The Development of a Water Quality Forecasting System for Recreational Coastal Bathing Waters in Ireland

Daniel Hawtree¹, John O'Sullivan¹, Gregory O'Hare², Levent Görgü², Conor Muldoon², Wim G. Meijer³, Bartholomew Masterson³, Aurora Gitto³, Malcolm Taylor⁴, and Elaine Mitchell⁴

¹) School of Civil Engineering, University College Dublin
²) School of Computer Science, University College Dublin
³) School of Biomolecular & Biomedical Science, University College Dublin
⁴) Agri-Food and Biosciences Institute, Belfast, United Kingdom
Project Objectives

• European Bathing Water Directive requires the implementation of *early warning systems for bathing waters* which are subject to short-term pollution events.

• Coastal water quality prediction models and alert systems are being developed which aim to provide *short-term forecasts of bathing water*.

• These forecasts are based on the (modeled) relationship between fecal indicator bacteria and *multiple environmental variables*.
Key Project Components

1) **Model Development & Testing:** UCD Civil Engineering

2) **Data & Model Infrastructure:** UCD Computer Science

3) **Water Quality Sampling:** UCD Microbiology & AgriFood and Biosciences Institute (AFBI)
Beaches where water quality, weather and river flow are being monitored by the SWIM Project.
Water Quality History Example @ Newcastle Beach
Environmental Variables from Previous Studies

Most Commonly Used Explanatory Variables
> 10% of Studies

- Rainfall
- Wind
- Turbidity
- Waves
- Water Temperature
- Stream Discharge
- Sunlight
- Date
- Air Temperature
- Tides
- Water Depth
- Conductivity
- Atmospheric Pressure

Statistical Models for Bathing Water Prediction are *Data Hungry*
MÉRA provides an excellent foundation for Bathing Water model development.
MÉRA Grid Points @ Newcastle Beach

- Many MÉRA grid points within a target catchment area
- Provides high spatial & temporal resolution (far exceeding what could be gathered by gauges)
MÉRA Grid Points @ Newcastle Beach

- Precipitation (& soil moisture) used from ALL points within the catchment area.
- Only the point closest to the sampling point is used for the other variables.
MÉRA Grid Points @ Newcastle Beach

- Precipitation (& soil moisture) used from ALL points within the catchment area.
- Only the point closest to the sampling point is used for the other variables.

**MERA Variables**
- Wind direction
- Wind speed
- Atmospheric pressure
- Air Temperature
- Direct Normal Irradiance

**Non-MERA Variables**
- Tides
- Streamflow
- Rain Gauge
- Rain Radar
# Summary of Modelling Approaches

<table>
<thead>
<tr>
<th>APPROACH</th>
<th>DESCRIPTION</th>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Threshold Optimizer</td>
<td>Determines what level of rainfall / streamflow has correlated with past FIB exceedance levels to predict future occurrences.</td>
<td>• Uses readily available data (e.g. rainfall, streamflow)</td>
<td>• Low correlations between single variables &amp; FIB levels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Implemented in Excel</td>
<td>• Does not consider multiple-variable drivers.</td>
</tr>
<tr>
<td>Decision Tree Models</td>
<td>Trains models based on past relationship between environmental variables &amp; FIB concentrations to predict future occurrences.</td>
<td>• Can utilize many variables</td>
<td>• Higher data requirements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can represent non-linear responses.</td>
<td>• Can suffer from “over-training”</td>
</tr>
<tr>
<td>Ensemble Decision Tree Models</td>
<td>Generates probabilistic predictions of FIB concentrations, based on many individual Decision-Tree models.</td>
<td>• Less susceptible to “over-training”</td>
<td>• Driving variables are more difficult to interpret</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Improved predictive power</td>
<td>• Higher data &amp; technical requirements.</td>
</tr>
</tbody>
</table>
1. Model Development

Response Variable

- Historical WQ Samples

Model Training / Testing

Predictor Variable

- MERA Data
1. Model Development

MÉRA Data

Model Training

Decision Tree for IE at Newcastle (2007 to 2014)

Model Testing

Validation Time-Series for Intestinal Enterococci

Log10 IE (cfu/100 ml)

Sample Day

Data
- Modeled
- Observed

2015

2016

2017
2. Model Implementation

Model Training / Testing

Response Variable
Compliance Samples

Model Implementation (Predictions)

Predictor Variable
HARMONIE Data
## 2. Model Implementation

### Predictions are for Saturday, August 31st 2019

<table>
<thead>
<tr>
<th>Location</th>
<th>General Status</th>
<th>IE Related</th>
<th>EC Related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballyholme</td>
<td>Excellent</td>
<td>1.539645765</td>
<td>1.582206637</td>
</tr>
<tr>
<td>Ballywalter</td>
<td>Excellent</td>
<td>1.646271057</td>
<td>1.582206637</td>
</tr>
<tr>
<td>Castlekock</td>
<td>Excellent</td>
<td>1.340914353</td>
<td>1.340914353</td>
</tr>
<tr>
<td>Clogherhead</td>
<td>Excellent</td>
<td>1.094602186</td>
<td>1.097259954</td>
</tr>
<tr>
<td>Enniscrone</td>
<td>Excellent</td>
<td>1.138361069</td>
<td>1.138361069</td>
</tr>
<tr>
<td>Ladyshay</td>
<td>Excellent</td>
<td>1.620812357</td>
<td>1.620812357</td>
</tr>
<tr>
<td>Newcastle</td>
<td>Poor</td>
<td>2.964792230</td>
<td>3.611563157</td>
</tr>
<tr>
<td>Portrush</td>
<td>Excellent</td>
<td>1.281588040</td>
<td>1.776059005</td>
</tr>
<tr>
<td>Waterfoot</td>
<td>Excellent</td>
<td>2.522965162</td>
<td>2.311484358</td>
</tr>
</tbody>
</table>

### Bathing Water Quality Forecasts utilizing HARMONIE Data

- **Ballyholme**: General Status: Excellent, IE Related: 1.539645765, EC Related: 1.582206637
- **Ballywalter**: General Status: Excellent, IE Related: 1.646271057, EC Related: 1.582206637
- **Castlekock**: General Status: Excellent, IE Related: 1.340914353, EC Related: 1.340914353
- **Clogherhead**: General Status: Excellent, IE Related: 1.094602186, EC Related: 1.097259954
- **Enniscrone**: General Status: Excellent, IE Related: 1.138361069, EC Related: 1.138361069
- **Ladyshay**: General Status: Excellent, IE Related: 1.620812357, EC Related: 1.620812357
- **Newcastle**: General Status: Poor, IE Related: 2.964792230, EC Related: 3.611563157
- **Portrush**: General Status: Excellent, IE Related: 1.281588040, EC Related: 1.776059005
- **Waterfoot**: General Status: Excellent, IE Related: 2.522965162, EC Related: 2.311484358
2. Model Implementation: Public Notification

**Website**

- Ballyholme: EXCELLENT @ 28th-Aug
- Ballywalter: GOOD @ 28th-Aug
- Castlerock: EXCELLENT @ 28th-Aug
- Clogherhead: EXCELLENT @ 27th-Aug
- Enniscrone: POOR @ 30th-Aug
- Lady's Bay: POOR @ 29th-Aug
- Newcastle: EXCELLENT @ 30th-Aug
- Portrush (Curran): EXCELLENT@27th-Aug
- Waterfoot: EXCELLENT @ 30th-Aug

*click for map view*

Information based on real water quality test results

**Mobile App**

- EU SWIM Project
- Live Water Quality
- Events
- Support
- Contact Us
- About Us
- Partners

*Mobile app screenshots showing map and test results,*

*POOR* and *EXCELLENT* ratings are indicated.
3. Model Refinement

Response Variable

- Historical WQ Samples
- Compliance Samples

+ additional WQ samples

Predictor Variable

- MERA Data
- HARMONIE Data

+ additional variables

Model Implementation (Predictions)

Model Training / Testing

+ additional variables
Key Challenge

  - **Total Water Quality Samples**: 560 to 130 (most sites ~ 300)
  - **Poor Water Quality Samples**: 40 to 2 (most sites ~ 20 to 30)
  - Relatively high proportion of non-meteorologically driven “Poor” samples (~ 20% to 30% at some sites)

- Impact:
  - Too few samples to adequately train the model at some sites.
  - Model is highly sensitive to the train / test split at other locations.
  - Model is confounded by non-meteorologically driven events.
    - Dogs, Birds, Horses, etc...
Non-Meteorologically Driven WQ Failures can’t be predicted (by this type of model)

Source of Contamination?

The Usual Suspects

Model Development – Next Steps

Multi-Model Development Framework

- A wide range of non-linear classification and tree-based methods are available which can utilize multi-variate data (e.g. MERA, rain radar, tide).
- A framework for training and testing multiple different models in parallel is under development – utilizing the “Caret” package in R, which contains ~ 240 different machine learning models.

No Free Lunch Theorem

“There is no such thing as a single, universally-best machine learning algorithm, and there are no context or usage-independent (a priori) reasons to favor one algorithm over all others.”
<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Variable Type</th>
<th>Location</th>
<th>Temporal Aggregation*</th>
<th>Variables @ Newcastle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>mm</td>
<td>Numeric</td>
<td>Catchment</td>
<td>Sum</td>
<td>138**</td>
</tr>
<tr>
<td>Soil Moisture</td>
<td>kg/m³</td>
<td>Numeric</td>
<td>Catchment</td>
<td>Mean</td>
<td>138**</td>
</tr>
<tr>
<td>Temperature</td>
<td>°C</td>
<td>Numeric</td>
<td>Sample Point</td>
<td>Mean</td>
<td>6</td>
</tr>
<tr>
<td>Atmospheric Pressure</td>
<td>kPA</td>
<td>Numeric</td>
<td>Sample Point</td>
<td>Mean</td>
<td>6</td>
</tr>
<tr>
<td>Direct Normal Irradiance</td>
<td>kW/m²</td>
<td>Numeric</td>
<td>Sample Point</td>
<td>Sum</td>
<td>6</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>Beaufort scale</td>
<td>Categorical</td>
<td>Sample Point</td>
<td>Mean</td>
<td>6</td>
</tr>
<tr>
<td>Wind Direction</td>
<td>Cardinal Direction</td>
<td>Categorical</td>
<td>Sample Point</td>
<td>Mode</td>
<td>6</td>
</tr>
</tbody>
</table>

* Data was aggregated over periods of 1, 6, 12, 24, 48, and 96 hours from the time of the sample.

** 22 MERA Points in Newcastle Catchment + 1 Catchment Mean x 6 Time Aggregations = 138 variables
There is typically a trade-off between model sensitivity and specificity, and increasing one results in a decrease in the other.

Bathing Water Quality models typically achieve high specificity, while high sensitivity is more difficult to achieve. This is due to the relatively low frequency of WQ failures (at most sites), complex driving conditions, and the occurrence of non-meteorological drivers.
# Model Performance Standards: Sensitivity & Specificity

<table>
<thead>
<tr>
<th>Source</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thoe et al. (2014)</strong></td>
<td>&gt;30%</td>
<td>&gt;80%</td>
</tr>
<tr>
<td>“Predicting water quality at Santa Monica Beach: Evaluation of five different models for public notification of unsafe swimming conditions”</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>California’s “Nowcast” System</strong></td>
<td>&gt;50%</td>
<td>&gt;85%</td>
</tr>
<tr>
<td><a href="https://beachreportcard.org/">https://beachreportcard.org/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Scottish EPA</strong></td>
<td>&gt;50%</td>
<td>-</td>
</tr>
<tr>
<td><em>R. Stidson, personal communication</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UK Environment Agency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>D. Tyrell, personal communication</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Following the levels set out by these standards, in our models we seek to maximize *sensitivity* while maintaining a minimum *specificity* of 0.80.
### Example of Model Performance at Different Thresholds

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>19 mm 2 days</td>
<td>0.17</td>
<td>0.98</td>
</tr>
<tr>
<td>8 mm 1 day</td>
<td>0.37</td>
<td>0.95</td>
</tr>
<tr>
<td>6 mm 1 day</td>
<td>0.49</td>
<td>0.89</td>
</tr>
<tr>
<td>4 mm 1 day</td>
<td>0.51</td>
<td>0.87</td>
</tr>
<tr>
<td>7 mm 2 day</td>
<td>0.57</td>
<td>0.82</td>
</tr>
</tbody>
</table>

More Conservative – Less Conservative