Model-based analysis of the potential of macroinvertebrates as indicators for microbial pathogens in rivers

Rubén Jerves-Cobo, I. Nopens, P. Goethals

Ruben.JervesCobo@UGent.Be

13 February 2018
1-BACKGROUND AND OBJECTIVE (I)

• The quality of the water must accomplish standards.
  • drinking water, recreational purpose, Irrigation

• The indicators used to verify microbial contamination of water are:
  • total coliforms and fecal coliforms and/or 
  Escherichia coli

Netherlands Annual Ecology Meeting 2018 (NAEM2018), Lunteren, The Netherlands, Date (13/02/2018)
Department of Applied Ecology and Environmental Biology, Research Group Aquatic Ecology (AECO)
Ruben.JervesCobo@UGent.be
1-BACKGROUND AND OBJECTIVE (II)

• Checking the fulfillment requires expensive and highly trained personnel in laboratories

• Biota works as a permanent monitor of water

• Biological samples can:
  • reflect an increase in pollution.
  • predict average values of chemical parameters
Objective

- Introduce a quick way of checking the fulfillment of fecal coliforms standards using macrobenthos.
2- MATERIALS AND METHODS

LOCATION

[Map showing the location of Machangara River basin in Ecuador, with markers for Quito, Guayaquil, and Cuenca.]

Legend:
- Waterbody
- River
- Ravine

South America

0 2.5 5 10 15 20 Km

Ecuador
LAND USE
DATA COLLECTION:
- Completed information on 33 locations

Physicochemical, hydraulic, microbiological

- **Laboratory**
  - \( \text{BOD}_{5} \), COD, Nitrate + Nitrite, Ammonia, Organic Nitrogen, Phosphates, Total Phosphorus, Fecal and Total Coliforms, Real Color, Turbidity, Total Solids

- **Field**: Flow Velocity, Ph, Conductivity, Temperature, Dissolved Oxygen

**Macrobenthos**
- 39 families (taxa) found
Variables Variation

Boxplot of Flow_Velocity

Boxplot of Dissolved_Oxygen

Boxplot of BOD5

Boxplot of pH

Boxplot of Nitrate

Boxplot of Log_Fecal_Coliforms
- Ecuadorian Water Quality Regulation for Fecal Coliforms

<table>
<thead>
<tr>
<th>Regulations</th>
<th>Water used to</th>
<th>Fecal Coliforms Limited Value MPN/100 ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recreational</td>
<td>Recreational with primary contact</td>
<td>≤200</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Agriculture and Livestock</td>
<td>≤1,000</td>
</tr>
<tr>
<td>Raw water</td>
<td>raw water previous to non-conventional treatment*</td>
<td>≤2,000</td>
</tr>
</tbody>
</table>

* Conventional treatment refers to chemical addition, rapid mixing, flocculation and sedimentation

Netherlands Annual Ecology Meeting 2018 (NAEM2018), Lunteren, The Netherlands, Date (13/02/2018)
Department of Applied Ecology and Environmental Biology, Research Group Aquatic Ecology (AECO)
Ruben.JervesCobo@UGent.be
**BIOLOGICAL WATER QUALITY**

- Biological Monitoring Working Party Index - Col
- BMWP-Col = f(Sensitivity of Macrobenthos)
- Sensitivity -> 1-10 (Low – High Sensitivity)

<table>
<thead>
<tr>
<th>Class</th>
<th>Quality</th>
<th>BMWP</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Very Good</td>
<td>&gt; 100</td>
<td>Light Blue</td>
</tr>
<tr>
<td>II</td>
<td>Good</td>
<td>61 - 100</td>
<td>Green</td>
</tr>
<tr>
<td>III</td>
<td>Moderate</td>
<td>36 - 60</td>
<td>Yellow</td>
</tr>
<tr>
<td>IV</td>
<td>Deficient</td>
<td>16 - 35</td>
<td>Orange</td>
</tr>
<tr>
<td>V</td>
<td>Bad</td>
<td>≤ 15</td>
<td>Red</td>
</tr>
</tbody>
</table>

Netherlands Annual Ecology Meeting 2018 (NAEM2018), Lunteren, The Netherlands, Date (13/02/2018)
Department of Applied Ecology and Environmental Biology, Research Group Aquatic Ecology (AECO)
Ruben.JervesCobo@UGent.be
Model development

Ecosystem

Training dataset

Validation dataset
Methodology
Pruned Multi-target Clustering Trees (PMCT)

- Classification trees: searching for if-then rules (threshold values): 100% reliable and ‘safe’ models
Model Performance

- Models must be evaluated based on statistical and ecological criteria.
- Models must be as clear and simple as possible.

Settings

- Machine learning software: Waikato Environment for Knowledge Analysis (WEKA)
- Three, five, ten-fold cross validation (k fcv)
- Pruning process
  - Pruning confidence factors (PCF): 0.25, 0.10
Model Performance

- Confusion matrix from Decision Tree Models:

<table>
<thead>
<tr>
<th>Observed Class</th>
<th>Predicted Class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>TP</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>FN</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>FP</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>TN</td>
</tr>
</tbody>
</table>

- Correctly Classified Instances (CCI): > 0.70
- Cohen’s Kappa Statistic: > 0.40
- False positive (FP = 0)
Model Optimization:

- **Cost sensitive classifier (CSC):** gives new weights in training instances FN, FP

- **Overall confusion entropy of a confusion matrix (CEN):** evaluates the confusion level of the class distribution of misclassified samples:

\[ CEN = (P_1 + P_2)CEN_j \]

\[ P_1 = \frac{TP + FN}{2(TP + FN + FP + TN)} \quad \text{and} \quad P_2 = \frac{FP + TN}{2(TP + FN + FP + TN)} \]
Model Optimization:

\[ CEN_j = -P_{FN} \log_2 P_{FN} - P_{FP} \log_2 P_{FP} \]

- \( P_j \): confusion probability of class \( j \)
- \( CEN_j \): confusion entropy of class \( j \).

\[ P_{FP} = \frac{FP}{FN + FP + 2TP} \quad \text{and} \quad P_{FN} = \frac{FN}{FN + FP + 2TN} \]

- \( P_{FP} \) and \( P_{FN} \) are the misclassification probability of classifying the samples of class \( i \) to class \( j \) subject to class \( j \)
- Higher accuracy corresponds to lower confusion entropy
3- RESULTS

BIOLOGICAL WATER QUALITY

BMWP-Col:
- 9 good
- 15 moderate
- 6 poor
- 3 bad
Fulfillment of Fecal Coliforms limits in relationship to water use

(a) Recreational with primary contact
(b) Agricultural - Livestock use and (c) raw water
Model Performance

**First Model:** Recreational water use with Primary Contact – Fecal Coliforms Regulation.

CCI = 77%
Kappa = 0.56
FP = 0
CEN = 0.64
Model Optimization (1)

Second Model: Recreational water use with Primary Contact – Fecal Coliforms Regulation.

CCI = 79%
Kappa = 0.58
FP = 0
CEN = 0.61
Model Optimization (2)

Third model: Irrigation use and raw water previous to non-conventional treatment use - Fecal Coliforms Regulation.

CCI = 87%
Kappa = 0.69
FP = 0
CEN = 0.53
4- CONCLUSION

- Three models were selected.
  • Two Models: fecal coliforms threshold in recreational with primary contact water use.
  • One Model: fecal coliforms limits in agricultural livestock water use, or raw water for drinking water treated with non-conventional processes.

- The cost-sensitive classifier (CSC) in the Weka can reduce false positives (FP) in the confusion matrix, improving the reliability of the resulting models.
Acknowledgement

- VLIR-UOS IUC Programme - University of Cuenca
- VLIR Ecuador Biodiversity Network Project
- Council of the Machangara River Basin
Thank you