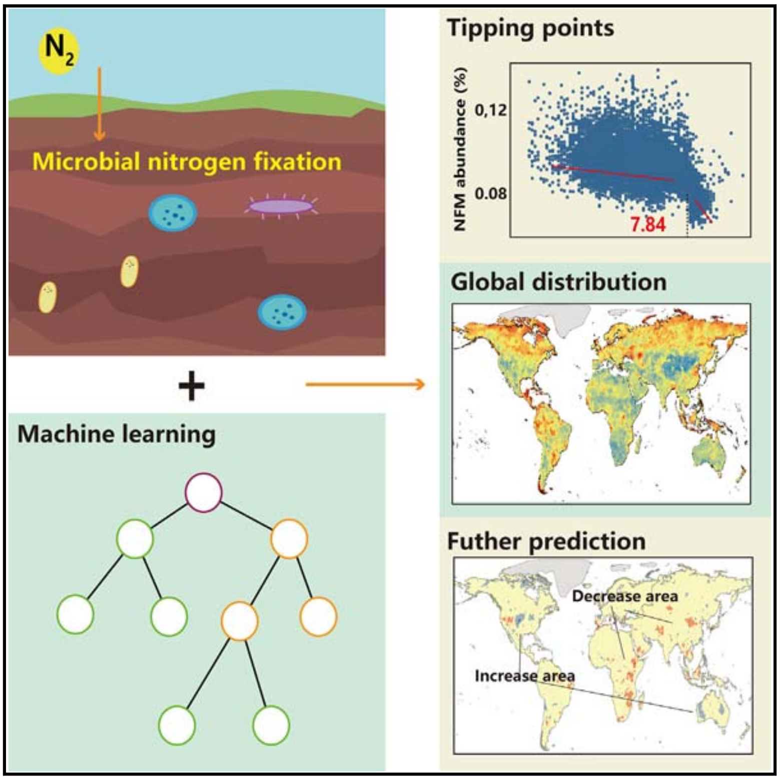
iScience

Article

Environmental tipping points for global soil nitrogen-fixing microorganisms

Graphical abstract



Highlights

- pH, precipitation, and organic carbon are domain factors for N-fixing microbes worldwide
- The pH tipping point for N-fixing microbes is 7.84
- The organic carbon tipping point for N-fixing microbes is 2.7%
- The contribution of precipitation to N-fixing microbes peaked at 1,265.65 mm

Authors

Yueqi Hao, Hao Liu, Jiawei Li, Li Mu

Correspondence

muli@caas.cn

In brief

Biogeochemistry; Biogeoscience; Global change; Global Nutrient Cycle; Microbiology

Hao et al., 2025, iScience *28*, 111634
January 17, 2025 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.111634

CellPress

iScience

CellPress OPEN ACCESS

Article

Environmental tipping points for global soil nitrogen-fixing microorganisms

Yueqi Hao,[1,2] Hao Liu,[2] Jiawei Li,[2] and Li Mu[1,3,*]

[1]Key Laboratory for Environmental Factors Control of Agro-product Quality Safety (Ministry of Agriculture and Rural Affairs), Tianjin Key Laboratory of Agro-environment and Safe-product, Institute of Agro-environmental Protection, Ministry of Agriculture and Rural Affairs, Tianjin 300191, China

[2]Key Laboratory of Pollution Processes and Environmental Criteria (Ministry of Education)/Tianjin Key Laboratory of Environmental Remediation and Pollution Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300080, China

[3]Lead contact

*Correspondence: muli@caas.cn

https://doi.org/10.1016/j.isci.2024.111634

SUMMARY

Nitrogen-fixing microorganisms (NFMs) are important components of soil N sinks and are influenced by multiple environmental factors. We established a random forest model optimized by the distributed delayed particle swarm optimization (RODDPSO) algorithm to analyze the global NFM data. Soil pH, organic carbon (OC), mean annual precipitation (MAP), altitude, and total phosphorus (TP) are factors with contributions greater than 10% to NFMs. pH, OC, and MAP are the top three factors at the global scale. The tipping points of pH and OC for the NFMs were 7.84 and 2.71%, respectively. The contribution of MAP first increased but then decreased with peak value at 1,265.65 mm. Under the scenario SSP 8.5, 12% of the NFMs increase occur in Africa in 2100; 16% and 36% of the NFMs decrease in North America and Oceania in 2100, respectively. Our work created a global NFMs map and identified the critical tipping points.

INTRODUCTION

Nitrogen-fixing microorganisms (NFMs) in soils are critical contributors to ecosystem functions since they bring atmospheric N_2 into the soil N cycle.[1] Microbe-associated N fixation provides approximately 0.9–1.3 × 10^{14} g N$\cdot$y^{-1} in terrestrial areas of the world.[2] NFMs are directly associated with the soil nitrogen fixation process and further affect soil N storage and the fertility of terrestrial ecosystems.[3] Nitrogenase plays a crucial role in biological nitrogen fixation. Nitrogenase exists in molybdenum (Mo)-dependent, vanadium (V)-dependent, and iron (Fe)-dependent or heterometal-independent forms. Mo-dependent nitrogenase, encoded by the *nif* gene, is the most well-characterized and most commonly occurring form of nitrogenase.[4]

Many factors are reported to affect soil N fixation efficiency and its associated microbial community (e.g., climate, nutrients, land use, agronomic activities, and natural environments).[5–7] Most previous studies on NFMs have been laboratory or experimental field studies. The results of these studies may exhibit constrained applicability in real environments on a global scale.[3,8,9] Moreover, environmental changes have profound implications for natural ecosystems and may lead to their modification, degradation, or collapse.[10] Many species, such as plants and plankton, have environmental tipping points.[11,12] The existence of NFM tipping points means that small changes in environmental factors can cause substantial and irreversible alterations of NFMs, and their impact on the soil nitrogen fixation capacity deserves attention. However, whether the impact of environmental factors is gradual or abrupt and whether the responses reveal multiple tipping points remain largely unknown in the context of NFM abundance.

The machine learning approach is a data-driven model with a high capability for learning complicated patterns[13,14] that can be used to determine important environmental factors and make sound predictions of the relationships between the responses to these factors.[11,15] Then, the conditions and locations of the tipping points may be identified. The identification of tipping points affecting NFM abundance is urgently needed to maintain soil function and mitigate climate change.

In this study, 1,659 microbial community samples were collected from 595 locations worldwide (Figure S1). The relative abundance of nitrogen-fixing microorganisms in the soil was taken as the research object. Sixteen environmental factors that may threaten soil NFMs according to previous articles were selected as independent variables.[16,17] The factors included those related to climate (e.g., mean annual temperature [MAT], mean annual precipitation [MAP], and aridity index [AI]), soil properties (e.g., pH, soil total carbon [TC], soil organic carbon [OC], soil organic matter [OM], soil total nitrogen [TN], soil total phosphorus [TP], available soil water [ASW], and texture), agricultural management (e.g., use of N/P/K fertilizers), altitude, and sampling depth. A random forest model optimized by the

© 2024 The Author(s). Published by Elsevier Inc.

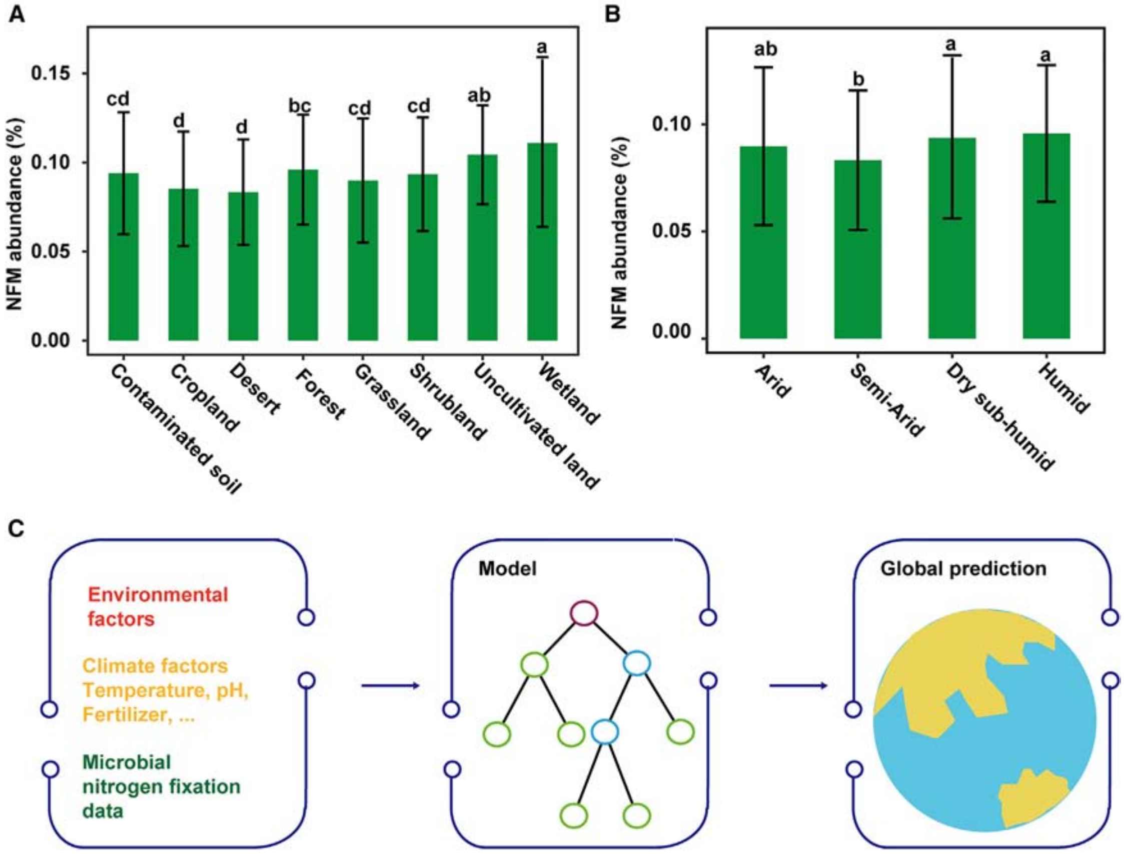


Figure 1. Basic distribution and workflow diagram of soil NFMs

(A) Abundance of NFMs in different habitats (contaminated soil $n = 103$, cropland $n = 619$; desert $n = 60$, forest $n = 403$, grassland $n = 303$, shrubland $n = 75$, uncultivated land $n = 46$, wetland $n = 50$), different letters indicate significant differences between columns ($p < 0.05$); (B) abundance of NFMs in different aridity classes (arid $n = 74$, semiarid $n = 582$, subhumid $n = 152$, and humid $n = 851$), aridity index (AI) < 0.2, semiarid 0.2<AI<0.5, dry subhumid 0.5<AI<0.65, and humid 0.65<AI, different letters indicate significant differences between columns ($p < 0.05$); (C) workflow for predicting the distribution pattern of soil microbial NFM abundance on a global scale. The error bars represent the standard errors of the columns.

distributed delayed particle swarm optimization (RODDPSO) algorithm was established to analyze the above information. A workflow for predicting the global distribution pattern of soil NFM abundance is presented in Figure 1B. This work predicted the current and future distributions and development of NFM abundance in soil and identified the crucial factors and ecological tipping points affecting NFM abundance. Our work provides quantitative information for understanding soil nitrogen sinks and mitigating climate change driven by microbes.

RESULTS AND DISCUSSION

Heterogeneous nitrogen-fixing microorganisms in the global soil

NFMs are important components of soil N sinks.[18] The global distribution pattern of soil NFMs provides a foundation for the predictive understanding of future soil nutrient storage and soil fertility.[1,19] A total of 1,659 datasets with soil microbial information from 595 locations worldwide were mined and analyzed (Figure S1). After thoroughly considering the main influencing factors of soil NFMs,[19–21] data for 16 environmental factors were collected, including data related to altitude, AI, MAT, MAP, pH, TC, OM, OC, TN, TP, ASW, texture, sampling depth, and artificial management status (i.e., use of N/P/K fertilizers).

According to data from published studies, the relative abundance range of global soil NFMs is 0.09%–0.28%. Considering that the rates of N_2 fixation are positively correlated with the population of N_2-fixing microorganisms ($R^2 = 0.85$, $p < 0.005$),[22] there are obvious differences in N fixation efficiency in soils around the world. The relative abundances of NFM in different habitats are presented in Figure 1A. The abundance of NFM in wetlands is slightly greater than that in other areas, whereas croplands and deserts are habitats with lower NFM abundances. The NFM abundance in wetlands and deserts may be influenced by soil moisture and nutrient conditions in these habitats.[23,24] The relatively lower NFM abundance in croplands may be caused by their inhibition by long-term fertilization, since fixing nitrogen is known to be an energy-intensive process. Additional nitrogen can suppress the development of N_2-fixing microorganisms.[25] The relative abundances of NFMs in different aridity classes are similar, whereas semiarid areas have slightly lower NFM contents (Figure 1B). In general, aridity has a negative influence on

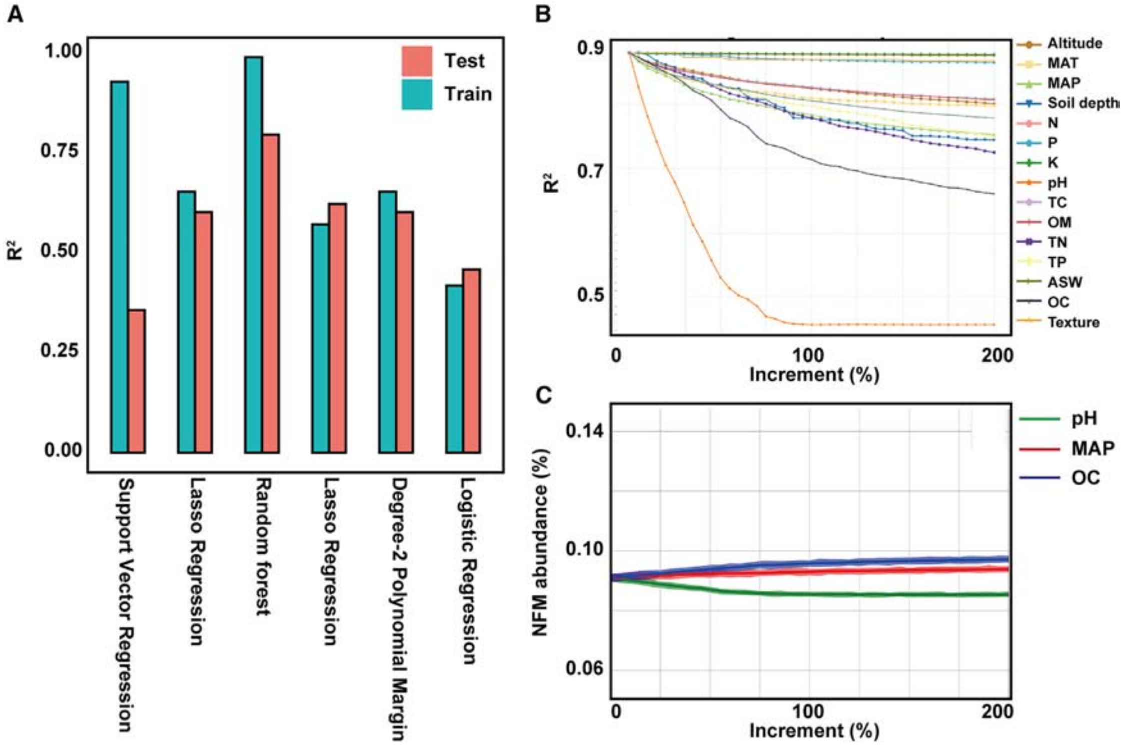


Figure 2. Performance of the machine learning models

(A) R^2 values of the training and testing sets of the six machine learning algorithms; (B) robustness evaluation of the random forest model; (C) uncertainty assessment of the random forest model.

soil diazotrophs.[26] The competitive advantage of free-living N_2-fixing microorganisms in arid ecosystems may balance the adverse conditions in arid regions.[27] The combined effects of these two patterns allow semiarid areas to become cold zones for NFMs.

The abundance of NFMs was not strongly related to all 16 environmental factors; only altitude and pH had correlations of −0.12 and −0.14 with NFM abundance (Figure S2), indicating that the linear relationships between the environmental factors and the NFMs were weak. Machine learning has a powerful ability to handle nonlinear problems. In the following sections, we establish a machine learning approach to further predict NFM abundance and perform feature analysis (Figure 1B).

Machine learning model building and critical factor screening

The coefficient of determination (R^2) values of the training and testing sets of the different models are presented in Figure 2A. The random forest model performed the best among the six tested machine learning models, with a training set R^2 of 0.97 and a testing set R^2 of 0.78. The R^2 values of the logistic regression, degree-2 polynomial margin, ridge regression, and LASSO regression range from 0.4 to 0.7. The random forest and support vector regression performed relatively well, although the R^2 of the training set and the testing set showed great differences (Figure 2A). This is due to overfitting caused by the model falling into a local optimal solution. We propose a randomly occurring distributed delayed particle swarm optimization (RODDPSO) algorithm to overcome this problem.[28] The random forest model was chosen as the base model. The particle swarm optimization (PSO) algorithm is an evolutionary computing technique that seeks the optimal solution through collaboration and information sharing among individuals in a population. To avoid the problem of the PSO algorithm falling into local optima, this study introduced a certain distributed time delay on the basis of the PSO algorithm to increase the ability of the particles to escape from local optima and overcome the problem of premature convergence. In addition, RODDPSO was also used to achieve automatic hyperparameter tuning of the model, which can effectively improve the processing efficiency of the model.[28] The R^2 values of the training and testing sets of the improved random forest model were 0.98 and 0.80, respectively. The mean squared error (MSE) was 0.0004. The mean absolute error (MAE) was 0.0144. Figure 2B presents the robustness of the model. The R^2 values of all of the variables remain above 0.6 as the data size increases, except for texture. Figure 2C presents the uncertainty of the model. The change proportions of the three variables are less than 10% as the data size increases. The results of the robustness evaluation and uncertainty assessment indicate that our model is stable.

Critical factor screening and tipping point identification

To assess the important factors associated with soil nitrogen-fixing microorganisms, the SHAP values for 16 factors were

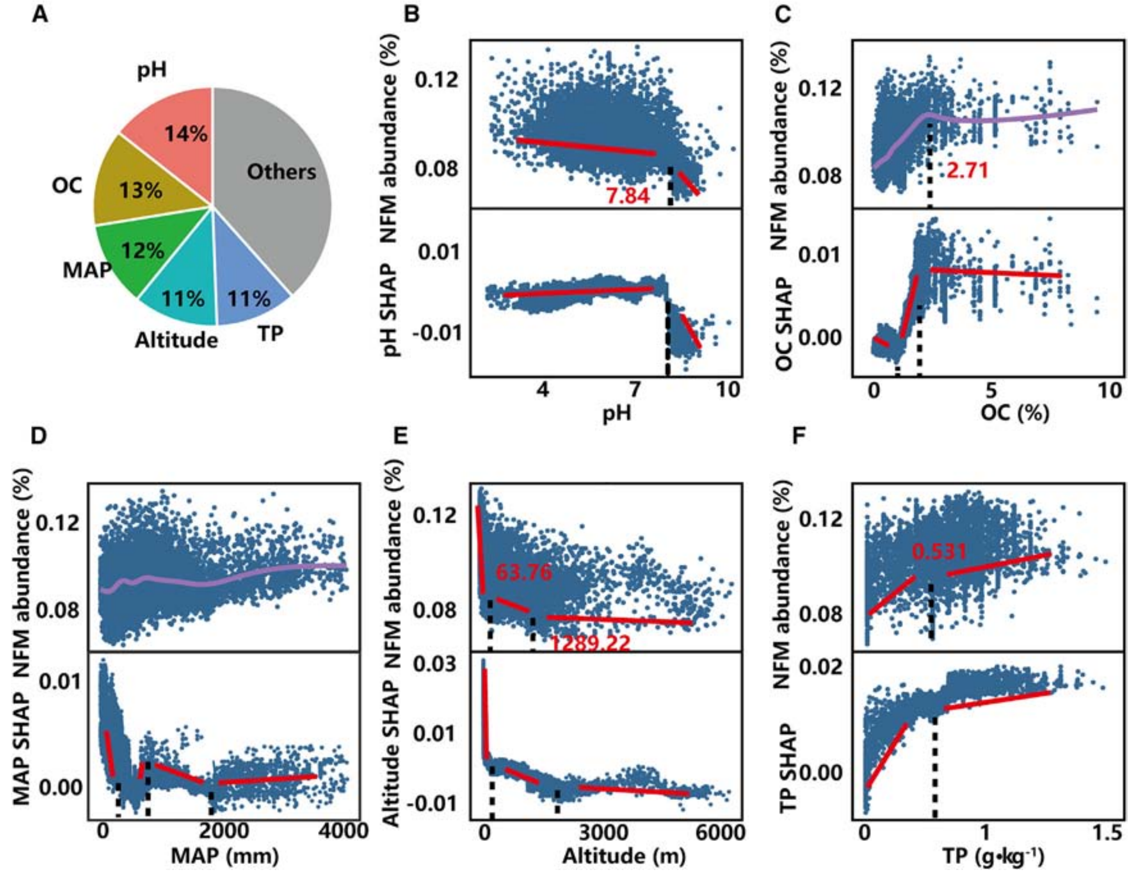


Figure 3. Quantitative contributions of critical environmental factors and their response to soil NFMs

(A) Quantitative contribution of critical environmental factors to soil NFM abundance in 2022; (B) response of global NFMs and SHAPs to pH; (C) response of global NFMs and SHAPs to OC; (D) response of global NFMs and SHAPs to MAP; (E) response of global NFMs and SHAPs to altitude; (F) response of global NFMs and SHAPs to TP. The black dashed lines and numbers in red font represent the identified tipping points; the purple line represents the smoothed trend fitted by the generalized additive model (GAM), and the red lines are the fitted lines obtained from the segmented linear regression (SLR) model.

calculated. The SHAP values estimate the contribution of each factor to the model output.[29] In our study, screening for critical factors laid the foundation for tipping point analysis. An ecological tipping point is a state or condition where a minor change in a critical factor triggers another variable or phenomenon to undergo an abrupt change or relationship reversal[30] (Figure S3). This information can become the premise of subsequent management. The mean SHAP values revealed that pH, OC, MAP, altitude, and soil TP were the most important factors affecting the soil NFMs (Figure 3A). These five factors are the factors with contributions greater than 10% to the NFMs on a global scale. The complete mean SHAP values of the different environmental factors are shown in Figure S5A.

pH is a critical factor affecting NFMs; 7.84 was the global pH tipping point we identified on the basis of the predicted global NFM abundance and SHAP value. The differences in NFM abundance on both sides of the pH tipping point were significant ($p < 0.01$; Figure S5B). Soil pH plays an essential role in the diazotrophic community. Some common diazotrophs, such as *Bradyrhizobium*, are more abundant at acidic pH values (4.5–5.5), whereas *Azospirillum* and *Rhizobium* increase in abundance with increasing pH.[31] As *Bradyrhizobium* species have wide niches and are excellent survivors across diverse conditions, high abundance in acidic soil is expected.[32] Neutral pH (6.5–7.5) is the optimal pH range for most soil microbes. Alkaline soil has become a cold zone for nitrogen-fixing microorganisms. The NFM abundance in areas with pH values higher than 7.84 was significantly different from that in other areas ($p < 0.01$; Figure S5B).

Soil nutrients are considered important factors affecting biotic nitrogen fixation. Our model selected soil OC and TP as critical factors. The relative abundance of NFM increased with OC and TP at the global scale (Figure 3CF). OC and TP in soil act as energy and ATP sources for diazotrophs since biological nitrogen fixation is a highly energy-consuming process.[33] The abundance of NFM barely changed after the tipping points (2.7% and 0.351 g kg^{-1}), indicating that the N fixers and the nonfixers in the soil had reached a balance.[34]

The response pattern also revealed that the abundance of NFMs did not change obviously with MAP (Figure 3D). However, we identified three tipping points (153.54; 1,265.64; and 2,968.57 mm) that delineated the MAP SHAP values into three

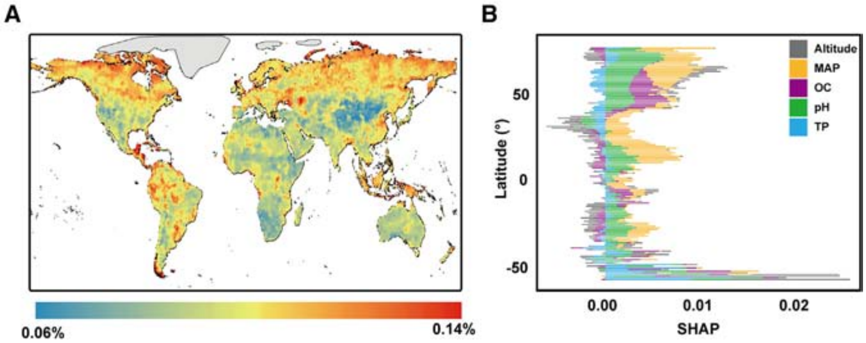


Figure 4. The global distribution of NFM abundance

(A) Prediction of the global distribution of NFM abundance in 2022; (B) SHAP value distribution for the NFM with latitude.

precipitation intervals (Figure 3D). When the precipitation is less than 153.53 mm, the contribution of MAP (SHAP value) to nitrogen-fixing microorganisms is high. Free-living nitrogen-fixing microorganisms have a competitive advantage in arid regions, and free-living fixation is a substantial contributor to biotic nitrogen fixation in areas with low precipitation.[35,36] The contribution of MAP to NFM abundance first increased but then decreased between 153.53 and 2,968.57 mm, and the peak value reached approximately 1,265.65 mm (Figure 3D). Partial dependence analysis also revealed that the response of the NFM to MAP presented a hump shape, with the highest value occurring at 1,200 mm (Figure S6C). This pattern is consistent with the response of soil microorganisms to precipitation.[37] The threshold of MAP SHAP ranged from −0.003 to 0.015, which is lower than that of the other factors (Figure S7). This may explain why the response pattern does not apply to NFM abundance at the global scale (Figure 3D). However, MAP has a broad influence globally and is the most important environmental factor affecting 16% of territorial land (Figure S8). Therefore, we still consider MAP as a critical environmental factor for NFMs, and its tipping point is at 1,265.65 mm.

According to our research, areas at low altitudes have relatively higher NFM abundance. The first altitude tipping point is 63.76 m, and the abundance of NFM remains unchanged after reaching the second tipping point at 1,289.22 mm (Figure 3E). The hot zone for altitude SHAP appeared in the coastal zone of continents (Figure S9). Coastal ecosystems are important carbon sinks for the Earth and have sufficient soil moisture.[38] The abundance of NFMs along the coastline is the result of the combined action of water and nutrients. NFMs also supplement oligotrophic waters with nitrogen through nitrogen fixation, harmonizing the distribution of nitrogen.[39]

Prediction of global NFMs under present and future climate scenarios

On the basis of the abovementioned established model, the relative abundance of global microbial NFMs is presented in Figure 4A. The relative abundance range of global soil NFM was 0.06%–0.14% (Figure 4A). Latitude 50–80°N is a hot zone for soil NFMs, with an average abundance of 0.104 ± 0.009% (Figure 4A). The high abundance of NFM is determined by the combined effects of pH, MAP, and OC, which all have positive SHAP values in this region (Figure 4B). High-latitude soil is characterized by acidic soil, low precipitation, and high OC, which make it the most favorable region for nitrogen-fixing microorganisms. Some studies indicate that free-living NFMs and NFMs in soil crusts are important sources of nitrogen in cold areas.[35,40] Latitude 30–50°N is a cold zone for the soil NFMs, with an average abundance of 0.090 ± 0.010% (Figure 4A). This area has 50% of the soil pH higher than 7 and an average precipitation of 500 mm. High altitude also has negative effects on nitrogen-fixing microorganisms. The abundance of NFMs at 10–30°N and −30 to 10°N were 0.092 ± 0.008% and 0.091 ± 0.008%, respectively (Figure 4A). The NFMs in these two regions are driven mainly by precipitation and pH, and hotspots occur mainly in Asia and America. In the regions of 10–30°N and −30 to 10°N, the abundance of NFMs in Asia and the Americas are 0.099 ± 0.010% and 0.097 ± 0.009%, respectively. There are no NFM hotspots in the same latitudinal zone of Africa or Oceania. The MAP in Asia and the Americas within this latitude is between 1,000 and 1,500 mm, with a relatively high contribution value (SHAP value) to the NFMs, whereas the MAP in Africa and Oceania is less than 1,000 mm. Acidic soils also increased the abundance of NFMs in Asia and America. The abundance of NFMs at −10 to 10°N is 0.096 ± 0.010%, which is similar to the global average NFM abundance of 0.096 ± 0.011%. Some studies suggest that high nitrogen availability in tropical areas inhibits microbial nitrogen fixation.[41] Highly weathered soil is also prone to lacking Mo.[42] However, our research revealed that tropical regions are not cold spots for nitrogen-fixing microorganisms. This may be due to the suitable MAP and OC conditions.

pH and MAP are the environmental factors that affect the widest latitudinal range (Figure 4B). pH, OC, and MAP were the domain factors (highest SHAP values) for 58% of the terrestrial soils (Figure S8). Double-variable partial dependence diagrams show the combined effect of the two critical factor pairs (Figures 5A–5C). The abundance of NFMs would clearly change when both factors were near the tipping points. Therefore, when any environmental factor reaches a tipping point, notable changes in the NFMs occur, and this effect intensifies when both factors coexist. The abundance of N-fixing microorganisms is highest when both factors are within the optimal range. The optimal ranges for pH, OC, and MAP are 6∼7, 2∼2.5%, and 1,000–1,200 mm, respectively. Figures 5D–5F show that the effects of individual factors on NFMs can be additive. The interaction effects between pH and MAP and pH and OC were not significant (*p* value for interaction >0.01), and only the interaction effect between MAP and OC was significant (*p* value for interaction <0.01). The range of the high interaction effect mainly occurs in areas with 10%–22% OC and 1,000–2,800 mm MAP. The proportion of this region on global land is less than 1%. Therefore,

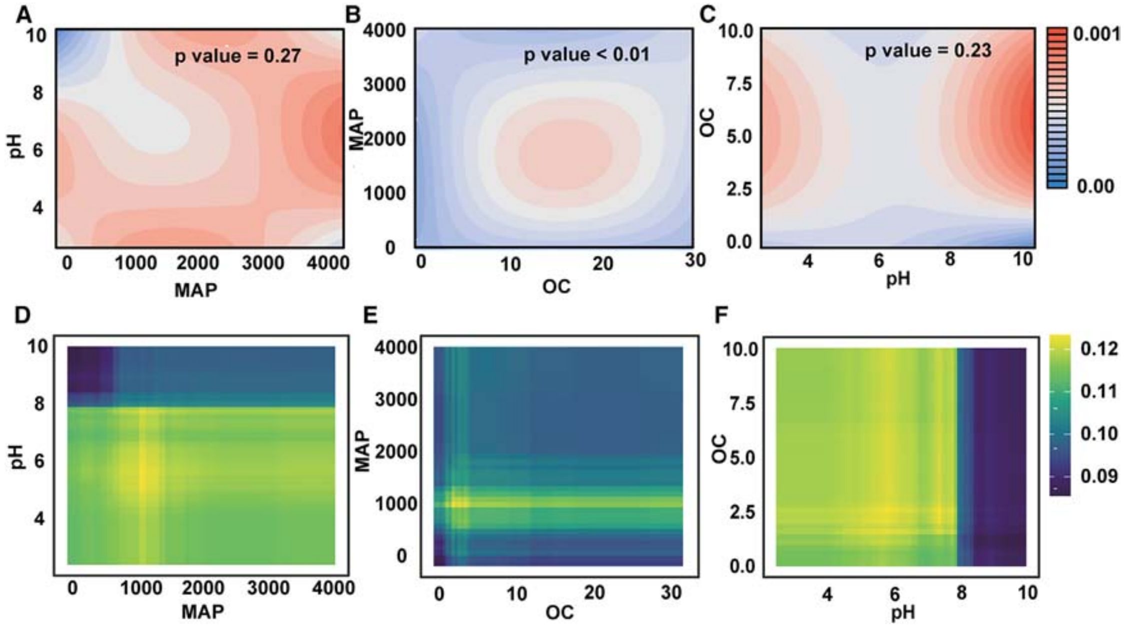


Figure 5. Double-variable partial dependence and interaction effects of domain factors

(A) Double-variable partial dependence diagram of pH and MAP; (B) double-variable partial dependence diagram of OC and MAP; (C) double-variable partial dependence diagram of pH and OC; (D) interaction effect for pH and MAP; (E) interaction effect for OC and MAP; (F) interaction effect for pH and OC.

the main interactions among pH, OC, and MAP are independent of the global scale.

Figure 6 shows the proportional changes in the NFMs forecasted from year 2021–2100 under the different climate scenarios. The change scales of the climate data were based on the sustainable economy (SSP 2.6) and fossil-based economy (SSP 8.5) scenarios in the CMIP6.[43,44] Under SSP 2.6 and SSP 8.5, 6.54% and 7.03%, respectively, of terrestrial soil has obvious increases in NFMs (more than 5%) in year 2100. The regions with obvious NFM decreases (more than 5%) accounted for 5.78% and 8.98% of the terrestrial soil under SSP 2.6 and SSP 8.5 in year 2100, respectively. The areas with increases in Africa and Oceania are greater than 10% under SSP 2.6 in year 2100. The areas with decreases in North America and Oceania are 9% and 27%, respectively, under SSP 2.6 in year 2100. The increase in area proportion under SSP 8.5 is similar to that under SSP 2.6, whereas the decrease in area proportion is larger than that under SSP 2.6 on every continent. Under the SSP 8.5 scenario, 16% and 36% of the NFMs in North America and Oceania, respectively, show a decreasing pattern at year 2100.

Because there is no suitable soil property database that is based on future climate models, current soil properties were used for future predictions in our analysis. Laboratory experiments have shown that elevated CO_2 can cause a decrease in soil pH.[45] Global-scale soil acidification is induced by nitrogen deposition and fertilizer.[46,47] Soil acidification is beneficial for the abundance of nitrogen-fixing microorganisms; however, considering the impact of acidic soil on soil nutrient loss and plant production,[48,49] maintaining the soil pH under neutral conditions is the best strategy. Soil OC is sensitive to climate change, and soil respiration caused by warming leads to OC loosening.[50] The global SOC stock will decrease by 6.0% under 1°C air warming despite stable carbon decomposition, and increased plant litter can be supplied as a carbon source.[51] The loss of OC directly affects the abundance of nitrogen-fixing microorganisms. Global warming can indirectly have negative impacts on nitrogen-fixing microorganisms through OC.

Conclusions

Elucidating the potential and trend of NFM abundance in soil is important for understanding the soil N cycle and the ecological functions of nitrogen. Through machine learning, pH, OC, and MAP were identified as the key factors with high contributions that affected the scope of soil NFMs at the global scale. Precipitation near tipping points may maximize the nitrogen fixation ability of soil NFM. The migration of future climate zones led to an increase in NFM abundance in Africa and a decrease in Oceania. Adjusting soil properties is a more feasible way to increase the soil microbial nitrogen fixation capacity. Owing to the tipping point of pH, there is an obvious difference in the abundance of NFM between alkaline and neutral/acidic soils. When soil OC exceeds 2.7%, OC is no longer a limiting factor for NFMs. Our study provides quantitative information at the global scale to assist in soil management.

Limitations of the study

We collected soil microbial data from related articles based on 16S rRNA sequencing technology; more soil metadata can be pulled from soil metagenomes in further research. Because there is no suitable soil property database that is based on future

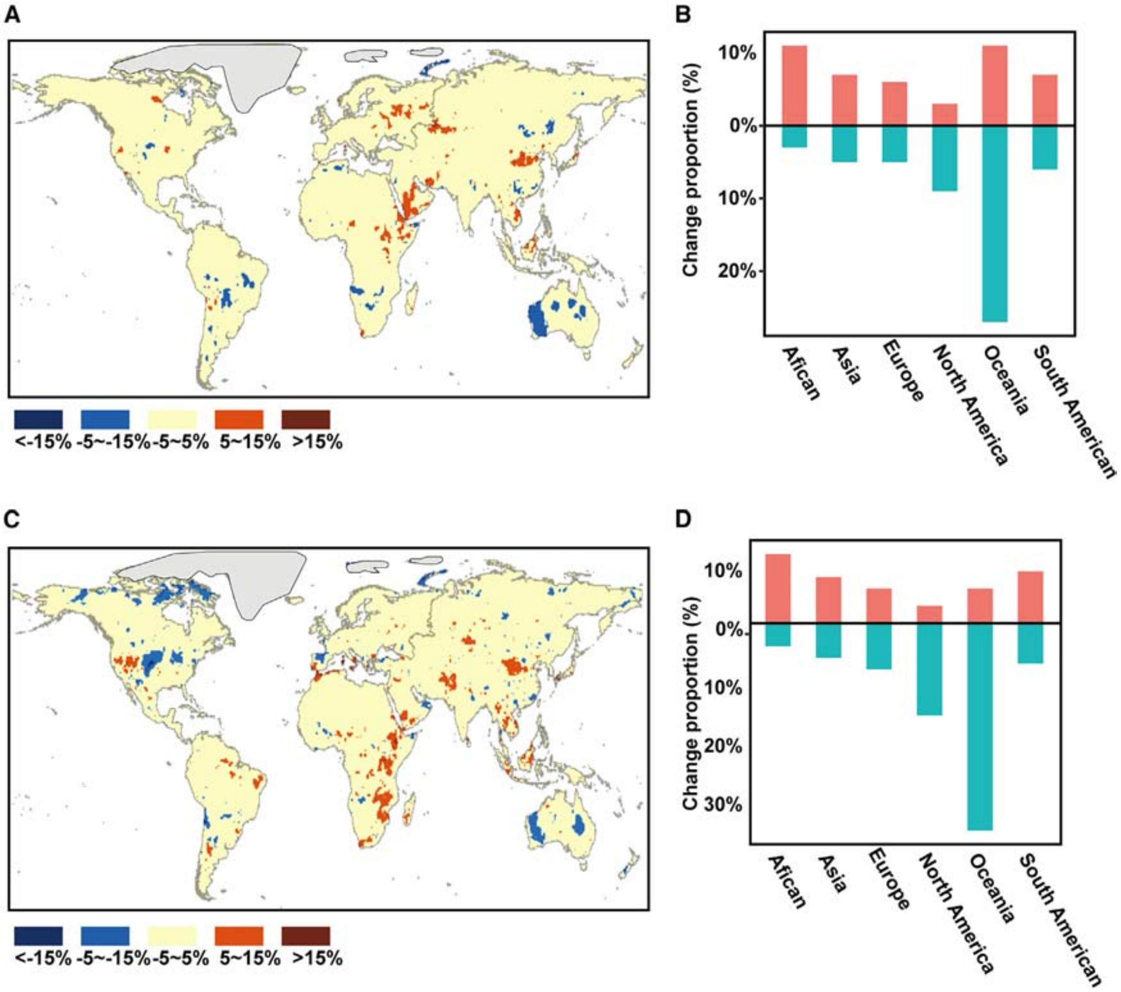


Figure 6. Change proportions of the NFMs from year 2022 to 2100 under different climate scenarios

(A) Changes in the proportion of NFM abundance under SSP 2.6.

(B) Changes in the proportion of NFM abundance on different continents under SSP 2.6.

(C) Changes in the proportion of NFM abundance under SSP 8.5.

(D) Changes in the proportion of NFM abundance on different continents under SSP 8.5.

climate models, current soil properties were used for future predictions in our analysis. The development of authoritative databases on soil properties under future climate scenarios can make predictions more accurate.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Li Mu (muli@caas.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

- The experimental data have been presented in Tables S1 and S2.
- The Python codes used in this study are available at Figshare (https://figshare.com/articles/dataset/Environmental_Tipping_Points_for_Global_Soil_Nitrogen_Fixation_Microorganisms/25549570).
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was financially supported by the National Key Research and Development Program of China (2024YFD1700100), the National Natural Science Foundation of China (grant nos. 22176103 and U22A20615) and the Ministry of Education of China (grant no. T2017002).

AUTHOR CONTRIBUTIONS

L.M., designed the project and improved the manuscript; Y.H. and H.L., collected the data, ran the models, and wrote the manuscript; J.L., collected the data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - ○ Data collection
  - ○ Relative abundance of microbial NFMs in soil
  - ○ Data preprocessing
  - ○ Machine learning regression and validation
  - ○ Model robustness evaluation and uncertainty assessment
  - ○ Model interpretability analysis
  - ○ Global prediction
  - ○ Identification of tipping points
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111634.

Received: June 2, 2024
Revised: August 3, 2024
Accepted: December 17, 2024
Published: December 19, 2024

REFERENCES

1. Nelson, M.B., Martiny, A.C., and Martiny, J.B.H. (2016). Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil. Proc. Natl. Acad. Sci. USA *113*, 8033–8040.
2. Yuan, H., Ge, T., Chen, C., O'Donnell, A.G., and Wu, J. (2012). Significant Role for Microbial Autotrophy in the Sequestration of Soil Carbon. Appl. Environ. Microbiol. *78*, 2328–2336.
3. Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., Averill, C., and Maynard, D.S. (2019). The Global Soil Community and Its Influence on Biogeochemistry. Science *365*, eaav0550.
4. Harwood, C.S. (2020). Iron-Only and Vanadium Nitrogenases: Fail-Safe Enzymes or Something More? Annu. Rev. Microbiol. *74*, 247–266.
5. Wall, D.H., Nielsen, U.N., and Six, J. (2015). Soil Biodiversity and Human Health. Nature *528*, 69–76.
6. Widdig, M., Heintz-Buschart, A., Schleuss, P., Alexander, G., Borer, E., Seabloom, E., and Spohn, M. (2020). Effects of Nitrogen and Phosphorus Addition on Microbial Community Composition and Element Cycling in a Grassland Soil. Soil Biol. Biochem. *151*, 108041.
7. Zheng, M., Zhou, Z., Zhao, P., Luo, Y., Ye, Q., Zhang, K., Song, L., and Mo, J. (2020). Effects of Human Disturbance Activities and Environmental Change Factors on Terrestrial Nitrogen Fixation. Glob. Chang. Biol. *26*, 6203–6217.
8. Canfield, D.E., Glazer, A.N., and Falkowski, P.G. (2010). The Evolution and Future of Earth's Nitrogen Cycle. Science *330*, 192–196.
9. Jiang, J., Li, Z., Xiao, H., Wang, D., Liu, C., Zhang, X., Peng, H., and Zeng, G. (2018). Labile Organic Matter Plays a More Important Role Than the Autotrophic Bacterial Community in Regulating Microbial CO_2 Fixation in an Eroded Watershed. Land Degrad. Dev. *29*, 4415–4423.
10. Carrier-Belleau, C., Pascal, L., Nozais, C., and Archambault, P. (2022). Tipping Points and Multiple Drivers in Changing Aquatic Ecosystems: A Review of Experimental Studies. Limnol. Oceanogr. *67*, S312–S330.
11. Ban, Z., Hu, X., and Li, J. (2022). Tipping Points of Marine Phytoplankton to Multiple Environmental Stressors. Nat. Clim. Chang. *12*, 1045–1051.
12. Wang, Y.-P., and Houlton, B.Z. (2022). Climate Tipping Point of Nitrogen Fixation. Nat. Plants *8*, 196–197.
13. Ban, Z., Yuan, P., Yu, F., Peng, T., Zhou, Q., and Hu, X. (2020). Machine Learning Predicts the Functional Composition of the Protein Corona and the Cellular Recognition of Nanoparticles. Proc. Natl. Acad. Sci. USA *117*, 10492–10499.
14. Yu, F., Wei, C., Deng, P., Peng, T., and Hu, X. (2021). Deep Exploration of Random Forest Model Boosts the Interpretability of Machine Learning Studies of Complicated Immune Responses and Lung Burden of Nanoparticles. Sci. Adv. *7*, eabf4130.
15. Karimi, B., Terrat, S., Dequiedt, S., Saby, N.P.A., Horrigue, W., Lelièvre, M., Nowak, V., Jolivet, C., Arrouays, D., Wincker, P., et al. (2018). Biogeography of Soil Bacteria and Archaea across France. Sci. Adv. *4*, eaat1808.
16. Buchanan, P.J., Chase, Z., Matear, R.J., Phipps, S.J., and Bindoff, N.L. (2019). Marine Nitrogen Fixers Mediate a Low Latitude Pathway for Atmospheric CO_2 Drawdown. Nat. Commun. *10*, 4611.
17. Zheng, M., Chen, H., Li, D., Luo, Y., and Mo, J. (2020). Substrate Stoichiometry Determines Nitrogen Fixation Throughout Succession in Southern Chinese Forests. Ecol. Lett. *23*, 336–347.
18. Nelson, M.B., Martiny, A.C., and Martiny, J.B.H. (2016). Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil. Proc. Natl. Acad. Sci. USA *113*, 8033–8040.
19. Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., et al. (2018). Structure and Function of the Global Topsoil Microbiome. Nature *560*, 233–237.
20. Li, D., Zhang, Q., Xiao, K., Wang, Z., and Wang, K. (2018). Divergent Responses of Biological Nitrogen Fixation in Soil, Litter and Moss to Temperature and Moisture in a Karst Forest, Southwest China. Soil Biol. Biochem. *118*, 1–7.
21. Rillig, M.C., Ryo, M., Lehmann, A., Aguilar-Trigueros, C.A., Buchert, S., Wulf, A., Iwasaki, A., Roy, J., and Yang, G. (2019). The Role of Multiple Global Change Factors in Driving Soil Functions and Microbial Biodiversity. Science *366*, 886–890.
22. Das, S., and De, T.K. (2018). Microbial Assay of N_2 Fixation Rate, a Simple Alternate for Acetylene Reduction Assay. Methods *5*, 909–914.
23. Zhang, X., Jia, X., Wu, H., Li, J., Yan, L., Wang, J., Li, Y., and Kang, X. (2020). Depression of Soil Nitrogen Fixation by Drying Soil in a Degraded Alpine Peatland. Sci. Total Environ. *747*, 141084.
24. Chen, D., Hou, H., Zhou, S., Zhang, S., Liu, D., Pang, Z., Hu, J., Xue, K., Du, J., Cui, X., et al. (2022). Soil Diazotrophic Abundance, Diversity, and Community Assembly Mechanisms Significantly Differ between Glacier Riparian Wetlands and Their Adjacent Alpine Meadows. Front. Microbiol. *13*, 1063027.
25. Fan, K., Delgado-Baquerizo, M., Guo, X., Wang, D., Wu, Y., Zhu, M., Yu, W., Yao, H., Zhu, Y.G., and Chu, H. (2019). Suppressed N Fixation and Diazotrophs after Four Decades of Fertilization. Microbiome *7*, 143.
26. Lei, S., Wang, X., Wang, J., Zhang, L., Liao, L., Liu, G., Wang, G., Song, Z., and Zhang, C. (2024). Effect of Aridity on the β-Diversity of Alpine Soil Potential Diazotrophs: Insights into Community Assembly and Co-Occurrence Patterns. mSystems *9*, e01042.
27. Scheibe, A., and Spohn, M. (2022). N_2 Fixation Per Unit Microbial Biomass Increases with Aridity. Soil Biol. Biochem. *172*, 108733.
28. Liu, W., Wang, Z., Liu, X., Zeng, N., and Bell, D. (2019). A Novel Particle Swarm Optimization Approach for Patient Clustering from Emergency Departments. IEEE Trans. Evol. Comput. *23*, 632–644.
29. Lundberg, S.M., and Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777.
30. Berdugo, M., Delgado-Baquerizo, M., Soliveres, S., Hernández-Clemente, R., Zhao, Y., Gaitán, J.J., Gross, N., Saiz, H., Maire, V., Lehmann, A., et al.

(2020). Global Ecosystem Thresholds Driven by Aridity. Science *367*, 787–790.

31. Fan, K., Weisenhorn, P., Gilbert, J.A., Shi, Y., Bai, Y., and Chu, H. (2018). Soil Ph Correlates with the Co-Occurrence and Assemblage Process of Diazotrophic Communities in Rhizosphere and Bulk Soils of Wheat Fields. Soil Biol. Biochem. *121*, 185–192.
32. Griebsch, A., Matschiavelli, N., Lewandowska, S., and Schmidtke, K. (2020). Presence of Bradyrhizobium Sp. Under Continental Conditions in Central Europe. Agriculture *10*, 446.
33. Chen, H., Zheng, C., Qiao, Y., Du, S., Li, W., Zhang, X., Zhao, Z., Cao, C., and Zhang, W. (2021). Long-Term Organic and Inorganic Fertilization Alters the Diazotrophic Abundance, Community Structure, and Co-Occurrence Patterns in a Vertisol. Sci. Total Environ. *766*, 142441.
34. Sheffer, E., Batterman, S.A., Levin, S.A., and Hedin, L.O. (2015). Biome-Scale Nitrogen Fixation Strategies Selected by Climatic Constraints on Nitrogen Cycle. Nat. Plants *1*, 15182.
35. Davies-Barnard, T., and Friedlingstein, P. (2020). The Global Distribution of Biological Nitrogen Fixation in Terrestrial Natural Ecosystems. Global Biogeochem. Cy. *34*, e2019GB006387.
36. Smercina, D.N., Evans, S.E., Friesen, M.L., and Tiemann, L.K. (2021). Temporal Dynamics of Free-Living Nitrogen Fixation in the Switchgrass Rhizosphere. GCB Bioenergy *13*, 1814–1830.
37. Yu, H., Li, L., Ma, Q., Liu, X., Li, Y., Wang, Y., Zhou, G., and Xu, Z. (2023). Soil Microbial Responses to Large Changes in Precipitation with Nitrogen Deposition in an Arid Ecosystem. Ecology *104*, e4020.
38. Hu, W., Wang, X., Xu, Y., Wang, X., Guo, Z., Pan, X., Dai, S., Luo, Y., and Teng, Y. (2024). Biological Nitrogen Fixation and the Role of Soil Diazotroph Niche Breadth in Representative Terrestrial Ecosystems. Soil Biol. Biochem. *189*, 109261.
39. Hou, L., Wang, R., Yin, G., Liu, M., and Zheng, Y. (2018). Nitrogen Fixation in the Intertidal Sediments of the Yangtze Estuary: Occurrence and Environmental Implications. JGR. Biogeosciences *123*, 936–944.
40. Pushkareva, E., Pessi, I.S., Wilmotte, A., and Elster, J. (2015). Cyanobacterial Community Composition in Arctic Soil Crusts at Different Stages of Development. FEMS Microbiol. Ecol. *91*, fiv143.
41. Wong, M.Y., Neill, C., Marino, R., Silvério, D., and Howarth, R.W. (2021). Molybdenum, Phosphorus, and Ph Do Not Constrain Nitrogen Fixation in a Tropical Forest in the Southeastern Amazon. Ecology *102*, e03211.
42. Reis, C.R.G., Pacheco, F.S., Reed, S.C., Tejada, G., Nardoto, G.B., Forti, M.C., and Ometto, J.P. (2020). Biological Nitrogen Fixation across Major Biomes in Latin America: Patterns and Global Change Effects. Sci. Total Environ. *746*, 140998.
43. O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci. Model Dev. (GMD) *9*, 3461–3482.
44. Hurtt, G.C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B.L., Calvin, K., Doelman, J.C., Fisk, J., Fujimori, S., Klein Goldewijk, K., et al. (2020). Harmonization of Global Land Use Change and Management for the Period 850–2100 (Luh2) for CMIP6. Geosci. Model Dev. (GMD) *13*, 5425–5464.
45. Ferdush, J., Paul, V., Varco, J., Jones, K., and Sasidharan, S.M. (2023). Consequences of Elevated CO_2 on Soil Acidification, Cation Depletion, and Inorganic Carbon: A Column-Based Experimental Investigation. Soil Tillage Res. *234*, 105839.
46. Tian, D., and Niu, S. (2015). A Global Analysis of Soil Acidification Caused by Nitrogen Addition. Environ. Res. Lett. *10*, 024019.
47. Guo, J.H., Liu, X.J., Zhang, Y., Shen, J.L., Han, W.X., Zhang, W.F., Christie, P., Goulding, K.W.T., Vitousek, P.M., and Zhang, F.S. (2010). Significant Acidification in Major Chinese Croplands. Science *327*, 1008–1010.
48. Neina, D. (2019). The Role of Soil pH in Plant Nutrition and Soil Remediation. Appl. Environ. Soil Sci. *9*, 5794869.
49. Hartemink, A.E., and Barrow, N.J. (2023). Soil pH-Nutrient Relationships: The Diagram. Plant Soil *486*, 209–215.
50. Walker, T.W.N., Kaiser, C., Strasser, F., Herbold, C.W., Leblans, N.I.W., Woebken, D., Janssens, I.A., Sigurdsson, B.D., and Richter, A. (2018). Microbial Temperature Sensitivity and Biomass Change Explain Soil Carbon Loss with Warming. Nat. Clim. Chang. *8*, 885–889.
51. Wang, M., Guo, X., Zhang, S., Xiao, L., Mishra, U., Yang, Y., Zhu, B., Wang, G., Mao, X., Qian, T., et al. (2022). Global Soil Profiles Indicate Depth-Dependent Soil Carbon Losses under a Warmer Climate. Nat. Commun. *13*, 5514.
52. Clarke, A.C., Prost, S., Stanton, J.A.L., White, W.T.J., Kaplan, M.E., and Matisoo-Smith, E.A.; Genographic Consortium (2014). From Cheek Swabs to Consensus Sequences: An a to Z Protocol for High-Throughput DNA Sequencing of Complete Human Mitochondrial Genomes. BMC Genom. *15*, 68.
53. Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., Williams, R.J., Flater, J., Tiedje, J., Hofmockel, K., Gelder, B., et al. (2016). Refsoil: A Reference Database of Soil Microbial Genomes. bioRxiv *4*, 053397. https://doi.org/10.1101/053397.
54. Topcuoglu, B.D., Lesniak, N.A., Ruffin, M.T., 4th, Wiens, J., and Schloss, P.D. (2020). A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. mBio *11*, e00434.
55. Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. (2002). Smote: Synthetic Minority over-Sampling Technique. J. Artif. Intell. Res. *16*, 321–357.
56. Ilyas, I.F., Beskales, G., and Soliman, M.A. (2008). A Survey of Top-K Query Processing Techniques in Relational Database Systems. ACM Comput. Surv. *40*, 1–58.
57. Strumbelj, E., Kononenko, I.J.K., and Systems, I. (2014). Explaining Prediction Models and Individual Predictions with Feature Contributions. Knowl. Inf. Syst. *41*, 647–665.
58. Still, C.J., Berry, J.A., Collatz, G.J., and DeFries, R.S. (2003). Global Distribution of C3 and C4 Vegetation: Carbon Cycle Implications. Global Biogeochem. Cy. *17*, 1006.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
The Python codes used in this study are available at Figshare	This paper	https://figshare.com/articles/dataset/Environmental_Tipping_Points_for_Global_Soil_Nitrogen_Fixation_Microorganisms/25549570
Software and algorithms		
Random forests regression	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
Logistic regression	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
Degree-2 polynomial margin regression	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
Ridge regression	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
LASSO regression	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
SHapley Additive exPlanations (SHAP) model	Scikit-learn (version1.0.1): Machine Learning in Python 3.8	https://scikit-learn.org/stable/index.html
R software version 4.1.2	R software	https://www.r-project.org/
Python version 3.8	Python Software	https://www.python.org/
ArcGIS 10.7	ArcGIS Desktop	https://desktop.arcgis.com/

METHOD DETAILS

Data collection

The data compilation focused on data from the literature related to soil microbial communities. Given that the second-generation sequencing technology established in the year 2005 gradually matured and began to be applied in a variety of studies by the year 2010,[52] the data we collected from relevant studies were published from January 2010 to December 2022. The keywords searched in the Web of Science database were "soil," "bacteri*" and "fung*", and the initial search returned 1,931,418 studies. The following criteria were used to screen for appropriate studies: (1) the study was a field study, not a laboratory experimental study; (2) the microbial composition at the phylum level was reported; and (3) the soil microbial community information was quantified via high-throughput sequencing techniques. Ultimately, 285 articles (shown in Table S1) were selected, which made 1659 observations from 595 locations worldwide (the sample locations are shown in Figure S1). Sixteen environmental variables, including altitude, AI, MAT, MAP, pH, TC, OM, OC, TN, TP, ASW, texture, sampling depth and artificial management status (i.e., use of N/P/K fertilizers), were recorded for analysis. The N/P/K fertilizer data were only collected during the experimental period. The historical fertilizer data were not collected because of a lack of data. Different types of fertilizers, such as chemical fertilizers, organic fertilizers, and straw, were converted into N/P/K contents and added together. Longitude and latitude data were also collected. In cases where the studies did not report MAT or MAP, the values were derived from historical monthly weather data from the Database: WorldClim (https://www.worldclim.org) database using site geographic location (i.e., latitude and longitude). The missing data pertaining to the properties of global soils (e.g., pH, soil total C, soil organic matter, soil total N and soil total P) were filled in by retrieving data from the Database: International Soil Reference and Information Center (ISRIC) World Soil Information Data Hub (http://www.tpdc.ac.cn). The missing data for soil organic carbon, available soil water and texture were filled with data from the Database: Harmonized World Soil Database (HWSD) (http://www./soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). In cases where the studies did not report the latitude or longitude, the approximate latitude and longitude were derived by geocoding the name of the location in Google Earth 7.0. For studies that did not report altitude, the values were also derived from Google Earth 7.0 using site geographic location (i.e., latitude and longitude). The aridity index data were collected from the Database: Global Aridity Index and Potential Evapotranspiration (ET0) Database, Version 3 (Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 (figshare.com)). We considered that various climate and soil properties exist in any given habitat and that further subdivision would lead to insufficient data. Therefore, we did not include habitat in the machine learning model.

Relative abundance of microbial NFMs in soil

The relative abundance of NFM was defined as the proportion of microbes with N fixation ability (shown in Table S2):

$$K = \frac{\sum_{i=1}^{n} B_i * \frac{N_k}{N}}{\sum_{i=1}^{n} B_i} \qquad \text{(Equation 1)}$$

where N_k is the number of NFMs in phylum i; N is the number of species in phylum i; B_i is the relative abundance of bacterial phylum i; and n is the number of phyla.

Information about common soil microbial species (n = 851) was obtained from the RefSoil database.[53] *nifH* was selected as the marker gene since molybdenum (Mo)-dependent (Nif) is the most well-characterized and most commonly occurring form of nitrogenase worldwide.[4] Microbial functional traits were derived from the Database: National Center for Biotechnology Information (NCBI) (nih.gov). Functional microbes were defined as species with relevant genes or verified functions according to the published literature Table S2).

Data preprocessing

Given that null and extreme data lead to unreliable conclusions and undermine the robustness of a model, the interquartile range (IQR) criterion was adopted to exclude outliers.[54] The IQR is defined as Q3-Q1, and (Q1, Q3) covers the middle 50% of the data values. If the data are located in (Q1-1.5·IQR, Q3+1.5·IQR), then they are considered normal data. Otherwise, the data are recognized as outliers. Before building the machine learning model, the synthetic minority oversampling technique (SMOTE) was used to solve the problem of unbalanced positive and negative samples in the dataset.[55] The positive samples were upsampled to obtain a balanced dataset. Before upsampling the dataset, the training set and test set were divided by 5-fold cross-validation and only the training set was upsampled. Consequently, the data dependence caused by upsampling and data division was eliminated, making the results credible.

Machine learning regression and validation

A random forest model with the *scikit-learn* package (version 1.0.1) in Python (version 1.1.3) was used as the base model. Using the randomly occurring distributed delayed particle swarm optimization (RODDPSO) algorithm,[28] hyperparameter optimization and loss function optimization were conducted for the random forest. The best *max_features* and *n_estimators* were 6 and 670, respectively. The parameters c_{1i}, c_{2i}, c_{1f} and c_{2f} were set as 2.5, 0.5, 0.5 and 2.5, respectively. c_{1i} and c_{2i} represent the initial values of the acceleration coefficients. c_{1f} and c_{2f} denote the final values of the cognitive acceleration coefficient c_1 and the social acceleration coefficient c_2, respectively. Parameters m_l and m_g were set as $0.m_L$ and m_g represent the intensity factors of the distributed time delay terms. The parameters w_{max} and w_{min} were set as 0.8 and 0.3, respectively. w_{max} and w_{min} represent the maximum and minimum values of the inertia weight, respectively. The population size *sizepop* was 50. The iterative times *maxgen* was 20.

Model robustness evaluation and uncertainty assessment

The models were based on the above preprocessed data. However, in the process of data cleaning, it is possible to remove data that are useful but not robust after removing outlier data.[56] Because chaos in models is sensitive to initial values, a trained model may not maintain the same performance as a new dataset after data preprocessing. Thus, adversarial samples were introduced to test the robustness of the models. The independent variables of the model are increased in a certain proportion to test the model's tolerance for data noise. To determine whether the model R^2 changes sharply with increasing data size, the independent variables of the model are increased in a certain proportion to test the uncertainty of the model and to determine the trend and confidence interval of the dependent variable.

Model interpretability analysis

Because the model was based on a black box model, the process from input to output was not clear, and the importance function made it difficult to reasonably explain the output and conclusion.[57] When evaluating the robustness of a model, it is important to identify which factors lead to low robustness. All of the abovementioned factors make it necessary to ensure the interpretability of the machine learning model. SHAPley Additive exPlanations (SHAPs) estimate the contribution of each feature by averaging over all of the possible marginal contributions to a prediction task and constitute a unified framework for interpreting machine learning models.[29] The SHAP value model is a relatively versatile method of model interpretability that can be used not only for global interpretation but also for local interpretation. The possible relationship between the predicted value given by the model and some features can be explained by the SHAP value model.[29]

The SHAP value model is a method of post hoc model interpretation.[29] Its core principle is to calculate the marginal contribution of features to the model output and then explain the black box model at the global and local levels. It constructs an additive interpretation model in which all features are regarded as contributors. For each prediction sample, the model generates a prediction value.

The SHAP value is the value assigned to each feature in the sample, and it can be defined as the average marginal contribution of eigenvalues in all possible coalitions. The SHAP value calculation of a single feature eliminating cross effects is as follows:

$$\Phi_{i,i} = \phi_i - \sum_{j \neq i} \Phi_{i,j} \qquad \text{(Equation 2)}$$

where $\Phi_{i,i}$ represents the contribution of a single feature; ϕ_i represents the solution of the SHAP value of the tree-based model; and $\Phi_{i,j}$ represents the cross influence of two features.

The Python SHAP package was derived from the *scikit-learn* package (version 1.0.1). On the basis of the created and fitted model, the SHAP package was used to construct an additive interpretation model, and the dataset X_train was used to calculate the SHAP value.

Global prediction

The Scenario Model Intercomparison Project (ScenarioMIP), which is part of this project, provides multimodel climate projections based on alternative scenarios of future emissions and land use changes.[43] The established models were used to evaluate the distribution patterns of NFM abundance for the year 2021 and two predicted shared socioeconomic pathway climate scenarios (SSP2.6 and SSP8.5) for the year 2100. SSP2.6 represents a sustainability scenario, and SSP8.5 represents a fossil fuel development scenario.

The monthly temperature and precipitation data for year 2021 were obtained from the Database: National Oceanic and Atmospheric Administration (National Oceanic and Atmospheric Administration (noaa.gov)), with resolutions of 0.5° × 0. 5°. The future MAT and MAP data were obtained from the Database: Climate Model Intercomparison Project Phase 6 (CMIP6) at a 1 ° × 1 ° resolution (cmip6 Data Search | cmip6 | ESGF-CoG (llnl.gov)). The current and future fertilizer data were derived from the Database: Land-Use Harmonization (LUH2) project (Land Use Harmonization (umd.edu)). The resolution was 1 ° × 1 °. In the latitudinal range of −30°–10°, C_4 plants are dominant,[58] so C_4 plant fertilizer data were used. C_3 plant data were employed for the other regions. Global soil properties (pH, TC, OM, TN, and TP) with a 10 km resolution were obtained from the Database: Global Soil Dataset for Earth System Modeling (2014) (data.tpdc.ac.cn). The global data of soil organic carbon, available soil water and texture at a 1 ° × 1 ° resolution were derived from the Database: Harmonized World Soil Database HWSD (http://www./soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). Because there is no suitable soil property database that is based on future climate models, current soil properties were used for future predictions in our analysis. The data were resampled to a spatial resolution of 1 ° × 1 °, and all terrestrial grid cells were input into the established models to obtain the global distribution pattern.

Identification of tipping points

Two types of tipping points (discontinuous and continuous) were identified and analyzed. Continuous tipping points indicate that the relationship between independent and dependent variables changes significantly; discontinuous tipping points indicate that the value of the dependent variable changes abruptly because of a small change in the independent variable.[30] Two models were used for regression analysis: a generalized additive model (GAM) and a segmented linear regression (SLR) model. The tipping point of the GAM was defined as the point with the second derivative as 0 in the continuous curve; the tipping point of the SLR was defined as the overall change in the intercept and slope in the linear regression from before to after the tipping point.[30] We used the *segmented* and *mgcv* packages in R to fit the SLR and GAM regressions, respectively. The tipping points obtained from the model with the best performance (highest R^2) were used for further analysis. Because the SHAP values reflect the contribution of the independent variable to the dependent variable and can represent the action of the independent variable alone, we also used the GAM and SLR models to calculate the tipping points of the independent variables and the corresponding SHAP values. Only when the tipping points came from raw data and the SHAP values were similar were they considered to be the actual tipping points.

QUANTIFICATION AND STATISTICAL ANALYSIS

All of the statistical analyses in this work were conducted in R software (version 4.1.2) or Python 3.8.6. All of the global maps were built using ArcGIS 10.7. All confidence levels not otherwise specified were 0.95.