ABSTRACT

MATHEMATICAL MODELS TO PREDICT *Escherichia coli* CONCENTRATIONS AT WISCONSIN BEACHES (MULTIPLE YEAR COMPARISIONS)

By Sreenivas Manchirala

Beach water quality is determined using *Escherichia coli* as an indicator organism for fecal contamination. Traditional methods may result in improper beach closures and openings, as they base their decisions on previous day *E. coli* concentrations. To overcome these problems mathematical models were developed using various environmental and water quality variables, such as water temperature, wave height, wind speed and direction, turbidity, and rainfalls to predict *E. coli* concentrations. Mathematical models were developed using “Virtual Beach” (VB) software, which uses multiple linear regressions to determine the relationship between *E. coli* concentrations and environmental variables. Data were collected from eleven Wisconsin beaches in Ashland, Bayfield, and Door counties during summer 2008. Mathematical models were developed using 2008 and the 2007 – 2008 combined data, which were compared with 2007 models. Mathematical models were used to calculate the variability in $\log_e$ *E. coli* concentrations in terms of adjusted R^2 (coefficient of determination). The Kreher Park Beach – 2008 model was the only one, which predicted $\log_e$ *E. coli* concentrations 100% accurately, without false positives or false negatives. Single variable models were developed for Ephraim Park, Lakeside Park, and Otumba Park beaches for the 2008 recreational season. The combination of years of data (2007 & 2008) did not improve the fitness of the mathematical models (decreased adjusted R^2) for all the beaches except the Kreher Park Beach. Explanatory variables that were included in the mathematical models were unique for each beach. Therefore, it was concluded that the mathematical models were beach specific and that combination of years of data did not necessarily result in more robust mathematical models, due to annual changes in swim season physical parameters.

MATHEMATICAL MODELS TO PREDICT *Escherichia coli* CONCENTRATIONS AT WISCONSIN BEACHES (MULTIPLE YEAR COMPARISIONS)

by

Sreenivas Manchirala

A Thesis Submitted
In Partial Fulfillment of the Requirements
For the Degree of

Master of Science – Biology

at

The University of Wisconsin Oshkosh
Oshkosh WI 54901-8621

August 2010

COMMITTEE APPROVAL

_____ Advisor

8-2-10 Date Approved

_____ Member

8-2-10 Date Approved

_____ Member

8/2/2010 Date Approved

PROVOST AND VICE CHANCELLOR

_____

August 3, 2010 Date Approved

FORMAT APPROVAL

_____

7/23/10 Date Approved

ACKNOWLEDGEMENTS

First, I would like to express my deepest appreciation to my Thesis Advisor, Dr. Colleen M. McDermott, who continually encouraged me for doing my Masters program successfully. Without her guidance and persistent help this thesis would not have been possible. I would like to thank my thesis committee members, Dr. Gregory Kleinheinz and Dr. Sabrina Mueller-Spitz, who asked me good questions and gave insightful comments on my work. Special thanks go to my co-graduate, Mr. Sashidhar R. Yerram, who is most responsible for helping me complete the writing of this thesis report. Sashidhar, has been a friend and always there to meet and talk about my ideas. I extend my heartfelt thanks to the academic staff, who taught me the courses in my graduate program, at University of Wisconsin Oshkosh. In addition, I thank all the folks, who made my graduate school wonderful.

Last, but not least, I would like to thank my family: my parents, Laxmi narayana, and Sulochana for giving me life in the first place, for educating me with aspects from both arts and sciences, for support and encouragement to pursue my Masters. Finally, I would like to thank my B-Tech Biotechnology batch (2002-2006) for their friendship and encouragement over the years.

TABLE OF CONTENTS

Page #

LIST OF TABLES

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF TABLES (Continued)

Page #

LIST OF FIGURES

Page #

LIST OF FIGURES (Continued)

Page #

LIST OF FIGURES (Continued)

Page #

LIST OF FIGURES (Continued)

Page #

INTRODUCTION

The Importance of Recreational Beaches

Beaches are one of the most enjoyable summer vacation destinations for people in the United States of America. United States Environmental Protection Agency (USEPA) reported that Americans make a total of 910 million trips and spend about $44 billion on beach related tourism (United States Environmental Protection Agency [USEPA], 2000). The United States has nearly 23,000 miles of shoreline and more than 5,500 miles of Great Lakes shoreline (United States Commission on Ocean Policy, 2004). In Wisconsin, Lake Superior has over 300 miles of shoreline spanning Ashland, Bayfield, Douglas and Iron counties, all important tourist destinations (Sampson, Swiatnicki, McDermott, & Kleinheinz, 2005). Door County is one of the most popular summer tourist destinations in Wisconsin, and has more than 250 miles of Lake Michigan shoreline ((United States Environmental Protection Agency [USEPA], 2006). In the summer of 2006, the BEACH Act grant was used to monitor 28 beaches in Door County (Wisconsin Department of Natural resources [WDNR], 2006). These 28 beaches are important resources to the Door County economy.

Recreational waters include natural waters such as marine and fresh water beaches, as well as artificial waters such as swimming pools and spas. Uses of recreational water include activities ranging from whole body water contact sports such as swimming, diving and surfing to non-contact activities such as fishing, walking and picnicking (World Health Organization [WHO], 2003). However, drinking water is given more importance than recreational water quality, but water-borne disease outbreaks are due to contact with recreational waters only. In order to assure the quality of beaches, the Beaches Environmental Assessment and Coastal Health (BEACH) Act was established by the Environmental Protection Agency (EPA) in 2000.

Beaches Environmental Assessment and Coastal Health Act (BEACH Act)

Beaches are ecologically, economically and psychologically essential, because they provide recreation, relaxation, and a chance to renew spirit for all people (USEPA, 2006). Over half of the U.S. population lives in coastal areas and approximately one-half of the nation's gross domestic product ($ 4.5 trillion in 2000) is generated from these areas (US Commission on Ocean Policy, 2004). Public health is seriously affected if recreational water users contact fecal contaminated waters. The United States government recognized the need for improved protection of public health at beaches. In order to assure the beach water quality, the US Congress passed the BEACH Act (Beaches Environmental Assessment and Coastal Health Act) in October 2000. The main goal of this Act is to reduce the risk of diseases to users of the nation's coastal recreational waters (United States Congress, 2000). The BEACH act was signed into law, amending the Clean Water Act (CWA) by adding several new sections (303(i), 104(v), 304(a) etc.,) to mandate conducting studies at beaches associated with pathogens and human health (USEPA, 2006).

The objectives of the BEACH Act are to monitor microbial contamination at beaches, to better inform the public of health concerns at the beaches, and to promote scientific research for better protection of the health of recreational water users (United States Environmental Protection Agency [USEPA], 2002).

Health Concerns at the Beaches

Epidemiological studies have revealed that contact with polluted recreational water can result in adverse human health effects, such as mild to moderate gastrointestinal diseases (e.g., diarrhea) and non-gastrointestinal diseases (respiratory, eye, ear and skin infections), to more serious illnesses, such as meningitis or hepatitis (Rose et al., 1999). A study conducted at New York City beaches as part of the United States Environmental Protection Agency (USEPA) program revealed that gastrointestinal (GI) and respiratory illnesses were higher among swimmers than non-swimmers (Cabelli, Dufour, Levin, McCabe, & Haberman, 1979). Various studies have reported the adverse human health effects that might result from human exposure to fecally contaminated water (USEPA, 2006). Fecal contamination of recreational waters is due to a variety of sources, such as agricultural runoff, storm water runoff, spring snow runoff, and pet waste contamination, leakage from faulty septic systems, sanitary sewer overflows and broken sewer lines (Cabelli, Hoardley, & Dutka, 1976). Feces can contain a variety of disease-causing microorganisms, such as bacteria, viruses and protozoans. These disease-causing microorganisms generally are spread by the fecal-oral route (Henrickson, Wong, Allen, Ford, & Epstein, 2001)

Bacterial Contamination at Beaches

Fecally contaminated water can contain many types of bacteria, such as coliforms, *Streptococcus, Lactobacillus, Staphylococcus* and *Clostridium* (USEPA, 2006). Members of the Enterobacteriaceae family are most likely to cause disease due to their high concentration in human and animal waste (Kleinheinz, McDermott, & Sampson, 2003). Of the Enterobacteriaceae family, *Salmonella* and *Shigella* are the most well known pathogens identified in gastrointestinal disease outbreaks upon contact with recreational water (Lee et al., 2002; Levy et al., 1998). *Shigella sonnei* was implicated in a diarrheal outbreak in children playing in an interactive fountain at a beachside park in Florida, in 1999 (Minshew et al., 2000). In 1999, *E. coli* O157:H7, the most notorious of the *E. coli* strains was involved in a diarrheal outbreak in children who were swimming at the lake in Finland (Paunio et al., 1999). Other bacteria such as *Campylobacter jejuni* (not a member of Enterobacteriaceae) have been identified in gastrointestinal outbreak associated with recreational water (St. Louis, 1988). *Pseudomonas aeruginosa* has been isolated from hot tub and pool-associated cases of follicular dermatitis, and *Legionella pneumophila* from whirlpool tub-associated cases of pneumonia (Pontiac Fever) (Yoder, Blackburn, Craun, Hill, & Levy, 2004).

Protozoal Contamination at Beaches

Protozoans are unicellular organisms found primarily in aquatic environments (USEPA, 2006). Pathogenic protozoans, which constitute ~ 30% of 35,000 known protozoan species that, are found in the feces of human and other warm-blooded animals

(Mitchell, Mutchmor, & Dolphin, 1988; North Carolina State University, 1977). *Cryptosporidium parvum* and *Giardia intestinalis* have been involved in recreational water-borne diarrheal outbreaks (Lee et al., 2002; Yoder et al., 2004). Various studies have reported an estimated 2.5 million cases of *giardiasis* in United States of America every year in late summer (Fattal, Vasl, Katzenelson, & Shuval, 1983). *Naegleria fowleri*, also a member of the Protozoa, has been involved in primary amoebic meningoencephalitis (PAM) associated with recreational water (Yoder et al., 2004).

Viral Contamination

Viruses are a group of infectious agents that require a host in which to survive and reproduce. Viruses can cause diseases like gastroenteritis and infectious hepatitis, when individuals have contact with contaminated recreational water. The most significant viral groups affecting water quality and human health are Adenoviruses (31 types), Rotaviruses, Reoviruses, Noroviruses, and 67 types of Enteroviruses (USEPA, 2006). Norovirus and Hepatitis A virus have been associated with recreational disease outbreaks (Lee et al., 2002; Yoder et al., 2004). *Norovirus* was likely the cause of swimming related illness in Door County, Wisconsin during the 2002 swimming season (Furness, Beach, & Roberts, 2000).

Indicators of Fecal Contamination

It is not practical to monitor all of the pathogenic microbes potentially found in surface water because these organisms are found in low number in fecal material and the techniques to monitor these pathogens are difficult and expensive. An ideal fecal indicator organism should be a microbe found in high concentrations in human and animal feces. The detection methods for the indicator should be easy, simple and cost-effective. The indicator microbe must survive in the secondary environment and should indicate recent fecal contamination (Sayler, Nelson, Justice, & Colwell, 1975; United States Geological Survey [USGS], 2006). Beginning of 1976, fecal coliforms were used to evaluate microbial contamination in recreational waters (United States Environmental Protection Agency [USEPA], 1976). A series of studies performed at New York City beaches revealed that a strong correlation was observed between swimming-associated gastrointestinal symptoms and *E. coli* and enterococci densities in waters, but not with fecal coliform concentrations (Cabelli et al., 1979). Enterococci and *E. coli* concentrations, however, have shown better correlation with potential for disease in recreational water users (Cabelli et al., 1976).

Therefore, the Environmental Protection Agency (EPA) recommended enterococci (*Enterococcus faecalis* and *Enterococcus faecium*) and *E. coli* as indicators of fecal contamination of recreational waters. *Bacteroides,* an anaerobic bacterium, also was considered for use as an indicator organism due to its high concentration in feces (Allsop & Stickler, 1985; Fiksdal, Make, LaCroix, & Staley, 1985), but isolation and identification of this organism is technically difficult.

***Escherichia coli* Evaluation**

The State of Wisconsin implemented the BEACH Act recommendations in 2003 for *E. coli* monitoring at public, Great Lakes beaches. If *E. coli* concentrations are between 0 – 234 CFU/ 100mL (colony forming units per 100 milliliter of water), this indicates that the recreational water quality is safe for swimming and the beach remain open. If *E. coli* concentrations are between 235 – 999 CFU/ 100mL of water, this indicates that a poor water quality advisory should be issued. If *E. coli* concentration are above 1000 CFU/ 100mL of water, this indicates that the beach should be closed for recreational use until concentrations decline.

Monitoring Techniques

There are two traditional methods to evaluate *E. coli* concentrations in recreational waters, classical membrane filtration (MF) and defined substrate (DS) methods.

Membrane filtration (MF). This is the older and most prevalently used method. In this method, 100mL of water is filtered through a sterile 0.45μm membrane filter, which retains the bacteria. After filtration, the membrane filter retaining the target bacteria is placed onto a selective and differential medium, such as modified mTEC (membrane thermo-tolerant *E. coli*) agar (United States Environmental Protection Agency [USEPA], 2002). The modified mTEC (Difco) plates are incubated for 2 hours at 35±0.5°C and then plates are transferred to a 44.5±0.5°C incubator and incubated for 22 hours. After 24 hours of incubation, red or magenta colonies are counted and expressed as colony forming units (CFU) per 100mL of water sample.

The principle behind the membrane filtration method is that the modified mTEC agar medium (Difco) contains 5-bromo-6-chloro-3-indolyl-β-D-glucuronide, a chromogenic substance catabolized by β-D-glucuronidase, an enzyme from *E. coli,* into glucuronic acid, which responsible for red or magenta colored colonies. This method has disadvantages. It is time consuming, error-prone, laborious, and expensive.

Defined substrate (DS) technology. It is a newer and alternative method for enumerating *E. coli* concentrations in water samples. It is a two-step assay that detects the presence of total coliforms and *E. coli* simultaneously (Colilert^{R}, IDEXX Corp., Portland, ME). In this method, 100mL of water sample is mixed with reagent (e.g. Colilert) and the sample is poured into an IDEXX Quanti-Tray/2000 which is incubated for 24 hours at 35°C. After 24 hours of incubation, the yellow color in wells in the Quanti-tray indicates the presence of total coliforms and yellow wells that fluorescence under UV light at 366nm indicates the presence of *E. coli,* expressed as most probable number (MPN)/100 mL (Ederberg, Allen, Smith, & Kriz, 1990). This method is a relatively uncomplicated process, less error-prone but requires 24 hours for enumeration of *E. coli* numbers.

The principle behind the defined substrate technology is that the Colilert reagent contains ortho-nitrophenyl-β-D-galactopyranoside (ONPG), which is turned from colorless to yellow by the β–galactosidase enzyme, indicating the presence of coliforms in the water sample. In addition, the 4-methylumbelliferyl-β-glucuronide (MUG) is metabolized by the β-glucuronidase enzyme and creates fluorescence under UV light, indicating the presence of *E. coli* (Paunio et al., 1999).

The Drawbacks of MF and DS methods

These two methods require at least 24 hours for enumerating *E. coli* concentrations. This may result in improper beach closures and openings. Beaches may be closed when they should have remained open or beaches may be open when they should have been closed. The beach advisories are issued based on previous day's *E. coli* concentrations indicating the problem with the culture based assessments of water quality. This may adversely affect public health or needlessly result in loss of tourism revenue. In order to reduce the time to evaluate coliform and *E. coli* concentrations, multiplex Polymerase Chain Reaction (PCR) and gene probe detection targeting the *lacZ* and *uidA* genes can be used (Bej, McCarty, & Atlas, 1991). This is a rapid method that takes approximately 2-3 hours to determine *E. coli* concentrations, but requires skilled personnel and sophisticated technology. It is not yet approved by the US government for enumeration of *E. coli* under the BEACH Act.

In order to overcome these problems, predictive mathematical models have been developed to estimate the *E. coli* concentrations, and thereby to assess beach water quality. Predictive modeling is a real-time prediction method uses multiple linear regressions to evaluate *E. coli* concentrations based on environmental and water quality variables such as water temperature, air temperature, wave height, turbidity, rainfall variables, and amount of algae etc., at the beach or weather forecasting websites. The steps to develop predictive models were data collection, data analysis, model development, followed by model diagnostics and selection. The predictive models were used to calculate the variability in *E. coli* concentrations. Predictive models are useful because they can guide decision makers during the period of time between sampling the beach water and obtaining water quality results (USEPA, 2002).

Previous Studies

In Southern California, at Santa Monica Bay beaches, a study was done to investigate the relationship between rainfall and *E. coli* concentrations in beach waters (Ackerman & Weisberg, 2003). The rainfall amounts were categorized into <2.5mm, 2.5 - 6mm, 6 - 25mm, and >25mm and the study showed that rainfall greater than 6mm was associated with elevated *E. coli* concentrations. County Health Departments now warn the public to avoid contact with beach water for three days after such rainfalls. The statistical analyses done for this study showed satisfactory results using multiple linear regressions to forecast beach water quality and have given insight to other researchers to develop predictive models using multiple linear regressions (Ackerman & Weisberg, 2003).

A study was done at 63rd Street Beach, Chicago, Illinois (Olyphant & Whitman, 2004), to develop a mathematical model to predict *E. coli* concentrations in beach water. The predictive model was developed from data collected on 57 occasions during the summer of 2000, using hydrological, meteorological and water quality variables as predictive variables. Variables included in the mathematical model to predict *E. coli* concentrations were wind direction and speed, rainfall, insolation (incoming solar radiation), lake stage, water temperature, and turbidity. The predictive model accounted for 71% variability in the log transformed *E. coli* concentrations with 88% accuracy (the percentage of correct predictions (Olyphant & Whitman, 2004).

A similar study was conducted at Huntington Beach, Bay Village, Ohio (Francy, Darner, & Bertke, 2006), and developed a mathematical model using data collected during 2000 – 2004 recreational seasons. In total, 248 water samples were collected and the model developed used explanatory variables such as wave height, weighted rainfall 48 hours (Rw 48) and $\log_{10}$ turbidity to predict *E. coli* concentrations. Mathematical models accounted for 38% of variability in the log *E. coli* concentrations with 84.7% accuracy. The model predicted 210 total correct predictions, which include correct non-exceedances and exceedances, 20 false positives, and 18 false negatives. The group also developed mathematical models for Edgewater, Villa Angela, Lakeview and Lakeshore beaches for the 2004 – 2005 recreational seasons (Francy et al., 2006).

Another study was done at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania (Zimmerman, 2006), and developed a mathematical model using data collected during the 2004 – 2005 recreational seasons. Mathematical models were developed for 2004, 2005 and combined 2004 – 2005 year data sets. Variables included in the combined 2004-2005 model were $\log_{10}$ turbidity, rain weight (sum of weighted rainfall amounts), wave height, and wind direction, accounting for 64% of log transformed *E. coli* variation with 94% accuracy. The combined 2004-2005 mathematical model predicted 156 correct predictions, 4 false negatives, and 6 false negatives (Francy, Darner, & Bertke, 2006).

All these predictive models described above have shown satisfactory results in forecasting beach water quality and have given insight to other researchers to develop predictive models for their local beaches. Hence, our goal was to develop predictive models for various beaches in Wisconsin from Ashland, Bayfield, and Door Counties using environmental and water-quality variables data collected during the recreational season 2008, from May 15th to August 31st.

The specific objectives of the study are,

1. To investigate the relationship between environmental variables and *E. coli* concentrations at selected beaches in Ashland, Bayfield, Door, and Manitowoc counties.
2. To develop predictive models, using multiple linear regressions for 2008 data, and combined 2007-2008 data from these beaches.
3. To compare the predictive models for each beach for 2007, 2008, and 2007-2008 combined data sets.
4. To compare the variables that are useful for predicting *E. coli* concentrations at each beach for similarities and differences and determine if a single model can be used for a variety of beaches.
5. To determine if combining data sets from multiple years will result in production of a more successful predictive model.

MATERIALS AND METHODS

Sites of Study

The Wisconsin Department of Natural Resources (WDNR) operates Wisconsin's Beach Monitoring Program to measure bacterial concentrations from beach water along Lake Michigan and Lake Superior. WDNR identified 195 public beaches along Lake Michigan and Lake Superior. Ashland and Bayfield counties have ~ 200 miles of Lake Superior shoreline and Door County has more than 250 miles of Lake Michigan shoreline (USEPA, 2006). These counties are among the most popular summer vacation destinations in Wisconsin. The local health departments from the selected counties and University of Wisconsin Oshkosh worked together to monitor beach water quality.

The sites selected for the study were nine beaches along the Lake Michigan shoreline and eight beaches along the Lake Superior shoreline. Eight beaches were selected from Door County, WI on the Lake Michigan shore (Figure 1). These included Bailey's Harbor, Ephraim, Fish Creek, Lakeside Park, Murphy Park, Otumba Park, Sunset Park and Whitefish Dunes State Park from Door County (Figure 2). In addition, three selected beaches along the Lake Superior shoreline were Kreher Park and Maslowski Park from Ashland County and Thompson's West End Park beach from Bayfield County, WI (Figure 1). These eleven study beaches were chosen for monitoring *E. coli* concentrations and predictive model development.

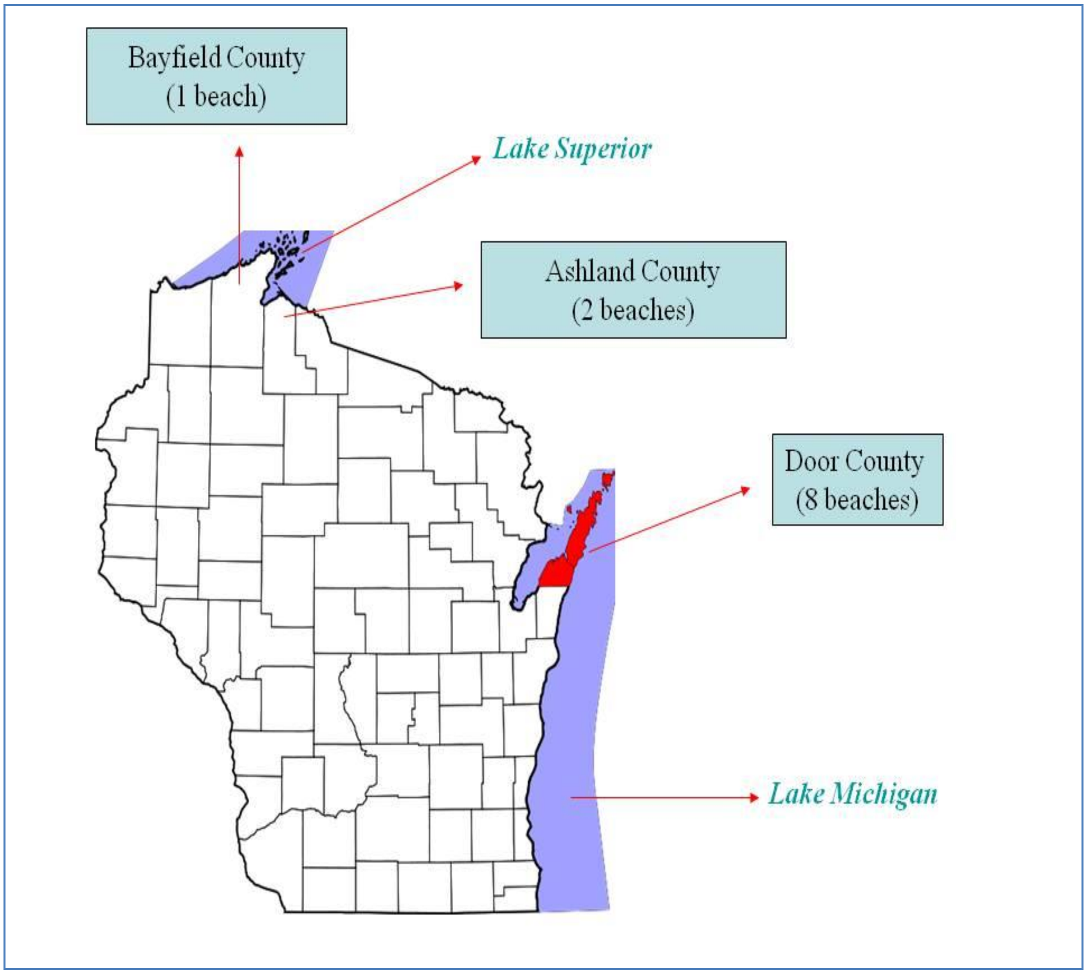

Figure 1: The map shows the study sites in the State of Wisconsin along the Lake Michigan and Lake Superior shorelines.

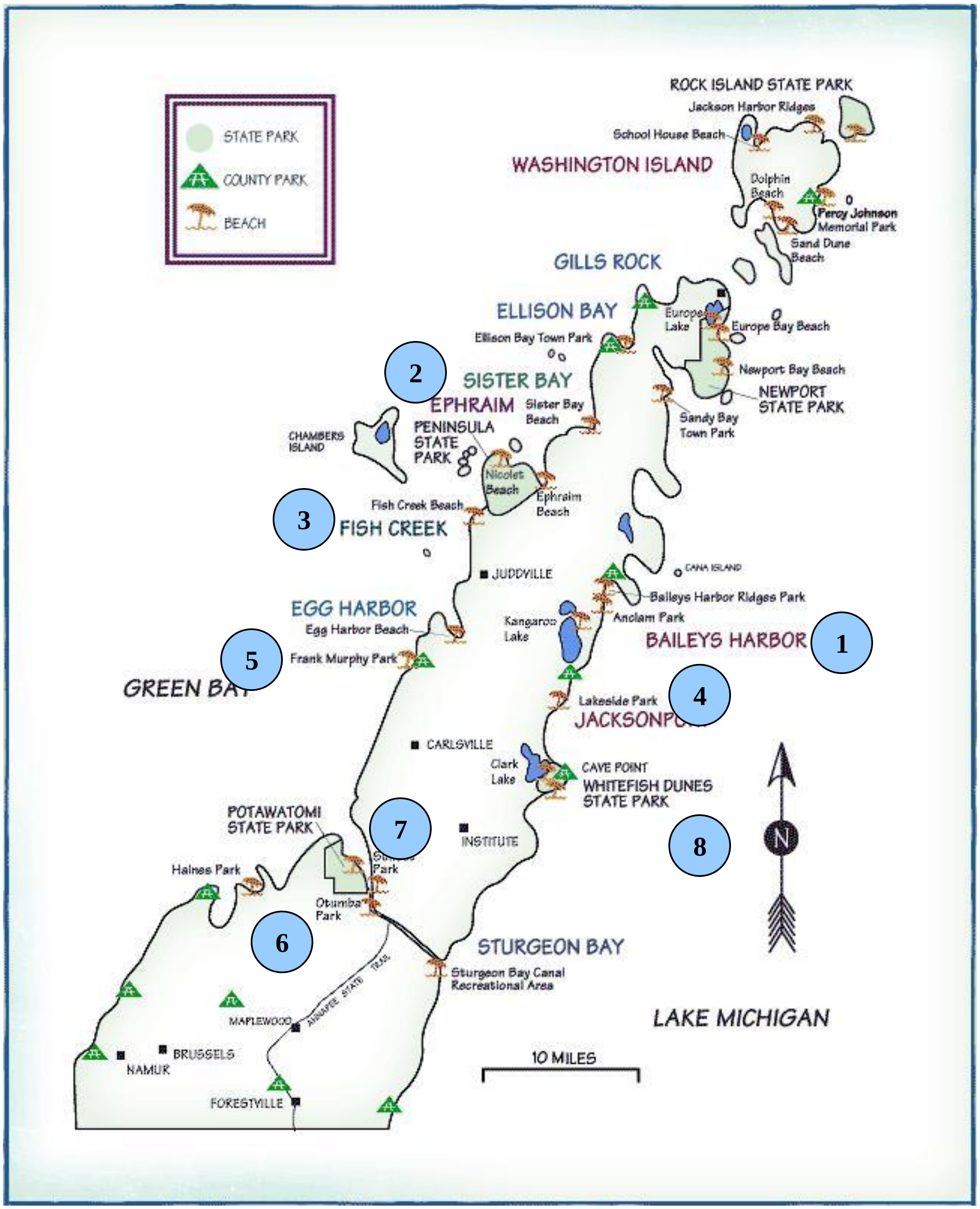

Figure 2: Door County map showing the eight study beaches in Door County labeled within circles.

Predictive modeling

Predictive models are analytical methods that provide more rapid estimates of beach water quality than the traditional microbiological methods. Predictive modeling is a real-time prediction method that can be used to predict *Escherichia coli* concentrations based on environmental and water quality variables. The variables found to be significant and related to *E. coli* concentrations are used as explanatory variables in multiple linear regression (MLR) models. Predictive modeling starts with data collection and analysis, model development, followed by model diagnostics and selection.

Data collection

Data collection included collection and analysis of water samples for *Escherichia coli* concentrations and measurement of environmental and water-quality variables. Data were collected four to five days a week by field personnel at beach study sites during the May through August recreational season of 2008. Data from past studies (2007) were also included in analysis in order to develop combined year (2007 – 2008) predictive models.

Collection and analysis of water samples

Water samples were collected from all the study beaches at a water depth of 61-76 centimeters (24-30 inches), as specified by the requirements of the Wisconsin Beach monitoring program. Over the entire duration of 2008 recreational season, 502 (n) beach water samples were collected. All the beach water samples were collected into sterile 100 mL polystyrene bottles and were labeled with date, time and location of the sample by the sampler. Immediately after collection, water samples were stored at 4°C until analysis.

Water samples were transported to the University of Wisconsin Oshkosh (UWO) remote laboratories (Ashland, Bayfield, and Sturgeon Bay) for analyzing *E. coli* concentrations, which were begun within four hours of collection. The University of Wisconsin Oshkosh (UWO) is a Wisconsin State Certified Laboratory with a Quality Assurance Plan utilized for *E. coli* analysis on file with the WI Department of Agriculture, Trade and Consumer Protection (DATCP). After returning to the UWO labs, the beach water samples were analyzed for their *E. coli* concentrations by the Defined Substrate (DS) technique, Colilert^{Tm} (IDEXX Corp., Portland, ME). *E. coli* concentrations were reported in Most Probable Number (MPN) of *E. coli* per 100mL of water using the MPN tables (IDEXX Corp., Portland, ME).

Collection of explanatory variables

Environmental and water-quality variables were considered as explanatory variables. The explanatory variables were measured during the 2008 recreational season either in the field by the field personnel at the time of sampling or through a weather forecasting website (http://www.wunderground.com/us/wi). Explanatory variables explained below were used in Multiple Linear Regression (MLR) analyses to develop mathematical models to predict *E. coli* concentrations (Table 1).

Air and water temperatures.

Air and water temperatures were measured in degrees Celsius (°C) using a thermometer. Air temperature was included as significant predictive variable for Indiana Dunes State Park Beach and West End Park Beach in Indiana (Olyphant, 2004). Water temperature was included as a significant model variable for Lake Erie beaches like Villa Angela (2004 – 2005) and Lakeshore (2004 – 2005) in Ohio (Francy et al., 2006). Several studies reported that *E. coli* may survive longer in cooler water temperatures than in warmer water temperatures (Bogosian et al., 1996; Brettar & Hofle, 1992; Sampson et al., 2005; Sampson et al., 2006).

Turbidity.

Beach water turbidity was measured using a turbidometer in Nephlometric Turbidity Units (NTU) with the help of turbidity reference standards. In this study, turbidity measurements were recorded as clear, slightly turbid, turbid, and opaque. These descriptors were then assigned a numerical value (clear = 0, slightly turbid = 1, turbid = 2, and opaque = 3) and used in model development. Turbidity can be contributed by clay, silt, finely divided organic matter, plankton, microscopic organisms and dyes (Anderson, & Wilde, 2005). Other researchers have found that turbidity (Log_{10} turbidity) could be one of the predictive variables for Lake Erie beaches (Huntington, Edgewater, Villa Angela, Lakeshore and Lakeview) in Ohio (Francy et al., 2006) and for Presque Isle Beach 2, Erie County in Pennsylvania (Zimmerman, 2006).

Wave height.

Wave height was measured in inches using a yard stick at the time of sampling. Wave heights were divided into four categories and denoted as low (0 - <1 feet), medium (≥1 - <2 feet) and high (≥2 feet). Wave height was considered as a predictive variable for Lake Erie beaches like Huntington, Edgewater and Villa Angela beaches in Ohio (Francy et al., 2006). At Huntington Beach, *E. coli* concentrations increased with increasing wave height during 2000-2005 (Francy et al., 2006). Because, as wave heights increased beach water may wash *E. coli* present in the sand and bring in to the water.

Wind current direction and speed.

Wind direction and speed were measured in the field using an anemometer in degrees and miles per hour (mph), respectively. Wind direction was reported as one of the predictive variable in models for 2004, 2005, and 2004 – 2005 combined for Presque Isle Beach 2, Erie County, Pennsylvania (Zimmerman, 2006). The long shore current speed (Lcs) was measured in centimeters per second (cps) and long shore current direction (Lsd) in degrees.

Onshore wind.

Onshore wind is another explanatory variable derived from the direction of the wind and classified into three terms. If the wind was blowing towards the beach (onshore) then it was coded as '1'. Along shore winds were coded as '0.5' and all other winds were coded as '0' (offshore).

In addition, wind directions also were coded as East = 1, Southeast = 2, South = 3, Southwest = 4, West = 5, Northwest = 6, North = 7 and Northeast = 8. Onshore wind vector was included in the regression model for 63rd Street Beach, Chicago (Olyphant & Whitman, 2004).

Gulls.

The numbers of gulls present on the beach were counted manually by the sampler at the time of sampling. Gulls were enumerated because waterfowl waste is a suspected source of fecal contamination at Wisconsin beaches (Kleinheinz, McDermott, & Chomeau, 2006). Other researchers have noted that higher concentrations of indicator bacteria such as *E. coli* and enterococci associated with gull feces (Fogarty, Haack, Wolcott, & Whitman, 2003). Other variables were derived from the gulls counted on the day of sampling. Gull counts from one day (one day late gull = “olagulls”) and two days (two day late gulls = “tlagulls”) prior to water sampling were tested for use as significant variables in models.

Algae.

The amount of algae (i.e. stranded *Cladophora*) present on the beach was classified into four categories as none (0), low (1), moderate (2), and high (3). Samplers were given photographs (Figure 3) to use for categorizing the deposited algal amounts. Recent studies have revealed that the *Cladophora* mats harbor *E. coli* in greater concentrations than found in the surrounding waters and provide suitable environment to persist for longer periods (Byappanahalli et al., 2003, Vanden Heuvel et al., 2009).

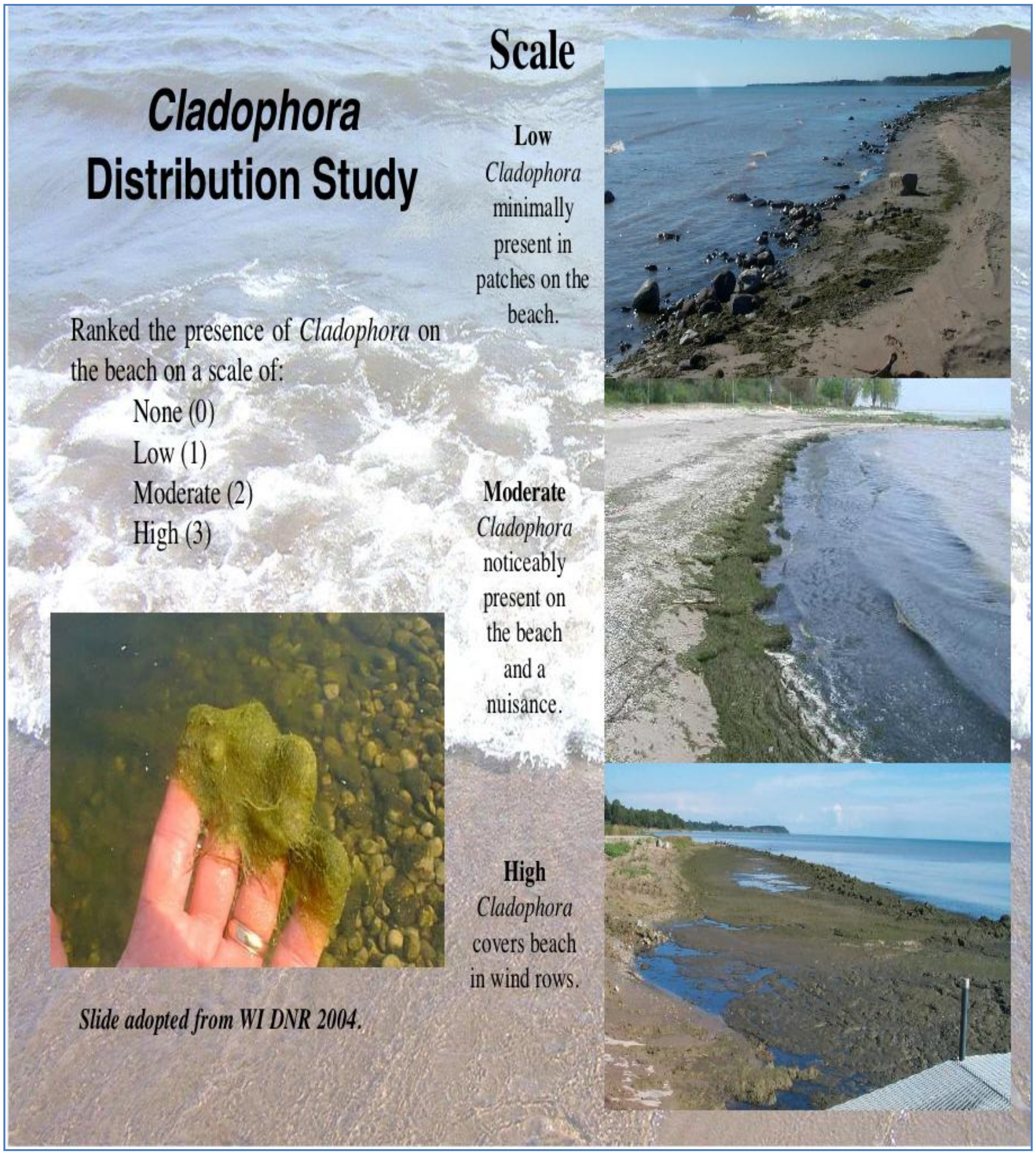

Figure 3: Photograph showing the *Cladophora* distribution study.

Barometric Pressure.

Barometric pressure was recorded in inches of mercury for the previous day and on the day of sampling, from a weather forecasting website (http://www.wunderground.com/us/wi).

Rainfall variables.

Rainfall is an important variable accounting for elevated *E. coli* concentrations. Previous studies revealed that rainfall events (> 6mm) were associated with elevated *E. coli* concentrations (Ackerman & Weisberg, 2003). During storms, sewage overflows caused increased *E. coli* concentrations generally occurring during rising stream-flow (Olyphant, Thomas, Whitman, & Harper, 2003). Because of the importance of rainfall variable as a predictor of *E. coli* concentrations, several different rainfall variables were developed and used in these predictive models. Twenty-four hour rainfall (tfrain) was the amount of rain, in inches, that fell in the day preceding the day of sampling. In the same fashion, forty-eight hour rainfall (ferain) and seventy-two hour rainfall (strain) were the amounts of rain that fell two days and three days before the day of sampling, respectively. Event rainfall (everain) was the amount of rain on the day of sampling. All the above rainfall variables were obtained from the weather underground website (http://www.wunderground.com/us/wi); using rain gauges installed at John F. Kennedy Memorial Airport (KASX) Ashland, WI and Door County Cherry land Airport (KSUE) Sturgeon Bay, WI.

Weighted hour rainfall variables (e.g. wferain, wstrain, and cstrain) were also estimated (Figure 4) and used in predictive model development. Weighted forty-eight hour rainfall (wferain) was the amount of rain that fell in the total forty-eight hour period preceding the day of sampling and calculated as,

Weighted forty-eight hour rainfall (wferain) = [2 * (tfrain + everain) + ferain].

In other words, rainfall occurring closer to the time of sampling was assumed to have greater impact of *E. coli* concentration in water than rainfall occurring further from the time of sampling.

Weighted seventy-two hour rainfall (wstrain) was the amount of rain that fell in the total seventy two-hour period preceding the day of sampling and calculated as,

Weighted seventy two hour rainfall (wstrain) = [3 * (tfrain + everain) + 2 * (ferain) + strain].

Combined seventy-two hour rainfall (cstrain) was the amount of rain that fell in the seventy-two hour period preceding sampling and calculated as,

Combined seventy-two hour rainfall (cstrain) = (tfrain + ferain + strain + everain).

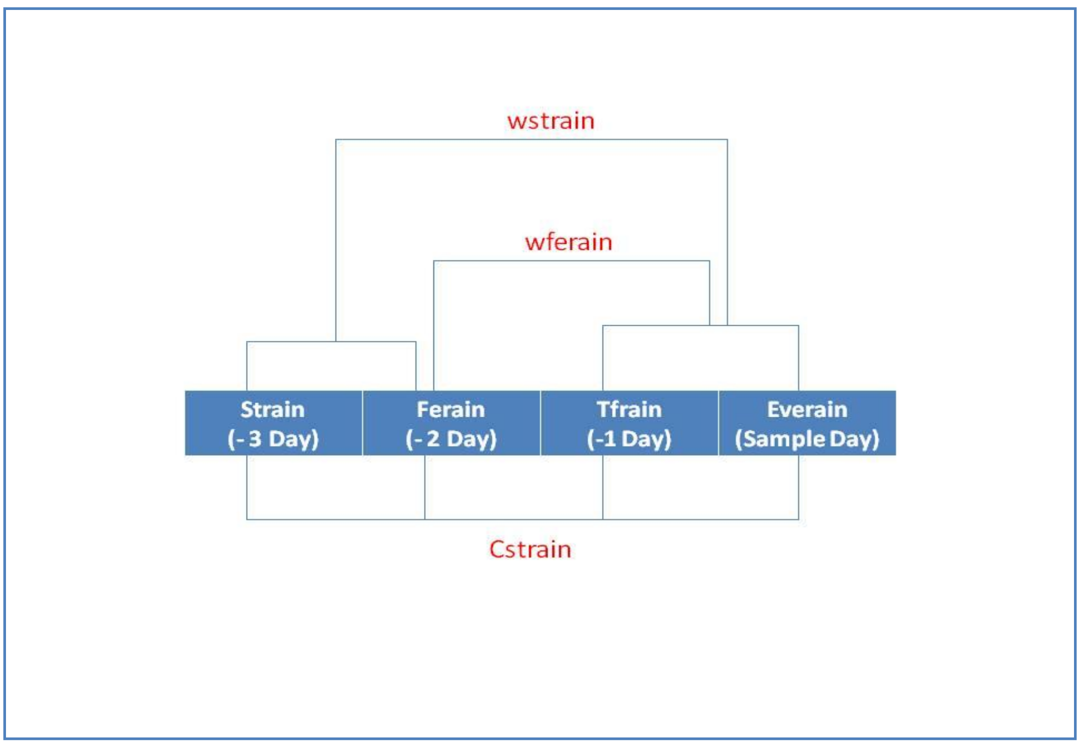

Figure 4: Diagram illustrating the relationship between the rainfall variables.

Table 1: List of environmental and water-quality variables considered for inclusion in a predictive model.

Variable	**Term used**	**Unit**
Water temperature	Wt	°C
Air temperature	At	°C
Turbidity	Trbdty	NTU
Wave height	Wh	Inches
Algae	Algae	Categorized
One day lag algae	Oalgae	Derived
Gulls	Gulls	Manual count
Gull variables	Olagulls, tlagulls	Derived
Wind direction	Wd	°
Wind speed	Ws	Mph
Long shore current speed	Lcs	Cps
Long shore current direction	Lsd	°
Onshore wind	Onwind	Categorized
Barometric pressure	Bp	Inches of mercury
Rainfall (24h)	Tferain	Inches
Rainfall (48h)	Ferain	Inches
Rainfall (72h)	Strain	Inches
Weighted hour rainfalls	wstrain, wferain and cstrain	Inches

Data management

All the collected data were entered into a Microsoft Excel 4.0 spreadsheet, starting with the sampling date and sampling time in the second column of the worksheet and then the third column was filled with the response variable i.e. *E. coli* concentration, determined by the defined substrate (DS) technique. Explanatory variables collected in the field and through the weather related websites were entered into the remaining columns.

Data analysis

Data analysis was started by examining the summary statistics for each beach for the year 2008 and 2007 – 2008 years of data combined which gave information about mean, median, minimum and maximum *E. coli* concentrations, and the number of days *E. coli* concentrations exceeded the water quality standard (235 CFU/ 100ml of water). Summary statistics were obtained by performing the statistical program – R and the statistics gave information regarding general water quality and explained year-to-year differences in explanatory variables.

Upon examining the yearly summary statistics, the explanatory variables related to *E. coli* were identified by constructing scatter plots. With a scatter plot, each explanatory variable measured was displayed on the X-axis and the response variable (*E. coli* concentration) was displayed on the Y-axis. These scatter plots show whether the relationships between *E. coli* and the environmental conditions were linear or nonlinear. Nonlinear relations were linearized by transforming the explanatory variables (log, square root, etc). In cases where relationships could not be linearized, variables responsible for nonlinear relations were eliminated.

Correlation analysis (Pearson's r correlation analysis) was used after identifying the explanatory variables related to *E. coli* concentration. R – Program was performed to estimate the quantitative measure of the relationship between the response variable (*E. coli*) and the explanatory variable. Pearson's 'r' correlation coefficient value measures the linear association between the variables and *E. coli* concentration statistically. As the correlation coefficient (r) approaches '1', there is a positive relation between the explanatory variable and response variable. If it approaches '-1' then there is a strong negative relation between the explanatory variable and response variable. In cases where 'r' approached '0' there is no relation or a weak relation between the explanatory variable and the response variable. Variables that are not related or weakly related to *E. coli* concentrations were not used in the predictive models. Correlation coefficients were considered statistically significant if the p – value was < 0.05. Upon identifying the best and most significant explanatory variables, mathematical models were generated using Virtual Beach (VB) software (Zhongfu & Frick, 2007).

Model development using Virtual Beach-Model Builder (VB-MB) software

Virtual Beach (VB-MB) is a model building software (Zhongfu & Frick, 2007), which internally uses multiple linear regressions to develop predictive models. Multiple linear regressions (MLR) are used to identify the relationships between two or more variables.

The general form of multiple linear regression equation is $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 +\ldots + \beta_kX_k + \varepsilon$, where Y is the response variable (i.e. Log *E. coli* concentration), β_0 is the intercept, β_1, β_2 and β_k are the slopes of first, second and k^{th} explanatory variables, and ε is the error. Virtual Beach Model Builder (VB-MB) software reads only Excel 4.0 files. Newer .xls formats were converted into Excel 4.0 and then used in modeling. The Excel 4.0 spread sheet data were imported into the VB software and the response variable (*E. coli* concentration) was log transformed due to the wide range of possible *E. coli* concentration values. Incomplete data of the sampling day and other variables were excluded from the model. Data then were evaluated using scatter plots. In MLR models, two or more explanatory variables that are highly correlated i.e., multicollinearity, can result in erroneous partial MLR coefficients. Virtual Beach-Model Builder checks multicollinearity by reporting the Variance Inflation Factors (VIF). A value of VIF > 10 indicates the presence of severe multicollinearity. In such cases, variables responsible for VIF values >10 were either excluded or created as an "Interaction term".

VB software runs a "backwards regression" program to develop various mathematical models. A backwards regression is a variable selection procedure used to eliminate insignificant variables. First, all the explanatory variables are considered and then variables are eliminated one at a time, starting with least statistical significance value. If the p-value is > 0.05, the variable is eliminated from the model. The backward regression process continues until a three or four variable model is obtained, because in general the simpler the model, the higher the predictive ability. VB software calculates the adjusted R squared (R^2) of the model for each beach, which is used to explain the percent of variation in the log transformed *E. coli* concentration.

Model diagnostics and selection

To estimate the accuracy of the developed mathematical model, scatter plots were constructed between observed *E. coli* concentrations and estimated (predicted) *E. coli* concentrations. The scatter plot graphically displays how well the model estimates $\log_e$ (*E. coli*) compared to the observed $\log_e$ (*E. coli)* concentrations for the various observations included in the model. The plot was divided into four quadrants by a vertical line through 5.46 ($\log_e$235 CFU/100mL) on the X-axis and a horizontal line through 5.46 ($\log_e$235 CFU/100mL) on the Y-axis. The four quadrants represent the number of correct predictions above the *E. coli* water quality standard (>235 CFU/100mL of water), the number of predictions below the *E. coli* water quality standard (<235 CFU/100mL of water), false positives (incorrect predictions that water quality was below the standard for a given day, when there was in fact an exceedance) and false negatives (incorrect predictions of exceedances) (Figure 5).

Ideally, a best-fit model should not predict any false negative responses, because swimming would be allowed when *E. coli* concentrations were above the water quality standard (235 CFU/100mL). These false negatives and false positives were minimized by altering and refitting the model.

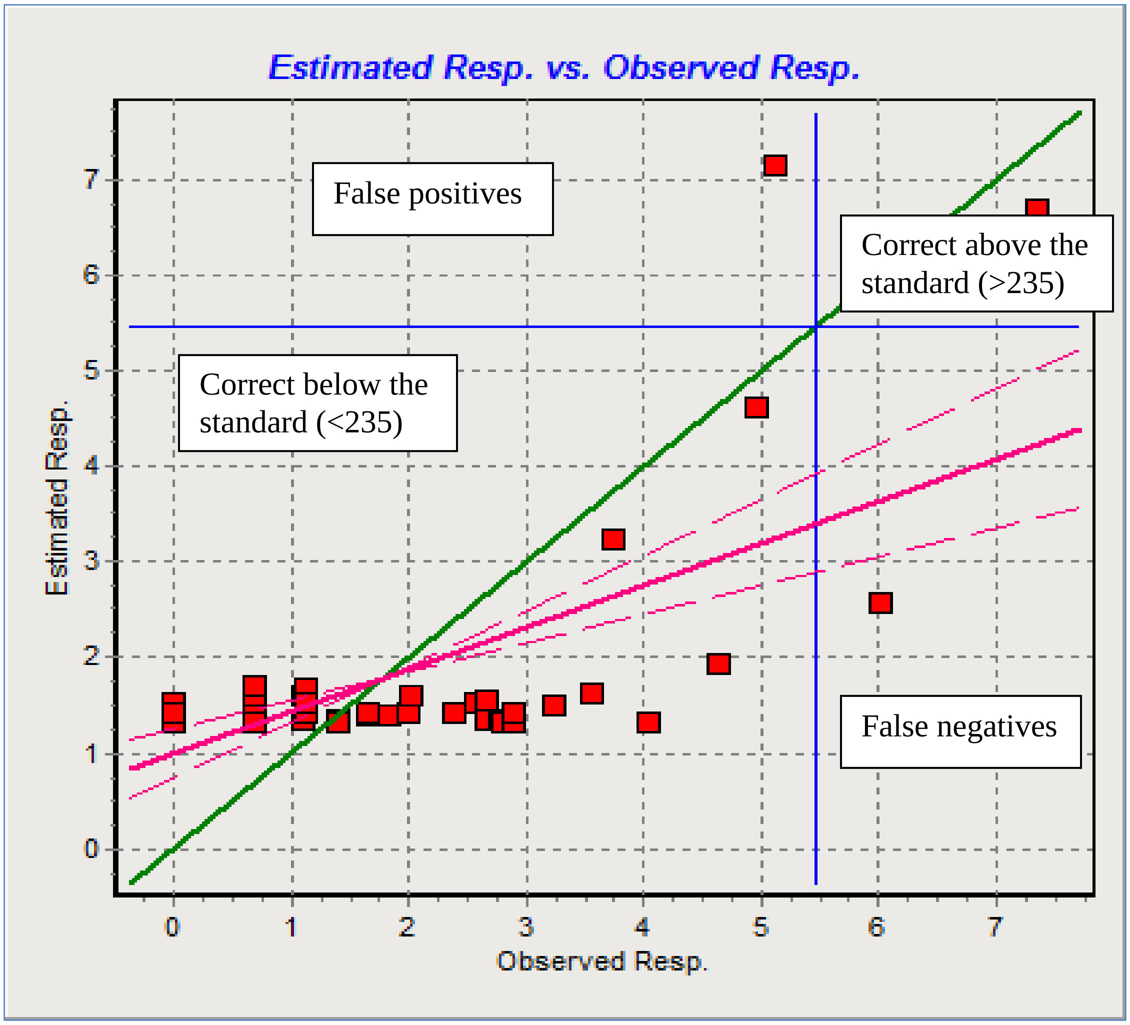

Figure 5: A model scatter plot showing the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 5, the correct above standard, correct below standard, false positive and false negative predictions were shown in four quadrants. Blue lines indicate bathing water quality standard 5.46 ($\log_e$ 235 *E. coli*/ 100mL water). The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it. The green diagonal represents a perfect model for comparison.

In addition to visual inspection of scatter plot, VB software automatically detects the influential outliers (those likely to skew the model as fitted). In such cases, those observations were excluded and the model was refitted. After identifying influential outliers, model selection was accomplished. The best-fit model was selected based on adjusted coefficient of determination (R^2) and Mallow's Cp. Mallow's Cp is a measure of the error in the best-fit model, relative to error incorporating all variables. The model with the smallest possible number of explanatory variables and the highestR^2 and lowest Cp was considered the best-fit model.

RESULTS

Ashland County

Kreher Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum log_e *E. coli* concentrations for Kreher Park beach 2008 season were (2.28 ± 1.56), 2.19, and 6.01, respectively. Correlations between *E. coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Kreher Park Beach were shown in Table 2. The explanatory variables that showed positive correlation with *E. coli* concentration for Kreher Park Beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were wave height ($r = 0.41$, $p < 0.05$), algae ($r = 0.32$, $p < 0.05$) and event rainfall ($r = 0.46$, $p < 0.05$). Only these were included in the final predictive model, and are highlighted in Table 2. The scatter plot shown in Figure 6 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant explanatory variables (wave height, algae and event rainfall) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (1 day), as seen in Figure 6 and Table 3. The adjusted R-squared (R^2) of the mathematical model was 43.7% (Table 5) with lowest Cp was – 1.279 (Table 5 and Figure 7) and the model was considered statistically significant ($p<0.05$), as seen in Table 5 using significant explanatory variables (Table 4).

The predictive equation for Kreher Park Beach – 2008 model is as follows: *E. coli* [Log_e (MPN)] **= [**1.6721 + (0.4033*wh) + (1.1375*algae) + (10.985*everain)].

Table 2: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Kreher Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.14
Wind speed	-0.08
Onshore wind	-0.09
Water temperature	-0.10
Air temperature	0.05
Turbidity	0.56
Wave height	**0.41****
Algae	**0.32****
Gulls	-0.13
One day late gulls	0.19
Two days late gulls	0.05
Long shore current speed	-0.08
24 hour rainfall	-0.16
48 hour rainfall	-0.05
72 hour rainfall	-0.19
Event rainfall	**0.46****
Weighted 72hr rainfall	0.03
Weighted 48hr rainfall	0.06
Combined 72hr rainfall	-0.12
Barometric pressure	-0.02

**Variable considered statistically significant i.e. $p < 0.05$.

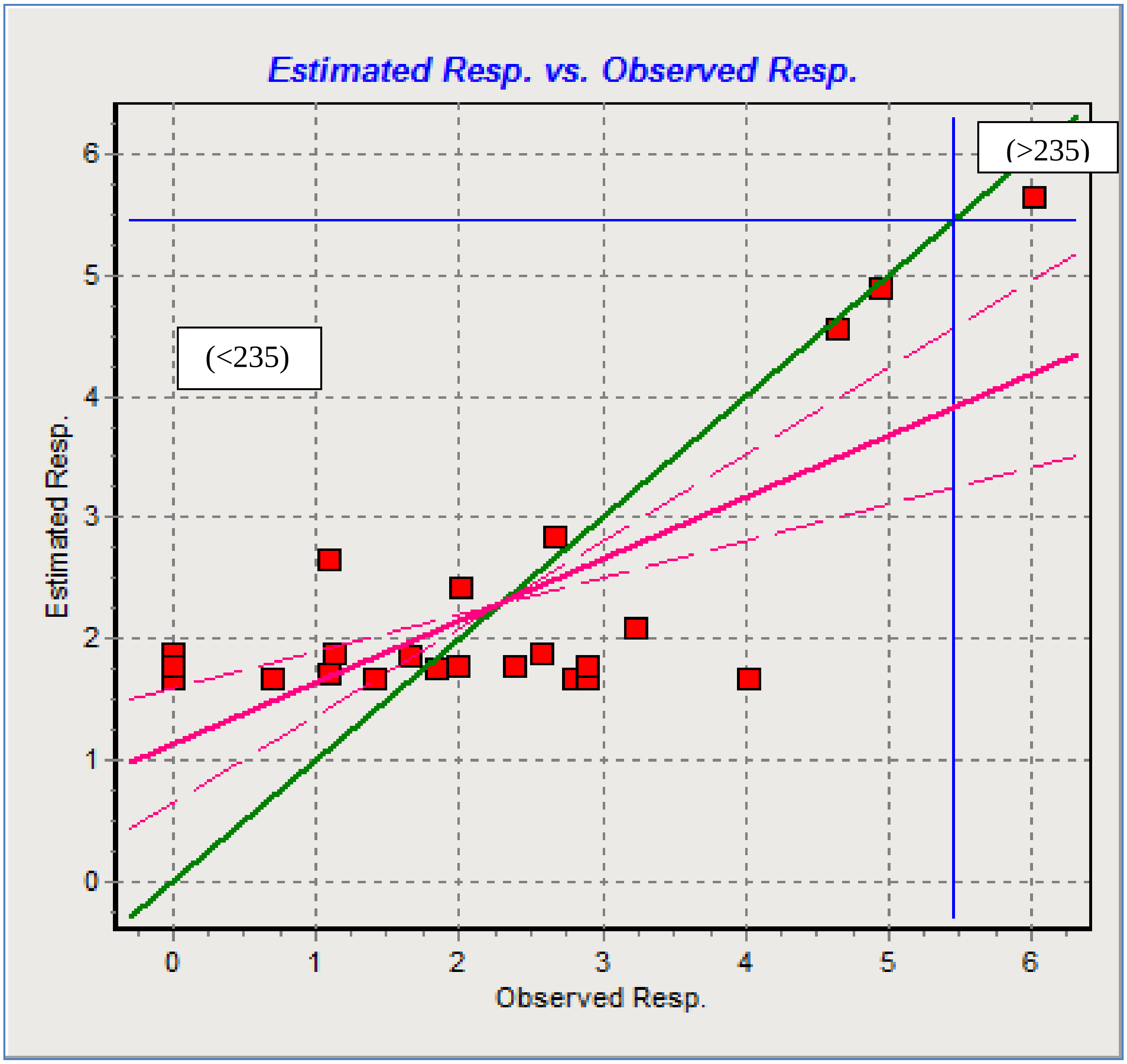

Figure 6: Kreher Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 6, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235 CFU/100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 43.7%. The green diagonal represents a perfect model for comparison.

Table 3: Kreher Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	**Number**	**Percentage**
Total predictions	24	100
Correct above standard	1	4.17%
Correct below standard	23	95.8%
False positives	0	0 %
False negatives	0	0 %

Table 4: Kreher Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	**Parameter estimate**	**Standard error**	**t-value**	**p-value**
Intercept	1.6721	0.2751	6.0776	6.1293E-6
Wave height	0.4033	0.1521	2.651	0.01538
Algae	1.1375	0.5487	2.0732	0.05142
Event rainfall	10.985	3.3589	3.2704	0.003841

Table 5: Kreher Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	**Value**
F-statistic	6.949
Degrees of freedom (DF)	20
Residual standard error (RSE)	1.172
Adjusted R^2	43.7%
Mallow’s Cp	-1.279
p-value	0.002181

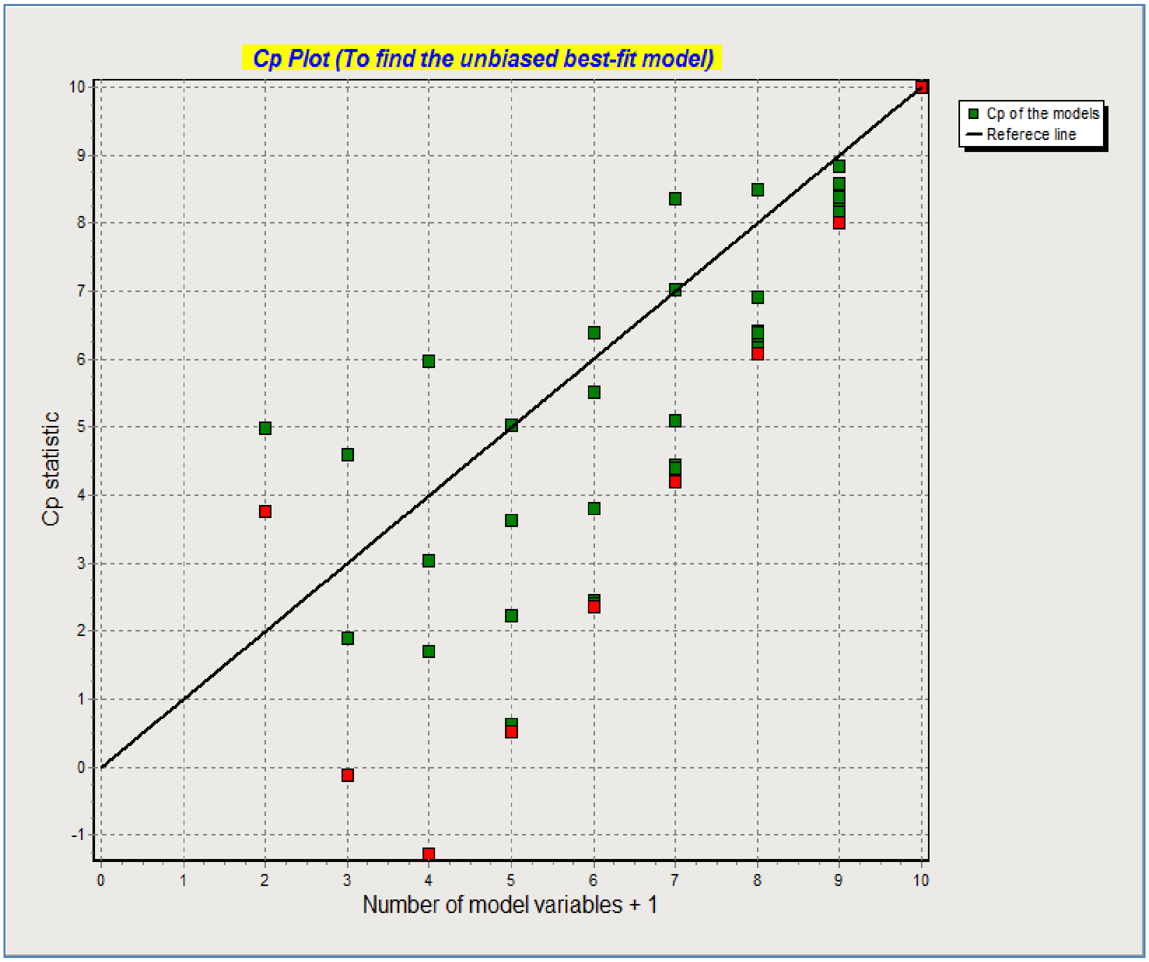

Figure 7: Kreher Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (-1.279) was considered best fit.

Kreher Park Beach 2007 – 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Kreher Park beach 2007 – 2008 combined season were (1.75 ± 1.72), 1.41, and 7.34 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Kreher Park Beach for 2007-2008 combined season were included in the modeling process (Table 6). The variables that showed significant positive correlation were wave height ($r = 0.27$, $p<0.05$) and event rainfall ($r = 0.58$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 6. The scatter plot shown in Figure 8 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (wave height and event rainfall) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (1 day), as seen in Figure 8 and Table 7. The adjusted R-squared (R^2) of the mathematical model was 47.6% (Table 9) with lowest Cp was – 17.16 (Table 9 and Figure 9) and the model was considered statistically significant ($p<0.05$), as seen in Table 9 using significant explanatory variables (Table 9). The predictive equation for Kreher Park Beach 2007 – 2008 combined model is as follows:

E. coli [Log_e (MPN)] = [1.3094 + (0.4138*wh) + (3.2649*everain)].

Models were compared for Kreher Park beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 10. Mathematical model for 2008 season was predicted $\log_e$ *E. coli* concentrations 100% accurately (Table 10).

Table 6: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Kreher Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.08
Wind speed	-0.12
Onshore wind	-0.14
Water temperature	-0.02
Air temperature	-0.20
Turbidity	0.07
Wave height	**0.27****
Algae	0.004
One day late algae	-0.12
Gulls	0.03
Long shore current speed	-0.16
24 hour rainfall	-0.02
48 hour rainfall	-0.06
72 hour rainfall	0.04
Event rainfall	**0.58****
Weighted 72hr rainfall	0.39
Weighted 48hr rainfall	0.41
Combined 72hr rainfall	0.31
Barometric pressure	-0.11

**Variable considered statistically significant i.e. $p < 0.05$.

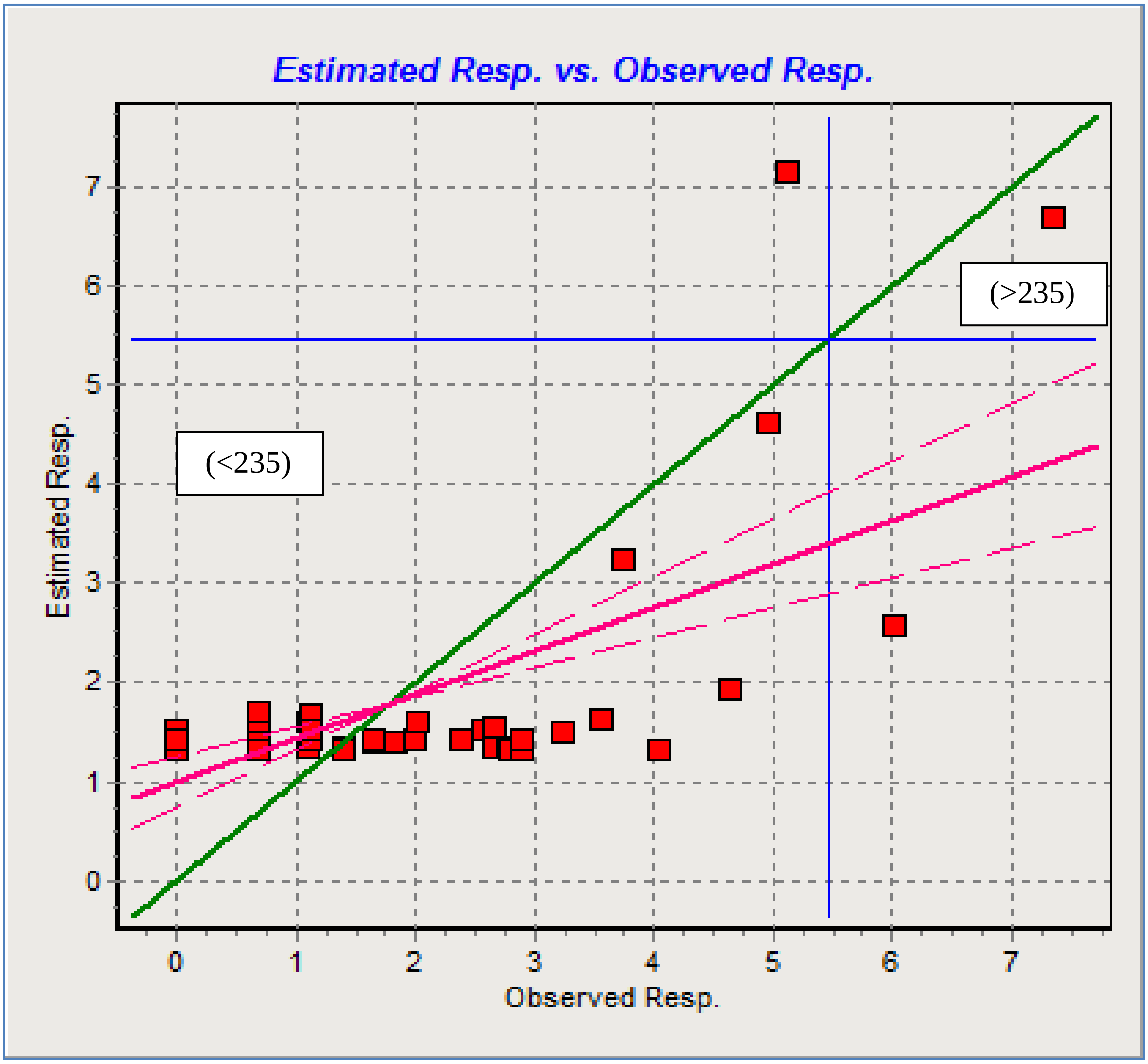

Figure 8: Kreher Park Beach 2007 – 2008 combined model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 8, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 47.6%. The green diagonal represents a perfect model for comparison.

Table 7: Kreher Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	53	100
Correct above standard	1	1.89%
Correct below standard	50	94.3%
False positives	1	1.89%
False negatives	1	1.89%

Table 8: Kreher Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	1.3094	0.1865	5.053	6.4960E-6
Wave height	0.4138	0.1611	2.5681	0.01338
Event rainfall	3.2649	0.4968	6.5718	3.0798E-8

Table 9: Kreher Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	19.35
Degrees of freedom (DF)	50
Residual standard error (RSE)	1.324
Adjusted R^2	47.6%
Mallow's Cp	-17.16
p-value	5.957e-07

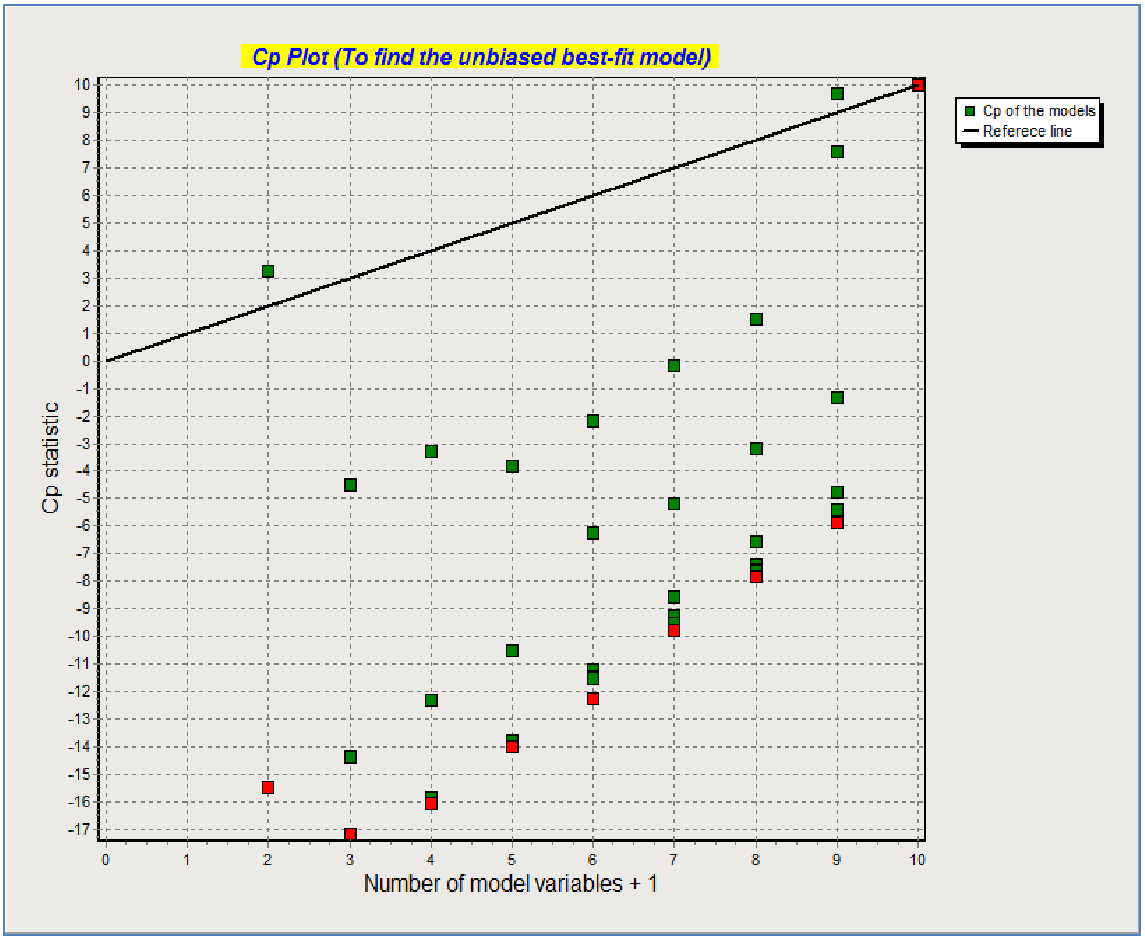

Figure 9: Kreher Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (-17.16) was considered best fit.

Table 10: Comparative models for Kreher Park Beach.

Kreher Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Weighted 72hr rainfall and √event rainfall	Wave height, algae, and event rainfall	Wave height, and event rainfall
Number of samples	29	24	53
Correct predictions	28 (96.5%)	24 (100%)	52 (98.1%)
Adjusted R^2	70%	43.7%	47.6%
Mallow's Cp	3.0	- 1.279	- 17.16
Residual standard error	0.953	1.172	1.324
p-value	<0.001	0.002181	5.957e-07
Predictive equation	*E. coli* [Log (MPN)] = 1.555 + (0.295 * wstrain) + (3.157 * √event rainfall)	*E. coli* [Log (MPN)] = 1.6721 + (0.4033 * wh) + (1.1375 * algae) + (10.985 * everain)	*E. coli* [Log (MPN)] = 1.3094 + (0.4138 * wh) + (3.2649 * everain)

Maslowski Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Maslowski Park beach 2008 season were (3.79 ± 1.65), 3.95, and 6.78 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season were shown in Table 11. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Maslowski Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were wave height ($r = 0.51$, $p<0.05$) and weighted 48-hour rainfall (wferain, $r = 0.28$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 11. The scatter plot shown in Figure 10 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant explanatory variables (wave height and weighted 48 hour rainfall) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (2 days), as seen in Figure 10 and Table 12. The adjusted R-squared (R^2) of the mathematical model was 40.8% (Table 14) with lowest Cp was – 17.16 (Table 14 and Figure 11) and the model was considered statistically significant ($p<0.05$), as seen in Table 14 using significant explanatory variables (Table 13). The predictive equation for Maslowski Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] = 2.1059 + (2.7186*wh) + (5.0543*wferain).

Table 11: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Maslowski Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	0.46
Wind speed	-0.17
Onshore wind	0.26
Water temperature	-0.07
Air temperature	-0.25
Turbidity	0.31
Wave height	**0.51****
Algae	0.32
One day late algae	0.07
Gulls	0.05
One day late gulls	0.11
Two days late gulls	-0.27
Long shore current speed	0.36
24 hour rainfall	0.13
48 hour rainfall	0.02
72 hour rainfall	0.007
Event rainfall	0.21
Weighted 72hr rainfall	0.19
Weighted 48hr rainfall	**0.28****
Combined 72hr rainfall	0.08
Barometric pressure	0.07

**Variable considered statistically significant i.e. $p < 0.05$.

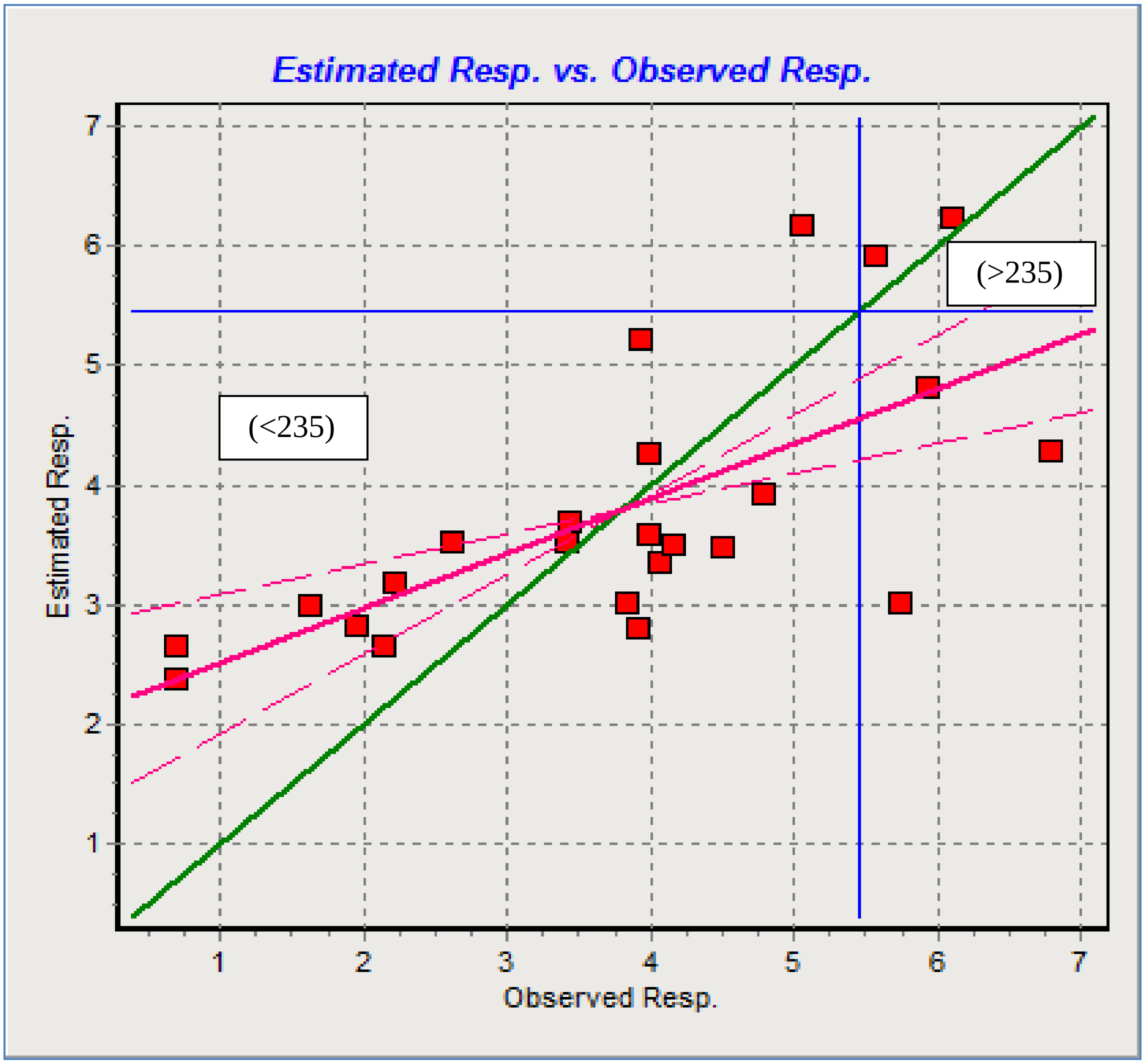

Figure 10: Maslowski Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 10, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 40.8%. The green diagonal represents a perfect model for comparison.

Table 12: Maslowski Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	**Number**	**Percentage**
Total predictions	24	100
Correct above standard	2	8.33%
Correct below standard	18	75%
False positives	1	4.17%
False negatives	3	12.5%

Table 13: Maslowski Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	**Parameter estimate**	**Standard error**	**t-value**	**p-value**
Intercept	2.1059	0.4801	4.3868	0.0002587
Wave height	2.7186	0.7074	3.8429	0.000949
Weighted 48hr rainfall	5.0543	1.8373	2.7509	0.01202

Table 14: Maslowski Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	**Value**
F-statistic	8.921
Degrees of freedom (DF)	21
Residual standard error (RSE)	1.277
Adjusted R^2	40.8%
Mallow’s Cp	- 11.87
p-value	0.001569

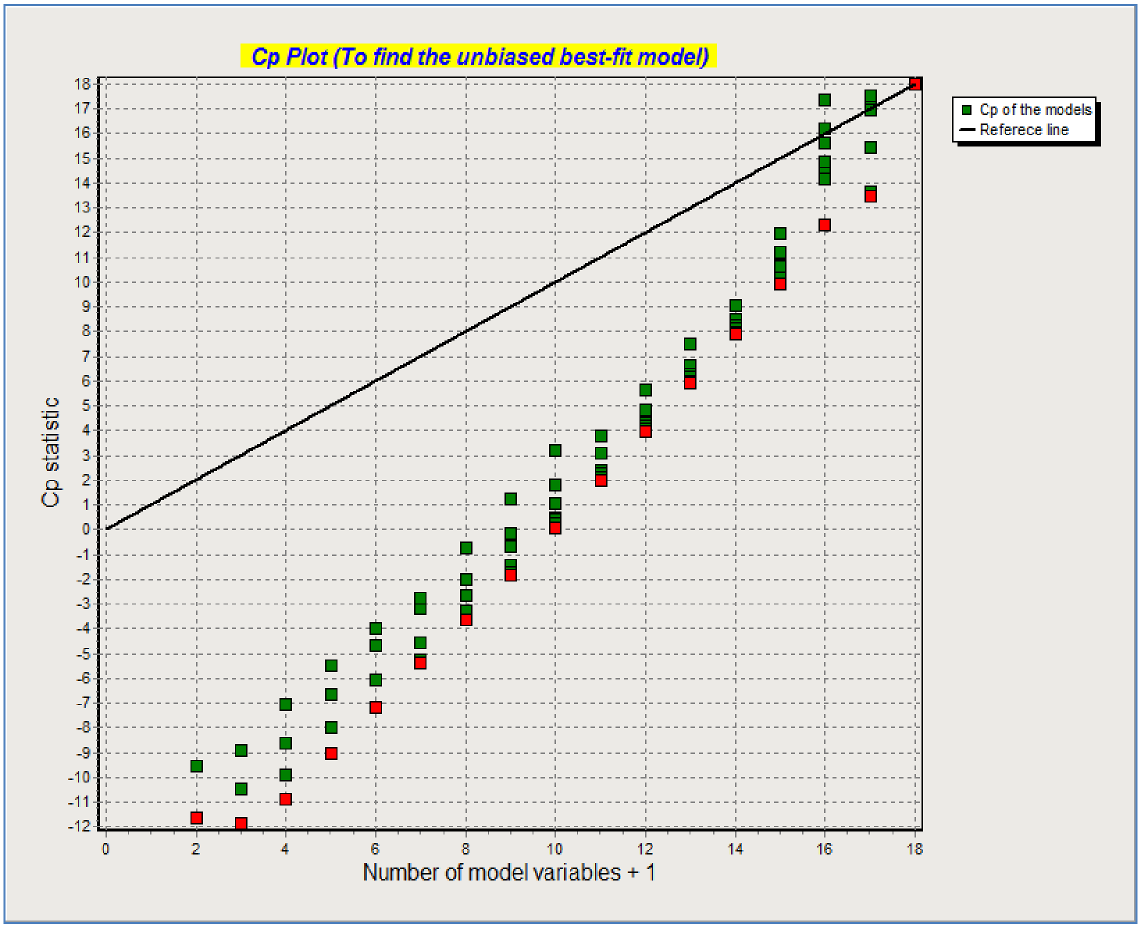

Figure 11: Maslowski Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (-11.87) was considered best fit.

Maslowski Park Beach 2007 – 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Maslowski Park beach 2007 – 2008 combined season were (3.53 ± 1.72), 3.49, and 7.21 respectively. The explanatory variables that showed positive correlation with *E. coli* concentration for Maslowski Park beach for 2007-2008 combined season model, were included in the modeling process shown in Table 15. The variables that showed significant positive correlations with *E. coli* concentrations were wave height (r = 0.22, p < 0.05), algae (r = 0.27, p < 0.05), one day late gulls (r = 0.41, p < 0.05), and weighted 78hour rainfall (r = 0.22, p < 0.05). Only these were included in the final predictive model, and are highlighted in Table 15. The scatter plot shown in Figure 12 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (wave height, algae, one day late gulls, and weighted 72 hour rainfall) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (1 day), as seen in Figure 12 and Table 16. The adjusted R-squared (R^2) of the mathematical model was 37.5% (Table 18) with lowest Cp was – 11.22 (Table 18 and Figure 13) and the model was considered statistically significant (p<0.05), as seen in Table 18 using significant explanatory variables (Table 17). The predictive equation is as follows: *E. coli* [Log (MPN)] = [1.3267 + (1.5702*wh) + (0.497*algae) + (0.01689*olagulls) + (0.4458*wstrain)].

Models were compared for Maslowski Park beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 19.

Table 15: Pearson's correlation coefficient (r) values between *E. coli* concentrations and explanatory variables for the Maslowski Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	0.23
Wind speed	-0.10
Onshore wind	0.01
Water temperature	-0.17
Air temperature	-0.23
Turbidity	-0.01
Wave height	**0.22****
Algae	**0.27****
One day late algae	0.15
Gulls	0.24
One day late gulls	**0.41****
Two days late gulls	0.10
24 hour rainfall	0.06
48 hour rainfall	0.05
72 hour rainfall	-0.006
Event rainfall	0.36
Weighted 72hr rainfall	**0.22****
Weighted 48hr rainfall	0.24
Combined 72hr rainfall	0.17
Barometric pressure	-0.005

**Variable considered statistically significant i.e. $p < 0.05$

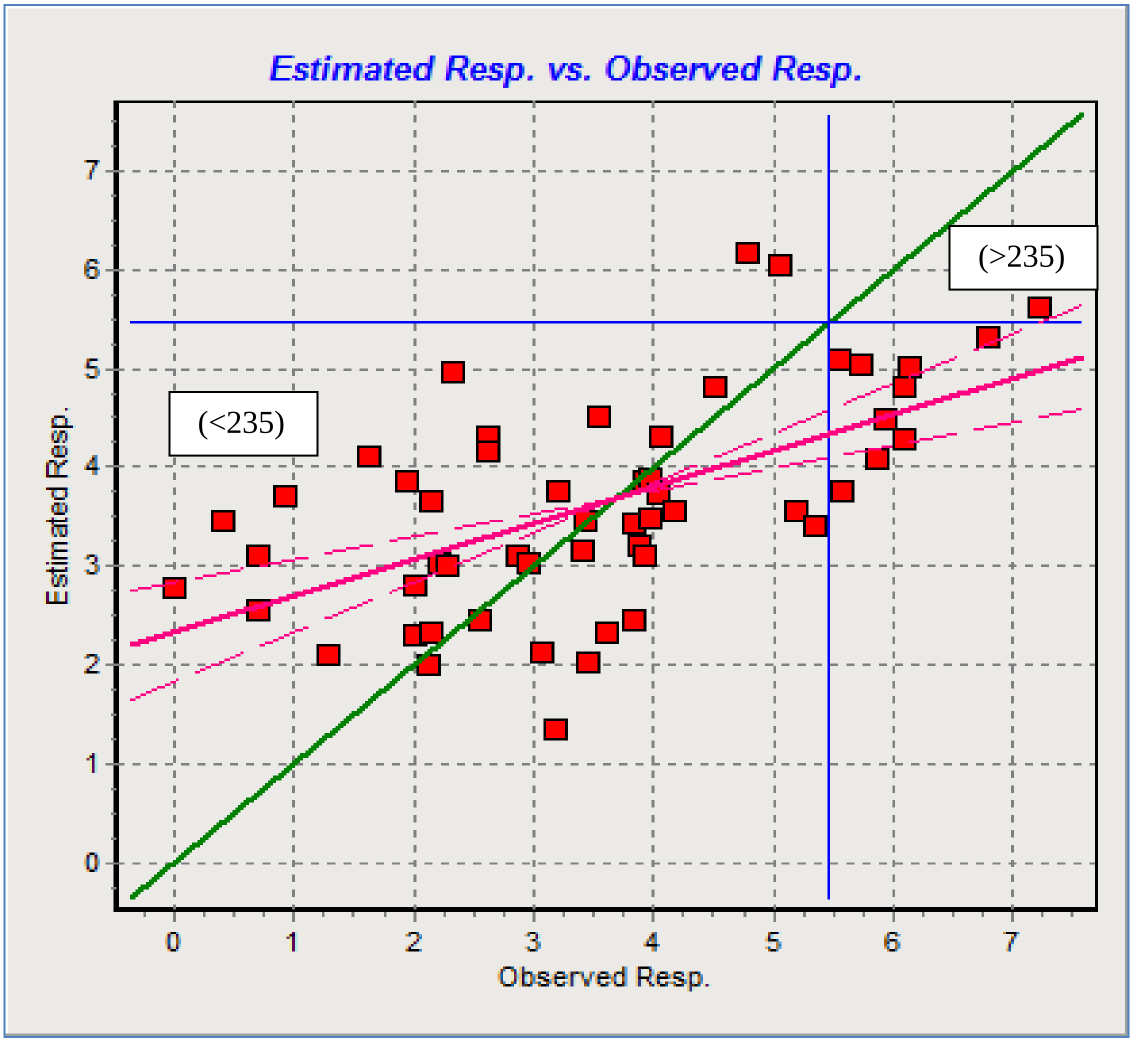

Figure 12: Maslowski Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 12, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the “best fit” line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 37.5%. The green diagonal represents a perfect model for comparison.

Table 16: Maslowski Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	54	100
Correct above standard	1	1.85%
Correct below standard	42	77.8%
False positives	2	3.7%
False negatives	9	16.7%

Table 17: Maslowski Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	1.3267	0.2877	2.5374	0.01453
Wave height	1.5702	0.51	3.0789	0.003449
Algae	0.497	0.1958	2.5387	0.01449
One day late gulls	0.01689	0.004917	3.4347	0.001239
Weighted 72hr rainfall	0.4458	0.1668	2.6723	0.01031

Table 18: Maslowski Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	6.641
Degrees of freedom (DF)	49
Residual standard error (RSE)	1.2429
Adjusted R^2	37.5%
Mallow’s Cp	- 11.22
p-value	0.0002365

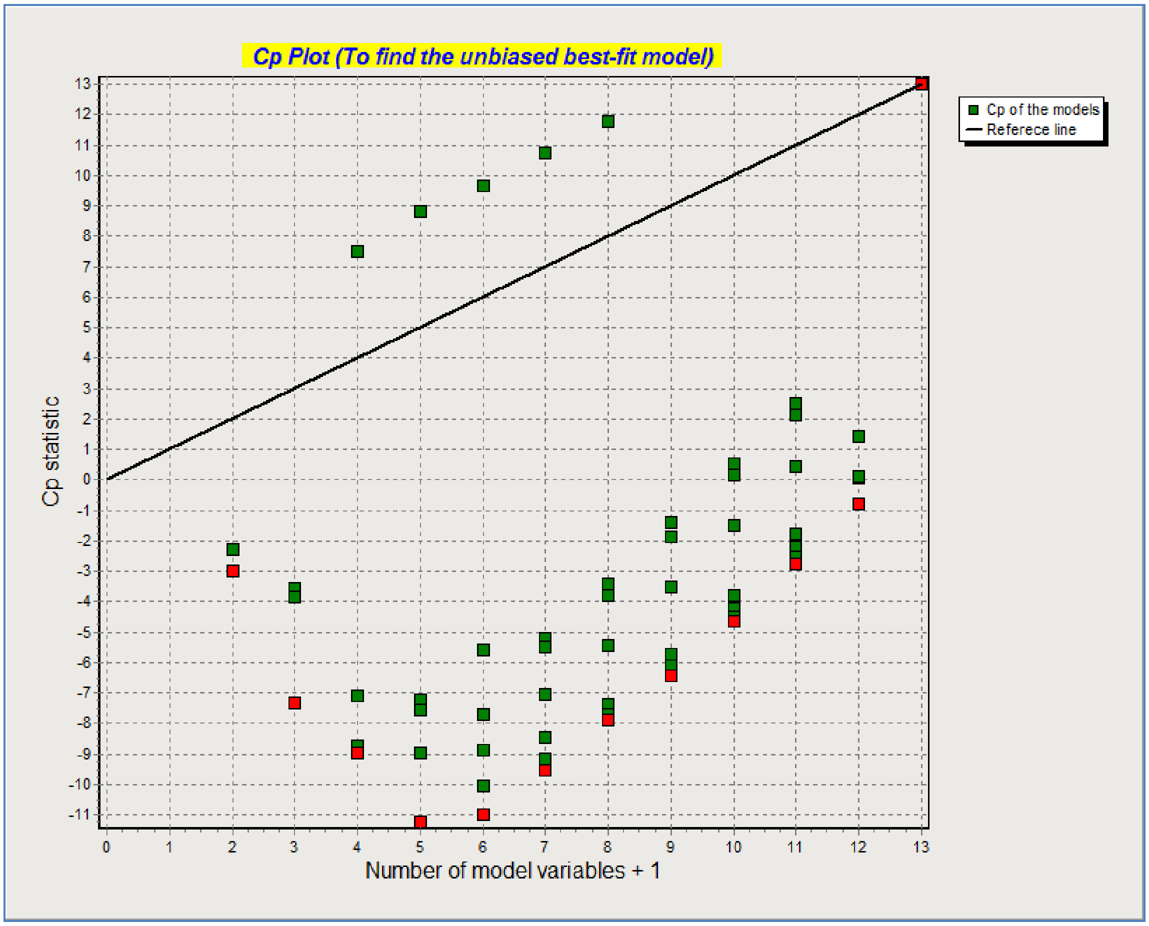

Figure 13: Maslowski Park Beach 2007 – 2008 Cp-Plot combined mathematical model. The model with lowest Cp (-11.22) was considered best fit.

Table 19: Comparative models for Maslowski Park Beach.

Maslowski Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Two days lag algae, one day late gulls, and weighted 48hr rainfall	Wave height and weighted 48hr rainfall	Wave height, algae, one day late gulls, and weighted 72hr rainfall
Number of samples	30	24	54
Correct predictions	28 (93.3%)	20 (83.3%)	43 (79.6%)
Adjusted R^2	54%	40.8%	37.5%
Mallow's Cp	4.0	- 11.87	- 11.22
Residual standard error	1.192	1.277	1.243
p-value	<0.001	0.001569	0.0002365
Predictive equation	*E. coli* [Log (MPN)] = 1.321 + (0.895 * tlalgae) + (0.025 * olagulls) + (0.459 * wferain)	*E. coli* [Log (MPN)] = 2.1059 + (2.7186 * wh) + (5.0543 * wferain)	*E. coli* [Log (MPN)] = 1.3267+ (1.5702 * wh) + (0.497 * algae) + (0.01689 * olagulls) + (0.4458 * wstrain)

Bayfield County

Thompson's West End Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Thompson West End Park beach 2008 season were (2.89 ± 1.86), 2.64, and 7.79 respectively. Correlations between *E. coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season were shown in Table 20. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Thompson's West End Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were water temperature ($r = 0.48$, $p<0.05$), one day late algae ($r = 0.17$, $p<0.05$) and gulls ($r = 0.71$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 20. The scatter plot shown in Figure 14 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variables (water temperature, one day late algae, and gulls) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (2 days), as seen in Figure 14 and Table 21. The adjusted R-squared (R^2) of the mathematical model was 71.5% (Table 23) with lowest Cp was 4.0 (Table 23 and Figure 15) and the model was considered statistically significant ($p<0.05$), as seen in Table 23 using significant explanatory variables (Table 22). The predictive equation for Thompson's West End Park Beach – 2008 model is as follows: *E. coli* [Log (MPN)] = [-2.1272 + (0.24*wt) + (0.9433*oalgae) + (0.08636*gulls)].

Table 20: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for Thompson West End Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.16
Wind speed	-0.02
Onshore wind	0.16
Water temperature	**0.48****
Air temperature	0.35
Turbidity	0.41
Wave height	0.43
Algae	0.20
One day late algae	**0.17****
Gulls	**0.71****
One day late gulls	0.12
Two days late gulls	0.11
Long shore current speed	0.21
24 hour rainfall	0.16
48 hour rainfall	-0.07
72 hour rainfall	-0.11
Event rainfall	-0.006
Weighted 72hr rainfall	-0.07
Weighted 48hr rainfall	0.06
Combined 72hr rainfall	-0.111
Barometric pressure	0.201

**Variable considered statistically significant i.e. $p < 0.05$.

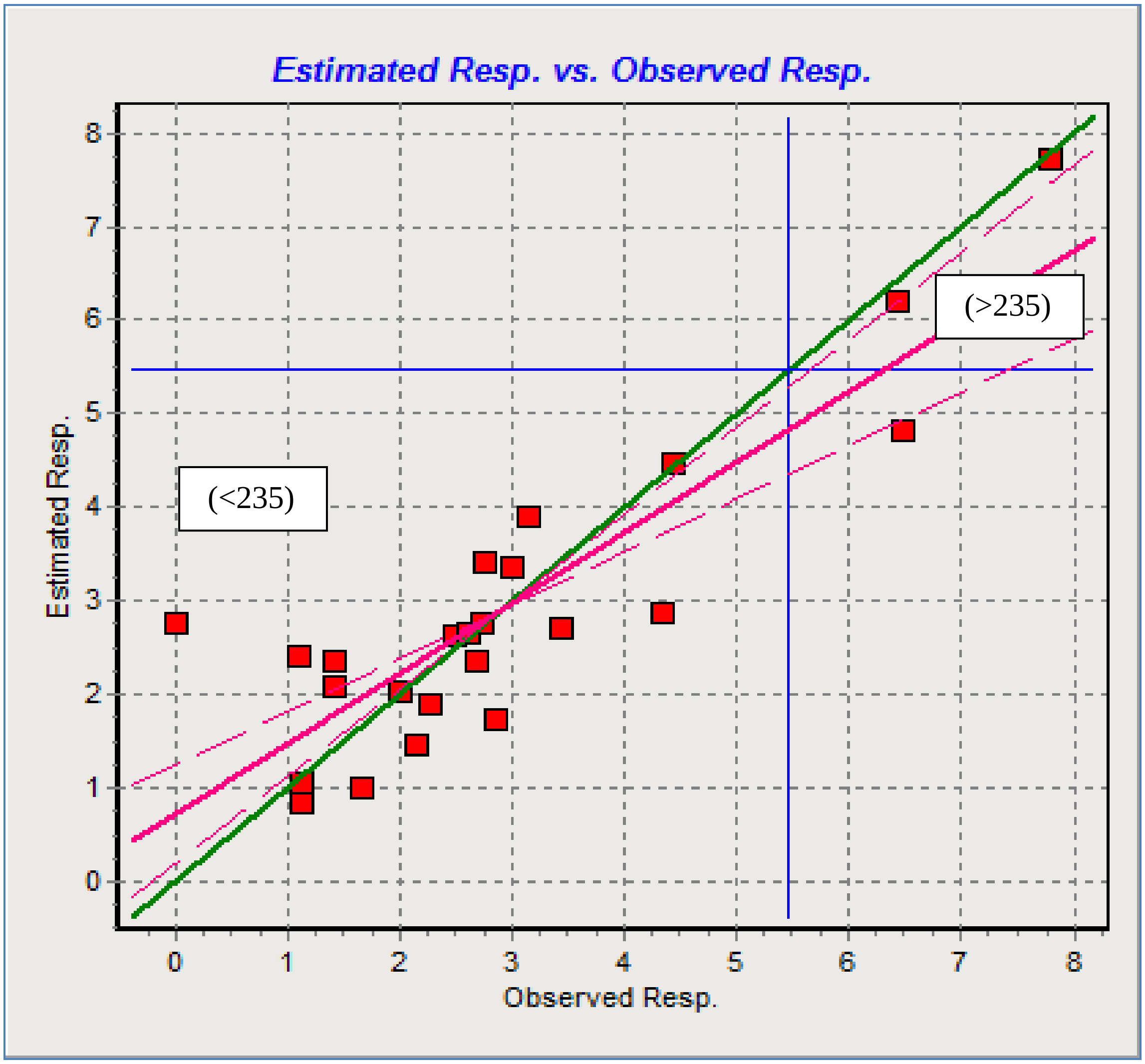

Figure 14: Thompson's West End Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 14, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 71.5%. The green diagonal represents a perfect model for comparison.

Table 21: Thompson's West End Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	24	100
Correct above standard	2	8.33%
Correct below standard	21	87.5%
False positives	0	0%
False negatives	1	4.17%

Table 22: Thompson's West End Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-2.1272	1.1585	-1.8361	0.08144
Water temperature	0.24	0.06371	3.7666	0.001218
One day late algae	0.9433	0.4109	2.2955	0.03275
Gulls	0.08636	0.0137	6.3046	3.7438E-6

Table 23: Thompson's West End Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	20.25
Degrees of freedom (DF)	20
Residual standard error (RSE)	0.9946
Adjusted R^2	71.5%
Mallow's Cp	4.0
p-value	2.832e-06

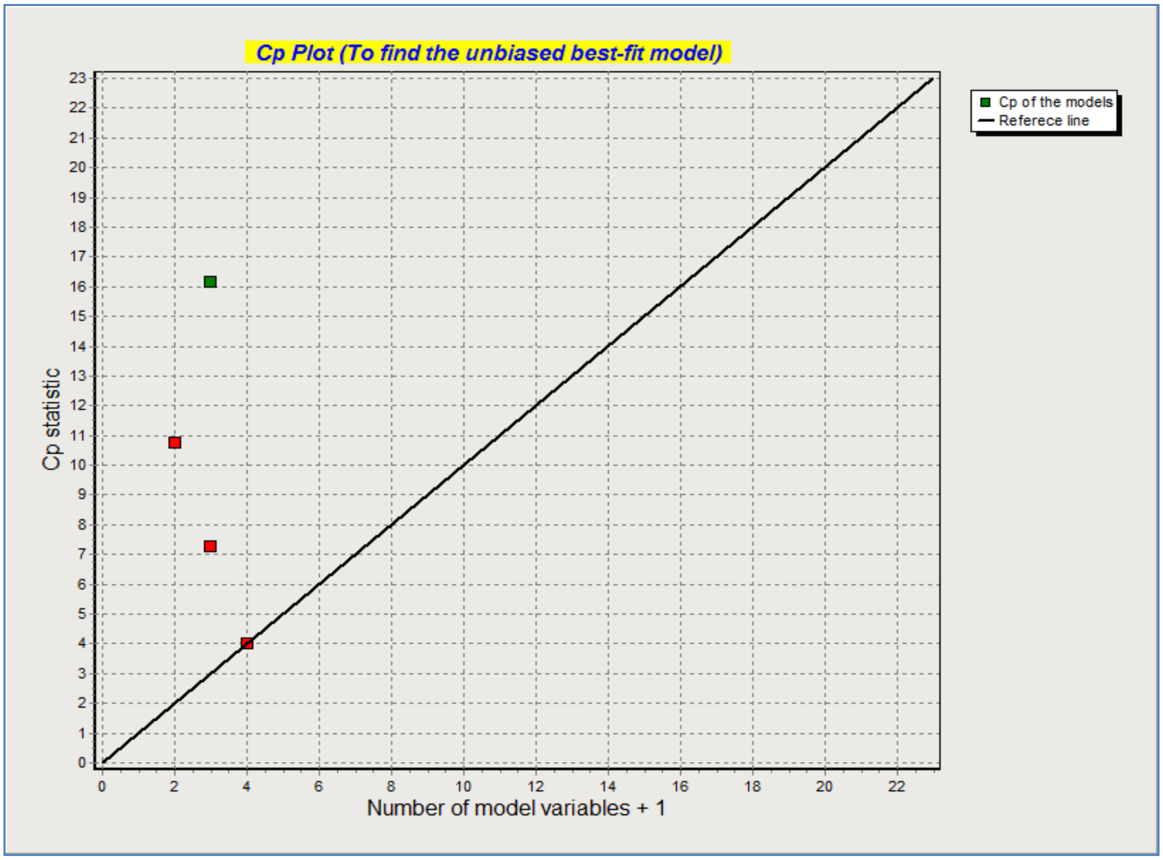

Figure 15: Thompson’s West End Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (4.0) was considered best fit.

Thompson's West End Park Beach 2007 – 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Thompson West End Park beach 2007 – 2008 combined season were (3.15 ± 1.98), 2.71, and 7.79 respectively. The explanatory variables that showed positive correlation with *E. coli* concentration for Thompson's West End Park beach for 2007 – 2008 combined data were included in the modeling process shown in Table 24. The variables that showed significant positive correlation with *E. coli* concentrations were onshore wind ($r = 0.45$, $p<0.05$), gulls ($r = 0.49$, $p<0.05$) and event rain ($r = 0.52$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 24. The scatter plot shown in Figure 16 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (onshore wind, gulls and event rainfall) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (4 days), as seen in Figure 16 and Table 25. The adjusted R-squared (R^2) of the mathematical model was 51.5% (Table 27) with lowest Cp was 4.0 (Table 27 and Figure 17) and the model was considered statistically significant ($p<0.05$), as seen in Table 27 using significant explanatory variables (Table 26). The predictive equation for the combined 2007 – 2008 model is as follows: *E. coli* [Log (MPN)] = [2.1869 + (1.1609*onwind) + (0.06615*gulls) + (2.4377*everain)].

Models were compared for Thompson's West End Park beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 28.

Table 24: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for Thompson's West End Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.23
Wind speed	-0.19
Onshore wind	**0.45****
Water temperature	0.21
Air temperature	0.07
Turbidity	0.31
Wave height	-0.04
Algae	0.04
One day late algae	0.04
Gulls	**0.49****
One day late gulls	0.05
Two days late gulls	0.32
Long shore current speed	0.005
24 hour rainfall	0.19
48 hour rainfall	-0.02
72 hour rainfall	-0.22
Event rainfall	**0.52****
Weighted 72hr rainfall	0.43
Weighted 48hr rainfall	0.51
Combined 72hr rainfall	0.25
Barometric pressure	0.15

**Variable considered statistically significant i.e. $p < 0.05$.

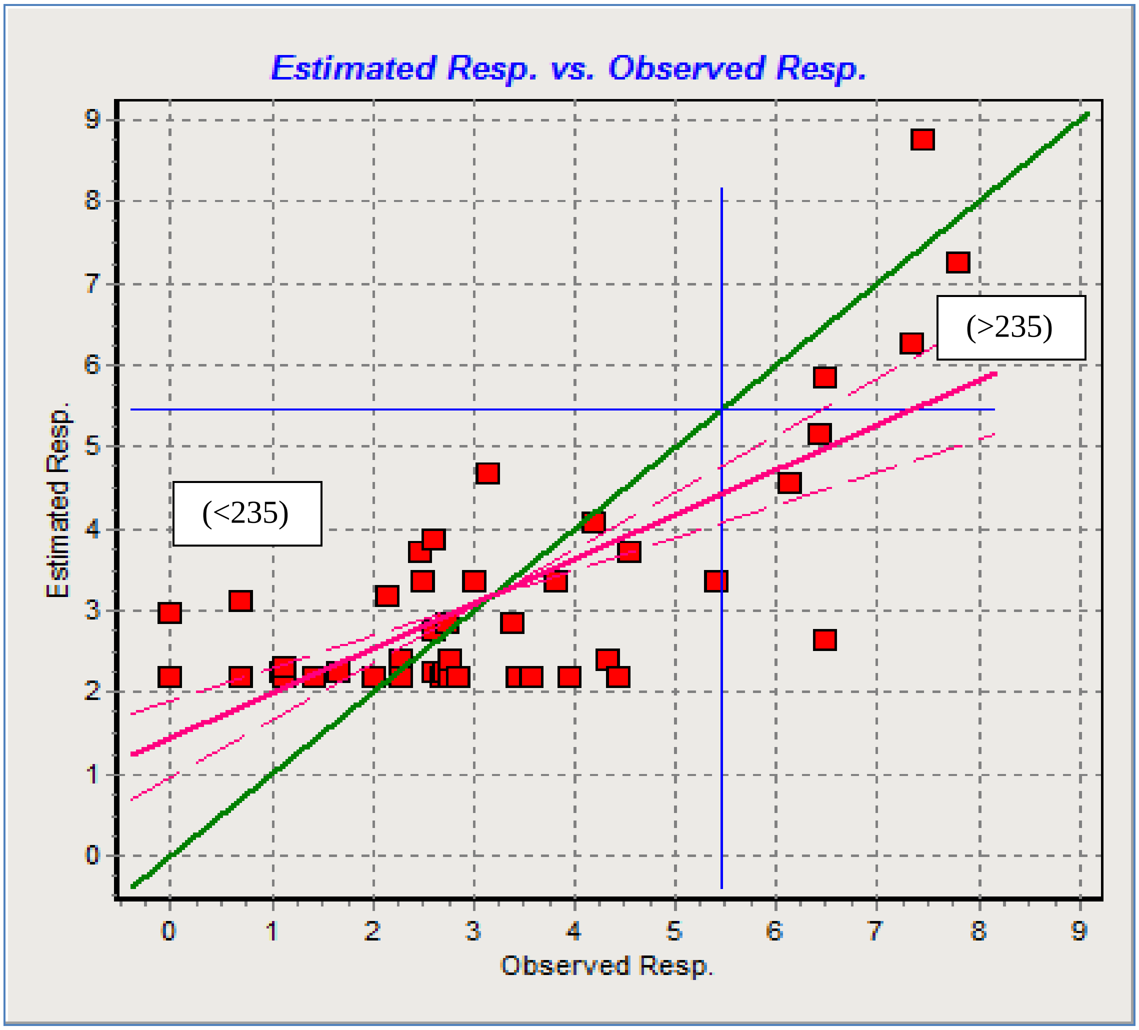

Figure 16: Thompson's West End Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 16, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 51.5%. The green diagonal represents a perfect model for comparison.

Table 25: Thompson's West End Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	46	100
Correct above standard	4	8.7%
Correct below standard	39	84.8%
False positives	0	0%
False negatives	3	6.52%

Table 26: Thompson's West End Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	2.1869	0.2568	8.5175	1.0791E-10
Onshore wind	1.1609	0.5072	2.2888	0.02729
Gulls	0.06615	0.01757	3.7653	0.0005142
Event rain	2.4377	0.5764	4.2291	0.0001249

Table 27: Thompson's West End Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	16.93
Degrees of freedom (DF)	42
Residual standard error (RSE)	1.382
Adjusted R^2	51.5%
Mallow's Cp	4.0
p-value	2.337e-07

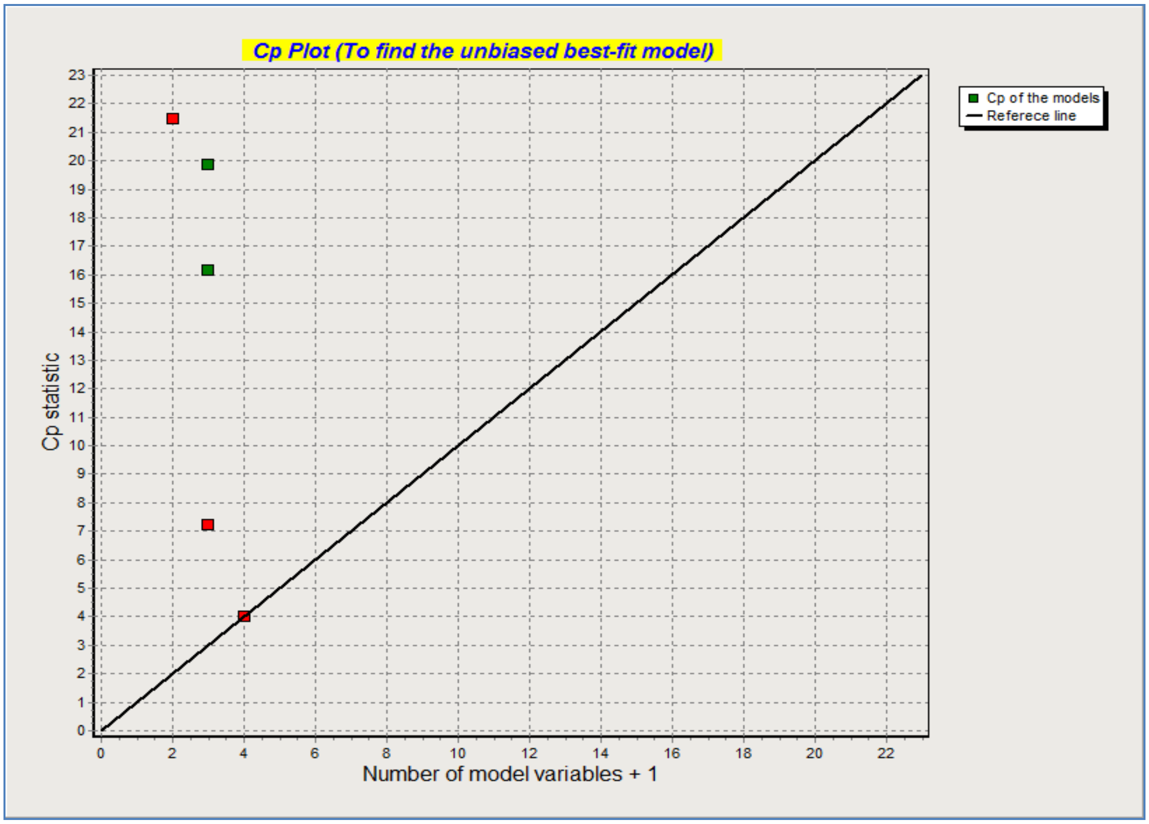

Figure 17: Thompson's West End Park Beach 2007 – 2008 Cp-Plot combined mathematical model. The model with lowest Cp (4.0) was considered best fit.

Table 28: Comparative models for Thompson's West End Park Beach.

Thompson West End Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Onshore wind, √Event rainfall, and weighted 48hr rainfall	Water temperature. One day late algae, and gulls	Onshore wind, Gulls, and Event rain
Number of samples	23	24	47
Correct predictions	22 (95.6%)	23 (95.8%)	44 (93.6%)
Adjusted R^2	77%	71.5%	51.5%
Mallow's Cp	4.0	4.0	4.0
Residual standard error	1.079	0.878	1.382
p-value	<0.001	2.742e-06	2.337e-07
Predictive equation	*E. coli* [Log (MPN)] = 1.939 + (1.824 *onwind) + (2.146*√everain) + (0.460* wferain)	*E. coli* [Log (MPN)] = -2.1272 + (0.24*wt) + (0.9433*oalgae) + (0.08636*gulls)	*E. coli* [Log (MPN)] = 2.1869+ (1.1609*onwind) + (0.06615*gulls) + (2.4377*everain)

Door County

Baileys Harbor Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Baileys Harbor Beach 2008 season were (3.99 ± 1.91), 4.32, and 6.86 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Baileys Harbor Beach are shown in Table 29. The explanatory variables that showed positive correlation with *E. coli* concentration for Baileys Harbor beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were water temperature (r = 0.62, $p<0.05$) and wave height (r = 0.35, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 29. The scatter plot shown in Figure 18 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variables (water temperature and wave height) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (4 days), as seen in Figure 18 and Table 30. The adjusted R-squared (R^2) of the mathematical model was 46.2% (Table 32) with lowest Cp was 4.0 (Table 32 and Figure 19) and the model was considered statistically significant ($p<0.05$), as seen in Table 32 using significant explanatory variables (Table 31). The predictive equation for Baileys Harbor Beach – 2008 model is as follows:

E. coli [Log (MPN)] = [-2.3898 + (0.301*wt) + (0.2608*wh)].

Table 29: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Baileys Harbor Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.06
Wind speed	-0.02
Onshore wind	0.02
Water temperature	**0.62****
Air temperature	0.37
Turbidity	0.23
Wave height	**0.35****
Algae	-0.12
One day late algae	-0.12
Gulls	0.21
One day late gulls	0.08
Two days late gulls	0.09
Long shore current speed	-0.21
Long shore current direction	0.01
24 hour rainfall	0.13
48 hour rainfall	0.07
72 hour rainfall	0.12
Event rainfall	0.25
Weighted 72hr rainfall	0.21
Weighted 48hr rainfall	0.19
Combined 72hr rainfall	0.21
Barometric pressure	-0.07

**Variable considered statistically significant i.e. $p < 0.05$.

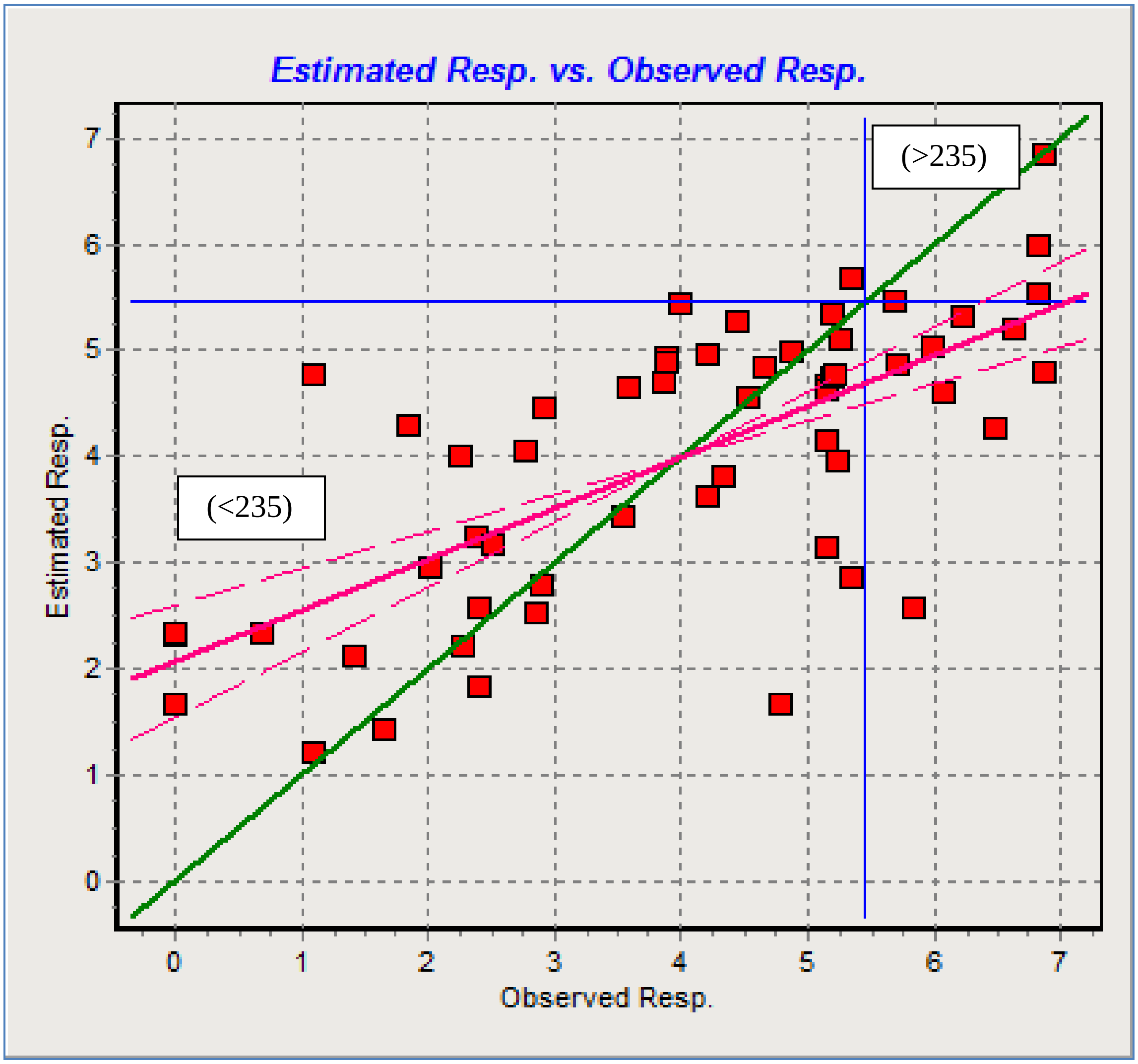

Figure 18: Baileys Harbor Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 18, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 46.2%. The green diagonal represents a perfect model for comparison.

Table 30: Baileys Harbor Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	57	100
Correct above standard	4	7.02%
Correct below standard	44	77.2%
False positives	1	1.75%
False negatives	8	14.0%

Table 31: Baileys Harbor Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-2.3898	0.9449	-2.5292	0.01445
Water temperature	0.301	0.04935	6.0986	1.1940E-7
Wave height	0.2608	0.08652	3.0142	0.003939

Table 32: Baileys Harbor Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	25.05
Degrees of freedom (DF)	54
Residual standard error (RSE)	1.402
Adjusted R^2	46.2%
Mallow's Cp	3.0
p-value	$2.009e^{-08}$

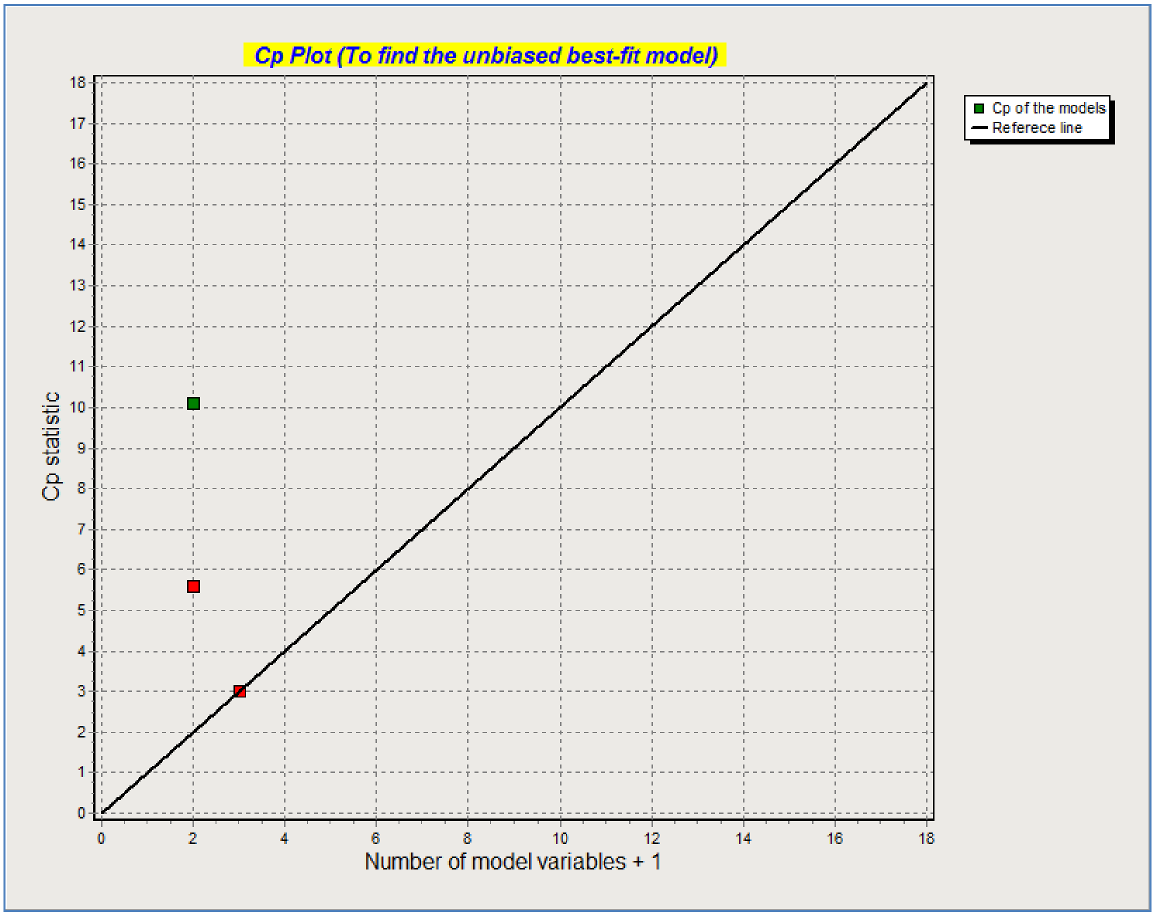

Figure 19: Baileys Harbor Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (3.0) was considered best fit.

Baileys Harbor Beach 2007 – 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Baileys Harbor Beach 2007 – 2008 combined season were (3.27 ± 2.06), 3.09, and 7.79 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Baileys Harbor Beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 33. The variables that showed significant positive correlation were water temperature ($r = 0.34$, $p<0.05$), wave height ($r = 0.49$, $p<0.05$), and gulls ($r = 0.35$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 33. The scatter plot shown in Figure 20 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant explanatory variables (water temperature, wave height, and gulls) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (2 days), as seen in Figure 20 and Table 34. The adjusted R-squared (R^2) of the mathematical model was 37.2% (Table 36) with lowest Cp was 4.0 (Table 36 and Figure 21) and the model was considered statistically significant ($p<0.05$), as seen in Table 36 using significant explanatory variables (Table 35). The predictive equation for the combined 2007 – 2008 model is as follows: *E. coli* [Log (MPN)] = [-0.8025 + (0.1455*wt) + (0.4845*wh) + (0.02503*gulls)].

Models were compared for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 37.

Table 33: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Baileys Harbor Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.15
Wind speed	0.006
Onshore wind	0.18
Water temperature	**0.34****
Air temperature	0.12
Turbidity	0.29
Wave height	**0.49****
Algae	-0.10
One day late algae	-0.27
Gulls	**0.35****
One day late gulls	0.21
Two days late gulls	0.19
Long shore current speed	-0.06
Long shore current direction	0.30
24 hour rainfall	0.07
48 hour rainfall	-0.01
72 hour rainfall	0.11
Event rainfall	0.27
Weighted 72hr rainfall	0.11
Weighted 48hr rainfall	0.08
Combined 72hr rainfall	0.11
Barometric pressure	-0.01

**Variable considered statistically significant i.e. $p < 0.05$.

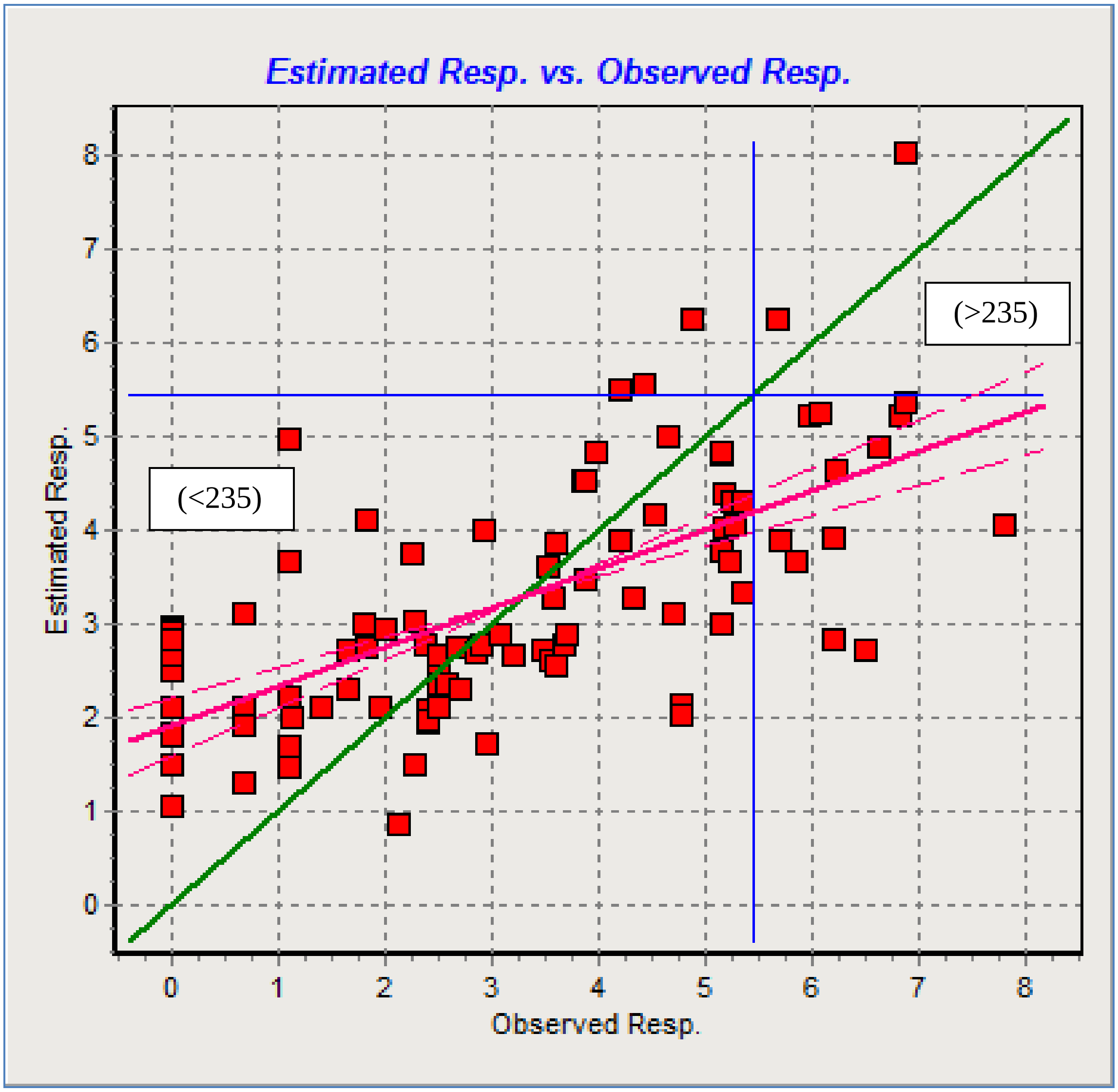

Figure 20: Baileys Harbor 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 20, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 37.2%. The green diagonal represents a perfect model for comparison.

Table 34: Baileys Harbor Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	98	100
Correct above standard	2	2.04%
Correct below standard	80	81.6%
False positives	3	3.06%
False negatives	13	13.3%

Table 35: Baileys Harbor Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-0.8025	0.6726	-0.8852	0.3787
Water temperature	0.1455	0.04677	3.1104	0.002496
Wave height	0.4845	0.08107	5.9767	4.2120E-8
Gulls	0.02503	0.009231	2.7115	0.008019

Table 36: Baileys Harbor Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	23.44
Degrees of freedom (DF)	95
Residual standard error (RSE)	1.514
Adjusted R^2	37.2%
Mallow's Cp	4.0
p-value	$1.929e^{-11}$

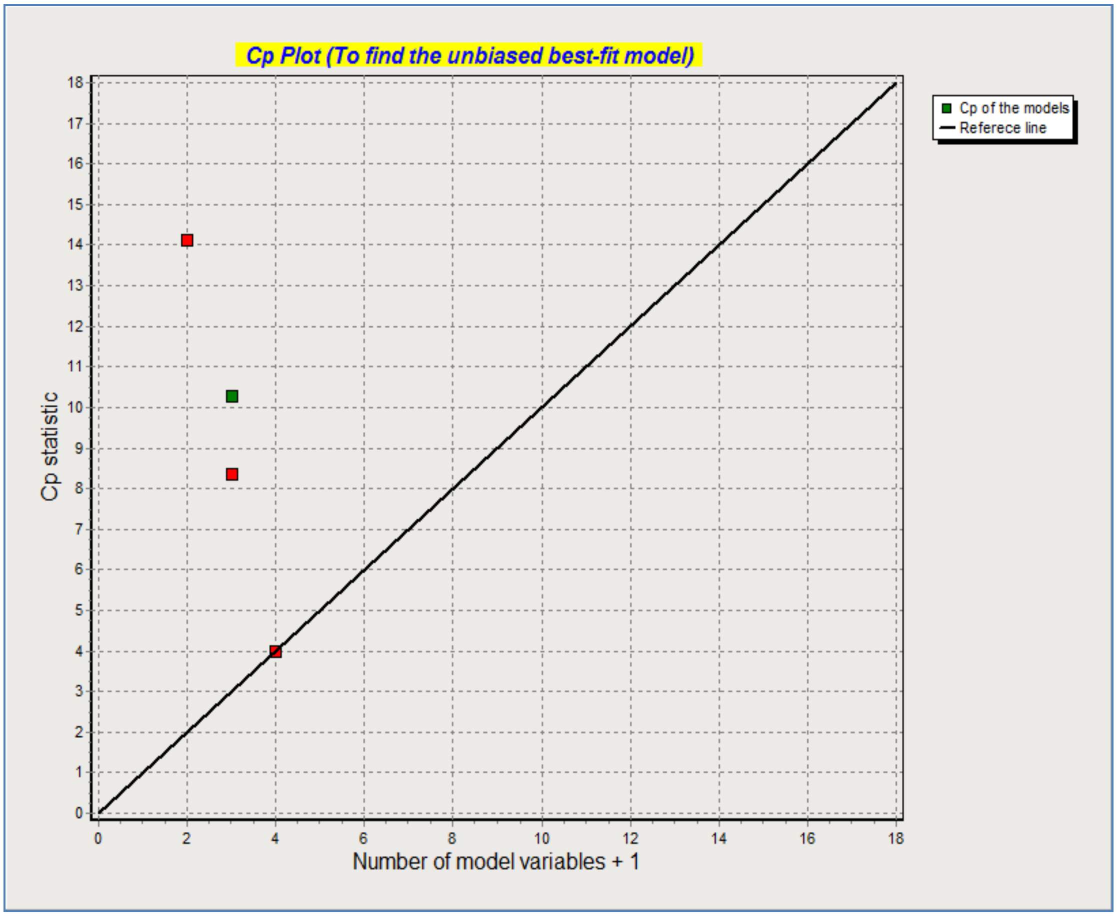

Figure 21: Baileys Harbor Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (4.0) was considered best fit.

Table 37: Comparative models for Baileys Harbor Beach.

Baileys Harbor Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Wave height and Onshore wind	Water temperature, and Wave height	Water temperature, Wave height, and Gulls
Number of samples	40	57	98
Correct predictions	39 (97.5%)	48 (84.2%)	82 (83.6%)
Adjusted R^2	57%	46.2%	37.2%
Mallow's Cp	3.0	3.0	4.0
Residual standard error	1.24	1.402	1.514
p-value	0.0001	$2.009e^{-08}$	$1.929e^{-11}$
Predictive equation	*E. coli* [Log(MPN)] = [0.67 + (1.87*wh) + (1.69*onwind)]	*E. coli* [Log (MPN)] = [-2.3898 + (0.301*wt) + (0.2608*wh)]	*E. coli* [Log (MPN)] = [-0.8025 + (0.1455*wt) + (0.4845*wh) + (0.02503*gulls)]

Ephraim Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Ephraim Park Beach 2008 season were (3.46 ± 1.65), 3.62, and 6.49 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Ephraim Park beach were shown in Table 38. The explanatory variables that showed positive correlation with *E. coli* concentration for Ephraim Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were turbidity ($r = 0.39$, $p<0.05$). Only it was included in the final predictive model, and it is highlighted in Table 38. The scatter plot shown in Figure 22 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variable, turbidity in the mathematical model. The mathematical model predicted six false negative $\log_e$ *E. coli* concentrations, as seen in Figure 22 and Table 39. The adjusted R-squared (R^2) of the mathematical model was 13.4% (Table 41) with lowest Cp was – 3.859 (Table 41 and Figure 23) and the model was considered statistically significant ($p<0.05$), as seen in Table 41 using significant explanatory variables (Table 40). The predictive equation for Ephraim Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] **= [**2.9875 + (0.9063*trbdty)].

Table 38: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Ephraim Park Beach for 2008.

Variable	'r' value
Wind direction (degrees)	-0.001
Wind speed	0.21
Onshore wind	-0.06
Water temperature	0.04
Air temperature	-0.14
Turbidity	**0.39****
Wave height	0.33
Algae	0.08
One day late algae	0.002
Gulls	-0.12
One day late gulls	-0.08
Two days late gulls	0.04
Long shore current speed	0.12
Long shore current direction	0.03
24 hour rainfall	0.08
48 hour rainfall	0.18
72 hour rainfall	0.02
Event rainfall	-0.23
Weighted 72hr rainfall	-0.02
Weighted 48hr rainfall	-0.04
Combined 72hr rainfall	0.02
Barometric pressure	-0.03

**Variable considered statistically significant i.e. $p < 0.05$.

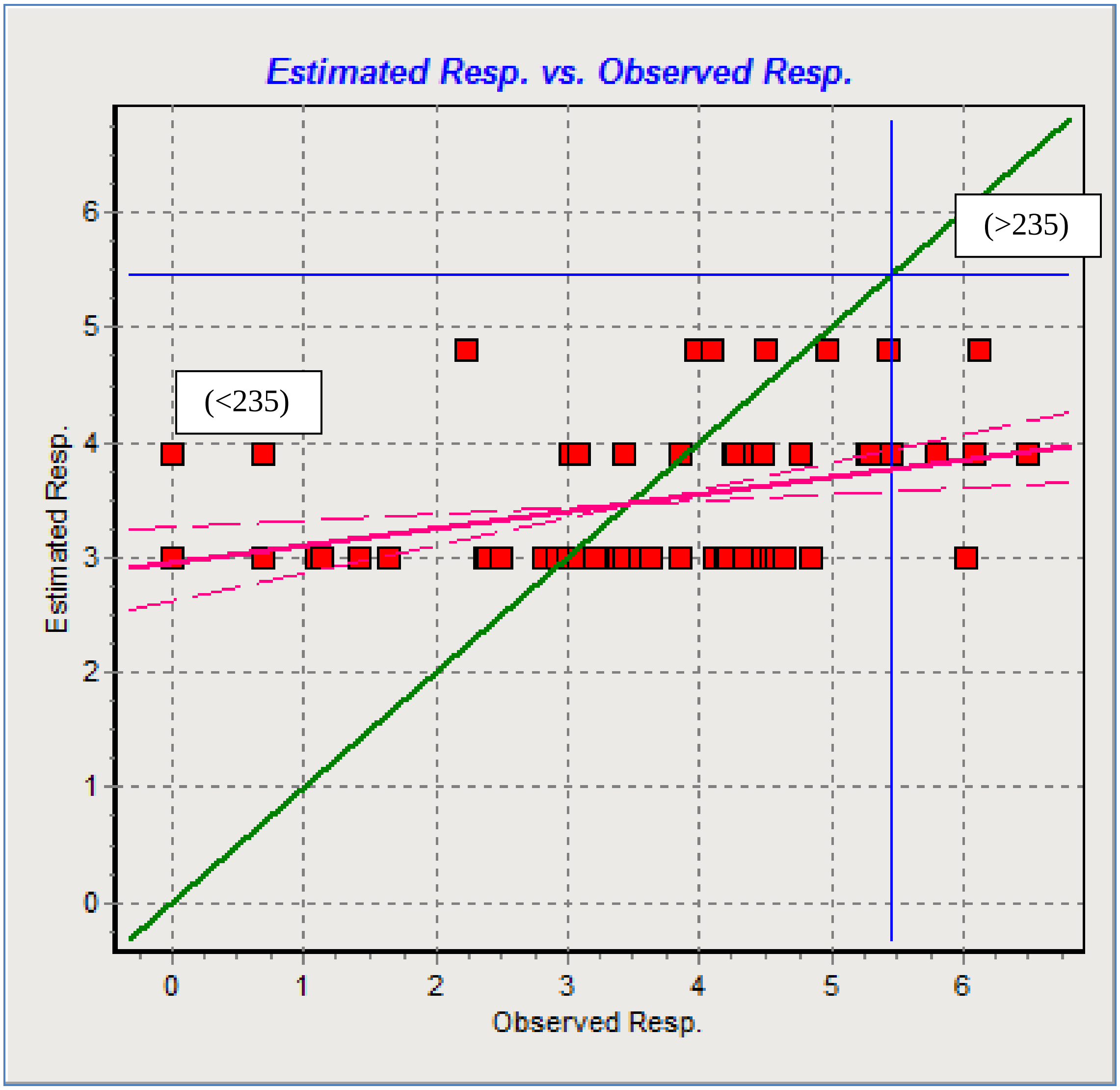

Figure 22: Ephraim Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 22, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the “best fit” line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 13.4%. The green diagonal represents a perfect model for comparison.

Table 39: Ephraim Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	59	100
Correct above standard	0	0 %
Correct below standard	53	89.8%
False positives	0	0 %
False negatives	6	10.2%

Table 40: Ephraim Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	2.9875	0.2508	11.913	4.2147E-17
Turbidity	0.9063	0.2871	3.1561	0.002569

Table 41: Ephraim Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	9.961
Degrees of freedom (DF)	57
Residual standard error (RSE)	1.539
Adjusted R^2	13.4%
Mallow’s Cp	-3.859
p-value	0.002554

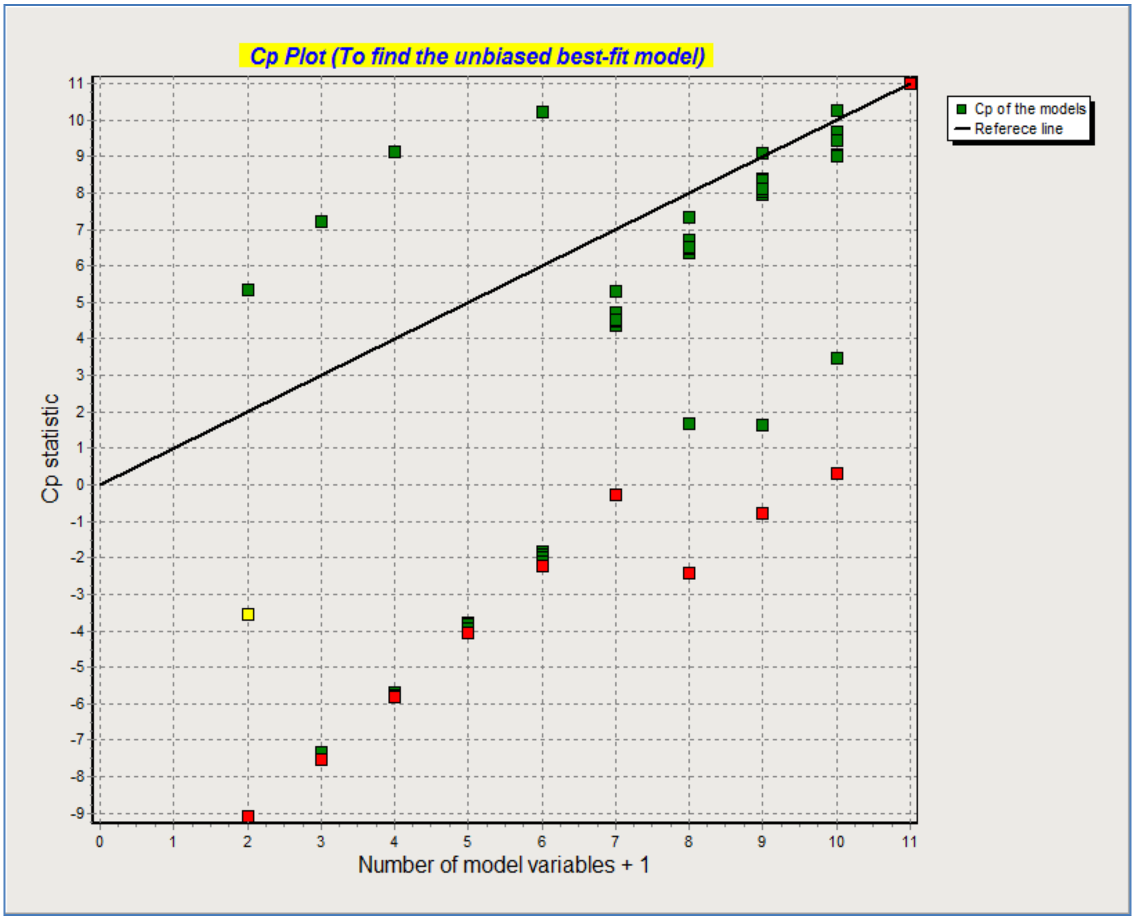

Figure 23: Ephraim Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (-3.859) was considered best fit.

Ephraim Park Beach 2007 - 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Ephraim Park Beach 2007 – 2008 combined season were (2.55 ± 1.90), 2.69, and 6.49 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Ephraim Park beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 42. The variable that showed significant positive correlation was wave height ($r = 0.52$, $p<0.05$). Only it was included in the final predictive model, and it is highlighted in Table 42. The scatter plot shown in Figure 24 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variable, wave height in the combined mathematical model. The mathematical model predicted six false negative $\log_e$ *E. coli* concentrations and *E. coli* concentrations exceeding the bathing water standards (0 days), as seen in Figure 24 and Table 43. The adjusted R-squared (R^2) of the mathematical model was 22.2% (Table 45) with lowest Cp was 2.576 (Table 45 and Figure 25) and the model was considered statistically significant ($p<0.05$), as seen in Table 45 using significant explanatory variable (Table 44). The predictive equation for Ephraim Park Beach 2007 – 2008 combined model is as follows:

E. coli [Log (MPN)] = [1.802 + (0.6377*wh)].

Models were compared for Ephraim Park Beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 46.

Table 42: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Ephraim Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.13
Wind speed	0.13
Onshore wind	-0.02
Water temperature	-0.01
Air temperature	-0.38
Turbidity	0.28
Wave height	**0.52****
Algae	-0.12
One day late algae	-0.26
Gulls	-0.13
One day late gulls	-0.14
Two days late gulls	-0.04
Long shore current speed	-0.17
Long shore current direction	0.39
24 hour rainfall	-0.02
48 hour rainfall	-0.08
72 hour rainfall	-0.13
Event rainfall	-0.002
Weighted 72hr rainfall	-0.10
Weighted 48hr rainfall	-0.06
Combined 72hr rainfall	-0.14
Barometric pressure	0.02

**Variable considered statistically significant i.e. $p < 0.05$.

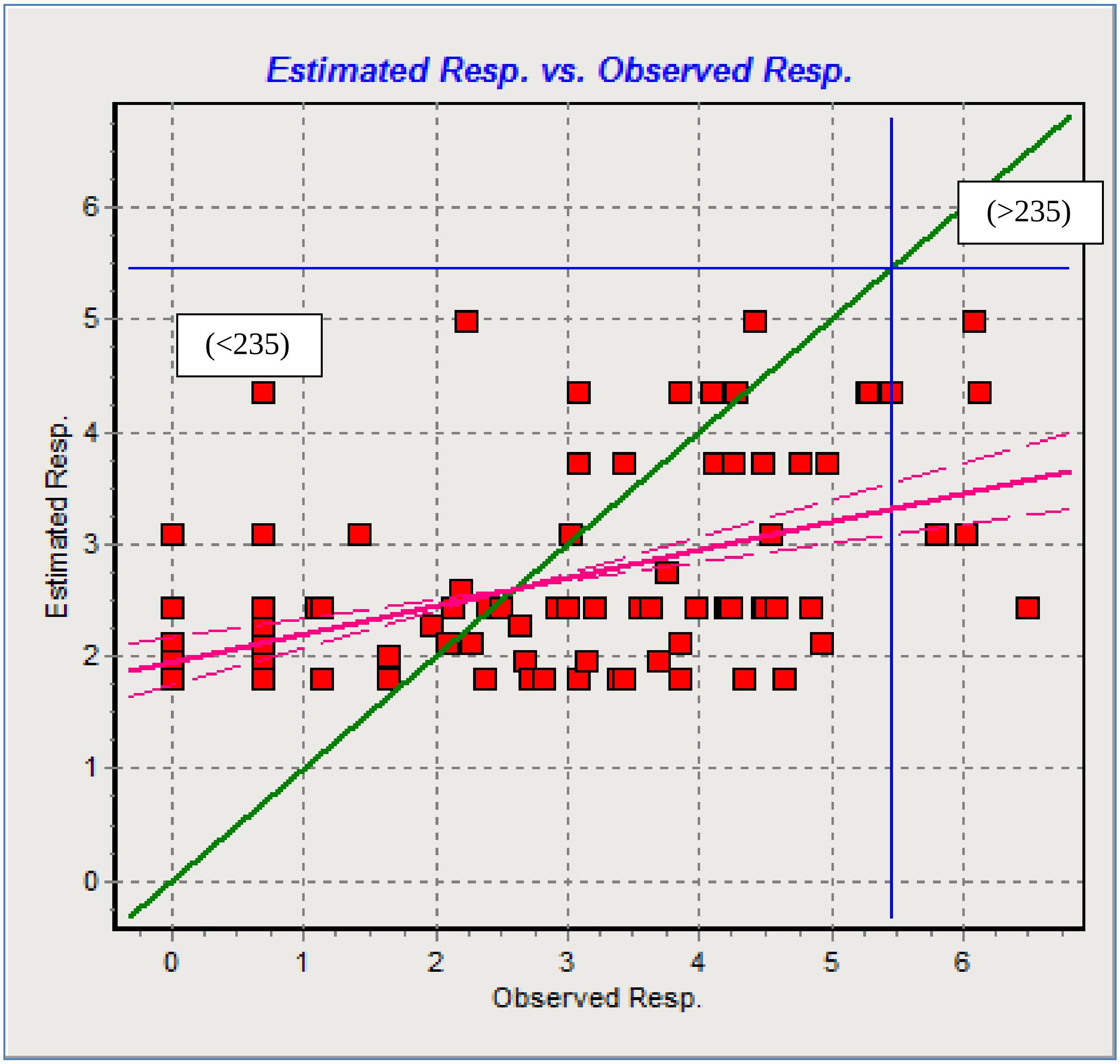

Figure 24: Ephraim Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between estimated and observed *E. coli* concentrations.

In the Figure 24, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 22.2%. The green diagonal represents a perfect model for comparison.

Table 43: Ephraim Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	100	100
Correct above standard	0	0 %
Correct below standard	94	94 %
False positives	0	0 %
False negatives	6	6 %

Table 44: Ephraim Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	1.802	0.1783	6.3395	7.3753E-9
Wave height	0.6377	0.1185	5.3823	5.1700E-7

Table 45: Ephraim Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	37.13
Degrees of freedom (DF)	98
Residual standard error (RSE)	1.509
Adjusted R^2	22.2%
Mallow’s Cp	2.576
p-value	$2.176e^{-08}$

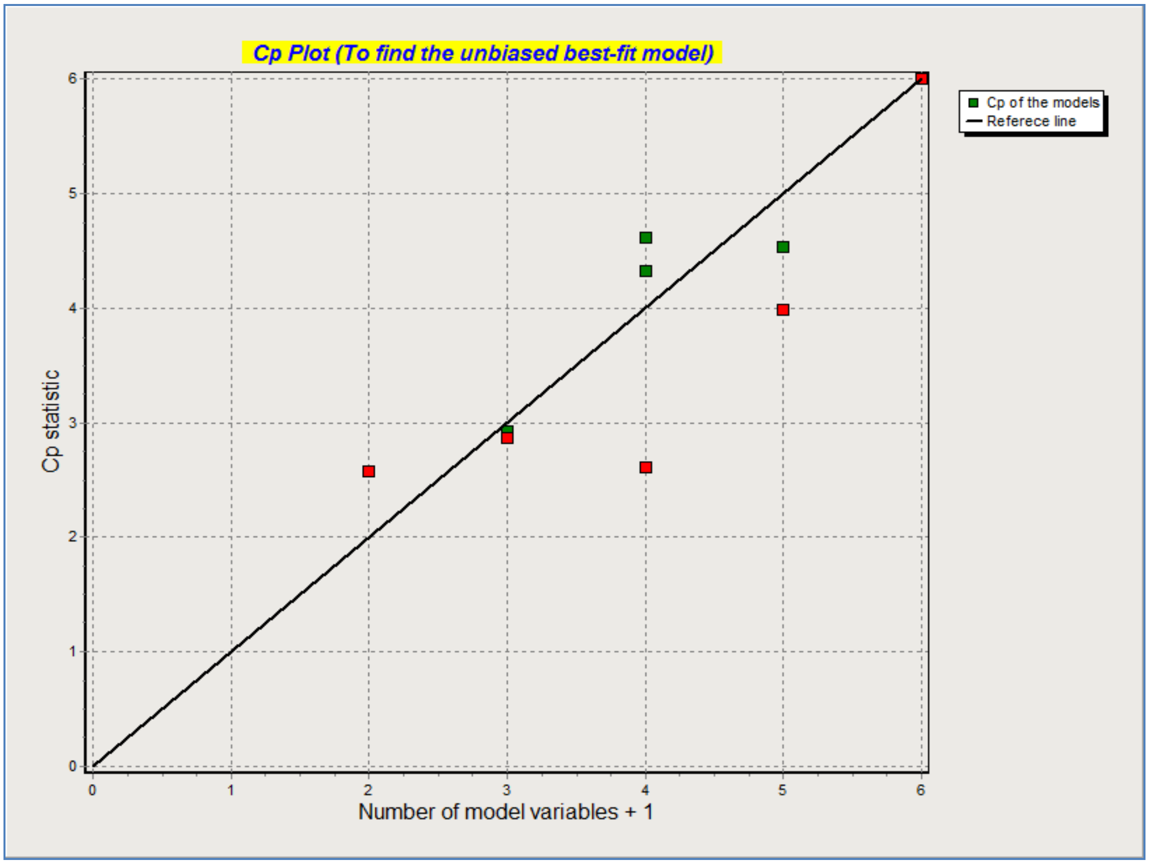

Figure 25: Ephraim Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (2.576) was considered best fit.

Table 46: Comparative models for Ephraim Park Beach.

Ephraim Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Turbidity and Gulls	Turbidity	Wave height
Number of samples	41	59	100
Correct predictions	41 (100%)	53 (89.8%)	94 (94%)
Adjusted R^2	36.9%	13.4%	22.2%
Mallow's Cp	3.0	-3.859	2.576
Residual standard error	1.152	1.539	1.509
p-value	$5.913e^{-05}$	0.002554	$2.176e^{-08}$
Predictive equation	*E. coli* [Log (MPN)] = [0.6038 + (0.9312*trbdty) + (0.2948*gulls)]	*E. coli* [Log (MPN)] **=** **[**2.9875 + (0.9063*trbdty)]	*E. coli* [Log (MPN)] = [1.802 + (0.6377*wh)]

Fish Creek Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Fish Creek Park Beach 2008 season were (3.57 ± 1.45), 3.68, and 6.86 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Fish Creek Park beach were shown in Table 47. The explanatory variables that showed positive correlation with *E. coli* concentration for Fish Creek Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were water temperature (r = 0.31, p<0.05), and turbidity (r = 0.38, p<0.05). Only these were included in the final predictive model, and are highlighted in Table 47. The scatter plot shown in Figure 26 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (water temperature and turbidity) in the mathematical model. The mathematical model predicted two false negative $\log_e$ *E. coli* concentrations, as seen in Figure 26 and Table 48. The adjusted R-squared (R^2) of the mathematical model was 28.8% (Table 50) with lowest Cp was 3.0 (Table 50 and Figure 27) and the model was considered statistically significant (p<0.05), as seen in Table 50 using significant explanatory variables (Table 49). The predictive equation for Fish Creek Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] **= [-** 0.1979 + (0.1805*wt) + (1.4224*trbdty)].

Table 47: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Fish Creek Park Beach for 2008.

Variable	'r' value
Wind direction (degrees)	-0.03
Wind speed	0.31
Onshore wind	-0.09
Water temperature	**0.31****
Air temperature	-0.06
Turbidity	**0.38****
Wave height	0.35
Algae	-0.21
One day late algae	-0.18
Gulls	-0.12
One day late gulls	0.18
Two days late gulls	-0.12
Long shore current speed	-0.07
Long shore current direction	-0.12
24 hour rainfall	0.04
48 hour rainfall	0.12
72 hour rainfall	-0.03
Event rainfall	-0.06
Weighted 72hr rainfall	0.04
Weighted 48hr rainfall	0.03
Combined 72hr rainfall	0.05
Barometric pressure	-0.13

**Variable considered statistically significant i.e. $p < 0.05$.

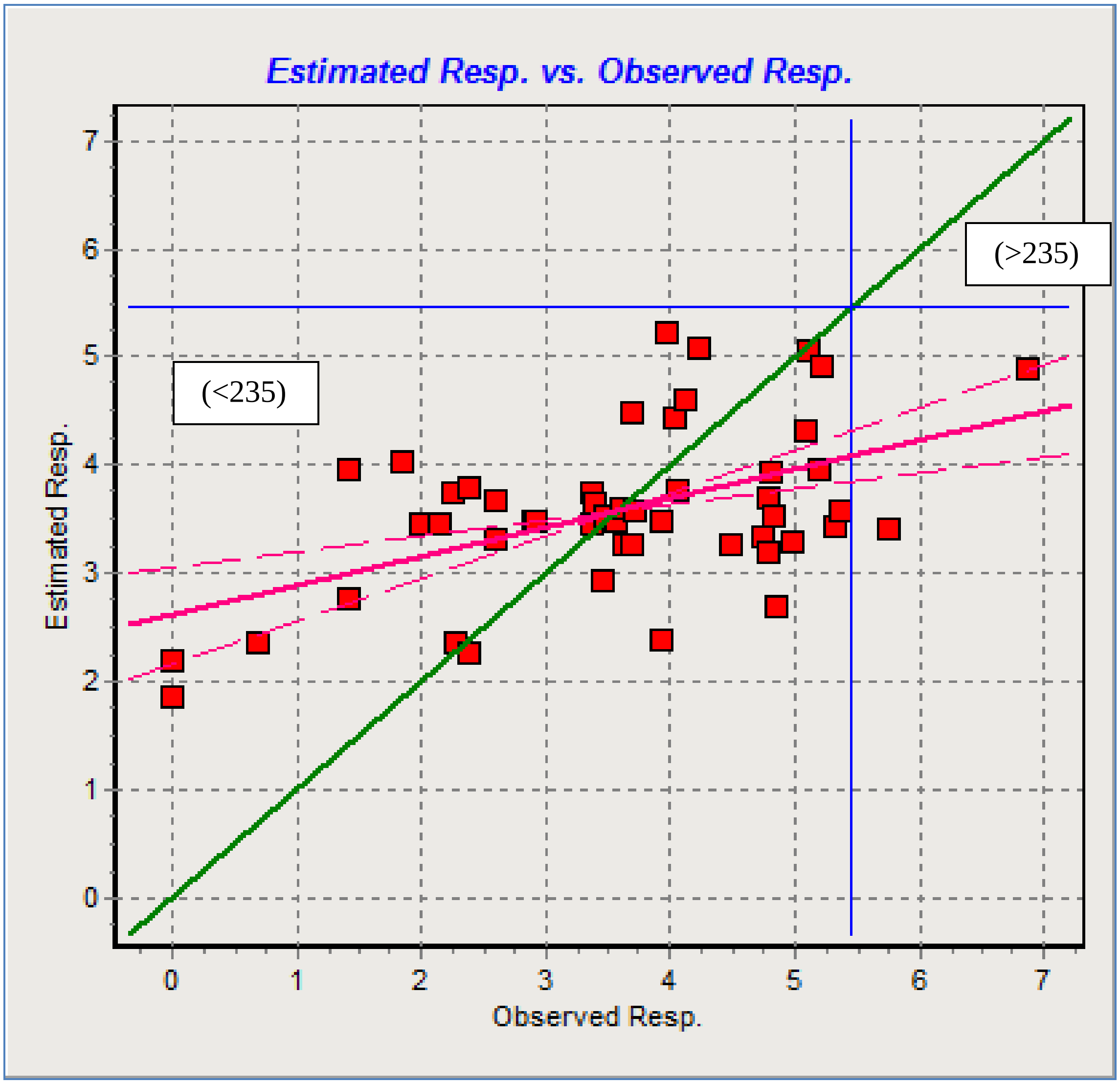

Figure 26: Fish Creek Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 26, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 28.8%. The green diagonal represents a perfect model for comparison.

Table 48: Fish Creek Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	**Number**	**Percentage**
Total predictions	51	100
Correct above standard	0	0 %
Correct below standard	49	96.1%
False positives	0	0 %
False negatives	2	3.92%

Table 49: Fish Creek Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	**Parameter estimate**	**Standard error**	**t-value**	**p-value**
Intercept	-0.1979	1.2327	-0.1605	0.8732
Water temperature	0.1805	0.06334	2.8491	0.00647
Turbidity	1.4224	0.4212	3.3769	0.001469

Table 50: Fish Creek Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	**Value**
F-statistic	8.829
Degrees of freedom (DF)	48
Residual standard error (RSE)	1.269
Adjusted R^2	28.8%
Mallow's Cp	3.0
p-value	0.000543

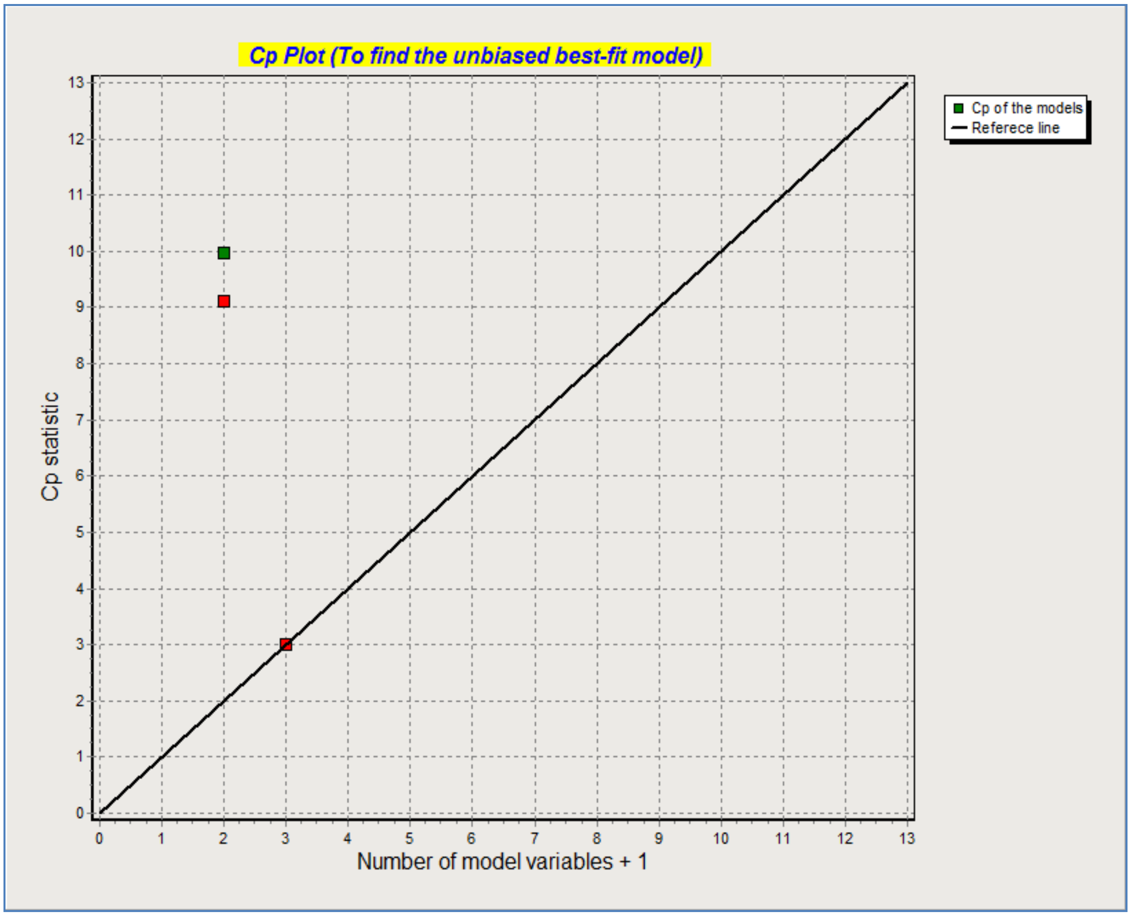

Figure 27: Fish Creek Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (3.0) was considered best fit.

Fish Creek Park Beach 2007 – 2008 combined season model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Fish Creek Park Beach 2007 – 2008 combined season were (2.82 ± 1.64), 2.9, and 6.86 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Fish Creek Park beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 51. The variable that showed significant positive correlation was wave height ($r = 0.49$, $p<0.05$). Only it was included in the final predictive model, and it is highlighted in Table 51. The scatter plot shown in Figure 28 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variable, wave height in the mathematical model. The mathematical model predicted three false negative $\log_e$ *E. coli* concentrations, as seen in Figure 28 and Table 52. The adjusted R-squared (R^2) of the mathematical model was 19.3% (Table 54) with lowest Cp was 3.631 (Table 54 and Figure 29) and the model was considered statistically significant ($p<0.05$), as seen in Table 54 using significant explanatory variables (Table 53). The predictive equation for Fish Creek Park Beach 2007 – 2008 combined model is as follows:

E. coli [Log (MPN)] = [2.3552 + (0.6485*wh)].

Models were compared for Fish Creek Park Beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 55.

Table 51: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Fish Creek Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.17
Wind speed	0.14
Onshore wind	-0.02
Water temperature	-0.04
Air temperature	-0.45
Turbidity	0.06
Wave height	**0.49****
Algae	-0.22
One day late algae	-0.24
Gulls	-0.23
One day late gulls	-0.11
Two days late gulls	-0.15
Long shore current speed	-0.11
Long shore current direction	0.28
24 hour rainfall	0.07
48 hour rainfall	-0.09
72 hour rainfall	-0.02
Event rainfall	0.09
Weighted 72hr rainfall	0.02
Weighted 48hr rainfall	0.04
Combined 72hr rainfall	0.001

**Variable considered statistically significant i.e. $p < 0.05$.

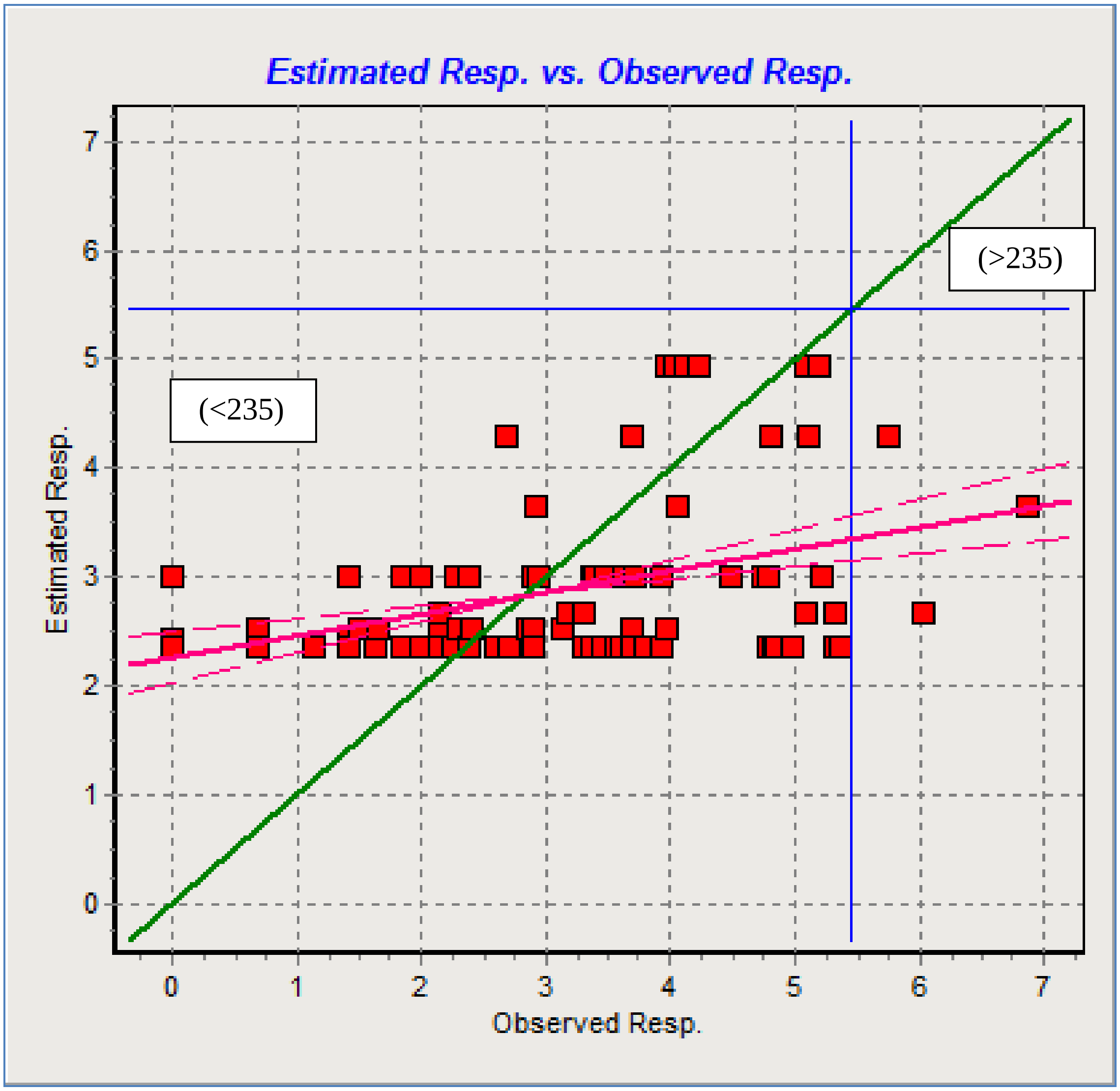

Figure 28: Fish Creek Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 28, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 19.3%. The green diagonal represents a perfect model for comparison.

Table 52: Fish Creek Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	97	100
Correct above standard	0	0 %
Correct below standard	94	96.9%
False positives	0	0 %
False negatives	3	3.09%

Table 53: Fish Creek Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	2.3552	0.1778	13.244	0.0
Wave height	0.6485	0.1327	4.888	4.1735E-6

Table 54: Fish Creek Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	23.89
Degrees of freedom (DF)	95
Residual standard error (RSE)	1.474
Adjusted R^2	19.2%
Mallow’s Cp	3.631
p-value	$4.131e^{-06}$

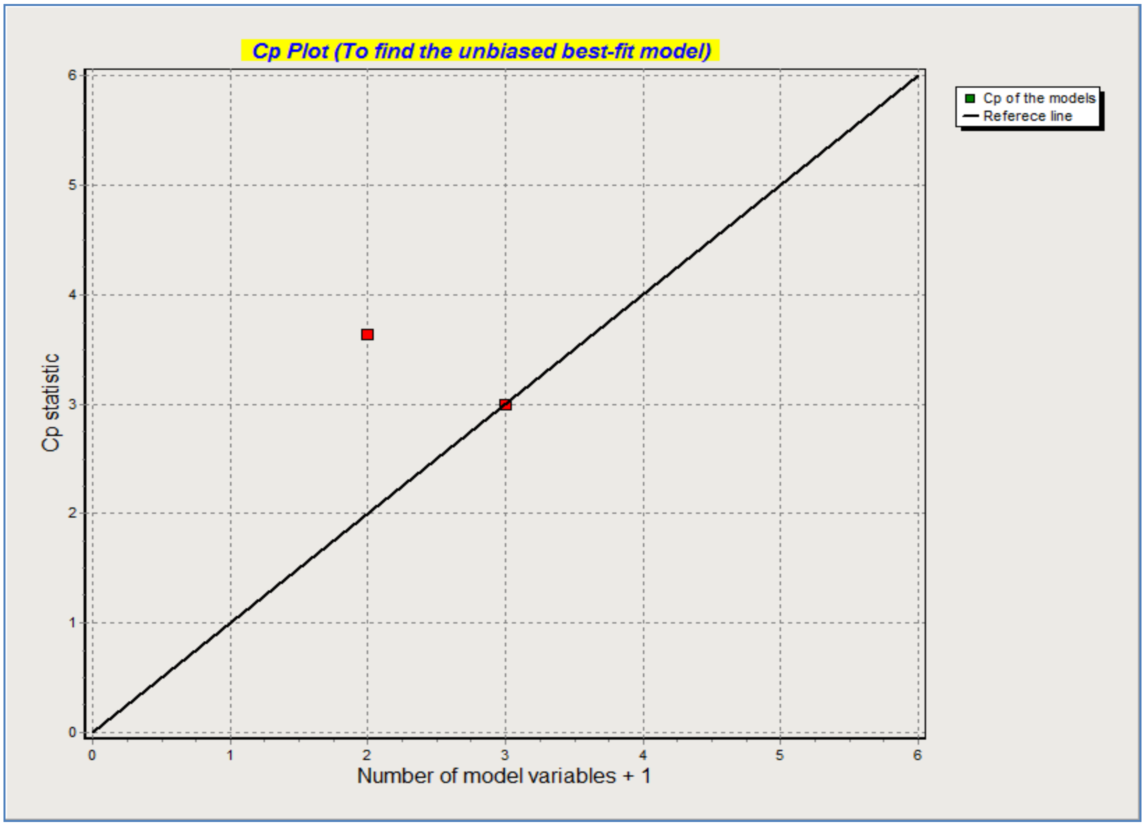

Figure 29: Fish Creek Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (3.631) was considered best fit.

Table 55: Comparative models for Fish Creek Park Beach.

Fish Creek Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Turbidity and Long shore current speed	Water temperature and Turbidity	Wave height
Number of samples	45	51	97
Correct predictions	44 (97.7%)	49 (96%)	94 (96.9%)
Adjusted R^2	28.5%	28.8%	19.2%
Mallow's Cp	3.0	3.0	3.631
Residual standard error	1.217	1.269	1.474
p-value	0.004723	0.000543	$4.131e^{-06}$
Predictive equation	*E. coli* [Log (MPN)] = [0.05246 + (0.7567*trbdty) + (0.1548*lcs)]	*E. coli* [Log (MPN)] **=** **[- 0**.1979 + (0.1805*wt) + (1.4224*trbdty)]	*E. coli* [Log (MPN)] = 2.3552 + (0.6485*wh)

Lakeside Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Lakeside Park beach 2008 season were (3.81 ± 1.41), 3.99, and 6.67 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Lakeside Park beach are shown in Table 56. The explanatory variables that showed positive correlation with *E. coli* concentration for Lakeside Park beach for 2008 were included in modeling process. The variable that showed the significant positive correlation was water temperature (r = 0.37, p<0.05). Only it was included in the final predictive model, and it is highlighted in Table 56. The scatter plot shown in Figure 30 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variable, water temperature, in the mathematical model. The mathematical model predicted three false negative $\log_e$ *E. coli* concentrations, as seen in Figure 30 and Table 57. The adjusted R-squared (R^2) of the mathematical model was 11.4% (Table 59) with lowest Cp was - 28.05 (Table 59 and Figure 31) and the model was considered statistically significant (p<0.05), as seen in Table 59 using significant explanatory variables (Table 58). The predictive equation for Lakeside Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] **= [**1.6021 + (0.1245*wt)].

Table 56: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Lakeside Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.004
Wind speed	-0.24
Onshore wind	-0.21
Water temperature	**0.37****
Air temperature	0.10
Turbidity	0.004
Wave height	0.21
Algae	-0.38
One day late algae	-0.30
Gulls	-0.18
One day late gulls	0.01
Two days late gulls	-0.06
Long shore current speed	-0.17
Long shore current direction	-0.35
24 hour rainfall	0.15
48 hour rainfall	-0.14
72 hour rainfall	0.07
Event rainfall	0.08
Weighted 72hr rainfall	0.13
Weighted 48hr rainfall	0.13
Combined 72hr rainfall	0.09
Barometric pressure	-0.02

**Variable considered statistically significant i.e. $p < 0.05$.

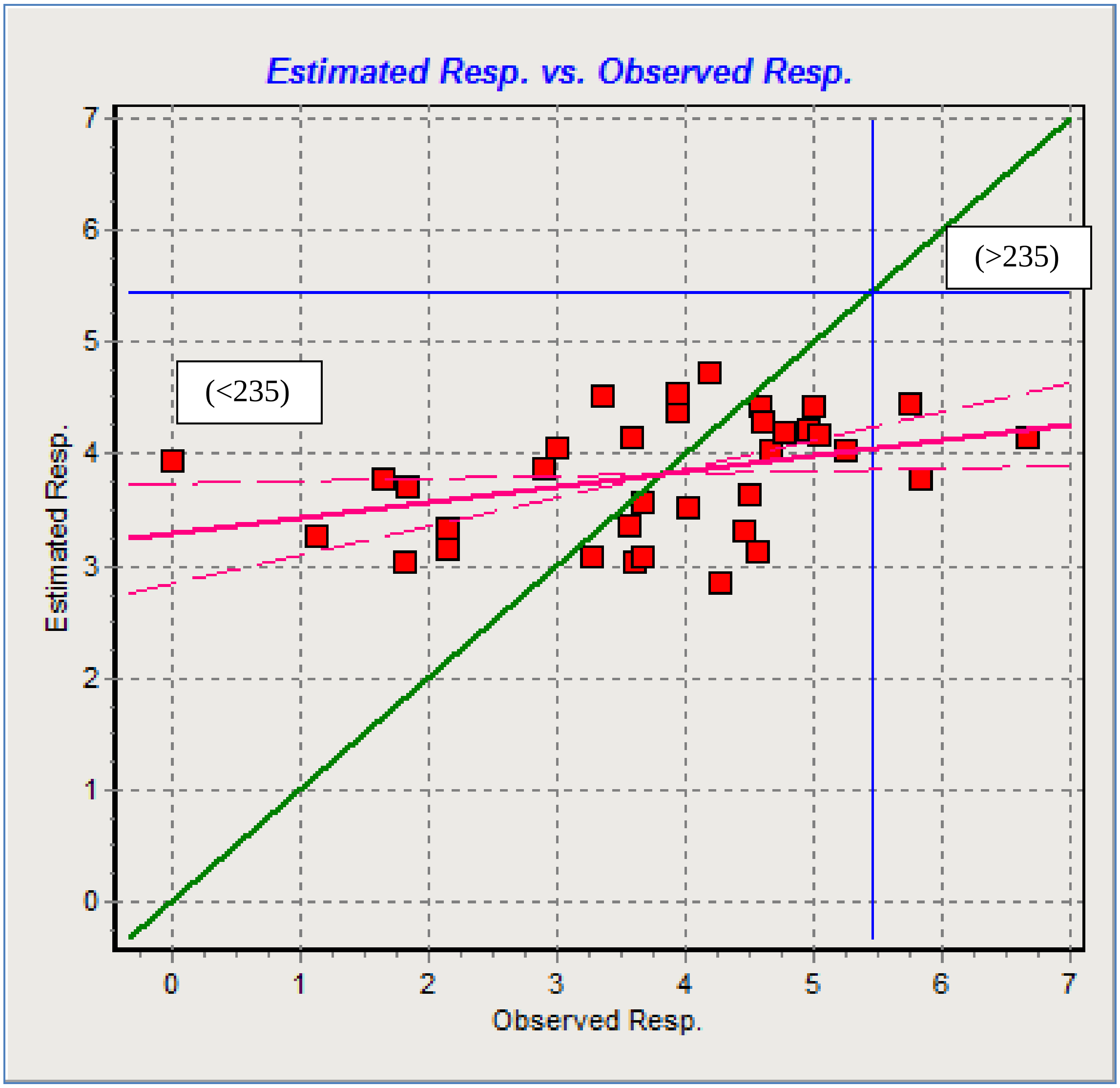

Figure 30: Lakeside Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 30, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 11.4%. The green diagonal represents a perfect model for comparison.

Table 57: Lakeside Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	36	100
Correct above standard	0	0 %
Correct below standard	33	91.7%
False positives	0	0 %
False negatives	3	8.33%

Table 58: Lakeside Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	1.6021	0.9705	1.6507	0.1083
Water temperature	0.1245	0.05307	2.3458	0.02504

Table 59: Lakeside Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	5.503
Degrees of freedom (DF)	34
Residual standard error (RSE)	1.335
Adjusted R^2	11.4%
Mallow's Cp	-28.05
p-value	0.02496

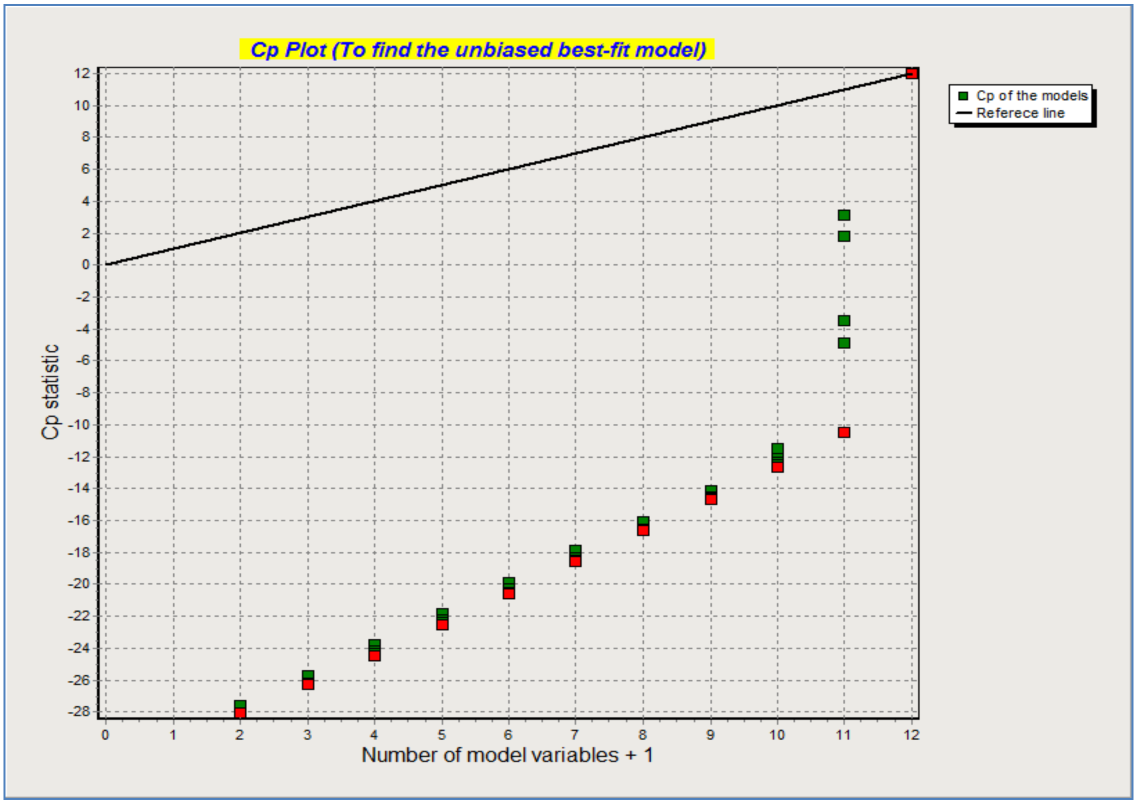

Figure 31: Lakeside Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (-28.05) was considered best fit.

Lakeside Park Beach 2007 – 2008 combined season model.

The mean (MPN±SD), median, and maximum $\log_e$ *E. coli* concentrations for Lakeside Park Beach 2007 – 2008 combined season were (3.11 ± 1.71), 3.36, and 6.67 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Lakeside Park beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 60. The variables that showed significant positive correlation were wave height (r = 0.40, $p<0.05$), turbidity (r = 0.26, $p<0.05$), and twenty-four hour (24hr) rainfall (r = 0.29, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 60. The scatter plot shown in Figure 32 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variables (wave height, turbidity, and twenty four-hour rainfall) in the mathematical model. The mathematical model predicted three false negative $\log_e$ *E. coli* concentrations and three false positive $\log_e$ *E. coli* concentrations, as seen in Figure 32 and Table 61. The adjusted R-squared (R^2) of the mathematical model was 33.3% (Table 63) with lowest Cp was -3.497 (Table 63 and Figure 33) and the model was considered statistically significant ($p<0.05$), as seen in Table 63 using significant explanatory variables (Table 62). The predictive equation for Lakeside Park Beach 2007 – 2008 combined model is as follows: *E. coli* [Log (MPN)] = [1.7624 + (0.642*trbdty) + (0.4035*wh) + (0.7586*tfrain)].

Models were compared for Lakeside Park Beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 64.

Table 60: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Lakeside Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.05
Wind speed	0.03
Onshore wind	-0.07
Water temperature	0.14
Air temperature	-0.08
Turbidity	**0.26****
Wave height	**0.40****
Algae	0.004
One day late algae	-0.15
Gulls	-0.46
One day late gulls	-0.22
Two days late gulls	-0.10
Long shore current speed	0.07
Long shore current direction	0.19
24 hour rainfall	0.29
48 hour rainfall	-0.02
72 hour rainfall	0.10
Event rainfall	0.19
Weighted 72hr rainfall	0.28
Weighted 48hr rainfall	0.32
Combined 72hr rainfall	0.18
Barometric pressure	0.02

**Variable considered statistically significant i.e. $p < 0.05$.

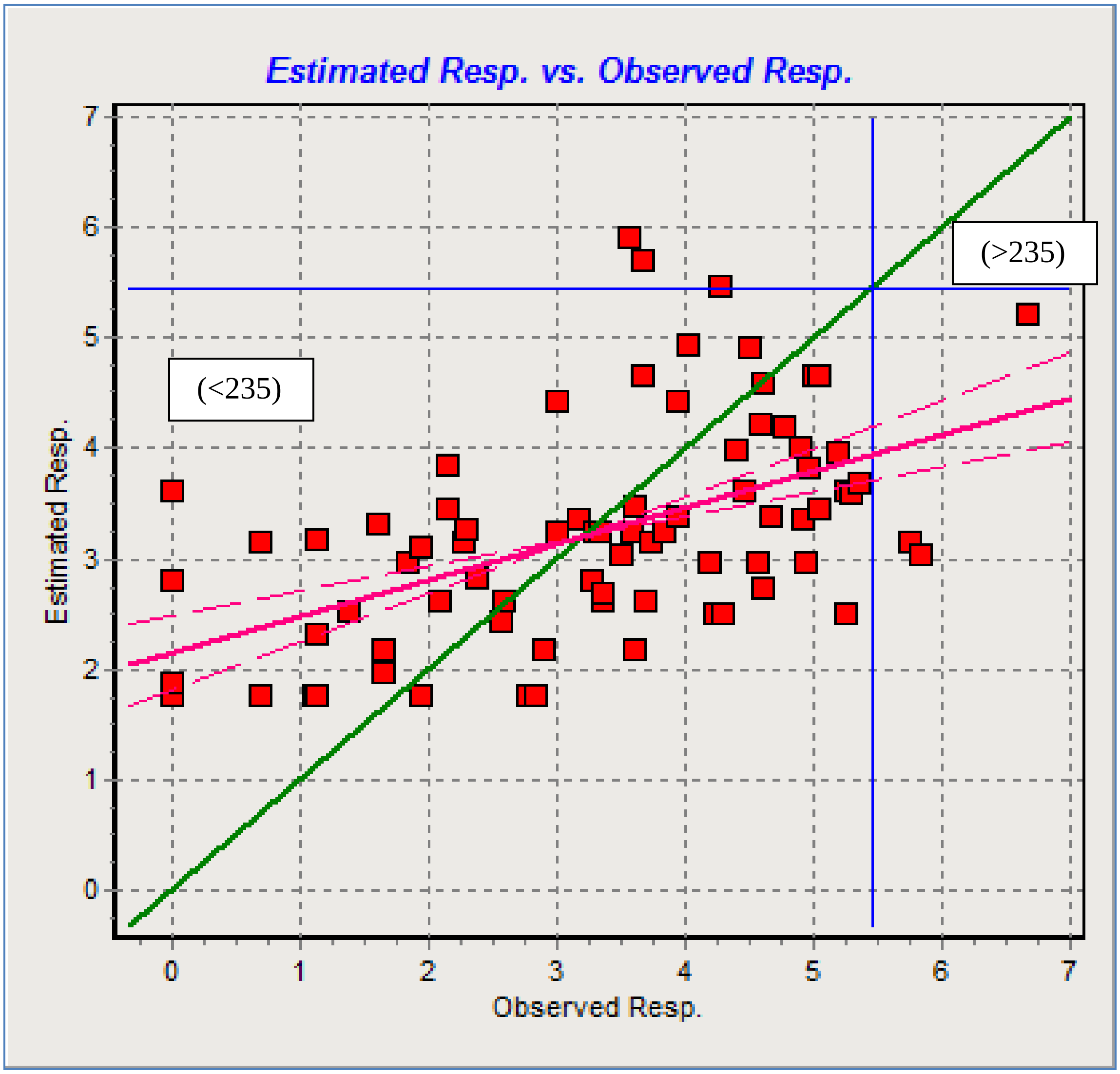

Figure 32: Lakeside Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 32, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it. The green diagonal represents a perfect model for comparison.

Table 61: Lakeside Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	82	100
Correct above standard	0	0 %
Correct below standard	76	92.7%
False positives	3	3.66%
False negatives	3	3.66%

Table 62: Lakeside Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	1.7624	0.213	5.5889	3.3528E-7
Turbidity	0.642	0.184	3.4895	0.0008094
Wave height	0.4035	0.1046	3.8578	0.0002379
24 hour rainfall	0.7586	0.3143	2.4135	0.01825

Table 63: Lakeside Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	10.25
Degrees of freedom (DF)	79
Residual standard error (RSE)	1.320
Adjusted R^2	33.3%
Mallow's Cp	-3.497
p-value	$8.93e^{-06}$

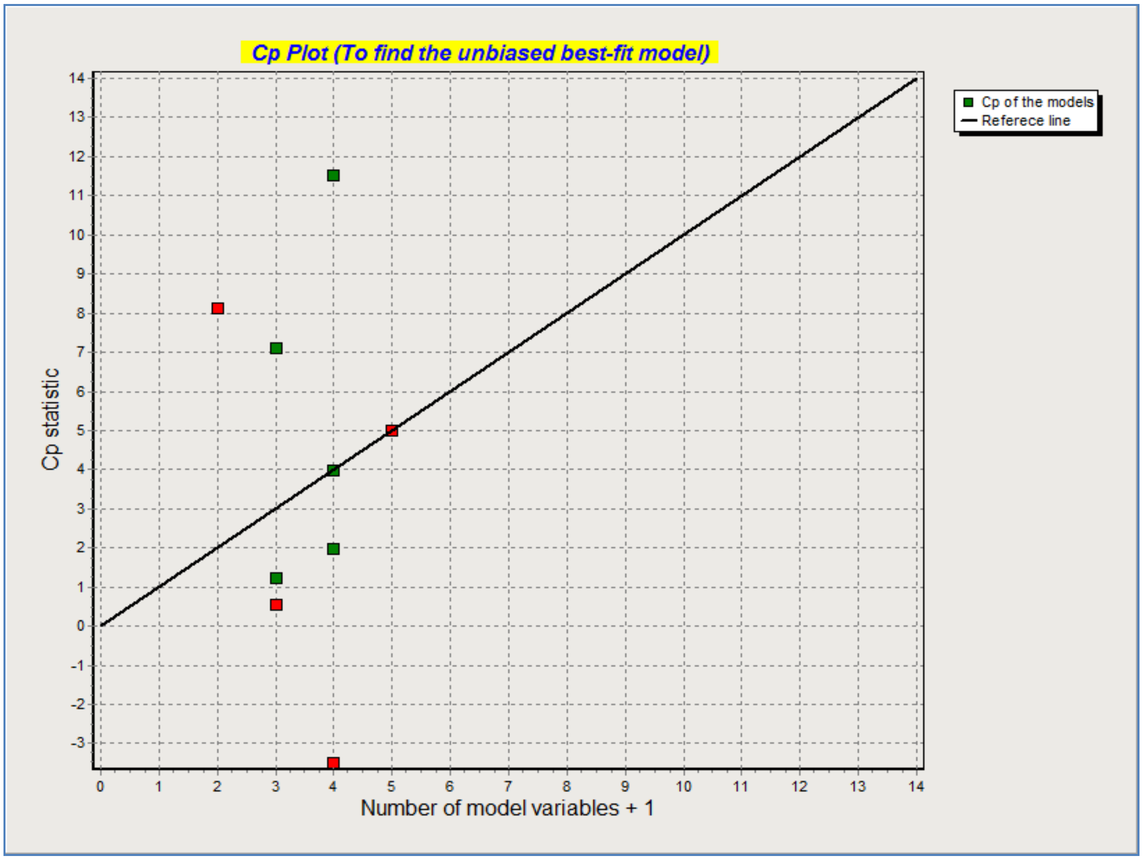

Figure 33: Lakeside Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (-3.497) was considered best fit.

Table 64: Comparative models for Lakeside Park Beach.

Lakeside Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Turbidity, Wave height, and Algae	Water temperature	Turbidity, Wave height, and 24hr rainfall
Number of samples	47	36	82
Correct predictions	44 (93.6%)	33 (91.6%)	76 (92.6%)
Adjusted R^2	47%	11.4%	33.3%
Mallow's Cp	4.0	-28.05	-3.497
Residual standard error	1.31	1.335	1.320
p-value	<0.005	0.02496	$8.93e^{-06}$
Predictive equation	*E. coli* [Log(MPN)] = [0.41+ (0.49*trbdty) + (0.6*algae) + (1.44*wh)]	*E. coli* [Log (MPN)] **=** **[**1.6021 + (0.1245*wt)].	*E. coli* [Log (MPN)] = [1.7624 + (0.642*trbdty) + (0.4035*wh) + (0.7586*tfrain)]

Murphy Park Beach – 2008 model.

The explanatory variables were not significantly correlated ($p > 0.05$) with *E. coli* concentration for data collected in the 2008 recreational season from Murphy Park beach. Therefore, the mathematical model was not developed.

Murphy Park Beach 2007 – 2008 combined season model.

The mean (MPN±SD), median, and maximum $\log_e$ *E. coli* concentrations for Murphy Park Beach 2007 – 2008 combined season were (2.46 ± 1.83), 2.58, and 7.59 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Murphy Park beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 65. The variable that showed significant positive correlation was wave height ($r = 0.43$, $p<0.05$). Only it was included in the final predictive model, and it is highlighted in Table 65. The scatter plot shown in Figure 34 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variable, wave height, in the mathematical model. The mathematical model predicted three false negative $\log_e$ *E. coli* concentrations, as seen in Figure 34 and Table 66. The adjusted R-squared (R^2) of the mathematical model was 10.1% (Table 68) with lowest Cp was – 1.994 (Table 68 and Figure 35) and the model was considered statistically significant ($p<0.05$), as seen in Table 68 using significant explanatory variables (Table 67). The predictive equation for Murphy Park Beach 2007 – 2008 combined model is as follows:

E. coli [Log (MPN)] = [2.0264 + (0.4076*wh)].

Models were compared for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 69.

Table 65: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Murphy Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.06
Wind speed	-0.16
Onshore wind	0.0008
Water temperature	0.07
Air temperature	-0.19
Turbidity	0.16
Wave height	**0.43****
Algae	-0.02
One day late algae	0.03
Gulls	0.14
One day late gulls	0.16
Two days late gulls	0.16
Long shore current speed	-0.13
Long shore current direction	0.33
24 hour rainfall	0.11
48 hour rainfall	-0.11
72 hour rainfall	-0.13
Event rainfall	0.14
Weighted 72hr rainfall	0.08
Weighted 48hr rainfall	0.11
Combined 72hr rainfall	0.03
Barometric pressure	-0.02

**Variable considered statistically significant i.e. $p < 0.05$.

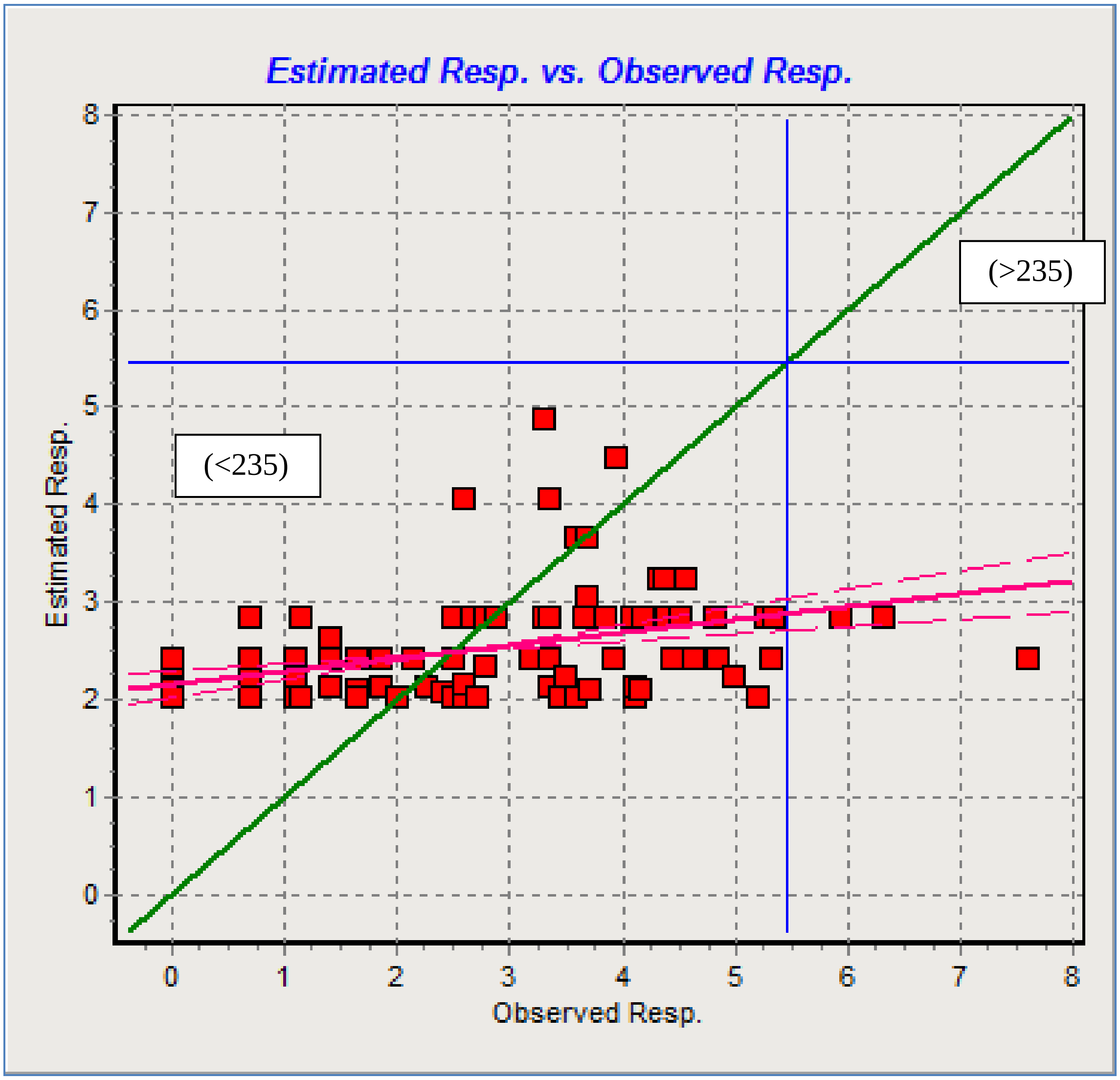

Figure 34: Murphy Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 34, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 10.1%. The green diagonal represents a perfect model for comparison.

Table 66: Murphy Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	**Number**	**Percentage**
Total predictions	97	100
Correct above standard	0	0 %
Correct below standard	94	96.9%
False positives	0	0 %
False negatives	3	3.09%

Table 67: Murphy Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	**Parameter estimate**	**Standard error**	**t-value**	**p-value**
Intercept	2.0264	0.1788	7.2548	1.1444E-10
Wave height	0.4076	0.1191	3.4224	0.0009272

Table 68: Murphy Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	**Value**
F-statistic	20.9
Degrees of freedom (DF)	94
Residual standard error (RSE)	1.529
Adjusted R^2	10.1%
Mallow's Cp	-1.994
p-value	1.470e-^{05}

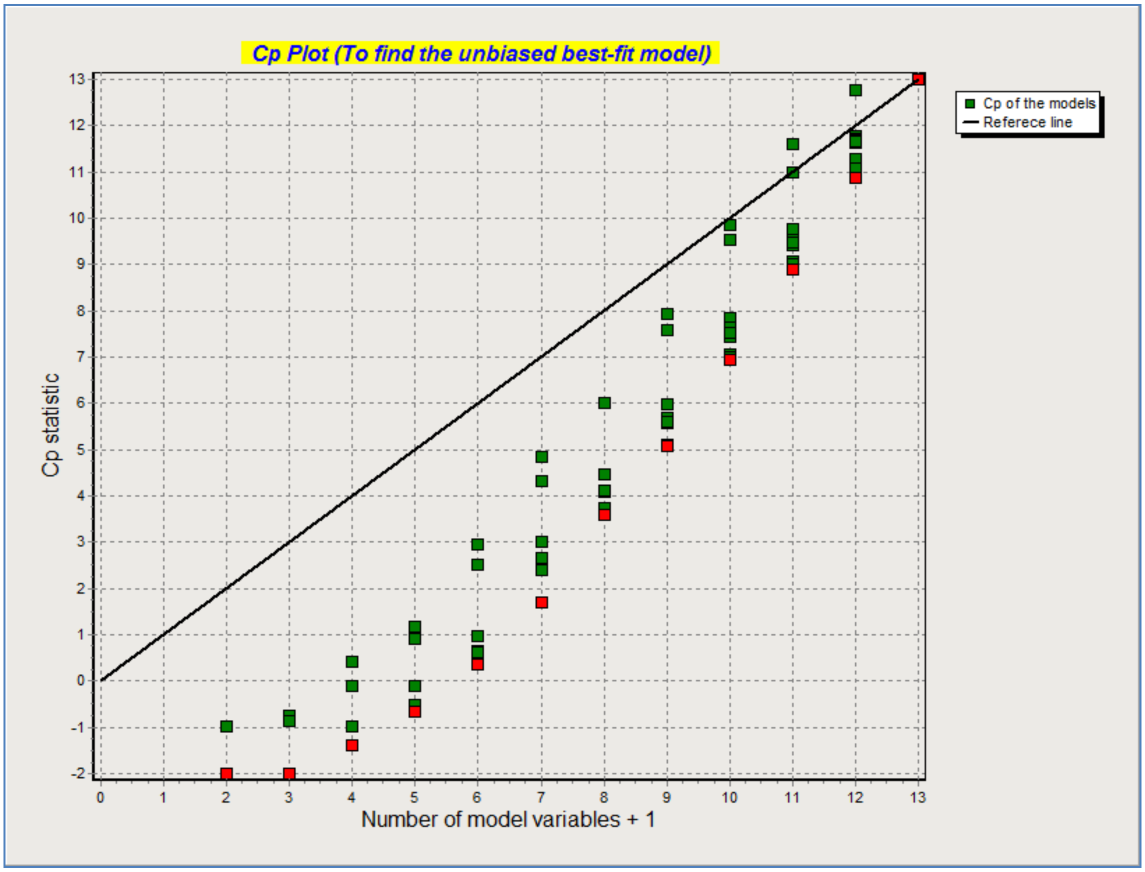

Figure 35: Murphy Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (-1.994) was considered best fit.

Table 69: Comparative models for Murphy Park Beach.

Murphy Park Beach			
Model	**2007**	**2008**	**2007-2008**
Variables	Turbidity, One day late gulls, and √Long shore current speed	Insignificant ($p > 0.05$)	Wave height
Number of samples	46	52	97
Correct predictions	46 (100%)	-	94 (96.9%)
Adjusted R^2	43%	-	10.1%
Mallow's Cp	4.0	-	-1.994
Residual standard error	1.35	-	1.529
p-value	<0.001	-	1.470e-^{05}
Predictive equation	*E. coli* [Log(MPN)] = [-0.31+ (0.28*trbdty) + (0.14*olagulls) + (0.35*√Lcs)]	-	*E. coli* [Log (MPN)] = 2.0264 + (0.4076*wh).

Otumba Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Otumba Park beach 2008 season were (3.84 ± 1.75), 4.24, and 6.86 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Otumba Park Beach are shown in Table 70. The explanatory variables that showed positive correlation with *E. coli* concentration for Otumba Park beach for 2008 were included in the modeling process. The variable that showed significant positive correlation was water temperature ($r = 0.35$, $p<0.05$). Only this is the variable included in the final predictive model, and are highlighted in Table 70. The scatter plot shown in Figure 36 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variable, water temperature in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (0 days) and predicted eight false negative $\log_e$ *E. coli* concentrations, as seen in Figure 36 and Table 71. The adjusted R-squared (R^2) of the mathematical model was 17.5% (Table 73) with lowest Cp was 2.245 (Table 73 and Figure 37) and the model was considered statistically significant ($p<0.05$), as seen in Table 73 using significant explanatory variable (Table 72). The predictive equation for Otumba Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] = [-0.8733 + (0.2334*wt)].

Table 70: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Otumba Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.01
Wind speed	0.11
Onshore wind	0.07
Water temperature	**0.35****
Air temperature	0.25
Turbidity	0.04
Wave height	-0.12
Algae	-0.49
One day late algae	-0.18
Gulls	0.08
One day late gulls	0.03
Two days late gulls	-0.12
Long shore current speed	0.04
Long shore current direction	-0.01
24 hour rainfall	0.09
48 hour rainfall	0.12
72 hour rainfall	0.11
Event rainfall	0.06
Weighted 72hr rainfall	0.18
Weighted 48hr rainfall	0.15
Combined 72hr rainfall	0.20
Barometric pressure	-0.05

**Variable considered statistically significant i.e. $p < 0.05$.

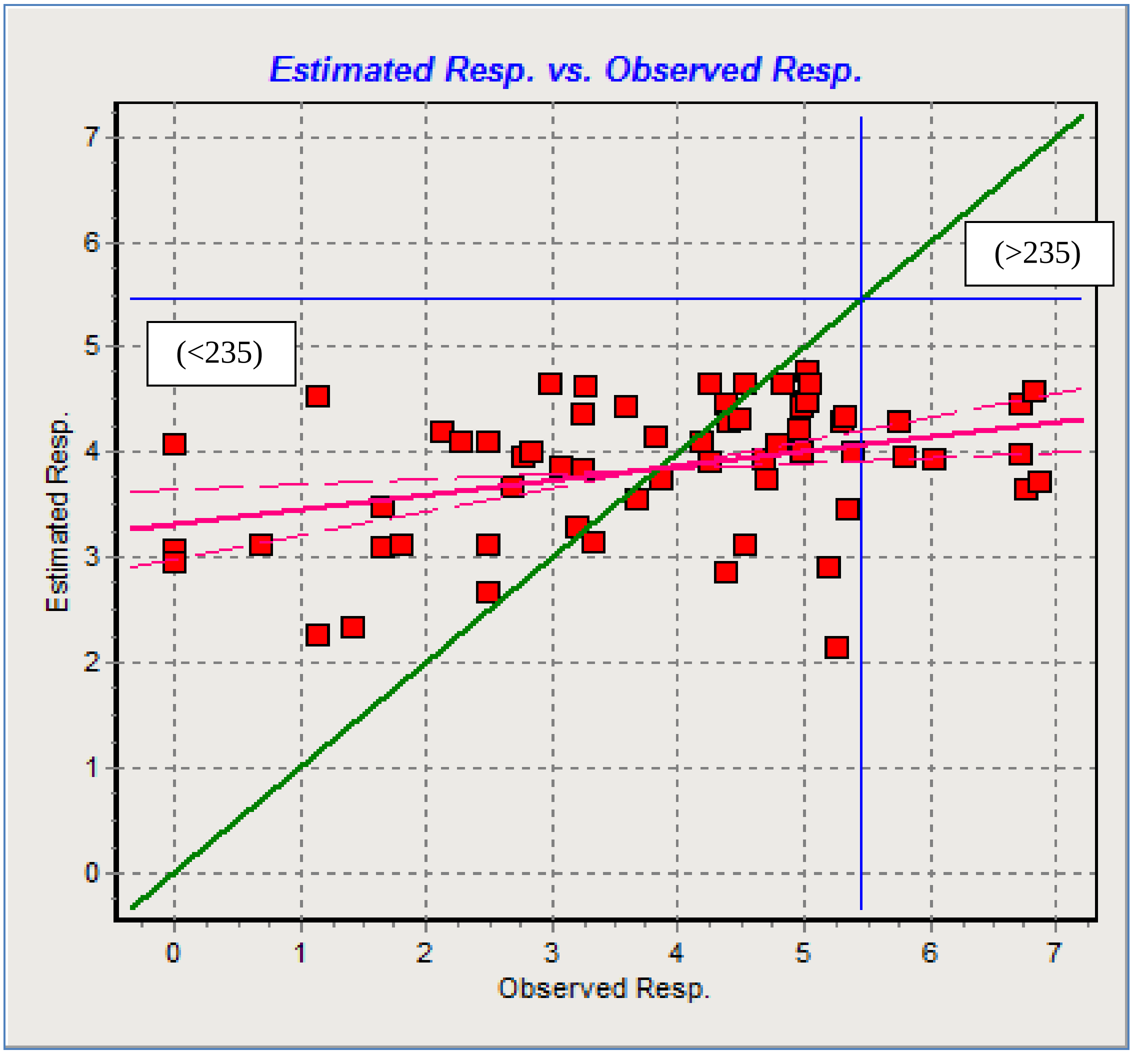

Figure 36: Otumba Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 36, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 17.5%. The green diagonal represents a perfect model for comparison.

Table 71: Otumba Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	61	100
Correct above standard	0	0%
Correct below standard	53	86.9%
False positives	0	0%
False negatives	8	13.1%

Table 72: Otumba Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-0.8773	1.2855	-0.6825	0.4979
Water temperature	0.2334	0.063	3.7051	0.0004709

Table 73: Otumba Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	9.583
Degrees of freedom (DF)	60
Residual standard error (RSE)	1.6003
Adjusted R^2	17.5%
Mallow's Cp	2.245
p-value	0.002999

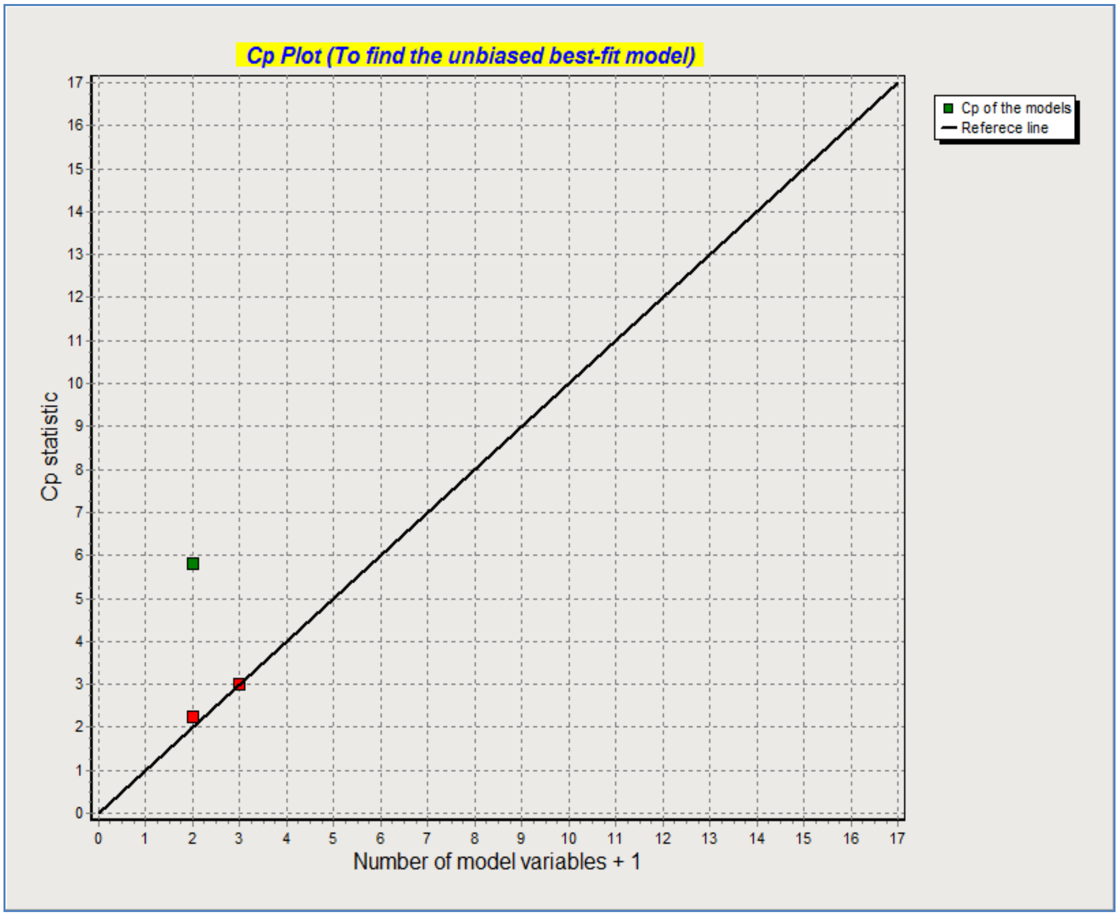

Figure 37: Otumba Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (2.245) was considered best fit.

Otumba Park Beach 2007 – 2008 combined season model.

The mean (MPN±SD), median, and maximum $\log_e$ *E. coli* concentrations for Otumba Park Beach 2007 – 2008 combined season were (3.31 ± 1.78), 3.24, and 6.86 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for the Otumba Park Beach for 2007 – 2008 combined season model, were included in the modeling process shown in Table 74. The variables that showed significant positive correlation were water temperature (r = 0.29, p<0.05), onshore wind (r = 0.26, p<0.05). Only these were included in the final predictive model, and are highlighted in Table 74. The scatter plot shown in Figure 38 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant explanatory variables (onshore wind and water temperature) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (0 days), as seen in Figure 38 and Table 75. The adjusted R-squared (R^2) of the mathematical model was 12.5% (Table 77) with lowest Cp was 3.0 (Table 77 and Figure 39) and the model was considered statistically significant (p<0.05), as seen in Table 77 using significant explanatory variables (Table 76). The predictive equation for Otumba Park Beach 2007 – 2008 combined model is as follows:

E. coli [Log (MPN)] **=** [0.08245 + (0.94583*onwind) + (0.14375*wt)].

Models were compared for Otumba Park Beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 78. Water temperature was the only variable conserved across all the three models (Table 78).

Table 74: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Otumba Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.11
Wind speed	-0.14
Onshore wind	**0.26****
Water temperature	**0.29****
Air temperature	-0.05
Turbidity	0.10
Wave height	0.11
Algae	-0.32
One day late algae	-0.25
Gulls	0.07
One day late gulls	0.02
Two days late gulls	-0.05
Long shore current speed	-0.22
Long shore current direction	-0.16
24 hour rainfall	-0.01
48 hour rainfall	0.04
72 hour rainfall	0.09
Event rainfall	0.16
Weighted 72hr rainfall	0.05
Weighted 48hr rainfall	0.09
Combined 72hr rainfall	0.07

**Variable considered statistically significant i.e. $p < 0.05$.

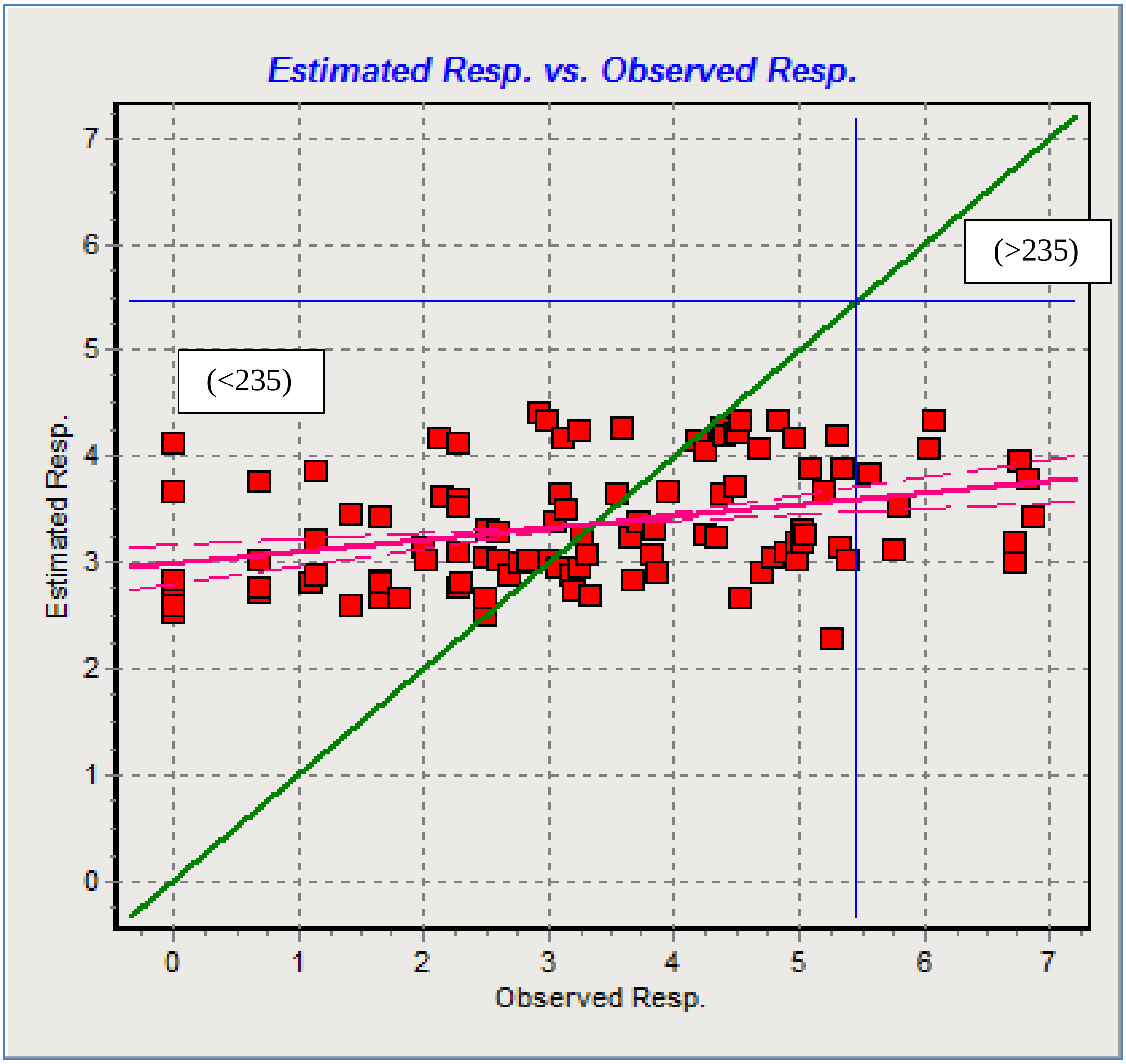

Figure 38: Otumba Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 38, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 12.5%. The green diagonal represents a perfect model for comparison.

Table 75: Otumba Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	106	100
Correct above standard	0	0%
Correct below standard	96	90.6%
False positives	0	0%
False negatives	10	9.4%

Table 76: Otumba Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	0.08245	0.97683	0.084	0.93290
Onshore wind	0.94583	0.38424	2.462	0.01549
Water temperature	0.14375	0.04835	2.973	0.00367

Table 77: Otumba Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	8.543
Degrees of freedom (DF)	103
Residual standard error (RSE)	1.672
Adjusted R^2	12.5%
Mallow's Cp	3.0
p-value	0.0003692

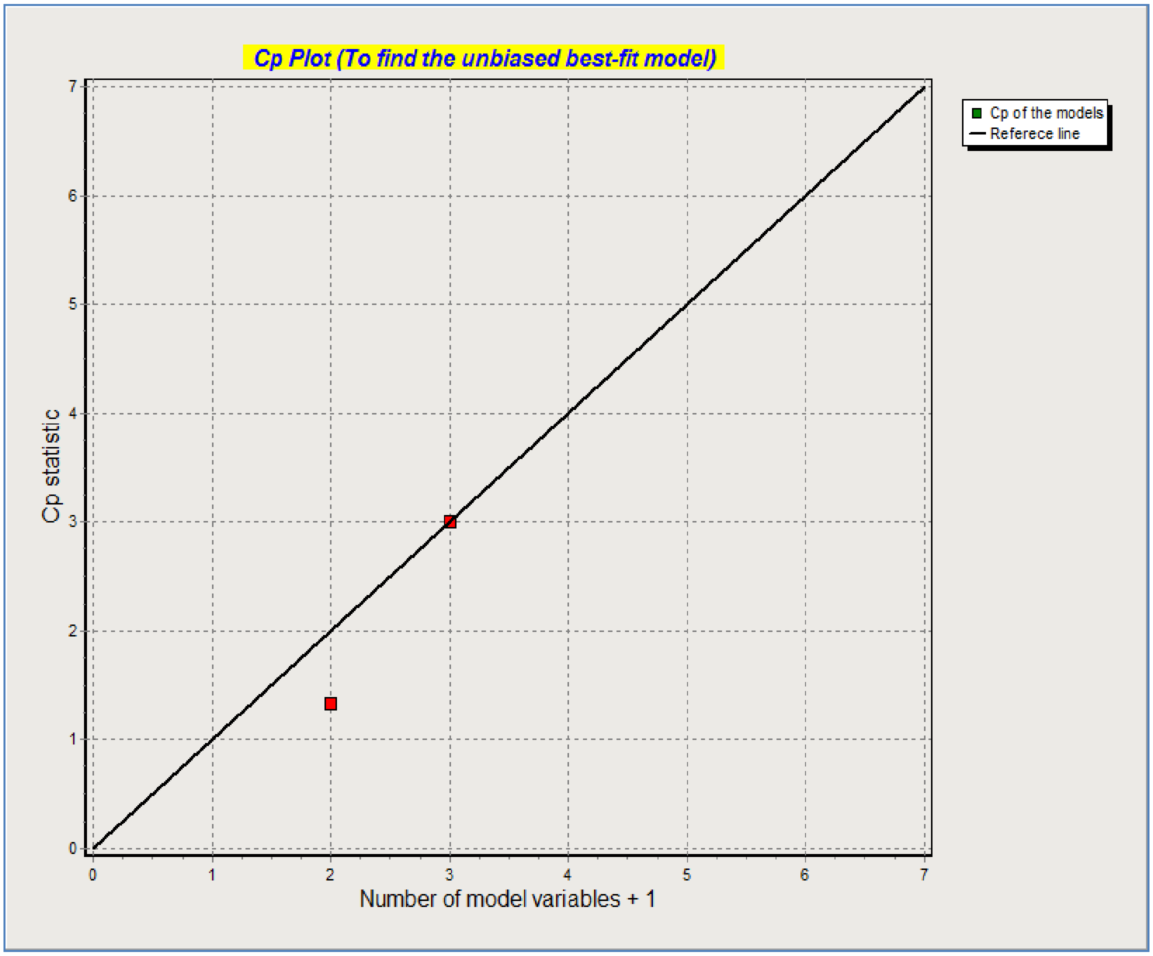

Figure 39: Otumba Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (3.0) was considered best fit.

Table 78: Comparative models for Otumba Park Beach.

Otumba Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Onshore wind, Water temperature, and Turbidity	Water temperature	Onshore wind and Water temperature
Number of samples	44	61	106
Correct predictions	43 (97.7%)	53 (86.9%)	96 (90.5%)
Adjusted R^2	48%	17.5%	12.5%
Mallow's Cp	4.0	2.245	3.0
Residual standard error	1.16	1.6003	1.672
p-value	0.001	0.002999	0.0003692
Predictive equation	*E. coli* [Log(MPN)] = [-1.03 + (2.15*onwind) + (0.09*wt) + (0.99*trbdty)]	*E. coli* [Log (MPN)] == [-0.8733 + (0.2334*wt)]	*E. coli* [Log (MPN)] = [0.08245 + (0.94583*onwind) + (0.14375*wt)]

Sunset Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Sunset Park beach 2008 season were (4.09 ± 1.31), 4.26, and 6.82 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Sunset Park beach are shown in Table 79. The explanatory variables that showed positive correlation with *E. coli* concentration for Sunset Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were water temperature (r = 0.58, p<0.05) and turbidity (r = 0.21, p<0.05). Only these were included in the final predictive model, and are highlighted in Table 79. The scatter plot shown in Figure 40 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (water temperature and turbidity) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (1 day) and four false negative $\log_e$ *E. coli* concentrations, as seen in Figure 40 and Table 80. The adjusted R-squared (R^2) of the mathematical model was 39.9% (Table 82) with lowest Cp was 3.0 (Table 82 and Figure 41) and the model variables were shown in Table 81 with their p-values. The predictive equation for the Sunset Park Beach–2008 model is as follows:

E. coli [Log (MPN)] **= [**-0.8802 + (0.2153*wt) + (0.4796*trbdty)].

Table 79: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for Sunset Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	0.12
Wind speed	0.18
Onshore wind	0.13
Water temperature	**0.58****
Air temperature	0.32
Turbidity	**0.21****
Wave height	0.29
Algae	-0.25
One day late algae	-0.29
Gulls	0.23
One day late gulls	0.22
Two days late gulls	0.30
Long shore current speed	0.24
Long shore current direction	0.01
24 hour rainfall	0.11
48 hour rainfall	-0.004
72 hour rainfall	0.002
Event rainfall	-0.003
Weighted 72hr rainfall	0.08
Weighted 48hr rainfall	0.09
Combined 72hr rainfall	-0.06
Barometric pressure	-0.02

**Variable considered statistically significant i.e. $p < 0.05$.

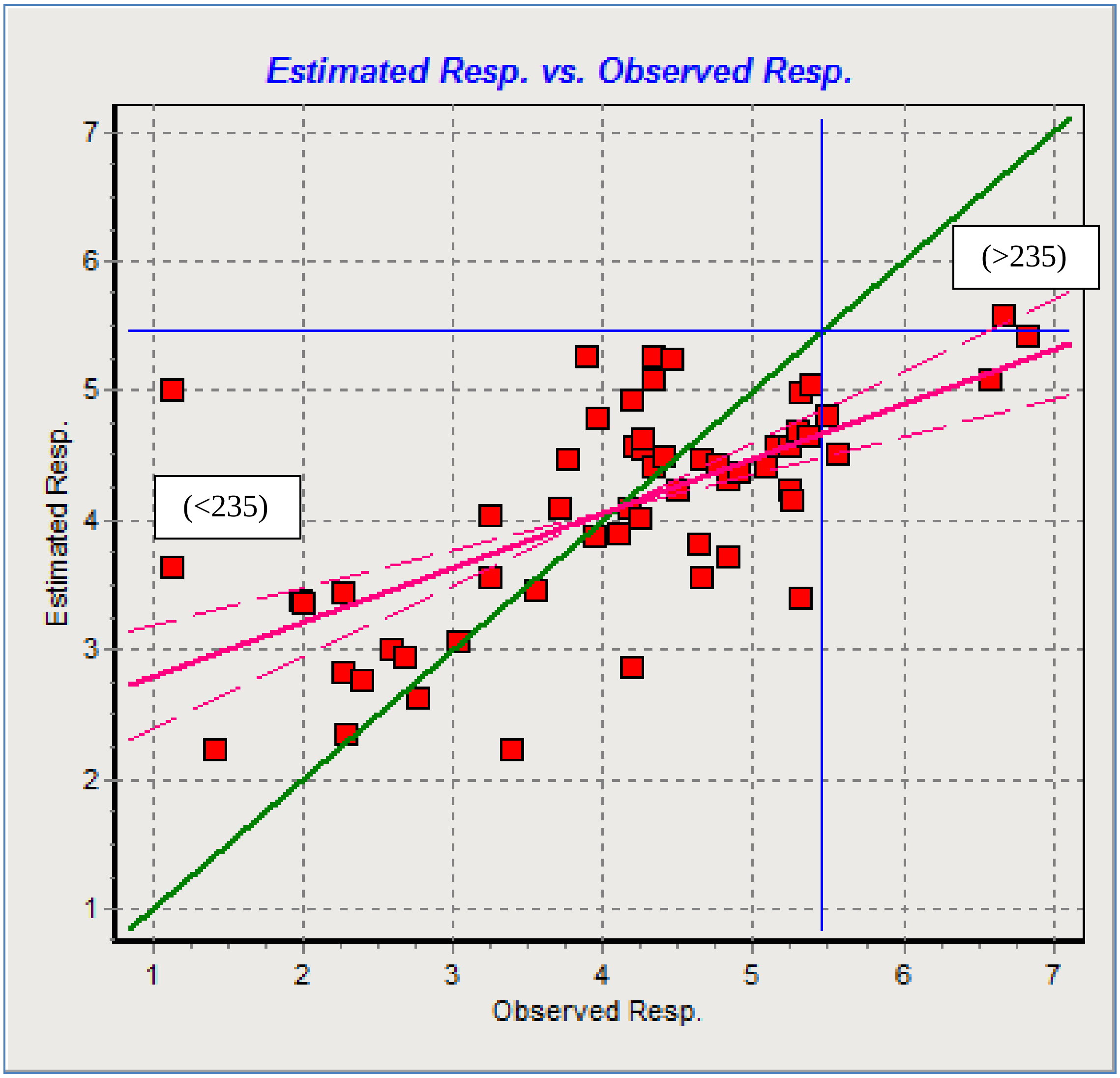

Figure 40: Sunset Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 40, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 39.9%. The green diagonal represents a perfect model for comparison.

Table 80: Sunset Park Beach – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	58	100
Correct above standard	1	1.72%
Correct below standard	53	91.4%
False positives	0	0 %
False negatives	4	6.9 %

Table 81: Sunset Park Beach – 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-0.8802	0.871	-1.0106	0.317
Water temperature	0.2153	0.04176	5.1565	3.5663E-6
Turbidity	0.4796	0.1325	3.6199	0.0006465

Table 82: Sunset Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	-
Degrees of freedom (DF)	-
Residual standard error (RSE)	1.017
Adjusted R^2	39.9%
Mallow's Cp	3.0
p-value	< 0.05

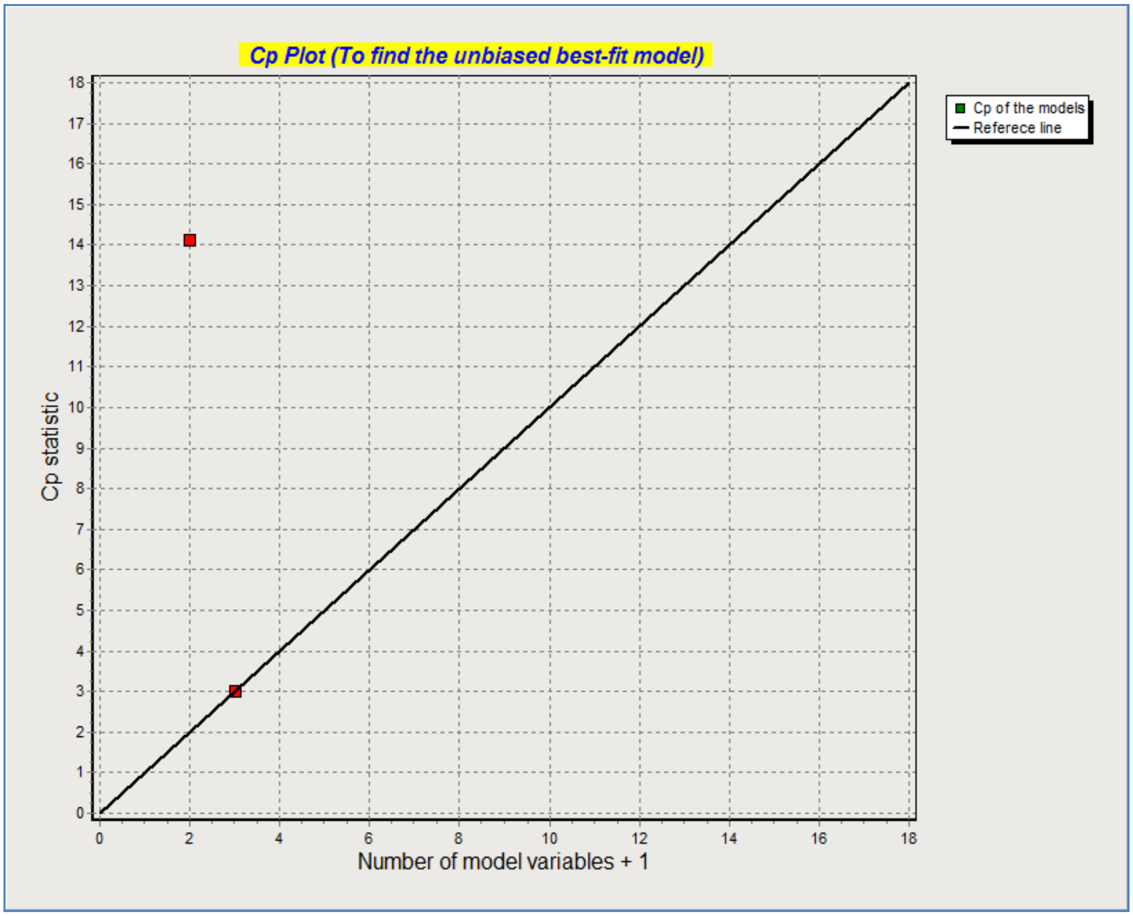

Figure 41: Sunset Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (3.0) was considered best fit.

Sunset Park Beach 2007 - 2008 combined season model.

The mean (MPN±SD), median, and maximum $\log_e$ *E. coli* concentrations for Sunset Park Beach 2007 – 2008 combined season were (3.60 ± 1.70), 3.96, and 6.82 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for Sunset Park beach for 2007 – 2008 combined season model, were included in the modeling process and are shown in Table 83. The variables that showed significant positive correlation were water temperature ($r = 0.55$, $p<0.05$), air temperature ($r = 0.11$, $p<0.05$), and algae ($r = 0.20$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 83. The scatter plot shown in Figure 42 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variables (water temperature, air temperature and algae) in the mathematical model. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (2 days), as seen in Figure 42 and Table 84. The adjusted R-squared (R^2) of the mathematical model was 33.9% (Table 86) with lowest Cp was – 1.372 (Table 86 and Figure 43) and the model was considered statistically significant ($p<0.05$), as seen in Table 86 using significant explanatory variables (Table 85). The predictive equation for Sunset Park 2007 – 2008 combined model is as follows: *E. coli* [Log (MPN)] = [-0.2033 + (0.3336*wt) - (0.1555*at) + (0.6705*algae)].

Models were compared for Sunset Park Beach for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 87.

Table 83: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Sunset Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	-0.03
Wind speed	-0.17
Onshore wind	0.22
Water temperature	**0.55****
Air temperature	**0.11****
Turbidity	-0.005
Wave height	0.35
Algae	**0.20****
One day late algae	0.05
Gulls	0.19
One day late gulls	0.09
Two days late gulls	0.10
Long shore current speed	-0.08
Long shore current direction	0.25
24 hour rainfall	0.11
48 hour rainfall	-0.03
72 hour rainfall	0.03
Event rainfall	0.09
Weighted 72hr rainfall	0.09
Weighted 48hr rainfall	0.09
Combined 72hr rainfall	0.07

**Variable considered statistically significant i.e. $p < 0.05$.

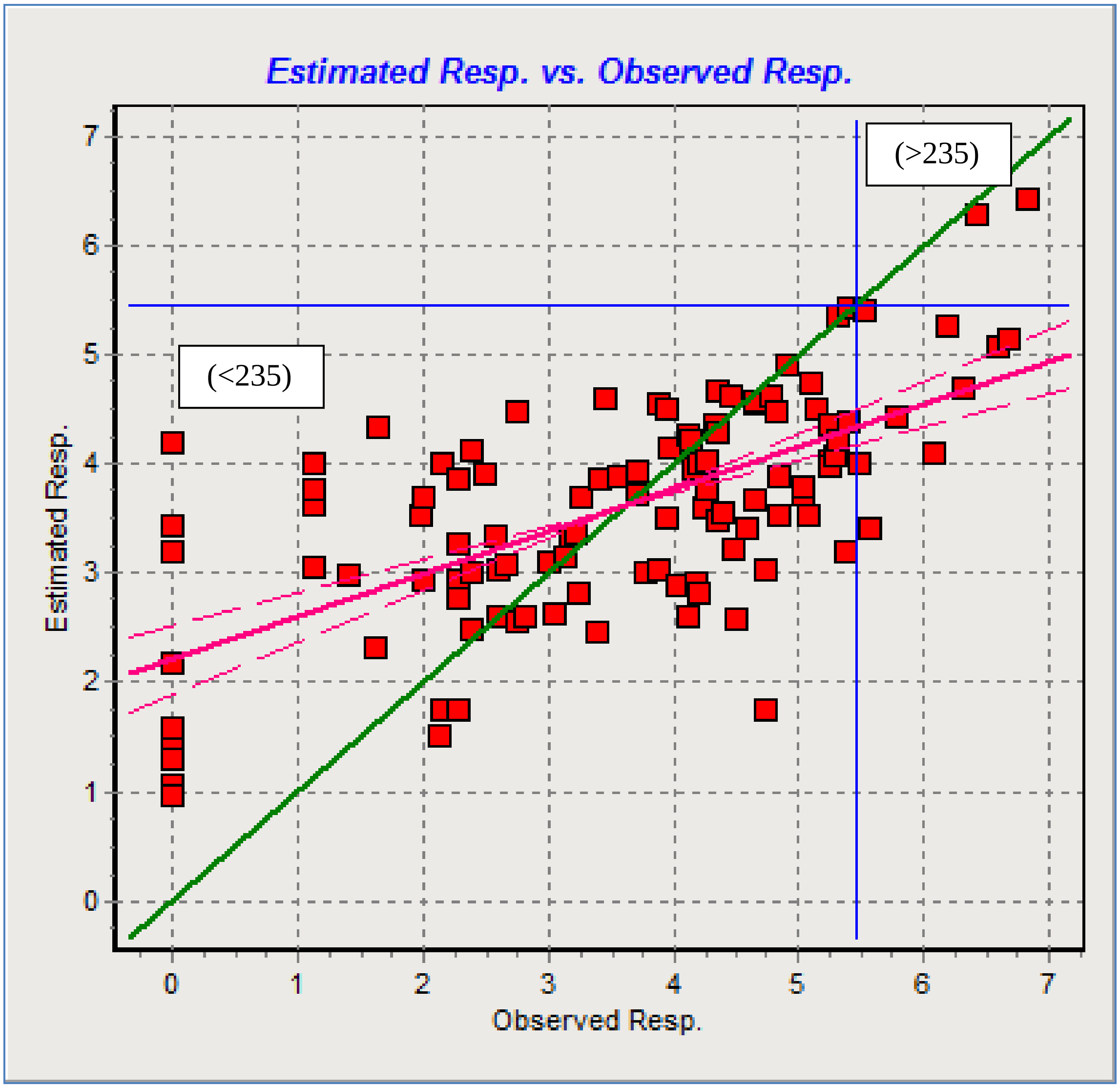

Figure 42: Sunset Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 42, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 33.9%. The green diagonal represents a perfect model for comparison.

Table 84: Sunset Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	**Number**	**Percentage**
Total predictions	113	100
Correct above standard	2	1.77%
Correct below standard	102	90.3%
False positives	0	0 %
False negatives	9	7.96%

Table 85: Sunset Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	**Parameter estimate**	**Standard error**	**t-value**	**p-value**
Intercept	-0.2033	0.6249	-0.2109	0.8335
Water temperature	0.3336	0.0477	6.9948	2.3548E-10
Air temperature	-0.1555	0.04022	-3.8672	0.0001901
Algae	0.6705	0.2249	2.9816	0.003565

Table 86: Sunset Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	**Value**
F-statistic	26.79
Degrees of freedom (DF)	109
Residual standard error (RSE)	1.229
Adjusted R^2	33.9%
Mallow’s Cp	-1.372
p-value	$4.678e^{-13}$

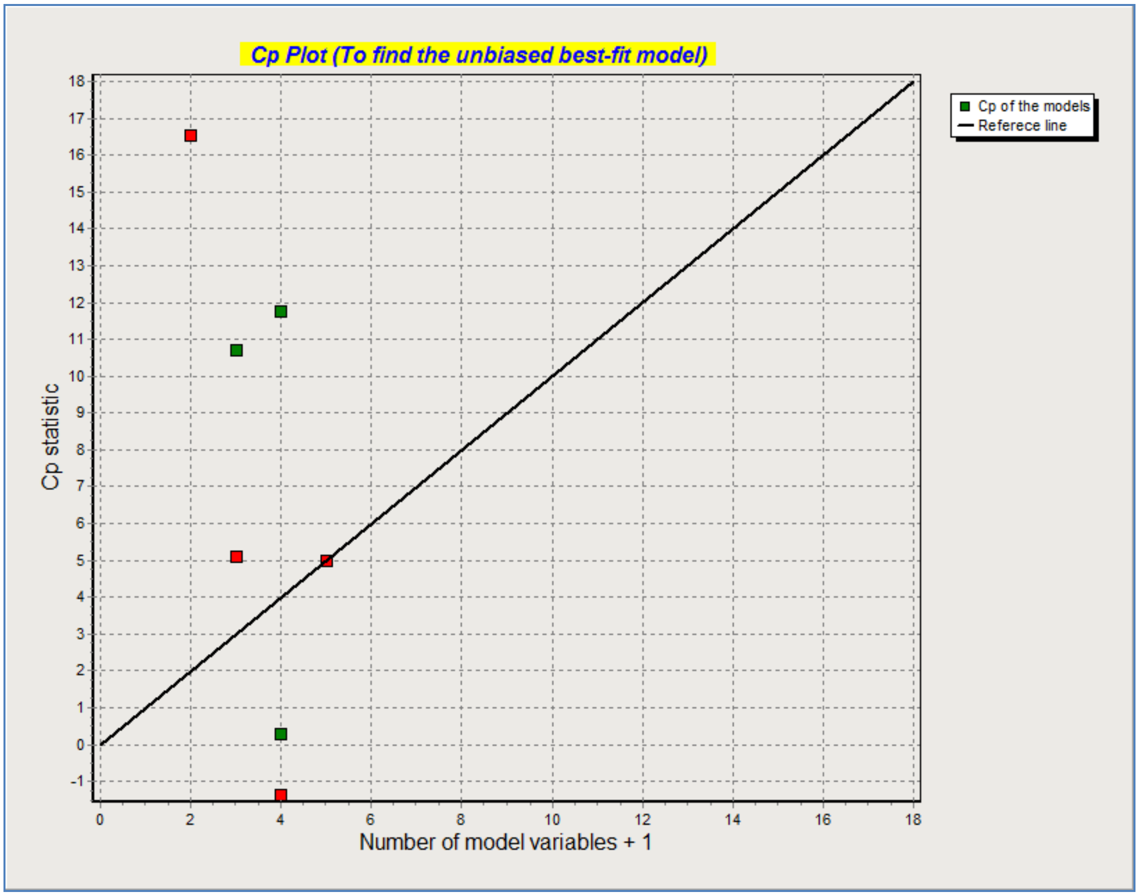

Figure 43: Sunset Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (-1.372) was considered best fit.

Table 87: Comparative models for Sunset Park Beach.

Sunset Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Onshore wind and Algae	Water temperature, and Turbidity	Water temperature, Air temperature and Algae
Number of samples	54	58	114
Correct predictions	48 (88.8%)	54 (93.1%)	104 (91.2%)
Adjusted R^2	47.2%	39.9%	33.9%
Mallow's Cp	3.0	3.0	-1.372
Residual standard error	1.273	1.017	1.229
p-value	<0.05	<0.05	$4.678e^{-13}$
Predictive equation	*E. coli* [Log(MPN)] = [0.8491+ (1.6788*onwind) + (1.4084*algae)]	*E. coli* [Log (MPN)] **=** **[**-0.8802 + (0.2153*wt) + (0.4796*trbdty)]	*E. coli* [Log (MPN)] = -0.2033 + (0.3336*wt) **-** (0.1555*at) + (0.6705*algae)]

Whitefish Dunes State Park Beach – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Whitefish Dunes State Park beach 2008 season were (3.51 ± 2.25), 1.72, and 6.62 respectively. Correlations between *Escherichia coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season for Whitefish Dunes State Park beach are shown in Table 88. The explanatory variables that showed positive correlation with *E. coli* concentration for Whitefish Dunes State Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were water temperature ($r = 0.49$, $p<0.05$), wave height ($r = 0.44$, $p<0.05$), and gulls ($r = 0.59$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table 88. The scatter plot shown in Figure 44 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using significant variables (water temperature, wave height, and gulls) in the mathematical model. The mathematical model predicted four false negative $\log_e$ *E. coli* concentrations, as seen in Figure 44 and Table 89. The adjusted R-squared (R^2) of the mathematical model was 52.7% (Table 91) with lowest Cp was 4.0 (Table 91 and Figure 45) and the model was considered statistically significant ($p<0.05$), as seen in Table 91 using significant explanatory variables (Table 90). The predictive equation for Whitefish Dunes State Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] **= [-** 0.7509 + (0.107*wt) + (0.108*wh) + (0.01245*gulls)].

Table 88: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Whitefish Dunes State Park Beach for 2008.

Variable	**'r' value**
Wind direction (degrees)	-0.14
Wind speed	0.27
Onshore wind	-0.03
Water temperature	**0.49****
Air temperature	0.21
Turbidity	0.09
Wave height	**0.44****
Algae	-0.10
One day late algae	0.20
Gulls	**0.59****
One day late gulls	0.24
Two days late gulls	0.30
Long shore current speed	-0.13
Long shore current direction	0.03
24 hour rainfall	0.14
48 hour rainfall	-0.06
72 hour rainfall	-0.13
Event rainfall	0.25
Weighted 72hr rainfall	0.21
Weighted 48hr rainfall	0.24
Combined 72hr rainfall	0.14
Barometric pressure	-0.08

**Variable considered statistically significant i.e. $p < 0.05$.

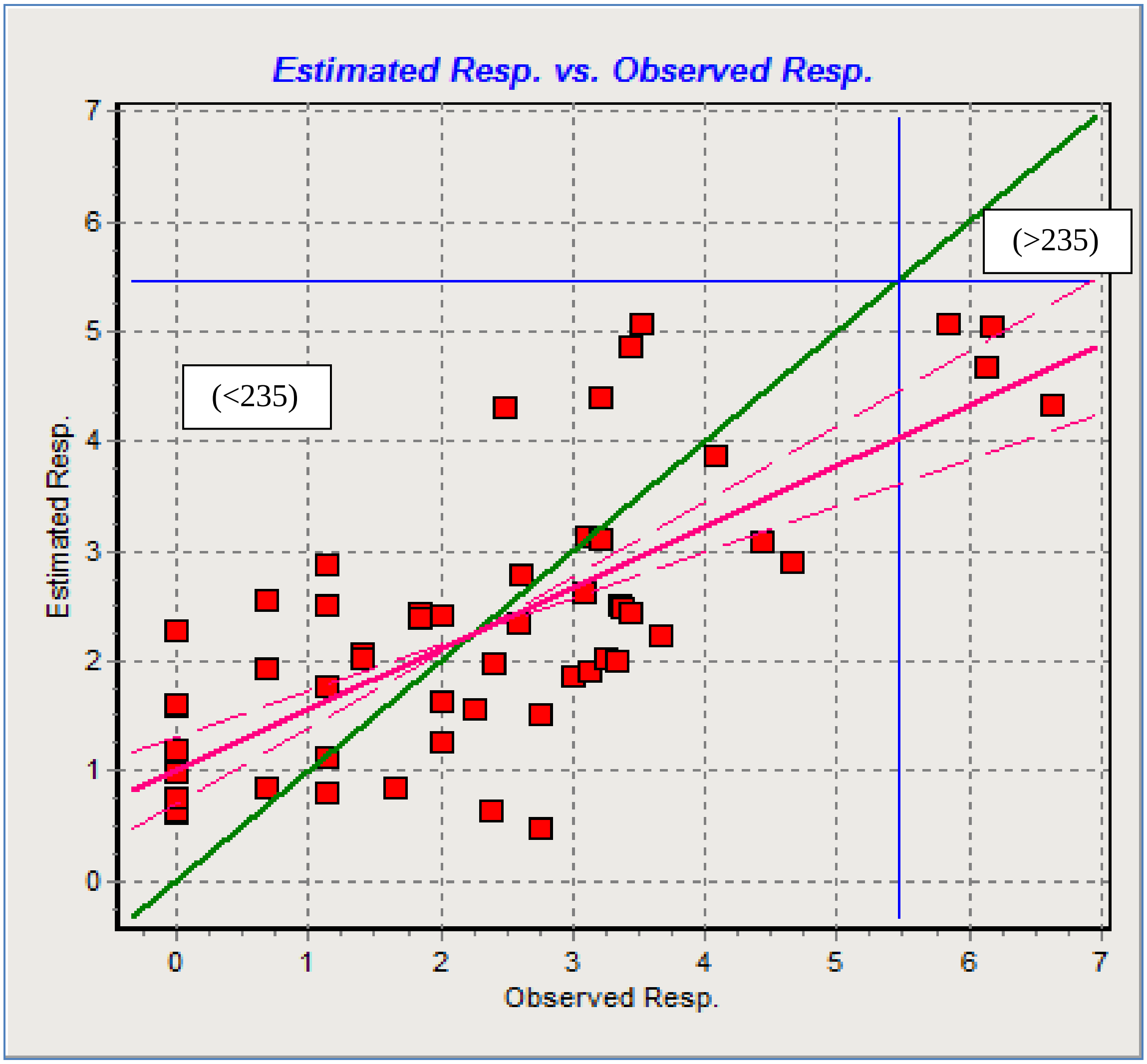

Figure 44: Whitefish Dunes State Park Beach – 2008 model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 44, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 52.7%. The green diagonal represents a perfect model for comparison.

Table 89: Whitefish Dunes State Park Beach 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	56	100
Correct above standard	0	0 %
Correct below standard	52	92.9%
False positives	0	0 %
False negatives	4	7.14%

Table 90: Whitefish Dunes State Park Beach 2008 model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-0.7509	0.5589	-1.3435	0.1853
Water temperature	0.107	0.03459	3.0929	0.003202
Wave height	0.108	0.02894	3.7335	0.0004724
Gulls	0.01245	0.003075	4.0504	0.0001722

Table 91: Whitefish Dunes State Park – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	21.4
Degrees of freedom (DF)	52
Residual standard error (RSE)	1.189
Adjusted R^2	52.7%
Mallow’s Cp	4.0
p-value	$3.662e^{-09}$

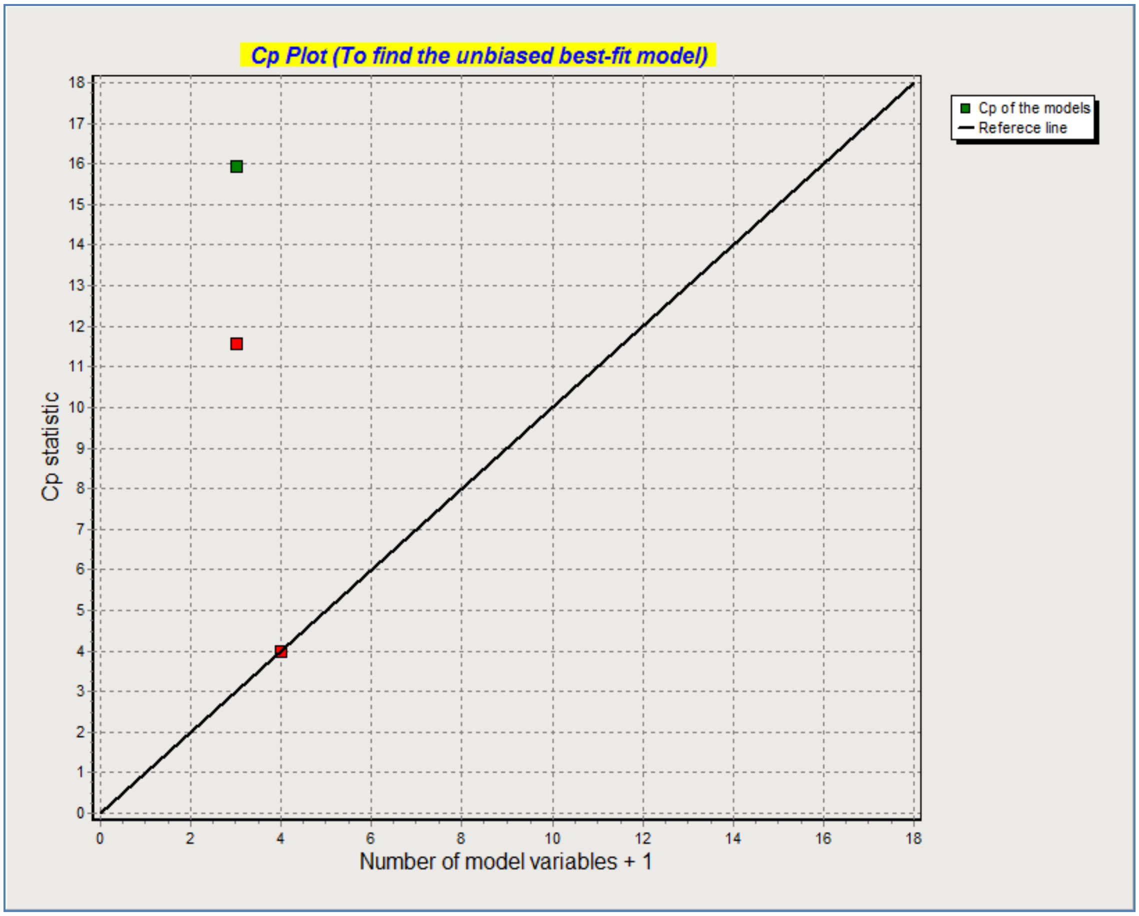

Figure 45: Whitefish Dunes State Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (4.0) was considered best fit.

Whitefish Dunes State Park Beach 2007 - 2008 combined season model.

The mean (MPN±SD), median, and maximum $\log_e$ *E. coli* concentrations for Whitefish Dunes State Park (WDSP) Beach 2007 – 2008 combined season were (2.12 ± 1.77), 2.00, and 7.25 respectively. The explanatory variables that showed positive correlation with *E. coli* concentrations for WDSP beach for the 2007 – 2008 combined season were included in the modeling process and are shown in Table 92. The variables that showed significant positive correlation were water temperature (r = 0.41, p<0.05), wave height (r = 0.32, p<0.05), gulls (r = 0.54, p<0.05), and long shore current speed (r = 0.21, p<0.05). Only these were included in the final predictive model, and are highlighted in Table 92. The scatter plot shown in Figure 46 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variables (water temperature, wave height, gulls and long shore current speed). The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (1 day), as seen in Figure 46 and Table 93. The adjusted R-squared (R^2) of the mathematical model was 43.4% (Table 95) with lowest Cp was 5.0 (Table 95 and Figure 47) and the model was considered statistically significant (p<0.05), as seen in Table 95 using significant explanatory variables (Table 94). The predictive equation for 2007 – 2008 combined model is as follows: *E. coli* [Log (MPN)] = [-0.6918 + (0.09462*wt) + (0.101*wh) + (0.01297*gulls) + (0.04085*lcs)]. Models were compared for each data set – 2007 only, 2008 only, and combined 2007 – 2008 presented in Table 96. Water temperature and wave height were the variables conserved across all the three models (Table 96).

Table 92: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for the Whitefish Dunes State Park Beach for 2007 – 2008 combined season.

Variable	**'r' value**
Wind direction (degrees)	0.02
Wind speed	0.19
Onshore wind	0.05
Water temperature	**0.41****
Air temperature	0.20
Turbidity	0.23
Wave height	**0.31****
Algae	-0.06
One day late algae	0.07
Gulls	**0.54****
One day late gulls	0.09
Two days late gulls	0.01
Long shore current speed	**0.21****
Long shore current direction	0.07
24 hour rainfall	0.06
48 hour rainfall	-0.13
72 hour rainfall	-0.07
Event rainfall	0.18
Weighted 72hr rainfall	-0.06
Weighted 48hr rainfall	-0.03
Combined 72hr rainfall	-0.09

**Variable considered statistically significant i.e. $p < 0.05$.

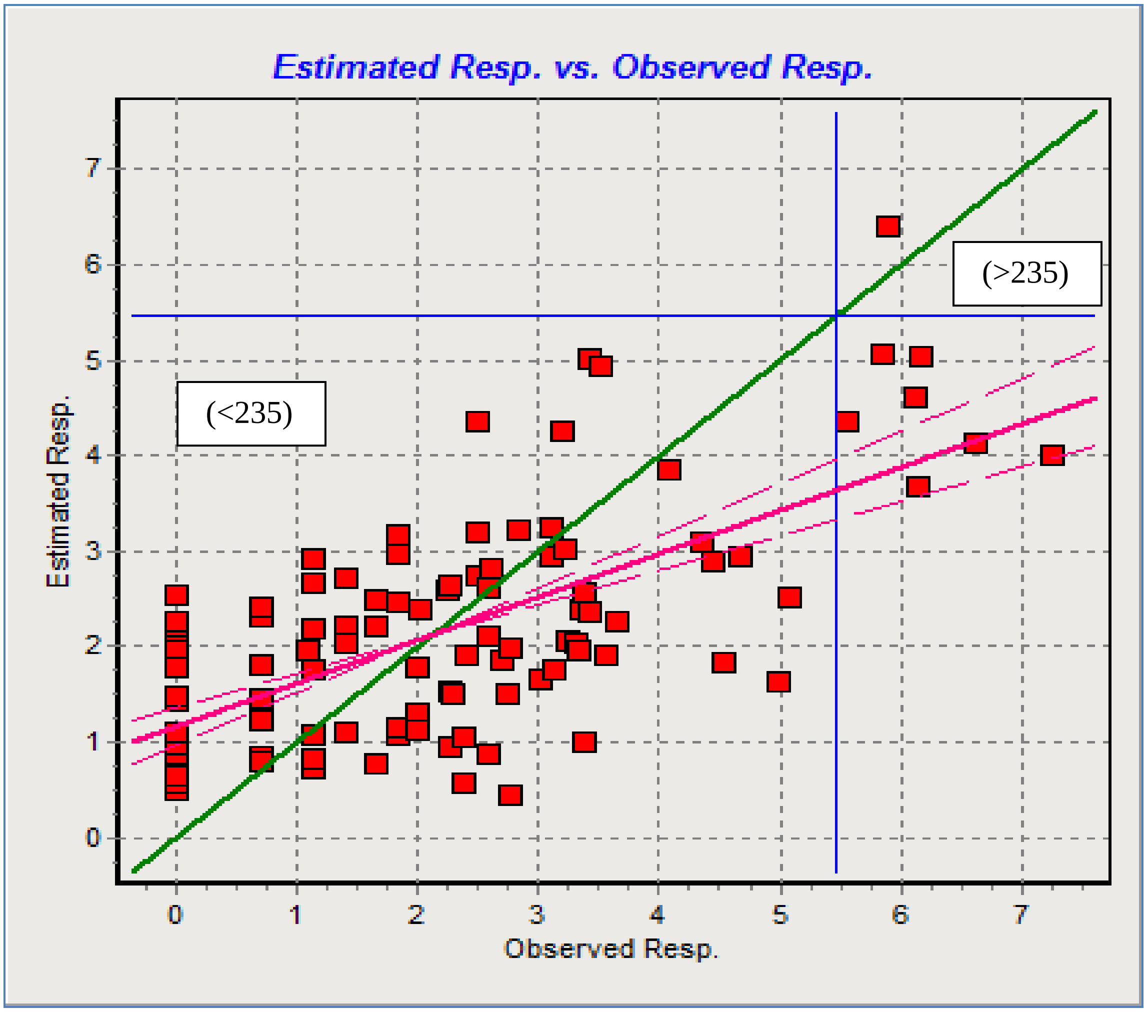

Figure 46: Whitefish Dunes State Park Beach 2007 – 2008 combined season model scatter plot shows the relationship between the estimated and observed *E. coli* concentrations.

In the Figure 46, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235 CFU/ 100mL water. The purple line shows the “best fit” line of the predicted values with a 95% confidence interval (dashed) around it with an adjusted R^2 of 43.4%. The green diagonal represents a perfect model for comparison.

Table 93: Whitefish Dunes State Park Beach 2007 – 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	110	100
Correct above standard	1	0.91%
Correct below standard	102	92.7%
False positives	0	0 %
False negatives	7	6.36%

Table 94: Whitefish Dunes State Park Beach 2007 – 2008 combined season model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	-0.6918	0.4532	-1.5266	0.1302
Water temperature	0.09462	0.02816	3.36	0.001095
Wave height	0.101	0.0256	3.9469	0.0001446
Gulls	0.01297	0.002405	5.3933	4.3467E-7
Long shore current speed	0.04085	0.01752	2.3325	0.02167

Table 95: Whitefish Dunes State Park Beach 2007 – 2008 combined season model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	21.89
Degrees of freedom (DF)	105
Residual standard error (RSE)	1.334
Adjusted R^2	43.4%
Mallow's Cp	5.0
p-value	$3.699e^{-13}$

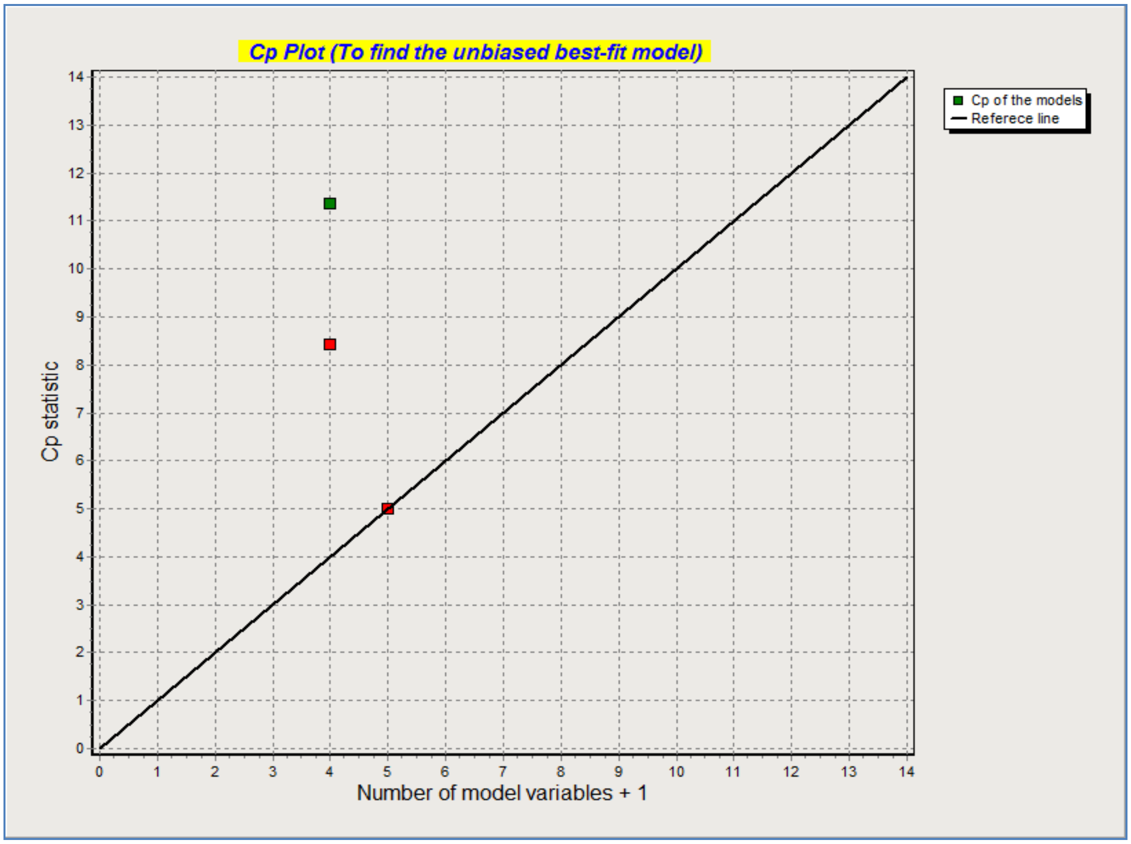

Figure 47: Whitefish Dunes State Park Beach 2007 – 2008 Cp-Plot of the combined mathematical model. The model with lowest Cp (5.0) was considered best fit.

Table 96: Comparative models for Whitefish Dunes State Park Beach.

Whitefish Dunes State Park Beach			
Model	**2007**	**2008**	**2007 – 2008**
Variables	Water temperature and wave height	Water temperature, Wave height, and Gulls	Water temperature, Wave height, gulls, and Long shore current speed
Number of samples	54	56	110
Correct predictions	52 (96.2%)	52 (92.8%)	103 (93.6%)
Adjusted R^2	38.9%	52.7%	43.4%
Mallow's Cp	3.0	4.0	5.0
Residual standard error	1.426	1.189	1.334
p-value	<0.001	$3.662e^{-09}$	$3.699e^{-13}$
Predictive equation	*E. coli* [Log(MPN)] = [-1.496 + (0.1386*wt) + (1.5363*wh)]	*E. coli* [Log (MPN)] **=** **[-** 0.7509 + (0.107*wt) + (0.108*wh) + (0.01245*gulls)]	*E. coli* [Log (MPN)] = [-0.6918 + (0.09462*wt) + (0.101*wh) + (0.01297*gulls) + (0.04085*lcs)]

DISCUSSION

Virtual Beach – Model Builder (VB-MB) was user friendly software for developing mathematical models for 2008 and combined 2007/2008 recreational seasons to predict *E. coli* concentrations in nearshore water. It has several components, including the multiple linear regressions (MLR) tool, for developing statistical models. The software automatically detects and excludes influential outliers that would likely to skew the model. The R^2 and Cp-statistic, as joint criteria, are used for the model selection process. VB facilitates another option allowing model results to be exported to MS Excel for further evaluation (Zhongfu & Frick, 2007).

MLR models can be fit to a variety of data sets ranging from a few days to a season or more (Frick, Ge, & Zepp, 2008). A minimum of five observations per variable are necessary to develop a model. Ideally, twenty observations per variable (with an overall *N* of 100) are required to develop a model (Francis, 2007). Periods as short as 10 days have been used to fit models at marine beaches (Hou, Rabinovici, & Boehm, 2006). Some researchers have maintained that MLR models should be based on long-term data sets, because as the dataset grows, the predictive capacity of the model can be improved (Nevers & Whitman, 2005; Francy & Darner, 2006). The adjusted R^2 values for Huntington Beach 2006 season, improved predictive ability with more data points, for example, are 50.0, 45.7, 61.0, 53.0, and 60.7% for models generated on 21, 28, 35, 42, and 49 days, respectively (Frick et al., 2008). These models for determining recreational water quality could aid beach managers in more rapidly determining when waters are not safe for recreational use. The models, however, should be tested with real time *E. coli* concentrations.

E. coli concentrations were monitored and mathematical models were developed for eight beaches selected from Door County (WI) along the Lake Michigan shoreline, and for three beaches selected from Ashland and Bayfield counties (WI) along the Lake Superior shoreline for recreational season of 2008. Duel year predictive mathematical models also were developed for these selected beaches using 2007/2008 combined years data.

In Ashland County, the Kreher Park Beach exceeded the bathing water quality standard once during the 2008 recreational season. The explanatory variables included in the Kreher Park Beach mathematical model were wave height, algae, and event rainfall (Tables 4), which explained 43.7% $\log_e$ *E. coli* variation. The mathematical model predicted $\log_e$ *E. coli* concentrations 100% accurately, without false positives or false negatives (Figure 6 and Table 3). The storm event rainfall and wave height were the important predictive variables for the single and duel year models (Table 10). Square root of event rainfall and weighted 72 hour rainfall were important variables for the 2007 model (Table 10).

The Maslowski Park Beach (Ashland County) exceeded the bathing water quality standard twice during the 2008 recreational season. The explanatory variables included in the Maslowski Park Beach mathematical model were wave height, and weighted 48 hour rainfall (Table 13), which explained 40.8% $\log_e$ *E. coli* variation with 83.3% accuracy. The weighted 48 hour rainfall was significantly correlated with *E. coli* concentrations in the 2007, 2008, and 2007/2008 combined models for Maslowski Beach (Table 19).

Overall, the inclusion of at least one rainfall variable was conserved across the years 2007 – 2008 at Ashland County beaches. These rainfall variables were positively correlated with $\log_e$ *E. coli* concentrations, because of the assumption that *E. coli* concentrations were observed to be higher after rainfall. The elevated *E. coli* concentrations are due to stormwater outfalls discharge, surface runoff, and pet waste discharged into the water. Stormwater runoff from heavy rainfalls has been associated with increasing microbial loads in beach waters and coastal areas in Wisconsin (Sampson, Swiatnicki, McDermott, & Kleinheniz, 2006). Following heavy rainfall events, the county health departments warn the public to avoid contact with beach water in Southern California for as much as three days after such rainfalls to protect human health (Ackerman & Weisberg, 2003). The long-term persistence and leaching of *E. coli* in temperate maritime soils provide a steady source of the fecal indicator after rainfalls that can impact water quality (Brennan, O'Flaherty, Kramers, Grant, & Richards, 2010).

In Bayfield County, the Thompson's West End Park Beach's 2008 model used water temperature, one day late algae and gulls as variables to predict *E. coli* concentrations explaining 71.5% $\log_e$ *E. coli* variation with 95.8% accuracy (Table 22).

However, this model predicted one false negative event (Figure 14 and Table 21). False negative responses are more problematic than false positive responses because the model would indicate safe recreational water instead of an advisory when *E. coli* concentrations exceeded the water quality standards. Under these conditions the public is potentially exposed to water-borne pathogens that would impact human health. However, these false negative and false positive predictions were minimized by altering and refitting the model and could be minimized by adding additional variable data. The algal (*Cladophora)* mats harbor *E. coli* in greater concentrations than found in the surrounding waters (Vanden Heuvel et al., 2009). It has been observed that the *E. coli* present in algal mats may take time to wash from the algal mats into water and changing the water quality. Number of gulls was positively correlated with $\log_e$ *E. coli* concentrations (Table 20). Other researchers have noted that higher concentrations of indicator bacteria, such as *E. coli* and enterococci, are associated with gull feces (Fogarty et al., 2003). Fecal material from gulls on the shore likely was in part responsible, for elevated *E. coli* concentrations in beach water. The lag between increased gull number and elevated *E. coli* concentrations in beach water can be explained by the transport time of waste being washed into beach water via wave action and rainfall (Ackerman & Weisberg, 2003).

In Door County, water temperature was included as a significant model variable for all the beaches, except Ephraim and Murphy Park beaches, for the 2008 recreational season. Water temperature also was included in the 2007/2008 combined models for four of eight beaches. Water temperature was positively correlated with $\log_e$ *E. coli* concentrations at these beaches, which means *E. coli* concentrations increased as water temperature increased. This may not seem reasonable because several studies have reported that cooler water temperatures can increase the ability of *E. coli* to survive in the aquatic environment (Bogosian et al., 1996; Brettar & Hofle, 1992; Sampson et al., 2005; & Sampson et al., 2006). On the other hand, some studies have supported our findings that water temperature can be positively correlated with *E. coli* concentrations (Francy et al., 2006). At two Lake Erie beaches, Villa Angela (2004-2005) and Lakeshore (2004-2005) in Ohio, *E. coli* concentrations increased as water temperature increased (Francy et al., 2006). Therefore, water temperature plays an important role in determining *E. coli* survival/persistence in the recreational water quality, but temperature alone is not be capable of predicting *E. coli* concentrations. This evidence indicates that beach models should be individually developed for individual beaches to take into account unique features of difference bodies of water.

Turbidity was another model variable found in common for three Door County beaches (Ephraim, Fish Creek, and Sunset Park), which was significantly and positively correlated with $\log_e$ *E. coli* concentrations. This indicated an increase in turbidity corresponded to an increase in *E. coli* concentrations. Three factors (UV radiation, sediment resuspension and transport) can be used to explain this association. First, as turbidity increases, penetration of ultraviolet radiation into the water column decreases, and *E. coli* survival should increase in beach water (Whitman, Nevers, Korineck, & Byappanahalli, 2004). This might result in elevated *E. coli* concentrations in beach waters. Storm generated outflows are associated with increase in suspended solids, and thus higher water turbidity (Olyphant et al., 2003). Storm events resuspend and transport bacteria-laden sediments into lakes that could contain elevated concentrations of fecal bacteria (Mueller-Spitz, Stewart, Klump, & McLellan, 2010; Olyphant & Whitman, 2004). Other researchers have found that turbidity could be one of the predictive variables for Lake Erie beaches in Ohio (Francy et al., 2006) and Presque Isle Beach 2 in Pennsylvania (Zimmerman, 2006).

A unique set of significant explanatory variables was included in the mathematical models for Lake Michigan and Lake Superior beaches. The results of this study indicate that the individual beach models could not be extended between lakes or beaches with close proximity to another. The models also changed annually due to changes in physical parameters measured (Francy et al., 2006; Zimmerman, 2006). The same thing occurred for the combined year models with different variables being used to predict *E. coli* concentrations.

Other researchers have shown that additional years of data (e.g. three or more years) could be used for developing a more robust predictive model to improve predictions (Francy et al., 2006). The Huntington Beach 2000-2004 model yielded 84.7% correct predictions with an adjusted R^2 of 0.38, whereas the model developed for 2000-2005 seasons, predicted 85.9% correct predictions with an adjusted R^2 of 0.42, which was an improvement over 200-2004 model (Francy & Darner, 2006). In this study, combining data from two consequent beach seasons did not dramatically improve the fitness of the model for most of the Wisconsin beaches. This can be explained by the vastly different weather conditions (e.g. temperature, turbidity, and rainfall totals) that occurred during the two beach seasons. Therefore, three or more years of data for model development may yield better results than two years, because a larger data set would provide a wider range of environmental and water quality conditions to improve the fitness of the model. Therefore, further studies could focus on adding more physical and biological parameters to improve the predictive ability of the models.

Some variables (e.g. wave height, turbidity, and algae) used in our models were not discrete measurements, but rather used a comparative scale to indicate the magnitude of the variable. For example, field turbidity measurements were recorded as clear, slightly turbid, turbid, and opaque. These descriptors were then assigned a numerical value (clear = 0, slightly turbid = 1, turbid = 2, and opaque = 3). In the future, more accurately measured variables, could be added to the models to improve predictions. Other variables used in other studies such as lake stage (Francy et al., 2006; Olyphant & Whitman, 2004), insolation (incoming solar radiation) (Olyphant, 2004; Olyphant & Whitman, 2004), and day of the year (Francy et al., 2006) could be included our modeling process to improve predictions and to better protect public health.

Overall outcomes derived from the work were,

1. Mathematical models were developed for all the beaches, except Murphy Park Beach (variables were not significant) for the recreational season 2008.

2. The Kreher Park Beach 2008 mathematical model was the only one which predicted *E. coli* concentrations in beach water 100% accurately, without false positives or false negatives (Figure 6).
3. Single variable predictive models were developed for Ephraim, Lakeside, and Otumba Park beaches for the 2008 recreational season.
4. The combination of years of data (2007 & 2008) usually did not improve the fitness of the mathematical models (decreased adjusted R^2).
5. Explanatory variables that were included in the predictive models were unique for each beach and often for each year. A common mathematical model that could not be derived for any of the Wisconsin beaches.

APPENDIX

Mathematical Model for Fischer Park Beach 2008 Season from Manitowoc County

Manitowoc County

Fischer Park – 2008 model.

The mean (MPN+SD), median, and maximum $\log_e$ *E. coli* concentrations for Fischer Park beach 2008 season were (4.46 ± 2.48), 4.54, and 7.79 respectively. Correlations between *E. coli* concentrations and potential explanatory variables for data collected in the 2008 recreational season are shown in Table 88. The explanatory variables that showed positive correlation with *E. coli* concentration at Fischer Park beach for 2008 were included in the modeling process. The variables that showed significant positive correlation were wind speed ($r = 0.16$, $p<0.05$), wave height ($r = 0.07$, $p<0.05$), algae ($r = 0.32$, $p<0.05$), 24-hour rainfall (tfrain) ($r = 0.42$, $p<0.05$), and 72-hour rainfall (strain) ($r = 0.30$, $p<0.05$). Only these were included in the final predictive model, and are highlighted in Table-28 and shown in Table 88. The scatter plot shown in Figure 62 was constructed to show the relationship between the estimated and observed *E. coli* concentrations using the significant variables wind speed, wave height, algae, 24-hour rainfall, and 72-hour rainfall. The mathematical model predicted *E. coli* concentrations exceeding the bathing water standards (7 days), as seen in Figure 62 and Table 89. The adjusted R^2 of the mathematical model was 55.8% (Table 91) with lowest Cp was 6.0 (Table 91 and Figure 64) and the model was considered statistically significant ($p<0.05$), as seen in Table 91 using significant explanatory variables (Table 90, Figure 63). The predictive equation for Fischer Park Beach – 2008 model is as follows:

E. coli [Log (MPN)] = [3.4716 - (0.4514*ws) + (0.4168*wh) + (1.2503*algae) + (1.3731*tfrain) + (0.9239*strain)].

Table 1: Pearson's correlation coefficient (r) values between $\log_e$ *E. coli* concentrations and explanatory variables for Fischer Park Beach for 2008.

Variable	'r' value
Wind direction (degrees)	0.06
Wind speed	**0.16****
Onshore wind	-0.23
Water temperature	0.34
Air temperature	0.20
Turbidity	0.39
Wave height	**0.07****
Algae	**0.32****
One day late algae	0.36
Gulls	-0.35
One day late gulls	0.11
Two days late gulls	0.24
Long shore wind direction	-0.23
24 hour rainfall	**0.42****
48 hour rainfall	0.39
72 hour rainfall	**0.30****
Event rainfall	0.22
Weighted 72hr rainfall	0.56
Weighted 48hr rainfall	0.52
Combined 72hr rainfall	0.55
Barometric pressure	-0.33

**Variable considered statistically significant i.e. $p < 0.05$.

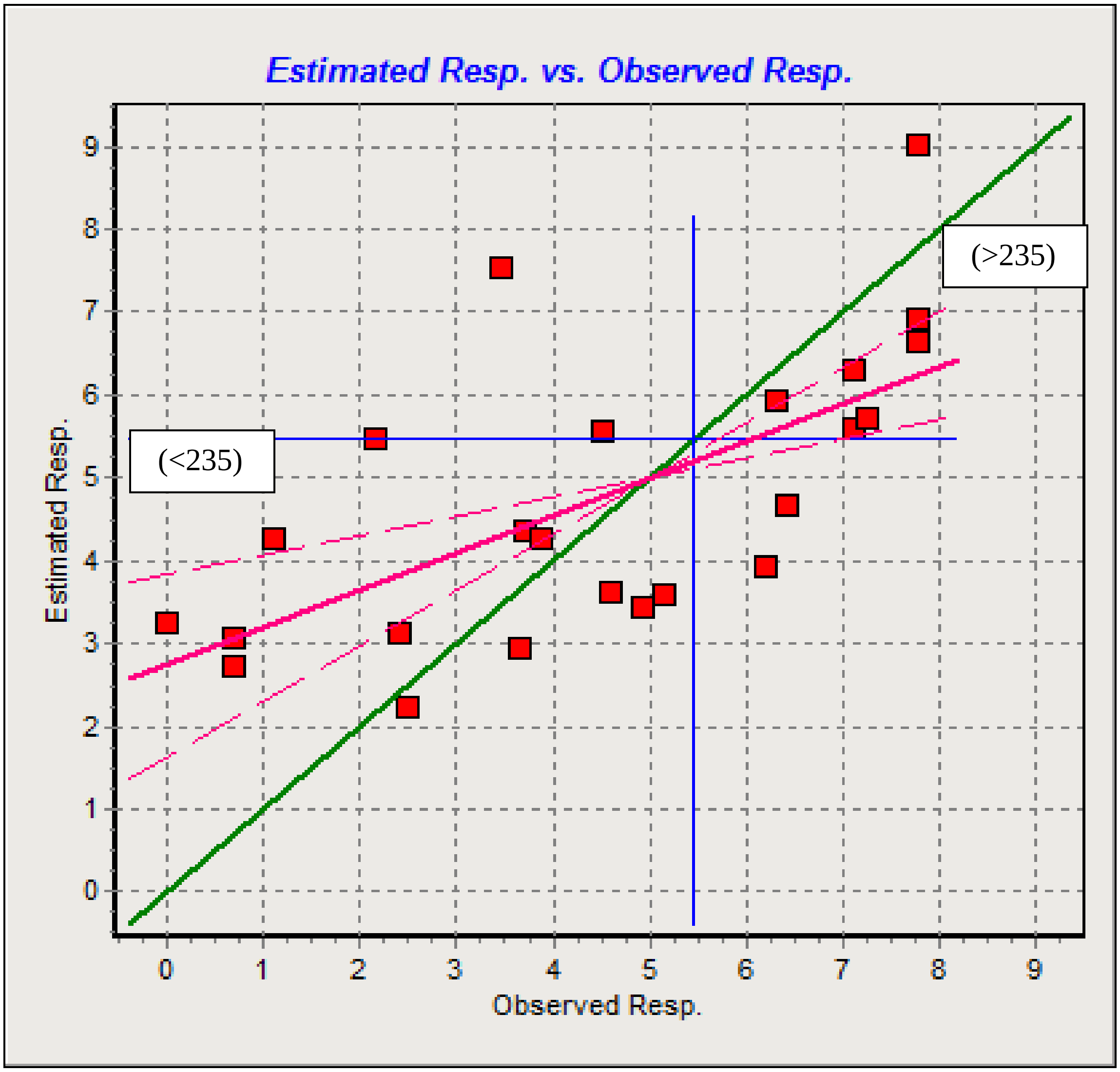

Figure 1: The Fischer Park Beach scatter plot shows the relationship between the estimated (predicted) and observed *E. coli* concentrations.

In the Figure 1, blue lines indicate bathing water quality standard (5.46) $\log_e$ 235CFU/ 100mL water. The purple line shows the "best fit" line of the predicted values with a 95% confidence interval (dashed) around it an adjusted R^2 of 55.8%. The green diagonal represents a perfect model for comparison.

Table 2: Fischer Park Beach 2008 model predictions and accuracy of data.

Predictions/Responses	Number	Percentage
Total predictions	24	100
Correct above standard	7	29.2%
Correct below standard	12	50 %
False positives	3	12.5%
False negatives	2	8.33%

Table 3: Fischer Park Beach 2008 mathematical model parameter estimates. P-values (<0.05) of t-tests indicate that each of parameter estimates is statistically different from zero.

Variable	Parameter estimate	Standard error	t-value	p-value
Intercept	3.4716	0.4313	3.8072	0.001413
Wind speed	-0.4514	0.2034	-2.2192	0.04047
Wave height	0.4168	0.1142	3.651	0.001984
Algae	1.2503	0.5524	2.2635	0.03708
24hr rainfall	1.3731	0.314	4.3735	0.0004154
72hr rainfall	0.9239	0.3764	2.4547	0.02524

Table 4: Fischer Park Beach – 2008 model statistical parameters. Adjusted R^2 indicates the fraction of variation in *E. coli* concentrations explained by the model. P-value (<0.05) indicates the model is statistically significant.

Statistic	Value
F-statistic	3.964
Degrees of freedom (DF)	18
Residual standard error (RSE)	1.408
Adjusted R^2	55.8%
Mallow’s Cp	6.0
p-value	0.01338

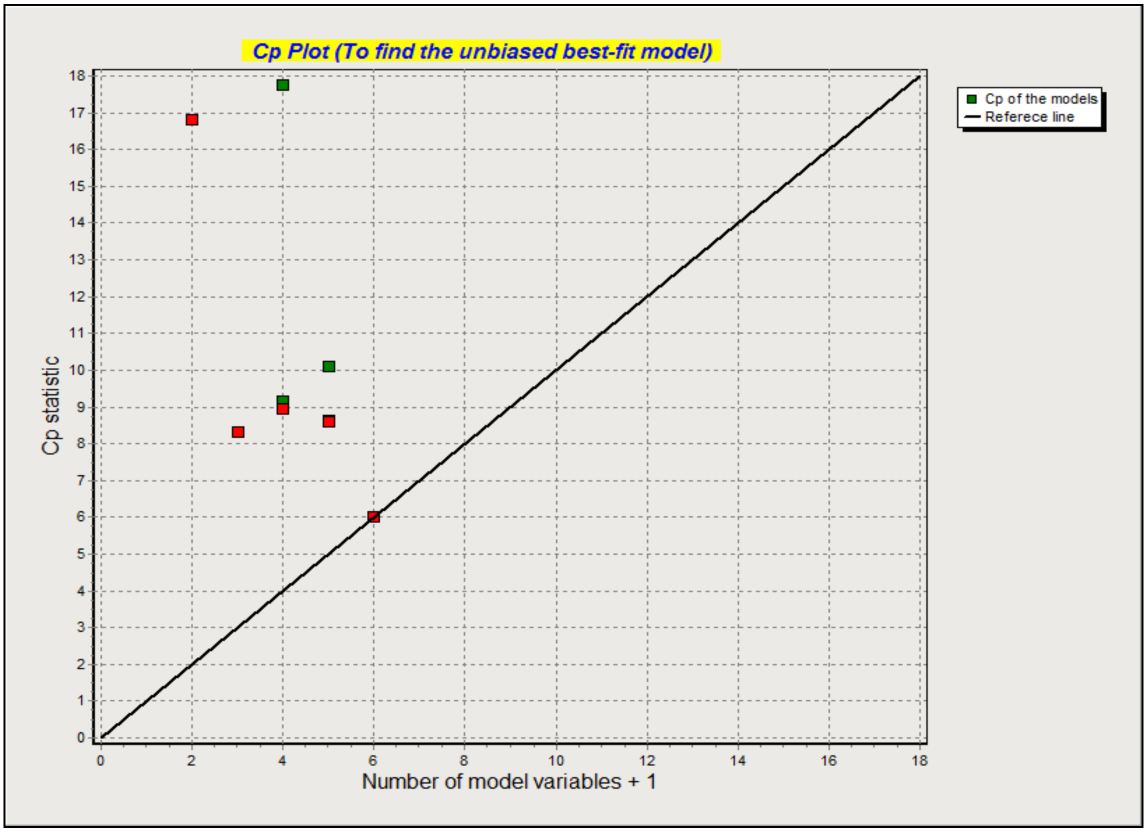

Figure 3: Fischer Park Beach – 2008 Cp-Plot of the mathematical model. The model with lowest Cp (6.0) is considered best fit.

In Manitowoc County, Fischer Park Beach was the only beach monitored and model developed for 2008 recreational season. Explanatory variables that were included in the mathematical model were wind speed, wave height, algae, 24hour rainfall and 72-hour rainfall explained 55.8% $\log_e$ *E. coli* variation. The total percentage of correct predictions (predictions correct above standard and correct below standard) is 79.2%. Additional sampling (data) would have improved the fitness of the model.

REFERENCES

Ackerman, D., and Weisberg, S.B. (2003). Relationship between rainfall and beach bacterial concentrations on Santa Monica Bay Beaches. *J. Water Health,* 1(2): 85-89.

Allsop, K., and Stickler, J.D. (1985). An Assessment of *Bacteroides fragilis* group organisms as Indicators of human fecal pollution. *J. Appl. Bacteriol.,* 58(1): 95-99.

Anderson, C.W.**, and Wilde, F.D., eds.,** (2005). Turbidity (version 2.1): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6, section 6.7.

Bej, A.K., McCarty, S.C., and Atlas, R.M. (1991). Detection of coliform bacteria and *Escherichia coli* by multiplex polymerase chain reaction: Comparison with defined substrate and plating methods for water quality monitoring. *Appl. Environ. Microbiol.,* 57(8): 2429-2432.

Bogosian, G., Sammons, L.E., Morris, P.J.L., O'Neil, J.P., Heitkamp, M.A., and Weber, D.B. (1996). Death of the *Escherichia coli* K-12 strain W3110 in soil and water. *Appl. Environ. Microbiol.,* 62, 4114-4120.

Brennan, F.P., O'Flaherty, V., Kramers, G., Grant, J., and Richards, K.G. (2010). Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. *Appl. Environ. Microbiol.,* 76(5): 1449-55.

Brettar, I., and Hofle, M.G. (1992). Influence of ecosystematic factors on survival of *Escherichia coli* after large scale release into lake water mesocosms. *Appl. Envorn. Microbiol.,* 58, 2201-2210.

Byappanahalli, M., **Shively, D., Nevers, M., Sadowsky, M., and Whitman, R.** (2003). Growth and Survival of *Escherichia coli* and enterococci populations in the macro-algae *Cladophora* (Chlorophyta). *FEMS Microbiol. Ecology,* 1575: 1-9.

Cabelli, V.J., Dufour, A.P., Levin, M.A., McCabe, L.J. & Haberman, P.W. (1979). Relationship of microbial indicators to health affects at marine bathing beaches. *Am. J. Public Health,* 69, 690-696.

Cabelli, V.J., Hoardley, A.W., and Dutka, B.J. (1976). Indicators of recreational water quality. In: Bacterial Indicators/Health Hazards Associated With Water. *ASTM,* p. 222-238, Philadelphia, PA.

Ederberg, S.C., Allen, M.J., Smith, D.B., and Kriz, N.J. (1990). Enumeration of total coliforms and *Escherichia coli* from source water by the defined substrate technology. *Appl. Environ. Microbiol.,* 56(2): 366-369.

Fattal, B., Vasl, R.J., Katzenelson, E., and Shuval, H.I. (1983). Survival of bacterial indicator organism and enteric viruses in Mediterranean coastal waters off Tel-Aviv. *Water Res.,* 17(4): 397-402.

Fiksdal, L., Make, J.S., LaCroix, S.J., and Staley, J.T. (1985). Survival and detection of *Bacteroides* spp., prospective indicator bacteria. *Appl. Environ. Microbiol.,* 49(1): 148-150.

Fogarty, L., **Haack, S.K., Wolcott, M.J., and Whitman, R.** (2003). Abundance and characteristics of the recreational water quality indicator bacteria *Escherichia coli* and enterococci in gull faeces. *J. Appl. Microbiol.,* 94(5): 865-878.

Francis, **G.** (2007). 5.1.4 Practical Issues and Assumptions (pp. 126-128). Retrieved from http://en.wikiversity.org/wiki/Multiple_linear_regression.

Francy, D.S., and Darner, R.A. (2006). Procedures for developing models to predict exceedances of recreational water-quality standards at coastal beaches. U.S. Geological Survey Techniques and Methods 6-B5.

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

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

Furness, B.W., Beach, M.J., and Roberts, J.M. (2000). Giardiasis surveillance – United States, 1992-1997. *MMWR*, 49(SS07): 1-13.

Henrickson, S.E., Wong, T., Allen, P., Ford, T., and Epstein, P.R. (2001). Marine swimming-related illness: Implications for monitoring and environmental policy. *Environ. Health Perspect.*, 109 (7, July), 645-650.

Hou, D., Rabinovici, S.J.M., and Boehm, A.B. (2006). Enterococci predictions from partial least squares regression models in conjunction with a single-sample standard improve the efficacy of beach management advisories. *Environ. Sci. Technol.*, 40, 1737-1743.

Kleinheinz, G.T., McDermott, C.M., and Chomeau, V. (2006). Evaluation of avian waste and bird counts as predicators of *Escherichia coli* contamination at Door County, Wisconsin beaches. *J. Great Lakes Res.*, 32(1): 117-123.

Kleinheinz, G.T., McDermott, C.M., and Sampson, R.W. (2003). Recreational water: Microbial contamination and human health. Wisconsin's Waters: A Confluence of Perspectives, Transactions, 90: 75-86.

Lee, S.H., Levy, D.A., Craun, G.F., Beach, M.J., and Calderon, R.L. (2002). Surveillance for waterborne-disease outbreaks-United States, 1999-2000. *MMWR*, 51 (SS08): 37-44.

Levy, D.A., Bens, M.S., Craun, G.F., Calderon, R.L., and Herwaldt, B.L. (1998). Surveillance for waterborne-disease outbreaks-United States, 1995-1996. *MMWR*, 47 (SS5): 1-34.

Minshew, P., Ward. K., Mulla, Z., Hammond, R., Johnson, D., Heber, S., and Hopkins, R. (2000). Outbreak of gastroenteritis associated with an interactive water fountain at a beachside park - Florida, 1999. *MMWR*, 49(25): 565-568.

Mitchell, L.G., Mutchmor, J.A., and Dolphin, W.D. (1988). Zoology. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Mueller-Spitz, S.R., Stewart, L.B., Klump, V., and McLellan, S.L. (2010). Freshwater suspended sediments and sewage are reservoirs for enterotoxin positive *Clostridium perfringens*. *Appl. Environ. Microbiol.*, doi:10.1128/AEM.01702-09.

NCSU (North Carolina State University). (1977). Water quality group, NCSU, watersheds Protozoan. Retrieved from http://www.water.ncsu.edu/watershedss/info/protozo.html.

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

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

Olyphant, G.A. (2004). Developing a prototypical predictive model for beach closings in Indiana. Indiana Geological Survey, Bloomington, IN.

Olyphant, G.A., and Whitman, R.L. (2004). Elements of a predictive model for determining beach closures on a real time basis: The Case of 63rd Street Beach Chicago. *Environ. Monit. Assess.*, 98(1-3): 175-190.

Paunio, M., Pebody, R., Keskimaki, M., Kokki, M., Ruutu, P., Oinonen, S., Vuotari, V., Siitonin, A., Lahti, E., and Leinikki, P. (1999). Swimming-associated outbreaks of *Escherichia coli* O157:H7. *Epidemiol Infect.*, 122(1): 1-5.

Rice, E.W., Allen, M.J., and Edberg, S.C. (1990). Efficacy of 1-Glucuronidase assay for identification of *Escherichia coli* by the defined substrate technology. *Appl. Environ. Microbiol.*, 56(5): 1203-1205.

Rose, B., R.M. Atlas, C.P. Gerba, M.R. Gilchrist, M.W. LeChevallier, M.D. Sobsey, M.V. Yates, G.H. Cassell, and J.M. Tiedje. (1999). Microbial pollutants in our nation's water: Environmental and public health issues. *American Society for Microbiology* (ASM), Washington, DC.

Sampson, R.W., Swiatnicki, S.A., McDermott, C.M., and Kleinheinz, G. T. (2005). *E. coli* at Lake Superior recreational beaches. *J. Great Lakes Res.*, 31, 116-121.

Sampson, R.W., Swiatnicki, S.A., McDermott, C.M., and Kleinheinz, G.T. (2006). The effects of rainfall on *Escherichia coli* and total coliform levels at 15 Lake Superior recreational beaches. *Water Resources Management*, 20: 151-159.

Sampson, R., Swiatnicki, S., Osinga, V., Supita, J., McDermott, C. & Kleinheinz, G. T. (2006). Effects of temperature and sand on *E. coli* survival in a northern lake water microcosm. *J. Water Health*, 04(3): 389-393.

Sayler, G.S., Nelson, J.D., Justice, A., and Colwell, R.R. (1975). Distribution and significance of fecal indicator organisms in the upper Chesapeake Bay. *Appl. Environ. Microbiol.*, 30(4): 625-638.

St. Louis, M.E. (1988). Water-related disease outbreaks, 1985. *MMWR*, 37(SS-2): 15-24.

U.S. Congress. (2000). Beaches Environmental Assessment and Coastal Health Act of 2000. Public Law 106-284-2000 Oct.10.

U.S. Commission on Ocean Policy. (2004). An ocean blueprint for the 21st century. Final Report. U.S. Commission on Ocean Policy, Washington, DC.

USEPA (U.S. Environmental Protection Agency). (1976). Quality criteria for water. U.S. Environmental Protection Agency, Washington, DC.

USEPA (U.S. Environmental Protection Agency). (2000). Liquid assets 2000: America's water resources at a turning point. EPA-840-B-00-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). (2002). Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using Modified membrane-Thermotolerant *Escherichia coli* agar (Modified mTEC). EPA-821-R-02-023. U.S. Environmental Protection Agency, Office of Water (4303T), Washington, D.C.

USEPA (U.S. Environmental Protection Agency). (2002). National beach guidance and required performance criteria for grants. EPA-823-B-02-004. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). (2006). Report to Congress 2006: implementing the BEACH Act of 2000. EPA-823-R-06-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USGS (United States Geological Survey). (2006). Fecal indicator bacteria and sanitary water quality. U.S. Geological Survey, Lansing, MI.

Vanden Heuvel. A., McDermott, C.M., Pillsbury, R., Sandrin, T., Kinzelman, J., Ferguson, J., Sadowsky, M., Byappanahalli, M., Whitman, R., **and Kleinheinz, G.T. (**2009). The green alga, Cladophora, promotes *Escherichia coli* growth and contamination of recreational waters in Lake Michigan. *J. Environ. Qual.*, 39: 333-344.

Weather forecasting website – Weather Underground. Retrieved from http://www.wunderground.com/us/wi.

Whitman. R.L., Nevers, M.B., Korineck, G.C., and Byappanahalli, M.R. (2004). Solar and temporal effects on *Escherichia coli* concentrations at Lake Michigan swimming beach. *Appl. Environ. Microbiol.*, 70(7): 4276-4285.

Wisconsin Department of Natural Resources. (2006). Wisconsin's Great Lakes Beach Monitoring and Notification Program Annual report: beach season 2006. Department of Natural Resources, Madison, WI.

Wisconsin Department of Natural Resources. (2008). Wisconsin's Great Lakes Beach Monitoring and Notification Program Annual report: beach season 2008. Department of Natural Resources, Madison, WI.

World Health Organization. (2003). Guidelines for safe recreational water environments 2003. Volume 1: Coastal and fresh waters. World Health Organization, Geneva.

Yoder, J.S., Blackburn, B.G., Craun, G.F., Hill, V., and Levy, D.A. (2004). Surveillance for waterborne-disease outbreaks associated with recreational water – United States, 2001-2002. *MMWR*, 53(SS08): 1-22.

Zhongfu, G., and Frick, W.E. (2007). Some statistical issues related to multiple linear regressions modeling of beach bacteria concentrations. *Environ. Res.,* 103 (3): 358-364.

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