Supplementary Materials for

Forecasting the soil microbiome at a continental scale

Zoey Werbin, Colin Averill, Jennifer Bhatnagar, Michael Dietze

Corresponding author: zrwerbin@bu.edu

The PDF file includes:

Figs. S1 to S19
Tables S1 to S2
Appendices S1 to S3

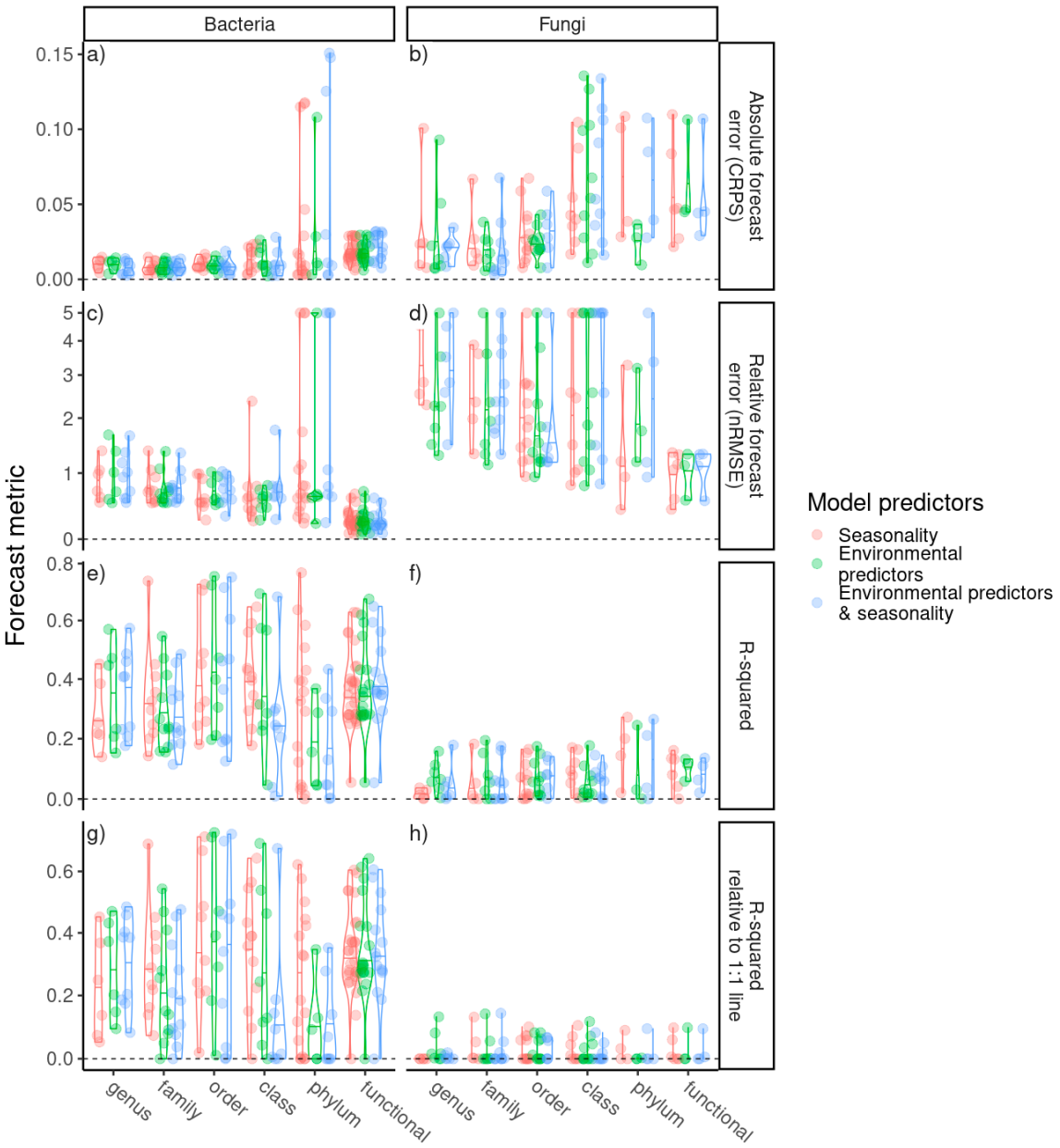


Figure S1. Forecast metrics across taxonomic ranks and functional groups of soil bacteria (left column) and fungi (right column), for three different sets of model predictors.

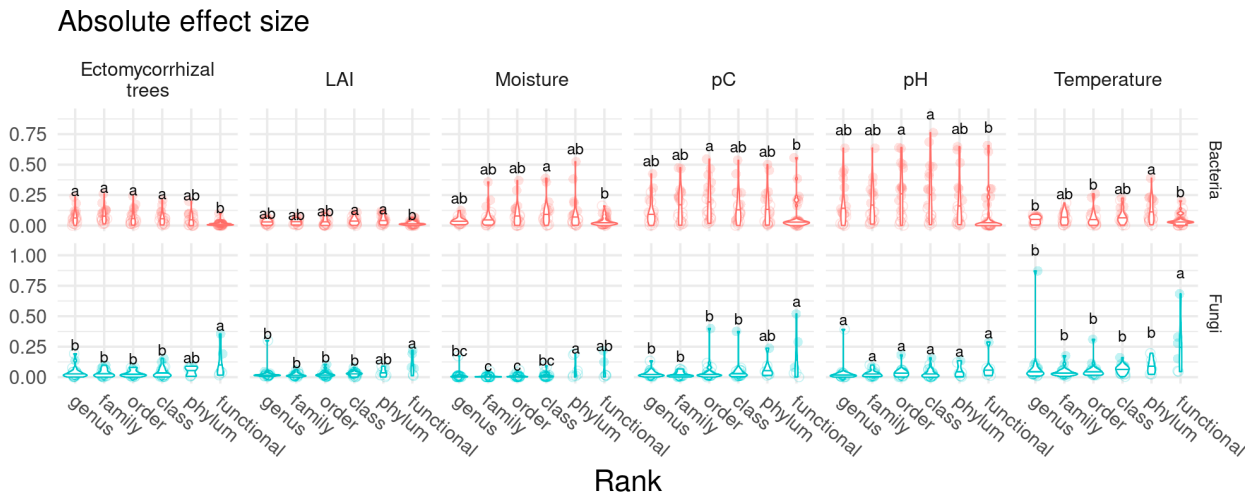
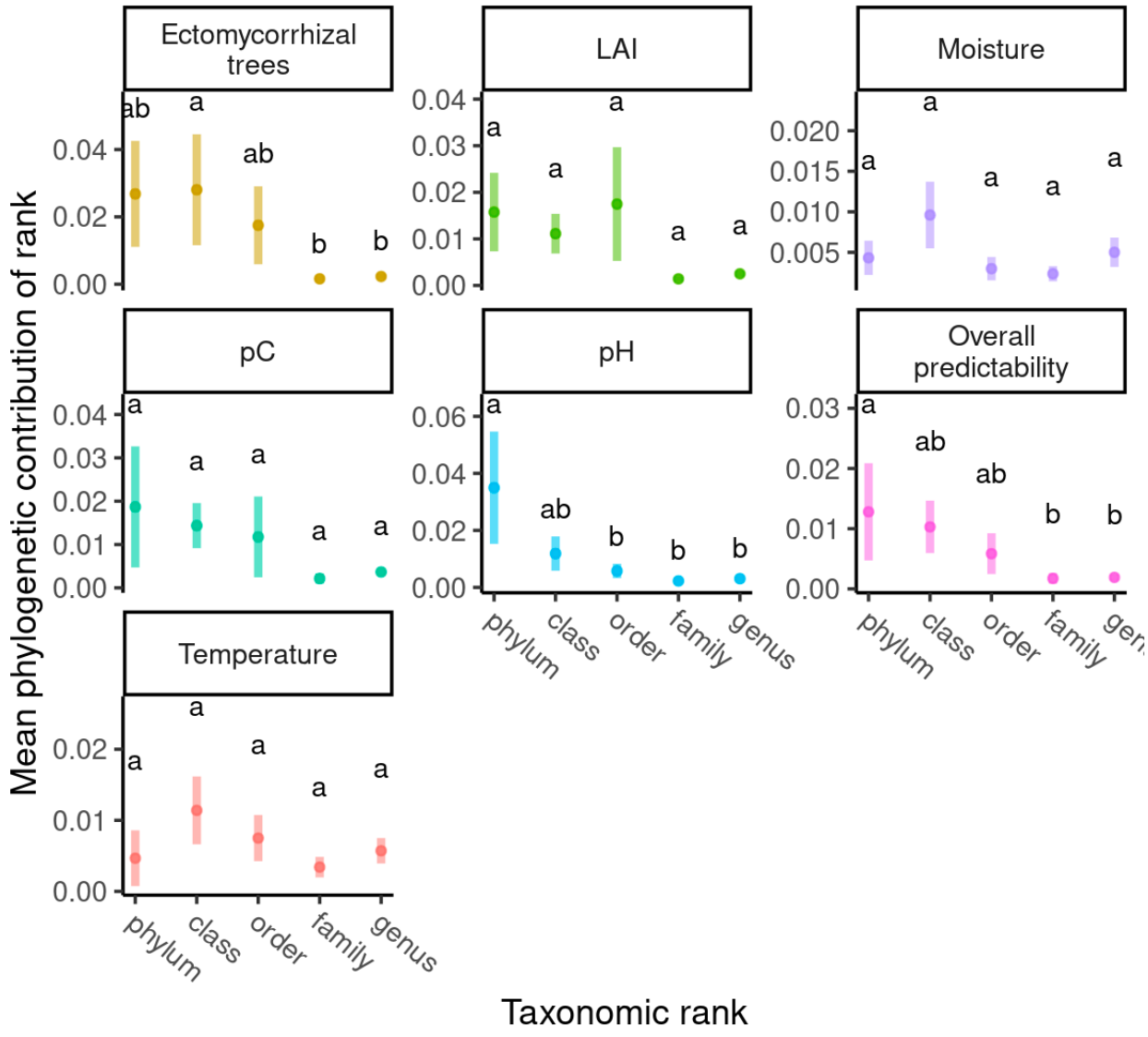


Figure S2. Absolute value of bacterial (top row) and fungal (bottom row) model estimates for taxonomic groups of soil microbes. Letters indicate distinct groups returned by Tukey's post-hoc test after ANOVA returned $p < .05$. For visualization, significance to each parameter was assigned if zero did not fall within the upper and lower 95% credible interval for the parameter.



552
553 **Figure S3.** Mean phylogenetic contribution index of taxonomic ranks to variation in one of 6
554 environmental predictors, or overall predictability (RSQ) estimated via the phylocomr package (*69*).

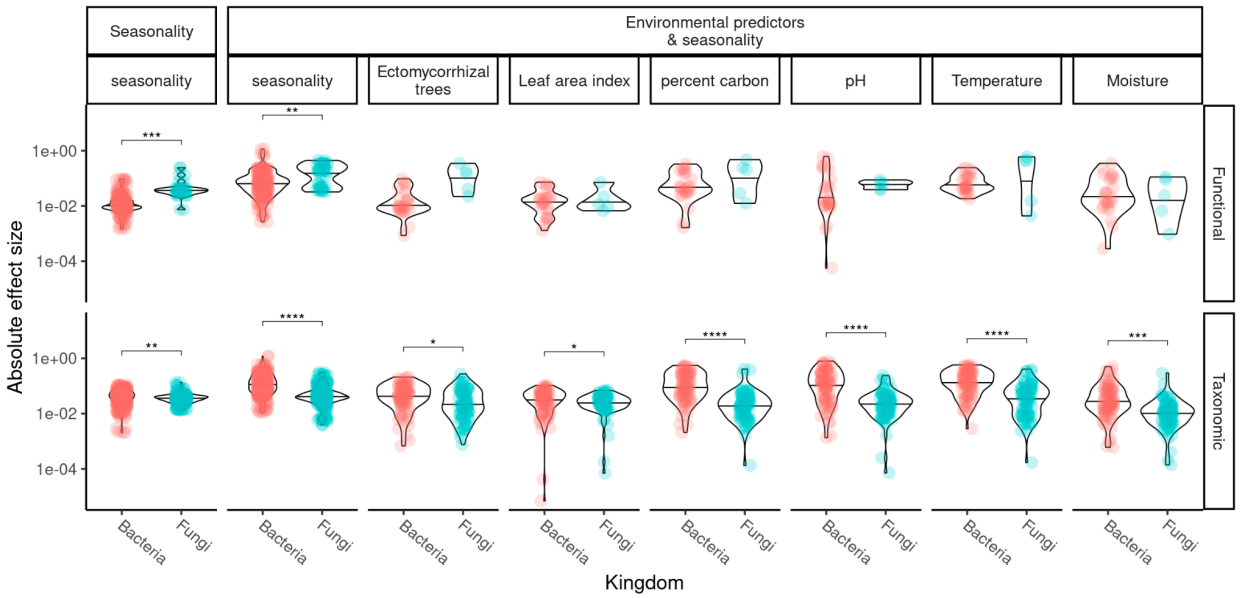


Figure S4. Violin plots showing parameter estimates for functional groups (top row) and taxonomic groups (bottom row) of soil microbes. T-tests show differences between fungi and bacteria for seasonal amplitude (estimated from seasonality-only model), seasonal amplitude (estimated from environmental predictor and seasonality model), and absolute sizes of environmental predictors, with asterisks indicating tests significant after false discovery rate p-value correction (*83*).

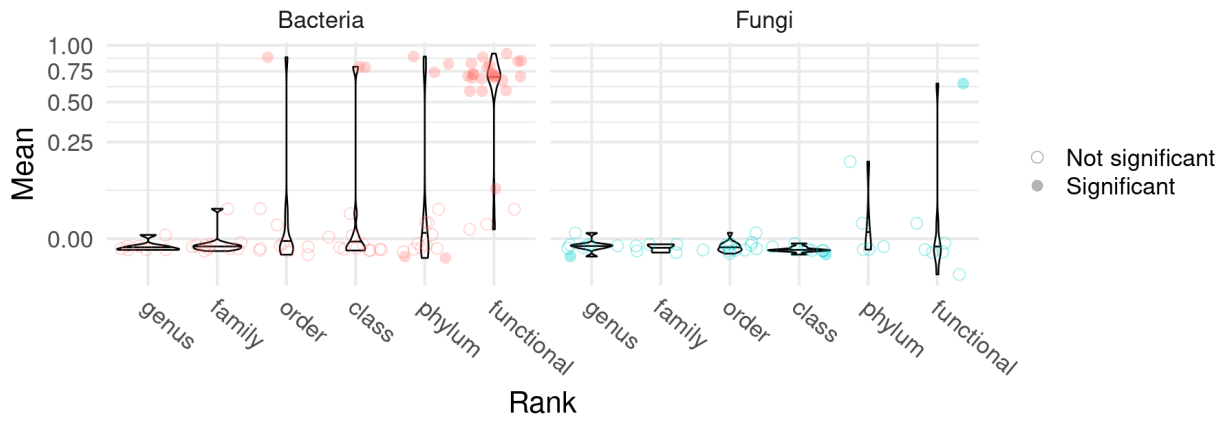


Figure S5. Mean of seasonality-only model estimates for $\boldsymbol{\rho}$ parameter, representing temporal memory (Equation 2). $\boldsymbol{\rho}$ is multiplied by abundance at previous timepoint. Significance was assigned if zero did not fall within the upper and lower 95% credible interval for the parameter.

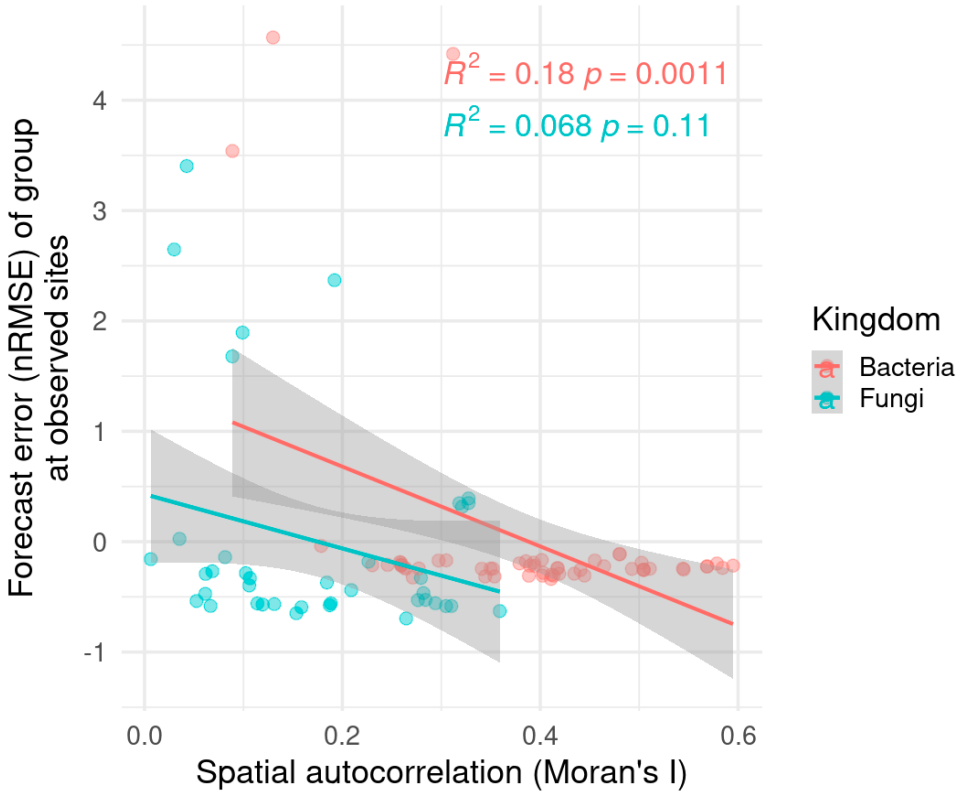


Figure S6. Scatterplot of positive relationship between forecast prediction RSQ and spatial autocorrelation, measured using Moran's I, with higher values indicating stronger clustering of abundances over space.

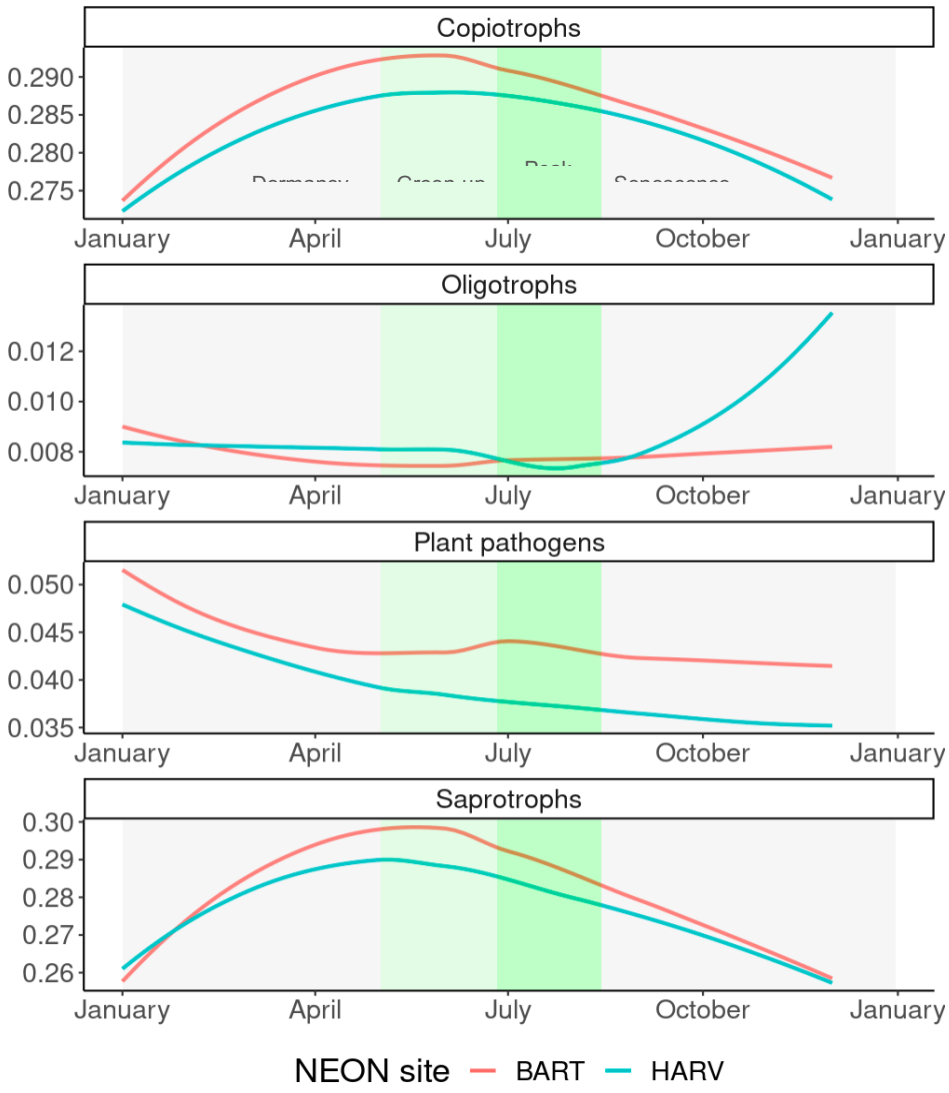


Figure S7. Seasonal trends in abundances of functional groups that are expected to be associated with plant growing seasons. Lines indicate loess curves over the modeled abundances for two deciduous New England NEON sites (Bartlett Forest and Harvard Forest) with similar phenophase transition points (dormancy, greenup, peak greenness, and senescence). Copiotrophic bacteria are thought to thrive in high nutrient environments, such as soils during the growing season when plants are exchanging simple sugars with microbes, while oligotrophic bacteria thrive in lower-nutrient environments (*64*). Plant pathogens are fungi whose dominant trophic guild is plant pathogenic, and saprotrophs are free-living fungi that consume decaying organic matter.

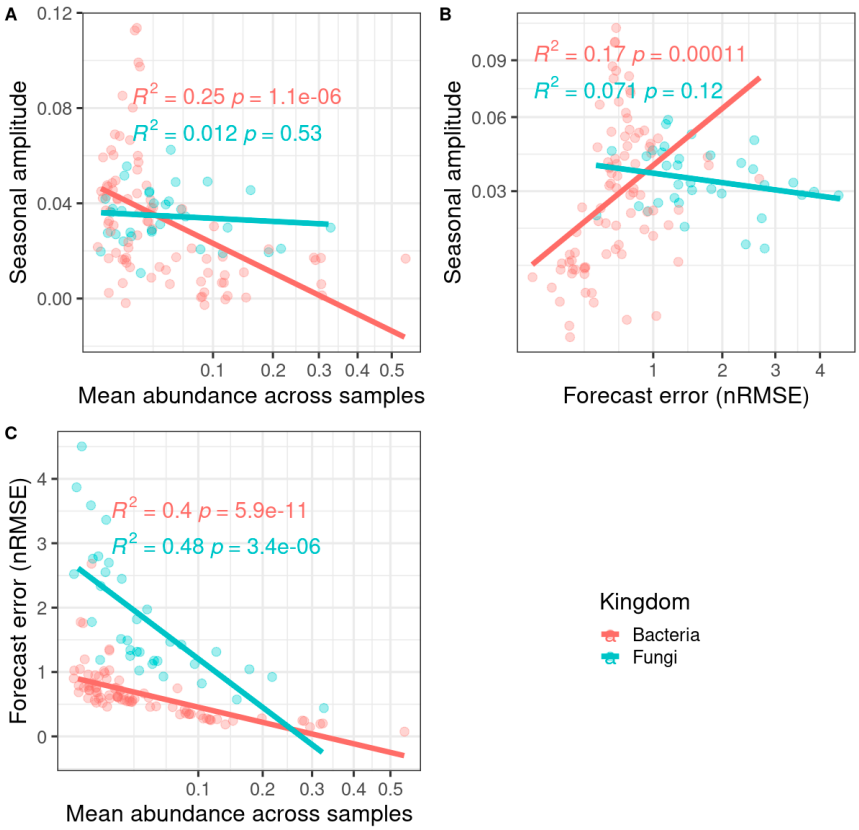


Figure S8. Scatterplot showing relationships between mean group abundance, forecast error (normalized root mean squared error, nRMSE), and seasonal amplitude, estimated from seasonality-only model.

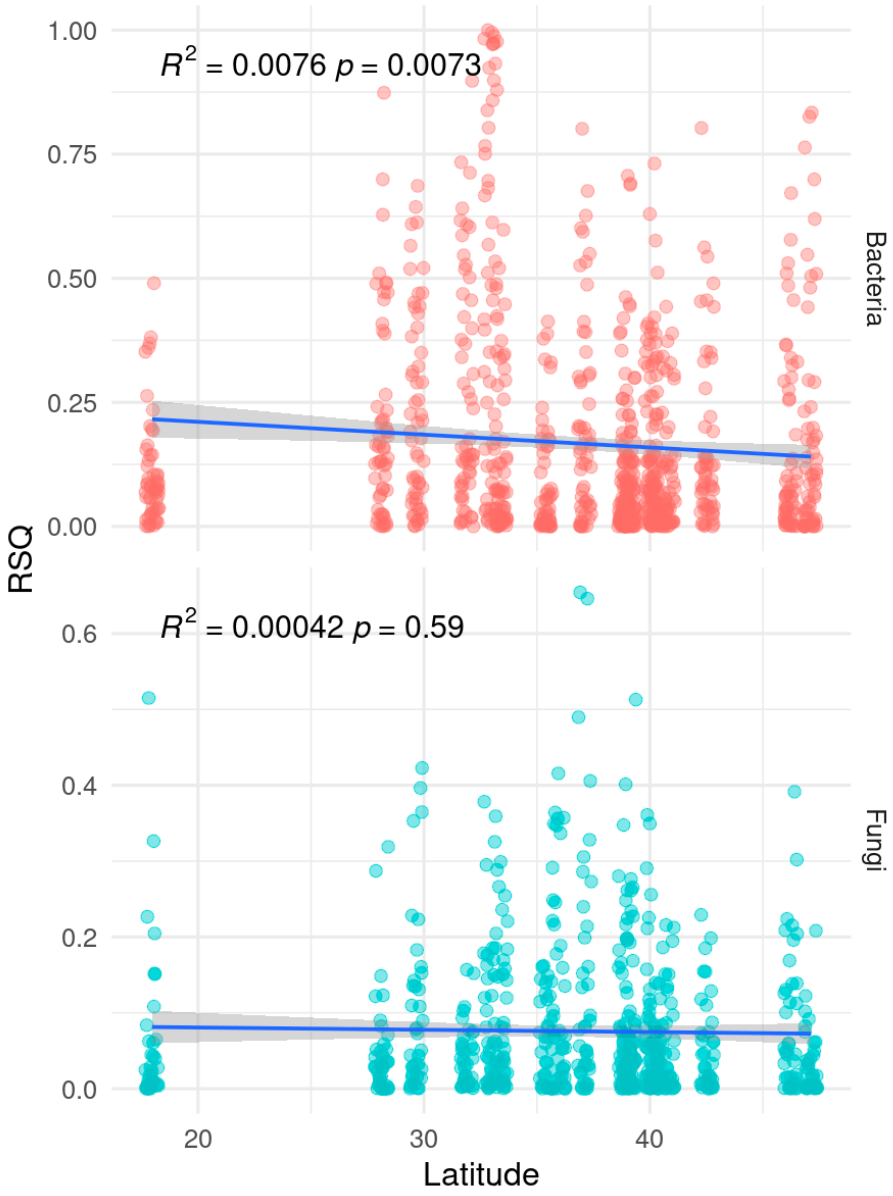


Figure S9. Scatterplot of relationship between site latitude and site forecast accuracy (R-squared of predicted and observed values) for fungi (teal) and bacteria (orange) for seasonality-only model, with similar patterns for other predictor sets.

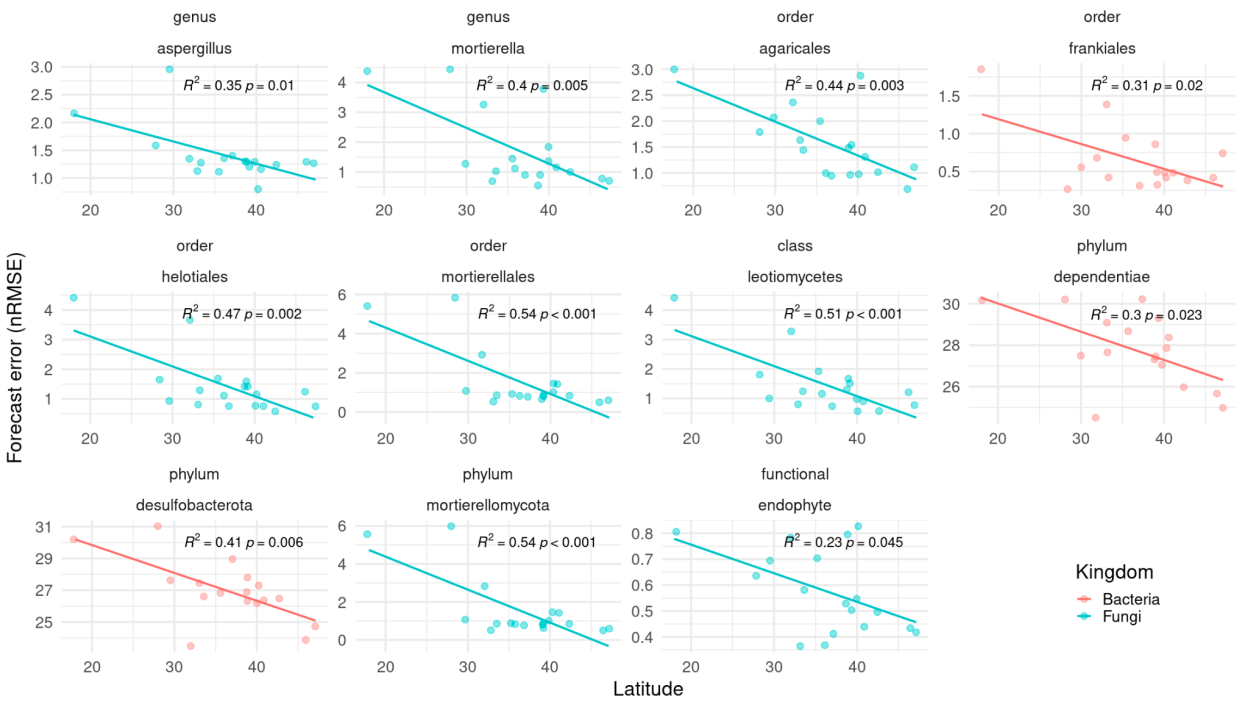


Figure S10. Forecast error (normalized root mean squared error, nRMSE) by latitude of observed sites, for groups with significant latitude-error relationships. Forecasts are from linear models with environmental predictors and seasonality.

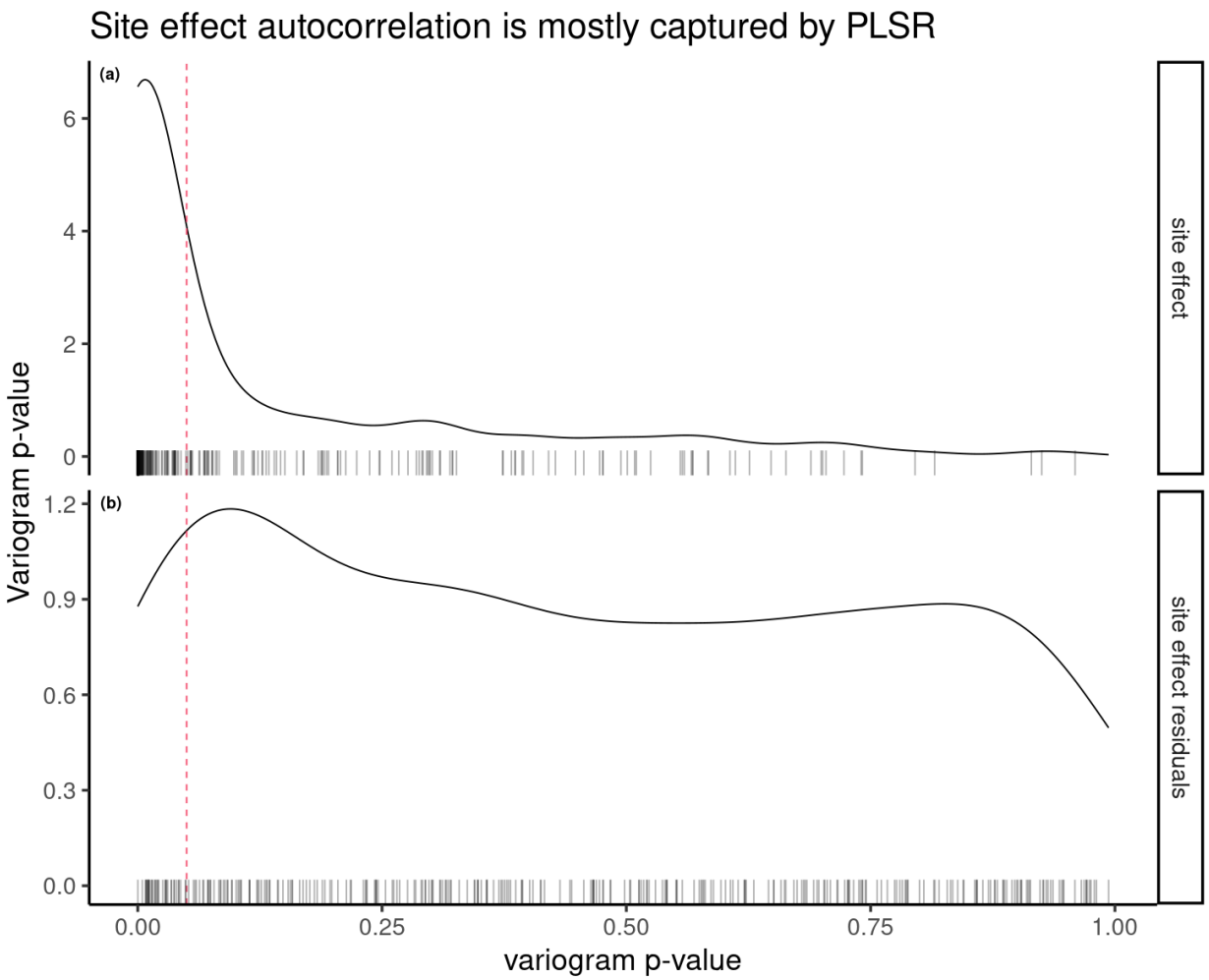


Figure S11. P-values of variograms created for the random site effects of every statistical model. While the site effects estimated by the models tend to have a spatial signal (median p-value below .05), but this spatial signal is rarely significant in the residuals of site effects modeled by partial least-squares regression (PLSR). P-values of PLSR residuals were not significantly different from a uniform distribution, except for the models without seasonality predictors, which retained a weak spatial signal. Variograms generated by the "variogram" function in the gstat R package (*84*), and significance estimated using "envelope" and "envsig" in the variosig R package (*85*).

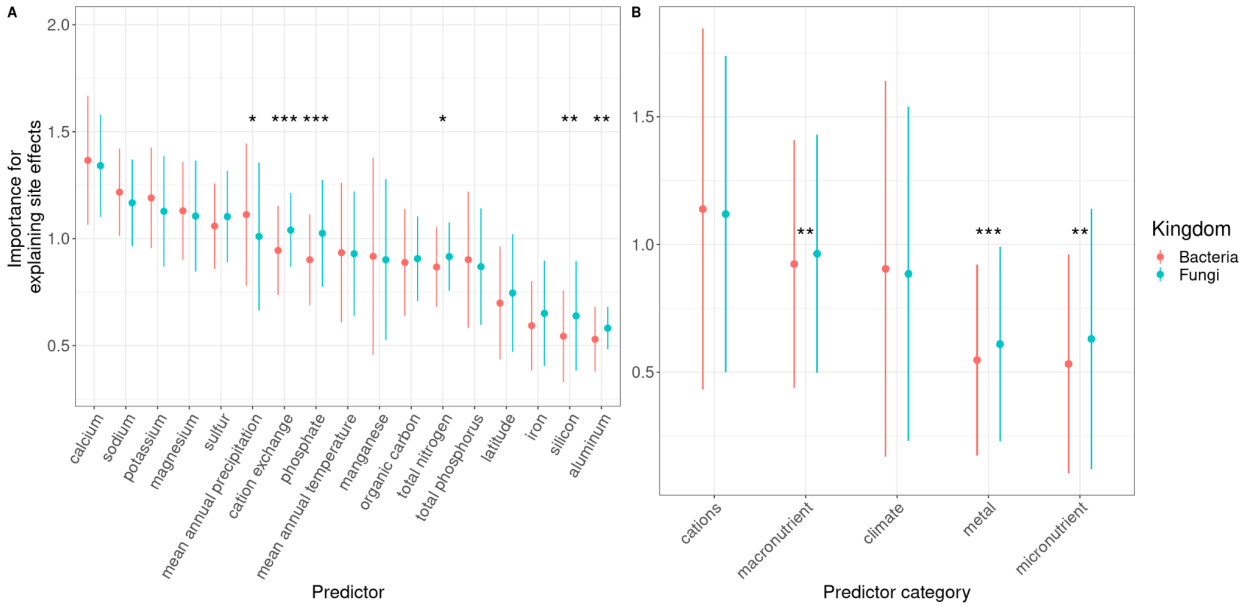


Figure S12. Importance of site characteristics for explaining the site random effects of each soil fungal (orange) or bacterial (teal) group over time. Partial least squares regression models were fit to 30 site effect estimates using the plsr function from the pls package, then importance was calculated from using the "vip" function from the spectratrait R package (*82*). Asterisks indicate significant T-tests at alpha=.05, comparing importance values for all fungal and bacterial groups.

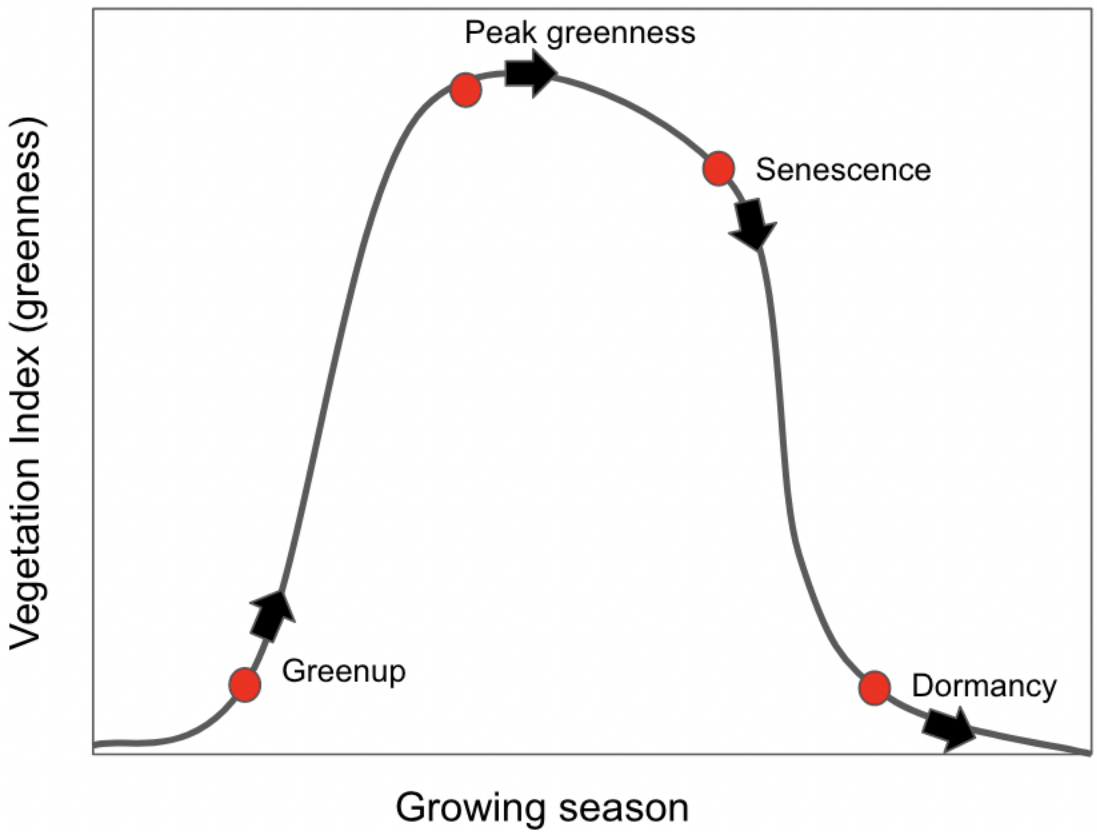


Figure S13. Simplified diagram of plant community phenophases assigned to each site and month by the MODIS Land Cover Dynamics data product (*73*). Among all 5362 site-month combinations within model calibration period, dormancy was assigned 2235 times, greenup 1115 times, peak greenness 492 times, and senescence 1520 times.

Table S1. Parameter and performance summary. Estimated parameters and performance for 673 models fit to the calibration time period of 2015-11-01 to 2018-01-01. Significant linear model predictors are in bold text. NA values represent covariates not estimated by a given model, e.g. "seasonality" models included coefficients for the sine and cosine of the sampling month, but no other environmental predictors. Sigma is the process error associated with the plot-level linear model, and core_sd is the standard deviation of samples taken from a single plot and timepoint. Rho is the temporal stability or autocorrelation between timepoints. (CSV)

Table S2. Summary of sites included in forecasting models. NEON site descriptions and number of high-quality samples per site used in calibration and validation datasets for bacteria and fungi. Sample quality filtering is described in Methods. (CSV)

Supplementary text

Appendix S1: Extended data preparation methods.

Scripts for downloading and preparing these datasets are located at https://github.com/zoey-rw/microbialForecasts. Soil moisture and soil temperature data are released with approx. 1 month lag time by NEON, while microbial data takes multiple months to process before release. Therefore, "forecasts" of microbial abundance can be made before microbial data are released, but these forecasts do not continue into the future. While soil moisture and temperature are components of many weather and agricultural forecasting systems (i.e. Syngenta's "Greencast", National Weather Service Arkansas Red Basin forecasts), there are no publicly available forecast products that span the United States. If forecasts for model covariates become available, they will allow our microbial forecast system to extend into the future.

Soil temperature

Soil temperature is measured continuously at all NEON sites via an array of five soil sensors at nine depths, and released as 30-minute averages (NEON DP1.00041.001). Soil temperatures represent point measurements, so the two sensors shallower than 30 cm were retained, to match the depths of microbial soil samples.

As described in the above section, sensor depth was chosen through comparison between sensor temperature data and temperature measurements associated with the soil sub-samples used for microbial analyses (NEON DP1.10047.001). The resulting measurements from samples and soil sensors were compared via linear regression R^2 values, and the shallowest sensor layer (0 cm deep) had the strongest relationship ($R^2 = 0.63$) and was used for soil temperature inputs.

Temperature data were gap-filled using daily air temperature data from DAYMET (*86*), which was strongly correlated with NEON soil sensor data (Figure S14).

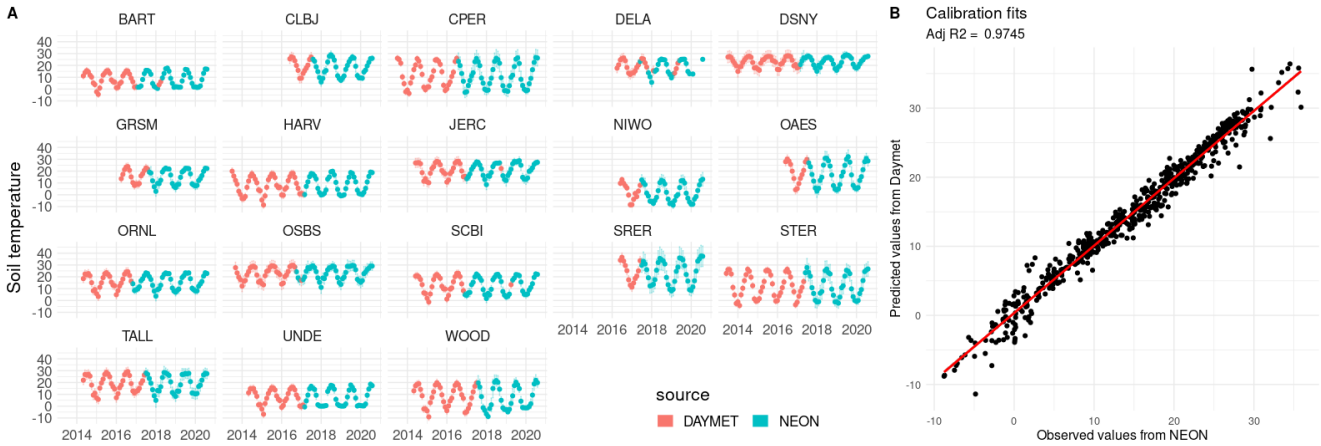


Figure S14. A) Gap-filled soil temperature time-series for NEON sites. **B)** DAYMET air temperature was modeled as a linear function of NEON soil temperature sensor data, then the regression was used to predict soil temperature at sites or monthly with missing soil temperature values. Uncertainty from sensor uncertainty and monthly averaging was propagated into monthly predictions using an errors-in-variables (EIV) approach in the NIMBLE statistical software.

<u>Soil water content (i.e. soil moisture)</u>

Volumetric soil water content (SWC) is measured continuously at all NEON sites via an array of five soil sensors at eight depths, and released as 30-minute averages (NEON DP1.00094.001). SWC sensors vertically integrate moisture from 5 cm above and below sensor midpoint, so sensors shallower than 25 cm were retained, to match the depths of microbial soil samples.

To decide which sensor depths were most relevant for microbial communities, we compared sensor data to gravimetric SWC measurements, which were associated with the soil sub-samples used for microbial analyses (NEON DP1.10047.001). To match the units of volumetric SWC sensor data, gravimetric SWC, θ_g, was converted to volumetric SWC, θ_v, by multiplying SWC by soil bulk density, *BD*:

$$\theta_v = \theta_g * BD$$

Soil bulk density was derived from each site from an initial characterization product (NEON DP1.10047.001); data were subsetted to soil horizons no deeper than 35 cm and mean bulk density was calculated for the site. The resulting moisture measurements from samples and soil sensors were compared via linear regression R^2 values, and the second shallowest sensor layer (approx. 6 cm deep at most sites) had the strongest relationship and was used for soil moisture inputs.

Moisture data were gap-filled using daily soil moisture values from soil sensor networks when available, and from satellite-derived values otherwise. Gap-filling used data from Soil Climate Analysis Network (SCAN), NASA's Soil Moisture Active Passive (SMAP) satellite, and ESA's Soil Moisture and Ocean Salinity (SMOS) satellite, in that order (*87–89*). Correlations with NEON soil sensor data varied by source (Figure S15).

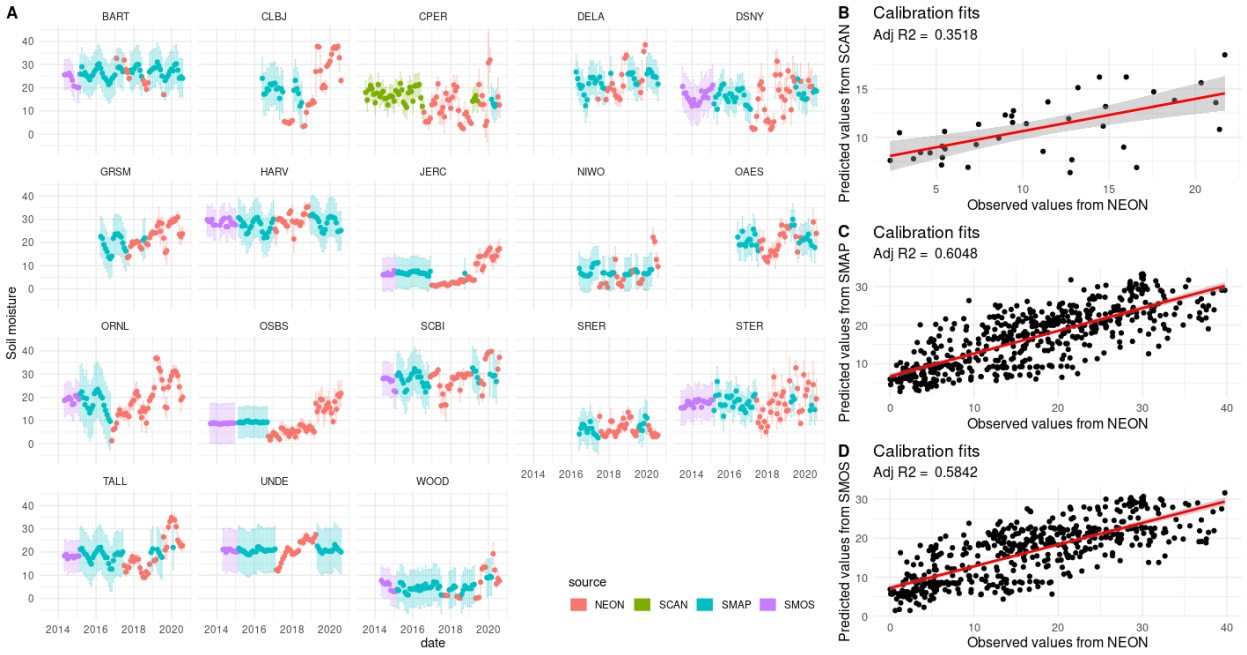


Figure S15. A) Gap-filled soil moisture time-series for NEON sites. **B-D)** Soil moisture from each source was modeled as a linear function of NEON soil moisture sensor data, then the regression was used to predict soil moisture at sites or monthly with missing soil moisture values. Uncertainty from sensor uncertainty and monthly averaging was propagated into monthly predictions using an errors-in-variables (EIV) approach in the NIMBLE statistical software.

<u>Soil moisture/temperature averaging and uncertainty propagation:</u>

NEON calculates expanded uncertainty estimates, which incorporate individual measurement uncertainty into time-averaged mean uncertainties for 30-minute means (described in (*90*, *91*)). Because we averaged data by the month, we scaled up the uncertainties associated with each measurement (reported by NEON as "Expanded Uncertainty" and "Standard Error") by sampling random biases and errors from a month's 30-minute measurements to add to each month's mean values, resulting in a site-level time-series with varying amounts of uncertainty depending on data source and reported measurement uncertainty.

<u>Soil pH and percent carbon</u>

Many soil cores used for microbial samples had associated measurements of soil pH and percent carbon reported in DP1.10086.001 (*92*). For forecasts, however, we do not expect to know where in a plot the soil cores will come from; therefore, plot-level mean values were constructed using

all previous soil pH and percent carbon measurements, and these plot-level values were used for prediction, with site-level values used to gap-fill any missing plot-level values. One site with no soil chemistry measurements from any year was excluded from models (ABBY).

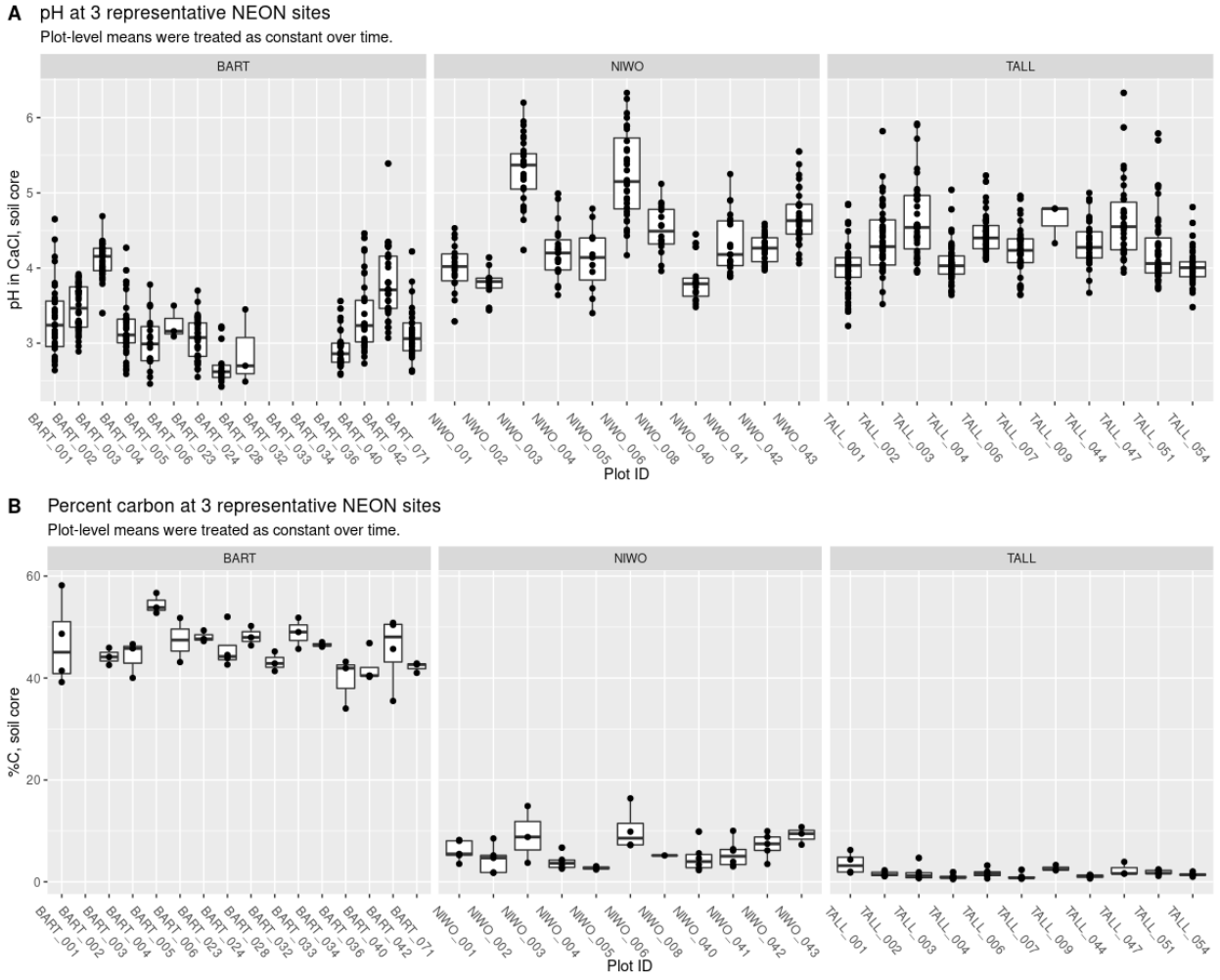


Figure S16. Variation across sampling plots and timepoints, within 3 example NEON sites, for observed measurements of A) soil pH and B) soil percent carbon. Points indicate individual measurements that were aggregated at the plot level.

Ectomycorrhizal basal area percent

Percent basal area of ectomycorrhizal trees was estimated using stem diameter from the NEON woody plant vegetation structure (DP1.10098.001) (*93*)), with ectomycorrhizal status assigned to individual trees using species names. After accounting for insect-damaged and fallen trees, basal area of trees was scaled to the plot level and values were averaged over all years of measurements. The following site identifiers had a land cover type of "Cultivated Crops",

"Shrub/Scrub", or "Grassland/Herbaceous" and consequently had no ectomycorrhizal tree abundances: BARR, CPER, DCFS, JORN, KONA, OAES, STER, TOOL, and WOOD.

Leaf area index (LAI)

To calculate the monthly LAI for each site, we first use the site latitude and longitude as input to the call_modis function from the PecanTools R package (*94*) to download the satellite-based MOD15A2H Version 6 Moderate Resolution Imaging Spectroradiometer (MODIS) combined Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (FPAR) product (*95*). LAI values were then averaged from each site and month from the 8-day composite values.

Appendix S2: Determination of calibration time period

Provisional data from NEON was available to extend the time series to additional years, but inclusion of these data led to higher forecast bias. This had a known cause for bacterial sequences (a methodological shift in 16S amplicon sequencing primers (*17*)) but the bias could not be explained for fungi. After evaluating the hindcasts, we increased the model calibration period to include all available data (spanning 2018-2020, but only for a subset of sites), which increased precision in environmental predictor covariates and reduced overall confidence intervals (Figure S17).

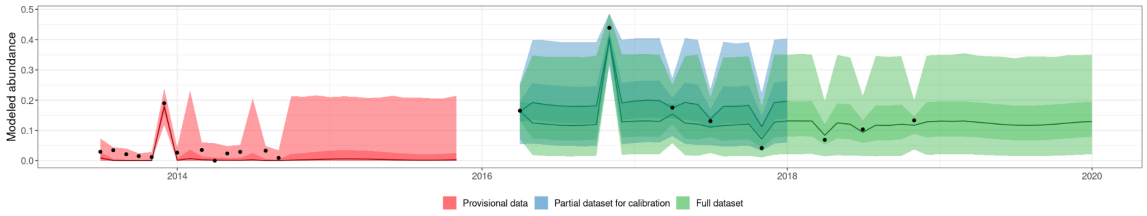


Figure S17. Calibration periods used in this analysis, shown as a time-series at one plot. Provisional data (red) was dropped due to bias caused by methodological changes. Partial dataset (blue) was used to evaluate forecasts. Full dataset (green) was used to create forecasts that can be evaluated in the future.

Appendix S3: Nimble Statistical Modeling Code.

The Nimble code below (embedded in R) represents a dynamic linear model of a single group, with environmental covariates and seasonal predictors included. Nimble code for the two simpler models is hosted at https://github.com/zoey-rw/microbialForecasts.

Nimble is a Bayesian statistical language and software with benefits for ecological modeling (*96*). It creates custom statistical distributions that represent common ecological data formats, such as counts, zero-inflation, and correlated data. For iterative forecasting in particular, the availability of cutting-edge data assimilation algorithms (e.g. sequential monte carlo filters) enables states to efficiently be updated with new data.

751 Data (observation) model

752 In constructing the statistical model, we evaluated the following distributions for the data model
753 (lines 3-6 in Example Code below), representing the microbial relative abundances observed in
754 multiple soil samples taken from the same plot (about $20m^2$). To increase computational
755 efficiency, we leveraged the advantage of conjugate Gibbs sampling whenever possible:
756 parameterizations of variance terms (e.g. distributions such as gamma, Inverse gamma, Normal
757 truncated).
758
759 *Dirichlet distribution*
760 The Dirichlet distribution is a multivariate generalization of the beta distribution, allowing
761 multiple fractional abundances to be modeled at once (*17*, *97*). This accounts for covariance
762 among microbial taxa arising simply from the "sum-to-one" constraint. However, the number of
763 model-estimated parameters increases linearly with the number of taxa included, and model
764 fitting is much less efficient (evaluated using Effective Sample Size (ESS) / computation time,
765 from the compareMCMCs R package (*98*)). When low-abundance taxa are included, model-
766 fitting occurred on the scale of weeks or months, rather than hours or days. Additionally, the
767 Dirichlet distribution does not account for correlation between increased sequencing depth and
768 precision, which can lead to heteroskedastic observation error. We found that predicted
769 coefficients were qualitatively similar between the Dirichlet and Beta regression approaches.
770
771 *Multinomial Dirichlet distribution*
772 The multinomial Dirichlet distribution has the properties of the Dirichlet distribution, with an
773 additional set of parameters representing observation error. This is a natural choice for microbial
774 data, since it can model fractional abundances as well as the effects of sequencing depth.
775 However, it had the same practical drawbacks as standard Dirichlet, with even lower model-
776 fitting efficiency and high sensitivity to model priors and initial conditions.
777
778 *Normal distribution with CLR transformation*
779 The center log-ratio (CLR) transformation is a common choice for microbial relative abundance
780 data because it reduces the impact of the sum-to-one constraint by scaling all abundances against
781 the geometric mean of a "reference" taxon (*99*). Because transformed values enter an unbounded
782 space (-Inf, Inf), the Normal distribution can be used for modeling observations. However, this
783 approach can make forecasts harder to interpret: forecast users may have to transform values
784 back to (0,1) for interpretability, or apply the CLR-transformation to all subsequent microbial
785 samples to compare observations with forecasts. We found that models fit to CLR-transformed
786 data produced similar parameter estimates to models fit with Beta-distributed relative
787 abundances, but the overall convergence rate was lower, especially for rare (low-abundance) taxa
788 (Fig S18).
789

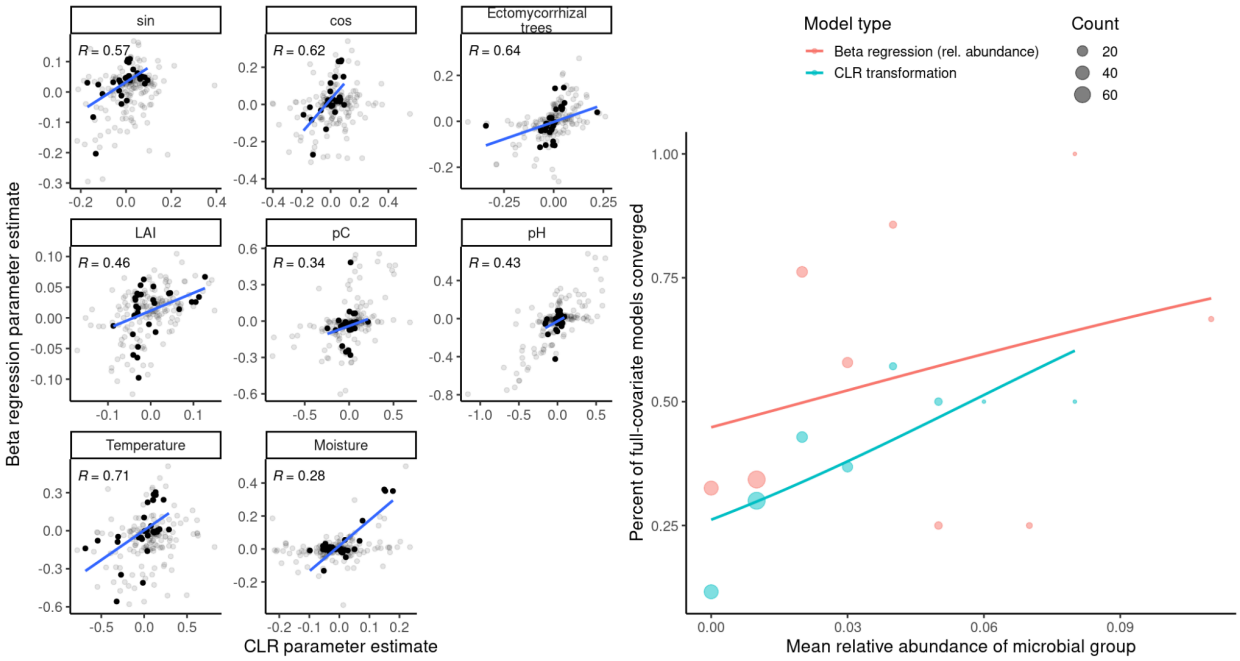


Figure S18. Comparison of models fit to microbial data transformed via centered-log ratio (CLR) or modeled as beta-distributed relative abundances. Left set of figures shows correlation between parameter estimates from the two approaches, calculated from the black points, which represent models that converged with both approaches; gray points are models that did not converge with one of the approaches. Right panel shows the percent of models converged as a function of the mean relative abundance of a group, for all abundance levels with more than one group representative. Colors indicate model approach and point size indicates the sample size of models at each abundance level.

Link function

Considerations for link function (which links the observed relative abundance values to the underlying model of the ecological process): the logit-link is commonly used link function, and improved model convergence by reducing proposed MCMC values outside the (0,1) interval. We also evaluated models with no link function, or with a log-link, which are both statistically valid, but led to much longer convergence times.

Bayesian prior considerations

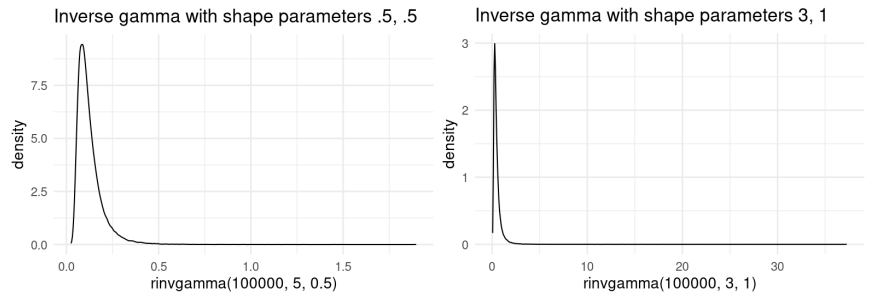


Figure S19. Distributions chosen as weakly informative priors in hierarchical Bayesian models.

Priors for process error (sigma, lines 46) and core-level variation (core_sd, line 45) were represented by an Inverse gamma distribution (Fig S19A). Priors for site effect variation (sig, line 47) were represented by a more lenient inverse gamma distribution (Fig 19B).

Error around environmental predictor values (lines 55-81) represents a combination of multiple error sources, which were modeled separately to reduce computational effort. This includes instrument error, reported by some of the data sources (see Appendix S2), and uncertainty around the specific sampling date within a month (represented by the standard deviation of the entire month's values).

Priors for coefficients associated with environmental predictors (lines 53) were given relatively flat priors with a mean of zero and a precision of .0001, corresponding to a standard deviation of 100. These priors could be improved with domain-specific knowledge from experimental or field manipulation, e.g. heat-loving soil extremophiles may be assigned a strong positive prior distribution, representing scientific evidence for an effect. However, in this analysis, uninformative priors were used for all environmental predictors to facilitate comparisons between model structures.

Code Example 1.

```
nimbleMod <- nimble::nimbleCode({

        # Loop through core observations (column 1 of input df "y") ----
        for (i in 1:N.core) {
                y[i, 1] ~ T(dnorm(mean = plot_mu[plot_num[i], timepoint[i]], sd = core_sd), 0, 1)
        }

        # Plot-level process model ----
        for (p in 1:N.plot) {
                for (t in plot_start[p]) {
                        # Plot means for first date
                        logit(Ex[p, t]) ~ dnorm(mean = X_init, sd = .2)
                        plot_mu[p, t] ~ T(dnorm(mean = Ex[p, t], sd = sd_init), 0, 1)
                }

                # Second date onwards
                for (t in plot_index[p]:N.date) {
                        # Dynamic linear model
                        logit(Ex[p, t]) <- rho * logit(plot_mu[p, t - 1]) +
                                beta[1] * temp_est[plot_site_num[p], t] +
                                beta[2] * mois_est[plot_site_num[p], t] +
```

```
                                beta[3] * pH_est[p, plot_start[p]] +
                                beta[4] * pC_est[p, plot_start[p]] +
                                beta[5] * relEM[p, t] +
                                beta[6] * LAI[plot_site_num[p], t] +
                                beta[7] * sin_mo[t] +
                                beta[8] * cos_mo[t] +
                                site_effect[plot_site_num[p]] +
                                intercept

                        # Define shape parameters for process error (sigma) model
                        shape1[p, t] <- Ex[p, t] * ((Ex[p, t] * (1-Ex[p, t]))/sigma^2 - 1)
                        shape2[p, t] <- (1 - Ex[p, t]) * ((Ex[p, t] * (1-Ex[p, t]))/sigma^2 - 1)

                        # Add process error (sigma)
                        logit(plot_mu[p, t]) ~ dLogitBeta(shape1[p, t], shape2[p, t])
                }
        }

        # Priors for site effects----
        for (k in 1:N.site) {
                site_effect[k] ~ dnorm(0,  sd = sig)
        }

        core_sd ~ dinvgamma(5,.5)
        sigma ~ dinvgamma(5,.5)
        sig ~ dinvgamma(3,1)

        intercept ~ dnorm(0, sd = 1)
        rho ~ dnorm(0, sd = 1) # Autocorrelation
        X_init <- .5 # Initial plot mean - will be constrained by data
        sd_init <- .01 # Initial process error
        beta[1:8] ~ dmnorm(zeros[1:8], omega[1:8, 1:8])

        # Add driver uncertainty if specified ----
        if (temporalDriverUncertainty) {
                for (k in 1:N.site) {
                        for (t in site_start[k]:N.date) {
                                mois_est[k, t] ~ dnorm(mois[k, t], sd = mois_sd[k, t])
                                temp_est[k, t] ~ dnorm(temp[k, t], sd = temp_sd[k, t])
                        }
                }
        } else {
                for (k in 1:N.site) {
                        for (t in site_start[k]:N.date) {
```

```
                                        mois_est[k, t] <- mois[k, t]
                                        temp_est[k, t] <- temp[k, t]
                                }
                        }
                }
                if (spatialDriverUncertainty) {
                        for (p in 1:N.plot) {
                                pH_est[p, plot_start[p]] ~ dnorm(pH[p, plot_start[p]], sd = pH_sd[p,
plot_start[p]])
                                pC_est[p, plot_start[p]] ~ dnorm(pC[p, plot_start[p]], sd = pC_sd[p,
plot_start[p]])
                        }
                } else {
                        for (p in 1:N.plot) {
                                pH_est[p, plot_start[p]] <- pH[p, plot_start[p]]
                                pC_est[p, plot_start[p]] <- pC[p, plot_start[p]]
                        }
                }

}) #end NIMBLE model
```

REFERENCES

1. J. K. Jansson, K. S. Hofmockel, Soil microbiomes and climate change. *Nat. Rev. Microbiol.* **18**, 35–46 (2020).
2. C. Gougoulias, J. M. Clark, L. J. Shaw, The role of soil microbes in the global carbon cycle: Tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems. *J. Sci. Food Agric.* **94**, 2362–2371 (2014).
3. R. Hooke, *Micrographia, or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses with Observations and Inquiries Thereupon / by R. Hooke ...* (1665; https://quod.lib.umich.edu/e/eebo/A44323.0001.001?view=toc).
4. A. van Leewenhoeck, Observations, communicated to the publisher by Mr. Antony van Leewenhoeck, in a dutch letter of the 9th Octob. 1676. here English'd: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infused. *Philos. Trans. R. Soc. Lond.* **12**, 821–831 (1677).
5. Winogradsky, S, Researches sur les Organismes de la Nitrification. *Ann. L'Institut Pasteur*, 577-616. (1891).
6. W. R. Wieder, G. B. Bonan, S. D. Allison, Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change* **3**, 909–912 (2013).
7. K. Lajtha, V. Bailey, K. McFarlane, K. Paustian, D. Bachelet, R. Abramoff, D. Angers, S. A. Billings, D. Cerkowniak, Y. G. Dialynas, A. Finzi, N. French, S. Frey, N. Gurwick, J. Harden, J. M. F. Johnson, K. Johnson, J. Lehmann, S. (Leo) Liu, B. McConkey, U. Mishra, S. Ollinger, D. Paré, F. Paz, D. deB. Richter, S. M. Schaeffer, J. Schimel, C. Shaw, J. Tang, K. Todd-Brown, C. Trettin, M. Waldrop, T. Whitman, K. Wickland, Chapter 12: Soils. Second State of the Carbon Cycle Report. doi: 10.7930/SOCCR2.2018.Ch12 (2018).
8. N. W. Schaad, "Emerging Plant Pathogenic Bacteria and Global Warming" in *Pseudomonas Syringae Pathovars and Related Pathogens – Identification, Epidemiology and Genomics*, M. Fatmi, A. Collmer, N. S. Iacobellis, J. W. Mansfield, J. Murillo, N. W. Schaad, M. Ullrich, Eds. (Springer Netherlands, Dordrecht, 2008; https://doi.org/10.1007/978-1-4020-6901-7_38), pp. 369–379.
9. M. Delgado-Baquerizo, C. A. Guerra, C. Cano-Díaz, E. Egidi, J. Wang, N. Eisenhauer, B. K. Singh, F. T. Maestre, The proportion of soil-borne pathogens increases with warming at the global scale. *Nat. Clim. Change*, doi: 10.1038/s41558-020-0759-3 (2020).
10. FAO, ITPS, GSBI, CBD, EC., *State of Knowledge of Soil Biodiversity - Status, Challenges and Potentialities* (Food and Agriculture Organization of the United Nations, Rome, 2020).
11. M. C. Dietze, Prediction in ecology: A first-principles framework. *Ecol. Appl.* **27**, 2048–2060 (2017).
12. M. C. Dietze, A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers, E. P. White, Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1424–1432 (2018).
13. C. C. Carey, W. M. Woelmer, M. E. Lofton, R. J. Figueiredo, B. J. Bookout, R. S. Corrigan, V. Daneshmand, A. G. Hounshell, D. W. Howard, A. S. L. Lewis, R. P. McClure, H. L. Wander, N. K. Ward, R. Q. Thomas, Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*, doi: 10.1080/20442041.2020.1816421 (2020).
14. M. E. Lofton, J. A. Brentrup, W. S. Beck, J. A. Zwart, R. Bhattacharya, L. S. Brighenti, S. H. Burnet, I. M. McCullough, B. G. Steele, C. C. Carey, K. L. Cottingham, M. C. Dietze, H. A. Ewing, K. C. Weathers, S. L. LaDeau, Using near-term forecasts and uncertainty partitioning to inform prediction of oligotrophic lake cyanobacterial density. *Ecol. Appl.* **32**, e2590 (2022).
15. D. J. Currie, Where Newton might have taken ecology. *Glob. Ecol. Biogeogr.* **28**, 18–27 (2019).
16. M. C. Dietze, *Ecological Forecasting* (Princeton University Press, 2019).

17. C. Averill, Z. R. Werbin, K. F. Atherton, J. M. Bhatnagar, M. C. Dietze, Soil microbiome predictability increases with spatial and taxonomic scale. *Nat. Ecol. Evol.* **5**, 747–756 (2021).
18. N. J. Bouskill, H. C. Lim, S. Borglin, R. Salve, T. E. Wood, W. L. Silver, E. L. Brodie, Pre-exposure to drought increases the resistance of tropical forest soil bacterial communities to extended drought. *ISME J.* **7**, 384–394 (2013).
19. S. M. Hatosy, J. B. H. Martiny, R. Sachdeva, J. Steele, J. A. Fuhrman, A. C. Martiny, Beta diversity of marine bacteria depends on temporal scale. *Ecology* **94**, 1898–1904 (2013).
20. S. Ferrenberg, S. P. O'neill, J. E. Knelman, B. Todd, S. Duggan, D. Bradley, T. Robinson, S. K. Schmidt, A. R. Townsend, M. W. Williams, C. C. Cleveland, B. A. Melbourne, L. Jiang, D. R. Nemergut, Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *ISME J.* **7**, 1102–1111 (2013).
21. A. Bissett, A. E. Richardson, G. Baker, S. Wakelin, P. H. Thrall, Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol. Ecol.* **19**, 4315–4327 (2010).
22. M. Goberna, M. Verdú, Phylogenetic-scale disparities in the soil microbial diversity–ecosystem functioning relationship. *ISME J.* **12**, 2152–2162 (2018).
23. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A phylogenetic perspective. *Science* **350** (2015).
24. C. Averill, L. A. L. Cates, M. C. Dietze, J. M. Bhatnagar, Spatial vs. temporal controls over soil fungal community similarity at continental and global scales. *ISME J.* **13**, 1–34 (2019).
25. A. C. Finzi, R. Z. Abramoff, K. S. Spiller, E. R. Brzostek, B. A. Darby, M. A. Kramer, R. P. Phillips, Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles. *Glob. Change Biol.* **21**, 2082–2094 (2015).
26. K. Zhalnina, K. B. Louie, Z. Hao, N. Mansoori, U. N. Da Rocha, S. Shi, H. Cho, U. Karaoz, D. Loqué, B. P. Bowen, M. K. Firestone, T. R. Northen, E. L. Brodie, Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* **3**, 470–480 (2018).
27. K. L. Matulich, C. Weihe, S. D. Allison, A. S. Amend, R. Berlemont, M. L. Goulden, S. Kimball, A. C. Martiny, J. B. H. Martiny, Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. *ISME J.* **9**, 2477–2489 (2015).
28. S. K. Chapman, G. S. Newman, Biodiversity at the plant-soil interface: Microbial abundance and community structure respond to litter mixing. *Oecologia* **162**, 763–769 (2010).
29. R. Z. Abramoff, A. C. Finzi, Are above- and below-ground phenology in sync? *New Phytol.* **205**, 1054–1061 (2015).
30. J. Voříškova, V. Brabcová, T. Cajthaml, P. Baldrian, Seasonal dynamics of fungal communities in a temperate oak forest soil. *New Phytol.* **201**, 269–278 (2014).
31. A. C. Martiny, K. Treseder, G. Pusch, Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* **7**, 830–838 (2013).
32. J. Cavender-Bares, A. Keen, B. Miles, PHYLOGENETIC STRUCTURE OF FLORIDIAN PLANT COMMUNITIES DEPENDS ON TAXONOMIC AND SPATIAL SCALE. *Ecology* **87**, S109–S122 (2006).
33. J. Cavender-Bares, K. H. Kozak, P. V. A. Fine, S. W. Kembel, The merging of community ecology and phylogenetic biology. *Ecol. Lett.* **12**, 693–715 (2009).
34. A. Shade, R. R. Dunn, S. A. Blowes, P. Keil, B. J. M. Bohannan, M. Herrmann, K. Küsel, J. T. Lennon, N. J. Sanders, D. Storch, J. Chase, Macroecology to Unite All Life, Large and Small. *Trends Ecol. Evol.* **33**, 731–744 (2018).
35. C. Kaiser, M. Koranda, B. Kitzler, L. Fuchslueger, J. Schnecker, P. Schweiger, F. Rasche, S. Zechmeister-Boltenstern, A. Sessitsch, A. Richter, Belowground carbon allocation by trees drives seasonal patterns of extracellular enzyme activities by altering microbial community composition in a beech forest soil. *New Phytol.* **187**, 843–858 (2010).
36. X. Shan, A. Goyal, R. Gregor, O. X. Cordero, Annotation-free discovery of functional groups in

microbial communities. *bioRxiv*, 2022.08.02.502537 (2022).
37. X. Xiao, N. Zhang, H. Ni, Y. Yang, J. Zhou, B. Sun, Y. Liang, A latitudinal gradient of microbial β-diversity in continental paddy soils. *Glob. Ecol. Biogeogr.* **30**, 909–919 (2021).
38. C. P. Andam, J. R. Doroghazi, A. N. Campbell, P. J. Kelly, M. J. Choudoir, D. H. Buckley, A Latitudinal Diversity Gradient in Terrestrial Bacteria of the Genus *Streptomyces*. *mBio* **7**, e02200-15 (2016).
39. S. Bickel, X. Chen, A. Papritz, D. Or, A hierarchy of environmental covariates control the global biogeography of soil bacterial richness. *Sci. Rep.* **9**, 12129 (2019).
40. S. Bickel, D. Or, Soil bacterial diversity mediated by microscale aqueous-phase processes across biomes. *Nat. Commun.* **11**, 116 (2020).
41. K. Bodner, C. Rauen Firkowski, J. R. Bennett, C. Brookson, M. Dietze, S. Green, J. Hughes, J. Kerr, M. Kunegel-Lion, S. J. Leroux, E. McIntire, P. K. Molnár, C. Simpkins, E. Tekwa, A. Watts, M. Fortin, Bridging the divide between ecological forecasts and environmental decision making. *Ecosphere* **12**, e03869 (2021).
42. R. Tecon, D. Or, Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiol. Rev.* **41**, 599–623 (2017).
43. C. Averill, B. G. Waring, C. V. Hawkes, Historical precipitation predictably alters the shape and magnitude of microbial functional response to soil moisture. *Glob. Change Biol.* **22**, 1957–1964 (2016).
44. A. E. Zanne, K. Abarenkov, M. E. Afkhami, C. A. Aguilar-Trigueros, S. Bates, J. M. Bhatnagar, P. E. Busby, N. Christian, W. K. Cornwell, T. W. Crowther, H. Flores-Moreno, D. Floudas, R. Gazis, D. Hibbett, P. Kennedy, D. L. Lindner, D. S. Maynard, A. M. Milo, R. H. Nilsson, J. Powell, M. Schildhauer, J. Schilling, K. K. Treseder, Fungal functional ecology: bringing a trait-based approach to plant-associated fungi. *Biol. Rev.* **95**, 409–433 (2020).
45. S. Põlme, K. Abarenkov, R. Henrik Nilsson, B. D. Lindahl, K. E. Clemmensen, H. Kauserud, N. Nguyen, R. Kjøller, S. T. Bates, P. Baldrian, T. G. Frøslev, K. Adojaan, A. Vizzini, A. Suija, D. Pfister, H. O. Baral, H. Järv, H. Madrid, J. Nordén, J. K. Liu, J. Pawlowska, K. Põldmaa, K. Pärtel, K. Runnel, K. Hansen, K. H. Larsson, K. D. Hyde, M. Sandoval-Denis, M. E. Smith, M. Toome-Heller, N. N. Wijayawardene, N. Menolli, N. K. Reynolds, R. Drenkhan, S. S. N. Maharachchikumbura, T. B. Gibertoni, T. Læssøe, W. Davis, Y. Tokarev, A. Corrales, A. M. Soares, A. Agan, A. R. Machado, A. Argüelles-Moyao, A. Detheridge, A. de Meiras-Ottoni, A. Verbeken, A. K. Dutta, B. K. Cui, C. K. Pradeep, C. Marín, D. Stanton, D. Gohar, D. N. Wanasinghe, E. Otsing, F. Aslani, G. W. Griffith, T. H. Lumbsch, H. P. Grossart, H. Masigol, I. Timling, I. Hiiesalu, J. Oja, J. Y. Kupagme, J. Geml, J. Alvarez-Manjarrez, K. Ilves, K. Loit, K. Adamson, K. Nara, K. Küngas, K. Rojas-Jimenez, K. Bitenieks, L. Irinyi, L. L. Nagy, L. Soonvald, L. W. Zhou, L. Wagner, M. C. Aime, M. Öpik, M. I. Mujica, M. Metsoja, M. Ryberg, M. Vasar, M. Murata, M. P. Nelsen, M. Cleary, M. C. Samarakoon, M. Doilom, M. Bahram, N. Hagh-Doust, O. Dulya, P. Johnston, P. Kohout, Q. Chen, Q. Tian, R. Nandi, R. Amiri, R. H. Perera, R. dos Santos Chikowski, R. L. Mendes-Alvarenga, R. Garibay-Orijel, R. Gielen, R. Phookamsak, R. S. Jayawardena, S. Rahimlou, S. C. Karunarathna, S. Tibpromma, S. P. Brown, S. K. Sepp, S. Mundra, Z. H. Luo, T. Bose, T. Vahter, T. Netherway, T. Yang, T. May, T. Varga, W. Li, V. R. M. Coimbra, V. R. T. de Oliveira, V. X. de Lima, V. S. Mikryukov, Y. Lu, Y. Matsuda, Y. Miyamoto, U. Kõljalg, L. Tedersoo, FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers.* **105** (2020).
46. J. D. Hoeksema, C. Averill, J. M. Bhatnagar, E. Brzostek, E. Buscardo, K.-H. Chen, H.-L. Liao, L. Nagy, N. Policelli, J. Ridgeway, J. A. Rojas, R. Vilgalys, Ectomycorrhizal Plant-Fungal Co-invasions as Natural Experiments for Connecting Plant and Fungal Traits to Their Ecosystem Consequences. *Front. For. Glob. Change* **3**, 1–13 (2020).
47. S. Kramer, D. Dibbern, J. Moll, M. Huenninghaus, R. Koller, D. Krueger, S. Marhan, T. Urich, T. Wubet, M. Bonkowski, F. Buscot, T. Lueders, E. Kandeler, Resource partitioning between bacteria, fungi, and protists in the detritusphere of an agricultural soil. *Front. Microbiol.* **7**, 1–12 (2016).

48. D. S. Maynard, K. R. Covey, T. W. Crowther, N. W. Sokol, E. W. Morrison, S. D. Frey, L. T. A. van Diepen, M. A. Bradford, Species associations overwhelm abiotic conditions to dictate the structure and function of wood-decay fungal communities. *Ecology* **99**, 801–811 (2018).
49. B. L. Turner, E. Laliberté, Soil Development and Nutrient Availability Along a 2 Million-Year Coastal Dune Chronosequence Under Species-Rich Mediterranean Shrubland in Southwestern Australia. *Ecosystems* **18**, 287–309 (2015).
50. L. R. Walker, D. A. Wardle, R. D. Bardgett, B. D. Clarkson, The use of chronosequences in studies of ecological succession and soil development. *J. Ecol.* **98**, 725–736 (2010).
51. L. F. Stanish, J. Parnell, NEON.DOC.000908: TOS Science Design for Terrestrial Microbial Diversity. *NEON Doc. Libr.* (2018).
52. L. Stanish, S. Weintraub, E.-L. S. Hinckley, J. Parnell, K. Thibault, M. Stewart, TOS PROTOCOL AND PROCEDURE: SOIL BIOGEOCHEMICAL AND MICROBIAL SAMPLING.
53. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
54. C. Qin, R. Bartelme, Y. A. Chung, D. Fairbanks, Y. Lin, D. Liptzin, C. Muscarella, K. Naithani, K. Peay, P. Pellitier, A. St. Rose, L. Stanish, Z. Werbin, K. Zhu, From DNA sequences to microbial ecology: Wrangling NEON soil microbe data with the neonMicrobe R package. *Ecosphere* **12** (2021).
55. S. Theil, E. Rifa, rANOMALY: AmplicoN wOrkflow for Microbial community AnaLYsis. *F1000Research* **10**, 1–24 (2021).
56. R. H. Nilsson, K. H. Larsson, A. F. S. Taylor, J. Bengtsson-Palme, T. S. Jeppesen, D. Schigel, P. Kennedy, K. Picard, F. O. Glöckner, L. Tedersoo, I. Saar, U. Kõljalg, K. Abarenkov, The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D264 (2019).
57. S. Theil, E. Rifa, UTOPIA: an automatically UpdaTed, cOmPlete and consistent ITS reference dAtabase. *Colloq. Génomique Environnementale*, 0–1 (2019).
58. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
59. D. H. Parks, M. Chuvochina, C. Rinke, A. J. Mussig, P.-A. Chaumeil, P. Hugenholtz, GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **202**, 1–10 (2021).
60. Y. Jia, S. Zhao, W. Guo, L. Peng, F. Zhao, L. Wang, G. Fan, Y. Zhu, D. Xu, G. Liu, R. Wang, X. Fang, H. Zhang, K. Kristiansen, W. Zhang, J. Chen, Sequencing introduced false positive rare taxa lead to biased microbial community diversity, assembly, and interaction interpretation in amplicon studies. *Environ. Microbiome* **17**, 43 (2022).
61. A. Scoma, Functional groups in microbial ecology: updated definitions of piezophiles as suggested by hydrostatic pressure dependence on temperature. *ISME J.* **15**, 1871–1878 (2021).
62. N. H. Nguyen, Z. Song, S. T. Bates, S. Branco, L. Tedersoo, J. Menke, J. S. Schilling, P. G. Kennedy, FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
63. D. Naylor, S. Fansler, C. Brislawn, W. C. Nelson, K. S. Hofmockel, J. K. Jansson, R. McClure, Deconstructing the soil microbiome into reduced-complexity functional modules. *mBio* **11**, 1–19 (2020).
64. A. Ho, D. P. Di Lonardo, P. L. E. Bodelier, Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol. Ecol.* **93**, 1–14 (2017).
65. R. Berlemont, A. C. Martiny, Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* **79**, 1545–1554 (2013).
66. M. B. N. Albright, B. Timalsina, J. B. H. Martiny, J. Dunbar, Comparative Genomics of Nitrogen Cycling Pathways in Bacteria and Archaea. *Microb. Ecol.*, doi: 10.1007/s00248-018-1239-4 (2018).
67. R. Lücking, M. C. Aime, B. Robbertse, A. N. Miller, H. A. Ariyawansa, T. Aoki, G. Cardinali, P.

W. Crous, I. S. Druzhinina, D. M. Geiser, D. L. Hawksworth, K. D. Hyde, L. Irinyi, R. Jeewon, P. R. Johnston, P. M. Kirk, E. Malosso, T. W. May, W. Meyer, M. Öpik, V. Robert, M. Stadler, M. Thines, D. Vu, A. M. Yurkov, N. Zhang, C. L. Schoch, Unambiguous identification of fungi: Where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* **11** (2020).

68. B. J. Callahan, K. Sankaran, J. A. Fukuyama, P. J. McMurdie, S. P. Holmes, Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research* **5**, 1492 (2016).
69. J. Ooms, S. Chamberlain, L. S. Reyes, phylocomr: Interface to "Phylocom," (2023); https://cran.r-project.org/package=phylocomr.
70. S. Vivelo, J. M. Bhatnagar, An evolutionary signal to fungal succession during plant litter decay. *FEMS Microbiol. Ecol.* **95**, fiz145 (2019).
71. S. L. P. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31**, 799–815 (2004).
72. A. M. Stolwijk, H. Straatman, G. A. Zielhuis, Studying seasonality by using sine and cosine functions in regression analysis. *J. Epidemiol. Community Health* **53**, 235–238 (1999).
73. M. Friedl, J. Gray, D. Sulla-Menashe, MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V061 [Data set] (2022). https://doi.org/10.5067/MODIS/MCD12Q2.061.
74. J. Gray, D. Sulla-Menashe, M. A. Friedl, User Guide to Collection 6 MODIS Land Cover Dynamics (MCD12Q2) Product.
75. P. De Valpine, D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, R. Bodik, Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *J. Comput. Graph. Stat.* **26**, 403–413 (2017).
76. C. K. Enders, D. Tofighi, Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *Psychol. Methods* **12**, 121–138 (2007).
77. M. Plummer, N. Best, K. Cowles, K. Vines, {CODA}: Convergence Diagnosis and Output Analysis for {MCMC}. *R News* **6**, 7–11 (2006).
78. A. Jordan, F. Krüger, S. Lerch, Evaluating probabilistic forecasts with scoringRules. *J. Stat. Softw.* **90** (2019).
79. R Development Core Team, R: A language and environment for statistical computing., R Foundation for Statistical Computing (2008).
80. NEON (National Ecological Observatory Network), Soil temperature (DP1.00041.001), RELEASE-2023.; https://doi.org/10.48443/9g10-ep60.
81. K. H. Liland, B.-H. Mevik, R. Wehrens, pls: Partial Least Squares and Principal Component Regression (2023). https://cran.r-project.org/package=pls.
82. J. Lamour, S. Serbin, spectratrait: A simple add-on package to aid in the fitting of leaf-level spectra-trait PLSR models (2023).
83. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate : A practical and powerful approach to multiple testing author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. *J. R. Stat. Soc.* **57**, 289–300 (1995).
84. B. Gräler, E. Pebesma, G. Heuvelink, Spatio-Temporal Interpolation using gstat. *R J.* **8**, 204–218 (2016).
85. C. Wang, R. Furrer, Monte Carlo Permutation Tests for Assessing Spatial Dependence at Difference Scales. *Nonparametric Stat.* (2018).
86. M. M. Thornton, R. Shrestha, Y. Wei, P. E. Thornton, S. Kao, B. E. Wilson, *Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 4* (ORNL Distributed Active Archive Center, 2020; https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1840).
87. G. L. Schaefer, M. H. Cosh, T. J. Jackson, The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). *J. Atmospheric Ocean. Technol.* **24**, 2073–2077 (2007).
88. and Q. L. Reichle, R., G. De Lannoy, R. D. Koster, W. T. Crow, J. S. Kimball, SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 4 (2017).

https://doi.org/10.5067/60HB8VIP2T8W.
89. J. Font, C. Gabarró, J. Ballabrera, A. Turiel, J. Martínez, M. Umbert, F. Pérez, N. Hoareau, M. Portabella, V. González, J. Gourrion, S. Guimbard, M. Piles, A. Camps, M. Vall-Llossera, SMOS CP34 soil moisture and ocean salinity maps. *2012 12th Spec. Meet. Microw. Radiom. Remote Sens. Environ. MicroRad 2012 - Proc.*, doi: 10.1109/MicroRad.2012.6185236 (2012).
90. E. Ayres, J. Roberti, T. Meehan, S. Petroy, T. Meehan, V. Aleksiev, NEON ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD) TIS SOIL TEMPERATURE.
91. E. Ayres, J. Roberti, K. Thibault, NEON ALGORITHMTHEORETICAL BASIS DOCUMENT (ATBD): TIS SOIL WATER CONTENT AND WATER SALINITY.
92. NEON (National Ecological Observatory Network), Soil physical and chemical properties, periodic (DP1.10086.001), RELEASE-2023.; https://doi.org/10.48443/0phb-j505.
93. NEON (National Ecological Observatory Network), Vegetation structure (DP1.10098.001), RELEASE-2023.; https://doi.org/10.48443/73zn-k414.
94. D. S. LeBauer, The Predictive Ecosystem Analyzer (PEcAn) is an integrated ecological bioinformatics toolbox.: PecanProject/pecan, (2019); https://github.com/PecanProject/pecan.
95. T. Myneni, R., Knyazikhin, Y., Park, MOD15A2H MODIS Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Process. DAAC* (2015).
96. L. C. Ponisio, P. de Valpine, N. Michaud, D. Turek, One size does not fit all: Customizing MCMC methods for hierarchical models using NIMBLE. *Ecol. Evol.* **10**, 2385–2416 (2020).
97. V. Pawlowsky-Glahn, J. J. Egozcue, R. Tolosana-Delgado, Modeling and Analysis of Compositional Data. *Model. Anal. Compos. Data*, 1–247 (2015).
98. P. De Valpine, S. Paganin, D. Turek, compareMCMCs: An R package for studying MCMC efficiency. *J. Open Source Softw.* **7**, 3844 (2022).
99. T. P. Quinn, I. Erb, G. Gloor, C. Notredame, M. F. Richardson, T. M. Crowley, A field guide for the compositional analysis of any-omics data. *GigaScience* **8**, 1–14 (2019).