

Comisión Interamericana del Atún Tropical Inter-American Tropical Tuna Commission



A MACHINE LEARNING SPECIES DISTRIBUTION MODEL FOR THE CRITICALLY ENDANGERED EAST PACIFIC LEATHERBACK TURTLE (*DERMOCHELYS CORIACEA*)

Jon Lopez, Shane Griffiths, Bryan Wallace, Veronica Caceres, et al.
Document-Documento BYC-11-01

Data – Observations / Datos - observaciones

Summary of the observation data used in the model

Resumen de los datos utilizados en el modelo

No. Countries <i>Países</i>	No. Gears/ <i>Artes</i>	Range <i>Periodo</i>	Data type <i>Tipos de datos</i>	Effort <i>Esfuerzo</i>	No presences <i>presencias</i>	Total ind.	Total obs.	% av. Presences <i>Presencias</i>	Source <i>Fuente</i>
6 + IATTC	Several (LL, PS, gillnet, etc.)	1995 (2001)-2020	Abundance, presence only (Y/N)	Some	1,145	1,190	594,563	0.19	Various (observer, NGO, logbooks, etc.)



Data - Environmental information / Datos – variables ambientales

All available variable

Todas las variables disponibles

	Environmental										
Spatio-temporal	Surface						Subsurface			Static	
	Abiotic	Sp. Res	T. Res	Biotic	Sp. Res	T. Res		Sp. Res	T. Res		
Latitude	SST	1/4	D	CHL	1/4	W		Temp 100m	1/4	D	Bathymetry
Longitude	Δ SST	1/4	D	Δ CHL	1/4	W		MLD	1/4	D	Distance to coast
Day of the year	Salinity	1/4	D								
	SSH	1/4	D								
	Speed	1/4	D								
	Heading	1/4	D								
	EKE	1/4	D								
	FSLE	1/25	D								
	Front index	1/4	W								

Methods – modelling / Métodos - modelado

Machine learning algorithm – Boosted Regression Trees *Algoritmo de aprendizaje automatizado – Árboles de regresión*

Boosted
Regression
Trees

Tune the models

Simplify (if needed)

K-fold (10,
75 vs 25)

AUC/TSS/Dev

Sensitivity analysis and Final
Models / *Análisis de
sensibilidad y modelos finales*

Different
presence/absence ratios

50% to 0.5%

Compare different
performance metrics

Select candidates for
final model(s)

Ensemble
models-
predictions /
*Modelos-
predicciones
conjuntas*

Most likely models
based on statistics and
expert knowledge

Estimate statistical
threshold values

Select final distribution
map for EASI-Fish

Hierarchical
model building
(all data)/
*Modelado
jerarquizado*

Spatio-temporal

Surface

Subsurface

Environmental

Static

Full

Prediction /
Predicción

Daily since 2002
(>6500 days)

Averages per year
and prediction
period



Results – performance metrics / Resultados – métricas de desempeño

		lr	n.trees	Dev	AUC	TSS	Drop
1	fit	0.01	4500	38.64	0.89	0.66	
ST	simp	-	-	-	-	-	No drop
2	fit	0.01	5150	32.28	-	-	
SUR	simp		5150	32.40	0.92	0.71	EkE
3	fit	0.01	3600	23.62	0.88	0.65	
SUB	simp	-	-	-	-	-	No drop
4	fit	0.01	5600	38.14	0.94	0.74	
ENV	simp	-	-	-	-	-	No drop
5	fit	0.01	4400	19.71	0.79	0.51	
STA	simp	-	-	-	-	-	No drop
6	fit	0.01	7000	40.57	0.94	0.76	
FULL	simp	-	-	-	-	-	No drop

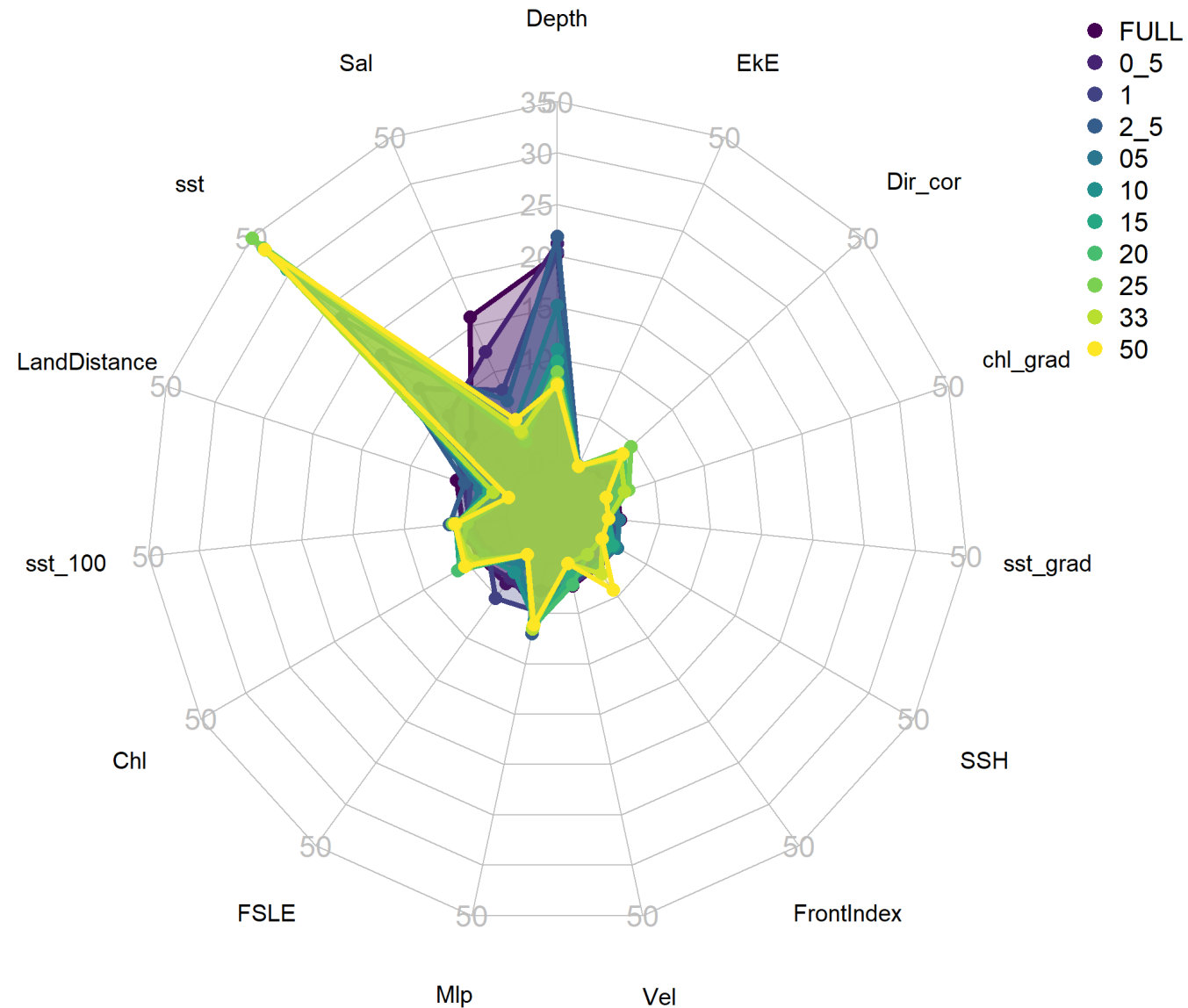


Results – Sensitivity Analysis / Resultados – Análisis de sensibilidad

		lr	n.trees	Dev	AUC	TSS	Drop
50	fit	0.005	2900	58.10	-	-	
n = 2176	simp	0.005	3650	59.37	0.92	0.71	7 variables
33	fit	0.01	2050	59.16	-	-	
n = 3264	simp	0.01	2650	61.54	0.94	0.76	5 variables
25	fit	0.01	2700	59.70	-	-	
n = 4352	simp	0.01	3250	60.42	0.96	0.81	7 variables
20	fit	0.01	3200	60.29	-	-	
n = 5440	simp	0.01	3350	59.14	0.94	0.77	6 variables
15	fit	0.01	3950	60.21	-	-	
n = 7250	simp	0.01	3650	58.96	0.94	0.76	3 variables
10	fit	0.01	5050	60.93	-	-	
n = 10880	simp	0.01	4800	60.16	0.95	0.78	2
5	fit	0.01	5300	56.36	0.95	0.78	
n = 21760	simp	-	-	-	-	-	No drop
2.5	fit	0.01	6400	54.67	-	-	
n = 43520	simp	0.01	5650	52.08	0.95	0.77	7 variables
1	fit	0.01	6650	49.24	0.94	0.78	
n = 108800	simp	-	-	-	-	-	No drop
0.5	fit	0.01	6100	45.23	0.94	0.76	
n = 217600	simp	-	-	-	-	-	No drop
0.19	fit	0.01	7000	40.57	0.94	0.76	
n = 573889	simp	-	-	-	-	-	No drop

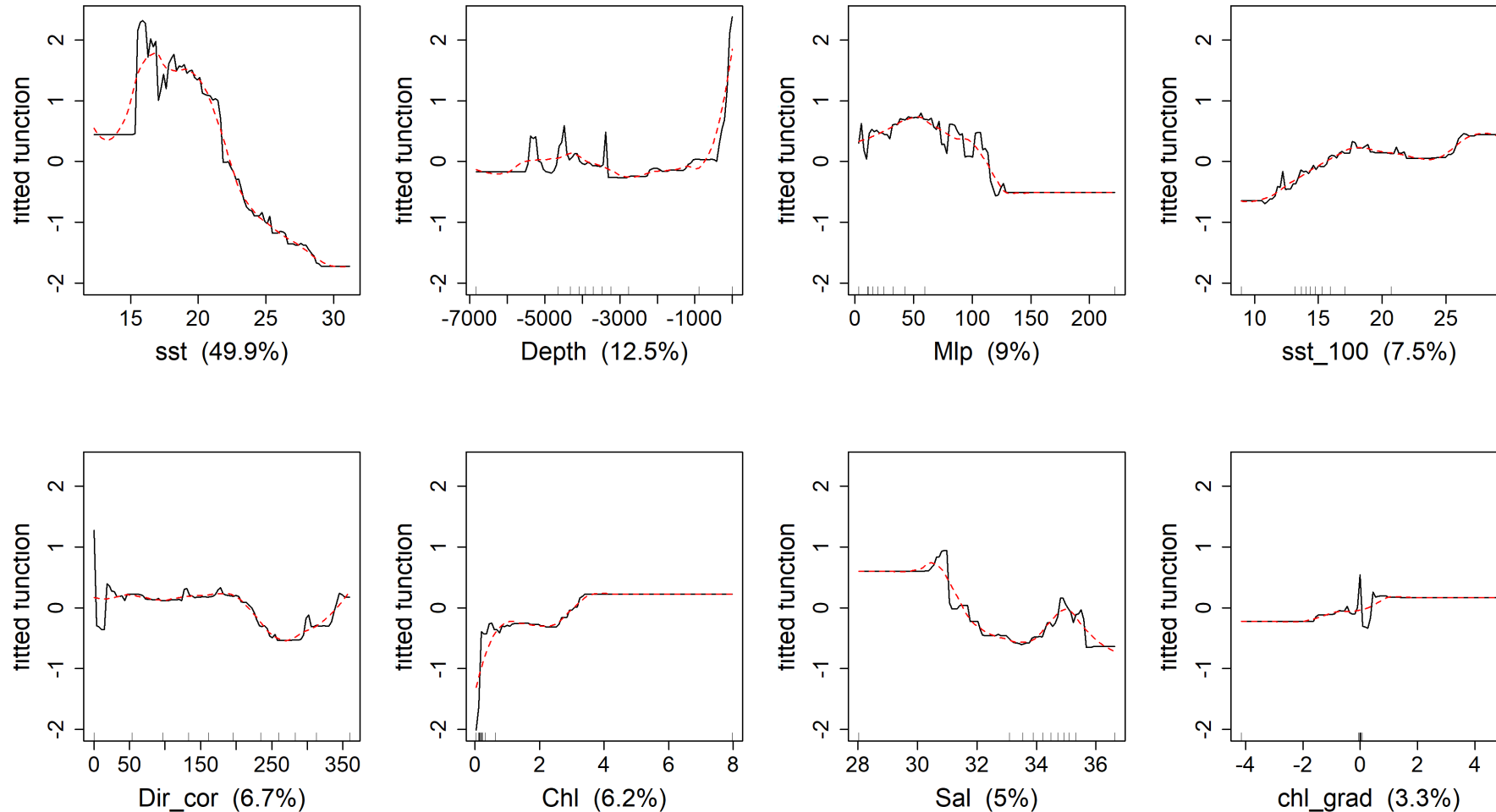


Results – variable importance / Resultados – importancia de las variables

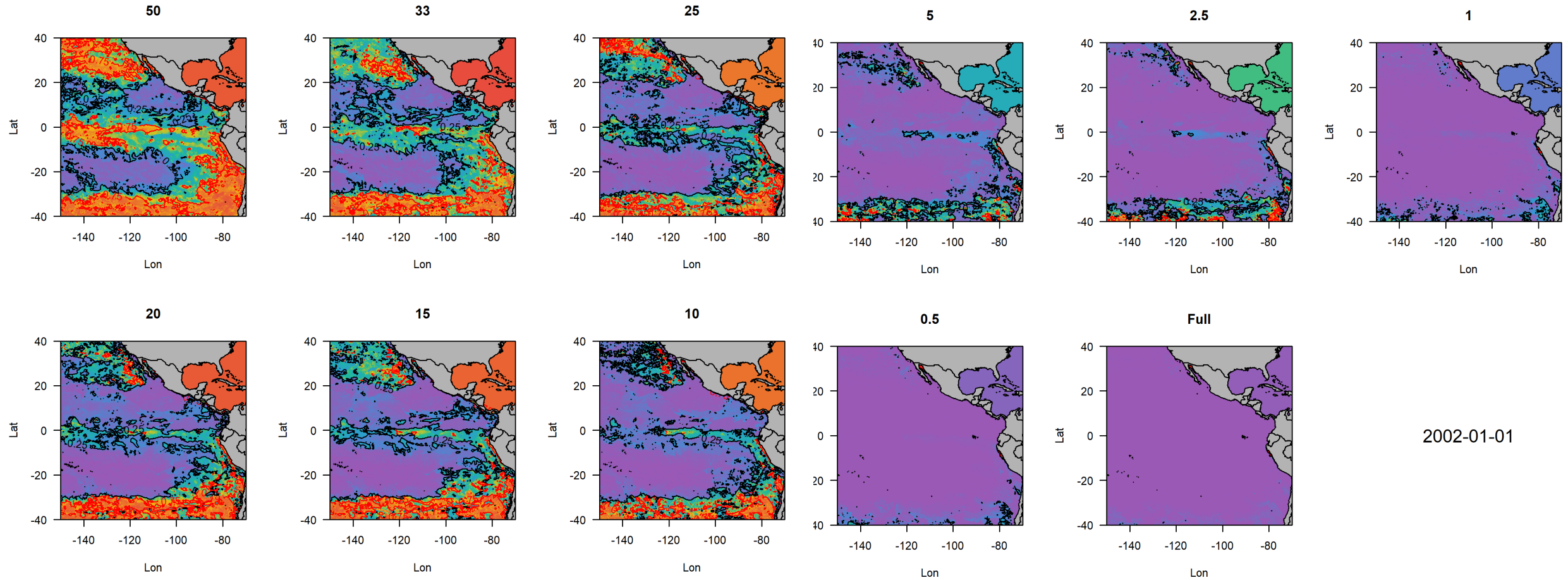


Results – partial dependence plots / Resultados – figuras de dependencia parcial

Example – 25% presence-absence ratio model
Ejemplo – modelo con un ratio de presencia-ausencia del 25%



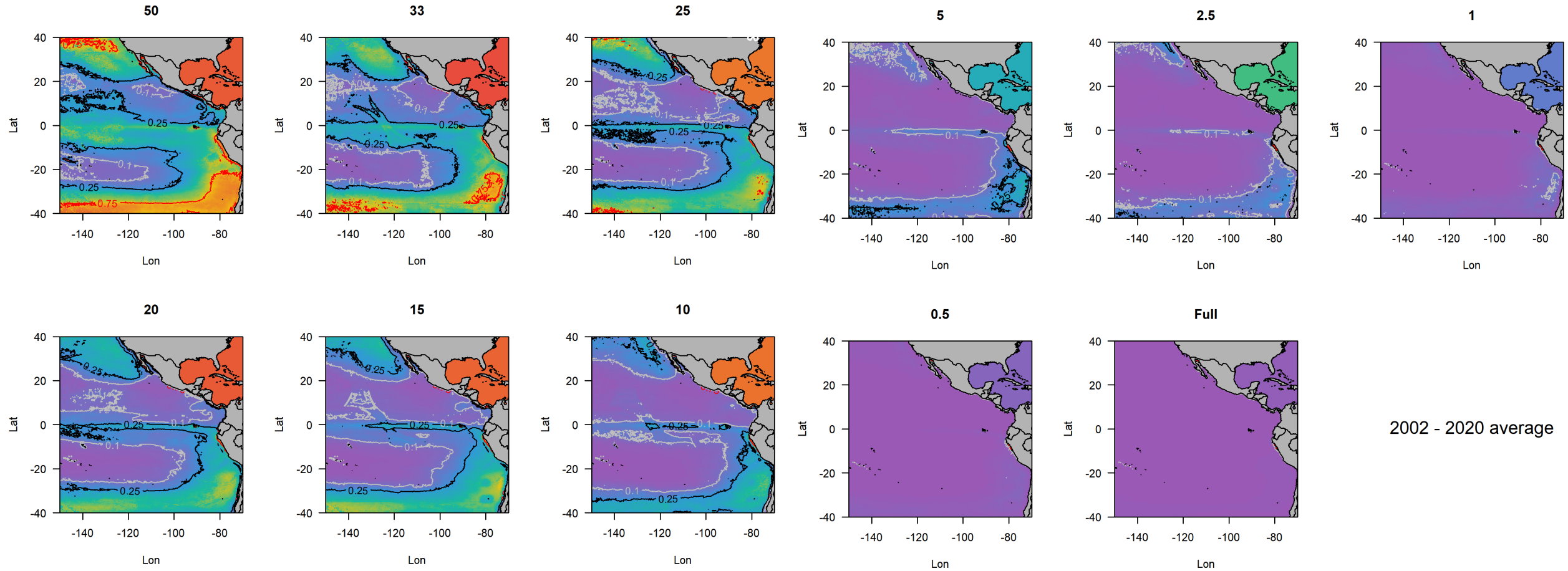
Results – daily predictions / *Resultados – predicciones diarias*



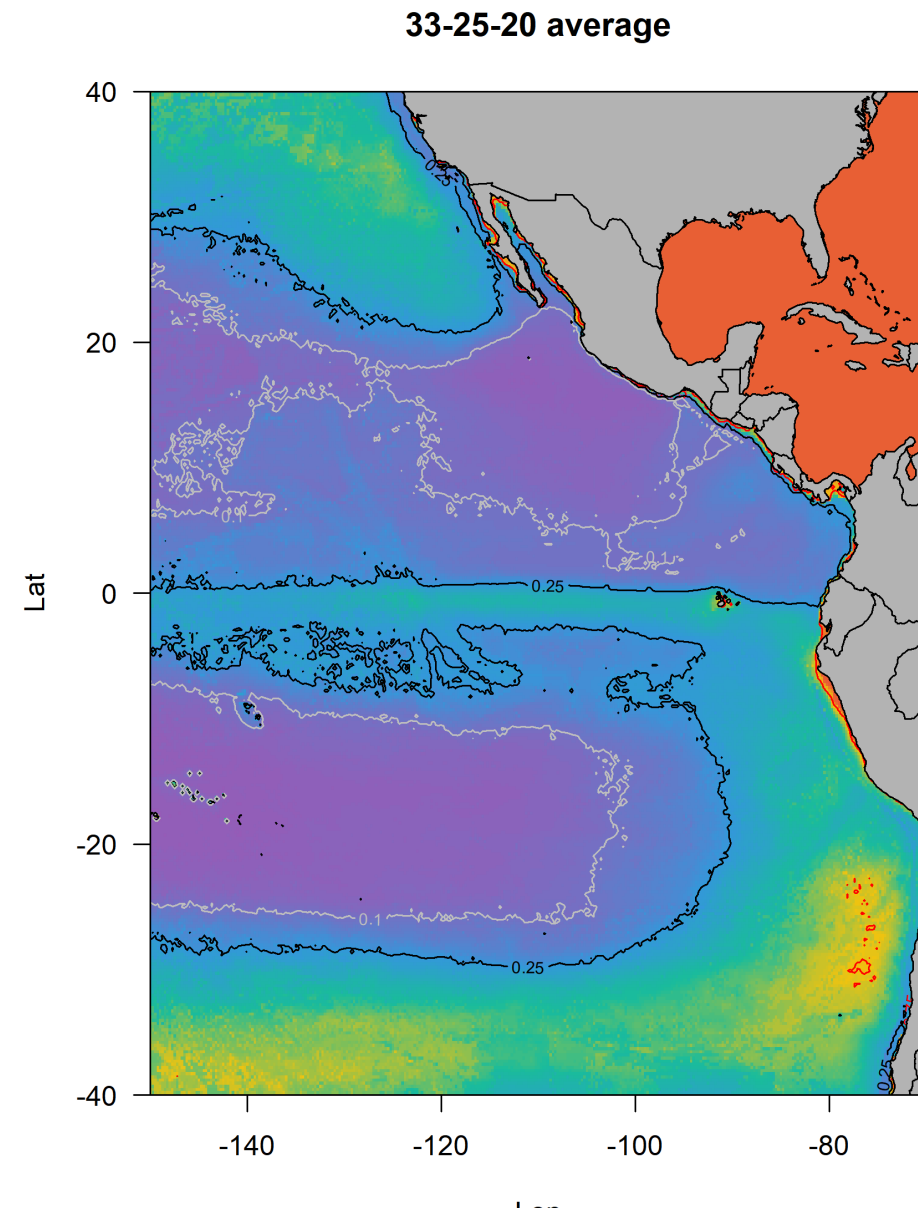
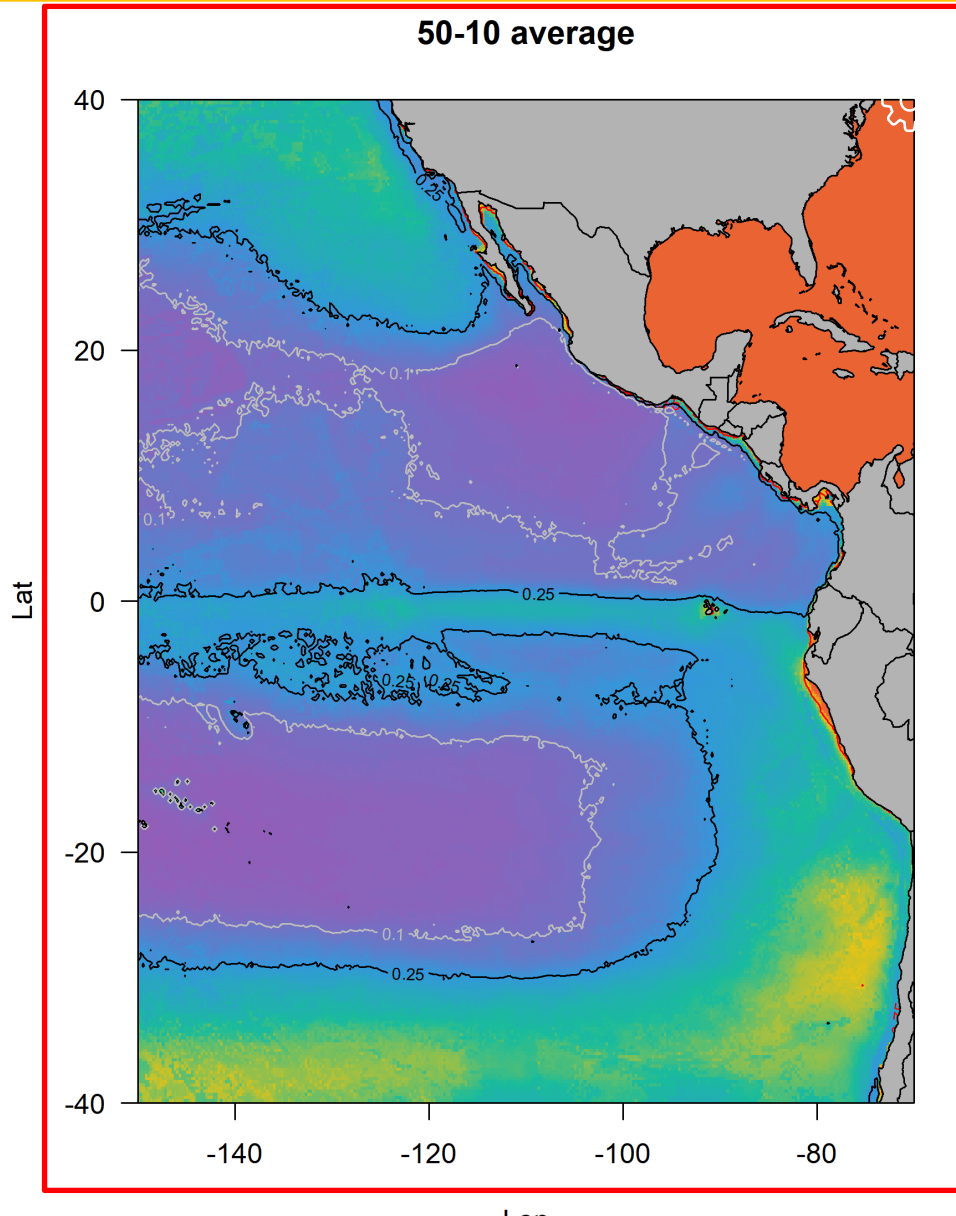
2002-01-01



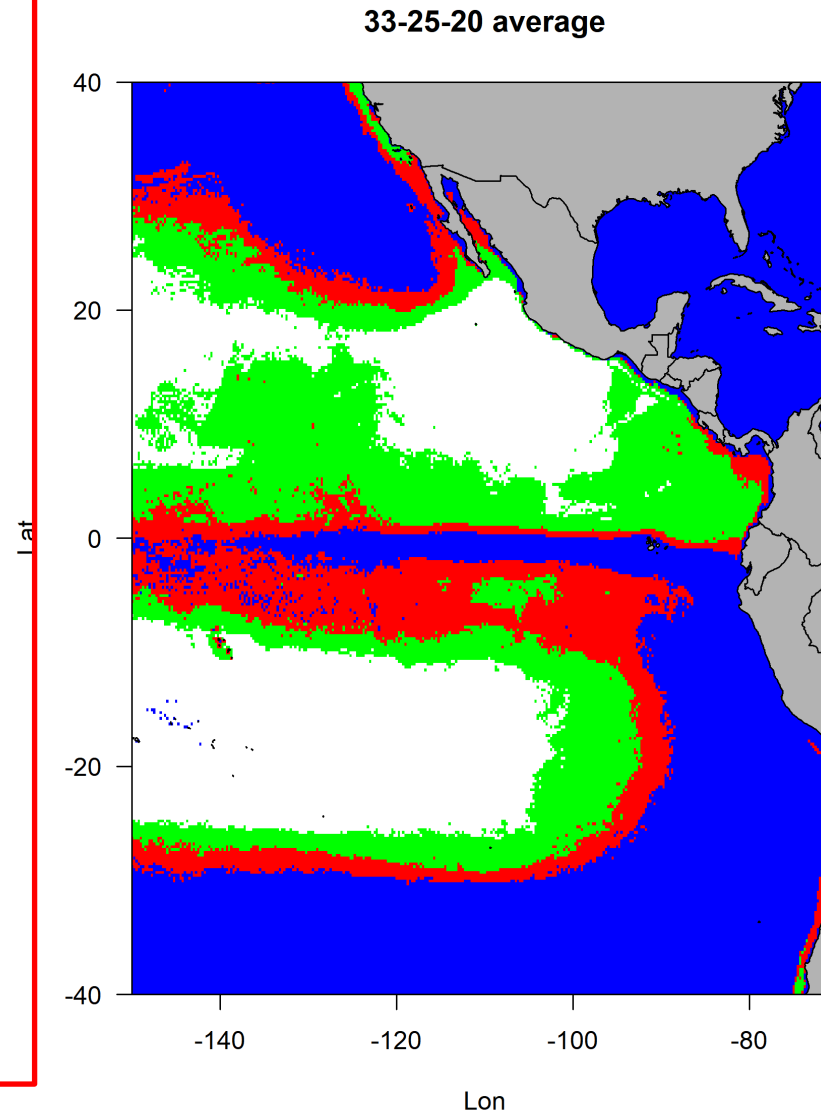
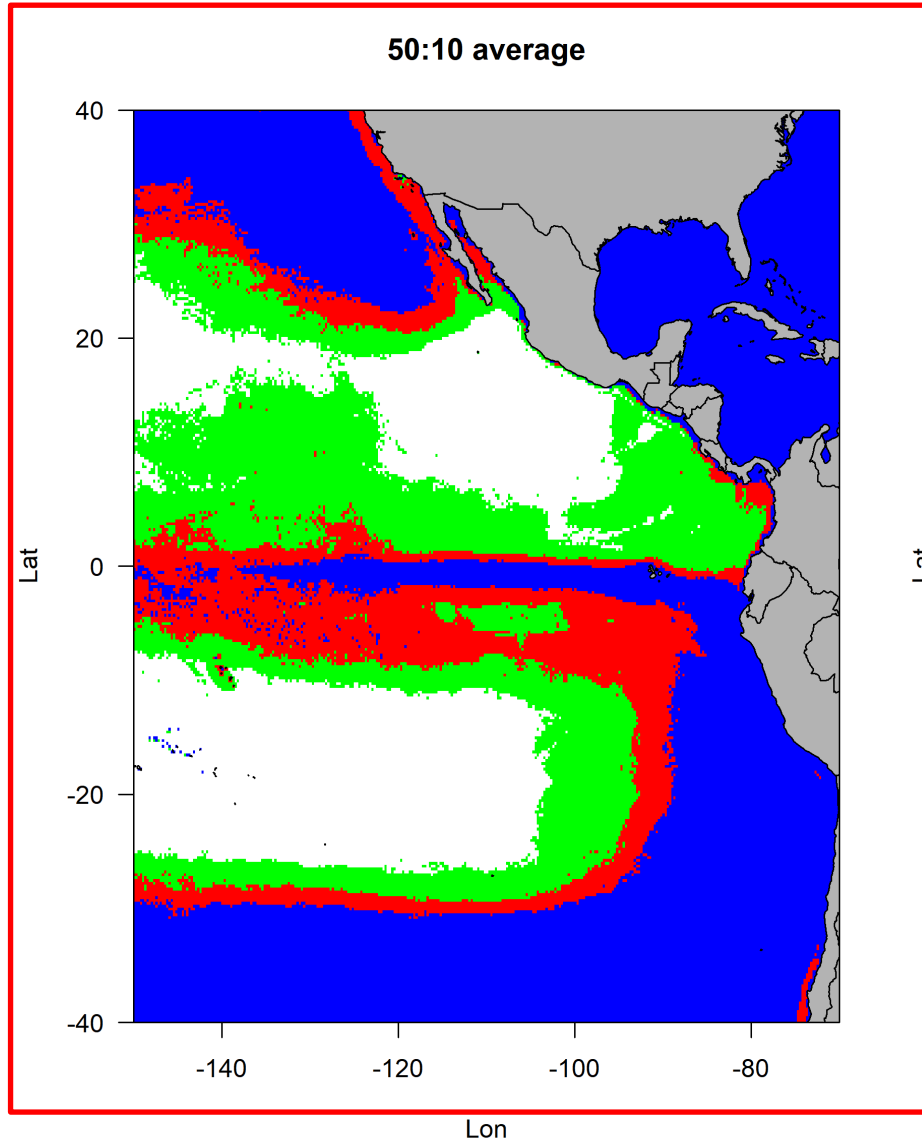
Results – average predictions / Resultados – predicciones medias (2002-2020)



Results – ensemble predictions / Resultados – predicciones conjuntas



Results – ensemble predictions / Resultados – predicciones conjuntas



Thresholds:
0.1 – Green
0.2 – Red
0.3 – Blue





Questions? / Preguntas?