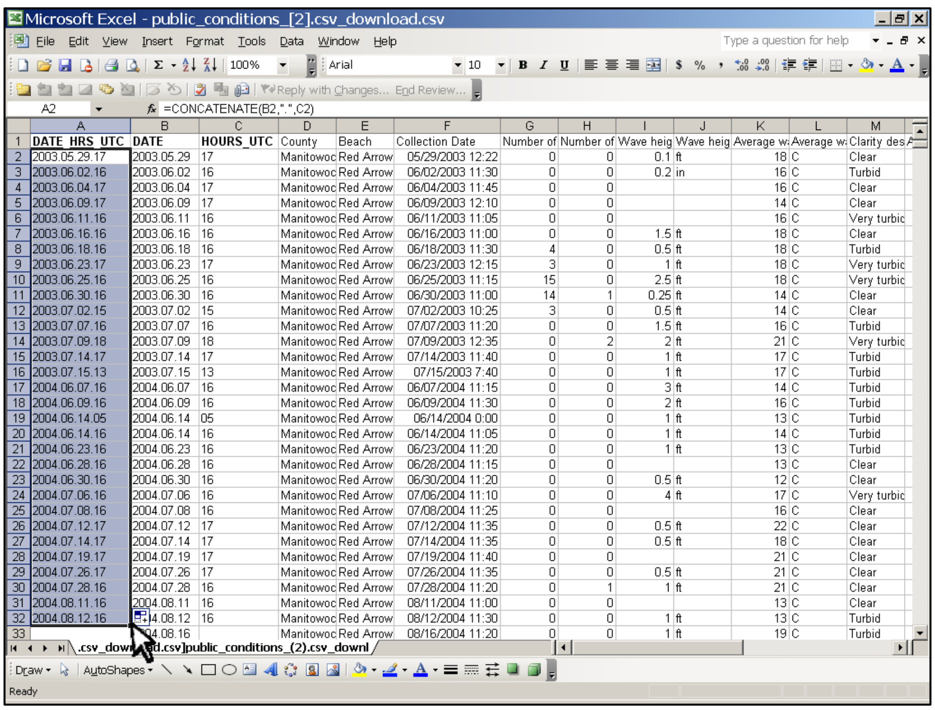
=CONCATENATE((YEAR(F2)),".",(TEXT((MONTH(F2)),"00")),".",(TEXT((DAY(F2)),"00")))

Hit "Enter." This should return a date value (e.g., "2003.05.29" for May 29, 2003). Copy this function by clicking in the lower right corner of cell B2 and dragging to the end of the table.

A.1.11. To calculate date-plus-hour in UTC ("DATE_HRS_UTC"; column A), click on cell A2 and type the following function:

=CONCATENATE(B2,".",C2)

Hit "Enter." This should return a date-plus-hour (e.g., "2003.05.29.13" for 1300 hours UTC on May 29, 2003). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DATE_HRS_UTC	DATE	HOURS_UTC	County	Beach	Collection Date	Number of	Number of	Wave heig	Wave heig	Average w	Average w	Clarity des
2	2003.05.29.17	2003.05.29	17	Manitowoc	Red Arrow	05/29/2003 12:22	0	0	0.1	ft	18	C	Clear
3	2003.06.02.16	2003.06.02	16	Manitowoc	Red Arrow	06/02/2003 11:30	0	0	0.2	in	16	C	Turbid
4	2003.06.04.17	2003.06.04	17	Manitowoc	Red Arrow	06/04/2003 11:45	0	0			16	C	Clear
5	2003.06.09.17	2003.06.09	17	Manitowoc	Red Arrow	06/09/2003 12:10	0	0			14	C	Clear
6	2003.06.11.16	2003.06.11	16	Manitowoc	Red Arrow	06/11/2003 11:05	0	0			16	C	Very turbic
7	2003.06.16.16	2003.06.16	16	Manitowoc	Red Arrow	06/16/2003 11:00	0	0	1.5	ft	18	C	Clear
8	2003.06.18.16	2003.06.18	16	Manitowoc	Red Arrow	06/18/2003 11:30	4	0	0.5	ft	18	C	Turbid
9	2003.06.23.17	2003.06.23	17	Manitowoc	Red Arrow	06/23/2003 12:15	3	0	1	ft	18	C	Very turbic
10	2003.06.25.16	2003.06.25	16	Manitowoc	Red Arrow	06/25/2003 11:15	15	0	2.5	ft	18	C	Very turbic
11	2003.06.30.16	2003.06.30	16	Manitowoc	Red Arrow	06/30/2003 11:00	14	1	0.25	ft	14	C	Clear
12	2003.07.02.15	2003.07.02	15	Manitowoc	Red Arrow	07/02/2003 10:25	3	0	0.5	ft	14	C	Clear
13	2003.07.07.16	2003.07.07	16	Manitowoc	Red Arrow	07/07/2003 11:20	0	0	1.5	ft	16	C	Turbid
14	2003.07.09.18	2003.07.09	18	Manitowoc	Red Arrow	07/09/2003 12:35	0	2	2	ft	21	C	Very turbic
15	2003.07.14.17	2003.07.14	17	Manitowoc	Red Arrow	07/14/2003 11:40	0	0	1	ft	17	C	Turbid
16	2003.07.15.13	2003.07.15	13	Manitowoc	Red Arrow	07/15/2003 7:40	0	0	1	ft	17	C	Turbid
17	2004.06.07.16	2004.06.07	16	Manitowoc	Red Arrow	06/07/2004 11:15	0	0	3	ft	14	C	Turbid
18	2004.06.09.16	2004.06.09	16	Manitowoc	Red Arrow	06/09/2004 11:30	0	0	2	ft	16	C	Turbid
19	2004.06.14.05	2004.06.14	05	Manitowoc	Red Arrow	06/14/2004 0:00	0	0	1	ft	13	C	Turbid
20	2004.06.14.16	2004.06.14	16	Manitowoc	Red Arrow	06/14/2004 11:05	0	0	1	ft	14	C	Turbid
21	2004.06.23.16	2004.06.23	16	Manitowoc	Red Arrow	06/23/2004 11:20	0	0	1	ft	13	C	Turbid
22	2004.06.28.16	2004.06.28	16	Manitowoc	Red Arrow	06/28/2004 11:15	0	0			13	C	Clear
23	2004.06.30.16	2004.06.30	16	Manitowoc	Red Arrow	06/30/2004 11:20	0	0	0.5	ft	12	C	Clear
24	2004.07.06.16	2004.07.06	16	Manitowoc	Red Arrow	07/06/2004 11:10	0	0	4	ft	17	C	Very turbic
25	2004.07.08.16	2004.07.08	16	Manitowoc	Red Arrow	07/08/2004 11:25	0	0			16	C	Clear
26	2004.07.12.17	2004.07.12	17	Manitowoc	Red Arrow	07/12/2004 11:35	0	0	0.5	ft	22	C	Clear
27	2004.07.14.17	2004.07.14	17	Manitowoc	Red Arrow	07/14/2004 11:35	0	0	0.5	ft	18	C	Clear
28	2004.07.19.17	2004.07.19	17	Manitowoc	Red Arrow	07/19/2004 11:40	0	0			21	C	Clear
29	2004.07.26.17	2004.07.26	17	Manitowoc	Red Arrow	07/26/2004 11:35	0	0	0.5	ft	21	C	Clear
30	2004.07.28.16	2004.07.28	16	Manitowoc	Red Arrow	07/28/2004 11:20	0	1	1	ft	21	C	Clear
31	2004.08.11.16	2004.08.11	16	Manitowoc	Red Arrow	08/11/2004 11:00	0	0			13	C	Clear
32	2004.08.12.16	4.08.12	16	Manitowoc	Red Arrow	08/12/2004 11:30	0	0	1	ft	13	C	Turbid
33		04.08.16		Manitowoc	Red Arrow	08/16/2004 11:20	0	0	1	ft	19	C	Turbid

A.1.12. Before closing the table, save it in MS Excel format. From the MS Excel pull-down menu in the upper left-hand corner, select "File" > "Save as…", browse to a directory to save in, name the file something like "[location]_conditions_[year(s)]", and select the most recent version of MS Excel as the file type (*.xls).

Notes

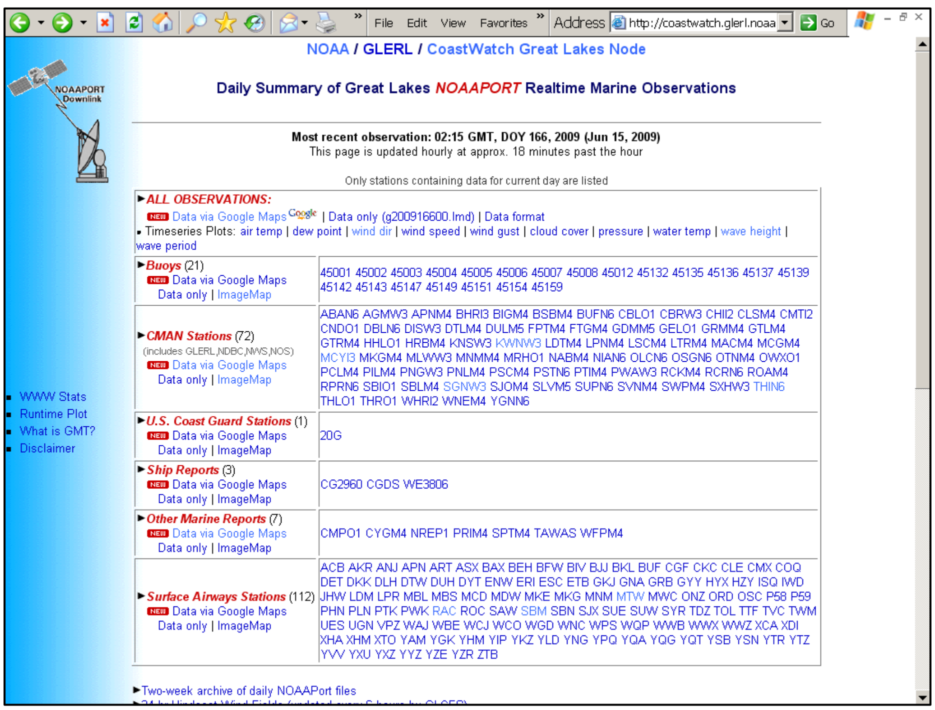
B. Meteorological and Nearshore Conditions

B.1. CoastWatch - Great Lakes Node

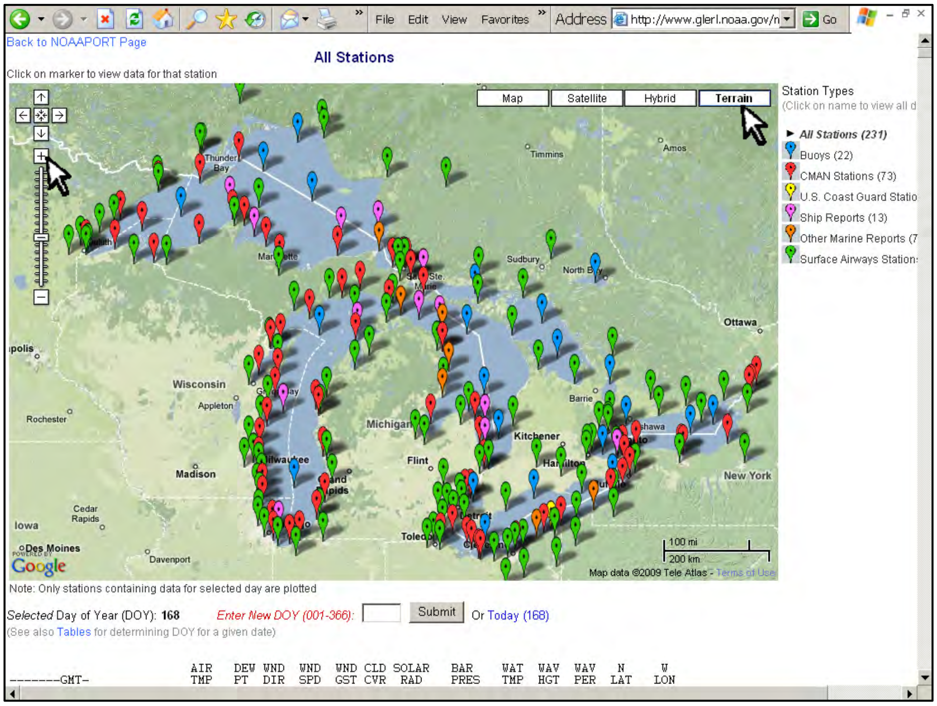
Provides access to real-time data from automated observation stations around the Great Lakes through a Google Maps interface. Includes automated airport weather observing stations (AWOS and ASOS), coastal marine automated network (C-MAN) stations, data buoys, and other sources. Depending on the source, pertinent variables may include air temperature, water temperature, wind speed, wind direction, and wave height.

CoastWatch is a good way to identify different types of observation stations in the vicinity of your beach. Accessible data are in metric units and are listed by Day-of-Year (as opposed to calendar date). Once you have located a station and determined that data collected there are useful for building your model, or for making real-time predictions, you may want to switch to one of the other data gateways listed below in sections B.2. – B.7.

B.1.1. Open the **CoastWatch** "Realtime Marine Observations" page (http://coastwatch.glerl.noaa.gov/marobs) and click the first link under "ALL OBSERVATIONS: Data via Google Maps". (If you are interested in one particular type of station, explore the Google Maps links under the individual categories.)



B.1.2. A standard Google Maps interface will open.



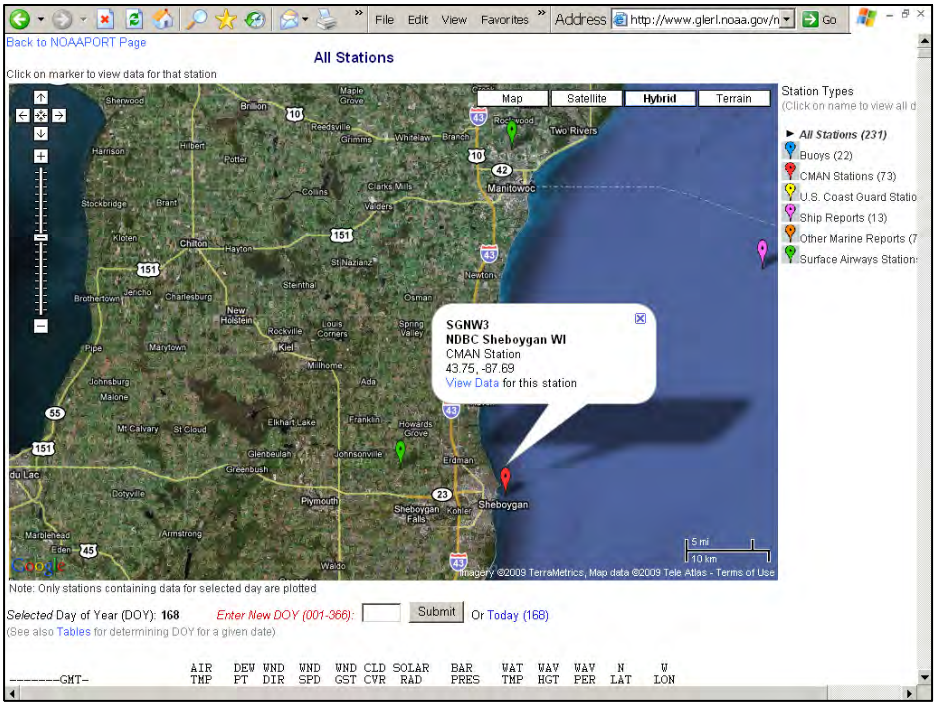
To zoom in or out, use the "+/-" bar in the upper left-hand corner of the map.

To pan, use the 4-direction arrows in the upper left-hand corner, or simply click inside the map, hold, drag, and release.

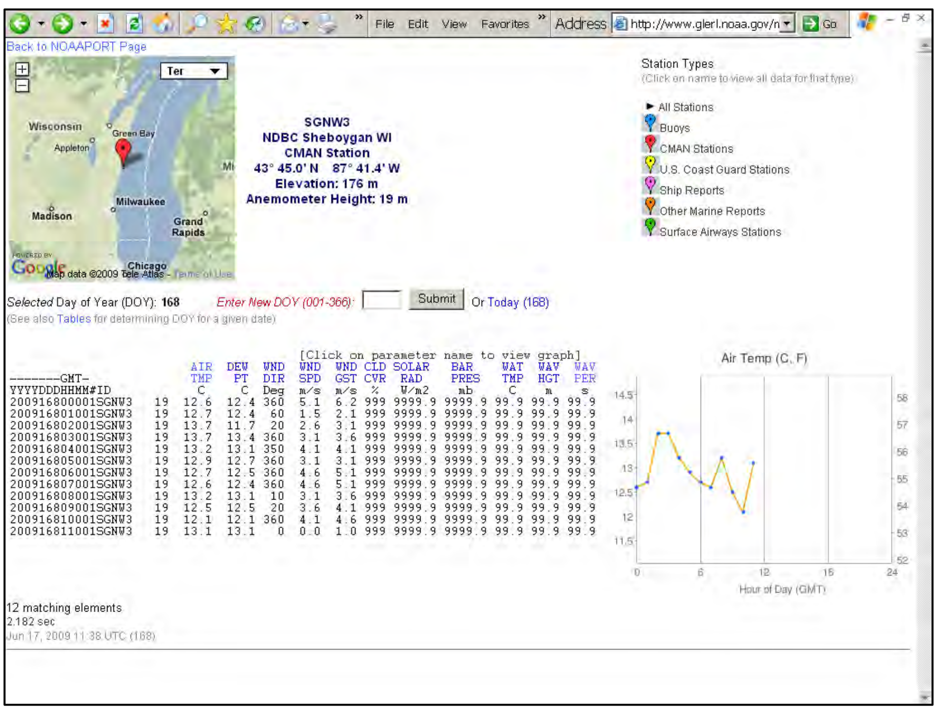
To change the background of the map, click on one of the buttons in the upper right-hand corner of the map.

- "Map" = standard map view (roads, water, parks, city and place names).
- "Satellite" = satellite imagery.
- "Hybrid" = satellite imagery with roads and city/place names on top.
- "Terrain" shows topographic relief.

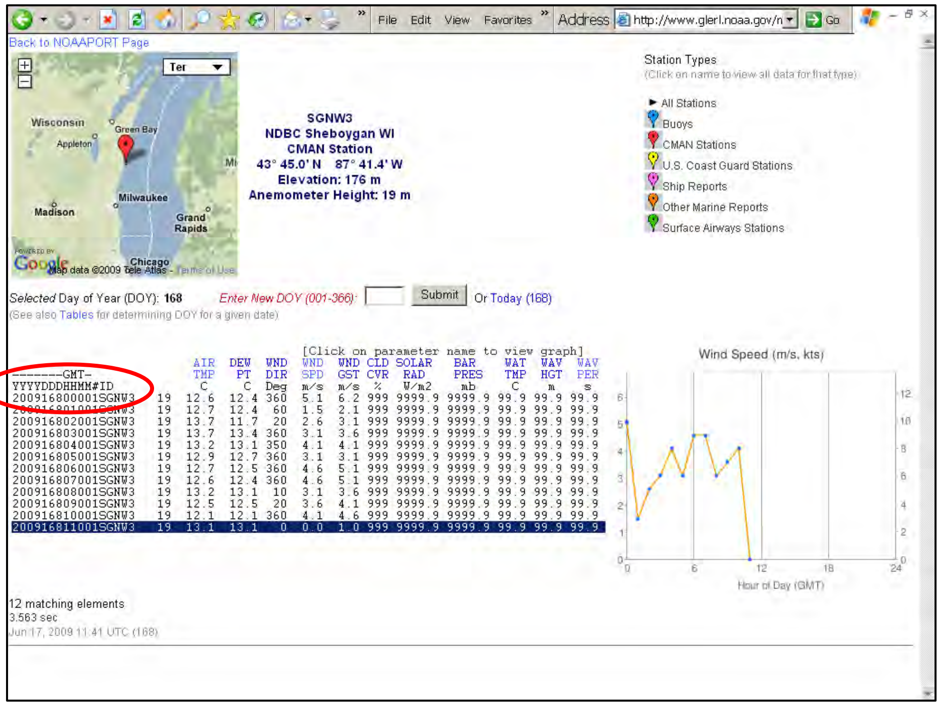
B.1.3. Once you have navigated to the vicinity of your beach, click on a point of interest. This will open a window with basic information about the selected station.



B.1.4. Clicking on the "View Data" link within the pop-up window will take you to a web page summarizing current and recent conditions at that station.



⚠ **Note:** Data are listed by year, day-of-year (DOY, 1-366), and Coordinated Universal Time (UTC). The last entry is the most recent of the list – typically within the last hour. Values of 999 represent unmeasured variables. A DOY converter is available at http://disc.gsfc.nasa.gov/julian_calendar.shtml. A UTC converter is available at http://hurricanes.noaa.gov/zulu-utc.html.



⚠ **Note:** If you leave a station data page open for an hour or longer without clicking the refresh button in your Web browser, the most recent listing will become out-of-date. Be sure to hit the refresh button if you intend to use the data as real-time predictors.

B.1.5. To view data for previous dates – limited to the same calendar year – enter a new day-of-year in the box next to "Enter New DOY (001-366)".

ⓘ For more extensive access to historic/archived data from surface weather observing stations, use the **Online Climate Data Directory** (B.5). For historic data from Great Lakes buoys or C-MAN coastal stations use the **National Data Buoy Center** (B.6).

B.2. Internet Weather Source

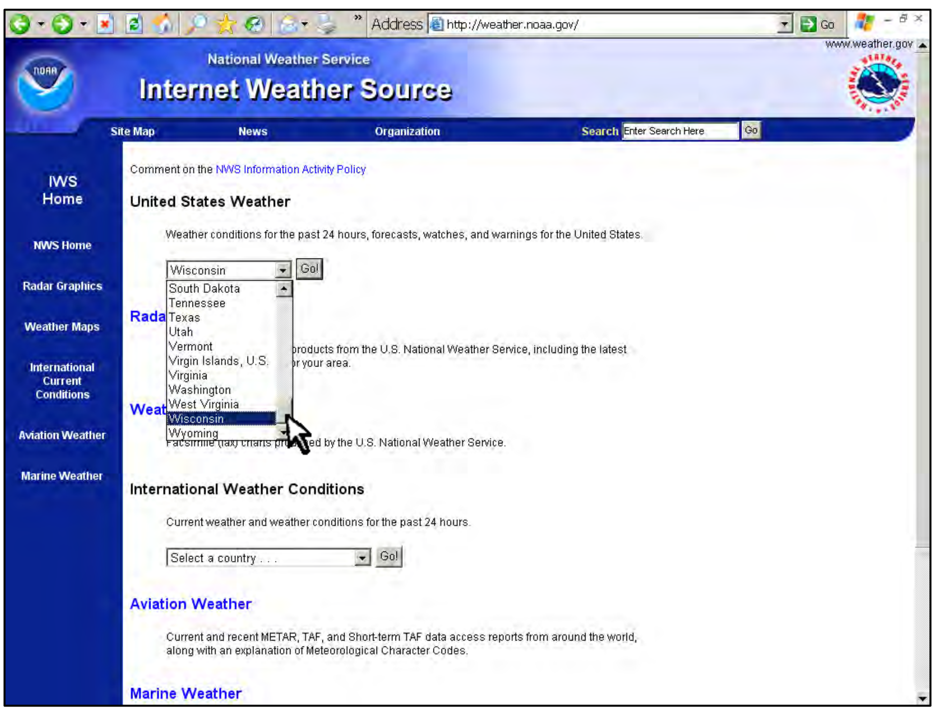
Provides quick access to real-time data from surface weather observing stations (typically located at regional airports). Pertinent variables include air temperature, wind speed, wind direction, 24-hour rainfall, and sky conditions (e.g., clear, scattered, overcast…).

The **Internet Weather Source** does not have a map interface. To locate or map an automated weather observing station in relation to your beach use the **CoastWatch Great Lakes Node** (B.1). (Click on the link to the Google Maps interface under "Surface Airways Stations" or go directly to: www.glerl.noaa.gov/marobs/php/data.php?sta=6.)

The **Internet Weather Source** does not provide access to data from Great Lakes buoys or C-MAN coastal stations. For these data, see the **National Data Buoy Center** (B.6).

Data accessible through the **Internet Weather Source** are limited to the last 24 hours only. To access historic/archived data for model-building or other purposes, use the **Online Climate Data Directory** (B.5).

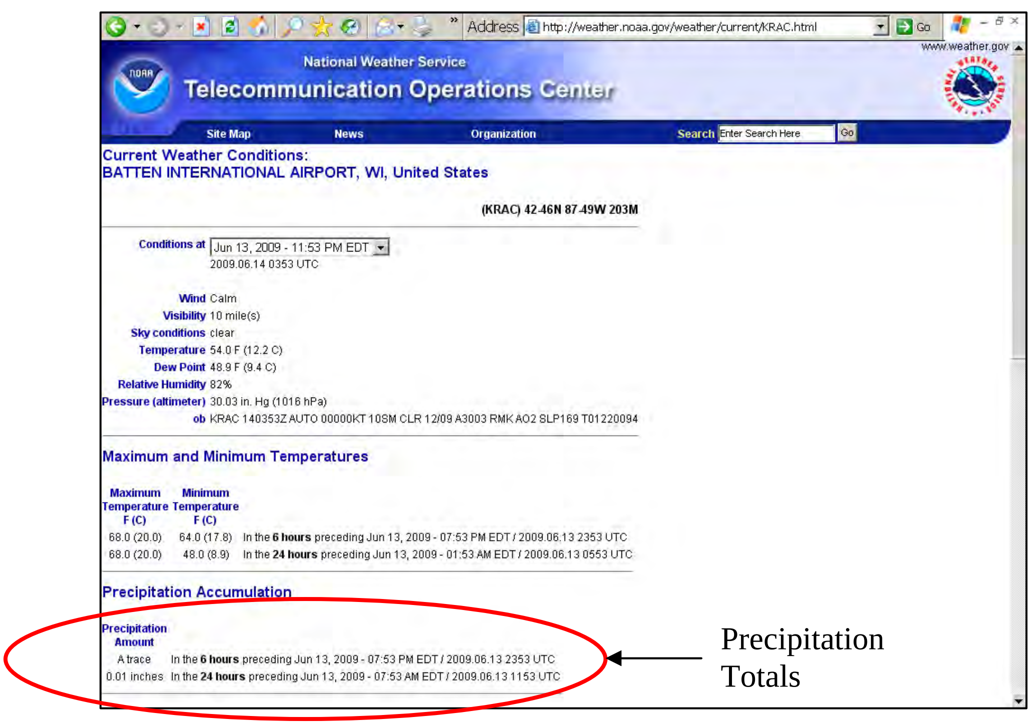
B.2.1. Open the main page (http://weather.noaa.gov) and scroll down the first pull-down menu ("Select a state…") to your state, select it, and click "Go."



Note: The nearest automated weather observing station may, in fact, be in a neighboring state. If you are unsure, use the **CoastWatch – Great Lakes Node** (B.1) for "Surface Airways Stations" (www.glerl.noaa.gov/marobs/php/data.php?sta=6).

B.2.2. On the page that opens for your state, under “Current Weather Conditions,” scroll down the first pull-down menu (“Select a location...”) to the weather station of interest, then click “Go.”

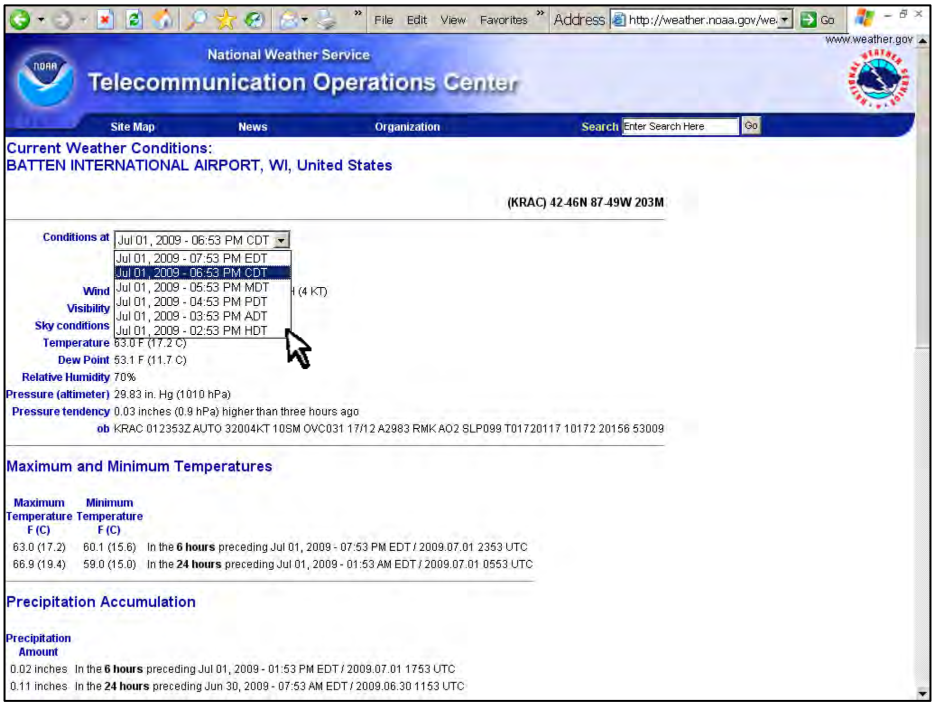
B.2.3. A “Current Weather Conditions” page will open for the selected station, listing the most recent (within the hour) wind speed and direction, temperature, sky conditions, and 6- and 24-hour rainfall totals.



Precipitation Totals

Note: Reporting times listed for 6- and 24-hour precipitation are ***not*** real-time. A single value for 24-hour precipitation is reported each day – at or around 12:00 UTM (7:00 am CDT, 8:00 am EDT). This marks the start of the "hydrologic day." Reporting times for 6-hour precipitation vary.

B.2.4. For Great Lakes beaches in the central time zone, change the time listed for the different variables from the default (EDT) to CDT using the "Conditions at" pull-down menu.



B.2.5. Add the "Current Weather Conditions" page for the selected location to your list of favorites in your Web browser. (This will enable you to return to the station page to receive updated conditions at any time.)

Notes

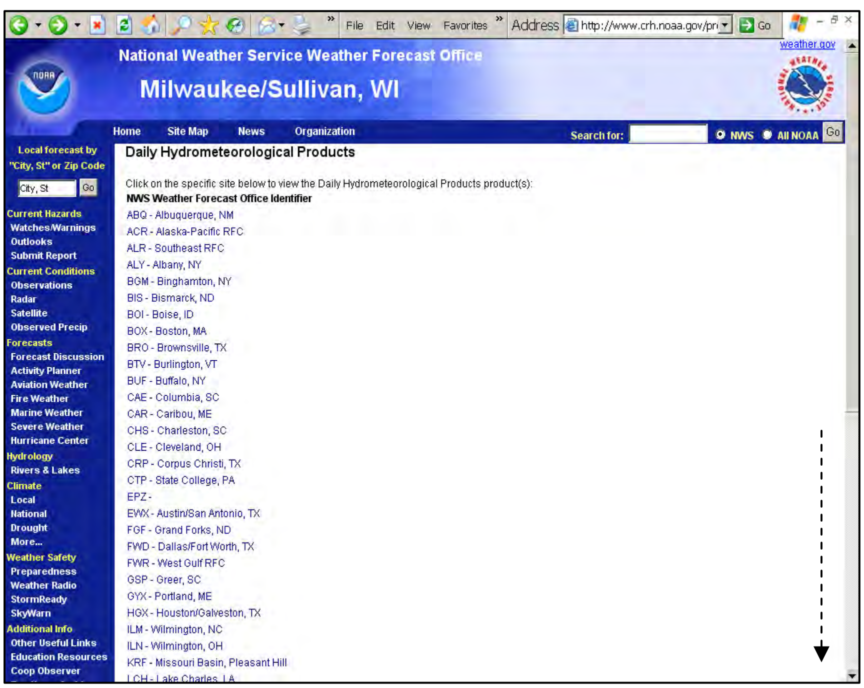
B.3 Daily Hydrometeorological Data

Provides links to daily recorded 24-hour precipitation totals measured at Cooperative Observer Program (COOP) sites in Wisconsin and other states. Maintained by the National Weather Services' Weather Forecast Office at Milwaukee/Sullivan, Wisconsin.

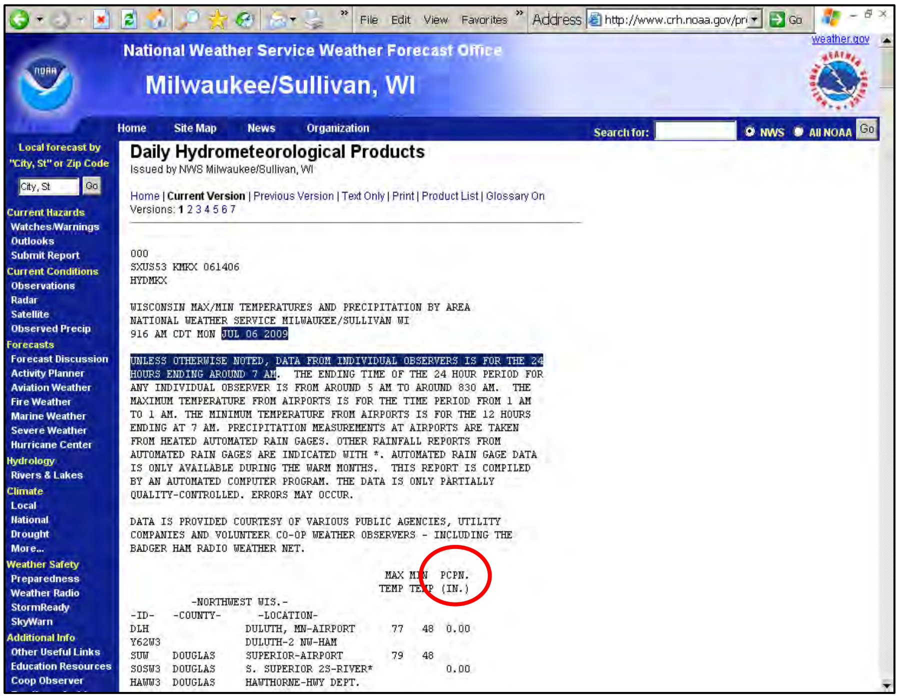
While their data are collected and reported by volunteers and are available only at 24-hour intervals, COOP stations are more common than AWOS and ASOS airport stations and are therefore more likely to occur in close proximity to your beach. Unlike the airport stations, though, COOP sites are not included in the **CoastWatch - Great Lakes Node** maps (B.1). To see if there is a COOP station nearer to your beach use the **Online Climate Directory - *Daily*** GIS link (B.4): http://gis.ncdc.noaa.gov/website/ims-cdo/sod/viewer.htm.

Note: The 24-hour rainfall totals are reported each day as of 12:00 UTM (7:00 am CDT). Depending on how long after this hour you collect your beach water sample, and how far away the nearest airport weather observing station is relative to the COOP site, antecedent rainfall based on COOP data may be more or less accurate than data available from the **Internet Weather Source** (B.2). As discussed in the "Using Online Sources vs. Field-Collected Data" section of this report, partial regression slope coefficients for similar variables measured at different locations, and their associated *p* values (measures of statistical significance), will determine which of the two rainfall totals is a better predictor of *E. coli* at your beach.

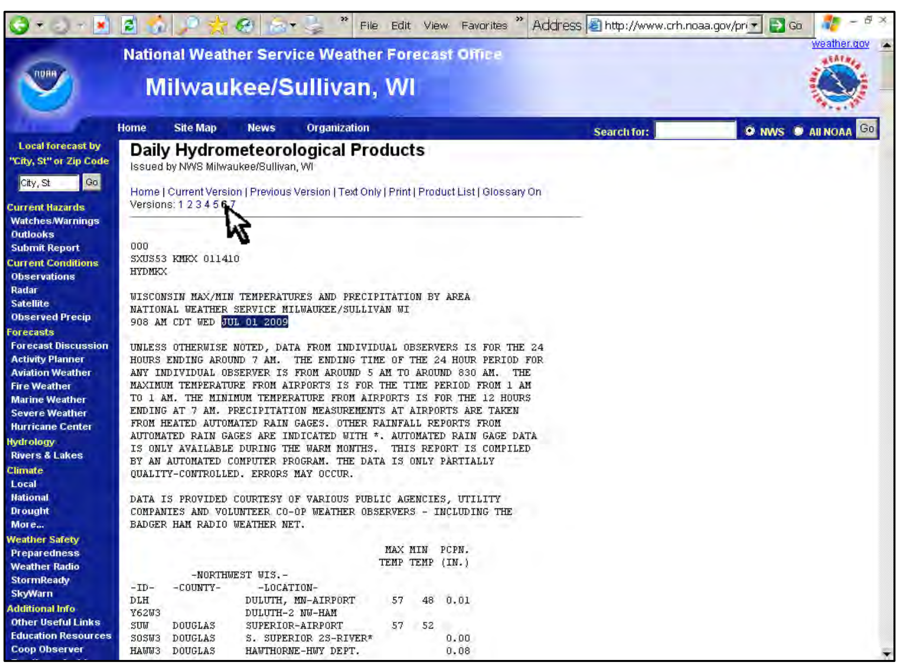
B.3.1. Open the "Daily Hydrometeorological Products" page at www.crh.noaa.gov/product_sites.php?site=MKX&product=HYD. This page lists a series of links to National Weather Service Forecast Offices, including today's COOP data for the different areas.



B.3.2. For Wisconsin stations, scroll down and click the link for “MKX - Milwaukee/Sullivan, WI.” This will open a new page (shown below), listing COOP sites by region (e.g., “NORHTEAST”). 24-hr totals (as of 7:00 am CDT) are listed under “PCPN”.



B.3.2. Data are available for the previous six days, as well. Clicking on the numbered links to the right of “Versions:” will list the data for previous days. Version “1” is today. Clicking “6” will return data five days prior to today.



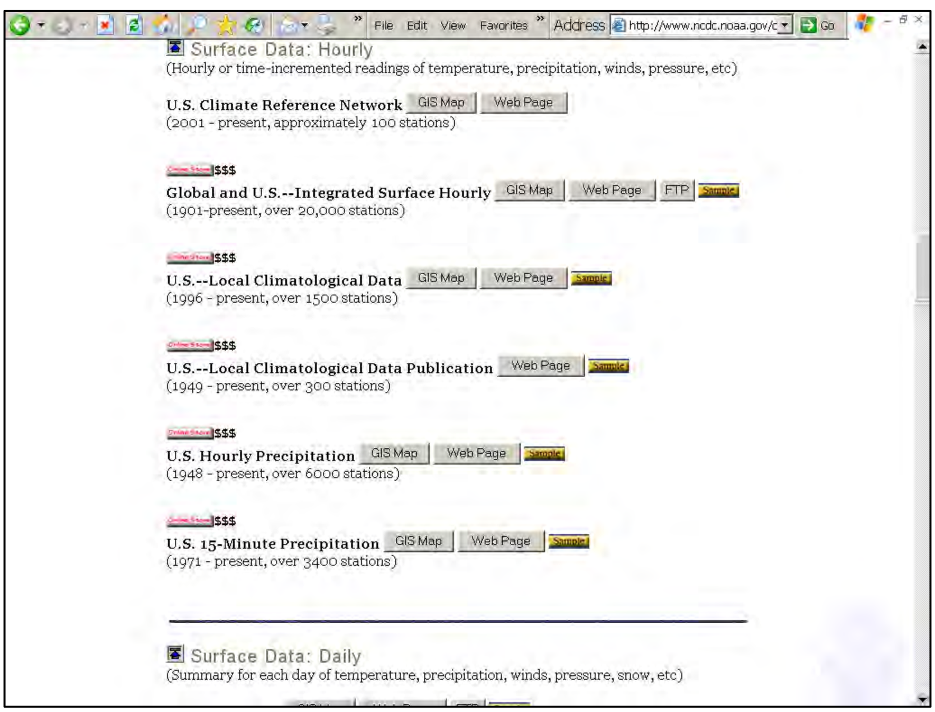
B.4. Online Climate Data Directory – *Hourly*

Provides access through a map interface to archived/historic *hourly* data from airport weather observing stations, among other sources. Part of the National Climate Data Center (NCDC). Pertinent variables include air temperature, wind speed and direction, antecedent rainfall, and sky conditions (e.g., clear, scattered, overcast, etc.).

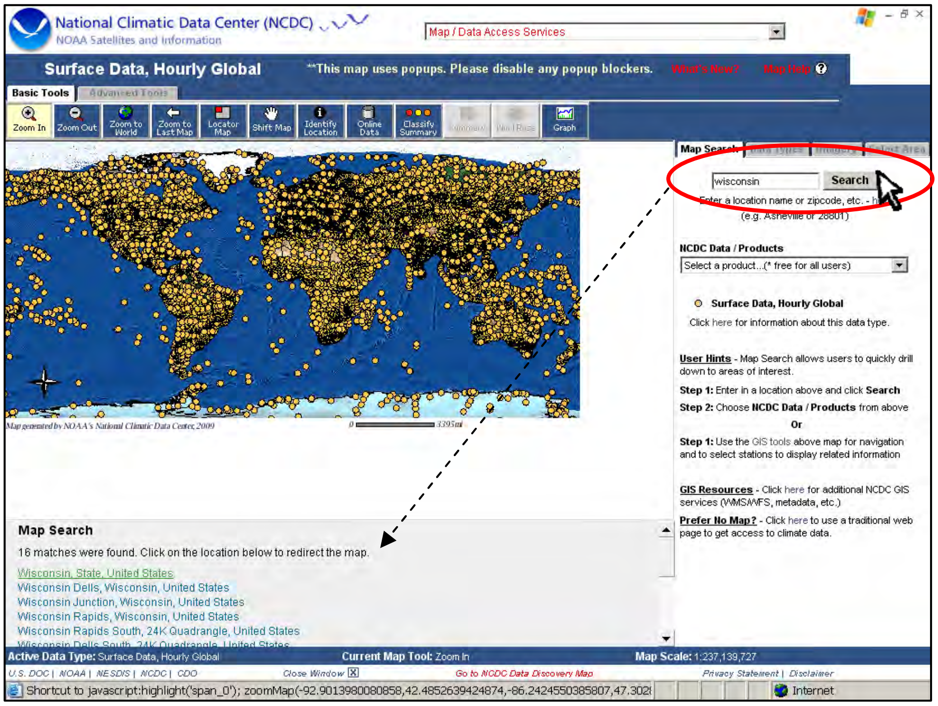
Note: The **Climate Data Directory – *Hourly*** is the only online access point for historic hourly data from AWOS and ASOS airport weather observing stations. Depending on your IP address, access may entail a fee. (If you have a web address ending in ".gov" or ".edu", the download will be free; otherwise, there will be a fee, which is payable online). Historic data collected at buoys and C-MAN stations are *available for free* from the **National Data Buoy Center** (B.6).

Note: NCDC data are reported at different intervals: monthly, daily, hourly, and sub-hourly. Generally-speaking, hourly data are the most relevant for building beach water quality models, since *E. coli* concentrations and other beach conditions vary over relatively short intervals. That said, daily-reported data on 24-hour precipitation may be more useful than hourly-reported data, in situations where the station in question is closer to your beach than the nearest station reporting in hourly increments. This may be the case with COOP stations – which are not included in the hourly version of the directory. To see whether any such stations exist near your beach, see the **Online Climate Directory - *Daily*** (B.5).

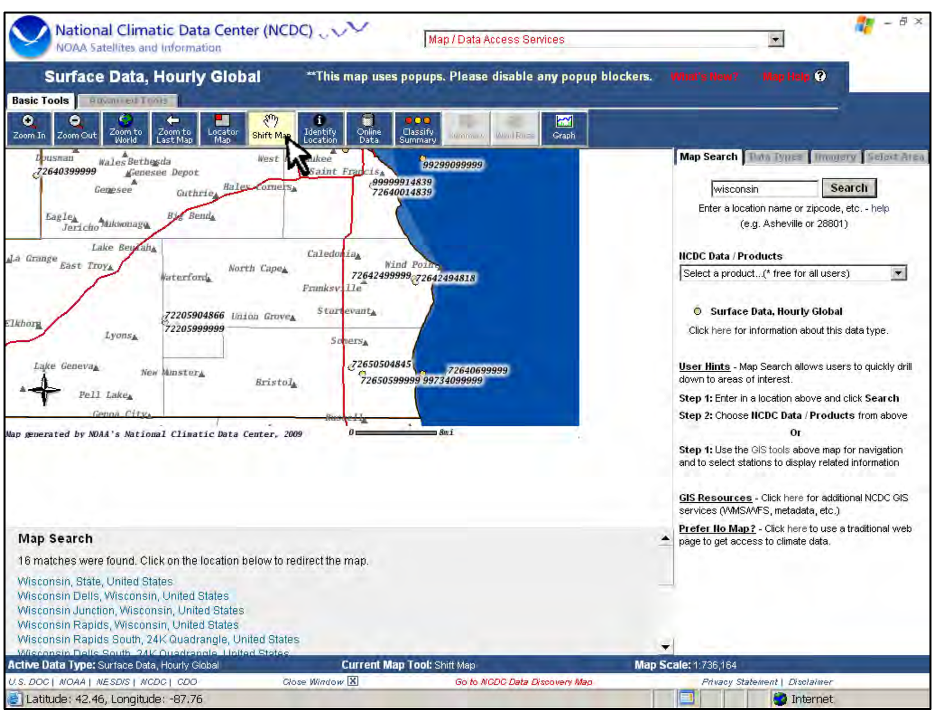
B.4.1. Open the "Surface Data: Hourly" section of the NCDC data download page at www.ncdc.noaa.gov/oa/climate/climatedata.html#hourly. Next to the heading "Global and U.S.--Integrated Surface Hourly", click the button for "GIS Map".



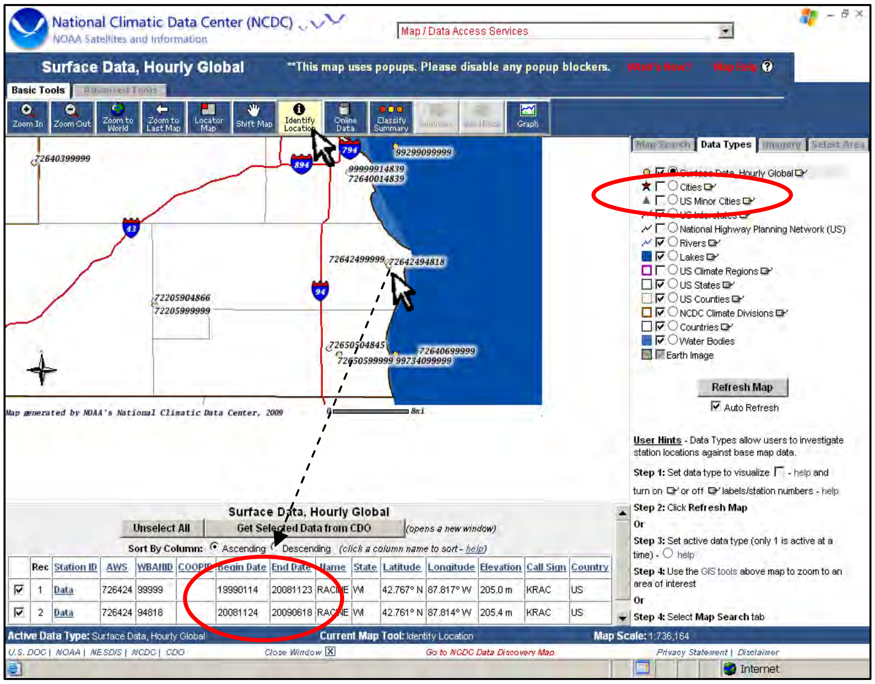
B.4.2. In the map page that opens, zoom to the vicinity of your beach by one of two methods: (1) Type the name of a location and click the “Search” button. Beneath the heading “Map Search”, click the result that matches your desired location (the map will automatically zoom to that location). (2) Click the “Zoom In” button on the upper left corner and draw a box around your area of interest.



B.4.3. To move around, click “Shift Map”, click any point, hold, and drag.



B.4.4. To make the map easier to read, click the “Data Type” tab on the right side of the page, un-check the boxes next to “Cities” and “US Minor Cities”, click the “Refresh Map” button, and check the box next to “Auto Refresh”.



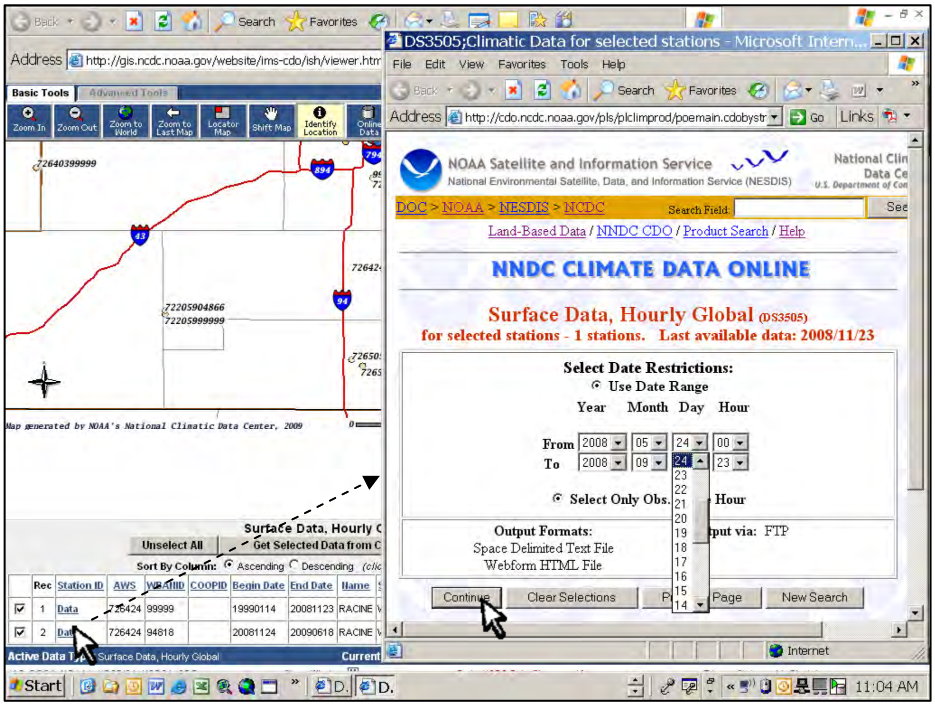
B.4.5. To begin the process of accessing station data, click the “Identify Location” button (see above) and then click on your station of interest. This will open a table, just below the map, summarizing basic information on the station or stations that you select. Summary information includes various station identifiers, geographic coordinates, a “Begin Date” (i.e. the start date of data collection at each station), and an “End Date”, as well as links to “Data”.

Note: As illustrated above, two stations may appear to occur in the same location. Typically in such cases, one of the stations replaced the other at some point in time. Check the “Begin Date” and “End Date” to determine which station/source you want to access. Depending on the dates, you may want both.

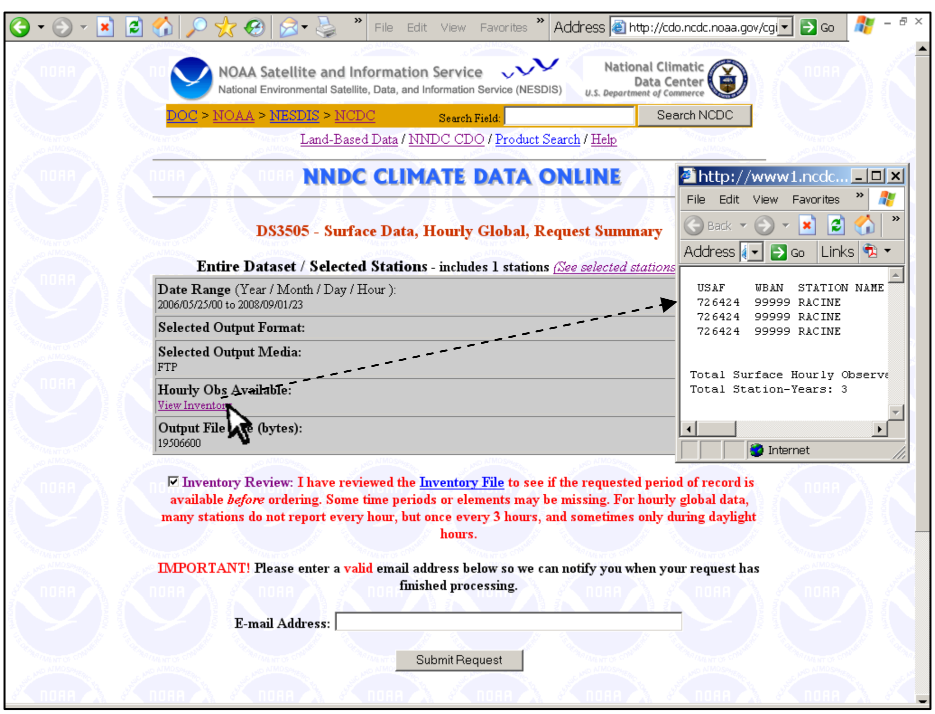
Note: If you are building your model *mid-season*, the “End Date” for currently-operated stations is typically *within the past week*, so there should be no significant gap between the most recent archived data and the real-time data, and therefore no need to use preliminary or unverified data. If you want seamless data (no gap) use the **Internet Weather Source** (B.2) to gather data for a few days prior to accessing archived hourly data.

Note: Stations that appear to be located immediately at (or just beyond) the shoreline may be C-MAN coastal stations or Great Lakes buoys. Check the map interface of the **National Data Buoy Center** (B.6) to see if they are. If so, you can access their data, for free, through that site.

B.4.6. Click the “Data” link in the station summary table to open the “Data” page. Using the pull-down menus, select a start date (“From”) and an end date (“To”).

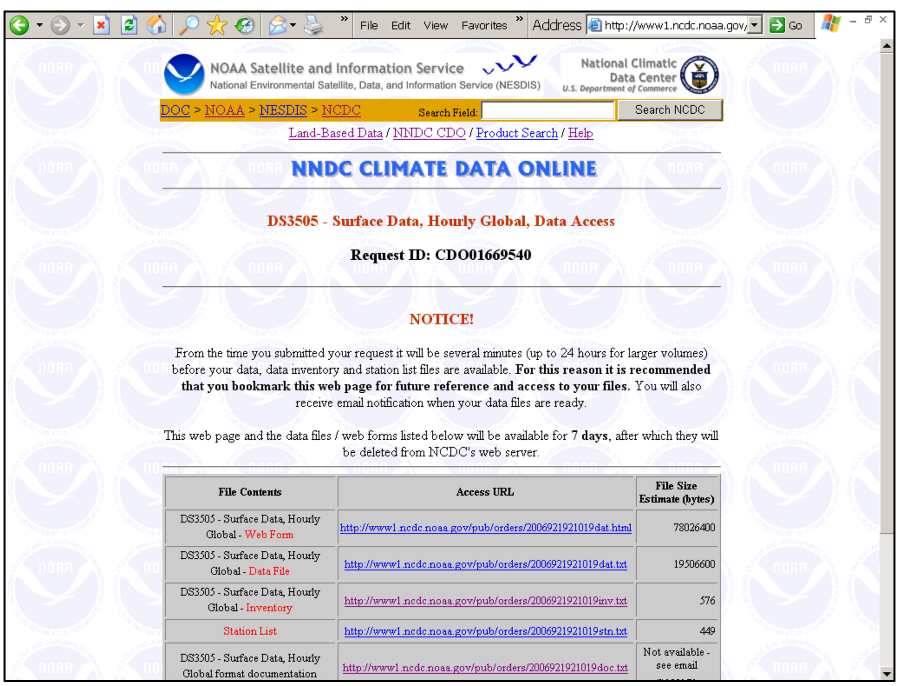


B.4.7. Check the circle next to “Select Only Obs. on the Hour” (above) and click “Continue”. In the page that opens (below) click the “View Inventory” link to review a list of the available data before proceeding.

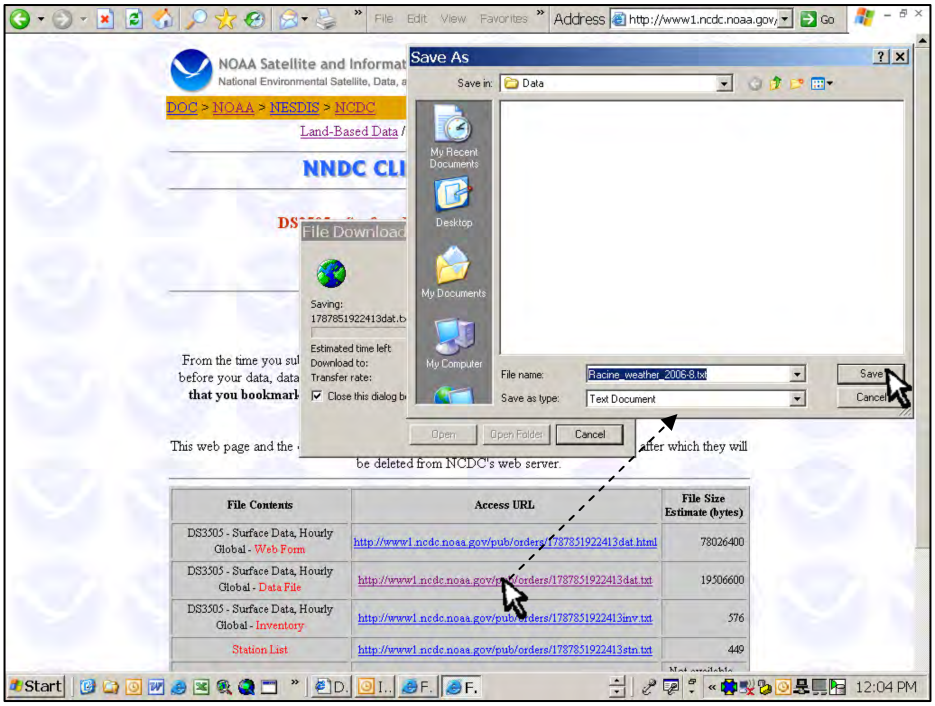


B.4.8. After viewing the inventory, check the box next to “Inventory Review”, type your email address in the space provided, and click the “Submit Request” button. This will open the verification window shown below. Leave this window open – minimized is okay – until you receive a verification email.

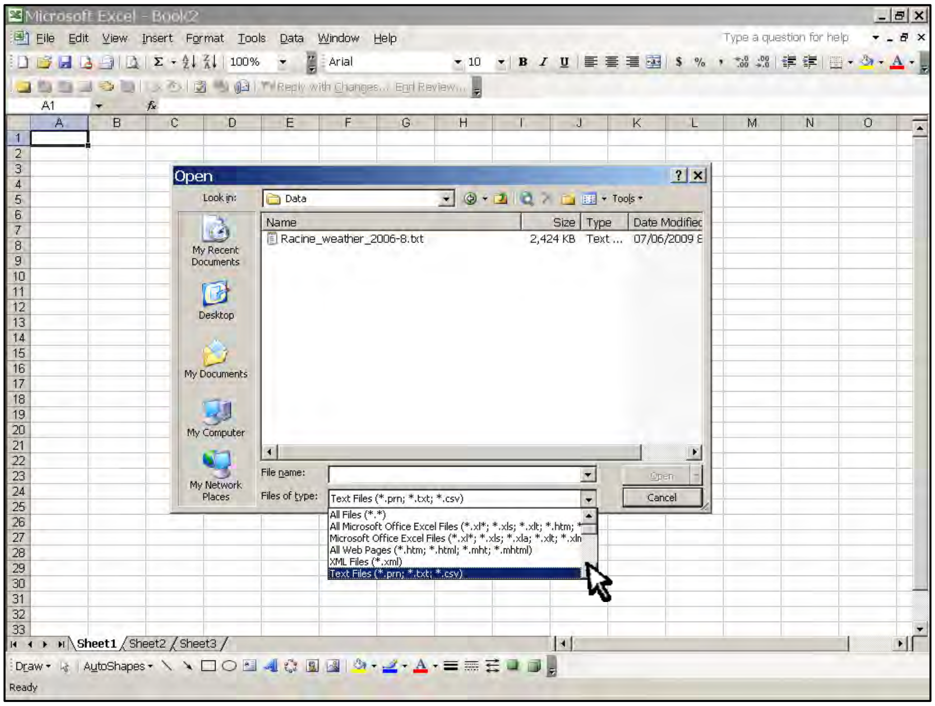
B.4.9. When you receive your verification email, click the URL link in the center of the Verification window (above). This will open a Data Access page (below).



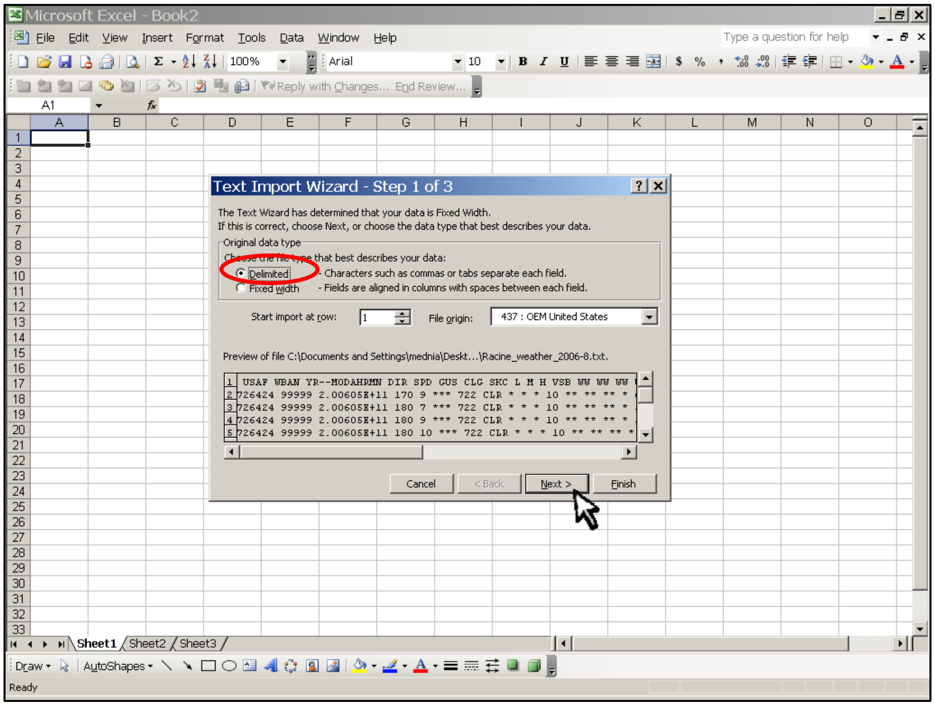
B.4.10. On the "Data Access" page, right-click on the second URL link ("... Surface Data, Hourly Global - Data File"), select "Save Target As..." from the right-click menu, name the file, set the file type to Text Document (*.txt), and click "Save".



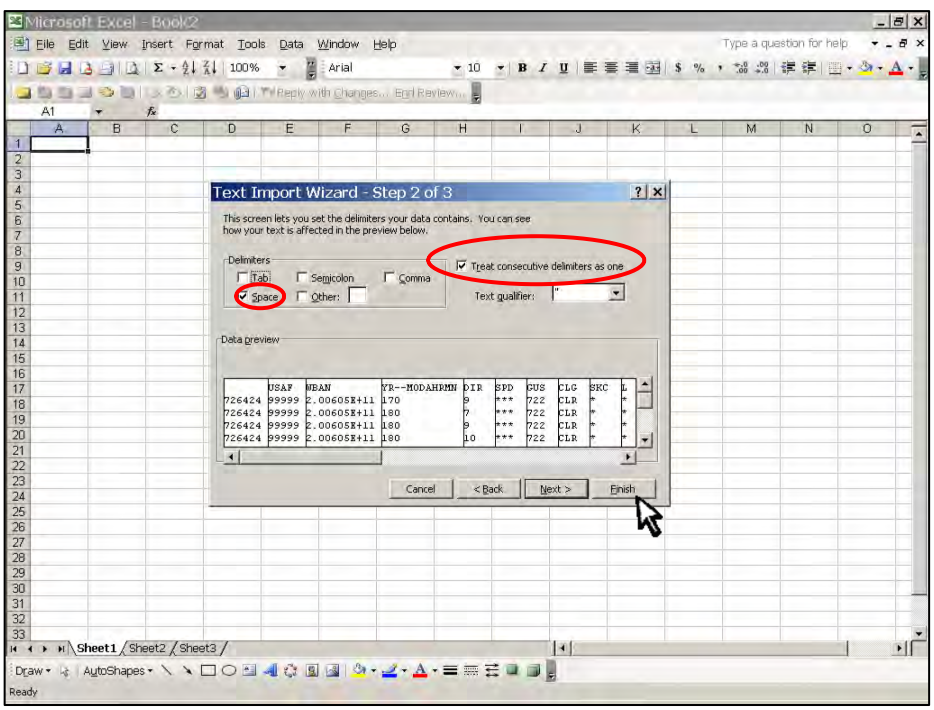
B.4.11. Open MS Excel. From the MS Excel pull-down menu (in the upper left corner), select "File" > "Open". In the "Open" window, select "Text Files" from the "Files of type:" pull-down menu. Double-click on the text file you created in step B.4.10.



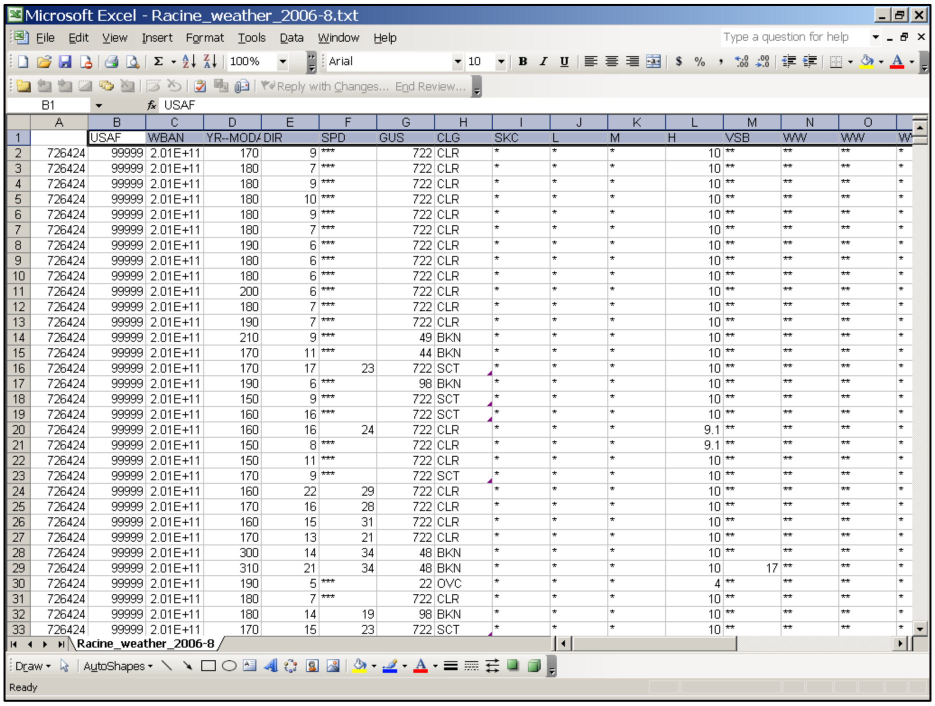
B.4.12. A “Text Import Wizard” will open. In Step 1, check the circle for “Delimited” and click the “Next>” button.



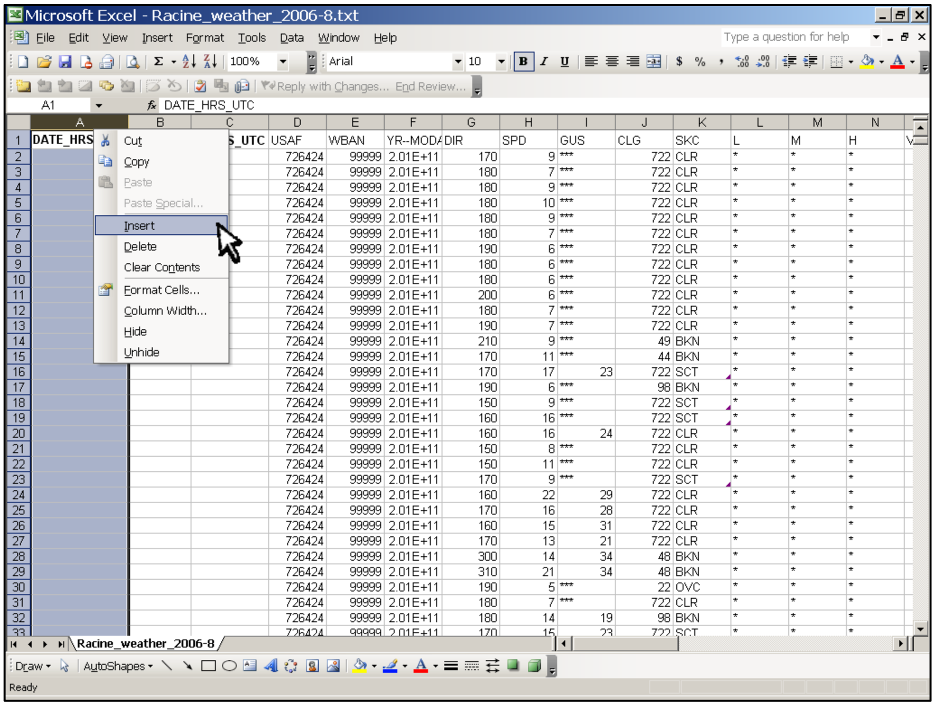
B.4.13. In Step 2 of the “Text Import Wizard”, check the boxes for “Space” and “Treat consecutive delimiters as one”. Click the “Finish” button.



B.4.14. Column names will need to be shifted one column to the left. To do this, click on first named cell (B1), hold the "Shift" key, and press the right arrow key until all of the named cells are highlighted. From the MS Excel pull-down menu, select "Edit" > "Cut". Then click on the empty cell (A1), and paste ("Edit" > "Paste").



B.4.15. Following the procedure in A.1.7, insert three new columns (A-C) named, respectively, "DATE_HRS_UTC", "DATE", and "HOURS_UTC".



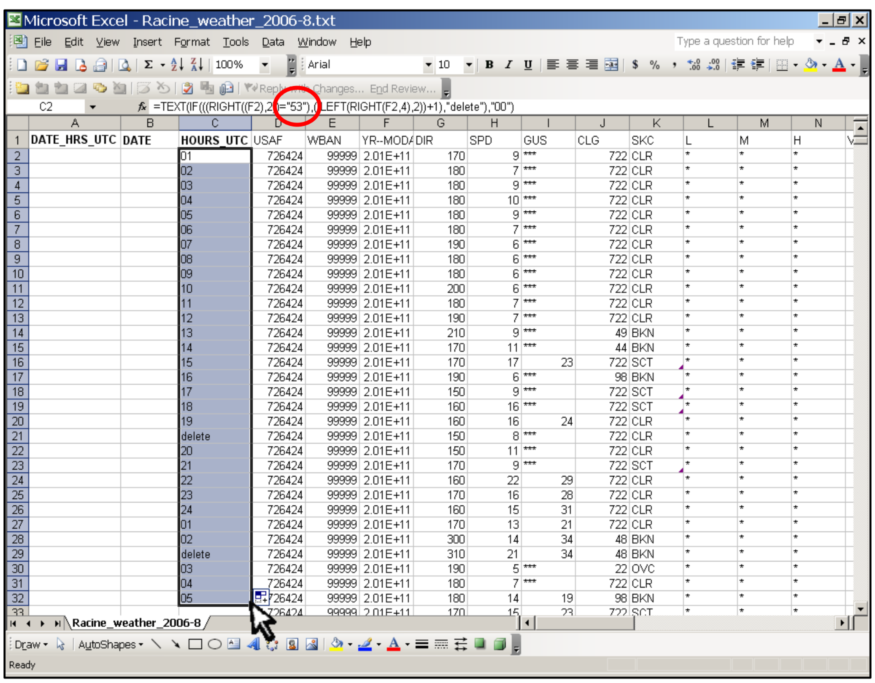
⚠ **Note:** As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to “join” the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and “joining” tables using queries in MS Access, see Section 6 of this report, “Formatting Data for Model-Building.”)

B.4.16. Hourly observations reported in NCDC data tables do not always fall “on-the-hour.” In the example shown here, the station reports regular observations at 53 minutes after each hour. In some cases additional (redundant) observations are reported in between the regular observations. To flag these, while creating values for “HOURS_UTC” (column C) type the following function in cell C2:

=TEXT(IF(((RIGHT((F2),2))="XX"),((LEFT(RIGHT(F2,4),2))+1),"delete"),"00")

(Where “XX” is the regular minute of the hour (00-59) at which data are reported, as indicated by the last two digits in the date and time values reported in column D, “YR--MODAHRMN”.)

Hit “Enter”. The result should be a two-digit, rounded value for the hour-of-day in UTC (Central Daylight Time + 5 hours). Copy this function by clicking in the lower right corner of cell C2 and dragging to the end of the table.



Note: In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Be sure to include period-separators between the year, month, day, and hour (e.g., 2003.05.29.12 as opposed to 2003052912). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

B.4.17. To calculate "DATE" (column B), click on cell B2 and type the following function:

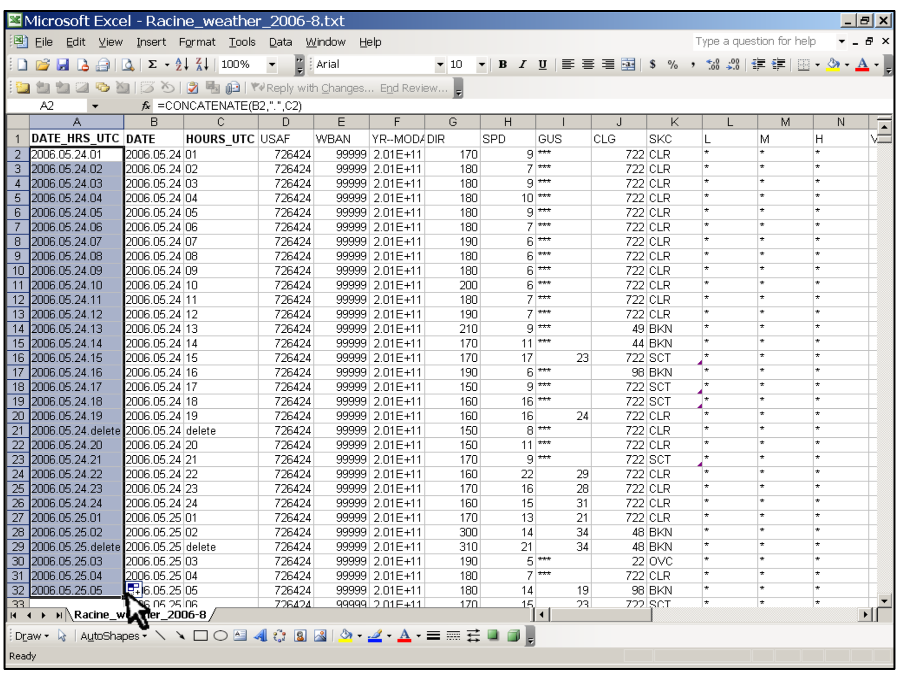
=CONCATENATE((LEFT((F2),4)),".",(RIGHT((LEFT(F2,6)),2)),".",(RIGHT((LEFT(F2,8)),2)))

Hit "Enter". This should return a date value (e.g., "2008.05.24" for May 24, 2008). Copy this function by clicking in the lower right corner of cell B2 and dragging to the end of the table.

B.4.18. To calculate "DATE_HRS_UTC" (column A), click on cell A2 and type the following function:

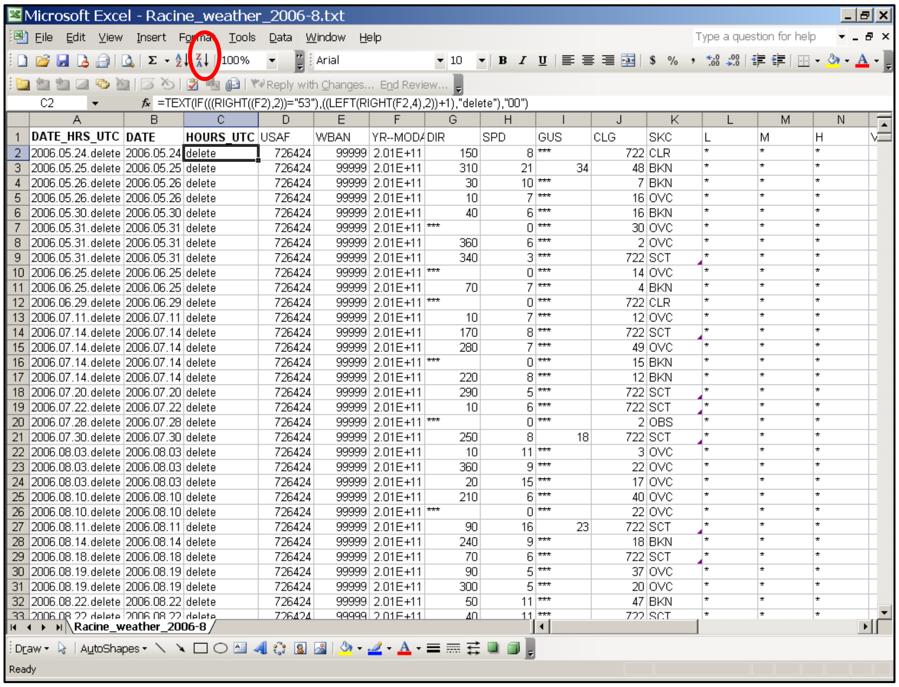
=CONCATENATE(B2,".",C2)

Hit "Enter". This should return a date-plus-hour value (e.g., "2008.05.24.01" for 01:00 hours UTC on May 24, 2008). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.

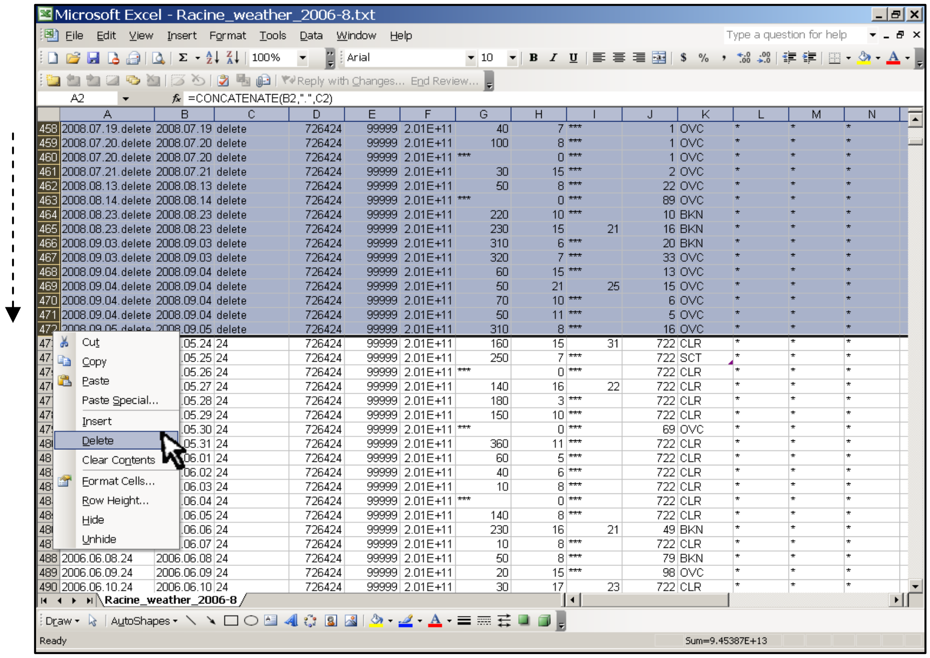


For additional column and row editing and formatting procedures, and for different functions, see the MS Excel Help menu.

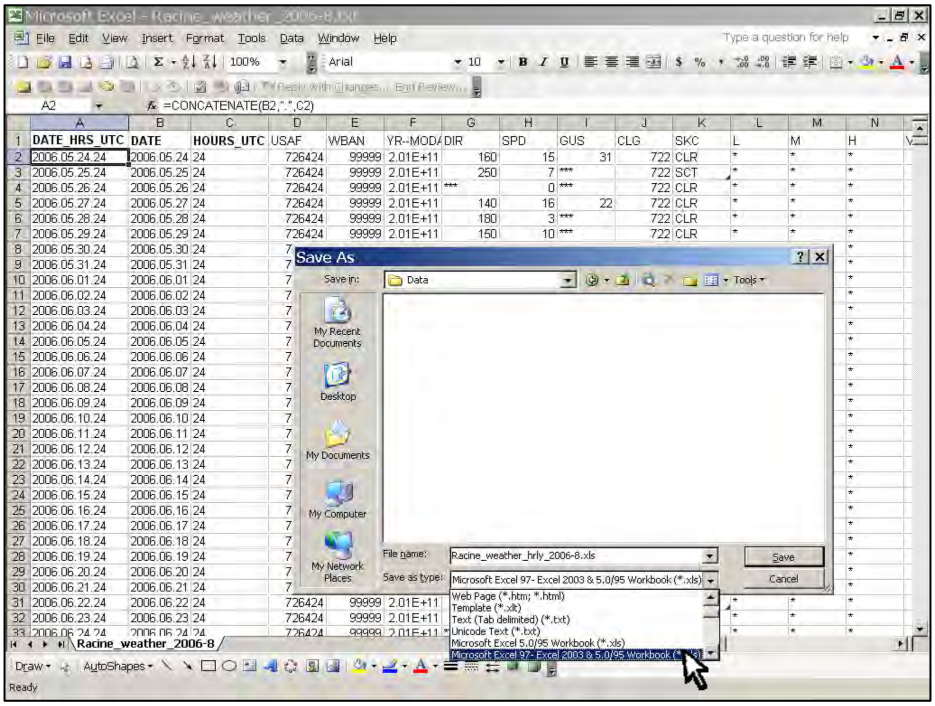
B.4.19. To remove the redundant hourly observations that you flagged in step B.4.16, first click on the first cell in column C and click on reverse-sort button ( ). The redundant observations will be at the top of the table.



B.4.20. Next, click to the left of row 2 and, holding the “Shift” key, push the down-arrow on the keyboard until all of the redundant observations are selected (i.e. column C values = “delete”). Right-click on the last-selected row and select “Delete”.



B.4.21. Save the table as an MS Excel file by selecting “File” > “Save as...” from the MS Excel pull-down menu, choosing the most recent version of MS Excel as the file type (*.xls), and naming the file something like “[location]_ weather_hrly_[year(s)].xls”.



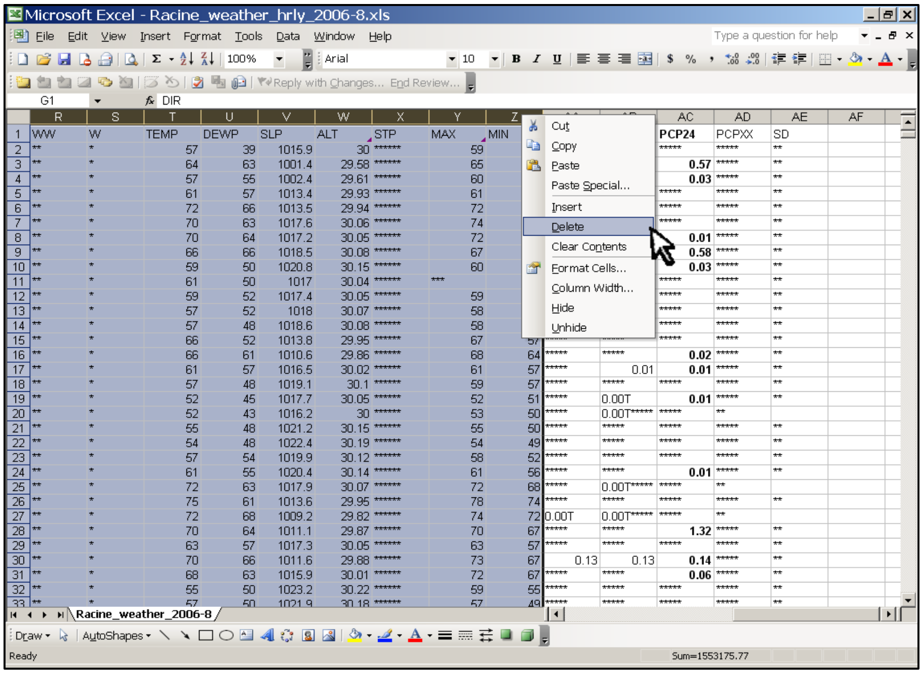
B.4.22. For descriptions of the different variables listed in the NCDC data tables, go to the “Data Access” page (B.4.9) and click on the last URL link: “Surface Data, Hourly Global format documentation”, or go directly to http://cdo.ncdc.noaa.gov/cdo/3505doc.txt.

Note: For wind direction (“DIR”, column G), values of “***”represent the *absence* of wind (i.e. wind speed = 0). They do *not* represent 0°, or due-northerly winds.

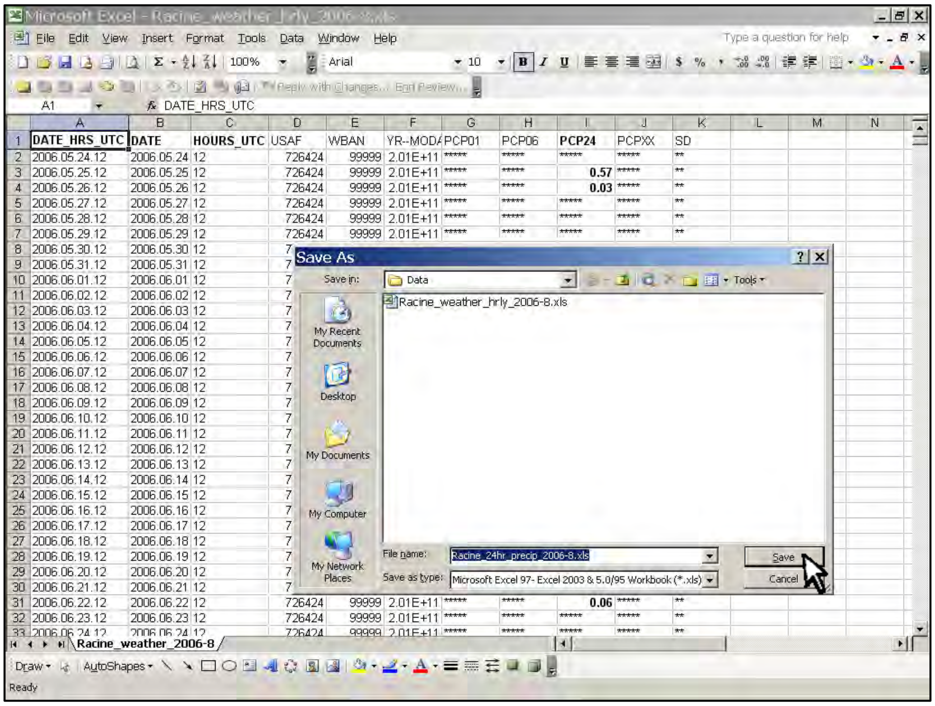
Note: 24-hour precipitation (“PCP24”, column AC) is reported only once per day, at or around 12:00 UTC, or 7:00 am CDT, which is the standard beginning of the “hydrologic day” for modeling purposes. Values of “*****” represent either no data reported (i.e. for hours of the day other than 12:00 UTC) or 0 precipitation (i.e. no rainfall recorded over the 24 hours prior to 12:00 UTC). MS Excel treats cell values of * (or multiple **’s) as 0.

B.4.23. Following the procedures described in step B.4.17, select and delete all rows where 24-hour precipitation was not reported (i.e. all cells were "HOURS_UTC" is less than 12, as well as all rows where it is *greater* than 12. (After doing so, confirm that all cells under "HOURS_UTC" have a value of 12, and that at least some of the cells under "PCP24" [column AC] are populated with numeric values, as opposed to all cells having a value of "*****".)

B.4.24. Select columns G ("DIR") through Z ("MIN") by clicking on the column G header, and while holding the "Shift" key, pushing the right-arrow key until column Z is highlighted as well. Right-click and select "Delete".



B.4.25. Re-save the edited table by selecting “File” > “Save as…” from the MS Excel pull-down menu and naming the file something like “[location]_24hr_precip_[year(s)].xls”. Click “Save” and close MS Excel.



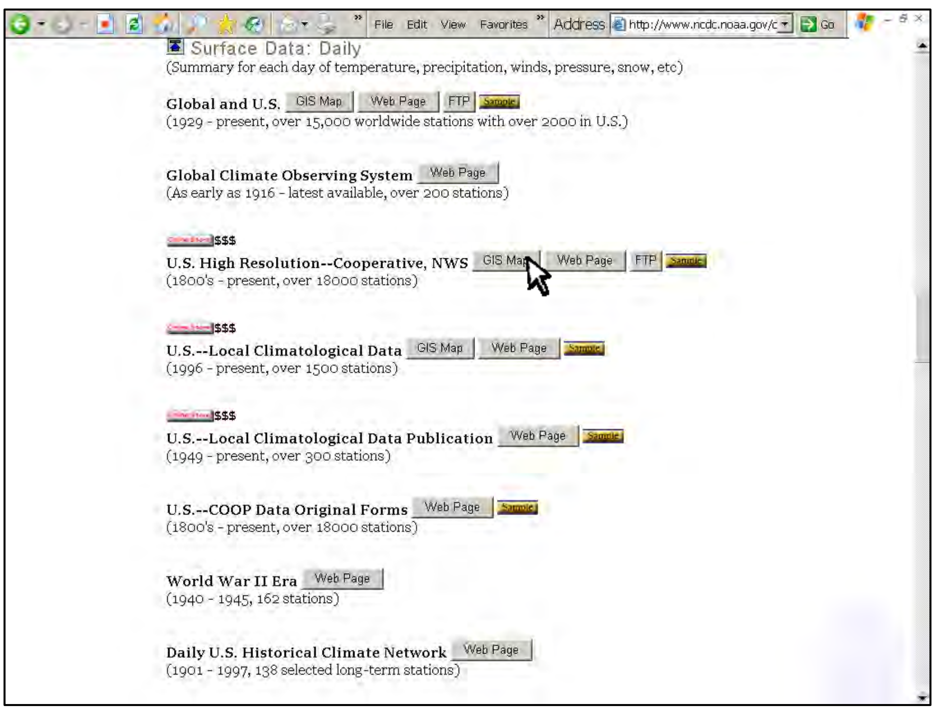
B.5. Online Climate Data Directory – *Daily*

Provides access through a map interface to *daily* archived/historic data from Cooperative Observer Program (COOP) stations and other sources. Part of the National Climate Data Center (NCDC). The principal variable for water quality modeling is 24-hour precipitation.

While hourly data are typically more relevant for building real-time water quality models, 24-hour rainfall reported each morning at COOP weather stations may prove to be a better predictor when the nearest airport station reporting hourly precipitation is far from your beach. (Use the GIS link [B.5.1 below] to see whether there are any stations reporting daily precipitation located closer to your beach than the nearest AWOS or ASOS station, as mapped under "Surface Airways Stations" in the **CoastWatch - Great Lakes Node** (B.1): www.glerl.noaa.gov/marobs/php/data.php?sta=6).

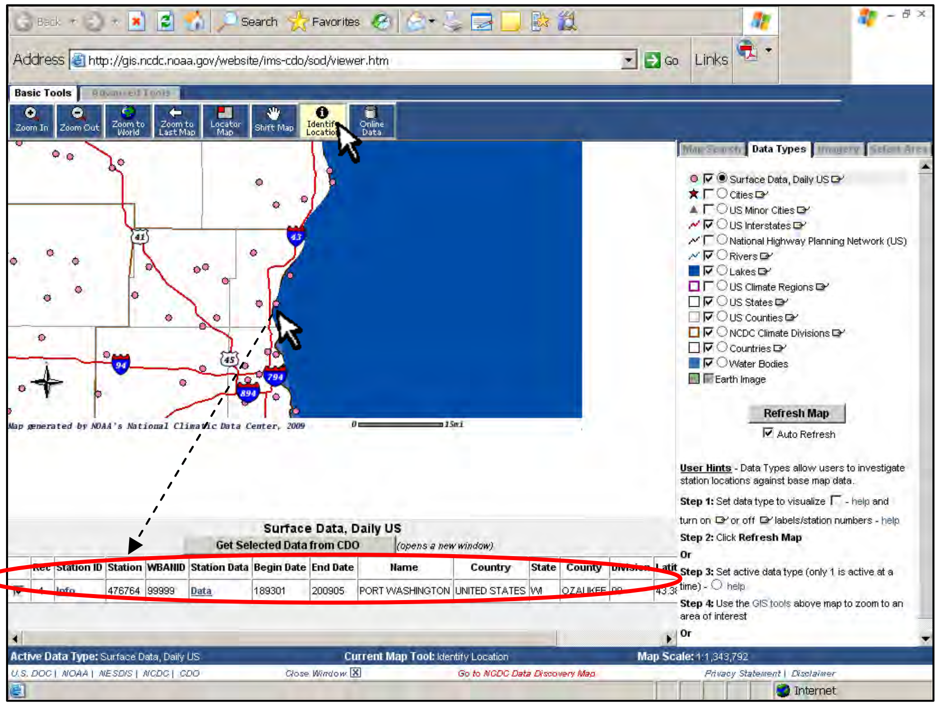
The **Climate Data Directory** is the only online access point for historic daily data from COOP stations. Depending on your IP address, access may entail a fee. (If you have a web address ending in ".gov" or ".edu", the download will be free; otherwise, there will be a fee, which is payable online).

B.5.1. Open the "Surface Data: Daily" section of the NCDC data download page at www.ncdc.noaa.gov/oa/climate/climatedata.html#daily. Next to the heading "U.S. High Resolution—Cooperative, NWS" click the button for "GIS Map".

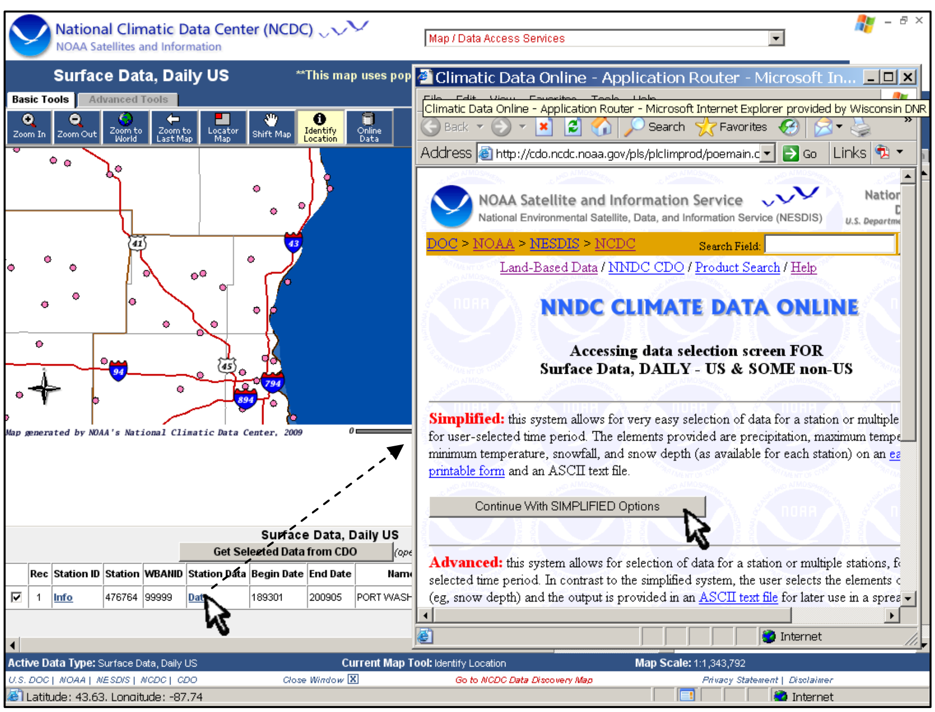


Note: The process of accessing *daily* data through the **Online Climate Data Directory – *Daily*** site is similar to the process for *hourly* data. Instructions below are shortened in places, referring back to identical procedures in B.4.

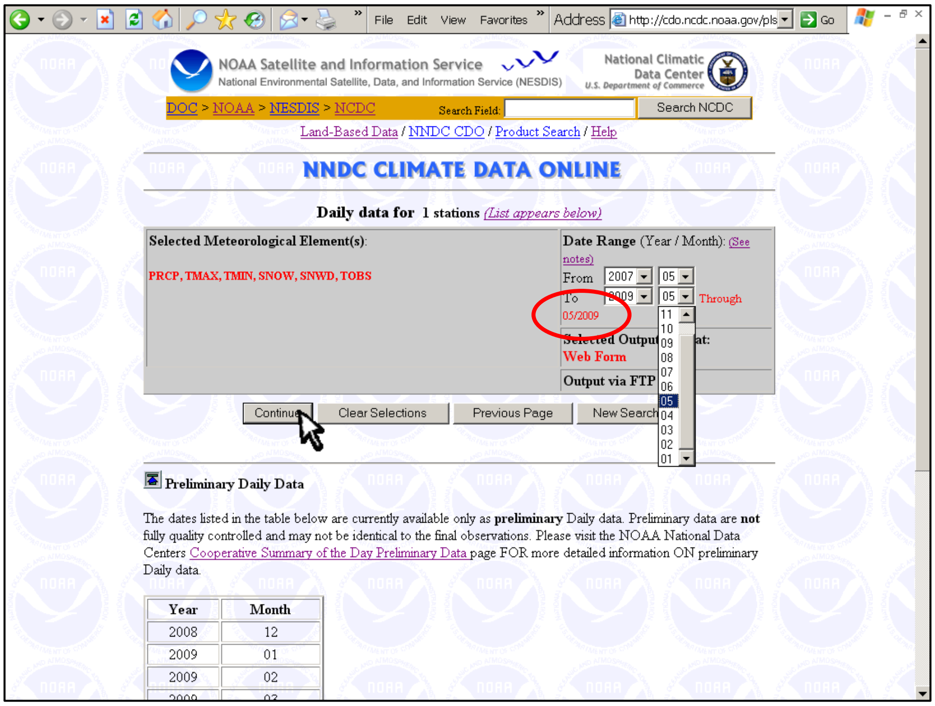
B.5.2. Following the procedures in steps B.4.2 – B.4.5, navigate to and identify your station of interest, click on the “Identify Location” button, and then click on your station of interest. This will open a station summary table just below the map.



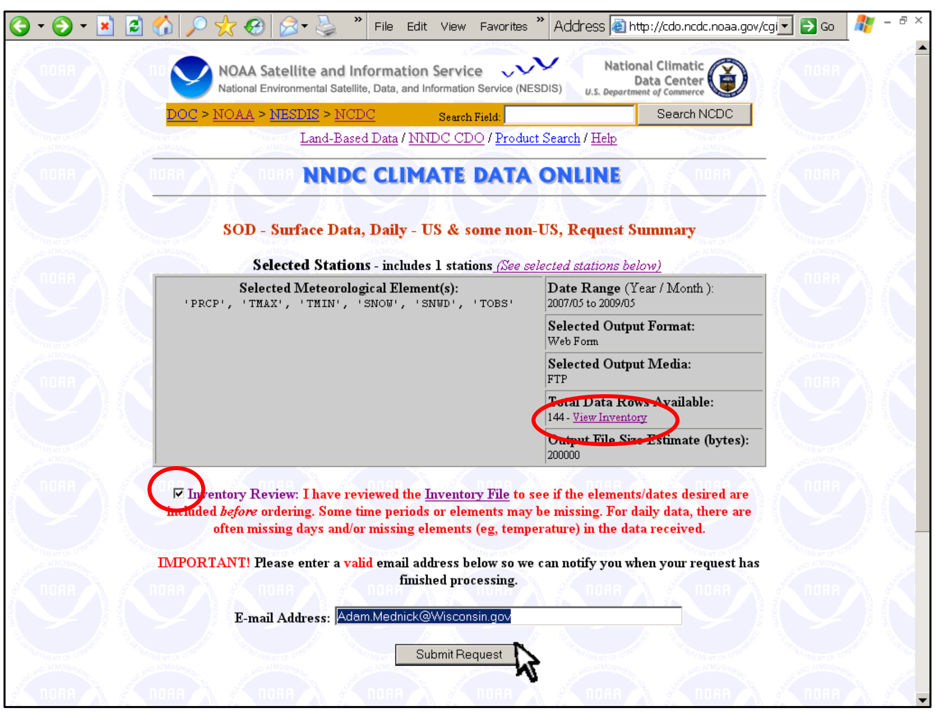
B.5.3. Click on the “Data” link in the station summary table to launch the “Data” page in a separate window. Click on the “Continue with SIMPLIFIED Options” button.



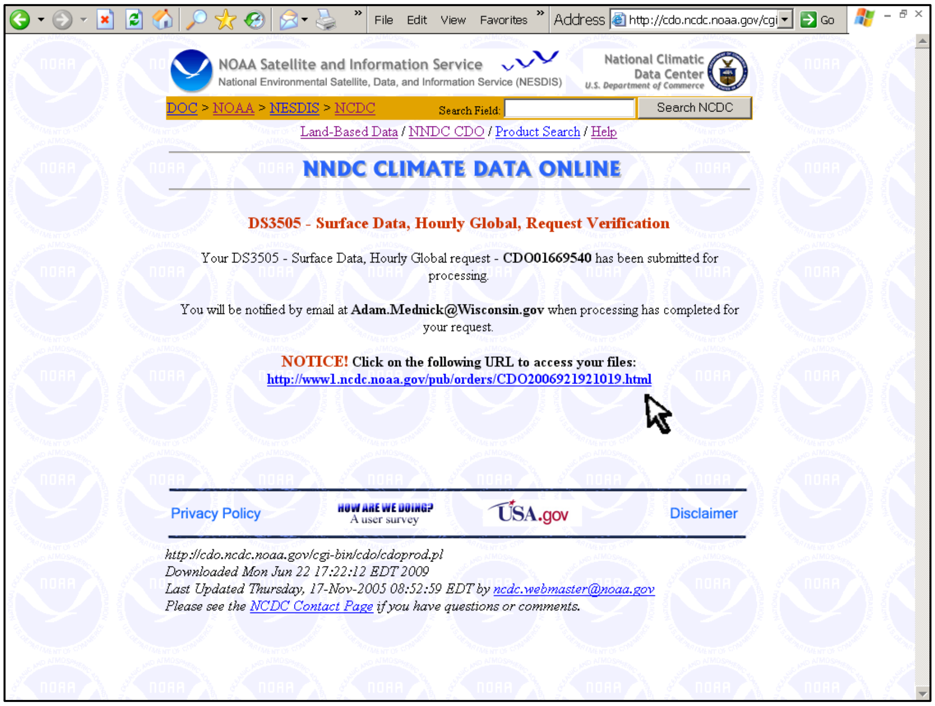
B.5.4. On the “Data” page, select a “Date Range” using the “To” and “From” menus. (Note_that the date in red, under “To,” indicates the most recent *verified* data. To access more recent (*unverified*) data, see B.5.16 below.) Click the “Continue” button.



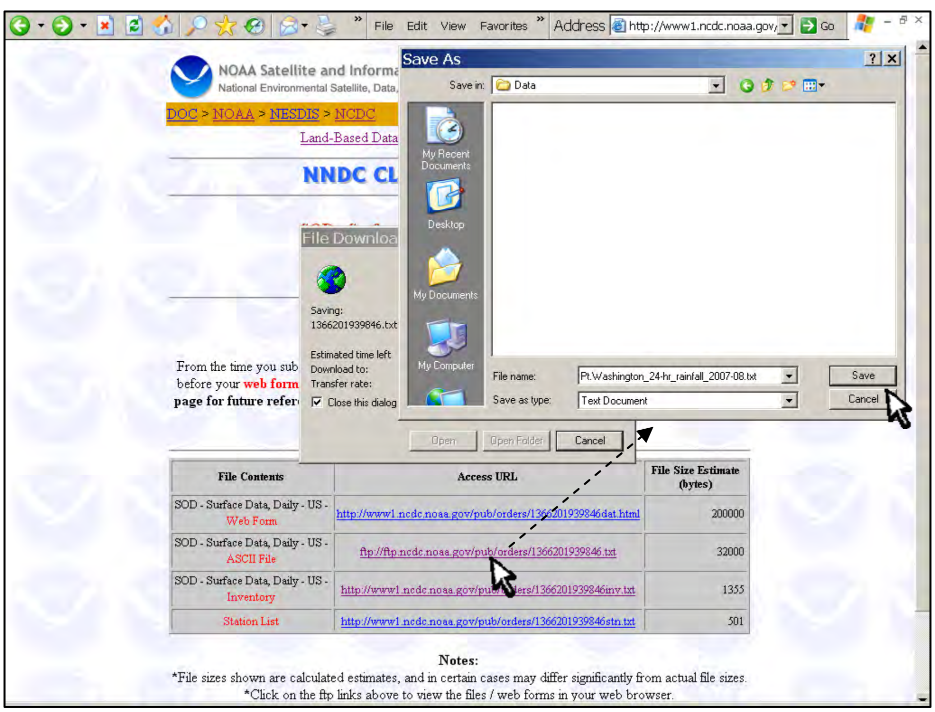
B.5.5. On the “Request” page that opens, click “View Inventory” to read a list of the available data. Then check the box next to “Inventory Review”, type your email address in the space provided, and click the “Submit Request” button.



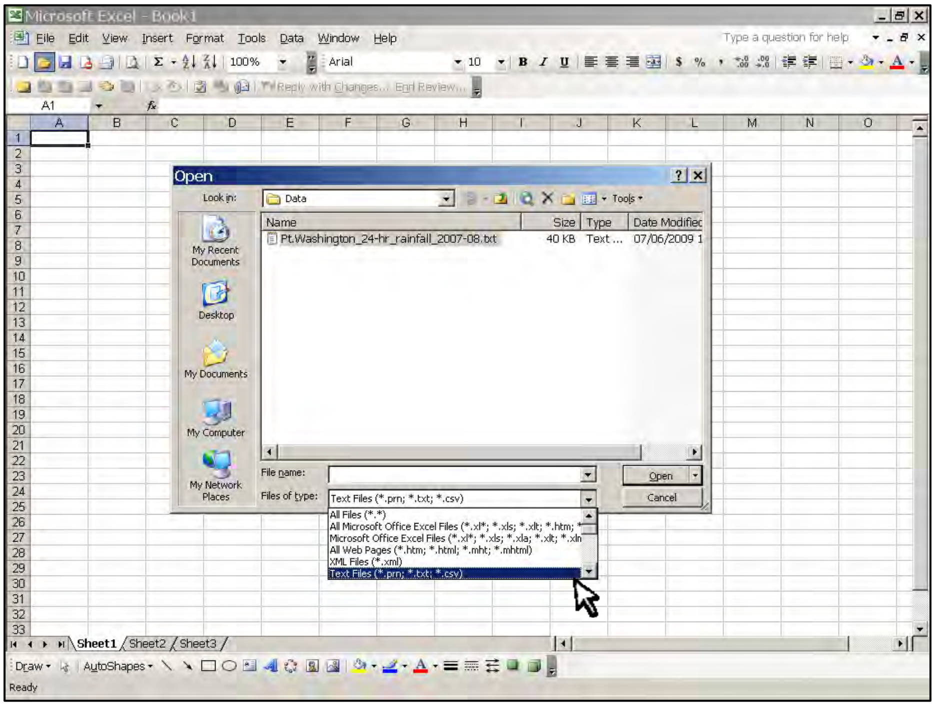
B.5.6. After you submit your request, a “Verification” window will open. Leave it open until you receive your verification email. Then click the URL link in the center of the window. This will open a “Data Access” page.



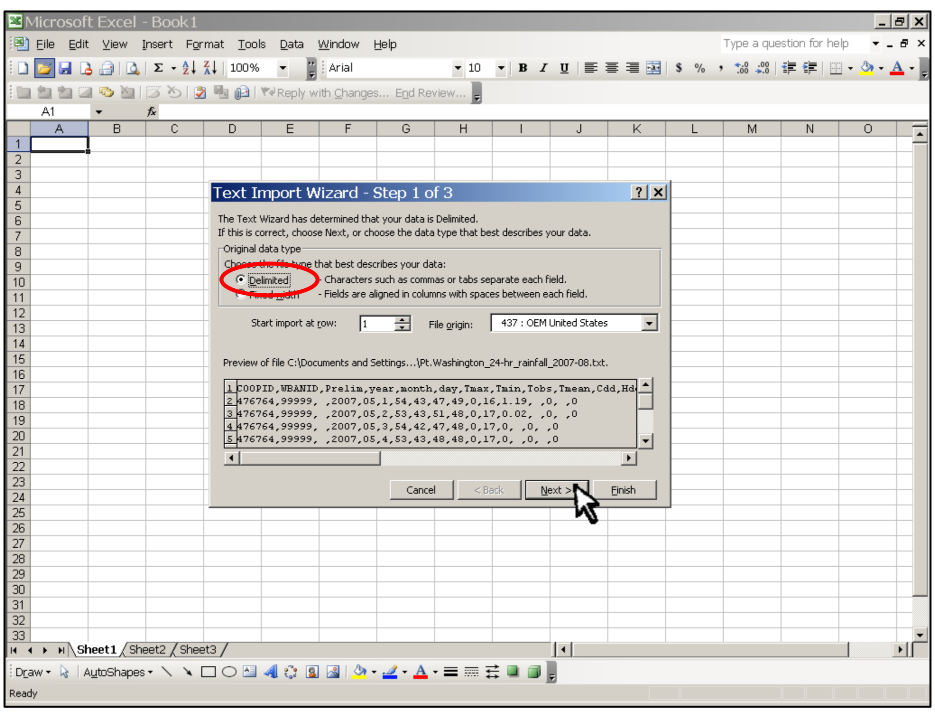
B.5.7. On the “Data Access” page, right-click on the second URL link (“… Surface Data, Daily… ASCII File”), select “Save Target As…” from the right-click menu, name the file, set the file type as “Text Document” (*.txt), and click “Save”.



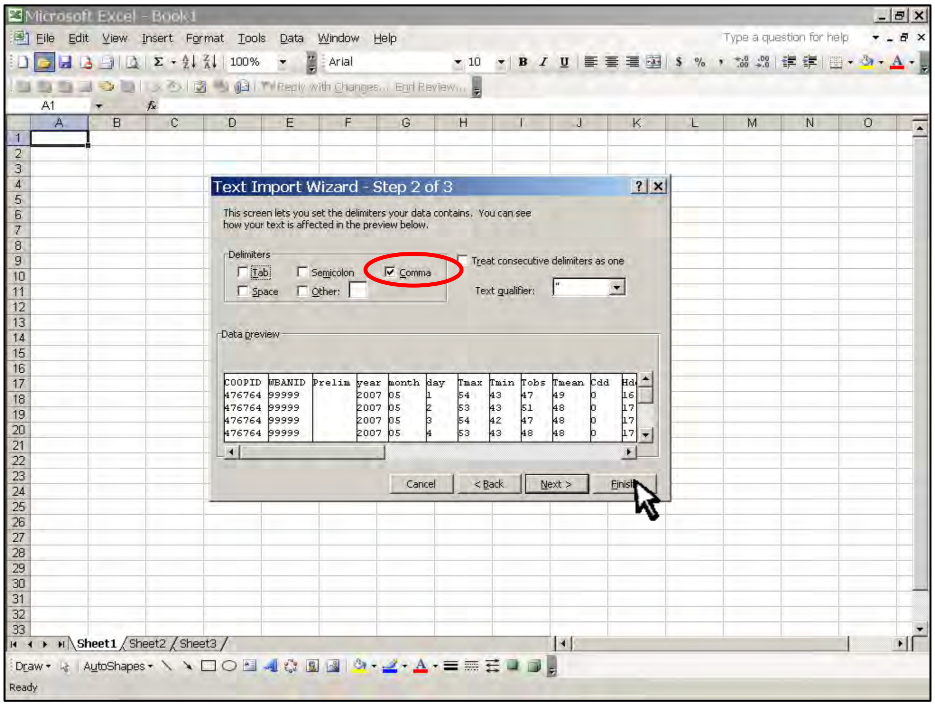
B.5.8. Open MS Excel. From the MS Excel pull-down menu (in the upper left corner), select “File” > “Open”. In the “Open” window, select “Text Files” from the “Files of type:” pull-down menu. Double-click on the text file you created in B.5.7.



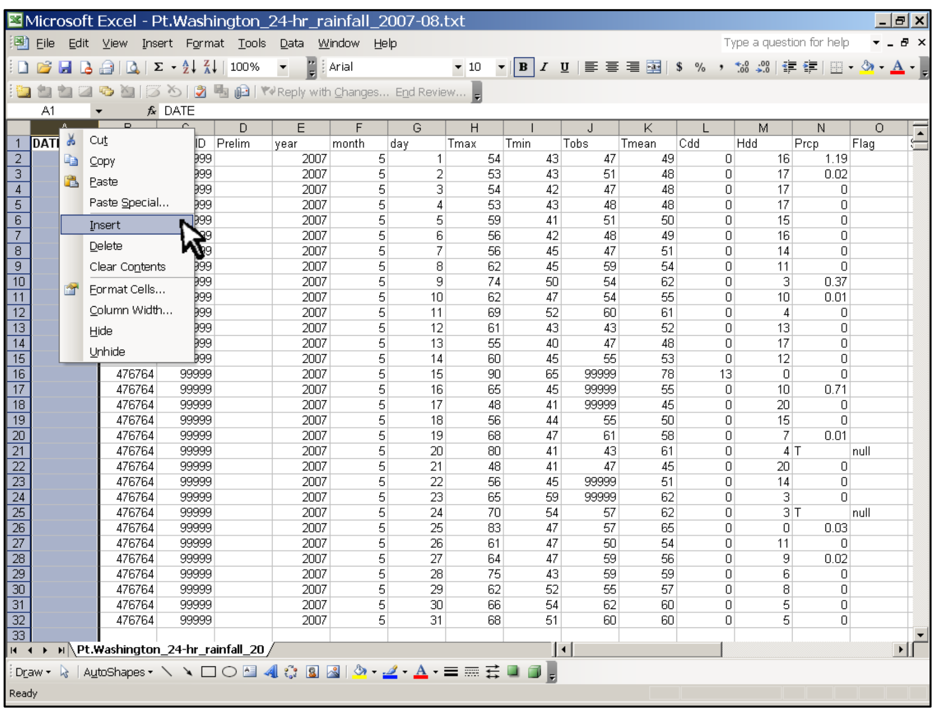
B.5.9. A “Text Import Wizard” will open. In Step 1, check the circle for “Delimited” and click the “Next>” button.



B.5.10. In Step 2 of the “Text Import Wizard”, check the box for “Comma” and click the “Finish” button.



B.5.11. Insert a new column called “DATE” by right-clicking on the first column header (A) and selecting “Insert” from the pull-down menu. Click on the first cell of the new column (cell A1) and type “DATE”.



Note: As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to “join” the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows.

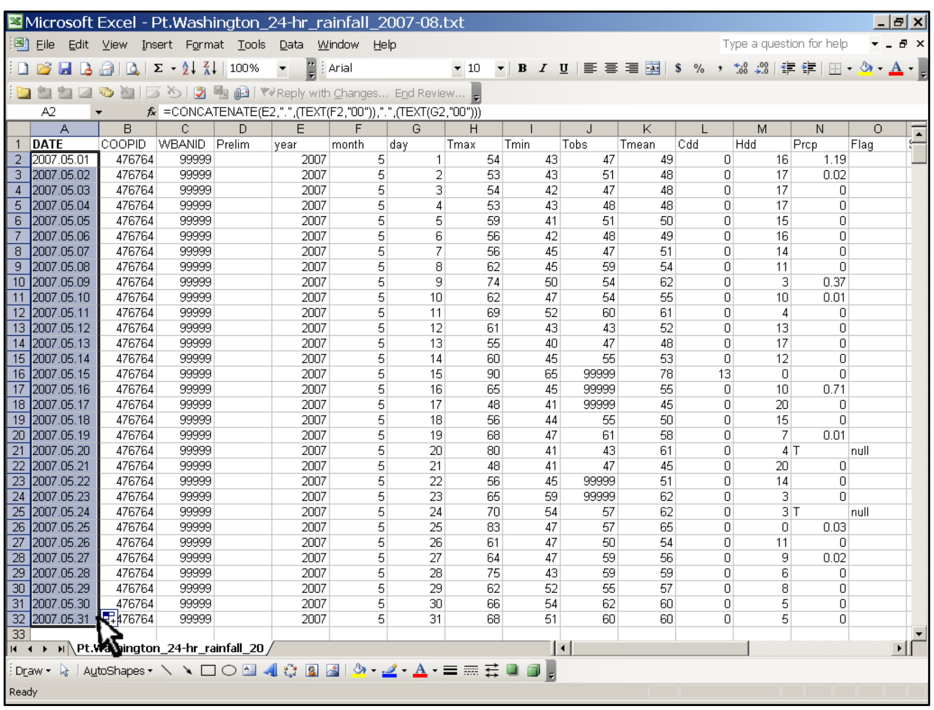
Note: In the case of data from COOP stations, common time-reporting values will not include the hour, since data are reported on a daily basis. When performing a table join to *E. coli* monitoring data – or to data on other explanatory variables, such as hourly winds – join on those tables’ “DATE” columns, rather than “DATE_HRS_UTC”.

For procedures on importing and “joining” tables using queries in MS Access, see Section 6 of this report, “Formatting Data for Model-Building.”

B.5.12. To create values for “DATE” (column A), type the following function in cell A2:

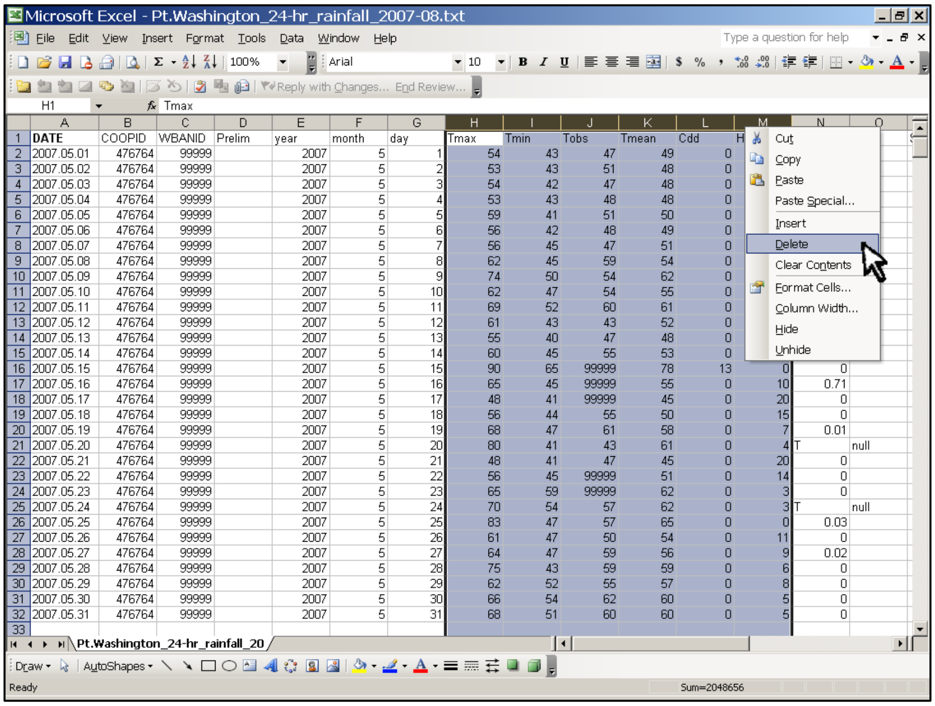
= CONCATENATE(E2,".",(TEXT(F2,"00")),".",(TEXT(G2,"00")))

Hit “Enter”. This should return a date value (e.g., “2007.05.01” for May 1, 2007). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.

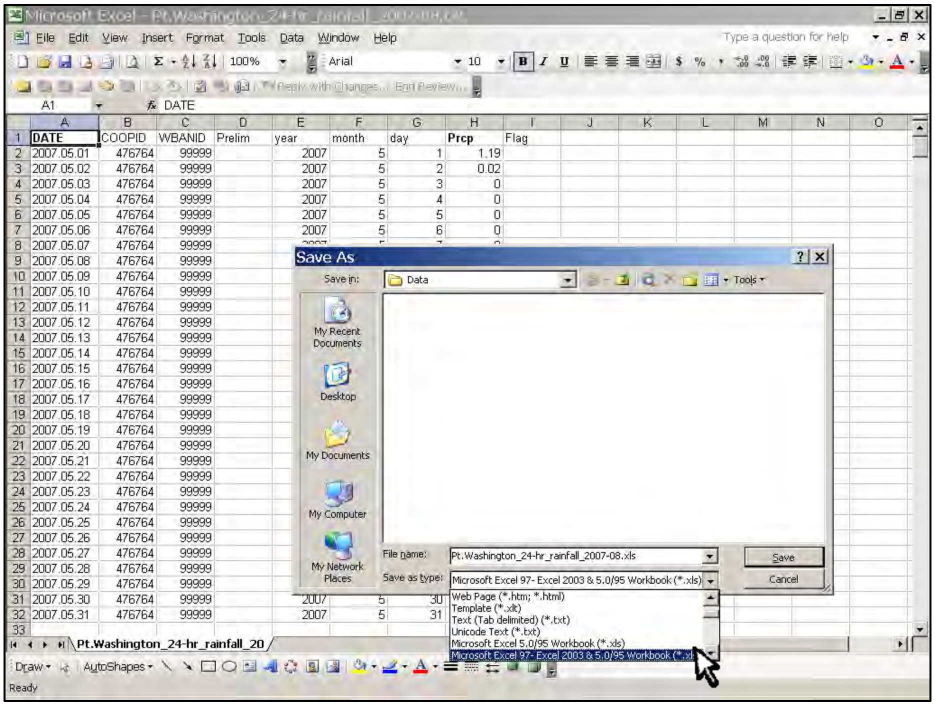


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	DATE	COOPID	WBANID	Prelim	year	month	day	Tmax	Tmin	Tobs	Tmean	Cdd	Hdd	Prcp	Flag
2	2007.05.01	476764	99999		2007	5	1	54	43	47	49	0	16	1.19	
3	2007.05.02	476764	99999		2007	5	2	53	43	51	48	0	17	0.02	
4	2007.05.03	476764	99999		2007	5	3	54	42	47	48	0	17	0	
5	2007.05.04	476764	99999		2007	5	4	53	43	48	48	0	17	0	
6	2007.05.05	476764	99999		2007	5	5	59	41	51	50	0	15	0	
7	2007.05.06	476764	99999		2007	5	6	56	42	48	49	0	16	0	
8	2007.05.07	476764	99999		2007	5	7	56	45	47	51	0	14	0	
9	2007.05.08	476764	99999		2007	5	8	62	45	59	54	0	11	0	
10	2007.05.09	476764	99999		2007	5	9	74	50	54	62	0	3	0.37	
11	2007.05.10	476764	99999		2007	5	10	62	47	54	55	0	10	0.01	
12	2007.05.11	476764	99999		2007	5	11	69	52	60	61	0	4	0	
13	2007.05.12	476764	99999		2007	5	12	61	43	43	52	0	13	0	
14	2007.05.13	476764	99999		2007	5	13	55	40	47	48	0	17	0	
15	2007.05.14	476764	99999		2007	5	14	60	45	55	53	0	12	0	
16	2007.05.15	476764	99999		2007	5	15	90	65	99999	78	13	0	0	
17	2007.05.16	476764	99999		2007	5	16	65	45	99999	55	0	10	0.71	
18	2007.05.17	476764	99999		2007	5	17	48	41	99999	45	0	20	0	
19	2007.05.18	476764	99999		2007	5	18	56	44	55	50	0	15	0	
20	2007.05.19	476764	99999		2007	5	19	68	47	61	58	0	7	0.01	
21	2007.05.20	476764	99999		2007	5	20	80	41	43	61	0	4	T	null
22	2007.05.21	476764	99999		2007	5	21	48	41	47	45	0	20	0	
23	2007.05.22	476764	99999		2007	5	22	56	45	99999	51	0	14	0	
24	2007.05.23	476764	99999		2007	5	23	65	59	99999	62	0	3	0	
25	2007.05.24	476764	99999		2007	5	24	70	54	57	62	0	3	T	null
26	2007.05.25	476764	99999		2007	5	25	83	47	57	65	0	0	0.03	
27	2007.05.26	476764	99999		2007	5	26	61	47	50	54	0	11	0	
28	2007.05.27	476764	99999		2007	5	27	64	47	59	56	0	9	0.02	
29	2007.05.28	476764	99999		2007	5	28	75	43	59	59	0	6	0	
30	2007.05.29	476764	99999		2007	5	29	62	52	55	57	0	8	0	
31	2007.05.30	476764	99999		2007	5	30	66	54	62	60	0	5	0	
32	2007.05.31	476764	99999		2007	5	31	68	51	60	60	0	5	0	
33															

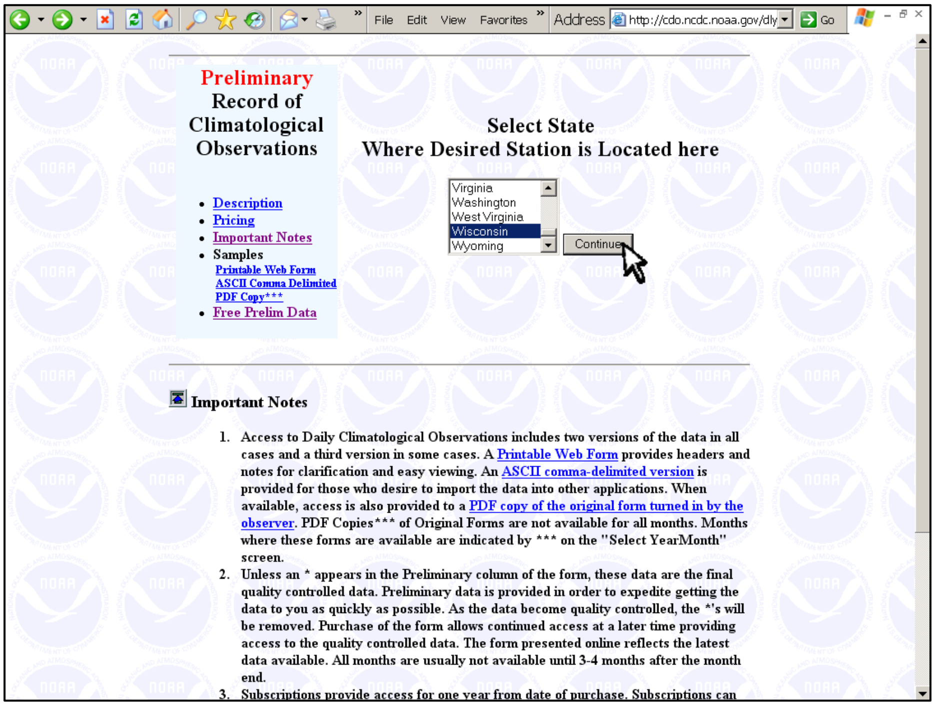
B.5.13. Select columns H (“Tmax”) through M (“Hdd”) by clicking on the header over column H and, while holding the “Shift” key, hitting the arrow key until column M is highlighted as well. Right-click and select “Delete”.



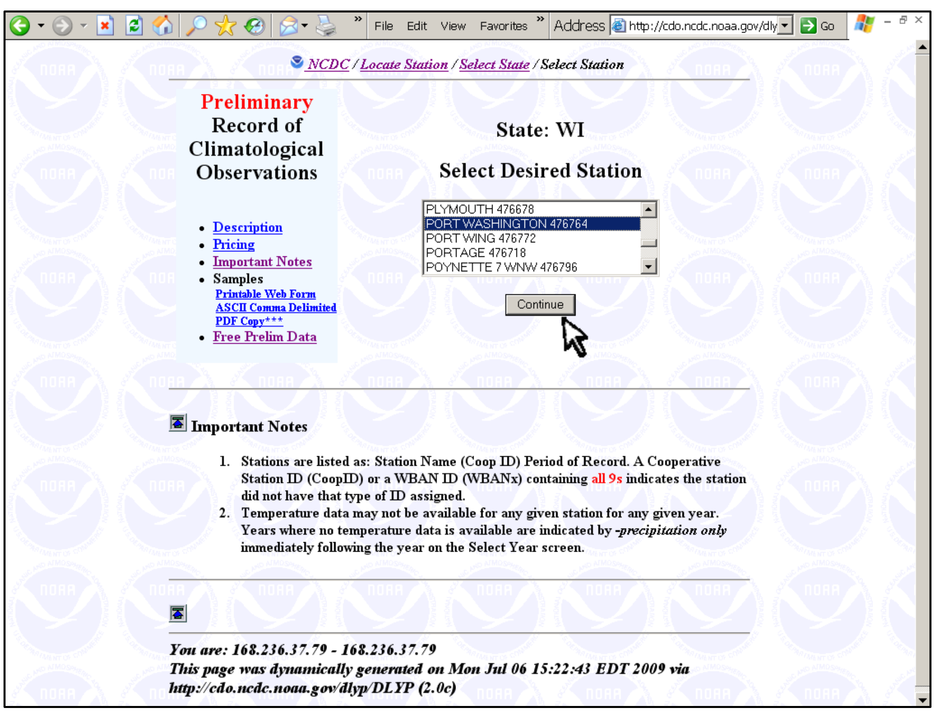
B.5.14. Save the table as an MS Excel file by selecting “File” > “Save as…” from the MS Excel pull-down menu, choosing the most recent version of MS Excel as the file type (*.xls), and naming the file something like “[location]_24-hr_precip_[year(s)].xls”.



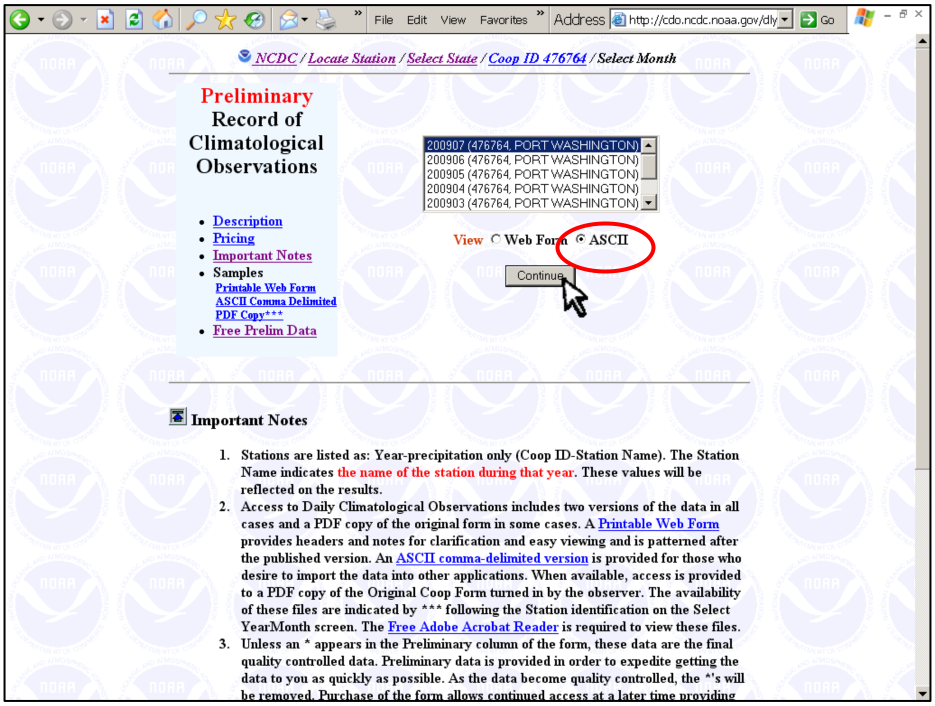
B.5.15. If you are building your model *mid-season*, there may be a gap between the most recent verified data (see section B.5.4) and today's real-time data. These gaps can be filled by downloading preliminary data. To access these data, open the "Preliminary Record of Climatological Observations" page at http://cdo.ncdc.noaa.gov/dlyp/DLYP. Select Wisconsin and click the "Continue" button.



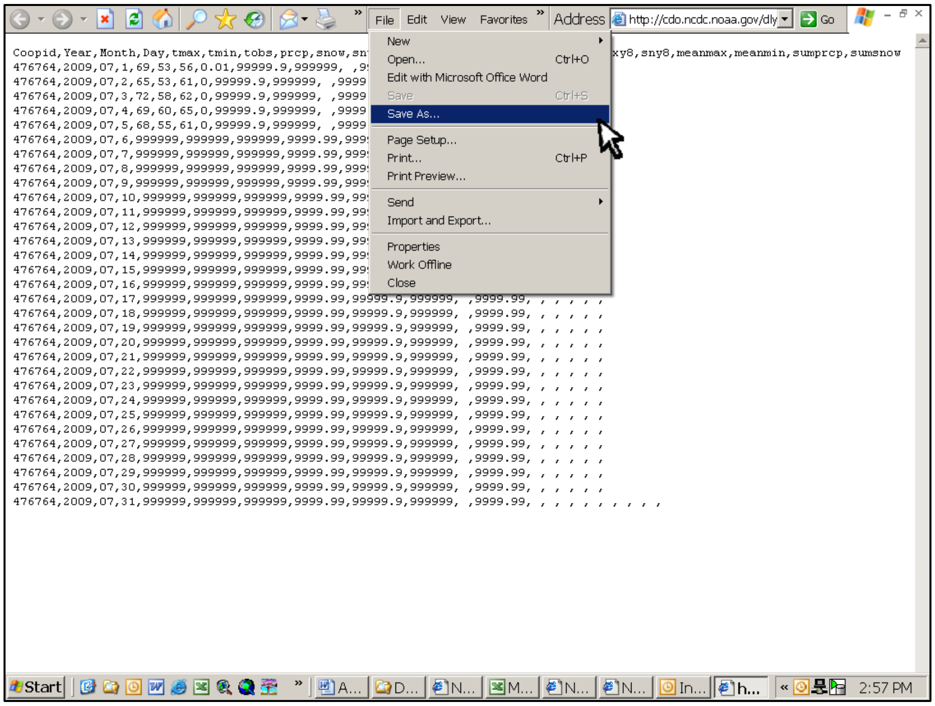
B.5.16. Select your station of interest and click the "Continue" button.



B.5.17. Select the year/month of interest, check the circle for "ASCII" (text), and click the "Continue" button (**Note:** Preliminary data can only be accessed one month at a time.)



B.5.18. A table of comma-delimited text will open in a new window. In your web browser, select "File" > "Save As...", name and save the file as a Text Document (.txt). To open, format, and save in MS Excel, follow the procedures in steps B.5.8 – B.5.14.



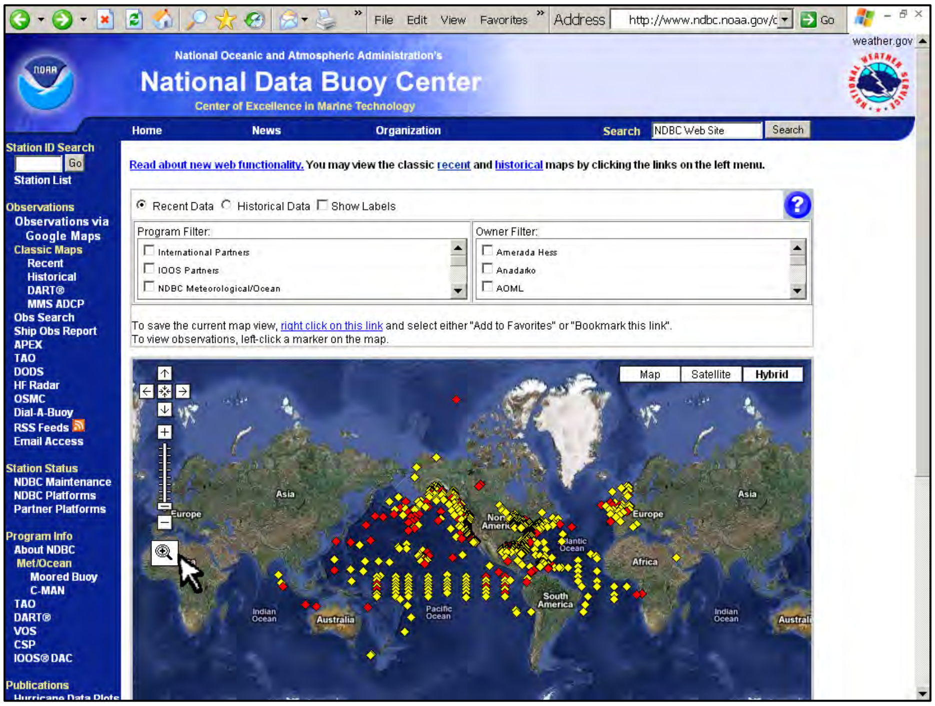
B.6. National Data Buoy Center

Provides access through a map interface to both real-time and historic/archived data from Great Lakes buoys and coastal marine automated network (C-MAN) stations. Pertinent variables may include air temperature, water temperature, wind speed, wind direction, and wave height.

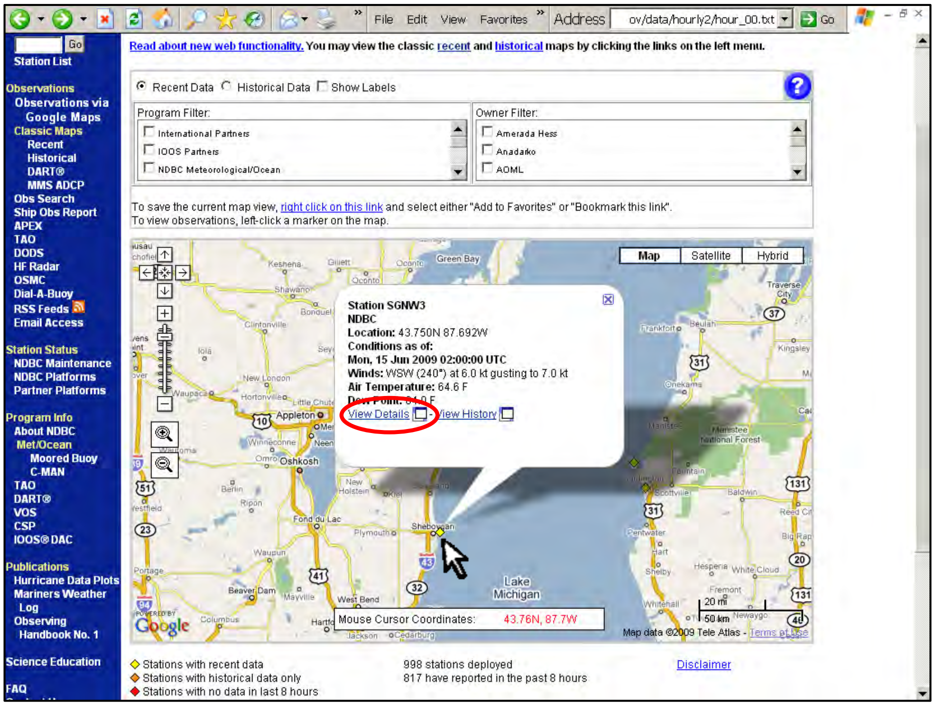
The National Data Buoy Center (NDBC) does not include surface weather observing stations. Real-time data from these stations can be accessed through the **Internet Weather Source** (B.2) or through the **CoastWatch Great Lakes Node** (B.1), under "Surface Airways Stations" (www.glerl.noaa.gov/marobs/php/data.php?sta=6). Historic surface data can be accessed via the **Online Climate Data Directory** (B.4).

The default NDBC entryway is a Google Maps view displaying marine and coastal observation points around the world. This page can be slow to draw. Faster-drawing, non-Google Maps interfaces are available for individual regions, including the Western- and Eastern Great Lakes at www.ndbc.noaa.gov/rmd.shtml (for current data) or www.ndbc.noaa.gov/hmd.shtml (for historic/archived data)

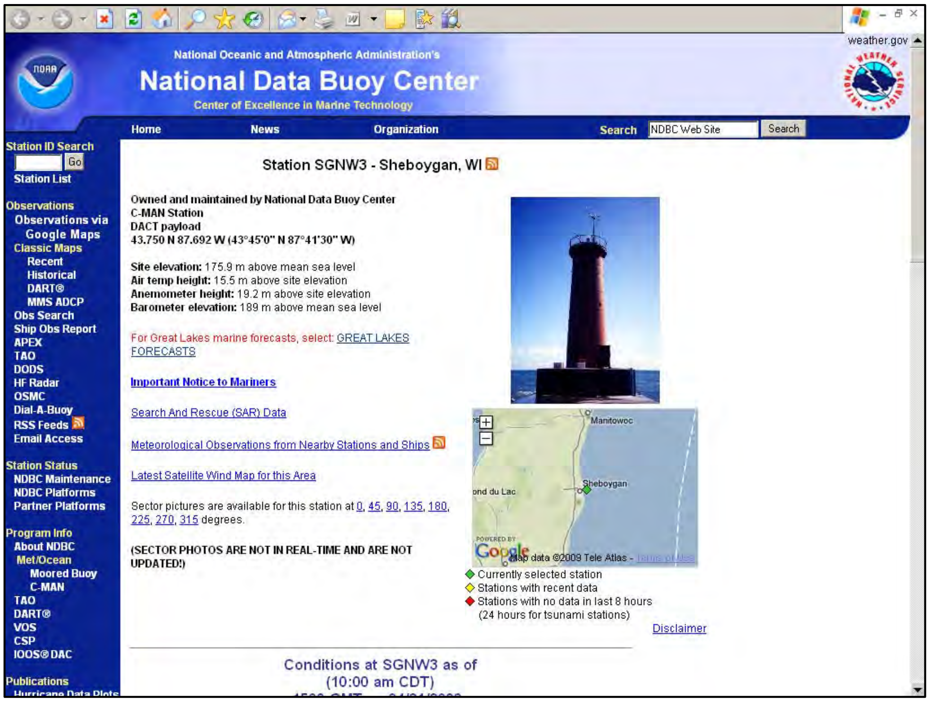
B.6.1. Open the NDBC main page at www.ndbc.noaa.gov. Zoom into the Great Lakes and/or the vicinity of your beach by clicking on the magnifying glass icon on the left side of the map and then drawing a box (click and drag) around the area. Pan the map in any direction by clicking a point with the map, holding, and dragging it. If desired, you can change the map background from the default "Satellite" image to "Map" (roads, parks, city/place names) or "Hybrid" (satellite + map) by clicking the corresponding button in the upper right corner of the map.



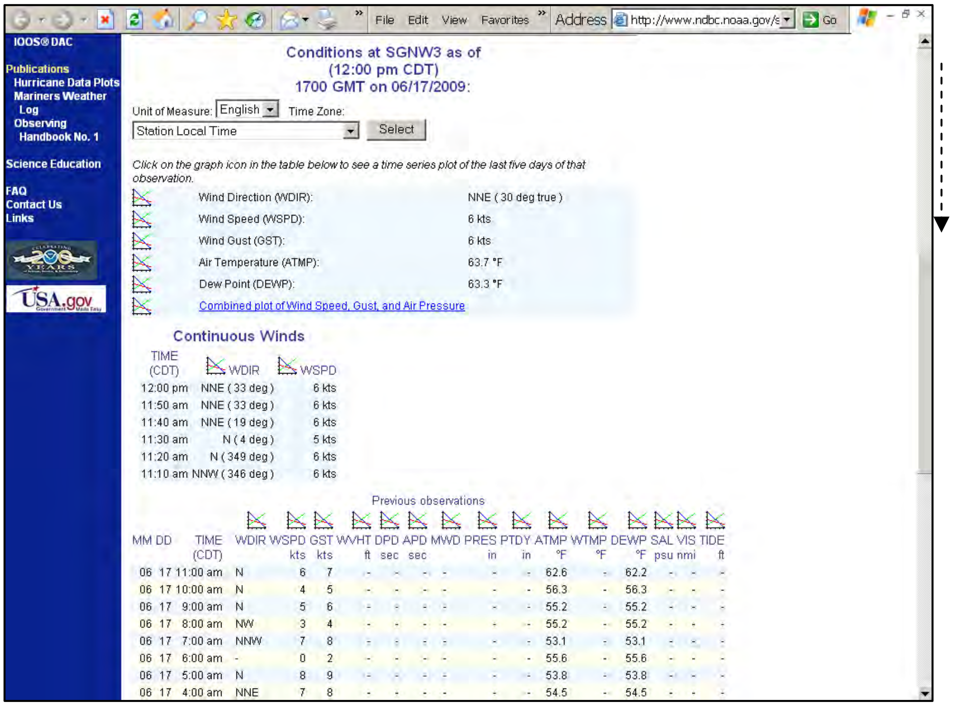
B.6.2. Once you have navigated to the vicinity of your beach, select a buoy or station by clicking on it in the map. This will open a pop-up window with a summary of the most recent conditions at that site. (**Note:** the time listed in this window is in UTC; however, data listed once you access the station page are in local time.)



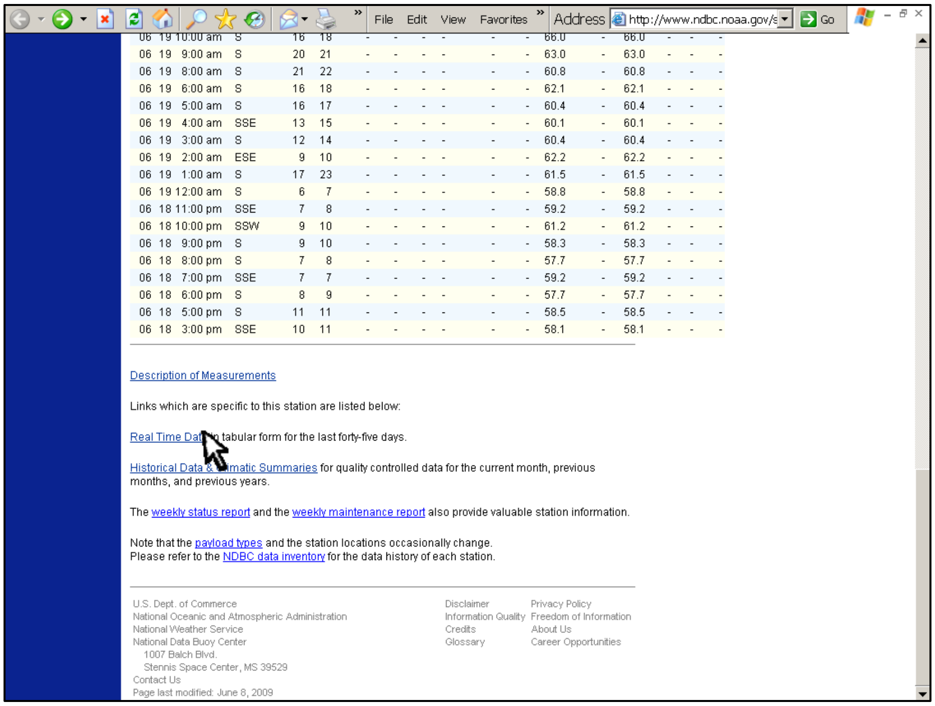
B.6.3. For more detailed data, click the "View Details" link in the pop-up window (above). This will open a new web page with station information.



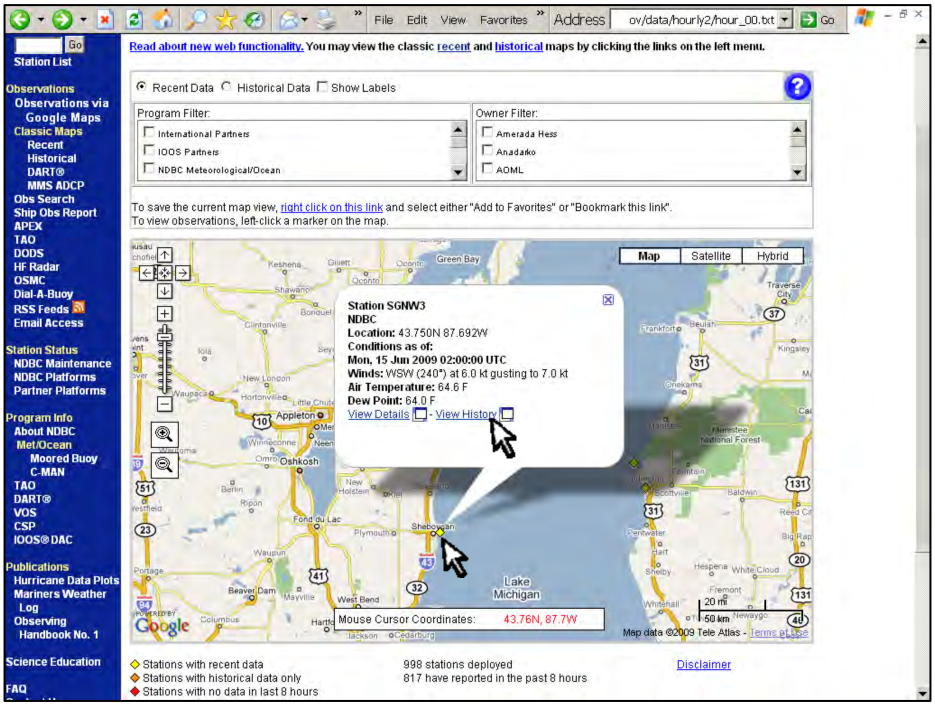
B.6.4. Scroll down the station page to view a detailed list of current and recent conditions – for those variables measured at the site – listed by local time.



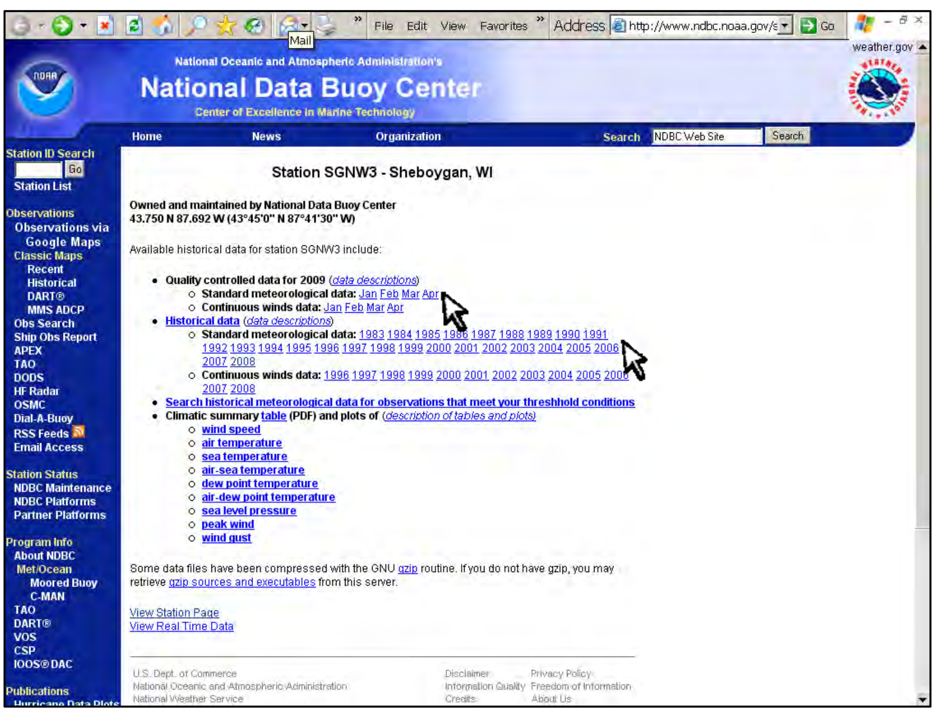
B.6.5. For continuous data over the previous 5 or 45 days scroll to the bottom of the station page and click the “Real-Time Data” link. (See B.6.8 – B.6.15 below for directions on downloading and formatting tabular data from NBCD.)



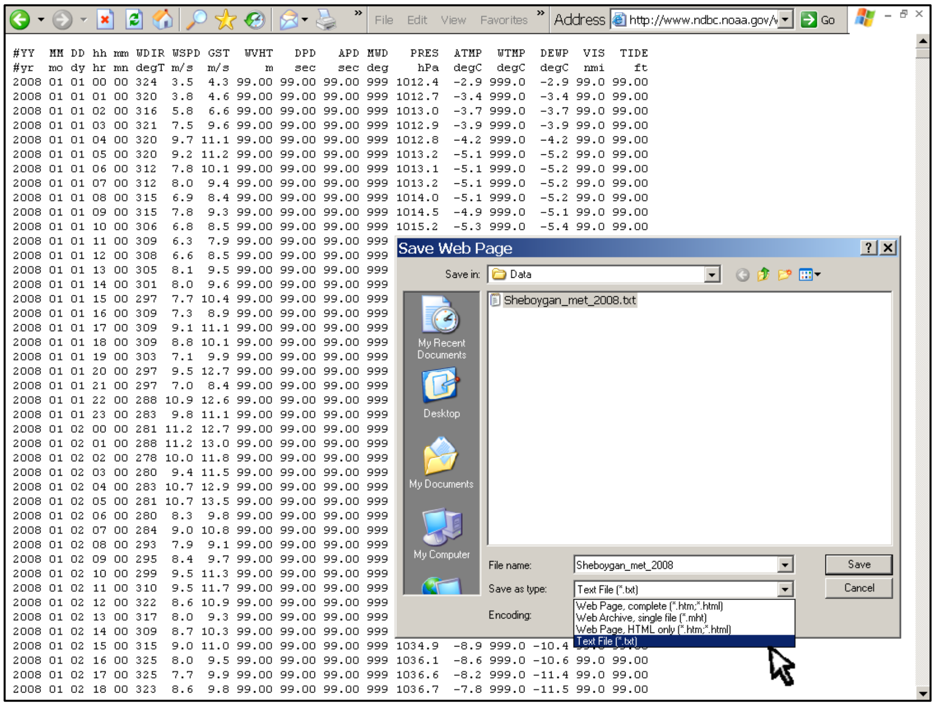
B.6.6. To download historic data return to the main (map) page, and in the pop-up window for the station, click the “View History” link. A new page will open.



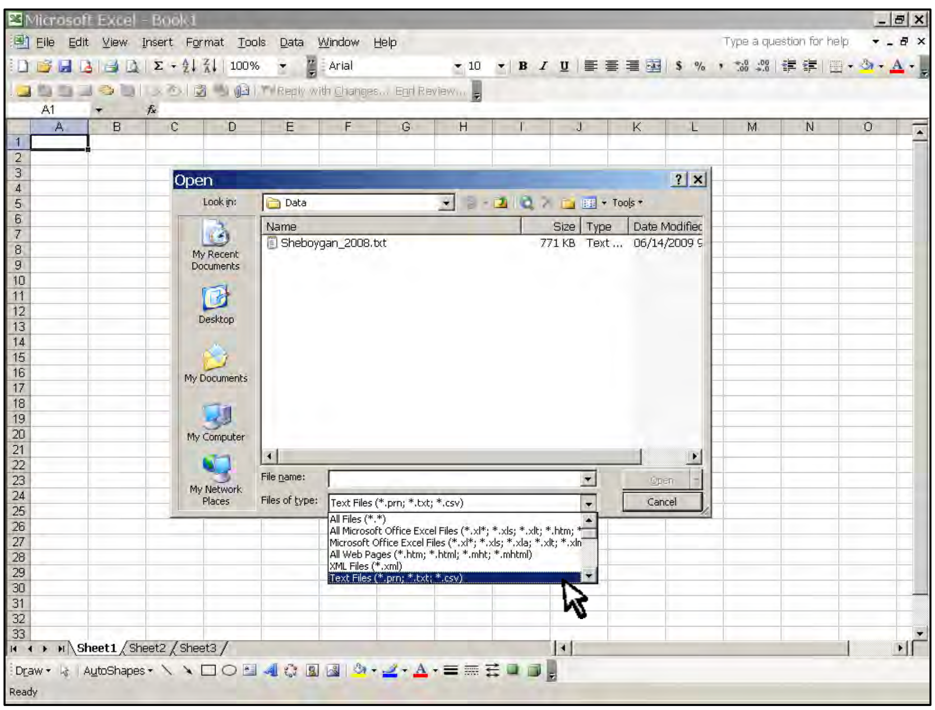
B.6.7. In the new page that opens, click on either a month in the current year (“Quality controlled data for…”) or a previous year (“Historic Data”) to the right of “Standard meteorological data”.



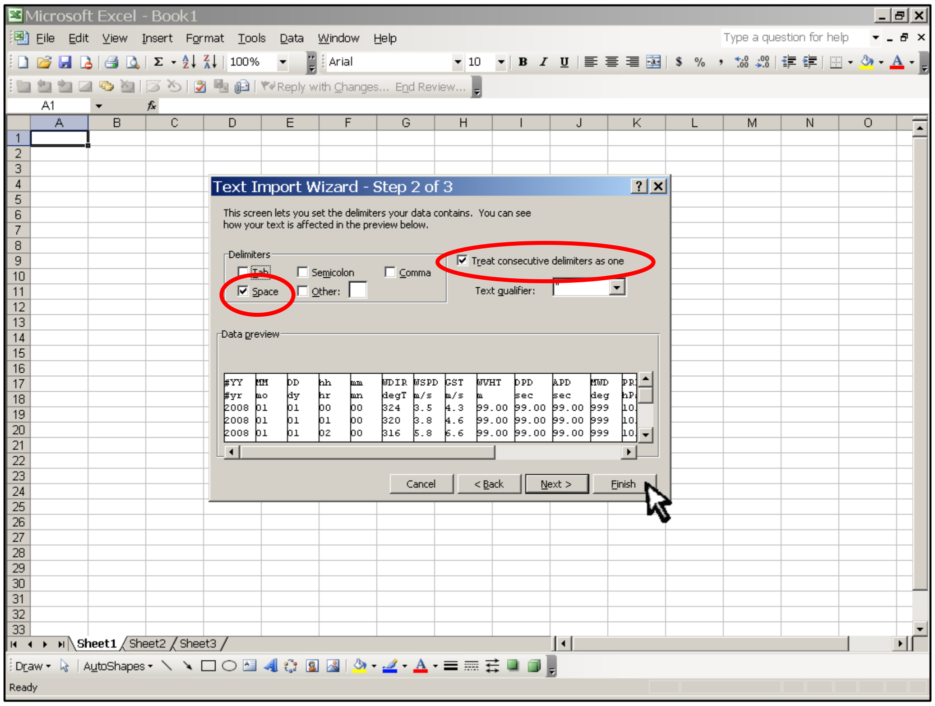
B.6.8. A new web page will open listing the data in tabular form. To save this data to your hard drive, click “File” > “Save as…” in your Web browser, name something like “[location]_met_[year]”, select “Text File” (*.txt) as the type, and click “Save”.



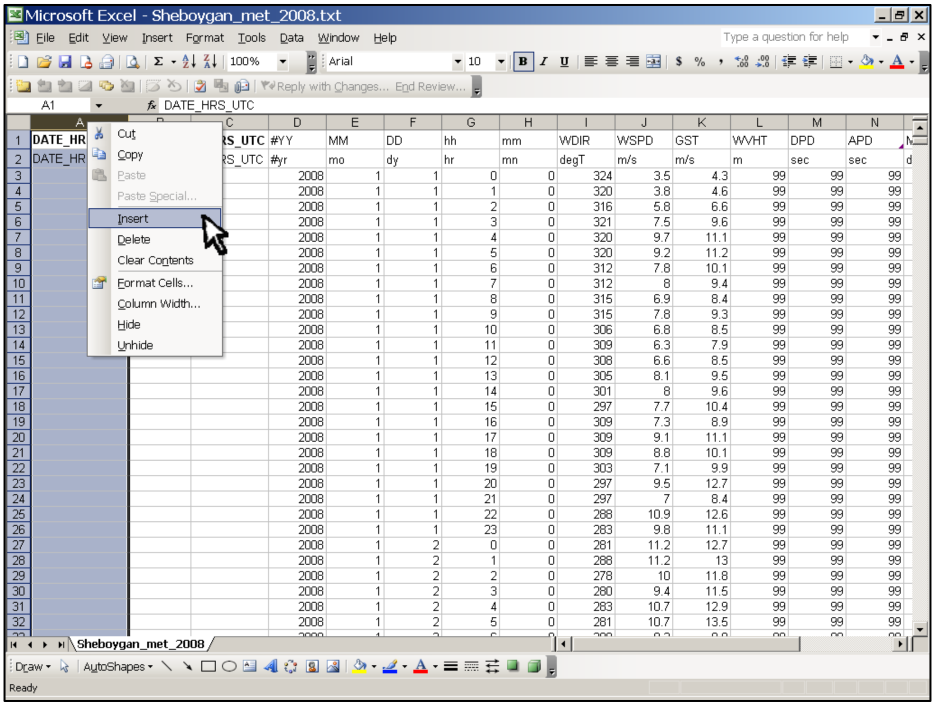
B.6.9. Open MS Excel. From the pull-down menu in the upper left corner, select “File” > “Open”. In the “Open” window, select “Text Files” from the “Files of type:” pull-down menu. Double-click on the text file you created in B.6.8.



B.6.10. A “Text Import Wizard” will open. In Step 1, check the circle for “Delimited” and click the “Next>” button. In Step 2 (below), check the boxes for “Space” and “Treat consecutive delimeters as one”. Click the “Finish” button.



B.6.11. Following the procedure in A.1.7, insert three new columns (A-C) named “DATE_HRS_UTC”, “DATE”, and “HOURS_UTC”. If you choose to keep row 2 [units], repeat column names in row 2 of columns A-C.

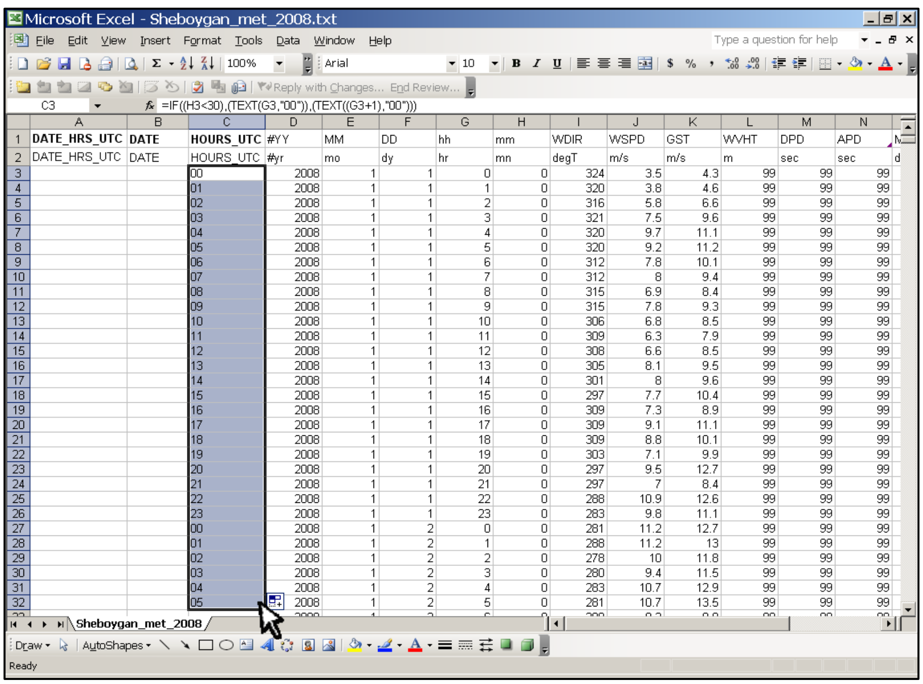


Note: As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to “join” the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and “joining” tables using queries in MS Access, see Section 6 of this report, “Formatting Data for Model-Building.”)

B.6.12. To create values for “HOURS_UTC” (column C) type the following function in cell C3:

=IF((H3<31),(TEXT(G3,"00")),(TEXT((G3+1),"00")))

Hit “Enter”. The result should be a two-digit, rounded value for the hour-of-day in UTC (Central Daylight Time + 5 hours.). Copy this function by clicking in the lower right corner of cell C3 and dragging to the end of the table.



Note: In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Be sure to include period-separators between the year, month, day, and hour (e.g., “2003.05.29.12” as opposed to “2003052912”). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

B.6.13. To calculate “DATE” (column B), click on cell B2 and type the following function:

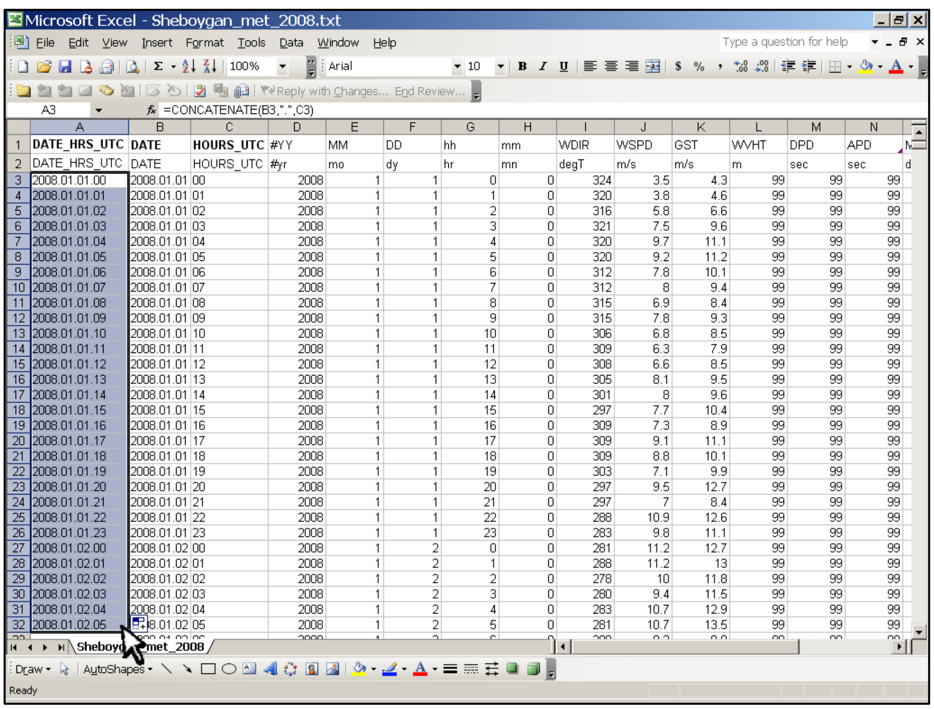
=CONCATENATE(D3,".",(TEXT(E3,"00")),".",(TEXT(F3,"00")))

Hit “Enter”. This should return a date value (e.g., “2008.01.01” for January 1, 2008). Copy this function by clicking in the lower right corner of cell B3 and dragging to the end of the table.

B.6.14. To calculate “DATE_HRS_UTC” (column A), click on cell A2 and type the following function:

=CONCATENATE(B3,".",C3)

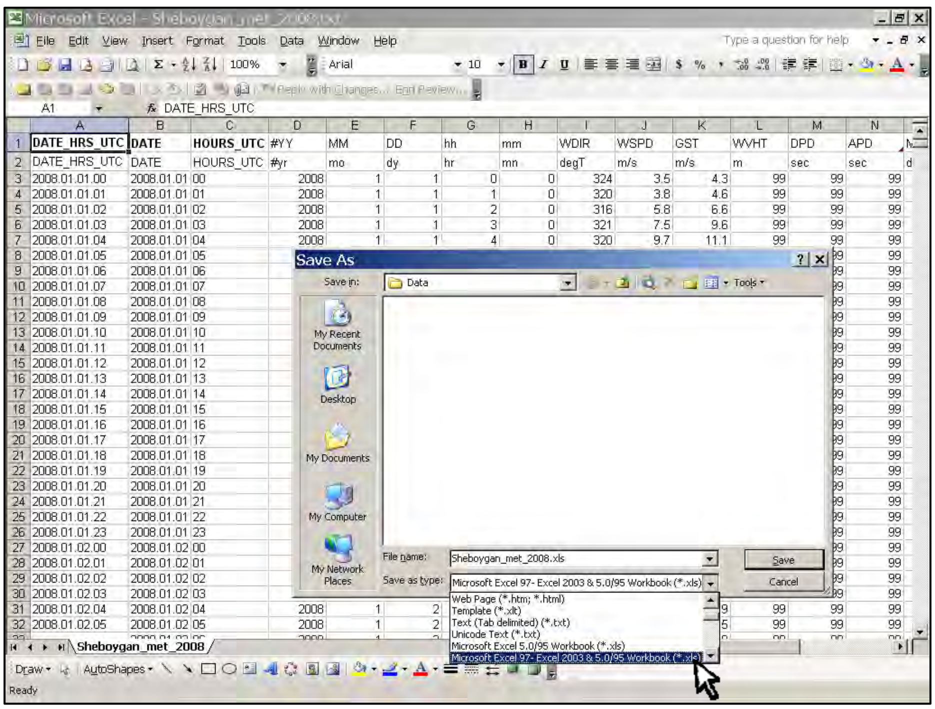
Hit “Enter”. This should return a date-plus-hour value (e.g., “2008.01.01.01” for 01:00 hours UTC on January 1, 2008). Copy this function by clicking in the lower right corner of cell A3 and dragging to the end of the table.



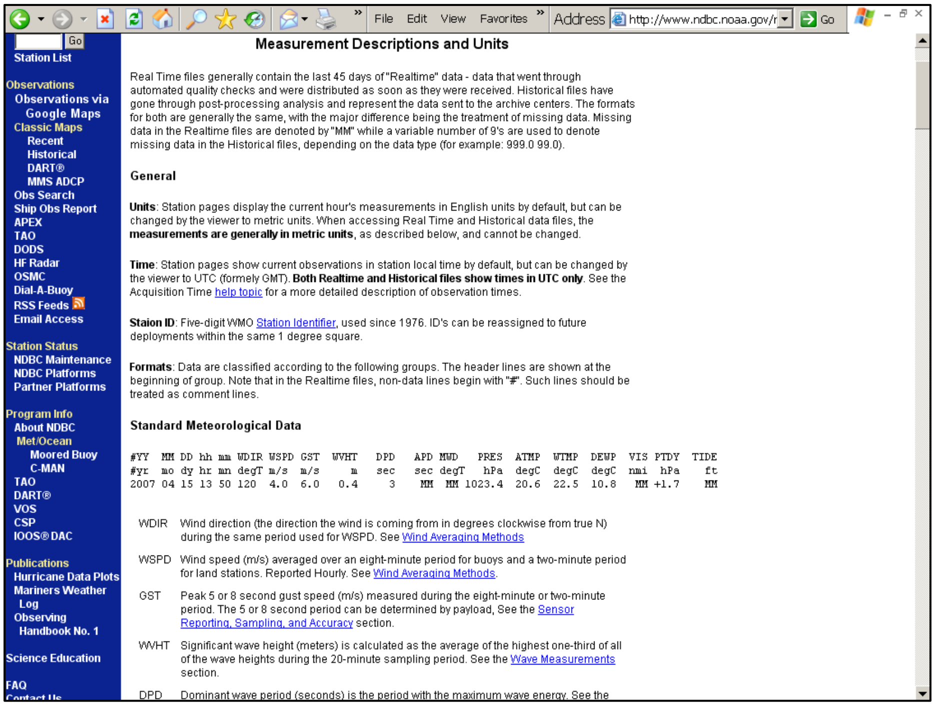
For additional column and row editing and formatting procedures, and for different functions, see the MS Excel Help menu.

Note: If you are building your model *mid-season*, there will be a gap of several weeks between the most up-to-date, quality controlled archived data (accessible through the “Historic Data” page) and today’s real-time data. You can cover this gap by downloading preliminary data for the past 45 days, which are accessible through the “Station” page (see step B.6.5).

B.6.15. Before closing, use the MS Excel pull-down menu to select “File” > “Save As…”, set the file type as the latest version of MS Excel (*.xls), and click the “Save” button.



B.6.16. For descriptions of the different variables listed in the NDCB data tables, click on the “Descriptions of Measurements” link on the station page, or alternatively on the “data descriptions” links on the historic data page. These links lead to the following page: www.ndbc.noaa.gov/measdes.shtml.

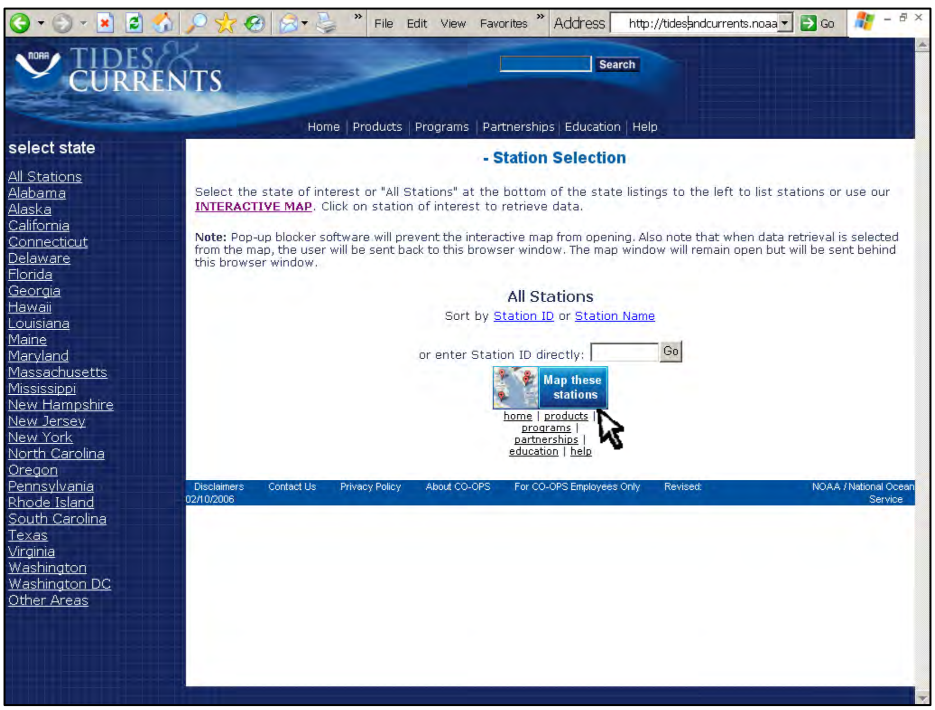


B.7. **Tides & Currents**

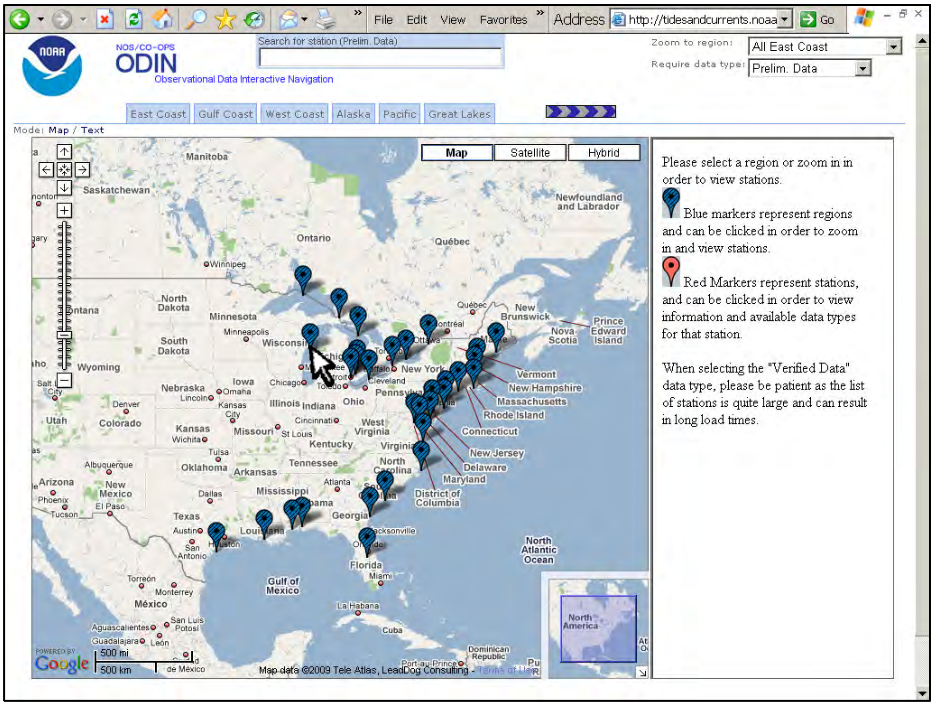
Provides access through a map interface to both real-time and archived/historic data on lake levels, as measured at national water level observation network (NWLON) stations.

Not including connecting channels between the Great Lakes, there are less than three-dozen stations measuring water levels across the Great Lakes. Since water levels are not uniform across a given lake, readings from a particular station may be more-or-less relevant to your beach. As discussed in Section 4 of this report ("Using Online Sources vs. Field-Collected Data"), the partial regression slope coefficient for water levels in your model and the coefficient's associated *p* value (measure of statistical significance) will determine whether or not data collected at a remote NWLON station will be a useful predictor of *E. coli* at your beach.

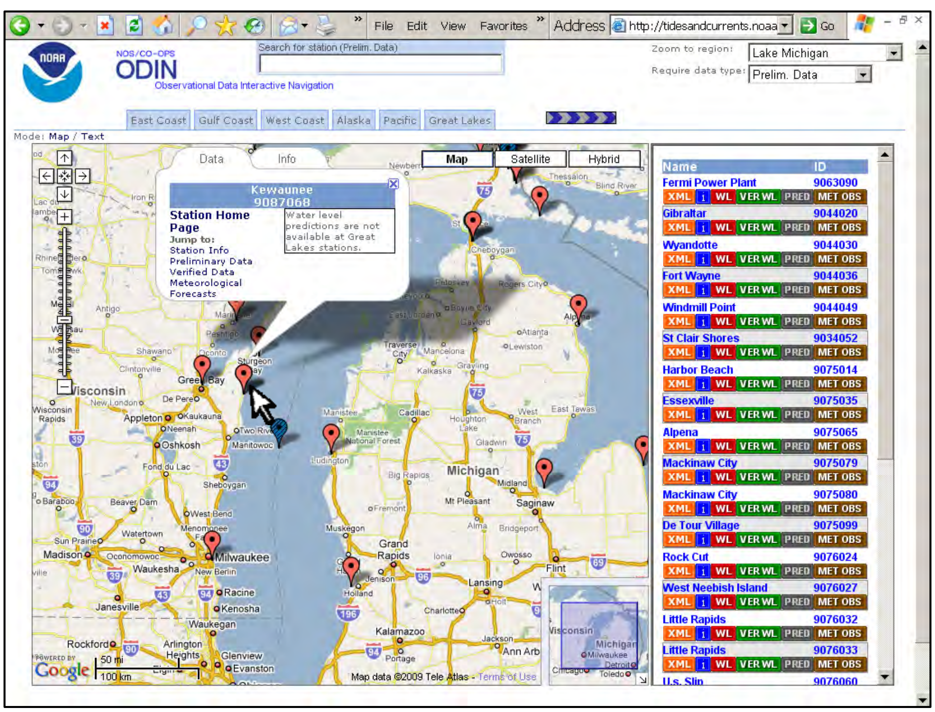
B.7.1. Open the **Tides & Currents** "Station Selections" page at http://tidesandcurrents.noaa.gov/station_retrieve.shtml. Click the icon in the center of the page titled "Map these stations". This will open a Google Maps interface.



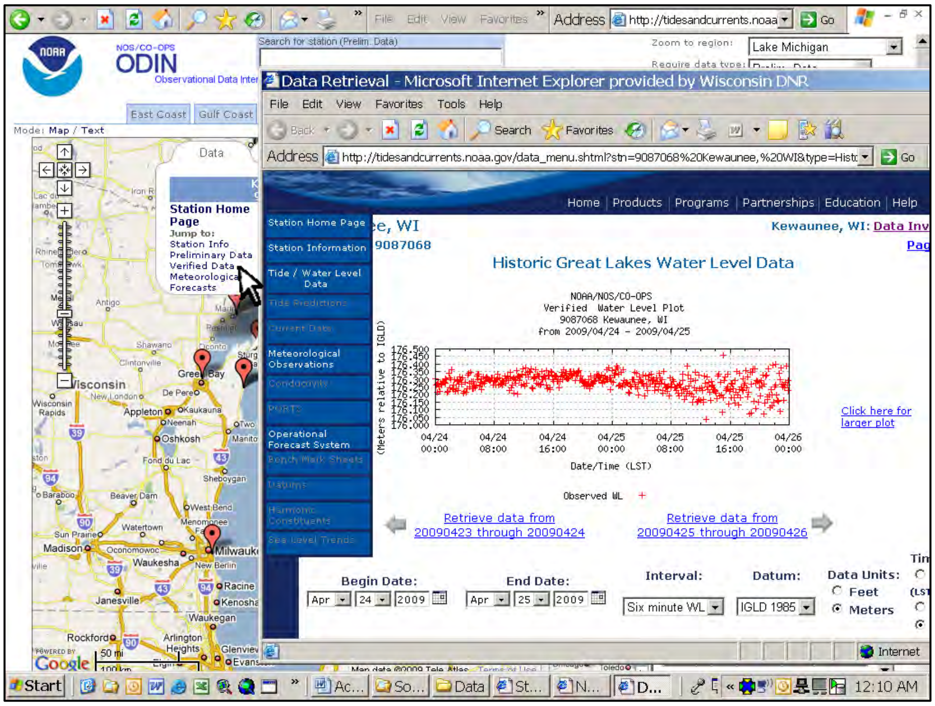
B.7.2. In the map interface, locate the blue marker closest to your region (e.g., Lake Michigan) and click on it to automatically zoom-in to that region.



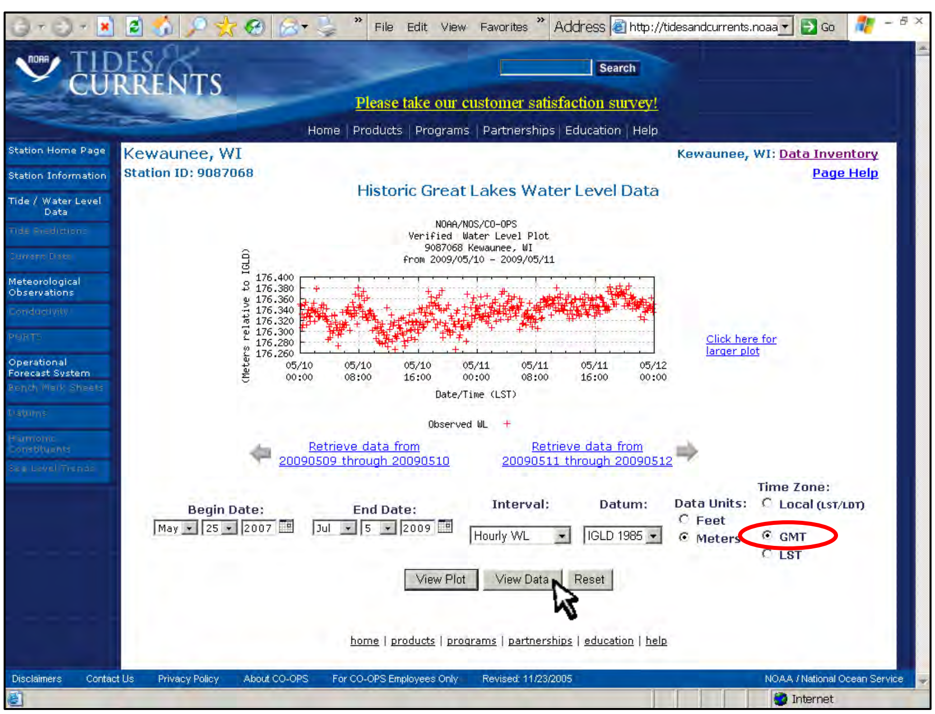
B.7.3. In the zoomed-in map, select a station by clicking on a red marker. This will open a pop-up window with a "Data" tab, containing links to *preliminary* (recent) and *verified* (historic) data on lake levels, as well as an "Info" tab listing the station's geographic coordinates.



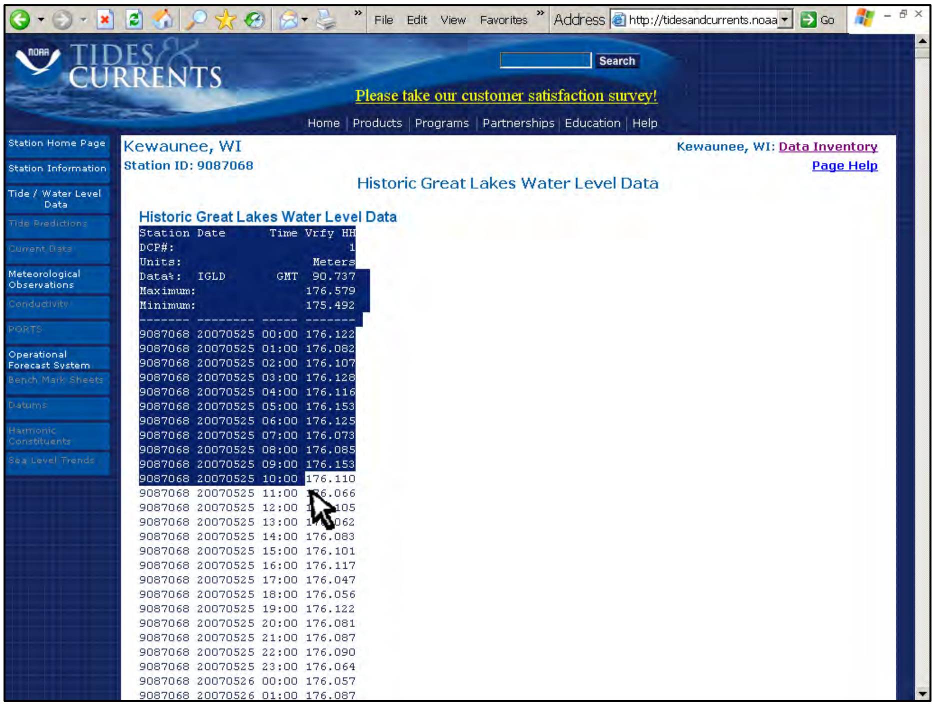
B.7.4. In the pop-up window, click on the “Verified Data” link. This will open a separate “Historic Data” page showing lake level, graphically, as measured over time in meters relative to the 1985 International Great Lakes Datum (IGLD).



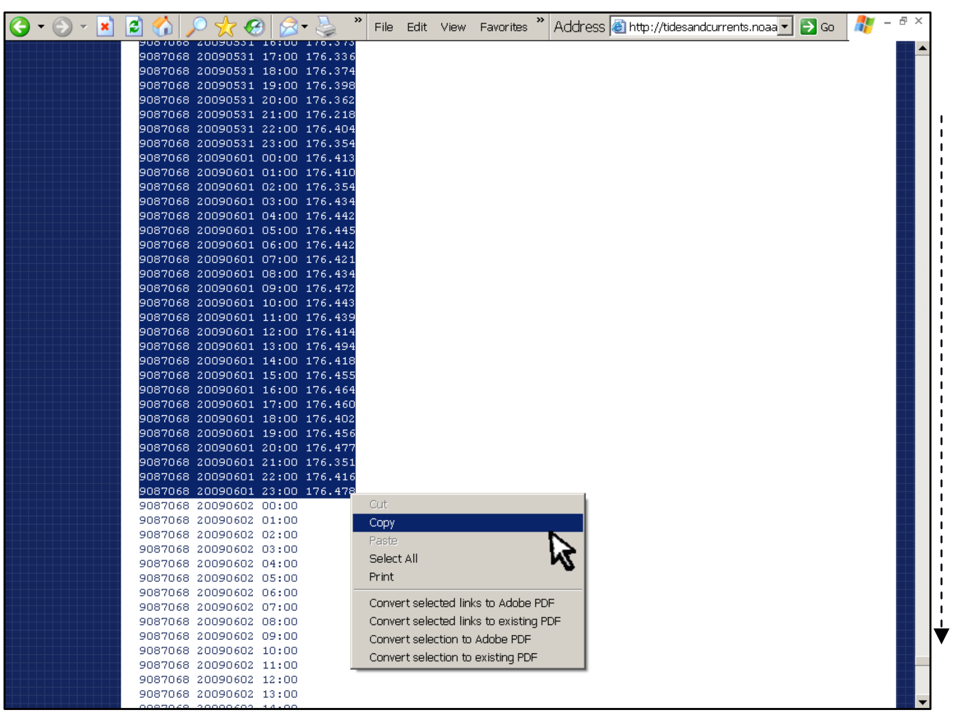
B.7.5. Using the pull-down menus, select a “Begin Date” and an “End Date” and your desired “Data Units”. Set “Interval” to “Hourly WL,” “Datum” to “IGLD 1985,” and “Time Zone” to “GTM” (effectively the same as UTM). Click the “View Data” button.



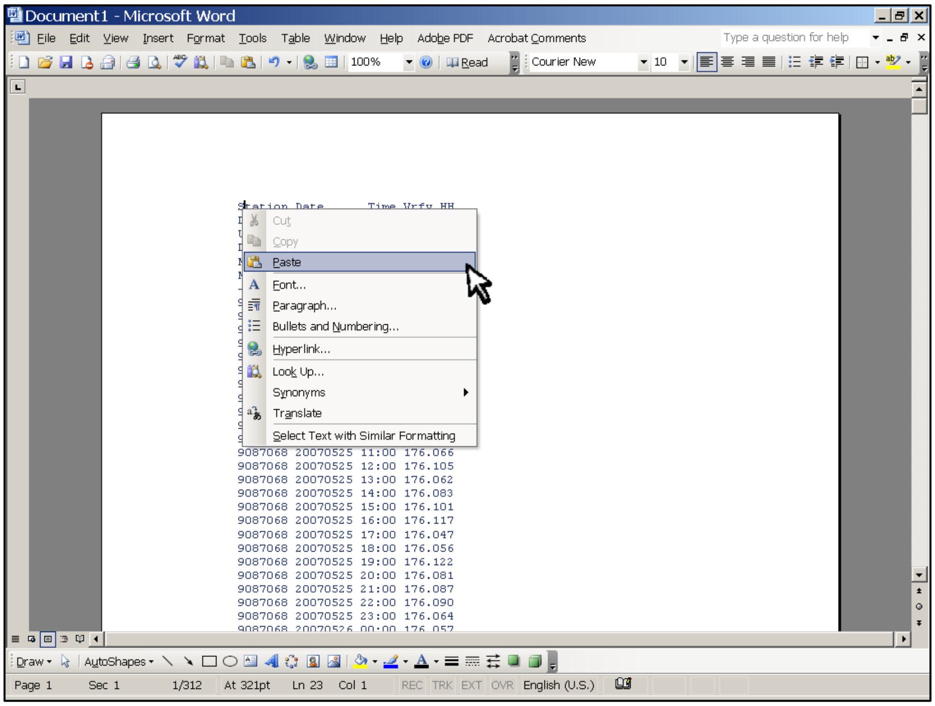
B.7.6. A table will open, with lake level values in the right column. Click just to the left of the word "Station", hold and drag the mouse down and to the right, until two or more rows are highlighted.



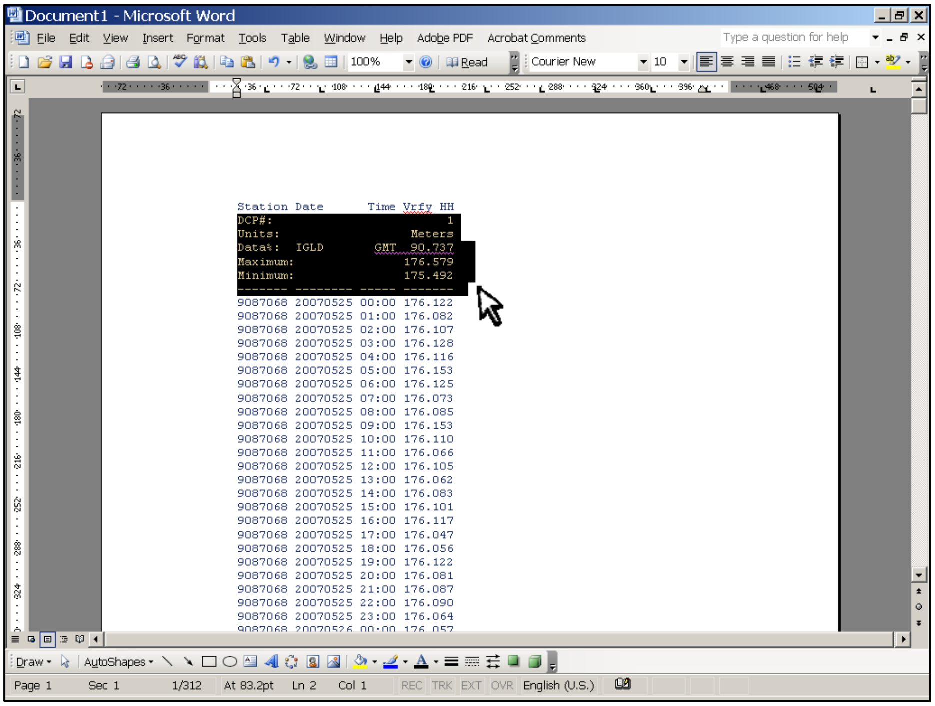
B.7.7. Scroll to the last row with lake level values. Hold the "Shift" key, and click to the right of the row to highlight. Right-click and select "Copy". (**Note:** *Verified* data end roughly a month prior to today. For more recent data, see B.7.17 – B.7.28.)



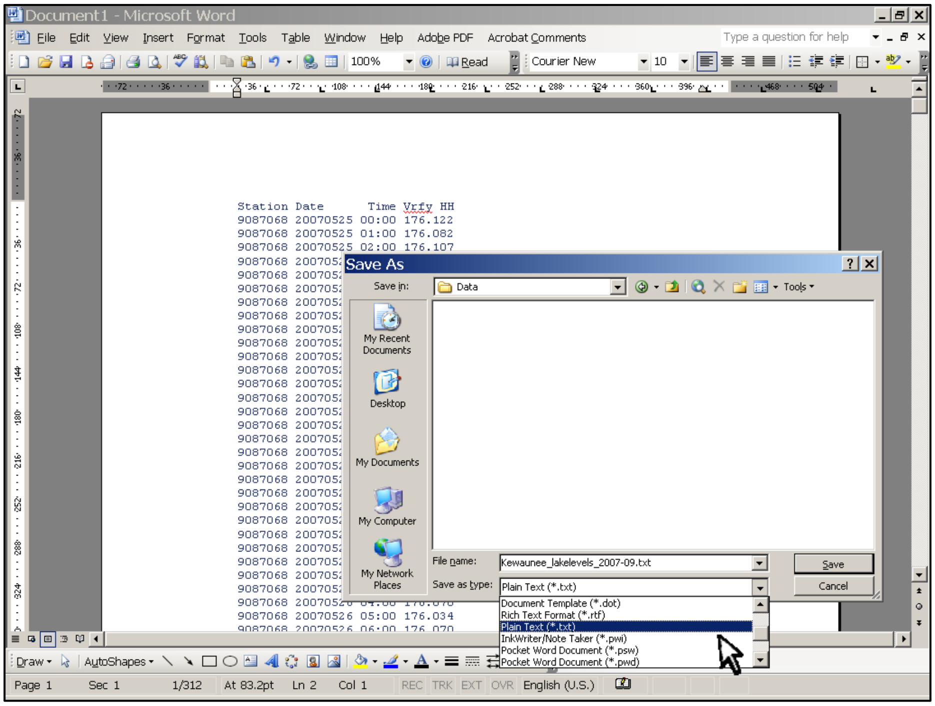
B.7.8. Open MS Word. Right-click at the top of a blank document and select “Paste”.



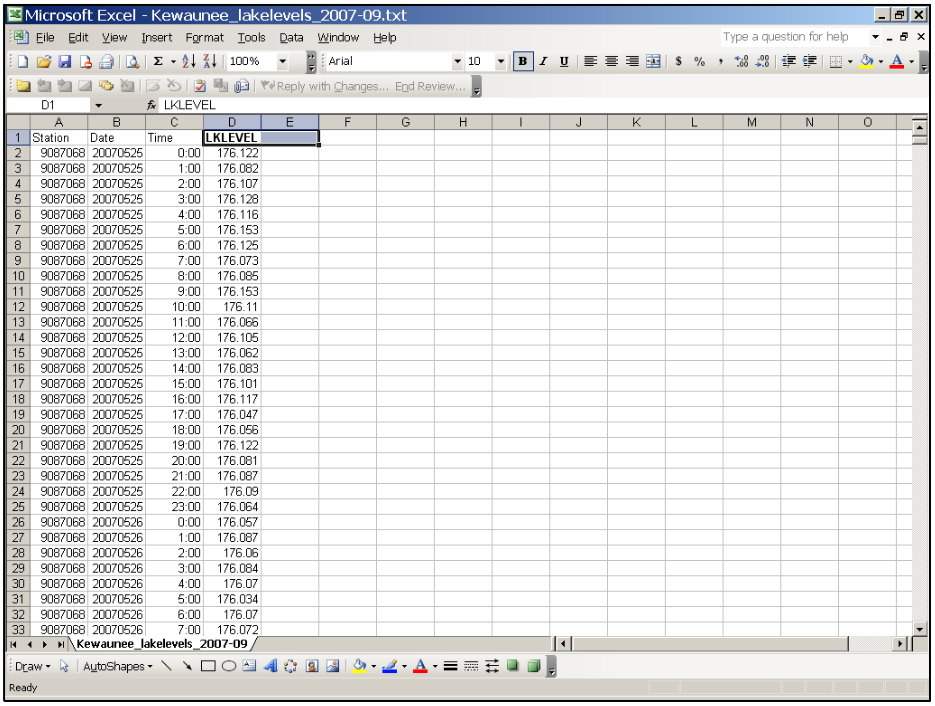
B.7.9. Click just to the left of the word “DCP#”, hold and drag the mouse downward and to the right, until rows two through seven (-------) arte highlighted. Hit the “Delete” key.



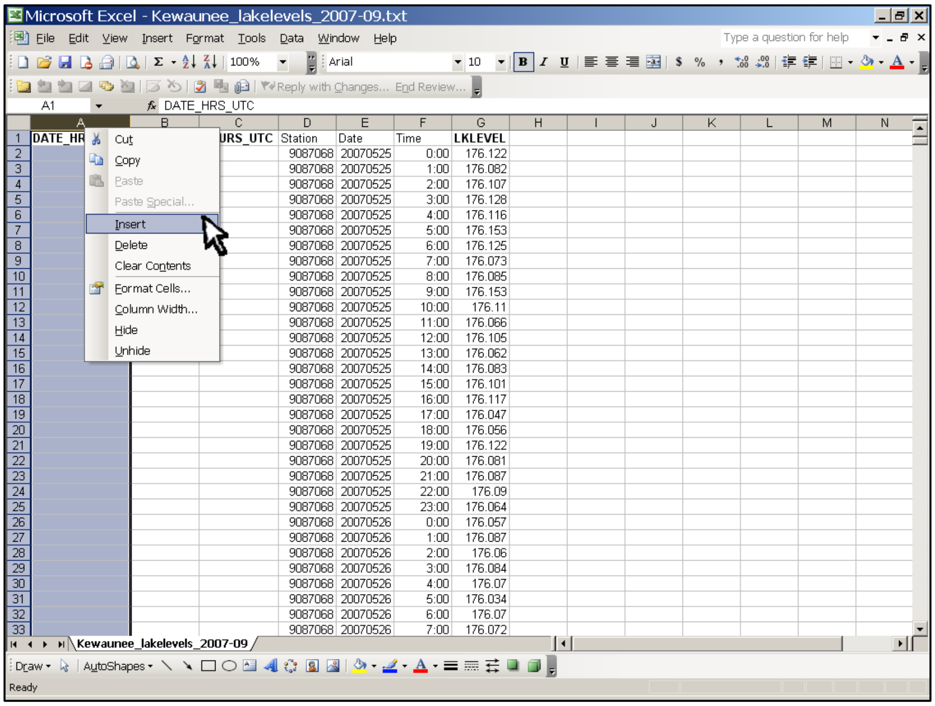
B.7.10. From the MS Word pull-down menu, select “File” > “Save as…”, name the file something like “[location]_lakelevels_[year(s)]”, and set the file type to “Plain Text” (*.txt). (In the “File Conversion” window that opens, select “Windows (Default)” as the text encoding system and click “OK”.)



B.7.11. Following the procedure in steps B.6.9-6.10, open the file in MS Excel. Rename cell D1 something like “LKLEVEL”. Delete text in cell E1.



B.7.12. Following the procedure in A.1.7, insert three new columns (A-C) named “DATE_HRS_UTC”, “DATE”, and “HOURS_UTC”.



For additional column and row editing and formatting procedures, and for different functions, see the MS Excel Help menu.

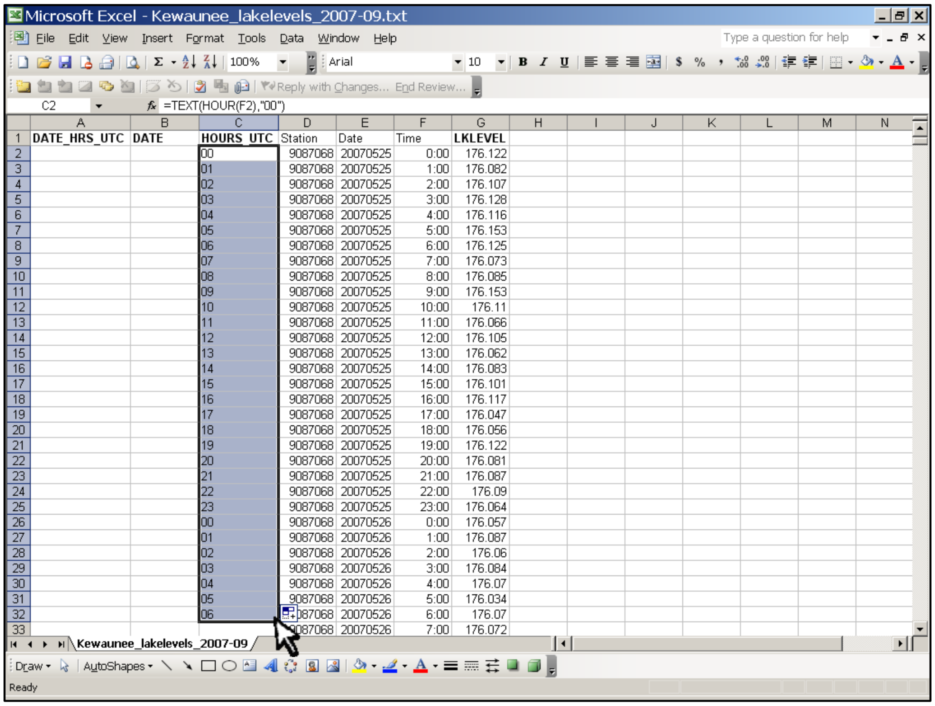
Note: The **Tides & Currents** data are reported in Greenwich Mean Time (GMT), which is effectively the same as UTC (Central Daylight Time + 5 hours.)

Note: As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to “join” the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and “joining” tables using queries in MS Access, see Section 6 of this report, “Formatting Data for Model-Building.”)

B.7.13. To create values for “HOURS_UTC” (column C) type the following function in cell C2:

=TEXT(HOUR(F2),"00")

Hit “Enter”. The result should be a two-digit, rounded value for the hour-of-day, in UTC (Central Daylight Time + 5 hours.). To copy, click on the lower right hand corner of the cell, hold and drag the mouse to the bottom of the table.



B.7.14. To calculate “DATE” (column B), click on cell B2 and type the following function:

=CONCATENATE((LEFT((E2),4)),".",(RIGHT((LEFT(E2,6)),2)),".",(RIGHT((LEFT(E2,8)),2)))

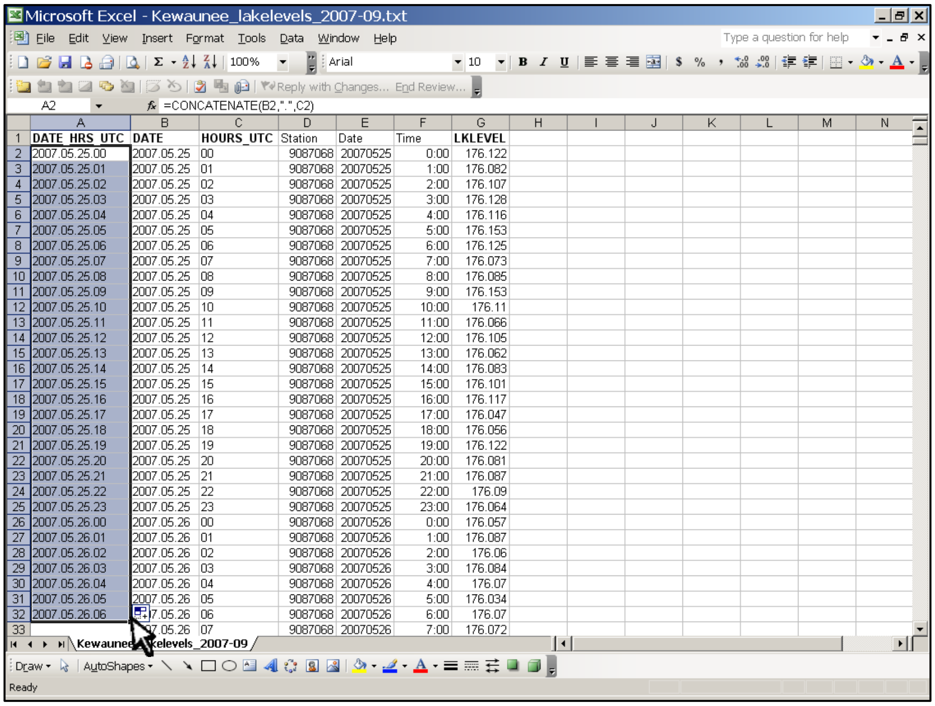
Hit “Enter”. This should return a date value (e.g., “2007.05.25” for May 25, 2007). Copy this function by clicking in the lower right corner of cell B2 and dragging to the end of the table.

Note: In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Be sure to include period-separators between the year, month, day, and hour (e.g., “2003.05.29.12” as opposed to “2003052912”). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

B.7.15. To calculate “DATE_HRS_UTC” (column A), click on cell A2 and type the following function:

=CONCATENATE(B2,".",C2)

Hit “Enter”. This should return a date-plus-hour value (e.g., “2006.05.25.01” for 01:00 hours UTC on May 25, 2006). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.

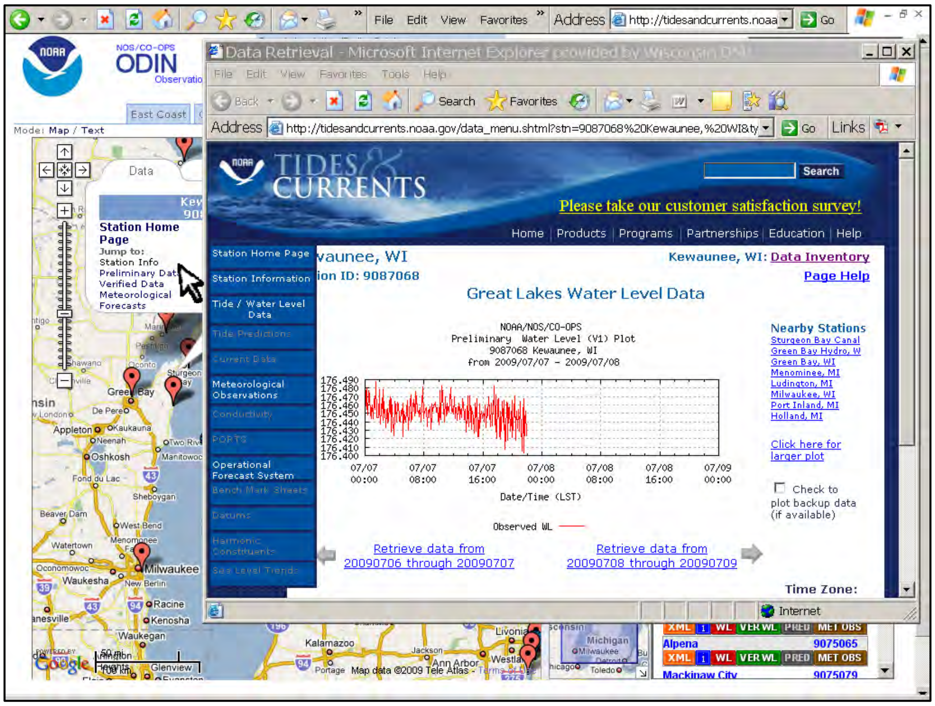


B.7.16. Save the table as an MS Excel file by selecting “File” > “Save as...” from the MS Excel pull-down menu in the upper left corner, choosing the most recent version of MS Excel as the file type (*.xls), and clicking the “Save” button.

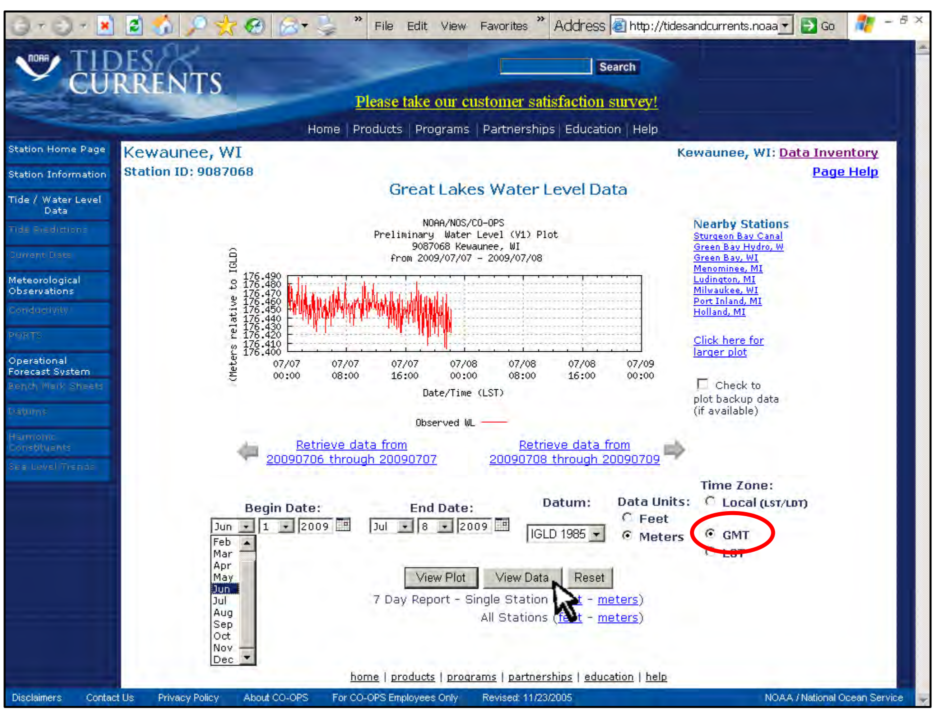
⚠ **Note:** If you are building your model *mid-season*, there will be a gap of approximately one month between the most up-to-date, *verified* data and today’s real-time data. You can cover this gap by downloading *preliminary* data, as described below.

B.7.17. For real-time and recently-collected data: Return to the map window (7.B.3). If it is not still open, locate the red marker for your station again and click on it to re-open the station pop-up window.

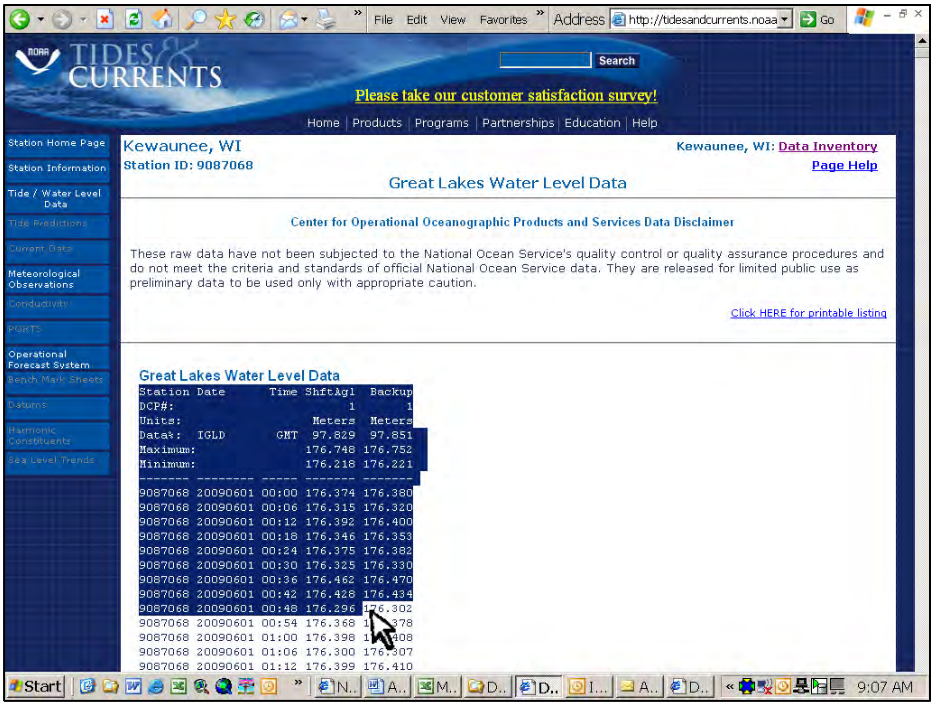
B.7.18. Click on the “Preliminary Data” link in the pop-up window to open the “Preliminary Data” page.



B.7.19. Using the pull-down menus, choose the “Begin Date” to correspond with the latest available verified data (see B.7.7), while leaving the “End Date” at today’s date (the default). Pick your desired “Data Units”, set “Datum” to “IGLD 1985,” and set “Time Zone” to “GTM” (effectively the same as UTC). Click on “View Data”.



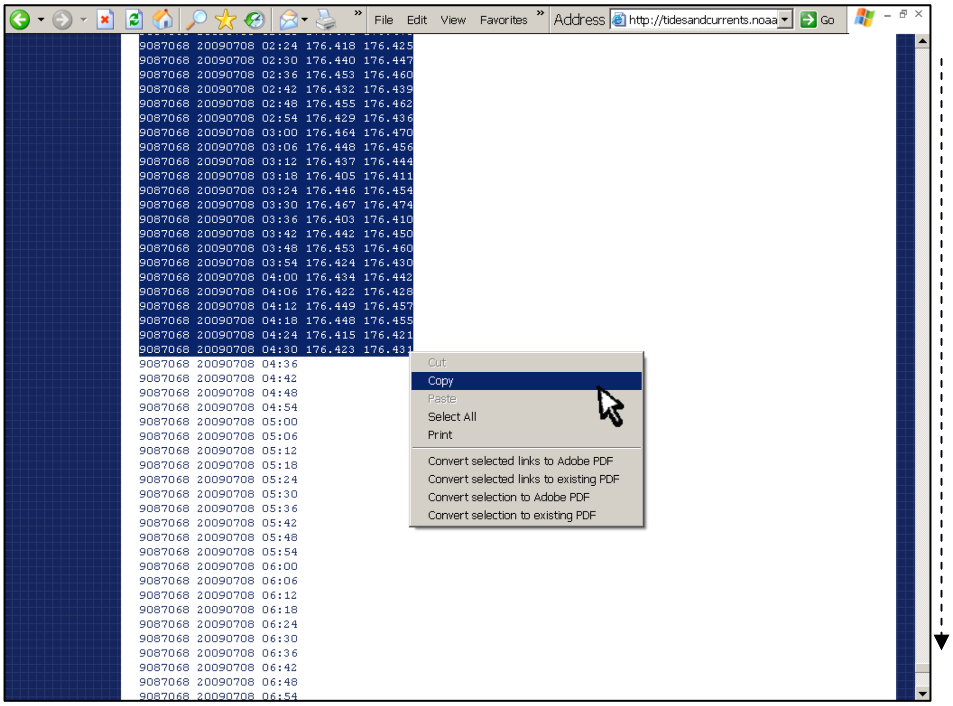
B.7.20. A table will open, with lake level values in the *two* right columns. Click just to the left of the word “Station”, hold and drag the mouse down and to the right, until two or more rows are highlighted.



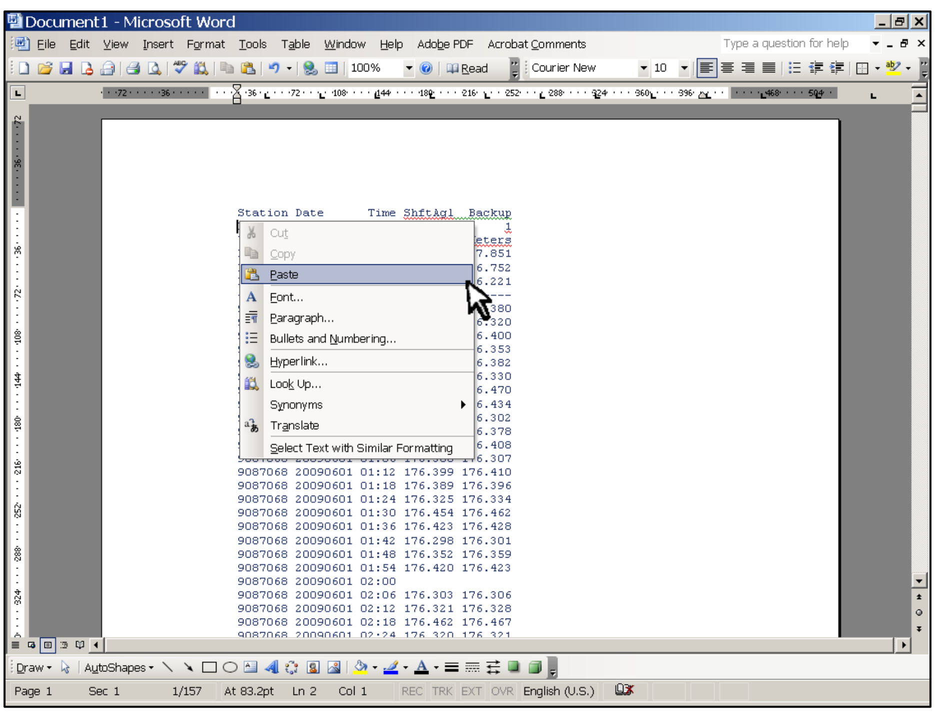
Lake level is typically measured by two instruments at each station: a principal sensor and a back-up. The back-up serves as a redundancy in the event that the principal sensor fails or otherwise reports an anomalous reading. The preferred instrument for measuring lake levels is known as a shaft angle encoder – listed in preliminary data tables as “ShftAgl” (see B.7.20). Typically, the back-up sensor is also a shaft angle encoder, and the readings are not reported. In some cases, however, the back-up sensor is what is known as a pressure sensor – an older and less accurate technology. In these stations (as in the example illustrated here) pressure sensor data are listed in addition to the preferred ShftAgl data, under the heading “Backup” (see B.7.20).

Note: Do *not* use data listed under the heading “Backup”.

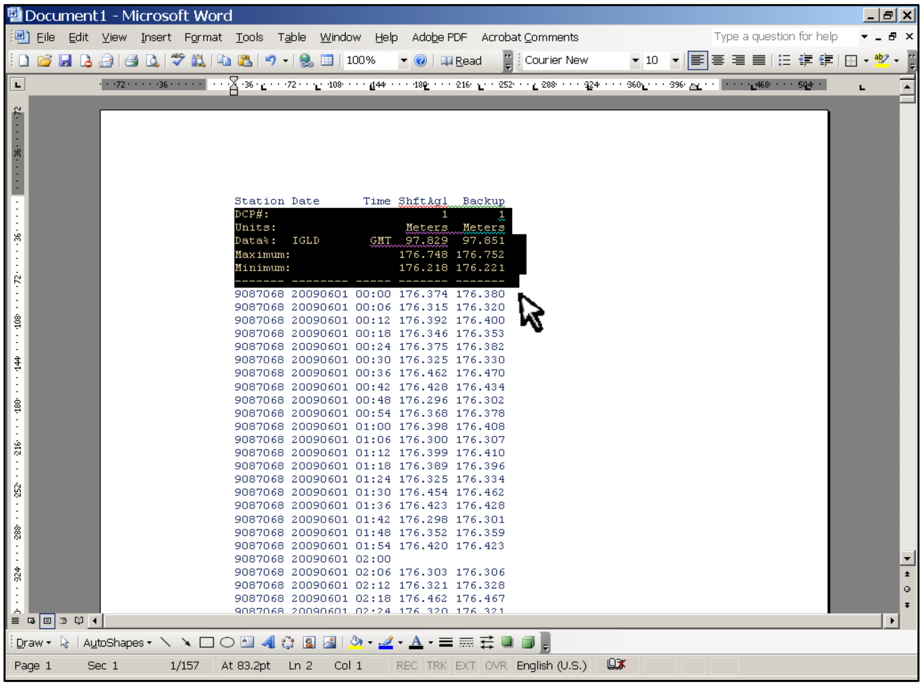
B.7.21. Scroll to the last row containing lake level values, which is the most recent (i.e. real-time) data. Hold the “Shift” key, and click to the right of the row to highlight. Right-click and select “Copy”.



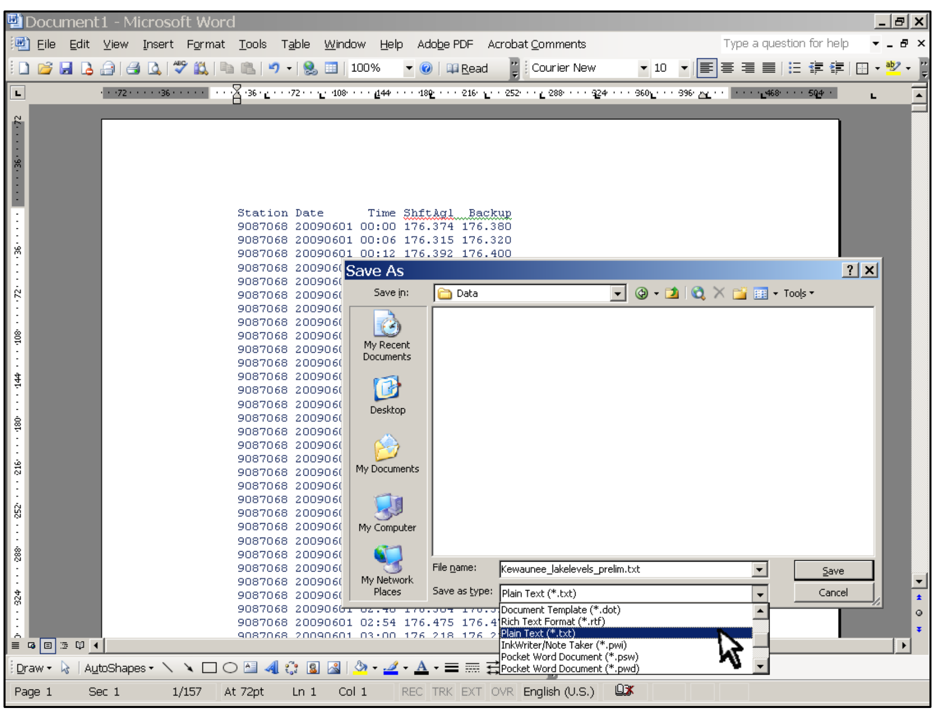
B.7.22. Open MS Word. Right-click at the top of a blank document and select “Paste”.



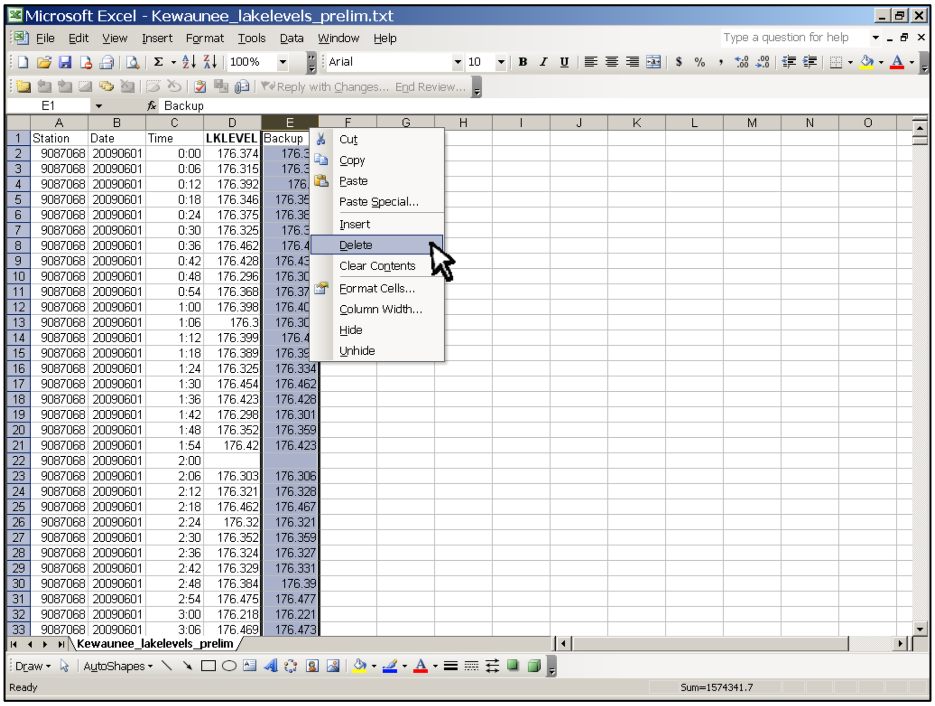
B.7.23. Click just to the left of "DCP#", hold and drag the mouse downward and to the right, until rows two through seven (-------) are highlighted. Hit the "Delete" key.



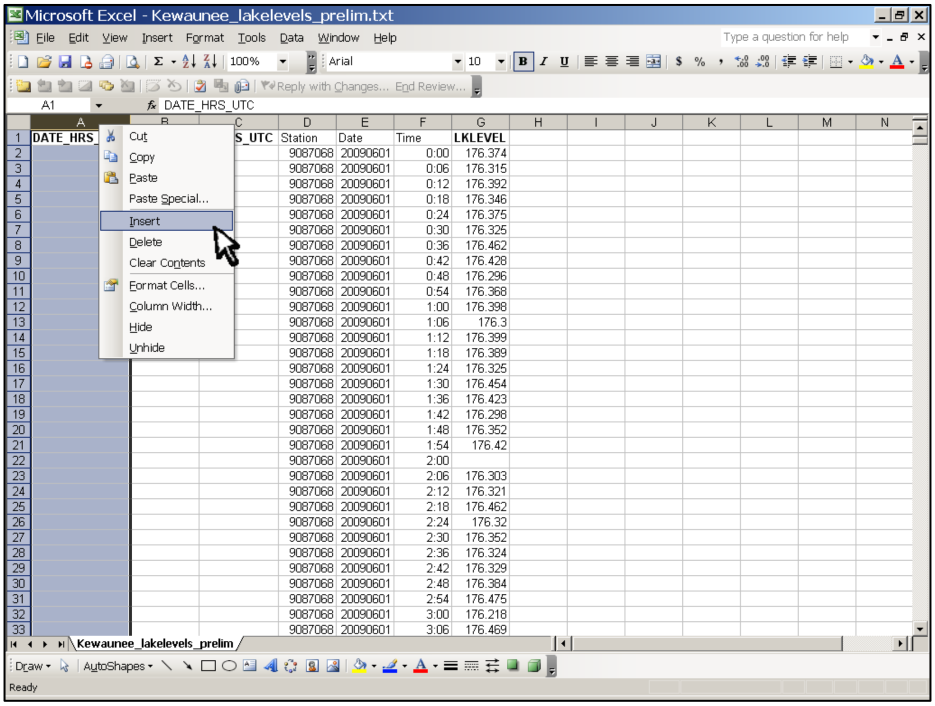
B.7.24. From the MS Word pull-down menu, select "File" > "Save as…", name the file something like "[location]_lakelevels_prelim", and set the file type to "Plain Text" (*.txt). (In the "File Conversion" window that opens, select "Windows (Default)" as the text encoding system and click the "OK" button.)



B.7.25. Following the procedures in steps B.6.9-6.10, open the file in MS Excel. Rename cell D1 something like "LKLEVEL". Delete column E ("Backup"), by right-clicking on the column header and selecting "Delete".



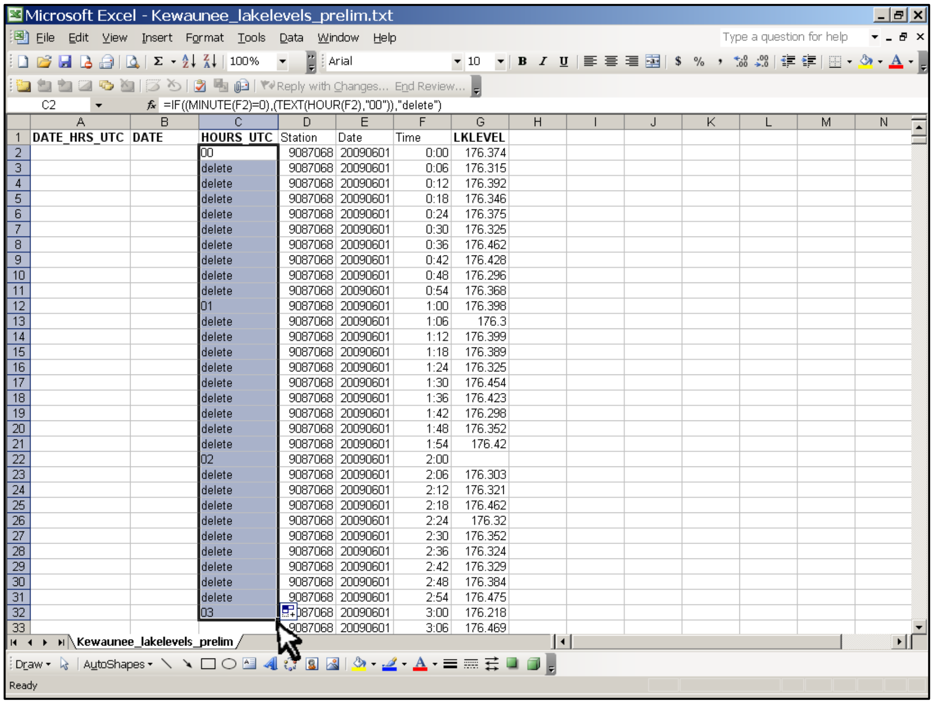
B.7.26. Following the procedures in A.1.7, insert new columns (A-C) named "DATE_HRS_UTC", "DATE", and "HOURS_UTC".



B.7.27. Preliminary lake level data are reported in 6-minute increments, as opposed to one-hour increments. As such, there will be 10 different entries per hour. To flag redundant entries and create values for “HOURS_UTC” (column C) type the following function in cell C2:

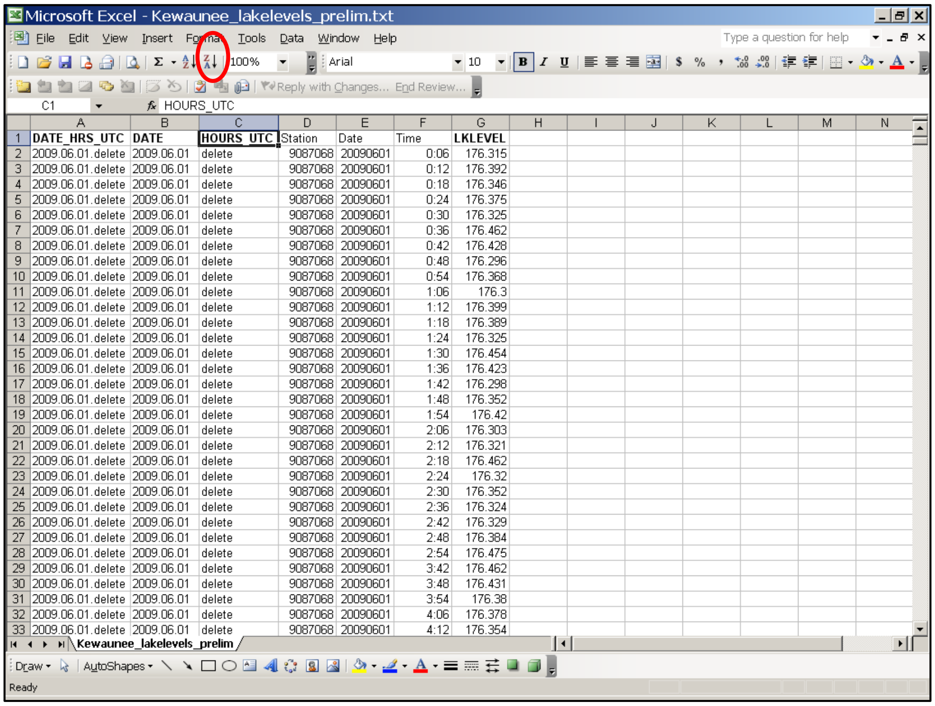
=IF((MINUTE(F2)=0),(TEXT(HOUR(F2),"00")),"delete")

Hit “Enter”. The result should be a two-digit value for the hour-of-day, in UTC (Central Daylight Time + 5 hours.). To copy, click on the lower right hand corner of the cell, hold and drag the mouse to the bottom of the table.

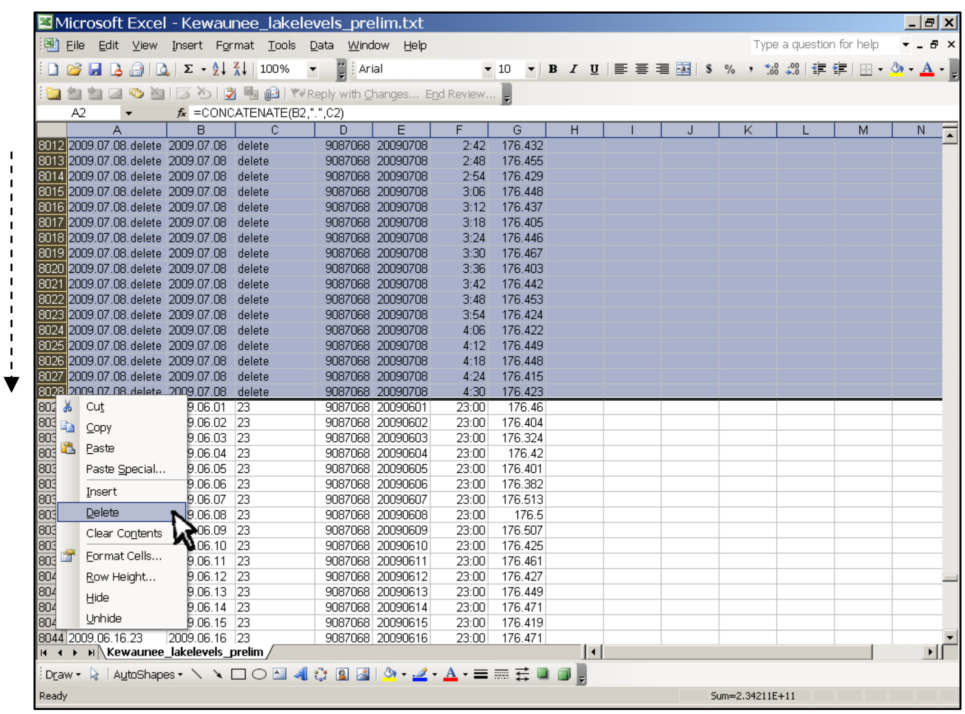


B.7.28. Follow the procedures and functions listed in steps B.7.14 – B.7.15 to calculate “DATE” (column B) and “DATE_HRS_UTC” (column A).

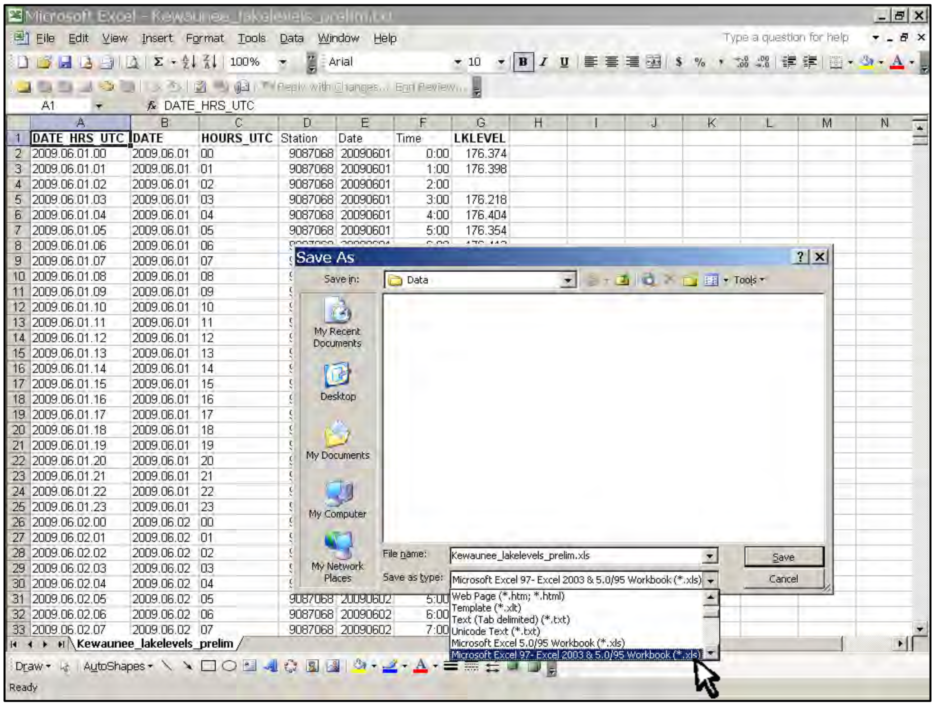
B.7.29. To remove redundant observations, first click on the first cell in column C (“HOURS_UTC”) and click on reverse sort button ( ). The redundant observations will be at the top of the table.



B.7.30. Next, click to the left of row 2 and, holding the “Shift” key, push the down-arrow on the keyboard until all of the redundant observations are selected (i.e. column C values = “delete”). Right-click on the last-selected row and select “Delete”.



B.7.31. Save the table as an MS Excel file by selecting "File" > "Save as…" from the MS Excel pull-down menu, choosing the most recent version of MS Excel as the file type (*.xls), and clicking the "Save" button.



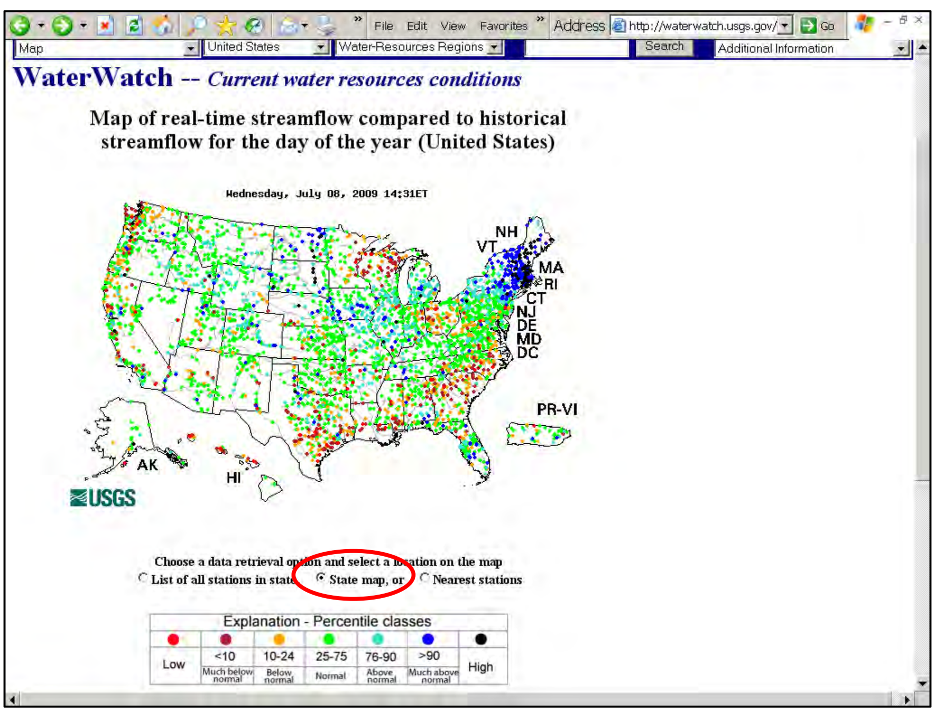
Notes

C. Real-Time and Historic Watershed Conditions

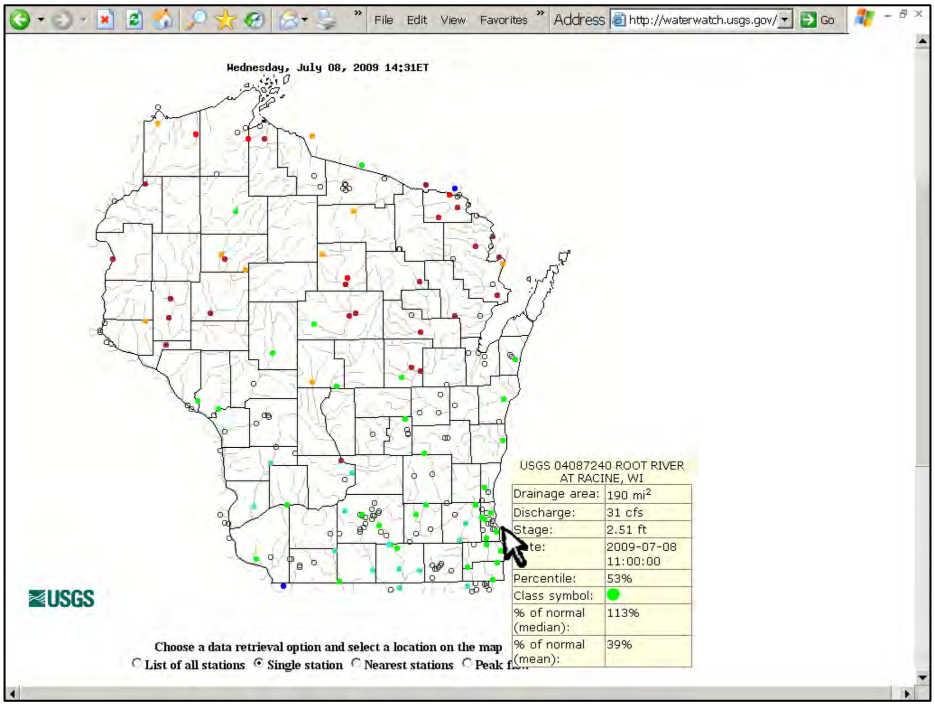
C.1. Waterwatch

Provides access through a map interface to near real-time, as well as archived/historic data from USGS stream gauges, which measure and report runoff and in some cases various surface water quality parameters as well. Data from stream gauges reflect not only rainfall, but watershed conditions, such as soil moisture, land cover imperviousness, and non-point source pollution, all of which can affect beach water quality. Where a gauged-stream outlets at or near your beach, stream discharge (stream flow) may be a useful explanatory variable for your model. Depending on the gauge, pertinent stream water quality parameters may also be available, such as specific conductance (related to suspended solids) and dissolved oxygen (related to concentrations of organic matter).

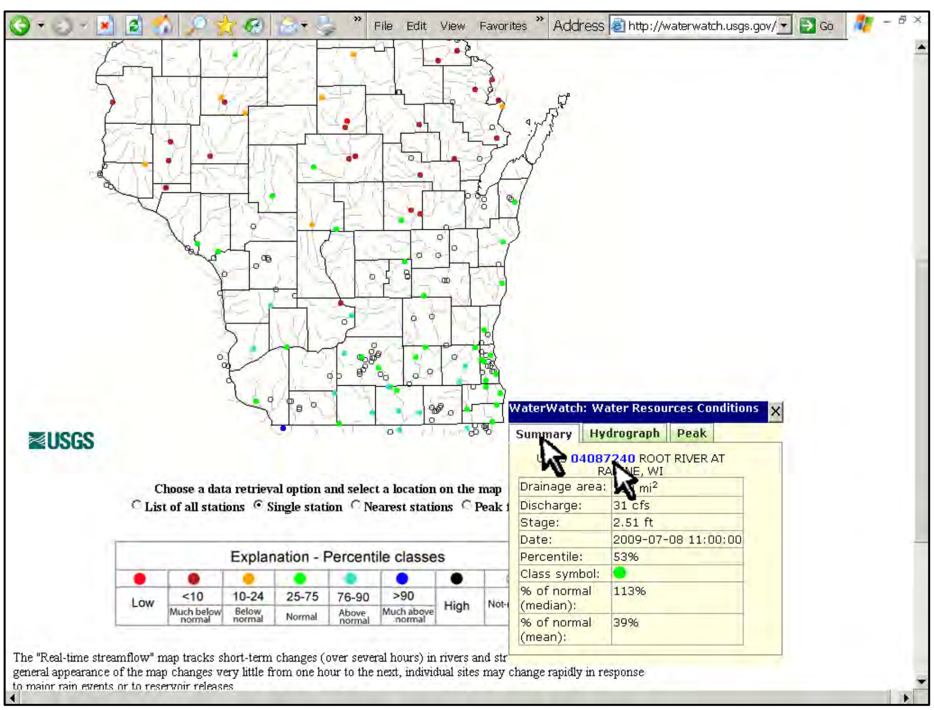
C.1.1. Open the **Waterwatch** page (http://water.usgs.gov/waterwatch). A map of the U.S. will appear, with various points representing stream gauges that are currently operated by USGS and its partners. Gauges (points) are colored-coded according to the most recent data on stream water discharge, ranging from red (Low) to black (High). Check the circle for "State map…" above the map key, as your desired means of data retrieval. Within the national map, click on your state (or neighboring state) of interest to launch a zoomed-in map, from which data can be accessed.



C.1.2. On the statewide map, move the mouse arrow over any gauge (point) to display a temporary window showing gauge name, current data on discharge and stage, and drainage area (i.e. size of the watershed or catchment discharging at that point).



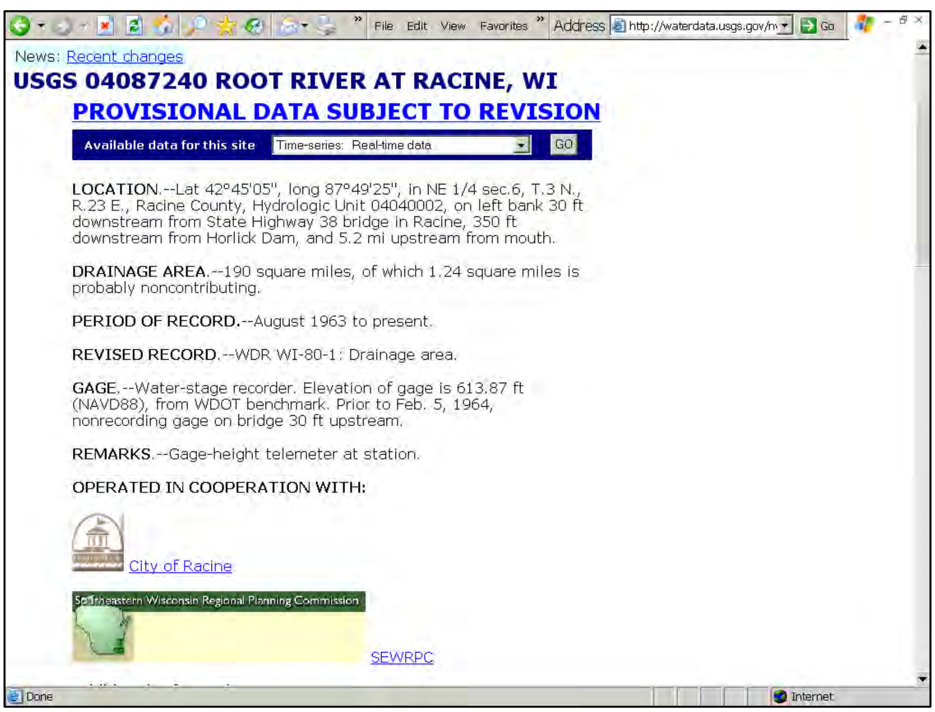
C.1.3. Click on your gauge of interest to open a data access window. In the first tab ("Summary"), click on the 8-digit gauge number (in **blue**) to open the main data access page for that gauge.



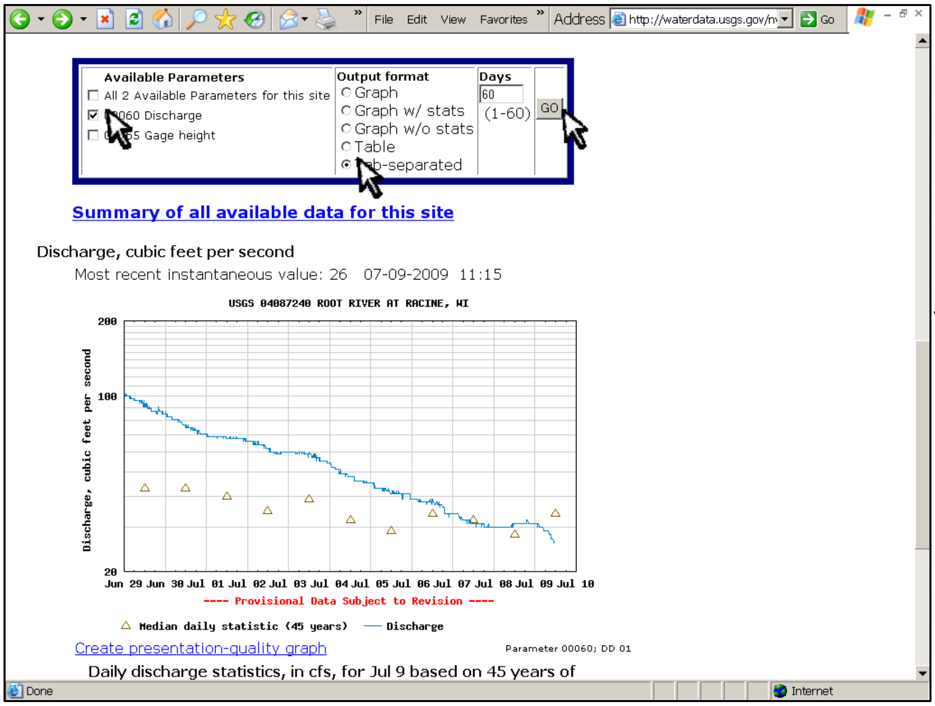
C.1.4. Under “AVAILABLE DATA”, click on “Real-time”.

Alternative “Daily Data” are comprised of 24-hour maxima, minima, and averages (for the period between midnight and midnight, local time.)

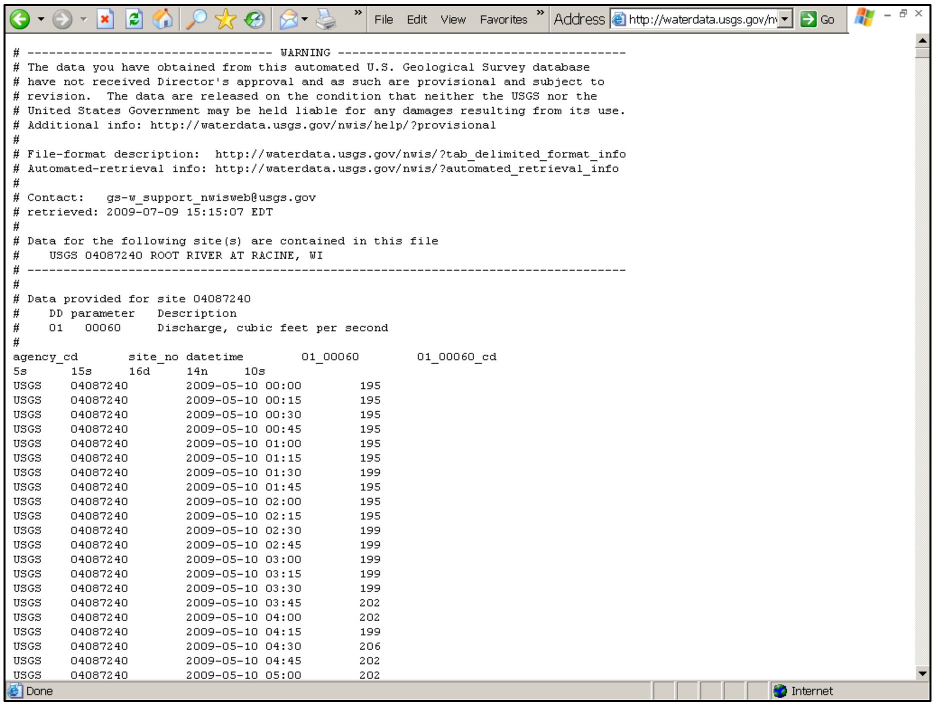
C.1.5. View the summary information that opens for the gauge.



C.1.6. Scroll down the page (to just below the purple box listing "Available Parameters") and note the current "Discharge, cubic feet per second" (reported in local time).



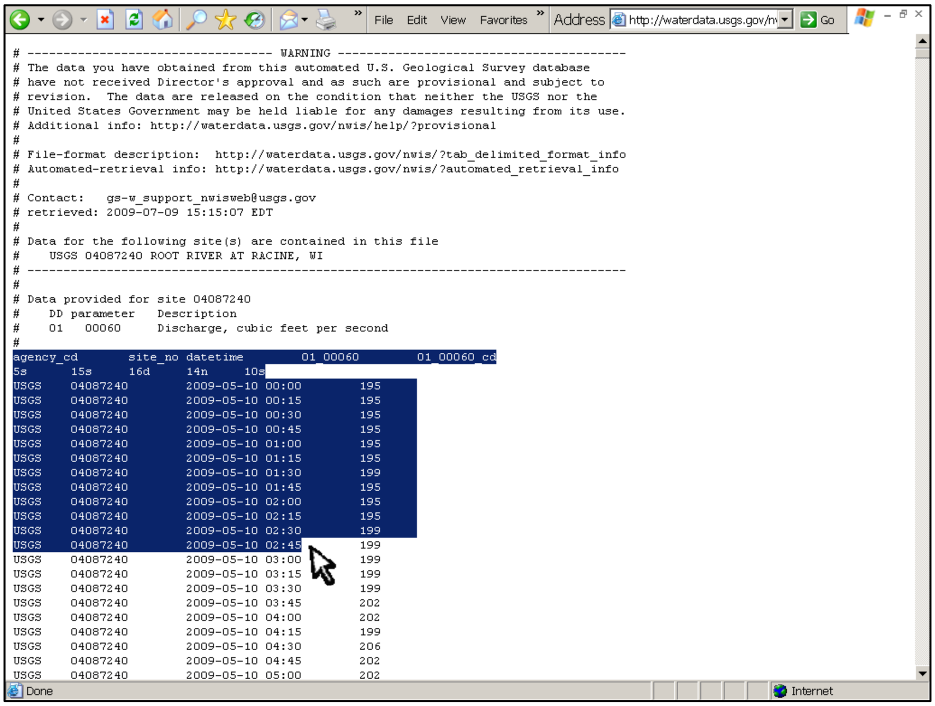
C.1.7. For *recently-archived* data: Under "Available Parameters" in the purple box (see above) check the box for "Discharge" (and any other variable of interest). Check "Tab-separated" under "Output format". Type the number of "Days" you want to go back (0-60), and click "Go". This will open an archive of instantaneous data.



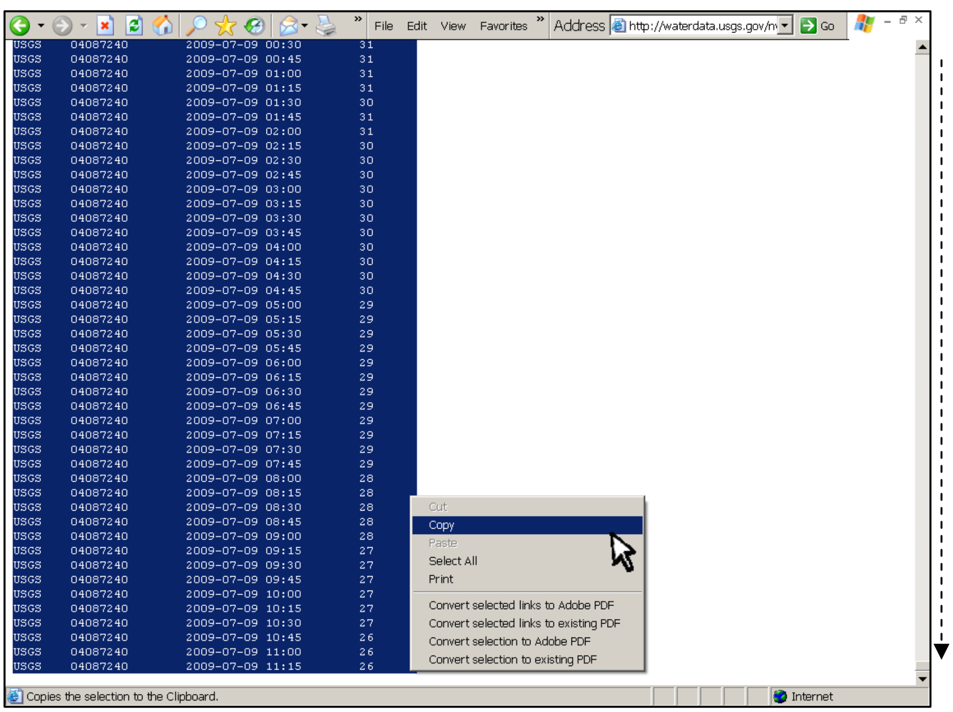
```
# ---------------------------------- WARNING ----------------------------------------
# The data you have obtained from this automated U.S. Geological Survey database
# have not received Director's approval and as such are provisional and subject to
# revision.  The data are released on the condition that neither the USGS nor the
# United States Government may be held liable for any damages resulting from its use.
# Additional info: http://waterdata.usgs.gov/nwis/help/?provisional
#
# File-format description:  http://waterdata.usgs.gov/nwis/?tab_delimited_format_info
# Automated-retrieval info: http://waterdata.usgs.gov/nwis/?automated_retrieval_info
#
# Contact:   gs-w_support_nwisweb@usgs.gov
# retrieved: 2009-07-09 15:15:07 EDT
#
# Data for the following site(s) are contained in this file
#    USGS 04087240 ROOT RIVER AT RACINE, WI
# -----------------------------------------------------------------------------------
#
# Data provided for site 04087240
#    DD parameter   Description
#    01   00060     Discharge, cubic feet per second
#
agency_cd       site_no datetime        01_00060        01_00060_cd
5s      15s     16d     14n     10s
USGS    04087240        2009-05-10 00:00        195
USGS    04087240        2009-05-10 00:15        195
USGS    04087240        2009-05-10 00:30        195
USGS    04087240        2009-05-10 00:45        195
USGS    04087240        2009-05-10 01:00        195
USGS    04087240        2009-05-10 01:15        195
USGS    04087240        2009-05-10 01:30        199
USGS    04087240        2009-05-10 01:45        195
USGS    04087240        2009-05-10 02:00        195
USGS    04087240        2009-05-10 02:15        195
USGS    04087240        2009-05-10 02:30        199
USGS    04087240        2009-05-10 02:45        199
USGS    04087240        2009-05-10 03:00        199
USGS    04087240        2009-05-10 03:15        199
USGS    04087240        2009-05-10 03:30        199
USGS    04087240        2009-05-10 03:45        202
USGS    04087240        2009-05-10 04:00        202
USGS    04087240        2009-05-10 04:15        199
USGS    04087240        2009-05-10 04:30        206
USGS    04087240        2009-05-10 04:45        202
USGS    04087240        2009-05-10 05:00        202
```

For data going back further than 60 days, see s C.1.24 - C.40.

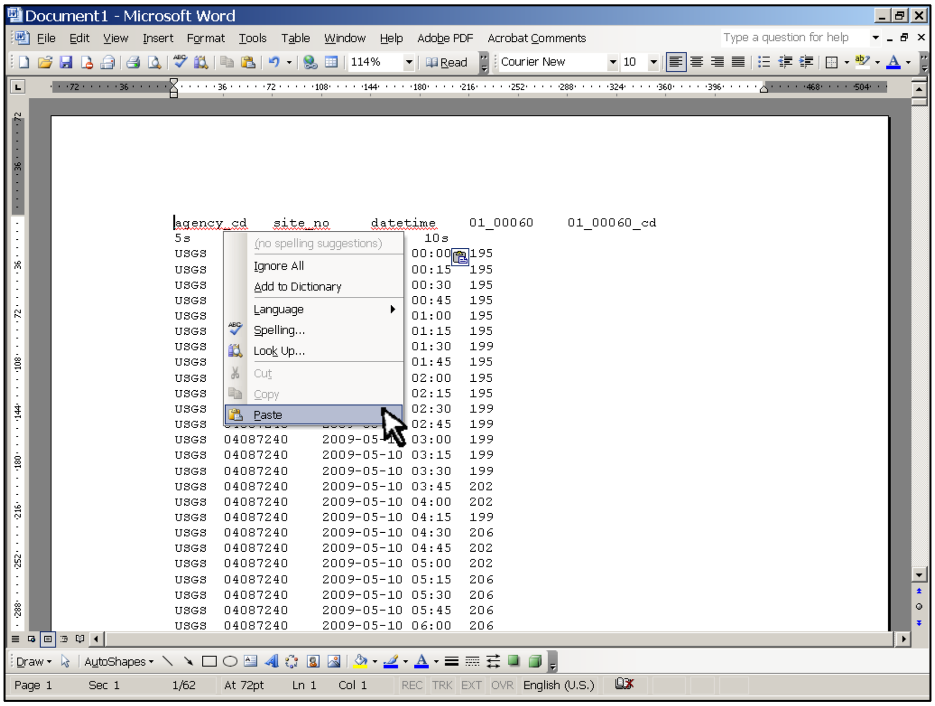
C.1.8. In the table of recent data, click just to the left of the word “agency”, then hold and drag the mouse down and to the right, until two or more rows are highlighted.



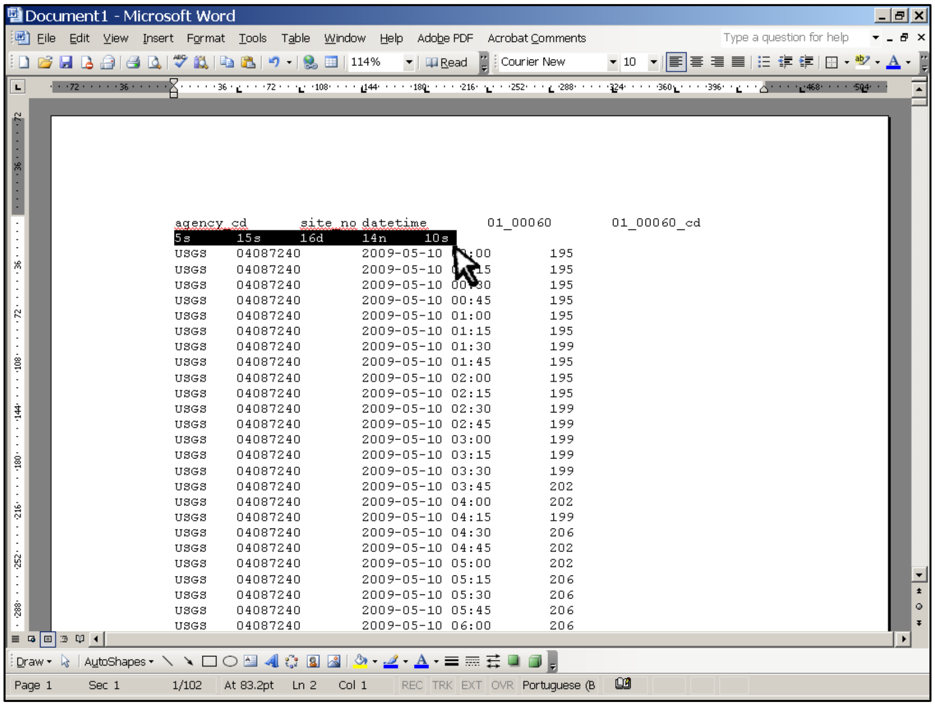
C.1.9. Scroll to the last row. Hold the “Shift” key, and click to the right of the row to highlight. Right-click and select “Copy”.



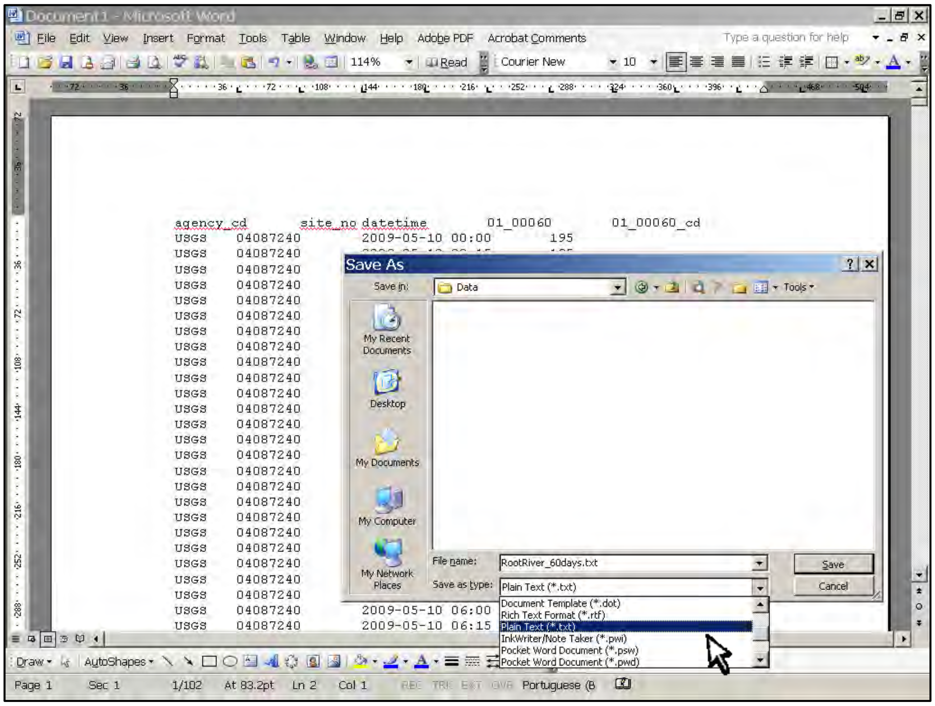
C.1.10. Open MS Word. Right-click at the top of a blank document and select "Paste".



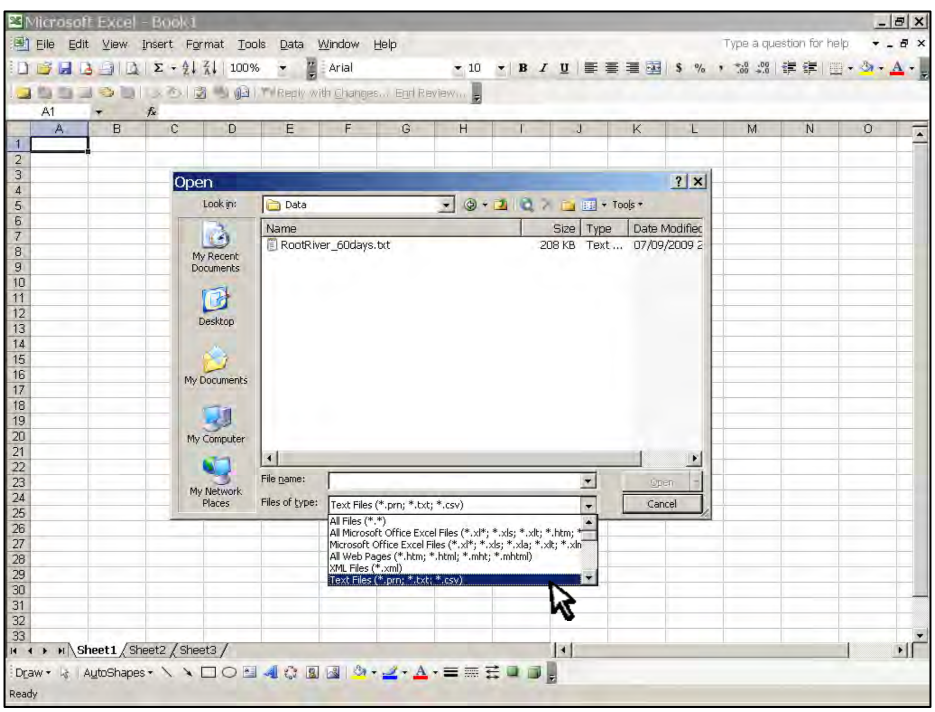
C.1.11.Click just to the left of the second row, hold the "Shift" key, and hit the down arrow key to highlight just that row. Click the "Delete" key.



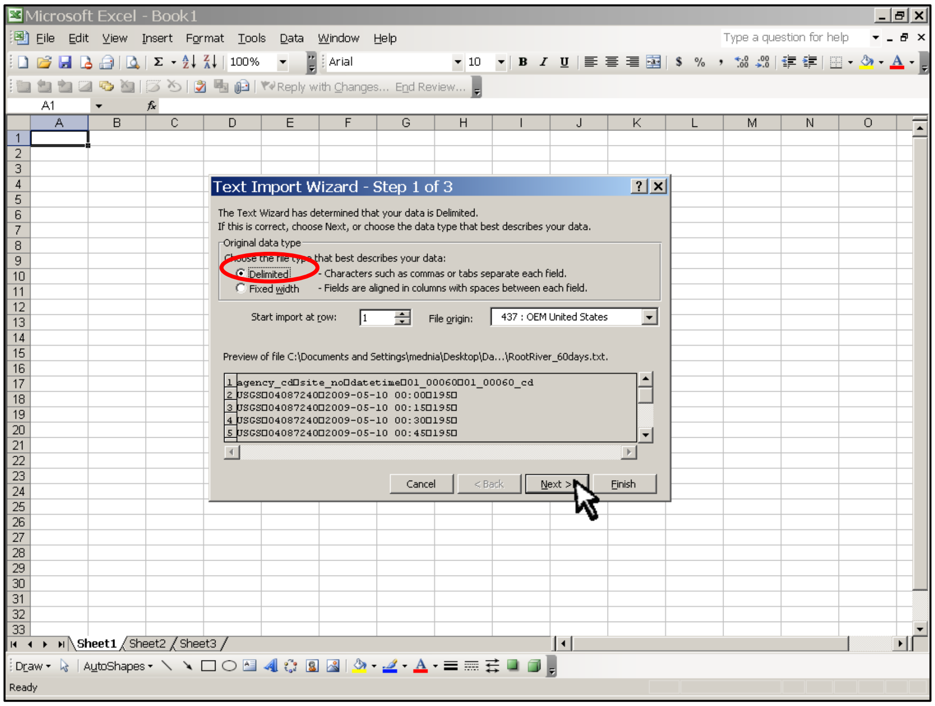
C.1.12. Following the procedure in step B.7.24, use the MS Word pull-down menu ("File" > "Save as...") save the table as a "Plain Text" (*.txt) file named something like "[river name]_60days".



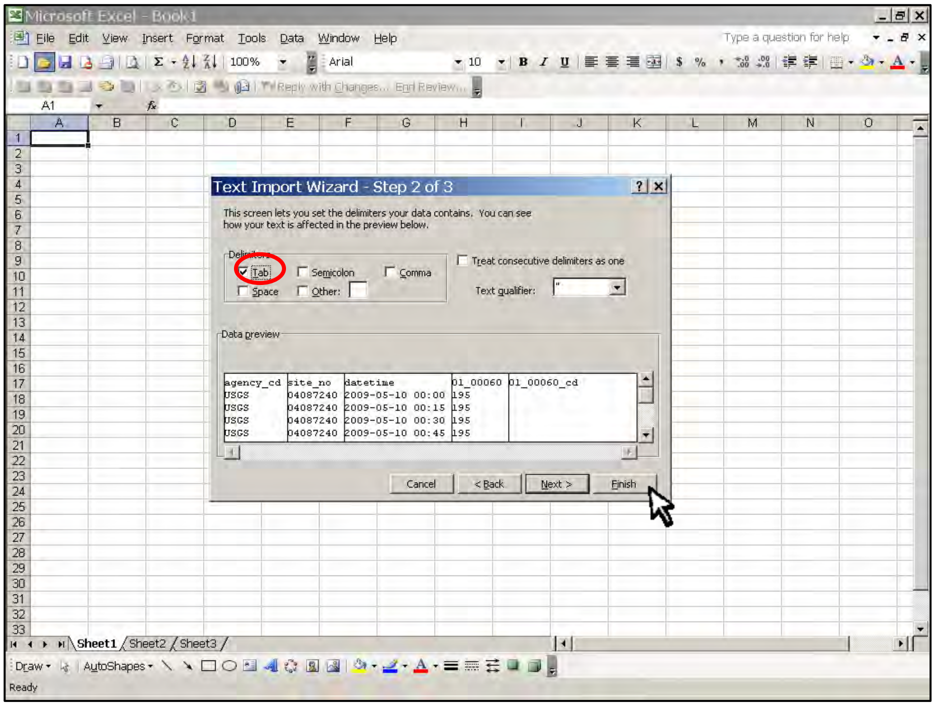
C.1.13. Open MS Excel. From the MS Excel pull-down menu in the upper left corner, select "File" > "Open". In the "Open" window, select "Text Files" from the "Files of type:" pull-down menu. Double-click on the text file you created in C.1.12.



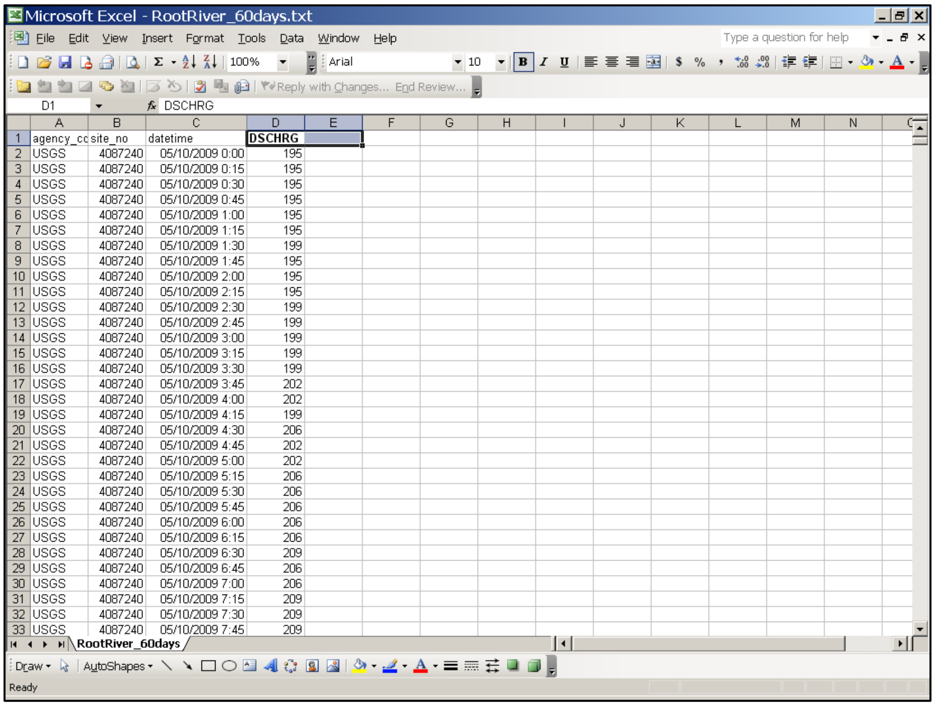
C.1.14. A “Text Import Wizard” will open. In Step 1, check the circle for “Delimited” and click the “Next>” button.



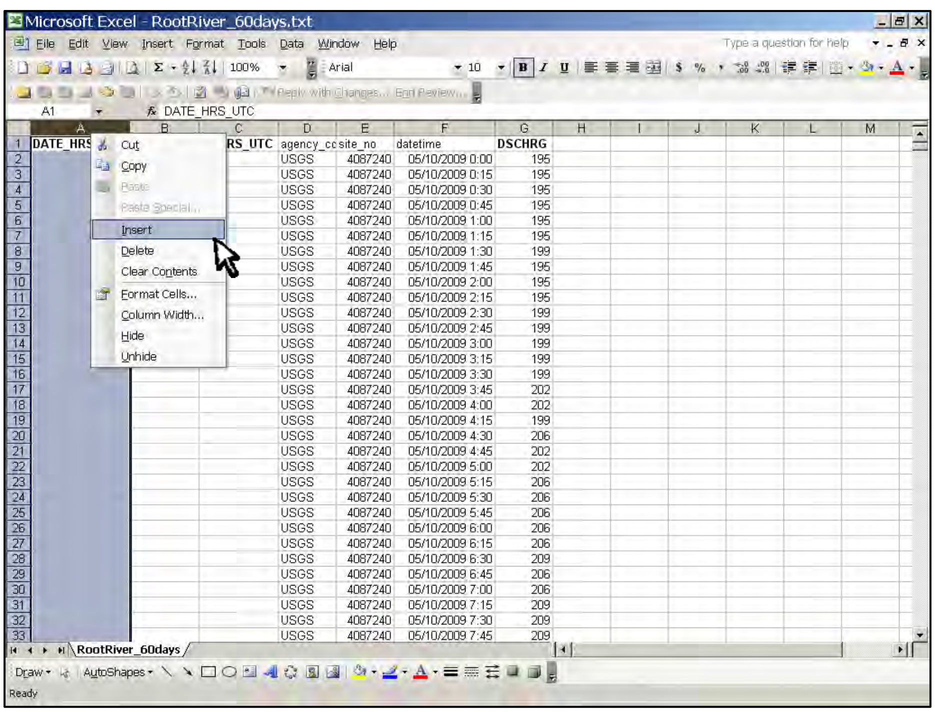
C.1.15. In Step 2 of the “Text Import Wizard”, check the box for “Tab” and click “Finish”.



C.1.16. After the table appears, rename cell D1 something like “DSCHRG” and delete the text in cell E1.



C.1.17. Following the procedure in A.1.7, insert three new columns (A-C) named “DATE_HRS_UTC”, “DATE”, and “HOURS_UTC”.



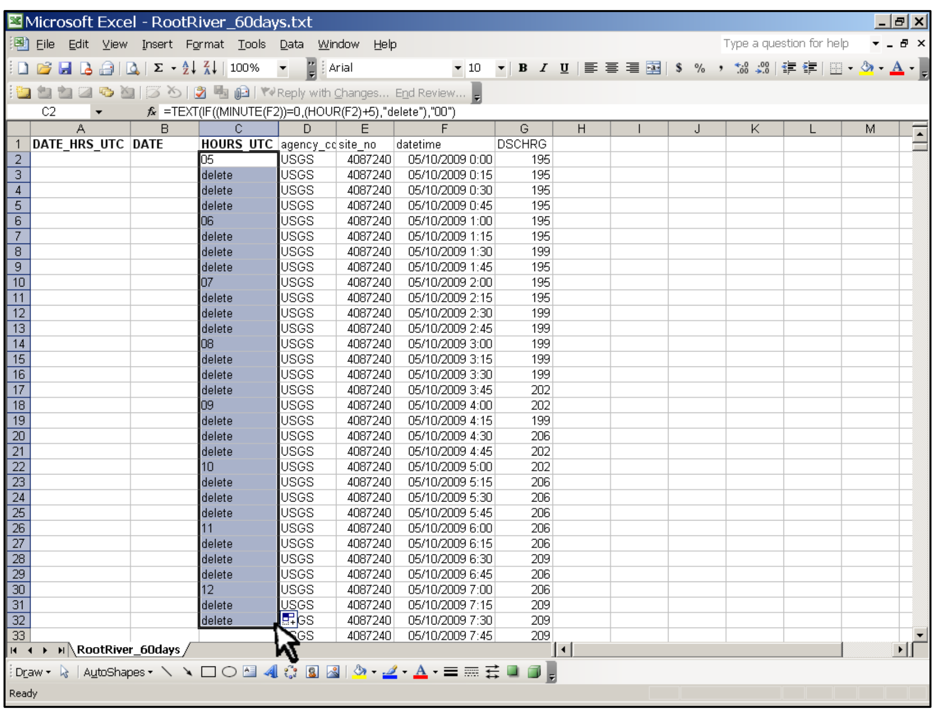
Note: As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to "join" the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and "joining" tables using queries in MS Access, see Section 6 of this report, "Formatting Data for Model-Building.")

USGS stream gauge data are reported in *local* time (UTC minus 5 hours in the central time zone).

C.1.18. USGS stream gauge data are reported in sub-hourly increments. As such, there will be several entries per hour. To flag redundant entries and create values for "HOURS_UTC" (column C) type the following function in cell C2:

=TEXT(IF((MINUTE(F2))=0,(HOUR(F2)+5),"delete"),"00")

Hit "Enter". The result should be a two-digit value for the hour-of-day, in UTC (Central Daylight Time + 5 hours.). To copy, click on the lower right hand corner of the cell, hold and drag the mouse to the bottom of the table.



	A	B	C	D	E	F	G
1	DATE_HRS_UTC	DATE	HOURS_UTC	agency_cc	site_no	datetime	DSCHRG
2			05	USGS	4087240	05/10/2009 0:00	195
3			delete	USGS	4087240	05/10/2009 0:15	195
4			delete	USGS	4087240	05/10/2009 0:30	195
5			delete	USGS	4087240	05/10/2009 0:45	195
6			06	USGS	4087240	05/10/2009 1:00	195
7			delete	USGS	4087240	05/10/2009 1:15	195
8			delete	USGS	4087240	05/10/2009 1:30	199
9			delete	USGS	4087240	05/10/2009 1:45	195
10			07	USGS	4087240	05/10/2009 2:00	195
11			delete	USGS	4087240	05/10/2009 2:15	195
12			delete	USGS	4087240	05/10/2009 2:30	199
13			delete	USGS	4087240	05/10/2009 2:45	199
14			08	USGS	4087240	05/10/2009 3:00	199
15			delete	USGS	4087240	05/10/2009 3:15	199
16			delete	USGS	4087240	05/10/2009 3:30	199
17			delete	USGS	4087240	05/10/2009 3:45	202
18			09	USGS	4087240	05/10/2009 4:00	202
19			delete	USGS	4087240	05/10/2009 4:15	199
20			delete	USGS	4087240	05/10/2009 4:30	206
21			delete	USGS	4087240	05/10/2009 4:45	202
22			10	USGS	4087240	05/10/2009 5:00	202
23			delete	USGS	4087240	05/10/2009 5:15	206
24			delete	USGS	4087240	05/10/2009 5:30	206
25			delete	USGS	4087240	05/10/2009 5:45	206
26			11	USGS	4087240	05/10/2009 6:00	206
27			delete	USGS	4087240	05/10/2009 6:15	206
28			delete	USGS	4087240	05/10/2009 6:30	209
29			delete	USGS	4087240	05/10/2009 6:45	206
30			12	USGS	4087240	05/10/2009 7:00	206
31			delete	USGS	4087240	05/10/2009 7:15	209
32			delete	GS	4087240	05/10/2009 7:30	209
33				GS	4087240	05/10/2009 7:45	209

C.1.19. To calculate “DATE” (column B), click on cell B2 and type the following function:

=CONCATENATE((YEAR(F2)),".",(TEXT((MONTH(F2)),"00")),".",(DAY(F2)))

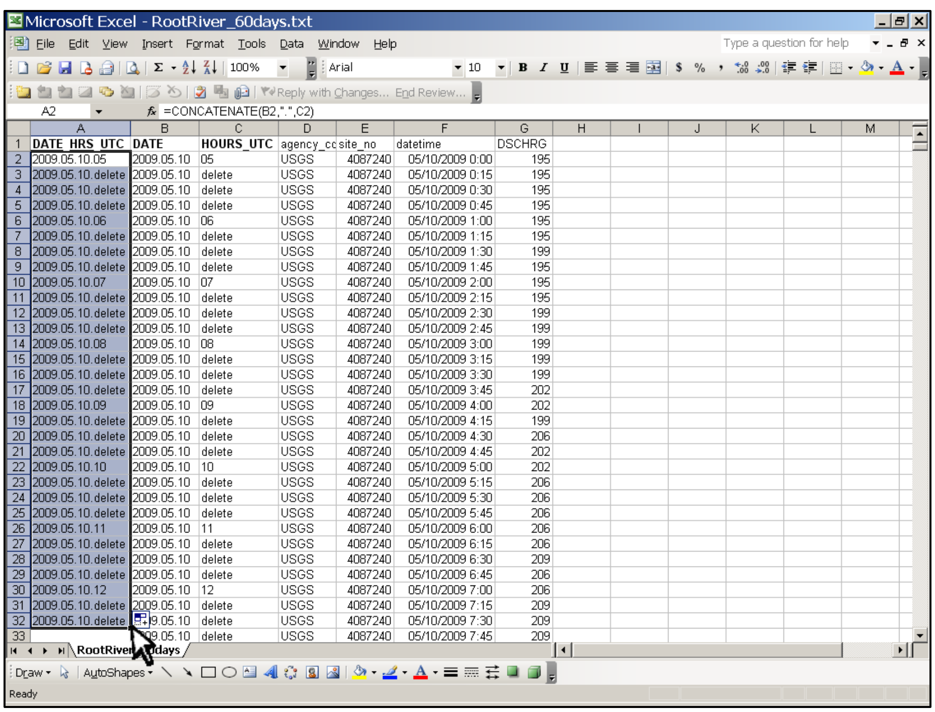
Hit “Enter”. This should return a date value (e.g., “2009.05.10” for May 10, 2009). Copy this function by clicking in the lower right corner of cell B2 and dragging to the end of the table.

Note: In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Be sure to include period-separators between the year, month, day, and hour (e.g., “2003.05.10.12” as opposed to “2003051012”). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

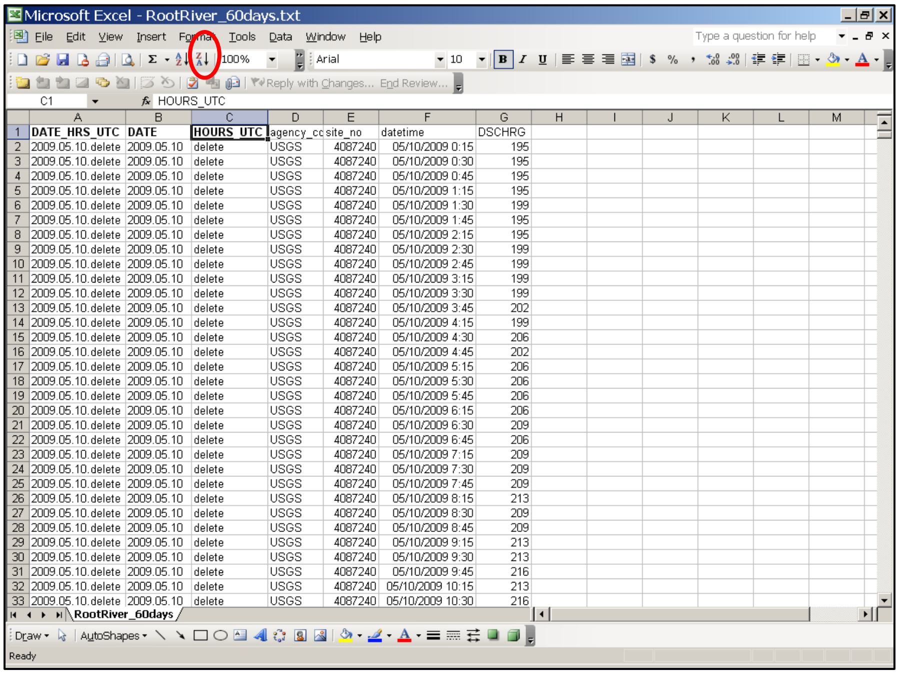
C.1.20. To calculate “DATE_HRS_UTC” (column A), click on cell A2 and type the following function:

=CONCATENATE(B2,".",C2)

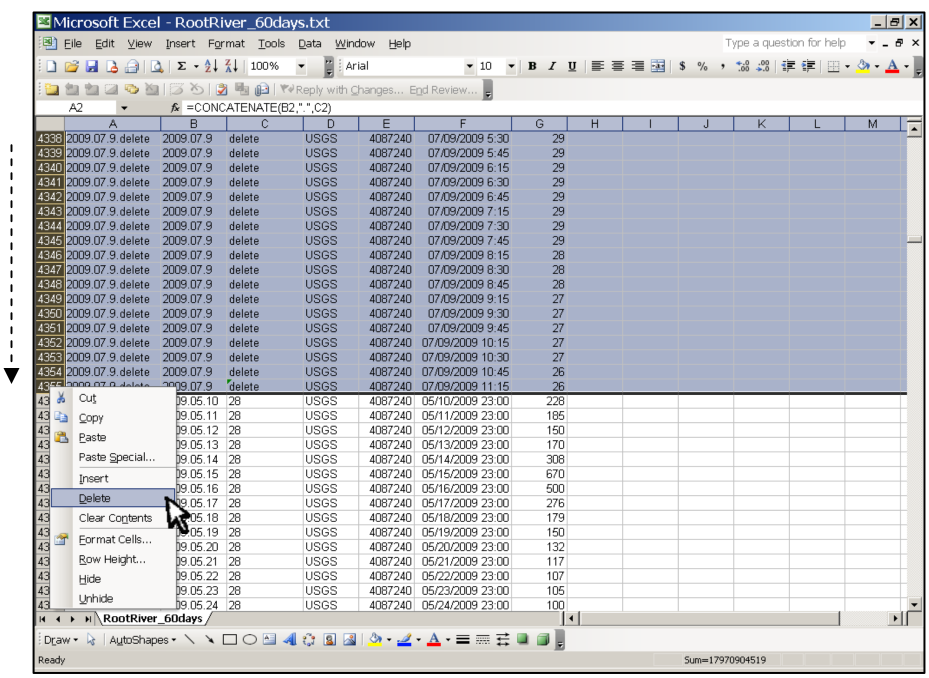
Hit “Enter”. This should return a date-plus-hour value (e.g., “2009.05.10.01” for 01:00 hours UTC on May 10, 2009). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.



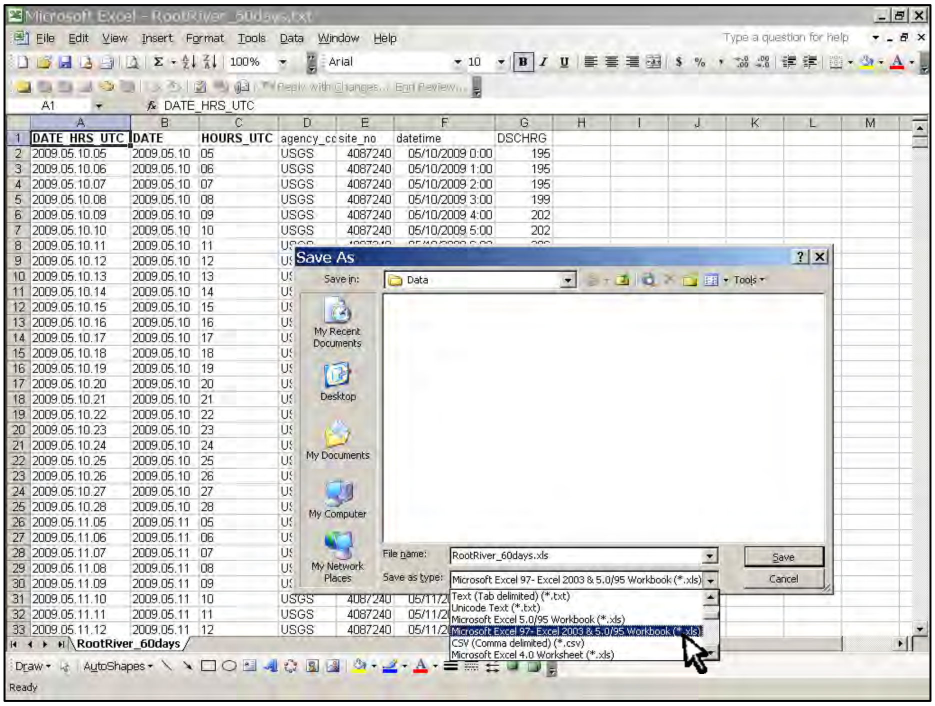
C.1.21. To remove the redundant hourly observations that you flagged in step C.1.18, first click on the first cell in column C and click on the reverse-sort button ( ). The redundant observations will be at the top of the table.



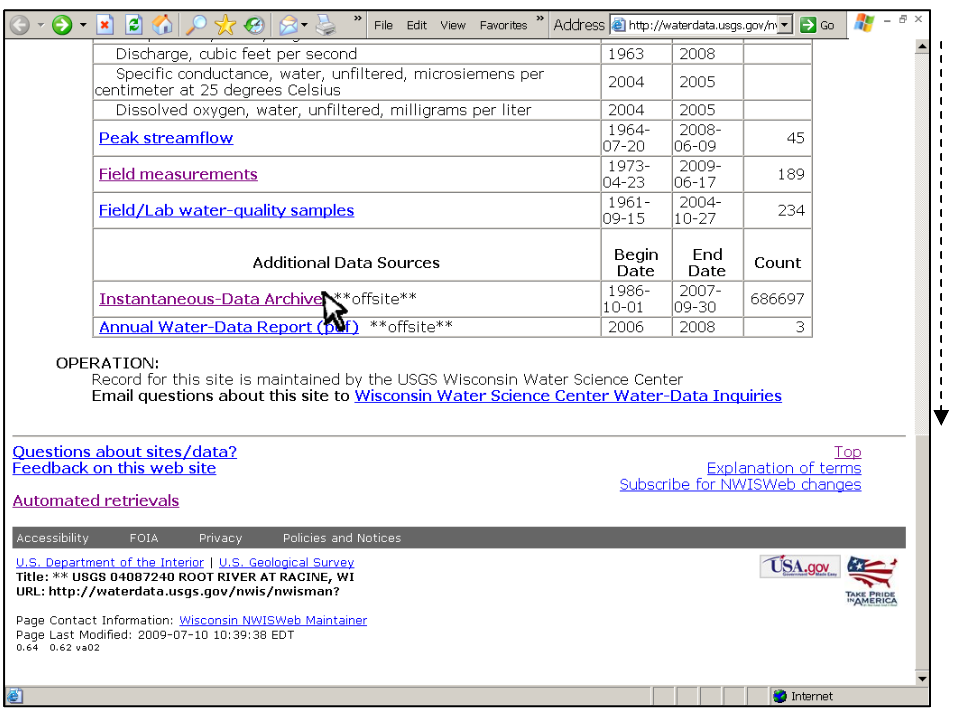
C.1.22. Next, click to the left of row 2 and, holding the “Shift” key, push the down-arrow on the keyboard until all of the redundant observations are selected (i.e. column C values = “delete”). Right-click on the last-selected row and select “Delete”.



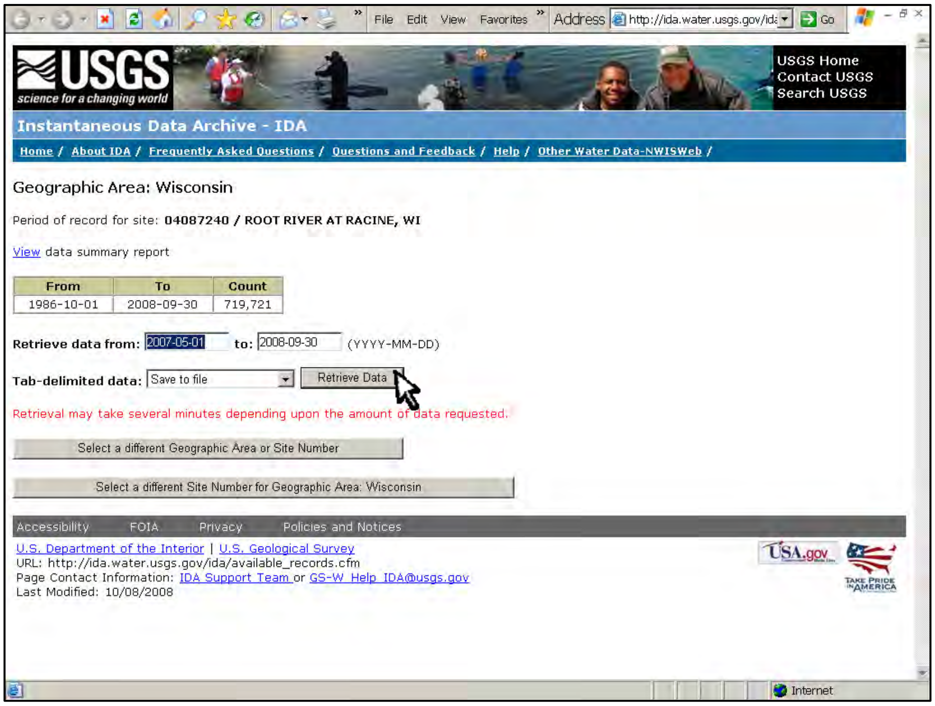
C.1.23. Save the table as an MS Excel file by selecting “File” > “Save as...” from the pull-down menu in the upper left corner, choosing the most recent version of MS Excel as the file type (*.xls), and clicking “Save”.



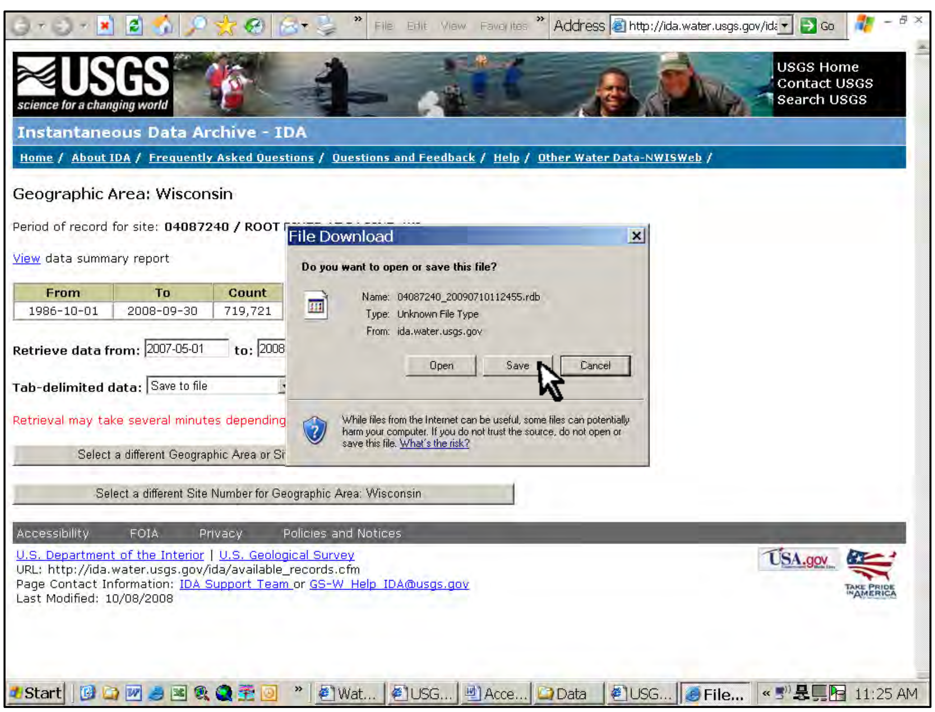
C.1.24. For historic/archived data older than 60 days: Return to the main data access page for your selected gauge (C.1.4), scroll to the bottom of the page and click the link to “Instantaneous-Data Archive”.



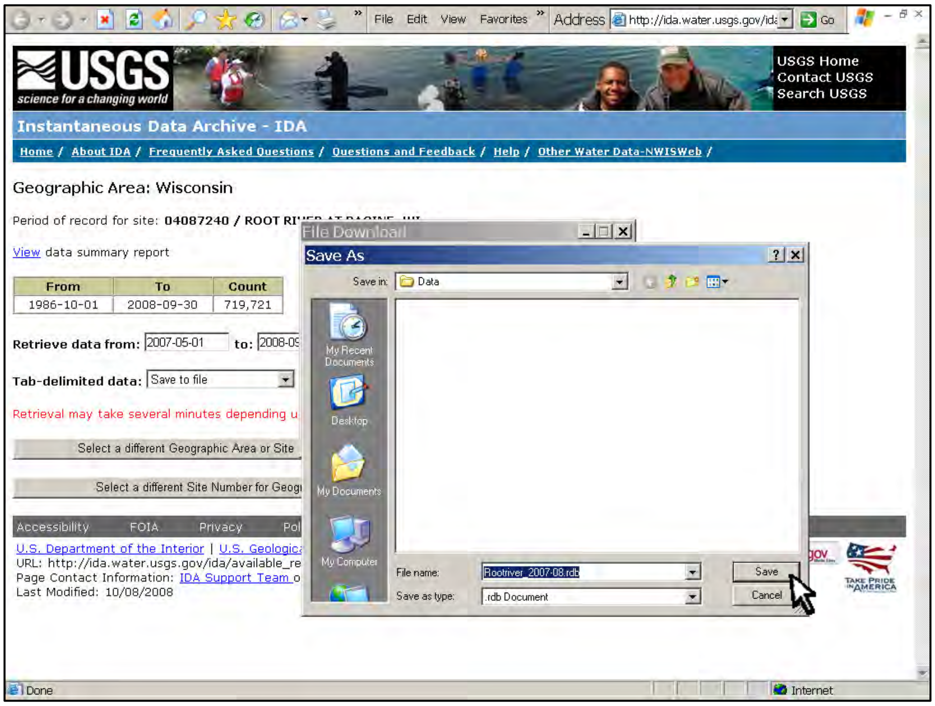
C.1.25 A separate “Instantaneous Data Archive” page will open. In the spaces next to “Retrieve data from:” and “To:” type the beginning and end dates you want data for. (The most recent data typically end in the fall of *the previous season*. If you are building your model *mid-season*, see step C.1.17 to access data going back 60 days.)



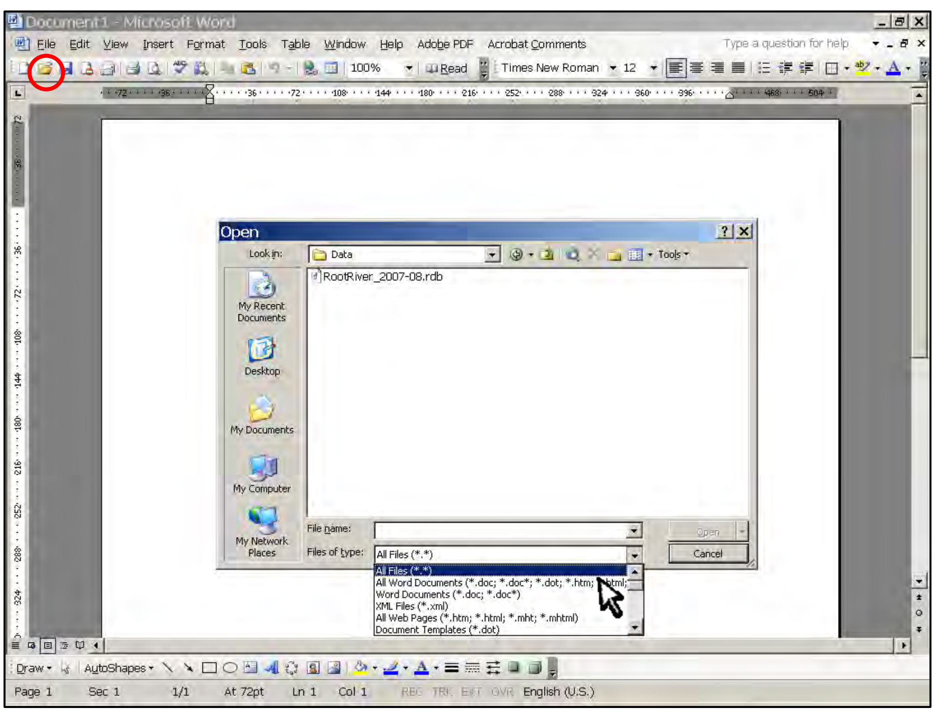
C.1.26. Click the “Retrieve Data” button, then “Save” (in the “File Download” window).



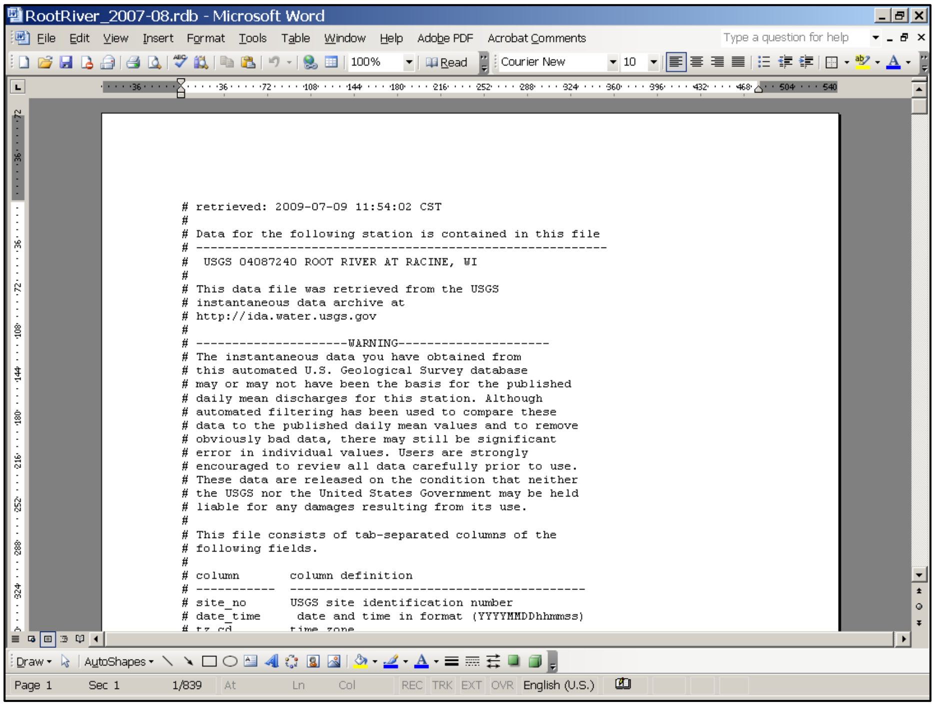
C.1.27. In the “Save As” window, navigate to the folder where you wish to save the data, name the file something like “[river name]_[year(s)]”, and click “Save”. The file type will – by default – be Retrospect backup data, or .rbd.



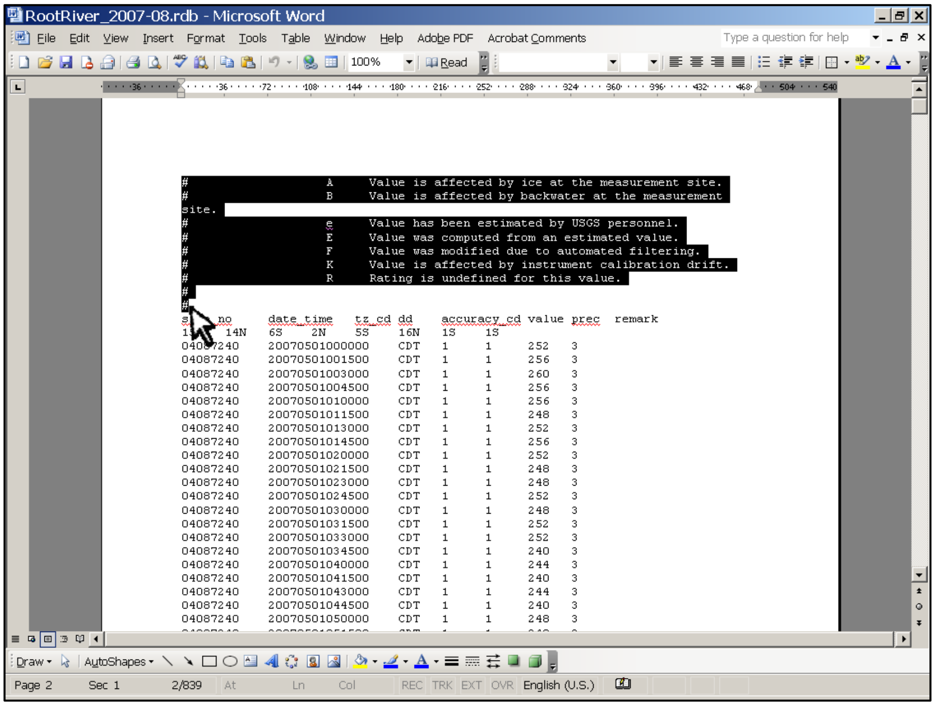
C.1.28. Open MS Word. Click the “Open File” ( ) button. In the “Open” window, select “All files” as the file type, and then double-click on the .rbd file you created in step C.1.27.



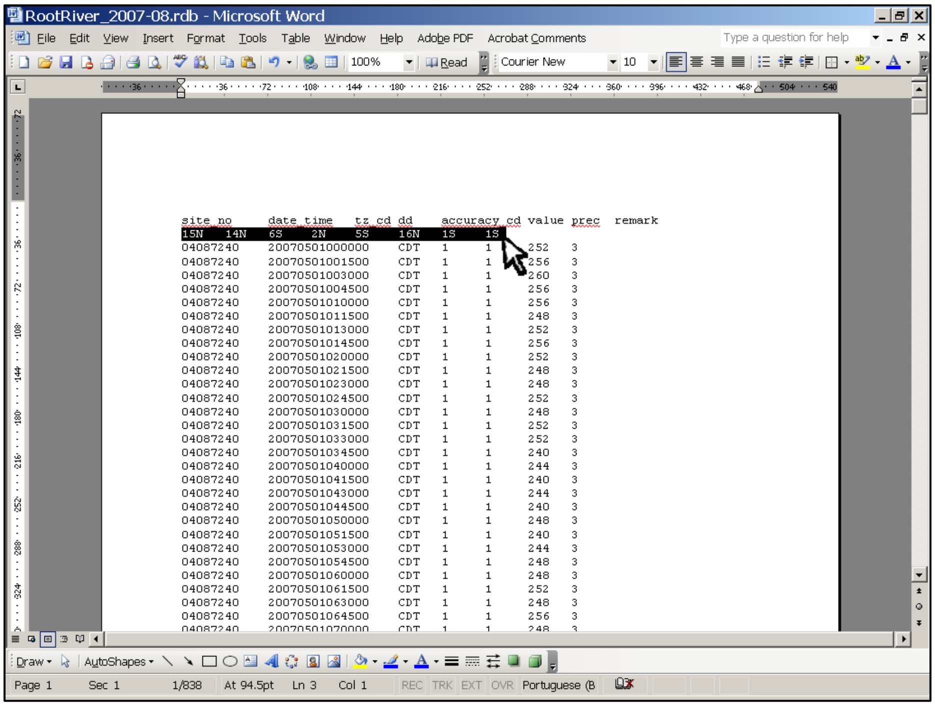
C.1.29. At the top of the file will be a data summary, disclaimer, and discriptions. An explanation of different “remarks” begins near the bottom of the first page (e.g., “e” = estimated values.)



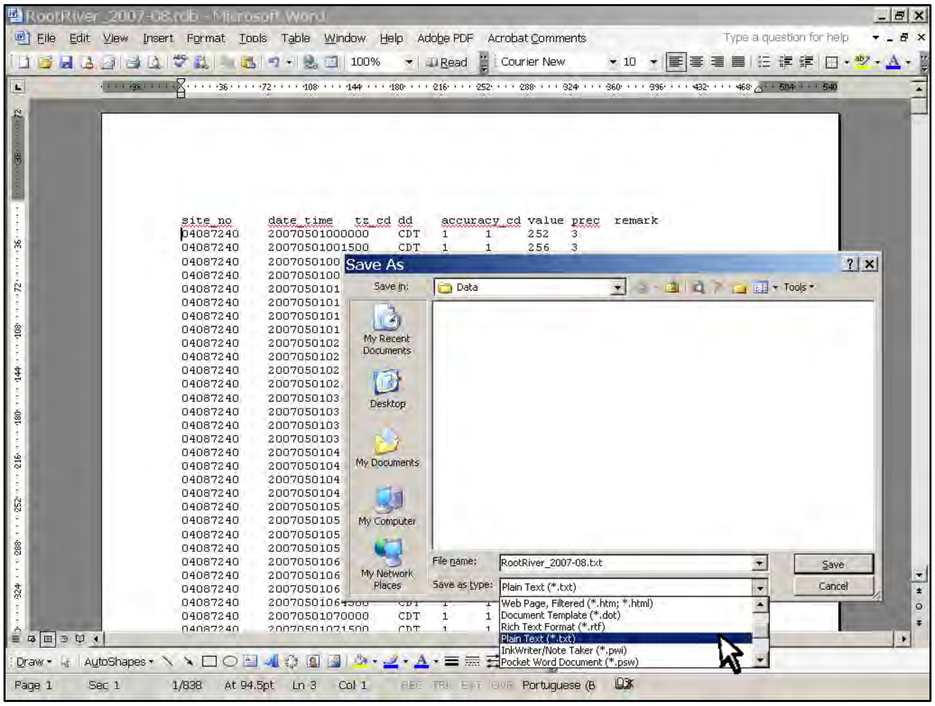
C.1.30. Click just to the left of the first row (to the left of the top most # sign) and hit the down arrow key until the last # sign is highlighted. Hit the “Delete” key.



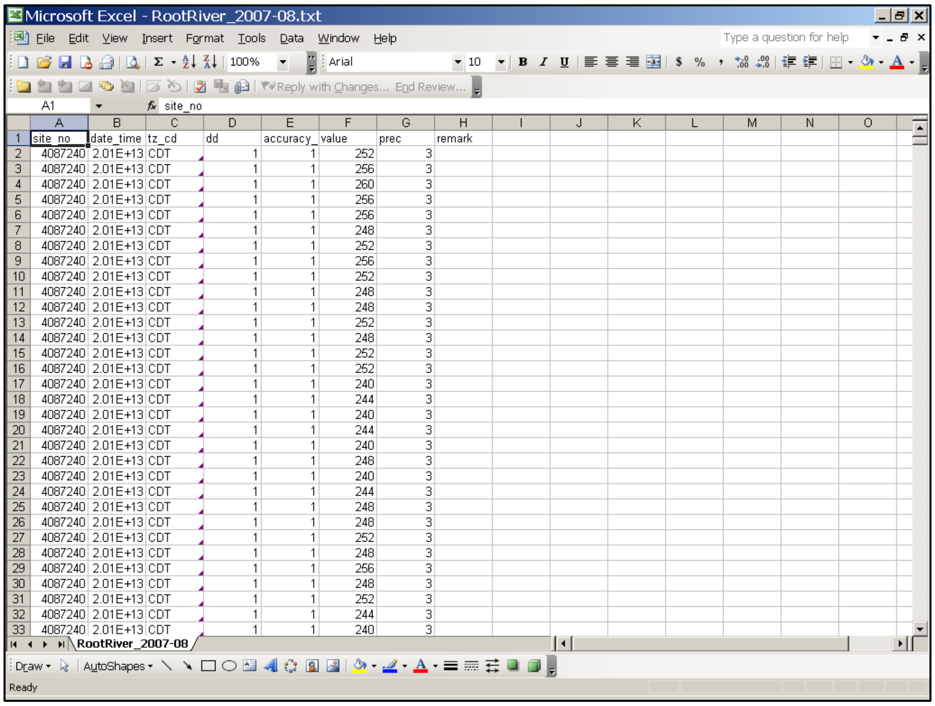
C.1.31. Click just to the left of the second row, hold the “Shift” key, and hit the down arrow key to highlight just that row. Click the “Delete” key.



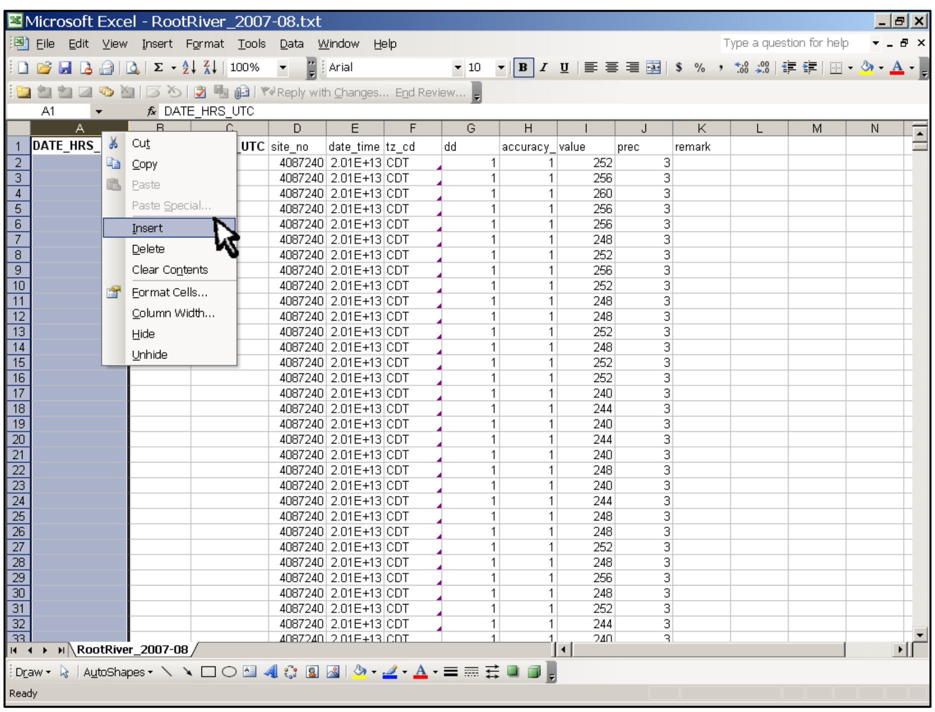
C.1.32. Following the procedure in B.7.24, use the MS Word pull-down menu (“File” > “Save as...”) save the table as a “Plain Text” (*.txt) file.



C.1.33. Following the procedures in steps C.1.13 – C.1.15, open MS Excel and import the text file you created in step C.1.32.



C.1.34. Following the procedures in step A.1.7, insert new columns (A-C) named “DATE_HRS_UTC”, “DATE”, and “HOURS_UTC”.



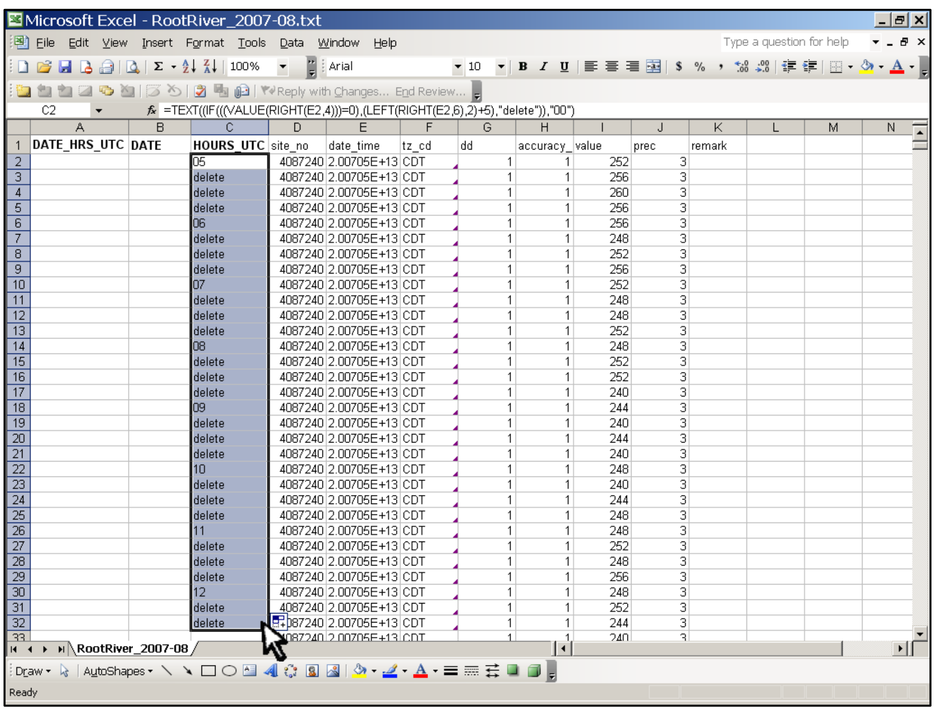
Note: As you build your input data set, you will need to match *E. coli* monitoring data to the measured values of your chosen explanatory variables, according to date-and-time of collection. Matching individual entries can be difficult, however, on account of differences in the time intervals at which different variables are reported (e.g., daily versus hourly). A *more efficient* approach is to create common time-reporting values in your respective data tables, which can then be used to “join” the tables into a single table of just those dates and times for which *E. coli* was measured. Various MS Excel functions can be used to combine time-interval values (i.e. year, month, day, and hour) into a single value for this purpose. An example follows. (For procedures on importing and “joining” tables using queries in MS Access, see Section 6 of this report, “Formatting Data for Model-Building.”)

USGS stream gauge data are reported in *local* time (UTC minus 5 hours in the central time zone).

C.1.35. USGS stream gauge data are reported in sub-hourly increments. As such, there will be several entries per hour. To flag redundant entries and create values for “HOURS_UTC” (column C) type the following function in cell C2:

```
=TEXT((IF(((VALUE(RIGHT(E2,4)))=0),(LEFT(RIGHT(E2,6),2)+5),"delete")),"00")
```

Hit “Enter”. The result should be a two-digit value for the hour-of-day, in UTC (Central Daylight Time + 5 hours.). To copy, click on the lower right hand corner of the cell, hold and drag the mouse to the bottom of the table.



C.1.36. To calculate “DATE” (column B), click on cell B2 and type the following function:

=CONCATENATE((LEFT(E2,4)),".",(RIGHT((LEFT(E2,6)),2)),".",(RIGHT((LEFT(E2,8)),2)))

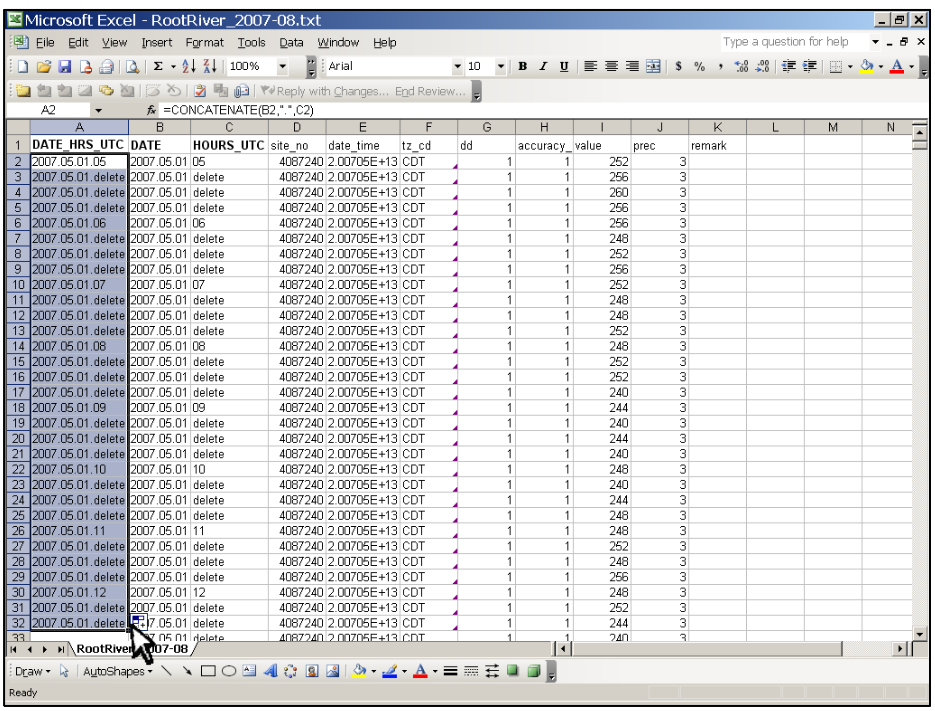
Hit “Enter”. This should return a date value (e.g., “2007.05.01” for May 1, 2007). Copy this function by clicking in the lower right corner of cell B2 and dragging to the end of the table.

⚠ **Note:** In order to correctly specify your model, your input data must be sorted in the correct chronological order. Your time-reporting values should therefore begin with the year, followed by the month, day, and hour. Be sure to include period-separators between the year, month, day, and hour (e.g., “2003.05.10.12” as opposed to “2003051012”). This will prevent *Virtual Beach* from misinterpreting them as long integers, instead of dates.

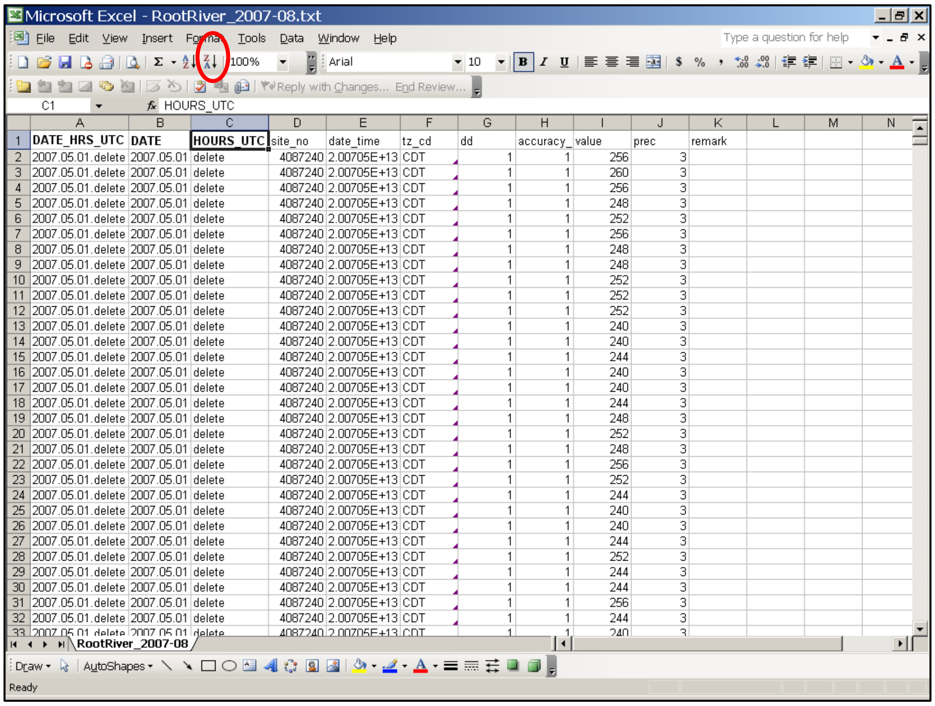
C.1.37. To calculate “DATE_HRS_UTC” (column A), click on cell A2 and type the following function:

=CONCATENATE(B2,".",C2)

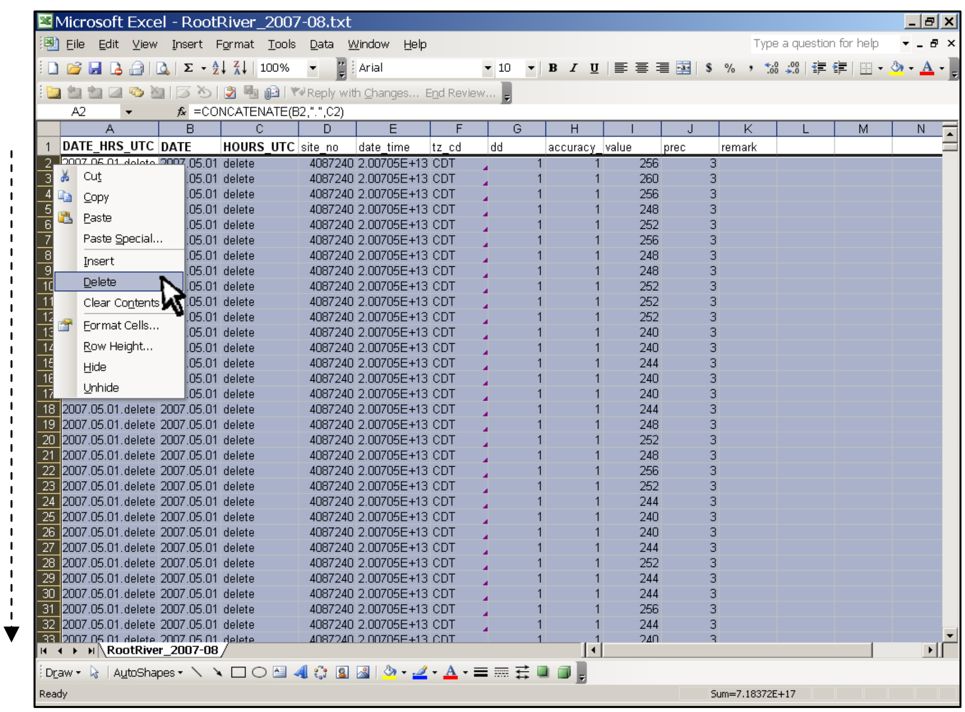
Hit “Enter”. This should return a date-plus-hour value (e.g., “2009.05.10.01” for 01:00 hours UTC on May 10, 2009). Copy this function by clicking in the lower right corner of cell A2 and dragging to the end of the table.



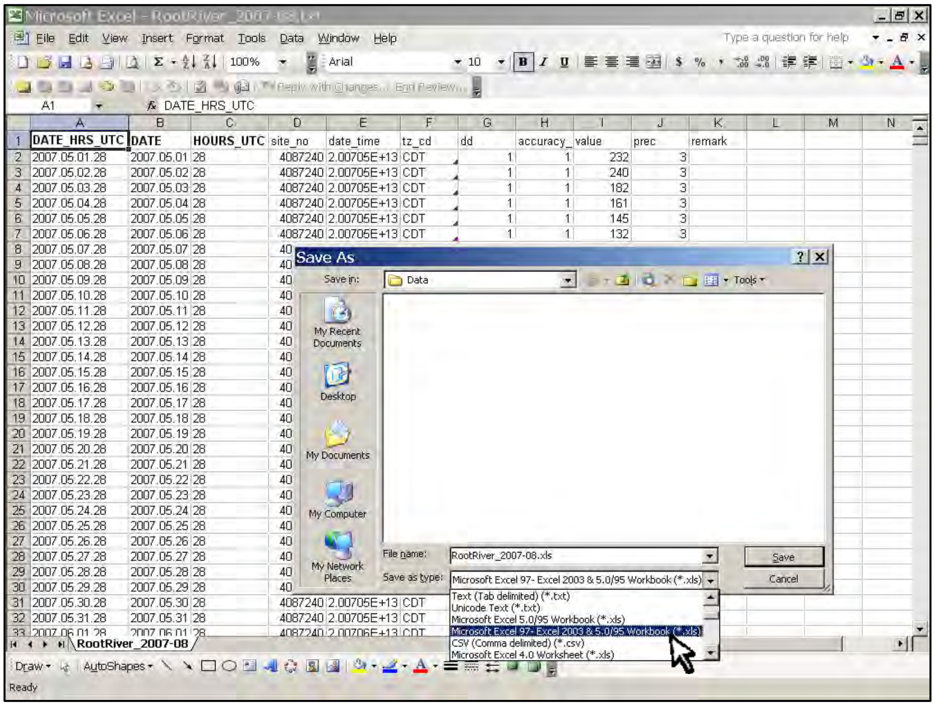
C.1.38. To remove the redundant hourly observations that you flagged in step C.1.37, first click on the first cell in column C and click on the reverse-sort button (). The redundant observations will be at the top of the table.



C.1.39. Next, click to the left of row 2 and, holding the "Shift" key, push the down-arrow on the keyboard until all of the redundant observations are selected (i.e. column C values = "delete"). Right-click on the last-selected row and select "Delete".



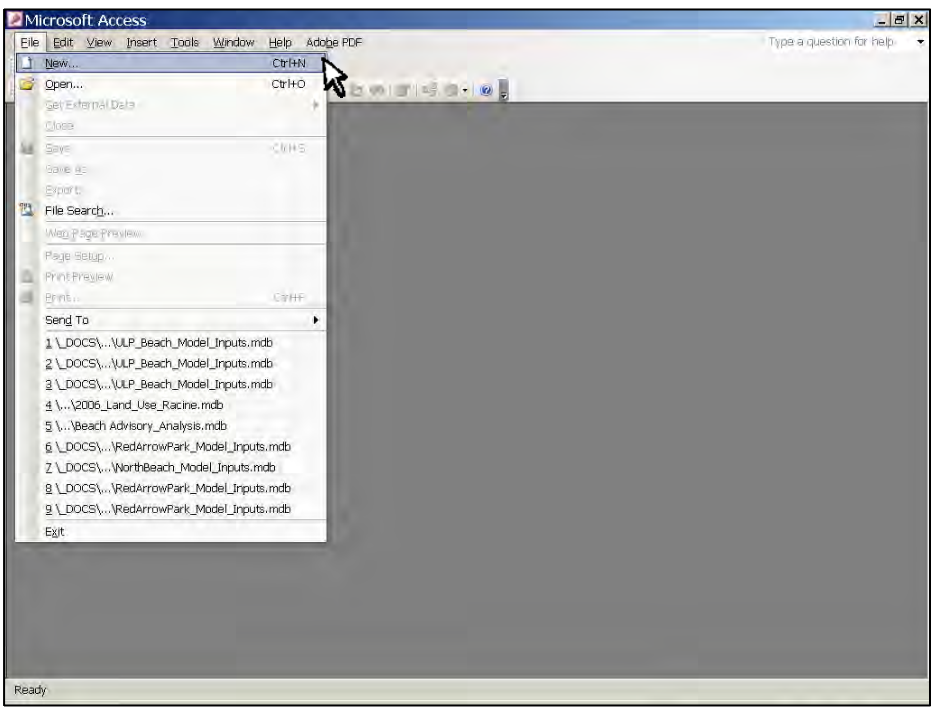
C.1.40. Save the table as an MS Excel file by selecting “File” > “Save as…” from the pull-down menu in the upper left corner, choosing the most recent version of MS Excel as the file type (*.xls), and clicking “Save”.



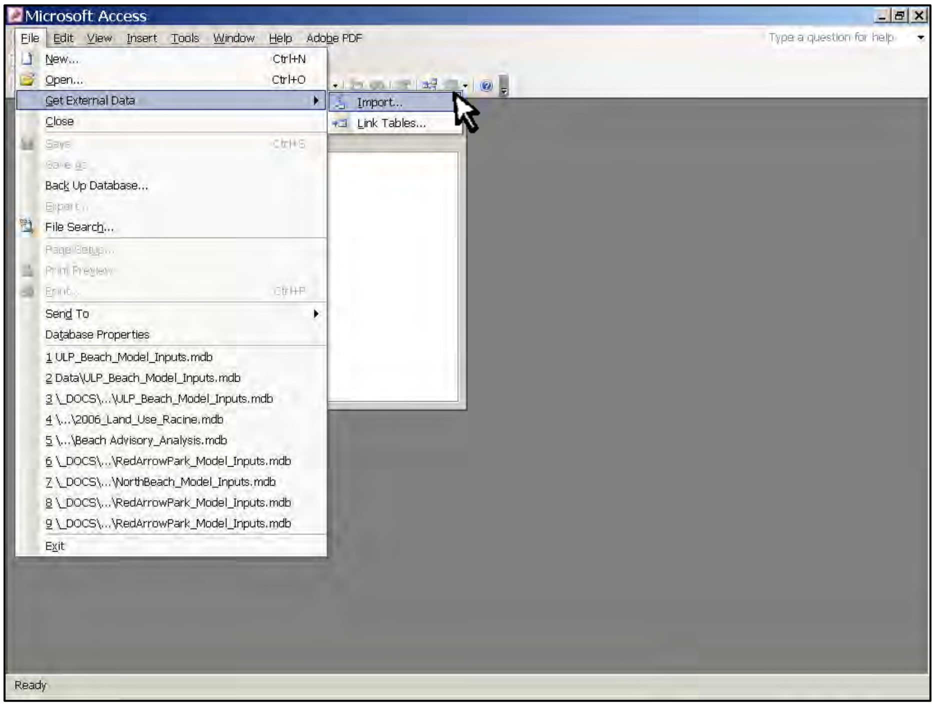
6. Formatting Data for Model-Building

As noted in the directions for downloading and formatting data described in Section 5, building a multivariate statistical model for predicting recreational water quality requires that you match your response variable (e.g., *E. coli* monitoring) to the measured values of your chosen explanatory variables, according to date-and-time of collection. Considering that different data are often collected/reported at different intervals (e.g., hourly versus daily), matching these data manually and combining them into a single table for model-building can be an arduous task. This process can be avoided, however, provided you have created common time-reporting values (table columns) using a system such as that described above (e.g., "DATE, HOURS_UTC"). If so, you can automatically "join" your various data tables into a single master table that contains all of the desired data, for only those dates and times when your response variable (e.g., *E. coli* concentration) was measured. The following is a step-by-step description of this process using MS Access database software.

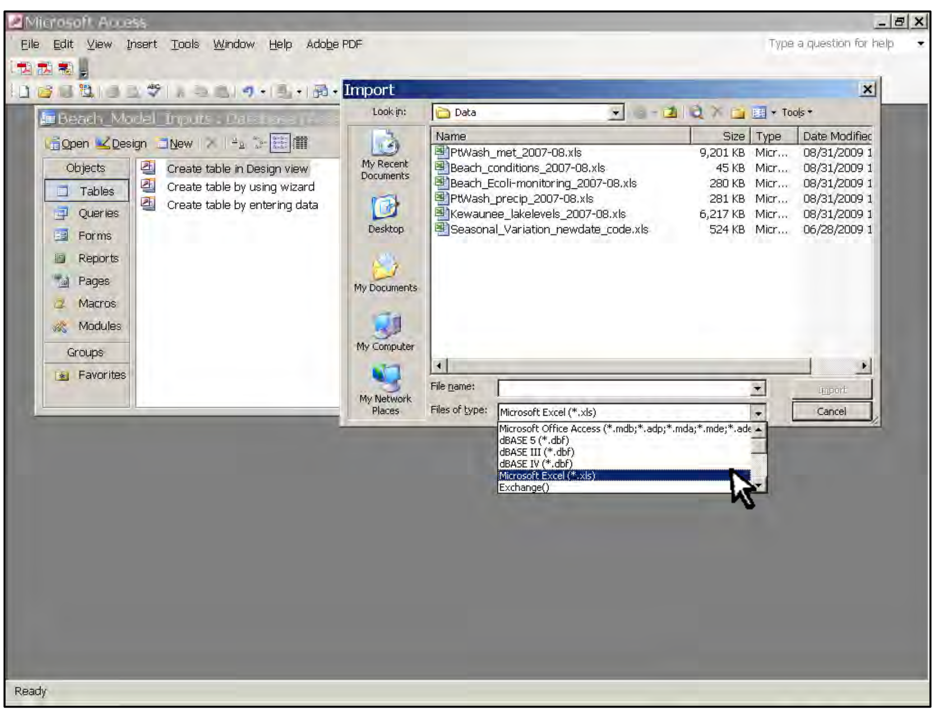
6.1. Open MS Access. When prompted, select "Blank database..." Browse to the directory and folder where you would like the database to reside, and name it something like "[beach name]_model_inputs.mbd."



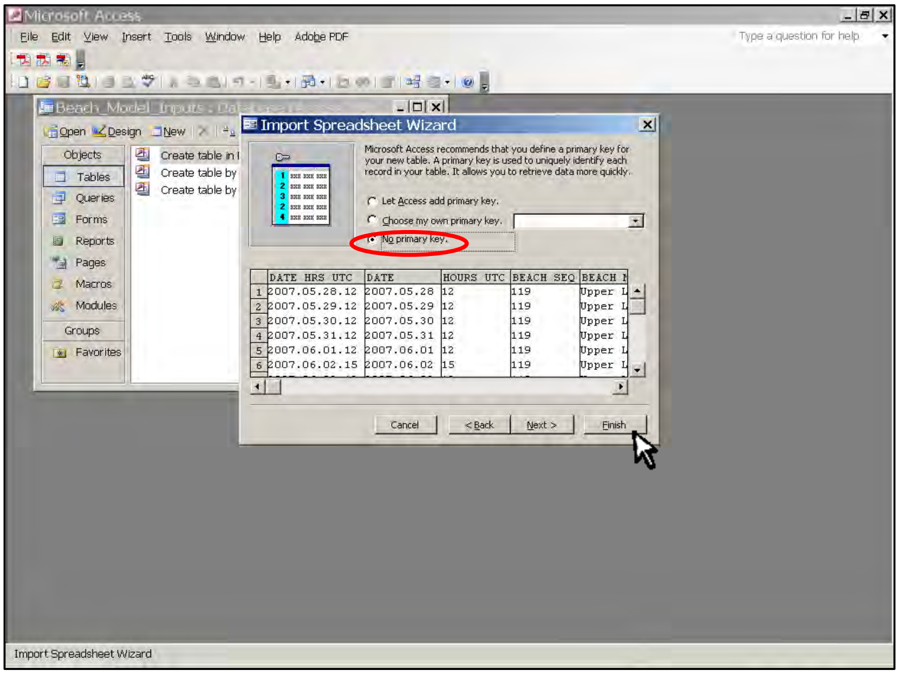
6.2 From the MS Access pull-down menu, select “File” > “Get External Data” > “Import”.



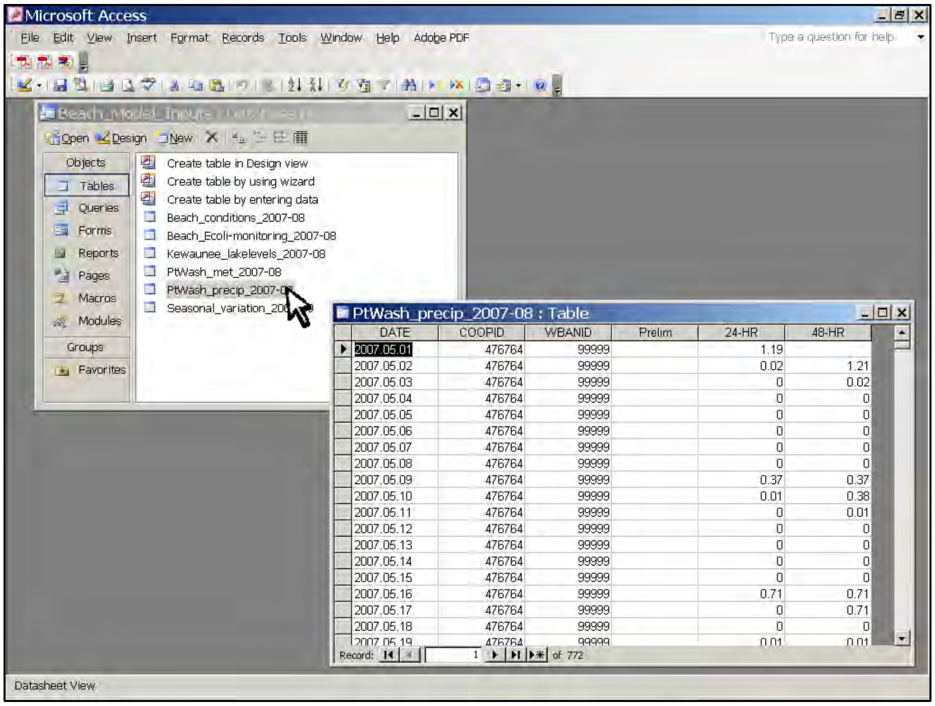
6.3 Browse to the directory where you have saved your data tables (spreadsheet files), select Microsoft MS Excel (*.xls) as the file type, and double-click the first file you want to include in your master input data table.



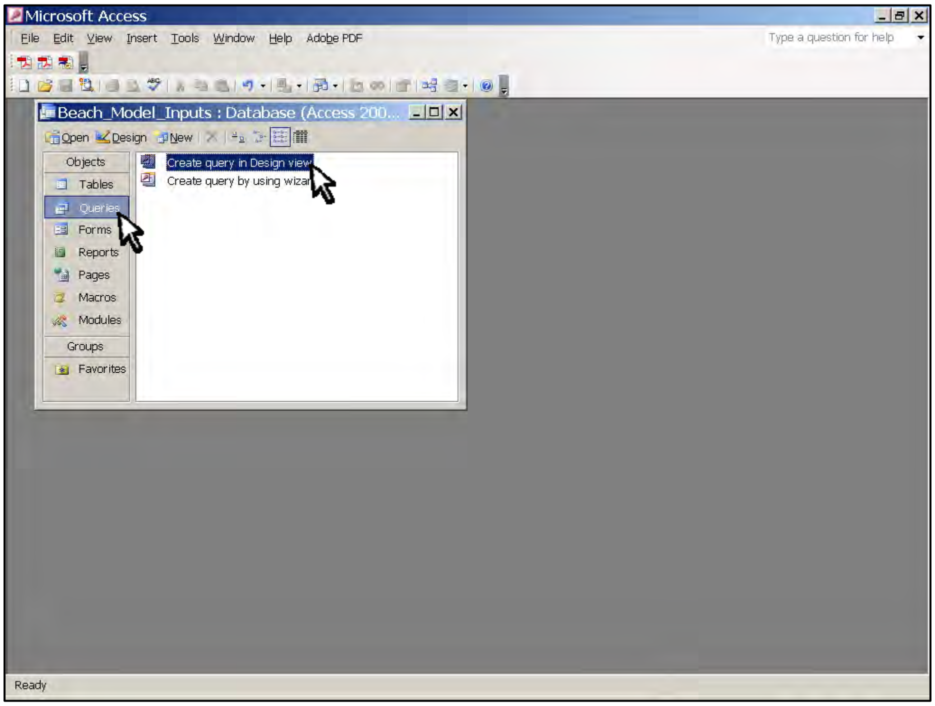
6.4. An “Import Spreadsheet Wizard” will open. Click the “Next>” button three times, until you get to “primary key” window shown below. Check the circle next to “No primary key”, then click the “Finish” button.



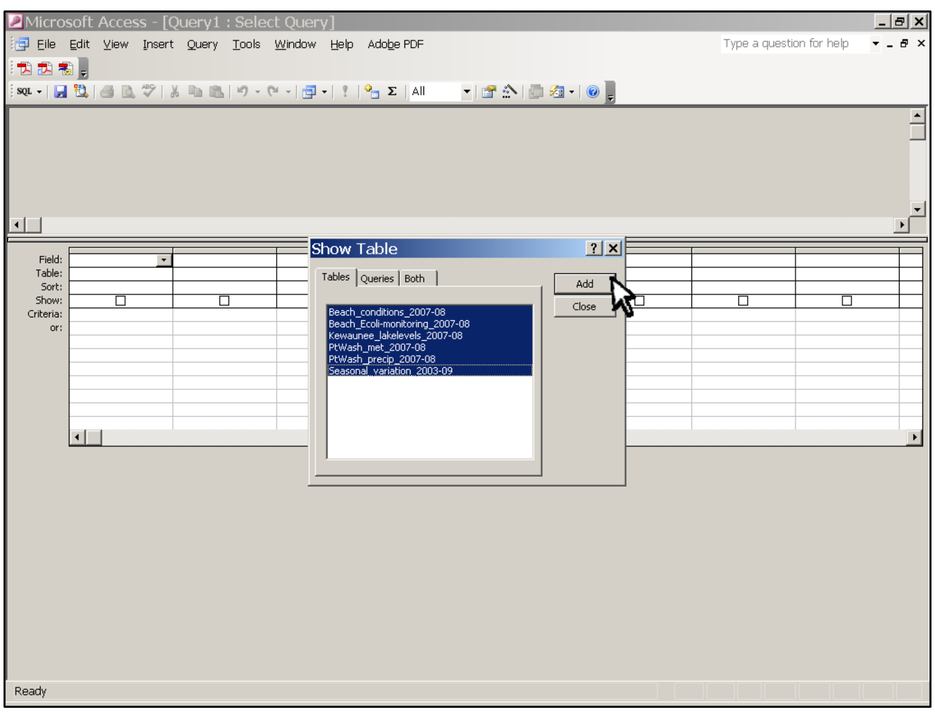
6.5. Repeat step 6.4 until all of the data tables you wish to merge are listed in the “Tables” tab of the main MS Access window (left-hand side, under “Objects”). To view any of the tables, double-click on their name in this window.



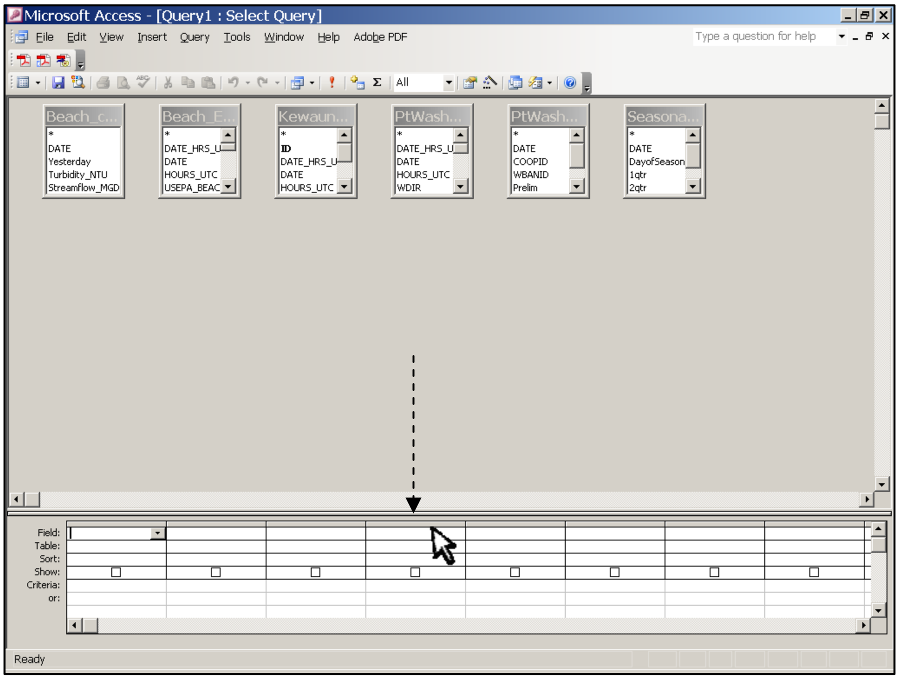
6.6. In the main MS Access window, click on the “Queries” tab, and double-click on “Create query in Design view”.



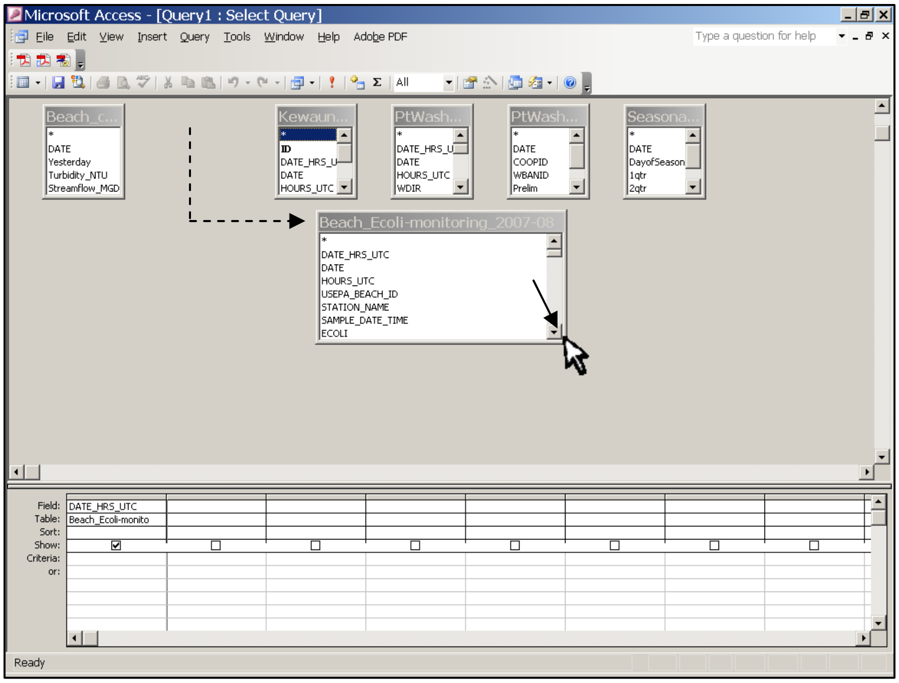
6.7. Two windows will appear: a split “Query Design” window and “Show Table,” containing a list of tables in the database file. Select (highlight) all of the tables and click the “Add” button. A series of small windows will appear within the top portion of the “Query Design” window, representing the selected tables (see step 6.8).



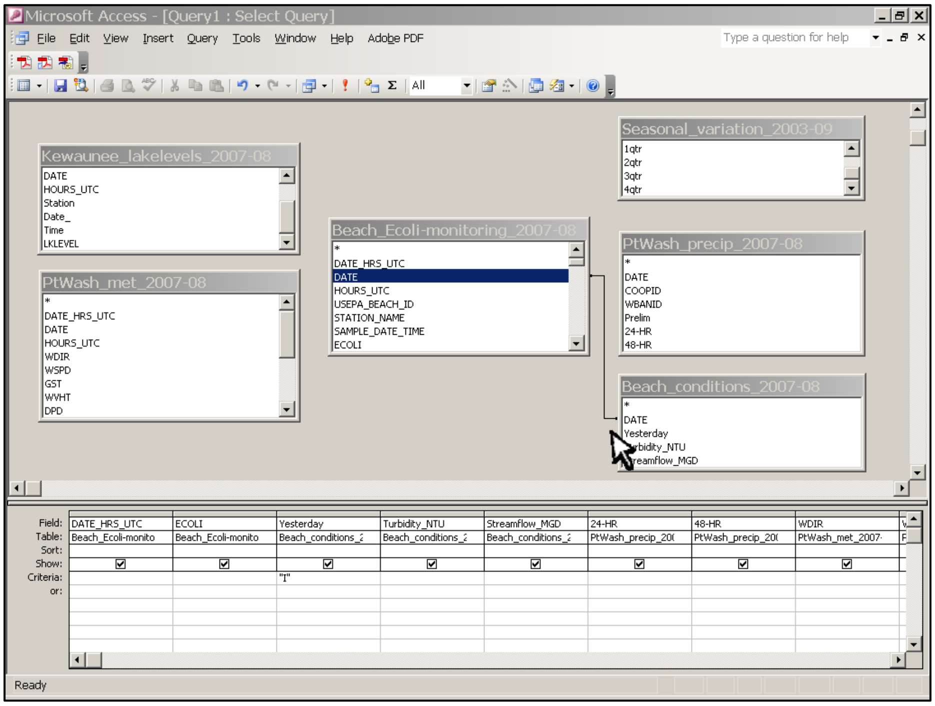
6.8. For easier viewing, maximize the “Query Design” window and expand the top section by clicking on the horizontal threshold and dragging it downward (be sure to leave the top few lines of the bottom portion of the window visible).



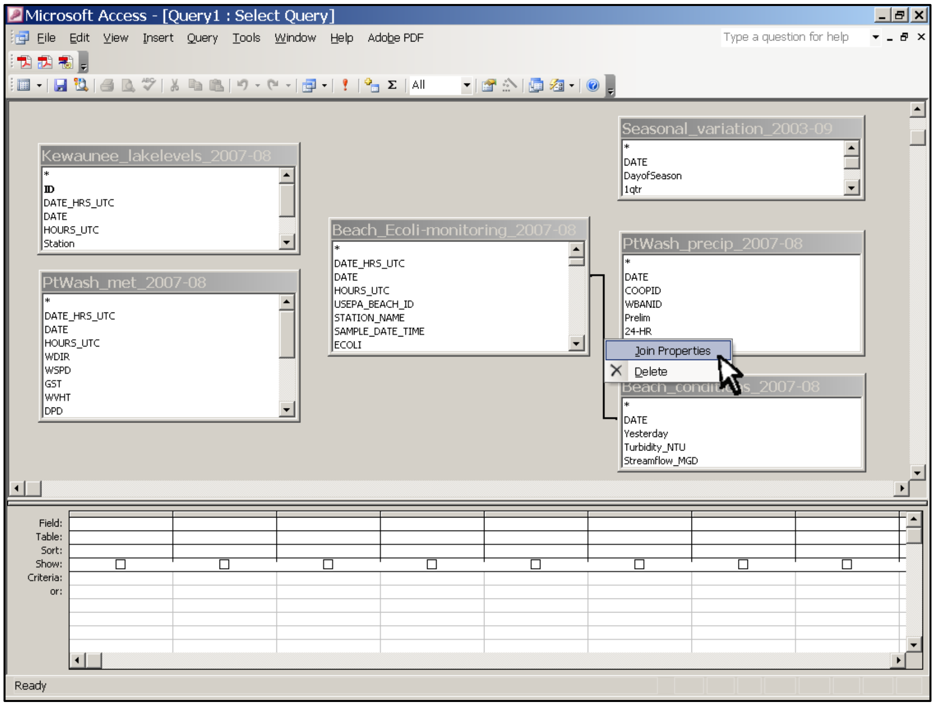
6.9. Click and hold the title bars of the windows representing the data tables and move (drag) them around so that the response variable (e.g., *E. coli* monitoring data) is in the center. Resize the windows by clicking and dragging their corners or sides.



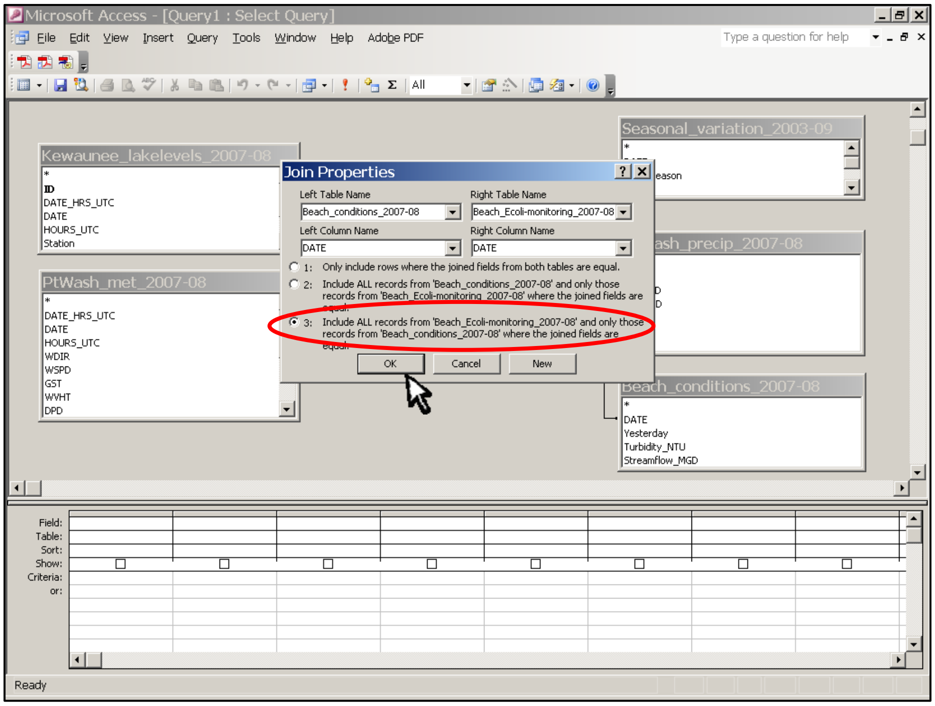
6.10. After re-sizing and arranging the data table windows (step 6.9), create "joins" (links) between the explanatory data tables and the response table, one at a time. First click on the common date-and-time field in the explanatory data table (e.g., "DATE"). Hold and drag the mouse to the corresponding field in the response table, and let go. A "join" line will appear.



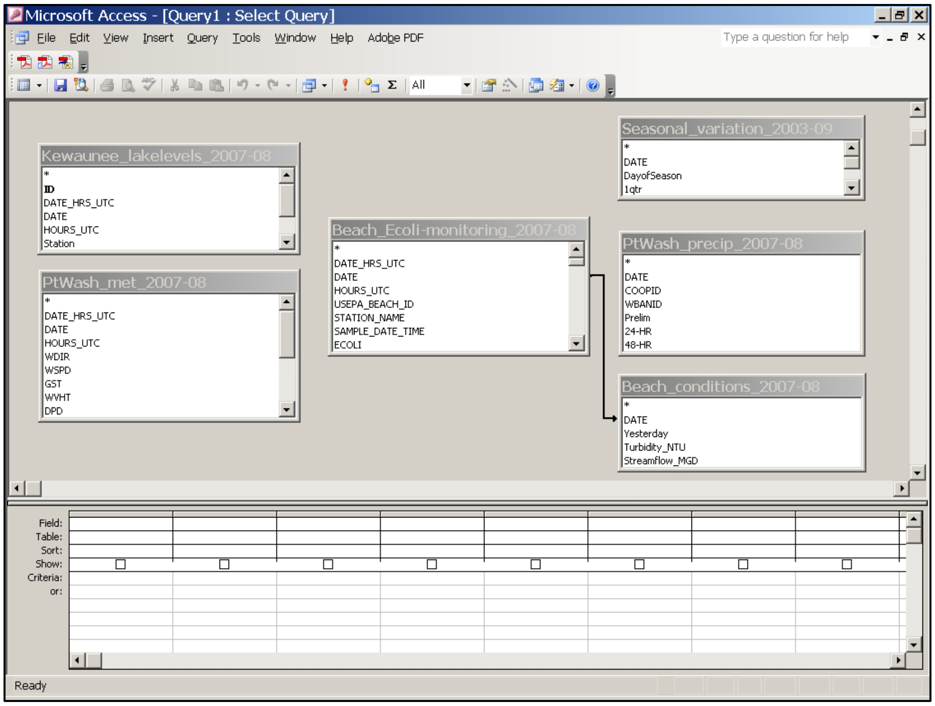
6.11. Right-click on the join line and select "Join Properties."



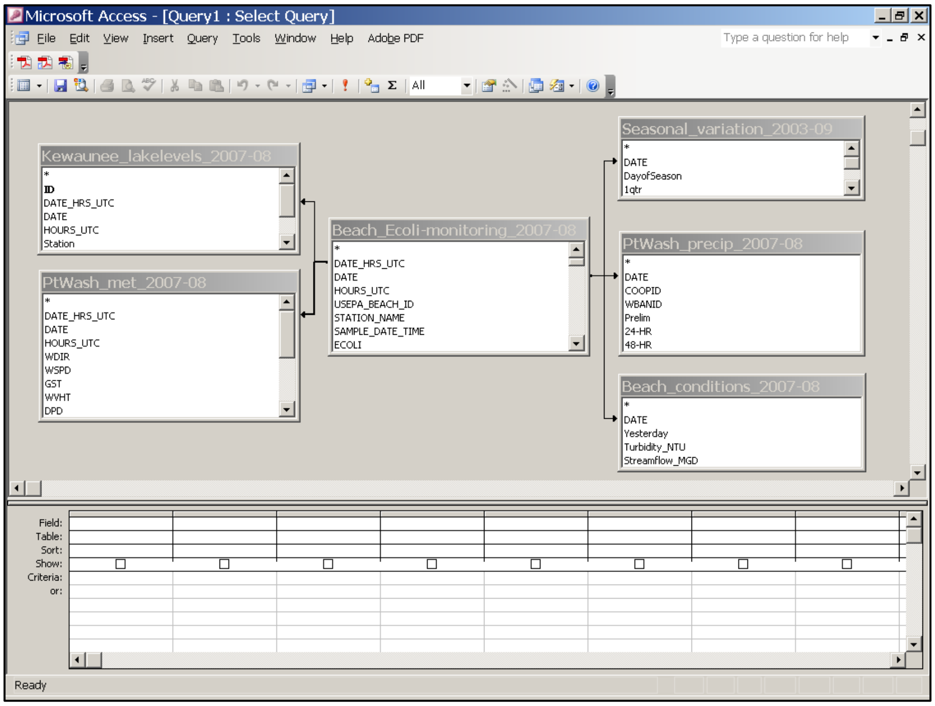
6.12. In the “Join Properties” window, check the circle next to option #3 – whereby the combined table will contain ALL of the records in the response variable table (e.g., recorded *E. coli*) but only those records in the explanatory data table with corresponding date-and-time values.



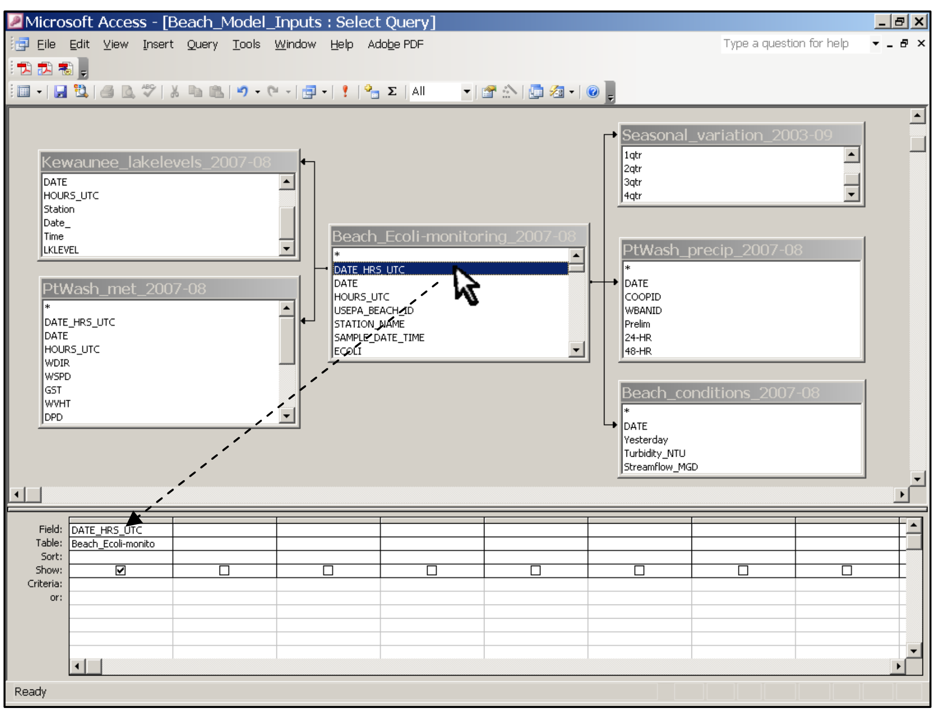
6.13. After completing step 6.12, the line between the table windows will turn into an arrow. Make sure the arrow is pointing *away* from the response variable table.



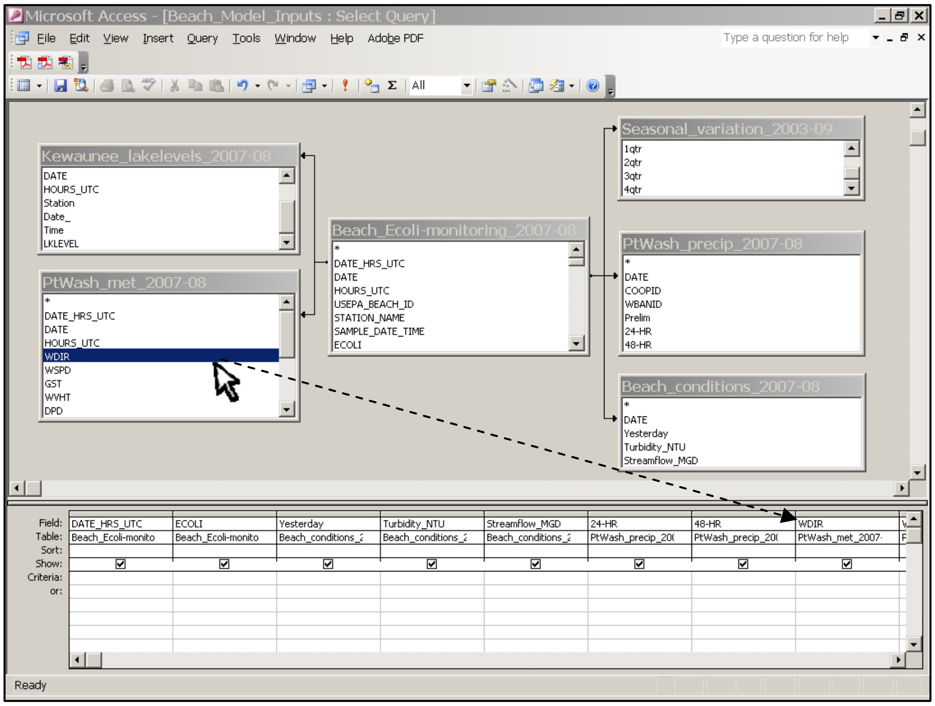
6.14. Repeat steps 6.11-6.13 until each of the explanatory data tables are linked to the response variable table with the arrows pointing away. (**Note:** Where explanatory data tables have a “DATE_HRS_UTC” field, be sure the lines you draw connect that field, and *not* “DATE”.)



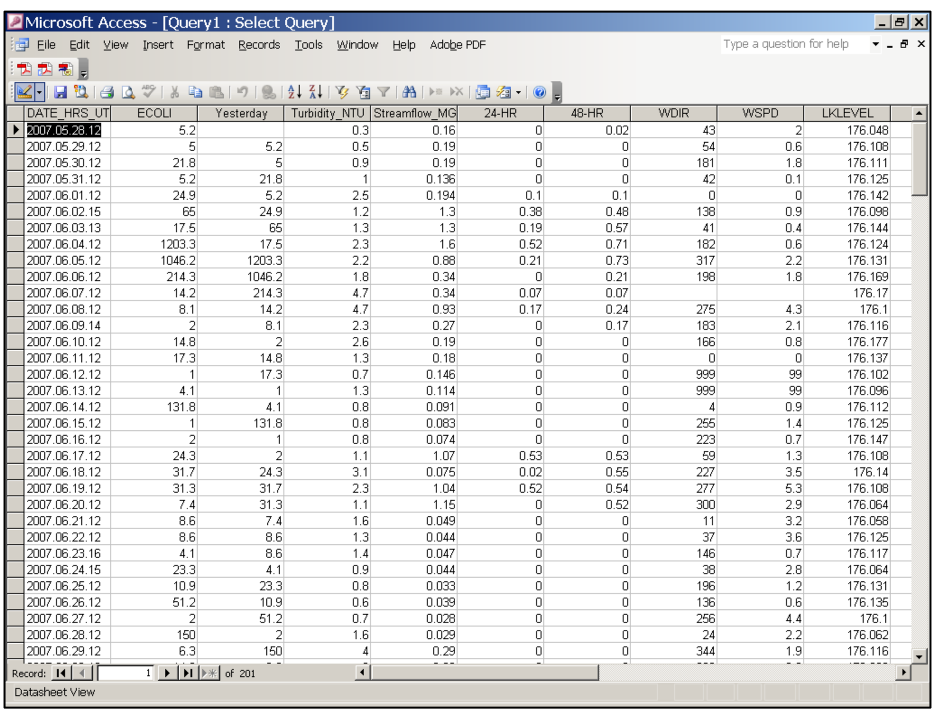
6.15. Click on the shortest interval time field in the response variable table (e.g., “DATE_HRS_UTC”) and drag it to the far left side of the query summary table.



6.16. Repeat step 6.15 for the response variable (e.g., *E. coli*), and then each of the desired explanatory variables – filling the query table from left to right.



6.17. Click the Run Query button ( ) shown in step 6.15. This will open a complete query table, with all of the selected fields from the various joined data tables.

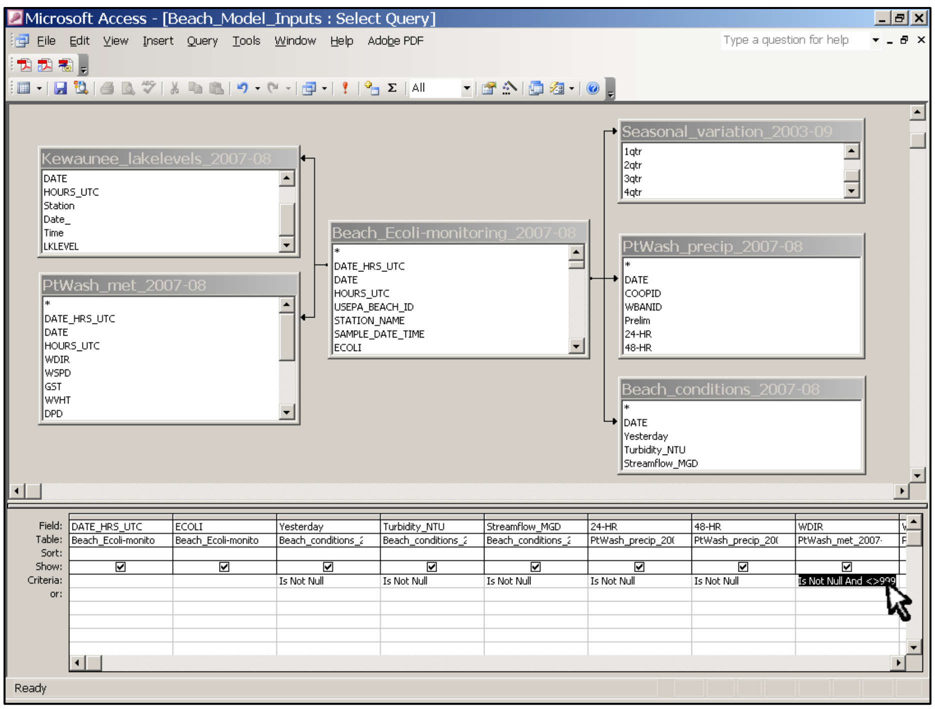


DATE_HRS_UT	ECOLI	Yesterday	Turbidity_NTU	Streamflow_MG	24-HR	48-HR	WDIR	WSPD	LKLEVEL
2007.05.28.12	5.2		0.3	0.16	0	0.02	43	2	176.048
2007.05.29.12	5	5.2	0.5	0.19	0	0	54	0.6	176.108
2007.05.30.12	21.8	5	0.9	0.19	0	0	181	1.8	176.111
2007.05.31.12	5.2	21.8	1	0.136	0	0	42	0.1	176.125
2007.06.01.12	24.9	5.2	2.5	0.194	0.1	0.1	0	0	176.142
2007.06.02.15	65	24.9	1.2	1.3	0.38	0.48	138	0.9	176.098
2007.06.03.13	17.5	65	1.3	1.3	0.19	0.57	41	0.4	176.144
2007.06.04.12	1203.3	17.5	2.3	1.6	0.52	0.71	182	0.6	176.124
2007.06.05.12	1046.2	1203.3	2.2	0.88	0.21	0.73	317	2.2	176.131
2007.06.06.12	214.3	1046.2	1.8	0.34	0	0.21	198	1.8	176.169
2007.06.07.12	14.2	214.3	4.7	0.34	0.07	0.07			176.17
2007.06.08.12	8.1	14.2	4.7	0.93	0.17	0.24	275	4.3	176.1
2007.06.09.14	2	8.1	2.3	0.27	0	0.17	183	2.1	176.116
2007.06.10.12	14.8	2	2.6	0.19	0	0	166	0.8	176.177
2007.06.11.12	17.3	14.8	1.3	0.18	0	0	0	0	176.137
2007.06.12.12	1	17.3	0.7	0.146	0	0	999	99	176.102
2007.06.13.12	4.1	1	1.3	0.114	0	0	999	99	176.096
2007.06.14.12	131.8	4.1	0.8	0.091	0	0	4	0.9	176.112
2007.06.15.12	1	131.8	0.8	0.083	0	0	255	1.4	176.125
2007.06.16.12	2	1	0.8	0.074	0	0	223	0.7	176.147
2007.06.17.12	24.3	2	1.1	1.07	0.53	0.53	59	1.3	176.108
2007.06.18.12	31.7	24.3	3.1	0.075	0.02	0.55	227	3.5	176.14
2007.06.19.12	31.3	31.7	2.3	1.04	0.52	0.54	277	5.3	176.108
2007.06.20.12	7.4	31.3	1.1	1.15	0	0.52	300	2.9	176.064
2007.06.21.12	8.6	7.4	1.6	0.049	0	0	11	3.2	176.058
2007.06.22.12	8.6	8.6	1.3	0.044	0	0	37	3.6	176.125
2007.06.23.16	4.1	8.6	1.4	0.047	0	0	146	0.7	176.117
2007.06.24.15	23.3	4.1	0.9	0.044	0	0	38	2.8	176.064
2007.06.25.12	10.9	23.3	0.8	0.033	0	0	196	1.2	176.131
2007.06.26.12	51.2	10.9	0.6	0.039	0	0	136	0.6	176.135
2007.06.27.12	2	51.2	0.7	0.028	0	0	256	4.4	176.1
2007.06.28.12	150	2	1.6	0.029	0	0	24	2.2	176.062
2007.06.29.12	6.3	150	4	0.29	0	0	344	1.9	176.116

Note: In the example shown in step 6.17, several cells are blank, representing missing data. Predictive models *cannot* be built from datasets with missing (null) values.

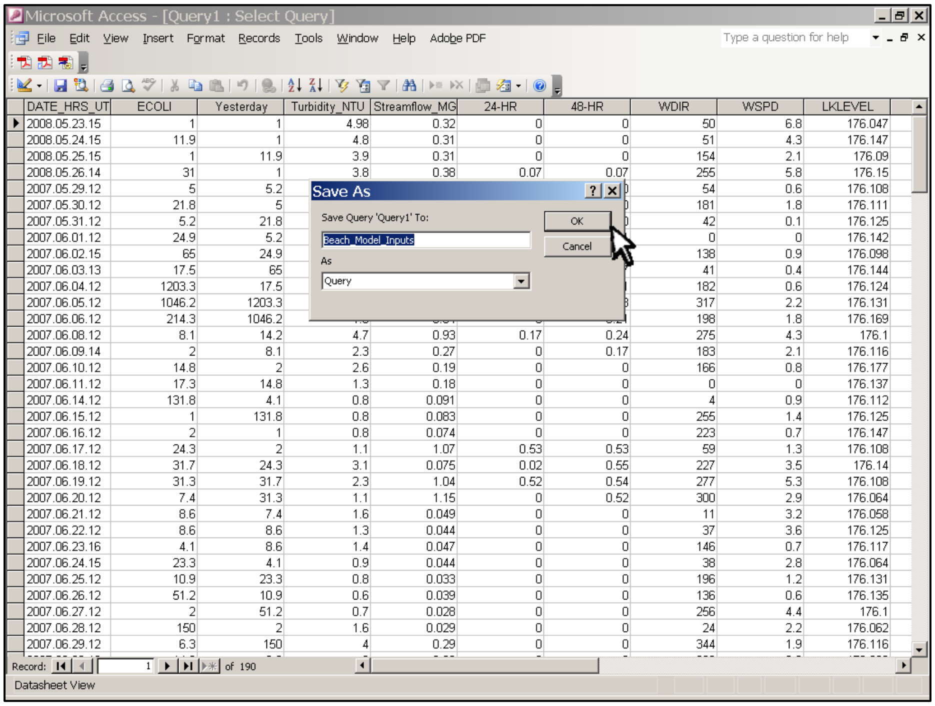
Note: Also in the example shown in step 6.17, several cells under the field "WDIR" (wind direction) have a value of "999," which also represents missing values, as reported in the **National Data Buoy Center** (see 5.B.6). Leaving these records in the final data table could skew the resulting model and lessen its predictive power.

6.18. To eliminate records with missing data, click the "Design View" button ( ) shown in step 6.17. In the "Criteria" row of the query summary table at the bottom of the window, type "Is Not Null" under each of the explanatory variables. For data from the **National Data Buoy Center** (e.g., "WDIR"), type "Is Not Null and <> 999" to eliminate records with "999" values.

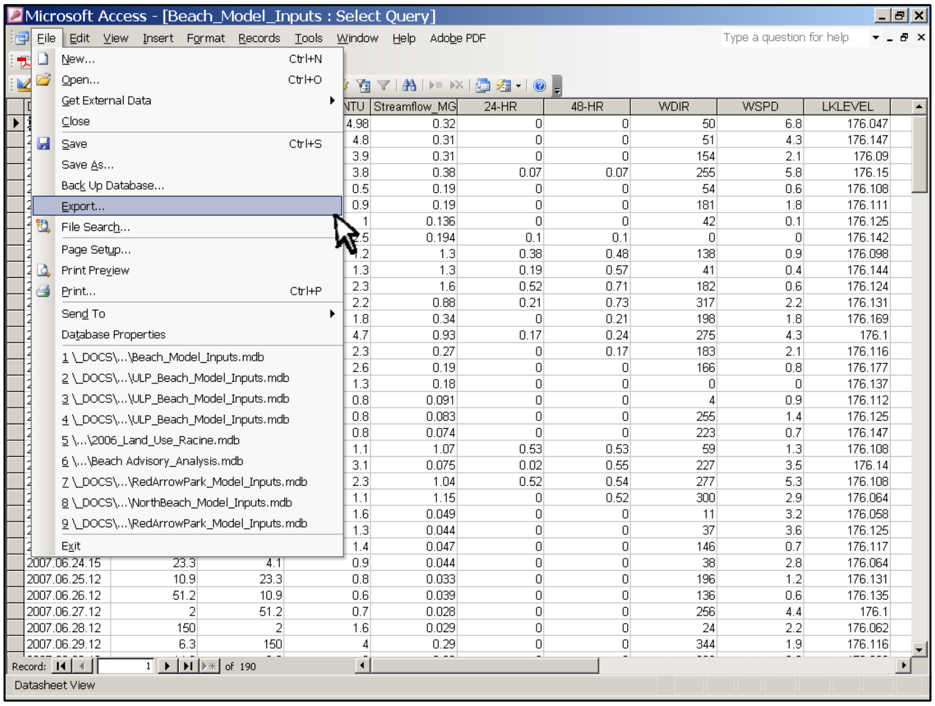


6.19. Click the Run Query button ( ) shown in step 6.18 to view the refined query table. Note that the total number of fields (listed in the lower left corner of the Table View) will have decreased, since records with missing data have been filtered-out.

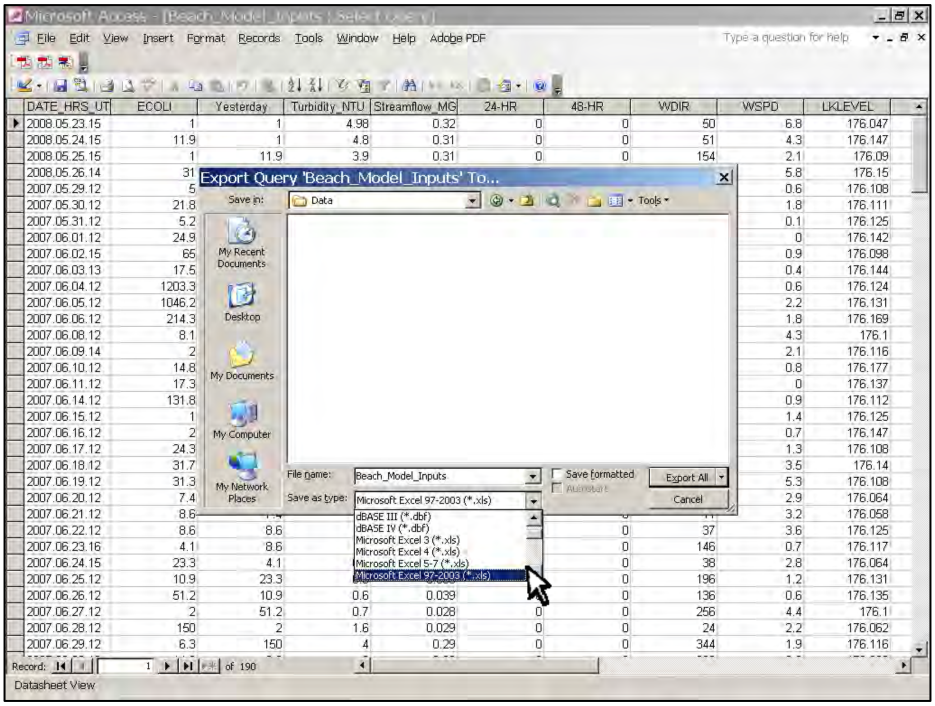
6.20. From the MS Access pull-down menu, select “File” > “Save As...”. Name the table something like “[beach name]_model_inputs”, select “Query” (the default) from the pull-down menu, and click “OK”.



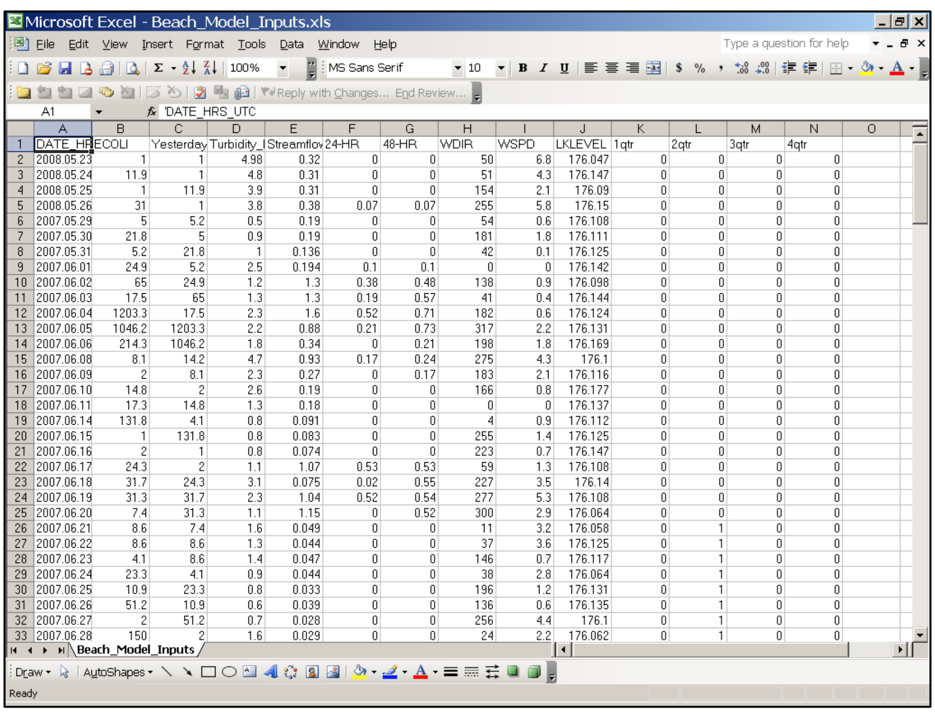
6.21. After the query has been saved, from the MS Access pull-down menu, select “File” > “Export...”.



6.22. In the “Export Query” window, browse to the directory where you would like the master table to reside, select Microsoft MS Excel (*.xls) as the file type, and click the “Export All” button.



6.23. Open the exported table in MS Excel to examine, and if necessary, to make additional formatting changes prior to using the dataset to build your predictive model using *Virtual Beach* or another statistical package.



Science Services

Center for Excellence –
providing expertise for science-based decision-making

We develop and deliver science-based information, technologies, and applications to help people make well-informed decisions about natural resource management, conservation, and environmental protection.

Our Mission: The Bureau of Science Services supports the Wisconsin Department of Natural Resources and its partners by:

- conducting research and acquiring original knowledge.
- analyzing new information and emerging technologies.
- synthesizing information for policy and management decisions.
- applying the scientific method to the solution of environmental and natural resources problems.
- providing science-based support services for department initiatives.
- collaborating with local, state, regional, and federal agencies and academic institutions in Wisconsin and around the world.
