RESEARCH GROUP
AQUATIC ECOLOGY

# 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

GHENT
UNIVERSITY

UNIVERSIDAD DE CUENCA
desde 1867

# OUTLINE

Background and objective

Materials and methods

Results

Conclusion

Department of Applied Ecology and Environmental Biology, Research Group Aquatic Ecology (AECO)
Ruben.JervesCobo@UGent.be

# 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

# 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.

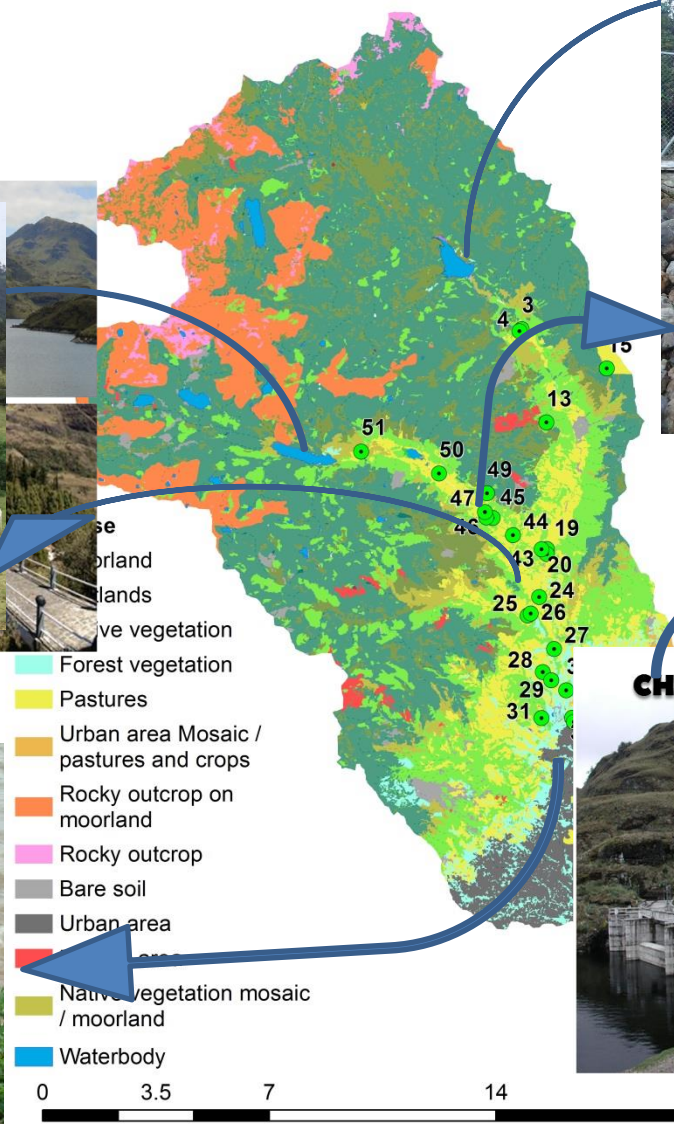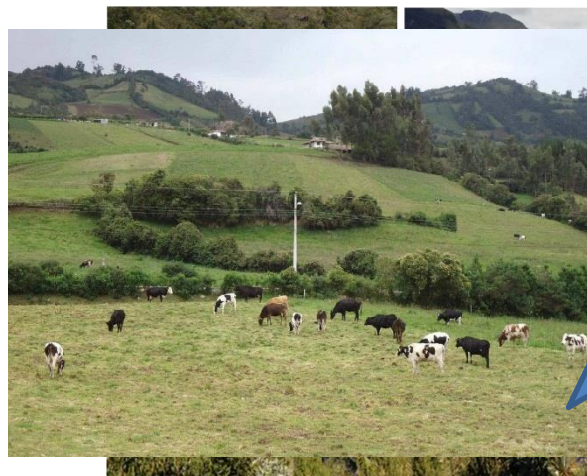  - predict average values of chemical parameters

# Objective

-   Introduce a quick way of checking the fulfillment of fecal coliforms standards using macrobenthos.

## LOCATION

# LAND USE
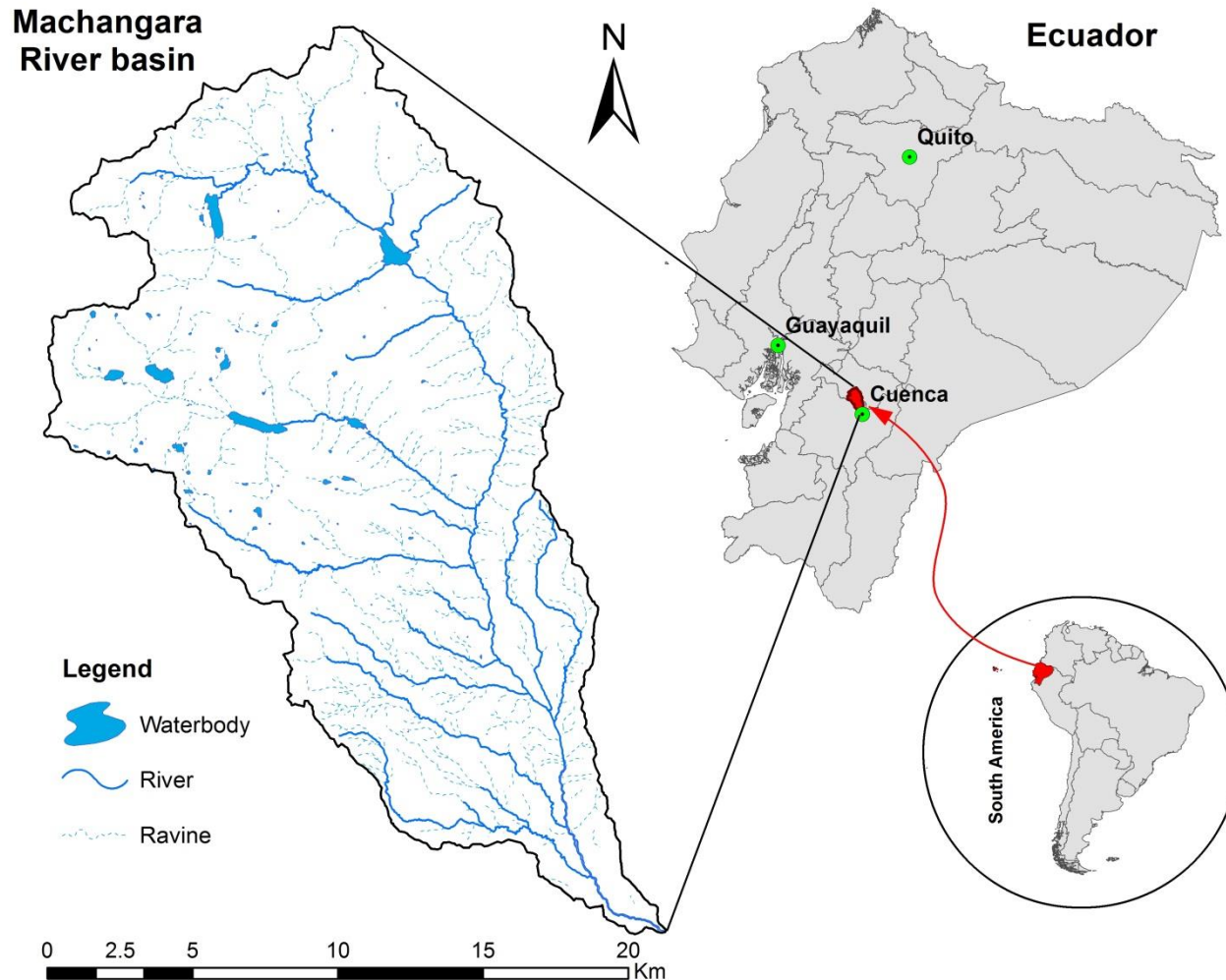


Land use legend:
- Moorland
- [grass]lands
- [Nati]ve vegetation
- Forest vegetation
- Pastures
- Urban area Mosaic / pastures and crops
- Rocky outcrop on moorland
- Rocky outcrop
- Bare soil
- Urban area
- [___ area]
- Native vegetation mosaic / moorland
- Waterbody

0   3.5   7   14

CHA

GHENT UNIVERSITY

UNIVERSIDAD DE CUENCA
desde 1867

RESEARCH GROUP
AQUATIC ECOLOGY

# DATA COLLECTION:

- Completed information on 33 locations

## Physicochemical, hydraulic, microbiological

- **Laboratory**
  - $BOD_5$, COD, Nitrate + Nitrite, Ammonia, Organic Nitrogen, Phosphates, Total Phosphorus, Fecal and Total Coliforms, Real Color, Turbidity, Total Solids

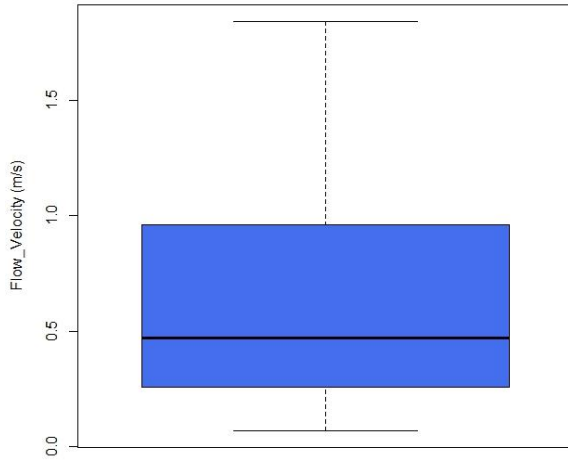- **Field:** Flow Velocity, Ph, Conductivity, Temperature, Dissolved Oxygen
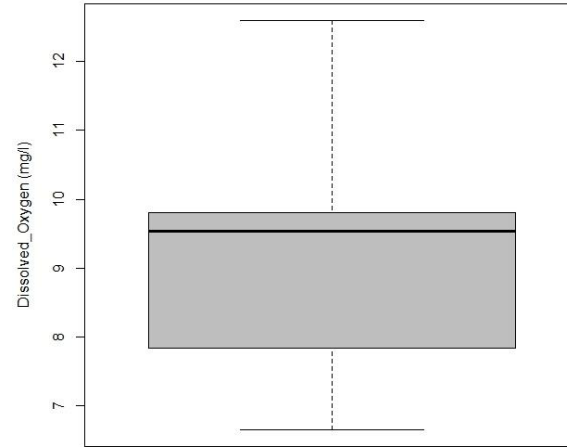
## Macrobenthos

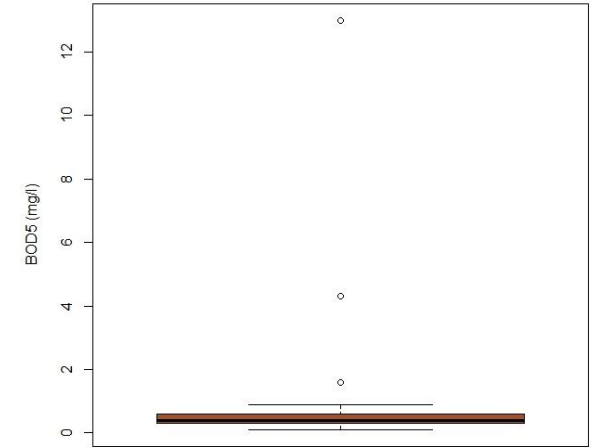39 families (taxa) found

# Variables Variation
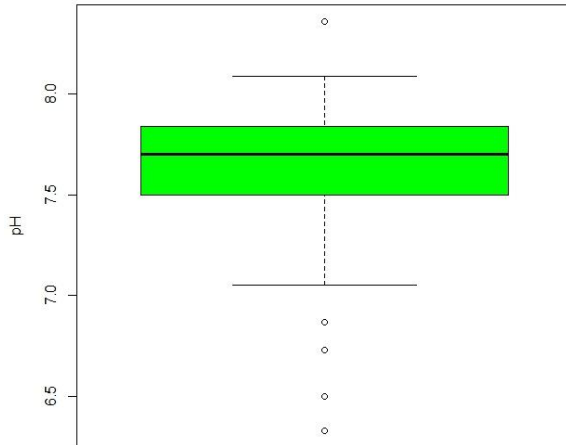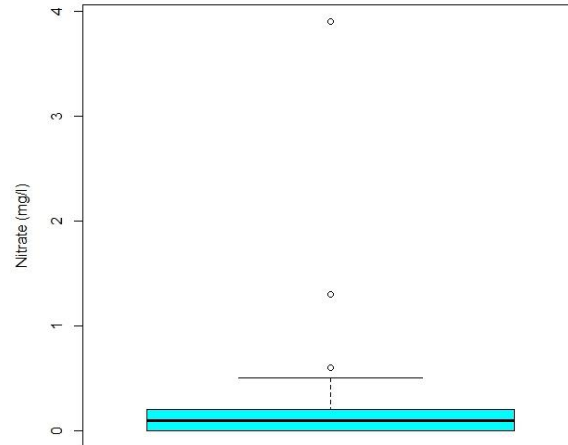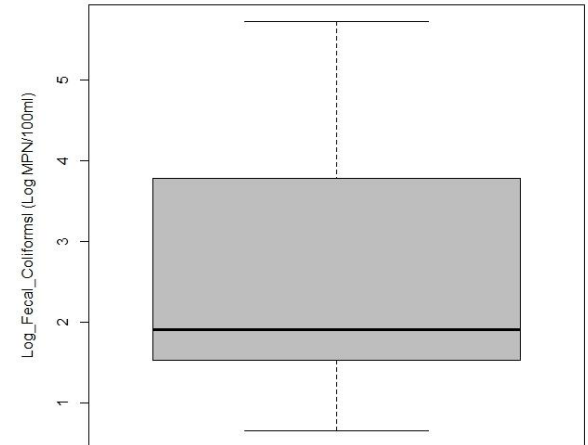
# - Ecuadorian Water Quality Regulation for Fecal Coliforms

| Regulations | Water used to | Fecal Coliforms Limited Value MPN/100 ml |
|---|---|---|
| Recreational | Recreational with primary contact | ≤200 |
| Agriculture | Agriculture and Livestock | ≤1,000 |
| Raw water | raw water previous to non-conventional treatment* | ≤2,000 |
| * Conventional treatment refers to chemical addition, rapid mixing, flocculation and sedimentation | | |

# BIOLOGICAL WATER QUALITY

- Biological Monitoring Working Party Index - Col

- BMWP-Col = f(Sensitivity of Macrobenthos)

- Sensitivity -> 1-10 (Low – High Sensitivity)

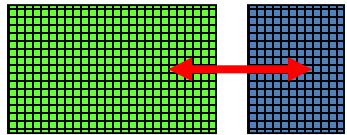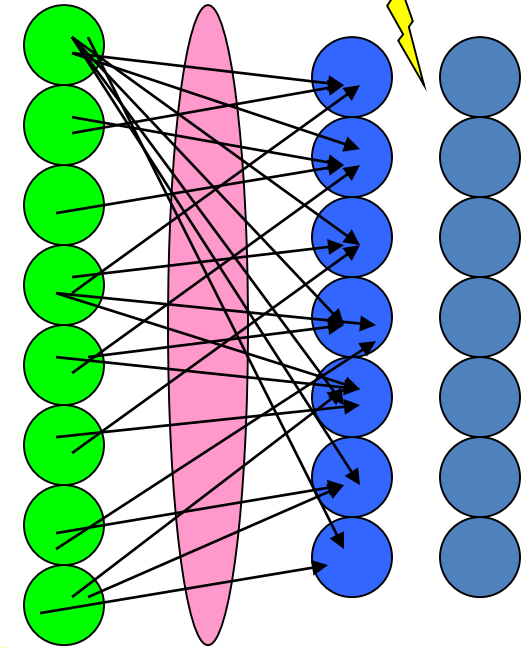| Class | Quality | BMWP | Color |
|---|---|---|---|
| I | Very Good | > 100 | |
| II | Good | 61 - 100 | |
| III | Moderate | 36 - 60 | |
| IV | Deficient | 16 - 35 | |
| V | Bad | ≤ 15 | |

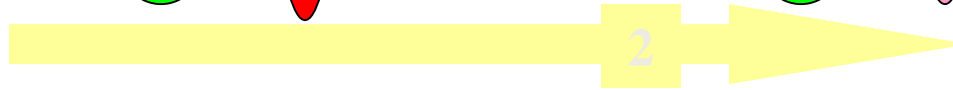# Model development

Ecosystem
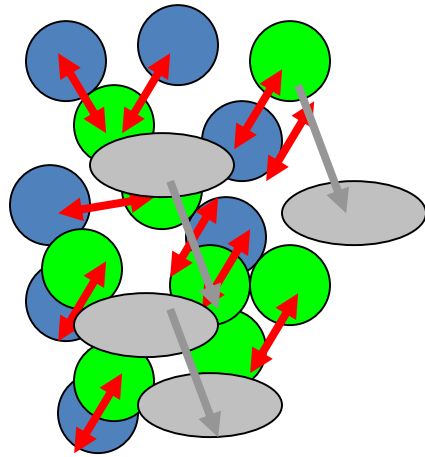
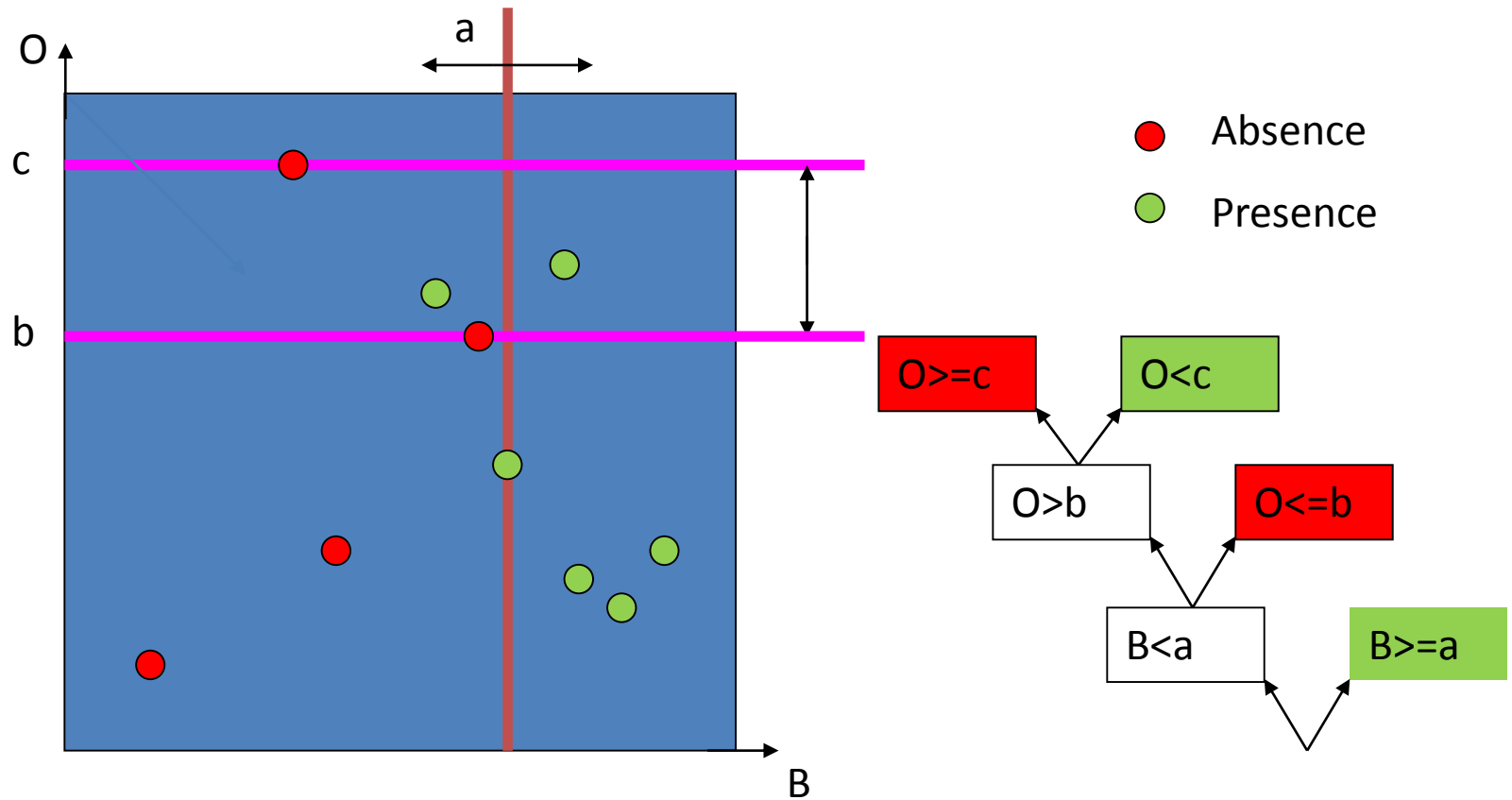Training dataset
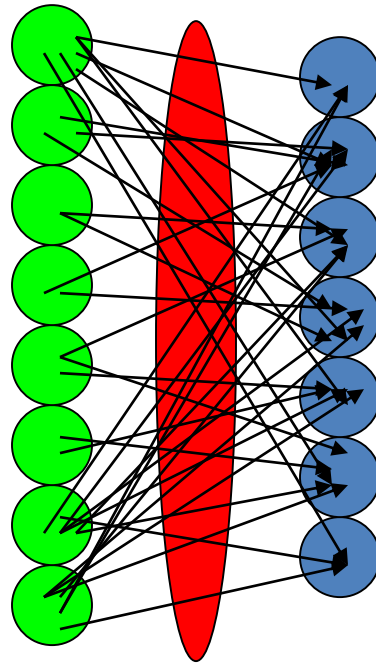
Validation dataset

Measurement set

# 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:

|  |  | Predicted Class | |
|---|---|---|---|
|  |  | Yes | No |
| **Observed Class** | Yes | TP | FN |
|  | No | FP | TN |

- 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:

$$CEN = (P_1 + P_2)CEN_j$$

$$P_1 = \frac{TP + FN}{2(TP + FN + FP + TN)} \ and \ P_2 = \frac{FP + TN}{2(TP + FN + FP + TN)}$$
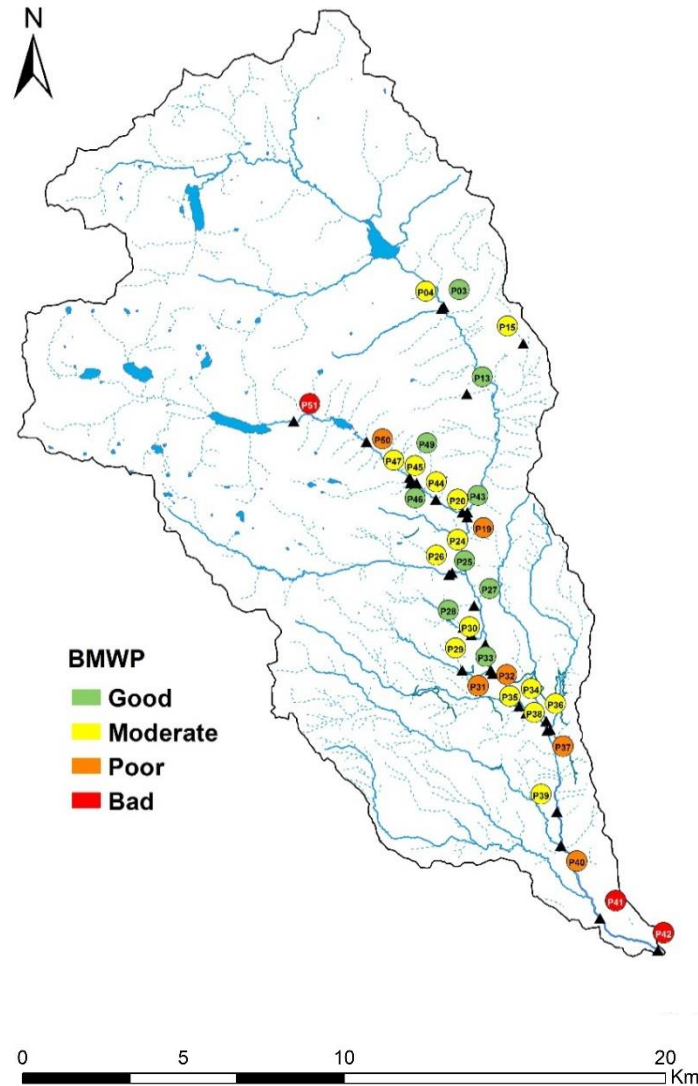
# Model Optimization:

$$CEN_j = -P_{FN}log_2P_{FN} - P_{FP}log_2P_{FP}$$

- $P_j$: confusion probability of class $j$
- $CEN_j$: confusion entropy of class $j$.

$$P_{FP} = \frac{FP}{FN + FP + 2TP} \ and \ P_{FN} = \frac{FN}{FN + FP + 2TN}$$

- $P_{FP}$ and $P_{FN}$ are the misclassification probability of classifying the samples of class $i$ to class $j$ subject to class $j$

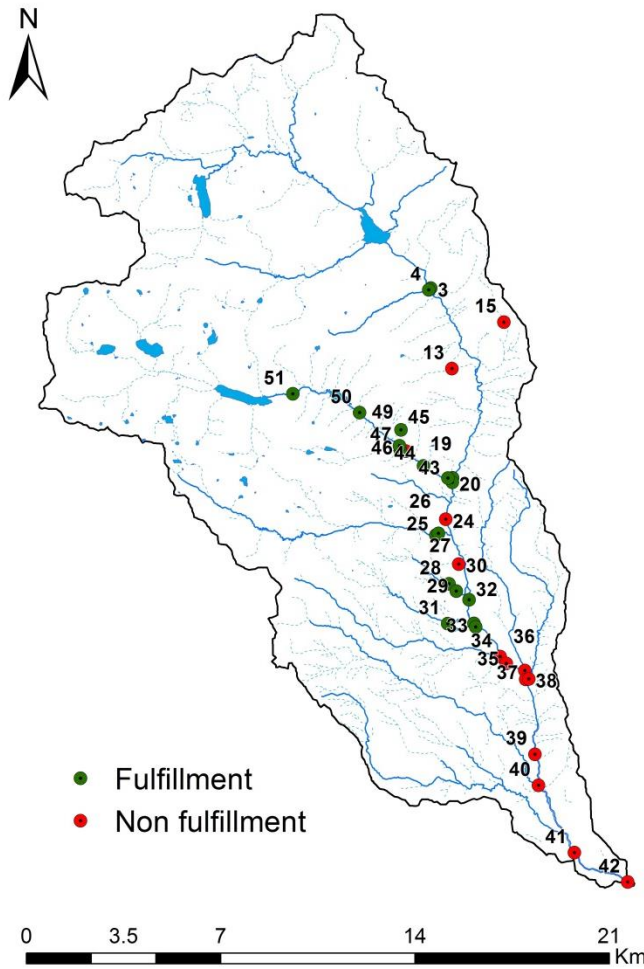- Higher accuracy corresponds to lower confusion entropy
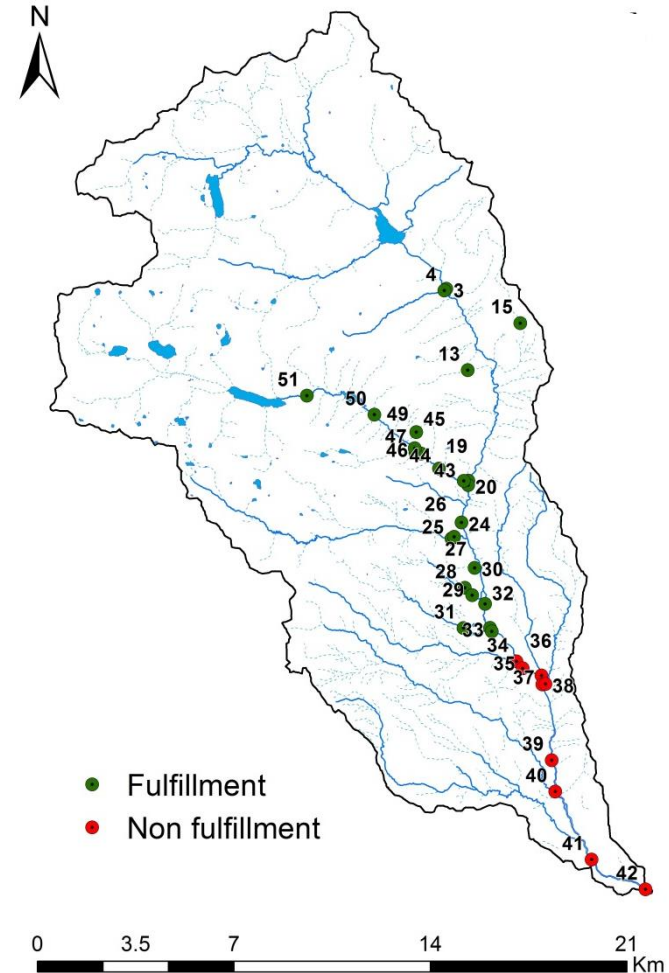
**BIOLOGICAL WATER QUALITY**

BMWP-Col:

- 9 good

- 15 moderate

- 6 poor

- 3 bad

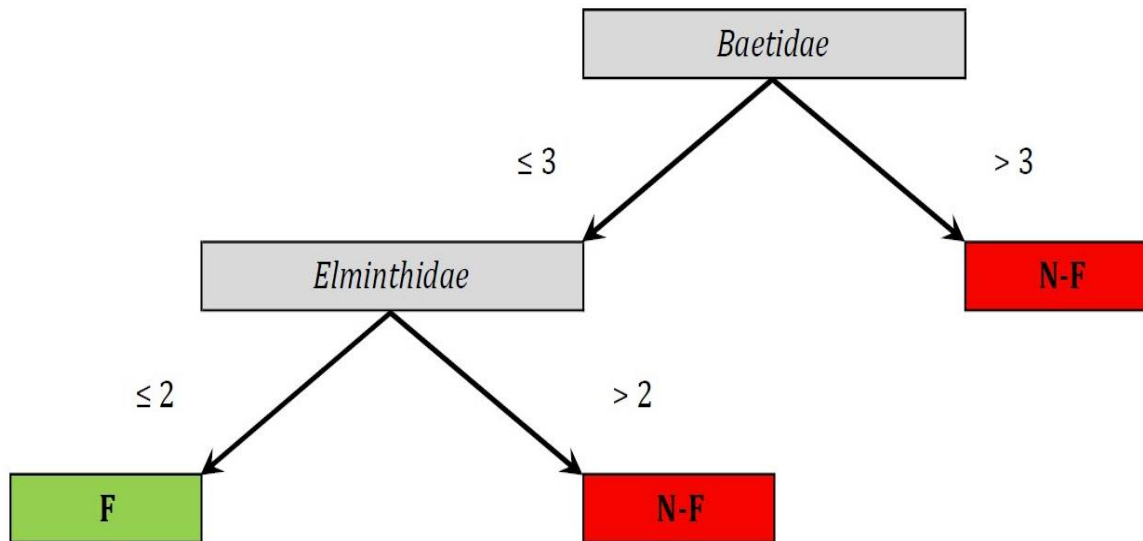# Fulfillment of Fecal Coliforms limits in relationship to water use



(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.*
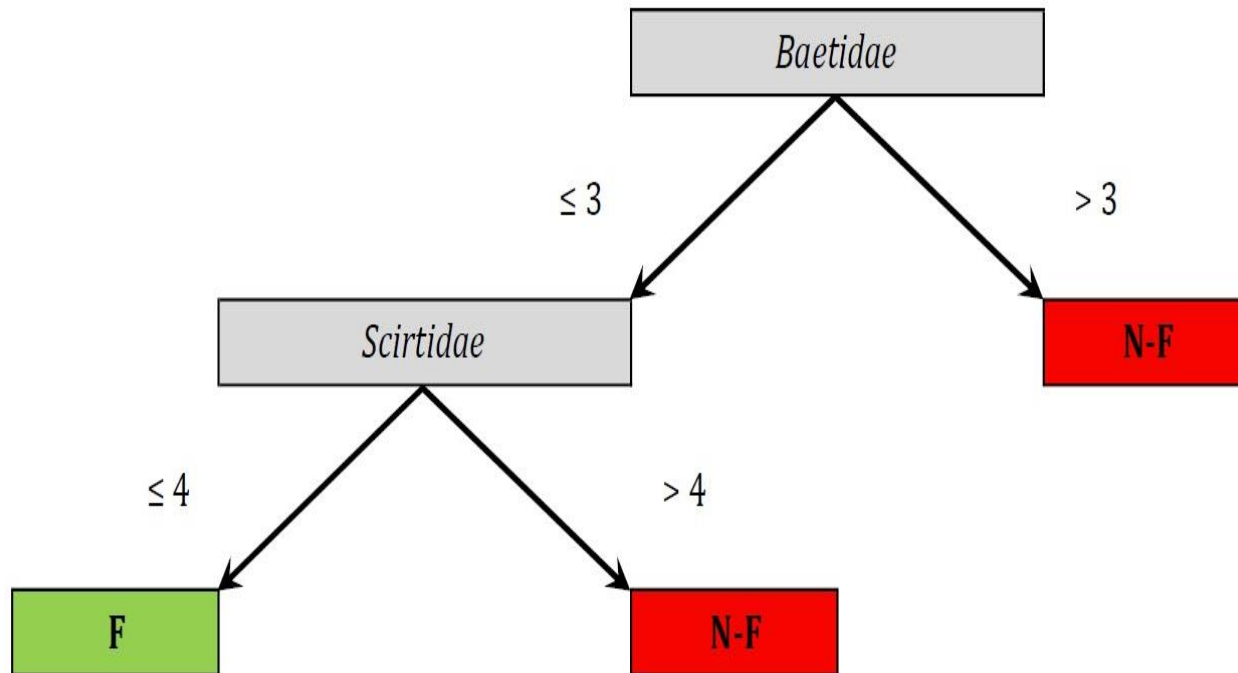


CCI = 77%

Kappa = 0.56

FP = 0

CEN = 0.64

# Model Optimization (1)
**Second Model:** *Recreational water use with Primary Contact – Fecal Coliforms Regulation.*
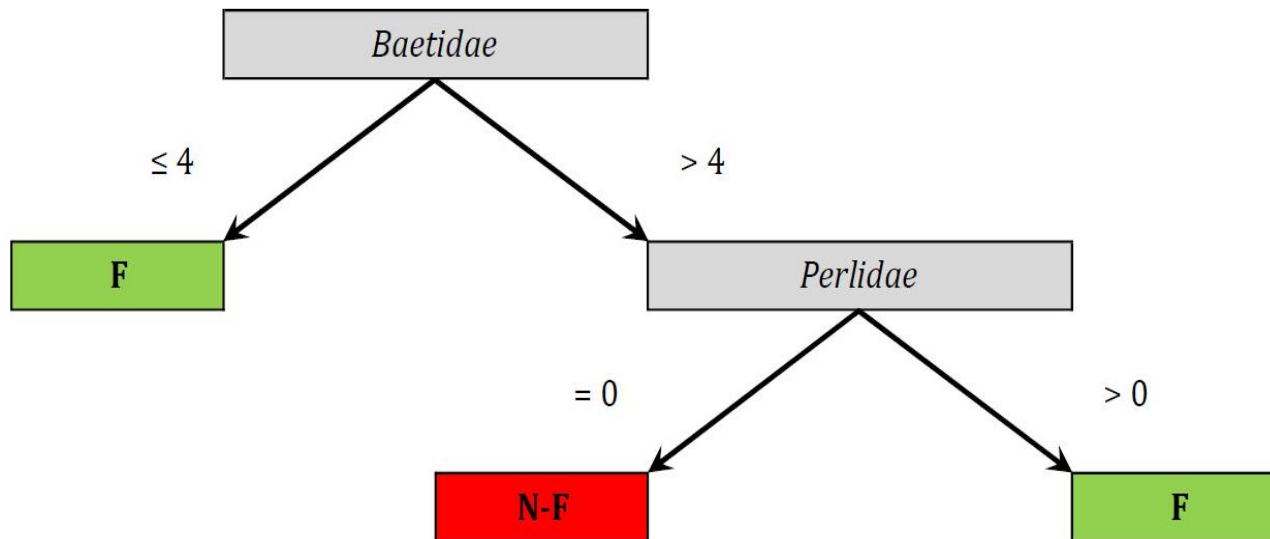


CCI = 79%

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

Department of Applied Ecology and Environmental Biology, Research Group Aquatic Ecology (AECO)
Ruben.JervesCobo@UGent.be

# Thank you

# ?