

Mapping local and global variability in plant trait distributions

Ethan E. Butler^{1,a,b}, Abhirup Datta^{2,a,b}, Habacuc Flores-Moreno^{1,3}, Ming Chen¹, Kirk R. Wythers¹, Farideh Fazayeli⁴, Arindam Banerjee⁴, Owen K. Atkin^{5,6}, Jens Kattge^{7,8}, Bernard Amiaud⁹, Benjamin Blonder¹⁰, Gerhard Boenisch⁷, Ben Bond-Lamberty¹¹, Kerry A. Brown¹², Chae-ho Byun¹³, Giandiego Campetella¹⁴, Bruno E.L. Cerabolini¹⁵, Johannes H.C. Cornelissen¹⁶, Joseph M. Craine¹⁷, Dylan Craven^{8,18}, Franciska T. de Vries¹⁹, Sandra Díaz²⁰, Tomas Domingues²¹, Estelle Forey²², Andres Gonzalez²³, Nicolas Gross^{24,25,26}, Wenxuan Han^{27,28}, Wesley N. Hattingh²⁹, Thomas Hickler^{30,31}, Steven Jansen³², Koen Kramer³³, Nathan J.B. Kraft³⁴, Hiroko Kurokawa³⁵, Daniel C. Laughlin³⁶, Patrick Meir^{6,37}, Vanessa Minden³⁸, Ülo Niinemets²⁹, Yusuke Onoda⁴⁰, Josep Peñuelas^{41,42}, Quentin Read⁴³, Fernando Valladares Ros⁴⁴, Lawren Sack³⁴, Brandon Schamp⁴⁵, Nadejda A. Soudzilovskaia⁴⁶, Marko J. Spasojevic⁴⁷, Enio Sosinski⁴⁸, Peter Thornton⁴⁹, Peter M. van Bodegom⁴⁶, Mathew Williams³⁷, Christian Wirth^{7,8,50}, and Peter B. Reich^{1,51}

¹Department of Forest Resources, University of Minnesota, St. Paul, MN 55108; ²Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205; ³Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, 55108; ⁴Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, 55455; ⁵ARC Centre of Excellence in Plant Energy, Research School of Biology, The Australian National University, Building 134, Canberra, ACT 2601, Australia; ⁶Division of Plant Sciences, Research School of Biology, The Australian National University, Building 134, Canberra, ACT 2601, Australia; ⁷Max Planck Institute for Biogeochemistry, Hans Knoell Strasse 10, 07745, Jena, Germany; ⁸German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany; ⁹UMR 1137 Ecologie et Ecophysiologie Forestières, Université de Lorraine – INRA, 54506 Vandoeuvre-lès-Nancy, France; ¹⁰Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3BJ, United Kingdom; ¹¹Joint Global Change Research Institute, DOE Pacific Northwest National Laboratory, College Park, MD USA; ¹²Department of Geography and Geology, School of Natural and Built Environments, Kingston University London, Penrhyn Road, Surrey, KT1 2EE; ¹³School of Biological Sciences, Seoul National University, Seoul 08826, South Korea; ¹⁴School of Biosciences & Veterinary Medicine, Plant Diversity and Ecosystems Management unit, University of Camerino, Italy; ¹⁵Department of Theoretical and Applied Sciences, University of Insubria, Via J.H. Dunant 3, I-21100 Varese, Italy; ¹⁶Systems Ecology, Department of Ecological Science, Vrije Universiteit, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; ¹⁷Jonah Ventures, Manhattan KS 66502; ¹⁸Department of Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser Straße 4, 06120, Halle (Saale), Germany; ¹⁹School of Earth and Environmental Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom; ²⁰Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET) and Departamento de Diversidad Biológica y Ecología, FCEFYn, Universidad Nacional de Córdoba, CC 495, Córdoba, Argentina; ²¹Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av Bandeirantes, 3900, CEP 14040-901, Bairro Monte Alegre, Ribeirão Preto, São Paulo, Brazil; ²²Normandie University, UNIROUEN, IRSTEA, ECODIV, FR-76000 Rouen, France; ²³Universidad del Rosario. Facultad de Ciencias Naturales y Matemáticas. Carrera 26 No 63B-48, Bogotá, Colombia; ²⁴Departamento de Biología y Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, C/ Tulipán s/n, 28933 Móstoles, Spain; ²⁵INRA, USC1339 Chizé (CEBC), F-79360, Villiers en Bois, France; ²⁶Centre d'étude biologique de Chizé, CNRS - Université La Rochelle (UMR 7372), F-79360, Villiers en Bois, France; ²⁷College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China; ²⁸Key Laboratory of Biogeography and Bio-resource in Arid Land, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; ²⁹School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, WITS 2050, Johannesburg, South Africa; ³⁰Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt/Main, Germany; ³¹Department of Physical Geography at Goethe-University, Frankfurt/Main; ³²Ulm University, Institute of Systematic Botany and Ecology, Albert-Einstein-Allee 11, 89081 Ulm, Germany; ³³Wageningen Environmental Research (Alterra); ³⁴Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, 90095; ³⁵Forestry and Forest Products Research Institute 1 Matsunosato, Tsukuba 305-8687, Japan; ³⁶Department of Botany, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, US; ³⁷School of Geosciences, University of Edinburgh, Edinburgh, EH9 3FF, UK; ³⁸Institute of Biology and Environmental Science, University of Oldenburg, Carl von Ossietzky-Straße 9-11, 26111, Oldenburg, Germany; ³⁹Department of Plant Physiology, Estonian University of Life Sciences, Kreutzwaldi 1, 51014 Tartu, Estonia; ⁴⁰Graduate School of Agriculture, Kyoto University, Kyoto, 606-8502 Japan; ⁴¹CSIC, Unitat d'Ecologia Global CREA-CSIC-UAB, Bellaterra 08193, Barcelona, Catalonia, Spain; ⁴²CREAF, Cerdanyola del Vallès 08193, Barcelona, Catalonia, Spain; ⁴³Michigan State University, Department of Forestry, 480 Wilson Rd., East Lansing, MI 48824; ⁴⁴Museo Nacional de Ciencias Naturales, CSIC Serrano 115 dpdo, E-28006 Madrid Spain; ⁴⁵Dept. of Biology, Algoma University, Sault Ste. Marie, Ontario, Canada P6A 2G4; ⁴⁶Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC Leiden, The Netherlands; ⁴⁷Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, Riverside, CA. 92521; ⁴⁸Embrapa Clima Temperado, Rodovia BR 392, Km 78, Pelotas, RS, Brasil, 96010-971; ⁴⁹Environmental Sciences Division, Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA; ⁵⁰Department Systematic Botany and Functional Biodiversity, University of Leipzig, 04103 Leipzig, Germany; ⁵¹Hawkesbury 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 (N_m) and phosphorus (P_m), we characterize how traits vary within and among over 50,000 $\approx 50 \times 50$ 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.

plant traits | Bayesian modeling | spatial statistics | global | climate

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 (N_m) and specific leaf area (SLA) (5–7). More recently, variation in leaf phosphorus concentration (P_m) 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 N_m and mean annual temperature) (1, 16–20), or 3) restricting the analysis to species whose

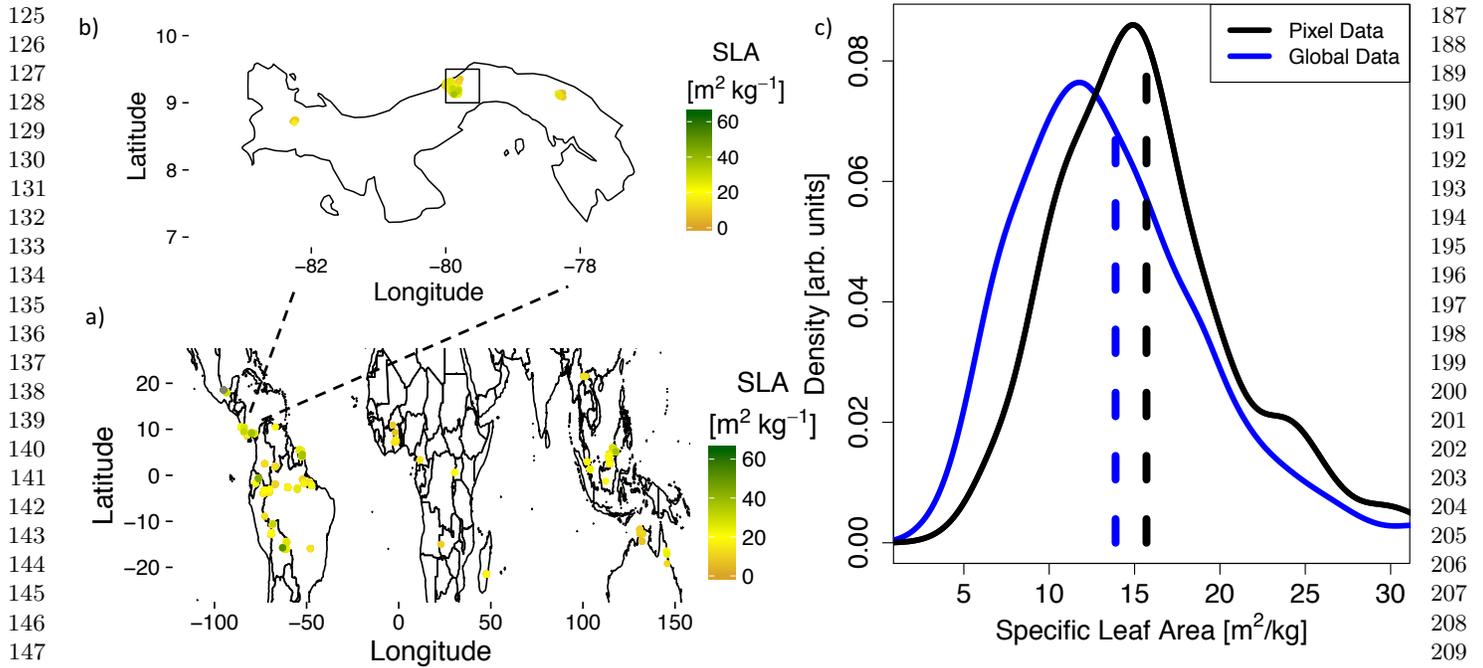


Fig. 1. Trait data a) Global locations and values of specific leaf area measurements for the PFT Tropical Broadleaf Evergreen Trees. b) Locations and values of specific leaf area measurements for the Tropical Broadleaf Evergreen Trees in Panama. The central square indicates a $0.5^\circ \times 0.5^\circ$ 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 the 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.

presence has been widely estimated on the ground (21–24). Each of these methods has limitations - for example, trait-environment relationships do not well explain observed trait spatial patterns(1, 25), while species-based approaches limit the scope of extrapolation to only areas with well measured species abundance. More critically, the first two global method-

ologies emphasized estimating a single trait value per PFT at every location, whereas both ground based (5, 14) and remotely sensed (26) observations suggest that at ecosystem or landscape scales traits would be better represented by distributions. Here, we use an updated version of the largest global database of plant traits (14) coupled with modern Bayesian spatial statistical modeling techniques (27) to capture local and global variability in plant traits. This combination allows the representation of trait variation both within pixels on a gridded land surface as well as across global environmental gradients.

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.

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 $65.2 \text{ m}^2 \text{ kg}^{-1}$). Even after limiting the scope to a single well measured $0.5^\circ \times 0.5^\circ$ pixel within Panama (Fig. 1b,c), there is still a wide range of SLA values (4.7 to $37.7 \text{ m}^2 \text{ kg}^{-1}$) with a local mean of $15.7 \text{ m}^2 \text{ kg}^{-1}$, and a local standard deviation of $5.4 \text{ m}^2 \text{ kg}^{-1}$ - over $1/3$ of the local mean. By contrast, the mean SLA value of all species associated with Broadleaf Evergreen Tropical trees is $13.9 \text{ m}^2 \text{ kg}^{-1}$, 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)

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 that 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.

^aE.E.B. and A.D. contributed equally to this work.

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

249 and global (31) scales. It has been shown that using trait
 250 distributions leads to different estimates of carbon dynamics
 251 (32) and that higher order moments of trait distributions
 252 contribute to sustaining multiple ecosystem functions (33).
 253 While species level mapping (21, 23, 24) does capture trait
 254 distributions, it has been limited geographically and restricted
 255 to subsets of functional groups.

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

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

289 The above mentioned methods allow for a representation
 290 of global vegetation that enables a more accurate formulation
 291 of functional diversity than the single trait value per PFT
 292 paradigm that is widely employed (4). The traits studied here
 293 - SLA, N_m , and P_m - are central to predicting variation in
 294 rates of plant photosynthesis (5, 6, 9, 11) and foliar respira-
 295 tion (10, 36). The importance of these traits and the more
 296 advanced representation of functional diversity developed here
 297 may be used to better capture the response of the land surface
 298 component of the Earth System to environmental change.

300 Results and Discussion

301 **Model Evaluation.** Given the full suite of nine models proposed,
 302 we conducted extensive model evaluation (see Table 1) to
 303 determine the trade-offs associated with each methodology
 304 and resolution of PFT. We assessed the predictive capability
 305 of the models using the root mean squared predictive error
 306 (RMSPE) based on out-of-sample data (SI Appendix, Section
 307 S6). Among the nine models, the spatial narrow 14-PFT model
 308 emerged as the best predictor of mean trait values for SLA and
 309 N_m , and the second best for P_m (Table 1). However, the spatial

311 broad 4-PFT model performed nearly as well (Table 1). The
 312 models' abilities to correctly estimate the spread of the trait
 313 distributions were assessed using the out-of-sample coverage
 314 probabilities (CP) – the proportion of instances the model
 315 predicted 95% confidence intervals contained the observed
 316 trait values. Most of the models provided adequate coverage
 317 (CP of around 90% or more). See the SI Appendix, Section S4,
 318 for more detailed definitions of the model comparison metrics.

320 **Table 1. Model evaluation**

SLA			
Model	ps-R ²	RMSPE	CP
Cf	NA	8.13	91.2%
Cb	16.9%	7.13	94.7%
Cn	26.0%	6.66	95.8%
Lf	4.6%	7.99	91.3%
Lb	23.4%	6.93	94.0%
Ln	30.7%	6.53	95.2%
Sf	45.5%	7.54	93.6%
Sb	58.5%	6.31	97.7%
Sn	60.2%	6.13	97.7%
N_m			
Model	ps-R ²	RMSPE	CP
Cf	NA	7.16	93.3%
Cb	12.5%	6.95	93.2%
Cn	19.4%	6.47	92.7%
Lf	5.2%	7.28	93.2%
Lb	16.7%	6.71	94.3%
Ln	24.1%	6.42	94.6%
Sf	44.2%	7.19	93.6%
Sb	53.7%	6.36	96.1%
Sn	54.8%	6.18	96.1%
P_m			
Model	ps-R ²	RMSPE	CP
Cf	NA	0.86	90.5%
Cb	5.3%	0.86	90.5%
Cn	28.1%	0.78	91.1%
Lf	25.6%	0.84	87.2%
Lb	32.8%	0.85	85.3%
Ln	35.4%	0.82	87.0%
Sf	62.0%	0.83	90.7%
Sb	66.7%	0.81	92.0%
Sn	67.6%	0.80	91.3%

319 The pseudo-R² (ps-R²), RMSPE and CP statistics for all nine models,
 320 for each of the three traits. The bold entries correspond to the model
 321 producing highest ps-R², lowest RMSPE, or CP closest to 0.95. The
 322 categorical PFT-free model (Cf) produces a constant estimate and
 323 hence ps-R² is not defined. Each model is indicated by a two-letter
 324 abbreviation: C=Categorical (no regression), L=Linear (linear
 325 regression), S=Spatial (linear regression with spatial term) and the
 326 accompanying PFT resolution: f=PFT-free (no PFT information),
 327 b=broad (4-PFT), n=narrow (14-PFT).

328 The improvement in prediction afforded by the inclusion of
 329 (1) a spatial term and (2) PFT information (Table 1) invites
 330 further examination. First, the spatial term in our model likely
 331 incorporates some of the finer scale variation that is unavailable
 332 given the relatively large grid cell size of the environmental
 333 covariates used in global studies. Thus, the spatial term allows
 334 for adjustment of trait values among neighboring or regional

373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434

435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496

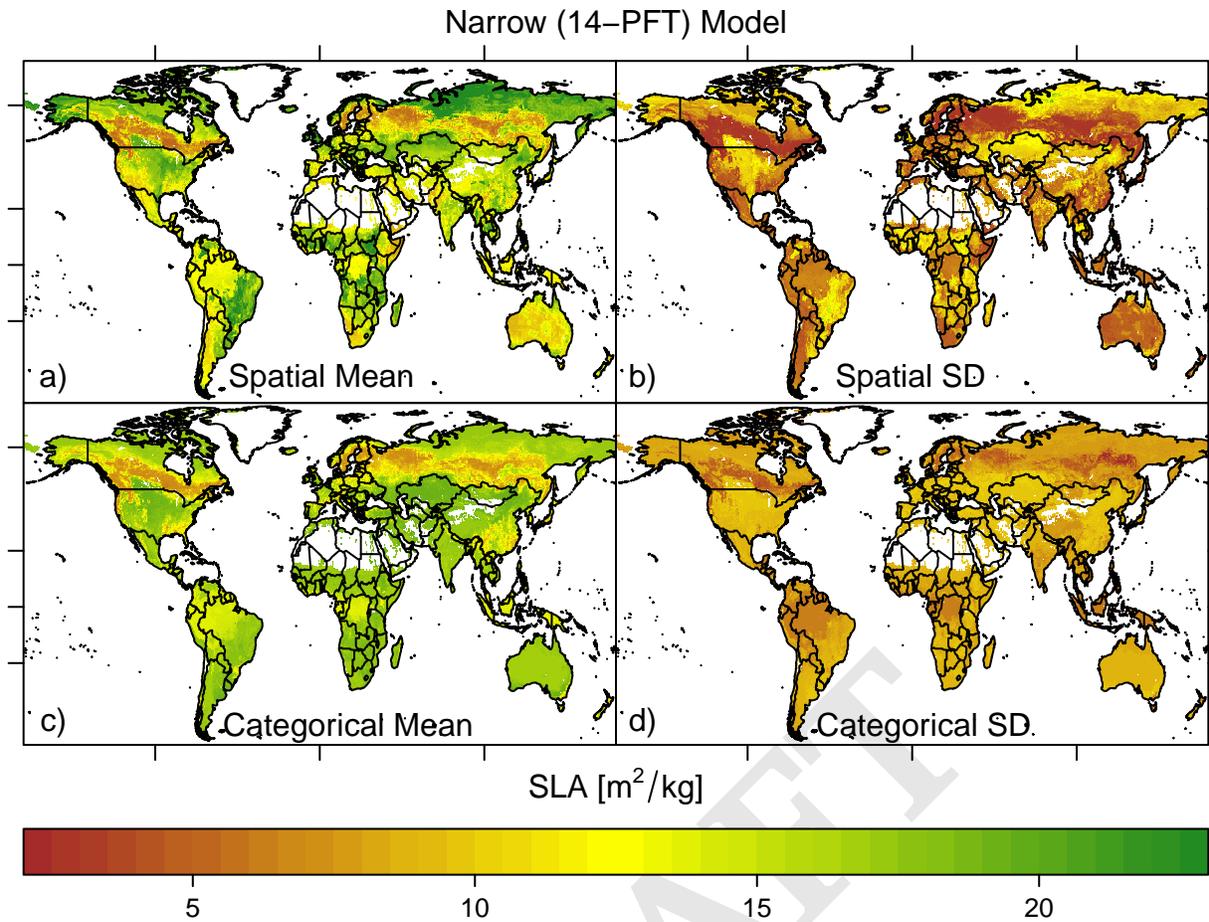


Fig. 2. Specific Leaf Area maps a,b) Narrow (14-PFT) Bayesian spatial model pixel mean and standard deviation estimates, respectively c,d) Narrow (14-PFT) Categorical model pixel mean estimates and standard deviation estimates, respectively. For clarity, the color bars have been truncated at the compound 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.

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.

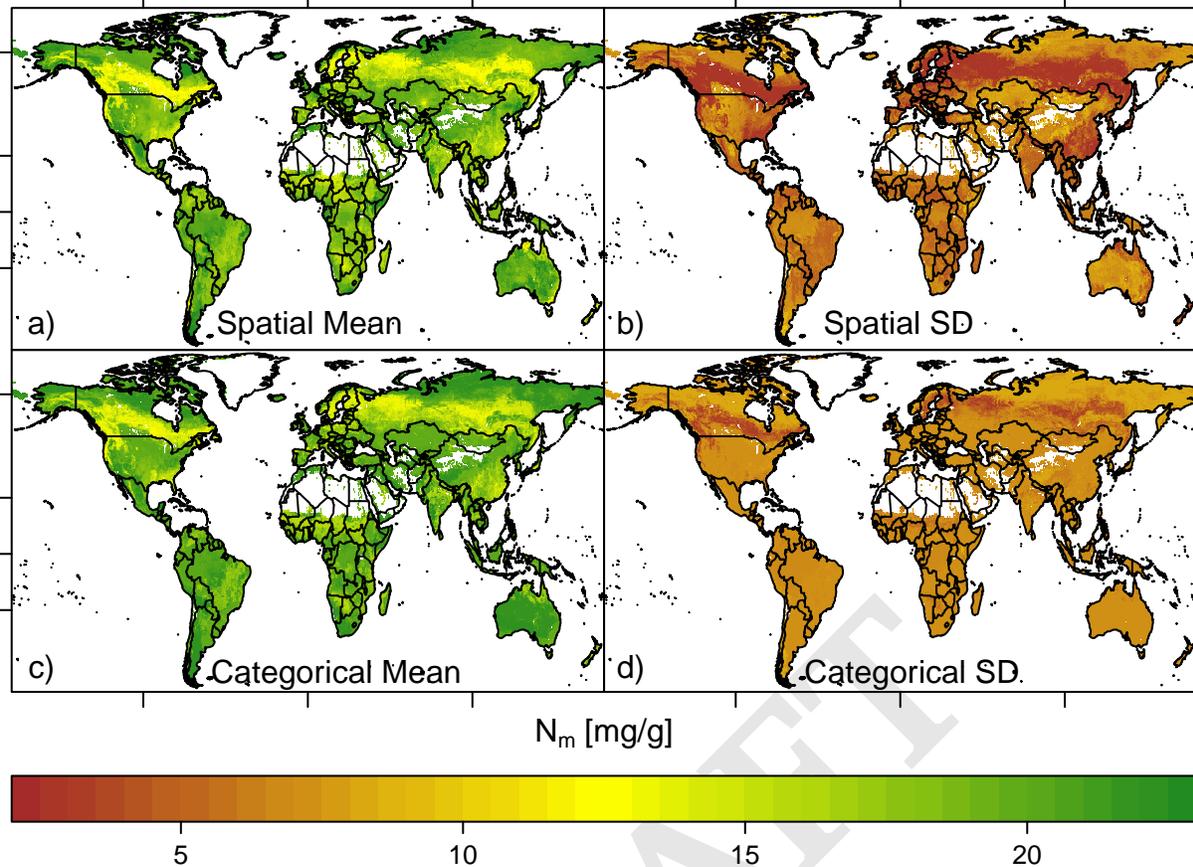
Global Maps. We selected two sets of maps to describe, in broad strokes, how trait distributions vary across the land surface: the narrow 14-PFT spatial model and its categorical counterpart. The narrow 14-PFT spatial model is the best predictor of mean trait values, and provided adequate

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. N_m (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 N_m in the spatial model across the Sahel in sub-Saharan Africa (3a,c). The lack of variation in the standard

497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558

Narrow (14-PFT) Model



559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620

Fig. 3. Nitrogen [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.

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 swathes 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

621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682

683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744

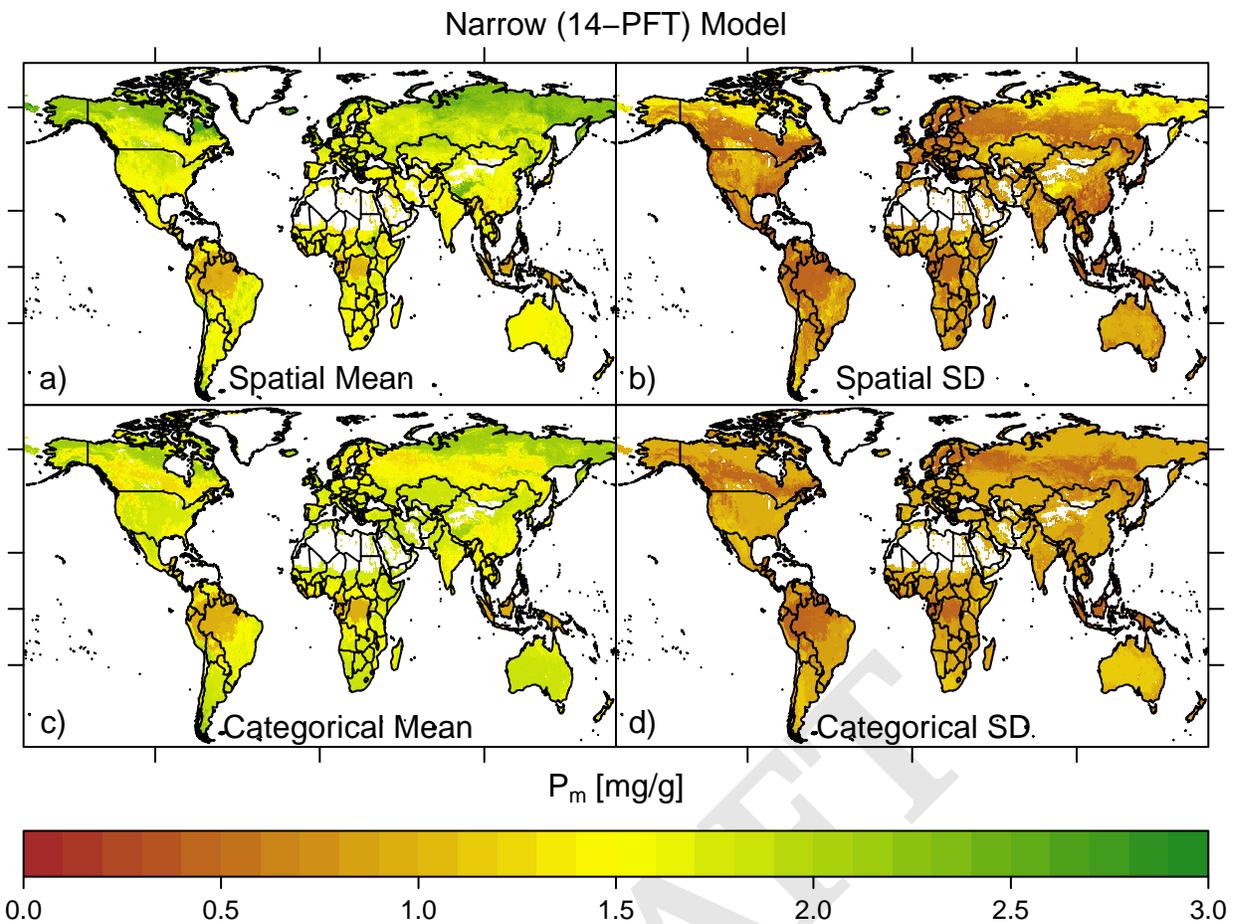


Fig. 4. Phosphorus [mass] maps a,b) Narrow (14-PFT) Bayesian spatial model pixel mean and standard deviation estimates, respectively, c,d) Narrow (14-PFT) Categorical model pixel mean estimates and standard deviation estimates, respectively. For clarity, the color bars have been truncated at the compound 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.

variation in total trait space, suggesting these environments permit a wider range of strategies than in both the boreal and tropical regions. Most broadly, both the data and the spatial model suggest (SI Appendix, Figs. S24,S25) lowest leaf nitrogen values in temperate climates; that increase in both cooler and warmer regions; this may indicate a more complicated leaf biochemistry-temperature relationship than has previously been suggested (16).

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

745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806

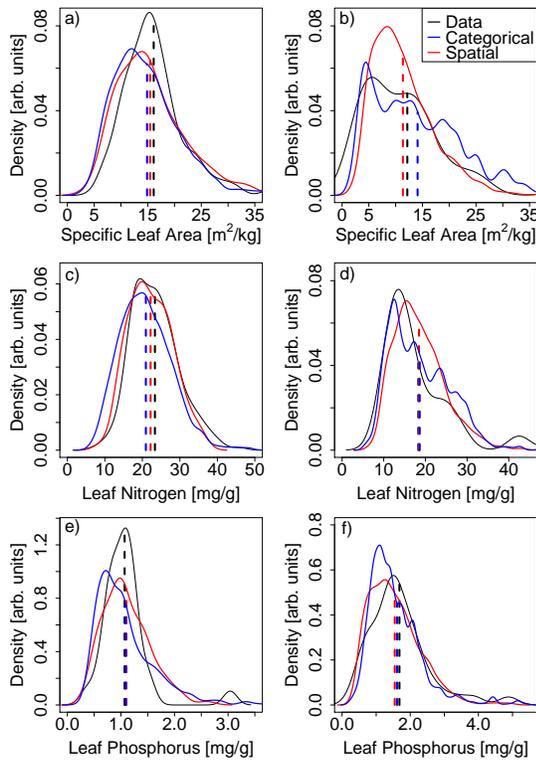


Fig. 5. Empirical trait distributions Barro Colorado Island on the left (a, c, e) and the US Southwest on the right (b, d, f). The first row is SLA (a, b), the second is leaf nitrogen (c, d) and the third is leaf phosphorus (e, f). Each panel depicts the distribution of the data in solid black, the categorical model in blue and the Bayesian spatial model in red. The vertical lines indicate mean values.

it was the PFT-free spatial model that fit best (SI Appendix, Table S6).

The differences between the trait distributions of the categorical and Bayesian spatial models were stark in the southwestern United States, although the mean estimates for N_m and P_m were close (Fig. 5b, d, f). This may be a result of the topographic complexity of this region and the resulting difficulty of aggregating climate and soil covariates at the 0.5° pixel scale and the sparser sampling than at BCI. To get enough data to approximate a distribution, we aggregated 18 pixels with nine PFTs including every temperate category, though many of them are only marginally present. The inclusion of so many PFTs produced a noisier distribution in the categorical model than suggested by the data and estimated by the spatial model. Neither of the models produced distributions that matched as well with the observations; however, it is notable how close the mean values for both models matched the observations for N_m and P_m , and the spatial model did well for the mean SLA.

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

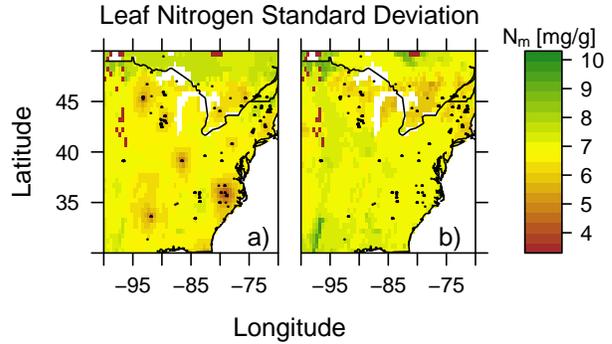


Fig. 6. Spatial learning a) the spatial model standard deviation of N_m . The predicted variation near the data locations (black dots) are much lower than variation at locations away from any data point. b) the linear model standard deviation which does not account for local spatial information has no such pattern.

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

807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868

869 representation eliminates the need to separately model the
870 future locations of PFTs (or species) when inferring the future
871 distribution of traits; hence, the output of a model like that
872 developed here could be updated with future environmental
873 covariates, with the caveats that ‘out of sample prediction’
874 may entail. At the same time, this method would allow for
875 greater functional diversity than multiple PFTs with single
876 trait values, as is currently used in most ESMs. Adopting
877 this approach does, however, raise the issue of how to deal
878 with the paucity of surface observations in some regions, as
879 evidenced by the greater errors associated with estimating out
880 of sample values with this model (Table 1). Complementary
881 work has retrieved leaf trait maps from a global carbon cycle
882 model fused with Earth observations (38), providing another
883 method that could be used for direct comparison against the
884 trait maps produced here. While the methodology outlined in
885 our analysis brings the possibility of a PFT-free land surface
886 closer, we remain several steps away from being able to make
887 such maps as accurately as we do using PFT characterizations
888 for trait prediction. Several actions can bring us closer to that
889 goal. First, incorporation of additional information (such as
890 phylogenetic relatedness and trait-trait covariance) will likely
891 improve trait maps, even using existing observations. Second,
892 as the current level of observations is extremely sparse in some
893 regions, and sparse in most, expanded trait databases will also
894 aid in development of PFT-free trait maps.

896 Conclusions

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

915 The global land surface is perhaps the most heterogeneous
916 component of the Earth System. Reducing vegetation to a col-
917 lection of PFTs with fixed trait values has been the preferred
918 method to constrain this heterogeneity and group similar bio-
919 chemical and biophysical properties; however, this has been at
920 the expense of functional diversity. This analysis quantifies the
921 substantial magnitude of this ignored trait variation. The ap-
922 proach and methods presented here retain the simplicity of the
923 PFT representation, but capture a wider range of functional
924 diversity.

928 Materials and Methods

929
930

Data. The TRY database (www.try-db.org) (14) provided all
931 data for leaf traits and the categorical traits to aggregate
932 PFTs (TRY – Categorical Traits Dataset, https://www.try-
933 db.org/TryWeb/Data.php#3, January 2016) used in the analysis.
934 See SI Appendix (Appendix 1) for a complete list of the original
935 publications associated with this subset of TRY. The extract from
936 TRY used here has just under 45,000 measurements of individuals
937 from 3,680 species with measurements of at least one of specific
938 leaf area (SLA), leaf nitrogen per dry leaf mass (N_m), and/or leaf
939 phosphorus per leaf dry mass (P_m). The number of individual
940 measurements varies from 32,315 for SLA on 2,953 species to 19,282
941 for N_m on 3,053 species down to 8,052 for P_m on 1,810 species; see
942 Table S4 for the number of unique measurements and species found
943 in all categorizations used in the analysis. The species taxonomy
944 was standardized using The Plant List (44). Measurements were
945 associated with environmental categories through Köppen-Geiger
946 climate zones (45). All environmental variables are on a $0.5^\circ \times 0.5^\circ$
947 grid. Climate variables use 30 year climatologies from 1961-1990
948 as estimated by the Climate Research Unit (46, 47). Soil variables
949 are from the International Soil Reference and Information Center
950 - World Inventory of Soil Emission Potentials (ISRIC-WISE) (48).
951 The spatial extent of PFTs have been previously estimated through
952 satellite estimates of land cover around the year 2005 (49), and
953 these estimates have been refined into climatic categories (15, 35).
954 While TRY, and thus the data used here, represents the largest col-
955 lection of plant traits in the world most of the measurements come
956 from a subset of global regions: North America, Europe, Australia,
957 China, Japan, and Brazil. There are still large sections of the planet
958 with extremely sparse measurements, notably: much of the tropics
959 outside of the Americas, large swathes of Central Asia, the Russian
960 Federation, South Asia and much of the Arctic (SI Appendix, fig.
961 S17). Improving data collection in these regions will greatly improve
962 future modeling efforts. Improving data collection in these regions
963 will greatly improve future modeling efforts. Until observations are
964 more complete there remains the possibility of spurious patterns,
965 though we have found little evidence to suggest their presence in
966 this analysis, even in comparison to detailed regional studies (SI
967 Appendix, fig. S26) (50).

Classification of PFTs and Categorical Model. We used three nested
968 levels of PFT classification. In the first level, all plants are cate-
969 gorized into a single group (‘PFT-free’). In the second level (‘broad’),
970 all plants are categorized into PFTs based on categorical traits as-
971 sociated with growth form (grass, shrub, tree) and leaf type (broad
972 and needle-leaved) leading to the following four PFTs: grasses,
973 shrubs, broad-leaved trees and needle-leaved trees (Fig. S1). In
974 the third level (‘narrow’), the broad PFTs are further refined by
975 their climatic region – tropical, temperate, boreal – as well as leaf
976 phenology, and, for the grasses, photosynthetic pathway (C_3 or C_4).
977 This produces 14 PFTs (Fig. S2), which correspond exactly to those
978 found in the community land model (CLM) (4). Note that these
979 PFT classifications exclude non-woody eudicots (‘herbs’), which
980 were excluded from the analysis, on account of their lack of domi-
981 nance within these PFT categories (51) and therefore, on account
982 of being widely measured could overly influence the structure of the
983 trait distributions if they were included. Satellite estimates of the
984 PFT abundance that correspond to the “narrow” PFT categories
985 defined above have already been calculated (15, 49) and we used
986 these to assign a percentage of each $0.5^\circ \times 0.5^\circ$ pixel to each PFT
987 present according to the fraction of the land surface within that
988 pixel occupied by the PFT. The “broad” PFT fractions are calcu-
989 lated by summing the narrow PFT categories within each “broad”
990 classification.

The categorical model uses the PFT categories and averages
991 trait values for each species across individual measurements at each
992 measured location. This defines the PFT as the interspecies range
993 of trait values and ignores all local environmental factors. The
994 results of the categorical model are summarized by the mean and
995 standard deviation of each PFTs trait values (Table S4) for all three
996 resolutions of the model. Note that in the PFT-free case where no
997 PFT information is used, the categorical model produces a constant
998 trait distribution across the entire vegetated world. The categorical
999 model, and the Bayesian models described in the following section
1000 all use location specific species mean values to estimate trait dis-
1001 tributions. We assume no intra-specific variation in trait values.

993 However, in regions dominated by a small number of species this
994 may lead to biased predictions. The hyper dominance of a small
995 group of species in the Amazon has recently been demonstrated
996 (52) and thus serves as a case study to evaluate our assumption of
997 equal species weighting (S8, fig. S23). We found that equal weights
998 (species means) produced trait distribution estimates closest to
999 those of the hyper dominant trait abundances and this reinforces
1000 the use of this assumption globally. Further, as noted above, the
1001 omission of herbaceous species from tropical regions in this analysis
1002 (and (52)) may unduly limit trait diversity, and calls for further
1003 research.

1003 **Bayesian Models.** A more fine-tuned depiction of geographical or spatial
1004 variation of plant trait values within each PFT can be achieved
1005 by leveraging environmental and location information, which allows
1006 trait values to adjust based on local conditions. Data for 17 climate
1007 (46, 47) and soil based (48) environmental predictors were available
1008 at the $0.5^\circ \times 0.5^\circ$ pixel resolution used to create the trait maps. To
1009 avoid overfitting and collinearity issues, these seventeen predictors
1010 were screened (see Section S7) based on correlations amongst pre-
1011 dictors, their individual correlation with the traits, and to include
1012 climate covariates along different axes of environmental stress and
1013 both chemical and physical soil covariates. We finally selected five
1014 predictors – mean annual temperature [MAT], total annual radiation
1015 [RAD], moisture index (precipitation/evapotranspiration) [MI],
1016 percent hydrogen (aqueous) [pH], and percent clay content [CLY].
1017 Remote sensing data products, such as Normalized Difference Vegetation
1018 Index (53)), are not used as covariates, to allow for inference
1019 outside of the historical observation period through perturbations
1020 of environmental covariates.

1021 We utilized environment-trait relationships to obtain predictions
1022 of trait values (1, 16–18, 37, 43) in a linear regression setup. The
1023 formal details of the initial model are as follows. We denote log-
1024 transformed trait values at a geographical location s as $y_{trait}(s)$.
1025 This set of five predictors at a location s is denoted by the vector
1026 $x(s) = (x_1(s), x_2(s), \dots, x_5(s))'$. A linear regression model relating
1027 the trait to the environmental predictors is specified as:

$$1028 \quad y_{trait}(s) = b_0 + b_1x_1(s) + b_2x_2(s) + \dots + b_5x_5(s) + \epsilon(s) \quad [1]$$

1029 where b_i are the regression coefficients and $\epsilon(s)$ is the error term
1030 explaining residual variation. Estimation of model parameters and
1031 prediction were achieved with a fully Bayesian hierarchical model.
1032 This enables inclusion of prior information and prediction of full trait
1033 distributions instead of representative values (like mean or median)
1034 thereby ensuring that the uncertainty associated with the estimation
1035 of model parameters is fully propagated into the predictive trait
1036 distributions.

1037 We then generalized the above model into a Bayesian spatial linear
1038 regression model that borrows information from geographically
1039 proximal regions to capture residual spatial patterns beyond what
1040 is explained by environmental predictors. A customary specification
1041 of a spatial regression model is obtained by splitting up the error
1042 term $\epsilon(s)$ in Equation (1) into the sum of a spatial process $w(s)$
1043 and an error term $\eta(s)$, that accounts for the residual variation
1044 after adjusting for the spatial effects $w(s)$. The underlying latent
1045 process $w(s)$ accounts for local nuances beyond what is captured
1046 by the environmental predictors and is often interpreted as the net
1047 contribution from unobserved or unusable predictors. Gaussian
1048 Processes (GP) are widely used for modeling unknown spatial surfaces
1049 such as $w(s)$, due to their convenient formulation as a multivariate
1050 Gaussian prior for the spatial random effect, unparalleled predictive
1051 performance (54) and ease of generating uncertainty quantified
1052 predictions at unobserved locations. We use the computationally
1053 effective Nearest Neighbor Gaussian Process (27) which nicely embeds
1054 into the Bayesian hierarchical setup as a prior for $w(s)$ in the
1055 second stage of the model specification. All technical specifications
1056 of the Bayesian spatial model are provided in Section S1 of the
1057 supplementary materials.

1058 The linear regression models used in previous studies (1, 16–18)
1059 and both the spatial and non-spatial Bayesian models described
1060 above assume a global relationship between the traits and environ-
1061 ment. Given the goal of predicting trait values for the entire land
1062 surface, the assumption of a universal trait-environment relationship
1063 may be an oversimplification (55). Moreover, if there is significant
1064 variation in plant trait values among different PFTs, the estimated

1055 parameters will be skewed towards values from abundantly sampled
1056 PFTs, such as broad-leaved trees. Additional information about
1057 plant characteristics at a specific location, if available, can poten-
1058 tially be used to improve predictions. As mentioned earlier, we have
1059 PFT classifications for each observation of the dataset used here
1060 and satellite estimates of PFT abundance at all pixels. The global
1061 regression approaches described above ignores this information and
1062 can yield biased predictions at locations dominated by PFTs poorly
1063 represented in the data, such as shrubs. Hence, we also incorpo-
1064 rate the PFT information in these regression models by allowing
1065 the trait-environment relationship to vary between different PFTs.
1066 Finally, the PFT specific distributions from the Bayesian models
1067 were weighted by the satellite based PFT abundances to create a
1068 landscape scale trait distribution, thereby enabling straightforward
1069 comparison between all three categorizations of PFT. Details of the
1070 PFT based Bayesian models are provided in Section S2. The use
1071 of a Gaussian Process based spatial model as well as the Bayesian
1072 implementation of the regression models were novel to this applica-
1073 tion of plant trait mapping and, as results indicated, were critical
1074 to improving model predictions as well as properly quantifying trait
1075 distributions.

1076 All the code and public data are available from the authors upon
1077 request. The TRY data may be requested from the TRY database
1078 custodians.

1079 **ACKNOWLEDGMENTS.** E.E.B., H.F.M., M.C., K.R.W., and
1080 P.B.R. acknowledge funding from the United States Department of
1081 Energy, Office of Science (DE-SC0012677). O.K.A. acknowledges
1082 the support of the Australian Research Council (CE140100008).
1083 P.B.R. also acknowledges support from two University of Minnesota
1084 Institute on the Environment Discovery Grants. The study has been
1085 supported by the TRY initiative on plant traits ([http://www.try-
1087 db.org](http://www.try-
1086 db.org)). The TRY initiative and database is hosted, developed and
1088 maintained at the Max Planck Institute for Biogeochemistry, Jena,
1089 Germany. TRY is currently supported by DIVERSITAS/Future
1090 Earth and the German Centre for Integrative Biodiversity Research
1091 (iDiv) Halle-Jena-Leipzig. BB acknowledges a NERC independent
1092 research fellowship NE/M019160/1. JP would like to acknowledge
1093 the financial support from the European Research Council Synergy
1094 grant ERC-SyG-2013-610028 IMBALANCE-P, the Spanish Govern-
1095 ment grant CGL2013-48074-P and the Catalan Government grant
1096 SGR 2014-274. B.B.-L. was supported by the Earth System Model-
1097 ing program of the U.S. Department of Energy, Office of Science,
1098 Office of Biological and Environmental Research. KK acknowledges
1099 the contribution of the WUR Investment theme Resilience for the
1100 project Resilient Forest (KB-29-009-003). PM acknowledges support
1101 from ARC grant FT110100457 and NERC NE/F002149/1. WH
1102 acknowledges support from the National Natural Science Founda-
1103 tion of China (#41473068) and “Light of West China” Program
1104 of Chinese Academy of Sciences. We would also like to thank the
1105 improvements suggested by two anonymous referees, which improved
1106 the clarity and depth of the manuscript.

- 1107 1. Van Bodegom PM, Douma JC, Verheijen LM (2014) A fully traits-based approach to modeling
1108 global vegetation distribution. *Proceedings of the National Academy of Sciences of the United
1109 States of America* 111(38):13733–8.
- 1110 2. Maire V, et al. (2015) Global effects of soil and climate on leaf photosynthetic traits and rates.
1111 *Global Ecology and Biogeography* 24(6):706–717.
- 1112 3. DeFries RS, et al. (1995) Mapping the land surface for global atmosphere-biosphere models:
1113 Toward continuous distributions of vegetation's functional properties. *Journal of Geophysical
1114 Research* 100(D10):20867.
- 1115 4. Bonan GB, et al. (2011) Improving canopy processes in the Community Land Model version
1116 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geo-
1117 physical Research* 116(G2):1–22.
- 1118 5. Reich PB, Ellsworth DS, Walters MB (1998) Leaf structure (specific leaf area) modulates
1119 photosynthesis – nitrogen relations : evidence from within and across species and functional
1120 groups. *Functional Ecology* 12:948–958.
- 1121 6. Kattge J, Knorr W, Raddatz T, Wirth C (2009) Quantifying photosynthetic capacity and its
1122 relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Global
1123 Change Biology* 15(4):976–991.
- 1124 7. Crous KY, et al. (2017) Nitrogen and phosphorus availabilities interact to modulate leaf trait
1125 scaling relationships across six plant functional types in a controlled-environment study. *New
1126 Phytologist*.
- 1127 8. Wright IJ, et al. (2004) The worldwide leaf economics spectrum. *Nature* 428(6985):821–827.
- 1128 9. Reich PB, Oleksyn J, Wright IJ (2009) Leaf phosphorus influences the photosynthesis-
1129 nitrogen relation: A cross-biome analysis of 314 species. *Oecologia* 160(2):207–212.

1117 10. Atkin OK, et al. (2015) Global variability in leaf respiration in relation to climate, plant functional types and leaf traits. *New Phytologist* 206(2):614–636. 1179

1118 11. Bahar N, et al. (2016) Leaf-level photosynthetic capacity in lowland Amazonian and high elevation, Andean tropical moist forests of Peru. *New Phytologist*. 1180

1119 12. Rowland L, et al. (2016) Scaling leaf respiration with nitrogen and phosphorus in tropical forests across two continents. 1181

1120 13. Reich PB (2005) Global biogeography of plant chemistry: Filling in the blanks. *New Phytologist* 168(2):263–266. 1182

1121 14. Kattge J, et al. (2011) TRY - a global database of plant traits. *Global Change Biology* 17(9):2905–2935. 1183

1122 15. Oleson KW, et al. (2013) Technical Description of version 4.5 of the Community Land Model (CLM), Technical Report Climate and Global Dynamics Division. 1184

1123 16. Reich PB, Oleksyn J (2004) Global patterns of plant leaf N and P in relation to temperature and latitude. *Proceedings of National Academy of Sciences* 101(30):11001–11006. 1185

1124 17. Ordoñez JC, et al. (2009) A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Global Ecology and Biogeography* 18(2):137–149. 1186

1125 18. Simpson AH, Richardson SJ, Laughlin DC (2016) Soil-climate interactions explain variation in foliar, stem, root and reproductive traits across temperate forests. *Global Ecology and Biogeography* 25(8):964–978. 1187

1126 19. Reich PB, Wright IJ, Lusk CH (2007) Predicting Leaf Physiology from Simple Plant and Climate Attributes : A Global GLOPNET Analysis. *Ecological Applications* 17(7):1982–1988. 1188

1127 20. Reich PB, Rich RL, Lu X, Wang YP, Oleksyn J (2014) Biogeographic variation in evergreen conifer needle longevity and impacts on boreal forest carbon cycle projections. *Proceedings of the National Academy of Sciences* 111(38):13703–13708. 1189

1128 21. Swenson NG, et al. (2012) The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography* 21(8):798–808. 1190

1129 22. Hawkins BA, Rueda M, Rangel TF, Field R, Diriz-Filho JAF (2014) Community phylogenetics at the biogeographical scale: Cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography* 41(1):23–38. 1191

1130 23. Šimová I, et al. (2015) Shifts in trait means and variances in North American tree assemblages: Species richness patterns are loosely related to the functional space. *Ecography* 38(7):649–658. 1192

1131 24. Swenson NG, et al. (2017) Phylogeny and the prediction of tree functional diversity across novel continental settings. *Global Ecology and Biogeography* pp. 1–12. 1193

1132 25. Douma JC, de Haan MWA, Aerts R, Witte JPM, van Bodegom PM (2012) Succession-induced trait shifts across a wide range of NW European ecosystems are driven by light and modulated by initial abiotic conditions. *Journal of Ecology* 100(2):366–380. 1194

1133 26. Asner GP, Knapp DE, Anderson CB, Martin RE, Vaughn N (2016) Large-scale climatic and geophysical controls on the leaf economics spectrum. *Proceedings of National Academy of Sciences*. 1195

1134 27. Datta A, Banerjee S, Finley A, Gelfand A (2016) Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association* 111(514):800–812. 1196

1135 28. Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149(1):78–90. 1197

1136 29. Scheiter S, Higgins SI (2009) Impacts of climate change on the vegetation of Africa: An adaptive dynamic vegetation modelling approach. *Global Change Biology* 15(9):2224–2246. 1198

1137 30. Scheiter S, Langan L, Higgins SI (2013) Next-generation dynamic global vegetation models: learning from community ecology. *The New Phytologist* 198(3):957–69. 1199

1138 31. Pavlick R, Drewry DT, Bohn K, Reu B, Kleidon a (2012) The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional trade-offs. *Biogeosciences Discussions* 9(4):4627–4726. 1200

1139 32. Pappas C, Faticchi S, Burlando P (2014) Terrestrial water and carbon fluxes across climatic gradients: does plant diversity matter? *New Phytologist* 16(i):3663. 1201

1140 33. Gross N, et al. (2017) Functional trait diversity maximizes ecosystem multifunctionality. *Nature Ecology & Evolution* 1(5):0132. 1202

1141 34. Clark DB, et al. (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development* 4(3):701–722. 1203

1142 35. Bonan GB (2002) Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles* 16(2):5.1–5.18. 1204

1143 36. Meir P, Grace J, Miranda AC (2001) Leaf respiration in two tropical rainforests: Constraints on physiology by phosphorus, nitrogen and temperature. *Functional Ecology* 15(3):378–387. 1205

1144 37. Verheijen LM, et al. (2015) Inclusion of ecologically based trait variation in plant functional types reduces the projected land carbon sink in an earth system model. *Global Change Biology* 21(8):3074–3086. 1206

1145 38. Bloom AA, Exbrayat JF, van der Velde IR, Feng L, Williams M (2016) The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences* (22):1–6. 1207

1146 39. Meir P, Levy PE, Grace J, Jarvis PG (2007) Photosynthetic parameters from two contrasting woody vegetation types in West Africa. *Plant Ecology* 192(2):277–287. 1208

1147 40. Domingues TF, et al. (2010) Co-limitation of photosynthetic capacity by nitrogen and phosphorus in West Africa woodlands. *Plant, Cell & Environment* 33(6):959–980. 1209

1148 41. Zhang Q, Wang YP, Pitman AJ, Dai YJ (2011) Limitations of nitrogen and phosphorus on the terrestrial carbon uptake in the 20th century. *Geophysical Research Letters* 38(22):1–5. 1210

1149 42. Medlyn B, et al. (2016) Using models to guide experiments: a priori predictions for the CO₂ response of a nutrient- and water-limited mature Eucalypt woodland. *Global Change Biology* pp. 2834–2851. 1211

1150 43. Verheijen LM, et al. (2013) Impacts of trait variation through observed trait-climate relationships on performance of an Earth system model: A conceptual analysis. *Biogeosciences* 10(8):5497–5515. 1212

1151 44. 1.1 TPLV (2013) January. <http://www.theplantlist.org/>. 1213

1152 45. Peel B, Finlayson BL, McMahon Ta (2007) Updated world map of the Köppen-Geiger climate classification.pdf. *Hydrology and Earth System Sciences* 11:1633–1644. 1214

1153 46. New M, Hulme M, Jones P (1999) Representing Twentieth-Century Space – Time Climate Variability . Part I : Development of a 1961 – 90 Mean Monthly Terrestrial Climatology. *Journal of Climate* 12:829–856. 1215

1154 47. Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* 34(3):623–642. 1216

1155 48. Batjes N (2005) ISRIC-WISE: Global data set of derived soil properties on a 0.5 by 0.5 degree grid (Version 3.0). (December):1–64. 1217

1156 49. Lawrence PJ, Chase TN (2007) Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *Journal of Geophysical Research: Biogeosciences* 112(1). 1218

1157 50. Asner GP, et al. (2017) Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science* 355(6323):385–389. 1219

1158 51. Gibson DJ (2009) *Grasses & Grassland Ecology*. (Oxford University Press, New York). 1220

1159 52. ter Steege H, et al. (2013) Hyperdominance in the Amazonian tree flora. *Science* 342(6156):1243092. 1221

1160 53. Ollinger SV, et al. (2008) Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences of the United States of America* 105(49):19336–41. 1222

1161 54. Rasmussen C (1996) *Evaluation of Gaussian Processes and other methods for non-linear regression*. (University of Toronto), Phd thesis edition. 1223

1162 55. Verheijen LM, Aerts R, Bönnisch G, Kattge J, Van Bodegom PM (2016) Variation in trait trade-offs allows differentiation among predefined plant functional types: Implications for predictive ecology. *New Phytologist* 209(2):563–575. 1224

1163 1225

1164 1226

1165 1227

1166 1228

1167 1229

1168 1230

1169 1231

1170 1232

1171 1233

1172 1234

1173 1235

1174 1236

1175 1237

1176 1238

1177 1239

1178 1240