Overcoming obstacles to sharing data on tree allometric equations

Miguel Cifuentes Jara, Matieu Henry, Maxime Réjou Méchain, Omar R. Lopez, Craig Wayson, José María Michel Fuentes, Edwin Castellanos, et al.

Springer

INRA
SCIENCE & IMPACT

Springer

OPINION PAPER

Overcoming obstacles to sharing data on tree allometric equations

Miguel Cifuentes Jara · Matieu Henry · Maxime Réjou Méchain ·
Omar R. Lopez · Craig Wayson · José María Michel Fuentes ·
Edwin Castellanos · Mauricio Zapata-Cuartas · Daniel Piotto ·
Federico Alice Guier · Héctor Castañeda Lombis · Ruby Cuenca Lara ·
Kelvin Cueva Rojas · Jhon del Águila Pasquel · Álvaro Duque Montoya ·
Javier Fernández Vega · Abner Jiménez Galo · Lars Gunnar Marklund ·
Fabián Milla · José de Jesús Návar Chaidez · Edgar Ortiz Malavassi ·
Johnny Pérez · Carla Ramírez Zea · Luis Rangel García · Rafael Rubilar Pons ·
Laurent Saint-André · Carlos Sanquetta · Charles Scott · James Westfall

1 Introduction

Advances in ecology face the complexity of ecosystems with
dynamics longer than a single scientist's career. In forestry and
REDD+ practice, in particular, our ability to understand forest
ecosystem dynamics and to manage them for mitigation and
adaptation strongly relies on the combination of long-term
research efforts and on data sharing. However, data collected
by many measurement campaigns are regularly lost because
of a lack of capacity to archive and maintain such information.
Much progress would be achieved by encouraging researchers
to provide access to primary data or publish "data-papers"
(Chavan and Penev 2011; Cifuentes Jara et al. 2013; Fady
et al. 2014).

Archiving, sharing, and harmonizing data among re-
searchers allows replication of analyses among researchers
and thus ensures consistency of measurements over time
and, ultimately, measurement accuracy (IPCC 2006). For al-
lometric equations in particular, data sharing avoids the dupli-
cation of expensive and time-consuming field data collection.

M. Cifuentes Jara (✉)
CATIE, Climate Change & Watersheds Program, CATIE 7170,
Cartago, Turrialba 30501, Costa Rica
e-mail: mcifuentes@catie.ac.cr

M. Henry
UN-REDD Programme, Food and Agriculture Organization of the
United Nations (FAO), Viale delle Terme di Caracalla,
00153 Rome, Italy

M. Réjou Méchain
Laboratoire Evolution et Diversite Biologique, UMR 5174 CNRS,
Université Paul Sabatier, 31062 Toulouse, France

O. R. Lopez
INDICASAT-AIP, Edificio 219, Clayton, Ciudad del Saber,
Apartado Postal 0843-03092, Balboa, Ancon, Panamá

C. Wayson
USDA Forest Service, International Programs – SilvaCarbon,
Lima, Perú

J. M. Michel Fuentes
FAO-México, Periférico Poniente 5360 Col. San Juan de Ocotan,
Zapopan, Jalisco, Mexico

E. Castellanos
Universidad del Valle de Guatemala, 18 AV. 11-95 Zona 15,
Guatemala, Guatemala

M. Zapata-Cuartas
SMURFIT KAPPA Cartón de Colombia, Km 15 Carretera Cali,
Yumbo, Colombia

D. Piotto
Instituto Floresta Viva, Av. Litorânea Norte 208, Ilhéus 45651-610,
Brazil

More importantly, it also increases the size of datasets, which directly enhances the quality of the resulting allometric equations in terms of diameter at breast height (DBH) range, goodness of fit indicators, and geographic range where the equations are valid (Chave et al. 2014). Furthermore, the calculation of uncertainty cannot be properly carried out if the original data are not available (Chave et al. 2004; Molto et al. 2013). As robust allometric equations are critical for calculating baseline biomass and carbon stocks for REDD+ and other climate change mitigation initiatives, the need for sharing the best available data becomes more relevant. Here, we identify constraints and propose solutions to facilitate data sharing of allometric equations in forestry research. We analyze which factors limit data sharing among researchers and propose solutions to overcome those limitations.

These recommendations are the result of expert discussions held during the "Regional Technical Workshop on Tree Volume and Biomass Allometric Equations in South and Central America" in Costa Rica, on May 21–24, 2014. The workshop brought together 30 scientists from throughout Latin America, the USA, and Europe to discuss the state of the art on allometric equations, identify knowledge gaps, and offer potential ways forward.

2 Limitations and solutions to strengthening data sharing on allometric equations

We identified uncertainty in intellectual property rules, unclear data use protocols, cultural issues, lack of data sharing structures, insufficient human, technological, and financial capacities, and limited data quality as the main constraints preventing transparent and useful sharing of data related to allometric equations (Table 1). Each of these limitations is unique but may overlap with each other or may prevail differently depending on a country or region. We offer recommendations to overcome these restrictions and facilitate data sharing protocols among researchers.

2.1 Intellectual property

Unless transferred, intellectual property usually belongs to the institution or the project that funded the research and not the people who collected the data. In the context of public research (i.e., conducted through public funding), data should be public but are often not made available by the authors. Intellectual property for data that are collected as part of research funded by the public sector or under the auspice of

F. Alice Guier
Universidad Nacional de Costa Rica, Campus Omar Dengo, Heredia, Costa Rica

H. Castañeda Lombis · A. Jiménez Galo
REDD/CCAD-GIZ, Boulevard Orden de Malta, San Salvador, El Salvador

R. Cuenca Lara · L. Rangel García
Comisión Nacional Forestal (CONAFOR), Periférico Poniente 5360 Col., San Juan de Ocotan, Zapopan, Jalisco, México

K. Cueva Rojas
FAO-Ecuador, Av. Eloy Alfaro y Amazonas. Edificio Ministerio de Agricultura, Ganadería, Acuacultura y Pesca, Quito, Ecuador

J. del Águila Pasquel
Instituto de Investigaciones de la Amazonia Peruana (IIAP), Av. José Abelardo Quiñones km 2.5, Iquitos, Perú

Á. Duque Montoya
Universidad Nacional de Colombia, Calle 59A #63-20, Medellin, Colombia

J. Fernández Vega
Fondo Nacional de Financiamiento Forestal (FONAFIFO), Oficinas Centrales, Moravia, San José, Costa Rica

L. G. Marklund
Universidad de Concepción, Campus Los Ángeles, J.A. Coloma 0201, Los Ángeles, Chile

F. Milla
Instituto Tecnológico de Costa Rica, Apartado 159-7050, Cartago, Costa Rica

J. d. J. Návar Chaidez
FAO-SLM, Cuidad del Saber, ED 238, Ciudad de Panamá, Panamá

E. Ortiz Malavassi
CIIDIR-IPN Unidad Durango, Sigma # 119 Fracc. 20 de Noviembre 11, Durango, DGO 34220, México

J. Pérez
Escuela Nacional de Ciencias Forestales (ESNACIFOR), Colonia las Américas, Siguatepeque, Comayagua, Honduras

C. Ramírez Zea
FAO-PERÚ, Manuel Oleachea 414, Miraflores, Lima, Perú

R. Rubilar Pons
Universidad de Concepción, Víctor Lamas 1290, Concepción, VIII Región del Biobío, Chile

L. Saint-André
INRA, UR1138, Unité Biogéochimie des Ecosystèmes Forestiers (BEF), Centre INRA de Nancy, 54280 Champenoux, France

L. Saint-André
CIRAD, UMR ECO&SOLS, 34000 Montpellier, France

C. Sanquetta
Federal University of Paraná, Ave.Lothario Meissener 900, Jardin Botánico, Brazil

C. Scott · J. Westfall
US Forest Service, 11 Campus Blvd., Newtown Square, PA, USA

O. R. Lopez
Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Balboa, Ancon, República de Panamá

Table 1 Limitations and proposed solutions to promote sharing data on tree allometric equations

Identified limitation	Impacts	Solutions
Intellectual property	Hinders collaboration and reduces data availability.	Encourage the clarification of intellectual property rules and establishing incentives for data sharing.
Data use policy	Work undertaken by owner(s) and author(s) is not properly recognized. Creates mistrust among parties involved.	Promote agreements that clarify what the data will be used for and what constitutes "fair use." Journals should request a statement regarding data provenance.
Lack of culture exposure to collaborative environments	Not sharing data negatively affects professional performance. Reduces equation availability.	Implement guidelines and encourage actions to foster collaboration and clarify authorship issues. Promote publication of equations and data sharing through regional workshops that create international collaborations.
Data sharing structures	Incompatible data sets and standards limit scaling out and up. Reduces equation availability, utilization, and recognition of local scientist efforts.	Develop frameworks for data sharing that includes clear rules, standardized procedures, formats, platforms, and funding streams for data sharing. Promote the use and implementation of regional or global database networks with reliable availability structures and clear use agreements.
Sustainability of efforts	Limits data collection, analysis, quality assurance and control, valorization, and sharing. Decreases local, national, or regional contributions.	Establish national and institutional policies to encourage data sharing and provide funding incentives for research and publication of local allometric equations.
Data quality and completeness	Uncertainty of estimates remains unresolved. Precludes researchers from drawing broader inferences from larger, pooled datasets.	Standards and tools for quality control need to be provided to facilitate quality assurance and control procedures.

government projects needs to be further clarified and rules established. When private institutions collect data, access can be more complicated and may necessitate the exchange of financial services. In the context of private data, national policies can support data sharing with national research institutions.

The intellectual property for the data collected by public and private, and national and international projects needs to be clarified. In the case of public research, it is necessary to inform authors and clarify data ownership and to ensure the collection and archiving of data. Ideally, this needs to be done at project inception, with clear rules laid out and made public. National capacities on intellectual property rules need to be enhanced adequately. Private sector data, guarded by commercial interests, may be pooled with other similar data sources and be eventually shared without critical proprietary information. Legal agreements, contracts, or letters of understanding (which may or may not include funding) may facilitate data sharing and clarify data ownership. Offering the opportunity to engage as coauthor in any potential publications may offer additional incentives for sharing data.

2.2 Data use policies

Data can be used for different purposes including commercial activities. It is important to clarify, at project inception, how and for what the data will be used for to ensure that the work undertaken by the owner(s) and the author(s) is acknowledged and/or compensated. When sharing data, owners, authors, and users have to mutually agree on the definition and terms of what constitutes "fair use." We recommend agreements that are drafted and accepted among relevant parties previous to data sharing. These agreements should at least clearly spell out the duration of the agreement, specific expectations of authorship, rules of fair use, options for third party sharing, and steps for resolving potential disputes.

2.3 Lack of cultural exposure to collaborative environments

Although a culture of collaboration seems to be developing worldwide, scientists are still not used to sharing their data, mainly because of the intense competition between institutions and/or individuals. Data owners are afraid to be excluded from publications resulting from the data they share. This affects their professional performance because, in the research community, they are principally evaluated by the number of publications they produce. In the field of allometric equations, scientists fear that data are used to develop new models with no real improvement over the original ones. Also, authors and data owners are afraid of losing the exclusive opportunities to answer further questions that they want to address in future works.

INRA SCIENCE & IMPACT ⌂ Springer

To develop a culture of data sharing and to enhance the recognition of researchers' work, collaboration between research institutions and between authors should be encouraged and facilitated to allow appropriate use of all results (not only published ones) in a transparent and timely manner. Collaborations are also needed to facilitate technical, financial, and human support and a more equitable allocation of resources and exploitation of data collected. Data sharing agreements between researchers need to be facilitated and implemented. Data owners tend to think that data sharing is a "win or lose" situation most of the time, i.e., sharing leads to loss of ownership while not sharing is necessary to retain ownership. However, as shown by the creative common license agreement (http://creativecommons.org/licenses/), there is a wide range of possible agreements and negotiations among data owners and users. These types of agreements should be encouraged to promote "win-win" situations where parties find it advantageous to share their data. Potentially useful guidelines to break through the cultural barrier, foster collaboration, and clarify authorship issues are presented by Fine and Lawrence (1993), Dreyfuss (2000), and Weltzin et al. (2006).

2.4 Data sharing structures

Scientists share their data using their own networks and database formats, some of which have been agreed upon within scientific working groups or research networks. However, beyond those groups, there are no rules, standards, or expectations for data sharing. There is a clear need to develop a broadly applicable framework for data sharing that includes clear rules and standardized procedures, formats, and even platforms for data sharing. This framework needs to be trusted by authors and owners alike, and may also include sharing of analysis scripts and the summary data that results from those scripts (Sist et al. 2015).

A structure for data sharing should have standards for data integrity as well as the actual software and tools used to archive and share data. Several networks have been developed to support data sharing agreements for forest inventory data (Lopez-Gonzalez et al. 2011) and for plant traits (Kattge et al. 2011a, b). These networks and structures may include access to metadata, which is a good FJluy 201 first step forward. However, although authors and owners outside those networks are encouraged to share their metadata, they are challenged by the absence of clearly defined platforms to publish them. The increasing availability of cloud storage and readily available online database software and tools may soon overcome such technological barrier. A network and structure for data sharing must be developed for tree allometric equations as well. The GlobAllomeTree platform (http://www.globallometree.org/) is a first step in that direction, allowing researchers to access available tree allometric equations and compare their equations with existing ones. An expanded version of the platform (due July 2014) will allow authors to share their data, including wood-specific gravity data and biomass expansion factors. We hope this platform will grow even further to address the needs outlined here and provide authors and data owners a hub that promotes collaboration and the advancement of science in the area of tree allometry. However, resources are needed to implement data sharing protocols and other platforms and to spread the common language across research communities.

2.5 Sustainability of efforts

Insufficient technical, human, and financial capacities may limit data collection, analysis, quality assurance and control, valorization, and sharing. For example, not all actors have the financial or the scientific capacity to publish their work; as many as 47 % of all allometric equations are only reported in gray literature (Cifuentes Jara et al. 2013). A first step to remedy this situation would be that national policies encourage scientific publishing and data sharing and eventually provide incentives for it. First, national research councils, ministries for science and technology, and similar government entities in charge of science need to earmark funds in research grants to finance the publication of research findings. Universities then need to recognize efforts that promote innovation and knowledge generation and reward researchers appropriately; unfortunately, data sharing is still not part of this recognition. While data sharing structures are being constructed, the parties involved should address current and future capacity deficiencies and the funding needed for long-term permanence and the continued technical feasibility of these efforts.

2.6 Data quality and completeness

We need to ensure that data quality and completeness is adequate among studies. However, there are no agreements upon standards or good practice guidelines that would facilitate that process (but see Cifuentes Jara et al. 2013, this issue). Individual researchers have complete independence as to the amount of details they may share about their data. Data and metadata reporting often lacks quality assurance and is rarely addressed during the peer-review process. There are documented instances of published allometric equations being improperly transcribed or misquoted, and inadequate descriptions of the tree components included in a given equation (Cifuentes Jara et al. 2013).

Standardized procedures for collecting field data to develop allometric equations should be readily available (Picard et al. 2012). Also, tools for quality control need to be provided to facilitate quality assurance and control procedures. The IPCC emissions factor database (http://www.ipcc-nggip.iges.or.jp/EFDB/main.php), for example, allows researchers to access emission factors and compare them to their results. Having

similar tools available for allometric equations would increase the impact factor of journals each time data are used by another article, increasing visibility, and recognition of the work undertaken. Authors would thus be encouraged to publish their results, including the raw data, which has the added advantage of making the calculation of uncertainties possible.

3 Conclusion

Several obstacles prevent widespread dissemination and sharing of tree allometric equation data. Some are cultural, while others are legal and technological in nature; all can be overcome with varying degrees of ease. Beyond the mechanics of data storage, metadata, and quality control/assurance protocols, our key message is that a change in culture promoted by evolving scientific practices is needed among governments, institutions, and researchers such that data sharing and collaborations are actively sought. We are encouraged by the increasing number of emerging collaborative networks as our suggestions are meant to facilitate the enabling environment necessary for their further development. FLUXNET (http://fluxnet.ornl.gov) is an example where a group of researchers, backed with clear data policies, metadata, quality assurance and control, and a platform for sharing data have already been able to advance the understanding of carbon cycling at a rate much faster than if they were working in an isolated manner. This is consistent with an ongoing trend in ecology, genetics, zoology, and botany, where the greater scientific impact of collaborative efforts is increasing (Nabout et al. 2015).

A long-term goal for researchers developing tree allometric equations should be to create an online repository to store raw data and metadata for allometric equations. Such a repository should allow management of data sharing among users. More importantly, a data platform such as GlobAllomeTree will be the "seed" to encourage the culture, legal aspects, and technologies associated with ideal data sharing processes. For starters, a reliable framework for the equitable sharing of data is necessary for that platform to work. Also, it is crucial that data authorship, ownership, and use of the data are clarified for such a sharing system to work. In addition, the data repository would promote the establishment of a network of users and facilitate collaboration among researchers (i.e., responding to a "build it and they will come" philosophy). Furthermore, governments can have a centralized location where the most current equations for calculating carbon stocks can be found, and donors could more easily identify funding needs.

References

Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinf 12:S2

Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R (2004) Error propagation and scaling for tropical forest biomass estimates. Philos T R Soc B 359:409–420

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WBC, Duque A, Eid T, Fearnside PM, Goodman RC, Matieu H, Martínez-Yrízar A, Mugasha WA, Muller-Landau HC, Mencuccini M, Nelson BW, Ngomanda A, Nogueira EM, Ortiz-Malavassi E, Pélissier R, Ploton P, Ryan CM, Saldarriaga JG, Vieilledent G (2014) Improved pantropical allometric models to estimate the above ground biomass of tropical forests. Glob Chang Biol 20:3177–3190

Cifuentes Jara M, Henry M, Morales D (2013) Inventory of volume and biomass tree allometric equations for Central and South America, UN-REDD MRV report 11. CATIE, Turialba, Costa Rica, Food & Agriculture Organization of the United Nations, Rome, Italy

Dreyfuss RC (2000) Collaborative research: conflicts on authorship, ownership, and accountability. Vanderbilt Law Rev 53:1162

Fady B, Benard A, Pichot C, Peiffer M, Leban JM, Dreyer E (2014) The open data debate: a need for accessible and shared data in forest science. Ann For Sci 71:523–525. doi:10.1007/s13595-014-0375-3

Fine MA, Lawrence AK (1993) Reflections on determining authorship credit and authorship order on faculty-student collaborations. Am Psychol 48:1141–1147. doi:10.1037/0003-066X.48.11.1141

IPCC (2006) 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Simon Eggleston, Leandro Buendia, Kyoko Miwa, Todd Ngara, Kiyoto Tanabe (eds.). Institute for Global Environmental Strategies (IGES), Hayama, Japan

Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ, Cornelissen JHC, Violle C, Harrison SP, Van Bodegom PM, Reichstein M, Enquist BJ, Soudzilovskaia NA, Ackerly DD, Anand M, Atkin O, Bahn M, Baker TR, Baldocchi D, Bekker R, Blanco CC, Blonder B, Bond WJ, Bradstock R, Bunker DE, Casanoves F, Cavender-Bares J, Chambers JQ, Chapin Iii FS, Chave J, Coomes D, Cornwell WK, Craine JM, Dobrin BH, Duarte L, Durka W, Elser J, Esser G, Estiarte M, Fagan WF, Fang J, FernáNdez-Méndez F, Fidelis A, Finegan B, Flores O, Ford H, Frank D, Freschet GT, Fyllas NM, Gallagher RV, Green WA, Gutierrez AG, Hickler T, Higgins SI, Hodgson JG, Jalili A, Jansen S, Joly CA, Kerkhoff AJ, Kirkup D, Kitajima K, Kleyer M, Klotz S, Knops JMH, Kramer K, Kühn I, Kurokawa H, Laughlin D, Lee TD, Leishman M, Lens F, Lenz T, Lewis SL, Lloyd J, Llusià J, Louault F, Ma S, Mahecha MD, Manning P, Massad T, Medlyn BE, Messier J, Moles AT, Müller SC, Nadrowski K, Naeem S, Niinemets U, Nöllert S, Nüske A, Ogaya R, Oleksyn J, Onipchenko VG, Onoda Y, Ordoñez J, Overbeck G, Ozinga WA, Patiño S, Paula S, Pausas JG, Peñuelas J, Phillips OL, Pillar V, Poorter H, Poorter L, Poschlod P, Prinzing A, Proulx R, Rammig A, Reinsch S, Reu B, Sack L, Salgado-Negret B, Sardans J, Shiodera S, Shipley B, Siefert A, Sosinski E, Soussana JF, Swaine E, Swenson N, Thompson K, Thornton P, Waldram M, Weiher E, White M, White S, Wright SJ, Yguel B, Zaehle S, Zanne AE, Wirth C (2011a) TRY—a global

database of plant traits. Glob Chang Biol 17:2905–2935. doi:10.1111/j.1365-2486.2011.02451.x

Kattge J, Ogle K, Bönisch G, Díaz S, Lavorel S, Madin J, Nadrowski K, Nöllert S, Sartor K, Wirth C (2011b) A generic structure for plant trait databases. Method Ecol Evol 2:202–213. doi:10.1111/j.2041-210X.2010.00067.x

Lopez-Gonzalez G, Lewis SL, Burkitt M, Phillips OL (2011) ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. J Veg Sci 22:610–613

Molto Q, Rossi V, Blanc L (2013) Error propagation in biomass estimation in tropical forests. Method Ecol Evol 4:175–183. doi:10.1111/j.2041-210x.2012.00266.x

Nabout JC, Parreira MR, Teresa FB, Carneiro FM, da Cunha HF, Ondei LS, Caramori SS, Soares TN (2015) Publish (in a group) or perish (alone): the trend from single- to multi-authorhip in biological papers. Scientometrics 102:357–364

Picard N, Saint André L, Henry M (2012) Manual for building tree allometric equations: from the field to the prediction. Food and Agriculture Organization of the United Nations, Centre de Coopération Internationale en Recherche Agronomique

Sist P, Rutishauser E, Peña-Claros M, Shenkin A, Hérault B, Blanc L, Baraloto C, Baya F, Benedet F, Emidio da Silva K, Descroix L, Nunes Ferreira J, Gourlet-Fleury S, Carneiro Guedes M, Bin Harun I, Jalonen R, Kanashiro M, Krisnawati H, Kshatriya M, Lincoln P, Mazzei L, Medjibé V, Nasi R, d'Oliveira MVN, de Oliveira LC, Picard N, Pietsch S, Pinard M, Priyadi H, Putz FE, Rodney K, Rossi V, Roopsind A, Ruschel AR, Shari NHZ, Rodrigues de Souza C, Herry Susanty F, Doff Sotta E, Toledo M, Vidal E, West TAP, Wortel V, Yamada T (2015) The Tropical managed Forests Observatory: a research network addressing the future of tropical logged forests. Appl Veg Sci 18:171–174

Weltzin JF, Belote RT, Williams LT, Keller JK, Engel EC (2006) Authorship in ecology: attribution, accountability, and responsibility. Front Ecol Environ 4:435–441. doi:10.1890/1540-9295(2006)4[435:aieaaa]2.0.co;2