Deep machine learning applications to monitor marine Essential Biodiversity Variables of rocky shore communities

G. Bravo 1, **N. Moity** 2, **E. Londoño-Cruz** 3, **F. E. Muller-Karger** 4, **G. Bigatti** 1, 5 , **E. Montes** 4

1 Instituto de Biología de Organismos Marinos, Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina

2 Charles Darwin Research Station, Charles Darwin Foundation, Santa Cruz, Galapagos , Ecuador

3 Departamento de Biología, Universidad del Valle, Cali, Colombia

4 College of Marine Science, University of South Florida, St Petersburg, Florida, USA

5 Universidad Espíritu Santo, Guayaquil, Ecuador

Abstract

Monitoring marine ecosystems and biodiversity are necessary to understand ecological patterns and processes but also to detect natural or human induced changes such as those resulting from climate change or coastal pollution. A standard technique used in rocky shores is the estimation of cover of sessile organisms and macro-invertebrates. Photoquadrats are becoming standard practice for surveying biodiversity of intertidal and subtidal habitats. They allow to collect large volumes of reliable data efficiently and rapidly in addition to provide a permanent record of the sample. Despite known limitation in taxonomic resolution compared to visual quadrats, photoquadrats have demonstrated to perform well when estimating percent cover of functional groups. Nonetheless, photoquadrat analyses are time consuming and may lead to insufficient accuracy due to low sampling. Cutting-edge machine learning tools are now being used by marine ecologists to annotate species records from photoquadrat imagery. They allow the automatic identification of species, or functional groups, to examine community composition and biodiversity of rocky intertidal and subtidal habitats with high certainty. The use of these tools can significantly reduce the processing time of photo-quadrat imagery and optimize biodiversity survey programs. In this study we present results from visual versus photoquadrat assessments of rocky shores from Argentina, Galapagos Islands (Ecuador) and the Pacific Colombian coast using the CoralNet software. Photo-quadrat imagery was collected during visual surveys carried out at these sites following the South American Research Group on Coastal Ecosystems (SARCE) protocol implemented across the continent by the Marine Biodiversity Observation Network of the Americas (MBON Pole to Pole) program. We apply an ad hoc standardized list of benthic biota and substrata (i.e. CATAMI) as a common label set to enable the comparison between locations. Preliminary results show that CoralNet is able to identify key benthic species and substrate types with high levels of confidence (Pearson correlation coefficient (r) from computer vs visual annotations: Substrate Consolidated, r = 0.91, Molluscs Bivalves, r = 1, Macroalgae filamentous, r = 0.79, Macroalgae sheet-like r = 0.87). We conclude that the CoralNet software can be used to extract presence and percent cover of CATAMI categories. Change detection was tested with an unsupervised configuration of the CoralNet software successfully detecting changes in percent cover of bivalves at 3 sites in Puerto Madryn, Argentina, between Nov/2018 and Nov/2019. This method brings together two programs that are already working to facilitate data analyses over large latitudinal gradients. We suggest the suitability of this method to establish a protocol to rapidly describe rocky shore biodiversity and to detect changes in the biota in the time frame of the MBON Pole to Pole project along the American continent.

Objetives

Test if unsupervised configuration of the CoralNet software identifies CATAMI categories on intertidal photoquadrats from 3 sites of Argentina.

Methods

The robot was trained with 90 photos (30 per site) with annotations performed by an experienced observator. An independent set of photoquadrats (n=90) from the same sites and date was analyised with CoralNet artificial inteligence. Both results (Human annotator and Robot annotator) were compared with visual quadrats results. The quadrtas analyised visually were the same as the photoquadrats annotated by human.

The same number of CATAMI categories were used for visual and photoquadrats:

	CATAMI CODE	Frequency
7	SC	7831
3	MOB	7742
11	MAFG	3416
16	MASG	3371
8	MAA	2547
12	MAFR	1017
2	CRB	385
9	MAEN	359
14	MALCB	171
5	WPOT	94
13	MAG	38
15	MASB	13
1	CNTR	10
10	MAENRC	5
4	SP	1
6	SU	0

Results

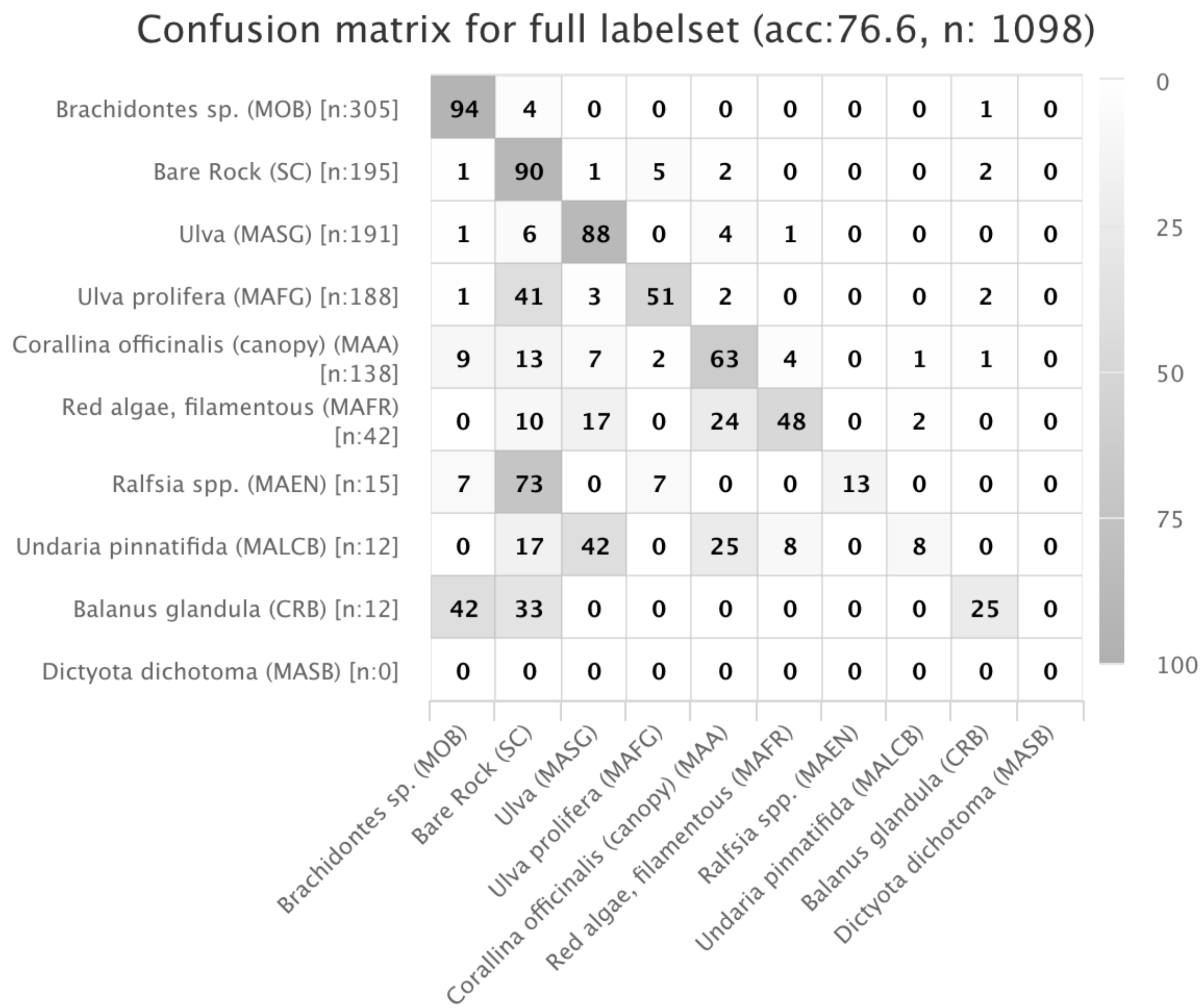

Figure 1: Confusion Matrix, robot trained with 90 photos, Expresed in percentage

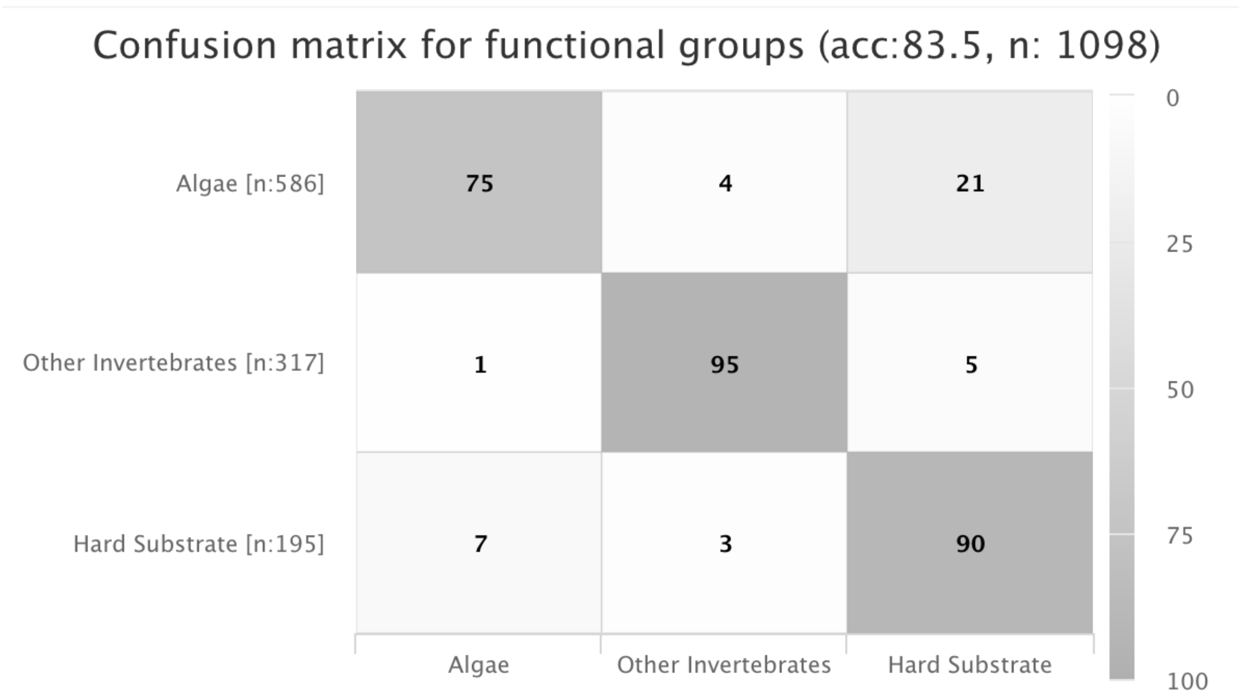

Figure 2: Confusion Matrix by funtional groups, robot trained with 90 photos, Expresed in percentage

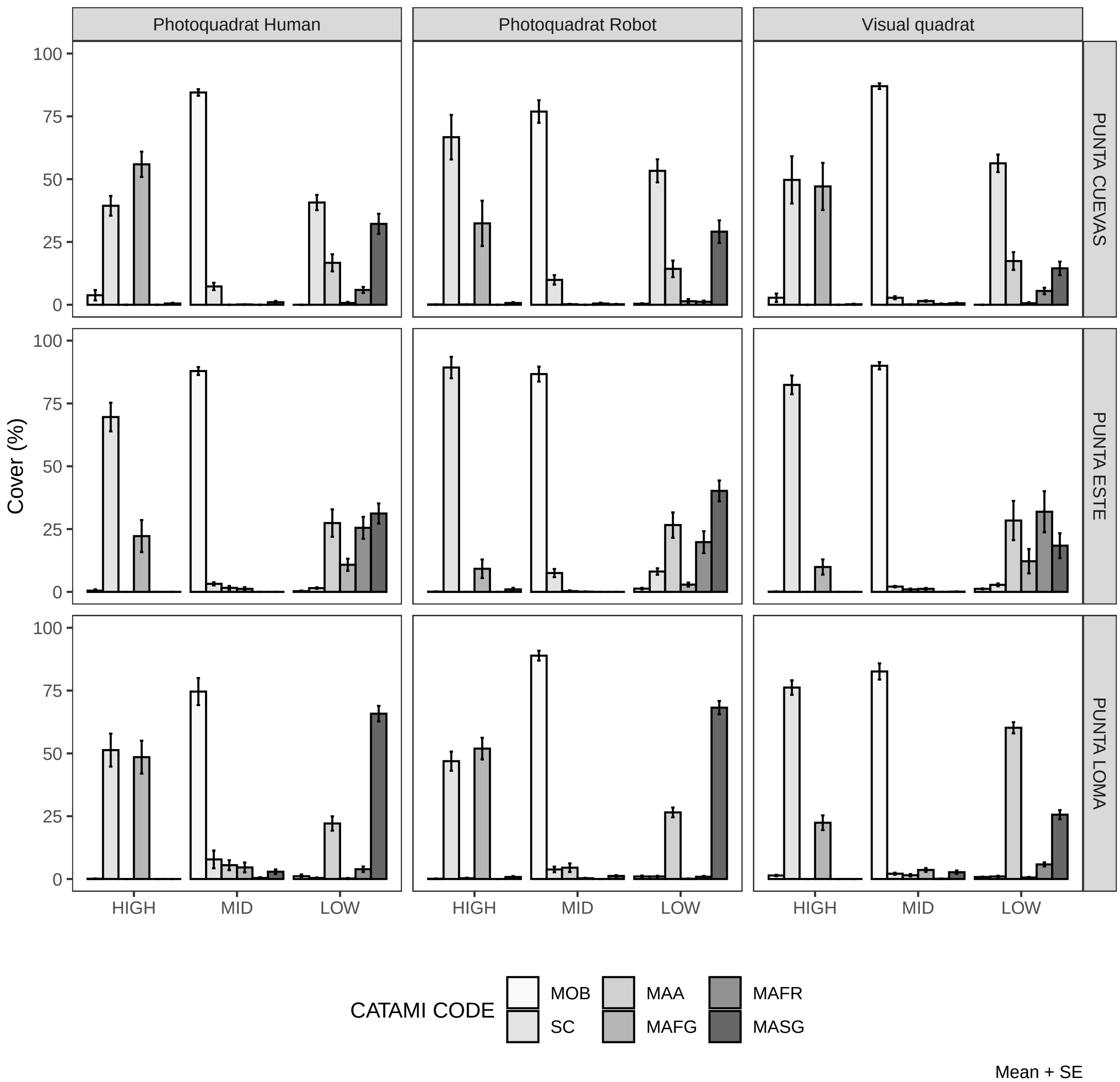

Figure 3: Overall view of differences between methods

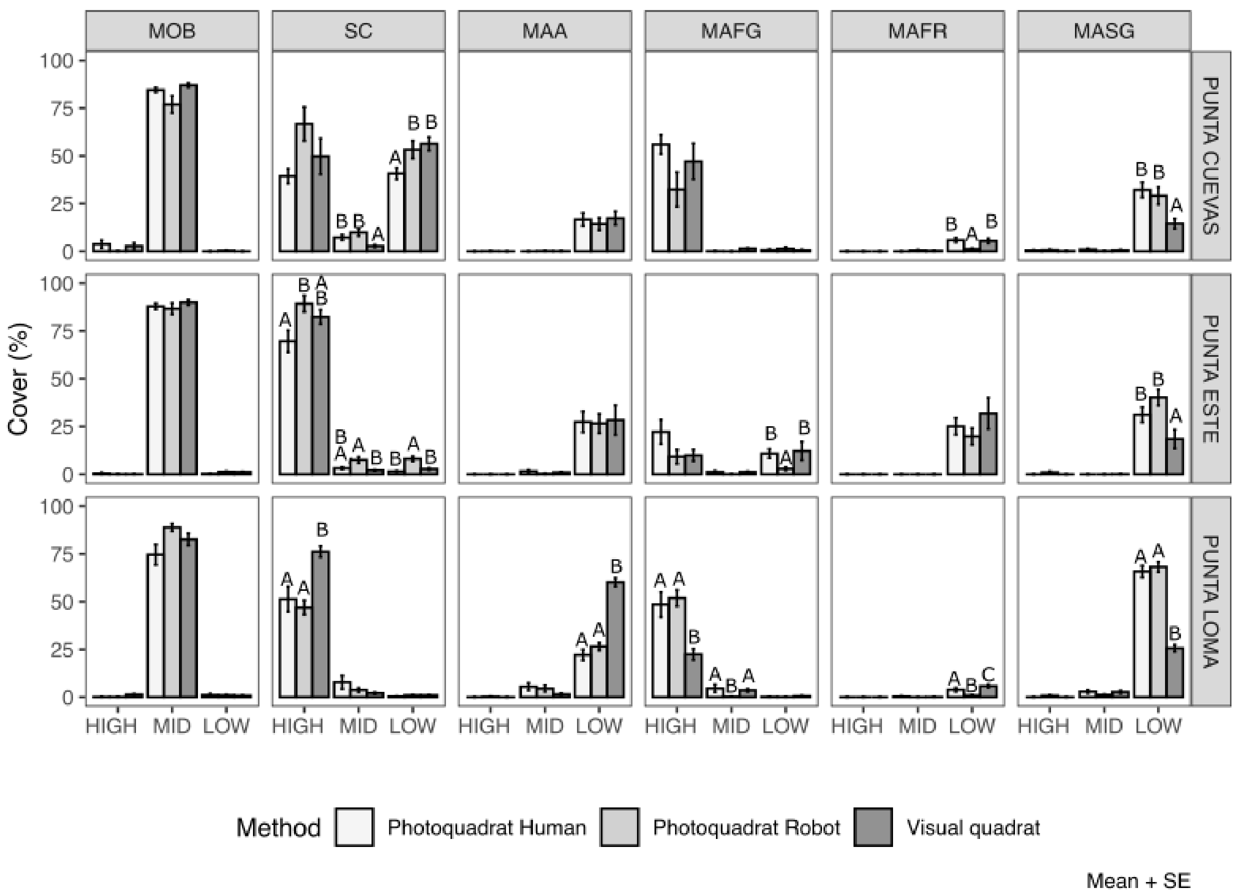

Figure 4: Diferences among method for percentage cover estimation of mayor CATAMI categories. Different letters indicate p<0.05 (Mann-Whitney test)

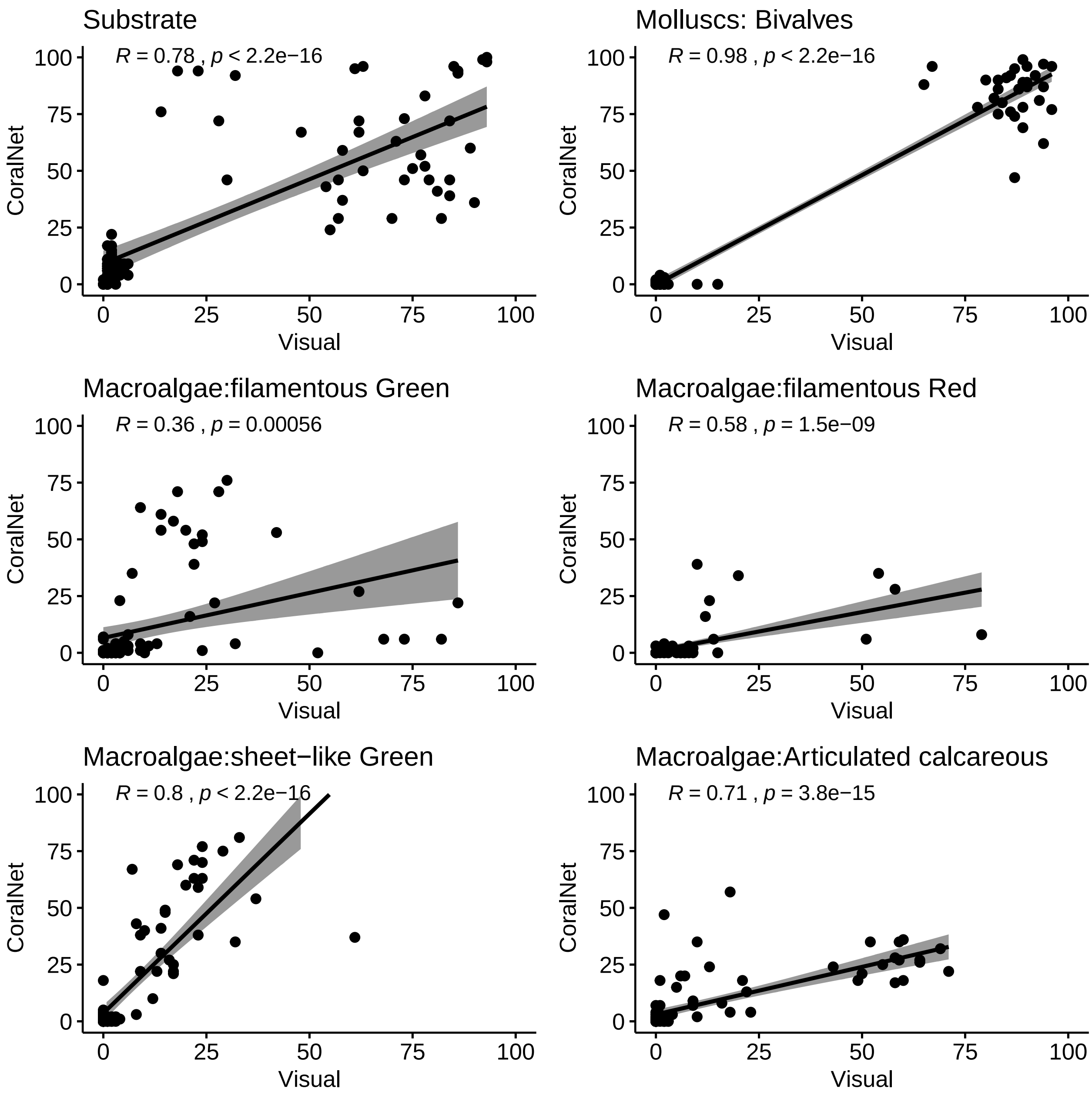

Figure 5: Correlations between CoralNet robot and visual estimations

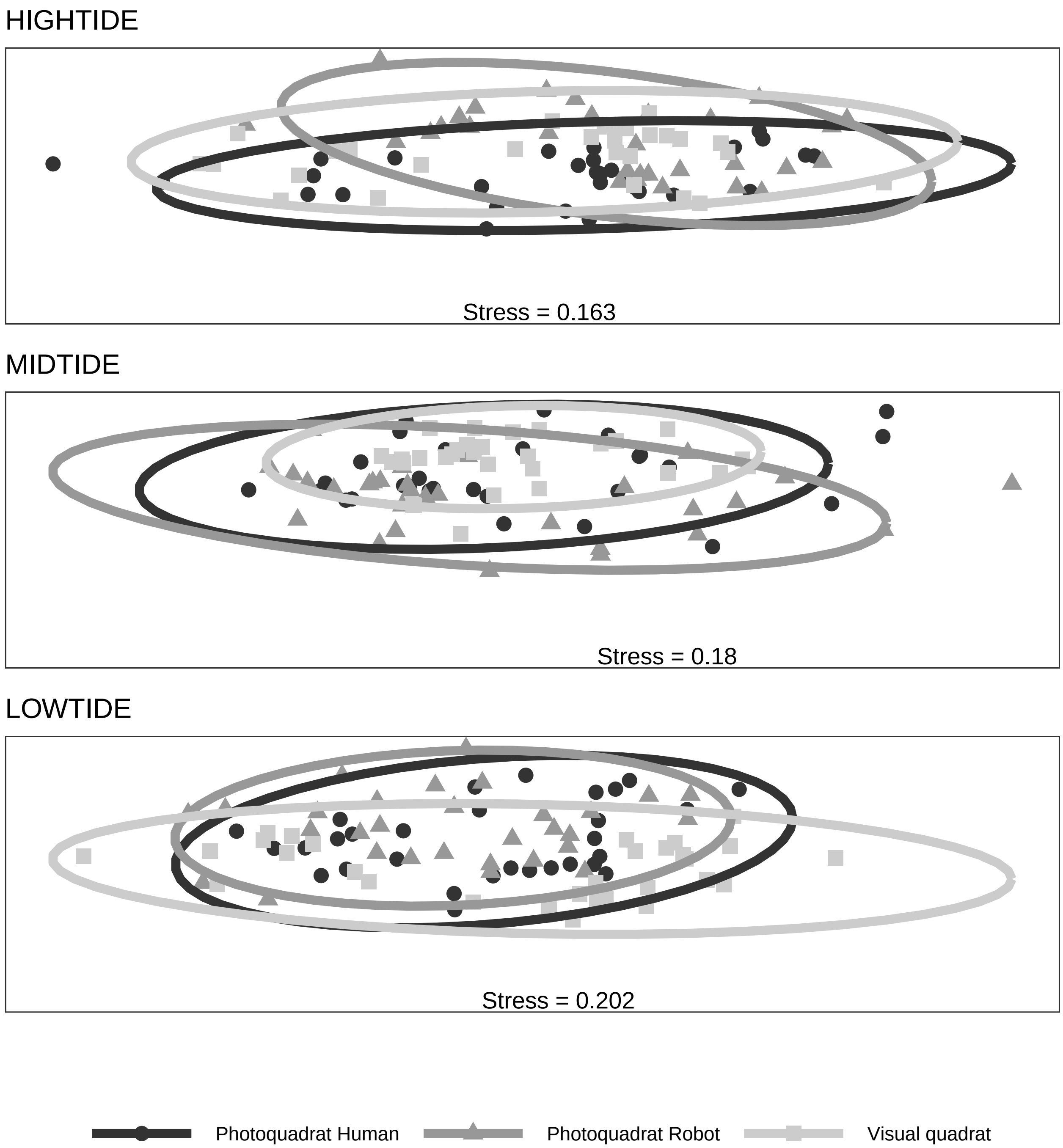
HIGHTIDE
Stress = 0.163
MIDTIDE
Stress = 0.18
LOWTIDE
Stress = 0.202
Photoquadrat Human
Photoquadrat Robot
Visual quadrat

Table 2: PERMANOVA Hightide

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
Photoquadrat Human vs Photoquadrat Robot	1	0.262835	4.0504880	0.0652773	0.036	0.108	
Photoquadrat Human vs Visual quadrat	1	0.394670	6.2484320	0.0972542	0.014	0.042	.
Photoquadrat Robot vs Visual quadrat	1	0.054405	0.8110187	0.0137903	0.367	1.000	

Table 3: PERMANOVA Midtide

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
Photoquadrat Human vs Photoquadrat Robot	1	0.0209133	1.122805	0.0189911	0.313	0.939	
Photoquadrat Human vs Visual quadrat	1	0.0367067	2.638117	0.0435059	0.046	0.138	
Photoquadrat Robot vs Visual quadrat	1	0.0662567	5.479375	0.0863174	0.005	0.015	.

Table 4: PERMANOVA Lowtide

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
Photoquadrat Human vs Photoquadrat Robot	1	0.1375217	1.350154	0.0227490	0.256	0.768	
Photoquadrat Human vs Visual quadrat	1	0.9176817	7.188003	0.1102657	0.003	0.009	*
Photoquadrat Robot vs Visual quadrat	1	1.1250550	9.013208	0.1344990	0.001	0.003	*

Remarks

Training the robot with 90 photos seems to be enought to get good estimations of Bivalves (MOB), Ulva (MASG) and substrate (SC) which have high number of annotations.

Multivariated analysis showed no difference between robot and human photoquadrats. But in some cases showed differences with visual quadrats (table 2, 3, 4).

Next step: train the robot with 180 photos and compare results.

In order to repeat the same analysis in other sites we need to have 90 extra photos from the same date. If not we continue with the idea of making comparations with photos from other year. (2018 vs 2019)

Some of the species recorded on Visual quadrats are from the understory (below algae), imposible to get with photoquadrats. So, we need to clean that data.

Code and data files

All the codes and data files are in github

https://github.com/gonzalobravoargentina/CoralNet_MBON

Files:

MBON_Argentina_metadata.csv (metada from photoquadrats)

MBON_Argentina_percent_covers.csv (% cover photoquadrats (human and robot))

percent_cover_visual_2018.csv (visual quadrat data)

photo_visual_dataframe.csv (merge dataframe photo + visual)

correation_matrix.csv (dataframe for correlation analysis)