

Modelado de capturas incidentales lobo marino pesquería demersal centro - sur

Joaquin Cavieres
PhD (c) in Statistics

Objetivos

Objetivo general

Desarrollo y capacitación sobre alternativas de modelamiento estadístico, a través de la plataforma estadística R, que permitan estimar capturas incidentales (incluida mortalidad), reconociendo los probables factores que inciden en sus magnitudes.

Objetivo específico 1

- Proponer alternativas de modelamiento a través de aproximaciones probabilísticas y/o Bayesiana, que permitan estimar capturas incidentales, identificando los eventuales factores que inciden en ello.

Objetivo específico 2

- Desarrollar un modelo espacio-temporal a través de aproximaciones probabilísticas o Bayesiana en ambiente R, que permita la estimación de las capturas incidentales y/o interacciones.

Objetivo específico 3

- Generar e implementar una rutina (script) con los modelos propuestos en R, desarrollando una capacitación teórico-práctica para los investigadores. Se considerará como caso de estudio, la pesquería de arrastre sur austral. El IFOP facilitará al consultor que se adjudique el trabajo, toda la información colectada en los monitoreos de pesquerías desarrollados, al igual que estudios previos internos realizados sobre el tema.

Metodología

Objetivo específico 1

- Se proponen los clásicos Modelos Lineales Generalizados (GLM's, por sus siglas en Inglés) para determinar los factores con mayor relevancia en la explicación de las capturas incidentales de los mamíferos marinos.
- Adicionalmente se creó un GLM bajo el enfoque Bayesiano el cual nos permitirá obtener distribuciones a posterior de los parámetros de interés. Se utilizó lo propuesto por Rue et al., 2009 y Martins et al. 2013 "Integrated Nested Laplace Approximation"

INLA <https://www.r-inla.org/>

Presentación

https://github.com/jcavieresg/tutoriales/blob/main/Intro_INLA_JCavieres.pdf

Objetivo específico 2

- El modelo explorado fue diseñado en R (R Core Team, 2017), con la librería **INLA** (www.r-.org) y la metodología propuesta por Rue et al. (2009), Martins et al. (2013) y Lindgren et al. (2011).

Presentación <https://github.com/jcavieresg/tutoriales/blob/main/gmrf.pdf>

Objetivo específico 3

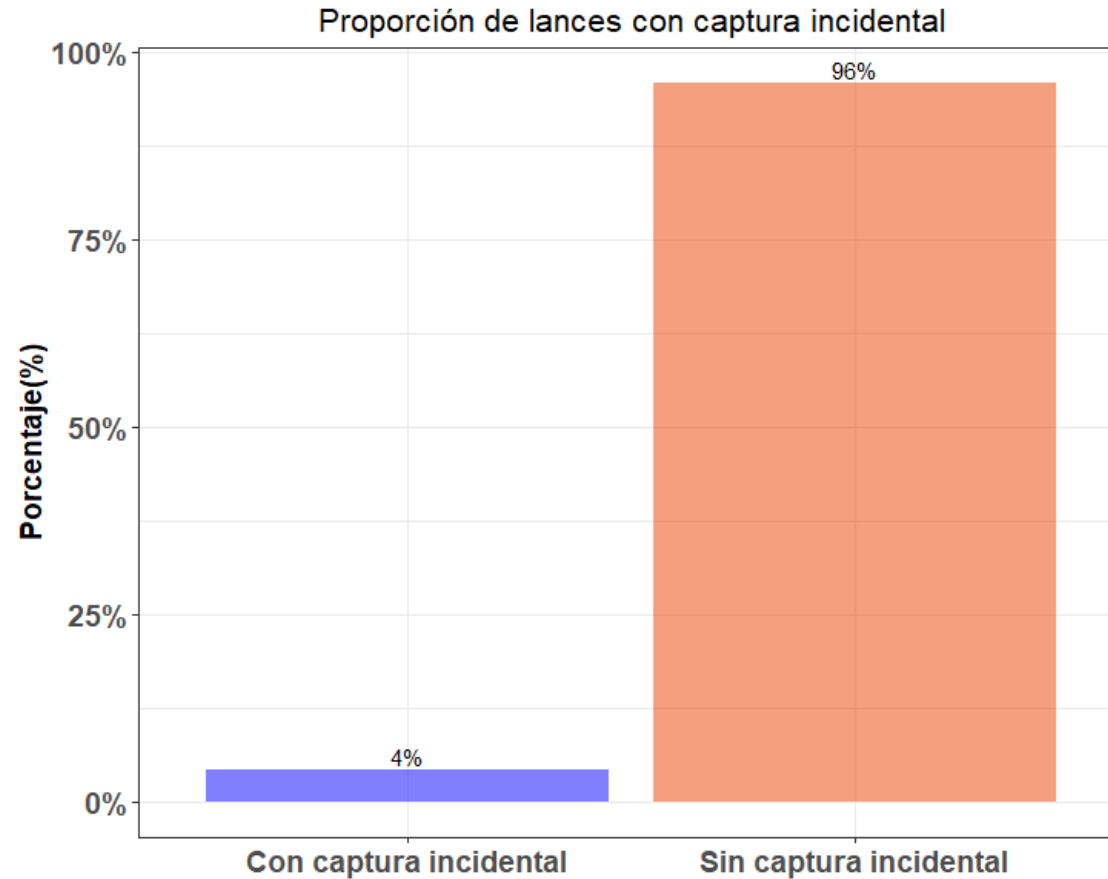
- Implementación de distintas aproximaciones de GLM's para determinar principales variables que pueden responder a las capturas incidentales de mamíferos marinos.
- Exploración y construcción de un modelo espacio-temporal en donde la parte espacial estará compuesta por un SPDE/GRMF y la parte temporal a través de la variable años modelada como un efecto aleatorio.
- Capacitación para comprender el algoritmo de construcción de los modelos propuestos junto a los datos relacionados a la pesquería de arrastre sur austral.

Análisis preliminares

La información considerada está relacionada a la variable 'OBS_CIAMT'. Este campo contiene dos posibles valores, '1' es cuando el lance fue observado efectivamente por observador científico a bordo del barco. El otro valor corresponde a '0' que es cuando el lance no fue observado efectivamente por el observador. En este análisis sólo se considera el valor '1' que es cuando efectivamente se observó el lance.

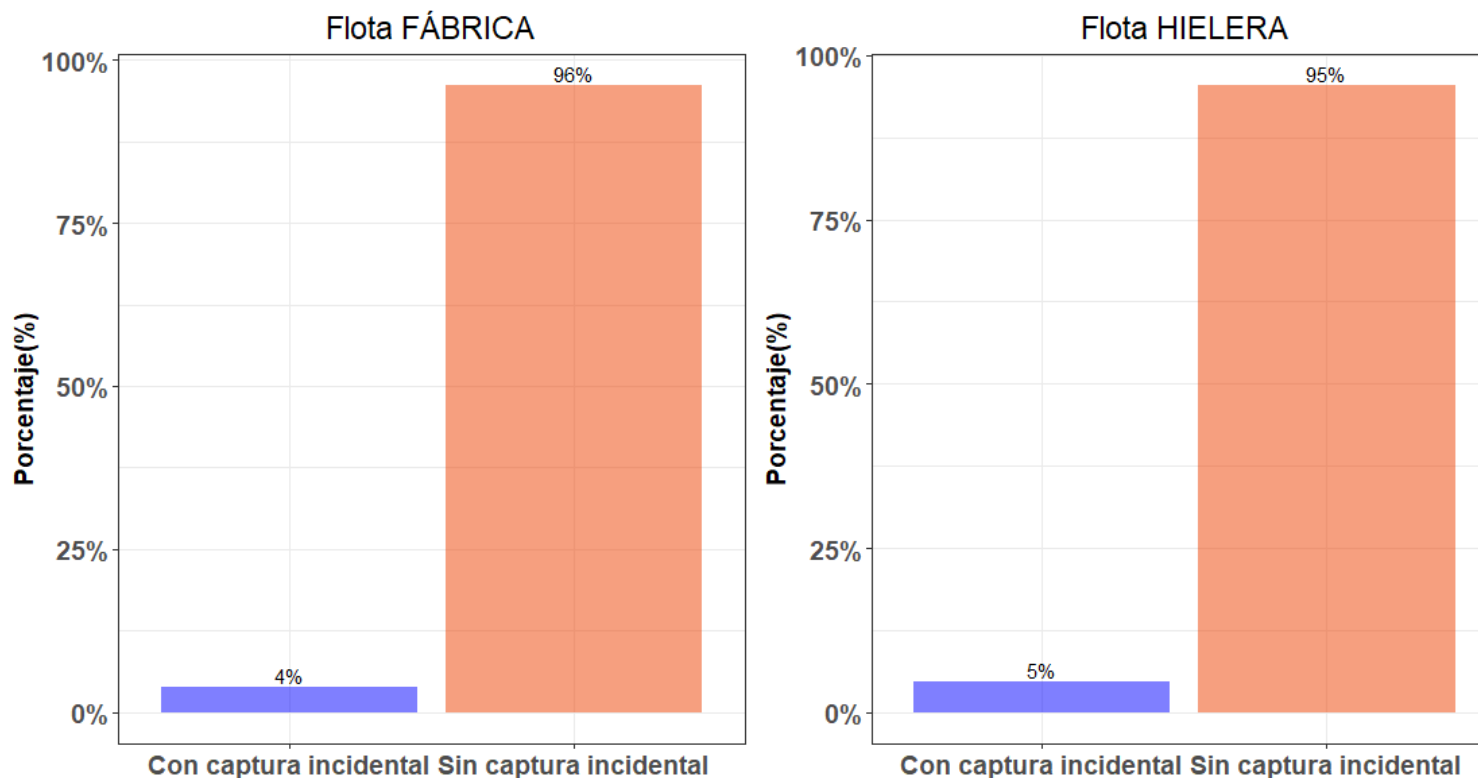
La base de datos inicial contiene 17910 lances observados pero luego del filtro, sólo considerando el campo 'OBS_CIAMT' = '1', la nueva base contiene 10560 lances observados.

Proporción total de lances con captura incidental

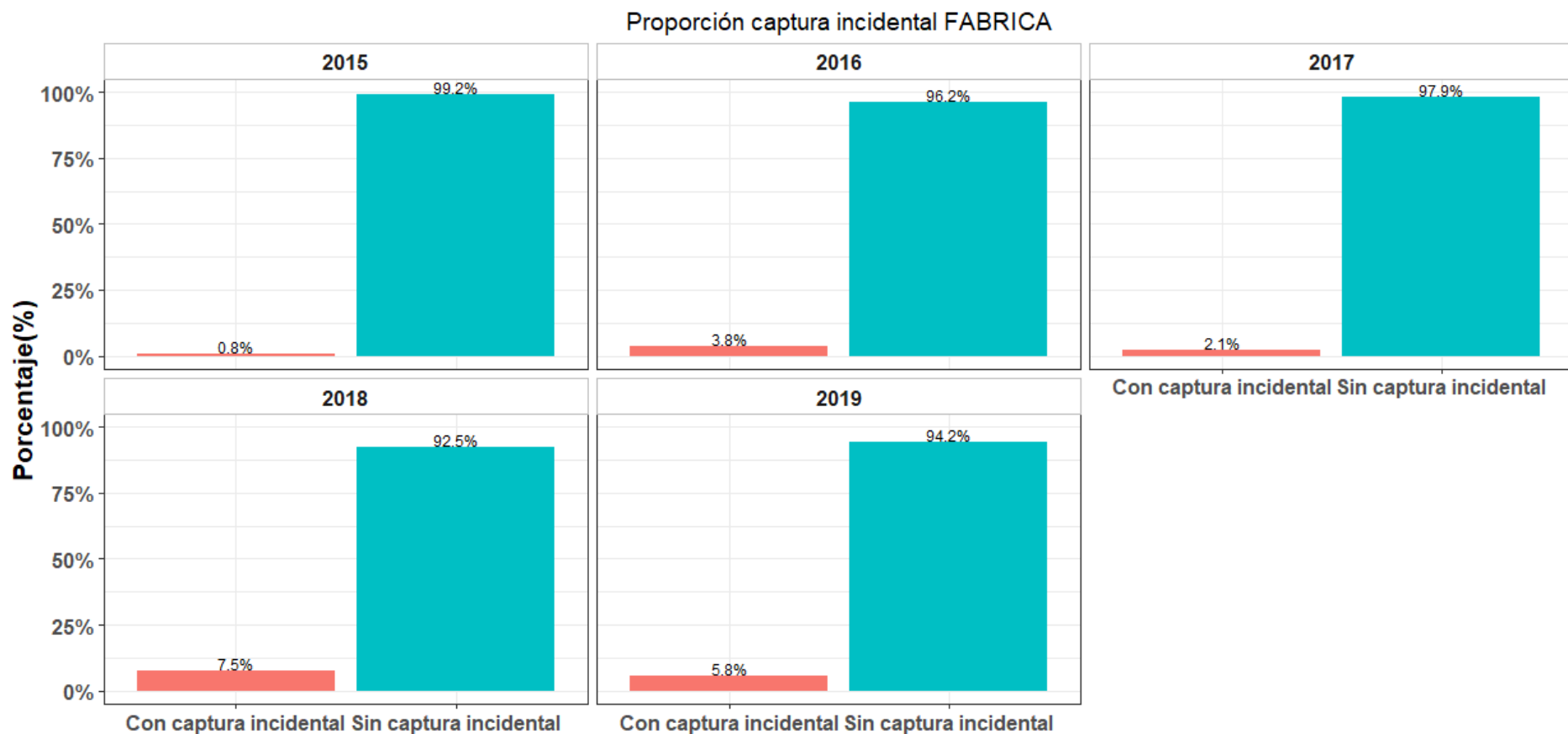


Porcentaje de lances con captura incidental de lobo marino y lobo fino, Pesquería Sur Austral, Chile.

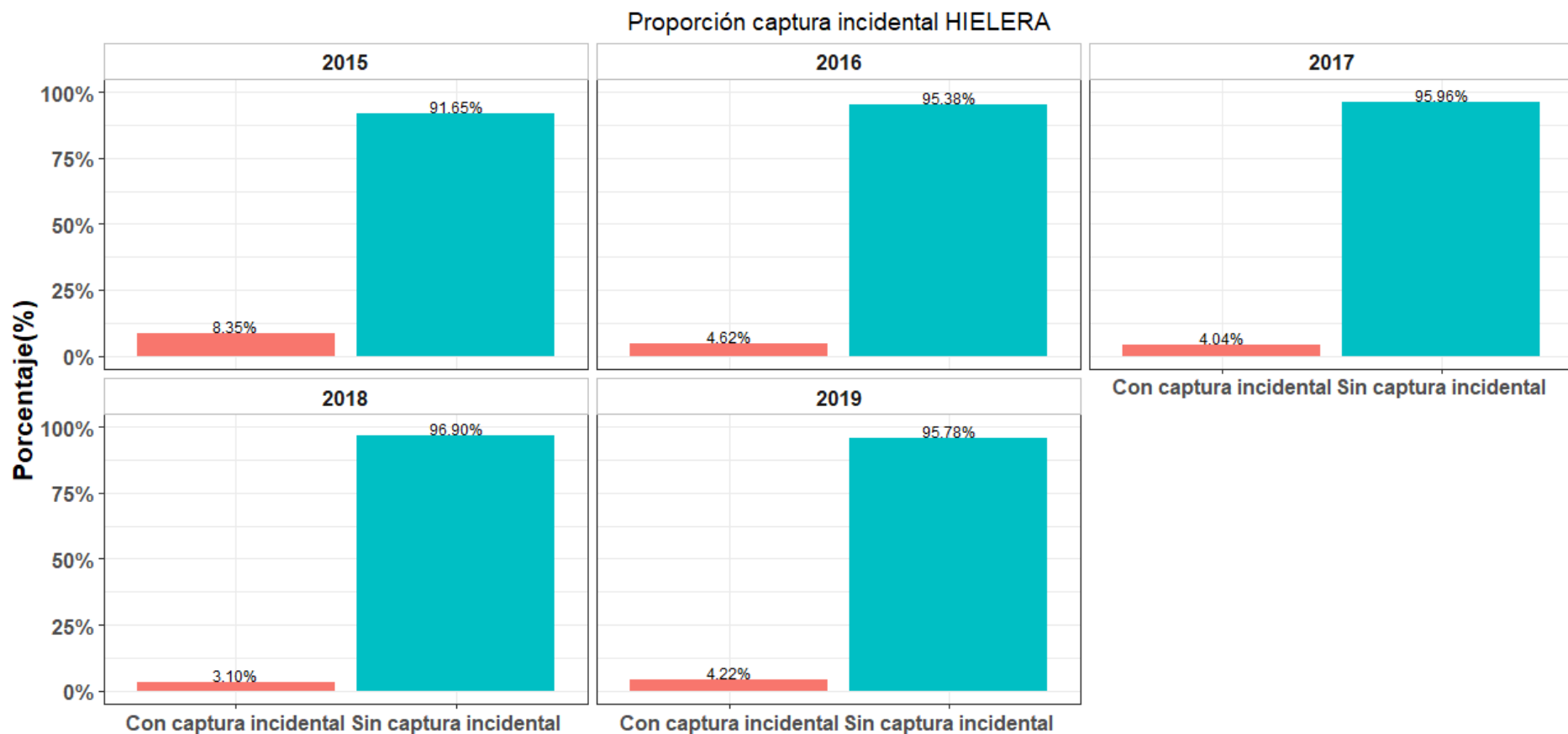
Proporción total de lances con captura incidental por flota



Porcentaje de lances observados con capturas incidentales y sin captura incidental por tipo de flota que participa en la Pesquería Sur Austral, Chile.



Porcentaje de lances observados con captura incidental y sin captura incidental por los barcos 'Fábrica' durante la serie temporal registrada, Pesquería Sur Austral, Chile.



Porcentaje de lances observados con captura incidental y sin captura incidental por los barcos 'Hieleros' durante la serie temporal registrada, Pesquería Sur Austral, Chile.

Lances de pesca con captura incidental por tipo de pesquería.

Nº Lobos	Nº lances Fábrica	Nº lances Hielero	% del total Fábrica	% del total Hielero
0	6417	3697	60.78%	35.02%
1	151	97	1.43%	0.92%
2	64	39	0.61%	0.37%
3	17	12	0.16%	0.11%
4	19	10	0.18%	0.09%
5	4	7	0.04%	0.07%
6	5	6	0.05%	0.06%
7	1	3	0.01%	0.03%
8	2	1	0.02%	0.01%
9	1	0	0.01%	0.00%
10	1	0	0.01%	0.00%
11	0	1	0.00%	0.01%
12	0	1	0.00%	0.01%
14	0	1	0.00%	0.01%
20	0	1	0.00%	0.01%

Media en número de ejemplares capturados, máximo número de ejemplares capturados y número de lances observados por año para cada flota.

Año	Barco	media_ejemplar_capt	max_n_ejemplar	n_lances_obs
2015	Fábrica	0.01	6	1153
2015	Hielero	0.24	20	503
2016	Fábrica	0.10	9	1574
2016	Hielero	0.11	7	1105
2017	Fábrica	0.04	4	1463
2017	Hielero	0.06	4	989
2018	Fábrica	0.12	10	1267
2018	Hielero	0.06	8	710
2019	Fábrica	0.09	8	1225
2019	Hielero	0.11	14	569

Resultados OE 1

Modelado Estadístico

En primera instancia se modelaron los datos de conteo de capturas incidentales a través de una distribución de Poisson considerando a las siguientes variables dentro del predictor lineal:

- ANO = Año en el cual fue registrada la observación. La serie contempla la información desde el año 2015 a 2019. Cada año es una categoría o nivel dentro del modelo.
- COD_PESQUERIA = Código de las pesquerías que tienen participación en las capturas incidentales de lobos marinos. las asignaciones para cada pesquería son: 9 = Flota Fábrica y 4 = Flota Hielera.
- TIPO_DE_RED = Tipo de red utilizada por los barcos. Las categorías asociadas a esta variable son: 1 = Arrastre de Fondo, 2 = Media agua y 3 = Multipropósito.

- CLASE_LANCE = Asignación del lance de pesca en el cual se observó efectivamente la captura incidental. Las categorías para esta variable son: 1 = Lance proyecto de seguimiento y 2 = Lance para proyecto de Descarte.
- MODELO_EXCLUSION = Esta variable describe el tipo de dispositivo utilizado en las redes de pesca para evitar las capturas incidentales de lobos marinos. Las categorías asociadas a esta variable son: 1 = Con dispositivo, 2 = Sin dispositivo, 3 = Con dispositivo modificado, 4 = Sin información y 9 = No observado.
- INTENSIDAD_VIENTO_AR = Esta variable tiene 13 categorías: 0 = Calmo, 1 = Brisa Tenue, 2 = Brisa Ligera, 3 = Brisa Suave, 4 = Brisa Moderada, 5 = Brisa, 6 = Brisa Fuerte, 7 = Borrascoso, 8 = Tempestad, 9 = Tempestad Fuerte, 10 = Temporal, 11 = Temporal Violento, 12 = Huracán, 13 = Sin información.
- ESPECIE_OBJETIVO_LANCE: Esta variable contiene las siguientes categorías: 2 = Merluza del Sur, 3 = Merluza de tres aletas, 4 = Merluza de cola, 6 = Congrio dorado, 27 = Reineta, 5 = Brótola, 29 = Cojinoba azul y 96 = Cojinoba del sur

Modelos propuestos

- Poisson
- Binomial Negativo
- Zero-Inflated Poisson
- Zero-Inflated Binomial Negativo

Variables de entrada

- ANO
- COD_PESQUERIA
- TIPO_DE_RED
- CLASE_LANCE
- MODELO_EXCLUSION
- INTENSIDAD_VIENTO_AR
- ESPECIE_OBJETIVO_LANCE
- PROFUNDIDAD_FROM

Se construyeron dos tipos de predictores:

- **Predictor 1**
Contiene a todas las variables de entrada.
- **Predictor 2**
No contiene las variables 'INTENSIDAD_VIENTO_AR' y 'ESPECIE_OBJETIVO_LANCE.

Resultados GLM (frecuentista)

Modelos	Variable respuesta	Predictor lineal	Baja/Sobre-dispersión
Poisson	N_EJEMPLAR	Predictor_1	2.405042
Binomial-negativa	N_EJEMPLAR	Predictor_1	1.051372
Zero-inflated Poisson	N_EJEMPLAR	Predictor_2	1.135581
Zero-inflated Binomial negativa	N_EJEMPLAR	Predictor_2	0.966073

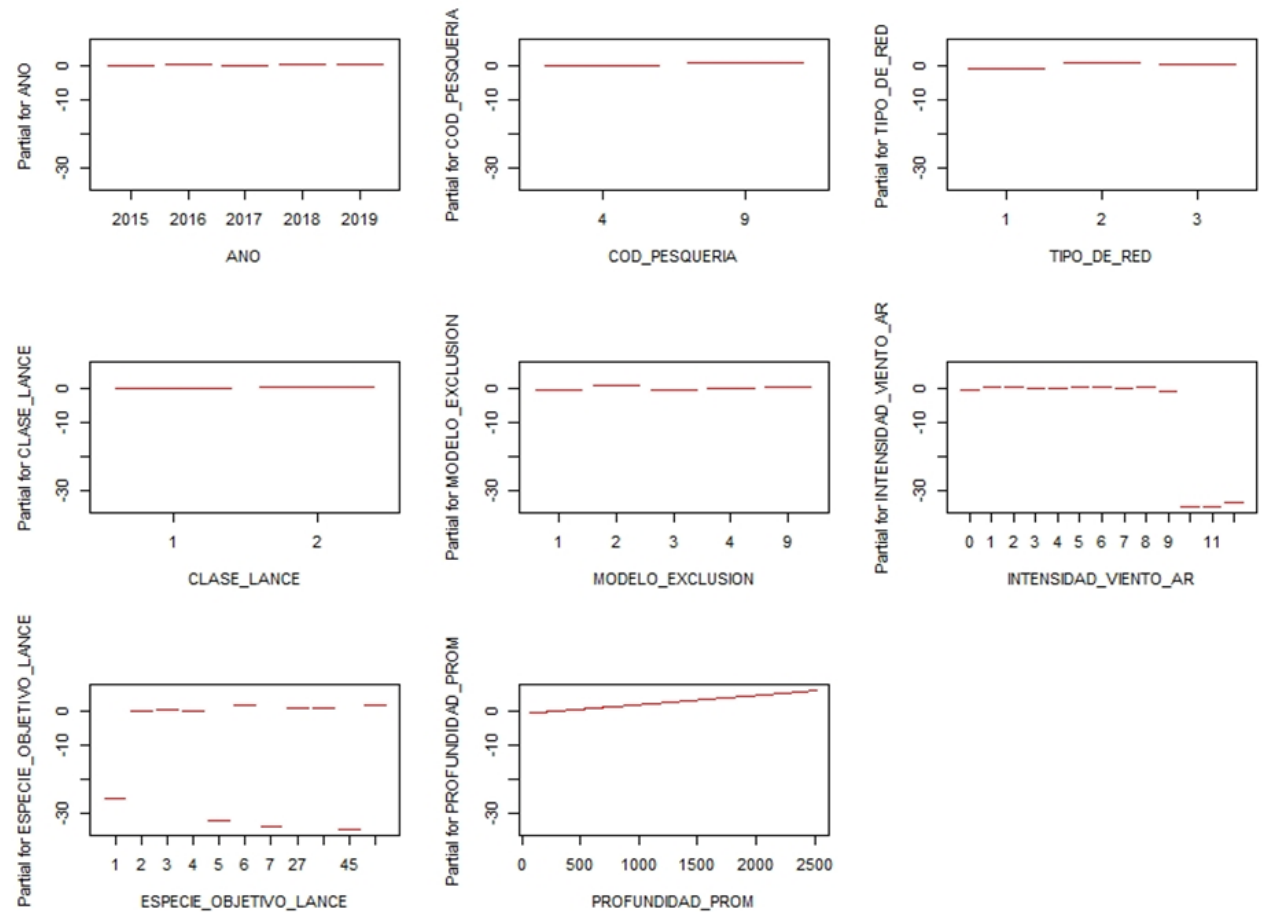
Tabla ANOVA para Predictor 1

	Df	Deviance	Resid. DF	Resid. Dev	Pr(>Chi)	% Deviance
NULL			10554	1773		
ANO	4	26.938	10550	1746	<2.05E-05***	1.52%
COD_PESQUERIA	1	14.379	10549	1731	1.50E-04***	0.81%
TIPO_DE_RED	2	144.878	10547	1587	2.2E-16***	8.17%
CLASE_LANCE	1	0.037	10546	1587	8.47E-01	0.00%
MODELO_EXCLUSION	4	100.507	10542	1486	<2.2E-16***	5.67%
INTENSIDAD_VIENTO_AR	12	18.677	10530	1467	9.66E-02	1.05%
ESPECIE_OBJETIVO_LANCE	10	17.321	10520	1450	6.76E-02	0.98%
PROFUNDIDAD_PROM	1	15.147	10519	1435	9.95E-05***	0.85%

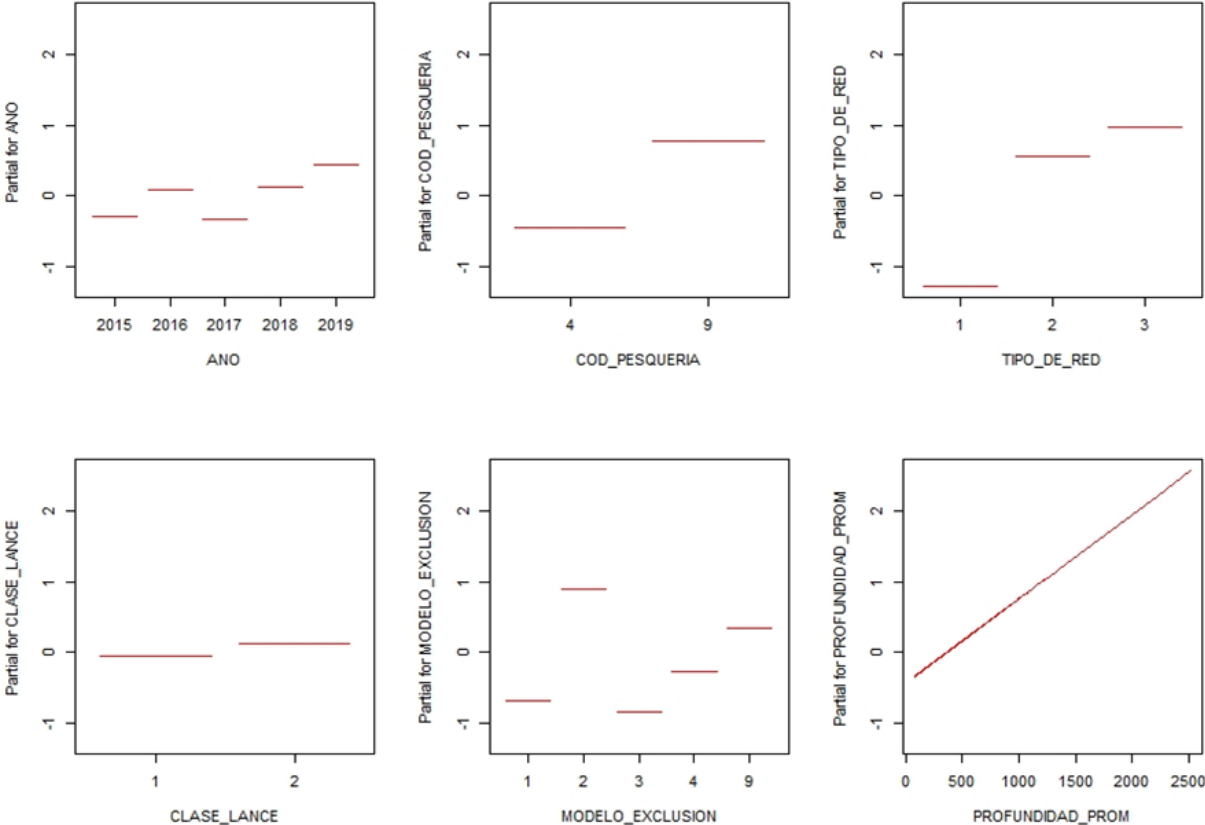
Tabla ANOVA para Predictor 2

	Df	Deviance	Resid. DF	Resid. Dev	Pr(>Chi)	% Devianza
NULL			10554	1705.3		
ANO	4	25.798	10550	1679.5	3.48E-05***	1.51%
COD_PESQUERIA	1	13.881	10549	1665.6	1.95E-04***	0.81%
TIPO_DE_RED	2	139.563	10547	1526	<2.2E-16***	8.18%
CLASE_LANCE	1	0.039	10546	1526	8.44E-01	0.00%
MODELO_EXCLUSION	4	96.861	10542	1429.1	<2.2E-16***	5.68%
PROFUNDIDAD_PROM	1	5.489	10541	1423.6	1.91E-02*	0.32%

Importancias relativas con Predictor 1



Importancias relativas con Predictor 2



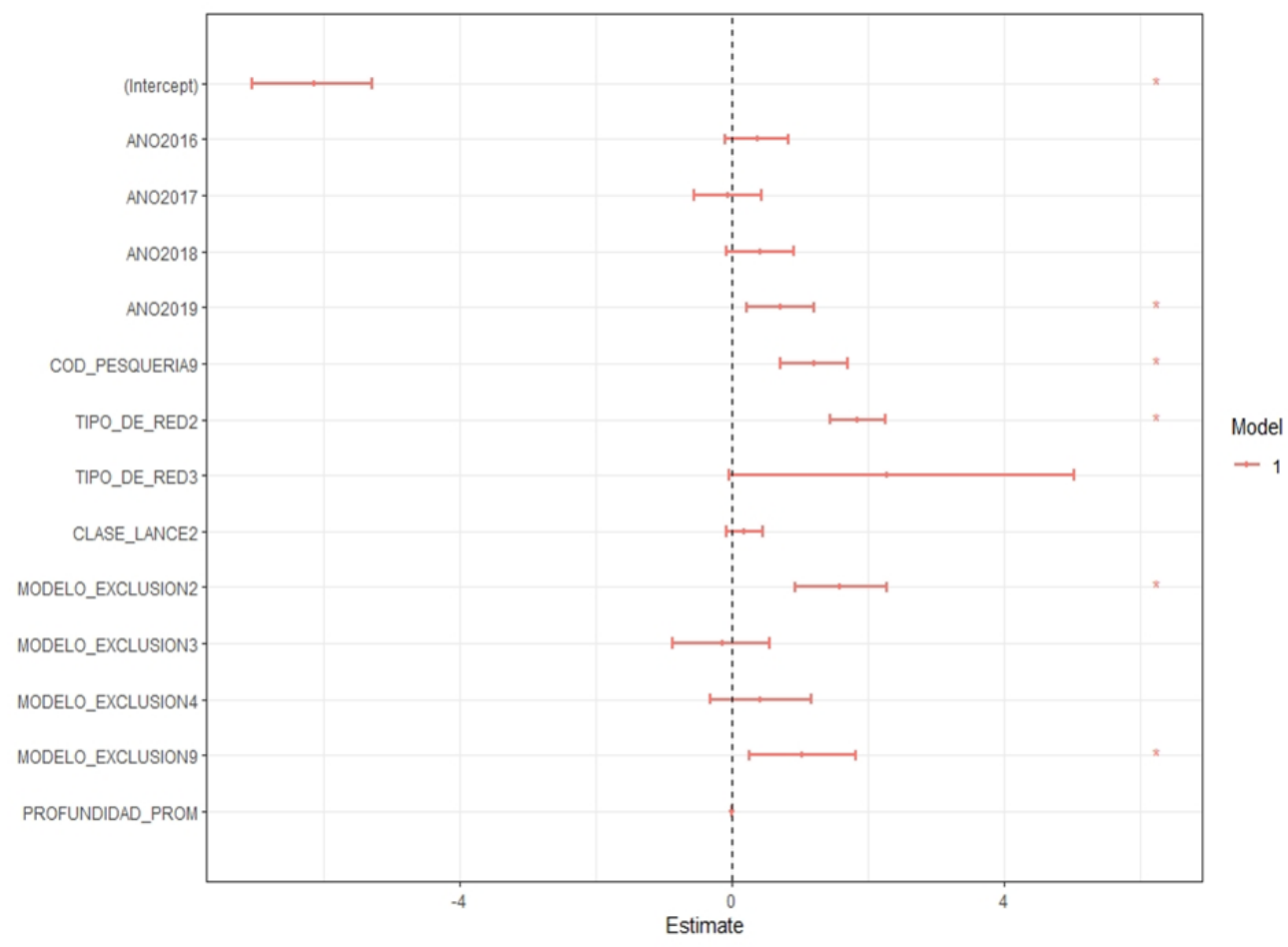
AIC para los modelos propuestos

Modelos	Predictor	AIC
Binomial Negativo	Predictor_1	4689
Binomial Negativo	Predictor_2	4690
Zero-Inflated Binomial Negativo	Predictor_1	4617
Zero-Inflated Binomial Negativo	Predictor_2	4601

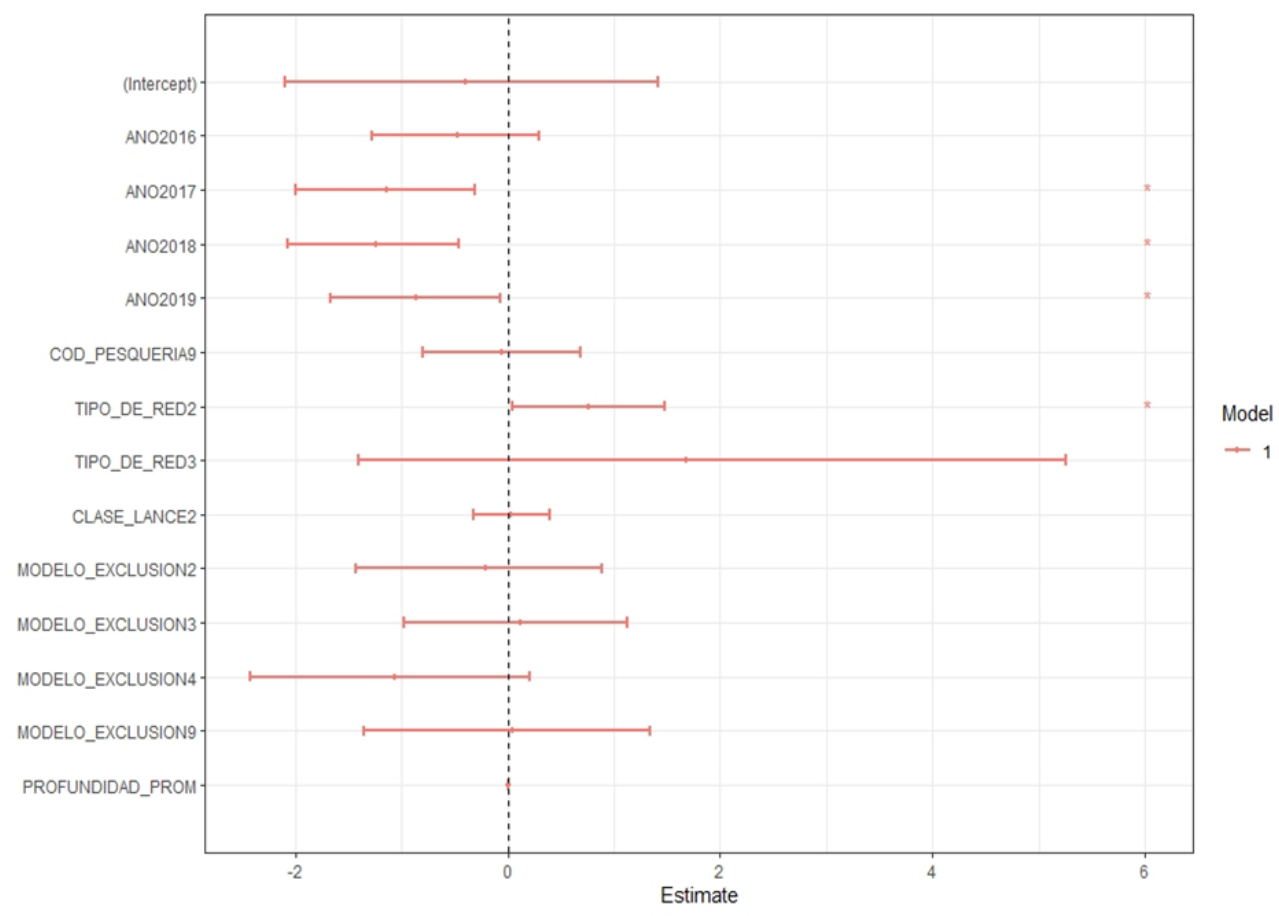
Resultados GLM con INLA

Modelo	Predictor	WAIC	DIC	CPO
Binomial Negativo	Predictor_1	4725.404	4732.945	2364.978
Binomial Negativo	Predictor_2	4693.622	4689.764	2347.627
Zero-Inflated Binomial Negativo	Predictor_1	-Inf	-Inf	2449.437
Zero-Inflated Binomial Negativo	Predictor_2	4906.420	4903.062	2448.986

Variables de importancia modelo Binomial Negativo con Predictor 2



Variables de importancia modelo Zero-Inflated Binomial Negativo con Predictor 2



Resumen

GLM frecuentista

- El análisis de la varianza con 'Predictor_1' y 'Predictor_2' para el modelo Binomial Negativo indica que las variables de importancia relativa son 'TIPO_DE_RED' y 'MODELO_EXCLUSION', mientras que con una importancia relativa menor aparecen 'ANO', 'COD_PESQUERIA', 'PROFUNDIDAD_PROM'.

GLM Bayesiano

- Bajo el enfoque Bayesiano el modelo con mejor desempeño basado en el WAIC fue el modelo Binomial Negativo con 'Predictor_2' y un valor del intercepto estimado de -6,155, mientras que el modelo Zero-Inflated Binomial Negativo con 'Predictor_2' estimó un intercepto de -0.394. El modelo Binomial Negativo con 'Predictor_1' fue el segundo con mejor WAIC pero el valor del intercepto fue bastante sesgado hacia los valores '0' (-27.187).

Conclusiones OE1

- 'Predictor 1' se ajusta bien para un modelo con distribución Binomial Negativa pero produciéndose problemas de estimación en los otros tipos de modelos. Lo anterior se puede deber a la gran cantidad de niveles en las variables 'INTENSIDAD_DEL_VIENTO' y 'ESPECIE_OBJETIVO_LANCE' que son tratados como factores dentro del predictor lineal.
- Las variables relevantes que podrían explicar el número de capturas incidentales de lobos marinos considerando los resultados expuestos anteriormente bajo los dos enfoques parecen coincidir, el factor año (ANO), los tipos de barcos que participan en la pesquería (COD_PESQUERIA), el tipo de red utilizada en los lances de pesca (TIPO_DE_RED), el modelo de exclusión (MODELO_EXCLUSION) y en una menor medida la profundidad promedio de las capturas (PROFUNDIDAD_PROM).

Bibliografía

- Wang, X., Yue, Y. R., & Faraway, J. J. (2018). Bayesian regression modeling with INLA. CRC Press.
- Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology*, 4, 33-49.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4, 395-421.
- Gómez-Rubio, V. (2020). Bayesian inference with INLA. CRC Press.
- Wood, S. N. (2017). Generalized additive models: an introduction with R. CRC press.

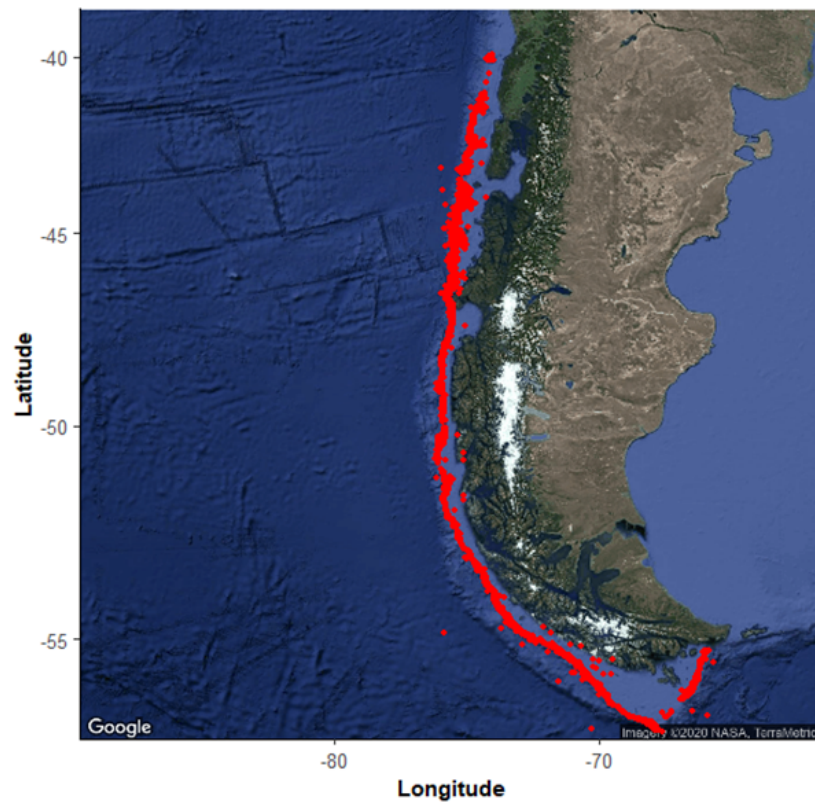
Resultados OE2

Los siguientes resultados están orientados a la presentación de una aproximación Bayesiana para modelar la variable respuesta asociadas a una dependencia espacial y/o temporal a través de INLA.

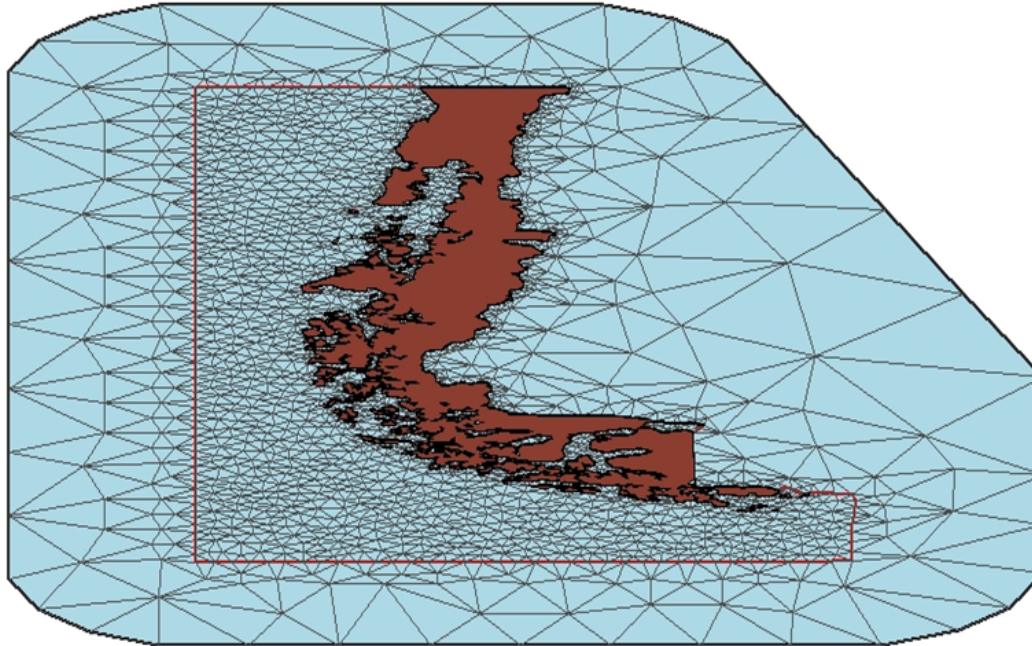
Variables modelo espacio-temporal con INLA

Nombre variable	Codificación
ANO	ano
COD_PESQUERIA	pesq
TIPO_DE_RED	red
PROFUNDIDAD_PROM	prof
MODELO_EXCLUSION	mode

Observaciones referenciadas geográficamente de capturas incidentales desde el año 2015 a 2019



Discretización del dominio espacial mediante método SPDE



Modelos propuestos

	Modelos	Efectos fijos	Efectos aleatorios	Observación
Binomial Negativo	Model_1	ano + pesq + red + prof + mode	-	-
	Model_2	ano + pesq + red + prof + mode	i.z	spde
	Model_3	pesq + red + prof + mode	ano + i.z	iid + spde
	Model_4	pesq + red + prof + mode	ano + i.z	rw1 + spde
	Model_5	pesq + red + prof + mode	ano + mode + i.z	rw1 + iid + spde
	Model_6	pesq + red + prof + mode	ano + mode + i.z	iid + rw1 + spde
Zero- inflated Binomial- negativo	Model_1	ano + pesq + red + prof + mode	-	-
	Model_2	ano + pesq + red + prof + mode	i.z + i.y	spde(2)
	Model_3	pesq + red + prof + mode	ano + i.z + i.y	iid + spde (2)
	Model_4	pesq + red + prof + mode	ano + i.z + i.y	rw1 + spde (2)
	Model_5	pesq + red + prof + mode	ano + mode + i.z + i.y	rw1 + iid + spde (2)
	Model_6	pesq + red + prof + mode	ano + mode + i.z + i.y	iid + rw1 + spde (2)

Resultados

	Modelos	WAIC	CPO
Binomial Negativo	Model_1	4693.60	2347.42
	Model_2	4600.12	2303.02
	Model_3	4609.67	2306.68
	Model_4	4609.68	2306.69
	Model_5	4608.41	2306.09
	Model_6	4608.70	2306.40
Zero-inflated Binomial-negativo	Model_1	4685.26	2342.73
	Model_2	4593.69	2297.29
	Model_3	4594.75	2297.79
	Model_4	4602.26	2301.52
	Model_5	4601.63	2301.20
	Model_6	4594.92	2297.87

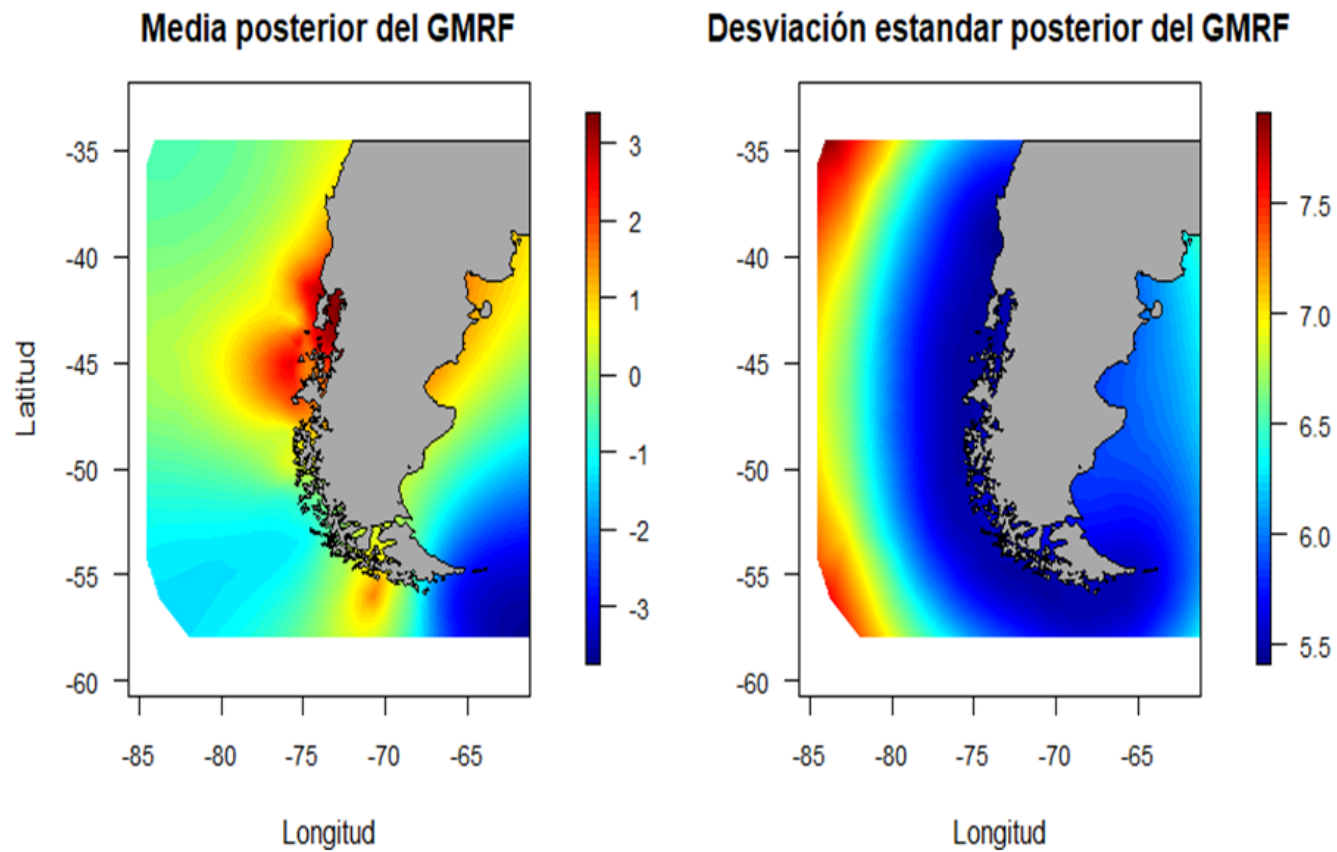
Ya que no se pudo contar con toda la base de datos completa en un inicio se evaluaron 3 modelos adicionales contemplando integrar las siguientes variables a 'Modelo_base':

1. 'día' que corresponde al día de la observación y modelada como una función de suavizamiento en escala temporal.
2. 'trim' que corresponde a un factor en el cual se pueda incorporar alguna medida las temporadas de pesca dentro de un año.
3. 'espe' la cual considera a la especie objetivo del lance observado.
4. 'dist' que es la distancia a la lobera más cercana a la observación espacial.

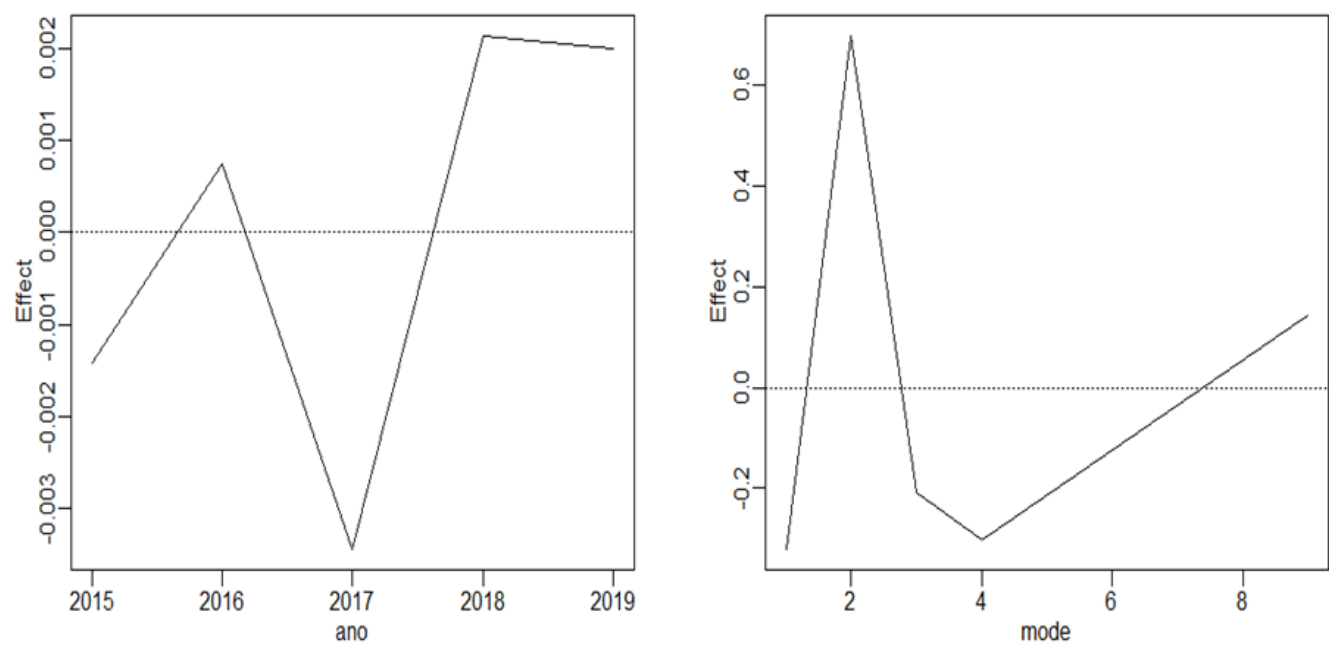
Resultados complementarios

	Modelos	WAIC	CPO
Zero-inflated Binomial- negativo	M1 = Modelo_base	9093.992	4547.096
	M2 = Modelo_base + trim	8833.776	4416.975
	M3 = Modelo_base + dia	9081.392	4540.843
	M4 = Modelo_base + espe	9091.948	4546.072
	M5 = Modelo_base + dist	9105.165	4563.930

Media y desviación estándar del GMRF (efecto aleatorio espacial)



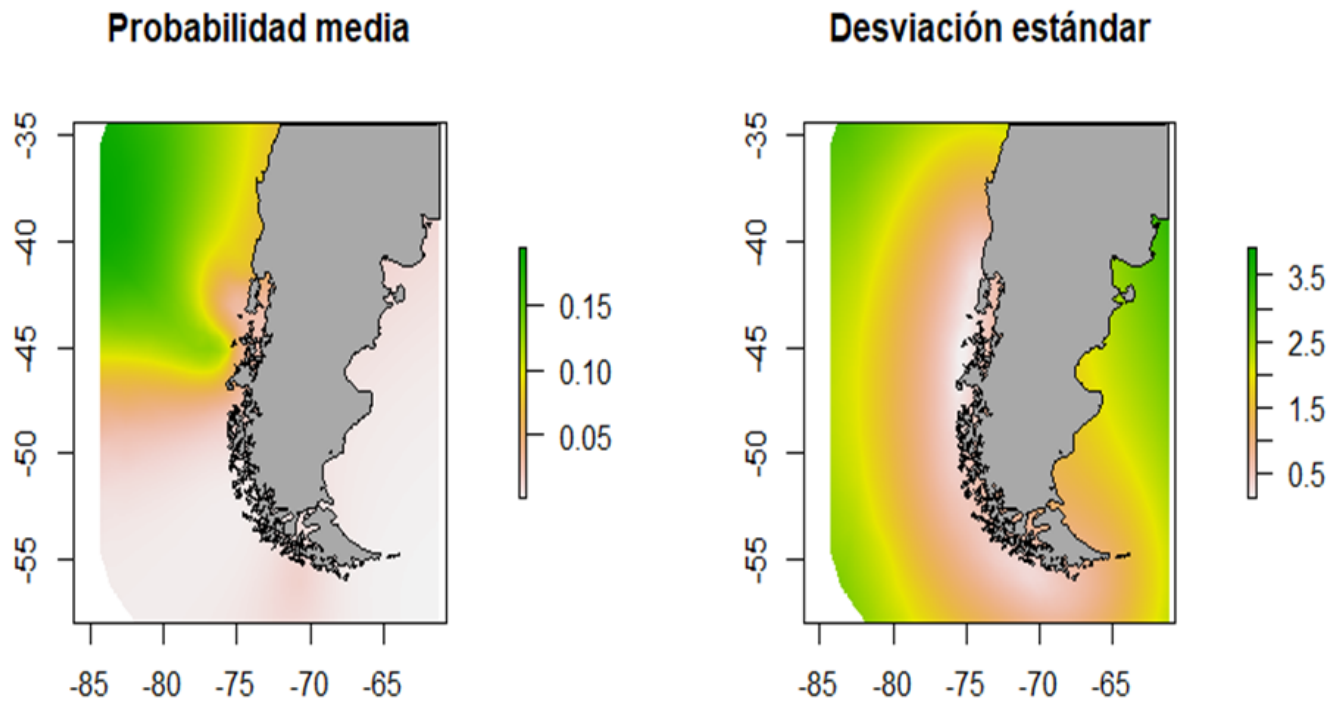
Importancias relativas de cada efecto aleatorio sobre el número de capturas incidentales de lobos marinos



Predicciones medias, inferiores y superiores de capturas incidentales modelo espacio-temporal.

	Mínimo	Mediana	Media	Máximo
media_pred	1.00	1.03	1.04	1.49
inf_media_pred	1.00	1.02	1.03	1.24
sup_media_pred	1.00	1.04	1.06	1.96

Probabilidad de ocurrencia asociada a las capturas incidentales de lobos marinos.



Resumen

- La mediana del modelo se encuentra en 1.03 lobos capturado por lance, mientras que la media es de 1.04 lobos capturado por lance, con un máximo estimado de 1.49 lobos por lance. La estimación está basada en las posteriores predichas para la media (`media_pred`), el intervalo de credibilidad del 0.25 (`inf_media_pred`) y el intervalo de credibilidad del 0.975 (`sup_media_pred`).
- El valor de probabilidad media más alto se concentra frente a la zona de la Isla de Chiloé y frente a la región de Aysén. Dentro de este dominio espacial es donde se puede identificar la mayor probabilidad de captura incidental de los lobos marinos. Los números indican que, la probabilidad media de que haya un lance con captura incidental positivo (>0), es de 0.01.

Conclusiones OE2

- En primera instancia se modelaron 12 distintos modelos con distintas combinaciones de predictores en base a una distribución Zero-Inflated Poisson y una distribución Zero-Inflated Binomial-Negativo. La comparación mediante el estadístico 'WAIC' y 'CPO' permitió identificar que el modelo con mejor desempeño estadístico fue el Zero-Inflated Binomial-Negativo.
- Considerando lo anterior se integraron nuevas variables que no se encontraron disponibles inicialmente, en donde la variable trimestre ('trim') tuvo una gran significancia ya que permitió disminuir considerablemente el 'WAIC' del modelo elegido previamente.
- El modelo que mejor explica las capturas incidentales de lobos marinos dentro de un modelo espacio-temporal incluye a las variables 'pesq' (Pesquería), 'red' (tipo de red) y 'trim' (trimestre) como efectos fijos. Además, las variables 'ano' (año) y 'mode' (modelo de exclusión), son incluidas como efectos aleatorios junto a los efectos espaciales asociados al proceso de conteo y de Bernoulli.

Bibliografía

- Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D. & Rue, H. (2018). Advanced spatial modeling with stochastic partial differential equations using R and INLA. CRC Press.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1-25.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J. & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), e1443.

Gracias