Predicting 'very poor' beach water quality gradings using classification tree

Wai Thoe, King Wah Choi and Joseph Hun-wei Lee

ABSTRACT

A beach water quality prediction system has been developed in Hong Kong using multiple linear regression (MLR) models. However, linear models are found to be weak at capturing the infrequent 'very poor' water quality occasions when *Escherichia coli* (*E. coli*) concentration exceeds 610 counts/100 mL. This study uses a classification tree to increase the accuracy in predicting the 'very poor' water quality events at three Hong Kong beaches affected either by non-point source or point source pollution. Binary-output classification trees (to predict whether *E. coli* concentration exceeds 610 counts/100 mL) are developed over the periods before and after the implementation of the Harbour Area Treatment Scheme, when systematic changes in water quality were observed. Results show that classification trees can capture more 'very poor' events in both periods when compared to the corresponding linear models, with an increase in correct positives by an average of 20%. Classification trees are also developed at two beaches to predict the four-category Beach Water Quality Indices. They perform worse than the binary tree and give excessive false alarms of 'very poor' events. Finally, a combined modelling approach using both MLR model and classification tree is proposed to enhance the beach water quality prediction system for Hong Kong.

Key words | beach management, classification tree, coastal beach, very poor water quality, water quality prediction

Wai Thoe (corresponding author)
Department of Civil and Environmental
   Engineering, Environmental and Water Studies,
Stanford University,
Stanford,
CA 94305,
USA
E-mail: awthoe@stanford.edu

King Wah Choi
Joseph Hun-wei Lee
Department of Civil and Environmental
   Engineering,
Hong Kong University of Science and Technology,
Clear Water Bay,
Hong Kong

INTRODUCTION

To protect bathers from swimming in faecally polluted waters, bathing beaches are typically monitored for bacterial concentrations in water samples. However, bacterial concentrations change rapidly with environmental conditions in a matter of hours (Boehm *et al.* 2002). The traditional beach monitoring method, which collects water samples mostly on a weekly basis and takes 18–24 hours to measure the concentration, may result in outdated beach advisories (Lee *et al.* 2008). Predictive modelling becomes an alternative approach to issue real-time beach advisories based on the most recent environmental conditions, and is endorsed by the US Environmental Protection Agency as a rapid and inexpensive tool to assist beach management (USEPA 2011). The Ohio nowcast system for beaches in the Great Lakes, USA is one of the successful predictive systems

(Francy *et al.* 2013). As from 2014, daily nowcasts of water quality at nine beaches are available online (http://www.ohionowcast.info/), and the system is potentially extending to a total of 49 beaches if their corresponding models obtain sensitivity and specificity greater than 50% and 85%, respectively, in an independent validation year.

In Hong Kong, a beach water quality prediction system has been developed using the multiple linear regression (MLR) model (Thoe *et al.* 2012). Beach Water Quality Index (BWQI) of 1 to 4, each associated with a health risk from negligible to high, is issued based on the MLR predicted *Escherichia coli* (*E. coli*) concentration (Table 1). The thresholds of *E. coli* concentrations for different BWQI and the associated health risks are obtained from a local epidemiological study conducted in the late 1980s (Cheung *et al.* 1990). Since August 2011,

Table 1 │ Beach grading system in Hong Kong and the corresponding Beach Water Quality Index (BWQI). The thresholds of *E. coli* concentrations for different BWQI and the associated health risks are obtained from a local epidemiological study conducted in the late 1980s (Cheung *et al.* 1990)

Grade/BWQI*	Beach water quality	*E. coli* count per 100 mL	Minor illness rates** (cases per 1,000 swimmers)	Water quality objective violation
1	Good	≤24	Undetectable	No
2	Fair	25–180	≤10	No
3	Poor	181–610	11–15	Yes
4	Very poor	>610	>15	Yes (beach closure)

*BWQI = Beach Water Quality Index.
**Skin and gastro-intestinal illnesses.

daily BWQI of 16 representative marine beaches are disseminated through the Project WATERMAN webpage (http://www.waterman.hku.hk/beach). Through a real-time validation in 2010–2012, the system has been proved to outperform the traditional beach monitoring method in capturing exceedances of water quality objectives (WQO = 180 counts/100 mL, corresponding to BWQI-3 or 4) (Thoe & Lee 2014). Swimming-associated illnesses (e.g., gastro-intestinal diseases) can be reduced. However, it is found that MLR models are generally weak at predicting pollution events reaching 'very poor' grading (>610 counts/100 mL, BWQI-4). Even for relatively polluted beaches such as Big Wave Bay and Silvermine Bay (non-point source dominated beaches), 'very poor' grading is rarely reached and comprises only <10% of all events. Although models of these beaches can explain a great portion of *E. coli* concentration variance (adjusted R^2 = 0.4–0.5), model accuracy in predicting 'very poor' events are lower than predicting 'WQO exceedance'. Also, for beaches where pollution comes mainly from the complicated sources at their catchment (e.g., luxurious houses and foul sewers), their MLR models can only explain a small proportion of *E. coli* variance; adjusted R^2 for the model of Silverstrand, a typical beach of this type, is only about 0.2, significantly lower than other beaches in Hong Kong (0.3–0.5). These models are unable to capture the observed variation of *E. coli* concentration, failing to predict both very clean and very poor water quality.

According to the WQO, a beach with a 'very poor' water quality grading (corresponds to a high risk of contracting waterborne diseases: >15 cases per 1,000 swimmers) should be immediately closed for public use, and its water quality should be continuously monitored until it is safe for swimming (*E. coli* concentration <610 counts/100 mL). Predictive

models sensitive to these 'very poor' events facilitate early public notification of potential beach closure and remedial actions to the pollution events. This study uses a categorical model – the classification tree (CT), to enhance the accuracy to capture 'very poor' water quality events. A few studies applied regression tree to predict average bacterial concentrations (Parkhurst *et al.* 2005; Boehm *et al.* 2007; Bae *et al.* 2010), while CT, which gives discrete categorical outputs, has received relatively less attention. An actual application of CT can be found at coastal beaches in Scotland (Stidson *et al.* 2012), where the beaches are mainly affected by local freshwater inputs. CT has also been applied to 25 coastal beaches in California, and is found to generally outperform other continuous models, including MLR models, in capturing beach postings due to faecal contamination (Thoe *et al.* 2014, 2015).

In this study, binary-output CT models are developed to predict 'very poor' water quality grading (BWQI-4) at two Hong Kong beaches mainly affected by non-point source pollution (Big Wave Bay and Silvermine Bay) and one beach mainly affected by the complicated pollution sources from its catchment (Silverstrand). Model performances are compared with the MLR models developed by Thoe & Lee (2014). The modelling period covers both before and after the implementation of the Harbour Area Treatment Scheme (HATS), when systematic changes in coastal water quality were observed. As a comparison, CT is also used to predict multiple categorical outputs (BWQI-1 to 4) at Big Wave Bay and Silverstrand after HATS implementation. A new modelling approach to better capture 'very poor' pollution events using both CT and MLR is proposed. It is the first time a beach water quality prediction system is enhanced to better predict a specific category of interest using CT.

METHODS

Study beach

CT models are developed for three beaches in Hong Kong (Figure 1). Big Wave Bay (BW) and Silvermine Bay (SIL) are non-point source dominated beaches in Hong Kong Island and Lantau Island, respectively. Both beaches receive freshwater from a natural stream flowing through an unsewered catchment, and are facing relatively unpolluted ocean waters (Tathong Channel and West Lamma Channel, respectively). Silverstrand (SS) is a beach in Sai Kung district facing the enclosed Port Shelter; the beach is mainly affected by the local point source pollution from its catchment with luxurious houses, restaurants and foul sewers.

Data and study period

E. coli concentrations at the study beaches in the years 1990–2009 with sampling intervals of 3–14 days (median

7 days) are used as the dependent variable for model development. The *E. coli* data are collected by the Environmental Protection Department (EPD) of Hong Kong Special Administrative Region Government under the beach monitoring programme. The independent variables used in this study are the same as those used to develop the MLR models (Thoe & Lee 2014): daily rainfall in the past 3 calendar-days (day-1, day-2 and day-3 rain), tide level during the typical time when water samples are collected (around 10:00 am), previous day's solar radiation and wind speed, *in-situ* measured water temperature and salinity, and geometric mean of *E. coli* concentration in the past five water samples in natural logarithm (lnEC5). Considering the possible impact on water quality from the most recent rain, a new input variable, 9-hour rain is also used. 9-hour rain is defined as the total rain that occurs in the first 9 hours starting from 0:00 (midnight) on the day of prediction. 9-hour is chosen because the typical sampling time is about 10–11 am for most of the beaches in Hong Kong.

Figure 1 │ Locations of Big Wave Bay, Silvermine Bay and Silverstrand.

The models are calibrated and validated for two different periods. First, they are calibrated with data in the years 1990–1997, and validated against data in the years 1998–2001. This period is chosen because water quality was relatively stable and reflected conditions before the implementation of the Harbour Area Treatment Scheme (HATS) in 2002 (pre-HATS period). HATS is a major environmental infrastructure in Hong Kong to collect sewage from both sides of Victoria Harbour to the centralized Stonecutters Island sewage treatment works for chemically enhanced primary treatment. Second, the models are developed with data in the years 2002–2006, and validated against daily data obtained during two beach water quality surveys conducted in June–July 2007 (Big Wave Bay only, $n = 60$) and July–October 2008 (all three beaches, $n = 89$) (post-HATS period). Details of the surveys can be found in Thoe (2010). During the post-HATS period, water quality is improved at most of the beaches. The post-HATS period is studied to investigate if the models can (1) continue to capture the reduced number of exceedances and (2) be applied on a daily basis by validating against an independent set of daily data.

Development of classification tree

CTs predicting both binary outputs and four-category outputs are developed. A binary-dependent variable is developed to calibrate the binary CT: '1' when the observed *E. coli* concentration is higher than 610 counts/100 mL ('very poor' events), and '0' otherwise. For the four-category CT, the BWQI (1–4) based on the observed *E. coli* concentration in Table 1 is used as the dependent variable.

CT is developed using the 'Classification Tree' package in MATLAB (version R2012b, Natick, MA). CT is a non-parametric and non-linear model to predict categorical outputs. Details of CT can be found in Breiman *et al.* (1983). Input variables for CT can be either continuous or categorical. CT starts with a parent node containing all observations; binary branching is carried out to yield two leaves based on an independent variable criterion, such that more observations in the same output category are grouped in one leaf. The leaves then become parent nodes for further branching in the next level. Branching from a parent node is based on the Gini's Diversity Index (GDI), a measure of node impurity (Jost 2006):

$$GDI = 1 - \sum_i p^2(i) \tag{1}$$

where p is the observed fraction of categories with category i that reach the node. For example, if all cases in a node are in the same category ($p = 1$), $GDI = 0$. Branching from a parent node will occur if a lower GDI can be obtained in the next level. As all input variables are continuous in this study, the optimal branching is obtained after comparing all input variables, and splitting halfway between any two adjacent unique values based on two major CT model parameters:

1. Minimum number of observations in a parent node (MP). If a parent node has observations less than MP, branching will be terminated.
2. Minimum number of observations in a leaf (ML). Branching will not occur if the number of observations that goes to one leaf is less than ML.

Values of MP and ML are determined through a sensitivity analysis: different CTs are developed based on a combination of different MP and ML values. The ranges of MP and ML tested are 10–50 (with an interval of 5) and 1–8, respectively.

The performances of the CTs are evaluated based on the following assessment criteria: (i) the percentage of observed *E. coli* concentration >610 counts/100 mL ('very poor' water quality) that is actually predicted (correct positive); (ii) the percentage of observed *E. coli* concentration <610 counts/100 mL that is actually predicted (correct negative); (iii) the percentage of predicted exceedance/non-exceedance that is actually observed (predictive value ±); and (iv) the percentage of correct prediction (overall accuracy). The best model that obtains the highest correct positive and overall accuracy in both calibration and validation periods is selected. CT model performances are compared with the results obtained by the MLR models developed by Thoe *et al.* (2012).

As an exploratory study to evaluate CT model's ability to capture 'very poor' gradings when compared to MLR model, for simplicity, all CT and MLR model predictions are conducted in 'hindcast' application, i.e., real-time availability of the input variables is not considered.

RESULTS

Frequency of 'very poor' gradings in the pre-HATS and post-HATS periods

Table 2 shows the frequency of the three study beaches reaching 'very poor' gradings based on EPD data in the pre-HATS (1990–2001) and post-HATS (2002–2009) periods. In the pre-HATS period, 'very poor' grading comprises 13–18% of all events, and is dropped to 6–8% in the post-HATS period. Among the beaches, Silvermine Bay has the highest frequency of 'very poor' grading (18 and

Table 2 │ Percentage of 'very poor' water quality gradings at the three study beaches in pre-HATS and post-HATS periods, according to EPD monitoring data

	Pre-HATS 1990–2001	Post-HATS 2002–2009
Big Wave Bay	15.6%	7.3%
Silvermine Bay	18.1%	7.6%
Silverstrand	13.0%	6.3%

8% pre- and post-HATS), while Silverstrand has the lowest (13 and 6% pre- and post-HATS).

Model performances: pre-HATS period

Figure 2 shows the performance tables for the binary CT and MLR models at the three beaches in the pre-HATS period during calibration (1990–1997) and validation (1998–2001). During calibration, CT gives a correct positive of 49% and 66%, respectively, at Big Wave Bay and Silvermine Bay, as compared to 23 and 34% obtained by MLR models. The advantage of CT is maintained in validation, with a correct positive of 37% and 64% for Big Wave Bay and Silvermine Bay, respectively, when MLR models achieve 15% and 29%, respectively. Correct negatives obtained by different models for the two beaches are similar and are consistently higher than 85%. CT for Silverstrand also gives higher correct positive (47% in calibration and 56% in validation) than the MLR model (29 and 31%). Correct negative for CT in the validation period is 92%, slightly lower than
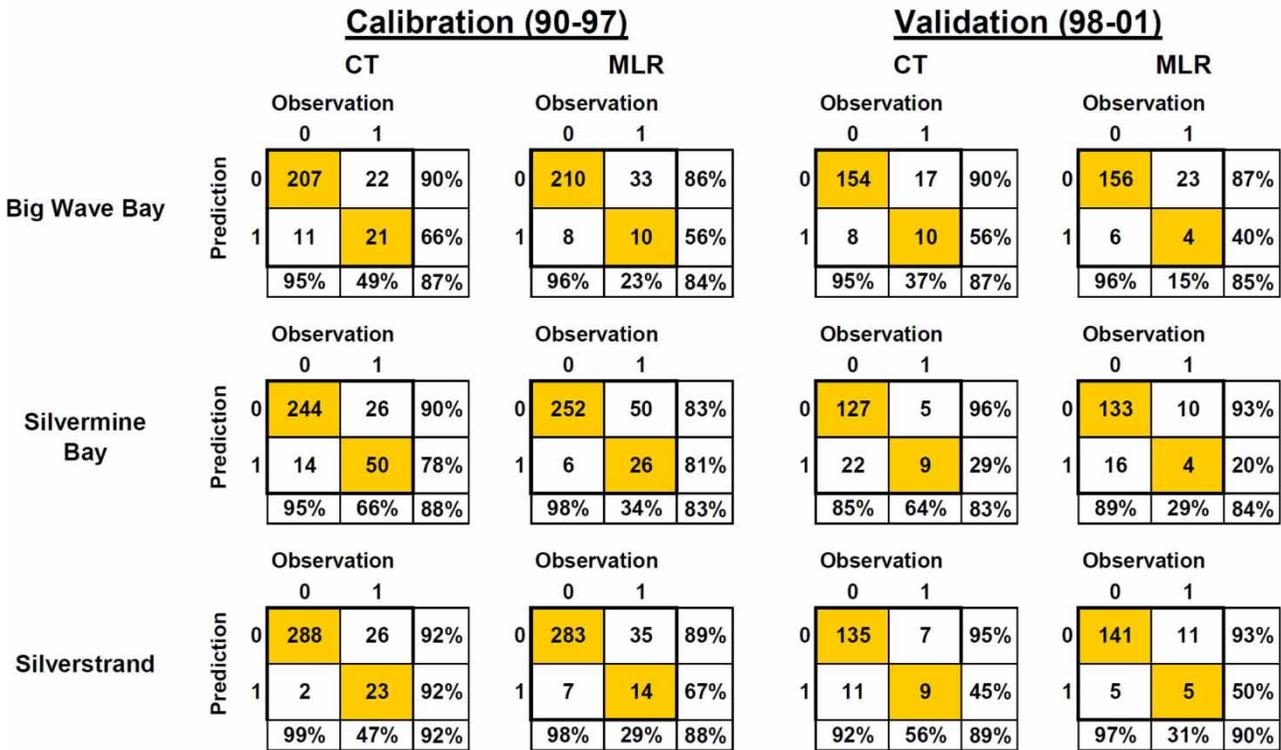

Figure 2 │ CT and MLR model performances in the pre-HATS period: calibration period (1990–1997) and validation period (1998–2001). *Note:* in the table, '1' = 610 threshold exceedance; '0' = 610 threshold non-exceedance.

the MLR model (97%). The corresponding total correct predictions for the two models are similar at around 90%.

The CT structures in the pre-HATS period are shown for (Figure 3(a)) Big Wave Bay, (Figure 3(b)) Silvermine Bay and (Figure 3(c)) Silverstrand, respectively. CT for Big Wave Bay has three criteria and are all rainfall related. Silvermine Bay has six criteria, including salinity, solar radiation, water temperature and 9-hour rainfall as inputs, with the first two variables appearing twice in the model. The CT structure for Silverstrand is more complicated, and has a total of eight criteria in five levels. There are four leaves out of nine predicting a 'very poor' event. The first three levels of the CT use day-1 rain, water temperature/salinity, and lnEC5 as inputs, respectively.

Model performances: post-HATS period

Figure 4 shows the CT and MLR performance tables in the post-HATS period during calibration (2002–2006) and

Figure 3 | Structures of the CTs for (a) Big Wave Bay, (b) Silvermine Bay and (c) Silverstrand in the pre-HATS period. *Note:* in the end-leaves '1' = 610 threshold exceedance; '0' = 610 threshold non-exceedance.

validation (2007 and 2008 daily data combined for Big Wave Bay, and 2008 for Silvermine Bay and Silverstrand). CT continues to perform better in capturing the 'very poor' events at non-point source beaches over MLR model. During calibration, CT for Big Wave Bay has a higher correct positive (45%) than the MLR model (40%), at the expense of a slightly lower correct negative (97% versus near 100%). CT for Silvermine Bay has slightly lower correct positive (50% versus 56%) but higher correct negative (98% versus 92%) and overall accuracy (94% versus 89%). During validation, CT models for both beaches show considerably higher correct positive (69% for Big Wave Bay and 44% for Silvermine Bay) than the MLR model (46% and 22%, respectively), with a similar level of correct negative (>93%). For Silverstrand, CT gives 29% correct positive during calibration with only one false alarm (close to 100% correct negative). During validation, CT can capture two out of four 'very poor' events (50% correct positive). In contrast, the MLR model for Silverstrand fails to predict any 'very poor' events during both calibration and validation.

The structures of CT in the post-HATS period are shown for (Figure 5(a)) Big Wave Bay, (Figure 5(b)) Silvermine Bay and (Figure 5(c)) Silverstrand, respectively. CT for Big Wave Bay has only one criterion (salinity), and Silvermine Bay has three (day-1 rain, 9-hour rain, tide level). For Silverstrand, there are five criteria leading to six end-leaves, with four leaves predicting a 'very poor' event. The first three levels of the CT use lnEC5, tide level and salinity for decision, followed by water temperature and day-3 rain in the fourth level.

CT to predict four-category BWQI

To compare CT performances in predicting binary and multiple-category outputs, two additional CTs are developed to predict the four-category BWQI for Big Wave Bay and Silverstrand in the post-HATS period. The 4×4 performance tables (Figure 6) show the observed and CT predicted BWQI-1 to 4 at the two beaches during validation. Calibration results are not shown for simplicity. All the assessment criteria in the performance table (e.g., correct positive and correct negative) are calculated based on the WQO threshold value of 180 *E. coli* counts/100 mL (i.e.,
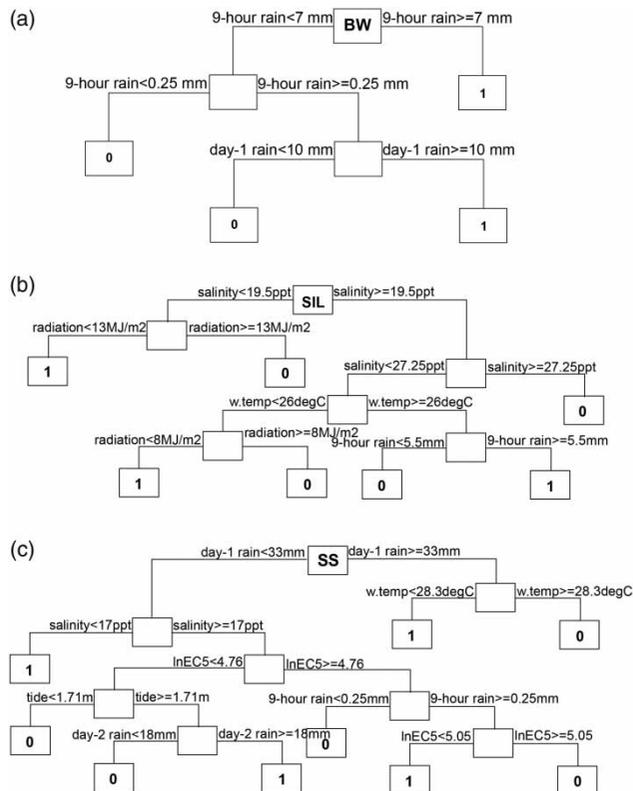
Figure 4 │ CT and MLR model performances in the post-HATS period: calibration period (2002–2006) and validation period (2007 and 2008 daily data combined). *Note:* in the table, '1' = 610 threshold exceedance; '0' = 610 threshold non-exceedance.
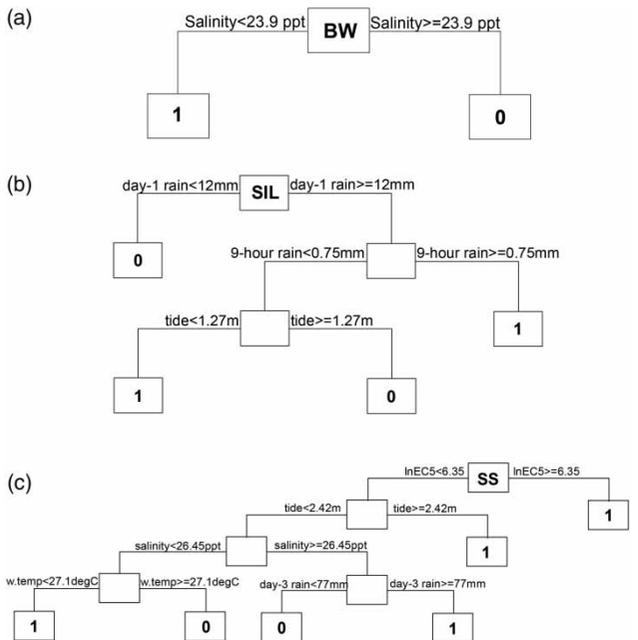

Figure 5 │ Structures of the CTs for (a) Big Wave Bay, (b) Silvermine Bay and (c) Silverstrand in the post-HATS period. *Note:* in the end-leaves, '1' = 610 threshold exceedance; '0' = 610 threshold non-exceedance.

Figure 6 │ CT performances in predicting four-category BWQI at Big Wave Bay and Silverstrand in the post-HATS validation period (2007 and 2008 daily data). All assessment criteria shown in this performance table are calculated based on *E. coli* threshold of 180 counts/100 mL (WQO).

positive = BWQI-3 and 4, negative = BWQI-1 and 2). Although the CT for Big Wave Bay has a very high correct positive in predicting WQO violation (BWQI-3 and 4, 81%), only 38% of 'very poor' events can be captured (five

predicted out of 13 events), considerably lower than the binary tree (69%). In addition, it gives excessive false alarms of 610 exceedance (33 times, as compared to 10 times by the binary CT), leading to unnecessary beach closures. For Silverstrand, CT can capture two out of eight WQO violation events during validation (25% correct positive), but cannot capture any 'very poor' events. The structure of the four-category CT for Big Wave Bay is more complicated than the corresponding binary CT: it has 24 end-leaves, two of which predict BWQI-4 (Figure 7). The CT for Silverstrand is even more complicated with 35 end-leaves, in which only two leaves predict BWQI-4 (detailed structure of the CT not shown).

DISCUSSION

CT model structures in pre-HATS and post-HATS periods

CT models for the same beach developed for the pre-HATS and post-HATS periods show different structures and decision rules. This is because beach water quality has significantly improved after the implementation of HATS, factors affecting beach water quality may have changed accordingly. This also implies a need for regular model update using the most recent data to capture any changes in water quality trend (e.g., due to upgrades of sewage

treatment facilities as in this study). For all three study beaches, their CT models in the post-HATS period have fewer decision rules than in the pre-HATS period. As an illustration, the model simplification for Big Wave Bay is possibly due to the fact that its water quality (in the form of *E. coli* concentration in natural logarithm) is much more correlated with salinity in the post-HATS (−0.61) than pre-HATS period (−0.36): the freshwater input from the nearby unsewered catchment becomes the dominating pollution source after all other sources have been mitigated in the post-HATS period (Thoe *et al.* 2012).

In the post-HATS period, CT structures for Big Wave Bay and Silvermine Bay are surprisingly simple with only one and three rules, respectively, to reach a decision. The corresponding [MP, ML] values adopted for Big Wave Bay and Silvermine Bay are [50, 5] and [10, 5], respectively. The simple CT structure identifies a few key criteria to trigger a 'very poor' water quality event. For Big Wave Bay, *E. coli* concentration is predicted to exceed 610 counts/100 mL only when salinity is lower than 23.9 ppt. For Silvermine Bay, exceedance is predicted when day-1 rain is greater than 12 mm and (1) 9-hour rain is greater than 0.75 mm, or (2) 9-hour rain is less than 0.75 mm but there is a low tide (tide level <1.27 m). Having rainfall and salinity as the major decision rules reveals that water quality at non-point source beaches are primarily driven by freshwater inputs, consistent with the observations based on MLR results (Thoe & Lee
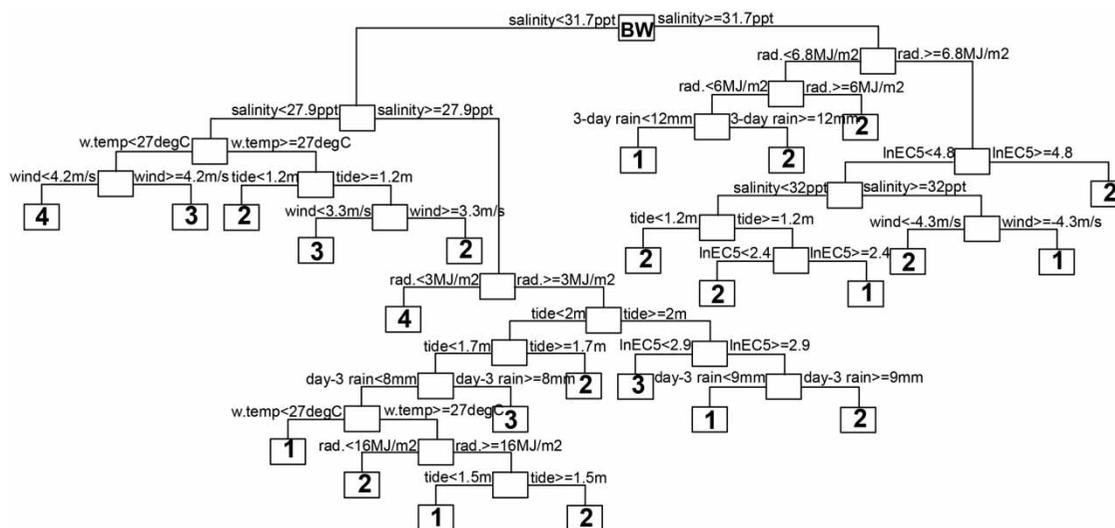

Figure 7 | Structure of the four-category CT for Big Wave Bay in the post-HATS period. *Note:* in the end-leaves, '1' to '4' correspond to BWQI-1 to 4, respectively.

2014). The inclusion of 9-hour rain in some models suggests a rapid deterioration of water quality after a recent storm event, largely due to the fact that catchments in Hong Kong are generally small in size ($<10\,km^2$) and have a short time of concentration (Thoe 2010). The very low 9-hour rain threshold (0.75 mm) adopted in the CT of Silvermine Bay also reflects that a 'very poor' grading can be triggered by a mild rain if the catchment is saturated by a preceding rain event (day-1 rain >12 mm is a prerequisite for a 9-hour rain of 0.75 mm to predict an exceedance). The more complicated CT structure for Silverstrand is brought about by a low ML value ([MP ML] = [25 1]). It reveals a complicated source of pollution, and partly explains why simple linear models do not perform well at this beach.

CT structures are also found to be more complicated when they are used to predict outputs with multiple categories. It gives an additional advantage to the binary CT for its relatively simple structure, making it easier to be used for real-time beach management.

Comparing classification trees with linear models

This study shows that CT holds promise in predicting 'very poor' events at beaches affected by either non-point source or local point source pollution, and also gives satisfactory results when used on a daily basis (based on the post-HATS validation results). This type of event, which can lead to immediate beach closure, has two characteristics: (1) they occur infrequently and (2) a binary decision of whether a beach should be open or closed is more important than the predicted concentrations. CT appears to be a good model for this application. First, CT can better capture the rare exceedances than MLR models. During the post-HATS period, 'very poor' events only contribute to about 6–8% of all events; despite the water quality improvement from the pre-HATS period, CT continues to give fairly consistent correct positive (30–69%, average 48%) during both calibration and validation, around 20% higher than that achieved by MLR models (0–56%, average 27%). The good performance at Silverstrand is particularly encouraging, because its MLR model cannot predict any of the 'very poor' events. Second, a binary dependent variable can be designed specifically to only the events of interest. In this study, the models are only used to predict whether the 610 counts/100 mL

threshold is exceeded. Since model calibration is not dependent on the exact concentration, events that only marginally exceed the 610 thresholds (which are more easily missed by MLR models) are treated equally with other pollution events that well exceed the threshold.

Another advantage of using CT over MLR models is its ability to model non-linear and threshold-type relationships between dependent and independent variables. Furthermore, categorical inputs can be easily used to develop CT, when they are relatively difficult to be incorporated into MLR models (i. e., need to convert into dummy variables). This increases the types of possible input variables such as weather and UV forecasts in categories, and tidal conditions (e.g., spring/neap cycle, flood/ebb/slack tide, semi-diurnal/diurnal tide). The tidal conditions are the most critical for beaches affected by the tidal currents. Chan *et al.* (2013) conducted a three-dimensional hydrodynamic modelling study, and found that beach water quality in Tsuen Wan and Tuen Mun districts in Hong Kong is generally worse under a semi-diurnal and flood tide over a spring cycle. Further investigation is necessary to develop CT models for this type of beach; a model to capture the diurnal variation of beach water quality based on the changing tidal condition may also be developed.

The four-category CTs developed at Big Wave Bay and Silverstrand are weaker than the binary CTs in capturing 'very poor' events. The excessive false alarms predicted by the four-category CT at Big Wave Bay is of particular concern. This may be due to the increase in the number of output categories (BWQI-1 to 4) without sufficient data for model calibration. In addition, the threshold concentrations (24, 180 and 610 counts/100 mL) for different BWQI are based on the associated number of excess illnesses obtained in a local epidemiological study conducted in the late 1980s (Cheung *et al.* 1990). These categories, especially BWQI-2 and 3 which lie between the two ends, may not bear a clear cause–effect relationship with the independent variables. Having BWQI-2 and 3 as the majority of the cases, model performance can be affected.

A combined modelling approach for beach management

A combined MLR and CT modelling approach is proposed to better capture 'very poor' pollution events while keeping

accuracy in predicting WQO violations (>180 counts/ 100 mL). The MLR model is used to predict the *E. coli* concentration and the corresponding BWQI, while the binary CT model is used to predict specifically BWQI-4. Based on the original beach management operation rules (Thoe & Lee 2014), the following 'enhanced' operation rules are suggested:

1. Beach water quality is predicted using both MLR and binary CT models everyday in the morning at 9:00 am.
2. BWQI-1 to 3 are issued based on MLR modelling results.
3. When MLR predicts BWQI-4 but CT predicts non-exceedance, BWQI-3 is issued. Additional sampling is conducted at the beach to cross-check the modelling results. BWQI-4 is re-issued if *E. coli* level exceeds 610 counts/100 mL in the water samples.
4. When CT predicts exceedance, BWQI-4 is issued irrespective of MLR modelling results. Additional sampling is conducted.

A flow diagram of the operation rules is shown in Figure 8. This combined modelling system uses essentially the same input variables as the MLR models to provide daily beach water quality forecasts (Thoe & Lee 2014). Only minor modification to the existing MLR-based system is needed, but higher correct positive to 'very poor' events by about 20% can be achieved. Especially for Big Wave

Bay and Silvermine Bay, the very simple CT structure enables even manual decisions without the need of computational power. As well, the system also gives insights into the design of beach monitoring programmes, by informing occasions when additional beach sampling may be necessary using very simple rules. As an illustration, whenever beach salinity at Big Wave Bay decreases to 24 ppt, additional sampling is recommended to ensure beach water quality is good for swimming. This also suggests a continuous monitoring of beach salinity can be a good management strategy for Big Wave Bay. In terms of *E. coli* measurement for additional sampling, rapid detection method such as qPCR method (Noble *et al.* 2010) can be adopted whenever possible. While culture-based method requires 24 hours to obtain the results, qPCR method can reduce the time to 4–6 hours, providing data to validate within the same day if the beach is polluted. However, care should be taken as qPCR and culture-based methods give different units for concentration (calibrator cell equivalents/100 mL and counts/100 mL, respectively), and the two methods may not hold a universal empirical relationship under different environmental conditions (Converse *et al.* 2012).

CONCLUSIONS

This study uses CT to predict pollution events of 'very poor' grading with *E. coli* concentration higher than 610 counts/ 100 mL at three Hong Kong beaches in both pre-HATS and post-HATS periods. The beaches are affected either by non-point source pollution (Big Wave Bay and Silvermine Bay) or local point source pollution (Silverstrand). These 'very poor' events are significant not only for public health protection but also beach management, because they may lead to immediate beach closure. Results show that binary CT can capture 44–69% of 'very poor' events on a daily basis while maintaining a minimum number of false alarms, even during the post-HATS period when beach water quality has significantly improved. Correct positives achieved by CTs is about 20% higher than their corresponding MLR models. CT's ability to better capture rare exceedances over linear models also suggests that it can be a suitable management tool for any clean beach with
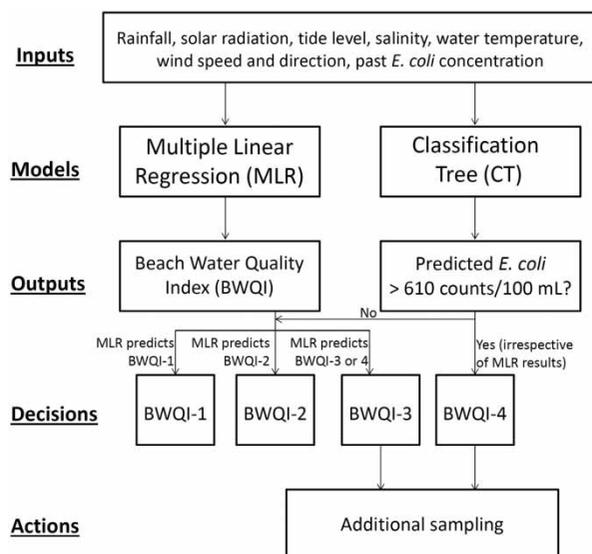

Figure 8 | Flow diagram of the proposed operation rules for beach management using a combined MLR and CT modelling approach.

infrequent exceedances. During the post-HATS period, CT structures are surprisingly simple for non-point source dominated beaches: rainfall or salinity are identified in the models as the major driving factors. Very simple rules can be used to inform potential beach pollution and closures. For beaches affected by the complicated local point source pollution, CT has a slightly more complicated structure, but its performance is encouraging with a high sensitivity to 'very poor' events. The corresponding four-category CT is found to be weaker in capturing 'very poor' pollution events than the binary CT, and also gives excessive false alarms at Big Wave Bay.

Based on the strengths and weaknesses of different predictive models, a combined MLR + CT modelling approach is proposed with operation rules for beach management. MLR models are used to predict *E. coli* concentration and the associated Beach Water Quality Indices, while CT models are used to increase the accuracy to capture the infrequent BWQI-4. The system gives improved predictions of both the four-category BWQI and 'very poor' pollution events relative to the MLR-based system, and also informs sampling occasions. This combined modelling system is easy to develop and is potentially a useful tool for other beaches having a wide range of bacterial concentrations. As all the comparisons in this study are conducted in 'hindcast' application, further investigations are needed to test the 'real-time' system performance in forecasting daily beach water quality when input variables such as salinity measurements are not available; adoption of a real-time salinity monitoring system to assist the modelling system can also be considered.

REFERENCES

Bae, H. K., Olson, B. H., Hsu, K. L. & Sorooshian, S. 2010 Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a California coastal area. *Water Sci. Technol.* **61** (2), 545–553.

Boehm, A. B., Grant, S. B., Kim, J. H., Mowbray, S. L., McGee, C. D., Clark, C. D., Foley, D. M. & Wellman, D. E. 2002 Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environ. Sci. Technol.* **36** (18), 3885–3892.

Boehm, A. B., Whitman, R. L., Nevers, M. B. & Weisberg, S. B. 2007 Now-casting recreational water quality. In: *Statistical Framework for Water Quality Criteria and Monitoring* (L. J. Wymer, ed.). John Wiley & Sons Ltd, Chichester, UK.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. 1983 *Classification and Regression Trees*. Wadsworth, Belmont, CA, USA.

Chan, S. N., Thoe, W. & Lee, J. H. W. 2013 Real-time forecasting of Hong Kong beach water quality by 3D deterministic model. *Water Res.* **47** (4), 1631–1647.

Cheung, W. H. S., Chang, K. C. K., Hung, R. P. S. & Kleevens, J. W. L. 1990 Health effects of beach water pollution in Hong Kong. *Epidemiol. Infect.* **105**, 139–162.

Converse, R. R., Griffith, J. F., Noble, R. T., Haugland, R. A., Schiff, K. C. & Weisberg, S. B. 2012 Correlation between quantitative PCR and culture-based methods for measuring *Enterococcus* spp. over various temporal scales at three California marine beaches. *Appl. Environ. Microbiol.* **78** (4), 1237–1242.

Francy, D. S., Brady, A. M. G., Carvin, R. B., Corsi, S. R., Fuller, L. M., Harrison, J. H., Hayhurst, B. A., Lant, J., Nevers, M. B., Terrio, P. J. & Zimmerman, T. M. 2013 *Developing and implementing predictive models for estimating recreational water quality at Great Lakes beaches*. US Geological Survey Scientific Investigations Report 2013-5166, p. 68. http://dx.doi.org/10.3133/sir20135166/.

Jost, L. 2006 Entropy and diversity. *Oikos* **113** (2), 363–375.

Lee, J. H. W., Choi, K. W., Thoe, W. & Wong, H. C. 2008 *Identification of critical factors affecting the bacteriological water quality of Hong Kong beaches*. Technical report prepared for Hong Kong Environmental Protection Department, December 2008.

Noble, R. T., Blackwood, A. D., Griffith, J. F., McGee, C. D. & Weisberg, S. B. 2010 Comparison of rapid quantitative PCR-based and conventional culture-based methods for enumeration of *Enterococcus* spp. and *Escherichia coli* in recreational waters. *Appl. Environ. Microbiol.* **72** (2), 1604–1612.

Parkhurst, D. F., Brenner, K. P., Dufour, A. P. & Wymer, L. J. 2005 Indicator bacteria at five swimming beaches – analysis using random forests. *Water Res.* **39** (7), 1354–1360.

Stidson, R. T., Gray, C. A. & McPhail, C. D. 2012 Development and use of modelling techniques for real-time bathing water quality predictions. *Water Environ. J.* **26**, 7–18.

Thoe, W. 2010 A daily forecasting system of marine beach water quality in Hong Kong. PhD thesis, The University of Hong Kong.

Thoe, W. & Lee, J. H. W. 2014 Daily forecasting of Hong Kong beaches water quality by multiple linear regression (MLR) models. *ASCE J. Environ. Eng.* **140** (2), 04013007.

Thoe, W., Wong, S. H. C., Choi, K. W. & Lee, J. H. W. 2012 Daily prediction of marine beach water quality in Hong Kong. *J. Hydro-Environ. Res.* **6** (3), 164–180.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M. L. & Boehm, A. B. 2014 Predicting water quality at Santa Monica Beach: evaluation of five different models for public notification of unsafe swimming conditions. *Water Res.* **67**, 105–117.

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M. L. & Boehm, A. B. 2015 Sunny with a chance of gastroenteritis: predicting swimmer risk at California beaches. *Environ. Sci. Technol.* **49**, 423–431.

USEPA 2011 *Recreational Water Quality Criteria*. US Environmental Protection Agency. Washington, DC, USA. 820-F-12-058.

First received 30 April 2015; accepted in revised form 18 August 2015. Available online 8 October 2015