Submitted Manuscript: Confidential

Template revised November 2022

Title: Predicting global organic-matter decomposition in flowing waters

Authors: Scott D. Tiegs[1]*#, Krista A. Capps[2]*#, John Paul Schmidt*[2], David M. Costello*[3], Christopher J. Patrick*[4], Jennifer J. Follstad Shah[5], Carri J. LeRoy[6], Vicenç Acuña[7], Ricardo Albariño[8], Daniel C. Allen[9], Cecilia Alonso[10], Patricio Andino[11], Clay Arango[12], Jukka Aroviita[13], Marcus V.M. Barbosa[14], Leon A. Barmuta[15], Colden Baxter[16], Brent Bellinger[17], Luz Boyero[18], Lyubov Bragina[19], Lee E. Brown[20], Andreas Bruder[21], Denise A. Bruesewitz[22], Francis Burdon[23], Marcos Callisto[24], Antonio Camacho[25], Cristina Canhoto[26], María M. Castillo[27], Eric Chauvet[28], Joanne Clapcott[29], Fanny Colas[30], Checo Colón-Gaud[31], Julien Cornut[21], Verónica Crespo-Pérez[11], Wyatt F. Cross[32], Joseph Culp[33], Michael Danger[34], Olivier Dangles[35], Elvira de Eyto[36], Alison M. Derry[37], Veronica Díaz Villanueva[8], Michael M. Douglas[38], Arturo Elosegi[18], Andrea C. Encalada[39], Sally Entrekin[40], Rodrigo Espinosa[41], Verónica Ferreira[26], Carmen Ferriol[25], Kyla M. Flanagan[42], Alexander S. Flecker[43], Tadeusz Fleituch[44], André Frainer[45,46], Nikolai Friberg[45,47], Paul C. Frost[48], Erica A. Garcia[49], Liliana García-Lago[50], Pavel Ernesto García Soto[51], Mark O. Gessner[52,53], Sudeep Ghate[54], Darren P. Giling[52,55,56], Alan Gilmer[57], José Francisco Gonçalves Jr[58], Rosario Karina Gonzales[59], Manuel A.S. Graça[26], Mike Grace[60], Natalie A. Griffiths[61], Hans-Peter Grossart[52,62], François Guérold[34], Vlad Gulis[63], Pablo E. Gutiérrez-Fonseca[64], Luiz U. Hepp[65], Scott Higgins[66], Takuo Hishi[67], Joseph Huddart[68], John Hudson[69], Samantha Imberger[70], Carlos Iñiguez-Armijos[71], Mark W. Isken[1], Tomoya Iwata[72], David J. Janetski[73], Andrea E. Kirkwood[74], Aaron A. Koning[75], Sarian Kosten[76], Kevin A. Kuehn[77], Hjalmar Laudon[78], Peter R. Leavitt[79], Aurea L Lemes da Silva[80], Shawn Leroux[81], Peter J. Lisi[82], Richard MacKenzie[83], Amy M. Marcarelli[84], Frank O. Masese[85], Peter B. McIntyre[43], Brendan G. McKie[78], Adriana Medeiros[86], Kristian Meissner[13], Marko Miliša[87], Shailendra Mishra[88], Yo Miyake[89], Ashley Moerke[90], Shorok Mombrikotb[68], Rob Mooney[82], Tim Moulton[91], Timo Muotka[92], Junjiro Negishi[93], Vinicius Neres-Lima[94], Mika L. Nieminen[95], Jorge Nimptsch[96], Jakub Ondruch[97], Riku Paavola[92], Isabel Pardo[50], Edwin T.H.M. Peeters[98], Jesus Pozo[18], Aaron Prussian[99], Estefania Quenta[59], Brian Reid[100], John S. Richardson[101], Anna Rigosi[102], José Rincón[103], Geta Risnoveanu[104], Christopher T. Robinson[105], Lorena Rodríguez-Gallego[106], Todd V. Royer[107], James A. Rusak[108], Anna C. Santamans[25], Géza B. Selmeczy[109], Gelas Simiyu[85], Agnija Skuja[110], Jerzy Smykla[44], Ryan Sponseller[111], Kandikere R. Sridhar[112], Aaron Stoler[113], Christopher M. Swan[114], Franco Teixeira-de Mello[115], Jonathan D. Tonkin[116], Sari Uusheimo[13], Allison M. Veach[117], Sirje Vilbaste[118], Lena B.-M. Vought[119], Chiao-Ping Wang[120], Jackson R. Webster[40], Paul B.Wilson[121], Stefan Woelfl[96], Guy Woodward[68], Marguerite A. Xenopoulos[48], Adam G. Yates[122], Chihiro Yoshimura[123], Catherine M. Yule[124], Yixin Zhang[125], Jacob A. Zwart[126]

Affiliations: CELLDEX Consortium. See Supplemental Files for author affiliations

*Corresponding authors: tiegs@oakland.edu, kcapps@uga.edu, jps@uga.edu, dcostel3@kent.edu, cpatrick@vims.edu

#These authors contributed equally to this work.

Abstract:

Rivers and streams contribute to the global carbon cycle by decomposing vast quantities of organic matter, but decomposition rates are highly heterogenous, and our understanding of large-scale patterns and drivers of this process remains limited. Using fine-scale climate, land-use, and water-quality data, we generated a predictive model that explains most (81%) of the variation in cellulose-decomposition rates across 514 streams spanning 135° of latitude. Our model reveals key environmental controls of decomposition, including geologic, climatic, and, importantly, anthropogenic attributes. Projections of cellulose decomposition, when combined with genus-level litter-quality attributes, predict leaf-litter-decomposition rates accurately at the global scale (70% of variance explained). High-resolution predictions of cellulose and natural-leaf-litter decomposition provide novel insight into carbon cycling in flowing waters worldwide, including vast, unstudied areas of Earth.

One-Sentence Summary: A distributed experiment yields robust, predictive models of organic-matter decomposition in rivers worldwide.

Main Text:

Earth's terrestrial ecosystems produce over 100 billion tons of plant detritus annually (*1*), and the fates of this organic matter (for example, long-term storage, mineralization to greenhouse gasses, or incorporation into biomass) depend on the rate at which it is decomposed. River ecosystems are carbon-processing hotspots, receiving and decomposing vast quantities of terrestrial plant matter, especially relative to the small fraction of the landscape they cover (*2*, *3*). By connecting terrestrial ecosystems with storage compartments, including floodplains, lakes, and oceans, rivers play a unique role in the global carbon cycle, functioning both as organic-matter conduits and reactors. Despite the widely recognized importance of flowing waters in global carbon cycling (*4*–*6*), our understanding of patterns in organic-matter decomposition and its drivers at large spatial scales is still limited (*7*).

Large-scale spatial patterns of organic-matter decomposition have been estimated by comparing leaf-litter decomposition rates from studies conducted in regions with contrasting climate regimes (*8*, *9*), conducting literature reviews of local field studies (*10*), developing conceptual models (*11*, *12*) and performing meta-analyses (*13*, *14*). Coordinated, distributed experiments (*15*–*19*) have been particularly insightful by generating directly comparable data across broad geographic areas and identifying coarse-resolution predictors of decomposition rates in rivers, including differences in decomposer communities and biomes. Still, we lack a comprehensive understanding of how drivers such as climate, geology, vegetation, water quality, and soils interact to govern organic-matter decomposition at large scales. Such knowledge gaps are particularly evident across the tropics and in lower-income economies – ecologically important areas where rivers are grossly understudied relative to those in northern temperate zones. Quantifying patterns and controls of decomposition in these areas is vital, however, because much of Earth's terrestrial plant matter is produced in the tropics (*20*), resulting in substantial transfers of terrestrially derived carbon to the ocean (*21*).

Effectively modeling carbon dynamics at the global scale – including areas where field data are scarce – requires a mechanistic understanding of the many environmental and biotic factors that drive organic-matter decomposition. Robust estimates generated by combining existing empirical measurements with fine-scale geospatial and environmental data have the potential to reduce the need for additional data in remote or difficult-to-access regions, contribute to a higher-resolution understanding of decomposition, support global modeling efforts of carbon dynamics, and generate baseline estimates for decomposition in understudied areas of the world. Of particular value are models that can accurately predict current *in-situ* decomposition rates across space, enabling manipulation of environmental drivers *in silico* to predict impacts under scenarios of future global environmental change.

Here, we present a predictive model fitted with global data from a coordinated, distributed experiment on cellulose decomposition in rivers to reveal previously undocumented patterns in decomposition and the key factors driving this fundamental ecosystem-level process. Decomposition of cellulose, the most abundant organic polymer on the planet and a main constituent of plant litter, was quantified by over 150 investigators by using a common cellulose-decomposition assay (*22*). The assay was performed in 514 flowing-water ecosystems at georeferenced field sites on all seven continents, spanning 135º of latitude and each of Earth's major terrestrial biomes (*18*, *19*). We used high-resolution (15 arcsecond) climate, soil, geology, vegetation, and water-quality data (101 predictors) in a boosted-regression tree (BRT) algorithm to develop the first global, high-resolution predictive model of organic-matter decomposition in rivers. We then independently validated the predictive power of the model by generating interpolated cellulose-decomposition rates and genus-level leaf-litter chemistry traits to predict leaf-litter decomposition at 861 locations across the globe. We found that cellulose is an excellent proxy for predicting litter-decomposition rates and that physicochemical factors at river and watershed scales interact with characteristics of the organic matter being decomposed to create heterogenous patterns in riverine decomposition across the planet. This unique approach produced the first truly data-driven perspective of organic-matter decomposition in flowing waters worldwide.

Results and Discussion

Predicting decomposition rates. Cellulose-decomposition rates are remarkably predictable across flowing waters worldwide. The selected climate, geology, soil, and water-quality predictors explain 81% of variance in field measurements of *in-situ* cellulose decomposition. Because an identical substrate was used at all field sites, observed variation in decomposition rates can be attributed unequivocally to the activity of microbial communities and environmental drivers. Prior efforts have explained broad variation in decomposition rates across riverine ecosystems as a function of exogenous factors such as temperature (*13*, *18*) and concentrations of dissolved nutrients (*16*, *23*), as well as litter traits (*14*, *24*, *25*). Our model supports those prior efforts and found that climatic and water-quality parameters are among the most important predictors of decomposition rates (Fig 1). However, a relatively large number of predictor variables (n=26) have importance values greater than 1.0 (Supplemental Table 1), and no single predictor contributes >15% to the predictive power of the model (Supplemental Table 1). This result reflects the complexity of predicting decomposition-related processes at the global scale.

Top predictors of cellulose decomposition include expected attributes like mean daily stream water temperature (importance value 14.0; Fig. 1A), nitrogen and phosphorus availability (6.7 and 4.9, respectively; Fig. 1C&D), and mean annual air temperature (2.5; Fig. 1F). Our data and approach also highlight watershed-level characteristics that have been given little attention previously, such as sub-watershed lake area (importance value 6.9, Fig. 1B), actual evapotranspiration in the watershed (4.4, Fig. 1E), and the chemical and physical properties of soil (Supplemental Table 1). Although the study sites were selected to have minimal human impacts relative to their region of study (*18*), variables associated with anthropogenic development, such as dissolved nutrient yields, crop-land extent (2.0), population count (1.3), and river regulation (1.3), still emerge as important predictors (Supplemental Table 1). Notably, relationships between predictors and decomposition rates are frequently non-linear, often revealing thresholds beyond which there are abrupt changes in decomposition rates (e.g., Fig. 1B, D, & E). Temperature affects cellulose decomposition in multiple ways, including a strong positive effect of water temperature (Fig. 1A), and an optimal range (5-13 °C) of annual air temperature with predicted lower rates in both cooler and warmer watersheds (Fig. 1F).

Extrapolating to global patterns of decomposition rates. Our model of riverine cellulose decomposition reveals pronounced, large-scale spatial patterns of organic-matter processing (Fig. 2). Rates generally increase with decreasing latitude, with most rapid rates in tropical regions (e.g., Central America, Amazon Basin, Western Africa, Indo-Pacific) and areas characterized by volcanic activity and young soils, an effect previously documented only at more local scales (*26*). Fluvial ecosystems in these regions are among the least studied (Fig. 2, inset) on the planet, yet they have high rates of terrestrial primary production (*20*) and carbon export to the ocean (*21*). Vast areas in middle latitudes with ubiquitous human impacts – Central Europe, eastern China, central North America, southeast South America, and Japan – also support elevated decomposition rates, strongly suggesting continental-scale human impacts on carbon cycling in rivers. In contrast, areas of boreal forests, characterized by short growing seasons, low temperatures, and peaty, acidic, water-logged soils, exhibit lower rates of organic-matter decomposition, especially in northern Asia, eastern Scandinavia, and northeast Canada.

Validating predicted cellulose-decomposition rates with leaf-litter-decomposition rates. Recognizing that the cotton strips used in our standardized decomposition assay lack the chemical complexity of organic matter that naturally enters running waters, we tested whether our modeling approach could predict variation in the decomposition rates of terrestrial leaf litter in rivers reported by ecologists worldwide. Model forecasts were independently validated using 891 unique litter-decomposition rates from 559 locations and representing 35 genera of terrestrial plants (*24*). Leaf and litter-trait data at the genus level (*27*, *28*) and experimental conditions (*13*, *24*) were also used as predictors to account for variation among decomposition estimates resulting from differences in leaf and litter quality (e.g., lignin, hemicellulose, tannin, nutrient content) and the feeding activity of invertebrates (Figure 3A, Supplemental Table 2). Our validation model accounts for 70% of the variation in leaf-litter decomposition. The predictive power of this model is overwhelmingly driven by predicted rates of cellulose decomposition (relative importance 39.5), despite the striking differences in quality between the cotton-strip substrate and natural litter (Fig 3A, Supplemental Table 2). This result provides strong support for the critical role that exogenous drivers play in riverine litter decomposition. Prior research at large scales has stressed the importance of litter quality as the predominant control of decomposition rates in rivers (*14*); our results highlight the critical role of environmental factors, such as temperature and nutrient availability, at regulating decomposition rates at large scales. Our validation model also revealed that invertebrate access to leaves, as assessed by experimentally manipulating litter-bag mesh size, greatly increased the rate of decomposition in all but the fastest decomposing leaves (Fig. 3A). Finally, litter chemistry contributes to the predictive power of the model in expected ways, with genera that are characterized by high lignin content and high carbon to nitrogen ratios decomposing more slowly (importance values 11.9 and 5.5; Figs. 3B,C). Litter nitrogen content (importance value 5.2; Fig 3D) was the only other litter trait that provides more explanatory power than expected by chance, including, surprisingly, whether leaves were fresh or senesced when entering the river (Supplemental Table 2). It is well recognized that leaf-litter chemistry can vary among individuals within a species (*29*) and even individual leaves from a tree (*30*); thus, our model may underestimate the importance of individual-level variation in leaf chemistry in driving decomposition. To advance our understanding of endogenous controls at global scales, we advocate for greater measurement and reporting of litter chemistry, especially nitrogen and lignin content. Despite limitations in available data, we have shown that cellulose decomposition can be an excellent proxy for litter decomposition, and our composite model of environmental drivers makes reasonable estimates of litter decomposition at a global scale.

Forecasting decomposition under global environmental change. The high predictive power of our cellulose and leaf-litter decomposition models enables forecasting of decomposition rates under altered climate, land cover, soil conditions, and nutrient-loading scenarios. These predictions can identify locations across the globe where decomposition may be particularly susceptible or resistant to global change, thereby informing freshwater-conservation efforts. As proof of concept, we examined potential changes in predicted litter-decomposition rates associated with changes in pine-oak forest composition in Mexican watersheds invaded by pine bark beetle (*31*). Our forecasts predict that insect-induced canopy replacement from pine to oak would nearly double mean decomposition rates given expected changes in litter quality, with particularly marked effects in the Rio Grande de Santiago watershed, a major conduit of organic matter to the Pacific Ocean (Fig. 4). To promote the use of our models for forecasting, we created an easy-to-use, open-source online application where users can estimate both cotton-strip and leaf-litter decomposition rates for any river across the globe (*https://costello.shinyapps.io/celldex_map/*).

Conclusions and implications. By pairing a distributed field experiment with publicly available environmental data, we created the first high-resolution predictions of organic-matter decomposition rates in flowing waters worldwide. Our model demonstrates that cellulose-decomposition rates result from diverse, interacting, and non-linear environmental forcings that can best be described with complex, data-rich models. Our relatively simple organic-matter substrate serves as an excellent proxy for litter decomposition, which should encourage the use of the cotton-strip method and this distributed experimental approach. Simplification of the leaf-litter-bag assay allowed us to both achieve standardized results and fill important geographic gaps in remote and low-resourced areas, which should further encourage coordinated, distributed experiments (*32*). Although our datasets were large when compared against other studies of organic-matter decomposition, the field data used were both spatially and temporally sparse, which makes our strong predictive power all the more striking. Thus, this work underscores the power of machine-learning algorithms and large geographic databases of environmental data (e.g., Hydrobasins) plus the critical value of geographically and temporally extensive data from simple but standardized coordinated experiments (e.g., CELLDEX).

Cellulose decomposition is strongly influenced by many environmental drivers that are and will continue to be impacted by anthropogenic activities. Undoubtably, climate change, increased nutrient loading, intensified land-use modification, and water extraction will continue to alter organic-matter processing in rivers and streams. Notably, the human-influenced drivers of cellulose decomposition – especially nutrient loading and temperature – are positively related to cellulose-decomposition rates. A critical implication is that, in the presence of continued environmental change, organic-matter decomposition rates will likely increase in rivers, resulting in declines in shorter-term carbon storage (*33*) and reductions in carbon transfer to longer-term storage compartments, such as reservoirs, floodplains, and oceans. This study provides new insight into ecological processes in rivers and a global understanding of the drivers of carbon fluxes between land and water. Given the interest in using ecosystem functions for biomonitoring (*34*, *35*), our estimated decomposition rates for immense, unstudied areas of the planet provide baseline data for supporting the development of biomonitoring networks in areas where they are most needed (*36*), and we have made this model accessible through an open-source online tool. Application of the models to current and future environmental threats (e.g., invasive species, biodiversity loss, land-use change, climate change) will empower scientists and natural-resource managers to better predict how those threats may alter the structure and function of river networks worldwide.

References

1. J. Cebrian, Patterns in the fate of production in plant communities. American Naturalist **154**, 449-468 (1999).
2. M. E. McClain, E. W. Boyer, C. L. Dent, S. E. Gergel, N. B. Grimm, P. M. Groffman, S. C. Hart, J. W. Harvey, C. A. Johnston, E. Mayorga, W. H. McDowell, G. Pinay, Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems* **6**, 301-312 (2003).
3. E. R. Hotchkiss, R. O. Hall, R. A. Sponseller, D. Butman, J. Klaminder, H. Laudon, M. Rosvall, J. Karlsson, Sources of and processes controlling CO_2 emissions change with the size of streams and rivers. Nature Geoscience **8**, 696-699 (2015).
4. P. Regnier, P. Friedlingstein, P. Ciais, F. T. Mackenzie, N. Gruber, I. A. Janssens, G. G. Laruelle, R. Lauerwald, S. Luyssaert, A. J. Andersson, S. Arndt, C. Arnosti, A. V. Borges, A. W. Dale, A. Gallego-Sala, Y. Goddéris, N. Goossens, J. Hartmann, C. Heinze, T. Ilyina, F. Joos, D. E. Larowe, J. Leifeld, F. J. R. Meysman, G. Munhoven, P. A. Raymond, R. Spahni, P. Suntharalingam, M. Thullner, Anthropogenic perturbation of the carbon fluxes from land to ocean. Nature Geoscience **6**, 597-607 (2013).
5. A. Marx, J. Dusek, J. Jankovec, M. Sanda, T. Vogel, R. van Geldern, J. Hartmann, J. A. C. Barth, A review of CO_2 and associated carbon dynamics in headwater streams: A global perspective. Reviews of Geophysics **55**, 560-585 (2017).
6. P. A. Raymond, J. Hartmann, R. Lauerwald, S. Sobek, C. McDonald, M. Hoover, D. Butman, R. Striegl, E. Mayorga, C. Humborg, P. Kortelainen, H. Dürr, M. Meybeck, P. Ciais, P. Guth, Global carbon dioxide emissions from inland waters. Nature **503**, 355-359 (2013).
7. T. J. Battin, R. Lauerwald, E. S. Bernhardt, E. Bertuzzo, L. G. Gener, R. O. Hall, E. R. Hotchkiss, T. Maavara, T. M. Pavelsky, L. Ran, P. Raymond, J. A. Rosentreter, P. Regnier, River ecosystem metabolism and carbon biogeochemistry in a changing world. Nature **613**, 449–459 (2023).
8. V. Ferreira, A. C. Encalada, M. A. S. Graça, Effects of litter diversity on decomposition and biological colonization of submerged litter in temperate and tropical streams. Freshwater Science **31**, 945-962 (2012).
9. A. Bruder, M. H. Schindler, M. S. Moretti, M. O. Gessner, Litter decomposition in a temperate and a tropical stream: The effects of species mixing, litter quality and shredders. Freshwater Biology **59**, 438-449 (2014).
10. J. C. Marks, Revisiting the fates of dead leaves that fall into streams. Annual Review of Ecology, Evolution, and Systematics **50**, 547-568 (2019).
11. M. A. S. Graça, The role of invertebrates on leaf litter decomposition in streams - A review. International Review of Hydrobiology **86**, 383-393 (2001).
12. T. V. Royer, G. W. Minshall, Controls on leaf processing in streams from spatial-scaling and hierarchical perspectives. Journal of the North American Benthological Society **22**, 352–358 (2003).
13. J. J. Follstad Shah, J. S. Kominoski, M. Ardón, W. K. Dodds, M. O. Gessner, N. A. Griffiths, C. P. Hawkins, S. L. Johnson, A. Lecerf, C. J. LeRoy, D. W. P. Manning, A. D. Rosemond, R. L. Sinsabaugh, C. M. Swan, J. R. Webster, L. H. Zeglin, Global synthesis of the temperature sensitivity of leaf litter breakdown in streams and rivers. Global Change Biology **23**, 3064-3075 (2017).

14. M. Zhang, X. Cheng, Q. Geng, Z. Shi, Y. Luo, X. Xu, Leaf litter traits predominantly control litter decomposition in streams worldwide. Global Ecology and Biogeography **28**, 1469-1486 (2019).

15. L. Boyero, R. G. Pearson, M. O. Gessner, L. A. Barmuta, V. Ferreira, M. A. S. Graça, D. Dudgeon, A. J. Boulton, M. Callisto, E. Chauvet, J. E. Helson, A. Bruder, R. J. Albariño, C. M. Yule, M. Arunachalam, J. N. Davies, R. Figueroa, A. S. Flecker, A. Ramírez, R. G. Death, T. Iwata, J. M. Mathooko, C. Mathuriau, J. F. Gonçalves, M. S. Moretti, T. Jinggut, S. Lamothe, C. M'Erimba, L. Ratnarajah, M. H. Schindler, J. Castela, L. M. Buria, A. Cornejo, V. D. Villanueva, D. C. West, A global experiment suggests climate warming will not accelerate litter decomposition in streams but might reduce carbon sequestration. Ecology Letters **14**, 289-294 (2011).

16. G. Woodward, M. O. Gessner, P. S. Giller, V. Gulis, S. Hladyz, A. Lecerf, B. Malmqvist, B. G. McKie, S. D. Tiegs, H. Cariss, M. Dobson, A. Elosegi, V. Ferreira, M. A. S. Graça, T. Fleituch, J. O. Lacoursière, M. Nistorescu, J. Pozo, G. Risnoveanu, M. Schindler, A. Vadineanu, L. B.-M. Vought, E. Chauvet, and others. Continental-scale effects of nutrient pollution on stream ecosystem functioning. Science **336**, 1438-1440 (2012).

17. I. T. Handa, R. Aerts, F. Berendse, M. P. Berg, A. Bruder, O. Butenschoen, E. Chauvet, M. O. Gessner, J. Jabiol, M. Makkonen, B. G. McKie, B. Malmqvist, E. T. H. M. Peeters, S. Scheu, B. Schmid, J. Van Ruijven, V. C. A. Vos, S. Hättenschwiler, Consequences of biodiversity loss for litter decomposition across biomes. Nature **509**, 218-221 (2014).

18. S.D. Tiegs, D. M. Costello, M. W. Isken, G. Woodward, P. B. McIntyre, M. O. Gessner, E. Chauvet, N. A. Griffiths, A. S. Flecker, V. Acuña, R. Albariño, D. C. Allen, C. Alonso, P. Andino, C. Arango, J. Aroviita, M. V. M. Barbosa, L. A. Barmuta, C. V. Baxter, T. D. C. Bell, B. Bellinger, L. Boyero, L. E. Brown, A. Bruder, D. A. Bruesewitz, F. J. Burdon, M. Callisto, C. Canhoto, K. A. Capps, M. M. Castillo, J. Clapcott, F. Colas, C. Colón-Gaud, J. Cornut, V. Crespo-Pérez, W. F. Cross, J. M. Culp, M. Danger, O. Dangles, E. De Eyto, A. M. Derry, V. D. Villanueva, M. M. Douglas, A. Elosegi, A. C. Encalada, S. Entrekin, R. Espinosa, D. Ethaiya, V. Ferreira, C. Ferriol, K. M. Flanagan, T. Fleituch, J. J. F. Shah, A. F. Barbosa, N. Friberg, P. C. Frost, E. A. Garcia, L. G. Lago, P. E. G. Soto, S. Ghate, D. P. Giling, A. Gilmer, J. F. Gonçalves, R. K. Gonzales, M. A. S. Graça, M. Grace, H.-P. Grossart, F. Guérold, V. Gulis, L. U. Hepp, S. Higgins, T. Hishi, J. Huddart, J. Hudson, S. Imberger, C. Iñiguez-Armijos, T. Iwata, D. J. Janetski, E. Jennings, A. E. Kirkwood, A. A. Koning, S. Kosten, K. A. Kuehn, H. Laudon, P. R. Leavitt, A. L. L. Da Silva, S. J. Leroux, C. J. LeRoy, P. J. Lisi, R. MacKenzie, A. M. Marcarelli, F. O. Masese, B. G. McKie, A. O. Medeiros, K. Meissner, M. Miliša, S. Mishra, Y. Miyake, A. Moerke, S. Mombrikotb, R. Mooney, T. Moulton, T. Muotka, J. N. Negishi, V. Neres-Lima, M. L. Nieminen, J. Nimptsch, J. Ondruch, R. Paavola, I. Pardo, C. J. Patrick, E. T. H. M. Peeters, J. Pozo, C. Pringle, A. Prussian, E. Quenta, A. Quesada, B. Reid, J. S. Richardson, A. Rigosi, J. Rincón, G. Rîşnoveanu, C. T. Robinson, L. Rodríguez-Gallego, T. V. Royer, J. A. Rusak, A. C. Santamans, G. B. Selmeczy, G. Simiyu, A. Skuja, J. Smykla, K. R. Sridhar, R. Sponseller, A. Stoler, C. M. Swan, D. Szlag, F. Teixeira-De Mello, J. D. Tonkin, S. Uusheimo, A. M. Veach, S. Vilbaste, L. B. M. Vought, C.-P. Wang, J. R. Webster, P. B. Wilson, S. Woelfl, M. A. Xenopoulos, A. G. Yates, C. Yoshimura, C. M. Yule, Y. X. Zhang, J. A. Zwart, Global patterns and drivers of ecosystem functioning in rivers and riparian zones. Science Advance **5**, 1-8 (2019).

19. D. M. Costello, S. D. Tiegs, L. Boyero, C. Canhoto, K. A. Capps, M. Danger, P. C. Frost, M. O. Gessner, N. A. Griffiths, H. M. Halvorson, K. A. Kuehn, A. M. Marcarelli, T. V. Royer, D. M. Mathie, R. J. Albariño, C. P. Arango, J. Aroviita, C. V. Baxter, B. J. Bellinger, A. Bruder, F. J. Burdon, M. Callisto, A. Camacho, F. Colas, J. Cornut, V. Crespo-Pérez, W. F. Cross, A. M. Derry, M. M. Douglas, A. Elosegi, E. de Eyto, V. Ferreira, C. Ferriol, T. Fleituch, J. J. Follstad Shah, A. Frainer, E. A. Garcia, L. García, P. E. García, D. P. Giling, R. K. Gonzales-Pomar, M. A. S. Graça, H.-P. Grossart, F. Guérold, L. U. Hepp, S. N. Higgins, T. Hishi, C. Iñiguez-Armijos, T. Iwata, A. E. Kirkwood, A. A. Koning, S. Kosten, H. Laudon, P. R. Leavitt, A. L. Lemes da Silva, S. J. Leroux, C. J. LeRoy, P. J. Lisi, F. O. Masese, P. B. McIntyre, B. G. McKie, A. O. Medeiros, M. Miliša, Y. Miyake, R. J. Mooney, T. Muotka, J. Nimptsch, R. Paavola, I. Pardo, I. Y. Parnikoza, C. J. Patrick, E. T. H. M. Peeters, J. Pozo, B. Reid, J. S. Richardson, J. Rincón, G. Risnoveanu, C. T. Robinson, A. C. Santamans, G. M. Simiyu, A. Skuja, J. Smykla, R. A. Sponseller, F. Teixeira-de Mello, S. Vilbaste, V. D. Villanueva, J. R. Webster, S. Woelfl, M. A. Xenopoulos, A. G. Yates, C. M. Yule, Y. Zhang, J. A. Zwart, Global patterns and controls of nutrient immobilization on decomposing cellulose in riverine ecosystems. Global Biogeochem Cycles **36**, 1-15 (2022).

20. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. Science **281**, 237-240 (1998).

21. T. H. Huang, Y. H. Fu, P. Y. Pan, C. T. A. Chen, Fluvial carbon fluxes in tropical rivers. Current Opinion in Environmental Sustainability **4**, 162-169 (2012).

22. S. D. Tiegs, J. E. Clapcott, N. A. Griffiths, A. J. Boulton, A standardized cotton-strip assay for measuring organic-matter decomposition in streams. Ecological Indicators **32**, 131-139 (2013).

23. M. Ardón, L. H. Zeglin, R. M. Utz, S. D. Cooper, W. K. Dodds, R. J. Bixby, A. S. Burdett, J. Follstad Shah, N. A. Griffiths, T. K. Harms, S. L. Johnson, J. B. Jones, J. S. Kominoski, W. H. McDowell, A. D. Rosemond, M. T. Trentman, D. Van Horn, A. Ward, Experimental nitrogen and phosphorus enrichment stimulates multiple trophic levels of algal and detrital-based food webs: a global meta-analysis from streams and rivers. Biological Reviews **96**, 692-715 (2020).

24. C. J. LeRoy, A. L. Hipp, K. Lueders, J. J. Follstad Shah, J. S. Kominoski, M. Ardón, W. K. Dodds, M. O. Gessner, N. A. Griffiths, A. Lecerf, D. W. P. Manning, R. L. Sinsabaugh, J. R. Webster, Plant phylogenetic history explains in-stream decomposition at a global scale. Journal of Ecology **108**, 17-35 (2020).

25. K. Yue, P. De Frenne, K. Van Meerbeek, V. Ferreira, D. A. Fornara, Q. Wu, X. Ni, Y. Peng, D. Wang, P. Heděnec, Y. Yang, F. Wu, J. Peñuelas, Litter quality and stream physicochemical properties drive global invertebrate effects on instream litter decomposition. Biological Reviews **97**, 2023-2038 (2022).

26. A. D. Rosemond, C. M. Pringle, A. Ramírez, M. J. Paul, J. L. Meyer, Landscape variation in phosphorus concentration and effects on detritus-based tropical streams. Limnology and Oceanography **47**, 278-289 (2002).

27. L. Boyero, M. A. S. Graça, A. M. Tonin, J. Pérez, A. J. Swafford, V. Ferreira, A. Landeira-Dabarca, M. A. Alexandrou, M. O. Gessner, B. G. McKie, R. J. Albariño, L. A. Barmuta, M. Callisto, J. Chará, E. Chauvet, C. Colón-Gaud, D. Dudgeon, A. C. Encalada, R. Figueroa, A. S. Flecker, T. Fleituch, A. Frainer, J. F. Gonçalves, J. E. Helson, T. Iwata, J. Mathooko, C. M'Erimba, C. M. Pringle, A. Ramírez, C. M. Swan, C. M. Yule, R. G. Pearson, Riparian plant litter quality increases with latitude. Scientific Reports **7**, 1-10 (2017).

28. J. Kattge, S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, G. Bönisch, E. Garnier, M. Westoby, P. B. Reich, I. J. Wright, J. H. C. Cornelissen, C. Violle, S. P. Harrison, P. M. Van Bodegom, M. Reichstein, B. J. Enquist, N. A. Soudzilovskaia, D. D. Ackerly, M. Anand, O. Atkin, M. Bahn, T. R. Baker, D. Baldocchi, R. Bekker, C. C. Blanco, B. Blonder, W. J. Bond, R. Bradstock, D. E. Bunker, F. Casanoves, J. Cavender-Bares, J. Q. Chambers, F. S. Chapin, J. Chave, D. Coomes, W. K. Cornwell, J. M. Craine, B. H. Dobrin, L. Duarte, W. Durka, J. Elser, G. Esser, M. Estiarte, W. F. Fagan, J. Fang, F. Fernández-Méndez, A. Fidelis, B. Finegan, O. Flores, H. Ford, D. Frank, G. T. Freschet, N. M. Fyllas, R. V. Gallagher, W. A. Green, A. G. Gutierrez, T. Hickler, S. I. Higgins, J. G. Hodgson, A. Jalili, S. Jansen, C. A. Joly, A. J. Kerkhoff, D. Kirkup, K. Kitajima, M. Kleyer, S. Klotz, J. M. H. Knops, K. Kramer, I. Kühn, H. Kurokawa, D. Laughlin, T. D. Lee, M. Leishman, F. Lens, T. Lenz, S. L. Lewis, J. Lloyd, J. Llusià, F. Louault, S. Ma, M. D. Mahecha, P. Manning, T. Massad, B. E. Medlyn, J. Messier, A. T. Moles, S. C. Müller, K. Nadrowski, S. Naeem, Ü. Niinemets, S. Nöllert, A. Nüske, R. Ogaya, J. Oleksyn, V. G. Onipchenko, Y. Onoda, J. Ordoñez, G. Overbeck, W. A. Ozinga, S. Patiño, S. Paula, J. G. Pausas, J. Peñuelas, O. L. Phillips, V. Pillar, H. Poorter, L. Poorter, P. Poschlod, A. Prinzing, R. Proulx, A. Rammig, S. Reinsch, B. Reu, L. Sack, B. Salgado-Negret, J. Sardans, S. Shiodera, B. Shipley, A. Siefert, E. Sosinski, J. F. Soussana, E. Swaine, N. Swenson, K. Thompson, P. Thornton, M. Waldram, E. Weiher, M. White, S. White, S. J. Wright, B. Yguel, S. Zaehle, A. E. Zanne, C. Wirth, TRY - a global database of plant traits. Global Change Biology **17**, 2905-2935 (2011).

29. A. Lecerf, E. Chauvet, Intraspecific variability in leaf traits strongly affects alder leaf decomposition in a stream. Basic Applied Ecology **9**, 598-605 (2008).

30. T. Sariyildiz, J. M. Anderson, Decomposition of sun and shade leaves from three deciduous tree species, as affected by their chemical composition. Biology and Fertility of Soils **37**, 137-146 (2003).

31. A. González-Hernández, R. Morales-Villafaña, M. E. Romero-Sánchez, B. Islas-Trejo, R. Pérez-Miranda, Modelling potential distribution of a pine bark beetle in Mexican temperate forests using forecast data and spatial analysis tools. Journal Forestry Research **31**, 649-659 (2018).

32. L. H. Fraser, H. Al Henry, C. N. Carlyle, S. R. White, C. Beierkuhnlein, J. F. Cahill, B. B. Casper, E. Cleland, S. L. Collins, J. S. Dukes, A. K. Knapp, E. Lind, R. Long, Y. Luo, P. B. Reich, M. D. Smith, M. Sternberg, R. Turkington, Coordinated distributed experiments: An emerging tool for testing global hypotheses in ecology and environmental science. Frontiers in Ecology and Environment **11**,147-155 (2013).

33. A. D. Rosemond, J. P. Benstead, P. M. Bumpers, V. Gulis, J. S. Kominoski, D. W. P. Manning, K. Suberkropp, J. B. Wallace, Experimental nutrient additions accelerate terrestrial carbon loss from stream ecosystems. Science **347**, 1142-1145 (2015).

34. M. O. Gessner, E. Chauvet, A case for using litter breakdown to assess functional stream integrity. Ecological Applications **12**, 498-510 (2002).

35. M. C. Jackson, O. L. F. Weyl, F. Altermatt, I. Durance, N. Friberg, A. J. Dumbrell, J. J. Piggott, S. D. Tiegs, K. Tockner, C. B. Krug, P. W. Leadley, G. Woodward, Recommendations for the next generation of global freshwater biological monitoring tools. Advances in Ecological Research **55**, 615-636 (2016)

36. K. A. Wilson, N. A. Auerbach, K. Sam, A. G. Magini, A. S. L. Moss, S. D. Langhans, S. Budiharta, D. Terzano, E. Meijaard, Conservation research is not happening where it is most needed. PLoS Biol **14**, 1-5 (2016).

37. A. N. Strahler, Quantitative analysis of watershed geomorphology. Transactions American Geophysical Union **38**, 913-920 (1957).

38. R. W. McDowell, A. Noble, P. Pletnyakov, L. M. Mosley, Global database of diffuse riverine nitrogen and phosphorus loads and yields. Geoscience Data Journal **8**, 132-143 (2020).

39. D. Ackerman, D. B. Millet, X. Chen, Global estimates of inorganic nitrogen deposition across four decades. Global Biogeochemical Cycles **33**, 100-107 (2019).

40. J. Brahney, N. Mahowald, D. S. Ward, A. P. Ballantyne, J. C. Neff, Is atmospheric phosphorus pollution altering global alpine lake stoichiometry? Global Biogeochemical Cycles **29**, 1369-1383 (2015).

41. N. Mahowald, T. D. Jickells, A. R. Baker, P. Artaxo, C. R. Benitez-Nelson, G. Bergametti, T. C. Bond, Y. Chen, D. D. Cohen, B. Herut, N. Kubilay, R. Losno, C. Luo, W. Maenhaut, K. A. McGee, G. S. Okin, R. L. Siefert, S. Tsukuda, Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. Global Biogeochemical Cycles **22**, 1-19 (2008).

42. B. Lehner, G. Grill, Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. Hydrological Processes **27**, 2171-2186 (2013).

43. R Development Core Team, R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, (2019).

44. L. Breiman, Random forests. Machine Learning **45**, 1-33 (2001) .

45. A. M. Thorn, J. R. Thompson, J. S. Plisinski, Patterns and predictors of recent forest conversion in New England. Land **5**, 1-17 (2016).

Acknowledgments: We are grateful for the efforts of the many people who assisted with the CELLDEX project in the lab and in the field, for Jasmine Mancuso for edits on an earlier version of this manuscript, and Diane Ethaiya for logistical assistance during the CELLDEX project.

Funding: This work was sponsored by an Ecuadorian National Science Foundation PROMETEO award to ST.

Author contributions:

Conceptualization: DC, JPS, KC, ST

Methodology: CP, DC, JPS, KC, ST

Investigation: All co-authors. CL and JFS also provided data from a literature review

Visualization: DC, JPS

Funding acquisition: ST

Writing – original draft: KC, ST

Writing – review and editing: BM, DC, FB, JPS, KC, MG, NG, ST, GW

Competing interests/Disclaimer: Authors declare that they have no competing interests.

This draft manuscript is distributed solely for purposes of courtesy review and comments received will be addressed and treated as appropriate to ensure there is no conflict of interest. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the U.S. Geological Survey (USGS), it does not represent any official USGS finding or policy.

Data and materials availability: All data and code for analyses and figures are available at the public repository *https://github.com/dmcostello/CELLDEX_geospatial*. This repository will be permanently archived at Zenodo and will receive a DOI upon final acceptance.

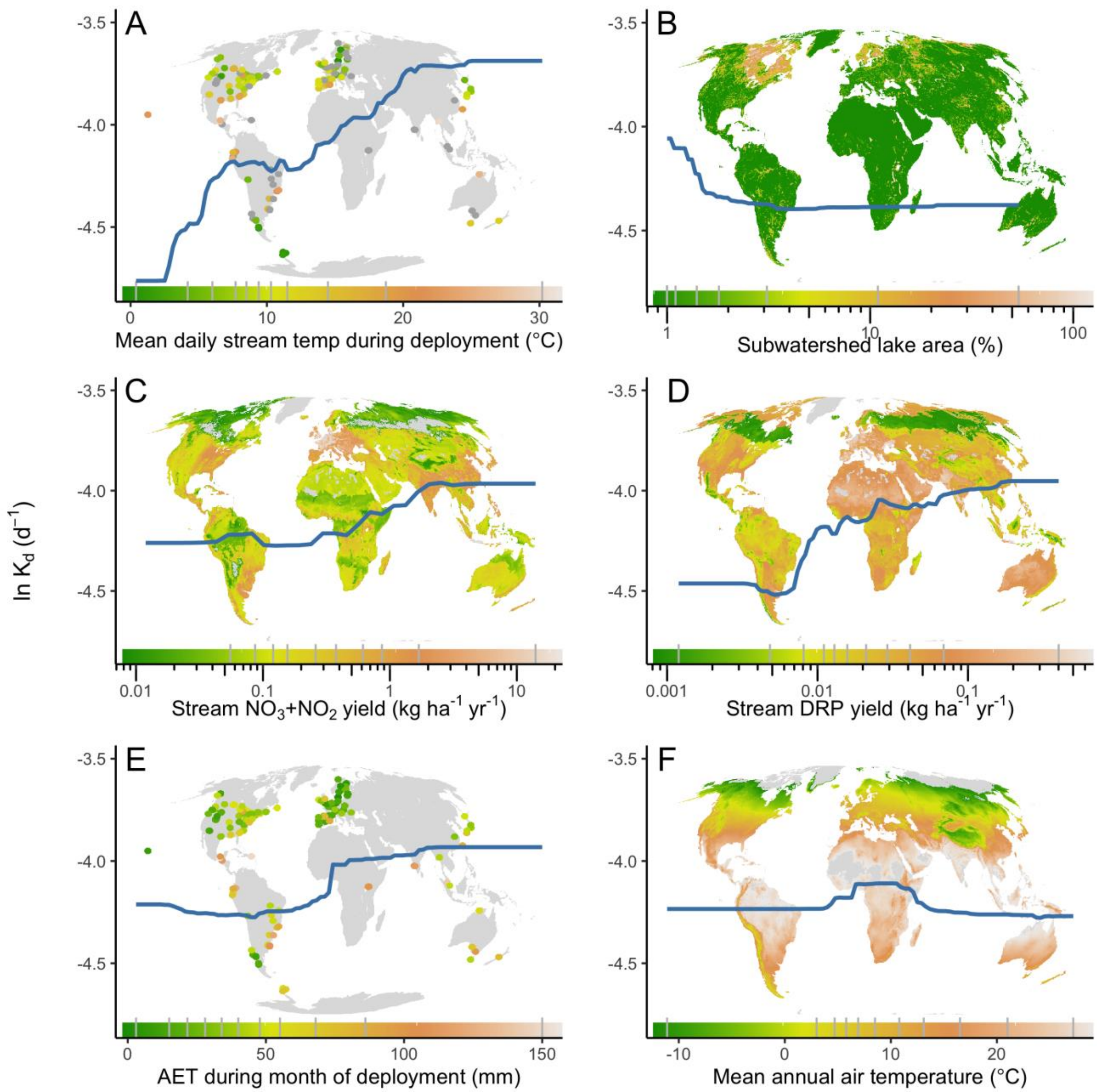

Fig. 1. Partial-dependence plots (blue lines) of the top variables that predict cellulose-decomposition rates (K_d, ln transformed). Background maps show global distributions of predictor variables. The boosted-regression tree model explains 81% of the variance in decomposition rates across the 514 streams used in our study. Most top variables relate to climate and water quality, and effects exhibit non-linear threshold responses. Grey ticks on *x*-axis indicate decile breaks.

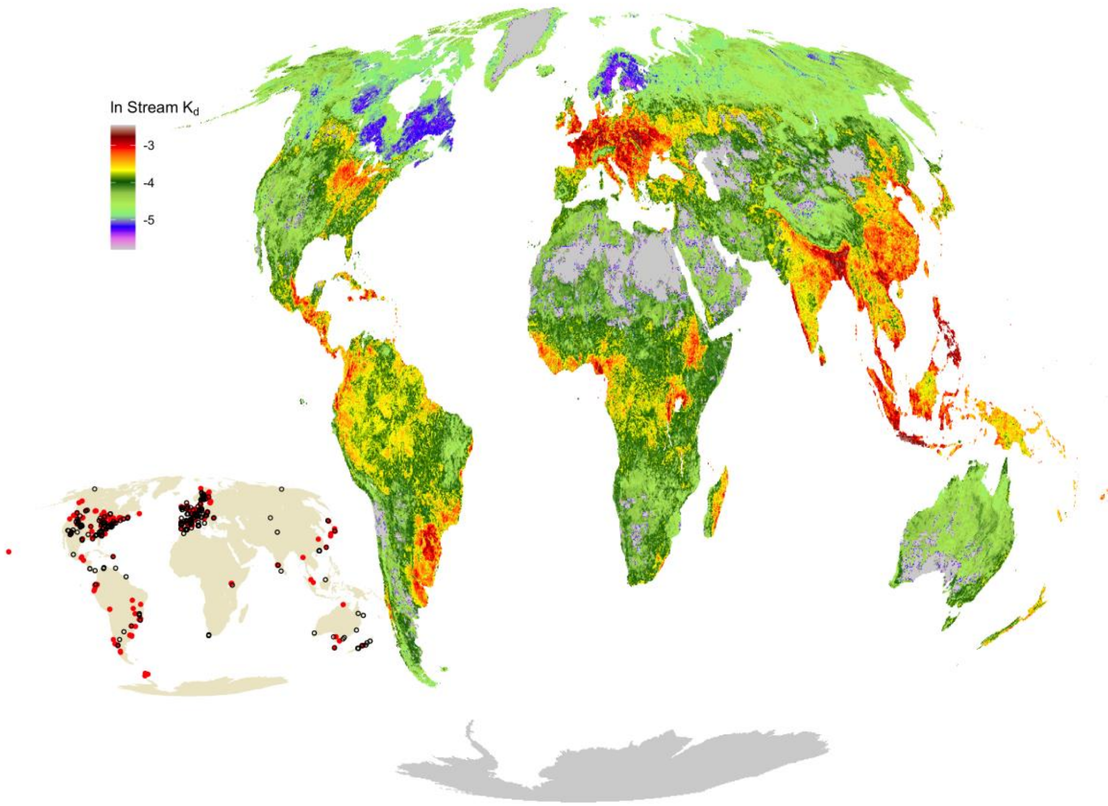

Fig. 2. Predicted mean annual cellulose-decomposition rates (K_d, ln transformed) revealing broad spatial patterns in processing rates. We did not predict K_d for sub-watersheds with ≤10 ha of sub-basin area, nor for Antarctica, which is not included in HydroAtlas. Inset shows study sites for cotton-strip (red circles) and leaf-litter (black circles) decomposition measurements.

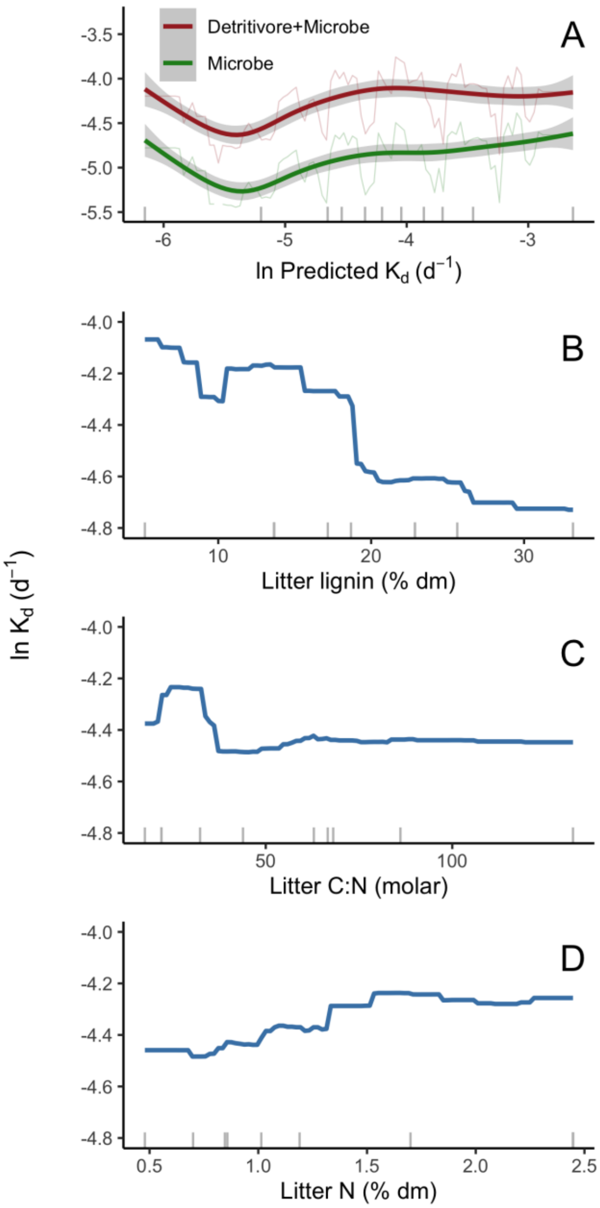

Fig. 3. Partial-dependence plots of the top variables that predict leaf-litter-decomposition rates (K_d, ln transformed). The boosted-regression-tree model explains 70% of the deviance in rates across 895 published values of leaf-litter decomposition and leaf quality (*24*). Top predictors were our modeled cellulose-decomposition rates, invertebrate access to the leaf material, and attributes related to litter quality. Smooth fits (GAM) show the relationship between cellulose-decomposition rate and litter decomposition for the two different common litter-bag mesh sizes that allow or exclude invertebrates. The smooth fits capture the general environmental effects on decomposition, whereas the partial dependency plots (thin lines) are noisier due to covariation in leaf quality and environmental conditions (i.e., certain leaf types are used in certain regions). Grey ticks on *x*-axis indicate decile breaks. Note the change in *y*-axis between panel A and B-C.

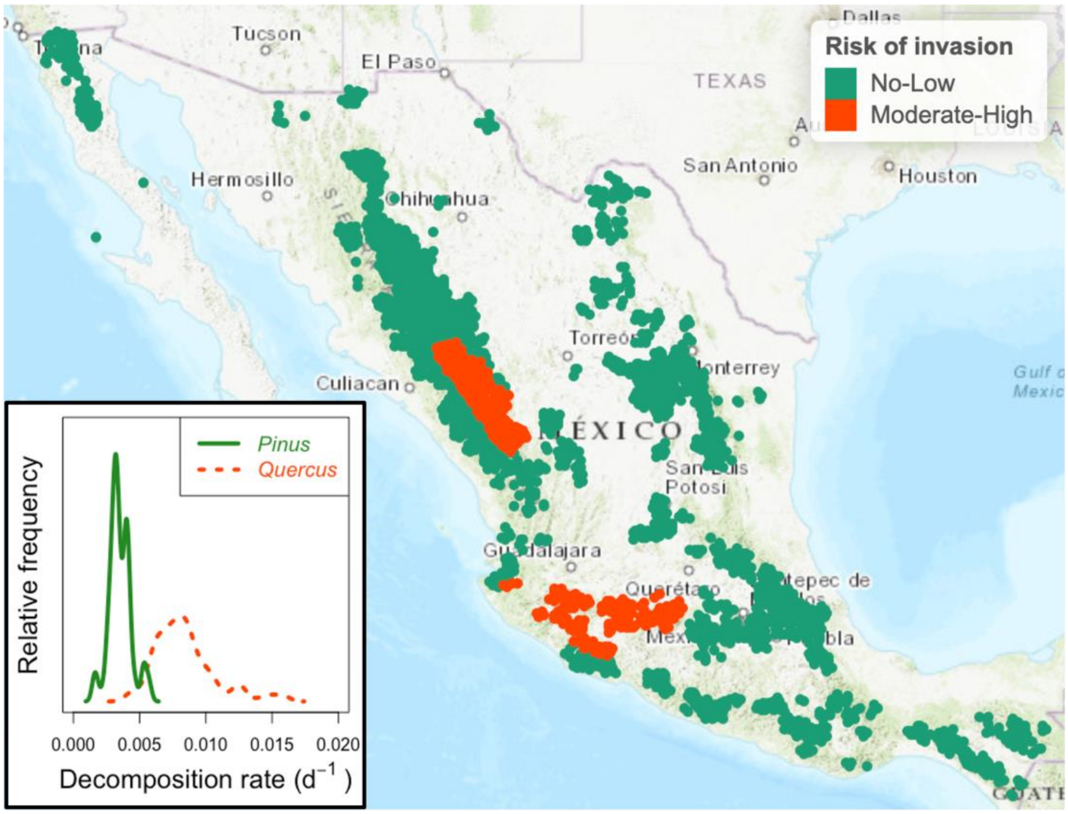

Fig. 4. Distribution of temperate-coniferous forests in Mexico (all points) and locations (orange) where there is a moderate-to-high risk of pine bark beetle (*Dendroctonus mexicanus*) invasion (adapted from (*31*)) that drives a shift from coniferous to deciduous forest. Inset shows the density distribution of predicted litter-decomposition rates for streams in areas of moderate-to-high invasion risk both for pine litter (green solid line) and oak litter (orange dashed line). Our model predicts that full canopy replacement from pine to oak would lead to a doubling of mean leaf-litter decomposition rates (from 0.004 to 0.008 d^{-1}).

Submitted Manuscript: Confidential

Template revised November 2022

Supplementary Materials for

Predicting global organic-matter decomposition in flowing waters

Authors: Scott D. Tiegs[1]*#, Krista A. Capps[2]*#, John Paul Schmidt*[2], David M. Costello*[3], Christopher J. Patrick*[4], Jennifer J. Follstad Shah[5], Carri J. LeRoy[6], Vicenç Acuña[7], Ricardo Albariño[8], Daniel C. Allen[9], Cecilia Alonso[10], Patricio Andino[11], Clay Arango[12], Jukka Aroviita[13], Marcus V.M. Barbosa[14], Leon A. Barmuta[15], Colden Baxter[16], Brent Bellinger[17], Luz Boyero[18], Lyubov Bragina[19], Lee E. Brown[20], Andreas Bruder[21], Denise A. Bruesewitz[22], Francis Burdon[23], Marcos Callisto[24], Antonio Camacho[25], Cristina Canhoto[26], María M. Castillo[27], Eric Chauvet[28], Joanne Clapcott[29], Fanny Colas[30], Checo Colón-Gaud[31], Julien Cornut[21], Verónica Crespo-Pérez[11], Wyatt F. Cross[32], Joseph Culp[33], Michael Danger[34], Olivier Dangles[35], Elvira de Eyto[36], Alison M. Derry[37], Veronica Díaz Villanueva[8], Michael M. Douglas[38], Arturo Elosegi[18], Andrea C. Encalada[39], Sally Entrekin[40], Rodrigo Espinosa[41], Verónica Ferreira[26], Carmen Ferriol[25], Kyla M. Flanagan[42], Alexander S. Flecker[43], Tadeusz Fleituch[44], André Frainer[45,46], Nikolai Friberg[45,47], Paul C. Frost[48], Erica A. Garcia[49], Liliana García-Lago[50], Pavel Ernesto García Soto[51], Mark O. Gessner[52,53], Sudeep Ghate[54], Darren P. Giling[52,55,56], Alan Gilmer[57], José Francisco Gonçalves Jr[58], Rosario Karina Gonzales[59], Manuel A.S. Graça[26], Mike Grace[60], Natalie A. Griffiths[61], Hans-Peter Grossart[52,62], François Guérold[34], Vlad Gulis[63], Pablo E. Gutiérrez-Fonseca[64], Luiz U. Hepp[65], Scott Higgins[66], Takuo Hishi[67], Joseph Huddart[68], John Hudson[69], Samantha Imberger[70], Carlos Iñiguez-Armijos[71], Mark W. Isken[1], Tomoya Iwata[72], David J. Janetski[73], Andrea E. Kirkwood[74], Aaron A. Koning[75], Sarian Kosten[76], Kevin A. Kuehn[77], Hjalmar Laudon[78], Peter R. Leavitt[79], Aurea L Lemes da Silva[80], Shawn Leroux[81], Peter J. Lisi[82], Richard MacKenzie[83], Amy M. Marcarelli[84], Frank O. Masese[85], Peter B. McIntyre[43], Brendan G. McKie[78], Adriana Medeiros[86], Kristian Meissner[13], Marko Miliša[87], Shailendra Mishra[88], Yo Miyake[89], Ashley Moerke[90], Shorok Mombrikotb[68], Rob Mooney[82], Tim Moulton[91], Timo Muotka[92], Junjiro Negishi[93], Vinicius Neres-Lima[94], Mika L. Nieminen[95], Jorge Nimptsch[96], Jakub Ondruch[97], Riku Paavola[92], Isabel Pardo[50], Edwin T.H.M. Peeters[98], Jesus Pozo[18], Aaron Prussian[99], Estefania Quenta[59], Brian Reid[100], John S. Richardson[101], Anna Rigosi[102], José Rincón[103], Geta Risnoveanu[104], Christopher T. Robinson[105], Lorena Rodríguez-Gallego[106], Todd V. Royer[107], James A. Rusak[108], Anna C. Santamans[25], Géza B. Selmeczy[109], Gelas Simiyu[85], Agnija Skuja[110], Jerzy Smykla[44], Ryan Sponseller[111], Kandikere R. Sridhar[112], Aaron Stoler[113], Christopher M. Swan[114], Franco Teixeira-de Mello[115], Jonathan D. Tonkin[116], Sari Uusheimo[13], Allison M. Veach[117], Sirje Vilbaste[118], Lena B.-M. Vought[119], Chiao-Ping Wang[120], Jackson R. Webster[40], Paul B.Wilson[121], Stefan Woelfl[96], Guy Woodward[68], Marguerite A. Xenopoulos[48], Adam G. Yates[122], Chihiro Yoshimura[123], Catherine M. Yule[124], Yixin Zhang[125], Jacob A. Zwart[126]

Affiliations: CELLDEX Consortium. See Supplemental Files for author affiliations

*Corresponding authors: tiegs@oakland.edu, kcapps@uga.edu, *jps@uga.edu*, *dcostel3@kent.edu*, cpatrick@vims.edu

#These authors contributed equally to this work.

Materials and Methods

Data Sources

Cellulose decomposition. We used a global dataset of cellulose-decomposition rates generated by a coordinated field experiment (Cellulose Decomposition Experiment (CELLDEX)) (*18*). Cotton strips were incubated in 514 flowing waters spanning 135 degrees of latitude by a consortium of over 150 peer-sourced researchers globally. Cotton strips are composed of cellulose, the primary constituent of most terrestrially derived leaf litter and the most abundant organic polymer on Earth; as such, cellulose is a plant polymer that is highly relevant for global biogeochemical cycles. Cotton strips were deployed in 2015-16 during periods of peak organic-matter inputs (e.g., autumn in temperate zones, dry season in tropical deciduous forests) at sites relatively free of major anthropogenic impacts (i.e., reference systems). We typically chose stream orders 1-3 (*37*) located in each of Earth's major terrestrial biomes (*18*), and the cellulose-decomposition rate at each stream was summarized as the exponential decay rate (K_d) of tensile-strength loss:

$$K_d=-\ln(T_f/T_i)/t$$

where T_f is the final tensile strength of each cotton strip after incubation in the field, T_i is an average tensile-strength value of control strips not incubated in the field to establish initial tensile strength, and t the field incubation time in days (usually 21-30 days). The loss of tensile strength corresponds to the decomposition of the cotton fabric and is driven mostly by the activity of microbes. Field and laboratory methods are detailed in (*18*, *22*).

Environmental data sources. For data on environmental variables other than *in-situ* stream water temperature, we relied on publicly available datasets with global coverage: 1) (*38*)for estimates of river yields of dissolved reactive phosphorus (kg DRP-P ha^{-1} yr^{-1}) and nitrate+nitrite (kg NO_X-N ha^{-1} yr^{-1}); 2) (*39*) for estimates of nitrogen (N) deposition; 3) (*40*) and (*41*) for estimates of phosphorus (P) deposition that we then interpolated; and 4) (*42*) for data on 96 variables summarized at the 12-digit hydrological scale or for the area upstream (HydroRIVERS: River ATLAS_v10_lev12; HydroBASINS: BasinATLAS_v10_lev12) for either river reaches or corresponding sub-watersheds, though all variables were not populated for all sub-watersheds). We excluded variables from HydroBASINS that were composite measures where we already included confounded variables (e.g., biome, human development index, and human footprint). We recorded temperature data with loggers for a subset (n=360) of the 514 rivers to determine the mean daily temperature of the river water during the cotton-strip incubation period.

Litter decomposition trait data. We used a global dataset of 3,216 unique estimates of litter-decomposition rates (as K_d) for 125 plant genera and multiple experimental conditions (*24*) to independently validate the predictive accuracy of our cellulose-decomposition model. These data are an expanded version of the data published by LeRoy et al. (2020) (see data repository for complete data)(*24*). For each unique river reach sampled in the dataset, we averaged K_d estimates by each unique combination of leaf condition (i.e., leaves picked from the trees while still living or collected from the ground after senescence), plant genus, and direct feeding by detritivorous invertebrates (i.e., coarse-mesh which included invertebrates or fine-mesh litter bags which excluded invertebrates). We excluded any data for which we had 3 or less measurements of decomposition for a genus. The final dataset included 895 unique observations of 35 genera from 559 river reaches. All but 7 estimates of litter decomposition also included mean temperature during deployment, which we included as a predictor variable.

Leaf- and litter-trait data sources. We downloaded 384,252 records from 21,100 plant species and 4,557 genera of leaf traits related to nutrient, micronutrient, and structural compounds for leaves from the TRY plant-trait database (*28*). After filtering for traits describing the chemical constituency of plant leaves that we felt were most relevant for decomposition, the resulting database included average values for 7 traits representing 64 genera. Litter traits were assembled from a literature review of 114 studies comprising 602 litter deployments of 172 genera in rivers. These trait values were joined by genus to the aforementioned empirical data on leaf litter. All genera for which we had litter-decomposition rates had data on either leaf or litter traits, and most included complete values for both. Details on filtering, aggregating, and variable selection as well as full datasets can be found in the data repository.

Data Analysis

Environmental data processing. At each river sampling location in the CELLDEX dataset, we combined temperature recorded during the experiment, extracted values from nutrient yield and deposition rasters, and attributes from HydroBASINS summarized by upstream watershed as well as the containing sub-watershed. For HydroBASINS fields that were additionally available as monthly summaries (e.g., air temperature, potential evapotranspiration, snow coverage), we used both annual summaries and those from the month of deployment at each site as predictors in the BRT model. Variables from HydroBASINS were back-transformed into original units, and predictors with log-normal distributions were $\log_{10}$ transformed. In total, we had 101 predictor variables for our cellulose-decomposition model.

Boosted-regression tree models. We used the gbm package in R to run boosted-regression tree (BRT) models (R Development Core Team, 2019, version 4.2.2)(*43*)which accommodate missing data and allow for complex interactions and non-linear responses. The BRT model was fit with Gaussian distributions, a maximum of 50,000 trees, a shrinkage rate of 0.001, and an interaction depth of 5. The optimal number of trees was selected using cross-validation (20 folds). While BRT models handle variables with large differences in ranges, we ln-transformed K_d to facilitate the interpretation of results. We assessed model predictive accuracy by calculating a pseudo-R^2 for each model and determined variable importance via permutation tests (*44*) (Supplemental Table 1). Predictor variables with importance values greater than $1/n^{\text{variable}}*100$ were included in trees more than would be expected from random chance and identified for further discussion (threshold = 0.99)(*45*).

Output rasters of predicted cellulose-decomposition rates. Using the BRT models and data from the assembled spatial data layers, we predicted river K_d at the extent and at the resolution of the WorldClim rasters (global with 30 arc-second resolution; www.worldclim.org) using the raster package in R. In these output rasters, we did not predict K_d for sub-watersheds with ≤10 ha of sub-basin area, nor for Antarctica, which is not included in HydroATLAS. Importantly, we predicted K_d using a BRT model that included variables measured at each site in the original CELLDEX experiment (i.e., stream temperatures and month of deployment), but those variables were not included in the generation of an interpolated global K_d map.

Independent validation of predicted cellulose-decomposition rates. To validate the predictive model of cellulose decomposition and test the environmental realism of the cellulose substrate, we used BRT to predict published rates of leaf-litter decomposition in rivers (ln K_d). The leaf-litter K_d BRT model included 17 predictor variables, which included cellulose K_d values predicted by the BRT model, 2 experimental conditions (coarse and fine mesh litter bags), 7 genus-level leaf traits (TRY dataset), and 7 genus-level litter traits (literature review). The BRT model was fit using the same model parameters as the cellulose-decomposition model, with cross-validation to optimize the final number of trees.

Data. See repository at https://github.com/dmcostello/CELLDEX_geospatial

Supplemental Table 1. Boosted-regression tree model importance values for cellulose-decomposition rates (ln[K_d])), their description, and the source of data. Importance values greater than 1 were deemed to be important.

Boosted-regression tree - predicting cellulose-decomposition rate (ln[K_d])

Pseudo R^2 0.81

Predictor variable	**Relative importance**	**Description**	**Source**
mean_mean_daily_temp	14.02	Measured water temp	This study
log10lka_pc_sse	6.94	Lake area	HydroBASINS
log10NO3c	6.70	NO_3 yield	McDowell et al. 2021
log10DRPc	4.89	DRP yield	McDowell et al. 2021
AETmonth	4.40	AET month of deployment	HydroBASINS
tmp_dc_uyr	2.48	Air temperature	HydroBASINS

snowmonth	2.26	Snow cover month of deployment	HydroBASINS
tmp_dc_smx	2.25	Air temp maximum	HydroBASINS
soc_th_uav	2.16	Organic C %	HydroBASINS
run_mm_syr	2.10	Runoff	HydroBASINS
crp_pc_sse	2.03	Cropland extent	HydroBASINS
log10dis_m3_pmn	1.94	Discharge minimum	HydroBASINS
gdp_ud_sav	1.71	GDP	HydroBASINS
slp_dg_sav	1.54	Slope	HydroBASINS
pet_mm_uyr	1.44	Potential ET	HydroBASINS
log10sgr_dk_sav	1.30	Stream gradient	HydroBASINS
log10pop_ct_usu	1.29	Population count	HydroBASINS
log10dor_pc_pva	1.29	Degree of regulation	HydroBASINS
tmp_dc_syr	1.28	Air temperature	HydroBASINS
tempmonth	1.28	Air temp month of deployment	HydroBASINS
crp_pc_use	1.24	Cropland extent	HydroBASINS
ele_mt_smx	1.22	Elevation maximum	HydroBASINS
log10dis_m3_pyr	1.09	Discharge average	HydroBASINS
snd_pc_uav	1.09	Sand content	HydroBASINS
log10rdd_mk_uav	1.07	Road density	HydroBASINS
tmp_dc_smn	1.01	Air temp minimum	HydroBASINS
TNdep	0.95	TN deposition	Ackerman et al. 2019
pac_pc_sse	0.90	Protected area extent	HydroBASINS
pre_mm_uyr	0.89	Precipitation	HydroBASINS
aet_mm_uyr	0.88	Actual ET	HydroBASINS
log10gdp_ud_usu	0.88	GDR	HydroBASINS
log10lkv_mc_usu	0.85	Lake volume	HydroBASINS
pre_mm_syr	0.85	Precipitation	HydroBASINS

TPdep	0.79	TP deposition	Brahney et al. 2015 & Mahowald et al. 2008
log10ppd_pk_uav	0.78	Population density	HydroBASINS
cly_pc_sav	0.77	Clay %	HydroBASINS
moist_indexmonth	0.77	Moisture index month of deployment	HydroBASINS
nli_ix_sav	0.70	Nighttime lights	HydroBASINS
ele_mt_sav	0.68	Elevation average	HydroBASINS
ari_ix_uav	0.67	Aridity index	HydroBASINS
soc_th_sav	0.66	Organic C %	HydroBASINS
snw_pc_uyr	0.65	Snow cover	HydroBASINS
log10rev_mc_usu	0.63	Reservoir volume	HydroBASINS
log10gdp_ud_ssu	0.62	GDP	HydroBASINS
PETmonth	0.61	PET month of deployment	HydroBASINS
log10rdd_mk_sav	0.59	Road density	HydroBASINS
slt_pc_sav	0.59	Silt content	HydroBASINS
pac_pc_use	0.58	Protected area extent	HydroBASINS
slp_dg_uav	0.57	Slope	HydroBASINS
for_pc_use	0.56	Forest cover extent	HydroBASINS
gwt_cm_sav	0.55	Groundwater depth	HydroBASINS
log10inu_pc_ult	0.55	Inundation extent	HydroBASINS
ele_mt_uav	0.52	Elevation average	HydroBASINS
cly_pc_uav	0.51	Clay %	HydroBASINS
for_pc_sse	0.50	Forest cover extent	HydroBASINS
log10ria_ha_ssu	0.50	River area	HydroBASINS
precipmonth	0.48	Precipitation month of deployment	HydroBASINS
snd_pc_sav	0.47	Sand content	HydroBASINS

pst_pc_use	0.47	Pasture extent	HydroBASINS
log10pop_ct_ssu	0.45	Population count	HydroBASINS
aet_mm_syr	0.43	Actual ET	HydroBASINS
snw_pc_smx	0.43	Snow cover	HydroBASINS
log10ero_kh_sav	0.38	Erosion rate	HydroBASINS
ari_ix_sav	0.36	Aridity index	HydroBASINS
swc_pc_syr	0.35	Soil water %	HydroBASINS
log10ero_kh_uav	0.35	Erosion rate	HydroBASINS
log10dis_m3_pmx	0.33	Discharge maximum	HydroBASINS
cmi_ix_uyr	0.32	Climate moisture index	HydroBASINS
slt_pc_uav	0.32	Silt content	HydroBASINS
log10ppd_pk_sav	0.31	Population density	HydroBASINS
ele_mt_smn	0.30	Elevation minimum	HydroBASINS
pet_mm_syr	0.29	Potential ET	HydroBASINS
log10riv_tc_usu	0.26	River volume	HydroBASINS
log10lka_pc_use	0.26	Lake area	HydroBASINS
soilwatermonth	0.26	Soil water % month of deployment	HydroBASINS
log10ria_ha_usu	0.21	River area	HydroBASINS
log10riv_tc_ssu	0.20	River volume	HydroBASINS
nli_ix_uav	0.18	Nighttime lights	HydroBASINS
pst_pc_sse	0.17	Pasture extent	HydroBASINS
kar_pc_sse	0.17	Karst extent	HydroBASINS
urb_pc_sse	0.16	Urban extent	HydroBASINS
log10inu_pc_umx	0.16	Inundation extent	HydroBASINS
log10inu_pc_smx	0.14	Inundation extent	HydroBASINS
snw_pc_syr	0.14	Snow cover	HydroBASINS
ire_pc_sse	0.12	Irrigated extent	HydroBASINS

swc_pc_uyr	0.12	Soil water %	HydroBASINS
kar_pc_use	0.12	Karst extent	HydroBASINS
ire_pc_use	0.10	Irrigated extent	HydroBASINS
log10inu_pc_slt	0.10	Inundation extent	HydroBASINS
wet_pc_ug1	0.10	Wetland extent - All types	HydroBASINS
cmi_ix_syr	0.10	Climate moisture index	HydroBASINS
urb_pc_use	0.08	Urban extent	HydroBASINS
log10inu_pc_umn	0.07	Inundation extent	HydroBASINS
wet_pc_ug2	0.07	Wetland extent - No lakes, reservoirs, rivers	HydroBASINS
log10inu_pc_smn	0.06	Inundation extent	HydroBASINS
wet_pc_sg1	0.05	Wetland extent - All types	HydroBASINS
prm_pc_use	0.02	Permafrost extent	HydroBASINS
wet_pc_sg2	0.01	Wetland extent - No lakes, reservoirs, rivers	HydroBASINS
prm_pc_sse	0.00	Permafrost extent	HydroBASINS
gla_pc_sse	0.00	Glacier extent	HydroBASINS
gla_pc_use	0.00	Glacier extent	HydroBASINS

Supplemental Table 2. Boosted-regression tree model importance values for leaf-litter-decomposition rates ($\ln[K_d]$)), their description, and the source of data. Importance values greater than 5 were deemed to be important.

BRT predicting litter bag decomposition rate (ln[kd])

Pseudo R^2 0.70

Predictor variable	**Relative importance**	**Description**	**Source**
ln_pred_k	39.47	Model predicted cotton kd	This study
Mesh.size	20.84	Mesh size	LeRoy 2020
Lignin_Litter_Mn	11.96	Litter lignin content	Literature review
CtoN_Litter_Mn	5.45	Litter C:N	Literature review

N_Litter_Mn	5.23	Litter N content	Literature review
P_Litter_Mn	3.59	Litter P content	Literature review
C_Litter_Mn	2.37	Litter C content	Literature review
Cellulose_Litter_Mn	2.20	Litter cellulose content	Literature review
Ca_Leaf_Mn	2.03	Leaf Ca content	TRY database
NtoP_Leaf_Mn	1.19	Leaf N:P	TRY database
Thick_Mn	1.04	Leaf thickness	TRY database
Leaf.condition	1.03	Leaf condition	TRY database
NtoP_Litter_Mn	0.95	Litter N:P	Literature review
P_Leaf_Mn	0.76	Leaf P content	TRY database
CtoN_Leaf_Mn	0.69	Leaf C:N	TRY database
C_Leaf_Mn	0.66	Leaf C content	TRY database
N_Leaf_Mn	0.53	Leaf N content	TRY database