



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

Robin Pouteau<sup>1,2\*</sup>, Élise Bayle<sup>2,3</sup>, Élodie Blanchard<sup>1</sup>, Philippe Birnbaum<sup>1,4</sup>, Jean-Jérôme Cassan<sup>3</sup>, Vanessa Hequet<sup>5</sup>, Thomas Ibanez<sup>1</sup> and Hervé Vandrot<sup>1</sup>

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

\*Correspondence: Robin Pouteau, Diversité biologique et fonctionnelle des écosystèmes terrestres, Laboratoire de botanique et d'écologie végétale appliquées, Institut Agronomique néo-Calédonien (IAC), BP A5, 98848 Noumea, New Caledonia.  
E-mail: robin.pouteau@ird.fr

## 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; Cañadas *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 ( $\alpha$ -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 (Cañadas *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 the fact that 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  $\times$  20 m (400 m<sup>2</sup>). 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  $\times$  2 m (4 m<sup>2</sup>), 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 tro-

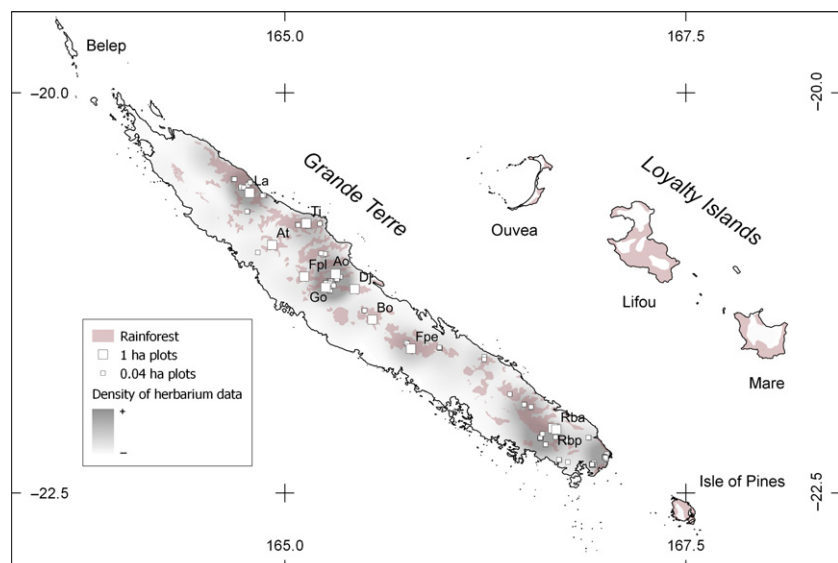
pic of Capricorn (20–23°S – 164–167°E) *c.* 1500 km west of Australia and 2000 km north of New Zealand. It comprises the main island of Grande Terre, the Loyalty Islands, the Belep archipelago and the Isle of Pines (Fig. 1). Grande Terre is 350 km in length and 50–70 km wide oriented north-west to south-east 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

**Figure 1** Map of the New Caledonian archipelago, its rain forest remnants and location of herbarium and forest inventory data used in this study. Independent one-hectare forest inventories have been labelled according to their abbreviation in Table 3.



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

	Herbarium specimens	NC-PIPPN	
		0.04 ha plots	1 ha plots
Compiling a list tree species		x	x
Building individual SDMs	x	x	
Evaluating the S-SDM			x

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 individual 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 species. Intraspecific 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 assessment) 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. Vegetation plots captured the most widespread plant species often abandoned by collectors focusing on species of greater botanical 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 questionable accuracy in the location where the oldest specimens were collected. To address this issue, we used only specimens 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 elevation extracted from a digital elevation model (DEM) using geographical coordinates derived from specimen localities. 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 sampling 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 number 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 provided by WorldClim may not be reliable on remote oceanic islands with scarce meteorological stations (Hijmans *et al.*, 2005).

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

Variable	Abbrev.	Type	Min value	Max value	Unit
Phytogeographical sectors	PHY	Categorical	(north-east, north-west, centre-east, centre-west, south)		
Elevation	ELE	Continuous	0	1626	m
Substrates	SUB	Categorical	(volcano-sedimentary, ultramafic, calcareous)		
Distance to the nearest road	DNR	Continuous	0	35	km
Windwardness	WIN	Continuous	0	69	km
Slope steepness	SLO	Continuous	0	203	%
Potential insolation	PSR	Continuous	0	2175	kWh m <sup>-2</sup>
Topographical Wetness Index	TWI	Continuous	1	24	Dimensionless

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,  $\beta$ , included to reduce overfitting were selected 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-PIPPN 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  $\times$  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  $\gamma$ -diversity found in each elevational band because of the basic species richness/habitat size relationship equated by Arrhenius:  $\gamma_{\text{emp}} = \gamma_{\text{cor}} A^z$ , where  $\gamma_{\text{emp}}$  is the empirical  $\gamma$ -diversity,  $\gamma_{\text{cor}}$  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  $\alpha_{\text{emp}}$  is known to be correlated to  $\gamma_{\text{emp}}$  (Graves & Rahbek, 2005; He *et al.*, 2005),  $\alpha_{\text{cor}}$ , the theoretical local species richness if all elevational bands would have the same area, should logically be correlated to  $\gamma_{\text{cor}}$ . As a consequence of Arrhenius' equation,  $\alpha_{\text{emp}}$  should then be interpreted as correlated to  $\alpha_{\text{cor}} A^z$ . Because S-SDMs ignore the regional area covering elevational bands, S-SDM estimates appear to reflect the definition of  $\alpha_{\text{cor}}$ , 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  $\gamma$ -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  $\gamma_{\text{emp}}$ .  $A$  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 hypothetical indirect area effect on S-SDM estimates. Prior to multiplication by  $A^z$ , S-SDM estimates were rescaled such that adjusted modelled species richness associated with the evaluation plot located at the lowest elevation (Rivière Bleue pente, 159 m) matched species richness measured in the field. This rescaling method allowed bringing closer species richness estimations and measurements without affecting S-SDM 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 evaluation inventories ranged from 39 (Gohapin) to 131 (Rivière Bleue pente) with a mean of  $90 \pm 22$  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  $153 \pm 49$  species. Species richness was well predicted in inventories located in lowland: Rivière Bleue pente (159 m), Rivière 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 was linearly correlated with elevation ( $r^2 = 0.71$ ;  $P$ -value  $< 0.05$ ) but not with any other continuous environmental 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 elevational 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 significant at a 5% threshold for values ranging from 0.34 to 0.63.

## DISCUSSION

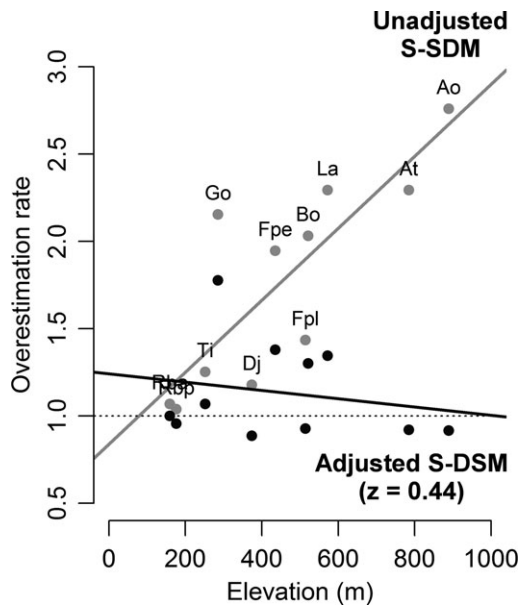
### Accuracy of stacked species distribution models varies with elevation

We documented a linear decrease in the ability of an S-SDM to predict tree species richness in New Caledonian rain forests 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

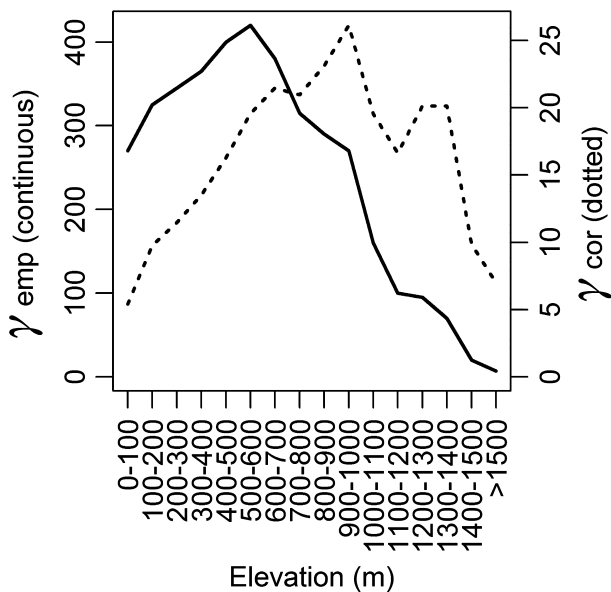
**Table 3** Summary of the 11 independent one-hectare forest inventories used to assess the stacked species distribution model (S-SDM). See Table 2 for environmental variable abbreviations.

Inventory	Abbrev.	Lat.	Lon.	PHY	ELE	SUB	SLO	WIN	DNR	PSR ( $\times 10^6$ )	TWI	Obs. SR	Unadjusted S-SDM	Adjusted S-SDM
Rivière Bleue alluvions*	Rba	-22.103	166.692	S	159	UM	1	13	0.6	1.61	10.8	103	110	103
Rivière Bleue pente*	Rbp	-22.096	166.671	S	176	UM	41	13	0.1	1.64	6.7	131	136	125
Tiwae	Ti	-20.815	165.136	CE	252	VS	38	7	1.0	1.44	7.1	91	114	97
Gohapin	Go	-21.257	165.241	CE	285	VS	33	31	0.9	1.46	6.5	39	84	69
Djeve	Dj	-21.226	165.435	CE	374	VS	23	11	1.8	1.54	7.9	95	112	84
Forêt Persant†	Fpe	-21.599	165.786	CE	435	VS	18	23	0.1	1.64	7.1	99	179	127
Forêt Plate	Fpl	-21.148	165.120	CE	514	VS	11	37	1.2	1.67	11.3	92	142	92
Bouirou	Bo	-21.417	165.545	CE	521	VS	11	21	0.2	1.70	6.8	96	195	125
Laguen	La	-20.625	164.780	NE	572	VS	6	7	2.0	1.63	12.1	75	172	101
Ateu	At	-20.951	164.920	CE	785	VS	42	31	1.3	1.58	6.5	82	188	75
Aoupinié	Ao	-21.178	165.277	CE	889	VS	2	22	0.1	1.73	7.7	90	251	83

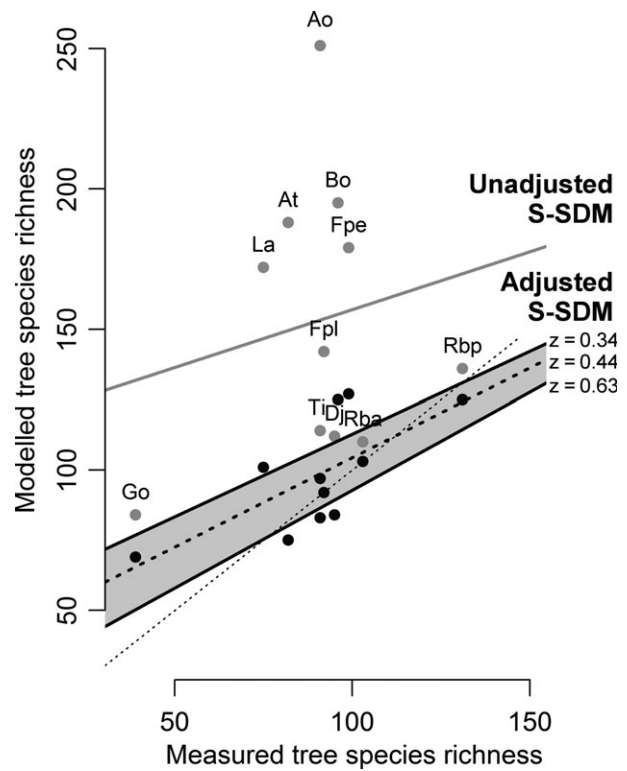
References: \*Jaffré & Veillon (1995); †Jaffré & Veillon (1991).



**Figure 2** Overestimation by the stacked species distribution model (S-SDM) of tree species richness measured in one-hectare plots along an elevational gradient in New Caledonia. The overestimation rate of the unadjusted S-SDM (continuous light grey line) is correlated to elevation while adjustment of S-SDM estimates by accounting for the indirect area effect (continuous black line) de-correlated the overestimation from elevation. The thin dashed black line is the  $\gamma = 1$  line. Forest inventories have been labelled according to their abbreviation in Table 3.



**Figure 3** Patterns of empirical  $\gamma$ -diversity (regional species richness) in tree species ( $\gamma_{emp}$ ) and area-adjusted  $\gamma$ -diversity ( $\gamma_{cor}$ ) along the elevational gradient in New Caledonia as estimated from records of the Noumea Herbarium and the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN) network of 0.04 ha plots.



**Figure 4** Correlation between empirical and predicted species richness for the unadjusted (continuous light grey line) and the area-adjusted (thick dashed black line) stacked species distribution models (S-SDM). Continuous black lines denote adjustments using the minimum and maximum  $z$  values for which S-SDM estimations remain significantly correlated to plot measurements. The thin dashed black line is the identity line ( $y = x$ ). Forest inventories have been labelled according to their abbreviation in Table 3.

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

significantly influenced the number of tree species found in each band. Interpolation of species ranges between its highest and lowest elevations might partially control the shape of the  $\gamma$ -diversity pattern we reported, with edges of the elevational gradient containing observed species while midelevations contain species observed plus species interpolated (McCain & Grytnes, 2010). However, other methods that address this potential sampling bias such as rarefaction analyses often produce a similar  $\gamma$ -diversity pattern as showed in Mount Kinabalu, Borneo (Grytnes & Beaman, 2006). The effect of the regional area has been evidenced to reduce plant  $\gamma$ -diversity in other mountain regions such as in the Gaoligong Mountains in Tibet (Wang *et al.*, 2007) and Taibai Mountain in central China (Tang *et al.*, 2014). The area has also been identified as a major driver of the  $\alpha$ -diversity in the Baekdudaegan Mountains in South Korea (Lee *et al.*, 2012). These studies converge towards the idea that area can be a better proxy of plant species richness than climate itself along elevational gradients, and this would be especially the case for tree species (Wang *et al.*, 2007; Lee *et al.*, 2012).

In New Caledonia, our area-based method for adjusting S-SDM overestimation empirically demonstrated the inherent inability of the method to account for the influence of this predominant factor on  $\alpha$ -diversity. Compared to adjusting bias by linking S-SDMs to macro-ecological models (Calabrese *et al.*, 2014), our method needs no extensive plot data and can directly address the underlying mechanisms causing such bias. When the indirect area effect was accounted for, the S-SDM successfully distinguished richness differences between plots located at similar elevations (e.g. Tiwae versus Gohapin, Forêt Plate versus Bouirou) on the basis of environmental variables. Despite remaining difficulties of predicting tree species richness in New Caledonian tropical rain forests when explicitly accounting for area at different elevations, our findings represent a significant step to improve S-SDM design and identify local hotspots within a larger biodiverse area.

### 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 *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

- Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.
- Bhattarai, K.R. & Vetaas, O.R. (2003) Variation in plant species richness of different life forms along a subtropical elevation gradient in the Himalayas, east Nepal. *Global Ecology and Biogeography*, **12**, 327–340.
- Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.
- Cañadas, E.M., Fenu, G., Peñas, J., Lorite, J., Mattana, E. & Bacchetta, G. (2014) Hotspots within hotspots: endemic plant richness, environmental drivers, and implications for conservation. *Biological Conservation*, **170**, 282–291.
- Cayuela, L., Golicher, D.J., Newton, A.C., Kolb, M., Alburquerque, F.S.D., Arets, E.J.M.M., Alkemade, J.R.M. & Pérez, A.M. (2009) Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science*, **2**, 319–352.
- Chen, J., Saunders, S.C., Crow, T.R., Naiman, R.J., Broszofski, K.D., Mroz, G.D., Brookshire, B.L. & Franklin, J.F. (1999) Microclimate in forest ecosystem and landscape ecology. *BioScience*, **49**, 288–297.
- Droissart, V., Hardy, O.J., Sonké, B., Dahdouh-Guebas, F. & Stévant, T. (2012) Subsampling herbarium collections to



- assess geographic diversity gradients: a case study with endemic Orchidaceae and Rubiaceae in Cameroon. *Biotropica*, **44**, 44–52.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677–697.
- Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.
- Fielding, A. & Bell, J. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Franklin, J. (2010) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, **16**, 321–330.
- Fritsch, E. (2012) Les sols. *Atlas de la Nouvelle-Calédonie* (ed. by J. Bonvallot, J.-C. Gay and E. Habert), pp. 73–76. IRD-Congrès de la Nouvelle-Calédonie, Marseille.
- Fu, P. & Rich, P.M. (2002) A geometric solar radiation model with applications in agriculture and forestry. *Computers and Electronics in Agriculture*, **37**, 25–35.
- Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578–587.
- Graves, G.R. & Rahbek, C. (2005) Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences USA*, **102**, 7871–7876.
- Grytnes, J.A. & Beaman, J.H. (2006) Elevational species richness patterns for vascular plants on Mount Kinabalu, Borneo. *Journal of Biogeography*, **33**, 1838–1849.
- Guillera-Arroita, G., Lahoz-Monfort, J.J. & Elith, J. (2014) Maxent is not a presence–absence method: a comment on Thibaud *et al.* *Methods in Ecology and Evolution*, **5**, 1192–1197.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Guisan, A., Weiss, S.B. & Weiss, A.D. (1999) GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology*, **143**, 107–122.
- Guisan, A., Tingley, R., Baumgartner, J.B., *et al.* (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.
- He, F.L., LaFrankie, J.V. & Song, B. (2002) Scale dependence of tree abundance and richness in a tropical rain forest, Malaysia. *Landscape Ecology*, **17**, 559–568.
- He, F.L., Gaston, K.J., Connor, E.F. & Srivastava, D.S. (2005) The local–regional relationship: immigration, extinction, and scale. *Ecology*, **86**, 360–365.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hortal, J., Lobo, J.M. & Jimenez-Valverde, A. (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology*, **21**, 853–863.
- Ibanez, T., Munzinger, J., Dagostini, G., Hequet, V., Rigault, F., Jaffré, T. & Birnbaum, P. (2014) Structural and floristic characteristics of mixed rainforest in New Caledonia: new data from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN). *Applied Vegetation Science*, **17**, 386–397.
- Jaffré, T. (1993) The relationship between ecological diversity and floristic diversity in New Caledonia. *Biodiversity Letters*, **1**, 82–87.
- Jaffré, T. & Veillon, J.-M. (1989) Morphology, distribution and ecology of palms in New Caledonia. *Palms of the South-West Pacific* (ed. by J.L. Dowe), pp. 158–169. Fund Palm and Cycad Society of Australia, Milton.
- Jaffré, T. & Veillon, J.-M. (1991) Etude floristique et structurale de deux forêts denses humides sur roches ultrabasiques en Nouvelle-Calédonie. *Adansonia*, **4**, 243–273.
- Jaffré, T. & Veillon, J.-M. (1995) Structural and floristic characteristics of a rain forest on schist in New Caledonia: a comparison with an ultramafic rain forest. *Adansonia*, **17**, 201–226.
- Jaffré, T., Rigault, F. & Munzinger, J. (2012) La végétation. *Atlas de la Nouvelle-Calédonie* (ed. by J. Bonvallot, J.-C. Gay and E. Habert), pp. 73–76. IRD-Congrès de la Nouvelle-Calédonie, Marseille.
- Lee, C.-B., Chun, J.-H., Song, H.K. & Cho, H.J. (2012) Altitudinal patterns of plant species richness on the Baekdu-daegan Mountains, South Korea: mid-domain effect, area, climate, and Rapoport's rule. *Ecological Research*, **28**, 67–79.
- Mateo, R.G., de la Estrella, M., Felicísimo, Á.M., Muñoz, J. & Guisan, A. (2013) A new spin on a compositionalist predictive modelling framework for conservation planning: a tropical case study in Ecuador. *Biological Conservation*, **160**, 150–161.
- McCain, C.M. (2007) Area and mammalian elevational diversity. *Ecology*, **88**, 76–86.
- McCain, C.M. & Grytnes, J.-A. (2010) Elevational gradients in species richness. *Encyclopedia of life science*. John Wiley & Sons, Chichester.
- Météo-France (2007) *Atlas climatique de la Nouvelle-Calédonie*. Météo-France in New Caledonia, Noumea.
- Moore, I.D., Grayson, R.B. & Ladson, A.R. (1991) Digital terrain modelling: a review of hydrological, geomorphological and biological applications. *Hydrological Processes*, **5**, 3–30.
- Moraes, M.M., Ríos-Uzeda, B., Moreno, L.R., Huanca-Huarachi, G. & Larrea-Alcazar, D. (2014) Using potential distribution models for patterns of species richness, endemism,

- and phytogeography of palm species in Bolivia. *Tropical Conservation Science*, **7**, 45–60.
- Morat, P., Jaffré, T., Tronchet, F., Munzinger, J., Pillon, Y., Veillon, J.-M. & Chalopin, M. (2012) The taxonomic database “FLORICAL” and characteristics of the indigenous flora of New Caledonia. *Adansonia*, **3**, 177–219.
- Murray-Smith, C., Brummitt, N.A., Oliviera-Filho, A.T., Bachman, S., Moat, J., Lughadha, E.M.N. & Lucas, E.J. (2009) Plant diversity hotspots in the Atlantic coastal forests of Brazil. *Conservation Biology*, **23**, 151–163.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Pärtel, M., Szava-Kovats, R. & Zobel, M. (2011) Dark diversity: shedding light on absent species. *Trends in Ecology and Evolution*, **26**, 124–128.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Townsend Peterson, A. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.
- Pérez, N. & Font, X. (2012) Predicting vascular plant richness patterns in Catalonia (NE Spain) using species distribution models. *Applied Vegetation Science*, **15**, 390–400.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Plotkin, J.B., Potts, M.D., Yu, D.W., Bunyavejchewin, S., Condit, R., Foster, R., Hubbell, S., LaFrankie, J., Manokaran, N., Lee, H.S., Sukumar, R., Nowak, M.A. & Ashton, P.S. (2000) Predicting species diversity in tropical forests. *Proceedings of the National Academy of Sciences USA*, **97**, 10850–10854.
- Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.
- Pouteau, R., Rambal, S., Ratte, J.-P., Gogé, F., Joffre, R. & Winkel, T. (2011) Downscaling MODIS-derived maps using GIS and boosted regression trees: the case of frost occurrence over the arid Andean highlands of Bolivia. *Remote Sensing of Environment*, **115**, 117–129.
- Raes, N., Roos, M.C., Slik, J.W.F., Van Loon, E.E. & ter Steege, H. (2009) Botanical richness and endemism patterns of Borneo derived from species distribution models. *Ecography*, **32**, 180–192.
- Romdal, T.S. & Grytnes, J.-A. (2007) An indirect area effect on elevational species richness patterns. *Ecography*, **30**, 440–448.
- Sánchez-González, A. & López-Mata, L. (2005) Plant species richness and diversity along an altitudinal gradient in the Sierra Nevada, Mexico. *Diversity and Distributions*, **11**, 567–575.
- Schmidt-Lebuhn, A.N., Knerr, N.J. & González-Orozco, C.E. (2012) Distorted perception of the spatial distribution of plant diversity through uneven collecting efforts: the example of Asteraceae in Australia. *Journal of Biogeography*, **39**, 2072–2080.
- Tang, L., Li, T., Li, D. & Meng, X. (2014) Elevational patterns of plant richness in the Taibai Mountain, China. *Scientific World Journal*, **2014**, 309053.
- Tomasetto, F., Duncan, R.P. & Hulme, P.E. (2013) Environmental gradients shift the direction of the relationship between native and alien plant species richness. *Diversity and Distributions*, **19**, 49–59.
- Tovaranonte, J., Blach-Overgaard, A., Pongsattayapipat, R., Svenning, J.-C. & Barfod, A.S. (2015) Distribution and diversity of palms in a tropical biodiversity hotspot (Thailand) assessed by species distribution modeling. *Nordic Journal of Botany*, **33**, 214–224.
- Wang, Z., Tang, Z. & Fang, J. (2007) Altitudinal patterns of seed plant richness in the Gaoligong Mountains, south-east Tibet, China. *Diversity and Distributions*, **13**, 845–854.
- Wang, X.G., Hao, Z.Q., Ye, J., Zhang, J., Li, B.H. & Yao, X.L. (2008) Spatial variation of species diversity across scales in an old-growth temperate forest of China. *Ecological Research*, **23**, 709–717.
- Wilson, J.P. & Gallant, J.C. (2000) *Terrain analysis: principles and applications*. John Wiley & Sons, New York, NY.
- Wulff, A.S., Hollingsworth, P.M., Ahrends, A., Jaffré, T., Veillon, J.-M., L’Huillier, L. & Fogliani, B. (2013) Conservation priorities in a biodiversity hotspot: analysis of narrow endemic plant species in New Caledonia. *PLoS ONE*, **8**, e73371.

## 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