ARTICLE

Check for updates

An equation of state unifies diversity, productivity, abundance and biomass

John Harte [1,2,3✉], Micah Brush[1,4], Erica A. Newman [5] & Kaito Umemura[1,6]

To advance understanding of biodiversity and ecosystem function, ecologists seek widely applicable relationships among species diversity and other ecosystem characteristics such as species productivity, biomass, and abundance. These metrics vary widely across ecosystems and no relationship among any combination of them that is valid across habitats, taxa, and spatial scales, has heretofore been found. Here we derive such a relationship, an equation of state, among species richness, energy flow, biomass, and abundance by combining results from the Maximum Entropy Theory of Ecology and the Metabolic Theory of Ecology. It accurately captures the relationship among these state variables in 42 data sets, including vegetation and arthropod communities, that span a wide variety of spatial scales and habitats. The success of our ecological equation of state opens opportunities for estimating difficult-to-measure state variables from measurements of others, adds support for two current theories in ecology, and is a step toward unification in ecology.

[1] The Energy and Resources Group, University of California, Berkeley, CA 94720, USA. [2] The Rocky Mountain Biological Laboratory, Gothic, CO 81224, USA. [3] The Santa Fe Institute, Santa Fe, NM 87501, USA. [4] The Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada. [5] School of Natural Resources & the Environment, University of Arizona, Tucson, AZ 85721, USA. [6] Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo 657-8501, Japan. ✉email: jharte@berkeley.edu

A major focus of ecology is the study of species diversity across ecosystems and its relationship to ecosystem structure and function[1–8]. Considerable effort has been directed at the challenge of finding other macro-level ecosystem variables, such as productivity, that might correlate with and potentially explain the wide range of values of species richness observed in different habitats, climates, and taxa across spatial scales. Species diversity and productivity are system-level descriptors of ecosystems. Other system descriptors include the total biomass and abundance of individuals.

Diversity, productivity, abundance, and biomass in ecology are loosely analogous to state variables in thermodynamics, such as the pressure, volume, temperature, and the number of moles of a container of gas. In thermodynamics, a universal relationship among state variables, also known as an equation of state, exists in the form of the ideal gas law: $PV = nRT$. Equations of state are common in physics and chemistry and derive from fundamental theory, but in macroecological studies of ecosystems, such framing has been lacking. A successful equation of state derived from ecological theory would deepen our understanding of ecology, allow prediction of diversity or productivity from knowledge of other system-level state variables, and potentially enhance applications of ecological theory to conservation and restoration[9].

In thermodynamics, it has proven useful to distinguish the micro-level and the macro-level descriptions of the system and then maximize Shannon information entropy[10,11] to infer phenomena at the micro-level from constraints imposed by state variables at the macro-level. For example, the Boltzmann distribution of molecular kinetic energies can be derived from knowledge of the total system energy and the number of molecules. Extending this concept to ecology, we take the micro-level variables to be the metabolic rates, ε, of individuals and the abundances, n, of species within an ecological community of, for example, plants, arthropods, or mammals. We take the macro-level state variables to be the total number of species, S, the total number of individuals, N, in the community, and the total metabolic rate, E, of all the individuals in a given area A. An application of MaxEnt then results in the Maximum Entropy Theory of Ecology (METE)[12–14], which we use to derive an equation of state.

At the core of METE is the "ecosystem structure function" $R(n, \varepsilon | S, N, E)$, a joint probability distribution over two micro-level variables, abundance n, and metabolism ε, that is constrained by values of the state variables S, N, and E. $R \cdot d\varepsilon$ is the probability that if a species is picked from the species pool, then it has abundance n, and if an individual is picked at random from all the species with abundance n, then its metabolic energy requirement is in the interval $(\varepsilon, \varepsilon + d\varepsilon)$. From the definition of R, total abundance, N, is S times the average of n, and total metabolic rate, E, is S times the average of $n\varepsilon$, where both averages are taken over the distribution R. The form of R is derived by maximizing its Shannon information entropy subject to the constraints imposed by the two independent ratios that can be formed from S, N, and E: N/S and E/S. The MaxEnt solution[12] is $R = e^{-\lambda_1 n} e^{-\lambda_2 n\varepsilon} / Z$ where the λ_i are Lagrange multipliers and Z is the normalization factor, all of which depend only on S, N, and E.

From the ecosystem structure function, predictions for the distribution of abundances across species and the distribution of metabolic rates across individuals, as well as a relationship between the abundance of a species and the average metabolism of its individuals, can be derived[12–14]. In spatially explicit applications, the area of the system is included as a fourth state variable and additional predictions follow, including a universal scale-collapse expression for the species-area relationship[15]. If higher taxonomic levels such as genus or family are included as additional state variables in addition to species, then the theory predicts the distribution of species over these added taxonomic categories and explicit dependence of the abundance-metabolism relationship on the taxonomic structure of the community[16].

Across ecosystems, no one of the three state variables, S, N, and E, that define METE at the macro-level is accurately predicted by the other two. Although models have been proposed that do relate species richness to the abundance or to either productivity or metabolic rate[6,17,18], empirically, there is considerable scatter around such relationships (Supplementary Note 1 and Supplementary Video). In a pioneering study[6], Fisher et al. derived a parameterized relationship between S and N from the widely-applicable log-series distribution of abundances. However, the parameter (Fisher's alpha) is observed to vary considerably from one ecosystem to another, from one taxonomic group to another, and across spatial scales. That variability has not been shown, either theoretically or empirically, to be explainable by abundance or productivity[17].

Various other semi-empirical relationships among biomass, species richness, abundance, and productivity or metabolic rate have been proposed as well[5,7,8]. Moreover, an effort has been directed at linking patterns in macroecology by searching for associations among hypothesized power-law exponents used to characterize various scaling relationships among, for example, abundance, body size, and spatial distribution[19,20]. None of these efforts have yielded the sought-after widely-applicable unification of state variables in ecology.

Here we derive an equation of state for ecology by combining results from the maximum entropy theory of ecology[12–14] with a mass-metabolism scaling rule from the metabolic theory of ecology[18,21]. The equation of state provides a relationship among four state variables: total biomass, B, total metabolic rate, E, the total number of individuals, N, and species richness, S. We demonstrate the accuracy of this equation of state across a wide variety of taxa, habitats, and spatial scales.

Results

To derive the relationship among macro-level ecological variables, which would constitute an ecological analog of the thermodynamic equation of state, we introduce a fourth state variable, B, the total biomass in the community. The ecological analog of the thermodynamic equation of state, an expression for biomass, B, in terms of S, N, and E, arises if we combine METE with a scaling result from the metabolic theory of ecology (MTE)[18,21]. In particular, we assume the MTE scaling relationship between the metabolic rate, ε, of an individual organism and its mass, m: $\varepsilon \sim m^{3/4}$. Without loss of generality[22], units are normalized such that the smallest mass and the smallest metabolic rate within a censused plot are each assigned a value of 1. With this units convention, the proportionality constant in this scaling relationship can be assigned a value of 1. From the definition of the structure-function, it follows[23] that averaging the biomass of individuals times the abundance of species, $n\varepsilon^{4/3}$, over the distribution R and multiplying by the number of species gives the total ecosystem biomass as a function of S, N, and E. Explicitly:

$$B = S \sum_n n \int d\varepsilon \, \varepsilon^{4/3} R(n, \varepsilon | S, N, E) \qquad (1)$$

Both the sum and integral in the above equation can be calculated numerically, and Python code to do so for a given set of state variables S, N, and E, is available at github.com/micbru/equation of_ state/.

We can also approximate the solution to Eq. 1 analytically (Supplementary Note 2) to reveal the predicted functional
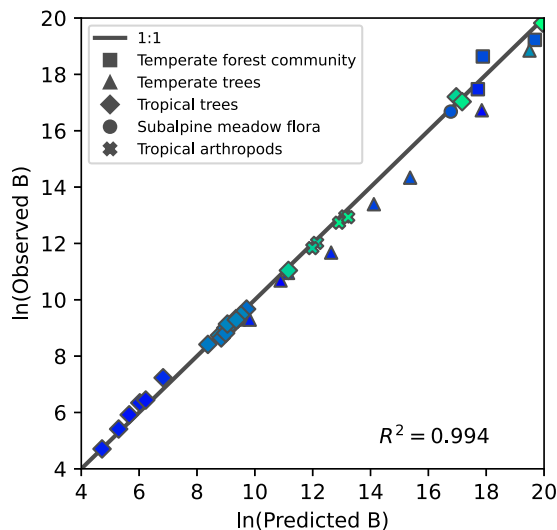
Fig. 1 A test of the ecological equation of state. Observed biomass is determined by either summing empirical masses of individuals or summing empirical metabolic rates raised to the ¾ power of each individual. Predicted biomass is determined from Eq. 1 using observed values of S, N, and E. The quantity ln(predicted biomass) explains 99.4% of the variance in observed biomass. Units of mass and metabolism are chosen such that the masses of the smallest individuals in each dataset are set to 1 and those individuals are also assigned a metabolic rate of 1. The shape of the marker indicates the type of data, and the lighter color corresponds to higher species richness. Data for all analyses come from tropical trees[39-45], temperate trees[30-33,46-48], temperate forest communities[27,49], subalpine meadow flora[28], and tropical island arthropods[50].

relationship among the four state variables. If $E \gg N \gg S \gg 1$:

$$B = c\frac{E^{4/3}}{S^{1/3}\ln(1/\beta)} \qquad (2)$$

where $c \approx (7/2)\Gamma(7/3) \approx 4.17$ and $\beta = \lambda_1 + \lambda_2$ is estimated[13,22] from the relationship $\beta\ln(1/\beta) \approx S/N$. Equation 2 approximates the numerical result to within 10% for 5 of the 42 datasets analyzed here, corresponding to N/S greater than ~100 and E/N greater than ~25. Multiplying the right-hand side of Eq. 2 by $1 - 1.16\beta^{1/3}$ approximates the numerical result to within 10% for 33 of the 42 datasets analyzed here, corresponding to N/S greater than ~3 and E/N greater than ~5. The inequality requirements are not necessary for the numerical solution of Eq. 1, which is what is used below to test the prediction.

Empirical values of E and B can be estimated from the same data. In particular, if measured metabolic rates of the individuals are denoted by ε_i, where i runs from 1 to N, then E is given by the sum over the ε_i and B is given by the sum over the $\varepsilon_i^{4/3}$. Similarly, if the mass, m_i, of each individual is measured, then B is the sum over the m_i and E is the sum over the $m_i^{3/4}$. In practice, for animal data, metabolic rate is often estimated by measuring mass and then using metabolic scaling, while for tree data, metabolic rate is estimated from measurements of individual tree basal areas, which are estimators[5] of the ε_i.

With E and B estimated from the same measurements, the question naturally arises as to whether a simple mathematical relationship holds between them, such as $E = B^{3/4}$. If all the measured m's, are identical, then all the calculated individual ε's are identical, and with our units convention we would have $E = B$. More generally, with variation in masses and metabolic rates, the only purely mathematical relationship we can write is inequality between E and $B^{3/4}$: $E = \sum\varepsilon_i \geq (\sum\varepsilon_i^{4/3})^{3/4} = B^{3/4}$. Our derived equation of state (Eq. 2) can be interpreted as expressing the
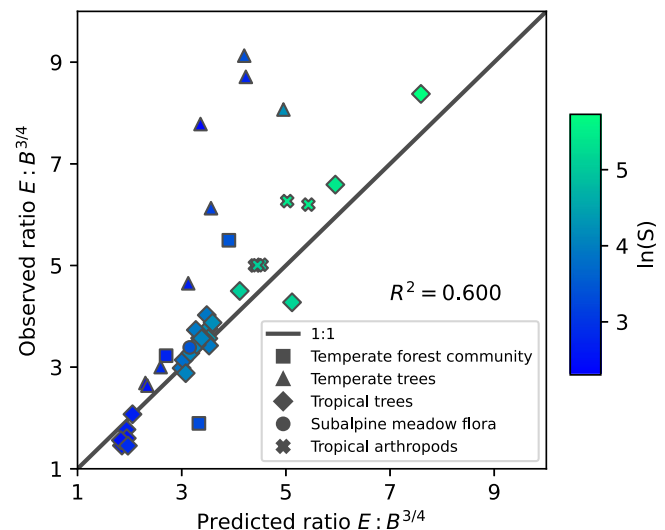

Fig. 2 The explanatory power of diversity and abundance. The observed ratio $E/B^{3/4}$ is plotted against the ratio predicted by Eq. 1. Of the fourfold variability across ecosystems in that ratio, 60% is explained by the variability in the predicted combination of diversity and abundance. The shape of the marker indicates the type of data, and the lighter color corresponds to higher species richness. Data for all analyses come from tropical trees[39-45], temperate trees[30-33,46-48], temperate forest communities[27,49], subalpine meadow flora[28], and tropical island arthropods[50].

theoretical prediction for the quantitative degree of inequality between E and $B^{3/4}$ as a function of S and N.

A test of Eq. 1 that compares observed and predicted values of biomass with data from 42 censused plots across a variety of habitats, spatial scales, and taxa is shown in Fig. 1. The 42 plots are listed and described in Table S2 and Supplementary Note 3. The communities censused include arthropods and plants, the habitats include both temperate and tropical, and the census plots range in area from 0.0064 to 50 ha. As seen in the figure, 99.4% of the variance in the observed values of B is explained by the predicted values of B.

Figure 2 addresses the possible concern that the success of Eq. 1 shown in Fig. 1 might simply reflect an approximate constancy, across all the datasets, of the ratio of E to $B^{3/4}$. If that ratio were constant, then S and N would play no effective role in the equation of state. Equation 1 predicts that variation in the ratio depends on S and N in the approximate combination $S^{1/4}\ln^{3/4}(1/\beta(N/S))$. In Fig. 2, the observed and predicted values of $E/B^{3/4}$ calculated from Eq. 1, are compared, showing a nearly fourfold variation in that ratio across the datasets. The equation of state predicts 60% of the variance in the ratio.

Figure 3 shows the dependence on S and N of the predicted ratio $E/B^{3/4}$ over empirically observed values of S, N, and E. We examined the case in which S is varied for two different fixed values of each of N and E (Fig. 3a) and N is varied for two different fixed values of S and E (Fig. 3b). The value of E does not have a large impact on the predicted ratio, particularly when $E \gg N$. On the other hand, the predicted ratio depends more strongly on N and S.

The total productivity of an ecological community is a focus of interest in ecology[1], as a possible predictor of species diversity[24] and more generally as a measure of ecosystem functioning[25]. By combining the METE and MTE frameworks, we can now generate explicit predictions for certain debated ecological relationships, including one between productivity and diversity. Interpreting total metabolic rate E in our theory as gross productivity, then in the limit $1 \ll S \ll N \ll E$, we can rewrite Eq. 2
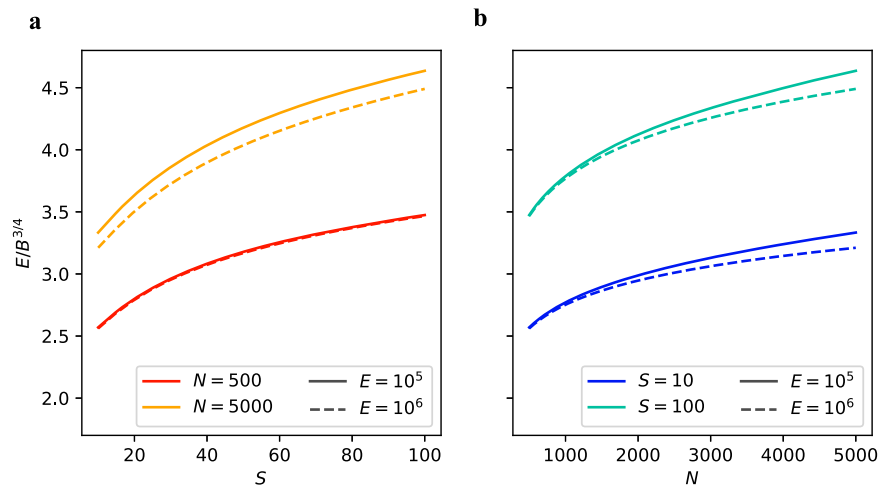
Fig. 3 The theoretical prediction for the ratio $E/B^{3/4}$ as a function of S and N. The biomass B is predicted by holding E fixed along with one other state variable. In **a** N is fixed and S is varied, and in **b** S is fixed and N is varied. The fixed values are chosen to be roughly consistent within a range of the data considered. The color of the lines represents the corresponding fixed value of N or S, while the solid and dashed lines represent different fixed values of E.

to highlight the role of diversity and the other state variables in determining this quantity:

$$E = c^{-3/4} S^{1/4} B^{3/4} \ln^{3/4}(1/\beta) \qquad (3)$$

here, as shown in Eq. 2, c is a universal constant (~ 4.17). From Eq. 3, and the fact that up to a logarithmic correction β varies as S/N, biomass exerts the strongest influence on E. With biomass fixed, productivity varies approximately as $S^{1/4}$, with a logarithmic correction. With biomass and species richness fixed, the dependence of productivity on abundance is logarithmic. Thus, the dependence of productivity on abundance is weakest, on biomass is strongest, and on species richness is of intermediate strength. Testing these predictions requires finding communities with the same values of pairs of state variables, and thus it is most feasible to compile further tests of Eq. 1 by allowing all four variables to vary as in Figs. 1 and 2. If $N \gg S$, then β is a function of N/S and we can re-express Eq. 3 entirely in terms of three ratios, E/S, B/S, and N/S, rather than the four state variables, and obtain:

$$E/S = c^{-\frac{3}{4}} \left(\frac{B}{S}\right)^{\frac{3}{4}} \ln^{\frac{3}{4}}\left(1/\beta\left(\frac{N}{S}\right)\right) \qquad (4)$$

To date, empirical cross-ecosystem surveys of relationships among community metabolism, biomass, species richness, and abundance have largely focused on uncontrolled pairwise relationships as, for example[1], between productivity and biomass but without controlling for species richness and abundance, and found considerably more scatter in the relationship than is shown in Fig. 1. In Supplementary Note 1, pairwise comparisons among state variables are shown for the datasets used here, illustrating the absence of strong relationships among pairs of state variables when the remaining pair of state variables is unconstrained. The largest R^2 value in all comparisons is 0.987, between $\ln(B)$ and $\ln(E)$, which is expected as they are computed from the same data. Noting $(1 - 0.994)/(1 - 0.987) \approx 0.46$, the unexplained variance in the observed $\ln(B)$ using the equation of state is less than half that of the second-best predictor, $\ln(E)$. All other tested relationships shown in Supplementary Note 1 were significantly worse.

Discussion

Far more testing of the ecological equation of state is warranted. Moreover, Eq. 1 raises many new questions. As with other

empirical successes of METE, it is unclear why an apparently mechanism-free theory should work at all in ecology. Our tentative answer is that in ecosystems in which the state variables are relatively constant in time, the myriad of mechanisms and traits, differing from organism to organism and species to species, confer upon them all sufficient fitness to co-exist[14,26]. In other words, just as in an ideal gas where interactions among molecules can be ignored in thermodynamic equilibrium except under extreme conditions, the mechanisms operating in ecosystems can be ignored because they simply represent different tactics for achieving the common goal of sufficient fitness.

Pursuing that idea, we might more generally expect the predictions of a top-down MaxEnt approach to fail under ecological disturbances that sufficiently alter the fitness landscape, resulting in dynamic, not static, state variables. This indeed appears to be the case for METE's prediction of the species-abundance distribution (SAD) and the species-area relationship (SAR), as discussed elsewhere[14,23,27–29]. An implication of this is that a survey of highly disturbed ecosystems with rapidly-changing state variables might reveal significant deviations from Eq. 1. As in thermodynamics, where a failure of the ideal gas law under extreme values of the state variables revealed the existence of van der Waals forces between polar molecules, so different types of failure of Eq. 1 in ecosystems with rapidly changing state variables might shed light on the causes of disturbance[23].

Applying those insights to our equation of state, four data points stand out as outliers in Fig. 1 and especially Fig. 2, These are the four temperate tree communities (blue triangles). One of the common characteristics of these four sites is that they are currently undergoing secondary succession processes from disturbances over the past decades to centuries[30–33]. Moreover, they stand out from the other sites of their relatively small S-to-N ratio (Supplementary Note 3). A more thorough survey of the validity of Eq. 1 across a spectrum of levels and kinds of natural and anthropogenic disturbance is suggested.

We note that the ecosystem area does not appear explicitly in Eq. 1; the area only enters implicitly through the area-dependence of the state variables. At least over the range of areas spanned by our datasets (50 ha/0.0064 ha $\sim 10^4$), this absence of explicit area-dependence in our equation of state is validated. However, just as some of METE's other predictions fail at large spatial scales encompassing multiple ecoregions[13,14], we expect to see significant deviations from the predicted equation of state at very large scales.

The full mass-metabolism scaling relationship in MTE[18] includes a temperature-dependent multiplicative term such that $\varepsilon \sim e^{-E_0/kT} m^{3/4}$, where E_0 is the activation energy, k is Boltzmann's constant, and T is the temperature in Kelvin. The effect of this temperature correction on the accuracy of the equation of state remains to be tested.

We have used a metabolic scaling exponent of 3/4 in all of the above, but there is both controversy over, and empirical variability in, its actual value[34]. If the scaling exponent is taken to be 2/3, as can be obtained from an energy budget model in which energy loss is proportional to the surface area of an organism[35] then the exponent of 4/3 in Eq. 1 becomes 3/2, and a different form of the equation of state is obtained. Moreover, the empirical value of E calculated from measured values of the masses of individuals will differ, and thus the empirical ratio of E to B will differ. In Supplementary Note 4 we show that the empirical validity of the equation of state is reduced if the metabolic scaling exponent is assumed to be 2/3 instead of 3/4. In particular, for the test shown in Fig. 1, the R^2 value drops from 0.994 to 0.986, and for Fig. 2, the R^2 drops from 0.600 to 0.477. On the other hand, if the metabolic scaling exponent is 1, then with our units convention, empirically, $E = B$ and Eq. 1 predicts exactly that. An example of a community in which $E \approx B$ might be a microbiome in which the masses of bacteria differ from one another by a much smaller factor than, for example, the masses of trees in a forest differ from one another. It is noteworthy that for microorganisms, there is evidence[18] that the metabolic scaling exponent is, in fact, closer to 1 than to 3/4.

Ecosystems exhibit numerous idiosyncratic phenomena, but ubiquitous patterns nevertheless exist. The latter motivates the search for general laws. We have provided evidence here for the validity of one such law, an ecological equation of state, which can be derived by combining two ecological theories, METE and MTE. Each of these theories had previously been shown to have broad explanatory power, and our result demonstrates the utility of combining theories that in combination yield more than the sum of the parts[36]. Our ecological equation of state is a simple property of complex ecosystems, and it appears to be valid across spatial scales, across types of habitat, and across different taxonomic groups. Parallel advances in our understanding of relationships among macro-scale variables in other types of complex systems, such as economies[37] or networks[38] may also be possible by combining the powerful MaxEnt inference procedure with appropriate scaling laws.

Methods

Data curation and processing. We searched several databases for data representing censused, species-level information for organisms from a well-defined area, such as a plot or single tree canopy, which included the size of individuals and abundances. We selected data sources in which at least ten species were represented due to the prediction accuracy constraints within METE[22]. Datasets were further investigated through metadata and associated publications. Any datasets that had a recent history of major natural disturbance or human alterations, such as logging and roadbuilding activities, were excluded. After identifying suitable candidate datasets, all data were then processed in the R programming language using a similar workflow.

For plant data, records were filtered to include only live individuals with size measurements. Where individual stem diameters were recorded, we grouped them by individual tree. Diameter at breast height (DBH) measurements were then combined using $\mathrm{DBH_{new}} = \sqrt{\mathrm{DBH}_1^2 + \ldots + \mathrm{DBH}_n^2}$. For plants, measurements of a size such as DBH and leaf area are considered proxies for the metabolic rates of those individuals. Plant data were processed such that the smallest individual metabolic rate measurement was rescaled to $\varepsilon_{\min} = 1$, and $B = \sum_1^N \text{metabolic rate}_{i,\text{rescaled}}^{4/3}$ (for the main manuscript) or $B = \sum_1^N \text{metabolic rate}_{i,\text{rescaled}}^{3/2}$ (Supplementary Note 4), depending on the scaling relationship being tested.

One modification was made for calculating E for the Point Reyes datasets. For both plots, only the sizes of the largest individuals (trees) were measured. Because the value of E depends almost entirely on the large individuals, we estimated the

size of the smallest individual from photographs for rescaling and employed those values for subsequent calculations.

Animal data is processed similarly to plant data, with the change that measurements of mass are directly measuring biomass, and therefore the metabolic rates are the calculated quantities. The individual with the smallest mass was therefore rescaled to $\varepsilon_{\min} = 1$; then all masses were rescaled using this convention. Rescaled masses were then summed to calculate B, and E was calculated using ¾ or 2/3 scaling, such that $E = \sum_1^N \text{mass}_{i,\text{rescaled}}^{3/4}$ or $E = \sum_1^N \text{mass}_{i,\text{rescaled}}^{2/3}$.

Statistics and reproducibility. Data were processed in R. All statistical analyses were then performed in Python using the stats package in SciPy.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data sources and permissions are detailed in Supplementary Note 3. Datasets used in these analyses are available publicly, except Traunstein Large Forest Dynamics Plot, Kellogg Biological Station, and Hawaii arthropod data, which were made available by permission of the data owners and can be requested directly from them. The processed data is available at https://doi.org/10.6084/m9.figshare.20288595.

Code availability

Code to reproduce the analyses and generate all figures in the text is available at github.com/micbru/equation_of_state/. Data cleaning code is available as R scripts on request from the authors.

References

1. Jenkins, D. G. Estimating ecological production from biomass. *Ecosphere* **6**, 1–31 (2015).
2. Currie, D. J. Energy and large-scale patterns of animal-and plant-species richness. *Am. Naturalist* **137**, 27–49 (1991).
3. O'Connor, M. I. et al. A general biodiversity–function relationship is mediated by trophic level. *Oikos* **126**, 18–31 (2017).
4. Gillman, L. N. et al. Latitude, productivity and species richness. *Glob. Ecol. Biogeogr.* **24**, 107–117 (2015).
5. Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
6. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
7. Enquist, B. J. & Niklas, K. J. Invariant scaling relations across tree-dominated communities. *Nature* **410**, 655–660 (2001).
8. Michaletz, S. T., Cheng, D., Kerkhoff, A. J. & Enquist, B. J. Convergence of terrestrial plant production across global climate gradients. *Nature* **512**, 39–43 (2014).
9. Santini, L. et al. The interface between Macroecology and Conservation: existing links and untapped opportunities. *Front. Biogeogr.* https://doi.org/10.21425/F5FBG53025 (2021).
10. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
11. Jaynes, E. T. On the rationale of maximum-entropy methods. *Proc. IEEE.* **70**, 939–952 (1982).
12. Harte, J., Zillio, T., Conlisk, E. & Smith, A. B. Maximum entropy and the state-variable approach to macroecology. *Ecology* **89**, 2700–2711 (2008).
13. Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* (Oxford Univ. Press, 2011).
14. Harte, J. & Newman, E. A. Maximum information entropy: a foundation for ecological theory. *Trends Ecol. Evol.* **29**, 384–389 (2014).
15. Harte, J., Smith, A. B. & Storch, D. Biodiversity scales from plots to biomes with a universal species-area curve. *Ecol. Lett.* **12**, 789–797 (2009).
16. Harte, J., Rominger, A. & Zhang, W. Integrating macroecological metrics and community taxonomic structure. *Ecol. Lett.* **18**, 1068–1077 (2015).
17. Hurlbert, A. & Jetz, W. More than 'more individuals': the nonequivalence of area and energy in the scaling of species richness. *Am. Naturalist* **176**, E50–E65 (2010).

18. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).

19. Blackburn, T. M. & Gaston, K. J. Linking patterns in macroecology. *J. Anim. Ecol.* **70**, 338–352 (2001).

20. Zaoli, S., Giometto, A., Maritan, A. & Rinaldo, A. Covariations in ecological scaling laws fostered by community dynamics. *Proc. Natl Acad. Sci. USA* **114**, 10672–10677 (2017).

21. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).

22. Brummer, B. & Newman, E. A. Derivations of the core functions of the maximum entropy theory of ecology. *Entropy* **21**, 712 (2019).

23. Harte, J., Umemura, K. & Brush, M. DynaMETE: a hybrid MaxEnt-plus-mechanism theory of dynamic macroecology. *Ecol. Lett.* **24**, 935–949 (2021).

24. Adler, P. B. et al. Productivity is a poor predictor of plant species richness. *Science* **333**, 1750–1753 (2011).

25. Loreau, M., Naeem, S. & Inchausti, P. (eds) *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ. Press, 2002).

26. Hubbell, S. Neutral theory and the evolution of ecological equivalence. *Ecology* **87**, 1387–1398 (2006).

27. Newman, E. A. et al. Disturbance macroecology: a comparative study of community structure metrics in a high-severity disturbance regime. *Ecosphere* **11**, e03022 (2020).

28. Newman, E. A., Harte, M. E., Lowell, N., Wilber, M. & Harte, J. Empirical tests of within-and across-species energetics in a diverse plant community. *Ecology* **95**, 2815–2825 (2014).

29. Franzman, J. et al. Shifting macroecological patterns and static theory failure in a stressed alpine plant community. *Ecosphere* **12** e03548 (2021).

30. Battles, J. J. et al. *Forest Inventory of a Northern Hardwood Forest: Watershed 6, 2017, Hubbard Brook Experimental Forest ver 1. Environmental Data Initiative.* https://doi.org/10.6073/pasta/0593ba15fb76a4f085797126a1bea3a7 (2019).

31. Bourg, N. A., McShea, W. J., Thompson, J. R., McGarvey, J. C. & Shen, X. Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot. *Ecology* **94**, 2111–2112 (2013).

32. Johnson, D. J. et al. Canopy tree density and species influence tree regeneration patterns and woody species diversity in a longleaf pine forest. *Forest Ecol. and Manag.* **490**, 119082 (2021).

33. Pretzsch, H. In *Forest Diversity and Function: Temperate and Boreal Systems* (eds Scherer-Lorenzen, M., Körner, C. & Schulze, E.-D.) Ch. 3 (Springer, 2005).

34. Glazier, D. S. A unifying explanation for diverse metabolic scaling in animals and plants. *Biol. Rev. Camb. Philos. Soc.* **85**, 111–138 (2010).

35. Dodds, P. S., Rothman, D. H. & Weitz, J. S. Re-examination of the "3/4-law" of metabolism. *J. Theor. Biol.* **209**, 9–27 (2001).

36. Marquet, P. A. et al. On theory in ecology. *Bioscience* **64**, 701–710 (2014).

37. Golan, A. *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information* (Oxford Univ. Press, 2018).

38. Williams, R. J. Simple MaxEnt models explain food web degree distributions. *Theor. Ecol.* **3**, 45–52 (2010).

39. Bradford, M. G., Murphy, H., Ford, A., Hogan, D. & Metcalfe, D. CSIRO permanent rainforest plots of North Queensland. v3. CSIRO. Data Collection. https://doi.org/10.4225/08/59475c67be7a4 (2014).

40. Hubbell, S. P. et al. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**, 554–557 (1999).

41. Condit, R. *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots* (Springer, 1998).

42. Condit, R. et al. Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. https://doi.org/10.15146/5xcp-0d46 (2019).

43. Condit, R. et al. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol.* **20**, 51–72 (2004).

44. Selhorst, D. & Brown, I. F. LBA-ECO LC-02 biophysical measurements of forests, Acre, Brazil: 1999-2002. ORNL DAAC. https://daac.ornl.gov/cgi-bin/download.pl?ds_id=1237&source=schema_org_metadata (2014).

45. Lourenço, J. Jr. et al. Soil-associated drivers of plant traits and functional composition in Atlantic Forest coastal tree communities. *Ecosphere* **12**, e03629 (2021).

46. Munger, W.& Wofsy, S. *Biomass Inventories at Harvard Forest EMS Tower Since 1993. Harvard Forest Data Archive: HF069 (v.36). Environmental Data Initiative.* https://doi.org/10.6073/pasta/5c2f17c295413da2a2a091fd7696af40 (2021).

47. Foster, D., Barker Plotkin, A. A. & Lyford, W. *Lyford Mapped Tree Plot at Harvard Forest since 1969. Harvard Forest Data Archive: HF032 (v.20). Environmental Data Initiative.* https://doi.org/10.6073/pasta/292e47940d0b0b07c3be7e9026c12c66 (2017).

48. Robertson, G. P. & Hamilton, S. K. *The Ecology of Agricultural Landscapes: Long-term Research on the Path to Sustainability* (Oxford Univ. Press, 2015).

49. Gilbert, G. S. et al. Beyond the tropics: forest structure in a temperate forest mapped plot. *J. Veg. Sci.* **21**, 388–405 (2010).

50. Gruner, D. S. Geological age, ecosystem development, and local resource constraints on arthropod community structure in the Hawaiian Islands. *Biol. J. Linn. Soc. Lond.* **90**, 551–570 (2007).

Author contributions

Conceptualization: J.H. Development: J.H., M.B., K.U., and E.A.N. Data curation and management: E.A.N. Visualization: M.B. Writing: J.H., M.B., E.A.N., and K.U.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03817-8.

Correspondence and requests for materials should be addressed to John Harte.

Peer review information *Communications Biology* thanks Pablo Marquet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Quan-Xing Liu, Luke R. Grinham, and Gene Chong. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information for

An Equation of State Unifies Diversity, Productivity, Abundance and Biomass

John Harte, Micah Brush, Erica A. Newman, Kaito Umemura

Correspondence to: jharte@berkeley.edu

This PDF file includes:

Supplementary Note 1: Pairwise correlations among state variables

While the four-variable (S, N, E, B) ecological equation of state appears to reliably describe a variety of ecosystems (Eq. 2 and Fig. 1 of main text), the question naturally arises as to whether simpler relationships among the state variables are equally successful.

Here we examine the pairwise comparisons among the observed values, or logarithms of observed values, of the four state variables and contrast the resulting regression coefficients with the regression coefficients obtained when observed species richness is compared with the value of species richness predicted by the equation of state, or when observed logarithm of biomass is compared with the value of $\ln(B)$ predicted by the equation of state.

Figures S1.a-f show the six possible pairwise comparisons between the natural logarithms of the observed values of the four state variables. Regression coefficients are summarized in Table S1.

Of the six pairwise comparisons, that of $\ln(B)$ versus $\ln(E)$ has the highest R^2 value (Fig. S1.f), which is expected because B and E are computed from the same data. We also note from Figs. S1.a-c, that comparing the three possible single-state variable predictors of $\ln(S)$, namely $\ln(N)$, $\ln(E)$ and $\ln(B)$, $\ln(N)$ has the highest explanatory power, although there is still much scatter around the regression line.
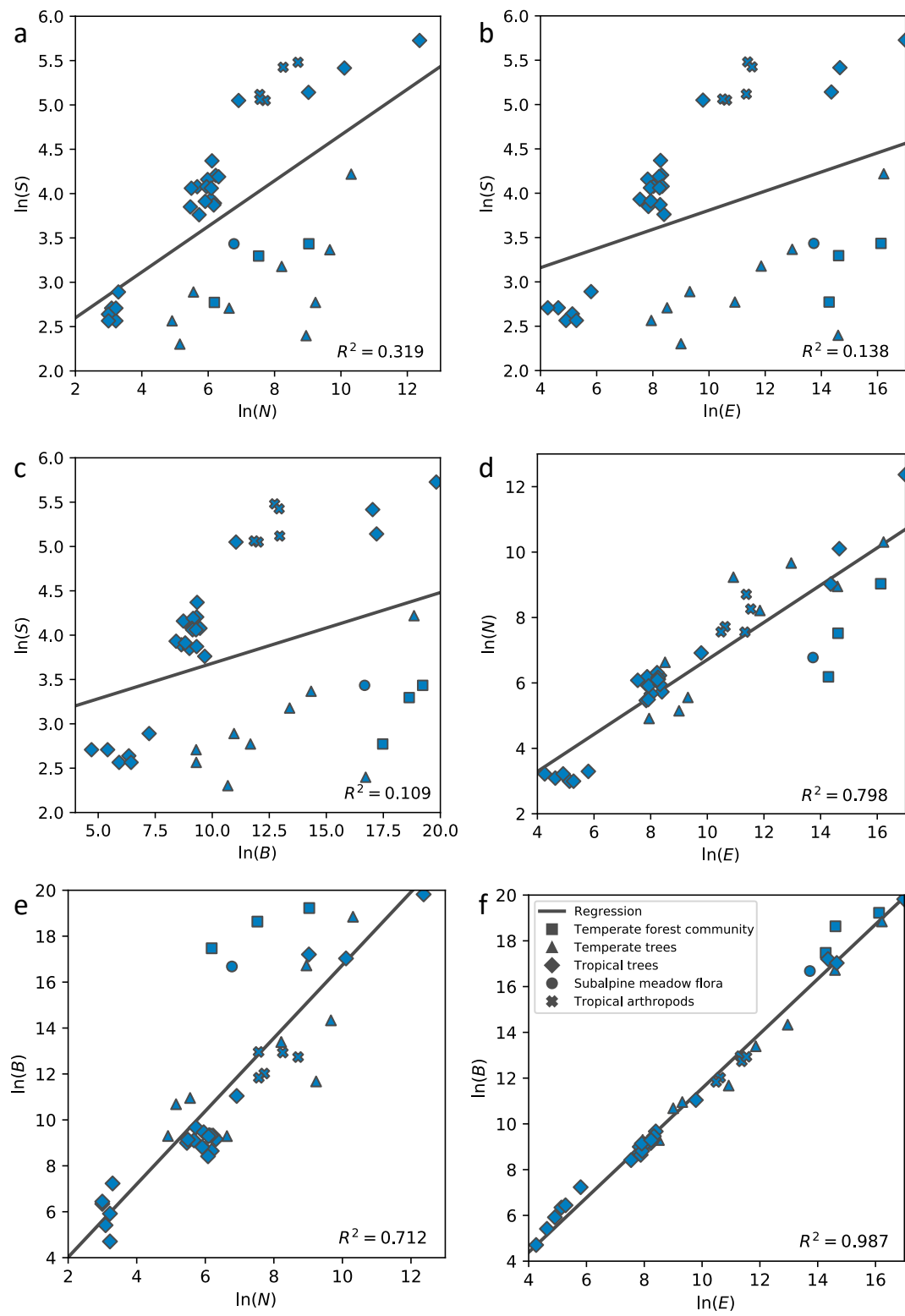
Fig. S1. The six (a, …, f) pairwise comparisons among the four state variables.

	ln(N)	ln(E)	ln(B)
ln(S)	0.319	0.138	0.109
ln(N)		0.798	0.712
ln(E)			0.987

Table S1. Summary of the R^2 values from pairwise comparisons among explanatory variables discussed above.

Supplementary Note 2. Deriving the closed form ecological equation of state

To derive a closed form for the equation of state, we need to find an analytic approximation for the equation for the biomass, $B = S \sum_n \int_\varepsilon n \ \varepsilon^{4/3} R(n, \varepsilon \mid S, N, E) \ d\varepsilon$. We use the methods from SI-E in ref. 23 in the main text to derive this approximation. In the notation used there, approximations to equations of the form $I = \sum_n \int_\varepsilon n^\nu \ \varepsilon^\sigma e^{-\lambda_1 n - \lambda_2 n\varepsilon} \ d\varepsilon$ are derived. The equation for B is of that form, with $\nu = 1$ and $\sigma = 4/3$. From Eq. E-10 in ref. S1, the leading order terms for this integral with $\sigma > \nu$ are

$$I \approx \frac{\Gamma(\sigma + 1)}{\lambda_2^{\sigma+1}} \left(\frac{1}{\sigma - \nu} + \frac{e^{-\lambda_1}}{2} + \Gamma(\nu - \sigma)\beta^{\sigma-\nu} \right) = \frac{\Gamma\left(\frac{7}{3}\right)}{\lambda_2^{\frac{7}{3}}} \left(3 + \frac{e^{-\lambda_1}}{2} + \Gamma\left(-\frac{1}{3}\right)\beta^{\frac{1}{3}} \right). \quad (S1)$$

Approximating $e^{-\lambda_1} \approx 1$ and substituting into the equation for B (Eq. 1 in main text),

$$B \approx \frac{4.17 \ S}{Z\lambda_2^{\frac{7}{3}}} \left(1 - 1.16\beta^{\frac{1}{3}} \right), \quad (S2)$$

where Z is the normalization and is approximately equal to $\ln (1/\beta)/\lambda_2$, and λ_2 is approximately S/E (ref 10 in the main text). Substituting these approximations into Eq. S2 gives

$$B \approx \frac{4.17 \ E^{\frac{4}{3}}}{S^{\frac{1}{3}} \ln \left(\frac{1}{\beta}\right)} \left(1 - 1.16\beta^{\frac{1}{3}} \right), \quad (S3)$$

which is Eq. 2 in the main text, with an additional first order correction factor of $1 - 1.16\beta^{1/3}$.

To see why the integral I gives this result, we summarize the arguments in SI-E of ref. S1 for this specific case, with $\nu = 1$ and $\sigma = 4/3$. To make this approximation, we first take the sum and the integral to go to infinity rather than N and E, and then additionally approximate the sum over

n as an integral. The first order correction for approximating a sum as an integral is half of the value of the function being summed over evaluated at the endpoints of the sum, that is $\sum_{n=1}^{\infty} f(n) \approx \int_{n=1}^{\infty} f(n)dn + \frac{f(1)+f(\infty)}{2}$. In this case, this leads to a correction term $\int_{\varepsilon} \varepsilon^{4/3} e^{-\lambda_1 - \lambda_2 e} \, d\varepsilon/2$, since the term at infinity is negligible. With these approximations,

$$I \approx \int_{n=1}^{\infty} \int_{\varepsilon=1}^{\infty} n \, \varepsilon^{\frac{4}{3}} e^{-\lambda_1 n - \lambda_2 n\varepsilon} \, dn \, d\varepsilon + \frac{e^{-\lambda_1}}{2} \int_{\varepsilon=1}^{\infty} \varepsilon^{\frac{4}{3}} e^{-\lambda_2} d\varepsilon. \tag{S4}$$

The integral over ε in both terms can be recognized as the generalized exponential integral $E_{-4/3}(\lambda_2 n)$, with $n = 1$ in the second term. The generalized exponential integral is defined as $E_p(z) = \int_1^{\infty} e^{-zt}/t^p dt$. In this case, we can take the first term from the expansion of this function using Eq. 8.19.10 from the Digital Library of Mathematical Functions (DLMF, ref S2) $(E_p(z) \approx z^{p-1}\Gamma(1-p))$. This means we can use $\int_{\varepsilon=1}^{\infty} \varepsilon^{4/3} e^{-\lambda_2 ne} \, d\varepsilon \approx \frac{\Gamma(7/3)}{(\lambda_2 n)^{7/3}}$. Note that this is only valid in this case because the exponent on ε is greater than that on n. This means that this approximation works well for large n, even though $\lambda_2 n$ is not small, because the subsequent integral over n has a factor of $n^{4/3}$ in the denominator, and since that power is larger than 1 that integral can be approximated again in the same way. Using the approximation of the integral over ε gives us

$$I \approx \frac{\Gamma\left(\frac{7}{3}\right)}{\lambda_2^{\frac{7}{3}}} \left(\int_{n=1}^{\infty} \frac{e^{-\lambda_1 n}}{n^{\frac{4}{3}}} \, dn + \frac{e^{-\lambda_1}}{2} \right). \tag{S5}$$

The integral over n is another generalized exponential integral, $E_{4/3}(\lambda_1)$, and we take the zeroth and first order terms from the same expansion as before (Eq. 8.19.10 from ref. S2). This gives us $\int_{n=1}^{\infty} \frac{e^{-\lambda_1 n}}{n^{4/3}} \, dn \approx \lambda_1^{1/3}\Gamma(-1/3) - \frac{1}{1-4/3} = 3 + \Gamma(-1/3)\lambda_1^{1/3}$. Plugging this in gives our final approximation of the integral,

$$I \approx \frac{\Gamma\left(\frac{7}{3}\right)}{\lambda_2^{\frac{7}{3}}} \left(3 + \frac{e^{-\lambda_1}}{2} + \Gamma\left(-\frac{1}{3}\right)\beta^{\frac{1}{3}} \right). \tag{S6}$$

Note that we have used β in place of λ_1 here as for most ranges of state variables $\lambda_1 \approx \beta$, but this also ensures that we can still use this approximation if $\lambda_1 < 0$. Note also that the more careful derivation in SI-E of ref. S1 obtains β rather than λ_1 here anyway. We can then plug this expression for I into the equation for B (Eq. 1 in the main text) to get an analytic approximation for B.

We then test the accuracy of this analytic expression compared to the direct numerical calculation for B for different values of the state variables. Fig. S2 shows direct comparisons of the analytical approximation to the numerical calculation for a wide range of E and N with S held fixed at 50 (see caption), and both Fig. S2**a** without the first order term (as Eq. 2 in the main text) and Fig. S2**b** with the first order correction. The contours in this plot are calculated as the negative of $\log_{10}$ of the percent difference between the analytic and numerical calculations, or $-\log_{10}\left|\frac{B_{\mathrm{approx}}-B_{\mathrm{num}}}{B_{\mathrm{num}}}\right|$, where B_{approx} is the biomass calculated using the analytic approximation, and B_{num} is calculated numerically. This means that the different contours correspond to the number of decimal places the approximation is good to. For example, the contour of level 1 corresponds to a 10% difference between the approximation and the numerical calculation, 2 corresponds to 1%, and so on. We see that for the zeroth order approximation (without the $1.16\beta^{1/3}$) our approximation is good to within less than 10% error for N/S greater than ~ 100 and E/N greater than ~ 25. With the first order correction, the approximation is good to within less than 10% for N/S greater than ~ 3 and E/N greater than ~ 5.
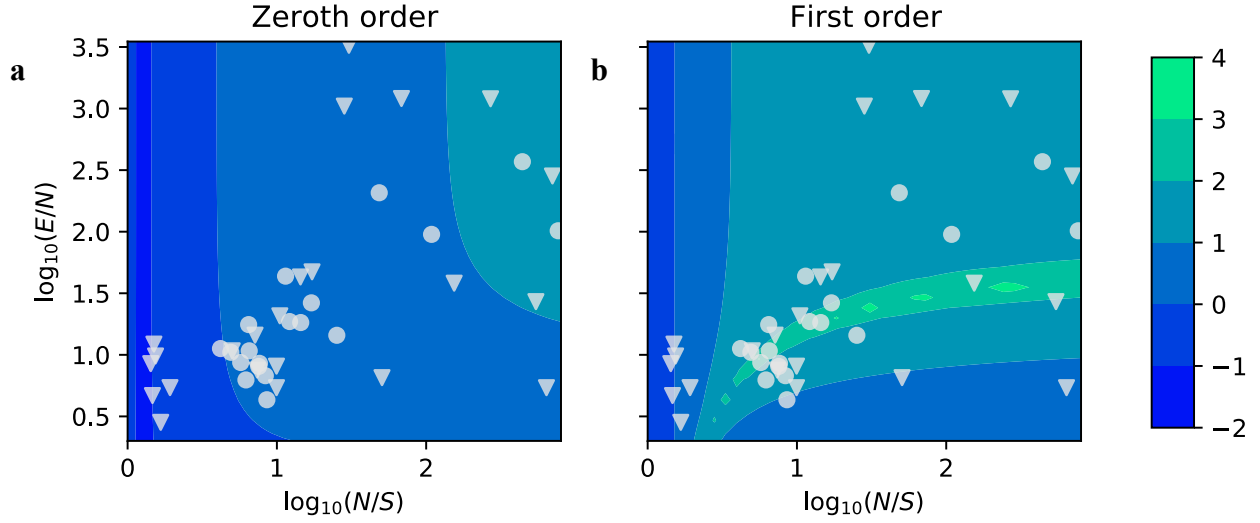
Fig. S2. Contour plots showing the accuracy of our analytic approximation for calculating biomass. The contour lines show the number of decimal places that are accurate between the zeroth order (Fig. S2**a**) or first order (Fig. S2**b**) approximations, and the exactly calculated numerical result. This is calculated as $-\log_{10}\left|\frac{B_{\text{approx}}-B_{\text{num}}}{B_{\text{num}}}\right|$, where B_{approx} is the biomass calculated using the analytic approximation, and B_{num} is calculated numerically. The axes correspond to varying E and N with S held constant at 50, and points correspond to the empirical data in Table 1. Note that S can be held constant because B/S depends only on E/S, and the constraints depend only on E/S and N/S, when S is large enough. However, for small S in this range, N can also be quite small, and we see the effects of the finite sum when calculating the constraints. This means that we must take S large enough that we do not see these effects, which in practice is about $S = 50$. We have therefore highlighted the data points where $S < 50$ by marking them with a downward triangle rather than a circle to show that the indicated accuracy of the approximation may not be correct for these points, though we note that in practice it is quite similar.

Supplementary Note 3: Data sources and extended acknowledgements

Data sources

A.

Data source and taxa	Plot	Year	Area (ha)	*S*	*N*	*E*	*B*
Tropical trees							
CSIRO permanent rainforest plots of North Queensland (Queensland, Australia) (S3)	EP3	1973	0.5	67	506	4074.955	11131.86
	EP18	1975	0.5	79	452	3915.177	11332.63
	EP19	1977	0.5	64	397	2482.458	6218.057
	EP29	1981	0.5	49	487	2634.251	5684.631
	EP30	1976	0.5	66	552	3737.412	9528.809
	EP31	1978	0.5	47	236	2544.609	8096.39
	EP32	1983	0.5	51	437	1887.393	4519.255
	EP33	1990	0.5	43	307	4435.325	15840.59
	EP34	1984	0.5	59	290	3040.91	9044.354
	EP40	1998	0.5	48	477	3859.423	11095.37
	EP41	1983	0.5	50	369	2775.138	6739.585
	EP42	1977	0.5	58	243	2732.479	9317.053
	EP43	1984	0.5	59	385	4151.287	12856.02
	EP44	1990	0.5	58	443	3779.391	10811.51
CTFS (Panama)	BCI (S4-S6)	1982	50	307	235338	23944860	405723649
	Cocoli (S7)	1997	2	171	8290	1717206	29636476
	Sherman (S7)	1997	5.96	225	24453	2325982	24937441
Catuba (Acre, Brazil) (S8)	Catuba	1999	20	156	1009	17760.1	62397.22

Data source and taxa	Plot	Year	Area (ha)	S	N	E	B
Atlantic Forest restinga (Espírito Santo, Brazil) (S9)	6L1	2015	0.00125	15	22	102.7092	224.3061
	6L3	2015	0.00125	14	20	169.6116	568.2694
	5L1	2015	0.00125	13	25	135.3977	371.2233
	5L2	2015	0.00125	13	20	195.7322	627.4894
	5L3	2015	0.00125	15	25	70.73095	110.6893
	5L5	2015	0.00125	18	27	329.9215	1383.537

B.

Data source and taxa	Plot	Year	Area (ha)	*S*	*N*	*E*	*B*
Temperate trees							
Harvard Forest (Massachusetts, USA)	EMS Tower (S10)	1993	1.26	15	759	4953.7	10880.94
	Lyford mapped tree plot (S11)	1969	2.88	24	3696	141101	655151.6
Hubbard Brook Experimental Forest (New Hampshire, USA) (S12)	Watershed 6	2017	13.23	16	10230	55324	117597.1
Kellogg Biological Station (Michigan, USA) (S13)	Deciduous Forest 1	2018	0.81 (*)	18	259	11105	57231.16
	Deciduous Forest 2	2018	0.5 (*)	10	172	8115.8	43636.65
	Deciduous Forest 3	2018	0.35 (*)	13	136	2820.1	10936.14
SCBI Large Forest Dynamics Plot (Virginia, USA) (S14)	Front Royal	2012	25.6	68	29986	11102334	153035741
Ordway Swisher Forest Dynamics Plot (Florida, USA) (S15)	unburned	2019	17.6	11	7714	2185400	18386158
Traunstein Forest Dynamics Plot (Germany) (S16)	CTFS	2016	25	29	15758	425994	1679991

C.

Data source and taxa	Plot	Year	Area (ha)	*S*	*N*	*E*	*B*
Temperate forest communities							
UC Santa Cruz (California, USA) (S17)	FERC	2006	6	31	8370	10046216	223493837
Point Reyes National Seashore Bishop Pines (California, USA) (S18)	Mt. Vision	2012	0.0256	27	1844	2223879 (*)	123940275
	Bayview	2012	0.0256	16	486	1585384 (*)	38793345
Subalpine meadow vascular plants							
Rocky Mountain Biological Laboratory (Colorado, USA) (S19)	Belleview	2012	0.0064	31	877	917872	17532137

D.

Site	Plot	Year	*S*	*N*	*B*	*E*
Tropical island arthropods Hawaii, USA (S20)	Volcano (200 y)	1997	167	1909	424260.2	83121.76
	Lanai (1500 y)	1997	156	2253	165838.1	41210.11
	Kohala (150 ky)	1997	240	6048	339471.8	87168.36
	Molokai (1.2 My)	1997	227	3865	413508.2	102189.2
	Kauai (4My)	1997	158	1922	137916.7	35818.52

Table S2. Sources of data and empirical values of state variables for the data used to test the Ecological Equation of State. Tables show values associated with Tropical trees (A), Temperate trees (B), Vascular plant communities (C), and Tropical island arthropods (D). "Year" denotes year the survey was started. Area is reported in hectares (ha). For Hawaii arthropods, the

geologic age of each site is given in parentheses and sampling was carried out by fumigating tree canopies of varying areas. Plant size data were converted to metabolic rates and rescaled such that $\varepsilon_{min} = 1$; rescaled metabolic rate measurements of individuals were summed to calculate E; and $B = \sum_1^N metabolic\ rate_{i,rescaled}^{4/3}$. For animal data, mass measurements directly measure biomass, and therefore the metabolic rates are the calculated quantities. The individual with the smallest mass was therefore rescaled to $\varepsilon_{min} = 1$; and rescaled masses were then summed to calculate B, and E was calculated as $E = \sum_1^N mass_{i,rescaled}^{3/4}$. The symbol (*) indicates an estimated value. All reference numbers in the Table refer to the Main Text reference list.

Supplementary Note 4: Sensitivity of the accuracy of the equation of state to the metabolic scaling law

In the main text, the ecological equation of state is derived under the assumption that the metabolic rate of individual organisms scale as mass to the 3/4 power. Although there is considerable evidence for such a relationship, alternative scaling exponents ranging from as low as 2/3 to 1 or higher have been proposed, depending upon type of taxonomic group and age of a cohort. The value 2/3 is motivated by a simple surface-to-volume argument.

Here we investigate the sensitivity of the equation of state to the choice of scaling exponent. As in Supplementary Note 1, we let σ be the inverse of the scaling exponent and then using $B = S \sum_n \int d\varepsilon \, n\varepsilon^\sigma R(n, \varepsilon|S, N, E)$, for $\sigma > 1$ we derive:

$$B \approx \frac{\Gamma(\sigma + 2)}{2(\sigma - 1)} \frac{E^\sigma}{S^{\sigma-1} \ln\left(\frac{1}{\beta}\right)} \qquad (\sigma > 1) \qquad \text{(S7)}$$

where β is calculated from $\beta\ln(1/\beta) \approx S/N$. For the special case of $\sigma = 1$ we derive:

$$B = E \qquad (\sigma = 1). \qquad \text{(S8)}$$

That $B = E$ if $\sigma = 1$ can be seen from the fact that one of the METE constraint conditions is $E = S \sum_n \int d\varepsilon \, n\varepsilon \, R(n, \varepsilon|S, N, E)$.

Because both mass and metabolic rate of all individuals are rarely measured, comparison of the above predictions with empirical data, requires estimating observed metabolic rate from measured mass, or vice versa. If masses are measured, then metabolic rates are calculated from $\varepsilon \sim m^{1/\sigma}$, while if metabolic rate is measured, then mass is computed using $m \sim \varepsilon^\sigma$. For trees, we will continue to assume that metabolic rate scales isometrically with basal area.

We first observe that if $\sigma = 1$ is assumed, then the equation of state, $B = E$, is an identity and the observed B will always equal the predicted B. To see this, assume that metabolic rate is measured and the empirical mass of each individual is calculated, using $\sigma = 1$, to be equal to the metabolic rate. Or equivalently, if mass is measured, then the empirical metabolic rate of each individual is calculated to equal its mass. In both cases, $B = E$. Parenthetically, we note that another trivial

result, this time holding for any σ, is that if all individuals have the same measured metabolic rate, then in our units the metabolic rate values are 1 for all individuals, resulting again in a $B = E$ equation of state.

We next compare the validity of the equation of state for the two scaling exponents 2/3 and 3/4. Comparing the same style figures in the main text, the prediction of B value by the 3/4 law is more accurate than the 2/3 law (Fig. S3 and Fig. 1). We also note that the 3/4 law predicts more of the variance in the ratio of $E : B^{1/\sigma}$ than does the 2/3 law (Fig. S4 and Fig. 2). In both results, the R^2 values of the simple regression of the 2/3 case are smaller than those of the 3/4 case.

Although the equation of state becomes an identity if $\sigma = 1$, it was not assured that the equation of state with a 3/4 scaling rule would outperform that with a 2/3 scaling rule under the assumption that our METE starting point is correct. If the true scaling exponent was exactly 2/3, then the equation of state derived under that condition might have outperformed the 3/4 result. For the data sets considered here, and under the assumption that the METE structure function, R, is valid, it thus appears more plausible to choose 3/4, rather than 2/3, metabolic scaling of biomass for both practical and empirical reasons: it predicts an equation of state relationship among the four state variables with high accuracy and is consistent with a model of the physiological features of the vascular system of plants and animals.
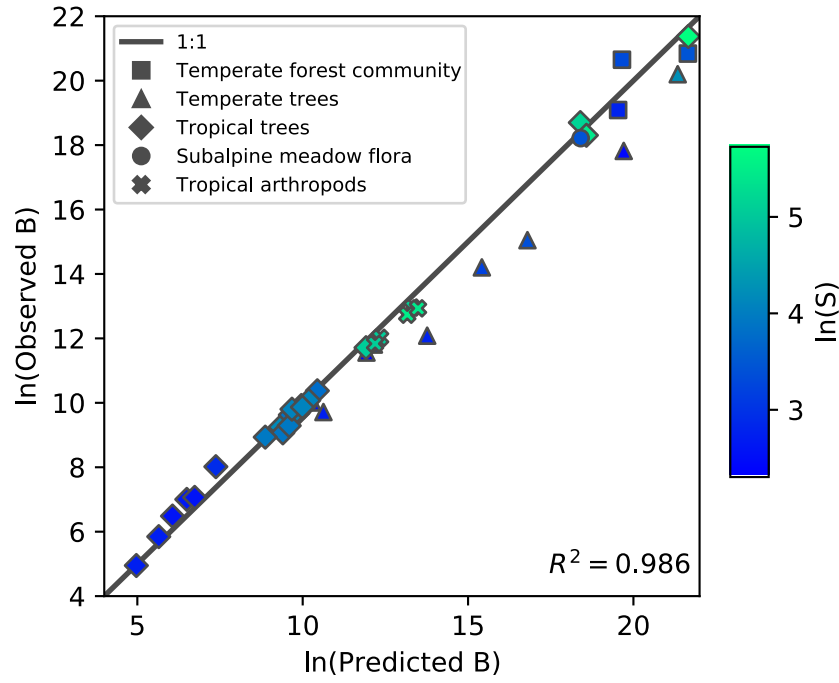
Fig. S3. Comparing observed and predicted values of *B* of each site. The same style plots as Fig. 1 in the main text is shown but here a different scaling relationship is used between metabolic rate $\boldsymbol{\varepsilon}$ and biomass $\boldsymbol{m}$ of an individual, $\boldsymbol{\varepsilon \sim m^{2/3}}$, namely $\boldsymbol{\sigma = 3/2}$. Lighter color corresponds to higher species richness.
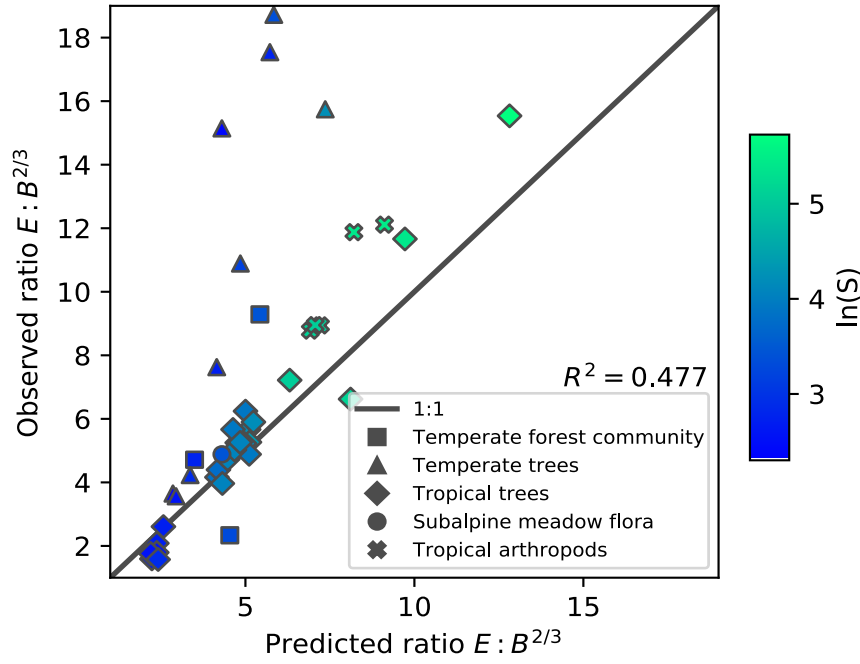
Fig. S4. Comparing observed and predicted values of the ratio of *E* and $B^{1/\sigma}$ of each site.
The same style plots as Fig. 2 in the main text is shown but here a different scaling relationship is used between metabolic rate ε and biomass m of an individual, $\varepsilon \sim m^{2/3}$, namely $\sigma = 3/2$. Lighter color corresponds to higher species richness.

Supplementary References

S1. Harte, J., Umemura, K. Brush, M. DynaMETE: a hybrid MaxEnt-plus-mechanism theory of dynamic macroecology. *Ecol. Lett*. **24**, 935–949 (2021).

S2. Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., Saunders, B. V., Cohl, H. S., McClain, M. A., eds., NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/8.19.E25, Release 1.0.28 of 2020-09-15.

S3. Bradford, M. G., Murphy, H., Ford, A., Hogan, D., Metcalfe, D. CSIRO Permanent Rainforest Plots of North Queensland. v3. (2014). CSIRO. Data Collection. https://doi.org/10.4225/08/59475c67be7a4

S4. Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., Wright, S. J., de Lao S. L. Light-Gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**, 554–557 (1999).

S5. Condit, R. Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots. Springer Science & Business Media (1998).

S6. Condit, R. Pérez, R., Aguilar, S., Lao, S., Foster, R., Hubbell, S. Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. URL https://doi.org/10.15146/5xcp-0d46 (2019).

S7. Condit, R., Aguilar, S., Hernandez, A., Perez, R., Lao, S., Angehr, G., Hubbell, S. P., Foster, R. B. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol*. **20**, 51–72 (2004).

S8. Selhorst, D., Brown, I. F.. LBA-ECO LC-02 Biophysical Measurements of Forests, Acre, Brazil: 1999-2002. ORNL DAAC (2014) (https://daac.ornl.gov/cgi-bin/download.pl?ds_id=1237&source=schema_org_metadata).

S9. Lourenço Jr. J., Newman E. A., Ventura J. A., Milanez C. R., Thomaz L. D., Wandekoken D. T., Enquist B. J. Soil-associated drivers of plant traits and functional composition in Atlantic Forest coastal tree communities. *Ecosphere* **12(7)**: e03629 (2021).

S10. Munger, W., Wofsy, S. Biomass Inventories at Harvard Forest EMS Tower since 1993. Harvard Forest Data Archive: HF069 (v.36). (2021). Environmental Data Initiative: https://doi.org/10.6073/pasta/5c2f17c295413da2a2a091fd7696af40.

S11. Foster, D., Barker Plotkin, A. A., Lyford, W. Lyford Mapped Tree Plot at Harvard Forest since 1969. Harvard Forest Data Archive: HF032 (v.20). (2017). Environmental Data Initiative: https://doi.org/10.6073/pasta/292e47940d0b0b07c3be7e9026c12c66.

S12. Battles, J. J., Cleavitt, N., Johnson, C., Hamburg, S., Fahey, T., Driscoll, C., Likens, G. Forest Inventory of a Northern Hardwood Forest: Watershed 6, 2017, Hubbard Brook Experimental Forest ver 1. Environmental Data Initiative (2019). (https://doi.org/10.6073/pasta/0593ba15fb76a4f085797126a1bea3a7 (Accessed 2021-10-02).

S13. Robertson, G. P., Hamilton, S. K. Long-term ecological research at the Kellogg Biological Station LTER site. The ecology of agricultural landscapes: Long-term research on the path to sustainability, 1–32 (2015).

S14. Bourg, N. A., McShea, W. J., Thompson, J. R., McGarvey, J. C., Shen, X. Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot. *Ecology*. **94**, 2111–2112 (2013).

S15. Johnson, D. J., Magee, L., Pandit, K., Bourdon, J., Broadbent, E., Glenn, K., Kaddoura, Y., Machado, S., Nieves, J., Wilkinson, B., Zambrano, A., Bohlman, S. Canopy tree density and species influence tree regeneration patterns and woody species diversity in a longleaf pine forest. *Forest Ecol. and Manag.* **490**, 119082 ISSN 0378-1127 (2021).

S16. Pretzsch, H. "Diversity and productivity in forests: Evidence from long-term experimental plots", pp. 41-64 in *Forest Diversity and Function: Temperate and Boreal Systems*, Scherer-Lorenzen, M. ,Körner, C., Schulze, E.-D., Eds., Springer Berlin, Heidelberg (2005).

S17. Gilbert, G. S., Howard, E., Ayala-Orozco, B., Bonilla-Moheno, M., Cummings, J., Langridge, S., Parker, I. M., Pasari, J., Schweizer, D., Swope, S. Beyond the tropics: forest structure in a temperate forest mapped plot. *J. Veg. Sci*. **21**, 388–405 (2010).

S18. Newman, E. A., Wilber, M. Q., Kopper, K. E., Moritz, M. A., Falk, D., McKenzie, D., Harte, J. Disturbance macroecology: a comparative study of community structure metrics in a high-severity disturbance regime, Ecosphere **11**(1), e03022 (2020).

S19. Newman, E. A., Harte, M. E., Lowell, N., Wilber, M., Harte, J. Empirical tests of within- and across-species energetics in a diverse plant community. *Ecology* **95**(10), 2815-2825 (2014).

S20. Gruner, D. S. Geological age, ecosystem development, and local resource constraints on arthropod community structure in the Hawaiian Islands. *Biol. J. Linn. Soc. Lond*. **90**, 551–570 (2007).