



## **MPI 18607 Project Report**

### **Species distribution models of the native New Zealand Myrtaceae**

Biosecurity New Zealand Technical Paper No: 2019/22

Prepared for the Ministry for Primary Industries

By James K. McCarthy<sup>1</sup>, Sarah J. Richardson<sup>1</sup>, Jerry A. Cooper<sup>1</sup>, Peter J. Bellingham<sup>1</sup>, Susan K. Wiser<sup>1</sup>

<sup>1</sup>Manaaki Whenua Landcare Research Limited

ISBN No: 978-0-9951272-6-5

ISSN No: 2624-0203

**July 2019**

# Disclaimer

While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

Requests for further copies should be directed to:

Publications Logistics Officer  
Ministry for Primary Industries  
PO Box 2526  
WELLINGTON 6140

Email: [brand@mpi.govt.nz](mailto:brand@mpi.govt.nz)  
Telephone: 0800 00 83 33  
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries website at <http://www.mpi.govt.nz/news-and-resources/publications/>

© Crown Copyright - Ministry for Primary Industries

# Executive summary

## The problem

Myrtle rust was discovered in New Zealand on Raoul Island in April 2017, and a month later it was found in Northland. New Zealand has 27 recognised native species of Myrtaceae that may be susceptible to myrtle rust, but little work has been done to understand the distribution of these species at a national scale. This is especially true for the distributions of the new species described in the recent taxonomic revision of kānuka (*Kunzea* spp.), whose distributions are only currently known from limited point records. The objective of this study was to develop national-scale species distribution models for all native Myrtaceae species using a consistent, standardised approach to predict the likelihood of finding an individual Myrtaceae species at a particular location, and to understand what environmental variables drive their national distributions.

## Key results

Using boosted regression trees as our modelling method, we parameterised models for 22 Myrtaceae species that occur on New Zealand's three main islands, surrounding inshore islands, and a limited number of offshore islands. From these models we predicted each species' distribution at a fine (100 m) resolution. The remaining five species could not be modelled effectively because they either occurred on offshore islands for which critical environmental data are not available, had too few records in our occurrence data, or genuinely occur very infrequently in the landscape.

The models generally performed well, with correlation coefficients ranging from 0.58 to 0.80 or 0.26 to 0.85, depending on the type of species observation data used. Climate variables were the most informative predictors for most species, but landscape and soil variables also contributed. The predicted distributions superficially followed the known distributions for most species, including depicting the northern populations of southern rātā (i.e. on the Coromandel Peninsula), which previous modelling approaches have found challenging to replicate. Species distribution models were also produced for seven (of the ten) recently revised/described species of kānuka for the first time. The predicted distributions for all species are presented in this report as separately supplied spatial data files, and through an online visualisation interface ([https://landcare.shinyapps.io/Myrtaceae\\_SDM\\_WebApp/](https://landcare.shinyapps.io/Myrtaceae_SDM_WebApp/)).

## Implications of results for the client

These predictions have a range of potential applications. First, they can be used, either on their own or in conjunction with available climate-based myrtle rust risk maps, to identify where the disease is likely to spread. They can also be used to prioritise areas for further conservation, management or regular monitoring. Having a nationwide prediction for species distributions can also be used to inform seed collection and seed banking efforts by identifying sampling locations which could be used to more effectively capture genetic diversity.

## Further work

These predictions could be further refined through an increase in the quality and representation of the input data, as follows.

- Species locational data could be greatly improved through additional plot-level sampling of under-sampled areas (additional analyses are required to identify these areas), collecting additional point-level records (i.e. herbarium samples, citizen science identifications), or digitising herbarium and plot survey data that have not yet been digitised.
- Environmental predictor data could be greatly improved and expanded through the generation of additional predictor layers characterising disturbance history, quantitative soil measurements such as pH and soil carbon, and remotely sensed variables.

The utility of these predictions could also be improved through integration with existing climate-based myrtle rust risk maps, and through a better understanding of species-level myrtle rust susceptibility.

Finally, these maps are fit for use at the national scale from which they were developed. Accurate regional- or local-scale maps may require further, detailed sampling and more tailored models, or alternative modelling approaches.

<b>Contents</b>		<b>Page</b>
<b>1</b>	<b>Project background</b>	<b>6</b>
<b>2</b>	<b>Introduction</b>	<b>7</b>
<b>3</b>	<b>Materials and methods</b>	<b>9</b>
3.1	An introduction to boosted regression trees	9
3.2	Presence–absence models vs presence-only models	9
3.3	Study region	10
3.4	Species data	11
3.5	Environmental spatial predictors	14
3.6	Model fitting, prediction and evaluation	15
	<b>Results</b>	<b>17</b>
<b>4</b>	<b>Results and discussion</b>	<b>21</b>
4.1	Predicted distributions	21
4.2	Presence–absence and presence-only predictions	21
4.3	Limitations and assumptions	21
4.4	Conclusions and future research	22
<b>5</b>	<b>Acknowledgements</b>	<b>24</b>
<b>6</b>	<b>References</b>	<b>25</b>
<b>7</b>	<b>Appendix B</b>	<b>34</b>
<b>8</b>	<b>Appendix C</b>	<b>38</b>
<b>9</b>	<b>Appendix D</b>	<b>42</b>
9.1	<i>Kunzea amathicola</i>	43
9.2	<i>Kunzea ericoides sensu lato</i>	45
9.3	<i>Kunzea ericoides sensu stricto</i>	47
9.4	<i>Kunzea linearis</i>	49
9.5	<i>Kunzea robusta</i>	51
9.6	<i>Kunzea serotina</i>	53
9.7	<i>Kunzea sinclairii</i>	55
9.8	<i>Kunzea tenuicaulis</i>	57
9.9	<i>Leptospermum scoparium</i>	59
9.10	<i>Lophomyrtus bullata</i>	61
9.11	<i>Lophomyrtus obcordata</i>	63
9.12	<i>Metrosideros albiflora</i>	65
9.13	<i>Metrosideros carminea</i>	67
9.14	<i>Metrosideros colensoi</i>	69
9.15	<i>Metrosideros diffusa</i>	71
9.16	<i>Metrosideros excelsa</i>	73
9.17	<i>Metrosideros fulgens</i>	75
9.18	<i>Metrosideros parkinsonii</i>	77
9.19	<i>Metrosideros perforata</i>	79
9.20	<i>Metrosideros robusta</i>	81
9.21	<i>Metrosideros umbellata</i>	83
9.22	<i>Neomyrtus pedunculata</i>	85
9.23	<i>Syzygium maire</i>	87
<b>10</b>	<b>Appendix E</b>	<b>89</b>

# 1 Project background

To better understand myrtle rust and limit its impact in New Zealand, the Ministry for Primary Industries commissioned a comprehensive research programme in 2017 with more than 20 projects valued at over \$3.7 million. Projects in this programme were completed by June 2019.

The projects covered research in the following themes:

- Theme 1 - Understanding the pathogen, hosts, and environmental influence.
- Theme 2 – Building engagement and social licence: Improved understanding of public perceptions and behaviours to allow better decisions about investment, improved design of pathway control strategies and maintain social license for use of management tools.
- Theme 3 – Te Ao Māori: Greater understanding of Te Ao Māori implications of myrtle rust in order to support more effective investments, and improved use of Mātauranga, specific Māori knowledge, and kaupapa Māori approaches in management regimes.
- Theme 4 – Improving management tools and approaches: Improved diagnostic and surveillance speed, accuracy and cost-effectiveness, supporting eradication efforts and enabling scaling up of surveillance efforts for a given resource. More effective treatment toolkits to avoid emergences of MR resistance to treatments and to enable disease control over increasingly large scales that will lead to reduced or avoided impacts.
- Theme 5 - Evaluating impacts and responses: Improved understanding of environmental, economic, social and cultural, impacts to inform risk assessment and management and to communicate implications to decision/makers and stakeholders.

This report is part of the MPI commissioned research under contract MPI18607 which addressed research questions within Theme 2, 4 and 5.

Text in the report may refer to other research programmes carried out under the respective theme titles.

## 2 Introduction

Myrtle rust is a disease caused by the fungus *Austropuccinia psidii* (formerly *Puccinia psidii*, Beenken 2017), which infects species within the plant family Myrtaceae. The fungus is native to South and Central America and was restricted to those areas and neighbouring Caribbean islands until 1977, when it was detected on planted trees in Florida (Marlatt & Kimbrough 1979). Further spread was minimal until it was detected in Hawaii in 2005 (Killgore & Heu 2007), after which it was found in Japan in 2007 (Kawanishi et al. 2009), Australia in 2010 (Carnegie et al. 2010), and South Africa in 2013 (Roux et al. 2013). Unlike most fungal pathogens, myrtle rust is unusual because it has a very wide host range within the Myrtaceae family. This makes it a significant threat to forests world-wide, such as those in Australia, which is home to over half the world's 3,000 Myrtaceae species (Glen et al. 2007), and in New Zealand, which has many functionally and culturally important Myrtaceae.

After the disease was detected in Australia, New Zealand officials feared its possible arrival (Clark 2011; Teulon et al. 2015), and in April 2017 the disease was detected on Raoul Island infecting a species of pōhutukawa endemic to the Kermadec Islands (Kermadec pōhutukawa, *Metrosideros kermadecensis*). It is presumed the fungus was carried from Australia by wind (Beresford et al. 2018). In May 2017 the disease was first reported by a nursery in Northland and shortly after this it was also found in Taranaki and the Bay of Plenty. Within 12 months it spread across most of the North Island and in April 2018 reached the Nelson–Marlborough regions of the South Island (Beresford et al. 2018).

Due to its recent arrival little is known about which of the 27 Myrtaceae species (Breitwieser et al. 2019) native to New Zealand are likely to be seriously affected by the disease. At present, species within the New Zealand endemic genera *Lophomyrtus* and *Neomyrtus* are considered most vulnerable due to their close relationship to susceptible species from Australia (Carnegie et al. 2016; de Lange et al. 2018). In accordance with this uncertainty, a conservative approach has been adopted and in 2018 all native Myrtaceae species were elevated to at least 'Threatened' status in the most recent revision of the conservation status of New Zealand's indigenous plants (de Lange et al. 2018).

In Australia severe cases of myrtle rust have caused local species extinctions, which, given the short timeframe since infection, are likely to culminate in rapid and fundamental changes to plant community structure (Pegg et al. 2017). The degree to which New Zealand forests are dominated by Myrtaceae varies, but many forest types are characterised by dominant and emergent Myrtaceae trees, with previous Myrtaceae dieback events significantly altering forest structure and composition (e.g. Allen & Rose 1983). Examples of New Zealand ecosystems dominated by Myrtaceae include widespread mānuka and kānuka (*Kunzea ericoides sensu lato*) shrublands, often (but not exclusively) present in recently disturbed environments; pōhutukawa (*Metrosideros excelsa*) forests found on the northern coastlines; and southern rātā-kāmahi forests of the South Island west coast. Native and exotic Myrtaceae species are also popular choices for garden, street and restoration plantings, and several native Myrtaceae are considered taonga (treasured, sacred) in Māori culture (Lambert et al. 2018). Significant dieback of Myrtaceae species is likely to have an impact on forest succession, ecosystem functionality, amenity, and cultural values. Many commercial sectors are also reliant on Myrtaceae species; for example, mānuka is vital to the New Zealand honey industry (Ministry for Primary Industries 2018), and *Eucalyptus* species are planted by the forest industry for timber (Forest Owners Association 2018).

The development of the response by New Zealand officials to myrtle rust is ongoing, but includes a range of activities such as monitoring various populations likely to be susceptible, controlling the spread of the disease, and collection and preservation of seed for conservation efforts. Future efforts may include species translocation, identification of isolated regions for refugia, or additional conservation protection for particularly important forests. These activities require a detailed knowledge of the distribution of Myrtaceae species across the landscape. New Zealand has a substantial network of vegetation survey plots, herbarium collections and citizen-science records that can be used to generate range maps by modelling. Species distribution models (SDMs) predict species' geographical ranges from occurrence records (species presence, or presence/absence) and associated environmental data.

Maps of New Zealand native Myrtaceae species have already been produced as part of the response to myrtle rust by assigning occurrence records from various sources to units based on land cover and ecological districts (Wiser et al. 2017). These were produced in a short timeframe to support the immediate response without directly considering the environmental gradients present in

New Zealand or quantifying distributions across a continuous scale. Here we build on this previous work by using SDMs to produce range maps for all native species within the Myrtaceae family that occur on the New Zealand mainland and surrounding inshore islands, including the recent taxonomic revisions to kānuka (*Kunzea ericoides sensu lato*) (de Lange 2014). SDMs based solely on environment predict potential distributions; here we attempt to produce predictions of 'actual occurrence' by clipping predictions to known species-level range limits. We use boosted regression trees, a machine-learning approach that performs well compared to many other popular SDM methods (Elith et al. 2006).

## 3 Materials and methods

### 3.1 An introduction to boosted regression trees

Most contemporary species distribution modelling techniques share common conceptual and technical underpinnings: they work by relating species occurrence data to information about the environment (Elith & Leathwick 2009). These models can be used to understand the relationship between a species and the environment, and, with sufficient spatial data, predict a species' distribution across a landscape.

Many different analytical approaches are used to fit SDMs (Franklin 2009), each with its own advantages and drawbacks. Here we use boosted regression trees (BRTs; Friedman 2001), a technique that ecologists began using to predict species distributions in the mid-2000s (e.g. Elith et al. 2006; Leathwick et al. 2006). The approach is known to perform well when used for prediction (Elith et al. 2006; Guisan et al. 2007), and has been in wide use by ecologists since publication of a working guide to the technique (Elith et al. 2008). Unlike SDM techniques with statistical origins (e.g. generalised linear models, generalised additive models, generalised dissimilarity models), BRTs are a machine-learning method whereby an algorithm 'learns' the nature of the relationship between a response (species data) and predictors (environment) (Breiman 2001). The popular SDM technique Maxent is also a machine-learning approach (Phillips et al. 2006).

BRTs involve fitting many simple models, which are then combined. This contrasts with statistical techniques, where a single, parsimonious model is fitted to parameterise the relationship between a response and its predictors. In BRTs, each simple model consists of a decision tree, which 'partitions' predictor space using a series of recursive binary splits, or rules (Elith et al. 2008). Decision trees are popular because they are intuitive and can handle any form of data (numeric, categorical, binary), they automatically capture non-linear responses, and they are not affected by extreme outliers or differing scales of measurement among predictors. Due to the data-driven nature of the technique, BRTs are insensitive to the inclusion of non-relevant predictors, because they are infrequently selected by the model. The hierarchical structure of each decision tree means that the response of one variable depends on those higher in the tree, so complex interactions are automatically captured (the number of steps in a decision tree is referred to as 'tree size'). However it should be noted that individual trees, and approaches that do not incorporate an ensemble of trees, are inherently poor at prediction (Elith et al. 2008). With BRTs this is overcome by 'boosting', which involves sequentially fitting an ensemble of thousands of individual trees (or models), with each iteration focusing on relationships that are poorly captured by the existing trees.

Given the flexibility and data-driven nature of the BRT approach, the method can be prone to overfitting. This occurs when trees continue to be added to the point where, eventually, all observations are perfectly explained. Since most modelling exercises aim to have a high level of generality (Hastie et al. 2001), procedures can be implemented to ensure that overfitting does not occur (Elith et al. 2008). Furthermore, individual component decision trees with more than two predictors automatically fit interactions. These interactions can be both difficult to detect and difficult to interpret due to the sheer number of trees in most models. The automatic fit of interactions also means that pairwise relationships between an individual variable and the predicted response can be difficult to interpret because their form can be influenced dramatically by the values of other predictors that are not shown in pairwise comparisons.

### 3.2 Presence–absence models vs presence-only models

Typically, species distribution models work by correlating not just species presences but also species absences, with relevant environmental covariates. In plant ecology, species absences can usually be inferred from plot-based systematic surveys, whereby all species in a small, defined area (e.g. 400 m<sup>2</sup>) are recorded. Species that were not recorded are assumed to have been absent. In plots, species may be recorded with some measure of abundance, but for species distribution models these values are usually reduced to 1 (species present) or 0 (species absent), and the models applied are called 'presence–absence' models.

Unfortunately, plot data tend to be sparsely spread across areas of interest, potentially missing the most optimal environments and range limits. To overcome this, ecologists often draw upon

occurrences associated with museum and herbarium collections, and, increasingly, from citizen science initiatives (Schmeller et al. 2009). While these data convey known presences, are often plentiful, and, through initiatives such as GBIF<sup>1</sup>, are usually straightforward to obtain, they lack information about confirmed absences. Models that use these data are often referred to as 'presence-only' models. To overcome the lack of true absences, ecologists often compare presences with a sample of the entire region's environment, referred to as the 'background' (replacing the absences from presence-absence models, e.g. Ferrier et al. 2002). Several techniques are used to create a background sample, and how best to do this is often debated (e.g. Phillips et al. 2009; Barbet-Massin et al. 2012).

A recent review found that there was no clear preference between presence–absence and presence-only models, with 47% and 53% of peer-reviewed species distribution modelling papers using each technique, respectively (Guillera-Arroita et al. 2015). Presence–absence models are considered the richest type of model in terms of information content because they compare environments between locations where a species was found and where it was genuinely absent, but absent plot data often prevents their use (Guillera-Arroita et al. 2015). These data can be used to distinguish whether a species is rare and well surveyed, or common but under-surveyed, and enable models to predict *probability of species occurrence*. Values predicted give an indication of prevalence, which in species distribution models are bound between 0 and 1, indicating the probability of observing a species at a randomly chosen site (Phillips et al. 2009). With presence–absence models, predictions are on a common scale across species, allowing direct comparison (Guillera-Arroita et al. 2015), and they can be combined for community-level analyses.

Presence-only models may be supported by more data, but predictions are less flexible in their application. These models do not estimate actual probabilities, but a *relative likelihood of species occurrence or observation* (Guillera-Arroita et al. 2015). This means that it is not possible to identify whether a species is rare and well surveyed or common but under-surveyed, and predictions of relative likelihood are not comparable across species, ruling out further community-level analyses. Because the data applied in presence-only models are usually collected opportunistically rather than through a designed survey scheme, occurrences are weighted toward areas that systematists, ecologists or members of the public frequently visit, such as urban areas, popular national parks, or near roads or tracks, introducing geographical bias (Dennis & Thomas 2000; Schulman et al. 2007). Search effort is also not standardised across locations (observation bias), and observers may not report all species present at a location, but rather those they find interesting (reporting bias) (van Strien et al. 2013). However, the advantages provided by presence-only models (incorporation of additional data, often at range limits) cannot be discounted, and for some under-sampled species their use cannot be avoided. It can also be useful to compare predictions from both models. Hybrid approaches are emerging (e.g. Fithian et al. 2015; Ovaskainen et al. 2016; Wilkinson et al. 2019) but have not been fully developed at this stage.

Here, for species with sufficient presence–absence data, we adopt a combined approach whereby presence–absence models are used to predict occurrence probability, and presence-only models are used to identify potential range limits. Some New Zealand Myrtaceae species are so rare, or, in the case of the recent *Kunzea* spp. revision (de Lange 2014), so recently described that they are not adequately represented in the New Zealand network of vegetation plots for presence–absence analyses (Table 1). In these cases, we produce presence-only models, only. See 'Model fitting, prediction and evaluation', below, for a technical description of our modelling approach.

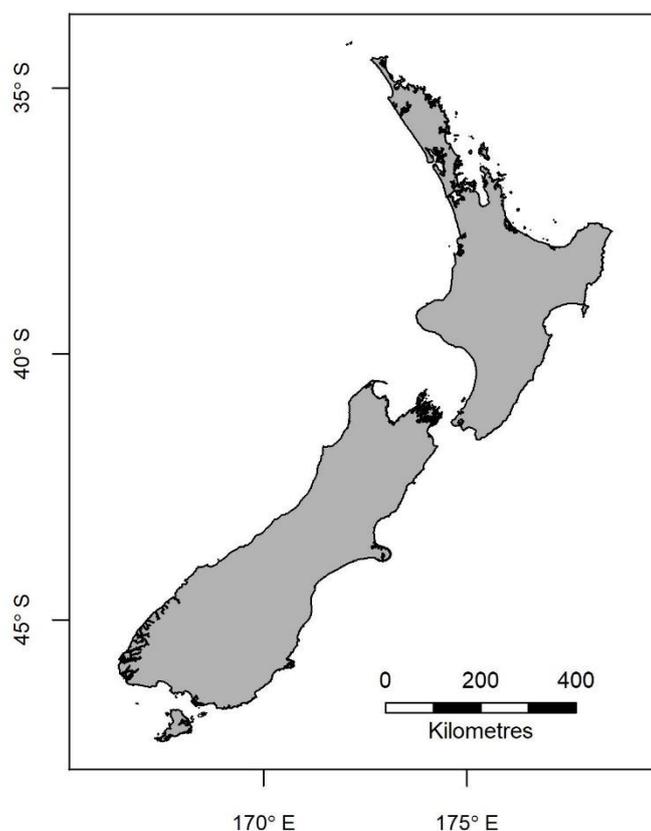
### 3.3 Study region

For this report, predictions were restricted to the extent of New Zealand within which reliable and consistent environmental predictor data were available. As a result, we constrained our study region to the coverage of Land Environments of New Zealand (LENZ), the country's most extensive characterisation of environment, covering 16 climate, landform, and soil attributes (Leathwick et al. 2002). This data set covers the three main islands (North Island, South Island, Stewart Island), surrounding inshore islands, and a limited number of offshore islands (Figure 1). Inshore and offshore islands considered in the study include the Three Kings Islands, islands in the Hauraki Gulf (including Great Barrier, Little Barrier, Waiheke and Rangitoto Islands), the Marlborough

---

<sup>1</sup> The Global Biodiversity Information Facility (GBIF) is an international collaborative effort to provide open access to species locational data (see <https://www.gbif.org/>). There are a range of R packages that can be used to interrogate GBIF data directly (e.g. *rgbif*, Chamberlain et al. 2018).

Sounds islands, among others (Figure 1), but unfortunately the Kermadec Islands, the Chatham Islands and the subantarctic islands were excluded.



**Figure 1.** Predictions of Myrtaceae species distributions were restricted to the extent of the country's primary source of spatially complete environmental predictors, the Land Environments of New Zealand (LENZ; Leathwick et al. 2002). This includes New Zealand's three main islands, and several inshore and offshore islands (see main text), but excludes the Kermadec, Chatham, and Auckland Island groups.

### 3.4 Species data

There are 27 species within the family Myrtaceae family that are recognised as being native to New Zealand. This number includes the revision to kānuka (the genus *Kunzea*) that increased the number of distinct species from three to ten because of revision to the *K. ericoides* complex (de Lange 2014). These 27 species belong to five genera: *Kunzea* (ten species), *Leptospermum* (one species), *Lophomyrtus* (two species), *Metrosideros* (twelve species), *Neomyrtus* (one species), and *Syzygium* (one species).

The species occurrence records used to parameterise our species distribution models were extracted from the following five sources:

- the **New Zealand National Vegetation Survey Databank (NVS)** – New Zealand's primary archive for storing vegetation survey plot data (Wiser et al. 2001)
- **BioWeb** – a proprietary database administered by the Department of Conservation, which holds natural heritage data of importance to the Department
- **iNaturalist** – formerly known as NatureWatch, a citizen science project that allows the public to map and share observations of biodiversity world-wide (<https://www.inaturalist.org>)
- the **Australasian Virtual Herbarium (AVH)** – an online resource that provides access to digitised records of specimens held in most of New Zealand's herbaria (<https://avh.chah.org.au/>)

- **de Lange (2014)** – the publication detailing the recent split of kānuka (*Kunzea ericoides sensu lato*) containing maps of herbarium specimens that have been renamed following the updated taxonomy; records were also retrieved from the appendices of an associated thesis (de Lange 2007).

Given that model accuracy is known to decline severely with fewer than 30 presence observations (Hernandez et al. 2006; Guisan et al. 2017, p. 116), we use this as a cut-off for any given species, below which we did not attempt any predictions. Also, some species either fell outside our study region or are too rare to produce useful predictions. Of the 27 Myrtaceae species native to New Zealand, we were unable to produce predictions for five species. *Kunzea salterae* (found on islands in the Bay of Plenty), *K. toelkenii* (found on the Bay of Plenty coast) and *K. triregensis* (endemic to the Three Kings Islands) all had fewer than 25 occurrence records, probably due to unfamiliarity with the updated taxonomy among botanists and the public, their restricted range, and the limited time for these species to be captured in NVS plots. Until this data deficiency improves, the maps in de Lange (2014) provide a useful indication of these species' distributions. *Metrosideros bartlettii* (Bartlett's rātā) is an extremely rare tree, confined to three small forest fragments in North Cape (Dawson 1985; Drummond et al. 2001). We identified 128 occurrence records for this species, but with only around 30 trees known to exist (Drummond et al. 2001), there has probably been considerable oversampling of these individuals. Given these drawbacks, a prediction would not be useful for this species. Finally, *Metrosideros kermadecensis* (Kermadec pōhutukawa) is endemic to the Kermadec Islands, which falls outside our study region. This left 22 species for our analyses (Table 1).

### 3.4.1 Presence–absence

**NVS:** Presence–absence (plot) data were extracted from the NVS databank, which contains data from over 100,000 plots. These include those of the Land Use and Carbon Analysis System (LUCAS, administered by the Ministry for the Environment) and the National Biodiversity Monitoring and Reporting System (administered by the Department of Conservation), which consistently sample New Zealand forests and shrublands at intersections of an 8 km grid (Allen et al. 2003); and the National Forest Survey of New Zealand, consisting of approximately 14,000 plots measured over approximately two decades from the late 1940s (Thomson 1946; Cunningham 1953).

A data use request was submitted on 16 April 2018 for all available data containing any New Zealand native species of Myrtaceae. All data from plots to which we were granted access and that fit our criteria (see below) were retrieved on 7 September 2018. Plots were then examined to ensure they met certain criteria:

- they measured full floristics (all plant species)
- they had no experimental treatments applied (e.g. grazing exclusion, herbicide application)
- if a plot was permanent, we used data from the most recent survey only.

For the *Kunzea ericoides sensu lato* model (see Table 1), we combined records from all new species described in de Lange (2014). Data from 60,965 plots were employed for our presence–absence analyses.

### 3.4.2 Presence-only

**NVS:** For our presence-only analyses, all Myrtaceae presences captured for presence–absence analyses from NVS (described above) were reduced to presence-only records and included. Myrtaceae records from sites where full floristics were not measured were also included. In addition to the presence–absence NVS sites, an additional 2,486 sites were extracted for presence-only analyses.

**BioWeb:** BioWeb data were provided as a shapefile from the Department of Conservation. Almost all these data are from the North Island, primarily the west coast and Bay of Plenty. Records from cultivation were removed if the word 'cultivated' appeared in the associated notes of any record. A total of 985 records were extracted from BioWeb for our presence-only analyses.

**iNaturalist:** Records of Myrtaceae were downloaded from the New Zealand iNaturalist data repository by Jon Sullivan (Lincoln University) on 9 July 2018. We only considered records that

achieve research grade status, which means their identity has been verified by experts, the record is georeferenced with a date, and it isn't of a captive or cultivated organism (iNaturalist 2018). After examining the various notes fields of the iNaturalist data it was clear that some records tagged as research grade were actually cultivated. As a further check we manually removed records that had any reference to being planted or cultivated. A total of 1,372 records from iNaturalist were used for our presence-only analyses.

**AVH:** These data were downloaded directly from the AVH database through the Integrated Publishing Toolkit<sup>2</sup> containing all Myrtaceae records set up by Aaron Wilton (Manaaki Whenua – Landcare Research) on 14 August 2018 (New Zealand Virtual Herbarium Network 2018). To remove potentially cultivated records, records with the following terms were removed from the data: 'cultivated', 'cultivation', 'gardens', 'nursery', 'plantation', 'planted', 'shrubbery', and 'urban'. Data from the New Zealand Forest Research Institute collection (hosted by Scion) were provided with integer (rounded) values for latitude and longitude. The potential error of up to 0.5 degrees (c. 50 km) of latitude or longitude introduced too much uncertainty for our purposes, and so the Scion data were excluded. A total of 3,918 records from the NZVH were used for our presence-only analyses.

de Lange (2014): Due to the recent reclassification of *Kunzea ericoides sensu stricto*, and the retention of the original name in the new classification, we ran two models for *K. ericoides*: one for *K. ericoides sensu lato*, and one for *K. ericoides sensu stricto* (Table 1). Because we are not confident that this reclassification will be applied consistently to all post-2014 records, we adopted a conservative approach and used only the *K. ericoides sensu stricto* records examined in de Lange (2014) for the *K. ericoides sensu stricto* presence-only model. Records from an associated thesis by the same author were also used (de Lange 2007). A total of 81 records were used for this model.

### 3.4.3 Post-processing

**Location:** We applied several filters to remove records with coordinate locations that were potentially errors. Since we were predicting species distributions at high resolution (100 m), we first excluded plots with a coordinate uncertainty of greater than 50 m. Records without any coordinate information were also removed.

Second, to remove potentially cultivated records in BioWeb, iNaturalist and NZVH data, we removed records that fell in the following urban or predominantly exotic land cover categories in the New Zealand Land Cover Database (LCDB) Version 4.1 (Manaaki Whenua – Landcare Research 2018): 'Built-up Area (settlement)', 'Urban Parkland/Open Space', 'Transport Infrastructure', 'Surface Mine or Dump', or 'Deciduous Hardwoods'. In order to prevent this filter removing records from genuine forest fragments in urban areas (for example, both Riccarton Bush, a forest fragment, and the Christchurch Botanic Gardens are mapped in LCDB as Deciduous Hardwoods), we ensured that points that fell within the Department of Conservation Protected Areas layer (Department of Conservation 2018) remained in our data set regardless of their LCDB class.

We also removed points that were greater than 100 m away from the land-based raster cells of our study region. If a point did not fall in a land-based cell but was less than 100 m away from land, we assigned that point the coordinates of the centroid of the nearest land-based cell using the `nearestLand` function in the `seegSDM R` package (SEEG Research Group 2018). As a final check, we manually removed records that were clearly beyond the species range limits through consultation with local botanists (see Acknowledgments) and reliable literature sources.

**Taxonomy:** To harmonise differences in taxonomy, and to update the names of old records, we used the `TPL` function in the `Taxonstand R` package, which standardises plant names using The Plant List<sup>3</sup> (Cayuela et al. 2012). Taxonomic categories below the species level (subspecies, variety, etc.) were grouped under their species name, and records at genus level or higher were excluded.

---

<sup>2</sup> An internet publishing toolkit is a Java-based software tool that supports data harvesting from the GBIF network.

<sup>3</sup> The Plant List (<http://www.theplantlist.org/>) is a working list of 'accepted names' for all known plant species, compiled by the Royal Botanic Gardens, Kew, and Missouri Botanic Garden.

**Table 1.** The distributions of 22 New Zealand-native Myrtaceae species were predicted in this report. Depending on the occurrence data available, distributions were predicted using boosted regression trees and presence–absence (PA) and/or presence-only (PO) data. Five data sources were interrogated for these occurrence data: de Lange (2014), BioWeb, iNaturalist, The New Zealand National Vegetation Survey Databank (NVS), and the New Zealand Virtual Herbarium (NZVH).

<i>Species</i>	<i>Number of records</i>					
	<i>Models</i>	<i>de Lange</i>	<i>BioWeb</i>	<i>iNaturalist</i>	<i>NVS *</i>	<i>NZVH</i>
<i>Kunzea amathicola</i>	PO	0	0	8	2 (2 + 0)	63
<i>Kunzea ericoides s.l.</i>	PA & PO	0	15	172	3927 (3704 + 223)	856
<i>Kunzea ericoides s.s.</i>	PO	81	0	0	0	0
<i>Kunzea linearis</i>	PO	0	6	17	1 (1 + 0)	140
<i>Kunzea robusta</i>	PO	0	0	96	75 (46 + 29)	269
<i>Kunzea serotina</i>	PO	0	0	7	10 (10 + 0)	103
<i>Kunzea sinclairii</i>	PO	0	17	2	0	70
<i>Kunzea tenuicaulis</i>	PO	0	1	2	84 (0 + 84)	52
<i>Leptospermum scoparium</i>	PA & PO	0	0	306	5347 (5163 + 184)	546
<i>Lophomyrtus bullata</i>	PA & PO	0	3	63	533 (514 + 19)	171
<i>Lophomyrtus obcordata</i>	PA & PO	0	19	54	696 (694 + 2)	177
<i>Metrosideros albiflora</i>	PA & PO	0	0	5	263 (232 + 31)	123
<i>Metrosideros carminea</i>	PO	0	256	6	10 (10 + 0)	191
<i>Metrosideros colensoi</i>	PA & PO	0	2	27	348 (341 + 7)	195
<i>Metrosideros diffusa</i>	PA & PO	0	2	115	11081 (10957 + 124)	271
<i>Metrosideros excelsa</i>	PA & PO	0	4	87	311 (306 + 5)	186
<i>Metrosideros fulgens</i>	PA & PO	0	1	117	5869 (5802 + 67)	248
<i>Metrosideros parkinsonii</i>	PA & PO	0	15	1	109 (109 + 0)	84
<i>Metrosideros perforata</i>	PA & PO	0	5	120	5345 (5268 + 77)	258
<i>Metrosideros robusta</i>	PA & PO	0	528	91	3515 (3483 + 32)	223
<i>Metrosideros umbellata</i>	PA & PO	0	0	91	10002 (9918 + 84)	282
<i>Neomyrtus pedunculata</i>	PA & PO	0	2	69	8471 (8304 + 167)	140
<i>Syzygium maire</i>	PA & PO	0	80	34	130 (121 + 9)	146

\* If measured in standardised plot sizes (around 400 m<sup>2</sup>), data from NVS were used for presence–absence (PA) analyses. Occurrence records from plots of other sizes were only used for presence only (PO) models. The total NVS occurrences and the number of occurrences under each category are presented as: total (PA + PO).

### 3.5 Environmental spatial predictors

We focused on constructing parsimonious models by selecting environmental predictors that we hypothesised would be important drivers of Myrtaceae species distributions. Nineteen uncorrelated (<0.7 absolute Pearson's correlation), spatially complete environmental variables that covered our study were selected, of which nine were climate, seven were substrate, and three were landform variables (**Error! Reference source not found.**). Most of these predictors were obtained directly from the LENZ data set (Leathwick et al. 2002), except for the landform variables, which were generated using the digital elevation model from LENZ and various R packages (**Error! Reference source not found.**). All variables were analysed at 100 m spatial resolution.

### 3.6 Model fitting, prediction and evaluation

Boosted regression trees (BRTs) were constructed using the *dismo* package (Hijmans et al. 2017) in R 3.4.2 (R Core Team 2018), with adapted code from Elith et al. (2008). We ran presence–absence and presence-only models for each species wherever possible, but poor representation of some species in the presence–absence data set meant that only presence-only models could be fitted for a subset of species (Table 1). For the background sample in the presence-only models we used a target group background approach, whereby the background sample comprised the occurrences of all remaining Myrtaceae species (the ‘target group’), with the assumption that they share a similar geographical and reporting bias to the occurrence data (Phillips et al. 2009). This approach has been shown to perform better than random generation of background locations (Mateo et al. 2010).

When using BRTs, the user can adjust four main settings to optimise the model:

- *learning rate*, a proportion between 0 and 1 that controls the contribution of each tree to the growing model
- *number of trees*, the number of trees in the model
- *tree complexity*, the ‘depth’ or number of nodes of each tree
- *bag fraction*, the proportion of data to be selected at each step (some stochasticity is introduced to BRTs to improve accuracy and speed, and to reduce overfitting).

The function `gbm.step` (from *dismo*, Hijmans et al. 2017) is a cross-validation technique that assesses the optimum number of trees using k-fold cross-validation (as described in Hastie et al. 2001, p. 215). Using this function, all combinations of settings for tree complexity (1, 2, 3, 4, 5) and learning rate (0.05, 0.01, 0.005, 0.001) were tested for each species. We selected the learning rate and tree complexity values for the model with the smallest mean deviance and  $\geq 1,000$  trees, as suggested by Elith et al. (2008). Bag fraction was retained at 0.75, the default in `gbm.step` (Harris et al. 2014). Even though variable selection is largely achieved automatically in BRTs through non-selection of unimportant variables, to increase model parsimony we simplified each of these models using the `gbm.simplify` function allowing the model to drop a maximum of five unimportant variables (Elith et al. 2008).

For each species we evaluated the models by examining the model’s area under the receiver operator curve (area under curve, AUC), the point biserial correlation coefficient (see p. 136 of Elith et al. 2006 for details), the percentage deviance explained, and the relative contribution of predictor variables (% contribution). These values were all calculated as the mean value based on the internal k-fold (10 folds) cross-validation from the `gbm.step` function. All values were returned from the `gbm.step` function, except for percentage deviance explained, which was calculated using code adapted from Derville et al. (2016), and the point biserial correlation coefficient, which was calculated using the `evaluate` function from *dismo* (Hijmans et al. 2017).

We also plotted partial dependence plots to show the effect of the nine most important predictor variables. These plots show the effect of a predictor on the response (occurrence) after accounting for the average effect of all other predictors in the model; positive and negative values indicate the influence on the response. Values of zero indicate no influence of the predictor. For these plots we used a modified version of the `gbm.plot` function and produced plots as a local polynomial regression (loess) smooths with span set to 0.3 (the default in `gbm.plot`).

Finally, to evaluate the predictive performance of our models, we implemented a 10-fold block cross-validation procedure that, for each model, partitioned the full data set into 10 spatially separated ‘testing’ and ‘training’ data sets using the `spatialBlock` function in the *blockCVR* package (Valavi et al. 2019). Each block (or cell) within an 80 km grid covering the study region was randomly assigned to one of the 10 cross-validation folds, along with all presence and absence (or background) data that fell within it. For some narrower range species, the 80 km grid was too large for each fold to contain at least one presence, so the block size was reduced sequentially by 10 km until this was achieved. For one species (*Kunzea sinclairii*), the minimum block size of 10 km was still too large to adequately partition the data, so we implemented a standard, randomised k-fold cross-validation (10 folds). We evaluated each model using the mean AUC across all folds. AUC values range from 0 (no prediction) to 1 (perfect prediction), with 0.5 indicating that the model performs no better than random. A general rule of thumb to assess model

performance based on AUC is as follows: 'useful' (>0.7), 'good' (>0.8), and 'very good' (>0.9) (Swets 1988; Gherghel et al. 2018).

Predictions of species distributions were produced at the same 100 m resolution of the predictor surfaces. To account for anthropogenic disturbance (urban expansion, agriculture, and other sources of habitat loss) and dispersal limitations, we applied species-specific regional constraints to each of our predictions. First, range restrictions were defined as the known range of each species at the Ecological District scale<sup>4</sup> based on published range data, as implemented in Wiser et al. (2017) (Appendix B). Second, the New Zealand Land Cover Database (LCDB) Version 4.1 (Manaaki Whenua – Landcare Research 2018) was used to restrict predictions to areas with land cover suitable to New Zealand Myrtaceae. For this exercise, species were allocated to two groups: group 1 comprised successional species (*Kunzea* spp. and *Leptospermum scoparium*), and group 2 comprised all remaining taxa, which are associated with less-disturbed habitats. Each group was allocated a range of LCDB class codes within which their species' distributions were predicted (Appendix C).

---

<sup>4</sup> New Zealand has been divided into 253 unique Ecological Districts (McEwen 1987). Each of these represent geographical regions with characteristic landscapes and biological communities.

## Results

The primary output of these models is a set of maps indicating either probability of occurrence (for presence–absence models) or relative likelihood of occurrence (for presence-only models) for each species. We also present a ‘Prediction synthesis’, which displays probability of occurrence from the presence–absence model on a continuous scale, but also identifies areas predicted by the presence-only model that were *not* predicted by the presence–absence model. We refer to these areas as the ‘Presence only model extension’ and they are defined as areas with low predictions from the presence–absence model (<5% probability of occurrence) but relatively high predictions from the presence-only model (>10% relative likelihood of occurrence).

For species where a presence–absence prediction was not possible, we present the presence-only prediction. These maps are included in Appendix D. All predicted surfaces are also provided electronically as spatial raster layers (100 m resolution, NZMG projection). Also, predictions are available for viewing via an online visualisation tool at

[https://landcare.shinyapps.io/Myrtaceae\\_SDM\\_WebApp/](https://landcare.shinyapps.io/Myrtaceae_SDM_WebApp/). These have been downscaled to 5 km resolution to decrease file size and will remain online for at least 12 months from the scheduled due date of this report (through to 30 June 2020).

Of the seven recently revised *Kunzea* species modelled here, two have a wide distribution covering much of the country (*K. robusta* and *K. serotina*), four have a primarily northern distribution (*K. amathicola*, *K. linearis*, *K. sinclairii*, and *K. tenuicaulis*), and *K. ericoides sensu stricto* occurs centrally in the Nelson/Tasman region (Appendix D). *Leptospermum scoparium* is widespread, with an almost ubiquitous distribution, which reflects its life history as a primarily early successional species (Stephens et al. 2005).

The Myrtaceae species currently affected most by myrtle rust, *Lophomyrtus bullata*, has a primarily northern and central distribution, occurring from Northland to Nelson/Tasman, approximately within the area identified as being most climatically suitable to the disease (Beresford et al. 2018). Closely related species *Lophomyrtus obcordata* and *Neomyrtus pedunculata* are also present in the North Island but extend much further south than *L. bullata*.

Of the *Metrosideros* species, three have a wide distribution (*M. diffusa*, *M. fulgens*, and *M. umbellata*), six are primarily northern (*M. albiflora*, *M. carminea*, *M. colensoi*, *M. excelsa*, *M. perforata*, and *M. robusta*), and *M. parkinsonii* is primarily central. *Metrosideros excelsa* (pōhutukawa) is also predicted to be mainly coastal, in line with other published sources (Simpson 2005). The models were also able to predict the disjunct, northern populations of *M. umbellata* (southern rātā) that are traditionally difficult to explain from an environmental and biogeographical perspective (Wardle 1971; Gardner et al. 2004). Another species that may be susceptible to myrtle rust, *Syzygium maire*, is also confined to northern and central areas of New Zealand where the disease is most likely to occur (Beresford et al. 2018; Appendix D).

For species where both presence–absence and presence-only models were parameterised, most offered a general agreement between both models (Appendix D). When disagreements did occur, these were generally because the presence-only models predicted species distributions more widely, which could indicate marginal habitat, range limits, or areas that are under-sampled with vegetation plots. This was more evident in species with a wide range (e.g. *Kunzea ericoides sensu lato*, *Leptospermum scoparium*, *Lophomyrtus obcordata*), with predictions for narrower-ranged species generally more consistent between model types (e.g. *Metrosideros albiflora*, *M. excelsa*).

When comparing block cross-validated AUCs (AUC<sub>cv</sub>), all presence–absence models performed well, with 14 out of 15 achieving values >0.8 (classified as ‘good’: Swets 1988; Gherghel et al. 2018) (Table 2). *Leptospermum scoparium* had an AUC<sub>cv</sub> of 0.74, which indicates that the model is ‘useful’. The reduced model performance for this species is not surprising given the difficulties associated with modelling generalist species (Segurado & Araújo 2004). The presence-only models also performed well, with 17 out of 23 models having AUC<sub>cv</sub> values >0.8 (Table 2). *Leptospermum scoparium* again had the worst model performance, based on AUC<sub>cv</sub>, of 0.71. However, it should be noted that AUC can be inflated by large sample sizes, as we have here with over 60,000 locations from the NVS databank alone (see Methods), since a high value can be obtained simply by predicting absences (Jiménez-Valverde 2012). Because of this, consideration of other metrics is important when assessing our models, so we also consider the correlation coefficient (between predicted and observed values) of the model. All presence–absence models had high correlation coefficients, with a mean of 0.7 and only one model below 0.6 (*Syzygium maire* at 0.58; Table 2).

The presence-only models did not perform as well, with 15 out of 23 models below 0.5 and a mean of 0.47.

**Table 2:** A summary of the boosted regression tree model parameters for the final models selected for each species (NT = number of trees, LR = learning rate, TC = tree complexity). Also presented are the following model fit statistics: area under the receiver operator curve (AUC), point-biserial correlation coefficient (Cor.), % deviance explained (%Dev.). AUC and %Dev are means based on k-fold cross-validations. We also present the mean AUC from the block cross-validation exercise (AUCcv). Model performance assessments in the AUC column are based on criteria from Swets (1988) and Gherghel et al. (2018).

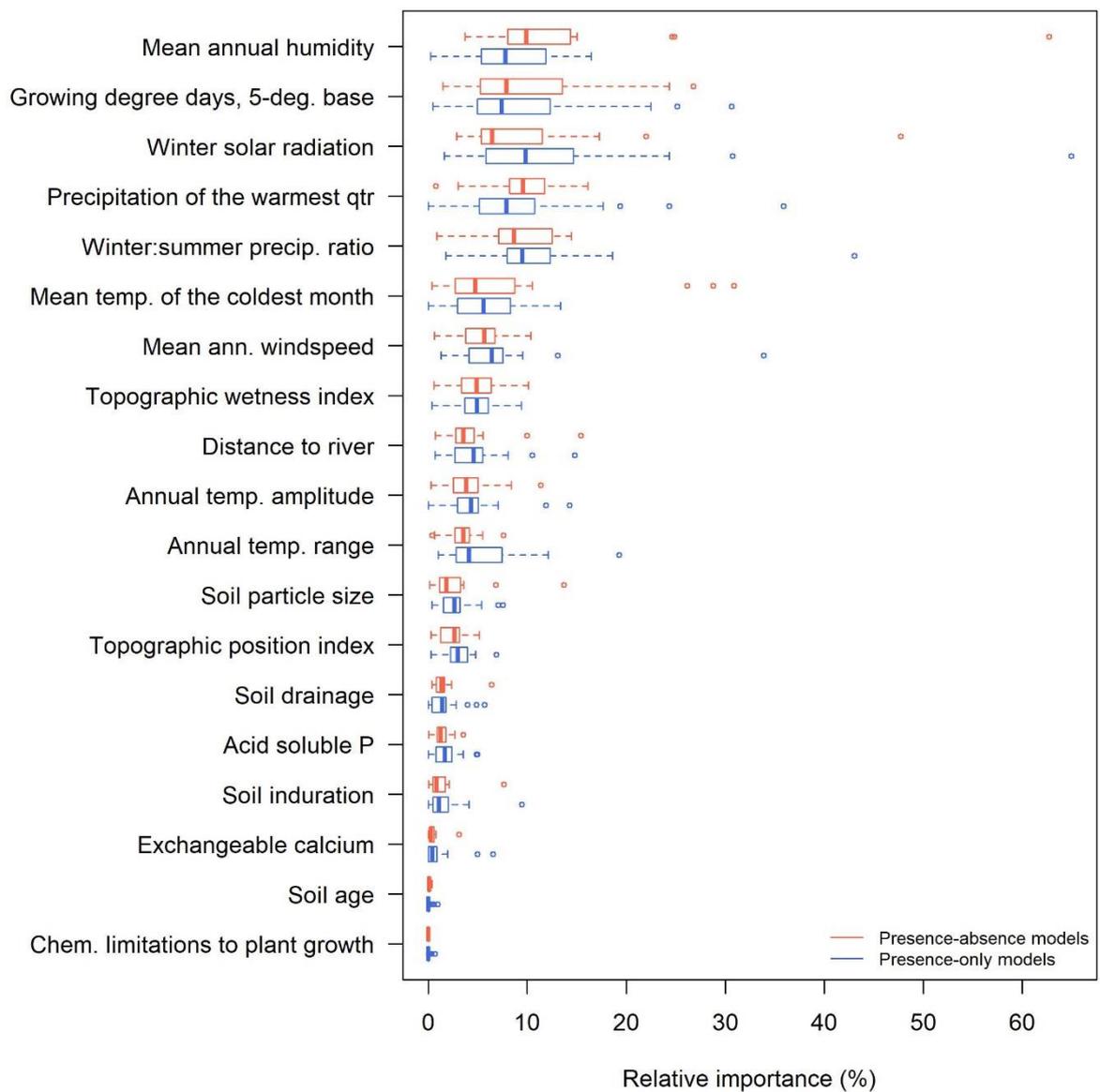
Species	NT	LR	TC	AUC	Cor.	%Dev.	AUCcv
<b>Presence-absence models</b>							
<i>Kunzea ericoides s.l.</i>	2950	0.05	4	0.94 (v. good)	0.73	44	0.89
<i>Leptospermum scoparium</i>	4450	0.05	4	0.87 (good)	0.69	29	0.74
<i>Lophomyrtus bullata</i>	3350	0.01	4	0.96 (v. good)	0.61	36	0.9
<i>Lophomyrtus obcordata</i>	5000	0.01	4	0.91 (v. good)	0.64	29	0.82
<i>Metrosideros albiflora</i>	4200	0.005	3	0.99 (v. good)	0.71	57	0.91
<i>Metrosideros colensoi</i>	7500	0.005	4	0.97 (v. good)	0.78	53	0.9
<i>Metrosideros diffusa</i>	3800	0.05	4	0.89 (good)	0.68	36	0.81
<i>Metrosideros excelsa</i>	7350	0.005	3	0.99 (v. good)	0.79	63	0.99
<i>Metrosideros fulgens</i>	3300	0.05	4	0.93 (v. good)	0.70	42	0.88
<i>Metrosideros parkinsonii</i>	8750	0.005	2	0.98 (v. good)	0.80	57	0.87
<i>Metrosideros perforata</i>	3200	0.05	4	0.91 (v. good)	0.66	36	0.85
<i>Metrosideros robusta</i>	2350	0.05	4	0.95 (v. good)	0.70	46	0.91
<i>Metrosideros umbellata</i>	4600	0.05	4	0.95 (v. good)	0.80	52	0.9
<i>Neomyrtus pedunculata</i>	3050	0.05	4	0.91 (v. good)	0.68	38	0.86
<i>Syzygium maire</i>	8700	0.001	3	0.96 (v. good)	0.58	42	0.89
<b>Presence-only models</b>							
<i>Kunzea amathicola</i>	7050	0.001	4	0.93 (v. good)	0.71	32	0.83
<i>Kunzea ericoides s.l.</i>	3450	0.05	4	0.88 (good)	0.42	37	0.7
<i>Kunzea ericoides s.s.</i>	2000	0.005	3	0.91 (v. good)	0.45	26	0.89
<i>Kunzea linearis</i>	2450	0.005	4	0.96 (v. good)	0.76	45	0.91
<i>Kunzea robusta</i>	3000	0.01	3	0.80 (good)	0.33	15	0.71
<i>Kunzea serotina</i>	1700	0.01	4	0.93 (v. good)	0.31	33	0.86
<i>Kunzea sinclairii</i>	1750	0.05	1	0.99 (v. good)	0.44	77	0.99
<i>Kunzea tenuicaulis</i>	3750	0.005	4	0.99 (v. good)	0.85	72	0.91
<i>Leptospermum scoparium</i>	4800	0.05	4	0.90 (v. good)	0.26	43	0.71
<i>Lophomyrtus bullata</i>	1350	0.05	4	0.91 (v. good)	0.38	37	0.78
<i>Lophomyrtus obcordata</i>	7600	0.01	4	0.91 (v. good)	0.28	41	0.78
<i>Metrosideros albiflora</i>	2450	0.01	4	0.97 (v. good)	0.50	55	0.91
<i>Metrosideros carminea</i>	7200	0.01	3	0.95 (v. good)	0.54	48	0.90
<i>Metrosideros colensoi</i>	7550	0.005	4	0.93 (v. good)	0.44	42	0.83
<i>Metrosideros diffusa</i>	4900	0.05	4	0.95 (v. good)	0.43	58	0.82

<i>Species</i>	<i>NT</i>	<i>LR</i>	<i>TC</i>	<i>AUC</i>	<i>Cor.</i>	<i>%Dev.</i>	<i>AUCcv</i>
<i>Metrosideros excelsa</i>	5100	0.005	4	0.96 (v. good)	0.65	52	0.91
<i>Metrosideros fulgens</i>	3000	0.05	4	0.96 (v. good)	0.46	61	0.88
<i>Metrosideros parkinsonii</i>	2150	0.01	4	0.98 (v. good)	0.36	64	0.87
<i>Metrosideros perforata</i>	3300	0.05	4	0.93 (v. good)	0.41	52	0.83
<i>Metrosideros robusta</i>	4100	0.05	4	0.94 (v. good)	0.53	54	0.82
<i>Metrosideros umbellata</i>	2400	0.05	4	0.99 (v. good)	0.51	82	0.96
<i>Neomyrtus pedunculata</i>	3150	0.05	4	0.98 (v. good)	0.43	73	0.92
<i>Syzygium maire</i>	7600	0.005	3	0.90 (v. good)	0.44	31	0.78

With BRTs, the relative contribution of different predictor variables in a model can be calculated based on the number of times a variable is selected across the individual trees and the resulting improvement to the model (Friedman 2001; Friedman & Meulman 2003; Elith et al. 2008). For our presence–absence models, mean annual humidity was the most important variable on average, followed by growing degree days (5°C base)<sup>5</sup>, winter solar radiation, and precipitation of the warmest quarter (Figure 2). Topographic wetness index and soil particle size were the most important landform and soil predictors on average, respectively.

For our presence-only models, winter solar radiation was the most important variable, followed by the winter:summer precipitation ratio and precipitation of the warmest quarter. Substrate/soil variables (**Error! Reference source not found.**) did not have a high importance for either presence–absence or presence-only models, with climate variables having the most influence when averaged across all species (Figure 2). Partial plots showing two-way relationships between predictor variables and species presence are shown in Appendix D, and species-specific values of relative importance are presented in Appendix E.

<sup>5</sup> Growing degree days is a measure of the amount of warmth available for plant and insect growth. Growing degrees are the total number of degrees Celsius each day above a threshold, in this case 5°C.



**Figure 2:** When using boosted regression trees, the relative importance of predictor variables is calculated based on the number of times a variable is selected in the model, weighted by its improvement to the overall model (Friedman 2001; Elith et al. 2008). Here we show the relative contribution for all predictor variables across our presence–absence and presence-only models. All species-specific values are presented in Appendix E.

## 4 Results and discussion

### 4.1 Predicted distributions

In this study we predicted the patterns of distribution for 23 New Zealand native Myrtaceae species around New Zealand, presented in Appendix D and via an online visualisation tool<sup>6</sup> (at reduced resolution) hosted by Manaaki Whenua – Landcare Research through to 30 June 2020. The spatial resolution of these predictions is based on a 100 m grid, and includes New Zealand’s three main islands, surrounding inshore islands, and a selection of offshore islands, including the Three Kings Islands and the islands in the Hauraki Gulf. Unfortunately, due to data limitations, the Kermadec Islands, the Chatham Islands and the subantarctic islands were unable to be included.

The model statistics suggest we can be relatively confident in the predictions, and that the presence-only models performed better than the presence–absence models (Table 2), although these models are less flexible in their application. In fact, given our use of a target group sample for the background records, it is likely that the presence only-models actually closely approximate the presence–absence models but with more data, which may partly explain their performance over the presence-only models.

### 4.2 Presence–absence and presence-only predictions

In this report we present predictions using presence-only models for all species, and presence–absence models when species-level data permitted. Because presence-only models can be parameterised using occurrence data without confirmed absences, they can incorporate records from a range of sources, including those from herbarium records or from citizen science initiatives. These models are often used because occurrence data are more readily available and accessible, reducing the cost and effort of sampling a species across their range (Gomes et al. 2018).

Presence–absence data are harder to obtain, but generally more desirable for species distribution modelling because they include information on where a species does not occur. For presence-only models, these absences are replaced with background (or ‘pseudo-absence’) data, usually captured through some random or stratified sample of the entire study area (Guillera-Arroita et al. 2015). In this case we used a ‘target group background’ in which, for the species being modelled, the occurrence records from all other species were used for the background. This helps overcome some of the bias issues with presence-only data (Phillips et al. 2009). Because presence–absence data naturally contain a measure of the proportion of sites occupied by a species (prevalence), an absolute probability of occurrence to be estimated using this approach. In contrast, presence-only models are only able to predict a relative likelihood of occurrence, which is not comparable across species/models.

For species where both presence–absence and presence-only models were completed, there are two different predictions for their distribution. Because the presence-only models include additional data from herbarium records and citizen science initiatives, these models were parameterised with a larger sample size than the presence–absence equivalent, which may explain their higher performance. To produce predictions where accuracy is maximised and a measure of prevalence is also included, we produced a hybrid map wherever possible, called a ‘prediction synthesis’. These maps used the presence–absence prediction for the quantitative scale, but we extended the prediction to include areas predicted by the presence-only model (but not predicted by the presence–absence model) (Appendix D).

### 4.3 Limitations and assumptions

As with any model, the species distribution models (SDMs) presented here have limitations and assumptions. They are the result of a complicated machine-learning algorithm (boosted regression trees; BRTs), which, although shown to be very effective at producing accurate species distribution

---

<sup>6</sup> [https://landcare.shinyapps.io/Myrtaceae\\_SDM\\_WebApp/](https://landcare.shinyapps.io/Myrtaceae_SDM_WebApp/)

models (Elith et al. 2006), can be difficult to interpret and evaluate. When compared to conventional statistical (i.e. regression) models, BRTs do not employ *P* values to denote statistical significance, and the technique relies on complicated interactions that can be difficult to interpret and visualise (Elith et al. 2008). Furthermore, these maps are not intended to be a definitive account of where a species can be found.

Although we used known distributions and habitat types to prevent predicting into areas where a species is known not to occur, the results are still a result of a mathematical model using species occurrence and environmental data that are themselves prone to error, and do not directly account for disturbance history, community-level factors such as competition and facilitation, and biogeographical factors such as dispersal barriers. These maps should also be considered in relation to the national scale at which they were produced (Elith & Leathwick 2009). They will be most effective at national and regional scales, where the broad patterns captured by our environmental predictors are adequately represented, and less effective at the site-level scale, where smaller-scale microhabitat drivers have greater influence.

Unfortunately, we lack readily accessible, spatially complete environmental layers that explain the complicated disturbance and geological history of the New Zealand landscape (glaciation, earthquake, volcanic activity, etc.) that are known to be important drivers of species distributions, often at smaller site-level scales. For example, mānuka (*Leptospermum scoparium*) is known to colonise disturbed habitats (Stephens et al. 2005), southern rātā (*Metrosideros umbellata*) is an early coloniser of sites affected by landslides (Stewart & Veblen 1982), and Parkinson's rātā's (*M. parkinsonii*) discontinuous distribution (it is known from the upper South Island and Great and Little Barrier Island) is thought to be driven by the more recent formation of the lower North Island (Simpson 2005). The inclusion of accurate and explicit measures of these geological and disturbance variables would probably improve the accuracy of these models.

#### 4.4 Conclusions and future research

There has been no concerted effort to map all New Zealand's native Myrtaceae species at a fine-resolution across the entire country, and this is the first attempt to predict distributions for the recently revised *Kunzea* species (de Lange 2014). While some studies present known occurrences, sometimes aggregated to spatial grid cells (e.g., Wardle 1971), this study represents the first attempt to predict these species' distributions at a fine resolution and broad spatial extent.

With regard to myrtle rust, these predictions have a range of potential applications. Aside from their use as a proxy for disease location through host mapping, these maps will allow investigation of species across their predicted range to assess disease progression and prevalence. These predictions may also be used to inform conservation efforts such as seed banking, where the entire range of a species could be systematically sampled, and disease control, where core habitat for particular species (or species assemblages) could be managed (e.g. removal of infected material, targeted fungicide application). It should be noted, however, that these maps are best suited for interpretation at the national scale at which they were developed. Local or regional applications may require further detailed sampling, and alternative models or modelling approaches.

These predictions could be further improved through updates to the input species presence (and absence) data, and to the environmental predictor data. Improving coverage by establishing additional plots in under-sampled areas could improve the accuracy of our predictions. Digitisation of additional historical records from un-digitised survey plots and herbarium specimens could also improve coverage. Citizen science records also have the potential for further application, possibly by further utilisation of existing data, or generation of new data through further refinements of interactive technology such as mobile phone apps and social media.

Peer-reviewed publication of the *Kunzea* species revision (de Lange 2014) occurred relatively recently, so there has been limited time for these species to be identified in survey plots, meaning that presence–absence predictions could only be performed for *Kunzea ericoides sensu lato*. Survey of additional plots – or further research to facilitate the update of *Kunzea* taxonomy from existing plots – would help to address this shortfall. Environmental predictor data could be improved through the generation, or increased availability, of additional quantitative predictor layers covering soil nutrients (e.g. soil nitrogen and phosphorus concentrations), soil pH, and soil depth, and also disturbance variables such as earthquake frequency or glaciation history. Remotely sensed variables such as canopy reflectance from satellite imagery or canopy structure from LiDaR could also be optimised for use in species distribution models.

Native hybrids, especially within genera known to be susceptible to myrtle rust (e.g. *Lophomyrtus*), may also benefit from further attention, although data from hybrids are not well represented in New Zealand databases. Accurately mapping, or even predicting, of the distribution of exotic Myrtaceae species could also be useful to the myrtle rust response. The research presented here could be extended further by producing predictions of Myrtaceae species richness to identify hot spots of Myrtaceae diversity. These could be further refined through identification and further exploration of those species known to be susceptible to myrtle rust, and incorporation of climate-based myrtle rust risk maps, such as those developed by Beresford et al. (2018). This might involve providing detailed risk maps to assist with targeting management and conservation efforts, or even identifying suitable locations outside the disease range that could be used for species translocation.

## 5 Acknowledgements

This work has been undertaken under contract MPI 18607 led by Scion, Plant and Food Research and Manaaki Whenua Landcare Research and the following collaborators: AsureQuality, Biosecurity Research Ltd., Learning for Sustainability and Te Tira Whakamātaki.

The information and opinions provided in the Report have been prepared for the Client and its specified purposes. Accordingly, any person other than the Client uses the information and opinions in this report entirely at its own risk. The Report has been provided in good faith and on the basis that reasonable endeavours have been made to be accurate and not misleading and to exercise reasonable care, skill and judgment in providing such information and opinions.

Neither Scion as the lead contractor, nor any of its employees, officers, contractors, agents or other persons acting on its behalf or under its control accepts any responsibility or liability in respect of any information or opinions provided in this Report.

We acknowledge the use of data drawn from the National Vegetation Survey Databank (NVS), the New Zealand Virtual Herbarium, iNaturalist and BioWeb, and thank all those who participated in the collection of these data and those who assisted in making them available. Plot data from NVS include data from the Natural Forest plot data collected between January 2002 and March 2007 by the LUCAS programme for the Ministry for the Environment, and data from the Department of Conservation, which is licensed by the Department of Conservation for re-use under a Creative Commons Attribution 4.0 International License.

Olivia Burge and Tom Etherington (Manaaki Whenua – Landcare Research, MWLR), and Jane Eliith (University of Melbourne), all provided invaluable modelling advice. We also thank Elise Arnst and Aaron Wilton (MWLR) and John Sullivan (Lincoln University) for assistance accessing species data. Species records were checked by Peter Bellingham, Larry Burrows and Alex Fergus (MWLR), and Shannel Courtney (Department of Conservation). We thank Ray Prebble for editing this report. This work was funded by the Ministry for Primary Industries as part of the Myrtle Rust research project 18607. Additional funding was provided by the Ministry for Business, Innovation and Employment's Strategic Science Investment Fund.

## 6 References

- Allen RB, Bellingham PJ, Wiser SK 2003. Forest biodiversity assessment for reporting conservation performance. *Science for Conservation* 216, Wellington, Department of Conservation.
- Allen RB, Rose AB 1983. Regeneration of southern rata (*Metrosideros umbellata*) and kamahi (*Weinmannia racemosa*) in areas of dieback. *Pacific Science* 37: 433-442.
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3: 327-338.
- Beenken L 2017. *Austropuccinia*: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). *Phytotaxa* 291: 53-61.
- Beresford RM, Turner R, Tait A, Paul V, Macara G, Yu ZD, Lima L, Martin R 2018. Predicting the climatic risk of myrtle rust during its first year in New Zealand. *New Zealand Plant Protection* 71: 332-347.
- Breiman L 2001. Statistical modeling: the two cultures. *Statistical Science* 16: 199-231.
- Breitwieser I, Brownsey PJ, Heenan PB, Nelson WA, Wilton AD 2010. Flora of New Zealand. <http://www.nzflora.info> (accessed 25 October 2018).
- Breitwieser I, Brownsey PJ, Heenan PB, Nelson WA, Wilton AD 2019. Flora of New Zealand Online – Taxon Profile: Myrtaceae. <http://www.nzflora.info/factsheet/Taxon/Myrtaceae.html> (accessed 19 June 2019).
- Brenning A, Bangs D, Becker M, Schratz P, Polakowski F 2018. RSAGA: SAGA geoprocessing and terrain analysis. R package version 1.1.1.
- Carnegie AJ, Kathuria A, Pegg GS, Entwistle P, Nagel M, Giblin FR 2016. Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. *Biological Invasions* 18: 127-144.
- Carnegie AJ, Lidbetter JR, Walker J, Horwood MA, Tesoriero L, Glen M, Priest MJ 2010. *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. *Australasian Plant Pathology* 39: 463-466.
- Cayuela L, Granzow-de la Cerda Í, Albuquerque FS, Golicher DJ 2012. taxonstand: an R package for species names standardisation in vegetation databases. *Methods in Ecology and Evolution* 3: 1078-1083.
- Chamberlain S, Barve V, McGlenn D, Oldoni D, Geffert L, Ram K 2018. rgbif: interface to the Global 'Biodiversity' Information Facility API. R package version 1.1.0.
- Clark S 2011. Risk analysis of the *Puccinia psidii*/Guava Rust fungal complex (including *Uredo rangelii*/Myrtle Rust) on nursery stock. Wellington, New Zealand, Ministry of Agriculture and Forestry.
- Coops N, Loughhead A, Ryan P, Hutton R 2001. Development of daily spatial heat unit mapping from monthly climatic surfaces for the Australian continent. *International Journal of Geographical Information Science* 15: 345-361.
- Cunningham A 1953. National Forest Survey: a large-scale ecological project. *Tuatara* 5: 39-48.
- Dawson JW 1985. *Metrosideros bartlettii* (Myrtaceae) a new species from North Cape, New Zealand. *New Zealand Journal of Botany* 23: 607-610.
- de Lange PJ 2007. Biosystematics of the New Zealand *Kunzea ericoides* (A.Rich.) Joy Thomps. complex. Unpublished PhD thesis, University of Auckland, Auckland, New Zealand.
- de Lange PJ 2014. A revision of the New Zealand *Kunzea ericoides* (Myrtaceae) complex. *PhytoKeys* 40: 1-185.
- de Lange PJ, Rolfe JR, Barkla JW, Courtney SP, Champion PD, Perrie LR, Beadel SM, Ford KA, Breitwieser I, Schönberger I, Hindmarsh-Walls R, Heenan PB, Ladley K 2018. Conservation status of New Zealand indigenous vascular plants, 2017. *New Zealand Threat Classification Series*, Wellington, New Zealand, Department of Conservation.
- Dennis RLH, Thomas CD 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation* 4: 73-77.

Department of Conservation 2018. Protected Areas. <https://data.linz.govt.nz/layer/53564-protected-areas/> (accessed 8 May 2018).

Derville S, Constantine R, Baker CS, Oremus M, Torres LG 2016. Environmental correlates of nearshore habitat distribution by the Critically Endangered Māui dolphin. *Marine Ecology Progress Series* 551: 261-275.

Drummond RSM, Keeling DJ, Richardson TE, Gardner RC, Wright SD 2001. Genetic analysis and conservation of 31 surviving individuals of a rare New Zealand tree, *Metrosideros bartlettii* (Myrtaceae). *Molecular Ecology* 9: 1149-1157.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Elith J, Leathwick JR 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.

Elith J, Leathwick JR, Hastie T 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-813.

Ferrier S, Watson G, Pearce J, Drielsma M 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity & Conservation* 11: 2275-2307.

Fithian W, Elith J, Hastie T, Keith DA 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution* 6: 424-438.

Forest Owners Association 2018. Facts and Figures 2016/17: New Zealand Plantation Forest Industry. [https://www.nzfoa.org.nz/images/stories/pdfs/Facts\\_Figures\\_2016\\_%C6%92a\\_web\\_version\\_v3.pdf](https://www.nzfoa.org.nz/images/stories/pdfs/Facts_Figures_2016_%C6%92a_web_version_v3.pdf) (accessed 12 February 2019).

Franklin J 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge, Cambridge University Press.

Friedman JH 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 29: 1189-1232.

Friedman JH, Meulman JJ 2003. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* 22: 1365-1381.

Gardner RC, De Lange PJ, Keeling DJ, Bowala T, Brown HA, Wright SD 2004. A late Quaternary phylogeography for *Metrosideros* (Myrtaceae) in New Zealand inferred from chloroplast DNA haplotypes. *Biological Journal of the Linnean Society* 83: 399-412.

Gherghel I, Brischoux F, Papeş M 2018. Using biotic interactions in broad-scale estimates of species' distributions. *Journal of Biogeography* 45: 2216-2225.

Glen M, Alfenas AC, Zauza EAV, Wingfield MJ, Mohammed C 2007. *Puccinia psidii*: a threat to the Australian environment and economy – a review. *Australasian Plant Pathology* 36: 1-16.

Gomes VHF, Ijff SD, Raes N, Amaral IL, Salomão RP, de Souza Coelho L, de Almeida Matos FD, Castilho CV, de Andrade Lima Filho D, López DC, Guevara JE, Magnusson WE, Phillips OL, Wittmann F, de Jesus Veiga Carim M, Martins MP, Irumé MV, Sabatier D, Molino J-F, Bánki OS, da Silva Guimarães JR, Pitman NCA, Piedade MTF, Mendoza AM, Luize BG, Venticinque EM, de Leão Novo EMM, Vargas PN, Silva TSF, Manzatto AG, Terborgh J, Reis NFC, Montero JC, Casula KR, Marimon BS, Marimon B-H, Coronado ENH, Feldpausch TR, Duque A, Zartman CE, Arboleda NC, Killeen TJ, Mostacedo B, Vasquez R, Schöngart J, Assis RL, Medeiros MB, Simon MF, Andrade A, Laurance WF, Camargo JL, Demarchi LO, Laurance SGW, de Sousa Farias E, Nascimento HEM, Revilla JDC, Quaresma A, Costa FRC, Vieira ICG, Cintra BBL, Castellanos H, Brienen R, Stevenson PR, Feitosa Y, Duivenvoorden JF, Aymard C GA, Mogollón HF, Targhetta N, Comiskey JA, Vicentini A, Lopes A, Damasco G, Dávila N, García-Villacorta R, Levis C, Schiatti J, Souza P, Emilio T, Alonso A, Neill D, Dallmeier F, Ferreira LV, Araujo-Murakami A, Praia D, do Amaral DD, Carvalho FA, de Souza FC, Feeley K, Arroyo L, Pansonato MP, Gribel R, Villa B, Licona JC, Fine PVA, Cerón C, Baraloto C, Jimenez EM, Stropp J, Engel J, Silveira M, Mora MCP, Petronelli P, Maas P, Thomas-Caesar R, Henkel TW, Daly D, Paredes MR, Baker TR, Fuentes A,

Peres CA, Chave J, Pena JLM, Dexter KG, Silman MR, Jørgensen PM, Pennington T, Di Fiore A, Valverde FC, Phillips JF, Rivas-Torres G, von Hildebrand P, van Andel TR, Ruschel AR, Prieto A, Rudas A, Hoffman B, Vela CIA, Barbosa EM, Zent EL, Gonzales GPG, Doza HPD, de Andrade Miranda IP, Guillaumet J-L, Pinto LFM, de Matos Bonates LC, Silva N, Gómez RZ, Zent S, Gonzales T, Vos VA, Malhi Y, Oliveira AA, Cano A, Albuquerque BW, Vriesendorp C, Correa DF, Torre EV, van der Heijden G, Ramirez-Angulo H, Ramos JF, Young KR, Rocha M, Nascimento MT, Medina MNU, Tirado M, Wang O, Sierra R, Torres-Lezama A, Mendoza C, Ferreira C, Baider C, Villarreal D, Balslev H, Mesones I, Giraldo LEU, Casas LF, Reategui MAA, Linares-Palomino R, Zagt R, Cárdenas S, Farfan-Rios W, Sampaio AF, Pauletto D, Sandoval EHV, Arevalo FR, Huamantupa-Chuquimaco I, Garcia-Cabrera K, Hernandez L, Gamarra LV, Alexiades MN, Pansini S, Cuenca WP, Milliken W, Ricardo J, Lopez-Gonzalez G, Pos E, ter Steege H 2018. Species Distribution Modelling: Contrasting presence-only models with plot abundance data. *Scientific Reports* 8: 1003.

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24: 276-292.

Guisan A, Thuiller W, Zimmerman NE 2017. *Habitat Suitability and Distribution Models With Applications in R*. Cambridge, UK, Cambridge University Press.

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77: 615-630.

Harris DB, Gregory SD, Brook BW, Ritchie EG, Croft DB, Coulson G, Fordham DA 2014. The influence of non-climate predictors at local and landscape resolutions depends on the autecology of the species. *Austral Ecology* 39: 710-721.

Hastie T, Tibshirani R, Friedman J 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York, NY, Springer.

Hernandez PA, Graham CH, Master LL, Albert DL 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773-785.

Hijmans RJ 2017. raster: Geographic Data Analysis and Modeling. R package version 2.6-7.

Hijmans RJ, Phillips SJ, Leathwick JR, Elith J 2017. dismo: Species Distribution Modelling. R package version 1.1-4.

iNaturalist 2018. iNaturalist Help. <https://www.inaturalist.org/pages/help> (accessed 9 July 2018).

Jiménez-Valverde A 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21: 498-507.

Kawanishi T, Uematsu S, Kakishima M, Kagiwada S, Hamamoto H, Horie H, Namba S 2009. First report of rust disease on ohia and the causal fungus, *Puccinia psidii*, in Japan. *Journal of General Plant Pathology* 75: 428.

Killgore EM, Heu RA 2007. Ohia rust: *Puccinia psidii* Winter. *New Pest Advisory* 05-04. <https://hdoa.hawaii.gov/pi/files/2013/01/npa05-04-ohiarust.pdf> (accessed 6 September 2018).

Lambert S, Waipara N, Black A, Mark-Shadbolt M, Wood W 2018. Indigenous Biosecurity: Māori Responses to Kauri Dieback and Myrtle Rust in Aotearoa New Zealand. In: Urquhart J, Marzano M, Potter C eds. *The Human Dimensions of Forest and Tree Health: Global Perspectives*. Cham, Switzerland, Springer International Publishing. Pp. 109-137.

Land Information New Zealand 2018. NZ River Centrelines (Topo, 1:500k). <https://data.linz.govt.nz/layer/50223-nz-river-centrelines-topo-1500k/> (accessed 26 June 2018).

Leathwick J, Morgan F, Wilson G, Rutledge D, McLeod M, Johnson K 2002. *Land Environments of New Zealand: A Technical Guide*. Wellington, Ministry for the Environment.

Leathwick JR, Elith J, Francis MP, 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.

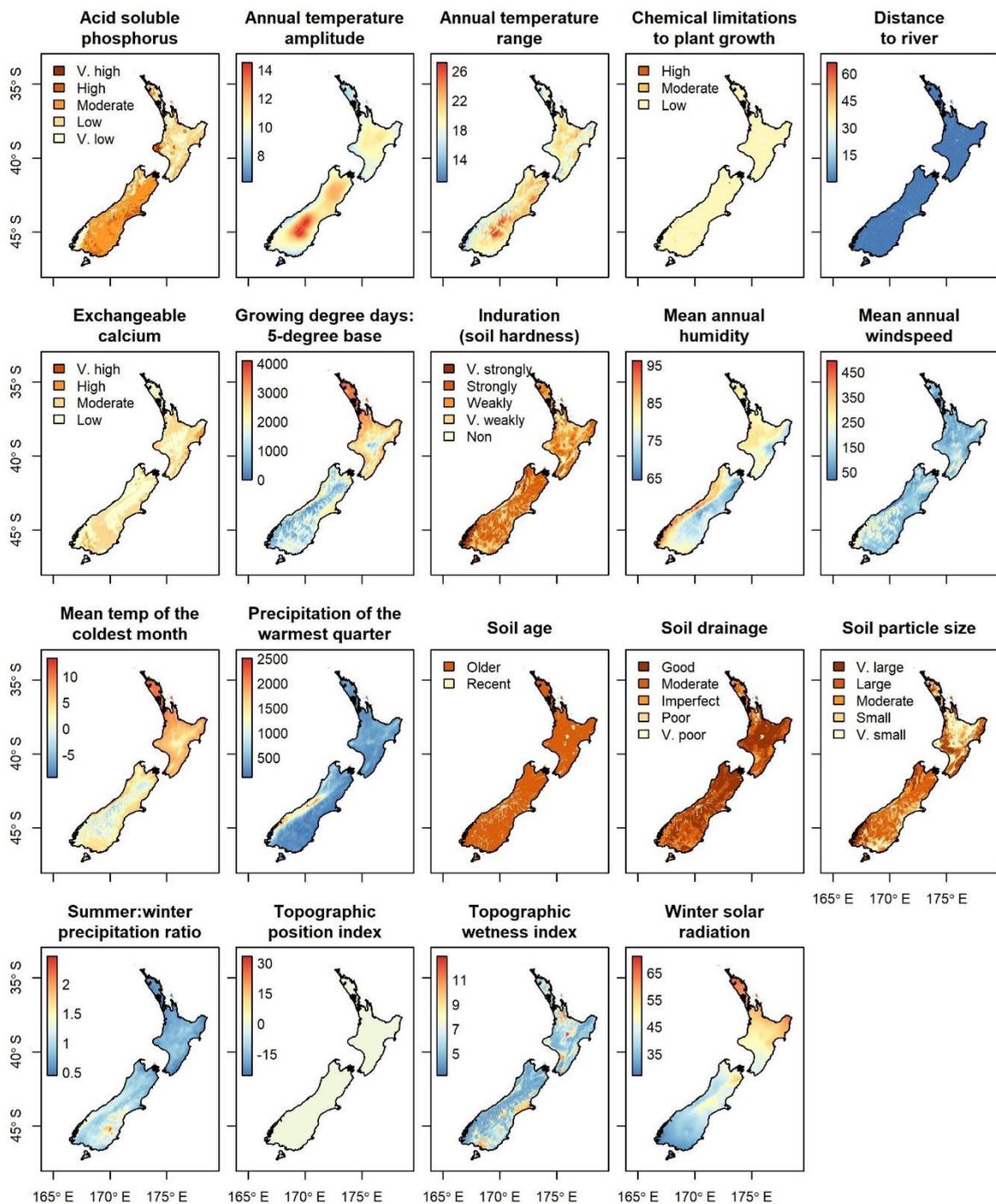
- Lee WG 1992. New Zealand Ultramafics. In: Roberts BA, Proctor J eds. The Ecology of Areas with Serpentinized Rocks: A World View. Dordrecht, Springer. Pp. 375-417.
- Manaaki Whenua – Landcare Research 2018. LCDB v4.1 – Land Cover Database version 4.1, Mainland New Zealand. <https://iris.scinfo.org.nz/layer/48423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand/> (accessed 6 June 2018).
- Marlatt RB, Kimbrough JW 1979. *Puccinia psidii* on *Pimenta dioica* in south Florida. Plant Disease Reporter 63: 510-512.
- Mateo RG, Croat TB, Felicísimo ÁM, Muñoz J 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. Diversity and Distributions 16: 84-94.
- McEwen WM ed. 1987. Ecological Regions and Districts of New Zealand. 3rd edn. Wellington, Department of Conservation.
- Ministry for Primary Industries 2018. Apiculture: Ministry for Primary Industries 2017 Apiculture Monitoring Programme. <https://landusenz.org.nz/wp-content/uploads/2018/03/2017-Apiculture-monitoring-report-FINAL.pdf> (accessed 12 February 2019).
- New Zealand Virtual Herbarium Network 2018. New Zealand Virtual Herbarium – Myrtaceae specimen data from members of the New Zealand National Herbarium Network.
- NZPCN 2017. New Zealand Plant Conservation Network. <http://www.nzpcn.org.nz> (accessed 28 October 2017).
- Ovaskainen O, Roy DB, Fox R, Anderson BJ 2016. Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. Methods in Ecology and Evolution 7: 428-436.
- Pegg G, Taylor T, Entwistle P, Guymer G, Giblin F, Carnegie A 2017. Impact of *Austropuccinia psidii* (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. PLOS ONE 12: e0188058.
- Phillips SJ, Anderson RP, Schapire RE 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19: 181-197.
- R Core Team 2018. R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing.
- Roux J, Greyling I, Coutinho TA, Verleur M, Wingfield MJ 2013. The Myrtle rust pathogen, *Puccinia psidii*, discovered in Africa. IMA Fungus 4: 155-159.
- Schmeller DS, Henry P-Y, Julliard R, Gruber B, Clobert J, Dziock F, Lengyel S, Nowicki P, DÉRi E, Budrys E, Kull T, Tali K, Bauch B, Settele J, Van Swaay C, Kobler A, Babij V, Papastergiadou EVA, Henle K 2009. Advantages of volunteer-based biodiversity monitoring in Europe. Conservation Biology 23: 307-316.
- Schulman L, Toivonen T, Ruokolainen K 2007. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. Journal of Biogeography 34: 1388-1399.
- SEEG Research Group 2018. seegSDM: streamlined functions for species distribution modelling in the SEEG research group. <https://github.com/SEEG-Oxford/seegSDM> (accessed 10 August 2018).
- Segurado P, Araújo MB 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555-1568.
- Simpson PG 2005. Pōhutukawa and Rātā: New Zealand's Iron Hearted Trees. Wellington, Te Papa Press.
- Simpson PG 2011. Lake Sumner Rata. Unpublished report for Hurunui Waitau Zone Committee.
- Stephens JMC, Molan PC, Clarkson BD 2005. A review of *Leptospermum scoparium* (Myrtaceae) in New Zealand. New Zealand Journal of Botany 43: 431-449.

- Stewart GH, Veblen TT 1982. Regeneration patterns in southern rata (*Metrosideros umbellata*) – kamahi (*Weinmannia racemosa*) forest in central Westland, New Zealand. *New Zealand Journal of Botany* 20: 55-72.
- Swets J 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.
- Teulon DAJ, Alipia TT, Ropata HT, Green JM, Viljanen-Rollinson SLH, Cromey MG, Arthur K, MacDiarmid RM, Waipara NW, Marsh AT 2015. The threat of myrtle rust to Māori taonga plant species in New Zealand. *New Zealand Plant Protection* 66: 66-75.
- Thomson AP 1946. Design for a forest survey. *New Zealand Journal of Forestry* 5: 191-199.
- Title PO, Bemmels JB 2017. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* 41: 291-307.
- Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G 2019. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* 10: 225-232.
- van Strien AJ, van Swaay CAM, Termaat T 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* 50: 1450-1458.
- Wardle P 1971. Biological flora of New Zealand 6. *Metrosideros umbellata* Cav. [Syn. *M. lucida* (Forst.f.) A. Rich.] (Myrtaceae) Southern rata. *New Zealand Journal of Botany* 9: 645-671.
- Wardle P 1991. *Vegetation of New Zealand*. Cambridge, UK, Cambridge University Press.
- Wellington Botanical Society 2008. Native vascular plants of Great Barrier Island. Unpublished species list. <https://www.thebarrier.co.nz/PDFs/ListNativeVascularPlantsGBI.pdf> (accessed 25 October 2017).
- Wilkinson DP, Golding N, Guillera-Arroita G, Tingley R, McCarthy MA 2019. A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution* 10: 198-211.
- Wiser SK, Bellingham PJ, Burrows LE 2001. Managing biodiversity information: development of New Zealand's National Vegetation Survey databank. *New Zealand Journal of Ecology* 25: 1-17.
- Wiser SK, Cooper JA, Arnst EA, Richardson SJ 2017. Mapping of native Myrtaceae species in New Zealand. Manaaki Whenua - Landcare Research Contract Report LC3065 prepared for Department of Conservation, Lincoln, New Zealand,

# Appendix A

## Environmental predictor data sources

Here we present the environmental variables used to predict the distributions of New Zealand's native Myrtaceae presented in this report. Nineteen variables were used in these analyses: nine climate, seven substrate and three landform variables (Figure A1, Table A1).



**Figure A1:** Environmental variables used to predict Myrtaceae species distributions. See Table A1 for units and variable descriptions.

**Table A1: Descriptions of environmental variables used to predict Myrtaceae species distributions as presented in this report**

<i>Environmental variable</i>	<i>Broad category</i>	<i>Unit</i>	<i>Short description</i>	<i>Reference</i>
Acid soluble phosphorus	Substrate	Ordinal category	A 5-factor variable indicating the phosphorus content of the soil parent material.	LENZ (Leathwick et al. 2002).
Annual temperature amplitude	Climate	°C	A measure of continentality, whereby low values indicate marine climates and high values continental climates. Calculated as the maximum mean monthly temperature minus the minimum mean monthly temperature.	Calculated by the authors using monthly temperature layers from the source LENZ temperature data (Leathwick et al. 2002).
Annual temperature range	Climate	°C	A measure of annual temperature variation experienced by a site. Calculated as the maximum temperature of the warmest month minus the minimum temperature of the coldest month.	Calculated by the authors using the source monthly LENZ temperature layers (Leathwick et al. 2002) and the biovars function in the <i>dismo</i> R package (Hijmans et al. 2017).
Chemical limitations to plant growth	Substrate	Ordinal category	A 3-factor variable (low, moderate, high) indicating the level of naturally occurring toxic chemicals in the soil parent material. Areas with high chemical limitations (sometimes referred to as ultramafic soils) are relatively sparse across New Zealand, but tend to support unique ecosystems (Lee 1992).	LENZ (Leathwick et al. 2002).
Distance to river	Landform	km	The horizontal distance to the nearest river.	Calculated by the authors using the NZ River Centrelines (Topo, 1:500k) spatial layer (Land Information New Zealand 2018) and the distance function in the <i>Raster</i> R package (Hijmans 2017).
Exchangeable calcium	Substrate	Ordinal category	A 4-factor variable indicating the concentration of calcium in the soil parent material. High levels of calcium generally occur in rocks of marine origin (e.g. limestone, marble).	LENZ (Leathwick et al. 2002).
Growing degree days: 5°C base	Climate	Index	A measure of how much warmth above 5°C is available for plant growth during a growing season. Calculated as the annual sum of mean daily temperatures minus 5 (days at <5°C are set to zero).	Calculated based on equations in Coops et al. (2001) applied to the source monthly LENZ temperature layers (Leathwick et al. 2002).
Induration (soil hardness)	Substrate	Ordinal category	A 5-factor variable indicating the hardness of the soil parent material. Strongly indurated soils include the igneous rocks formed under great heat and pressure, whereas weakly indurated rocks include the younger siltstones and mudstones.	LENZ (Leathwick et al. 2002).

<b>Environmental variable</b>	<b>Broad category</b>	<b>Unit</b>	<b>Short description</b>	<b>Reference</b>
Mean annual humidity	Climate	%	The average annual humidity, measured as a percentage of the maximum amount of moisture air can hold.	Calculated using the monthly LENZ source humidity layers (Leathwick et al. 2002).
Mean annual windspeed	Climate	km hr <sup>-1</sup>	The average wind speed experienced at a site.	Calculated using the monthly LENZ source windspeed layers (Leathwick et al. 2002).
Mean temperature of the coldest month	Climate	°C	The average temperature of the coldest month.	Calculated by the authors using monthly temperature layers from the source LENZ temperature data (Leathwick et al. 2002).
Precipitation of the warmest quarter	Climate	mm	The total precipitation experienced at a site during the warmest quarter (three months).	Calculated by the authors using the source monthly LENZ temperature and precipitation layers (Leathwick et al. 2002), and the <i>biovars</i> function in the <i>dismo</i> R package (Hijmans et al. 2017).
Soil age	Substrate	Ordinal category	A 2-factor variable indicating the age of the soil. Young soils are often formed on alluvial parent materials, which generally contain more nutrients due to smaller particle sizes.	LENZ (Leathwick et al. 2002).
Soil drainage	Substrate	Ordinal category	A 5-factor variable indicating the soil drainage. Areas with poor drainage are stressful to plants because they decrease oxygen supply to the roots.	LENZ (Leathwick et al. 2002).
Soil particle size	Substrate	Ordinal category	A 5-factor variable indicating the average size of different parent materials. Particle size is influenced by a range of processes including earthquakes, weathering, and volcanic activity.	LENZ (Leathwick et al. 2002).
Summer:winter precipitation ratio	Climate	mm	A measure of rainfall seasonality with high values experiencing more rain during summer than winter.	Calculated by the authors as the sum of summer rainfall (Dec, Jan, Feb) divided by the sum of winter rainfall (Jun, Jul, Aug) using the source monthly LENZ precipitation layers (Leathwick et al. 2002).
Topographic position index	Landform	Index	Compares the elevation of each location to the mean elevation of that location's neighbourhood. High values are associated with ridgetops and negative values are associated with valleys. Values close to zero indicate either hillsides or flat areas.	Calculated by the authors using the LENZ digital elevation model (Leathwick et al. 2002) and the <i>terrain</i> function in the <i>Raster</i> R package (Hijmans 2017).

<b><i>Environmental variable</i></b>	<b><i>Broad category</i></b>	<b><i>Unit</i></b>	<b><i>Short description</i></b>	<b><i>Reference</i></b>
Topographic wetness index	Landform	Index	Quantifies topographic control of hydrological processes by combining local upslope contributing area in the landscape with slope.	Calculated by the authors using the LENZ digital elevation model (Leathwick et al. 2002) and the topoWetnessIndex function in the <i>envirem</i> R package (Title & Bemmels 2017), in conjunction with the program SAGA GIS (version 6.2.0) accessed via R using the <i>rsaga.env</i> function from the <i>RSAGA</i> R package (Brenning et al. 2018).
Winter solar radiation	Climate	MJ m <sup>-2</sup> day <sup>-1</sup>	The amount of solar radiation experienced at a site during winter (June) when the sun is lowest in the sky. Values are higher in the north of New Zealand.	LENZ (Leathwick et al. 2002).

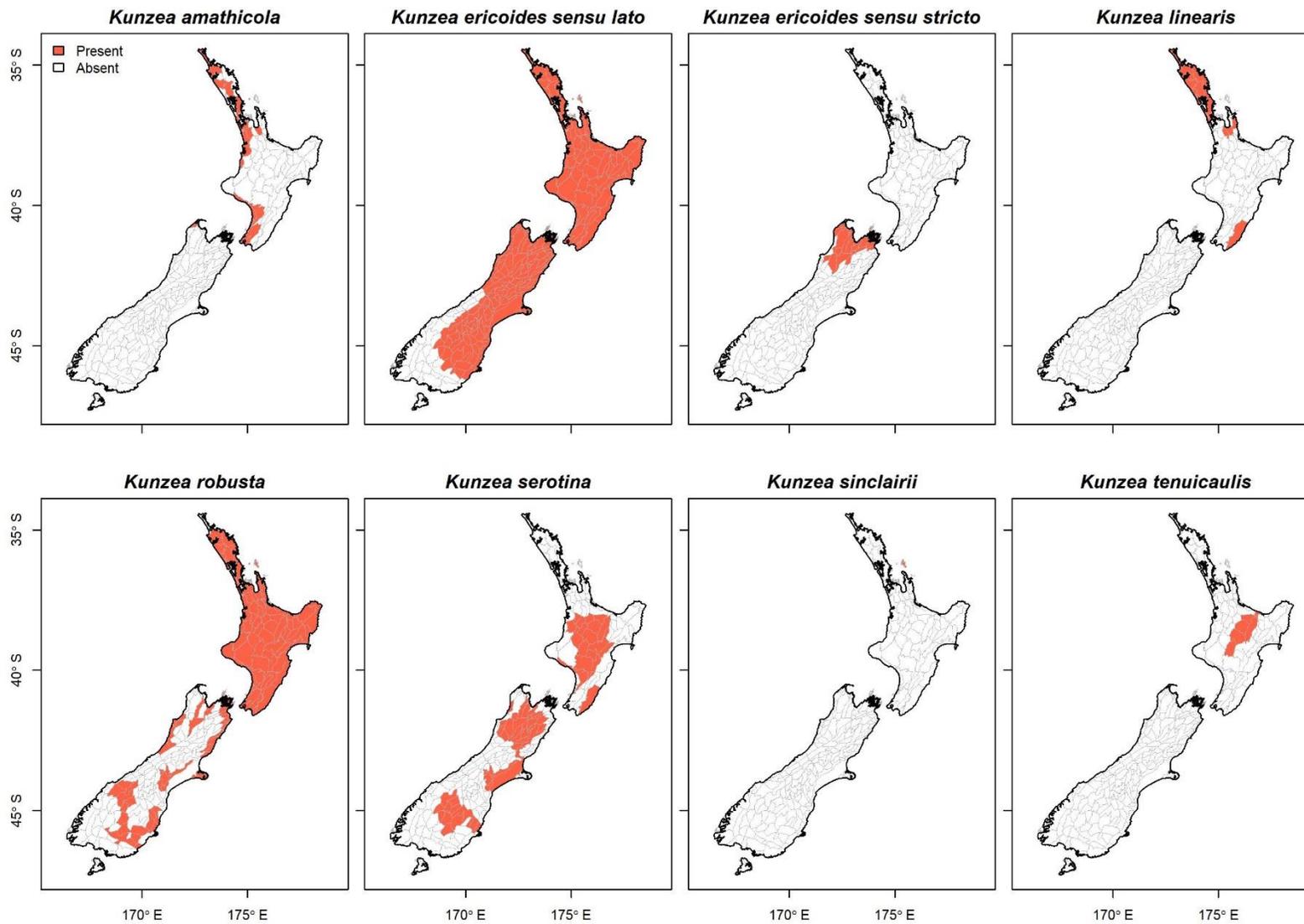
## 7 Appendix B

### Definition of species ranges at the Ecological District scale

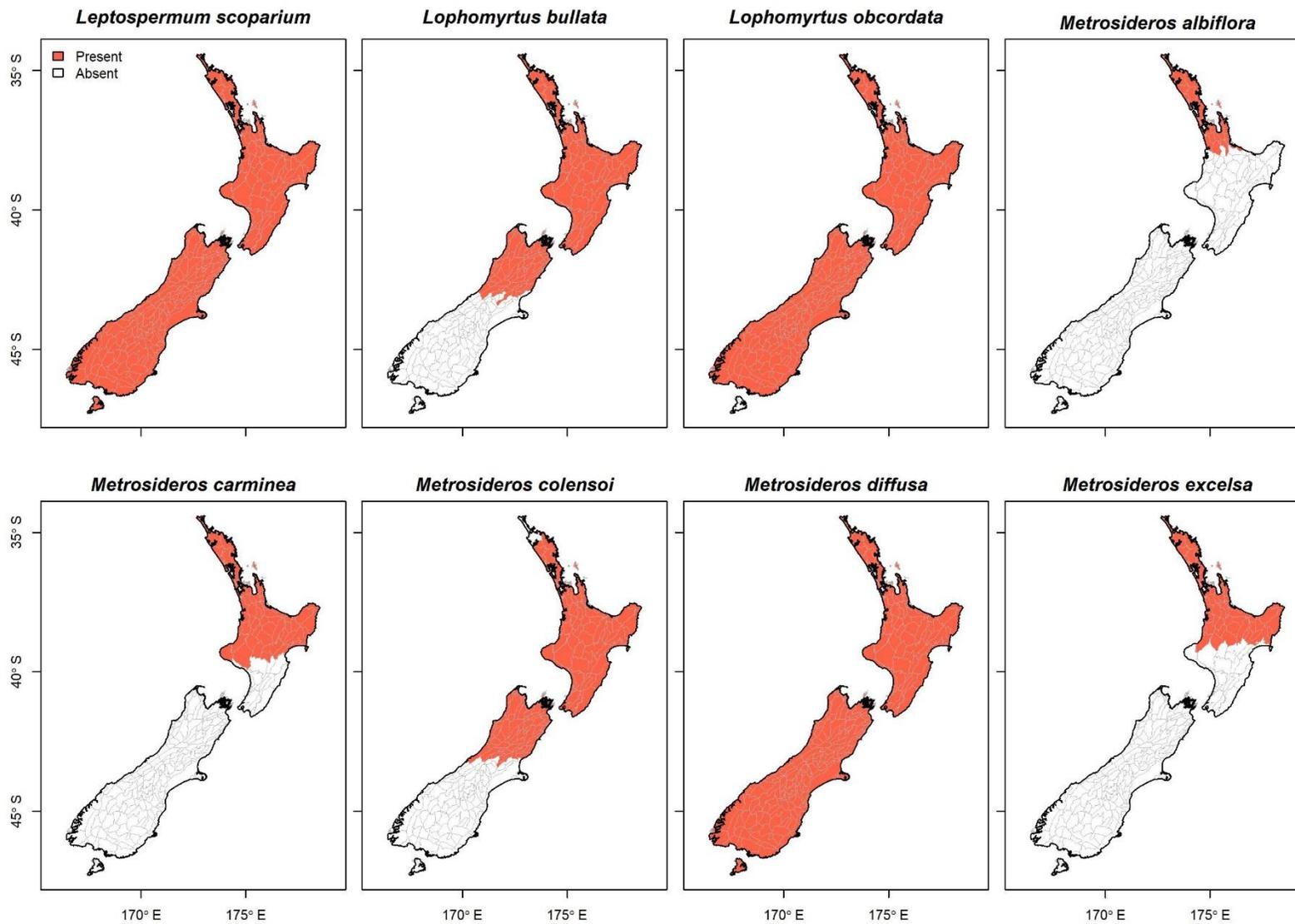
The range of each species in this report was defined using Ecological Districts after consulting various literature sources (Table B1, Figures B1–B3). These ranges were used to restrict our distribution predictions to the known range of these species. Here we use the same species ranges applied in earlier Myrtaceae distribution mapping exercises (Wiser et al. 2017), with the addition of the revised *Kunzea* spp., which were defined using the occurrence maps from de Lange (2014) (Table B1).

**Table B1:** Literature sources used to define species ranges at the Ecological District scale

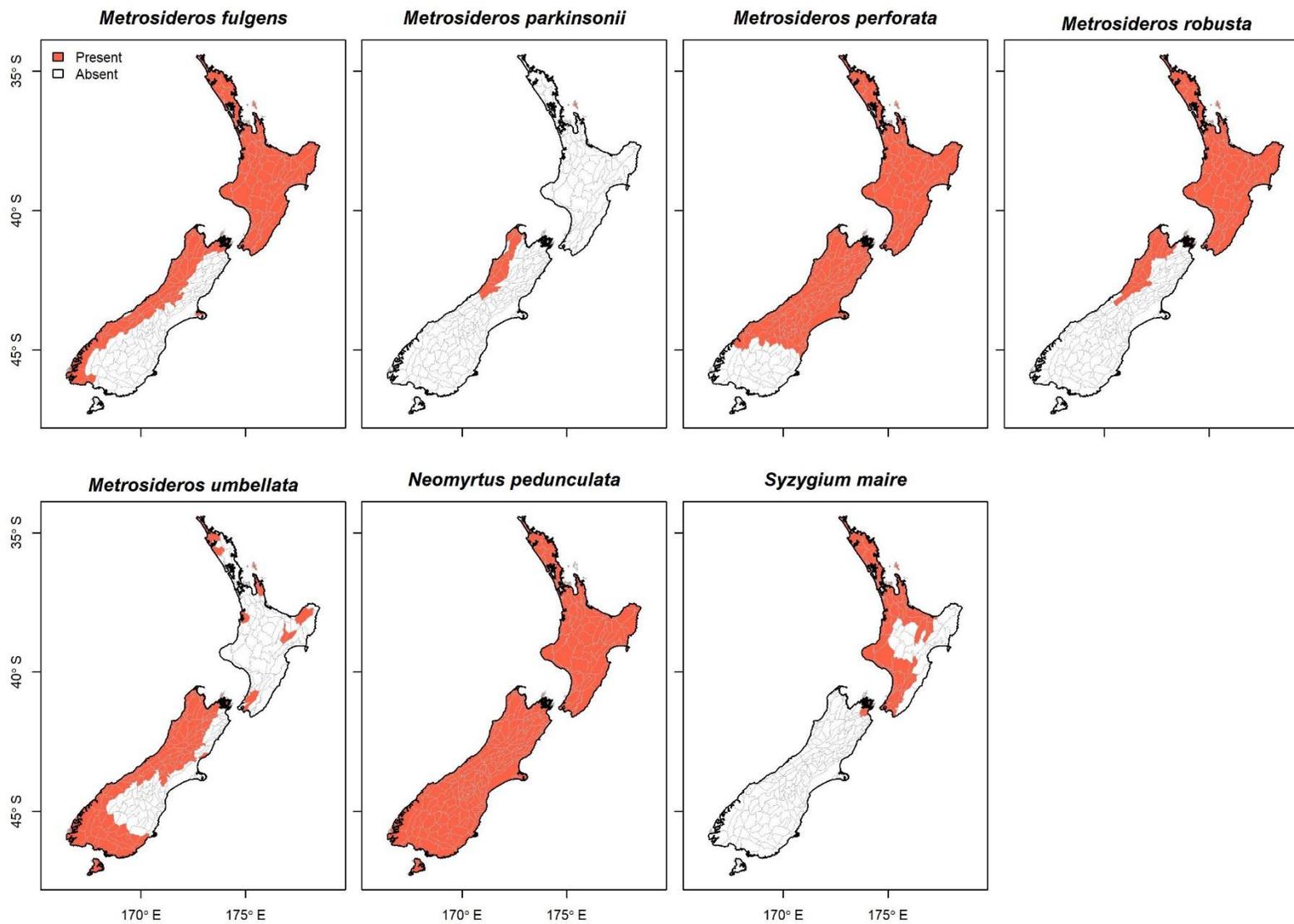
Species	Breitwieser et al. (2010)	de Lange (2014)	NZPCN (2017)	Simpson (2005)	Simpson (2011)	Wardle (1991)	Wellington Botanical Society (2008)
<i>Kunzea amathicola</i>		X					
<i>Kunzea ericoides sensu lato</i>	X	X	X			X	
<i>Kunzea ericoides sensu stricto</i>		X					
<i>Kunzea linearis</i>		X					
<i>Kunzea robusta</i>		X					
<i>Kunzea serotina</i>		X					
<i>Kunzea sinclairii</i>	X	X	X				
<i>Kunzea tenuicaulis</i>		X					
<i>Leptospermum scoparium</i>	X		X			X	
<i>Lophomyrtus bullata</i>	X		X				
<i>Lophomyrtus obcordata</i>	X		X				
<i>Metrosideros albiflora</i>	X		X				
<i>Metrosideros carminea</i>	X		X				
<i>Metrosideros colensoi</i>	X		X			X	
<i>Metrosideros diffusa</i>	X		X				
<i>Metrosideros excelsa</i>	X		X	X		X	
<i>Metrosideros fulgens</i>	X		X			X	
<i>Metrosideros parkinsonii</i>	X		X				
<i>Metrosideros perforata</i>	X		X				
<i>Metrosideros robusta</i>	X		X	X		X	
<i>Metrosideros umbellata</i>	X		X	X	X		
<i>Neomyrtus pedunculata</i>	X		X				X
<i>Syzygium maire</i>	X		X				



**Figure B1:** Species ranges at the Ecological District Scale (1 of 3).



**Figure B2:** Species ranges at the Ecological District Scale (2 of 3).



**Figure B3:** Species ranges at the Ecological District Scale (3 of 3).

## 8 Appendix C

### LCDB classes

The New Zealand Land Cover Database (LCDB) Version 4.1 (Manaaki Whenua – Landcare Research 2018) was used to restrict prediction of species' distributions to areas likely to contain suitable habitat (i.e. excluding intensive agriculture, urban areas, etc.; Table C1). All New Zealand Myrtaceae were split into two groups for this purpose: group 1 (successional species; *Kunzea* spp. and *Leptospermum scoparium*) and group 2 (all other species). Each group's predictions were restricted to a different set of LCDB classes depending on their associated habitats (Table C1, Figure C1).

**Table C1:** New Zealand Land Cover Database class codes and descriptions<sup>7</sup>

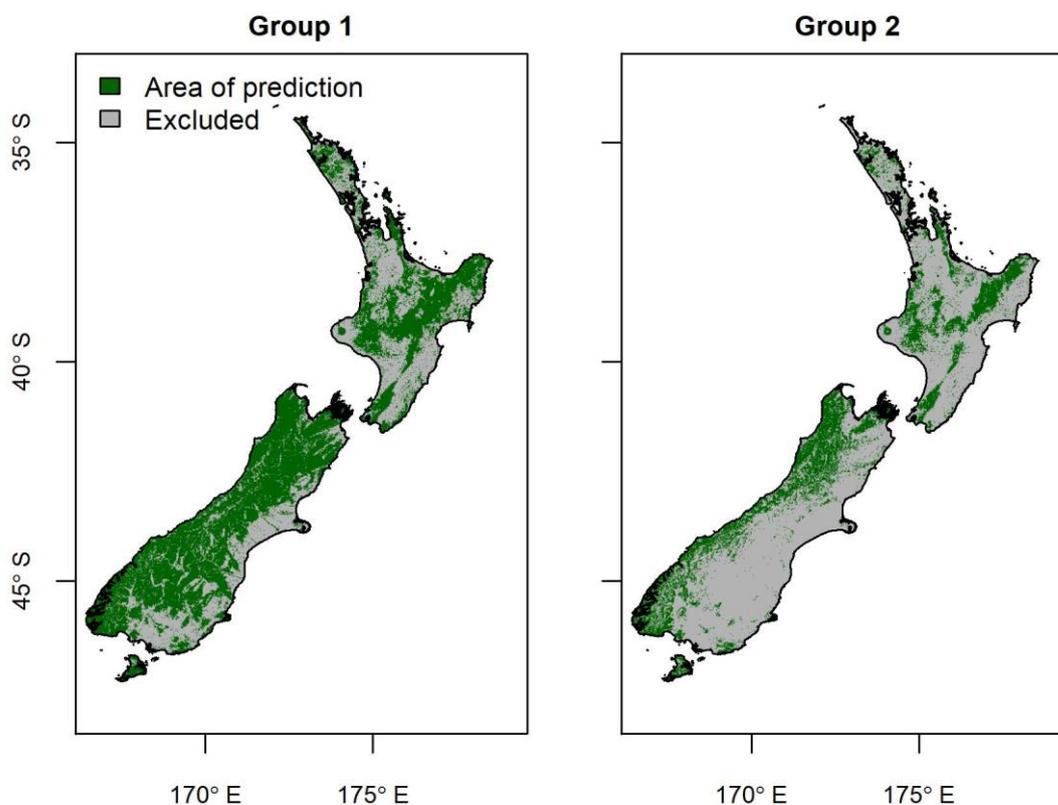
<b>Class Code</b>	<b>Class Name</b>	<b>Class Description</b>	<b>Group 1</b>	<b>Group 2</b>
1	Built-up Area (settlement)	Commercial, industrial or residential buildings, including associated infrastructure and amenities, not resolvable as other classes. Low density 'lifestyle' residential areas are included where hard surfaces, landscaping and gardens dominate other land covers.	-	-
2	Urban Parkland/Open Space	Open, mainly grassed or sparsely-treed, amenity, utility and recreation areas. The class includes parks and playing fields, public gardens, cemeteries, golf courses, berms and other vegetated areas usually within or associated with built-up areas.	-	-
5	Transport Infrastructure	Artificial surfaces associated with transport such as arterial roads, rail-yards and airport runways. Skid sites and landings associated with forest logging are sometimes also included.	-	-
6	Surface Mine or Dump	Bare surfaces arising from open-cast and other surface mining activities, quarries, gravel-pits and areas of solid waste disposal such as refuse dumps, clean-fill dumps and active reclamation sites	✓	-
10	Sand or Gravel	Bare surfaces dominated by unconsolidated materials generally finer than coarse gravel (60 mm). Typically mapped along sandy seashores and the margins of lagoons and estuaries, lakes and rivers and some areas subject to surficial erosion, soil toxicity and extreme exposure.	✓	-
12	Landslide	Bare surfaces arising from mass-movement erosion generally in mountain-lands and steep hill-country.	✓	-
14	Permanent Snow and Ice	Areas where ice and snow persists through late summer. Typically occurring above 1800 m but also at lower elevations as glaciers.	-	-
16	Gravel or Rock	Bare surfaces dominated by unconsolidated or consolidated materials generally coarser than coarse gravel (60 mm). Typically mapped along rocky seashores and rivers, sub-alpine and alpine areas, scree slopes and erosion pavements.	✓	-

<sup>7</sup> Reproduced from: <https://iris.scinfo.org.nz/layer/48423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand/>, retrieved 6/12/2018.

<b>Class Code</b>	<b>Class Name</b>	<b>Class Description</b>	<b>Group 1</b>	<b>Group 2</b>
15	Alpine Grass/Herbfield	Typically sparse communities above the actual or theoretical treeline dominated by herbaceous cushion, mat, turf, and rosette plants and lichens. Grasses are a minor or infrequent component, whereas stones, boulders and bare rock are usually conspicuous.	-	-
20	Lake or Pond	Essentially-permanent, open, fresh-water without emerging vegetation including artificial features such as oxidation ponds, amenity, farm and fire ponds and reservoirs as well as natural lakes, ponds and tarns.	-	-
21	River	Flowing open fresh-water generally more than 30m wide and without emerging vegetation. It includes artificial features such as canals and channels as well as natural rivers and streams.	-	-
22	Estuarine Open Water	Standing or flowing saline water without emerging vegetation including estuaries, lagoons, and occasionally lakes occurring in saline situations such as inter-dune hollows and coastal depressions.	-	-
30	Short-rotation Cropland	Land regularly cultivated for the production of cereal, root, and seed crops, hops, vegetables, strawberries and field nurseries, often including intervening grassland, fallow land, and other covers not delineated separately.	-	-
33	Orchards, Vineyards or Other Perennial Crops	Land managed for the production of grapes, pip, citrus and stone fruit, nuts, olives, berries, kiwifruit, and other perennial crops. Cultivation for crop renewal is infrequent and irregular but is sometimes practiced for weed control.	-	-
40	High Producing Exotic Grassland	Exotic sward grassland of good pastoral quality and vigour reflecting relatively high soil fertility and intensive grazing management. Clover species, ryegrass and cocksfoot dominate with lucerne and plantain locally important, but also including lower-producing grasses exhibiting vigour in areas of good soil moisture and fertility.	-	-
41	Low Producing Grassland	Exotic sward grassland and indigenous short tussock grassland of poor pastoral quality reflecting lower soil fertility and extensive grazing management or non-agricultural use. Browntop, sweet vernal, danthonia, fescue and Yorkshire fog dominate, with indigenous short tussocks (hard tussock, blue tussock and silver tussock) common in the eastern South Island and locally elsewhere.	✓	-
43	Tall Tussock Grassland	Indigenous snow tussocks in mainly alpine mountains and red tussock in the central North Island and locally in poorly-drained valley floors, terraces and basins of both islands.	✓	-
44	Depleted Grassland	Areas, of mainly former short tussock grassland in the drier eastern South Island high country, degraded by over-grazing, fire, rabbits and weed invasion among which <i>Hieracium</i> species are conspicuous. Short tussocks usually occur, as do exotic grasses, but bare ground is more prominent.	✓	-
45	Herbaceous Freshwater Vegetation	Herbaceous wetland communities occurring in freshwater habitats where the water table is above or just below the substrate surface for most of the year. The class includes rush, sedge, restiad, and sphagnum communities and other wetland species, but not flax nor willows which are mapped as Flaxland and Deciduous Hardwoods respectively.	-	-

<b>Class Code</b>	<b>Class Name</b>	<b>Class Description</b>	<b>Group 1</b>	<b>Group 2</b>
46	Herbaceous Saline Vegetation	Herbaceous wetland communities occurring in saline habitats subject to tidal inundation or saltwater intrusion. Commonly includes club rush, wire rush and glasswort, but not mangrove which is mapped separately	-	-
47	Flaxland	Areas dominated by New Zealand flax usually swamp flax ( <i>harakeke</i> ) in damp sites but occasionally mountain flax ( <i>wharariki</i> ) on cliffs and mountain slopes.	✓	-
50	Fernland	Bracken fern, umbrella fern, or ring fern, commonly on sites with low fertility and a history of burning. Manuka, gorse, and/or other shrubs are often a component of these communities and will succeed Fernland if left undisturbed.	✓	-
51	Gorse and/or Broom	Scrub communities dominated by gorse or Scotch broom generally occurring on sites of low fertility, often with a history of fire, and insufficient grazing pressure to control spread. Left undisturbed, this class can be transitional to Broadleaved Indigenous Hardwoods.	✓	-
52	Manuka and/or Kanuka	Scrub dominated by mānuka and/or kānuka, typically as a successional community in a reversion toward forest. Mānuka has a wider ecological tolerance and distribution than kānuka with the latter somewhat concentrated in the north with particular prominence on the volcanic soils of the central volcanic plateau.	✓	✓
54	Broadleaved Indigenous Hardwoods	Lowland scrub communities dominated by indigenous mixed broadleaved shrubs such as wineberry, mahoe, five-finger, <i>Pittosporum</i> spp., fuchsia, tutu, titoki and tree ferns. This class is usually indicative of advanced succession toward indigenous forest.	✓	✓
55	Sub Alpine Shrubland	Highland scrub dominated by indigenous low-growing shrubs including species of <i>Hebe</i> , <i>Dracophyllum</i> , <i>Olearia</i> , and <i>Cassinia</i> . Predominantly occurring above the actual or theoretical treeline, this class is also recorded where temperature inversions have created cooler micro-climates at lower elevations e.g. the 'frost flats' of the central North Island	✓	✓
56	Mixed Exotic Shrubland	Communities of introduced shrubs and climbers such as boxthorn, hawthorn, elderberry, blackberry, sweet brier, buddleja, and old man's beard.	✓	-
58	Matagouri or Grey Scrub	Scrub and shrubland comprising small-leaved, often divaricating shrubs such as matagouri, <i>Coprosma</i> spp., <i>Muehlenbeckia</i> spp., <i>Casinnia</i> spp., and <i>Parsonsia</i> spp. These, from a distance, often have a grey appearance.	✓	-
80	Peat Shrubland (Chatham Is)	Low-growing shrubland communities usually dominated by <i>Dracophyllum</i> spp. in association with <i>Cyathodes</i> spp. and ground ferns. Mapped only on the Chatham Islands.	-	-
81	Dune Shrubland (Chatham Is)	Low-growing shrubland communities dominated by <i>Leucopogon</i> spp., <i>Pimelia arenaria</i> and <i>Coprosma</i> spp., in association with sedges and scattered herbs and grasses. Mapped only on the Chatham Islands.	-	-

<b>Class Code</b>	<b>Class Name</b>	<b>Class Description</b>	<b>Group 1</b>	<b>Group 2</b>
70	Mangrove	Shrubs or small trees of the New Zealand mangrove ( <i>Avicennia marina</i> subspecies <i>australascia</i> ) growing in harbours, estuaries, tidal creeks and rivers north of Kawhia on the west coast and Ohiwa on the east coast.	-	-
64	Forest – Harvested	Predominantly bare ground arising from the harvesting of exotic forest or, less commonly, the clearing of indigenous forest. Replanting of exotic forest (or conversion to a new land use) is not evident and nor is the future use of land cleared of indigenous forest.	✓	-
68	Deciduous Hardwoods	Exotic deciduous woodlands, predominantly of willows or poplars but also of oak, elm, ash or other species. Commonly alongside inland water (or as part of wetlands), or as erosion-control, shelter and amenity plantings.	✓	-
69	Indigenous Forest	Tall forest dominated by indigenous conifer, broadleaved or beech species.	✓	✓
71	Exotic Forest	Planted or naturalised forest predominantly of radiata pine but including other pine species, Douglas fir, cypress, larch, acacia and eucalypts. Production forestry is the main land use in this class with minor areas devoted to mass-movement erosion-control and other areas of naturalised (wildling) establishment.	✓	-



**Figure C1:** LCDB coverage, by group.

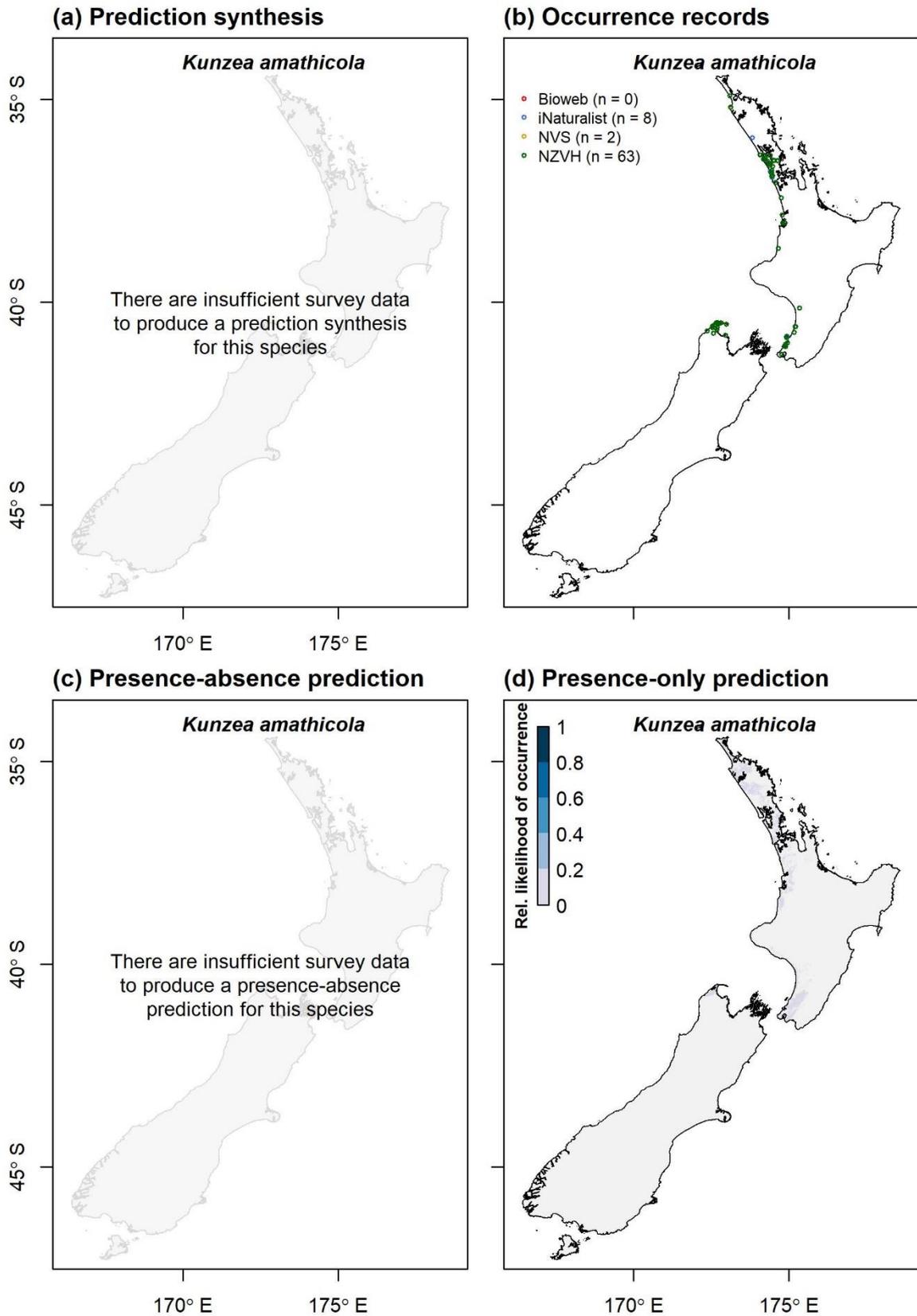
## 9 Appendix D

### Species predictions

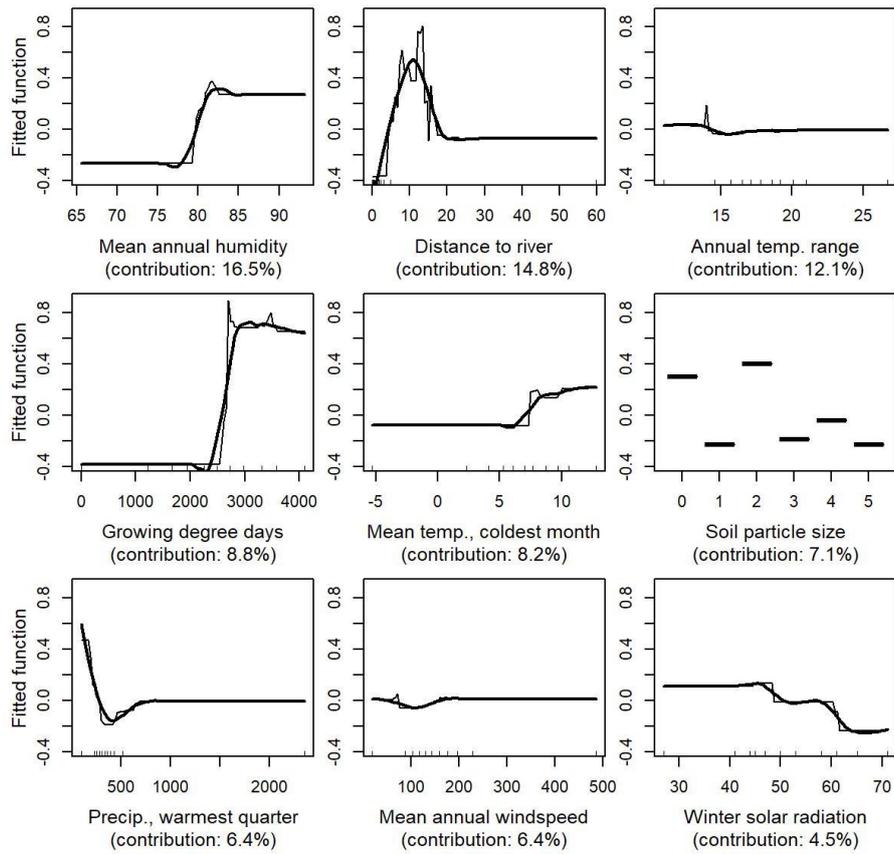
In this appendix, for each species we present predicted distribution maps for New Zealand native Myrtaceae species. For species where it was possible to run presence–absence and presence-only models we present predictions for both, as well as a ‘Prediction synthesis’, whereby the presence–absence model provides the continuous scale and areas predicted in the presence-only model are identified in grey (see details in text).

For each species we also include partial dependence plots for the nine variables with the highest relative contribution. These plots show two-way relationships between predictor variables and species presence, with all other variables held at their mean for both presence–absence and presence-only models. The thicker lines on these plots show a local polynomial regression (loess) smooth with span set to 0.3 (using the function `gbm.plot` from *dismo*, see text), and lighter lines show the actual response. Note that these plots should be interpreted with caution because they do not capture the sometimes-complex interactions common in boosted regression trees so are not a perfect representation of each variable’s effect on species presence and absence (Elith et al. 2008). See Table A1 for predictor variable units.

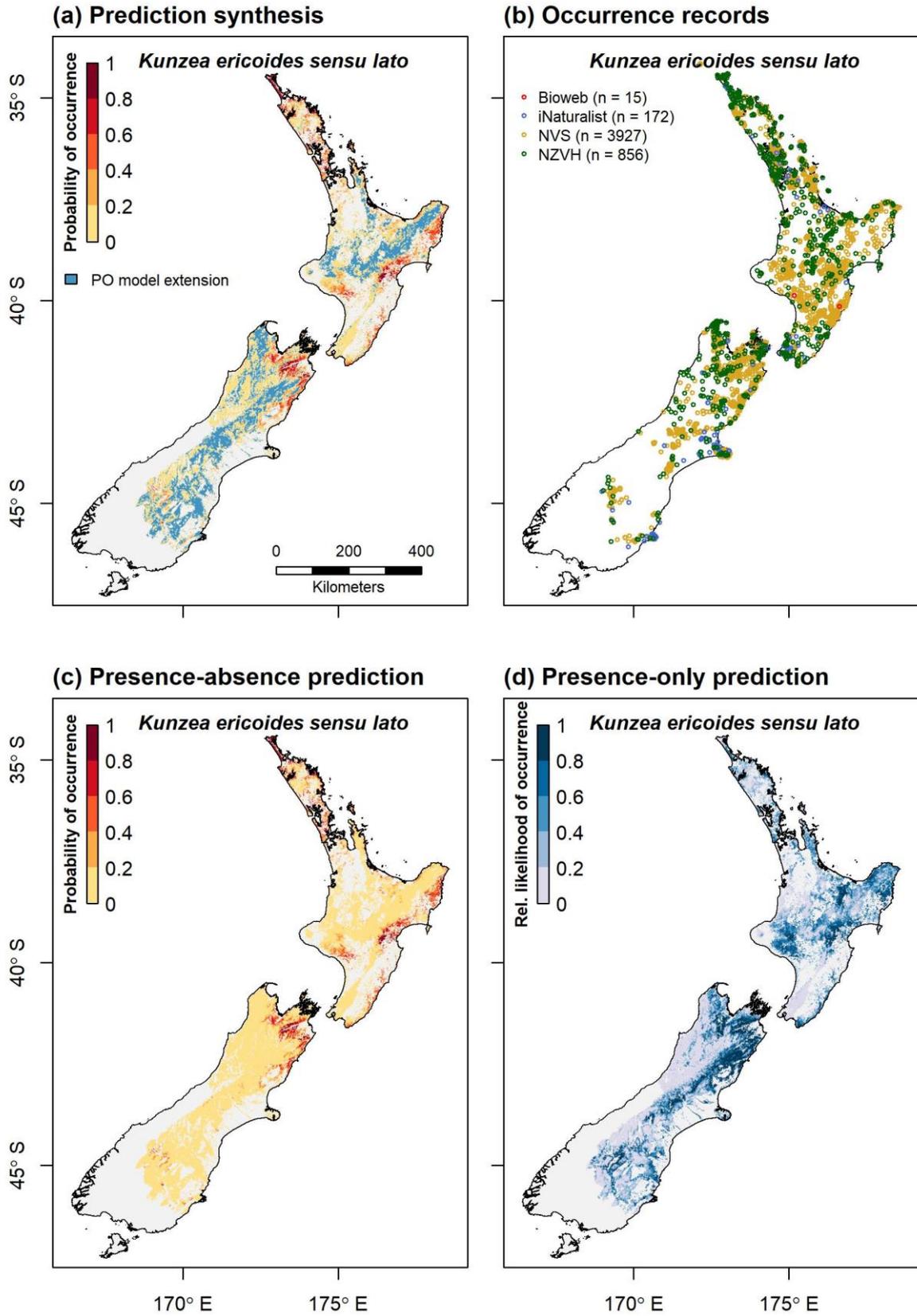
## 9.1 *Kunzea amathicola*



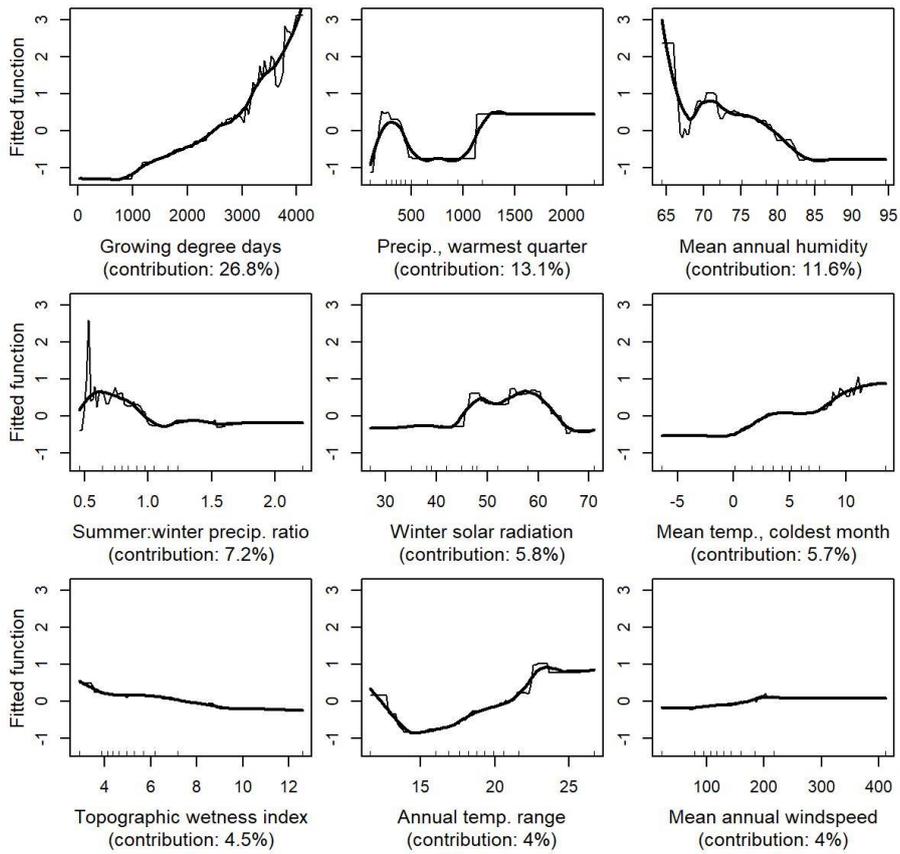
### Presence-only model



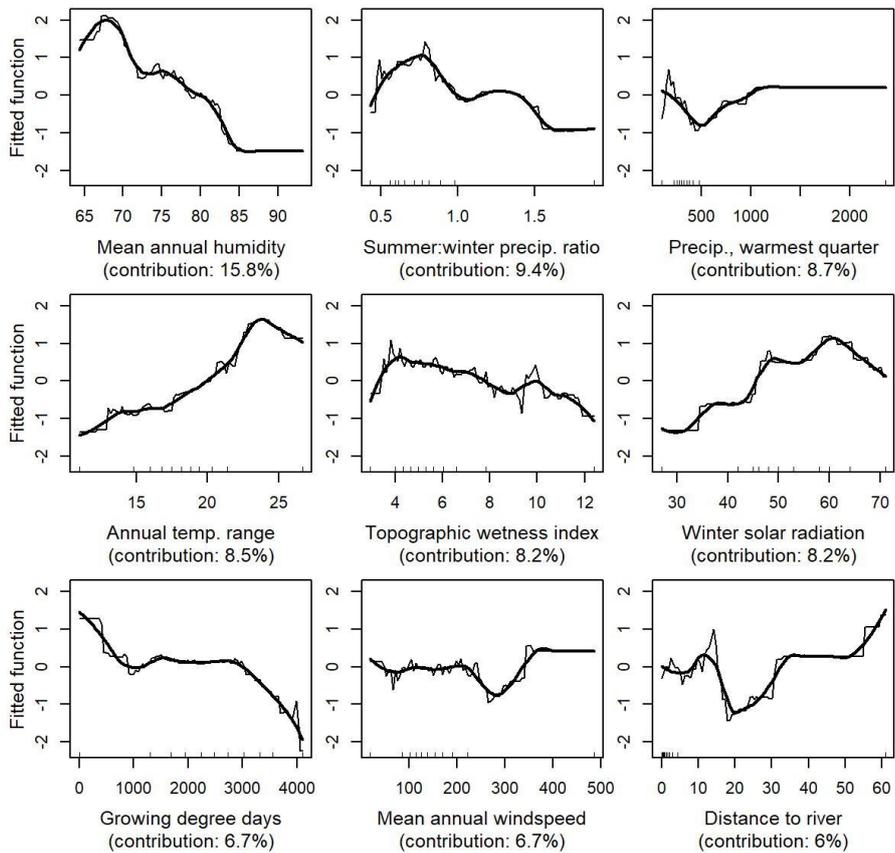
## 9.2 *Kunzea ericoides sensu lato*



### Presence-absence model

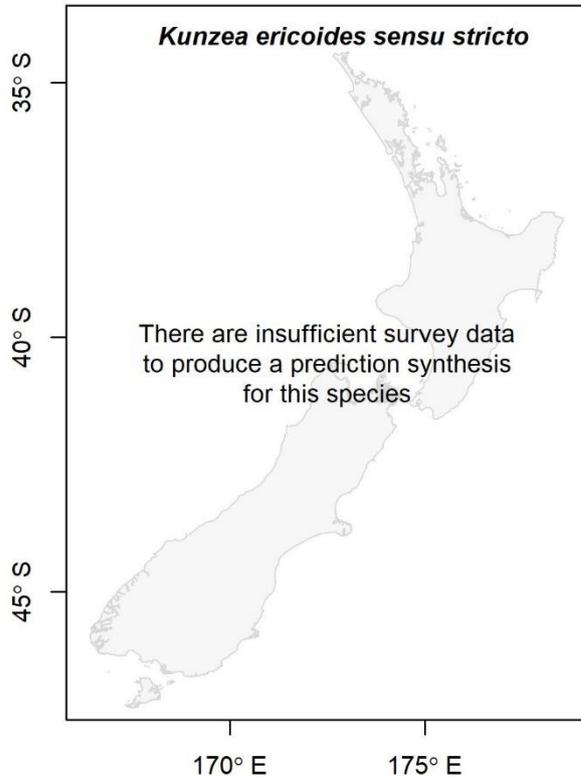


### Presence-only model

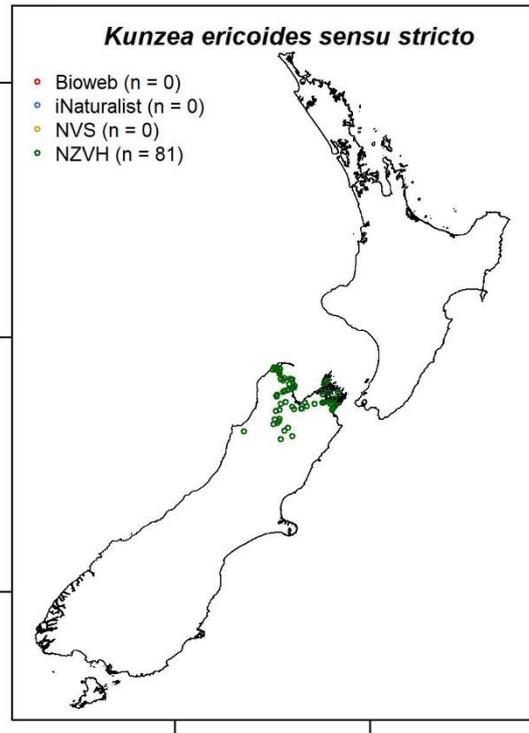


### 9.3 *Kunzea ericoides sensu stricto*

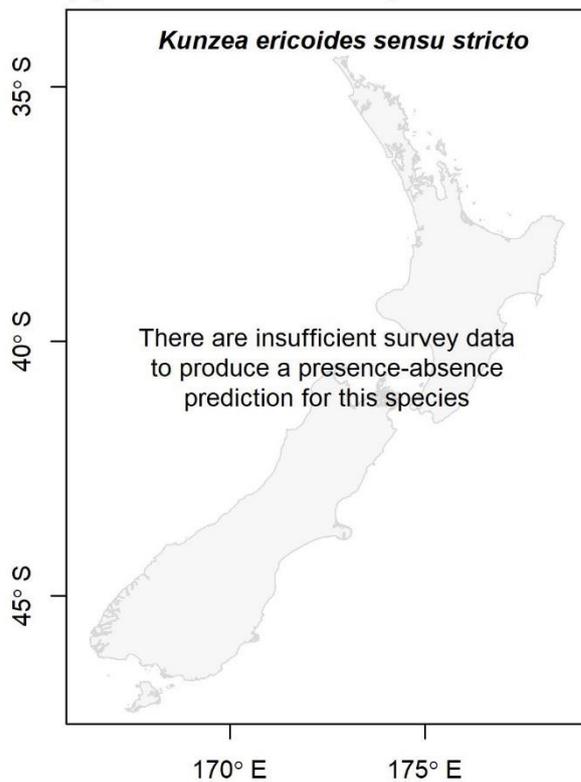
(a) Prediction synthesis



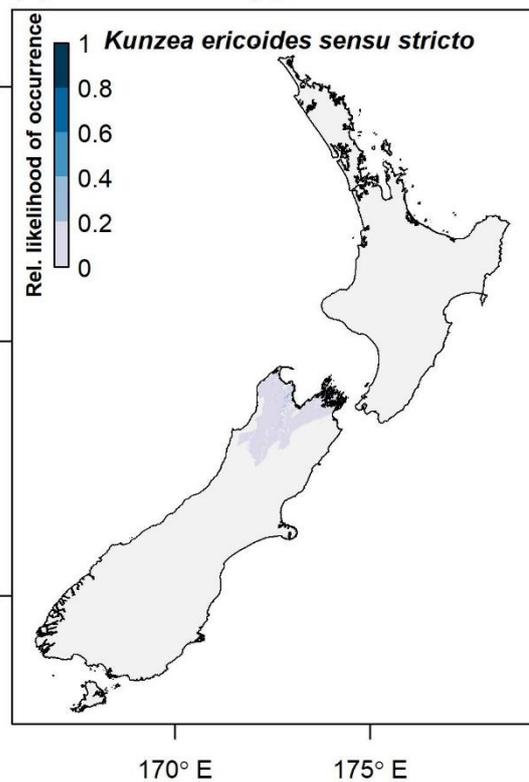
(b) Occurrence records



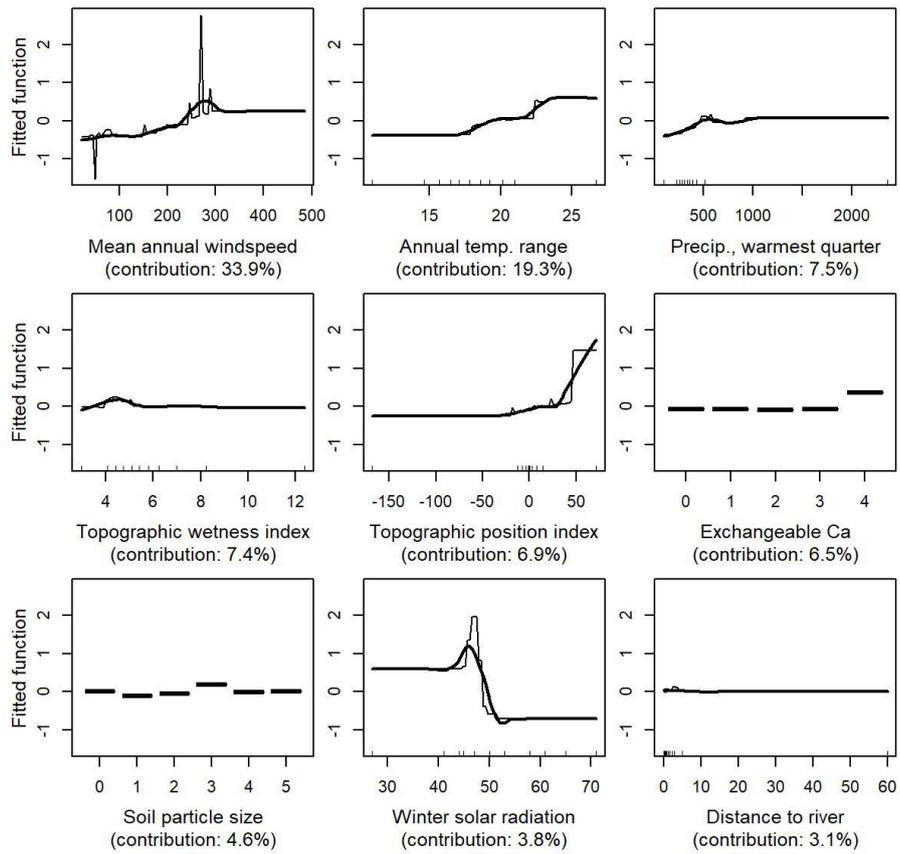
(c) Presence-absence prediction



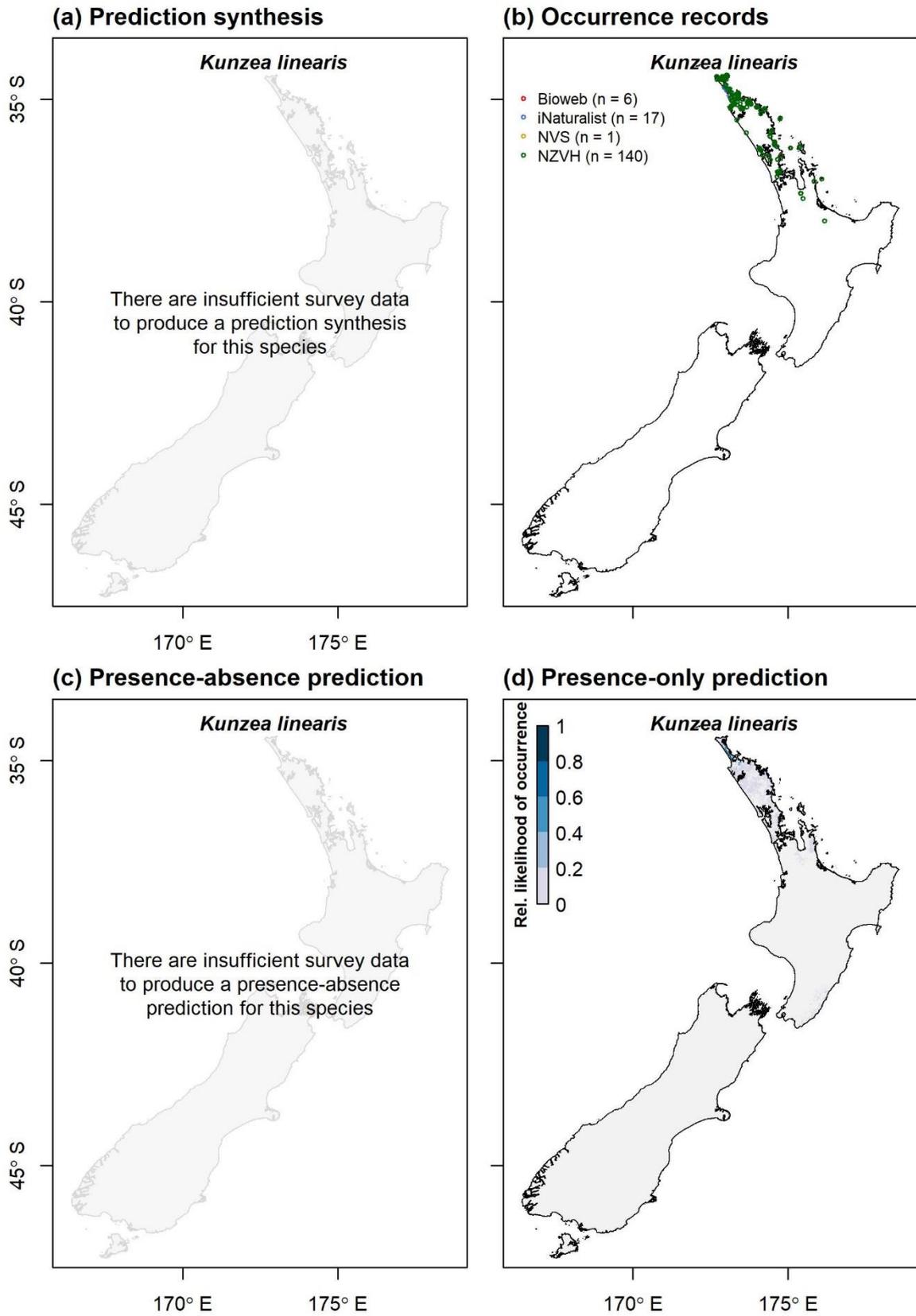
(d) Presence-only prediction



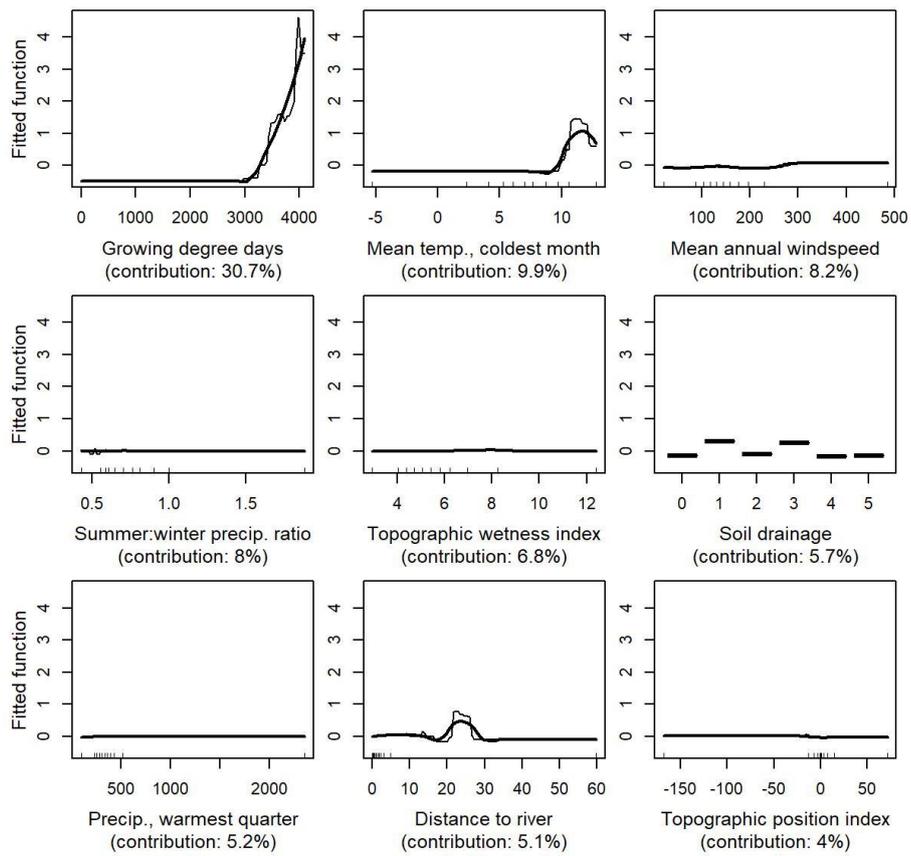
### Presence-only model



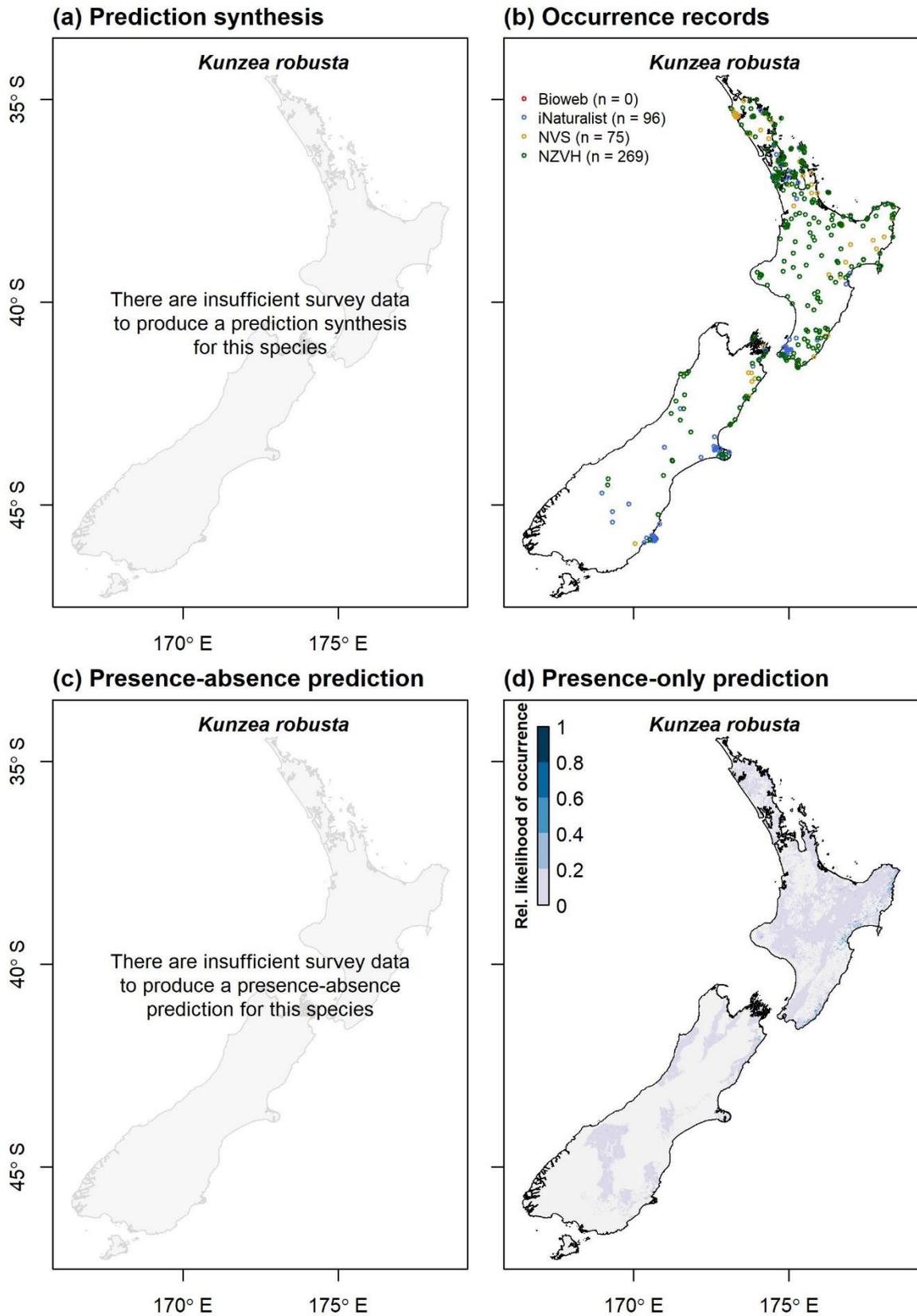
## 9.4 *Kunzea linearis*



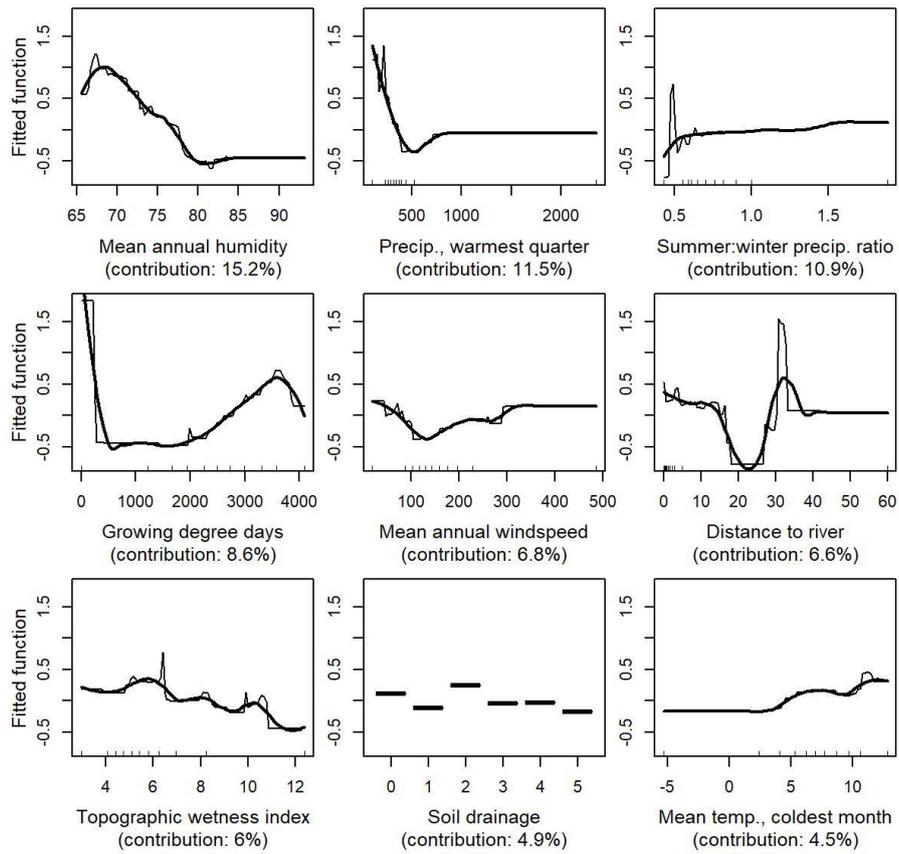
### Presence-only model



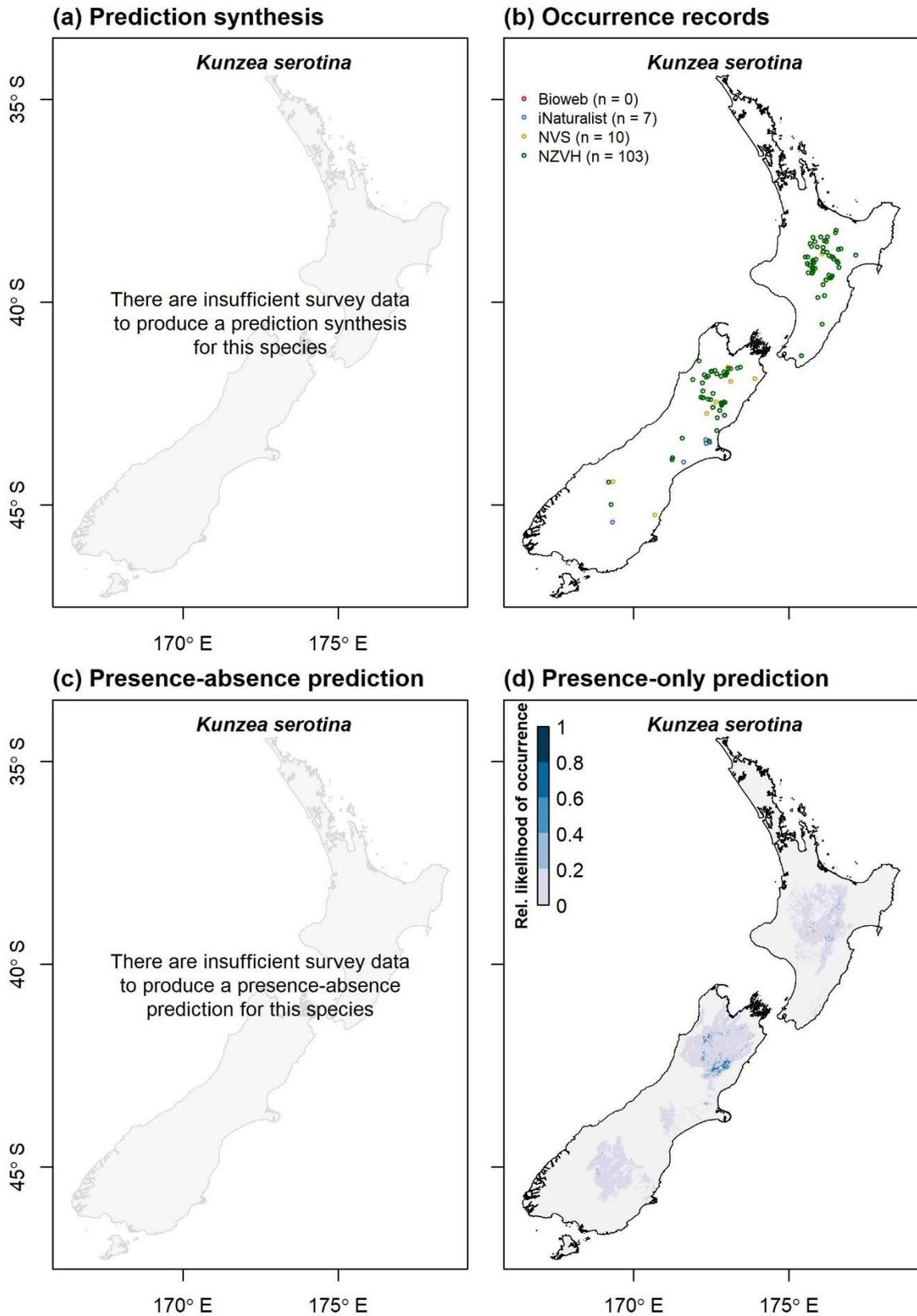
## 9.5 *Kunzea robusta*



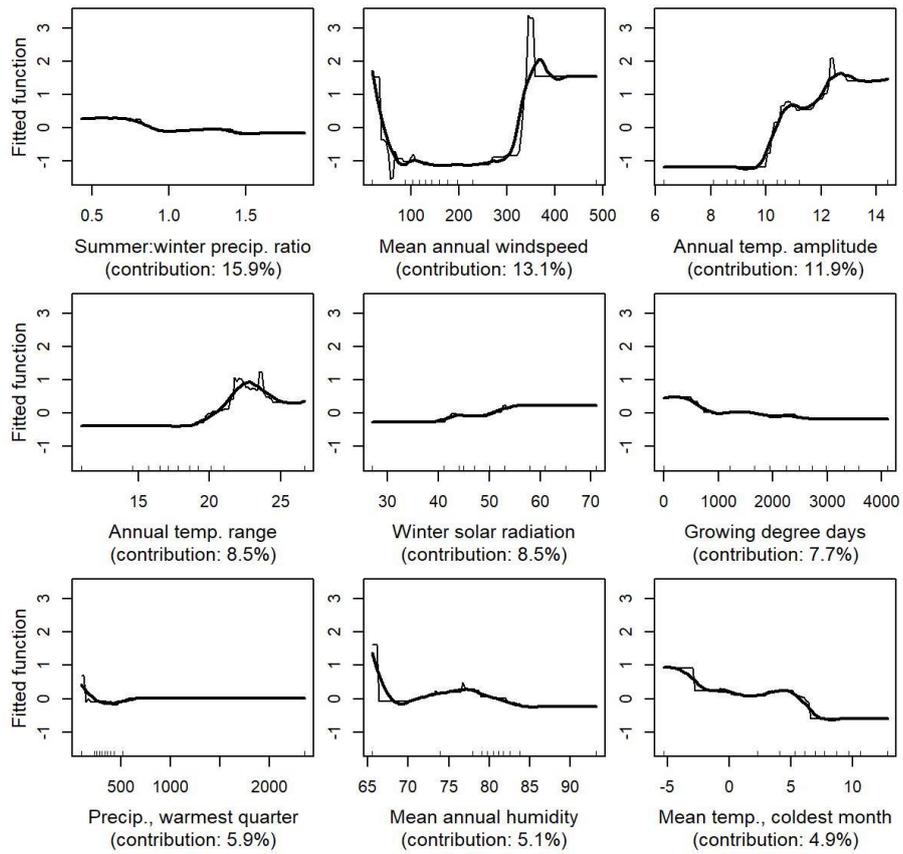
### Presence-only model



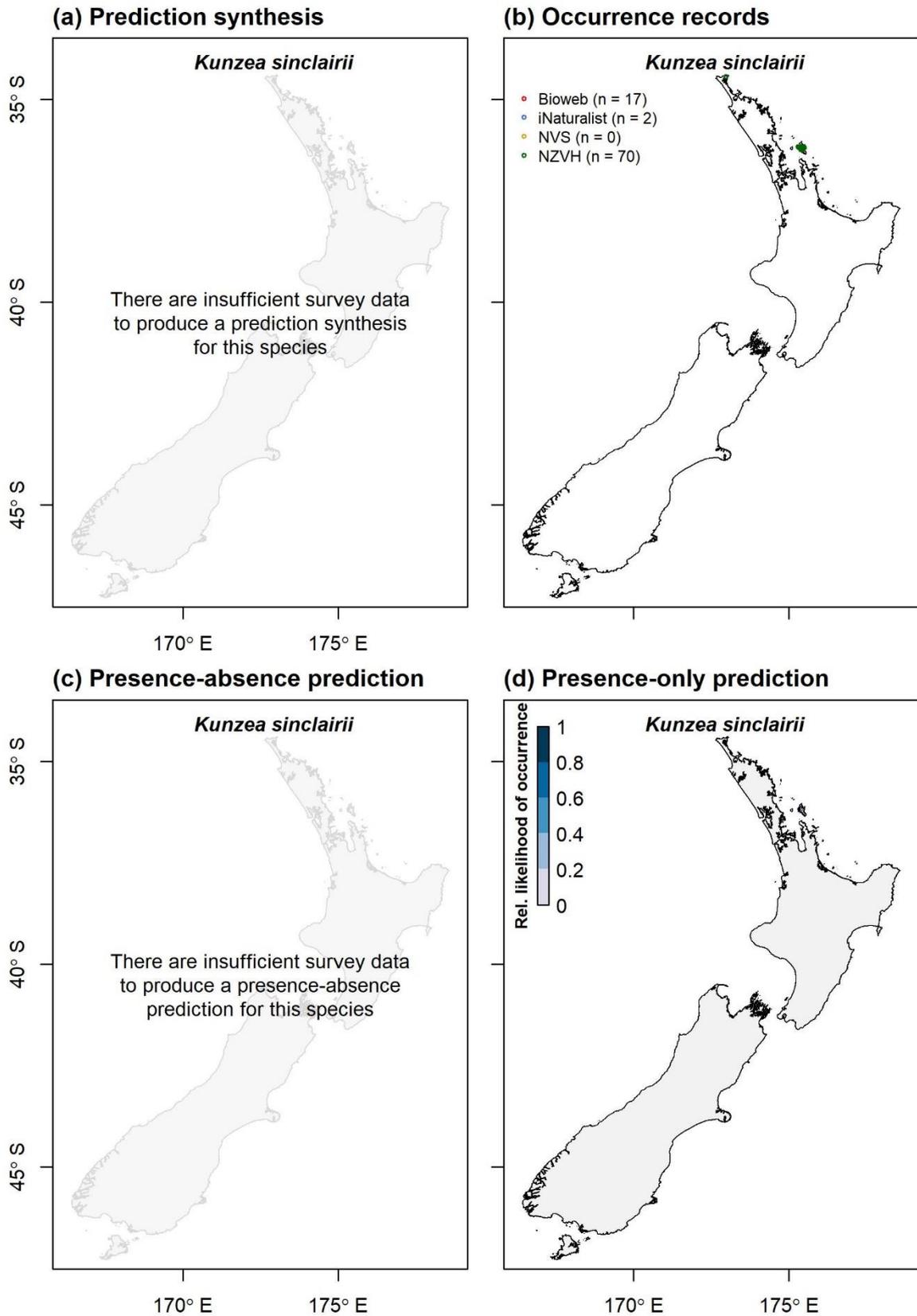
## 9.6 *Kunzea serotina*



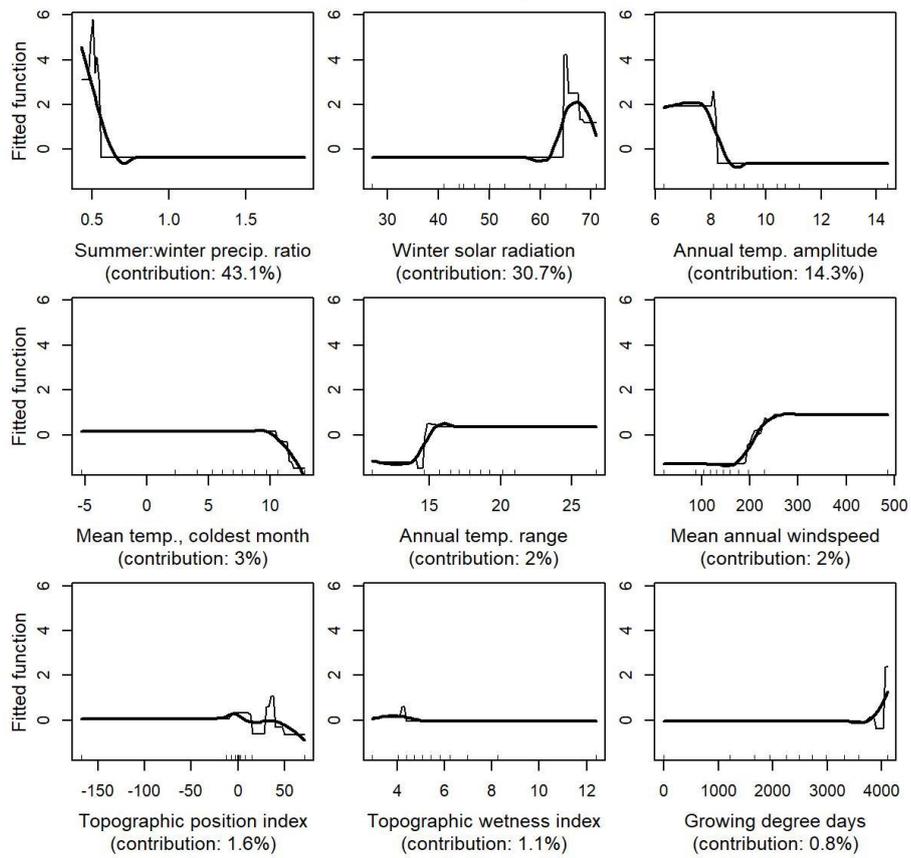
### Presence-only model



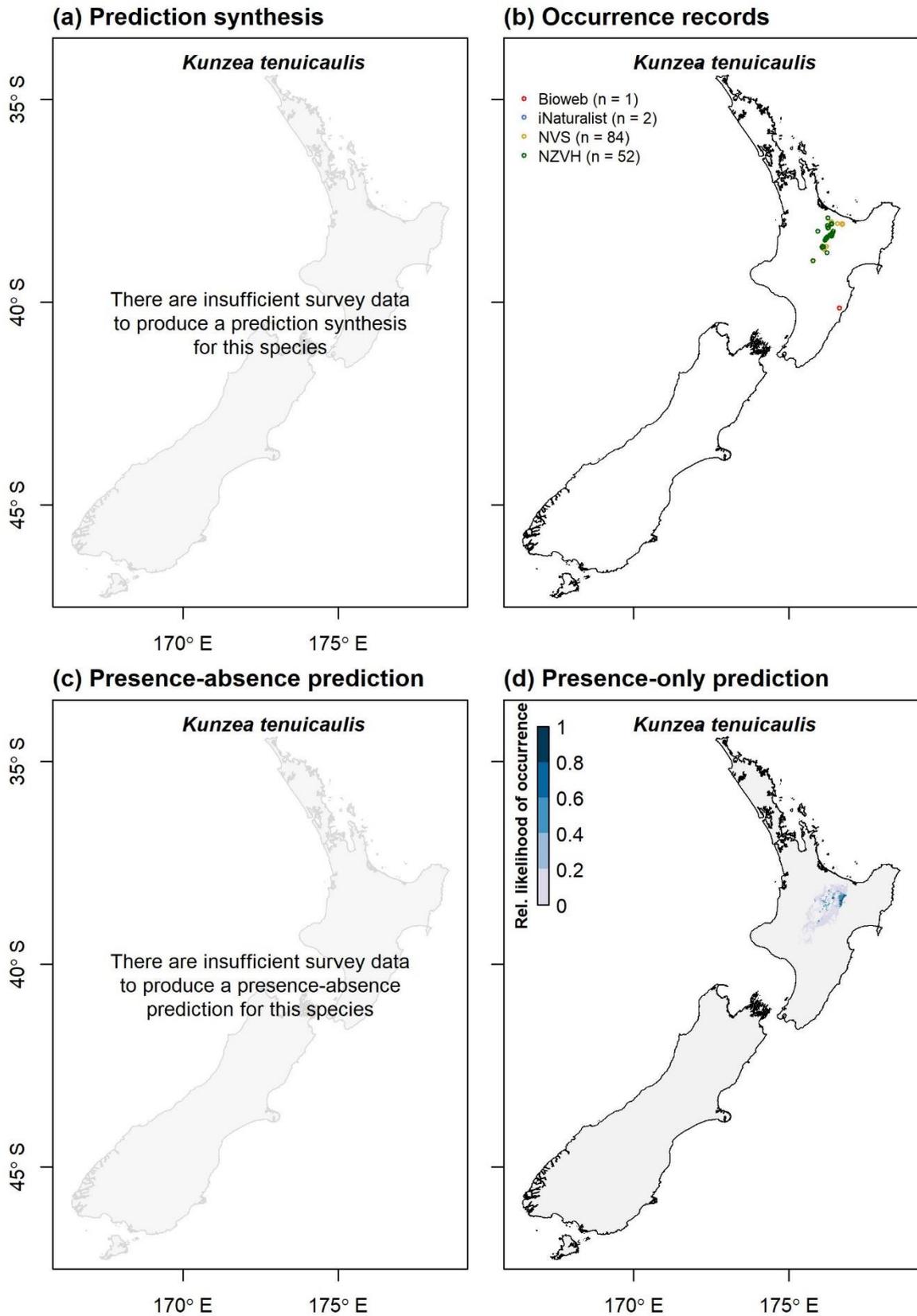
## 9.7 *Kunzea sinclairii*



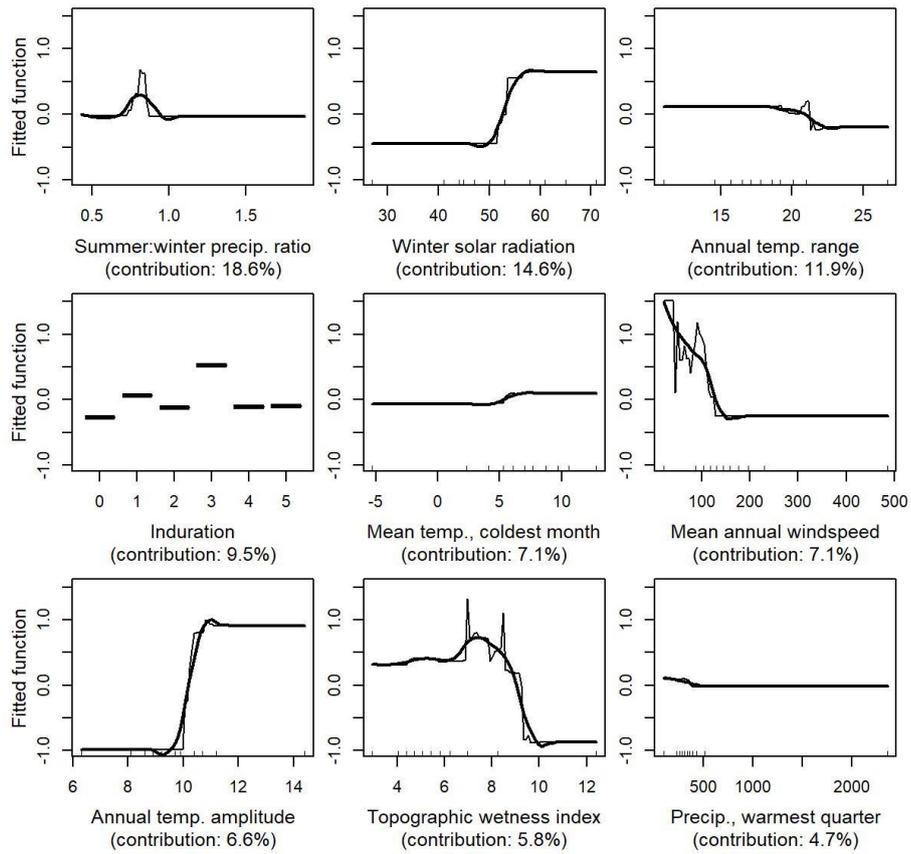
### Presence-only model



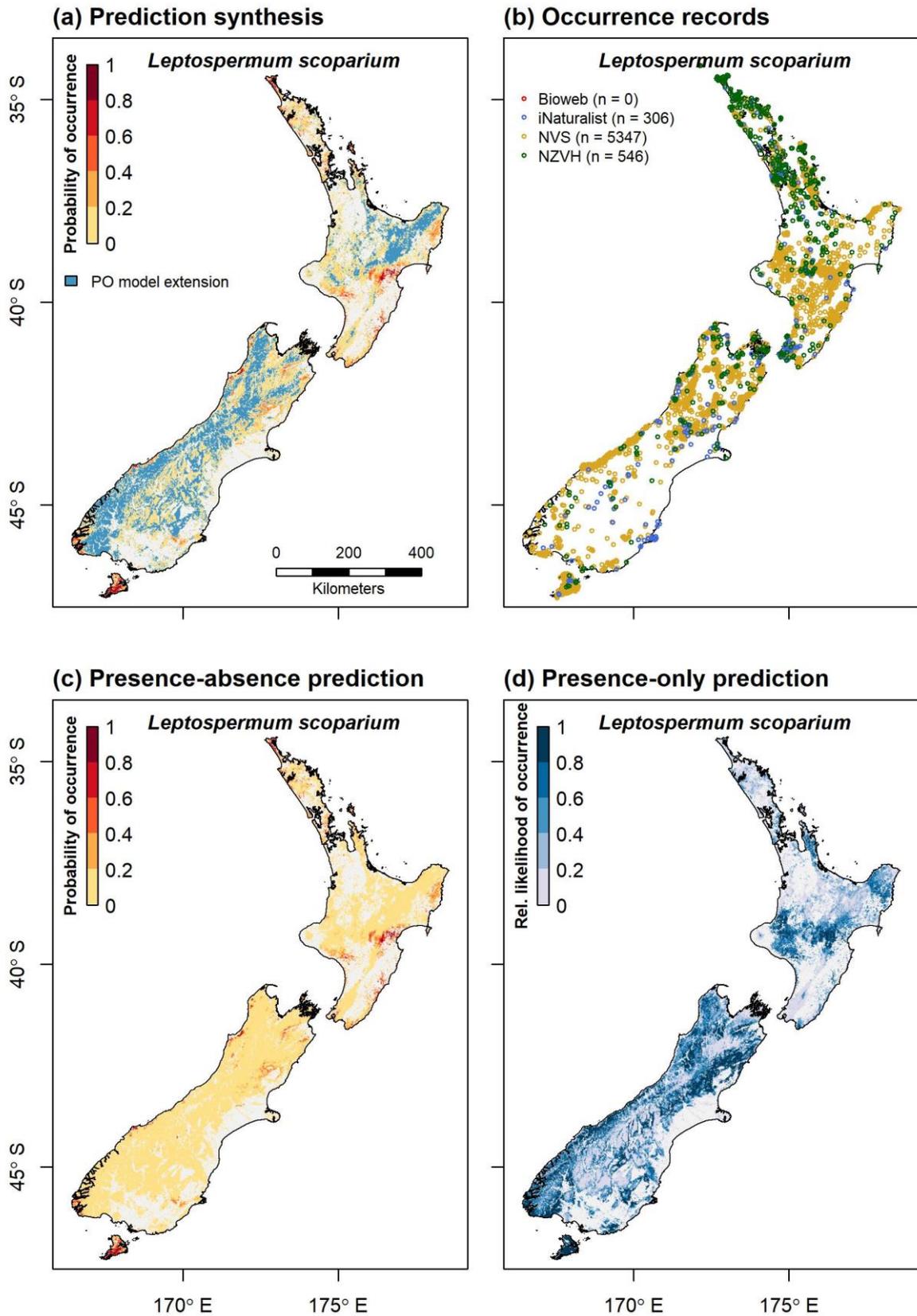
## 9.8 *Kunzea tenuicaulis*



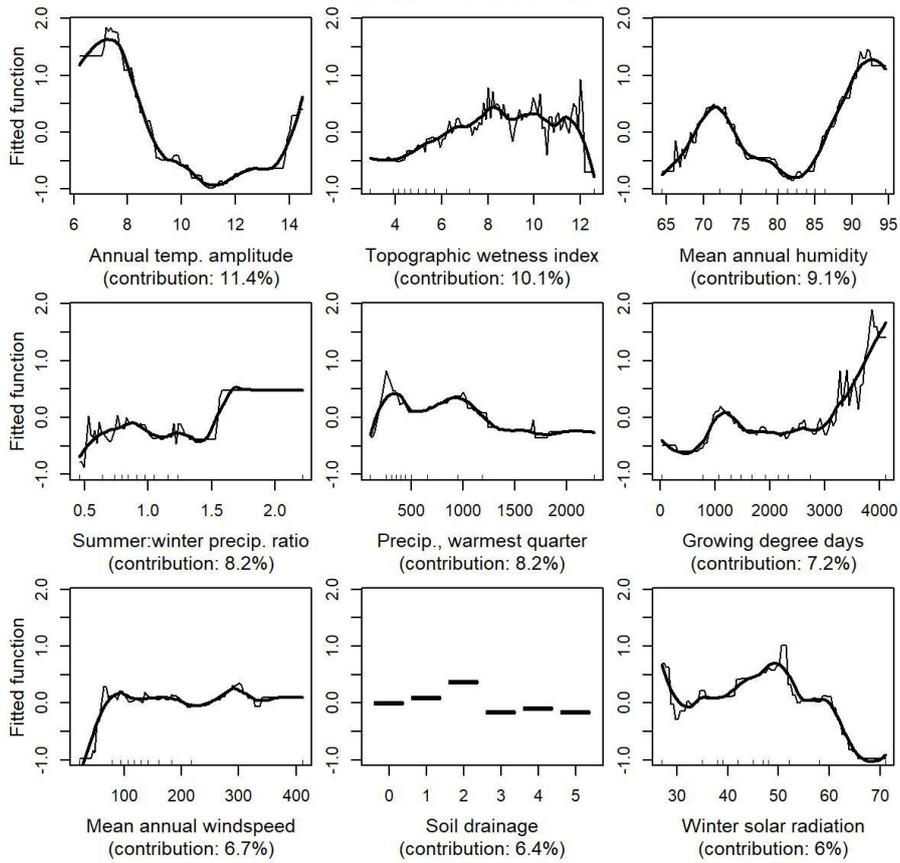
### Presence-only model



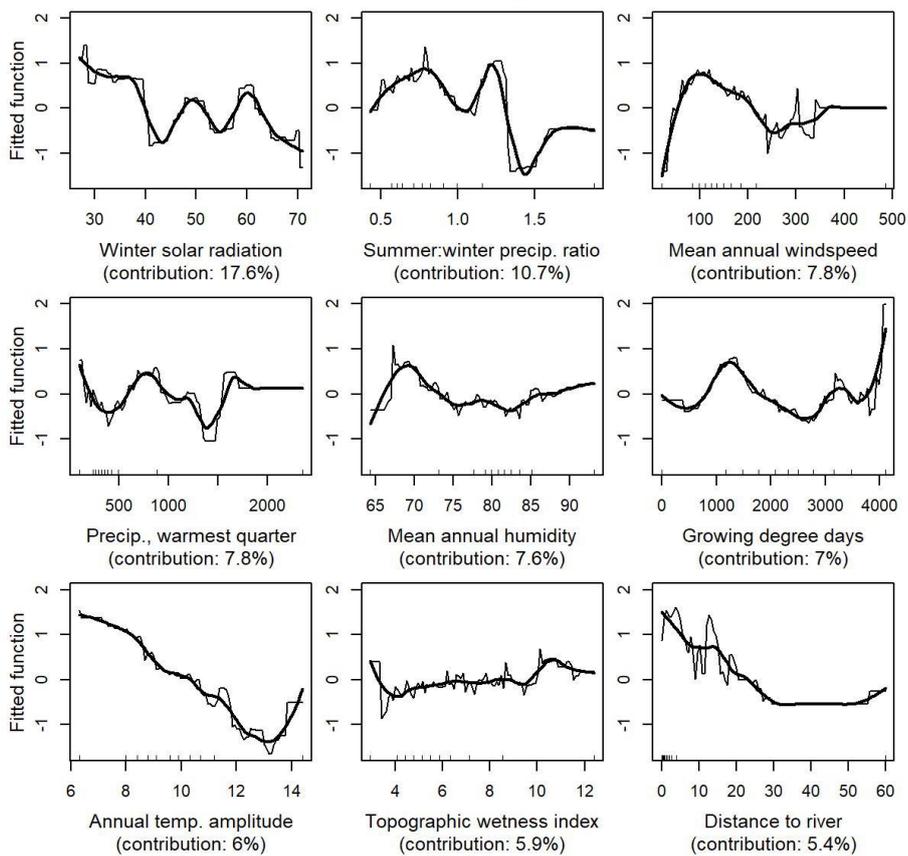
## 9.9 *Leptospermum scoparium*



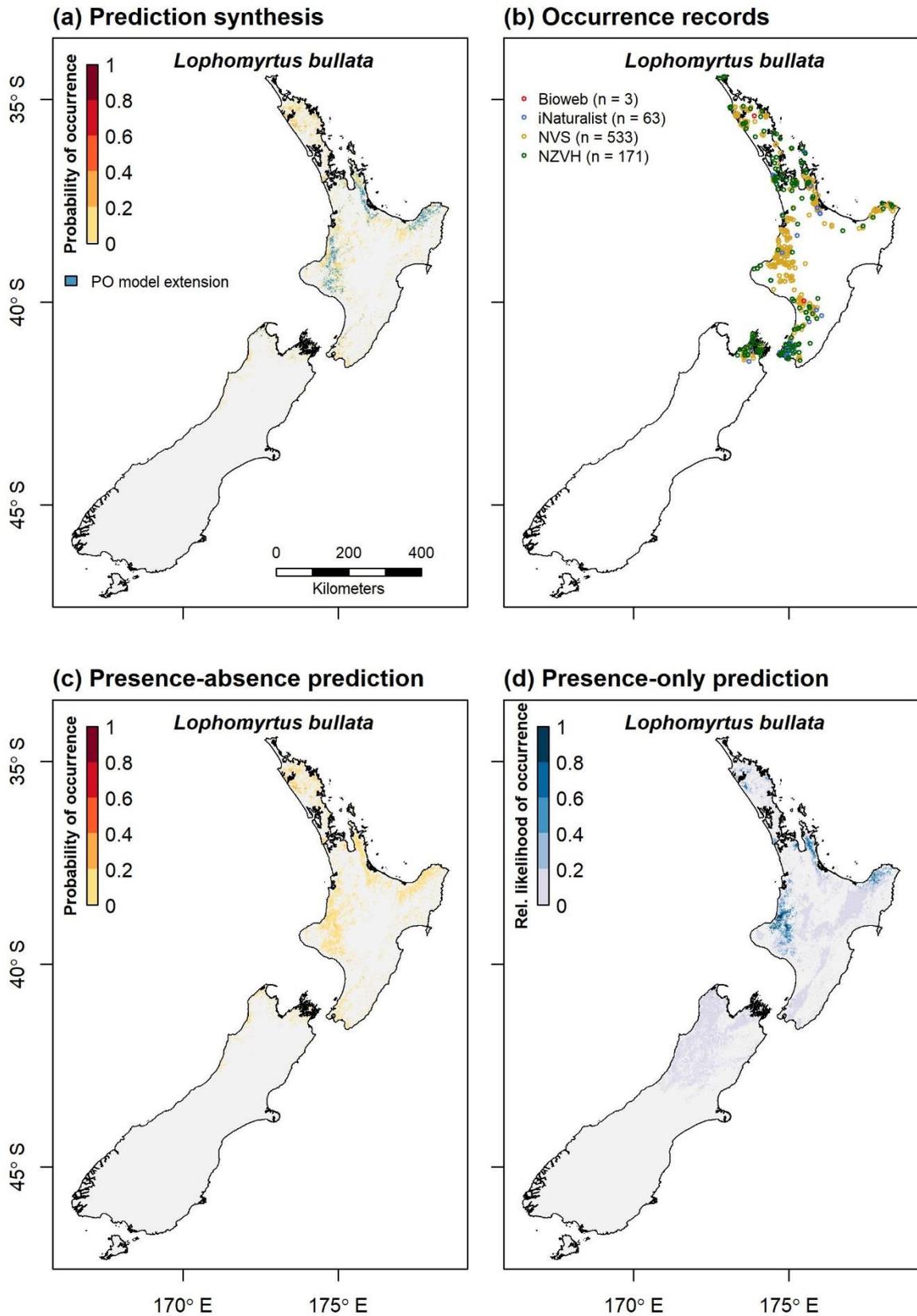
### Presence-absence model



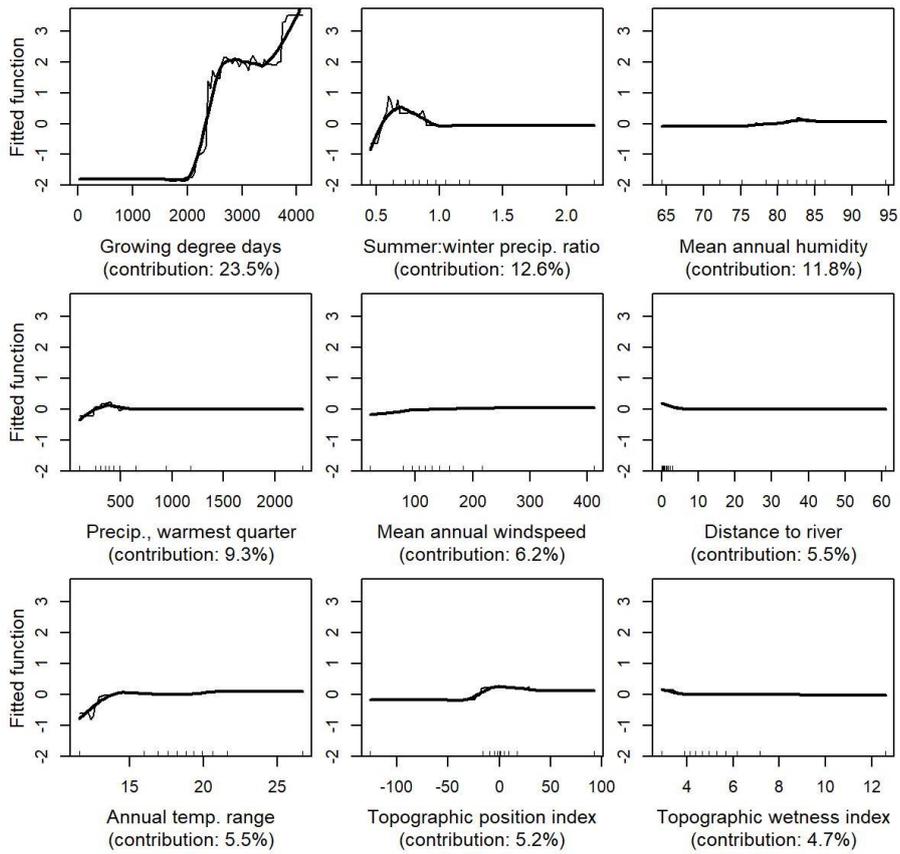
### Presence-only model



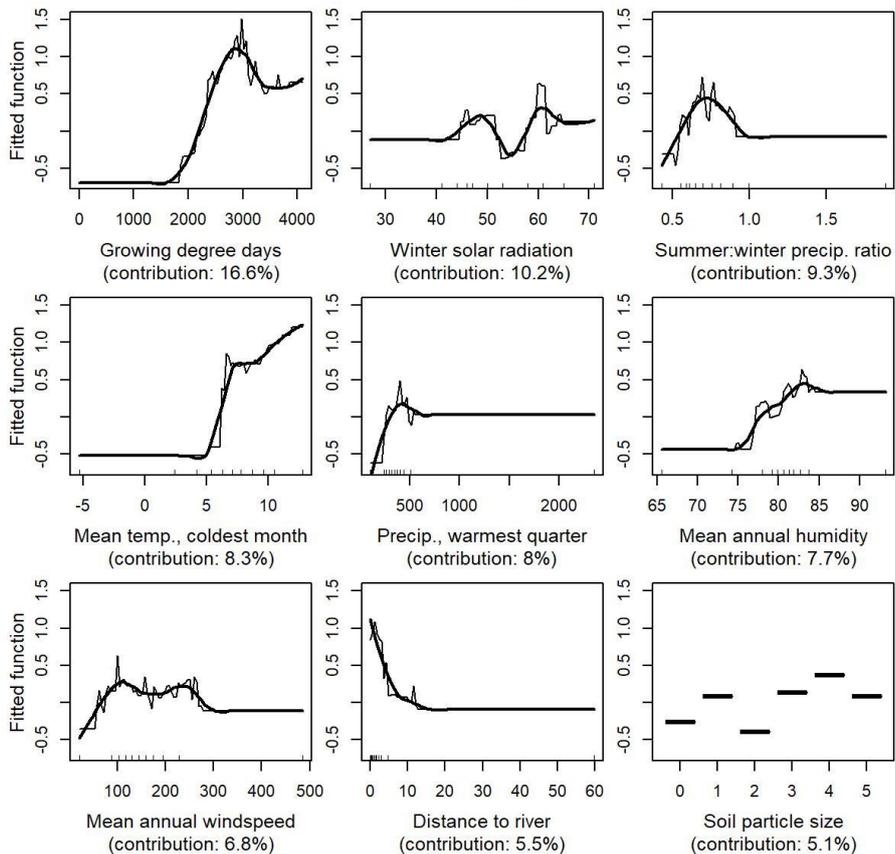
## 9.10 *Lophomyrtus bullata*



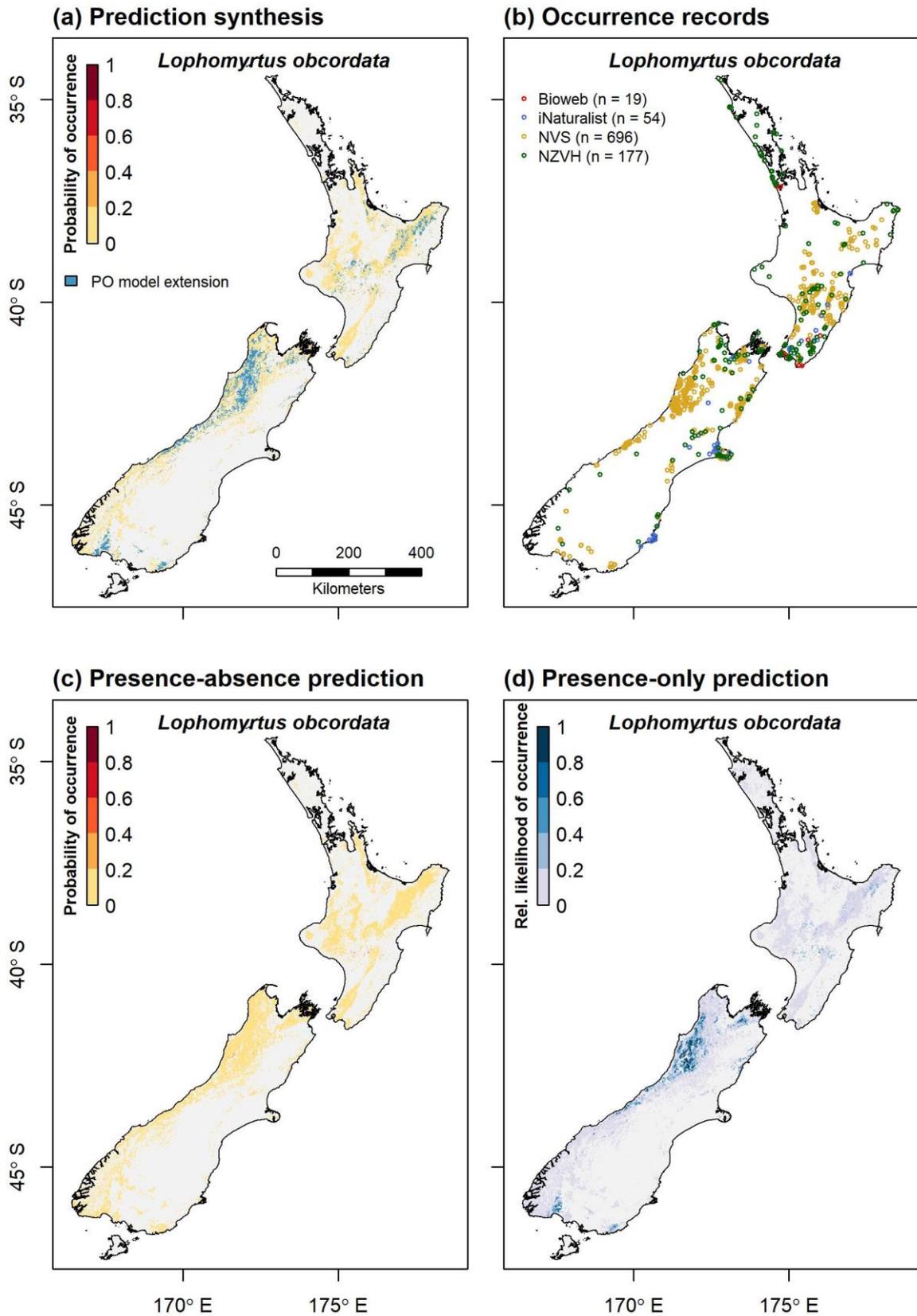
### Presence-absence model



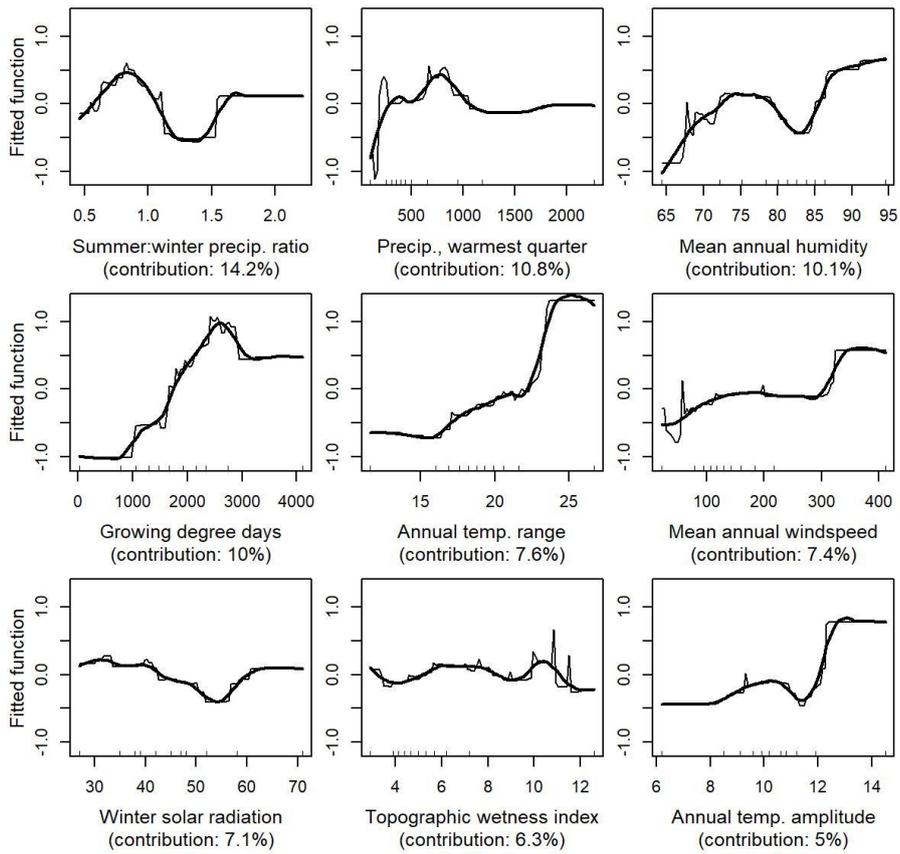
### Presence-only model



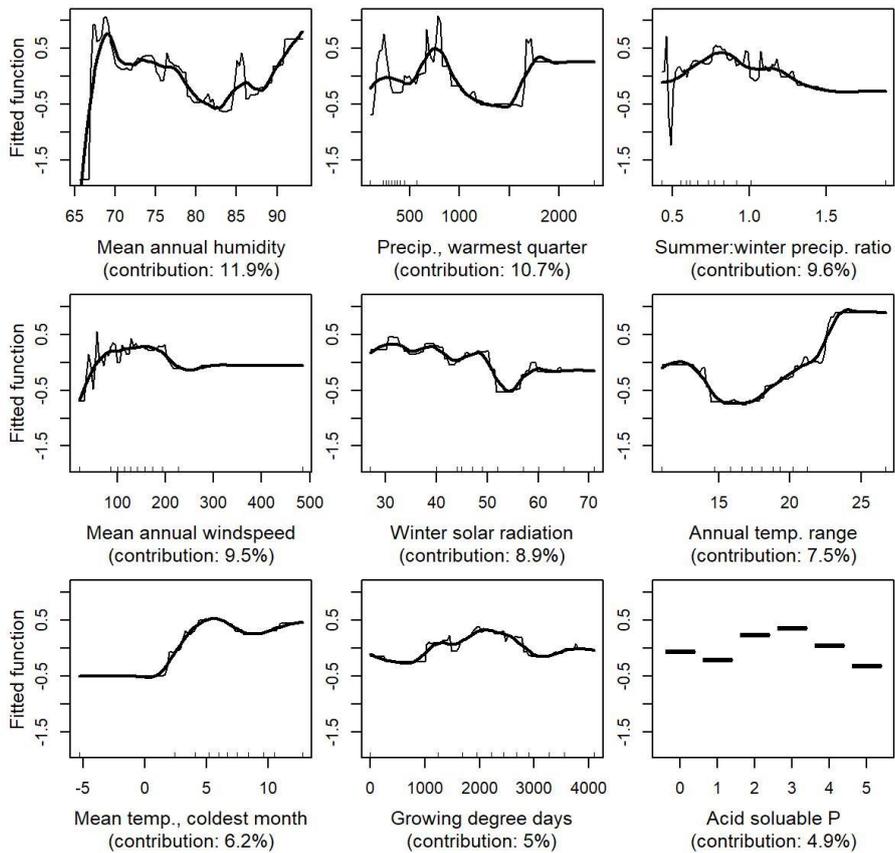
## 9.11 *Lophomyrtus obcordata*



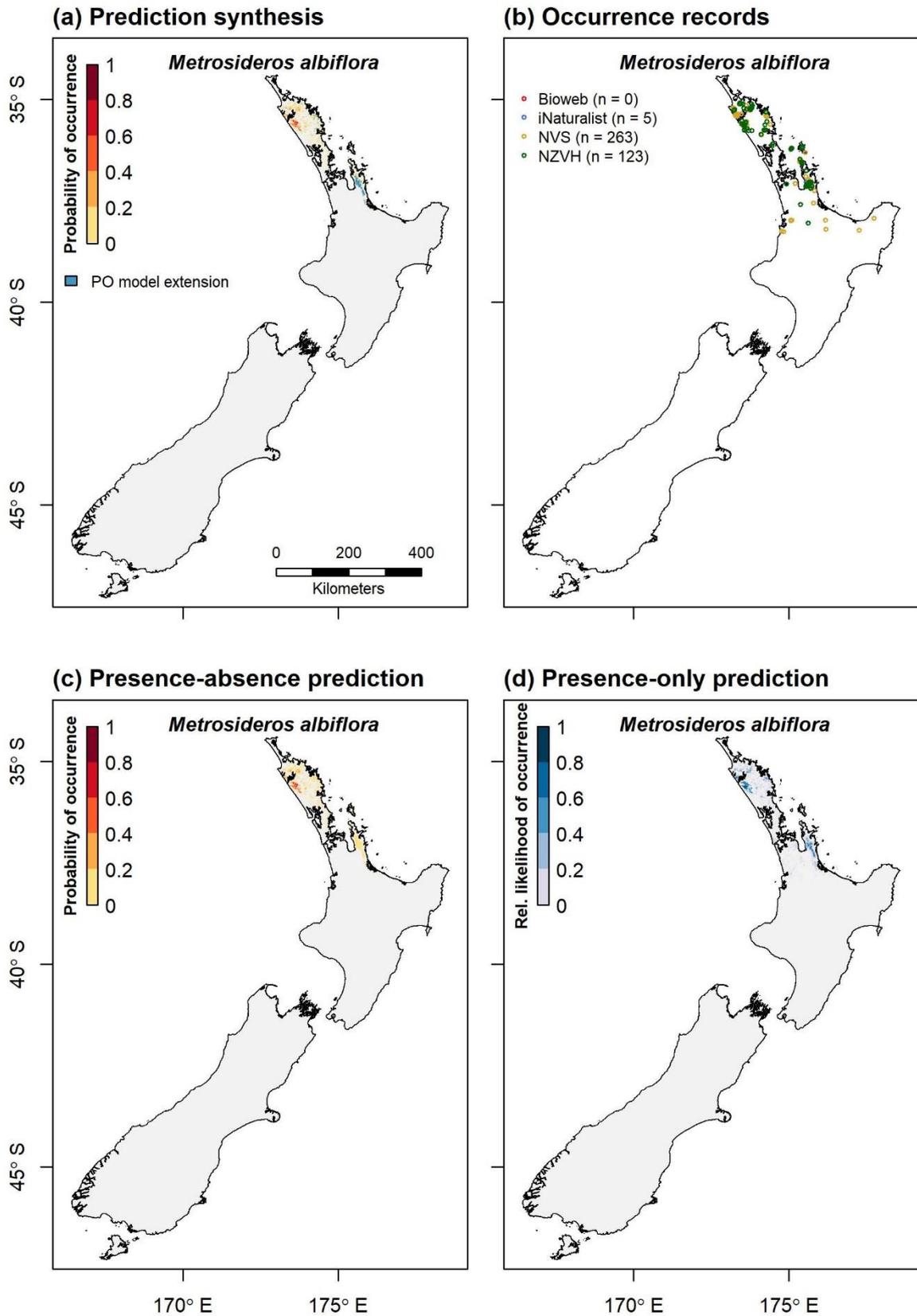
### Presence-absence model



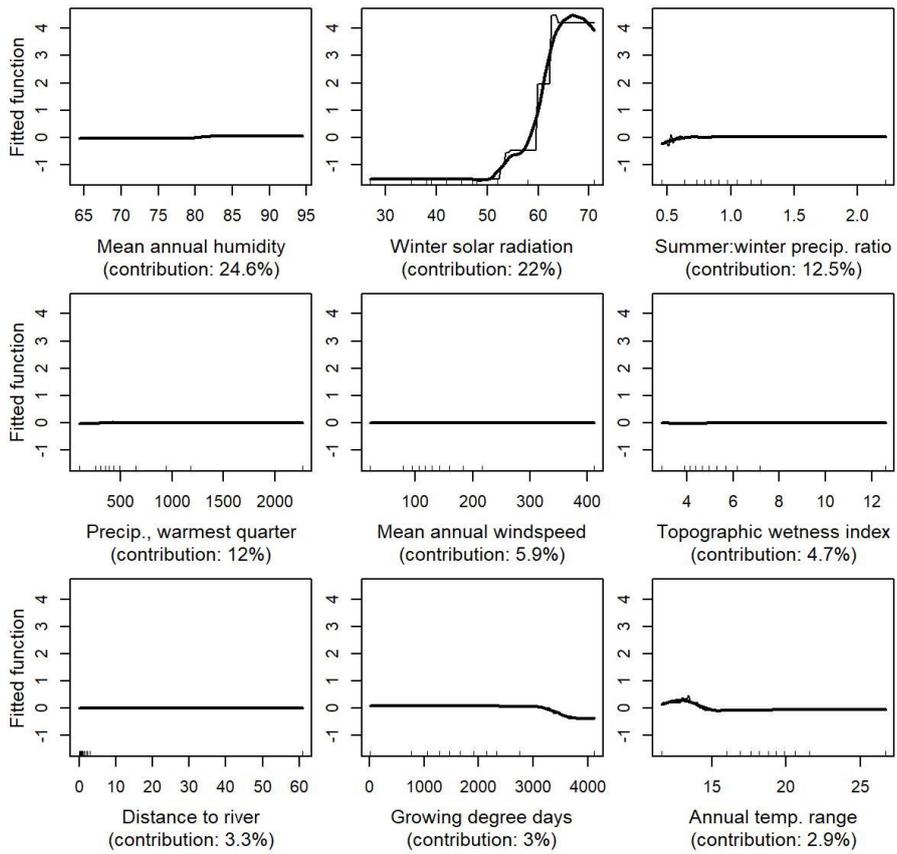
### Presence-only model



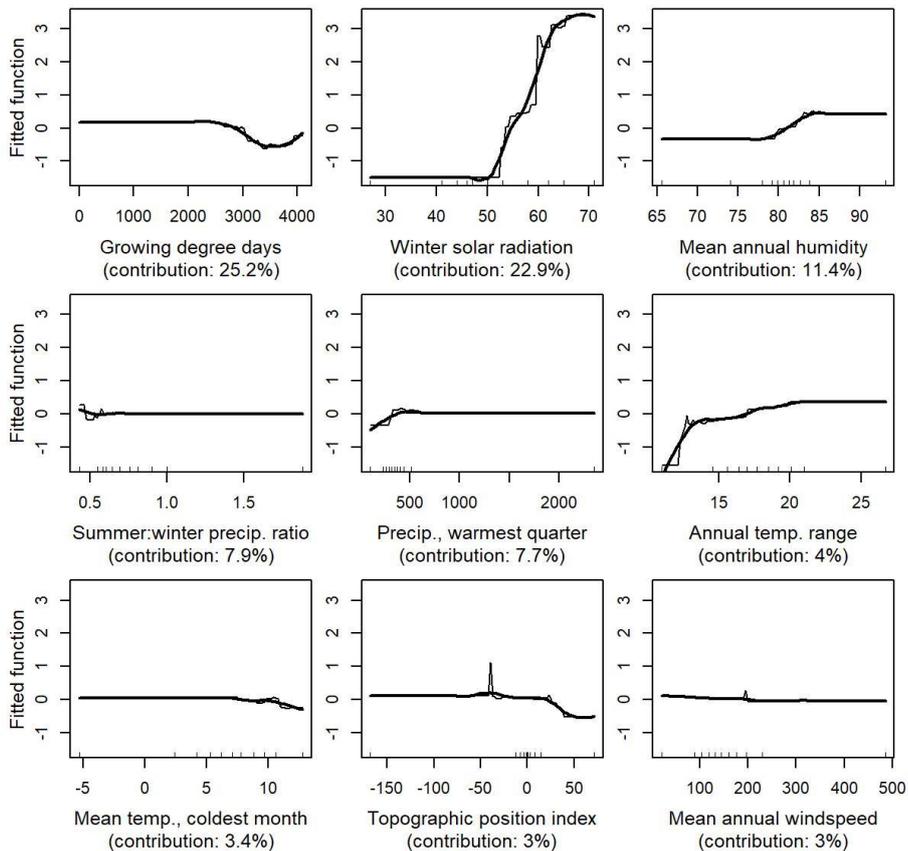
## 9.12 *Metrosideros albiflora*



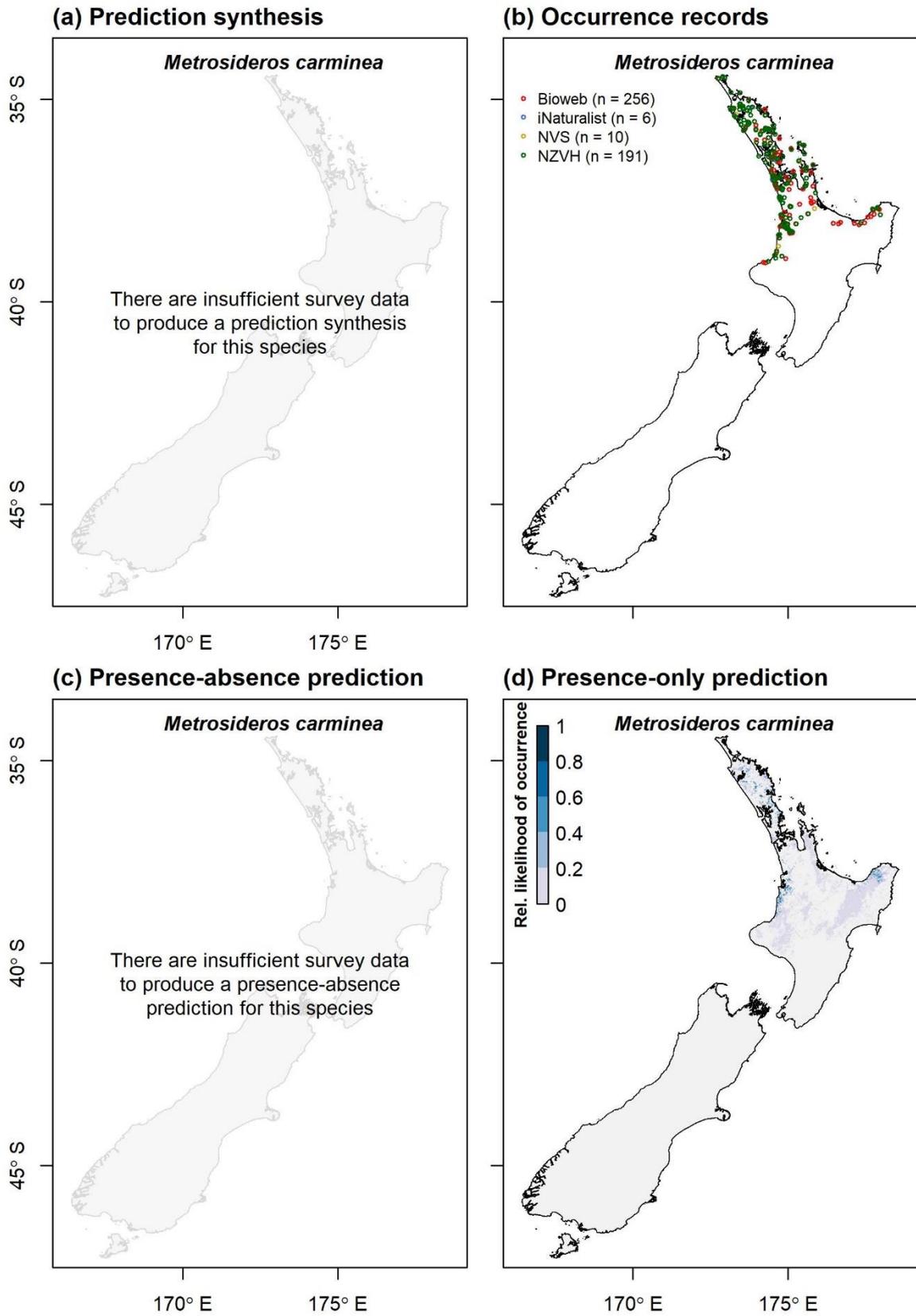
### Presence-absence model



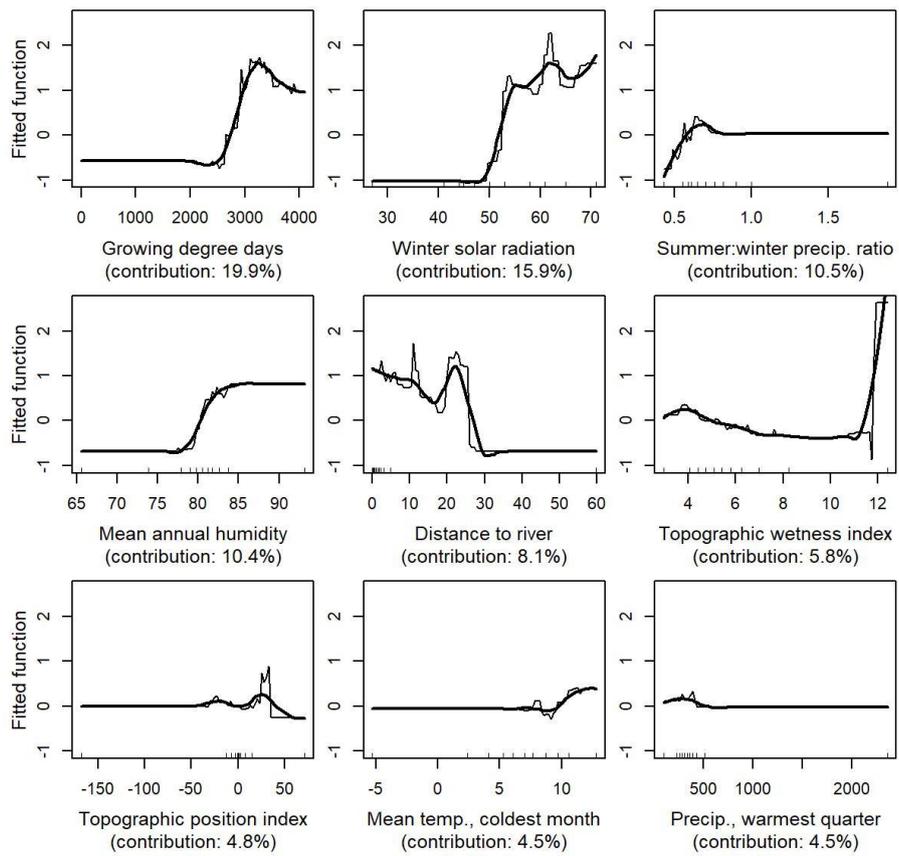
### Presence-only model



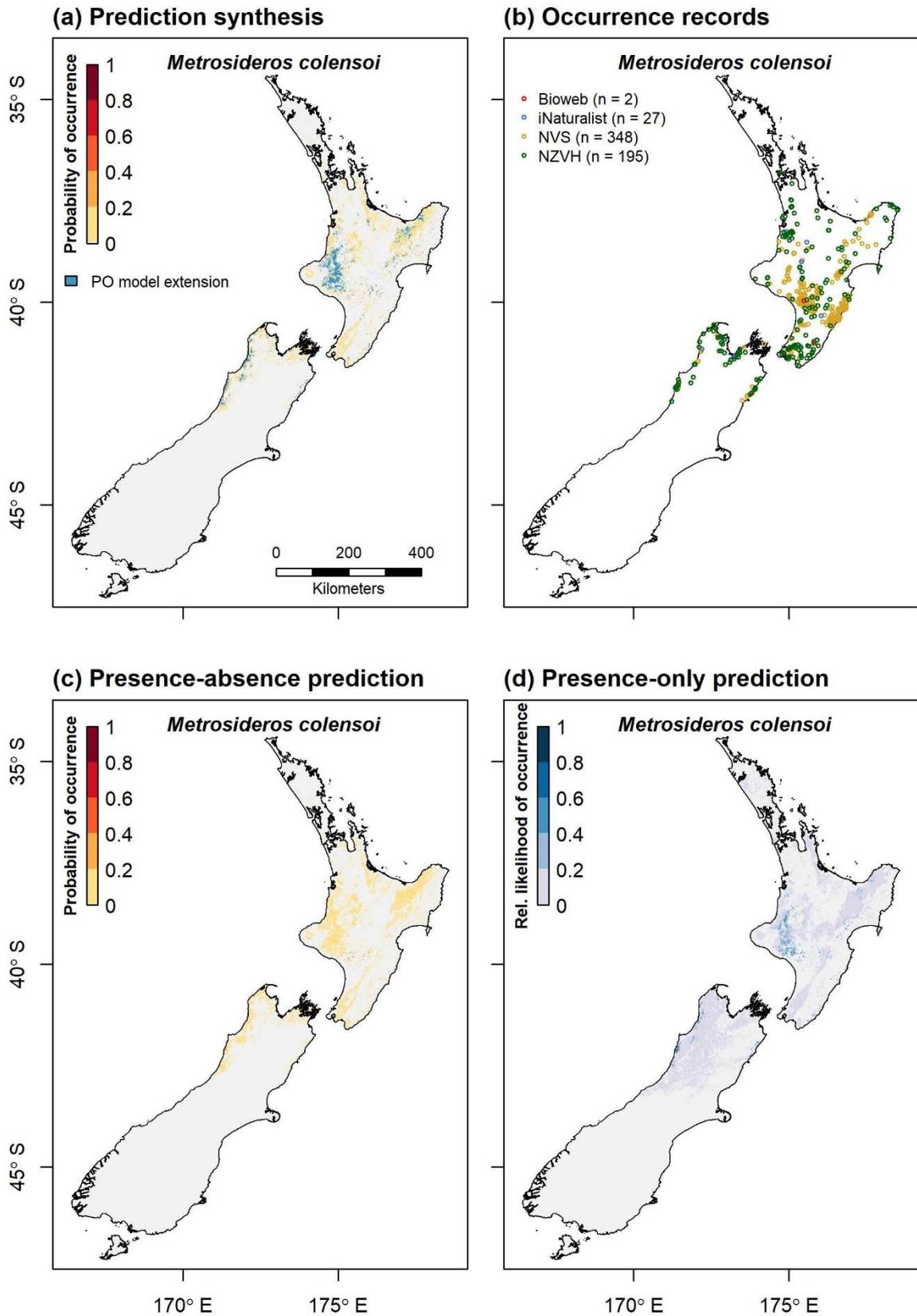
### 9.13 *Metrosideros carminea*



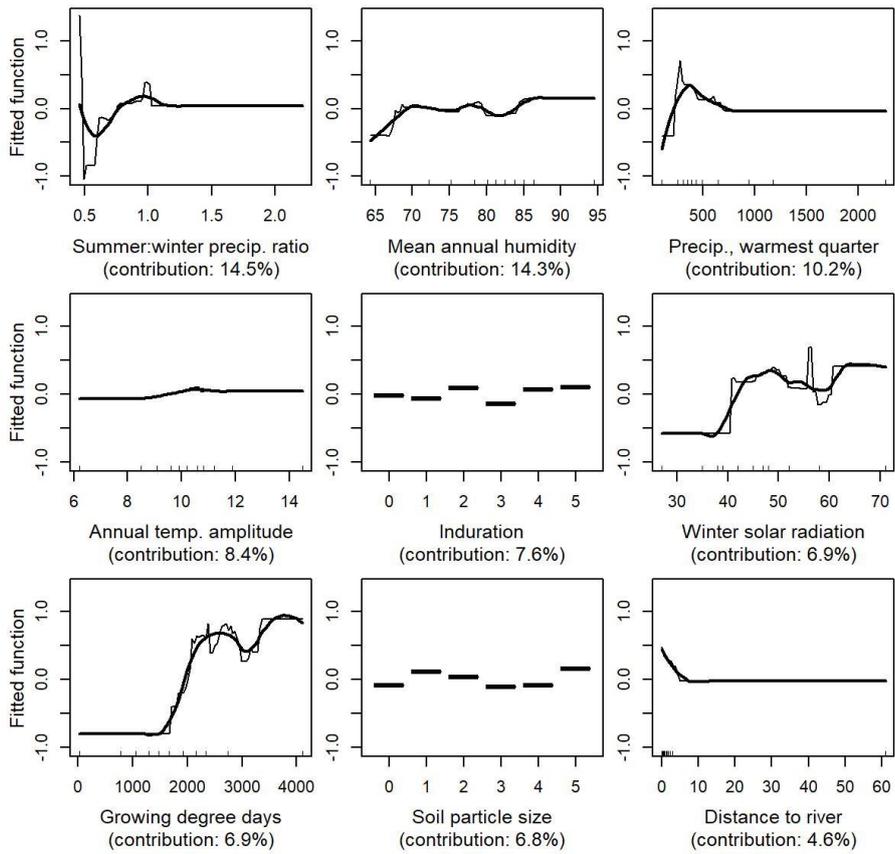
### Presence-only model



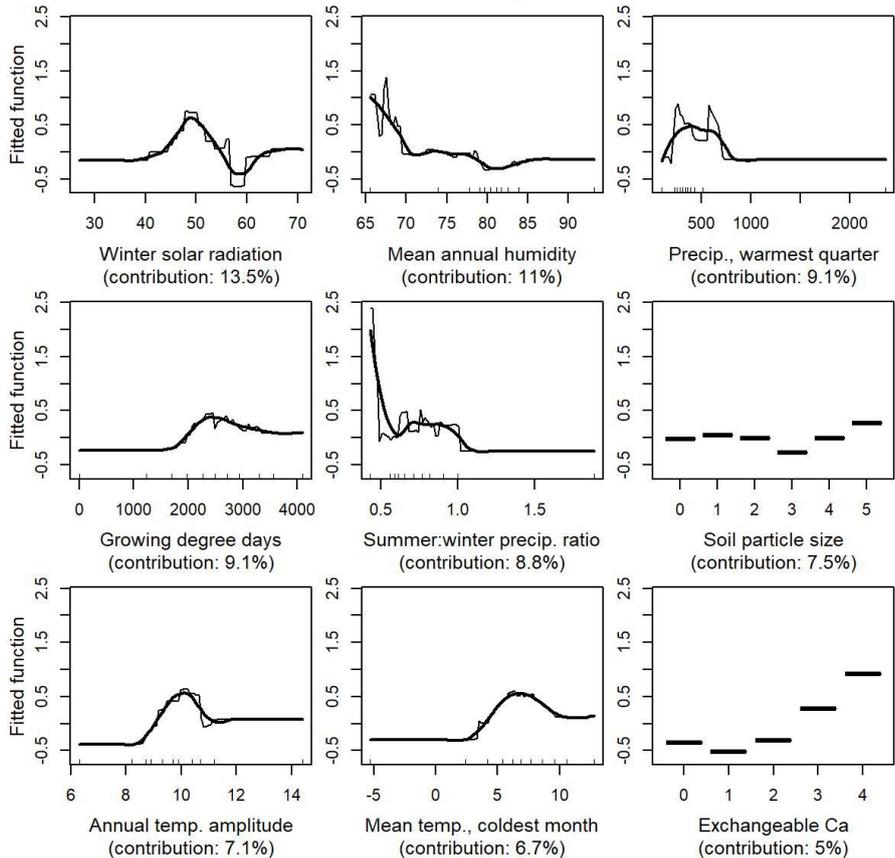
## 9.14 *Metrosideros colensoi*



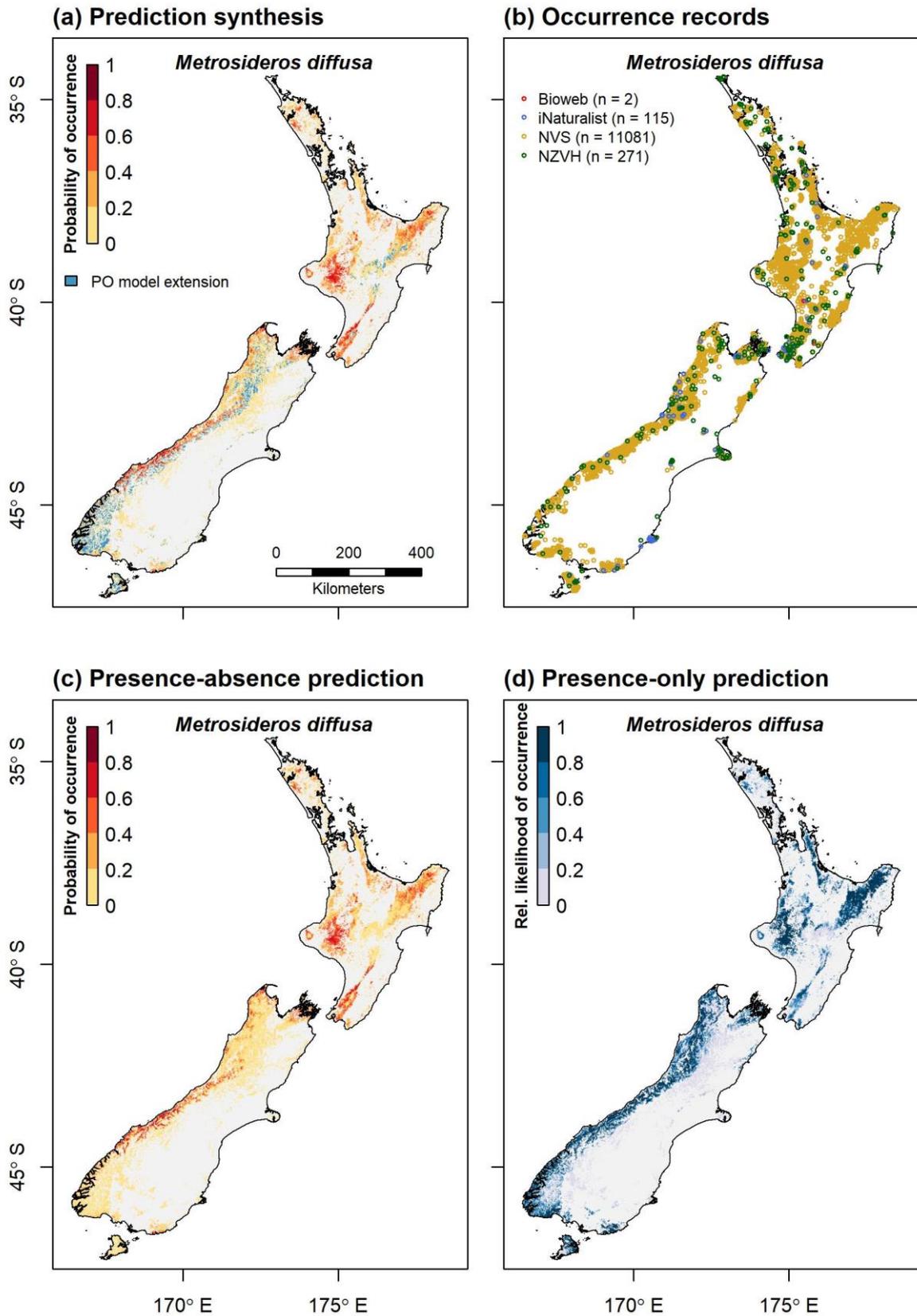
### Presence-absence model



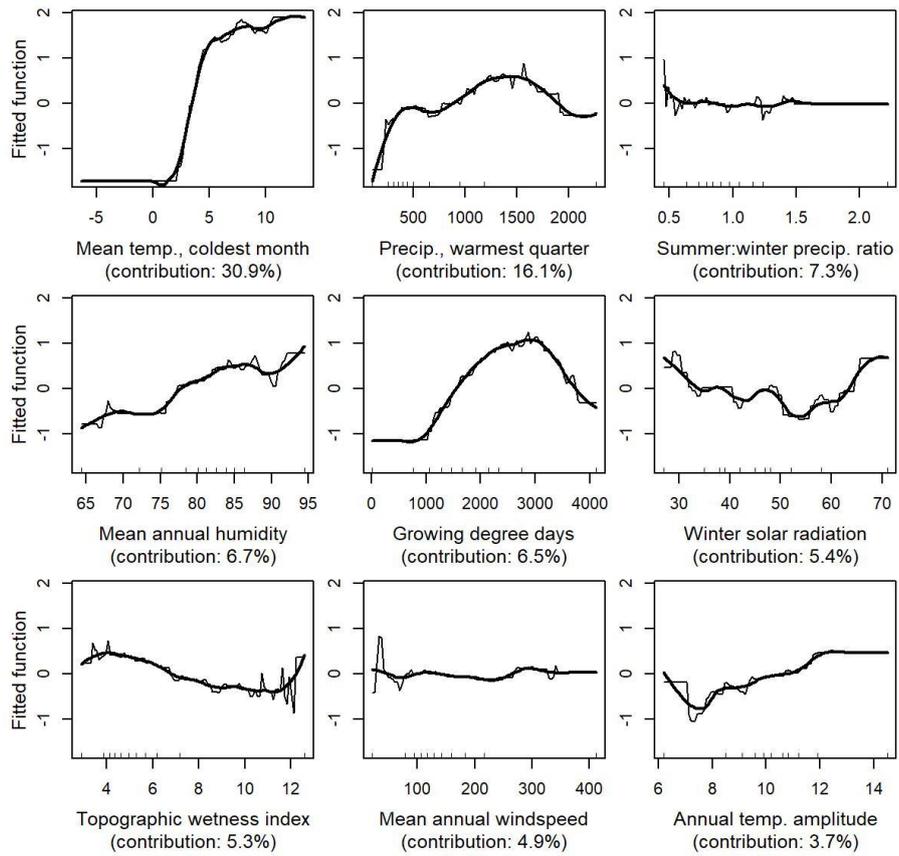
### Presence-only model



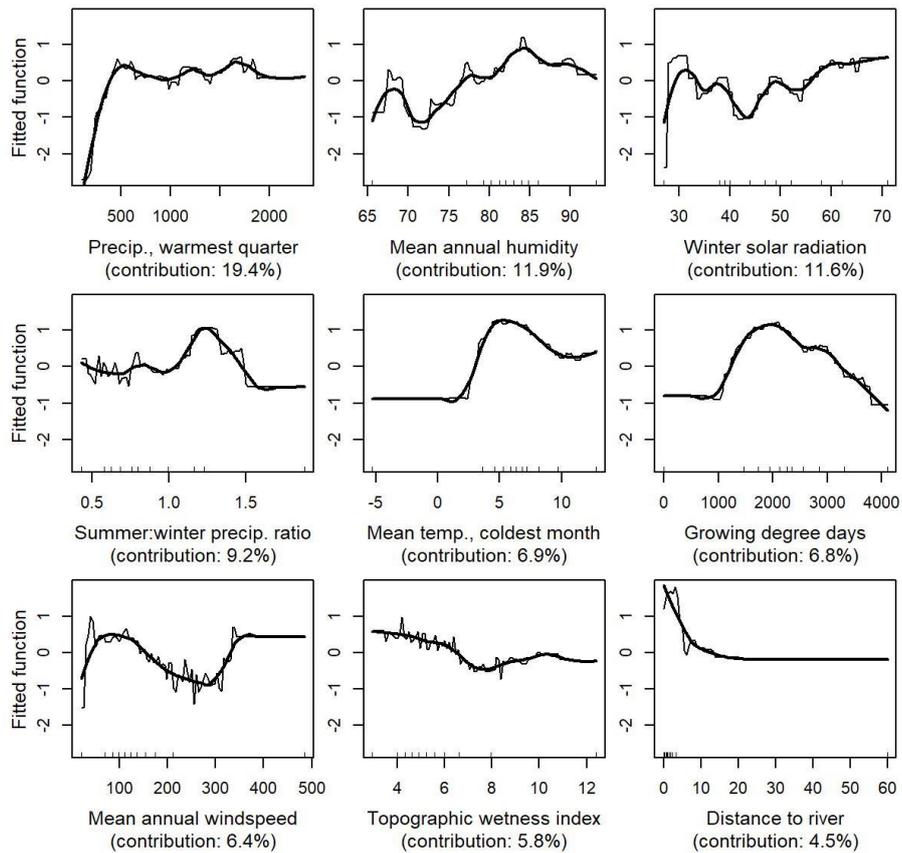
## 9.15 *Metrosideros diffusa*



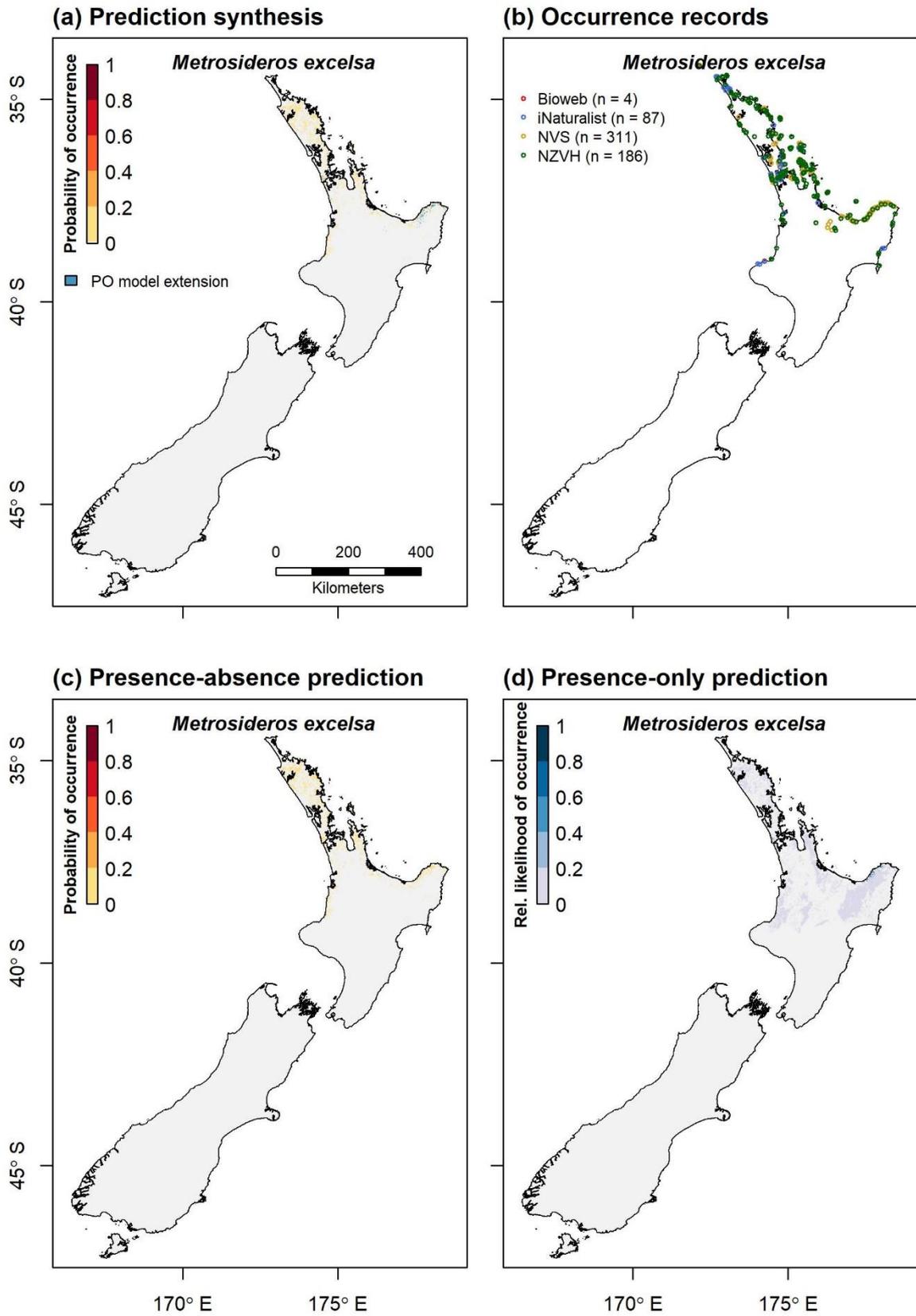
### Presence-absence model



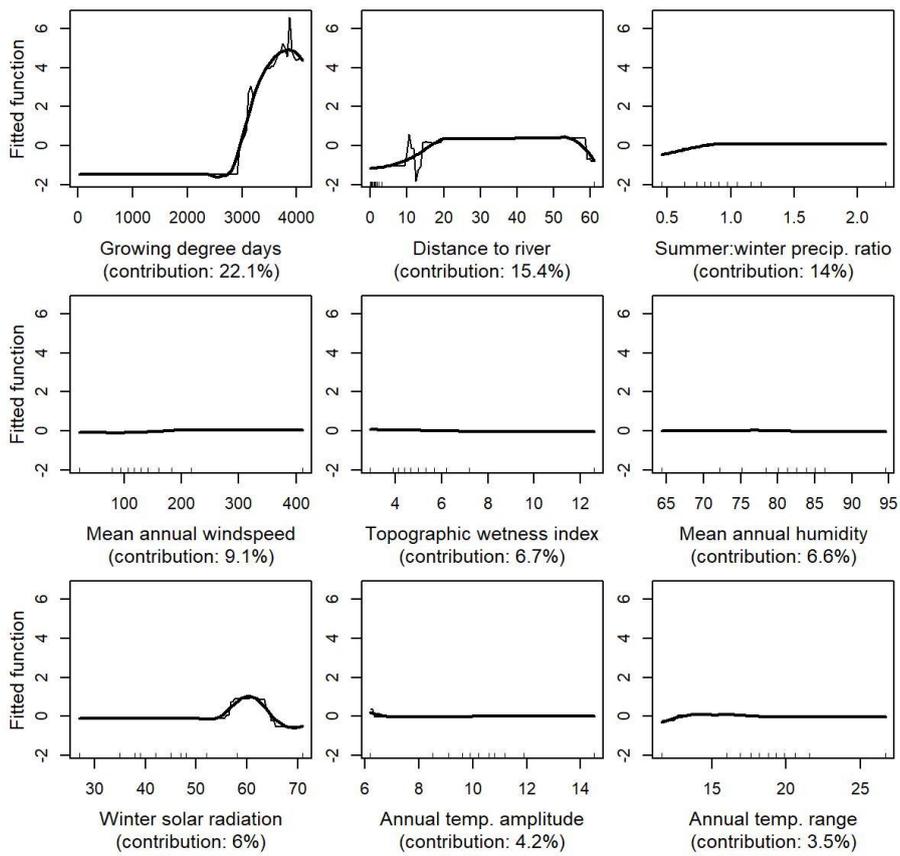
### Presence-only model



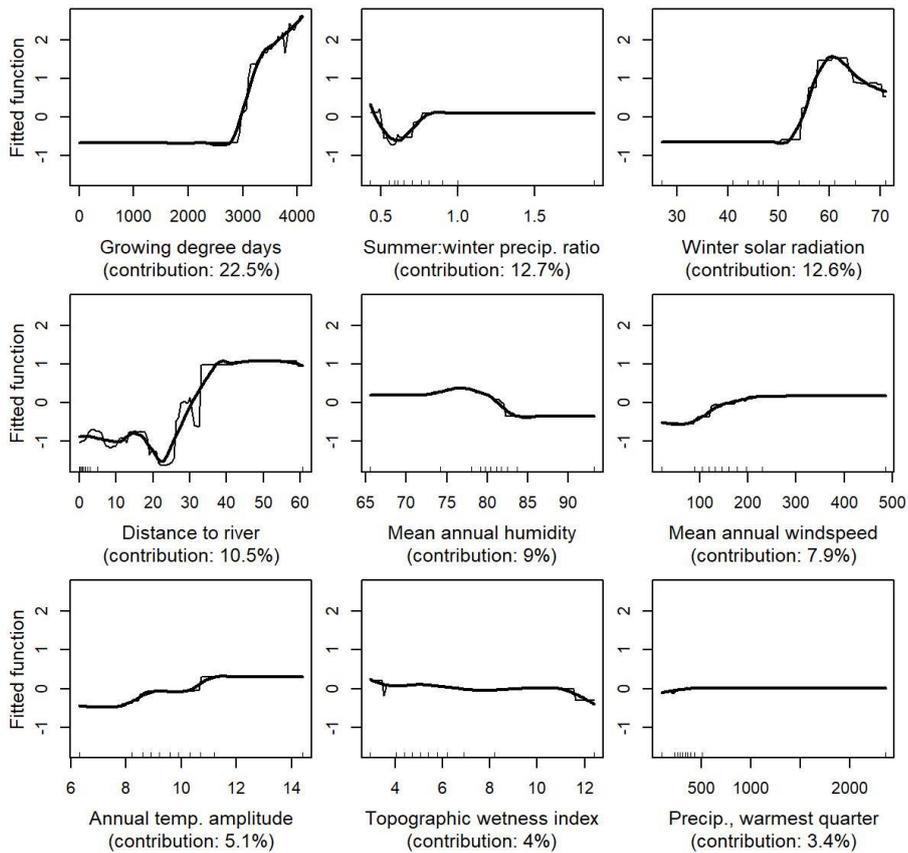
## 9.16 *Metrosideros excelsa*



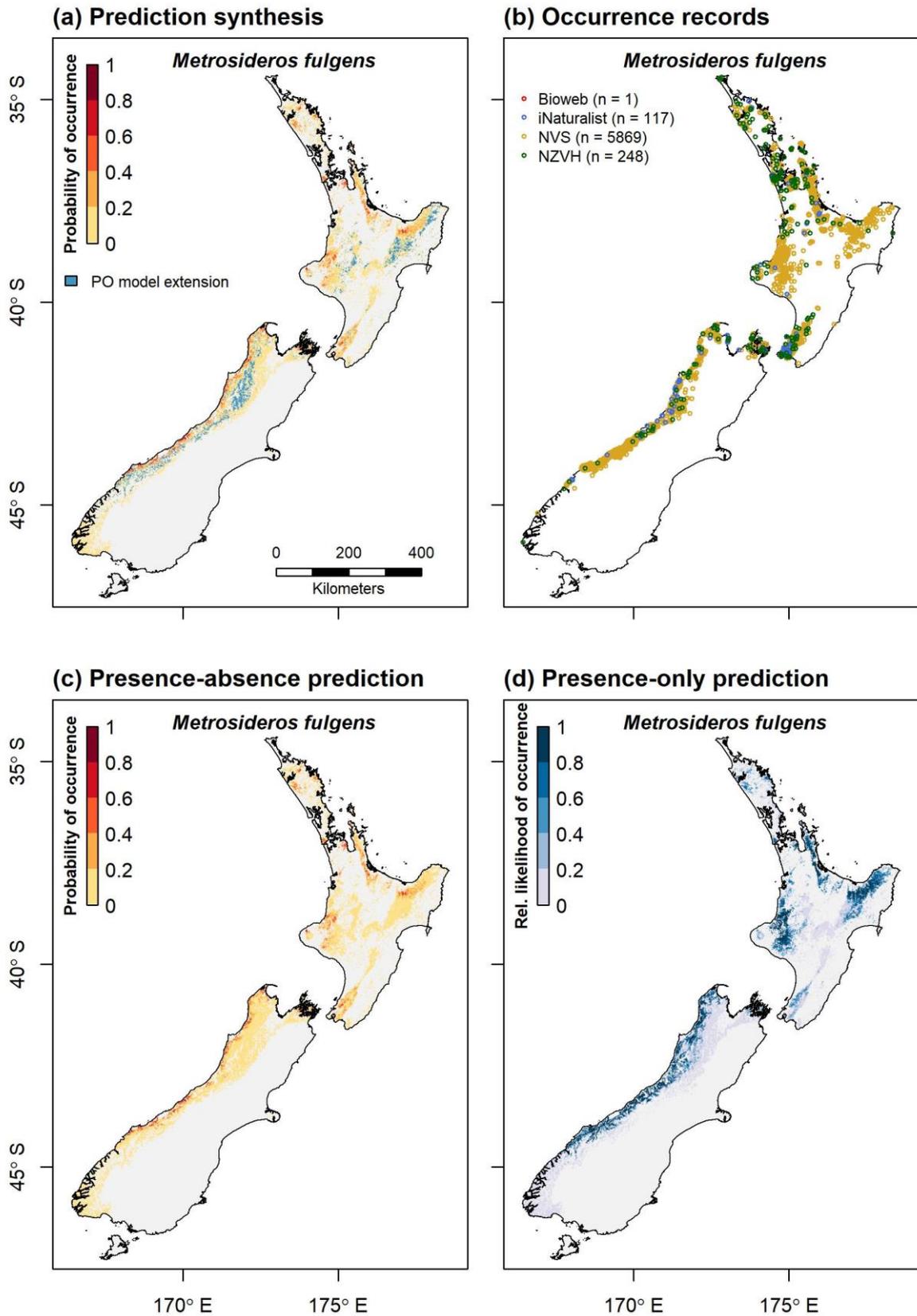
### Presence-absence model



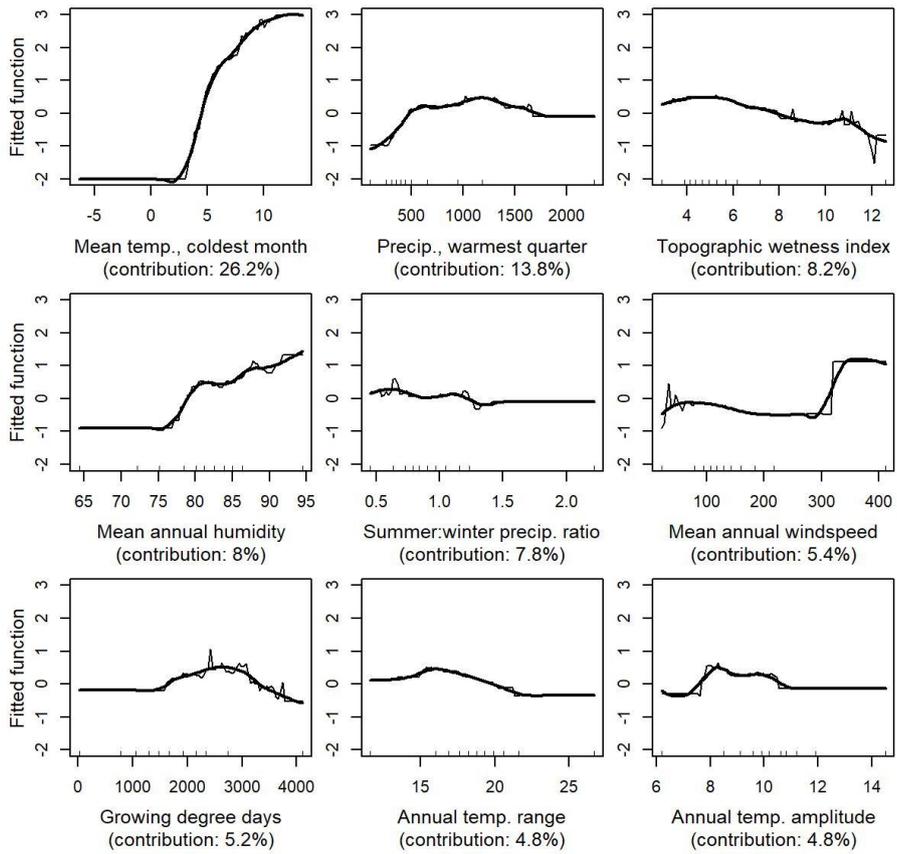
### Presence-only model



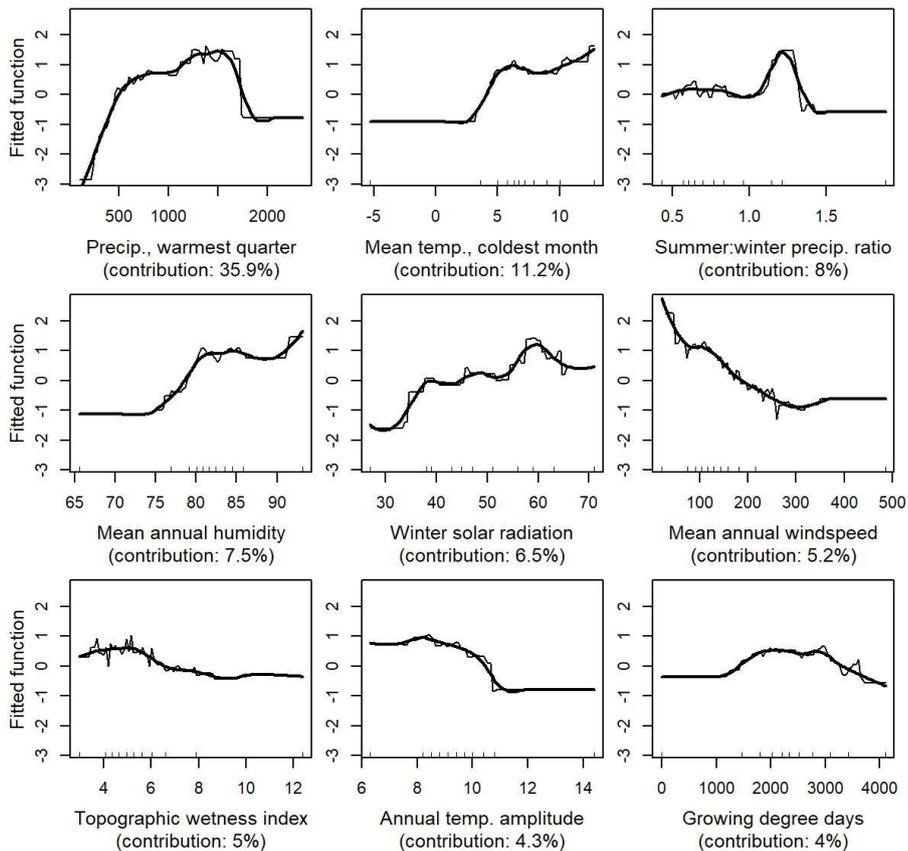
## 9.17 *Metrosideros fulgens*



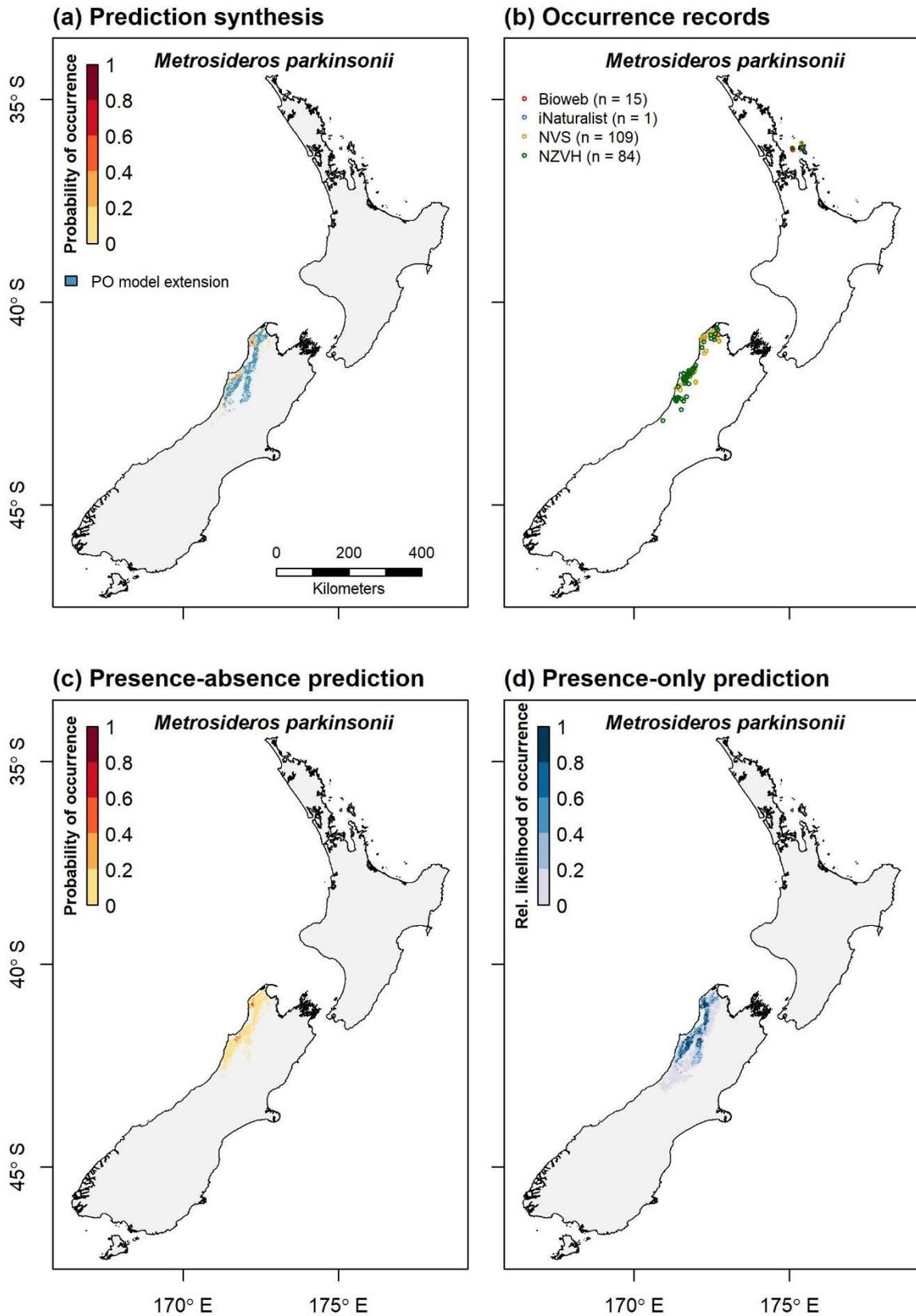
### Presence-absence model



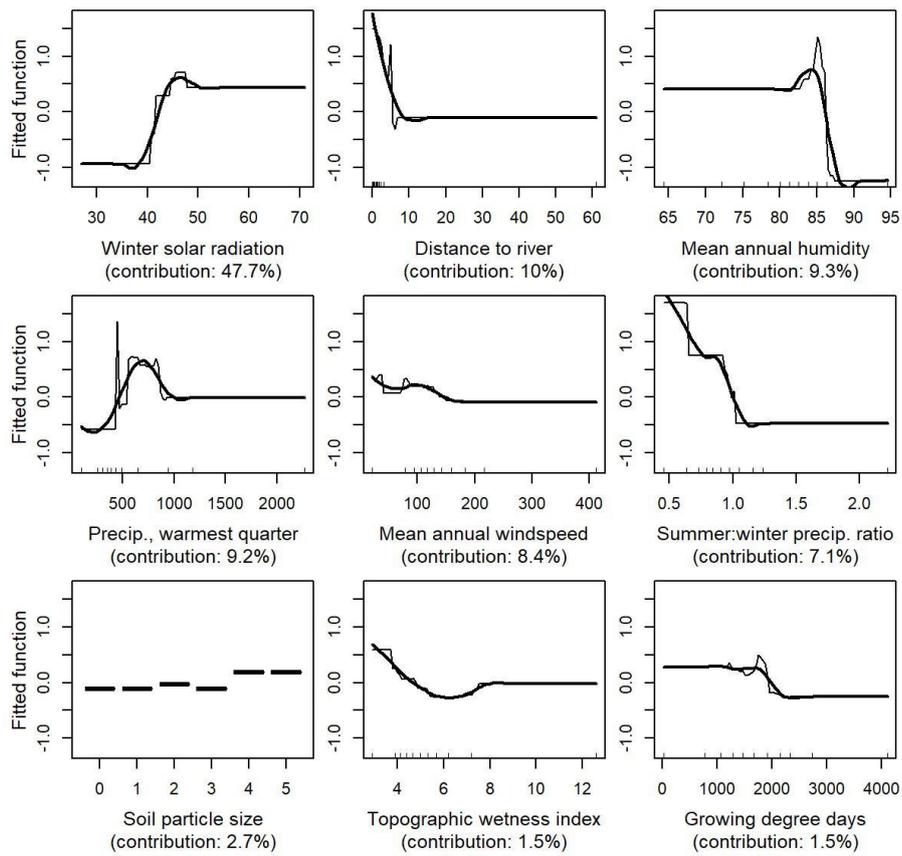
### Presence-only model



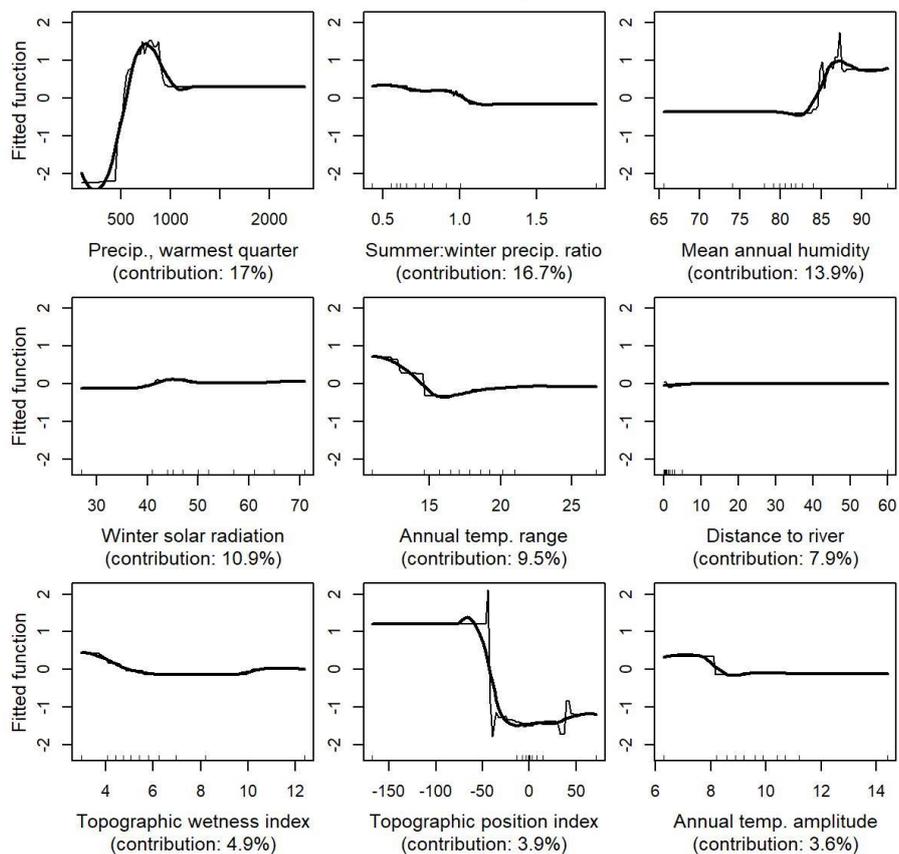
## 9.18 *Metrosideros parkinsonii*



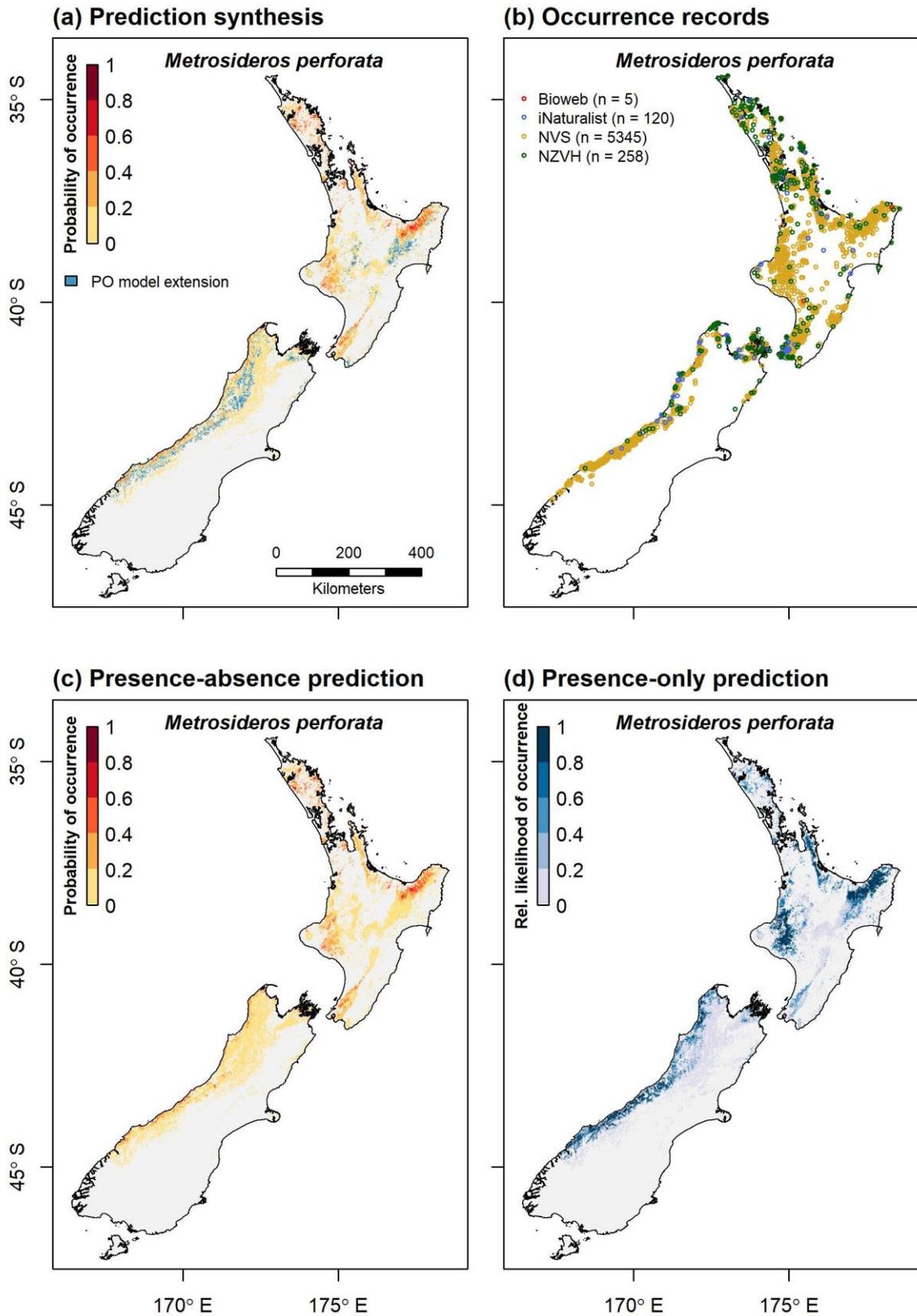
### Presence-absence model



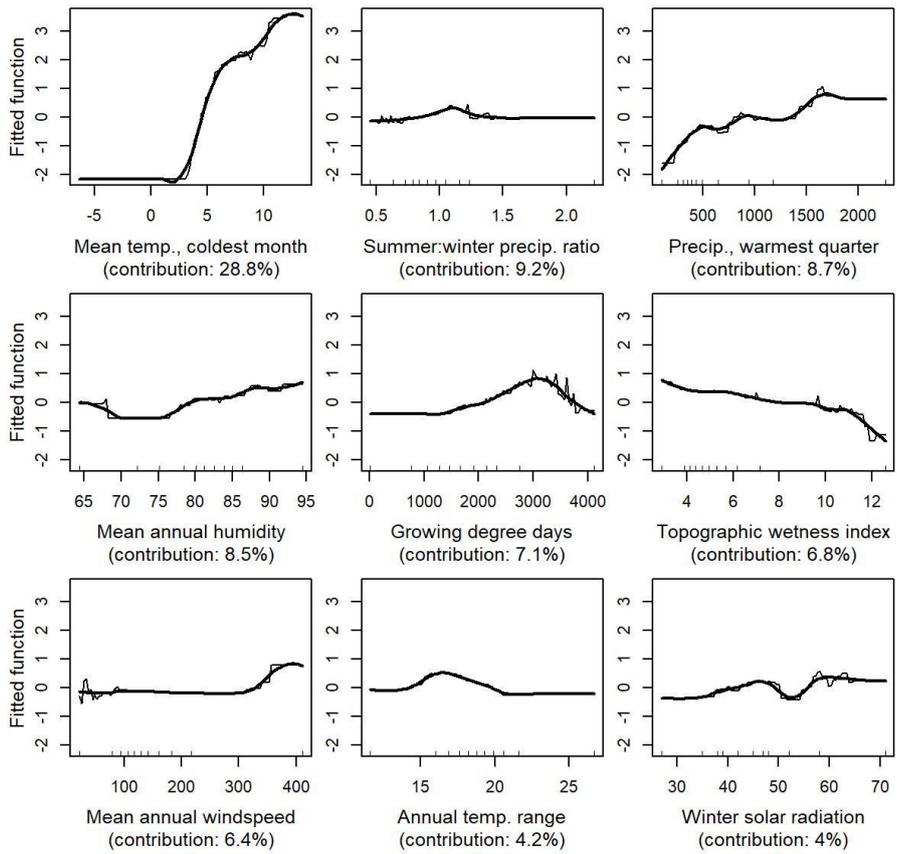
### Presence-only model



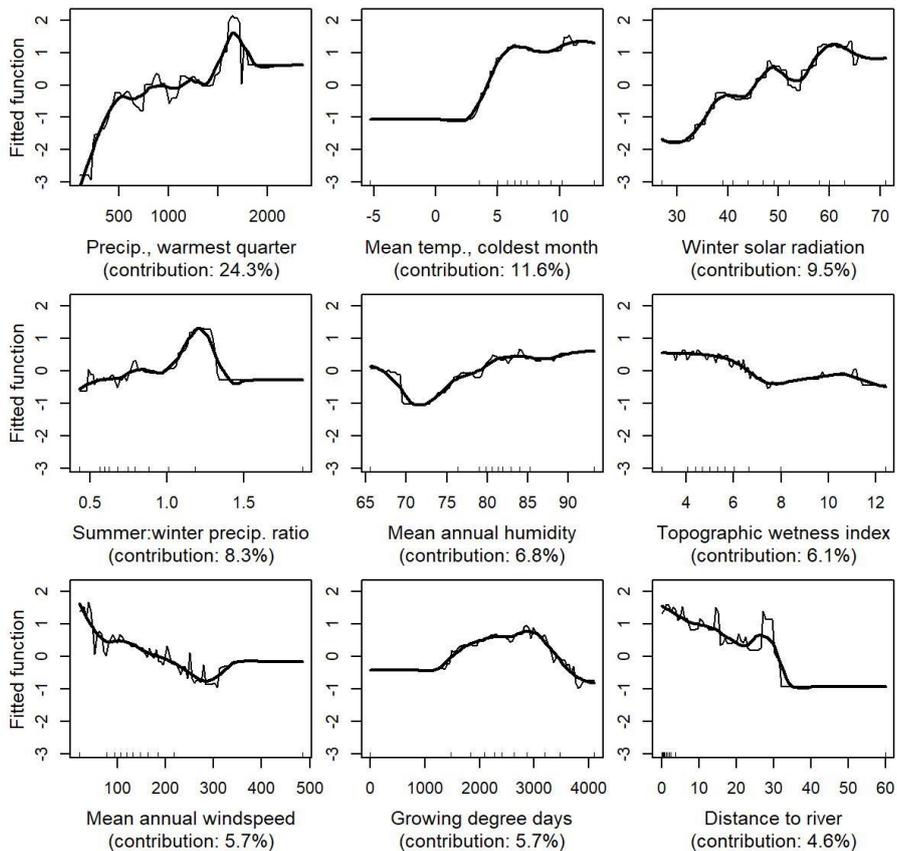
## 9.19 *Metrosideros perforata*



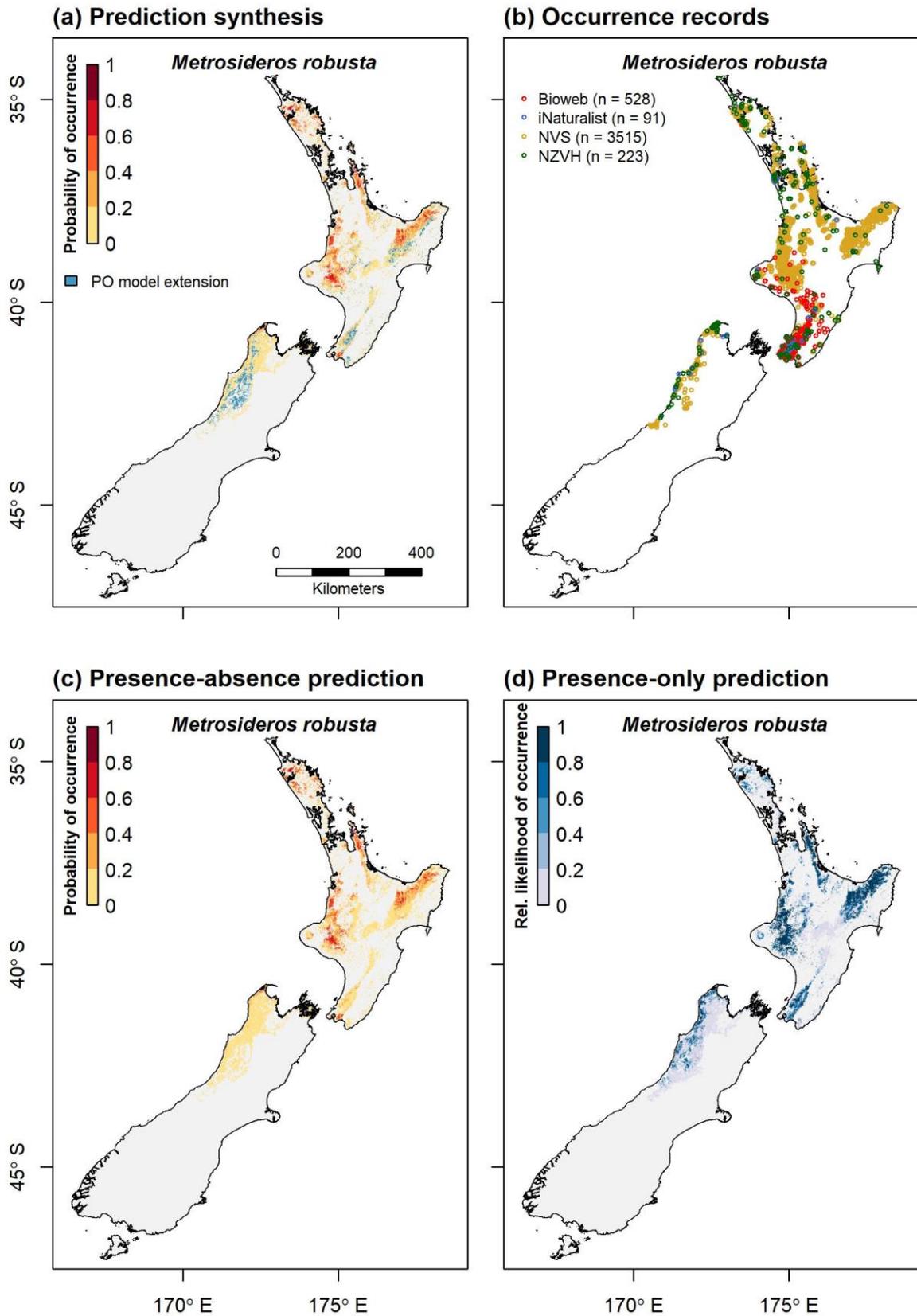
### Presence-absence model



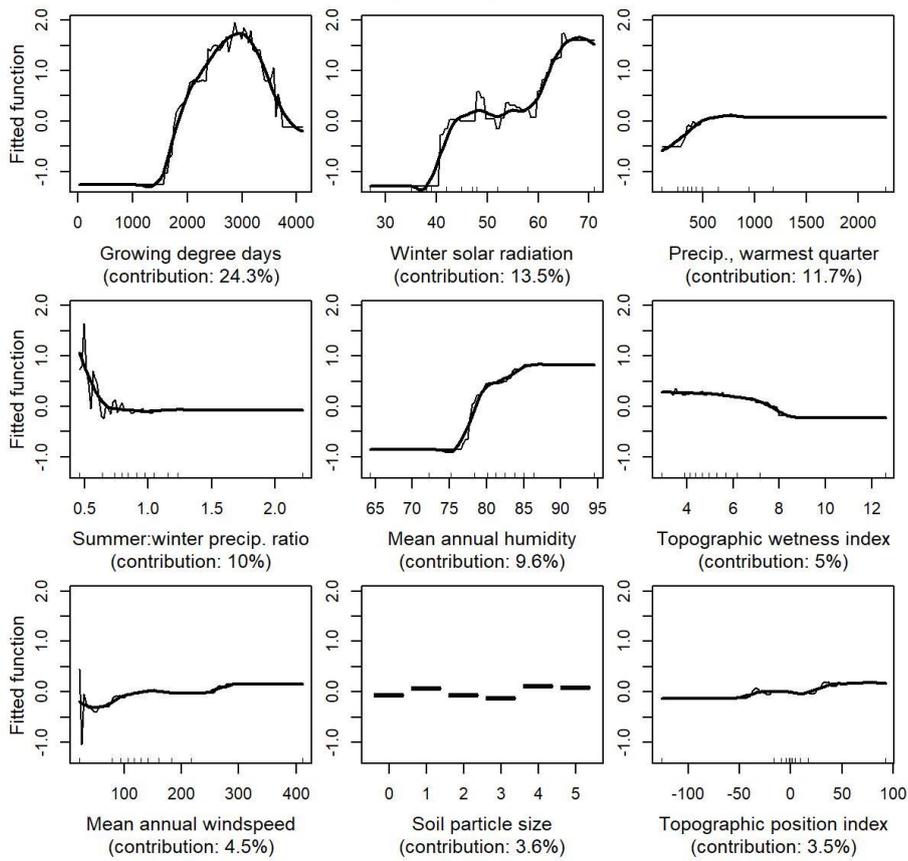
### Presence-only model



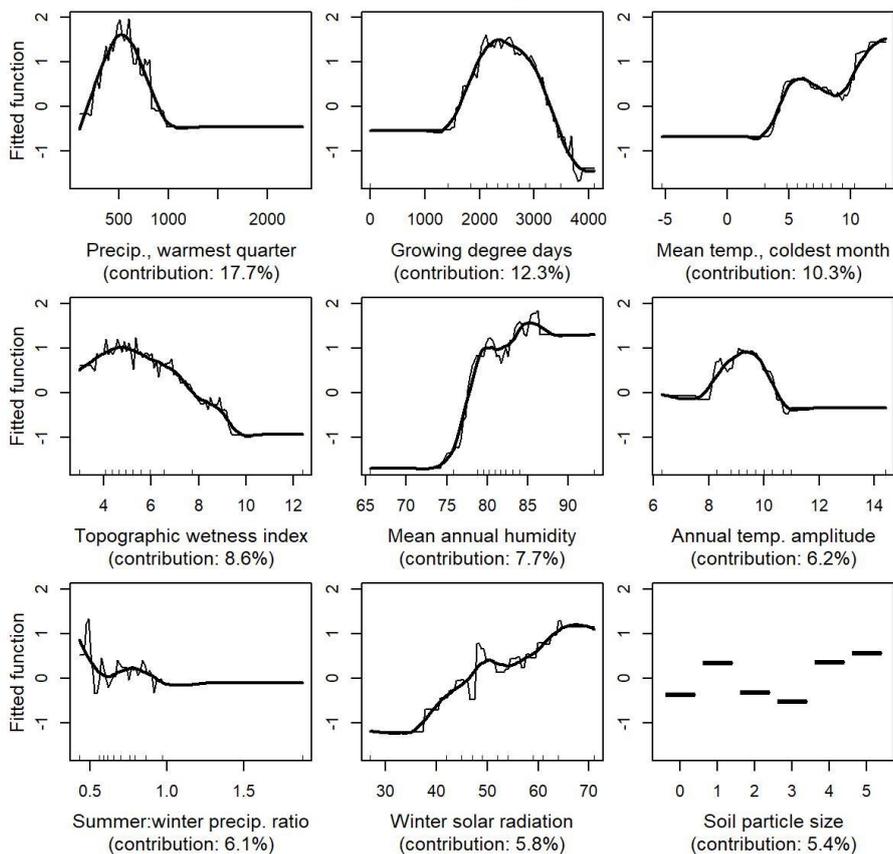
## 9.20 *Metrosideros robusta*



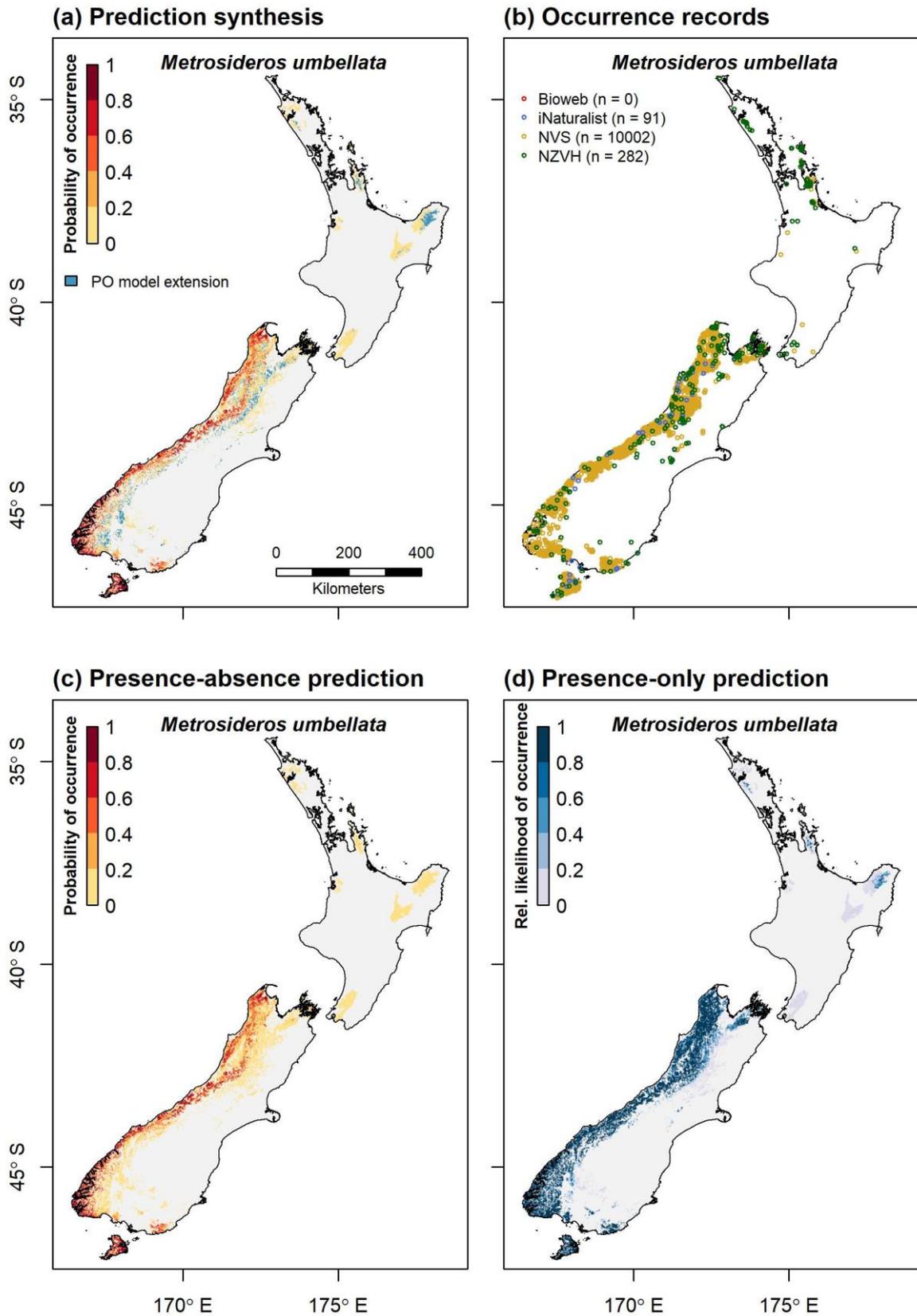
### Presence-absence model



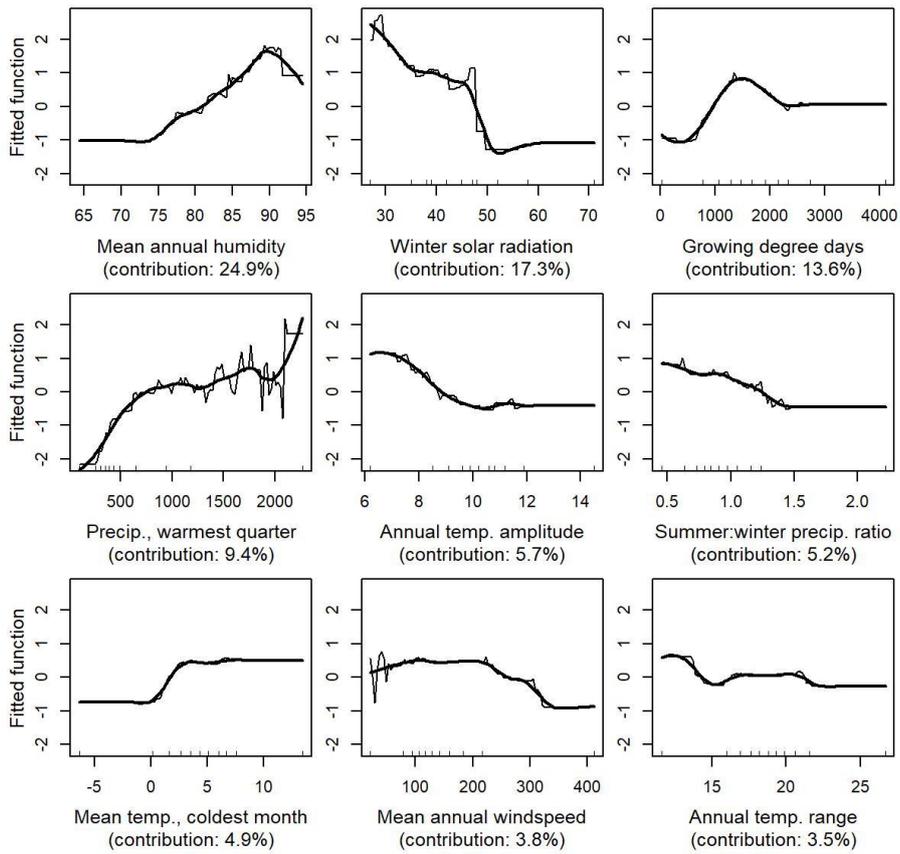
### Presence-only model



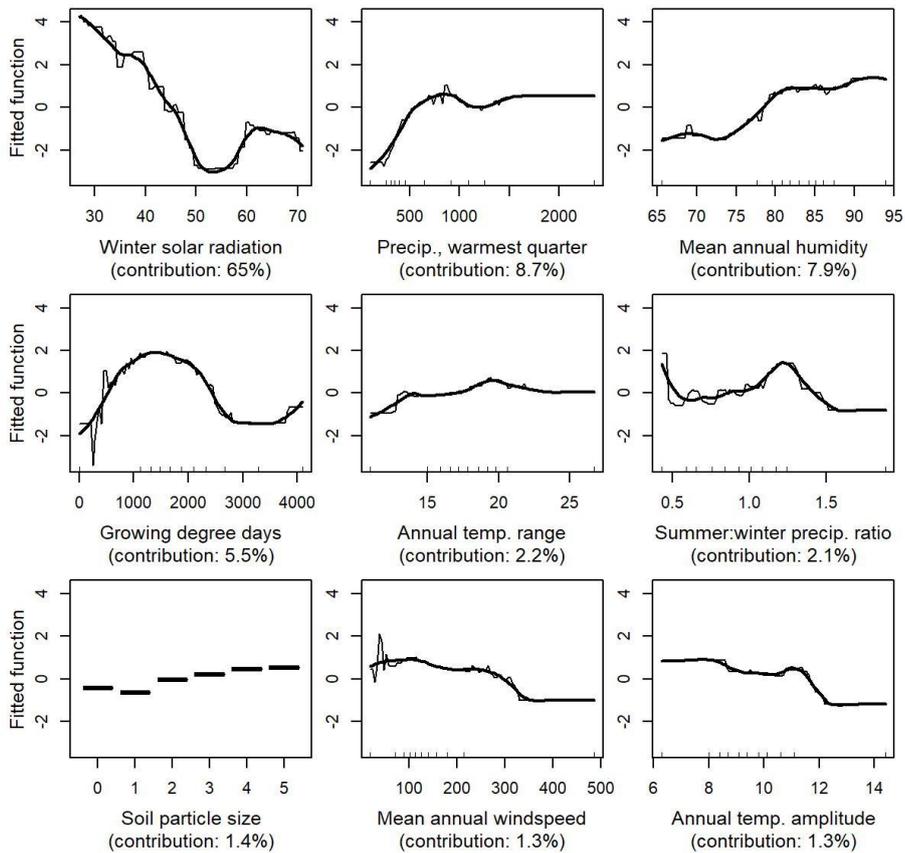
## 9.21 *Metrosideros umbellata*



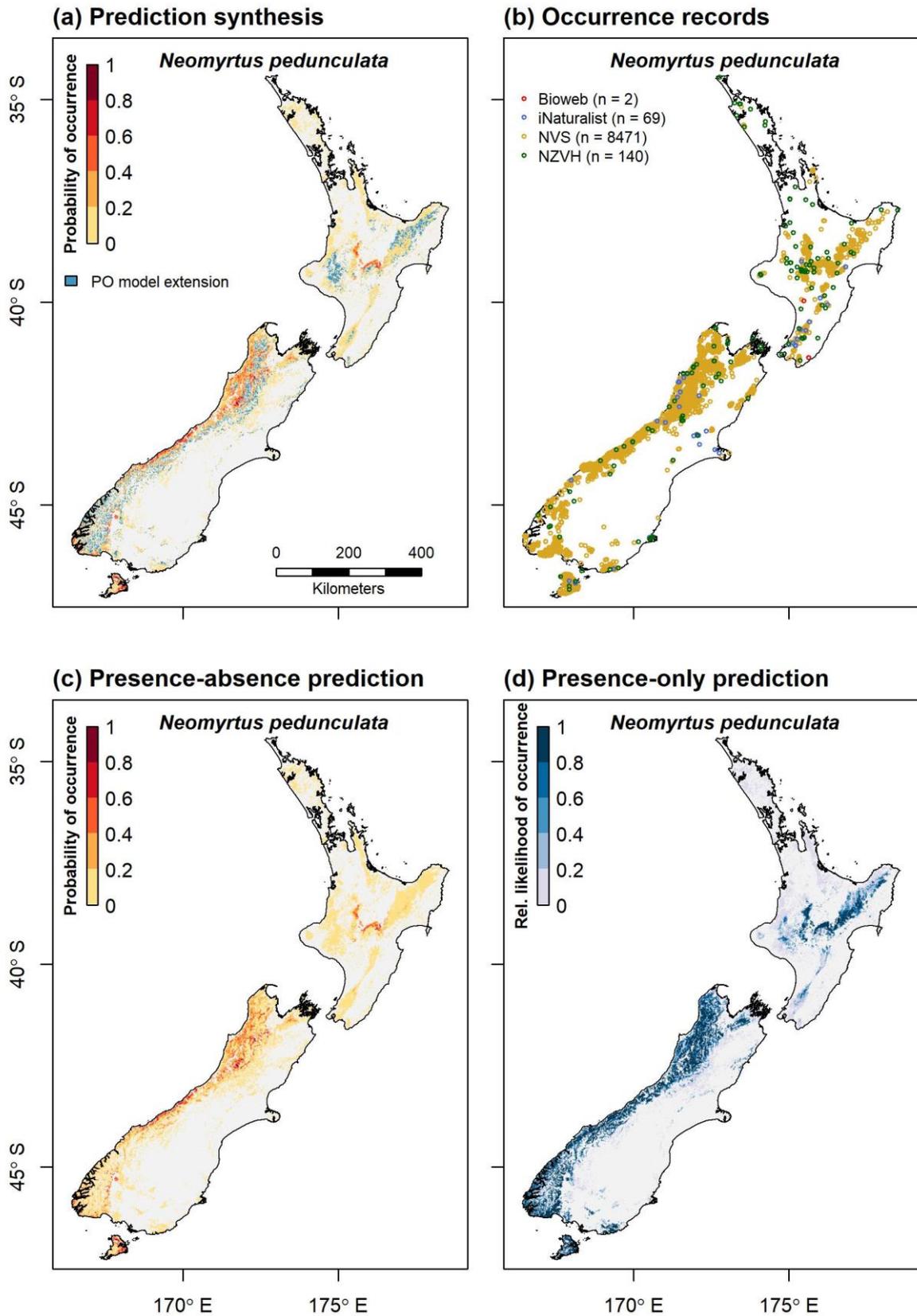
### Presence-absence model



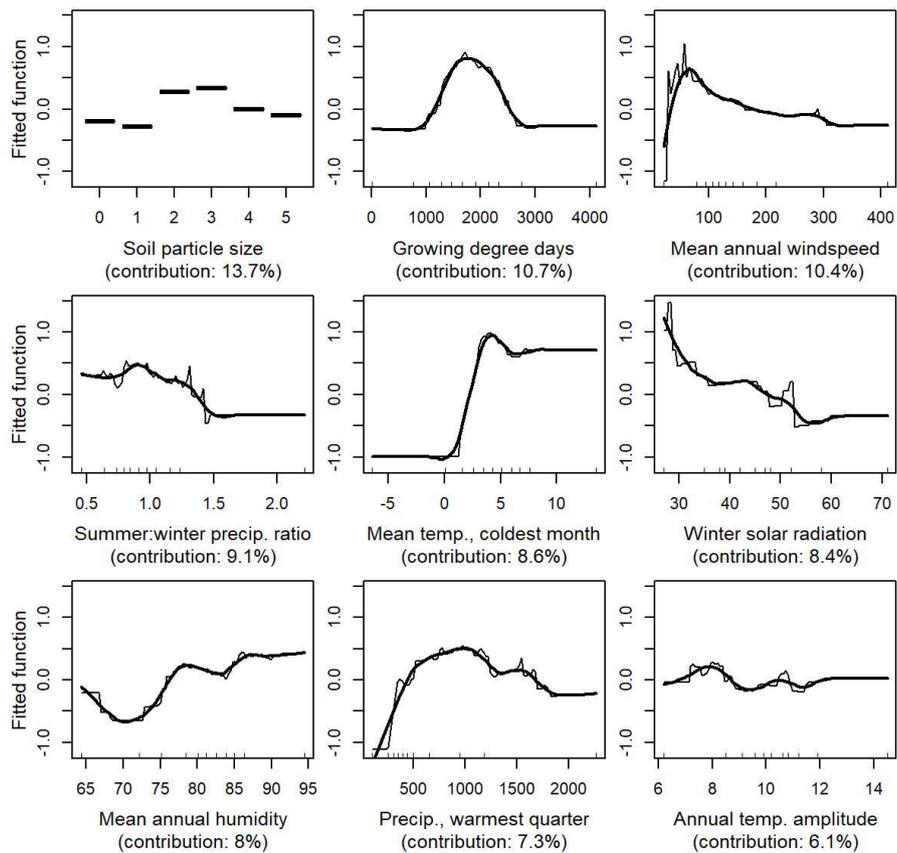
### Presence-only model



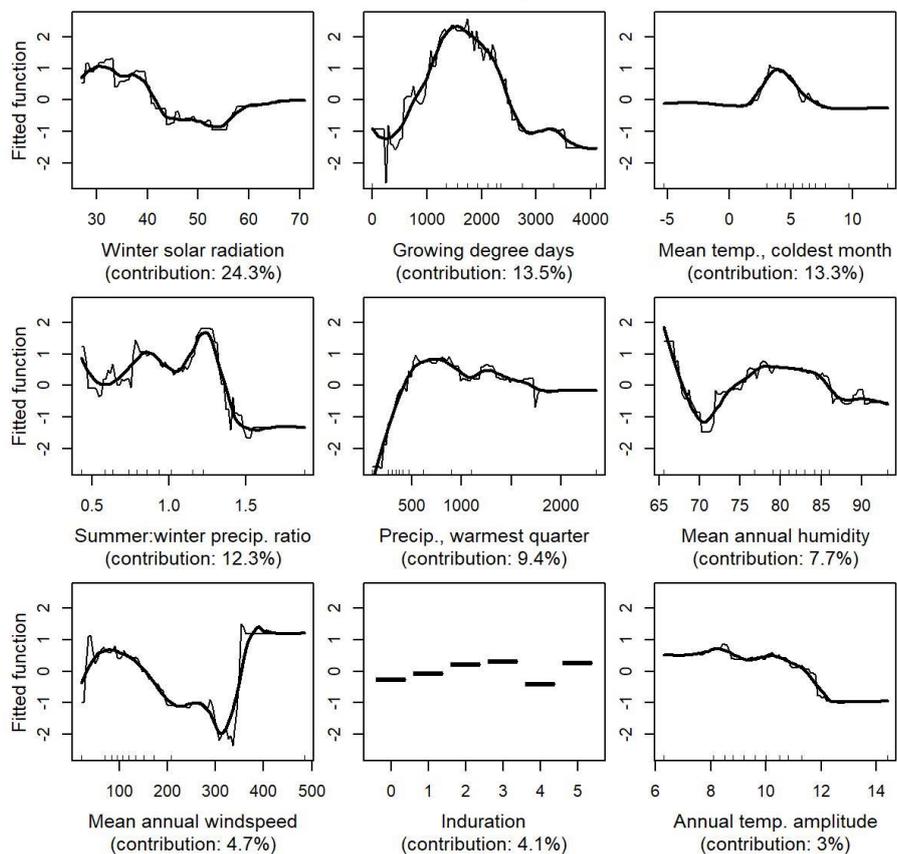
## 9.22 *Neomyrtus pedunculata*



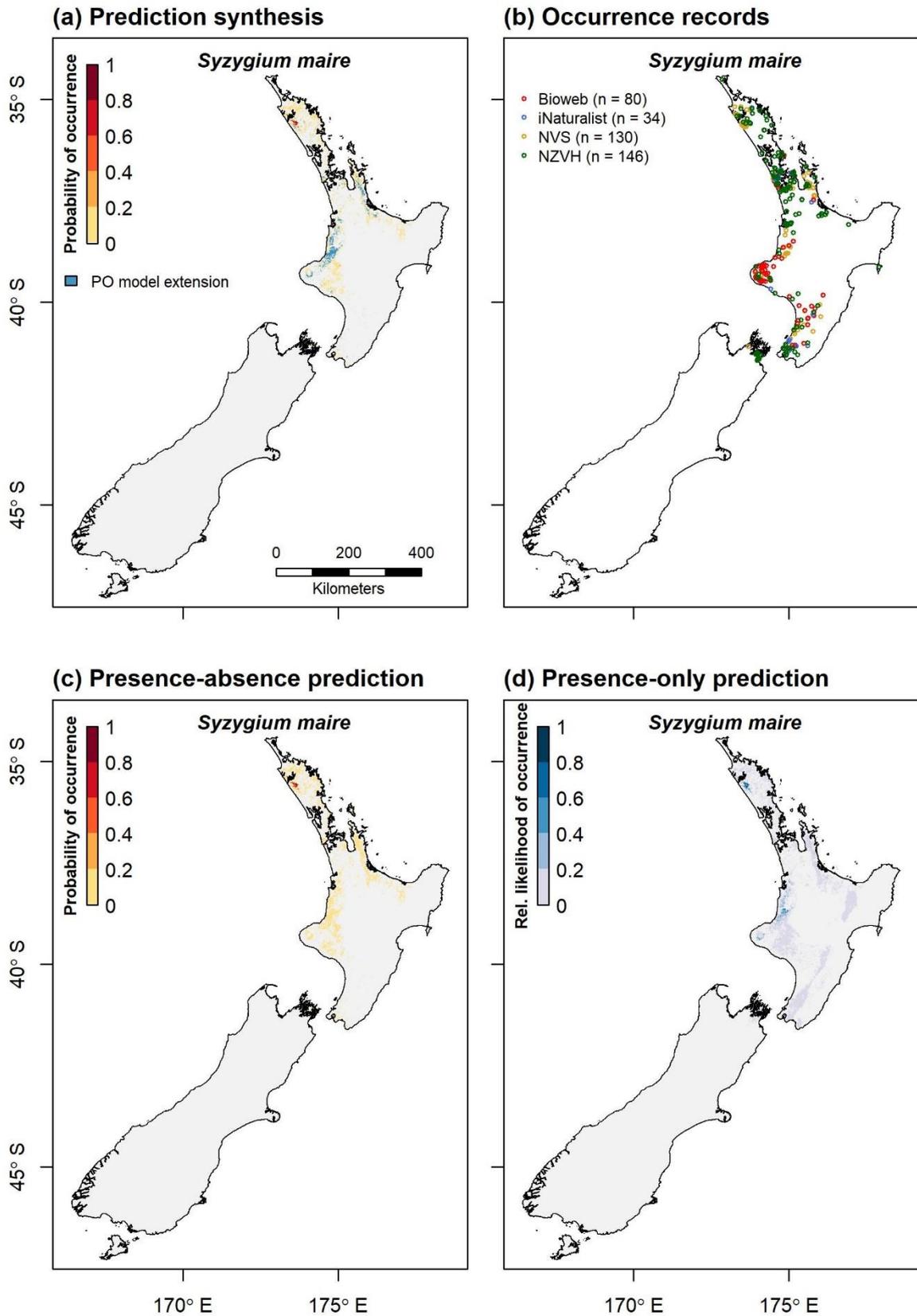
### Presence-absence model



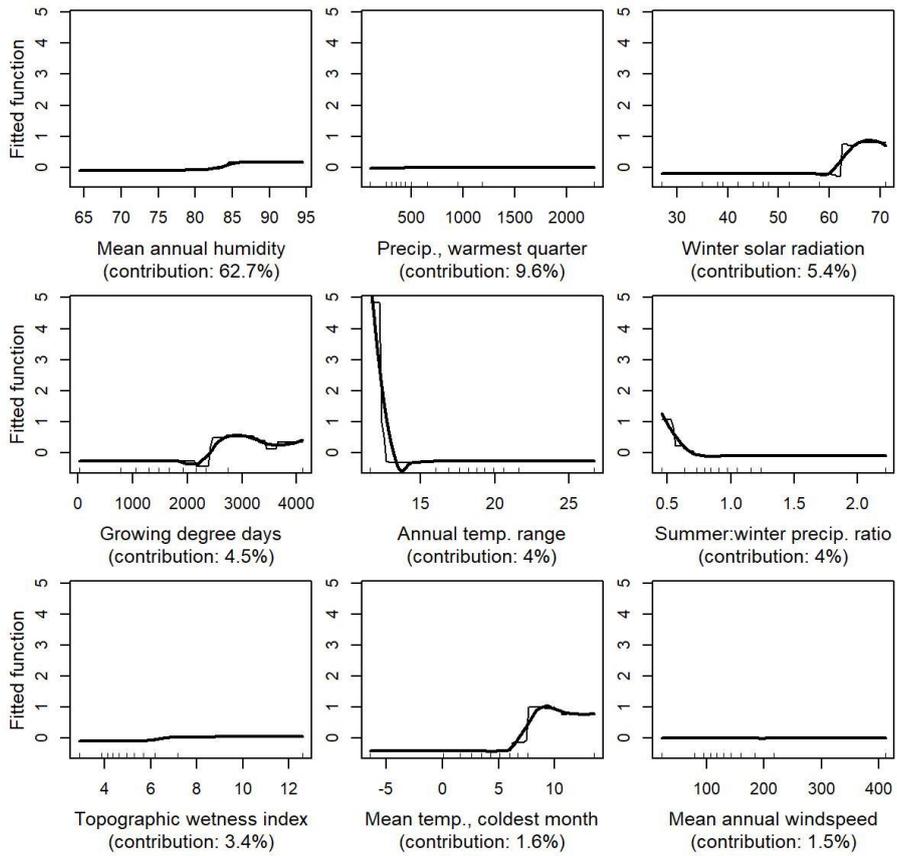
### Presence-only model



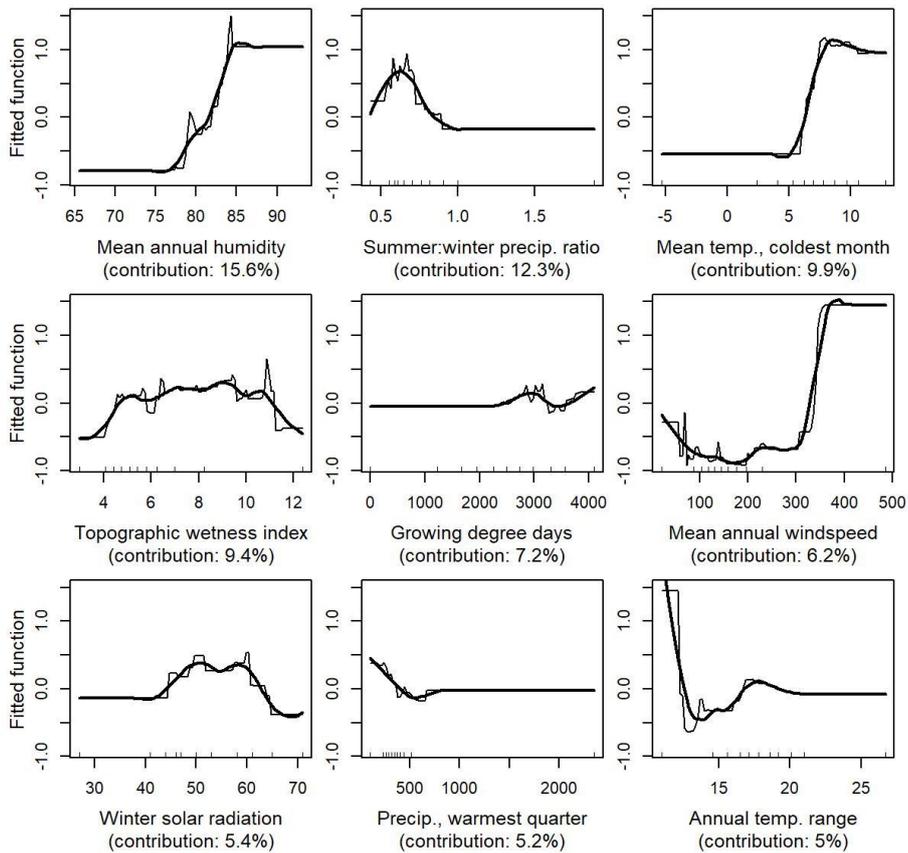
## 9.23 *Syzygium maire*



### Presence-absence model



### Presence-only model



## 10 Appendix E

### Relative importance of predictor variables

When using boosted regression trees, the relative importance, which sums to 100% for each model, can be calculated based on how often a predictor variable is selected across all trees, and the improvement to the overall model provided by the variable (Friedman 2001; Friedman & Meulman 2003; Elith et al. 2008). Here we provide the relative contribution of all 19 predictor variables to our presence–absence (Table E1) and presence-only models (Table E2).

**Table E1:** Relative contribution (%) of predictor variables to presence–absence boosted regression tree models. Decreasing shading intensity indicates the most important variables for each species.

Species	Acid soluble P	Annual temp. amplitude	Exchangeable calcium	Chem. lims plant growth	Growing deg. days	Soil induration	Mean temp. cold mth	Mean annual humidity	Mean ann. windspeed	Soil particle size	Win:sum precip. ratio	Precip. warmest qtr	Distance to river	Soil age	Soil drainage	Winter solar radiation	Ann. temp. range	Topo. position index	Topo. wetness index
<i>Kunzea ericoides s.l.</i>	1.6	3.5	0.7	0.0	26.8	0.8	5.7	11.6	4.0	1.4	7.2	13.1	3.9	0.4	2.3	5.8	4.0	2.7	4.5
<i>Leptospermum scoparium</i>	3.5	11.4	0.6	0.0	7.2	2.1	5.5	9.1	6.7	2.5	8.2	8.2	5.0	0.0	6.4	6.0	4.2	3.2	10.1
<i>Lophomyrtus bullata</i>	2.7	2.7	0.2	0.0	23.5	0.9	3.0	11.8	6.2	1.2	12.6	9.3	5.5	0.2	1.9	2.9	5.5	5.2	4.7
<i>Lophomyrtus obcordata</i>	2.1	5.0	0.6	0.0	10.0	2.0	4.1	10.1	7.4	3.3	14.2	10.8	4.3	0.0	1.6	7.1	7.6	3.4	6.3
<i>Metrosideros albiflora</i>	0.5	2.3	0.0	0.0	3.0	0.0	1.9	24.6	5.9	0.5	12.5	12.0	3.3	0.0	0.9	22.0	2.9	2.8	4.7
<i>Metrosideros colensoi</i>	1.8	8.4	3.1	0.0	6.9	7.6	4.6	14.3	3.3	6.8	14.5	10.2	4.6	0.0	1.2	6.9	2.9	1.3	1.7
<i>Metrosideros diffusa</i>	1.3	3.7	0.2	0.0	6.5	0.6	30.9	6.7	4.9	1.8	7.3	16.1	2.8	0.1	1.5	5.4	2.7	2.2	5.3
<i>Metrosideros excelsa</i>	1.0	4.2	0.6	0.0	22.1	0.8	2.1	6.6	9.1	0.4	14.0	3.4	15.4	0.3	0.8	6.0	3.5	2.9	6.7
<i>Metrosideros fulgens</i>	1.2	4.8	0.2	0.0	5.2	1.1	26.2	8.0	5.4	1.8	7.8	13.8	2.2	0.1	2.0	4.4	4.8	2.9	8.2
<i>Metrosideros parkinsonii</i>	0.0	0.3	0.0	0.0	1.5	0.4	0.4	9.3	8.4	2.7	7.1	9.2	10.0	0.0	0.4	47.7	0.6	0.6	1.5
<i>Metrosideros perforata</i>	1.0	3.9	0.5	0.0	7.1	1.0	28.8	8.5	6.4	1.7	9.2	8.7	3.4	0.1	1.3	4.0	4.2	3.4	6.8
<i>Metrosideros robusta</i>	2.2	2.5	0.0	0.0	24.3	0.2	3.0	9.6	4.5	3.6	10.0	11.7	2.9	0.0	0.7	13.5	2.7	3.5	5.0
<i>Metrosideros umbellata</i>	1.4	5.7	0.1	0.0	13.6	0.6	4.9	24.9	3.8	1.2	5.2	9.4	2.2	0.2	0.9	17.3	3.5	1.8	3.4
<i>Neomyrtus pedunculata</i>	0.9	6.1	0.2	0.0	10.7	1.7	8.6	8.0	10.4	13.7	9.1	7.3	3.0	0.2	1.4	8.4	3.1	2.1	5.2
<i>Syzygium maire</i>	0.5	0.6	0.1	0.0	4.5	0.1	1.6	62.7	1.5	0.1	4.0	9.6	0.7	0.0	0.5	5.4	4.0	0.9	3.4
Mean	1.5	4.3	0.5	0.0	11.5	1.3	8.7	15.0	5.9	2.8	9.5	10.2	4.6	0.1	1.6	10.9	3.7	2.6	5.2

**Table E2:** Relative contribution (%) of predictor variables to presence-only boosted regression tree models. Decreasing shading intensity indicates the most important variables for each species.

Species	Acid soluble P	Annual temp. amplitude	Exchangeable calcium	Chem. lims plant growth	Growing deg. days	Soil induration	Mean temp. cold mth	Mean annual humidity	Mean ann. windspeed	Soil particle size	Win:sum precip. ratio	Precip. warmest qtr	Distance to river	Soil age	Soil drainage	Winter solar radiation	Ann. temp. range	Topo. position index	Topo. wetness index
<i>Kunzea amathicola</i>	1.0	4.0	0.0	0.0	8.8	0.0	8.2	16.5	6.4	7.1	3.0	6.4	14.8	1.0	0.0	4.5	12.1	2.4	3.9
<i>Kunzea ericoides s.l.</i>	2.6	4.3	1.1	0.0	6.7	2.0	4.1	15.8	6.7	1.9	9.4	8.7	6.0	0.0	1.5	8.2	8.5	4.2	8.2
<i>Kunzea ericoides s.s.</i>	1.6	1.3	6.5	0.0	0.5	0.0	0.0	1.9	33.9	4.6	1.8	7.5	3.1	0.0	0.0	3.8	19.3	6.9	7.4
<i>Kunzea linearis</i>	3.5	0.0	0.0	0.0	30.7	0.0	9.9	3.9	8.2	3.2	8.0	5.2	5.1	0.0	5.7	2.2	3.6	4.0	6.8
<i>Kunzea robusta</i>	3.0	2.9	0.8	0.3	8.6	2.1	4.5	15.2	6.8	3.0	10.9	11.5	6.6	0.0	4.9	4.3	4.2	4.4	6.0
<i>Kunzea serotina</i>	2.0	11.9	0.0	0.7	7.7	1.1	4.9	5.1	13.1	1.3	15.9	5.9	4.6	0.0	2.0	8.5	8.5	2.8	4.2
<i>Kunzea sinclairii</i>	0.0	14.3	0.0	0.0	0.8	0.0	3.0	0.2	2.0	0.4	43.1	0.0	0.7	0.0	0.0	30.7	2.0	1.6	1.1
<i>Kunzea tenuicaulis</i>	0.0	6.6	0.0	0.0	4.5	9.5	7.1	1.3	7.1	2.8	18.6	4.7	2.1	0.0	2.2	14.6	11.9	1.4	5.8
<i>Leptospermum scoparium</i>	2.4	6.0	1.2	0.0	7.0	1.8	3.6	7.6	7.8	3.2	10.7	7.8	5.4	0.0	4.0	17.6	4.7	3.4	5.9
<i>Lophomyrtus bullata</i>	1.9	4.4	0.5	0.0	16.6	1.2	8.3	7.7	6.8	5.1	9.3	8.0	5.5	0.0	1.3	10.2	4.3	4.4	4.4
<i>Lophomyrtus obcordata</i>	4.9	4.7	0.9	0.0	5.0	3.2	6.2	11.9	9.5	3.5	9.6	10.7	3.3	0.6	2.8	8.9	7.5	3.1	3.7
<i>Metrosideros albiflora</i>	0.5	2.6	0.3	0.0	25.2	0.0	3.4	11.4	3.0	1.7	7.9	7.7	2.4	0.0	1.0	22.9	4.0	3.0	2.9
<i>Metrosideros carminea</i>	1.7	2.6	0.3	0.0	19.9	0.7	4.5	10.4	4.1	2.4	10.5	4.5	8.1	0.0	1.2	15.9	2.8	4.8	5.8
<i>Metrosideros colensoi</i>	2.3	7.1	5.0	0.0	9.1	3.2	6.7	11.0	4.0	7.5	8.8	9.1	3.2	0.0	0.0	13.5	3.5	2.4	3.7
<i>Metrosideros diffusa</i>	1.3	3.9	0.0	0.0	6.8	1.0	6.9	11.9	6.4	2.5	9.2	19.4	4.5	0.0	1.4	11.6	3.3	4.1	5.8
<i>Metrosideros excelsa</i>	0.8	5.1	0.0	0.0	22.5	0.5	2.3	9.0	7.9	0.8	12.7	3.4	10.5	0.4	1.7	12.6	3.4	2.5	4.0

<b>Species</b>	<b>Acid soluble P</b>	<b>Annual temp. amplitude</b>	<b>Exchangeable calcium</b>	<b>Chem. lims plant growth</b>	<b>Growing deg. days</b>	<b>Soil induration</b>	<b>Mean temp. cold mth</b>	<b>Mean annual humidity</b>	<b>Mean ann. windspeed</b>	<b>Soil particle size</b>	<b>Win:sum precip. ratio</b>	<b>Precip. warmest qtr</b>	<b>Distance to river</b>	<b>Soil age</b>	<b>Soil drainage</b>	<b>Winter solar radiation</b>	<b>Ann. temp. range</b>	<b>Topo. position index</b>	<b>Topo. wetness index</b>
<i>Metrosideros fulgens</i>	0.8	4.3	0.4	0.0	4.0	0.6	11.2	7.5	5.2	2.4	8.0	35.9	2.7	0.0	1.1	6.5	1.9	2.3	5.0
<i>Metrosideros parkinsonii</i>	2.1	3.6	0.6	0.0	2.3	1.5	1.4	13.9	2.4	1.5	16.7	17.0	7.9	0.0	0.0	10.9	9.5	3.9	4.9
<i>Metrosideros perforata</i>	1.0	4.3	0.5	0.0	5.7	1.1	11.6	6.8	5.7	2.7	8.3	24.3	4.6	0.0	1.7	9.5	3.1	2.9	6.1
<i>Metrosideros robusta</i>	2.5	6.2	0.0	0.0	12.3	0.7	10.3	7.7	5.2	5.4	6.1	17.7	4.9	0.0	1.0	5.8	2.6	2.9	8.6
<i>Metrosideros umbellata</i>	0.5	1.3	0.3	0.0	5.5	0.2	1.3	7.9	1.3	1.4	2.1	8.7	0.8	0.0	0.3	65.0	2.2	0.4	1.0
<i>Neomyrtus pedunculata</i>	0.9	3.0	0.4	0.0	13.5	4.1	13.3	7.7	4.7	0.6	12.3	9.4	1.1	0.1	0.5	24.3	1.4	0.7	1.9
<i>Syzygium maire</i>	5.0	4.8	0.5	0.0	7.2	1.8	9.9	15.6	6.2	2.9	12.3	5.2	3.7	0.0	1.7	5.4	5.0	3.3	9.4
Mean	1.8	4.7	0.8	0.0	10.0	1.6	6.2	9.0	7.1	3.0	11.1	10.4	4.9	0.1	1.6	13.8	5.6	3.1	5.1

