NL 2021 Selected Samples prediction using all regions training set,
Learning with all classes present in the selected samples, with extra regional training categories,
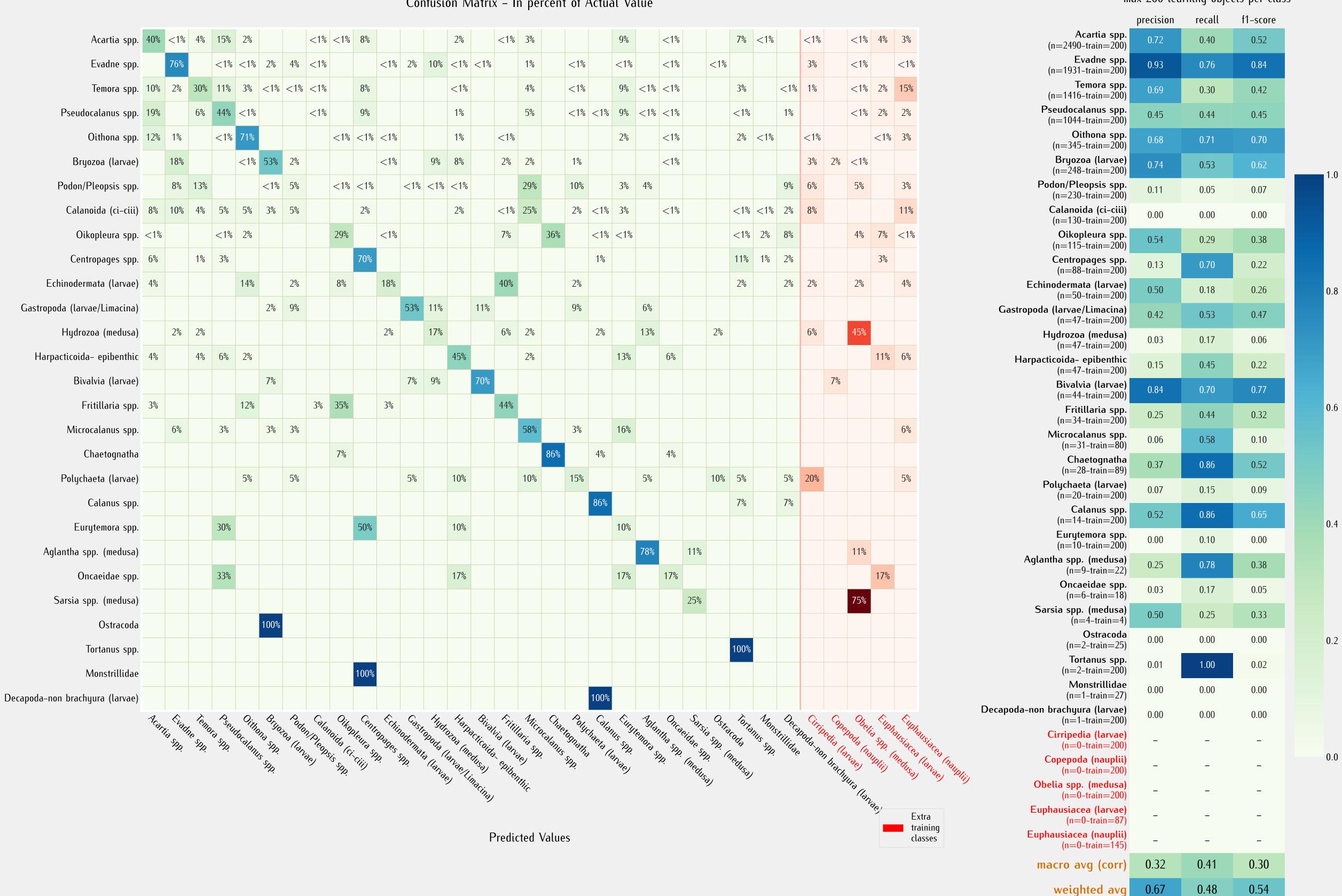No Calanoida (civ–vi), Cyclopoida, Zooplankton classes in learning set

Predictions of discarded taxa from training

Predicted Taxa	Calanoida (civ-vi) n=1370	Cyclopoida (unid) n=2	Foraminifera n=1	Zooplankton (unid) n=318
Acartia spp.	19.8% (n=271)	0.0%	0.0%	0.0%
Aglantha spp. (medusa)	0.1% (n=1)	0.0%	0.0%	2.2% (n=7)
Bivalvia (larvae)	0.0%	0.0%	0.0%	4.1% (n=13)
Bryozoa (larvae)	0.5% (n=7)	0.0%	0.0%	9.7% (n=31)
Calanoida (ci-ciii)	1.2% (n=17)	0.0%	0.0%	0.6% (n=2)
Calanus spp.	0.1% (n=1)	0.0%	0.0%	0.0%
Centropages spp.	2.0% (n=27)	0.0%	0.0%	0.0%
Cirripedia (larvae)	2.5% (n=34)	0.0%	0.0%	2.5% (n=8)
Copepoda (nauplii)	0.0%	0.0%	0.0%	3.1% (n=10)
Decapoda-non brachyura (larvae)	0.4% (n=5)	0.0%	0.0%	0.9% (n=3)
Echinodermata (larvae)	0.0%	0.0%	0.0%	0.3% (n=1)
Euphausiacea (larvae)	3.4% (n=47)	0.0%	0.0%	0.0%
Euphausiacea (nauplii)	8.5% (n=116)	0.0%	0.0%	0.3% (n=1)
Eurytemora spp.	12.8% (n=175)	0.0%	0.0%	0.3% (n=1)
Evadne spp.	3.9% (n=54)	0.0%	0.0%	11.3% (n=36)
Fritillaria spp.	0.4% (n=6)	0.0%	0.0%	1.9% (n=6)
Gastropoda (larvae/Limacina)	0.1% (n=1)	0.0%	0.0%	12.3% (n=39)
Harpacticoida- epibenthic	3.1% (n=43)	0.0%	0.0%	1.9% (n=6)
Hydrozoa (medusa)	0.2% (n=3)	0.0%	0.0%	11.9% (n=38)
Microcalanus spp.	8.5% (n=117)	100.0% (n=2)	0.0%	6.3% (n=20)
Monstrillidae	0.4% (n=5)	0.0%	0.0%	0.0%
Obelia spp. (medusa)	0.7% (n=9)	0.0%	0.0%	3.5% (n=11)
Oikopleura spp.	2.3% (n=32)	0.0%	0.0%	0.0%
Oithona spp.	10.4% (n=143)	0.0%	0.0%	2.8% (n=9)
Oncaeidae spp.	1.4% (n=19)	0.0%	0.0%	1.6% (n=5)
Ostracoda	0.3% (n=4)	0.0%	0.0%	3.5% (n=11)
Podon/Pleopsis spp.	0.4% (n=5)	0.0%	100.0% (n=1)	11.3% (n=36)
Polychaeta (larvae)	0.4% (n=6)	0.0%	0.0%	6.0% (n=19)
Pseudocalanus spp.	9.9% (n=136)	0.0%	0.0%	0.9% (n=3)
Temora spp.	4.9% (n=67)	0.0%	0.0%	0.3% (n=1)
Tortanus spp.	1.4% (n=19)	0.0%	0.0%	0.3% (n=1)

Actual discarded Taxa

Extra training classes

Relative Abundance of Top Taxonomic Instances per Sample