## Supplementary Information

## Appendix I: Methods for Building Database

*Selecting the literature from which to extract the data*

(1) First check to make sure that they have **ALL** of the below variables. If they do not, then exclude that paper and note the reason for exclusion in the "Code 2" column on the Literature to INCLUDE tab.

- Some measure of organic carbon (or metric that can be converted). These measures must be per unit area. Relevant variables include :
    - soil organic carbon (SOC)
    - soil organic matter (SOM)
    - %C + bulk density (preferred but not necessary)
    - Biomass
    - Carbon
    - NOTE, a carbon sequestration rate alone is insufficient as we are quantifying stocks.
    - NOTE, soil microbial biomass/carbon alone is insufficient since we are focusing on overall soil carbon
- A clear reforestation activity (see "refor.type" below), though we will only analyze papers that describe naturally regenerating forest here.
- At least one forest age between 5 and 30 (but can include younger and older forests within the same paper). If a range is given, that range must be less than 10 years.
- Latitude & Longitude (OR a findable latitude and longitude from site details (e.g., can look up the site name in googlemaps and get an approximate location)

*General rules*

(1) If raw data are re-analyzed more than once, use the reference with the more complete data to avoid pseudoreplication. If data quality is the same, then only include data from the most recent publication.
(2) For papers within meta-analyses, extract the data from the original paper except when:
    a. the meta-analysis cites an unpublished dataset or a paper that we cannot access (e.g., its written in Chinese or published in a journal we cannot access). In this case, treat the meta-analysis as the correct citation but note the source in the Sites tab under "other references" column.
    b. the meta-analysis acquired and re-analyzed the data with an improved approach (e.g., better allometric equations). In this case, enter the updated numbers under the original publication and note the meta-analysis in the Sites tab under "other references" column
(3) If a site referenced in multiple papers has an identical geolocation across papers, then it is considered the same site.
(4) If a range is given for a variable, calculate the average.

*General database structure*

1. Literature datasheet – includes general citation information

2. Site datasheet – includes all variables specific to a site (e.g., latitude/longitude, climate variables), where a site is defined as a place with a unique geolocation.
3. Measurements datasheet – includes all variables specific to a plot

*Explanation of variables – Literature tab*

| Column name | Description | Values/Units | Type | Notes |
|---|---|---|---|---|
| study.id | unique identifier per publication | 1-highest value needed | integer | |
| citations.author | last name of first author | | short text | |
| citations.year | year of publication | 1975-2018 | YYYY | |
| citations.journal | citation information | Journal Name Volume: pg-pg. | | |
| citations.title | full title from publication | | text | |
| Entered | date final data were entered | date | date | also include a date if you fully exclude the paper |
| citations.notes | any relevant details here such as ambiguities, key assumptions, or judgement calls | | | |
| | | | | |

*Explanation of variables – Site tab*

| Column name | Description | Values/Units | Type | Notes |
|---|---|---|---|---|
| citations.title | see above | | | |
| site.id | unique identifier for each site, defined as a unique geolocation | 1-highest value needed | integer | |
| study.id | see above | | | |
| site.sitename | final decision on exclusion or inclusion | yes = include, no = exclude, otherwise reason to set aside (e.g., basal only, > 30 years) | short text | |
| site.state | sub-national jurisdiction such as state, province etc. | short text | short text | |
| site.country | country name | short text | short text | |
| lat_deg | latitude | degrees.minutes.seconds | single | negative indicates S |
| long_deg | longitude | degrees.minutes.seconds | single | negative indicates W |
| lat_dec | latitude | decimal degrees | single | negative indicates S |
| long_dec | longitude | decimal degrees | single | negative indicates W |

| Column name | Description | Values/Units | Type | Notes |
|---|---|---|---|---|
| other reference | resources used to fill out site information, such as papers referenced in a meta-analysis, the meta-analysis or googlemaps | Author name Year; Place name googlemaps | short text | see General Rules section above |
| elevation | height above sea level | meters | | US system ends with "ols", European system ends with "ol" |
| AMT | annual mean precipitation | celcius | | If given, otherwise we can extract values from WorldClim |
| AMP | annual mean temperature | millimeters | | If given, otherwise we can extract values from WorldClim |
| soil.classification | | oxisols, antisols, andisols, etc | | US system ends with "ols", European system ends with "ol" |
| site.notes | any sources of ambiguity, judgement calls etc. | short text | short text | |

*Explanation of variables – Measurements tab*

| Column name | Description | Values/Units | Type | Notes |
|---|---|---|---|---|
| measurement.id | unique identifier per measurement | 1 to highest value needed | integer | |
| plot.id | unique identifier per plot | 1-highest value needed | integer | |
| site.id | see above | | | |
| study.id | see above | | | |
| citations.title | see above | | | |
| site.sitename | see above | | | |
| refor.type | reforestation type or reference condition | SNR = Spontaneous natural regeneration, ANR = assisted natural regeneration, ITP = initial tree planting, DP = diverse species | text | |

| | | | | |
|---|---|---|---|---|
| | | plantation, MP = monoculture plantation, TMC = intensive tree monocrop, MS = multistrata, TI = tree intercropping, SP = silvopastoral systems, TR = transitional, C = cropland (reference), PA = pasture (reference), OG = old growth (intact forest reference) | | |
| Species | name of dominant species | | text | |
| prior | prior land use | C = crop; SC = shifting cultivation/fallow agriculture/slash and burn agriculture; H = clearcut harvest; F= fire; D = non-fire disturbance such as landslide or hurricane; PA = pasture; OG = old growth (only relevant for intact stand which have no prior condition) | text | |
| stand.age | age of forest stand | C & PA = 0; OG = 999; otherwise as given in paper | integer | Prioritize extracting data from forests < 30 years in age. |
| date | collection date for data | NA, or as given in paper | | This column captures repeat measures as well. |
| n | number of plots per site and stand age | 1-highest value needed | integer | Make sure that the plot size corresponds to how biomass/carbon data are given and plot size |
| sub_n | number of subplots, e.g., soil samples that are pooled | 1-highest value needed | | |
| variables.name | name of stock pool | aboveground_biomass; aboveground_carbon; understory_biomass; | | see additional details below in "Measures and |

| | | | | |
|---|---|---|---|---|
| | | understory_carbon; litter_biomass; litter_carbon; belowground_biomass; belowground_carbon; SOC/SOM/soil_perC | | their covariates" |
| mean_ha | value of biomass or carbon estimate | Mg/ha | | |
| covar_1 | type of covariate | max_depth; max_height; min_height; min_dbh; min_diameter; min_height; min_length | text | see additional details below in "Measures and their covariates" |
| coV1_value | value of covariate 1 | | typically an integer | |
| covar_2 | type of covariate | min_depth; root_diameter_max; root_diameter_min | text | See additional details below in "Measures and their covariates" |
| coV2_value | value of covariate 2 | | typically an integer | |
| covar_3 | type of covariate | soil_type; other | text | See additional details below in "Measures and their covariates" |
| coV3_value | value of covariate 3 | | text, other | |
| plot.size | | hectares | | |
| density | number of individual trees | 1-highest value needed | | |
| sand.silt.clay | soil texture | %:%:% or description (e.g., clay, sandy clay, sandy clay loam, loamy sand, silty clay, silt loam) | short text | |
| pH | pH | | number | |
| allometry | | direct harvest = when all biomass is harvested; site-specific harvest = based on | | |

| | | | | |
|---|---|---|---|---|
| | | trees harvested at the site; species-specific; forest-type-specific; biome-specifi | | |
| measurements.notes | any ambiguities, assumptions, or judgement calls | | | |

*Extracting the data*

- If data are only show graphically, use WebPlotDigitizer to extract the variables. Best practice is to extract single figures at a time to protect clarity of image. Keep a record of the csv and png files.
- If there are multiple values per stand age, calculate the average and increase the value of n on the Measurements tab.
- If a range is given for a numeric value, calculate the average.

*Measurements and their covariates*

1. **Aboveground_biomass/carbon** refers to aboveground tree biomass and should not include herbaceous biomass/carbon. If the two pools are combined, note the presence of the latter by adding **"+ understory_biomass/carbon**" to the variables.name column. A **min_dbh** (covariate 1) should typically be listed with this measurement, with a "0" indicating all trees were sampled. Alternatively, sometimes a height cut-off is included instead, in which case note **min_height** (covariate 1).

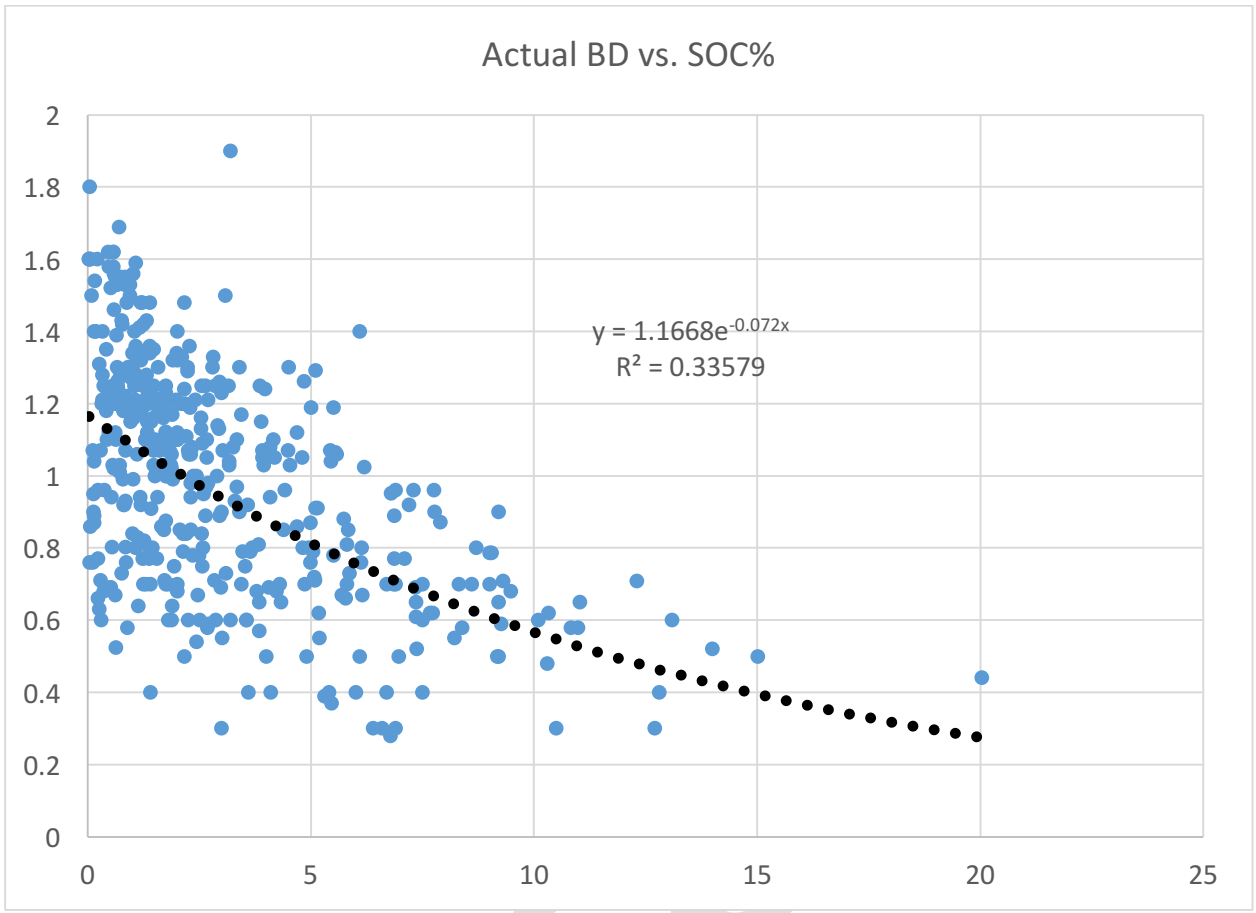
2. **Understory_biomass/carbon** refers to herbaceous biomass though this can also include small trees. If the latter, include **max_height** or **max_dbh** as a covariate (covariate 1).

3. Belowground_biomass refers to root biomass and when directly measured should include a **root_diameter_min** (covariate 1) and/or **root_diameter_max** (covariate 2), if relevant.

4. Soil biomass/carbon will be noted as **SOC, SOM** or **soil_perC**. When multiple types of measurements are included, select SOC if possible, then SOM and then %C. If %C is given, include bulk_density (covariable 3) where possible. For all soil measures, **max_depth** (covariate 1) should be recorded and **min_depth** (covariate 2) should be recorded.

5. **Litter_biomass/carbon** refers to litter and/or deadwood (necromass). Potential covariates include the **min_length** or **min_diameter**, whenever a subsample of the material was assessed.

Figure S1


Actual BD vs. SOC%
y = 1.1668e-0.072x
R² = 0.33579
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25

343