

Using chemometric methods to identify an adequate biotic index for monitoring the Paute river basin (Ecuador)

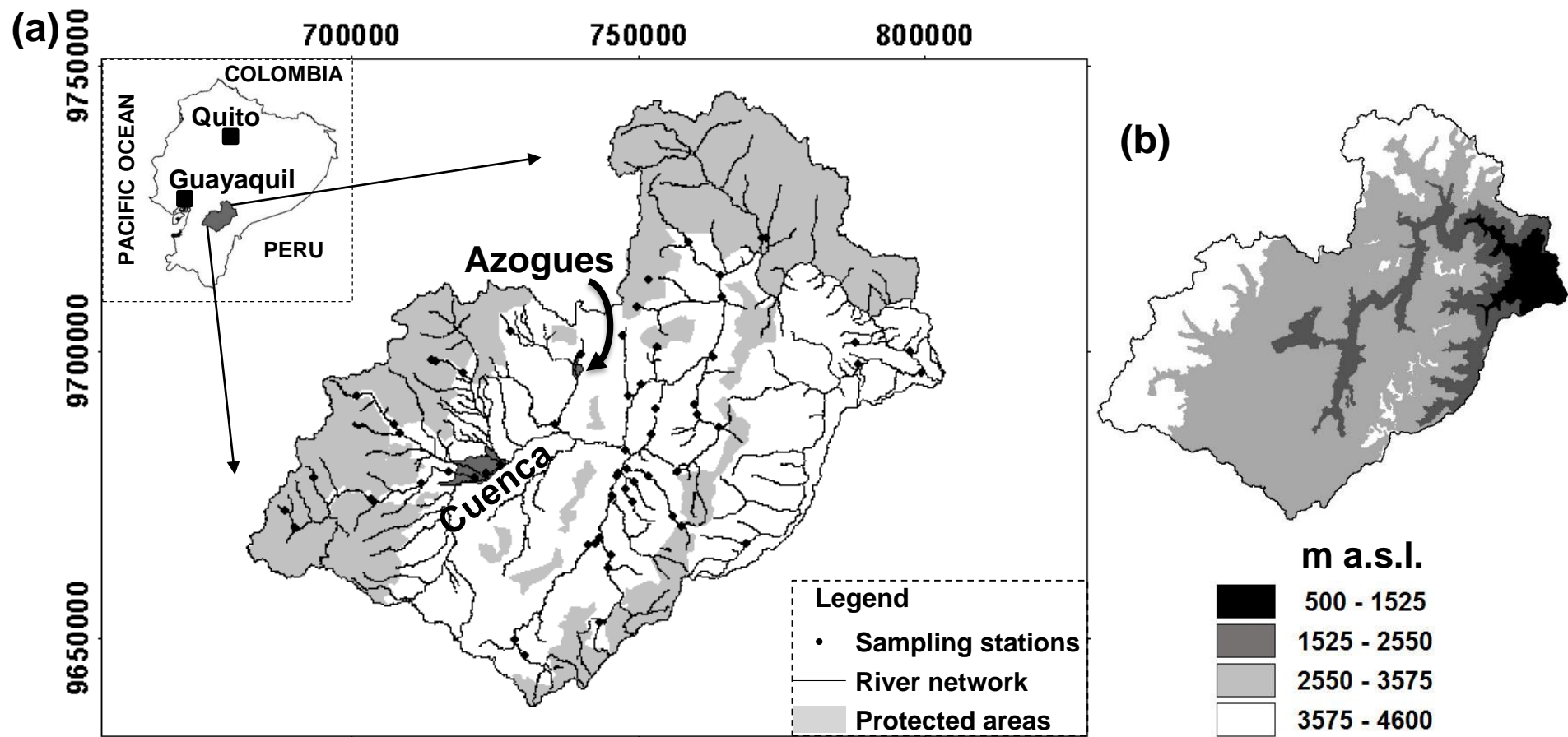
Gonzalo Sotomayor, Henrietta Hampel,
Raúl F. Vázquez, Peter Goethals

IWA-IDB INNOVATION CONFERENCE ON SUSTAINABLE USE OF
WATER: Cities, Industry and Agriculture
Guayaquil - Ecuador, October 2019

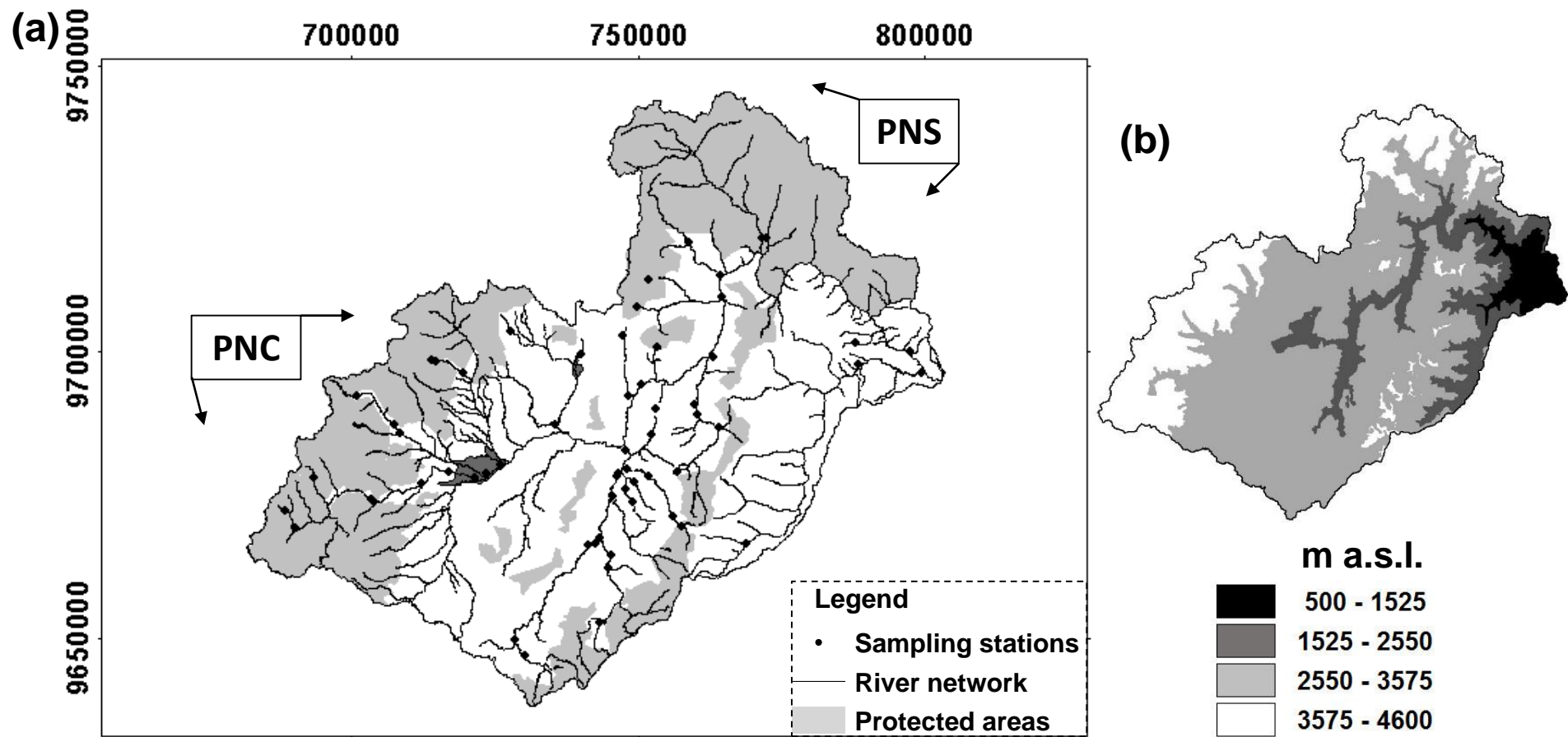


Co-funded by the
Erasmus+ Programme
of the European Union





(a) Paute River basin (PRB) in continental Ecuador, its two largest cities (Quito and Guayaquil) and the location of the 64 water monitoring stations (sampling sites); (b) a Digital Elevation Model (DEM) of the study basin.



The PRB includes important extents of protected areas. → the Cajas National Park (PNC), which is a Ramsar-Convention (RAMSAR) wetland site, and the Sangay National Park (PNS), both recognised by the UNESCO as World Heritage Sites.

BACKGROUND

- 27 WQ descriptive variables were sampled in 64 monitoring stations aiming at representing the WQ distribution in the study basin (SENAGUA, 2016).
- Benthic macroinvertebrates were sampled in 64 monitoring stations and identified to the family level.
- Also, the fluvial habitat index of the Environmental Protection Agency (FHI-EPA) was assessed.
 - The FHI-EPA is focused on the visual assessment of streambed and riparian habitat, the alteration of which is considered one of the major stressors of aquatic systems.

WQ Parameter	Mean	Median	STD	Range
Al (mg L ⁻¹)	0.06	0.00	0.23	0.00 - 1.59
BOD ₅ (mg L ⁻¹)	10.30	9.98	8.08	0.00 - 55.82
Cd (mg L ⁻¹)	0.01	0.00	0.05	0.00 - 0.61
CL (mg L ⁻¹)	5.26	0.73	27.32	0.00 - 363.86
Cu (mg L ⁻¹)	0.02	0.00	0.12	0.00 - 1.32
DO (mg L ⁻¹)	6.83	6.89	0.75	4.12 - 9.75
EC (μS cm ⁻¹)	122.43	70.00	197.36	2.96 - 1810.00
Elev (m a.s.l.)	2419.07	2420.00	730.01	480.00 - 3780.00
FC (bacteria 100 ⁻¹ ml ⁻¹)	5038.32	1600.00	6451.66	1.00 - 16000.00
Fe (mg L ⁻¹)	0.21	0.00	0.49	0.00 - 3.73
FHI-EPA	129.49	128.00	28.54	71.00 - 184.00
FL (mg L ⁻¹)	1.56	0.42	6.83	0.00 - 67.89
K (mg L ⁻¹)	1.63	0.37	5.29	0.00 - 69.86
N-NH ₄ (mg L ⁻¹)	0.78	0.00	1.57	0.00 - 15.00
Na (mg L ⁻¹)	5.28	3.37	9.15	0.00 - 112.89
Ni (mg L ⁻¹)	0.02	0.00	0.14	0.00 - 1.51
N-NO ₃ (mg L ⁻¹)	0.56	0.13	2.05	0.00 - 20.79
Pb (mg L ⁻¹)	0.02	0.00	0.07	0.00 - 0.85
pH	7.52	7.56	0.65	5.30 - 9.43
P-tot (mg L ⁻¹)	0.38	0.16	0.50	0.00 - 2.09
Shreve	334.15	51.00	1048.84	1.00 - 5760.00
Slp (%)	25.82	10.31	35.05	0.00 - 142.30
TALK (mg L ⁻¹)	0.74	0.08	1.35	0.00 - 7.80
TH (mg L ⁻¹)	33.34	23.60	36.50	0.00 - 263.56
TS (mg L ⁻¹)	2.00	0.01	10.42	0.00 - 116.00
TU (NTU)	19.48	0.92	88.16	0.00 - 1136.81
WT (°C)	14.57	14.00	3.27	8.70 - 23.50

A Shapiro-Wilk test was applied → none of the WQ descriptive variables (with exception of pH) were normally distributed

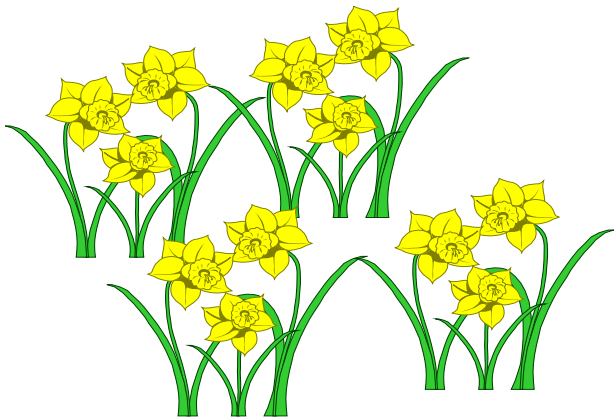
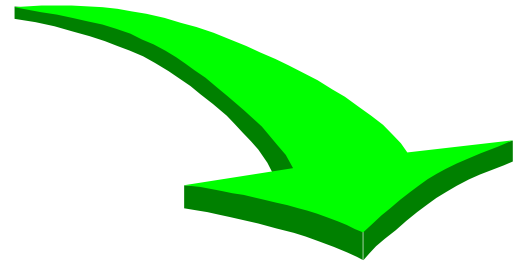
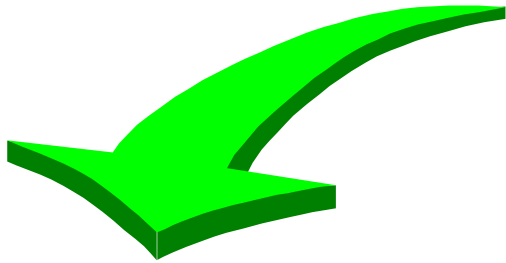
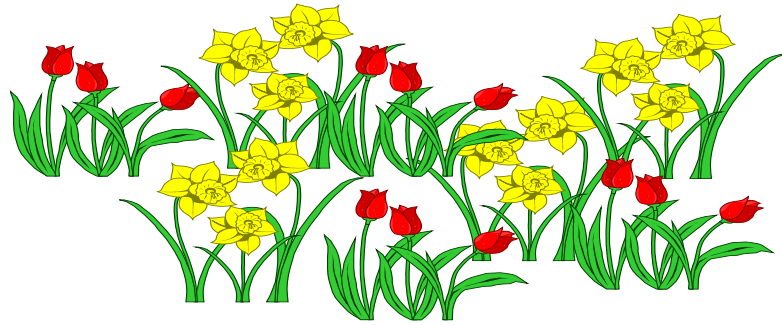
BENTHIC MACROINVERTEBRATES

- $n_{\text{fam}} = 65$ families were identified and grouped into $n_{\text{ord}} = 19$ superior taxonomic groups (in its great majority orders).
- 9 biotic indices (Bx) were calculated using n_{fam} and n_{ord} .
- Each Bx was calculated for each of the n_{rep} replicates

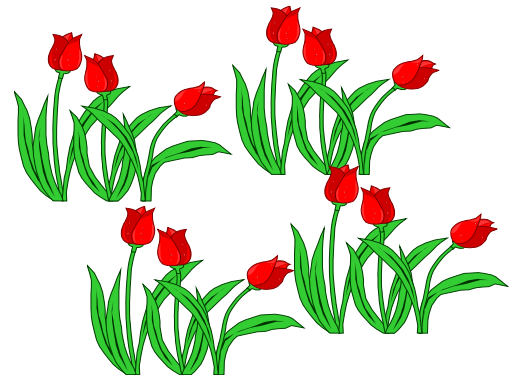


1. Biological Monitoring Working Party (BMWP) (Armitage et al., 1983), calibrated for Colombia (BMWP_Col)
2. Andean Biotic Index (ABI)
3. A combined ABI_BMWP_Col index, namely, (i) ABI for streams located above 2000 m a.s.l.; and (ii) BMWP_Col for lower elevations (< 2000 m a.s.l.)
4. Ephemeroptera-Plecoptera-Trichoptera (EPT)
5. Elmidae-Plecoptera-Trichoptera (ElmPT)
6. Average Score Per Taxon (ASPT)
 - ASPT_BMWP_Col
 - ASPT_ABI
 - ASPT_ABI_BMWP_Col
 - ABI_BMWP_Col index

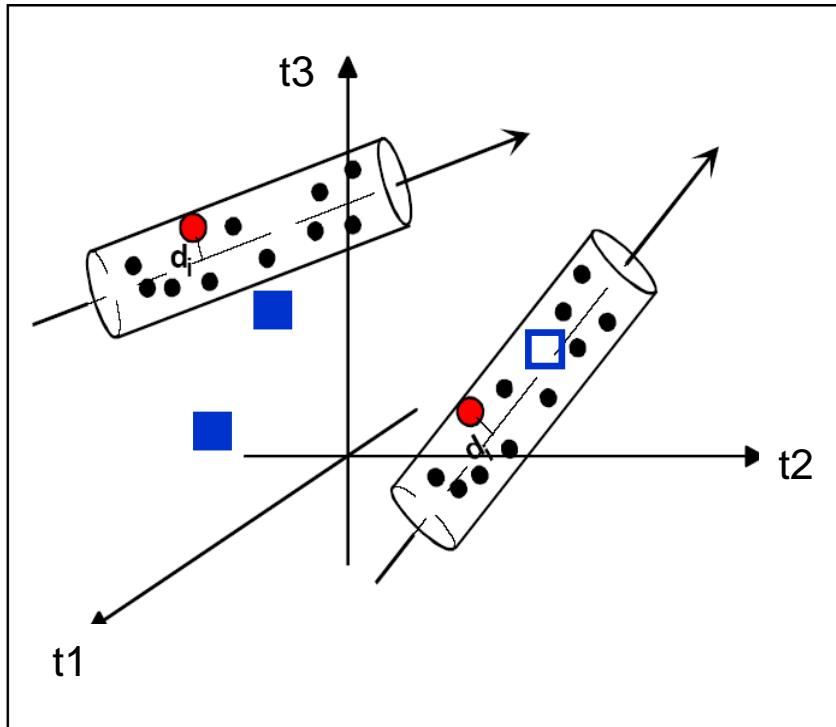
Framework



**Classification
models**



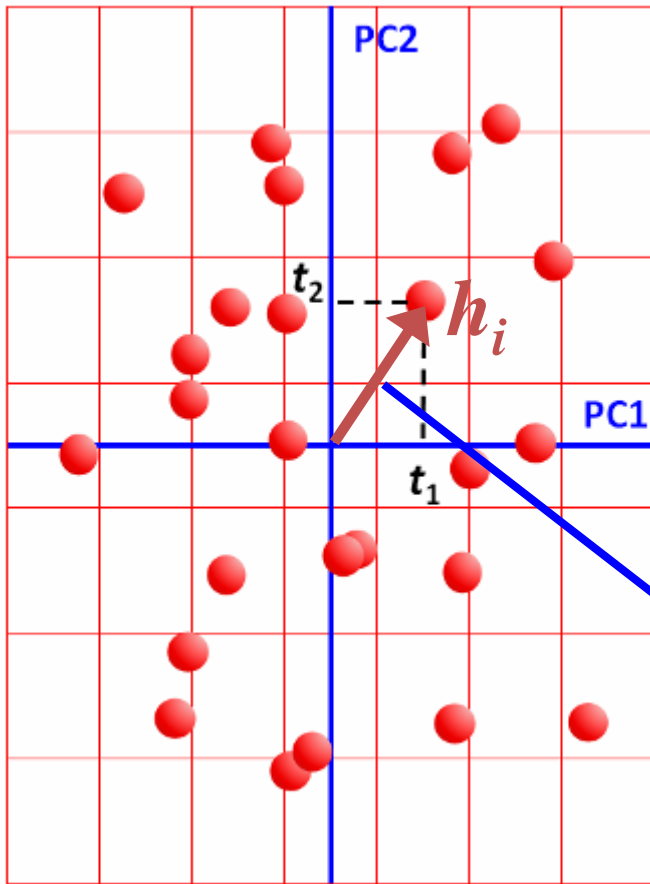
SIMCA (Soft Independent Modeling of Class Analogy)



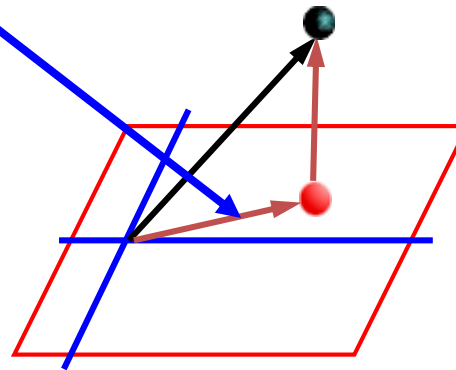
Disjoint Principal components Analysis (PCA) class modeling

New object is compared with each class

Score distance (SD)



- Normalised Q residual variances
- Normalised Hotelling T2 values



Evaluating the performance of SIMCA

non-error rate (*NER*)

% of correct classifications

$$NER\% = \frac{\sum_{g=1}^G c_{gg}}{n} \cdot 100$$

$$NER_g\% = \frac{c_{gg}}{n_g} \cdot 100$$

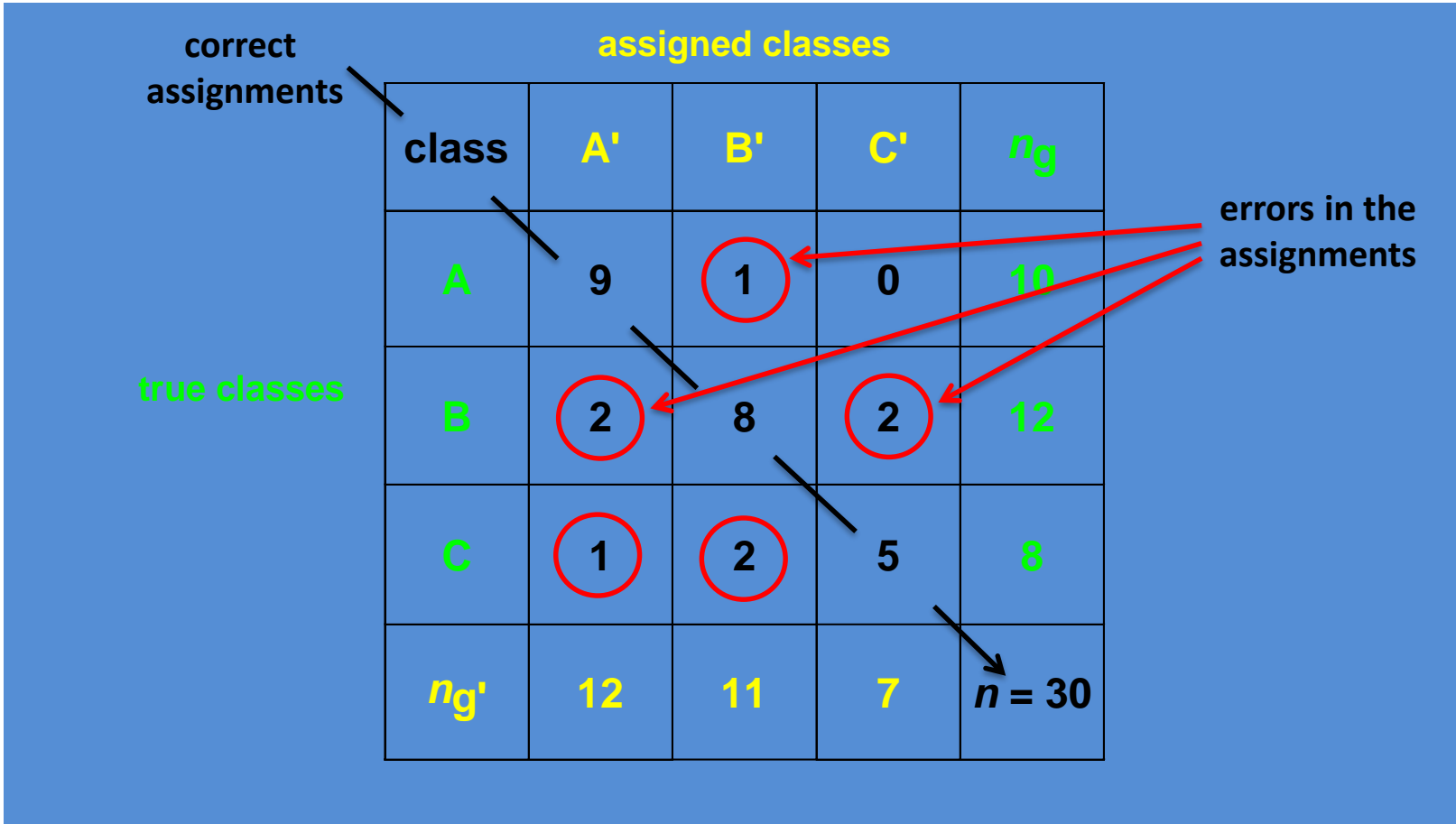
error rate (*ER*)

% of misclassifications

$$ER\% = 100 - NER\%$$

$$ER_g\% = 100 - NER_g\%$$

The confusion matrix

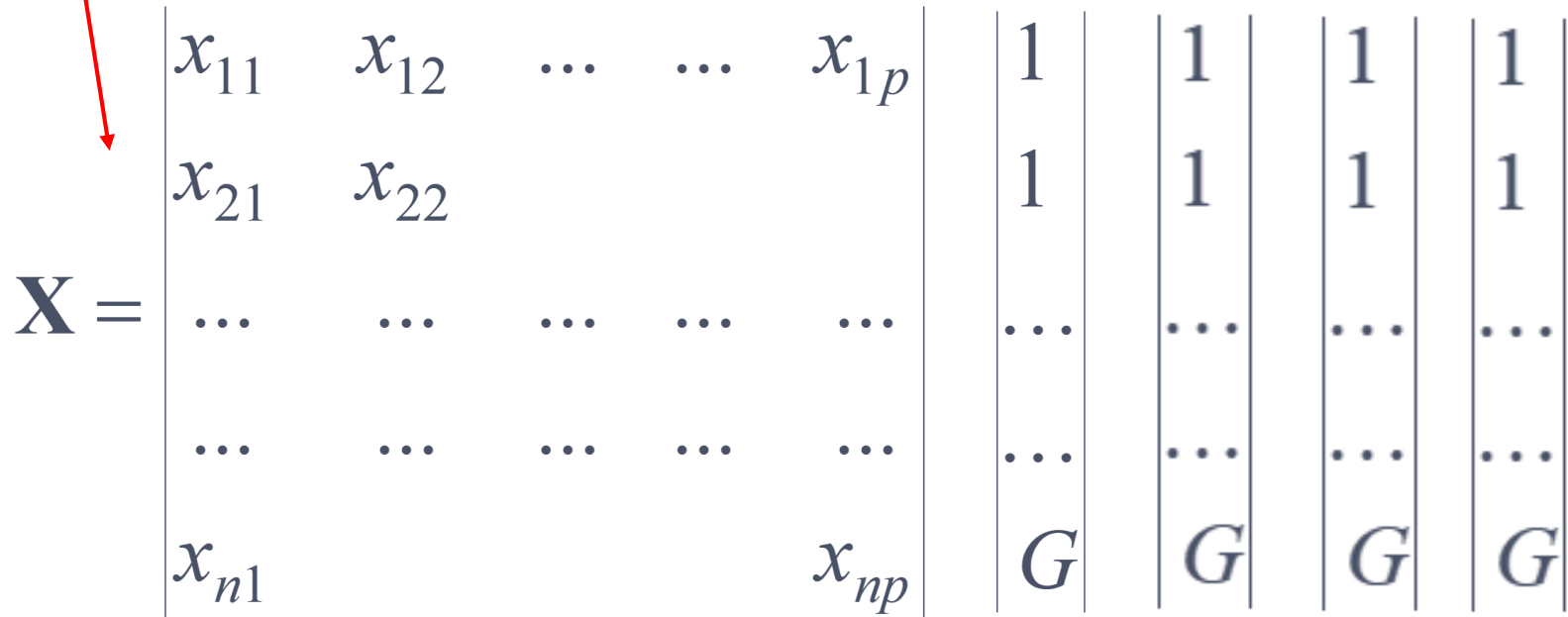


$NER\% (A) = (9/10) \times 100 = 90.0 \%$
 $NER\% (B) = (8/12) \times 100 = 66.7 \%$
 $NER\% (C) = (5/8) \times 100 = 62.5 \%$
 $NER\% = (22/30) \times 100 = 72.3 \%$

**n = sampling points
and their replicas**

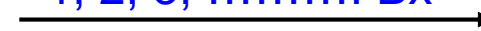
**p = WQ descriptive
variables**

Bx



**Dimension: X (n * p).
n = 301; p = 27.**

1, 2, 3, Bx



n = sampling points and their replicas

p = WQ descriptive variables

X =

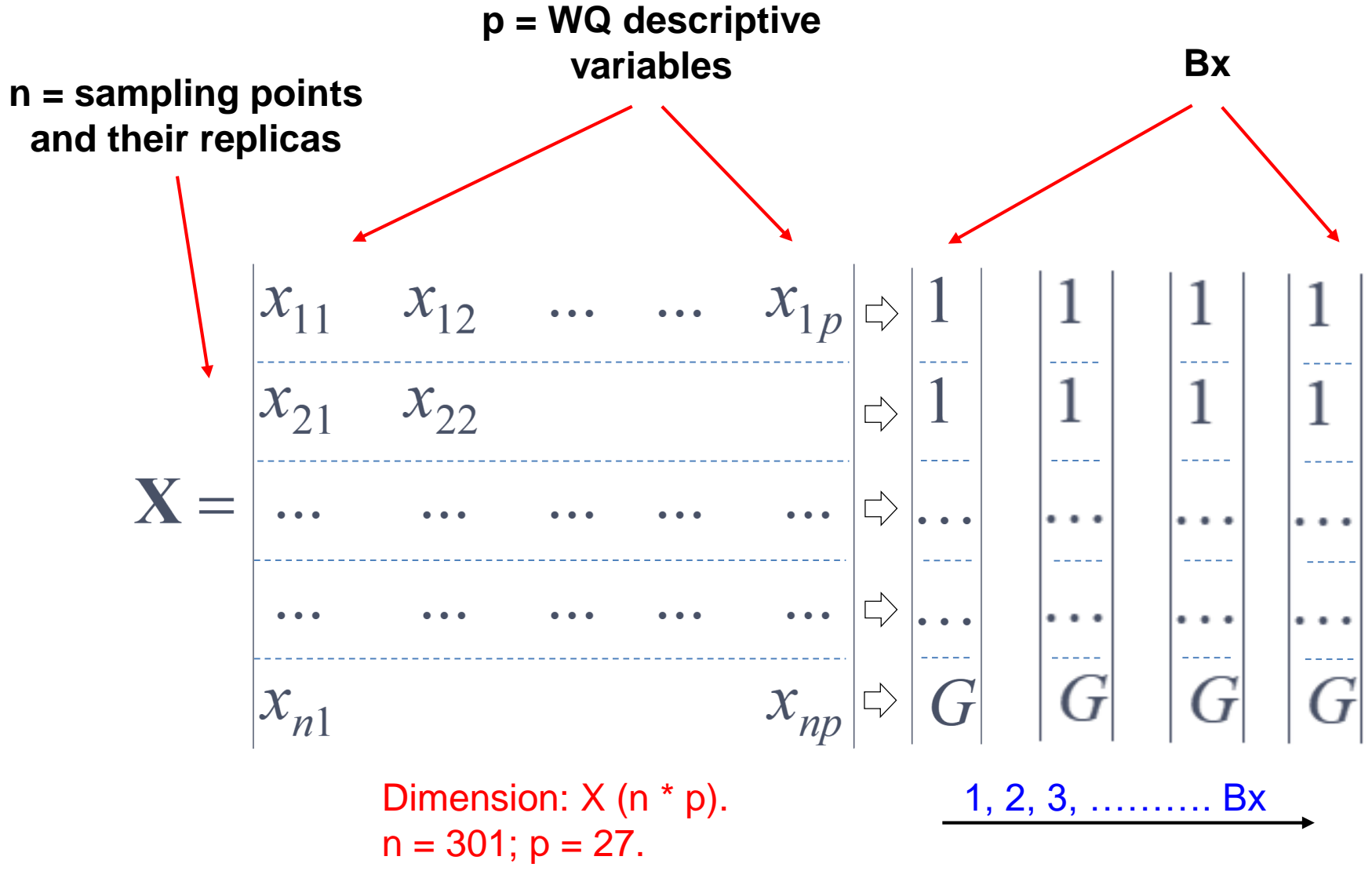
x_{11}	x_{12}	x_{1p}
x_{21}	x_{22}			
...
...
x_{n1}				x_{np}

Dimension: X (n * p).
 n = 301; p = 27.

Bx

1	1	1	1
1	1	1	1
...
...
G	G	G	G

1, 2, 3, Bx

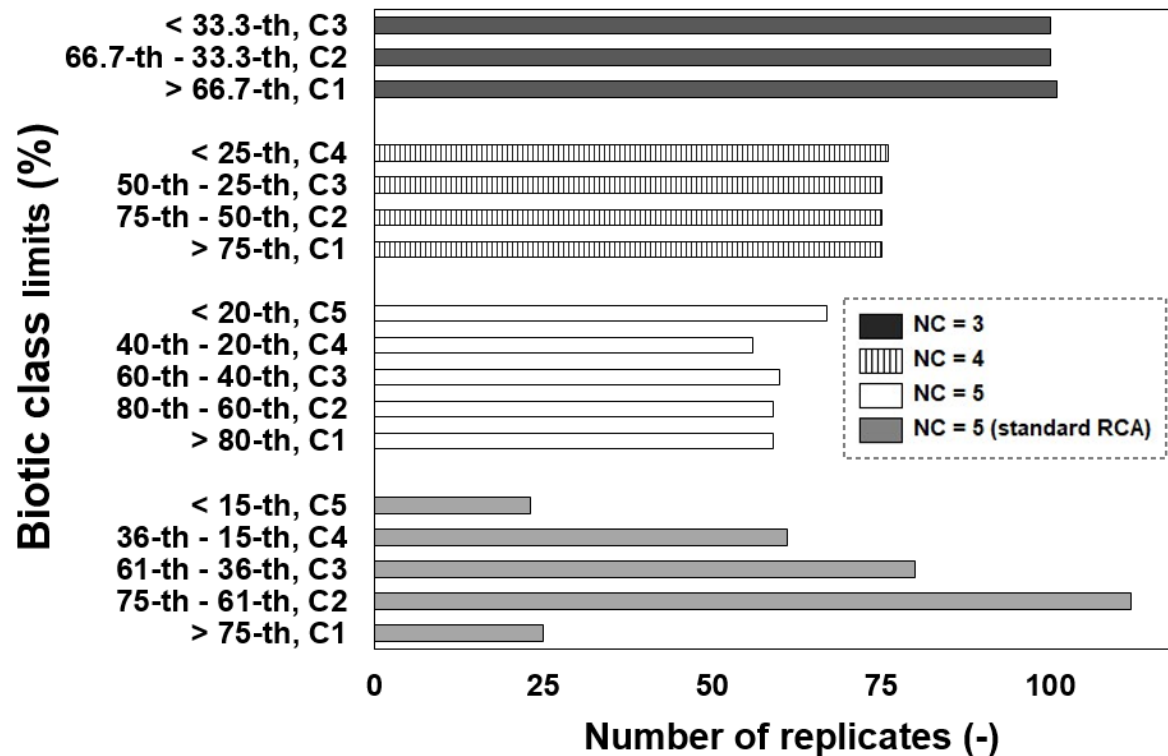


Results

- 53452 macroinvertebrate individuals were collected.
- Ephemeroptera (63.7%), Diptera (11.2%), Trichoptera (8.2%), Coleoptera (6.3%) and Oligochaeta (5.2%).

Results

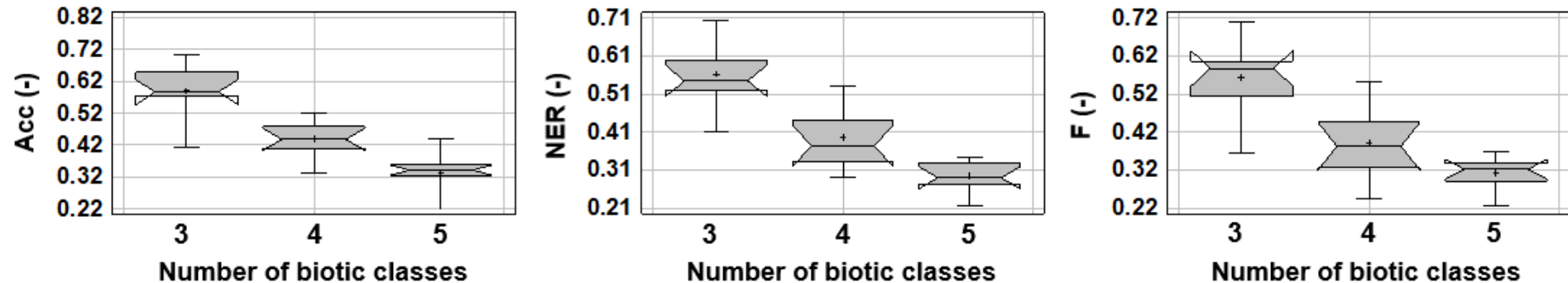
- Number and limits of biotic indices classes



Distribution of the replicates as a function of the number of biotic classes, i.e., NC = 3, NC = 4, NC = 5 and NC = 5 (standard methodology) for the BMWP_Col index.

Results

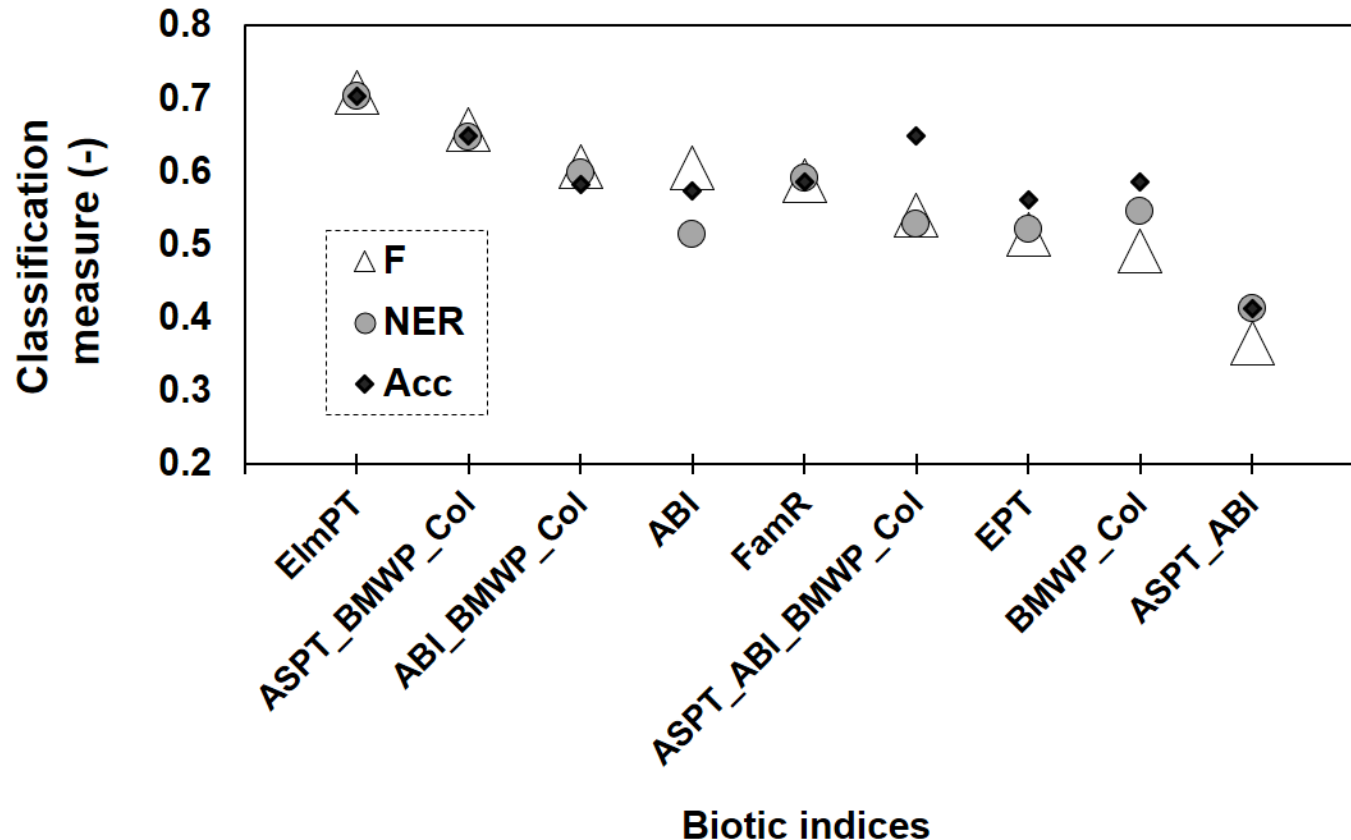
- Number and limits of biotic indices classes



Notched box plots of Acc, NER and F classification measures as a function of the number of biotic classes (NC), considered for every one of the nine inspected biotic indices

Results

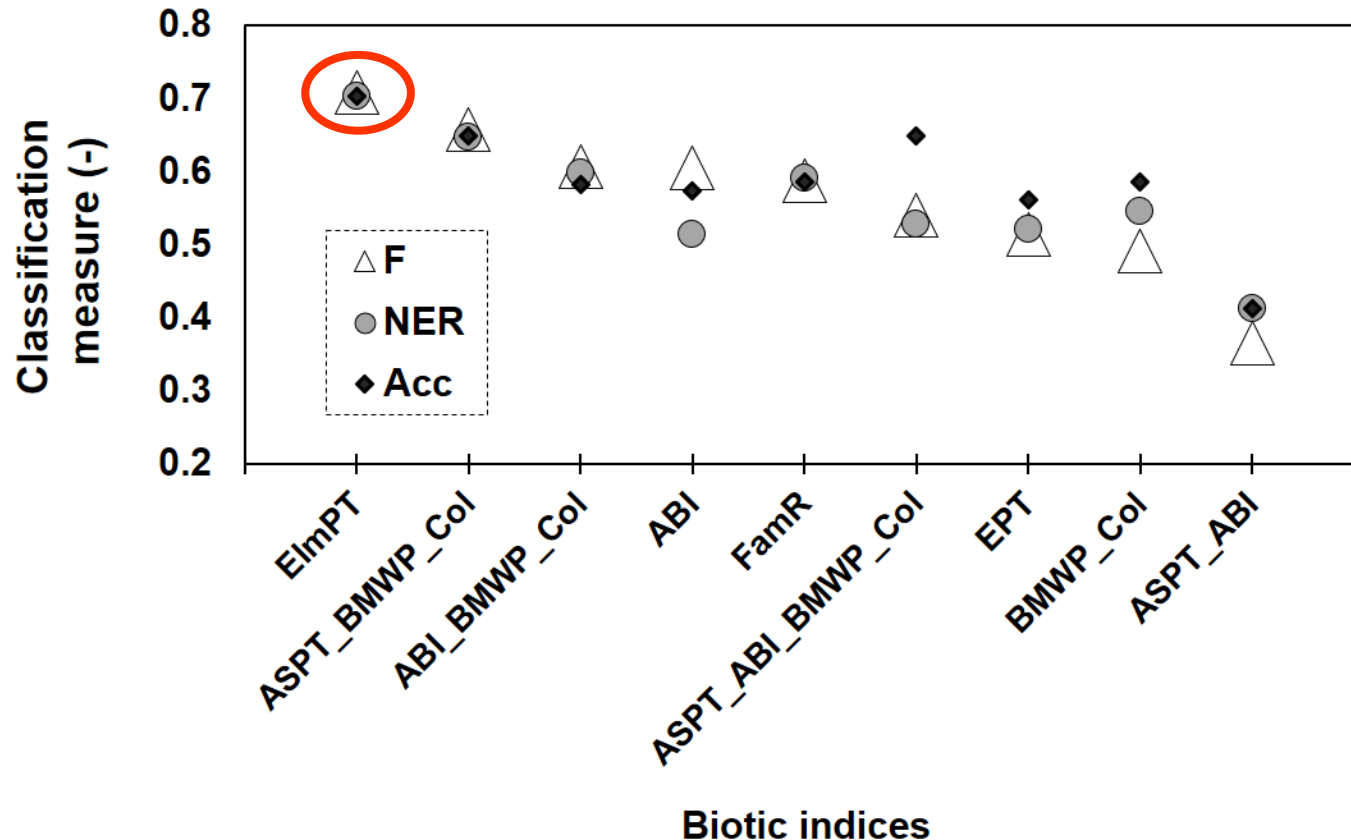
- Choosing the best biotic index through the SIMCA method



Variation of the classification measures (F, NER and Acc) as a function of the classification model (represented herein by the respective biotic index) after application of the SIMCA approach.

Results

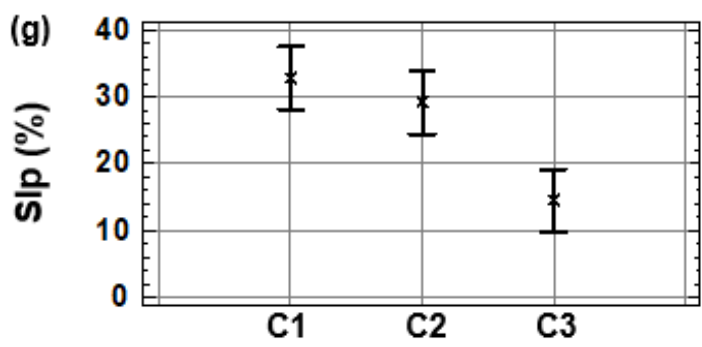
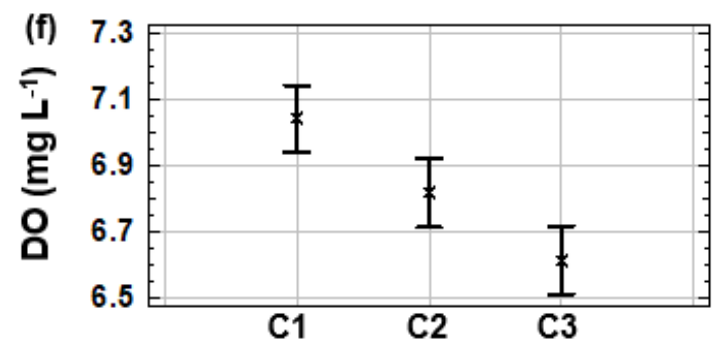
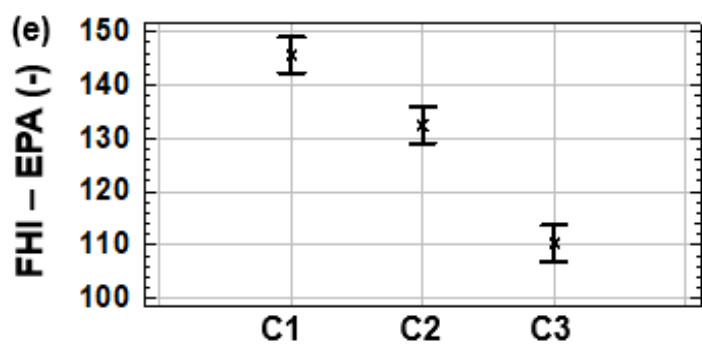
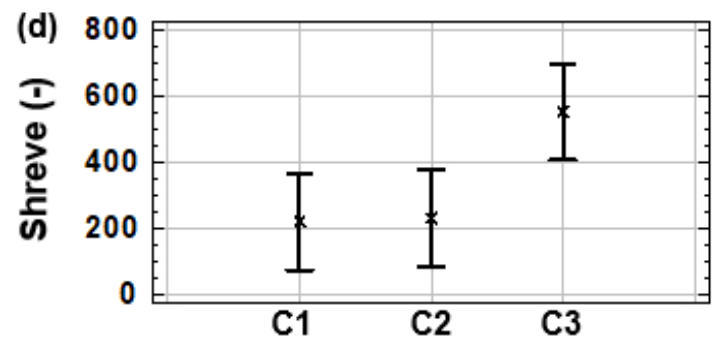
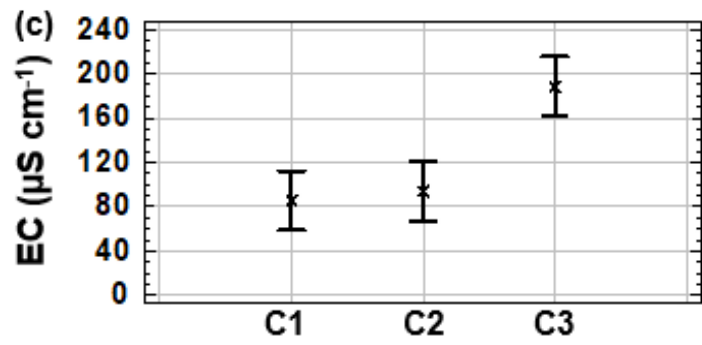
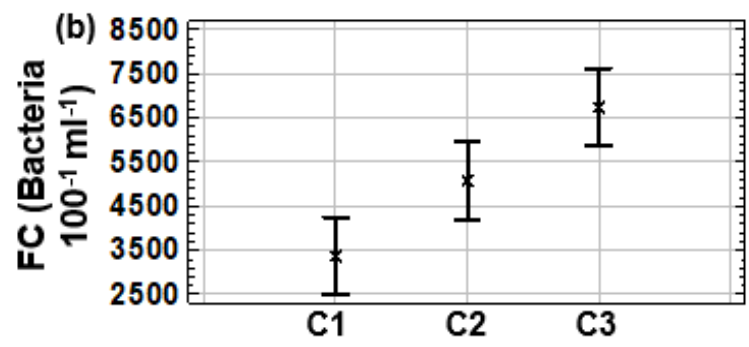
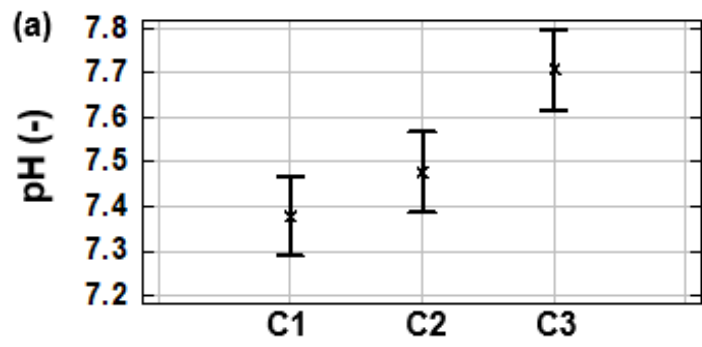
- Choosing the best biotic index through the SIMCA method



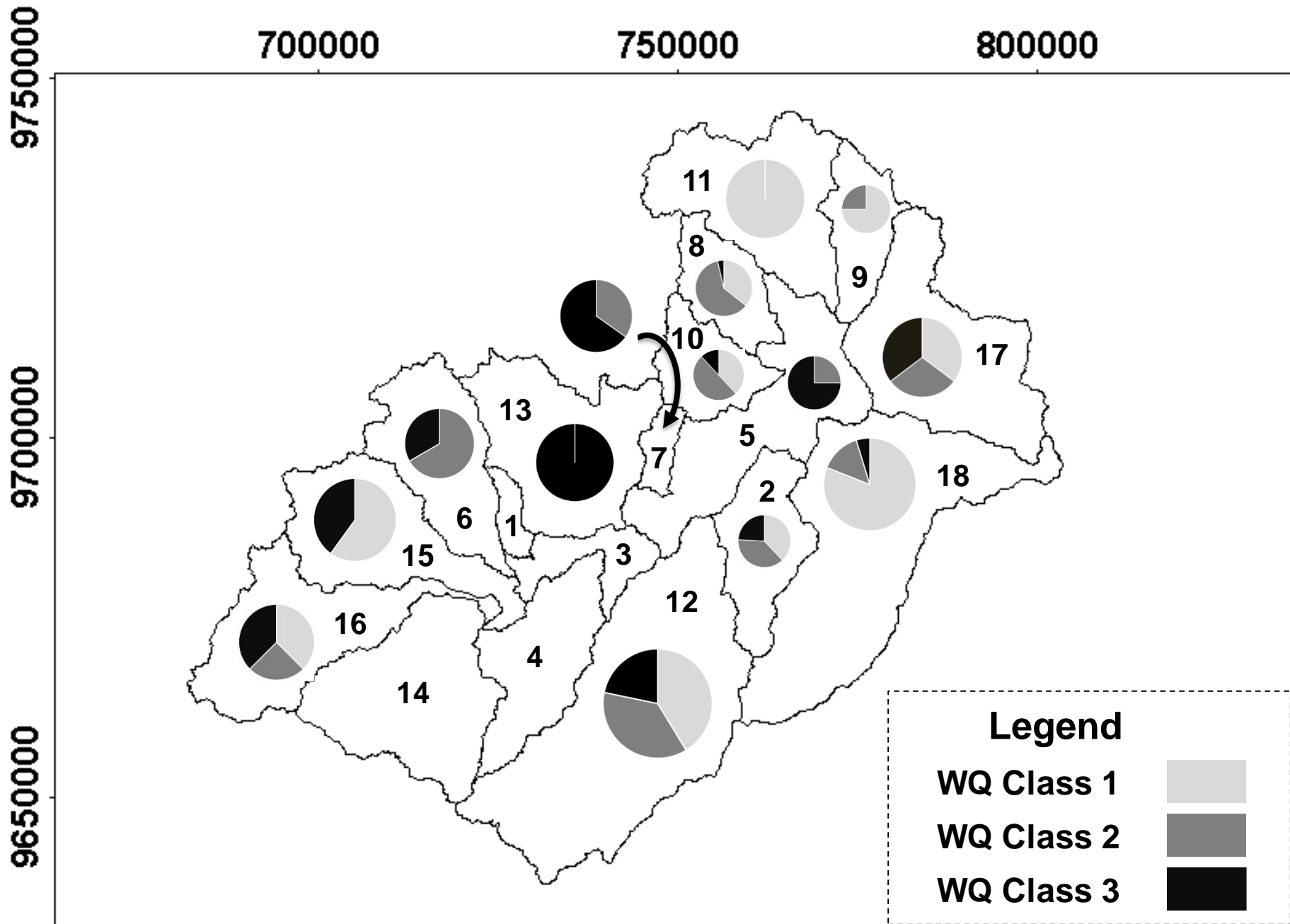
Variation of the classification measures (F, NER and Acc) as a function of the classification model (represented herein by the respective biotic index) after application of the SIMCA approach.

Results

- Assessing the most significant WQ descriptive variables.



Significant WQ descriptive variables as a function of the three biotic classes C1 (less polluted), C2 (moderate polluted) and C3 (highly polluted). The mean values are depicted through an "x" symbol. The intervals around the means are based on Fisher's least significant difference (LSD) procedure.



9750000

700000

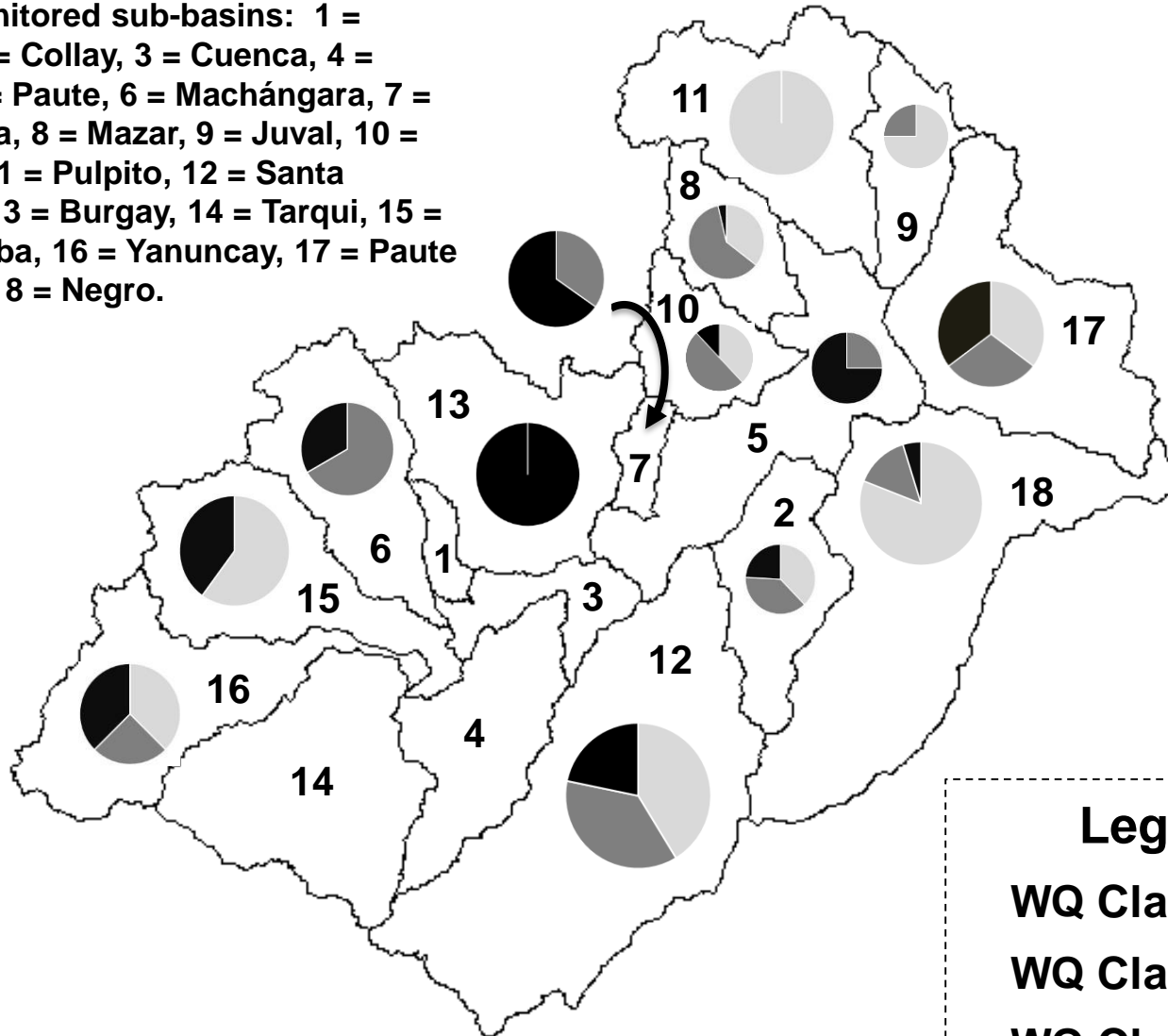
750000

800000

Proportion of WQ classes as a function of the monitored sub-basins: 1 = Sidcay, 2 = Collay, 3 = Cuenca, 4 = Jadán, 5 = Paute, 6 = Machángara, 7 = Magdalena, 8 = Mazar, 9 = Juval, 10 = Pindilig, 11 = Pulpito, 12 = Santa Bárbara, 13 = Burgay, 14 = Tarqui, 15 = Tomebamba, 16 = Yanuncay, 17 = Paute bajo and 18 = Negro.

9700000

9650000



Legend

WQ Class 1



WQ Class 2



WQ Class 3



A lush tropical forest scene featuring a rocky stream. The water flows over a bed of smooth, light-colored stones, creating white rapids. The surrounding environment is dense with green foliage, including large ferns and hanging vines. The walls of the stream are covered in moss and dark rocks. The overall atmosphere is serene and natural.

Thank you...