Mapping local and global variability in plant trait distributions

Citation for published version:

Digital Object Identifier (DOI):
10.1073/pnas.1708984114

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the National Academy of Sciences (PNAS)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Oct. 2020
Mapping local and global variability in plant trait distributions


1 Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; 2 Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205; 3 Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108; 4 Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455; 5 ARC Centre of Excellence for Plant Energy, Research School of Biology, The Australian National University, Building 134, Canberra, ACT 2601, Australia; 6 Division of Plant Sciences, Research School of Biology, the Australian National University, Building 134, Canberra, ACT 2601, Australia; 7 Max Planck Institute for Biogeochemistry, Hans Knoll Strasse 10, 77745 Jena, Germany; 8 German Centre for Integrative Biodiversity Research (Div) Halle-Jena-Leipzig. Deutscher Platz 5e, 04103 Leipzig, Germany; 9 UMR 1137 Ecologie et Ecophysiology Forestières, Université de Lorraine – INRA, 54506 Vandoeuvre-lès-Nancy, France; 10 Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3BL, United Kingdom; 11 Joint Global Change Research Institute, DOE Pacific Northwest National Laboratory, College Park, MD USA; 12 Department of Geography and Geology, School of Natural and Built Environments, Kingston University London, Penrhyn Road, Surrey, KT1 2EE; 13 School of Biological Sciences, Seoul National University, Seoul 08826, South Korea; 14 School of Biosciences & Veterinary Medicine, Plant Diversity and Ecosystems Management unit, University of Cambridge, England; 15 Department of Theoretical and Applied Sciences, University of Insbruck, Via J.D. Unzu 3, I-12100 Varese, Italy; 16 Systems Ecology Department, Ecological University, Vrijeschool, De Boelelaan 1058, 1081 HV Amsterdam, The Netherlands; 17 Jonath Ventures, Manhattan KS 66502; 18 Department of Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Thedor-Liebscher Straße 4, 06120, Halle (Saale), Germany; 19 School of Earth and Environmental Sciences, The University of Manchester, Michael Smith Building. Oxford Road, Manchester M13 9PT, United Kingdom; 20 Instituto Multidisciplinary of Biologia Vegetal (IMBIV-CONICET) and Departamento de Diversidad Biológica y Ecología, FCEFvN, Universidad Nacional de Córdoba, CC 495, Córdoba, Argentina; 21 Facultad de Filosofía Ciencias e Letras de la Universidad de Buenos Aires; 22 USF, School of the Environment, University of South Florida, 4801 Bruce B. Downs Boulevard, Tampa, FL 33620, USA; 23 Key Laboratory of Biogeography and Bio-resource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 24 Key Laboratory of Plant Physiology, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 25 Key Laboratory of Biogeography and Bio-resource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 26 Key Laboratory of Plant Physiology, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 27 Ecological School, Seoul National University, Seoul 08826, South Korea; 28 INRA, USC 1339 Chizé (CEBC), F-79360, Villiers en Bois, France; 29 IU Carlsberg Research Institute, Copenhagen 2100, Denmark; 30 Key Laboratory of Plant Physiology, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 31 Key Laboratory of Plant Physiology, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 32 University of Arizona, Tucson, AZ 85721; 33 University of California, Los Angeles, CA 90095; 34 Institute of Biology and Environmental Science, University of Oldenburg, Carl von Ossietzky-Straße 9-11, 26111, Oldenburg, Germany; 35 Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, USA; 36 College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China; 37 Key Laboratory of Biogeography and Bio-resource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 38 School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, WITS 2050, Johannesburg, South Africa; 39 Senckenberg Biodiversity and Climate Research Centre (BfK), Senckenberganlage 25, 60325 Frankfurt, Germany; 40 Department of Physical Geography at Goethe-University, Frankfurt/Main; 41 Ume University, Institute of Systems Ecology and Evolutionary Biology, Albert-Einstein-Allee 11, 90801 Umea, Sweden; 42 Environmental Research (Alterra); 43 Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, 90095; 44 Forest and Forest Products Research Institute 1 Matsuno-naka, Tsukuba 305-0867, Japan; 45 Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, USA; 46 School of Geosciences, University of Edinburgh, Edinburgh, EH9 3FF, UK; 47 Joint Global Change Research Institute, DOE Pacific Northwest National Laboratory, College Park, MD USA; 48 Institute of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, 90095; 49 Forest and Forest Products Research Institute 1 Matsuno-naka, Tsukuba 305-0867, Japan; 50 Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, USA; 51 Key Laboratory of Plant Physiology, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 52 University of Arizona, Tucson, AZ 85721; 53 University of California, Los Angeles, CA 90095; 54 Institute of Biology and Environmental Science, University of Oldenburg, Carl von Ossietzky-Straße 9-11, 26111, Oldenburg, Germany; 55 Department of Plant Physiology, Estonian University of Life Sciences, Kreutzwald 1, 51014 Tartu, Estonia; 56 Graduate School of Agriculture, Kyuto University, Kyoto, 606-8502 Japan; 57 CSIC, Unidad d'Ecologia Global CREAF-CSIC-UAB, Bellaterra 08193, Barcelona, Catalonia, Spain; 58 CREAF, Cerdanyola del Vallés 08193, Barcelona, Catalonia, Spain; 59 Michigan State University, Department of Forestry, 480 Wilson Rd., East Lansing, MI 48824; 60 Museo Nacional de Ciencias Naturales, CSIC Serrano 115 edificio, E-28006 Madrid Spain; 61 Dept. of Biology, Algoma University, Sault Ste. Marie, Ontario, Canada P6A 2C4; 62 Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands; 63 Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, Riverside, CA, 92521; 64 Embraço Clima Temperado, Rodovia BR 392, Km 78, Pelotas, RS, Brasil, 96010-971; 65 Environmental Sciences Division, Climate Change Science Institute, Oak Ridge National Laboratory. Oak Ridge, TN, USA; 66 Department Systematic Botany and Functional Biodiversity, University of Leipzig; 67 04103 Leipzig, Germany; 68 Hawkesbury Institute for the Environment, Western Sydney University, Penrith New South Wales 2751, Australia.

Our ability to understand and predict the response of ecosystems to a changing environment depends on quantifying vegetation functional diversity. However, representing this diversity at the global scale is challenging. Typically, in Earth Systems Models, characterization of plant diversity has been limited to grouping related species into Plant Functional Types (PFTs), with all trait variation in a PFT collapsed into a single mean value that is applied globally. Using the largest global plant trait database and state of the art Bayesian modeling, we created fine-grained global maps of plant trait distributions that can be applied to Earth System Models. Focusing on a set of plant traits closely coupled to photosynthesis and foliar respiration – specific leaf area (SLA), and dry mass-based concentrations of leaf nitrogen (Nm) and phosphorus (Pm), we characterize how traits vary within and among over 50,000 ~ 50 × 50 km cells across the entire vegetated land surface. We do this in several ways - without defining the PFT of each grid cell, and using 4 or 14 PFTs; each model's predictions are evaluated against out-of-sample data. This endeavor advances prior trait mapping by generating global maps that preserve variability across scales by using modern Bayesian spatial statistical modeling in combination with a database over three times larger than previous analyses (Van Bodegom, et al. (2014) PNAS 111(38):13733-8; Maire, et al. (2015), Global Ecol. Biogeogr. 24(6):706-17). Our maps reveal that the most diverse grid cells possess trait variability close to the range of global PFT means.

Modeling global climate and the carbon cycle with Earth System Models (ESMs) requires maps of plant traits that play key roles in leaf- and ecosystem-level metabolic processes (1–4). Multiple traits are critical to both photosynthesis and respiration, foremost leaf nitrogen concentration (Nm) and specific leaf area (SLA) (5–7). More recently, variation in leaf phosphorus concentration (Pm) has also been linked to variation in photosynthesis and foliar respiration (7–12). Estimating detailed global geographic patterns of these traits and corresponding trait-environment relationships has been hampered by limited measurements (13), but recent improvements in data coverage (14) allows for greater detail in spatial estimates of these key traits.

Previous work has extrapolated trait measurements across continental or larger regions through three methodologies: 1) grouping measurements of individuals into larger categories that share a set of properties (a working definition of plant functional types or PFTs) (4, 15); 2) exploiting trait-environment relationships (e.g. leaf Nm and mean annual temperature) (1, 16–20), or 3) restricting the analysis to species whose...
Fig. 1. Trait data a) Global locations and values of specific leaf area measurements for the PFT Tropical Broadleaf Evergreen Trees. b) Locations and values of specific leaf area measurements for the Tropical Broadleaf Evergreen Trees in Panama. The central square indicates a 0.5° × 0.5° pixel containing the Barro Colorado Island sites (see Fig. 5). These points have been jittered up to 0.05° to highlight the density of measurements. c) The full distribution of specific leaf area values for all species classified as tropical Evergreen Broadleaf Tropical Trees. The blue line is the global data while black is the local pixel, the dashed vertical lines are the respective means.

**Significance Statement**

Currently, Earth System Models (ESMs) represent variation in plant life through the presence of a small set of Plant Functional Types (PFTs), each of which accounts for hundreds or thousands of species across thousands of vegetated grid cells on land. By expanding plant traits from a single mean value per PFT to a full distribution per PFT that varies among grid cells, the trait variation present in nature is restored and may be propagated to estimates of ecosystem processes. Indeed, critical ecosystem processes tend to depend on the full trait distribution, which therefore needs to be represented accurately. These maps re-introduce substantial local variation and will allow for a more accurate representation of the land surface in ESMs.

The general idea for the study was developed by E.E.B., A.D., H.F.M., F.F., M.C., K.W., A.B., J.K., O.K.A., and P.B.R.; specifics were developed by E.E.B. and A.D. and refined with the rest of the team. Data were made available by the hundreds of contributors to the TRY database; with further data management and compilation by E.E.B., A.D., H.F.M. and J.K. E.E.B. and A.D. performed the analysis, with all authors contributing to interpretation. E.E.B. and A.D. wrote the first draft; all authors contributed to subsequent versions, including the submitted one.

The authors declare no known conflicts of interest.

*E.E.B. and A.D. contributed equally to this work.

*To whom correspondence should be addressed. E-mail: eebutler@umn.edu, abhidatta@jhu.edu

Information is lost when the range of measured trait values is compressed into a single PFT (Fig. 1). We observe that the global range of site level SLA values for a single PFT such as Broadleaf Evergreen Tropical trees (Fig. 1a,c) is quite large (2.7 to 65.2 m² kg⁻¹). Even after limiting the scope to a single well measured 0.5° × 0.5° pixel within Panama (Fig. 1b,c), there is still a wide range of SLA values (4.7 to 37.7 m² kg⁻¹) with a local mean of 15.7 m² kg⁻¹ and a local standard deviation of 5.4 m² kg⁻¹ — over 1/3 of the local mean. By contrast, the mean SLA value of all species associated with Broadleaf Evergreen Tropical trees is 13.9 m² kg⁻¹, over 10% lower than the local average (Fig. 1c). Thus, single trait values per PFT fail to capture variability in trait values within or among grid cells; i.e. over a wide range of spatial scales.

Transitioning from a single trait value per PFT (within or among grid cells) to a distribution may lead to significantly different modeling results (20) as critical plant processes, such as photosynthesis, are non-linear with respect to these traits (28). This is reinforced by recent modeling studies which have begun to incorporate distributions of traits at regional (29, 30)
and global (31) scales. It has been shown that using trait distributions leads to different estimates of carbon dynamics (32) and that higher order moments of trait distributions contribute to sustaining multiple ecosystem functions (33). While species level mapping (21, 23, 24) does capture trait distributions, it has been limited geographically and restricted to subsets of functional groups.

Even the largest plant trait database offers only partial coverage across the globe in terms of site level measurements. Hence, gap-filling approaches need to be adopted to extrapolate trait values at regions with no data coverage. Here, we overcome data limitations through PFT classification, trait-environment relationships, and additional location information to develop a suite of models capable of estimating trait distributions across the entire vegetated globe. The simplest is a categorical model, which assigns traits to maps of remotely sensed PFTs. Every species, with its corresponding trait values, is associated with a PFT and these trait distributions are extrapolated to the satellite estimated range of the PFT (SI Appendix, Figs. S1-S2). The second is a Bayesian linear model which complements the PFT information with trait-environment relationships. The third is a Bayesian spatial model which, in addition to PFTs and the trait-environment relationships, leverages additional location information via Gaussian Processes (Methods). The use of a spatial Gaussian Process in this context is novel and model evaluation reveals the superior predictive performance of this model.

Each of these methods interpolate (and extrapolate) both mean trait values and entire trait distributions across space (i.e., across grid cells on a global map). These models are further stratified by three different levels of PFT categorization: 1) PFT-free, all plants in a single group (i.e., no PFTs); 2) broad, four groups based on growth form and leaf type; 3) narrow, fourteen groups based on further environmental, phenological, and photosynthetic categories (Methods). The PFT-free categorization groups all plants into a single class, while the broad grouping (4-PFT) is similar to the vegetation classification used in the JULES land surface (34), and the narrow (14-PFT) category is equivalent to the classification used in the Community Land Model (4, 15, 35).

The above mentioned methods allow for a representation of global vegetation that enables a more accurate formulation of functional diversity than the single trait value per PFT paradigm that is widely employed (4). The traits studied here - SLA, N\textsubscript{m}, and \(\frac{\text{P}}{\text{m}}\) - are central to predicting variation in rates of plant photosynthesis (5, 6, 9, 11) and foliar respiration (10, 36). The importance of these traits and the more advanced representation of functional diversity developed here may be used to better capture the response of the land surface component of the Earth System to environmental change.

**Results and Discussion**

**Model Evaluation.** Given the full suite of nine models proposed, we conducted extensive model evaluation (see Table 1) to determine the trade-offs associated with each methodology and resolution of PFT. We assessed the predictive capability of the models using the root mean squared predictive error (RMSPE) based on out-of-sample data (SI Appendix, Section S6). Among the nine models, the spatial narrow 14-PFT model emerged as the best predictor of mean trait values for SLA and \(N_m\), and the second best for \(\frac{\text{P}}{\text{m}}\) (Table 1). However, the spatial broad 4-PFT model performed nearly as well (Table 1). The models’ abilities to correctly estimate the spread of the trait distributions were assessed using the out-of-sample coverage probabilities (CP) – the proportion of instances the model predicted 95% confidence intervals contained the observed trait values. Most of the models provided adequate coverage (CP of around 90% or more). See the SI Appendix, Section S4, for more detailed definitions of the model comparison metrics.

**Table 1. Model evaluation**

<table>
<thead>
<tr>
<th></th>
<th>ps-R\textsuperscript{2}</th>
<th>RMSPE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cf</td>
<td>NA</td>
<td>8.13</td>
<td>91.2%</td>
</tr>
<tr>
<td>Cb</td>
<td>16.9%</td>
<td>7.13</td>
<td>94.7%</td>
</tr>
<tr>
<td>Cn</td>
<td>26.0%</td>
<td>6.66</td>
<td>95.8%</td>
</tr>
<tr>
<td>Lf</td>
<td>4.6%</td>
<td>7.99</td>
<td>91.3%</td>
</tr>
<tr>
<td>Lb</td>
<td>23.4%</td>
<td>6.93</td>
<td>94.0%</td>
</tr>
<tr>
<td>Ln</td>
<td>30.7%</td>
<td>6.53</td>
<td>95.2%</td>
</tr>
<tr>
<td>Sf</td>
<td>45.5%</td>
<td>7.54</td>
<td>93.6%</td>
</tr>
<tr>
<td>Sb</td>
<td>58.5%</td>
<td>6.31</td>
<td>97.7%</td>
</tr>
<tr>
<td>Sn</td>
<td>60.2%</td>
<td>6.13</td>
<td>97.7%</td>
</tr>
<tr>
<td>N\textsubscript{m}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cf</td>
<td>NA</td>
<td>7.16</td>
<td>93.3%</td>
</tr>
<tr>
<td>Cb</td>
<td>12.5%</td>
<td>6.95</td>
<td>93.2%</td>
</tr>
<tr>
<td>Cn</td>
<td>19.4%</td>
<td>6.47</td>
<td>92.7%</td>
</tr>
<tr>
<td>Lf</td>
<td>5.2%</td>
<td>7.28</td>
<td>93.2%</td>
</tr>
<tr>
<td>Lb</td>
<td>16.7%</td>
<td>6.71</td>
<td>94.3%</td>
</tr>
<tr>
<td>Ln</td>
<td>24.1%</td>
<td>6.42</td>
<td>94.6%</td>
</tr>
<tr>
<td>Sf</td>
<td>44.2%</td>
<td>7.19</td>
<td>93.6%</td>
</tr>
<tr>
<td>Sb</td>
<td>53.7%</td>
<td>6.36</td>
<td>96.1%</td>
</tr>
<tr>
<td>Sn</td>
<td>54.8%</td>
<td>6.18</td>
<td>96.1%</td>
</tr>
<tr>
<td>(\frac{\text{P}}{\text{m}})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cf</td>
<td>NA</td>
<td>0.86</td>
<td>90.5%</td>
</tr>
<tr>
<td>Cb</td>
<td>5.3%</td>
<td>0.86</td>
<td>90.5%</td>
</tr>
<tr>
<td>Cn</td>
<td>28.1%</td>
<td>0.78</td>
<td>91.1%</td>
</tr>
<tr>
<td>Lf</td>
<td>25.6%</td>
<td>0.84</td>
<td>87.2%</td>
</tr>
<tr>
<td>Lb</td>
<td>32.8%</td>
<td>0.85</td>
<td>85.3%</td>
</tr>
<tr>
<td>Ln</td>
<td>35.4%</td>
<td>0.82</td>
<td>87.0%</td>
</tr>
<tr>
<td>Sf</td>
<td>62.0%</td>
<td>0.83</td>
<td>90.7%</td>
</tr>
<tr>
<td>Sb</td>
<td>66.7%</td>
<td>0.81</td>
<td>92.0%</td>
</tr>
<tr>
<td>Sn</td>
<td>67.6%</td>
<td>0.80</td>
<td>91.3%</td>
</tr>
</tbody>
</table>

The pseudo-R\textsuperscript{2} (ps-R\textsuperscript{2}), RMSPE and CP statistics for all nine models, for each of the three traits. The bold entries correspond to the model producing highest ps-R\textsuperscript{2}, lowest RMSPE, or CP closest to 0.95. The categorical PFT-free model (Cf) produces a constant estimate and hence ps-R\textsuperscript{2} is not defined. Each model is indicated by a two-letter abbreviation: C=Categorical (no regression), L=Linear (linear regression), S=Spatial (linear regression with spatial term) and the accompanying PFT resolution: f=PFT-free (no PFT information), b=broad (4-PFT), n=narrow (14-PFT).

The improvement in prediction afforded by the inclusion of (1) a spatial term and (2) PFT information (Table 1) invites further examination. First, the spatial term in our model likely incorporates some of the finer scale variation that is unavailable given the relatively large grid cell size of the environmental covariates used in global studies. Thus, the spatial term allows for adjustment of trait values among neighboring or regional...
grid cells that the relatively coarse environmental metrics are not able to capture. Finer scale studies that can evaluate local variations in climate, soil, or other relevant abiotic or biotic covariates may see less improvement from the inclusion of a spatial term, as they may directly measure local sources of variation. Second, the use of PFTs greatly improves the models, perhaps for similar reasons involving the degree of variation the raw data fail to incorporate. The greatest decrease in RMSPE occurs between the PFT-free grouping (a single category for all plants) and the broad (4-PFT) grouping across each of the models tested. If our trait data were perfectly predicted by environment, there would be no usefulness to including PFTs in mapping traits. That this not is so implies that the broad PFTs, based primarily on growth form and leaf type, offer superior predictive skill than environmental covariates on their own [19]. However, the extra information in the narrow (14-PFT) grouping does further improve the fit and produces the most accurate predicted trait surface.

Global Maps. We selected two sets of maps to describe, in broad strokes, how trait distributions vary across the land surface: the narrow 14-PFT spatial model and its categorical counterpart. The narrow 14-PFT spatial model is the best predictor of mean trait values, and provided adequate coverage probability (Figs. 2-4a,b). For comparison, we also include the 14-PFT categorical model, which is most similar to maps currently used in ESMs (Figs. 2-4c,d). Maps for the other models can be found in the supplemental material (SI Appendix, Figs. S8-S16). The mean and standard deviation are presented as a summary of the full log-normal distribution within each pixel, but there are full distributions estimated in each pixel, see Case Studies below.

The standard deviation maps (Figs. 2-4b,d) compared to the mean maps (Figs. 2-4a,c) highlight one of the central results of this analysis – the local standard deviations of trait values are of similar magnitudes as their respective means. Generally, we observed that the local standard deviation is close to half the local mean value, but can approach the global range of the trait mean values, e.g. Nm (Fig. 3) has a maximum local standard deviation of 9 mg N / g, and the global mean range is only ≈10 mg N / g. The maps of the trait standard deviations follow similar patterns to the means, though there are several regions where the mean varies more markedly than the standard deviation; such as SLA in the SE United States and China in the categorical model (Fig. 2c,d) and similarly for Nm in the spatial model across the Sahel in sub-Saharan Africa (3a,c). The lack of variation in the standard

**Fig. 2.** Specific Leaf Area maps. a,b) Narrow (14-PFT) Bayesian spatial model pixel mean and standard deviation estimates, respectively; c,d) Narrow (14-PFT) Categorical model pixel mean estimates and standard deviation estimates, respectively. For clarity, the color bars have been truncated at the compound 5th and 95th percentiles of both models. Latitude tick marks indicate the equator, tropics, and arctic circle and longitude is marked at 100° W, 0°, and 100° E.
deviation is most clear in the categorical model for $N_m$ while both models show relatively modest variation in $P_m$.

For each of the three traits, the broad features of both the categorical and spatial models are similar, but there are numerous marked differences across regional and fine spatial scales (Figs. 2-4). The shared broad features of the maps from both models include SLA (Fig. 2) and $P_m$ (Fig. 4) increasing from the tropics to the poles, while $N_m$ (Fig. 3) has more modest variation, except that it tends to be lower in regions dominated by needle-leaved trees. Some of the notable differences between the models include the spatial model’s greater range and more marked variability of SLA within equatorial regimes (e.g., Brazil or central Africa); it also better captures the low SLA of most of arid Australia than the categorical model (Fig. 2a); and more strongly highlights the gradient of $P_m$ from the tropics to the arctic (16) (Fig. 4a).

The most consistent estimates between the categorical and spatial models are in the boreal regions dominated by needle-leaved trees; the measurements in this region are relatively sparse which may have limited the ability of the spatial model to capture differences. On the other hand, broad-leaved trees span a wide range of environments, but a large portion of the measurements come from the tropics (66%), where there is a limited range of values among the climate covariates and therefore little variation with which to estimate a correlation. The grasses and shrubs have the largest standard deviations of the four broad PFTs (SI Appendix, Table S4) and dominate wide swaths of the land surface, but have fewer measurements – shrubs are the least measured of the broad PFTs in the database, and this appears to reduce the accuracy of the categorical model more than the spatial model (Table 1). The fact that shrubs are assumed to dominate in arid and boreal environments, which also tend to be under-sampled, also likely contributes to these differences.

Our results also suggest that the breadth of functional niche space is reduced in both boreal and tropical biogeographic regions. The low variation across all three traits within the boreal forest implies that there is strong filtering and smaller niche space available in this relatively harsh environment. Surprisingly, despite the high species diversity in tropical forests, we also find that SLA and $P_m$ have relatively low variation in these forests – suggesting that in this environment the trait space is reduced. This could be, in part, an artifact of the Earth System Model PFT classification omitting herbaceous species. Conversely, grasslands and savannahs exhibit large
Fig. 4. Phosphorus [mass] maps. a,b) Narrow (14-PFT) Bayesian spatial model pixel mean and standard deviation estimates, respectively. c,d) Narrow (14-PFT) Categorical model pixel mean estimates and standard deviation estimates, respectively. For clarity, the color bars have been truncated at the compound 5th and 95th percentiles of both models. Latitude tick marks indicate the equator, tropics, and arctic circle and longitude is marked at 100°W, 0°, and 100°E. 

Case Studies. We conducted two regional case studies to provide a more in-depth analysis of the true and predicted shapes of trait distributions than can be provided by the standard deviation maps and coverage probability. In these case studies trait data were pooled over an area to construct full trait distributions and then formally compared with the model predicted distributions.

We considered two areas with substantially different environmental conditions to evaluate the trait distributions obtained from the spatial and categorical models. We chose a single pixel that contained a highly studied site with numerous measurements of tropical trees, Barro Colorado Island (BCI), Panama; and a collection of pixels in an arid environment in which the mean estimates for SLA of the spatial and categorical models substantially disagreed, the southwestern United States. These areas were in the training data, and this analysis constituted a more detailed analysis of the models’ fit to the observed distribution of these locations. Here, the focus was on the structure of the full distribution of traits predicted at these sites; Fig. S17 is a map of the measurements that comprised these locations and other sites included in this analysis. Both areas offer further insight into the structure of the distributions estimated by the categorical and spatial models.

In the pixel containing BCI, the categorical and spatial models broadly agreed for all three traits (Fig. 5a, c, e), although the spatial model means were only half as distant from the observed means for SLA and $N_m$ (4% vs. 8% and 5% vs. 10%, respectively). There were only two PFTs present in this pixel: tropical broadleaf evergreen and deciduous trees. Despite the general similarity of the shapes of the distribution, the spatial model appears capable of capturing some subtle features. This is clearest for leaf nitrogen, where the peak of the distribution was quite broad. This is neatly captured in the narrow PFT model, and the pattern was detectable through the Kolmogorov-Smirnov (K-S) statistic, which evaluates the similarity of two full distributions. Indeed, the superiority of the spatial model was reinforced by a closer match for the Bayesian spatial model across all traits at BCI, though for $P_m$...
Fig. 5. Empirical trait distributions Barro Colorado Island on the left (a, c, e) and the US Southwest on the right (b, d, f). The first row is SLA (a, b), the second is leaf nitrogen (c, d) and the third is leaf phosphorus (e, f). Each panel depicts the distribution of the data in solid black, the categorical model in blue and the Bayesian spatial model in red. The vertical lines indicate mean values.

Environmental Covariates and the Spatial Term. The improvement in prediction from the linear model to the spatial model is partially explained by weak trait-environment relationships (SI Appendix, Tables S1-S3). The magnitude of spatial variation explained by the Gaussian process model is comparable to the unexplained trait variation. For most of the spatial models, the estimated spatial range was around 500 kilometers; this suggests a strong spatial effect, and implies that the spatial model can provide more precise information about the trait distribution near the locations where we have data. This was largely borne out in the case studies, and is illustrated more explicitly in Fig. 6 where the predicted trait standard deviation for the spatial model was up to 50% lower than the linear non-spatial model near locations with trait measurements. The spatial model leverages local information to reduce the uncertainty of trait estimation near data locations and may provide guidance for future data collection by identifying high uncertainty regions.

Applications for Trait Distributions. Plant traits vary across a range of spatial scales, and the spatial model best captures changes across large spatial gradients (such as in Amazonia and Australia) as well as the subtleties within pixels. Maps for all the models highlight how much information about local variability is lost when representing plant traits with a single value, and suggests that a first application of these maps will be for ESMs to incorporate these scales of variability. For process-based ESMs, the simplest model to incorporate will likely be the categorical model as it is closest to the current PFT approach, but this model is also the least flexible. The more sophisticated models developed here provide more accurate large scale variation, and may be used to infer new trait values in a novel climate by perturbing the climatic covariates (37). However, given the likelihood of non-linear trait-environment relationships, the spatial sparsity of the data, and the possibility of alternate strategies within a PFT that may alter the trait-environment relationship in a future climate some caution is called for when using these models for extrapolation. Future ecosystem models could also integrate the leaf level variation in these maps with canopy scale changes in leaf display traits - leaf angle, azimuth, and total area.

We have emphasized the quality of the Bayesian spatial model with narrow PFTs, but there is an intriguing possibility opened by the PFT-free model (SI Appendix, Fig. S8, S11, and S14) – that being the representation of vegetation without reference to PFTs (1). In this case the representation of vegetation would rely entirely on the structure of trait distributions at various landscape scales (1). Such a
representation eliminates the need to separately model the future locations of PFTs (or species) when inferring the future distribution of traits; hence, the output of a model like that developed here could be updated with future environmental covariates, with the caveats that ‘out of sample prediction’ may entail. At the same time, this method would allow for greater functional diversity than multiple PFTs with single trait values, as is currently used in most ESMs. Adopting this approach does, however, raise the issue of how to deal with the paucity of surface observations in some regions, as evidenced by the greater errors associated with estimating out of sample values with this model (Table 1). Complementary work has retrieved leaf trait maps from a global carbon cycle model fused with Earth observations (38), providing another method that could be used for direct comparison against the trait maps produced here. While the methodology outlined in our analysis brings the possibility of a PFT-free land surface closer, we remain several steps away from being able to make such maps as accurately as we do using PFT characterizations for trait prediction. Several actions can bring us closer to that goal. First, incorporation of additional information (such as phylogenetic relatedness and trait-trait covariance) will likely improve trait maps, even using existing observations. Second, as the current level of observations is extremely sparse in some regions, and sparse in most, expanded trait databases will also aid in development of PFT-free trait maps.

Conclusions

SLA and Nm are essential inputs into the land surface components of Earth System Models, and while phosphorus has not yet been as widely incorporated into ESMs, it has been shown - particularly across the tropics - to be important to photosynthesis (9, 11, 39–42) and respiration (11, 12, 36). The maps and trait-environment relationships presented here may be used by existing land surface models that use similar categories to classify vegetation. However, it should be noted that PFT-dependent models often have many other parameters that have been calibrated to historical estimates of particular trait values (4). Thus, the values developed here, while likely drawing from a larger pool of measurements than has been done previously can not necessarily be adopted without further modification of other model elements (37, 43). Nonetheless, these results can be incorporated into a wide class of models with relative ease. We can now provide global trait distributions at the pixel scale.

The global land surface is perhaps the most heterogeneous component of the Earth System. Reducing vegetation to a collection of PFTs with fixed trait values has been the preferred method to constrain this heterogeneity and group similar biochemical and biophysical properties; however, this has been at the expense of functional diversity. This analysis quantifies the substantial magnitude of this ignored trait variation. The approach and methods presented here retain the simplicity of the PFT representation, but capture a wider range of functional diversity.

Materials and Methods

Data. The TRY database (www.try-db.org) (14) provided all data for leaf traits and the categorical traits to aggregate PFTs (TRY – Categorical Traits Dataset, https://www.try-db.org/TryWeb/Data.php#3, January 2016) used in the analysis. See SI Appendix (Appendix 1) for a complete list of the original publications associated with this subset of TRY. The extract from TRY used here has just under 45,000 measurements of individuals from 3,680 species with measurements of at least one of specific leaf area (SLA), leaf nitrogen per dry leaf mass (Nm), and/or leaf phosphorus per leaf dry mass (Pm). The number of individual measurements varies from 32,315 for SLA on 2,953 species to 19,282 for Nm on 3,053 species down to 8,052 for Pm on 1,810 species; see Table S4 for the number of unique measurements and species found in any categories used in the analysis. The species taxonomy was standardized using The Plant List (44). Measurements were associated with environmental categories through Köppen-Geiger climate zones (45). All environmental variables are on a 0.5° × 0.5° grid. Climate variables use 30 year climatologies from 1961-1990 as estimated by the Climate Research Unit (46, 47). Soil variables are from the International Soil Reference and Information Centre - World Inventory of Soil Emission Potentials (ISRIC-WISE) (48).

The spatial extent of PFTs have been previously estimated through satellite estimates of land cover around the year 2005 (49), and these estimates have been refined in climatic categories (15, 35). While TRY, and thus the data used here, represents the largest collection of plant traits in the world most of the observations come from a subset of global regions: North America, Europe, Australia, China, Japan, and Brazil. There are still large sections of the planet with extremely sparse measurements, notably: much of the tropics outside of the Americas, large swathes of Central Asia, the Russian Federation, South Asia and much of the Arctic (SI Appendix, fig. S17). Improving data collection in these regions will greatly improve future modeling efforts. Improving data collection in these regions will greatly improve future modeling efforts. Until observations are more complete there remains the possibility of spurious patterns, though we have found little evidence to suggest their presence in this analysis, even in comparison to detailed regional studies (SI Appendix, fig. S26) (50).

Classification of PFTs and Categorical Model. We used three nested levels of PFT classification. In the first level, all plants are categorized into a single group (‘PFT-free’). In the second level (‘broad’), all plants are categorized into PFTs based on categorical traits associated with growth form (grass, shrub, tree) and leaf type (broad and needle-leaved) leading to the following four PFTs: grasses, shrubs, broad-leaved trees and needle-leaved trees (Fig. S1). In the third level (‘narrow’), the broad PFTs are further refined by their climatic region – tropical, temperate, boreal – as well as leaf phenology, and, for the grasses, photosynthetic pathway (C3 or C4). This produces 14 PFTs (Fig. S2), which correspond exactly to those in the Community Land Model (CLM). Note that these PFT classifications exclude non-woody eudicots (‘herbs’), which were excluded from the analysis, on account of their lack of dominance within these PFT categories (51) and therefore, on account of being widely measured could overly influence the structure of the trait distributions if they were included. Satellite estimates of the PFT abundance that correspond to the “narrow” PFT categories defined above have already been calculated (15, 49) and we used these to assign a percentage of each 0.5° × 0.5° pixel to each PFT present according to the fraction of the land surface within that pixel occupied by the PFT. The “broad” PFT fractions are calculated by summing the narrow PFT categories within each “broad” classification.

The categorical model uses the PFT categories and averages trait values for each species across individual measurements at each measured location. This defines the PFT as the interspecies range of trait values and ignores all local environmental factors. The results of the categorical model are summarized by the mean and standard deviation of each PFTs trait values (Table S4) for all three resolutions of the use where no PFT information is used, the categorical model produces a constant trait distribution across the entire vegetated world. The categorical model, and the Bayesian models described in the following section all use location specific species mean values to estimate trait distributions. We assume no intra-specific variation in trait values.
Bayesian Models. A more fine-tuned depiction of geographical or spatial variation of plant trait values within each PFT can be achieved by leveraging environmental and location information, which allows trait values to adjust based on local conditions. Data for 17 climate (46, 47) and soil based (48) environmental predictors were available at the 0.5° × 0.5° pixel resolution used to create the trait maps. To avoid overfitting and collinearity issues, these seventeen predictors were screened (see Section S7) based on correlations amongst predictors, their individual correlation with the traits, and to include climate covariates along different axes of environmental stress and both chemical and physical soil covariates. We finally selected five predictors – mean annual temperature [MAT], total annual radiation [RAD], moisture index (precipitation/evapotranspiration) [MI], percent nitrogen (aquaous) [pNH4], and percent clay content [CLY]. Remote sensing data products, such as Normalized Difference Vegetation Index (53)), are not used as covariates, to allow for inference outside of the historical observation period through perturbations of environmental covariates.

We utilized environment-trait relationships to obtain predictions of trait values, (1, 16–18, 37, 49) in a linear regression setup. The formal details of the initial model are as follows. We denote log-transformed trait values at a geographical location s as $y_{trait}(s)$. This set of five predictors at a location s is denoted by the vector $x(s) = (x_1(s), x_2(s), ..., x_5(s))$. A linear regression model relating the trait to the environmental predictors is specified as:

$$y_{trait}(s) = b_0 + b_1 x_1(s) + b_2 x_2(s) + ... + b_5 x_5(s) + \epsilon(s)$$

where $b_i$ are the regression coefficients and $\epsilon(s)$ is the error term explaining residual variation. Estimation of model parameters and prediction were achieved with a fully Bayesian hierarchical model. This enables inclusion of prior information and prediction of full trait distributions instead of representative values (like median or mean) thereby ensuring that the uncertainty associated with the estimation of model parameters is fully propagated into the predictive trait distributions.

We then generalized the above model into a Bayesian spatial linear regression model that borrows information from geographically proximal regions to capture residual spatial patterns beyond what is explained by environmental predictors. The formulation of a spatial regression model is obtained by splitting up the error term $\epsilon(s)$ in Equation (1) into the sum of a spatial process $w(s)$ and an error term $\eta(s)$, that accounts for the residual variation after adjusting for the spatial effects $w(s)$. The underlying latent process $w(s)$ accounts for local nuances beyond what is captured by the environmental predictors and is often interpreted as the net contribution from unobserved or unusable predictors. Gaussian Processes (GP) are widely used for modeling unknown spatial surfaces such as $w(s)$, due to their convenient formulation as a multivariate Gaussian prior for the spatial random effect, uncorrelated predictive performance (54) and ease of generating uncertainty quantified predictions at unobserved locations. We use the computationally effective Nearest Neighbor Gaussian Process (27) which nicely embeds into the Bayesian hierarchical setup as a prior for $w(s)$ in the second stage of the model specification. All technical specifications of the Bayesian spatial model are provided in Section S1 of the supplementary materials.

The linear regression models used in previous studies (1, 16–18) and both the spatial and non-spatial Bayesian models described above assume a global relationship between the traits and environment. Given the goal of predicting trait values for the entire land surface, the assumption of a universal trait-environment relationship may be an oversimplification (55). Moreover, if there is significant variation in plant trait values among different PFTs, the estimated parameters will be skewed towards values from abundantly sampled PFTs, such as broad-leaved trees. Additional information about plant characteristics at a specific location, if available, can potentially be used to improve predictions. As mentioned earlier, we have PFT classifications for each observation of the dataset used here and satellite estimates of PFT abundance at all pixels. The global regression approaches described above ignore this information and therefore the best predictions at locations dominated by PFTs poorly represented in the data, such as shrubs. Hence, we also incorporate the PFT information in these regression models by allowing the trait-environment relationship to vary between different PFTs.

Finally, the PFT specific distributions from the Bayesian models were weighted by the satellite based PFT abundances to create a larger-scale set of distributions that are closer to the global trait distribution. This provides a more realistic comparison between all three categorizations of PFT. Details of the PFT based Bayesian models are provided in Section S2. The use of a Gaussian Process based spatial model as well as the Bayesian implementation of the regression models were novel to this application of plant trait mapping and, as results indicated, were critical to improving model predictions as well as properly quantifying trait distributions.

All the code and public data are available from the authors upon request. The TRY data may be requested from the TRY database custodians.

ACKNOWLEDGMENTS. E.E.B., H.F.M., M.C., K.R.W., and P.B.R. acknowledge funding from the United States Department of Energy, Office of Science (DE-SC0012677). O.K.A. acknowledges the support of the Australian Research Council (CE140100008). P.B.R also acknowledges support from two University of Minnesota Institute on the Environment Discovery Grants. The study has been supported by the TRY initiative on plant traits (http://www.try-db.org). The TRY initiative and database is hosted, developed and maintained at the Max Planck Institute for Biogeochemistry, Jena, Germany. The TRY is currently supported by DIVERSTIES/Future Earth and the German Centre for Integrative Biodiversity Research (iDiv) Halle–Jena–Leipzig. BB acknowledges a NERC independent research fellowship NE/M019160/1. JP would like to acknowledge the financial support from the European Research Council Synergy grant ERC-SyG-2013-010028 IMBALANCE-P, the Spanish Government grant CGL2013-48074-P and the Catalan Government grant SGR 2014-274. B.B.-L. was supported by the Earth System Modeling program of the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. KK acknowledges the contribution of the WUR Investment theme Resilience for the project Resilient Forest (KB-29-009-003). PM acknowledges support from ARC grant FT140100457 and NERC NE/F002149/1. WH acknowledges funding from the European Commission’s Seventh Framework Programme (FP7/2007-2013) under Grant Agreement STREP-264294. SG acknowledges support from the National Natural Science Foundation of China (#81473068) and “Light of West China” Program of Chinese Academy of Sciences. We would also like to thank the improvements suggested by two anonymous referees, which improved the clarity and depth of the manuscript.


Butler and Datta et al.
PNAS | October 18, 2017 | vol. XXX | no. XX | 9


37.ather and Datta et al.