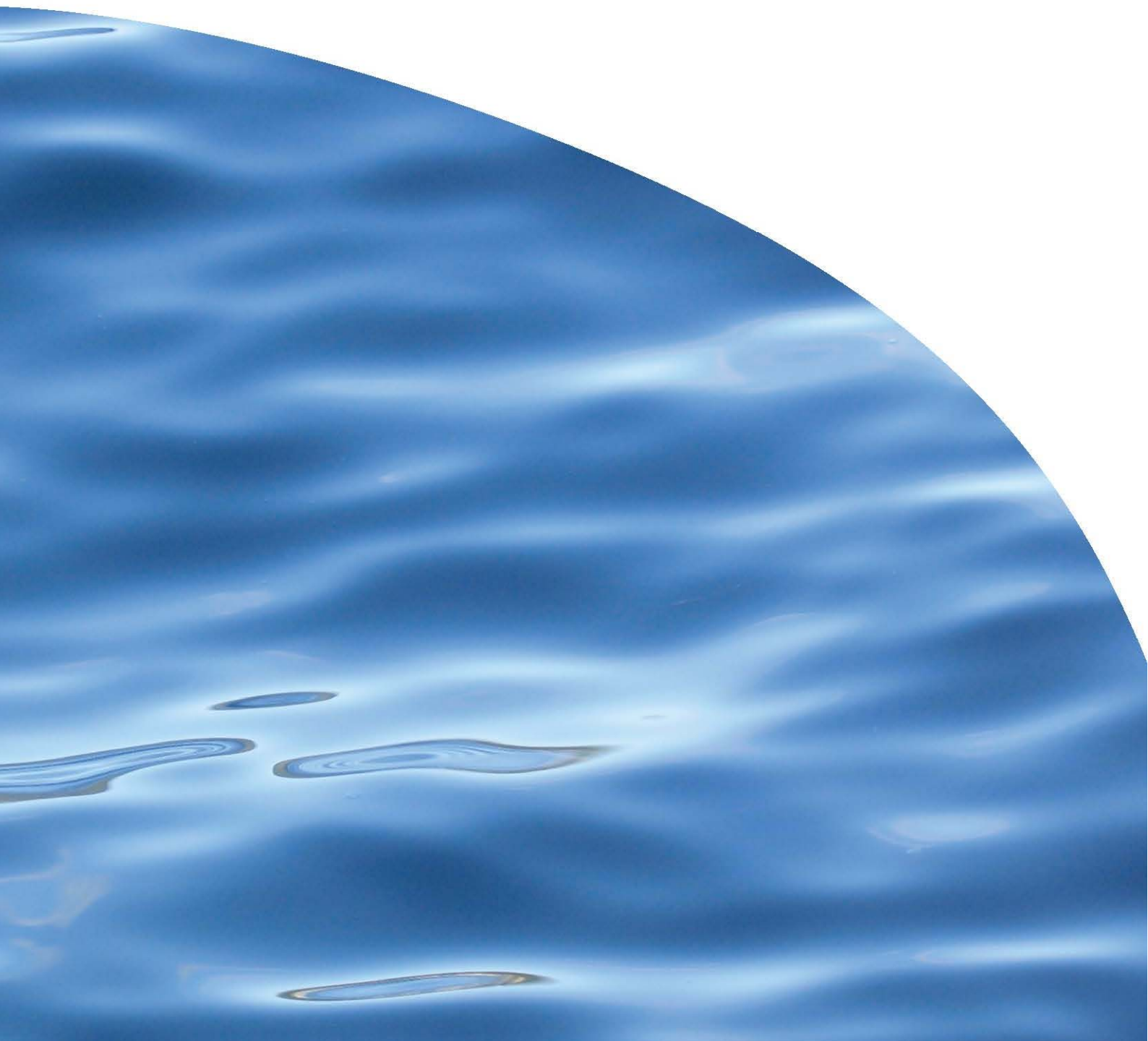




REPORT NO. 2122

**INDICATOR BACTERIA MODELLING DECISION
SUPPORT TOOL FOR RECREATIONAL BEACHES
AFFECTED BY THE MOTUEKA RIVER PLUME**



INDICATOR BACTERIA MODELLING DECISION SUPPORT TOOL FOR RECREATIONAL BEACHES AFFECTED BY THE MOTUEKA RIVER PLUME

BEN KNIGHT, WEIMIN JIANG

Prepared for Tasman District Council (TDC).

CAWTHRON INSTITUTE

98 Halifax Street East Nelson 7010 | Private Bag 2 Nelson 7042 | New Zealand
Ph. +64 3 548 2319 | Fax. + 64 3 546 9464
www.cawthron.org.nz

REVIEWED BY:
Chris Cornelisen



APPROVED FOR RELEASE BY:
Rowan Strickland



ISSUE DATE: May 2012

RECOMMENDED CITATION: Knight BR, Jiang WM. 2012. Indicator Bacteria Modelling Decision Support Tool for Recreational Beaches Affected by the Motueka River Plume. Prepared for Tasman District Council (TDC). Cawthron Report No. 2122. 26 p. plus appendices.

© **COPYRIGHT:** Apart from any fair dealing for the purpose of study, research, criticism, or review, as permitted under the Copyright Act, this publication must not be reproduced in whole or in part without the written permission of the Copyright Holder, who, unless other authorship is cited in the text or acknowledgements, is the commissioner of the report.

EXECUTIVE SUMMARY

This study aimed to construct a tool for Tasman District Council (TDC) to estimate real-time faecal indicator bacteria (FIB; enterococci) concentrations at a popular recreational beach (Kaiteriteri). This work was requested in response to a 2011 beach FIB closure over a popular Christmas period. Modelling tools were developed by considering potential drivers of FIB in the region, and tested to assess their potential use as decision tools following the approach undertaken by other international studies.

Customised FIB models for beaches at and around Kaiteriteri are based on candidate environmental data (northerly and southerly wind strength, Motueka River flow and solar radiation) and estimates of faecal indicator bacteria (FIB; *E. coli*) loading derived from a mechanistic faecal bacteria loading model developed for the Motueka River (Wilkinson, 2008).

Six candidate general logistic models were constructed to estimate the probability of detection FIB at the sites using the available explanatory and response data. The models were stratified by the direction of the northerly component of the wind and considered lags in the explanatory variables. The variance explained by all of the models was low (<2%) for estimating non-detection/detection of FIB at the sites. Although the variance explained by the models was low for the artificially for non-detection/detection of FIB, the model's performance were also tested at higher threshold exceedances to see if this could contribute additional information to improve decision making.

However, the diagnostic performance of the models for predicting the Ministry for the Environment (MfE 2003) bacterial guideline standards imply that the models were not able to substantially improve decision making; therefore we would not recommend the use of the models in an operational setting.

Factors that potentially contributed to poor model performance and possible solutions for improving forecasts of water quality conditions at Kaiteriteri Beach include the following:

1. **Data limitations:** The bathing water quality dataset used to build the model was limited to FIB data collected only during summer months and did not include many periods (and corresponding weather conditions and river flows) when FIB would have been elevated. Collection of more data during wet weather periods may provide better data for building robust models in future.
2. **Exclusion of local drivers:** Additional factors such as local conditions (wave action, tides), local inputs (e.g. from bathers, fauna and local streams) and resuspension of persistent bacteria in beach sands may be contributing to episodic events; inclusion of these factors in the model may improve forecasts. Microbial source tracking (MST) tools would also assist in discriminating between contamination driven primarily by fresh sources versus those that are associated with persistent FIB.

3. Model flexibility: The model techniques used, although successful in other systems, may not have allowed for non-linear interactions. Revisiting our existing dataset with different modelling techniques could produce more successful models in future.

These issues indicate that a mechanistic approach to forecasting contamination at the site could be more successful.

The stated objective of producing a robust FIB decision tools for Kaiteriteri Beach was not met, but the study provides a useful methodology which could be developed for other areas. Although not part of the original objectives, we show that a Motueka River FIB model (Wilkinson, 2008) can be used to successfully predict MfE guideline standards in the river. Although the river model alone was not able to provide a level of confidence that would enable its use in an operational sense, it was particularly useful when combined with the logistic modelling techniques presented in this study and its performance is classified as 'excellent'. We suggest this Motueka River FIB model may be of assistance to the Tasman District Council (TDC) in implementing near real time estimates (and possibly forecasts in future) of riverine FIB quality for internal and public users of the river and mudflat areas around the river in future.

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	III
1. INTRODUCTION	1
1.1. Background to faecal indicator bacteria drivers and die-off in the coastal environment	3
1.2. Review of approaches to quantifying drivers of FIB concentrations	4
2. METHODS	5
2.1. Analysis approach	5
2.2. Descriptions of explanatory and response data.....	5
2.2.1. <i>Response data</i>	5
2.2.2. <i>Explanatory data</i>	7
2.3. Data exploration	8
2.3.1. <i>Accounting for colinearity</i>	10
2.4. Logistic modelling approach	11
2.5. Time lag analysis.....	12
2.6. Performance assessment.....	14
3. RESULTS AND DISCUSSION	16
3.1. Model construction and evaluation	16
3.2. Performance analysis	16
4. CONCLUSIONS.....	21
5. ACKNOWLEDGEMENTS.....	23
6. REFERENCES	24
7. APPENDICES.....	27

LIST OF FIGURES

Figure 1.	Study area, showing proximity of Kaiteriteri Beach to the Motueka and surrounding rivers and sampling sites at Woodstock and Woodmans Bend.....	1
Figure 2.	Log-transformed data from each of the sites show a high number of zero counts and no significant differences.	6
Figure 3.	Comparison of modelled to measured riverine <i>E. coli</i> concentrations over the April 2009 Motueka flood event.....	8
Figure 4.	Comparison of response and explanatory variables, showing data distributions, and Pearson's <i>r</i> for all variables.....	9
Figure 5.	Comparison of log transformed response and explanatory variables, showing data distributions, and Pearson's <i>r</i> for all variables.	10
Figure 6.	Variation in lagged explanatory variable correlations for the whole dataset.	13
Figure 7.	Variation in lagged explanatory variable correlations for a southerly wind data.....	13
Figure 8.	Variation in lagged explanatory variable correlations for northerly wind data.	14
Figure 9.	ROC curves for All Data, Northerly and Southerly, using the detection of enterococci as a threshold.....	18
Figure 10.	ROC curves for All Data, Northerly and Southerly, using 30 cfu/100ml as a threshold. ..	19
Figure 11.	ROC curves for All Data, Northerly and Southerly, using the MfE Marine Guideline Standard of 140 cfu/100ml as a threshold.	20

LIST OF TABLES

Table 1.	Guideline levels and management response for marine water samples.....	2
Table 2.	Response and candidate explanatory variables used in the development of the model, the short names used in our models and the data source.....	5
Table 3.	Sample sizes of enterococci data used in model construction.	6
Table 4.	Results of final VIF analysis for each of the six different models	11
Table 5.	Summary of lags applied to the models.....	14
Table 6.	Explanatory variable coefficients and their significance from the final model selections. 16	

LIST OF APPENDICES

Appendix 1.	The Motueka River <i>E. coli</i> Model validation.....	27
Appendix 2.	R version of the Motueka River <i>E. coli</i> Model.....	30
Appendix 3.	Performance of the FIB River Model for operational use.	33

1. INTRODUCTION

Kaiteriteri Beach is a very popular bathing beach in the summer periods and is located to the north of the Motueka River. The beach is within the influence of the Motueka River plume (Figure 1, Gillespie *et al.* 2011) and the catchment surrounding the river has farming activities which contribute to observed bacterial concentrations in the river (Wilkinson *et al.* 2011) and Tasman Bay (Cornelisen *et al.* 2011).

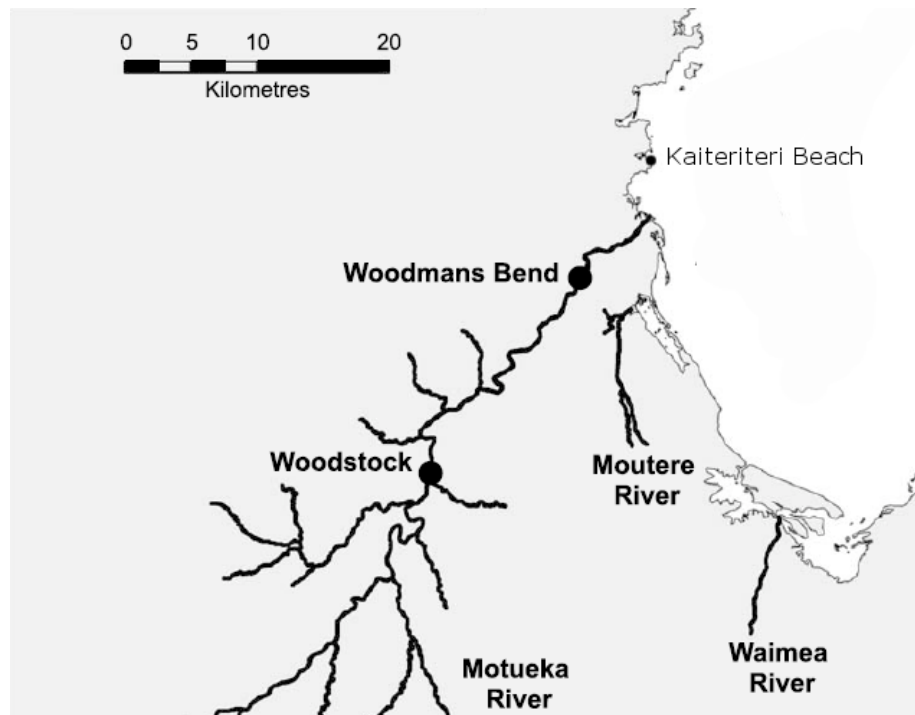


Figure 1. Study area, showing proximity of Kaiteriteri Beach to the Motueka and surrounding rivers and sampling sites at Woodstock and Woodmans Bend.

Regular water sample tests are carried out over the summer bathing season by Tasman District Council (TDC) to ensure that people using Kaiteriteri Beach for contact recreation are not exposed to significant health risks from poor water quality. Water samples are assessed against the Ministry for the Environment (MfE) and Ministry of Health (MoH) Microbial Water Quality Guidelines (MfE 2003; hereafter referred to as the 'MfE Marine Guidelines'). The MfE Marine Guidelines use enterococci bacteria as a faecal indicator bacteria (FIB) to detect the potential presence of harmful pathogens in marine waters and provide a framework for addressing appropriate management response (Table 1).

Table 1. Guideline levels and management response for marine water samples (MfE 2003).

Mode	Guideline (Enterococci count in colony-forming units (cfu) per 100 mL)	Management response
Green /Surveillance	Single sample \leq 140	Routine monitoring
Amber/Alert	Single sample $>$ 140	Increased monitoring, investigation of source and risk assessment
Red/Action	Two consecutive samples within 24 hours $>$ 280	Closure, public warnings, increased monitoring and investigation of source

Detectable concentrations of enterococci are common at Kaiteriteri Beach after rain events in summer, which suggests a link between riverine flows and beach FIB concentrations. However, the incidences of FIB levels exceeding the amber alert mode of the MfE Marine Guidelines are rare.

A recent beach closure due to an enterococci concentration exceedance event that occurred at Kaiteriteri Beach over the 2010 Christmas period led to a community demand for quicker decision making by TDC. Presently, laboratory testing of FIB (Enterococci) concentrations at the site introduces long delays ($>$ 24 hours) between sampling and the availability of results. This study aims to provide information that could aid quicker decisions whilst attempting to improve an understanding of environmental and catchment drivers of FIB concentrations in the region.

This study builds from existing knowledge obtained through 10 years of research in the Motueka catchment and river as part of the FRST funded integrated catchment management (ICM) programme. The study also leverages information from other State of the Environment (SoE) data collected by TDC around Kaiteriteri Beach from 1996 to present.

The following steps were proposed:

1. Review previously published modelling methods that could potentially be applied for predicting faecal contamination at Kaiteriteri beaches. .
2. Generate statistical regression models of riverine bacterial concentrations, based on knowledge of the processes affecting die-off and transport of FIB in the region for key recreational beaches where data is available.
3. Test the performance of the models ability to predict available FIB measurements at the sites and aid in the FIB decision process.
4. Develop an operational tool that may be automated in future by TDC to provide near real time estimates of FIO concentrations at coastal sites.
5. Enable any constructed model to be integrated with the Hilltop software package by the subcontractor (to be completed after publication of this report).

1.1. Background to faecal indicator bacteria drivers and die-off in the coastal environment

FIB concentrations are often highly correlated to rainfall and increased inputs of diffuse sources of faecal contamination (from farming) during and following rain events (e.g. Wilkinson *et al.* 2011; Cornelisen *et al.* 2011a). Commonly measured FIB such as *E. coli* and enterococci are known to persist in a terrestrial environment once released by the host (Ishi & Sadowsky 2008). These bacteria can survive for periods of days to weeks in river bed sediments, which can then be released in large pulses of contamination to coastal environments during flood events (Wilkinson *et al.* 2011). Hence, river flooding and antecedent rainfall patterns therefore play a major role in driving FIB concentrations observed near the coast.

With the exception of birds, the dominant sources of faecal contamination contributing to coastal waters are primarily associated with diffuse non-point source pollution and potentially point source pollution (sewage outfalls). Spatial gradients in FIB concentrations are likely to exist within coastal environments, with higher concentrations in proximity to incoming sources and lower concentrations observed with distance and dilution with marine waters.

In addition to river bed sediments, estuary sediments and beach sands can also act as reservoirs for FIB that can be released into the water column just following high tides (e.g. G. Lewis presentation at WaterMicro2011 on Bethells Beach; Boehm and Weisburg 2005; Yamahara *et al.* 2007). Hence, tidal movements can influence the extent of dilution of FIB populations in nearshore waters; this is particularly pertinent in shallow estuaries where strong currents can suspend fine sediments and associated bacterial populations (e.g. Sanders *et al.* 2005). Wave action can also enhance the resuspension of sediments and beach sands, and thereby also has the potential to release persistent populations of FIB to the marine environment (Yamahara *et al.* 2007).

FIB populations are also subject to a number of mortality pressures. Compared to fresh water, a faster die-off of *E. coli* and enterococci is observed when exposed to brackish and saline water. Temperature is also known to affect the viability of FIB, but the major driver of faecal bacteria mortality is exposure to UV radiation.

FIB concentrations in clear waters exposed to sunlight, typically decline to less than 90% of original concentrations over the course of the day (Rosenfield *et al.* 2006). As a result, the time of day, combined with water clarity characteristics, antecedent rainfall patterns and proximity to incoming freshwater sources, may all have large influences on variability in FIB concentrations. As a consequence of all of these processes and effects on the survival of FIB in the coastal environment, it can be difficult to find simple explanatory variables of FIB concentrations at coastal sites.

1.2. Review of approaches to quantifying drivers of FIB concentrations

Authorities around the world are under pressure to make quicker decisions about whether bathing water areas remain open or closed. Masopust (2005) notes that the delayed bacterial results generally have a poor correlation with the period they are meant to represent (e.g. Boehm *et al.* 2002; Whitman & Nevers 2004), and therefore beach closures made on the basis of 'yesterdays' data may not be reliable. Consequently there exist strong incentives to improve the predictability of short-term variability.

Whilst investigating solutions to address these issues, two possible approaches are evident in the literature:

1. The use of molecular or PCR tools to reduce the long incubation periods required under traditional approaches (e.g. Noble & Weiseberg 2005)
2. The use of models based on existing data to statistically model the likely FIB concentrations at a site (e.g. Diane & Ahlfield 2007; Eleria & Vogel 2005), or alternatively the use of an theoretical understanding of the system to mechanistically model FIB transport and survival (Wilkinson *et al.* 2011; Steets & Holden 2003).

In this study we were focused on a statistically based modelling approach.

A review of similar attempts to derive statistical predictive models from environmental forcing suggests this approach had a reasonable chance of successfully predicting FIB concentrations (e.g. Eleria & Vogel 2005; Olyphant & Whitman 2004; Maimone *et al.* 2007).

Our review also highlighted a number of statistical modelling methods that could be used in studies available from the literature; of note were two main approaches:

1. Machine learning or artificial intelligent methods, which includes the use of techniques, such as:
 - a. artificial neural networks (Diane & Ahlfield 2007)
 - b. genetic algorithms (Tufail *et al.* 2008).
2. Regression modelling methods (e.g. Eleria & Vogel 2005, Olyphant & Whitman 2004; Christensen *et al.* 2001; Diane & Ahlfield 2007).

Tufail *et al.* (2008) note that if logarithm transformed data are used that the performance of both machine learning and regression approaches are comparable and that both techniques are potentially suitable for estimating FIB water quality.

2. METHODS

2.1. Analysis approach

Given the primary objective of aiming to produce reliable models for the region, but also improving an understanding of FIB explanatory variables in the region, this study adopted the multiple linear regression modelling approach employed by Eleria & Vogel (2007).

The multiple linear regression modelling approach attempts to use a number of explanatory variables to improve the performance of the model at predicting observed bacterial concentrations at and around Kaiteriteri Beach. The model takes the form of:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots + E \quad (1)$$

Where, β_i are loadings on each explanatory variable (X_i) determined by linear regression, Y is a response variable (*i.e.* bacterial concentrations at the beach) and the X_i are the appropriate environmental explanatory variables (Table 2). E is an error term which accounts for the unexplained fraction of the model.

Table 2. Response and candidate explanatory variables used in the development of the model, the short names used in our models and the data source.

Variable name	Short name	Description	Source
Kaiteriteri Beach Enterococci counts	bacti	Response variable	TDC
Motueka River <i>E. coli</i> model estimates	bactiModel	Candidate explanatory variable	Appendix 1; Wilkinson (2008)
Motueka River flow	Flow	Candidate explanatory variable (m ³ /s)	TDC
Woodman's Bend Solar radiation (W/m ² , Riwaka)	Light	Candidate explanatory variable	NIWA CliFlo
North/South wind (m/s, going to)	U	Candidate explanatory variable – 6 hour smoothed North/South wind (m/s)	NIWA CliFlo
East/West wind (m/s, going to)	V	Candidate explanatory variable – 6 hour smoothed East/West wind (m/s)	NIWA CliFlo

2.2. Descriptions of explanatory and response data

2.2.1. Response data

FIB data obtained through state of the environment monitoring by TDC at and around Kaiteriteri Beach was used to investigate relationships between candidate environmental explanatory variables. Data from six sites were used in the construction

of the model, providing 351 samples from 1 January 2001 (Table 3). Sites were aggregated to increase the amount of data available for model construction due to the close geographical proximity of the sites and a lack of *E. coli* significant differences observed in the distributions between the sites (Figure 2).

Table 3. Sample sizes of enterococci data used in model construction.

Site name	Sample size
BW Breaker Bay	33
BW Kaiteriteri (North)	15
BW Kaiteriteri (South)	15
BW Kaiteriteri Beach	198
BW Little Kaiteriteri Beach	81
BW Stephens Bay	9
Total	351

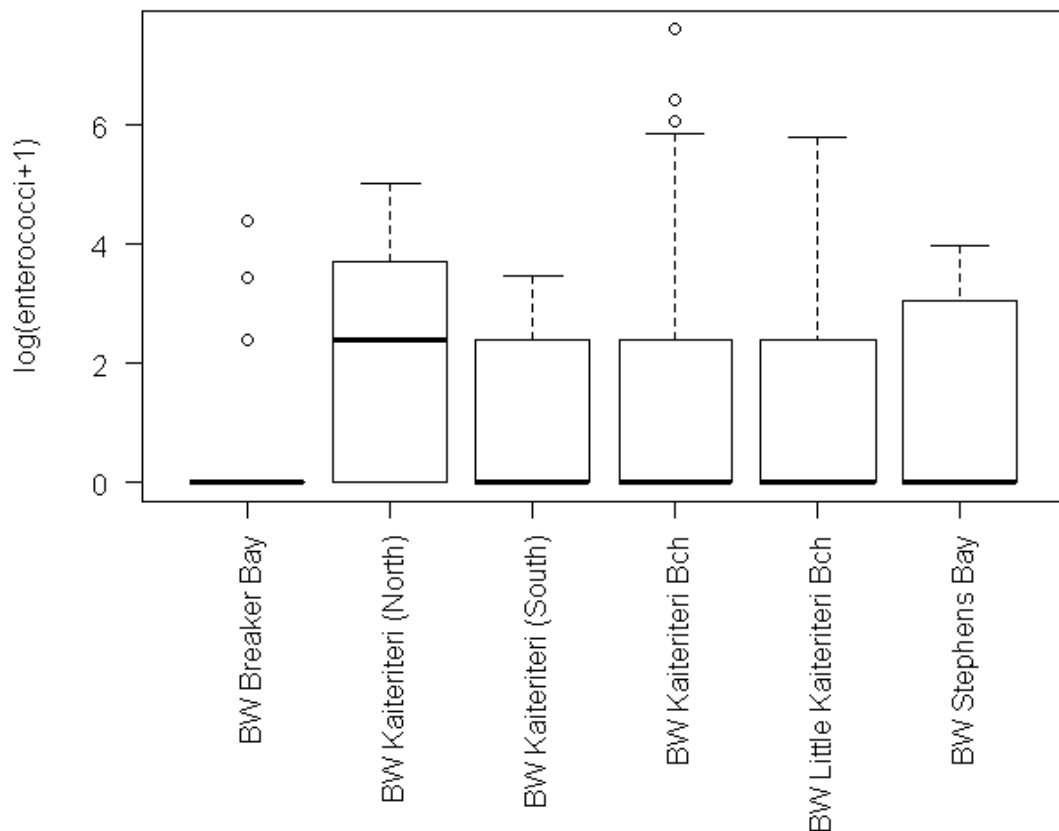


Figure 2. Log-transformed data from each of the sites show a high number of zero counts and no significant differences. Note that the edges of the boxes represent the upper and lower quartiles, whilst the dark line in the boxes represents the median value. The dashed lines represent the lesser of the maximum or 1.5 times the interquartile range, with outlying data points marked with an 'o'.

2.2.2. Explanatory data

Data used to build the models were composed of existing validated mechanistic river model estimates of *E. coli* (Wilkinson *et al.* 2011) and environmental forcing data supplied by TDC (River flow) and the NIWA Cliflo database (wind, radiation, temperature data) (Table 2).

Motueka River *E. coli* Model

A single point riverine *E. coli* model was developed in the ICM programme (Wilkinson 2008) and is also based on TDC (River flow) and meteorological NIWA Cliflo (NIWA, 2011) data from the region. This model successfully reproduced the timing, magnitude and variability of *E. coli* concentrations near the river mouth (Figure 3). As *E. coli* and enterococci concentrations are often correlated in riverine water (MfE 2003)¹, the model data was investigated as a possible explanatory variable of the Kaiteriteri beach enterococci concentrations in this project. The model performance was good and it was deemed suitable for inclusion as a candidate explanatory variable (model performance detailed in Appendix 1).

A 10 year dataset of predicted *E. coli* concentrations was constructed for this project and will be supplied with this report along with the model code (Appendix 2). It is envisaged these data and the model may find use in other council activities, possibly as a predictive tool to provide information to the public on possible bacterial contamination in the river in the absence of monitoring data. The single point model used here (Wilkinson 2008) has also since been updated to include multiple catchments (Wilkinson *et al.* 2011) and appears to offer improved performance but has not been applied to this study.

¹ Enterococci generally have a longer survival in seawater than *E. Coli*, hence its use is recommended as a marine FIO.

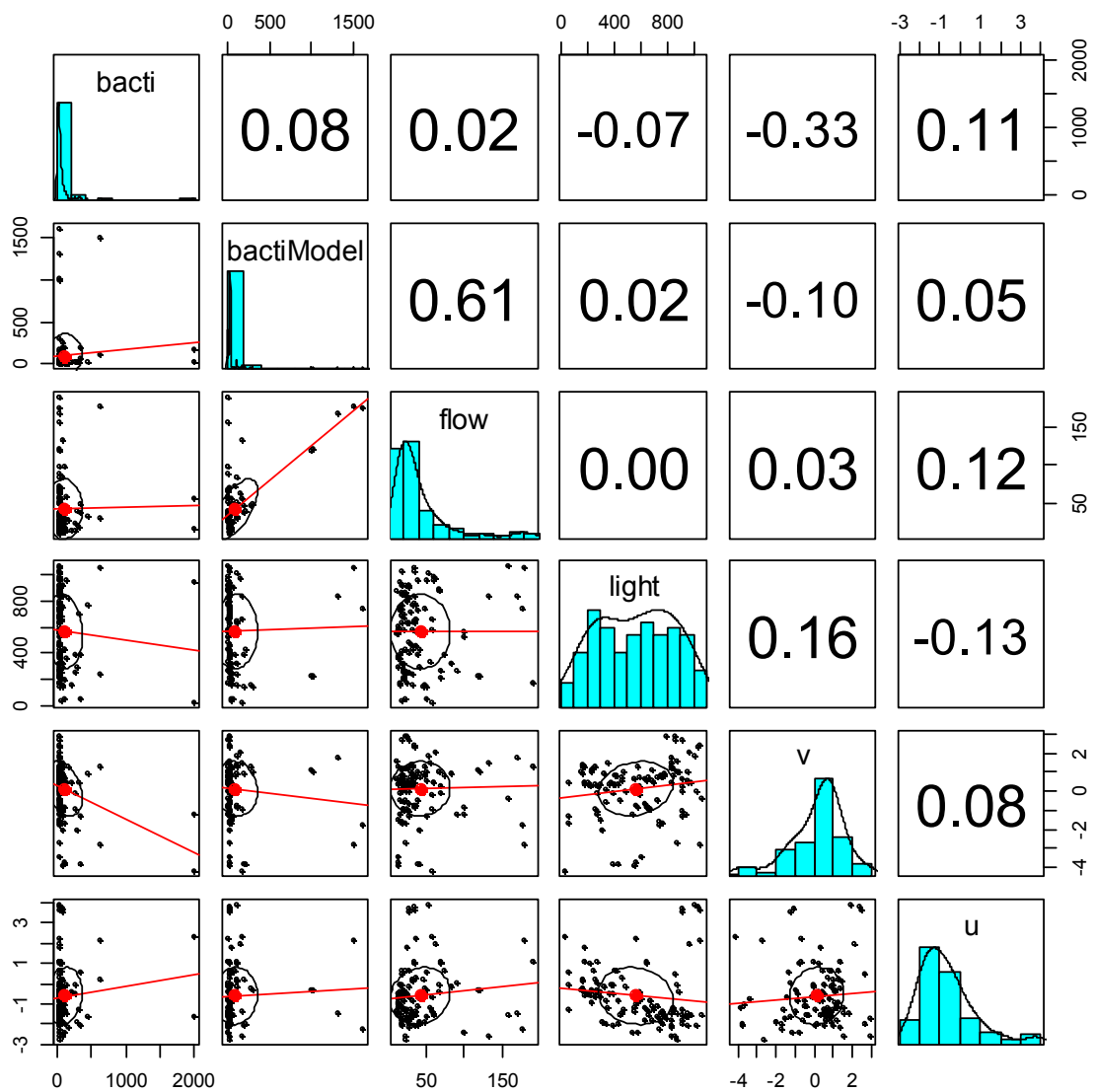


Figure 4. Comparison of response (bacti) and explanatory variables, showing data distributions, and Pearson's r for all variables. Abbreviations for variable names are provided in Table 2.

The initial analysis of the histograms in Figure 4 showed some outliers in the candidate explanatory variables dataset (*i.e.* bactiModel, Flow and v). In order to minimise the influence of outlier data on the beach model development, these were log transformed (*i.e.* $\log_{10}(X+1)$; Figure 5). It is recognised that although the underlying distributions of explanatory variables were not normally distributed, that the linear modelling approach is reasonably robust against violations of these assumptions (Fitzmaurice *et al.* 2004 as referenced by Zuur *et al.* 2010).

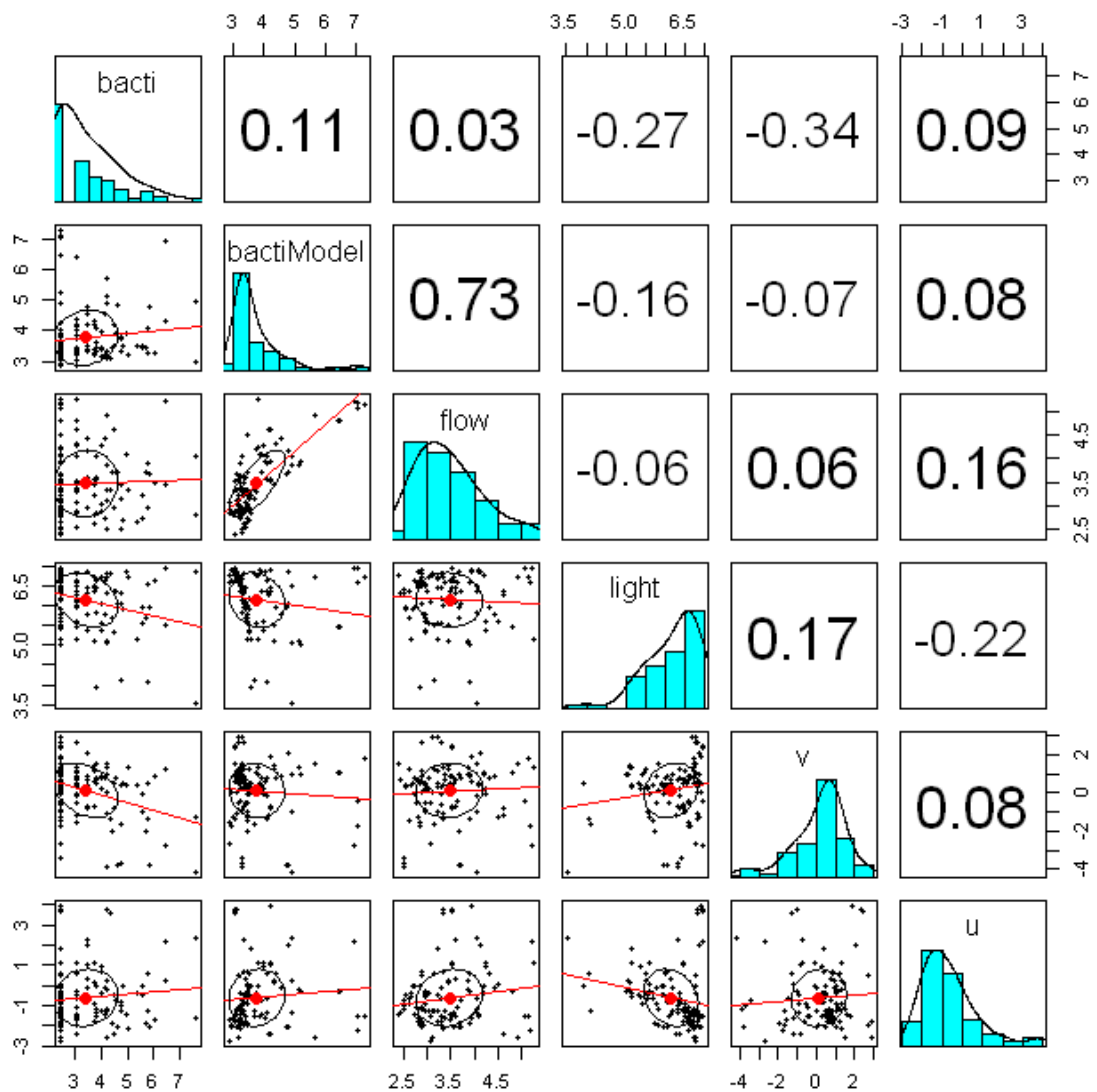


Figure 5. Comparison of log transformed response (bacti) and explanatory variables (bactiModel, light, and flow), showing data distributions, and Pearson's r for all variables.

The input data was also smoothed using a six hour moving average window to remove high frequency artefacts which could complicate the modelling process.

Southerly winds ($v > 0$) have been observed to drive the Motueka plume northwards towards Kaiteriteri Beach and northerly wind events ($v < 0$) away. Hence two separate north and south wind models based on a split in the smoothed northerly component of the wind field (v) were investigated.

2.3.1. Accounting for colinearity

The data exploration of the transformed data and lagged data (see Section 2.5 for a discussion of lags) also showed the potential for collinear explanatory variables to

cause issues in the construction of the model (Figure 5). In order to ensure that collinear explanatory variables were not included in the model construction, variables with the highest variance inflation factor (VIF) were sequentially excluded from the model, until VIF values were all below two (see e.g. Fox 1997, as cited by Blanchard *et al.* 2005).

Due to the use of flow data (flow) in the Motueka River *E. coli* Model (bactiModel) (*i.e.* a possible confounding variable) and their high correlation (e.g. Figure 5, Pearson's $r=0.73$), the variable with the higher VIF value was removed from each of the models, even if the VIF value was less than two. Table 4 shows the results of this analysis for each of the six models investigated (three lagged and three unlagged models), with the excluded explanatory variables highlighted in the table.

Table 4. Results of final VIF analysis for each of the six different models. Explanatory variables VIF values indicated in bold were removed prior to initiating the model selection process due to issues with colinearity.

Explanatory variable/model	Lagged	log(bactiModel + 1)	log(flow + 1)	log(light + 1)	u	log(v + 1) ¹
Southerly wind	Y	1.01	2.27	1.03	1.29	1.31
Northerly wind	Y	2.09	1.04	1.10	1.11	1.03
Whole dataset	Y	1.86	1.01	1.02	1.43	1.44
Southerly wind	N	1.01	2.27	1.03	1.29	1.31
Northerly wind	N	1.99	1.07	1.08	1.16	1.09
Whole dataset	N	1.00	1.85	1.00	2.11	1.01

¹Note that the logarithm of the northerly component of wind (v ; $\log(v+1)$) was not taken for the whole dataset analysis and the absolute value of v was used for the northerly wind scenario (when $v < 0$).

2.4. Logistic modelling approach

The histograms shown in Figure 4 also show that the response variable (bacti) was zero biased, with many samples returning non-detection results. Consequently, quantitative model construction was not possible and a logistic modelling approach was required. This approach aimed to model the probability of detection, rather than estimating FIB concentrations directly from the explanatory variables.

The form of the logistic model is similar to regular regression shown in equation (1), but the response variable (Y) is replaced by a logit function, and there is no error term, so the form of the equation becomes:

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots \quad (2)$$

The logit function is defined to be equivalent to the logarithm of the odds ratio for the probability (p) of a predefined outcome being successful:

$$\text{logit}(p) = \log(p/(1-p)) \quad (3)$$

In this study the detection of FIB would have $p = 1$ and $p=0$ for non-detection.

The model derivation of the loadings for the logistic models are determined by the base R statistical modelling function 'glm' with a binomial distribution applied (RDC Team 2011).

The relevant X_i and β_i were then determined by backward model selection to minimise the Aikeke Information Criterion (AIC; Aikeke 1974), whilst maintaining the explained variance (R^2) in the models. This process involves constructing a model and successively removing the least significant variable from the model until the AIC does not decrease. A further check was made to ensure that the standard error in the coefficient estimates was not greater than the estimates themselves. If this occurred, the least significant variable was removed from the model until this additional criteria was met.

2.5. Time lag analysis

A time lag between major sources of FIB and the Motueka River and Kaiteriteri Beach were possibly due to the distance between them. In order to allow for this, appropriate time lags between changes in candidate explanatory variables and the response variable was also explored. Comparison of lagged single explanatory variable correlations to observed bacterial beach concentrations was undertaken at hourly time steps (up to 12 hours). Maximum absolute correlations were used to determine lags for each variable to create an 'optimal' lagged model. It is recognised this is not a truly optimised lagged model, but it represents an objective method for providing lagged model development without requiring an exhaustive number of model permutations.

The lagged modelling approach can allow for the effects of travel time from riverine sources to the beach. Although qualitative evidence exists for riverine sourced bacteria causing detection of bacteria at and around Kaiteriteri Beach, the possibility of local sources is not discounted by this study. Any lagged model was therefore compared to an unlagged version – with the best model selected.

The results of the time lag analysis on the entire dataset, and northerly and southerly wind subsets of the data do show variability in the correlation to the response variable, with higher Pearson's correlations seen for some lagged variables (Figures 6, 7, and 8).

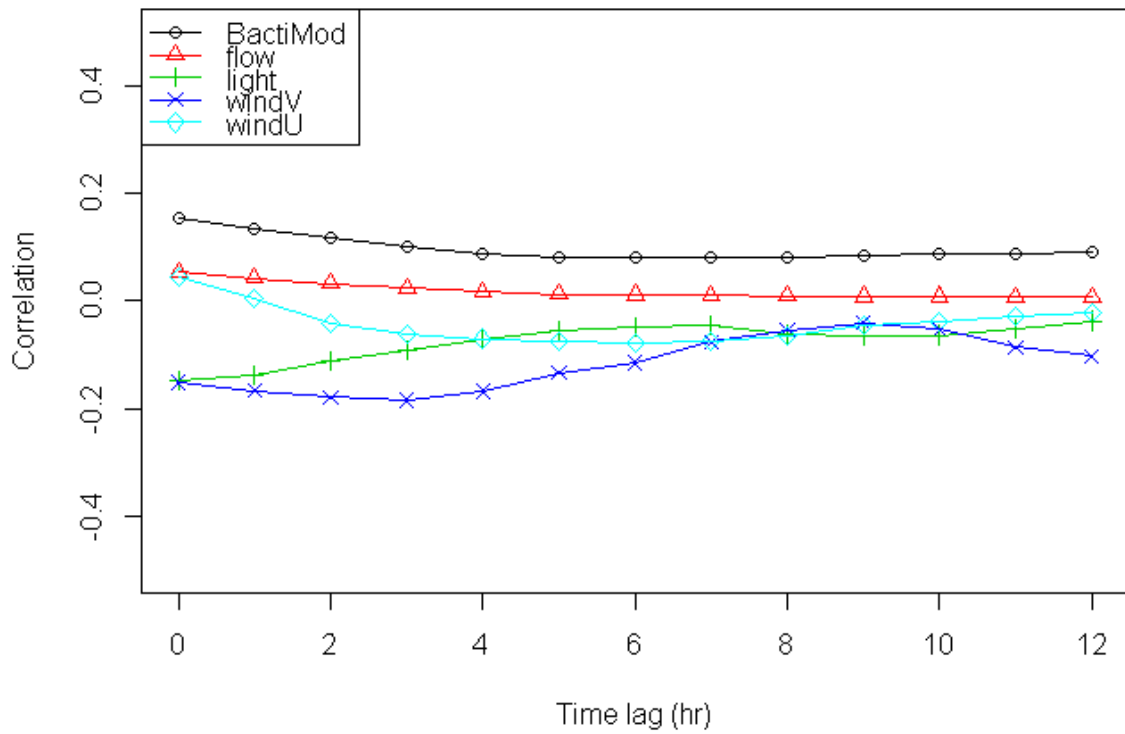


Figure 6. Variation in lagged explanatory variable correlations for the whole dataset.

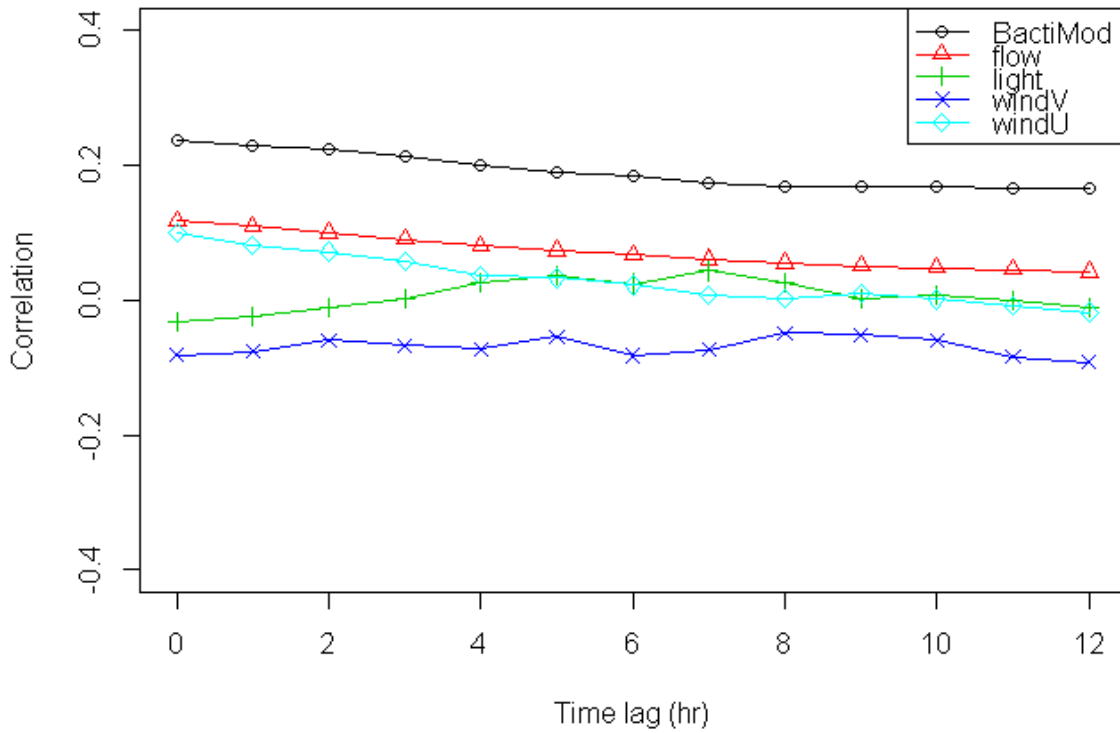


Figure 7. Variation in lagged explanatory variable correlations for a southerly wind data.

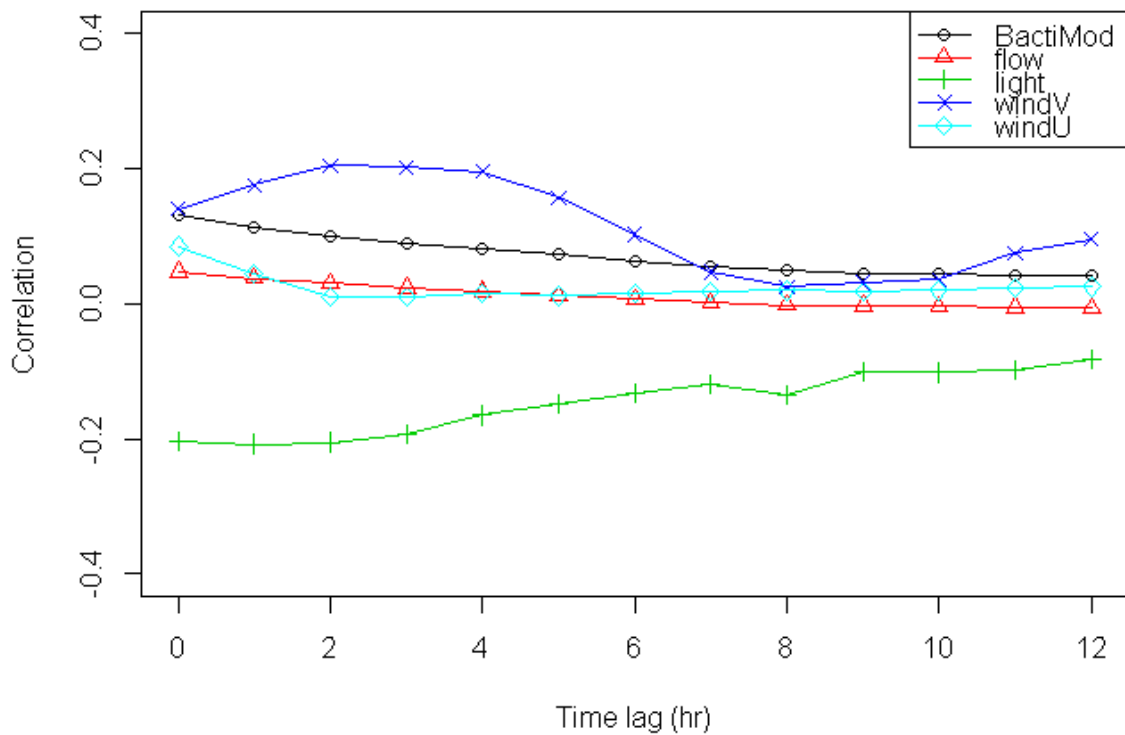


Figure 8. Variation in lagged explanatory variable correlations for northerly wind data.

On the basis of the lagged results, the following lags were applied to the three different lagged variable models (Table 5).

Table 5. Summary of lags (in hours) applied to the models.

Explanatory variable/model	$\log(\text{bactiModel} + 1)$	$\log(\text{flow} + 1)$	$\log(\text{light} + 1)$	u	$\log(v + 1)^1$
Southerly wind	0	0	7	0	0
Northerly wind	0	0	1	0	3
Whole dataset	0	0	0	6	3

¹ Note that the logarithm of the northerly component of wind (v ; $\log(v+1)$) was not taken for the whole dataset analysis and the absolute value of v was used for the northerly wind scenario (*i.e.* when $v < 0$).

2.6. Performance assessment

The potential utility of the models for decision making was ultimately determined using methods based on the receiver operating characteristic (ROC) performance results (*e.g.* sensitivity and specificity) for a range of bacterial concentration thresholds (Table 1).

Evaluation of the models through ROC analysis attempted to assess the ability of the models to estimate historical rates of true to false positives for different probability of bacterial detection thresholds for the model (P_{crit}). Depending on the application of the model (*i.e.* used for research, or determining human health issues) the rates of false negatives, in particular, may also be important (*e.g.* assessing the likelihood of predicting no exceedance, when MfE Marine Guidelines were actually exceeded).

The performance of the models is determined by assessing the area under a ROC curve (AUC) of the true positive to the false positive rate. A value of 0.5 represents a random model (not useful) whilst a value of 1 is excellent. This is more formally defined using the traditional academic point system (Swets 1988 as referenced by Thuiller *et al.* 2003), where: 0.90-1.00 = excellent; 0.80-0.90 = good; 0.70-0.80 = fair; 0.60-0.70 = poor; 0.50-0.60 = fail.

3. RESULTS AND DISCUSSION

3.1. Model construction and evaluation

Models were generally poor at explaining the variance in the response data ($R^2 < 0.04$), even with all candidate explanatory variables used in the model derivation.

Nevertheless improving the possible utility of the models through backward selection was undertaken, the results of this analysis are shown in Table 6.

Variability explained by the final model selections was also poor ($R^2 < 0.02$), although there was some evidence that Motueka River *E. coli* concentrations are a very weak, but significant predictor of enterococci concentrations around the Kaiteriteri Beach region. In the case of the best unlagged southerly wind and the whole dataset models, modelled riverine *E. coli* data was the only statistically significant explanatory variable.

In the case of the northerly wind model the only significant explanatory variable was the strength of the lagged northerly wind component ($\log(v+1)$).

Table 6. Explanatory variable coefficients and their significance from the final model selections; Variance explained by models (R^2) is also displayed. Note that the models with the highest R^2 values are highlighted in bold. The symbols used to indicate significance are: ^o, < 0.1 , ^{*}, < 0.05 , ^{**}, < 0.01 , ^{***}, < 0.001 .

Explanatory variable/model	Lagged	log (bactiModel + 1)	log(flow + 1)	log(light + 1)	u	log (v + 1) ¹	Intercept	R^2
Northerly wind	Y					0.9301 ^o	-0.9839 ^{***}	0.013
Whole dataset	Y		0.1885			-	-1.2271 ^o	0.010
Northerly wind	N					0.273 ^o	-0.9598 ^{**}	0.010
Whole dataset	N	0.4882 ^{**}					-2.4752 ^{***}	0.019
Southerly wind	N ²	0.441 ^o					-2.3738 ^{**}	0.015

¹ Note that the logarithm of the northerly component of wind (v ; $\log(v+1)$) was not taken for the whole dataset analysis and the absolute value of v was used for the northerly wind scenario (when $v < 0$).

² Although a candidate lagged variables were presented to model selection process, the final model only included the unlagged BactiModel variable.

3.2. Performance analysis

Despite the small variance explained was observed for the logistic non-detection/detection model (low value for R^2), there was a possibility that the models may have had some value for higher concentration threshold management decisions. In order to assess the performance of the logistic model for this purpose, the area under the ROC curve (AUROC) was calculated (see Section 2.5) for 0, 30 and 140 cfu/100ml thresholds.

The results of this additional assessment confirm that the performance of all the models generated for this study was generally poor, although improvements are seen in the model performance for higher threshold FIB concentrations (Figures 9, 10 and 11). The best performing model (lagged northerly wind model) may be considered 'fair' under the traditional academic grade point system, however given the small number of samples exceeding this threshold (13 out of 351 samples), a high degree of confidence could not be attributed to this result. Additionally, the model performance is calculated based on the data used to build the models, so these results should be considered to represent the upper bound of model performance.

Independent construction of models based on a subset of the data and testing on remaining data would have been undertaken if the results of these initial models were better (*i.e.* at least 'good' on the ROC scoring scale). This additional analysis would not have produced any useful outcomes for the study, as it requires building the models based on a smaller set of data producing 'worse' models and therefore would be expected to have with similarly poor performance. Consequently this additional work was not undertaken with a relocation of time devoted to producing a useful model for the Motueka River environment (see Appendix 3).

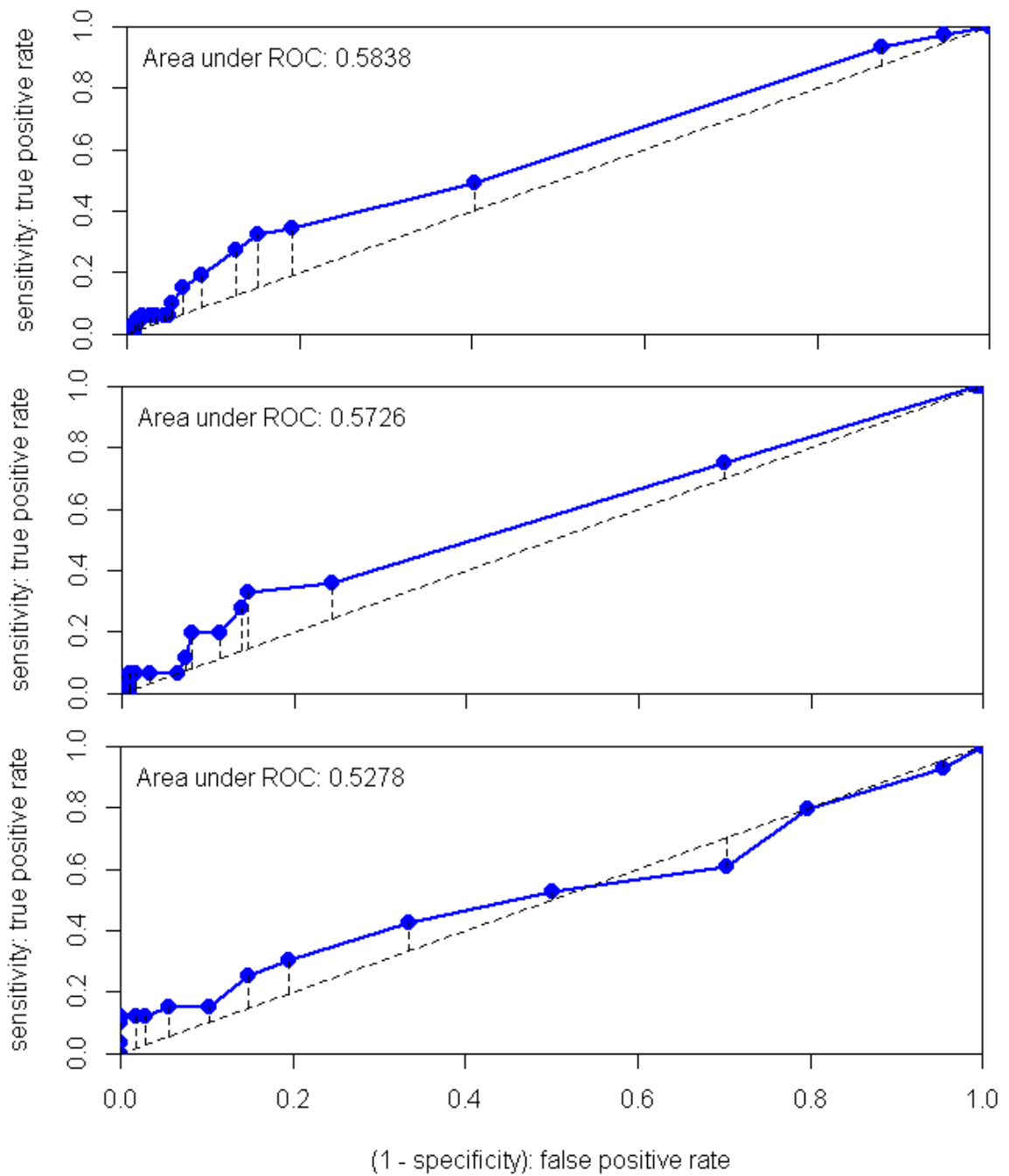


Figure 9. ROC curves for All Data (top), Northerly (middle) and Southerly (bottom), using the detection of enterococci (0 cfu/100ml) as a threshold.

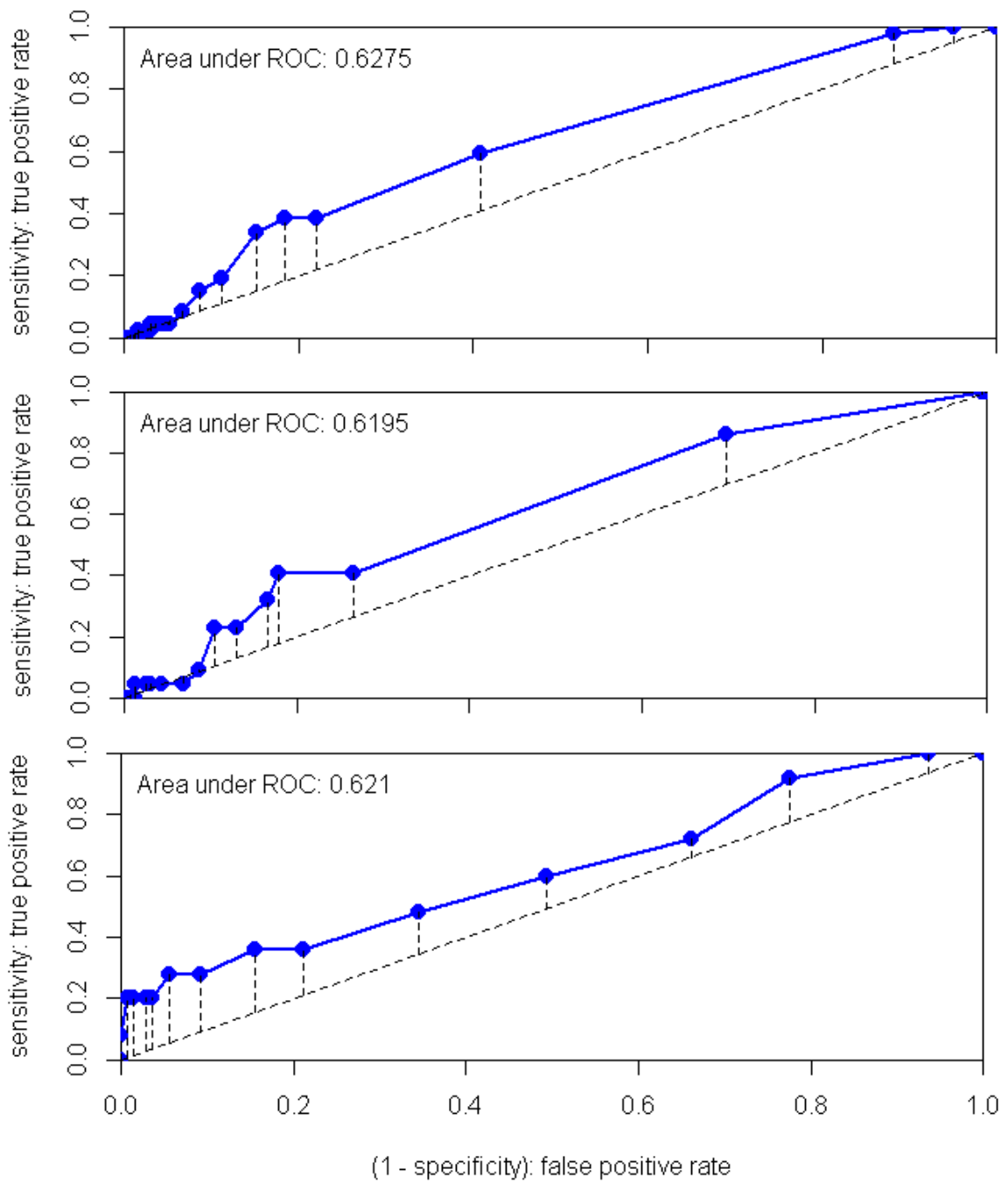


Figure 10. ROC curves for All Data (top), Northerly (middle) and Southerly (bottom), using 30 cfu/100ml as a threshold.

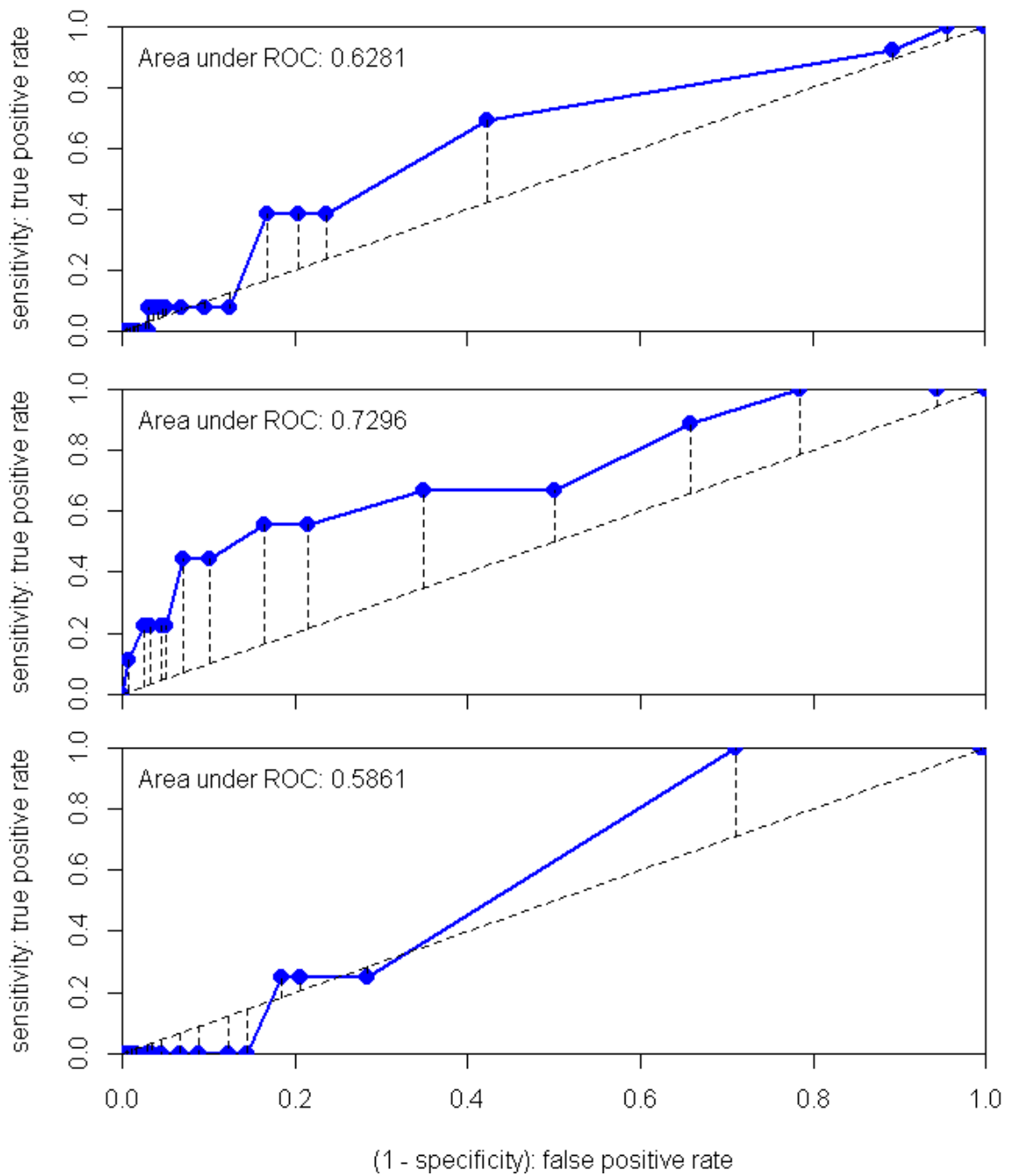


Figure 11. ROC curves for All Data (top), Northerly (middle) and Southerly (bottom), using the MfE Marine Guideline Standard of 140 cfu/100ml as a threshold.

4. CONCLUSIONS

This study has investigated the relationships between potential explanatory variables of enterococci concentrations measured at and around Kaiteriteri Beach. Although the study did not discover any strong relationships between the candidate explanatory variables and measured enterococci concentrations around Kaiteriteri Beach, it does show a weak, but significant effect of modelled *E. coli* concentrations from the Motueka River. Similarly, weak relationships with lagged northerly wind strength are also observed, suggesting waves may have an influence at the site under those conditions. However, the quality of the models is not sufficient to draw firm conclusions from this analysis.

The reasons for poor performance of the models are not entirely clear, but three possible reasons (and possible solutions for improvement) are summarised here:

1. The focus on sampling over the summer season means the dataset did not include many periods where high enterococci concentrations were observed (*i.e.* only 13 exceedences of the 140 enterococci /100mL threshold). Consequently, this bias may have influenced our attempts to build a robust model. Collection of more data during wet weather periods may provide better data for building robust models in future.
2. The Motueka River may not have been the major source of enterococci to the region. We cannot discount local sources of pollution (*e.g.* other animals or septic tank leakage) which may be important drivers of observed concentrations in the region. In order to assess the source of the pollution, microbial source tracking techniques to differentiate bacterial sources could be applied to samples with high enterococci concentrations to see if the source profiles were similar between the river and the beach.
3. The data-mining approach used in this project was not appropriate for the site. If the river is the main source of enterococci to the region, possibly a mechanistic approach may have been more successful as was undertaken for the river. This would involve tracking riverine sourced bacteria (as defined by the riverine model outputs) in a coastal hydrodynamic model to estimate their die-off, dilution and dispersal in the marine environment. As is seen in our application of the river model of Wilkinson (2008), using a logistic modelling approach can produce excellent results.

Although the primary aim of the study was not met, there are a number of useful outcomes from the study, namely:

1. The production of a 10 year synthesis of all the explanatory variables used to build the model, therefore facilitating any future efforts and possibly of further use in other studies undertaken by the Council.

2. Development of techniques that could be applied to generate similar classification tools for other rivers.
3. One application of these is the successful construction and testing of an 'excellent' Motueka River FIB water quality classification tool.

We applied a logistic modelling approach to the river model developed by Wilkinson (2008) as an additional task, to generate a Motueka River bacterial water quality classification tool (Appendix 3). A separate ROC analysis based on a simple logistic model incorporating only the river model estimates suggests the model is "excellent" at predicting MfE guideline exceedances for freshwater (see Appendix 3 for details). This raises the possibility that the model could be used in a similar manner to other models that are in operational use (e.g. Philly RiverCast model <http://www.phillyrivercast.org/>).

The Philly RiverCast web page states that:

"65% the time the RiverCast prediction was accurate. 35% of the time the prediction was conservative (higher bacteria levels were predicted than measured). There were no examples of predicted levels lower than the measured levels."

Our initial assessment suggests that the Motueka River model is better than the PhillyCast predictions (*i.e.* >80% true positive with a <10% false positive; Appendix 3 - Figure 3.1) – therefore this model is considered as a potential tool to provide information on the likely FIB water quality classification for the river.

Bacterial water quality for the Motueka River information could prove to be a very useful tool for providing information for river users (e.g. kayakers, rafters and swimmers) and shellfish gatherers using the mud flat regions close to the river to make informed choices about the degree of health risk they expose themselves to. Additionally, this information may also help council staff with planning river sampling over representative periods.

Our study highlights the difficulties involved in determining for explanatory variables of FIB in the coastal environment without an underlying mechanistic framework to support statistical model construction and the importance of unbiased and suitably stratified sampling data for producing these models. Efforts could be made to improve the utility of datasets in order to relate terrestrial explanatory variables to coastal FIB concentrations. One method is to increase sampling during and after heavy rainfalls, but it would need to be balanced against the costs and the existing needs for the operational use of this data (*i.e.* to protect human health at these sites).

In summary, although the initial objectives of this study have not been met, we suggest that the suite of outcomes from the project will be of direct relevance to the TDC and other councils. In particular the Motueka River tool has been a successful

outcome of the project, and datasets and techniques developed for this study will prove useful for other applications.

5. ACKNOWLEDGEMENTS

The Ministry for Science and Innovation are thanked for the provision of funds to undertake this work under the Medium Advice Grant scheme for the EnviroLink Programme (Regional Council Advice number: 1008-TSDC79).

Trevor James at the TDC is thanked for his timely provision of council bacterial data for the region which formed the basis of the modelling work in this study.

Jeremy Wilkinson is acknowledged for the provision of an excel version of his Motueka River Model which forms the basis of the R adaptation provided in Appendix 2 of this report.

Chris Cornelisen is acknowledged for his efforts in reviewing this work and Cherie Johansson for editing this report.

6. REFERENCES

- Akaike H 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716-723.
- Blanchard JL, Dulvy NK, Jennings S, Ellis JR, Pinnegar JK, Tidd A, Kell LT 2005. Do climate and fishing influence size-based indicators of Celtic Sea fish community structure? *ICES Journal of Marine Science: Journal du Conseil* 62 (3): 405-411.
- Boehm A, Grant S, Kim J, Mowbray S, McGee C, Clark C, Foley D, Wellman D 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environmental Science & Technology* 36 (18): 3885-3892.
- Boehm AB, Weisberg SB 2005. Tidal forcing of enterococci at marine recreational beaches at fortnightly and semidiurnal frequencies. *Environmental Science & Technology* 39 (15): 5575-5583.
- Bordalo AA 2003. Microbiological water quality in urban coastal beaches: the influence of water dynamics and optimization of the sampling strategy. *Water Research* 37 (13): 3233-3241.
- Brion GM; Lingireddy S 2003 Artificial neural network modeling: a summary of successful applications relative to microbial water quality. *Water Sci. Tech.* 47(3), 235–240.
- Cornelisen CD, PA Gillespie, M Kirs, RG Young, RW Forrest, PJ Barter, BR Knight, VJ Harwood. 2011a. Motueka River plume facilitates transport of ruminant faecal contaminants into shellfish growing waters, Tasman Bay, New Zealand. 45:477-495.
- Cornelisen CD, Jiang W, Griffiths R. 2011b. Interpreting Northland's coastal water quality monitoring results under different tide conditions. Prepared for Northland Regional Council. Cawthron Report No. 2026. 33 p. plus appendices
- Christensen VG, Jian X, Ziegler A 2001. Characterization of surface-water quality based on real-time monitoring and regression analysis, Quivira National Wildlife Refuge, south-central Kansas, December 1998 through June 2001, US Geol. Survey Water Resources Investigations Report 00-4126.
- Diane MLM, Ahlfeld DP 2007. Comparing artificial neural networks and regression models for predicting faecal coliform concentrations. *Hydrological Sciences Journal* 52 (4): 713-731.
- Eleria A, Vogel RM 2005. Predicting Fecal Coliform Bacteria Levels in the Charles River, Massachusetts, USA. *JAWRA Journal of the American Water Resources Association* 41 (5): 1195-1209.
- Fox, J. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, Thousand Oaks, CA. xxi C 597 pp.

- Gillespie PA, Forrest RW, Peake BM, Basher LR, Clement DM, Dunmore RA, Hicks DM 2011. Spatial delineation of the depositional footprint of the Motueka River outwelling plume in Tasman Bay, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 45 (3): 455-475.
- Ishi S, MJ Sadowsky 2008 *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes and Environments* 23:101-108
- Maimone M, Crockett CS, Cesanek WE 2007. PhillyRiverCast: A real-time bacteria forecasting model and Web application for the Schuylkill River. *Journal of Water Resources Planning and Management* 133: 542.
- Masopust P 2005. High-Resolution Spatial and Temporal Variability and Patterns of *Escherichia coli* in the Charles River. Unpublished M.Sc. Thesis, Northeastern University, Boston, 133 p.
- MfE (Ministry for the Environment) 2003. Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Area, ME Number: 474 159 p.
- NIWA 2011, Cliflo - NIWA National Climate Database Tool, <http://cliflo.niwa.co.nz/> (accessed November 2011).
- Noble RT, Weisberg SB 2005. A review of technologies for rapid detection of bacteria in recreational waters. *Journal of water and health* 3 (4): 381-392.
- Olyphant GA, Whitman RL 2004. Elements of a predictive model for determining beach closures on a real time basis: the case of 63rd Street Beach Chicago. *Environmental monitoring and assessment* 98 (1): 175-190.
- RDC Team (R Development Core Team) 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenfeld LK, McGee CD, Robertson GL, Noble MA, Jones BH 2006. Temporal and spatial variability of fecal indicator bacteria in the surf zone off Huntington Beach, CA. *Marine Environmental Research* 61 (5): 471-493.
- Sanders BF, Arega F, Sutula M 2005. Modeling the dry-weather tidal cycling of fecal indicator bacteria in surface waters of an intertidal wetland. *Water Research* 39 (14): 3394-3408.
- Schielzeth H 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* 1 (2): 103-113.
- Steets B, Holden P 2003. A mechanistic model of runoff-associated fecal coliform fate and transport through a coastal lagoon. *Water Research* 37 (3): 589-608.
- Swets JA 1988. Measuring the accuracy of diagnostic systems. *Science* 240 (4857): 1285.
- Tufail M, Ormsbee L, Teegavarapu R 2008. Artificial Intelligence-Based Inductive Models for Prediction and Classification of Fecal Coliform in Surface Waters. *Journal of Environmental Engineering* 134: 789.

- Thuiller W, Araújo MB, Lavorel S 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14 (5): 669-680.
- Whitman RL, Nevers MB 2004. *Escherichia coli* sampling reliability at a frequently closed Chicago Beach: monitoring and management implications. *Environmental Science & Technology* 38 (16): 4241-4246.
- Wilkinson J, Jenkins A, Wyer M, Kay D 1995. Modelling faecal coliform dynamics in streams and rivers. *Water Research* 29 (3): 847-855.
- Wilkinson J 2008. Faecal indicator organism modelling (FIB): application to Motueka River. Prepared for Integrated Catchment Management project. Cawthron Report No. 1454.
- Wilkinson RJ, McKergow LA, Davies-Colley RJ, Ballantine DJ, Young RG 2011. Modelling storm-event *E. coli* pulses from the Motueka and Sherry Rivers in the South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research* 45 (3): 369-393.
- Yamahara KM, Layton BA, Santoro AE, Boehm AB 2007. Beach sands along the California coast are diffuse sources of fecal bacteria to coastal waters. *Environmental Science & Technology* 41 (13): 4515-4521.
- Zuur AF, Ieno EN, Elphick CS 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1 (1): 3-14.

7. APPENDICES

Appendix 1. The Motueka River *E. coli* Model validation.

Model development based on single point model catchment of Wilkinson (2008) developed under research contract CO9X0014 for the New Zealand Foundation for Research Science and Technology (FRST). Although the report of Wilkinson (2008) includes the main details and methodology for the model, the model has been updated to include a new light attenuation model. This new model accounts for effects of dissolved yellow matter (yBD) and CDOM (k390) based on discussions with Rob Davies-Collie (pers. comm. 10/12/2009) and uses the approach taken by Smith et al. (1997). The analysis on yBD and k390 using Motueka River data from Woodstock yielded the following relationships.

$$yBD = -5.1476 \times \ln(Q) + 11.7575 \quad R^2=0.6423 \quad (1)$$

$$k390 = 0.7294 \times Q^{0.57707} \quad R^2=0.6436 \quad (2)$$

Where Q is the river flow. The total light attenuation coefficient (k) was estimated by:

$$k = 10^{(-0.5034 - 0.0649 \times \log_{10}(yBD) + 0.2145 \times \log_{10}(k390))} \quad (3)$$

It was assumed that these relationships also held for the lower reaches of the river (e.g. Woodsmans Bend). This assumption appears justified given the good performance of the model at Woodsmans Bend, with an adjusted R-squared of 0.6498 (Table 1.1).

Table 1.1 Linear model parameters and performance statistics of river model results to observations comparisons at Woodsmans Bend from the period 15 March 2001 to 30 June 2004 (N=272).

```
Call:
lm(formula = Log10Obs ~ Log10Model, data = calData)

Residuals:
    Min       1Q   Median       3Q      Max
-1.18633 -0.24306 -0.02053  0.22109  1.38303

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.41342    0.09932   4.163 4.24e-05 ***
Log10Model   0.75050    0.03350  22.405 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3776 on 269 degrees of freedom
Multiple R-squared:  0.6511,    Adjusted R-squared:  0.6498
F-statistic:  502 on 1 and 269 DF,  p-value: < 2.2e-16
```

The model does have a tendency to underestimate observed *E. coli* concentrations during peak events (Figure 1.1, slope = 0.75). Very little data was available to check the model performance against low concentration periods (*i.e.* low flows), although they too appear to be underestimated by the model (Figure 1.1). Nevertheless, the model appears to be good at predicting the timing and magnitude of flood events (Figure 1.2).

Given the performance of the model and its ability to estimate *E. coli* concentrations in the Motueka River, it appears to be a useful candidate variable for use in this study. Although not the focus of this study, it may also prove to be a useful tool for communicating bacterial contamination to river users in periods where sampling is not undertaken and is provided in Appendix 2 for future improvement and modification as required. Coupling of this model to a logistic classification of MfE freshwater FIB guidelines has also been undertaken, and suggests that this model has some value in predicting water quality classifications (see Appendix 3).

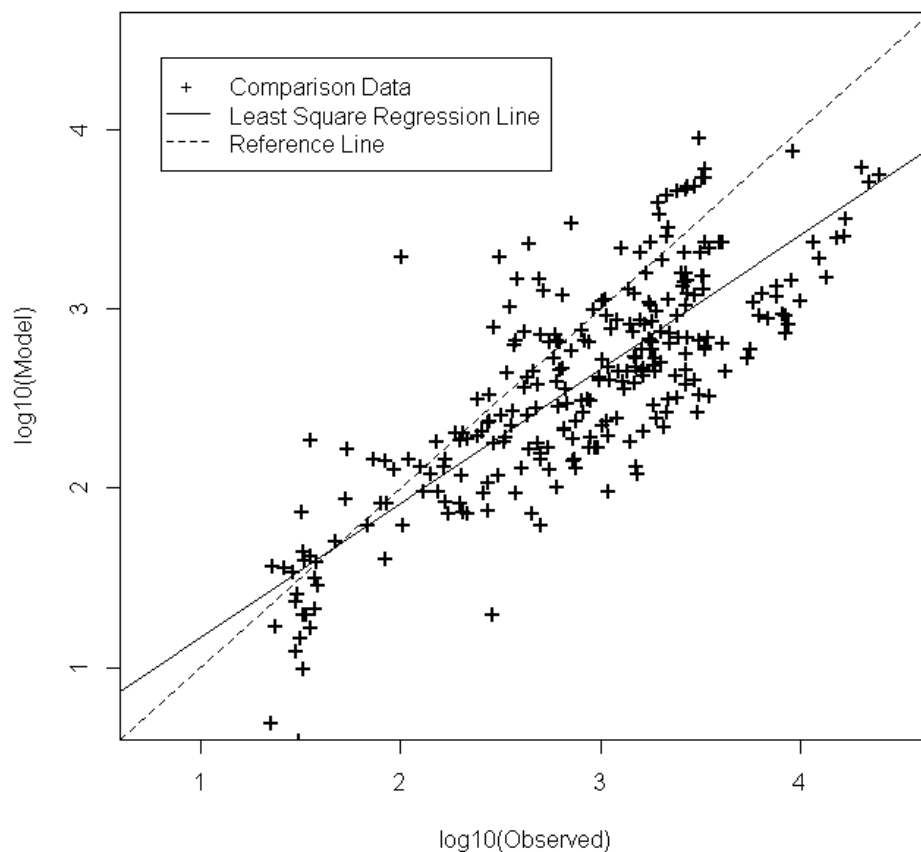


Figure 1.1 Comparison of river model results to measurements at Woodsman's Bend over the period 15 March 2001 to 30 June 2004 (N=272).

Appendix 2. R version of the Motueka River *E. coli* Model

```

motRiverFIBmodelFn<-function(flowIn, lightIn, tempIn, initBacti=NULL,
initChannelStore=NULL, Ni=NULL){
# R implementation of Motueka River Model developed by RJ Wilkinson (2008)
# Reference: Wilkinson J 2008. Faecal indicator organism modelling
# (FIB): application to Motueka River.
# Prepared for Integrated Catchment Management project.
# Cawthron Report No. 1454.
#
# Further improvements to the model have been undertaken and are described in:
# Wilkinson RJ, McKergow LA, Davies-Colley RJ, Ballantine DJ, Young RG 2011
#Modelling storm-event E. coli pulses from the Motueka and Sherry Rivers in the
#South Island, New Zealand. New Zealand Journal of Marine and Freshwater
#Research 45 (3): 369-393.
#
# Adapted for use in R by Weimin Jiang and Ben Knight 2011
#
# BRK: Note updated to new light shading model added to account for yellow matter
# (yBD) and CDOM (k390) as suggested by RD Collie (10/12/2009)
#
# yBD=-5.1476*ln(Q)+11.7575 (and =0 if yBD<0) r2=0.6423
# k390=0.7294*Q^0.57707, r2=0.6436
#
# see e.g. ref:
# Smith DG, Davies-Colley RJ, Knoef J, Slot GWJ 1997.
# Optical characteristics of New Zealand rivers in relation to flow.
# JAWRA Journal of the American Water Resources Association 33 (2): 301-312.

if(is.null(initBacti)) initBacti<-100 #intial bacterial concentration
if(is.null(initChannelStore)) initChannelStore<-6e9 #initial channel store for the
riverine bacteria
if(is.null(Ni)) Ni=3250000000 #model bacti input... per second

#input data static
riverlength=20000 #m
maxQ=1e3 #max river flow (m3)
dt=3600 #secs
bedSplit=0.945 #fraction of bacti going to settle in bed vs. water column
ErosionRate=0.02
smoothVal=0.91 #smoothing param for output
dieOff.mult=1 #adjustable dieoff multiplier
dieOff.TempConsts=c(0.1068, 0.0709)
dieOff.LightConsts=c(0.0089, 1.011)

```

```

#new light attn model data
light.CDOMConst=c(0.57707, 0.7294)
light.clarConst=c(-5.1476, 11.7575)
light.attnConsts=c(-0.5034, 0.2145, -0.0649)

## start model
ChannelStore=initChannelStore
#Bacti=Q(:)*0;
Bacti<-as.vector(flowIn)*0
toSuspension=Bacti
chanScour=Bacti
Bacti[1]=initBacti

for(it in 2:length(flowIn)){
  # transport vars
  #calc derived vars for the mot river
  vel=0.1061*flowIn[it]^0.5477
  traveltime=riverlength/(vel*3600) #time in hours
  crosssectionArea=12.577*flowIn[it]^0.3256 #m2
  wettedVol=crosssectionArea*riverlength #m3
  depth=0.37*flowIn[it]^0.1813 #m

  #calc clarity and attenuation
  clarity=light.clarConst[1]*log(flowIn[it])+light.clarConst[2]
  clarity[clarity<0.05]<-0.05 #minimum clarity
  CDOM=light.CDOMConst[2]*flowIn[it]^light.CDOMConst[1]

  lightAttn=10^(light.attnConsts[3]+light.attnConsts[1]*log10(clarity)+light.attnConsts[2]*
    log10(CDOM))

  #calc total dieoff from light + temp dieoff for surf, depth avg and bottom
  tmp=dieOff.TempConsts[1]*tempIn[it]^(dieOff.TempConsts[2])/24
  dieOff.surf=dieOff.LightConsts[1]*lightIn[it]^(dieOff.LightConsts[2])/24+tmp
  depAvgLight=(lightIn[it]/(lightAttn*depth))*(1-exp(-lightAttn*depth))
  dieOff.depAvg=dieOff.LightConsts[1]*(depAvgLight)^(dieOff.LightConsts[2])/24+tmp
  dieOff.bottom=dieOff.LightConsts[1]*(lightIn[it]*exp(-
    lightAttn*depth))^(dieOff.LightConsts[2])/24+tmp
  if (flowIn[it]>flowIn[it-1]){
    deltaQi=(flowIn[it]-flowIn[it-1])/maxQ
  } else{
    deltaQi=0
  }

  #Calc Channel store

```

```
channelDieoff=dieOff.bottom*dieOff.mult
Nin=(Ni*dt*(1-dieOff.surf))
Erosion=ErosionRate*ChannelStore*deltaQi
Erosion[Erosion>ChannelStore]<-ChannelStore #can't erode more than in channel
ChannelStore=ChannelStore+Nin*bedSplit*(1-log10(flowIn[it])/log10(maxQ))-
              ChannelStore*channelDieoff-Erosion
ChannelStore[ChannelStore<0]<-10 #keep min number in channel

#calc bacti concs
chanScour=flowIn[it]*Erosion/(wettedVol*1e4)
toSuspension=Nin*(1-bedSplit)*log10(flowIn[it])/log10(maxQ)/(wettedVol*1e4)
Bacti[it]=(smoothVal-dieOff.depAvg)*Bacti[it-1]+(toSuspension+chanScour)*(1-
              smoothVal)
if(Bacti[it]<0) Bacti[it]<-0
}
return(list(Bacti=Bacti,ChannelStore=ChannelStore))
}
```


Appendix 3. Performance of the FIB River Model for operational use.

Application of the mechanistic river model as a tool for predicting riverine *E. coli* concentrations was undertaken to see the mechanistic model output may be a useful tool for providing information on the water quality with respect to Ministry for the Environment (MfE) Microbiological Assessment Category (MAC) definitions (MfE, 2003; Table 3.1). The same process for testing the performance of the river model data was used as was described for the beach data (see section 2 of this report) assuming the application of a logistic GLM model, fitted to a model of the form:

$$Y = mX + C$$

Where Y is a binary form of the logged measured data, calculated on the basis of the MAC definitions (Table 3.1), X is the modelled data, m is the slope and C is the intercept.

Table 3.1. Microbiological Assessment Category (MAC) definitions (reproduced from MfE 2003).

MAC	Guideline Definition
A	Sample 95 percentile > 130 <i>Escherichia coli</i> per 100 mL
B	Sample 95 percentile 131 - 260 <i>Escherichia coli</i> per 100 mL
C	Sample 95 percentile 261 - 550 <i>Escherichia coli</i> per 100 mL
D	Sample 95 percentile >550 <i>Escherichia coli</i> per 100 mL.

The coefficients from the fitting process undertaken in R are shown in Table 3.2 below. The results from the new model were then compared to the data they were built on to get an initial idea of the model performance using the ROC analysis (see Section 2.5 of this report).

Table 3.2. Results of logistic model coefficients (slope and intercept) for differing MAC threshold definitions.

MAC Model	Slope/intercept	Estimate	Std. Error	z value	Pr(> z)
B	Intercept	-8.7028	1.2998	-6.696	2.15E-11
	Slope	3.9416	0.5244	7.517	5.60E-14
C	Intercept	-11.2208	1.4707	-7.629	2.36E-14
	Slope	4.1098	0.5115	8.034	9.40E-16
D	Intercept	-9.7228	1.2782	-7.606	2.82E-14
	Slope	3.0982	0.4066	7.621	2.52E-14

The result of the ROC analysis show that the models have the potential to provide 'excellent' prediction of MACs for the river at Woodmans Bend (Figure 3.1) based on the traditional academic point scale for assessing the Area under an ROC curve. Other similar approaches are in operational use (e.g. Philly RiverCast <http://www.phillyrivercast.org/>). The Philly RiverCast webpage states that: *"65% the time the RiverCast prediction was accurate. 35% of the time the prediction was conservative (higher bacteria levels were predicted than measured). There were no examples of predicted levels lower than the measured levels."*

Given our assessment suggests this model has a >80% true positive rate with a <10% false positive rate (Figure 3.1), on this basis we recommend this model is considered as a potential tool to provide information on the likely river quality classification in the region.

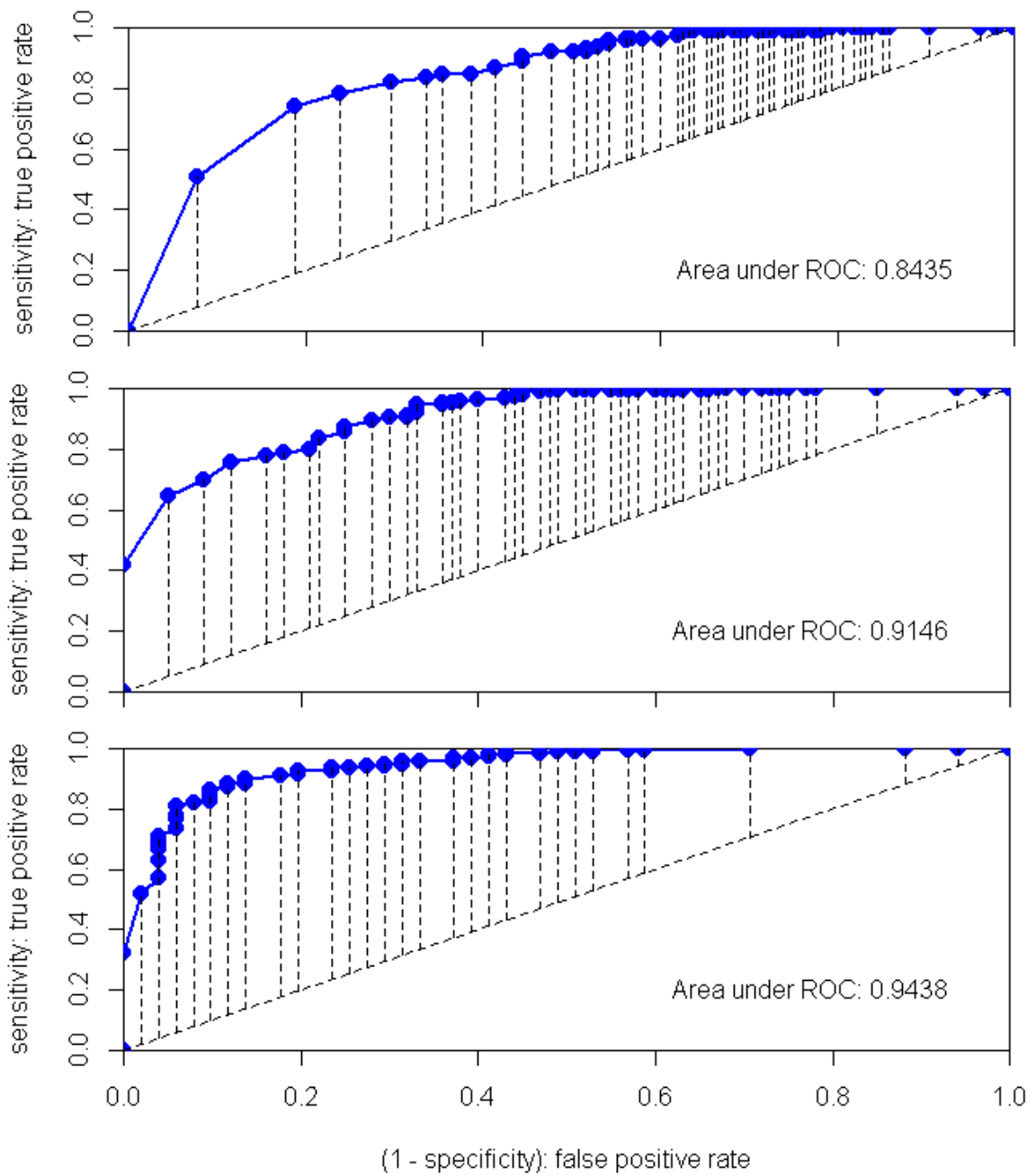


Figure 3.1. Receiver operator curves for logistic model based on FIB river model compared to all observed data at Woodmans Bend for MfE MAC definitions: B (>131 *E. coli*/100mL, bottom), C (>261 *E. coli*/100mL, middle), D (>550 *E. coli*/100mL, top).