

Appendix S1

Table S1.1: Number of megabenthic invertebrate taxa (species richness) collected per Geographical Sub-Areas (GSA) during MEDITS surveys.

Sub-basin	GSA	Species richness
Western Mediterranean	7	198
	5	184
	1	175
	6	158
	8	109
	9	93
	10	88
	2	85
	11	71
Central Mediterranean	16	115
	19	97
	15	8
Adriatic Sea	18	137
	17	97
Aegean Sea	22	78
	20	70

Table S1.2: Evaluation metrics of the RF models performed per sub-basin.

Sub-basin	Cluster	Sensitivity	Specificity	Balanced Accuracy	Global accuracy
Western Mediterranean	1	0.95	0.99	0.97	0.89
	2	0.96	0.98	0.97	
	3	0.64	1.00	0.82	
	4	0.86	0.92	0.89	
	5	0.89	0.94	0.92	
Central Mediterranean	6	0.73	0.95	0.84	0.87
	7	0.71	0.98	0.84	
	8	0.94	0.78	0.86	
Adriatic Sea	9	0.77	0.96	0.86	0.82
	10	0.75	0.90	0.83	
	11	0.83	0.92	0.87	
	12	0.83	0.98	0.90	
	13	0.94	1.00	0.97	
Aegean Sea	14	0.80	1.00	0.90	0.84
	15	0.67	0.96	0.81	
	16	0.96	0.79	0.88	

Table S1.3: Diversity index values for each cluster across the four sub-basins, with R corresponding to taxa richness, and J', Pielou's evenness index.

Sub-basin	Cluster	R	J'
Western Mediterranean	1	182	0.85
	2	195	0.83
	3	148	0.85
	4	243	0.76
	5	213	0.69
Central Mediterranean	6	132	0.81
	7	109	0.80
	8	123	0.69
Adriatic Sea	9	88	0.82
	10	99	0.76
	11	106	0.64
	12	121	0.84
	13	78	0.78
Aegean Sea	14	49	0.79
	15	50	0.92
	16	56	0.75

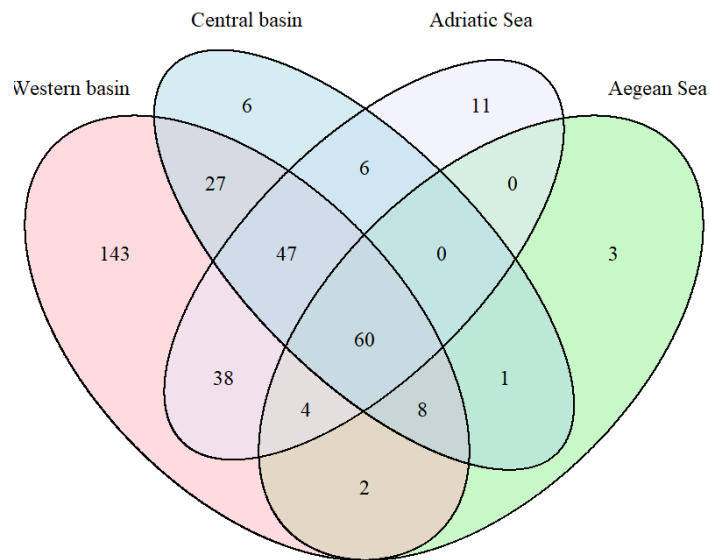


Figure S1.1: Venn diagram illustrating the comparison of megabenthic invertebrate taxa across the four Mediterranean sub-basins. Each ellipse represents a sub-basin, with overlapping areas showing shared taxa between sub-basins. Non-overlapping areas indicate taxa unique to a specific sub-basin.

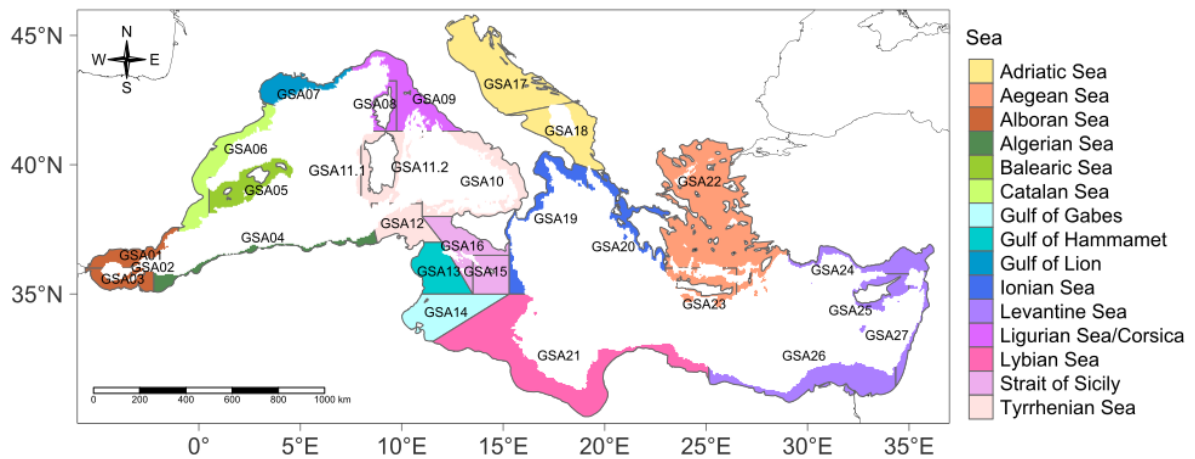


Figure S1.2: Map of Geographical Sub-Areas (GSA) defined by the GFCM in the Mediterranean.

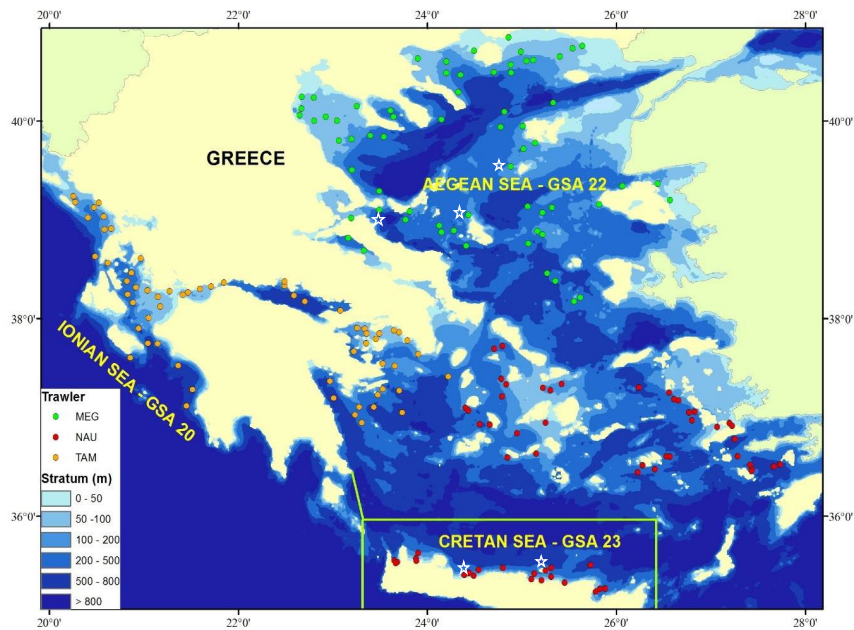


Figure S1.3: Sampling stations carried out during the 2020 MEDITS surveys in GSAs 20, 22, and 23. Three different vessels, operated by three separate teams, covered distinct areas. The colored dots indicate the stations completed by each vessel and team (source: Greece Annual Report 2020-DCF). For this study, only MEDITS records from the regions corresponding to the green and yellow stations were considered.

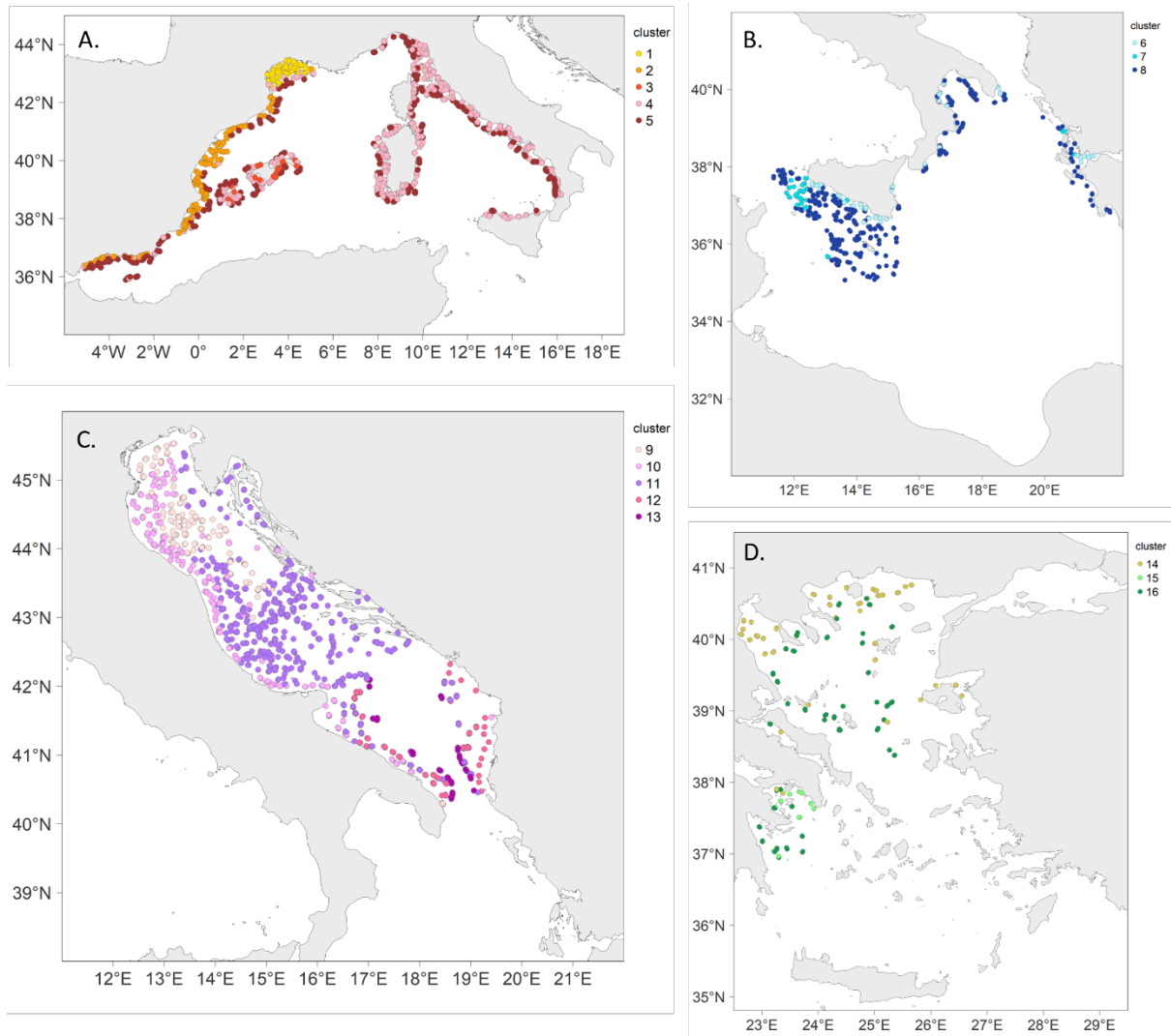


Figure S1.4: Clusters of MEDITS stations determined by the K-means algorithm for (A) the Western Mediterranean, (B) the Central Mediterranean, (C) the Adriatic Sea, and (D) the Aegean Sea.

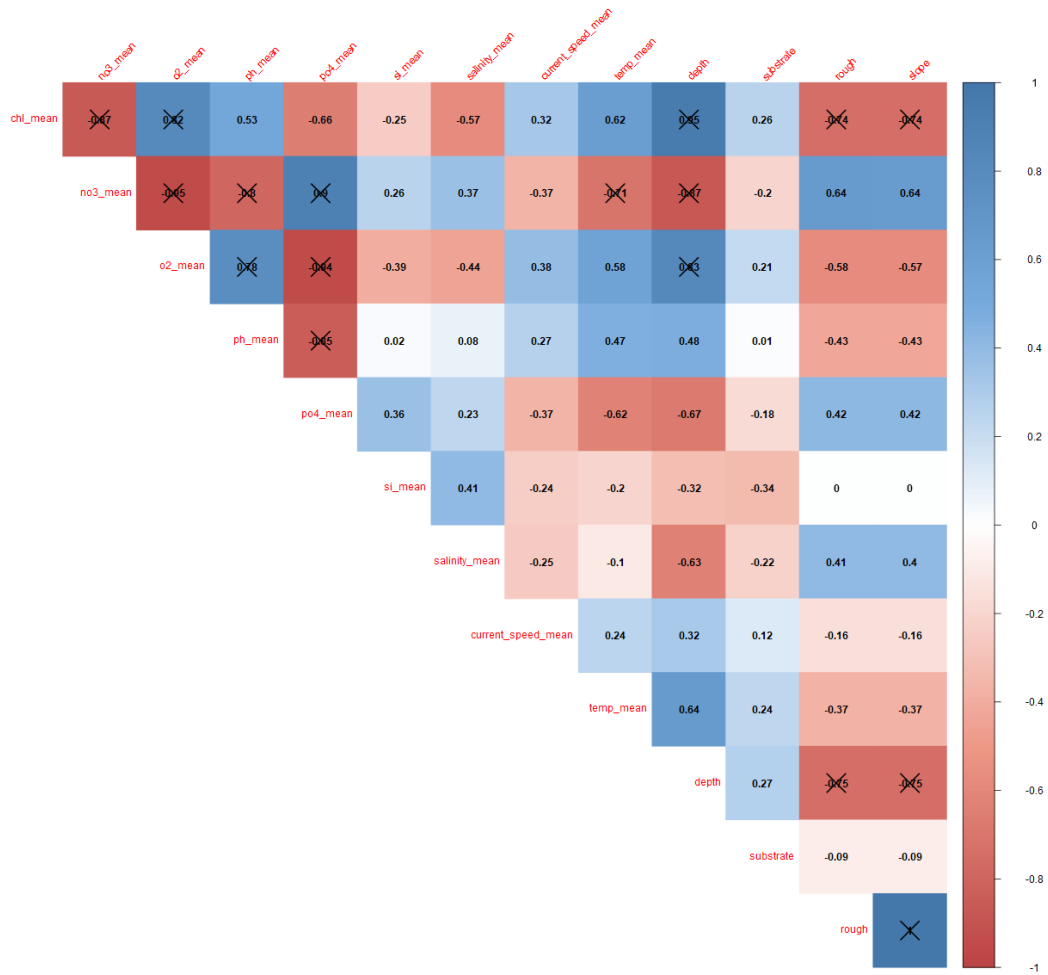


Figure S1.5: Correlation matrix based on Spearman's rank coefficient between environmental predictors. Positive correlations are colored in blue and negative correlations are colored in red. Coefficients higher than 0.7 are crossed out.

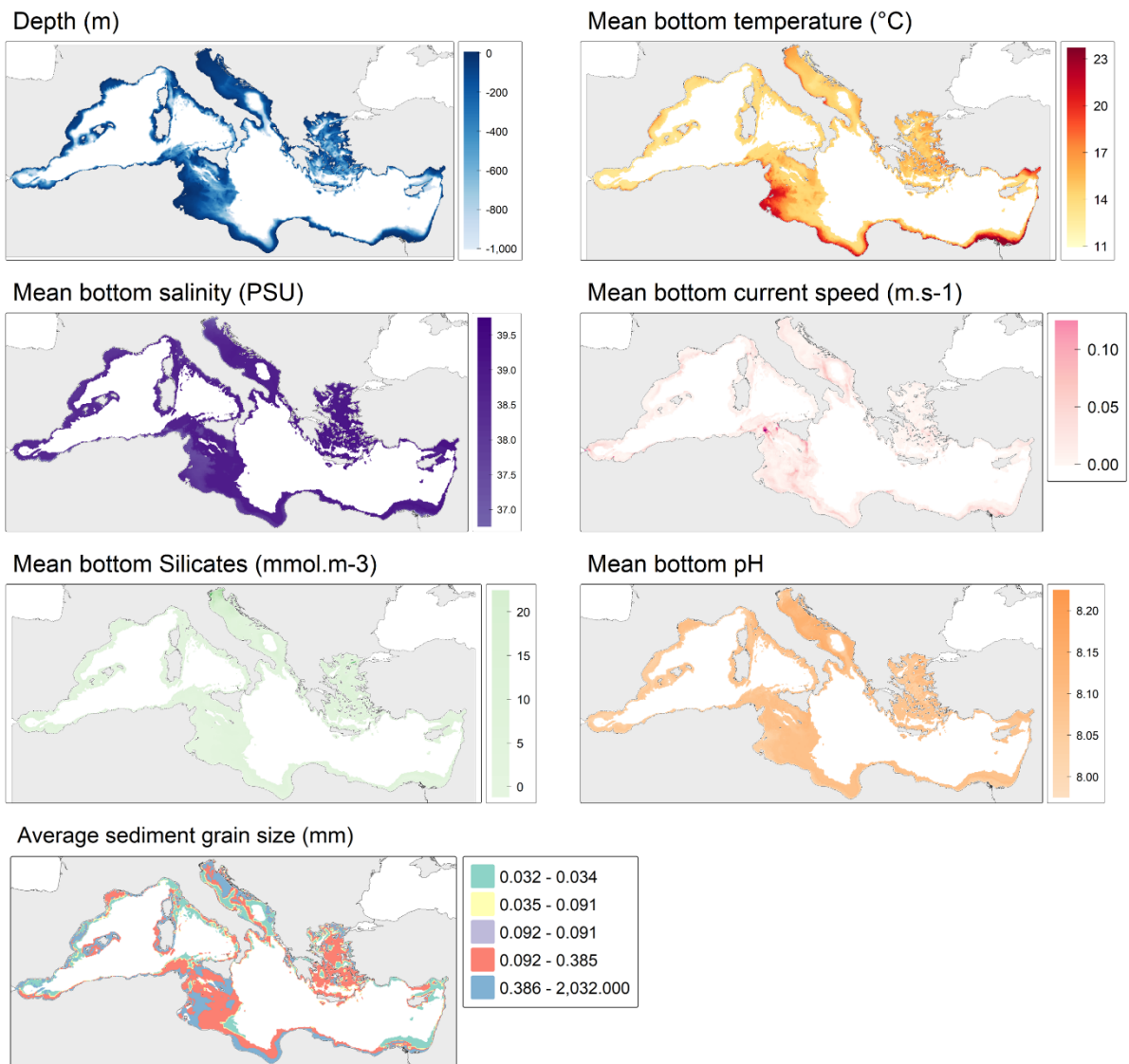


Figure S1.6: Maps of the seven environmental predictors selected to model the distribution of epibenthic megainvertebrate bioregions.

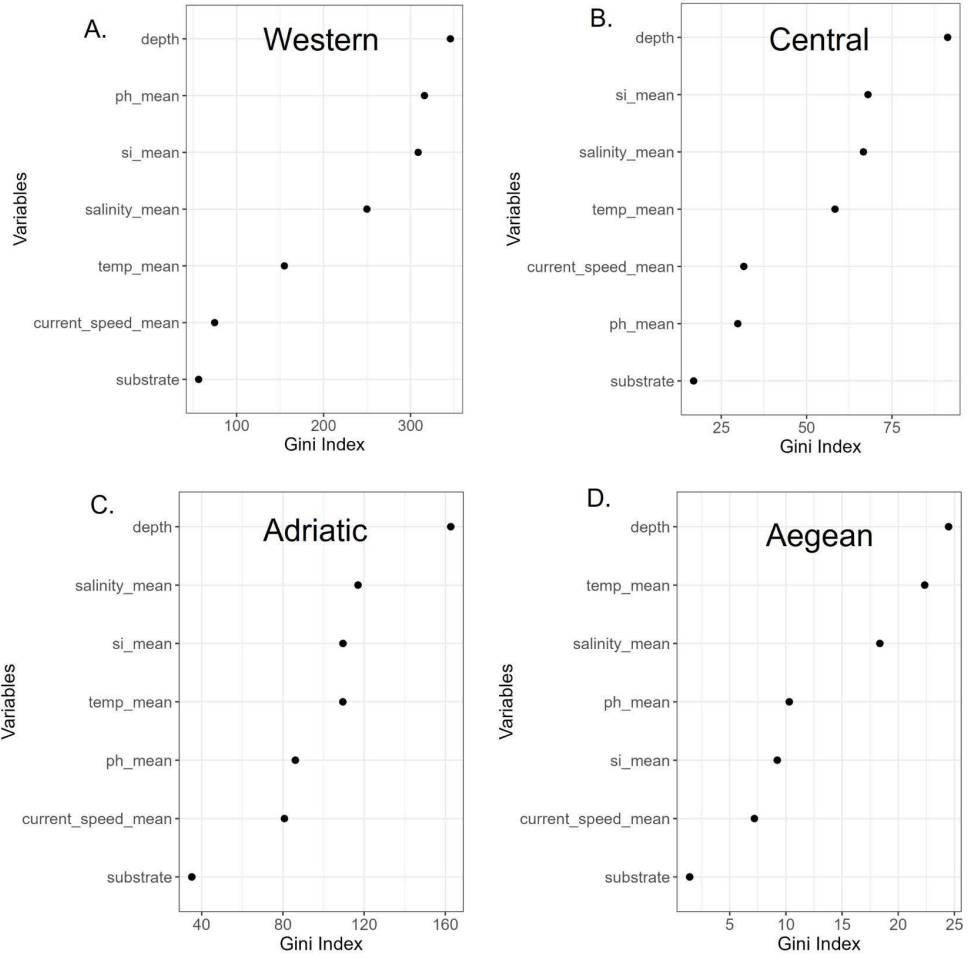


Figure S1.7: Environmental predictor contributions to Random Forest predictions in each sub-basin.

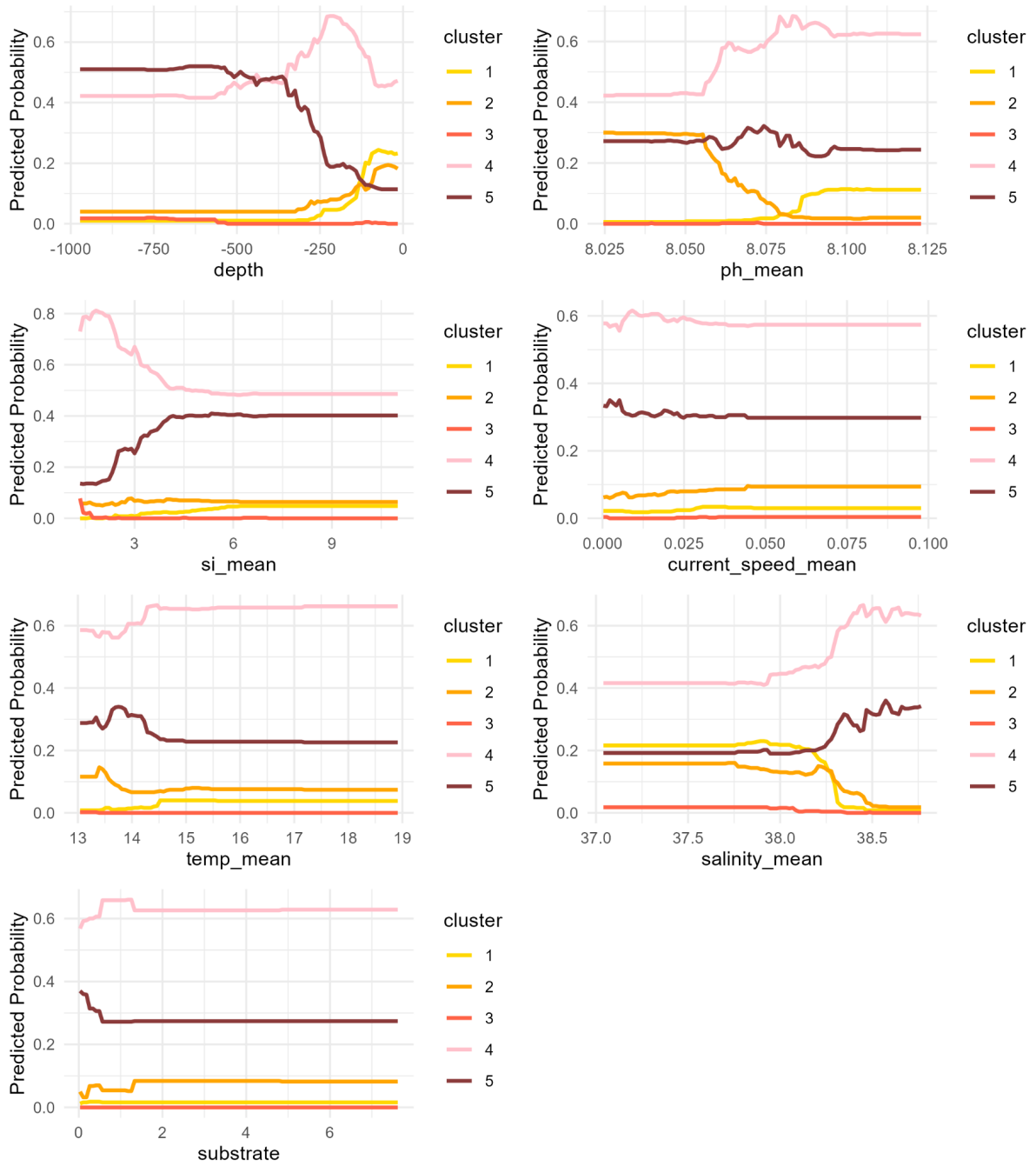


Figure S1.8: Partial response curves of the environmental predictors used in the Random Forest predictions for the Western Mediterranean.

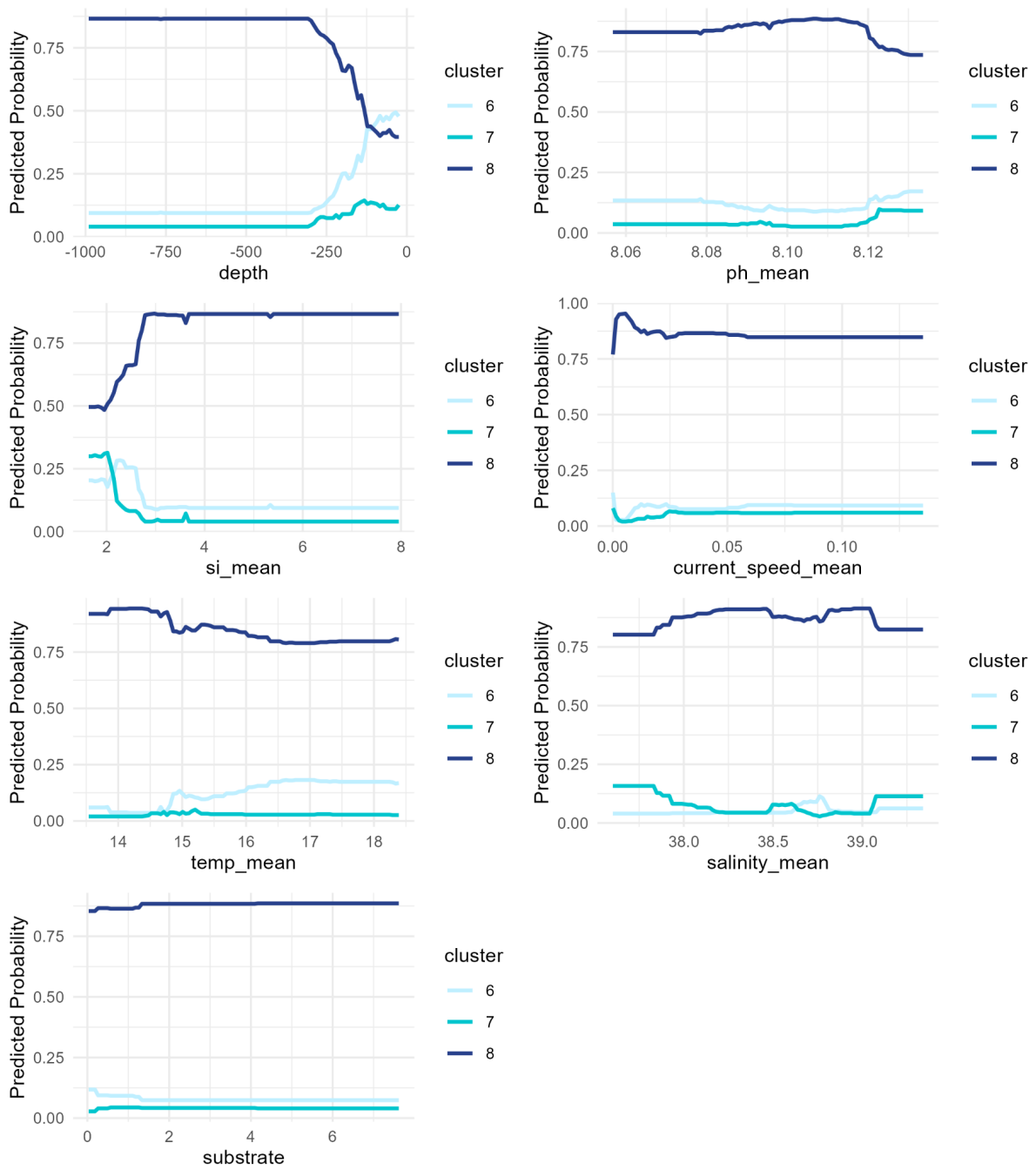


Figure S1.9: Partial response curves of the environmental predictors used in the Random Forest predictions for the Central Mediterranean.

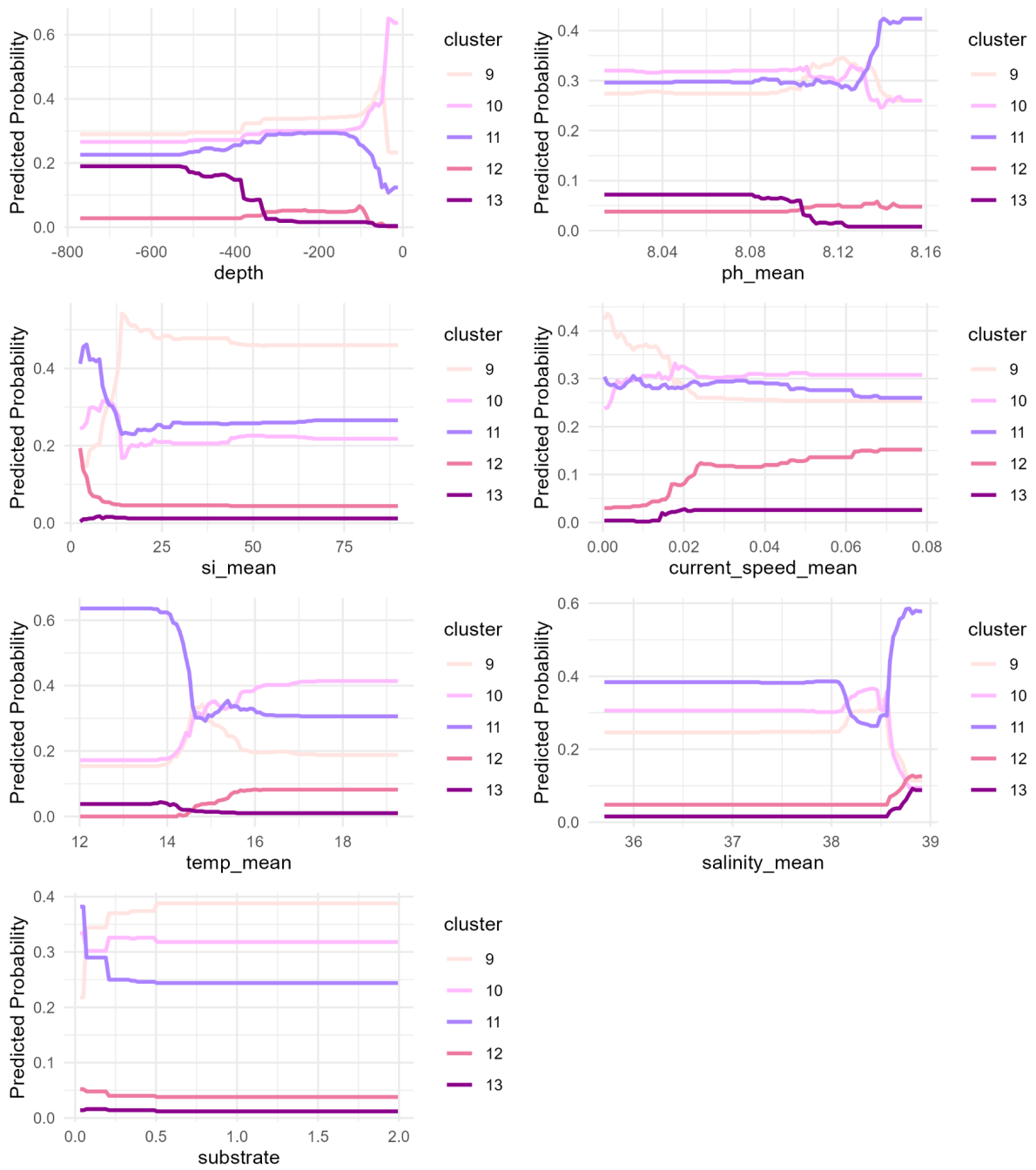


Figure S1.10: Partial response curves of the environmental predictors used in the Random Forest predictions for the Adriatic Sea.

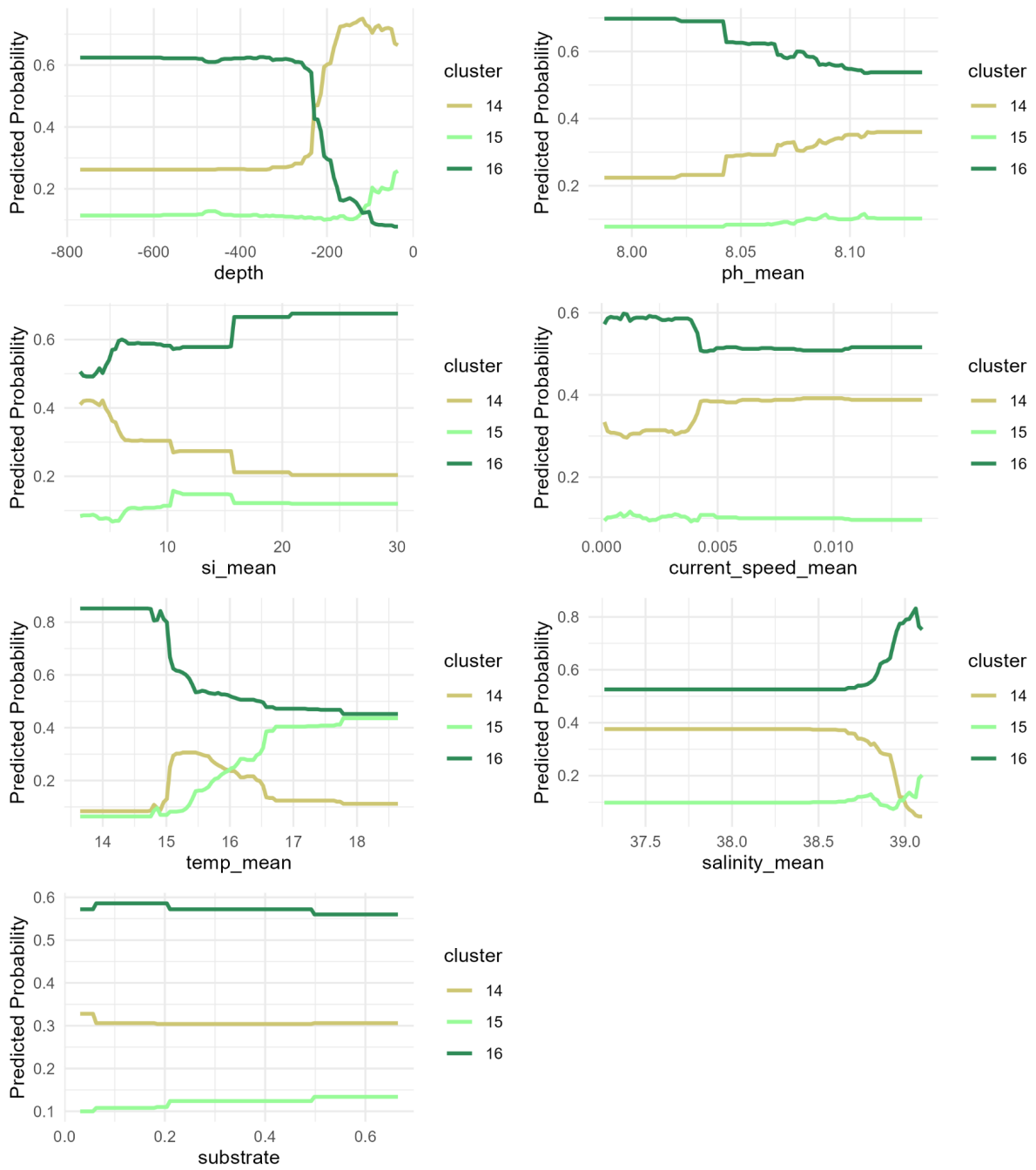


Figure S1.11: Partial response curves of the environmental predictors used in the Random Forest predictions for the Aegean Sea.

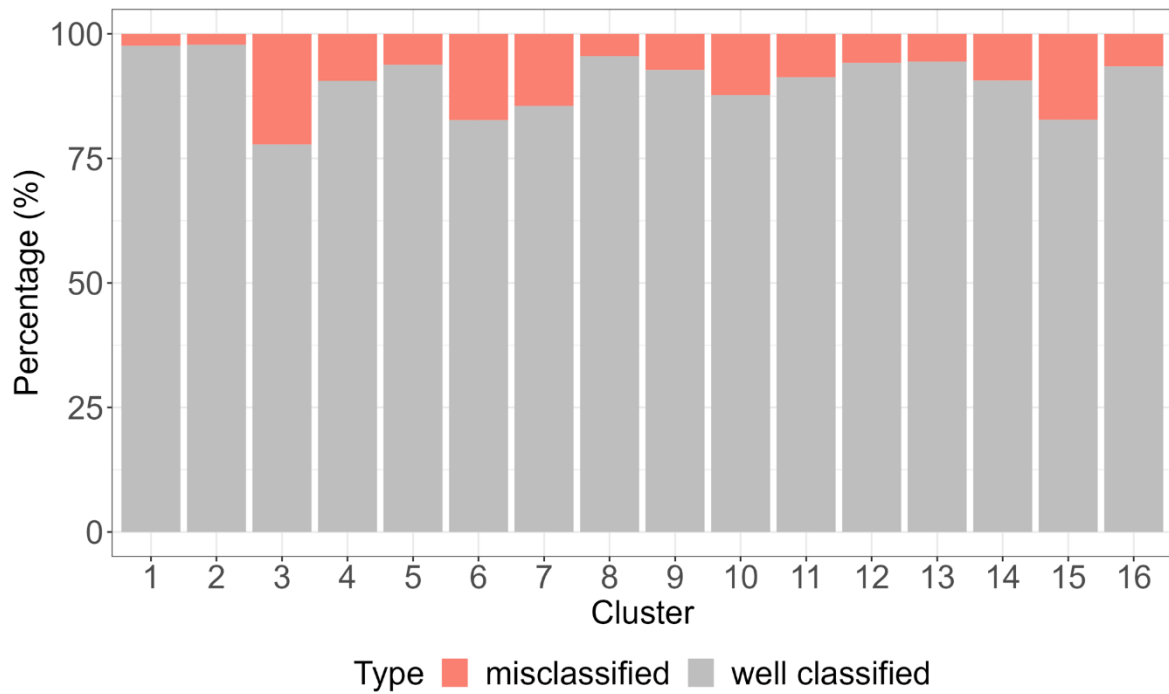


Figure S1.12: Percentage of misclassified stations per bioregion resulting from the Random Forest predictions compared to the k-means classification.