Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems

Peter M. Vitousek 1, Duncan N. L. Menge 2, Sasha C. Reed 3
and Cory C. Cleveland 4

1 Department of Biology, Stanford University, Stanford, CA 94305, USA
2 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
3 US Geological Survey, Southwest Biological Science Center, Canyonlands Research Station, Moab, UT 84532, USA
4 Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT 59812, USA

New techniques have identified a wide range of organisms with the capacity to carry out biological nitrogen fixation (BNF)—greatly expanding our appreciation of the diversity and ubiquity of N fixers—but our understanding of the rates and controls of BNF at ecosystem and global scales has not advanced at the same pace. Nevertheless, determining rates and controls of BNF is crucial to placing anthropogenic changes to the N cycle in context, and to understanding, predicting and managing many aspects of global environmental change. Here, we estimate terrestrial BNF for a pre-industrial world by combining information on N fluxes with 15N relative abundance data for terrestrial ecosystems. Our estimate is that pre-industrial N fixation was 58 (range of 40–100) Tg N fixed yr\(^{-1}\); adding conservative assumptions for geological N reduces our best estimate to 44 Tg N yr\(^{-1}\). This approach yields substantially lower estimates than most recent calculations; it suggests that the magnitude of human alternation of the N cycle is substantially larger than has been assumed.

1. Introduction
Prior to the industrial revolution, biological nitrogen fixation (BNF) was the dominant source of reactive N to the biosphere. More recently, human activity has increased the creation of reactive N substantially, both intentionally through the synthesis of industrial N fertilizers and the cultivation of crops that support BNF, and also unintentionally via fossil fuel combustion [1,2]. To put the magnitude and consequences of this human perturbation to the N cycle in context, we need to understand background rates of BNF—and we need to understand the ecological regulation of BNF in a rapidly changing world. Our knowledge of the rates and regulation of BNF affects our ability to understand, predict and ultimately manage many components of human-caused environmental change, from increasing atmospheric CO\(_2\) and consequent climate change, to alterations of other major biogeochemical cycles, to aspects of land use change. For example, Hungate et al. [3] calculated that Intergovernmental Panel on Climate Change-based carbon (C) storage scenarios for terrestrial ecosystems in 2100 would require an additional 2.3–37.5 Pg of N, whereas reactive N supply would only increase by 1.2–6.1 Pg of N. Therefore, N supply will likely be insufficient to support predicted levels of biological C sequestration unless increasing atmospheric CO\(_2\) drives a substantial increase in BNF. On finer spatial and temporal scales, enrichment experiments in terrestrial ecosystems frequently show that elevated CO\(_2\) enhances primary production and C storage in the short term, but progressive N limitation ultimately constrains productivity and C storage [4,5], in large part because BNF generally does not replenish all of the reactive N that is sequestered in accumulating organic matter under elevated CO\(_2\).
From an ecological perspective, BNF raises a fundamental question: how can widespread N limitation to primary production, biomass accumulation and other ecosystem processes [6,7] co-occur with the widespread occurrence of organisms capable of acquiring the essentially limitless pool of atmospheric \( \text{N}_2 \) [8,9]? Put another way, why do organisms with the ability to fix N not realize a competitive advantage when N supply limits non-fixers, and as a by-product of their activity reduce N limitation to a marginal and/or transient phenomenon?

In this paper, we address the rates and ecological regulation of BNF. We begin with a brief discussion of the ways that genetic and molecular tools have expanded the list of organisms and symbioses known to carry out BNF. We then draw on recent global N flux estimates and an understanding of N isotopes to propose a new estimate of pre-industrial rates of BNF, the rates that occurred in little-managed ecosystems before the global N cycle was transformed by human activity. It is difficult to assign a latest date to this estimate; 1750 (prior to substantial fossil fuel N fixation) is a reasonable possibility, though certainly N fixation in many regions had been altered by human land use at that time. Next, we discuss observations and experiments that provide insights into the patterns and controls of BNF, as these operate on genetic, physiological and ecological levels [10–12]. Finally, we summarize results from several models that offer powerful means for addressing ecosystem and regional controls of BNF on land. While the control of BNF in marine ecosystems represents a lively, important and controversial area of research, our analysis is confined to terrestrial systems. Moreover, our focus is on BNF in little-managed terrestrial ecosystems; we see little that can be added to Herridge et al.’s [13] thorough analysis of BNF in agricultural and intensively grazed ecosystems.

2. The growing diversity of recognized nitrogen fixers

Biological N fixation reflects the activity of a phylogenetically diverse list of bacteria, archaea and symbioses. Genetic and molecular techniques now provide efficient means to identify organisms with the potential to fix \( \text{N}_2 \), and the application of these techniques in an array of environments has broadened considerably our understanding of the suite of organisms that can carry out BNF. In the past, most evaluations of terrestrial BNF considered rhizobial and actinorhizal symbioses; free-living and symbiotic cyanobacterial fixers; and free-living and symbiotic (or associated) heterotrophic bacteria. More recent research has demonstrated that a dizzying array of microbes are capable of BNF, including archaea as well as many previously undiscovered bacteria [14]. Even the best-known system, the nodulated legume/rhizobial symbiosis, has been found to be much more diverse than previously understood—with new discoveries ranging from additional microbial partners (e.g. *Burkholderia*) that nodulate legumes and fix \( \text{N}_2 \) effectively [15] to a diversity of pathways by which nodules themselves are initiated and infected [16].

Our expanded ability to identify organisms capable of BNF has outpaced increases in our understanding of how BNF is regulated in these different organisms. Given that the genetic and biochemical machinery that carries out BNF is highly conserved across the broad array of organisms that express it, the environmental and physiological conditions that could constrain BNF are similar across taxa. These conditions include low levels of available energy, high levels of \( \text{O}_2 \) at the locus of fixation, high levels of reactive N in the milieu and low availability of other resources (especially phosphorus (P), iron (Fe), potassium (K) and molybdenum (Mo)). However, predicting which subset of these controls is most relevant in different taxa and environments remains a challenge. For example, there is evidence for variation in how reactive N affects BNF across N-fixing symbioses between plants and bacteria. Unlike relatively well-known legume/rhizobial systems, some actinorhizal symbioses (which involve a diverse group of angiosperms that form nodulated symbioses with *Frankia*) appear to have weak or no feedback regulation from reactive N in the environment; they continue to nodulate and fix N even when N availability in the environment is high [17–19]. Some legumes also appear to experience weak feedback from reactive N to rates of fixation [20]. How these variations and others map onto the broad diversity of recently identified N fixers is not yet known—and we are not likely to find out as long as most research into symbiotic BNF focuses on two model legume species [21].

3. Biological nitrogen fixation in terrestrial ecosystems

In this section, we propose a new estimate of pre-industrial BNF in terrestrial ecosystems. The most comprehensive global estimate of BNF to date was carried out by Cleveland et al. [22], who proposed regional and global estimates based on scaling up empirical measurements of BNF. They applied average rates of BNF measured in empirical studies within each biome to the biome as a whole, and assumed a range of values for the cover of plants with potential N-fixing symbioses (legumes and actinorhizal species) in each biome. With this approach, they calculated an intermediate (and preferred) estimate for total terrestrial BNF of 195 Tg N yr\(^{-1}\); a later publication [23] that Cleveland co-authored reduced that estimate to 128 Tg N yr\(^{-1}\), largely because any higher estimate suggested too large a contribution of BNF relative to the quantity of N cycling within ecosystems annually. The multiple assumptions and extrapolations underlying these calculations make it difficult to test their validity, but the study did provide a summary of ecosystem-level measurements up to that time.

Unlike the previous estimate [22], our new calculation does not scale up from field-based measurements of BNF. Rather, it assumes a steady state for the pre-industrial terrestrial N cycle and uses estimates and calculations of all major N input and loss fluxes (aside from BNF) to calculate BNF by difference (figure 1). The steady-state assumption is unlikely to be true at local scales, but for this global calculation, it need only hold at large spatial and temporal scales.

The inputs to terrestrial ecosystems we consider initially are BNF, N fixed by lightning and deposited on land (LNF), and atmospheric deposition of reactive N transported from the ocean (\( \text{N}_2 \)). Losses we consider are hydrologic (HL) and gaseous losses (GL) from land to oceans or gaseous losses to unreactive forms in the atmosphere (\( \text{N}_2 \) and \( \text{N}_2\text{O} \); figure 1). For reactive N, we consider transfers only from the ocean to the land or vice versa as inputs and losses; most emissions of ammonia from terrestrial ecosystems (for example) are rapidly deposited downwind on other terrestrial systems.
In addition, a fraction of the N that leaves terrestrial systems via hydrologic pathways is denitrified along the flowpath from below soils to river mouths (flowpath denitrification = HLF; HLF = HLO + HLG) [28]. That fraction is poorly constrained, even for current conditions; as low as 25 per cent of current NANI appears at the mouths of rivers [27] (a fraction that includes accretion in terrestrial systems and river and reservoir sediments as well as denitrification). Bouwman et al. [29] calculated that as of 1900, nearly 60 per cent of N lost from terrestrial ecosystems by hydrologic pathways was denitrified along flowpaths from below soils to streams. Therefore, we account for downstream losses via denitrification by adding HLF = 20 Tg N yr⁻¹ (reflecting a nearly 60% loss along flowpaths) to hydrologic losses that reach the ocean, so our central value for HLF is 35 Tg N yr⁻¹. However, we suspect that pre-industrial systems probably lost a smaller fraction of N to denitrification along flowpaths, in large part because nitrate was a smaller fraction of total losses [30]; for this reason, our BNF estimate may err on the high side.

The gaseous loss flux from land (GL) is more difficult to determine directly, but calculations based on N isotopes make it possible to estimate the fraction of total losses that are gaseous (fGL) at steady state [28,31]. The ratio of stable N isotopes (¹⁵N/¹⁴N) in the terrestrial biosphere depends on the degree to which the different loss pathways (hydrologic versus gaseous) fractionate against the heavier isotope and the relative magnitudes of the different loss pathways. At steady state (eqn (3.1) in Houlton & Bai [31]),

\[
\frac{f_{GL}}{f_{LNF}} = \frac{\delta^{15}N_{TB} - \delta^{15}N_{I}}{\delta^{15}N_{I} - \delta^{15}N_{LNF}}
\]  

The data required to evaluate equation (3.2) are δ¹⁵N₁₁, δ¹⁵N₀, the mean isotopic ratios of the terrestrial biosphere and terrestrial N inputs, respectively, and δ¹⁵N₁ and δ¹⁵N₁₅, the isotope effects associated with hydrologic and gaseous losses, respectively. All of these terms are expressed in per mil notation. With this information and a reasonable estimate of the pre-industrial hydrologic N loss flux, we can calculate the gaseous N loss flux. Therefore,

\[
BNF = \frac{HL}{(1 - f_{LNF})} - LNF - N_{d1}
\]

Several studies have summarized global patterns in the distribution of soil and plant δ¹⁵N as a function of climate (mean annual temperature and precipitation) [32–34]; we use Amundson et al. [33] here because they included soils to 50 cm depth. Houlton & Bai [31] used the information in Amundson et al. [33] to estimate denitrification; following their approach, we used regressions from Amundson et al. [33] and the relative global stocks of plant (3.5 Pg N) and soil (95–140 Pg N) N [35] to estimate δ¹⁵N₁₁. Our analysis differed from Houlton & Bai [31] in two minor ways—we do not exclude human-managed areas because we are interested in a pre-industrial estimate, and we use a different climate dataset with a higher spatial resolution (10° versus 0.5°) [36]. Our approach assumes that changes in the N cycle since the industrial revolution have had a negligible impact on terrestrial N isotopes, an assumption that seems reasonable because most terrestrial N is soil N with a long residence time, and because most of the information in Amundson et al. [33] was derived from little-managed soils. Using the best fit model
for soil N to 50 cm depth from Amundson et al. [33], we calculate a land surface-area-weighted mean $\delta^{15}$N$_{TB}$ = 5.9‰, slightly higher than the 5.3‰ from Houlton & Bai [31].

The isotopic ratio of inputs ($\delta^{15}$N$_{in}$) depends on the isotopic values of all inputs weighted by their fluxes. The isotopic values of BNF (0% in theory, −2 to 0% in practice [37] and lightning-fixed N (0–1‰ [38]) are close enough that the relative fluxes do not have a large effect on the weighted isotopic value of total N input. We do not know the isotopic value for $N_{HL}$, the pre-industrial transfer of reactive N from ocean to land, but as discussed below, the magnitude of this flux was small. We use as our central value $\delta^{15}$N$_{HL}$ = −1‰; see the electronic supplementary material for an exact version of the calculation that accommodates different isotopic signatures of the inputs. Houlton & Bai [31] show that $\epsilon_{HL}$ is between −1 and 0‰ globally; we use 0‰ as our central value. Their synthesis of field estimates of $\epsilon_{HL}$ for denitrification yields a mean of −17% for temperate forests and −13% for tropical forests; we use −15‰ as a central value.

With these values and equation (2.2), our estimate for the fraction of total N lost as unreactive gas (N$_2$ and N$_2$O) is 0.46, which yields a total N gas loss of 3 Tg N yr$^{-1}$ (figure 1). Independent estimates of N$_2$O fluxes can provide perspective on this calculation. Pre-industrial N$_2$O emissions are thought to have been around 6.5 Tg N yr$^{-1}$ [39,40], and the pre-industrial atmospheric flux of reactive N from the land to the ocean is thought to be around 6.5 Tg N yr$^{-1}$ [23], yielding N$_2$ emissions of around 17 Tg N yr$^{-1}$ (not including denitrification that occurs along the flowpath from below soils to the mouths of rivers). This ratio of N$_2$ to N$_2$O (2.7) is similar to other estimates (2.2–4.6) [31]. An alternative estimate of the pre-industrial flux of reactive N from land to the ocean [41] suggested 14 ± 10 Tg N yr$^{-1}$; the central value of this estimate would reduce our estimate of N$_2$ emissions to 9 Tg N yr$^{-1}$, but it would not change our estimate of BNF.

The one remaining value required to calculate BNF is the atmospheric flux of reactive N from the ocean to the terrestrial surface ($N_{HL}$). We know of no estimates for this flux, but it likely was much less than the flux of reactive N from the land to the ocean. We therefore use half of the land to ocean flux, or 3 Tg N yr$^{-1}$, as our central value.

Using these values and equations (3.1)–(3.3), our central estimate for pre-industrial terrestrial BNF is 58 Tg N yr$^{-1}$. Nearly all of the values used here have substantial uncertainty associated with them, so we evaluated the sensitivity of the BNF estimate to each of the parameters by calculating derivatives of equation (3.1) with respect to each parameter. The resulting range of estimates for BNF is summarized in table 1. Of the three flux measurements that enter directly into the calculation, the hydrologic loss term is most important for three reasons. First, changes in HL must be matched by nearly two-fold changes in BNF (for $f_{L}$ near 0.5), whereas changes in the input terms must only be matched by one-to-one changes in BNF. Second, there is substantial uncertainty in our HL estimate. Third, non-BNF input fluxes are small relative to the hydrologic loss flux, and so proportionally similar uncertainty in HL affects BNF much more than does LNF or $N_{HL}$ (table 1).

Of the four isotopic parameters, $\epsilon_{C}$ is the most important for two reasons. First, it is more uncertain than that of the other parameters. Second, because it is the main driver of terrestrial isotopic N enrichment, $\epsilon_{C}$ has a highly nonlinear effect. As the absolute value of the isotope effect gets small (as it approaches the difference between the terrestrial biosphere $\delta^{15}$N and the input $\delta^{15}$N$_{o}$), there must be an increasing large gas loss flux to account for the soil isotopic enrichment, and thus a correspondingly large amount of BNF to balance this gas loss.

Table 1. Sensitivity of biological N fixation to the parameters used to generate the global estimate. All flux values (hydrologic losses that reach the ocean (HL), lightning-fixed N (LNF) deposited on land, the atmospheric flux of reactive N from oceans to land (N$_{HL}$) and biological nitrogen fixation (BNF)) have units of Tg N yr$^{-1}$, whereas all isotope values (the $^{15}$N/$^{14}$N ratios of the total biosphere ($\delta^{15}$N$_{TB}$) and inputs to land ($\delta^{15}$N$_{in}$), and the isotope effects associated with hydrologic ($\epsilon_{HL}$) and gaseous ($\epsilon_{C}$) losses) have units of per mill (‰).

<table>
<thead>
<tr>
<th>parameter</th>
<th>value (range)</th>
<th>effect of change on BNF (dBNF/dparameter)</th>
<th>BNF estimate$^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL$_{o}$</td>
<td>15 (7–23)$^{b}$</td>
<td>$1/(1 - f_{L})$</td>
<td>58 (43–73)</td>
</tr>
<tr>
<td>HL$_{f}$</td>
<td>20 (10–30)$^{b}$</td>
<td>$1/(1 - f_{L})$</td>
<td>58 (39–76)</td>
</tr>
<tr>
<td>LNF</td>
<td>4 (2–6)$^{b}$</td>
<td>1</td>
<td>58 (60–56)</td>
</tr>
<tr>
<td>$N_{HL}$</td>
<td>3 (1.5–4.5)$^{b}$</td>
<td>−1</td>
<td>58 (59–56)</td>
</tr>
<tr>
<td>$\delta^{15}$N$_{TB}$</td>
<td>5.9 (4.9–6.9)</td>
<td>$K(f_{HL} - f_{C})$</td>
<td>58 (51–67)</td>
</tr>
<tr>
<td>$\delta^{15}$N$_{in}$</td>
<td>−1 (−2 to 0)</td>
<td>$K(f_{C} - f_{HL})$</td>
<td>58 (49–58)</td>
</tr>
<tr>
<td>$\epsilon_{HL}$</td>
<td>0 (−2 to 0)</td>
<td>$K(-\epsilon_{C} - \delta^{15}$N$<em>{TB} + \delta^{15}$N$</em>{in}$)</td>
<td>58 (46–106)</td>
</tr>
<tr>
<td>$\epsilon_{C}$</td>
<td>−15 (−20 to −10)</td>
<td>$K(f_{HL} + \delta^{15}$N$<em>{TB} - \delta^{15}$N$</em>{in}$)</td>
<td>58 (46–106)</td>
</tr>
</tbody>
</table>

$^{a}$Calculated for other parameters at their central values.

$^{b}$For flux data, we used ± 50% for the uncertainty range.

$^{c}K = HL/(1 - \epsilon_{C} - \delta^{15}$N$_{TB} + \delta^{15}$N$_{in})$. 

$^{d}f_{HL}$ is the pre-industrial transfer of reactive N from ocean to land, but as discussed below, the magnitude of this flux was small. We use as our central value $\delta^{15}$N$_{HL}$ = −1‰; see the electronic supplementary material for an exact version of the calculation that accommodates different isotopic signatures of the inputs. Houlton & Bai [31] show that $\epsilon_{HL}$ is between −1 and 0‰ globally; we use 0‰ as our central value. Their synthesis of field estimates of $\epsilon_{HL}$ for denitrification yields a mean of −17% for temperate forests and −13% for tropical forests; we use −15‰ as a central value.

With these values and equation (2.2), our estimate for the fraction of total N lost as unreactive gas (N$_2$ and N$_2$O) is 0.46, which yields a total N gas loss of 3 Tg N yr$^{-1}$ and a total N loss of 65 Tg N yr$^{-1}$ (figure 1). Independent estimates of N$_2$O fluxes can provide perspective on this calculation. Pre-industrial N$_2$O emissions are thought to have been around 6.5 Tg N yr$^{-1}$ [39,40], and the pre-industrial atmospheric flux of reactive N from the land to the ocean is thought to be around 6.5 Tg N yr$^{-1}$ [23], yielding N$_2$ emissions of around 17 Tg N yr$^{-1}$ (not including denitrification that occurs along the flowpath from below soils to the mouths of rivers). This ratio of N$_2$ to N$_2$O (2.7) is similar to other estimates (2.2–4.6) [31]. An alternative estimate of the pre-industrial flux of reactive N from land to the ocean [41] suggested 14 ± 10 Tg N yr$^{-1}$; the central value of this estimate would reduce our estimate of N$_2$ emissions to 9 Tg N yr$^{-1}$, but it would not change our estimate of BNF.

The one remaining value required to calculate BNF is the atmospheric flux of reactive N from the ocean to the terrestrial surface ($N_{HL}$). We know of no estimates for this flux, but it likely was much less than the flux of reactive N from the land to the ocean. We therefore use half of the land to ocean flux, or 3 Tg N yr$^{-1}$, as our central value.

Using these values and equations (3.1)–(3.3), our central estimate for pre-industrial terrestrial BNF is 58 Tg N yr$^{-1}$.

Nearly all of the values used here have substantial uncertainty associated with them, so we evaluated the sensitivity of the BNF estimate to each of the parameters by calculating derivatives of equation (3.1) with respect to each parameter. The resulting range of estimates for BNF is summarized in table 1. Of the three flux measurements that enter directly into the calculation, the hydrologic loss term is most important for three reasons. First, changes in HL must be matched by nearly two-fold changes in BNF (for $f_{L}$ near 0.5), whereas changes in the input terms must only be matched by one-to-one changes in BNF. Second, there is substantial uncertainty in our HL estimate. Third, non-BNF input fluxes are small relative to the hydrologic loss flux, and so proportionally similar uncertainty in HL affects BNF much more than does LNF or $N_{HL}$ (table 1).

Of the four isotopic parameters, $\epsilon_{C}$ is the most important for two reasons. First, it is more uncertain than that of the other parameters. Second, because it is the main driver of terrestrial isotopic N enrichment, $\epsilon_{C}$ has a highly nonlinear effect. As the absolute value of the isotope effect gets small (as it approaches the difference between the terrestrial biosphere $\delta^{15}$N and the input $\delta^{15}$N$_{o}$), there must be an increasing large gas loss flux to account for the soil isotopic enrichment, and thus a correspondingly large amount of BNF to balance this gas loss.

Examining all of these sensitivities, we suggest that a reasonable range for pre-industrial terrestrial BNF is 40–100 Tg N yr$^{-1}$, with a preferred single estimate near 60 Tg N yr$^{-1}$. This preferred value, and indeed the entire range of values, is substantially less than previous estimates of 195 (range of 100–290) or 128 Tg N yr$^{-1}$ [22,23]. If our estimate is nearer the true value, previous calculations have underestimated the magnitude of the anthropogenic influence on terrestrial N inputs. Anthropogenic N fixation in 2005 was approximately 187 Tg N yr$^{-1}$ [1], so our new estimate suggests a 320 per cent (range of 190–470%) increase in total N fixed rather than the 100–150% increase suggested by previous estimates.
Even BNF in cropping systems, which Herridge et al. [13] calculated to be 40–55 Tg N yr$^{-1}$ (exclusive of extensively grazed tropical savannahs), looms large relative to pre-industrial BNF. Modern BNF in little-managed systems is likely to be substantially less than our pre-industrial estimate owing to land conversion, and perhaps to downregulation of BNF caused by increased deposition of anthropogenic reactive N.

As is often the case with global biogeochemical calculations, our assumptions do not consider several potentially important factors. Perhaps most importantly, we do not include inputs of geological N derived from the weathering of fixed N in rocks [42, 43]. Such inputs make a significant contribution to the N economy of specific sites, including to their hydrologic losses of N [44], but the global significance of geological N is not known. Geological N is potentially important not only because every Tg of input would imply an equivalent decrease in BNF, but also because the limited evidence that does exist shows geological N to be $^{15}$N-enriched relative to BNF [43]. Enriched inputs imply that matching observed ecosystem $\delta^{15}$N would require a lower fraction of losses via denitrification and hence lower BNF. Reasonably conservative estimates of geological N inputs of 10 Tg yr$^{-1}$, with $\delta^{15}$N = 2‰, imply a reduction in our central estimate of BNF from 58 to 44 Tg N yr$^{-1}$; electronic supplementary material details this calculation and explores a range of values.

Another assumption that should be evaluated is that atmospheric transfer of reactive N only affects our BNF calculation when it results in a transfer of N from the land surface to the ocean or vice versa. In fact, under some conditions, the isotopic calculation of gaseous losses could be sensitive to source-sink dynamics within terrestrial ecosystems. For example, areas without hydrologic losses (such as many deserts) and fire-prone areas such as savannahs might transfer isotopically light reactive N to non-fire-prone ecosystems downwind, and these transfers could decrease $\delta^{15}$N$_{gr}$ and thus inappropriately lower our estimate of BNF. For this reason and others, a regional analysis of pre-industrial N fluxes and enrichments would be of considerable interest.

### 4. From patterns to ecological controls

Here, we use information on regional patterns in BNF, together with observed variations through disturbance–succession cycles and the results of selected field experiments, to evaluate potential ecological controls of BNF. Using patterns to infer putative controls has obvious limitations, not the least of which is that N cycling dynamics have been disrupted strongly and recently by human activity, but it offers a useful starting point.

A number of studies have reported peaks in the abundance and activity of symbiotic N fixers at some early stage of succession. These BNF peaks frequently occur following geological disturbances (such as volcanic eruptions and glacial recessions) that initiate primary succession [19, 45, 46]; they also occur during secondary succession after severe stand-replacing fires [47], and indeed they can be used as a management tool to restore reactive N to degraded sites [48]. These peaks are usually associated with a small supply of N relative to other resources such as light, water and P. Even less destructive disturbances can cause losses of N owing to its high mobility relative to other nutrients, especially P. Consequently, the growth of successional vegetation is often limited by N supply, even in ecosystems where N was abundant in the system prior to disturbance [49] — and consequently symbiotic N fixers may be systematically favoured in many successional ecosystems. For example, Menge et al. [50] used forest inventory data to show that symbiotic N-fixing trees are much more abundant early in succession in all regions of the coterminus United States. Moreover, they showed that the broad phylogenetic clade that contains all symbiotic N fixers (legumes and actinorhizal plants, as well as their close relatives that appear to be incapable of N fixation) is not disproportionately abundant in early successional forests; it is specifically putative N fixers that occur early in succession [50].

Enhanced rates of BNF can occur later in succession as well. For example, where extremely cold temperatures cause an accumulating sink for C and N in boreal soils, cyanobacteria in moss carpets are often an important source of N via N fixation [51]. Similarly, in desert soils where the potential for gaseous losses is high, biological soil crusts may fix meaningful quantities of N in late-successional systems [52]. However, the highest rates of BNF generally are associated with symbiotic N fixers during early stages of succession.

Geographically, tropical savannah ecosystems long have been considered ‘hot spots’ of BNF — and many measurements in a wide variety of savannahs do report a substantial cover by potentially fixing legumes, the possibility of fixation from several other sources and in some cases high rates of measured N fixation [53]. Relatively high rates of BNF in savannahs make sense, because frequent fires effectively remove large quantities of N relative to P and other nutrients [54]. However, N fixation rates vary substantially among savannahs even within small geographical areas, resulting from differences in edaphic characteristics, history, grazing practices and other features. For example, in a Tanzanian savannah region Cech et al. [54] measured rates of BNF ranging from approximately 0.3 to 7.5 kg N ha$^{-1}$ yr$^{-1}$. Lopez-Hernandez et al. [55] measured substantial rates of BNF in a Venezuelan savannah, especially in burned areas; they found most N fixation was carried out by microbial crusts and microbes associated with the roots of C$_4$ grasses (pathways that Cech et al. [54] did not evaluate). It is fair to conclude that the variability in BNF is just as noteworthy as the occasionally high rates of fixation measured in savannahs.

(a) Biological nitrogen fixation often is regulated by N supply, and often responds to N deficiency

These and other studies of geographical and ecological patterns in BNF support the generalization that BNF is often (but not invariably) active where reactive N supply is small relative to other resources. Experimental manipulations also provide evidence supporting this generalization. For example, many studies demonstrate that rates of BNF are suppressed following N additions. On the other hand, adding another resource (water, P, CO$_2$, light) to an ecosystem where growth and biomass accumulation are limited by that resource can induce a demand for N, because in the absence of N fertilization or anthropogenically enhanced N deposition, N rarely is available in great excess above the supply of other resources. Thus, additions of these limiting resources can stimulate enhanced N fixation. For example, Benner & Vitousek [56, 57] observed that adding P to a P-limited montane forest in
Hawaii led to enhanced lichen growth in the canopy, and most of the lichens that responded were N-fixing cyanobionts that increased N inputs from approximately 0.3 to 9 kg ha \(^{-1}\) yr \(^{-1}\). Similarly, CO\(_2\) enrichment often stimulates plant growth, and in some experiments, it also increases the growth of legumes and rates of BNF [58,59]—although any such stimulation may be short-lived [60].

Tropical rainforests appear to function differently from most other forests; putatively, N-fixing legumes are abundant in many late-successional lowland tropical forests, and many of these forests are relatively rich in available N [61,62]. Hedin et al. [63] referred to this situation as the ‘nitrogen paradox’ of tropical forests. A number of studies have used \(^{15}\)N natural abundance to determine whether late-successional legumes actively fix N—an approach made feasible by the fact that non-fixing plants in many tropical forests are substantially enriched in \(^{15}\)N as a consequence of their open N cycle [64]. In most sites, \(^{15}\)N natural abundance studies suggest that only a small fraction of canopy legumes are actively fixing N—even though many of them belong to genera that are known to nodulate and fix N under appropriate conditions [64–66]. An independent approach showed that very little symbiotic BNF was needed to balance N losses in an Amazonian forest [67]. However, a thorough study by Pons et al. [68] in Guyana found nine species that both nodulated and met a substantially fraction of their N requirements from fixation. They concluded that approximately 6 per cent of the N taken up by these forests could be derived from contemporaneous BNF—an amount that is sufficient to replace N losses and to recover from N-depleting disturbances over reasonably rapid time-scales, but not a dominant source of N input [68]. These results are consistent with those of a field study in Panama that measured low rates of symbiotic BNF in a mature tropical forest with relatively high N availability, but high rates in secondary or disturbed forests with lower N availability [69]. Overall, while the capacity to fix N symbiotically is present in many late-successional tropical forests in which N is relatively abundant, it appears that relatively little of that BNF capacity is realized in practice.

(b) Biological nitrogen fixation is constrained in some N-limited ecosystems

Despite the overall tendency for BNF to be active wherever biologically available N is in short supply, some important types of ecosystems have unmet demands for reactive N and very low rates of BNF. This observation applies most clearly to the many late-successional temperate forests and grasslands in which plant production is proximately limited by N supply [6–8]—often despite enhanced deposition of anthropogenic reactive N. Many of these ecosystems are relatively productive and reasonably rich in energy, water and nutrients other than N. Rates of BNF could be low and N limitation could be sustained in these ecosystems if a resource or process systematically constrains N fixers more than it does non-fixers. Possible constraints that apply primarily to symbiotic systems [8,9] include:

— lower shade tolerance of N fixers—a lower ability of N fixers to persist in and ultimately to grow up through the shade of an established forest canopy, owing to the high energetic costs of BNF. This constraint is consistent with models of competition between N fixers and non-fixers [9,70,71] and with observations across North America showing that temperate trees with N-fixing symbionts typically are shade-intolerant [50]. This relationship does not extend to tropical forests, in which shade-tolerant legumes are widespread [62];
— preferential grazing on N fixers—higher rates of grazing on N fixers could occur because N fixers generally have higher tissue N content than do non-fixers—and most grazing animals seek (and often require) N- and protein-rich food sources. There is good evidence for disproportionately intense grazing on N fixers, and evidence that such grazing can reduce N inputs and sustain N limitation to primary production on the ecosystem level [72]; and
— higher non-N nutrient demands by N fixers—certain nutrients are required in greater amounts by fixers than non-fixers (as clearly is true for Mo and Fe [11]; it may also be true for P, at least for some groups of N fixers). In this case, N fixers may be constrained by another nutrient, whereas non-fixers are constrained by N supply. Thus, N limitation to primary production on a whole-system level may be limitation by P (or another non-N nutrient) in disguise.

The first two mechanisms involve ecological interactions in that other organisms constrain N fixers; non-fixing shading or preferential grazing on plants with N-fixing symbioses would keep them from growing into or persisting within the plant canopy, where they could actively carry out BNF. The last mechanism also could be influenced by biotic interactions; for example, Cech et al. [73] demonstrated that a dominant C\(_4\) grass could outcompete an herbaceous legume for otherwise-available P (and N) in a tall-grass Tanzanian savannah ecosystem, thereby suppressing legume growth and N fixation.

With the exception of decomposers [14,74–76], there has been less research on patterns of and constraints to BNF in groups of organisms other than angiosperms with N-fixing symbioses. Where (and when) they are present in ecosystems, symbiotic N fixers are important because they can achieve extremely high rates of fixation. However, the cumulative effects of other groups of N fixers can provide the dominant input of fixed N where symbiotic BNF associated with higher plants is low or non-existent [14]. These contributions are included in our global estimate of N fixation; further analysis of their ecological regulation would be rewarding.

5. Models of ecological controls of biological nitrogen fixation

Given both the methodological challenges inherent to estimating BNF in complex ecosystems and the notorious spatial and temporal heterogeneity of the process, we are not likely to develop well-constrained estimates of regional/global BNF or its controls simply by accumulating and extrapolating the results of local studies. However, models that synthesize the dynamics of N-fixing organisms and symbioses, their interactions with other organisms and environmental factors, and the ecosystem-level feedbacks that influence resource availability offer a potentially productive path towards an ecological understanding of BNF. While a few high-quality empirical studies may not suffice to estimate biome-level BNF by extrapolation, they can provide the fundamental information needed to develop and test models.
Several models that focus on ecosystem-level controls on BNF have been developed. Vitousek & Field [9,77] evaluated constraints to BNF in N-limited ecosystems, simulating a N-fixing symbiosis that pays higher costs for N acquisition than does a non-fixed. Rastetter et al. [70] developed a similar model for competition between a N fixer and one or more non-fixers within the framework of a multiple element limitation model. Vitousek & Field [9] found that without additional constraints, the relatively high cost of BNF alone was insufficient to constrain N fixation enough to sustain N limitation. Adding other constraints to the model (an increased requirement for P, decreased shade tolerance by the N fixer and increased susceptibility to grazing for the N fixer) could constrain N fixation sufficiently to sustain N limitation indefinitely—as long as there were N losses from the system via pathways that could not be prevented by biotic uptake of available N [9]. Later variations on this model showed that N losses caused by environmental fluctuations (especially in water availability in seasonally dry environments) and by periodic disturbance (e.g. fire) also could constrain BNF sufficiently to yield long-term N limitation ([77,78], see also [79]). While these models demonstrate that modelled constraints to N fixation could inhibit BNF enough to sustain N limitation, they do not show which mechanisms constrain BNF in N-limited ecosystems.

More recently, Menge et al. [71,80] developed analytical models of a plant with N-fixing symbioses competing with a non-fixed. They concluded that a lower N use efficiency (less growth per unit of N acquired) can suffice to keep N fixers from entering a N-limited ecosystem dominated by non-fixers [71]—a result consistent with the greater N concentrations observed in legumes [81] and actinorhizal plants in comparison with most non-fixers. Menge et al. [80] also explored the importance of facultative versus obligate symbiotic N fixation; both Vitousek & Field [9] and Rastetter et al. [70] modelled a situation in which the N fixer always acquires N via fixation. It appears that a superior strategy would be to fix N when N is in short supply, and to function as a non-fixed when N is abundant or (especially for an understory plant) when N fixation is too costly. However, Menge et al. [80] found that the costs of the regulatory system that turns N fixation on and off (together with all of its structural and biochemical machinery), and the lag time before regulation becomes effective could make obligate N fixation, facultative N fixation or non-fixation the most effective strategy (depending on costs and conditions). As discussed earlier, there is some evidence for a diversity of regulation strategies in nature (see [82]). In theory, this diversity of strategies could simultaneously explain the preponderance of putative N fixers in late-successional tropical forests and the paucity of N fixers in late-successional temperate forests [80].

Houlton et al. [83] proposed a biome-level model of controls of BNF—suggesting that N-fixing legumes could be favoured in tropical forests because they acquire enough N to maintain higher extracellular phosphatase activity (relative to non-fixers) that help overcome potential P limitation in these often P-limited ecosystems, and because the temperature in tropical environments is near the optimum for BNF (which their review suggested to be approx. 26°C). The temperature dependence of the nitrogenase enzyme system is unusual in that it has strongly biphasic kinetics, with an activation energy of 0.65 eV above 22°C (similar to respiration), and a very high activation energy of 2.18 eV below 22°C [84]. Accordingly, the potential rate of BNF decreases relatively slowly as temperature decreases to 22°C, and then falls much more rapidly at lower temperatures. These kinetic differences may contribute to the global pattern of BNF summarized by Cleveland et al. [22], modelled by Houlton et al. [83] and implied by the 15N natural abundance patterns reported by Amundson et al. [33].

6. Conclusions

Research over the past decade has greatly expanded our knowledge of the diversity of organisms capable of carrying out BNF, but it has contributed less to our understanding of the rates and regulation of BNF. Here, we draw upon information from global summaries of 15N/14N in terrestrial ecosystems and estimates of pre-industrial N fluxes to suggest a new estimate of terrestrial BNF prior to extensive human alteration of the Earth system. Our estimate is 58 Tg N yr⁻¹ (44 Tg N yr⁻¹ accounting for geological N), with a plausible range from 40 to 100 Tg N—substantially lower than previous estimates. If accurate, then this new estimate suggests that the magnitude of anthropogenic alteration of the terrestrial N cycle has been understated. Further, we suggest that understanding BNF on ecosystem and regional scales is a key to understanding, predicting and managing the consequences of multiple components of anthropogenic global change—and that a new generation of models and associated experiments offer the best opportunity to achieve that understanding.

Research and manuscript preparation were supported by grant no. DEB-1020791 from the National Science Foundation; D.N.L.M. was supported by the Carbon Mitigation Initiative (cmi.princeton.edu). Conversations with T. Crews, C. B. Field, L. O. Hedlin, and B. Z. Houlton, contributed to the analyses here, and B.Z. Houlton commented on an earlier draft of the manuscript.

References

7. LeBauer DS, Treseder KK. 2008 Nitrogen limitation of net primary productivity in terrestrial ecosystems.


84. Ceuterick F, Peeters J, Heremans K, de Smet H, Olbrechts et al. 2006 The presence of nitrogen fixing legumes in terrestrial communities: evolutionary vs.