https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Mapping the global forest diameter spectrum using machine learning approach

Ankita Mitra[1], Nancy Harris[2], Javier G. P. Gamarra[3], Meredith L. Bastian[4,5], Thomas W. Crowther[6,7], Nicolas Picard[8], Martin Herold[9], Cang Hui[10,11,12], Sergio de-Miguel[13,14], W. Gabe Powell[15], Nathan R. Beane[15], John R. Shallock[15], Fons van der Plas[16], Michael J. Lawes[17,18], Mo Zhou[1], Mohammad Jahanshahi[19], Nadezhda M. Tchebakova[20], Elena I. Parfenova[20], Peter Reich[21,22], Cesar Ivan Alvarez[23], Daniel M. Griffith[24], Oliver L. Phillips[25], Ariane Mirabel[26,27,28], Marieke Sandker[3], WookJin Choi[1], Akane O. Abbasi[1], Giorgio Alberti[29,30], Jan Altman[31,32], Luciana F. Alves[33], Bienvenu H. K. Amani[34], Christian Amani[35], Christian Ammer[36], Valerio Avitabile[37], Gerardo A. Aymard C.[38,39], Akomian F. Azihou[40], Timothy R. Baker[25], Radomir Balazy[41], Jorcely G. Barroso[42], Jean-François Bastin[43], Marijn Bauters[44], Mahmoud Bayat[45], Hans Beeckman[46], Eduardo van den Berg[47], Jahangeer A. Bhat[48], Anupam Bhatt[49], Angoboy Ilondea Bhely[50], Luca Birigazzi[51], Philippe Birnbaum[52,53], Robert Bitariho[54], Pascal Boeckx[55], Jan Bogaert[43], Alexander Bondarev[56], Frans Bongers[57], Elise Bouchard[58], Nils Bourland[46,59], Pedro H. S. Brancalion[60,61,62], Francis Q. Brearley[63], Jennifer L. Bufford[64], José Javier Corral-Rivas[65], Jaime Briseno-Reyes[65], Eben North Broadbent[66], David F. R. P. Burslem[67], Plinio Barbosa de Camargo[68], Goran D. Češljar[69], Ravi K. Chaturvedi[70], Han Y. H. Chen[71,21], Jiaxin Chen[72], Francesco Chianucci[73], George Chuyong[74], Emil Cienciala[75,76], Zorayda Restrepo Correa[77], Antonio Lola da Costa[78], Anibal Cuchietti[3], Javid Ahmad Dar[79,80], Selvadurai Dayanandan[81], Mathieu Decuyper[82], Tran Van Do[83], Jiri Dolezal[31,84], Ilija Đorđević[85], Donald R. Drake[86], Giselda Durigan[87,88], Guy-Michel A. Elogne[89], Tsutomu Enoki[90], Javier Silva Espejo[91,92], Tom Maurice Fayle[93,94], Rinaldo Luiz Caraciolo Ferreira[95], Paul F. Foster[96,97], Lorenzo Frizzera[98], Orou G. Gaoue[99,100,101], Roberto Cazzolla Gatti[102], Damiano Gianelle[98], Mario Guevara[103], Jesús Aguirre Gutiérrez[104,65,105,106], Richard Habonayo[107], David J. Harris[108], Billy C. H. Hau[109], Bruno Hérault[110,111], Wannes Hubau[112,46], Nobuo Imai[113], Andrzej M. Jagodzinski[114,115], Štěpán Janeček[116], Patrick A. Jansen[117,118,119], Bogdan Jaroszewicz[120], Vivian Kvist Johannsen[121], Carlos A. Joly[122], Daniel J. Johnson[66], Tommaso Jucker[123], Ben Hur Marimon Junior[124], Raimundo Cosme de Oliveira Junior[125], Goytom A. Kahsay[121], Victor Karminov[126], Elizabeth Kearsley[44], Deborah K. Kennard[127], Sebastian Kepfer-Rojas[128], Gunnar Keppel[129], Mohammed Latif Khan[130], John N. Kigomo[131], Hyun Seok Kim[132,133,134], Sukyung

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Kim[135,136], Carine Klauberg[66], Henn Korjus[137], Subashree Kothandaraman[79,80], Florian Kraxner[138], Leonid Krivobokov[20], Amit Kumar[139], Gajendra Kumar[140], Anna E. Kvashnina[141], Diana Laarmann[142], Felix Laurent[46,112], Gaia Vaglio Laurin[143], Rodrigo Vieira Leite[144,145], Bas Lerink[82], Simon L. Lewis[25,146], Moses B. Libalah[147,148], Weiguo Liu[149,150], Jon C. Lovett[25], Huicui Lu[151], Maria de los Angeles Soriano-Luna[152], Anne Mette Lykke[153], Emma Jane Mackintosh[154], Jean-Remy Makana[50], Andrew R. Marshall[154,155,156], Emanuel H. Martin[157], Radim Matula[32], Musingo Mbuvi[131], Sandra Liliana Medina[152], Vincent P. Medjibe[158], Desmond Mensah[1], Sylvanus Mensah[159,160], Thiago Metzker[161], Stanisław Miścicki[162], Liudmila Mukhortova[163], Sharif A. Mukul[164,165,166], Agustinus Murdjoko[167], Gert-Jan Nabuurs[82], Anny Estelle N'GUESSAN[168], Michael R. Ngugi[169], Alain Senghor K. Ngute[154,170], Thomas Nord-Larsen[128], Thomas O. Ochuodho[171], Friday N. Ogana[172], Luzmila Arroyo Padilla[173], Nazir A. Pala[174], Alain Paquette[58], Pablo L. Peri[175,176], Sebastian Pfautsch[177], Maria Teresa F. Piedade[178], Daniel Piotto[179], John James Pipoly III[180,181], Nigel C. A. Pitman[182], John R. Poulsen[183,184,185], A. S. Raghubanshi[186], Dipika Rana[187,188], Mirco Rodeghiero[98], Samir Rolim[179], Purabi Saikia[189], Peter Schall[36], Dmitry Schepaschenko[138,56], Jochen Schöngart[190], Vladimír Šebeň[191], Giacomo Sellan[27], Carlos Alberto Silva[66], Marcos Silveira[192], J. S. Singh[193], Ferry Slik[194], Jarosław Socha[195], José Daniel Soto[173], Alexandre F. Souza[196], Prashant K. Srivastava[186], Hans ter Steege[197,198], Krzysztof Stereńczak[41], Somaiah Sundarapandian[199], Martin Svátek[200], Jens-Christian Svenning[153], Xiaolu Tang[201,202], Raquel S. Thomas[203], Antonio Tomao[29], Liam A. Trethowan[204], Robert Tropek[116,205], Kalidas Upadhyaya[206], Renato Valencia[207], Rodolfo Vásquez[208], Daniel Jose Vega-Nieva[65], Hans Verbeeck[44], Helder F. Viana[209,210], Alexander Christian Vibrans[211], Simone A. Vieira[212], Barbara Vinceti[213], Jason Vleminckx[214], Vincent Antoine Vos[215], Anna S. Vozmishcheva[56], Foma Vozmitel[216,217], Hayden Wagia[218], Hua-Feng Wang[219,220], Eric K. Wasingya[221], Edward L. Webb[222,223], Chemuku Wekesa[224], Verginia Wortel[225], Emmanuel Kasongo Yakusu[226,112,46], Zhen Yu[227,228], Angelica M. Almeyda Zambrano[66], Irié Casimir Zo-Bi[229], Jingjing Liang[1,3]

Corresponding author: Correspondence to Jingjing Liang (albeca.liang@gmail.com).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[1] Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA.
[2] Global Forest Watch, World Resources Institute, Washington, DC, USA.
[3] Forestry Division, Food and Agriculture Organization of the United Nations, Rome, Italy.
[4] Proceedings of the National Academy of Sciences, Washington, DC, USA.
[5] Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA.
[6] King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.
[7] BRANCH Institute, Zug, Switzerland.
[8] GIP Ecofor, Paris, France.
[9] GFZ Helmholtz Centre for Geosciences, Telegrafenberg, Potsdam, Brandenburg 14473, Germany.
[10] Department of Mathematical Sciences, Stellenbosch, South Africa.
[11] African Institute for Mathematical Sciences, Cape Town, South Africa.
[12] National Institute for Theoretical and Computational Sciences, Stellenbosch, South Africa.
[13] Department of Agricultural and Forest Sciences and Engineering, University of Lleida, 25198 Lleida, Spain.
[14] Forest Science and Technology Centre of Catalonia (CTFC), 25280 Solsona, Spain.
[15] U.S. Army Engineer Research and Development Center (ERDC), Environmental Laboratory, Forest Ecosystem Dynamics Team, USA.
[16] Plant Ecology and Nature Conservation, Wageningen University & Research, Wageningen, The Netherlands.
[17] School of Life Sciences, University of KwaZulu-Natal, Scottsville 3209, South Africa.
[18] Institute of Biodiversity and Environmental Conservation (IBEC), Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.
[19] Department of Civil and Construction Engineering, Purdue University, West Lafayette, IN, USA.
[20] Sukachev Institute of Forest, Siberian Branch, Russian Academy of Sciences, Federal Research Center “Krasnoyarsk Science Center”, Krasnoyarsk, Russia.
[21] Institute for Global Change Biology, and School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA.
[22] Department of Forest Resources, University of Minnesota, St. Paul, MN, USA.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[23] Centre for Climate Resilience, University of Augsburg, Universitätsstrasse 12a, 86159 Augsburg, Germany.
[24] Laboratorio de Ecología Tropical y Servicios Ecosistémicos (EcoSs Lab), Departamento de Ciencias Biológicas y Agropecuarias, Universidad Técnica Particular de Loja, Loja, Ecuador.
[25] School of Geography, University of Leeds, Leeds, UK.
[26] Cirad, UPR Forêts et Sociétés, University of Montpellier, Montpellier, France.
[27] Cirad, UMR EcoFoG (AgroParistech, CNRS, INRAE, Université des Antilles, Université de la Guyane), Campus Agronomique, Kourou, French Guiana.
[28] University of Brasilia, Department of Forestry, Brasília, Brazil.
[29] Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, Italy.
[30] National Biodiversity Future Centre (NBFC), University of Palermo, Palermo, Italy.
[31] Institute of Botany of the Czech Academy of Sciences, Dukelská 135, 379 01 Třeboň, Czech Republic.
[32] Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague 6–Suchdol, Czech Republic.
[33] Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA.
[34] Université Nangui Abrogoua, Unité de Formation et de Recherche en Sciences de la Nature, 02 BP 801 Abidjan 02, Laboratoire d'Écologie et du Développement Durable (LEDD), Abidjan, Côte d'Ivoire.
[35] Université Officielle de Bukavu, Bukavu, Democratic Republic of Congo.
[36] Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Göttingen, Germany.
[37] Independent researcher, Ispra, Italy.
[38] UNELLEZ-Guanare, Programa de Ciencias del Agro y el Mar, Herbario Universitario (PORT), Mesa de Cavacas, estado Portuguesa 3350, Venezuela.
[39] Jardín Botánico de Bogotá José Celestino Mutis, Cl. 63 #68-95, Bogotá DC, Colombia.
[40] Laboratory of Applied Ecology, Faculty of Agricultural Sciences, University of Abomey-Calavi, 01 BP 526, Cotonou, Benin.
[41] Forest Research Institute, Braci Leśnej 3, Sękocin Stary, 05-090, Poland.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[42] Universidade Federal do Acre, Campus de Cruzeiro do Sul, Acre, Brazil.
[43] TERRA Teaching and Research Centre, Gembloux Agro Bio-Tech, Université de Liège, Gembloux, Belgium.
[44] Q-ForestLab, Department of Environment, Ghent University, Ghent, Belgium.
[45] Research Institute of Forests and Rangelands, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran.
[46] Service of Wood Biology, Royal Museum for Central Africa, Tervuren, Belgium.
[47] Universidade Federal de Lavras, Departamento de Ecologia e Conservação, Lavras, Brazil.
[48] Piramal School of Educational Leadership, India.
[49] CSIR–Institute of Himalayan Bioresource Technology, Palampur (H.P.) 176061, India.
[50] Institut National pour l'Etude et la Recherche Agronomiques, Kinshasa, Democratic Republic of Congo.
[51] Forestry expert, Via Unione Sovietica 105, Grosseto, Italy.
[52] CIRAD, UMR AMAP, F-34398 Montpellier, France.
[53] AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France.
[54] Institute of Tropical Forest Conservation, Mbarara University of Science and Technology, Mbarara, Uganda.
[55] Isotope Bioscience Laboratory (ISOFYS), Department of Green Chemistry and Technology, Ghent University, Gent, Belgium.
[56] Siberian Federal University, Krasnoyarsk, Russia.
[57] Forest Ecology and Forest Management Group, Wageningen University & Research, Wageningen, The Netherlands.
[58] Centre for forest research, Département des sciences biologiques, Université du Québec à Montréal, Montreal, Canada.
[59] Climate Risk Assessment Center, Brussels, Belgium.
[60] Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, Brazil.
[61] Center for Carbon Research in Tropical Agriculture, University of São Paulo, Piracicaba, Brazil.
[62] Re.green, Rio de Janeiro, Brazil.
[63] Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

[64] Bioeconomy Science Institute, Lincoln, New Zealand.
[65] Facultad de Ciencias Forestales y Ambientales, Universidad Juárez del Estado de Durango, México.
[66] School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32611, USA.
[67] School of Biological Sciences and Interdisciplinary Institute, University of Aberdeen, Aberdeen, UK.
[68] Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, SP, Brazil.
[69] Institute of Forestry, Belgrade, Serbia (Department of Spatial regulation, GIS and Forest Policy).
[70] Center for Integrative Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests & Asian Elephant, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan 666303, P.R. China.
[71] Faculty of Natural Resources Management, Lakehead University, 955 Oliver Road, Thunder Bay, Ontario P7B 5E1, Canada.
[72] Ontario Forest Research Institute, Ministry of Natural Resources, Sault Ste. Marie, Ontario, Canada.
[73] CREA – Research Centre for Forestry and Wood, Arezzo, Italy.
[74] Department of Botany and Plant Physiology, University of Buea, Buea, Cameroon.
[75] IFER – Institute of Forest Ecosystem Research, Jílové u Prahy, Czech Republic.
[76] Global Change Research Institute CAS, Brno, Czech Republic.
[77] Grupo de Servicios Ecosistémicos y Cambio Climático, Corporación COL-TREE, Medellín, Colombia.
[78] Universidade Federal do Pará, Instituto de Geociências, Rua Augusto Correa, 01, CEP 66075-110, Belém, Pará, Brasil.
[79] Terrestrial Ecology and Modelling (TEaM) Lab, Department of Environmental Science and Engineering, SRM University-AP, Amaravati, India.
[80] Centre for Geospatial Technology, SRM University-AP, Amaravati, India.
[81] Biology Department, Concordia University, Montreal, Quebec, Canada.
[82] Wageningen Environmental Research, Wageningen University & Research, Wageningen, Netherlands.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[83] Silviculture Research Institute, Vietnamese Academy of Forest Sciences, Hanoi, Vietnam.
[84] Faculty of Science, University of South Bohemia, Branišovská 1760, CZ-370 05 České Budějovice, Czech Republic.
[85] Institute of Forestry, Belgrade, Serbia.
[86] Harold L. Lyon Arboretum, University of Hawai‘i at Mānoa, Honolulu, HI, USA.
[87] Instituto de Pesquisas Ambientais, São Paulo, Brazil.
[88] Universidade Estadual de Campinas, Campinas, Brazil.
[89] Institut National Polytechnique Félix Houphouët-Boigny (INPHB), Yamoussoukro, Côte d’Ivoire.
[90] Ashoro Research Forest, Kyushu University, 1-85 Kita 5, Ashoro, Hokkaido 089-3705, Japan.
[91] Departamento de Biologia, Universidad de La Serena, Chile.
[92] Instituto de Ecologia y Biodiversidad, La Serena, Chile.
[93] School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK.
[94] Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic.
[95] Federal Rural University of Pernambuco, Department of Forest Science, Avenida Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife, PE, Brazil.
[96] Bijagual Ecological Reserve, Sarapiquí, Costa Rica.
[97] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.
[98] Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38098 San Michele all’Adige (TN), Italy.
[99] Department of Ecology and Evolutionary Biology, University of Tennessee Knoxville, Knoxville, TN 37996, USA.
[100] Laboratoire d’Etude sur la Dynamique des Populations (PopDYN), Faculté d’Agronomie, Université de Parakou, Parakou, Benin.
[101] African Centre for DNA Barcoding, Department of Botany and Plant Biotechnology, University of Johannesburg, APK Campus, Johannesburg, South Africa.
[102] Department of Biological, Geological, and Environmental Sciences (BiGeA), Alma Mater Studiorum, University of Bologna, Bologna, Italy.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[103] Instituto de Geociencias, Universidad Nacional Autónoma de México, Campus Juriquilla, Qro., Mexico.
[104] Georgina Mace Centre for the Living Planet, Imperial College London, Ascot, UK.
[105] Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK.
[106] Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, UK.
[107] Université du Burundi, Faculté d'Agronomie et de Bio-Ingénierie, Centre de Recherche en Sciences des Productions Animales, Végétales et Environnementales (CRAVE), B.P. 2940 Bujumbura, Burundi.
[108] Royal Botanic Garden Edinburgh, Edinburgh, UK.
[109] School of Biological Sciences, The University of Hong Kong, Hong Kong, China.
[110] CIRAD, Forêts et Sociétés, F-34398 Montpellier, France.
[111] Forêts et Sociétés, Univ Montpellier, CIRAD, Montpellier, France.
[112] Ghent University, Faculty of Bioscience Engineering, Laboratory of Wood Technology, Ghent, Belgium.
[113] Tokyo University of Agriculture, Tokyo, Japan.
[114] Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland.
[115] Poznań University of Life Sciences, Department of Game Management and Forest Protection, Wojska Polskiego 71D, 60-625 Poznań, Poland.
[116] Department of Ecology, Faculty of Science, Charles University, Prague, Czechia.
[117] Smithsonian Tropical Research Institute, Balboa, Ancón, Republic of Panama.
[118] Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands.
[119] Department of Biology, Utrecht University, Utrecht, The Netherlands.
[120] University of Warsaw, Faculty of Biology, Białowieża Geobotanical Station, Białowieża, Poland.
[121] University of Copenhagen, Department of Geosciences and Natural Resource Management, Copenhagen, Denmark.
[122] Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil, 13083-862, Brazil.
[123] School of Biological Sciences, University of Bristol, Bristol, UK.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[124] State University of Mato Grosso (UNEMAT), Plant Ecology Laboratory (LABEV), Graduate Program in Ecology and Conservation, Nova Xavantina, Mato Grosso, Brazil.
[125] Embrapa Amazônia Oriental, Santarém, Pará, Brazil.
[126] Bauman Moscow State Technical University (BMSTU), Moscow, Russia.
[127] Colorado Mesa University, Grand Junction, USA.
[128] Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark.
[129] School of Biological Sciences and Environment Institute, The University of Adelaide, Adelaide, SA 5005, Australia.
[130] Department of Botany, Dr. Harisingh Gour Vishwavidyalaya (A Central University), Sagar 470003, Madhya Pradesh, India.
[131] Kenya Forestry Research Institute, P.O. Box 20412-00200, Nairobi, Kenya.
[132] Department of Agriculture, Forestry and Bioresources, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea.
[133] Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea.
[134] Interdisciplinary Program in Agricultural and Forest Meteorology, Seoul National University, Seoul 08826, Republic of Korea.
[135] Department of Agriculture, Forestry and Bioresources, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea.
[136] Agriculture and Life Sciences Research Institute, Kangwon National University, Chuncheon, Republic of Korea.
[137] Estonian University of Life Sciences, Institute of Forestry and Engineering, Tartu, Estonia.
[138] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.
[139] Department of Geography, Institute of Science, Banaras Hindu University, Varanasi 221005, Uttar Pradesh, India.
[140] Wildlife Institute of India, Dehradun, India.
[141] Denezhkin Kamen Russian Federal Nature Preserve (Zapovednik), Russia.
[142] Estonian University of Life Sciences, Estonia.
[143] National Research Council (CNR), Research Institute on Terrestrial Ecosystem (IRET), Montelibretti Research Area, Italy.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

[144] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.
[145] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.
[146] Department of Geography, University College London, London, UK.
[147] Department of Plant Biology, Faculty of Science, University of Yaoundé I, Yaoundé, Cameroon.
[148] Plant Systematics and Ecology Laboratory, Higher Teachers' Training College, University of Yaoundé, Yaoundé, Cameroon.
[149] College of Forestry, Northwest Agriculture & Forestry University, Yangling, China.
[150] Qinling National Forest Ecosystem Research Station, Yangling, Shaanxi 712100, China.
[151] Faculty of Forestry, Qingdao Agricultural University, Qingdao 266109, China.
[152] Comisión Nacional Forestal (CONAFOR), Zapopan, Jalisco, México.
[153] Department of Ecoscience, Aarhus University, Denmark.
[154] Forest Research Institute, University of the Sunshine Coast, Queensland, Australia.
[155] Reforest Africa, Mang'ula, Tanzania.
[156] Flamingo Land Ltd., Kirby Misperton, UK.
[157] College of African Wildlife Management, Mweka, Moshi, Tanzania.
[158] The Nature Conservancy (TNC-Gabon).
[159] Chair of Forest Growth and Dendroecology, Albert-Ludwigs-Universität Freiburg, Germany.
[160] Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Bénin.
[161] IBAM - Instituto Bem Ambiental, Belo Horizonte, Brazil. Grupo Myr - ESG Solutions.
[162] Warsaw University of Life Sciences, Warsaw, Poland.
[163] Sukachev Institute of Forest, Siberian Branch of the Russian Academy of Sciences, Federal Research Center "Krasnoyarsk Science Center SB RAS", Krasnoyarsk, Russia.
[164] Department of Environment and Development Studies, United International University, Dhaka 1212, Bangladesh.
[165] Tropical Forests and People Research Centre, University of the Sunshine Coast, Maroochydore DC, Queensland 4556, Australia.
[166] Department of Earth and Environment, Florida International University, Miami, FL, USA.
[167] Fakultas Kehutanan, Universitas Papua, Jalan Gunung Salju Amban, Manokwari, Papua

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Barat, Indonesia.
[168] UFR Biosciences, Félix Houphouët-Boigny University, Abidjan, Côte d'Ivoire.
[169] Queensland Herbarium and Biodiversity Science, Department of the Environment, Tourism, Science and Innovation (DETSI), Brisbane Botanic Gardens, Mt Coot-tha Road, Toowong, Qld 4066, Australia.
[170] Australian Wildlife Conservancy, Perth, Western Australia, Australia.
[171] Department of Forestry and Natural Resources, University of Kentucky, Lexington, KY, USA.
[172] Virginia Polytechnic Institute and State University, Department of Forest Resources and Environmental Conservation, 310 West Campus Dr., Blacksburg, VA 24061, USA.
[173] Museo de Historia Natural Noel Kempff Mercado; Universidad Autónoma Gabriel René Moreno, Bolivia.
[174] Faculty of Forestry, SKUAST-Kashmir, Jammu & Kashmir, India.
[175] Instituto Nacional de Tecnología Agropecuaria (INTA), Río Gallegos, Santa Cruz, Argentina.
[176] Universidad Nacional de la Patagonia Austral (UNPA)-CONICET, Río Gallegos, Santa Cruz, Argentina.
[177] Urban Transformations Research Centre, Western Sydney University, 6 Hassall Street, Parramatta, NSW 2150, Australia.
[178] National Institute for Amazon Research (INPA), Coordination of Environmental Dynamics (CODAM), Ecology, Monitoring, and Sustainable Use of Wetlands Group (MAUA), Manaus, Brazil.
[179] Centro de Formação em Ciências Agroflorestais, Universidade Federal do Sul da Bahia, BR 415, km 29, Ilhéus-BA, 45613-204, Brazil.
[180] Florida Atlantic University, Boca Raton, Florida, USA.
[181] Broward County Parks, Oakland Park, Florida, USA.
[182] Field Museum of Natural History, Chicago, IL, USA.
[183] The Nature Conservancy, Boulder, USA.
[184] Nicholas School of the Environment, Duke University, Durham, USA.
[185] Ecology and Evolutionary Biology, University of Colorado, Boulder, USA.
[186] Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi 221005, India.
[187] School of Biological & Environmental Sciences, Shoolini University, Solan, H.P., India.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[188] Department of Botany, Sri Sai University, Palampur, Kangra, Himachal Pradesh 176081, India.
[189] Department of Botany, Institute of Science, Banaras Hindu University, Varanasi 221005, India.
[190] National Institute for Amazonian Research (INPA), Manaus, Brazil.
[191] National Forest Centre, Zvolen, Slovakia.
[192] Universidade Federal do Acre, Centro de Ciências Biológicas e da Natureza, Rio Branco, Acre, Brasil.
[193] Ecosystems Analysis Laboratory, Department of Botany, Banaras Hindu University, Varanasi 221005, Uttar Pradesh, India.
[194] Environmental and Life Sciences, Faculty of Science, Universiti Brunei Darussalam, BE1410, Gadong, Brunei Darussalam.
[195] University of Agriculture in Krakow, Faculty of Forestry, Krakow, Poland.
[196] Departamento de Ecologia, Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN, Brazil.
[197] Naturalis Biodiversity Center, Leiden, The Netherlands.
[198] Quantitative Biodiversity Dynamics, Faculty of Science, Utrecht University, The Netherlands.
[199] Department of Ecology and Environmental Sciences, Pondicherry University, Puducherry, India.
[200] Department of Forest Botany, Dendrology and Geobiocoenology, Faculty of Forestry and Wood Technology, Mendel University in Brno, Brno, Czech Republic.
[201] State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China.
[202] Tianfu Yongxing Laboratory, Chengdu 610213, China.
[203] Iwokrama International Centre for Rain Forest Conservation and Development, Guyana, South America.
[204] Herbarium, Royal Botanic Gardens, Kew, London, UK.
[205] Institute of Entomology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czechia.
[206] Department of Forestry, Mizoram University, Aizawl, Mizoram, India.
[207] Pontificia Universidad Católica del Ecuador, Quito, Ecuador.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[208] Missouri Botanical Garden—Peru; Herbario Selva Central Oxapampa-HOXA, Peru.
[209] CERNAS-IPV Research Centre, Polytechnic Institute of Viseu, Campus Politécnico, Repeses, 3504-510 Viseu, Portugal.
[210] Centre for the Research and Technology of Agroenvironmental and Biological Sciences (CITAB), Inov4Agro, Universidade de Trás-os-Montes e Alto Douro (UTAD), Quinta de Prados, 5000-801 Vila Real, Portugal.
[211] Department of Forest Engineering, Universidade Regional de Blumenau, Santa Catarina, Brazil.
[212] Núcleo de Estudos e Pesquisas Ambientais, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 13083-970, Brazil.
[213] Bioversity International, Rome, Italy.
[214] Plant Ecology and Biogeochemistry lab, Faculty of Sciences, Université Libre de Bruxelles, 50 av. F.D Roosevelt, 1050 Brussels, Belgium.
[215] Instituto de Investigaciones Forestales de la Amazonía, Universidad Autónoma del Beni José Ballivián, Riberalta, Bolivia.
[216] Forest Research Institute of the Karelian Research Centre of the Russian Academy of Sciences (FRI KarRC RAS), Petrozavodsk, Russia.
[217] State Nature Reserve «Denezhkin Kamen», Russia.
[218] Papua New Guinea University of Technology, School of Forestry, Faculty of Natural Resources, Lae, Papua New Guinea.
[219] Collaborative Innovation Center of Nanfan and High-Efficiency Tropical Agriculture, School of Breeding and Multiplication, Hainan University, Sanya 572025, China.
[220] School of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, China.
[221] Université de Kisangani, Fac Gest Ressources Nat Renouvelables, Kisangani, Democratic Republic of Congo; Institut National pour l'Etude et la Recherche Agronomiques, Centre de Recherche de Yangambi, Yangambi, Democratic Republic of Congo.
[222] Viikki Tropical Resources Institute, Department of Forest Sciences, University of Helsinki, Finland.
[223] Helsinki Institute for Sustainability, University of Helsinki, Finland.
[224] Kenya Forestry Research Institute, Coast Eco-Region Research Programme, P.O. Box 1078-80200, Malindi, Kenya.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

[225] Department of Forest Management, Centre for Agricultural Research in Suriname (CELOS), Paramaribo, Suriname.
[226] Faculté de Gestion des Ressources Naturelles Renouvelables, Université de Kisangani, Kisangani, Democratic Republic of the Congo.
[227] Key Laboratory of Forest Ecology and Environment, China's National Forestry and Grassland Administration, Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing, China.
[228] State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, China.
[229] ESA & UMRI SAPT, Institut National Polytechnique Félix Houphouët-Boigny, BP 1093, Yamoussoukro, Côte d'Ivoire.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Abstract

Global forest assessments assist climate policy development, ecosystem science, and conservation planning, yet they rely on biomass and canopy data that do not explicitly represent the stand structural attributes derived from tree diameter measurements. This limits the ability to compare size-related structure and within-stand heterogeneity at large spatial scales. Here we present a global, spatially explicit dataset of stand-level tree diameter structure for forest cover in 2020 at 0.027° (~3 km) resolution, based on 1,203,524 georeferenced forest inventory plots comprising 54.6 million trees (≥10 cm DBH) integrated with more than 50 environmental and satellite-derived covariates into machine learning models. The dataset provides the first globally consistent maps of three complementary diameter-based metrics: arithmetic mean diameter (D_{mean}), quadratic mean diameter (D_{qm}), and the coefficient of variation of diameter (D_{cv}), representing average tree size, large-tree dominance, and within-stand size variability, respectively. Model performance of the ecozone-specific Random Forest framework ranged from R^2 = 0.41–0.82 (RMSE = 3.91–4.63 cm) for D_{mean}, R^2 = 0.43–0.83 (RMSE = 4.38–5.27 cm) for D_{qm}, and R^2 = 0.47–0.62 with (RMSE = 0.10–0.13) for D_{cv} across different forest ecozones. By jointly quantifying central tendency and variability in tree size, the dataset revealed spatial patterns of forest structural organization not captured by existing biomass or canopy-height products. It provides a consistent baseline for cross-biome comparison of forest structure, supporting parameterization and evaluation of vegetation and Earth system models, while offering an independent benchmark for remotely sensed structural proxies. Furthermore, it enables spatial assessment of stand structural attributes, including large-tree dominance and structural complexity, facilitating integration of diameter-based structure into global analyses of carbon dynamics and ecosystem functioning.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

1. Introduction

Forest structure is a central organizing feature of terrestrial ecosystems, shaping carbon storage, resource allocation, and habitat conditions across biomes (Bonan et al., 2008, Pan et al., 2011, Friedlingstein et al., 2022). Diameter at breast height (DBH) is the most widely recorded tree-level attribute in forest inventories and provides a consistent basis for quantifying forest stand structure across regions and inventory systems. Because DBH is measured using standardized protocols and is directly linked to key structural and functional attributes, it underlies the estimation of basal area, growing stock, and aboveground biomass and carbon. As a result, DBH forms a core variable for ecological analyses and forest resource assessments (Gianfranco et al., 2007; Pretzsch, 2009; Wu et al., 2015; Zhou et al., 2018). In addition, DBH measurements support a range of operational applications, including inventory reporting, stand characterization, and silvicultural decision-making (Hardiman et al., 2011; Danescu et al., 2016; Liang et al., 2016; Müller et al., 2000; Peck et al., 2014; Dieler et al., 2017; Young et al., 2017).

Stand-level structure is commonly summarized by using statistical descriptors derived from individual tree diameters from forest inventories: (i) mean tree diameter (D_{mean}) represents central tendency, (ii) quadratic mean diameter (D_{qm}) emphasizes larger trees through diameter-squared weighting, and (iii) the coefficient of variation of diameter (D_{cv}) quantifies variability in DBH among trees and reflects structural heterogeneity within stands. These metrics summarize complex diameter information into interpretable quantities that can be compared across forest types, management regimes, and environmental settings (Curtis & Marshall, 2000; Larson & Churchill, 2012; McElhinny et al., 2005; Pretzsch, 2009).

Despite the extensive collection of tree-level measurements by forest inventory programs, their spatial coverage remains uneven. Sampling intensity varies across regions, and large areas, particularly remote and less accessible landscapes, are sparsely represented. Consequently, inventory data do not provide continuous spatial information on forest structural attributes, limiting their application in large-scale analyses (McRoberts et al., 2010; Crowther et al., 2015; Liang et al., 2016; Zhou et al., 2018). This mismatch between detailed plot-level measurements and the need for spatially explicit information has long constrained regional to global assessments of forest stand structure.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Recent advances in international data synthesis and the availability of spatially explicit environmental datasets have begun to address this limitation. Large-scale collaborative networks have enabled the integration of forest inventory data across continents, improving both the scale and consistency of available observations (Liang et al., 2016). At the same time, globally consistent climate, topographic, and vegetation datasets derived from satellite observations provide environmental covariates of field measurements. Integrating these data sources enables extrapolation of inventory-based structural information beyond sampled plots and supports the development of spatially continuous representations of forest structure (Crowther et al., 2015; Harris et al., 2012; Santoro et al., 2018; Spawn et al., 2020; Potapov et al., 2021).

In this study, we integrate standardized forest inventory observations with environmental and remotely sensed covariates within a machine-learning framework to derive a spatially explicit global dataset of forest diameter structural attributes. The resulting product provides spatially continuous estimates of three complementary metrics: central tendency (D_{mean}), diameter-weighted structure (D_{qm}), and within-stand variability (D_{cv}), across the global forest domain at 0.027° resolution. The dataset reflects structural frequency distributions associated with forest cover in 2020 and is derived from single-measurement inventory observations. It captures spatial variation in forest structure rather than temporal changes in growth or stand development. By extending inventory-based information into a spatially continuous form, this work establishes a standardized global baseline of diameter-based forest structure. This baseline supports comparative analyses across biomes, facilitates integration with ecological and Earth system models, and enables evaluation of structural information derived from remote sensing and other global products.

2. Materials and Methods

2.1. Overview

Development of the global dataset containing forest diameter attributes involved four main stages: (1) compilation and standardization of tree-level forest inventory observations from multiple sources, (2) derivation of stand-level diameter metrics by forest ecotypes (based on FAO Global Ecological Zones classification-FAO, 2012) for use in spatial modelling, (3)

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

ecozone-specific model training and validation, and (4) mapping of the tree size metrics on forest extent, based on the model performance **(Figure 1)**

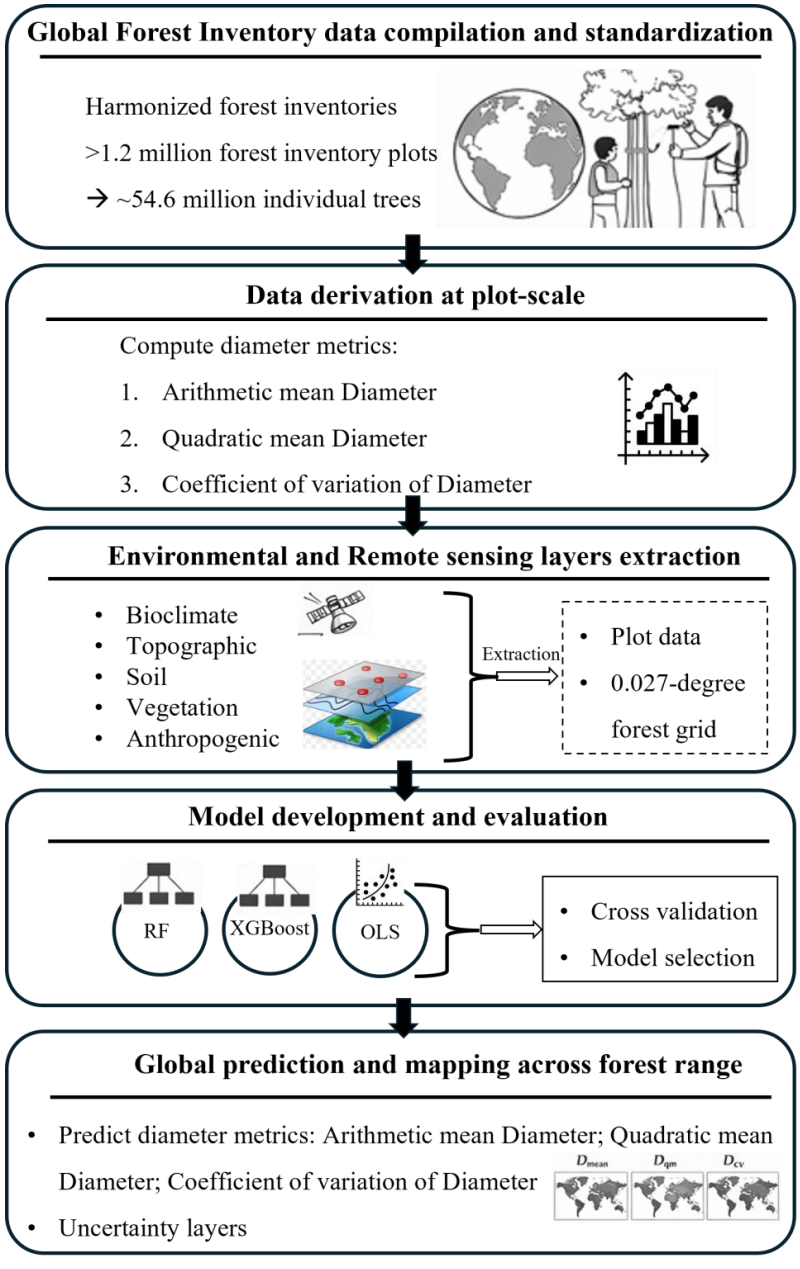


Figure 1. Workflow illustrates the key steps in generating global maps of forest structure in terms of tree stem diameter attributes.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

506 **2.2. Data preprocessing stages**

507 **(i) Data collection and assemblage:** Tree-level forest inventory measurements were compiled
508 from a wide range of sources, including global forest monitoring initiatives, national forest
509 inventory programs, and multiple collaborative research networks and data-sharing efforts. These
510 sources collectively provided extensive geographic coverage across forested regions worldwide
511 (Liang et al., 2016, 2022). The compiled dataset included 1,203,524 georeferenced forest
512 inventory plots representing approximately 54.6 million individual tree records and
513 encompassing approximately 600 km² of sampled forest area globally. Plot sizes varied across
514 contributing datasets due to differences in inventory design, ranging from small fixed-area plots
515 to larger inventory units. To ensure comparability across datasets, all plot-level measurements
516 were standardized to a per-hectare basis using plot-specific expansion factors (trees per hectare,
517 TPH) prior to analysis *(see Section 2.2-iii).*

518 For each tree observation, we retained information on plot identity, geographic coordinates,
519 taxonomic identity where available, DBH, and measurement year. DBH, measured at 1.3m
520 above ground level, represents the most consistently recorded structural variable across forest
521 inventory programs worldwide and serves as the primary measurement used to quantify tree size
522 within forest stands.

523

524 **(ii) Taxonomic standardization:** Forest inventory datasets compiled from multiple monitoring
525 programs often differ in taxonomic conventions and species naming formats. To ensure
526 taxonomic consistency across the compiled dataset, species nomenclature was standardized using
527 the Taxonomic Name Resolution Service (TNRS), which validates submitted names against
528 authoritative taxonomic repositories including World Flora Online and the World Checklist of
529 Vascular Plants (Boyle et al., 2013). During this process, scientific binomial names were retained
530 while subspecific ranks and author citations were removed to maintain consistent formatting
531 across records. For individuals identified only to morphospecies, standardized genus-level
532 identifiers were assigned. These identifiers were constructed by combining the genus name with
533 the label *"spp"*, the source dataset identifier, and a numeric suffix.

534

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

(iii) Data standardization: Forest inventory datasets compiled from multiple countries and monitoring programs vary substantially in sampling design, plot area, measurement protocols, and reporting formats. To ensure comparability, we harmonized and standardized data prior to analysis.

First, all records were screened for essential plot metadata, and only observations containing geographic coordinates, tree diameter measurements, plot identifiers, and measurement year were retained. To represent contemporary forest conditions, the dataset was restricted to plots measured from the year 2000 onward. Second, to standardize tree populations across datasets, we retained only stems with DBH ≥ 10 cm, a threshold widely adopted in global forest inventory analyses (Liang et al., 2022, Crowther et al., 2015). In cases where trees were recorded as multi-stem individuals, diameters were combined into an equivalent diameter based on cross-sectional area, ensuring that inclusion thresholds were applied consistently to the whole tree rather than individual stems. This consideration is particularly important in dry forests, where multi-stem growth forms are common. Stems smaller than this threshold were excluded to ensure that structural metrics were derived from comparable tree populations across regions. Where disturbance information was available, plots affected by recent stand-replacing or stand-altering disturbances (e.g., intensive harvesting, severe wildfire, or large-scale windthrow) were excluded, as these events can temporarily shift forest structure away from conditions representative of longer-term stand development. Third, because plot sizes and sampling approaches vary among inventories, tree counts were standardized using plot expansion factors that convert measured stems to trees per hectare (TPH). Expressing observations on a consistent per-area basis ensures that structural metrics derived from different inventories remain directly comparable despite differences in sampling design (Liang et al., 2022).

(iv) Data aggregation to plot-scale: Tree-level diameter measurements were aggregated to derive plot-level indicators describing forest diameter structure. Using all stems meeting the minimum threshold (DBH ≥ 10 cm), we calculated three complementary metrics of forest diameter structure, including (i) arithmetic mean diameter (D_{mean}), (ii) quadratic mean diameter (D_{qm}), and (iii) coefficient of variation of diameter (D_{cv}) **(Table 1).**

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

a. The arithmetic mean diameter represents the average DBH of trees within a plot and provides a measure of typical tree size. For each plot containing *n* trees, we calculated D_{mean} as:

$$D_{mean} = \frac{1}{n}\sum_{i=1}^{n} DB\,H_i \,,$$

where DBH_i denotes the diameter of the *i*th tree.

b. Quadratic mean diameter places greater weight on larger trees and therefore reflects the structural influence of dominant canopy individuals within forest stands. Because basal area scales with the square of stem diameter, this metric emphasizes the contribution of larger trees to stand structure. We calculated D_{qm} as:

$$D_{qm} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} DB\,H_i^2} \,,$$

where DBH_i denotes the diameter of the *i*th tree.

c. To quantify variability in tree size within each plot, we calculated the coefficient of variation of diameter (D_{cv}) as:

$$D_{cv} = \frac{SD(DBH)}{D_{mean}} \,,$$

where SD(DBH) is the standard deviation of tree diameter at breast height within a plot, and D_{mean} is the arithmetic mean diameter of all trees in that plot.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Table 1. Summary statistics of tree stem diameter structure metrics across global ecozones, including mean, standard deviation (SD), and sample size (n), derived from ground-based forest inventories.

Broader Group	Ecozone	D_{mean} (cm) (Mean ± SD)	D_{qm} (cm) (Mean ± SD)	D_{cv} (cm) (Mean ± SD)	n
Boreal	Boreal coniferous forest	17.64 ± 4.70	18.23 ± 5.07	0.26 ± 0.11	59,249
	Boreal tundra woodland	15.97 ± 4.42	16.51 ± 4.93	0.23 ± 0.11	1,295
	Boreal mountain system	17.31 ± 5.67	18.1 ± 6.31	0.29 ± 0.15	15,263
Temperate	Temperate continental forest	22.79 ± 6.45	24.18 ± 7.13	0.34 ± 0.13	297,427
	Temperate mountain system	25.76 ± 10.38	27.75 ± 11.69	0.38 ± 0.16	179,145
	Temperate desert	28.22 ± 8.61	30.38 ± 9.56	0.4 ± 0.14	7,976
	Temperate steppe	24.27 ± 7.42	25.95 ± 8.4	0.37 ± 0.16	10,554
	Temperate oceanic forest	30.02 ± 12.76	31.6 ± 13.33	0.32 ± 0.17	182,657
Subtropics	Subtropical mountain system	22.01 ± 9.10	23.4 ± 10.03	0.35 ± 0.16	100,115
	Subtropical humid forest	22.27 ± 6.20	23.8 ± 7.15	0.35 ± 0.15	108,850
	Subtropical dry forest	25.24 ± 12.06	26.36 ± 12.54	0.29 ± 0.16	56,023
	Subtropical desert	20.81 ± 9.31	21.75 ± 10.24	0.3 ± 0.17	7,524
	Subtropical steppe	21.39 ± 8.23	22.53 ± 9.06	0.32 ± 0.15	17,362
Tropics	Tropical shrubland	21.17 ± 7.41	23.48 ± 9.27	0.43 ± 0.20	575
	Tropical rainforest	22.50 ± 9.15	24.89 ± 10.62	0.45 ± 0.21	15,321
	Tropical moist forest	16.91 ± 4.96	17.91 ± 5.81	0.33 ± 0.16	26,727
	Tropical dry forest	18.18 ± 6.95	19.30 ± 7.90	0.34 ± 0.17	15,655
	Tropical mountain system	21.36 ± 7.58	22.98 ± 8.71	0.39 ± 0.18	8,438
	Global	**24.56 ± 10.24**	**26.06 ± 11.04**	**0.34 ± 0.15**	**1,203,524**

(v) Acquisition of covariates: To model spatial variation in forest diameter structure, we assembled a suite of environmental and remote sensing correlates representing major climatic, and physiographic gradients known to influence forest growth and stand development. These layers were compiled to represent climate, vegetation, topography, soil characteristics, and anthropogenic impacts, including global climate surfaces (Karger et al., 2017; Beck et al., 2019), structure and vegetation metrics derived from tree height and MODIS products (Potapov et al., 2021, Didan, 2021; Myneni et al., 2015; Running et al., 2017), forest age information (Besnard et

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

al., 2021), soil properties (Batjes, 1996), topographic variables (Amatulli et al., 2018), land-use and management datasets (Lesiv et al., 2022), human footprint indices (Venter et al., 2016), and ancillary global environmental datasets including aridity (Zomer et al., 2022) and livestock distribution (Gilbert et al., 2018) **(Table 2)**.

Climatic covariates describe long-term temperature and precipitation regimes that regulate forest productivity. Satellite-derived indicators of canopy condition were included to represent multiple dimensions of forest structure and function, including vegetation greenness (NDVI), leaf area index (LAI), light availability and use (photosynthetically active radiation), and structural attributes such as canopy height. Additional covariates representing terrain characteristics, soil properties, and anthropogenic pressures were included to represent environmental constraints influencing forest development. Because the objective of this study is spatial prediction rather than inference on individual drivers, we adopted a comprehensive predictor set to ensure that major environmental gradients influencing forest structure were represented across diverse biomes. In total, 55 covariates were initially assembled. Following screening for multicollinearity and limited contribution to model performance, two variables were removed **(Table 2).**

(vi) Spatial harmonization of covariates: Covariates, which served as the predictors, were originally available at different spatial resolutions and coordinate systems. To ensure spatial consistency between the covariates and forest inventory observations, we harmonized all the covariate layers to a common global grid with a spatial resolution of 0.027° referenced to the WGS84 geographic coordinate system. Spatial processing included reprojection, resampling, and raster alignment to ensure consistent spatial support across predictor datasets. Following this harmonization step, the values were extracted at the geographic coordinates of each forest inventory plot. The extracted values were then used as explanatory variables for training predictive models of plot-level forest diameter structure metrics **(Table 2).** Spatial processing and statistical analyses were conducted in R using the packages *terra*, *raster*, *sf*, *data.table*, *ranger*, and *xgboost* (Hijmans et al., 2018, Pebesma et al., 2018).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026


Open Access Earth System Science Data Discussions

(vii) Construction of the global prediction grid: To generate spatially explicit predictions of forest diameter structure, we constructed a global prediction grid representing forest areas in 2020. The grid was defined in the WGS84 geographic coordinate system with a spatial resolution of 0.027° × 0.027° (~3 km at the equator) and was derived from the global canopy height dataset developed by Potapov et al. 2021, which identifies forests using a minimum canopy height threshold of 5m. Only grid cells meeting this criterion were retained in the prediction domain, resulting in approximately 5.7 million forested grid cells worldwide. Values were extracted at the centroid of each forest grid cell to construct the explanatory dataset used for spatial prediction of stand forest diameter metrics *(see Section 2.2-vi)*. To address missing values in covariates, common in mountainous or data-sparse regions, we used predictive mean matching using the *aregImpute()* function from the *Hmisc* package in R (Harrell et al., 2020).

(viii) Classification of Ecozones: To facilitate ecological interpretation of global patterns, both forest inventory plots and prediction grid cells were assigned to Global Ecological Zones (GEZ) defined by the Food and Agriculture Organization of the United Nations -FAO 2012) **(Figure S1)**. The GEZ framework delineates global ecological regions based on climatic conditions, vegetation characteristics, and biogeographic patterns. Linking observations and spatial grid to FAO ecological zones provides a consistent ecological framework for summarizing and comparing patterns of forest diameter structure across major forest biomes. This classification enables regional analysis of tree size metrics and supports interpretation of global variation in forest structural attributes.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

662 **Table 2. Summary statistics for environmental and structure-vegetation-related variables**
663 **used as predictors, across 1,203,524 forest inventory plots to predict tree stem diameter**
664 **attributes globally.**
665

Category	Covariates	Variable	Unit	Resolution	Mean	SD	References
Structure, composition & vegetation	Plot size	Ps	ha	ha	0.06	0.14	Author-generated
	Species richness	R	ha^{-1}	3 km	10.14	11.20	Author-generated
	Forest age	A	Years	1 km	76.14	37.31	Besnard et al., 2021
	Normalized difference vegetation index	V_1	Unitless	1 km	0.82	0.11	Didan, 2021
	Enhanced vegetation index	V_2	Unitless	1 km	0.55	0.13	Didan, 2021
	Evapotranspiration	V_3	$mm.year^{-1}$	1 km	475	209	Running et al., 2017
	Leaf area index	V_4	Unitless	1 km	5.53	1.88	Myneni et al., 2015
	Photosynthetic radiation	V_5	Unitless	1km	0.48	0.13	Myneni et al., 2015
	Forest height	V_6	m	30 m	12.85	8.15	Potapov et al.,2021
Bioclimate	Mean annual air temperature	B_1	°C	1 km	10.31	6.01	Karger et al., 2017
	Mean diurnal air temperature range	B_2	°C	1 km	8.80	2.18	Karger et al., 2017
	Isothermality (ratio)	B_3	Unitless	1 km	31.76	11.71	Karger et al., 2017
	Temperature seasonality	B_4	0.01°C	1 km	729.46	256.44	Karger et al., 2017
	Mean daily maximum air temperature of the warmest month	B_5	°C	1 km	25.31	4.45	Karger et al., 2017
	Mean daily minimum air temperature of the coldest month	B_6	°C	1 km	-4.19	9.29	Karger et al., 2017
	Annual range of air temperature	B_7	°C	1 km	29.50	7.56	Karger et al., 2017
	Mean daily air temperatures of the wettest quarter	B_8	°C	1 km	14.96	7.01	Karger et al., 2017
	Mean daily air temperatures of the driest quarter	B_9	°C	1 km	6.20	11.29	Karger et al., 2017
	Mean daily air temperatures of the warmest quarter	B_{10}	°C	1 km	19.54	4.18	Karger et al., 2017

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

	Mean daily air temperatures of the coldest quarter	B_{11}	°C	1 km	0.81	8.84	Karger et al., 2017
	Annual precipitation amount	B_{12}	$kg.m^{-2}$	1 km	1082.83	523.32	Beck et al., 2019
	Precipitation amount of the wettest month	B_{13}	$kg.m^{-2}$	1 km	157.94	100.16	Karger et al., 2017
	Precipitation amount of the driest month	B_{14}	$kg.m^{-2}$	1 km	50.10	31.58	Karger et al., 2017
	Precipitation seasonality (coefficient of variation)	B_{15}	Unitless	1 km	36.77	26.66	Karger et al., 2017
	Mean monthly precipitation amount of the wettest quarter	B_{16}	$kg.m^{-2}$	1 km	425.13	253.50	Karger et al., 2017
	Mean monthly precipitation amount of the driest quarter	B_{17}	$kg.m^{-2}$	1 km	170.86	101.34	Karger et al., 2017
	Mean monthly precipitation amount of the warmest quarter	B_{18}	$kg.m^{-2}$	1 km	340.17	210.97	Karger et al., 2017
	Mean monthly precipitation amount of the coldest quarter	B_{19}	$kg.m^{-2}$	1 km	230.92	165.45	Karger et al., 2017
	Potential evapotranspiration	B_{20}	$mm.year^{-1}$	1 km	43.9	10.8	Zomer et al., 2022
	Indexed annual aridity	B_{21}	Unitless	1 km	9668.77	4996.84	Zomer et al., 2022
Topography	Terrain ruggedness Index	T_1	Unitless	1 km	18.59	19.00	Amatulli et al., 2018
	Roughness	T_2	Unitless	1 km	57.31	59.38	Amatulli et al., 2018
	Slope	T_3	Degree	1 km	4.85	5.12	Amatulli et al., 2018
	Aspect cosine	T_4	Unitless	1 km	-3.06	174.55	Amatulli et al., 2018
	Aspect sine	T_5	Unitless	1 km	-3.05	174.55	Amatulli et al., 2018
	First order partial derivative (N-S slope)	T_6	Unitless	1 km	0.00	0.08	Amatulli et al., 2018
	Second-order partial derivative (E-W slope)	T_7	Unitless	1 km	0.00	0.00	Amatulli et al., 2018
	First-order partial derivative (E-W slope)	T_8	Unitless	1 km	0.00	0.07	Amatulli et al., 2018

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

	Second-order partial derivative (N-S slope)	T_9	Unitless	1 km	0.00	0.00	Amatulli et al., 2018
	Profile curvature	T_{10}	Unitless	1 km	0.00	0.00	Amatulli et al., 2018
	Topographic Position Index	T_{11}	Unitless	1 km	0.14	4.82	Amatulli et al., 2018
	Tangential curvature	T_{12}	Unitless	1 km	0.00	0.00	Amatulli et al., 2018
	Elevation	T_{13}	M	1 km	555.43	632.23	Amatulli et al., 2018
Soil	Organic carbon content	S_1	$g.kg^{-1}$	1 km	41.61	45.63	Batjes,1996
	Percent clay	S_2	%	1 km	18.79	8.48	Batjes,1996
	pH	S_3	Unitless	1 km	5.86	1.04	Batjes,1996
	Bulk density	S_4	$g \cdot cm^{-3}$	1 km	1.25	0.24	Batjes,1996
	C/N ratio	S_5	Unitless	1 km	13.53	3.03	Batjes,1996
	Total nitrogen	S_6	$g.kg^{-1}$	1 km	2.20	2.26	Batjes,1996
	Electrical conductivity	S_7	$dS \cdot m^{-1}$	1 km	0.68	1.40	Batjes,1996
Anthropogenic	The Human Footprint (HF), a measure of the cumulative impact of direct human pressure	G_1	Unitless	1km	9.15	8.63	Venter et al., 2016
	Distribution of cattle	G_2	number of cattle per ~100 km^2	$100km^2$	865.83	1386.03	Gilbert et al, 2018
	Forest management	G_3	Unitless	100m	37.55	38.18	Lesiv et al., 2022

2.3. Model training

2.3.1. Spatial prediction of diameter structure

Spatial prediction of forest diameter structure was performed by relating the ground truth diameter metrics to gridded covariates used as predictors. Because relationships between

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

environmental conditions and forest structure vary across broad climatic regions, modelling was conducted separately within ecozones, rather than imposing a single global model. Within each ecozone, three structural indicators derived from tree diameter measurements were modelled: D_{mean}, D_{qm}, and D_{cv}. The covariates extracted for forest inventory plots served as explanatory variables for model development *(see Table 2 for a full list of covariates used)*. Three modelling approaches representing different statistical paradigms were evaluated: a linear regression model (ordinary least squares; OLS) and two nonlinear ensemble methods, Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The OLS model provides a transparent, parametric baseline by assuming linear relationships between predictors and response variables. In contrast, RF and XGBoost are tree-based ensemble learning methods that iteratively combine multiple decision trees to improve predictive performance. RF builds an ensemble of decorrelated trees using bootstrap aggregation and random feature selection, while XGBoost employs gradient boosting to sequentially minimize prediction errors. These approaches are widely used in ecological modelling because they can capture complex nonlinear relationships and interactions among predictors without requiring predefined functional forms (Breiman, 2001; Cutler et al., 2007; Chen & Guestrin, 2016, Liaw et al., 2022).

Model predictions were generated within a bootstrap framework *(see Section 2.5),* in which repeated model realizations were produced for each ecozone. The final predicted value for each grid cell was calculated as the mean of predictions across bootstrap iterations, providing a robust estimate of the forest diameter attributes.

2.4. Model tuning

Hyperparameters for the machine-learning approaches were optimized using a randomized search across candidate parameter combinations. For Random Forest, candidate configurations included variation in ensemble size (*10–250 decision trees*) and the number of predictors evaluated at each split (*mtry = 1–52*). Model error stabilized at *250 decision trees*, and the best-performing configuration used *mtry = 50*. For XGBoost, tuning explored parameters controlling model learning and complexity, including *eta, max_depth, subsample, colsample_bytree, min_child_weight*, and *nrounds*. The final parameter configuration for each ecozone was selected based on the lowest validation root mean square error.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026


Open Access Earth System Science Data Discussions

706

707 **2.5. Model performance evaluation**

708

709 The reliability of the global forest diameter structure dataset was evaluated using predictive
710 validation and comparison among candidate modelling approaches. Model performance was
711 evaluated using the independent validation dataset with three metrics: the coefficient of
712 determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). To
713 facilitate comparison across broader climate gradients, the ecozone-level predictions were
714 grouped into four major domains: boreal, temperate, subtropical, and tropical, by grouping the
715 corresponding ecozonal categories *(see Table 1 Broader Group column)*. Among the evaluated
716 algorithms, RF consistently produced the most reliable predictive performance across the groups
717 **(Tables 3-5).** Therefore, RF models trained separately within each ecozone were selected to
718 generate spatial predictions of D_{mean}, D_{qm}, and D_{cv} across the global forest prediction grid. The
719 final models were applied to ~5.7 million forested grid cells to produce spatially continuous
720 maps of forest diameter structure **(Figure 2a, Figure 3a, Figure 4a).**

721 To further evaluate model robustness, a bootstrap resampling procedure was implemented to
722 generate repeated model realizations, allowing estimation of prediction variability associated
723 with sampling and model fitting (Efron & Tibshirani, 1993; Hastie et al., 2009). For each
724 ecozone, we randomly partitioned the dataset into training (90%) and validation (10%) subsets
725 **(Figures S2-S4).** A fixed partitioning scheme was used across all model runs to ensure that
726 differences in predictive performance reflected model behavior rather than variation in data
727 splits. All modelling procedures were implemented in R, using the *ranger* package for Random
728 Forest and *xgboost* for gradient boosting models. Prediction uncertainty was quantified as the
729 standard deviation of predictions across bootstrap iterations, providing an estimate of variability
730 arising from model fitting and sampling **(Figure 2b, Figure 3b, Figure 4b).** Model performance
731 was evaluated using independent validation data by comparing observed and predicted values,
732 with agreement assessed relative to the 1:1 reference line and the fitted regression trend.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

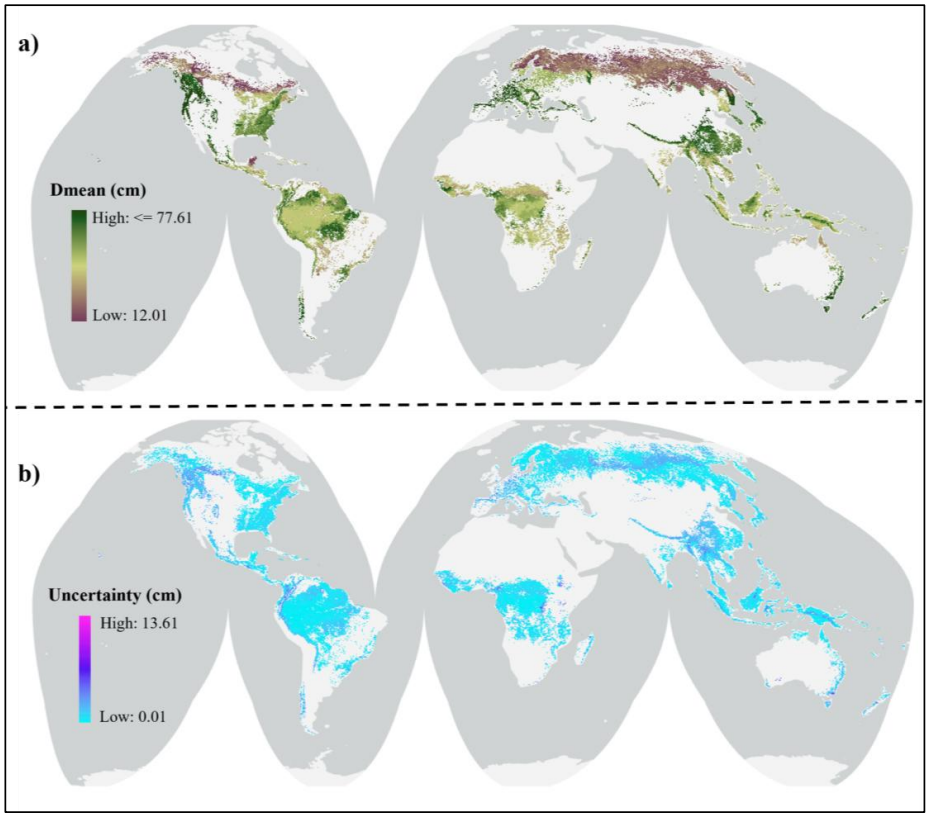


Figure 2. Global mapping and associated uncertainty of predicted forest arithmetic mean diameter (D_{mean}, cm). (a) Spatial distribution of predicted D_{mean} across global forests at 0.027° resolution, generated using Random Forest models trained on forest inventory observations and 53 covariates. (b) Prediction uncertainty in D_{mean} is expressed as the standard deviation (cm) of predicted D_{mean} values across Random Forest model iterations Both maps are displayed using the Mollweide equal-area projection (WGS84 datum).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

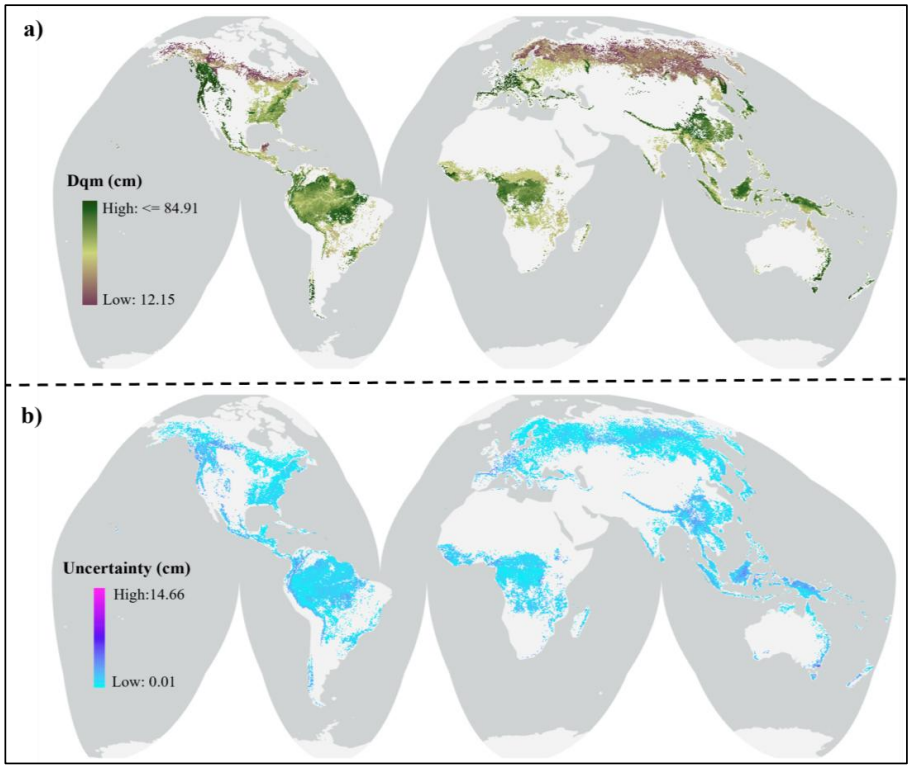
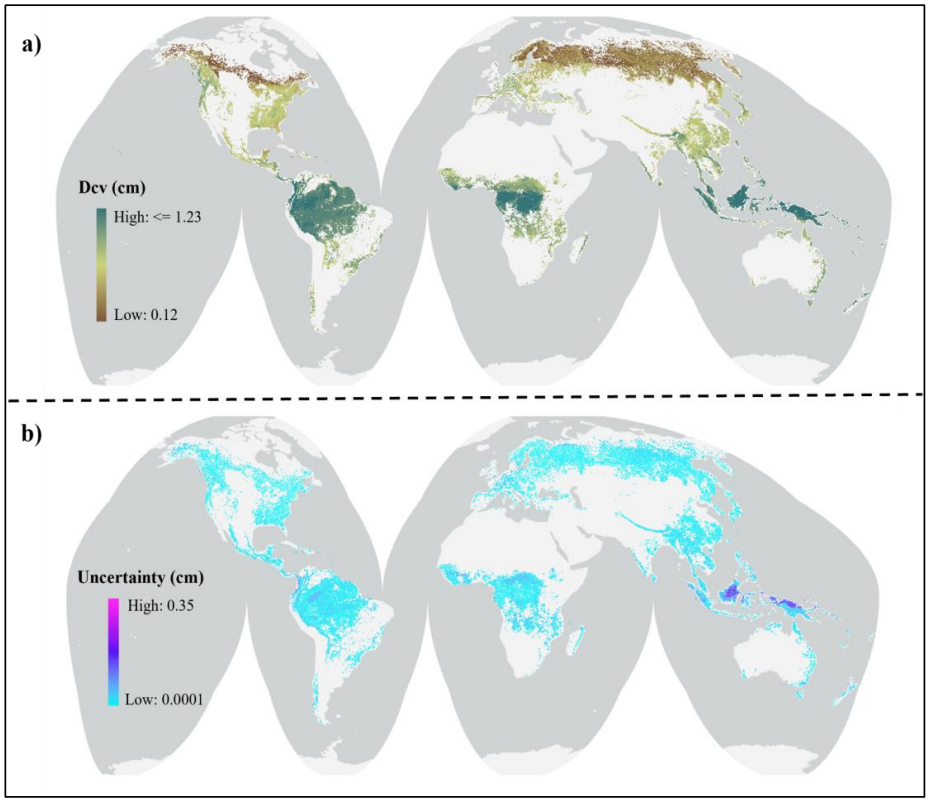


Figure 3. Global mapping and associated uncertainty of predicted forest quadratic mean diameter (D_{qm}, cm). (a) Spatial distribution of predicted D_{qm} across global forests at 0.027° resolution, generated using Random Forest models trained on forest inventory observations and 53 covariates. (b) Prediction uncertainty in D_{qm} is expressed as the standard deviation (cm) of predicted D_{qm} values across Random Forest model iterations Both maps are displayed using the Mollweide equal-area projection (WGS84 datum).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions



747

748 **Figure 4. Global mapping and associated uncertainty of predicted forest diameter**
749 **variability (D_{cv}).** (a) Spatial distribution of predicted D_{cv} across global forests at 0.027°
750 resolution, generated using Random Forest models trained on forest inventory observations and
751 53 covariates. (b) Prediction uncertainty in D_{cv} is expressed as the standard deviation (cm) of
752 predicted D_{cv} values across Random Forest model iterations Both maps are displayed using the
753 Mollweide equal-area projection (WGS84 datum).
754

755 ## 3. Results

756 ### 3.1. Validation and model selection

757 For the three complementary forest stand diameter metrics over the ecozonal categories, Random
758 Forest consistently demonstrated the most reliable predictive performance for different model
759 approaches **(Table 3-5, Figures S2-S4).** The superior performance of ensemble tree-based RF
760 algorithms likely reflects its ability to represent nonlinear relationships and interactions among

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

761 environmental predictors (Breiman, 2001; Cutler et al., 2007; Belgiu & Drăguţ, 2016). Forest
762 diameter structure arises from the combined influence of climate, vegetation dynamics, site
763 conditions, and disturbance, arising from complex ecological relationships that are difficult to
764 capture using linear models. Based on these validation results, RF predictions trained within each
765 ecozone within the broader category, were retained as the final gridded product distributed with
766 the dataset **(Figures 2-4).**

767 **a. Performance for arithmetic mean diameter (D_{mean}):** Across all broad climatic groupings,
768 RF consistently reproduced observed values of arithmetic mean diameter (D_{mean}) more accurately
769 than the alternative models **(Table 3).** In boreal forests, RF achieved strong performance (R^2 =
770 0.82; RMSE = 4.45; MAE = 3.07), substantially outperforming XGBoost and OLS, which
771 explained considerably less variance (R^2 = 0.45 and 0.35, respectively). In temperate forests, RF
772 explained 52% of the variance in observed D_{mean} (R^2 = 0.52), again exceeding the performance of
773 XGBoost (R^2 = 0.19) and OLS (R^2 = 0.11). Similar patterns were observed in subtropical
774 ecosystems, where RF achieved moderate explanatory power (R^2 = 0.41), while alternative
775 models showed near-zero performance. In tropical forests, RF and XGBoost performed
776 comparably (R^2 = 0.81 and 0.79, respectively), although RF produced lower prediction errors.

777

778

779

780

781

782

783

784

785

786

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Open Access Earth System Science Data Discussions

Table 3. Model performance comparison for predicting D_{mean} across ecozones grouped into major climatic zones.

Ecozone group	Candidate Model	R^2	RMSE	MAE
Boreal	RF	0.82	4.45	3.07
	XGBoost	0.45	7.68	5.14
	OLS	0.35	8.34	5.70
Temperate	RF	0.52	3.91	2.73
	XGBoost	0.19	5.04	3.76
	OLS	0.11	5.30	3.98
Subtropics	RF	0.41	4.63	3.26
	XGBoost	0.05	5.84	4.42
	OLS	0.01	5.97	4.53
Tropics	RF	0.81	4.11	2.57
	XGBoost	0.79	4.35	2.89
	OLS	0.61	5.90	3.77

b. Performance for quadratic mean diameter (D_{qm}): Model evaluation for D_{qm} showed consistent patterns across ecozones with the RF model achieving the highest predictive accuracy in all climatic groups **(Table 4).** In boreal forests, RF explained a large proportion of variance ($R^2 = 0.83$), outperforming XGBoost and OLS ($R^2 = 0.48$ and 0.39, respectively). In temperate regions, RF accounted for 54% of the observed variation in D_{qm} ($R^2 = 0.54$), again exceeding the performance of the alternative models. Subtropical ecozones showed lower overall predictive power, although RF still improved model fit ($R^2 = 0.43$) relative to XGBoost and OLS. In tropical forests, RF and XGBoost performed similarly ($R^2 = 0.78$ and 0.76, respectively), while RF consistently produced lower prediction errors (RMSE and MAE).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Table 4. Model performance comparison for predicting D_{qm} across ecozones grouped into major climatic zones.

Ecozone group	Candidate Model	R^2	RMSE	MAE
Boreal	RF	0.83	4.94	3.40
	XGBoost	0.48	8.74	5.91
	OLS	0.39	9.53	6.59
Temperate	RF	0.54	4.38	3.06
	XGBoost	0.21	5.73	4.30
	OLS	0.12	6.04	4.57
Subtropics	RF	0.43	5.27	3.72
	XGBoost	0.07	6.74	5.15
	OLS	0.01	6.95	5.31
Tropics	RF	0.78	5.09	3.18
	XGBoost	0.76	5.36	3.56
	OLS	0.61	6.80	4.41

c. **Performance for diameter variability (D_{cv}):** Predictive performance for the coefficient of variation of diameter (D_{cv}) was generally lower than for the mean diameter metrics reflecting the greater complexity of modeling within-stand structural variability. Nevertheless, the RF model consistently outperformed alternative approaches across all climatic zones **(Table 5).** In boreal forests, RF achieved moderate performance (R^2 = 0.62; RMSE = 0.10; MAE = 0.07), exceeding XGBoost and OLS (R^2 = 0.31 and 0.23, respectively). Similar patterns were observed in temperate ecosystems, where RF explained 47% of the observed variation in D_{cv}. In subtropical and tropical forests, RF again showed higher explanatory power (R^2 = 0.52 and 0.47, respectively) and lower prediction errors compared with the alternative models.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Table 5. Model performance comparison for predicting D_{cv} across ecozones grouped into major climatic zones.

Ecozone group	Candidate Model	R^2	RMSE	MAE
Boreal	RF	0.62	0.10	0.07
	XGBoost	0.31	0.13	0.10
	OLS	0.23	0.14	0.11
Temperate	RF	0.47	0.09	0.06
	XGBoost	0.15	0.11	0.09
	OLS	0.09	0.12	0.09
Subtropics	RF	0.52	0.11	0.07
	XGBoost	0.11	0.14	0.11
	OLS	0.05	0.15	0.12
Tropics	RF	0.47	0.13	0.10
	XGBoost	0.38	0.15	0.11
	OLS	0.22	0.16	0.12

3.2. Uncertainty Assessment

To support cautious use of the gridded diameter products, each prediction layer was accompanied by a map of pixel-level variability **(Figure 2b, Figure 3b, Figure 4b)**. For every grid cell, uncertainty is expressed as the standard deviation of predicted values generated across repeated model iterations (*see section 2.5. Model performance evaluation*). Small standard deviations indicate that estimates are stable across repeated fits, whereas larger values highlight locations where predictions shift more across iterations and are therefore less stable. Spatial patterns in prediction variability primarily track (i) the density and representativeness of training plots, and (ii) how unusual local environmental conditions are relative to those observed in the training data. Areas with substantial inventory coverage tend to show lower variability, while regions with sparse sampling and/or environmental conditions near the edges of the training domain more often exhibit higher variability. These layers flag areas where mapped diameter

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

metrics warrant added caution, to guide masking or weighting in downstream analyses, and to contextualize comparisons among regions. Because diameter metrics were computed directly from measured tree diameters prior to model fitting, the uncertainty reported here reflects model-based spatial prediction variability (and associated sensitivity to training data and covariates), rather than measurement error in the underlying field observations.

4. Data Availability

The datasets generated in this study provide a spatially explicit representation of forest tree diameter structure across the global forest extent 2020. All data products are publicly available through the *Figshare* repository (under license CC by 4.0) and are distributed in formats designed to support geospatial analysis, ecological modeling, and large-scale environmental assessments.

The repository contains both gridded raster layers and a tabular dataset describing predicted forest diameter metrics and associated uncertainty estimates. Spatial predictions were produced on a global forest grid with a spatial resolution of 0.027° × 0.027°, corresponding to ~3 km at the equator. Raster layers are provided in GeoTIFF format using the WGS 84 coordinate reference system, enabling direct use in standard geographic information systems and spatial analysis software (https://figshare.com/s/689e4be80a63d05d5189; Mitra et al., 2026). Three principal variables describing forest diameter structure are included in the dataset: (i) D_{mean} (ii) D_{qm} (iii) D_{cv}. For each diameter metric, two raster layers are provided: a prediction layer representing the average value derived from repeated model iterations (*"Dmean_ecozonal.tif, Dqm_mean_ecozonal.tif, Dcv_mean_ecozonal.tif"*), and a corresponding uncertainty layer representing the standard deviation of predictions across those iterations *("Uncertainty_Dmean.tif, Uncertainty_Dqm.tif, Uncertainty_Dcv.tif").*

In addition to raster products, the *Figshare* repository includes grid-level tabular datasets containing detailed information for each forested grid cell with the predictions of D_{mean}, D_{qm}, and D_{cv} (https://figshare.com/s/91c6cc4f2b9757d92da8; Mitra et al., 2026). Each record corresponds to a single grid cell and includes geographic coordinates (LON and LAT), assigned Global Ecological Zone (Ecozone), corresponding broader domain of ecozone (GEZ_GROUP),

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026


Open Access Earth System Science Data Discussions

covariates used as predictors used in the modeling framework as listed in Table 1, predicted diameter metrics from the ecozone-specific models (D_{mean}, D_{qm}, and D_{cv}), and their associated uncertainty estimates. This tabular format allows users to access the modeled variables directly and facilitates integration with statistical analyses or external datasets.

Comprehensive metadata describing the harmonized training dataset, including the characteristics and sources of all contributing in-situ forest inventory datasets, are available at https://figshare.com/s/34c21c03883a01ef708f; Mitra et al., 2026. A public version of the harmonized training data frame containing all predictor variables required to reproduce the modeling workflow is available at https://figshare.com/s/c795146ea9e8aa42f81f; Mitra et al., 2026.

All spatial products are additionally accessible through the Science-i international research platform (https://science-i.org). Raw forest inventory data underlying the harmonized dataset are subject to data-sharing agreements and confidentiality restrictions imposed by contributing data providers. Access to sensitive components (e.g., precise plot coordinates and tree-level records) may be requested from the corresponding author via Science-i or the Global Forest Biodiversity Initiative (GFBI), subject to approval by the original data owners.

5. Code availability

All data processing, harmonization, species abundance matrix construction, computation of SBA and average tree size, geostatistical imputation, model training, and cross-validation were implemented in R (v3.4.2) using command-line workflows. Computationally intensive analyses were performed on Purdue University's Bell and Negishi high-performance computing (HPC) clusters within Linux-based environments optimized for parallel and high-memory workloads (compute nodes up to 1 TB RAM; clusters up to 64 nodes). Additional covariate preparation, spatial mapping, and quality-control procedures were conducted on Windows-based workstations (up to 128 GB RAM). The Random Forest (RF) models required to reproduce the modeling framework presented in this study are publicly available under an MIT license at *Figshare* https://figshare.com/s/60eac94e4c64a95c6d13; Mitra et al., 2026.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

6. Conclusion

This study presents a spatially explicit dataset of forest diameter structure derived from standardized forest inventory observations and environmental predictors, providing estimates of mean diameter (D_{mean}), quadratic mean diameter (D_{qm}), and diameter variability (D_{cv}) at 0.027° resolution **(Figures 2-4).** Based on more than 1.2 million georeferenced forest inventory plots that includes 54.6 million tree records within an ecozone-specific modeling framework, the dataset extends field measurements into a continuous representation of forest structural attributes with associated uncertainty. The dataset captures variation in tree size, the influence of larger stems, and within-stand heterogeneity across ecological regions. Agreement between predicted and observed values, together with consistency among derived metrics **(Figures S2-S4)**, indicates that the mapped diameter attributes retain key structural properties of the underlying inventory data and complement existing biomass- and canopy-focused products.

The availability of spatially continuous diameter metrics supports applications in ecological analysis, vegetation and Earth system model parameterization and evaluation, benchmarking of remotely sensed structural indicators, and large-scale assessment of forest condition and structural variability **(Figures 2a, 3a, 4a).** The accompanying uncertainty layers allow users to evaluate prediction stability, identify areas requiring cautious interpretation, and incorporate uncertainty into downstream analyses **(Figures 2b, 3b, 4b)**.

Several limitations should be considered when applying the dataset. Predictions are influenced by the spatial distribution of inventory plots and the resolution and completeness of environmental predictors, which may lead to higher uncertainty in under-sampled regions or under-represented environmental conditions. In addition, the dataset represents structural conditions associated with forest cover 2020 and does not capture temporal dynamics or changes in stand development. Continued expansion of harmonized inventory data and advances in environmental and canopy-structure datasets, including lidar-based observations, will support further refinement and higher-resolution representations of forest structure. In conclusion, this work establishes a framework for representing forest diameter structure from standardized

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

observations, enabling improved integration of structural information into ecological analyses and modeling applications.

Author contributions

Conceptualization: A.M. and J. L

Methodology: A.M. and J. L

Data curation: A.M., J. L and W.J.C.; contributing co-authors curated their respective regional datasets as part of national or institutional monitoring programs.

Data provision: All contributing co-authors provided and maintained the original forest inventory datasets used in this study.

Writing, revision and editing: all.

Competing interests: At least one of the (co-)authors is a member of the editorial board of Earth System Science Data.

Additional information: Supplementary information is available for this paper.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

Acknowledgments

This study was supported, in part, by the World Resource Institute (WRI); U.S. Army Engineer Research and Development Center (US-ERDC); the Purdue University Office of Agricultural Research and Graduate Education and the Executive Vice President for Research through their joint Elevating the Visibility of Research initiative; the U.S. Department of Agriculture National Institute of Food and Agriculture McIntire Stennis Projects 1017711 and 1016676; and a faculty start-up grant from the Department of Forestry and Natural Resources, Purdue University (JL). JCS was supported by the Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), funded by the Danish National Research Foundation (grant DNRF173). RM was supported by the Czech Science Foundation (project No. 25-18519S). RT was supported by the Czech Science Foundation (project No. 21-24186M). J.A. and J.D. were supported by Research Grants of the Czech Science Foundation (25-15727S), the Czech Ministry of Education, Youth and Sports (LUAUS26250, program INTER-EXCELLENCE, subprogram INTER-ACTION), Mobility Plus between the Czech Republic and Taiwan (NSTC-26-05), and long-term research development project of Institute of Botany of the Czech Academy of Sciences (RVO 67985939). L.M. and L.K. acknowledge that plot data from Central Siberian forests were collected with support from the State Assignment of the Ministry of Science and Higher Education of the Russian Federation [L.M. was supported by project no. FWES-2024-0023 (state registration number 124012900558-7), and L.K. was supported by project no. FWES-2024-0028 (state registration number 124012900557-0)]. E.C. acknowledges funding from the Czech Ministry of Education, Youth and Sports - AdAgriF (CZ.02.01.01/00/22_008/0004635). We would like to thank the Italian and French National Forest Inventories for the work done to make forest inventory data publicly available.

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

963

964 **References**

965 1. Amatulli, G. et al. A suite of global cross-scale topographic variables for environmental
966 and biodiversity modelling. *Sci. Data* 5, 180040 (2018).
967 2. Batjes, N. H. Development of a world dataset of soil water retention properties using
968 pedotransfer rules. *Geoderma* 71, 31–52 (1996).
969 3. Beck, H. E. et al. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and
970 quantitative assessment. *Bull. Am. Meteorol. Soc.* 100, 473–500 (2019).
971 4. Belgiu, M. & Drăguţ, L. Random forest in remote sensing: a review of applications and
972 future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31 (2016).
973 5. Besnard, S. et al. Mapping global forest age from forest inventories, biomass and climate
974 data. *Earth Syst. Sci. Data* 13, 4881–4896 (2021).
975 6. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of
976 forests. *Science* 320, 1444–1449 (2008).
977 7. Boyle, B. et al. The taxonomic name resolution service: an online tool for automated
978 standardization of plant names. *BMC Bioinformatics* 14, 16 (2013).
979 8. Breiman, L. Random forests. *Mach. Learn.* 45, 5–32 (2001).
980 9. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proc. 22nd ACM*
981 *SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 785–794 (2016).
982 10. Chowdhury, T. A., Thiel, C. & Schmullius, C. Growing stock volume estimation from L-
983 band ALOS PALSAR polarimetric coherence in Siberian forest. *Remote Sens. Environ.*
984 155, 129–144 (2014).
985 11. Crowther, T. W. et al. Mapping tree density at a global scale. *Nature* 525, 201–205
986 (2015).
987 12. Curtis, R. O. & Marshall, D. D. Why quadratic mean diameter? *West. J. Appl. For.* 15,
988 137–139 (2000).
989 13. Cutler, D. R. et al. Random forests for classification in ecology. *Ecology* 88, 2783–2792
990 (2007).
991 14. Danescu, A., Albrecht, A. T. & Bauhus, J. Structural diversity promotes productivity of
992 mixed, uneven-aged forests in southwestern Germany. *Oecologia* 182, 319–333 (2016).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

15. Didan, K. MODIS/Terra vegetation indices 16-day L3 global 250 m SIN grid V061. NASA Land Processes Distributed Active Archive Center (2021). https://doi.org/10.5067/MODIS/MOD13Q1.061
16. Dieler, J. et al. Effect of forest stand management on species composition, structural diversity and productivity in the temperate zone of Europe. *Eur. J. For. Res.* 136, 739–766 (2017).
17. Efron, B. & Tibshirani, R. J. *An introduction to the bootstrap.* (Chapman & Hall/CRC, New York, 1993).
18. FAO. Global ecological zones for FAO forest reporting. Food and Agriculture Organization of the United Nations (2012).
19. Friedlingstein, P. et al. Global Carbon Budget 2022. *Earth Syst. Sci. Data* 14, 4811–4900 (2022).
20. Gianfranco, S., Laura, M. & David, G. Development of a neural network model to update forest distribution data for managed alpine stands. *Ecol. Model.* 206, 331–346 (2007).
21. Gilbert, M. et al. Global distribution data for livestock species in 2010. *Sci. Data* 5, 180227 (2018).
22. Hardiman, B. S., Bohrer, G., Gough, C. M., Vogel, C. S. & Curtis, P. S. The role of canopy structural complexity in wood net primary production of a maturing northern deciduous forest. *Ecology* 92, 1818–1827 (2011).
23. Harrell, F. E. & Dupont, C. Hmisc: Harrell miscellaneous. R package v4 (2020).
24. Harris, N. L. et al. Baseline map of carbon emissions from deforestation in tropical regions. *Science* 336, 1573–1576 (2012).
25. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference and prediction.* (Springer, New York, 2009).
26. Hijmans, R. J. raster: Geographic data analysis and modelling. R package v2.8 (2018).
27. Karger, D. N. et al. Climatologies at high resolution for the Earth's land surface areas. *Sci. Data* 4, 170122 (2017).
28. Larson, A. J. & Churchill, D. Tree spatial patterns in fire-frequent forests of western North America. *For. Ecol. Manage.* 267, 74–92 (2012).
29. Lesiv, M. et al. Global forest management data for 2015 at a 100 m resolution. *Sci. Data* 9, 199 (2022).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

30. Liang, J. et al. Positive biodiversity–productivity relationship predominant in global forests. *Science* 354, aaf8957 (2016).
31. Liang, J. et al. Co-limitation shapes global forest diversity gradients. *Nat. Ecol. Evol.* 6, 1423–1437 (2022).
32. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* 2, 18–22 (2002).
33. McElhinny, C., Gibbons, P., Brack, C. & Bauhus, J. Forest and woodland stand structural complexity: its definition and measurement. *For. Ecol. Manage.* 218, 1–24 (2005).
34. McRoberts, R. E., Tomppo, E. O., Finley, A. O. & Heikkinen, J. Estimating areal means and variances of forest attributes using the k-nearest neighbours technique and satellite imagery. *Remote Sens. Environ.* 114, 911–921 (2010).
35. Mitra, A. et al. Predictions of global forest diameter attributes with uncertainty raster layers. figshare https://figshare.com/s/689e4be80a63d05d5189 (2026).
36. Mitra, A. et al. Predictions of global forest diameter attributes tabular dataset. figshare https://figshare.com/s/91c6cc4f2b9757d92da8 (2026).
37. Mitra, A. et al. Harmonized forest inventory metadata. figshare https://figshare.com/s/34c21c03883a01ef708f (2026).
38. Mitra, A. et al. Harmonized training dataset for global forest diameter predictions. figshare https://figshare.com/s/c795146ea9e8aa42f81f (2026).
39. Mitra, A. et al. Random Forest models and relevant codes for global forest diameter predictions. figshare https://figshare.com/s/60eac94e4c64a95c6d13 (2026)
40. Müller, S., Ammer, C. & Nüßlein, S. Analyses of stand structure as a tool for silvicultural decisions: a case study in Quercus petraea–Sorbus torminalis stands. *Forstwiss. Centralbl.* 119, 32–42 (2000).
41. Myneni, R., Knyazikhin, Y. & Park, T. MOD15A2H MODIS/Terra leaf area index/FPAR 8-day L4 global 500 m SIN grid V006. NASA LP DAAC (2015). https://doi.org/10.5067/MODIS/MOD15A2H.006
42. Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* 333, 988–993 (2011).
43. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* 10, 439–446 (2018).

https://doi.org/10.5194/essd-2026-226
Preprint. Discussion started: 22 May 2026

1055 44. Peck, J. E., Zenner, E. K., Brang, P. & Zingg, A. Tree size distribution and abundance
1056 explain structural complexity differentially within stands of even-aged and uneven-aged
1057 structure types. *Eur. J. For. Res.* 133, 335–346 (2014).
1058 45. Potapov, P. et al. Mapping global forest canopy height through integration of GEDI and
1059 Landsat data. *Remote Sens. Environ.* 253, 112165 (2021).
1060 46. Pretzsch, H. *Forest dynamics, growth and yield: from measurement to model.* (Springer,
1061 Berlin, 2009).
1062 47. Running, S., Mu, Q. & Zhao, M. MOD16A3 MODIS/Terra net evapotranspiration yearly
1063 L4 global 500 m SIN grid V006. NASA LP DAAC (2017).
1064 https://doi.org/10.5067/MODIS/MOD16A3.006
1065 48. Santoro, M. et al. The BIOMASAR algorithm: global forest growing stock volume and
1066 biomass mapping from Envisat ASAR data. *Remote Sens. Environ.* 182, 164–178 (2018).
1067 49. Spawn, S. A., Sullivan, C. C., Lark, T. J. & Gibbs, H. K. Harmonized global maps of
1068 above- and belowground biomass carbon density in the year 2010. *Sci. Data* 7, 112
1069 (2020).
1070 50. Venter, O. et al. Global terrestrial human footprint maps for 1993 and 2009. *Sci. Data* 3,
1071 160067 (2016).
1072 51. Wu, D. S. & Ji, Y. Q. Dynamic estimation of forest volume based on multi-source data
1073 and neural network model. *J. Agric. Sci.* 7, 18–31 (2015).
1074 52. Young, B. D., D'Amato, A. W., Kern, C. C., Kastendick, D. N. & Palik, B. J. Seven
1075 decades of change in forest structure and composition in *Pinus resinosa* forests in
1076 northern Minnesota, USA: comparing managed and unmanaged conditions. *For. Ecol.*
1077 *Manage.* 395, 92–103 (2017).
1078 53. Zhou, R. Y., Wu, D. S., Fang, L. M., Xu, A. J. & Lou, X. W. A Levenberg–Marquardt
1079 backpropagation neural network for predicting forest growing stock based on least-
1080 squares equation fitting parameters. *Forests* 9, 757 (2018).
1081 54. Zomer, R. J., Xu, J. & Trabucco, A. Version 3 of the global aridity index and potential
1082 evapotranspiration
1083