



HAL
open science

Sensitivity of community-level trait-environment relationships to data representativeness: a test for functional biogeography

Benjamin Borgy, Cyrille Violle, Philippe Choler, Eric Garnier, Jens Kattge, Jessy Loranger, Bernard B. Amiaud, Pierre Cellier, Guilhem Debarros, Pierre Denelle, et al.

► To cite this version:

Benjamin Borgy, Cyrille Violle, Philippe Choler, Eric Garnier, Jens Kattge, et al.. Sensitivity of community-level trait-environment relationships to data representativeness: a test for functional biogeography. *Global Ecology and Biogeography*, 2017, 26 (6), pp.729 - 739. 10.1111/geb.12573 . hal-01465075

HAL Id: hal-01465075

<https://hal.science/hal-01465075v1>

Submitted on 25 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sensitivity of community-level trait–environment relationships to data representativeness: A test for functional biogeography

Benjamin Borgy^{1,2,3} | Cyrille Violle^{1,2} | Philippe Choler^{3,4}  | Eric Garnier¹  |
Jens Kattge^{4,5} | Jessy Loranger^{1,6} | Bernard Amiaud^{7,8} | Pierre Cellier^{1,9} |
Guilhem Debarros¹⁰ | Pierre Denelle¹ | Sylvain Diquelou^{11,12} | Sophie Gachet¹³ |
Claudy Jolivet¹⁴ | Sandra Lavorel³ | Servane Lemauviel-Lavenant^{11,12} |
Alexis Mikolajczak¹⁵ | François Munoz^{9,16}  | Jean Olivier¹⁰ | Nicolas Viovy¹⁷

¹Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175), CNRS – Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, Montpellier, France

²CESAB/FRB, Aix-en-Provence, France

³Université Grenoble Alpes, CNRS, LECA, Grenoble, France

⁴Max Planck Institute for Biogeochemistry, Jena, Germany

⁵German Centre for Integrative Biodiversity Research [iDiv] Halle-Jena-Leipzig, Leipzig, Germany

⁶Université de Sherbrooke, Sherbrooke, Canada

⁷Ecologie et Ecophysiologie Forestières, UMR 1137, Université de Lorraine, Vandoeuvre-les-Nancy, France

⁸INRA, Ecologie et Ecophysiologie Forestières, UMR 1137, Champenoux, France

⁹Université de Montpellier, UMR AMAP, Montpellier, France

¹⁰Fédération des Conservatoires Botaniques Nationaux, Montreuil-sous-Bois, France

¹¹Normandie Université, UMR Ecophysiologie Végétale Agronomie et Nutrition NCS, Caen, France

¹²INRA, UMR 950 NCS, Caen, France

¹³Aix Marseille Univ, Avignon Université, CNRS, IRD, IMBE, Marseille, France

¹⁴INRA, US1106 InfoSol, Orléans, France

¹⁵Conservatoire Botanique National Alpin, Domaine de Charance, Gap, France

¹⁶French Institute of Pondicherry, Pondicherry, India

¹⁷Laboratoire des Sciences du Climat et de l'Environnement [UMR 8212 CEA/CNRS/UVSQ], CEA Saclay, Gif-sur-Yvette, France

Correspondence

Cyrille Violle, CEFE UMR 5175, CNRS –
Université de Montpellier, Université
Paul-Valéry Montpellier – EPHE,
F-34293 Montpellier, Cedex 5, France.
Email: cyrille.violle@cefe.cnrs.fr

Editor: Dr. Brian McGill

Abstract

Aim: The characterization of trait–environment relationships over broad-scale gradients is a critical goal for ecology and biogeography. This implies the merging of plot and trait databases to assess community-level trait-based statistics. Potential shortcomings and limitations of this approach are that: (i) species traits are not measured where the community is sampled and (ii) the availability of trait data varies considerably across species and plots. Here we address the effect of trait data representativeness [the sampling effort per species and per plot] on the accuracy of (i) species-level and (ii) community-level trait estimates and (iii) the consequences for the shape and strength of trait–environment relationships across communities.

Innovation: We combined information existing in databases of vegetation plots and plant traits to estimate community-weighted means [CWMs] of four key traits [specific leaf area, plant height, seed mass and leaf nitrogen content per dry mass] in permanent grasslands at a country-wide

scale. We propose a generic approach for systematic sensitivity analyses based on random sub-sampling and data reduction to address the representativeness of incomplete and heterogeneous trait information when exploring trait–environment relationships across communities.

Main conclusions: The accuracy of the CWMs was little affected by the number of individual trait values per species [NIV] but strongly affected by the cover proportion of species with available trait values [P_{Cover}]. A P_{Cover} above 80% was required for all four traits studied to obtain an estimation bias below 5%. Our approach therefore provides more conservative criteria than previously proposed. Restrictive criteria on both NIV and P_{Cover} primarily excluded communities in harsh environments, and such reduction of the sampled gradient weakened trait–environment relationships. These findings advocate systematic measurement campaigns in natural environments to increase species coverage in global trait databases, with special emphasis on species occurring in under-sampled and harsh environmental conditions.

KEYWORDS

community functional structure, community-weighted mean [CWM], global database, plant trait, trait sampling, vegetation plot

1 | INTRODUCTION

A growing body of evidence suggests that understanding both the effects of environmental factors on vegetation and how vegetation affects ecosystem properties requires a specific focus on the functional structure of plant communities (Diaz et al., 2007; Enquist et al., 2015; Garnier & Navas, 2012; Lavorel, 2013; Mason & De Bello, 2013), defined by the distribution of trait values within these communities. A wealth of metrics have been developed to describe this distribution (De Bello, Carmona, Mason, Sebastia, & Leps, 2013; Mouchet, Villéger, Mason, & Mouillot, 2010; Schleuter, Daufresne, Massol, & Arguillier, 2010; Taudiere & Violle, 2016; Violle et al., 2012; Weiher, 2010), which essentially capture two complementary facets of functional structure: the community average trait value and trait dissimilarity among coexisting individuals. Among these metrics, the trait mean calculated for all species in a community weighted by species abundances [the community-weighted mean, hereafter CWM] (Garnier et al., 2004; Violle et al., 2007) appears in many aspects as a key descriptor of community functional structure. First, several reports have shown that trait–environment relationships are stronger when species abundances are taken into account [reviewed in Garnier, Navas, & Grigulis (2016)], suggesting a better adequacy between trait values and environment for the dominant species (Cingolani, Cabido, Gurvich, Renison, & Díaz, 2007). Second, consistent with the ‘mass ratio’ hypothesis, which states that the traits of species affect ecosystem properties in proportion to their local biomass (Grime, 1998), the CWMs of several core vegetative traits have been shown to be consistently associated with several key ecosystem processes (e.g., Diaz et al., 2007; Garnier & Navas, 2012; Lavorel, 2013).

The constant improvement of plant trait databases now offers opportunities to explore large-scale spatial patterns of plant functions, a primary objective of the emerging field of functional biogeography (Violle, Reich, Pacala, Enquist, & Kattge, 2014). A first generation of functional biogeography studies has used average species trait values

for all species occurring within the same geographical unit [usually a grid cell]. This approach gives the same weight to all species present [or expected to be present] in the defined spatial unit (e.g., Swenson et al., 2012). However, following our arguments above, a more functional approach to biogeography should combine species abundance taken from vegetation plots with trait data. Here, we address a number of issues that will be common to forthcoming functional biogeography analyses—if conducted across large scales—in which: (i) available plant traits have not been measured where the plant community was sampled and (ii) the plant community composition includes species for which no trait data are available. Thus, the main problem to overcome is that of missing data in the species-by-trait matrix of the different vegetation plots (Pakeman, 2014). More specifically, we will be confronted with an insufficient number of individual trait measurements to properly estimate a species’ mean trait value [MTV] (Hulshof & Swenson, 2010; Sandel et al., 2015; Violle, Borge, & Choler, 2015) and the heterogeneous availability of vegetation plots and trait data along the environmental gradient of interest. Both can affect the estimation of CWM and the potential to identify a trait–environment relationship. In this context, the equation classically given to calculate a CWM (Violle et al., 2007) for a given trait

$$\text{CWM} = \sum_i p_i T_i, \quad (1)$$

where p_i is the relative abundance of species i in the community and T_i is the trait value of species i observed in the plot, needs to be replaced by

$$\text{CWM} = \frac{\sum_i p_i \sum_{j=1}^{\text{NIV}_i} (t_{ij} / \text{NIV}_i)}{P_{\text{Cover}}} \quad \forall \text{NIV}_i \geq 1, \quad (2)$$

where p_i is the relative abundance of species i in the community, NIV is the number of individual values of the trait under scrutiny in the database for species i , t_{ij} is the value j of a given trait of species i in the database and P_{Cover} is, for a given community, the proportion

[cumulative relative abundance] of all species for which a trait value is available in the database.

In this study, we examine the reliability of predictions from a global plant trait database to assess community-level traits and their sensitivity to environmental variation by addressing three questions:

1. What is the sensitivity of species-level trait estimates to the number of available measurements per species [NIV] in the database and how well do these species trait estimates match locally measured trait values [at both species and community levels].
2. What is the impact of NIV and of the proportion of species with available trait data [P_{Cover}] on the accuracy of CWM estimates?
3. To what extent may the constraints on trait data availability blur relationships between CWM and climate in a real dataset?

Our study draws upon the combined analysis of two large databases covering the diversity of permanent grasslands in France. The first one consists of a dataset of around 50,000 vegetation plots combined into a single database, the DivGrass vegetation plot database (Violle, Choler, et al., 2015). The second is the TRY database, recognized as the main resource for plant functional traits worldwide (Kattge et al., 2011). For the first two questions, we considered four functionally relevant and well-studied traits: specific leaf area [SLA], plant height [height], seed mass [SM] and leaf nitrogen content per dry mass [LNC]. For the third question, we focused on the CWM of SLA and its variation along a gradient of growing degree days [GDD].

2 | MATERIALS AND METHODS

2.1 | Vegetation data

A dataset of 51,486 vegetation plots from French permanent grasslands [FPGs] was created using multiple data sources [see Table S1 in the Supporting Information online], combined into a data repository, the DivGrass database (Violle, Choler, et al., 2015). The data consist of visually estimated relative cover of all species present in plots using a six-level abundance scale: 0%–1%, 1%–5%, 5%–25%, 25%–50%, 50%–75% and 75%–100%. We used the median of each class to derive a percentage cover for each species, i.e., 0.5%, 3%, 15%, 37.5%, 62.5% and 87.5%, respectively.

We combined the European Corine Land Cover database (European Environment Agency, 2006) and the Registre Parcellaire Graphique to map the distribution of FPGs at 5-km resolution (Violle, Choler, et al., 2015). This map was used to select 26,586 plots [51.6% of the total] that occurred within grid cells covered by at least 20% permanent grassland. In regions where grasslands are relatively scarce, we noticed that the limited number of available vegetation plots often correspond to specialized micro-habitats [e.g., meadows on calcareous outcrops, wet grasslands] that are not captured by macroclimatic variables.

Species names from the different datasets were harmonized using the TaxRef4 national taxonomic referential database [Muséum National d'Histoire Naturelle, [\[nomique-taxref\]\(https://www.try-db.org/\)\]. To merge the taxonomy from vegetation plots and the trait database \[TRY, <https://www.try-db.org/>\], all infraspecific ranks \[subspecies, variety, etc.\] were replaced by their corresponding specific rank. The final dataset included 3,589 vascular plant species.](https://inpn.mnhn.fr/programme/referentiel-taxo-</p></div><div data-bbox=)

2.2 | Trait data

We extracted individual values for SLA [$\text{m}^2 \text{kg}^{-1}$], LNC [mg g^{-1}], SM [mg] and height [m] from the TRY database (Kattge et al., 2011), which contains plant trait data collected worldwide. Trait data from artificial conditions [e.g., greenhouses, growth chambers] were not retained for this analysis. For a given trait and a given species with more than three values we discarded values if their deviation from the species' mean was two times higher than the standard deviation of that species' trait distribution. We ended up with 17.4%, 14.2%, 55.2% and 30.2% of species with available traits for SLA, LNC, SM and height, respectively. The mean number of individual trait values [i.e., NIV] per species was 5.71, 5.66, 3.63 and 3.35 for SLA, LNC, SM and height, respectively.

2.2.1 | Species-level comparison

We compared species mean traits obtained from the TRY database with species mean traits derived from a database of local measurements, the DivHerbe database (Gardarin et al., 2014). This latter database includes 100 vegetation plots and trait measurements on 180 species covering a large range of climate and soil conditions within France, from subalpine to Mediterranean grasslands.

2.2.2 | Community-level comparison

We compared on-site CWM values of the 100 DivHerbe plots with CWM values calculated using local species abundances of the DivHerbe plots and species MTVs taken from TRY. For each comparison, Spearman rank correlation tests were performed to test whether ranking was conserved between datasets (Cordlandwehr et al., 2013; Kazakou et al., 2014; Violle, Choler, et al., 2015). Additionally, paired Wilcoxon signed rank tests were performed to test whether distributions were different between both datasets.

2.3 | Climate data

To relate CWM to climate, we used the AURELHY [Analyse Utilisant le RELief pour l'HYdrométéorologie] gridded dataset of Météo France (Benichou & Le Breton, 1986). Monthly mean temperature over the period 1961–1990 was daily interpolated and we calculated the number of GDD above 5°C, considered as a threshold for temperate grassland plant growth. The resolution of the climate grid cell was 5 km, which is in the same order of magnitude as the uncertainty of the location of most of the vegetation plots. To derive trait–climate relationships, we considered only grid cells containing at least three plots. An averaged CWM for each trait was calculated for each climate grid cell.

2.4 | Sensitivity of species mean trait estimates to NIV

We analysed the impact of the NIV value available in TRY on the estimation of species MTVs. Only herbaceous species with $NIV \geq 50$ were selected for this analysis, leading to 30, 8, 19 and 23 species for SLA, LNC, SM and height, respectively. For each species, 50 individual values were randomly selected and averaged to estimate a species' MTV [MTV₅₀]. We chose $NIV = 50$ as a compromise between a sufficient number of examined species and the robustness of the sensitivity analysis. Then, we varied NIV between 1 and 49 to calculate MTV_{NIV}, and the MTV₅₀ error was calculated as $[MTV_{NIV} - MTV_{50}] / MTV_{50}$. This random subsampling was performed 10 times for a given NIV value and a given species [Fig. S1]. Results were averaged across all species for each NIV value. This procedure was repeated 100 times giving a total of 1,000 simulations per NIV level. We calculated the NIV corresponding to a MTV₅₀ with an error smaller than 20%, 10% and 5% for at least 95% of these simulations.

2.5 | Sensitivity of CWMs to measurement intensity in trait databases [NIV] and P_{Cover}

Species with a $NIV \geq 10$ were selected to ensure a sufficient number of plots for the sensitivity analysis. For a given plot, the cover proportion [total cumulative percentage of relative abundance] represented by all the species for which trait data were available was defined as P_{Cover} . For the sensitivity analysis, only plots for which P_{Cover} was $\geq 95\%$ were selected.

For each trait, we analysed the sensitivity of CWM to P_{Cover} and NIV using 1,000 random subsamplings. For each simulation, 10 individual trait values were randomly drawn for each species and the CWM was calculated by using species relative abundances [p_i] and their MTVs [Equation 2]. These CWMs were used as reference values. Then we varied NIV between 1 and 10 for each species and calculated CWM_{NIV} by using 10%–100% of occurring species [in 10% increments]. The error was defined as $[CWM_{NIV} - CWM] / CWM$. This random subsampling allows us to assess the error of the calculated trait at community level with independent variation of NIV and P_{Cover} [Fig. S2].

2.6 | Sensitivity of the climate–CWM_{SLA} relationship to NIV and P_{Cover}

We finally focused on the reliability of climate–CWM relationships using SLA, the trait for which sufficient information [measurement intensity and data coverage] was available. We first analysed the impact of data selection criteria [variation in NIV and P_{Cover}] on the size of the dataset [the total number of vegetation plots and total number of grid cells for which the number of plots is at least three] and on the proportion of the climate gradient [i.e., GDD] covered. To that end, we first varied the minimum NIV between 1 and 10 and minimum P_{Cover} between 10% and 100% [10% increments]. Next, according to the preliminary results of sensitivity of CWM to data selection, we analysed the CWM_{SLA}–GDD relationship for two distinct data selection criteria: highly restrictive [$NIV \geq 10$ and $P_{Cover} \geq 90\%$] and moderately restrictive [$NIV \geq 2$ and $P_{Cover} \geq 60\%$].

3 | RESULTS

3.1 | Sensitivity of species mean trait estimates to NIV

For all traits, the ranking of species mean traits was consistent between the TRY database and the local DivHerbe database, with the Spearman correlation ranging from .51 to .92 [Figure 1a]. Distributions were not significantly different for SLA [$V = 2711$, p value = .11] and height [$V = 4536$, p value = .58].

Figure 2 shows the effect of NIV [between 1 and 49] on the estimation of species mean SLA, LNC, SM and height using trait data from the TRY database. The availability of only one individual value [$NIV = 1$] led to a maximum estimation error of 50% [positive or negative]. An error below 20% requires a minimum NIV of 7, 7, 8 and 15 for SLA, LNC, SM and height, respectively. To reach an error below 5%, the corresponding NIVs were 35, 36, 38 and 44. Table S2 summarizes the NIV availability per species and trait and the minimum NIV required for each error threshold.

3.2 | Sensitivity of CWM to NIV and P_{Cover}

For all traits, the ranking of CWM was consistent between on-site measurements of the plots from the DivHerbe database and CWM calculated for the same plots but using trait values from the TRY database, with the Spearman correlation ranging from .53 to .95 [Figure 1b]. Distributions were not significantly different for SM and height [Figure 1b].

The analysis of the combined effect of NIV and P_{Cover} on CWM estimation errors revealed a predominant effect of P_{Cover} for the four studied traits [Figure 3]. For all the traits, the impact of NIV was only noticeable for very low values. In our case study, when $NIV = 2$, an error on CWM estimation below 10% requires a trait coverage per plot [P_{Cover}] of 53%, 76%, 94% and 100% for SLA, LNC, SM and height, respectively [Figure 3]. Based on this result, we focused on SLA only in the following to test the sensitivity of climate–trait relationships to NIV and P_{Cover} .

3.3 | Sensitivity of climate–CWM_{SLA} relationships to NIV and P_{Cover}

We examined the extent to which uncertainties in CWM_{SLA} affect the performance of a linear regression model between CWM_{SLA} and GDD. Figure 4 provides a comprehensive assessment of data reduction in terms of number of plots, number of grid cells and extent of the GDD gradient when varying the threshold values for NIV and P_{Cover} .

A model using $NIV \geq 2$ and $P_{Cover} \geq 60\%$ was implemented with 1,686 grid cells representing 17,215 vegetation plots and showed an intercept of 14.36 [95% CI 13.8–14.9], a slope of 14.33 [95% CI 13.18–15.47] and an adjusted R^2 of .26 [Figure 5a]. This data subset covered 85% of the GDD gradient corresponding to the whole dataset. An example of the same bivariate plot using more restrictive criteria [$NIV \geq 10$ and $P_{Cover} \geq 90\%$] resulted in a smaller dataset with 488 grid cells representing 3,261 vegetation plots and in a restricted gradient of GDD [63% of the whole gradient] [Figure 5b]. In this restrictive case,

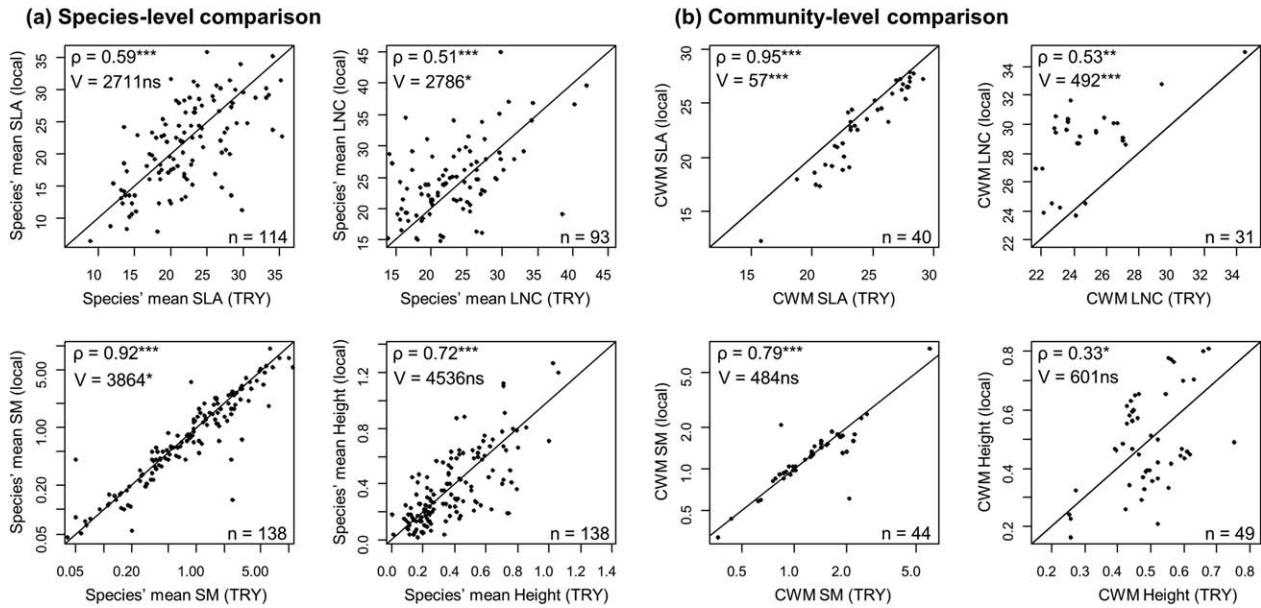


FIGURE 1 Species-level and community-level trait measurements: comparisons between on-site trait values and trait values extracted from the TRY database. (a) A species-level comparison. Each dot corresponds to the mean trait value of a given species extracted from the local DivHerbe dataset [y-axis] or from the TRY database [x-axis]. Analyses included 114, 93, 138 and 138 species found in both datasets for specific leaf area [SLA, $\text{m}^2 \text{kg}^{-1}$], leaf nitrogen content per dry mass [LNC, mg g^{-1}], seed mass [SM, mg] and height [m] respectively. (b) A community-level comparison. Each dot corresponds to the community-weighted mean [CWM] of traits of a DivHerbe vegetation plot using local [on-site] trait measurements [y-axis] or local species abundances and trait values extracted from TRY [x-axis] [$n = 100$ plots]. The line represents the 1:1 line. Spearman correlation coefficients [ρ and p values] between the two datasets and results of the paired Wilcoxon tests [V and p values] are given [$^{***}p < .001$; $^{**}p < .01$; $^*p < .05$; ns, not significant]. Scales were logged for SM

the model performance was diminished and the intercept of the model significantly differed from the previous case (intercept = 17.5 [95% CI 16.3–18.6]).

4 | DISCUSSION

Our study has investigated the impact of both the number of individual values used to calculate a species' MTV [NIV; i.e., measurement intensity; Sandel et al., 2015], and the cumulated cover proportion of species included in the calculation of a CWM [P_{Cover}] on the accuracy of estimates of species mean traits and CWM, and on CWM–climate relationships. We find that although a large NIV is necessary to estimate accurately a species' MTV, the accuracy of CWM is only weakly affected by NIV but strongly by P_{Cover} . The P_{Cover} needed to minimize the error on estimation of CWM varies among traits, which challenges the unique $P_{\text{Cover}} = 80\%$ threshold largely used in functional ecology studies (Pakeman & Queded, 2007). For SLA, for example, accepting a CWM estimation error of 10% allows the required P_{Cover} to be reduced to 60% with an NIV of at least two without affecting the CWM–environment relationship. In the following, we discuss in more detail the effect of data representativeness [measurement intensity and species coverage] on the accuracy of (i) species-level and (ii) community-level mean trait estimates and (iii) the consequences for the shape and strength of community-level trait–environment relationships.

4.1 | Sensitivity of trait means at species level

Extracting MTVs per species from regional and global databases necessarily leads to challenging issues when applying them in a more local context. A first issue is the effect of intraspecific variability (Albert, Grassein, Schurr, Vieilledent, & Violle, 2011; Baraloto et al., 2010). Consistent with previous studies showing the robustness of interspecific differences to intraspecific variability (Cordlandwehr et al., 2013; Kazakou et al., 2014; Violle, Choler, et al., 2015), we found that the interspecific ranking for MTVs from TRY or from local measurements was conserved for the four studied traits. Nevertheless the two datasets were overall significantly different for LNC and SM, showing a systematic bias. This difference between the two datasets for LNC confirms that [leaf] chemical traits are more variable than structural traits, most probably due to their dependence on local soil nutrient availability (Kazakou et al., 2014). The difference between the two datasets is more surprising for SM since it is generally considered to be one of the least variable traits within species (Harper, Lovell, & Moore, 1970; Kazakou et al., 2014). However, these differences seem to be explained by a few outliers which could potentially be associated with measurement errors, most probably in the TRY database.

A second aspect in using MTVs from global databases is the heterogeneous number of available NIVs per species in such datasets (Sandel et al., 2015). For instance, for SLA, NIV varies from 1 to 647 in TRY. Our test of the sensibility of species MTVs to NIV reveals that an accurate estimate [5% of estimation error] requires an NIV of at least 35 for the

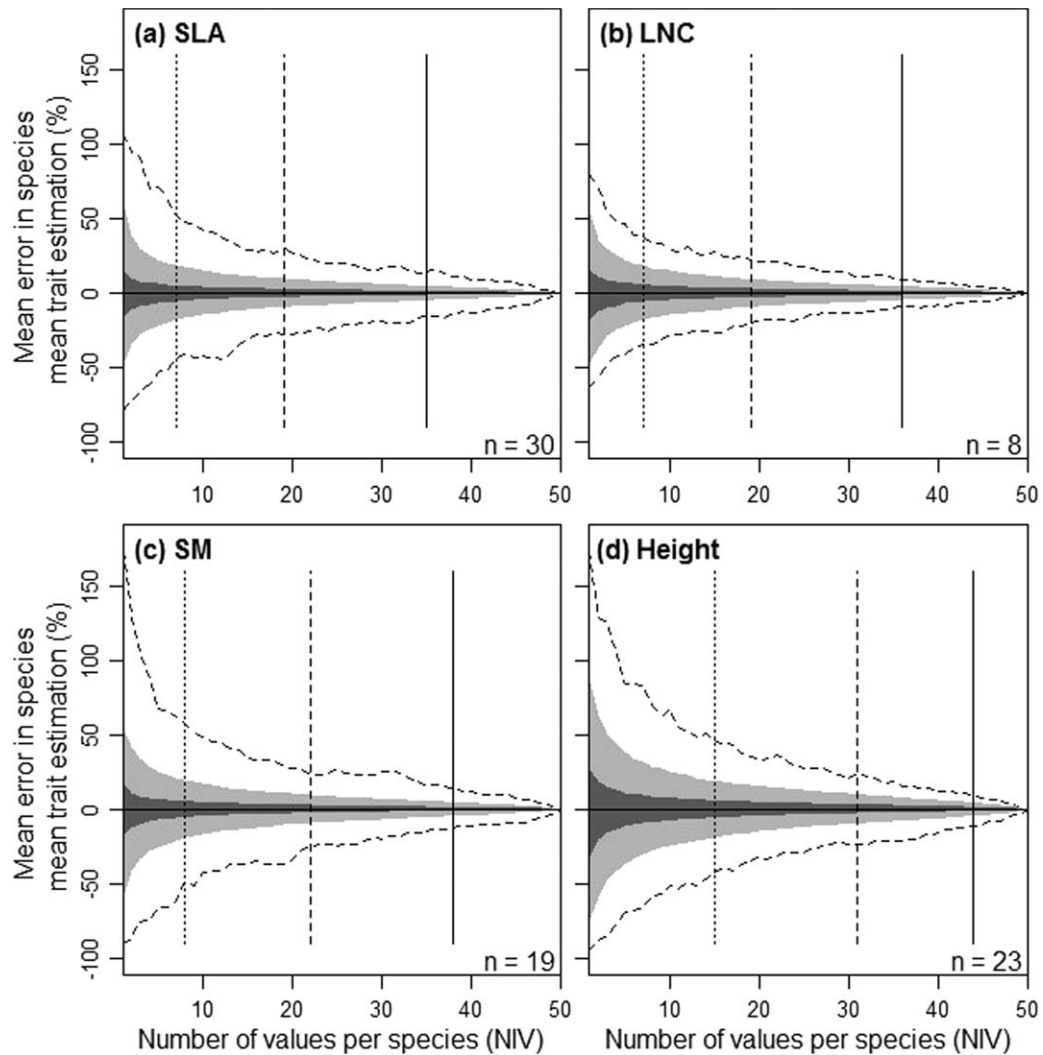


FIGURE 2 Impact of the number of individual values [NIV] on the estimates of species mean trait values using the TRY database. For each trait, the figure shows the mean error [%] in the estimation of mean trait value of species as a function of NIV randomly drawn [varying between 1 and 49] available in TRY. The reference mean trait value was estimated with 50 individual values taken in the TRY database. Black dashed lines represent maximum error, while light [dark] grey shapes represent 95% [50%] confidence intervals. Continuous, dashed and dotted vertical lines represent 5%, 10% and 20% of error, respectively. *n* represents the number of species used in the analysis

four studied traits [Figure 2]. This condition is not fulfilled for most species in current trait databases (Sandel et al., 2015). For instance, only ca. 3% of species possess at least 35 SLA observations in the TRY database (Sandel et al., 2015), the [currently] largest database of plant traits globally. On the other hand, our results suggest that a 20% error on species MTVs requires NIVs between NIV = 7 [SLA] and NIV = 15 [height].

From the perspective of comparative ecology, a core question is our ability to discriminate species based on their phenotypic features (Albert et al., 2011; Garnier et al., 2001; Keddy, 1992). In that case, the consequences of assessment errors of trait means may depend on the range of trait values covered among species. If individual species' values are predicted with 20% error, this seems far less problematic for SM, which varies by many orders of magnitude among species [4.5 orders of magnitude in our study], than predicting LNC, which varies far less widely among species [a 1.7 order of magnitude variation in our study]. To examine this idea more deeply, we extended the numerical simula-

tions provided in Figure 2 to test our ability to discriminate pairs of species along the gradient of measurement intensity [from NIV = 1 to NIV = 50] [Fig. S3]. With a 20% estimation error on trait means, the probability of discriminating two species varies between .70 for LNC and .98 for SM [Fig. S3]. This result supports the idea that the consequences of uncertainty on species trait means depend on species and trait coverages, as previously envisioned (Albert et al., 2011). Given that the DivGrass database covers the range of herbaceous grassland species occurring in France, those consequences may be less pronounced in functional biogeography exercises using a global coverage and/or both woody and non-woody species; conversely they may be more severe in local and ecosystem-specific studies. Overall, such sensitivity analyses need to be applied to other systems and using a more systematic comparison of rare and widespread species given a severe bias towards widespread species [more often measured] in current trait databases (Violle, Borghy, et al., 2015; Violle, Choler, et al., 2015).

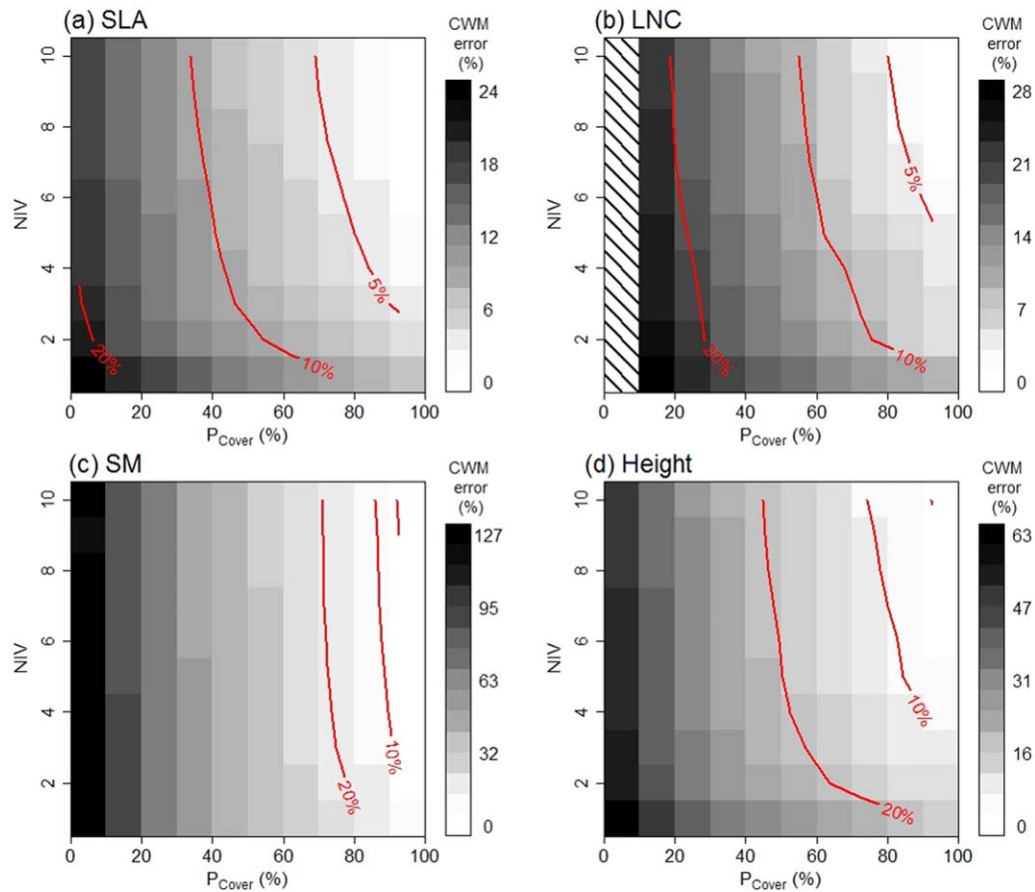


FIGURE 3 Impact of the number of individual trait values per species [NIV] in the TRY database and of the community-level cover proportion [P_{Cover}] of species for which trait values were available in TRY on the estimation of community weighted means [CWMs]. The figure shows the error [%] in the estimation of CWM for several individual values used in the calculation of species mean and several cover proportions for which mean trait values of species were known [P_{Cover} , %] for vegetation plots for which P_{Cover} was greater than or equal to 95% and for which occurring species had at least 10 individual trait values in TRY. Mean errors over the plots are represented by a grey shade gradient [lighter grey error of 0%, darker grey error of 25%]. Thresholds of 5%, 10% and 20% of mean error are denoted by red continuous lines. See Fig. S2 for details

4.2 | Sensitivity of CWMs to NIV and P_{Cover}

Regardless of the sensitivity of individual species' estimates to NIV, the CWMs of SLA, LNC, SM and height were estimated with an error <5% for $NIV > 5$ and P_{Cover} close to 80%, 90%, 100% and 100%, respectively [Figure 3]. A threshold P_{Cover} of 80% is in line with previous work evaluating the sampling effort required in the field for an accurate estimation of CWM (Pakeman & Quested, 2007). However, higher thresholds for the other traits highlight the importance of running more sensitivity analyses in the future and preclude the use of a unique threshold of 80%. Our study goes further by showing that even when a very low NIV of around two is retained, estimation of CWM remains relatively accurate [<20% error depending on the traits] even at much lower P_{Cover} values. Nevertheless, a particular effort needs to be made for the measurement of rare species, which are still under-represented in trait databases (Violle, Borge, et al., 2015; Violle, Choler, et al., 2015). This is all the more important as the omission of rare species can strongly affect other metrics of functional diversity (Pakeman, 2014).

The trade-off observed between NIV and P_{Cover} is not a trivial issue because these two variables are dependent: increasing the minimum

NIV necessarily decreases P_{Cover} [Figure 6]. In our study, the predominant role of P_{Cover} for CWM estimations suggests that, as long as intra-specific trait variability is lower than interspecific trait variability—which seems to be the case for several traits investigated at large scales (Kattge et al., 2011; Kazakou et al., 2014) as theoretically expected (Albert et al., 2011)—it seems to be preferable to favour P_{Cover} over NIV. In the case of traits showing high intraspecific variability and/or in studies considering a more restricted environmental gradient, we hypothesize that reducing NIV will have a larger effect on estimates of CWM.

The questions we address are of relevance in field studies: How many species should be covered with trait measurements in each community [P_{Cover}] of a study area? How many individuals should be measured to capture the trait variation of each species in the study area [NIV]? As stated in the Introduction, Equation 1 assumes a single trait value for each species in the calculation of CWM. Even in case of traits measured in a single field study, this value is most often a mean of measured values of several individuals from different plots (see, e.g., Bernard-Verdier et al., 2012). We proposed Equation 2 for a more systematic sensitivity analysis of CWM with regard to data quality and

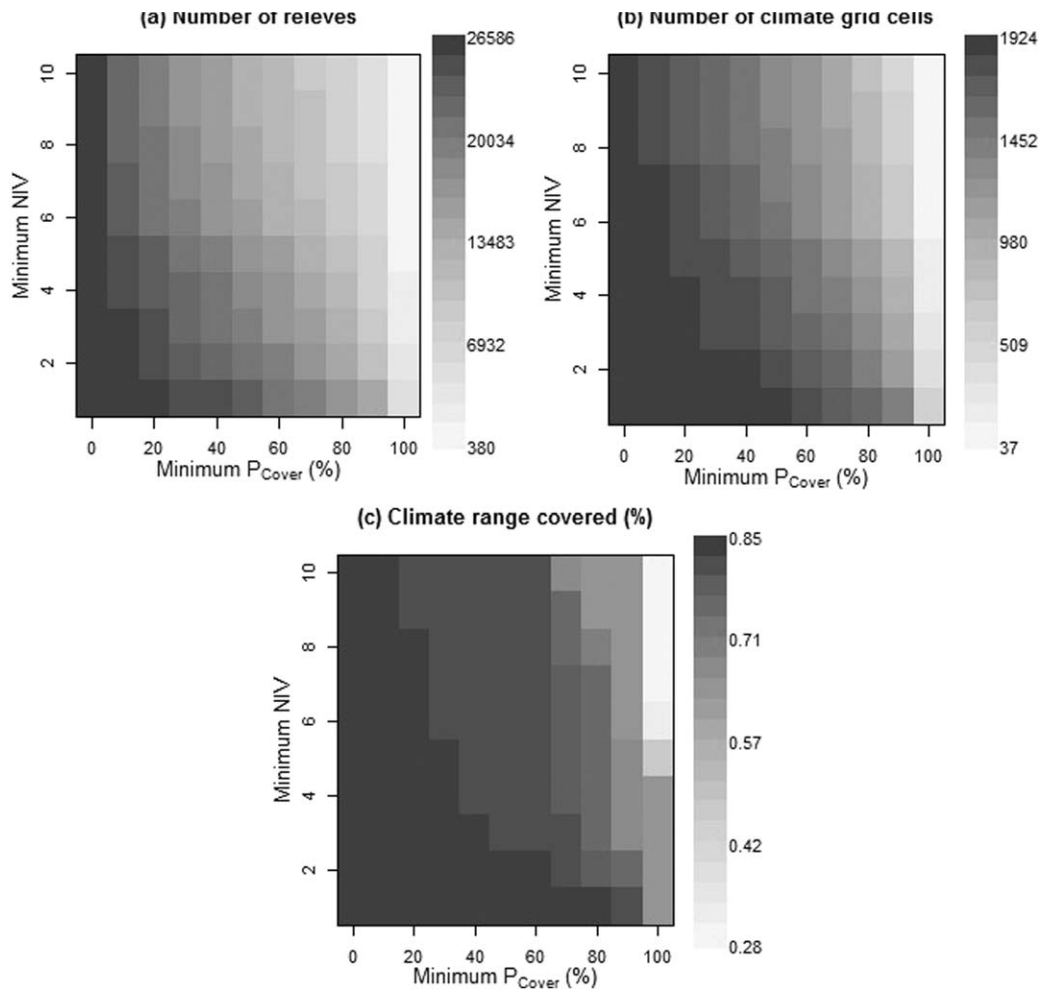


FIGURE 4 Effect of dataset selection criteria (i.e., minimum number of individual trait values per species [NIV] and community-level cover proportion [P_{Cover}] required) on the dataset size and gradient representativeness: (a) number of vegetation plots, (b) number of pixels and (c) total covered climate [GDD] gradient

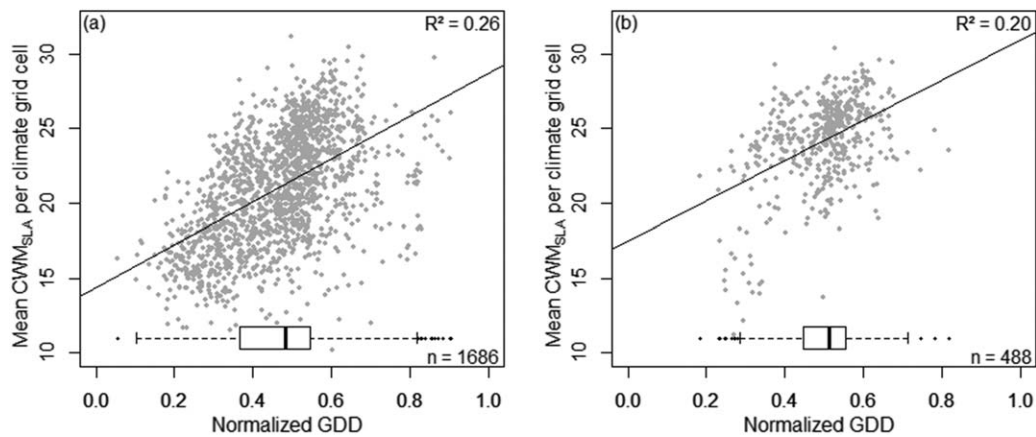


FIGURE 5 Impact of the dataset selection criteria on the growing degree day [GDD] gradient coverage and on the relationship between community-weighted mean for specific leaf area [CWM_{SLA}] and GDD at the climate grid-cell level. Both datasets were reduced to pixels with at least three vegetation plots. n represents the number of pixels. (a) The dataset was reduced to species with at least two individual values in TRY and plots with $P_{Cover} \geq 60\%$. In (b) the dataset was reduced to species with at least 10 individual values [NIV] in TRY and to plots with $P_{Cover} \geq 90\%$. The GDD gradient was normalized between 0 and 1. Boxplots represent the pixel distribution over the GDD gradient [maximum, mean and first and third quartiles]. Continuous black lines represent linear regressions

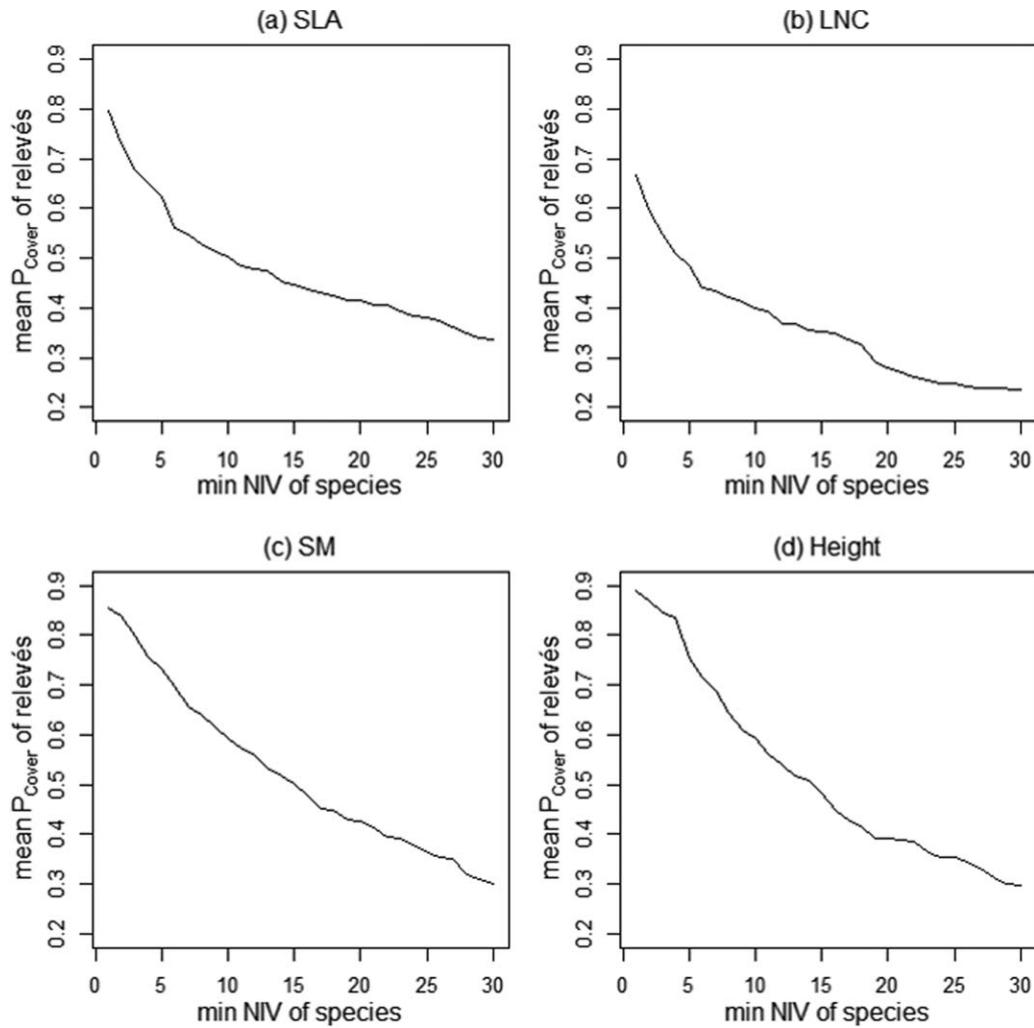


FIGURE 6 Trade-off between number of individual trait values per species [NIV] and the community-level cover proportion [P_{Cover}] across the whole dataset. Increasing the minimum NIV required for all species belonging to a vegetation plot reduces, on average, the P_{Cover} of this plot, for the four studied traits

coverage, even in local studies. In that case, NIV represents the number of individuals of a species where each individual is from the same study area [different plots or even a single plot], probably measured by the same researcher and with the same methodology. We prefigure that the overall relationships found in this study still hold for field studies even if the variation in trait values is smaller.

4.3 | Sensitivity of CWM–climate relationships to NIV and P_{Cover}

The results of this study indicate that the use of an incomplete trait dataset has implications for the ability to detect significant shifts in community functional structure along environmental gradients. This issue has been largely overlooked in the recent literature on “functional biogeography” (Lamanna et al., 2014; Šimová et al., 2015; Swenson & Enquist, 2007; Swenson & Weiser, 2010; Swenson et al., 2012). Our pilot study quantifies the effect of incomplete trait data on the relationship between CWM_{SLA} and GDD in temperate grasslands. It shows that setting a constraining high threshold for either [or both] NIV and

P_{Cover} (i) strongly reduces the width of the GDD gradient covered by the CWM_{SLA} estimates and (ii) tends to reduce the significance of the CWM_{SLA} –GDD relationship [Figure 5]. In our study, the reduced number of CWM_{SLA} in the low range of the GDD gradient mainly drove this effect of reduced significance. Given that the central parts of climate ranges are often well sampled while sampling is scarce in the margins, this calls for an intensification of sampling efforts of vegetation plots and traits at the extreme of the investigated climate gradients even if the area covered by these climate conditions is relatively small.

We here provide a generic approach for systematic sensitivity analyses to trade off between the magnitude of CWM error and the amount of available information. Overall, our findings highlight the trade-off between (i) the accuracy of estimates for CWM traits requiring high NIV and P_{Cover} and (ii) the ability to detect general trends of CWMs along a climate gradient requiring a large set of vegetation plots and, hence, more liberal criteria for NIV and P_{Cover} . We conclude that threshold values for NIV and P_{Cover} and trade-offs between NIV and P_{Cover} represent a critical feature of any functional biogeographical analysis with, often necessarily, limited data availability. In the case

study of permanent grasslands, P_{Cover} was the parameter with the greatest effect on CWM estimates and we consider this to be an argument for increasing species coverage in global trait databases. In the context of functional biogeography, a particular effort has to be made for those species occurring in under-sampled and harsh environmental conditions.

ACKNOWLEDGEMENTS

This research is supported by the French Foundation for Research on Biodiversity [FRB; <http://www.fondationbiodiversite.fr/>] in the context of the CESAB project “Assembling, analysing and sharing data on plant functional diversity to understand the effects of biodiversity on ecosystem functioning: a case study with French Permanent Grasslands” [DIVGRASS]. We are very grateful to the Fédération des Conservatoires Botaniques Nationaux for providing a useful dataset of vegetation relevés. C.V. was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Program [DiversiTraits project, no. 221060] and by the European Research Council [ERC] Starting Grant Project “Ecophysiological and biophysical constraints on domestication in crop plants” [grant ERC-StG-2014-639706-CONSTRAINTS]. The study has been supported by the TRY initiative on plant traits [<http://www.try-db.org>]. TRY is/has been supported by DIVERSITAS, IGBP, the Global Land Project, the UK Natural Environment Research Council [NERC] through its program QUEST [Quantifying and Understanding the Earth System], the French Foundation for Biodiversity Research [FRB] and GIS “Climat, Environnement et Société” France.

REFERENCES

- Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G., & Violle, C. (2011). When and how should intraspecific trait variability be considered in plant ecology? *Perspectives in Plant Ecology, Evolution and Systematics*, 13, 217–225.
- Baraloto, C., Paine, C., Patino, S., Bonal, D., Herault, B., & Chave, J. (2010). Functional trait variation and sampling strategies in species-rich plant communities. *Functional Ecology*, 24, 208–216.
- Benichou, P., & Le Breton, O. (1986). Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*, 7, 23–34.
- Bernard-Verdier, M., Navas, M. L., Vellend, M., Violle, C., Fayolle, A., & Garnier, E. (2012). Community assembly along a soil depth gradient: Contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. *Journal of Ecology*, 100, 1422–1433.
- Cingolani, A., Cabido, M., Gurvich, D., Renison, D., & Diaz, S. (2007). Filtering processes in the assembly of plant communities: Are species presence and abundance driven by the same traits? *Journal of Vegetation Science*, 18, 911–920.
- Cordlandwehr, V., Meredith, R., Ozinga, W., Bekker, R., van Groenendael, J., & Bakker, J. (2013). Do plant traits retrieved from a database accurately predict on-site measurements? *Journal of Ecology*, 101, 662–670.
- De Bello, F., Carmona, C., Mason, N., Sebastia, M. T., & Leps, J. (2013). Which trait dissimilarity for functional diversity: Trait means or trait overlap? *Journal of Vegetation Science*, 24, 807–819.
- Diaz, S., Lavorel, S., Chapin, F. III, Tecco, P., Gurvich, D., & Grigulis, K. (2007). Functional diversity—at the crossroads between ecosystem functioning and environmental filters. In J. Canadell, D. Pataki, & L. Pitelka (Eds.), *Terrestrial ecosystems in a changing world* (pp. 81–91). Berlin: Springer-Verlag.
- Diaz, S., Lavorel, S., De Bello, F., Quétier, F., Grigulis, K., & Robson, M. (2007). Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences USA*, 36, 20684–20689.
- Enquist, B. J., Norberg, J., Bonsor, S. P., Violle, C., Webb, C. T., Henderson, A., ... Savage, V. M. (2015). Scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Advances in Ecological Research*, 52, 249–318.
- Gardarin, A., Garnier, E., Carrère, P., Cruz, P., Andueza, D., Bonis, A., ... Kazakou, E. (2014). Plant trait–digestibility relationships across management and climate gradients in permanent grasslands. *Journal of Applied Ecology*, 51, 1207–1217.
- Garnier, E., & Navas, M. L. (2012). A trait-based approach to comparative functional plant ecology: Concepts, methods and applications for agroecology. A review. *Agronomy for Sustainable Development*, 32, 365–399.
- Garnier, E., & Navas, M. L., & Grigulis, K. (2016). *Plant functional diversity: organism traits, community structure, and ecosystem properties*. New York: Oxford University Press.
- Garnier, E., Cortez, J., Billes, G., Navas, M. L., Roumet, C., Debussche, M., ... Toussaint, J. P. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology*, 85, 2630–2637.
- Garnier, E., Laurent, G., Bellmann, A., Debain, S., Berthelie, P., Ducout, B., ... Navas, M. L. (2001). Consistency of species ranking based on functional leaf traits. *New Phytologist*, 152, 69–83.
- Grime, J. P. (1998). Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. *Journal of Ecology*, 86, 902–910.
- Harper, J. L., Lovell, P. H., & Moore, K. G. (1970). The shapes and sizes of seeds. *Annual Review of Ecology and Systematics*, 1, 327–355.
- Hulshof, C., & Swenson, N. (2010). Variation in leaf functional trait values within and across individuals and species: An example from a Costa Rican dry forest. *Functional Ecology*, 24, 217–223.
- Kattge, J., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönsch, G., ... Wirth, C. (2011). TRY: A global database of plant traits. *Global Change Biology*, 17, 2905–2935.
- Kazakou, E., Violle, C., Roumet, C., Navas, M. L., Vile, D., Kattge, J., & Garnier, E. (2014). Are trait-based species’ rankings consistent across datasets and spatial scales? *Journal of Vegetation Science*, 25, 235–247.
- Keddy, P. A. (1992). A pragmatic approach to functional ecology. *Functional Ecology*, 6, 621–626.
- Lamanna, C. A., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., Simova, I., ... Enquist, B. J. (2014). Functional space and the latitudinal species richness gradient. *Proceedings of the National Academy of Sciences USA*, 111, 13745–13750.
- Lavorel, S. (2013). Plant functional effects on ecosystem services. *Journal of Ecology*, 101, 4–8.
- Mason, N., & De Bello, F. (2013). Functional diversity: A tool for answering challenging ecological questions. *Journal of Vegetation Science*, 24, 777–780.
- Mouchet, M., Villéger, S., Mason, N., & Moullot, D. (2010). Functional diversity measures: An overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology*, 24, 867–876.

- Pakeman, R. (2014). Functional trait metrics are sensitive to the completeness of the species' trait data? *Methods in Ecology and Evolution*, 5, 9–15.
- Pakeman, R., & Queded, H. (2007). Sampling plant functional traits: What proportion of the species need to be measured? *Applied Vegetation Science*, 10, 91–96.
- Sandel, B., Gutiérrez, A. G., Reich, P. B., Schrod, F., Dickie, J., & Kattge, J. (2015). Estimating the missing species bias in plant trait measurements. *Journal of Vegetation Science*, 26, 828–838.
- Schleuter, D., Daufresne, M., Massol, F., & Arguillier, C. (2010). A user's guide to functional diversity indices. *Ecological Monographs*, 80, 469–484.
- Šímová, I., Violle, C., Kraft, N. J. B., Storch, D., Svenning, J. C., Boyle, B., ... Enquist, B. J. (2015). Shifts in trait means and variances in North American tree assemblages: Species richness patterns are loosely related to the functional space. *Ecography*, 38, 649–658.
- Swenson, N., & Enquist, B. (2007). Ecological and evolutionary determinants of a key plant functional trait: Wood density and its community-wide variation across latitude and elevation. *American Journal of Botany*, 94, 451–459.
- Swenson, N. G., Enquist, B. J., Pither, J., Kerkhoff, A. J., Boyle, B., Weiser, M. D., ... Noling, K. M. (2012). The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography*, 21, 798–808.
- Swenson, N., & Weiser, M. (2010). Plant geography upon the basis of functional traits: An example from eastern North American trees. *Ecology*, 91, 2234–2241.
- Taudiere, A., & Violle, C. (2016). cati: An R package using functional traits to detect and quantify multi-level community assembly processes. *Ecography*, 39, 699–708.
- Violle, C., Borge, B., & Choler, P. (2015). Trait databases: Misuses and precautions. *Journal of Vegetation Science*, 26, 826–827.
- Violle, C., Choler, P., Borge, B., Garnier, E., Amiaud, B., Debarros, G., ... Viovy, N. (2015). Vegetation ecology meets ecosystem science: Permanent grasslands as a functional biogeography case study. *Science for the Total Environment*, 534, 43–51.
- Violle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C., Hulshof, C., ... Messier, J. (2012). The return of the variance: Intraspecific variability in community ecology. *Trends in Ecology and Evolution*, 27, 244–252.
- Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892.
- Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences USA*, 111, 13690–13696.
- Weiher, E. (2010). A primer of trait and functional diversity. In A. E. Magurran, & P. J. McGill (Eds.), *Biological diversity: Frontiers in measurement and assessment* (pp. 175–193). Oxford: Oxford University Press.

BIOSKETCH

The DivGrass consortium is a CESAB working group including experts in grassland ecology, soil science, ecosystem modelling and conservation management. The objective of the consortium is to combine existing data on the plant diversity of French permanent grasslands to better understand the links between biodiversity and ecosystem functioning.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Borge B, Violle C, Choler P, et al. Sensitivity of community-level trait–environment relationships to data representativeness: A test for functional biogeography. *Global Ecol Biogeogr.* 2017;26:729–739. <https://doi.org/10.1111/geb.12573>