**Forest Carbon database (ForC-db)**

**PI:** Dr. Kristina J. Anderson-Teixeira, Smithsonian Institution; [teixeirak@si.edu](mailto:teixeirak@si.edu)

# Database overview:

The Forest C database (**ForC-db**) is an open access global C database that is updated and maintained on Github (<https://github.com/forc-db>). The database contains data on ground-based measurements of ecosystem-level C stocks and annual fluxes in forests globally, along with site information, disturbance history, and information on methodology. The tropical component of ForC-db has been published in the peer-reviewed literature (**TropForC-db**; [Anderson-Teixeira et al. 2016](http://dx.doi.org/10.1111/gcb.13226)). At present, this is the only publicly available portion of ForC-db, but additional data (both tropical and extra-tropical) have been compiled and will eventually be made publicly available (interested users may contact database PI Dr. Kristina Anderson-Teixeira to discuss collaboration). Moreover, we anticipate database growth as other investigators contribute to the database.

The structure of ForC-db 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. Original citations associated with all data are given in the measurements table (3), and a full bibliography may be found in the file “ForC\_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 “ForC\_Rcode.txt”.

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

# *References:*

Anderson-Teixeira, K.J. et al. **2016**. Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC-db). *Global Change Biology.* doi: [10.1111/gcb.13226](http://dx.doi.org/10.1111/gcb.13226)

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.

## **Data Use:**

ForC-db is an open access database, and we encourage use of the data for scientific research and educational purposes. All users are responsible to adhere to the data use policy: DATA USE POLICY.md (<https://github.com/forc-db/ForC/blob/master/DATA%20USE%20POLICY.md>).

## **Contributing to the database:**

We anticipate that ForC-db will be of great value for various initiatives seeking to understand and manage the role of tropical forests in the global C cycle, and we encourage authors to contribute data. Please see CONTRIBUTING.md(<https://github.com/forc-db/ForC/blob/master/CONTRIBUTING.md>).

# Overview of database structure and data files:

|  |  |  |
| --- | --- | --- |
| Data file | Description | Links to other data files |
| (1) ‘**sites**’  ForC\_sites.csv | Geographic, climatic, and edaphic site data | (2) and (3) |
| (2) ‘**plothistory**’  ForC\_plothistory.csv | Known history of each plot or set of replicate plots | (1), (3), and (5) |
| (3) ‘**measurements**’  ForC\_measurements.csv | Records of ecosystem-level measurements relevant to C cycling | (1), (2), (4), and (6) |
| (4) ‘**variables**’  ForC\_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) ‘**disttype**’  ForC\_disttype.csv | Definition of disturbance, management or regeneration history event types. | Defines codes for column ‘disttype’ in (2) |
| (6) ‘**pft’**  ForC\_pft.csv | Definitions of plant functional codes. | Defines codes for columns ‘acceptedsymbol’ and ‘dominantveg’ in (3) |
| (7) ‘**methodology**’  ForC\_methodology.csv | Description of methodologies. | Defines codes for column ‘method\_id’ in (3) |
| (8) ‘**allometry’**  ForC\_allometry.csv | Sources and description of allometric equations | Defines codes for the covariate ‘allometric\_equation’ in (3) |

**Entity Relationship Diagram for ForC-db:** 

Note: Diagram can be edited at the following public link (anyone with the link can edit)

<https://drive.google.com/file/d/0B9F3sC2fKyd3WS1lZzcwYXd4UmM/view?usp=sharing>

# Data File Contents:

## (1) ForC\_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 **plothistory** and **measurements** 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 |
| climate.notes | Climate details not included above |  |
| 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 | - |
| hydrology.notes | Notes on sites hydrology (e.g., terra firma, seasonally flooded) |  |
| sitenotes | Site details not included above | - |
| siteref | References—in addition to those associated with measurement data—used to obtain site data. When all site data were obtained from references associated with measurement data, this field is populated with ‘NA’. | - |
| 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. Note that groupings are subject to change. As of February 2017, all values were for the set of sites published in Anderson-Teixeira et al. (2016). | - |
| biogeog | Biogeographical region, extracted from map of Olson *et al.* (2001) | - |
| Koeppen | Köppen-Geiger classification, extracted from the ESRI Köppen-Geiger map (http://www.arcgis.com/home/item.html?id=7a53584fa55643df969f93cec83788e1). | - |
| FAOecozone | FAO global ecological zones classification, extracted from FAO’s GeoNetwork (http://www.fao.org:80/geonetwork). | - |
| siteID.v1 | Original siteID, where numbers ≤503 correspond to records released in the first published version of the database (Anderson-Teixeira et al., 2015) |  |

## (2) ForC\_plothistory.csv - ‘**plothistory**’

|  |  |  |
| --- | --- | --- |
| Column | Description | Units |
| plothistoryID | Unique numerical identifier for each record in the history table. | - |
| **sites.sitename** | Unique site identifier, sufficient to identify the site within the original publication. Links to **sites** and **measurements** tables. | - |
| **plot.name** | 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. | - |
| daten | Date number for disambiguation in case of Excel date formatting errors. Format YYYYMMDD. |  |
| dateloc | Level of confidence in date (codes given in Table 2). | - |
| **distcat** | Plot history event categories, namely ‘Disturbance’ (includes natural and anthropogenic disturbances, management), ‘Establishment’, ‘No.info’ (data not available), and ‘Regrowth’. | - |
| **disttype** | Plot history event types. See **disttype** 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. | - |
| plothistoryID.v1 | Original plothistoryID, where numbers ≤2441 correspond to records released in the first published version of the database (Anderson-Teixeira et al., 2015) |  |

## (3) ForC\_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 **sites** and **plothistory** tables. | - |
| **plot.name** | Unique plot identifier name for the plot in which the measurement was made. Links to **plothistory** 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 (≥80% dominance, if given). Identical to **acceptedsymbol** for multi-species forests. For plantations, this is the appropriate **pftcode** 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, **pftcode**, is as defined in the the **pft** table. For single species, species name acronyms are as listed in the USDA Plants Checklist (<http://plants.usda.gov/dl_all.html>), or listed as ‘nocode’ if not available for the species. | - |
| Regentype | Regeneration type of plot, as follows:  natcult – Natural regeneration following cultivation  natgrazed – Natural regeneration following grazing  natother – Natural regeneration following other forms of disturbance  plantation – Plantation/regeneration by planting or sowing  nodata – Missing data or none of the above | - |
| Managementtype | Indicates if the plot is unmanaged (“UM”), managed (“M”) or other/unknown (“O”). | - |
| **variables.name** | Code name for variable sampled. Codes defined in **variables** table. | - |
| **method\_id** | Code number for methodology. Codes defined in **methodology** 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. | - |
| mean | Mean value of measurement recorded. Units dependent on **variables.name**. See **variables** 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 **variables** table. | - |
| coV#\_value | Value of reported covariate. | - |
| coV#\_units | Units of reported covariate. | - |
| dupcode | Marks entries that meet one of the following criteria:   * Duplicate estimates of the same variable but with different values (**code D** for duplicate) * Replicate sampling of the same variables within the same study (**code R** for replicate) * Multiple estimates of the same variables within the same study but using different methods (**code M** for different measurement methods) * Does not have a duplicate (**code 0** for no duplicates) | - |
| 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 original data publication, if available. Studies relying on data from ≤20 original publications contained within the database should cite those publications. | - |
| citations.author | First author’s last name of original data publication. Studies relying on data from ≤20 original publications contained within the database should cite those publications. | - |
| citations.year | Year of original data publication. Studies relying on data from ≤20 original publications contained within the database should cite those publications. | - |
| citations.title | Complete title of original data publication. Studies relying on data from ≤20 original publications contained within the database should cite those publications. | - |
| loaded\_from | Reference to data compilations from which data and reference to primary source were obtained, if applicable. Studies relying on data from ≤10 previous data compilations contained within the database should cite those publications. | - |
| measurementID.v1 | Original plothistoryID, where numbers ≤3568 correspond to records released in the first published version of the database (Anderson-Teixeira et al., 2015) |  |
| ForC\_investigator | Name, institution, and email address of the investigator responsible for creating the database record (*i.e.,* the corresponding author for the database record). This is the individual who should be contacted with correspondence regarding that record (*e.g.,* questions, corrections, requests for collaboration). | - |

## (4) ForC\_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. | - |
| **variables.name** | Code name for variable sampled. Used in **measurements** 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) ForC\_disttype.csv - ‘**disttype**’

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

## (6) ForC\_pft.csv - ‘**pft**’

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

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

|  |  |  |
| --- | --- | --- |
| Column | Description | Units |
| **method\_id** | Code name for methodology. Used in **measurements** table. | - |
| method\_citation | Reference for the methodology used. | - |
| variables.name | The type of variable for which the methodology is described. May not exactly correspond to variables.name in **variables** table, as similar variables may be grouped together. | - |
| method\_notes | Description of the methodology used. | - |

## (8) ForC\_allometry.csv - ‘**allometry**’

|  |  |  |
| --- | --- | --- |
| Column | Description | Units |
| **allometric\_equation** | Code name for allometric equation. Used in **covariate** field in **measurements** 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.,* no ready access to the publication, language barrier). |
| 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".

# Database citations:

Anderson-Teixeira, K.J. et al. **2016**. Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC-db). *Global Change Biology.* <http://dx.doi.org/10.1111/gcb.13226>

Anderson-Teixeira, K.J. et al. **2016**. Data from: Carbon dynamics of mature and regrowth tropical forests derived from a pantropical database (TropForC-db).Dryad Digital Repository. [http://dx.doi.org/10.5061/dryad.t516f](https://datadryad.org/resource/doi:10.5061/dryad.t516f)