Accounting for the indirect area effect in stacked species distribution models to map species richness in a montane biodiversity hotspot

Robin Pouteau1,2*, Élise Bayle2,3, Élodie Blanchard1, Philippe Birnbaum1,4, Jean-Jérome Cassan3, Vanessa Hequet5, Thomas Ibanez1 and Hervé Vandrot1

1Institut Agronomique néo-Calédonien (IAC), Centre IRD de Nouméa, Noumea, New Caledonia, 2Institut de Recherche pour le Développement (IRD), Conservatoire des Espaces Naturels, Koolnhé (Koné), New Caledonia, 3Direction du Développement Économique et de l’Environnement (DDEE), Koolnhé (Koné), Province Nord, New Caledonia, 4Centre de Cooperation Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France, 5Institut de Recherche pour le Développement (IRD), Centre IRD de Nouméa, Noumea, New Caledonia

ABSTRACT

Aim Understanding how species richness is distributed is a critical prerequisite to implement efficient conservation strategies in biodiversity hotspots. Stacked species distribution models (S-SDMs) provide new opportunities to map species richness, but the accuracy of this method may decline with elevation. Here, we test whether variation in model accuracy arises from increasingly unpredictable environments or the decreasing availability of area with increasing elevation, which might affect the regional pool of species and, as a result, the local species richness (the indirect area effect).

Location The New Caledonian biodiversity hotspot (south-west Pacific Ocean).

Methods An individual MAXENT model was built for 562 tree species by combining eight 100-m resolution environmental variables with c. 10,000 occurrence records. For each species, a map was produced at a one-hectare scale indicating the estimated habitat suitability. All models were then summed, and the resulting estimates were compared with richness measured in 11 independent one-hectare inventories. To account for the indirect area effect, S-SDM estimates were adjusted through the Arrhenius’ equation linking the number of species hosted by a habitat with its surface area.

Results The S-SDM predictions ranged from 95 to 251 species (mean = 153) while field inventories were lower, ranging from 39 to 131 species (mean = 90). Overall, the S-SDM increasingly overestimated richness as elevation increased. Taking into account the indirect area effect de-correlated residuals from elevation and induced a significant correlation between modelled and measured species richness.

Conclusion The decreasing accuracy of the S-SDM with elevation was explained by the decreasing availability of habitat influencing regional diversity found in each elevational band. Despite remaining difficulties to predict species richness when addressing the indirect area effect, our findings represent a significant step forward towards improved S-SDM designing.

Keywords Alpha diversity, biodiversity mapping, gamma diversity, herbarium data, mountain ecology, tropical rain forest.

INTRODUCTION

Pragmatic conservation strategies are urgently needed to slow the global loss of biodiversity. Biodiversity and biodiversity losses are not evenly distributed, and the concept of biodiversity hotspots has been valuable for prioritizing regions of primary conservation concern (Myers et al., 2000). However, this concept might be of limited practical value for...
environmental managers working within hotspots who instead require guidelines to delineate landscape units that require conservation actions (Murray-Smith et al., 2009; Canadas et al., 2014). Thus, a key step to utilize the hotspot concept at the stakeholders’ scale must be to assess landscape-level variations of biodiversity, including species richness.

Species inventories are so expensive and time-consuming that the implementation of plot networks capturing the full variation of species richness is often unrealistic. As an alternative, macro-ecological models have been designed to statistically relate observed species richness in plots (α-diversity) with spatially explicit environmental variables (Ferrier & Guisan, 2006). These models have been used to map vascular plant species richness in a number of biodiversity hotspots such as in the Himalayas (Bhattarai & Vetaas, 2003), within the Trans-Mexican Volcanic Belt (Mesoamerica hotspot) (Sánchez-González & López-Mata, 2005) and in New Zealand grasslands (Tomasetto et al., 2013). Major weaknesses of these methods include the need for many plots and their inability to extrapolate beyond known communities as they may not accurately model the drivers of underlying species distributions (Ferrier & Guisan, 2006).

Another method for mapping species richness uses herbarium and museum records to assemble collection data within grid cells (point-to-grid maps). This method was used to map the distribution of palm species in Thailand (Indo-Burma) (Tovaranonte et al., 2015), endemic species of Orchidaceae and Rubiaceae in Cameroon (Guinean Forests of West Africa) (Droissart et al., 2012), narrow-range endemic plant species in New Caledonia (Wulff et al. 2013) and all vascular plants endemic to the Mediterranean Basin (Cañas et al., 2014). Biodiversity studies using such natural history records always face heterogeneity of sample comprehensiveness related to irregular accessibility and botanical attractiveness between sites (Hortal et al., 2007). As a result of this sampling bias, the accuracy of the method tends to decrease as the resolution of grids increases so it may reach its maximum reliability at a scale too coarse for local decision-makers (Graham & Hijmans, 2006).

Species distribution models (SDMs; also referred to as ‘ecological niche models’ or ‘habitat suitability models’) are recognized as a promising alternative to correct gaps and biases in biodiversity databases (Graham & Hijmans, 2006). These models combine individual species occurrence records with a set of environmental predictors to estimate species’ ecological requirements and thereby predict habitat suitability across landscapes (Guisan & Zimmermann, 2000; Elith & Leathwick, 2009). SDMs are increasingly used to guide conservation actions related to biological invasions, identification of critical habitats, translocation of endangered species and reserve selection (Guisan et al., 2013).

Stacked species distribution models (S-SDMs), the combination of multiple SDMs, have been proposed to produce a community-level model (Ferrier & Guisan, 2006) and to provide a first fine-scale richness assessment in the short term (Murray-Smith et al., 2009; Raes et al., 2009; Pérez & Font, 2012; Schmidt-Lebuhn et al., 2012; Mateo et al., 2013; Moraes et al., 2014). S-SDMs have been used to identify areas of conservation priority for 59 Bolivian palms (Tropical Andes) (Moraes et al., 2014), 64 endemic-threatened species of Myrtaceae found in the Atlantic coastal forest of Brazil (Murray-Smith et al., 2009) and 408 plant taxa in Ecuador (Mesoamerica) (Mateo et al., 2013). S-SDMs have also been used to guide conservation efforts in the threatened forests of Borneo (Sundaland) by identifying previously unknown areas of high botanical diversity and endemism on the basis of 1439 SDMs (Raes et al., 2009).

Unfortunately, in these studies conducted in tropical biodiversity hotspots, S-SDM assessment was only based on the global accuracy of individual models (e.g. using the area under the ROC curve (AUC) as evaluation criterion; Fielding & Bell, 1997), which does not estimate spatial or environmental variation in the model accuracy. The lack of studies in tropical biomes comparing S-SDM-derived communities from communities observed in the field may reflect the lack of biological data available in tropical areas (Cayuela et al., 2009) or that the species accumulation curves reach saturation in much larger areas than in temperate regions (Plotkin et al., 2000; He et al., 2002; Wang et al., 2008). To our knowledge, only three studies have evaluated S-SDMs using plot inventories and they all took place in temperate regions. In Guisan et al. (1999), S-SDM-derived richness in tree and shrub species widespread in Nevada, USA was compared to species richness observed in 71 plots measuring 20 m × 20 m (400 m²). In Dubuis et al. (2011) and Pottier et al. (2013), estimation of the non-woody plant species richness found in Swiss Alps was evaluated using 335 and 298 vegetation plots measuring 2 m × 2 m (4 m²), respectively. These studies found that the reliability of S-SDMs varies along elevational gradients, but the underlying reasons remain unclear (Dubuis et al., 2011; Pottier et al., 2013).

Pottier et al. (2013) demonstrated that S-SDM success increased from low to midelevations (up to 1700 m) and interpreted this tendency as a consequence of increased climate determinism (addressed by SDM) and decreased competition (difficult to assess through SDM). In contrast, the accuracy of their S-SDM decreased from mid to high elevations. The authors argued that environmental predictors they selected failed to capture the uneven distribution of snow, rocky and bare soils, and microclimatic refuges associated with alpine environments. Increasing errors at increasing elevation might therefore result in insufficient species filtering towards high elevations where the overestimation rate (ratio between model estimates and the actual species richness found in independent evaluation plots) of the S-SDM was up to 500%. A relevant method to test this hypothesis would be to reproduce such studies in a non-alpine tropical environment (fully vegetated and without snow, rocky and bare soils or particular microclimatic refuges) and test whether overprediction along an elevational gradient persists.
In addition to climate, another driver of the $\alpha$-diversity in montane ecosystems is the indirect area effect (Romdal & Grytnes, 2007). In most mountain regions, the regional pool of species (the $\gamma$-diversity) decreases with elevation as a reflection of the decline of elevational band areas (McCain, 2007; McCain & Grytnes, 2010). Yet, the $\alpha$-diversity is a sample drawn from the $\gamma$-diversity available in each elevational band such that both diversity metrics are usually correlated (Graves & Rahbek, 2005; He et al., 2005). Thus, the regional area is expected to affect the distribution of the $\alpha$-diversity. Because S-SDMs are per-pixel procedures (under their most classical implementation), they allocate species richness values irrespective of the regional area covering the elevational band to which a pixel belongs (Franklin, 2010). Hence, one might assume that the indirect area effect can also compromise the accuracy of S-SDMs.

The aim of this study was to determine whether S-SDM accuracy varies along an environmental gradient in a tropical biodiversity hotspot, thereby expanding the geographic scope of past studies in temperate regions, and, if so, whether this variation arises from increasingly unpredictable environments or the indirect area effect. For the first time, we assessed S-SDM estimates using independent exhaustive tree inventories covering one hectare (1 ha), an international standard in which tropical rain forest communities are adequately represented.

METHODS

Study site

New Caledonia (south-west Pacific Ocean) is well suited to examine whether S-SDM accuracy is context-dependent in a tropical hotspot by virtue of steep climate variations and a diversity of substrates along with a complex topography that creates a wide variety of environmental gradients (Jaffré, 1993). It is an archipelago situated slightly north of the tropic of Capricorn (20–23°S – 164–167°E) c. 1500 km west of Australia and 2000 km north of New Zealand. It comprises the main island of Grande Terre, the Loyalty Islands, the Belep archipelago and the Isle of Pines (Fig. 1). Grande Terre is 350 km in length and 50–70 km wide oriented north-west to south-east and bisected by an almost continuous mountain range reaching 1628 m a.s.l. in the north (Mont Panié) and 1618 m in the south (Mont Humboldt). The climate is tropical with annual mean temperature between 27 and 30°C from November to March and between 20 and 23°C from June to August, and annual precipitation ranging from 300 to 4200 mm with greater precipitation on the east coast (Météo-France 2007). The archipelago is roughly dominated by three types of substrates: ultramafic substrates covering the southern third of Grande Terre and isolated massifs of the north-western coast, volcano-sedimentary substrates roughly covering the northern two-thirds of Grande Terre, and calcareous substrates prevailing on the Loyalty Islands. The combination of this variety of environments along with a complex biogeographical history would be behind the exceptional biological diversity of the hotspot (more than 3300 plant species with an endemism rate of 75%) (Morat et al., 2012). Today, the main threat to the unique biota of New Caledonia is probably habitat loss and fragmentation caused by fire and mining activities such that rain forests now only cover a quarter of the archipelago (Fig. 1) (Jaffré et al., 2012).

Species occurrence data

This study focused on tree species. Although they represent barely a quarter of the total vascular flora of the hotspot (22%; Morat et al., 2012), tree species have been the object of particular attention in herbarium records and forest inventories over the last 50 years. A list of tree species was drawn from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN) made of 201 plots.
Herbarium with Hugh S. MacKee (the principal contributor to the NOUmens either: (1) geo-referenced with a GPS; (2) collected by were collected. To address this issue, we used only speci-
ationable accuracy in the location where the oldest specimens
much larger amount of data (62%).

To build SDMs, occurrence data were compiled from
0.04 ha plots (not 1 ha plots, set aside for S-SDM assess-
ment) as well as voucher specimens through the database of
the Herbarium of the IRD Centre of Noumea (NOU)
(Table 1; http://herbier-noumea.plantnet-project.org). These
two datasets offered interesting complementarities. Vegeta-
tion plots captured the most widespread plant species often
abandoned by collectors focusing on species of greater botan-
ical interest such as rare species. Plots were also positioned
according to a systematic sampling design while herbarium
data were derived from opportunistic sampling schemes
prone to spatial autocorrelation. In contrast, herbarium data
were collected by other botanists but with an elevation recorded
in the field matching with an accuracy of ± 50 m the ele-
vation extracted from a digital elevation model (DEM)
using geographical coordinates derived from specimen local-
ities. Occurrence data were computed in a 100 m × 100 m
grid. When several records of the same species occurred in
the same cell, a single record was conserved to avoid sam-
pling biases associated with taxa duplicated in the same grid
cell. The final dataset comprised 10,530 occurrence records.
The mean number of cells a species occupied was 14. A
total of 181 species had less than three occurrences, and the
most collected species Polyscias dioica (Araliaceae) occurred
in 90 cells. The minimum number of occurrences that a
species had to have to be included in the S-SDM was set to
four as used in Pearson et al. (2007) because this threshold
maximized the number of species included in the S-SDM
so that it provided a prediction closer to the actual num-
ber of species) without significantly reducing its overall
accuracy (Table S1). The final S-SDM encompassed 562
species including 517 dicots (92%), 19 monocots (palms
and Pandanus) (3%), 20 gymnosperms (4%) and 6 tree
ferns (1%) (Table S2). The endemism rate of the selected
species set was 95% which ensured that we were modelling
distributions within the full geographical area of most
species.

GIS environmental variables

Eight environmental variables, including six continuous and
two categorical variables, were selected to describe species
habitat (Table 2). Continuous variables were correlated with
an \( r^2 < 0.80 \). Grid cell size was chosen to be 100 m × 100 m
to match the resolution of occurrence data and the area of
independent evaluation plots. Distal climatic factors were
preferred to proximal factors as local climate maps were
unsuitable both in quantity (mean annual precipitation was
the sole climatic variable available) and resolution (3 km)
(Météo-France 2007), and global products such as those pro-
vided by WorldClim may not be reliable on remote oceanic
islands with scarce meteorological stations (Hijmans et al.,
2005).

Table 1 Summary of the different occurrence datasets and their respective uses.

<table>
<thead>
<tr>
<th></th>
<th>Herbarium specimens</th>
<th>NC-PIPPN 0.04 ha plots</th>
<th>NC-PIPPN 1 ha plots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compiling a list tree species</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Building individual SDMs</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Evaluating the S-SDM</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SDM, species distribution models; S-SDM, stacked species distribution models; NC-PIPPN, New Caledonian Plant Inventory and Permanent Plot Network.

measuring 20 m × 20 m (0.04 ha) and 8 plots measuring
100 m × 100 m (1 ha) distributed across rain forests of
Grande Terre (Table 1; Ibanez et al., 2014). All native tree,
palm and tree fern species having at least one recorded indi-
vidual with a diameter at breast height (d.b.h.; i.e. 1.3 m)
≥ 10 cm in the NC-PIPPN network were listed as tree spe-
cies. Infraspecific taxa (subspecies and varieties) were not
considered and merged at the species rank. This method led
to the selection of 743 tree species.

To build SDMs, occurrence data were compiled from
0.04 ha plots (not 1 ha plots, set aside for S-SDM assess-
ment) as well as voucher specimens through the database of
the Herbarium of the IRD Centre of Noumea (NOU)
(Table 1; http://herbier-noumea.plantnet-project.org). These
two datasets offered interesting complementarities. Vegeta-
tion plots captured the most widespread plant species often
abandoned by collectors focusing on species of greater botan-
ical interest such as rare species. Plots were also positioned
according to a systematic sampling design while herbarium
data were derived from opportunistic sampling schemes
prone to spatial autocorrelation. In contrast, herbarium data
targeted plants without a minimum d.b.h. and provided a
much larger amount of data (62%).

A major inconvenience of herbarium data was the ques-
tionable accuracy in the location where the oldest specimens
were collected. To address this issue, we used only speci-
mens either: (1) geo-referenced with a GPS; (2) collected by
Hugh S. MacKee (the principal contributor to the NOU
Herbarium with c. 45,000 specimens) and estimated to have
an horizontal accuracy lower than 500 m according to the
gazetteer available at http://phanero.novcal.free.fr/; or (3)
collected by other botanists but with an elevation recorded
in the field matching with an accuracy of ± 50 m the ele-
vation extracted from a digital elevation model (DEM)
using geographical coordinates derived from specimen local-
ities. Occurrence data were computed in a 100 m × 100 m
grid. When several records of the same species occurred in
the same cell, a single record was conserved to avoid sam-
pling biases associated with taxa duplicated in the same grid
cell. The final dataset comprised 10,530 occurrence records.
The mean number of cells a species occupied was 14. A
total of 181 species had less than three occurrences, and the
most collected species Polyscias dioica (Araliaceae) occurred
in 90 cells. The minimum number of occurrences that a
species had to have to be included in the S-SDM was set to
four as used in Pearson et al. (2007) because this threshold
maximized the number of species included in the S-SDM
so that it provided a prediction closer to the actual num-
ber of species) without significantly reducing its overall
accuracy (Table S1). The final S-SDM encompassed 562
species including 517 dicots (92%), 19 monocots (palms
and Pandanus) (3%), 20 gymnosperms (4%) and 6 tree
ferns (1%) (Table S2). The endemism rate of the selected
species set was 95% which ensured that we were modelling
distributions within the full geographical area of most
species.

Table 2 Summary of the environmental predictors used in this study to describe tree species habitats in New Caledonia.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Abbrev.</th>
<th>Type</th>
<th>Min value</th>
<th>Max value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phytogeographical sectors</td>
<td>PHY</td>
<td>Categorical</td>
<td>(north-east, north-west, centre-east, centre-west, south)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>ELE</td>
<td>Continuous</td>
<td>0</td>
<td>1626</td>
<td>m</td>
</tr>
<tr>
<td>Substrates</td>
<td>SUB</td>
<td>Categorical</td>
<td>(volcano-sedimentary, ultramafic, calcareous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to the nearest road</td>
<td>DNR</td>
<td>Continuous</td>
<td>0</td>
<td>35</td>
<td>km</td>
</tr>
<tr>
<td>Windwardness</td>
<td>WIN</td>
<td>Continuous</td>
<td>0</td>
<td>69</td>
<td>km</td>
</tr>
<tr>
<td>Slope steepness</td>
<td>SLO</td>
<td>Continuous</td>
<td>0</td>
<td>203</td>
<td>%</td>
</tr>
<tr>
<td>Potential insolation</td>
<td>PSR</td>
<td>Continuous</td>
<td>0</td>
<td>2175</td>
<td>kWh m⁻²</td>
</tr>
<tr>
<td>Topographical Wetness Index</td>
<td>TWI</td>
<td>Continuous</td>
<td>1</td>
<td>24</td>
<td>Dimensionless</td>
</tr>
</tbody>
</table>
Elevation was used as a surrogate of air temperature (a difference of c. 10°C is expected between sea level and highest summits) which is one of the most important factors controlling key ecological processes in mountain ecosystems such as evapotranspiration, carbon fixation and decomposition, plant productivity and mortality (Chen et al., 1999). Elevation was derived from a Shuttle Radar Topographic Mission DEM with 90 m resolution resampled at 100 m using the nearest neighbour method. Slope steepness was expected to determine flow velocity and water drainage (Wilson & Gallant, 2000). Potential insolation (mean annual incoming solar radiation) equated to the amount of radiative energy received over the ground (Fu & Rich, 2002). Windwardness expressed the asymmetrical precipitation gradient between the windward (east) and leeward (west) coasts imposed by Foehn winds (Météo-France 2007). The topographical wetness index (TWI) was a function of the upstream contributing area and the slope of the landscape. It was expected to be informative of soil properties and the terrain position (with low values in convex terrains like ridges and high values in concave terrains like thalwegs) and to quantify the drainage of fluids (including water and air masses) (Moore et al., 1991; Pouteau et al., 2011).

In addition to topographical variables, distance to the nearest road was linked to the degree of human disturbance associated with urbanization, fire, mining activities and alien species invasions. The road shapefile was downloaded from the OpenStreetMap database (http://www.openstreetmap.org). Substrate types (volcano-sedimentary, ultramafic, calcareous) and phytogeographical sectors (north-east, north-west, centre-east, centre-west, south) reflected entities known to harbour a distinctive portion of the New Caledonian flora. The former was based on a shapefile published in the Atlas of New Caledonia (Fritsch, 2012), and the latter was taken from Jaffré & Veillon (1989).

Stacked species distribution model design

For each species, an individual MAXENT model was built using the version 3.3.3k of the software available at http://www.cs.princeton.edu/~schapire/maxent/ (Phillips et al., 2006). Models were trained on 10,000 pseudo-absence points drawn at random from background pixels. The convergence threshold was set at 0.00005, and the maximum number of iterations at 500 and suitable regularization values, β, included to reduce overfitting were set automatically by the program (Phillips et al., 2006). Ten replicate runs were performed for model building. To assess the predictive capacity of the models, we randomly split the data at each run so that models were calibrated using 70% of species occurrences and evaluated for predictive accuracy using the remaining 30% of the dataset. The overall predictive performance of individual SDMs was evaluated by calculating the AUC given in Table S2 (Fielding & Bell, 1997). The MAXENT model yielded a map of estimated habitat suitability for each species and for each pixel covering rain forests of the archipelago. Species richness maps were obtained by summing individual model outputs and not by summing threshold-based presence-absence outputs as the latter generally yield incorrect results (Calabrese et al., 2014).

Stacked species distribution model assessment

Species richness predicted by the final S-SDM was compared to the number of species measured in 1 ha independent inventories located on the island of Grande Terre (Fig. 1). In inventories, all trees with a d.b.h. above the international standard of 10 cm were identified by local experts. Independent inventories included the eight 1 ha plots of the NC-PIPPPN network (Table 1), and three transects extensively described in Jaffré & Veillon (1991, 1995). Putative changes in taxonomic classifications used between both inventory campaigns (2013–2015 versus early 1990s) were assumed to have a negligible effect on measured species richness. To compare S-SDM estimates to species richness measured in the field, inventories were associated with the appropriate 100 m × 100 m pixel on which it was centred.

Testing for an indirect area effect on S-SDM accuracy

The shape of mountains provides smaller habitats for species as elevation increases, and this can influence the γ-diversity found in each elevational band because of the basic species richness/habitat size relationship equated by Arrhenius:

\[
\gamma_{emp} = \gamma_{cor} A^z,
\]

where \(\gamma_{emp}\) is the empirical γ-diversity, \(\gamma_{cor}\) a constant that equals the number of species that the elevational band would support if it was confined to a one square unit, \(A\) is the regional area of the elevational band and \(z\) a constant describing the slope in a double-log species-area plot (Arrhenius, 1921).

As the empirical local species richness \(\gamma_{emp}\) is known to be correlated to \(\gamma_{emp}\) (Graves & Rahbek, 2005; He et al., 2005), \(\gamma_{cor}\), the theoretical local species richness if all elevational bands would have the same area, should logically be correlated to \(\gamma_{cor}\). As a consequence of Arrhenius’ equation, \(\gamma_{emp}\) should then be interpreted as correlated to \(\gamma_{cor} A^z\). Because S-SDMs ignore the regional area covering elevational bands, S-SDM estimates appear to reflect the definition of \(\gamma_{cor}\), that is the area-independent local species richness. Hence, we attempted to add the effect of decreasing available area on S-SDM-derived diversity by multiplying it by the factor \(A^z\).

The full set of occurrences available for selected species was used to describe the elevational pattern of tree γ-diversity in New Caledonia. The elevational range of each species was determined using the minimum and maximum elevations where the species occurs (Wang et al., 2007; McCain & Grytnes, 2010; Tang et al., 2014). The total number of tree species predicted in each 100-m elevation band was used as estimate of \(\gamma_{emp}\). A was the planar area occupied by each 100-m elevational band. The constant \(z\) was empirically determined on a log-log scale (see McCain (2007) for further details).
We kept the same z value to account for the hypotheti-
cal indirect area effect on S-SDM estimates. Prior to multi-
plication by $A^2$, S-SDM estimates were rescaled such that
adjusted modelled species richness associated with the eval-
uation plot located at the lowest elevation (Rivièr Bleu pente, 159 m) matched species richness measured in the field. This rescaling method allowed bringing closer species
richness estimations and measurements without affecting S-
DM assessment as it was based on correlation, not on
comparison of absolute values. Indeed, the aim of the
study was to determine the between-pixel relative richness
and not to predict exactly the empirical number of species
found in independent plots.

RESULTS

The number of tree species found in 1 ha independent evalu-
ination inventories ranged from 39 (Gohapin) to 131 (Rivièr Bleu pente) with a mean of 90 ± 22 species (Table 3).

Unadjusted species richness estimated by the S-SDM in evaluation inventories ranged from 95 species (Tiwae) to 251 species (Aoupinié) with a mean of 153 ± 49 species. Species
richness was well predicted in inventories located in lowland: Rivièr Bleu pente (159 m), Rivièr Bleu alluvions (176 m) and Tiwae (252 m), but it was two to three times the richness
observed in the highest plots: Aoupinié (889 m), Ateu (785 m), Laguen (572 m) and Bouirou (521 m). Reflecting
this tendency, we noticed that S-SDM overestimation rate
varies with elevation ($r^2 = 0.71$; $P-value < 0.05$) but not with any other continuous environ-
mental variable ($r^2 < 0.15$; $P-value > 0.05$) (Fig. 2).

The empirical $\gamma$-diversity estimated by herbarium collection and 0.04 ha forest inventories along the elevational gra-
dient in New Caledonia showed a peak of 420 species in the
range 500–600 m while the number of species was c. 250
between sea level and 100 m and only seven in the highest
elevation band 1500–1628 m (Fig. 3). The log-transformed empirical $\gamma$-diversity was dependent on the log-transformed available area ($r^2 = 0.85$; $P-value < 0.05$), and this relation had a slope $z = 0.44$ that was used to obtain values of the
area-corrected $\gamma$-diversity peaking around the mid-domain,
at 900–1000 m.

The addition of the indirect area effect to adjust S-SDM estimates reduced overestimation based on elevation ($r^2 = 0.04$; $P-value = 0.56$) (Fig. 2). Correlation between empirical and modelled
$\alpha$-diversity was non-significant with the unadjusted S-SDM ($r^2 = 0.03$; $P-value = 0.60$) (Fig. 4). However, when the indirect area effect was addressed, adjusted model estimations were significantly correlated to
plot measurements ($r^2 = 0.48$; $P-value < 0.05$), and this was
not sensitive to z values as the correlation remained signifi-
cant at a 5% threshold for values ranging from 0.34 to
0.63.

DISCUSSION

Accuracy of stacked species distribution models
varies with elevation

We documented a linear decrease in the ability of an S-SDM
to predict tree species richness in New Caledonian rain for-
ests along a c. 750 m wide gradient of elevation. In contrast,
S-SDM prediction appears to be more accurate in European
mid-elevation sites than at low elevations because of a
harsher climate driving more strongly the local assembly of
species (Pottier et al., 2013). This contrast may mean that, in

Table 3 Summary of the 11 independent one-hectare forest inventories used to assess the stacked species distribution model (S-SDM).

<table>
<thead>
<tr>
<th>Inventory</th>
<th>Abbrev.</th>
<th>Lat.</th>
<th>Lon.</th>
<th>PHY</th>
<th>ELE</th>
<th>SUB</th>
<th>SLO</th>
<th>WIN</th>
<th>DNR</th>
<th>PSR ($\times10^6$)</th>
<th>TWI</th>
<th>Obs. SR</th>
<th>Unadjusted S-SDM</th>
<th>Adjusted S-SDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rivièr Bleu alluvions*</td>
<td>Rba</td>
<td>−22.103</td>
<td>166.692</td>
<td>S</td>
<td>159</td>
<td>UM</td>
<td>1</td>
<td>13</td>
<td>0.6</td>
<td>1.61</td>
<td>10.8</td>
<td>103</td>
<td>110</td>
<td>103</td>
</tr>
<tr>
<td>Rivièr Bleu pente*</td>
<td>Rbp</td>
<td>−22.096</td>
<td>166.671</td>
<td>S</td>
<td>176</td>
<td>UM</td>
<td>41</td>
<td>13</td>
<td>0.1</td>
<td>1.64</td>
<td>6.7</td>
<td>131</td>
<td>136</td>
<td>125</td>
</tr>
<tr>
<td>Tiwae</td>
<td>Ti</td>
<td>−20.815</td>
<td>165.136</td>
<td>CE</td>
<td>252</td>
<td>VS</td>
<td>38</td>
<td>7</td>
<td>1.0</td>
<td>1.44</td>
<td>7.1</td>
<td>91</td>
<td>114</td>
<td>97</td>
</tr>
<tr>
<td>Gohapin</td>
<td>Go</td>
<td>−21.257</td>
<td>165.241</td>
<td>CE</td>
<td>285</td>
<td>VS</td>
<td>33</td>
<td>31</td>
<td>0.9</td>
<td>1.46</td>
<td>6.5</td>
<td>39</td>
<td>84</td>
<td>69</td>
</tr>
<tr>
<td>Djève</td>
<td>Dj</td>
<td>−21.226</td>
<td>165.435</td>
<td>CE</td>
<td>374</td>
<td>VS</td>
<td>23</td>
<td>11</td>
<td>1.8</td>
<td>1.54</td>
<td>7.9</td>
<td>95</td>
<td>112</td>
<td>84</td>
</tr>
<tr>
<td>Forêt</td>
<td>Fpe</td>
<td>−21.599</td>
<td>165.786</td>
<td>CE</td>
<td>435</td>
<td>VS</td>
<td>18</td>
<td>23</td>
<td>0.1</td>
<td>1.64</td>
<td>7.1</td>
<td>99</td>
<td>179</td>
<td>127</td>
</tr>
<tr>
<td>Forêt Plate</td>
<td>Fpl</td>
<td>−21.148</td>
<td>165.120</td>
<td>CE</td>
<td>514</td>
<td>VS</td>
<td>11</td>
<td>37</td>
<td>1.2</td>
<td>1.67</td>
<td>11.3</td>
<td>92</td>
<td>142</td>
<td>92</td>
</tr>
<tr>
<td>Bouirou</td>
<td>Bo</td>
<td>−21.417</td>
<td>165.545</td>
<td>CE</td>
<td>521</td>
<td>VS</td>
<td>11</td>
<td>21</td>
<td>0.2</td>
<td>1.70</td>
<td>6.8</td>
<td>96</td>
<td>195</td>
<td>125</td>
</tr>
<tr>
<td>Laguen</td>
<td>La</td>
<td>−20.625</td>
<td>164.780</td>
<td>NE</td>
<td>572</td>
<td>VS</td>
<td>6</td>
<td>7</td>
<td>2.0</td>
<td>1.63</td>
<td>12.1</td>
<td>75</td>
<td>172</td>
<td>101</td>
</tr>
<tr>
<td>Ateu</td>
<td>At</td>
<td>−20.951</td>
<td>164.920</td>
<td>CE</td>
<td>785</td>
<td>VS</td>
<td>42</td>
<td>31</td>
<td>1.3</td>
<td>1.58</td>
<td>6.5</td>
<td>82</td>
<td>188</td>
<td>75</td>
</tr>
<tr>
<td>Aoupinié</td>
<td>Ao</td>
<td>−21.178</td>
<td>165.277</td>
<td>CE</td>
<td>889</td>
<td>VS</td>
<td>2</td>
<td>22</td>
<td>0.1</td>
<td>1.73</td>
<td>7.7</td>
<td>90</td>
<td>251</td>
<td>83</td>
</tr>
</tbody>
</table>

the tropics, interspecific competitive interactions are stronger and climate determinism weaker at higher elevations. Differences in rainfall along this New Caledonian elevational gradient potentially support this hypothesis (Météo-France 2007). However, windwardness also largely controls rainfall distribution in New Caledonia through the rain shadow effect (Météo-France 2007), but is not correlated with the model overestimation rate ($r^2 = 0.15; P$-value = 0.23). Moreover, Pottier et al. (2013) found a similar S-SDM overestimation in Europe from mid to high elevations (beyond 1700 m) which suggests that competition is not enough to explain S-SDM accuracy along a full elevational gradient. This stress hypothesis is therefore unlikely to drive overestimation in New Caledonia and may primarily reflect mountain morphology.

The effect of area on the accuracy of stacked species distribution models

As expected, the area of elevational bands declined with increasing elevation and our data show this decline...
Other causes that might influence the accuracy of stacked species distribution models

Our results show that overestimations of the empirical \( \alpha \)-diversity by the S-SDM were caused by the indirect area effect above other potential sources of error. Part of the residual error between area-adjusted S-SDM estimates and \textit{in situ} measurements might be attributable to (1) cumulative differences between relative suitability of habitats predicted by MAXENT and true occupancy probability that only presence–absence SDMs are able to provide (Guillera-Arroita et al., 2014); (2) microtopography, neglected by our 100-m resolution GIS variables; (3) the variety of forest dynamic stages observed in evaluation plots caused by contrasting disturbance histories associated with cyclonic events, Melanesian and early European transformation (through fire, logging and cultivation) not taken into account by our S-SDM implementation due to a lack of evident proxies (the Gohapin, Forêt Plate and Laguen plots show evidence of human occupation); (4) dispersal limitation and narrow-range endemism that can induce commission errors; (5) species that are present in independent inventories but with a d.b.h. < 10 cm; and (6) 'dark diversity', that is species that belong to a particular species pool but that are not locally present (Pärtel et al., 2011). Future work will try to refine models further to account for these elements. Our hope is that this method to explicitly account for area in S-SDM models can improve our predictions and provide more robust patterns on which to base management decisions in the New Caledonian biodiversity hotspot.

ACKNOWLEDGEMENTS

We are grateful to Benjamin A. Sikes (University of Kansas) and Richard P. Duncan (University of Canberra) for their comments on the manuscript and the English revision, to the NOU Herbarium curator Jacqueline Fambart-Tinel as well as all contributors without whom this work would have not been possible and to two anonymous referees for their insightful comments and criticisms. This study was cofounded by the Direction for Economic and Environmental Development (DDEE) of the North Province of New Caledonia and the French Research Institute for Development (IRD).

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

Table S1. Coefficients of determination ($r^2$) between species richness maps generated with various minimum sample sizes needed for a species to be included in the stacked species distribution model.

Table S2. Summary of the 562 plant species used in the stacked species distribution model.

**BIOSKETCH**

Robin Pouteau is a postdoctoral fellow at the Agronomy Institute of New Caledonia. He is interested in patterns and causes of plant distribution in high-elevation islands.

Author contributions: R.P. conceived the study. All the authors contributed to set up one-hectare plots. Elise Bayle implemented the S-SDM. R.P. designed the area-based adjustment method, performed the S-SDM assessment and wrote the first draft of the manuscript. P.B. and T.I. commented on the manuscript.

Editor: Mark Robertson