SUSTAINABLE SEAS

Ko ngā moana whakauka

National Science Challenges

# Innovation Fund Research Proposal Template

## A. PROJECT TITLE

Can we define marine habitat use by seabirds without costly at-sea observational data?

## B. PROJECT TEAM


| Project Leader: | Investigators: |
|---|---|
| Dr David Thompson, National Institute of Water and Atmospheric Research Ltd.<br>301 Evans Bay Parade, Hataitai, Wellington 6021<br>david.thompson@niwa.co.nz<br>04 386 0582 | Dr Kimberly Goetz, National Institute of Water and Atmospheric Research Ltd.<br>301 Evans Bay Parade, Hataitai, Wellington 6021<br>kim.goetz@niwa.co.nz<br>04 382 1623 |


## C. ABSTRACT


New Zealand waters support the greatest diversity of seabirds on Earth, yet quantitative information on where seabirds occur within New Zealand's EEZ, and how distributions vary temporally, remains sparse. This lack of data often means that managers, decision makers and the public alike are often faced with a lack of detailed and robust information to gauge how particular threats could potentially impact these high-profile, protected species. However, acquiring detailed, spatially and temporally resolved seabird distribution data through conventional approaches requires considerable resources. This project will examine the extent to which seabird location data, either sightings data or data derived from electronic tracking devices, can be used to significantly improve our understanding of seabird distributions over and above that derived from relative environmental suitability (RES) modelling. By comparing species-specific RES models to habitat suitability models produced from (1) observational data and (2) location data from bird-borne electronic tags, our approach will quantify the level of accuracy associated with each method and determine if RES models are sufficient to produce accurate estimates of seasonal seabird distributions or identify if additional data are needed. This tiered and entirely novel approach has the potential for huge resource savings.


## D. RELEVANCE TO CHALLENGE OBJECTIVE

This proposal is entirely aligned with the Challenge's Objective to 'enhance utilisation of our marine resources within environmental and biological constraints' by:

- Delivering novel information on how to most cost-effectively define spatio-temporal variation in marine habitat use by New Zealand seabirds
- Providing a framework through which potential interactions between seabirds and anthropogenic marine activities can be better understood, and
- Reducing uncertainty around New Zealand seabird habitat use that will be of direct and practical relevance to resource managers and policy makers faced with challenging decisions around the effects of increasing human resource use in New Zealand's marine realm.

## E. INTRODUCTION

Marine megafauna, including seabirds, enjoy an extremely high public profile. These charismatic species, perhaps more than any other group, define the character of New Zealand's diverse marine estate. New Zealand waters support the greatest diversity of seabirds on Earth, yet quantitative information on where and when seabirds occur within New Zealand's EEZ remains sparse and managers, decision makers and the public alike are often faced with a lack of detailed and robust information to gauge how particular threats could potentially impact these species. The considerable interest by public and managerial stakeholders in maintaining and protecting these iconic New Zealand species, while simultaneously balancing the needs of marine resource use, indicates an obvious need for high-quality spatial and temporal data on species distribution and abundance, as well as the environmental drivers underpinning distribution and abundance. For example, the Decision Making Committees of recent Environmental Protection Authority hearings have relied on incomplete and patchy information, with a high degree of uncertainty, in this space when deciding upon the potential of proposed resource extraction activities to impact seabirds.

However, the collection of robust at-sea observational data for seabirds through systematic surveys, even over relatively small areas, can be prohibitively expensive. Clearly, a tension exists between the need for high quality, temporally resolved seabird distribution data within areas of interest, and the relatively high costs of acquiring such information. The research proposal outlined here, entirely novel within the New Zealand context, aims to test the extent to which currently available location data can be used to (1) significantly improve our understanding of seabird distributions and therefore (2) prioritise future data collection so as to minimise resource expenditure and maximise information delivery. Importantly, our approach will be tiered, comparing the performance of seabird distribution models built without seabird location data to those utilising sightings data and others utilising location data acquired from bird-borne tracking devices.

Currently, research into marine megafauna in general, and seabirds in particular, is lacking within the Sustainable Seas National Science Challenge. This proposal would fill that gap and would, importantly, be of direct and practical relevance to the Challenge's objective to recognise environmental and biological constraints of enhanced utilisation of marine resources. This would be achieved by providing marine decision makers with the information needed to assess the effects that anthropogenic activities in our marine environment might have on New Zealand's diverse seabird assemblage.

## F. AIMS

The over-arching null hypothesis of this project is that seabird distribution models provide equally good information about when and where seabirds utilise marine resources, regardless of the information used to construct the models.

In order to test this hypothesis we will:

- Construct a suite of seabird distribution models, each with increasingly accurate seabird distribution data
- Quantify the extent to which high-quality tracking data can improve predictions of seabird habitat suitability, derived from our distribution models, over and above those produced with and without seabird sightings data, and
- Critically assess whether predictions from habitat suitability models produced with and without seabird location data can result in a reliable cost-effective tool to inform management

Which will then enable us to:

- Determine the costs of collecting additional seabird location information that might be needed, if any, to produce more robust model outputs, and
- Provide a framework for reducing uncertainty around decisions where seabird use of marine space may conflict with human activities

## G. PROPOSED RESEARCH

### Introduction

This project will test to what extent seabird location data, either sightings data or data derived from electronic tracking devices, can be used to significantly improve our understanding of seabird distributions over and above that derived from relative environmental suitability (RES) modelling.

Using a tiered approach, this research will model seabird species' distributions, introducing additional data at each step and comparing model performance. First, we will construct RES models, which do not require seabird location data for model construction. Next, we will use opportunistically-collected and publicly-available at-sea location data to predict species occurrence. Both the RES and the species distribution models will be evaluated by comparing model outputs to the 'truth', or, in this case, location data from electronic tracking devices - light-based geolocation and GPS tags. This structured approach will enable us to assess the extent to which currently available seabird sightings and tracking data can improve our prediction of species' distributions beyond which can be achieved from RES modelling.

### Study Species

We propose to focus on a suite of five procellariiform seabirds: Campbell albatross *Thalassarche impavida*, grey-headed albatross *T. chrysotoma*, white-capped albatross *T. steadi*, Buller's albatross *T. bulleri* and sooty shearwater *Ardenna grisea*. These species all range widely through New Zealand's EEZ and occur within the Sustainable Seas focal region, either when foraging (e.g., Shaffer et al., 2009) or transiting through the area on migration (e.g., Shaffer et al., 2006). Grey-headed albatross is classified as 'threatened', the remaining four species are classified as 'at-risk' (Robertson et al., 2013).

Species selection is constrained largely by the availability of both sightings data and location data acquired from electronic tracking devices. While the species noted above are all migratory, with at least some component of the population departing New Zealand waters following breeding, our aim is to focus upon the breeding season component of species' annual cycles.

Relative Environmental Suitability models

First, we will thoroughly examine the primary and secondary literature for known habitat associations for each of the five proposed seabird study species to produce relative environmental suitability (RES) models. Our RES models will not use seabird location data of any kind, but will incorporate preferred habitat parameters available in the published literature. Models will focus on the New Zealand breeding season for each species, but will additionally and separately include the non-breeding season for those species which remain in New Zealand waters year-round (e.g., white-capped albatross). In the absence of sightings and tag-derived location data, this approach uses information on species' overall distributions, foraging behaviour, and life history characteristics relative to a suite of environmental parameters to assign each species to broad-scale niche categories and creates a rule-based envelope model to map species-habitat relationships (Kaschner et al., 2006, 2011). We plan to incorporate a suite of environmental variables that have been used previously in this approach, including depth, bathymetric slope, sea surface temperature, chlorophyll a (chl a) concentration and distance from land (e.g., see Watson et al., 2013). Depending on the environmental data and the qualitative habitat descriptions available, other environmental variables may be included on a species by species basis. Following Kaschner et al. (2006), we will assign each species to a niche category for each environmental variable based on a species' response curve (also termed a resource selection function): here, each habitat predictor is assigned a preferred minimum and maximum value, between which suitability is assumed to be uniform and maximal (value = 1), with suitability tending to zero towards absolute minimum and maximum habitat predictor values. An index of RES will be derived, based on an appropriate grid size (dependent on the resolution of available environmental variables), on a scale from zero to one, by multiplying the suitability of each environmental predictor variable. In this way, if any one predictor variable falls outside of a species' absolute range the overall environmental suitability will be zero. This assumes that a species will not occur in an environment where one or more environmental variable is deemed unsuitable. Using ArcGIS, we will derive the 95%, 75%, and 50% data contours to determine areas where the five species are spending the majority of their time during both the breeding and non-breeding seasons.

## Habitat suitability models – sightings data

Secondly, we will use sightings data for each species as the response variable in boosted regression tree (BRT) models in order to predict species' preferred habitat, based on the same environmental predictor variables as included in the RES models.

Sightings data will be acquired from published primary and secondary literature sources (e.g., Jenkins, 1981, McQuaid & Ricketts, 1984), together with online, publicly-accessible seabird sightings databases (e.g. eBird: http://ebird.org/content/newzealand/). We will use BRTs since they are able to accommodate the non-linear relationships characteristic of many ecological systems and can additionally interpret highly complex relationships between species and their environment. BRTs have become increasingly popular in habitat suitability modelling across a wide array of study systems, including tortoise *Gopherus agassizii* in the central Mojave Desert (Andersen et al., 2000), demersal fish species richness around New Zealand (Leathwick et al., 2006), benthic assemblages off New Zealand (Compton et al., 2013), southern right whale *Eubalaena australis* in Australasia (Torres et al., 2013) and phosphorite nodule deposits on the Chatham Rise (Leduc et al., 2015). BRT models use two algorithms: 1) classification and regression trees and 2) boosting to combine a collection of models (Elith et al., 2008). By combining trees during boosting, the misclassification error is minimised and the overall predictive performance improved (Leathwick et al., 2006). Boosting is optimized by three parameters: the learning rate (lr) that determines the weight of individual trees, tree complexity (tc) that indicates the number of interactions, and the number of trees (nt), as used by Torres et al. (2015).

We will use the 'gbm' package (Ridgeway, 2006, 2007, 2010) within the statistical software R 3.2.2 (R Development Core Team, 2013) and methods proposed Elith et al. (2008) to fit a BRT model to our data. A binomial error distribution of the Bernoulli family will be used to predict the probability of seabird occurrence. As such, we will generate pseudo absence points and we will obtain environmental data at both presence and absence locations. Initially, 50 trees will be fit using recursive partitioning of the data after which residuals from the initial fit will be fit with another 50 trees, and so forth, until the model deviance is minimized. As recommended by Elith et al. (2008), we will test models with and without interactions while allowing the lr to vary so as to ensure that a minimum of 1000 trees are run. Environmental variables that contribute less than 5% to the model will be removed.

Based on the final BRT model for each species and season, we will generate spatial predictions. These maps will facilitate the visual assessment of habitat suitability. As with the RES predictions, we will derive the 50%, 75%, and 95% data contours for the habitat suitability prediction.

## Habitat suitability models – electronic tag data

Thirdly, we will use location data derived from a mix of bird-borne electronic tracking devices to assess the predictive performance of RES and BRT models derived from opportunistic sighting data. Location data from tracking devices have been collected and are available for all five study species: Table 1 summarises the number of tracked individuals, by species, for three different types of tracking data. GPS (Global Positioning System) are highly accurate and resolved data, typically locations are acquired every few minutes, but over relatively short-term deployments of days to weeks. PTT (Platform Transmitting Terminal) provide location data acquired via the ARGOS satellite array, which are relatively accurate and typically acquired at intervals of a few hours over deployments of weeks to months. GLS (Global Location Sensor) tracking tags provide long-term (months to years) data acquisition of lower accuracy and with only two location per 24-hour period.

**Table 1.** Numbers of individuals of the five study species for which tracking data are available. See text for explanation of tracking devices.

| Species | GPS | PTT | GLS |
|---|---|---|---|
| Campbell albatross | 81 | - | 73 |
| Grey-headed albatross | 53 | -- | 76 |
| White-capped albatross | 50 | 35 | 34 |
| Buller's albatross | 77 | 30 | 34 |
| Sooty shearwater | - | - | 28 |

We will use these data to test the predictive accuracy of both the RES and BRT models. Specifically, we will use Receiver Operator Characteristic (ROC) curves and the Area Under the Curve (AUC) to assess model performance. AUC values range from 0 (no discrimination ability between presence and absence values) and 1 (perfect discrimination) (Pearce & Ferrier, 2000). The AUC value will provide a measure of predictive ability and will allow us to compare the performance of the RES to BRT models and determine if a model informed by sightings data is superior to a model that is sightings-independent.

Finally, we will create a habitat suitability model for each species during the breeding and non-breeding season using only the data available from electronic tags. Similar to the opportunistic sightings data, we will use BRT models to fit a binary response variable (presence and absence) to a suite of environmental predictor variables. As such, we will create pseudo absence points using correlated random walks following the methods outlined in Goetz (2015). We will use the same methods outlined above for BRT models but will modify code from Crase et al. (2012) to account for residual spatial autocorrelation in the data.

Having determined the best BRT model created from the electronic tag data, we will create a spatial prediction map and determine the 50%, 75% and 95% data distribution contours. These contours will then be compared spatially and temporally (across seasons) with the contours from the RES (without sightings data) and BRT (with sightings data) model predictions. In this comparison, the predictions derived from the electronic animalborne tags are considered as close to the 'truth' as possible and the accuracy of the RES and BRT sightings models will be assessed based on how closely models predict habitat suitability relative to the electronic tagging data.

## H. RESEARCH ROLES

| Researcher | Organisation | Contribution |
|---|---|---|
| *Dr David Thompson* | *NIWA* | The development of distribution models and model comparison. |
| *Dr Kimberly Goetz* | *NIWA* | The development of distribution models and model comparison. |

## I. LINKAGES AND DEPENDENCIES

This project complements, rather than overlaps, research currently being conducted to support the over-arching Challenge Objective to enhance utilisation of marine resources within environmental constraints, and does not have any dependencies within the Challenge framework.

Specifically, this project proposed here will provide data and outputs that are synergistic to the theme of Sustainable Seas' Programme 5, Managed Seas. The Managed Seas Programme, will deliver innovative and effective decision support tools, which will integrate the knowledge generated by the Challenge to allow ecosystem based management, ensuring sustainable utilisation of our marine resources. While we are not proposing to develop an ecosystem based model, we will develop a tiered modelling approach to assess what additional data may be needed to accurately quantify the distribution of seabirds, a key component of all marine ecosystems in New Zealand. Furthermore, our proposed project will provide a template which will allow marine managers make better-informed decisions when faced with the sometimes conflicting requirements of seabird conservation and marine resource use.

## J. RISK AND MITIGATION

Because this project is primarily a desktop exercise, little risk is involved. The modelling described in this proposal draws upon data available in published sources, publicly available or published seabird sightings data and tracking data, which have been collected as part of previous NIWA research programmes or as part of programmes in which NIWA has been a key collaborator. In short, data availability is a very low risk to the successful outcome of this project.

## K. ALIGNED FUNDING AND CO-FUNDING

This project is aligned with an MPI-funded project (PRO2014-01, $109,900 NZD), which aims to improve our understanding of the distribution of marine mammals. This will be achieved by collating and examining the available spatial occurrence data of 47 potentially at-risk cetacean species and produce habitat suitability maps for those species with sufficient data. This information will be used to support future updates of the marine mammal risk assessment (PRO2012/02), which is currently being undertaken for the first time (Berkenbusch et al., 2013). Spatially comprehensive environmental data layers will be used to inform species distribution models in order to derive empirically-based protected species distribution maps for selected marine mammals.

There is a relatively large amount of synergy between the project outlined here and the MPI-funded marine mammal project. Furthermore, our project will directly add valuable information to this MPI funded project, especially for marine mammal species where little data are available. In these cases, models must be developed from a paucity of sightings data or from basic environmental parameters that are thought to determine marine mammal distributions directly or indirectly, as in the case of RES models. Knowing how well our seabird models preform using different levels of data quality will allow us to determine not only the level of confidence in our models but also where resources should be allocated to increase the level of data quality when engaging in ecosystem based modelling.

## L. VISION MĀTAURANGA (VM)

This proposal is likely to have broad appeal to Māori who regard seabirds as taonga species. There is likely to be considerable traditional knowledge of the five target species identified for inclusion in this project, knowledge that could inform the successful development of the RES models for each species. To do this we will aim to capture relevant information from published sources and will seek input from key informants and/or identified mana whenua representatives in the early stages of the project.

Our initial species selection was constrained largely by the availability of tracking data though the outputs from this project will be applicable to Māori with particular interests in species not initially covered in this research. We will seek input and guidance to potential future target species for this work, with the aim of providing a cost-effective solution to improving the cross-cultural understanding of seabird habitat use.

## M. CONSENTS AND APPROVAL

This project does not require any marine consents or ethics approvals.

## N. DATA MANAGEMENT

NIWA has an in-house IT department with expertise and proven capability in securely managing large quantities of data. The IT team will supply room on a dedicated server where project data can be securely stored and accessed.

## REFERENCES

Andersen, M. C., Watts, J. M., Freilich, J. E., Yool, S. R., Wakefield, G. I., McCauley, J. F., & Fahnestock, P. B. (2000). Regression-tree modeling of desert tortoise habitat in the central Mojave Desert. *Ecological Applications, 10*(3), 890-900.

Berkenbusch, K., Abraham, E. R., & Torres, L. (2013). New Zealand marine mammals and commercial fisheries *New Zealand Aquatic Environment and Biodiveristy Report No. 119* (pp. 110).

Compton, T. J., Bowden, D. A., Roland Pitcher, C., Hewitt, J. E., & Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand. *Journal of Biogeography, 40*(1), 75-89.
Crase, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography, 35*(10), 879-888.
Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813.
Goetz, K. T. (2015). Movement, Habitat, and Foraging Behavior of Weddell Seals (Leptonychotes Weddellii) in the Western Ross Sea, Antarctica.
Jenkins, J.A.F. (1981). Birds seen at sea around New Zealand during the 1978 BAAS expedition. Australasian seabird group newsletter no. 16: 3-16.
Kaschner, K., Tittensor, D. P., Ready, J., Gerrodette, T., & Worm, B. (2011). Current and future patterns of global marine mammal biodiversity. *Plos One, 6*(5), e19653.
Kaschner, K., Watson, R., Trites, A., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series, 316*(285-310), 2-3.
Leathwick, J., Elith, J., Francis, M., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series, 321*, 267-281.
Leduc, D., Rowden, A. A., Torres, L. G., Nodder, S. D., & Pallentin, A. (2015). Distribution of macro-infaunal communities in phosphorite nodule deposits on Chatham Rise, Southwest Pacific: Implications for management of seabed mining. *Deep Sea Research Part I: Oceanographic Research Papers, 99*, 105-118.
McQuaid, C.D.; Ricketts, L.H. (1984). Seabird records from the Transglobe Expedition, part 1: African and New Zealand sectors of the Southern Ocean, 1979-1981. Cormorant 12: 14-28.
Pearce, J., & Ferrier, S. (2000). An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling, 128*(2-3), 127-147. doi: Doi: 10.1016/s0304-3800(99)00227-6
R Development Core Team. (2013). R: A languate and enviromnment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org
Ridgeway, G. (2006). gbm: Generalized boosted regression models. *R package version, 1*(3).
Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. *Update, 1*(1), 2007.
Ridgeway, G. (2010). gbm: generalized boosted regression models. R package version 1.6-3.1.
Robertson, H.A., Dowding, J.E., Elliott, G.P., Hitchmough, R.A., Miskelly, C.M., O'Donnell, C.F.J., Powlesland, R.G., Sagar, P.M., Scofield, R.P., Taylor, G.T. (2013). Conservation status of New Zealand birds, 2012. New Zealand Threat Classification Series 4, Department of Conservation, Wellington, 22p.
Shaffer, S.A.; Tremblay, Y.; Weimerskirch, H.; Scott, D.; Thompson, D.R.; Sagar, P.M.; Moller, H.; Taylor, G.A.; Foley, D.G.; Block, B.A.; Costa, D.P. (2006). Endless summer: migratory shearwaters integrate oceanic resources across the Pacific Ocean. Proceedings of the National Academy of Sciences 103: 12799-12802.
Shaffer, S.A.; Weimerskirch, H.; Scott, D.; Pinaud, D.; Thompson, D.R.; Sagar, P.M.; Moller, H.; Taylor, G.A.; Foley, D.G.; Tremblay, Y.; Costa, D.P. (2009). Spatiotemporal habitat use by breeding sooty shearwaters Puffinus griseus. Marine Ecology Progress Series 391: 209-220.
Torres, L. G., Smith, T. D., Sutton, P., MacDiarmid, A., Bannister, J., & Miyashita, T. (2013). From exploitation to conservation: habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Diversity and Distributions, 19*(9), 1138-1152.

Watson, H.; Hiddink, J.G.; Hobbs, M.J.; Brereton, T.M.; Tetley, M.J. (2013). The utility of relative environmental suitability (RES) modelling for predicting distributions of seabirds in the North Atlantic. Marine Ecology Progress Series 485: 259-273.