

### Tropical Forest Carbon database (TropForC-db)

**Source:** Anderson-Teixeira, K.J. et al. 2016. Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC). *Global Change Biology*.

#### Database description:

The Tropical Forest C database (TropForC-db) is the tropical component of ForC-db, which is a global C database that is currently under development and maintained at <https://github.com/forc-db>, where a more recent version may be available. Its structure is derived from that of BETY-db ([www.betydb.org](http://www.betydb.org); LeBauer *et al.*, 2010). In brief, the database consists of a series of cross-referenced data tables describing (1) sites, (2) plots and their history, (3) measurements of C cycle variables, (4) variables, (5) disturbance/history event type, (6) plant functional types (PFTs)/ species, (7) methodology, and (8) allometries. We obtained data from multiple previous compilations and individual published studies, as detailed in Anderson-Teixeira et al. (2016). Original citations associated with all data are given in the measurements table (3), and full citations are in the file “TropForC\_bibliography.pdf”. Sample R code, showing how to load the data files into R and join the tables, is provided in a separate file, named “TropForC\_Rcode.txt”.

Throughout the database, missing data were coded to indicate the reason for the missing values (Table 1).

#### Overview of database structure and data files:

| Data file  | Description  | Links to other data files   |
|--|--|---|
| (1) ‘ <b>sites</b> ’<br>TropForC_sites.csv               | Geographic, climatic, and edaphic site data                                | (2) and (3)   |
| (2) ‘ <b>plothistory</b> ’<br>TropForC_plothistory.csv   | Known history of each plot or set of replicate plots                       | (1), (3), and (5)   |
| (3) ‘ <b>measurements</b> ’<br>TropForC_measurements.csv | Records of ecosystem-level measurements relevant to C cycling              | (1), (2), (4), and (6)  |
| (4) ‘ <b>variables</b> ’<br>TropForC_variables.csv       | Definitions of C cycle variables and covariates.                           | Defines codes for columns ‘variables.name’, ‘covariate_1’, ‘covariate_2’, and ‘covariate_3’ in (3). |
| (5) ‘ <b>disttype</b> ’<br>TropForC_disttype.csv         | Definition of disturbance, management or regeneration history event types. | Defines codes for column ‘disttype’ in (2)  |
| (6) ‘ <b>pft</b> ’<br>TropForC_pft.csv                   | Definitions of plant functional codes.                                     | Defines codes for columns ‘acceptedsymbol’ and ‘dominantveg’ in (3)                                 |
| (7) ‘ <b>methodology</b> ’<br>TropForC_methodology.csv   | Description of methodologies.  | Defines codes for column ‘method_id’ in (3)   |
| (8) ‘ <b>allometry</b> ’<br>TropForC_allometry.csv       | Sources and description of allometric equations                            | Defines codes for the covariate ‘allometric_equation’ in (3)  |

**Data File Contents:**

(1) TropForC\_sites.csv - ‘sites’

| Column              | Description  | Units |
|---------------------|--|-------|
| siteID              | Unique numerical identifier for each record in the sites table.  | -     |
| sites.sitename      | Site identifier, sufficient to uniquely identify the site within the paper. Links to <b>plothistory</b> and <b>measurements</b> tables.  | -     |
| city                | Nearest city   | -     |
| state               | State  | -     |
| country             | Country  | -     |
| lat                 | Latitude in decimal form   | -     |
| lon                 | Longitude in decimal form  | -     |
| masl                | Elevation (meters above sea level) (reported)  | m     |
| mat                 | Mean annual temperature (reported)   | °C    |
| min.temp            | Mean temperature of the coldest month (reported)   | °C    |
| max.temp            | Mean temperature of the warmest month (reported)   | °C    |
| map                 | Mean annual precipitation (reported)   | mm    |
| soil.texture        | Categorical variable describing soil texture, based on % sand-silt-clay. When original publication reports % sand-silt-clay, classification follows USDA soil classification.  | -     |
| soil.classification | Soil classification (USDA soil taxonomy or FAO soil classification)  | -     |
| soilnotes           | Soil details not included above  | -     |
| sitenotes           | Site details not included above  | -     |
| siteref             | Additional references used to obtain site data.  | -     |
| area                | Groups sites that are geographically proximate and edaphically similar. Defined as a group of sites where no site is removed from the rest in its group by >0.25 degrees latitude or longitude.  | -     |
| biogeog             | Biogeographical region, extracted from map of Olson <i>et al.</i> (2001)   | -     |
| Koeppen             | Köppen-Geiger classification, extracted from the ESRI Köppen-Geiger map ( <a href="http://www.arcgis.com/home/item.html?id=7a53584fa55643df969f93cec83788e1">http://www.arcgis.com/home/item.html?id=7a53584fa55643df969f93cec83788e1</a> ). | -     |
| FAOecozone          | FAO global ecological zones classification, extracted from FAO's GeoNetwork ( <a href="http://www.fao.org:80/geonetwork">http://www.fao.org:80/geonetwork</a> ).   | -     |

(2) TropForC\_plothistory.csv - '**plothistory**'

| Column                | Description   | Units |
|-----------------------|---|-------|
| plothistoryID         | Unique numerical identifier for each record in the history table.   | -     |
| <b>sites.sitename</b> | Unique site identifier, sufficient to identify the site within the original publication. Links to <b>sites</b> and <b>measurements</b> tables.                                      | -     |
| <b>plot.name</b>      | Unique plot name, sufficient to identify the plot within the original publication.  | -     |
| plotarea              | Area of plot or combined area of replicate plots.   | ha    |
| date                  | Date of plot history event.   | -     |
| dateloc               | Level of confidence in date (codes given in Table 2).   | -     |
| <b>distcat</b>        | Plot history event categories, namely 'Disturbance' (includes natural and anthropogenic disturbances, management), 'Establishment', 'No.info' (data not available), and 'Regrowth'. | -     |
| <b>disttype</b>       | Plot history event types. See <b>disttype</b> table for code definitions.   | -     |
| level                 | Intensity/frequency/amount applied of history event type, if available.   | -     |
| units                 | Level units.  | -     |
| distnotes             | History event details not included above.   | -     |

## (3) TropForC\_measurements.csv - ‘measurements’

| Column         | Description  | Units |
|----------------|--|-------|
| measurementID  | Unique identifier for each record in the measurements table.   | -     |
| sites.sitename | Unique site identifier name for the site at which the measurement was made. Links to <b>sites</b> and <b>plothistory</b> tables.   | -     |
| plot.name      | Unique plot identifier name for the plot in which the measurement was made. Links to <b>plothistory</b> table.   | -     |
| stand.age      | Age of stand in years as reported in the original publication or calculated based on the date of initiation of forest regrowth. When the publication reports a range of ages, the mean is recorded. For primary/old-growth/mature/ intact stands, the stand.age is recorded as “999”.  | -     |
| dominantveg    | Code used to identify plant functional trait of dominant vegetation ( $\geq 80\%$ dominance, if given). Identical to <b>acceptedsymbol</b> for multi-species forests. For plantations, this is the appropriate <b>pftcode</b> based on the planted species.  | -     |
| scientificname | Genus and species for single dominant species, if applicable (plantations only).   | -     |
| acceptedsymbol | Acronym for plant functional type (pft; multi-species stands) or species listed in scientificname (monospecific stands—i.e., plantations). Plant functional type code, <b>pftcode</b> , is as defined in the the <b>pft</b> table. For single species, species name acronyms are as listed in the USDA Plants Checklist ( <a href="http://plants.usda.gov/dl_all.html">http://plants.usda.gov/dl_all.html</a> ), or listed as ‘nocode’ if not available for the species. | -     |
| Regentype      | Regeneration type of plot, as follows:<br>natcult – Natural regeneration following cultivation<br>natgrazed – Natural regeneration following grazing<br>natother – Natural regeneration following other forms of disturbance<br>plantation – Plantation/regeneration by planting or sowing<br>nodata – Missing data or none of the above   | -     |
| UTBE           | Indicates if the plot is Unmanaged Tropical Broadleaf Evergreen (“utbe”) or otherwise (“not utbe”).  | -     |
| variables.name | Code name for variable sampled. Codes defined in <b>variables</b> table.   | -     |
| method_id      | Code number for methodology. Codes defined in <b>methodology</b> table.  | -     |
| date           | Date on which measurement was made. Usually found as the year of measurement in the methods section, or, when unavailable, approximated as the year before publication.  | -     |
| dateloc        | Level of confidence in date. See Table 2 for codes behind level of confidence.   | -     |
| start_date     | Date on which measurements were started.   | -     |
| start_dateloc  | Level of confidence in date. See Table 2 for codes behind level of confidence.   | -     |
| end_date       | Date on which measurements ended.  | -     |
| end_dateloc    | Level of confidence in date. See Table 2 for codes behind level of confidence.   | -     |

| Column           | Description  | Units                         |
|------------------|--|-------------------------------|
| mean             | Mean value of measurement recorded. Units dependent on <b>variables.name</b> . See <b>variables</b> table for the associated units.  | as defined in variables table |
| n                | Number of experimental replicates used to estimate mean and statistical summary.   | -                             |
| statname         | Name of reported statistic (SE- standard error; SD- standard deviation; 95%CI- 95% confidence interval).   | -                             |
| stat             | Value of reported statistic.   | -                             |
| notes            | Details of study.  | -                             |
| covariate_#      | Code name for covariate associated with the variable being measured. Codes defined in <b>variables</b> table.  | -                             |
| coV#_value       | Value of reported covariate.   | -                             |
| coV#_units       | Units of reported covariate.   | -                             |
| dupcode          | Marks entries that meet one of the following criteria: <ul style="list-style-type: none"> <li>- Duplicate estimates of the same variable but with different values (<b>code D</b> for duplicate)</li> <li>- Replicate sampling of the same variables within the same study (<b>code R</b> for replicate)</li> <li>- Multiple estimates of the same variables within the same study but using different methods (<b>code M</b> for different measurement methods)</li> <li>- Does not have a duplicate (<b>code 0</b> for no duplicates)</li> </ul> | -                             |
| dupnum           | Assigns a number to every entry in each set of duplicates/replicates, in chronological order according to the citation.year. If two sources were published in the same year, the primary source (as opposed to a synthesis) comes first.   | -                             |
| citations.doi    | Digital object identifier unique to publication, if available.   | -                             |
| citations.author | First author's last name of publication.   | -                             |
| citations.year   | Year of publication.   | -                             |
| citations.title  | Complete title of publication.   | -                             |
| loaded_from      | Reference to data compilations from which data and reference to primary source were obtained, if applicable.   | -                             |

(4) TropForC\_variables.csv - ‘**variables**’

| Column                | Description   | Units |
|-----------------------|---|-------|
| variables.type        | Variable type, where “primary” refers to a carbon cycle variable, “secondary” refers to a variable recorded because it can influence carbon cycle variables, and “covariate” refers to a variable required to interpret carbon cycle variables. | -     |
| <b>variables.name</b> | Code name for variable sampled. Used in <b>measurements</b> table.  | -     |
| units                 | Units of the variable.  | -     |
| description           | Definition and notes about the variable.  | -     |
| equations             | Equations describing relationships among variables. Parentheses indicate variables that are not included in the current database version.   | -     |
| notes                 | Notes of relevance to interpreting variables.   | -     |
| Associated.covariate  | Covariates associated with the variable listed. The codes for these covariates are also defined in this table.  | -     |

(5) TropForC\_disttype.csv - ‘**disttype**’

| Column          | Description   | Units |
|-----------------|---|-------|
| <b>distcat</b>  | Plot history event categories, i.e. Disturbance (includes natural and anthropogenic disturbances, management), Establishment, No.disturbance, No.info (data not available), Regrowth. Used in <b>plothistory</b> table. | -     |
| <b>disttype</b> | Code name for plot history event types. Used in <b>plothistory</b> table.   | -     |
| description     | Definition and notes about the disturbance, management, or regeneration type.   | -     |
| units           | Units of the variable.  | -     |

(6) TropForC\_pft.csv - ‘**pft**’

| Column         | Description  | Units |
|----------------|--|-------|
| <b>pftcode</b> | Code name for plant functional traits. Used in <b>measurements</b> table. Codes match those in BETY-db (LeBauer <i>et al.</i> , 2010). | -     |
| description    | Definition of the <b>pftcode</b> .   | -     |
| notes          | Details.   | -     |

(7) TropForC\_methodology.csv - ‘**methodology**’

| Column           | Description   | Units |
|------------------|---|-------|
| <b>method_id</b> | Code name for methodology. Used in <b>measurements</b> table. | -     |
| method_citation  | Reference for the methodology used.                           | -     |
| method_variable  | The type of variable for which the methodology is described.  | -     |
| method_notes     | Description of the methodology used.                          | -     |

(8) TropForC\_allometry.csv - '**allometry**'

| Column                     | Description   | Units |
|----------------------------|---|-------|
| <b>allometric_equation</b> | Code name for allometric equation. Used in <b>covariate</b> field in <b>measurements</b> table. | -     |
| allometry_citation         | Author name and year of publication describing the allometries.                                 | -     |
| allometry_citation.title   | Title of reference for the allometries.   | -     |
| allometry_notes            | Description of the allometries used.  | -     |

**Additional Tables:**

Table 1. Missing data codes

| Code | Definition            | Description   |
|------|-----------------------|---|
| NA   | Not Applicable        | -   |
| NAC  | Not Acquired          | Information may be available but has not been acquired.   |
| NRA  | Not Readily Available | Information was not readily available to the authors (e.g., publication not readily available). |
| NI   | No Information        | Publication does not contain the required information.  |

Table 2. Date level of confidence (**dateloc**) numbering convention, used in **plothistory** and **measurements** tables. Adapted from Table 4 in the BETY-db Data Entry Workflow (available at URL: <https://www.authorea.com/users/5574/articles/6800>; LeBauer *et al.*, 2010).

| Dateloc* | Definition  |
|----------|-------------|
| 9        | nodata      |
| 8        | year        |
| 7        | season      |
| 6        | month       |
| 5        | day         |
| 4        | time of day |

\* When the dateloc is followed by '.5', it indicates that a range of dates (year/month/day) were given; e.g. "late 1990s" would be given a dateloc of 8.5 for the year "1999".



**References:**

- LeBauer D, Dietze M, Kooper R, Long SP, Mulrooney P, Rhode GS, Wang D (2010) Biofuel Ecophysiological Traits and Yields Database (BETYdb). *Energy Biosciences Institute, University of Illinois at Urbana-Champaign*. doi:10.13012/J8H41PB9
- Olson DM, Dinerstein E, Wikramanayake ED et al. (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, **51**, 933.

**How to cite:**

Publications using these data should cite both the original publication and the Dryad data package:

Anderson-Teixeira, K.J. et al. **2016**. Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC). *Global Change Biology*.

Anderson-Teixeira, K.J. et al. **2016**. Data from: Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC). Dryad Digital Repository.