Forest migration outpaces tree species range shift across North America

Akane Abbasi  
Purdue University

Christopher Woodall  
United States Department of Agriculture, Forest Service

Javier Gamarra  
Food and Agriculture Organization of the United Nations  https://orcid.org/0000-0002-1290-9559

Thomas Ochuodho  
Department of Forestry and Natural Resources, University of Kentucky

Sergio de-Miguel  
Department of Crop and Forest Sciences, University of Lleida

Rajeev Sahay  
School of Electrical and Computer Engineering, Purdue University

Songlin Fei  
Purdue University West Lafayette  https://orcid.org/0000-0003-2772-0166

Alain Paquette  
Université du Québec à Montréal  https://orcid.org/0000-0003-1048-9674

Han Chen  
Lakehead University  https://orcid.org/0000-0001-9477-5541

Ann Christine Catlin  
Rosen Center for Advanced Computing, Purdue University

Jingjing Liang  (✉ albeca.liang@gmail.com)  
Purdue University

Biological Sciences - Article

Keywords: forest migration, migration patterns, North America, tree species range shift

Posted Date: August 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-840978/v1

License: ☛ ☀ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Forest migration outpaces tree species range shift across North America

Authors: Akane O. Abbasi\(^1\), Christopher W. Woodall\(^2\), Javier G. P. Gamarra\(^3\), Thomas Ochuodho\(^4\), Sergio de-Miguel\(^5,6\), Rajeev Sahay\(^7\), Songlin Fei\(^8\), Alain Paquette\(^9\), Han Y. H. Chen\(^10\), Ann Christine Catlin\(^11\), Jingjing Liang\(^1*\)

Affiliations:

1. Forest Advanced Computing and Artificial Intelligence Lab (FACAI), Department of Forestry and Natural Resources, Purdue University; West Lafayette, Indiana, USA.
2. United States Department of Agriculture, Forest Service; Durham, New Hampshire, USA.
3. National Forest Monitoring (NFM) Team, Forestry Division, Food and Agriculture Organization of the United Nations, Rome, Italy.
4. Department of Forestry and Natural Resources, University of Kentucky; Lexington, Kentucky, USA.
5. Department of Crop and Forest Sciences, University of Lleida, Lleida, Spain.
6. Joint Research Unit CTFC – AGROTECNIO – CERCA, Solsona, Spain.
7. School of Electrical and Computer Engineering, Purdue University; West Lafayette, Indiana, USA.
8. Department of Forestry and Natural Resources, Purdue University; West Lafayette, Indiana, USA.
9. Centre for Forest Research, Université du Québec à Montréal; Montréal, Québec, Canada.
10. Faculty of Natural Resources Management, Lakehead University; Thunder Bay, Ontario, Canada.
11. Rosen Center for Advanced Computing, Purdue University; West Lafayette, Indiana, USA.

* Corresponding author. Email: albeca.liang@gmail.com
Abstract:
Mounting evidence suggests that geographic ranges of tree species worldwide are shifting under global environmental change, but little is known about forest migration—the shift in the geographic ranges of forest types—and how it differs from individual tree species migration. Here, based on in situ records of more than 9 million trees from 596,282 sample plots, we quantified and compared the migration patterns of forests and tree species across North America between 1970 and 2019. On average, forests migrated at a mean velocity of 205.2 km·decade⁻¹, which is twice as fast as species-level migration (95.6 km·decade⁻¹), and 12 times faster than the average of previous estimates (16.3 km·decade⁻¹). Our findings suggest that as subtle perturbations in species abundance can aggregate to change an entire forest from one type to another, failing to see the forest for the trees may result in a gross underestimation of the impacts of global change on forest ecosystem functioning and services. With the first forest classification and quantification of forest migration patterns at a continental level, this study provides an urgently needed scientific basis for a new paradigm of adaptive forest management and conservation under a rapid forest migration.

Main text
Trees are immobile organisms, but tree species worldwide are found to undergo substantial changes in geographic distributions under global environmental change. Some tree species move to higher latitudes, tracking warming climate¹⁻⁴, while some move towards lower latitudes⁴⁻⁵, longitudinally⁶, or altitudinally⁷⁻⁸. Collectively, these changes can alter the relative abundance and dominance of tree species, causing a complete change in the type of local forest
communities. To differentiate from the term *tree species migration* which refers to the shift of
tree species ranges\(^3\)–\(^6\), we call the shift in the geographic range of a forest type *forest migration*.

Quantifying forest migration is crucial for the understanding of the impacts of global change on forest ecosystem functioning and services. A forest constitutes a foundational entity supporting most ecosystem services as well as human culture, customs, economies, and identity\(^9\)–\(^12\). In addition, a forest is a fundamental unit of sustainable forest management\(^11\). By shifting local forest types, forest migration can extensively change ecosystem functioning and services\(^13\)–\(^17\), causing massive ecological and socioeconomic impacts worldwide\(^18\). For instance, in the central United States, a diminishing supply of *Quercus alba*, *Q. macrocarpa*, and other white oak species caused by the shifting and shrinking ranges of oak-dominated forests is threatening the bourbon industry\(^19\), a staple of American culture and tradition. Meanwhile, the migration of maple-dominated forests has raised concerns over the sustainability of the maple syrup industry in North America\(^20\).

To see the forest for the trees is a major challenge in quantifying forest migration. Previous studies found that the geographic ranges of some tree species in North America shifted at a mean velocity of 16.3 km·decade\(^{-1}\), with a range of 0.03 –100.20 km·decade\(^{-1}\) (see Supplementary Table 1). However, because these studies were limited to a local or regional scale with inconsistent migration measurements (*e.g.*, some use marginal shifts, but some use centroid or latitudinal shifts), the patterns of forest migration at a continental scale still remain largely unknown\(^21\). Moreover, as most reported migration velocities were calculated from species-level shifts, how forest migration differs from tree species migration also remains largely unknown.

Here, we systematically quantified, for the first time, forest and tree species migration patterns at a continental scale, based on more than 9 million ground-surveyed tree records from
596,282 sample plots. Using these *in situ* data, we classified North American forests into a hierarchal system consisting of eight forest biomes and 51 underlying forest types (Table 1, sans forests in Mexico, Central America, and the Caribbean due to a lack of data). We then quantified the azimuth and velocity of forest migration between 1970–1999 and 2000–2019. Similarly, we quantified the azimuth and velocity of tree species migration across the continent during the same time periods.

To quantify forest migration, we first used an established machine learning algorithm to consistently classify all forested areas across the study region into 51 forest types (Table 1, see §Forest Classification in Methods). There are 49 forest types in eight biomes in the conterminous United States and Alaska, and 35 forest types in six biomes throughout Canada (Fig. 1, Extended Data Figs. 1, 2). The two countries share a total of six forest biomes. The Boreal Forest (total area 2,462,924 km\(^2\)) is the largest forest biome shared by the two countries, followed by the Eastern Mixed Forest (644,011 km\(^2\)). Mediterranean California (59,849 km\(^2\)) and Southern Plains (547,118 km\(^2\)) are only distributed in the United States. At the forest type level, black spruce–balsam fir (B-E, 750,121 km\(^2\)) is the largest forest type shared by the United States and Canada, followed by quaking aspen–balsam fir–paper birch (B-A, 349,949 km\(^2\)) and jack pine—black spruce (B-C, 294,291 km\(^2\)). The largest non-boreal forest types shared by the two countries are balsam fir–maple–yellow birch (E-I, 229,088 km\(^2\)) and subalpine fir—Engelmann spruce (W-K, 203,700 km\(^2\)).

Based on the temporal differences of the range of forest types classified above, we quantified the patterns of forest migration in terms of the velocity and azimuth. Among the 43 forest types in eight forest biomes that were present in both periods 1970–1999 and 2000–2019 across the continent (Table 1), quaking aspen—balsam fir—paper birch forest (B-A) migrated...
with the highest velocity at 683.3 km·decade⁻¹, moving eastward (Table 1, Supplementary Table 2). Among the twelve forest types that migrated at a speed between 100 and 440 km·decade⁻¹, five are in the Eastern Mixed Forest biome (E-A, E-C, E-H, E-J, and E-K), three in the Pacific Coastal Forest biome (W-A, W-B, and W-D), and one in the Western Cordillera (W-J), Mediterranean California (W-Q), Central Forest (E-M), and Boreal Forest biome (B-E), respectively (Fig. 1). The remaining forest types migrated at less than 100 km·decade⁻¹. In terms of the direction of migration, 16 out of 43 forest types migrated westward, 11 eastward, nine southward, and seven northward in the past 50 years (Table 1). Across the continent, forests migrated at a mean velocity of 205.2 km·decade⁻¹ (Fig. 2a).

At the tree species level, we estimated the geographic range of 150 tree species in North America for the same time period to quantify tree species migration. We found that tree species on average migrated at 95.6±1.7 km·decade⁻¹ (Fig. 2b). *Picea sitchensis* had the greatest migration velocity of all the tree species (504.8 km·decade⁻¹), followed by *Abies balsamea* (502.0 km·decade⁻¹) and *Alnus incana* (359.4 km·decade⁻¹). In contrast, *Platanus occidentalis* had the lowest migration velocity (4.3 km·decade⁻¹), followed by *Quercus macrocarpa* (4.9 km·decade⁻¹) and *Celtis laevigata* (5.3 km·decade⁻¹) (Supplementary Table 3). Across the continent, we found that tree species migrated at a mean velocity five times greater than the average of previous estimates (16.3 km·decade⁻¹, with a range of 0.03 –100.20 km·decade⁻¹, see Supplementary Table 1). This difference in species-level migration velocity between current and previous studies can be mainly attributed to the fact that very few boreal species have been covered in previous studies. Because boreal tree species were found here to migrate faster in general, and boreal region constitutes the largest forest biome in North America, a lack of boreal tree species coverage in previous studies has resulted in an underestimation of species-level
migration velocity at a continental scale. Nevertheless, for temperate biomes, the species migration velocity estimated here is generally consistent with previous estimates. For instance, we estimated that tree species on average migrated at 81.1±1.1 km·decade^{-1} in the eastern region (Fig. 2b), which is consistent with the previous estimates for eastern United States and Quebec, Canada (Supplementary Table 1).

Overall across the continent, forest migrated (205.2 km·decade^{-1}, Fig. 2a) more than twice as fast as tree species migration (95.6 km·decade^{-1}, Fig. 2b). The velocity was the highest for the Boreal Forest biome, where forest migrated almost three times faster than tree species migration (335.4 km·decade^{-1} vs. 113.4 km·decade^{-1}) (Fig. 2). We further examined potential drivers behind this geographic trend (Fig. 3a) from among three species diversity measures and 15 bioclimate variables. Precipitation seasonality, mean temperature of driest quarter, mean temperature of coldest quarter, and tree species evenness were identified as the most important variables (Extended Data Fig. 3). The ratio of forest migration velocity to tree species migration velocity was positively associated with climate change, an aggregated indicator of temporal changes in the top nine bioclimate variables. In contrast, the ratio was negatively associated with tree species evenness (Fig. 3b).

The substantial difference in the velocity of migration between forest type and individual tree species therein can also be attributed to the high sensitivity of forest type classification to changes in the abundance and dominance of underlying tree species. A small, local perturbation in species abundance and/or dominance, which has little impact on the overall shift of the species range, can potentially alter the local forest type and the overall forest migration pattern.

Our findings suggest that the impacts of global environmental change on forest ecosystem functioning and services may have been grossly underestimated. Since the mean
velocity of forest migration (205.2 km·decade\(^{-1}\)) estimated here is more than 12 times greater than the average of previous estimates (16.3 km·decade\(^{-1}\)), the associated impacts of on forest ecosystem functioning and services can be much more profound than previously thought.

Because forest ecosystem functioning\(^{22,23}\), productivity\(^{24}\), as well as phenology and population turnover\(^{25,26}\) are very sensitive to tree species composition and tree species diversity, subtle changes in relative abundance or relative dominance of tree species can aggregate to affect ecosystem services\(^{22–24,27}\) in a snowball effect. For example, in the eastern region, our results show that oak–hickory forest (E-M) and Appalachian oak–pine forest (E-N) migrated at 101.6 and 36.8 km·decade\(^{-1}\), respectively, despite a mere 0.02–0.1% reduction in their ranges (Table 1, Supplementary Table 2). Suppressed fire, land-use change, forest fragmentation, and climate change in this region have increased the proportion of competitive, late-successional mesophytic hardwood species (e.g., *Acer* and *Fagus* spp.), while suppressing fire-dependent xerophytic species (e.g., *Quercus* and *Pinus* spp.)\(^{28}\). This “mesophication” of the central eastern forests has already rendered profound ecological and economic impacts on soil processes, nutrient cycling, wildlife food and habitat, and local timber industry\(^{29}\). Moreover, since existing adaptive forest management regimes are based primarily on individual species range projections and associated environmental and social aspects\(^{30,31}\), it would be difficult for these regimes to fully address the consequences of rapid forest migration. To this end, the quantification of forest classification and associated forest migration patterns provided here can inform decision-making to better support assisted species migration strategies in balancing the deleterious effects of rapid forest migration\(^{32}\).

Rapid forest migration at the continental scale has profound economic and social implications. Changes in species mix would affect biophysical and environmental factors that
directly or indirectly affect timber supply\textsuperscript{33}, such as forest productivity\textsuperscript{24}, as well as frequency and severity of forest fires and pest infestations\textsuperscript{34-36}. Forest migration could potentially widen the breadth of timber baskets (\textit{i.e.}, timber procurement radii) associated with wood processing plants, hence increasing transportation costs with downstream financial implications regarding finished forest product prices. Such impacts have significant distributive (welfare) and economy-wide consequences through intersectoral linkages, making local forest industry less self-sustainable and more vulnerable to timber price fluctuations\textsuperscript{37}. Furthermore, the collective human experience (\textit{e.g.}, culture, customs, and identities) of rural communities embedded within these forested landscapes have a strong tie to surrounding forest types. From the Sitka spruce—western hemlock forests in the Pacific Coast to the oak–pine forests along the Appalachians (Table 1), the substantial decline of native forests can threaten the customs, identities, and culture of indigenous\textsuperscript{38} and other local communities, while jeopardizing the non-timber forest products supply and environmental justice overall\textsuperscript{39}. Rapid forest migration places an urgent call upon human communities, especially rural populations, to adapt their cultural norms and relationships with surrounding forests.

Our finding that on average forest migration outpaces tree species range shift by 115\% across North America can be attributed to two main factors, namely climate change and tree species evenness (Fig. 3b, Extended Data Fig. 3). Climate change is considered the top driver of forest migration, which impacts the movement, persistence, and competition within and between plant communities\textsuperscript{3,39-41}. In addition to a worldwide temperature increase by about 0.2°C per decade\textsuperscript{42}, alterations in precipitation patterns, diurnal timing, seasonal intensity, and season length are also evident across the globe\textsuperscript{42}. Consistent with previous studies of smaller geographic scales\textsuperscript{6}, we found that climate change accelerates forest migration more than it accelerates tree
species migration. In contrast, tree species evenness was found to reduce the difference between forest migration velocity and tree species migration velocity. This complements previous findings that biodiversity and species evenness in particular make forest communities more resilient to climate change\textsuperscript{43,44}.

The differed migration patterns between forests and tree species observed here represent a snapshot of a more prominent trend seen across the geological time scale. Forests, because of a high sensitivity to tree species composition changes, have over the millennia exhibited shorter life spans than individual species\textsuperscript{40}. While most tree species migrated at relatively low velocities across the continent, others went through substantial fluctuations, such as an 8\% reduction in the species range of the hemlocks between 5,400 and 4,800 BP\textsuperscript{41}. These sudden onsets of species-level range shifts have triggered forest migration across North America over the millennia.

Besides climate forcing which is generally seen as the main cause of these changes, anthropocentric disturbances, land use change, invasive species, and associated insect/diseases outbreaks are emerging as a suite of drivers that have permanently changed forest landscapes. For instance, the massive monospecific white pine (\textit{Pinus strobus}) forests that once dominated the northern forests have been replaced by mixed hardwoods, due to extensive logging since the European colonization. During the 20\textsuperscript{th} century, an outbreak of \textit{Cryphonectria parasitica} has destroyed nearly four billion American chestnut (\textit{Castanea dentata}) trees, and completely changed eastern hardwood forests of which American chestnut was a keystone species. This study supports the hypothesis that global environmental change is disrupting forests’ adaptive responses to climate change formed since the late Quaternary, and is pushing forests to migrate at an unprecedented rate\textsuperscript{41}.
Our findings depict the first continentally consistent and locally relevant record of forest classification and forest migration patterns. These results contribute fundamental insights into the rapid shifts in tree species assemblage distribution under global environmental change, and their underlying drivers. Our machine-learning analyses reveal strong effects of climate change and species evenness on forest migration patterns, and pinpointed forest communities with an extreme migration velocity, where assisted migration and other adaptive forest management efforts\textsuperscript{17} are critical in mitigating biodiversity loss, climate change, and associated socioeconomic impacts. Overall, this study provides an urgently needed scientific basis for a new paradigm of adaptive forest management and conservation, so that effective mitigation and intervention efforts can be developed in response to the rapid forest migration.

References

1. Boisvert-Marsh, L., Périé, C., & de Blois, S. Shifting with climate? Evidence for recent changes in tree species distribution at high latitudes. *Ecosphere* 5, 1-33 (2014)

2. Sittaro, F., Paquette, A., Messier, C., & Nock, C. A. Tree range expansion in eastern North America fails to keep pace with climate warming at northern range limits. *Glob. Change Biol.* 23, 3292-3301 (2017)

3. Woodall, C. W. et al. An indicator of tree migration in forests of the eastern United States. *Forest Ecol. Manag.* 257, 1434-1444 (2009)

4. Woodall, C. W. et al. Selecting tree species for testing climate change migration hypotheses using forest inventory data. *Forest Ecol. Manag.* 259, 778-785 (2010)
5. Zhu, K., Woodall, C. W., & Clark, J. S. Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Change Biol.* **18**, 1042-1052 (2012)

6. Fei, S. et al. Divergence of species responses to climate change. *Science advances*, **3**, e1603055 (2017)

7. Kelly, A. E., & Goulden, M. L. Rapid shifts in plant distribution with recent climate change. *P. Natl. A. Sci. USA* **105**, 11823-11826 (2008)

8. Lenoir, J., Gégout, J. C., Marquet, P. A., De Ruffray, P., & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**, 1768-1771 (2008)

9. Food and Agriculture Organization of the United Nations (FAO), & Platform for Agrobiodiversity Research (PAR) *Biodiversity for Food and Agriculture: Contributing to Food Security and Sustainability in a Changing World* (FAO & PAR, 2010)

10. Food and Agriculture Organization of the United Nations (FAO), & United Nations Environmental Programme (UNEP) *The State of the World’s Forests 2020. Forests, biodiversity and people* (United Nations, 2020)

11. Genin, D., Aumeeruddy-Thomas, Y., Balent, G., & Nasi, R. The multiple dimensions of rural forests: Lessons from a comparative analysis. *Ecol. Soc.* **18**, 27 (2013)

12. Watson J. E. M. et al. The exceptional value of intact forest ecosystems. *Nat. Ecol. Evol.* **2**, 599-610 (2018)

13. Neilson, R. P. et al. Forecasting regional to global plant migration in response to climate change. *BioScience* **55**, 749-759 (2005)
14. Kremen, C., & Merenlender, A. M. Landscapes that work for biodiversity and people. *Science* **362**, 304 (2018)

15. Wellstead, A., & Howlett, M. Assisted tree migration in North America: policy legacies, enhanced forest policy integration and climate change adaptation. *Scand. J. Forest Res.* **32**, 535-543 (2017)

16. Gitay, H., Suárez, A., & Watson, R. *Climate change and biodiversity* (Intergovernmental Panel on Climate Change, 2002)

17. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) *Report of the Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on the Work of its Sixth Session* (IPBES, 2018)

18. Hanewinkel, M., Cullmann, D. A., Schelhaas, M. J., Nabuurs, G. J., & Zimmermann, N. E. Climate change may cause severe loss in the economic value of European forest land. *Nat. Clim. Change* **3**, 203-207 (2013)

19. Conrad, A. O. et al. Threats to oaks in the eastern United States: Perceptions and expectations of experts. *J. Forest.* **118**, 14-27 (2020)

20. Rapp, J. M. et al. Finding the sweet spot: Shifting optimal climate for maple syrup production in North America. *Forest Ecol. Manag.* **448**, 187-197 (2019)

21. Knott, J. A., Jenkins, M. A., Oswalt, C. M., & Fei, S. Community-level responses to climate change in forests of the eastern United States. *Global Ecol. Biogeogr.* **29**, 1299-1314 (2020)

22. Loreau, M. Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* **91**, 3-17 (2000)
23. Cardinale, B. J. et al. The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572-592 (2011)

24. Liang, J. et al. Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016)

25. Pecl, G. T. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017)

26. Zhu, K., Woodall, C. W., Ghosh, S., Gelfand, A. E., & Clark, J. S. Dual impacts of climate change: forest migration and turnover through life history. *Glob. Change Biol.* **20**, 251-264 (2014)

27. Paquette, A., & Messier, C. The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecol. Biogeogr.* **20**, 170-180 (2011)

28. Nowacki, G. J., & Abrams, M. D. The demise of fire and “mesophication” of forests in the eastern United States. *BioScience* **58**, 123-138 (2008)

29. Ma, W. et al. Fundamental shifts of central hardwood forests under climate change. *Ecol. Model.* **332**, 28-41 (2016)

30. Iverson, L. R., Prasad, A. M., Peters, M. P., & Matthews, S. N. Facilitating adaptive forest management under climate change: A spatially specific synthesis of 125 species for habitat changes and assisted migration over the eastern United States. *Forests* **10**, 989 (2019)

31. Millar, C. I., Stephenson, N. L., & Stephens, S. L. Climate change and forests of the future: managing in the face of uncertainty. *Ecol. Appl.* **17**, 2145-2151 (2007)
32. Renwick, K. M., & Rocca, M. E. Temporal context affects the observed rate of climate-driven range shifts in tree species. *Global Ecol. Biogeogr*. **24**, 44-51 (2015)

33. Ochuodho, T. O., Lantz, V. A., Lloyd-Smith, P., & Benitez, P. Regional economic impacts of climate change and adaptation in Canadian forests: a CGE modeling analysis. *Forest Policy Econ.* **25**, 100-112 (2012)

34. Flannigan, M. D., Logan, K. A., Amiro, B. D., Skinner, W. R., & Stocks, B. J. Future area burned in Canada. *Climatic Change* **72**, 1-16 (2005)

35. Lemprière, T. C. et al. *The Importance of Forest Sector Adaptation to Climate Change* (Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, 2008)

36. van Wagner, C. E. *Fire behavior in northern conifer forests and shrublands in the Role of Fire in Northern Circumpolar Ecosystem* (Wiley, 1983)

37. Zhou, M. Input substitution and relative input price variability in timber markets. *Can. J. Forest Res.* **51**, 339-347 (2021)

38. Chamberlain, J. L., Emery, M. R., & Patel-Weynand T. *Assessment of Nontimber Forest Products in the United States under Changing Conditions* (U.S. Department of Agriculture, Forest Service, Southern Research Station, 2018)

39. Fleetwood, J. Social justice, food loss, and the sustainable development goals in the era of COVID-19. *Sustainability* **12**, 5027 (2020)

40. Williams, J. W., Shuman, B. N., Webb III, T., Bartlein, P. J., & Leduc, P. L. Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecol. Monogr.* **74**, 309-334 (2004)
41. Davis, M. B., & Shaw, R. G. Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673-679 (2001)

42. The Intergovernmental Panel on Climate Change (IPCC) *Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming Of 1.5°C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, In the Context Of Strengthening the Global Response to the Threat Of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (IPCC, 2018)

43. Chapin Iii, F. S. et al. Consequences of changing biodiversity. *Nature* **405**, 234-242 (2000)

44. Hisano, M., Searle, E. B., & Chen, H. Y. Biodiversity as a solution to mitigate climate change impacts on the functioning of forest ecosystems. *Biol. Rev.* **93**, 439-456 (2018)
## Table 1. Summary of forest types and biomes classified based on the present (2000–2019) and past (1970–1999) forest inventories. Only the top dominant species for each forest type are listed to save space. Forest types with “W-“ belongs to West region, “E-“ to East region, and “B-“ to Boreal region.

| Forest biome (present area km²) | Forest type | Forest type name | Time | Area (km²) | Mean centroid shift ± se (km·decade⁻¹) | Direction of shift | Key species (importance value) |
|---------------------------------|-------------|------------------|------|------------|----------------------------------------|-------------------|--------------------------------|
| **Pacific Coastal Forest**      | W-A         | Sitka spruce—western hemlock | present | 12,959 | 432.6±0.5 | W | Picea sitchensis (155), Tsuga heterophylla (21), Populus balsamifera (7) |
|                                 | W-B         | mountain hemlock—cedar—spruce | present | 49,228 | 270.7±0.2 | N | Tsuga mertensiana (123), Chamaecyparis nootkatensis (23), Tsuga heterophylla (9) |
|                                 | W-C         | western hemlock—cedar—spruce | present | 120,070 | 18.2±0.1 | S | Tsuga heterophylla (115), Pseudotsuga menziesii (29), Thuja plicata (14) |
|                                 | W-D         | red alder—Douglas-fir | present | 41,085 | 156.8±0.2 | N | Alnus rubra (72), Pseudotsuga menziesii (50), Acer macrophyllum (23) |
|                                 | W-E         | fir—hemlock | present | 26,463 |                      |                    | Abies amabilis (111), Tsuga heterophylla (27), Tsuga mertensiana (18) |
|                                 | W-F         | yellow-cedar—hemlock | present | 10,522 |                      |                    | Chamaecyparis nootkatensis (90), Tsuga heterophylla (40), Tsuga mertensiana (34) |
| **Western Cordillera**          | W-G         | pure and mixed lodgepole pine forest | present | 124,573 | 40.0±0.1 | S | Pinus contorta (152), Abies lasiocarpa (10), Pseudotsuga menziesii (9) |
|                                 | W-H         | Northern Rocky Mountains cedar—larch—hemlock | present | 33,435 | 36.6±0.1 | N | Thuja plicata (66), Pseudotsuga menziesii (36), Larix occidentalis (28) |
|                                 | W-I         | grand fir—Douglas-fir | present | 47,490 | 15.4±0.0 | N | Abies grandis (98), Pseudotsuga menziesii (40), Larix occidentalis (13) |
|                                 | W-J         | aspen-mixed conifer | present | 31,435 | 199.7±0.4 | W | Populus tremuloides (139), Abies lasiocarpa (13), Pseudotsuga menziesii (10) |
|                                 | W-K         | subalpine fir—Engelmann spruce | present | 230,700 | 19.0±0.1 | E | Abies lasiocarpa (107), Picea engelmannii (33), Pinus contorta (18) |
|                                 | W-L         | white fir—Douglas-fir—ponderosa pine | present | 27,668 | 40.6±0.2 | W | Abies concolor (111), Pseudotsuga menziesii (19), Abies magnifica (13) |
|                                 | W-M         | Engelmann spruce—subalpine fir | present | 34,246 | 75.9±0.2 | S | Picea engelmannii (109), Abies lasiocarpa (42), Pinus contorta (13) |
|                                 | W-N         | whitebark—lodgepole pine—subalpine fir | present | 12,349 |                      |                    | Pinus albicaulis (120), Abies lasiocarpa (42), Pinus contorta (19) |
|                                 | W-O         | western larch—Douglas-fir | present | 40,745 |                      |                    | Larix occidentalis (44), Pseudotsuga menziesii (43), Pinus contorta (23) |
| **Mediterranean California**    | W-P         | coastal redwood—tanoak | present | 23,266 | 6.8±0.1 | N | Lithocarpus densiflorus (77), Pseudotsuga menziesii (44), Sequoia sempervirens (35) |
|                                 | W-Q         | California mixed oak woodland | present | 36,583 | 123.0±0.1 | S | Pseudotsuga menziesii (42), Quercus chrysolepis (31), Calocedrus decurrens (25) |
| **Coastal-Interior Range**      | W-R         | Douglas-fir mixed forest | present | 34,915 | 48.0±0.1 | W | Pseudotsuga menziesii (158), Pinus ponderosa (7), Pinus contorta (5) |
|                                 | W-S         | ponderosa pine—fir | present | 30,524 | 21.2±0.3 | N | Pinus ponderosa (149), Pseudotsuga menziesii (24), Pinus contorta (4) |
| Region          | Forest Type                     | Present | Past  | Species                                                                 |
|-----------------|---------------------------------|---------|-------|-------------------------------------------------------------------------|
| Eastern Forest  | red oak mixed hardwood forest   | present | past  | Pinus ponderosa (174), Pseudotsuga menziesii (14), Abies concolor (2)   |
|                 |                                 | 56,652  | 27,961| Quercus rubra (26), Acer rubrum (23), Populus grandidentata (16)        |
|                 |                                 | 46,698  |       | Quercus rubra (32), Populus grandidentata (26), Acer rubrum (21)        |
|                 | Great Lakes tamarack—spruce     | present | past  | Larix laricina (135), Picea mariana (27), Thuja occidentalis (7)       |
|                 |                                 | 10,091  | 7,987 | Larix laricina (128), Picea mariana (27), Abies balsamea (8)           |
|                 | North-central maple forest      | present | past  | Acer saccharum (92), Acer rubrum (13), Betula alleghaniensis (11)      |
|                 |                                 | 80,059  | 49,747| Acer saccharum (107), Acer rubrum (12), Tilia americana (9)            |
|                 | beech—maple                    | present | past  | Fagus grandifolia (80), Acer saccharum (27), Acer rubrum (19)          |
|                 |                                 | 50,902  | 104,869| Fagus grandifolia (60), Acer saccharum (26), Tsuga canadensis (22)    |
|                 | Great Lakes pine forest         | present | past  | Pinus resinosa (89), Pinus banksiana (33), Quercus ellipsoidalis (10) |
|                 |                                 | 22,786  | 10,353| Pinus resinosa (119), Pinus banksiana (15), Populus tremuloides (12)  |
|                 | red maple—hardwood              | present | past  | Acer rubrum (101), Quercus rubra (7), Acer saccharum (6)               |
|                 |                                 | 46,720  | 61,350| Acer rubrum (98), Populus tremuloides (9), Quercus rubra (8), Acer saccharum (8) |
|                 | eastern hemlock—maple           | present | past  | Tsuga canadensis (79), Acer rubrum (25), Fagus grandifolia (11)       |
|                 |                                 | 49,452  | 27,402| Tsuga canadensis (69), Acer rubrum (27), Acer saccharum (13)           |
|                 | northern white-cedar—balsam fir | present | past  | Abies balsamea (33), Picea mariana (9)                                 |
|                 |                                 | 28,370  | 37,928| Abies balsamea (31), Picea mariana (10)                                |
|                 | balsam fir—maple—yellow birch   | present | past  | Acer saccharum (35), Abies balsamea (32), Betula alleghaniensis (24)  |
|                 |                                 | 229,088 | 264,717| Acer saccharum (29), Abies balsamea (31), Betula alleghaniensis (24)  |
|                 | eastern white pine—maple        | present | past  | Pinus strobus (87), Acer rubrum (24), Quercus rubra (10)               |
|                 |                                 | 9,842   | 37,905| Pinus strobus (92), Acer rubrum (23), Quercus rubra (9)                |
|                 | Great Lakes black ash—poplar    | present | past  | Fraxinus nigra (77), Populus tremuloides (15), Abies balsamea (14)     |
|                 |                                 | 9,210   | 17,210| Fraxinus nigra (62), Populus balsamifera (29), Populus tremuloides (23) |
|                 | black cherry—maple              | present | past  | Prunus serotina (78), Acer rubrum (20), Fraxinus americana (8)         |
|                 | oak—hickory                     | present | past  | Quercus alba (49), Acer rubrum (15), Quercus velutina (14), Nyssa sylvatica (10) |
|                 |                                 | 58,484  | 62,119| Quercus alba (49), Carpinus caroliniana (25), Carpinus caroliniana (25), Cornus florida (13) |
|                 | Appalachian oak—pine            | present | past  | Quercus prinus (46), Acer rubrum (26), Nyssa sylvatica (16), Quercus coccinea (11) |
|                 | mixed oak—hickory               | present | past  | Quercus prinus (45), Acer rubrum (24), Nyssa sylvatica (13), Quercus coccinea (12) |
|                 | yellow-poplar—maple             | present | past  | Quercus stellata (33), Pinus echinata (32), Quercus velutina (16), Quercus alba (12) |
|                 | eastern redcedar—white ash—American elm | present | past  | Quercus stellata (57), Carpinus caroliniana (27), Quercus velutina (20) |
|                 | slash pine mixed                | present | past  | Liriodendron tulipifera (46), Acer rubrum (21), Acer saccharum (11), Betula lenta (11) |
|                 | lobolly pine—sweetgum           | present | past  | Carpinus carpinus (18), Fraxinus americana (13), Ulmus americana (13) |
|                 | green ash-mixed floodplain forest | present | past  | Carpinus carpinus (18), Fraxinus americana (13), Ulmus americana (13) |
|                 | longleaf mixed pine             | present | past  | Pinus taeda (154), Pinus taeda (6), Taxodium ascendens (4)             |
|                 |                                 | 57,916  | 23.9±0.1| Pinus taeda (155), Pinus palustris (7), Pinus taeda (5)              |
|                 |                                 | 118,746 | 150,713| Pinus taeda (152), Liquidambar sylvatic (10), Quercus nigra (4)       |
|                 |                                 | 47,788  | 54,381| Pinus taeda (123), Liquidambar sylvatic (14), Pinus echinata (8)      |
|                 |                                 | 18,482  | 20,539| Fraxinus pennsylvanica (26), Celtis laevigata (14), Acer negundo (13)   |
|                 |                                 | 57,916  | 23.9±0.1| Fraxinus pennsylvanica (17), Liquidambar sylvatic (15), Carpinus carpinus (12) |
|                 |                                 | 118,746 | 150,713| Pinus palustris (98), Pinus clausa (18), Quercus laevis (11), Pinus taeda (11) |
|                 |                                 | 47,788  | 54,381| Pinus palustris (106), Pinus clausa (14), Pinus elliotii (11), Pinus taeda (10) |
|                 |                                 | 118,746 | 150,713| Nyssa biflora (30), Acer rubrum (15), Magnolia virginiana (15), Pinus elliotii (13) |
| E-W | sweetgum—water oak—loblolly pine | present | 131,721 | 22.2±0.1 | E | Liquidambar styraciflua (48), Quercus nigra (21), Pinus taeda (21), Acer rubrum (12) |
|-----|---------------------------------|---------|---------|---------|---|--------------------------------------------------|
| E-X | Virginia pine—maple             | present | 38,640  | 13.4±0.0 | S | Pinus virginiana (50), Acer rubrum (17), Oxydendrum arboreum (11) |
| E-Y | loblolly pine—sweetgum mixed     | present | 84,752  |         |    | Pinus taeda (86), Liquidambar styraciflua (25), Quercus nigra (10) |
| B-A | quaking aspen—balsam fir—paper birch | present | 349,949 | 683.3±0.1 | E | Populus tremuloides (105), Abies balsamea (17), Betula papyrifera (13) |
| B-B | paper birch—balsam fir           | present | 153,236 | 84.9±0.1 | E | Betula papyrifera (59), Abies balsamea (42), Picea mariana (20), Picea glauca (18) |
| B-C | jack pine—black spruce           | present | 294,291 | 72.3±0.1 | W | Pinus banksiana (118), Picea mariana (40), Populus tremuloides (10) |
| B-D | balsam fir—black spruce          | present | 200,950 | 51.6±0.0 | S | Abies balsamea (113), Picea mariana (22), Betula papyrifera (17), Picea glauca (11) |
| B-E | black spruce—balsam fir          | present | 750,121 | 313.3±0.1 | W | Picea mariana (150), Abies balsamea (20), Betula papyrifera (7) |
| B-F | white spruce—lodgepole pine      | present | 714,379 |         |    | Picea glauca (105), Pinus contorta (22), Betula neoalaskana (14), Picea mariana (12) |

Boreal Forest (2,462,924)

| B-A | quaking aspen—balsam fir—paper birch | present | 1,104,633 | 683.3±0.1 | E | Populus tremuloides (105), Abies balsamea (17), Betula papyrifera (13) |
| B-B | paper birch—balsam fir              | present | 123,495  | 84.9±0.1 | E | Betula papyrifera (59), Abies balsamea (42), Picea mariana (20), Picea glauca (18) |
| B-C | jack pine—black spruce              | present | 527,541  | 72.3±0.1 | W | Pinus banksiana (118), Picea mariana (40), Populus tremuloides (10) |
| B-D | balsam fir—black spruce             | present | 222,889  | 51.6±0.0 | S | Abies balsamea (113), Picea mariana (22), Betula papyrifera (17), Picea glauca (11) |
| B-E | black spruce—balsam fir             | present | 436,470  | 313.3±0.1 | W | Picea mariana (150), Abies balsamea (20), Betula papyrifera (7) |
| B-F | white spruce—lodgepole pine         | present | 714,379  |         |    | Picea glauca (105), Pinus contorta (22), Betula neoalaskana (14), Picea mariana (12) |
Fig. 1. Map of present (2000–2019) forest type classification, as well as the azimuth and velocity of each forest type. Forest migration was quantified based on the movement of weighted geographic centroids of each forest type. Forest type code corresponds to Table 1.
Fig. 2. Comparison of migration velocity (km·decade⁻¹) between forest types (a) and tree species (b), assessed at the 0.025° grid level. Migration velocity was measured by distance shift in kilometers per decade. Violin plots show the distribution of grid-level velocity by region and type (left: forest migration, right: tree species migration). Solid lines and surrounding bands represent the mean and 95% confidence interval, respectively (red represents forest migration velocity, and blue tree species migration velocity).
Fig. 3. Geographic distribution and main contributing factors of the ratio of forest migration velocity to tree species migration velocity. In the continental map (a), different colors represent different levels of the ratio and are consistent with background colors of the partial dependence plot (b) which shows the estimated relationship between the ratio and two top contributing variables: climate change and species evenness. Climate change is an aggregated indicator—normalized between 0 and 1—of changes in top nine bioclimate variables ranked by variable importance. Tree species evenness is the average of past (1970-1999) and present (2000-2019) surveys.
Acknowledgments: We thank the Global Forest Biodiversity Initiative (GFBI) for facilitating the international research collaboration. We thank Mo Zhou and John Dunning, Jr. for their feedback on this study.

Funding:

U.S. Department of Agriculture’s (USDA) Agricultural Marketing Service through grant AM200100XXXXG007

USDA National Institute of Food and Agriculture McIntire Stennis project 1017711

Start-up Fund provided by the Department of Forestry and Natural Resource and the College of Agriculture, Purdue University

Department of Forestry and Natural Resources, Purdue University

Takenaka Scholarship Foundation (AOA)

Author contributions:

Conceptualization: JL, AOA

Methodology: AOA, JL, RS

Investigation: AOA, JL, CWW, TO

Supervision: JL

Writing – original draft: AOA, JL

Writing – review & editing: All

Competing interests: Authors declare that they have no competing interests.
Methods

Data and data integration

For this study, we compiled and integrated *in situ* forest-tree data from independent and standard forest inventories. Data for the United States came from the Forest Inventory and Analysis (FIA)\(^4\) and the Cooperative Alaska Forest Inventory (CAFI)\(^5\). Data for Canada came from two independent sources: permanent sample plot networks\(^4\) and Canada’s National Forest Inventory ground plot network\(^5\). FIA is a nation-wide survey of the extent and status of forests\(^4\). The plots are permanent sample plots from which data were collected periodically. The FIA plots are approximately 0.1 ha in size and are placed on a hexagonal grid so that there is one plot for every 2,428 ha (6,000 acres) of forested land. In order to maintain the privacy of landowners, all plot coordinates are fuzzed under the passage of the Food Security Act of 1985. However, true coordinates are within 0.80 to 1.61 km of the fuzzed coordinates, so the impact is negligible\(^4\). CAFI provides a collection of permanent sample plots in southeast Alaska, and the plot size is 0.04 ha in a square shape\(^5\). The data from permanent sample plot networks of Canada is distributed across eight provinces – British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, Nova Scotia, and Newfoundland and Labrador. These plots are 0.04 ha in size, and their distribution over forested areas and re-measurement frequency vary slightly among provinces\(^5\). The data from Canada’s National Forest Inventory ground plot network is distributed across the forested areas in Canada, and plot size differs (125–500 m\(^2\))\(^5\).

We derived the following data integration protocol to harmonize the different forest inventory datasets described above into consistent continental data frames. From each dataset, we obtained tree-level information for all the trees with a minimum diameter at breast height (DBH) of 1 cm. We grouped these tree-level records by the year of inventory, and compiled one
data frame for 2000–2019 and another data frame for 1970–1999. For each period, we then summarized tree-level information into a plot-level species abundance matrix, which contained the percentage of the number of stems by species (i.e., relative abundance), as well as the percent basal area by species (i.e., relative dominance). Based on the species abundance matrix, we calculated the importance value of each tree species present on a sample plot, which equally weights relative abundance and relative dominance of a particular species\textsuperscript{51–53}.

The final continental data frames consisted of plot identification number and coordinates, as well as the importance values of all tree species present on each plot. The plots were uniformly distributed in the sampled areas across the continent (Extended Data Fig. 4). For the 1970-1999 data frame, because some trees in the genera of Aesculus, Amelanchier, Carya, Crataegus, Halesia, Malus, and Salix were recorded only to the genus level, we also calculated the importance values of these genera (Supplementary Table 3). Based on the continental data frames, we aggregated plot-level species importance values into a grid-based forest range map to harmonize the past and present survey data. The grid map consisted of 0.025 by 0.025-degree (approximately 3 by 3 km) grids with a minimum 10\% canopy cover, in accordance with FAO’s definition of ‘forest’\textsuperscript{54}. Based on the global forest range map\textsuperscript{55}, our study region encompassed 1,004,358 grids of forested area across North America, with a total of \(~5\) million km\textsuperscript{2}. The tropical regions of North America, \textit{i.e.} Mexico, Central America, and the Caribbean, were not included in this analysis due to a lack of remeasured \textit{in situ} data. Our study region covered 92 terrestrial ecoregions\textsuperscript{56} across the United States and Canada. These ecoregions were grouped into three distinct regions: West (39 ecoregions), East (33 ecoregions), and Boreal (20 ecoregions, Extended Data Fig. 4).

\textbf{Forest classification}
A lack of consistent classification of forest communities at a continental scale has been a major obstacle to the understanding of the patterns of forest migration. For over a century, forests have been classified based on tree species composition and structural characteristics\textsuperscript{21,51,52,57–59}. Conventional forest classification is manually defined by experts based on the similarity of forest communities in terms of species dominance\textsuperscript{57–59}. With the recent advancement of forest data availability\textsuperscript{60} and computational capacity, new data-driven forest classification schemes minimize subjective biases and exhibit greater accuracy than conventional approaches\textsuperscript{21,51,52}. To this date, however, little has been done to map forest types at a continental scale using a consistent classification scheme.

Our forest classification consists of two steps: defining forest types and mapping them. The definition of forest types was determined by the combination of autoencoder neural network and K-means cluster analysis. Autoencoder neural networks create a compressed representation of the original data, which is more suitable for K-means cluster analysis than the original data. Then, we mapped forest types determined by the K-means cluster analysis to the forested area using machine learning imputation models. Due to the random nature of K-means cluster analysis, we repeated the whole process 20 times to derive the final classification results.

For each region, based on the continental data frame and aggregated grid data described above, we first used an autoencoder neural network to calculate a latent space representation of the original input features\textsuperscript{61}. Autoencoder neural networks are unsupervised deep learning models, which use the nonlinear generalization of principal component analysis used to reduce dimensions in data\textsuperscript{61,62}. Autoencoders learn to decompose input data into alternate representations using an encoding function, $e: R^n \rightarrow R^k$, and then reconstruct an approximation of the input using a decoding function, $d: R^k \rightarrow R^n$\textsuperscript{63}, where the parameters of $e$ and $d$ are
simultaneously optimized. Both the encoding and decoding functions are comprised of one layer or more to perform operations between the input and model parameters, where \( z \) is the number of units in each layer. The result of all \( z \) operations in each layer can then be transformed using a non-linear activation function, \( \sigma(\cdot) \), to reveal characteristics of the data distribution in an alternate dimensional space\(^6\). This approach provides a more informative data distribution along with the data’s reduced dimensionality for efficient data transformations.

In this work, we used the autoencoders’ encoding function, \( e: \mathbb{R}^n \rightarrow \mathbb{R}^k \), where \( k < n \), to transform the input data into a reduced dimensional representation to conduct K-means cluster analysis. The reduced dimensional representation of the input information improves robust clustering results, and mitigates the computational complexity of the K-means algorithm \( O(n^2) \). We began by constructing a fully connected autoencoder comprised of an input layer (with \( n \) units) followed by three fully-connected layers (consisting of 150, \( 0.75 \times n \), and 150 units, respectively) and the \( n \)-dimensional output layer. The output of each fully-connected layer was given by \( \sigma(x \cdot w + b) \), where \( x, w \in \mathbb{R}^p \) and \( b \in \mathbb{R} \) denote the layer’s input, the number of units in each layer, and the threshold bias value, respectively. Note that \( p \) represents an arbitrary dimensionality of any given layer. The three hidden layers used a linear activation function and the output layer utilized a sigmoid activation function, which is given by \( \frac{1}{1+\exp(xw+b)} \). This sigmoid activation function in the output layer made the overall network non-linear while all the three hidden layers utilized a linear activation function. The encoder, \( e \), and the decoder, \( d \), were simultaneously optimized according to
\[
\min_{e,d} \left\| \frac{1}{n} \sum_{i=1}^{n} (x_i - d(e(x_i))) \right\|^2
\]
using the Adam optimizer\(^6\). After training the autoencoder, the output of the second hidden layer was used to
encode the input into its reduced dimensional representation, which was then inputted into the K-means clustering algorithm.

To avoid potential bias caused by insufficient sample sizes, we excluded the species that are present in less than 60 grids (Supplementary Table 3). Based on the reduced dimensional representation, we conducted a K-means cluster analysis to classify forests across North America. We conducted K-means cluster analysis in R (version 4.0.4) using the built-in function “kmeans”. We set the number of starts to 50 and the maximum iterations to 100. Choosing the number of dimensions from the autoencoder neural network and the number of clusters, as well as the evaluation methods are described in §Model evaluation below.

With the defined forest types (i.e., clusters) from the 20 repetitions, we manually matched the same forest type by calculating the Euclidean distance in terms of species importance value between all the combinations of forest types generated from the 20 repetitions. When 10 or more repetitions identified the given forest type, we recognized the forest type as a final forest type. Since we classified forest types for three regions separately (West, East, and Boreal), there were potential overlaps of forest types between regions. To identify and merge potential overlaps, we calculated the Euclidean distance of all combinations of the final forest types in terms of species importance value. If a Euclidean distance was less than 60, across-region forest types were merged. One exception was that western aspen-mixed conifer (W-J) and boreal quaking aspen—balsam fir—paper birch (B-A) remained separated due to the large expanse of *Populus tremuloides*.

**Mapping of forest types**

To map the distribution of forest types across the 4.9 million-km² study region, we considered two candidate imputation models to estimate the underlying forest type of unsampled
grids based on 38 predictor variables. The two candidate models were random forests and support-vector machines. Random forests are a non-parametric ensemble learning approach\textsuperscript{65}, which combines a variant of classification trees and an additional level of randomness by bootstrapping sub-data and different sets of predictor variables to mitigate the multicollinearity issues that most statistical models face\textsuperscript{66}. We used the “randomForest” package in R with the default hyperparameter setting\textsuperscript{67}. Support-vector machines are supervised learning models which construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space to help analyze data for classification and regression analysis\textsuperscript{68}. We used the “e1071” package in R with the default hyperparameter setting\textsuperscript{69}. We compared the performance of these two candidate models and selected random forests as the final imputation model (see §Model evaluation; Extended Data Fig. 5).

To train candidate models to predict forest type based on local environmental and biophysical conditions, we compiled a total of 38 predictor variables. The predictor variables we compiled consisted of 17 climate variables\textsuperscript{70–72}, 13 topographic variables\textsuperscript{73}, seven soil variables\textsuperscript{74}, and human footprint\textsuperscript{75}. These predictor variables were derived from open access satellite-based remote sensing and ground-based survey data layers, all of which have a nominal resolution of 1-km. Detailed information of the predictor variables is available in Extended Data Table 1 and Extended Data Fig. 6. We used “Hmisc” package in R to impute missing data in those predictor variables\textsuperscript{76}.

Model evaluation

To maximize the clustering performance, we calculated the silhouette width to determine the number of dimensions from the autoencoder neural network and the number of clusters. Silhouette width is an indicator of between-cluster heterogeneity\textsuperscript{77}. With a range between -1 and
1. Positive silhouette width values indicate that a given member of a cluster is closer to its own cluster’s centroid than to the nearest cluster’s centroid. Negative values indicate that a given member is closer to the nearest cluster’s centroid than to the centroid of its own cluster. Generally, higher silhouette width values indicate greater between-cluster heterogeneity. We used the silhouette width to fine-tune hyperparameters for the autoencoder (the number of dimensions) (Extended Data Fig. 7) and K-means cluster analysis (the number of clusters). We calculated silhouette width using the “cluster” package in R. The mean silhouette widths from our K-means cluster analyses were significantly greater than zero for all forest types ($p < 0.001$) in the West for the present dataset. Eighteen out of 19 forest types in the West, 22 out of 26 in the East, and all six forest types in the Boreal region were significantly greater than zero in the mean silhouette width. In summary, 90% of the forest types classified here were significantly distinct from one another in terms of species composition (Supplementary Table 2).

To assess the performance of the imputation model in mapping forest types across the continent, we conducted a rigorous 80/20 cross-validation using bootstrapping. In each iteration, we used stratified sampling to split the entire dataset into the training (80%) and testing (20%) set, and conducted the combination of under-sampling and oversampling of the training set for both random forests and support-vector machines. Stratified sampling was conducted using the “caret” package in R, and under-sampling and oversampling were conducted using the “UBL” package. Based on five random iterations with sample replacement in each of the 20 repetitions, we calculated the 95% confidence interval of classification accuracy, the Kappa statistic, and elements of confusion matrix, as well as predictor variable importance. For each candidate imputation model, the output was a matrix of class probability from five iterations. We
chose the forest type of majority vote from the five iterations, and thus, our final output was a
matrix of class probability from the 20 repetitions.

The classification accuracy, Kappa statistic, and elements of confusion matrices were
calculated based on the prediction on the testing set in each iteration. Compared with the support
vector machine model, random forests model was 10–17% more accurate in terms of overall
accuracy and 11–20% more precise in terms of the Kappa statistic (Extended Data Fig. 5).
Therefore, we selected random forests as the final imputation model. The confusion matrices
based on random forests models were based on the number of cases in class prediction,
standardized in percentage (Extended Data Figs. 8, 9). For the present dataset, the coastal
redwood—tanoak forest (W-P) had the highest classification accuracy (88%, Extended Data Fig.
8), and the red maple—hardwood forest (E-F) had the lowest accuracy (18%, Extended Data Fig.
8) among all the classes (i.e., forest types).

**Quantifying forest migration patterns**

We quantified migration patterns of forest type in terms of velocity (km⋅decade\(^{-1}\)) and
azimuth (°) of the mean geographic movement, as well as changes in area, based on *in situ* forest
inventory data aggregated into 0.025 by 0.025-degree grids. The first inventory was conducted
between 1970 and 1999 (past inventory), whereas the second inventory was conducted between
2000 and 2019 (present inventory). We ensured past and present forest types matched so that
forest migration can be quantified. To do this, we calculated the Euclidean distance of all
combinations between past and present forest types in terms of species importance value. Pairs
were considered matching when the forest type of minimum distance was the same between the
past-and-present pair. For example, if and only if present forest type X’s closest past forest type
is Y, and past forest type Y’s closest present forest type is also X, they were considered matching.

For each matching pair of past and present forest types, we determined its mean geographic movement and associated 95% confidence interval using a bootstrapping approach with 1,000 iterations. In each iteration, we randomly sampled 80% of past and present data with replacement and quantified the velocity and azimuth of forest migration, based on the past and present centroids of the geographic range of that forest type. The geographic centroid was calculated by weighting the grid geographic coordinates with percent forest type. After mapping forest types across the continent using the imputation random forest models, all the 1,004,358 grids contain a percentage for each forest type as well as the geographic coordinates (latitude and longitude) of that grid’s centroid. Percent forest type was determined by how many repetitions, out of 20 repetitions, returned the particular forest type. Geographic centroids for each forest type were then calculated by weighting the geographic coordinates and percentage in that grid with the following equations:

$$\bar{X}_j = \frac{\sum_{i=1}^{n} w_{ij} X_i}{\sum_{i=1}^{n} w_{ij}}, \bar{Y}_j = \frac{\sum_{i=1}^{n} w_{ij} Y_i}{\sum_{i=1}^{n} w_{ij}},$$

(1)

where $\bar{X}_j$ is the weighted mean longitude of forest type $j$, $\bar{Y}_j$ is the weighted mean latitude of forest type $j$, $X_i$ and $Y_i$ are the longitude and latitude for the centroid of grid $i$, and $w_{ij}$ is the grid cell level percentage of forest type $j$.

This geographic distance was calculated using the “sp” package in R$^{80}$, while the associated azimuth was determined using the “sfsmisc” package$^{81}$. The velocity of forest migration (km·decade$^{-1}$) was then calculated as the average distance of movement for each forest type (Table 1) per decade. We also determined area coverage of each forest type by weighting
grid area by percent presence of the forest type. Grid area was estimated using the “raster”
package in R$^82$.

**Comparison of forest migration and tree species migration**

To directly compare the geographic shift of forest types and tree species, we calculated
grid-level velocity for each entity. For forest type, we quantified grid-level velocity of forest
migration by weighting the forest type velocity by percent presence of the forest type in each
grid. Percent presence of the forest type here was based on how many models, out of five
models, returned the given forest type. Therefore, the output was a matrix of grid-level velocity
from the 20 repetitions.

We estimated tree species migration in a similar manner, using the same grid-level forest-
tree data for identical time periods. For each tree species and each time period, we estimated tree
species distribution range based on random forests models and the 38 predictor variables
(Extended Data Table 1). For each region (West, East, and Boreal), only species with sufficient
sample size ($\geq 60$ grids) in both time periods were included (Supplementary Table 3). Following
Iverson et al. (2019), we reported the mean predicted importance value for each species or zero
for species with zero median and a coefficient of variation no less than 2.75 among all predicted
values$^{53}$. We calculated weighted mean geographic centroids using predicted importance value,
and determined the species’ mean geographic shift using the identical method to the one stated
above. We then repeated this process 20 times to derive the mean and 95% confidence interval of
tree species migration velocity. To maximize the model performance while minimizing
computational time, we selected the number of trees = 100 after fine-tuning using the West
present dataset as an example. Specifically, we calculated root mean square error (RMSE) for
different number of trees with 10 iterations, and chose the number of trees where RMSE values converged.

**Modeling the ratio of forest migration velocity to tree species migration velocity**

Based on the grid-level velocities of forest types and tree species, we took the ratio of forest migration velocity to tree species migration velocity for each grid. We then trained a random forests regression model with the ratio being the response variables, and 18 predictor variables (Extended Data Fig. 3). Based on grid-level tree species abundance data, we calculated three biodiversity measures: species richness, Shannon’s index, and species evenness. Species richness \((S)\) represents the total number of tree species present in the grid. Shannon’s diversity index \((H)\) was calculated using the formula:

\[
H = -\sum_{i=1}^S p_i \ln p_i, \tag{2}
\]

where \(p_i\) is the proportion of importance value of species \(i\) relative to the sum of importance value of all species present in that grid. Species evenness \((E)\) was calculated using the measure proposed by Chao and Ricotta\(^8^4\):

\[
E = \frac{e^H - 1}{S - 1}. \tag{3}
\]

In addition, we calculated the temporal changes of 15 bioclimate variables \((C_1-C_{15},\) Extended Data Table 1) between the past and present surveys, and added these variables \((\Delta C_1, \Delta C_2, \ldots, \Delta C_{15})\) as predictor variables. With the total of 18 predictor variables, we conducted a bootstrapping of 100 random forests regression models, each trained with a random 80% subset of the full dataset with replacement. Variable importance was determined based on the Gini impurity, a measure that represents the probability of incorrect classification of randomly selected sample due to its distribution.
Finally, we assessed the partial dependence of the ratio on the predictors with the highest variable importance values, by plotting the predicted ratio values over the range of the predictor variable, holding other variables constant at their sample mean. For all temporal change-related bioclimate variables, we developed climate change as an aggregated indicator, which was calculated as the mean of the top nine most important bioclimate variables normalized to a common range between 0 and 1. We chose to calculate the climate change indicator based on the top nine most important bioclimate variables, because this subset of variables accounted for 68.4% of the total importance values of all the 21 bioclimate variables studied here.

Data and Code availability

All data, code, and materials used in the analysis will be deposited to Figshare and Purdue University Research Repository (PURR) upon the publishing of this paper.

Additional References

46. Burrill, E. A. et al. The Forest Inventory and Analysis Database: Database Description and User Guide Version 8.0 for Phase 2 (U.S. Department of Agriculture, Forest Service, 2018)

47. Malone, T. Liang, J. & Packee, E. C. Cooperative Alaska Forest Inventory (Boreal Ecology Cooperative Research Unit, Pacific Northwest Research Station, U.S. Department of Agriculture, Forest Service, 2009)
48. Chen, H. Y., Luo, Y., Reich, P. B., Searle, E. B., & Biswas, S. R. Climate change-associated trends in net biomass change are age dependent in western boreal forests of Canada. *Ecol. Lett.* **19**, 1150-1158 (2016)

49. Paquette, A., & Messier, C. The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecol. Biogeogr.* **20**, 170-180 (2011)

50. Zhang, Y., Chen, H. Y., & Taylor, A. R. Positive species diversity and above-ground biomass relationships are ubiquitous across forest strata despite interference from overstory trees. *Funct. Ecol.* **31**, 419-426 (2017)

51. Dyer, J. M. Revisiting the deciduous forests of eastern North America. *BioScience* **56**, 341-352 (2006)

52. Costanza, J. K., Coulston, J. W., & Wear, D. N. An empirical, hierarchical typology of tree species assemblages for assessing forest dynamics under global change scenarios. *PloS One* **12**, e0184062 (2017)

53. Iverson, L. R., Peters, M. P., Prasad, A. M., & Matthews, S. N. Analysis of climate change impacts on tree species of the eastern US: Results of DISTRIB-II modeling. *Forests* **10**, 302 (2019)

54. Food and Agriculture Organization of the United Nations (FAO) *Global forest resources assessment 2015 – How are the world’s forests changing?* (FAO, 2015)

55. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850-853 (2013)
56. Olson, D. M. et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51**, 933-938 (2001)

57. Eyre, F. H. *Forest cover types of the United States and Canada* (Society of American Foresters, 1980)

58. Küchler, A. W. *Potential natural vegetation of the conterminous United States* (American Geographical Society, 1964)

59. Rowe, J. S. *Forest regions of Canada* (Department of the Environment, Canadian Forestry Service, 1972)

60. Liang, J., & Gamarra, J. G. The importance of sharing global forest data in a world of crises. *Sci. Data* **7**, 1-5 (2020)

61. Song, C., Huang, Y., Liu, F., Wang, Z., & Wang, L. Deep auto-encoder based clustering. *Intell. Data Anal.* **18**, S65-S76 (2014)

62. Hinton, G. E., & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507 (2006)

63. Goodfellow, I., Bengio, Y., Courville, A. *Deep learning* (MIT Press, 2016)

64. R Core Team *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, 2021)

65. Breiman, L. Random forests. *Mach. Learn.* **45**, 5-32 (2001)

66. James, G., Witten, D., Hastie, T., & Tibshirani, R. *An introduction to statistical learning with applications in R* (Springer, 2013)
67. Liaw, A., & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18-22 (2002)

68. Cortes, C., & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995)

69. Meyer, D. et al. Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), vol. R package version 1.7-4. (2020)

70. Fick, S. E., & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302-4315 (2017)

71. Karger, D. N. et al. Climatologies at high resolution for the earth’s land surface areas. *Sci. Data* **4**, 1-20 (2017)

72. Trabucco, A., & Zomer, R. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. figshare. Fileset https://doi.org/10.6084/m9.figshare.7504448.v3 (2019)

73. Amatulli, G. et al. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci. Data* **5**, 1-15 (2018)

74. Batjes, N. World soil property estimates for broad-scale modelling (WISE30sec). (ISRIC-World Soil Information, 2015).

75. Venter, O. et al. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 1-11 (2016)

76. Harrell Jr., F. E. Harrell miscellaneous. R package version 4.4-2. (2020)

77. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. Cluster: Cluster Analysis Basics and Extensions. vol. R package version 2.1.0. (2019)
78. Kuhn, M. et al. Classification and regression training. R package version 6.0-86. (2020)

79. Branco, P., Ribeiro, R., & Torgo, L. An implementation of re-sampling approaches to utility-based learning for both classification and regression tasks. R package version 0.0.6. (2017)

80. Pebesma, E. J., & Bivand, R. S. Classes and methods for spatial data in R. *R News*, 5, 9-13 (2005)

81. Maechler, M. et al. Utilities from “Seminar fuer Statistik” ETH Zurich. R package version 1.1-8. (2021)

82. Hijjmans, R. J. et al. Geographic data analysis and modeling. R package version 3.4-5. (2020)

83. Shannon, C. E. A mathematical theory of communication. *Bell Syst. tech.* 27, 379-423 (1948)

84. Chao, A., & Ricotta, C. Quantifying evenness and linking it to diversity, beta diversity, and similarity. *Ecology* 100, e02852 (2019)

85. Iverson, L. R., Prasad, A. M., Matthews, S. N., & Peters, M. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecol. Manag.* 254, 390-406 (2008)

86. Zolkos, S. G. et al. Projected tree species redistribution under climate change: implications for ecosystem vulnerability across protected areas in the Eastern United States. *Ecosystems* 18, 202-220 (2015)

87. Hernández, L., Cañellas, I., Alberdi, I., Torres, I., & Montes, F. Assessing changes in species distribution from sequential large-scale forest inventories. *Ann. For. Sci.* 71, 161-171 (2014)
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ForestMigrationSITable1final.xlsx
- ForestMigrationSITable2final.xlsx
- ForestMigrationSITable3final.xlsx
- ForestMigrationExtendedData.docx