Disentangling the processes driving tree community assembly in a tropical biodiversity hotspot (New Caledonia)

Robin Pouteau1,2 | François Munoz3 | Philippe Birnbaum1,4

1Institut Agronomique néo-Calédonien (IAC), Nouméa, New Caledonia
2Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China
3Laboratoire d’Écologie Alpine, CNRS, Université Grenoble Alpes, Grenoble, France
4UMR botAnique et Modélisation de l’Architecture des Plantes et des végétations (AMAP), Université de Montpellier, CIRAD, CNRS, INRA, IRD, Montpellier, France

Correspondence
Robin Pouteau, Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, School of Life Science, Taizhou University, Zhejiang Province, China. Email: robin.pouteau@uni-konstanz.de

Funding information
Direction du Développement Économique et de l’Environnement (DDEE); Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD); Institut Agronomique néo-Calédonien (IAC)

Abstract
Aim: Understanding the drivers of community diversity and composition is a major question in ecology, which is of particular significance for the conservation of narrowly distributed taxa facing habitat alteration. In this study, we examined three ecological processes (environmental and biotic filtering, and dispersal limitation) hypothesized to drive community assembly in the island biodiversity hotspot of New Caledonia, and whether the relative importance of these processes changes along environmental gradients.

Methods: First, we assessed environmental filtering by stacking binary presence/absence species distribution models generated for 678 tree species from c. 40,000 occurrences. Second, we acknowledged the influence of dispersal limitation by applying a buffer around each occurrence. Third, we modelled a local hierarchy of performances by ranking species in decreasing order of their predicted habitat suitability, up to the predicted local species richness. Predictions of stacked species distribution models (S-SDMs) accounting for different combinations of the filtering processes were compared with the composition of 12 forest plots.

Results: The three filters appear to be important driving forces. Accounting for environmental filtering provided fairly accurate assemblage predictions but species richness was overestimated. Considering dispersal limitation increased assemblage specificity (the proportion of correctly predicted absences) while considering differences in local performances increased assemblage sensitivity (the proportion of correctly predicted presences). The predictive ability of all S-SDMs decreased with elevation, which suggests that unpredictable stochastic processes influenced biodiversity dynamics in more productive mountain habitats.

Main conclusions: New Caledonian tree communities appear to be similar to those of other tropical regions in their structuring processes (important role of dispersal and biotic filtering, increasing role of neutral processes with productivity). Our findings stress the potential of using S-SDMs to understand and predict current and future biodiversity patterns.

Keywords
abiotic niche, alpha-diversity, community assembly, community saturation, dispersal limitation, spatial ecology, stacked species distribution model, tropical rain forest
INTRODUCTION

Understanding and predicting the composition of ecological communities in space and time is crucial for determining the fate of biodiversity under climate and land use changes (Urban et al., 2016). A key prerequisite is to determine how ecological processes shape species assemblages (Thuiller et al., 2013). Three processes are thought to be of prime importance: species have to disperse to a given site (dispersal filter), to find a suitable environment there (environmental filter) and to survive in presence of other resident species (biotic filter; Lortie et al., 2004). In the extreme case of neutral theory, stochastic fluctuations and dispersal limitation alone determine biodiversity patterns, that is, biological differences between members of an ecological community are “neutral” and do not influence their relative success (Hubbell, 2001). Conversely, much work in community and functional ecology has emphasized the role of niche differences in species coexistence (McGill, Maurer, & Weiser, 2006). It is now widely acknowledged that communities are structured by both niche and neutral processes (Gravel, Canham, Beaudet, & Messier, 2006), but deciphering their relative influence on biodiversity patterns remains challenging (Munoz & Huneman, 2016).

There is a growing body of literature providing empirical evidence for a less prominent role of environmental filtering in species-sorting processes at low latitude than at higher latitude, where harsher climatic conditions constrain community assembly (Myers et al., 2013). As a result, climatic constraints are thought to exert stronger control over species assemblage in temperate regions (Laughlin & Laughlin, 2013), while the composition of tropical communities more likely reflects the influence of dispersal limitation, historical factors and interactions with neighbours of the same (conspecific) and other (heterospecific) species (Kraft & Ackerly, 2010). In addition, the imprint of neutral dynamics should prevail in the least disturbed and richest habitats because of a high frequency of rare species that are more influenced by stochastic events (Chisholm & Pacala, 2010). Accordingly, this should lead to higher α-diversity under benign conditions (with exceptions, Šimová & Storch, 2017). On the contrary, communities in less productive tropical ecosystems would be more constrained by environmental conditions and display lower α-diversity, unless the local environment also promotes high rates of adaptive speciation (Šimová & Storch, 2017).

Several modelling frameworks have been proposed to decipher the relative influences of these entangled processes (see D’Amen, Rahbek, Zimmermann, & Guisan, 2017 for a review). The integrated framework called ‘spatially explicit species assemblage modelling’ (SESAM, Guisan & Rahbek, 2011) applies three successive filters to community assembly as follows: (a) dispersal filtering: definition of the local (within-site) species source pool from a regional source pool to account for historical and current dispersal limitations. Using an empirical approach, this aims to determine which species in the regional source pool could have dispersed to a given site, ignoring habitat requirements and competitive interactions; (b) environmental filtering: using a species distribution model (SDM) to predict the suitability of the abiotic environment for each species and stacking predictions (then abbreviated S-SDM, see Calabrese, Certain, Kraan, & Dormann, 2014 for a review) to define the environmentally filtered species pool and (c) local biotic filtering: using species richness models to predict the distribution of macroecological constraints, that is, factors that control assemblage saturation (also called “α-capacity”, Gavish et al., 2017), then applying ecological assembly rules to select species that are the most likely to coexist in each site.

Characterizing processes that generate and maintain biodiversity is a major question in ecological research with particular resonance in environments where exceptional concentrations of endemic species are undergoing exceptional loss of habitat for agriculture, settlement or other human use (Mateo, Mokany, & Guisan, 2017; Thuiller et al., 2013). Although islands account for only 5% of the Earth’s land area, they harbour a disproportionately high share of global biodiversity: islands host one-quarter of the world’s vascular flora and a very high diversity of habitats, thus accounting for a quarter of global biodiversity hotspots (Whittaker & Fernández-Palacios, 2007).

In island biodiversity hotspots, we expect the role of environmental filtering, dispersal limitation, species interactions and stochastic dynamics in shaping diversity patterns to depend on specific adaptations and environmental conditions. Island biotas are thought to exhibit several particularities: Sherwin Carlquist hypothesized that selective pressure for competitive ability is lowered when a species migrates to an island due to biota impoverishment and unoccupied environmental space, while during evolution on an island various groups would tend to lose dispersal ability (e.g. less dispersive seeds, flightlessness in birds; Carlquist, 1974). Such “insular syndromes” could lead to community structuring on islands differing from that observed on continents (Carlquist, 1974). Examining the role of the different filtering processes on islands is therefore important for understanding and predicting species assembly dynamics (Warren et al., 2015).

In this study, we aimed to disentangle the drivers of community assembly in a tropical island biodiversity hotspot, and to determine the relative importance of the drivers along environmental gradients. We expected less influence of competition and dispersal in shaping communities on islands than reported on continents, as a consequence of insular syndromes (Carlquist, 1974). In contrast, the diversity of habitats found over short distances on oceanic islands should entail a greater role of environmental filtering (Whittaker & Fernández-Palacios, 2007). Moreover, we hypothesized that environmental filtering should predominate on lower-productive, leeward coasts experiencing temporary rainfall deficit and in plains altered by human activities, while competition and neutral processes should predominate in higher mountain areas that are less disturbed and more productive.

We focused on rain forest tree communities of the world’s smallest biodiversity hotspot: New Caledonia, an archipelago located in the south-west Pacific Ocean hosting more than 3300 vascular plant species with an endemicism rate of c. 75% (Morat et al., 2012). There, countless botanical studies have provided an extraordinary
collection of herbarium specimens, inventories and other georeferenced records, but the ecology of tree communities is still poorly known (Birnbaum et al., 2015; Ibanez et al., 2017; Pouteau et al., 2015).

2 | MATERIALS AND METHODS

2.1 | Study area

New Caledonia is an archipelago in the Melanesia region situated slightly north of the tropic of Capricorn (20°–23°S–164–167°E). The main island of Grande Terre is 350 km in length and 50–70 km wide, oriented northwest to southeast, and bisected by an almost continuous mountain range reaching 1,628 m a.s.l. (Mont Panié). A variety of environments along with a complex biogeographical history are hypothesized to be the main determinants of the exceptional biological diversity of the hotspot (Morat et al., 2012; Pillon, Munzinger, Amir, & Lebrun, 2010). New Caledonia has a tropical climate with annual mean temperature in lowland areas between 27 and 30°C from November to March and between 20 and 23°C from June to August. Annual precipitation ranges from 300 to 4,200 mm, with greater precipitation on the windward east coast (Météo-France, 2007). Three main types of substrate occur in New Caledonia: (a) ultramafic substrates covering the southern third of Grande Terre and isolated ultramafic massifs along the west coast; (b) substrates derived from volcano-sedimentary rocks roughly covering the northern two-thirds of Grande Terre and (c) calcareous substrates prevailing in the Loyalty Islands (Fritsch, 2012).

2.2 | Species occurrence data

This study focused on 678 native tree species with an endemism rate of 95%, representing 99% of the currently recorded tree flora (Birnbaum et al., 2015). We used the occurrence database compiled by Birnbaum et al. (2015), which includes four datasets: (a) specimens extracted from the Noumea herbarium database http://herbier-noumea.plantnet-project.org (23%); (b) the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN) made of 201 "small" plots measuring 20 m × 20 m (29%); (c) unpublished inventories for impact assessments or projects of protected area (44%) and (d) georeferenced photographs (4%). The final dataset comprised 38,798 occurrence records with a minimum of four occurrences per species, a mean of 57 occurrences, and a maximum of 504 occurrences for Polyscias dioica (Araliaceae; see Appendix S1).

2.3 | GIS environmental variables

We defined six environmental predictors, including five continuous and one categorical variable, to describe New Caledonian rain forest habitats (Table 1). Correlation between continuous variables did not exceed the threshold of $|r| = 0.70$ beyond which collinearity begins to severely distort model estimations and subsequent predictions (Dormann et al., 2013). A grid cell size of 100 m × 100 m (area of 1 ha) was chosen to match the size of independent evaluation "large" plots (see Section 2.4). Communities are predicted to bear the imprint of environmental constraints, dispersal limitation and biotic interactions at such a local scale (Mateo et al., 2017). 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, Cameron, Parra, Jones, & Jarvis, 2005).

Elevation was used as a surrogate for 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 plant productivity and mortality, evapotranspiration, carbon fixation and decomposition (Chen et al., 1999). Elevation was derived from a digital elevation model having a 10-m horizontal resolution (Juffroy, 2012), upsampled by averaging 10 × 10 pixels to the resolution of 100 m. A range of terrain attributes were subsequently calculated using the SAGA software 6.1.0 (Conrad et al., 2015). Slope steepness was expected to determine flow velocity and water drainage (Wilson & Gallant, 2000). Potential insolation (mean annual incoming solar radiation) was used to quantify the amount of radiative energy received over the ground (Fu & Rich, 2002). Distance to the east coast expressed the asymmetrical precipitation gradient between the windward (east) and leeward (west) coasts imposed by foehn winds (Météo-France, 2007). The topographical wetness index, a function of the upstream contributing area and the slope of the landscape, was expected to be informative of soil properties and terrain position (with low values in convex terrains such as ridges and high values in concave terrains such as thalwegs), and to reflect the influence of water

<p>| TABLE 1 | Summary of the environmental predictors used in species distribution models of tree species in New Caledonia. Terrain variables were derived from a digital elevation model (Juffroy, 2012) and the substrate classification from a published shapefile (Fritsch, 2012). Contribution stands for the mean relative increase in AUC when the predictor was included to a species distribution model |</p>
<table>
<thead>
<tr>
<th>Environmental predictor</th>
<th>Type</th>
<th>Min</th>
<th>Max</th>
<th>Unit</th>
<th>Contribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>Continuous</td>
<td>0</td>
<td>1628</td>
<td>m</td>
<td>36</td>
</tr>
<tr>
<td>Slope</td>
<td>Continuous</td>
<td>0</td>
<td>203</td>
<td>%</td>
<td>11</td>
</tr>
<tr>
<td>Potential solar radiation</td>
<td>Continuous</td>
<td>1245</td>
<td>2175</td>
<td>kWh/m²</td>
<td>10</td>
</tr>
<tr>
<td>Distance to the east coast</td>
<td>Continuous</td>
<td>0</td>
<td>69</td>
<td>km</td>
<td>16</td>
</tr>
<tr>
<td>Topographic wetness index</td>
<td>Continuous</td>
<td>1</td>
<td>24</td>
<td>None</td>
<td>10</td>
</tr>
<tr>
<td>Substrate</td>
<td>Categorical</td>
<td>Volcano sedimentary, ultramafic, calcareous</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and air fluxes (Moore, Grayson, & Ladson, 1991; Pouteau et al., 2011). Substrate classification (volcano sedimentary, ultramafic, and calcareous) reflected the contrasted soil conditions harbouring distinct floras in New Caledonia (Morat et al., 2012). This classification was based on an existing shapefile published in the Atlas of New Caledonia (Fritsch, 2012).

2.4 | Experimental design

We predicted assemblage composition for four different S-SDMs incorporating different constraints: (a) ‘Environment’, that is, accounting for environmental filtering by summing binary SDM predictions; (b) ‘+Dispersal’, that is, accounting for dispersal limitation by summing binary SDMs and constraining predictions around known occurrences; (c) ‘+Saturation’, that is, by using a probability ranking rule representing a hierarchy of local performance, up to a saturation in richness and without spatial constraint and (d) ‘All’, that is, accounting for both dispersal barriers and the hierarchy of local performance by using both spatial constraints and the probability ranking rule. S-SDMs ‘Environment’ and ‘+Dispersal’ were thus based on a thresholding scheme of predicted occurrences (without assembly saturation), while S-SDMs ‘+Saturation’ and ‘All’ were threshold free and considered a local saturation rule.

2.5 | S-SDMs accounting for environmental filtering

Nine SDMs were used to model the environmental preferences of each species: generalized linear models (GLM), general additive models (GAM), multivariate adaptive regression splines (MARS), classification tree analysis (CTA), generalized boosted models (GBM), maximum entropy (Maxent), artificial neural networks (ANN), random forests (RF) and support vector machines (SVM). The maximum number of iterations for GLM, GAM and ANN was set to 500. The final leaf size for CTA, GBM and RF was set to 1, and the maximum number of trees for GBM and RF was set to 2,500. The shrinkage threshold for GBM was set to 1e−3. The degree of interaction for MARS was set to 2.

Because our sampling scheme did not account for species absences, pseudoabsences (randomly selected sites where a species is assumed to be absent) were used in SDMs. Pseudoabsence selection was repeated 10 times to reduce the potential effect of randomization, then results of all runs were merged by averaging habitat suitability probabilities (Barbet-Massin, Jiguet, Cécile Hélène, & Wilfried, 2012). The number of selected pseudoabsences has been set to the recommendations of Barbet-Massin et al. (2012) adapted to each statistical method: 1,000 pseudoabsences for GLM, GAM and MARS, and the same number as available presences for CTA, GBM, Maxent, ANN, RF and SVM.

To assess the predictive ability of an SDM, we randomly split the data at each run so that the model was calibrated using 70% of species occurrences and evaluated for predictive accuracy using the remaining 30%. The overall predictive performance of an individual SDM was evaluated by calculating the area under the curve (AUC) of the receiver operating characteristic plot and the true skill statistic (TSS), the former being prevalence dependent but not the latter (Allouche, Tsoar, & Kadmon, 2006). We used a jackknife approach to evaluate the difference in AUC between a full SDM and one with each environmental variable omitted in turn, and this difference was used to assess the relative importance of environmental variables.

An ensemble forecasting approach was then performed to overcome the problem of variability in the prediction of each statistical method. If an SDM had an AUC <0.75, it was not included in the consensus SDM as it was considered not accurate enough. If the AUC was >0.75, it was used to weight the means of predicted probabilities and to combine ensembles of species range forecasts (Marmion, Parviainen, Luoto, Heikkinen, & Thuiller, 2009). We converted continuous habitat suitability maps into binary presence/absence maps by applying a threshold so that sensitivity was equal to specificity (Liu, Berry, Dawson, & Pearson, 2005). Species distribution modelling was performed using the package ‘SSDM’ (Schmitt et al., 2017) developed in R (R Development Core Team, 2017).

2.6 | S-SDMs accounting for dispersal limitation

The New Caledonian flora includes many narrow-range endemic plant species (also called ‘micro-endemics’, Wulff et al., 2013). Narrow-range endemism typically violates the SDM equilibrium assumption and can induce commission errors in biodiversity models if species distribution reflects dispersal limitation and historical factors rather than specific environmental conditions (Pouteau et al., 2015). To avoid such errors, SDM predictions for each species were restricted within a circular buffer area around the occurrences of the species, that is, sites located at distances greater than the buffer distance were assigned a habitat suitability of 0 (Boulangeat, Gravel, & Thuiller, 2012). As information regarding dispersal syndromes of most species was lacking, a single value of buffer distance was used for the entire tree flora (Carpenter, Read, & Jaffré, 2003). This distance should be large enough to distinguish between a true absence (the species is genuinely absent from the site due to dispersal limitation) and a false absence (the species is present but never detected due to insufficient sampling). The mean distance between nearest-neighbour presences of a species, 18.2 km, was expected to represent the average sampling effort of botanists within an evenly occupied habitat. On average, half of the validation plots were within the distance buffer of a species and half were outside, but this tendency significantly varied with species prevalence (the rarer the species, the smaller the proportion of masked validation plots (c. 15%), so that the results were little sensitive to this distance (see Appendix S2).

2.7 | S-SDMs accounting for assemblage saturation

There are several options for defining potential diversity patterns based either on occupancy probabilities or on predicted presences,
which should provide complementary insights on community saturation (Mateo et al., 2017). First, threshold-based presence/absence data are suited to represent the influence of niche preferences, as the binary prediction is designed to optimize the prediction of species presence/absence depending on environmental variables. Second, occupancy probability further reflects the prevalence of species in their habitat depending on varying dispersal and survival ability. Summing these probabilities thus represents the influence of restricted dispersal and survival ability, which limits local species richness and reflects constraints on community saturation.

We based this step on two primary assumptions: (a) environmental constraints limit the number of species that can coexist in a community (saturation hypothesis, see Mateo et al., 2017) and (b) a species that has a higher value of habitat suitability will have a greater probability of occurring in a site (D’Amen, Dubuis, et al., 2015; D’Amen, Pradervand, & Guisan, 2015; Gavish et al., 2017). Given the above assumptions, community composition was determined by ranking species in decreasing order of their predicted probability up to the local species richness prediction. This richness estimate was calculated by summing site-level probabilities of each individual species, as recommended by Calabrese et al. (2014).

2.8 Analyses of predicted assemblages

Predicted assemblage composition was compared with the composition of 12 100 m × 100 m (1 ha) independent inventories in which all trees with a diameter at breast height above the common threshold of 10 cm were identified (Table 2). The main limitations of this dataset include a weak representativeness of the c. 3,200 km² of rain forests remaining on the main island and a biased coverage of certain environments (e.g. overrepresentation of flat areas, underrepresentation of ultramafic substrates, see Appendix S3). This plot network captured half (50.3%) of the New Caledonian tree flora and a total of 872 species-plot combinations. A plot was associated with the appropriate pixel on which it was centred. A ‘true positive’ was a species both observed and predicted as being present in an assemblage, while a ‘true negative’ was a species both non-observed and predicted as being absent.

Six metrics were computed for each independent plot: (a) species richness error, that is, the difference between the modelled and observed species richness; (b) assemblage prediction success, that is, the proportion of correct predictions, including true positives and true negatives; (c) assemblage Cohen’s kappa, that is, the proportion of specific agreement; (d) assemblage specificity, that is, the proportion of true negatives over all absent species; (e) assemblage sensitivity, that is, the proportion of true positives over all present species and (d) the Jaccard index, a widely used metric of community similarity (Pottier et al., 2013). The mean prevalence of true positives was also calculated so as to detect a possible difference between S-SDMs in their ability to predict the rarest and most narrowly distributed species versus the most widespread species.

The accuracy of the four S-SDMs was first compared over all independent inventories taken together (pairwise comparison t tests if the assumption of a normal density was met, otherwise Wilcoxon signed-rank tests, followed by a Bonferroni post hoc correction). Then the variation in accuracy along the five non-categorical environmental variables (Table 1) was examined (GLM).

3 RESULTS

The SDMs showed good overall predictive accuracy (mean AUC = 0.84; mean number of SDMs with AUC >0.75 of 10
pseudoabsence selections = 6; mean TSS = 0.67). The AUC of SDMs was positively correlated with species prevalence ($r = 0.52; p < 0.001$), but this trend did not apply to the TSS ($r = -0.04; p = 0.27$). The performance of the different statistical methods was similar: the three top-ranked methods were RF (mean AUC = 0.87; mean number of kept SDMs = 8; TSS = 0.72), SVM (0.86; 8; 0.72) and GBM (0.85; 7; 0.69), while the lowest ranked methods were CTA (0.83; 6; 0.62), MAXENT (0.82; 3; 0.65) and GLM (0.82; 4; 0.65).

The most important environmental variables were elevation (mean relative increase in AUC when the predictor was included:

![Boxplot of the performance of tree species assemblage predictions derived from four stacked species distribution models (S-SDMs) among 12 forest inventory plots. 'Environment' = S-SDM accounting for environmental filtering only; '+Dispersal' accounting for environmental filtering and dispersal; '+Saturation' = accounting for environmental filtering and assemblage saturation; and 'All' = accounting for environmental filtering, dispersal and assemblage saturation. Letters (a, b, c) indicate significance between-S-SDM differences (pairwise comparison t tests if the assumption of a normal density was met, otherwise Wilcoxon signed-rank tests followed by a Bonferroni post hoc correction).]
overestimated observed richness by a factor 2–3, while stack-modelled richness estimates derived from threshold-based stacking increased logarithmically with elevation ($r^2 = 0.64$; $p < 0.05$). On average, 73 species per plot (see Appendix S4). Plot species richness increased logarithmically with elevation ($r = 0.25$; $p < 0.001$), and that of elevation remained unchanged ($r = 0.05$; $p = 0.21$). However, the importance of elevation (and no other variable) in shaping species distribution significantly varied among plant families (Kruskal–Wallis test; $p < 0.05$) with Apiaceae (mean elevation contribution = 63%), Taxaceae (58%) and Alseuosmiaceae (56%) being the most responsive to elevation and Bignoniaceae (4%), Malpighiaceae (3%) and Violaceae (2%) the least responsive.

Independent evaluation plots included between 14 species at Koumac (at 35 m elevation a.s.l.) and 95 species at Amos (493 m), with an average of 73 species per plot (see Appendix S4). Plot species richness increased logarithmically with elevation ($r = 0.64$; $p < 0.05$). The modelled richness estimates derived from threshold-based stacking methods, that is, ‘Environment’ (mean = 228 species) and ‘+Dispersal’ (163), overestimated observed richness by a factor 2–3, while stacking methods without thresholds in SDMs but with saturation, that is, ‘+Saturation’ (78) and ‘All’ (78), were significantly closer to field measurements (Appendix S4). Despite the overprediction, estimates from the four stacking methods reflected the actual gradient of biodiversity as they were all significantly correlated with observed species richness ($0.59 < r < 0.80$; $P < 0.05$; see Appendix S5).

Mean assemblage prediction success varied from 0.70 (‘Environment’; min = 0.55 at Laguen located at 572 m; max = 0.96 at Koumac situated at 35 m) to 0.87 (‘All’; min = 0.81 at Amos at 493 m; max = 0.97 again at Koumac; Figure 1a and Appendix S4). The S-SDM based on environmental filtering only (‘Environment’) performed worse at predicting communities than the S-SDM accounting for both dispersal barriers and saturation (‘All’) (Figure 1a–e). Those accounting for dispersal only (+Dispersal) and saturation only (+Saturation) performed similarly but the former had better specificity (Figure 1c), while the latter had on average better sensitivity, although more variable (Figure 1d). Irrespective of the stacking method, no difference in the number of occurrence records of true positive species was observed ($p > 0.61$), which denoted that the proportion of widespread versus rare species in the pool of correctly predicted species did not differ among the different S-SDMs (Figure 1f).

Accuracy metrics varied significantly along a single environmental gradient: elevation (Figure 2). Overall, S-SDM performance tended to decline as elevation increased. Based on environmental filtering only, this tendency was significant for two metrics: species richness error and assemblage prediction success. However, taking into account dispersal limitation and biotic filtering in S-SDMs decorrelated species richness error from elevation (Figure 2a). The difference in accuracy between the two stacking methods accounting for biotic filtering and the two neglecting biotic filtering tended to increase with elevation (GLM; $p < 0.001$; Figure 2).

**FIGURE 2** Variation in the accuracy of tree species assemblage predictions derived from four stacked species distribution models along an elevation gradient: red = ‘Environment’, that is, accounting for environmental filtering only; blue = ‘+Dispersal’, that is, accounting for environmental filtering and dispersal; green = ‘+Saturation’, that is, accounting for environmental filtering and assemblage saturation; and black = ‘All’, that is, accounting for environmental filtering, dispersal and assemblage saturation. Solid lines represent significant trends fitted with generalized linear models and dotted lines non-significant trends.

**4 | DISCUSSION**

We investigated the relative influence of environmental and biotic filtering, and dispersal on tree species assemblages in the megadiverse tropical island of New Caledonia. Our results suggest that the three ecological processes are important driving forces, but biotic filtering appears to play a more important role than dispersal limitation in shaping tree communities (Figure 1a–e). We also found that dispersal limitation does not help to improve predictions of the occurrence of rare and narrow-endemic species (Figure 1f). Moreover, the predictive ability of the alternative S-SDMs generally decreased with elevation, indicating that factors not accounted for in the models could influence biodiversity dynamics at higher elevation (Figure 2).

The New Caledonian biota exhibits a range of emblematic island attributes: for example, adaptive radiations (Barrabé et al., 2018; Pillon, Hopkins, Rigault, Jaffré, & Stacy, 2014), disharmony (Pillon et al., 2010) and ecological naïveté (Gérard, Jourdan, Millon, & Vidal, 2015). However, evidence for community-level features typical to
islands, including biota impoverishment or the vacancy of a portion of the environmental space, is lacking. Rather, New Caledonia has been shown to harbour the highest number of vascular plant species ($\gamma$-diversity) per unit area (0.17 species/km$^2$) among the biodiversity hotspots (Mittermeier et al., 2004). Local species richness ($\alpha$-diversity) is also similar to that of part of the closest continent (tropical Australia) and of neighbour large islands (e.g. New Guinea, the largest within the southern hemisphere), and stem density is among the highest reported in the southwest Pacific (Ibanez et al., 2017). Despite New Caledonia’s high species diversity, our study suggests that communities are at least in part limited by the carrying capacity of local environment due, for example, to limited available resources at local scale or geometric constraints (Calabrese et al., 2014; Gavish et al., 2017; Pouteau et al., 2015).

This was reflected by the finding that the observed $\alpha$-diversity was largely overestimated by S-SDMs accounting for biotic environment only and dispersal limitation only, while observations were well predicted by S-SDMs accounting for saturation. However, this difference in predicted species richness cannot be the only reason behind the observed difference in predicted species composition as the reported results did not vary significantly when the S-SDMs ‘Environment’ and ‘+Dispersal’ were randomized within the pool of species predicted as being present (see Appendices S6 and S7).

The method we selected to address the hierarchy of local performance (ranking the species in order of probability up to the local species richness prediction) is often associated with the assumption that a species that has a higher habitat suitability value in a given site will have better performance over another species with a lower predicted probability by being closest to optimal performance on a site (D’Amen, Dubuis, et al., 2015; D’Amen, Pradervand, et al., 2015). This assumption can be flawed for two reasons. First, habitat suitability predicted by pseudoabsence SDMs differ from true occupancy probability that only presence–absence SDMs can provide (Gomez et al., 2018). Here, thresholds used to convert continuous habitat suitability maps into binary presence/absence maps based on sensitivity and specificity were found to vary with species prevalence ($r = 0.10$; $p < 0.01$). Second, a species close to its optimal performance can be less competitive than another species further from its optimum.

A total of 22% of the New Caledonian vascular flora has been classified as narrow-range endemic species (known to occur at three or fewer locations, Wulff et al., 2013). This situation was primarily interpreted as mirroring a general difficulty for the flora to disperse in a topographically fragmented landscape, thereby promoting allopatric speciation (narrow-range endemism would thus originate from geographical vicariance). Our results reveal that dispersal limitation can play a substantial role in shaping community assembly but it did not better explain richness of low-prevalence tree species. If dispersal played a specific role in shaping the distribution of narrow-range species, the prevalence of true positive species would have been expected to decrease when accounting for dispersal, but the prevalence remained unchanged (Figure 1f).

If this outcome does not stem from an insufficient number of plots that would prevent from deciphering more subtle patterns, this would tend to support a second hypothesis to explain such richness: a variety of narrowly distributed environments (e.g. unique combinations of elevation, substrate and topographically driven climate) could promote ecological vicariance without the influence of dispersal limitation (Kozak & Wiens, 2006). The hypothesis is supported by a previous study showing that the distribution of hotspots for narrow-range endemic plant species in New Caledonia can be predicted solely on the basis of abiotic descriptors (using maxent modelling, Wulff et al., 2013). A more recent study on animal species native to New Caledonia showed that there is a higher proportion of species with restricted ranges ($\leq$5.2 km$^2$) in non-endemic species (22%) than in endemic species (12%), suggesting that environmental dissimilarity may play an important role in their distribution (Caesar, Grandcolas, & Pellens, 2017). The role of environmental filtering in shaping island communities is gaining increasing support in the literature, while recent biogeographical and phylogenetical analyses in the Galápagos and Macaronesia have provided increasing evidence against the hypothesis of lesser dispersal ability in islands (Carvajal-Endara, Hendry, Hendry, Emery, & Davies, 2017; García-Verdugo, Mairal, Monroy, Sajeva, & Caujapé-Castells, 2017).

Ecological vicariance on the New Caledonian archipelago may be boosted by high climatic diversity over short distances typical of oceanic islands due to orographic cloud formation and precipitation, foehn winds and other topographically determined climatic patterns resulting in heterogeneous distributions of microclimates, which contrast with the relatively buffered climates of large continental areas (Whittaker & Fernández-Palacios, 2007). Similarly, topographically driven climate has been found to be a major driver explaining spatial patterns of plant species diversity and endemism in Canary Islands (Irl et al., 2015). In addition to topoclimate, the distribution of plant species in New Caledonia is recognized as being constrained by a mosaic of different substrates, including New Caledonia’s unusual ultramafic substrates (Morat, 1993; Pillon et al., 2010), which provide small amounts of essential plant nutrients and are rich in heavy metals that are toxic to most species (Morat, 1993). Edaphic factors have also been shown to vary over short distances (e.g. on alluvium vs. on slopes) and to greatly influence tree communities (Jaffré & Veillon, 1991). Such a diversity of habitats is thus likely to have great influence on community composition.

Another key finding of this study was that the predictive power of the three filters examined together tended to decrease along the elevation gradient, suggesting an increasing role played by stochastic factors as elevation increases (Figure 2). This is the opposite pattern from what was reported in previous research focusing on both plants and animals along a c. 1,200 m elevation gradient in Japanese islands (Mori et al., 2013). The underlying reason is probably that rather than depending directly on the elevation gradient per se, the relative importance of deterministic versus stochastic factors arises from a productivity or energy gradient. In temperate regions, productivity is often lower at higher elevations due to lower temperatures while, in New Caledonia, productivity is higher at higher elevation due to lower aridity and disturbance intensity (at least in the range considered in this research, i.e. 35–759 m, but this pattern
is likely to be reversed from mid to high elevation, see Birnbaum et al., 2015; Pouteau et al., 2015). This productivity gradient was reflected by the significant positive correlation we found between observed $\alpha$-diversity and elevation, which suggests increasing overlap of species abiotic tolerance (see S-SDM ‘Environment’ in Figure 1) and increasing stochastic processes (e.g. chance colonization, random extinction) as elevation increases (Chase, 2010).

Saturation is known to decline with stress (Bertness & Callaway, 1994). As a result, while a maximum number of species find optimal climatic conditions at mid-elevation, we can expect species to become fewer and more specialised (e.g. reduced photosynthesis, deep root systems, increased wood density, low abundance of palm trees, gymnosperms and mosses) and communities less saturated in water-limited environments such as the leeward coastal plain of New Caledonia’s main island. These expectations regarding ecological patterns are likely to explain why the four stacking methods performed similarly at predicting observed communities in lowlands (e.g. in Koumac) while S-SDMs accounting for saturation performed much better than the two saturation-independent S-SDMs at higher elevation (Figure 2).

This study is the first attempt to disentangle the processes controlling community assembly in the island tropical biodiversity hotspot of New Caledonia. We found many similarities (important role of dispersal and differences in performance, increasing role of neutral processes with productivity) between New Caledonian and other tropical rain forests, suggesting similar community functioning. These similarities challenge several of Carlquist’s (1974) intuitions with regard to island communities, but structuring processes may diverge on smaller and more remote oceanic islands.

This study is also the first one analysing community assembly processes based on the SESAM framework in the tropics. We found the framework extremely useful from a theoretical perspective and relatively easy to implement, but we have also stressed several major limitations, including in the way to infer processes from spatial patterns (Warren, Cardillo, Rosauer, & Bolinck, 2014). For instance, individual SDMs are used to perform abiotic environmental filtering by correlating species occurrences with environmental predictor variables (Guisan & Rahbek, 2011). However, occurrences already result from competitive exclusion, dispersal limitation and stochastic processes, particularly in tropical rain forests (Kraft & Ackerly, 2010), which makes disentangling the driving processes difficult and might skew the conclusions. We therefore encourage further tests of other recently developed community models (e.g. Pollock et al., 2014) in megadiverse tropical rain forests.

ACKNOWLEDGEMENTS

We thank the three anonymous referees for their careful reading of the manuscript and suggestions. We also thank Wilfried Thuiller (Laboratoire d’Écologie Alpine, CNRS, Université Grenoble Alpes, France) for his insightful comments on an earlier version of this manuscript, and Dimitri Justeau (IAC) for his assistance in using the compute-optimized server on which S-SDMs were generated.

We are grateful to all people (colleagues, students, guides, rangers and many others) who contributed to the collection of field data, to Jérôme Munzinger (IRD, UMR AMAP) who initiated the implementation of the NC-PIPPN network and to the NOU Herbarium curator Jacqueline Fambart-Tinel. This study is part of the COGEFOR project jointly funded by the Direction du Développement Économique et de l’Environnement (DDEE) of the North Province of New Caledonia (convention 16C180), the Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) and the Institut Agronomique néo-Calédonien (IAC).

AUTHOR CONTRIBUTIONS

R.P. designed and performed the research. P.B. computed the occurrence database. R.P. and F.M. discussed the results. R.P. wrote the manuscript. F.M. and P.B. provided comments.

ORCID

Robin Pouteau https://orcid.org/0000-0003-3090-6551
François Munoz https://orcid.org/0000-0001-8776-4705
Philippe Birnbaum https://orcid.org/0000-0003-4830-5500

REFERENCES


**Supporting Information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Pouteau R, Munoz F, Birnbaum P. Disentangling the processes driving tree community assembly in a tropical biodiversity hotspot (New Caledonia). J Biogeogr. 2019:00:1-11. [https://doi.org/10.1111/jbi.13535](https://doi.org/10.1111/jbi.13535)