From Temperate to Boreal – Untangling the Importance of Environment, Space, and Species Co-distributions in Forest Plant Distributions

Gabrielle Bibeau Lapierre[1*], Marilou Cournoyer[1], and F. Guillaume Blanchet[1,2,3]

[1]Département de biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada

[2]Département de mathématiques, Université de Sherbrooke, Sherbrooke, Québec, Canada

[3]Département des sciences de la santé communautaires, Université de Sherbrooke, Sherbrooke, Québec, Canada

[*]Address correspondence to: Gabrielle.Bibeau@USherbrooke.ca

Keywords

Community assembly, Environmental filtering, Forest ecology, Plant distributions, Regional-scale, Spatial autocorrelation, Species co-occurrence, Variation partitioning.

Abstract

Aim

Quantify how environmental conditions, spatial autocorrelation, and species co-distributions shape forest plant community at regional scale.

Location

Southern Québec, Canada, between the 45^{th} and 53^{rd} parallels.

Period

From 1986 to 2000.

Major taxa studied

Vascular plants, mosses, and lichens.

Methods

We analyzed the presence-absence of 224 plant species across 28,210 ecological observation points using a multivariate hierarchical model. Environment was characterized by bioclimatic, edaphic, and topographic variables, while regional scale spatial structure was defined by distance-based Moran's eigenvector maps. Sampling unit level latent variables were used to characterize residual species co-distributions. Model parameters were estimated using Bayesian inference through Markov chain Monte Carlo (MCMC) and were studied using ordinations (a principal component analysis calculated on the parameter matrix) for environment and spatial autocorrelation or a correlation matrix to quantify species co-distributions. We then quantified the relative contributions of environmental conditions, spatial structure, and species co-distributions to the community using variation partitioning.

Results

Environment, spatial autocorrelation, and species co-distributions all contributed substantially to plant distribution patterns although broad scale spatial autocorrelation contributed slightly more to structuring the community. Environmental responses varied considerably among species, with drainage emerging as an important differentiation component. Residual species co-distributions were predominantly positive (56.7%), although these patterns may reflect unmeasured environmental conditions or ecological processes involving species co-distributions. Using, ordinations we could also quantify uncertainty in the parameter estimated for the environment and spatial autocorrelation. Using the ordination results, we highlighted the species for which we had the most confidence in the results and notably found a gradient in confidence (or uncertainty) in the species spatial autocorrelation.

Main conclusions

Environmental conditions, spatial autocorrelation, and species co-distributions all contributed to a similar extent in explaining regional patterns of plant distribution, with regional-scale spatial structure having the largest overall contribution. This study highlights the importance of taking a multifactorial perspective to better understand how vascular plants, mosses, and lichens are structured on the landscape.

1 Introduction

A community refers to a group of interacting species that occur in a given place at a given time [1]. This definition is subjective, as the species included vary depending on the scale considered [2]. In community ecology, understanding patterns of species distribution and the processes underlying them is a fundamental goal [3]. Selection, drift, dispersal and speciation are the four key processes that have been identified as the main drivers of community assembly [4, 5]. Briefly, selection favors species with traits that confer survival and reproductive advantages under the influence of particular abiotic or biotic pressures. It acts deterministically. For example, in boreal forests, more acidic and nutrient-poor conditions favor conifers over deciduous trees [6], illustrating an abiotic selective pressure. In plants, competition for nutrients and light is an example of biotic selective pressure [7]. Drift alters the relative abundances of species. It is a stochastic process. For instance, births and deaths are events affecting drift. Dispersal refers to the establishment distance or movement abilities of species and has high interspecific variations. It can be related to spatial autocorrelation, as a species that has short dispersal will be spatially correlated with itself. For example, *Tsuga canadensis*, the eastern hemlock, can have patchy spatial structure, as one adult hemlock will create favorable conditions for the establishment of seedlings, resulting in positive spatial autocorrelation in the species' distribution. Furthermore, the processes shaping community structure and dynamics are influenced by the spatial and temporal scales considered [8, 9].

Studies on species distribution have long failed to consider species associations [10]. However, the importance of these associations for species distribution has been illustrated across spatial scales [11]. It is important to note that species co-distributions are not necessarily direct indicators of biotic interactions. Although co-distribution patterns may suggest ecological relationships, they can also arise from shared environmental preferences, unmeasured abiotic factors, or stochastic processes such as ecological drift [12]. Therefore, interpreting co-distribution requires caution, as they may reflect indirect or confounding influences rather than true biological interactions. A noteworthy example when studying plant distribution is the importance of edaphic variables that, until recently, have commonly been neglected, leading to misinterpretation in the importance of interaction in structuring plant distributions [13, 14].

In the present study, our objective was to evaluate the determinants of the distribution of forest plant species at a regional scale by accounting for abiotic environmental factors, spatial autocorre-

lation, and species co-distribution. Speciation was excluded from this study because it operates at a temporal scale too broad to be measured with the available data. For this study, we focused on plant species found in Québec (Canada), south of the 53rd latitude.

Our first hypothesis proposed that the environment (abiotic factors) plays a predominant role in the distribution of plant communities at the regional scale. This hypothesis is based on previous work suggesting that climate, specifically temperature and precipitations, are the main drivers of plant community structure in Québec [15, 16] but also more generally [17–19]. At the regional scale, we expect environmental selection to explain a greater proportion of variation in species distributions than spatial autocorrelation for most species. Important environmental variables, such as temperature and precipitation, exhibit broad spatial gradients across our study area [16]. For this reason, we expect the environment to account for a substantial proportion of the spatial variation in species distributions, reducing the relative contribution of residual spatial structure. As a last hypothesis, we proposed that species co-distributions has little importance at regional scale. Species co-distributions tend to be more important in communities characterized by fewer species and reasonably narrow environmental gradients, conditions that contrast with those of our study area [20]. Moreover, broad spatial scales may be necessary to fully capture community clusters, suggesting that the regional extent of our study may influence our perception of the assemblage [21].

2 Materials and Methods

2.1 Data

The Ecological Observation Points (*Point d'observations écologique* in French) is an ecological inventory program carried out in the forests of southern Québec from May to October each year between 1986 and 2000. It is the main source of data used in the present study. The study area covered 760,000 km^2 between the 45th and 53rd latitude and between the 57th and 80th meridians. In all, it includes 28,425 ecological observation points. Each point consisted of a circular plot of 400 m^2 (radius of 11.28 m) in which information on vegetation, soil, and topography was collected. In total, 531 species of moss, lichen, herbaceous plants, shrub, and tree were recorded (see Appendix 4 in Supporting information). The ecological observation points were distributed along routes, each of which contained five to seven points. The number of points per route varied depending on the complexity of the terrain to be sampled. The ecological observation points were generally located

at intervals of 15 km^2 in deciduous forest, 20 km^2 in mixed forest and balsam fir–white birch forest, and 25 km^2 in boreal forest. Inventory routes were distributed according to a sampling plan that covers the characteristic relief of each of the ecological districts inventoried. An ecological district is the basic unit of Québec's ecological land classification, defined as an area of $\sim$ 100 km^2 with a unique relief, geology, geomorphology and vegetation [22]. Each point was sampled only once and points belonging to the same route were generally sampled on the same day. The technical standards for this inventory are detailed by [23].

In the present study, we considered 28,210 ecological observation points, choosing only the points for which all necessary data for the study were accessible. Specifically, 186 ecological points did not have soil organic matter information and two were not surveyed for vegetation. We focused on the presence and absence of species at each point. Since the plant species inventoried at each point represented the plant community present at the time of sampling, there may be a bias in the representativeness of species only occurring very early (spring) or very late (autumn) in the sampling season. Moreover, there may be a taxonomic bias due to identification difficulties; notably grasses and sedges had fewer mentions. To reduce the bias associated with species that appear early or late in the sampling season, we focused on species with at least 100 observations, which reduced the number of plant species to 224. Although we are aware that this does not fully correct for seasonal sampling bias, it at least partially accounts for extreme cases. As mentioned in the previous paragraph, the data collection spans 14 years. However, for the purpose of this study we assumed that environmental variations were minimal during this period.

To characterize the distribution of plants, we relied on 19 bioclimatic variables gathered from *WorldClim* historical data (1970-2000) at 30 seconds precision associated with each of the ecological observation points (See Appendix 3). To assess collinearity among environmental variables, we used the variance inflation factor (VIF) [24]. In our model, we only included explanatory variables with a VIF < 3 and encoded the categorical variables into polynomial contrast variables. For interpretation purposes, we performed a principal component analysis [25] on all environmental variables to assess the ecological importance of explanatory variables excluded from the model [26]. Figure 1 presents the principal component analysis we used to interpret the model results.

Additionally, distance-based Moran's eigenvector maps (dbMEM, [27]) were constructed to account for the spatial autocorrelation across the region. The dbMEM were constructed using a truncation distance of 144,455.3 meters. It is the minimum distance that link all ecological obser-

vation points together. Only the first 11 dbMEMs were used in the analysis as they were the ones with a Moran's I [28] larger than 0.5. The construction of dbMEM was carried out using the dbmem function and the Moran's I were calculated using the moran.randtest function, both implemented in the adespatial [29] R package. Note that, by definition, each spatial variable captures a unique independent spatial structure. We chose to focus on the first 11 dbMEMs, which are designed to capture broad scale ecological patterns because our aim was to study plant distribution at a regional scale.

2.2 Statistical Analysis

We relied on a multivariate hierarchical model implemented in the HMSC R package [30] to perform the statistical analysis. Since species presence-absence were considered, we used a probit link function to perform the analyses. To run the analyses, we had to create three matrices. The multivariate hierarchical model focuses on explaining the community matrix ($\mathbf{Y}$ of n points and p species) encompassing the presence-absence of plant species across all ecological observation points with a matrix of explanatory variables ($\mathbf{X}$ of n points and q explanatory variables) that contains the standardized edaphic and bioclimatic variables (blue variables in Figure 1) referred to as $\mathbf{X}^{\text{ENV}}$ in Equation 1 and the 11 spatial variables (dbMEM) referred to as $\mathbf{X}^{\text{SPAT}}$ in Equation 1. In addition, we accounted for a sampling unit level random effect. To perform the analysis, we followed the procedure described by Ovaskainen, Tikhonov, Norberg, et al. [31] and Ovaskainen and Abrego [32].

The full model is described as follows

$$\text{Probit}(\mathbf{Y}_{i,j} = 1) = \mathbf{X}^{\text{ENV}}_{i,k}\mathbf{B}^{\text{ENV}}_{k,j} + \mathbf{X}_{i,l}\mathbf{B}^{\text{SPAT}}_{l,j} + \mathbf{H}_{i,m}\mathbf{\Lambda}_{m,j} \tag{1}$$

where

$$\mathbf{B}^{\text{ENV}} \sim \mathcal{N}(\boldsymbol{\mu}^{\text{ENV}}, \mathbf{V}^{\text{ENV}}), \tag{2}$$

and

$$\mathbf{B}^{\text{SPAT}} \sim \mathcal{N}(\boldsymbol{\mu}^{\text{SPAT}}, \mathbf{V}^{\text{SPAT}}). \tag{3}$$

In words, $\mathbf{B}^{ENV}$ is a matrix of regression coefficients for the environmental variables across all species, $\mathbf{B}^{SPAT}$ is a matrix of regression coefficients associated to the selected spatial variables across the species, while $\mathbf{H}$ and $\mathbf{\Lambda}$ are, respectively, latent variables and their associated parameters

to account for sampling level random effect and thus assess species co-distributions. The estimation of the model's parameters (the $\mathbf{B}$'s in equation 1), of the latent variables and their associated parameter was carried out through Bayesian inference via a Markov Chain Monte Carlo (MCMC). The selection of the optimal number of latent variables to use relied on the procedure proposed by Bhattacharya and Dunson [33].

The model's priors we used were weakly informative. Specifically, $\boldsymbol{\mu}^{\mathrm{ENV}}$ and $\boldsymbol{\mu}^{\mathrm{SPAT}}$ were assumed to follow a multivariate normal distribution with a mean of 0 and a covariance that follows an inverse-Wishart distribution with $q+2$ degrees of freedom and a $(q+1)\times(q+1)$ identity matrix for the scale matrix. The covariance matrices $\mathbf{V}^{\mathrm{ENV}}$ and $\mathbf{V}^{\mathrm{SPAT}}$ were also assumed to follow an inverse-Wishart distribution with $q+2$ degrees of freedom and an $(q+1)\times(q+1)$ identity matrix for the scale matrix. Latent variables were assigned hierarchical shrinkage priors that increasingly penalizes higher-order factors, enabling automatic dimensionality selection and reducing overfitting [33].

To estimate the parameters, we ran the MCMC for a total of 15,000 iterations, including a burn-in phase of 10,000 iterations. To assess the convergence, we calculated the effective sample size [34] for each estimated parameters, which represents the independent samples in the chain.

To assess the importance of environment, space and co-distribution for the plant community, we performed a variation partitioning analysis [35, 36] where we studied how much variation is explained by each fraction for each species. The amount of variation explained was calculated using [37] R^2 specifically designed for probit models. In addition, following [36], all R^2 were adjusted to account for the number of parameters estimated for each sub-model estimated to calculate the amount of variation explained by the different fractions of the variation partitioning analysis. Because latent variables are very efficient in capturing the amount of variation in the data, we relied on the procedure proposed by [38] to compute each fraction in the variation partitioning analysis. Again relying on [38], we also calculated the contribution of each point to each fraction of the variation partitioning analysis.

All analyzes were carried out with the statistical language R [39].

3 Results

The effective sample size for all estimated model parameters are presented and briefly discussed in Appendix 1.

To visualize the combined effect of the environment, space, and species co-distributions (latent variables) on the distribution of the 224 plant species considered, we generated two ternary diagrams following the approach proposed [38] (Figure 2). This result reveals that neither environment, space nor co-distribution is a dominant factor structuring distribution of the species or the contribution of the points to the community. Indeed, among the proportion of variation associated to the explanatory variables included in the model, the environment accounts for 26.9% of the variation, while space and species co-distributions explain respectively 38.8% and 34.4%. As for the points' contribution to the explained variation, environment, space and co-distributions respectively accounts for 27.1%, 37.9% and 34.9% of the variation. The high relative density located roughly in the middle of the ternary plots further supports the importance of all three components structuring the plant community. However, the points' contribution to the plant community (Figure 2b) shows that space is slightly more important than the other factors, suggesting that either plants present an inherent spatial structure or that some important environmental variable that is spatially structured was not considered but should have been. Focusing on the species ternary plot, the variation explained by the model is generally reasonably high across the plant species (mean adjusted R^2 across all the species of 0.626) considered, with *Pyrola sp.* (PYS) presenting the lowest adjusted R^2 (0.147) and *Tilia americana* (TIL) the highest (adjusted $R^2 = 0.944$).

3.1 Environment

To better understand the effect of the environment for plant distribution, we carried out a principal component analysis (PCA) on the estimated species by environmental parameters matrix (i.e. $\mathbf{B}^{\text{ENV}}$). Since the model parameters were estimated with an MCMC, we carried out the PCA on every iteration of the estimated species by environmental parameters matrix and performed an orthogonal Procrustes rotation ([40]; Chapter 4) to align the results of each PCA (Figure 3). As such, each species polygon area represents the variation between the iterations. For instance, in the upper left corner, *Comptonia peregrina* (COP) has a higher variance between iterations value than *Pinus banksiana* (PIG). From this presentation, ecologically relevant clusters can be identified. For example, in the upper-right corner, a cluster of species including *Cassandra calyculata* (CAL), *Rhododendron groenlandicum* (LEG), *Kalmia polifolia* (KAP), and *Vaccinium oxycoccos* (VAO) can be identified. This cluster is characterized by herbaceous species typically associated with humid, acidic soils such as those found in bogs. These habitats are often dominated by the tree species *Picea*

mariana (EPN), which is unsurprisingly located close to this cluster. In contrast, a few species, such as *Tilia americana* (TIL) in the bottom-left corner, appear to have relatively little in common with the other species in terms of their environmental preferences. Most species, however, are clustered near the center of the ordination, reinforcing the idea that most of the species richness in our study region is associated with temperate forests, which are also the most represented in our data.

3.2 Space

The species by spatial variables parameters matrix (i.e. $\mathbf{B}^{\text{SPAT}}$) estimated with the multivariate hierarchical model was studied using the exact same procedure as the species by environment parameters matrix and the results are presented in Figure 4. A few pairs of species, such as *Viburnum lantanoides* (VIL) and *Betula alleghaniensis* (BOJ) on the left side of the plot, have a similar spatial autocorrelations structure across southern Québec. In contrast, some species, such as *Acer spicatum* (ERE), also located on the left side of the plot, appear more isolated, suggesting a distinct spatial structure that is not shared with other species at the regional scale, at least based on the 11 dbMEMs selected for this analysis. The PCA also highlights a notable pattern. For the species on the top left side of the plot, we have a very well defined spatial autocorrelation and the model described them with a high level of confidence. As we move towards the bottom right side of the plot, species tends to be more difficult to estimate. This gradient is associated with several spatial variables, particularly dbMEMs 4-7. The spatial scales represented by these dbMEMs may contribute to the observed variability in the estimated species coefficients, particularly for species with geographically restricted distributions. For instance, at opposite ends of this gradient, *Viburnum cassinoides* (VIC), which occurs throughout much of the study region, shows relatively low variability, whereas *Fraxinus americana* (FRA), which occurs primarily in southern Québec near its northern range limit, shows substantially greater variability. This pattern suggests that regional-scale spatial effects may be estimated with greater uncertainty for species whose distributions are restricted to only part of the study region. Additionally, most species are clustered near the center of the PCA, indicating broadly similar spatial autocorrelations across southern Québec.

3.3 Species Co-distribution

To assess plant co-distributions, we relied on a species correlation matrix (Figure 5) calculated from the estimated species by latent parameters matrix (i.e. $\mathbf{\Lambda}$). Although some species pairs show random correlations (represented in black in Figure 5), most are characterized by either negative (red pixels) or positive (blue pixels) correlations. Of these correlations, 56.7% of all possible species pairs were estimated as positive, whereas 29.4% were estimated as negative. The prevalence of positive correlations suggests that, after accounting for the environmental and spatial variables included in the model, many species still exhibited similar distributional patterns. However, the proportion of positive correlations varied substantially among species. For instance, *Trillium grandiflorum* (TRG) had the lowest proportion of positive correlations with other species (19.6%), whereas *Eurybia macrophylla* (ASM) had the highest (78.1%). These differences may partly reflect variation in species' geographic distributions and ecological niches. For example, the relatively restricted southern distribution of *T. grandiflorum* contrasts with the broader distribution of *E. macrophylla* across the study region. Importantly, these residual co-distributions should not be interpreted as direct evidence of biotic interactions. Rather, they represent co-distribution patterns remaining after accounting for the environmental and spatial predictors included in the model and may therefore also reflect responses to unmeasured environmental variables as well as other unaccounted ecological processes.

4 Discussion

The results presented in Figure 2 suggest that environment, spatial autocorrelation, and species co-distributions have similar importance in structuring the plant communities at regional scale both for species structure and points' contribution. More specifically, both species and points appear to be slightly more influenced by spatial autocorrelation than by species co-distributions, with environmental effects representing a lesser contribution, even if the contribution of the environment is still notable. These results contrast with our initial hypotheses, which predicted that environmental conditions would play a predominant role, and thus that spatial autocorrelation would be of less importance, and that species co-distributions would have the weakest influence at the regional scale.

4.1 Environmental effect

Although environmental conditions had the smallest overall contribution when all species were considered simultaneously, some species remained strongly structured by the environment. Notably, environmental variables alone accounted for 56.7% and 55.4% of the explained variation for *Picea rubens* (EPR) and *Carya cordiformis* (CAC), respectively (Figure 2). Both species have relatively restricted distributions within the study area, suggesting that their occurrence in southern Québec may be particularly constrained by a specific combination of environmental conditions.

Among the environmental variables represented in Figure 3, drainage and altitude were particularly important in structuring differences among species. Their influence is especially apparent for the distinct cluster formed by *Picea mariana* (EPN), *Empetrum nigrum* (EMN), *Cassandra calyculata* (CAL), *Rhododendron groenlandicum* (LEG), *Kalmia polifolia* (KAP), and *Vaccinium oxycoccos* (VAO). Interestingly, several species in this cluster belong to the Ericaceae and share functional adaptations to relatively harsh environmental conditions, including tough, often evergreen and waxy leaves [41]. Such traits can contribute to resource conservation and persistence under stressful conditions. Consistent with these adaptations, these species are commonly associated with environmentally restrictive habitats, including poorly drained and nutrient-poor environments such as bogs [42] and, for some species, higher-elevation habitats. Their positions in the ordination along gradients associated with drainage and altitude are therefore ecologically coherent with their shared adaptations. This clustering provides an example of how environmental filtering can bring together species with similar ecological strategies and environmental requirements, even when species identity itself is not considered.

Most species, however, were concentrated near the center of the environmental ordination, suggesting that their distributions are associated with a relatively similar range of environmental conditions. The densest portion of this central cluster appears particularly aligned with drainage and precipitation of the driest month (Bioclimatic variable 14). Both variables are closely related to water availability, an important constraint on plant establishment, growth, and persistence [43]. The concentration of species along these gradients therefore suggests that water availability may represent an important environmental axis structuring plant distributions across southern Québec. Furthermore, the large number of species occupying this region of environmental space suggests that environmental filtering alone may be insufficient to explain their local distributions, leaving greater potential for

spatial processes and species co-distributions to contribute to community structure. Nevertheless, these patterns represent realized distributions rather than the full range of environmental conditions that each species could potentially tolerate. Consequently, the environmental associations identified here should not necessarily be interpreted as species' physiological niche limits, particularly for boreal species whose realized distributions may occupy only part of their potential environmental niche [44].

Overall, these results demonstrate that the relative importance of environmental filtering varies considerably among species. While environmental conditions strongly constrain the distributions of some species and generate distinct groups of species with similar environmental responses, many species occupy broadly similar environmental space. Importantly, the prominence of drainage in these patterns reinforces the value of incorporating edaphic variables into studies of plant distributions rather than relying exclusively on climatic predictors, consistent with Ni and Vellend [13] and Dubuis, Giovanettina, Pellissier, Pottier, Vittoz, and Guisan [45].

4.2 Spatial effect

The predominant contribution of space to plant distributions suggests that regional-scale spatial autocorrelation plays an important role in organizing plant communities across southern Québec. However, part of this spatial signal may reflect environmental gradients that are themselves spatially structured. This is particularly relevant in Québec, where temperature varies strongly along a south-to-north latitudinal gradient [15, 16], while precipitation generally increases from west to east [16]. Consequently, although environmental and spatial effects were separated in our model and in the ternary plot (Figure 2), their ecological interpretation is not entirely independent.

Nevertheless, the contribution to the explanation of plant distribution, even though inconsistent in its autocorrelation (from 8.8% to 73.9% of explained variation by dbMEMs across species), indicates that their distributions are somewhat well captured by the regional-scale spatial structures represented by the selected dbMEMs. One possible explanation is that plant species do not necessarily occupy all environmentally suitable habitats across the study region. Postglacial colonization history, differences in dispersal abilities, and historical range expansion may all contribute to contemporary spatial structure [44]. Consequently, some portion of the spatial signal detected by the dbMEMs may reflect historical and biogeographic processes rather than present-day environmental filtering alone.

Similarities among species in their spatial structure are further illustrated by Figure 4. For instance, a large proportion of species are concentrated near the center of the ordination plot. Rather than indicating high spatial autocorrelation among these species, their clustering suggests that many species respond similarly to the regional spatial gradients represented by the selected dbMEMs. This result is consistent with the existence of broad geographical structures shared by a substantial proportion of the southern Québec flora, while species located toward the periphery of the ordination exhibit more distinctive spatial patterns.

Another notable pattern in Figure 4 is a gradient in the uncertainty of the estimated spatial responses, extending across the ordination and associated particularly with dbMEMs 4-7. Species at one end of this gradient tend to have more precisely estimated spatial responses, whereas species at the opposite end show substantially greater uncertainty. One possible explanation is that geographically restricted species provide less information for estimating the regional-scale spatial patterns represented by the selected dbMEMs. When a species occurs primarily within a limited portion of the study region, distinguishing its spatial autocorrelation from the geographic restriction of its distribution may become more difficult. Consequently, regional-scale spatial effects may be estimated with greater uncertainty for locally restricted species than for species distributed throughout a larger proportion of the study area. This limitation is not necessarily specific to our modeling framework and may be relevant more generally when spatial models are used to characterize the distributions of geographically restricted species.

In general, space emerged as the largest of the three components considered in our analysis, but its ecological interpretation remains inherently complex. Spatial autocorrelation can arise from species-specific processes such as dispersal and historical range dynamics, while also capturing the effects of spatially structured environmental conditions, including potentially important variables absent from the model. Thus, our results provide strong evidence that regional-scale spatial structure is important for understanding plant distributions, but they do not attribute this spatial signal to a single ecological mechanism.

4.3 Species co-distribution effect

Species co-distributions contributed substantially to the distribution of plant species (Figure 2) in our study, although their overall contribution was slightly lower than that of spatial autocorrelation. The latent component represents residual variation in species distributions that remains

after accounting for the environmental and spatial variables included in the model. Such residual co-distributions may arise from several ecological processes, including biotic interactions, responses to unmeasured environmental variables, unaccounted stochastic processes such as ecological drift [12, 32] or spatial autocorrelation at a scale finer than either of the 11 dbMEMs variables considered can capture. Species varied from having 79.2% of the total explained variation defined by co-distribution (*Gymnocarpium dryopteris*) to almost not at all (*Adiantum pedatum*). High contributions may partly reflect the comparatively smaller proportion of their distributions explained by the environmental and spatial variables considered in the model. In other words, important determinants of their occurrence may not have been represented by our measured environmental variables or by the regional-scale spatial structures captured by the selected dbMEMs.

Looking at Figure 5, the species correlation matrix contains a heterogeneous mixture of positive (56.7% of all species pairs) and negative (29.4%) with a reasonably small fraction of species pairs presenting negligible co-distribution structure (13.9%). The latter is quite interesting because it highlights the important signal related to co-distribution. Some of these species pairs were related to plants with a very small part of their distribution explained by latent variables (Figure 2). Negative residual co-distributions could potentially reflect competitive interactions or contrasting habitat requirements among species. However, because residual co-distributions may also emerge from omitted environmental variables or imperfectly captured spatial structure, they cannot be interpreted as direct evidence of competition. While positive residual co-distributions share the same limits, their predominance suggests that many species share similar distributional patterns. These interpretations are consistent with the idea that community assembly at regional scales may be influenced by multiple processes acting simultaneously, rather than solely by direct competitive interactions.

In summary, our results indicate that species co-distributions represent an important component of regional plant distribution patterns, contrary to our initial hypothesis that their contribution would be relatively weak at this spatial scale. However, the latent component should not be interpreted as direct evidence of biotic interactions. Instead, it identifies residual structure that is not explained by the measured environment or regional-scale spatial variables and therefore highlights potentially important ecological processes or environmental factors that remain unaccounted for. Further investigation of these residual co-distributions, particularly through the inclusion of additional explanatory variables, would help gain a deeper understanding of the underlying mechanisms

structuring the distribution of these species.

4.4 Uncertainty in parameter estimation

While the relative contributions of environmental filtering, spatial structure, and residual species co-distributions provide insight into the processes associated with plant distributions, an equally important consideration is the degree of certainty associated with these estimates. Across species, the total variation explained by the model varied substantially, ranging from an adjusted R^2 of 0.147 for *Pyrola sp.* to 0.944 for *Tilia americana* (Figure 2). This wide range indicates that the explanatory variables and latent structure included in the model captured species distributions with very wide range of confidence. Furthermore, species with a similar total amount of explained variation could differ markedly in the relative importance of environmental, spatial, and co-distribution components, indicating that predictive performance alone provides only a partial assessment of model certainty.

Additional information regarding parameter uncertainty is provided by, when presenting the results as ordinations as we did, using a polygon of confidence around each species (Figures 3 and 4). Because the principal component analyses were performed separately on each retained MCMC iteration and subsequently aligned using an orthogonal Procrustes rotation, the polygon associated with each species summarizes the variability of its ordination position across the posterior distribution. Smaller polygons therefore indicate greater agreement among MCMC iterations and more precise parameter estimates, whereas larger polygons indicate greater posterior uncertainty. Interestingly, in Figure 4, uncertainty seemed correlated with dbMEMs 4–7. From the upper left corner of the ordination to its lower right corner, polygons went from smaller to wider. Studying the species at each end, those with geographically restricted distributions generally tended to show greater uncertainty in their estimated spatial responses, likely because fewer observations were available to characterize their association with broad-scale spatial structures. This result is directly related to the framing of our study that focuses on regional level species distribution.

Together, these results highlight the importance of including a measure of uncertainty in studies such as ours. Although thresholding, such as what we did in Figure 5, brings valuable ecological insights, including uncertainty in ordination plots as we did, brings more detailed information that influences the interpretation of results. In our study, this way of accounting for parameter uncertainty highlighted that although the overall importance of environment, space, and species co-distributions was robust at the community level, confidence in the estimated responses varied considerably among

species, reflecting differences in the amount of information available to estimate species-specific parameters.

5 Conclusion

In conclusion, our study illustrates that plant distributions across southern Québec cannot be attributed to a single ecological component. Contrary to our initial hypotheses, environmental conditions, spatial autocorrelation, and species co-distributions all contributed substantially to explaining regional patterns of plant distribution, with spatial autocorrelation representing the largest overall component. These results suggest that environmental filtering alone is insufficient to explain plant distributions at regional scale and highlight the importance of considering spatial and residual community structure simultaneously. The prominence of drainage among the environmental predictors further emphasizes the value of incorporating edaphic variables alongside climatic predictors when studying plant distributions. Including additional ecologically relevant environmental variables, such as soil pH, nutrient availability, or disturbance history, could further improve model performance and help determine whether residual co-distributions reflect unmeasured environmental filtering, historical processes, dispersal limitation, or other ecological mechanisms. Overall, our results underscore the complexity of plant community assembly and illustrate that environmental, spatial, and residual co-distribution structures must be considered jointly to better understand the processes shaping plant communities at regional scale such as for a region of the size of Québec.

References

1. Mittelbach GG and McGill BJ. Community ecology. 2nd ed. Oxford: Oxford university press, 2019.
2. Leibold MA and Chase JM. 1. Introduction: The Rise, Fall, and Rise Again of Metacommunity Ecology. In: *Metacommunity Ecology*. Princeton University Press, 2017:1–22.
3. Morin PJ. Community Ecology. 2nd. Chichester, West Sussex, UK: Blackwell Science, 2011. DOI: 10.1002/9781444341966.
4. Vellend M. The Theory of Ecological Communities. Monographs in Population Biology 57. Princeton University Press, 2016.

5. Vellend M. Conceptual Synthesis in Community Ecology. The Quarterly Review of Biology 2010;85:183–206.

6. Carteron A, Parasquive V, Blanchard F, et al. Soil Abiotic and Biotic Properties Constrain the Establishment of a Dominant Temperate Tree into Boreal Forests. Journal of Ecology 2020;108:931–44.

7. Bengtsson J, Fagerström T, and Rydin H. Competition and Coexistence in Plant Communities. Trends in Ecology & Evolution 1994;9:246–50.

8. Chase JM, McGill BJ, Thompson PL, et al. Species Richness Change across Spatial Scales. Oikos 2019;128:1079–91.

9. Viana DS and Chase JM. Spatial Scale Modulates the Inference of Metacommunity Assembly Processes. Ecology 2019;100:e02576.

10. VanDerWal J, Shoo LP, Johnson CN, and Williams SE. Abundance and the Environmental Niche: Environmental Suitability Estimated from Niche Models Predicts the Upper Limit of Local Abundance. The American Naturalist 2009;174:282–91.

11. Wisz MS, Pottier J, Kissling WD, et al. The Role of Biotic Interactions in Shaping Distributions and Realised Assemblages of Species: Implications for Species Distribution Modelling. Biological Reviews 2013;88:15–30.

12. Blanchet FG, Cazelles K, and Gravel D. Co-Occurrence Is Not Evidence of Ecological Interactions. Ecology Letters 2020;23:1050–63.

13. Ni M and Vellend M. Soil Properties Constrain Predicted Poleward Migration of Plants under Climate Change. New Phytologist 2024;241:131–41.

14. Ni M and Vellend M. Space-for-Time Inferences about Range-Edge Dynamics of Tree Species Can Be Influenced by Sampling Biases. Global Change Biology 2021.

15. Berteaux D, Casajus N, De Blois S, Logan T, and Périé C. Changements climatiques et biodiversité du Québec : vers un nouveau patrimoine naturel. Les presses de l'Université du Québec, 2014.

16. Leboeuf M. Arbres et Plantes Forestières Du Québec et Des Maritimes - Réédition. Guides Nature Quintin. Éditions Michel Quintin, 2016.

17. Zhang Qp, Wang J, and Wang Q. Effects of Abiotic Factors on Plant Diversity and Species Distribution of Alpine Meadow Plants. Ecological Informatics 2021;61:101210.

18. Dunson WA and Travis J. The Role of Abiotic Factors in Community Organization. The American Naturalist 1991;138:1067–91.

19. Pyšek P, Jarošík V, Kropáč Z, Chytrý M, Wild J, and Tichý L. Effects of Abiotic Factors on Species Richness and Cover in Central European Weed Communities. Agriculture, Ecosystems & Environment 2005;109:1–8.

20. Soliveres S and Maestre FT. Plant–Plant Interactions, Environmental Gradients and Plant Diversity: A Global Synthesis of Community-Level Studies. Perspectives in Plant Ecology, Evolution and Systematics 2014;16:154–63.

21. Jin LS, Yin D, Fortin MJ, and Cadotte MW. The Mechanisms Generating Community Phylogenetic Patterns Change with Spatial Scale. Oecologia 2020;193:655–64.

22. Morneau C. Classification écologique du territoire québécois. Québec, Canada, 2021.

23. Saucier JP, Berger JP, D'Avignon H, and Racine P. Normes Techniques : Le Point d'observation Écologique. Gouvernement du Quéec, 1994.

24. Sadi AS and Chatterjee S. 10.4.2 Variance Inflation Factors. In: *Regression Analysis By Example Using R*. 6th. Wiley Series in Probability and Statistics. Wiley, 2023:480.

25. Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 1901;2:559–72.

26. Kyriazos T and Poga M. Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. Open Journal of Statistics 2023;13:404–24.

27. Dray S, Legendre P, and Peres-Neto PR. Spatial Modelling: A Comprehensive Framework for Principal Coordinate Analysis of Neighbour Matrices (PCNM). Ecological Modelling 2006;196:483–93.

28. Moran PAP. Notes on Continuous Stochastic Phenomena. Biometrika 1950;37:17.

29. Dray S, Bauman D, Blanchet G, et al. Adespatial: Multivariate Multiscale Spatial Analysis. 2025.

30. Blanchet FG, Tikhonov G, and Norberg A. HMSC: Hierarchical Modelling of Species Community. 2019.

31. Ovaskainen O, Tikhonov G, Norberg A, et al. How to Make More out of Community Data? A Conceptual Framework and Its Implementation as Models and Software. Ecology Letters 2017;20:561–76.

32. Ovaskainen O and Abrego N. Joint Species Distribution Modelling: With Applications in R. Ecology, Biodiversity and Conservation. Cambridge: Cambridge University Press, 2020. DOI: 10.1017/9781108591720.

33. Bhattacharya A and Dunson DB. Sparse Bayesian infinite factor models. Biometrika 2011;98:291–306.

34. Kong A. A Note on Importance Sampling Using Standardized Weights. Technical Report 348. University of Chicago, Dept. of Statistics, 1992:4.

35. Borcard D, Legendre P, and Drapeau P. Partialling out the Spatial Component of Ecological Variation. Ecology 1992;73:1045–55.

36. Peres-Neto PR, Legendre P, Dray S, and Borcard D. Variation partitioning of species data matrices: Estimation and comparison of fractions. Ecology 2006;87:2614–25.

37. Nakagawa S and Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. en. Methods in Ecology and Evolution 2013;4. Ed. by O'Hara RB:133–42.

38. Leibold MA, Rudolph FJ, Blanchet FG, et al. The Internal Structure of Metacommunities. Oikos 2022;2022.

39. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria, 2025. URL: https://www.R-project.org/.

40. Gower JC and Dijksterhuis GB. Procrustes Problems. Oxford Statistical Science Series 30. Oxford ; New York: Oxford University Press, 2004.

41. Aerts R. The Advantages of Being Evergreen. Trends in Ecology & Evolution 1995;10:402–7.

42. Rydin H and Jeglum JK. The Biology of Peatlands, 2e. Second Edition, Second Edition. Biology of Habitats Series. Oxford, New York: Oxford University Press, 2013.

43. Choat B, Jansen S, Brodribb TJ, et al. Global Convergence in the Vulnerability of Forests to Drought. Nature 2012;491:752–5.

44. Svenning JC and Skov F. Limited Filling of the Potential Range in European Tree Species. Ecology Letters 2004;7:565–73.

45. Dubuis A, Giovanettina S, Pellissier L, Pottier J, Vittoz P, and Guisan A. Improving the Prediction of Plant Species Distribution and Community Composition by Adding Edaphic to Topo-Climatic Variables. Journal of Vegetation Science 2013;24:593–606.

Data Availability

The data and R code that support the findings of this study are openly available in Zenodo at 10.5281/zenodo.23600586.

Acknowledgments

We thank Varina Crisfield and Mark Vellend for valuable comments on previous draft of the manuscript.

Competing interests

The authors declare no conflicts of interest.

Figures

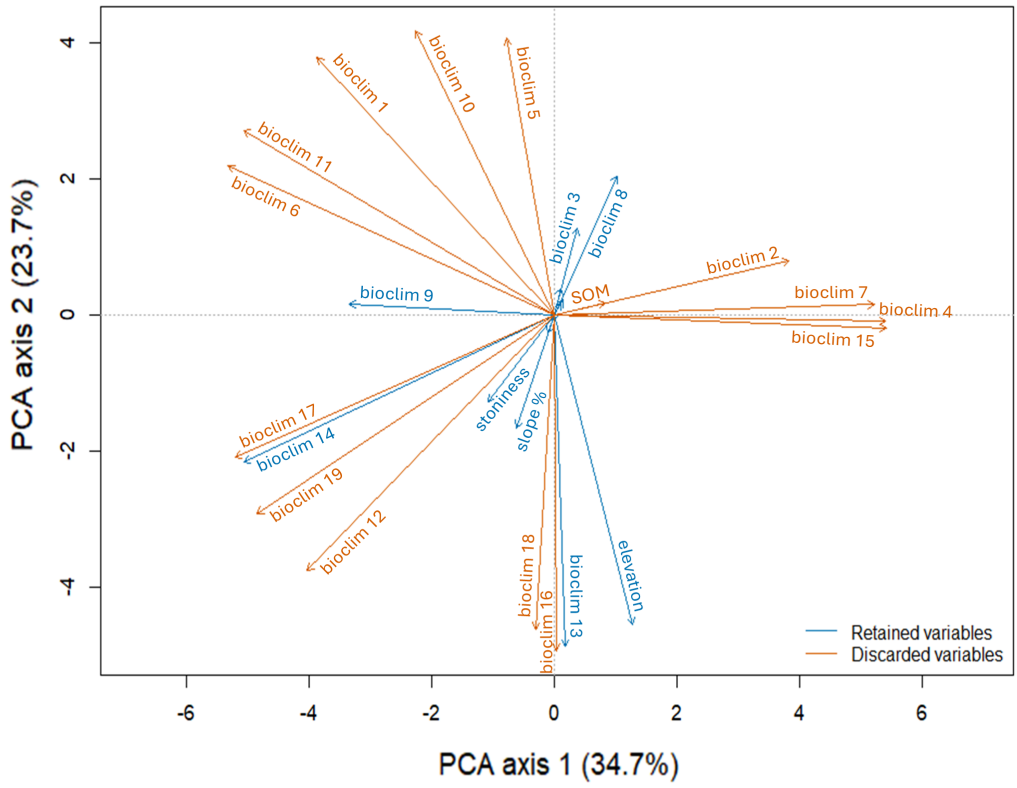


Figure 1: Principal component analysis illustrating the environmental variables included in the model (in blue) and the ones excluded from the model (in orange). Variables near the center of the plots were not labeled for visual clarity and because their interpretability is limited.

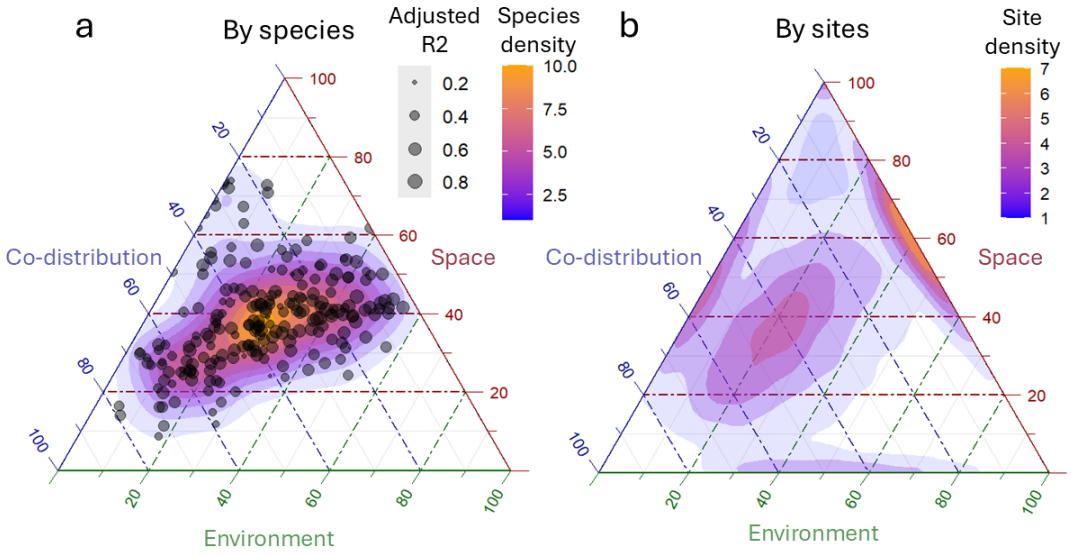


Figure 2: Ternary plots of the variation partitioning analysis showing the relative contributions of environment, space, and co-distribution (latent variables) to the distribution of plant species (a) and the points' contribution to each component of the analyses (b). The amount of variation explained by the model is presented using adjusted R^2. The color gradient (light blue to orange) indicates a smoothed density of species (a) or sites (b) at every location on the ternary plot, with warmer colors representing a greater number of species (a) or sites (b). The color gradient was calculated using the estimated kernel density within the ternary space. In (a), the points represent individual species. Point size is proportional to the total variation explained for each species (adjusted R^2), with larger points indicating higher explained variance. Points were not included in the site ternary plot because their number would make the plot unreadable.

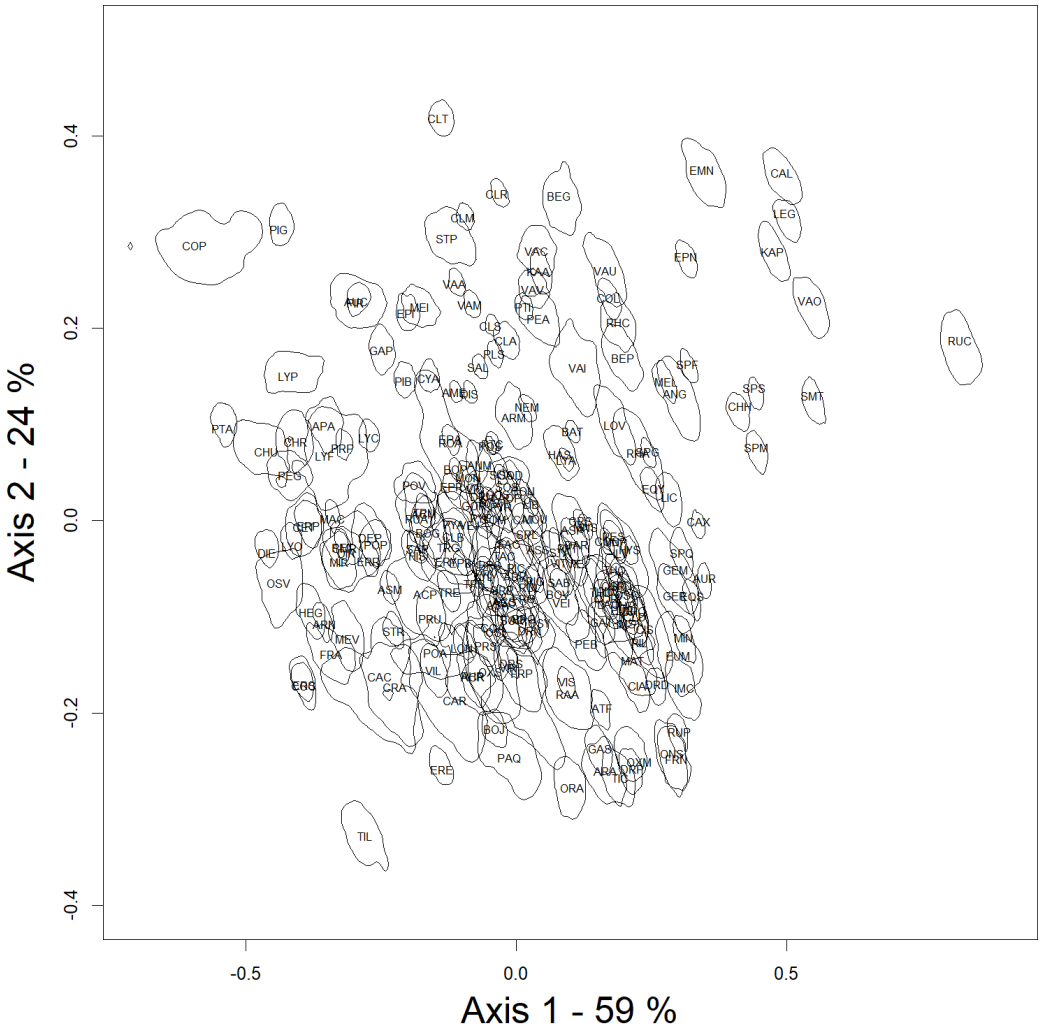


Figure 3: Principal component analysis (PCA) carried out on the species by environmental parameters matrix, $\mathbf{B}^{\mathrm{ENV}}$, (accounting for bioclimatic, topographic, and edaphic variables) for every iterations of the estimated model. For a clearer visualization, an orthogonal Procrustes rotation was applied on the PCA results and to account for the iterations, the species were presented as polygons including 50% of all iterations. The three-letter species codes indicate the mean position of each species (see Appendix 4 for full species names).

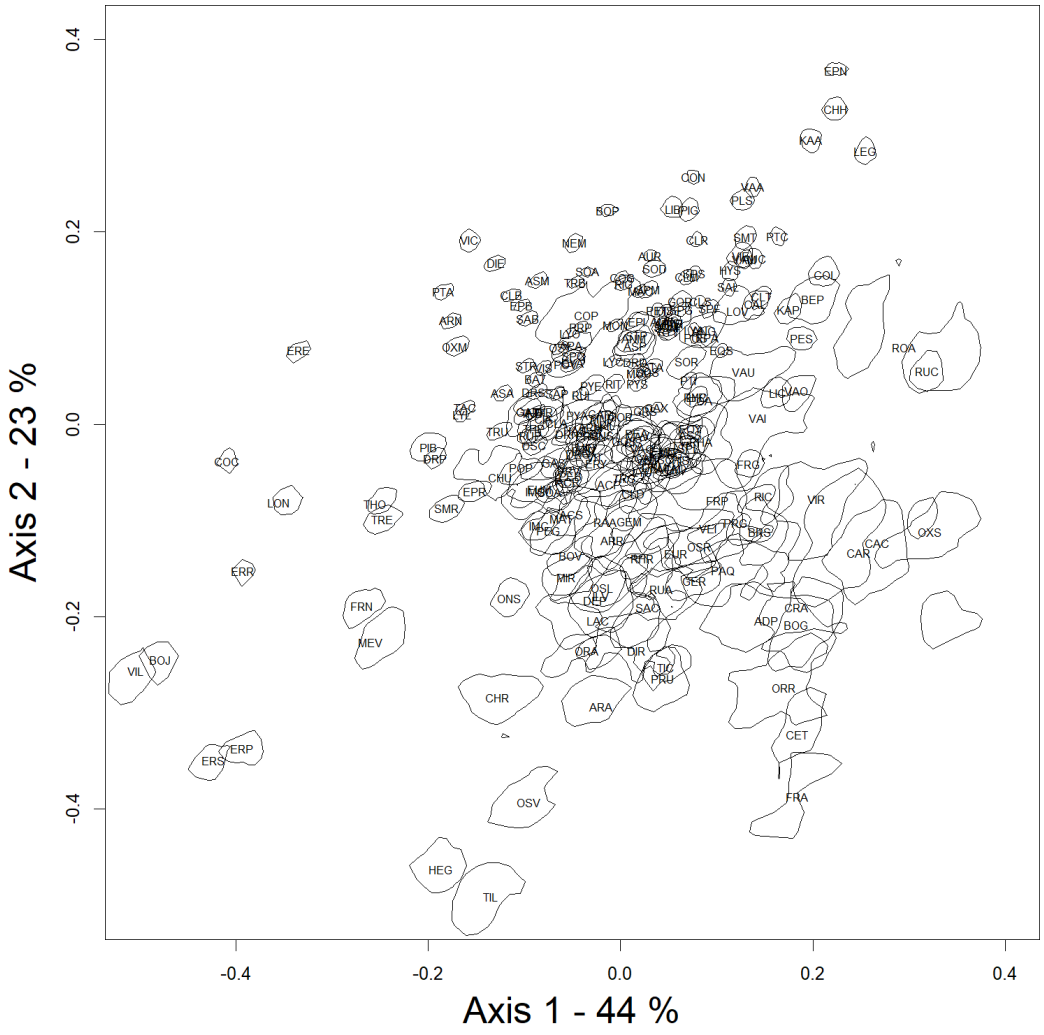


Figure 4: Principal component analysis (PCA) carried out on the species by spatial parameters matrix ($\mathbf{B}^{\mathrm{SPAT}}$, accounting for the 11 distance-based Moran's eigenvector maps) of every iterations of the estimated model. For a clearer visualization, an orthogonal Procrustes rotation was applied on the PCA results. To account for the iterations, the species were presented as polygons including 50% of all iterations. The three-letter species codes indicate the mean position of each species (see Appendix 4 for full species names).



Figure 5: Correlation matrix of the 224 plant species based on the species by latent parameters matrix ($\mathbf{\Lambda}$). Red pixels insinuate a negative effect on co-distribution and blue pixel, a positive effect on co-distribution. Black pixels were attributed to correlations with no significant effect on co-distribution. See Appendix 2 for the same figure with species' three-letter codes.

Supporting Information

From Temperate to Boreal – Untangling the Importance of Environment, Space, and Species co-distributions in Forest Plant Distributions

Appendix S1. Effective Sample Size

Figure S1.1 presents the effective sample sizes on a logarithmic scale for the multivariate hierarchical model run for 15,000 iterations. Effective sample sizes were calculated using the 5,000 retained iterations following a burn-in of 10,000 iterations. Of all effective sample sizes, the lowest were associated with species with a narrow spatial distribution and for which the model had difficulties estimating the intercept and the parameters associated to the broadest distance-based Moran's eigenvector maps (dbMEMs). As this is a situation that will remain present because of the nature of the data and dbMEMs used and it is at the core of our study that focuses on regional pattern of plant diversity, we elected not to alter the structure of our model, but to be careful in the interpretation we made of our results given this known particularity in our model.

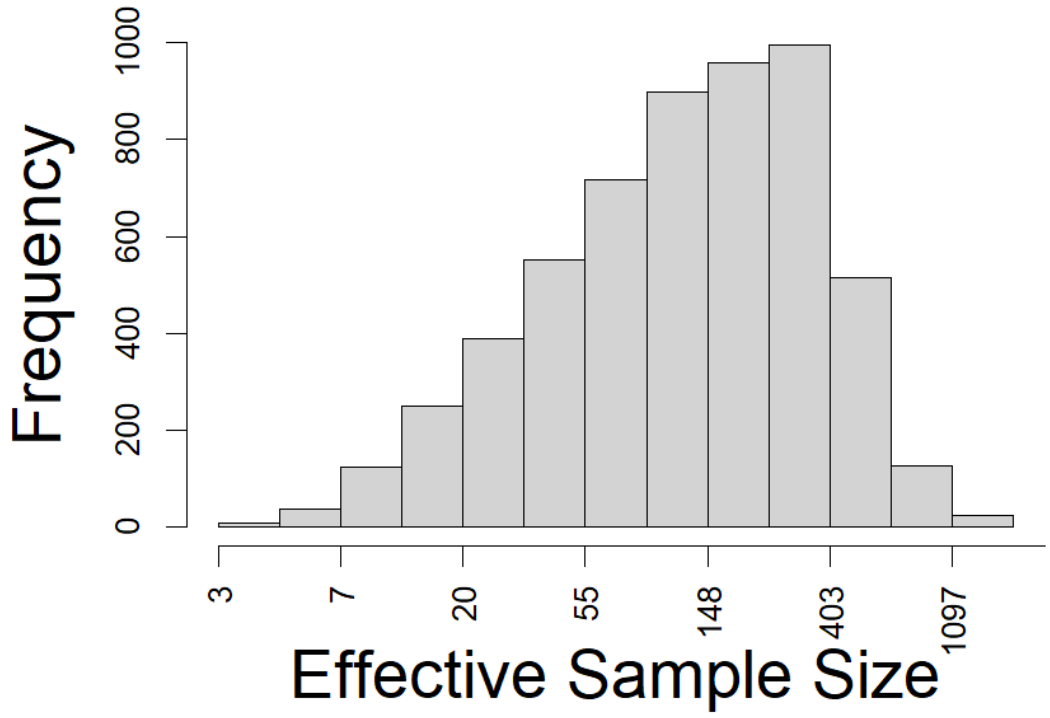


Figure S1.1: Histogram of the effective sample sizes for all 5,600 parameter–species pairs, calculated from the 5,000 retained MCMC iterations following a burn-in of 10,000 iterations. Effective sample sizes were put in the logarithmic scale for better visualization.

Appendix S2. Co-occurrence Matrix

Figure S2.1 is the same as Figure 5 found in the main text, except species' three-letter codes were added.

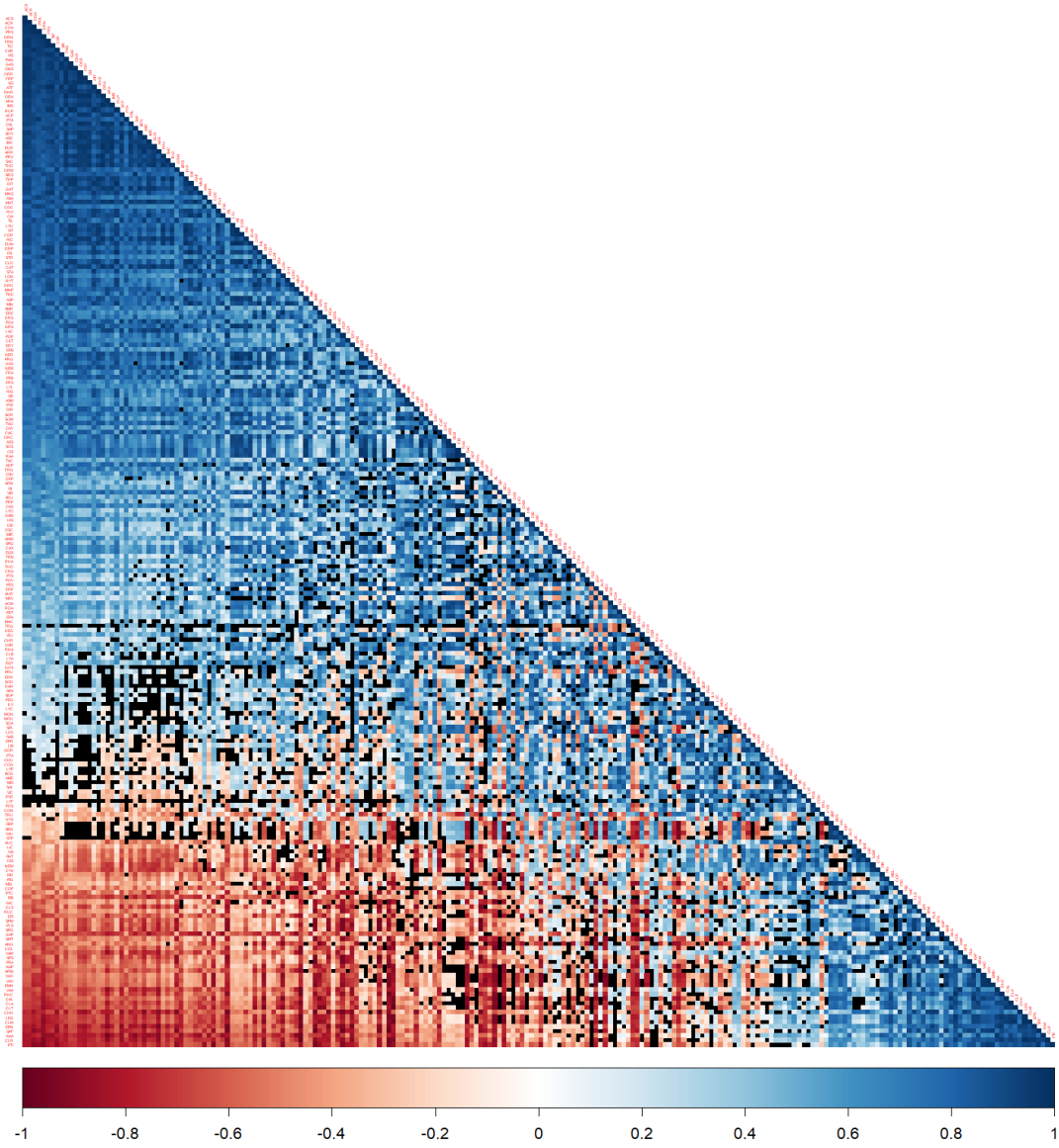


Figure S2.1: Correlation matrix of the 224 plant species based on the species' mean of all the latent variables estimated from the MCMC. Red pixels insinuate a negative effect on co-distribution and blue pixel, a positive effect on co-distribution. Black pixels were attributed to species with no significant effect on co-distribution. Three-letter species code are written in cex = 0.3 for optimal visualization (See Appendix 4 for full species name).

Appendix S3. Environmental Variables

Environmental variables were screened to reduce redundancy among predictors. Table S3.1 presents the environmental variables retained for subsequent analyses, while Table S3.2 presents the discarded variables.

Table S3.1: Environmental variables retained and used in subsequent analyses.

Variable	**Description**
Bioclim3	Isothermality, calculated as the ratio of mean diurnal temperature range (BIO2) to annual temperature range (BIO7), multiplied by 100.
Bioclim8	Mean temperature of the wettest quarter.
Bioclim9	Mean temperature of the driest quarter.
Bioclim13	Precipitation of the wettest month.
Bioclim14	Precipitation of the driest month.
Slope (%)	Slope measured at the center of the ecological observation point using a clinometer (0–99%).
Stoniness	Percentage of soil volume composed of rock fragments greater than 2 mm in diameter.
Elevation	Elevation interpolated from a topographic map (± 5 m).
Drainage	Soil drainage class, representing the capacity of the soil to remove excess water.

Table S3.2: Environmental variables discarded and excluded from subsequent analyses.

Variable	**Description**
Bioclim 1	Mean annual temperature.
Bioclim 2	Mean diurnal temperature range, calculated as the mean of the monthly differences between maximum and minimum temperature.
Bioclim 4	Temperature seasonality, calculated as the standard deviation of monthly temperature multiplied by 100.
Bioclim 5	Maximum temperature of the warmest month.
Bioclim 6	Minimum temperature of the coldest month.
Bioclim 7	Annual temperature range, calculated as the maximum temperature of the warmest month (Bioclim 5) minus the minimum temperature of the coldest month (Bioclim 6).
Bioclim 10	Mean temperature of the warmest quarter.
Bioclim 11	Mean temperature of the coldest quarter.
Bioclim 12	Annual precipitation.
Bioclim 15	Precipitation seasonality, expressed as the coefficient of variation.
Bioclim 16	Precipitation of the wettest quarter.
Bioclim 17	Precipitation of the driest quarter.
Bioclim 18	Precipitation of the warmest quarter.
Bioclim 19	Precipitation of the coldest quarter.
SOM	Thickness of the soil organic matter layer, measured in cm.

Appendix S4. Species Codes and Taxonomic Nomenclature

Table S4.1: Species codes and corresponding scientific names.

Code	Scientific name
ACM	*Achillea millefolium*
ACP	*Actaea pachypoda*
ACR	*Actaea rubra*
ACS	*Actaea sp.*
ADP	*Adiantum pedatum*
AME	*Amelanchier sp.*
ANG	*Andromeda glaucophylla*
ANM	*Anaphalis margaritacea*
APA	*Apocynum androsaemifolium*
ARA	*Arisaema triphyllum*
ARM	*Aronia melanocarpa*
ARN	*Aralia nudicaulis*
ARR	*Aralia racemosa*
ASA	*Oclemena acuminata*
ASC	*Asarum canadense*
ASM	*Eurybia macrophylla*
ASP	*Symphyotrichum puniceum*
ASS	Asteraceae spp.
ATF	*Athyrium filix-femina*
AUC	*Alnus viridis* subsp. *crispa*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
AUR	*Alnus incana* subsp. *rugosa*
BAT	*Bazzania trilobata*
BEG	*Betula glandulosa*
BEP	*Betula pumila*
BOG	*Betula populifolia*
BOJ	*Betula alleghaniensis*
BOP	*Betula papyrifera*
BOV	*Botrychium virginianum*
BRS	*Bromus sp.*
CAC	*Carya cordiformis*
CAL	*Chamaedaphne calyculata*
CAR	*Carpinus caroliniana*
CAT	*Caulophyllum thalictroides*
CAX	*Carex sp.*
CET	*Prunus serotina*
CHH	*Gaultheria hispidula*
CHR	*Quercus rubra*
CHU	*Chimaphila umbellata*
CIA	*Circaea alpina*
CIS	*Cirsium sp.*
CLA	*Cladonia sp.*
CLB	*Clintonia borealis*
CLD	*Climacium dendroides*
CLM	*Cladonia mitis*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
CLR	*Cladonia rangiferina*
CLS	*Cladonia sp.*
CLT	*Cladonia stellaris*
COA	*Cornus alternifolia*
COC	*Corylus cornuta*
COG	*Coptis trifolia*
COL	*Comandra umbellata*
CON	*Cornus canadensis*
COP	*Comptonia peregrina*
COR	*Cornus sericea*
CRA	*Crataegus sp.*
CYA	*Cypripedium acaule*
DAR	*Dalibarda repens*
DEP	*Dennstaedtia punctilobula*
DIE	*Diervilla lonicera*
DIR	*Dirca palustris*
DIS	*Dicranum sp.*
DRC	*Dryopteris cristata*
DRD	*Gymnocarpium dryopteris*
DRM	*Dryopteris marginalis*
DRN	*Parathelypteris noveboracensis*
DRP	*Phegopteris connectilis*
DRS	*Dryopteris carthusiana*
EMN	*Empetrum nigrum*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
EPA	*Chamaenerion angustifolium*
EPB	*Picea glauca*
EPI	*Epigaea repens*
EPN	*Picea mariana*
EPR	*Picea rubens*
EQS	*Equisetum sp.*
EQY	*Equisetum sylvaticum*
ERE	*Acer spicatum*
ERP	*Acer pensylvanicum*
ERR	*Acer rubrum*
ERS	*Acer saccharum*
ERY	*Erythronium americanum*
EUM	*Eutrochium maculatum*
EUR	*Ageratina altissima*
FRA	*Fraxinus americana*
FRG	*Fragaria sp.*
FRN	*Fraxinus nigra*
FRP	*Fraxinus pennsylvanica*
GAP	*Gaultheria procumbens*
GAS	*Galium sp.*
GAT	*Galium triflorum*
GEM	*Geum macrophyllum*
GER	*Geum rivale*
GOR	*Goodyera repens*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
GOS	*Goodyera sp.*
GRS	Poaceae spp.
HAS	*Platanthera sp.*
HEG	*Fagus grandifolia*
HIS	*Hieracium sp.*
HYS	*Hylocomium splendens*
ILV	*Ilex verticillata*
IMC	*Impatiens capensis*
IMS	*Impatiens sp.*
KAA	*Kalmia angustifolia*
KAP	*Kalmia polifolia*
LAC	*Laportea canadensis*
LEG	*Rhododendron groenlandicum*
LIB	*Linnaea borealis*
LIC	*Neottia cordata*
LON	*Lonicera hirsuta*
LOV	*Lonicera villosa*
LYA	*Spinulum annotinum*
LYC	*Lycopodium clavatum*
LYF	*Diphasiastrum digitatum*
LYL	*Huperzia lucidula*
LYO	*Dendrolycopodium obscurum*
LYP	*Diphasiastrum complanatum*
LYU	*Lycopodiella inundata*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
MAC	*Maianthemum canadense*
MAT	*Matteuccia struthiopteris*
MEI	*Melampyrum lineare*
MEL	*Larix laricina*
MEV	*Medeola virginiana*
MIN	*Mitella nuda*
MIR	*Mitchella repens*
MNP	*Mnium punctatum*
MNS	*Mnium sp.*
MON	*Monotropa uniflora*
MOU	*Moneses uniflora*
NEM	*Ilex mucronata*
ONS	*Onoclea sensibilis*
ORA	*Ulmus americana*
ORR	*Ulmus rubra*
OSC	*Osmundastrum cinnamomeum*
OSL	*Osmorhiza claytonii*
OSR	*Osmunda regalis*
OSV	*Ostrya virginiana*
OSY	*Osmunda claytoniana*
OXM	*Oxalis montana*
OXS	*Oxalis stricta*
PAQ	*Parthenocissus quinquefolia*
PEA	*Peltandra virginica*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
PEB	*Populus balsamifera*
PEG	*Populus grandidentata*
PES	*Petasites frigidus* var. *palmatus*
PET	*Populus tremuloides*
PIB	*Pinus strobus*
PIG	*Pinus banksiana*
PIR	*Pinus resinosa*
PLS	*Pleurozium schreberi*
POA	*Polystichum acrostichoides*
POP	*Polygonatum pubescens*
POS	*Polytrichum sp.*
POV	*Polypodium virginianum*
PRG	*Prunella vulgaris*
PRP	*Prunus pensylvanica*
PRS	*Prenanthes sp.*
PRU	*Tsuga canadensis*
PRV	*Prunus virginiana*
PTA	*Pteridium aquilinum*
PTC	*Ptilium crista-castrensis*
PTI	*Ptilidium ciliare*
PYA	*Pyrola asarifolia*
PYE	*Pyrola elliptica*
PYR	*Orthilia secunda*
PYS	*Pyrola sp.*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
RAA	*Ranunculus acris*
RHA	*Rhamnus alnifolia*
RHC	*Rhododendron canadense*
RHR	*Toxicodendron radicans*
RIC	*Ribes cynosbati*
RIG	*Ribes glandulosum*
RIL	*Ribes lacustre*
RIT	*Ribes triste*
ROA	*Rosa acicularis*
RUA	*Rubus allegheniensis*
RUC	*Rubus chamaemorus*
RUI	*Rubus idaeus*
RUP	*Rubus pubescens*
SAB	*Abies balsamea*
SAC	*Sambucus canadensis*
SAL	*Salix sp.*
SAP	*Sambucus racemosa*
SMR	*Maianthemum racemosum*
SMT	*Maianthemum trifolium*
SOA	*Sorbus americana*
SOD	*Sorbus decora*
SOM	*Solidago macrophylla*
SOR	*Solidago rugosa*
SOS	*Solidago sp.*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
SPF	*Sphagnum fuscum*
SPG	*Sphagnum girgensohnii*
SPL	*Spiraea alba* var. *latifolia*
SPM	*Sphagnum magellanicum*
SPQ	*Sphagnum squarrosum*
SPS	*Sphagnum sp.*
STA	*Streptopus amplexifolius*
STP	*Stereocaulon paschale*
STR	*Streptopus lanceolatus*
TAC	*Taxus canadensis*
TAO	*Taraxacum officinale*
THD	*Thalictrum dioicum*
THO	*Thuja occidentalis*
THP	*Thalictrum pubescens*
TIC	*Tiarella cordifolia*
TIL	*Tilia americana*
TRB	*Lysimachia borealis*
TRE	*Trillium erectum*
TRG	*Trillium grandiflorum*
TRU	*Trillium undulatum*
VAA	*Vaccinium angustifolium*
VAC	*Vaccinium cespitosum*
VAI	*Vaccinium sp.*
VAM	*Vaccinium myrtilloides*

Continued on next page

Table S4.1 – continued from previous page

Code	Scientific name
VAO	*Vaccinium oxycoccos*
VAU	*Vaccinium uliginosum*
VAV	*Vaccinium vitis-idaea*
VEI	*Veratrum viride*
VEJ	*Vicia cracca*
VIC	*Viburnum cassinoides*
VIE	*Viburnum edule*
VIL	*Viburnum lantanoides*
VIR	*Vitis riparia*
VIS	*Viola sp.*
VIT	*Viburnum opulus* var. *trilobum*