NOAA Technical Memorandum GLERL-156

BEACH WATER QUALITY DECISION SUPPORT SYSTEM

David Rockwell[1]
Kent Campbell[1]
Gregory Lang[2]
David Schwab[2,4]
Greg Mann[3]
Richard Wagenmaker[3]

[1]Cooperative Institute for Limnology and Ecosystems Research, 4840 S. State Rd., Ann Arbor, MI 48108

[2]NOAA, Great Lakes Environmental Research Laboratory, 4840 S. State Rd., Ann Arbor, MI 48108

[3]NOAA, National Weather Service Forecast Office, Detroit/Pontiac, 9200 White Lake Rd., White Lake, MI 48386

[4]Retired. Now at University of Michigan, Graham Institute, 214 S. State, Suite 200, Ann Arbor, MI 48104

March 2013
Revised June 2014

UNITED STATES
DEPARTMENT OF COMMERCE

Rebecca Blank
Acting Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

Jane Lubchenco
Under Secretary for Oceans & Atmosphere
NOAA Administrator

NOTICE

Mention of a commercial company or product does not constitute an endorsement by the NOAA. Use of information from this publication concerning proprietary products or the tests of such products for publicity or advertising purposes is not authorized. This is GLERL Contribution No. 1652.

This publication is available as a PDF file and can be downloaded from GLERL's web site: www.glerl.noaa.gov. Hard copies can be requested from GLERL Information Services, 4840 S. State Rd., Ann Arbor, MI 48108. pubs.glerl@noaa.gov.

NOAA's Mission – To understand and predict changes in Earth's environment and conserve and manage coastal and marine resources to meet our nation's economic, social, and environmental needs.

NOAA's Mission Goals:

- Protect, restore and manage the use of coastal and ocean resources through an ecosystem approach to management
- Understand climate variability and change to enhance society's ability to plan and respond
- Serve society's needs for weather and water information
- Support the Nation's commerce with information for safe, efficient, and environmentally sound transportation
- Provide critical support for NOAA's Mission

TABLE OF CONTENTS

LIST OF FIGURES

APPENDIX 1: Hydrodynamic Grid Cell Locations

LIST OF TABLES

LIST OF ATTACHMENTS

Beach Water Quality Decision Support System

D.C. Rockwell
K.B. Campbell
G.A. Lang
D.J. Schwab
G. Mann
R. Wagenmaker

ABSTRACT

The need for rapid assessment of bacterial contamination at beaches is well known. Bacterial concentrations change rapidly (Bohm, Whitman et al. 1995, Olyphant and Whitman 2004, Whitman and Nevers 2008). The "persistence model" regulates today's swimming with yesterday's *Escherichia coli* (*E. coli*) measurement. As a result, this beach management tool is known to be less protective of human health than desired but was still used at over 500 beaches in 2010.

One method for meeting this problem has been to develop rapid analytical methods taking two hours of laboratory analytical time. This approach is becoming operationally available but at higher analytical cost (Setty 2012) than the slower cultural methods currently employed. These rapid analytical methods provide beach managers with the capability to advise swimmers about *E. coli* concentrations on the same day, but only if sampling, transportation, and data management time components can be completed within four hours, giving an overall time of 6 hours. This requires early field sampling and laboratories close to the beaches being sampled.

Another approach is to use Nowcast predictive models that can provide estimations of *E. coli* during the same day (Francy 2009) by measuring easily determined variables that can be correlated with *E. coli* bacterial concentrations. The Nowcast predictive models were being used at 10 beaches in 2010 (Adam Mednick, personal communication). It is expected that Nowcast predictive models will be expanded in 2011 to more than 20 beaches because of Great Lake Restoration Initiative funding.

This technical memorandum is based primarily on the final grant report to Ed Pniak, GLRI Project Officer, Michigan Project Lead, U.S. EPA, Region 5, Water Division, State and Tribal Programs Branch for GL-00E00658. The grant is titled "60 Hour Beach Forecasting Models." The principle investigators are Allen Burton, Kent Campbell, and David Rockwell.

This technical memorandum provides beach managers with a tool to forecast several days in advance the likelihood of *E. coli* concentrations exceeding the state single sample regulatory standard. This predictive beach water quality management tool is the only beach management decision support system (DSS) capable of forecasting several days in advance the beach water quality bacterial concentrations because it limits the explanatory variables to those variables for which the National Weather Service (NWS) is able to make forecasts.

The forecast DSS was applied to 35 sampling locations at 24 beaches during the 2010 swimming season. The results of Grant GL-00E00658 showed the forecast DSS provided the same or better beach management decision support than the persistence model at 71% (25 of the 35) of the sampling sites. In addition, the stratification of beach sampling sites based on 5% of the samples exceeding the state regulatory standard single sample maximum of 235 counts/100 ml improved the forecast DSS. The forecast DSS for the beaches meeting this criteria provided better (70%) beach management support or the same (15%) beach management support than the persistence model for a total of 85% (23 of the 27) of the sampling sites. The evaluation of the percent of samples exceeding the state regulatory standard allows the beach manager to readily determine if the beach is more suitable for the

forecast DSS tool. Forecast DSS has the potential to provide the swimming community with information several days in advance, allowing the planning of their water recreational activities and reducing the risk of swimming in bacterially contaminated water while maximizing the opportunity to use the water recreational facilities.

Lastly, application of the forecast DSS to all monitored beaches with two or more samples per week in 2010 could have led to about a 23% reduction (862 swimming advisory days) of the 3766 reported (NRDC 2011). The lost value of swimming for the Great Lake recreational swimmers ranges from \$11.3M to \$117M for these days when swimming is banned (Shaikh 2006 and Rabinovici et al. 2004).

1. INTRODUCTION

This technical memorandum is based primarily on the final grant report to Ed Pniak, GLRI Project Officer, Michigan Project Lead, U.S. EPA, Region 5, Water Division, State and Tribal Programs Branch for GL-00E00658. The grant is titled "60 Hour Beach Forecasting Models". The principle investigators are Allen Burton, Kent Campbell, and David Rockwell.

Predictive models have served as a management tool for advisory or closure of swimming at some beaches since 1990 (Kuntz 1998). Early predictive models were rainfall-based alert curve models, were statistical in nature and did not distinguish between point and non-point sources. (DNREC 1997, Kuntz 1998, USEPA 1999). Numerous deterministic models explicitly incorporating advection, transport, and decay processes have been used (USEPA 1999) to predict water quality parameters as well as *E. coli* levels. Multiple Linear Regression (MLR) models began to be used in the Great Lakes in the late 1990's (Francy 2009, SwimCast, Whitman et al.). A systematic review of published Nowcasting/Forecasting studies in the Great Lakes basin employing various predictive models indicated that bona fide unbiased tests of predictive ability were almost invariably lacking. (Findlay et al. 2009). Nowcasting models have lower errors in making decisions compared to the persistence model. Now/forecasting models achieve moderate predictive accuracy using R2 and Cp- statistics as joint criteria (Ge and Frick 2007). Typical adjusted R2 for Nowcast models range from 38 to 44% (Francy et al. 2006b).

In the earliest reported scientific studies, (Veley et al. 1998, Francy and Darner 1998) USGS scientists looked at using turbidity, rainfall, and wave height in a beach-specific statistical model to estimate *E. coli* concentrations. Since characterization and statistical modeling of bacterial concentrations was introduced in the Great Lakes, mathematical modeling for bacterially induced beach closures has become a growing body of published research and scientific investigations (Olyphant et al. 2003, Francy and Darner 2003). Application of multiple linear regression (MLR) for single beaches (Olyphant and Whitman 2004, Pfister 2004, Francy et al. 2006, Zimmerman 2008, Przybyla-Kelly et al. 2008) has expanded to regional approaches (Nevers and Whitman 2005, Nevers and Whitman 2008). Other mathematical models have been employed using partial least squares regression (Hou et al. 2006) and neural network models (He and He 2008) for marine beaches. Nowcast models frequently use explanatory variables that can be forecasted with some confidence out to 60 hours. Most of the nowcast models use hydro-meteorological variables such as rainfall, wind direction and speed, and wave height as inputs. The Great Lakes Coastal Forecasting System's hydrodynamic models can be used for beach water quality condition forecasting (Schwab et al. 2006). Combinations of Nowcasting and Forecasting using MLR have been advanced (Frick et al. 2008) showing the possibility of "dynamic" forecasting using a moving data set based on daily beach recreational water quality measurements and meteorological data collected to update an existing model. If there is a tributary in the vicinity of the beach, a potential major source of bacteria could come from tributaries (Olyphant et al. 2003, Byappanahalli et al. 2003, Whitman et al. 2006). Process models involving the integration of river models, hydrodynamic models, and statistical models have been used (Nevers et al. 2007, Holtschlag et al. 2008, Wong et al. 2009) to improve the ability to forecast *E. coli* concentrations.

The transience of fecal coliform bacteria concentrations at the beach, the episodic nature of their introduction, the potential for re-growth, and the expanding understanding of the sources for the *E. coli* indictor bacteria makes the task of predicting safe swimming conditions challenging (Canale et al. 1993, Wilkes et al. 2009). Predictive models are used to inform the swimming public (Francy et al. 2006, Whitman 2005, Pfister 2004) through an internet-based system. Because of cost factors and ease of use, professional judgment models proceeded mathematical models and are still commonly used (USEPA 1999). Budget factors, generally good water conditions in the Great Lakes, and the lack of evidence of swimmer health impacts (there is no swimmer health data base in the Great Lakes) are moving some beach managers towards an always open policy connected with an informational system providing the swimming public *E. coli* concentrations and an evaluation of the health risk associated with the *E. coli* concentration through an Internet-based system. This information system, based on using real-time data correlated to *E. coli* concentrations, provides swimmers notification of elevated bacterial levels above the state standard meets Federal reporting requirements. Closing of swimming is required only under certain conditions such as high waves, lightning, or heavy rainfall releasing sewage into the beach (Breitenbach 2012). There is a weakness in this approach, since the *E. coli* correlations and weather conditions are based on samples collected once per day usually during the week, in the morning before bathers arrive. Application of this model outside of the times when samples were collected is an extrapolation to conditions not present during sample collection times and may not be warranted. Conditions undoubtedly are different during times when swimmers are present. Generally more swimmers are present during the weekend days, which are valued higher by recreational users of beaches.

2. DATA DESCRIPTION AND SOURCES

Factors known to influence the fate and transport of *E. coli* in the Great Lakes are sunlight, rainfall, waves, wind speed and direction, temperature, algae, birds, human bathers, farm animals, and domestic animals. (Boehm et al. 2007). A wide range of explanatory variables has been tested. When these explanatory variables could be forecast, they have been included as possible explanatory variables for the 24 beaches in this study (see Table 2.1). Parameters fall into three categories that are useful for explaining *E. coli* concentrations and can be forecast. Nearshore Beach Conditions (such as wave height, surface and bottom lake current speed and direction, water temperature, lake level), Weather Conditions (such as antecedent rainfall, air temperature, wind direction and velocity, dew point, cloud cover), and Tributary Discharge and Runoff. Data is available for these categories and can be obtained from NOAA–GLERL Great Lakes Coastal Forecasting System's hydrodynamic model, NWS meteorological stations, and USGS stream gauges.

At present, onshore conditions such as turbidity, presence of algae, number and type of birds, number of human bathers, and influence of domestic animals cannot be forecast.

Attachment 2 includes the 35 Excel files containing the hourly data for the Independent Variables used by Virtual Beach 2.3 to arrive at the forecast decision support system equations in Table 5.1.

2.1 Identification of Key Variables

Multi-linear regression models dominate the modeling landscape during the first decade of the 21st century. The determination of the "best" statistical model depends on the metric chosen during the descriptive parameter selection process (Boehm et al. 2007). The parameters retained can be a function of the evaluation process. The mathematical evaluation criteria employed in this study is the Bayesian Information Criterion available in Virtual Beach as one of several evaluation criteria. The decision support system usefulness or success is determined by the accuracy of the management decisions.

Decision support systems have been used at 97 beaches as reported in the literature. Both point and nonpoint sources were present at these beaches. Independent variables were selected from the above sources where studies resulted in equations to predict *E. coli* concentrations. The independent variables included as possible variables in our forecast models (see Table 2.1) were chosen based on work published in Francy and Darner 1998, Francy et al. 2006b, Francy and Darner 2007, Frick et al. 2008, Holtschlag et al. 2008, Przybyla-Kelly et al. 2008, Nevers et al. 2007, Nevers et al. 2009, Nevers and Whitman 2005a,b, Nevers and Whitman 2008a,b, Olyphant 2005, Olyphant and Whitman 2004, Whitman and Nevers 2004, Whitman and Nevers 2006, Vermette et al. 2008, Zimmerman 2006, Zimmerman 2008, and USEPA 2010. Table 2.1 summarizes these variables into several groups. Seven key variables for time were used to observe if Julian date,

Table 2.1. Independent variables and definitions.

Variable	Definition
Date	Time (GMT) at which the sample was taken (to the nearest hour) in the format: mm/dd/yyyy hh:mm GMT = EDST + 4 hours; GMT = CDST + 5 hours; 1-29 minutes round to hour, 30-59 minutes round to next hour)
ECOLI	E. coli measurement at beach sampling site. For Michigan Beaches measurement is a geo. mean of at least 3 samples
Log10Ecoli	Log to base 10 of E. coli
Days	Count of Days from the first sample to the last sample with first sample day = 0
Q1	Categorical Variable = 1 for the First 25% of the sampling season by year
Q2	Categorical Variable = 2 for the Second 25% of the sampling season by year
Q3	Categorical Variable = 3 for the Third 25% of the sampling season by year
Q4	Categorical Variable = 4 for the Fourth 25% of the sampling season by year
QTR	Categorical Variable with 1, 2, 3, and 4 values respectively representing the four quarters of the sampling season
GLCFS Hydrodynamic Model Data Variables	
AT0	Air temp at sample hour, $n \le 1$ (°C)
AT4	Air temp previous 4 hour avg., $n \le 4$, (°C)
AT24	Air temp previous 24 hour avg., $n \le 24$, (°C)
AT120	Air temp previous 120 hour avg., $n \le 120$, (°C)
AT240	Air temp previous 240 hour avg., $n \le 240$, (°C)
DP0	Dew point temp at sample hour, $n \le 1$ (°C)
DP4	Dew point temp previous 4 hour avg., $n \le 4$ (°C)
DP24	Dew point temp previous 24 hour avg., $n \le 24$ (°C)
SWT0	Surface water temp at nearest 3 hour, n=1 (°C)
SWT6	Surface water temp at nearest previous 6 hour ave., n=2 (°C)
SWT9	Surface water temp at nearest previous 9 hour ave., n=3 (°C)
CS0	Surface current speed at sample hour, n=1 (m/s)
CD0	Surface current direction at sample hour, n=1 (degrees towards True North)
ASC0	Alongshore current at sample hour, n=1 (m/s, positive clockwise)
OSC0	Onshore current at sample hour , n=1 (m/s, positive towards beach)
CSb0	Bottom current speed at nearest 3 hour, n=1 (m/s)
CDb0	Bottom current direction at nearest 3 hour, n=1 (degrees toward True North)
ASCb0	Alongshore bottom current at sample hour, n=1 (m/s, positive clockwise)
OSCb0	Onshore bottom current at sample hour , n=1 (m/s, positive towards beach)
WVH0	Wave height at sample hour, n=1 (m)
WVS0	Wave direction at sample hour, n=1 (degrees towards True North)
WVP0	Wave period at sample hour, n =1 (s)
ASWV0	Alongshore Waves at sample hour (m), positive= clockwise rotation
OSWV0	Onshore Waves at sample hour (m), positive towards shore, negative away from shore
CC0	Cloud Cover at sample hour, $n \le 1$ (fraction, 0-1)
CC4	Cloud Cover previous 4 hour avg., $n \le 4$ (fraction, 0-1)
CC24	Cloud Cover previous 24 hour avg., $n \le 24$ (fraction, 0-1)
WS0	Wind speed at sample hour, n=1 (m/s)
WD0	Wind Direction at sample hour, n=1 (degrees from True North)
ASW0	Alongshore wind at sample hour , n=1 (m/s, positive clockwise)
OSW0	Onshore wind at sample hour , n=1 (m/s, positive towards beach)
Nearby Meteorological Station (NMS) Data Variables: (denoted by “m” at the end of variable name)	
AT0m	Air temperature (Dry Bulb Celsius) at NMS at sample hour of measurement $n \le 1$ (°C)
AT4m	Air temperature (Dry Bulb Celsius) at NMS previous 4 hour avg. $3 \le n \le 4$ (°C)
AT24m	Air temperature (Dry Bulb Celsius) at NMS previous 24 hour avg. $18 \le n \le 24$ (°C)
AT120m	Air temperature (Dry Bulb Celsius) at NMS previous 120 hour avg. $90 \le n \le 120$ (°C)
AT240m	Air temperature (Dry Bulb Celsius) at NMS previous 240 hour avg. $180 \le n \le 240$ (°C)

Table 2.1. Independent variables and definitions (cont.).

ATmin24m	Minimum Temperature at NMS (based on minimum of the hourly night time temps) $18 \le n \le 24$ (°C)
ATmax24m	Maximum Temperature at NMS (based on maximum of the hourly night time temps) $18 \le n \le 24$ (°C)
DP0m	Dew Point Celsius from NMS at sample hour, $n \le 1$ (°C).
DP4m	Dew Point Celsius from NMS previous 4 hour avg., $3 \le n \le 4$ (°C)
DP24m	Dew Point Celsius from NMS previous 24 hour avg., $18 \le n \le 24$ (°C)
CC0m	Cloud cover at sample hour from NMS, $n \le 1$ (fraction, 0-1)
CC4m	Cloud cover previous 4 hour avg., $3 \le n \le 4$ (fraction, 0-1)
CC24m	Cloud cover previous 24 hour avg., $18 \le n \le 24$ (fraction, 0-1)
WSP0m	Wind speed at sample time (hour) from NMS, n=1 (m/s)
WSP2m	Wind speed previous 2 hour avg., from NMS,$1 \le n \le 2$) (m/s)
WSP3m	Wind speed previous 3 hour avg., from NMS,$2 \le n \le 3$) (m/s)
WSP4m	Wind speed previous 4 hour avg., from NMS,$3 \le n \le 4$) (m/s)
WG0m	Wind Gust at sample hour from NMS, n=1 (m/s)
WG2m	Wind Gust previous 2 hour avg., from NMS, $1 \le n \le 2$ (m/s)
WG3m	Wind Gust previous 3 hour avg., from NMS, $2 \le n \le 3$ (m/s)
WG4m	Wind Gust previous 4 hour avg., from NMS, $3 \le n \le 4$ (m/s)
WD0m	Wind Direction at sample hour, n=1 (degrees from True North)
ASW0m	Alongshore wind at sample hour from NMS, n=1 (m/s, positive clockwise)
OSW0m	Onshore wind at sample hour from NMS, n=1 (m/s, positive towards beach)
ASWG0m	Alongshore wind gust at sample hour from NMS, n=1 (m/s, positive clockwise)
OSWG0m	Onshore wind gust at sample hour from NMS, n=1 (m/s, positive towards beach)
TP0m	Total precipitation at sample hour at NMS, n=1 (inches)
TP4m	Total precipitation previous 4 hour total at NMS, $3 \le n \le 4$ (inches)
TP24m	Total precipitation previous 24 hour total at NMS, $18 \le n \le 24$ (inches)
TP48m	Total precipitation previous 48 hour total at NMS, $36 \le n \le 48$ (inches)
TP72m	Total precipitation previous 72 hour total at NMS, $54 \le n \le 72$ (inches)
River Discharge Data: “RvrName” is a place holder for any USGS gauged river. Units=Daily Mean Discharge (ft^3/s)	
RvrNameRD_0d	Same Day River Discharge (ft^3/s)
RvrNameRD_1d	Previous Day River Discharge (ft^3/s)
RvrNameRD_2d	Previous 2nd Day River Discharge (ft^3/s)
RvrNameRD_3d	Previous 3rd Day River Discharge (ft^3/s)
RvrNameRD_4d	Previous 4th Day River Discharge (ft^3/s)
RvrNameRD_5d	Previous 5th Day River Discharge (ft^3/s)
RvrNameRD_6d	Previous 6th Day River Discharge (ft^3/s)
RvrNameRD_7d	Previous 7th Day River Discharge (ft^3/s)
RvrNameRD_8d	Previous 8th Day River Discharge (ft^3/s)
RvrNameRD_9d	Previous 9th Day River Discharge (ft^3/s)
RvrNameRO_0d	Runoff day of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_1d	Runoff previous day of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_2d	Runoff two days previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_3d	Runoff three days previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_4d	Runoff fourth day previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_5d	Runoff fifth day previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_6d	Runoff sixth day previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_7d	Runoff seventh day previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_8d	Runoff eighth day previous of sample = Discharge - Baseflow (ft^3/s)
RvrNameRO_9d	Runoff ninth day previous of sample = Discharge - Baseflow (ft^3/s)
NOTE: Averages include sample at time = zero. Four hour average involves previous 3 hours and time of sample.	

calendar date, or seasonal effects were significant for *E. coli* variation. Thirty-one variables are obtained from a nearby meteorological station. These variables are denoted by attaching an "m" e.g. AT0m to distinguish from the same parameter obtained from the GLCFS hydrodynamic model. Of the 31 meteorological variables, the four wind gustiness parameters and five precipitation parameters are not duplicated by the GLCFS. There are 31 GLCFS independent variables. The GLCFS produces readings every hour for all of the parameters. Meteorological stations are not as reliable. In this study, except for the unique meteorological observations, GLCFS variables could be used in place of the meteorological observations without changing the beach management decision outcomes. However, some loss of statistical significance did occur in the Forecast DSS equation. The last group of independent variables is the river discharge and runoff variables for those rivers where USGS maintained a gauging site. Twenty parameters were used. Ten variables were used for river discharge and 10 variables were used for runoff. These variables included the discharge or runoff for the day of the measurement, plus nine lags respectively. The lags were included to see if previous day(s) discharge or runoff affected the beach suggesting the source of bacterial contamination was higher up in the watershed. The lagged variables would represent the additional time needed for transporting the bacteria from more distant sources to the beach water sampling location.

2.2 Beach Locations and Orientations

Each beach's location (Figure 2.2.1) was determined either by the beach's manager supplying the latitude and longitude coordinates for their beach or by manually locating the beach in Google Earth (Table 2.2). Beach Orientation/Beach Angle is determined in accordance with the methodology used by the developers of Virtual Beach and is demonstrated in Figure 2.2.2. The angle (in degrees) follows compass readings.

2.2.1 Alongshore & Onshore Variable Calculations

The variables winds, currents and waves all had alongshore and onshore variables derived from them. The values derived were done using formulae that correspond to those used by the developers of the Virtual Beach Software. Alongshore and onshore winds were determined using the following formulae:

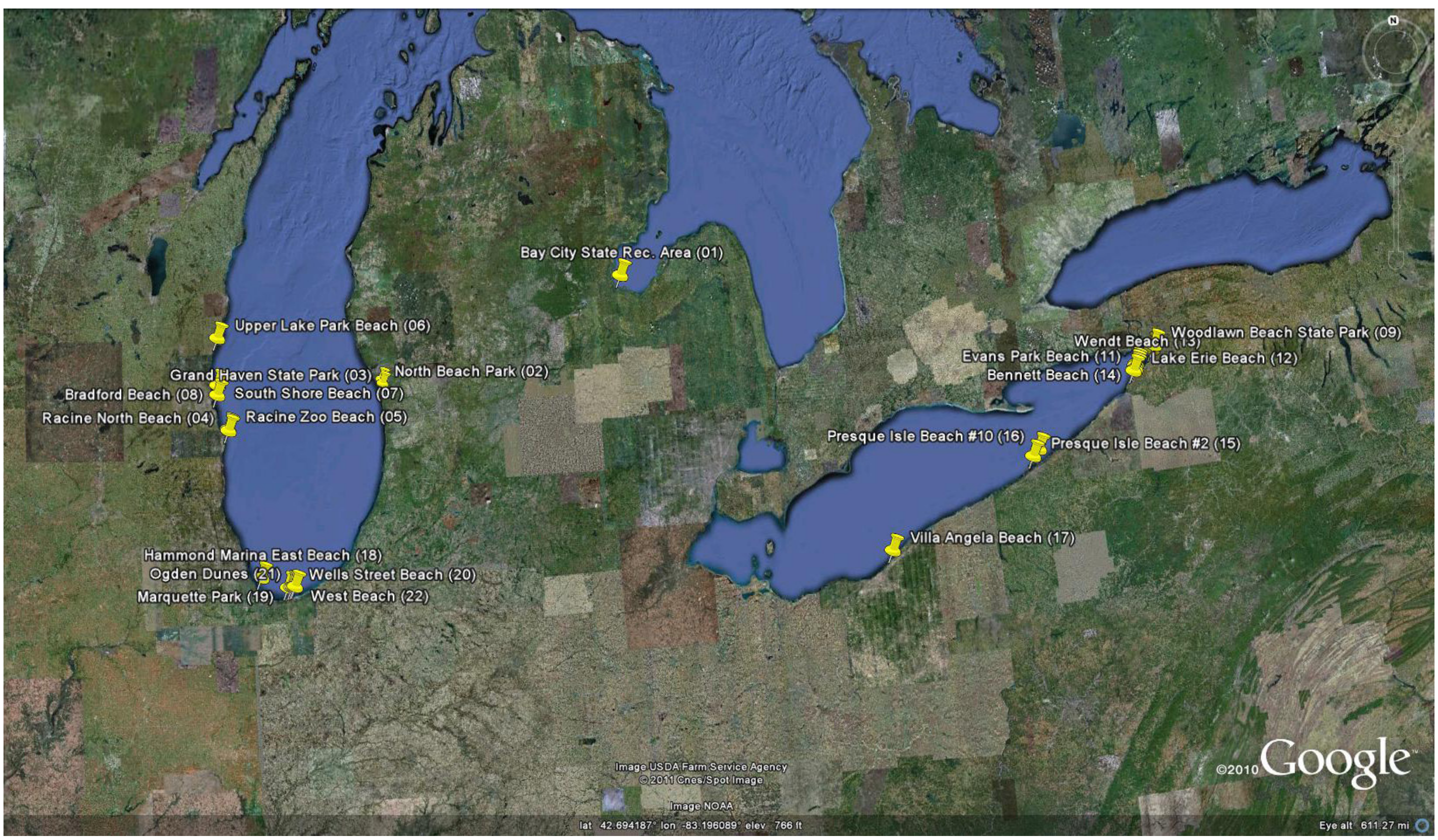


Figure 2.2.1. Beach locations.

$$ASW = -wind\_speed * \cos((wind\_direction - beach\_angle) * (\pi/180))$$
$$OSW = wind\_speed * \sin((wind\_direction - beach\_angle) * (\pi/180))$$

Alongshore & onshore currents were determined using the following formulae:

$$ASC = current\_speed * \cos((current\_direction - beach\_angle) * (\pi/180))$$
$$OSC = -current\_speed * \sin((current\_direction - beach\_angle) * (\pi/180))$$

Alongshore & onshore waves were determined using the following formulae:

$$ASWv = -wave\_speed * \cos((wave\_direction - beach\_angle) * (\pi/180))$$
$$OSWv = wave\_speed * \sin((wave\_direction - beach\_angle) * (\pi/180))$$

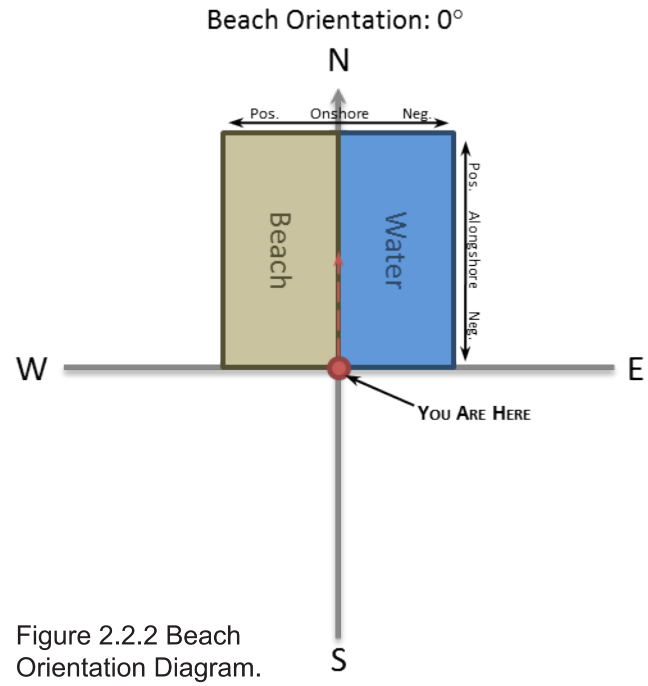


Figure 2.2.2 Beach Orientation Diagram.

Table 2.2. Beach list and information.

60 Hour Beach Forecasting Model: Beach List								
Beach Number	Beach Name	Lake	Beach Orientation	Latitude	Longitude	Primary NMS	Secondary NMS	USGS Gauge Number
1	Bay City State Rec. Area	Huron	-44.71	43.672557	-83.906193	HYX	MBS	04157000
2	North Beach Park	Michigan	165.82	43.081800	-86.255000	MKG		04119000
3	Grand Haven State Park	Michigan	147.60	43.053600	-86.247300	MKG		04119000
4	North Beach	Michigan	-17.53	42.739783	-87.778314	RAC		04087240
5	Zoo Beach	Michigan	-2.62	42.749851	-87.780914	RAC		04087240
6	Upper Lake Park Beach	Michigan	31.39	43.394475	-87.863171	ETB		
7	South Shore Beach	Michigan	-29.05	42.993389	-87.878818	MKE		04087000
8	Bradford Beach	Michigan	37.91	43.060442	-87.873579	MKE		04087000
9	Woodlawn Beach	Erie	-179.07	42.790245	-78.854552	BUF		
10	Hamburg Bathing Beach	Erie	-143.52	42.765617	-78.879825	BUF		
11	Evans Town Park Beach	Erie	-123.39	42.642016	-79.068108	DKK		
12	Erie Beach	Erie	-160.02	42.631319	-79.086144	DKK		
13	Wendt Beach	Erie	-153.16	42.676947	-79.054067	DKK		
14	Bennett Beach	Erie	178.28	42.662876	-79.064441	DKK		
15	Presque Isle Beach #2	Erie	-164.31	42.128252	-80.149890	ERI		
16	Presque Isle Beach #10	Erie	-82.17	42.173621	-80.087003	ERI		
17	Villa Angela	Erie	-140.45	41.586592	-81.566836	BKL		
18	Hammond Marina East Beach	Michigan	-55.22	41.697817	-87.511261	GYY	VPZ	
19	Marquette Park	Michigan	-96.67	41.620769	-87.260131	GYY	VPZ	
20	Washington Park Beach	Michigan	-106.96	41.729003	-86.903352	GYY	VPZ	
21	Ogden Dunes	Michigan	-101.75	41.628791	-87.191905	GYY	VPZ	04095090
22	Indiana Dunes West Beach	Michigan	-117.04	41.662749	-87.064451	GYY	VPZ	
23	Memorial Park Beach	St. Clair	17.49	42.527245	-82.871297	MTC		04164000
24	Metro Park Beach	St. Clair	-251.11	42.571055	-82.796232	MTC		04164000

2.3 Hydrodynamic Data Grid Cell Locations

The hydrodynamic grid cells for each beach were determined programmatically by their proximity to each beach individually. Beaches whose location was within the bounds of a grid cell used hydrodynamic data from the cell in which they were located. Beaches that were not located within the bounds of a grid cell used the closest grid cell to their location. The proximity was determined using the distance formula ($d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$).

A Google Earth '.kmz' overlay of the latest grids for the Great Lakes and Lake St. Clair can be found at: http://www.glerl.noaa.gov/res/glcfs/kml/glgrid-regions.kmz. A Google Earth '.kmz' overlay of the 2006-2008 Lake Huron grid can be found at: http://www.glerl.noaa.gov/res/glcfs/kml/glgrid-regions-5km.kmz. The Beach Forecast Program (see section 3) allows a user to manually select a different grid cell than the one the program selected using arrow buttons to move the grid cell marker on a map that is displayed.

2.4 *E. coli* Data Sources and Data Processing

The *E. coli* data for the beaches within this project was provided by beach managers and local health departments (Table 2.4). To utilize the *E. coli* data provided, some of the data had to be processed. Where an *E. coli* value had a "<" (less than sign) associated with it, the value used was equal to half of the recorded numeric value. Where an *E. coli* value had a ">" (greater than sign) associated with it, the value used was equal to the recorded numeric value.

Table 2.4. *E. coli* data sources.

Beach No.	Beach Name	Source of *E. coli* Data Name	Organization
1	Bay City State Rec. Area	Robert Hill	Bay County Health Department
2	North Beach Park	Adeline Hambley	Ottawa County Health Department
3	Grand Haven State Park	Adeline Hambley	Ottawa County Health Department
4	North Beach	Adam Mednick	City of Racine Health Department & Lab
5	Zoo Beach	Adam Mednick	City of Racine Health Department & Lab
6	Upper Lake Park Beach	Adam Mednick	Ozaukee County Public Health Department
7	South Shore Beach	Adam Mednick	City of Milwaukee/Wisconsin Water Science Center
8	Bradford Beach	Adam Mednick	City of Milwaukee/Wisconsin Water Science Center
9	Woodlawn Beach	Kristen Husson	New York State Office of Parks, Recreation, and Historical Preservation
10	Hamburg Bathing Beach	John Finster	Erie County (NY) Department of Health
11	Evans Town Park Beach	John Finster	Erie County (NY) Department of Health
12	Erie Beach	John Finster	Erie County (NY) Department of Health
13	Wendt Beach	John Finster	Erie County (NY) Department of Health
14	Bennett Beach	John Finster	Erie County (NY) Department of Health
15	Presque Isle Beach #2	Dr. Doug Range	Erie County (PA) Department of Health/Presque Isle State Park, PA
16	Presque Isle Beach #10	Dr. Doug Range	Erie County (PA) Department of Health/Presque Isle State Park, PA
17	Villa Angela	Mark Citrigilia	Northeast Ohio Regional Sewer District/USGS Real Time Database
18	Hammond Marina East Beach	Michelle Caldwell	Indiana Department of Environmental Management
19	Marquette Park	Michelle Caldwell	Indiana Department of Environmental Management
20	Washington Park Beach	Michelle Caldwell	Indiana Department of Environmental Management
21	Ogden Dunes	Michelle Caldwell	Indiana Department of Environmental Management
22	Indiana Dunes West Beach	Michelle Caldwell	Indiana Department of Environmental Management
23	Memorial Park Beach	Steve Lichota	Macomb County Health Department
24	Metro Park Beach	Steve Lichota	Macomb County Health Department

2.5 Hydrodynamic Model Data Sources

The hydrodynamic model data was obtained from two online locations. Data from 2002-2005 for Lakes Erie, Huron, Michigan, Ontario and Superior and for 2007-2010 for Lake St. Clair was collected from the GLERL website located at http://www.glerl.noaa.gov/res/glcfs/gridded_fields/. Data from 2006-present (except for Lake St. Clair), was accessed through GLOS's point query tool for the Great Lakes Coastal Forecasting System (http://data.glos.us/glcfs/). Lake St. Clair data is scheduled to be available through GLOS's web tool in the near-future.

2.6 Meteorological Data Sources

All meteorological data is retrieved from the National Climatic Data Center (NCDC). The data is collected through NCDC's ftp site located at ftp://ftp.ncdc.noaa.gov/pub/data/noaa/. The selection of meteorological station(s) for a given beach was accomplished by calculating the closest surface airway stations as determined by the distance formula ($d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$). The availability of key variable data at the closest surface airway stations then narrowed the list for a given beach to a single station. In some cases, a secondary station was identified. A secondary station was used for locations where the primary station had insufficient or erratic precipitation data, in these cases the secondary station was only used for precipitation data.

2.7 USGS River Gauging Station Discharge Data Sources

Daily mean river discharge values are acquired from USGS through their water data website located at http://waterdata.usgs.gov/nwis/.

2.7.1 River Runoff Computation

The values for river baseflow are calculated using a USGS computer program developed by A.T. Rutledge called PART. The PART computer program uses a method of determining baseflow that is based on antecedent streamflow recession. Runoff is calculated by subtracting the calculated baseflow values from the river's daily mean discharge.

3. BEACH FORECAST PROGRAM

To aggregate the data needed to create the models for each beach using Virtual Beach, a graphically driven program was written. The Python programming language was used to write the scripts and Tkinter was used in conjunction with Python to provide the graphical user interface. Python and Tkinter were chosen because they are both freely downloadable and distributable tools (Figure 3.1).

4. VIRTUAL BEACH 2.3 SOFTWARE

Virtual Beach 2.3 (VB2) series is a software package designed to construct beach sampling site-specific beach decision support tools using multiple linear regression. These equations correlate *E. coli* measurements with independent variables (IVs) measured at the beach or from nearby meteorological stations, river gage sites, or Great Lakes Coastal Forecasting Hydrodynamic cells. The VB2 modeling interface is designed to help find the best model amongst a large number of candidate models, based on criteria selected by the user. As the number of IVs increases, the number of possible models in the solution space increases exponentially. The user is able to select all, or a subset of, the IVs for consideration in the model to reduce the size of the solution space. For complete user instructions, Virtual Beach User guide (Cyterski et al. 2012) is found on the web at http://www.epa.gov/ceampubl/swater/vb2/Virtual Beach 2 User Guide.pdf.

5. BEACH WATER QUALITY DECISION SUPPORT SYSTEM (DSS)

The technical and scientific merit of the beach water quality decision support system is based on well-known observations that general seasonal, weather, and hydrological conditions greatly influence the physical, chemical, and biological characteristics of large water bodies such as the Great Lakes. These factors in turn affect the occurrence, distribution,

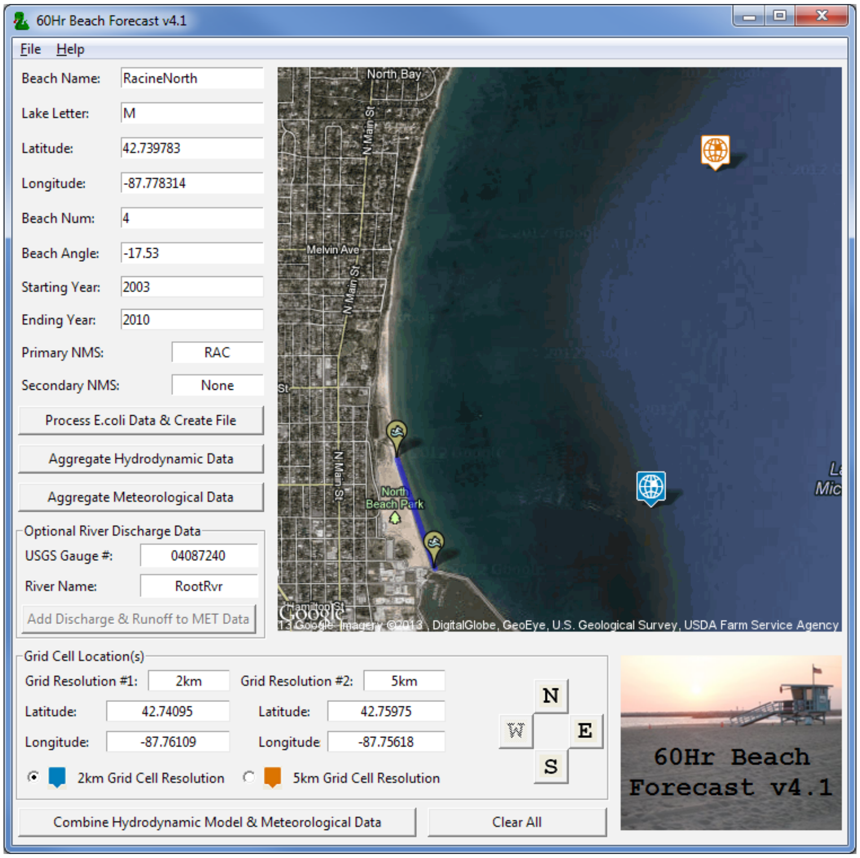


Figure 3.1 Beach forecast program screenshot.

and survival of microbiological contaminants in the water. Forecast models using data from deterministic models (hydrodynamic and meteorological) quantify influences that are common to a geographical region and at no additional analytical expense to the beach manager.

The development and application of this concept by Whitman and Nevers (2004) demonstrated in a regional model from Milwaukee, WI to Indiana Dunes State Park, IN of Southern Lake Michigan that many beaches in a common geographical area respond similarly. These large-scale patterns cause simultaneous fluctuations in fecal indicating bacteria at beaches throughout a region (Nevers and Whitman 2008). Predictive nowcast models frequently include IVs such as current speed and direction, wave height, sunlight, and rainfall (Francy et al. 2006, Francy 2009, Olyphant 2005, Nevers and Whitman 2005); all of which can be forecasted.

The following description uses Indiana Dunes State Park West Beach (IDSP West) data to demonstrate the process used to build a multiple linear regression model for making forecasts predicting *E. coli* using VB2.

The IDSP West 2006-2009 data (Attachment 2) was assembled as described in section 2. The data processing required to arrive at a typical forecast DSS equation follows. All Table 2.1 IVs were computed hourly for each *E. coli* measurement using the nearest hour IVs value. Sample times between 1-29 minutes were rounded back to the hour. Sample times between 30-59 minutes were rounded to next hour. All local times were converted to GMT using GMT=EDST+4 hours and GMT=CDST+5 hours. The recommended minimum number of samples in a training data set is 100 or more samples obtained over a preceding time interval of at least one year. However, data was assembled over years assuming beach environmental conditions were stable.

To load the data set into VB2, open VB2 and use the Data Processing Tab to import the data file containing the IVs which have been assembled for the beach *E. coli* sampling times. These values represent the training data set for the forecast DSS. Typically, they represent measurements from one or more years prior to the year for which the forecast will be made.

Validate the imported data set and correct any errors detected. VB2.3 identifies categorical variables which cannot be transformed and automatically disables them. The response variable (LOG10Ecoli) is in the second column. Sample times for the *E. coli* measurements are in column one. The second column is a LOG10 transform of *E. coli* and VB2 requires manual identification of the transform by right clicking on column 2 and selecting the appropriate transform being used. Within define transform, highlight transform needed, which in this case is log10 and hit enter key. See Figure 5.0.1. Note ECOLI is highlighted in red. VB2 denotes disabled columns by changing the color of the column of numbers from black to red.

In preparing variables for transformation, the next step is to disable IVs which would be correlated. In this study, these variables included *E. coli* (disabled in above screenshot as noted by red numbers) which would be correlated with LOG10Ecoli. IVs correlated with onshore and along shore components of the surface and bottom currents, wind, and wave were CD0, CDb0, WVD0, WD0, and WD0m. QTR was disabled because individual Q1, Q2, Q3, and Q4 categorical variables were retained. WVP0 was disabled.

Meteorological variables duplicate the Great Lakes Coastal Forecasting System (GLCFS) hydrodynamic model except for precipitation and gustiness variables. In some cases, individual meteorological data sets have instrument failures resulting in missed hourly readings. Meteorological variables representing faulty readings or those with fewer than 15% real numbers were disabled if a GLCFS variable could be used instead. VB2 will automatically disable some variables that have one value. In this example, eleven meteorological parameters were disabled. Altogether, 30 meteorological IVs were disabled.

In general, wind gustiness and precipitation variables are not disabled, because the hydrodynamic deterministic model cannot provide these measurements. However, instrument failure represents a challenge to incorporate wind gustiness and precipitation variables in the forecast DSS as occurred in this example.

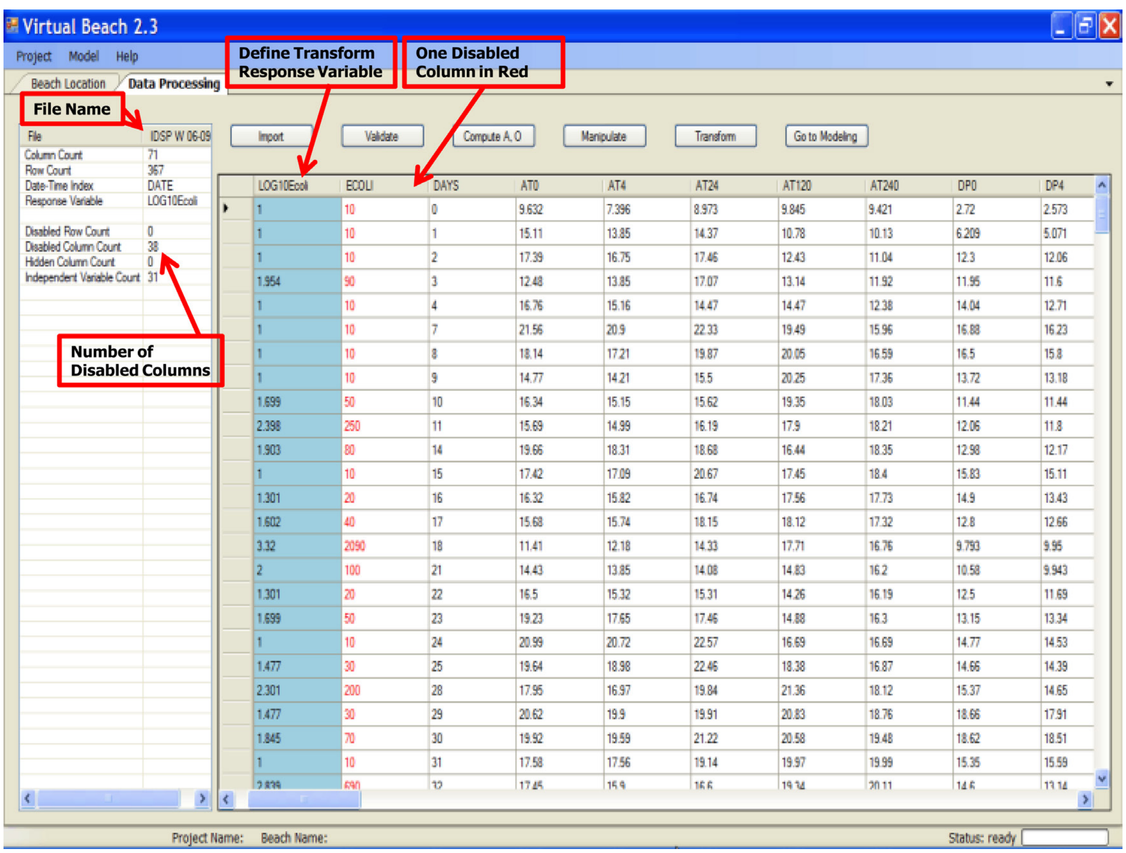


Figure 5.0.1:VB2.3 Data Processing Screenshot

Second order or interaction terms are possible using the Manipulate Tab. VB2 provides the user an option to include two-way interactions between IVs. This study did not employ this option (Figure 5.0.1).

This study employed three transforms when using the Transform Tab. The transforms used were square, square root, and polynomial. Other transformations are available (such as Log10, Natural Log, and Inverse) were not used in this study. Zero sample values would result in a singularity in these transforms. VB2 avoids these transform singularities by adding a small value to a zero sample measurement.

VB2 uses a 20% threshold as a default value for the Transform Tab. The 20% transform was rerun by selecting the 40% threshold as shown in Figure 5.0.2. A transformed variable was selected if the Pearson coefficient exceeded the un-transformed variable by 40%. In Figure 5.0.2, no transform of OSWO exceeded the un-transformed variable by 40%, but the POLY transform of OSWO was selected to have the same transform for both ASWO and OSWO. This process was applied for all grouped variables so that a single transform was used to aid in interpretation of the forecast DSS.

The 40% threshold (r=.4) provides a stronger linear relationship with the transformed variable and minimized the number of transformations (Cyterski, M personal communication). This choice is arbitrary. This effect size variance represents at least 16% (r2=.16) of the dependent variable variance is attributable to the independent variable (Cohen, J. 1988). The goal was to seek stronger linear relationships with the transformed variables and to minimize the number of transformations. This would allow for ease in interpretation of the forecast DSS equation.

Groups of independent variables (e.g. RvrNameRD_0d, RvrNameRD_1d, RvrNameRD_2d…) may have more than one transform. To facilitate interpretation of the decision support system, only one transform was selected for such a data group. The transform that had the highest average Pearson coefficient was chosen. This was illustrated in Figure 5.0.2.

Figure 5.0.3 shows the Data Processing Tab after the threshold transformation is completed. Note VB2 allows for inclusion of the variable and the transformed variable as illustrated by the DAYS and POLY(DAYS) variables. The POLY(DAYS) variable is one of eight transforms added to the data set which will be used in the MLR model. Note the column count in Figure 5.0.1 was 71. Figure 5.0.3 has a column count of 79 reflecting the 8 additional transforms included in the IV data set.

Click on the Go To Modeling Tab in ribbon above data located on the right hand side of Figure 5.0.3. Figure 5.0.3 shows the number of IVs available for the MLR. There were 79 IVs (columns) – 38 disabled IVs (columns) – Dependent Variable (LOG10Ecoli) - Time Variable (Date) leaving 39 IVs available for multiple linear regression. In the Modeling Tab, all the transformed and un-transformed variables obtained from the Data Processing Tab are selected for multiple linear regression evaluation and are moved into the Available Variables (under the Variable Selection sub-tab).

The evaluation criteria used in this study was the Bayesian Information Criterion (BIC). This control criterion was selected from eight possible options in the pull down screen. BIC is based, in part, on the likelihood function, and it is closely related to Akaike information criterion (AIC). When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in over fitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC (Schwarz, G. E. 1978). This is expected to minimize the number of key variables identified in the forecast DSS equation. In addition to specification of the evaluation criterion, the selection of the Genetic Algorithm is needed. Figure 5.0.4 shows the layout of the Modeling Tab.

In Figure 5.0.5, the Genetic Algorithm was selected because the possible models for most beaches exceeded 9.2e18. In this case the number of possible models is 5.5e11. The Decision Criterion (235 or 300 counts/100 m) and State Regulatory Standard (235 or 300 counts/100 m) appropriate for each state was used. Default control values available in VB2 were accepted. Default values were used for VIF (5), mutation rate (0.05) and cross over rate (0.5). This study set the seed value to 1 to fix the random number generator. The population number was set to 5000. The number of generations was

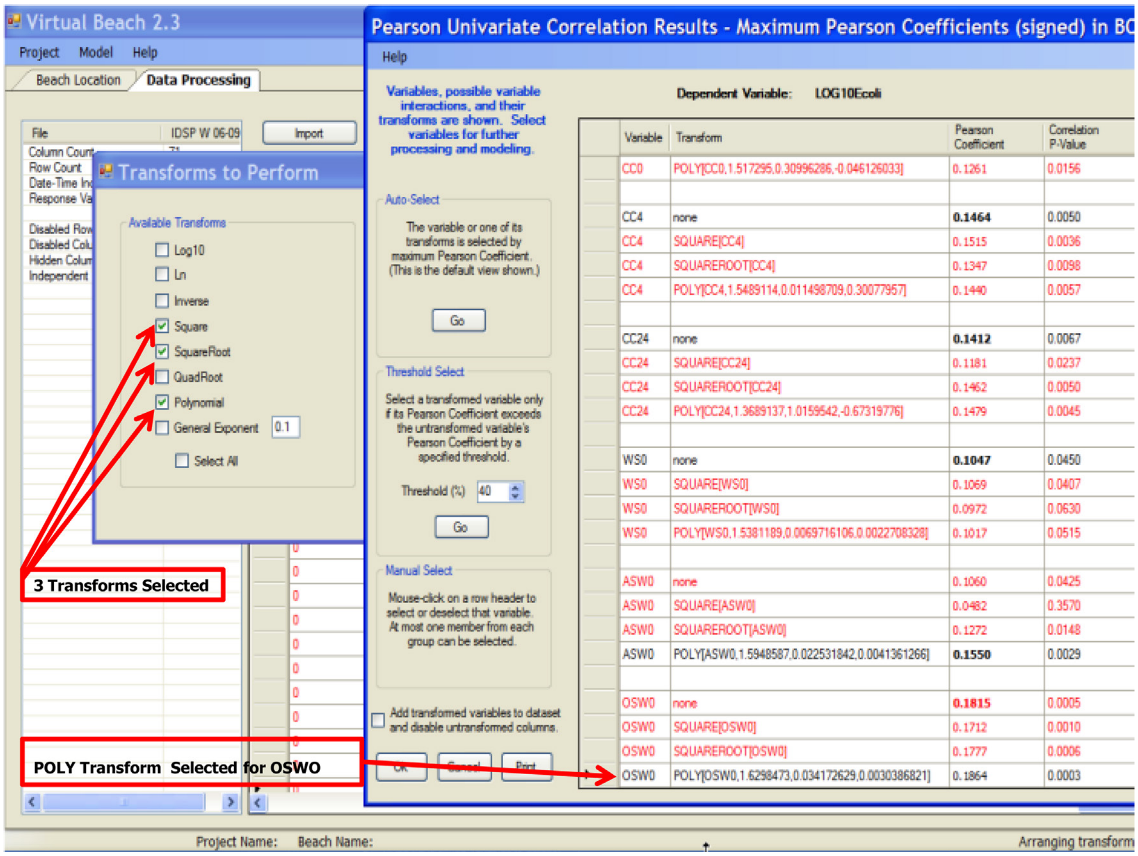


Figure 5.0.2: VB2.3 Transform Screenshot

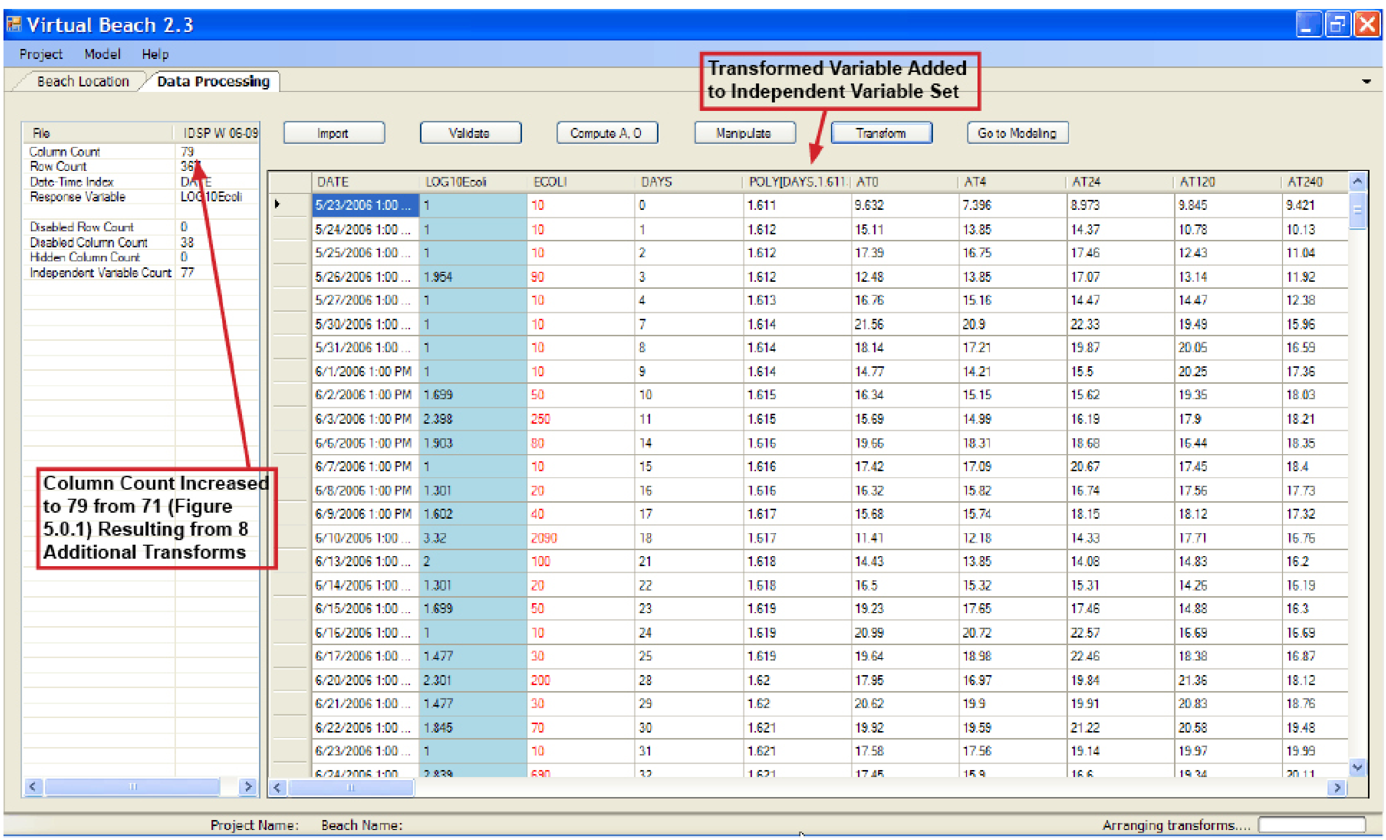


Figure 5.0.3: VB2.3 Data Processing Screenshot after Completion of Transform Variable Selection

set at 50. VB2 requires a minimum of 25 generations. It was observed that additional generations often did not improve the value of the evaluation criteria, but increased the time to complete the MLR run. Typical run times for the program was 1-2 hours on a HP Intel 2 Duo CPU P8400 running at 2.26 GHz with Microsoft Windows XP Professional Version 2002 Service Pack 3.

Upon completion of the Genetic Algorithm run, click on Clear List Tab to remove all IVs. Highlight each of the 10 BIC values representing the best equations in the Model Information Best Fit screen, one at a time and click Add To List Tab. This will accumulate all the unique IVs in the ten best fit equations in the Independent Variable Screenshot (Figure 5.0.6). Note the BIC value was unchanged after six generations in the Forecast DSS for Indiana Dunes State Park West Beach.

Figure 5.0.7 shows the Modeling Manual Tab and Cross Validation screenshot. The resulting number of IVs will generally allow the use of the Manual Tab to search all the possible models for the ten best fits available in the reduced set of IV obtained from the Genetic Algorithm. Check box for Run all combinations and click on Run Tab.

This will result in another set of ten best fit equations. Run the cross validation step by clicking on Cross Validation Tab. In this study, 25% of the samples were held out for this procedure, rounding to the nearest whole number if the number of samples were not divisible by 4. The number of trials in this study was set at 5000. On repeats of the cross validation step, 5000 trials resulted in less variation in the values for MSEP when repeated runs were attempted.

In the Cross Validation screen, the 10 resulting equations can be ranked by MSEP and number of variables. The best fit equation with the most variables has the lowest MSEP. In this study, the equation with the minimum MSEP and most variables was selected as the best Forecast DSS. This selection is highlighted in blue in the left most column (fitness) of the Cross Validation screen above. The equation with the fewest variables and minimum MSEP was also selected for this study. This choice was made to keep the Forecast DSS as simple as possible. However, one of the 34 sites had the same number of variables for all 10 best fit equations. In this case, the equation with the maximum MSEP was selected to represent the second choice.

The “fitness” column is the value of the evaluation BIC criterion chosen in this study. This allows identification of the model in the Modeling Tab. The MSEP column stands for the “mean squared error of prediction.” Across all the random cross-validation trials (in this case, 5000), it summarizes the average squared predictive error for all of the testing data. It is calculated by taking the difference between the actual observation and the model prediction, square the differences, sum them up for all testing data set observations, and then divide by the number of observations to obtain the mean value. The equation with the smallest MSEP did the very best at predicting “new” observations – that is the 25% of the observations not included in fitting the model’s regression coefficients. Models with the largest MSEP did the worst at making predictions. Cross-validation is a means to choose the “best” model to be used in prediction (Figure 5.0.7).

Return to the Modeling Tab we get Figure 5.0.8 by highlighting the BIC value from the Cross Validation screen. Select the variable statistics tab and check all IV’s P-values. Note in the Variables Statistics screen, all IV’s have statistically significant P-values. In this study when one of the IVs did not have a statistically significant P value, the IV was removed and the above steps rerun to get a best fit model with all IVs having significant P values (Figure 5.0.8).

The Residuals Tab allows you to evaluate the normality of the residuals resulting from the model. Note Figure 5.0.9 has the standard cutoff set at 0.65. This choice was made early in VB2 development. A p/n ratio of 0.1 (10 observations for each parameter) was chosen as reasonable. If p/n is 0.1, then $2*\sqrt{(p/n)}=0.632$ and was rounded to 0.65. If $2*\sqrt{(p/n)}$ becomes too small, it will often result in too many observations being removed from a normally distributed residuals dataset. This happens more frequently when the p/n ratio drops below 0.1.

The Anderson Darling statistical test tests for normality of the residuals can be found by clicking the Predicted vs. Residuals tab located in the left most position of the ribbon located to the left of the residuals plot. Anderson Darling (A.D.) statistical test <0.05 indicates the distribution of studentized residuals is not normal. See Figure 5.0.10.

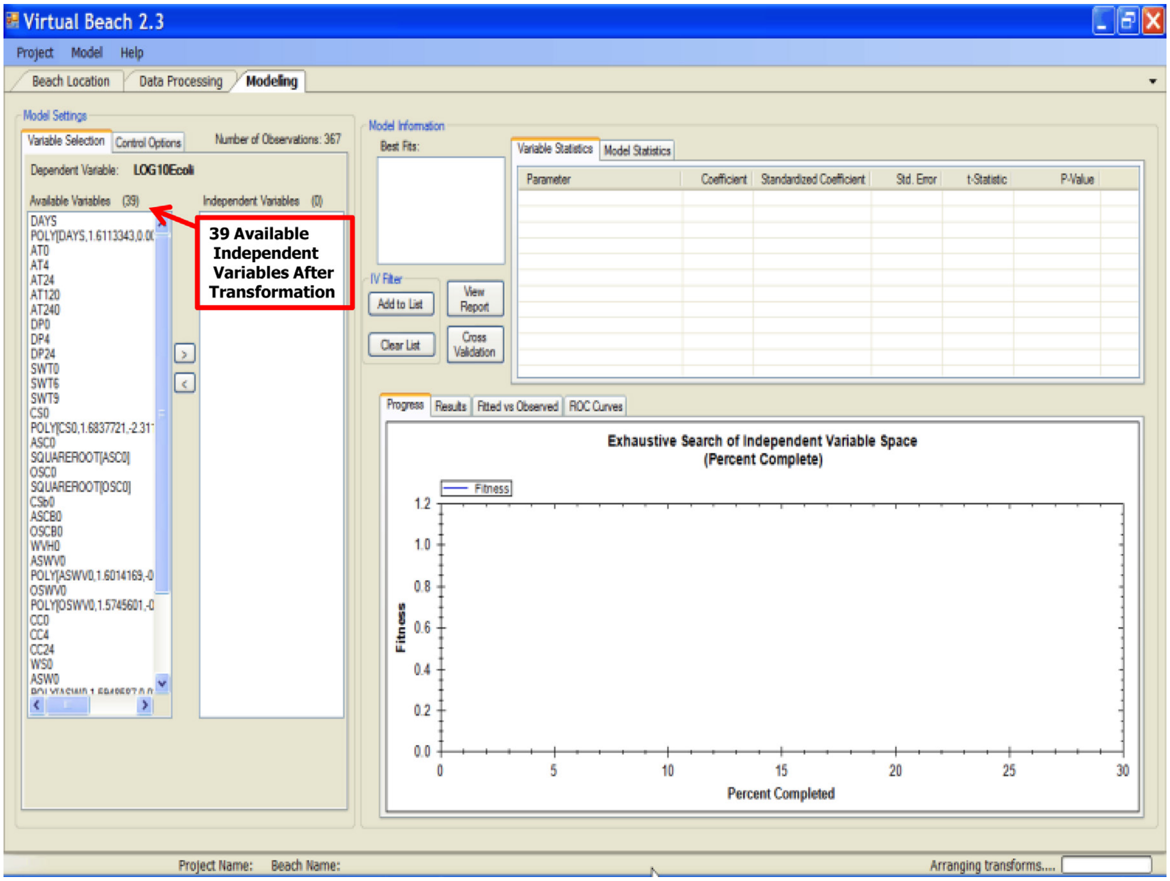


Figure 5.0.4: VB2.3 Initial Modeling Screenshot

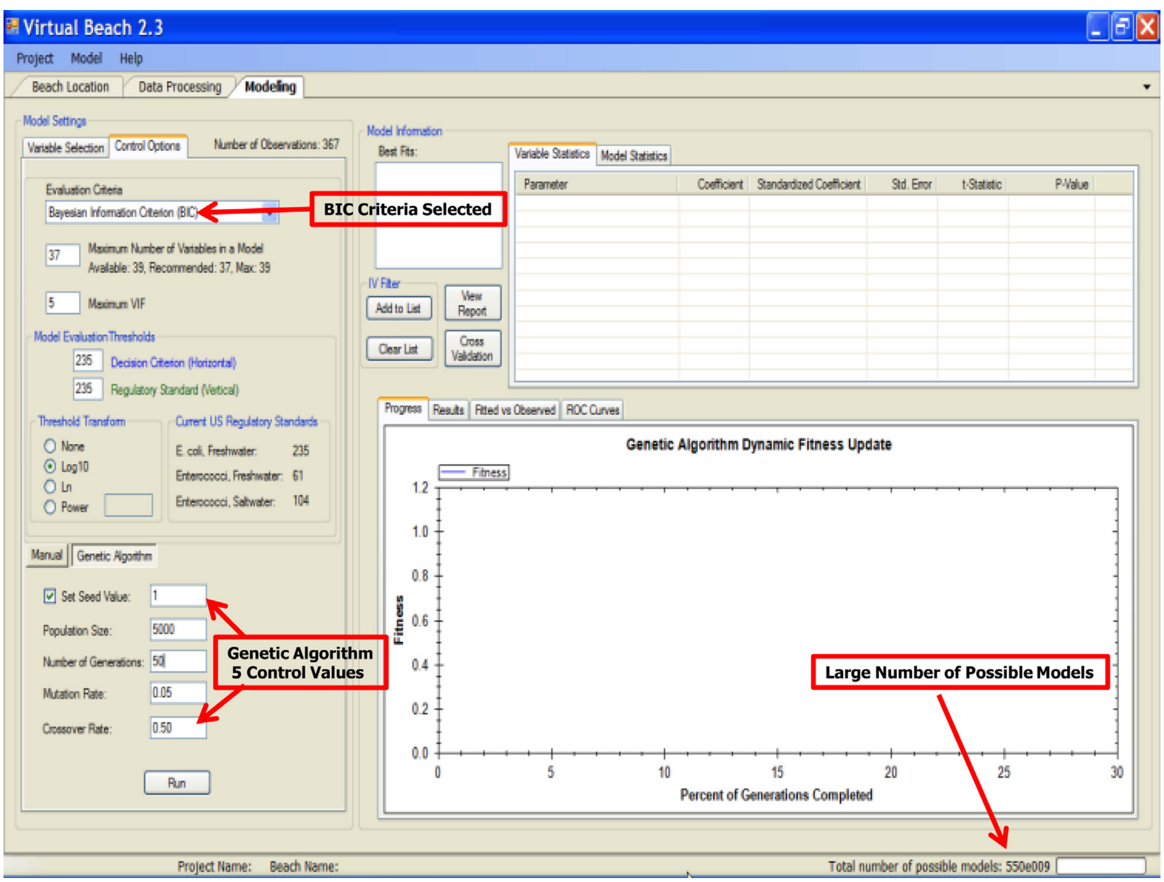


Figure 5.0.5: VB2.3 Modeling Genetic Algorithm Screenshot

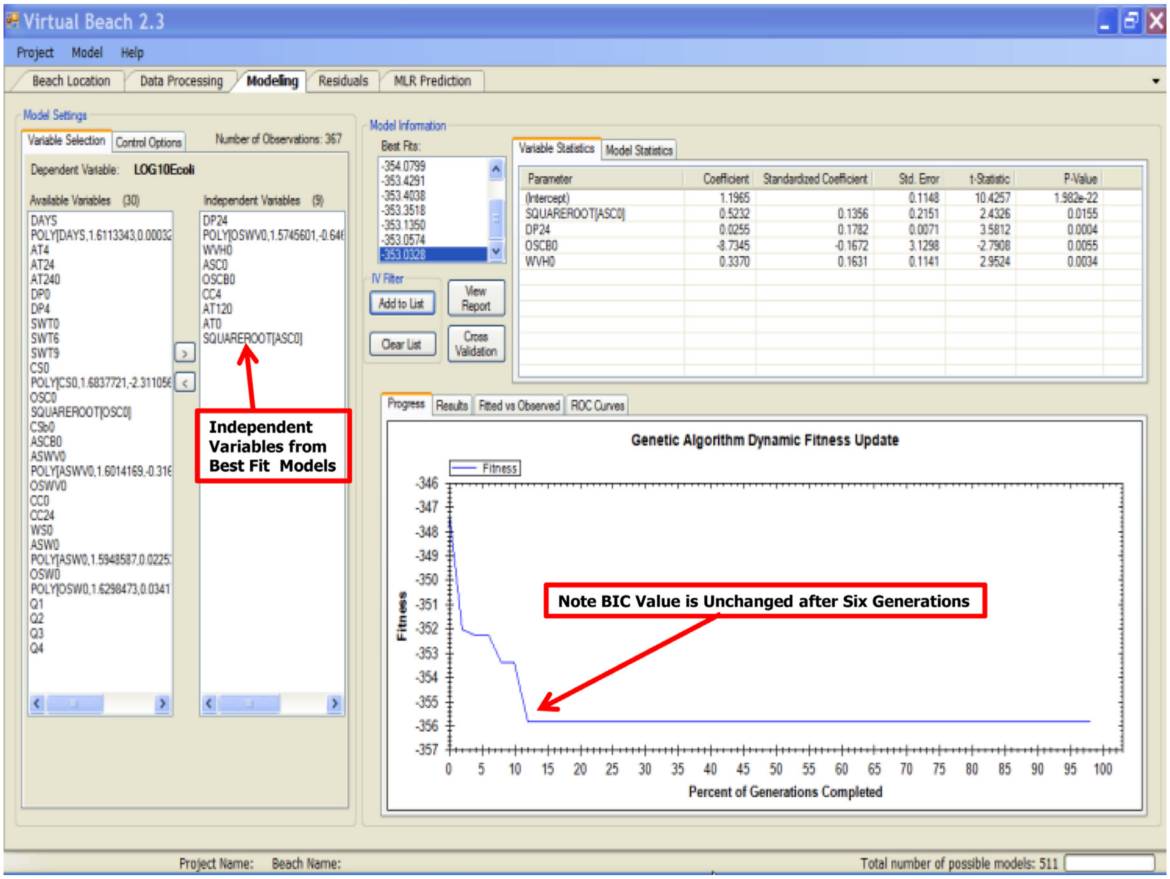


Figure 5.0.6: VB2.3 IVs from Ten Best Fit Genetic Algorithm Models Screenshot

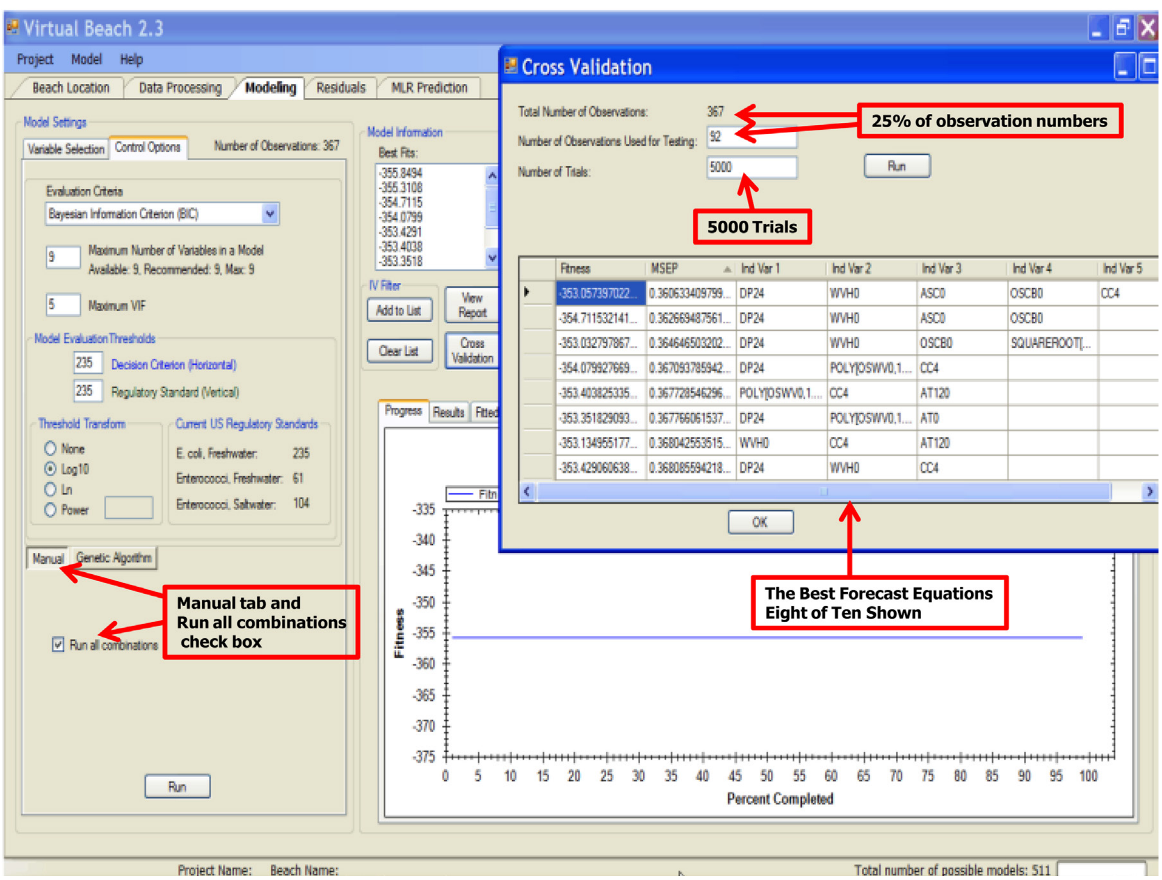


Figure 5.0.7: VB2.3 Modeling Manual Tab and Cross Validation Screenshot

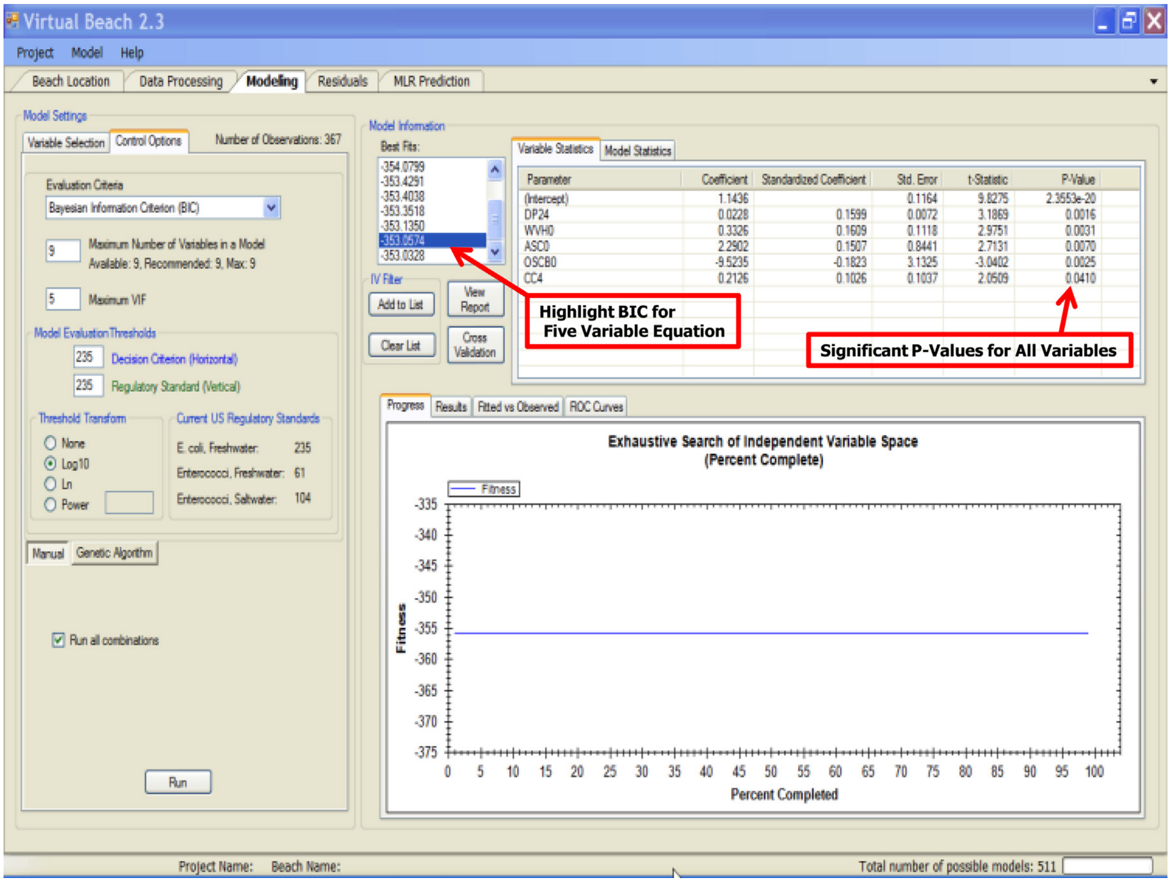


Figure 5.0.8: VB2.3 Display of FDSS Equation Variabale P-Values Screenshot

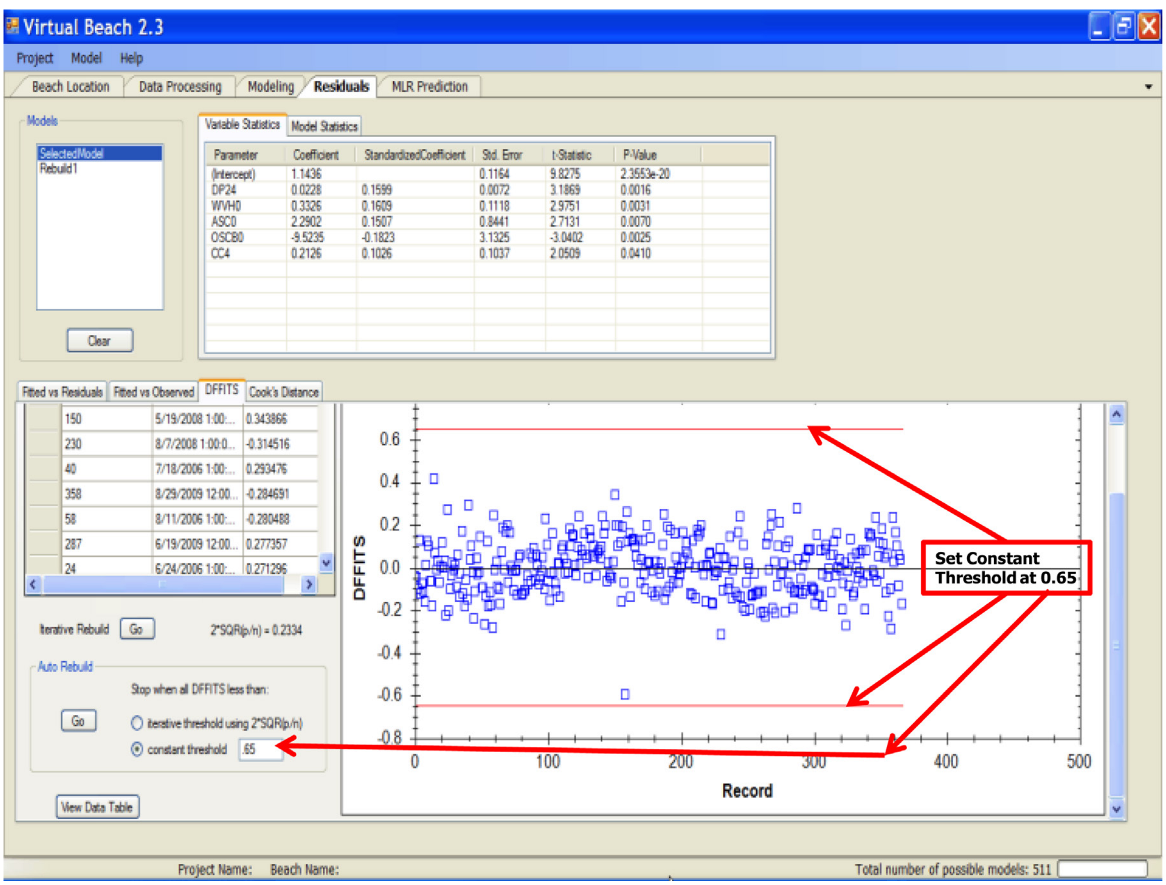


Figure 5.0.9: VB2.3 Display Model of DFFITS Residuals Screenshot

For large data sets, which are typical in this study, the A.D. statistic is sensitive to minor variations in the normality of the residuals (Cyterski personal communication) and frequently evaluates the residuals as non-normal. To meet this criterion would result in too many samples being discarded by applying the iterative threshold options. However, the loss of more than 5% of the samples is not recommended to attain normality as measured by the A.D. statistic. The approach applied in this study is to remove outliers that have DFFITS absolute values greater than 0.65 and run the rebuild model resulting from the outlier removal.

To find the optimum Decision Control (Horizontal) (DCH) setting that minimizes the total number of errors and the number of false negative errors (Type 2) in the training data set, the following procedure is used (Francy et al. 2006b).

A plot of the predictions vs. the observations for the training data is obtained by clicking the Fitted vs. Observed tab located in the second left most position of the ribbon located to the left of the residuals plot. Doing this replaces the predictions versus the Studentized Residuals plot with the Fitted vs. the Observed in the next screenshot (Figure 5.0.11).

Note the total number of errors are 58 with 56 type 2 errors using DCH=RSV=235. The accuracy of the model is 84.3%. By adjusting the DCH to lower values we can reduce the number of type 2 errors at the risk of increasing type 1 errors. This process will not improve the accuracy of the model if the results are a one for one trade off. Exposure to higher bacterial concentrations increases the risk of greater human health problems. The trade-off between not swimming in water with higher bacterial concentrations for not swimming in water when bacterial levels are below the state regulatory standard is a tradeoff that county health departments favor despite the loss of economic benefits when swimming is not permitted.

In this case you can observe in Figure 5.0.11, a cluster of type 2 errors near the DCH=235. By selection of DCH=155 we find the cluster of 4 false negatives errors (type 2) are located above the DCH line without incurring an increase in the number of false positive errors (type 1). By re-plotting the Fitted vs. Observed with the DCH=155 we get Figure 5.0.12. Note the accuracy of the model has increased to 85.3% with the reduction in total errors to 54.

Having found the optimum DCH for the training data, we move to the MLR Prediction tab to apply the model developed from the 2006-2009 training data (Figure 5.0.13).

Using IDSP-West Beach 2010 data we can see in Figure 5.0.13, the Forecast DSS equation under consideration in the Model text box at the top. You can navigate to Figure 5.0.13, by clicking on the MLR Prediction tab.

The IVs have been imported. The 2010 *E. coli* observation data has been imported. After passing VB2's required data validation step, prediction of model values to compare with observations can be done. First the DCH is set to 155 based on the optimization determined by the training data. If the environmental conditions in effect during the training data years have not been altered, the predictions developed from the model should compare well with the observations taken in 2010.

The next step is to click the Make Prediction tab. Note the IVs from which the predictions are made are displayed on the left. In the middle are the observations. On the right side of Figure 5.0.13 are the model predictions and evaluations of the type of errors resulting from the model predictions.

The last step is to plot the model predictions verses the observations. Figure 5.0.14 illustrates the results for the 2010 observations. The accuracy of the forecast model in this case was 81.1% which is slightly below the training data. This is a usual outcome since the validation data set contains new variability which the training dataset did not contain.

The persistence model for 2010 would have had 28 errors evenly divided between type 1 and type 2 errors. The forecast model reduced the number of type 1 errors by 13 and increased the number of type 2 errors by 5. The accuracy of the persistence model is 78.8%. The forecast DSS accuracy is 81.1% with a total of 20 errors.

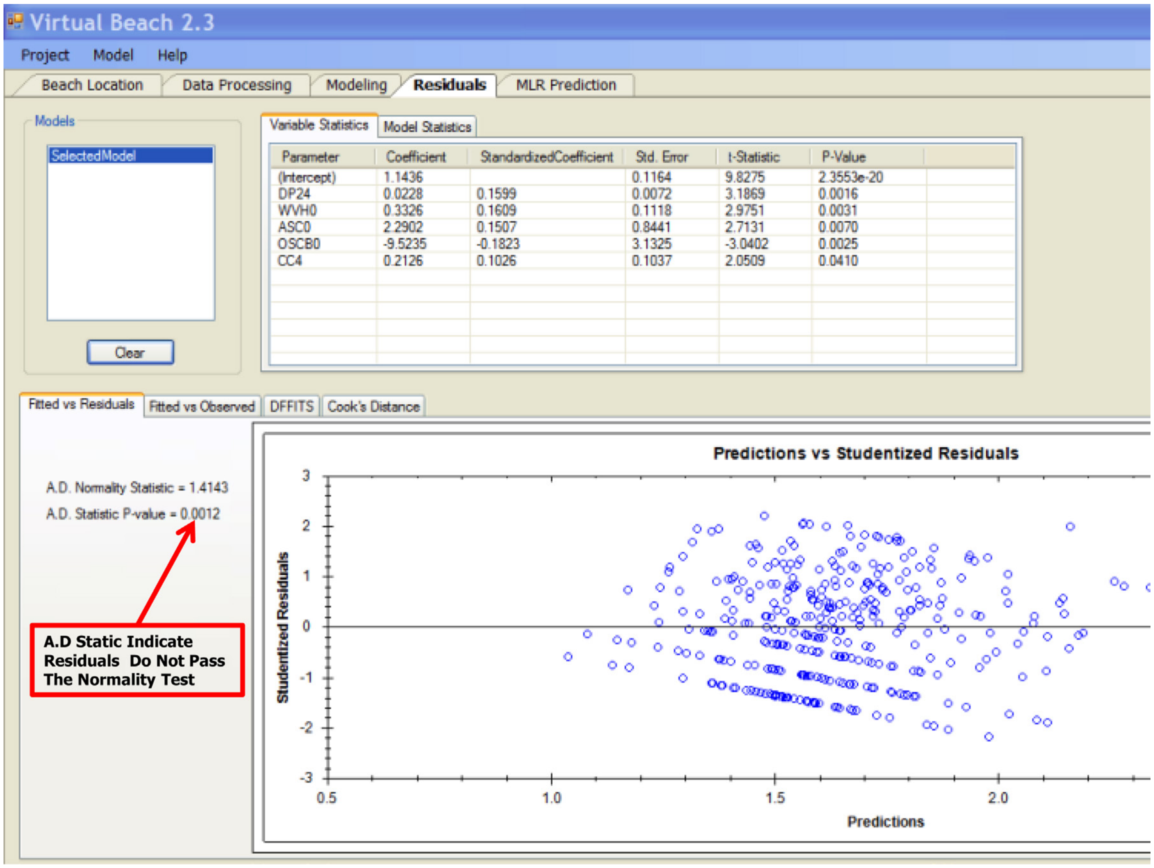


Figure 5.0.10: VB2.3 Residuals, Fitted vs Residuals, Anderson Darling Statistic Screenshot

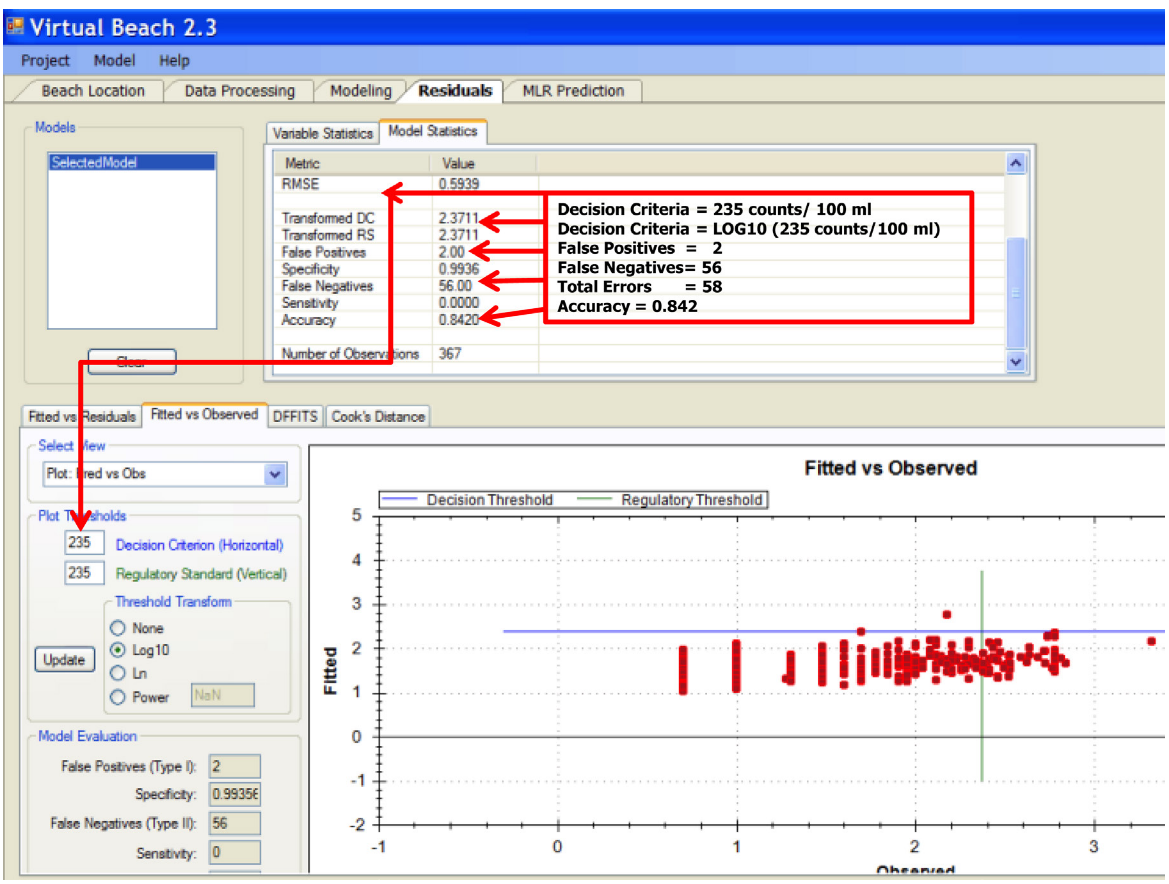


Figure 5.0.11: VB2.3 Residuals Fitted vs Observed, Model Statistics, Total Errors Screenshot

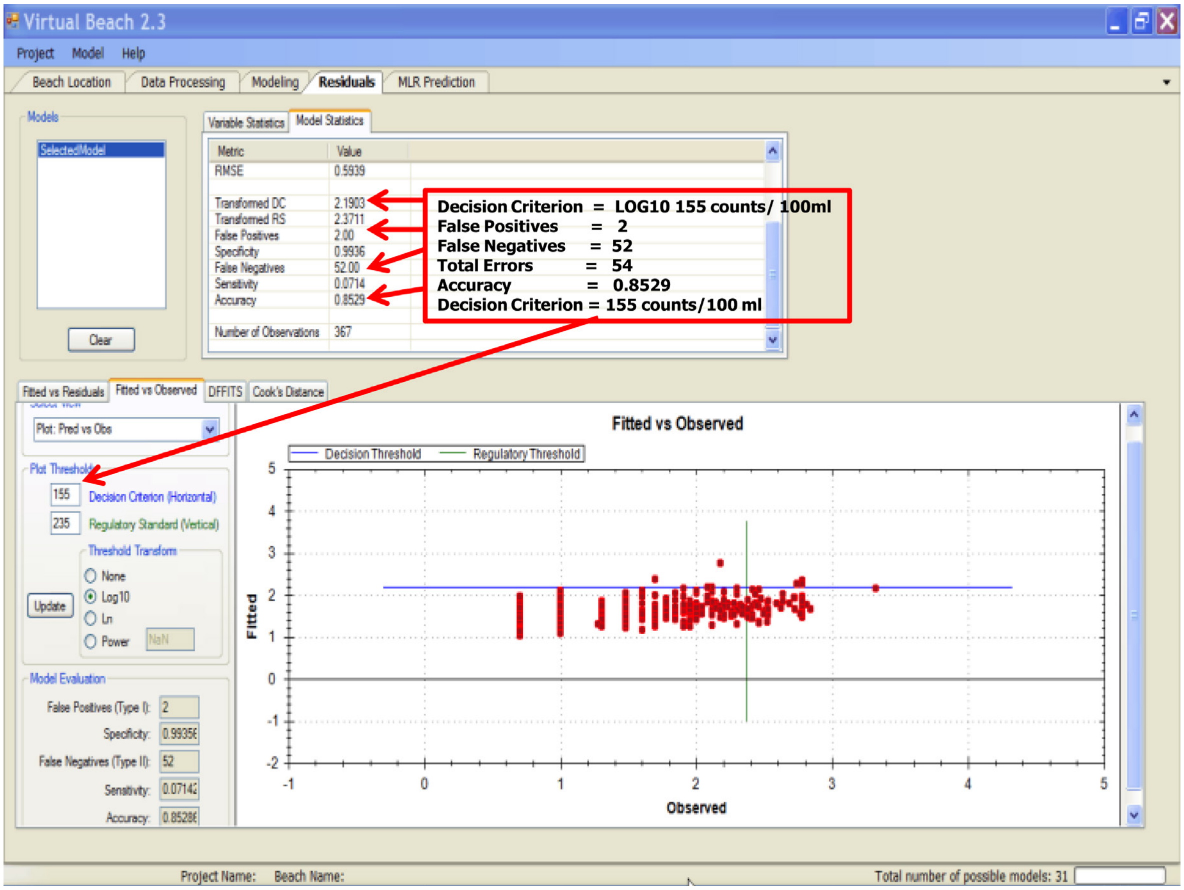


Figure 5.0.12: VB2.3 Residuals Fitted vs Observed, Optimize Decision Criterion Screenshot

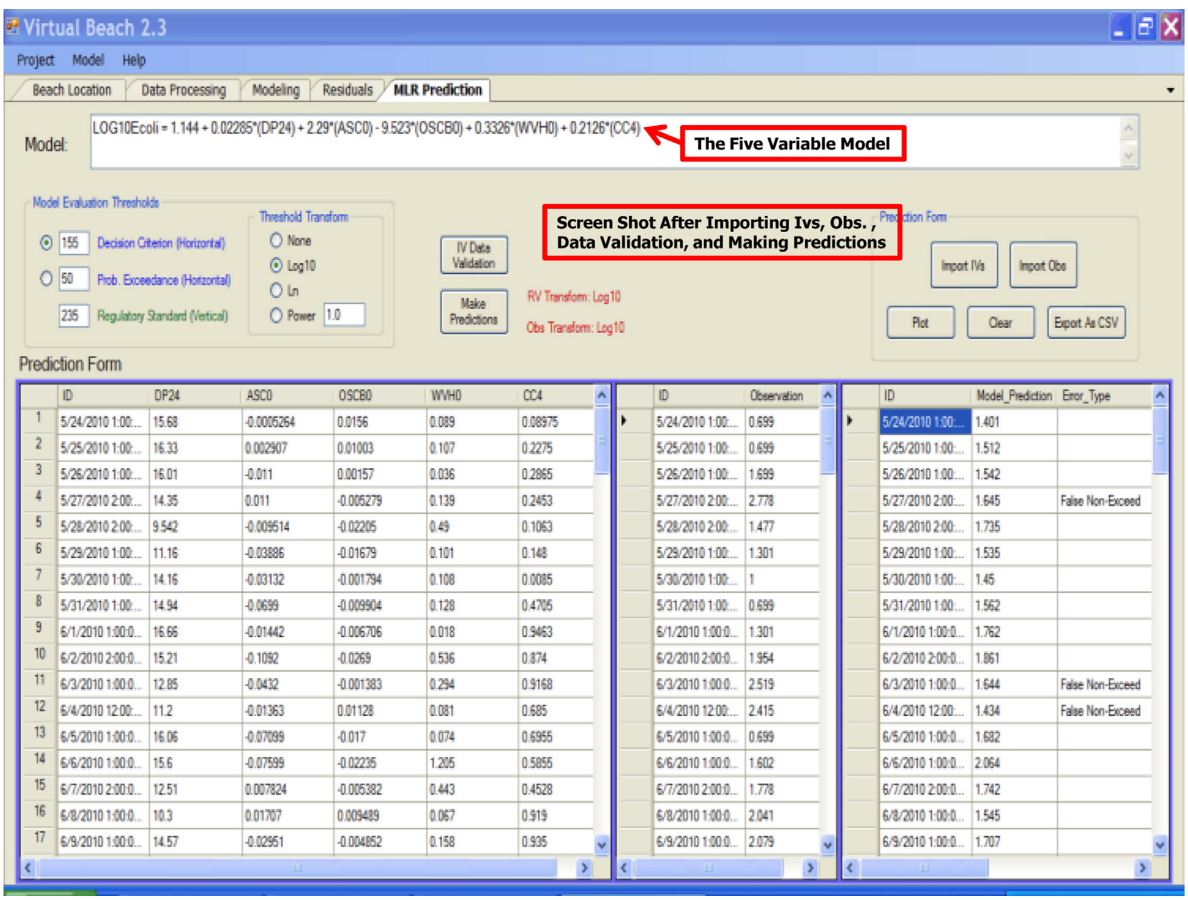


Figure 5.0.13: VB2.3 MLR Prediction Screenshot

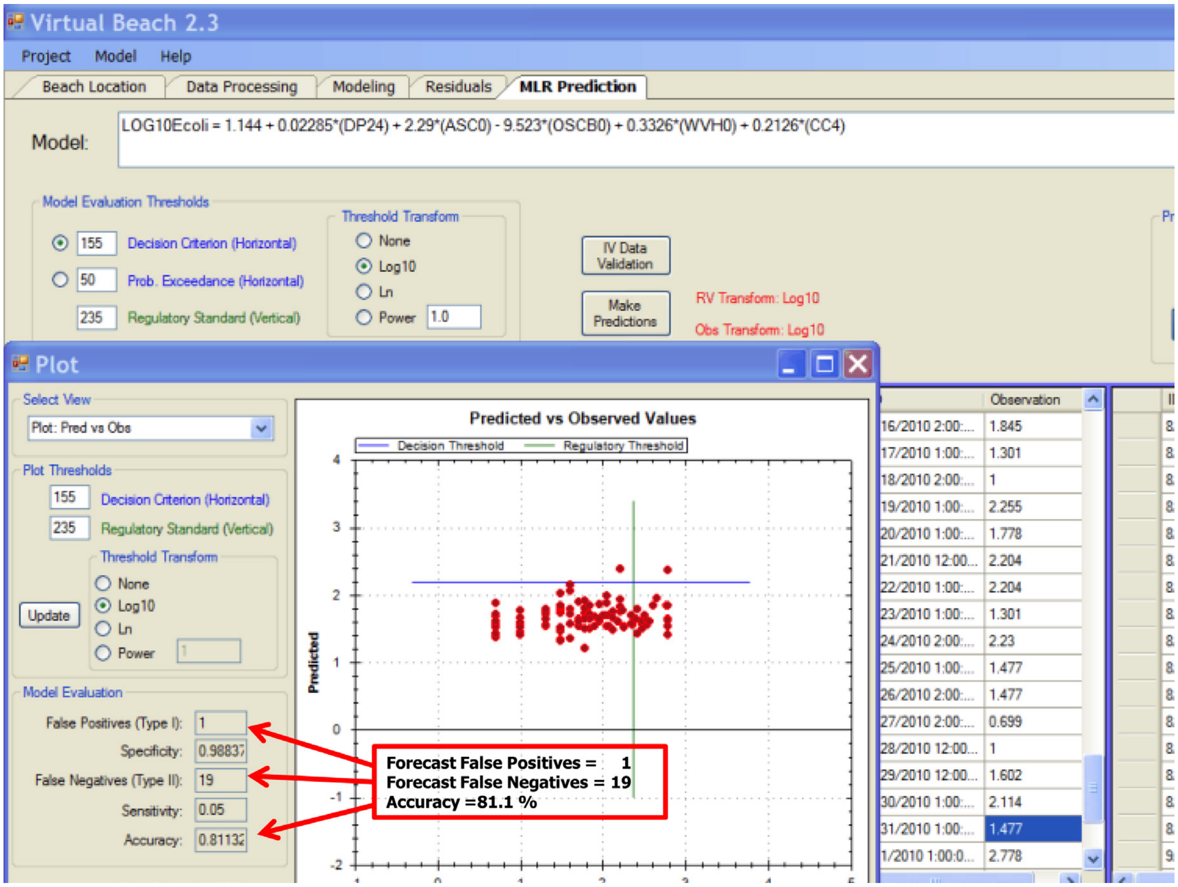


Figure 5.0.14: VB2.3 2010 Model Predictions vs 2010 Observed Values

It should also be noted that the DCH=155 was an optimal selection for the 2010 data. A lower DCH value would have increased type 1 errors (prevented swimming) with no corresponding decrease in type 2 errors (allowing swimming when bacterial contamination actually exceeded the state regulatory standard).

5.1 2010 Forecast Decision Support Systems for 35 Sampling Sites at Twenty-Four Beaches

The grant proposed to take predictive modeling to the next step. The grant developed forecast DSS for Twenty-Four Beaches. This resulted in 35 forecast DSS equations since some of the beaches were sampled at more than one location. Table 5.1 contains the 2010 forecast DSS equations for the 35 sites. The beach managers used these additional sampling locations to manage a portion of the beach. Table 5.1 lists the six beaches where this occurred. The key parameters identified by the multiple linear regressions are from the complete list of key parameters identified in Table 2.1. Fifty-two different key parameters were identified as being explanatory variables for *E. coli* variation at these 35 locations.

5.2 Key Variables in the Forecast Decision Support Systems for the 35 Sites

Table 5.2 summarizes the key variables in the Thirty-Five forecast DSS at the Twenty-Four Beaches and the number of times they were selected by the MLR. As a group, precipitation was identified most often as a key parameter for these beaches. Precipitation variables have been reported in several papers as a key explanatory variable. The physical process indicated by precipitation is the flushing of bacterial laden materials into the beach watershed (Francy et al. 2006, Holtschlag et al. 2008, Nevers and Whitman 2005). Precipitation key variables averaged over 48 and 72 hours suggest watershed areas further from the beach are sources of bacterial contamination. Onshore waves, along shore waves, and wave height variables are also identified as explanatory variables. The physical process involves the re-suspension of beach sand and release of bacteria in the sand (Nevers and Whitman 2005, Olyphant 2005, Frick et al. 2008). Although not identified as frequently, variables relating to tributaries are also important. Those beaches in the study near gauged

Table 5.1. 2010 Forecast DSS equations for the 35 sites.

Beach Name	Training Data Years	No. of Samples	2010 Decision Support System Equation (Using Minimum MSEP)	
Bay City State Rec. Area	2009	50	Log10E.coli =	-0.468 + 0.917(POLY(WSP0m,0.484,0.442,-0.0570)) + 1.01(POLY(SaginawRO_1d,1.18,-0.000347,4.62e-08)) + 0.180(AT0) - 0.214(AT120) - 1.77(OSWV0) + 18.9(OSC0)
Grand Haven North Beach Park	2002-2009	119	Log10E.coli =	-1.062 + 0.0786(DP24m) + 0.783(POLY(ASW 0m,1.04,0.00538,0.0173)) + 0.000112(GdRiverRD_5d) - 0.000296(GdRiverRO_8d)
Grand Haven State Park Beach	2002-2009	113	Log10E.coli =	-1.37 + 0.104(DP24) + 0.906(POLY(AS WV0,1.02,0.737,1.46)) + 0.842(POLY(CC 0m,1.02,1.15,-0.905)) - 6.62(√(TP0m)) - 0.103(AT240m) + 0.919(POLY(AT24 0m,3.70,-0.326,0.00975))
Racine North Beach Park	2003-2009	516	Log10E.coli =	-0.208 - 0.000186(DAYS) + 0.391(WVH0) + 0.301(ASWV0) + 0.316(CC4) + 0.740(POLY(TP24m,1.43,1.58,-0.705)) + 0.0938(TP72m) + 0.149(Q3) + 0.0295(SWT9) - 1.27(TP0m)
Racine Zoo Beach	2003-2009	517	Log10E.coli =	0.430 - 8.66e-05(DAYS) + 1.37(CS0) - 0.817(POLY(AS CB0,1.47,0.749,66.1)) + 0.490(WVH0) + 0.426(CC4) + 0.601(POLY(TP24m,1.38,1.71,-0.611)) + 0.426(POLY(TP72m,1.37,0.543,-0.0711)) + 0.118(Q3) + 0.0174(SWT9)
Upper Lake Park Beach	2003-2009	539	Log10E.coli =	-0.206 + 0.568(POLY(DAYS,1.25,0.00122,-5.47e-06)) + 0.0618(DP24) + 0.0537(SWT9) + 0.904(WVH0) - 1.95(POL Y(AT240m,1.53,-0.0183,0.00116)) - 0.0347(DP4m) + 0.569(CC4m) + 0.740(POLY (TP72m,1.52,0.55,-0.0183)) + 0.654(POLY(TP 0m,1.64,55.2,-679))
South Shore Beach	2003-2009	602	Log10E.coli =	-13.6 + 0.693(POLY(DAYS,2.08,0.000597,-2.26e-07)) + 0.474(WVH0) + 0.486(POLY(ASW0,2.24,0.0267,0.00971)) - 0.193(Q1) - 0.0547(AT24m) + 0.983(POLY(AT12 0m,0.605,0.145,-0.00287)) + 0.0404(DP24m) + 0.501(POLY (TP24m,2.26,0.966,-0.263)) + 3.85(POLY(MlwkeRD_0d,2.29,1.63e-05,-4.73e-10)) + 0.526(POLY(TP0m,2.32,11.0,-50.0))
Bradford Beach	2003-2009	604	Log10E.coli =	1.45 - 0.000190(DAYS) + 0.0710(SWT0) + 0.819(CS0) + 0.429(CC4) - 0.100(ATmin24) + 0.0673(DP24m) + 0.0338(WSP2m)
Woodlawn Beach State Park	2003-2009	468	Log10E.coli =	0.354 + 0.000167(DAYS) + 0.0808(DP24) + 2.88(ASC0) - 0.785(OSWV0) + 0.206(Q4) + 0.580(TP24m) - 0.0647(AT0) + 0.556(POLY(ASWV0,2.01,-0.445,1.99))
Hamburg Bathing Beach	2003-2009	221	Log10E.coli =	0.438 + 7.50(CS0) - 1.19(OSWV0) + 0.426(POLY(TP2 4m,1.75,1.65,-0.402)) + 0.396(TP72m) - 0.192(Q2)
Evans Town Park Beach	2003-2009	212	Log10E.coli =	0.229 + 0.0474(DP24) - 5.45(OSC0) + 1.31(WVH0) + 0.346(TP72m)
Wendt Beach	2003-2009	209	Log10E.coli =	0.444 + 0.100(DP24) - 0.0592(SWT0) - 3.30(ASC0) + 0.110(WS0) + 0.0765(OSW0) + 0.400(TP72m) - 0.194(Q2)
Lake Erie Park Beach	2003-2009	222	Log10E.coli =	-2.58 + 0.632(POLY(ASW0,1.52,0.00331,0.0125)) + 0.665(POLY(OSW0,1.67,0.0582,0.00308)) + 0.685(POLY(T P24m,1.57,1.51,-0.229)) + 0.437(POLY(TP72m,1.52,0.615,0.0101)) + 2.63(CS0) - 0.000116(DAYS)

Table 5.1. 2010 Forecast DSS equations for the 35 sites (cont.).

Bennett Beach	2003-2009	208	Log10E.coli =	-3.24 + 0.964(POLY(ASWV0,1.58,-0.169,2.93)) + 0.957(POLY(OSW0,1.56,0.0494,0.00525)) + 0.885(POLY(TP72m,1.44,0.846,-0.105)) - 0.212(Q2) + 0.0120(AT240m)
Presque Isle Beach 2 - 1	2006-2009	193	Log10E.coli =	-1.45 + 17.9(CSb0) + 0.672(POLY(OSWV0,0.662,-2.17,-0.738)) + 0.0733(WS0) + 0.0649(DP24m) + 0.0809(WSP4m) - 0.223(CC0)
Presque Isle Beach 2 - 2	2006-2009	192	Log10E.coli =	-1.75 + 0.000248(DAYS) + 5.69(CS0) + 0.735(POLY(OSWV0,0.581,-2.17,-0.649)) + 0.0525(AT240m) + 0.217(TP48m) + 0.0415(WSP4m) + 0.0221(DP24m)
Presque Isle Beach 10 - 1	2006-2009	164	Log10E.coli =	1.65 + 0.681(POLY(OSC0,1.11,-1.36,125)) + 1.02(WVH0) + 0.685(POLY(AT120m,-0.451,0.041,0.00198)) + 0.705(POLY(DP4m,1.29,-0.0797,0.00500)) - 2.85(POLY(CC4m,1.22,0.408,-0.324)) + 0.315(ASW0m) + 0.726(POLY(CC24m,0.977,1.35,-0.801)) - 0.0444(DP0) - 0.349(Q1)
Presque Isle Beach 10 - 2	2006-2009	170	Log10E.coli =	-1.47 - 3.86(ASCB0) + 0.710(WVH0) + 0.630(CC24) + 0.740(POLY(AT120m,-0.143,0.00497,0.00266)) + 0.0848(WSP4m) + 0.584(POLY(DAYS,1.24,-0.000888,8.24e-07)) + 0.264(POLY(AT24m,0.983,-0.0663,0.00344))
Presque Isle Beach 10 - 3	2006-2009	171	Log10E.coli =	-2.07 - 4.32(ASCB0) + 1.06(WVH0) + 0.528(CC24) + 0.0828(AT24m) + 0.702(POLY(TP0m,1.11,61.1,-490))
Villa Angela East	2002-2009	852	Log10E.coli =	-1.26 + 0.0604(DP24) + 0.323(WVH0) + 0.503(POLY(OSWV0,1.8,-1.55,-0.440)) + 0.248(CC4) + 0.338(Q1) + 0.417(Q2) + 0.479(TP24m) + 6.18e-05(DAYS) + 0.351(POLY(ASW0,1.98,0.0156,0.00662))
Villa Angela West	2002-2009	824	Log10E.coli =	0.722 + 0.0751(DP24) - 3.90(OSC0) + 0.183(ASWV0) + 0.779(POLY(OSWV0,1.76,-1.77,-0.533)) + 0.0700(WS0) - 0.374(Q3) - 0.370(Q4) - 0.908(POLY(AT240m,1.94,-0.0152,0.000958)) + 0.398(TP24m) + 0.306(TP48m) + 0.125(CC4)
Hammond Marina East Beach	2006-2009	413	Log10E.coli =	-2.75 - 0.0655(AT24) + 0.840(POLY(CC4,1.84,-0.730,0.879)) - 0.518(Q4) + 0.0481(SWT9) + 1.45(POLY(DAYS,1.68,0.000502,-3.66e-07)) + 0.0504(DP0) - 3.38(ASC0) - 0.492(OSWV0)
Marquette Park Beach - 1	2006-2009	176	Log10E.coli =	-6.03 + 0.0576(DP24) + 0.889(POLY(ASW0,1.66,0.00287,0.00703)) + 0.830(POLY(OSW0,1.72,0.0541,0.00529)) + 1.10(POLY(TP0m,1.77,-22.5,200)) + 1.12(POLY(TP24m,1.71,1.62,-1.35)) - 8.44(OSCB0)
Marquette Park Beach - 2	2006-2009	165	Log10E.coli =	-4.74 + 0.0387(DP24) + 3.93(OSC0) + 0.787(POLY(OSWV0,1.32,-1.54,-0.900)) + 0.199(CC24m) + 0.798(POLY(ASW0m,1.41,1.74,0.516)) + 1.30(POLY(TP4m,1.43,-0.839,-0.322)) + 0.988(POLY(TP48m,1.35,0.912,-0.485))
Marquette Park Beach - 3	2006-2009	157	Log10E.coli =	-3.56 + 0.0298(AT120) + 1.542(POLY(ASW0,1.18,0.0198,0.00811)) + 0.988(POLY(OSW0,1.27,0.0631,0.00448)) + 0.31(CC24m) + 0.705(POLY(TP48m,1.2,1.1,-0.506))
Marquette Park Beach - 4	2006-2009	163	Log10E.coli =	-2.57 + 1.39(POLY(ASW0,1.21,0.0230,0.00569)) + 0.803(POLY(OSW0,1.25,0.0749,0.00565)) + 0.810(POLY(TP48m,1.21,0.955,-0.506)) - 9.82(OSCB0)

Table 5.1. 2010 Forecast DSS equations for the 35 sites (cont.).

Washington Park Beach -1	2006-2009	201	Log10E.coli =	-0.482 + 0.0690(DP24) + 0.193(Q3) - 0.0743(AT24) + 0.0483(SWT0) + 0.915(POLY(DAYS,2.03,0.000566,-5.85e-07))
Washington Park Beach -3	2006-2009	199	Log10E.coli =	-1.42 + 0.0696(AT120) + 0.495(CC24) - 0.214(Q2) + 0.967(POLY(TP24m,2.03,0.669,-0.434)) - 0.139(Q4)
Washington Park Beach -5	2006-2009	200	Log10E.coli =	-1.90 + 0.0648(AT120) + 0.548(CC24) + 1.19(POLY(TP24m,2.05,0.454,-0.369)) - 0.617(ASC0)
OgdenDunes - 1	2006-2009	239	Log10E.coli =	-1.50 + 0.934(POLY(DAYS,1.20,-0.00279,2.53e-06)) + 1.06(POLY(ASC0,0.773,3.01,70.3)) + 0.886(POLY(OSW0,0.879,0.0504,0.00181)) - 0.0175*(DP24m) + 0.0145(AT4m)
OgdenDunes - 2	2006-2009	240	Log10E.coli =	-4.41 + 0.969(POLY(OSWV0,1.28,-0.859,-0.243)) + 0.872(POLY(TP72m,1.24,0.483,-0.0468)) + 1.04(POLY(BrnsDtchRO_1d,1.27,0.000209,-3.20e-08)) + 1.38(POLY(BrnsDtchRO_8d,1.40,-9.53e-05,9.83e-09))
OgdenDunes - 3	2006-2009	239	Log10E.coli =	-3.68 + 0.824(POLY(DAYS,1.08,-0.00220,2.25e-06)) + 1.11(POLY(OSWV0,0.862,-1.20,-0.409)) + 1.75(POLY(DP24m,1.09,-0.0233,0.000312)) + 0.439(CC24m) + 0.933(POLY(TP72m,0.862,0.784,-0.360)) + 0.000121(BrnsDtchRO_1d) + 0.0150(AT4m)
IDSP West Beach	2006-2009	367	Log10E.coli =	1.14 + 0.0228(DP24) + 0.333(WVH0) + 2.29(ASC0) - 9.52(OSCB0) + 0.213(CC4)
Memorial Beach	2007-2009	130	Log10E.coli =	-3.76 + 0.000424(DAYS) + 0.550(CC24) - 0.491(Q4) + 6.07e-05(ClintonRD_0d) + 1.42(POLY(AT120m,-0.911,0.310,-0.00806)) + 1.23(POLY(OSWV0,1.93,0.657,3.62))
Metropolitan Beach	2007-2009	130	Log10E.coli =	-1.30 + 0.000809(DAYS) + 0.588(POLY(AT24m,-0.472,0.229,-0.00595)) + 0.998(POLY(ASWV0,1.65,2.09,-4.49))
Poly (IV,a,b,c)= a + b(IV) + c $(IV)^2$ where IV is the independent variable, and a,b,c are coefficients in the equation.				

Rivers had key variables identified from the tributaries either as discharge or runoff variables. Tributaries carry bacteria to the beach watershed from nearby watersheds (Nevers and Whitman 2005) or from storm or waste water treatment plants discharging to the rivers.

5.3. Comparison of 2010 Forecast DSS with Persistence and Always Open Management Tools

The grant proposed to compare the forecast models with the persistence and nowcast models. Only one Nowcast Model was found available for 2010. In place of the nowcast model, the other common beach management methods used in the Great Lakes is the Always Open (AO) management tool. This tool is used to keep the beach swimming always available except during times of physical danger such as high waves, lightning, or known sewage spills. AO is used by default when no bacterial water quality monitoring is done. AO is used when swimming at your own risk is permitted irrespective of elevated bacterial levels based on real time models.

Attachment 5.3 contains comparison of 2010 forecast DSS with Always Open (AO) and Persistence Model (PM) Beach Management Models for 24 Beaches at 35 locations. VB2 provides 10 best equations during the cross validation step described in section 5. For this study, the best equation with the most variables and minimum mean square error of prediction was chosen and call FDSS-MV. The second selection from the 10 best equations available in the cross validation step was the equation with the fewest variables and minimum mean square error of prediction and called FDSS-LV.

Table 5.2. Key variables in the 2010 forecast DSS for the Twenty-Four Beaches.

Independent Variable	No. of Occurrances	Independent Variable	No. of Occurrances
DAYS	16	WVH0	9
Q1	3	ASWV0	6
Q2	5	OSWV0	13
Q3	4	CS0	7
Q4	5	ASC0	6
AT0	2	OSC0	4
AT4m	2	CSb0	1
AT24/AT24m	6	ASCb0	3
AT120/AT120m	8	OSCb0	3
AT240m	6	SWT0	3
ATmin24m	1	SWT9	4
DP0	2	TP0m	5
DP4m	2	TP4m	1
DP24/DP24m	17	TP24m	12
CC0	1	TP48m	5
CC4/CC4m	9	TP72m	9
CC24/CC24m	9	ClintonRD_0d	1
WSP0/WSP0m	4	MlwkeRD_0d	1
ASW0/ASW0m	9	GDRiverRD_5d	1
OSW0	7	SaginawRO_1d	1
WSP2m	1	BrnsDtchRO_1d	2
WSP4m	3	GDRiverRO_8d	1
		BrnsDtchRO_8d	1

In actual practice of forecasting for a beach, evaluation of the 10 best equations might result in another selection using the beach manager's professional judgment. The explanatory variables considered and ultimately used in the forecast DSS selected would depended on the type of sources found in the beach watershed and the transport mechanisms by which the bacterial contamination is suspected to reach the beach waters.

Attachment 5.3.1 stratifies the comparison of 2010 Forecast DSS with AO and PM into those beaches using the *E. coli* standard for Michigan's single sample maximum state standard of 300 counts/100 ml and the single sample maximum of 235 counts/100 ml state standard for all the other Great Lakes states.

For the beaches with 5% or fewer samples exceeding the single sample maximum state regulatory standard, the forecast DSS system does not do better than the PM. AO beach management control system does have fewer total errors when compared to forecast DSS or PM at beaches with 30% or fewer samples exceeding the single sample maximum state regulatory standard. However, the AO exposes swimmers to elevated bacterial levels more frequently than PM or forecast DSS management control systems.

Figure 5.3.1 uses data from 30 sampling sites from 19 beaches where the single sample *E. coli* standard is 235 counts/100 ml. The PM is compared with the FDSS-MV using the equation with the minimum Mean Square Error of Prediction (MSEP). For beaches where the percentage of the number of samples above the state regulatory standard are less than or equal to 5%, the FDSS-MV has more errors than the PM. For beaches where the percentage of the number of samples above the state regulatory standard are greater than 5%, the FDSS-MV had fewer errors in 19 of 27 sampling sites or 70% of the sites. FDSS-MV had more errors in 4 of 27 sampling sites or 15% of the sites. At 4 of the 27 sites FDSS-MV had

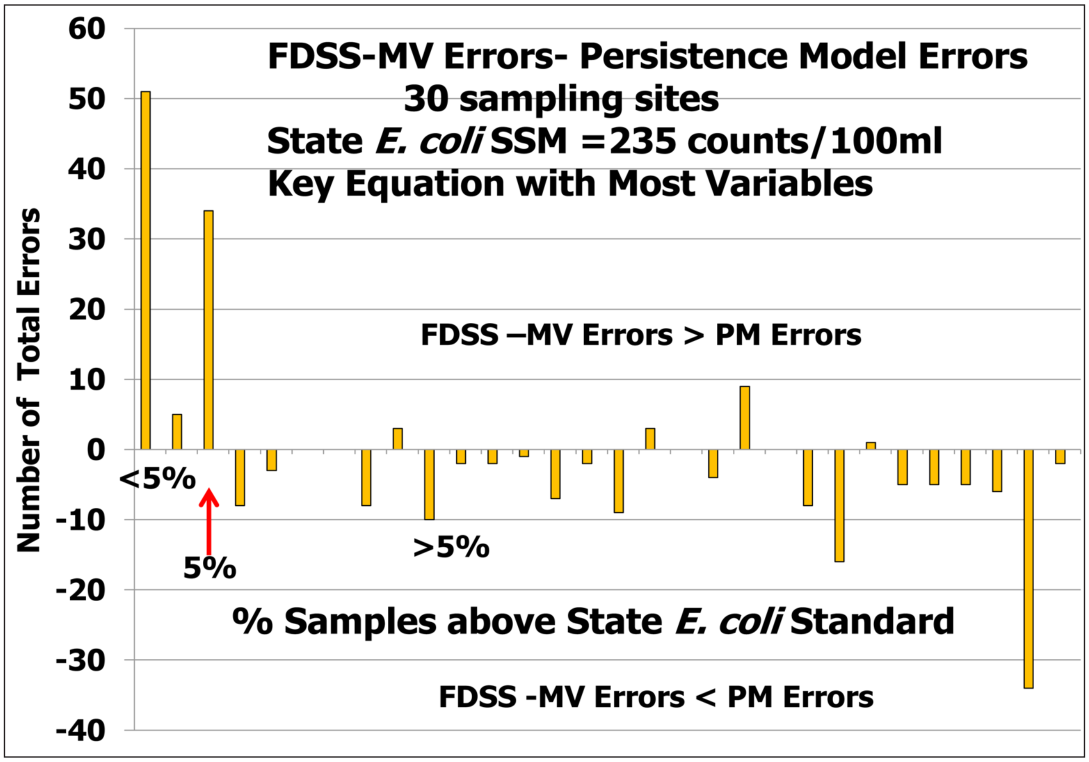


Figure 5.3.1. FDSS-MV Errors Compared to PM Errors.

the same number of errors as PM. FDSS-MV did as well or better in 85% of the sites with more than 5% of the samples exceeding the state regulatory standard.

Figure 5.3.2 uses data from 30 sampling sites from 19 beaches where the single sample *E. coli* standard is 235 counts/100 ml. The PM is compared with the FDSS-LV using the equation with the minimum Mean Square Error of Prediction (MSEP). This option is considered because the FDSS-LV equation generally had 2 fewer variables to forecast. For beaches where the percentage of the number of samples above the state regulatory standard are less than or equal to 5%, the FDSS-LV has more errors than the PM similar to FDSS-MV. For beaches where the percentage of the number of samples above the state regulatory standard are greater than 5%, the FDSS-LV had fewer errors in 18 of 27 sampling sites or 67% of the sites. FDSS-LV had more errors in 5 of 27 sampling sites or 18% of the sites. At 4 of the 27 sites FDSS-LV had the same number of errors as PM. FDSS-LV did as well or better than the PM at 82% of the sites with more than 5% of the samples exceeding the state regulatory standard.

These two approaches result in similar outcomes. FDSS-MV and minimum MSEP performed 3% better in this study than the FDSS-LV. This resulted in one additional site with fewer errors than the PM model. Based on this small sample, FDSS-MV equations were used for the five 2012 forecast DSS models.

5.4 Comparison of 2010 Forecast DSS for adjacent Hydrodynamic Grid Cells at Three Beaches

A comparison of forecast DSS results for three beaches was done to test the impact of using hydrodynamic model data from adjacent grid cells. The object of this test was to determine if adjacent hydrodynamic cells would result in different beach management decisions and if a method could be identified to optimize the selection.

The three beaches were selected that fell outside the hydrodynamic grids. These beaches are Racine Zoo Beach (Wisconsin, Lake Michigan), Lake Erie Beach (New York, Lake Erie), and Memorial Park (Michigan, Lake St. Clair).

Figures 5.4.1 to 5.4.3 shows the two hydrodynamic grid cells with one of the cells identified as optimal for each of the four beaches. The optimal cell was selected based on criteria given in section 2.3. For Racine's Zoo Beach in Figure 5.4.1,

Table 5.4 Comparison of Beach 2010 FDSS-MV Equation Results to Always Open & Persistence Beach Management Models at Adjacent Hydrodyamic Grid Nodes.

			Training Data for Model						2010 Forecast Decision Support System					
Beach	**Grid Nodes Optimal or Other**	**Variables in VB2.2 Decision Support System**	**% Accuracy**	**% R^2**	**% adj R^2**	**MSEP**	**Number Training Samples (Start-End YY)**	**% (Number) Data Equal or Greater than State Standard**	**% Accuracy**	**% Specificity**	**False Positives**	**% Sensitivity**	**False Negatives**	**Total Errors**
Zoo Beach	Optimal	9	91	44	43	0.246	517 (2003-2009)	11.4 % (59)	91	99	1	0	4	5
	Other	7	91	43	42	0.246	517 (2003-2009)	11.4 % (59)	91	94	3	33	2	5
Racine County, WI	PM								93	96	3	25	3	6
	AO								95	100	0	0	4	4
Lake Erie Beach	Optimal	6	80	42	41	0.362	222 (2003-2009)	22.5 % (50)	83	96	1	29	5	6
	Other	6	80	44	42	0.353	222 (2003-2009)	22.5 % (50)	77	93	2	14	6	8
Erie County, NY	PM								71	82	5	17	5	10
	AO								80	100	0	0	7	7
Memorial Park Beach	Optimal	6	82	32	29	0.348	130 (2007-2009)	20.0 % (26)	46	38	21	86	1	22
	Other	7	83	38	35	0.342	128 (2007-2009)	20.0 % (26)	54	47	18	86	1	19
Macomb County, MI	PM								80	88	4	43	4	8
	AO								83	100	0	0	7	7

FDSS MV = Forecast Decision Support System using Equations with most variables

FDSS LV = Forecast Decision Support System using equations with least variables

PM = Persistence Model

AO = Always Open Model

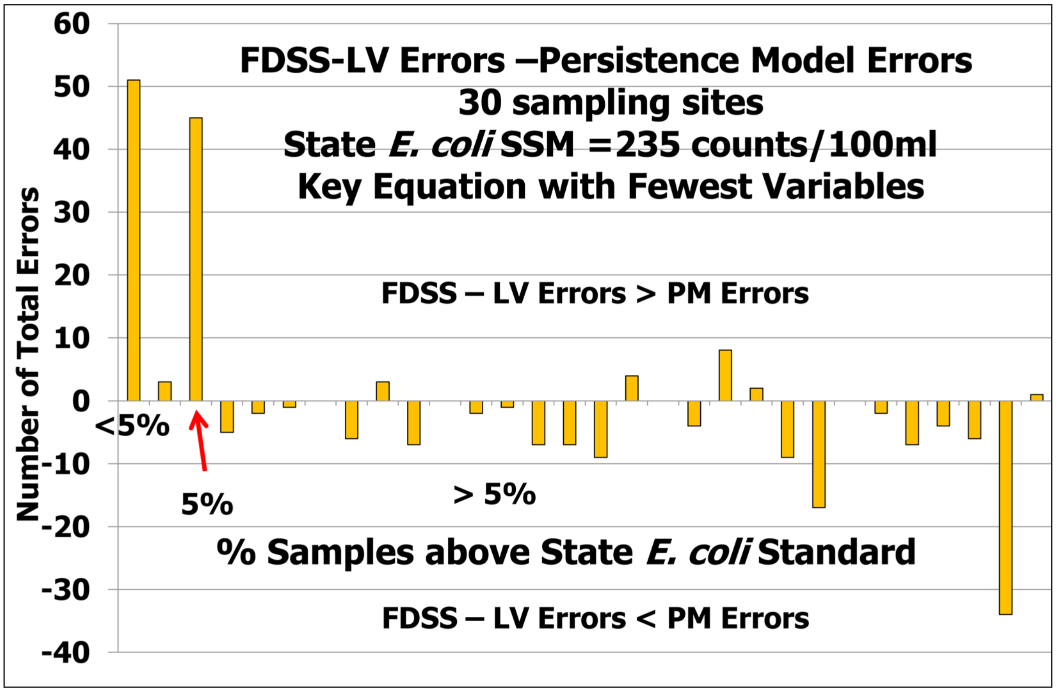


Figure 5.3.2. FDSS-LV Errors Compared to PM Errors.

Figure 5.4.1. Racine Zoo Beach Adjacent Hydrodynamic Grid Cells.

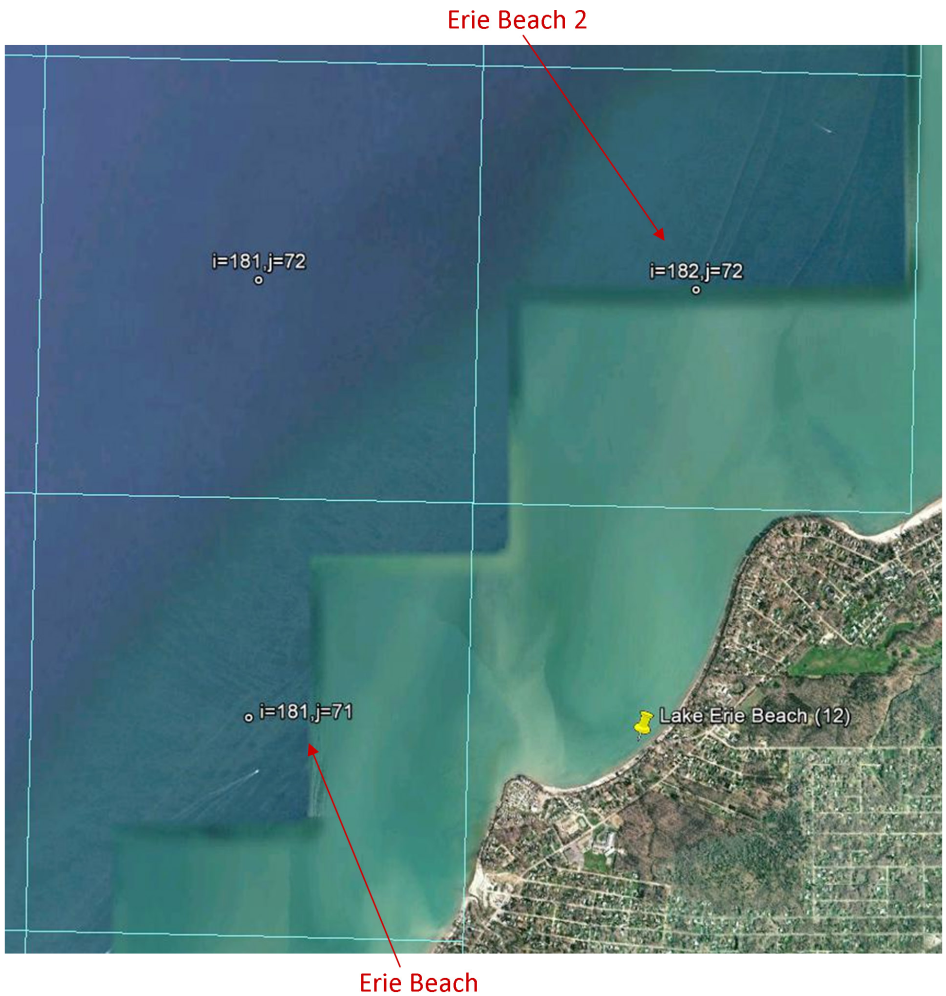


Figure 5.4.2. Lake Erie Beach Adjacent Hydrodynamic Grid Cells.

Figure 5.4.3. Memorial Beach Adjacent Hydrodynamic Grid Cells.

the difference in the distance calculation from each cell is small. The distance is 1.87 km to grid cell "A" and 1.90 km to grid cell "B" (see Figure 5.4.1).

Table 5.4 summarizes the results for FDSS-MV equations for adjacent beach cells. The training data sets for each hydrodynamic cell for Racine Zoo Beach and Lake Erie beach have the same accuracy. The accuracy for the Memorial Beach training data sets differed by one percent. For the three comparisons, the total errors in the 2010 forecasts were the same for Racine's Zoo beach but differed with respect to type 1 and type 2 errors. The total errors for Lake Erie beach were two fewer errors for the optimal cell, but for Memorial Beach the optimal cell had three more errors. Based on this small number of cases, it appears the adjacent nearest cells, have some effect on the total number of errors in the forecast, but the training data does not clearly indicate which cell would be preferred for developing the forecast DSS. For this study, the nearest hydrodynamic cell was used.

6. NOAA BEACH WATER QUALITY EXPERIMENTAL FORECASTS

6.1 Five Beaches 2012 Forecast Decision Support Systems

One forecast DSS was developed for the 2012 swimming season for each of five beaches. The beaches involved are located in Michigan. Three of the beaches are in the geographical forecast area of the Detroit Pontiac NWS. Two Ottawa County beaches are outside of the Detroit Pontiac Forecast Office geographical domain. The 2012 forecast DSS equations for each of the five beaches in the State of Michigan are found in Table 6.1. The beaches are Bay City State Recreation Area, Bay County MI, Memorial and Metro Beach, Macomb County, and North Beach Park Beach, and Grand Haven State Park Beach, Ottawa County, MI.

The Detroit Pontiac National Weather Service Office provided four daily forecasts for these five beaches. These forecasts were issued at Midnight, 6 am, Noon, and 6 pm. The forecasts were provided to cooperating beach managers via the web during the swimming season.

6.2 Key Variables in 2012 Forecast Decision Support System Equations

6.2.1 Bay City State Rec. Area Beach 2012 Forecast DSS

The 2012 forecast DSS equation for the Bay City State Rec. beach based on training data from 2009-2011 is:
LOG10Ecoli = 0.453 - 0.000422 * DAYS + 6.25 * ASC0 + 0.0993 * WS0 + 0.103 * OSW0 + 4.99e-05 * SaginawRD_9d.

Table 6.1. 2012 forecast DSS equations for five beaches in the State of Michigan.

Beach Name	Training Data Years	2012 Forecast Decision Support System Equations	
Bay City State Rec. Area	2009-2011	LOG10Ecoli =	0.453 - 0.000422*(DAYS) + 6.25*(ASC0) + 0.0993*(WS0) + 0.103*(OSW0) + 4.99e-05*(SaginawRD_9d)
Memorial Beach	2007-2011	LOG10Ecoli =	-9.72 + 0.999*POLY(DAYS,1.61,0.00120,-6.61e-07) + 2.44*POLY(AT24,1.47,0.0798,-0.00263) + 0.880*POLY(OSWV0,1.89,-0.228,2.56) - 0.568*(Q4) + 1.03*POLY(ClintonRD_0d,1.623,0.000932,-1.42e-07) + 0.0675*(AT120)
Metro Beach	2007-2011	LOG10Ecoli =	-4.88 + 0.996*POLY(DAYS,1.18,0.00130,-6.70e-07) + 0.0351*(AT0) + 2.051*(ASWV0) - 0.269*(Q2) + POLY(AT240m,0.578,0.093,-0.00193) + 1.66*POLY(ClintonRD_0d,1.51,0.000316,-6.93e-08)
Grand Haven State Park	2002-2011	LOG10Ecoli =	-0.893 + 0.0393*(DP0) + 0.0632*(WS0) + 0.790*POLY(AT 0m,1.827,-0.141,0.00487) + 9.95e-05*(GdRiverRD_0d) - 0.000118*(GdRiverRO_0d)
North Beach Park Beach	2002-2011	LOG10Ecoli =	-0.570 + 0.116*(DP24) - 0.0849*(SWT0) + 2.242*(CS0) + 1.065*PO LY(OSW0m,1.27,0.0551,-0.00834) + 0.328*(TP48m)

The five key variables are: DAYS (Count of Days from the first sample to the last sample with first sample day = 0), ASCO(Alongshore current at sample hour, n=1 (m/s, positive clockwise)), WSO(Wind speed at sample hour, n=1 (m/s)), OSWO(On-shore wind at sample hour , n=1 (m/s, positive towards beach)), and SaginawRD_9d(Discharge nine days prior to sample day (ft^3/s)).

The DAYS independent variable has a negative coefficient indicating over time *E. coli* concentrations are diminishing. ASCO has a positive coefficient indicating currents from the south of the beach increase *E. coli* by bringing higher bacterial concentration water towards the beach from the mouth of the Bay where the Saginaw River discharges. WSO has a positive coefficient indicating stronger winds increase *E. coli* concentrations at the beach. OSWO has positive coefficient indicating onshore winds increase *E. coli* concentrations at the beach. SaginawRD_9d has a positive coefficient indicating Saginaw River Discharge 9 days previous to the sample day increases *E. coli* concentration at the beach.

The combination of ASCO and SaginawRD_9d paint a picture of river discharge into the Saginaw Bay that is gradually circulated clockwise to the beach. The other key parameters indicate winds blowing Saginaw Bay water towards the beach increase *E. coli* concentrations. Stronger winds can push water more effectively towards the beach and re-suspend more sediment.

6.2.2 Memorial Beach 2012 forecast DSS

The 2012 forecast DSS equation for Memorial Beach based on training data from 2007-2011 is:
LOG10Ecoli = -9.72 + 0.999 * POLY(DAYS, 1.61, 0.00120, -6.61e-07) + 0.0675 * AT120
+ 2.44 * POLY(AT24, 1.47, 0.0798, -0.00263) + 0.88 * POLY(OSWV0, 1.89, -0.228, 2.56)
– 0.568 * Q4 + 1.03 * POLY(ClintonRD_0d, 1.623, 0.000932, -1.42e-07).

This forecast DSS involves the Polynomial transformation which is defined in Section 4 VB2 software as POLY(KV,a,b,c) and is the following equation $P(KV)= a + b(KV) + c (KV)^2$ where KV stands for Key Variable or IV in this report. This transformation is more difficult to interpret because the contribution to LOG10Ecoli can change sign over the time period of the training data and the forecast year. Several illustrations will be provided with this forecast DSS.

The six key variables are: DAYS (Count of Days from the first sample to the last sample with first sample day = 0), AT120 (Air temp previous 120 hour average, n≤120, (°C)), AT24 (Air temp previous 24 hour average, n≤24, (°C), OSWVO (Onshore Waves (m) at sample hour, positive towards shore, negative away from shore), Q4 (Categorical Variable for the fourth quartile of the sampling season by year), and ClintonRD_0d (Same Day River Discharge ft^3/s).

A plot of 0.999 * POLY(DAYS, 1.61, 0.0012, -6.61e-07) is shown in Figure 6.2.1. Note the inflection point during 2009. After 2009, this IV of the 2012 forecast DSS equation contributes less to the total LOG10Ecoli. This is signaling a declining bacterial load to the beach after 2009. Macomb County has been eliminating bacterial sources to the Clinton River after the 2002-2005 USGS study (Holtschlag et al. 2008) identified the Clinton River discharge as a key variable in explaining *E. coli* variation at this beach.

The contribution to *E. coli* concentration from this IV is greater for Memorial Beach than Metro Beach. This may be due to a cutoff in the Clinton River which takes river water closer to Memorial Beach (see discussion in Metro Beach following).

A plot of 2.44 * POLY(AT24, 1.47, 0.0798, -0.00263) is shown in Figure 6.2.2. All AT24 values contribute to increasing LOG10Ecoli. However, the optimal temperature for bacterial growth appears to be about 15°C.

The other IV for air temperature is the un-transformed variable AT120. The coefficient of AT120 is positive indicating that with increasing air temperatures averaged over five days, *E. coli* bacterial concentrations would be increasing.

Figure 6.2.1. Plot of POLY(DAYS) for Memorial Beach.

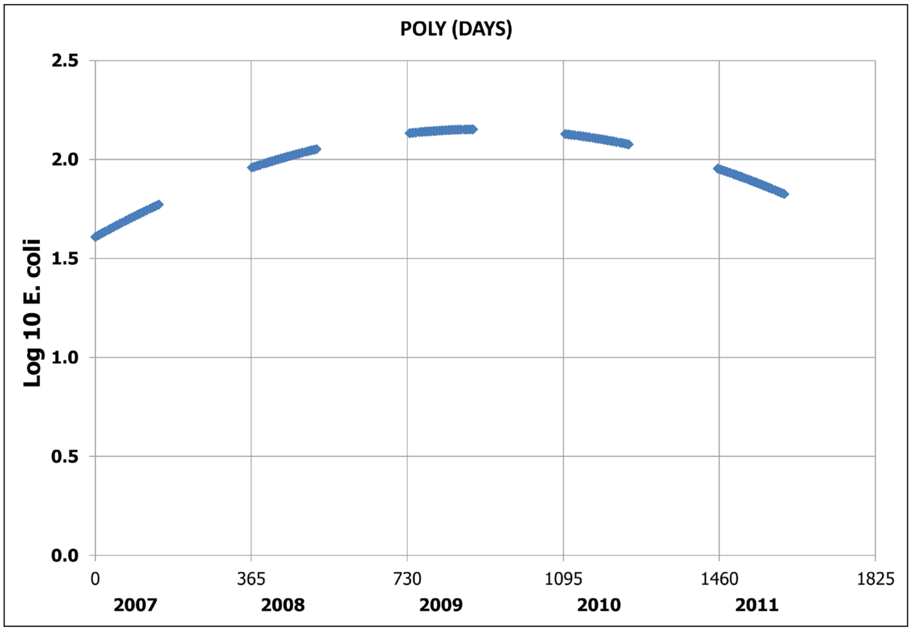


Figure 6.2.2. Plot of POLY(AT24) for Memorial Beach.

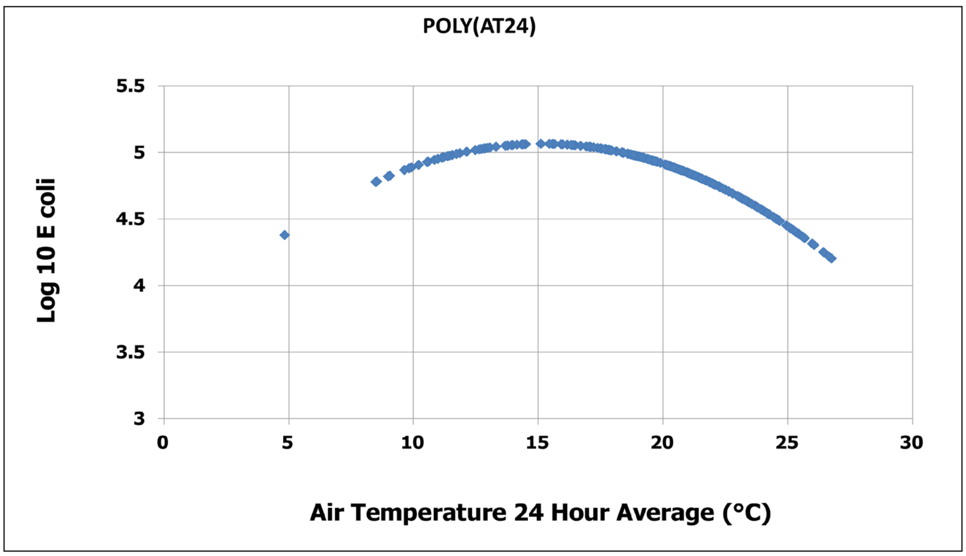
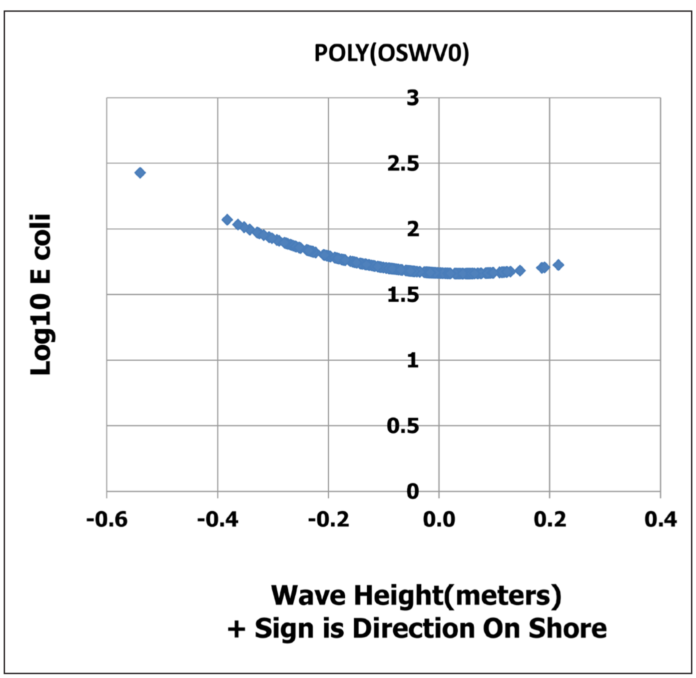


Figure 6.2.3. Plot of POLY(OSWO) for Memorial Beach.



The coefficient of Q4 is negative indicating that during the last quarter of the swimming season, *E. coli* bacterial concentrations were declining. The last quarter of the season would include August when air temperatures would be highest and reduced rainfall. Diminished rainfall would adversely affect bacterial growth and figure 6.2.2 POLY(AT24) graph shows less bacterial growth at higher temperatures.

Figure 6.2.3 POLY(OSWVO) is positive and adds to the *E. coli* concentration regardless of direction of the waves toward or away from shore. Waves during the time of sampling are not large with maximum height near 0.6 meter (under 2 ft).

On the next page Figure 6.2.4 POLY(ClintonRD_0d) shows the discharge of the Clinton River at the time of sampling as measured by the USGS gauge 04164000. Increased discharge generally results in higher bacterial concentrations at the beach up to 3000 ft^3/s. The bacterial concentration declined at the highest discharge rate (between 4000 and 5000 ft^3/s).

6.2.3 Metro Beach 2012 FDSS

The 2012 forecast DSS equation for Metro Beach based on training data from 2007-2011 is:
LOG10Ecoli = -4.88 + 0.996 * POLY(DAYS, 1.18, 0.00130, -6.70e-07) + 0.0351 * AT0 +
2.051 * ASWV0 - 0.269 * Q2 + POLY(AT240m, 0.578, 0.093, -0.00193) +
1.66 * POLY(ClintonRD_0d, 1.51, 0.000316, -6.93e-08).

The five key variables are: DAYS (Count of Days from the first sample to the last sample with first sample day = 0), AT0 (Air temp at sample hour, $n \le 1$ (°C), ASWVO (Along Shore Waves at sample hour (m), positive = clockwise rotation), Q2 (Categorical Variable =2 for the second 25% of the sampling season by year), AT240m (Air temperature (Dry Bulb Celsius) at Selfridge Airport (MTC) over previous 240 hour average $180 \le n \le 240$ (°C)), and ClintonRD_0d (Same Day River Discharge ft^3/s).

Figure 6.2.5 shows a plot of 0.996 * POLY(DAYS, 1.18, 0.00130, -6.70e-07). Note the inflection point in the contribution to *E. coli* for Metro Beach during 2009 similar to Memorial Beach. After 2009, Metro's POLY(DAYS) contributes less to the total LOG10Ecoli. The magnitude of this IV to Metro Beach bacterial concentration is less than its contribution to Memorial Beach. There is a cut off in the Clinton River which takes river water near to the Memorial Beach. This may account for the lower concentrations at Metro Beach since the source of the bacterial contamination at Memorial Beach may not be coming from the main Clinton river outlet into Lake St. Clair.

The coefficient of AT0 is positive indicating that with increasing air temperatures at the time of sampling, *E. coli* bacterial concentrations would be increasing.

The coefficient of ASWVO (Along Shore Waves at sample hour (m), positive= clockwise rotation) is positive indicating that waves approaching the beach from the west would increase bacterial concentrations. A dominant clockwise circulation cell is observed near Metro Beach (Anderson et al. 2010, Anderson and Schwab 2011).

The coefficient of Q2 is negative indicating that during the second quarter of the swimming season, *E. coli* bacterial concentrations were declining. The second quarter of the season would include July when air temperatures would be increasing, and rainfall declining. Increasing air temperatures would increase bacterial growth while declining rainfall would adversely affect bacterial growth. Other factors could be in play such as bacteriaphage increased activity could consume bacterial concentrations faster than elevated temperatures could increase the bacterial concentrations.

In Figure 6.2.6, POLY(AT240m) is plotted. The coefficient of the 240 hour average air temperature is positive indicating that increasing average air temperatures over the previous ten days will increase bacterial concentrations at the beach resulting in increased *E. coli* concentrations. From the plot bacterial growth is near its maximum around 25°C.

Figure 6.2.4 Plot of POLY(ClintonRD_0d) for Memorial Beach.

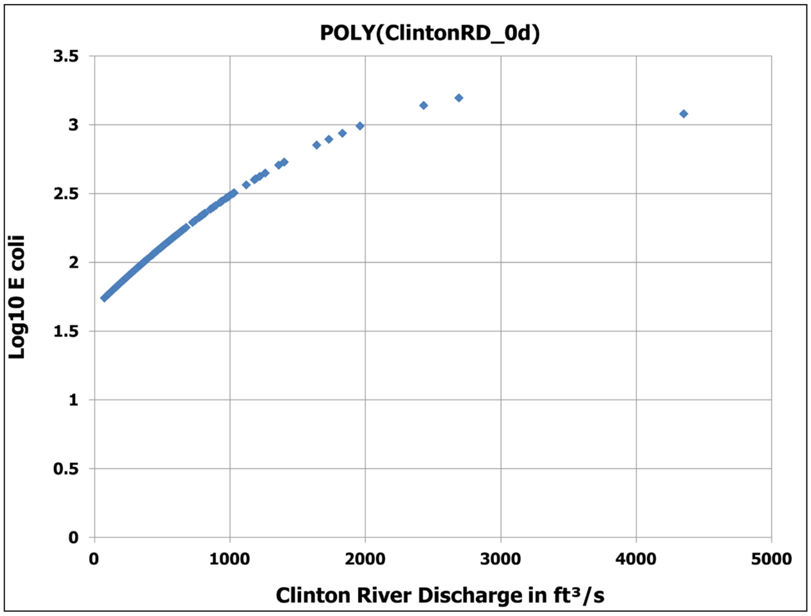


Figure 6.2.5. Plot of POLY(DAYS) for Metro Beach.

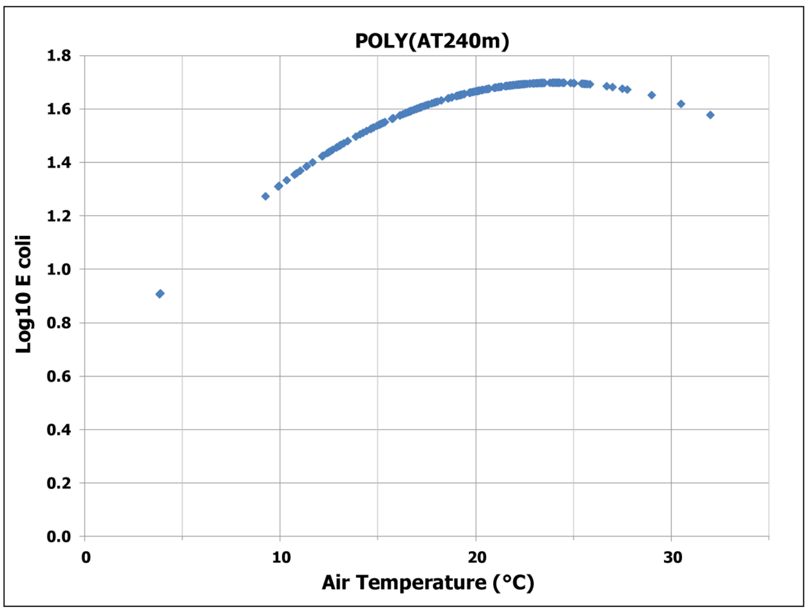


Figure 6.2.6. Plot of POLY(AT240m) for Metro Beach.

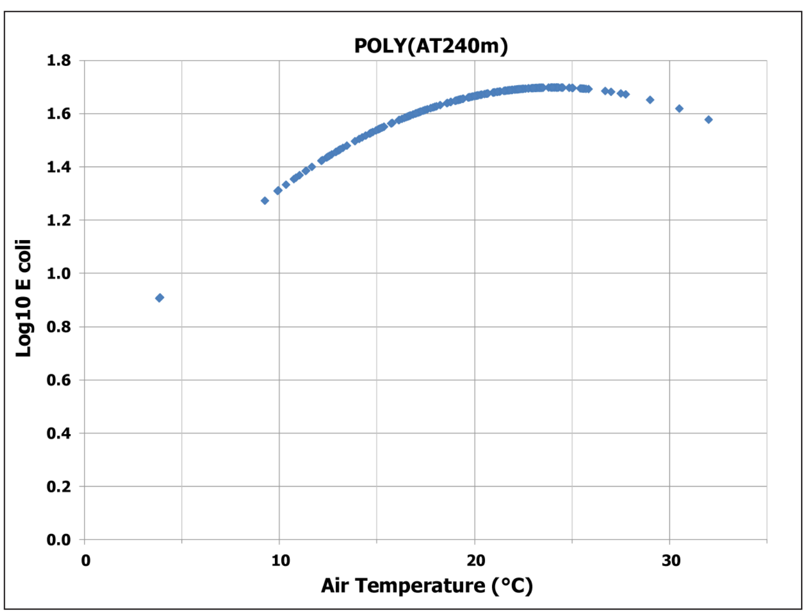


Figure 6.2.7 POLY(ClintonRD_0d) shows the discharge of the Clinton River at the time of sampling as measured by the USGS gauge 04164000. Increased discharge results in increased bacterial concentrations at the beach for all discharge volumes up to 2500 ft^3/s. Above 2500 ft^3/s the bacterial concentration appears to decline. The highest discharge above 4000 ft^3/s contribution to *E. coli* concentrations declines to levels similar to low flow levels.

6.2.4 Grand Haven State Park 2012 forecast DSS

The 2012 forecast DSS equation for Grand Haven State Park Beach based on training data from 2002-2011 is:
LOG10Ecoli = -0.893 + 0.0393 * DP0 + 0.0632 * WS0 + 0.790 * POLY(AT0m, 1.827, -0.141, 0.00487) + 9.95e-05 * GdRiverRD_0d - 0.000118 * GdRiverRO_0d.

The five key variables are DP0, Dew point temp at sample hour, $n \leq 1$ (°C); WS0, Wind speed at sample hour, n=1 (m/s); AT0m, (Air temperature (Dry Bulb Celsius) at Muskegon County Airport (MKG) at sample hour of measurement $n \leq 1$ (°C); GdRiverRD_0d and GdRiverRO_0d, Grand River Discharge and Runoff at same day of sample (ft^3/s).

The coefficient of DP0 is positive indicating that increasing amounts of water vapor in the air lead to increased bacterial concentration via impeding solar radiation which induces microbial inactivation. Dew point frequently replaces cloud cover which suggests that dew point and cloud cover are moderately collinear or redundant. Dr. Richard Zepp (personal communication) believes that air quality and/or poor ventilation in the atmospheric boundary layer may explain dew point as a key parameter. Aerosols and other air pollutants such as ozone that are particularly prevalent in stagnant urban air are effective attenuators of solar radiation, especially in the UV region which is particularly important in inducing microbial inactivation. Here in a non-urban atmospheric environment, we are seeing the same affect with increasing bacterial concentrations due to the lack of solar radiation attenuation with increasing dew point (increasing cloud cover).

The coefficient of WS0 is positive indicating increased wind speed increases bacterial concentrations.

Figure 6.2.8 POLY(AT0M) plot with increasing air temperature at sample time resulting in increased bacterial concentration for all temperatures. The positive coefficient of the second order term results in continuing increased bacterial concentrations with temperature rather than an optimal temperature.

Grand River discharge during the day of sampling has a positive coefficient which reflects that increasing discharge results in higher levels of *E. coli* at Grand Haven State Park.

Grand River runoff during the day of sampling has a negative coefficient which indicates that higher runoff decreases *E. coli* concentration at Grand Haven State Park.

The coefficient for the Grand River discharge is an order of magnitude smaller than the coefficient for Grand River Runoff. However, the Grand River discharge is more than an order of magnitude greater than the Grand River runoff. The combined effect of these two key parameters is to increase bacterial concentration in all years (see Figure 6.2.9).

6.2.5 North Beach Park Beach 2012 forecast DSS

The 2012 forecast DSS equation for North Beach Park Beach based on training data from 2002-2011 is:
LOG10Ecoli = -0.570 + 0.116 * DP24 - 0.0849 * SWT0 + 2.242 * CS0 + 1.065 * POLY(OSW0m, 1.27, 0.0551, -0.00834) + 0.328 * TP48m.

The five key variables are: DP24, Dew point temp previous 24 hour average, $n \leq 24$(°C); SWT0, Surface water temp at nearest 3 hour, n=1 (°C); CS0, Surface current speed at sample hour, n=1 (m/s); OSW0m, Onshore wind at sample hour at Muskegon County Airport (MKG), n=1 (m/s, positive towards beach) and TP48m,Total precipitation previous 48 hour total at Muskegon County Airport (MKG), $36 \leq n \leq 48$ (inches).

Figure 6.2.7 Plot of POLY(Clinton RD_0d) for Metro Beach.

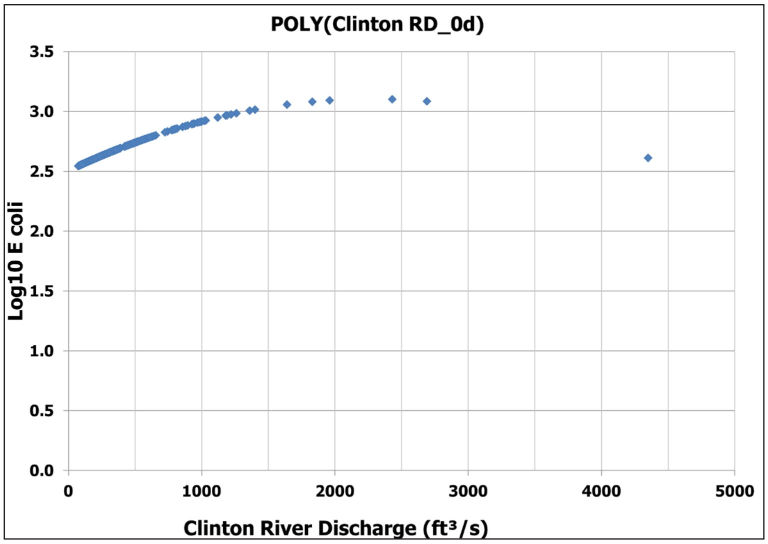


Figure 6.2.8. Plot of POLY(AT0m)) for Grand Haven State Park.

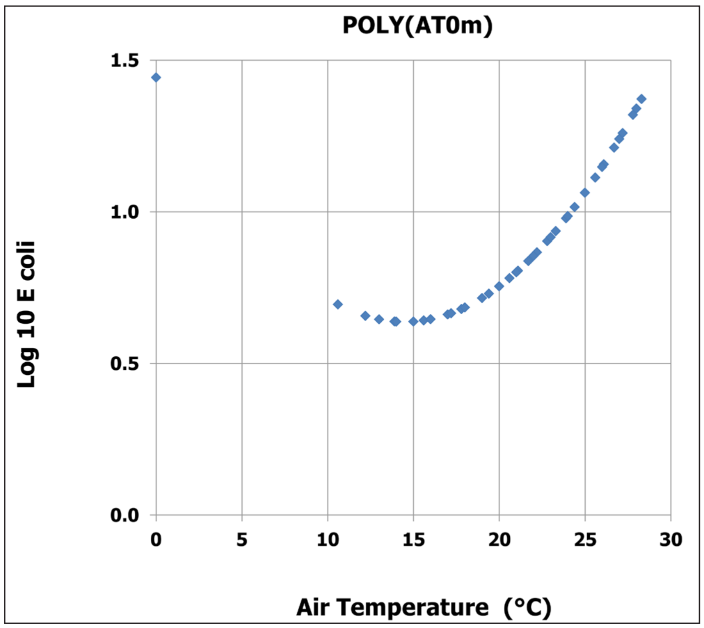
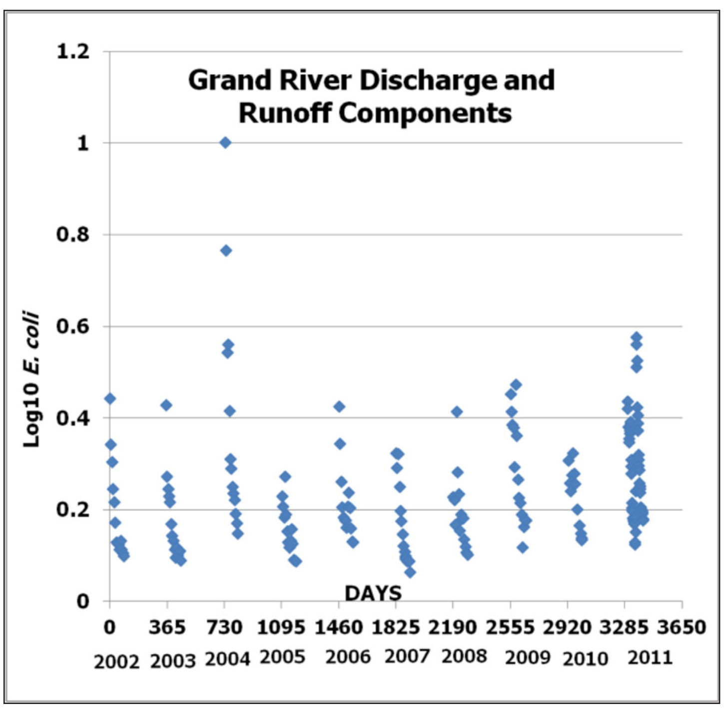


Figure 6.2.9. Plot of Grand River Discharge and Grand River Runoff for Grand Haven State Park.



The coefficient of DP24 is positive indicating increasing bacterial concentrations due to increasing values of the 24 hour average of dew point. This key variable acts in a similar manner to cloud cover which decreases solar radiation attenuation.

The coefficient of SWT0 is negative indicating decreasing bacterial concentrations due increasing water temperature. This result is counter intuitive since bacterial concentrations generally increase with increasing temperature. It is possible that increasing water temperature stimulates bacteriophage which reduces the bacteria concentrations more rapidly than the bacteria can grow due to higher water temperatures (Muruleedhara Byappanahalli personal communication).
The coefficient of CS0 is positive indicating increasing bacterial concentrations due to faster surface currents.
The coefficient of TP48m is positive indicating increasing bacterial concentrations due to increasing rainfall over the previous forty-eight hours. Runoff from the beach watershed would carry *E. coli* to the beach. The 48-hour average indicates the impact at the beach is from portions of the watershed further away from the beach.

Figure 6.2.10 POLY(OSW0m) shows the effect of onshore winds increasing bacterial contamination. Strong offshore winds have minimal impact while winds onshore have the maximum impact even at lower velocities.

6.3 2012 Forecast DSS Results Compared to Monitoring and Other Management Methods

Timely accurate forecasts of beach water quality are critical to protect human health against adverse exposure situations. The Center of Excellence for Great Lakes and Human Health, Great Lakes Environmental Research Laboratory, the National Weather Service, Detroit Pontiac Office, and the Cooperative Institute for Limnology and Ecosystems Research, University of Michigan developed and tested beach management forecast DSS at five beaches in Michigan. The NOAA Beach Water Quality Experimental Forecasts are possible because Bay, Macomb, and Ottawa County Health Departments have provided their *E. coli* monitoring data. The beaches involved are Bay City State Rec. Area, Bay Co. MI, Metro and Memorial Beaches, Macomb Co. MI, and North Beach Park and Grand Haven State Park, Ottawa Co. MI. (Table 6.1) The first three beaches were monitored approximately four times per week, and the last two were monitored approximately one time per week during the 2012 swimming season between Memorial and Labor Day. The NWS was generally successful in being able to run the forecast DSS in forecast mode by finding grid point and values for the forecast IVs. In one case, NWS required a different forecast equation because the precipitation independent variable needed to be replaced.

The results of the beach water quality forecasts was the ability to accurately forecast null events, i.e. keeping swimming available when bacterial counts were low. The accuracy ranged from 83 to 100% (average = 93.4 +/- 6.6%) when compared with the County Health Departments *E. coli* monitoring. The weakness was the inability to forecast any of the seven *E. coli* 2012 events occurring at four of the beaches (Table 6.3).

Figure 6.2.10. Plot of POLY(OSW0m) for North Beach Park Beach.

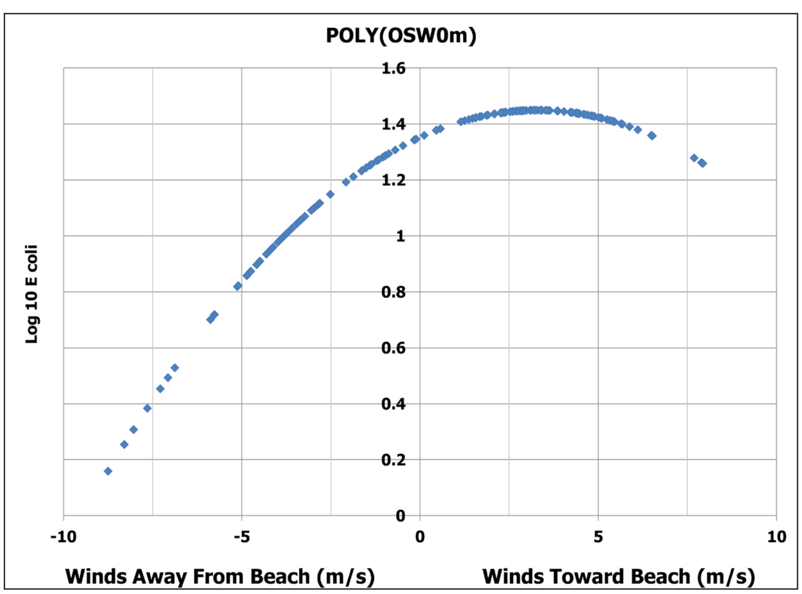


Table 6.3. Summary of 2012 FDSS Results for the Forecast Beaches.

Beach Name	Number of 2012 Samples	Monitored FDSS False Positives	Monitored FDSS False Negatives	FDSS Elevated Counts No Monitoring	% Exceedances of State Single Sample Maximum
Bay City State Rec Area	54	0	0	0	0
Memorial Beach	55	1	3	6	5.5
Metro Beach	55	0	1	0	1.8
Grand Haven State Park	12	0	2	0	13.3
North Beach Park Beach	15	0	1	0	6.7

Several graphics are provided for Memorial beach to illustrate the work performed. Similar graphics are available for the other beaches. For Memorial beach, the forecast DSS specificity was 98%. Specificity is the percentage of correctly forecasting bacterial concentrations lower than 300 counts/100 ml. The forecast DSS sensitivity was zero. Sensitivity is the percentage of correctly forecasting bacterial concentrations at or above 300 counts/100 ml. This is the important measure for forecasting skill used by the NWS. Although forecast DSS had fewer errors than Macomb County's presently used persistence beach management model, an always open management model would have had 100% specificity. The NWS requires skill in forecasting rare events. In beach water quality forecasting the rare event is an elevated *E. coli* concentration. Correctly forecasting at least one high bacterial concentration (sensitivity > 0) is required to exceed minimal skill in forecasting (Doswell et al. 1990). Doswell's 2 x 2 contingency table is directly comparable to VB predicted *E. coli* versus observed *E. coli* graphic. NWS cannot make forecast DSS available for widespread use because sensitivity was zero for all beaches where elevated *E. coli* events (≥300 counts/100 ml) were observed.

Figure 6.4.2 is used to show the results of the forecasting for Memorial beach. The blue, green, and gold traces are the minimum, expected, and high *E. coli* values generated at 06:00 EDT by the Detroit Pontiac NWS to forecast beach water quality for the 8:00 am EDT sampling period. Red dots are observed *E. coli* values provided by the Macomb County Health Department.

Note that six of the seven forecasted elevated *E. coli* concentrations occurred on the weekends when sampling was not conducted. As a result it is not known if these forecasted *E. coli* concentrations were correctly forecasted on not correctly forecasted.

Beach managers are evaluating the results of the 2012 season. Bay County will continue their four times a week monitoring in 2013 and is seeking NWS support to continue Bay City State Rec. Area beach forecast DSS in 2013.

6.4 Communication Plan for Distribution of Forecasts

The National Weather Service Forecast Office in White Lake, Michigan (WFO DTX), executed the beach forecast regression models for five Michigan Beaches in Bay, Macomb and Ottawa Counties four times per day through the 2012 beach and boating season (generally Memorial Day to Labor Day). Forecasts generated at midnight and 6 am EDT on a given day will be valid for 8 am expected morning sampling time that day and the 8 am expected morning sampling time the following day. Forecasts generated at noon and 6 pm EDT on a given day will be valid for 8 am expected morning sampling times the following two days. Forecast output will include a minimum value, a most likely value, and maximum expected values, and the likelihood that *E. coli* counts will exceed 300 parts per ml and 600 parts per ml respectively. These forecasts were available to beach managers via the National Weather Service dissemination infrastructure, and specifically via specified URLs through the National Weather Service Information Dissemination System (NIDS). Memorial Park: http://www.crh.noaa.gov/dtx/?n=beachwaterquality_mp is provided as an example of the URL available during the 2012 summer. Figure 6.4.1 shows the format and information provided by the NOAA Forecast for Memorial Beach Macomb County, MI for July 29, 2012 Midnight forecast at the website. All beach URL sites were taken down after the 2012 swimming season.

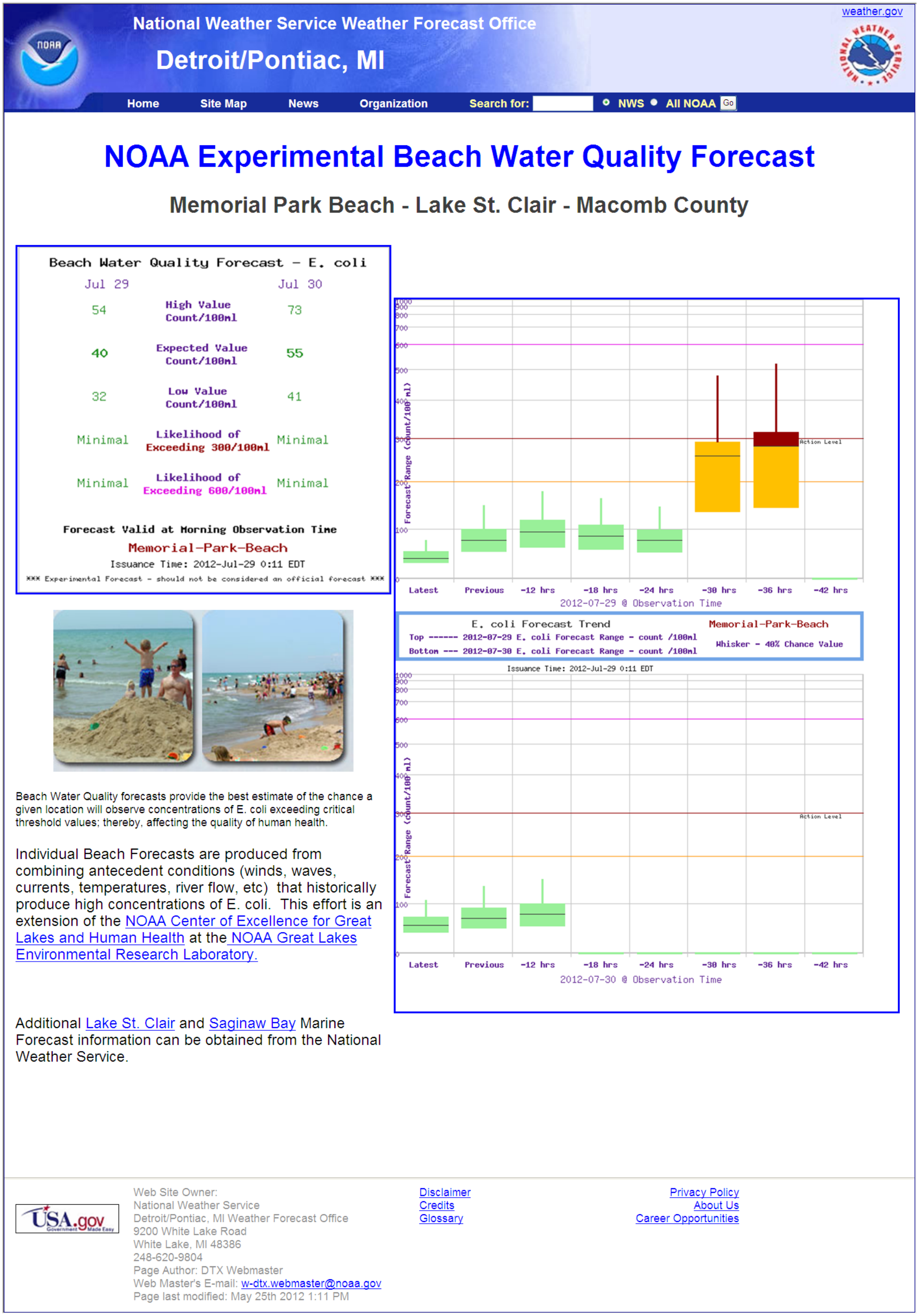


Figure 6.4.1. NOAA forecast for Memorial Beach, Macomb County, MI.

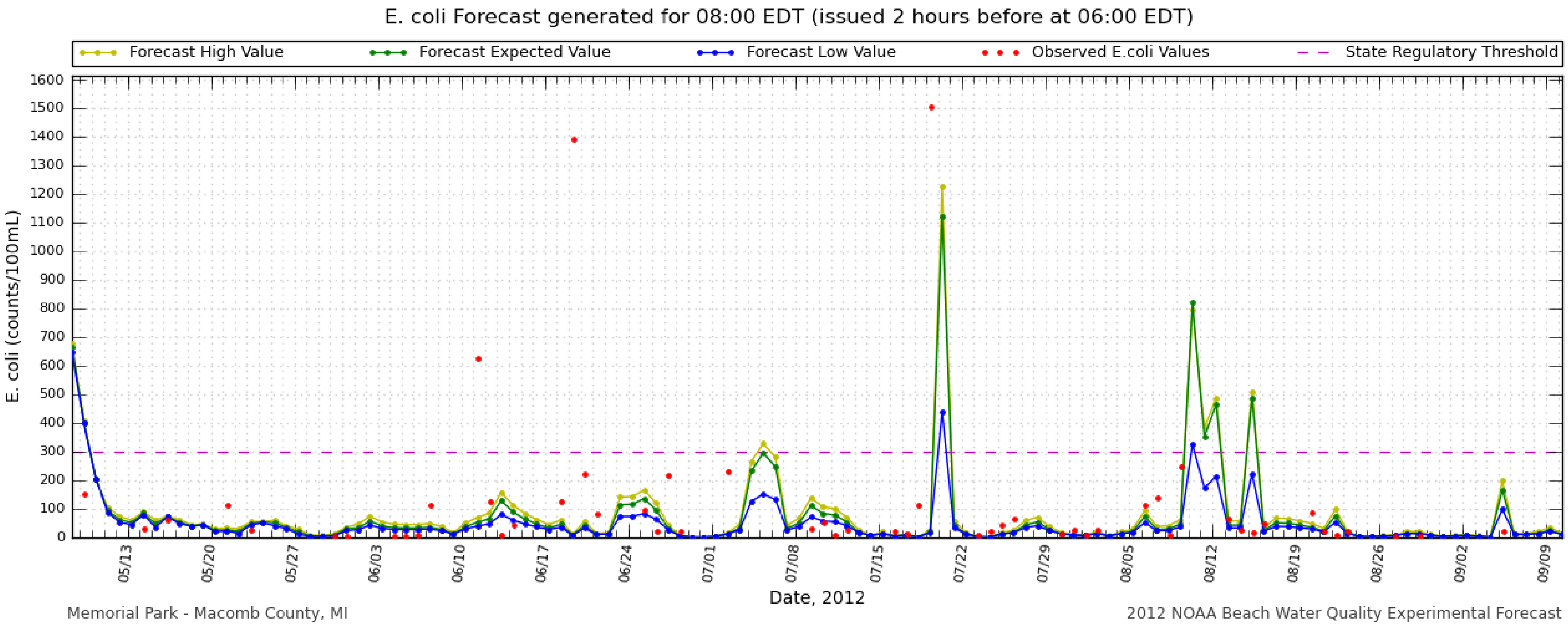


Figure 6.4.2. Memorial Beach *E. coli* forecasts for swimming season 2012.

The NWS color codes for the forecast websites illustrated in part by Figure 6.4.1 are presented in four levels depending on the likelihood of exceeding the *E. coli* threshold values of 300 counts/100 ml and 600 counts/100 ml (Minimal, Moderate, High, and Very High). Not all outcomes occur in any one forecast. These likelihood probabilities are calculated using the high *E. coli* value forecast for each time period. The expected value and low value complete the box plot.

Cross over values for each category:

300 counts/100 ml		600 counts/100 ml	
Minimal - High :	< 200	Minimal - High :	< 300
Moderate - High:	≥ 200	Moderate - High:	≥ 300
High - High:	≥ 300	High - High:	≥ 600
Very High - High:	≥ 400	Very High - High:	≥ 700

The whisker values are based on the 40% chance level and are displayed on the trend graphs. The colors on the trend graphics correspond to passing the thresholds of 200, 300, and 600 by the whisker, high value (top half of box) and expected value (lower half of box).

The color coding is based on the following:

< 200	- green
200 - 299	- orange (essentially approaching critical levels)
300 - 499	- red (exceeding critical levels)
≥ 500	- magenta (very high levels)

The whiskers are the 40% chance values based on the expected value (this is using the neighborhood in the Weibull distribution). The box is the range from the tables low, expected, and high values.

7. DISCUSSION AND CONCLUSIONS

The technical and scientific merit of this proposal is based on well-known observations that general seasonal, weather, and hydrological conditions greatly influence the physical, chemical and biological characteristics of large water bodies such as the Great Lakes. These factors in turn affect the occurrence, distribution, and survival of microbiological contaminants in the water. Forecast DSS equations use data from deterministic models. The NOAA Great Lakes Environmental Research Lab's Great Lakes Coastal Forecasting System (GLERL/GLCFS) hydrodynamic model and NOAA/National Weather Service National Digital Forecast Database (NWS/NDFD) meteorological model provide variables which quantify influences common to a geographical region. The goal of this study was to develop and operationally test the utility of 60 hour forecasting of beach water quality at a variety of beaches using *E. coli* as the indicator variable of bacterial contamination. The developmental goal was accomplished at 24 beaches in Lake Erie, Lake Michigan, Saginaw Bay and Lake St. Clair. The operational testing goal was accomplished at five beaches. There are 35 monitoring sites at the 24 beaches each of which is used to develop a forecast DSS. Of these 35 forecast DSS equations, 30 are in states where the state regulatory single sample maximum for *E. coli* is 235 counts/100 ml. For the five beaches in Michigan, the state regulatory single sample maximum for *E. coli* is 300 counts/100 ml. The NOAA/NWS Detroit Weather Forecast Office ran the 2012 forecast DSS equations successfully in operational forecasting mode for these beaches as part of NOAA Beach Water Quality Experimental Forecasting Initiative. The NWS office in Detroit Pontiac MI provided the beach managers four daily forecasts for each beach. Demonstration of forecast DSS at these five beaches was partially successful in 2012 due to minimal forecasting skill of seven *E. coli* events. This may be due to the beaches exhibiting a low percentage of *E. coli* events in 2012. In 2010 forecast DSS performed better when exceedances were above 5%. In addition, six forecasted elevated *E. coli* concentrations occurred during the weekends when no monitoring was done to verify if the forecast was correct.

CILER, University of Michigan developed and tested MLR predictive decision support equations in cooperation with NOAA's Center for Excellence for Great Lakes and Human Health, the National Weather Service, Bay Co. Health Dept. MI, Macomb Co. Health Dept. MI, Ottawa Co. Health Dept. MI, Indiana Dept. of Environmental Management, Northeast Ohio Regional Sewer District, Erie Co. (PA) Dept. of Health, Presque Isle State Park PA, Erie Co. (NY) Dept. of Health, New York State Office of Parks, Recreation, and Historical Preservation, Cities of Racine & Milwaukee, WI, Ozaukee Co. Health Dept. WI and the USGS Columbus Ohio Science Center.

Virtual Beach provides ten equations for consideration as potential forecast DSS equations during the cross validation check. In this study, we found the equation with the minimum square error of prediction (MSEP) produced the best accuracy when used to predict *E. coli* concentrations. This equation also had the most key variables. Each of the 30 forecast DSS equations represented a sampling site from one of 19 beaches. These sites are located in states using the single sample maximum for *E. coli* regulatory standard of 235 counts/100 m. The 30 FDSS-MV equations were evaluated together.

For the 30 sampling sites, we found forecast DSS equations for sites having 5% or fewer samples exceeding the single sample maximum, were not as accurate as the persistence model. This observation allows selection of beaches for forecasting beach water quality. A forecast DSS is more likely to be successful if the samples have more than 5% of the values exceeding the state single sample regulatory standard. Unless a beach has water quality issues exceeding the state regulatory standard more than 5% of the time, it does not appear to be a good candidate for forecasting beach water quality problems. Francy, 2009 noted that beaches are not good candidates for predictive modeling if fewer than five exceedances occur per sampling season. This criterion suggests a higher percentage of elevated *E. coli* sample results may be needed. Most health departments do not collect one hundred samples per swimming season which is approximately daily monitoring during the swimming season.

Similarly, Francy, 2009 noted beaches are also not good candidates for predictive modeling if the observations exceed the state regulatory standard for a majority of the samples. In this study we had two beaches with respectively 40% and 43% of the samples exceeding the state regulatory standard. In both cases the forecast DSS had better accuracy than the persistence model.

For the remaining 27 equations, the forecast DSS was superior to the persistence model at 19 sites (70%) with better accuracy, i.e. fewer type 1 and type 2 errors. The forecast DSS averaged 2.4 fewer false positive (type 1) errors and 1 fewer false negative (type 2) errors than the persistence model. For four sites, the forecast DSS was equivalent to the persistence model (15%) in terms of the total number of errors. For the remaining four sites, the forecast DSS had more errors than the persistence model.

In developing the key variables, we found the GLERL-GLCFS hydrodynamic model played a key role. Since the hydrodynamic model can provide data for all meteorological parameters except rainfall and wind gustiness, a beach having clear interaction with the lake would be a good candidate for the forecast DSS. Further, the hydrodynamic model provides reliable hourly data. Meteorological station data can have gaps in their records for a variety of reasons, i.e. sensor problems, budgetary problems, or data management issues.

This study would recommend not using the forecast DSS equation development for beaches where the GLERL-GLCFS hydrodynamic model cannot be used. Thus, beaches in a river system or in a bay away from the main body of the adjacent lake would not be good candidates. For a beach with a narrow tea cup opening to the lake may possibly make it inappropriate to use the hydrodynamic cell information which may be otherwise adjacent to the beach. For a beach where there are no GLERL-GLCFS cells nearby, the hydrodynamic model cannot be used.

A sufficient number of *E. coli* samples collected over one year or more is essential for forecast DSS development. In this study, all sites except one, had 100 or more samples. Frick et al. 2008 showed predictive models could be developed with fewer than 100 samples. It is generally recommended to have two years of data and around 100 samples (Francy et al. 2006b). Depending on the frequency of sampling, it will generally take more than one year to accumulate 100 or more *E. coli* samples. In these situations, you are inherently making the assumption that environmental processes are stable and the collection of *E. coli* data are representing the same processes. These assumptions include stable funding to allow consistent monitoring at the established frequency, usually on the order of 1-4 times per week using trained personnel (or adequately trained new hires) over the entire period of record.

This study demonstrated that beaches near a major tributary were influenced by the tributary. This suggests that tributary and or watershed runoff information should be available for all beaches.

Lastly, the economic impact of unnecessarily closing swimming at beaches can be large due to the wide spread use of the persistence model as the beach management decision support tool. The persistence model averages 2.4 more type 1 errors per beach than the forecast DSS. Once a beach is posted, the normal time needed to clear the posting is a minimum of two days. The number of days swimming is banned would be twice the number of days the beach has swimming posted. The forecast DSS makes one less error when advising people not to swim, i.e. shuts swimming at a beach correctly when the persistence model would have permitted swimming. The net number of additional days swimming would be permitted is 3.8 per monitored beach if the forecast DSS beach management tool was used in place of the persistence model. If the persistence model was replaced at the 227 beaches where monitored occurred two to seven times per week in 2010 (USEPA 2010b), a total 866 additional days of swimming would have been available in 2010 at Great Lake beaches. This represents about 23% of the swimming days banned in 2010 (NRDC 2011).

Rabinovici et al. 2004 estimates the loss of swimming at an Indiana Dunes beach ranged between a low ($18,859) and high ($37,030) value when mild swimmer health costs are included. Shaikh 2006 estimated the value of swimming day loss in 2004 for Chicago was $135,000. Applying the range of values for the days in 2010 when swimming was prohibited in error in the Great Lakes, the loss of value swimmers could have received might be substantial and range from 16M to 117M.

The elimination of swimming in water with high bacterial contamination would reduce the adverse health impacts for those exposed unnecessarily by use of the persistence model. However, there is no swimmer health registry available in the Great Lakes to assess adverse swimmer health impacts resulting from swimming in bacterial contaminated waters. Rabinovici et al. 2004 estimate individual costs from a mild health effects in (2000 dollars) at $250 to a moderate health effects at $1,125. The Center for Disease Control is proposing to track Human Health in the Great Lakes where such registries are not currently available. The purpose is to expand Great Lakes-associated state health department capacity to 1) detect, investigate, and report waterborne disease outbreaks and HAB events associated with inland and Great Lake beach use, 2) connect these data to other environmental data (e.g., beach sanitary surveys) through a unique identifier system, and 3) use these data to inform decision making by public health officials, beach managers, environmental agencies, researchers, and the general public. This is especially important with the decline in funding for monitoring which will result in the swimming public being less informed about the bacterial levels.

Principle conclusions and implications for future work are the following:

- At the sites studied, *E. coli* variation was influenced by weather patterns. Forecasted IVs explain up to almost 40% of the variation in beach *E. coli* concentrations.
- Using VB software, forecast DSS equations can be developed making fewer errors than the persistence model at 70% of the sites.
- This study reinforced the concept that a beach with 5% or fewer samples exceeding the State Regulatory *E. coli* standard may not be suitable for predictive modeling. This allows a beach manager to quickly determine if a forecast DSS should be attempted. The study also suggests that beaches with samples exceeding the state regulatory standard at 40 to 42% are still suitable for forecast DSS development.

- Because forecast DSS for all beaches adjacent to USGS gauged tributaries used key variables of river discharge or runoff, the development of a Great Lakes wide watershed model would be valuable. Such a model would be capable of providing water discharge and bacterial concentrations for all 121 watersheds.
- The procedures detailed in this report may be used by beach managers to develop predictive models at coastal beaches where the GLERL-GLCFS hydrodynamic data are available.
- The economic impact of unnecessarily banning swimming at Great Lake beaches is potentially substantial.
- Federal funding for beach monitoring is potentially ending in 2013. Swimmer health impacts are not currently registered in the Great Lakes. Establishment of a Great Lakes-swimmer health registry and associated state health department capacity will inform beach management decisions made by public health officials, beach managers, environmental agencies, researchers, and the general public.
- Minimal forecasting skill was observed for *E. coli* events during 2012 when Michigan's single sample standard was exceeded 7 times in 191 samples.

8. ACKNOWLEDGMENTS

The authors thank all the personnel involved in collecting, analyzing, and providing the *E. coli* data from the county health departments, state parks, and cities. Without this support there would have been no data. We want to acknowledge the fine working relationship with the following personnel who provided the *E. coli* data to us. This data was received from Robert Hill, Joel Strasz, Bay County Health Department, MI; Adeline Hambley, Vijayavel Kannappan, Ottawa County Health Department, MI;Steve Lichota, Cole Shoemaker, Stacy Mcfarlane, Laura Pobanz, Macomb County Health Department, MI, Michelle Caldwell, Indiana Dept. of Environmental Management, IN; Mark Citrigilia, Northeast Ohio Regional Sewer District, OH; Doug Range, Erie Co. Department of Health, and Presque Isle State Park PA; John Finster, Erie Co. Department of Health, NY; Kristen Husson, New York State, Office of Parks, Recreation, and Historical Preservation; Adam Mednick, Department of Natural Resources, WI. Partial Funding Support for this work came from USEPA's Great Lakes Restoration Initiative Grant GL-00E00658 capably managed by Ed Pniak, GLRI Project Officer, Michigan Project Lead, U.S. EPA, Region 5, Water Division, State and Tribal Programs Branch. We thank Holly Wirick for her review.

9. REFERENCES

Anderson, E.J., D.J. Schwab, and G. A. Lang (2010). Real-time hydraulic and hydrodynamic model of the St. Clair River, Lake St. Clair, Detroit River system. *J. Hydraul. Eng*. 136(8), 507–518.

Anderson, E.J., and D.J. Schwab (2011). Relationships between wind driven and hydraulic flow in Lake St. Clair and the St. Clair River Delta. *J. Great Lakes Res*. 37:147–158.

Byappanahalli, M,. M. Fowler, D. Shively, and R. Whitman (2003). Ubiquity and Persistence of *Escherichia coli* in a Midwestern Coastal Stream. *Applied and Environmental Microbiology* 69(8):4549-4555.

Breitenbach, C. (2012). Chicago Tribune, Chicago Beaches *E. coli* Bans Eliminated: Park District Transitions To Bacteria 'Advisories', May 24, 2012.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Lawrence Erlbaum Associations, Publishers, Hillsdale, New Jersey.

Cyterski, M., M. Galvin, R. Parmar, and K. Wolfe (2012). Virtual Beach User's Manual, Version 2.2. EPA/600/R-12/024. USEPA Office of Research and Development, Washington, DC.

Doswell III, C.A., D. Davis-Jones, and D.L. Keller, 1990. On summary measures of skill in rare events forecasting based on contingency tables. *Weather and Forecasting* 5:576-585.

Findlay, X., Giroud-Bougard, C. Losier, and A. Rasiulis (2009). Meta analysis of beach closure predictive models in freshwater systems. Report For the Recreational Water Quality Work Group, International Joint Commission, 24 pp.

Francy, D.S., and R.A. Darner (1998). Factors affecting *Escherichia coli* concentrations at Lake Erie public bathing beaches: U.S. Geological Survey Water-Resources Investigations Report 98-4241, 41 pp.

Francy, D.S., and R.A. Darner (2002). Forecasting bacteria levels at bathing beaches in Ohio. U.S. Geological Survey, Fact Sheet 132-02, 4 pp.

Francy, D.S. and A.M. Gifford (2002). *Escherichia coli* in the swash zone at four Ohio bathing beaches. U.S. Geological Survey Fact Sheet FS-134-02, 4 pp.

Francy, D.S., and R.A. Darner (2006a). Procedures for developing models to predict exceedances of recreational water-quality standards at coastal beaches. U. S. Geological Survey, Techniques and Methods 6-5B, 34 pp.

Francy, D.S., R.A. Darner, and E.E. Bertke (2006b). Models for predicting recreational water quality at Lake Erie beaches. U.S. Geological Survey, Scientific Investigations Report 2006-5192, 13 pp.

Francy, D.S., and R.A. Darner (2007). Nowcasting beach advisories at Ohio Lake Erie beaches. U.S. Geological Survey Open File Report 2007-1427.

Francy, D.S. (2009). Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. *Aquatic Ecosystem Health and Management* 12(2):177-182.

Frick, W.E., Z. Ge, and R.G. Zepp (2008). Nowcasting and forecasting concentrations of biological contaminants at beaches: A feasibility and case study. *Environ. Sci. Technol.* 42:4818-4824.

Ge, Z, and W.E. Frick (2007). Some statistical issues related to multiple linear regression modeling of beach bacteria concentrations. *Environmental Research* 103:358-364.

Holtschlag, D. J., D. Shively, R.L. Whitman, S.K. Haack, L.R. Fogarty (2008). Environmental factors and flow paths related to *Escherichia coli* Concentrations at two beaches on Lake St. Clair, Michigan 2002-2005. U.S. Geological Survey Scientific Report 2008-5028, 38 pp.

Kuntz, J.E. (1998). Non-point Sources of Bacteria at Beaches. City of Stamford Health Department, Stamford, Connecticut.

NRDC (2011). Natural Resources Defense Council Testing the Waters 2011, downloaded July 17. http://www.nrdc.org/water/oceans/ttw/greatlakes.pdf.

Nevers, M.B., and R.L. Whitman (2005a). Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan. *Water Research* 39:5250-5260.

Nevers, M.B. and R.L. Whitman (2005b). Empirical evaluation of the influences of hydrometeorology, Kintzele Ditch, and Trail Creek on Central Avenue and Mount Baldy Beaches' *E. coli* Concentrations. Report to IDNL.

Nevers, M.B., R.L. Whitman, W.A. Frick, and Z. Ge (2007). Interaction and influence of two creeks on *E. coli* concentrations of nearby beaches: Exploration of predictability and mechanisms. *J. Environmental Quality* 36:1337-1345.

Nevers, M.B., and R.L. Whitman (2008). Coastal strategies to predict *E. coli* concentrations for beaches along a 35 km stretch of southern Lake Michigan. *Environ. Sci. Technol.* 42:4454-4460.

Nevers, M.B., D.A. Shively, G.T. Kleinheinz, C.M. McDermott, W. Schuster, V. Chomeau, and R.L. Whitman (2009). Geographic relatedness and predictability of *Escherichia coli* along a peninsular beach complex of Lake Michigan. *J. Environ. Qual.* 38:2357–2364.

Olyphant, G.A. (2003). Beach closure decisions: the case for developing and using predictive models. *Lake Line* 23(2):19-22.

Olyphant, G.A., J. Thomas, R.L. Whitman, and D. Harper (2003). Characterization and statistical modeling of bacterial (*Escherichia coli*) outflows from watersheds that discharge into southern Lake Michigan. *Environmental Monitoring and Assessment* 81(1-3):289-300.

Olyphant, G.A, and R.L. Whitman (2004). Elements of a predictive model for determining beach closures on a real time base: The case of 63rd Street Beach Chicago. *Environmental Monitoring and Assessmen*t 98:175-190.

Olyphant, G.A. (2005). Statistical basis for predicting the need for bacterially induced beach closures: Emergence of a paradigm. *Water Research* 39(20):4953-4960.

Przybyla-Kelly, K., R.L. Whitman, and M.B. Nevers (2008). Examination of *E. coli* fluctuations, sources, and predictive modeling potential at Fair Haven Beach State Park, NY. Report to New York State Parks, Recreation, and Historic Preservation, Karen Terbush and Kristen Husson, Albany, NY, 39 pp.

Rabinovici, S.J.M., R.L. Bernknopf, A.M. Wein, D.L. Coursey, R.L. Whitman (2004). Economic and health risk trade-offs of swim closures at a Lake Michigan beach. *Environ. Sci. Technol.* 38(10):2737–2745.

Rutledge, A.T. (1998). Computer programs for describing the recession of ground-water discharge and for estimating mean ground-water recharge and discharge from streamflow data–update. U.S. Geological Survey, Water-Resources Investigations Report 98-4148, 43 pp.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics* 6(2):461–464.

Setty, K. (2012). Adapting and adopting rapid molecular methods for beach water quality monitoring. *Urban Coast* 3:48-53.

Shaikh, S.L. (2006). Fact Sheet: The value of Chicago beaches, University of Chicago, Chicago, IL.

US EPA (1999). 1999 review of potential modeling tools and Approaches to dupport the BEACH Program Office of Science and Technology, USEPA 823-R-99-002.

US EPA (2010a). Predictive modeling at beaches, Volume II: Predictive tools for beach notification. EPA-600-R-10-176.

US EPA (2010b). State Beach Data, http://water.epa.gov/type/oceb/beaches/seasons_2010_index.cfm.

Veley, R.J., D.S. Francy, and R.A. Darner (1998). How do we determine when the beaches are safe for swimming? U.S. Geological Survey, Fact Sheet FS-112-98, 4 pp.

Vermette, Niziol, Drakes (2008). Assessment of beach closures attributed to rainfall using archived radar data: Erie County, New York, April 30, 2008.

Waschbusch, R., S. Corsi, K. Sorsa, J. Walker, J. Standridge, and T. Schnieder (2004). Data collection and modeling of enteric pathogens, fecal indicators, and real-time environmental data at Madison, Wisconsin recreational beaches for timely public access to water quality information. Final Report for the EMPACT Project R-82933901-0 USEPA Project Officer Madalene Stevens, 84 pp.

Whitman, R.L. and M.B. Nevers (2004). Regional and local factors affecting patterns of *E. coli* distribution in southern Lake Michigan. Project Report on Grant GL-98500001, GLNPO-MIRB David Rockwell Project Officer.

Whitman, R.L. and M.B Nevers (2008). Summer *E. coli* patterns and responses along 23 Chicago beaches. *Environmental Science and Technology* 42(24):9217-9224.

Zimmerman, T.M. (2006). Monitoring and modeling to predict *E. coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania, U.S. Geological Survey Scientific Investigations Report 2006-5159, 15 pp.

Zimmerman, T.M. (2008). Modeling to predict *E. coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania. U.S. Geological Survey Scientific Investigations Report 2008-5039, 13 pp.

APPENDIX 1: HYDRODYNAMIC GRID CELL LOCATIONS

Figure A1: Bay City State Recreation Area, Bay County, MI and Hydrodynamic Grid Cell Locations.

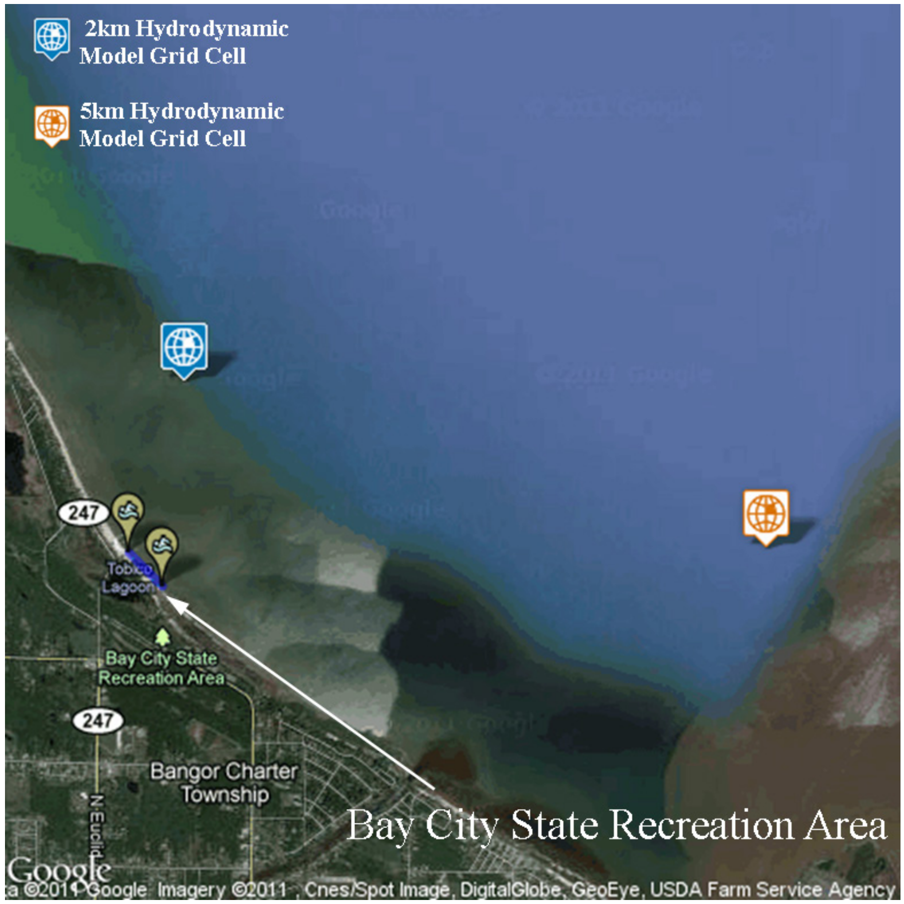


Figure A2: North Beach Park Beach, Ottawa County, MI and Hydrodynamic Grid Cell Locations.

Figure A3: Grand Haven State Park Beach, Ottawa County, MI and Hydrodynamic Grid Cell Locations.

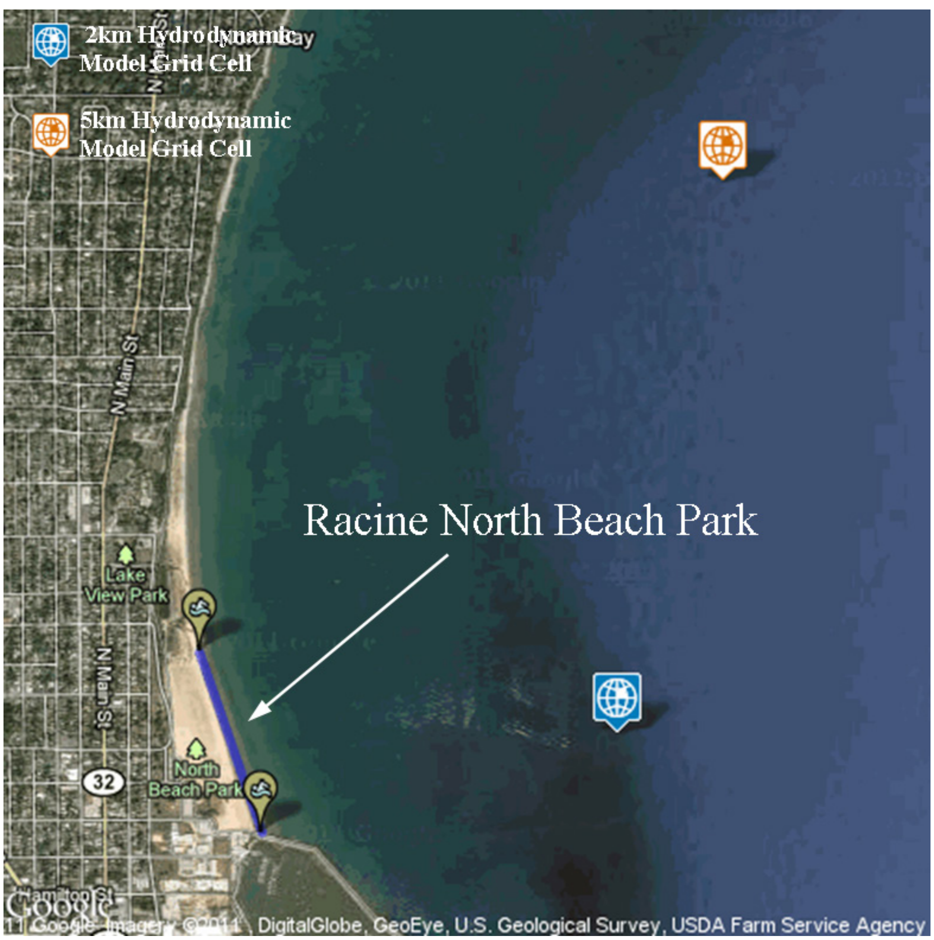


Figure A4: North Beach, Racine County WI and Hydrodynamic Grid Cell Locations.

Figure A5: Zoo Beach, Racine County, WI and Hydrodynamic Grid Cell Locations.

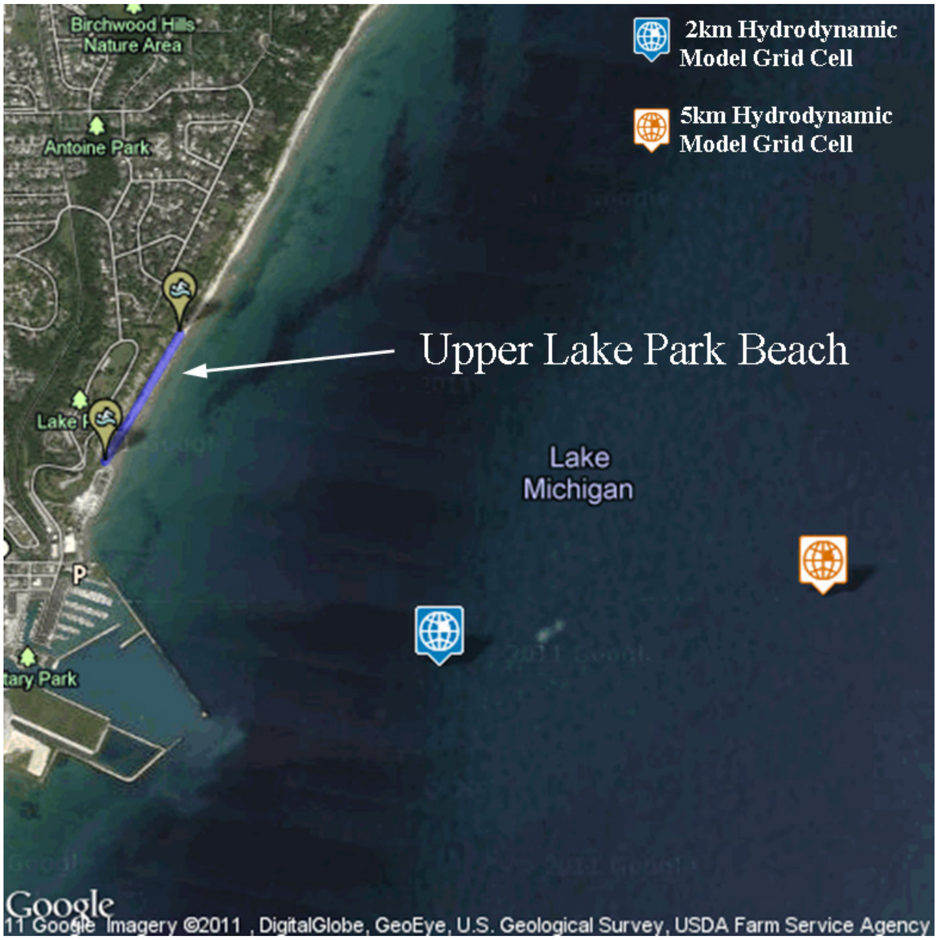
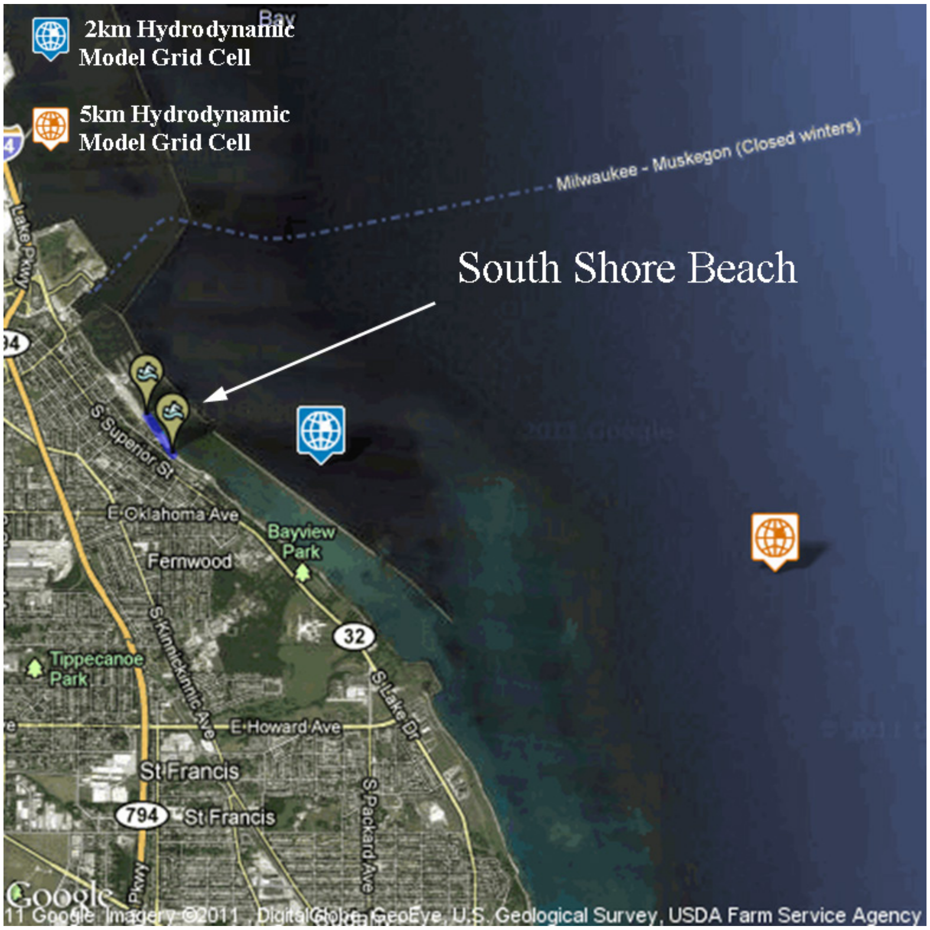


Figure A6: Upper Lake Park Beach, Ozaukee County, WI and Hydrodynamic Grid Cell Locations.



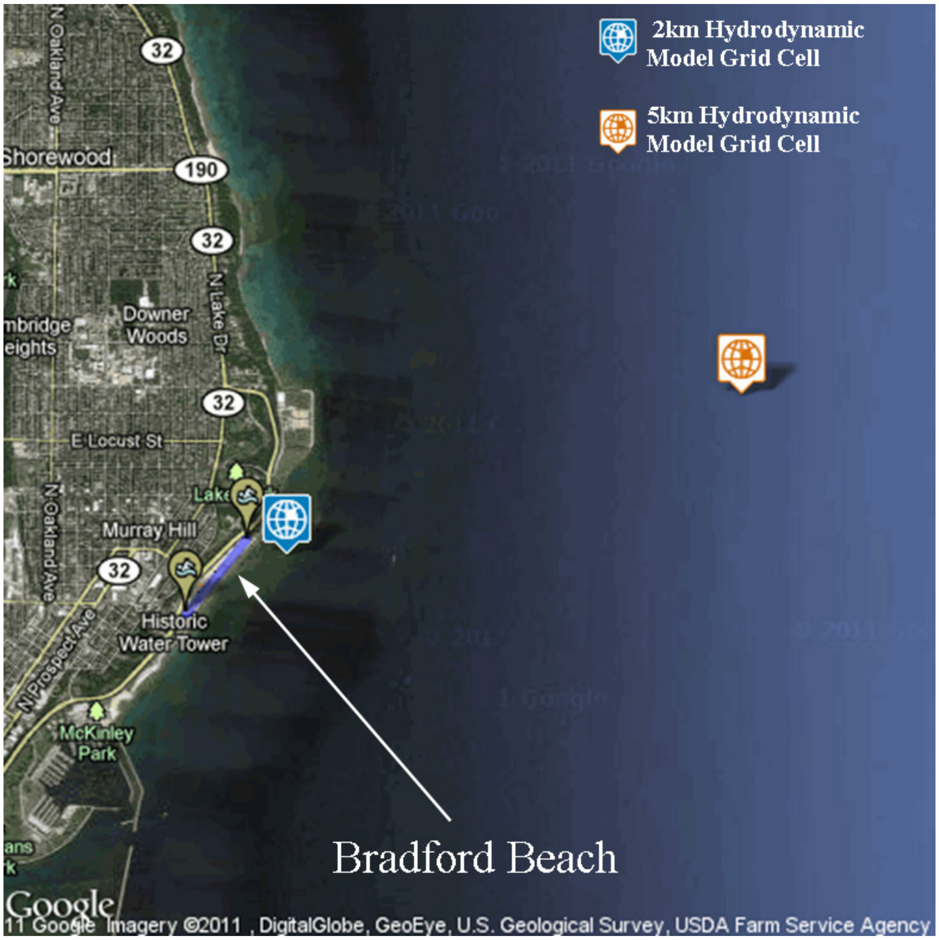


Figure A7: South Shore Beach, Milwaukee County, WI and Hydrodynamic Grid Cell Locations.

Figure A8: Bradford Beach, Milwaukee County, WI and Hydrodynamic Grid Cell Locations.

Figure A9: Woodlawn Beach State Park, Erie County, NY and Hydrodynamic Grid Cell Locations.

Figure A10: Hamburg Bathing Beach, Erie County, NY and Hydrodynamic Grid Cell Locations.

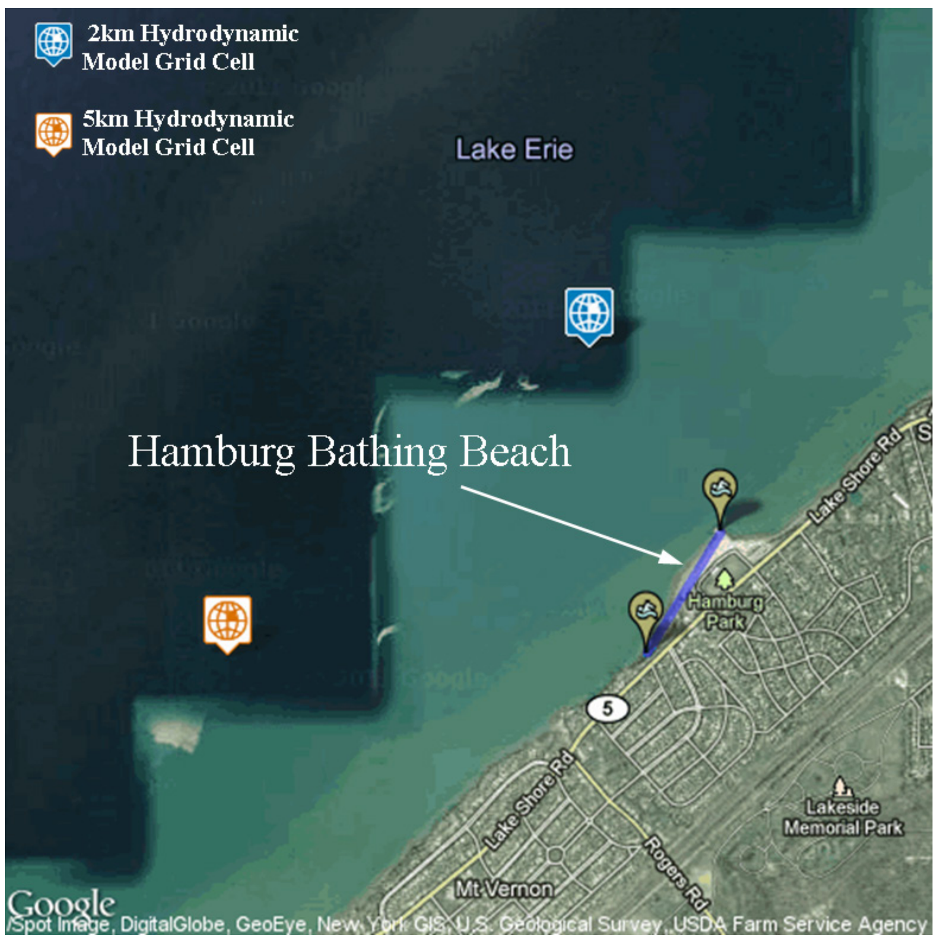


Figure A11: Evans Town Park Beach, Erie County, NY and Hydrodynamic Grid Cell Locations.

Figure A12: Wendt Beach, Erie County, NY and Hydrodynamic Grid Cell Locations.

Figure A13: Lake Erie Beach, Erie County, NY and Hydrodynamic Grid Cell Locations.

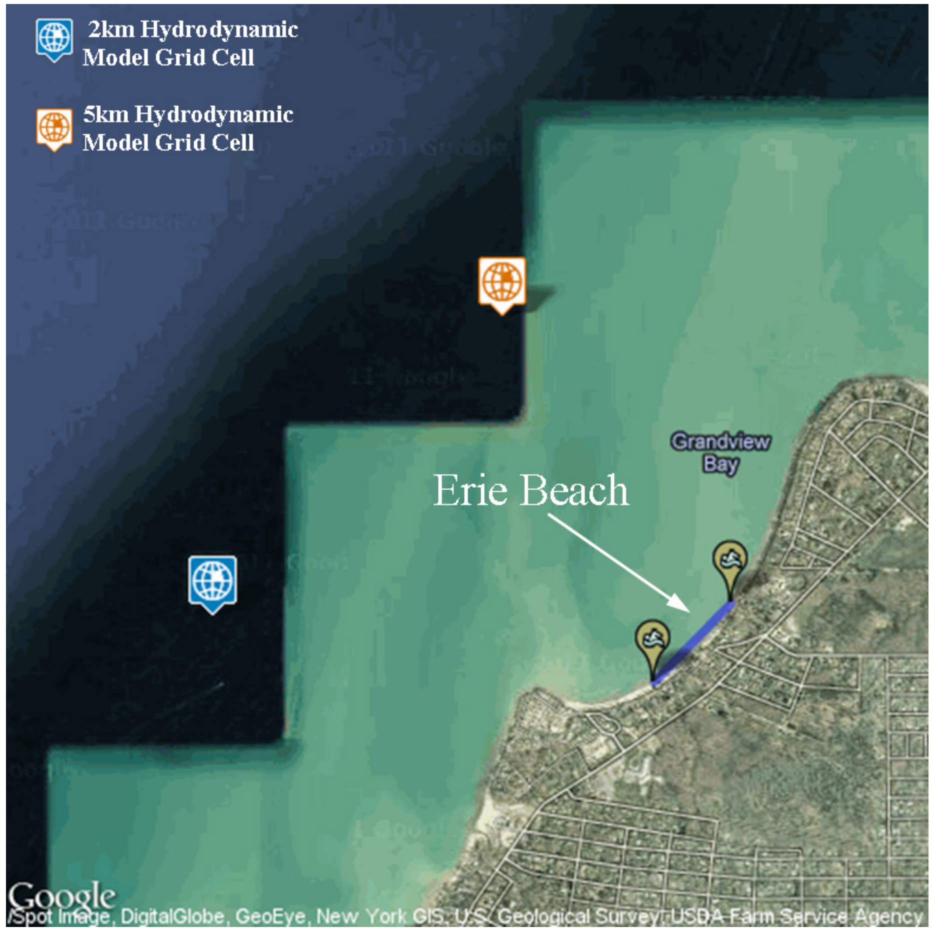
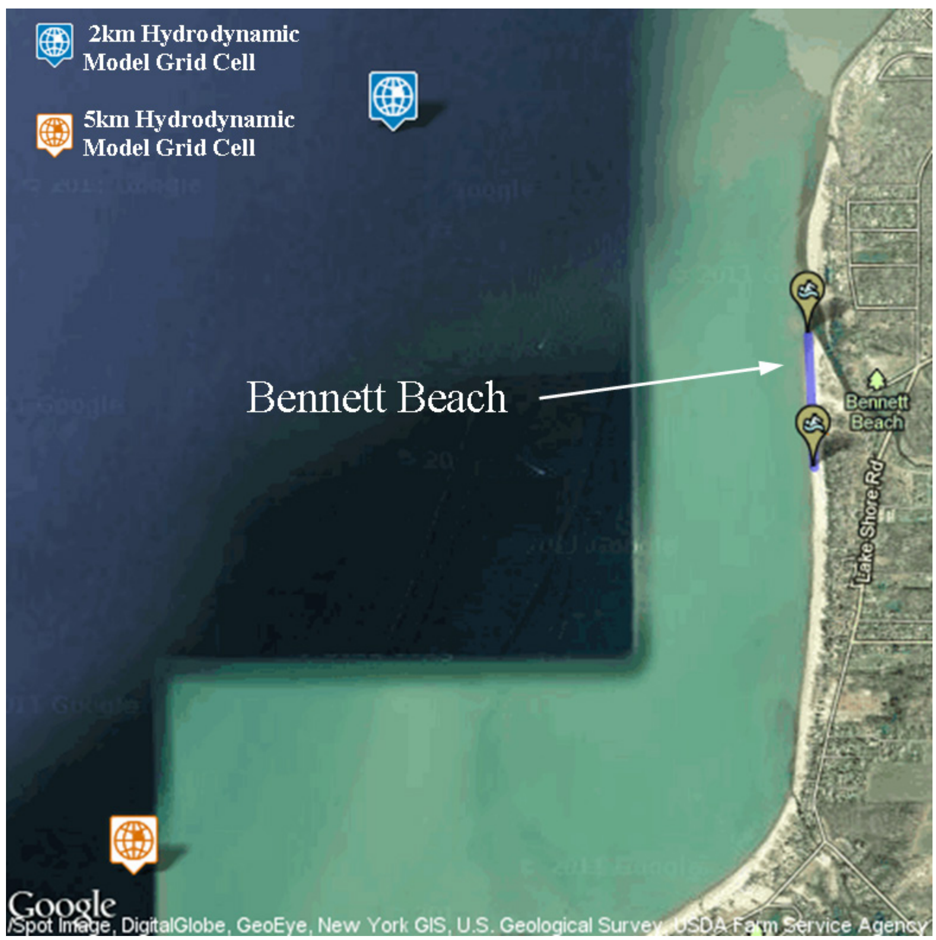


Figure 11.14: Bennett Beach, Erie, County, NY and Hydrodynamic Grid Cell Locations.



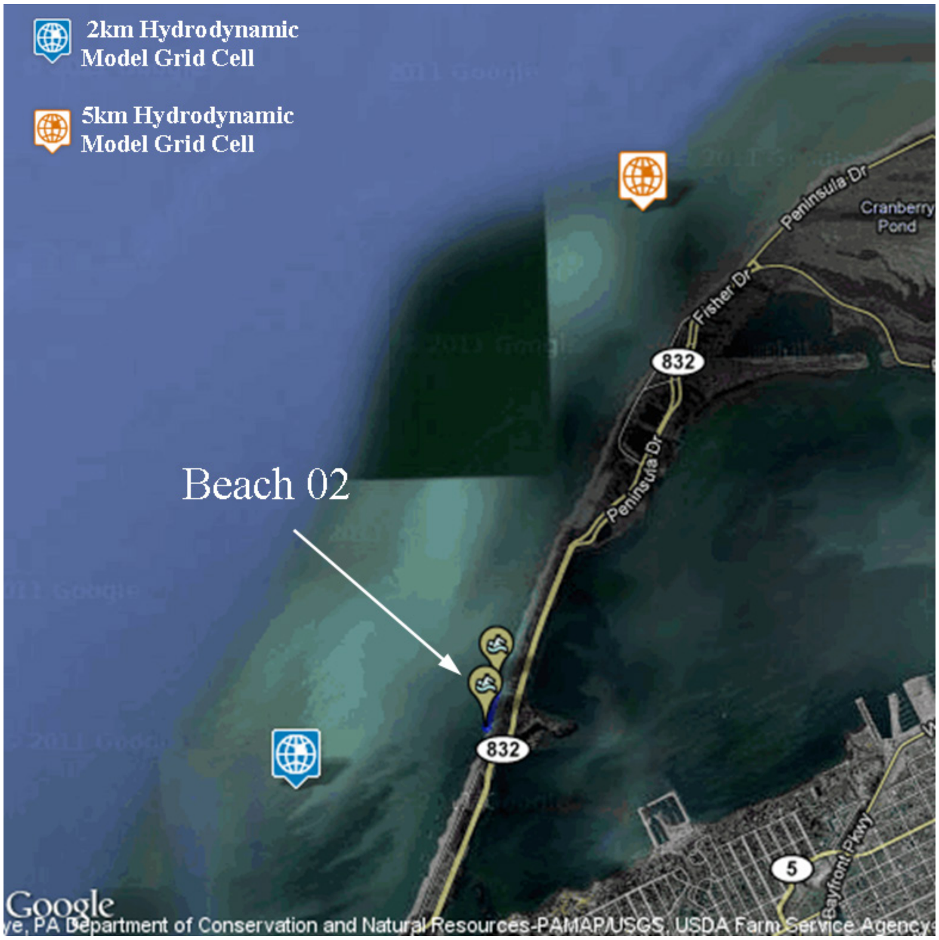
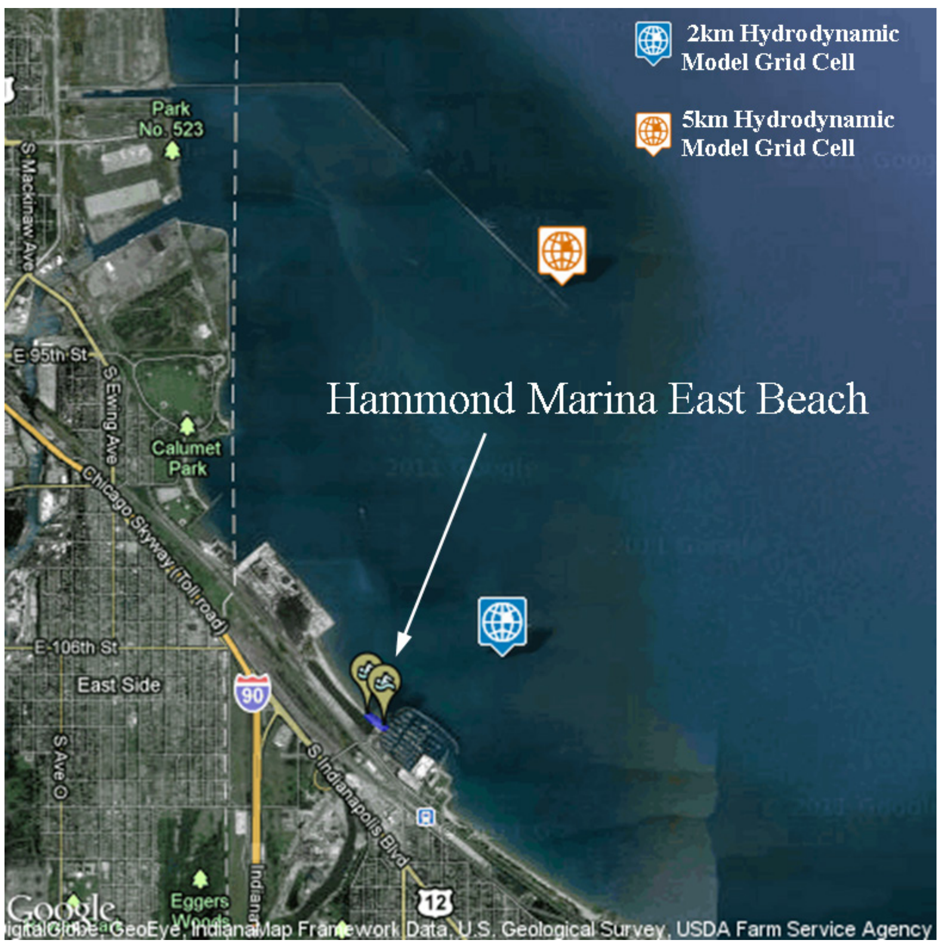


Figure 11.15: Presque Isle Beach #2, Erie, County, PA and Hydrodynamic Grid Cell Locations.

Figure 11.16: Presque Isle Beach #10, Erie, County, PA and Hydrodynamic Grid Cell Locations.

Figure 11.17: Villa Angela, Cuyahoga County, OH and Hydrodynamic Grid Cell Locations.

Figure 11.18: Hammond Marina East Beach, Lake County, IN and Hydrodynamic Grid Cell Locations.



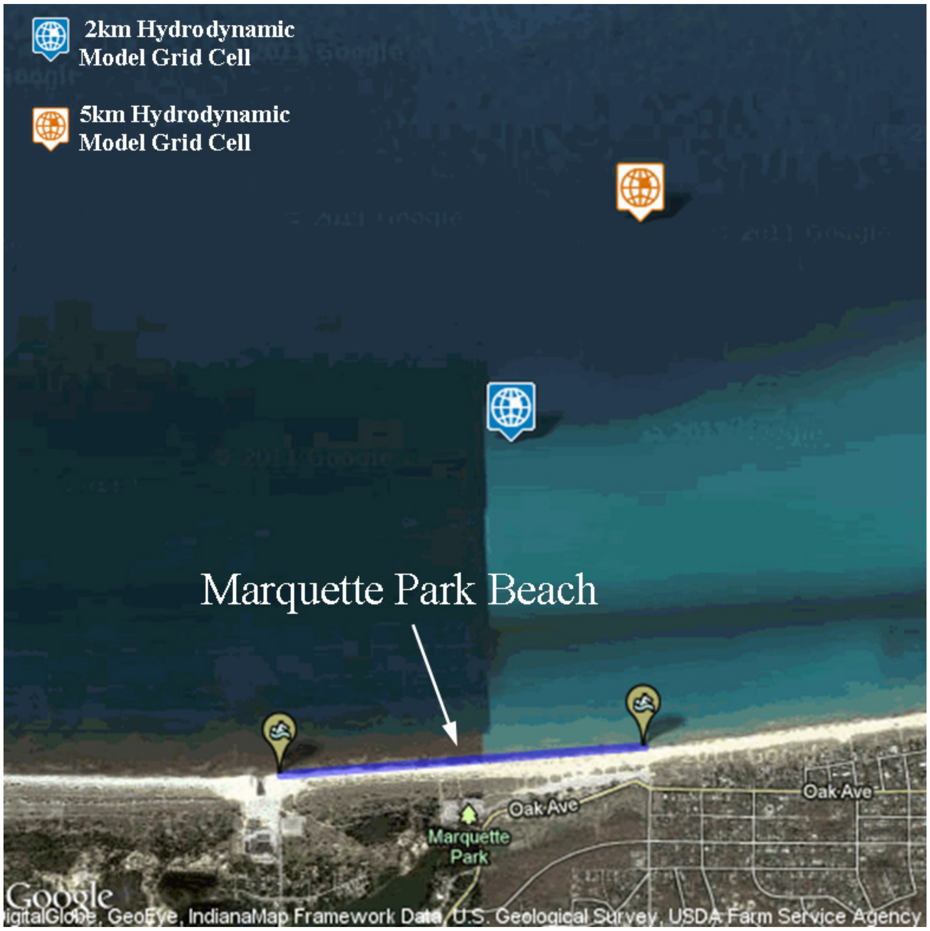


Figure 11.19: Marquette Park Beach, Lake County, IN and Hydrodynamic Grid Cell Locations.

Figure 11.20: Washington Park Beach, Lake County, IN and Hydrodynamic Grid Cell Locations.

Figure 11.21: Ogden Dunes, Lake County, IN and Hydrodynamic Grid Cell Locations.

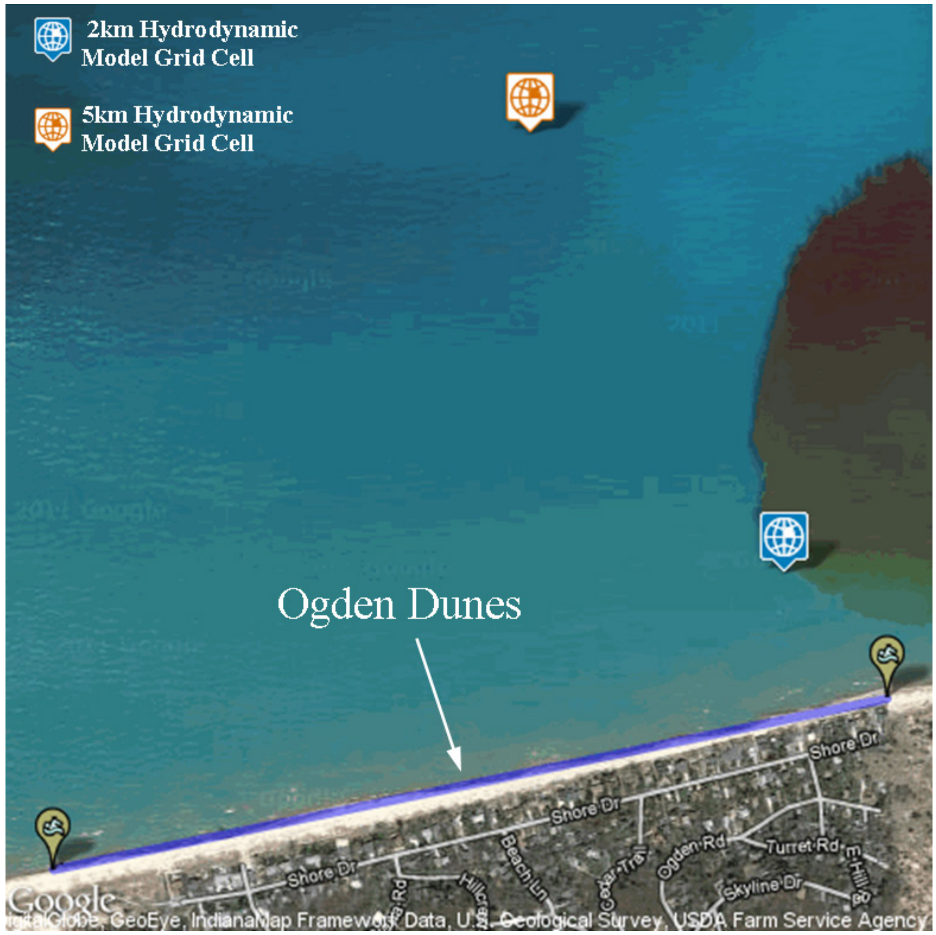


Figure 11.22: IDSP West Beach, Lake County, IN and Hydrodynamic Grid Cell Locations.

Figure 11.23: Memorial Beach, Macomb County, MI and Hydrodynamic Grid Cell Location.

Figure 11.24: Metropolitan Beach, Macomb County, MI and Hydrodynamic Grid Cell Location.