# Appendix S1 - Cetacean conservation planning in a global diversity hotspot: dealing with uncertainty and data deficiencies

*Ecosphere*

Fabrice Stephenson, Judi E Hewitt, Leigh G Torres, Théophile L Mouton, Tom Brough, Kimberly T. Goetz, Carolyn J Lundquist, Alison B MacDiarmid, Joanne Ellis, and Rochelle Constantine

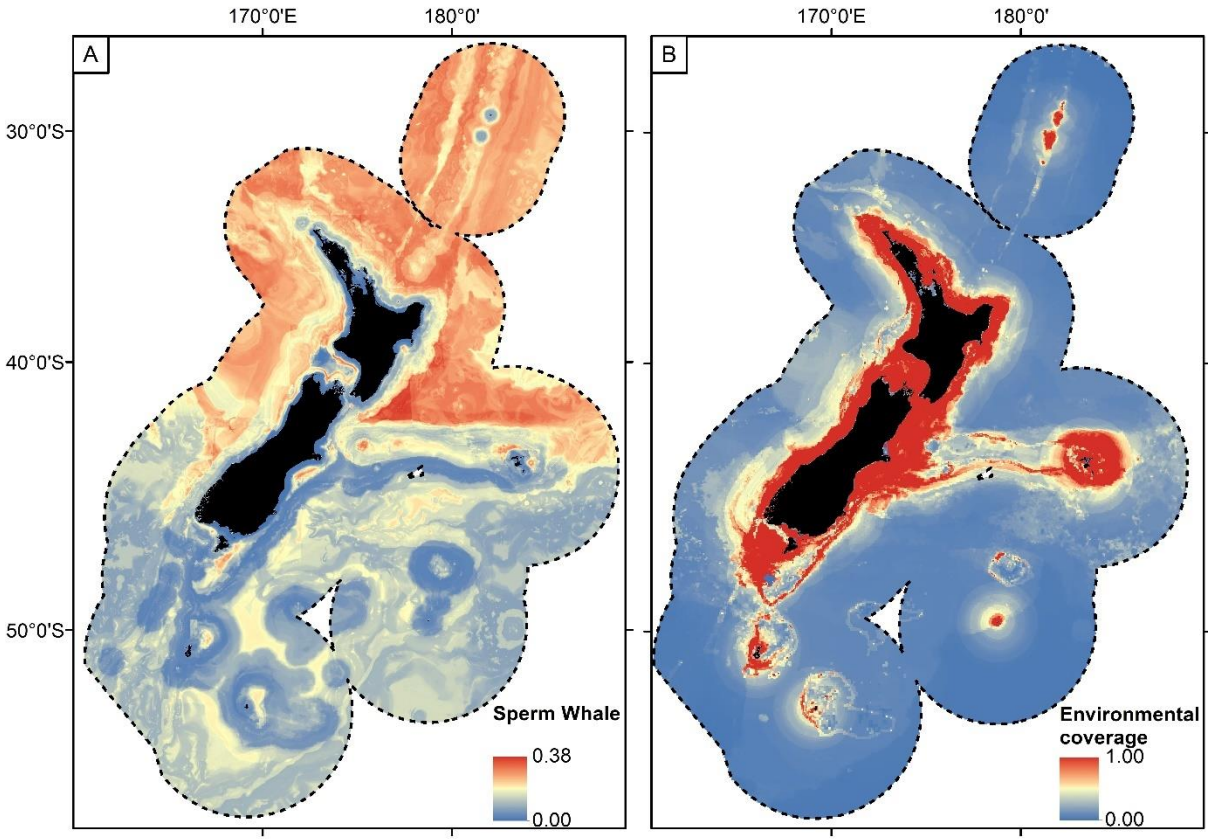


Figure. S1. Example of uncertainty estimates (standard deviation, SD): SD of sperm whales probability of occurrence in the New Zealand EEZ (A); and Environmental coverage (B).

## Brief methods for the prediction of environmental coverage based on Stephenson et al., 2020a.

The degree to which the environmental conditions of each predictive site was covered by the sightings records was quantified by randomly sampling 50,000 values from the

environmental space (where no cetacean sightings were recorded) and assigning a ‘records’ value of 0 to these, indicating that these were ‘absent’ sample sites. These were combined with the true sightings records ($n$ = 14,513), to which a ‘records’ value of 1 was assigned (‘present’). A BRT analysis was then used to model the relationship between ‘absent’ (random) records and ‘present’ (true) records for the 14 environmental predictor variables, using a Bernoulli error distribution. Predictions using this model yielded estimates of the probability of a site occurring in each part of the environmental space. A learning rate that yielded 2,000 trees with an interaction depth of 2 was used (so that only pair-wise combinations of the environmental variables were considered). Predictions were then made spatially, generating values between 0 and 1 (where 0 indicated little understanding of the environmental space and 1 a perfect understanding), according to how well each cell was represented by the sighting records