water

MDPI

Article

Unsupervised Machine Learning and Data Mining Procedures Reveal Short Term, Climate Driven Patterns Linking Physico-Chemical Features and Zooplankton Diversity in Small Ponds

Nicolò Bellin *, Erica Racchetti, Catia Maurone, Marco Bartoli and Valeria Rossi

Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Viale delle Scienze 11/A, I-43124 Parma, Italy; erica.racchetti@unipr.it (E.R.); geomcatia@hotmail.it (C.M.); marco.bartoli@unipr.it (M.B.); valeria.rossi@unipr.it (V.R.)

* Correspondence: nicolo.bellin@unipr.it; Tel.: +39-0521-905612

Table S1. Reported the latitude and longitude in WGS84 coordinate reference system of the 24 ponds under study, the number of years since origin.

Id	Ponds	Age	Latitude	Longitude
1	Pastore3	218	10.324003	45.000660
2	Pastore1	69	10.320438	45.000116
3	Pastore4	218	10.328615	45.000381
4	Temporanea	5	10.322755	45.003484
5	Bosco Braca	69	10.388445	44.994899
6	Pavarini	47	10.391577	44.997471
7	Bosco Valloni	218	10.355710	44.985057
8	San Giorgio	219	10.365307	44.987526
9	Forche	218	10.358832	45.000701
10	Martignana	219	10.363053	44.991424
11	Santa Maria Maddalena	300	10.329035	45.014149
12	Bosco Bodini	46	10.313195	44.995203
13	Cascina Mortara	35	10.313565	44.999700
14	Bosco Piazza	218	10.305016	45.002634
15	Cascina Tavernelle	300	10.309694	45.014147
16	Vecchio	300	10.288462	45.014119
17	Bazzi	64	10.284595	45.011462
18	Motta	219	10.256552	45.052146
19	Ronchetto	218	10.240978	45.033728
20	Rita	300	10.242026	45.052845
21	Bicocca	300	10.228081	45.048045
22	Pescaroli West	218	10.194924	45.046146
23	Pescaroli East	218	10.196969	45.044955
24	Sabbie	300	10.327130	45.006925

Table S2. Showed the physical and chemical environmental features measured for each pond, the unity of measure, the symbols adopted in the study and the laboratory assay or the method of estimation.

Enviromental feature	Type	Unity of measure	Symbol	Laboratory Assay – Method of estimation
Water temperature	Physical	Celsius (°C)	wT	Multiparameters probe (YSI model 566 MPS)
pH	Chemical	dimensionless	pH	Multiparameters probe (YSI model 566 MPS)
Oxygen	Chemical	milligrams per liter (mg l^{-1})	O_2	Multiparameters probe (YSI model 566 MPS)
Conductivity	Chemical	microsiemens per meter ($\mu S.m^{-1}$)	ES	Multiparameters probe (YSI model 566 MPS)
Dissolved inorganic carbon	Chemical	millimolars (mM)	DIC	Anderson et al., 1986
Ammonia	Chemical	milligrams per liter (mg.l^{-1})	NH_4^+	A.P.H.A, 1981
Soluble reactive phosphorus	Chemical	milligrams per liter (mg.l^{-1})	SRP	Valderrama et al., 1977
Nitrate	Chemical	milligrams per liter (mg.l^{-1})	NO_3^-	Rodier, 1987
Chlorophyll a	Chemical	micrograms per liter (µg. l^{-1})	Chla	A.P.H.A, 1981
Reactive silica	Chemical	milligrams per liter (mg.l^{-1})	SiO_2	A.P.H.A, 1981
Depth	Physical	meters (m)	Depth	D'Auria & Zavagno, 1999
Perimeter	Physical	meters (m)	Perimeter	D'Auria & Zavagno, 1999

Table S3. Showed the descriptive statical parameters Range, Mean, Median and Standard deviation (SD) of the nine chemical and the three physical enviromental features, in 2014 and 2015. Water temperature (wT) was expressed in Celsius (°C); Oxygen (O_2), ammonia (NH_4^+), soluble active phosphorus (SRP), nitrate (NO_3^-) and soluble reactive silica (SiO_2) were expressed in $mg.l^{-1}$; dissolved inorganic carbon (DIC) was expressed in mM and clorophyll a (Chla) in $\mu g.l^{-1}$. Depth and perimeter were expressed in meters (m).

Enviromental feature	Range	Mean	Median	SD
wT	10–27.7	21.31	21.40	4.18
pH	6.9–9.6	7.75	7.70	0.45
O_2	0–31.6	7.07	6.20	4.84
ES	180-961	532.6	494	235.65
DIC	1.28–11.73	4.71	4.23	2.78
NH_4^+	0–3.44	0.27	0.09	0.57
SRP	0–1.59	0.09	0.02	0.25
NO_3^-	1–1.09	0.16	0.07	0.23
Chla	0–11.20	2.86	1.95	2.78
SiO_2	0.01–24.41	6.21	3.31	6.39
Depth	0.70–7.50	4.20	4.40	1.64
Perimeter	16–408	210.1	196	75.68

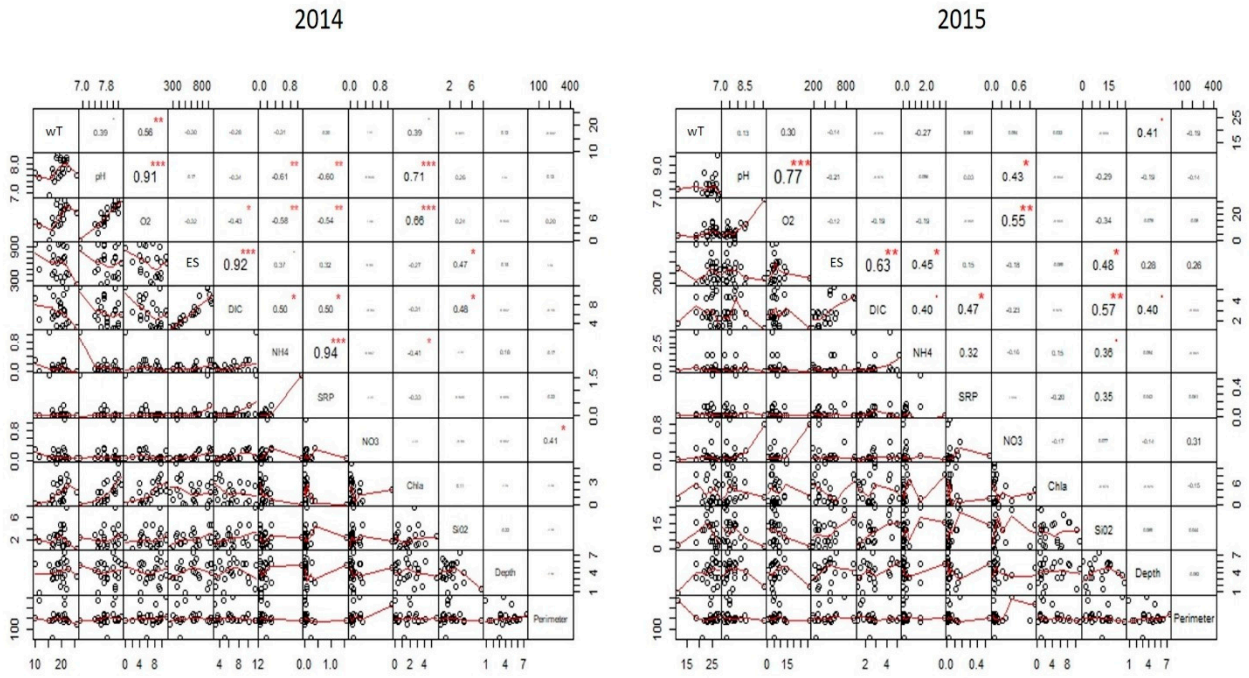

Figure S1. The panels showed the correlation matrix of the enviromental features for the years 2014 and 2015. The diagonal showed the labels of the eleven enviromental features: water temperature (wT), pH, Oxygen (O_2), conductivity (EC), dissolved inorganic carbon (DIC), ammonia (NH_4), soluble reactive phospourus (SRP), nitarte (NO_3^-), clorophyll a (Chla), silica (SiO_2), depth and perimeter. In the upper diagonal part was present the pearson correlation coefficient between pairs of variablels proportional in size to the magnitude of the value, the stars showed the significance level of the correlation test (* $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$). The lower diagonal part showed the scatterplot between pairs of variables. The red line reported the lowess smoother.

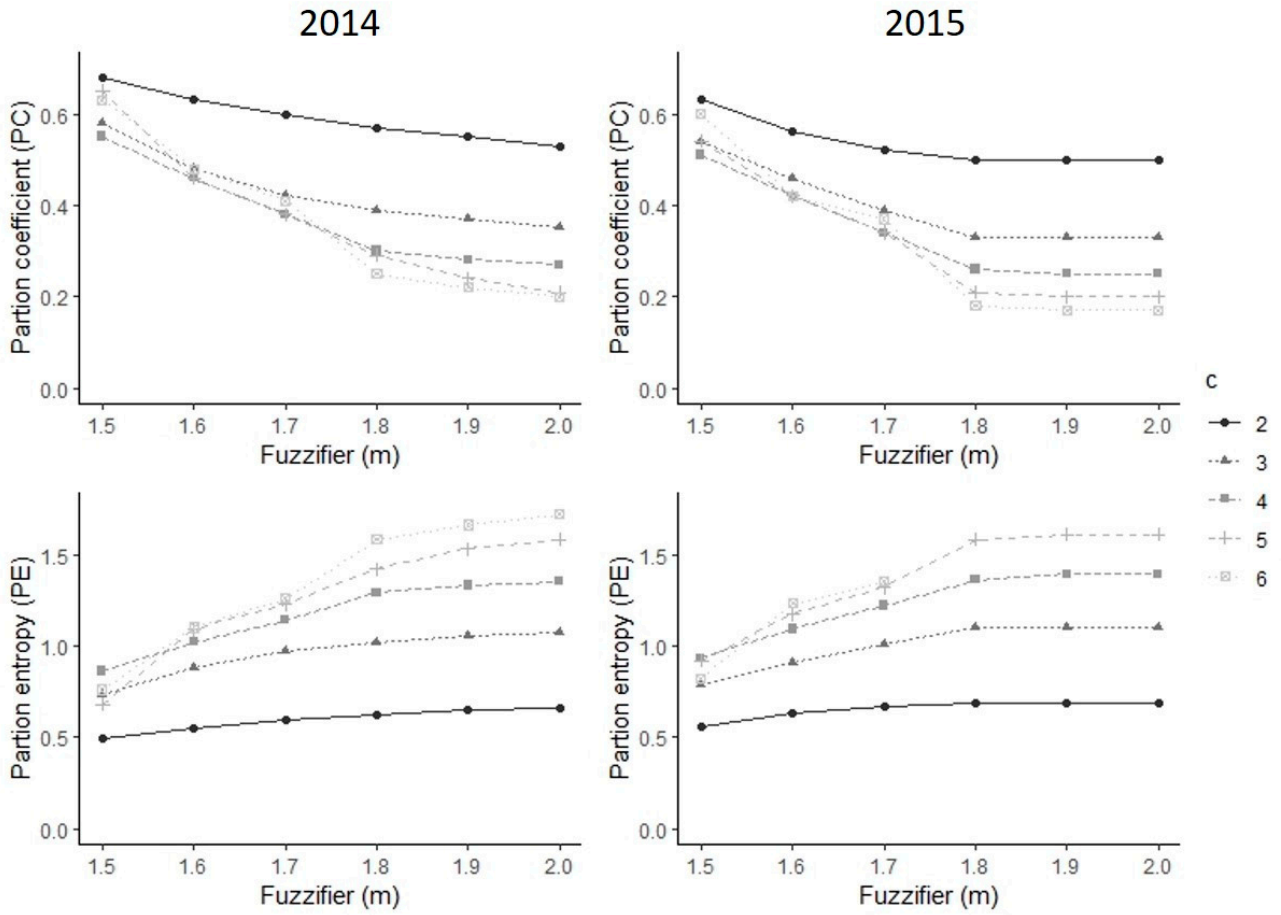

Figure S2. Values of partition coefficient (PC) and partition entropy (PC) for the years 2014 (left panel) and 2015 (right panel). For each plot, the grey scale colors, the point's shape and the geometry of the lines were relative to the different number of clusters (c) in the range 2–6.

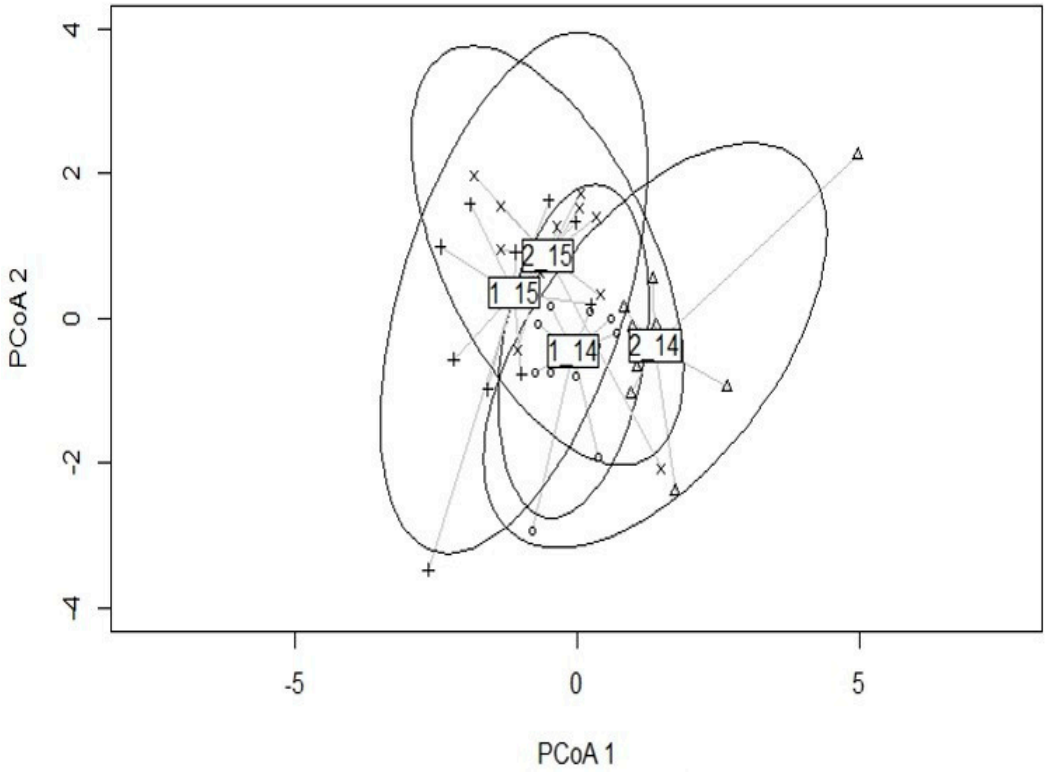

Figure S3. Showed the ordination plot of the average euclidean distance of the scaled environmental features of the water chemistry, from the median of each cluster found by fuzzy c-means. Each ponds were reported on the first two principal coordinate axis and the symbols were relative to the ponds of Cluster 1 (○) and Cluster 2 (Δ) in 2014, Cluster 1 (+) and Cluster 2 (x) in 2015. The ellypses represent 1 standard deviation of the euclidean distances from the median of the clusters. The PERMIDISP analysis after permutation test, revealed not significant difference between groups in habitat heterogeneity.

Table S4. Showed the results of Kolgomorov-Smirnov test between distributions of beta diversity indices after the resampling procedure. The stars represents the p-value (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$).

	2014	2015	
Beta diversity index	**Cluster 1–Cluster 2**	**Cluster 1–Cluster 2**	**2014–2015**
βSOR	D = 0.75****	D = 0.60****	D = 0.39****
βSNE	D = 0.41****	D = 0.16****	D = 0.31****
βSIM	D = 0.81****	D = 0.43****	D = 0.37****

Table S5. Showed the association rules from presence/absence data mined with frequent pattern growth algorithm, in 2014 and 2015. The association rules highlight the frequency and the correlations between taxa co-occurences. For each rule were reported the quantitative measures of interestigness: support, confidrence and lift.

N°	Association rules	Support	Confidence	Lift
1	*Pleuroxus*, Cyclopoida, Calanoida => *Simocephalus*	0.10	1.00	3.92
2	*Pleuroxus*, Cyclopoida => *Simocephalus*	0.10	0.83	3.26
3	*Daphnia*, *Simocephalus* => Calanoida	0.10	1.00	1.52
4	*Chydorus*, *Pleuroxus* => Calanoida	0.10	1.00	1.52
5	*Simocephalus*, *Pleuroxus* => Calanoida	0.17	1.00	1.52
6	*Pleuroxus* => Calanoida	0.25	0.85	1.29
7	*Simocephalus*, Cyclopoida => Calanoida	0.12	0.85	1.29
8	*Simocephalus* => Calanoida	0.21	0.83	1.26
9	*Chydorus*, Cyclopoida => Calanoida	0.10	0.83	1.26

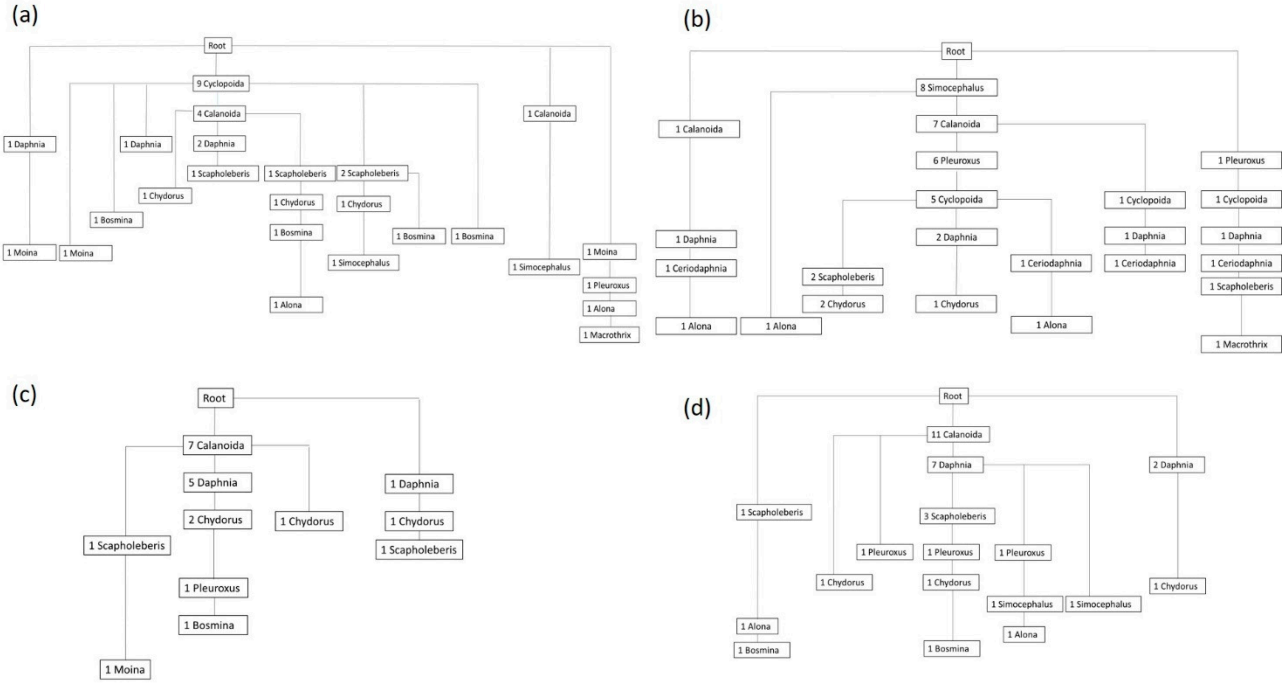

Figure S4. Frequent pattern trees (FPt) for the community structure in Cluster 1 (a) and 2 (b) in 2014 and in Cluster 1 (c) and 2 (d) in 2015. Each node represents a specific taxon and its absolute frequency (number of ponds where the taxon was found). The branches join the co-occurrence of taxa.