Standardized Data to Improve Understanding and Modeling of Soil Nitrogen at Continental Scale

Samantha R. Weintraub-Leff[1] , **Steven J. Hall[2]** , **Matthew E. Craig[3], Debjani Sihi[4]** , **Zhuonan Wang[4]** , **and Stephen C. Hart[5]**

[1]National Ecological Observatory Network, Battelle, Boulder, CO, USA, [2]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA, [3]Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [4]Department of Environmental Sciences, Emory University, Atlanta, GA, USA, [5]Department of Life and Environmental Sciences, Sierra Nevada Research Institute, University of California Merced, Merced, CA, USA

Abstract Nitrogen (N) is a key limiting nutrient in terrestrial ecosystems, but there remain critical gaps in our ability to predict and model controls on soil N cycling. This may be in part due to lack of standardized sampling across broad spatial–temporal scales. Here, we introduce a continentally distributed, publicly available data set collected by the National Ecological Observatory Network (NEON) that can help fill these gaps. First, we detail the sampling design and methods used to collect and analyze soil inorganic N pool and net flux rate data from 47 terrestrial sites. We address methodological challenges in generating a standardized data set, even for a network using uniform protocols. Then, we evaluate sources of variation within the sampling design and compare measured net N mineralization to simulated fluxes from the Community Earth System Model 2 (CESM2). We observed wide spatiotemporal variation in inorganic N pool sizes and net transformation rates. Site explained the most variation in NEON's stratified sampling design, followed by plots within sites. Organic horizons had larger pools and net N transformation rates than mineral horizons on a sample weight basis. The majority of sites showed some degree of seasonality in N dynamics, but overall these temporal patterns were not matched by CESM2, leading to poor correspondence between observed and modeled data. Looking forward, these data can reveal new insights into controls on soil N cycling, especially in the context of other environmental data sets provided by NEON, and should be leveraged to improve predictive modeling of the soil N cycle.

Plain Language Summary Nitrogen (N) is not only a key limiting nutrient that often constrains plant growth but also a major pollutant in places where supply exceeds demand. However, we lack the ability to accurately predict and model rates of soil N cycling. This paper introduces a first of its kind, standardized and publicly available data set of soil inorganic N pools and net transformation rates spanning the United States. We describe the data set in detail, then examine spatiotemporal trends in N pools and fluxes and how those measurements compare to predictions from an Earth system model.

1. Introduction

Nitrogen (N) is a critical element in the Earth system and acts as a limiting nutrient in many biomes (LeBauer & Treseder, 2008), especially those located in higher latitudes (Du et al., 2020; P. Vitousek & Howarth, 1991). In these ecosystems, N availability is likely to play a role in limiting carbon (C) uptake in the face of rising atmospheric carbon dioxide (Craine et al., 2018; Feng et al., 2015; Groffman et al., 2018; Reich et al., 2006), constraining the size of the land C sink (Wang & Houlton, 2009; Zaehle et al., 2015) and influencing Earth's climate. At the same time, industrial and agricultural activities have greatly increased the amount of reactive N flowing through the biosphere (Gruber & Galloway, 2008; Sullivan et al., 2014); for certain ecosystems, this has led to eutrophication, declines in biodiversity, and increased emissions of heat-trapping gasses such as nitrous oxide (N_2O; Bergström & Jansson, 2006; Davidson, 2009; Firestone & Davidson, 1989; Galloway et al., 2003; Payne et al., 2017; Sutton et al., 2011; P. M. Vitousek et al., 1997). Understanding where there is insufficient N or an excess compared to biological demand, what drives those trends, and the consequences for ecosystem structure and function is relevant to a wide range of ecological and conservation issues.

Whether N is limiting or in excess depends on the balance between N supply and demand and how both are influenced by environmental factors including variation in climate and growing seasons. In regions with seasonal

snowpack, there is a strong positive influence of snow cover on N availability in spring (Brooks et al., 1998; Durán et al., 2014), while in warmer ecosystems, temporal trends in N availability are often linked to seasonality of rainfall (Austin et al., 2004; Davidson et al., 1990; Homyak et al., 2014). But many regions are experiencing less snowpack and earlier growing season arrival, along with more variable rainfall and prolonged droughts (Gergel et al., 2017; Kapnick & Hall, 2010). How will these changes influence seasonal patterns in N supply and demand?

Along with seasonality, other factors such as vegetation cover and soil type are known to influence N availability. For instance, herbaceous cover tends to promote higher soil N availability compared to forests due to higher-quality plant litter inputs (Chapman et al., 2006; Weintraub et al., 2017), highly weathered soils such as those found in the humid tropics tend to be relatively N-rich (Brookshire et al., 2011; P. Vitousek & Matson, 1988), and organic soils found in colder climates tend to have high N stocks and turnover rates compared to underlying mineral horizons (Hart & Gunther, 1989). In the context of our rapidly changing climate and resulting alterations in seasonality and vegetation phenology, we need a better understanding of the trajectory of seasonal N dynamics and the relative importance of this compared to controls by vegetation composition, soil horizon type, and other ecosystem factors.

Given the many variables that influence N availability, this property is difficult to predict and model across scales. A common proxy for available N is net nitrogen mineralization, the net microbial production of inorganic N from organic matter. Nitrogen mineralization provides a major N source for plant and microbial growth in terrestrial ecosystems, although soluble organic N can be equally important in low N systems (Schimel & Bennett, 2004). N release from plant litter is well predicted using initial litter chemical characteristics and site climate (Parton et al., 2007). Yet we still lack the ability to predict broad-scale controls on the rates of N cycle processes in soil, where mineralogy and mineral–organic matter interactions are likely to figure prominently (Bingham & Cotrufo, 2016; Jilling et al., 2018). This holds true for N mineralization as well as nitrification, the conversion of ammonium to nitrate with a byproduct of N_2O and other N trace gasses. For instance, Colman and Schimel (2013) measured rates of net N mineralization via laboratory incubations of continentally distributed mineral soils under constant temperature and moisture and found that edaphic and site factors could only explain ~33% of the observed variation. Using data synthesis, Li et al. (2020) found increasing rates of net nitrification with decreasing latitude and increasing temperature, and a positive relationship with N_2O production and net nitrification rate. However, they observed wide variation in nitrification rates within biomes and low explanatory power overall. Soils display high spatial heterogeneity from local to regional and continental scales, with land use history contributing to variation in N dynamics. Explicitly capturing this cross-scale variance may be key to improved understanding and modeling of ecosystem function and response to global change (Bradford et al., 2021). Indeed, N cycle processes have proved challenging to represent in Earth system models (ESMs; Davies-Barnard et al., 2020; Thomas et al., 2015), and this makes it difficult to predict N availability and related carbon cycle processes, now and in the future.

A suite of standardized, broadly distributed measurements of soil inorganic N pools and net transformation rates could shed new insights into controls on N cycling in soils. The National Ecological Observatory Network (NEON) provides a community resource to deliver such measurements, along with a rich set of ancillary biogeochemical, ecological, and climatic variables. To our knowledge, there has been no other effort of this type to collect standardized data on soil N pools and transformation rates at sites distributed strategically across the U.S., with stratified within-site sampling across land-cover types, intra-annual sampling to capture seasonal dynamics, and plans for repeated sampling every 5 years for several decades. The spatial and temporal axes covered by this sampling should allow for new understanding and improve our ability to understand and model the forces shaping soil N availability in response to global change.

The aim of this paper is to introduce the NEON soil inorganic N data set to the biogeochemistry research community and illustrate how a unified data set such as this can address important research questions. First, we describe the spatial and temporal sampling design along with analytical approaches and challenges in generating a robust inorganic N data set, even for a project with highly standardized protocols. After exploring solutions to overcome the challenges, we use the data to investigate sources of variation in the context of the NEON sampling design, to probe whether NEON measurements are capturing heterogeneity in N cycling along dimensions of seasonality, soil horizon type, vegetation cover, and across spatial scales. Additionally, we compare NEON soil N measurements with outputs from an ESM (Community Earth System Model version 2, CESM2) to evaluate the degree
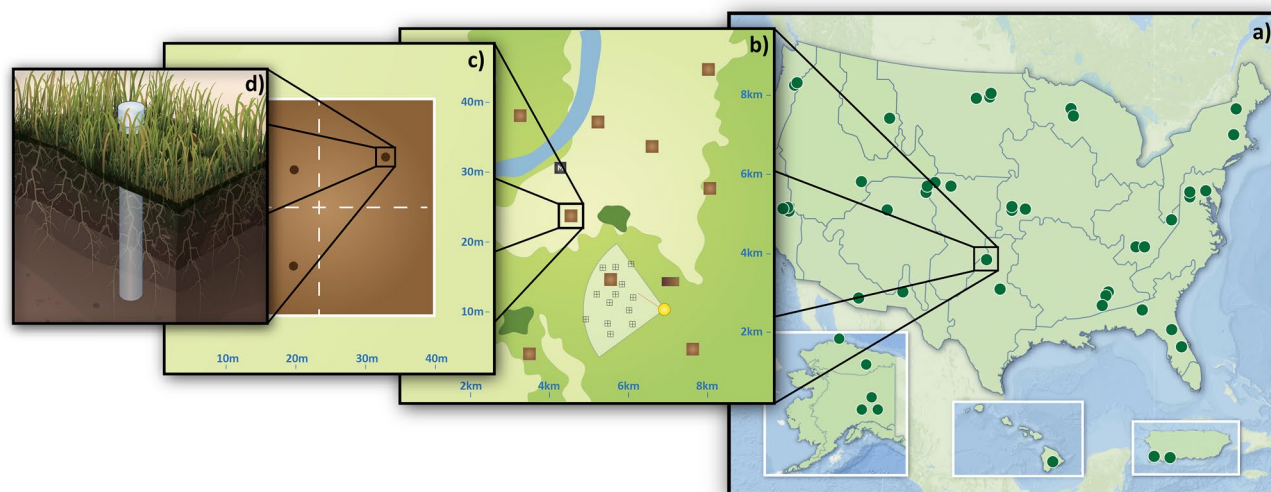
Figure 1. Nested scales of the National Ecological Observatory Network soil sampling design. (a) Forty-seven terrestrial sites (colored circles) are distributed across 20 eco-climatic domains. (b) Within each site, 10 plots are used for soil sampling (brown boxes), with a random-stratified distribution across dominant land-cover types. (c) Each plot is divided into four subplots, three of which are sampled in any given bout at random, preselected locations (brown circles). (d) At each subplot location, an initial core is collected and an incubated core is installed for in situ, net N transformation rate measurements.

of agreement between the two. Finally, we discuss the many opportunities to leverage these data, pairing them with other NEON data sets, additional measurements, and modeling efforts to yield new insights into terrestrial N availability.

2. Materials and Methods

2.1. NEON Sampling and Data

The NEON is a U.S. based monitoring networking that provides open ecological data at a continental scale (www.neonscience.org). The NEON provides 182 data products that deliver standardized measurements of biodiversity, biogeochemistry, micrometeorology, disease ecology, and other environmental variables. The observatory is composed of 47 terrestrial and 34 aquatic sites, distributed across 20 eco-climatic domains (Figure 1). Sensor infrastructure and observational sampling are used to observe ecosystems on the scale of minutes to years and from microns to kilometers (Balch et al., 2020; Keller et al., 2008).

Measurements of soil inorganic N pools and net transformation rates are a key component of the terrestrial biogeochemistry sampling design (Hinckley et al., 2016). These measurements are provided as part of NEON data product DP1.10086.001, soil physical and chemical properties, periodic (https://data.neonscience.org/data-products/DP1.10086.001). At each NEON terrestrial site, ten 40×40-m plots are used for long-term soil sampling. These plots are located using stratified random sampling to encompass the major vegetation/land-cover types within each site, according to National Land Cover Database (NLCD) classification (Barnett et al., 2019). Each plot is divided into quadrants, and soils are sampled from three of these four quadrants from a random, preselected location during each field campaign, which in NEON terminology is a "bout" (Figure 1). Thus, at each site, each of the 10 plots and 3 of the 4 subplots within each plot are resampled during each sampling bout, and the exact location within each subplot rotates for each bout. Samples are collected from 0 to 30 cm depth (or refusal), a standardized depth that was selected by NEON to facilitate within and across site comparisons and which includes the highest rates of soil N cycling activity and the majority of fine root biomass in most ecosystems (Binkley & Hart, 1989; Jackson et al., 1996), including NEON sites (Table S1 in Supporting Information S1). Soils are generally sampled using 2 or 2¼ in. (5.1 or 5.7 cm) soil augers, but with some adjustments to accommodate difficult to sample soils (see Appendix table in the NEON soil protocol, Stanish, 2022). If a surficial organic horizon is present, it is separated from mineral material and processed as a distinct sample.

Most NEON sites have three soil sampling bouts per year, except for the five sites located in Alaska that only have one due to the short growing season. All sites are sampled during their period of peak greenness, and most

are also sampled during two seasonal transition periods, the timing of which differs between sites as a function of climate. Seasonally cold sites are sampled during their respective spring and fall, while sites with milder temperature with greenness driven by rainfall are sampled during rainy season onset and dry down. Inorganic N pools and net transformation rates are measured during all bouts in a given sampling year, but each site only conducts inorganic N analyses one out of every 5 years, with other years reserved for nonbiogeochemistry soil sampling. This allows for the examination of seasonal trends in N dynamics, as well as how these dynamics may be changing over time at the continental scale.

Within 1 day of collection, field moist mineral horizon soils are sieved to 2 mm, and organic horizon soils are hand-picked to remove roots, rocks, and other non-soil material. Then, inorganic N pools are measured using 2 M potassium chloride (KCl) extractions with 10:1 solution to fresh soil weight ratio (Bremner & Keeney, 1966). The NEON includes up to three procedural blanks for each day of extractions to account for any contaminant N introduced in the procedure. Aliquots from "initial" soil samples are used to measure bacterial and fungal community composition, microbial biomass, pH, and gravimetric soil moisture, as well as total organic C and total N concentrations, C and N stable isotope ratios, and soil metagenomic content during peak greenness. To measure net N transformation rates, covered cores (PVC cylinders with loose-fitting caps) are installed to 30 cm depth (or refusal) when initial soil samples are taken. The exception is for wetland plots where buried bags are used instead to prevent the water table from interfering with N accumulation (Hart et al., 1994). These covered cores (or buried bags) are left to incubate in situ, then recovered 2–4 weeks later (depending on site and time of year), and extracted with 2 M KCl, similar to the initial samples.

The KCl extracts are frozen following collection and shipped to a centralized facility contracted by NEON (2017–2018, EcoCore Analytical Services at Colorado State University; 2019–present, University of Florida Wetland Biogeochemistry Laboratory) for analysis of ammonium (NH_4^+) and nitrate + nitrite ($NO_3^- + NO_2^-$). Ammonium is measured using the salicylate nitroferricyanide method, and $NO_3^- + NO_2^-$ is measured using cadmium reduction and sulfanilamide, both on an autoanalyzer. Due to resource constraints, NEON samples are not run with and without the cadmium reduction step, thus all measurements for nitrate include nitrite. However, nitrite is often a minor and short-lived constituent of natural soils (Hart et al., 1994); hereafter, we refer to $NO_3^- + NO_2^-$ as nitrate. Difference between final and initial inorganic N is net N mineralization, while difference in final versus initial nitrate is net nitrification. Data are reported as milligrams N per liter KCl (mg-N/L). In order to blank-correct the data and then calculate milligrams N per kilogram dry soil (per day for net rates), data must be combined from several tables that come in the data product download, followed by a series of basic calculations. The NEON developed a simple R package to perform these calculations (https://github.com/NEON-Science/NEON-Nitrogen-Transformations). Data are generally available within 6 months of sample collection on the NEON data portal, along with field and laboratory metadata, sampling and analytical protocols, and other documentation.

2.2. Challenges to Continental-Scale Inorganic N Sampling

To our knowledge, this is the first time a monitoring network has attempted repeated, standardized, continental-scale soil inorganic N sampling, and it has not been without challenges. The NEON employs a large cohort of technicians every season, and while most have some experience with ecology or field work, rarely is this specific to soils or biogeochemistry. The NEON has learned that all details required to ensure high-quality soil inorganic N pool and flux data must be included explicitly in protocols and training materials. Examples of details that were added over time include setting a minimum and maximum distance between the initial and incubated core, as well as a requirement for similar depths (within 5 cm) and adding detailed instructions for laboratory equipment preparation and clean lab techniques. Moreover, the Domain Support Facility (DSF) laboratories where extractions occur are set up for general ecological field work, not biogeochemistry per se, and the in-house water purification systems only achieve Type II (1 MOhm) grade. This water was used in 2017 to create KCl solutions but, upon examination of the first year of data, it was clear that certain DSF purification systems were not removing enough NH_4^+ or NO_3^- and blanks were unacceptably high. To solve this, all DSFs began purchasing Type I (18.2 MOhm) water to use for solution preparation and final equipment rinses starting in 2018.

Another issue involves "negative" sample concentrations after accounting for the inorganic N in analytical blanks. After observing this in several sites, troubleshooting revealed the cause to be nitrite contamination (up to 0.15 mg N/L) in certain batches of KCl used for extractions. In the presence of acidic soil, NO_2^- is abiotically

converted to nitrogen gasses and lost (chemodenitrified), but this process does not occur in blanks (Homyak et al., 2015). This often leads to significant negative $NO_3^- + NO_2^-$ values upon blank-correction for highly contaminated KCl batches and greatly reduces sensitivity of the method in low inorganic N concentration NEON sites where nitrate-N tends to be <0.05 mg/kg, about one third of the total. NEON continues to troubleshoot this issue. It has been difficult to source KCl powder that is consistently low in nitrite, so NEON has considered options such as in-house purification of KCl or custom order of KCl free from contaminants—but these solutions have proved challenging. For the 2022 field season, the workflow was to test each batch of KCl ordered from standard vendors, then use only batches where nitrite concentrations were verified to be low (generally <0.02 mg NO_2^--N/L, although a few batches were used with up to 0.05 mg NO_2^--N/L due to supply chain issues). The plan is to continue with this procedure unless or until an alternate solution can be implemented. Any sample values that blank-correct negative beyond the level of background noise, which is considered less than −0.02 mg N/L (more than twice the method detection limit of 0.01 mg/L), are flagged in the data with "blanks exceed sample value." Below, we discuss possible approaches to account for these values.

2.3. Data Download and Blank-Corrections

To examine the inorganic N data, we downloaded DP1.10086.001 soil physical and chemical properties, periodic from the NEON data portal on 9 June 2022 (NEON, 2022a) using the neonUtilities R package (Lunch et al., 2022). We chose to download "RELEASE-2022" data that includes a static version of inorganic N measurements from 2017 to 2020. More information about the sites included in the download along with relevant site metadata are summarized in Table S1 in Supporting Information S1. This includes fine root biomass fraction in the top 30 cm of soil, derived from DP1.10066.001 root biomass and chemistry, Megapit (NEON, 2022b). We used the *def.calc.ntrans* function in the neonNTrans R package (Weintraub, 2017) to calculate blank-corrected inorganic N concentrations and net N transformation rates on a per soil mass basis. Expressing pools and rates per gram soil is common practice in the biogeochemical literature (Liu et al., 2017), should be reflective of what plants and microbes "see" when competing for N, and is relevant to N losses to water and the atmosphere. Recent work has linked soil C with variation in net N transformation rates (Gill et al., 2023; Liu et al., 2017), and it could be interesting to examine rates per gram soil C. However, this would require assumptions and extrapolations, as the peak green sampling bout is the only one where total soil C is measured on the NEON soil cores, thus this would introduce more uncertainty (especially for organic horizons). We avoid this given our aim to assess sources of variation in the context of the NEON sampling design but acknowledge this could be relevant for future studies.

Given the issues with high NO_2^- blanks described above, we explored four possible ways to account for negative blank-corrected concentration data, as outlined in Table 1. Each has certain advantages, disadvantages, and assumptions. In the end, we used Option 4 because it yields a larger data set with all sites and samples retained, while avoiding substantial inflation with zeros (Figure S1 in Supporting Information S1). We suggest that the combination of a quantile regression for nitrate to estimate and add back the NO_2^- from KCl lost from samples but not blanks, plus use of an empirical cutoff for blank values for NH_4^+ (Figure S2 in Supporting Information S1) produces a data set highly suitable for further analyses. However, we recognize that there are other ways these data can (and will) be analyzed. Over time, these issues should become less prominent as NEON is now paying attention to and remediating the nitrite contamination issue.

2.4. Mixed Model Analyses to Examine Sources of Variation

Taking the resulting data set generated from the Option 4 blank-correction procedure, we fit mixed-effects models to test which aspects of the NEON sampling design were associated with variation in inorganic N dynamics. We fit separate models for the following response variables, leading to four final models, one each for: net N mineralization, net nitrification, and the initial concentrations of NH_4^+ and NO_3^- measured in the field. The net N mineralization and net nitrification data were strongly skewed and included large positive and negative values, so we used a "neglog" transformation to reduce skewness while preserving the sign of each value (Whittaker et al., 2005). Positive values were summed with one prior to a natural log transformation, and negative values were multiplied by negative one and summed with one prior to a natural log transformation, and were again multiplied by negative one to preserve the negative sign of the original value (Whittaker et al., 2005). The NH_4^+ and NO_3^- concentrations were natural log-transformed.

In each model, we included soil horizon type (organic or mineral), seasonal timing, and horizon by timing interactions as candidate fixed effects along with candidate random effects to account for the spatial structure of the

Table 1
Four Possible Ways to Adjust for Negative Blank-Corrected Concentration Data in National Ecological Observatory Network (NEON) Potassium Chloride Extraction Samples

Option description	Mechanics	Pros and cons
Option 1: drop samples that blank-correct to less than −0.02 mg N/L	Use *def.calc.ntrans* function with "drop flags" parameters set to exclude samples flagged as "blanks exceed sample value"	Conservative with respect to removing values that are negative, but liberal insofar as samples with high blanks remain as long as the blank-corrected value is positive
Option 2: set blank-corrected negative values to zero, no matter how negative they are	Use *def.calc.ntrans* without "drop flags" settings. The default behavior of the function is to convert all negative values to zero	Most liberal approach as all data are retained, but inflates the data set with many zeroes (Figure S1 in Supporting Information S1)
Option 3: exclude samples associated with high blanks, then set remaining blank-corrected negative values to zero	Drop all samples associated with blanks that exceed 0.13 mg N/L for $NO_3^- + NO_2^-$ and 0.1 for NH_4^+, then use default behavior of *def.calc.ntrans* to set negative values to zero	Leads to the smallest data set as certain sites are removed completely given site-specific issues with contaminated water or batches of KCl (Figure S1 in Supporting Information S1)
Option 4: for $NO_3^- + NO_2^-$, apply a quantile regression to estimate and add back blank concentrations; for NH_4^+, drop samples that blank-correct to less than −0.06 mg N/L	Correct for the systematic trend of increasingly negative blank-corrected values with increasing blank $NO_3^- + NO_2^-$ concentrations (Figure S2 in Supporting Information S1). Estimate slope of the 1% quantile regression (−0.75), then add this fraction of blank NO_2^--N back to sample values. For NH_4^+, use a threshold for discarding negative values (−0.06 mg N/L), acknowledging idiosyncratic analytical error (Figure S2 in Supporting Information S1)	Allows for the largest data set, all sites and majority of samples retained, yet the data set is not inflated with zeros

NEON sampling design. Subplot, plot, site, and domain were all coded as unique spatially explicit variables (i.e., there were observations from 1,328 subplots, 349 plots, 35 sites, and 19 domains). Given the NEON design of random stratification of plots within sites across NLCD vegetation class, this parameter was also included as a random effect with 10 levels: deciduous forest, pasture/hay, cultivated crops, shrub/scrub, evergreen forest, mixed forest, grassland/herbaceous, woody wetlands, dwarf scrub, and sedge/herbaceous. NLCD class varies at the plot scale, but not within plots. Because seasonal timing of sampling intentionally differed among sites as a function of climate (e.g., rainfall transitions in some sites compared to spring and fall in others), we generated a timing variable that was more comparable among sites. Specifically, seasonal timing was recoded from the initial five levels (peak greenness, winter/spring transition, fall/winter transition, dry/wet, and wet/dry) to a simpler scheme, where dry/wet and winter/spring were recoded as "ramp up" and wet/dry and fall/winter were recoded as "ramp down," and were then compared relative to "peak greenness." In the initial phase of analysis, we explored adding other fixed effects such as mean annual temperature, soil moisture, pH, and soil organic C concentration, but they resulted in only weak increases in explanatory power of the statistical models (no more than a 0.1 increase in R^2). This was not pursued further in order to focus on evaluating components of the sampling design.

We fit the linear mixed models using the lmer function in the "lme4" package in R (Bates et al., 2015). Singular fits were obtained in some cases where a random effect had variance near zero, so models were refit after removing that particular random effect. We assessed variable importance by plotting estimates of fixed effects coefficients, the standard deviations of random effects, and approximate *p*-values of fixed effects using a Type II Wald chi-square test with the ANOVA function in the "car" package (Fox & Weisberg, 2018). We report the marginal and conditional R^2 values, which represent the approximate variance explained by fixed and fixed plus random effects, respectively (Nakagawa et al., 2017). We evaluated model assumptions using plots of residual versus predicted values and sample quantiles versus theoretical quantiles. Moderate heteroscedasticity and deviations from normality were observed, so models should be treated with caution.

2.5. ESM Outputs and Data–Model Comparison

We used modeled monthly rates of net N mineralization from the CESM2, from simulations conducted as part of the Coupled Model Intercomparison Project 6 (CMIP6; O'Neill et al., 2016). Specifically, we used output from the Shared Socioeconomic Pathway (SSP) scenario 5, Representative Concentration Pathway (RCP) 8.5,

which represents the high end of plausible future pathways in terms of radiative forcing (O'Neill et al., 2014). Compared with CESM1, the land model used in CESM2 (Community Land Model version 5, CLM5) includes a more detailed and dynamic representation of the ecosystem N cycle (Lawrence et al., 2019). For example, CLM5 allows for flexible plant tissue stoichiometry, interactions between plant N and C, and more dynamic N fixation responses. Observations were compared with the simulated variable "fNnetmin" (i.e., net N mineralization) from soil and litter pools across the whole soil profile. Thus, modeled rates are not a perfect match with the NEON observations, which were collected from soil to 30 cm depth. However, there is a wealth of data to suggest that rates of N mineralization decline substantially with depth (Chen et al., 2019; Darby et al., 2020; Dessureault-Rompré et al., 2016). Since it is reasonable to assume that surface soils are making a significant contribution to modeled rates, the data–model comparison should prove useful. Moreover, temporal comparisons are still quite relevant; if the model represents seasonal patterns in N mineralization correctly, the temporal trends in data versus model should align, even if absolute magnitudes differ. Model data have a nominal 100 km spatial resolution. Site-level net N mineralization data were extracted based on latitude and longitude for each CESM2 grid cell that contained a site. We used the "ncdf4" package v.1.17 in R to work with netCDF files (Pierce, 2021) and the "raster" package v.3.4–5 (Hijmans, 2022) to match site locations to grid cells.

In order for the data set of net N mineralization observations to be comparable to CESM2 outputs, measurements were converted from a per-kg soil basis to per area fluxes, according to the following equation:

$$N_{min} \text{ areal basis} \left(g\,N\,m^{-2}\,day^{-1} \right) = N_{min} \text{ concentration (mg/kg)} \times \text{bulk density} \left(Mg/m^3 \right)$$
$$\times \text{ sample depth (m)} \times (1 - \text{percent coarse fragments} \times 0.01) \tag{1}$$

This required bulk density and percent coarse fragments (particles >2 mm diameter) for each sample. For a subset of plots used for periodic soil sampling, plot-level measurements of these variables were available by downloading DP1.10047.001 Soil physical and chemical properties, distributed initial characterization (NEON, 2022c). This data product includes physical and geochemical properties from 8 to 26 soil pits, generally 1-m deep, dug by the U.S. Department of Agriculture Natural Resource Conservation Service at each site to characterize the diversity of local soils. For this effort, soils were sampled by pedogenic horizon; to convert these data to a fixed depth interval (0–30 cm) to better match inorganic N measurements, data from each pit were separated into "mineral" and "organic" horizons, then the "slab" function in the aqp package (Beaudette et al., 2022) was used to estimate 0–30 cm values. Where periodic sampling and initial characterization data were measured in the same plots, we assigned bulk density and coarse fragment data to all inorganic N concentration values from that plot. For inorganic N samples from plots that lack these measurements, we estimated bulk density using a pedotransfer function (Drew, 1973; Guevara et al., 2020):

$$\text{Bulk density} \left(Mg/m^3 \right) = 1/(0.6268 + 0.0361 \times \text{organic carbon} (\%) \times 1.724) \tag{2}$$

As percent coarse fragments cannot be estimated from a pedotransfer function, we instead calculated mean of percent coarse fragments by horizon type (mineral vs. organic) and NLCD cover classes within each site, then assigned these values to inorganic N data within the same site, horizon, and NLCD cover. Net N mineralization rates were thus estimated for all organic and mineral horizon samples on an areal basis. Then, horizon values were summed for each within-plot sampling location where both were present. Finally, we took the average and standard error for each site and temporal sampling bout ($n = 89$), assigned month to each bout of NEON sampling, and assigned dominant NLCD class at the site level according to whichever cover class accounted for the greatest area of the NEON soil plots.

First, we examined the seasonal time course of net N mineralization in observations compared to modeled results from the same year to visually assess how well the two agreed. Then, we plotted average net N mineralization by month and site in observation versus model data and examined the relationships between them using several metrics: Spearman rank correlation, a linear model, and a mixed model with site as a random effect. We calculated the Nash-Sutcliffe efficiency of the data–model comparison using the NSE function in the "hydroGOF" package (Zambrano-Bigiarini, 2020) and the root mean square error (RMSE) using the rmse function in the "Metrics" package (Hamner & Frasco, 2018). Lastly, we extracted the residuals from the linear model and used analysis of variance to see if residuals could be explained by dominant NLCD class or seasonal timing. All analyses and data visualizations were performed using R Statistical Software (v 4.0.5; R Core Team, 2021) and statistical tests were considered significant with p-values <0.05.
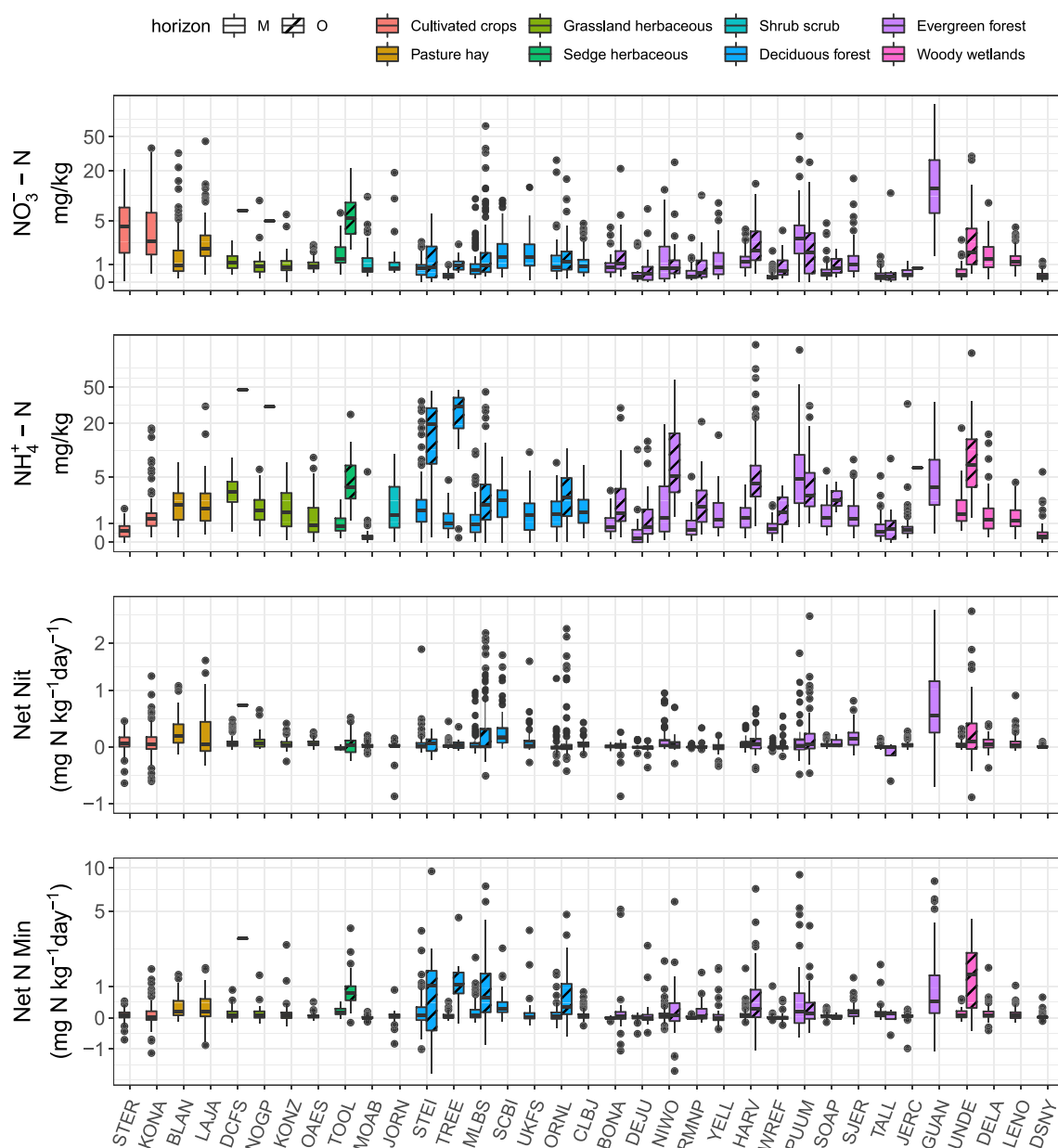
Figure 2. Boxplots of inorganic nitrogen (N) pools and net transformation rates across National Ecological Observatory Network (NEON) sites, with colors for dominant National Land Cover Database (NLCD) class. Diagonal lines within boxes denote data from surficial organic horizons. Net Nit, net nitrification; Net N Min, net N mineralization. Note log scale of *y* axis. Sites are arranged in increasing order of mean annual temperature within each NLCD class. For more information on site characteristics, see Table S1 in Supporting Information S1.

3. Results

The 35 NEON sites examined here displayed wide variation in rates of net N transformations and inorganic N pool sizes (Figure 2). Overall, organic horizons showed higher rates and pool sizes compared to adjacent mineral horizons when expressed on a weight basis. The largest portion of variance in soil extractable pools and net N transformation rates was explained by site, except for net N mineralization where domain was slightly statistically more important (Table 2 and Figure 3b). Plot explained the next largest fraction of variance for most variables, followed by domain, with subplot and NLCD class contributing the least explanatory power. The fixed effects did not explain a high proportion of the variation compared to the random effects, and each model had a substantiation amount of variation in residuals. For the fixed effects, horizon was important for all four variables, with

Table 2

Variance Explained (%) at Each Level of the Linear Mixed Model for Each of the Variables Classified as Random Effects, As Well As for All Random Effects Combined (i.e., Subplot, Plot, Site, Domain, and NLCD Class) or All Fixed Effects Variables Combined (i.e., Horizon Type, Sample Timing, and Their Interaction)

|	NO_3^-	NH_4^+	Net nit.	Net N min.
Subplot			5	4
Plot	12	15	14	6
Site	27	18	15	5
Domain	12	8	9	7
NLCD class	2	1		2
All random Effects	53	42	43	24
All fixed effects	4	16	0	1
Total	57	58	43	25
Residual	46	49	56	67

Note. A blank cell means a variable was not a significant predictor. Net nit., net nitrification; net N min., net N mineralization; NLCD, National Land Cover Database; residual, unexplained variance in each model.

higher levels in organic layers compared to mineral layers on a per mass basis (Figure 3a). Soil NH_4^+ concentrations and net mineralization rates were also influenced by seasonality. Soil NH_4^+ concentrations displayed lowest values during peak greenness compared to the transition seasons (e.g., positive coefficients for ramp up and ramp down when compared to peak greenness, Figure 3a), while the net rates were higher during the ramp up and slower during the ramp down compared to peak greenness—but only in organic horizons.

Most NEON sites displayed some degree of seasonality in net N mineralization rates, and whether temporal patterns aligned between measured and modeled data depended on the site (Figure 4). In some cases, dynamics observed in the field were captured reasonably well by CESM2, such as at the sites STER, HARV, and SCBI (see Table S1 in Supporting Information S1 for site names and metadata). For other sites, the temporal patterns did not coincide, for example, in GUAN, DCSF, and WREF. There was no clear visual indication that the model was better at capturing seasonality in certain types of ecosystems.

The Spearman rank correlation between mean observed net N mineralization and modeled net N mineralization in the same month was significant (*p*-value <0.001, $r = 0.38$; Figure 5), as was the linear correlation between modeled versus observed values ($Nmin_{obs} = Nmin_{mod} \times 0.323 + 0.022$, $p = 0.003$, $r^2 = 0.09$), but the explanatory power was low. A higher fraction of the variation could be explained by using a mixed model accounting for site as a random effect ($p = 0.004$, conditional $r^2 = 0.22$). The RMSE of the data–model comparison was 0.05 g N m^{-2} day^{-1}, and the Nash-Sutcliffe efficiency was −0.51. Analysis of variance
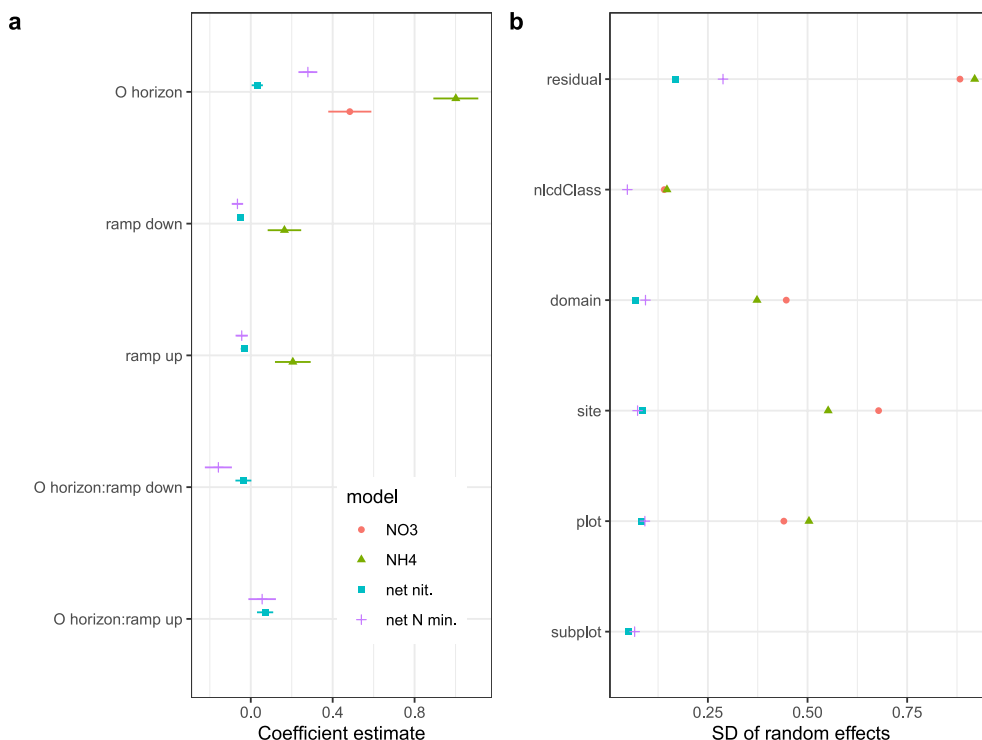
Figure 3. Estimated coefficients of fixed effects (a) and standard deviations of random effects (b) in the four separate mixed-effects models of soil net nitrogen mineralization (net N min.), net nitrification (net nit.), and initial ammonium (NH_4) and nitrate (NO_3) concentrations measured in field samples. For the fixed effects (a), the row for O horizon reflects the coefficient in relation to the mineral horizon, and ramp up or ramp down rows reflect coefficients during transition season sampling in relation to the peak green period.
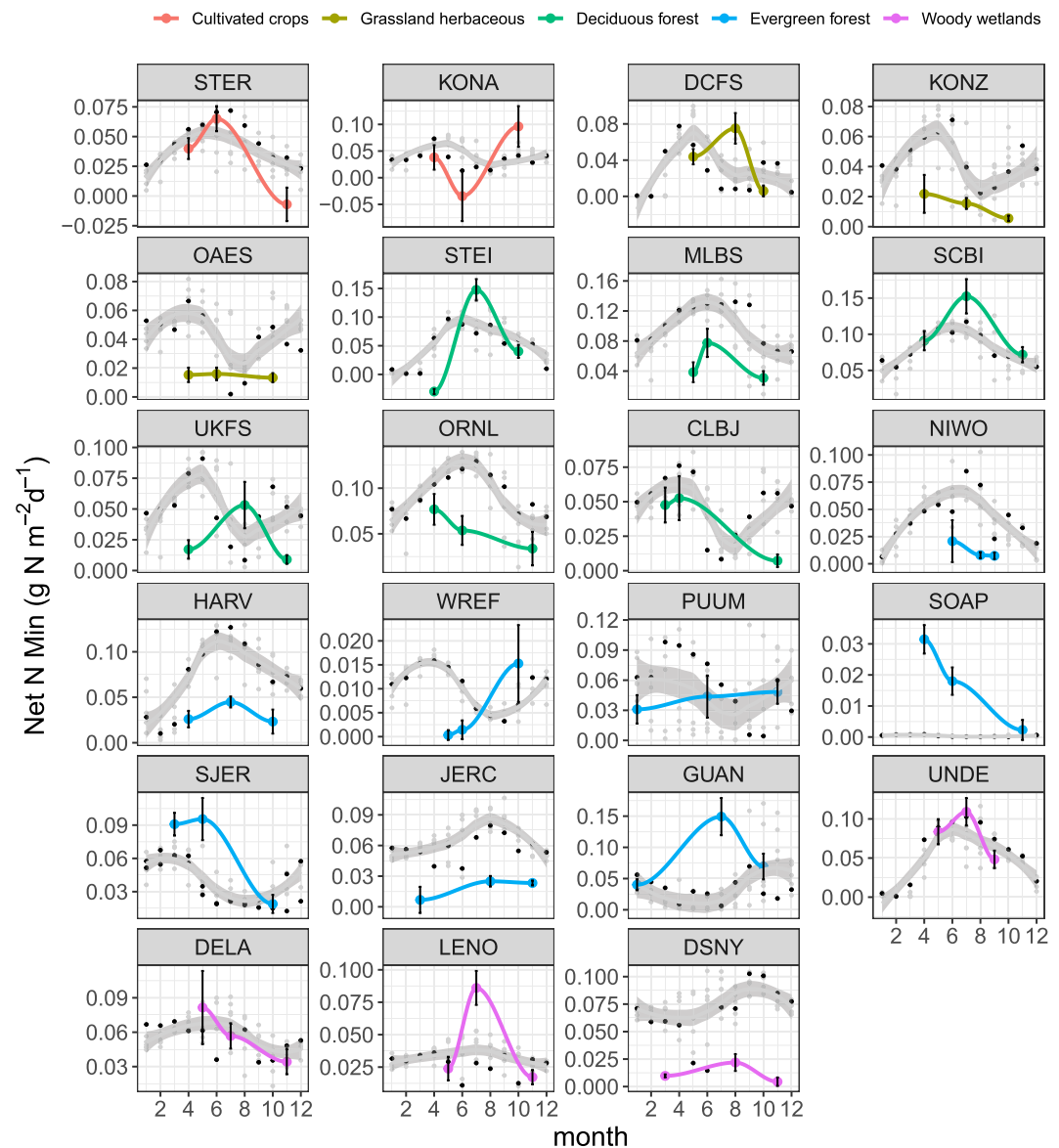
Figure 4. Measured and simulated seasonal patterns of soil net nitrogen (N) mineralization (Net N Min) by site. Colored circles are mean measured rates, color coded by dominant National Land Cover Database class, with vertical lines indicating standard errors ($n = 30$ unless bouts had incomplete sampling). Black and gray circles are simulated values from the Community Earth System Model 2 (CESM2) from 2015 to 2020 in years where measurements were and were not made, respectively. Gray lines show the confidence interval of the simulated temporal trends. Sites are arranged in increasing order of mean annual temperature within each land-cover class.

suggested that dominant NLCD cover was a significant predictor of data versus model residuals, but only for the "Pasture Hay" class, which had larger residuals compared to other land-cover types ($p = 0.006$). The "ramp down" period had marginally significantly lower residuals compared to peak greenness ($p = 0.08$), but overall this statistical model had low explanatory power ($r^2 = 0.11$, $p = 0.03$).

4. Discussion

Despite the importance of N mineralization and nitrification for terrestrial N cycling, we have a poor understanding of the controls on these processes at a continental scale. Data collected by NEON provide a potential opportunity to make progress in this area, and the standardized data sets can be useful for validating and improving
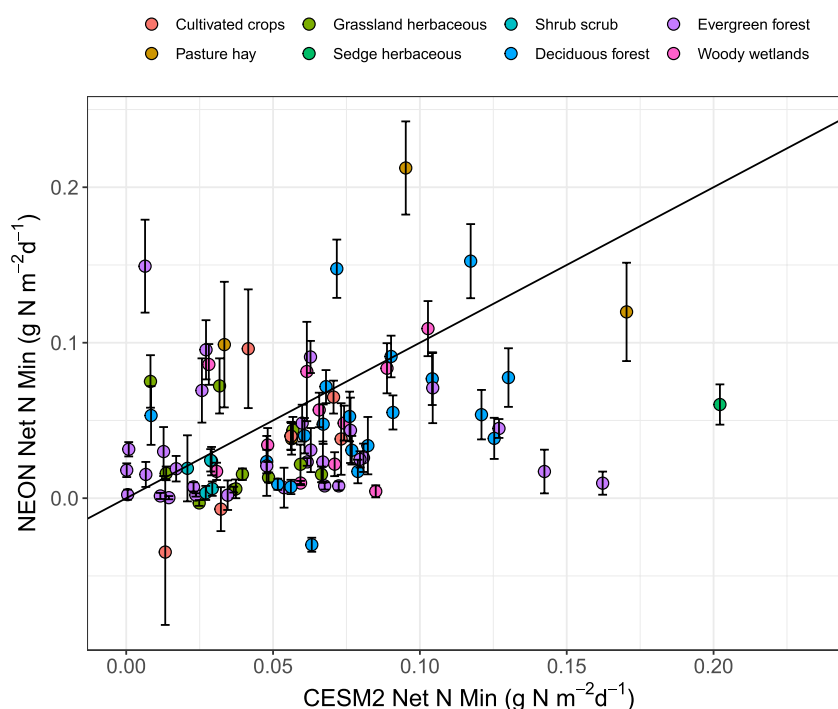
Figure 5. Net nitrogen (N) mineralization rates measured by the National Ecological Observatory Network (NEON) compared to simulated values from Community Earth System Model 2 (CESM2) at the same site, in the same month and year. The 1:1 line is displayed for reference. Vertical lines are standard errors of measurements.

predictive models. The NEON measurements capture high spatiotemporal heterogeneity in inorganic N pools and fluxes, some of which are predictable as a function of the spatial structure of sampling and core aspects of the NEON design.

4.1. Variation in N Pool and Net Flux Rates Across Sites

Certain types of sites conformed to expectations. For example, those with greatest soil nitrate pools and highest net N transformation rates were either in the tropics (GUAN) or under agricultural management (KONA, STER). Most forested sites outside the tropics had larger ammonium pools and very little nitrate, which is typical (although on its own, not diagnostic) for N-limited ecosystems. In contrast, some sites showed unexpected patterns; for example, relatively large soil inorganic N pools and high net transformation rates in the Alaskan tundra (TOOL) where we expected very low N availability, and low net N transformation rates and N pool sizes in subtropical woody wetlands in the Everglades (DSNY). Sites with low inorganic N pools might have low rates of N transformations, or alternatively, high plant N uptake, and this question could potentially be further investigated using additional NEON data.

4.2. Sampling Design Captures Important Sources of Variation, But Still Much Is Unexplained

Given that this is a first of its kind effort, it is important to evaluate the NEON design to ensure key sources of variation are being captured. According to the mixed model analysis, the nested components of the NEON design (Figure 1) all have some explanatory power for inorganic N pools and processes, suggesting that the sampling approach is generally robust. Sites tend to account for the most variation (Table 2), hence adding more sites could yield more insight into soil N cycling compared to adding more plots within sites or subplots within plots. While NEON is unlikely to make this expansion in the near future, and adding more sites is no assurance that fundamental understanding will increase, if others use NEON's standardized, publicly available protocols and analysis methods, the community can assess whether new insight can be gained by filling in geographic, climatic, and other kinds of gaps. The next most important source of variation is at the level of plots, which suggests that NEON should preserve the existing number of plots even if it means sacrificing within-plot replication. This is

helpful given budgetary and other logistical pressures and can help to focus efforts when sampling resources are strained (e.g., during the COVID-19 pandemic). Not only are these insights useful for NEON, but also potentially for researchers seeking to set up cross-site soil monitoring efforts or establish within-site sampling designs.

Surficial organic (O) horizons tended to have greater extractable inorganic N pools and higher mineralization rates compared to mineral horizons on a per weight basis (Figures 2 and 3). This is consistent with findings from previous studies, which tend to be limited to single sites or regions (Darby et al., 2020; Hart & Gunther, 1989) and implies that the pattern holds when looking across a wide variety of ecosystems. Given higher N concentrations and faster net mineralization rates per unit soil, organic horizons are likely to be "hotspots" for provision of mineral N and it may be useful for investigators to separate organic and mineral layers, even if otherwise pursuing fixed-depth sampling, to understand the contributions of the organic layer to soil nutrient availability. Moreover, as the climate warms and dries, organic horizons are likely to be reduced in areal density due to increased decomposition rates (Melillo et al., 2017), as their persistence is governed by temperature and moisture constraints as opposed to physical protection of organic matter as in the mineral soil (Schmidt et al., 2011). Because these organic horizons have high rates of microbial decomposition and inorganic N production, especially in the early part of the growing season (Figure 3a, ramp up), their decline might exacerbate N limitation. On the other hand, there could be counterbalancing forces in the kinds of ecosystems with thick organic horizons (e.g., high latitude or altitude), such as longer growing seasons and an expansion of the active layer, that could yield higher N availability in mineral layers. The net effects on N availability will be important to assess going forward, and NEON data can be leveraged for this.

There was a significant amount of unexplained variation in the statistical models used for predicting inorganic N pools and net flux rates, similar to previous findings (Colman & Schimel, 2013; Li et al., 2020). Clearly, understanding the sources of this unexplained variation in N dynamics remains a ripe area for research, and the NEON samples provide a unique opportunity to integrate a suite of data sets collected at a variety of spatiotemporal scales to assist in uncovering what drives these patterns. Our exploratory analyses suggested that snapshots of soil moisture and temperature (i.e., measurements at the time of soil sampling) did not provide additional statistical insight. However, NEON monitors soil moisture and temperature continuously via automated sensors, and these temporally resolved measurements might shed light on controls on N pool sizes and transformation rates. Sensor data could be especially useful if paired with information on microbial community abundance and composition, which NEON provides. We did not find NLCD class to be a strong driver of N cycle variation, yet this is a very coarse metric of vegetation cover. Within the same NLCD class, various attributes of plant communities can strongly influence N dynamics (e.g., mycorrhizal type, Phillips et al., 2013). Information on plant biomass, composition, phenology, or chemistry, derived from on-the-ground measurements and/or high-resolution airborne remote sensing, are also available from NEON and may provide new insights into how vegetation influences soil N dynamics across diverse ecosystems.

4.3. Simulations Do Not Capture Observed Rates of Net N Mineralization

Overall, we observed a relatively poor fit between NEON measured and CESM2 modeled net N mineralization, suggesting room for growth in our understanding of the soil N cycle and its representation in large-scale models. There was a large RMSE and a negative Nash-Sutcliffe efficiency value suggesting that the process model was worse for prediction than taking only the mean of the observed values. However, the Spearman rank correlation was higher than linear model fits, suggesting that although prediction of the absolute values was difficult, there was some correspondence in the general patterns observed. As noted earlier, there are limitations in both the data and model side of the comparison. For observations, we had to make assumptions about bulk density and coarse fragment content in some plots, and soils below 30 cm depth were not included. For the modeled data, the observed climate during net N transformation measurements may have differed from the one simulated in the CESM2 outputs. Nonetheless, it is likely that the way N cycling is mechanistically represented in the model is not sufficient to capture spatial and temporal variation in N cycling processes. We were unable to explain much of the residual in the observations versus simulations linear model with dominant NLCD class or sample timing, but a next step would be to probe more systematically where and why the model does better or worse and seek to improve the underlying process representation from there. Additionally, NEON's standardized, continental-scale data on soil inorganic N pools and fluxes can be used to evaluate process-based model estimates of soil N_2O emissions. Even without the N_2O flux data, soil pools and net N transformation rates can

serve as standardized benchmarks and play important roles in the N₂O Model Intercomparison Project (NMIP, Tian et al., 2018; Xu et al., 2021).

4.4. Opportunities for Future Research and Engagement

It is critical to understand how cycling of a key limiting nutrient—N—will respond to increasing climatic and other ecological stressors. The NEON data provide a baseline for monitoring changes in soil N cycling going forward. Changes in the timing and magnitude of N pools and process rates will be mediated by complex, cross-scale interactions among different ecosystem components. NEON helps meet this challenge by monitoring many ecosystem dynamics at the same time. Nonetheless, there remain some obvious information gaps. This is where individual researchers can propose to collaborate with NEON (SanClements et al., 2020) to uncover new, continental-scale insights. For instance, we now recognize that minerals play an important role in soil N availability (Bingham & Cotrufo, 2016; Jilling et al., 2018). There is potential to not only leverage existing NEON data on soil mineralogy, geochemistry, and N cycle metrics but also propose new collaborative projects at NEON sites. Our ability to benchmark models and upscale measured rates would be improved by having estimates of bulk density and coarse fragments in all NEON soil plots. This is not part of the baseline NEON design, but could be added through an initiative driven by the scientific community.

There are several ways for researchers to engage with the NEON data sets described in this paper. This ranges from simply downloading and using data from the NEON portal, to taking NEON data tutorials, contributing code resources to the NEON code hub, or proposing collaborations to conduct additional research (Nagy et al., 2021; SanClements et al., 2020). To fully utilize these large, integrated, publicly available data sets, techniques such as machine learning, deep learning, information theory, and others will be needed to uncover previously hidden connections between N cycling and ecological dynamics. Moreover, the biogeochemistry research community will derive the most benefit by wholly embracing the FAIR data principles (Wilkinson et al., 2016), to ensure all derived data sets, code, and analyses are findable, reproducible, and enable new insights to build on the most recent contributions. By leveraging these unique data sets and opportunities, we can unlock new insights into the controls on soil N cycling.

Data Availability Statement

Measurements of inorganic N pools, fine root biomass by depth interval, and bulk density analyzed in this study are available by downloading "RELEASE-2022" of DP1.10086.001, DP1.10066, and DP1.10047, respectively, using the NEON data portal or NEON's application programming interface (API). Simulated net N mineralization rates from CESM2 can be found at https://doi.org/10.22033/ESGF/CMIP6.7768 (Danabasoglu, 2019). Derived data sets and all code used for data analysis and visualization are available at https://doi.org/10.6073/pasta/ec9e2e7c1fa40f2b61c2e930bd5a4ad1 (Weintraub-Leff et al., 2023).

References

Austin, A. T., Yahdjian, L., Stark, J. M., Belnap, J., Porporato, A., Norton, U., et al. (2004). Water pulses and biogeochemical cycles in arid and semiarid ecosystems. *Oecologia*, *141*(2), 221–235. https://doi.org/10.1007/S00442-004-1519-1

Balch, J. K., Nagy, R. C., & Halpern, B. S. (2020). NEON is seeding the next revolution in ecology. *Frontiers in Ecology and the Environment*, *18*(1), 3. https://doi.org/10.1002/FEE.2152

Barnett, D. T., Duffy, P. A., Schimel, D. S., Krauss, R. E., Irvine, K. M., Davis, F. W., et al. (2019). The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. *Ecosphere*, *10*(2), e02540. https://doi.org/10.1002/ECS2.2540

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaudette, D., Roudier, P., & Brown, A. (2022). aqp: Algorithms for Quantitative Pedology. R package version 1.42. Retrieved from https://CRAN.R-project.org/package=aqp

Bergström, A. K., & Jansson, M. (2006). Atmospheric nitrogen deposition has caused nitrogen enrichment and eutrophication of lakes in the northern hemisphere. *Global Change Biology*, *12*(4), 635–643. https://doi.org/10.1111/J.1365-2486.2006.01129.X

Bingham, A. H., & Cotrufo, M. F. (2016). Organic nitrogen storage in mineral soil: Implications for policy and management. *Science of the Total Environment*, *551–552*, 116–126. https://doi.org/10.1016/j.scitotenv.2016.02.020

Binkley, D., & Hart, S. (1989). The components of nitrogen availability assessments in forest soils. In *Advances in soil science* (pp. 57–112). Springer.

Bradford, M. A., Wood, S. A., Addicott, E. T., Fenichel, E. P., Fields, N., González-Rivero, J., et al. (2021). Quantifying microbial control of soil organic matter dynamics at macrosystem scales. *Biogeochemistry*, *156*(1), 19–40. https://doi.org/10.1007/S10533-021-00789-5

Bremner, J. M., & Keeney, D. R. (1966). Determination and isotope-ratio analysis of different forms of nitrogen in soils: 3. Exchangeable ammonium, nitrate, and nitrite by extraction–distillation methods. *Soil Science Society of America Journal*, *30*(5), 577–582. https://doi.org/10.2136/SSSAJ1966.03615995003000050015X

Brooks, P. D., Williams, M. W., & Schmidt, S. K. (1998). Inorganic nitrogen and microbial biomass dynamics before and during spring snowmelt. *Biogeochemistry*, *43*(1), 1–15. https://doi.org/10.1023/A:1005947511910

Brookshire, E. N. J., Gerber, S., Menge, D. N. L., & Hedin, L. O. (2011). Large losses of inorganic nitrogen from tropical rainforests suggest a lack of nitrogen limitation. *Ecology Letters*, *15*(1), 9–16. https://doi.org/10.1111/j.1461-0248.2011.01701.x

Chapman, S. K., Langley, J. A., Hart, S. C., & Koch, G. W. (2006). Plants actively control nitrogen cycling: Uncorking the microbial bottleneck. *New Phytologist*, *169*(1), 27–34. https://doi.org/10.1111/j.1469-8137.2005.01571.x

Chen, D., Saleem, M., Cheng, J., Mi, J., Chu, P., Tuvshintogtokh, I., et al. (2019). Effects of aridity on soil microbial communities and functions across soil depths on the Mongolian Plateau. *Functional Ecology*, *33*(8), 1561–1571. https://doi.org/10.1111/1365-2435.13359

Colman, B. P., & Schimel, J. P. (2013). Drivers of microbial respiration and net N mineralization at the continental scale. *Soil Biology and Biochemistry*, *60*, 65–76. https://doi.org/10.1016/J.SOILBIO.2013.01.003

Craine, J. M., Elmore, A. J., Wang, L., Aranibar, J., Bauters, M., Boeckx, P., et al. (2018). Isotopic evidence for oligotrophication of terrestrial ecosystems. *Nature Ecology & Evolution*, *2*(11), 1735–1744. https://doi.org/10.1038/s41559-018-0694-0

Danabasoglu, G. (2019). *NCAR CESM2 model output prepared for CMIP6 ScenarioMIP ssp585*. Version 20220420. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.7768

Darby, B. A., Goodale, C. L., Chin, N. A., Fuss, C. B., Lang, A. K., Ollinger, S. V., & Lovett, G. M. (2020). Depth patterns and connections between gross nitrogen cycling and soil exoenzyme activities in three northern hardwood forests. *Soil Biology and Biochemistry*, *147*, 107836. https://doi.org/10.1016/J.SOILBIO.2020.107836

Davidson, E. A. (2009). The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nature Geoscience*, *2*(9), 659–662. https://doi.org/10.1038/ngeo608

Davidson, E. A., Stark, J. M., & Firestone, M. K. (1990). Microbial production and consumption of nitrate in an annual grassland. *Ecology*, *71*(5), 1968–1975. https://doi.org/10.2307/1937605

Davies-Barnard, T., Meyerholt, J., Zaehle, S., Friedlingstein, P., Brovkin, V., Fan, Y., et al. (2020). Nitrogen cycling in CMIP6 land surface models: Progress and limitations. *Biogeosciences*, *17*(20), 5129–5148. https://doi.org/10.5194/BG-17-5129-2020

Dessureault-Rompré, J., Zebarth, B. J., Burton, D. L., & Grant, C. A. (2016). Depth distribution of mineralizable nitrogen pools in contrasting soils in a semi-arid climate. *Canadian Journal of Soil Science*, *96*(1), 1–11. https://doi.org/10.1139/cjss-2015-0048

Drew, L. A. (1973). Bulk density estimation based on organic matter content of some Minnesota soils, Minnesota Forestry Research Notes, 243. Retrieved from http://conservancy.umn.edu/handle/11299/58293

Du, E., Terrer, C., Pellegrini, A. F. A., Ahlström, A., van Lissa, C. J., Zhao, X., et al. (2020). Global patterns of terrestrial nitrogen and phosphorus limitation. *Nature Geoscience*, *13*(3), 221–226. https://doi.org/10.1038/s41561-019-0530-4

Durán, J., Morse, J. L., Groffman, P. M., Campbell, J. L., Christenson, L. M., Driscoll, C. T., et al. (2014). Winter climate change affects growing-season soil microbial biomass and activity in northern hardwood forests. *Global Change Biology*, *20*(11), 3568–3577. https://doi.org/10.1111/gcb.12624

Feng, Z., Rütting, T., Pleijel, H., Wallin, G., Reich, P. B., Kammann, C. I., et al. (2015). Constraints to nitrogen acquisition of terrestrial plants under elevated CO_2. *Global Change Biology*, *21*(8), 3152–3168. https://doi.org/10.1111/GCB.12938

Firestone, M., & Davidson, E. (1989). Microbiological basis of NO and N_2O production and consumption in soils. In M. Andreae & D. Schimel (Eds.), *Exchanges of trace gases between terrestrial ecosystems and the atmosphere*. John Wiley & Sons.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (3rd ed.). SAGE Publications, Inc. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., & Cosby, B. J. (2003). The nitrogen cascade. *BioScience*, *53*(4), 341–356. https://doi.org/10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., & Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, *141*(2), 287–299. https://doi.org/10.1007/s10584-017-1899-y

Gill, A. L., Grinder, R. M., See, C. R., Chapin, F. S., DeLancey, L. C., Fisk, M. C., et al. (2023). Soil carbon availability decouples net nitrogen mineralization and net nitrification across United States Long Term Ecological Research sites. *Biogeochemistry*, *162*(1), 13–24. https://doi.org/10.1007/S10533-022-01011-w

Groffman, P. M., Driscoll, C. T., Durán, J., Campbell, J. L., Christenson, L. M., Fahey, T. J., et al. (2018). Nitrogen oligotrophication in northern hardwood forests. *Biogeochemistry*, *141*(3), 523–539. https://doi.org/10.1007/s10533-018-0445-y

Gruber, N., & Galloway, J. N. (2008). An Earth-system perspective of the global nitrogen cycle. *Nature*, *451*(7176), 293–296. https://doi.org/10.1038/nature06592

Guevara, M., Arroyo, C., Brunsell, N., Cruz, C. O., Domke, G., Equihua, J., et al. (2020). Soil organic carbon across Mexico and the Conterminous United States (1991–2010). *Global Biogeochemical Cycles*, *34*, e2019GB006219. https://doi.org/10.1029/2019GB006219

Hamner, B., & Frasco, M. (2018). Metrics: Evaluation metrics for machine learning. R package version 0.1.4. Retrieved from https://CRAN.R-project.org/package=Metrics

Hart, S., & Gunther, A. (1989). In situ estimates of annual net nitrogen mineralization and nitrification in a subarctic watershed. *Oecologia*, *80*(2), 284–288. https://doi.org/10.1007/bf00380165

Hart, S., Stark, J., Davidson, E., & Firestone, M. (1994). Nitrogen mineralization and immobilization. In R. Weaver (Ed.), *Methods of soil analysis, Part 2: Microbiological and biochemical properties* (pp. 985–1018). Soil Science Society of America.

Hijmans, R. J. (2022). raster: Geographic data analysis and modeling. R package version 3.5-15. Retrieved from https://CRAN.R-project.org/package=raster

Hinckley, E. L. S., Bonan, G. B., Bowen, G. J., Colman, B. P., Duffy, P. A., Goodale, C. L., et al. (2016). The soil and plant biogeochemistry sampling design for the National Ecological Observatory Network. *Ecosphere*, *7*(3), e01234. https://doi.org/10.1002/ECS2.1234

Homyak, P. M., Sickman, J. O., Miller, A. E., Melack, J. M., Meixner, T., & Schimel, J. P. (2014). Assessing nitrogen-saturation in a seasonally dry chaparral watershed: Limitations of traditional indicators of N-saturation. *Ecosystems*, *17*(7), 1286–1305. https://doi.org/10.1007/s10021-014-9792-2

Homyak, P. M., Vasquez, K. T., Sickman, J. O., Parker, D. R., & Schimel, J. P. (2015). Improving nitrite analysis in soils: Drawbacks of the conventional 2 M KCl extraction. *Soil Science Society of America Journal*, *79*(4), 1237–1242. https://doi.org/10.2136/SSSAJ2015.02.0061N

Jackson, R., Canadell, J., Ehleringer, J., Mooney, H., Sala, O., & Schulze, E. (1996). A global analysis of root distributions for terrestrial biomes. *Oecologia*, *108*(3), 389–411. https://doi.org/10.1007/bf00333714

Jilling, A., Keiluweit, M., Contosta, A. R., Frey, S., Schimel, J., Schnecker, J., et al. (2018). Minerals in the rhizosphere: Overlooked mediators of soil nitrogen availability to plants and microbes. *Biogeochemistry*, *139*(2), 103–122. https://doi.org/10.1007/s10533-018-0459-5

Kapnick, S., & Hall, A. (2010). Observed climate–snowpack relationships in California and their implications for the future. *Journal of Climate*, *23*(13), 3446–3456. https://doi.org/10.1175/2010JCLI2903.1

Keller, M., Schimel, D., Hargrove, W., & Hoffman, F. (2008). A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment*, *6*(5), 282–284. https://doi.org/10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*, 4245–4287. https://doi.org/10.1029/2018MS001583

LeBauer, D., & Treseder, K. (2008). Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, *89*(2), 371–379. https://doi.org/10.1890/06-2057.1

Li, Z., Zeng, Z., Tian, D., Wang, J., Fu, Z., Zhang, F., et al. (2020). Global patterns and controlling factors of soil nitrification rate. *Global Change Biology*, *26*(7), 4147–4157. https://doi.org/10.1111/gcb.15119

Liu, Y., Wang, C., He, N., Wen, X., Gao, Y., Li, S., et al. (2017). A global synthesis of the rate and temperature sensitivity of soil nitrogen mineralization: Latitudinal patterns and mechanisms. *Global Change Biology*, *23*(1), 455–464. https://doi.org/10.1111/GCB.13372

Lunch, C., Laney, C., Mietkiewicz, N., Sokol, E., & Cawley, K., & NEON (National Ecological Observatory Network). (2022). neonUtilities: Utilities for working with NEON data. R package version 2.1.4. Retrieved from https://CRAN.R-project.org/package=neonUtilities

Melillo, J. M., Frey, S. D., DeAngelis, K. M., Werner, W. J., Bernard, M. J., Bowles, F. P., et al. (2017). Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, *358*(6359), 101–105. https://doi.org/10.1126/science.aan2874

Nagy, R. C., Balch, J. K., Bissell, E. K., Cattau, M. E., Glenn, N. F., Halpern, B. S., et al. (2021). Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. *Ecosphere*, *12*(12), e03833. https://doi.org/10.1002/ECS2.3833

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/RSIF.2017.0213

NEON (National Ecological Observatory Network). (2022a). Soil physical and chemical properties, periodic (DP1.10086.001), RELEASE-2022 [Dataset]. Retrieved from https://data.neonscience.org

NEON (National Ecological Observatory Network). (2022b). Root biomass and chemistry, Megapit (DP1.10066.001), RELEASE-2022 [Dataset]. Retrieved from https://data.neonscience.org

NEON (National Ecological Observatory Network). (2022c). Soil physical and chemical properties, distributed initial characterization (DP1.10047.001), RELEASE-2022 [Dataset]. Retrieved from https://data.neonscience.org

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., et al. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. https://doi.org/10.1007/S10584-013-0905-2

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/GMD-9-3461-2016

Parton, W., Silver, W. L., Burke, I. C., Grassens, L., Harmon, M. E., Currie, W. S., et al. (2007). Global-scale similarities in nitrogen release patterns during long-term decomposition. *Science*, *315*(5810), 361–364. https://doi.org/10.1126/science.1134853

Payne, R. J., Dise, N. B., Field, C. D., Dore, A. J., Caporn, S. J. M., & Stevens, C. J. (2017). Nitrogen deposition and plant biodiversity: Past, present, and future. *Frontiers in Ecology and the Environment*, *15*(8), 431–436. https://doi.org/10.1002/FEE.1528

Phillips, R. P., Brzostek, E., & Midgley, M. G. (2013). The mycorrhizal-associated nutrient economy: A new framework for predicting carbon–nutrient couplings in temperate forests. *New Phytologist*, *199*(1), 41–51. https://doi.org/10.1111/nph.12221

Pierce, D. (2021). ncdf4: Interface to Unidata netCDF (version 4 or earlier) format data files. R package version 1.19. Retrieved from https://CRAN.R-project.org/package=ncdf4

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Reich, P. B., Hobbie, S. E., Lee, T., Ellsworth, D. S., West, J. B., Tilman, D., et al. (2006). Nitrogen limitation constrains sustainability of ecosystem response to CO_2. *Nature*, *440*(7086), 922–925. https://doi.org/10.1038/nature04486

Sanclements, M., Lee, R. H., Ayres, E. D., Goodman, K., Jones, M., Durden, D., et al. (2020). Collaborating with NEON. *BioScience*, *70*(2), 107. https://doi.org/10.1093/BIOSCI/BIAA005

Schimel, J. P., & Bennett, J. (2004). Nitrogen mineralization: Challenges of a changing paradigm. *Ecology*, *85*(3), 591–602. https://doi.org/10.1890/03-8002

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., et al. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, *478*(7367), 49–56. https://doi.org/10.1038/nature10386

Stanish, L. (2022). TOS protocol and procedure: SLS—Soil biogeochemical and microbial sampling. NEON.DOC.014048vO. NEON (National Ecological Observatory Network).

Sullivan, B., Smith, W., Townsend, A., Nasto, M., Reed, S., Chazdon, R., & Cleveland, C. (2014). Spatially robust estimates of biological nitrogen (N) fixation imply substantial human alteration of the tropical N cycle. *Proceedings of the National Academy of Science of the United States of America*, *111*(22), 8101–8106. https://doi.org/10.1073/pnas.1320646111

Sutton, M. A., Oenema, O., Erisman, J. W., Leip, A., van Grinsven, H., & Winiwarter, W. (2011). Too much of a good thing. *Nature*, *472*(7342), 159–161. https://doi.org/10.1038/472159a

Thomas, R. Q., Brookshire, E. N. J., & Gerber, S. (2015). Nitrogen limitation on land: How can it occur in Earth system models? *Global Change Biology*, *21*(5), 1777–1793. https://doi.org/10.1111/GCB.12813

Tian, H., Yang, J., Lu, C., Xu, R., Canadell, J. G., Jackson, R. B., et al. (2018). The global N_2O Model Intercomparison Project. *Bulletin of the American Meteorological Society*, *99*(6), 1231–1251. https://doi.org/10.1175/BAMS-D-17-0212.1

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., et al. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Applications*, *7*(3), 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Vitousek, P., & Howarth, R. (1991). Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry*, *13*(2), 87–115. https://doi.org/10.1007/bf00002772

Vitousek, P., & Matson, P. (1988). Nitrogen transformations in a range of tropical forest soils. *Soil Biology and Biochemistry*, *20*(3), 361–367. https://doi.org/10.1016/0038-0717(88)90017-x

Wang, Y. P., & Houlton, B. Z. (2009). Nitrogen constraints on terrestrial carbon uptake: Implications for the global carbon-climate feedback. *Geophysical Research Letters*, *36*, L24403. https://doi.org/10.1029/2009GL041009

Weintraub, S. R. (2017). neonNTrans: NEON nitrogen transformations. R package version 0.2.0.

Weintraub, S. R., Brooks, P. D., & Bowen, G. J. (2017). Interactive effects of vegetation type and topographic position on nitrogen availability and loss in a temperate montane ecosystem. *Ecosystems*, *20*(6), 1073–1088. https://doi.org/10.1007/s10021-016-0094-8

Weintraub-Leff, S. R., Hall, S. J., & Craig, M. (2023). NEON soil inorganic nitrogen measurements 2017–2020, derived data and code for Earth's Future manuscript ver 2. Environmental Data Initiative. https://doi.org/10.6073/pasta/ec9e2e7c1fa40f2b61c2e930bd5a4ad1

Whittaker, J., Whitehead, C., & Somers, M. (2005). The neglog transformation and quantile regression for the analysis of a large credit scoring database. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(5), 863–878. https://doi.org/10.1111/J.1467-9876.2005.00520.X

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Xu, R., Tian, H., Pan, N., Thompson, R. L., Canadell, J. G., Davidson, E. A., et al. (2021). Magnitude and uncertainty of nitrous oxide emissions from North America based on bottom-up and top-down approaches: Informing future research and national inventories. *Geophysical Research Letters*, *48*, e2021GL095264. https://doi.org/10.1029/2021GL095264

Zaehle, S., Jones, C. D., Houlton, B., Lamarque, J. F., & Robertson, E. (2015). Nitrogen availability reduces CMIP5 projections of twenty-first-century land carbon uptake. *Journal of Climate*, *28*(6), 2494–2511. https://doi.org/10.1175/JCLI-D-13-00776.1

Zambrano-Bigiarini, M. (2020). hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.4-0. https://doi.org/10.5281/zenodo.839854