What we talk about when we talk about climate services
Serna Chavez, H.M.

Citation for published version (APA):
Serna Chávez, H. M. (2016). What we talk about when we talk about climate services

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
CHAPTER 3

Global patterns and determinants of microbial abundance in soil

Based on article:
Summary

Soil microorganisms play key roles in Earth's biogeochemical cycles, still, little is known on the spatial distribution of soil microbial biomass. In this chapter, using a global dataset of georeferenced soil microbial biomass estimates, and climatic and soil data, we present a comprehensive assessment of the abiotic determinants of soil microbial biomass at a global scale. We show that microbial biomass carbon ($C_{\text{Mic}}$) is primarily driven by moisture availability, with this single factor accounting for 34% of the global variance. For the microbial carbon-to-soil organic carbon ratio ($C_{\text{Mic}}/C_{\text{Org}}$), soil nitrogen content was an equally important determinant as moisture. Temperature was not a strong constraint of microbial biomass patterns at a global scale, while temperature likely has an indirect effect on microbial biomass by influencing rates of evapotranspiration and decomposition. As our models explain an unprecedented 50% of the global variance in $C_{\text{Mic}}$ and $C_{\text{Mic}}/C_{\text{Org}}$, we were able to leverage gridded abiotic information to build the first, robust spatially explicit global estimates of microbial biomass, quantifying the global soil microbial carbon pool to equal 14.6 Pg (Pentagrams). Our unbiased models allowed us to build the first global spatially explicit predictions of microbial biomass. These patterns show that soil microbial biomass is not primarily driven by temperature, but instead, its spatial distribution is more heterogeneous through the effects of moisture availability and soil nutrients. Our global estimates provide important data for integration into large-scale carbon and nutrient models that may imply a major step forward in our ability to predict the global carbon exchanges between ecosystems and the atmosphere, now and in a future climate.
3.1 Introduction

Ecologists have long been intrigued by the distribution of life on Earth. Information on the distribution and abundance of aboveground plant biomass has been available for decades, improving our understanding of the processes structuring plant communities, influencing species abundances and ecosystem processes. In contrast, efforts to estimate global patterns in abundance and the size of the soil microbial biomass pool have lagged considerably. Given the key role soil microorganisms play in ecosystem processes, elucidating such patterns will represent a critical step as we seek to improve our ability to predict and understand microbial controls on biogeochemical cycles.

Local and regional-scale studies have identified environmental determinants of microbial biomass carbon ($C_{\text{Mic}}$) and of the biologically active fraction of the soil organic carbon pool ($C_{\text{Mic}}/C_{\text{Org}}$) (Insam, 1990; Wardle, 1992; Franzluebbers et al., 2001; Bachar et al., 2010; Drenovsky et al., 2010; Dequiedt et al., 2011). However, the effect and direction of these determinants are rarely consistent across studies. Further identification and quantification of determinants of soil microbial abundance and activity have been hampered by lack of representative data collected using uniform (standard) methodologies. Global patterns may be presumed, however, as evidenced by stoichiometric and metabolic constraints determining patterns of microbial abundance and activity across all soils (Manzoni et al., 2008; Sinsabaugh et al., 2008; Cleveland & Liptzin, 2007). Fierer et al. (2009) presented results that suggest the variability in $C_{\text{Mic}}$ across major biomes is strongly correlated to soil organic carbon (SOC) content and plant productivity. However, by using biome-level means they omitted intra-biome variation, which may hinder the identification of those factors influencing microbial biomass levels across and within biomes. There have been only few attempts to quantify, let alone map, global patterns of soil microbial biomass. Such quantifications are critical given that current global estimates of the soil microbial biomass pool vary widely, from 13.9, 23.2 to 26 Pg C (Wardle, 1992; Xu et al., 2013; Whitman et al., 1998; 1 Pg=$1\times10^{15}$ g). Given that the data used to generate these estimates were largely derived from extrapolations based on temperate forest and cropland soils, a more accurate determination of the size of the microbial biomass pool is needed.

Uncertainties in previous estimates leave an important gap in our understanding of the soil microbial world and soil carbon dynamics. Elucidating the abiotic factors determining the size of the soil microbial pool is essential to better evaluate global fluxes and cycling of soil carbon and soil nutrients alike, as all are
strongly influenced by soil microbial communities. Also, the interactions between above and belowground communities are increasingly understood; however, it is still unknown if the spatial patterns of soil microbial abundance mirror those of plant biomass. In this chapter, we focus on quantifying and explaining global patterns of soil microbial biomass and its relationship to climate and soil factors. By subsequently applying our multivariate models to gridded climate and soil information, we derive and present the first robust spatially explicit global estimates of $C_{\text{Mic}}$ in the soil profile and the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio. Our spatially explicit estimates of the soil microbial pool can substantially improve predictions of global biogeochemical and vegetation models and, thus, allowing better estimations of the carbon balance.

### 3.2 Materials & Methods

#### 3.2.1 Soil microbial biomass carbon estimates

We based our study on a dataset of soil microbial biomass estimates across all major biomes (Cleveland & Liptzin, 2007, as extended and modified by Fierer et al., 2009). To improve latitudinal representation, we complemented the dataset with estimates from arctic and tropical regions. The final dataset comprised 414 georeferenced estimates (in mg $C_{\text{Mic}}$ kg$^{-1}$ soil), most of which had been obtained from mineral soils and only a few from organic soils. All estimates were from the A horizons obtained under control (non-manipulated) conditions. No estimates from litter layers were included, as determinants of microbial biomass in litter layers are presumably different from those in soil. Also, because seasonal variation in microbial biomass was seldom reported, we only present annual mean microbial biomass. The majority of estimates (92%) were obtained using the Chloroform Fumigation-Extraction (CFE) method (Vance et al., 1987); only 8% of estimates in the dataset were obtained from Chloroform Fumigation-Incubation (CFI) (Jenkinson & Powlson, 1976). For both methods, microbial biomass in the depth-interval sampled was calculated as:

$$C_{\text{Mic}} = E_C / k_{EC}$$  \[1\]

where $E_C$ represents the difference between fumigated and unfumigated soil extractable Carbon or CO$_2$ produced; and $k_{EC}$ the extraction efficiency accounting for
incomplete mineralization of biomass (CFI) or incomplete extraction of the killed biomass (CFE). While CFE and CFI are subject to biases in soils with high organic matter content and high acidity, respectively (Martens, 1995), no other method to quantify microbial biomass is as widely used nor bias-free (Martens, 1995; Joergensen et al., 2011). Moreover, we found no systematic deviations in microbial biomass estimates under potentially biasing conditions for CFE or CFI (see Appendix B1).

To facilitate comparisons across studies, we calculated EC using the reported $C_{\text{Mic}}$, the applied extraction coefficient ($k_{\text{EC}}$) and equation [1]. Afterwards, a $k_{\text{EC}}$ of 0.4 was applied to standardize all estimates. In addition, considering soil bulk density and the depth interval sampled, we extrapolated microbial biomass content to the first meter of soil profile ($C_{\text{Mic}}$ in g C m$^{-2}$). This extrapolation was made assuming that soil conditions change continuously and similarly across all soils down to the first meter (Fierer et al., 2009) (Appendix B1). Since bulk density was usually not given for the individual studies, thus we applied the bulk density reported in the global database for the specific soil type (Batjes, 1995) if the estimate was from a mineral soil. Given that the bulk density of organic soils varies much more strongly, we used the value reported in the original publication (or from other publications describing the same site). In addition, because all depth intervals sampled for microbial biomass represented soil conditions in the upper 30 cm, we calculated the topsoil microbial carbon–to-soil organic carbon ratio ($C_{\text{Mic}}/C_{\text{Org}}$), considered an index of microbial activity. First, we normalized all estimates in the dataset to the amount of microbial biomass (g C$_{\text{Mic}}$ m$^{-2}$) in the topsoil (Appendix B1). Then for mineral soils we extracted topsoil SOC from the soil profile database (Batjes, 1995), whereas for organic soils SOC was obtained from the original publication.

3.2.2 Soil data

We used the ISRIC-WISE 0.5$^\circ$-lat × 0.5$^\circ$-long global database of soil profiles (Batjes, 1995) to extract soil variables, including: bulk density (g cm$^{-3}$), soil organic carbon (SOC, % mass); cation exchange capacity (CEC, meq 100 g$^{-1}$); C-N ratio (g C g$^{-1}$ N); pH (in H$_2$O solution); total nitrogen (% mass), and texture variables (% of sand and clay). These parameters reflect the physical-chemical characteristics of mineral soils (0-30 cm). Parameters for organic soils (0-30 cm) were extracted from the original or from related publications. Most commonly, studies reported SOC, total nitrogen, bulk density and pH. The missing parameters for organic soils were excluded from
the analysis. These actions were a pragmatic solution to minimize biases due to missing organic soil data. Not considering microbial biomass estimates from organic layers may have led to underrepresentation or misrepresentation of particular global regions.

We used topsoil parameters to assess environmental determinants of soil profile \( C_{\text{Mic}} \) since they had higher correlation coefficients and accounted for more of the global variation in microbial biomass. We also ran analyses calculating soil parameters as (a) arithmetic mean of topsoil (0-30 cm) and subsoil (30-100 cm) data as reported in the ISRIC-WISE database, and (b) as weighted-means according to their assumed relative importance to microbial biomass (87% topsoil, 13% subsoil). In both cases, neither correlation coefficients nor the explanatory power of multivariate models was higher than when using topsoil parameters alone.

3.2.3 Climate data

We used the climate research unit (CRU) 0.5º-lat \( \times \) 0.5º-long surface mean monthly climatology database of global land areas excluding Antarctica (New et al., 1999). Using the georeferences of the biomass sampling locations, we calculated mean annual precipitation (MAP, mm H\(_2\)O yr\(^{-1}\)); maximum monthly frequencies of wet and frost days (MaxWD and MaxFD, respectively); lowest annual minimum monthly temperature (MinAT, in °C); highest annual maximum monthly temperature (MaxAT, °C); mean annual temperature (MAT, °C); and annual temperature range (TAR = MaxAT – MinAT, in °C). MaxWD is defined as the highest monthly number of days where precipitation exceeded 1.0 or 0.1 mm. Similarly, MaxFD represents the highest monthly number of days with gross minimum temperature below 0° C. MinAT and MaxAT are defined as the lowest minimum and highest maximum monthly temperatures in a year, respectively. Additionally, we calculated mean monthly and annual evaporative demand (ETo, mm H\(_2\)O day\(^{-1}\) and mm H\(_2\)O yr\(^{-1}\)), mean annual soil moisture deficit (MASMD, mm H\(_2\)O) and the annual ratio of MAP and ETo (MAP/ETo) for each biomass estimate. MASMD is the sum of differences between mean monthly ETo and mean monthly precipitation (Fierer & Jackson, 2006). ETo was calculated according to the FAO Penman-Montheith equation using climate data and assuming constant vegetation and roughness attributes, which makes the estimate independent of the standing vegetation (Allen et al., 1998).
3.2.4 Statistical analysis

Multivariate regression analyses were performed to quantify global abiotic determinants and to model soil microbial biomass as a linear function of multiple explanatory variables. Estimates of soil profile $C_{Mic}$ (g $C_{Mic}$ m$^{-2}$) and topsoil $C_{Mic}/C_{Org}$ ratio were used as response variables and analysed individually. Log$_{10}$-transformations were applied to both response variables to approach normal data distribution. Subsequently, all soil and climate variables were plotted against the transformed response variables to determine which to include in the regression analysis. In each case, Pearson’s correlation coefficients ($r$, $\alpha$=0.05) were calculated. We log$_{10}$-transformed some soil and climate variables to approach normal data distribution, enhance $r$, and improve linear relations between the explanatory and response variables. These transformations were performed separately for each analysis, as both response variables showed different patterns with explanatory variables. Upon the log$_{10}$-transformations, essentially linear relations were obtained.

To identify and select the most parsimonious set of abiotic variables that explain $C_{Mic}$ and $C_{Mic}/C_{Org}$ variance, we performed one-by-one backward stepwise regressions. To manage multicollinearity between environmental variables we only included in our models explanatory variables of which the $r$ among them varied between 0.7 and -0.7. Additionally, variance inflation factors (VIF) were calculated for all regressors in multivariate models and a limit of VIF $\leq$ 4 was set. Criteria to drop variables were: collinearity with other explanatory variables, significance of regression coefficients, single term deletion $F$-test, and the overall improvement of model’s fit after removal. The best multivariate models were selected according to the adjusted coefficient of determination ($R^2_{adj}$), Akaike’s Information Criteria (AIC), and regression diagnostic plots (studentized residuals versus fitted values, and distribution of studentized residuals). For the developed models, we assessed the relative variance accounted for by each explanatory variable using a hierarchical partitioning metric that accounts for the direct effect of each regressor and the adjusted effect for all other regressors ($lmg$ metric, Groemping, 2006).

As a control, we developed a null model to explain soil profile $C_{Mic}$ and topsoil $C_{Mic}/C_{Org}$ using ‘biome’ as sole explanatory variable. All estimates in our dataset are classified as boreal forest, desert, temperate coniferous forest, temperate deciduous forest, temperate grassland, tropical forest, or tundra. This simple biome model served to analyse inter-biome differences and to compare further multivariate models with, that is, to assess whether or not we improved the proportion of explained variance by using continuous relationships with, abiotic variables. The proportion of explained variance by inter-biome differences alone was calculated as:
\[ \eta^2 = \left( \frac{SS_{\text{treatment}}}{SS_{\text{treatment}} + SS_{\text{residuals}}} \right) \times 100 \]  

Furthermore, to cross-validate the developed multivariate models we compared our estimates to a dataset of substrate-induced respiration (SIR) estimates (\(\mu g \text{ C g}^{-1} \text{ h}^{-1}\)) from 77 soil samples. This dataset comprised soil samples from an array of temperate soils in North America and a few tropical soils in Puerto Rico, as described in Fierer & Jackson (2006). SIR estimates were obtained for the top 30 cm of mineral soil using the methodology described in Fierer et al. (2003). For each location where SIR had been measured we predicted soil profile \(C_{\text{Mic}}\) based on its lat-long locations to derive soil and climate variables, using the 0.5º-lat \(\times\) 0.5º-long global soil profile (Batjes, 1995) and climate (New et al., 1999) databases, and our multivariate models. We then assessed the significance of correlations (Pearson’s \(r\), \(a=0.05\)) between our \(C_{\text{Mic}}\) estimates and the SIR-responsive biomass data. Finally, we mapped the global soil \(C_{\text{Mic}}\) and \(C_{\text{Mic}}/C_{\text{Org}}\) ratio applying our multivariate models to the gridded climate and the soil profile information for each 0.5º\(\times\)0.5º grid cell, following appropriate back-transformations. All statistical analyses were conducted in R v.2.13.0, while the application of the developed regression models to derive the world-maps was conducted in Fortran using Force 2.0 v.2.0.9p.

3.3 Results

3.3.1 Quantification of global abiotic determinants

Overall, moisture variables were the main individual determinants of soil profile \(C_{\text{Mic}}\). Correlation with the MAP-ETo ratio (MAP/ETo) was the strongest \((r=0.46, P<0.0001)\) followed by MAP \((r=0.45, P<0.0001)\), both common measures of moisture availability and supply. Also, the interaction between moisture availability and MAT (MAP/ETo \(\times\) MAT) exhibited a strong linear relationship with \(C_{\text{Mic}}\) \((r=0.57, P<0.0001)\). This interaction indicates that with increasing MAT, the linear relationship between moisture availability (MAP/ETo) and \(C_{\text{Mic}}\) becomes stronger. From our temperature-related variables only annual TAR had a strong correlation with \(C_{\text{Mic}}\) \((r=-0.33, P<0.0001)\). All other variables, including MAT, exhibited low correlation coefficients. Soil characteristics also showed an important influence; while the relation between SOC and \(C_{\text{Mic}}\) was not as strong as expected \((r=0.14, \ldots)\).
P=0.004), total nitrogen and pH were quite strongly correlated to C\textsubscript{Mic} (r=0.25 and -0.26, respectively; P<0.0001).

The analysis of topsoil C\textsubscript{Mic}/C\textsubscript{Org} ratio suggests that soil and temperature variables exert a stronger influence on this ratio than for C\textsubscript{Mic}. Variables associated with soil moisture availability had the also strongest correlation coefficients with topsoil C\textsubscript{Mic}/C\textsubscript{Org} ratio. However, it was MAP instead of MAP/ETo \times MAT that exhibited the strongest correlation (r=0.43 and 0.32, respectively, P<0.0001). The influence of nutrients was exemplified by the strong correlation with total nitrogen content (r= -0.37, P<0.0001), and to a lesser extent with the carbon-nitrogen (C-N) ratio (r= -0.25, P<0.0001). MAT and the MinAT exhibited strong correlation with C\textsubscript{Mic}/C\textsubscript{Org} (r= 0.33; P<0.0001, in both cases). Also, the MaxFD strongly constrained C\textsubscript{Mic}/C\textsubscript{Org} (r= -0.36, P<0.0001). For both microbial biomass variables, all correlates and scatterplots are presented in Appendix B2.

3.3.2 Multivariate models to predict global patterns of microbial biomass

We selected the best multivariate linear models that explained soil profile C\textsubscript{Mic} and topsoil C\textsubscript{Mic}/C\textsubscript{Org} (Table 1). Selected models had a distribution of studentized residuals close to normal, with no distinct patterns between residuals and the explanatory variables in the models, or with those variables not included in the model. Also, no true outliers were detected (Cook’s D<0.08; Hat-values<0.09; Bonferroni’s adjusted P>0.05).

Soil profile C\textsubscript{Mic} was best explained by a combination of MAP/ETo \times MAT, the MaxWD, soil pH and total nitrogen, accounting for 39.0% of the total variance (Model 1, F\textsubscript{4,386}=63.2, P<0.0001, AIC\textsubscript{M1}= -875.6; Fig. 1A). A relative importance analysis of regressors supported the finding that moisture availability variables are the main determinants of C\textsubscript{Mic} alone accounting for more than 80% of the explained variance (Fig. 2A). For topsoil C\textsubscript{Mic}/C\textsubscript{Org}, the best model included MAP/ETo \times MAT, MaxWD, soil total nitrogen, pH, C-N ratio and cation exchange capacity (CEC). It explained 50.7% of the total variance in C\textsubscript{Mic}/C\textsubscript{Org} (Model 4, F\textsubscript{6,360}=63.7, P<0.0001; AIC\textsubscript{M4}= -802.9; Fig. 1B). Here, the relative importance analysis suggested a more balanced contribution between moisture availability and soil nutrient variables as main determinants, together accounting for more than half of the explained variance (Fig. 2B). Alternative multivariate models, presenting the second best fit, are presented in Appendix B1.
Table 1. Global determinants of soil profile $C_{\text{Mic}}$ (Model 1 and 2) and the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio (Model 3 and 4).

<table>
<thead>
<tr>
<th>Determinant</th>
<th>$F$-value</th>
<th>$\beta$ (±1SEM)</th>
<th>$b$</th>
<th>$P$</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP/ETo$^\dagger \times$MAT</td>
<td>230.0</td>
<td>0.04 (±0.004)</td>
<td>0.61</td>
<td>***</td>
<td>2.6</td>
</tr>
<tr>
<td>MaxWD</td>
<td>4.8</td>
<td>0.01 (±0.005)</td>
<td>0.15</td>
<td>**</td>
<td>2.1</td>
</tr>
<tr>
<td>pH</td>
<td>14.1</td>
<td>0.09 (±0.022)</td>
<td>0.21</td>
<td>***</td>
<td>1.7</td>
</tr>
<tr>
<td>Total N$^\dagger$</td>
<td>4.5</td>
<td>0.13 (±0.062)</td>
<td>0.09</td>
<td>*</td>
<td>1.1</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP/ETo$^\dagger \times$MAT</td>
<td>237.3</td>
<td>0.04 (±0.004)</td>
<td>0.57</td>
<td>***</td>
<td>2.8</td>
</tr>
<tr>
<td>MaxWD</td>
<td>7.4</td>
<td>0.02 (±0.005)</td>
<td>0.23</td>
<td>***</td>
<td>2.3</td>
</tr>
<tr>
<td>pH</td>
<td>11.4</td>
<td>0.07 (±0.026)</td>
<td>0.17</td>
<td>**</td>
<td>2.3</td>
</tr>
<tr>
<td>C-N ratio$^\dagger$</td>
<td>20.7</td>
<td>-1.35 (±0.432)</td>
<td>-0.14</td>
<td>**</td>
<td>1.3</td>
</tr>
<tr>
<td>CEC</td>
<td>5.5</td>
<td>0.01 (±0.004)</td>
<td>0.10</td>
<td>*</td>
<td>1.1</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP$^\dagger$</td>
<td>131.4</td>
<td>0.65 (±0.082)</td>
<td>0.46</td>
<td>***</td>
<td>2.4</td>
</tr>
<tr>
<td>ETo$^\dagger$</td>
<td>24.9</td>
<td>-0.47 (±0.129)</td>
<td>-0.18</td>
<td>***</td>
<td>1.7</td>
</tr>
<tr>
<td>MaxWD</td>
<td>0.5</td>
<td>0.02 (±0.006)</td>
<td>0.24</td>
<td>***</td>
<td>2.5</td>
</tr>
<tr>
<td>Total N$^\dagger$</td>
<td>113.2</td>
<td>-0.82 (±0.103)</td>
<td>-0.50</td>
<td>***</td>
<td>2.8</td>
</tr>
<tr>
<td>C-N ratio$^\dagger$</td>
<td>61.4</td>
<td>-2.94 (±0.448)</td>
<td>-0.28</td>
<td>***</td>
<td>1.3</td>
</tr>
<tr>
<td>pH$^\dagger$</td>
<td>21.5</td>
<td>1.23 (±0.343)</td>
<td>0.19</td>
<td>***</td>
<td>2.1</td>
</tr>
<tr>
<td>CEC$^\dagger$</td>
<td>4.6</td>
<td>0.38 (±0.176)</td>
<td>0.12</td>
<td>*</td>
<td>2.1</td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP/ETo$^\dagger \times$MAT</td>
<td>83.5</td>
<td>0.04 (±0.004)</td>
<td>0.55</td>
<td>***</td>
<td>2.6</td>
</tr>
<tr>
<td>MaxWD</td>
<td>1.7</td>
<td>0.02 (±0.005)</td>
<td>0.23</td>
<td>***</td>
<td>2.4</td>
</tr>
<tr>
<td>Total N$^\dagger$</td>
<td>225.0</td>
<td>-0.96 (±0.095)</td>
<td>-0.59</td>
<td>***</td>
<td>2.5</td>
</tr>
<tr>
<td>C-N ratio$^\dagger$</td>
<td>43.1</td>
<td>-2.07 (±0.441)</td>
<td>-0.19</td>
<td>***</td>
<td>1.3</td>
</tr>
<tr>
<td>pH$^\dagger$</td>
<td>24.1</td>
<td>1.27 (±0.334)</td>
<td>0.20</td>
<td>***</td>
<td>2.0</td>
</tr>
<tr>
<td>CEC$^\dagger$</td>
<td>5.2</td>
<td>0.40 (±0.173)</td>
<td>0.12</td>
<td>*</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Significance codes: $P<0.001$ (***); <0.001 (**); <0.05 (*); $^\dagger$ Variable is log$_{10}$-transformed. $\beta$: Regression coefficients; SEM: Standard error; b: Standardized coefficients; VIF: variance inflation factors; MAP: mean annual precipitation; ETo: annual evaporative demand; MAT: mean annual temperature; MaxWD: maximum monthly frequency of wet days; Total N: soil total nitrogen content; CEC: soil cation exchange capacity; pH: soil pH; C-N ratio: soil Carbon-Nitrogen ratio.

For comparison, we also assessed inter-biome differences for our dataset (see Methods). Biome level means of $C_{\text{Mic}}$ were significantly different (ANOVA $F_{6,407}=10.6$, $P<0.0001$) particularly due to tropical forests exhibiting higher microbial biomass than several other biomes (Fig. 3A). A similar result was obtained for the biome level means of $C_{\text{Mic}}/C_{\text{Org}}$ (ANOVA $F_{6,390}=11.2$, $P<0.0001$; Fig. 3B). These inter-biome differences reflect inherent differences in abiotic conditions.
among biomes and account for 13.5% (7-19%) and 14.7% [8-20%; $\eta^2$ and 95% confidence interval (C.I.)] of the total $C_{\text{Mic}}$ and $C_{\text{Mic}}/C_{\text{Org}}$ variation, respectively. This level of variance is about three times less than the variance explained when directly coupling abiotic factors to biomass estimates. By directly estimating the impacts of abiotic factors on soil microbial biomass, we were able to account for the interbiome and intrabiome differences in microbial biomass.

Figure 1. Observed versus predicted soil microbial biomass. (A) Soil profile $C_{\text{Mic}}$ estimates from Model 1 (as provided in Table 1). Regression line: slope of 1.0 ($\pm 6.2 \times 10^{-2}; \beta \pm 1\text{SEM}$); $R^2$ of 39.8%, $P<0.0001$ ($n=391$). (B) Topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio estimated using Model 4 (as provided in Table 1). Regression line: slope of 1.0 ($\pm 6.2 \times 10^{-2}; \beta \pm 1\text{SEM}$); $R_{\text{adj}}^2$ of 50.7%, $P<0.0001$ ($n=367$). For both panels shadowed area represents the 95% C.I. and all axes are in log$_{10}$ scale. Legend and biome classification as follows: boreal forest (square, BF), desert (circle, D), tundra (triangle, T), temperate coniferous forest (plus, TCF), temperate deciduous forest (times, TDF), tropical forest (diamond, TF), and temperate grassland (inverted triangle, TG).
Figure 2. Relative variance component analysis of explanatory variables for (A) Model 1 estimating global soil profile $C_{\text{Mic}}$ ($R^2_{\text{adj}} = 39\%$) and (B) Model 4 estimating global topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio ($R^2_{\text{adj}} = 50.7\%$). In both cases, metrics are normalized to sum 100%. Confidence intervals (95%) were obtained after 1,000 bootstrap runs. Equal lower-case letter indicate that confidence intervals for the difference between the contributions of each predictor includes zero ($P<0.05$). Abbreviations as explained in Table 1.

Figure 3. Inter-biome differences in (A) soil profile $C_{\text{Mic}}$ ($n=414$) and (B) topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio ($n=397$). Values presented are arithmetic means ±1SEM. Similar lower-case letter indicate that means are not significantly different from each other (Tukey’s post hoc test, $\alpha=0.05$). Biome classification as follows: boreal forest (BF), desert (D), tundra (T), temperate coniferous forest (TCF), temperate deciduous forest (TDF), tropical forest (TF), and temperate grassland (TG).
We cross-validated our model estimates using microbial biomass estimated using a different methodology: SIR (µg C g⁻¹ h⁻¹; see Methods). Predicted C_{Mic} and SIR were better related to estimates from our best model (r = 0.39; P < 0.001) than to our alternative second-best model (r = 0.34; P < 0.05), although correlation coefficients were not statistically different from each other (Z_{calculated} = 0.35; P > 0.05). Given that this cross-validation compares local SIR estimates with grid-average microbial biomass, these correlation coefficients are reasonably high. More importantly, the residual variance of the cross-validation did not show deviations from null patterns, indicating unbiased estimates.

3.3.3 Global estimates of microbial biomass

The spatially explicit global patterns of microbial biomass derived from our multivariate models are presented in Fig. 4. Hotspots of soil profile C_{Mic} were in tropical regions; however, comparable abundances were evident in some subtropical, temperate, and boreal regions. The lowest C_{Mic} was estimated for arid and semi-arid regions (Fig. 4A). Our multivariate models led to an estimated global soil profile microbial biomass pool of 14.6 Pg C_{Mic} (± 0.007 Pg; 95% C.I.). For the topsoil C_{Mic}/C_{Org}, patterns indicated highest ratios in tropical and subtropical regions where soil mineralization rates are typically higher (Aerts & Chapin, 2000). Conversely, low ratios persisted in arid regions but also in most boreal and tundra regions (Fig. 4B). For the latter regions, low ratios were expected given substrate accumulation because of temperature and moisture constraints (Deyn et al., 2008), as were also apparent in our models. Our model estimated the global topsoil C_{Mic}/C_{Org} at 1.2% (0 – 2; 75% C.I.).

3.4 Discussion

3.4.1 Global determinants of microbial abundance

By analysing the continuous relations between abiotic factors and microbial biomass, we were able to explore the variations between and within biomes without any a priori classification, and reduced the residual variance in biomass estimates. In our analysis, moisture-related variables exhibited the strongest linear relationships with C_{Mic} and C_{Mic}/C_{Org}. This may be no surprise, as moisture availability can strongly
affect microbial activity in both wet and dry climates (Bachar et al., 2010; Blankinship et al., 2011) while also influencing nutrient availability. Although high soil moisture is known to restrict microbial activity (Drenovsky et al., 2010; Wu et al., 2011), we found no evidence for such limitations likely because anaerobiosis is sufficiently infrequent to limit microbial biomass in surface soils. Furthermore, because both C_Mic and C_Mic/C_Org were positively related to annual moisture availability, this suggests that microbial biomass increases more with precipitation than SOC. By including the inherent influence of temperature on annual moisture availability (MAP/ETo × MAT interaction) we were able to better explain the variability in microbial biomass. Because MAT alone showed little correlation with C_Mic, while annual ETo, MAP/ETo and the MASMD were all strongly correlated (Appendix B2), this suggests that temperature predominantly affects C_Mic indirectly by influencing evapotranspiration rates. Conversely, both temperature and moisture were important determinants of C_Mic/C_Org in contrast to effects found for C_Mic. This suggests that the influence of temperature on C_Mic/C_Org is mainly through its effects on SOC, which is known to be strongly dependent on temperature (Parton et al., 2007). Variables that expressed the seasonality of precipitation (e.g. coefficient of variation, or precipitation in the wettest quarter), as well as of temperature (e.g. temperature of the coldest quarter), was less able to account for the variance in either C_Mic or C_Mic/C_Org than mean annual variables (Appendix B1).

Soil parameters were also important determinants of C_Mic and especially of C_Mic/C_Org, suggesting that soil conditions have a stronger influence on the amount of organic carbon immobilized in microbial biomass than on biomass itself. The positive relationships we found between C_Mic, SOC and total nitrogen content are in line with previous evidence (Wardle, 1992; Cleveland & Liptzin, 2007; Fierer et al., 2009). Multivariate models that included total nitrogen instead of SOC explained a higher variance of C_Mic and were a better fit despite total nitrogen being highly correlated with SOC \( (R^2 = 95\%; \, F_{1,392} = 7281; \, P<0.0001) \). Thus, while the relation between C_Mic and total nitrogen also reflects an influence of SOC, our results suggest that effects of nitrogen limitation are more consistent. However, these effects seem to decrease at high substrate concentrations because low microbial biomass was found in tundra soils at high total nitrogen (and coinciding high SOC; Appendix B2).
Figure 4. Global 0.5º-lat × 0.5º-long spatially explicit estimates of soil microbial abundance. (A) Patterns depict soil profile C_{Mic} derived from Model-1 (as provided in Table 1). Areas in black represent estimates greater than 250 g C_{Mic} m^{-2}. The average C_{Mic} content was 97.2 g C_{Mic} m^{-2} (±0.45; 95% C.I.). (B) Patterns depict global estimates for topsoil C_{Mic}/C_{Org} derived from Model-4 (as provided in Table 1). Areas in black represent estimates greater than 5%. For both panels gray areas represent ice-covered surfaces.
These areas are characterized by climatic stress, such as prolonged low temperatures and long periods of moisture shortage because of frost, which can restrict decomposition leading to accumulation of organic matter (Wardle, 1992; Cleveland & Liptzin, 2007; Fierer et al., 2009). Also, this accumulated carbon may be more recalcitrant (lower substrate quality; Franzluebbers et al., 2001; Fierer et al., 2009). As depicted in our maps, low Cmic/Corg ratios are evident for these regions under climatic stress, while the estimated Cmic is similar to that found in tropical regions.

Unlike for Cmic, the effect of soil nitrogen on Cmic/Corg was negative. Given that Cmic/Corg is generally considered an index of microbial activity, these results suggest that in addition to limitations by soil nitrogen on microbial biomass, nitrogen may also hamper microbial activity. A biome-deviation analysis revealed that for all biomes, the relationship was indeed negative except for desert soils (see Appendix B1). The latter is expected because in arid regions, soil nitrogen is considered a limiting resource (Gallardo & Schlesinger, 1992). Thus, increased nitrogen availability could allow for an increase in Cmic/Corg if other environmental conditions are favourable. A negative relationship does not seem to be due to impaired decomposition by anoxic conditions with accumulated soil organic matter (as occurs in, e.g., humid tropical forest, Pett-Ridge et al., 2006), given that this cannot explain the negative impacts in other biomes. However, the negative relationship is consistent with nitrogen fertilization studies (Dijkstra et al., 2005; Treseder, 2008; Ramirez et al., 2010). This may be explained by a shift in the soil microbial community to one with lower standing biomass but higher turnover rates given higher nutrient availabilities (Allison et al., 2008; Ramirez et al., 2010; Eisenhauer et al., 2012) or a reduction in the allocation of belowground labile carbon with increasing soil nitrogen contents (Dijkstra et al., 2005; Allison et al., 2008; Eisenhauer et al., 2012), which may subsequently increase the carbon demand of microbial biomass and induce shifts in community structure (Allison et al., 2008). Given that aforementioned processes are intimately linked, they may represent complementary responses (with potentially different importance to subsets of our data) and may together help explain the overall negative global relationship between total nitrogen and Cmic/Corg.

Soil resource quality (C-N ratio) also exhibited an important influence on both microbial variables. High C-N ratios tend to indicate low organic matter decomposability, whereas higher substrate quality enhances microbial activity and growth (Parton et al., 2007; Dequiedt et al., 2011). Also, environmental stress can limit decomposition causing an accumulation of soil carbon, and a higher C-N ratio because accumulation of undecomposed plant litter. This might restrain microbial biomass abundance by reducing substrate quality. These mechanisms account for
the negative relationship between our response variables and soil C-N ratio. Furthermore, according to our models, $C_{\text{Mic}}/C_{\text{Org}}$ and $C_{\text{Mic}}$ increased with pH. Acidic conditions are known to restrict organic matter decomposition and microbial (enzymatic) activity. As pH increases, microbial activity follows, allowing the immobilization of a bigger proportion of organic carbon in microbial biomass (Wardle, 1992; Sinsabaugh et al., 2008). In accordance with previous studies (Anderson & Domsch, 1993; Baath & Anderson, 2003), this can translate into an overall positive effect of pH on microbial biomass. Soil texture variables also exhibited important influence on $C_{\text{Mic}}/C_{\text{Org}}$ and on $C_{\text{Mic}}$ (Appendix B2). It is likely that the relation between CEC and $C_{\text{Mic}}/C_{\text{Org}}$ reflects an enhanced retention of organic matter and reduced quality (Wardle, 1992; Six et al., 2002).

The patterns mentioned above may be the result of complex interacting effects in the soil environment and with climate (e.g., the climate, soil organic matter and pH relationship; Insam, 1990; Jobbagy & Jackson, 2000; Franzluebbers et al., 2001). Nonetheless, the determinants described here are likely the main actors due to the imposed constraints on collinearity between predictors in our models (see section 3.2).

### 3.4.2 Future improvements in soil microbial biomass estimates

We recognize two important sources of uncertainty. First, given the scale of the abiotic information used to build our models, local heterogeneity was not included, which likely explains an important proportion of residual variance in our models, particularly in the validation outcome. However, we expect that this did not affect our predicted global patterns because our predictions accurately portrayed average observations (see 1:1 patterns in Fig. 1). And, in spite of this discrepancy in scale, our validation showed similar patterns to the observed SIR as reported earlier for local datasets (Anderson & Joergensen, 1997; Fierer et al., 2003, 2009). Considering uncertainties in our models, the correlation between our biomass estimates and SIR data are in line with previous studies, further supporting the contention that our estimates and spatial patterns are unbiased. Second, to standardize measurements we extrapolated microbial biomass estimates from the depth-interval sampled to the first meter. This might have biased estimates for specific sites, e.g. where data was obtained from organic layers on top of mineral soils. However, even though depth distributions will differ among sites, and biomes, it presents a shape similar to the global average SOC depth distribution (Jobbagy & Jackson, 2000) and likely only contributed to residual variance in our models. Still, including soil parameters of
subsoil conditions did not improve our models (see section 3.2). Future work elucidating microbial biomass depth distribution in different soil environments will help making suitable extrapolations and, thus, better comparisons. A better characterization of soil parameters along the soil profile might help improving the resolution of our multivariate models.

3.4.3 Spatial patterns and estimates of microbial biomass

Our unbiased multivariate models allowed us to build global spatially explicit predictions of microbial biomass in soils. These patterns show that soil microbial biomass is not primarily driven by temperature (as reflected in latitude), but instead biomass is more heterogeneous through the effects of moisture availability and soil nutrients (Fig. 4). Similarly, precipitation and soil nutrients have been shown to influence patterns of plant biomass (Moles et al., 2009; Vicca et al., 2012). To assess whether global patterns in plant and soil microbial abundance coincide, we tested the relation between net primary production of potential vegetation (NPP in g m\(^{-2}\) year\(^{-1}\), from the Lund-Potsdam-Jena-Dynamic Global Vegetation Model, Sitch et al., 2003) and soil C\(_{\text{Mic}}\). Their correlation coefficient was comparable to those of annual moisture variables (\(r = 0.50, P<0.0001\); Appendix B2). However, a large portion of the variance in NPP is already accounted for by annual moisture supply and availability variables (e.g., MAP and MAP/ETo × MAT, in bivariate regressions: \(R^2=0.52\) and 0.45, respectively; \(P<0.0001\)). Indeed, including NPP in our models did not increase the amount of explained variance nor their fit. Instead, replacing the main determinants (i.e., MAP/ETo × MAT in the C\(_{\text{Mic}}\) model) with NPP led to a decrease in explained variance from \(R_{\text{adj}}^2=0.39\) (\(P<0.0001\); AIC =-876), to \(R_{\text{adj}}^2=0.28\) (\(P<0.0001\); AIC =-816). Compared to soil microbial biomass, plant biomass seems to be more constrained by temperature (Law et al., 2002), since MAT was more strongly correlated to NPP (\(r=0.39; P<0.0001\)) than to C\(_{\text{Mic}}\) (\(r=0.07; P>0.05\)). Together, this seems to reveal that (1) the extent and direction of these relationships with abiotic factors is different for plant NPP and soil microbial biomass, and (2) that annual climatic variables are better able than NPP to account for the global variance in soil microbial abundance.

Our comprehensive analysis provides a global estimate of 14.6 Pg C\(_{\text{Mic}}\) in the soil profile. This estimate falls within the range of previous estimates (Wardle, 1992; Whitman et al., 1998; Xu et al., 2013), which had been hampered by low representation in particular biomes, and lack of spatially explicit analysis of determinants. In addition, according to the well-conserved C : N : P ratio of
microbial biomass (Cleveland & Liptzin, 2007), our findings suggest that the global soil microbial nitrogen and phosphorous pools are around 2.0 and 0.6 Pg, respectively. Our study also estimates the global topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio at 1.2% (0-2%; 75% C.I.), which is in line with previous studies that estimated it for topsoil (cumulatively, 0.3 - 5%; Wardle, 1992; Insam, 1990; Zak et al., 1994; Xu et al., 2013) and for the soil profile (0.6 - 1.1%; Fierer et al., 2009), respectively.

Microbial biomass estimates used in our study are mostly representative of the active season, which matters most for ecosystem-atmosphere exchanges and related ecosystem services like climate regulation. Thereto, our estimates are well suited to understand and to further improve quantitative estimates of ecosystem-atmosphere exchanges, now and in the face of global environmental change. In fact, our quantitative estimates may serve many applications as soil microorganisms mediate the rates of carbon and nutrient cycling in terrestrial ecosystems (Wardle, 1992; Falkowski et al., 2008). Consequently, microbial biomass serves as a critical control on the feedbacks between global change, organic matter decomposition, and ecosystem productivity (Manzoni & Porporato, 2009; Todd-Brown et al., 2012).

Current global biogeochemical and vegetation models have only a coarse, if any, representation of soil microbial properties. Most commonly models frame decomposition of organic matter as first order decay, that is, only as a function of environment (temperature and moisture), and not of microbial biomass (Parton et al., 1993; Sitch et al., 2003), hampering the predictive ability of these models (Manzoni & Porporato, 2009; Todd-Brown et al., 2012). The problems with scaling up information on microbial abundance and activity from sparse local data (Todd-Brown et al., 2012) have made it difficult to integrate soil microbial biomass into global-scale decomposition models. Our global estimates of soil microbial biomass provide important data for integration into large-scale carbon and nutrient models, which may imply a major step forwards in our ability to predict the global carbon balance, now and in a future climate. Thereto, our global $0.5^\circ \times 0.5^\circ$ maps of $C_{\text{Mic}}$ and topsoil $C_{\text{Mic}}/C_{\text{Org}}$ are available online, as supplementary information of the published manuscript.

In summary, our study shows how using average abiotic variables allows us to explain a large portion of the global variance in soil microbial biomass. By identifying generic abiotic controls on soil microbial abundance we show the stronger constraints exerted by soil variables and temperature on $C_{\text{Mic}}/C_{\text{Org}}$ than on $C_{\text{Mic}}$, suggesting that topsoil microbial activity is further constrained by local soil conditions and low temperatures. Also, by deriving a spatially explicit global map of

---

1 Available at: http://onlinelibrary.wiley.com/doi/10.1111/geb.12070/suppinfo
soil profile $C_{Mic}$ and topsoil $C_{Mic}/C_{Org}$ distribution, we provide an improved spatially explicit estimate of the global biological soil carbon pool. Our estimates, in combination with new understanding of substrate and decomposer stoichiometry (Cleveland & Liptzin, 2007; Manzoni et al., 2008; Sinsabaugh et al., 2008), will together improve our ability to predict the controls and rates of organic matter decomposition, nutrient mineralization and soil-atmosphere carbon exchange.

3.5 Acknowledgements

The author H.S.C. would like to thank the European Commission's Education, Audiovisual and Culture Executive Agency (EACEA) for its support throughout the research period with an Erasmus Mundus Scholarship (2009-2011). Also to thank E. van Loon for comments on an earlier version of the manuscript.

3.6 References


Appendix B1
Supporting information, methods, and results

Soil microbial biomass estimates

The 414 georeferenced mean annual microbial biomass carbon estimates (mg \( C_{\text{Mic}} \) kg\(^{-1}\) soil) compiled in our dataset were first transformed to per area estimates (g \( C_{\text{Mic}} \) m\(^{-2}\)) using the bulk density value for the specific soil profile (g cm\(^{-3}\)) and the reported depth interval sampled (m). Subsequently, we extrapolated the microbial biomass carbon content below the sampling depth until the top meter of the soil profile using the following equation (Fierer et al., 2009):

\[
\text{Microbial biomass (to 1 m)} = [-0.132 \times \ln(d) + 0.605] \times B \tag{1}
\]

where \( \ln \) = natural logarithm; \( d \) = depth interval sampled (m), and \( B \) = the measured microbial biomass carbon (g \( C_{\text{Mic}} \) m\(^{-2}\)). The final soil profile estimate (\( C_{\text{Mic}} \), in g C m\(^{-2}\)) resulted from the sum of the biomass in the depth interval sampled and the biomass below the sampling depth obtained from equation [1]. Similarly, to obtain the topsoil microbial carbon-to-soil organic carbon ratio (\( C_{\text{Mic}}/C_{\text{Org}} \)) we used the per area estimates for the depth interval sampled (estimated as explained above). After, to extrapolate from the depth interval sampled to the first 30 cm of soil we used the following equation (adapted from Fierer et al., 2009):

\[
\text{Microbial biomass (to 0.3 m)} = [-0.12 \times \ln(d) + 0.597] \times B \tag{2}
\]

where \( \ln \) = natural logarithm; \( d \) = depth interval sampled (m), and \( B \) = the measured microbial biomass carbon (g \( C_{\text{Mic}} \) m\(^{-2}\)). The topsoil \( C_{\text{Mic}} \) was the sum of the biomass in the depth interval sampled and the biomass below the sampling depth from equation [2]. Some of the compiled estimates already represented the \( C_{\text{Mic}} \) in the top 30 cm of soil; here no extrapolations were performed and per area estimates
were used. Finally, the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratios were calculated dividing the calculated topsoil $C_{\text{Mic}}$ by the corresponding topsoil organic carbon (SOC) content for the specific soil profile. For mineral soils we extracted SOC from the soil profile database (Batjes, 1995), whereas for organic soils SOC was obtained from the original publication or from other publications describing the same site.

All estimates in our dataset, calculated as explained above, represent microbial carbon in soils, for none of the sites microbial carbon content in litter layers was accounted. Moreover, we only included “annual means” in our study since we seldom encountered a site sampled multiple times across seasons. Nonetheless, since soils are commonly sampled during the active season when microbial abundance, and activity, matters most for ecosystem fluxes and services, we think that such microbial biomass estimates (like those in our dataset) may be highly suitable to understand and to further improve quantitative estimates of ecosystem fluxes and services.

Most of estimates in our dataset were obtained using the Chloroform Fumigation-Extraction (CFE) method (Vance et al., 1987) with only a minor proportion from Chloroform Fumigation-Incubation (CFI) (Jenkinson & Powlson, 1976). As any method of microbial biomass, these methods may lead to biased estimates at particular conditions: CFE is known to make over- and under-estimations in soils with high content of organic matter and in those with low porosity, respectively. CFI is known to fail in acidic soils and in waterlogged soils (Jenkinson et al., 2005). However, the CFI estimates in our dataset were neither from acidic (all sampled $pH > 5$) nor in highly organic soils (all sampled soil organic carbon content $\leq 5\%$). On the other hand, for microbial biomass from the CFE method, around fifteen estimates were obtained from soils with soil organic carbon content $\geq 20\%$ (but $\leq 50\%$). To test whether these points deviate from the patterns described in the main text, we re-fitted our models omitting them. Overall, we found no significant deviations, for instance the difference between the re-fitted and the original model for soil profile $C_{\text{Mic}}$ was only minor (Original: $R^2_{\text{adj}} = 0.39$, $P<0.0001$; re-fitted: $R^2_{\text{adj}} = 0.40$, $P<0.0001$). Therefore, we consider that the known biases of both CFE and CFI methods do not affect the results of our study.

Environmental drivers of soil profile $C_{\text{Mic}}$ and topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio

To assess the relationship between $C_{\text{Mic}}$ and $C_{\text{Mic}}/C_{\text{Org}}$ with climate and soil variables we calculated Pearson’s correlation coefficients ($r; \alpha=0.5$). To test whether variables that expressed seasonality in precipitation and temperature were better drivers of
we compared coefficients of determination ($R^2$) from bivariate regressions. Seasonality variables were calculated using CRU 1.0 mean monthly climatologies (New et al., 1999) and the climates package v0.1.1.3 (VanderWal et al., 2012) for R v2.15.2 software. Annual variables were calculated from mean monthly climatologies (New et al., 1999), as described in the main text.

Bivariate regressions showed how variables that express the seasonality of precipitation and temperature are less able to account for the variance of microbial biomass than mean annual precipitation (see Table B1) and, were also highly correlated to mean annual climate variables. This supports our finding that annual moisture supply and availability, as well as soil nitrogen content, are the main individual drivers of soil profile $C_{\text{Mic}}$ and topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio. Appendix B2 presents scatterplots for the secondary drivers of soil microbial variables.

**Multivariate models for soil microbial biomass**

For the developed multivariate models that predict soil profile $C_{\text{Mic}}$ (Model 1 and 2) and the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio (Model 3 and 4) we conducted a test of model performance and also assessed the relative contribution of each predictor to the total explained variance using the $\text{lm}g$ function in the ‘relaimpo’ package (Groemping, 2006) for R v2.15.2 software (R Core Team, 2012). Figure B1 presents the results from such analyses for secondary models.

Model-2 included the interaction between moisture availability and mean annual temperature (MAP/ETo×MAT), the maximum monthly number of wet days (MaxWD), soil pH, carbon-nitrogen (C-N) ratio and cation exchange capacity (CEC). Although CEC was not significantly correlated to $C_{\text{Mic}}$ ($r=0.07$, $P>0.05$), it explained a significant proportion of the residual variance. Model-2 explained 43.1% of the total variation in soil profile $C_{\text{Mic}}$ ($F_{5,361}=56.5$, $P<0.0001$; $\text{AIC}_{\text{M2}}=-836.04$; Fig. B1A). The relative importance analysis for this model presented similar results as for the main model: MAP/ETo×MAT and MaxWD alone accounted for more than 80% of the explained variance and MAP/ETo×MAT alone accounted for more than half of the explained variance (Fig. B1C). Soil profile $C_{\text{Mic}}$ estimates from our Models-1 and 2 exhibited a mean of differences of 3.2 g $C_{\text{Mic}}$ m$^{-2}$ ($\pm 0.92$; 95% C.I.). Model-2 predicted a global microbial biomass carbon pool of 14.9 Pg C ($\pm 0.007$, 95% C.I.; 1 Pg=$1\times10^{15}$ g).

Model-3 included mean annual precipitation (MAP), MaxWD, annual evaporative demand (ETo), soil CEC, total nitrogen, pH and C-N ratio, and explained 48.9% of the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ variance ($F_{7,359}=51.1$, $P<0.0001$; $\text{AIC}_{\text{M3}}=-$
The relative importance analysis showed a balanced contribution between the main drivers: MAP and soil total nitrogen, together accounting for more than 50% of the explained variance (Fig. B1D). Model-3 estimated the global mean topsoil $C_{\text{Mic}}/C_{\text{Org}}$ at 1.03% (0 - 2; 75% C.I.). Global estimates of topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio from Models-3 and 4 had a mean of differences of 0.19% (±0.006; 95% C.I.).

Table B1.1 Environmental drivers of soil profile microbial biomass carbon (g $C_{\text{Mic}}$ m$^{-2}$). Bivariate regressions of annual and seasonal climatic variables vs. microbial biomass.

<table>
<thead>
<tr>
<th>Annual variables</th>
<th>$R^2$</th>
<th>$\beta$ (±SE)</th>
<th>$F$-stat</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual precipitation$^\dagger$</td>
<td>0.21</td>
<td>0.58 (±6.0×10$^{-2}$)</td>
<td>106.7</td>
<td>***</td>
</tr>
<tr>
<td>Maximum monthly frequency of wet days</td>
<td>0.19</td>
<td>0.04 (±3.9×10$^{-3}$)</td>
<td>99.6</td>
<td>***</td>
</tr>
<tr>
<td>MAP/ETo$^\dagger$ ratio</td>
<td>0.21</td>
<td>0.51 (±5.0×10$^{-2}$)</td>
<td>110.1</td>
<td>***</td>
</tr>
<tr>
<td>Mean annual temperature</td>
<td>0.01</td>
<td>3.4×10$^{-3}$ (±2.3×10$^{-3}$)</td>
<td>2.2</td>
<td>n.s.</td>
</tr>
<tr>
<td>MAP/ETo$^\dagger$ × MAT</td>
<td>0.32</td>
<td>0.04 (±3.0×10$^{-3}$)</td>
<td>193.4</td>
<td>***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Seasonality variables</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean precipitation of the wettest quarter$^\dagger$</td>
<td>0.17</td>
<td>0.5 (±5.4×10$^{-2}$)</td>
<td>84.1</td>
<td>***</td>
</tr>
<tr>
<td>Mean precipitation of the driest quarter</td>
<td>6.6×10$^{-2}$</td>
<td>7.6×10$^{-3}$ (±1.4×10$^{-3}$)</td>
<td>29.0</td>
<td>***</td>
</tr>
<tr>
<td>Mean precipitation of the warmest quarter</td>
<td>8.6×10$^{-2}$</td>
<td>1.0×10$^{-3}$ (±1.6×10$^{-4}$)</td>
<td>38.6</td>
<td>***</td>
</tr>
<tr>
<td>Mean precipitation of the coldest quarter</td>
<td>4.4×10$^{-2}$</td>
<td>1.3×10$^{-3}$ (±8.8×10$^{-4}$)</td>
<td>19.1</td>
<td>***</td>
</tr>
<tr>
<td>Precipitation seasonality (Coeff. of variation)$^\dagger$</td>
<td>1.0×10$^{-2}$</td>
<td>-0.15 (±6.0×10$^{-2}$)</td>
<td>5.5</td>
<td>*</td>
</tr>
<tr>
<td>Mean temperature of the wettest quarter</td>
<td>1.0×10$^{-3}$</td>
<td>-2.0×10$^{-3}$ (±2.5×10$^{-3}$)</td>
<td>0.4</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mean temperature of the driest quarter</td>
<td>5.1×10$^{-6}$</td>
<td>7.5×10$^{-5}$ (±7.5×10$^{-5}$)</td>
<td>0.0</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mean temperature of the warmest quarter</td>
<td>4.0×10$^{-3}$</td>
<td>-4.2×10$^{-3}$ (±3.2×10$^{-3}$)</td>
<td>1.7</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mean temperature of the coldest quarter</td>
<td>1.7×10$^{-2}$</td>
<td>4.5×10$^{-3}$ (±1.7×10$^{-3}$)</td>
<td>7.1</td>
<td>**</td>
</tr>
<tr>
<td>Temperature seasonality (Coeff. of variation)</td>
<td>1.0×10$^{-2}$</td>
<td>-1.9×10$^{-3}$ (±1.0×10$^{-3}$)</td>
<td>3.4</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Note: response variable (soil microbial biomass) was log$_{10}$ transformed to approach normal data distribution. Results for the topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratio are presented in Appendix S2.

$^\dagger$Variable is log$_{10}$-transformed to improve data distribution and overall linear fit of bivariate regressions.

$\beta$: standardized regression coefficient; SE: Standard error; MAP: mean annual precipitation; ETo: annual evaporative demand; MAT: mean annual temperature. Significance codes: $P<0.001$ (**); $<0.001$ (**); $<0.05$ (*); $>0.05$ (n.s.)
Fig. B1.1 Secondary multivariate models that predict patterns of $C_{\text{Mic}}$ and $C_{\text{Mic}}/C_{\text{Org}}$. (A) Observed and predicted soil profile $C_{\text{Mic}}$ estimates from Model-2: slope = 1.0 (±5.9×10^{-2}, β±1SEM), intercept: -1.6×10^{-8} (±1.2×10^{-9}), $R^2$ = 43.9%, $F_{1,365}$=285; $P<0.0001$. (B) Observed and predicted topsoil $C_{\text{Mic}}/C_{\text{Org}}$ ratios from Model-3: Slope =1.0(±5.2×10^{-2}; β±1SEM), intercept: -1.8×10^{-5}, $R^2$ = 49.9%, $F_{1,365}$=363; $P<0.0001$. Shadowed area is the 95% C.I. and all axes are in log_{10} scale. Panels (C) and (D) are the relative importance of regressors included in Model 2 and 3, respectively. In both, panels equal lower-case letter indicate that confidence intervals for the difference between the contributions of each predictor includes zero ($P<0.05$). Also, metrics are normalized to sum 100% of explained variance and 95% C.I. were obtained after 1,000 bootstrap runs. Legend and biome classification as follows: boreal forest (⬛, BF), desert (○, D), tundra (Δ, T), temperate coniferous forest (+, TCF), temperate deciduous forest (×, TDF), tropical forest (◊, TF), and temperate grassland (◇, TG). MAP: mean annual precipitation; ETo: annual evaporative demand; MAT: mean annual temperature; MaxWD: maximum monthly frequencies of wet days; Total N: soil total nitrogen content; CEC: soil cation exchange capacity; pH: soil pH; C-N ratio: soil Carbon-Nitrogen ratio.
Relationship between topsoil $\text{C}_{\text{Mic}}/\text{C}_{\text{Org}}$ ratio and total nitrogen content

The overall effect of soil nitrogen on topsoil $\text{C}_{\text{Mic}}/\text{C}_{\text{Org}}$ ratio was negative ($r = -0.37$, $P<0.0001$). To explore possible mechanisms that explained this relationship we performed linear regression analyses for the $\text{C}_{\text{Mic}}/\text{C}_{\text{Org}}$ ratios and soil total nitrogen for each biome category separately. As exemplified in figure B2, the relationship between total nitrogen and the topsoil $\text{C}_{\text{Mic}}/\text{C}_{\text{Org}}$ ratio is indeed negative in all biome categories except for desert. Possible mechanisms to explain this relationship are discussed in the main text.

Fig. B1.2. Relationship between topsoil total nitrogen and $\text{C}_{\text{Mic}}/\text{C}_{\text{Org}}$. Biome category is presented on top of each panel, label is as follows: (BF) boreal forest, (D) desert, (T) tundra, (TCF) temperate coniferous forest, (TDF) temperate deciduous forest, (TF) tropical forest and (TG) temperate grassland. Regression coefficients and standard error ($\beta\pm\text{SEM}$) from the corresponding bivariate regressions are as follows: BF: $-0.79(\pm0.23)$, $P<0.001$; D: $+1.22(\pm0.44)$, $P<0.001$; T: $-0.58(\pm0.13)$, $P<0.0001$; TCF: $-0.40(\pm0.41)$, $P>0.05$; TDF: $-1.14(\pm0.13)$, $P<0.0001$; TF: $-0.38(\pm0.13)$, $P<0.001$; TG: $-0.76(\pm0.14)$, $P<0.0001$. In all graphs, grey lines represent the bivariate regression line and the shaded-area the 95% C.I.
References


Appendix B2

Scatterplots with primary and secondary determinants
The graphs show the relationship between soil profile $C_{\text{EC}}$, log$_{10}$ and precipitation in the wettest, warmest, coldest, and driest quarters, respectively. Correlation coefficients and significance levels are as follows:

- Wettest Quarter: $r = 0.41$, $P < 0.0001$, $n = 414$
- Warmest Quarter: $r = 0.20$, $P < 0.0001$, $n = 414$
- Coldest Quarter: $r = 0.21$, $P < 0.0001$, $n = 414$
- Driest Quarter: $r = 0.26$, $P < 0.0001$, $n = 414$