Robustness and limitations of maximum entropy in plant community assembly

Jelyn Gerkema [1],*, Daniel E. Bunker [2], Andrew M. Cunliffe [3], Erika Bazzato [4], Michela Marignani [4], Tommaso Sitzia [5,6], Isabelle Aubin [7], Stefano Chelli [8], Julieta A. Rosell [9], Peter Poschlod [10], Josep Penuelas [11,12], Arildo S. Dias [13], Christian Rossi [14], Tanvir A. Shovon [15], Juan A. Campos [16], Mark C. Vanderwel [17], Sharif A. Mukul [18,19], Bruno E.L. Cerabolini [20], Thomas Sibret [21], Bruno Hérault [22,23], Sylvain Schmitt [22,23], Pedro Higuchi [24], James L. Tsakalos [25], Decky I. Junaedi [26], Yun-Peng Zhao [27], Vanessa Minden [28], Ana Carolina da Silva [24], Tereza Mašková [29,30], Roberto Canullo [8], Ning Dong [31], Edwin T. Pos [1,32]

[1] Quantitative Biodiversity Dynamics, Ecology and Biodiversity, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands
[2] Department of Biological Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ, USA
[3] Department of Geography, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK
[4] Department of Life and Environmental Sciences, University of Cagliari, Via Sant'Ignazio da Laconi, 13, 09123, Cagliari, Italy
[5] Università degli Studi di Padova, Department of Land, Environment, Agriculture and Forestry, Viale dell'Università 16, 35020, Legnaro (PD), Italy
[6] National Biodiversity Future Center, Palermo, Italy
[7] Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, Sault Ste Marie, Ontario, Canada
[8] School of Biosciences and Veterinary Medicine, University of Camerino, Italy
[9] Laboratorio Nacional de Ciencias de la Sostenibilidad, Instituto de Ecología, Universidad Nacional Autónoma de Mexico, A.P. 70-275, Ciudad Universitaria, Coyoacán, 04510, Mexico City, Mexico
[10] Ecology and Conservation Biology, Institute of Plant Sciences, University of Regensburg, Germany
[11] CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Spain
[12] CREAF, Cerdanyola del Vallès, Spain
[13] Goethe University, Department of Physical Geography, Altenhöferallee 1, 60438, Frankfurt am Main, Germany
[14] Swiss National Park, Runatsch 124-Chastè Planta-Wildenberg, 7530, Zernez, Switzerland
[15] Center for Boreal Research, Northern Alberta Institute of Technology, 8102 99 Ave, Peace River, AB T8S 1R2, Canada
[16] Department of Plant Biology and Ecology, Faculty of Science and Technology, University of the Basque Country UPV/EHU, Bilbao, Spain
[17] Department of Biology, University of Regina, Regina, Canada
[18] Department of Environment and Development Studies, United International University, Dhaka, 1212, Bangladesh
[19] Tropical Forests and People Research Centre, University of the Sunshine Coast, Maroochydore DC, Queensland, 4556, Australia
[20] Department of Biotechnology and Life Science, University of Insubria, via J.H. Dunant, 3, 21100, Varese, Italy
[21] Q-forestlab, Department of Environment, Ghent University, Ghent, Belgium
[22] CIRAD, Forêts et Sociétés, Montpellier, F-34398, France
[23] Forêts et Sociétés, Univ Montpellier, CIRAD, Montpellier, France
[24] Departamento de Engenharia Florestal, Universidade do Estado de Santa Catarina, Lages, SC, Brazil
[25] Harry Butler Institute, Murdoch University, Murdoch, Perth, WA, Australia
[26] Research Centre for Ecology and Ethnobiology, National Research and Innovation Agency (BRIN), Indonesia
[27] Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, 310058, Hangzhou, China
[28] Department of Biology, Research group WILD, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050, Brussels, Belgium
[29] Institute of Ecology and Evolution with Herbarium Haussknecht and Botanical Garden, Plant Biodiversity Group, Friedrich Schiller University Jena, D-07743, Jena, Germany
[30] Senckenberg Institute for Plant Form and Function Jena (SIP), D-07743, Jena, Germany
[31] Georgina Mace Centre for the Living Planet, Imperial College London, Department of Life Sciences, Silwood Park Campus, Ascot SL5 7PY, UK
[32] Utrecht University Botanic Gardens, P.O. Box 80162, 3508 TD, Utrecht, The Netherlands

* Corresponding author.
  E-mail address: J.Gerkema@uu.nl (J. Gerkema).

A B S T R A C T

An in-depth understanding of local plant community assembly is critical to direct conservation efforts to promising areas and increase the efficiency of management strategies. This, however, remains elusive due to the sheer complexity of ecological processes. The maximum entropy-based Community Assembly via Trait Selection (CATS) model was designed to quantify the relative contributions of trait-based filtering, dispersal mass effects, and stochastic processes on community assembly. As a maximum entropy model, it does so without introducing additional bias or assumptions. Despite its increasing use, questions regarding its robustness and potential limitations remain. Here, we compared model predictions using either local or database-derived trait values, across different levels of species richness and between different taxonomic levels. A total of 19 datasets and 790 plots were analysed, spanning multiple habitat types (n = 18) and biomes (n = 7). Results indicate trait value origin does indeed influence model outcomes, warranting caution in selecting the method for obtaining trait data. We hypothesise that, for example, intraspecific trait variation combined with trait-based filtering or stochastic processes causes local and database trait values to deviate, potentially even further exacerbated by imputing missing trait data. Furthermore, trait-related information obtained from the model decreased with increasing species richness. We further hypothesise this could signal that stochastic processes are more dominant within species-rich systems, for example, due to functional redundancy or the existence of multiple fitness strategies. This general pattern was conserved across biomes, although with varying strength, showing CATS' robustness despite these challenges.

1. Introduction

Understanding plant community assembly remains challenging (Grace, 1999; Shipley, 2010a), despite decades of research into the relationships between plant species occurrence, functional traits, and environmental features (e.g. Brown et al., 2014; Laughlin and Laughlin, 2013; Shipley et al., 2006; Warton et al., 2015). While the general mechanisms – trait-based filtering, dispersal and stochastic processes – are well-understood in principle, quantifying their relative contribution to local community assembly can be difficult. Nevertheless, this information is considered invaluable for optimising conservation efforts and increasing the efficiency of management strategies, thereby helping to meet conservation and restoration targets (Chase, 2007).

Many ecological models incorporate a priori mechanistic assumptions, incorporating pre-existing ideas about local processes based on established ecological theories or empirical data. However, different models can be constructed to represent the same community, especially when the system is complex. Choosing a model could thus introduce unintentional biases (Harte, 2011). The CATS model (Community Assembly via Trait Selection) stands out as a specific application of the principle of Maximum Entropy that allows objective statistical inference without incorporating a priori assumptions (Shipley, 2010a). It was designed to quantify the relative importance of so called constraints in the form of Community Weighted Means (CWMs) of functional traits to make inferences regarding processes driving community assembly. It does so without any a priori assumptions about the importance of these explanatory mechanisms.

CATS predicts the relative abundance distribution of species in a local plot, taking into account only what is really known from empirical data. This information is condensed into summarised values, forming constraints or requirements to which the predicted abundance distribution must adhere (Shipley, 2010a). The predicted distribution must yield the same summarised values. From a set of possible distributions that satisfy these constraints, the algorithm selects the one that maximises relative entropy, a measure for uncertainty or information content. Within the CATS framework, this is the distribution closest to a uniform or previously specified prior distribution. The selected distribution reflects *only* the information specified by the constraints and nothing else. Consequently, CATS, and other maximum entropy models analyse data without introducing any a priori bias or unwarranted assumptions (Dewar and Porté, 2008; Harte, 2011; Phillips et al., 2006; Pueyo et al., 2007; Shipley, 2010a; Volkov et al., 2009).

Specifically, CATS incorporates constraints based on CWMs and optionally variance (CWV) of functional trait values and integrates a prior distribution that describes the relative abundances of species within a regional community surrounding the local plot of interest (meta-community) (Shipley, 2010a, 2014). It operates on the assumption that *if* there is a relationship between the functional traits of species and their relative abundances, indicating that trait-based filtering plays a role in community assembly, CWMs should be skewed to reflect the species with more favourable traits in such environments. Including these constraints should result in more accurate predictions compared with CWMs without such a relationship. Similarly, if the local community reflects the regional distribution through dispersal mass effects, incorporating its distribution as a prior should improve the prediction. CATS is a four-step model that alternates the input of constraints, either permuted or not, and a prior (uniform or a representation of the meta-community). By using the output of each of these steps in several decomposition functions, the model can quantify the relative contributions of dispersal mass effects, trait-based filtering, and their joint contribution to community assembly (Shipley, 2014). The part of the variance that is left unexplained, or the portion of the community structure that cannot be predicted from the constraints (traits) and prior (meta-community), is usually attributed to stochastic processes or unspecified constraints.

In previous studies, CATS has provided useful insights into local ecosystem functioning (Arim et al., 2023; Baastrup-Spohr et al., 2015; Cunillera-Montcusí et al., 2020; Frenette-Dussault et al., 2013; Harabiš and Dolný, 2018; He et al., 2022; Janssen et al., 2017; Laliberté et al., 2012; Niu et al., 2016; Pos et al., 2023; Rech et al., 2019; Shipley et al., 2012; Sonnier et al., 2012; Strahan et al., 2018), despite being met with scrutiny and controversy upon its introduction as a reaction to the underlying mathematical principles (Haegeman and Loreau, 2009). These concerns have since been addressed (Shipley, 2009b,a) but questions regarding CATS's limitations and robustness remain.

Functional trait data can either be measured in the field, which is very labour-intensive, or obtained from databases (Lavorel et al., 2008; Kattge et al., 2020). Intraspecific differences cause these trait values to diverge, especially in more extreme environments (Cordlandwehr et al., 2013) and this could impact model outcomes. For example, a review investigating mechanisms of community assembly trait-based null models found that the mechanism of competitive dominance was more frequently detected with local trait measurements than with database values (Perronne et al., 2017). However, it remains unclear whether such differences between local details and aggregated database-derived trait values could lead to misinterpretations or false conclusions within

the CATS framework. Additionally, differences in taxonomic resolution (i.e. species vs. genus level) could influence overall conclusions, especially when taxonomic units are not identified to species level or when trait data are unavailable at that level. A large-scale comparison across different habitats and geographical locations, including variations in species richness, will help clarify the overall application of CATS.

In a previous study, the predictive abilities of CATS were explored under specific conditions and compared with other methods such as linear regression analysis and Hubbell's neutral model (Merow et al., 2011). Here, we assessed the generalisability of the CATS model across different taxonomic resolutions, gradients of species richness, biomes, and methods of obtaining functional trait data, by analysing data from several independent studies and the TRY trait database (Kattge et al., 2020). Understanding whether CATS' results depend on these factors will enhance its current applications or guide future developments, making it possible to understand, explain, and predict community structure with much better accuracy.

2. Methods

2.1. Data collation

We compiled data from published, open-access studies using several criteria, resulting in 19 datasets, some split across multiple locations, spanning multiple continents, and a wide range of habitats (see Table A.1, A2, and A3 in the supplementary material). The selection criteria for the studies were as follows: (1). An assessment of community composition was available, recording the abundance of all species present or those representing approximately 80% or more of the total abundance (rather than solely recording species richness). (2). Data from at least two functional traits was obtained. (3). The trait data were measured on individuals collected in the same research area as the inventory plots. (4). The regional trait coverage was at least 60%, indicating that the species for which trait data was available covered more than 60% of the regional abundance. Data were either obtained directly from their respective publication or via the CESTES dataset (Jeliazkov et al., 2020). To ensure that the datasets were indeed varied in their environmental conditions, we used the mean annual temperature (MAT) and precipitation (MAP) to determine their biomes (Whittaker, 1975). Seven out of the nine biomes are represented in the collected studies (Fig. 1). If the MAT and/or MAP were not mentioned in the original studies, we extracted them from the WorldClim data (Hijmans et al., 2024).

Only species for which the traits were measured were retained in their respective datasets. Plots with less than 60% trait coverage were excluded from the analysis (Fig. A1). For the studies where plots were recorded at different sites, the data were divided accordingly, and trait data were supplemented if recorded at one site but not the other. The CATS model was run separately for each dataset. In subsequent statistical analysis, data from different sites were regrouped.

To assess the effect of measuring trait values in the field versus obtaining them from databases, we applied the CATS model with both the trait values measured in their respective studies (hereafter referred to as 'local values') and the values obtained from TRY database (hereafter referred to as 'TRY values'). The collected datasets in total contained a total of 27 different traits for which (standardised) TRY data were available. To maximise data availability for the gap-filling process for species with incomplete trait data, data were requested for all species in TRY (data requested on 27-10-2023). This resulted in a dataset with 78 021 unique "Accepted species names". Data from juvenile, immature, or unhealthy individuals and moss species were removed, as were data originating from experimental studies and entries with an error risk of greater than 3 (package: rtry, version 1.1.0) (Lam et al., 2023). Entries that were not actual species names but referred to a larger taxonomic unit or a broader group of species (e.g., "Dwarf shrub") or were suspected cultivars (i.e., Carex 'chocolate' or Rosa

'wedding day') were also excluded (n = 1585). Each trait in TRY includes a multitude of subtraits; those not fitting the objective of the trait (e.g., subtrait 1739, wood density measured after drying at 100 °C) were removed (n = 56). One trait was removed as it only had 17 data points (TraitID 116, root density). Lastly, to avoid circularities in the analysis, data from the collected datasets that were also present in TRY were omitted in the latter, leaving the final TRY dataset with 26 traits (Table B.1) and 67 842 unique, accepted species names.

2.2. Data cleaning & taxonomic harmonisation

To reconcile taxonomic nomenclature across the datasets used for the analysis and the TRY database, and to obtain taxonomic information at higher hierarchical levels (genus, family and order), the 'accepted species names' in TRY were matched against the WFO backbone (v.2023.12) (Elliott et al., 2023), using the WorldFlora R package version 1.13.2 (Kindt, 2020). The author citation is not included in the standardised species names in the TRY database leading to some entries having multiple matches, due to different species having the same genus and epithet but different author citations, and some entries did not have an initial match. Matching the TRY names to the WFO backbone was therefore done in a step-wise manner using the taxonomic and nomenclatural status of the entries in the WFO backbone. The taxonomic status of each WFO entry was marked as "Accepted", "Unchecked", or "Synonym", while the nomenclatural status was either "Conserved", "Illegitimate", "Invalid", "Rejected", "Superfluous", "Valid" or NA.

Of the accepted TRY names, 63 302 had direct matches with the WFO backbone. For the unmatched TRY names (n = 3076), subspecies, varieties and hybrid (_x) names were extracted and their appendices were removed (n = 894), resulting in 864 unique names that were matched again, yielding 751 new matches. The remaining unmatched names (n = 2298) were fuzzy matched to obtain an accepted species name using the fuzzy.match function of the WorldFlora R package (Kindt, 2020). The Levenshtein distance between the old and new name was calculated using the default distance of the fuzzy.match function. They were removed if they (1), had no or multiple valid matches with a species, subspecies or variety with a Levenshtein distance lower than 3 or, (2) were indistinguishable from another accepted species name except for the author citation, which is not included in TRY. This was to ensure that data belonging to different species were not incorrectly regarded as belonging to the same one. The resulting matches (n = 854) were checked for plausibility and removed if deemed not plausible (n = 20).

The species names within the TRY database for which one or multiple match(es) were found, can be categorised into five categories: (1). Names with a single accepted/unchecked match in the WFO backbone (n = 58 616). (2). Names with multiple accepted/unchecked matches in the WFO backbone (n = 2468), due to different species having the same genus and epithet but different author citations. From these duplicates, the names that had an "Unchecked" taxonomic status were removed. For the remaining duplicates, entries with anything other than a "Valid" nomenclatural status were removed, resulting in 1579 additional TRY names with only one WFO match and 889 remaining duplicates. (3). Names that occurred as an accepted/unchecked name and as a synonym for another species (n = 3699), for which we assumed the former as the correct one. (4). Names that were a synonym for a single taxonomically accepted species name (n = 4435) or (5), for multiple taxonomically accepted species names, distinguishable only by author citation (n = 194). For the latter, it was first checked whether retaining only entries with a nomenclatural status of NA or "Valid" led to a single match. Next, for the remaining names, entries with a nomenclatural status of NA were removed, resulting in a total of 127 new matches. Species that only had NA's or multiple "valid" entries were retained in the dataset with their original synonym name only if the accepted species names were absent in the TRY database (n =
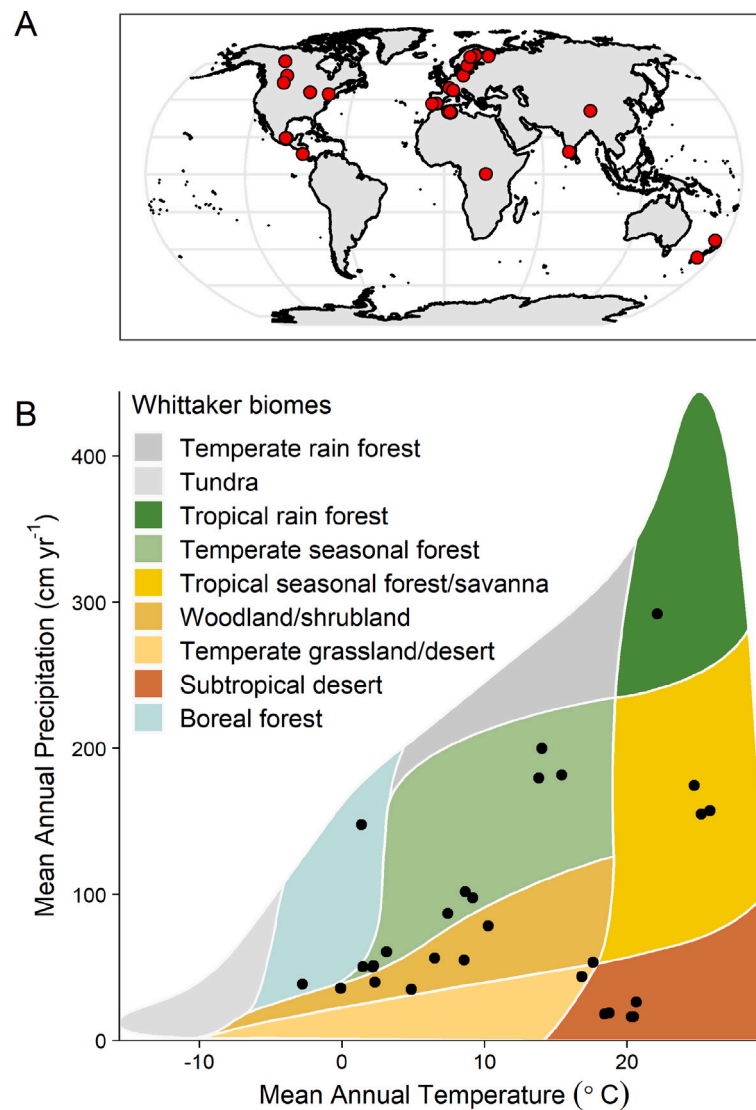
Fig. 1. Visualisation of the global distribution and biomes of the collected datasets. The studies were spread across 16 countries (A) and from seven different biomes (Whittaker, 1975) (B). Some studies encompassed multiple sites, sometimes in different countries. All are shown here. Mean annual temperature and precipitation were either obtained from the studies directly or from WordClim data (Hijmans et al., 2024). Code for the visualisation adapted from Cunliffe et al. (2022). See Table A.1 in the supplementary material for a full list of the included studies, their countries, habitats, biome, and corresponding references.

30). The remaining TRY names were removed, as there was no way to validate if the data listed under the synonym belonged to the data listed under the corresponding accepted name. Finally, we verified whether the newly found accepted names corresponding to the synonyms could be confused with another accepted species name, such names were removed if they could (n = 21).

For the resulting harmonised TRY names, taxonomic information (genus, family, order) was obtained (Kindt, 2020). For four genera, species were found to belong to two families. (1). *Ardisia* to Primulaceae (n = 50) and Ericaceae (n = 3). (2). *Athyrium* to Aspleniaceae (n = 5) and Athyriaceae (n = 1). (3). *Baeckea* to Myrtaceae (n = 11) and Bruniaceae (n = 1), and (4). *Libertia* to Iridaceae (n = 7) and Poaceae (n = 1). In all cases, the outliers belonged to species entries with a taxonomic status of "Unchecked" and the family was changed to the family to which the majority of the species belonged.

Missing trait values within the full TRY database were imputed using Bayesian hierarchical probabilistic matrix factorisation (BHPMF) which uses the correlation between functional trait values and higher taxonomic information (genus, family, order) to fill the gaps (Schrodt

et al., 2015). This method is suitable to deal with the sparsity of the TRY dataset (Schrodt et al., 2015) (see Table D.7 for the percentage of species present in TRY for each trait and study), as opposed to other common gap-filling methods such as calculating a taxonomic mean or the usage of multivariate imputation using chained equations (MICE) which is only effective up to 30% gaps (Taugourdeau et al., 2014). Trait values were log-transformed and Z-score normalised before the imputation (Joswig et al., 2023). Values were back-transformed afterwards.

We matched the species names in the other collected datasets to the harmonised TRY database following a step-wise procedure as well. First, we checked whether the species had a direct match in the harmonised TRY database. Species without a match, possibly due to the use of a synonym, were then checked against the WFO database. If a synonym match was found there, and it was not also the name of an accepted species name, the corresponding WFO number was used to check for a match in the TRY database. In some cases, species names in the collected datasets were adjusted in spelling to obtain a match (n = 58).

2.3. CATS model

To quantify the relative contribution or importance of trait-based filtering, dispersal mass effects, and their joint contribution to local community assembly, the CATS framework fits a model four times, each time with different information (Shipley, 2014). For each step, the prior distribution representing the meta-community (q_i), the predicted (p_i), and observed (o_i) relative abundances are used to calculate the fit of the model using the Kullback–Leibler (KL) index (Eq. (1)). The index is the ratio of two KL-divergences, a metric that measures the loss of information when approximating one distribution with another (Colin Cameron and Windmeijer, 1997).

$$R_{\mathrm{KL}}^2 = \frac{\sum_{i=1}^{S} o_i \cdot \ln(\frac{o_i}{p_i})}{\sum_{i=1}^{S} o_i \cdot \ln(\frac{o_i}{q_i})} \tag{1}$$

First, a null model is tested, providing an estimation of how well the model performs without any information. In an iterative process, the actual trait values are permuted, breaking any possible association between them and the respective relative abundances of species. These shuffled trait values are used to calculate new CWMs, which are used, together with a uniform prior, as input for the model. By repeating this process, calculating the fit for multiple times, and taking the average fit, model bias was estimated ($\bar{R}_{\mathrm{KL}}^2(u)$) (Shipley, 2014). The second model was run with the original, unpermuted trait values and uniform prior ($R_{\mathrm{KL}}^2(u, t)$), representing the situation where trait-based filtering plays a role but the relative abundances in the meta-community do not matter. The third model used the permuted trait values and the meta-prior ($R_{\mathrm{KL}}^2(m)$), while the fourth used the unpermuted values and the meta-prior ($R_{\mathrm{KL}}^2(m, t)$), thereby incorporating all available information.

After fitting the model four times, the resulting four KL indexes were used in four decomposition equations (Eqs. (2)–(5)) (Shipley, 2014). This leads to an estimation of the part of the observed community structure solely explained by trait-based filtering (pure trait effects), the portion attributed solely to dispersal mass effects (pure meta-community effects), their joint contribution, and the portion that remains unexplained (Shipley et al., 2012). Values were standardised by dividing them by: $1 - \bar{R}_{\mathrm{KL}}^2(u)$, so that the outcomes of Eqs. (2)–(5) sum up to 1.

$$\text{Pure trait effects}: \frac{R_{\mathrm{KL}}^2(m, t) - R_{\mathrm{KL}}^2(m)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} = \frac{\Delta R_{\mathrm{KL}}^2(t|m)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} \tag{2}$$

$$\text{Pure meta-community effects}: \frac{R_{\mathrm{KL}}^2(m, t) - R_{\mathrm{KL}}^2(u, t)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} = \frac{\Delta R_{\mathrm{KL}}^2(m|t)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} \tag{3}$$

$$\text{Joint effects}: \frac{\Delta R_{\mathrm{KL}}^2(m + t)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} = \frac{(\Delta R_{\mathrm{KL}}^2(m) - R_{\mathrm{KL}}^2(u)) - \Delta R_{\mathrm{KL}}^2(m|t)}{1 - \bar{R}_{\mathrm{KL}}^2(u)}$$
$$= \frac{(\Delta R_{\mathrm{KL}}^2(u, t) - R_{\mathrm{KL}}^2(u)) - \Delta R_{\mathrm{KL}}^2(t|m)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} \tag{4}$$

$$\text{Unexplained deviance}: \frac{1 - R_{\mathrm{KL}}^2(m, t)}{1 - \bar{R}_{\mathrm{KL}}^2(u)} \tag{5}$$

Previously, adjustments were made to avoid nonsensical negative values when ($\bar{R}_{\mathrm{KL}}^2(u)) > R_{\mathrm{KL}}^2(u, t)$, when ($\bar{R}_{\mathrm{KL}}^2(u)) > R_{\mathrm{KL}}^2(m)$ and/or when ($\bar{R}_{\mathrm{KL}}^2(u, t)) > R_{\mathrm{KL}}^2(m, t)$ (Shipley, 2014). However, negative values for the pure trait-based filtering effects can still occur when $R_{\mathrm{KL}}^2(m, t) < R_{\mathrm{KL}}^2(m)$. This might occur by chance if trait-based filtering does not play a significant role and the permuted trait values accidentally provide a better average fit. To address this, the definition of $R_{\mathrm{KL}}^2(m, t)$ was modified to $R_{\mathrm{KL}}^2(m, t) = \max(R_{\mathrm{KL}}^2(m, t), R_{\mathrm{KL}}^2(m))$. The R functions needed to perform CATS were obtained from Shipley et al. (2016).

We used a permutation test to determine for which plots the trait constraints contained any relevant information regarding the observed relative abundances, and thus where trait-based filtering presumably played a role in shaping the local community (Shipley, 2010b). Hence,

the following error function was computed: $\sum_{i=1}^{S} o_i \cdot \log\left(\frac{p_i}{q_i}\right)$. This error function, which is structurally similar to the KL-divergence, measures the divergence between the predicted values and a prior distribution, scaled by the observed probability. If the information contained in the constraints is relevant for the observed abundances, then it is likely that the predicted probability (p_i) will differ from the prior (q_i). In that case, the value of the statistic will be (relatively) large. The hypothesis testing algorithm as proposed by Shipley (2010b) calculates the statistic using the unpermutated trait values and then, in an iterative process, permutes the trait values and recalculates it. Then, the proportion of times that the permutated statistic is larger than the unpermutated one is determined. This proportion is considered an estimate of "the true probability of observing at least this degree of improvement in fit due to the stated constraints under the null hypothesis" (Shipley, 2010b). Small values indicate that the number of permuted datasets that provide a better fit is low as well, suggesting that the constraints do contain information. Trait values were permuted 300 times.

Additionally, the CATS model determines the strength of association between each trait and the observed relative abundances of species by calculating Lambda values (Shipley et al., 2011). A negative value indicates a negative relationship between the probability of occurrence of species and their corresponding trait value, potentially suggesting that lower trait values result in higher species dominance. Conversely, a positive value indicates a positive relationship. The higher or lower a value is, the stronger the relationship.

2.4. CATS application

We applied the CATS framework to each dataset in four configurations: at both species and genus level, using either local or TRY values. To ensure a valid comparison, we included only species for which trait data were available from both sources.

For each dataset, we Z-score normalised the TRY values separately, after selecting the local species. To avoid negative values and values of 0, which are unsuitable for the CATS algorithm, the absolute value of the lowest value plus a small constant (0.01) was added afterwards. We did the same for the local values. Standardising the trait values ensures that they will be approximately in the same range, allowing for a comparison of their lambda values. Smaller values tend to lead to an inflation of the lambda values (results not shown).

We assessed the difference between the TRY and local values by calculating correlation coefficients between the CWMs of each trait in each dataset (Table D.7). Furthermore, we explored the impact of imputation on trait data quality by comparing the correlation coefficients between the local and TRY trait data at species level, for each trait in each dataset, separately for the imputed and non-imputed trait data. We used a paired t-test to test the significance of the difference.

We evaluated the impact of taxonomic resolution and trait data origin by analysing the fit (R_{KL}^2, Eq. (1)) of each step of the CATS framework using Major Axis Regression (MA). This method is appropriate when there is measurement error in both variables, no clear distinction between predictor and response variables, and a bivariate normal distribution (Legendre and Legendre, 2012; Legendre, 2022). To determine whether the use of TRY values could potentially lead to different conclusions, plots where trait-based filtering significantly contributed to community assembly (according to the permutation test for the metacommunity-trait (m,t) model) were extracted. The impact of the origin of the trait values on the outcome of the permutation test was then tested using MA regressions.

Potential differences in model fit are likely driven by differences in the summarised trait values. To investigate this, we examined the relationship between the differences in the output of the trait (u,t) model and the mean differences in CWM and CWV for plots that were significant when analysed using local data but not with TRY values. Linear mixed models were used to assess these relationships, with the study as a random effect, using the lmerTest package version

3.1.3 (Kuznetsova et al., 2017). To determine the optimal structure of the fixed effects, we used the Akaike Information Criterion (AIC), which seeks to minimise information loss (Anderson, 2007). The model including only the mean difference in CWV provided a significantly better fit, compared to the model with only the CWM (ΔAIC: 61.69). Including both CWV and CWM, with and without interaction, did not improve model fit.

The influence of species richness on the outcomes of the decomposed model was evaluated, separately for plots undergoing significant trait-based filtering and for all plots together. To account for the under-dispersed nature of the data, which was constrained between 0 and 1, linear mixed models with a quasibinomial family link were employed for all but the joint trait and dispersal mass effect using the glmmPQL function from the MASS package version 7.3.60 (Venables and Ripley, 2002). For the latter, the lmerTest package was used with the standard Gaussian family link (Kuznetsova et al., 2017), as these values were not constrained between 0 and 1. To determine the best structure of the random effects, models were also fitted with the lmerTest package, since the glmmPQL function does not calculate the AIC nor a similar metric.

For the significant plots, including the random slope improved the model for the joint and unexplained effects (ΔAIC: 10.996; ΔAIC: 8.479, respectively). For all plots together, the random slope improved the model for trait-based filtering, joint, and the unexplained effects (ΔAIC: 15.05; ΔAIC: 6.1; and ΔAIC: 40.513, respectively).

Lastly, we assessed the robustness, or consistency, of CATS performance, across different biomes. First, we tested whether the total amount of variance explained by the model (1-unexplained effects) differed between biomes. Second, we assessed whether observed patterns between model performance and species richness were consistent across biome types. This was done using a linear mixed model as before with a quasibinomial family link but now including a random slope to test whether biome type as random effect significantly improved model fit. A lack of improvement would suggest that the observed patterns are independent of biome type, whereas a significant improvement would indicate variation in these patterns across biomes. The best structure of the random effects of the biome type was determined with the Gaussian family link.

2.5. Case study — Rolhauser dataset

The Rolhauser was the largest amongst the collected datasets in regards to the number of plots (Rolhauser et al., 2021a). Moreover, the original study collected data on several environmental variables and quantified their relationships with the functional traits (Rolhauser et al., 2021b). We therefore used this dataset for a more in-depth comparison. We decided to focus solely on the traits' relationship with MAT, as they found the strongest relationship between MAT and height. We tested whether our results were congruent with theirs by comparing the relationship between the importance of each trait in each plot (the lambda values) and the Rolhauser MAT data, with the relationships they found.

All analyses were performed in R version 4.3.0 (R Core Team, 2021). See Table E.1 for a complete list of the used packages.

3. Results

3.1. The impact of the origin of trait values

In total, 790 plots were analysed at the species level using the CATS framework (Table A.1). The origin of the trait values influenced the outcome of the model assessing the effect of trait-based filtering (u,t) and the model incorporating both traits and meta-community prior (m,t) (Fig. 2). Relationships between the prediction fit using either local or TRY values were positive (slope: 1.059, 95% CI: 1.017–1.103, R^2: 0.749; slope: 0.977, 95% CI: 0.941–1.015, R^2: 0.774, respectively).

However, there were differences at plot level, both positive and negative, indicating that in some cases the use of TRY values improved the fit, while it led to a decrease in others. Correlations between the CWMs of local and imputed TRY values were significantly lower than with non-imputed data (mean difference = 0.31, paired t-test, p < 0.0001) (Fig. C5).

The origin of the trait values did not substantially impact the estimated model bias (u) and the estimated effect of dispersal mass effects (m). Generally speaking, the local trait and TRY data produced similar results, with strong, positive relationships between the fits (slope: 1.027, 95% CI: 1.019–1.035, R^2: 0.988; slope: 0.977, 95% CI: 0.941–1.015, R^2: 0.992, respectively; see Table D.1 for further statistical details). Similar patterns were observed for the comparison at genus level (Fig. C1, Table D.4).

The origin of the trait values impacted the outcome of the permutation tests as well, leading to different plots being identified as having their community structures significantly influenced by trait-based filtering, hereafter referred to as 'significant plots' (Fig. 3). Of the 790 total plots analysed, 117 plots were only significant with the local data, 31 plots were only significant with TRY data and 45 plots were significant with both datasets (Table C.1).

Analyses of the decomposition outcomes for significant plots with the local trait values (n = 162) showed that the contribution of trait-based filtering was severely underestimated with the TRY data (slope: 1.805, 95% CI: 1.332-2.596, R^2: 0.188). In addition, the estimated relative contribution of dispersal mass effects varied greatly between plots, although the slope did not significantly differ from 1 (slope: 1.033, 95% CI: 0.855–1.250, R^2: 0.406; see Table D.2 for further statistical details).

For the plots that were only significant with the local trait values, there was a significant positive relationship between the difference in the fit of the trait model (u,t) and mean difference in CWV (slope: 0.083, t-value: 8.454, p-value < 0.0001).

3.2. Performance across taxonomic levels, biomes, and along a species richness gradient

A comparison of the fit of the data analysed at the species level versus the same data analysed at genus level, showed strong, significantly positive relationships but a substantial amount of variation around the fitted lines (Fig. C2). The weakest relationship was between the fit of the (m,t) model (slope: 1.094, 95% CI: 1.059–1.131, R^2: 0.816), followed by the (m) model (slope: 1.09, 95% CI: 1.059–1.122, R^2: 0.851; see Table D.5 for further statistical details).

For the significant plots (n = 162), analysed with the local trait data, a negative relationship was observed between the relative importance of trait-based filtering and species richness (slope: −0.056, t-value: −6.784, p-value < 0.0001). Consequently, the unexplained deviance increased (slope: 0.171, t-value: 3.528, p-value < 0.001; see Table D.3 for further statistical details) (Fig. 4). Slopes are reported on the logit scale, except for the joint effect. Species richness itself was not a constraint for the occurrence of trait-based filtering, as non-significant plots were distributed along the same range (results not shown).

For all plots together, the same relationships were observed for trait-based filtering (Fig. C3) (slope: −0.058, t-value: −5.627, p-value < 0.0001) and the unexplained effects (slope: 0.114, t-value: 4.339, p-value < 0.0001). Additionally, there were weak but significant negative relationships for the dispersal mass effect and the joint effect (slope: −0.013, t-value: −2.188, p-value: 0.029; slope: −0.006, t-value: −2.698, p-value: 0.03, respectively; see Table D.6 for further statistical details). Slopes are reported on the logit scale, except for the joint effect.

The total amount of variance explained by CATS significantly differed between the biomes (Fig. 5) (p-value < 0.001). Furthermore, including both a random intercept and slope significantly improved model fit (ΔAIC: 13.05, p-value < 0.001). Although the strength of the relationship between the total amount of variance explained and species richness varied between biomes, the direction was conserved (Fig. 5).
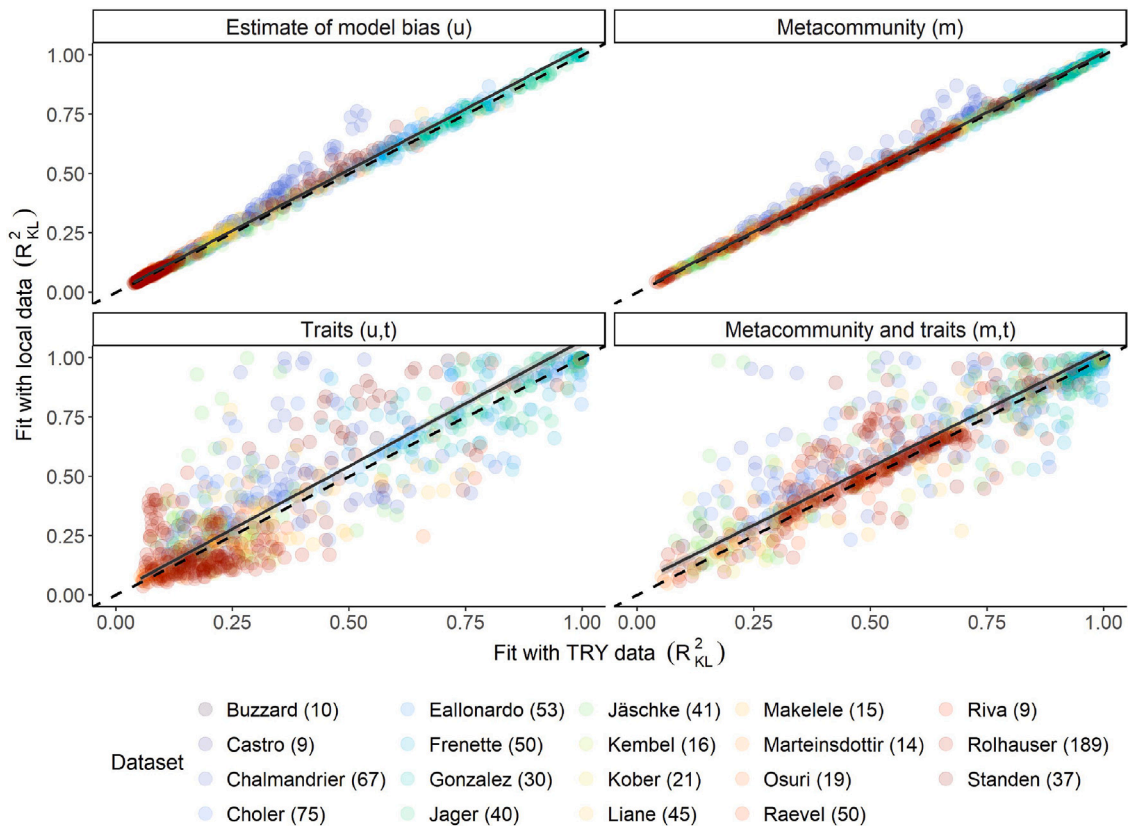
Fig. 2. The relationship between the fit of the four CATS framework models ran with local and TRY trait data. The fit of the CATS framework models was determined with the Kullback–Leibler (KL) index. Relationships between the fit of the models ran with local and TRY trait data were quantified using major axis regressions. Although all relationships were significantly positive, the origin of the trait values affected the fit of the models incorporating the trait values (u,t) and (m,t), as evident by the large spread (slope: 1.059, 95% CI: 1.017–1.103, R^2: 0.749; slope: 0.977, 95% CI: 0.941–1.015, R^2: 0.774, respectively). The fit of the models estimating model bias and the contribution of the dispersal mass effects were nearly the same (slope: 1.027, 95% CI: 1.019–1.035, R^2: 0.988; slope: 0.977, 95% CI: 0.941–1.015, R^2: 0.992, respectively; see Table D.1 for further statistical details).

3.3. Case study — Rolhauser dataset

For the majority of the plots (n = 628), including the CWMs of the traits did not significantly improve the model compared to the model using permutated trait-values and the meta-prior, indicating that trait-based filtering had little to no importance (Table C.1). A notable exception is the Rolhauser dataset, where 67 of 189 plots were significant, allowing for a more in depth exploration and a comparison between our results and the results of the original study (Rolhauser et al., 2021b).

Using the local trait values, a significant positive relationship was observed between MAT and the lambda of Height and Specific Leaf Area (SLA) (slope: 0.631, p-value < 0.0001; slope: 0.317, p-value < 0.0001, respectively) while a significant negative relationship was found with leaf carbon content (LCC) (slope: −0.19, p-value < 0.0001), similar to the original study (Fig. 6). Contrary to the original study where the relationship was non-significant, a significant negative relationship was also found for the Leaf Area (LA) (slope: −0.238, p-value < 0.0001). For the local data, there is an outlier in the LA panel. This particular plot also had the smallest CWM value, which could have caused the lambda value to increase (see also Section 2.4). For the TRY data, there was a significant positive relationship for Height (slope: 0.255, p-value < 0.0001) and a negative one for LCC (slope: −0.216, p-value < 0.0001). There was no significant relationship for LA and SLA (slope: −0.009, p-value: 0.797; slope: 0.034, p-value: 0.204, respectively).

In addition, the correlation between the local and TRY CWMs for these plots was higher for Height (r: 0.897) and LCC (r: 0.895) than for LA (r: −0.243) and SLA (r: 0.295). The correlation at the species level

was higher for SLA (r: 0.559) than for LA (r: 0.068) and LCC (r: 0.337), and similar to height (r: 0.572).

4. Discussion

In this study, we compared predictions from the Maximum Entropy based Community Assembly via Trait Selection (CATS) model (Shipley, 2010a), using either locally collected or TRY trait values, across different levels of species richness and between different taxonomic levels (species vs. genus level). A total of 19 different plant community datasets were analysed, spanning multiple countries (n = 16), habitat types (n = 18) and biomes (n = 7). At both species and genus level, a comparison of the four models within the CATS framework, using either local or TRY trait values, showed minimal differences for the estimated model bias (u) and the presumed dispersal mass effect (m). In contrast, the model using either only the trait values (u,t) or both trait and the meta-community prior for its predictions (m,t), showed both positive and negative differences in predictive accuracy. In some cases, using local data provided more accurate predictions than database-derived trait values while they resulted in a worse fit for other plots. Substantial differences were observed between the outcome of the analysis at species versus genus level taxonomy. This suggests that patterns of community assembly are inconsistent across different taxonomic scales or that the CATS model is not able to pick these up.

4.1. Effect of the origin of trait values on trait-based filtering and dispersal mass effects

In addition to assessing prediction accuracy of species relative abundances per site, permutation tests showed that ~20% of the plots (162
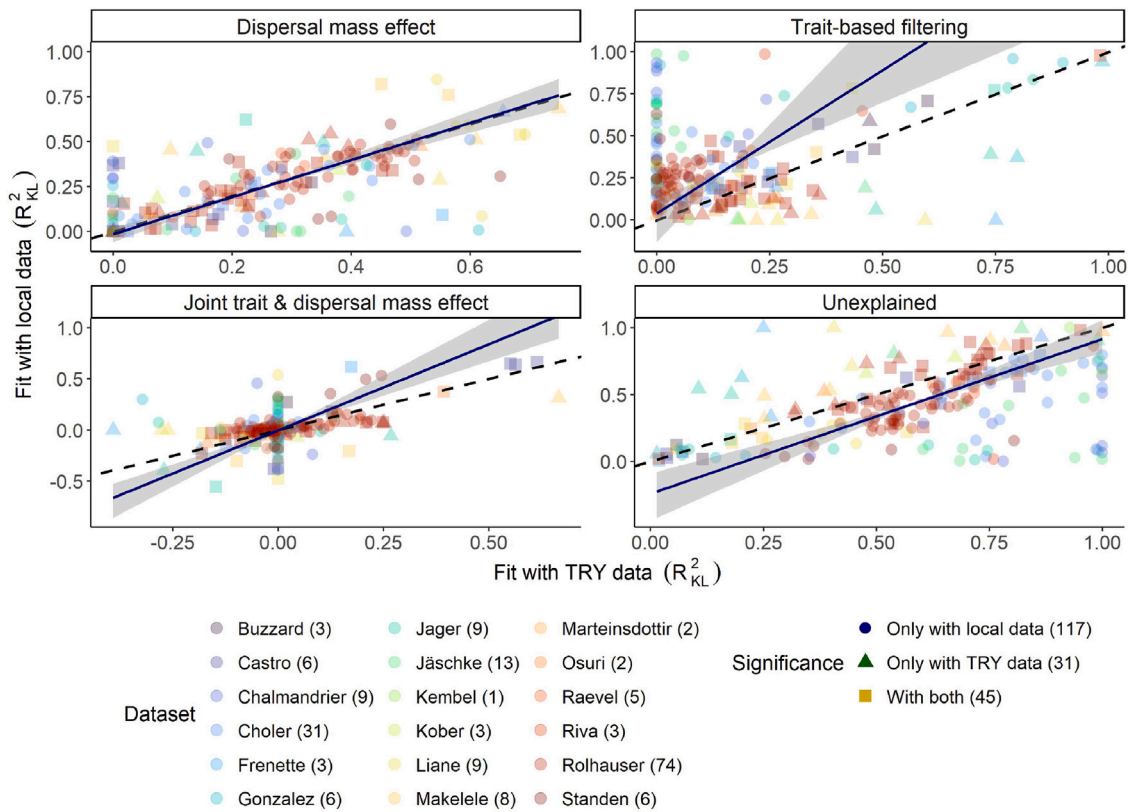
Fig. 3. The relationship between the relative contributions of the assembly mechanisms as obtained by running the CATS framework with local and TRY trait values for only the significant plots. The fit of the CATS framework models was determined with the Kullback–Leibler (KL) index. Relationships between the fit of the models ran with local and TRY trait data were quantified using major axis regressions. 117 plots were only significant with the local data (circles), 31 were only significant with TRY data (rectangles) and 45 were significant with both (squares). For the plots significant with the local values (n = 162), the relationship between the estimated relative contribution of trait-based filtering of the local and TRY values was weak (slope: 1.805, 95% CI: 1.332-2.596, R^2: 0.188), although still significantly positive. The estimated relative contribution of dispersal mass effects varied greatly between plots, although the slope did not significantly differ from 1 (slope: 1.033, 95% CI: 0.855–1.25, R^2: 0.406; see Table D.2 for further statistical details).

out of 790) had a community composition significantly related to trait-based filtering when analysed with the local traits. For those plots, only a weak relationship was found for the contribution of trait-based filtering as estimated with the local trait values and trait values derived from the TRY database. In general, the TRY values resulted in a severe underestimation compared to the first. A total of 117 plots (∼15% of total plots and ∼72% of those initially considered as being significantly influenced by trait-based filtering) were no longer deemed significant when analysed with trait values derived from TRY.

Our results indicate that the method behind obtaining trait values substantially impacts the conclusions drawn from the model. Moreover, the need for imputation worsens trait data quality as the correlations with local values were significantly lower for imputed TRY values. It is reasonable to assume that the differences in model output are driven by intraspecific variation as well, causing the relative differences between species trait means to deviate. Stochastic differences might also skew trait values, especially with relatively small sample sizes (Pérez-Llorca et al., 2018). For the plots where the outcome of the permutation tests assessing the significance of trait-based filtering differs, this raises the question of which result should be deemed correct.

Locally measured trait values are potentially better able to capture local processes. However, trait values fluctuate locally as well and differ depending on the time of year (Römermann et al., 2016). The moment of measuring might then not capture the process of trait-based filtering. For traits from a database to be used accurately, they should represent the environmental conditions of the study location and include a sample size large enough to provide a reliable average for the species. This is particularly important in extreme or disturbed environments, where trait means from databases do not accurately reflect local conditions,

especially for traits with high plasticity (Cordlandwehr et al., 2013).

Overall, the results indicate that the CATS' decomposition functions emphasise existing differences between local and database-derived trait values and their relationship with relative abundances on site level. These discrepancies warrant caution when using CATS to assess the importance of assembly mechanisms. This presents a challenge for studies interested in determining which mechanism is the most dominant on a local or regional scale, hampering the necessary understanding of ecological dynamics for nature conservation or restoration (Chase, 2007).

4.2. Case study of Rolhauser dataset — relating traits to the environment

To gain a deeper understanding, we analysed the largest dataset (Rolhauser) in more detail by focusing on the Lambda values. These values represent the proportional importance and direction of each functional trait for predicting relative abundances of species (Shipley et al., 2011). Rolhauser et al. (2021b) investigated the relationships between relative abundances of species, their functional traits, and environmental characteristics. We related the Lambda values of the local and TRY values back to the mean annual temperature (MAT), allowing for a comparison of our and their results.

Our results showed no difference in relationships between lambda values of plant height and leaf carbon content (LCC) with MAT using either local or TRY traits. In all cases, higher plant height and lower LCC were positively linked to a higher MAT. Lambda values for the Specific Leaf Area (SLA) and Leaf Area (LA), however, were inconsistent between the two. Locally measured trait values matched the results from the original study for SLA but not for LA, while the TRY values
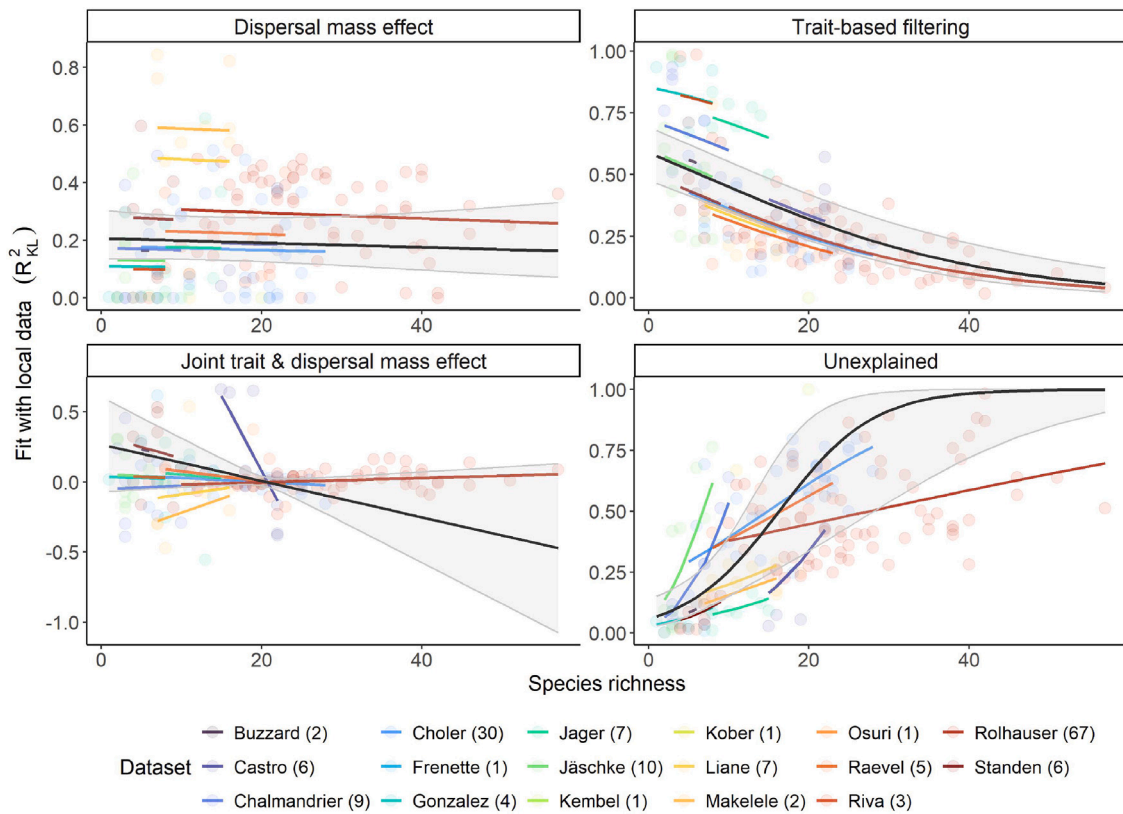
Fig. 4. The relationship between species richness and the relative contributions of the assembly mechanisms as obtained by running the CATS framework with local trait values for only the significant plots. The fit of the CATS framework models was determined with the Kullback–Leibler (KL) index. Relationships between the fit of the models ran with local trait data and species richness were quantified using linear mixed models with the dataset as random effect. There was a significant negative relationship between the contribution of trait-based filtering (slope: −0.056, t-value: −6.784, p-value < 0.0001) while the unexplained effect significantly increased (slope: 0.171, t-value: 3.528, p-value < 0.001; see Table D.3 for further statistical details). Slopes are reported on the logit scale, except for the joint effect.

aligned with the results for LA but not for SLA. There was minimal correlation between the CWMs of the local and database SLA values but the correlation at species level was comparable with other traits, such as height values. This implies that differences in CWM were more likely driven by differences in trait values for the more abundant species, and not because the TRY data were overall not representable for the Rolhauser dataset. Previous studies have shown that SLA has relatively high intraspecific variation (Siefert et al., 2015; Kichenin et al., 2013), which could be the underlying driver of these species specific differences. These differences may be driven by the higher intraspecific variation of SLA compared to other traits (Siefert et al., 2015; Kichenin et al., 2013). Although with the TRY values, the relationship between LA and MAT was correctly identified as non-significant, the correlation between the local and TRY values was negative at CWM level, and almost zero at species level. This might suggest that the relationship is non-significant because the LA TRY values do not sufficiently reflect the local situation.

The positive relationship between the difference in fit and the difference in mean CWV for all significant plots together, suggests that deviations in the CWVs are a driving factor. Potential follow-up studies could focus on exploring this relationship further by, for example, replacing the local trait values with TRY values in a stepwise manner and assessing the impact on the CWVs and predictions of CATS. While the Rolhauser dataset may serve as an example, this specific case study does support the validity of the CATS framework itself. Even without incorporating environmental data, CATS was still able to identify the same relationships for most traits as the models used in the Rolhauser study that did include it.

4.3. Community assembly across a species richness gradient

For the plots that were found to have significant trait-based filtering using local trait values, the relative importance of trait-based filtering was shown to decrease along an exponential curve with increasing species richness whereas the dispersal mass effect and the joint effects of trait and dispersal showed no significant relationship. Conversely, the unexplained effect increased with increasing species richness when using field-derived trait values. This suggests that stochastic processes are more dominant within diverse systems (Arim et al., 2023). These same patterns were observed for all plots together, although in that case, the importance of dispersal mass effects and the joint effect also significantly decreased.

We propose two potential hypotheses to explain these patterns. First, with increasing species richness, average population sizes usually decrease, increasing their vulnerability to demographic and environmental stochasticity (Lande et al., 2003). In addition, a higher species richness could potentially dilute the effect of traits specifically related to certain species-abundance relationships which CATS tries to quantify, as it increases the chance of functional redundancy (Rosenfeld, 2002). Consequently, in the most extreme case, competition might result in some species with similar trait values showing opposite patterns of abundance. Alternatively, increased species richness is often linked with niche differentiation, where (abundant) species fill different niches, thereby avoiding competition (Tang and Zhou, 2011; Zuppinger-Dingley et al., 2014). The CATS model, however, is based on the notion that in multivariate trait-space, one optimum corresponds with the highest fitness, meaning that species clustered around that average, should be more abundant in their local community (CWM-optimality hypothesis) (Shipley, 2010a). If multiple multivariate fitness
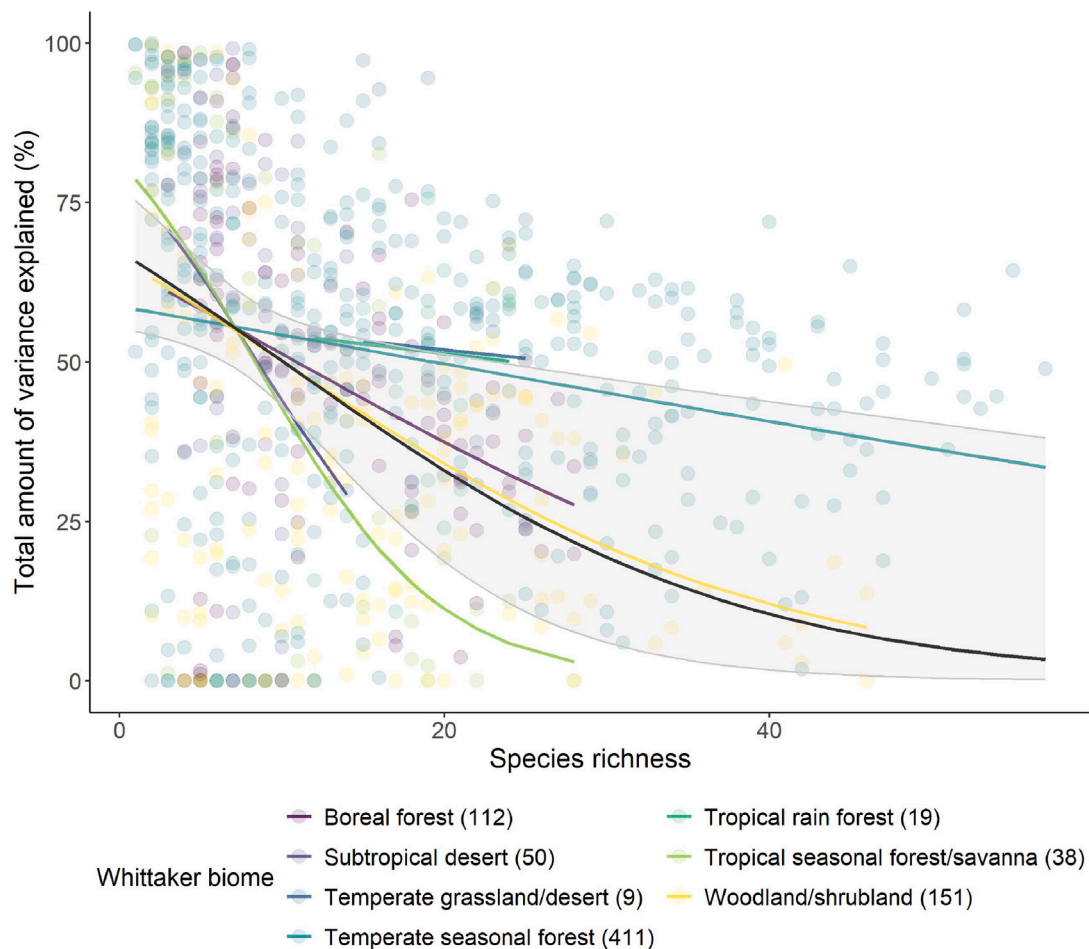
Fig. 5. Relationship between the total variance explained (%) by CATS and species richness across biomes. The relationship was quantified using linear mixed models with biome and dataset as random effects. There was a significant negative relationship between the total amount of variance explained and species richness (p-value < 0.016). Although the strength of the relationship varied between biomes, the direction remained the same.

optimums exist due to coexisting functional strategies, the CATS model might fail in its attempt to accurately predict the relative abundances. In theory, in the current application, trait-based filtering is thus only detected when trait convergence across all or most of the considered traits takes place and where there is one preferred functional strategy (Bernard-Verdier et al., 2012).

This pattern of decreased information yield from CATS at higher species richness levels is largely conserved across biomes as the direction of the relationship did not differ between them. The strength of the relationship, however, did vary. We hypothesise that these slope differences are partly due to differences in data quality. For example, the temperate grassland/desert and tropical rainforest both showed the weakest relationships. These are also the biomes less well represented in our dataset, with both only having one study with a lower amount of plots (Castro and Osuri with 9 and 19 respectively, compared to, for example, Standen with 37 plots). Furthermore, the Osuri dataset has a lower trait coverage compared to most other datasets (61%). In addition to dataset properties, it is also likely that differences in ecological processes, both at the dataset and biome level, cause the relative contributions of the assembly mechanisms to differ. For instance, the contribution of stochastic processes, or ecological drift, was found to change alongside a successional gradient (Måren et al., 2018) and to be more important in harsher environments (Silva et al., 2015). Improving dataset representation to uncover such relationships and increased trait coverage to more clearly identify signals of selection now attributed to unexplained information, will be crucial for elucidating the changing dynamics of community assembly across biomes.

For the majority of the plots, trait-based filtering was either non-significant and/or of lesser importance in comparison to estimated dispersal mass effects. It should be noted, however, that failure of accurate prediction using traits does not necessarily mean there is no trait-based filtering. It might simply mean that those traits on which filtering is acting, are not included in the model. Trait-based filtering will be underestimated if relevant traits are missing from the model. For example, floral and clonal traits were absent in all datasets, and root traits were underrepresented while in some instances, they could have provided functional information not covered by other traits (E-Vojtkó et al., 2020; Chelli et al., 2024; Aubin et al., 2024). Missing information and the ecological processes previously described, might subsequently lead to the (false) conclusion that trait-based filtering is of no, limited, or lesser importance compared to dispersal mass effects.

Additional analysis should clarify the hypothesised underlying dynamics and provide results either supporting or contradicting the conclusions drawn from the CATS framework. A possible focus for future research could be to further zoom in on the different traits separately, testing whether trait convergence or divergence occurs at a local scale, following work by Bernard-Verdier et al. (2012). The plots for which trait-based filtering is deemed non-significant in the assembly process may be the ones for which trait divergence is more prevalent. In addition, the model provides an alternative way of assessing the significance of trait-based filtering, which would further consolidate or dispute the results of the CATS framework.
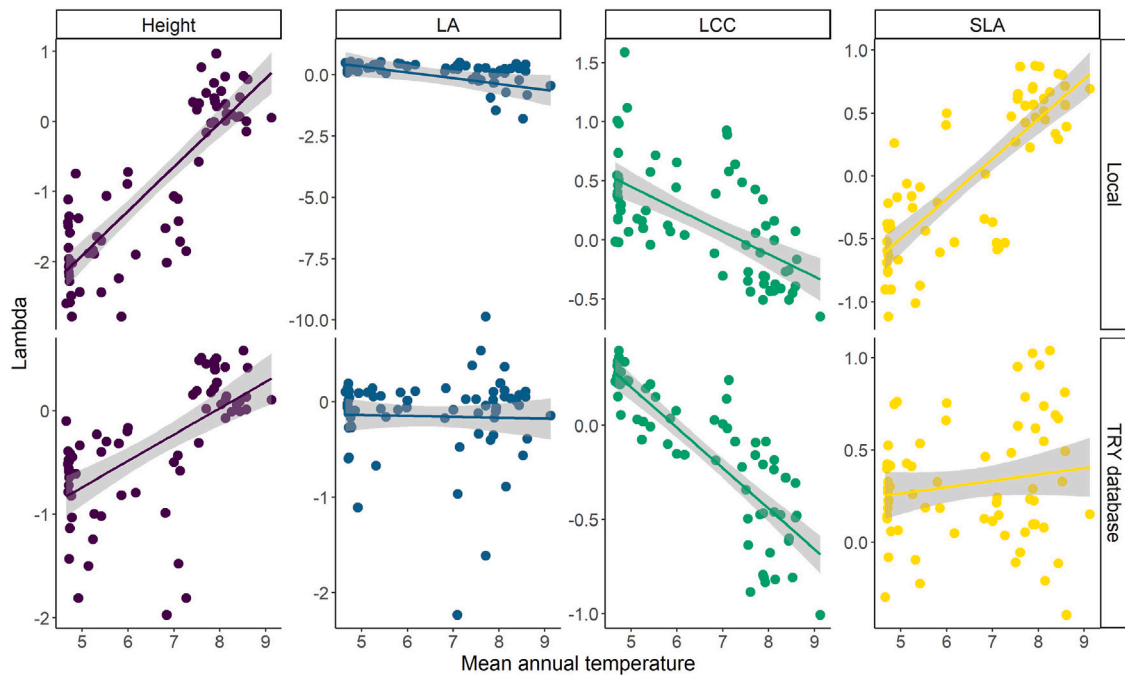
Fig. 6. The relationship between the Lambda values as obtained with the local trait values (above) and the TRY derived values (below), with the mean annual temperature from the Rolhauser dataset. For the local trait values, a significant positive relationship was observed for Height and Specific Leaf Area (SLA) (slope: 0.631, p-value < 0.0001; slope: 0.317, p-value < 0.0001, respectively) while a significant negative relationship was found with the Leaf Area (LA) and leaf carbon content (LCC) (slope: −0.238, p-value < 0.0001; slope: −0.19, p-value < 0.0001, respectively). For the TRY data, there was a significant positive relationship for Height (slope: 0.255, p-value < 0.0001) and a negative one for LCC (slope: −0.216, p-value < 0.0001). There was no significant relationship for LA and SLA (slope: −0.009, p-value: 0.797; slope: 0.034, p-value: 0.204, respectively).

4.4. Conclusion

Our results highlight the need for caution when deciding whether to use locally measured or database-derived trait values in the CATS model. Results show they can yield different outcomes, as demonstrated in our plot-level comparison. These differences might be driven by intraspecific variation, causing database values to diverge from local observations. Furthermore, imputing data that is missing can exacerbate these differences. Despite these challenges, we demonstrated that CATS was able to extract most of the same relationships between specific trait values and environmental conditions without including data on the latter, demonstrating its robustness. The amount of information obtained from the model, however, does decrease with increasing species richness. This pattern was conserved across biomes although with varying strength. We hypothesise this negative relationship reflects an increase in stochastic processes due to, on average, reduced population sizes of each species with higher species richness. Additionally, a higher species richness could correspond with increased functional trait redundancy, and thus nice overlap, leading to a disjoint between trait values and abundance. Higher species richness levels could also coincide with the emergence of multiple fitness-optima (e.g. species with similar trait values can have opposite patterns in abundance due to competition or stochastic processes). Future research should explore complementary modelling techniques, focused on trait divergence and convergence, to elucidate these processes of community assembly of more diverse, species-rich communities. These approaches could supplement the conclusions drawn from CATS, providing the necessary information needed to apply this framework to nature conservation or restoration.

CRediT authorship contribution statement

Jelyn Gerkema: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daniel E. Bunker:** Writing – review & editing, Investigation, Data curation. **Andrew M. Cunliffe:** Writing – review & editing, Software, Investigation, Data curation. **Erika Bazzato:** Writing – review & editing, Investigation, Data curation. **Michela Marignani:** Writing – review & editing, Investigation, Data curation. **Tommaso Sitzia:** Writing – review & editing, Investigation, Data curation. **Isabelle Aubin:** Writing – review & editing, Investigation, Data curation. **Stefano Chelli:** Writing – review & editing, Investigation, Data curation. **Julieta A. Rosell:** Writing – review & editing, Investigation, Data curation. **Peter Poschlod:** Writing – review & editing, Investigation, Data curation. **Josep Penuelas:** Writing – review & editing, Investigation, Data curation. **Arildo S. Dias:** Writing – review & editing, Investigation, Data curation. **Christian Rossi:** Writing – review & editing, Investigation, Data curation. **Tanvir A. Shovon:** Writing – review & editing, Investigation, Data curation. **Juan A. Campos:** Writing – review & editing, Investigation, Data curation. **Mark C. Vanderwel:** Investigation, Data curation. **Sharif A. Mukul:** Investigation, Data curation. **Bruno E.L. Cerabolini:** Investigation, Data curation. **Thomas Sibret:** Investigation, Data curation. **Bruno Hérault:** Investigation, Data curation. **Sylvain Schmitt:** Investigation, Data curation. **Pedro Higuchi:** Investigation, Data curation. **James L. Tsakalos:** Investigation, Data curation. **Decky I. Junaedi:** Investigation, Data curation. **Yun-Peng Zhao:** Investigation, Data curation. **Vanessa Minden:** Investigation, Data curation. **Ana Carolina da Silva:** Investigation, Data curation. **Tereza Mašková:** Investigation, Data curation. **Roberto Canullo:** Investigation, Data curation. **Ning Dong:** Investigation, Data curation. **Edwin T. Pos:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the main author, Jelyn Gerkema, used ChatGPT to improve sentences already written by drawing inspiration from suggestions for improvement. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103031.

Data availability

R scripts are available on the GitHub repository of Jelyn Gerkema (https://github.com/JGerkema/CATS_robustness_limitations).

References

Anderson, David R., 2007. Model Based Inference in the Life Sciences: A Primer on Evidence. Springer Science & Business Media.

Arim, Matías, Pinelli, Verónica, Rodríguez-Tricot, Lucía, Ortiz, Esteban, Illarze, Mariana, Fagúndez-Pachón, César, Borthagaray, Ana I., 2023. Chance and necessity in the assembly of plant communities: Stochasticity increases with size, isolation and diversity of temporary ponds. J. Ecol. 111 (8), 1641–1655.

Aubin, Isabelle, Deschênes, Élise, Santala, Kierann, Emilson, Erik J.S., Schoonmaker, Amanda L., McIntosh, Anne C.S., Bourgeois, Bérenger, Cardou, Françoise, Dupuch, Angélique, Handa, I. Tanya, Lapointe, Mélanie, Lavigne, Jonathan, Maheu, Audrey, Nadeau, Solange, Naeth, M. Anne, Neilson, Eric W., Wiebe, Philip A., 2024. Restoring forest ecosystem services through trait-based ecology. Environ. Rev. (4), 498–524.

Baastrup-Spohr, Lars, Sand-Jensen, Kaj, Nicolajsen, Sascha V., Bruun, Hans H., 2015. From soaking wet to bone dry: predicting plant community composition along a steep hydrological gradient. J. Veg. Sci. 26 (4), 619–630.

Bernard-Verdier, Maud, Navas, Marie-Laure, Vellend, Mark, Violle, Cyrille, Fayolle, Adeline, Garnier, Eric, 2012. Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. J. Ecol. 100 (6), 1422–1433.

Brown, Alexandra M., Warton, David I., Andrew, Nigel R., Binns, Matthew, Cassis, Gerasimos, Gibb, Heloise, 2014. The fourth-corner solution – using predictive models to understand how species traits interact with the environment. Methods Ecol. Evol. 5 (4), 344–352.

Chase, Jonathan M., 2007. Drought mediates the importance of stochastic community assembly. Proc. Natl. Acad. Sci. 104 (44), 17430–17434.

Chelli, Stefano, Klimešová, Jitka, Tsakalos, James L., Puglielli, Giacomo, 2024. Unravelling the clonal trait space: Beyond above-ground and fine-root traits. J. Ecol. 112, 730–740.

Colin Cameron, A., Windmeijer, Frank A.G., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. J. Econometrics 77 (2), 329–342.

Cordlandwehr, Verena, Meredith, Rebecca L., Ozinga, Wim A., Bekker, Renée M., van Groenendael, Jan M., Bakker, Jan P., 2013. Do plant traits retrieved from a database accurately predict on-site measurements? J. Ecol. 101 (3), 662–670.

Cunillera-Montcusí, David, Arim, Matías, Gascón, Stéphanie, Tornero, Irene, Sala, Jordi, Boix, Dani, Borthagaray, Ana I., 2020. Addressing trait selection patterns in temporary ponds in response to wildfire disturbance and seasonal succession. J. Anim. Ecol. 89 (9), 2134–2144.

Cunliffe, Andrew M., Anderson, Karen, Boschetti, Fabio, Brazier, Richard E., Graham, Hugh A., Myers-Smith, Isla H., Astor, Thomas, Boer, Matthias M., Calvo, Leonor G., Clark, Patrick E., Cramer, Michael D., Encinas-Lara, Miguel S., Escarzaga, Stephen M., Fernández-Guisuraga, José M., Fisher, Adrian G., Gdulová, Kateřina, Gillespie, Breahna M., Griebel, Anne, Hanan, Niall P., Hanggito, Muhammad S., Haselberger, Stefan, Havrilla, Caroline A., Heilman, Phil, Ji, Wenjie, Karl, Jason W., Kirchhoff, Mario, Kraushaar, Sabine, Lyons, Mitchell B., Marzolff, Irene, Mauritz, Marguerite E., McIntire, Cameron D., Metzen, Daniel, Méndez-Barroso, Luis A., Power, Simon C., Prošek, Jiří, Sanz-Ablanedo, Enoc, Sauer, Katherine J., Schulze-Brüninghoff, Damian, Šímová, Petra, Sitch, Stephen,

Smit, Julian L., Steele, Caiti M., Suárez-Seoane, Susana, Vargas, Sergio A., Villarreal, Miguel, Visser, Fleur, Wachendorf, Michael, Wirnsberger, Hannes, Wojcikiewicz, Robert, 2022. Global application of an unoccupied aerial vehicle photogrammetry protocol for predicting aboveground biomass in non-forest ecosystems. Remote. Sens. Ecol. Conserv. 8 (1), 57–71.

Dewar, Roderick C., Porté, Annabel, 2008. Statistical mechanics unifies different ecological patterns. J. Theoret. Biol. 251 (3), 389–403.

E-Vojtkó, Anna, de Bello, Francesco, Durka, Walter, Kühn, Ingolf, Götzenberger, Lars, 2020. The neglected importance of floral traits in trait-based plant community assembly. J. Veg. Sci. 31 (4), 529–539.

Elliott, Alan, Hyam, Roger, Ulate, William, Watson, Mark, Anderson, Gregory, Andrella, Giovani Carlos, Anguiano, Marco, Ardi, Wisnu Handoyo, Atkins, Hannah, Atwood, John J., Aubriot, Xavier, Baker, William, Balan, Anoop Puthuparampil, Bartolucci, Fabrizio, Berendsohn, Walter G., Bonifacino, Mauricio, Borges, Leonardo, Brinda, John C., Brown, Gillian, Bruneau, Anne, Cardoso, Domingos, Carvalho, Catarina Silva de, Chiarini, Franco, Couvreur, Thomas, Crowley, Dan, Duan, Lei, Durán, Juan David Tovar, Egan, Ashley N., Eggli, Urs, Escallón, Eugenio Valderrama, Estrella, Manuel de la, Falcão, Marcus, Franck, Alan, Fritsch, Peter, Gagnon, Edeline, Galasso, Gabriele, Galdo, Gianpietro Giusso del, Giacomin, Leandro, Girmansyah, Deden, Gissi, Danilo Soares, Gonzáles, Paúl, Gouvêa, Yuri, Govaerts, Rafaël Herman Anna, Gregório, Bernarda, Guerrero, Pablo C., Haegi, Laurence, Haevermans, Thomas, Harris, David J., Hilario, Robin Fernandez, Hopkins, Helen Fortune, Hua-Feng, Hughes, Mark, Iganci, João Rando, J.G., Jara-Muñoz, Orlando Adolfo, Javadi, Firouzeh, Jiang, Kai-Wen, Jiménez-Mejías, Pedro, Jordão, Lucas Sá Barreto, Keller, Héctor, Klitgaard, Bente, Klopper, R.R., Korotkova, Nadja, Köhler, Matias, Larridon, Isabel, Lavandero, Nicolás, Leong-Skornickova, Jana, Lewis, Gwilym, Liede-Schumann, Sigrid, Lima, Alexandre Gibau de, Martine, Chris, Maslin, Bruce, Mering, Sabine von, Metzing, Detlev, Meve, Ulrich, Miller, Joseph T., Moeller, Michael, Moonlight, Peter, Morales, J. Francisco, Moreira-Munoz, Andres, Murphy, Daniel, Mustaqim, Wendy A., Müller, Andreas, Nelson, Charles, Nersesyan, Anush, Newman, Mark, Orejuela, Andres, Pandey, Arun, Pellegrini, Marco O.O., Pennington, Toby, Pinto, Rafael, Pirie, Michael D., Poulsen, Axel Dalberg, Puglisi, Carmen, Quintanar, Alejandro, Rakotonirina, Nivohenintsoa, Ramos, Gustavo, Rapini, Alessandro, Rodda, Michele, Rodrigues, Rodrigo Schütz, Roux, Margaretha Marianne Le, Sampaio, Valéria da Silva, Santilli, Ludovica, Schlumpberger, Boris, Schneider, Julio V., Silva, Rafael Costa, Simon, Marcelo, Smith, Stacey, Sokoloff, Dmitry, Sotuyo, Solange, Stirton, Charles, Straub, Shannon, Strijk, Joeri Sergej, Sánchez, Daniel, Tebbitt, Mark, Tepe, Eric, Thomas, Daniel Caspar, Thulin, Mats, Trad, Rafaela, Vorontsova, Maria (Bat) S., Wilding, Nicholas, Wilson, Hannah, Zamora-Tavares, Pilar, Zhang, Rong, Zizka, Georg, 2023. World flora online plant list december 2023 (2023-12).

Frenette-Dussault, Cédric, Shipley, Bill, Meziane, Driss, Hingrat, Yves, 2013. Trait-based climate change predictions of plant community structure in arid steppes. J. Ecol. 101 (2), 484–492.

Grace, James B., 1999. The factors controlling species density in herbaceous plant communities: an assessment. Perspect. Plant Ecol. Evol. Syst. 2 (1), 1–28.

Haegeman, Bart, Loreau, Michel, 2009. Trivial and non-trivial applications of entropy maximization in ecology: a reply to Shipley. Oikos 118 (8), 1270–1278.

Harabiš, Filip, Dolný, Aleš, 2018. Military training areas as refuges for threatened dragonfly species: Effect of spatial isolation and military activity. Biol. Cons. 217, 28–35.

Harte, John, 2011. Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics. Oxford University Press.

He, Yun-Yun, Srisombut, Kwansupa, Xing, Ding-Liang, Swenson, Nanthan G., Asefa, Mengesha, Cao, Min, Song, Xiao-Yang, Wen, Han-Dong, Yang, Jie, 2022. Ontogenetic trait variation and metacommunity effects influence species relative abundances during tree community assembly. Plant Divers. 44 (4), 360–368.

Hijmans, Robert J., Barbosa, Márcia, Ghosh, Aniruddha, Mandel, Alex, 2024. geodata: Download geographic data. R package version 0.6-2.

Janssen, Philippe, Fuhr, Marc, Cateau, Eugénie, Nusillard, Benoit, Bouget, Christophe, 2017. Forest continuity acts congruently with stand maturity in structuring the functional composition of saproxylic beetles. Biol. Cons. 205, 1–10.

Jeliazkov, Alienor, Mijatovic, Darko, Chantepie, Stéphane, Andrew, Nigel, Arlettaz, Raphaël, Barbaro, Luc, Barsoum, Nadia, Bartonova, Alena, Belskaya, Elena, Bonada, Núria, Brind'Amour, Anik, Carvalho, Rodrigo, Castro, Helena, Chmura, Damian, Choler, Philippe, Chong-Seng, Karen, Cleary, Daniel, Cormont, Anouk, Cornwell, William, de Campos, Ramiro, de Voogd, Nicole, Doledec, Sylvain, Drew, Joshua, Dziock, Frank, Eallonardo, Anthony, Edgar, Melanie J., Farneda, Fábio, Hernandez, Domingo Flores, Frenette-Dussault, Cédric, Fried, Guillaume, Gallardo, Belinda, Gibb, Heloise, Gonçalves-Souza, Thiago, Higuti, Janet, Humbert, Jean-Yves, Krasnov, Boris R., Saux, Eric Le, Lindo, Zoe, Lopez-Baucells, Adria, Lowe, Elizabeth, Marteinsdottir, Bryndis, Martens, Koen, Meffert, Peter, Mellado-Díaz, Andres, Menz, Myles H.M., Meyer, Christoph F.J., Miranda, Julia Ramos, Mouillot, David, Ossola, Alessandro, Pakeman, Robin, Pavoine, Sandrine, Pekin, Burak, Pino, Joan, Pocheville, Arnaud, Pomati, Francesco, Poschlod, Peter, Prentice, Honor C., Purschke, Oliver, Raevel, Valerie, Reitalu, Triin, Renema, Willem, Ribera, Ignacio, Robinson, Natalie, Robroek, Bjorn, Rocha, Ricardo, Shieh, Sen-Her, Spake, Rebecca, Staniaszek-Kik, Monika, Stanko, Michal, Tejerina-Garro, Francisco Leonardo,

Braak, Cajo ter, Urban, Mark C., Klink, Roel van, Villéger, Sébastien, Wegman, Ruut, Westgate, Martin J., Wolff, Jonas, Żarnowiec, Jan, Zolotarev, Maxim, Chase, Jonathan M., 2020. A global database for metacommunity ecology, integrating species, traits, environment and space. Sci. Data 7 (1), 6, Knowledge cutoff: 2022-01.

Joswig, Julia S., Kattge, Jens, Kraemer, Guido, Mahecha, Miguel D., Rüger, Nadja, Schaepman, Michael E., Schrodt, Franziska, Schuman, Meredith C., 2023. Imputing missing data in plant traits: A guide to improve gap-filling. Glob. Ecol. Biogeogr. 32 (8), 1395–1408.

Kattge, Jens, Bönisch, Gerhard, Díaz, Sandra, Lavorel, Sandra, Prentice, Iain Colin, Leadley, Paul, Tautenhahn, Susanne, Werner, Gijsbert D.A., Aakala, Tuomas, Abedi, Mehdi, Acosta, Alicia T.R., Adamidis, George C., Adamson, Kairi, Aiba, Masahiro, Albert, Cécile H., Alcántara, Julio M., Alcázar C, Carolina, Aleixo, Izabela, Ali, Hamada, Amiaud, Bernard, Ammer, Christian, Amoroso, Mariano M., Anand, Madhur, Anderson, Carolyn, Anten, Niels, Antos, Joseph, Apgaua, Deborah Mattos Guimarães, Ashman, Tia-Lynn, Asmara, Degi Harja, Asner, Gregory P., Aspinwall, Michael, Atkin, Owen, Aubin, Isabelle, Baastrup-Spohr, Lars, Bahalkeh, Khadijeh, Bahn, Michael, Baker, Timothy, Baker, William J., Bakker, Jan P., Baldocchi, Dennis, Baltzer, Jennifer, Banerjee, Arindam, Baranger, Anne, Barlow, Jos, Barneche, Diego R., Baruch, Zdravko, Bastianelli, Denis, Battles, John, Bauerle, William, Bauters, Marijn, Bazzato, Erika, Beckmann, Michael, Beeckman, Hans, Beierkuhnlein, Carl, Bekker, Renee, Belfry, Gavin, Belluau, Michael, Beloiu, Mirela, Benavides, Raquel, Benomar, Lahcen, Berdugo-Lattke, Mary Lee, Berenguer, Erika, Bergamin, Rodrigo, Bergmann, Joana, Bergmann Carlucci, Marcos, Berner, Logan, Bernhardt-Römermann, Markus, Bigler, Christof, Bjorkman, Anne D., Blackman, Chris, Blanco, Carolina, Blonder, Benjamin, Blumenthal, Dana, Bocanegra-González, Kelly T., Boeckx, Pascal, Bohlman, Stephanie, Böhning-Gaese, Katrin, Boisvert-Marsh, Laura, Bond, William, Bond-Lamberty, Ben, Boom, Arnoud, Boonman, Coline C.F., Bordin, Kauane, Boughton, Elizabeth H., Boukili, Vanessa, Bowman, David M.J.S., Bravo, Sandra, Brendel, Marco Richard, Broadley, Martin R., Brown, Kerry A., Bruelheide, Helge, Brumnich, Federico, Bruun, Hans Henrik, Bruy, David, Buchanan, Serra W., Bucher, Solveig Franziska, Buchmann, Nina, Buitenwerf, Robert, Bunker, Daniel E., Bürger, Jana, Burrascano, Sabina, Burslem, David F.R.P., Butterfield, Bradley J., Byun, Chaeho, Marques, Marcia, Scalon, Marina C., Caccianiga, Marco, Cadotte, Marc, Cailleret, Maxime, Camac, James, Camarero, Jesús Julio, Campany, Courtney, Campetella, Giandiego, Campos, Juan Antonio, Cano-Arboleda, Laura, Canullo, Roberto, Carbognani, Michele, Carvalho, Fabio, Casanoves, Fernando, Castagneyrol, Bastien, Catford, Jane A., Cavender-Bares, Jeannine, Cerabolini, Bruno E.L., Cervellini, Marco, Chacón-Madrigal, Eduardo, Chapin, Kenneth, Chapin, F. Stuart, Chelli, Stefano, Chen, Si-Chong, Chen, Anping, Cherubini, Paolo, Chianucci, Francesco, Choat, Brendan, Chung, Kyong-Sook, Chytrý, Milan, Ciccarelli, Daniela, Coll, Lluís, Collins, Courtney G., Conti, Luisa, Coomes, David, Cornelissen, Johannes H.C., Cornwell, William K., Corona, Piermaria, Coyea, Marie, Craine, Joseph, Craven, Dylan, Cromsigt, Joris P.G.M., Csecserits, Anikó, Cufar, Katarina, Cuntz, Matthias, da Silva, Ana Carolina, Dahlin, Kyla M., Dainese, Matteo, Dalke, Igor, Dalle Fratte, Michele, Dang-Le, Anh Tuan, Danihelka, Jirí, Dannoura, Masako, Dawson, Samantha, de Beer, Arend Jacobus, De Frutos, Angel, De Long, Jonathan R., Dechant, Benjamin, Delagrange, Sylvain, Delpierre, Nicolas, Derroire, Géraldine, Dias, Arildo S., Diaz-Toribio, Milton Hugo, Dimitrakopoulos, Panayiotis G., Dobrowolski, Mark, Doktor, Daniel, Dřevojan, Pavel, Dong, Ning, Dransfield, John, Dressler, Stefan, Duarte, Leandro, Ducouret, Emilie, Dullinger, Stefan, Durka, Walter, Duursma, Remko, Dymova, Olga, E-Vojtkó, Anna, Eckstein, Rolf Lutz, Ejtehadi, Hamid, Elser, James, Emilio, Thaise, Engemann, Kristine, Erfanian, Mohammad Bagher, Erfmeier, Alexandra, Esquivel-Muelbert, Adriane, Esser, Gerd, Estiarte, Marc, Domingues, Tomas F., Fagan, William F., Fagúndez, Jaime, Falster, Daniel S., Fan, Ying, Fang, Jingyun, Farris, Emmanuele, Fazlioglu, Fatih, Feng, Yanhao, Fernandez-Mendez, Fernando, Ferrara, Carlotta, Ferreira, Joice, Fidelis, Alessandra, Finegan, Bryan, Firn, Jennifer, Flowers, Timothy J., Flynn, Dan F.B., Fontana, Veronika, Forey, Estelle, Forgiarini, Cristiane, François, Louis, Frangipani, Marcelo, Frank, Dorothea, Frenette-Dussault, Cedric, Freschet, Grégoire T., Fry, Ellen L., Fyllas, Nikolaos M., Mazzochini, Guilherme G., Gachet, Sophie, Gallagher, Rachael, Ganade, Gislene, Ganga, Francesca, García-Palacios, Pablo, Gargaglione, Verónica, Garnier, Eric, Garrido, Jose Luis, de Gasper, André Luís, Gea-Izquierdo, Guillermo, Gibson, David, Gillison, Andrew N., Giroldo, Aelton, Glasenhardt, Mary-Claire, Gleason, Sean, Gliesch, Mariana, Goldberg, Emma, Göldel, Bastian, Gonzalez-Akre, Erika, Gonzalez-Andujar, Jose L., González-Melo, Andrés, González-Robles, Ana, Graae, Bente Jessen, Granda, Elena, Graves, Sarah, Green, Walton A., Gregor, Thomas, Gross, Nicolas, Guerin, Greg R., Günther, Angela, Gutiérrez, Alvaro G., Haddock, Lillie, Haines, Anna, Hall, Jefferson, Hambuckers, Alain, Han, Wenxuan, Harrison, Sandy P., Hattingh, Wesley, Hawes, Joseph E., He, Tianhua, He, Pengcheng, Heberling, Jacob Mason, Helm, Aveliina, Hempel, Stefan, Hentschel, Jörn, Hérault, Bruno, Hereş, Ana-Maria, Herz, Katharina, Heuertz, Myriam, Hickler, Thomas, Hietz, Peter, Higuchi, Pedro, Hipp, Andrew L., Hirons, Andrew, Hock, Maria, Hogan, James Aaron, Holl, Karen, Honnay, Olivier, Hornstein, Daniel, Hou, Enqing, Hough-Snee, Nate, Hovstad, Knut Anders, Ichie, Tomoaki, Igić, Boris, Illa, Estela, Isaac, Marney,

Ishihara, Masae, Ivanov, Leonid, Ivanova, Larissa, Iversen, Colleen M., Izquierdo, Jordi, Jackson, Robert B., Jackson, Benjamin, Jactel, Hervé, Jagodzinski, Andrzej M., Jandt, Ute, Jansen, Steven, Jenkins, Thomas, Jentsch, Anke, Jespersen, Jens Rasmus Plantener, Jiang, Guo-Feng, Johansen, Jesper Liengaard, Johnson, David, Jokela, Eric J., Joly, Carlos Alfredo, Jordan, Gregory J., Joseph, Grant Stuart, Junaedi, Decky, Junker, Robert R., Justes, Eric, Kabzems, Richard, Kane, Jeffrey, Kaplan, Zdenek, Kattenborn, Teja, Kavelenova, Lyudmila, Kearsley, Elizabeth, Kempel, Anne, Kenzo, Tanaka, Kerkhoff, Andrew, Khalil, Mohammed I., Kinlock, Nicole L., Kissling, Wilm Daniel, Kitajima, Kaoru, Kitzberger, Thomas, Kjøller, Rasmus, Klein, Tamir, Kleyer, Michael, Klimešová, Jitka, Klipel, Joice, Kloeppel, Brian, Klotz, Stefan, Knops, Johannes M.H., Kohyama, Takashi, Koike, Fumito, Kollmann, Johannes, Komac, Benjamin, Komatsu, Kimberly, König, Christian, Kraft, Nathan J.B., Kramer, Koen, Kreft, Holger, Kühn, Ingolf, Kumarathunge, Dushan, Kuppler, Jonas, Kurokawa, Hiroko, Kurosawa, Yoko, Kuyah, Shem, Laclau, Jean-Paul, Lafleur, Benoit, Lallai, Erik, Lamb, Eric, Lamprecht, Andrea, Larkin, Daniel J., Laughlin, Daniel, Le Bagousse-Pinguet, Yoann, le Maire, Guerric, le Roux, Peter C., le Roux, Elizabeth, Lee, Tali, Lens, Frederic, Lewis, Simon L., Lhotsky, Barbara, Li, Yuanzhi, Li, Xine, Lichstein, Jeremy W., Liebergesell, Mario, Lim, Jun Ying, Lin, Yan-Shih, Linares, Juan Carlos, Liu, Chunjiang, Liu, Daijun, Liu, Udayangani, Livingstone, Stuart, Llusià, Joan, Lohbeck, Madelon, López-García, Álvaro, Lopez-Gonzalez, Gabriela, Lososová, Zdeňka, Louault, Frédérique, Lukács, Balázs A., Lukeš, Petr, Luo, Yunjian, Lussu, Michele, Ma, Siyan, Maciel Rabelo Pereira, Camilla, Mack, Michelle, Maire, Vincent, Mäkelä, Annikki, Mäkinen, Harri, Malhado, Ana Claudia Mendes, Mallik, Azim, Manning, Peter, Manzoni, Stefano, Marchetti, Zuleica, Marchino, Luca, Marcilio-Silva, Vinicius, Marcon, Eric, Marignani, Michela, Markesteijn, Lars, Martin, Adam, Martínez-Garza, Cristina, Martínez-Vilalta, Jordi, Mašková, Tereza, Mason, Kelly, Mason, Norman, Massad, Tara Joy, Masse, Jacynthe, Mayrose, Itay, McCarthy, James, McCormack, M. Luke, McCulloh, Katherine, McFadden, Ian R., McGill, Brian J., McPartland, Mara Y., Medeiros, Juliana S., Medlyn, Belinda, Meerts, Pierre, Mehrabi, Zia, Meir, Patrick, Melo, Felipe P.L., Mencuccini, Maurizio, Meredieu, Céline, Messier, Julie, Mészáros, Ilona, Metsaranta, Juha, Michaletz, Sean T., Michelaki, Chrysanthi, Migalina, Svetlana, Milla, Ruben, Miller, Jesse E.D., Minden, Vanessa, Ming, Ray, Mokany, Karel, Moles, Angela T., Molnár V, Attila, Molofsky, Jane, Molz, Martin, Montgomery, Rebecca A., Monty, Arnaud, Moravcová, Lenka, Moreno-Martínez, Alvaro, Moretti, Marco, Mori, Akira S., Mori, Shigeta, Morris, Dave, Morrison, Jane, Mucina, Ladislav, Mueller, Sandra, Muir, Christopher D., Müller, Sandra Cristina, Munoz, François, Myers-Smith, Isla H., Myster, Randall W., Nagano, Masahiro, Naidu, Shawna, Narayanan, Ayyappan, Natesan, Balachandran, Negoita, Luka, Nelson, Andrew S., Neuschulz, Eike Lena, Ni, Jian, Niedrist, Georg, Nieto, Jhon, Niinemets, Ülo, Nolan, Rachael, Nottebrock, Henning, Nouvellon, Yann, Novakovskiy, Alexander, Network, The Nutrient, Nystuen, Kristin Odden, O'Grady, Anthony, O'Hara, Kevin, O'Reilly-Nugent, Andrew, Oakley, Simon, Oberhuber, Walter, Ohtsuka, Toshiyuki, Oliveira, Ricardo, Öllerer, Kinga, Olson, Mark E., Onipchenko, Vladimir, Onoda, Yusuke, Onstein, Renske E., Ordonez, Jenny C., Osada, Noriyuki, Ostonen, Ivika, Ottaviani, Gianluigi, Otto, Sarah, Overbeck, Gerhard E., Ozinga, Wim A., Pahl, Anna T., Paine, C.E. Timothy, Pakeman, Robin J., Papageorgiou, Aristotelis C., Parfionova, Evgeniya, Pärtel, Meelis, Patacca, Marco, Paula, Susana, Paule, Juraj, Pauli, Harald, Pausas, Juli G., Peco, Begoña, Penuelas, Josep, Perea, Antonio, Peri, Pablo Luis, Petisco-Souza, Ana Carolina, Petraglia, Alessandro, Petritan, Any Mary, Phillips, Oliver L., Pierce, Simon, Pillar, Valério D., Pisek, Jan, Pomogaybin, Alexandr, Poorter, Hendrik, Portsmuth, Angelika, Poschlod, Peter, Potvin, Catherine, Pounds, Devon, Powell, A. Shafer, Power, Sally A., Prinzing, Andreas, Puglielli, Giacomo, Pyšek, Petr, Raevel, Valerie, Rammig, Anja, Ransijn, Johannes, Ray, Courtenay A., Reich, Peter B., Reichstein, Markus, Reid, Douglas E.B., Réjou-Méchain, Maxime, de Dios, Victor Resco, Ribeiro, Sabina, Richardson, Sarah, Riibak, Kersti, Rillig, Matthias C., Riviera, Fiamma, Robert, Elisabeth M.R., Roberts, Scott, Robroek, Bjorn, Roddy, Adam, Rodrigues, Arthur Vinicius, Rogers, Alistair, Rollinson, Emily, Rolo, Victor, Römermann, Christine, Ronzhina, Dina, Roscher, Christiane, Rosell, Julieta A., Rosenfield, Milena Fermina, Rossi, Christian, Roy, David B., Royer-Tardif, Samuel, Rüger, Nadja, Ruiz-Peinado, Ricardo, Rumpf, Sabine B., Rusch, Graciela M., Ryo, Masahiro, Sack, Lawren, Saldaña, Angela, Salgado-Negret, Beatriz, Salguero-Gomez, Roberto, Santa-Regina, Ignacio, Santacruz-García, Ana Carolina, Santos, Joaquim, Sardans, Jordi, Schamp, Brandon, Scherer-Lorenzen, Michael, Schleuning, Matthias, Schmid, Bernhard, Schmidt, Marco, Schmitt, Sylvain, Schneider, Julio V., Schowanek, Simon D., Schrader, Julian, Schrodt, Franziska, Schuldt, Bernhard, Schurr, Frank, Selaya Garvizu, Galia, Semchenko, Marina, Seymour, Colleen, Sfair, Julia C., Sharpe, Joanne M., Sheppard, Christine S., Sheremetiev, Serge, Shiodera, Satomi, Shipley, Bill, Shovon, Tanvir Ahmed, Siebenkäs, Alrun, Sierra, Carlos, Silva, Vasco, Silva, Mateus, Sitzia, Tommaso, Sjöman, Henrik, Slot, Martijn, Smith, Nicholas G., Sodhi, Darwin, Soltis, Pamela, Soltis, Douglas, Somers, Ben, Sonnier, Grégory, Sørensen, Mia Vedel, Sosinski Jr., Enio Egon, Soudzilovskaia, Nadejda A., Souza, Alexandre F., Spasojevic, Marko, Sperandii, Marta Gaia, Stan, Amanda B., Stegen, James, Steinbauer, Klaus, Stephan, Jörg G., Sterck, Frank, Stojanovic, Dejan B., Strydom, Tanya, Suarez, Maria Laura, Svenning, Jens-Christian, Svitková, Ivana, Svitok, Marek,

Svoboda, Miroslav, Swaine, Emily, Swenson, Nathan, Tabarelli, Marcelo, Takagi, Kentaro, Tappeiner, Ulrike, Tarifa, Rubén, Tauugourdeau, Simon, Tavsanoglu, Cagatay, te Beest, Mariska, Tedersoo, Leho, Thiffault, Nelson, Thom, Dominik, Thomas, Evert, Thompson, Ken, Thornton, Peter E., Thuiller, Wilfried, Tichý, Lubomír, Tissue, David, Tjoelker, Mark G., Tng, David Yue Phin, Tobias, Joseph, Török, Péter, Tarin, Tonantzin, Torres-Ruiz, José M., Tóthmérész, Béla, Treurnicht, Martina, Trivellone, Valeria, Trolliet, Franck, Trotsiuk, Volodymyr, Tsakalos, James L., Tsiripidis, Ioannis, Tysklind, Niklas, Umehara, Toru, Usoltsev, Vladimir, Vadeboncoeur, Matthew, Vaezi, Jamil, Valladares, Fernando, Vamosi, Jana, van Bodegom, Peter M., van Breugel, Michiel, Van Cleemput, Elisa, van de Weg, Martine, van der Merwe, Stephni, van der Plas, Fons, van der Sande, Masha T., van Kleunen, Mark, Van Meerbeek, Koenraad, Vanderwel, Mark, Vanselow, Kim André, Vårhammar, Angelica, Varone, Laura, Vasquez Valderrama, Maribel Yesenia, Vassilev, Kiril, Vellend, Mark, Veneklaas, Erik J., Verbeeck, Hans, Verheyen, Kris, Vibrans, Alexander, Vieira, Ima, Villacís, Jaime, Violle, Cyrille, Vivek, Pandi, Wagner, Katrin, Waldram, Matthew, Waldron, Anthony, Walker, Anthony P., Waller, Martyn, Walther, Gabriel, Wang, Han, Wang, Feng, Wang, Weiqi, Watkins, Harry, Watkins, James, Weber, Ulrich, Weedon, James T., Wei, Liping, Weigelt, Patrick, Weiher, Evan, Wells, Aidan W., Wellstein, Camilla, Wenk, Elizabeth, Westoby, Mark, Westwood, Alana, White, Philip John, Whitten, Mark, Williams, Mathew, Winkler, Daniel E., Winter, Klaus, Womack, Chevonne, Wright, Ian J., Wright, S. Joseph, Wright, Justin, Pinho, Bruno X., Ximenes, Fabiano, Yamada, Toshihiro, Yamaji, Keiko, Yanai, Ruth, Yankov, Nikolay, Yguel, Benjamin, Zanini, Kátia Janaina, Zanne, Amy E., Zelený, David, Zhao, Yun-Peng, Zheng, Jingming, Zheng, Ji, Ziemińska, Kasia, Zirbel, Chad R., Zizka, Georg, Zo-Bi, Irié Casimir, Zotz, Gerhard, Wirth, Christian, 2020. TRY plant trait database – enhanced coverage and open access. Global Change Biol. 26 (1), 119–188.

Kichenin, Emilie, Wardle, David A., Peltzer, Duane A., Morse, Chris W., Freschet, Grégoire T., 2013. Contrasting effects of plant inter- and intraspecific variation on community-level trait measures along an environmental gradient. Funct. Ecol. 27 (5), 1254–1261.

Kindt, Roeland, 2020. WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. Appl. Plant Sci. 8 (9), e11388.

Kuznetsova, Alexandra, Brockhoff, Per B., Christensen, Rune H.B., 2017. lmerTest package: Tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26.

Laliberté, Etienne, Shipley, Bill, Norton, David A., Scott, David, 2012. Which plant traits determine abundance under long-term shifts in soil resource availability and grazing intensity? J. Ecol. 100 (3), 662–677.

Lam, Olee H.Y., Tautenhahn, Susanne, Walther, Gabriel, Boenisch, Gerhard, Baddam, Pramod, Kattge, Jens, 2023. rtry: Preprocessing plant trait data. R package version 1.1.0.

Lande, Russell, Engen, Steinar, Sæther, Bernt-Erik, 2003. Demographic and environmental stochasticity. In: Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press.

Laughlin, Daniel C., Laughlin, David E., 2013. Advances in modeling trait-based plant community assembly. Trends Plant Sci. 18 (10), 584–593.

Lavorel, Sandra, Grigulis, Karl, McIntyre, Sue, Williams, Nick S.G., Garden, Denys, Dorrough, Josh, Berman, Sandra, Quétier, Fabien, Thébault, Aurélie, Bonis, Anne, 2008. Assessing functional diversity in the field – methodology matters!. Funct. Ecol. 22 (1), 134–147.

Legendre, Pierre, 2022. Model II regression user's guide, R edition. R package version 1.7-3.

Legendre, P., Legendre, L., 2012. Numerical Ecology, second ed. In: Developments in Environmental Modelling, (v. 24), Elsevier.

Måren, Inger Elisabeth, Kapfer, Jutta, Aarrestad, Per Arild, Grytnes, John-Arvid, Vandvik, Vigdis, 2018. Changing contributions of stochastic and deterministic processes in community assembly over a successional gradient. Ecology 99 (1), 148–157.

Merow, Cory, Latimer, Andrew M., Silander, Jr., John A., 2011. Can entropy maximization use functional traits to explain species abundances? A comprehensive evaluation. Ecology 92 (7), 1523–1537.

Niu, Kechang, He, Jin-Sheng, Lechowicz, Martin J., 2016. Foliar phosphorus content predicts species relative abundance in P-limited tibetan alpine meadows. Perspect. Plant Ecol. Evol. Syst. 22, 47–54.

Pérez-Llorca, Marina, Fenollosa, Erola, Salguero-Gómez, Roberto, Munné-Bosch, Sergi, 2018. What Is the Minimal Optimal Sample Size for Plant Ecophysiological Studies? Plant Physiol. 178 (3), 953–955.

Perronne, Rémi, Munoz, François, Borgy, Benjamin, Reboud, Xavier, Gaba, Sabrina, 2017. How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. Perspect. Plant Ecol. Evol. Syst. 25, 29–44.

Phillips, Steven J., Anderson, Robert P., Schapire, Robert E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3), 231–259.

Pos, Edwin T., de Souza Coelho, Luiz, de Andrade Lima Filho, Diogenes, Salomão, Rafael P., Amaral, Iêda Leão, de Almeida Matos, Francisca Dionízia, Castilho, Carolina V., Phillips, Oliver L., Guevara, Juan Ernesto, de Jesus Veiga Carim, Marcelo, et al., 2023. Unraveling amazon tree community assembly using maximum information entropy: a quantitative analysis of tropical forest ecology. Sci. Rep. 13 (1), 2859.

Pueyo, Salvador, He, Fangliang, Zillio, Tommaso, 2007. The maximum entropy formalism and the idiosyncratic theory of biodiversity. Ecol. Lett. 10 (11), 1017–1028.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rech, Morgana, Carlucci, Marcos B., Bergamin, Rodrigo S., Seger, Guilherme D.S., Müller, Sandra C., Duarte, Leandro, Hartz, Sandra M., 2019. Mass effects explain sapling community assembly in araucaria mixed forest metacommunities. J. Veg. Sci. 30 (4), 664–673.

Rolhauser, Andres, Waller, Donald, Tucker, Caroline, 2021a. Complex trait–environment relationships underlie the structure of forest plant communities. http://dx.doi.org/10.5061/dryad.4qrfj6qb0, Dryad.

Rolhauser, Andrés G., Waller, Donald M., Tucker, Caroline M., 2021b. Complex trait–environment relationships underlie the structure of forest plant communities. J. Ecol. 109 (11), 3794–3806.

Römermann, Christine, Bucher, Solveig Franziska, Hahn, Melanie, Bernhardt-Römermann, Markus, 2016. Plant functional traits – fixed facts or variable depending on the season? Folia Geobot. 51 (2), 143–159.

Rosenfeld, Jordan S., 2002. Functional redundancy in ecology and conservation. Oikos 98 (1), 156–162.

Schrodt, Franziska, Kattge, Jens, Shan, Hanhuai, Fazayeli, Farideh, Joswig, Julia, Banerjee, Arindam, Reichstein, Markus, Bönisch, Gerhard, Díaz, Sandra, Dickie, John, Gillison, Andy, Karpatne, Anuj, Lavorel, Sandra, Leadley, Paul, Wirth, Christian B., Wright, Ian J., Wright, S. Joseph, Reich, Peter B., 2015. BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. Glob. Ecol. Biogeogr. 24 (12), 1510–1521.

Shipley, Bill, 2009a. Limitations of entropy maximization in ecology: a reply to Haegeman and Loreau. Oikos 118 (1), 152–159.

Shipley, Bill, 2009b. Trivial and non-trivial applications of entropy maximization in ecology: Shipley's reply. Oikos 118 (8), 1279–1280.

Shipley, Bill, 2010a. From Plant Traits to Vegetation Structure: Chance and Selection in the Assembly of Ecological Communities. Cambridge University Press.

Shipley, Bill, 2010b. Inferential permutation tests for maximum entropy models in ecology. Ecology 91 (9), 2794–2805.

Shipley, Bill, 2014. Measuring and interpreting trait-based selection versus meta-community effects during local community assembly. J. Veg. Sci. 25 (1), 55–65.

Shipley, Bill, Laughlin, Daniel C., Sonnier, Grégory, Otfinowski, Rafael, 2011. A strong test of a maximum entropy model of trait-based community assembly. Ecology 92 (2), 507–517.

Shipley, Bill, Paine, C.E. Timothy, Baraloto, Christopher, 2012. Quantifying the importance of local niche-based and stochastic processes to tropical tree community assembly. Ecology 93 (4), 760–769.

Shipley, Bill, Paine, C.E. Timothy, Baraloto, Christopher, 2016. Supplement 1. R script and two R functions to implement the analyses described in our paper.

Shipley, Bill, Vile, Denis, Garnier, Éric, 2006. From plant traits to plant communities: A statistical mechanistic approach to biodiversity. Science 314 (5800), 812–814.

Siefert, Andrew, Violle, Cyrille, Chalmandrier, Loïc, Albert, Cécile H., Taudiere, Adrien, Fajardo, Alex, Aarssen, Lonnie W., Baraloto, Christopher, Carlucci, Marcos B., Cianciaruso, Marcus V., de L. Dantas, Vinícius, de Bello, Francesco, Duarte, Leandro D.S., Fonseca, Carlos R., Freschet, Grégoire T., Gaucherand, Stéphanie, Gross, Nicolas, Hikosaka, Kouki, Jackson, Benjamin, Jung, Vincent, Kamiyama, Chiho, Katabuchi, Masatoshi, Kembel, Steven W., Kichenin, Emilie, Kraft, Nathan J.B., Lagerström, Anna, Bagousse-Pinguet, Yoann Le, Li, Yuanzhi, Mason, Norman, Messier, Julie, Nakashizuka, Tohru, Overton, Jacob McC., Peltzer, Duane A., Pérez-Ramos, I.M., Pillar, Valério D., Prentice, Honor C., Richardson, Sarah, Sasaki, Takehiro, Schamp, Brandon S., Schöb, Christian, Shipley, Bill, Sundqvist, Maja, Sykes, Martin T., Vandewalle, Marie, Wardle, David A., 2015. A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. Ecol. Lett. 18 (12), 1406–1419.

Silva, J.L.A., Souza, A.F., Jardim, J.G., Goto, B.T., 2015. Community assembly in harsh environments: the prevalence of ecological drift in the heath vegetation of south america. Ecosphere 6 (7), art111.

Sonnier, Grégory, Navas, Marie-Laure, Fayolle, Adeline, Shipley, Bill, 2012. Quantifying trait selection driving community assembly: a test in herbaceous plant communities under contrasted land use regimes. Oikos 121 (7), 1103–1111.

Strahan, Robert T., Laughlin, Daniel C., Moore, Margaret M., 2018. An experimental test of the community assembly by trait selection (CATS) model. PLoS One 13 (11), 1–15.

Tang, Junfeng, Zhou, Shurong, 2011. The importance of niche differentiation for coexistence on large scales. J. Theoret. Biol. 273 (1), 32–36.

Taugourdeau, Simon, Villerd, Jean, Plantureux, Sylvain, Huguenin-Elie, Olivier, Amiaud, Bernard, 2014. Filling the gap in functional trait databases: use of ecological hypotheses to replace missing data. Ecol. Evol. 4 (7), 944–958.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York, ISBN 0-387-95457-0.

Volkov, Igor, Banavar, Jayanth R., Hubbell, Stephen P., Maritan, Amos, 2009. Inferring species interactions in tropical forests. Proc. Natl. Acad. Sci. 106 (33), 13854–13859.

Warton, David I., Shipley, Bill, Hastie, Trevor, 2015. CATS regression – a model-based approach to studying trait-based community assembly. Methods Ecol. Evol. 6 (4), 389–398.

Whittaker, Robert H., 1975. Communities and Ecosystems, second ed. MacMillan Publishing Co., New York, NY.

Zuppinger-Dingley, Debra, Schmid, Bernhard, Petermann, Jana S., Yadav, Varuna, De Deyn, Gerlinde B., Flynn, Dan F.B., 2014. Selection for niche differentiation in plant communities increases biodiversity effects. Nature 515 (7525), 108–111.