Supplementary materials

Table of Contents

1. At-sea cetacean sightings records

Table A1. Summary of sightings data contained in the various databases. The oldest and most recent records for each database are shown in columns ‘Start’ ‘End’ respectively. The number of sightings in each database is shown in the final column.

Database	Source	Start	End	Number
Cawthorn	Martin Cawthorn (independent consultant)	January 1980	November 1999	1 150
COD	Centralised Observer Database (COD): NIWA maintained database on behalf of MPI	January 2009	April 2017	4 768
NIWA	National Institute of Water and Atmospheric Research (NIWA)	April 2011	July 2016	6
OMV	OMV limited	March 2005	April 2015	416
DOC	Department of Conservation (DOC)	January 1970	July 2017	7 867

The Cawthorn database contains opportunistic sightings collected by trained ship officers that were on standard routes around NZ, across the Pacific or Tasman. Cetacean sightings were quality controlled by matching species with the fin/fluke/body shape indicators circled on a sightings form, as well as any other notes provided about behaviour, colours, and markings. If there was doubt about the correct identification the sighting was eliminated. The COD cetacean database contains sightings reported by trained fisheries observers aboard inshore fishing vessels between 2009 and 2017. The NIWA database primarily contained opportunistic sightings recoded by scientists conducting research onboard one of NIWA’s vessels. The OMV cetacean database contained sightings collected by trained observers onboard various platforms associated with oil and gas activities. Finally, the database maintained by DOC contains data from several sources (including NIWA) but also opportunistic sighting reported by the public from the following platforms: fixed wing plane, helicopter, land, and private, research or non-motorised coastal vessel. Note that small sample size for several of the databases is the result of eliminating records that were duplicated in the DOC database. Locations of at-sea cetacean sightings for each cetacean taxa are shown in supplementary materials 4.

2. Variable correlations and distributions

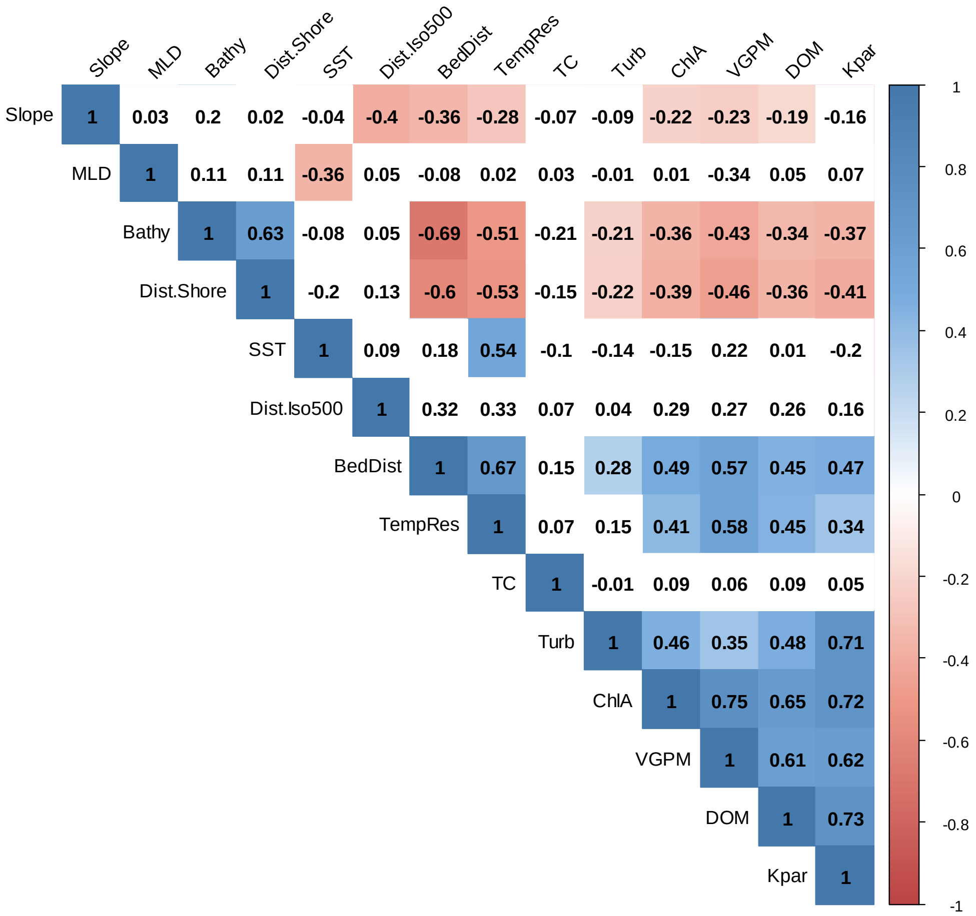

Fig A 2-1. Correlation matrix of environmental variables (Pearson correlation). Non-significant correlations are shown in white.

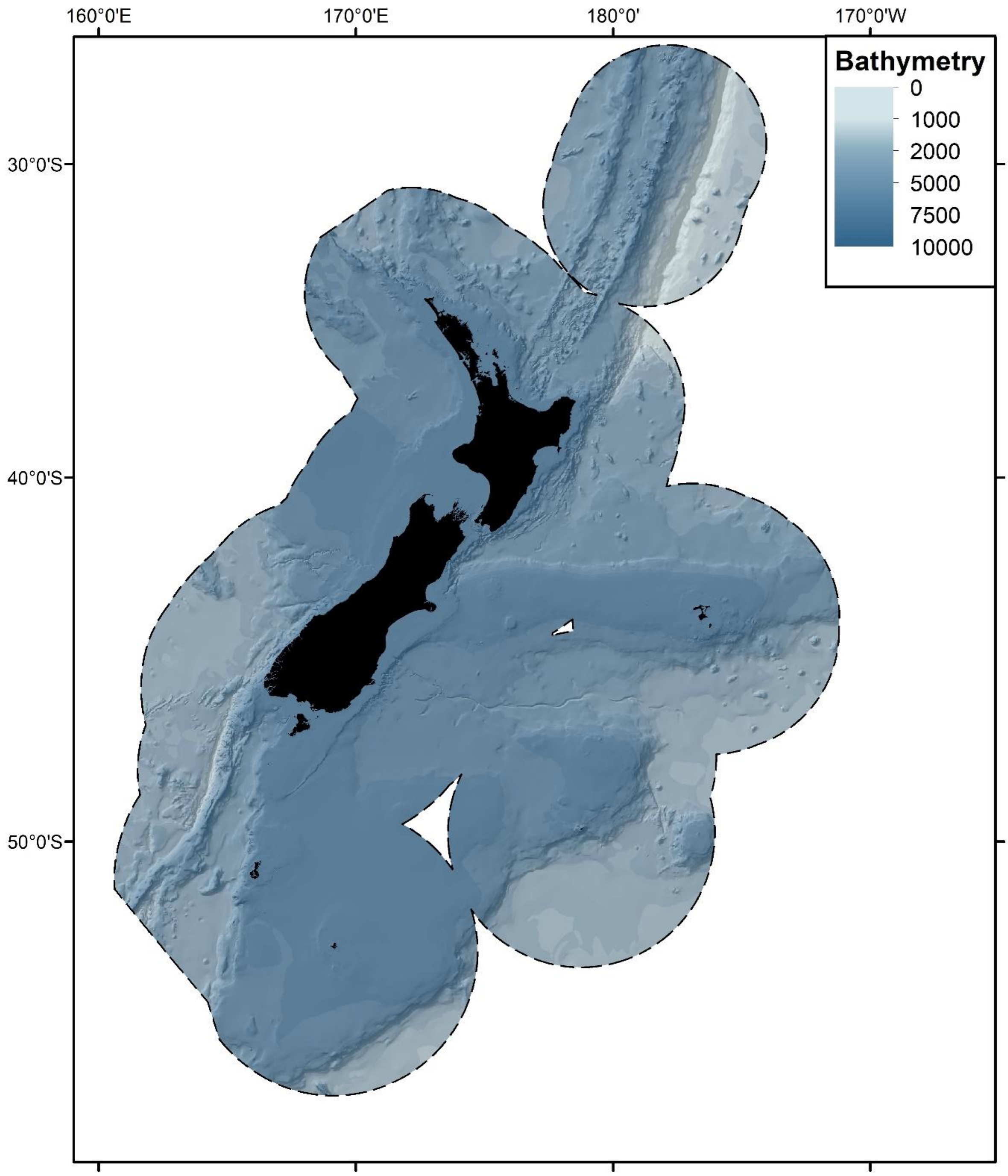

Fig A 2-2. Bathymetry (*Bathy*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

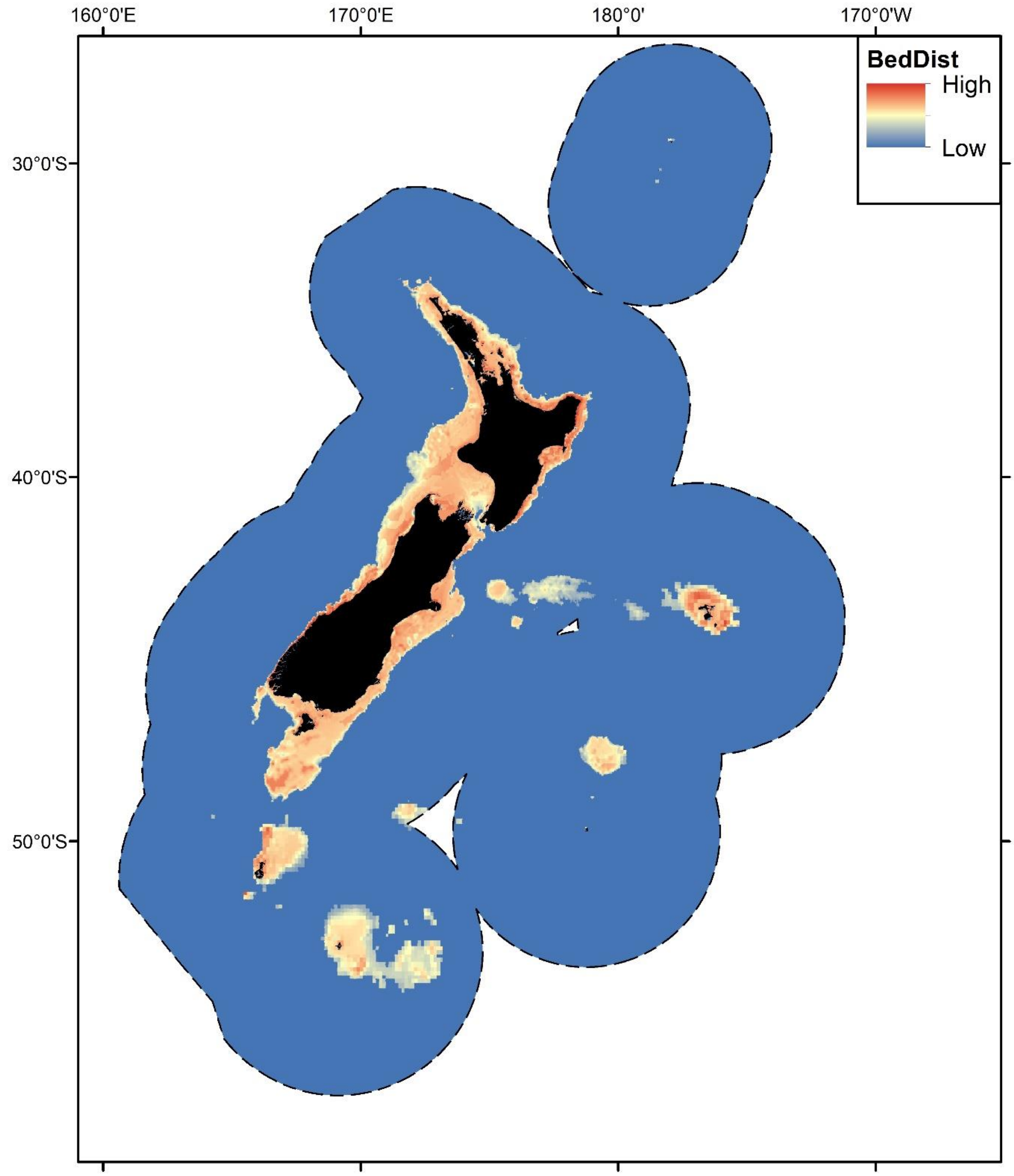

Fig A 2-3. Benthic sediment disturbance (*BedDist*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

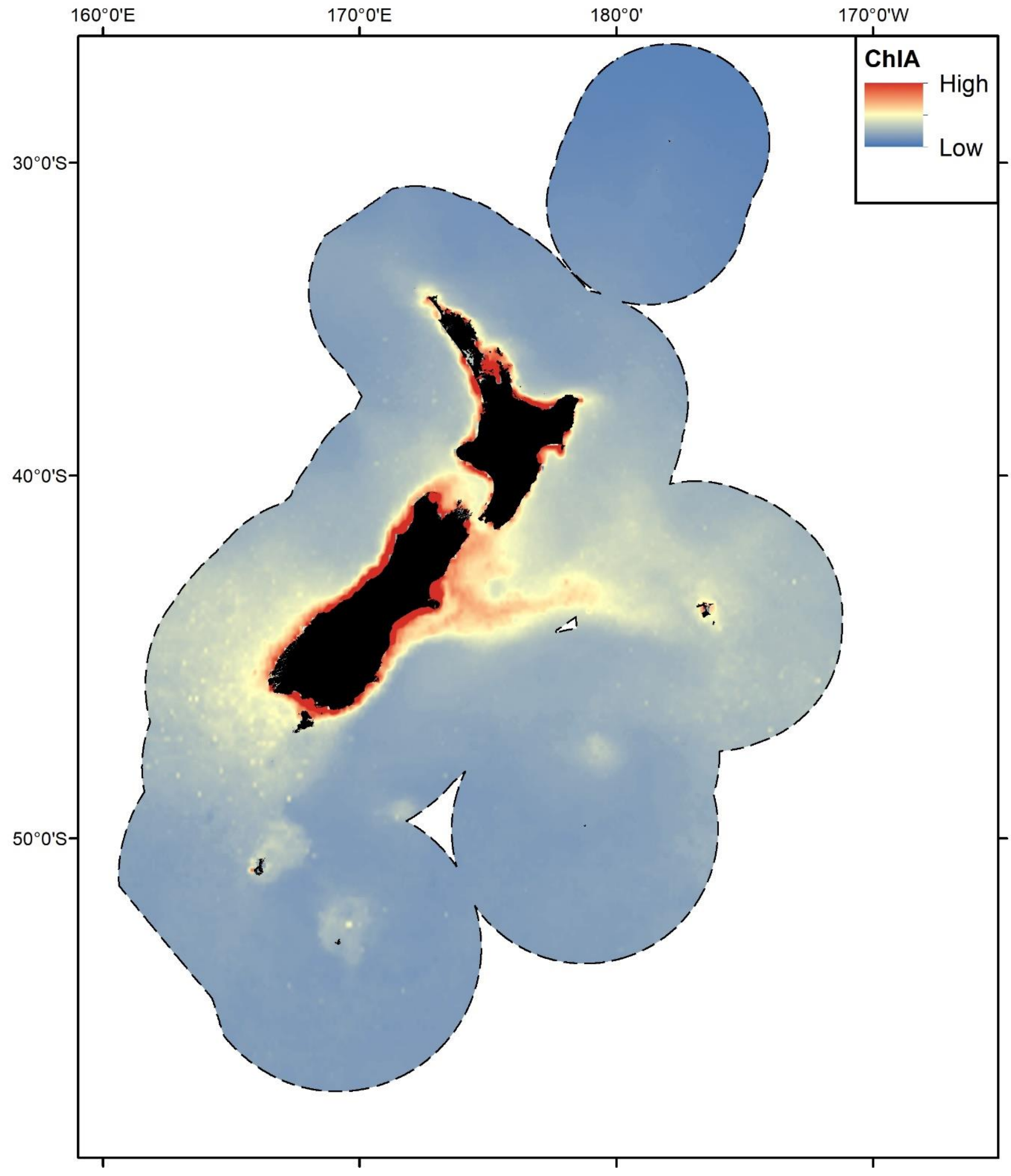

Fig A 2-4. Annual mean Chlorophyll-a concentration (*ChlA*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

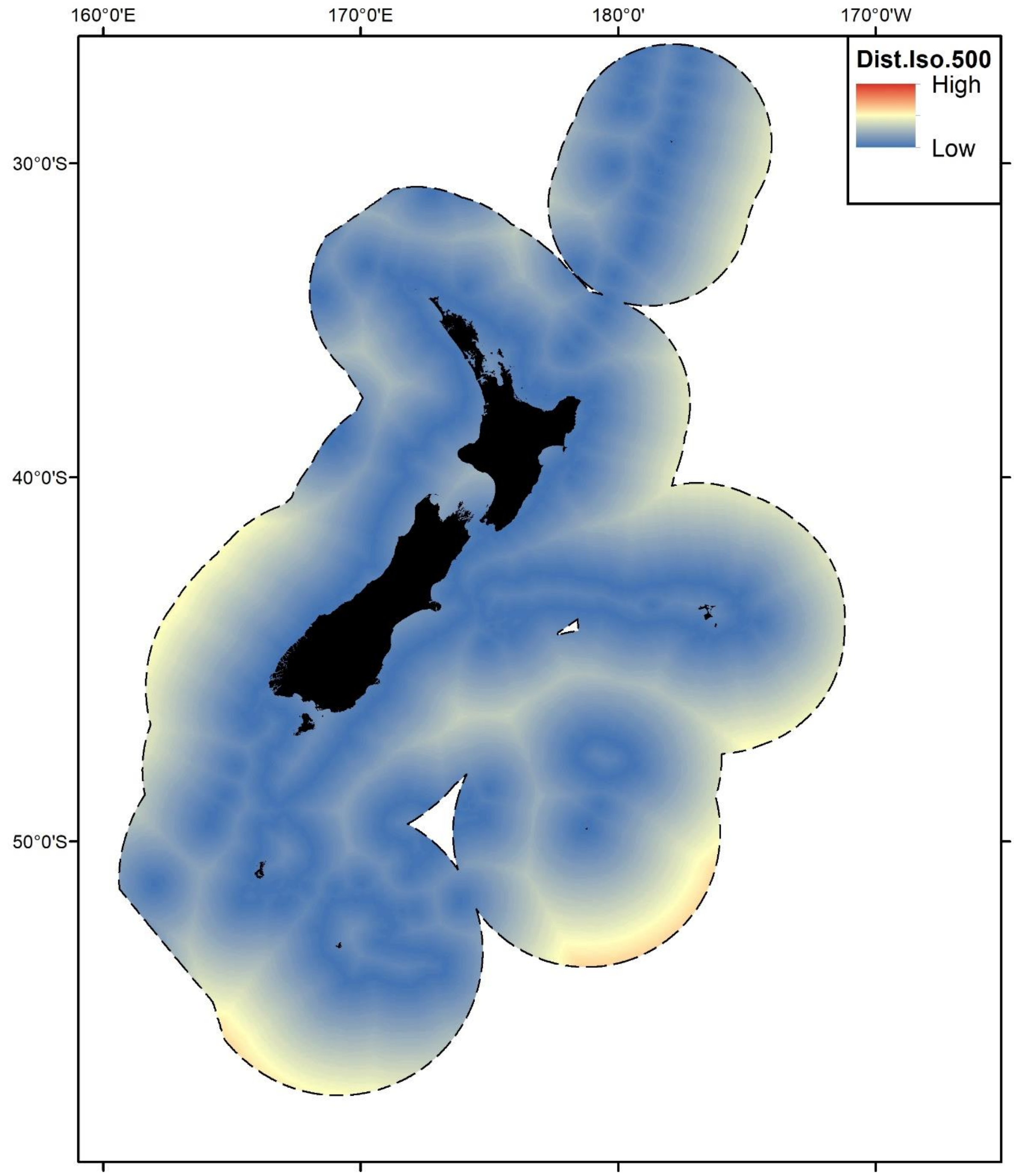

Fig A 2-5. Distance to 500 m isobath (*Dist.Iso.500*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

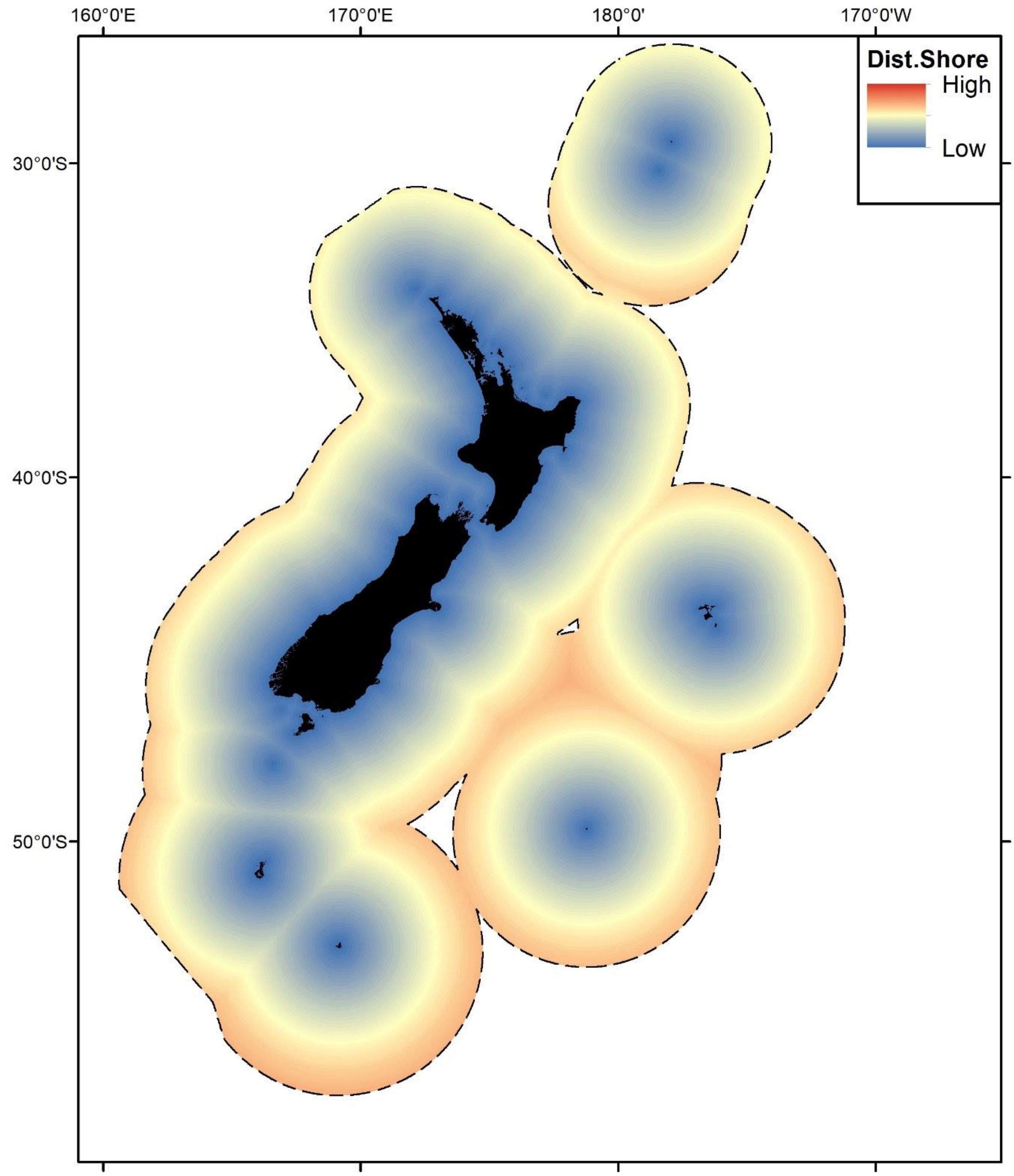

Fig A 2-6. Distance to shore (*Dist.Shore*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

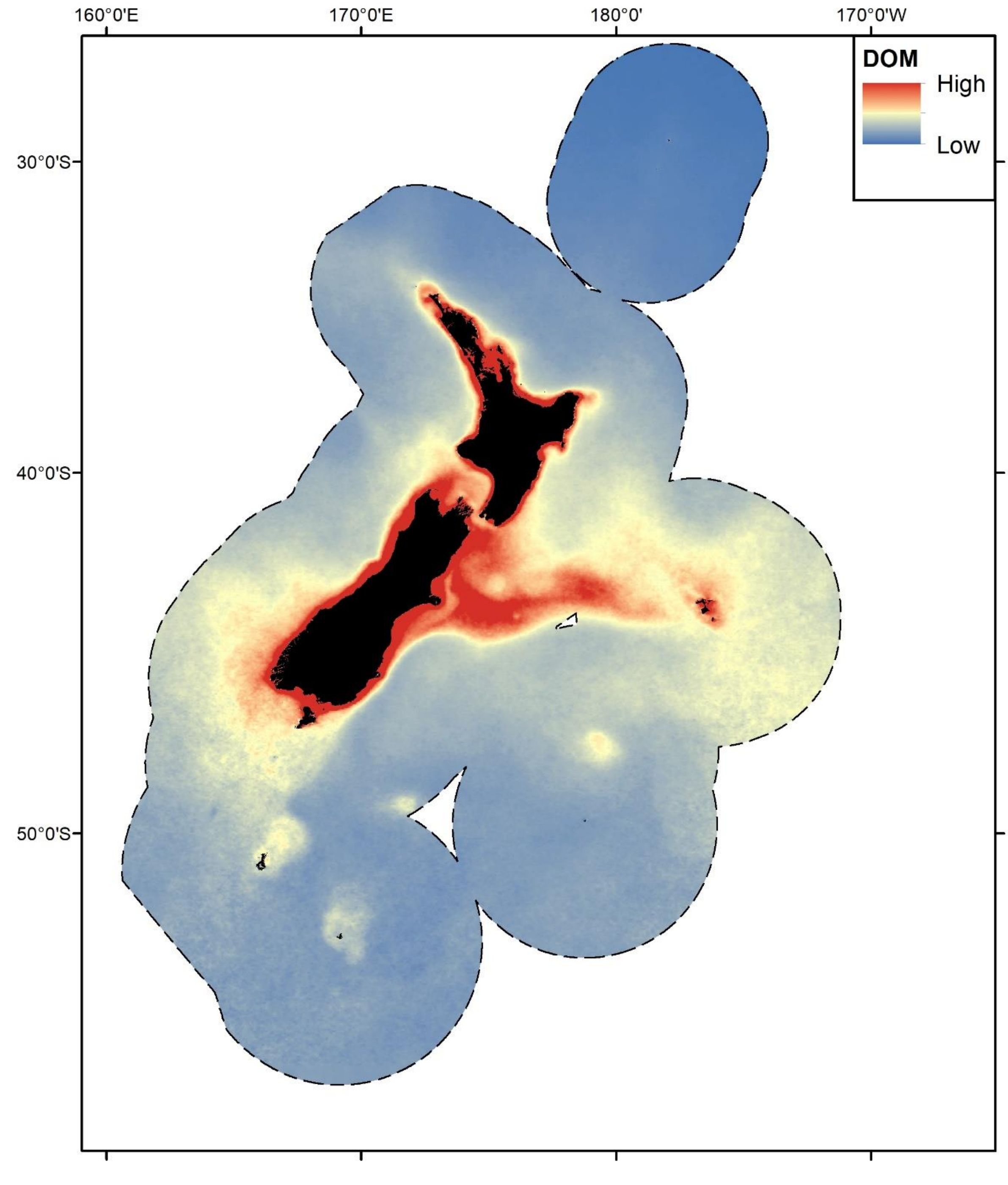

Fig A 2-7. Coloured dissolved organic matter (*DOM*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

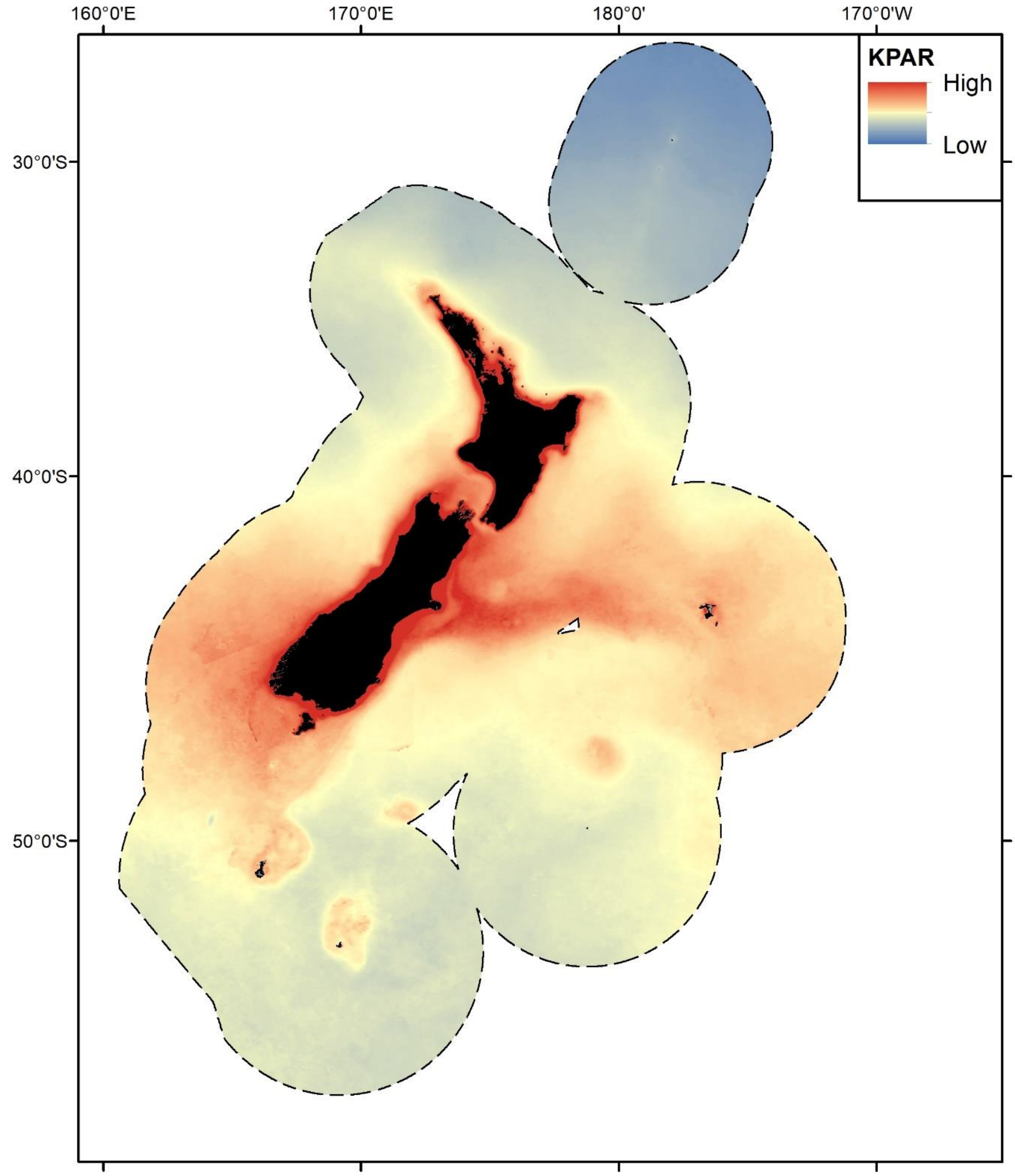

Fig A 2-8. Annual mean KPAR (*Kpar*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

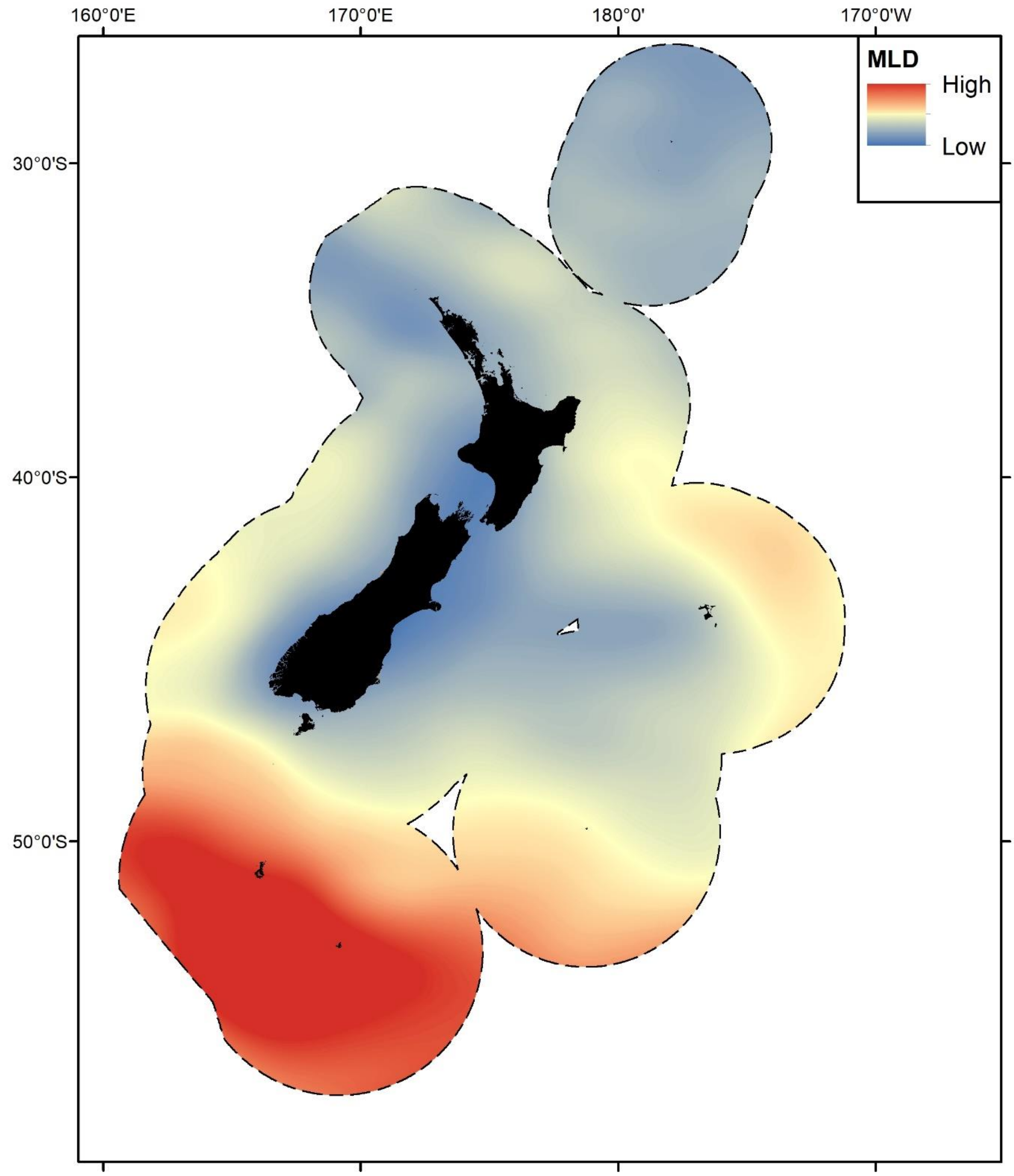

Fig A 2-9. Annual mean Mixed layer depth (*MLD*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

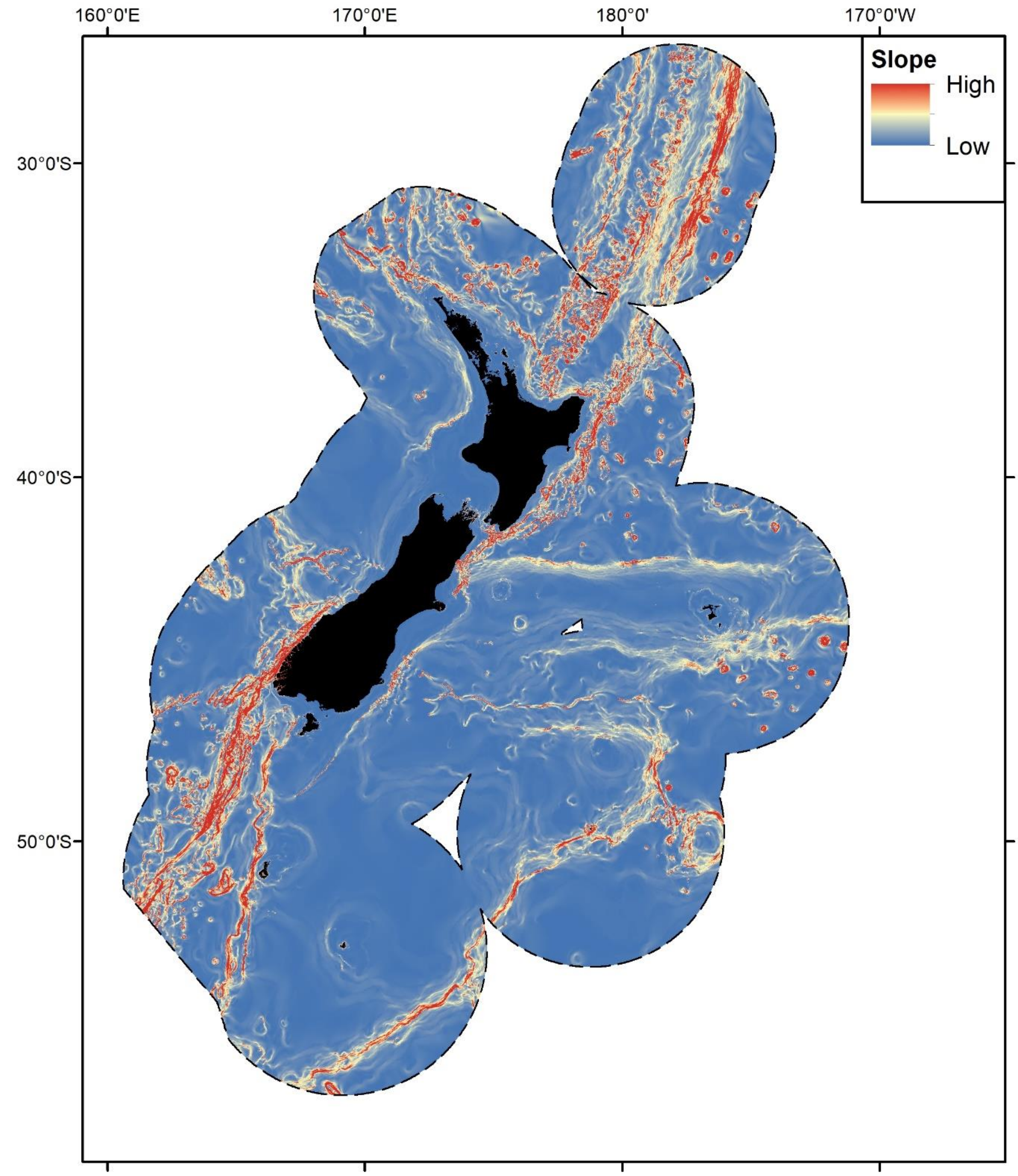

Fig A 2-10. Slope (*Slope*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

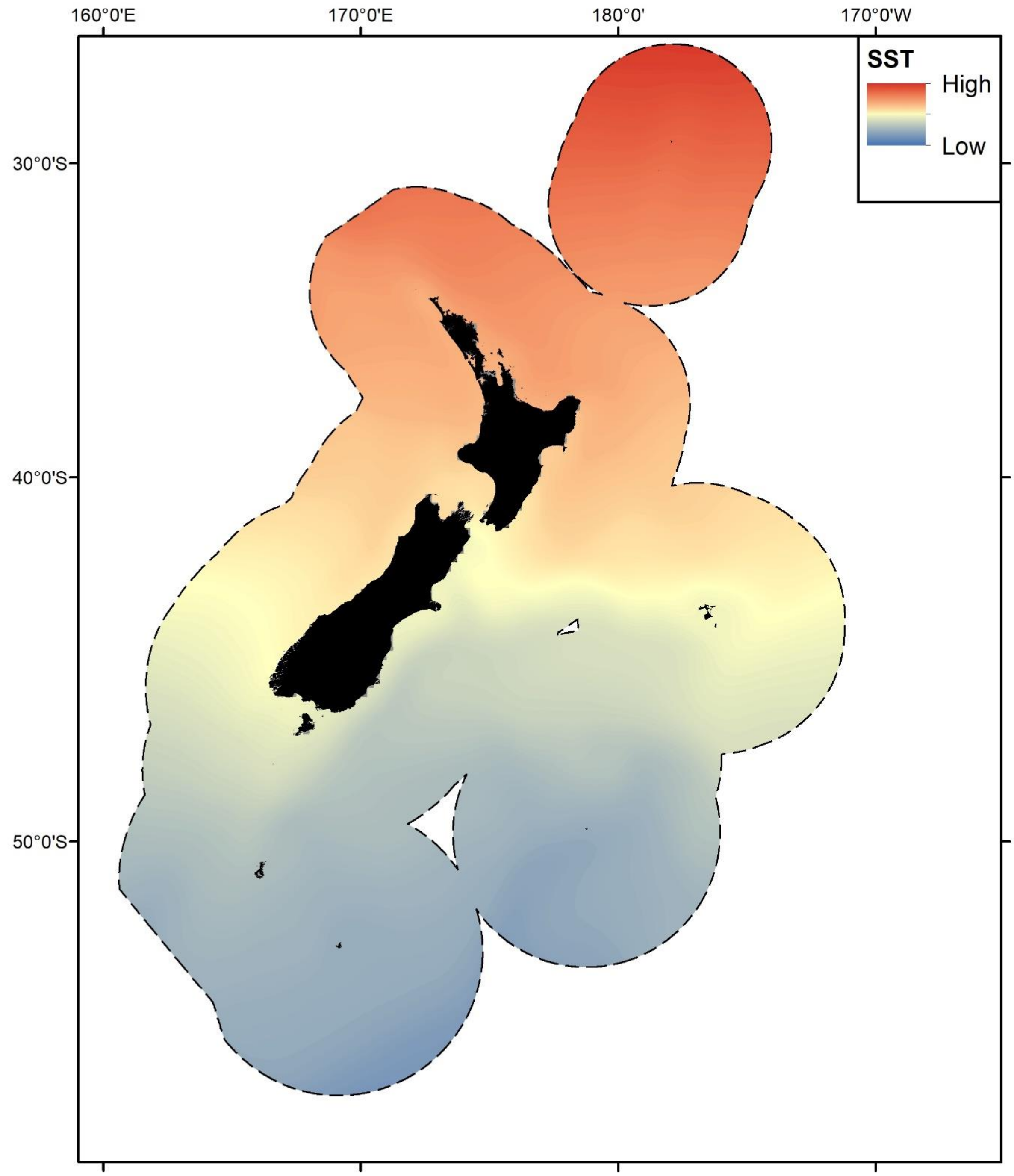

Fig A 2-11. Annual mean Sea surface temperature (*SST*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

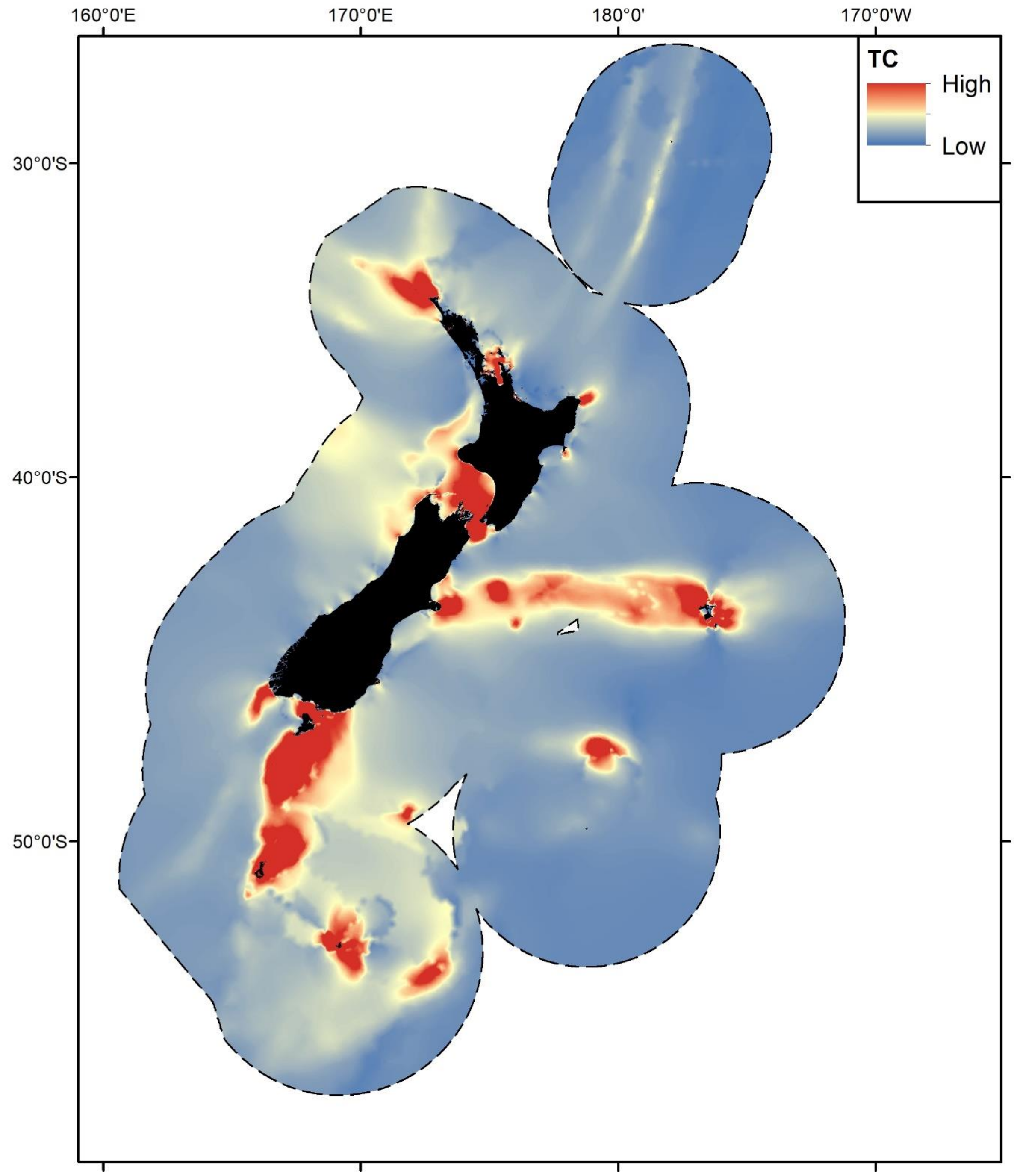

Fig A 2-12. Tidal Current speed (*TC*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

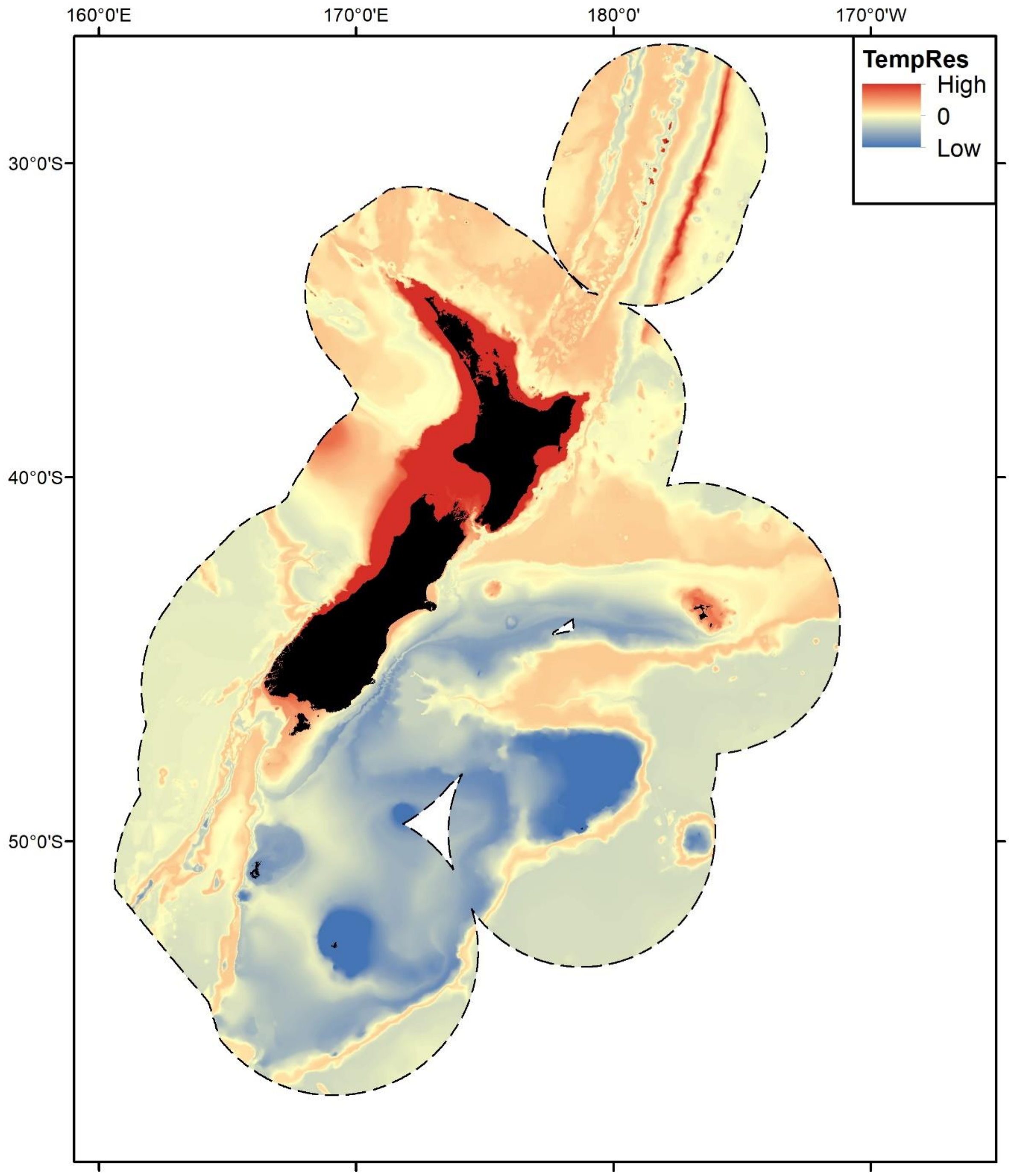

Fig A 2-13. Temperature residuals (*TempRes*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

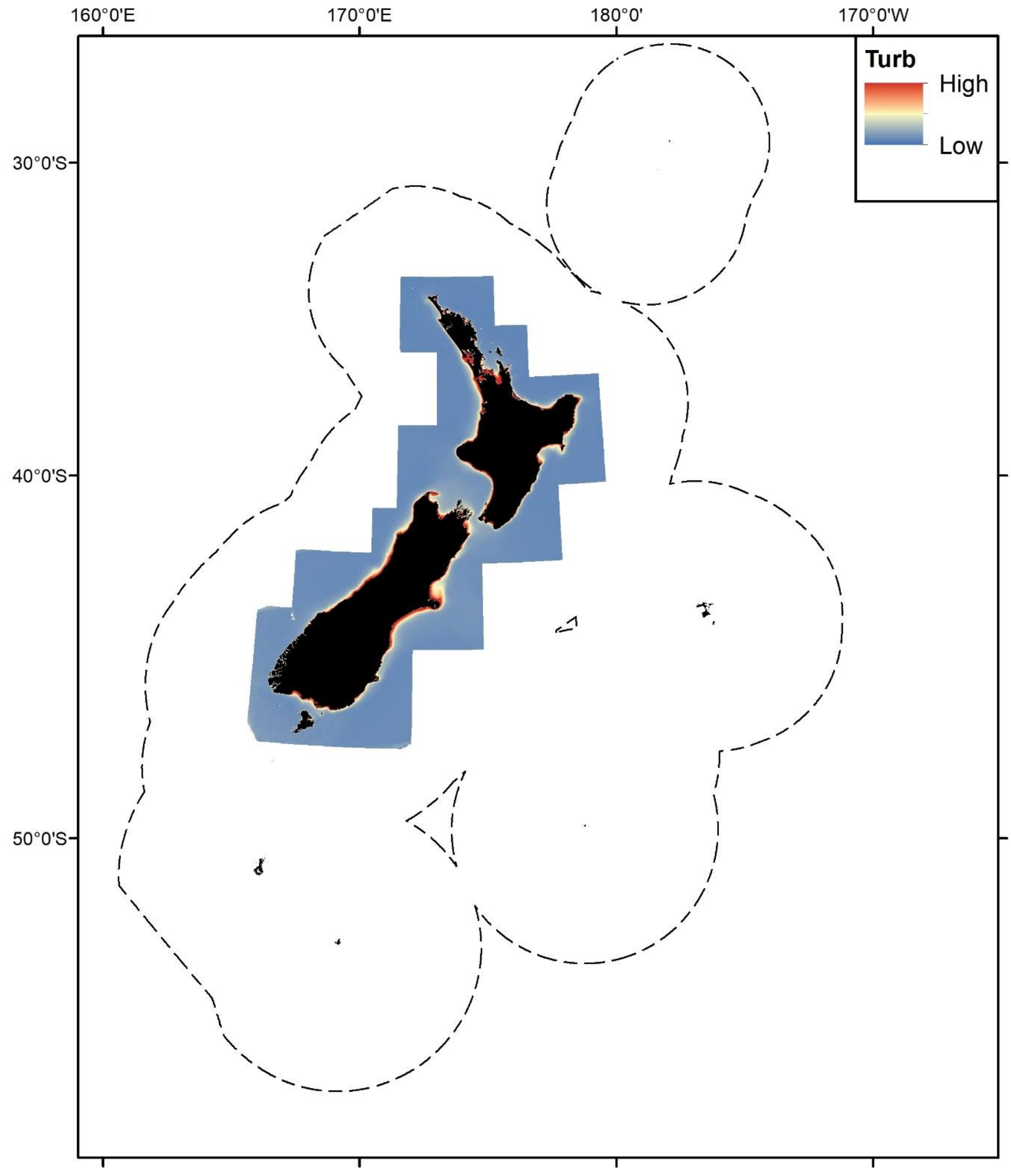

Fig A 2-14. Annual mean Turbidity (*Turb*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

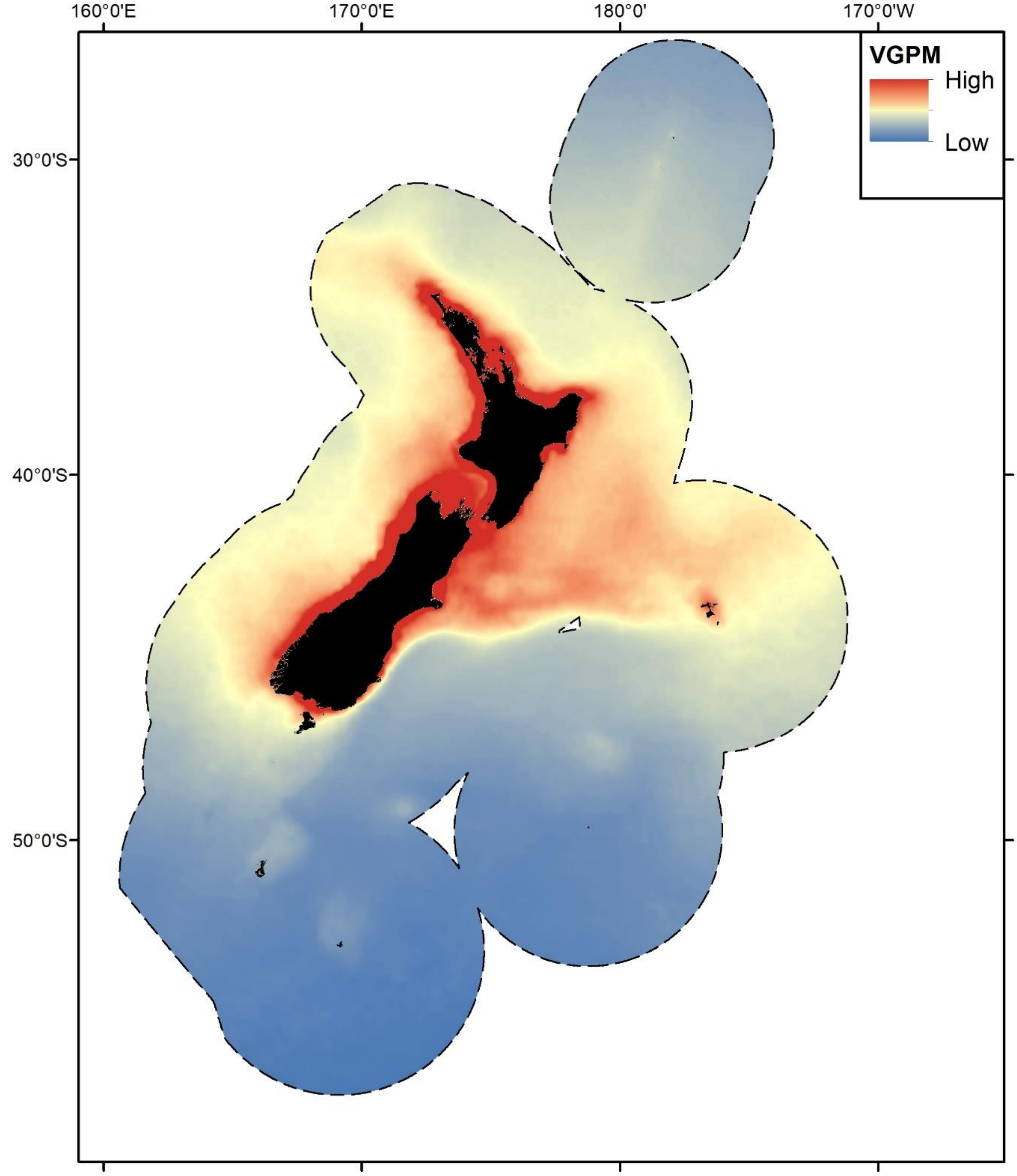

Fig A 2-15. Annual mean Productivity Model (*VGPM*) within the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line).

3. Relative environmental suitability parameters

Table A 3-1. RES values for variable Sea Surface Temperature for cetacean species with low observation records

Common name	Species name	**Sea Surface Temperature (degrees C)**				
		Min	Preferred min	Preferred max	Max	Description
Andrew's beaked whale	*Mesoplodon bowdoini*	5	10	20	25	Cold temperate - subtropical
Arnoux's beaked whale	*Berardius arnuxii*	-2	0	20	25	Polar-subtropical
Blainville's beaked whale	*Mesoplodon densirostris*	10	`15	30	35	Warm temperate - full tropical
Cuvier's beaked whale	*Ziphius cavirostris*	5	10	30	35	Cold temperate - full tropical
Dwarf minke whale	*Balaenoptera acutorostrata*	-2	0	25	30	Polar - tropical
False killer whale	*Pseudorca crassidens*	10	15	30	35	Warm temperate - full tropical
Gray's beaked whale	*Mesoplodon grayi*	0	5	20	25	Subpolar - subtropical
Hourglass dolphin	*Lagenorhynchus cruciger*	-2	0	15	20	Polar-warm temperate
Pygmy sperm whale	*Kogia breviceps*	10	15	30	35	Warm temperate - full tropical
Risso's dolphin	*Grampus griseus*	5	10	30	35	Cold temperate - full tropical
Shepherd's beaked whale	*Tasmacetus shepherdi*	0	5	15	20	Subpolar - warm temperate
Short finned pilot whale	*Globicephala macrorhynchus*	15	20	25	30	Subtropical - tropical

Short-beaked common dolphin	*Delphinus delphis*	5	10	30	35	Cold temperate - full tropical
Southern bottlenose whale	*Hyperoodon planifrons*	-2	0	20	25	Polar - subtropical
Southern right whale dolphin	*Lissodelphis peronii*	-2	0	20	25	Polar - subtropical
Spectacled porpoise	*Phocoena dioptrica*	-2	0	10	15	Polar - cold temperate
Striped dolphin	*Stenella coeruleoalba*	5	10	25	30	Cold temperate - tropical

Table A 3-2. RES values for variable Bathymetry for cetacean species with low observation records

Common name	Species name	**Bathymetry (m)**				
		Min	Preferred min	Preferred max	Max	Description
Andrew's beaked whale	*Mesoplodon bowdoini*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Arnoux's beaked whale	*Berardius arnuxii*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Blainville's beaked whale	*Mesoplodon densirostris*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Cuvier's beaked whale	*Ziphius cavirostris*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Dwarf minke whale	*Balaenoptera acutorostrata*	0	-10	-2000	-8000	Mainly coast. Cont. slope to v deep water
False killer whale	*Pseudorca crassidens*	0	-1000	-2000	-8000	Mainly low. Cont. slope to v deep water
Gray's beaked whale	*Mesoplodon grayi*	0	-200	-2000	-8000	Mainly cont. slope to very deep water

Hourglass dolphin	*Lagenorhynchus cruciger*	0	-1000	-4000	-8000	Mainly low cont. slope - abyss plains to v deep water
Pygmy sperm whale	*Kogia breviceps*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Risso's dolphin	*Grampus griseus*	0	-200	-1000	-6000	Mainly up cont. slope to deep water
Shepherd's beaked whale	*Tasmacetus shepherdi*	0	-1000	-2000	-8000	Mainly low. Cont. slope to v deep water
Short finned pilot whale	*Globicephala macrorhynchus*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Short-beaked common dolphin	*Delphinus delphis*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Southern bottlenose whale	*Hyperoodon planifrons*	0	-1000	-4000	-8000	Mainly low cont. slope - abyss plains to v deep water
Southern right whale dolphin	*Lissodelphis peronii*	0	-200	-2000	-8000	Mainly cont. slope to very deep water
Spectacled porpoise	*Phocoena dioptrica*	0	-10	-2000	-8000	Mainly coast. Cont slope to v deep water
Striped dolphin	*Stenella coeruleoalba*	0	-200	-2000	-8000	Mainly cont. slope to very deep water

Table A 3-3. RES values for variable Distance to Shore for cetacean species with low observation records

Common name	Species name	**Distance to Shore (km)**				
		Min	Preferred min	Preferred max	Max	Description
Andrew's beaked whale	*Mesoplodon bowdoini*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Arnoux's beaked whale	*Berardius arnuxii*	10	100	1000	4000	Oceanic - continental shelf/pelagic species

Blainville's beaked whale	*Mesoplodon densirostris*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Cuvier's beaked whale	*Ziphius cavirostris*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Dwarf minke whale	*Balaenoptera acutorostrata*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
False killer whale	*Pseudorca crassidens*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Gray's beaked whale	*Mesoplodon grayi*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Hourglass dolphin	*Lagenorhynchus cruciger*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Pygmy sperm whale	*Kogia breviceps*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Risso's dolphin	*Grampus griseus*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Shepherd's beaked whale	*Tasmacetus shepherdi*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Short finned pilot whale	*Globicephala macrorhynchus*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Short-beaked common dolphin	*Delphinus delphis*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Southern bottlenose whale	*Hyperoodon planifrons*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Southern right whale dolphin	*Lissodelphis peronii*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Spectacled porpoise	*Phocoena dioptrica*	10	100	1000	4000	Oceanic - continental shelf/pelagic species
Striped dolphin	*Stenella coeruleoalba*	10	100	1000	4000	Oceanic - continental shelf/pelagic species

4. At-sea cetacean sightings records, RES and BRT model outputs

Blainville's beaked whale (*Mesoplodon densirostris*)

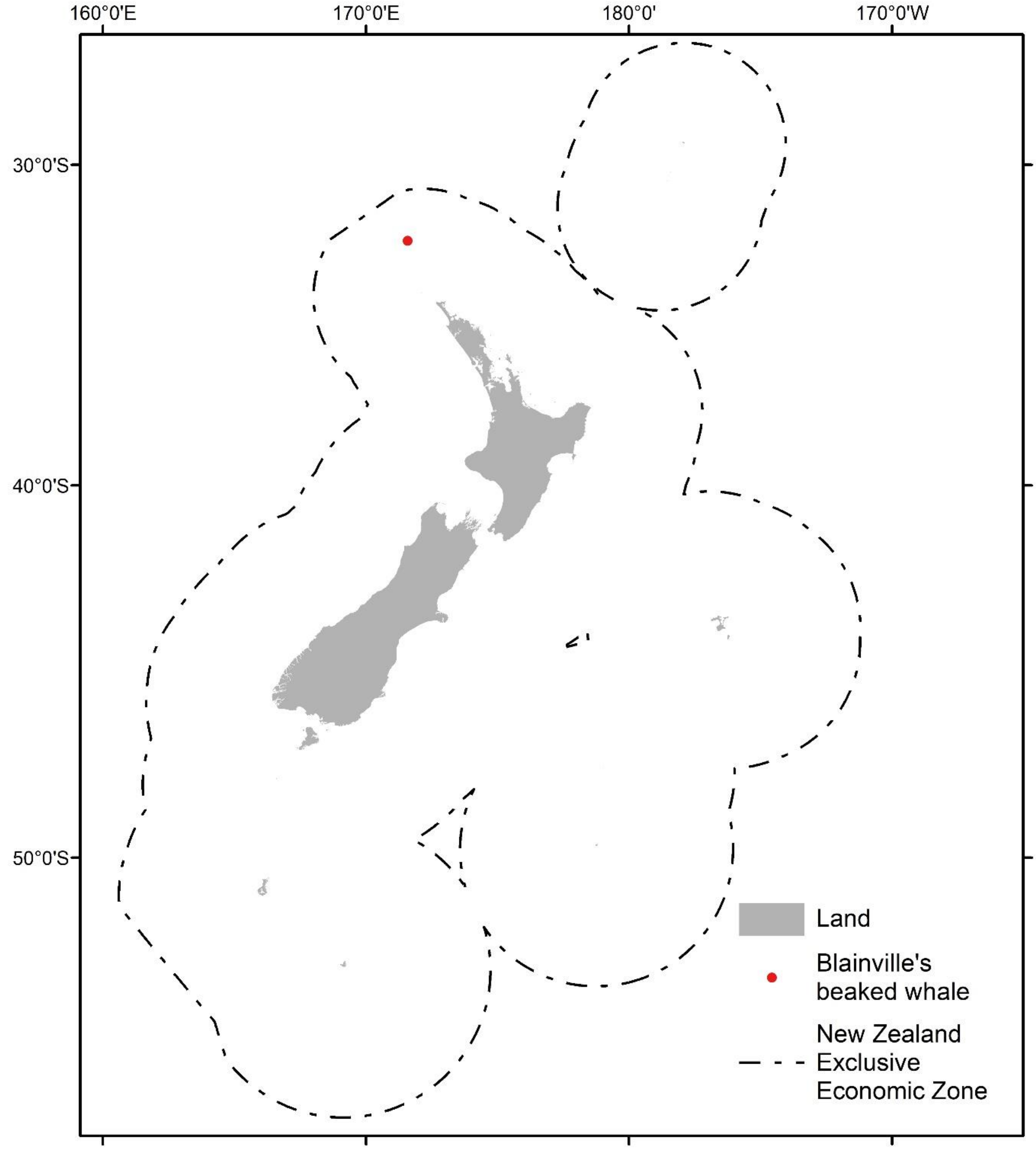

Fig A 4-1. Location of at-sea sightings of Blainville's beaked whale (n = 1) in the New Zealand EEZ.

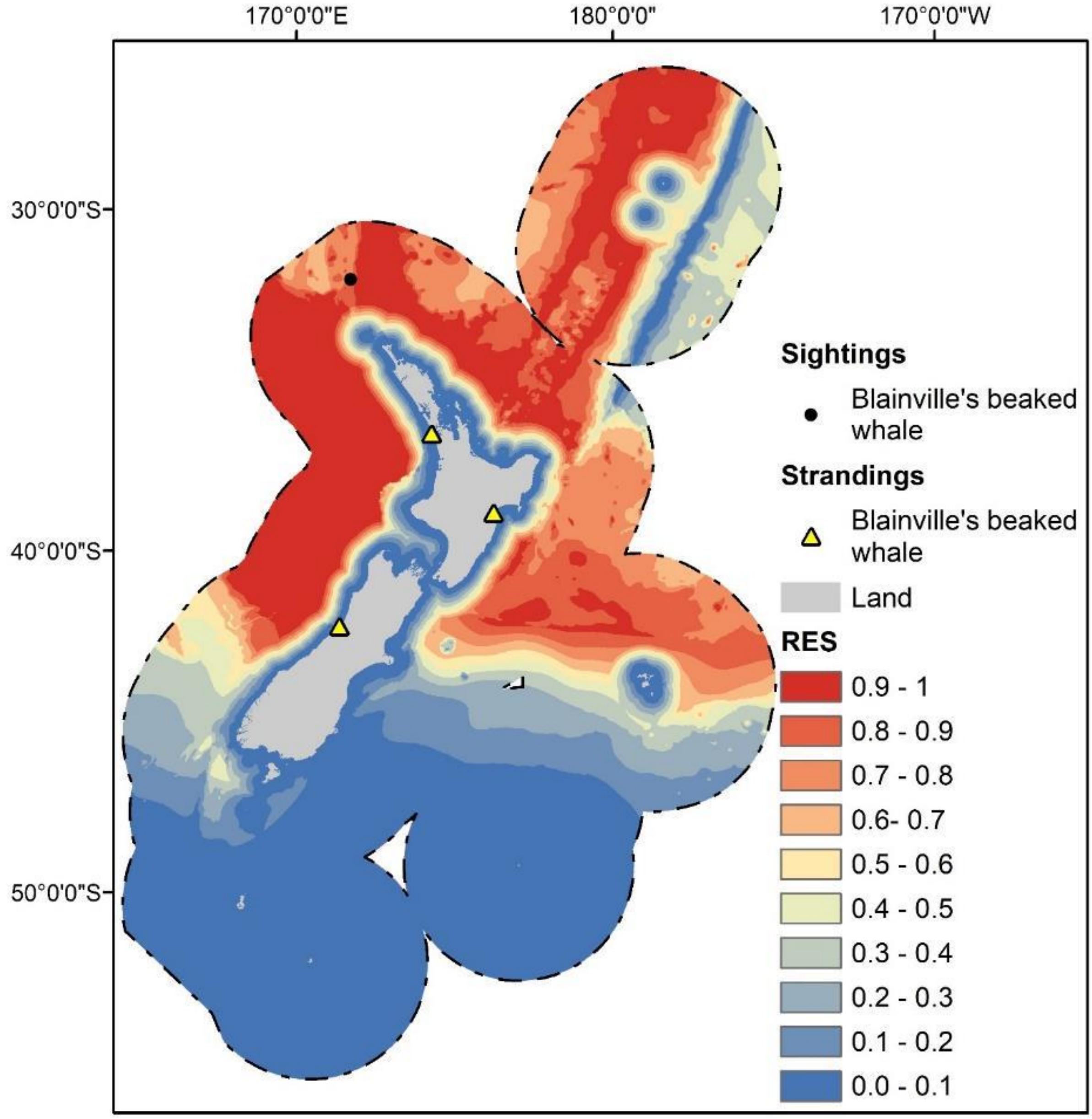

Fig A 4-2. Predicted RES scores for Blainville's beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Dwarf minke whale (*Balaenoptera acutorostrata*)

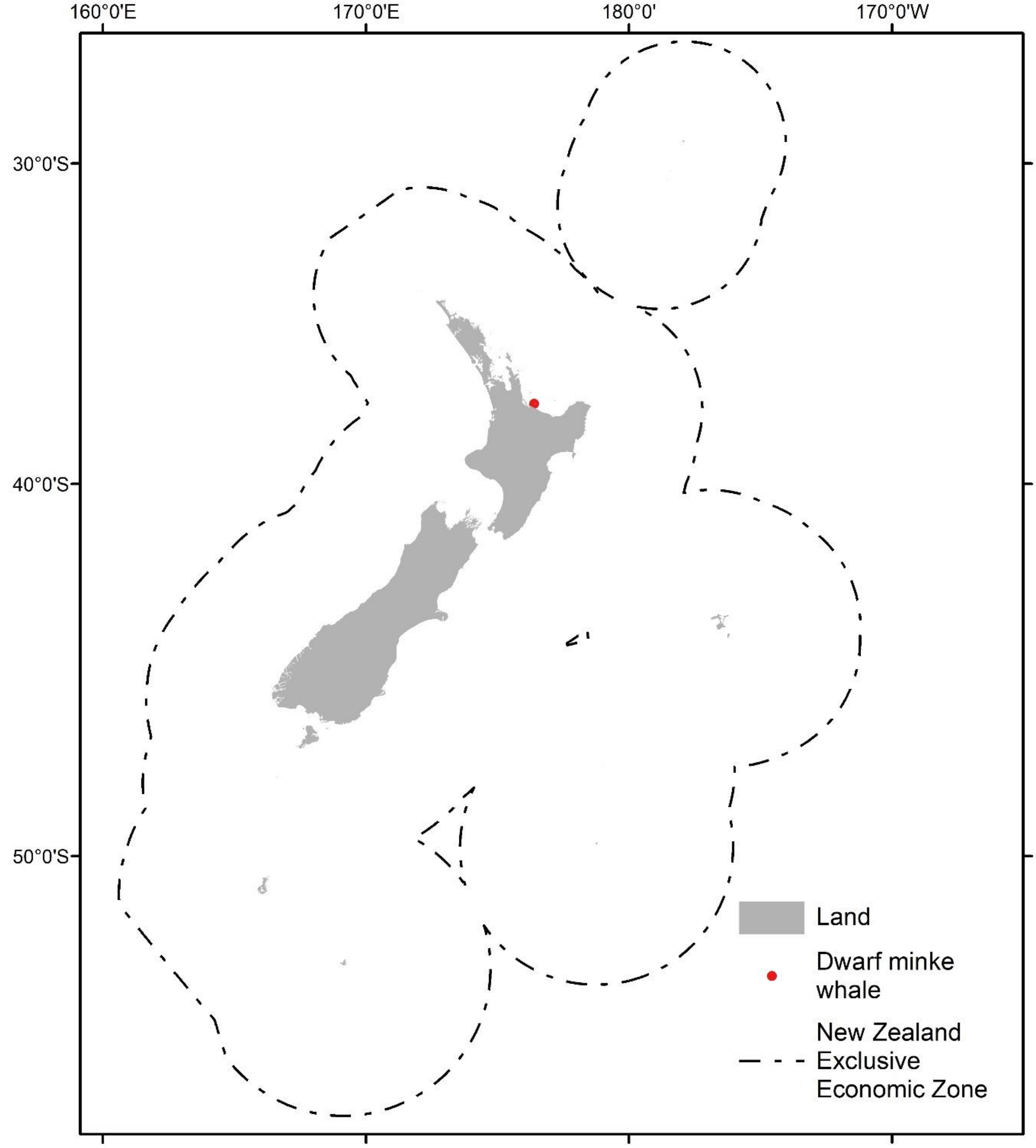

Fig A 4-3. Location of at-sea sightings of dwarf minke whale (*n* = 1) in the New Zealand EEZ.

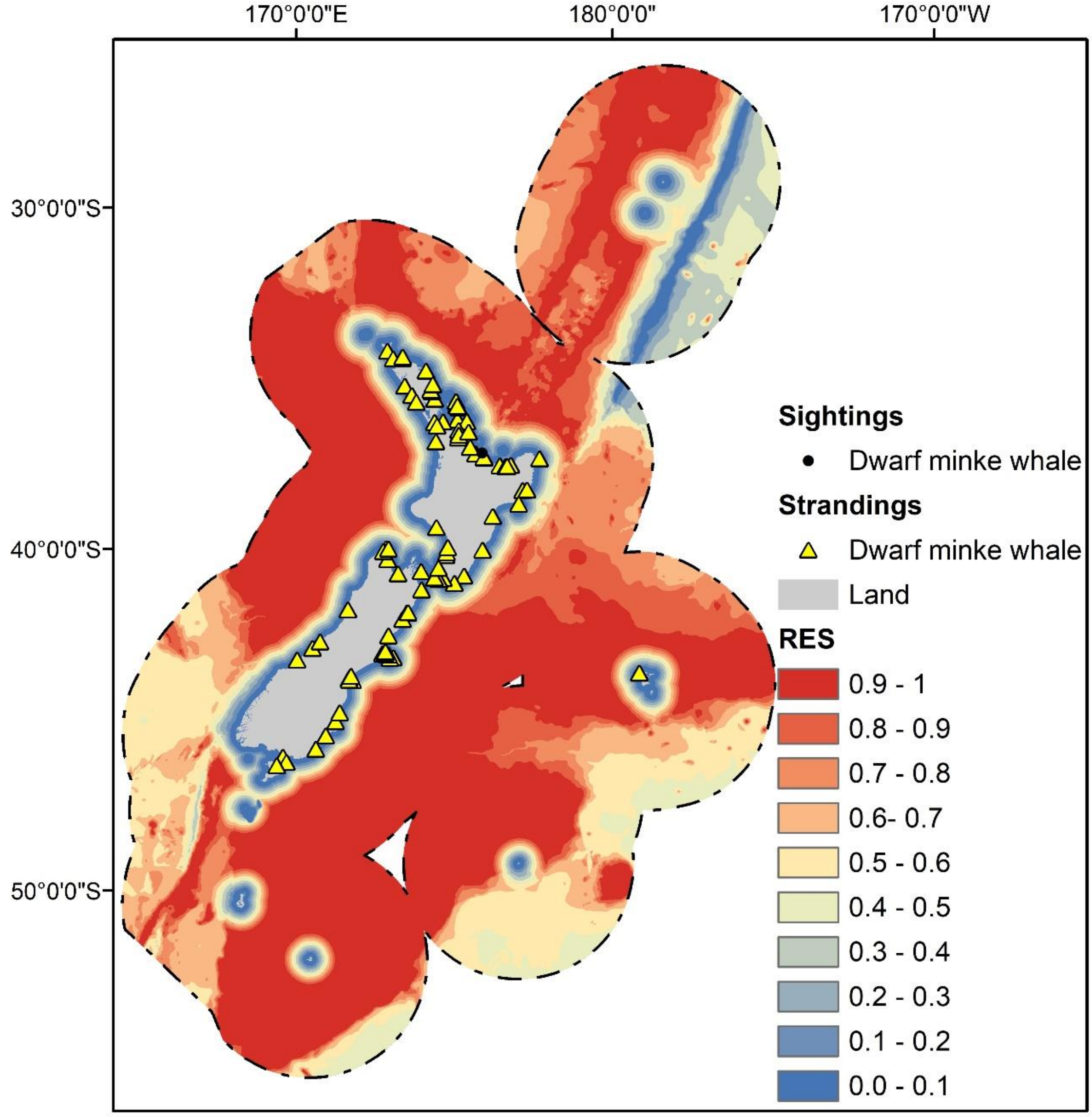

Fig A 4-4. Predicted RES scores for dwarf minke whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Spectacled porpoise (*Phocoena dioptrica*)

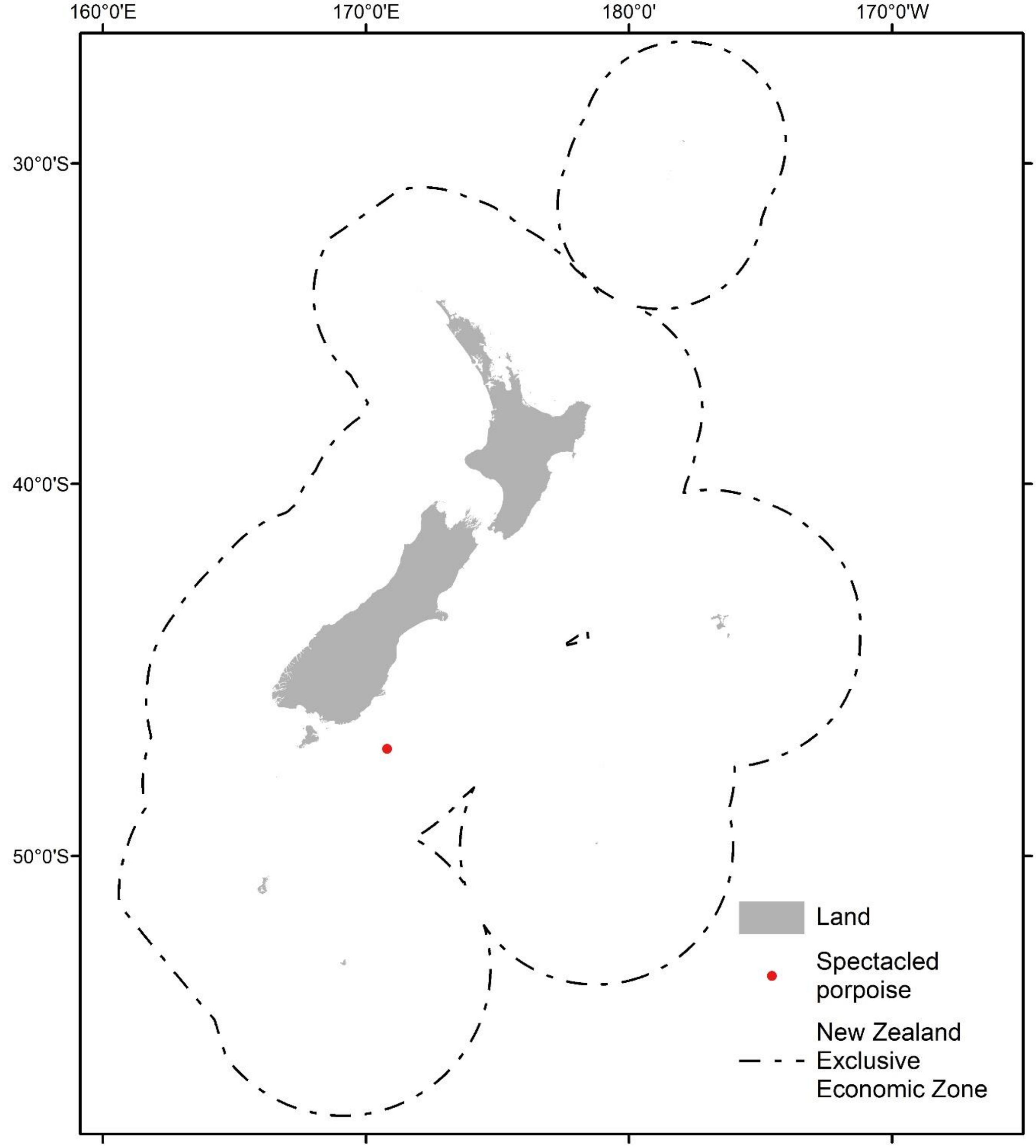

Fig A 4-5. Location of at-sea sightings of spectacled porpoise (*n* = 1) in the New Zealand EEZ.

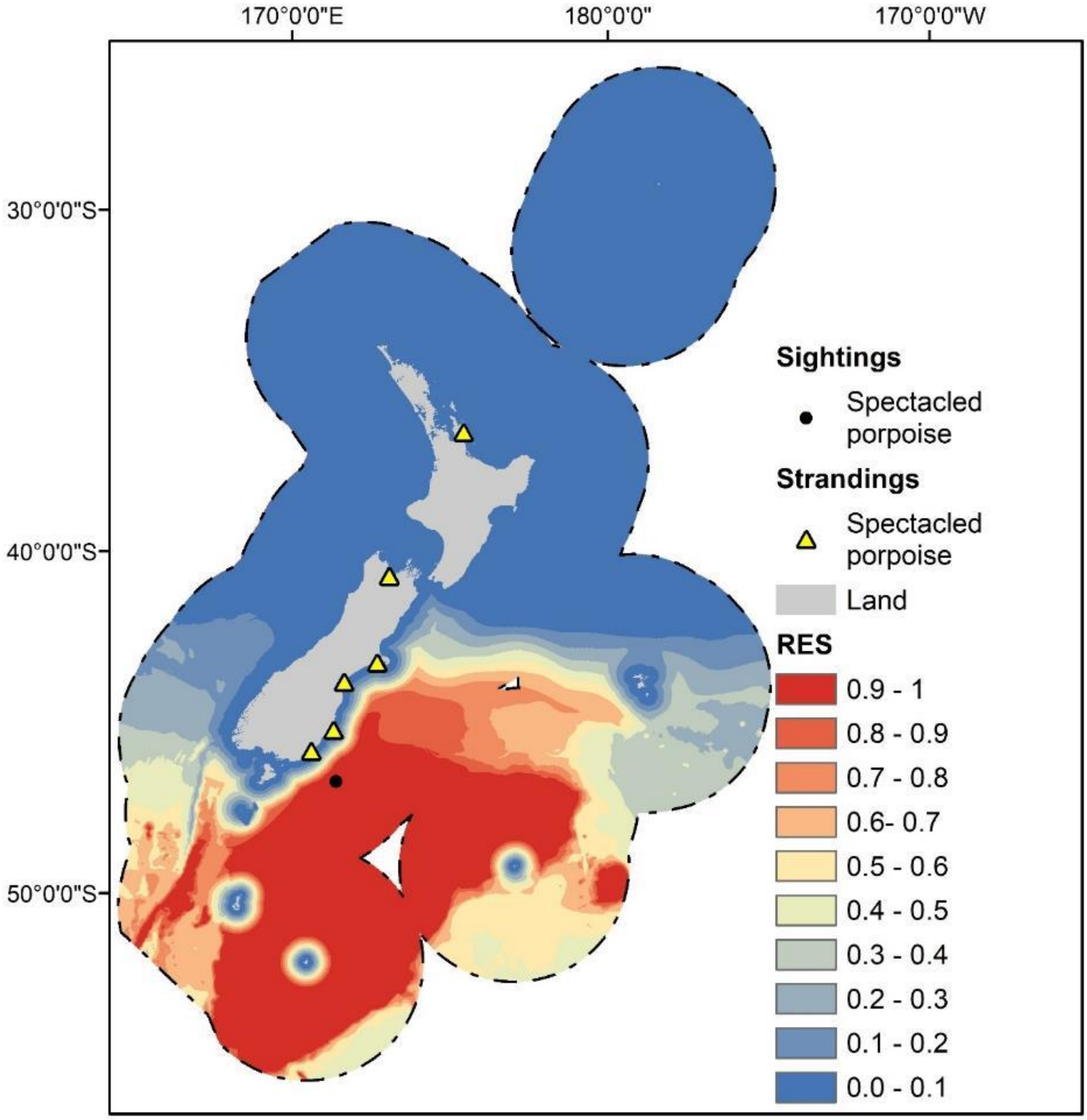

Fig A 4-6. Predicted RES scores for spectacled porpoise, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Striped dolphin (*Stenella coeruleoalba)*

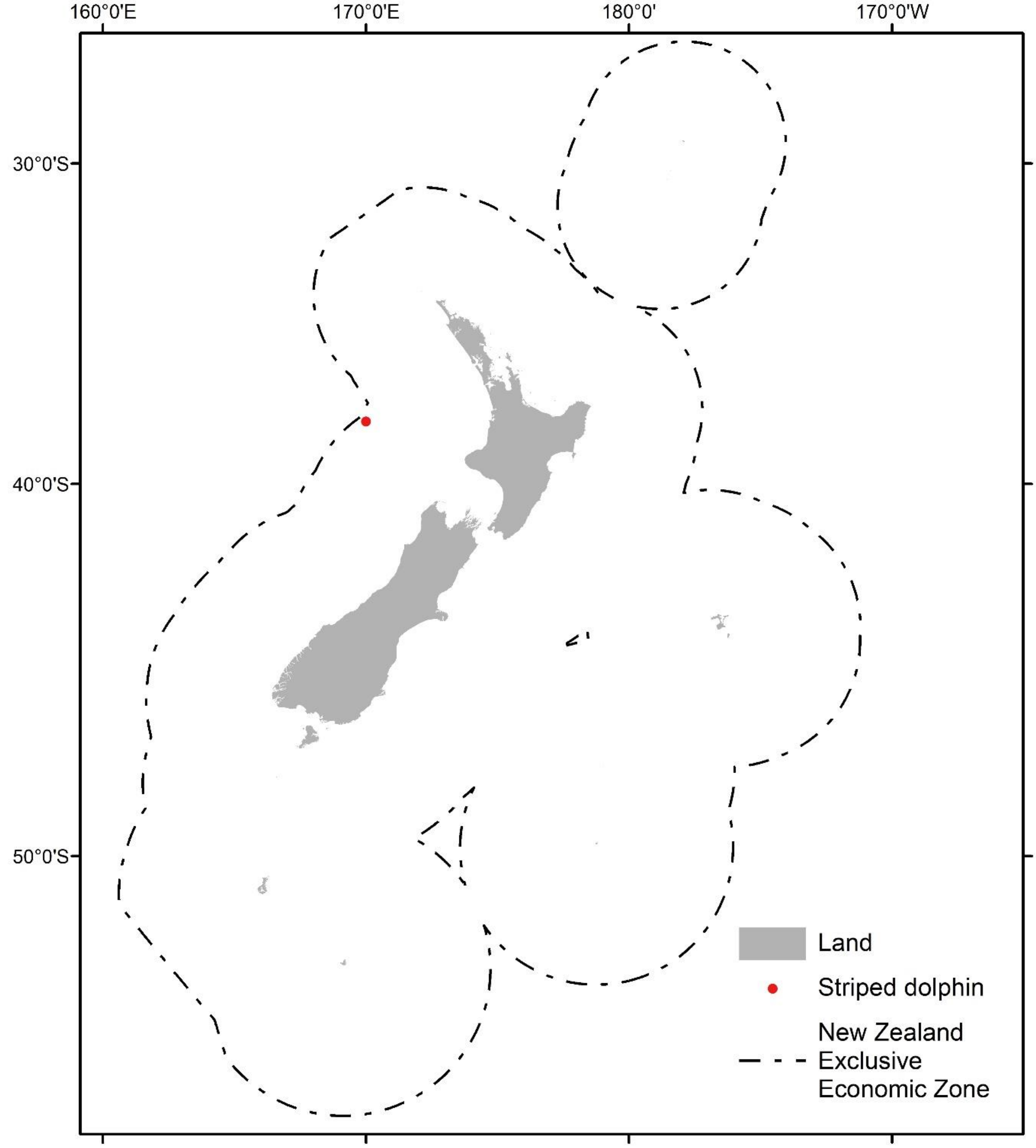

Fig A 4-7. Location of at-sea sightings of striped dolphin (*n* = 1) in the New Zealand EEZ.

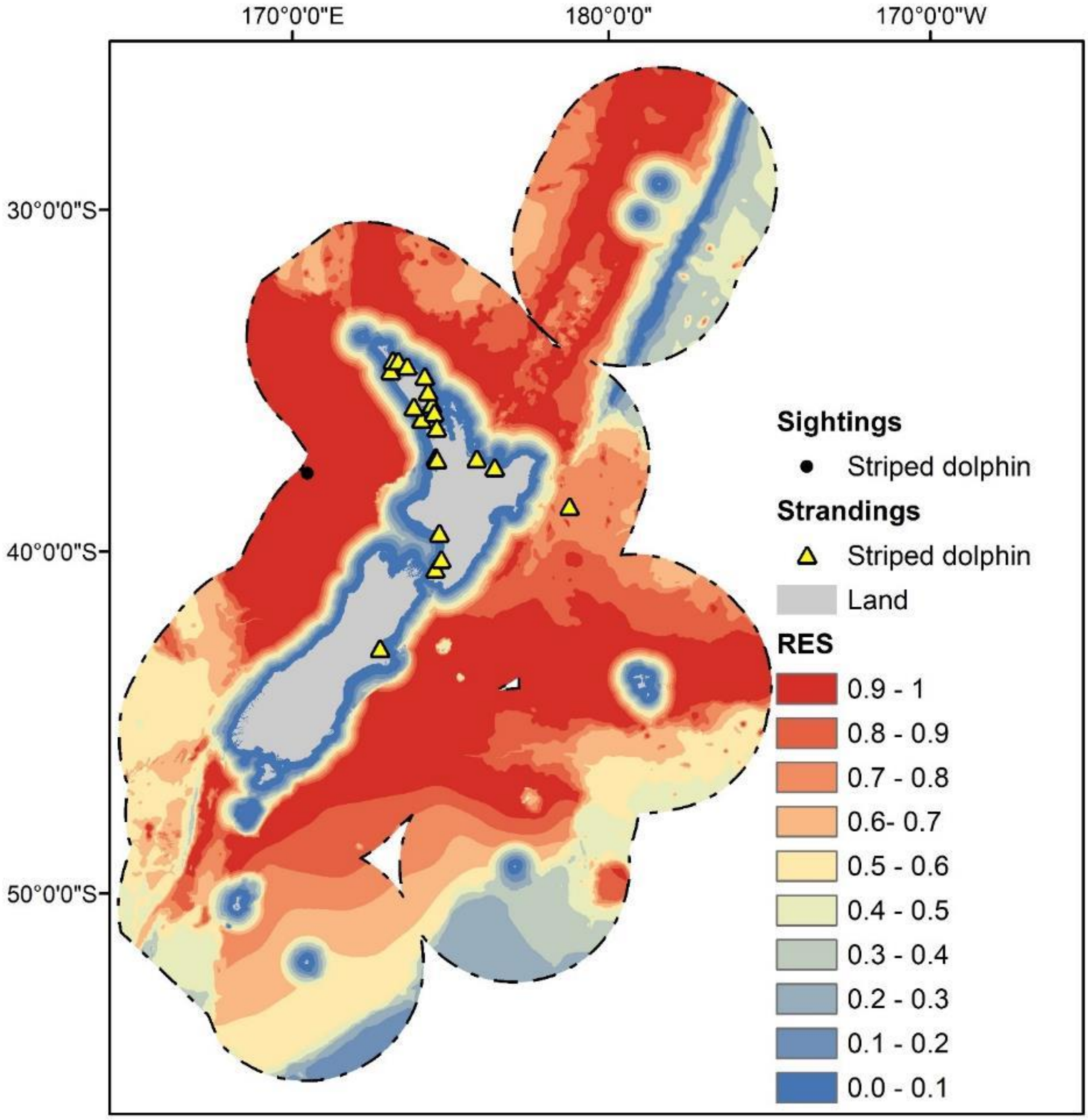

Fig A 4-8. Predicted RES scores for striped dolphin, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Andrew's beaked whale (*Mesoplodon bowdoini*)

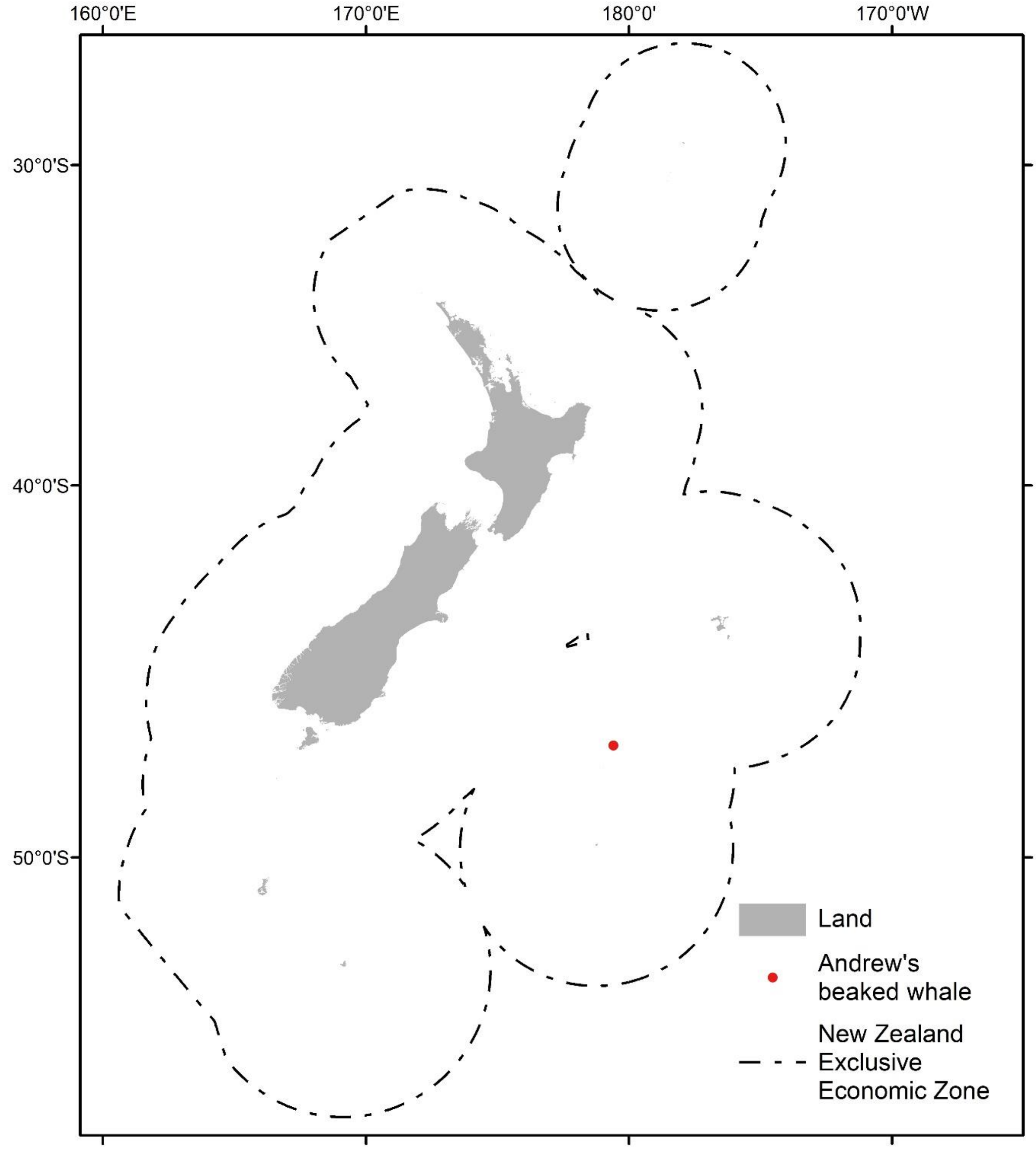

Fig A 4-9. Location of at-sea sightings of Andrew's beaked whale (*n* = 2) in the New Zealand EEZ.

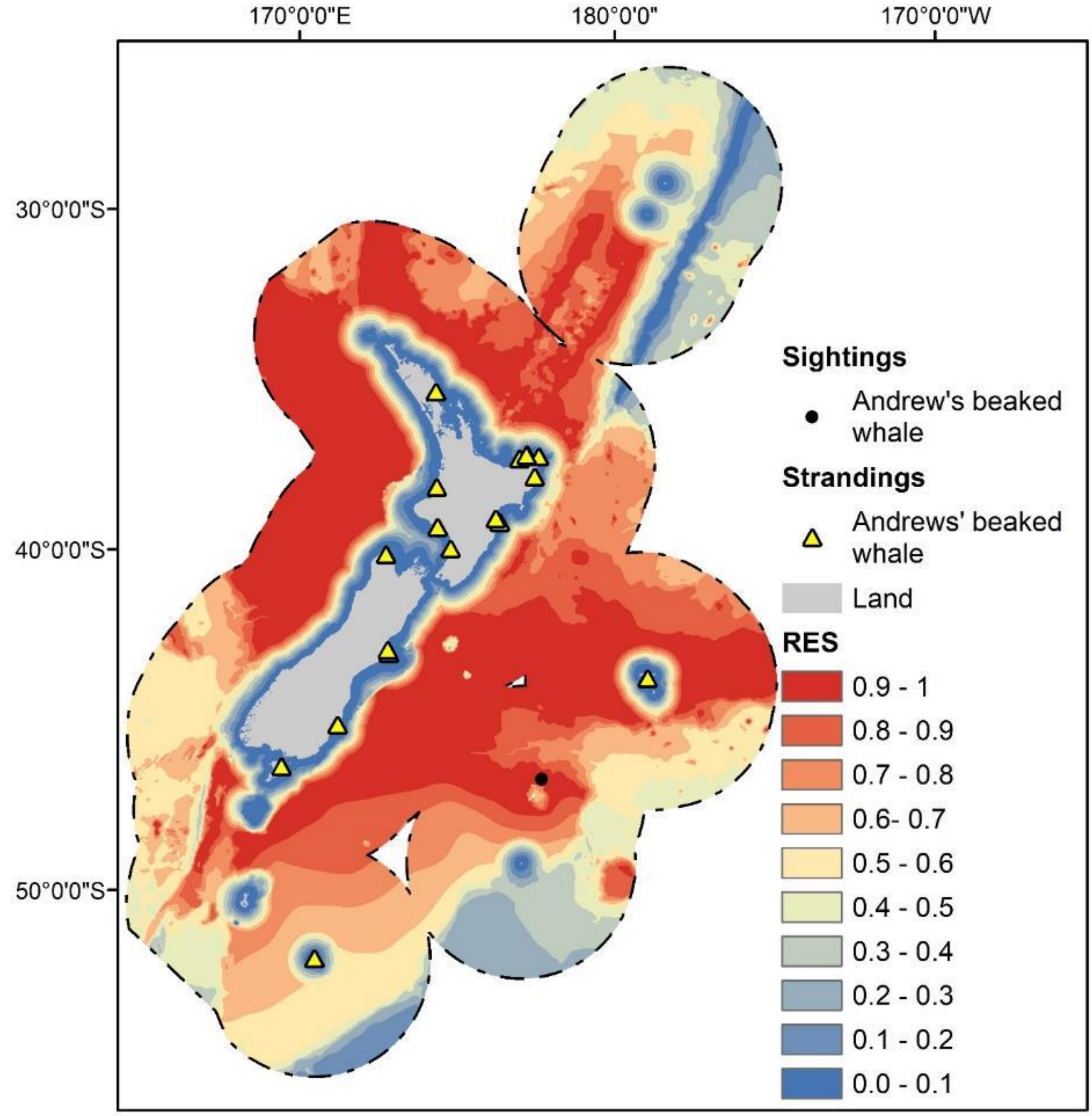

Fig A 4-10. Predicted RES scores for Andrew's beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Hourglass dolphin (*Lagenorhynchus cruciger*)

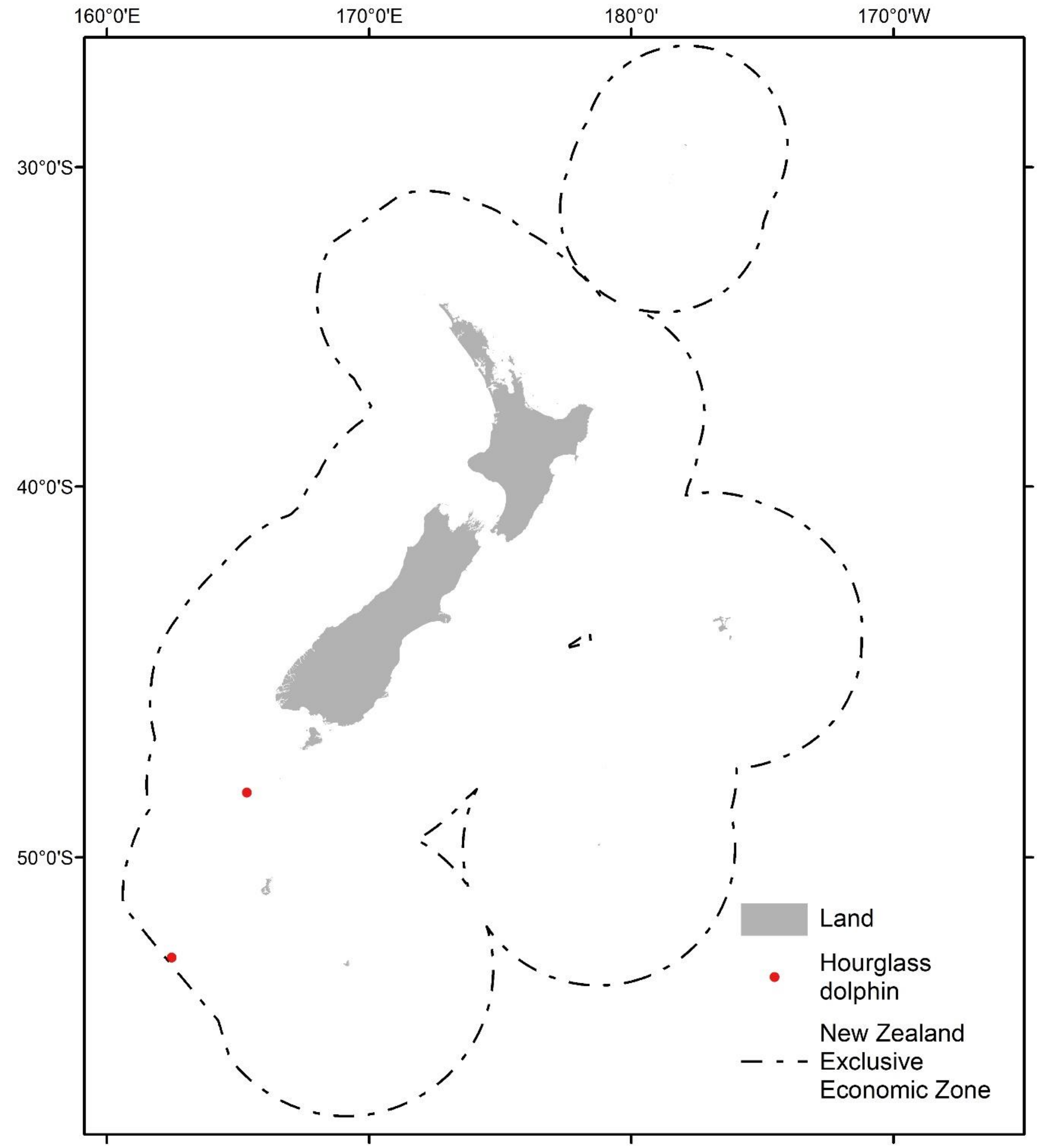

Fig A 4-11. Location of at-sea sightings of hourglass dolphin (*n* = 2) in the New Zealand EEZ

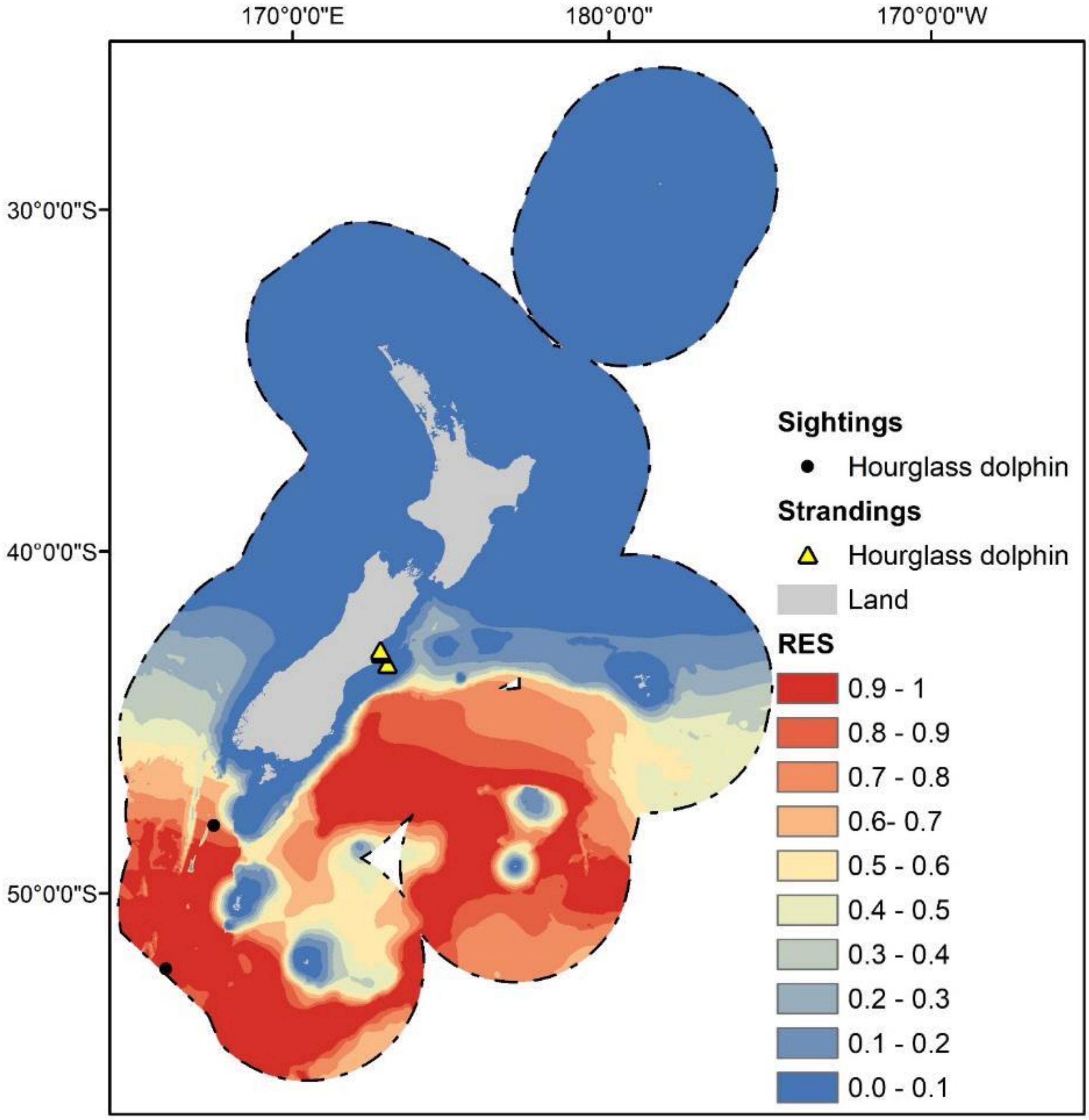

Fig A 4-12. Predicted RES scores for hourglass dolphin, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Pygmy sperm whale (*Kogia breviceps*)

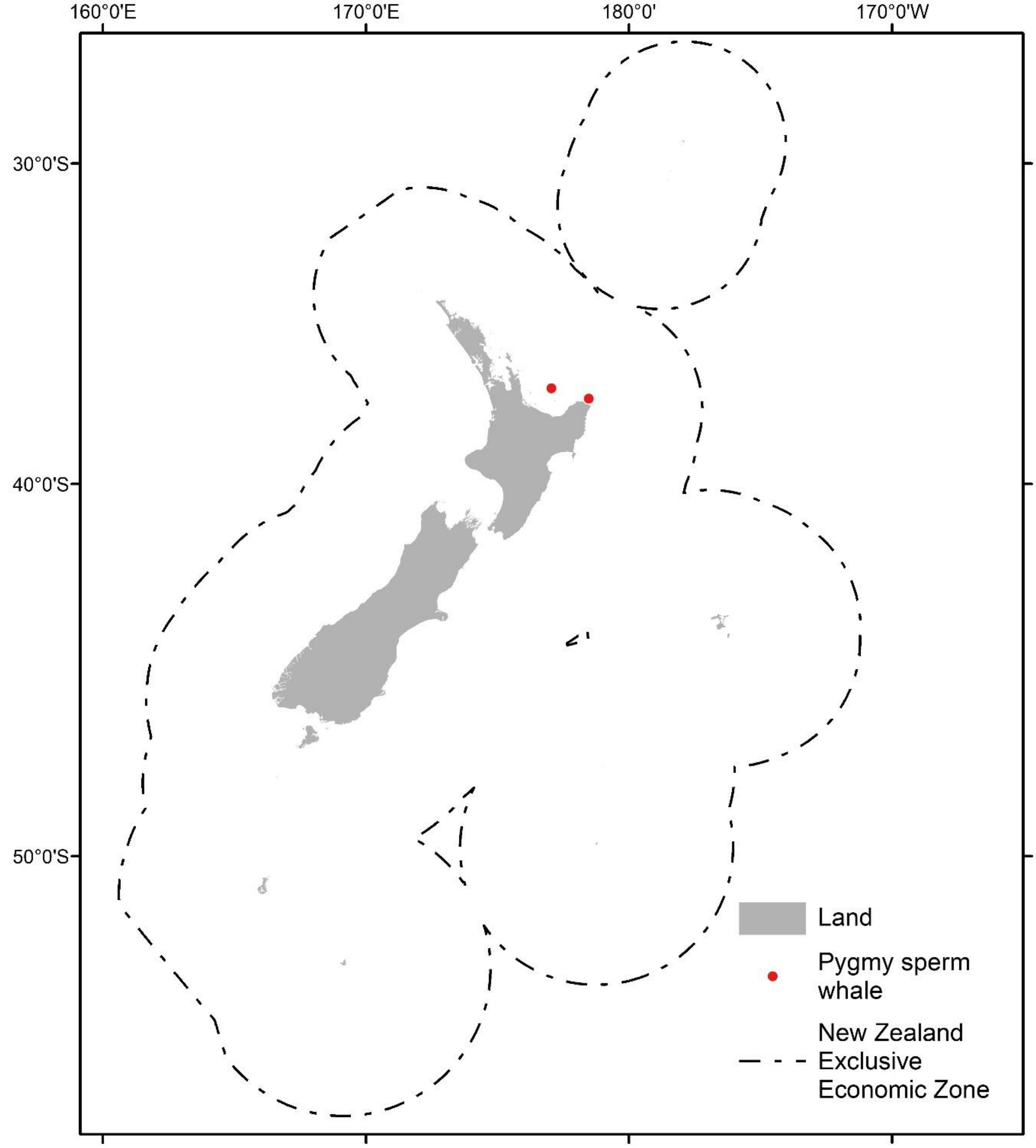

Fig A 4-13. Location of at-sea sightings of pygmy sperm whale (*n* = 2) in the New Zealand EEZ.

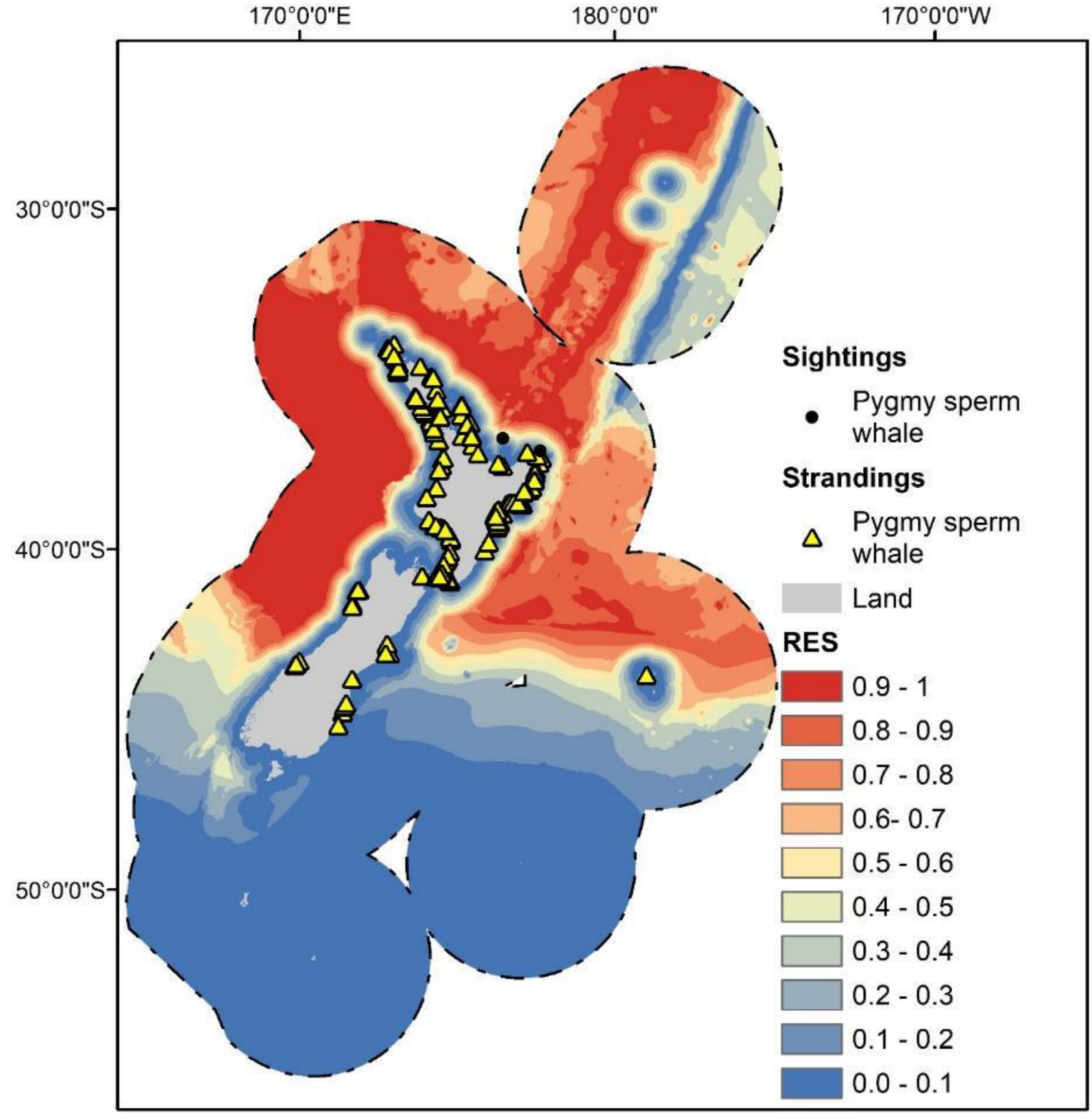

Fig A 4-14. Predicted RES scores for pygmy sperm whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Southern bottlenose whale (*Hyperoodon planifrons*)

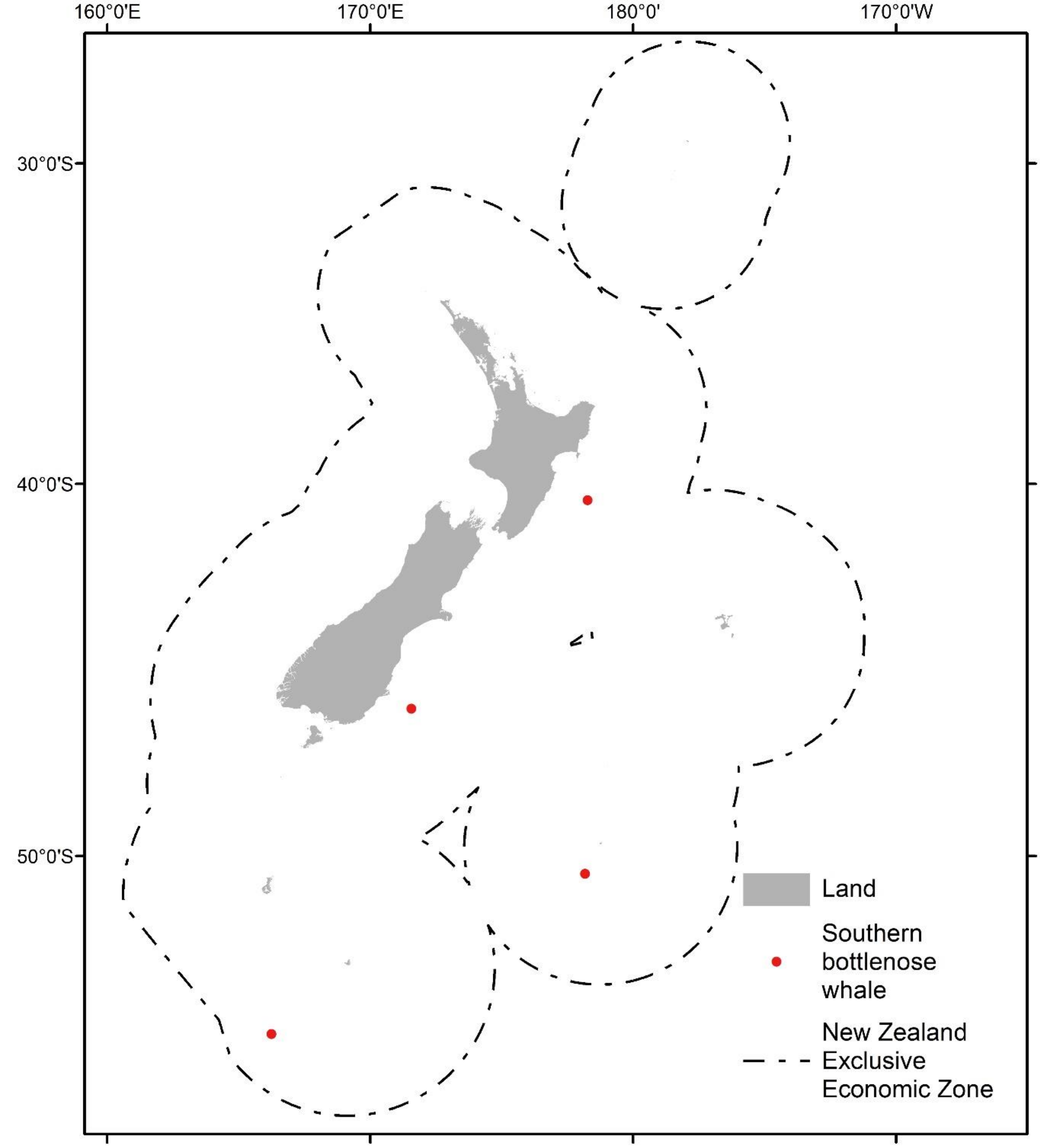

Fig A 4-15. Location of at-sea sightings of southern bottlenose whale (n = 4) in the New Zealand EEZ.

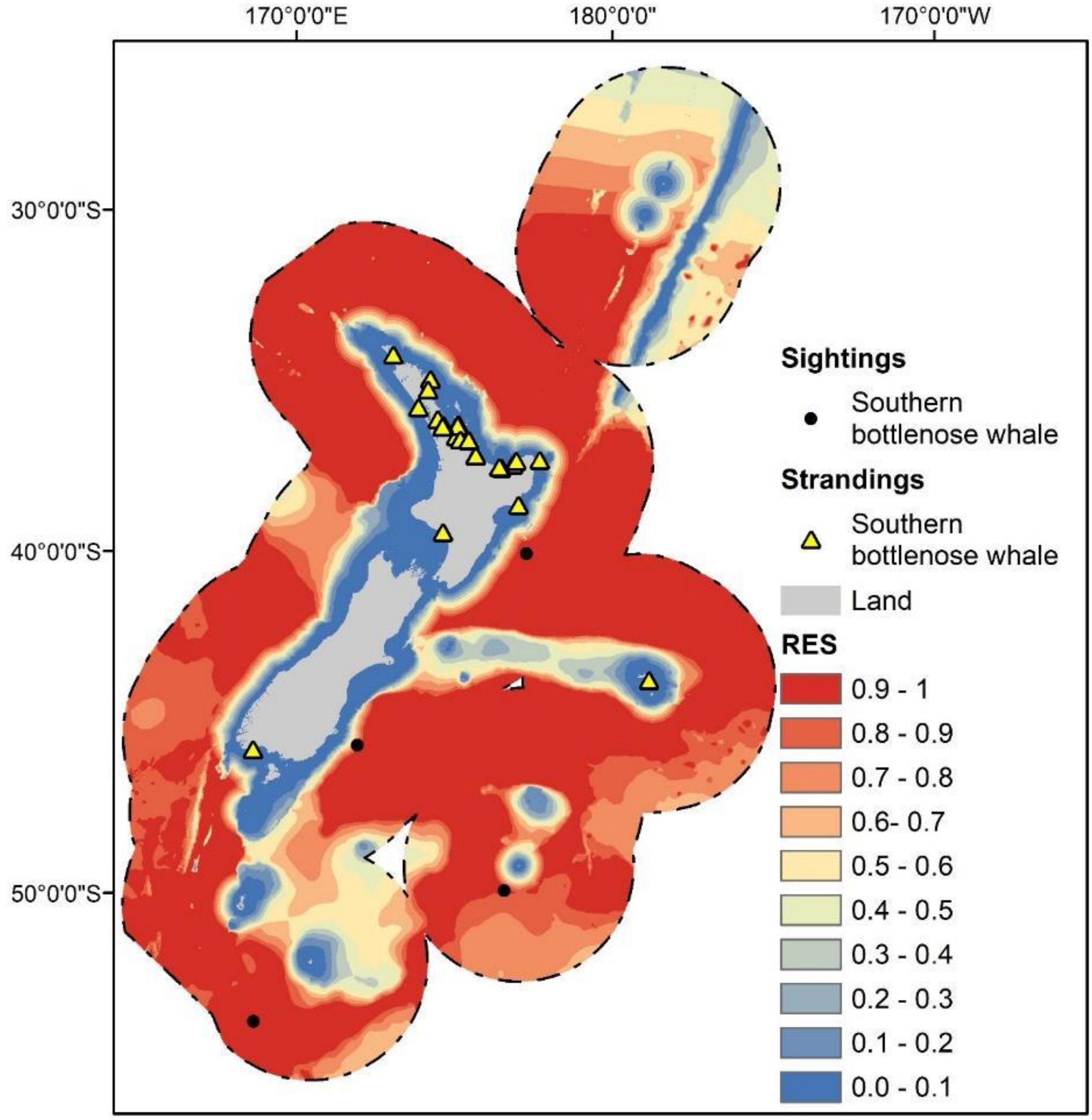

Fig A 4-16. Predicted RES scores for southern bottlenose whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Risso's dolphin (*Grampus griseus*)

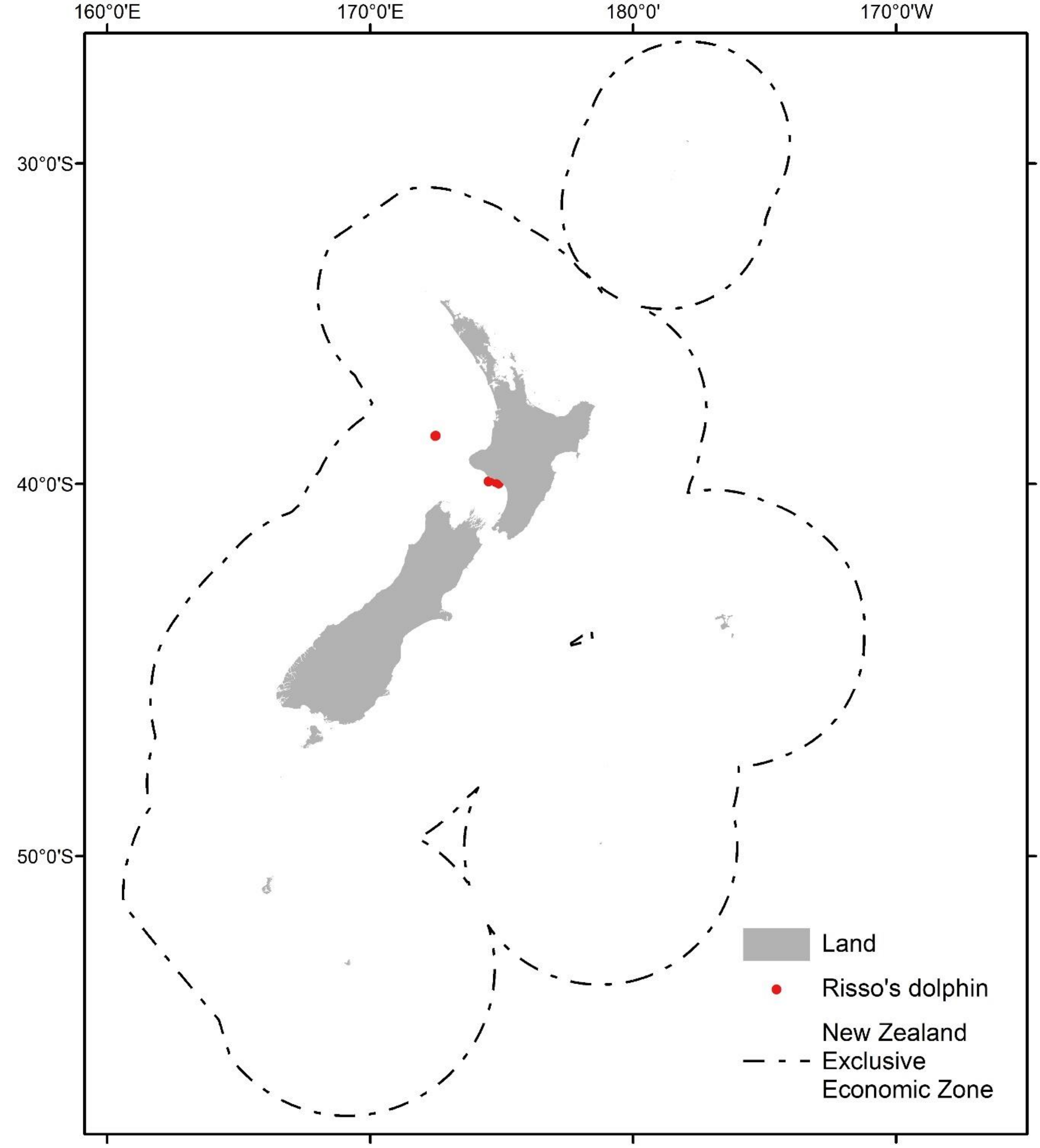

Fig A 4-17. Location of at-sea sightings of Risso's dolphin (*n* = 5) in the New Zealand EEZ.

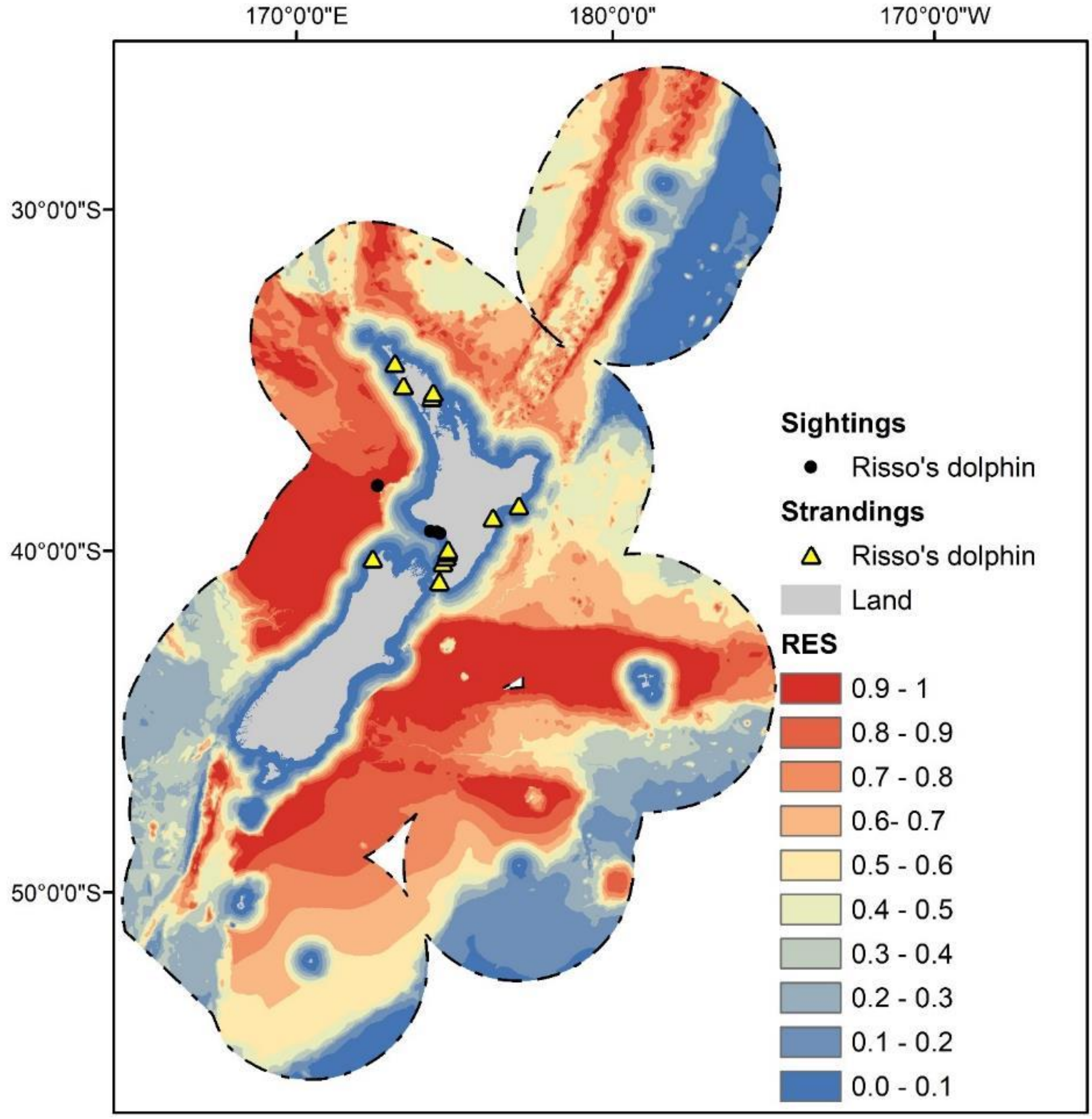

Fig A 4-18. Predicted RES scores for Risso's dolphin, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Shepherd's beaked whale (*Tasmacetus shepherdi*)

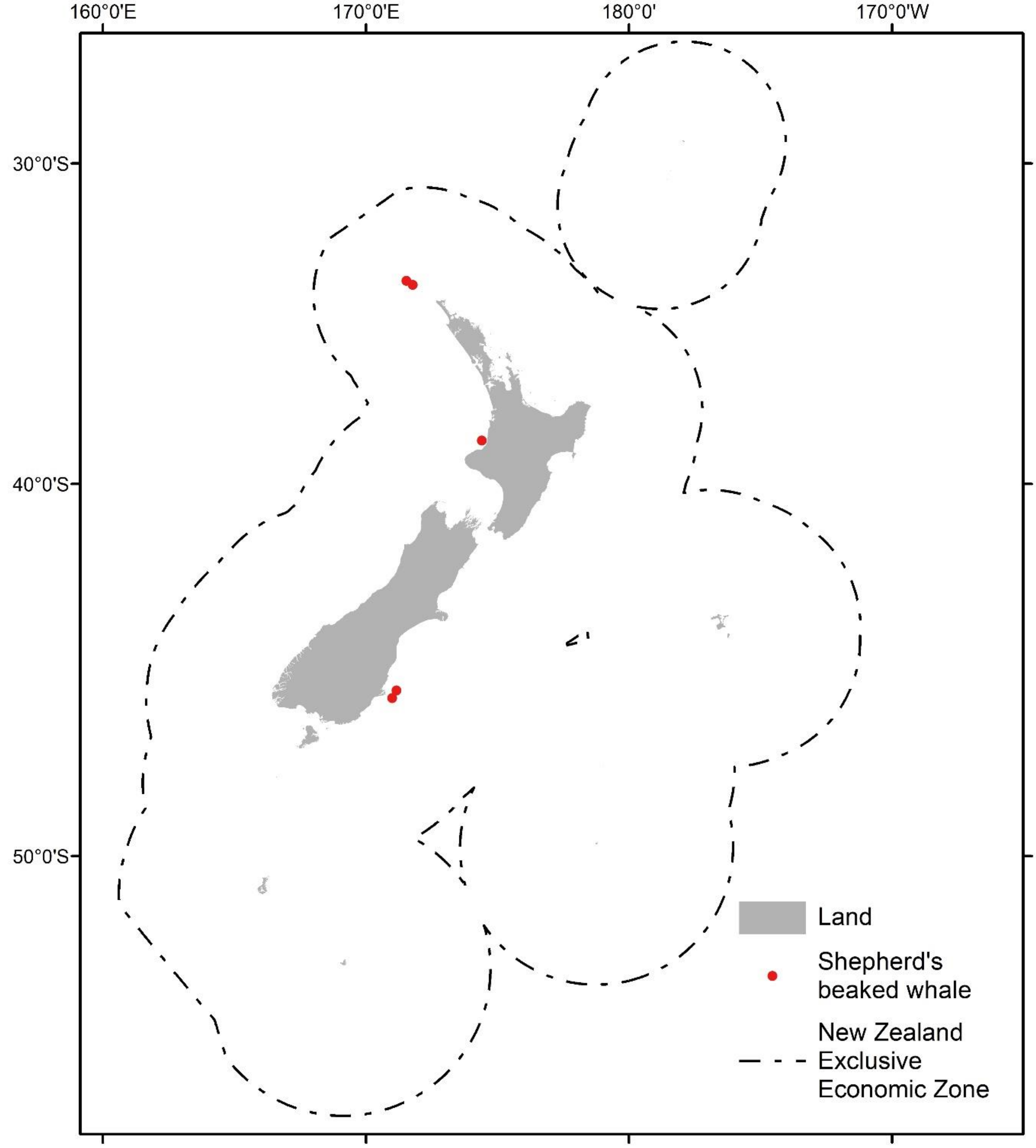

Fig A 4-19. Location of at-sea sightings of Shepherd's beaked whale (n = 5) in the New Zealand EEZ.

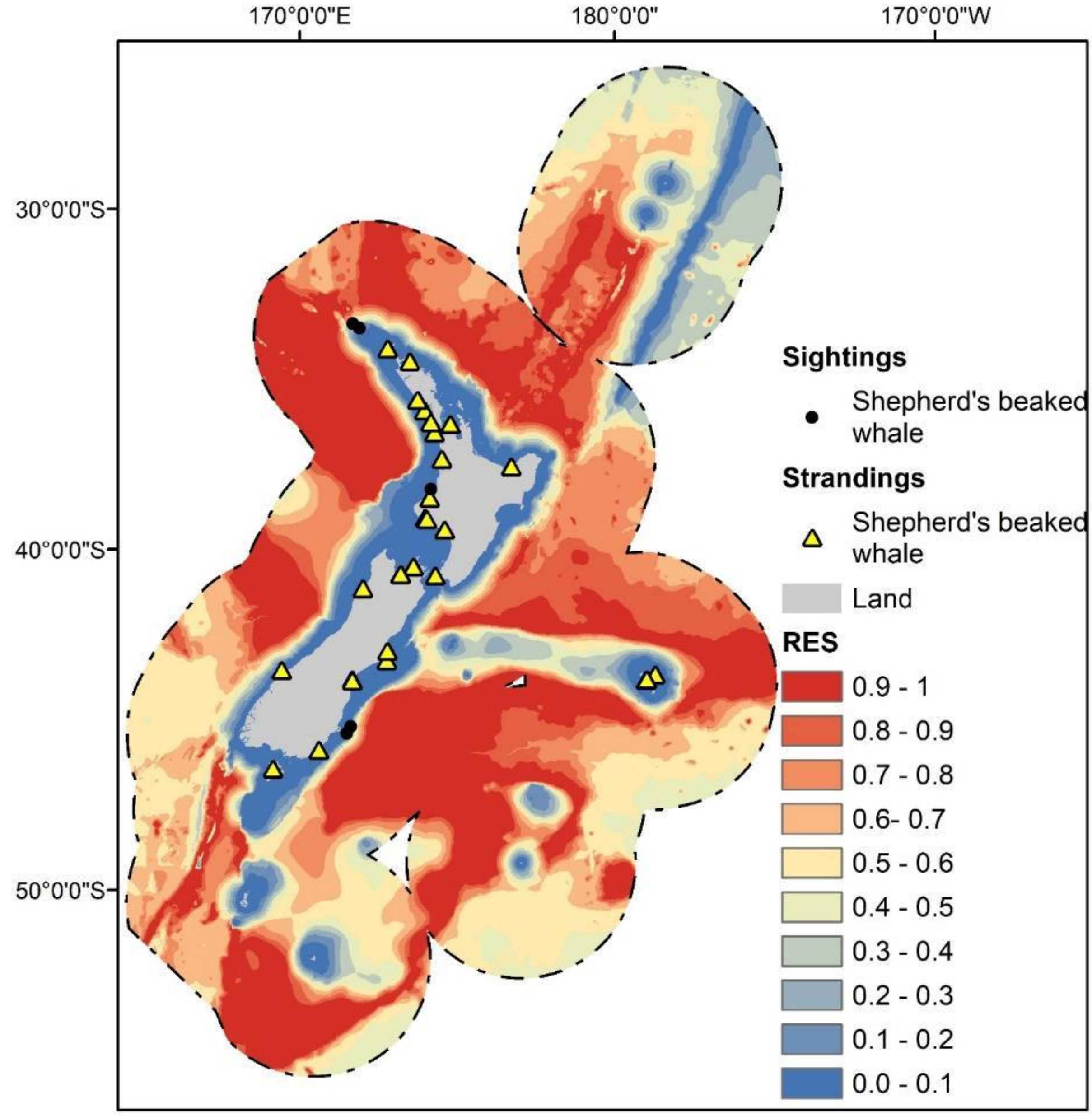

Fig A 4-20. Predicted RES scores for Shepherd’s beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Cuvier's beaked whale (*Ziphius cavirostris*)

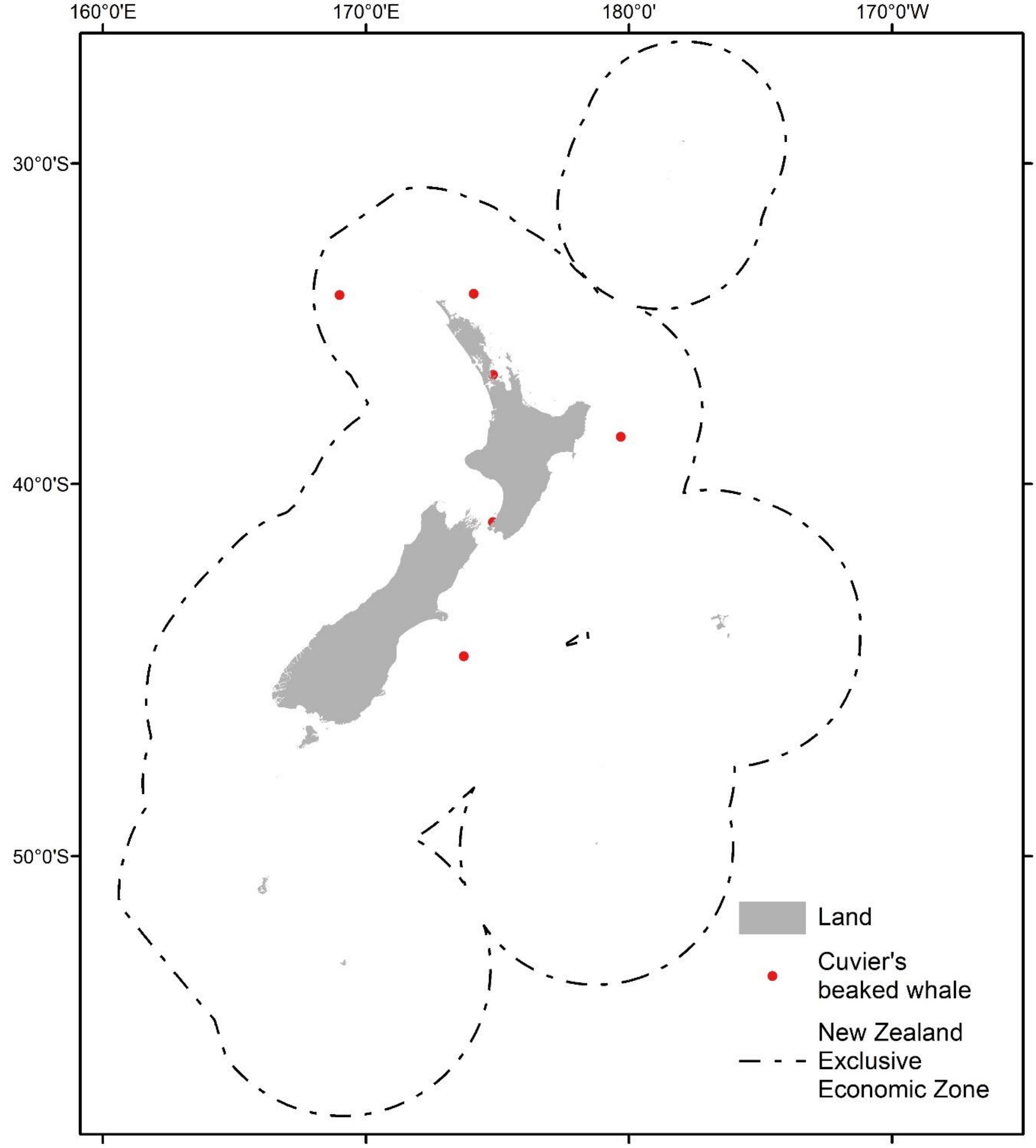

Fig A 4-21. Location of at-sea sightings of Cuvier's beaked whale (*n* = 7) in the New Zealand EEZ.

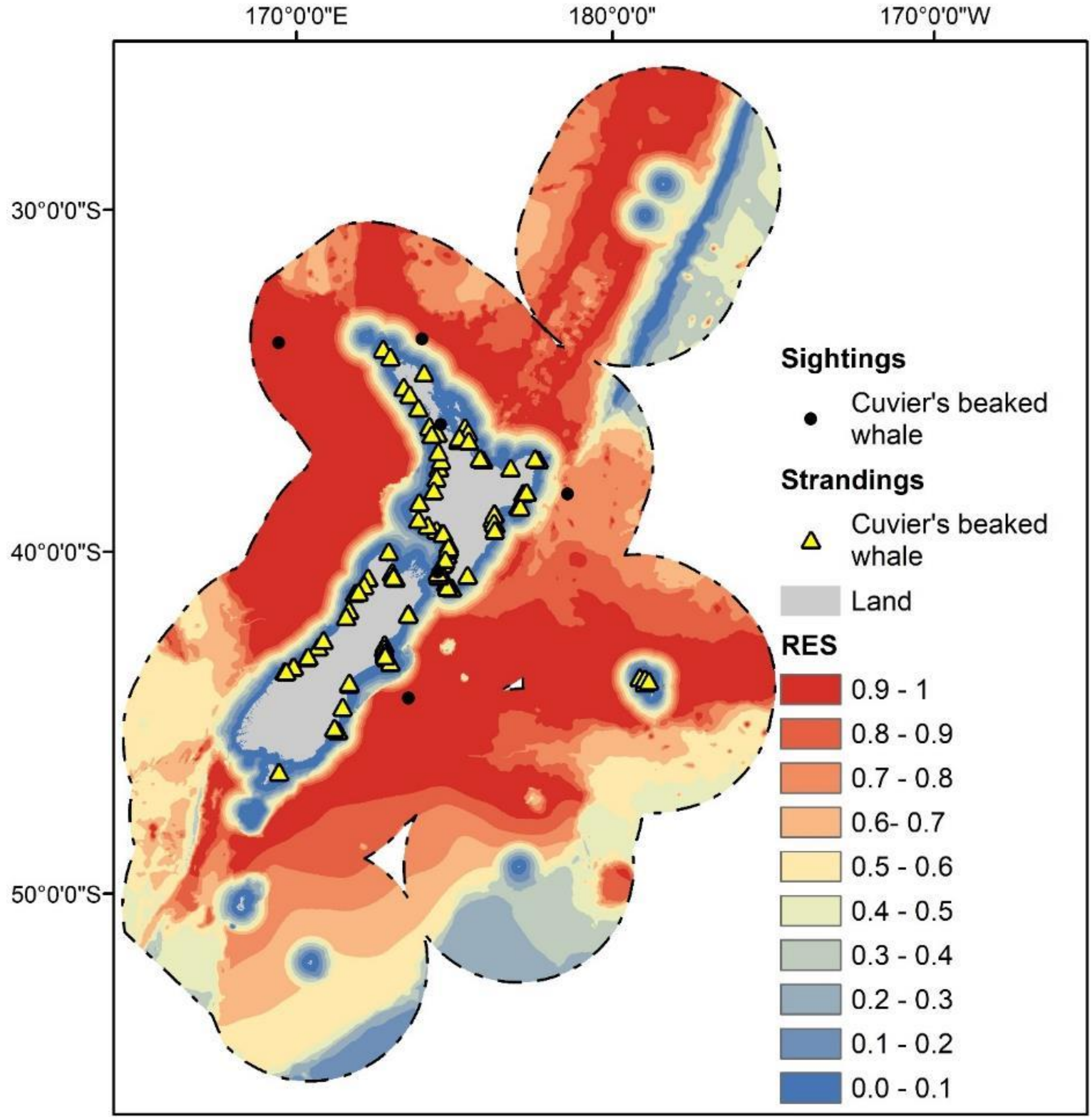

Fig A 4-22. Predicted RES scores for Cuvier's beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Gray's beaked whale (*Mesoplodon grayi*)

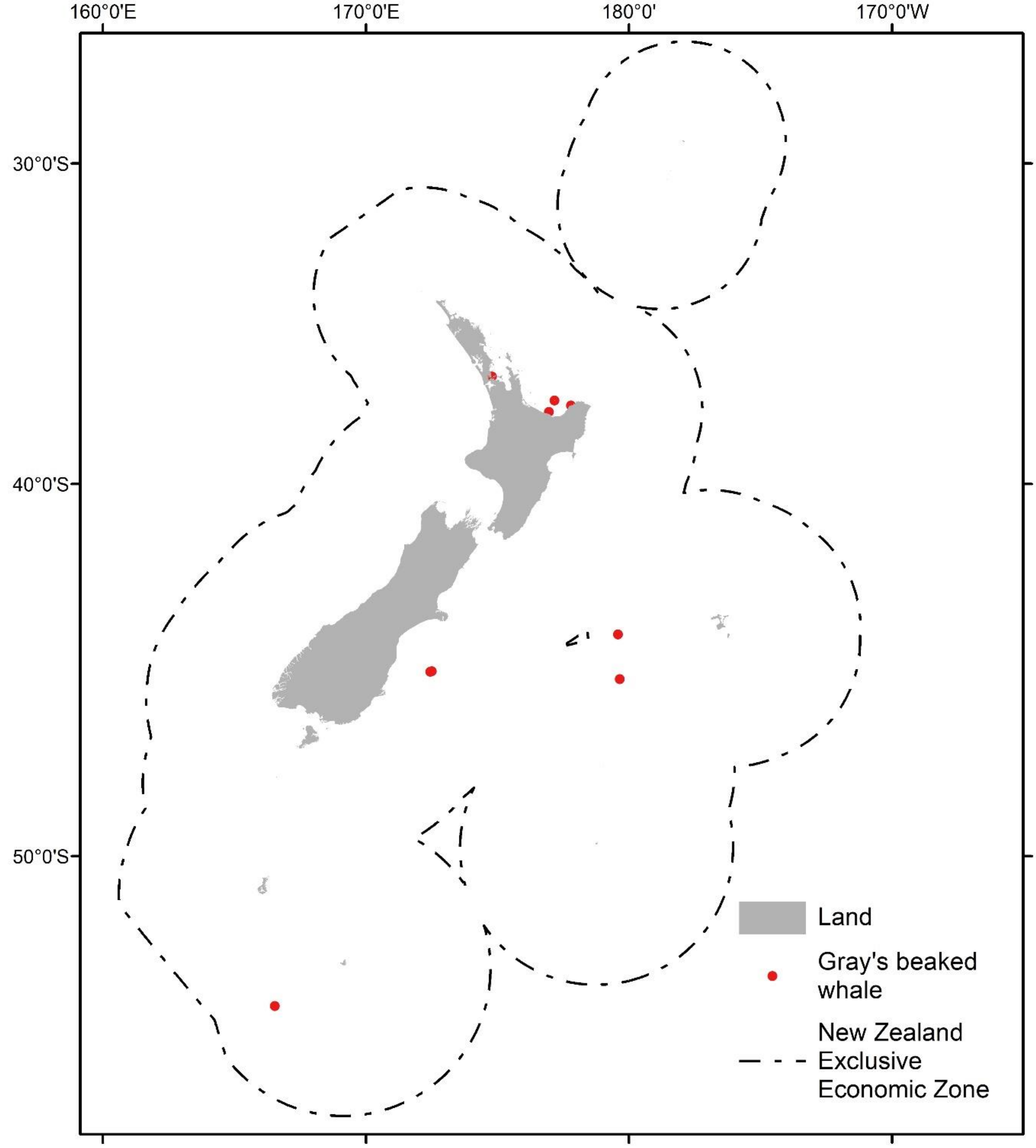

Fig A 4-23. Location of at-sea sightings of Gray's beaked whale ($n = 9$) in the New Zealand EEZ.

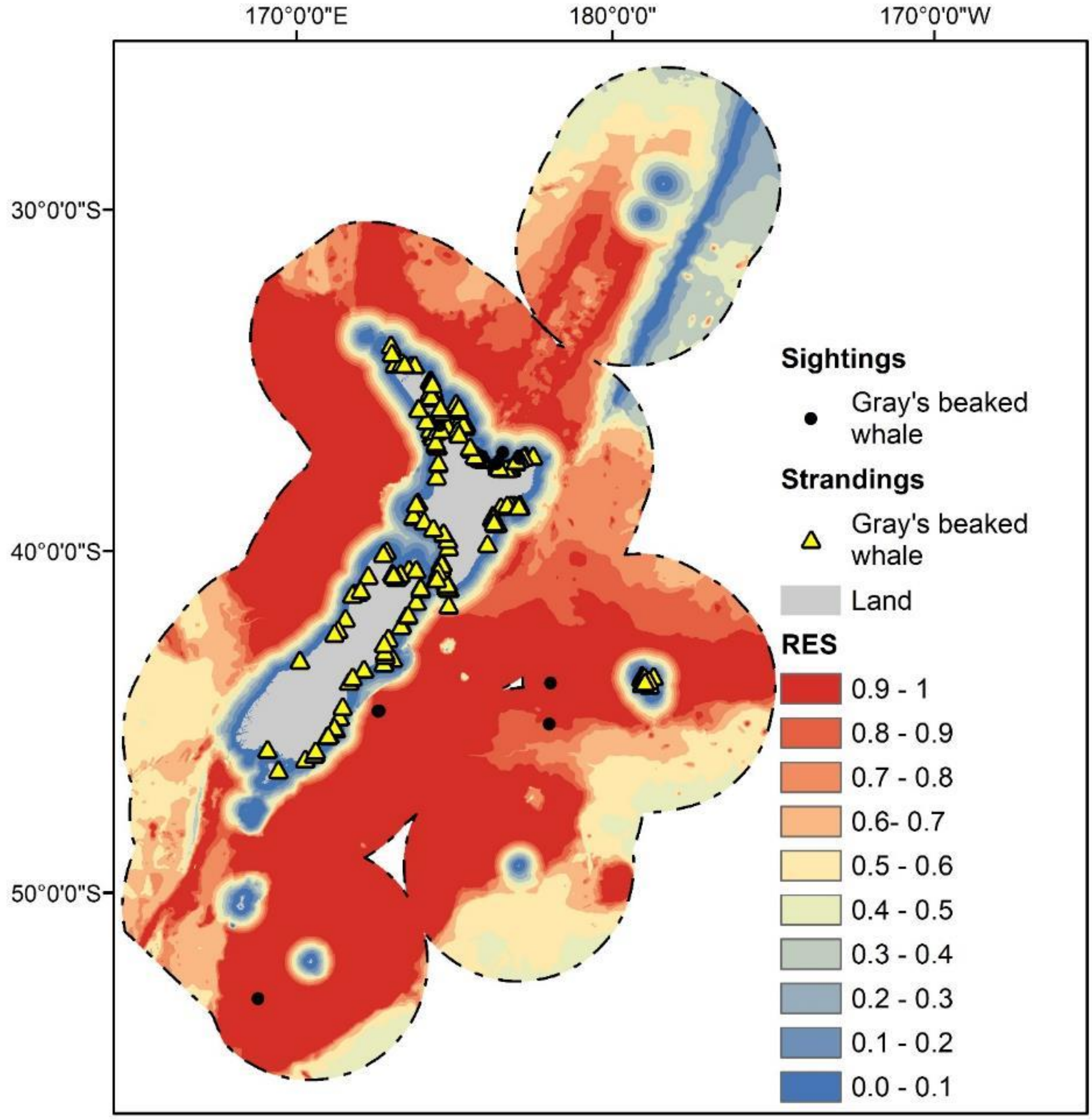

Fig A 4-24. Predicted RES scores for Gray's beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Southern right whale dolphin (*Lissodelphis peronii*)

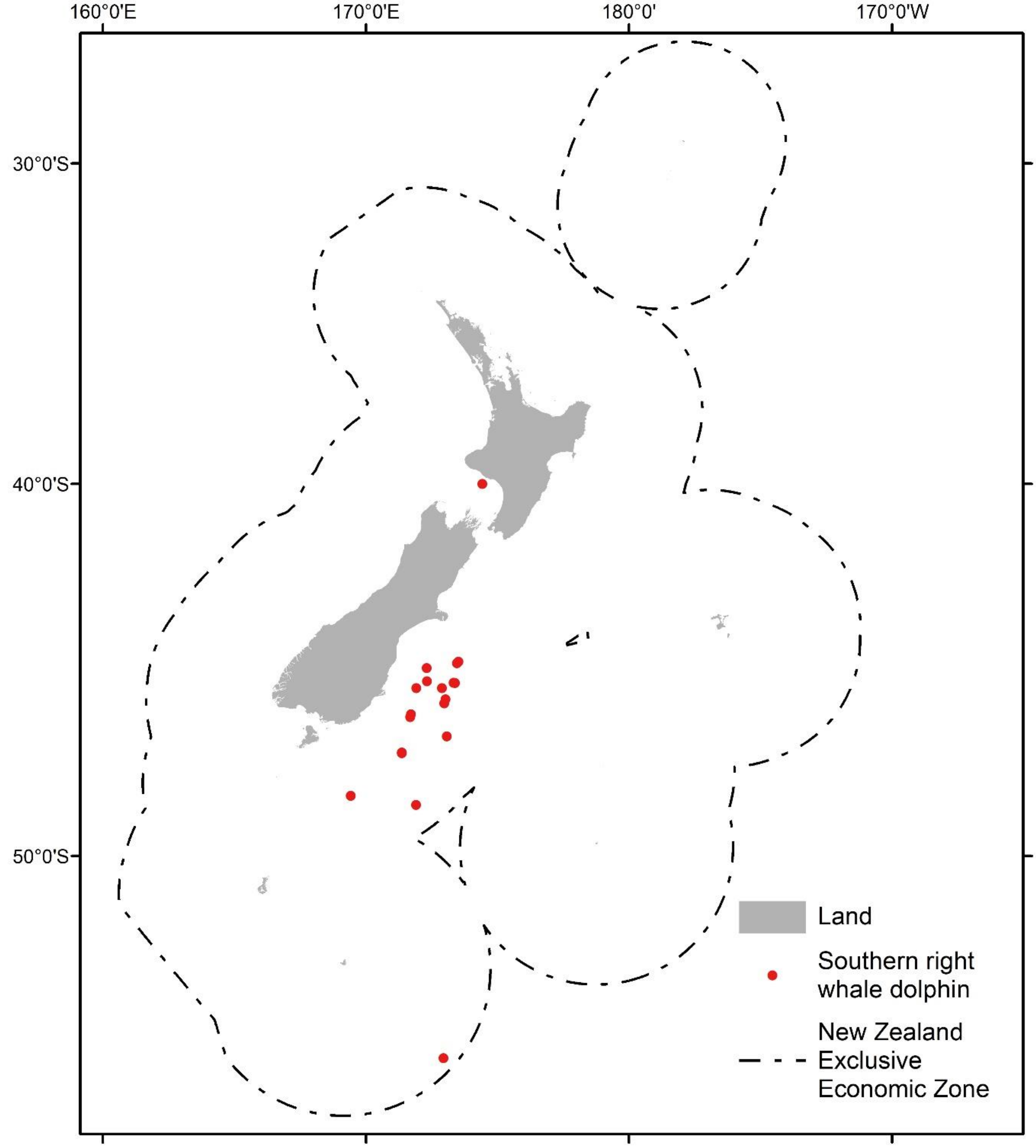

Fig A 4-25. Location of at-sea sightings of southern right whale dolphin (n = 27) in the NZ EEZ.

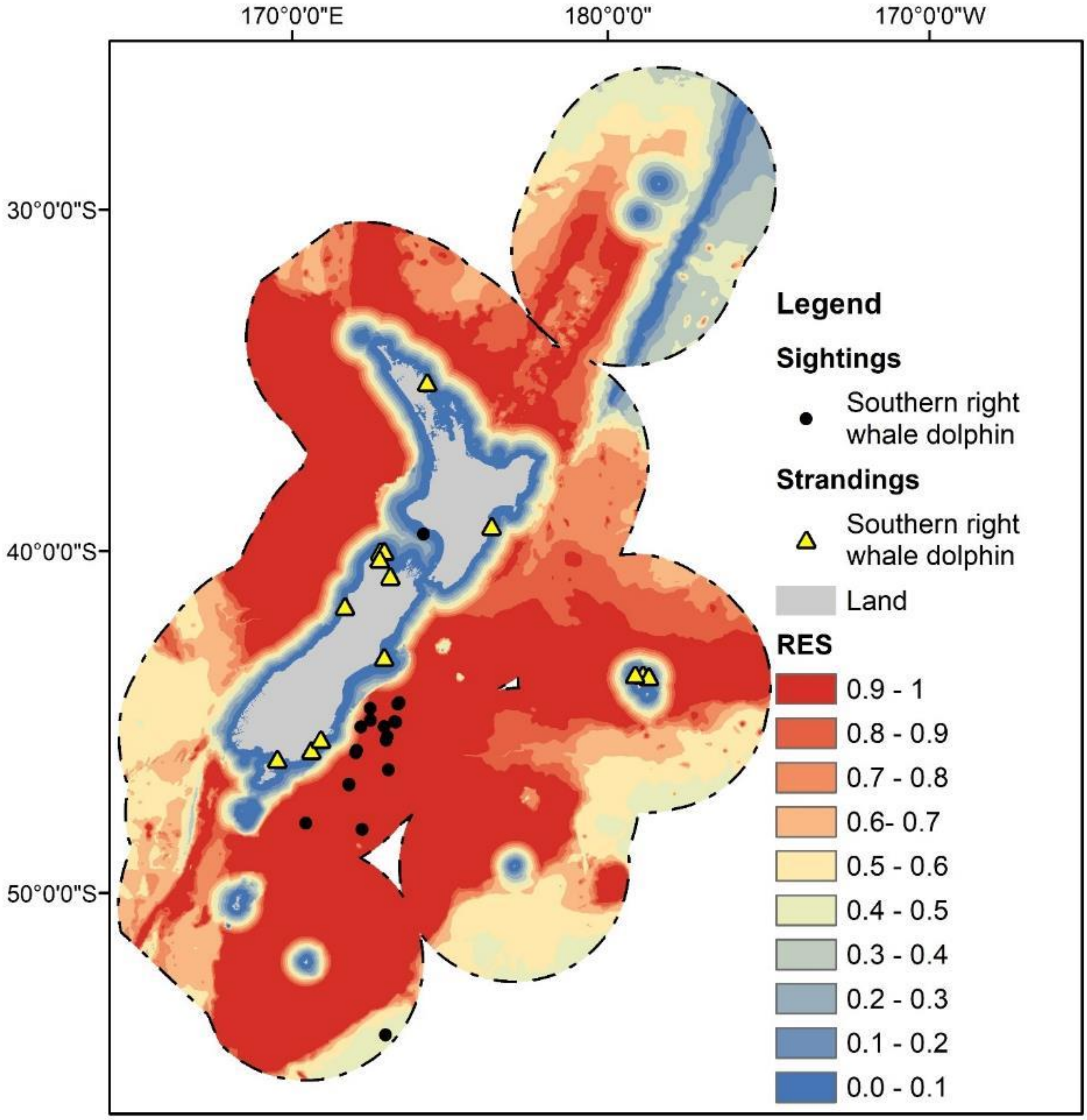

Fig A 4-26. Predicted RES scores for southern right whale dolphin, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

False killer whale (*Pseudorca crassidens*)

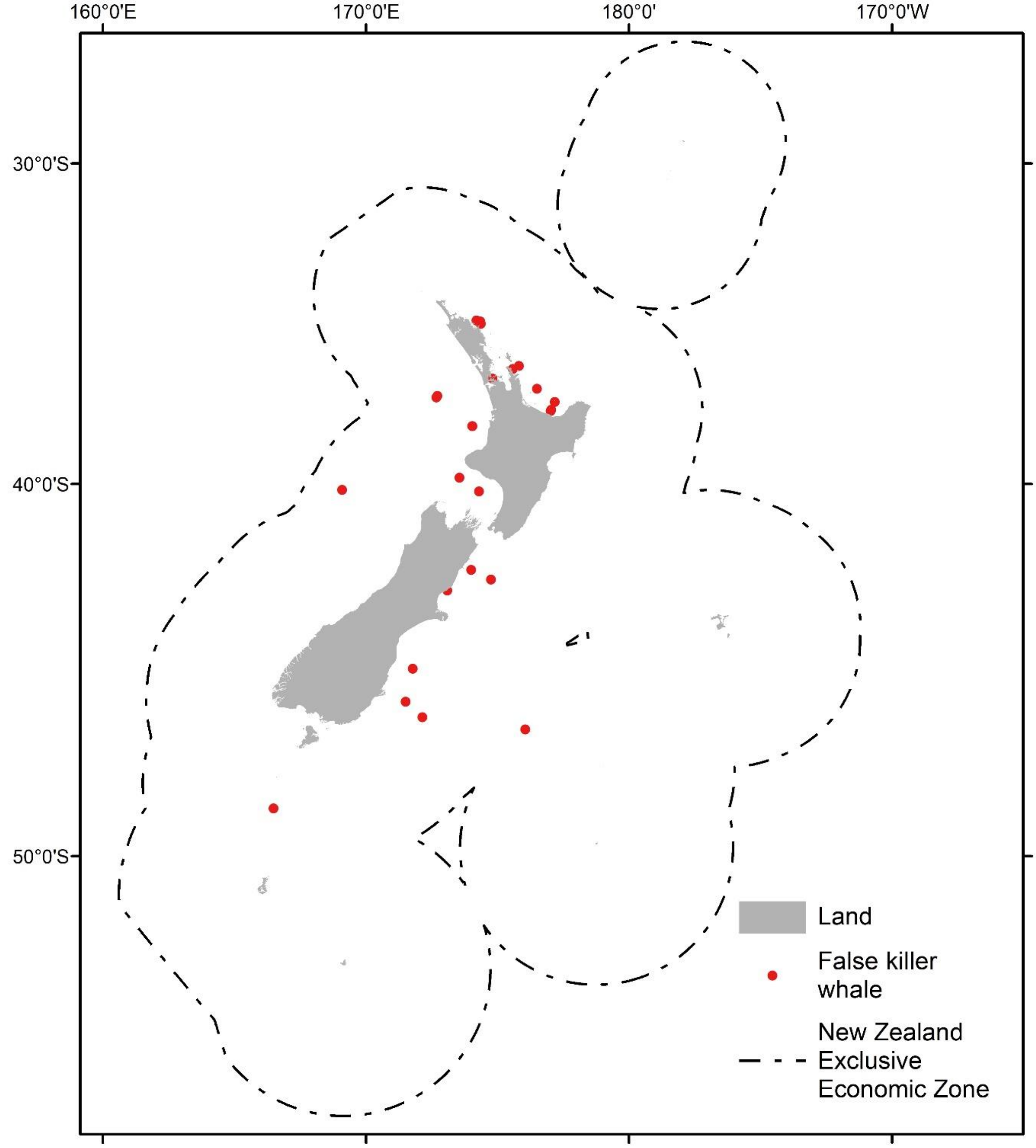

Fig A 4-27. Location of at-sea sightings of false killer whale (*n* = 28) in the NZ EEZ.

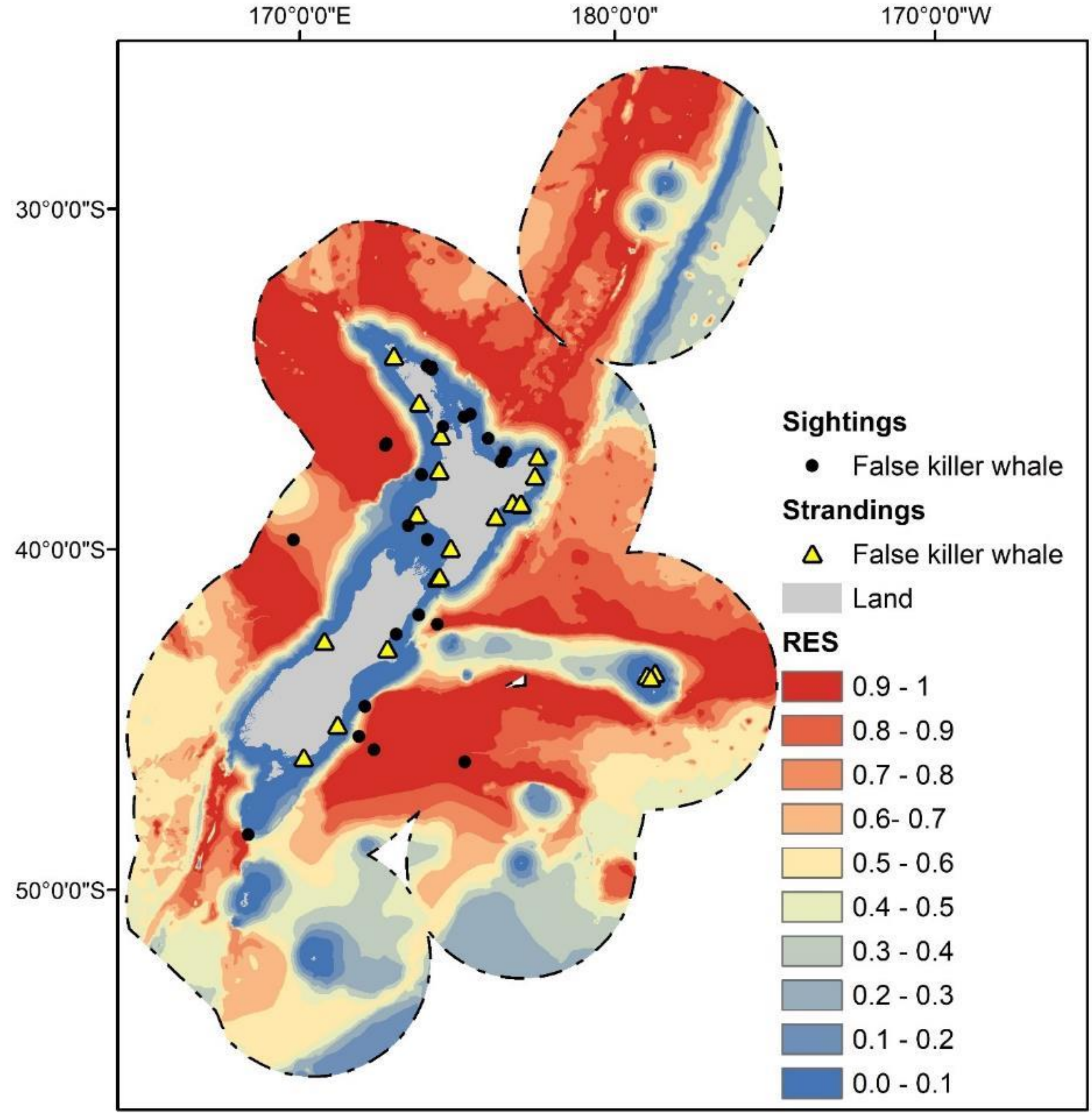

Fig A 4-28. Predicted RES scores for false killer whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Arnoux's beaked whale (*Berardius arnuxii)*

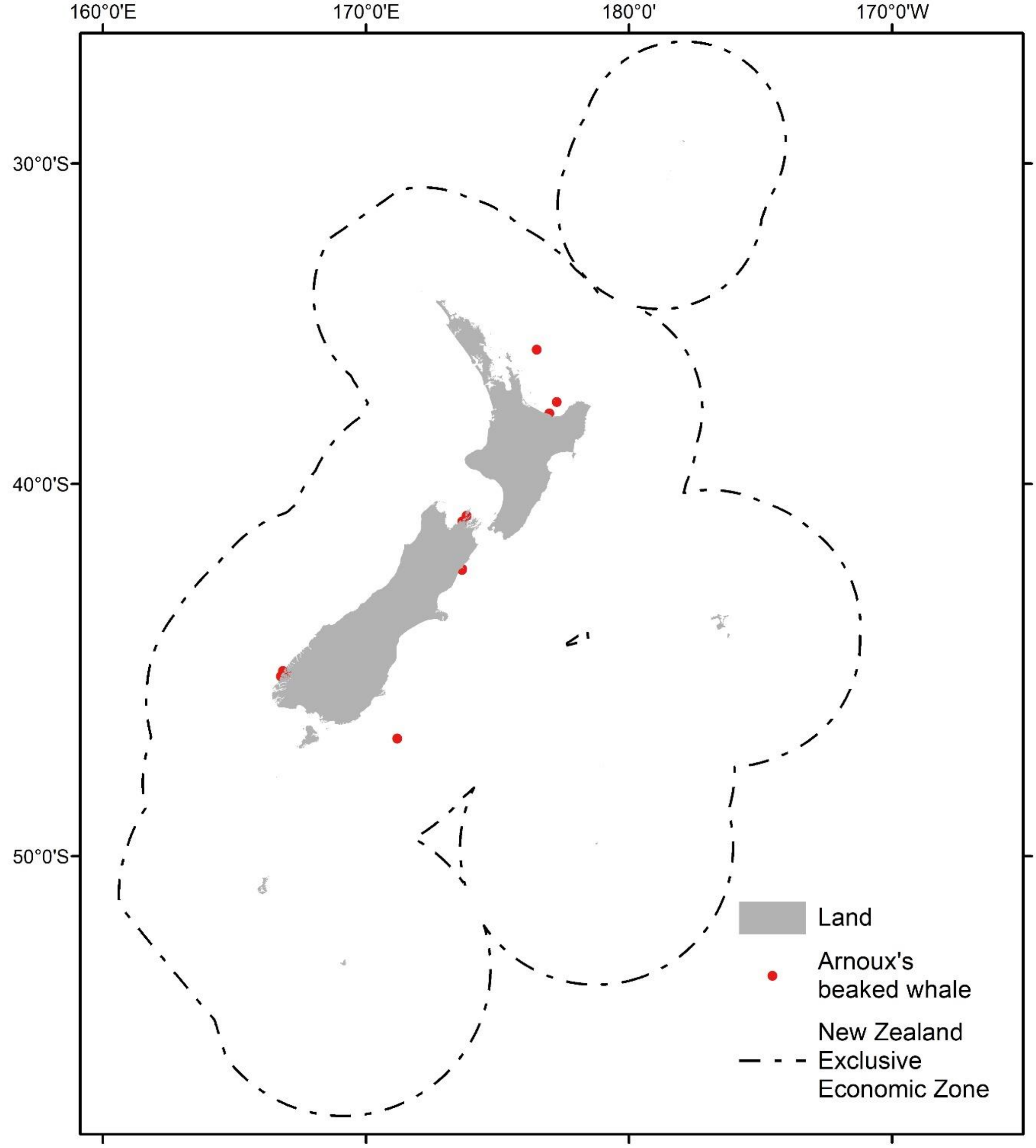

Fig A 4-29. Location of at-sea sightings of Arnoux's beaked whale (*n* = 31) in the NZ EEZ.

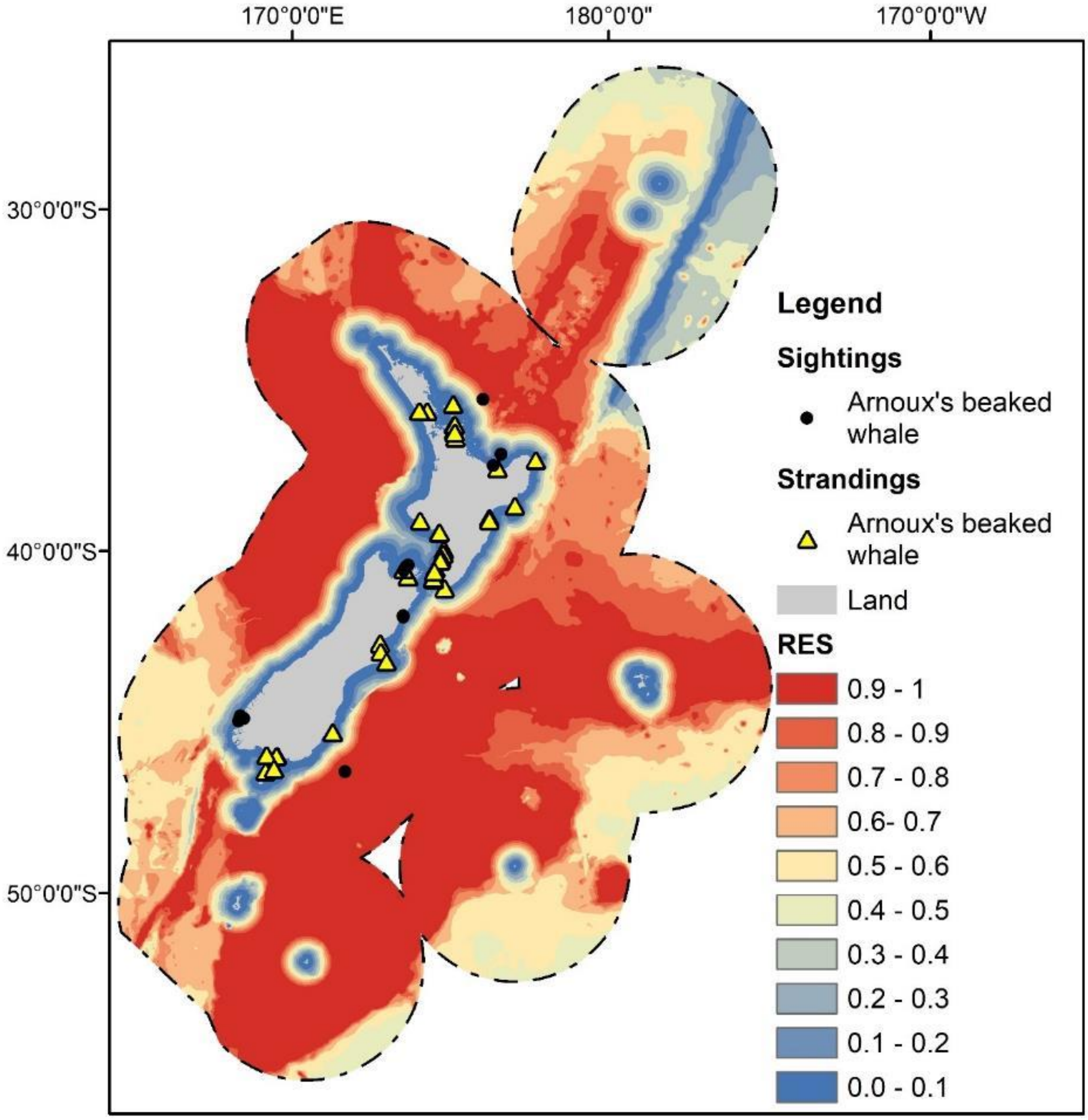

Fig A 4-30. Predicted RES scores for Arnoux's beaked whale, ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with at-sea sightings location and recorded strandings (from the DOC marine mammal strandings database).

Minke whale (*Balaenoptera acutorostrata*)

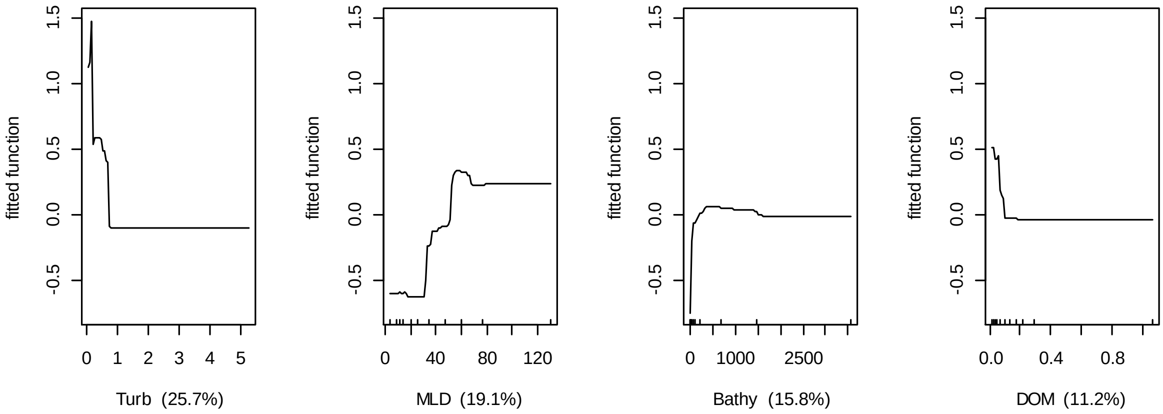

Fig A 4-31. Partial dependence plots showing the relationships between predictor variables and probability presence of minke whale modelled using BRTs. The four most influential environmental predictors in the model are shown. Quantiles of each environmental predictor are shown on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

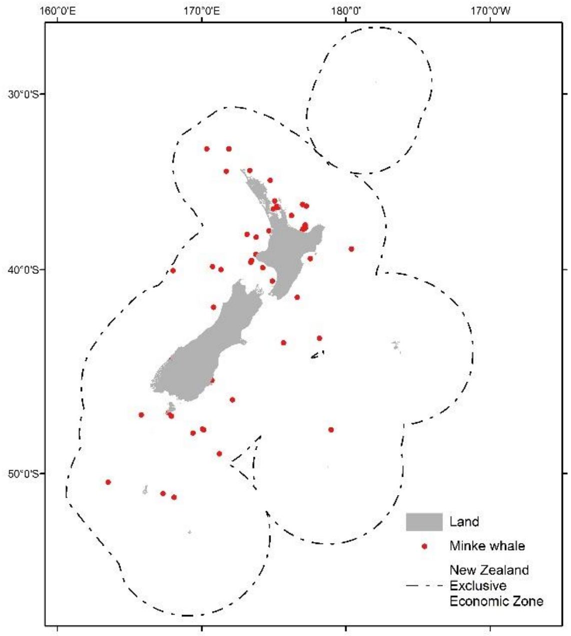

Fig A 4-32. Location of at-sea sightings of minke whale (*n* = 57) in the NZ EEZ.

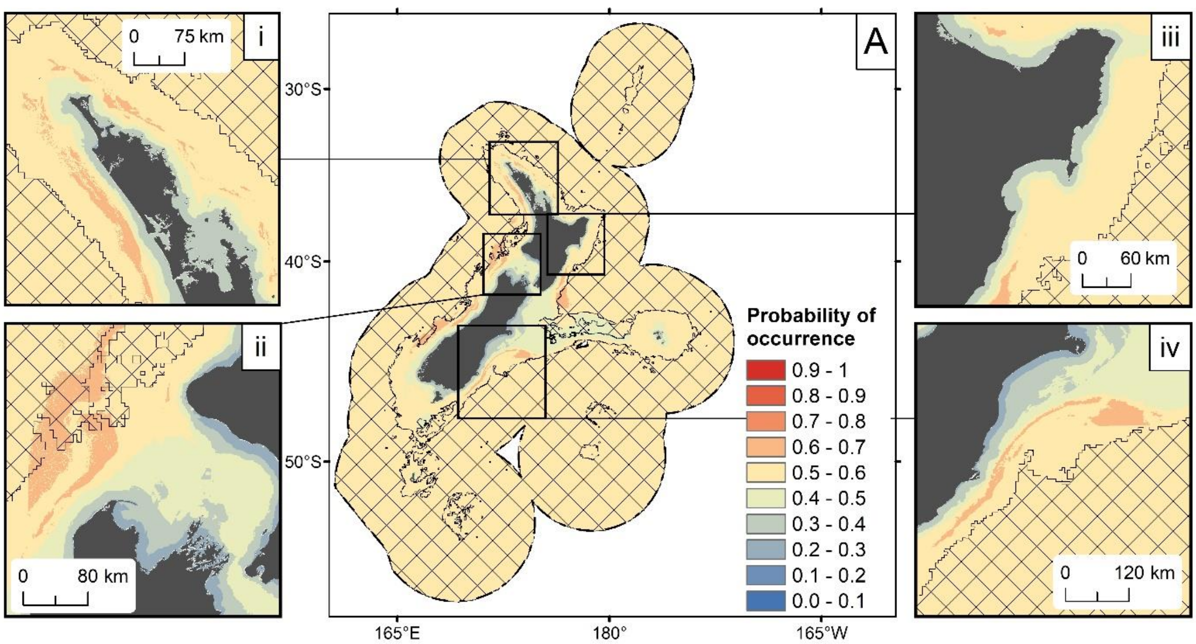

Fig A 4-33. The predicted probability occurrence of minke whale (*Balaenoptera acutorostrata*) in the New Zealand EEZ modelled using BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) south of the North Island and north of the South Island, including the Cook Strait, South Taranaki Bight and Tasman and Golden Bays; iii) central North Island including the Bay of Plenty; iv) east coast of the South Island. See figure 1 in the main body of the text for place names.

Fin whale (*Balaenoptera physalus)*

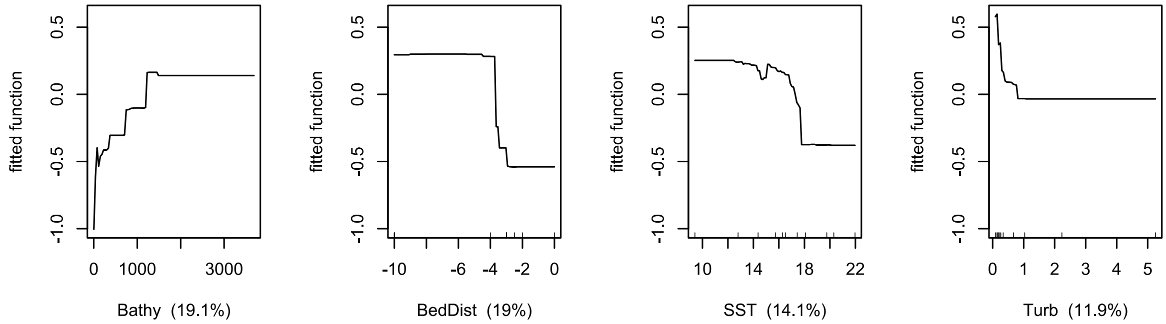

Fig A 4-34. Partial dependence plots showing the relationships between predictor variables and probability presence of fin whale modelled using BRTs. The four most influential environmental predictors in the model are shown. Quantiles of each environmental predictor are shown on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

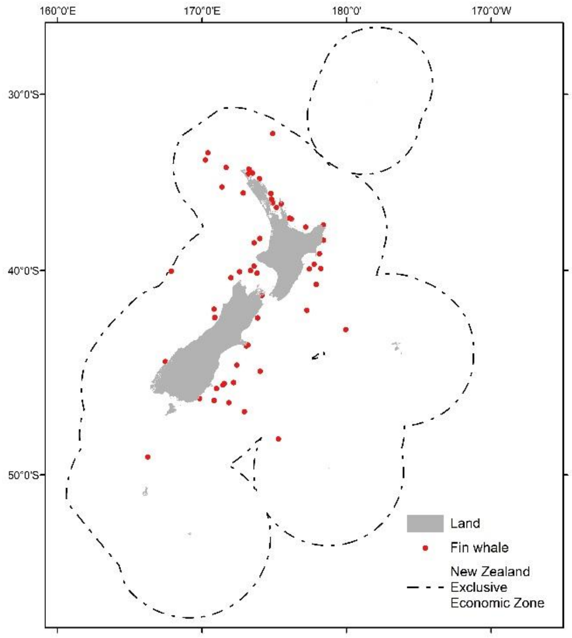

Fig A 4-35. Location of at-sea sightings of fin whale (*n* = 61) in the NZ EEZ.

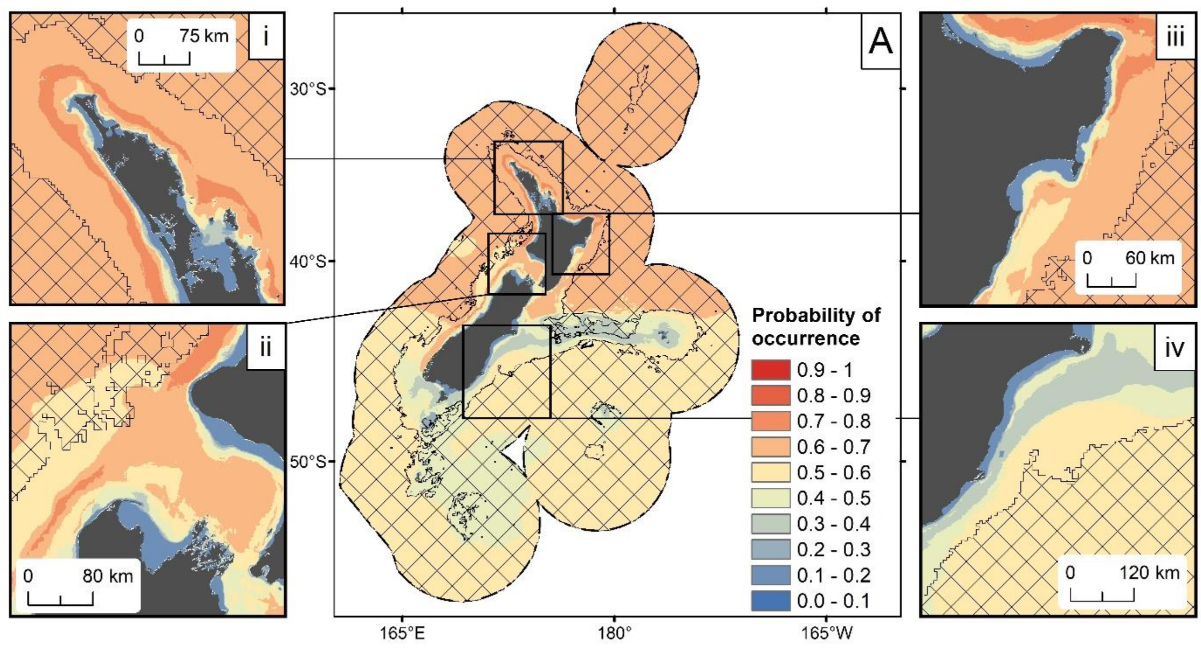

Fig A 4-36. The predicted probability occurrence of fin whale (*Balaenoptera physalus*) in the New Zealand EEZ modelled using BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) south of the North Island and north of the South Island, including the Cook Strait, South Taranaki Bight and Tasman and Golden Bays; iii) central North Island including the Bay of Plenty; iv) east coast of the South Island. See figure 1 in the main body of the text for place names.

Sei whale (*Balaenoptera borealis*)

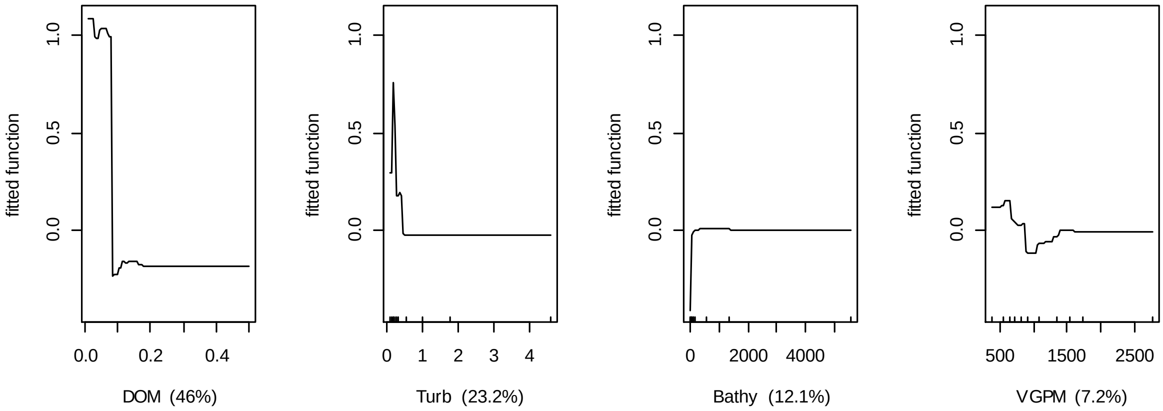

Fig A 4-37. Partial dependence plots showing the relationships between predictor variables and probability presence of sei whale modelled using BRTs. The four most influential environmental predictors in the model are shown. Quantiles of each environmental predictor are shown on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

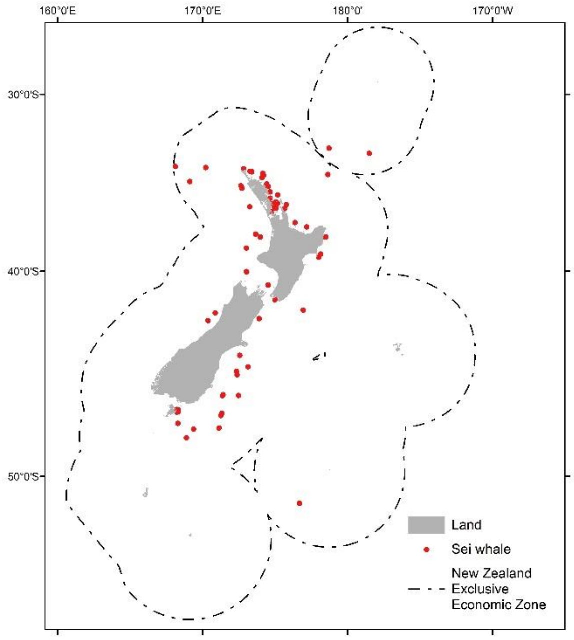

Fig A 4-38. Location of at-sea sightings of sei whale (*n* = 70) in the NZ EEZ.

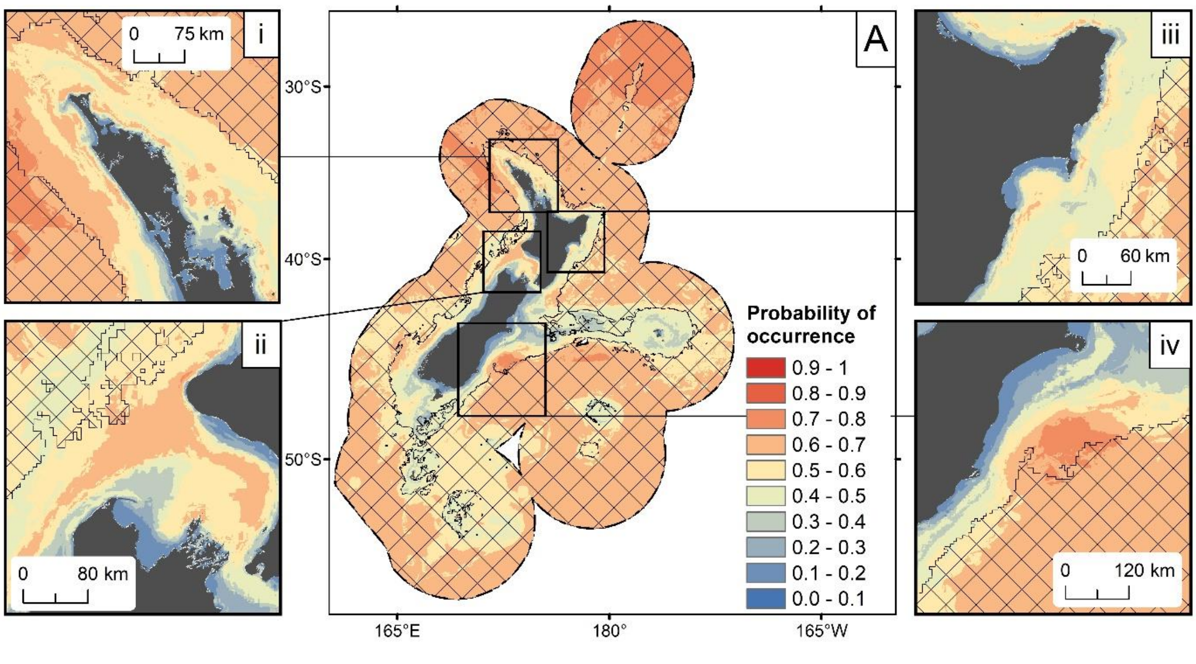

Fig A 4-39. The predicted probability occurrence of sei whale (*Balaenoptera borealis*) in the New Zealand EEZ modelled using BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) south of the North Island and north of the South Island, including the Cook Strait, South Taranaki Bight and Tasman and Golden Bays; iii) Kermadec islands, Lau-Colville Ridge and Kermadec Ridge; iv) east coast of the South Island. See figure 1 in the main body of the text for place names.

Blue whales (spp. & sub spp.) (*Balaenoptera musculus musculus* and *Balaenoptera m. brevicauda)*

The predicted probability occurrence of Blue whale (*Balaenoptera musculus)* can be found in the main body of the research article.

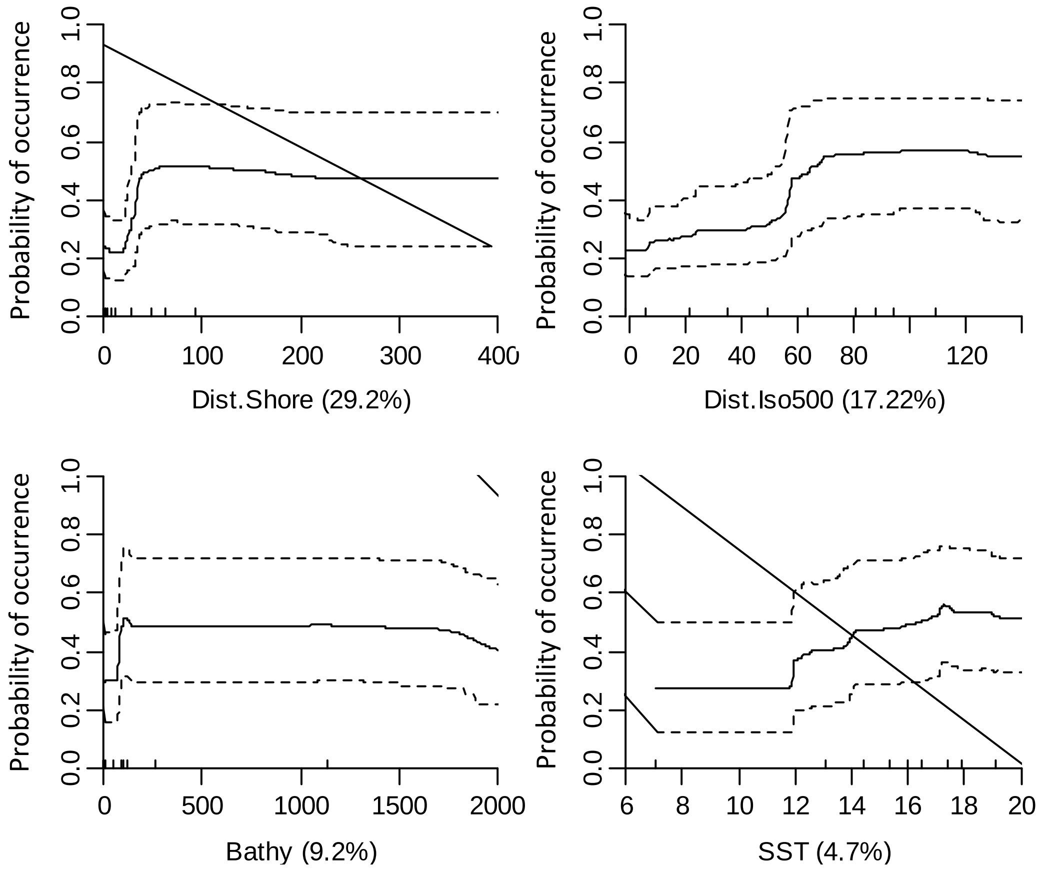

Fig A 4-40. Partial dependence plots showing the relationships between predictor variables and probability of blue whales occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

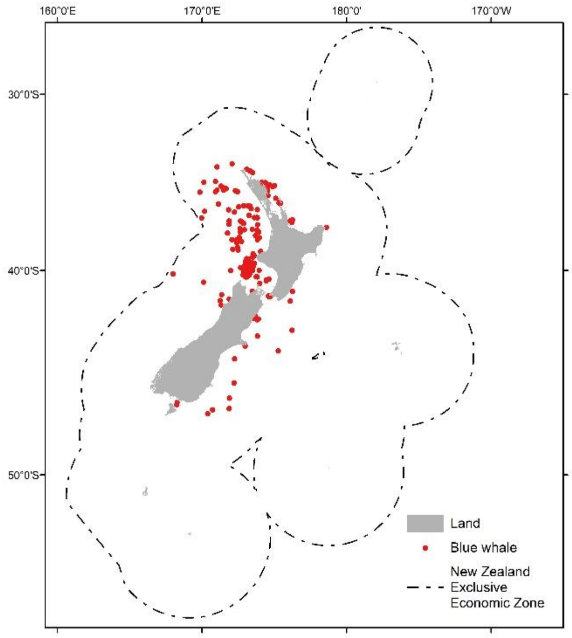

Fig A 4-41. Location of at-sea sightings of blue whale (*n* = 354) in the NZ EEZ.

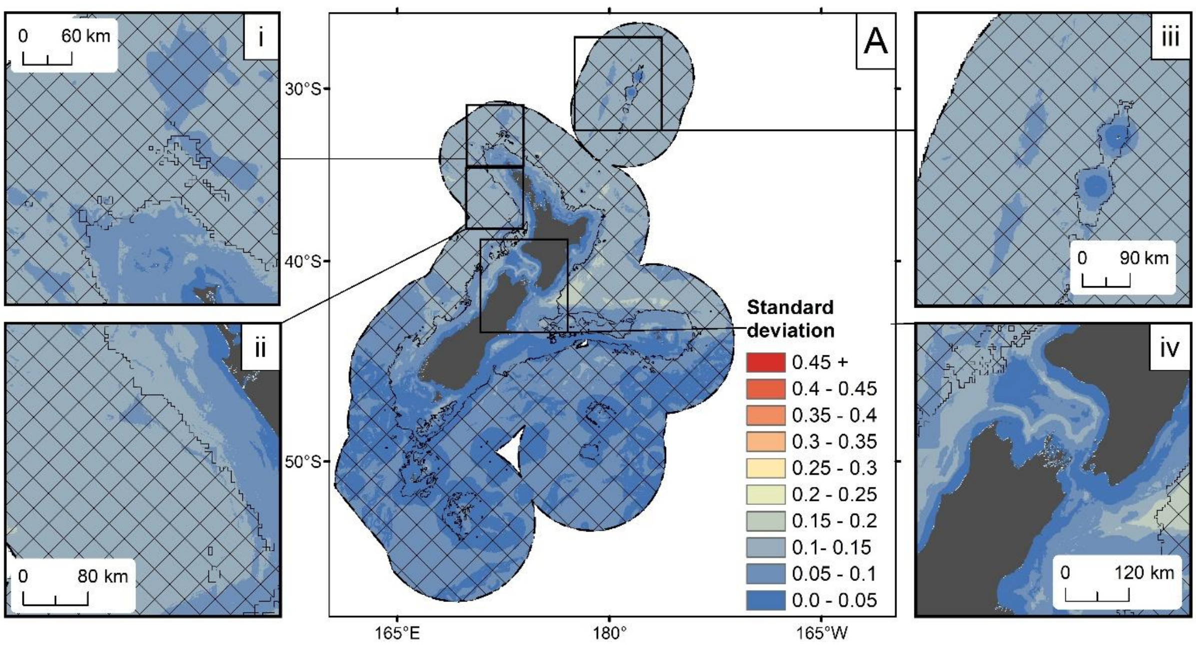

Fig A 4-42. Uncertainty estimates (standard deviation, SD) of blue whales probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and West Norfolk Ridge; ii) offshore waters of west coast of North Island including a northern section of the Challenger Plateau; iii) Kermadec islands, Lau-Colville Ridge and Kermadec Ridge; iv) south of the North Island and north of the South Island including Tasman and Golden Bays, South Taranki Bight and the Cook Strait. See figure 1 in the main body of the text for place names.

Right whale (*Eubalaena australis*)

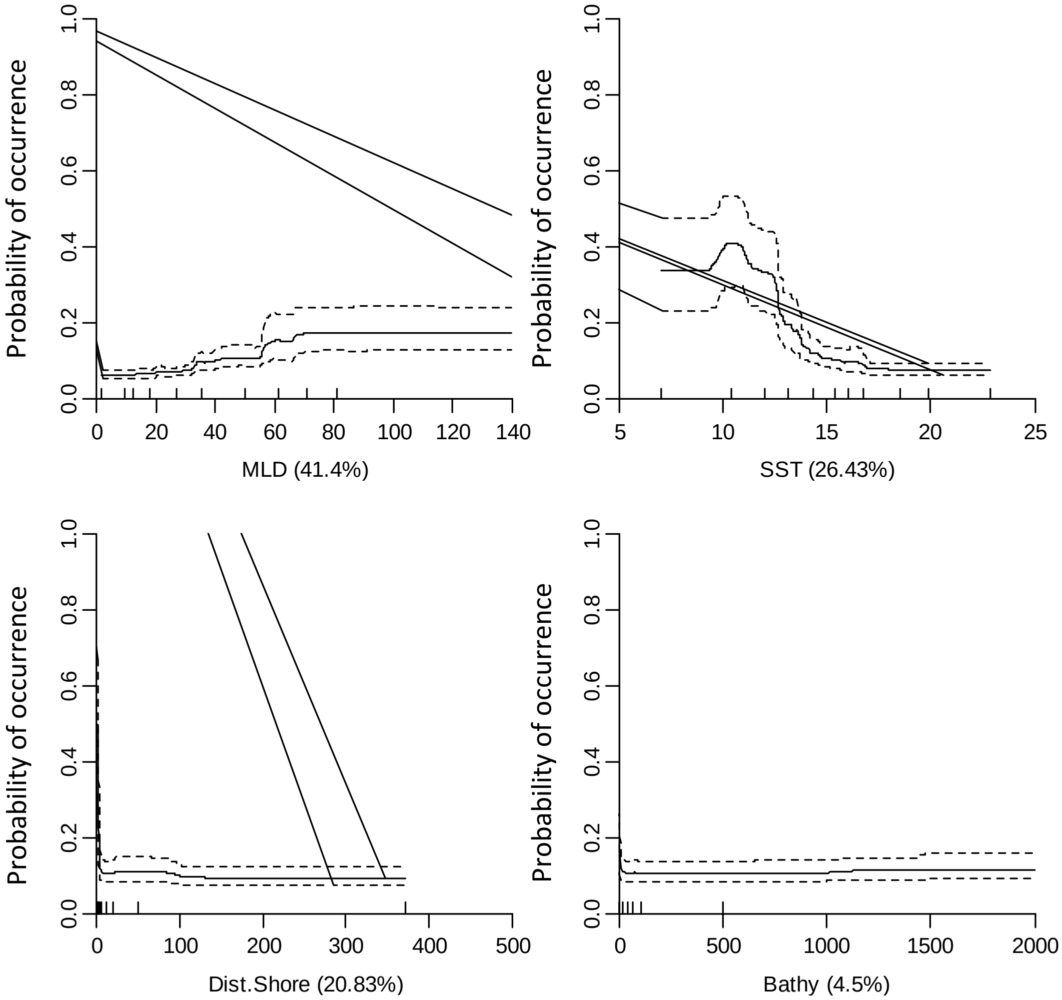

Fig A 4-43. Partial dependence plots showing the relationships between predictor variables and probability of southern right whale (*Eubalaena australis*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

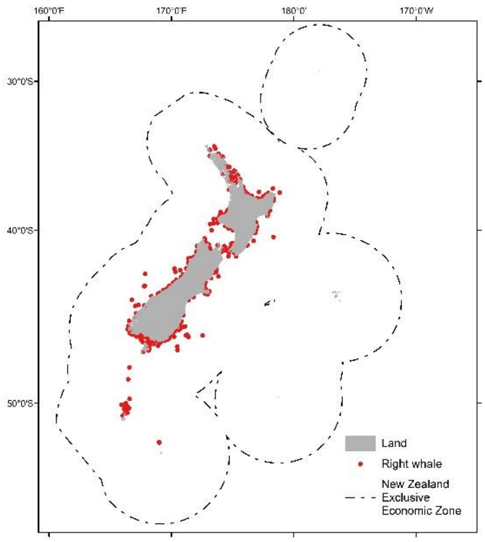

Fig A 4-44. Location of at-sea sightings of right whale (n = 477) in the NZ EEZ.

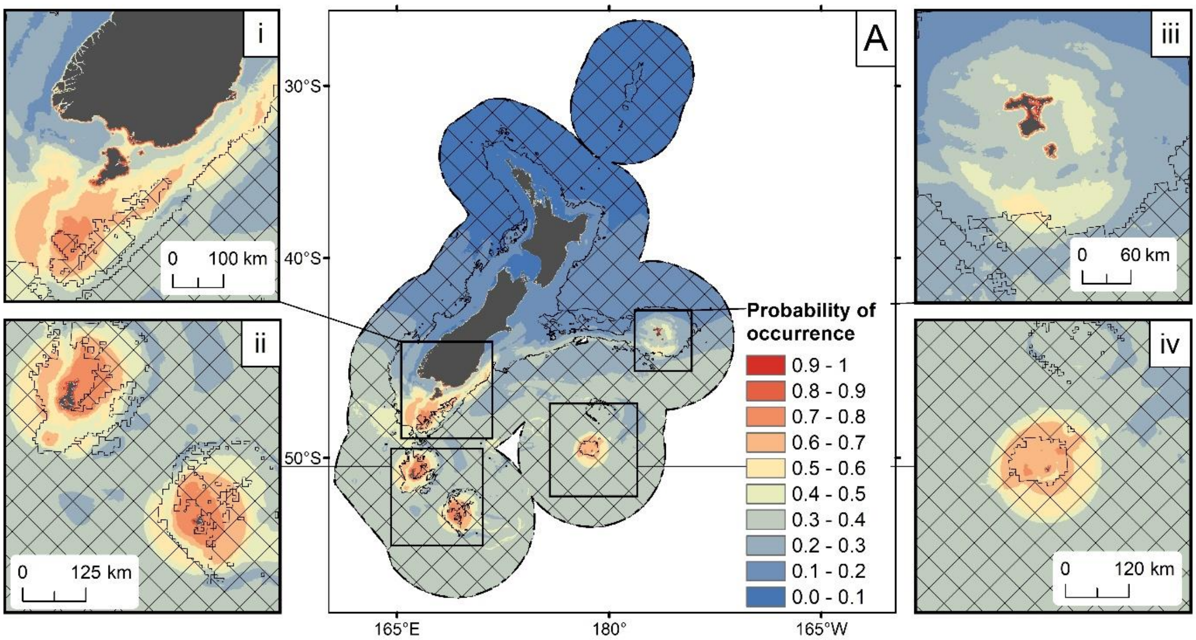

Fig A 4-45. The predicted probability occurrence of southern right whale (*Eubalaena australis*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) south of the South Island including Stewart Island / Rakiura and Campbell Plateau; ii) Auckland and Campbell Islands; iii) Chatham Island and Chatham Rise; iv) Antipodes Island and Bounty Plateau. See figure 1 in the main body of the text for place names.

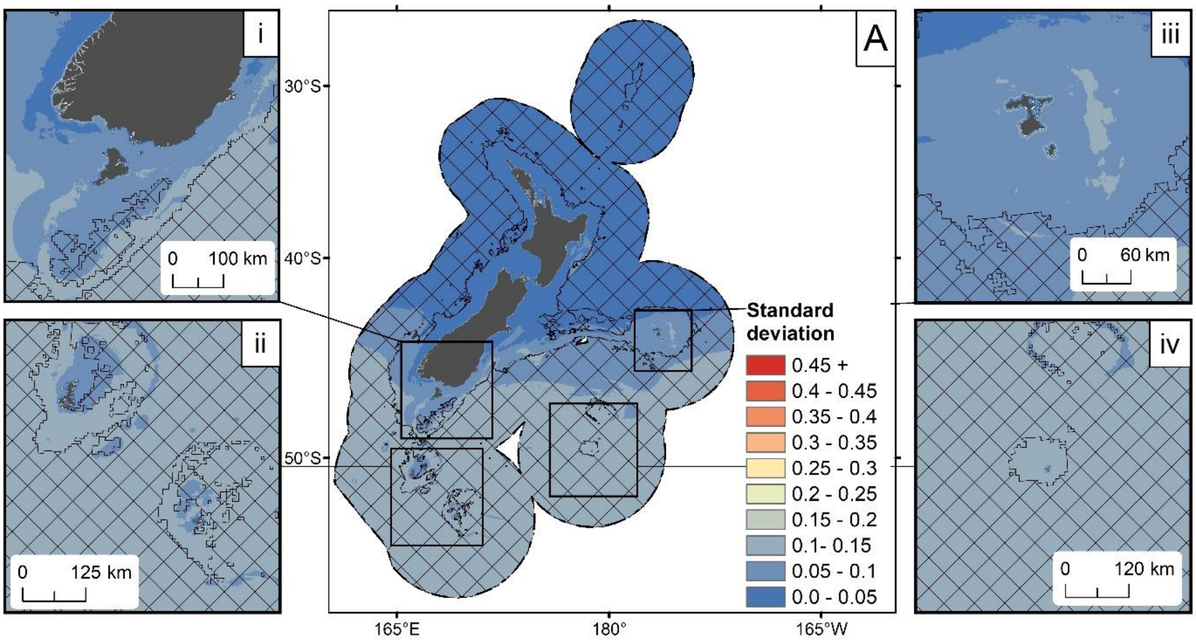

Fig A 4-46. Uncertainty estimates (standard deviation, SD) of southern right whale (*Eubalaena australis*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) south of the South Island including Stewart Island / Rakiura and Campbell Plateau; ii) Auckland and Campbell Islands; iii) Chatham Island and Chatham Rise; iv) Antipodes Island and Bounty Plateau. See figure 1 in the main body of the text for place names.

Sperm whale (*Physeter macrocephalus*)

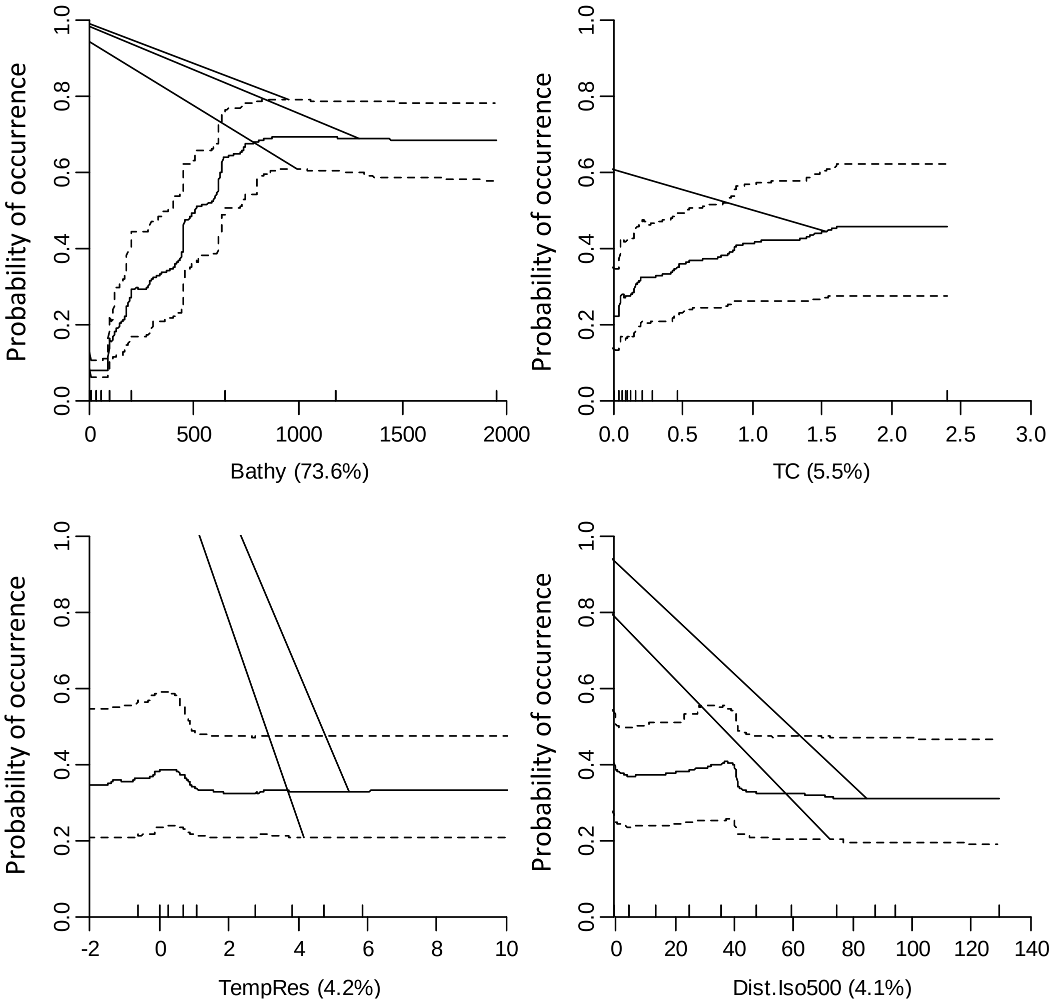

Fig A 4-47. Partial dependence plots showing the relationships between predictor variables and probability of sperm whale (*Physeter macrocephalus*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

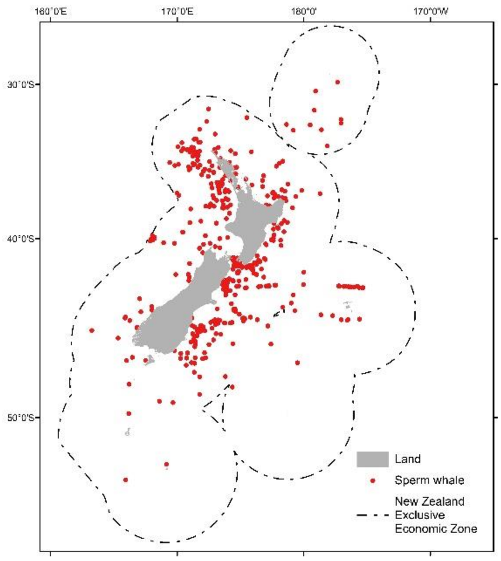

Fig A 4-48. Location of at-sea sightings of sperm whale (*n* = 497) in the NZ EEZ.

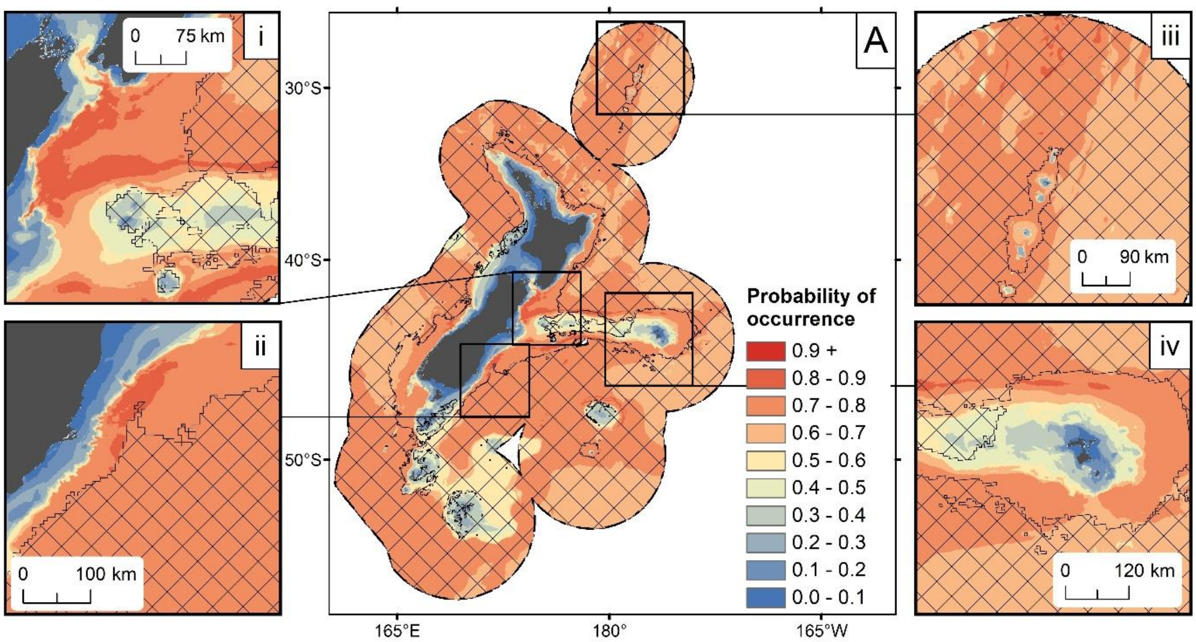

Fig A 4-49. The predicted probability occurrence of sperm whale (*Physeter macrocephalus*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) West of the Chatham Rise, Kaikoura coast and Cook strait; ii) east coast of South Island including the Canterbury Bight; iii) Kermadec islands and Kermadec Ridge; iv) east Chatham Rise and Chatham Islands. See figure 1 in the main body of the text for place names.

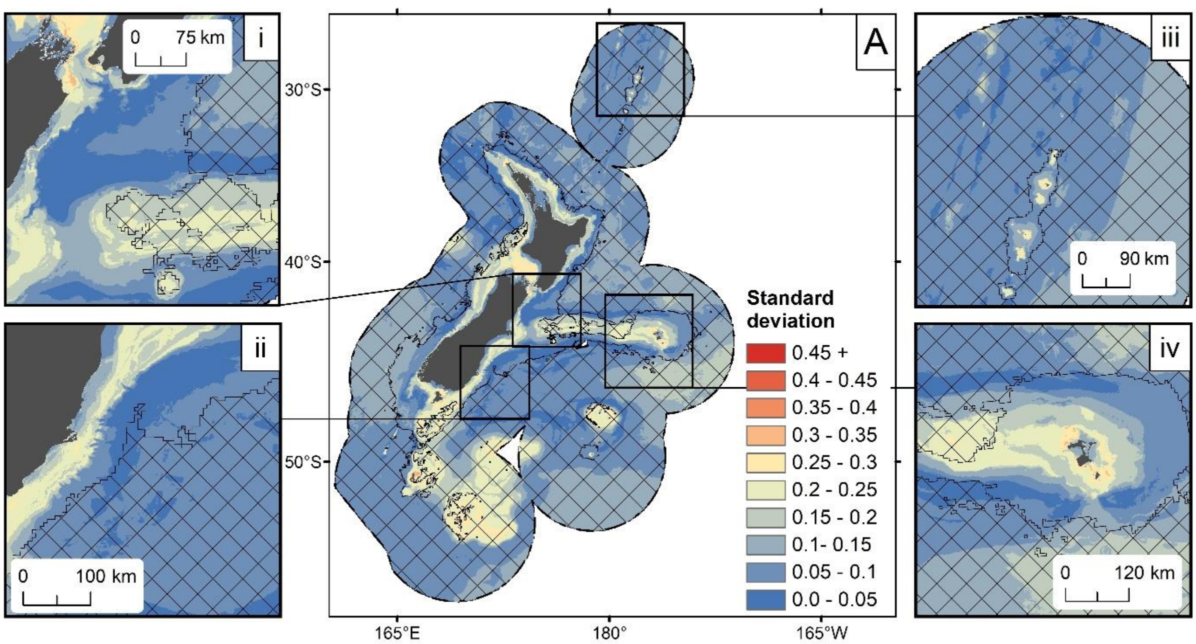

Fig A 4-50. Uncertainty estimates (standard deviation, SD) of sperm whale (*Physeter macrocephalus*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) West of the Chatham Rise, Kaikoura coast and Cook strait; ii) east coast of South Island including the Canterbury Bight; iii) Kermadec islands and Kermadec Ridge; iv) east Chatham Rise and Chatham Islands. See figure 1 in the main body of the text for place names.

Bottlenose dolphin (*Tursiops truncates*)

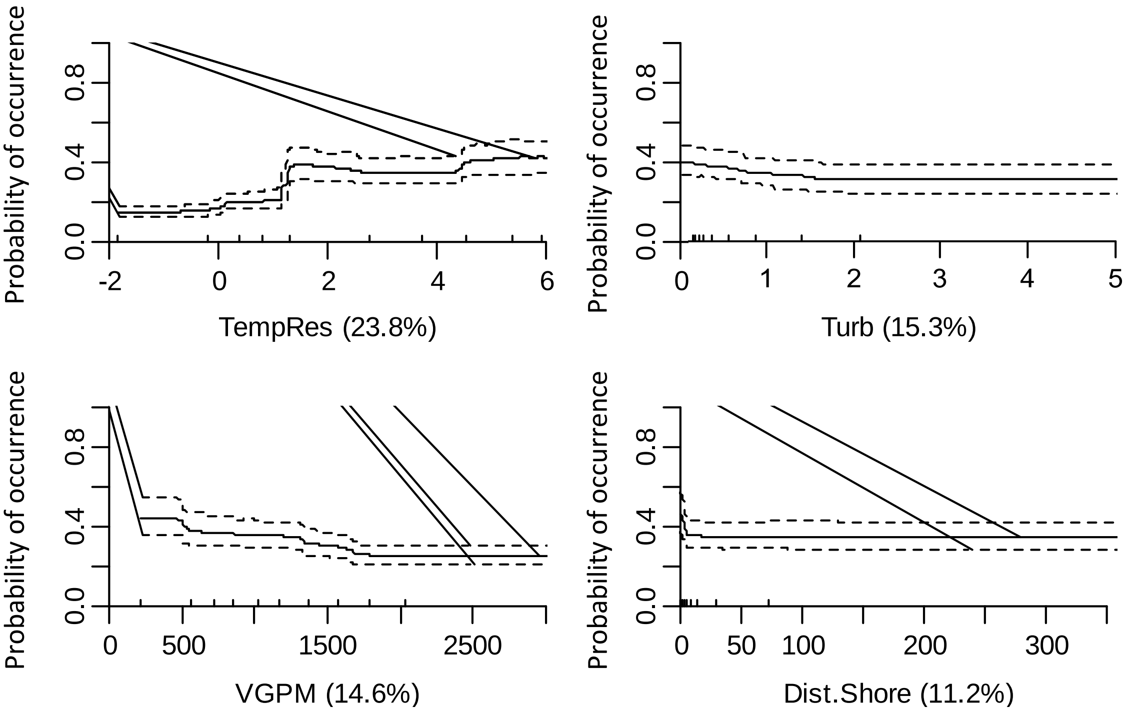

Fig A 4-51. Partial dependence plots showing the relationships between predictor variables and probability of bottlenose dolphin (*Tursiops truncates*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

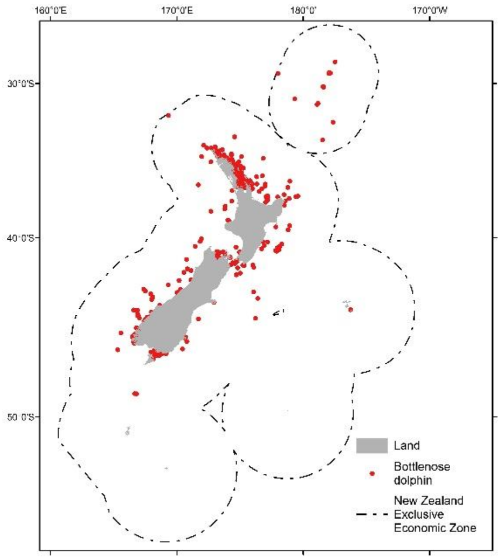

Fig A 4-52. Location of at-sea sightings of bottlenose dolphin (*n* = 498) in the NZ EEZ.

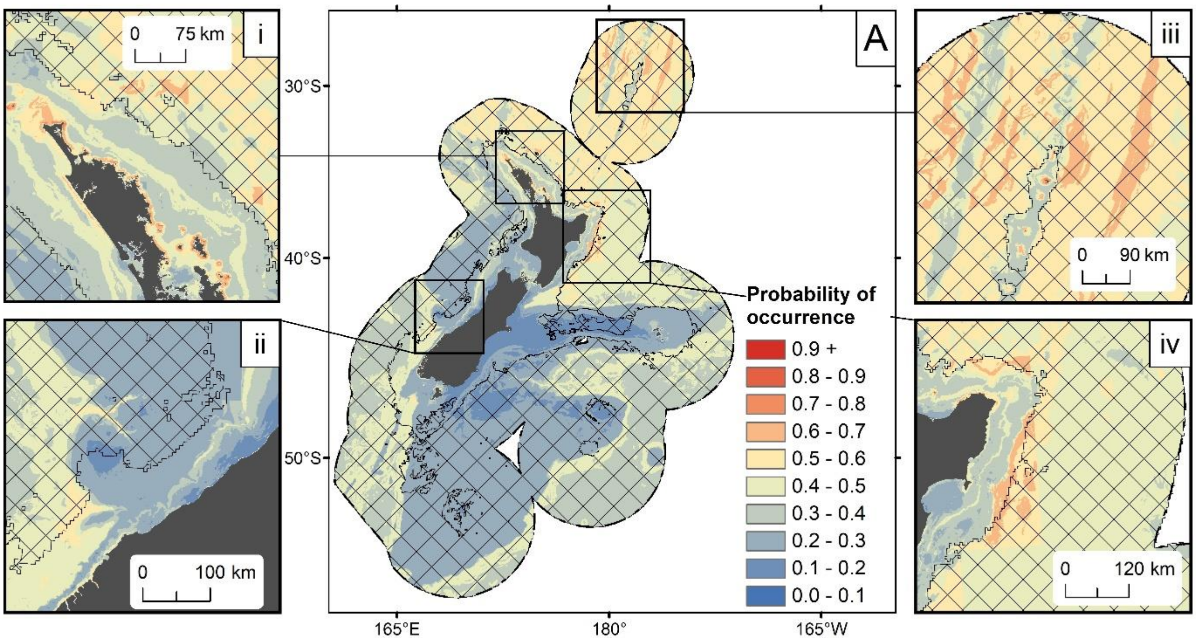

Fig A 4-53. The predicted probability occurrence of bottlenose dolphin (*Tursiops truncatus*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) west coast of South Island including the Challenger Plateau and Fiordland Coast; iii) Kermadec islands and Kermadec Ridge; iv) east of the North Island including the Bay of Plenty. See figure 1 in the main body of the text for place names.

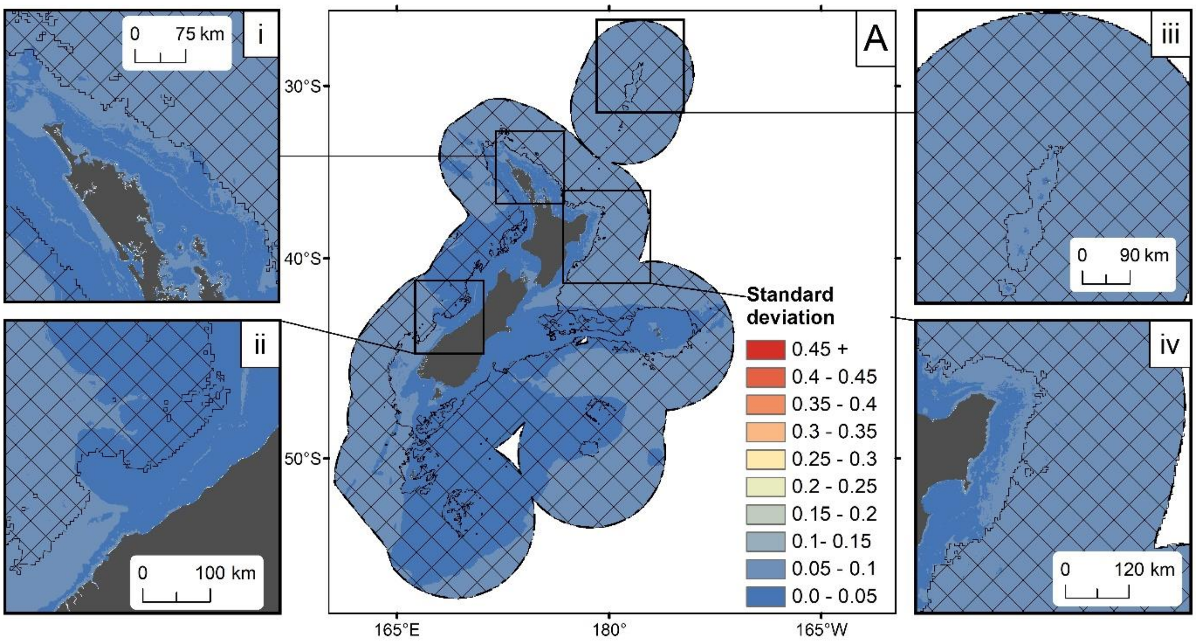

Fig A 4-54. Uncertainty estimates (standard deviation, SD) of bottlenose dolphin (*Tursiops truncatus*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) west coast of South Island including the Challenger Plateau and Fiordland Coast; iii) Kermadec islands and Kermadec Ridge; iv) east of the North Island including the Bay of Plenty. See figure 1 in the main body of the text for place names.

Killer whale (*Orcinus orca*)

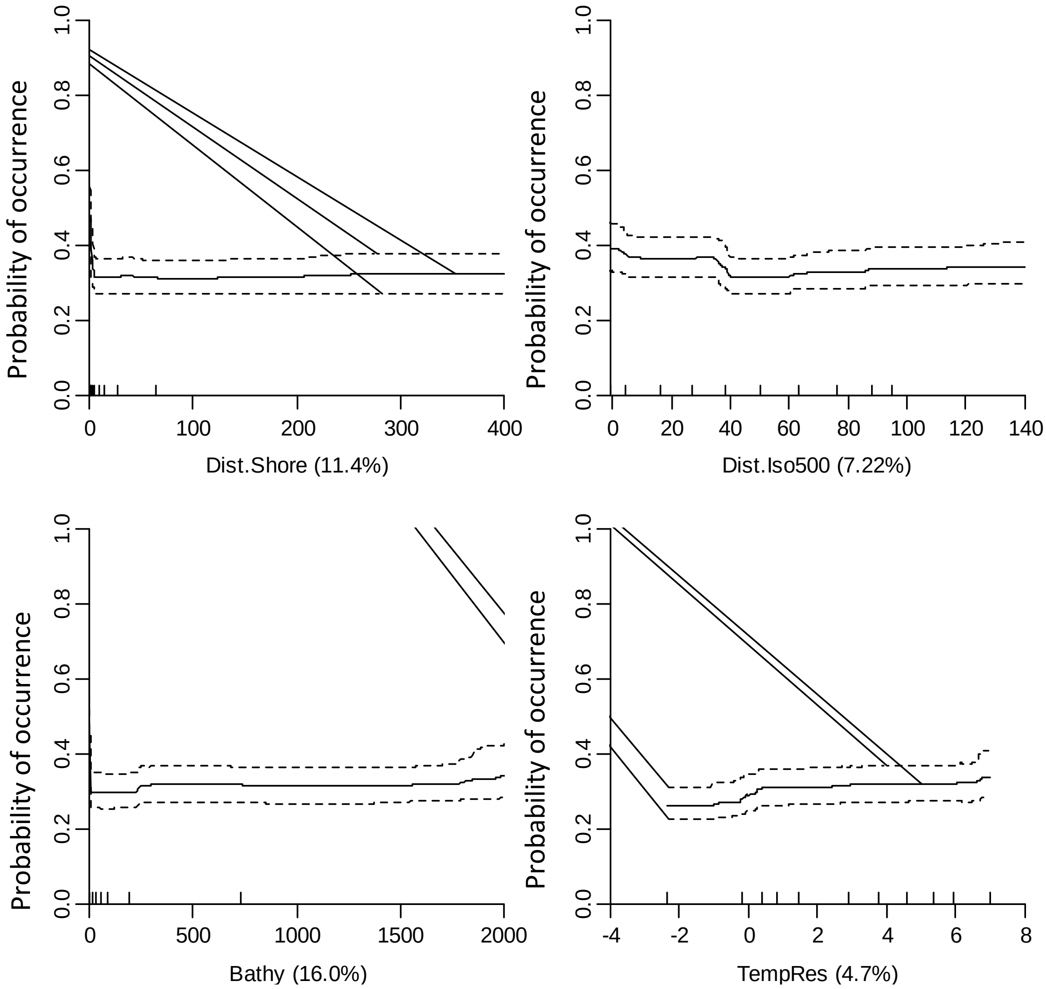

Fig A 4-55. Partial dependence plots showing the relationships between predictor variables and probability of killer whale (*Orcinus orca*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

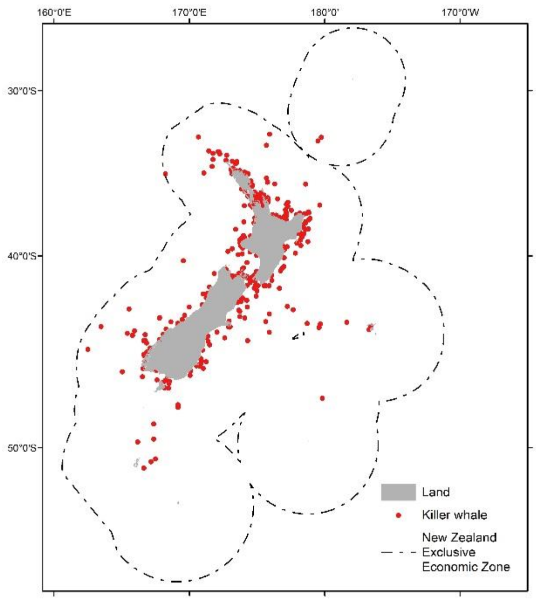

Fig A 4-56. Location of at-sea sightings of killer whale (n = 569) in the NZ EEZ.

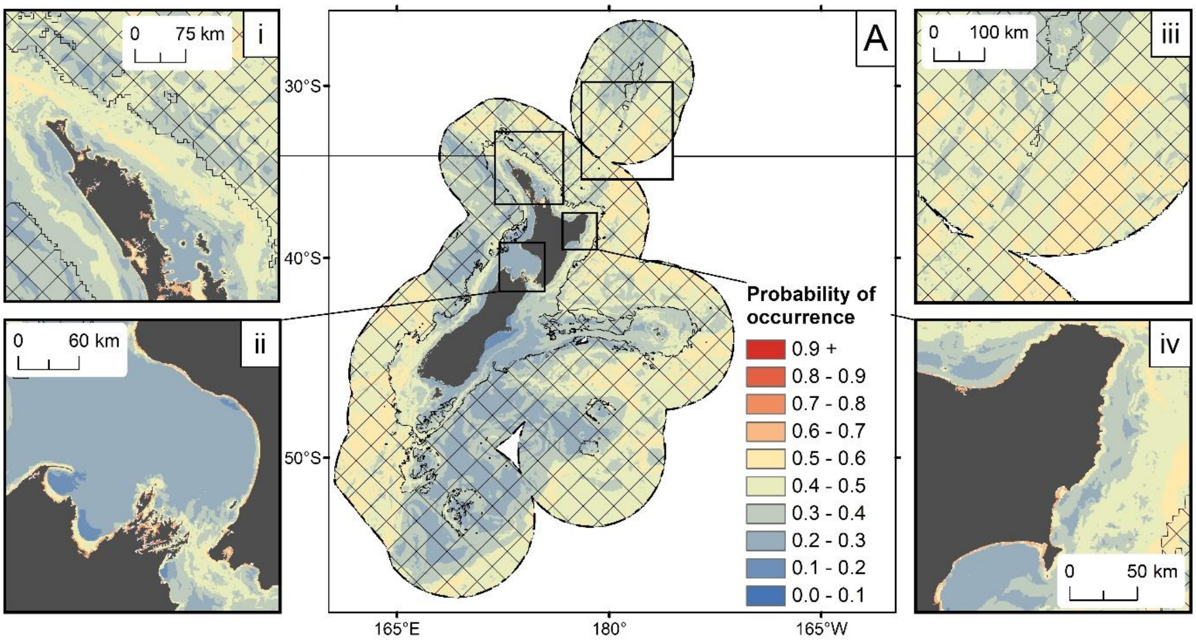

Fig A 4-57. The predicted probability occurrence of killer whale (*Orcinus orca*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) south of the North Island and north of the South Island including Tasman and Golden Bays, South Taranki Bight and the Cook Strait; iii) Kermadec islands and Kermadec Ridge; iv) central North Island including the Bay of Plenty and Hawke's Bay. See figure 1 in the main body of the text for place names.

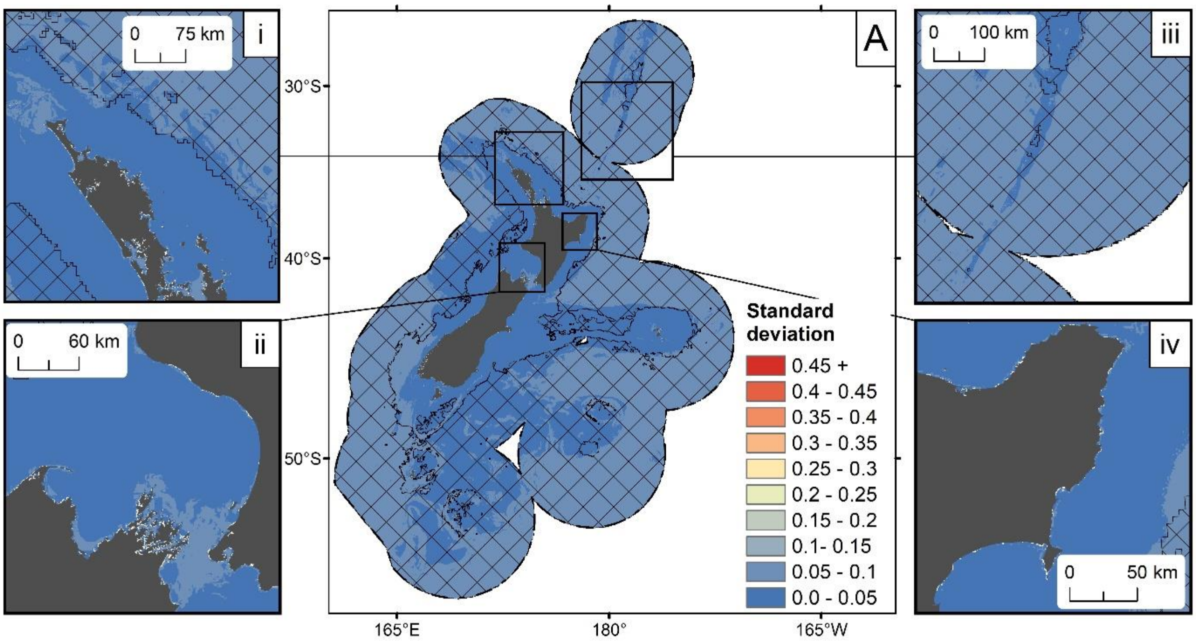

Fig A 4-58. Uncertainty estimates (standard deviation, SD) of killer whale (*Orcinus orca*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) south of the North Island and north of the South Island including Tasman and Golden Bays, South Taranki Bight and the Cook Strait; iii) Kermadec islands and Kermadec Ridge; iv) central North Island including the Bay of Plenty and Hawke's Bay. See figure 1 in the main body of the text for place names.

Bryde's whale (*Balaenoptera edeni brydei*)

The predicted probability occurrence of Bryde's whale (*Balaenoptera edeni brydei*) can be found in the main body of the research article.

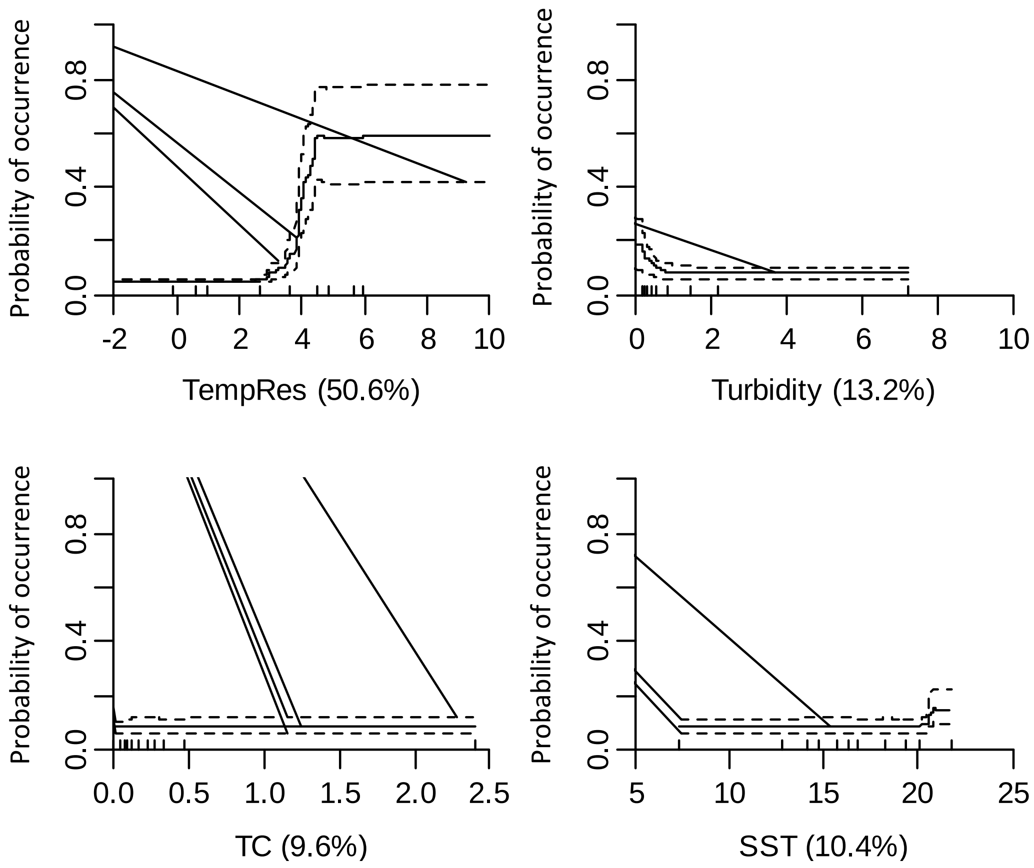

Fig A 4-59. Partial dependence plots showing the relationships between predictor variables and probability of Bryde's whale (*Balaenoptera edeni brydei*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

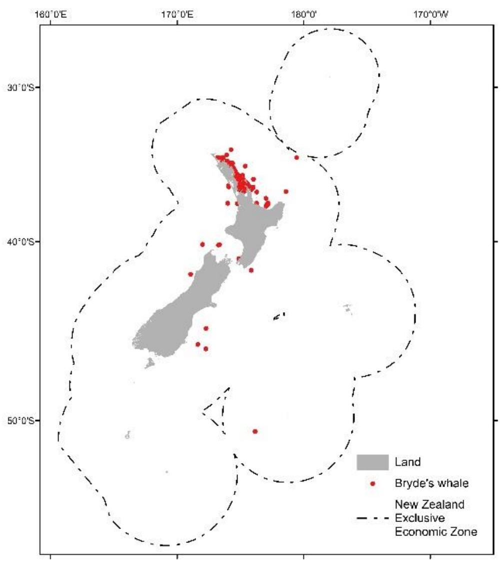

Fig A 4-60. Location of at-sea sightings of Bryde's whale (*n* = 593) in the NZ EEZ.

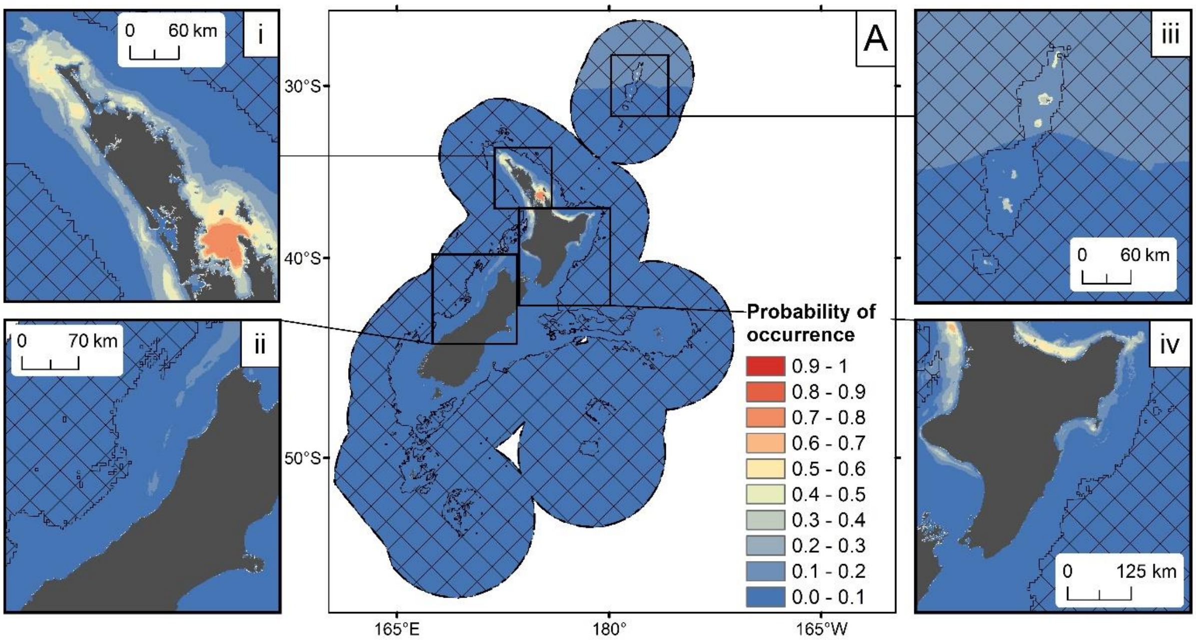

Fig A 4-61. The predicted probability occurrence of Bryde's whales (*Balaenoptera edeni brydei*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) west coast of South Island including the Fiordland Coast; iii) Kermadec islands and Kermadec Ridge; iv) central North Island including the Bay of Plenty, South Taranaki Bight and Cook Strait. See figure 1 in the main body of the text for place names.

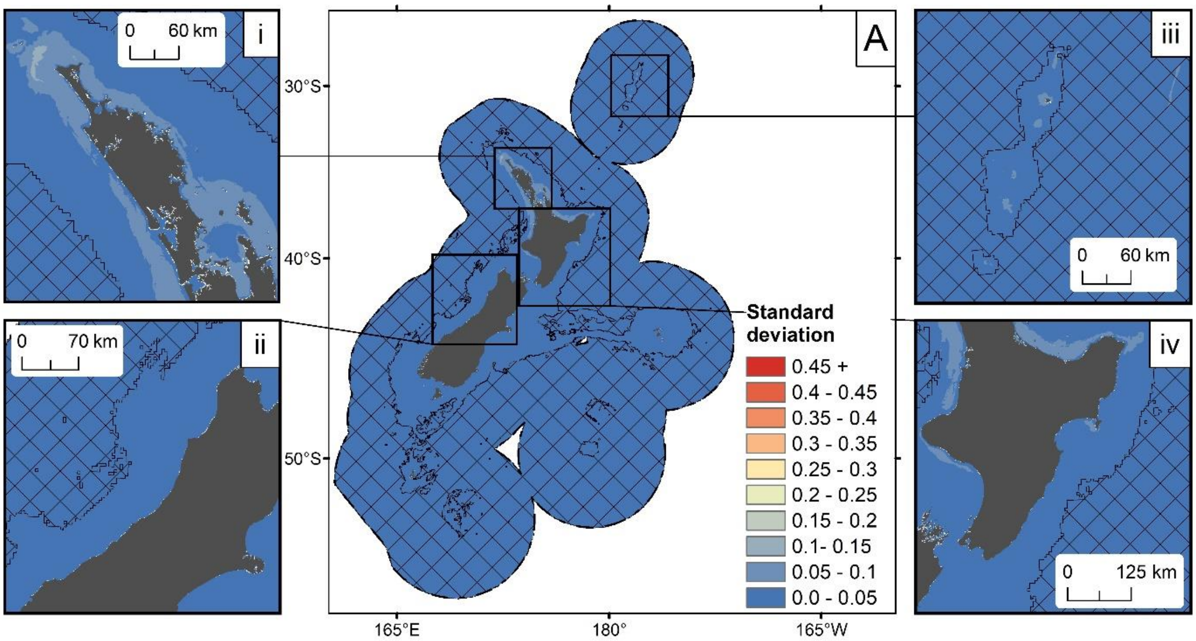

Fig A 4-62. Uncertainty estimates (standard deviation, SD) of Bryde's whale (*Balaenoptera brydei*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) west coast of South Island including the Fiordland Coast; iii) Kermadec islands and Kermadec Ridge; iv) central North Island including the Bay of Plenty, South Taranaki Bight and Cook Strait. See figure 1 in the main body of the text for place names.

Humpback whale (*Megaptera novaeangliae*)

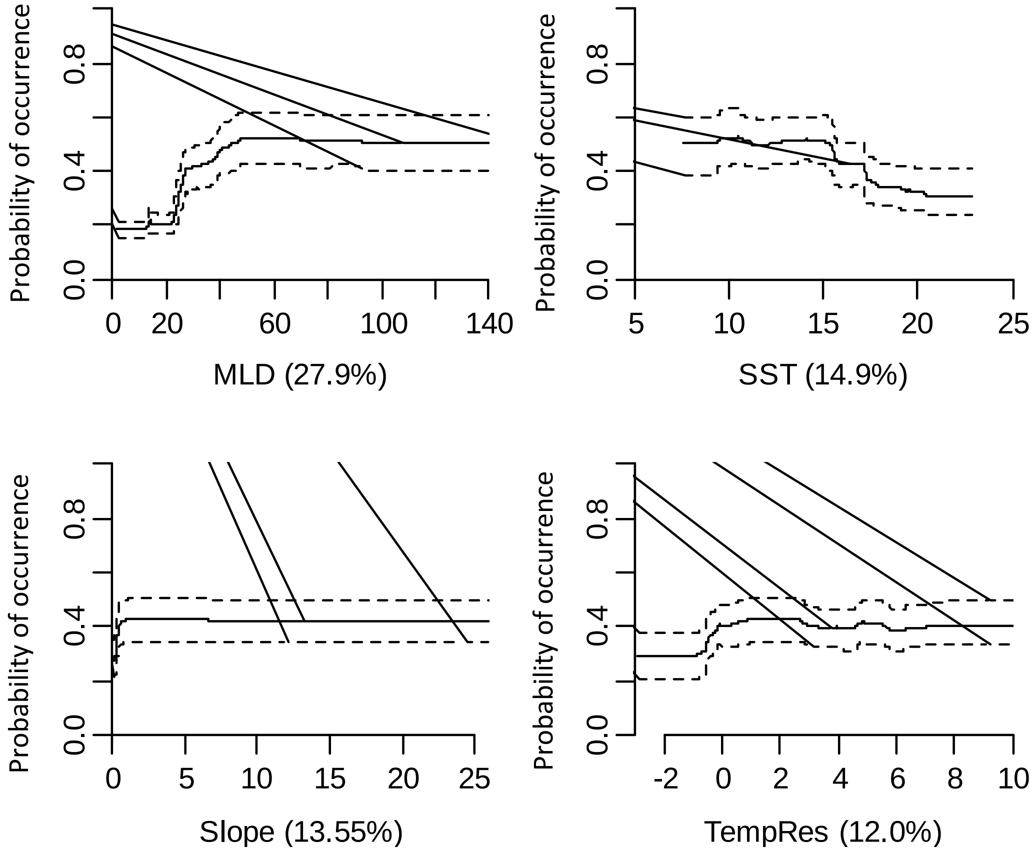

Fig A 4-63. Partial dependence plots showing the relationships between predictor variables and probability of humpback whale (*Megaptera novaeangliae*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

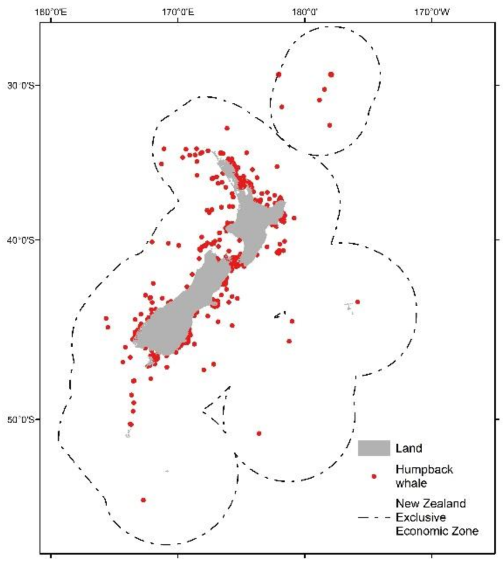

Fig A 4-64. Location of at-sea sightings of humpback whale (*n* = 629) in the NZ EEZ.

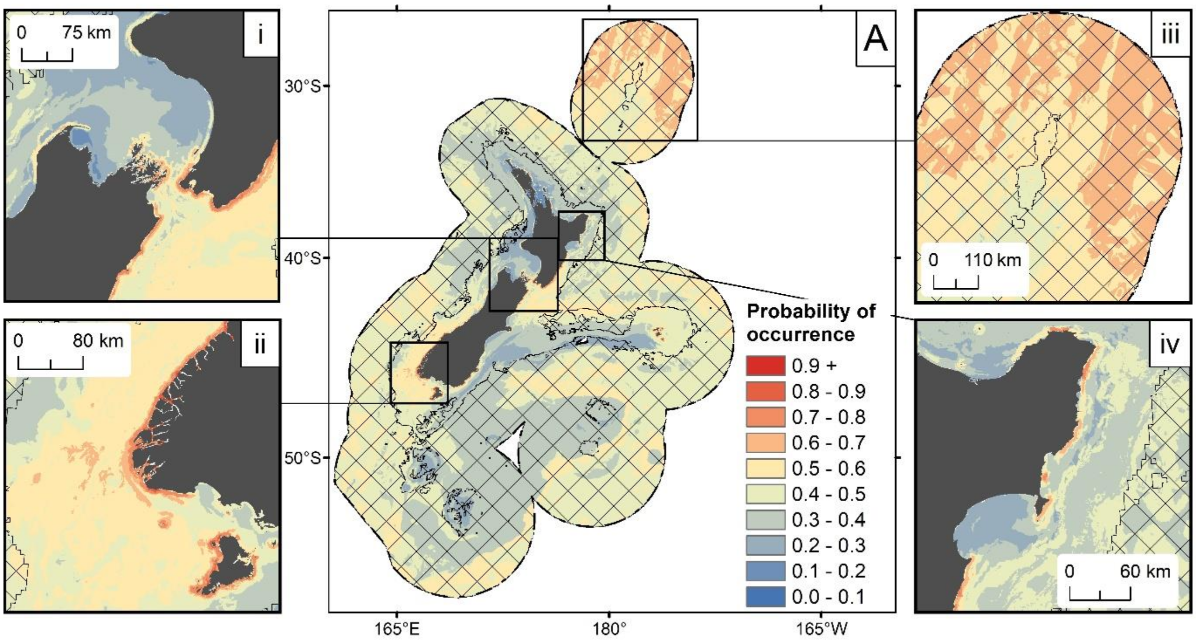

Fig A 4-65. The predicted probability occurrence of humpback whale (*Megaptera novaeangliae*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) south of the North Island and north of the South Island including the South Taranaki Bight, Tasman and Golden Bays and Cook Strait; ii) west coast of South Island including the Fiordland Coast and Stewart Island / Rakiura; iii) Lau-Colville and Kermadec Ridges and Kermadec islands; iv) central North Island including the Bay of Plenty and Hawke's Bay. See figure 1 in the main body of the text for place names.

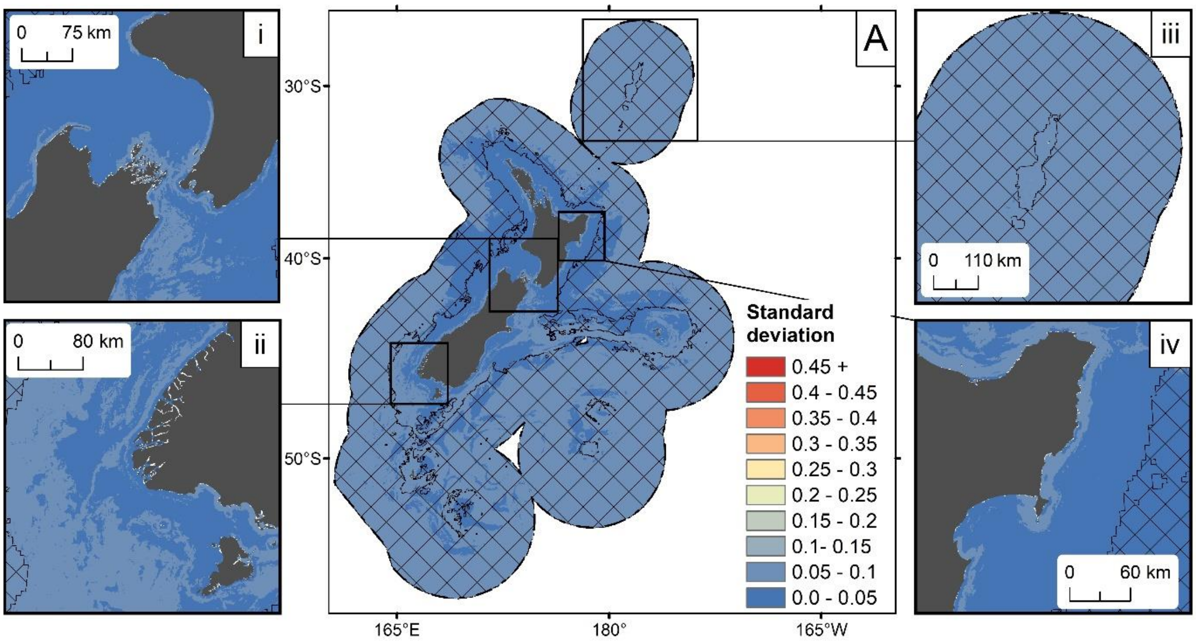

Fig A 4-66. Uncertainty estimates (standard deviation, SD) of humpback whale (*Megaptera novaeangliae*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) south of the North Island and north of the South Island including the South Taranaki Bight, Tasman and Golden Bays and Cook Strait; ii) west coast of South Island including the Fiordland Coast and Stewart Island / Rakiura; iii) Lau-Colville and Kermadec Ridges and Kermadec islands; iv) central North Island including the Bay of Plenty and Hawke's Bay. See figure 1 in the main body of the text for place names.

Pilot whales (2 spp.) (*Globicephala melas* & *Globicephala macrorhynchus*)

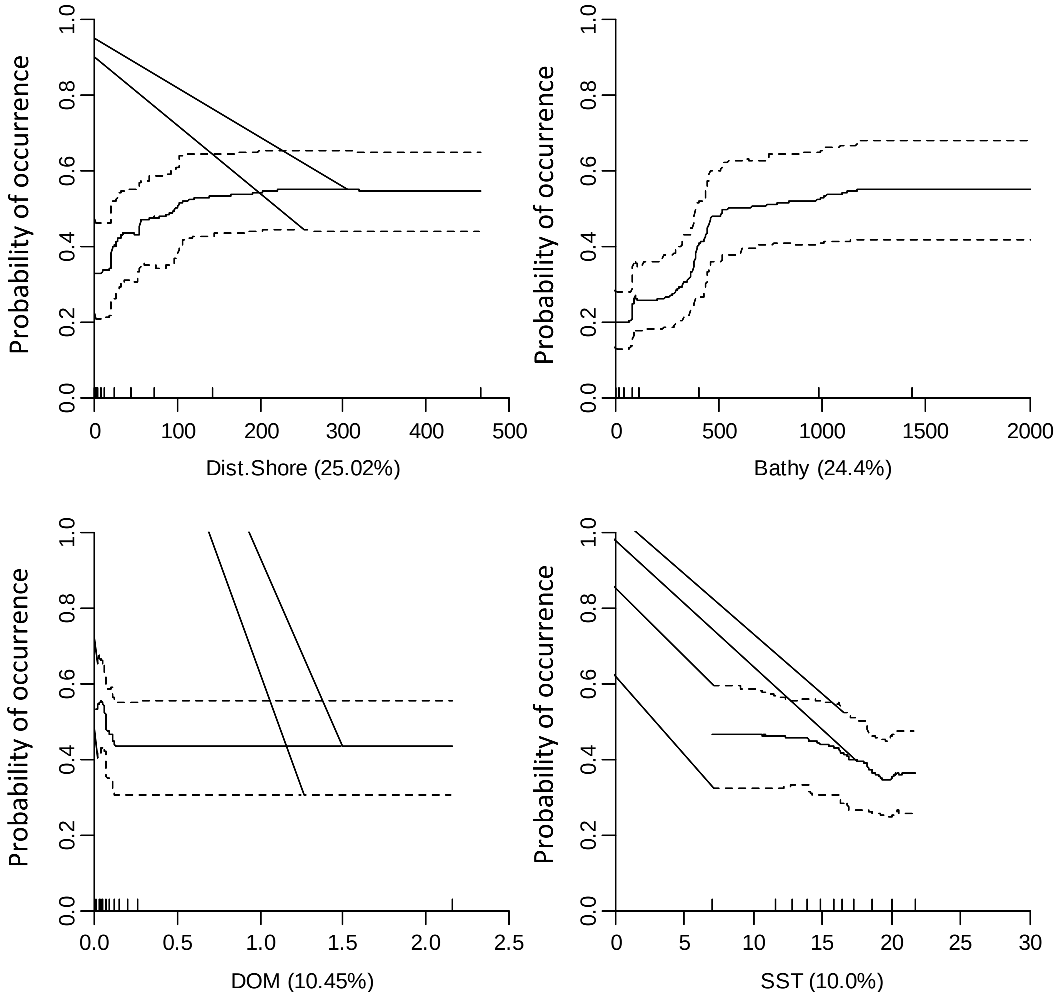

Fig A 4-67. Partial dependence plots showing the relationships between predictor variables and probability of pilot whales (*Globicephala melas* & *Globicephala macrorhynchus*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

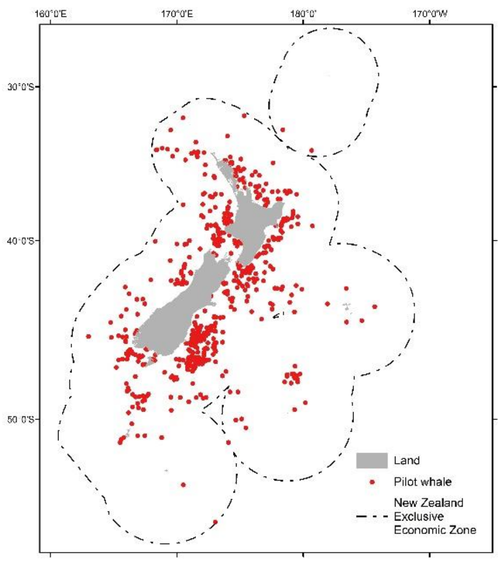

Fig A 4-68. Location of at-sea sightings of pilot whales (n = 679) in the NZ EEZ.

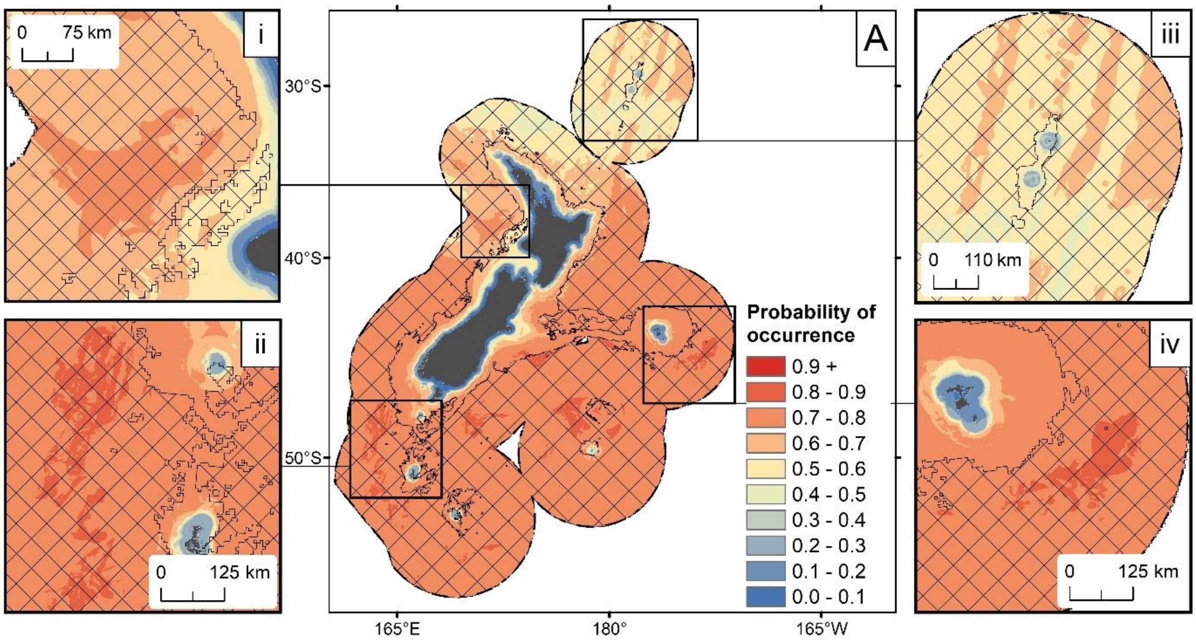

Fig A 4-69. The predicted probability occurrence of pilot whales (*Globicephala melas* & *Globicephala macrorhynchus*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) west of North Island including the Challenger Plateau; ii) North East Island, Auckland Island and Puysegur Trench; iii) Kermadec islands and Lau-Colville and Kermadec Ridges; iv) east of the Chatham Rise and Chatham Islands. See figure 1 in the main body of the text for place names.

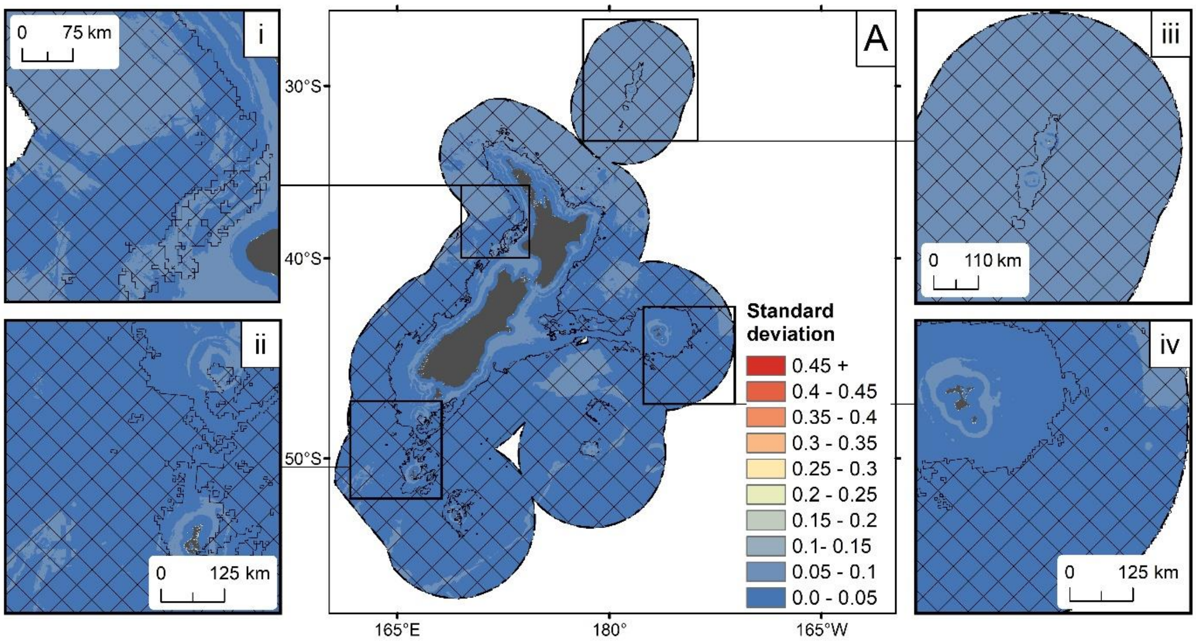

Fig A 4-70. Uncertainty estimates (standard deviation, SD) of pilot whales (*Globicephala melas* & *Globicephala macrorhynchus*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) west of North Island including the Challenger Plateau; ii) North East Island, Auckland Island and Puysegur Trench; iii) Kermadec islands and Lau-Colville and Kermadec Ridges; iv) east of the Chatham Rise and Chatham Islands. See figure 1 in the main body of the text for place names.

Dusky dolphin (*Lagenorhynchus obscurus)*

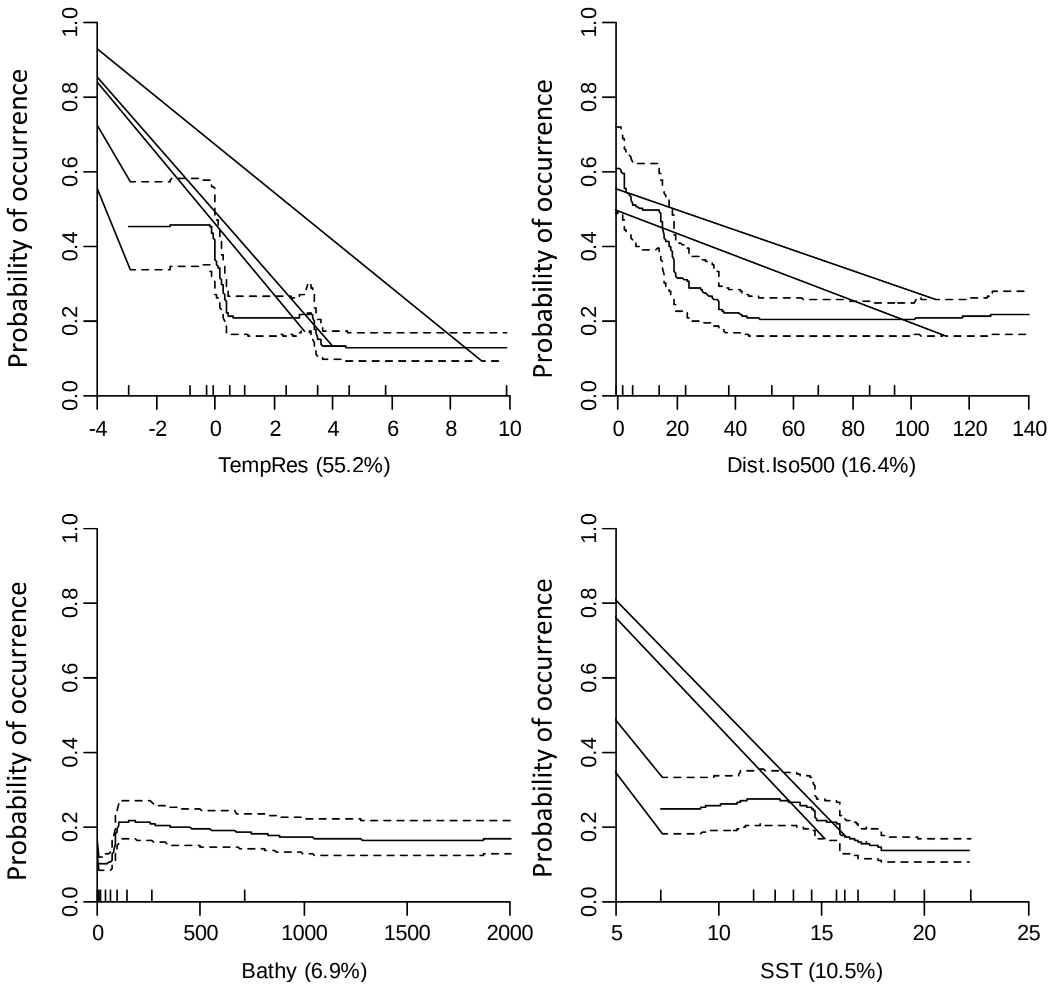

Fig A 4-71. Partial dependence plots showing the relationships between predictor variables and probability of dusky dolphin (*Lagenorhynchus obscurus)* occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

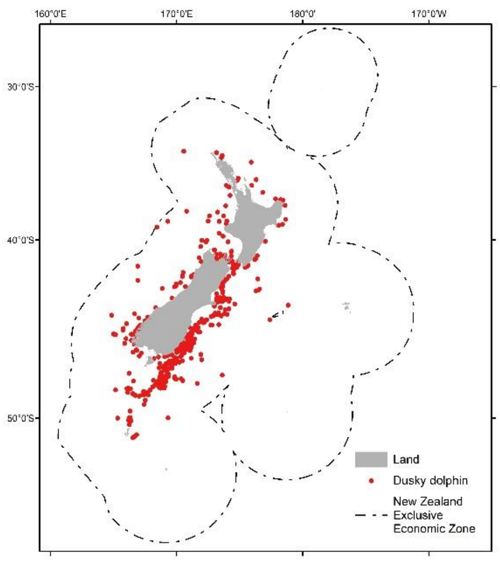

Fig A 4-72. Location of at-sea sightings of dusky dolphin (*n* = 823) in the NZ EEZ.

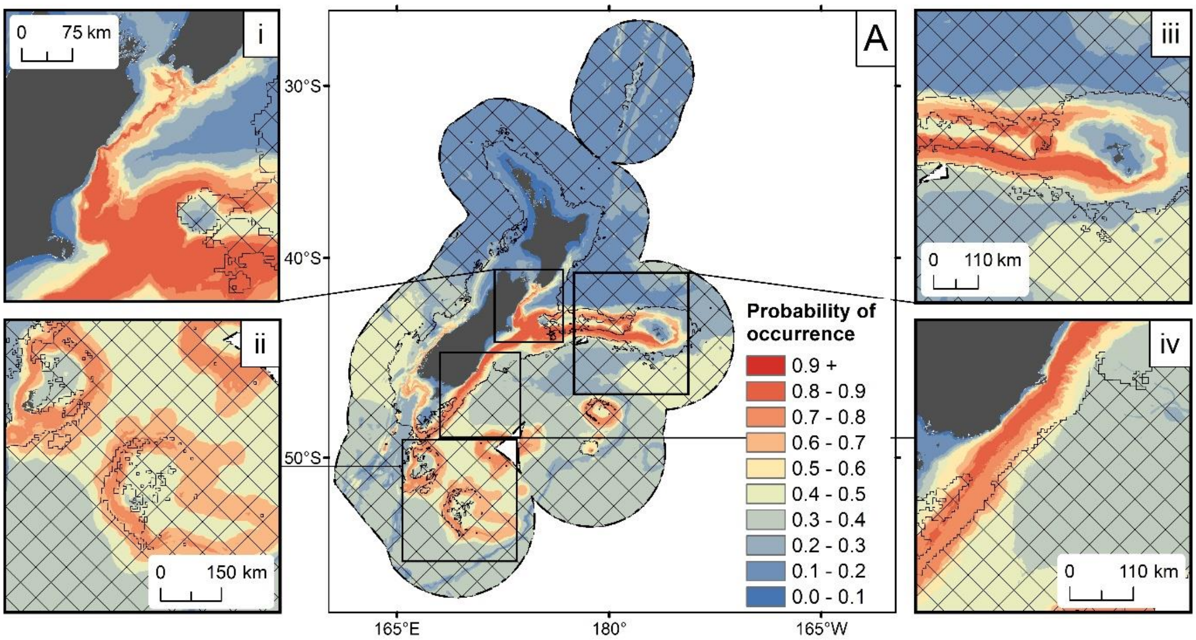

Fig A 4-73. The predicted probability occurrence of dusky dolphin (*Lagenorhynchus obscurus)* in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) Cook Strait, Kaikoura coast and west of the Chatham Rise; ii) Campbell Plateau, including Auckland Island and Campbell Island and Pukaki Rise ; iii) east of the Chatham Rise and Chatham Islands; iv) south east coast of the South Island. See figure 1 in the main body of the text for place names.

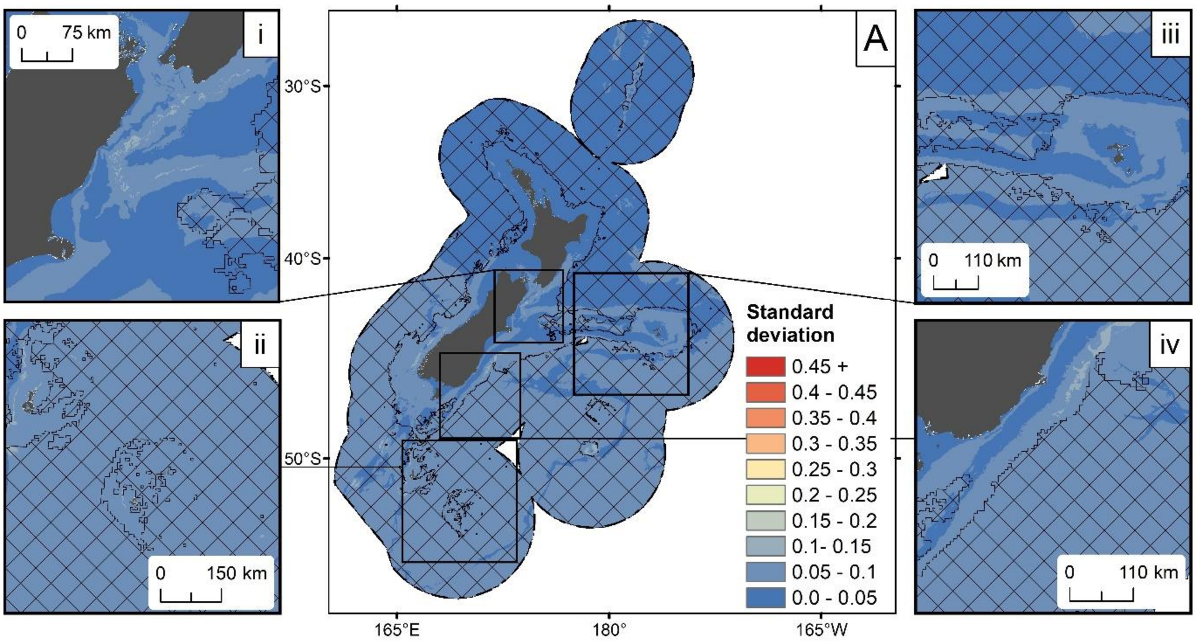

Fig A 4-74. Uncertainty estimates (standard deviation, SD) of dusky dolphin (*Lagenorhynchus obscurus)* probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) Cook Strait, Kaikoura coast and west of the Chatham Rise; ii) Campbell Plateau, including Auckland Island and Campbell Island and Pukaki Rise ; iii) east of the Chatham Rise and Chatham Islands; iv) south east coast of the South Island. See figure 1 in the main body of the text for place names.

Māui dolphin (*Cephalorhynchus hectori maui*)

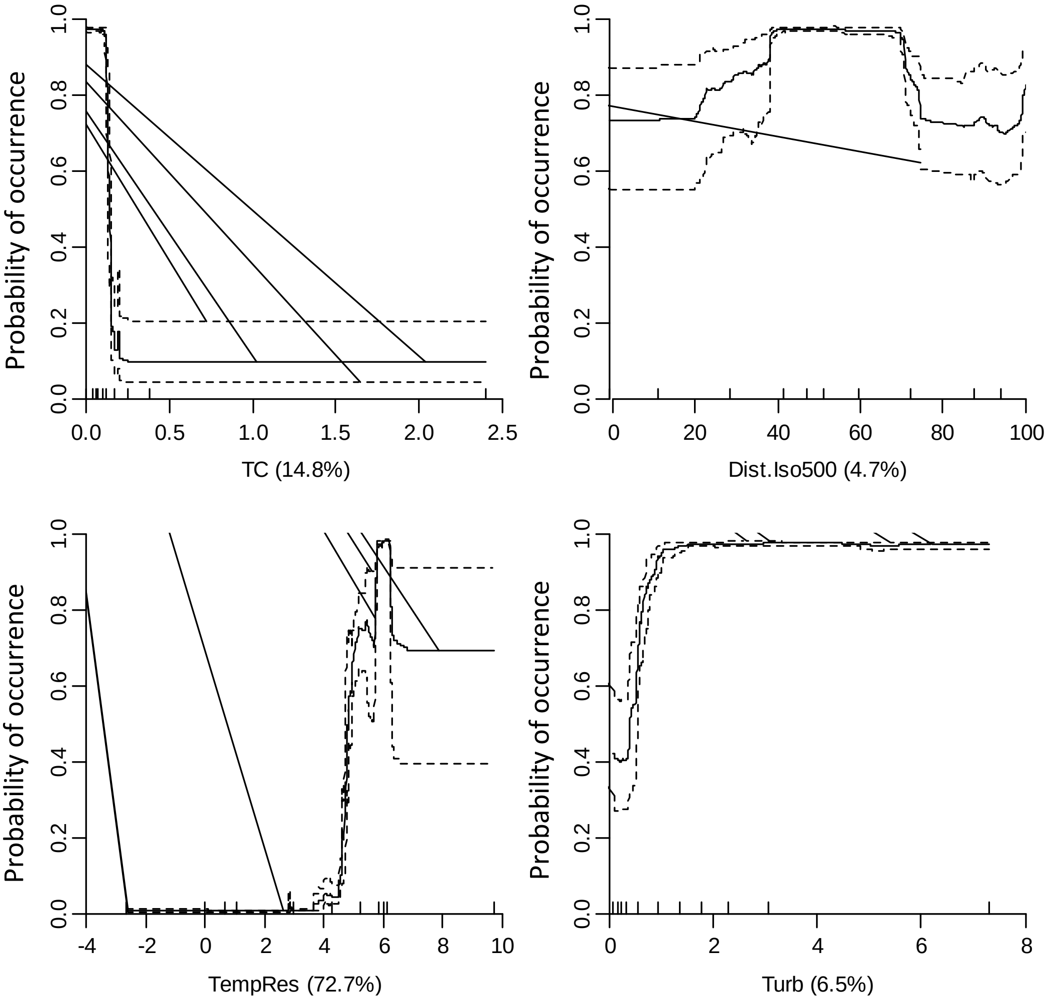

Fig A 4-75. Partial dependence plots showing the relationships between predictor variables and probability of Māui dolphin (*Cephalorhynchus hectori maui*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

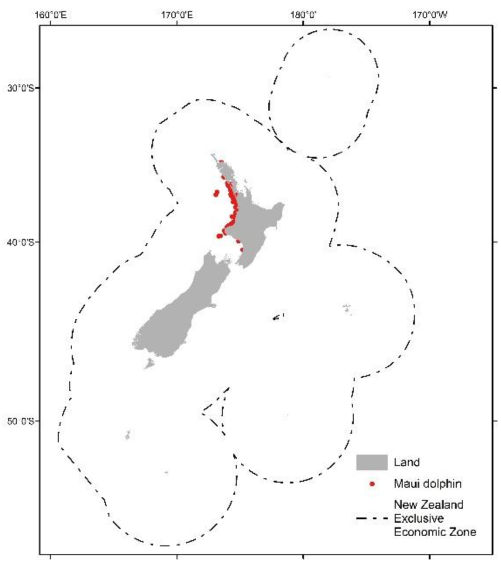

Fig A 4-76. Location of at-sea sightings of Māui dolphin (*n* = 1051) in the NZ EEZ.

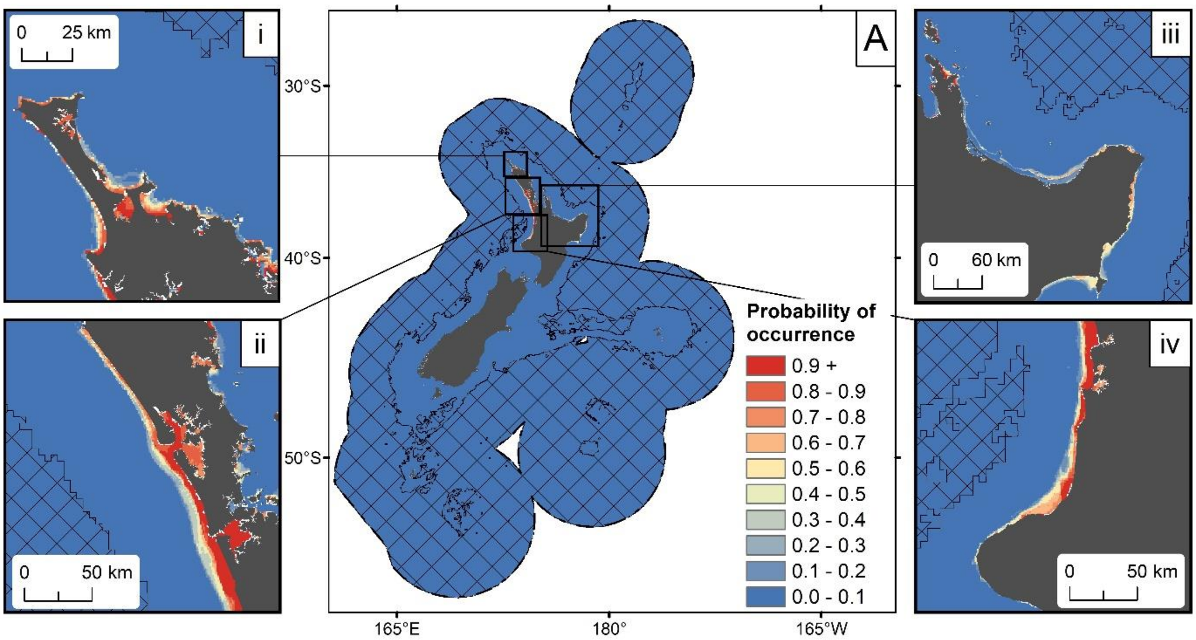

Fig A 4-77. The predicted probability occurrence of Māui dolphin (*Cephalorhynchus hectori maui*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape; ii) north of North Island including the Hauraki Gulf and north west coast; iii) central North Island including the Bay of Plenty; iv) central North Island including the North Taranaki Bight. See figure 1 in the main body of the text for place names.

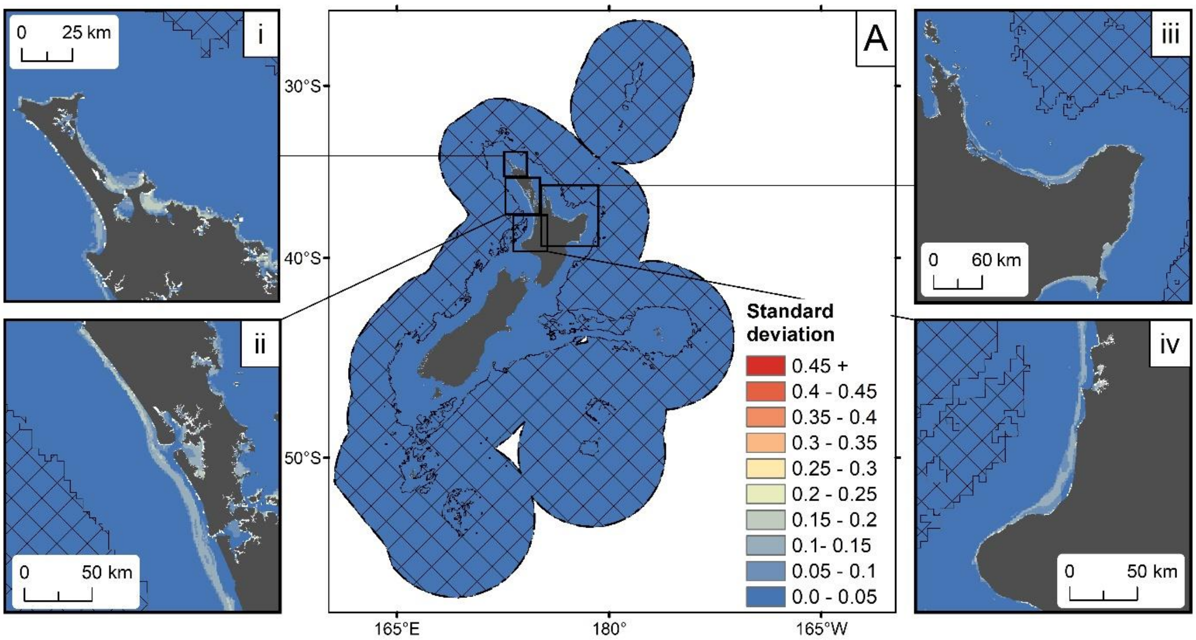

Fig A 4-78. Uncertainty estimates (standard deviation, SD) of Māui dolphin (*Cephalorhynchus hectori maui*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape; ii) north of North Island including the Hauraki Gulf and north west coast; iii) central North Island including the Bay of Plenty; iv) central North Island including the North Taranaki Bight. See figure 1 in the main body of the text for place names.

Hector's dolphin (*Cephalorhynchus hectori hectori)*

The predicted probability occurrence of Hector's dolphin (*Cephalorhynchus hectori hectori)* can be found in the main body of the research article.

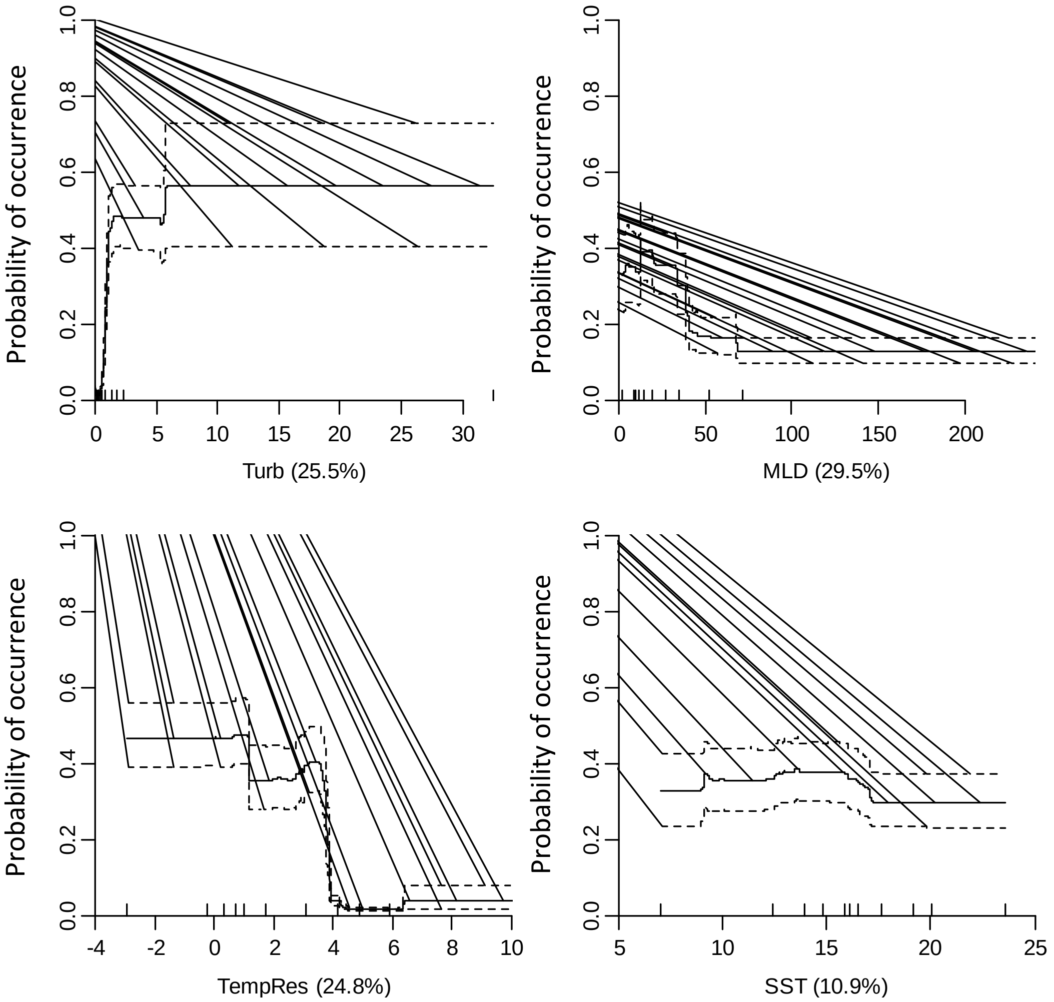

Fig A 4-79. Partial dependence plots showing the relationships between predictor variables and probability of Hector's dolphin (*Cephalorhynchus hectori hectori)* occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

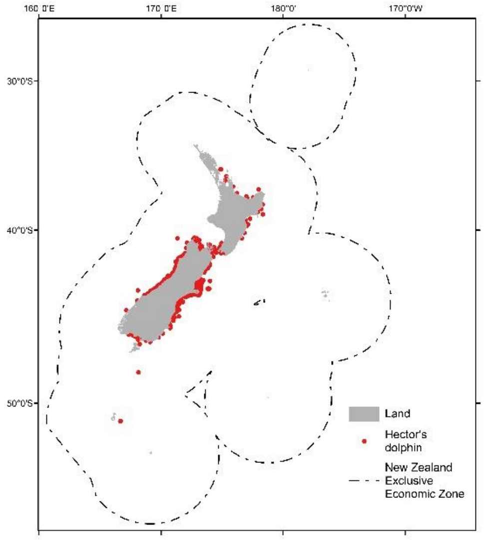

Fig A 4-80. Location of at-sea sightings of Hector's dolphin (*n* = 3688) in the NZ EEZ.

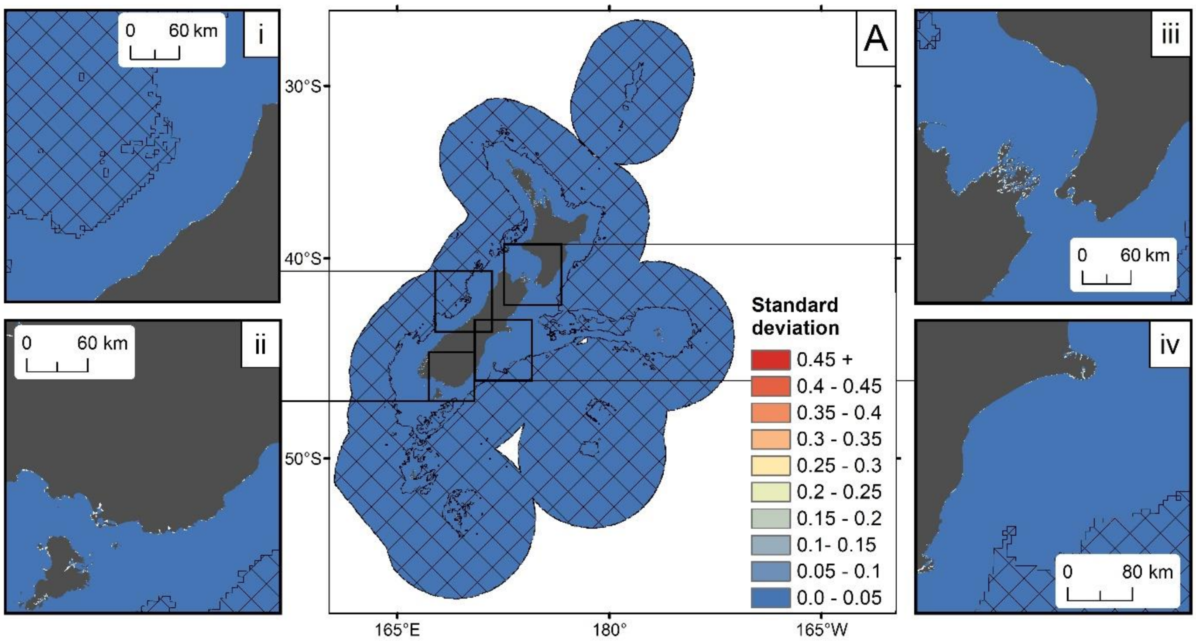

Fig A 4-81. Uncertainty estimates (standard deviation, SD) of Hector's dolphin (*Cephalorhynchus hectori hectori)* probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) west coast of South Island including the Fiordland Coast; ii) south of the South Island including Stewart Island / Rakiura; iii) south of the North Island and north of the South Island including Tasman and Golden Bays and Cook Strait; iv) East of the South Island including Canterbury Bight. See figure 1 in the main body of the text for place names.

Common dolphin (*Delphinus delphis*)

Example interpretation of partial dependence plots (Fig A 4-82) for common dolphin (*Delphinus delphis*) is provided here. The most important environmental predictor variables for the probability of common dolphin occurrence were temperature residuals (*TempRes*, 42.9%), water turbidity (*Turb*, 21.1%), distance to shore (*DistShore*, 17.0%) and sea surface temperature (*SST*, 8.9%) (Fig A 4-82). There was a strong positive relationship of predicted common dolphin probability occurrence with the variable temperature residuals (*TempRes* – i.e. areas where average temperature at depth is higher or lower than would be expected for any given depth). Probability occurrence was highest in areas with 1.5°C bottom water temperature greater than would be expected for the given depth (Fig A 4-82). Higher probability of occurrence was predicted for low turbidity areas (*Turb*) and areas 1 – 50 km from shore (*Dist.Shore*) (Fig A 4-82). A less clear (and weaker) relationship was observed between the probability of common dolphin occurrence and sea surface temperature (*SST*), although a small increase in probability occurrence was predicted for areas with *SST* greater than 17°C (Fig A 4-82). Overall, there was low uncertainty in predictions as shown by the 95% prediction intervals (dashed lines in Fig A 4-82).

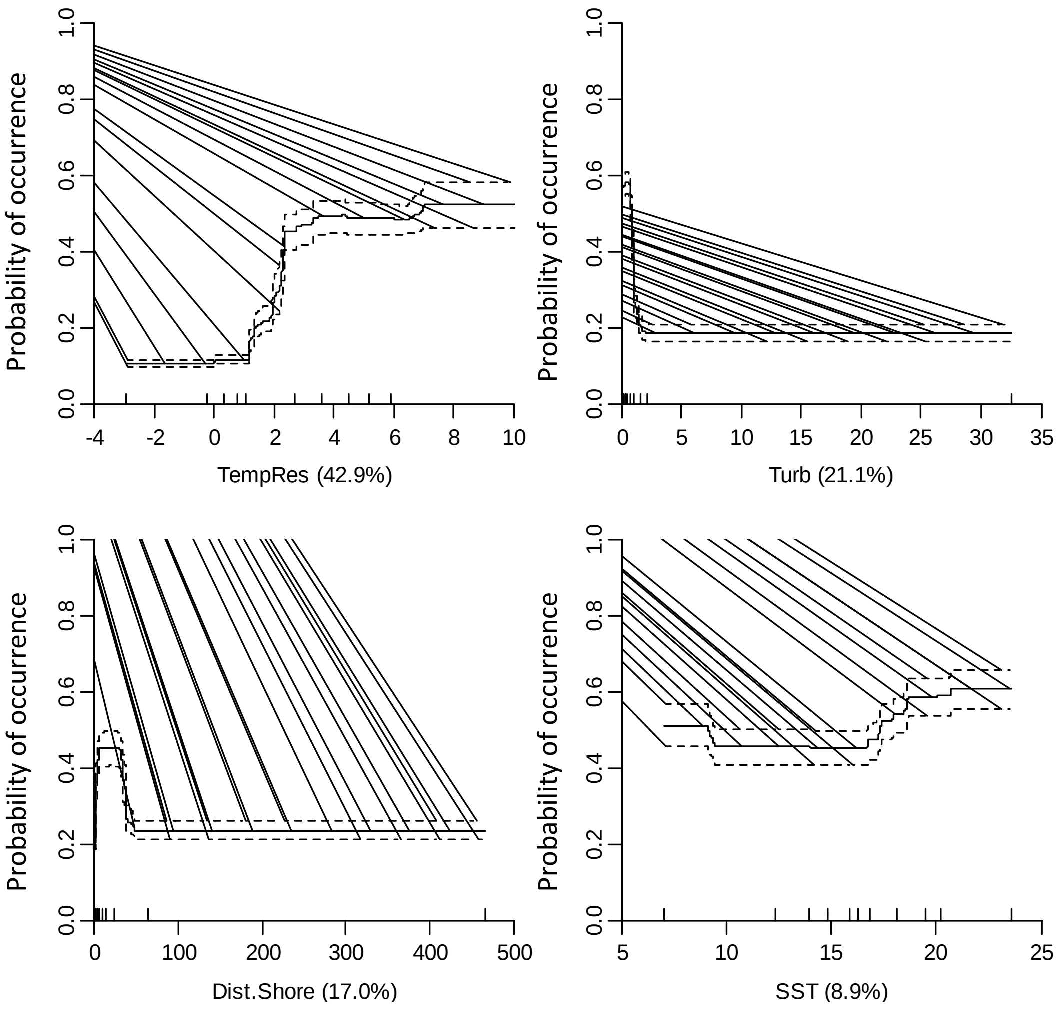

Fig A 4-82 Partial dependence plots showing the relationships between predictor variables and probability of common dolphin (*Delphinus delphis*) occurrence modelled using bootstrapped BRTs. The four most influential environmental predictors in the model are shown. Solid lines represent the mean of 100 bootstrap predictions and dashed lines the 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Each plot represents a predictor variable (labels and relative percentage contribution in parentheses are shown on the x-axes).

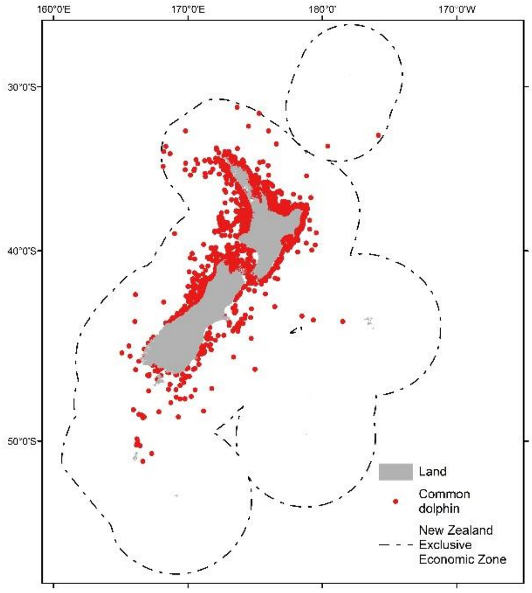

Fig A 4-83. Location of at-sea sightings of common dolphin (*n* = 4411) in the NZ EEZ.

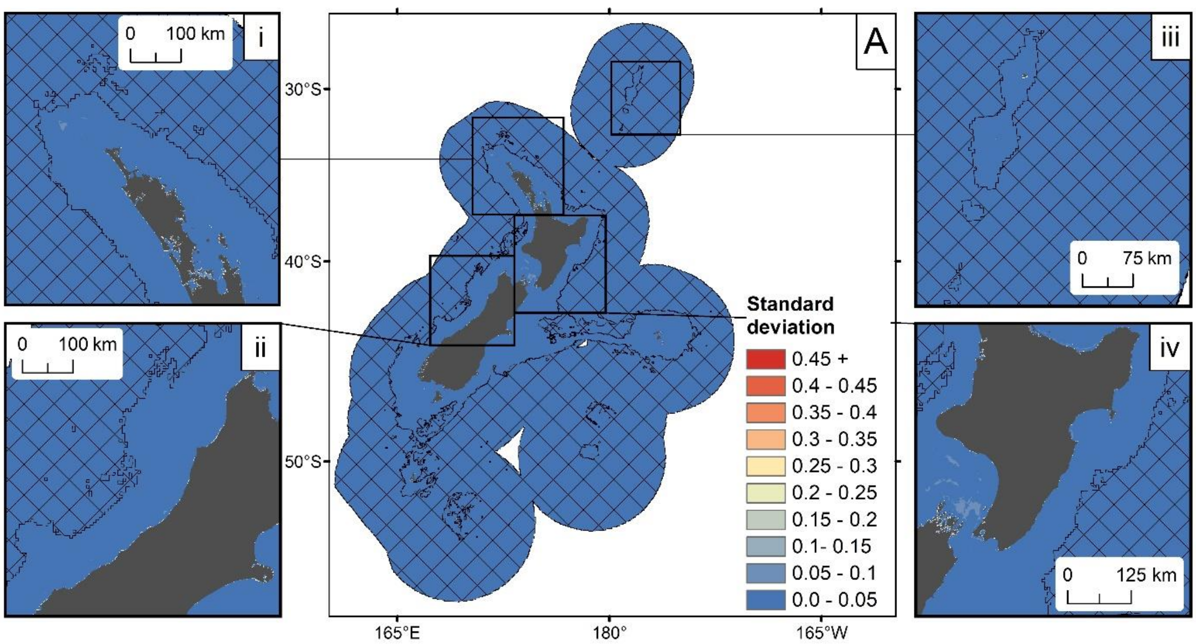

Fig A 4-84. Uncertainty estimates (standard deviation, SD) of common dolphin (*Delphinus delphis*) probability of occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: i) north of North Island including the North Cape and Hauraki Gulf; ii) west coast of South Island including the Fiordland Coast; iii) Kermadec islands, Lau-Colville Ridge and Kermadec Ridge; iv) south of the North Island including the South Taranaki Bight and Cook Strait. See figure 1 in the main body of the text for place names.

5. R code: BRT, RES and coverage of the environmental space

In the following section example R code is provided for BRT, RES and coverage of the environmental space models. At-sea cetacean sighting records and environmental variables will be available from the MPI managed NABIS website (https://maps.mpi.govt.nz) or upon request.

```
##================ Modelling the spatial distribution of cetaceans in New Zealand waters  ====================##

##------------------------------  Fabrice Stephenson, Théophile L Mouton, Kimberly Goetz, Fenna L Beets
##------------------------------  Project : MPI project PRO201401
##------------------------------  Start date : 01/06/2018
##------------------------------  End date : 15/12/2019

##=========================================================================================##

require(raster);require(dismo);require(gbm); require(sp); require(pROC);require(FinCal);require(ggplot2);
require(tidyverse); require(dplyr)

##### LOAD CETACEAN AT-SEA SIGHTINGS DATA ######
load("MM.LOC.Rdata")
head(MM.LOC)
# A dataframe with 14,513 rows (at sea sightings) and 17 columns (Species: factor with species names; X Y coordinates and
# env. variables: "Bathy", "BedDist", "ChlA","Dist.Iso500", "Dist.Shore", "DOM", "Kpar","MLD", "Slope", "SST", "TC",
# "TempRes","Turb", "VGPM")
MM.LOC <- MM.LOC[,c(1:3,7:20)]

##### LOAD ENVIRONEMTNAL DATA ######
load("predStack1km.Rdata")
# A raster stack of environmental variables used: "Bathy", "BedDist", "ChlA","Dist.Iso500",
# "Dist.Shore", "DOM", "Kpar","MLD", "Slope", "SST", "TC", "TempRes","Turb", "VGPM"

# add unique spatial identifier from raster stack to cetacean records
xy <- as.matrix(cbind(MM.LOC$X, MM.LOC$Y))
MM.LOC$FID <- cellFromXY(predStack1km[[1]], xy) # extract unique location from raster stack

# Turn raster stack in to dataframe for model prediction
FullEEZPreds.1km <- rasterToPoints(predStack1km, spatial=TRUE)
FullEEZPreds.1km <- data.frame(FullEEZPreds.1km@data, X =coordinates(FullEEZPreds.1km)[,1],
                Y =coordinates(FullEEZPreds.1km)[,2])
colnames(FullEEZPreds.1km) <- c("Bathy", "BedDist", "ChlA","Dist.Iso500", "Dist.Shore", "DOM", "Kpar","MLD", "Slope", "SST",
                        "TC", "TempRes","Turb", "VGPM", "X", "Y")

##### EXAMPLE CODE FOR HECTOR'S DOLPHIN #####
Species <- "hector's dolphin"
HD <- MM.LOC[MM.LOC$Species == Species,] #extract sightings for Hector's dolphin
HD$pres <- 1
# weighting of occurrences aggregated to 1km2
HD <- HD %>%
  group_by(FID) %>%
  summarize(wg = (1/sum(pres))) %>%
  select(wg = wg, FID) %>%
  left_join(HD,wg,by= "FID")
# shuffle order so that wg and FID are last columns
HD <- HD[,c(3:20,1:2)]

# generate relative absences
absence <- droplevels(MM.LOC[-which(MM.LOC$Species == Species), ] ) # create DF with all other species
absence <- absence[!absence$FID %in% HD$FID, ] # ensure absences are not in same location as presence
absence$pres <- 0
length.abs <- length(absence$Species)

# BOOTSTRAPING
# setup objects to save information from bootstraps
n.boot <- 100
length.map <- nrow(FullEEZPreds.1km)
boot_mat <- array(0, c(length.map,n.boot)) # matrix where rows will store predictions for each bootstrap (columns)
deviance_mat <- array(0, c(n.boot,4)) # for saving model fit metrics
```

```
influence_mat <- array(0,c(9,n.boot)) # for saving information on predictors

for (i in 1:n.boot) # bootstrap loop
{ # create training and evaluation presence/ relative absence dataframes
  train_ind <- sample(seq_len(nrow(HD)), size = floor(0.75 * nrow(HD))) # index of rows for 75% of presence data
  # creat training and evaluation presences
  HD_train <-HD[train_ind, ]
  HD_eval <-HD[-train_ind, ]

  #creation of absences
  rnd <- sort(sample(seq(1,length.abs),nrow(HD_train)*2))
  rnd.ev <- sort(sample(seq(1,length.abs),nrow(HD_eval)*2))
  absence.rnd <- absence[rnd,]
  absence.rnd.ev <- absence[rnd.ev,]

  # weighting of absence data 1) training 2 )evaluation
  absence.rnd <- absence.rnd %>%
    group_by(FID) %>%
    summarize(wg = (1/sum(pres+1))) %>%
    select(wg = wg, FID) %>%
    left_join(absence.rnd,wg,by= "FID")
  # shuffle order so that wg = at end
  absence.rnd <- absence.rnd[,c(3:ncol(absence.rnd),1)]

  absence.rnd.ev <- absence.rnd.ev %>%
    group_by(FID) %>%
    summarize(wg = (1/sum(pres+1))) %>%
    select(wg = wg, FID) %>%
    left_join(absence.rnd.ev,wg,by= "FID")
  # shuffle order so that wg = at end
  absence.rnd.ev <- absence.rnd.ev[,c(3:ncol(absence.rnd.ev),1)]

  # creation of presence and absence
  HD.PresAbs.B <- rbind(HD_train,absence.rnd) # final presence/absence data
  HD_eval <- rbind(HD_eval,absence.rnd.ev)

  # Training model
  M1<-gbm.step(data=HD.PresAbs.B, gbm.x = c(4,7,11,13,15,16,17), # column number of env preds to be used in analysis
             gbm.y = 18, # column number of presence/ relative absence data
             weight = 19, # column number of weighting
             family = "bernoulli", tree.complexity = 5,
             learning.rate = 0.005, bag.fraction = 0.6, n.folds=10,
             max.trees = 7500, step.size = 25, plot.main = F, tolerance.method = "fixed",
             tolerance = 0.01, verbose = F)

  #internal model fit metrics
  int.null.deviance <- M1$self.statistics$mean.null
  int.residual.deviance <- M1$cv.statistics$deviance.mean
  deviance_mat[i,1] <- (int.null.deviance-int.residual.deviance)/int.null.deviance
  #internal AUC
  deviance_mat[i,3] <- M1$cv.statistics$discrimination.mean
  #model fit comparison with withheld evaluation data
  pred <- predict.gbm(M1, HD_eval, n.trees = M1$gbm.call$best.trees, type = "response")
  ext.residual.deviance <- calc.deviance(HD_eval$pres, pred, family = "bernoulli" ,calc.mean=T)
  ext.null.deviance <- calc.deviance(HD_eval$pres,family = "bernoulli", rep(mean(HD_eval$pres),nrow(HD_eval)),calc.mean=T)
  deviance_mat[i,2]<-(ext.null.deviance - ext.residual.deviance)/ext.null.deviance
  # external AUC
  deviance_mat[i,4] <- roc(HD_eval$pres, pred)$auc

  # Env pred contribution
  M1_contrib <- as.data.frame(M1$contributions) # M1 = your BRT model object
  env_var_ord <- M1_contrib[order(M1_contrib$var),]
  influence_mat[,i]<-env_var_ord[,2]

  # predict spatially to map
  pred.map <- predict.gbm(M1, FullEEZPreds.1km, n.trees = M1$gbm.call$best.trees, family = "bernoulli", type = "response")
  boot_mat[,i] <- round(pred.map, digits = 2) # round to save space
}

save(boot_mat, file="HD_boot_mat.Rdata")
```

```
save(deviance_mat, file="HD_dev_mat.Rdata")
save(influence_mat, file="HD_inf_mat.Rdata")

# Importance of Environmental predictors
row.names(influence_mat)<-as.character(env_var_ord[,1])
preds_influences<-apply(influence_mat, 1, function(x) c(mean = mean(x), sd = sd(x)))   #Calculate mean and standard error of
# the relative influecne of each preds
write.csv(preds_influences, file="preds.influences_HD.csv")

# Model fit metrics
colnames(deviance_mat) <- c("Dev.Exp.int","AUC.int","Dev.Exp.eval","AUC.eval")
mean.model.fit <-apply(deviance_mat, 2, function(x) c(mean = mean(x), sd = sd(x)))
write.csv(mean.model.fit, file="mean.model.fit.csv")

# caculate mean suitability and CV spatially
HD.boot.mean<-apply(boot_mat,1,mean)
HD.boot.sd<-apply(boot_mat,1,sd)
HD.boot.cv<-HD.boot.sd/HD.boot.mean # calculation of Coefficient of variation if needed

# export the map
HD.map.mean <- cbind(FullEEZPreds.1km[, c("X", "Y")],HD.boot.mean) # mean prediciton
HD.map.UC <- cbind(FullEEZPreds.1km[, c("X", "Y")],HD.boot.sd) # uncertainty

# convert to raster
BRT_HD.mean <- rasterFromXYZ(data.frame(x =HD.map.mean[,1],
                                        y =HD.map.mean[,2],
                                        z =HD.map.mean[,3]),
                             crs = crs("+init=epsg:3851")) # same proj as orginal env variables tiff files
BRT_HD.cv <- rasterFromXYZ(data.frame(x =HD.map.CV[,1],
                                      y =HD.map.CV[,2],
                                      z =HD.map.CV[,3]),
                           crs = crs("+init=epsg:3851")) # same proj as orginal env variables tiff files
plot(BRT_HD.mean)
plot(BRT_HD.cv)

################# SAVE FILE
writeRaster(BRT_HD.mean,filename = "BRT_HD_MEAN.tif", overwrite=T, progress = "window")
writeRaster(BRT_HD.cv,filename = "BRT_HD_CV.tif", overwrite=T)



##### RES MODELLING OF SPECIES WITH < 50 SIGHTINGS ###########

# function for linear decay between rmin and r pref, ideal conditions, and linear decay between rprefmax and r max. all other
# areas= o
rasterRescale<-function(r,r.min, rpref.min, rpref.max, r.max){

  ideal <- r >=rpref.min & r<=rpref.max

  lower <- (r >=r.min & r<rpref.min)*r
  lower[lower == 0] <- NA
  lower.r <- (lower-r.min)/(rpref.min-r.min)

  upper <- (r >rpref.max & r<=r.max)*r
  upper[upper == 0] <- NA
  upper.r<-(r.max-upper)/(r.max-rpref.max)

  upper.r[is.na(upper.r[])]<-0
  lower.r[is.na(lower.r[])]<-0
  ideal[is.na(ideal[])]<-0
  final <- lower.r+upper.r+ideal
}

#### environmental data ####
bathy <- predStack1km[[1]] # save individual rasters from the stack
SST <- predStack1km[[10]]
Dist.Shore <- predStack1km[[5]]

# RES LOOP
```

```
# Common name; Species/sub-species name; Abbreviation used in model
# Blainville's beaked whale; Mesoplodon densirostris; BBW
# Dwarf minke whale; Balaenoptera acutorostrata; DMW
# Spectacled porpoise	Phocoena dioptrica	SP
# Striped dolphin	Stenella coeruleoalba	SD
# Andrew's beaked whale	Mesoplodon bowdoini	ANBW
# Hourglass dolphin	Lagenorhynchus cruciger	HGD
# Pygmy sperm whale	Kogia breviceps	PSW
# Southern bottlenose whale	Hyperoodon planifrons	SBNW
# Risso's dolphin	Grampus griseus	RS
# Shepherd's beaked whale	Tasmacetus shepherdi	SBW
# Cuvier's beaked whale	Ziphius cavirostris	CBW
# Gray's beaked whale	Mesoplodon grayi	GBW
# Southern right whale dolphin	Lissodelphis peronii	SRWD
# False killer whale	Pseudorca crassidens	FKW
# Arnoux's beaked whale	Berardius arnuxii	ARBW

#Create a vector of species names
species <- c("ANBW", "ARBW", "BBW","CBW", "DMW", "FKW", "GBW", "HGD", "PSW", "RS",
        "SBW", "SFPW", "SBNW", "SRWD", "SP", "SD")
length_species <- length(species)

#####Create a list of the parameters (Bathy, SST, Dist.Shore) for each species

list_bathy <- list(
  ANBW = c(0, 200, 2000, 8000),
  ARBW = c(0, 200, 2000, 8000),
  BBW = c(0, 200, 2000, 8000),
  CBW = c(0, 200, 2000, 8000),
  DMW = c(0, 10, 2000, 8000),
  FKW = c(0, 1000, 2000, 8000),
  GBW = c(0, 200, 2000, 8000),
  HGD = c(0, 1000, 4000, 8000),
  PSW = c(0, 200, 2000, 8000),
  RS = c(0, 200, 1000, 6000),
  SBW = c(0, 1000, 2000, 8000),
  SFPW = c(0, 200, 2000, 8000),
  SBNW = c(0, 1000, 4000, 8000),
  SRWD = c(0, 200, 2000, 8000),
  SP = c(0, 10, 2000, 8000),
  SD = c(0, 200, 2000, 8000))


list_SST <- list(
  ANBW = c(5, 10, 20, 25),
  ARBW = c(-2, 0, 20, 25),
  BBW = c(10, 15, 30, 35),
  CBW = c(5, 10, 30, 35),
  DMW = c(-2, 0, 25, 35),
  FKW = c(5, 10, 30, 35),
  GBW = c(0, 5, 20, 25),
  HGD = c(-2, 0, 10, 15),
  PSW = c(10, 15, 30, 35),
  RS = c(5, 10, 30, 35),
  SBW = c(0, 5, 20, 25),
  SFPW = c(15, 20, 30, 35),
  SBNW = c(-2, 0, 20, 25),
  SRWD = c(-2, 0, 20, 25),
  SP = c(-2, 0, 10, 15),
  SD = c(5, 10, 25, 30))

list_Dist.Shore <- list(
  ANBW = c(10, 100, 1000, 4000),
  ARBW = c(10, 100, 1000, 4000),
  BBW = c(10, 100, 1000, 4000),
  CBW = c(10, 100, 1000, 4000),
  DMW = c(10, 100, 1000, 4000),
  FKW = c(10, 100, 1000, 4000),
  GBW = c(10, 100, 1000, 4000),
  HGD = c(10, 100, 1000, 4000),
```

```
  PSW = c(10, 100, 1000, 4000),
  RS = c(10, 100, 1000, 4000),
  SBW = c(10, 100, 1000, 4000),
  SFPW = c(10, 100, 1000, 4000),
  SBNW = c(10, 100, 1000, 4000),
  SRWD = c(10, 100, 1000, 4000),
  SP = c(10, 100, 1000, 4000),
  SD = c(10, 100, 1000, 4000))

for (i in 1:length_species) {
  model_name <- species[i]
  bathymetry <- rasterRescale(bathy,list_bathy[[i]][1], list_bathy[[i]][2],list_bathy[[i]][3],list_bathy[[i]][4])
  SeaTemp <- rasterRescale(SST,list_SST[[i]][1], list_SST[[i]][2],list_SST[[i]][3],list_SST[[i]][4])
  Distance.Shore <- rasterRescale(Dist.Shore,list_Dist.Shore[[i]][1],
list_Dist.Shore[[i]][2],list_Dist.Shore[[i]][3],list_Dist.Shore[[i]][4])
  RES <- SeaTemp*bathymetry*Distance.Shore
  writeRaster(RES,filename = paste(model_name, "_RES",".tif", sep= ""), overwrite=T)
}



##### BRT MODELLING OF THE COVERAGE OF THE ENVIRONMENTAL SPACE ###########
# Coverage by samples
MM_SAMPLE <- MM.LOC[,c(4:18)] #extract environmental predictors and FID
MM_SAMPLE$sample.sites <- 1 # sample present

# coverage by absences
xy <- as.matrix(cbind(FullEEZPreds.1km$X, FullEEZPreds.1km$Y))
FullEEZPreds.1km$FID <-  cellFromXY(predStack1km[[1]], xy)
FullEEZPreds.1km <- FullEEZPreds.1km[!FullEEZPreds.1km$FID %in% MM.LOC$FID,] # no overlap with presences
train_ind <- sample(seq_len(nrow(FullEEZPreds.1km)), size = 50000)
preddat_EC <- FullEEZPreds.1km[train_ind, ]
preddat_EC <- preddat_EC[,-c(15:16)] # replace NAs with the mean value
preddat_EC$sample.sites <- 0 # absent

# bind by row
MM_ES <- rbind(preddat_EC,MM_SAMPLE)

MM_BRT_M2 <- gbm.step(data=MM_ES, gbm.x = 1:14, # environmental varibale columns
                      gbm.y = 16, # presence absence of samples
                      family = "bernoulli", tree.complexity = 2,
                      learning.rate = 0.05, bag.fraction = 0.75, n.folds=10,
                      max.trees = 5000, plot.main = T, tolerance.method = "fixed",
                      tolerance = 0.01, verbose = T)

pred.map <- predict.gbm(MM_BRT_M2, FullEEZPreds.1km,
                        n.trees = MM_BRT_M2$gbm.call$best.trees, type = "response")
# export the map
ES.map.mean <- cbind(FullEEZPreds.1km[, c("X", "Y")],pred.map)

# convert to raster
BRT_ES.mean <- rasterFromXYZ(data.frame(x = ES.map.mean[,1],
                                        y = ES.map.mean[,2],
                                        z = ES.map.mean[,3]),
                             crs = crs("+init=epsg:3851")) # same proj as orginal env variables tiff files

plot(BRT_ES.mean) # plot map
points(MM.LOC[,2],MM.LOC[,3], col = "black", pch = ".", cex = .3) # plot presence
writeRaster(BRT_ES.mean,filename = "Env_cov_MM.tif", overwrite=T) # write raster file
```