**Appendix S1.** Additional methods and results describing sensitivity analyses undertaken.

To determine how each of the five climate models used (ACCESS1.0, BNU-ESM, CESM1-BGC, CSIRO Mk3.6.0, GFDL-ESM2M) altered predictions of climate suitability, compared against the 5-model average, ensemble-SDMs were built and projected across each climate dataset using default settings and a full ensemble of 10 SDMs. To test the robustness of the outcomes of the 10-model ensemble to variations in SDM modelling methods, a subset of the four most commonly used techniques (Generalized Linear Model, Generalized Additive Model, Boosted Regression Trees, Maximum Entropy) was taken from the full 10-model ensemble on averaged climate data and SDM ensemble analyses and projections were repeated.

Additionally, we tested the robustness of our models with regard to using default settings, as the use of these for SDMs may risk overfitting. Ideally, SDMs should be fine-tuned for each species modelled, but, when working with a large number of species, simple modifications to settings can reduce overfitting (Radosavljevic & Anderson, 2014), which may decrease model performance measured against evaluation data but increase transferability (Wenger & Olden, 2012). Modelling was, therefore, also repeated again using slightly modified settings (using BIC instead of AIC for stepwise GLM procedure and increasing the β-penalisation for Maximum Entropy modelling to 3; Radosavljevic & Anderson, 2014) to see if and how such modifications would adjust model outcomes.

Owing to climate value differences between averaged and individual climate models, and across all five individual climate models, the exact number of species for which the climate of each island was identified as suitable did vary depending upon the climate model examined (Tables 1 and 2). Results of performing SDM-analyses on each climate model are presented, in detail, in Tables S3-S18. Adjusting model settings (Table S10) and taking a subset of four of the 10-model ensemble-SDM also altered species-specific model outcomes (Table S9). The absolute maximum difference in number of species for which any island was suitable between the ensemble model of the 69 species from the 100 worst list using default settings and the ensemble model using modified settings was four (Auckland Island, current conditions). This was the exception rather than the norm and the mean absolute difference between default settings and modified setting ensemble predictions across all islands and years was 0.75 species. Differences between predictions of the full 10-model ensemble and the four-model subset were also relatively small (absolute maximum = 6 species, Campbell Island, 2050; absolute mean = 1.94 species).

Radosavljevic, A. and Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of Biogeography, 41, 629–643. doi:10.1111/jbi.12227

Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3, 260–267. doi:10.1111/j.2041-210X.2011.00170.x