

# 海水浴场中粪便污染指示细菌的浓度预测模型研究进展

黄慧玲<sup>1,2</sup>, 明红霞<sup>2</sup>, 樊景凤<sup>2</sup>, 王斌<sup>1</sup>

(1.大连海洋大学水产与生命学院, 辽宁大连 116023; 2.国家海洋环境监测中心, 辽宁大连 116023)

**摘要:** 构建粪便污染指示细菌的浓度预测模型有助于保障海水浴场水质, 避免因传统的病原微生物监测方法的时间滞后性而影响海水浴场管理的有效性, 可及时给予公众健康指导, 为海水浴场管理提供技术支撑。本文基于国内外海水浴场水质中粪便污染指示细菌的浓度预测模型技术发展现状, 综合评述了模型构建方法, 归纳了各种模型的特点、应用范围以及在实际应用中的优缺点, 分析了在构建我国海水浴场粪便污染指示细菌的浓度预测模型中所面临的困难及存在的差距, 对模型的发展趋势进行了展望, 为开展我国海水浴场中粪便污染指示细菌的浓度预测提供参考。

**关键词:** 海水浴场; 粪便污染指示细菌; 浓度预测模型

**中图分类号:**      **文献标识码:** A      **文章编号:** 1007-6336(\$article\_year)00-0001-07

## Overview on indicator bacterial concentration prediction model of bathing beaches

HUANG Hui-ling<sup>1,2</sup>, MING Hong-xia<sup>2</sup>, FAN Jing-feng<sup>2</sup>, WANG Bin<sup>1</sup>

(1.College of Fisheries and Life Science, Dalian Ocean University, Dalian 116023, China; 2.National Marine Environmental Monitoring Center, Dalian 116023, China)

**Abstract:** It will help to ensure the safety of the bathing water quality and avoid the time-lag of the traditional monitoring method affecting the effectiveness of the beach management using a concentration prediction model for fecal indicator bacteria (FIB). And it also can provide the health guidance and technical supporting to the public and beach manager in timely. Based on the technical development status of the indicator bacterial concentration prediction model of bathing beaches, the model construction methods were comprehensively reviewed in this paper, and the characteristics, application scopes, advantages and disadvantages of various models were also summarized. In addition, the difficulties and bottleneck in constructing the indicator bacterial concentration prediction model in bathing beaches of China were analyzed. Finally, the development trend of the model was prospected, which will provide a valuable reference for administrative department of China.

**Key words:** bathing beach; fecal indicator bacteria; concentration prediction model

近年来,随着沿海经济的发展,人类活动加剧,海洋环境中的病原微生物污染愈发严重<sup>[1]</sup>。海水浴场作为重要的旅游活动区域,是传播疾病的敏感地带。流行病学研究表明,公众在被病原微生物污染的浴场中活动极易患上胃肠道和呼吸道等疾病<sup>[2-4]</sup>,及时对浴场水质进行分析评估

和预报对保护游客的健康安全至关重要<sup>[4]</sup>。20世纪70年代末到80年代初,美国、法国、加拿大、澳大利亚、日本、欧盟等国家和组织相继颁布了娱乐用水的微生物监测标准<sup>[5]</sup>。我国自2002年起,也在大连、秦皇岛、烟台、威海等地开展了全国主要海水浴场的海洋环境监测工作<sup>[6]</sup>。

收稿日期:2019-08-12, 修订日期:2019-11-14

基金项目:国家重点研发计划“海洋环境安全保障”重点专项基金项目(2017YFC1404500),辽宁省兴辽英才计划青年拔尖人才计划项目(XLYC1807186)

作者简介:黄慧玲(1995-),女,辽宁铁岭人,硕士,主要研究方向为病原微生物, E-mail: huilhuang@outlook.com

通讯作者:樊景凤(1972-),女,研究员,主要研究方向为海洋微生物学与生态学, E-mail: jffan@nmemc.org.cn

国内外现有的浴场水质监测主要是基于粪便指示细菌 (fecal indicator bacteria, FIB) 的浓度评估浴场病原微生物的感染风险, 保护游泳者的健康<sup>[7]</sup>, 典型的指示细菌包括肠球菌 (enterococci, ENT)、总大肠菌 (total coliform, TC) 和大肠杆菌 (*E. coli*)<sup>[7-8]</sup> 等。在美国, 通常将 *E. coli* 或 ENT 作为淡水或海水的粪便指示细菌, 并根据其监测结果决定是否发布关闭海水浴场的公告。监测粪便指示细菌浓度的传统手段主要是基于可培养的方法, 至少需要 18~24 h 才能得到结果<sup>[9]</sup>, 因此海水浴场当天通报的水质状况主要依据的是前一两天指示细菌的浓度<sup>[10]</sup>, 水文气象条件的变化会导致指示细菌浓度随之变化, 造成浴场健康管理存在时间滞后性<sup>[10-11]</sup>。由此可见, 依靠传统的 FIB 的检测方法不能准确地评估浴场环境的微生物安全状况。

建立指示细菌浓度预测模型可以起到及时预报的目的, 有效降低监测误差, 从而更好地保护公众的健康安全<sup>[12]</sup>。2003 年世界卫生组织 (WHO) 首次引入浴场水质预测模型<sup>[13]</sup>, 欧盟也于 2006 年将其纳入浴场用水指导文件中<sup>[14]</sup>, 2012 年美国环境保护署 (U.S. Environmental Protection Agency, USEPA) 也建议将浴场水质预测模型作为快速、高效的监测工具, 以弥补传统检测方法带来的时间滞后的缺陷<sup>[15]</sup>。

目前, 常见的浴场水质预测模型有多元线性回归模型 (multiple least-squares regression, MLR)、人工神经网络 (artificial neural network, ANN)、偏最小二乘回归模型 (partial least squares, PLS)、决策树 (decision tree) 和其他模型等<sup>[12]</sup>。在实际应用中, 研究学者们发现通过预测模型所呈现的结果普遍优于传统的监测系统<sup>[10,16-25]</sup>。美国<sup>[17]</sup>、英国苏格兰<sup>[18]</sup> 和中国香港<sup>[19]</sup> 等国家和地区已使用预测模型构建了水质预测监测系统以加强浴场的健康安全管理。本文拟通过总结上述模型的构建方法和实际应用中的优缺点, 为构建我国海水浴场粪便污染指示细菌的浓度预测模型提供借鉴。

## 1 常见的浴场水质预测模型

### 1.1 多元线性回归模型 (MLR)

动态多元线性回归 (MLR) 模型是基于 SPSS

Statistics (version 20, IBM, Chicago, IL) 软件研发的模型, 采用逐步回归算法, 避免了因方差扩大因子导致共线性, 从而降低模型准确度的情况<sup>[26]</sup>。MLR 模型易构建, 变量多, 预测准确度较高, 是使用最广泛的一类模型, 目前已用于湖泊、海水浴场等环境中的病原微生物安全研究<sup>[18,27-28]</sup>。如 Olyphant 等人<sup>[18]</sup> 在对芝加哥海水浴场的研究中, MLR 模型对浴场指示细菌是否超标的预测准确度达 88%。

MLR 模型可以通过环境变量, 如降雨量、太阳辐射、风速、潮位、盐度、水温、指示细菌的实测浓度等数据<sup>[29]</sup> 预测海水浴场水质的每日动态变化, 如因污水溢流<sup>[30]</sup> 和非点源污染导致的海水中指示细菌浓度上升<sup>[31-32]</sup>, 或因紫外线照射浓度下降<sup>[33-34]</sup> 等变化。Thoe 等人<sup>[24]</sup> 以 *E. coli* 作为环境监测指标, 使用 MLR 模型, 成功预测到四个水文条件不同的海水浴场微生物水质状况和 *E. coli* 的日动态变化 (以 180 *E. coli* CFU/100 mL 作为临界值), 监测预测的结果准确度分别是 70%、92%、82% 和 85%。

MLR 模型的预测准确度主要取决于指示细菌的浓度。如 Francy 等<sup>[9]</sup> 使用 MLR 模型对俄亥俄州亨廷顿海水浴场 3 个月的监测数据进行评估测试, 结果显示该模型的准确度为 82.7%, 并且发现该模型对较为干净的沙滩预测效果并不好, 对中等污染的沙滩预测效果最好。此外, 当 MLR 模型的环境参数变量的数值发生剧烈变化时, 也会导致模型的预测准确度不高。例如在暴风雨前后盐度会剧烈变化, 导致模型预测的误差变大<sup>[29]</sup>, 所以 MLR 模型会配合人工神经网络 (ANN) 模型一同使用<sup>[24,29]</sup>, 例如香港开发的浴场海水水质预测系统 (WATERMAN) (<http://www.waterman.hku.hk/>)<sup>[35]</sup>, 构建了一个综合的每日浴场海水水质预报系统, 该系统于 2010—2011 年期间预测了 16 个具有代表性的海水浴场的水质, 可以获得 78%~100% 的预测准确度, 同时通过对关键参数进行分析发现, 降雨和盐度是非点源污染类型的海水浴场水质的主要影响参数, 该系统对这类非点源污染的浴场水质预测效果最好。

### 1.2 人工神经网络 (ANN)

人工神经网络模型是由 MATLAB“神经网

络”工具箱建立的神经网络模型(ANN),也可以使用MLR模型中的变量开发ANN模型<sup>[36]</sup>。ANN模型的预测效果好、用途广泛,近几年,逐渐应用到各科学研究领域<sup>[37]</sup>,如水文和水利工程方面的污水处理、河流和海岸线水质监测、泥沙输送、洪水演进、藻类繁殖问题<sup>[18,38]</sup>等。ANN属于数值模型,建立在溶质运移问题和动力学方程基础之上,该模型假设功能变量间不存在任何交互关系,特别适合处理变量之间的非线性关系<sup>[17]</sup>,因此该模型也被应用于复杂的系统中预测海水浴场细菌浓度,分析指示细菌浓度与其他容易测定的如温度、电导率、水流量、降雨量等水质变量之间的关系<sup>[39-40]</sup>。

有研究通过使用ANN模型成功预测了淡水或海水水域的水质状况<sup>[41-43]</sup>,该模型现已发展成为预测水环境中病原微生物动态的主要工具之一。该模型的优势是在模拟期间的任何时间段均可预测出指示细菌的变化。Lee等人<sup>[21]</sup>对加利福尼亚南部的海水浴场进行了水质检测,并使用ANN模型预测了加利福尼亚州圣地亚哥两个海水浴场的FIB浓度(TC, FC和ENT),用于弥补检测数据和实际水质之间的差距。结果表明,开发的模型能够同时准确预测三种FIB的浓度水平,得出的假阳性或阴性率低于10%,即开发的ANN模型的准确度为90%。

### 1.3 偏最小二乘法回归模型(PLS)

偏最小二乘法回归模型(PLS)是由MATLAB(version R2012b, Natick, MA)软件研发的模型,采用回归算法将所有的独立变量转换为主成分用以建立模型<sup>[44-45]</sup>。该模型是一种解释共线性的分析方法,能够削减或抑制那些对输出建模无意义的预测因子的影响<sup>[46-47]</sup>。模型的变量数据可以从公共数据库中获取,使用模型的用户只需提供检测的指示细菌浓度,就可在软件中自动进行模型的建立和检验,如Brooks等人<sup>[44]</sup>利用Python 2.6.6<sup>[48]</sup>和Rpy2 2.0.8<sup>[49]</sup>编写的统计软件包就可以通过该模型进行自动化回归建模。

Hou等人<sup>[45]</sup>使用PLS模型模拟预测五大湖水域的四个海水浴场站点的FIB浓度,与四个站点当前使用的普通最小二乘法(ordinary least squares, OLS)模型相比,PLS模型准确度与其相

当,但其无需选择变量,因此模型构建过程更加简化。另外作者还创建了动态偏最小二乘回归(dynamic partial least squares regression, DPLSR)模型,用于监测加利福尼亚州亨廷顿州立海(Huntington State Beach, HSB)和亨廷顿市海水浴场(Huntington City Beach, HCB)的水质,从公开数据库中获取了待测站点的环境数据作为因子变量,包括降雨量、海面温度、上升流指数、风速、海浪高度和方向、游客数量、大气压力、太阳辐射、采样时间、潮位和范围等<sup>[50-53]</sup>影响海水浴场FIB浓度的因素<sup>[11,54-56]</sup>。通过对模型的评估发现,与现有的OLS模型相比,DPLSR模型的预测能力更好,但在分析这两种模型的预测结果与浴场每日疾病发生率的关系时,二者没有显著差异,而且PLS模型的预测值和观察值之间仍存在差异。模型预测的目的是减少海水浴场游客的疾病发生率,所以使用PLS模型对水质进行预测时需要进一步的完善。

### 1.4 决策树模型

决策树模型(decision tree)是一种持续将数据分组、分类直到不能从后续的拆分中获得更多信息的模型。决策树建模在许多行业中都有广泛使用,也常用于浴场水质预测。浴场水质是随着时间持续变化着的,这就需要模型的预测结果进行不断的改进和修正,而决策树在一定程度上是进化的<sup>[57]</sup>,所以很适合用于构建浴场水质预测模型,预测的准确度较高。如Stidson等人使用软件“S-plus”和“Insightful Miner”决策树工具,以TC, FC和粪便链球菌(fecal streptococci, FS)作为指示细菌,对苏格兰四个海水浴场站点的水质进行预测,结果得出,该模型能够快速对数据进行分类,每天和每小时的预测准确度分别为88%和89%<sup>[57]</sup>。

根据决策树预测输出结果的不同,可以分为分类决策树和回归决策树。当预测结果为类别时,为分类决策树;当预测结果为一个实数时,为回归决策树。

分类决策树是利用MATLAB建立的“分类树”(classification tree, CT)模型,由Breiman等人于1984年开发而成<sup>[58]</sup>,该模型可以轻松处理分类和数字变量,因为其分裂算法能够隔离单个节

点中的异常值<sup>[59-60]</sup>,从而将其识别和剔除;另外为了验证每棵树,其采用留一法交叉验证,每个数据集仅限于分成两个不同的组,用于构建和验证树的结构<sup>[61-62]</sup>。Bae 等人通过 CART Ex Version 6.0 分析软件(Salford Systems 2006)<sup>[63]</sup>,利用现有的指示细菌数据和一些物理化学参数,使用分类决策树对加利福尼亚海水浴场指示细菌的浓度进行预测分析,结果显示,该模型能够成功预测各指示细菌的浓度,准确度在 60% 以上,该模型的优势在于可以利用简单快速监测到的水质变量来预测指示细菌的浓度,从而进行浴场健康管理<sup>[22]</sup>。

回归决策树的代表模型为随机树模型<sup>[58]</sup>。“随机树”提供了一种相对较新的方法来分析数据变量,其与多重线性回归相比较,不需要假设关系是线性的,该模型能够自动计算变量之间可能存在的交互关系<sup>[19]</sup>。Parkhurst 等人运用“随机树”模型进行水中指示细菌浓度的预测时,发现水中指示细菌的浓度与许多潜在变量存在相关性,而随机树与单树相比,可使预测误差降低 30%。另外除实时预测外,该模型还可以对未来一段时间内的指示细菌浓度进行预测, Parkhurst 等人<sup>[63-65]</sup>以大肠杆菌和肠球菌作为指示细菌,使用“随机树”模型研究美国五个海水浴场的指示细菌浓度与其他众多变量之间的关系,发现使用前 52d 测量的数据可预测未来 10d 内的指示细菌的浓度,证明该模型在分析具有许多变量的大型数据集时具有优势<sup>[64]</sup>。

### 1.5 其他模型

除上述常用的指示细菌浓度预测模型外,其他模型包括二元回归模型(binary logistic regression, BLR), 普通最小二乘法回归模型(OLS), SEPA (the scottish environment protection agency) 预测模型等。BLR 模型通过 SPSS 软件研发,采用“向前:条件算法”对变量进行选择<sup>[7]</sup>,具有二元因变量(1: posting, 0: no posting),能预测 SSS(SSS: 一定的指示细菌浓度值,超出该值时水质不适合游泳)超标的可能性(模型输出范围从 0 到 100%)<sup>[12]</sup>。普通最小二乘法回归模型(OLS)是由威斯康辛州自然资源部创建的<sup>[66]</sup>,通常 OLS 模型会和 PLS 模型共同使用进行比较分析<sup>[44-45]</sup>。

SEPA 是指示预测工具,是一个以电子表格为基础的操作模型,通常被用于开发以一系列特定变量比如降雨量和入海河流流量的历史数据来预测浴场水质的模型<sup>[57]</sup>。另外还有 Seneque/River strahler 模型,是一种与大型排水网络水质和生物化学功能相关的模型<sup>[67]</sup>,首次被研发使用于塞纳河分析河流中浮游植物爆发的相关因素,以控制水华的发生<sup>[68-70]</sup>,随后该模型被应用于其他大型流域如东欧的多瑙河<sup>[71]</sup>。除此之外还有水动力模型,比如模拟微生物运输和衰变的流体力学模型 GEMSS(GEMSS-HDM), Edinger 和 Buchak 曾介绍过该模型的理论基础<sup>[72]</sup>,该模型使用 GLLVHT(广义,纵向-横向-垂直水动力和运输)来计算不同水体中的微生物随时间的衰变速度、水质成分如盐分浓度等变量。

## 2 总结与展望

通过对上述各个模型的特点进行总结(表 1)可以看出,所述模型各有优缺点。如 MLR 模型是最常用的模型<sup>[16,27,30,73]</sup>,该模型是线性模型,但其缺陷在于模型灵敏度较低,未来研究中可尝试借鉴类似于 BLR 模型中的“阈值调整”方法来提高 MLR 模型的灵敏度。PLS 既可以是线性又可以是非线性的模型,但同样 PLS 模型的灵敏度也较低,也可尝试“阈值调整”的方法来提高灵敏度。决策树中 CT 模型是一种非线性的视觉模型,具有较高的灵敏度,但有研究表明,当受测的海水浴场受到更多污染时,影响水质的相互作用变量增加,CT 模型结构会变得更加复杂,从而影响计算的准确度。ANN 是非线性模型,可以得到依赖变量和自变量之间的非线性关系, Thoe 等人研究发现 ANN 模型是继 CT 模型后,灵敏度第二高的模型,ANN 对 ENT 和 FC 的灵敏度

表 1 浴场水质预测模型特点比较

Tab.1 Comparison of bathing beaches water quality prediction model

模型	线性	非线性	优点	缺点
MLR	√		准确度较高	灵敏度较低
ANN		√	灵敏度高	不能显示变量间的关系
PLS	√	√	准确度较高	灵敏度较低
CT		√	灵敏度高	结构复杂

分别达到 30% 和 28%, 校准和验证期间的特异性非常高(85%~99%)<sup>[7]</sup>。然而, 人工神经网络的缺点是它是一种“黑箱”模式, 即下层关系变量之间不能直接表示<sup>[7]</sup>, 在外推输出中也较弱, 即当输入变量的值超出校准范围时, 它预测的 FIB 浓度是不准确的<sup>[24]</sup>。

模型的性能主要根据环境的特殊性和模型本身的局限性来进行评估, 同时模型的选择也与模型本身的预测程度有关, 通常使用者会就多个模型去计算进而比较, 选出最优模型。如 Thoe 等人在加利福尼亚海水浴场对五种模型的测试中得出, CT 和 BLR-T<sup>[24]</sup>(调整后的 BLR 模型) 始终是基于单一样本标准预测海水浴场水质的最佳表现模型。而在另一篇关于圣莫尼亚卡海水浴场最佳模型选择测试中, CT 模型是最佳模型, ANN 和 BLR-T 位次佳<sup>[7]</sup>。

指示细菌浓度预测模型的构建有助于实现海水浴场污染的早期预警, 是确定预测是否对公众开放的标准参考, 大大方便了浴场管理。目前模型的使用范围在不断扩大, 但仍有局限性, 模型看似是一个独立的个体, 但实际上它涉及面广, 综合性很强, 在建模过程中无论是水文气象因素还是物理因素都要考虑到。

目前应用模型开展浴场水质监测的报道大多来自国外, 虽然近年来中国香港的浴场监测模型发展迅速, 实际应用效果明显, 但大陆浴场水质预测模型的研究还存在着如下几点问题: 一是我国水质预测研究起步较晚, 自主研发浴场水质预测模型的能力较弱, 已建立的预测模型也存在着描述对象单一、空间维数低、预测范围窄、概化参数多等问题, 因此缺乏推广性。二是受限于信息源, 构建模型时使用的变量参数数据不充分; 三是没有建立起系统完整的监测系统, 浴场风险管理不足。

对于国内大陆建立系统完善的浴场水质监测模型有如下几点建议: 一是加强水质预测模型的理论研究, 学习国外浴场构建水质预测模型的先进经验。二是加强动力学、微生物学以及数学等领域的学科交叉。三是集合多方力量, 收集、整理长期的浴场水质监测数据。四是完善现有监测系统, 有针对性地监测、补充模型所需

重要参数。

## 参考文献:

- [1] 王迎春. “十一五”大连市海水浴场水质状况分析[J]. 科技传播, 2011, (22): 77-78.
- [2] PRÜSS A. Review of epidemiological studies on health effects from exposure to recreational water[J]. *International Journal of Epidemiology*, 1998, 27(1): 1-9.
- [3] HAILE R W, WITTE J S, GOLD M, et al. The health effects of swimming in ocean water contaminated by storm drain runoff[J]. *Epidemiology*, 1999, 10(4): 355-363.
- [4] 马祖友, 张文斌, 夏煜, 等. 2005年高罗海水浴场现状分析与评价[J]. *海洋开发与管理*, 2006, 23(6): 141-144.
- [5] 张微微, 王菊英, 王燕, 等. 海水浴场水质监测与评价研究进展[J]. *海洋开发与管理*, 2014, 31(7): 99-104.
- [6] 周连翔. 海水浴场环境监测预报工作进展概况[J]. *海洋预报*, 2005, 22(S1): 161-166.
- [7] THOE W, GOLD M, GRIESBACH A, et al. Predicting water quality at Santa Monica Beach: evaluation of five different models for public notification of unsafe swimming conditions[J]. *Water Research*, 2014, 67: 105-117.
- [8] 陈昕. 新旧版《生活饮用水标准检验方法》中微生物指标及检测方法比较[J]. *环境与职业医学*, 2008, 25(3): 297-299.
- [9] FRANCY D S. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches[J]. *Aquatic Ecosystem Health & Management*, 2009, 12(2): 177-182.
- [10] GONZALEZ R A, NOBLE R T. Comparisons of statistical models to predict fecal indicator bacteria concentrations enumerated by qPCR- and culture-based methods[J]. *Water Research*, 2014, 48: 296-305.
- [11] BOEHM A B, GRANT S B, KIM J H, et al. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California[J]. *Environmental Science & Technology*, 2002, 36(18): 3885-3892.
- [12] THOE W, GOLD M, GRIESBACH A, et al. Sunny with a chance of gastroenteritis: predicting swimmer risk at California Beaches[J]. *Environmental Science & Technology*, 2015, 49(1): 423-431.
- [13] World Health Organization. Guidelines for safe recreational-water environments: coastal and fresh-waters, vol 1[M]. Geneva: World Health Organization, 2003.
- [14] TEPatCotE U. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC[J]. *Official Journal of the European Union*, 2006, 64: 14.
- [15] Fujioka R, Solo-Gabriele H, Byappanahalli M, et al. US recreational water quality criteria: a vision for the future[J]. *International journal of environmental research and public health*, 2015, 12(7): 7752-7776.
- [16] CROWTHER J, KAY D, WYER M D. Relationships between

- microbial water quality and environmental conditions in coastal recreational waters: the Fylde Coast, UK[J]. *Water Research*, 2001, 35(17): 4029-4038.
- [17] LIN B, KASHEFIPOUR S M, FALCONER R A. Predicting near-shore coliform bacteria concentrations using ANNS[J]. *Water Science & Technology*, 2003, 48(10): 225-232.
- [18] OLYPHANT G A, WHITMAN R L. Elements of a predictive model for determining beach closures on a real time basis: the case of 63rd street beach Chicago[J]. *Environmental Monitoring and Assessment*, 2004, 98(1/2/3): 175-190.
- [19] PARKHURST D F, BRENNER K P, DUFOUR A P, et al. Indicator bacteria at five swimming beaches—analysis using random forests[J]. *Water Research*, 2005, 39(7): 1354-1360.
- [20] WYMER L J. *Statistical framework for recreational water quality criteria and monitoring*[M]. Hoboken: John Wiley & Sons, 2007.
- [21] HE L M, HE Z L. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA[J]. *Water Research*, 2008, 42(10/11): 2563-2573.
- [22] BAE H K, OLSON B H, HSU K L, et al. Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a Californian coastal area[J]. *Water Science & Technology*, 2010, 61(2): 545-553.
- [23] GONZALEZ R A, CONN K E, CROSSWELL J R, et al. Application of empirical predictive modeling using conventional and alternative fecal indicator bacteria in eastern North Carolina waters[J]. *Water Research*, 2012, 46(18): 5871-5882.
- [24] THOE W, WONG S H C, CHOI K W, et al. Daily prediction of marine beach water quality in Hong Kong[J]. *Journal of Hydro-environment Research*, 2012, 6(3): 164-180.
- [25] ZHANG Z H, DENG Z Q, RUSCH K A. Development of predictive models for determining *enterococci* levels at Gulf Coast beaches[J]. *Water Research*, 2012, 46(2): 465-474.
- [26] BOWERMAN B L. *Statistical design and analysis of experiments with applications to engineering and science*[J]. *Technometrics*, 1991, 33(1): 105-106.
- [27] FRICK W E, GE Z F, ZEPP R G. Nowcasting and forecasting concentrations of biological contaminants at beaches: a feasibility and case study[J]. *Environmental Science & Technology*, 2008, 42(13): 4818-4824.
- [28] Francy D S, Darner R A. *Nowcasting beach advisories at Ohio Lake Erie beaches*[R]. Geological Survey (US), 2007.
- [29] THOE W, LEE J H W. Real-time forecast of marine beach water quality in Hong Kong[C]//*Proceedings of the 6th International Conference on APAC 2011*. Hong Kong, China: World Scientific Publishing Co Pte Ltd, 2011.
- [30] OLYPHANT G A, THOMAS J, WHITMAN R L, et al. Characterization and statistical modeling of bacterial (*Escherichia coli*) outflows from watersheds that discharge into southern lake Michigan[J]. *Environmental Monitoring and Assessment*, 2003, 81(1/2/3): 289-300.
- [31] WHITMAN R L, NEVERS M B, GEROVAC P J. Interaction of ambient conditions and fecal coliform bacteria in southern Lake Michigan beach waters: monitoring program implications[J]. *Natural Areas Journal*, 1999, 19(2): 166-171.
- [32] Haack S K, Fogarty L R, Wright C. *Escherichia coli* and enterococci at beaches in the Grand Traverse Bay, Lake Michigan: sources, characteristics, and environmental pathways[J]. *Environmental science & technology*, 2003, 37(15): 3275-3282.
- [33] FUJIOKA R S, HASHIMOTO H H, SIWAK E B, et al. Effect of sunlight on survival of indicator bacteria in seawater[J]. *Applied and Environmental Microbiology*, 1981, 41(3): 690-696.
- [34] DAVIES-COLLEY R J, BELL R G, DONNISON A M. Sunlight inactivation of enterococci and fecal coliforms in sewage effluent diluted in seawater[J]. *Applied and Environmental Microbiology*, 1994, 60(6): 2049-2058.
- [35] THOE W, LEE J H W. Daily forecasting of Hong Kong beach water quality by multiple linear regression models[J]. *Journal of Environmental Engineering*, 2014, 140(2): 04013007.
- [36] RECKNAGEL F. ANNA-Artificial Neural Network model for predicting species abundance and succession of blue-green algae[J]. *Hydrobiologia*, 1997, 349(1/2/3): 47-57.
- [37] WIDROW B, RUMELHART D E, LEHR M A. Neural networks: applications in industry, business and science[J]. *Communications of the ACM*, 1994, 37(3): 93-105.
- [38] ANDERSON D M. Toxic red tides and harmful algal blooms: a practical challenge in coastal oceanography[J]. *Reviews of Geophysics*, 1995, 33(S2): 1189-1200.
- [39] Liang D, Falconer R A, Lin B. Linking one-and two-dimensional models for free surface flows[C]//*Proceedings of the Institution of Civil Engineers-Water Management*. London: [Published for the Institution of Civil Engineers by Thomas Telford Ltd.], c2004-, 2007, 160(3): 145-152.
- [40] HE L M, KEAR-PADILLA L L, LIEBERMAN S H, et al. Rapid in situ determination of total oil concentration in water using ultraviolet fluorescence and light scattering coupled with artificial neural networks[J]. *Analytica Chimica Acta*, 2003, 478(2): 245-258.
- [41] BRION G M, NEELAKANTAN T R, LINGIREDDY S. A neural-network-based classification scheme for sorting sources and ages of fecal contamination in water[J]. *Water Research*, 2002, 36(15): 3765-3774.
- [42] KUO J T, HSIEH M H, LUNG W S, et al. Using artificial neural network for reservoir eutrophication prediction[J]. *Ecological Modelling*, 2007, 200(1/2): 171-177.
- [43] LEE J H W, HUANG Y, DICKMAN M, et al. Neural network modelling of coastal algal blooms[J]. *Ecological Modelling*, 2003, 159(2/3): 179-201.

- [44] BROOKS W R, FIENEN M N, CORSI S R. Partial least squares for efficient models of fecal indicator bacteria on Great Lakes beaches[J]. *Journal of Environmental Management*, 2013, 114: 470-475.
- [45] HOU D Y, RABINOVICI S J M, BOEHM A B. Enterococci predictions from partial least squares regression models in conjunction with a single-sample standard improve the efficacy of beach management advisories[J]. *Environmental Science & Technology*, 2006, 40(6): 1737-1743.
- [46] EPA. Ambient water quality criteria for bacteria-1986[M]. Washington: EPA Office of Water, 1986.
- [47] Van Rossum G, Drake Jr F L. Python tutorial[M]. Amsterdam: Centrum voor Wiskunde en Informatica, 1995.
- [48] GAUTIER L, Moreira W, Warnes G R. rpy2 - R in Python[EB/OL]. (2018)[2019-08-12] <https://rpy2.bitbucket.io/>
- [49] R Core Team. R: a language and environment for statistical computing[EB/OL]. (2012)[2019-08-12] <http://www.R-project.org>.
- [50] ERGON R. Principal component regression (PCR) and partial least squares regression (PLSR)[M]//GRANATO D, ARES G. *Mathematical and Statistical Methods in Food Science and Technology*. Chichester: John Wiley & Sons, Ltd, 2014.
- [51] WADE T J, PAI N, EISENBERG J N S, et al. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis[J]. *Environmental Health Perspectives*, 2013, 111(8): 1102-1109.
- [52] CABELLI V J, DUFOUR A P, MCCABE L J, et al. A marine recreational water quality criterion consistent with indicator concepts and risk analysis[J]. *Journal (Water Pollution Control Federation)*, 1983, 55(10): 1306-1314.
- [53] GIBBARD M D, FLYNN C. E. coli prediction tools are on guard at beaches[EB/OL]. (2005-07-12). <https://www.chicagotribune.com/news/ct-xpm-2005-07-12-0507120180-story.html>.
- [54] SHERRY J P. Temporal distribution of faecal pollution indicators and opportunistic pathogens at a Lake Ontario Bathing Beach[J]. *Journal of Great Lakes Research*, 1986, 12(3): 154-160.
- [55] BOEHM A B, KEYMER D P, SHELLENBARGER G G. An analytical model of enterococci inactivation, grazing, and transport in the surf zone of a marine beach[J]. *Water Research*, 2005, 39(15): 3565-3578.
- [56] BOEHM A B, LLUCH - COTA D B, DAVIS K A, et al. Covariation of coastal water temperature and microbial pollution at interannual to tidal periods[J]. *Geophysical Research Letters*, 2004, 31(6): L06309.
- [57] STIDSON R T, GRAY C A, MCPHAIL C D. Development and use of modelling techniques for real - time bathing water quality predictions[J]. *Water and Environment Journal*, 2012, 26(1): 7-8.
- [58] BREIMAN L, FRIEDMAN J, OLSHEN R A, et al. *Classification and regression trees*[M]. Boca Raton: Chapman & Hall/CRC, 2017.
- [59] TIMOFEEV R. *Classification and regression trees (CART) theory and applications*[D]. Berlin: Humboldt University, 2004.
- [60] Steinberg D, Colla P. CART: classification and regression trees[J]. *The top ten algorithms in data mining*, 2009, 9: 179.
- [61] SHAO J. Linear model selection by cross-validation[J]. *Journal of the American Statistical Association*, 1993, 88(422): 486-494.
- [62] DELSOLE T, SHUKLA J. Linear prediction of Indian monsoon rainfall[J]. *Journal of Climate*, 2002, 15(24): 3645-3658.
- [63] CRUMP K S. On summarizing group exposures in risk assessment: is an arithmetic mean or a geometric mean more appropriate?[J]. *Risk Analysis*, 1998, 18(3): 293-297.
- [64] HAAS C N. How to average microbial densities to characterize risk[J]. *Water Research*, 1996, 30(4): 1036-1038.
- [65] PARKHURST D F. Peer reviewed: arithmetic versus geometric means for environmental concentration data[J]. *Environmental Science & Technology*, 1998, 32(3): 92A-98A.
- [66] Cyterski M, Brooks W, Galvin M, et al. *Virtual Beach 3: User's Guide*[J]. United States Environmental Protection Agency, 2013, .
- [67] CAUSSE J, BILLEN G, GARNIER J, et al. Field and modelling studies of *Escherichia coli* loads in tropical streams of montane agro-ecosystems[J]. *Journal of Hydro-environment Research*, 2015, 9(4): 496-507.
- [68] BILLEN G, GARNIER J, HANSET P. Modelling phytoplankton development in whole drainage networks: the RIVERSTRAHLER Model applied to the Seine river system[J]. *Hydrobiologia*, 1994, 289(1/2/3): 119-137.
- [69] BILLEN G, GARNIER J. Nitrogen transfers through the seine drainage network: a budget based on the application of the 'Riverstrahler' model[J]. *Hydrobiologia*, 1999, 410: 139-150.
- [70] RUELLAND D, BILLEN G, BRUNSTEIN D, et al. SENEQUE: a multi-scaling GIS interface to the Riverstrahler model of the biogeochemical functioning of river systems[J]. *Science of the Total Environment*, 2007, 375(1/2/3): 257-273.
- [71] GARNIER J, BILLEN G, HANNON E, et al. Modelling the transfer and retention of nutrients in the drainage network of the Danube River[J]. *Estuarine, Coastal and Shelf Science*, 2002, 54(3): 285-308.
- [72] EDINGER J E, BUCHAK E M. *Numerical hydrodynamics of estuaries*[M]//HAMILTON P, MACDONALD K B. *Estuarine and Wetland Processes: With Emphasis on Modeling*. Boston, MA: Springer, 1980: 115-146.
- [73] FRANCY D S, BRADY A M G, CARVIN R B, et al. Developing and implementing predictive models for estimating recreational water quality at great lakes beaches[R]. New York: US Geological Survey, 2013.