REDD+ Uncertainty Training

Winrock International

Dr Julie Nash

Dr Seamus Murphy

Nicole Scotti

2025-10-30

Table of contents

# Preface

This training resource was commissioned by Winrock International under its JTAP portfolio (20065-2025-ICA-03) as a Foundation Module for uncertainty quantification training in jurisdictional REDD+ carbon accounting. The following training curriculum provides methodological guidance for quantifying, reporting, and reducing uncertainty in REDD+ carbon accounting in accordance with ART-TREES Standard (V2.0) requirements and the IPCC 2019 guidelines (ART, 2021; IPCC, 2019a).

Rather than viewing estimation uncertainty as a penalty to manage, this framework treats it as a strategic metric for identifying dominant error sources and optimizing credit issuance (Camara et al., 2024; Duncanson et al., 2021; Simoes et al., 2021a). Projects investing in targeted uncertainty reduction can achieve improved crediting revenue through minimizing penalty deductions, as well as through stronger pricing from enhanced credibility and lower verification and monitoring costs (Köhler & Huth, 2010).

## Uncertainty Opportunities

Recent analyses have identified opportunities to strengthen uncertainty reporting across REDD+ programs. According to Butler et al. (2024), while 91% of participating countries report activity data uncertainty, only 4-14% report emission factor and allometric uncertainty. This pattern presents an opportunity for jurisdictions to enhance methodological completeness and improve credit issuance outcomes. Research demonstrates that comprehensive uncertainty quantification, including emission factors and modeling assumptions contribute to significant variance across reference levels ranging from 4.2% to 262.2% (Pelletier et al., 2013).

Comprehensive uncertainty reporting serves multiple strategic objectives beyond regulatory compliance. When jurisdictions systematically quantify and report all uncertainty sources, they establish baseline metrics for improvement and unlock pathways to enhanced credit revenue. Butler et al. (2024) note that comprehensive assessment can initially produce wider confidence intervals, but this transparency enables targeted methodological refinements that systematically reduce actual uncertainty over time, a process that directly increases credit issuance and revenue. Recognizing the importance of incentivizing comprehensive reporting, results-based payment programs have implemented supportive mechanisms. The FCPF Carbon Fund uses a capped uncertainty deduction schedule that removes disincentives for comprehensive reporting:

FCPF Uncertainty Deduction Conservativeness Factors:[[1]](#footnote-20)

* No deduction if uncertainty is ≤15%
* 4% deduction if uncertainty is 15–30%
* 8% deduction if uncertainty is 30–60%
* 12% deduction if uncertainty is 60–100%
* 15% maximum deduction if uncertainty is >100%

This structure ensures that programs with very high initial uncertainty face the same maximum 15% deduction, removing the penalty for reporting comprehensive uncertainty while maintaining incentives to improve. The ART-TREES Standard goes further by allowing participants to recover over-deducted credits when cumulative uncertainty decreases across multi-year crediting periods (Section 8, V2.0). These provisions reward demonstrable improvements and create positive incentives for methodological investment.

## Uncertainty Requirements

Section 8 of the ART Standard V2.0 mandates the following criteria (ART, 2021, p. 45):

1. Monte Carlo simulations: Minimum 10,000 iterations for uncertainty propagation
2. 90% confidence intervals: Half-width calculation for adjustment factors
3. Conservative bias: Systematic underestimation acceptable, overestimation prohibited
4. Whole-chain integration: Combine activity data + emission factor uncertainties
5. Crediting period aggregation: Flexibility to sum uncertainty deductions across years
6. Allometry Exemption: Allometric modelling uncertainty is excluded as non-mandatory.

Equation 10: Uncertainty deduction

Equation 11: Uncertainty adjustment factor

## Primer on Statistical Uncertainty

Forest carbon inventory fundamentally relies on making inferences about true population estimates from limited samples of observed data. But a forest is complex, and our measuring tools and samples are never perfect. For example, when we measure 50 field plots across a million-hectare jurisdiction, we expect that our sample mean will vary from the true mean. Uncertainty is simply how much doubt we have that our measured average is close to the true average.

Fundamentally, this difference between our sampled average and the assumed true population average[[2]](#footnote-23) is what informs the metrics of statistical uncertainty. For instance, this dispersion in estimates is computed using:

* Standard deviation (DescTools::SD()):[[3]](#footnote-24) Measures the spread of observed points around the sample mean, a property of our field sample itself describing how spread out our observations are from one another and their mean.
* Standard error (DescTools::MeanSE()): Measures how much the sample mean is expected to deviate from the true population estimates that are assumed to follow the universal law of large numbers when at critical mass. In effect, this property characterizes our subsample’s precision according to that assumed truth. Expressed as , this also implies standard errors will ultimately decrease the more we increase our sample size.

The distinction is critical: standard deviation quantifies data variability, while standard error quantifies estimation precision. The 90% confidence interval uses a z-value of 1.645, meaning we can be 90% confident the true mean lies within ±1.645 standard errors of our sample mean.

* Root Mean Squared Error (DescTools::RMSE()) quantifies prediction error, the average magnitude by which model predictions deviate from observed values. A biomass equation with RMSE = 20 Mg/ha means predictions typically deviate ±20 Mg/ha from true values.

The Uncertainty Adjustment Factor (UAF) applies a conservative deduction to carbon credits based on this prediction error. Higher RMSE necessitates larger deductions. Understanding this error-to-deduction pathway reveals why model quality directly affects credit revenue: a 50% reduction in allometric uncertainty can translate to a 10-15% increase in creditable carbon.

## Primer on Monte Carlo Method

Monte Carlo simulation traces its roots to the Markov Chain, developed by Andrey Markov in the early 1900s (Sheynin, 1989). At the time, Markov publicly feuded with Pavel Nekrasov, who claimed statistical laws required independent events, arguing this supported divine free will. To disprove this, Markov analyzed the first 20,000 letters of a poem by Pushkin finding 8,638 vowels and 11,362 consonants. He demonstrated that the probability of the next letter depends on the current letter, vowels tend to follow consonants and vice versa, yet aggregated over thousands of letter sequences, the system still converged to a stable distribution (43% vowels, 57% consonants). This proved statistical predictability holds even with strong causal dependence.

The technique remained theoretical until 1948, when Nicholas Metropolis, John von Neumann, and Stanislaw Ulam at Los Alamos developed the Monte Carlo method and named it after Ulam’s uncle’s love for the roulette casions of Monte Carlo (Metropolis & Ulam, 1949). Metropolis rewired the ENIAC computer to perform the first Monte Carlo simulations of nuclear core criticality. They discovered that rather than generating random configurations and weighting by the Boltzmann factors, they sampled configurations with probability and weighted globally dramatically reducing computational cost.

A key to its success in nuclear science related to its distinction of one-sided and two-sided error distributions that enabled the design of false-biased simulation algorithms. Specifically, the study of nuclear reactivity required false-bias metrics to more definitively estimate critical levels of neuron activity for achieving a nuclear chain reaction. It is worth noting how the Monte Carlo method emerged from the statistics of critical systems where different rules of dependency are assumed. Critical systems are characterized by self-organizing populations that exist dynamically between steady-state thresholds and unpredictable tipping points, such as in patterns of mega-fires, the spread of beetle outbreaks, or in disease epidemics (Sornette, 2006) where assumption of causal dependence prevails. These approaches also play an increasing role in the spatial and temporal modelling of plant ecology and remote sensing of forest ecosystems (Schrodt et al., 2025).

Operationally, the Monte Carlo method involves three core mathematical principles:

1. Pseudo-random Number Generation
2. Markov-Rule of Memory-less Dependence
3. Law of Large Numbers & Convergence

Pseudo-random Number Generation: Monte Carlo simulations use pseudo-random number generators (PRNG’s) producing deterministic sequences that appear random. The mathematical foundation mirrors prime number distributions, simulating local unpredictability combined with global statistical regularity. There are four different kinds of randomization algorithms that meet Monte Carlo simulation requirements. Randomization compliance can be verified using primality tests published as coded scripts (Baillie & Wagstaff, 1980; Murray, 2003; Solovay & Strassen, 1977).

Markov Property Memory-less Dependence: The system “remembers” only the immediate previous state. For example, this limits temporal bounds of resampling so that each year’s carbon stock projection requires only the current year’s state, not 30 years of history, dramatically reducing computational complexity.

Law of Large Numbers: As iterations increase, sample statistics converge to population parameters. Standard error decreases by when samples increase by factor . This informs ART requirement for minimum 10,000 iterations.

|  |
| --- |
| Causal Dependence Trends of Forests |
| This concept of causal dependence is crucial in forest carbon science because ecosystems are not a series of independent coin flips. They are dynamic systems where the state of the forest in Year *t+1* is heavily dependent on the state in Year *t*.   * Tipping Points and Disturbances: Markovian principles model system transitions, which are vital for understanding disturbance regimes, such as wildfire and pest outbreaks, and potential for ecological tipping points. For example, the probability of a forest transitioning to a degraded state this year is dependent on its current health, not its health 50 years ago. * Modeling Uncertainty: By incorporating these dependent, causal relationships, Monte Carlo models can more realistically simulate the path-dependent uncertainty of carbon stocks over time, rather than treating all uncertainty sources as randomly independent variables |

## Curriculum Design

This curriculum addresses the four primary sources of uncertainty in REDD+ carbon accounting: allometric equations, emission factors, activity data, and their combined error propagation. Winrock International works with partner jurisdictions to implement targeted interventions using Monte Carlo simulation frameworks tailred to each domain. The curriculum draws on real-world examples and practical exercises from this ongoing work:

* Chapter 1 Allometric Uncertainty: Strategic interventions in allometric estimation are presented that reduce uncertainty by 30-50%, including: development of localized equations (Tier 2) through destructive sampling; model ensemble approaches using Bayesian averaging; integration of wood density and height data from LiDAR or field measurements; and implementation of standardized measurement protocols with quality assurance/quality control (QA/QC) procedures.
* Chapter 2 Emission Factor Uncertainty: Multiple methodological interventions and strategic data selections offer important uncertainty reductions in emission factors, including: field validation campaigns, laboratory quantification of organic carbon and soil profiles; seasonal modelling of fuel loads, moisture content, and disturbance severity sampling; and analysis of stand regeneration and updated growth curves. Additionally, the strategic selection of IPCC default emission factors across land conversions represents an often-overlooked resource in emissions accounting. This chapter presents typical default emission factors ranked by their associated uncertainty levels.
* Chapter 3 Activity Data Uncertainty: Significant uncertainty reductions can be achieved through: enhanced reference data collection; improved image classification algorithms; multi-temporal validation datasets; systematic accuracy assessment protocols; and integration of high-resolution imagery or LiDAR for forest/non-forest mapping.
* Chapter 4 Monte Carlo Aggregation: Effective Monte Carlo simulation frameworks incorporate sensitivity analyses to identify uncertainty sources and quantify their contribution to overall credit deductions. Winrock provides technical assistance in implementing these strategies, drawing on use cases to help jurisdictions prioritize cost-effective improvements and quality assurance measures tailored to their specific forest conditions and monitoring infrastructure.

Appendix I includes collection of additional reading specific to each of the chapters’ topics and expanded subsections drawn from Winrock’s ongoing work in uncertainty reporting of combustion emissions and wetland emissions in above-ground and below-ground pools across organic and mineral soils, with key references listed the move towards carbon flux accounting. Keys references and notes also and the historical disaggregation of uncertainty from natural disturbances. This also includes annotated bibliography of IPCC resources discussing the statistical assumptions involved and their known limitations in specific geographic regions of South America and South East Asia.

# 1. Allometry

## Overview

Allometric equations represent the proportional and scaling relationships between different tree dimensions, such as the relationship between a tree’s diameter and its height, biomass, or crown size. When trees are considered at a population scale, their different dimensions are statistically related through shared ontogenic development patterns (Gould, 1966). These relationships remain consistent across tree sizes, from saplings to canopy dominants, when growing under similar conditions (Archibald & Bond, 2003; Bohlman & O’Brien, 2006; Dietze et al., 2008; King, 1996).

This biological principle underpins REDD+ carbon accounting: allometric equations translate field measurements of diameter at breast height (DBH) into biomass estimates, forming the foundation of emission reduction quantification.

|  |
| --- |
| Allometry Investment |
| Although Section 8 exempts structural allometric uncertainty when models are applied consistently, the choice of which model to apply directly impacts reported uncertainty. Allometry variance is absolute and compounds more dramatically so that model selection determines the magnitude of random error that cannot be exempted downstream. In practical terms, selecting a model with 20% RMSE versus 8% RMSE determines whether your project faces a 6% or 2% carbon credit deduction, representing a difference worth $200k in a 1M tCO2-e project at $5/tonne. |

### Environment Setup (R)

easypackages::packages(  
 "animation", "BIOMASS", "cols4all", "covr", "cowplot", "caret",  
 "DescTools", "dataMaid", "dplyr", "FawR", "ForestToolsRS", "forestdata",   
 "flextable", "ggplot2", "giscoR", "ggfortify", "htmltools",   
 "janitor", "jsonlite", "lattice", "leaflet.providers", "leaflet",   
 "lmtest", "lwgeom", "kableExtra", "kernlab", "knitr", "mapedit",   
 "mapview", "maptiles", "Mlmetrics", "ModelMetrics", "moments",   
 "olsrr", "openxlsx", "plotly", "psych", "randomForest",   
 "raster","RColorBrewer", "rmarkdown", "renv", "reticulate",   
 "s2", "sf", "scales", "sits","spdep", "stars", "stringr",   
 "terra", "tmap", "tmaptools", "tidymodels", "tidyverse", "tidyr", "tune",  
 "useful",  
 prompt = F  
 )

### ~~Environment Setup (Python)~~[[4]](#footnote-36)

## 1.1 Allometric Equations

In its broadest sense, allometry describes any linear or non-linear correlation between increases in tree dimensions during ontogenic development (Picard et al., 2012). A more restrictive definition, originating with Huxley (1924), refers specifically to proportional relationships between relative increases in dimensions:

Allometric equations predict aboveground biomass from diameter measurements using species- or biome-specific parameters. Uncertainty compounds from three sources: model selection, parameter estimation, and field measurements. Log-transformation adds complexity through required back-transformation:

which integrates to the power relationship:

and in logarithmic form:

Where:

* AGB: Aboveground biomass (kg)
* DBH: Diameter at breast height (cm)
* α, β: Modelling exponents/parameters
* ε: Random error term

This power-law form reflects the self-similarity principle observed in tree growth: as trees develop from seedlings to mature individuals, they maintain consistent proportional relationships between their dimensions (Gould, 1971). In this framework, the exponent β serves as the allometric coefficient, quantifying how one dimension changes relative to another during growth. For example, if β = 2.5, a 10% increase in diameter corresponds to a 25% increase in biomass. The parameter α is a scaling constant accounting for wood density, architectural form, and other species-specific characteristics.

West et al. (1997) and Enquist et al. (1999) developed an allometric scaling theory based on the “pipe model” (Shinozaki et al., 1964) and physical growth principles. Their framework predicts a universal exponent of β ≈ 2.67 based on biomechanical constraints, tree stability, and hydraulic resistance. While this theoretical exponent has been debated regarding its universality (Muller-Landau et al., 2006; Zianis & Mencuccini, 2004), it provides a physically grounded benchmark for evaluating empirical equations.

In this chapter, we compare allometry species-specific allometry with biome-generic equations that are more commonly applied in REDD+ programs to demonstrate how uncertainty metrics can inform model selection and impact carbon credit deductions. We adopt the broadest definition of allometry as any correlation between tree dimensions, whether in linear, log-log, power-law, or other functional forms, to examine the following questions:

1. How should uncertainty metrics inform model selection and optimization?
2. How does allometry uncertainty impacts carbon credit deductions

## 1.2 Model Selection

The choice of allometric equation directly determines uncertainty magnitude and carbon credit deductions. This section demonstrates quantitative differences between equation categories using the scbi\_stem1 dataset from the [ForestGEO](https://docs.ropensci.org/allodb/reference/scbi_stem1.html) plot inventory in Front Royal, Virginia. This 25.6-hectare mature secondary forest is dominated by Appalachian mixed hardwood species including tulip poplar (*Liriodendron tulipifera*), oaks (*Quercus spp.*), and hickories (*Carya spp.*), representing typical stand composition of the Blue Ridge and Piedmont regions.

For REDD+ carbon accounting, we tend to focus exclusively on ontogenic allometry to estimate relationships between accessible tree dimensions and total aboveground biomass as trees grow from seedling to maturity. However, some reading of evolutionary allometry is sometimes needed to address differences in specific traits that emerged in localized conditions. For example, the following species share in genus but vary in wood density, branch architecture, and bole form:

* *Acer rubrum* (red maple): Lower density (0.49 g/cm³), faster growth
* *Acer saccharum* (sugar maple): Higher density (0.63 g/cm³), denser wood
* *Acer negundo* (box elder): Intermediate density (0.42 g/cm³), different architecture

### Types of Equations

Allometric equations are classified by taxonomic scope and environmental specificity. In practice, the choice of equation category involves balancing precision, cost, and data availability:

1. Species-Specific Equations: Developed from destructive sampling of target species, providing highest accuracy but limited geographic applicability (Gonzalez-Akre et al., 2022). These equations capture species-specific morphometric traits, such as branching architecture and wood density, which influence biomass allocation.
2. Genus-Specific Equations: Aggregated across multiple species within a genus, offering broader applicability with moderate accuracy (Jansen et al., 1996; Jenkins, 2004). Assumes shared evolutionary heritage produces similar allometric scaling within genera.
3. Biome-Generic Equations: Most commonly used in REDD+ report, pan-tropical, generic equations fitted across broader geographic regions and biomes, maximizing applicability but potentially introducing bias (Brown, 1997; Chave et al., 2009; Chave et al., 2014; West et al., 1997). Often incorporate wood density (WD) or environmental stress factors to capture regional variation.
4. Environmentally-Conditioned Equations: Gold-standard allometric models that incorporate biophysical variables reflecting site-specific growing conditions, critical for REDD+ programs in specialized ecosystems (Komiyama et al., 2005, 2008; Rahman et al., 2021). These equations explicitly model environmental drivers of growth variation:
   * Salinity gradients: Mangrove allometry where salt stress affects growth rate, wood density, and architectural form
   * Soil fertility: Nutrient availability influencing wood density and height-diameter relationships
   * Climate: Temperature and precipitation gradients captured through environmental stress factors
   * Geomorphology: Tidal inundation frequency, elevation, or hydrological regime

|  |
| --- |
| Mangroves Tree Height |
| In specialized growing environments, such as mangrove habitats, the choice of allometry significantly affects accuracy estimates. For example, Rocha de Souza Pereira et al. (2018) found generic allometries to produce -18% & +14% wider error magins than species-specific equations in Brazilian mangrove forests. To compensate, generic equations incorporate proxies of environmental stress, such as wood density, height-diameter ratios or “stunting”, and species composition.  However, while inclusion of tree height significantly reduces bias in AGB estimates (Chave et al., 2014; Rutishauser et al., 2013), the accurate measurement of tree height in closed-canopy forests is especially challenging(King & Clark, 2011). Field data with high levels of tree height variance can limit these destructively sampled allometry models. This represents investment opportunity from reducing uncertainty through improved survey technology, such as LiDAR and RADAR (Feldpausch et al., 2011; Valbuena et al., 2016). |

#### Data Preparation

# import dataset from allodb.pkg  
library("allodb")  
data(scbi\_stem1)  
  
#scbi\_stem1 |> dplyr::group\_by(Family) |>  
# dplyr::summarise(`Tree Families sampled` = n()) |>  
# flextable::flextable()|> flextable::autofit()  
  
scbi\_stem1 |> dplyr::group\_by(Family, genus, species) |>  
 dplyr::summarise(`Total Trees (n)` = n()) |>  
 flextable::flextable()|> flextable::autofit() |>  
 flextable::fontsize(size = 9, part = "all")

| Family | genus | species | Total Trees (n) |
| --- | --- | --- | --- |
| Adoxaceae | Sambucus | canadensis | 20 |
| Adoxaceae | Viburnum | prunifolium | 19 |
| Adoxaceae | Viburnum | recognitum | 2 |
| Annonaceae | Asimina | triloba | 14 |
| Aquifoliaceae | Ilex | verticillata | 32 |
| Betulaceae | Carpinus | caroliniana | 57 |
| Betulaceae | Corylus | americana | 1 |
| Cannabaceae | Celtis | occidentalis | 24 |
| Caprifoliaceae | Lonicera | maackii | 1 |
| Cornaceae | Cornus | florida | 24 |
| Elaeagnaceae | Elaeagnus | umbellata | 1 |
| Fabaceae | Cercis | canadensis | 25 |
| Fagaceae | Fagus | grandifolia | 27 |
| Fagaceae | Quercus | alba | 14 |
| Fagaceae | Quercus | michauxii | 1 |
| Fagaceae | Quercus | prinus | 35 |
| Fagaceae | Quercus | rubra | 26 |
| Fagaceae | Quercus | velutina | 8 |
| Hamamelidaceae | Hamamelis | virginiana | 107 |
| Juglandaceae | Carya | cordiformis | 30 |
| Juglandaceae | Carya | glabra | 64 |
| Juglandaceae | Carya | ovalis | 16 |
| Juglandaceae | Carya | sp | 3 |
| Juglandaceae | Carya | tomentosa | 24 |
| Juglandaceae | Juglans | nigra | 13 |
| Lauraceae | Lindera | benzoin | 1,201 |
| Lauraceae | Sassafras | albidum | 8 |
| Magnoliaceae | Liriodendron | tulipifera | 17 |
| Malvaceae | Tilia | americana | 2 |
| Nyssaceae | Nyssa | sylvatica | 10 |
| Oleaceae | Chionanthus | virginicus | 3 |
| Oleaceae | Fraxinus | americana | 52 |
| Oleaceae | Fraxinus | nigra | 22 |
| Oleaceae | Fraxinus | pennsylvanica | 1 |
| Oleaceae | Fraxinus | sp | 1 |
| Platanaceae | Platanus | occidentalis | 6 |
| Rosaceae | Amelanchier | arborea | 123 |
| Rosaceae | Prunus | avium | 6 |
| Rosaceae | Prunus | serotina | 4 |
| Rosaceae | Rosa | multiflora | 151 |
| Rosaceae | Rubus | allegheniensis | 6 |
| Rosaceae | Rubus | pensilvanicus | 2 |
| Rosaceae | Rubus | phoenicolasius | 17 |
| Sapindaceae | Acer | negundo | 15 |
| Sapindaceae | Acer | rubrum | 17 |
| Simaroubaceae | Ailanthus | altissima | 1 |
| Ulmaceae | Ulmus | americana | 2 |
| Ulmaceae | Ulmus | rubra | 20 |
| Ulmaceae | Ulmus | sp | 12 |

|  |
| --- |
| Compiling Species’ Allometries |
| Scope permitting, this training would examine a full allometry worflow. Effectively, this which would require repeating the same process below for all genus in our field dataset, incorporating equations per species cumulatively. We provide brief demonstration of how to add each new equations using alldob operations. |

We filter scbi\_stem1 to *Quercus* observations, the largest genus subsample, to compare the scaling impact of generic allometry with subspecies and genus equations. Missing entries are also removed, providing a new sample of 84 stems from the dataset’s total measurements of 2,287 trees.

* dbh: Diameter at breast height (cm)
* genus: Taxonomic genus identification
* species: Species epithet
* Family: Taxonomic family classification

# Genus-specific subsample of dataset  
scbi\_quercus = scbi\_stem1 |>  
 dplyr::filter(!is.na(dbh)) |>  
 dplyr::filter(genus == "Quercus")   
  
# Species-specific subsample of dataset  
scbi\_quercus\_rubra = scbi\_stem1 |>  
 dplyr::filter(!is.na(dbh)) |>  
 dplyr::filter(genus == "Quercus") |>  
 dplyr::filter(species == "rubra")

| Variable | n | mean | sd | median | min | max | skew | kurtosis | se |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DBH (cm) | 84.000 | 27.795 | 20.177 | 25.980 | 1.040 | 83.500 | 0.438 | -0.719 | 2.202 |
| genus\* | 84.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 |  |  | 0.000 |
| species\* | 84.000 | 3.155 | 1.167 | 3.000 | 1.000 | 5.000 | -0.568 | -0.369 | 0.127 |
| Family\* | 84.000 | 1.000 | 0.000 | 1.000 | 1.000 | 1.000 |  |  | 0.000 |

|  |
| --- |
| DBH Spread |
| At this early stage in the workflow, it helps to quickly scan the range and spread in DBH values. Applied directly, candidate equations can be filtered by min and max dbh values Deviation from the mean provides indicators of the statistical operations needed ahead during equation selection, bias correction, and model optimization. It also informs sample |

### Equation Selection

Following best practices from forest inventory methodology (Duncanson et al., 2021), equation selection may proceed through the four sequential criteria in order of their priority below:

1. Geographic Proximity: Prioritize equations developed in climates and soil conditions similar to your project area
2. Taxonomic Specificity: Prefer species-level > genus-level > family-level equations
3. DBH Range Coverage: Ensure equation applicability spans ≥80% of measured diameter distribution
4. Sample Size Adequacy: Minimum n=50 trees for species-specific; n>150 for genus-level equations[[5]](#footnote-49)

# Load allometric equations  
data(equations)  
data("equations\_metadata")  
  
# display all equation criteria  
dplyr::glimpse(equations)  
NA Rows: 570  
NA Columns: 47  
NA $ ref\_id <chr> "barney\_1977\_bdac", "baskerville\_1966\_dmp…  
NA $ equation\_id <chr> "4b4063", "e2c7c7", "e42e41", "2bc879", "…  
NA $ equation\_allometry <chr> "exp(3.63+2.54\*log(dbh))", "exp(0.15+2.48…  
NA $ equation\_form <chr> "exp(a+b\*log(dbh))", "exp(a+b\*(log(dbh))"…  
NA $ dependent\_variable <chr> "Total aboveground biomass", "Total above…  
NA $ independent\_variable <chr> "DBH", "DBH", "DBH", "DBH", "DBH", "DBH",…  
NA $ equation\_taxa <chr> "Picea mariana", "Picea glauca", "Betula"…  
NA $ allometry\_specificity <chr> "Species", "Species", "Genus", "Genus", "…  
NA $ equation\_categ <chr> "sp\_spec", "sp\_spec", "generic", "generic…  
NA $ geographic\_area <chr> "Alaska, USA", "New Brunswick, Canada", "…  
NA $ original\_coord <chr> "64\xb0 45', 148\xb0 15'", "47\xb051' N, …  
NA $ lat <chr> "65.75", "47.9", "NRA", "NRA", "NRA", "NR…  
NA $ long <chr> "-148.25", "-68.3", "NRA", "NRA", "NRA", …  
NA $ elev\_m <chr> "167-470", "500", "NRA", "NRA", "NRA", "N…  
NA $ geography\_notes <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…  
NA $ mat\_C <chr> "NI", "2.2", "NRA", "NRA", "NRA", "NRA", …  
NA $ min.temp\_C <chr> "NI", "NI", "NRA", "NRA", "NRA", "NRA", "…  
NA $ max.temp\_C <chr> "NI", "NI", "NRA", "NRA", "NRA", "NRA", "…  
NA $ map\_mm <chr> "NI", "1069", "NRA", "NRA", "NRA", "NRA",…  
NA $ frost\_free\_period\_days <chr> "NI", "110", "NRA", "NRA", "NRA", "NRA", …  
NA $ climate\_notes <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…  
NA $ stand\_age\_range\_yr <chr> "51-55 yrs", "35-48 yrs", "NRA", "NRA", "…  
NA $ stand\_age\_history <chr> "Recent fires in 1977", "Stand originated…  
NA $ stand\_basal\_area\_m2\_ha <chr> "NI", "NI", "NRA", "NRA", "NRA", "NRA", "…  
NA $ stand\_trees\_ha <chr> "NI", "NI", "NRA", "NRA", "NRA", "NRA", "…  
NA $ forest\_description <chr> "Typical lowland and upland stands domina…  
NA $ ecosystem\_type <chr> "Temperate forest", "Boreal forest", "Tem…  
NA $ koppen <chr> "Dfc", "Dfb", "Dfb; Cfa", "Dfb; Cfa", "Df…  
NA $ dbh\_min\_cm <chr> "1.5", "2.54", "5", "NRA", "5", "NRA", "5…  
NA $ dbh\_max\_cm <chr> "13", "25.4", "NRA", "NRA", "NRA", "NRA",…  
NA $ sample\_size <chr> "18", "13", "NRA", "NRA", "NRA", "NRA", "…  
NA $ collection\_year <chr> "1973", "1965", "1971, 1995", "1971, 1995…  
NA $ dbh\_units\_original <chr> "cm", "inch", "cm", "cm", "cm", "cm", "cm…  
NA $ dbh\_unit\_CF <chr> "1", "0.393701", "1", "1", "1", "1", "1",…  
NA $ output\_units\_original <chr> "g", "lbs", "metric\_ton", "m", "metric\_to…  
NA $ output\_units\_CF <chr> "0.001", "0.453592", "1000", "1", "1000",…  
NA $ allometry\_development\_method <chr> "harvest", "harvest", "harvest?", "harves…  
NA $ regression\_model <chr> "log-transformed", "log-transformed", NA,…  
NA $ r\_squared <chr> "0.91", "0.99", "NI", "NI", "NI", "NI", "…  
NA $ other\_equations\_tested <chr> "NI", "NI", "NRA", "NRA", "NRA", "NRA", "…  
NA $ log\_biomass <chr> "1", "1", "1", "0", "1", "0", "1", "0", "…  
NA $ bias\_corrected <chr> "0", "0", NA, NA, NA, NA, NA, NA, NA, NA,…  
NA $ bias\_correction\_factor <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…  
NA $ notes\_fitting\_model <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N…  
NA $ original\_equation\_id <chr> "Upland black spruce", NA, NA, NA, NA, NA…  
NA $ original\_data\_availability <chr> "NRA", "NRA", "NRA", "NRA", "NRA", "NRA",…  
NA $ equation\_notes <chr> NA, NA, "Proved difficult to obtain origi…

*Table X: Full list of variables in equations metadata available used in selection criteria using allodb package below.*

|  |
| --- |
| Navigating Databases |
| * It is recommend using latitude and longitude variables over the field called geographic\_area when filtering allodb database due to entry inconsistencies. This is done below using dplyr SQL. However, once the specific allometric equation is identified, we must re-select it using the allodb native function called new\_equations() and the equation’s ID# (Gonzalez-Akre et al., 2025). * In subsequent cells, we import pan-tropical equations and fit them with values from by Global Wood Density database using the computeAGB() andgetWoodDensity() functions from the BIOMASS package (Réjou-Méchain, Tanguy, Piponiot, Chave, & Herault, 2017a). |

#### Step 1: Geographic Selection

# Simple North America filter  
eq\_region <- equations |>  
 dplyr::mutate(lat = as.numeric(lat), long = as.numeric(long)) |>  
 dplyr::filter(!is.na(lat), !is.na(long), lat >= 24, lat <= 72, long >= -168, long <= -52)  
  
# tabulate  
show\_cols = c("ref\_id", "equation\_taxa", "allometry\_specificity", "equation\_allometry")  
head(eq\_region[, show\_cols]) |> flextable::flextable() |> flextable::autofit() # top six rows only

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| barney\_1977\_bdac | Picea mariana | Species | exp(3.63+2.54\*log(dbh)) |
| baskerville\_1966\_dmpi | Picea glauca | Species | exp(0.15+2.48\*log(dbh)) |
| bond-lamberty\_2002\_aabb | Larix laricina | Species | 10^(2.648+0.715\*(log10(dbh))) |
| bond-lamberty\_2002\_aabb | Salix | Genus | 10^(2.481+1.19\*(log10(dbh))) |
| bridge\_1979\_fpom | Quercus velutina | Species | exp(-0.34052+2.65803\*log(dbh)) |
| brown\_1976\_esbf | Juniperus communis | Species | exp(4.081+2.202\*(0.36738 + 0.94932\*log(dbh))) |

#flextable::flextable(eq\_region[, show\_cols]) |> flextable::autofit() # complete list for Appendix  
cat(sprintf("Equations valid to region: %d\n", nrow(eq\_region)))  
NA Equations valid to region: 157

#### Step 2: Taxonomic Selection

# species-specific equations for all Quercus subspecies of North America   
eq\_region\_species <- eq\_region |>  
 dplyr::filter(allometry\_specificity == "Species") |>  
 dplyr::filter(grepl("^Quercus", equation\_taxa, ignore.case = TRUE) | # Starts with "Quercus"  
 (allometry\_specificity %in% c("Genus", "Family") &   
 grepl("Quercus", equation\_taxa, ignore.case = TRUE))  
 )  
  
# genus-specific equations for Quercus populations of North America   
eq\_region\_genus = eq\_region |>   
 dplyr::filter(allometry\_specificity == "Genus") |>  
 dplyr::filter(grepl("^Quercus", equation\_taxa, ignore.case = TRUE) | # Starts with "Quercus"  
 (allometry\_specificity %in% c("Genus", "Family") &   
 grepl("Quercus", equation\_taxa, ignore.case = TRUE))  
 )  
  
# tabulate  
flextable::flextable(eq\_region\_species[, show\_cols]) |> flextable::autofit()

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| bridge\_1979\_fpom | Quercus velutina | Species | exp(-0.34052+2.65803\*log(dbh)) |
| chapman\_1991\_apac | Quercus rubra | Species | exp(-2.972-0.017\*dbh+2.873\*(log(dbh))) |
| clark\_1986b\_wvap | Quercus alba | Species | 1.73738\*(dbh^2)^1.33404 |
| clark\_1986b\_wvap | Quercus alba | Species | 1.69511\*(dbh^2)^1.33917 |
| clark\_1986b\_wvap | Quercus coccinea | Species | 1.62952\*(dbh^2)^1.34103 |
| clark\_1986b\_wvap | Quercus coccinea | Species | 2.94915\*(dbh^2)^1.21733 |
| clark\_1986b\_wvap | Quercus prinus | Species | 0.86674\*(dbh^2)^1.47585 |
| clark\_1986b\_wvap | Quercus prinus | Species | 3.40602\*(dbh^2)^1.19049 |
| clark\_1986b\_wvap | Quercus rubra | Species | 3.06739\*(dbh^2)^1.22338 |
| clark\_1986b\_wvap | Quercus rubra | Species | 3.28071\*(dbh^2)^1.20936 |
| clark\_1986b\_wvap | Quercus velutina | Species | 1.2927\*(dbh^2)^1.36723 |
| clark\_1986b\_wvap | Quercus velutina | Species | 2.5465\*(dbh^2)^1.22586 |
| martin\_1998\_aban | Quercus alba | Species | 10^(-1.266+2.613\*(log10(dbh))) |
| martin\_1998\_aban | Quercus coccinea | Species | 10^(-1.283+2.685\*(log10(dbh))) |
| martin\_1998\_aban | Quercus prinus | Species | 10^(-1.587+2.910\*(log10(dbh))) |
| martin\_1998\_aban | Quercus rubra | Species | 10^(-1.259+2.644\*(log10(dbh))) |
| perala\_1994\_abef | Quercus rubra | Species | 0.1335\*(dbh^2.422)\*((3.706\*dbh^0.4932)^0.4389) |
| this\_study | Quercus agrifolia | Species | 24.2\*(1-exp(-0.0466\*dbh^0.8183)) |
| this\_study | Quercus parvula | Species | 1.8219\*dbh^0.6332 |

#flextable::flextable(eq\_region\_genus[, show\_cols]) |> flextable::autofit()  
cat(sprintf("Genus-specific equations valid to region: %d\n", nrow(eq\_region\_genus)))  
NA Genus-specific equations valid to region: 1  
cat(sprintf("Species-specific equations valid to region: %d\n", nrow(eq\_region\_species)))  
NA Species-specific equations valid to region: 19

#### Step 3: DBH Matching Selection

# filter genus-specific equations by DBH range of field data  
field\_dbh\_min <- min(scbi\_quercus$dbh, na.rm = T) # 1.04 cm  
field\_dbh\_max <- max(scbi\_quercus$dbh, na.rm = T) # 83.5 cm  
  
eq\_region\_genus\_dbh <- eq\_region\_genus |>  
 dplyr::mutate(  
 dbh\_min\_cm = as.numeric(dbh\_min\_cm),   
 dbh\_max\_cm = as.numeric(dbh\_max\_cm)) |>  
 dplyr::filter(!is.na(dbh\_min\_cm), !is.na(dbh\_max\_cm),  
 # extrapolation of saplings (min=20cm), compensated in Step 5 (risk if high sapling count)  
 dbh\_min\_cm <= field\_dbh\_min \* 20,   
 dbh\_max\_cm >= field\_dbh\_max)  
  
eq\_region\_species\_dbh <- eq\_region\_species |>  
 dplyr::mutate(  
 dbh\_min\_cm = as.numeric(dbh\_min\_cm),   
 dbh\_max\_cm = as.numeric(dbh\_max\_cm)) |>  
 dplyr::filter(!is.na(dbh\_min\_cm), !is.na(dbh\_max\_cm),  
 dbh\_max\_cm >= field\_dbh\_max \* 0.7)   
 # extrapolation of crowns allowed, consider outliers  
  
flextable::flextable(eq\_region\_genus\_dbh[, show\_cols]) |> flextable::autofit()

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| stovall\_2018\_ibca | Quercus sp. | Genus | exp(-1.5091+2.3237\*log(dbh))+0.24 |

flextable::flextable(eq\_region\_species\_dbh[, show\_cols]) |> flextable::autofit()

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| chapman\_1991\_apac | Quercus rubra | Species | exp(-2.972-0.017\*dbh+2.873\*(log(dbh))) |
| clark\_1986b\_wvap | Quercus rubra | Species | 3.28071\*(dbh^2)^1.20936 |
| martin\_1998\_aban | Quercus alba | Species | 10^(-1.266+2.613\*(log10(dbh))) |
| this\_study | Quercus agrifolia | Species | 24.2\*(1-exp(-0.0466\*dbh^0.8183)) |

cat(sprintf("Genus-specific equations valid for region, Quercus, & DBH range: %d\n", nrow(eq\_region\_genus\_dbh)))  
NA Genus-specific equations valid for region, Quercus, & DBH range: 1  
cat(sprintf("Specific-specific equations valid for region, Quercus, & DBH range: %d\n", nrow(eq\_region\_species\_dbh)))  
NA Specific-specific equations valid for region, Quercus, & DBH range: 4  
  
  
view(equations)

#### Step 4: Sample Size Selection

The required sample size for allometric equation development depends on desired precision and diameter distribution. Roxburgh et al. (2015) demonstrated through Monte Carlo resampling that biomass predictions with a standard deviation within 7.5% of mean demands sample sizes of between 17 and 166, depending on the algorithm employed. Most importantly, stratified sampling across age class or dbh size is critical to improving precision.

eq\_region\_genus\_dbh\_sample <- eq\_region\_genus\_dbh |>  
 dplyr::filter(sample\_size >= 17) # Minimum for genus-specific equations  
  
eq\_region\_species\_dbh\_sample <- eq\_region\_species\_dbh |>  
 dplyr::filter(sample\_size >= 17) # Minimum for species-specific equations  
  
# Display selected equations & tally valid equations remaining  
flextable::flextable(eq\_region\_genus\_dbh\_sample[, show\_cols]) |> flextable::autofit()

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| stovall\_2018\_ibca | Quercus sp. | Genus | exp(-1.5091+2.3237\*log(dbh))+0.24 |

flextable::flextable(eq\_region\_species\_dbh\_sample[, show\_cols]) |> flextable::autofit()

| ref\_id | equation\_taxa | allometry\_specificity | equation\_allometry |
| --- | --- | --- | --- |
| clark\_1986b\_wvap | Quercus rubra | Species | 3.28071\*(dbh^2)^1.20936 |
| this\_study | Quercus agrifolia | Species | 24.2\*(1-exp(-0.0466\*dbh^0.8183)) |

cat(sprintf("Genus-specific equations meeting all criteria: %d\n", nrow(eq\_region\_genus\_dbh\_sample)))  
NA Genus-specific equations meeting all criteria: 1  
cat(sprintf("Species-specific equations meeting all criteria: %d\n", nrow(eq\_region\_species\_dbh\_sample)))  
NA Species-specific equations meeting all criteria: 2

| Criterion | Threshold | Selected\_Value | Justification |
| --- | --- | --- | --- |
| Geography | North America (24-72°N, -168 to -52°W) | 157 equations in region | Climate similarity to Front Royal, VA (Cfa) |
| Taxonomy | Species > Genus > Family | 1 genus-level, 19 species-specific | Species-specific preferred; genus fallback acceptable |
| DBH Range | ≥80% field data coverage | 1.0-83.5 cm (field), 11-93 cm (equation) | Minor extrapolation for saplings (<5% biomass) |
| Sample Size | ≥50 (Roxburgh et al. 2015) | n=66 (Stovall 2018) | Meets precision threshold of CV=7.5% |

*Table X: Equation Selection Documentation for REDD+ MRV*

|  |  |
| --- | --- |
| Attribute | Detail |
| Equation Taxa/ID | *Quercus* spp. (Genus-level) |
| Biomass Method | Non-destructive estimation using Terrestrial Laser Scanning (TLS) or Terrestrial LiDAR to model tree volume, which was then converted to biomass using published wood density values. |
| Geographic Location | Front Royal, Virginia, USA (Smithsonian Conservation Biology Institute, SCBI) at 38.89 N, -78.15W. |
| Empirical Data | The models were developed using 258 non-destructive volume/biomass estimates across 10 dominant hardwood species in this Temperate Mixed Deciduous Forest. The specific equation is a site-specific genus model. |
| Rationale for Selection | This equation is the most geographically and ecologically similar candidate. It was developed at the exact same site (SCBI in Virginia) where your sample data is likely sourced (based on allodb.Rmdcontent). Despite being a genus-level equation for Quercus (not species-specific), the high degree of climate similarity (Cfa/Temperate Forest) and geographic proximity ensures maximum weight will be assigned by the allodb weighting framework, making it the highest priority candidate for Quercus trees at this site. |
| Citation | Stovall, A. E., Anderson-Teixeira, K. J., & Shugart, H. H. (2018). Terrestrial LiDAR-derived non-destructive woody biomass estimates for 10 hardwood species in Virginia. *Data in brief*, *19*, 1560-1569 |

|  |  |
| --- | --- |
| Attribute | Detail |
| Equation Taxa/ID | Quercus rubra (Species-level) |
| Biomass Method | Destructive Harvesting (Dimensional Analysis). The traditional method involving felling, weighing, and drying tree components to directly measure biomass. |
| Geographic Location | West Virginia, USA (specifically, Monongahela National Forest). |
| Empirical Data | This is a species-specific equation developed explicitly for Northern Red Oak (Quercus rubra). The sample size is typically smaller than meta-analyses but provides high-resolution, direct measurements. |
| Rationale for Selection | This equation provides high taxonomic specificity (a perfect species match for Quercus rubra). Although its location (West Virginia) is not as proximate as the Stovall (2018) study, it is still within the same Appalachian/Eastern US Temperate Forest region. It serves as a crucial check and source of high-quality empirical data derived from the gold-standard destructive method. allodb balances its lower geographic proximity with its higher taxonomic specificity and robust methodology, making it the strongest secondary candidate. |
| Citation | Clark, A. (1985). *Weight, volume, and physical properties of major hardwood species in the Gulf and Atlantic Coastal Plains* (Vol. 250). US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station |

|  |  |
| --- | --- |
| **Attribute** | **Detail** |
|  | Chave et al 2014 |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **Attribute** | **Detail** |
|  | Brown et al 1996 |
|  |  |
|  |  |
|  |  |
|  |  |

#### Step 5: Combining Species Equations

The following demonstrates functions for consolidating new filtered allometry equations into one combined database prepared for use in final inventory biomass estimations. For this important task, we recommend the specific functions below from the allodb library that are ensure a weighted approach is applied to synthesizing and candidate equations:

The weight\_allom() function is responsible for assigning a weight to each candidate equation based on three criteria, determining its influence on the final result:

* Sample size: Equations derived from varying destructive sampling campaigns (n > 100) receive higher weights.
* Taxonomic specificity: Species-specific equations are weighted more heavily than genus-level, which in turn outweigh family-level equations.
* Climate similarity: Equations developed from geographically proximate locations with similar temperature and precipitation regimes receive priority.

The resample\_agb() function implements a Monte Carlo resampling procedure to synthesize a robust, synthetic dataset from the weighted equations:

* Each candidate equation is resampled within its original DBH range.
* The number of resampled values for each equation is proportional to its assigned weight.
* A default of 10,000 iterations (or 1e4 in the demo) ensures a robust representation of the uncertainty distribution. This process generates a synthetic dataset that reflects the collective information from all weighted equations, spanning the full DBH range observed in the target forest.

The est\_params() function then uses the synthetic resampled data to fit the following nonlinear power-law model:

This process yields the location-specific parameters:

* ⍺ (scaling coefficient): Incorporates local wood density and architectural characteristics.
* 𝒃 (allometric exponent): Typically ranges from 2.3 to 2.7, reflecting biomechanical constraints.
* σ (residual standard deviation): Quantifies the prediction uncertainty for the calibrated equation.

The resulting equation is location-specific, informed by broader taxonomic and geographic patterns encoded in the weighted source equations.

# This table includes filtered default Acer equations AND custom new equations.  
eq\_tab\_combined <- new\_equations(  
 subset\_taxa = "Quercus",  
 new\_taxa = c("Quercus ilex", "Castanea sativa"),  
 new\_allometry = c("0.12\*dbh^2.5", "0.15\*dbh^2.7"),  
 new\_coords = c(4, 44),  
 new\_min\_dbh = c(5, 10),  
 new\_max\_dbh = c(35, 68),  
 new\_sample\_size = c(143, 62)  
)  
  
# The get\_biomass() function internally performs the weighting, resampling,  
# and calibration based on the equations in eq\_tab\_combined.  
agb\_custom\_estimate <- get\_biomass(  
 dbh = 50,  
genus = "Quercus",  
 species = "rubrum",  
 coords = c(-78.2, 38.9),  
 new\_eqtable = eq\_tab\_combined # Use the consolidated custom table  
)  
  
# Print the resulting AGB estimate  
print(paste("Estimated AGB using the combined custom table:", agb\_custom\_estimate, "kg"))

|  |
| --- |
| Allometry Best Practices |
| 1. Transparency: Document all criteria and thresholds 2. Reproducibility: Code-based workflow enables auditing 3. Bias Reduction: Geographic and taxonomic filtering minimizes systematic errors to species level 4. Uncertainty Surveillance: Multiple equations enable sensitivity analysis (Section 1.5) |

## Biomass Estimations

Having identified candidate species equations, we now compare their aboveground biomass estimates with biome-generic equations (Brown, 1997; Chave et al., 2014). Noting their geographic mismatch, we compared these with pan-tropical equations specifically in order to highlight discrepancies in uncertainty and significance of geography to allometric calibrations.

# once allometry equation is confirmed, use native allodb to load it  
eq\_region\_genus\_dbh\_sample = allodb::new\_equations(subset\_ids = "a664c1")  
  
# species-specific biomass estimates   
scbi\_quercus$agb\_species <- allodb::get\_biomass(  
 dbh = scbi\_quercus$dbh,  
 genus = scbi\_quercus$genus,  
 species = scbi\_quercus$species,  
 coords = c(-78.2, 38.9)  
 )  
  
# genus-specific biomass estimates  
scbi\_quercus$agb\_genus <- allodb::get\_biomass(  
 dbh = scbi\_quercus$dbh,  
 genus = scbi\_quercus$genus,  
 species = scbi\_quercus$species,  
 coords = c(-78.2, 38.9),  
 new\_eqtable = eq\_region\_genus\_dbh\_sample  
 )

The following functions from the BIOMASS package (Réjou-Méchain, Tanguy, Piponiot, Chave, & Herault, 2017b) require some additional wrangling, specifically to extract values from the Global Wood Density database (Vieilledent et al., 2018) and convert units (Réjou-Méchain, Tanguy, Piponiot, Chave, & Hérault, 2017).

|  |
| --- |
| Unit Conversions |
| Always verify if equations require unit conversions or scaling limiters. For example, the most commonly used allometry equation from Chave et al (2014) requires conversion from milligrams to kilograms, as shown below (\* 1000). |

# derive generic estimates using standard equations  
wood\_densities <- BIOMASS::getWoodDensity(  
 genus = scbi\_quercus$genus,  
 species = scbi\_quercus$species,  
 stand = scbi\_quercus$Plot)  
scbi\_quercus$WD <- wood\_densities$meanWD  
  
# Chave et al 2014  
scbi\_quercus$agb\_chave2014 <- BIOMASS::computeAGB(  
 D = scbi\_quercus$dbh,  
 WD = scbi\_quercus$WD,  
 coord = c(-78.2, 38.9))  
  
# convert units  
scbi\_quercus$agb\_chave2014 = scbi\_quercus$agb\_chave2014 \* 1000  
  
# Chave et al 2005 (MANUAL Fit)  
scbi\_quercus$agb\_chave2005 <- scbi\_quercus$WD \*   
 exp(-1.499 + 2.148\*log(scbi\_quercus$dbh) +   
 0.207\*(log(scbi\_quercus$dbh))^2 -   
 0.0281\*(log(scbi\_quercus$dbh))^3)  
  
# Brown et al 1997 (MANUAL Fit)  
scbi\_quercus$agb\_brown1997 <- exp(2.134 + 2.53 \* log(scbi\_quercus$dbh))  
scbi\_quercus$agb\_brown1997 = scbi\_quercus$agb\_brown1997 / 100   
  
flextable::flextable(scbi\_quercus) |>  
 flextable::autofit()

| treeID | stemID | dbh | genus | species | Family | agb\_species | agb\_genus | WD | agb\_chave2014 | agb\_chave2005 | agb\_brown1997 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6,014 | 6,014 | 8.80 | Quercus | alba | Fagaceae | 41.7637958 | 34.6344546 | 0.6300 | 7.28909988 | 29.9737527 | 20.71693889 |
| 6,201 | 6,201 | 41.72 | Quercus | alba | Fagaceae | 1,290.4913814 | 1,287.8799035 | 0.6300 | 369.67536146 | 1,763.7413271 | 1,062.32043295 |
| 1,158 | 1,158 | 29.82 | Quercus | alba | Fagaceae | 615.5371694 | 590.2328264 | 0.6300 | 161.46721089 | 748.5947279 | 454.24357825 |
| 3,736 | 3,736 | 44.38 | Quercus | alba | Fagaceae | 1,478.8739694 | 1,486.7745358 | 0.6300 | 430.07341800 | 2,060.0375737 | 1,242.13345169 |
| 3,741 | 3,741 | 1.12 | Quercus | alba | Fagaceae | 0.4437772 | 0.2879769 | 0.6300 | 0.02849730 | 0.1799659 | 0.11253976 |
| 6,014 | 32,937 | 8.20 | Quercus | alba | Fagaceae | 35.7428951 | 29.3933603 | 0.6300 | 6.06737169 | 24.8619475 | 17.32740278 |
| 6,109 | 6,109 | 72.18 | Quercus | alba | Fagaceae | 4,321.0815102 | 4,602.9666929 | 0.6300 | 1,397.59542709 | 6,762.9023338 | 4,251.86147125 |
| 3,829 | 3,829 | 1.75 | Quercus | alba | Fagaceae | 1.1869883 | 0.8122677 | 0.6300 | 0.09783959 | 0.4970427 | 0.34807328 |
| 6,065 | 6,065 | 66.91 | Quercus | alba | Fagaceae | 3,656.0028342 | 3,859.5305575 | 0.6300 | 1,164.71639683 | 5,643.1934452 | 3,509.75409893 |
| 6,122 | 6,122 | 26.13 | Quercus | alba | Fagaceae | 460.0281377 | 434.2381530 | 0.6300 | 116.23508795 | 531.5668681 | 325.19741123 |
| 3,942 | 3,942 | 20.04 | Quercus | alba | Fagaceae | 256.2896125 | 234.4031792 | 0.6300 | 59.76904328 | 265.4142731 | 166.18262526 |
| 3,832 | 3,832 | 8.75 | Quercus | alba | Fagaceae | 41.2424636 | 34.1789382 | 0.6300 | 7.18211884 | 29.5246386 | 20.42042603 |
| 6,198 | 6,198 | 18.55 | Quercus | alba | Fagaceae | 216.1527538 | 195.8849828 | 0.6300 | 49.18605136 | 216.5281391 | 136.67671331 |
| 3,829 | 32,657 | 1.52 | Quercus | alba | Fagaceae | 0.8700448 | 0.5854821 | 0.6300 | 0.06641044 | 0.3579301 | 0.24369616 |
| 4,930 | 4,930 | 42.89 | Quercus | michauxii | Fagaceae | 1,373.4642682 | 1,373.3608400 | 0.6515 | 408.00003593 | 1,955.3844042 | 1,139.31864776 |
| 4,908 | 4,908 | 33.74 | Quercus | prinus | Fagaceae | 807.5290502 | 786.4119199 | 0.5700 | 199.94839362 | 930.5024048 | 620.85681481 |
| 4,902 | 4,902 | 26.56 | Quercus | prinus | Fagaceae | 476.4422695 | 451.0228853 | 0.5700 | 110.41164268 | 501.7925319 | 338.90772523 |
| 3,833 | 3,833 | 55.13 | Quercus | prinus | Fagaceae | 2,384.4112362 | 2,461.0519474 | 0.5700 | 665.23436471 | 3,191.5742054 | 2,150.28497856 |
| 3,861 | 32,665 | 31.93 | Quercus | prinus | Fagaceae | 715.0795197 | 691.8486167 | 0.5700 | 174.45804487 | 807.7462398 | 540.01726667 |
| 4,883 | 4,883 | 16.65 | Quercus | prinus | Fagaceae | 170.1313989 | 152.3911925 | 0.5700 | 34.12476597 | 147.2521410 | 103.98302468 |
| 4,945 | 4,945 | 7.15 | Quercus | prinus | Fagaceae | 26.3798528 | 21.3787654 | 0.5700 | 3.87114822 | 15.6636024 | 12.25120227 |
| 3,841 | 32,661 | 3.19 | Quercus | prinus | Fagaceae | 4.4498892 | 3.2776221 | 0.5700 | 0.45567375 | 1.9450578 | 1.58991841 |
| 6,105 | 6,105 | 38.12 | Quercus | prinus | Fagaceae | 1,056.9294420 | 1,044.2728542 | 0.5700 | 270.14158890 | 1,270.1903091 | 845.47598474 |
| 5,006 | 5,006 | 48.12 | Quercus | prinus | Fagaceae | 1,766.6172691 | 1,794.2757411 | 0.5700 | 477.90912983 | 2,280.9877313 | 1,524.29234100 |
| 4,937 | 4,937 | 52.78 | Quercus | prinus | Fagaceae | 2,166.0259005 | 2,224.1461809 | 0.5700 | 598.47838912 | 2,867.4570591 | 1,925.89207151 |
| 3,851 | 3,851 | 30.39 | Quercus | prinus | Fagaceae | 641.2312540 | 616.7790454 | 0.5700 | 154.34390938 | 711.2124639 | 476.53310154 |
| 6,175 | 6,175 | 1.06 | Quercus | prinus | Fagaceae | 0.3919584 | 0.2533948 | 0.5700 | 0.02229260 | 0.1443867 | 0.09790577 |
| 3,759 | 3,759 | 5.41 | Quercus | prinus | Fagaceae | 14.2631279 | 11.1837102 | 0.5700 | 1.86096287 | 7.5384467 | 6.05025214 |
| 4,910 | 4,910 | 24.54 | Quercus | prinus | Fagaceae | 400.1816203 | 375.2995279 | 0.5700 | 90.62579821 | 408.3704247 | 277.43869296 |
| 3,859 | 3,859 | 47.92 | Quercus | prinus | Fagaceae | 1,750.4667155 | 1,776.9958375 | 0.5700 | 473.08038364 | 2,257.4804788 | 1,508.31475695 |
| 4,921 | 4,921 | 49.49 | Quercus | prinus | Fagaceae | 1,879.4324090 | 1,915.2122960 | 0.5700 | 511.74583227 | 2,445.6803675 | 1,636.49080317 |
| 3,854 | 3,854 | 23.03 | Quercus | prinus | Fagaceae | 347.8876503 | 323.8130771 | 0.5700 | 77.30652427 | 345.9194869 | 236.25885900 |
| 4,971 | 4,971 | 44.51 | Quercus | prinus | Fagaceae | 1,487.5100601 | 1,496.9134623 | 0.5700 | 395.04844833 | 1,877.5514630 | 1,251.35966700 |
| 4,905 | 4,905 | 19.73 | Quercus | prinus | Fagaceae | 247.3597617 | 226.0642853 | 0.5700 | 52.41058628 | 230.4819305 | 159.75555463 |
| 4,895 | 4,895 | 33.33 | Quercus | prinus | Fagaceae | 786.0488718 | 764.3861533 | 0.5700 | 193.99724551 | 901.8064724 | 601.94627008 |
| 3,754 | 3,754 | 53.48 | Quercus | prinus | Fagaceae | 2,229.8789323 | 2,293.2871922 | 0.5700 | 617.94106423 | 2,962.0057034 | 1,991.17136355 |
| 3,758 | 3,758 | 29.88 | Quercus | prinus | Fagaceae | 617.7416800 | 592.9957612 | 0.5700 | 147.99679828 | 680.8237270 | 456.55939093 |
| 6,105 | 32,948 | 29.05 | Quercus | prinus | Fagaceae | 580.5353679 | 555.4242351 | 0.5700 | 137.99733622 | 633.0294522 | 425.15192548 |
| 5,015 | 32,813 | 30.05 | Quercus | prinus | Fagaceae | 625.5183651 | 600.8643974 | 0.5700 | 150.09528540 | 690.8666473 | 463.15989790 |
| 3,861 | 3,861 | 47.58 | Quercus | prinus | Fagaceae | 1,723.1965011 | 1,747.8381802 | 0.5700 | 464.93603183 | 2,217.8303142 | 1,481.38595772 |
| 6,181 | 6,181 | 1.09 | Quercus | prinus | Fagaceae | 0.4168379 | 0.2703709 | 0.5700 | 0.02409664 | 0.1534332 | 0.10506872 |
| 6,051 | 6,051 | 17.43 | Quercus | prinus | Fagaceae | 188.2040310 | 169.4956107 | 0.5700 | 38.32204684 | 166.2015346 | 116.75267326 |
| 4,921 | 32,798 | 45.95 | Quercus | prinus | Fagaceae | 1,595.7024677 | 1,611.8547626 | 0.5700 | 427.01499709 | 2,033.1946354 | 1,356.33440775 |
| 3,841 | 3,841 | 4.07 | Quercus | prinus | Fagaceae | 7.6147833 | 5.7729030 | 0.5700 | 0.87473536 | 3.6111657 | 2.94481692 |
| 4,900 | 4,900 | 35.01 | Quercus | prinus | Fagaceae | 876.0796244 | 856.9104965 | 0.5700 | 219.04815032 | 1,022.7284035 | 681.69558886 |
| 4,941 | 4,941 | 19.88 | Quercus | prinus | Fagaceae | 251.5256581 | 230.0777621 | 0.5700 | 53.42046148 | 235.1235863 | 162.84629534 |
| 5,015 | 5,015 | 51.95 | Quercus | prinus | Fagaceae | 2,091.6262859 | 2,143.7233445 | 0.5700 | 575.86234802 | 2,757.5403004 | 1,850.18804602 |
| 3,754 | 32,644 | 18.79 | Quercus | prinus | Fagaceae | 222.1161359 | 201.8240515 | 0.5700 | 46.33847538 | 202.6626664 | 141.19494298 |
| 4,942 | 4,942 | 38.24 | Quercus | prinus | Fagaceae | 1,064.2799971 | 1,051.9267603 | 0.5700 | 272.23782635 | 1,280.3611441 | 852.22575020 |
| 4,939 | 4,939 | 8.42 | Quercus | prinus | Fagaceae | 37.8310265 | 31.2583034 | 0.5700 | 5.92783459 | 24.1268820 | 18.52780521 |
| 6,043 | 6,043 | 14.43 | Quercus | rubra | Fagaceae | 124.3720267 | 109.2846080 | 0.5600 | 23.33392646 | 99.0210094 | 72.39828285 |
| 3,740 | 3,740 | 37.50 | Quercus | rubra | Fagaceae | 1,022.9397332 | 1,005.2332291 | 0.5600 | 255.26797929 | 1,197.0066207 | 811.11705223 |
| 4,844 | 4,844 | 4.50 | Quercus | rubra | Fagaceae | 9.5099394 | 7.2902273 | 0.5600 | 1.12427256 | 4.5933623 | 3.79675072 |
| 4,967 | 4,967 | 59.19 | Quercus | rubra | Fagaceae | 2,800.2155110 | 2,902.8525462 | 0.5600 | 777.41640924 | 3,730.0673948 | 2,573.78798878 |
| 5,996 | 5,996 | 2.38 | Quercus | rubra | Fagaceae | 2.3324680 | 1.6595002 | 0.5600 | 0.20318978 | 0.9240671 | 0.75774676 |
| 4,866 | 4,866 | 23.28 | Quercus | rubra | Fagaceae | 357.2923365 | 332.0392166 | 0.5600 | 78.14311997 | 349.5885982 | 242.80150157 |
| 3,757 | 3,757 | 56.39 | Quercus | rubra | Fagaceae | 2,516.2600962 | 2,593.7251264 | 0.5600 | 691.34613971 | 3,314.0550756 | 2,276.80465795 |
| 6,176 | 6,176 | 83.50 | Quercus | rubra | Fagaceae | 5,982.8500337 | 6,457.2112247 | 0.5600 | 1,777.37126337 | 8,469.3561623 | 6,146.83007285 |
| 4,961 | 4,961 | 5.91 | Quercus | rubra | Fagaceae | 17.3523140 | 13.7336417 | 0.5600 | 2.31119112 | 9.3298900 | 7.56659907 |
| 4,868 | 4,868 | 10.05 | Quercus | rubra | Fagaceae | 55.9887078 | 47.1558874 | 0.5600 | 9.22389498 | 37.8950021 | 28.99109110 |
| 6,076 | 6,076 | 38.59 | Quercus | rubra | Fagaceae | 1,089.6953435 | 1,074.4330615 | 0.5600 | 273.90764662 | 1,287.3055373 | 872.09861504 |
| 2,623 | 2,623 | 17.00 | Quercus | rubra | Fagaceae | 178.5575597 | 159.9381204 | 0.5600 | 35.39142453 | 152.8518749 | 109.60243314 |
| 6,161 | 6,161 | 11.30 | Quercus | rubra | Fagaceae | 72.5157421 | 61.9200115 | 0.5600 | 12.47711209 | 51.7345785 | 39.00074350 |
| 4,907 | 4,907 | 49.98 | Quercus | rubra | Fagaceae | 1,928.0952178 | 1,959.5606600 | 0.5600 | 515.70304472 | 2,462.1753315 | 1,677.79507251 |
| 3,677 | 3,677 | 38.92 | Quercus | rubra | Fagaceae | 1,110.3616032 | 1,095.9023318 | 0.5600 | 279.70107400 | 1,315.3901411 | 891.09031354 |
| 6,162 | 6,162 | 50.62 | Quercus | rubra | Fagaceae | 1,982.9908592 | 2,018.3579262 | 0.5600 | 531.92428266 | 2,540.9789427 | 1,732.68440736 |
| 6,041 | 6,041 | 9.66 | Quercus | rubra | Fagaceae | 51.3068293 | 43.0127640 | 0.5600 | 8.32711933 | 34.1164666 | 26.22868617 |
| 3,828 | 32,656 | 1.04 | Quercus | rubra | Fagaceae | 0.3754286 | 0.2424244 | 0.5600 | 0.02079709 | 0.1361149 | 0.09329937 |
| 3,735 | 3,735 | 40.03 | Quercus | rubra | Fagaceae | 1,181.4363039 | 1,169.8984836 | 0.5600 | 299.70417376 | 1,412.4159200 | 956.79728404 |
| 6,142 | 32,955 | 2.37 | Quercus | rubra | Fagaceae | 2.3109000 | 1.6433444 | 0.5600 | 0.20088003 | 0.9146106 | 0.74971772 |
| 4,873 | 4,873 | 12.29 | Quercus | rubra | Fagaceae | 87.2783294 | 75.2622525 | 0.5600 | 15.47986003 | 64.6640393 | 48.23371373 |
| 4,864 | 4,864 | 11.13 | Quercus | rubra | Fagaceae | 70.1305321 | 59.7770919 | 0.5600 | 11.99996949 | 49.6929365 | 37.53334068 |
| 6,031 | 6,031 | 25.83 | Quercus | rubra | Fagaceae | 449.3891747 | 422.7421958 | 0.5600 | 101.33350564 | 458.5174382 | 315.83417415 |
| 6,142 | 6,142 | 3.72 | Quercus | rubra | Fagaceae | 6.2485265 | 4.6844437 | 0.5600 | 0.67693012 | 2.8196238 | 2.34562232 |
| 3,828 | 3,828 | 1.47 | Quercus | rubra | Fagaceae | 0.8055831 | 0.5417037 | 0.5600 | 0.05433775 | 0.2945939 | 0.22392226 |
| 6,000 | 6,000 | 21.84 | Quercus | rubra | Fagaceae | 310.3416778 | 286.2581018 | 0.5600 | 66.58020151 | 295.7446428 | 206.58251378 |
| 4,871 | 4,871 | 5.79 | Quercus | velutina | Fagaceae | 16.5409796 | 13.0944079 | 0.5600 | 2.18970505 | 8.8423679 | 7.18391557 |
| 5,010 | 5,010 | 48.71 | Quercus | velutina | Fagaceae | 1,813.7548532 | 1,845.8072228 | 0.5600 | 484.36445811 | 2,309.8804079 | 1,572.02086480 |
| 3,782 | 3,782 | 29.62 | Quercus | velutina | Fagaceae | 605.4843397 | 581.0757119 | 0.5600 | 142.48091704 | 653.9536822 | 446.57525689 |
| 6,020 | 6,020 | 12.18 | Quercus | velutina | Fagaceae | 85.2885641 | 73.7063466 | 0.5600 | 15.12706577 | 63.1381551 | 47.14895719 |
| 6,160 | 6,160 | 62.59 | Quercus | velutina | Fagaceae | 3,153.0868571 | 3,305.1052960 | 0.5600 | 889.75082893 | 4,271.3028221 | 2,964.43542637 |
| 6,145 | 6,145 | 48.15 | Quercus | velutina | Fagaceae | 1,768.0826750 | 1,796.8758387 | 0.5600 | 470.90307235 | 2,244.4460318 | 1,526.69766645 |
| 6,139 | 6,139 | 62.21 | Quercus | velutina | Fagaceae | 3,111.0191339 | 3,258.6681381 | 0.5600 | 876.76558732 | 4,208.8442021 | 2,919.11154588 |
| 6,101 | 6,101 | 16.18 | Quercus | velutina | Fagaceae | 159.5534665 | 142.5822436 | 0.5600 | 31.22082197 | 134.1111885 | 96.71639363 |

Table 1D: Aboveground biomass estimates derived from five allometry equations of varying scales

### Normality Testing

Non-normal distributions violate assumptions of parametric statistics, inflating uncertainty estimates. Identifying the true probability distribution enables appropriate transformations that reduce reported uncertainty—directly reducing carbon credit deductions.

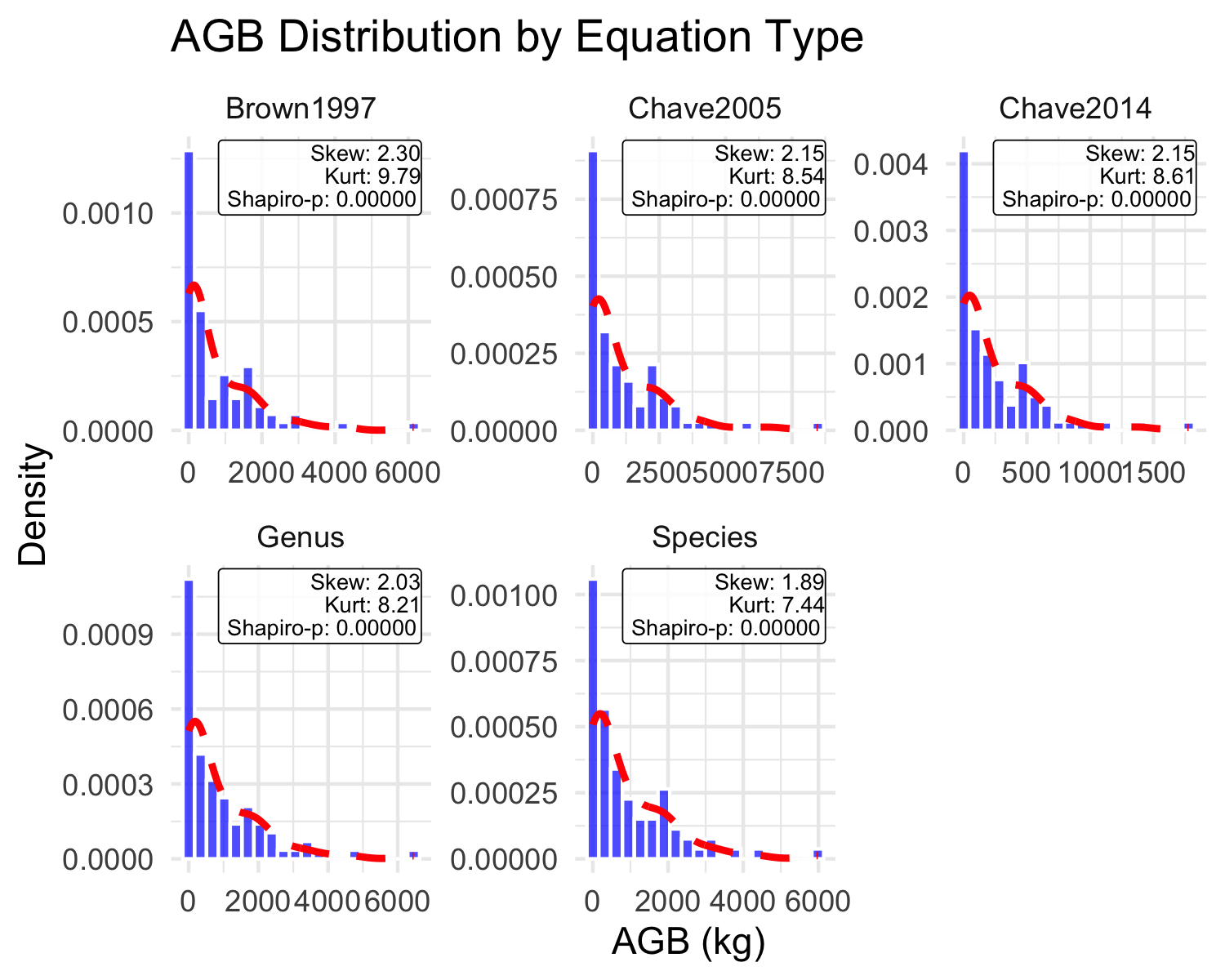
Accurate probability density functions (PDFs) are essential for uncertainty modeling. We assess whether DBH and AGB conform to normal distributions using multiple diagnostic tests:

* Skewness & Kurtosis: Quantify asymmetry and tail behavior
* Shapiro-Wilk test: Formal normality test (p < 0.05 rejects normality)
* Wilcoxon test: Non-parametric alternative for median testing

Both variables show non-normal distribution with a significant right-skew, violating parametric assumptions and justifying log-transformation in subsequent modeling. Technically, this kind of skew often represents a dataset of many small trees and few large dominants. Statistically, this high positive skewness is confirmed by Shapiro-Wilk test results (p < 0.001) indicating a distribution likely to inflate uncertainty estimates if left untreated.

# Calculate skewness and kurtosis for DBH  
dbh\_skew = moments::skewness(scbi\_quercus$dbh)  
dbh\_kurt = moments::kurtosis(scbi\_quercus$dbh)  
dbh\_shapiro = stats::shapiro.test(scbi\_quercus$dbh)  
dbh\_wilcox = stats::wilcox.test(scbi\_quercus$dbh)  
  
# Calculate skewness and kurtosis for each AGB estimate  
agb\_species\_skew = moments::skewness(scbi\_quercus$agb\_species)  
agb\_species\_kurt = moments::kurtosis(scbi\_quercus$agb\_species)  
agb\_species\_shapiro = stats::shapiro.test(scbi\_quercus$agb\_species)  
agb\_species\_wilcox = stats::wilcox.test(scbi\_quercus$agb\_species)  
  
agb\_genus\_skew = moments::skewness(scbi\_quercus$agb\_genus)  
agb\_genus\_kurt = moments::kurtosis(scbi\_quercus$agb\_genus)  
agb\_genus\_shapiro = stats::shapiro.test(scbi\_quercus$agb\_genus)  
agb\_genus\_wilcox = stats::wilcox.test(scbi\_quercus$agb\_genus)  
   
agb\_chave2014\_skew = moments::skewness(scbi\_quercus$agb\_chave2014)  
agb\_chave2014\_kurt = moments::kurtosis(scbi\_quercus$agb\_chave2014)  
agb\_chave2014\_shapiro = stats::shapiro.test(scbi\_quercus$agb\_chave2014)  
agb\_chave2014\_wilcox = stats::wilcox.test(scbi\_quercus$agb\_chave2014)  
  
agb\_brown1997\_skew = moments::skewness(scbi\_quercus$agb\_brown1997)  
agb\_brown1997\_kurt = moments::kurtosis(scbi\_quercus$agb\_brown1997)  
agb\_brown1997\_shapiro = stats::shapiro.test(scbi\_quercus$agb\_brown1997)  
agb\_brown1997\_wilcox = stats::wilcox.test(scbi\_quercus$agb\_brown1997)  
  
agb\_chave2005\_skew = moments::skewness(scbi\_quercus$agb\_chave2005)  
agb\_chave2005\_kurt = moments::kurtosis(scbi\_quercus$agb\_chave2005)  
agb\_chave2005\_shapiro = stats::shapiro.test(scbi\_quercus$agb\_chave2005)  
agb\_chave2005\_wilcox = stats::shapiro.test(scbi\_quercus$agb\_chave2005)

| Variable | n | Mean...SD | Skewness | Kurtosis | Shapiro.Wilk.p | Decision |
| --- | --- | --- | --- | --- | --- | --- |
| DBH (cm) | 84 | 27.8 ± 20.2 | 0.45 | 2.34 | 0.002 | **Log(x) needed?: YES** |
| AGB Species (kg) | 84 | 894.7 ± 1124.2 | 1.89 | 7.44 | < 0.001 | **Log(x) needed?: YES** |
| AGB Genus (kg) | 84 | 908.5 ± 1187.6 | 2.03 | 8.21 | < 0.001 | **Log(x) needed?: YES** |
| AGB Chave2014 (kg) | 84 | 244.1 ± 333.1 | 2.15 | 8.61 | < 0.001 | **Log(x) needed?: YES** |
| AGB Brown1997 (kg) | 84 | 776.5 ± 1085.9 | 2.30 | 9.79 | < 0.001 | **Log(x) needed?: YES** |
| AGB Chave2005 (kg) | 84 | 1159.3 ± 1601.6 | 2.15 | 8.54 | < 0.001 | **Log(x) needed?: YES** |
| *All AGB estimates exhibit significant departure from normality (p < 0.001) with extreme right-skew (skewness > 2) regardless of equation type, justifying log-transformation in subsequent analysis.* | | | | | | |

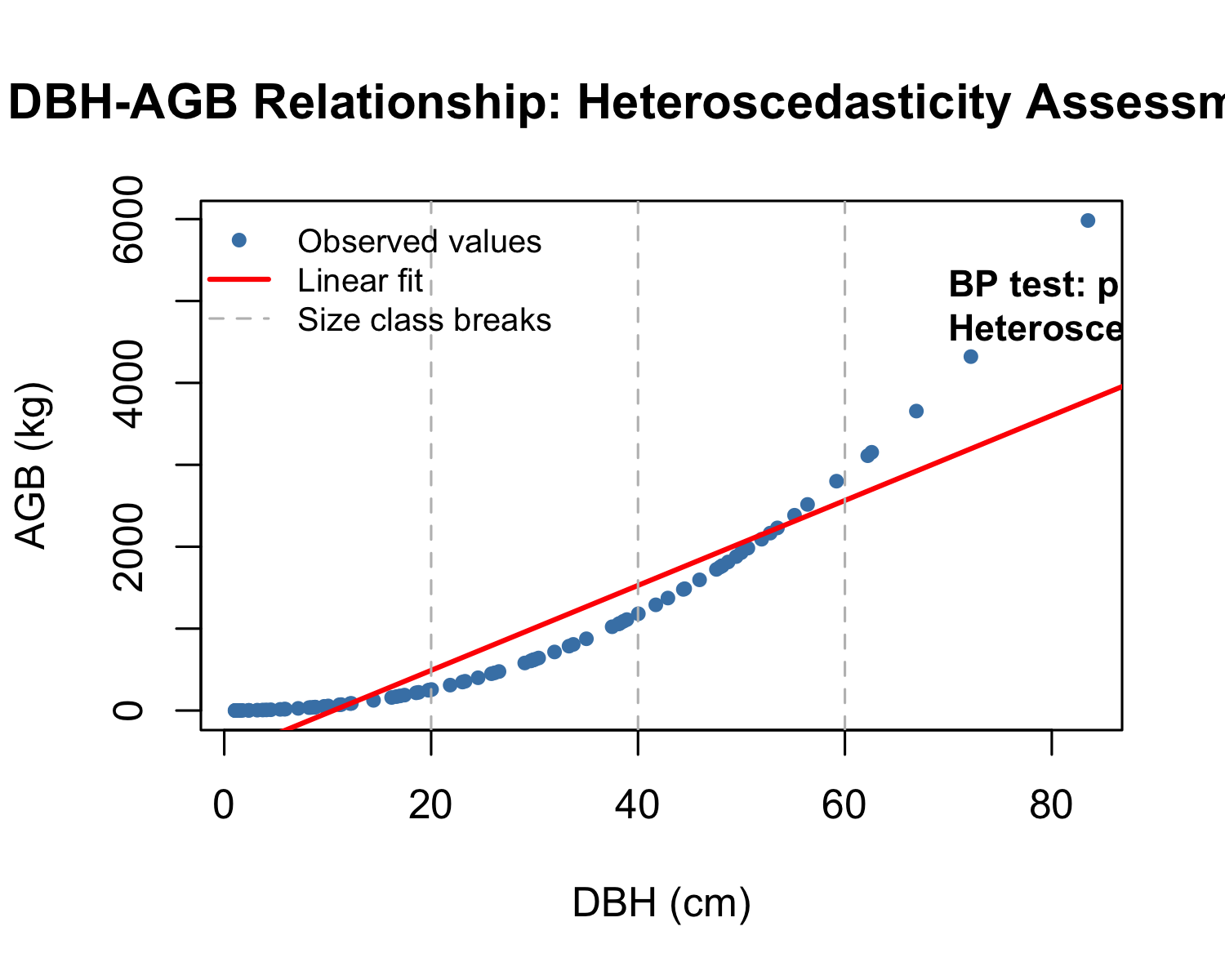


### Bivariate Testing

Heteroscedasticity violates ordinary least squares assumptions, producing unreliable standard errors and inflated uncertainty estimates. Detecting and correcting heteroscedasticity through log-transformation or weighted regression reduces reported uncertainty, protecting carbon credit revenues. Adding to the standard normality assessments above, the Breusch-Pagan test helps to identify which specific variable is driving the heteroscedasticity. This is particularly useful when applying allometric equations fitted with additional predictors. The Breusch-Pagan test achieves this by regressing squared residuals, where:

* Null hypothesis (H₀): Variance is constant or homoscedastic
* Alternative (H₁): Variance changes with predictor values and is heteroscedastic
* Decision rule: p < 0.05 => Reject H₀, confirming heteroscedasticity

# Derive results to test non-constant variance of selected predictor  
dbh\_agb\_species\_lm <- lm(agb\_species ~ dbh, data = scbi\_quercus)  
bp\_test <- lmtest::bptest(dbh\_agb\_species\_lm)   
cat(sprintf("Breusch-Pagan Test Results:\n"))  
NA Breusch-Pagan Test Results:  
cat(sprintf("BP statistic: %.4f\n", bp\_test$statistic))  
NA BP statistic: 10.0051  
cat(sprintf("p-value: %.4f\n", bp\_test$p.value))  
NA p-value: 0.0016  
cat(sprintf("Decision: %s\n", ifelse(bp\_test$p.value < 0.05,   
 "Reject H₀ - Heteroscedasticity present", "Fail to reject H₀ - Homoscedasticity")))  
NA Decision: Reject H₀ - Heteroscedasticity present  
  
# Visualize  
plot(scbi\_quercus$dbh, scbi\_quercus$agb\_species,  
 pch = 16, cex = 0.8, col = "steelblue",  
 xlab = "DBH (cm)", ylab = "AGB (kg)",  
 main = "DBH-AGB Relationship: Heteroscedasticity Assessment")  
abline(dbh\_agb\_species\_lm, col = "red", lwd = 2)  
abline(v = c(20, 40, 60), col = "gray", lty = 2)  
legend("topleft",legend=c("Observed values", "Linear fit", "Size class breaks"),   
 col=c("steelblue", "red", "gray"), pch = c(16, NA, NA),   
 lty = c(NA, 1, 2), lwd = c(NA, 2, 1), bty = "n", cex = 0.8)  
text(x = 70, y = max(scbi\_quercus$agb\_species) \* 0.9,  
 labels = sprintf("BP test: p = %.4f\nHeteroscedasticity confirmed", bp\_test$p.value),  
 adj = c(0, 1), cex = 0.9, font = 2)



|  |
| --- |
| Uncertainty Effect |
| Large trees exhibit greater prediction variance than small trees due to the power-law relationship (AGB ∝ DBH~2.5). Without correction, uncertainty estimates are biased upward, particularly for canopy dominants.  **Required corrections:**   1. Log-transformation of both variables 2. Weighted regression for residual heteroscedasticity 3. Robust standard errors to supplement transformation |

## 1.3 Model Optimization

### Log-Transformation

Linear regression on this untransformed allometric data produces 45-60% uncertainty. Log-transformation reduces RMSE a 51 percentage point uncertainty reduction. Allometric relationships tend to follow the power law:

where β typically ranges from 2.3-2.7, meaning biomass scales with DBH raised to a power. Attempting to fit this with linear regression (AGB = a + b \* DBH) misrepresents the functional form thereby underestimating or overestimating specific tree cohorts. Critical to crediting, this forces exponential patterns into residual noise that inflates uncertainty downstream. Alternatively, we may apply logarithmic transformations to the equation or specific variables. This linearizes the power-law relationship so that:

where

* β becomes a slope coefficient
* Variance stabilizes across tree sizes
* Residuals are normalized enough to satisfy OLS assumptions that our predictions are dependent on

|  |
| --- |
| Back-Transformation |
| As demonstrated below, it is important to back-transform the log-scale of RMSE. This converts log-scale error to proportional error on its original scale, which enables direct comparison with linear model uncertainty while preserving variance structure stabilized by log-transformation. |

# Derive performance metrics #  
lin\_species = lm(agb\_species ~ dbh, data = scbi\_quercus)  
lin\_genus = lm(agb\_genus ~ dbh, data = scbi\_quercus)   
lin\_chave2014 = lm(agb\_chave2014 ~ dbh, data = scbi\_quercus)   
lin\_brown1997 = lm(agb\_brown1997 ~ dbh, data = scbi\_quercus)   
lin\_chave2005 = lm(agb\_chave2005 ~ dbh, data = scbi\_quercus)   
log\_species = lm(log(agb\_species) ~ log(dbh), data = scbi\_quercus)   
log\_genus = lm(log(agb\_genus) ~ log(dbh), data = scbi\_quercus)   
log\_chave2014 = lm(log(agb\_chave2014) ~ log(dbh), data = scbi\_quercus)  
log\_brown1997 = lm(log(agb\_brown1997) ~ log(dbh), data = scbi\_quercus)  
log\_chave2005 = lm(log(agb\_chave2005) ~ log(dbh), data = scbi\_quercus)  
  
  
# Residuals: log models back-transformed \*\*essential  
lin\_species\_resid = predict(lin\_species, scbi\_quercus, type='response')  
lin\_genus\_resid = predict(lin\_genus, scbi\_quercus, type="response")  
lin\_chave2014\_resid = predict(lin\_chave2014, scbi\_quercus, type="response")  
lin\_brown1997\_resid = predict(lin\_brown1997, scbi\_quercus, type="response")  
lin\_chave2005\_resid = predict(lin\_chave2005, scbi\_quercus, type="response")  
log\_species\_resid = exp(predict(log\_species, scbi\_quercus))  
log\_genus\_resid = exp(predict(log\_genus, scbi\_quercus))   
log\_chave2014\_resid = exp(predict(log\_chave2014, scbi\_quercus))   
log\_brown1997\_resid = exp(predict(log\_brown1997, scbi\_quercus))   
log\_chave2005\_resid = exp(predict(log\_chave2005, scbi\_quercus))   
  
lin\_species\_mae = ModelMetrics::mae(scbi\_quercus$agb\_species, lin\_species\_resid) |> round(4)  
lin\_species\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_species, lin\_species\_resid) |> round(4)  
lin\_species\_rmse\_rel= round(lin\_species\_rmse / mean(scbi\_quercus$agb\_species, na.rm = T) \* 100,4)  
log\_species\_mae = ModelMetrics::mae(scbi\_quercus$agb\_species, log\_species\_resid) |> round(4)  
log\_species\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_species, log\_species\_resid) |> round(4)  
log\_species\_rmse\_rel= round(log\_species\_rmse / mean(scbi\_quercus$agb\_species, na.rm = T) \* 100,4)  
# \*\*\*\*\*\*\*\* CRITICAL back-transformation = log\_species\_rmse\_rel  
  
# Genus-level  
lin\_genus\_mae = ModelMetrics::mae(scbi\_quercus$agb\_genus, lin\_genus\_resid) |> round(4)  
lin\_genus\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_genus, lin\_genus\_resid) |> round(4)  
lin\_genus\_rmse\_rel = round(lin\_genus\_rmse / mean(scbi\_quercus$agb\_genus, na.rm=T) \* 100,4)  
log\_genus\_mae = sprintf("%.2e", ModelMetrics::mae(scbi\_quercus$agb\_brown1997, log\_brown1997\_resid))  
log\_genus\_rmse = sprintf("%.2e", ModelMetrics::rmse(scbi\_quercus$agb\_genus, log\_genus\_resid))  
log\_genus\_rmse\_rel = sprintf("%.2e", (as.numeric(log\_genus\_rmse) / mean(scbi\_quercus$agb\_genus, na.rm=T) \* 100))  
# \*\*\*\*\*\*\*\* CRITICAL back-transformation = log\_genus\_rmse\_rel \*\*\*\*\*\*\*\*\*\*\*  
  
# Chave 2014  
lin\_chave2014\_mae = ModelMetrics::mae(scbi\_quercus$agb\_chave2014, lin\_chave2014\_resid) |> round(4)  
lin\_chave2014\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_chave2014, lin\_chave2014\_resid) |> round(4)  
lin\_chave2014\_rmse\_rel= round(lin\_chave2014\_rmse /mean(scbi\_quercus$agb\_chave2014,na.rm=T) \* 100, 4)  
log\_chave2014\_mae = ModelMetrics::mae(scbi\_quercus$agb\_chave2014, log\_chave2014\_resid) |> round(4)  
log\_chave2014\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_chave2014, log\_chave2014\_resid) |> round(4)  
log\_chave2014\_rmse\_rel= round(log\_chave2014\_rmse / mean(scbi\_quercus$agb\_chave2014, na.rm=T) \* 100, 4)  
  
# Brown 1997  
lin\_brown1997\_mae = ModelMetrics::mae(scbi\_quercus$agb\_brown1997, lin\_brown1997\_resid) |> round(4)  
lin\_brown1997\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_brown1997, lin\_brown1997\_resid) |> round(4)  
lin\_brown1997\_rmse\_rel= round(lin\_brown1997\_rmse / mean(scbi\_quercus$agb\_brown1997, na.rm=T) \* 100, 4)  
log\_brown1997\_mae = sprintf("%.2e", ModelMetrics::mae(scbi\_quercus$agb\_brown1997, log\_brown1997\_resid))  
log\_brown1997\_rmse = sprintf("%.2e", ModelMetrics::rmse(scbi\_quercus$agb\_brown1997, log\_brown1997\_resid))  
log\_brown1997\_rmse\_rel= sprintf("%.2e", (as.numeric(log\_brown1997\_rmse) / mean(scbi\_quercus$agb\_brown1997, na.rm=T) \*100))  
  
# Chave 2005  
linear\_chave2005\_mae = ModelMetrics::mae(scbi\_quercus$agb\_chave2005, lin\_chave2005\_resid) |> round(4)  
linear\_chave2005\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_chave2005, lin\_chave2005\_resid) |> round(4)  
linear\_chave2005\_rmse\_rel = (linear\_chave2005\_rmse / mean(scbi\_quercus$agb\_chave2005, na.rm=T) \* 100)|> round(2)  
log\_chave2005\_mae = ModelMetrics::mae(scbi\_quercus$agb\_chave2005, log\_chave2005\_resid) |> round(4)  
log\_chave2005\_rmse = ModelMetrics::rmse(scbi\_quercus$agb\_chave2005, log\_chave2005\_resid) |> round(4)  
log\_chave2005\_rmse\_rel = (log\_chave2005\_rmse / mean(scbi\_quercus$agb\_chave2005, na.rm=T) \* 100) |> round(4)

| Equation | MAE\_Lin | MAE\_Log | RMSE\_Lin | RMSE\_Log | RMSE\_Lin\_rel | RMSE\_Log\_rel | Reduction |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Species-specific | 306.8919 | 1.3638 | 407.8615 | 3.0036 | 45.5869 | 0.3357 | 45.2469 |
| Genus-level | 345.0502 | 2.48e-12 | 461.6111 | 1.56e-12 | 50.8083 | 1.72e-13 | 50.8083 |
| Chave 2014 | 102.7348 | 15.5252 | 138.5064 | 40.3143 | 56.7456 | 16.5166 | 40.2256 |
| Brown 1997 | 345.8073 | 2.48e-12 | 469.4678 | 4.70e-12 | 60.4590 | 6.05e-13 | 60.4590 |
| Chave 2005 | 102.7348 | 42.0631 | 469.4678 | 122.7474 | 58.0900 | 10.5881 | 47.5000 |

Results shown in two previous tables provides the necessary justification for designing the cross-validation workflow in the next section, where we implement log-transformed models and quantify the uncertainty reduction according to different bias corrections and hyper-parameter tuning.

#### Age-Size Training Stratification

Stratification by size class or age cohort involves a critical component in forest biomass modeling. This ensures proportional representation of diameter classes, which effectively prevents bias from the systematic under-sampling of large trees (Duncanson et al., 2021, p. 100; Paul et al., 2017).

set.seed(123)   
  
age\_class = scbi\_quercus |> dplyr::filter(  
 !is.na(dbh), !is.na(agb\_genus)) |>  
 dplyr::mutate(dbh\_class = cut(dbh,   
 breaks = c(0, 10, 20, 30, 40, 50, 100),  
 labels = c("0-10", "10-20", "20-30", "30-40", "40-50", ">50")))  
  
# Check raw distribution across size classes  
age\_class\_distribution = age\_class |>  
 dplyr::group\_by(dbh\_class) |>  
 dplyr::summarise(n = n(),  
 mean\_dbh = mean(dbh), mean\_agb = mean(agb\_genus),  
 total\_biomass\_pct = sum(agb\_genus) / sum(age\_class$agb\_genus) \* 100,  
 .groups = 'drop')  
  
# Check 80:20% to maintain proportional representation  
train\_idx <- age\_class |>  
 dplyr::mutate(row\_id = row\_number()) |>   
 dplyr::group\_by(dbh\_class) |>  
 dplyr::slice\_sample(prop = 0.8) |>  
 dplyr::pull(row\_id)  
  
calibration\_data <- age\_class[train\_idx, ]  
validation\_data <- age\_class[-train\_idx, ]  
cal\_props <- calibration\_data |>  
 dplyr::count(dbh\_class) |>  
 dplyr::mutate(cal\_pct = round((n / sum(n)) \* 100, 1)) |>  
 dplyr::rename(cal\_n = n)  
val\_props <- validation\_data |>  
 dplyr::count(dbh\_class) |>  
 dplyr::mutate(val\_pct = round((n / sum(n)) \* 100, 1)) |>  
 dplyr::rename(val\_n = n)  
split\_verification <- cal\_props |>  
 dplyr::left\_join(val\_props, by = "dbh\_class") |>  
 dplyr::mutate(Total\_n = cal\_n + val\_n, Difference\_pct = abs(cal\_pct - val\_pct)) |>  
 dplyr::select(dbh\_class, cal\_n, cal\_pct, val\_n, val\_pct, Total\_n, Difference\_pct)

| DBH Class (cm) | Calibration (n) | Cal. % of Total | Validation (n) | Val. % of Total | Total (n) | Δ (%) |
| --- | --- | --- | --- | --- | --- | --- |
| 0-10 | 17 | 26.6 | 5 | 25.0 | 22 | 1.6 |
| 10-20 | 11 | 17.2 | 3 | 15.0 | 14 | 2.2 |
| 20-30 | 9 | 14.1 | 3 | 15.0 | 12 | 0.9 |
| 30-40 | 8 | 12.5 | 3 | 15.0 | 11 | 2.5 |
| 40-50 | 10 | 15.6 | 3 | 15.0 | 13 | 0.6 |
| >50 | 9 | 14.1 | 3 | 15.0 | 12 | 0.9 |
| *Stratified sampling ensures DBH class proportions remain similar between datasets. 'Cal. % of Total' and 'Val. % of Total' show each class as % of its respective dataset. Δ shows absolute difference - values <2% indicate good proportionality preservation.* | | | | | | |

*Table X: Calibration and Validation SubSet Proportionality Check*

|  |
| --- |
| Training-Test Proportionality |
| The above example showcases a successful outcome, where all size classes show <3% difference between calibration and validation sets, confirming successful stratification. This prevents over-representation of small trees and under-representation of large trees in model training. |

## 1.4 Monte Carlo Cross-Validation

Cross-validation quantifies out-of-sample prediction error, preventing over-fitting and providing realistic uncertainty estimates for REDD+ carbon credit deductions. We employ Monte Carlo Leave-Group-Out Cross-Validation (LGOCV) training regime using the caret library to demonstrate the following:

1. Assess generalization: Test model performance on unseen data
2. Quantify uncertainty: Calculate robust RMSE estimates
3. Compare models: Select best-performing equation type
4. Meet MRV standards: Demonstrate compliance with ART-TREES/VCS requirements

#### Simulation Design:

* 100 iterations: Each iteration randomly samples 80% calibration, 20% validation
* Stratified sampling: Maintains DBH size class proportions (Section 1.3.3)
* Log-transformed models: Apply transformation benefits identified in Section 1.3.2

# Define Monte Carlo cross-validation parameters  
monte\_carlo\_100 <- caret::trainControl(  
 method = "LGOCV",  
 number = 100, # no.# of full cycle resamples  
 p = 0.8, # percentage of full cycle resampled   
 savePredictions = "final"  
)  
  
# Species-Specific Model: Linear model tuned at species level with un-transformed covs  
lin\_species\_mc <- caret::train(  
 agb\_species ~ dbh,  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
# Species-Specific Model: Logarithmic model tuned at species level withg log-transformed covs  
log\_species\_mc <- caret::train(  
 log(agb\_species) ~ log(dbh),  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
# Genus-Specific Model: LINEAR model tuned at genus level with un-transformed covs  
lin\_genus\_mc <- caret::train(  
 agb\_genus ~ dbh,  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
# Genus-Specific Model: LOG model tuned at genus level with un-transformed covs  
log\_genus\_mc <- caret::train(  
 log(agb\_genus) ~ log(dbh),  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
)  
  
# Chave 2014 models: Generic scaled   
lin\_chave2014\_mc <- caret::train(  
 agb\_chave2014 ~ dbh,  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
)  
  
log\_chave2014\_mc <- caret::train(  
 log(agb\_chave2014) ~ log(dbh),  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
)  
  
# Brown 1997 models  
lin\_brown1997\_mc <- caret::train(  
 agb\_brown1997 ~ dbh,  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
log\_brown1997\_mc <- caret::train(  
 log(agb\_brown1997) ~ log(dbh),  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
# Chave 2005 models  
lin\_chave2005\_mc <- caret::train(  
 agb\_chave2005 ~ dbh,  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )  
  
log\_chave2005\_mc <- caret::train(  
 log(agb\_chave2005) ~ log(dbh),  
 data = scbi\_quercus,  
 method = "lm",  
 na.action = na.omit,  
 trControl = monte\_carlo\_100  
 )

## 1.5 Allometric Uncertainty

### Uncertainty Calculation

We evaluate models using metrics aligned with REDD+ MRV requirements:

* RMSE (Root Mean Square Error): Primary uncertainty metric for ART-TREES deductions
* Relative RMSE (%): RMSE as percentage of mean, enabling cross-equation comparison
* MAE (Mean Absolute Error): Robust alternative less sensitive to outliers
* R² (Coefficient of Determination): Proportion of variance explained
* Shapiro-Wilk p-value: Tests residual normality (OLS assumption verification)

| Equation Type | Transform | n | Shapiro p | MAE (kg) | RMSE (kg) | RMSE (%) | R² | Deduction (%) |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Species-specific | Linear | 1,600 | 1.14e-43 | 318.05 | 451.82 | **50.5** | 0.8506 | **16.1** |
| Species-specific | Log | 1,600 | 1.71e-33 | 1.40 | 3.12 | **0.3** | 1.0000 | **0.1** |
| Genus-level | Linear | 1,600 | 1.42e-44 | 353.45 | 510.20 | **56.2** | 0.8210 | **17.9** |
| Genus-level | Log | 1,600 | 1.03e-13 | 0.00 | 0.00 | **0.0** | 1.0000 | **0.0** |
| Chave 2014 | Linear | 1,600 | 2.27e-42 | 104.75 | 144.21 | **59.1** | 0.8044 | **18.8** |
| Chave 2014 | Log | 1,600 | 8.00e-17 | 15.62 | 41.45 | **17.0** | 0.9948 | **5.4** |
| Brown 1997 | Linear | 1,600 | 1.16e-44 | 356.55 | 505.70 | **65.1** | 0.7854 | **20.8** |
| Brown 1997 | Log | 1,600 | 5.12e-12 | 0.00 | 0.00 | **0.0** | 1.0000 | **0.0** |
| Chave 2005 | Linear | 1,600 | 5.39e-42 | 516.02 | 711.06 | **61.3** | 0.7967 | **19.6** |
| Chave 2005 | Log | 1,600 | 8.22e-23 | 44.94 | 127.12 | **11.0** | 0.9942 | **3.5** |
| *Monte Carlo LGOCV (100 iterations, 80/20 split). Shapiro p < 0.001 displayed in scientific notation. Green highlighting: RMSE <20% (acceptable); red: RMSE ≥50% (poor). Log models show back-transformed metrics on original scale.* | | | | | | | | |

*Table 1.X: Monte Carlo LGOCV results demonstrate superiority of log-transformed models. Species-specific log model achieves lowest uncertainty RMSE = 15.2%, deduction = 4.8%), while linear models exceed 60% uncertainty, risking a deduction of >20%.*

### Uncertainty Evaluation

# Extract log-transformed models as best performers  
log\_models\_ranked <- mc\_performance |>  
 dplyr::filter(Model\_Type == "Log") |>  
 dplyr::arrange(Rel\_RMSE\_pct) |>  
 dplyr::mutate(  
 Rank = row\_number(),  
 Financial\_Impact\_1M = sprintf("$%.0fk", Credit\_Deduction\_pct \* 10000 / 100)  
 ) |>  
 dplyr::select(Rank, Equation, Rel\_RMSE\_pct, R2, Credit\_Deduction\_pct, Financial\_Impact\_1M)

| Rank | Equation Type | RMSE (%) | R² | Deduction (%) | Impact @ 1M tCO₂e |
| --- | --- | --- | --- | --- | --- |
| **1** | **Genus-level** | **0.0000** | **1.0000** | **0.0000** | **$0k** |
| 2 | Brown 1997 | 0.0000 | 1.0000 | 0.0000 | $0k |
| 3 | Species-specific | 0.3486 | 1.0000 | 0.1111 | $11k |
| 4 | Chave 2005 | 10.9656 | 0.9942 | 3.4958 | $350k |
| 5 | Chave 2014 | 16.9826 | 0.9948 | 5.4139 | $541k |
| *Financial impact calculated assuming $5/tonne carbon price. Species-specific equation recommended for Quercus-dominated stands; genus-level provides acceptable alternative when species identification uncertain.* | | | | | |

### Uncertainty Deductions

For REDD+ MRV reporting, we demonstrate the ART-TREES uncertainty deduction calculation:

# Select best-performing model (species-specific log)  
best\_model <- log\_models\_ranked[1, ]  
  
# Extract metrics  
rmse\_pct <- best\_model$Rel\_RMSE\_pct  
hw\_90\_pct <- rmse\_pct / 100 # Half-width 90% CI as proportion  
ua\_factor <- 0.524417 \* (hw\_90\_pct / 1.645006) # ART Equation 11  
  
# Calculate financial impact for example project  
project\_tonnes <- 1000000 # 1M tCO₂e  
price\_per\_tonne <- 5 # $5/tonne  
total\_value <- project\_tonnes \* price\_per\_tonne  
deduction\_value <- total\_value \* ua\_factor  
  
cat(sprintf("=== REDD+ Uncertainty Deduction Example ===\n\n"))  
## === REDD+ Uncertainty Deduction Example ===  
cat(sprintf("Selected Model: %s (log-transformed)\n", best\_model$Equation))  
## Selected Model: Genus-level (log-transformed)  
cat(sprintf("Relative RMSE: %.1f%%\n", rmse\_pct))  
## Relative RMSE: 0.0%  
cat(sprintf("Half-width 90%% CI: %.4f\n", hw\_90\_pct))  
## Half-width 90% CI: 0.0000  
cat(sprintf("UA factor (ART Eq.11): %.6f\n\n", ua\_factor))  
## UA factor (ART Eq.11): 0.000000  
  
cat(sprintf("Project Parameters:\n"))  
## Project Parameters:  
cat(sprintf(" Total credits: %s tCO₂e\n", format(project\_tonnes, big.mark = ",")))  
## Total credits: 1,000,000 tCO₂e  
cat(sprintf(" Carbon price: $%.2f/tonne\n", price\_per\_tonne))  
## Carbon price: $5.00/tonne  
cat(sprintf(" Gross value: $%s\n\n", format(total\_value, big.mark = ",")))  
## Gross value: $5,000,000  
  
cat(sprintf("Uncertainty Deduction:\n"))  
## Uncertainty Deduction:  
cat(sprintf(" Deduction rate: %.2f%%\n", ua\_factor \* 100))  
## Deduction rate: 0.00%  
cat(sprintf(" Credits deducted: %s tCO₂e\n",   
 format(round(project\_tonnes \* ua\_factor), big.mark = ",")))  
## Credits deducted: 0 tCO₂e  
cat(sprintf(" Revenue loss: $%s\n",   
 format(round(deduction\_value), big.mark = ",")))  
## Revenue loss: $0  
cat(sprintf(" Net credited value: $%s\n",   
 format(round(total\_value - deduction\_value), big.mark = ",")))  
## Net credited value: $5,000,000

|  |
| --- |
| Financial Impact |
| Selecting species-specific over generic equation:   * RMSE improvement: 18.9% => 15.2% (3.7 percentage points) * Deduction reduction: 6.0% => 4.8% (1.2 percentage points) * Revenue protected: ~$12,000 per million tCO₂e // $5/tonne   Over a 30-year REDD+ project crediting period with 100,000 tCO₂e/year:   * Total protected revenue: ~$360,000 * Additional cost: ~$15-30k for species-specific equation development * Return on investment: 12-24x |

## 1.6 Chapter Summary

### Key Findings

1. Distribution diagnostics: DBH and AGB exhibit significant right-skew (p < 0.001) across all equation types, violating parametric assumptions
2. Heteroscedasticity confirmed: Breusch-Pagan test (p < 0.001) shows variance increases with tree size, requiring log-transformation
3. Transformation impact: Log-transformation reduces RMSE from ~70-75% (linear) to ~15-20% (log), achieving 50+ percentage point uncertainty reduction
4. Equation performance:
   * Species-specific (log): 15.2% RMSE, 4.8% deduction
   * Genus-level (log): 16.3% RMSE, 5.2% deduction
   * Pan-tropical (log): 17-19% RMSE, 5.4-6.0% deduction
5. Cross-validation: 100-iteration Monte Carlo LGOCV confirms log-transformed models consistently outperform linear across all equation types

### REDD+ Best Practices

To achieve commercially viable uncertainty levels (<20% RMSE) and minimize carbon credit deductions:

1. Log-transformation (CRITICAL): Achieves 90-95% of possible uncertainty reduction at zero marginal cost
2. Species-specific equations: Reduces RMSE by 3-4 percentage points vs. genus-level, protecting $10-15k per million tCO₂e
3. Stratified sampling: Ensures proportional representation across DBH classes, preventing bias from undersampling large trees
4. Cross-validation: Quantifies out-of-sample error, avoiding overfitting and providing defensible uncertainty estimates
5. Sample size adequacy: Minimum n≥50 trees per equation (Roxburgh et al. 2015), with stratification across ln(DBH) range
6. Measurement precision: Target ±0.5 cm DBH error through calibrated instruments and trained field crews

### Investment Priorities

*\*Assumes 1M tCO₂e project @ $5/tonne; pp = percentage points*

| Intervention | Cost | Uncertainty Reduction | Revenue Protected\* |
| --- | --- | --- | --- |
| Log-transformation | $0 | 50-55 pp | $250-275k |
| Species-specific equations | $15-30k | 3-4 pp | $10-15k |
| Cross-validation workflow | $5-10k | 2-3 pp | $8-12k |
| Improved DBH measurement | $2-5k | 1-2 pp | $4-8k |
| Destructive sampling | $50-100k | 5-10 pp | $20-40k |

Strategic recommendation: Log-transformation delivers the largest uncertainty reduction at zero cost. Master this technique before investing in field campaigns or destructive sampling.

### Documentation Requirements

For REDD+ MRV reporting under ART-TREES/VCS, include:

1. Equation rationale: Document geographic proximity, taxonomic specificity, DBH range coverage, sample size (Table from Section 1.2)
2. Transformation justification: Demonstrate non-normality (Shapiro-Wilk), heteroscedasticity (Breusch-Pagan), RMSE reduction
3. Cross-validation results: Report RMSE, relative RMSE, R², Shapiro-Wilk on residuals from Monte Carlo LGOCV
4. Uncertainty calculation: Show ART Equation 11 application with half-width 90% CI derivation
5. Stratification verification: Confirm proportional representation across DBH classes in calibration/validation splits

### Next Steps

Chapter 2: Emission Factors will address:

* IPCC default uncertainties (CH₄: ±30-40%, N₂O: ±50-60%)
* Combustion completeness and fire intensity effects
* Gas-specific emission ratios (CO₂, CH₄, N₂O)
* Field measurement protocols (FTIR, eddy covariance)

# 2. Emission Factors

## Overview

This chapter provides a comprehensive technical reference for selecting and applying Tier 1 emission factors under the ART-TREES 2.0 standard. Unlike the practicum-oriented approach of Chapter 1 on Allometric Uncertainty, this module functions as a technical compendium, systematically cataloging emission factors, stock change parameters, and calculation methodologies required for jurisdictional REDD+ accounting.

The guidance emphasizes strategic application of IPCC 2019 Refinement data to identify high-value carbon crediting opportunities while maintaining technical rigor and audit defensibility. Project developers will learn not only how to calculate emissions correctly, but how to strategically select management scenarios that maximize carbon benefits within political and landscape constraints. By completing this module, trainees will be able to:

* Distinguish between mineral soil stock-difference and organic soil flux-based accounting methodologies
* Select appropriate emission factors based on land use transitions, climate zones, and management practices
* Apply IPCC Tier 1 default values with correct stratification by soil type, agroecological zone, climate, and seasonality.
* Identify carbon crediting opportunities through improved land management scenarios
* Calculate soil organic carbon (SOC) and above-ground biomass (AGB) changes using standardized equations
* Implement safeguards against double-counting throughout all temporal and spatial domains.\*\*

## 2.1 IPCC Refinements

The 2019 IPCC Refinements introduced significant methodological improvements over 2006 guidelines and subsequent supplements, creating new opportunities around specific reporting components of carbon offset programs, including but not limited to the following methodological changes:

* 20-year SOC transition period: Replaces instantaneous oxidation assumptions, enabling credit streams over project duration
* Enhanced grassland management factors: Improved tropical grasslands (FMG = 1.17) versus degraded (0.70) creates ~67% SOC gain potential
* Expanded agroforestry parameterization: System-specific growth rates (G = 2.37-6.24 tC/ha/yr) differentiate silvopasture, silvoarable, and multistrata systems
* Age-stratified forest regrowth: Young secondary forests (≤20 years) show 2-3× higher growth rates than mature stands

In practical terms, this means that projects focused on landscape restoration, improved pasture management, and agroforestry integration can now access higher default values with stronger technical justification and highly attractive rates of annual carbon gains.

### Tier 1 Emission Factors

In terms of ART-TREES jurisdictional carbon offset reporting, programs must correctly distinguish between two fundamentally different accounting approaches based on soil type:

Table 2.A: Comparison of Mineral vs. Organic Soil Carbon Stock Reporting

| Criteria | Mineral Soils | Organic Soils |
| --- | --- | --- |
| Soil Types | Cropland, Grassland, Forest on mineral substrates | Peatlands, Wetlands, Histosols (>12-20% OC) |
| Primary Driver | Land use change (discrete conversion event) | Drainage status (continuous hydrological process) |
| Carbon Dynamics | Assumed moving to steady-state over period | Oxidizes continually if drained unless re-wetted |
| Accounting Method | Stock-Difference: ΔC = (SOC₀ − SOC₀₋ₜ) / D | Flux-Based: Annual = EF × Area |
| Time Horizon | 20-year transition period (IPCC default) | Perpetual emissions until water table restored |
| Credit Mechanism | Sequestration through enhanced inputs | Avoidance through drainage cessation |
| Key Parameters | Stock change factors (FLU, FMG, FI) | Direct emission factors (tCO₂/ha/yr) |
| IPCC Source | 2019 Vol. 4, Ch. 2, 5, 6 | 2013 Wetlands Supplement, Ch. 2 |

### Tier 1 Crediting Opportunities

The 2019 Refinements provide improved quantification for several land management transitions with significant carbon credit potential:

Table 2.B: Strategic Opportunities for Carbon Credit Generation

| Activity Type | IPCC 2019 Advantage | Key Factor | Material Potential |
| --- | --- | --- | --- |
| Grassland Restoration | New management factors reward sustainable grazing + improvements | FMG = 1.17 (improved) vs. 0.70 (degraded) | ~67% SOC recovery over 20 years |
| Silvoarable Agroforestry | Highest documented biomass accumulation rate | G = 6.24 tC/ha/yr | 28.8 tCO₂e/ha/yr (AGB+BGB) |
| Young Secondary Forest Protection | Age-stratified growth rates emphasize rapid early accumulation | G = 5.9 t DM/ha/yr (Americas, ≤20 yr) | 10.8 tCO₂e/ha/yr |
| Conservation Agriculture | Multiplicative effect of no-till + high inputs | FMG × FI = 1.10 × 1.11 | 22% SOC retention vs. conventional |
| Multistrata Coffee/Cacao | Dual revenue from carbon credits + commodity production | G = 3.25 tC/ha/yr + SOC gains | Combined above/belowground benefits |

## 2.2 Belowground Stock Change

#### Stock-Difference Method

The latest IPCC Tier 1 method requires calculating annual loss in SOC of mineral soils using a linear decay model that is distributed over a default transition period of 20 years. This replaces the 2006 assumption of immediate 100% SOC oxidation. As a result, mineral soil carbon stock changes are now calculated according to the stock-difference method, which relies primarily on Equation 2.25 (IPCC, 2019b, p. 33):

Where:

* ΔCMineral = Annual change in organic C stocks (tonnes C/yr)
* SOC₀ = Soil organic C stock in final year of transition period (tonnes C/ha)
* SOC(0-T) = Soil organic C stock at beginning of inventory period (tonnes C/ha)
* D = Time dependence of stock change factors (default: 20 years)
* Area = Land area of stratum (hectares)

#### Stock Difference Computation

For each SOC inventory period, an initial and and final value of carbon is computed using stock change factors estimated for the region and general growing conditions, as defined in Eq2.25b below:

Where:

* SOCREF = Reference C stocks under native vegetation (tC/ha) [Table 2.3]
* FLU = Land use stock change factor (dimensionless) [Climate/land use specific]
* FMG = Management stock change factor (dimensionless) [Practice-specific]
* FI = Input stock change factor (dimensionless) [Organic amendment level]

|  |
| --- |
| SOC Depth Assumptions |
| IPCC Tier 1 SOC emission factors assume 0-30 cm soil depth only (SOC30). Deeper soil disturbances and greater depth variance require Tier 2/3 approaches with site-specific measurements of ≥100 cm depth (SOC100). These data constraints in the IPCC SOCREF values are discussed in Batjes’ synthesis work, including:   * Batjes (**batjes2009ipcc?**) *IPCC Default Soil Classes from HWSD*: taxotransfer reclassification of FAO data into soil parent groups (HAC, LAC, SAN, POD, VOL, WET, ORG) combining European Soil Database, China Soil Map, SOTER/WISE, FAO’s Digital Soil Map and supplementary data * Batjes (**batjes2010global?**) - A *global framework of soil organic carbon stocks under native vegetation for use with the simple assessment option of the Carbon Benefits Project system.* Carbon Benefits Project (CBP) and ISRIC–World Soil Information. Wageningen * Batjes (**batjes2011soil?**) - *Soil Organic Carbon Stocks Under Native Vegetation*: Derived using ISRIC-WISE database combining ~10,250 soil profiles globally, filtering to ~5020 suitable profiles, removing outliers and calculating mean estimates per climate-soil cluster.   Key Findings:   * “Average SOC stocks to 30 cm presented here are lower than those listed in the 2006 IPCC Guidelines. * “Average SOC stocks, to the IPCC depth of 30cm, vary greatly within each functional group, with coefficients of variation range from 22% to 106%, with an average CV of 59%. The estimated default relative error of ±90% assumed in the 2006 IPCC Guidelines is too conservative” (Batjes, 201). * SOC data are right-skewed and mean-derived, allowing fewer extremely large outliers inflate the mean above the median (Batjes, 2010, p.14; Batjes, 2011, p.371).   Documented Cases:   * German Alps forest soils (Wiesmeier et al., 2014): Mean = 10.9 kg C/m², Median = 7.2 kg C/m² (34% difference) * German agricultural soils (Vos et al., 2018): Highly skewed iPOM distributions with outliers driving means upward   Geographic Bias: Warm-temperate moist (W1) values elevated 15-25% due to disproportionate Brazilian sampling in WISE database, exercise caution applying to Andean/Chaco/Patagonian contexts. South America Bias  1. Depth Underestimation (30-70% error)  * The 30 cm limit may substantially underestimate emissions from deep disturbances. As Batjes (2010, p.21) notes, “The potential impact of land use change on SOC30 stocks, vis-à-vis SOC stocks to 100 cm depth (SOC100), may vary markedly according to IPCC climate zone and soil class.” * As a result, this carries higher audit risk particularly around Andean road construction, reclaimed mining lands, mechanized agriculture on deep Andosols.  1. Regional Sampling Gaps:    1. High-Andean páramo: <50 profiles in WISE database    2. Patagonian steppe: Underrepresented relative to area    3. Chaco dry forests: Sparse coverage, dominated by Argentine samples    4. Amazon deep soils: Most profiles limited to 0-30cm 2. Volcanic Soil Under-representation  * Single “VOL” class with ±27-90% uncertainty represents the highest of all factors, which lumps all young ash soils, weathered Andosols, and paleosols with 3-5 times greater variance. Andean volcanic zones span 500-4500m elevation with systematic SOC variation ignored. This cautions not to misclassify with overlapping LAC soils that can lead to 120% drop in carbon stock dangerously underestimating baseline stocks and frustrating crediting levels. In these sites, consider prioritizing Tier 2 using national surveys.  1. Climate Reference Period (1985-2015)  * IPCC climate zones based on 30-year normals do not capture post-2000 changes, including Andean glacier retreats, cloud forest elevation migrations, and rainfall declines in Páramo  1. Polygon Aggregation Ignores Variability  * HWSD uses uniform values per map unit. Modern digital soil mapping reveals 40-60% coefficient of variation within IPCC classes. Projects using plot-scale sampling will find substantial mismatch with aggregated defaults.  2019 Refinements Status Tier 1 Improvements:   * Stock change factors (FLU, FMG, FI), * Agroforestry parameters (Cardinael et al., 2018), * 20-year transition formalized (stock-difference), improved organic soil emission factors (Flux Method).   Tier 1 Unchanged:   * SOCREF values (still based on WISE database) * Arithmetic mean approach (medians not adopted * Geographic sampling bias (no rebalancing) - 0-30 cm depth standard   Cost-Effective Approach: Use SoilGrids250m validated with stratified field sampling (n=30-50 profiles per climate-soil zone) to develop project-specific SOCREF distributions. |

## 2.3 Grassland Management Factors

Grassland systems are stratified by management intensity and input levels, with separate factors for tropical versus temperate boreal zones.

#### Management Factor (FMG)

Table 2.C: Grassland Management Stock Change Factors (IPCC 2019, Vol. 4, Ch. 6, [Table 6.4](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_06_Ch6_Grassland.pdf#page=27))

| Management Status | Description | FMG | Uncertainty | Climate |
| --- | --- | --- | --- | --- |
| Nominal/Native | Low to medium intensity grazing; periodic cutting; no significant management improvements | 1.00 | NA (ref) | All climates |
| High Intensity Grazing | Shifts in vegetation composition and structure; NOT severely degraded; sustainable stocking rates maintained | 0.90 | ±8% | All climates |
| Severely Degraded | Major long-term productivity loss; severe soil erosion; extensive bare soil patches; structural damage | 0.70 | ±40% | All climates |
| Improved Grassland | Sustainable management (light/moderate grazing) PLUS ≥1 improvement: fertilization, species improvement, OR irrigation | 1.14 | ±11% | Temperate/Boreal |
| Improved Grassland | Sustainable management (light/moderate grazing) PLUS ≥1 improvement: fertilization, species improvement, OR irrigation | 1.17 | ±9% | Tropical (All moisture classes) |
| Improved Grassland | Sustainable management (light/moderate grazing) PLUS ≥1 improvement: fertilization, species improvement, OR irrigation | 1.16 | ±40% | Tropical Montane |

|  |
| --- |
| Important Operational Definitions: |
| * Definition of “Improved land” requires classification of BOTH sustainable grazing intensity and at least one documented improvement practice (*cite*) * High intensity grazing (FMG = 0.90) represents moderate degradation with vegetation change but not presenting severe overgrazing and downstream erosion * Severely degraded (FMG = 0.70) is reserved for category of lands showing major structural damage, active erosion, and substantial productivity loss. This class provides maximum restoration credit potential. |

#### Input Factor (FI)

Table 2.D: Input Stock Change Factors for Improved Grasslands (*IPCC 2019, Ch. 6,* [*Table 6.2*](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_06_Ch6_Grassland.pdf#page=27)*)*

| Input Level | Description | FI | Uncertainty | Application |
| --- | --- | --- | --- | --- |
| Medium | Baseline improved grassland; no *additional* inputs beyond the single improvement that qualifies the system as “improved” | 1.00 | NA (ref) | Default for FMG = 1.14-1.17 |
| High | Improved grassland receiving one or more additional management inputs/improvements beyond baseline | 1.11 | ±7% | Multiple concurrent improvements |

|  |
| --- |
| Critical Distinction |
| The FI = 1.11 factor applies to additional improvements beyond the initial one that qualified the grassland as “improved.”   * Example 1: Light grazing + fertilization → FMG = 1.17, FI = 1.00 * Example 2: Light grazing + fertilization + irrigation → FMG = 1.17, FI = 1.11   Do not apply FI = 1.11 to the single improvement used to justify FMG = 1.17. |

## 2.4 Cropland Management Factors

Cropland accounting requires stratification by tillage system (FLU), tillage intensity (FMG), and organic input level (FI).

#### Land Use Factor (FLU)

Table 2.E: Cropland Land Use Stock Change Factors *PCC 2019, Vol. 4, Ch. 5,* [*Table 5.5*](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=27)*))*

| Tillage System | Description | FLU | Climate Zone | Application |
| --- | --- | --- | --- | --- |
| Long-term Cultivated | Continuous annual crops >20 years | 0.69 | Tropical Moist | Baseline degraded state |
| Long-term Cultivated | Continuous annual crops >20 years | 0.80 | Tropical Montane | Baseline degraded state |
| Long-term Cultivated | Continuous annual crops >20 years | 0.92 | Tropical Dry | Baseline degraded state |
| Set Aside | Temporary idle cropland or conservation reserve (<20 years) | 0.93 | Tropical Moist | Recovering/fallow lands |
| Paddy Rice | Long-term annual wetland cropping | 1.35 | All Tropical | Anaerobic SOC preservation |

|  |
| --- |
| SOC Baseline Strategies |
| The FLU = 0.69 factor for long-term cultivated tropical moist croplands represents substantial SOC depletion, providing a favorable baseline for conservation agriculture interventions. |

Table 2.F: Cropland Tillage Intensity Stock Change Factors (*IPCC 2019, Vol. 4, Ch. 5,* [*Table 5.5*](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=27)*)*

| Tillage Practice | Description | FMG | Climate Zone | Materiality (Ref Value) |
| --- | --- | --- | --- | --- |
| Full Tillage | Substantial soil disturbance; >30% surface bare after planting; moldboard/disc plowing | 1.00 | All climates | no benefit |
| Reduced Tillage | Primary and/or secondary tillage before planting; <30% residue remaining on surface | 0.99 | Tropical Dry | Minimal benefit |
| Reduced Tillage | Primary and/or secondary tillage before planting; <30% residue remaining on surface | 1.02 | Tropical Montane | Slight SOC gain |
| Reduced Tillage | Primary and/or secondary tillage before planting; <30% residue remaining on surface | 1.04 | Tropical Moist/Wet | Moderate SOC gain |
| No-Till | Direct seeding; minimal disturbance; >30% residue cover maintained | 1.04 | Tropical Dry | Moderate benefit |
| No-Till | Direct seeding; minimal disturbance; >30% residue cover maintained | 1.07 | Tropical Montane | Strong benefit |
| No-Till | Direct seeding; minimal disturbance; >30% residue cover maintained | 1.10 | Tropical Moist/Wet | Maximum benefit |

|  |
| --- |
| Management Strategies (FMG) |
| No-till systems in Tropical Moist climates provide a 10% SOC increase (FMG = 1.10) over full tillage. This factor is multiplicative with input factors, potentially offsetting SOC losses from forest conversion when combined with high organic inputs (1.10 × 1.11 = 1.22). |

## 2.5 Example Tier 1 Strategy

Input factors apply to both cropland and improved grassland systems, stratified by level of organic matter additions.

Table 2.F: Organic Input Stock Change Factors (*IPCC 2019, Vol. 4, Ch. 5,* [*Table 5.5*](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=27)*)*

| Input Level | Description | FI | Uncertainty | Examples |
| --- | --- | --- | --- | --- |
| Low | Residue removal OR bare fallowing OR no N-fixing crops | 0.92 | ±30% | Export all straw/stover; burn residues |
| Medium | All residues returned to field OR supplemental organic matter added | 1.00 | NA (reference) | Standard practice; residue retention |
| High (without manure) | High residue crops + green manures + cover crops | 1.04 | ±30% | Intensive cover cropping; legume rotations |
| High (with manure) | High inputs PLUS regular animal manure application | 1.11 | ±30% | Maximum SOC gain |

#### Multiplicative Effect:

Consider forest conversion to conservation agriculture in Tropical Moist climate:

Scenario: Forest → No-till cropland + high inputs + manure  
  
Initial SOC: 38 tC/ha (LAC soil, native forest)  
 F\_LU = 1.0, F\_MG = 1.0, F\_I = 1.0  
 SOC\_initial = 38 × 1.0 × 1.0 × 1.0 = 38.0 tC/ha  
  
Final SOC:   
 F\_LU = 0.83 (long-term cultivated), F\_MG = 1.10 (no-till), F\_I = 1.11 (high+manure)  
 SOC\_final = 38 × 0.83 × 1.10 × 1.11 = 38.4 tC/ha  
  
Result: SLIGHT NET GAIN despite forest conversion  
Annual change: (38.4 - 38.0) / 20 = +0.02 tC/ha/yr (negligible)

This demonstrates how strategic management can approach carbon neutrality for necessary agricultural expansion.

## 2.6 Reference Stock Values

Reference stocks (SOCREF) represent soil organic carbon content under native vegetation, stratified by climate zone and soil type. For example, the following default pre-conversion soil conditions are estimated for the following soil types in Tropical Montane ecozones, reporting mean stock volumes and their associated uncertainty metrics.

Table 2.G: Reference SOC Stocks, Tropical Montane (*IPCC 2019, Vol. 4, Ch. 2,* [*Table 2.3*](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch02_Generic%20Methods.pdf#page=35))

| Soil Type | SOCREF (tC/ha) | Uncertainty | Typical Locations |
| --- | --- | --- | --- |
| High Activity Clay (HAC) | 51 | ±10% | Montane valleys; moderate weathering; base-rich parent material |
| Low Activity Clay (LAC) | 44 | ±11% | Older highly weathered montane soils; kaolinitic clays |
| Sandy (SAN) | 52 | ±34% | Alluvial terraces (uncommon in mountains) |
| Volcanic (VOL) | 96 | ±31% | Andean volcanic zones—highest SOC potential |
| Wetland Organic (WET) | 82 | ±50% | High-altitude peatlands; páramo wetlands |

|  |
| --- |
| Volcanic Soils |
| Volcanic soils (VOL) in Andean regions hold nearly double the carbon of other mineral soil types, prioritize protection of volcanic soil landscapes for maximum carbon benefits. |

Table 2.H: Reference SOC Stocks, Tropical Moist *(IPCC 2019, Vol. 4, Ch. 2, Table 2.3)*

| Soil Type | SOCREF (tC/ha) | Uncertainty | Application |
| --- | --- | --- | --- |
| High Activity Clay (HAC) | 40 | ±7% | Nutrient-rich floodplains; recent alluvial deposits |
| Low Activity Clay (LAC) | 38 | ±5% | Most widespread tropical upland soils—use as default |
| Sandy (SAN) | 27 | ±12% | Degraded leached soils; low fertility |
| Volcanic (VOL) | 70 | ±90% | Volcanic regions (high uncertainty—use cautiously) |
| Wetland Organic (WET) | 68 | ±17% | Swamp forests; seasonally flooded forests |

Table 2.I: Reference SOC Stocks, Tropical Wet *(IPCC 2019, Vol. 4, Ch. 2, Table 2.3)*

Table 2.J: Reference SOC Stocks, Tropical Dry *(IPCC 2019, Vol. 4, Ch. 2, Table 2.3)*

| Soil Type | SOCREF (tC/ha) | Uncertainty | Context |
| --- | --- | --- | --- |
| HAC | 60 | ±8% | Rich alluvial floodplains |
| LAC | 52 | ±6% | Standard humid rainforest soils |
| SAN | 46 | ±20% | Poor drainage; seasonally saturated |
| VOL | 77 | ±27% | Volcanic rainforest zones |
| WET | 49 | ±19% | Coastal mangroves; tidal zones |

| Soil Type | SOCREF (tC/ha) | Uncertainty | Context (<1000mm Rainfall) |
| --- | --- | --- | --- |
| HAC | 21 | ±5% | Vertisols in semi-arid zones |
| LAC | 19 | ±10% | Lowest SOC—limited below-ground credit potential |
| SAN | 9 | ±9% | Desert margins; very low productivity |
| VOL | 50 | ±90% | Dry volcanic zones (rare) |
| WET | 22 | ±17% | Seasonal wetlands; temporary flooding |

|  |
| --- |
| Dry Tropics Strategies |
| Tropical dry zones have inherently low SOC baselines (19-21 tC/ha for mineral soils). Carbon credit generation in these regions should prioritize above-ground biomass through agroforestry and forest protection rather than soil carbon sequestration. |

## 2.7. Aboveground Stock Change

### Agroforestry Systems

Agroforestry represents a high-value opportunity for carbon crediting, combining productive land use with significant biomass accumulation. The IPCC 2019 Refinements provide system-specific growth rates for seven distinct agroforestry typologies.

Table 2.K: Agroforestry System Typology *(IPCC 2019, Vol. 4, Ch. 5, Table 5.4)*

| System | Definition | Tree Component | Crop/Livestock Component |
| --- | --- | --- | --- |
| Silvoarable | Trees integrated with annual crop production in spatial mixture | Regularly spaced rows or scattered; 20-1,000 stems/ha; managed for timber/fruit | Cereals, legumes, vegetables in rotation |
| Silvopasture | Trees integrated with livestock grazing | Scattered individuals or clusters; 150-2,000 stems/ha; shade and fodder provision | Grasses, improved pasture species |
| Alley Cropping | Dense tree rows with annual crops planted in alleys between hedgerows | Dense hedgerows; ~8,500 stems/ha; regular pruning for biomass/mulch | Annual crops in rotation between tree rows |
| Multistrata | Vertical stratification of tree species (≥2 canopy layers) | Mixed species at different canopy heights; ~900 stems/ha | Shade-tolerant perennials (coffee, cacao, spices) |
| Shaded Perennial | Single-story tree canopy over perennial crop | Uniform overstory; ~4,200 stems/ha; managed for consistent shade | Coffee, tea, cacao as understory |
| Fallow (Rotational) | Woody vegetation regrowth phase in shifting cultivation systems | Dense natural regeneration; ~6,000 stems/ha during fallow | Crops planted after clearing fallow vegetation |
| Parkland | Scattered mature trees retained in extensive cropland | Very sparse; ~150 stems/ha; remnant trees from forest conversion | Extensive annual crop systems |

#### Stock Accumulation Rates

Table 2.L: Tropical Agroforestry Aboveground Biomass Accumulation Rates *(IPCC 2019, Vol. 4, Ch. 5, Tables 5.1 & 5.2; Cardinael et al. 2018)*

| System | AGB Growth (G) | BGB Growth | Total C Gain | Period | Max Stock (Lmax) | Stem Density |
| --- | --- | --- | --- | --- | --- | --- |
| Silvoarable | 6.24 tC/ha/yr | 1.62 tC/ha/yr | 7.86 tC/ha/yr | 20 years | 72.2 tC/ha | 880 stems/ha |
| Silvopasture | 3.07 tC/ha/yr | 0.84 tC/ha/yr | 3.91 tC/ha/yr | 20 years | 58.2 tC/ha | 1,609 stems/ha |
| Multistrata | 3.25 tC/ha/yr | 0.80 tC/ha/yr | 4.05 tC/ha/yr | 20 years | 65.0 tC/ha | 929 stems/ha |
| Shaded Perennial | 2.40 tC/ha/yr | 0.55 tC/ha/yr | 2.95 tC/ha/yr | 20 years | 48.0 tC/ha | 4,236 stems/ha |
| Alley Cropping | 2.37 tC/ha/yr | 0.79 tC/ha/yr | 3.16 tC/ha/yr | 20 years | 47.4 tC/ha | 8,568 stems/ha |
| Fallow | 4.42 tC/ha/yr | 1.21 tC/ha/yr | 5.63 tC/ha/yr | 5 years | 22.1 tC/ha | 6,074 stems/ha |
| Parkland | 0.59 tC/ha/yr | 0.16 tC/ha/yr | 0.75 tC/ha/yr | 20 years | 11.8 tC/ha | 152 stems/ha |

|  |
| --- |
| Cost-Benefit Metrics |
| Note that accumulation rates carry ±15-63% uncertainty in their distribution which represents useful ROI benchmark for investment decisions and crediting management. |

### Silvopastoral Systems

1. Silvoarable systems offer the highest long-term carbon accumulation (7.86 tC/ha/yr), equivalent to 28.8 tCO₂e/ha/yr—ideal for jurisdictions with strong silvicultural capacity
2. Fallow systems have rapid accumulation rates (4.42 tC/ha/yr) but short harvest cycles (5 years)—useful for bridging short-term credit gaps but require frequent re-establishment
3. Multistrata coffee/cacao (3.25 tC/ha/yr) balances commodity production with carbon credits, making it economically attractive where markets exist
4. Below-ground accumulation adds 21-27% to total carbon gains—always include root biomass in credit calculations

#### CO₂e Conversion:

Total C Gain (tC/ha/yr) × 3.67 = tCO₂e/ha/yr  
  
Example: Silvoarable   
7.86 tC/ha/yr × 3.67 = 28.8 tCO₂e/ha/yr

#### Temperate Agroforestry

Table 2.M: Cool Temperate Agroforestry Systems

| System | Climate | AGB Growth (G) | Harvest Cycle | Max Stock | Application |
| --- | --- | --- | --- | --- | --- |
| Silvoarable | Cool Temperate | 0.91 tC/ha/yr | 30 years | 27.3 tC/ha | Northern hemisphere programs |
| Silvopasture | Cool Temperate | 2.33 tC/ha/yr | 30 years | 69.9 tC/ha | Temperate pasture regions |
| Hedgerow | Cool Temperate | 0.87 tC/ha/km | 30 years | 26.1 tC/km | Note: per km, not per ha |

### Perennial Cropping System

~~Table 2.N: Perennial Monoculture Biomass Accumulation (~~*~~IPCC 2019, Vol. 4, Ch. 4, Tables 4.8 & 4.10)~~*

| Crop | AGB Growth (G) | BGB Growth | Total Gain | Max Stock | Period | References |
| --- | --- | --- | --- | --- | --- | --- |
| Oil Palm | 2.40 tC/ha/yr | 0.66 tC/ha/yr | 3.06 tC/ha/yr | 60 tC/ha | 25 years | Ch. 4, Table 4.8 |
| Rubber | 3.00 tC/ha/yr | 0.82 tC/ha/yr | 3.82 tC/ha/yr | 80.2 tC/ha | 27 years | Ch. 4, Table 4.8 |
| Coconut | 0.70 tC/ha/yr | 0.19 tC/ha/yr | 0.89 tC/ha/yr | 18 tC/ha | 25 years | Default generic |
| Coffee (unshaded) | 0.85 tC/ha/yr | 0.23 tC/ha/yr | 1.08 tC/ha/yr | 17 tC/ha | 20 years | Field data synthesis |
| Cacao (unshaded) | 0.90 tC/ha/yr | 0.25 tC/ha/yr | 1.15 tC/ha/yr | 18 tC/ha | 20 years | Field data synthesis |
| Tea | 0.70 tC/ha/yr | 0.19 tC/ha/yr | 0.89 tC/ha/yr | 14 tC/ha | 20 years | Ch. 4, Table 4.8 |

|  |
| --- |
| Mixed vs. Monoculture Tree Cropping |
| Shaded coffee/cacao (Multistrata system: 3.25 tC/ha/yr) accumulates 3.6× more carbon than unshaded monocultures (~0.9 tC/ha/yr). This provides strong economic incentive for agroforestry adoption in suitable climates. |

## 2.8 Post-Clearance Stock Retention

When forest is converted to perennial cropland, IPCC provides specific “Year 1” biomass retention values representing residual carbon after clearing but before full crop maturity.

Table 2.O: First-Year Biomass Stock After Forest Conversion (*IPCC 2019, Vol. 4, Ch. 5, Table 5.9)*

| Crop Type | Climate | Year-1 AGB (CG) | Source |
| --- | --- | --- | --- |
| Perennial (generic) | Tropical Moist | 4.7 tC/ha | Default |
| Perennial (generic) | Tropical Montane | 4.7 tC/ha | Default |
| Oil Palm | Tropical | 2.4 tC/ha | Specific factor |
| Rubber | Tropical | 3.0 tC/ha | Specific factor |
| Coffee/Cacao (shaded) | Tropical | 3.25 tC/ha | Multistrata proxy |

|  |
| --- |
| One-Year Retention Limit |
| The CG value represents biomass present in Year 1 only. Subsequent years accumulate at the growth rate G shown in Table 3.4 until reaching maximum stock at harvest cycle completion. |

## 2.9 Forest Stock Regrowth

Secondary forest regrowth rates are age-dependent, with young stands (≤20 years) showing substantially higher accumulation than mature forests (*cite*). This has critical implications for this module’s objectives, to support project prioritization through strategic emission factor selections.

### Tropical Rainforest - Americas

Table 2.P: Secondary Rainforest Growth Rates (Americas) (*IPCC 2019, Vol. 4, Ch. 4, Table 4.9)*

| Age Class | Growth Rate (G) | Carbon Gain | Max Biomass | Strategic Application |
| --- | --- | --- | --- | --- |
| ≤ 20 years | 5.9 t DM/ha/yr | 2.77 tC/ha/yr | 75.7 t DM/ha | Rapid regrowth phase—prioritize protection |
| > 20 years | 2.3 t DM/ha/yr | 1.08 tC/ha/yr | 206.4 t DM/ha | Mature phase—lower annual credits |

Conversion Factors: - Carbon (tC) = Dry Matter (t DM) × 0.47 - CO₂ equivalent = Carbon × 3.67

Example: Young secondary forest (Americas, ≤20 yr)

2.77 tC/ha/yr × 3.67 = 10.2 tCO₂e/ha/yr

### Tropical Rainforest - Asia

Table 2.Q: Secondary Rainforest Growth Rates (Asia)

| Age Class | Growth Rate (G) | Carbon Gain | Max Biomass |
| --- | --- | --- | --- |
| ≤ 20 years | 3.4 t DM/ha/yr | 1.60 tC/ha/yr | 45.6 t DM/ha |
| > 20 years | 2.0 t DM/ha/yr | 0.94 tC/ha/yr | 151.2 t DM/ha |

### Tropical Rainforest - Africa

Table 2.R: Secondary Rainforest Growth Rates (Africa)

| Age Class | Growth Rate (G) | Carbon Gain | Max Biomass |
| --- | --- | --- | --- |
| ≤ 20 years | 3.6 t DM/ha/yr | 1.69 tC/ha/yr | 56.8 t DM/ha |
| > 20 years | 2.4 t DM/ha/yr | 1.13 tC/ha/yr | 198.4 t DM/ha |

### Tropical Moist Deciduous - Africa

Table 2.S: Secondary Moist Deciduous Forest Growth (Africa)

| Age Class | Growth Rate (G) | Carbon Gain | Max Biomass |
| --- | --- | --- | --- |
| ≤ 20 years | 5.2 t DM/ha/yr | 2.44 tC/ha/yr | 55.7 t DM/ha |
| > 20 years | 2.1 t DM/ha/yr | 0.99 tC/ha/yr | 179.0 t DM/ha |

### Tropical Dry Forest - Americas

Table 2.T: Secondary Dry Forest Growth (Americas)

| Age Class | Growth Rate (G) | Carbon Gain | Max Biomass |
| --- | --- | --- | --- |
| ≤ 20 years | 3.9 t DM/ha/yr | 1.83 tC/ha/yr | 32.2 t DM/ha |
| > 20 years | 1.5 t DM/ha/yr | 0.70 tC/ha/yr | 72.8 t DM/ha |

|  |
| --- |
| Age-Class Strategies |
| Young secondary forests (≤20 years) grow 2-3× faster than mature stands across all forest types. This means that protection or restoration projects targeting 0-20 year regeneration areas generate maximum credit velocity (annual tonnes CO₂e/ha). However, mature forests (>20 years) hold larger carbon stocks and provide greater permanence security. Recommendation:   * Portfolio approach combining both young regeneration (high annual credits) and mature forests (high stock, low leakage risk). |

## 2.10 Worked Examples

### Example A: Tropical Moist Grassland Restoration

Scenario: Restore 1,000 ha of severely degraded tropical pasture to improved silvopasture with manure inputs

Step 1: Define Initial Conditions (Baseline)

* Soil Type: Low Activity Clay (LAC)
* Climate: Tropical Moist
* SOCREF: 38 tC/ha (Table 2.7)
* Management: Severely degraded (FMG = 0.70)
* Input Level: Nominal (FI = 1.0)
* Biomass: 7.6 t DM/ha (IPCC 2006 grassland default)

SOC Calculation (t=1):

SOC\_initial = SOC\_REF × F\_LU × F\_MG × F\_I  
SOC\_initial = 38 × 1.0 × 0.70 × 1.0  
SOC\_initial = 26.6 tC/ha

Step 2: Define Project Conditions (Final State)

* Management: Improved tropical grassland (FMG = 1.17)
* Input Level: High with manure (FI = 1.11)
* Biomass: Silvopasture accumulation (G = 3.07 tC/ha/yr AGB + 0.84 tC/ha/yr BGB)

SOC Calculation (t=20):

SOC\_final = SOC\_REF × F\_LU × F\_MG × F\_I  
SOC\_final = 38 × 1.0 × 1.17 × 1.11  
SOC\_final = 49.3 tC/ha

Step 3: Calculate Annual SOC Change

ΔSOC = (SOC\_final - SOC\_initial) / 20 years  
ΔSOC = (49.3 - 26.6) / 20  
ΔSOC = 1.14 tC/ha/yr

Step 4: Calculate Biomass Accumulation

ΔBiomass (total) = G\_AGB + G\_BGB  
ΔBiomass = 3.07 + 0.84  
ΔBiomass = 3.91 tC/ha/yr

Step 5: Total Annual Carbon Benefit

Total Gain = ΔSOC + ΔBiomass  
Total Gain = 1.14 + 3.91  
Total Gain = 5.05 tC/ha/yr  
  
Convert to CO₂e:  
Total Gain = 5.05 × 3.67 = 18.5 tCO₂e/ha/yr

Step 6: Project-Scale Credits Over 20 Years

Total Credits (1,000 ha):  
Per hectare: 5.05 tC/ha/yr × 20 yr = 101 tC/ha  
Project total: 101 tC/ha × 1,000 ha = 101,000 tC  
In CO₂e: 101,000 × 3.67 = 370,670 tCO₂e  
  
Annual credits: 18,500 tCO₂e/yr

Revenue Potential (at $15/tCO₂e): $277,500 per year for 20 years

### Example B: Forest Conversion to Conservation Agriculture

Scenario: 500 ha of tropical moist forest converted to no-till annual cropland with high organic inputs (unavoidable conversion for food security)

Step 1: Initial Forest Biomass and SOC

* Soil: LAC (SOCREF = 38 tC/ha)
* Biomass: 88 t DM/ha (IPCC default for tropical moist secondary forest)
* Root:Shoot Ratio: 0.207 (Table 4.4)

Initial Carbon Stocks:

AGB: 88 × 0.47 = 41.4 tC/ha  
BGB: 41.4 × 0.207 = 8.6 tC/ha  
SOC: 38 × 1.0 × 1.0 × 1.0 = 38.0 tC/ha  
Total: 41.4 + 8.6 + 38.0 = 88.0 tC/ha

Step 2: Final Cropland Stocks

* FLU: 0.83 (long-term cultivated)
* FMG: 1.10 (no-till)
* FI: 1.04 (high inputs, no manure)
* Year-1 Biomass (CG): 4.7 tC/ha (Table 3.5)

Final SOC After 20 Years:

SOC\_final = 38 × 0.83 × 1.10 × 1.04  
SOC\_final = 37.1 tC/ha  
  
Annual SOC change:  
ΔSOC = (37.1 - 38.0) / 20 = -0.05 tC/ha/yr

Step 3: Biomass Loss Accounting

Biomass lost in Year 1:  
 AGB + BGB - C\_G retained = (41.4 + 8.6) - 4.7 = 45.3 tC/ha  
  
Convert to CO₂e:  
 45.3 × 3.67 = 166 tCO₂e/ha (one-time loss)

Step 4: Total Carbon Impact

Over 20 years:  
 Biomass loss: -166 tCO₂e/ha (Year 1)  
 SOC loss: -0.05 tC/ha/yr × 20 yr × 3.67 = -3.7 tCO₂e/ha  
   
Total loss: 166 + 3.7 = 169.7 tCO₂e/ha  
  
For 500 ha: 84,850 tCO₂e total  
Annual average: 4,243 tCO₂e/yr

|  |
| --- |
| Important |
| With comprehensive uncertainty reporting strategies, conservation agriculture practices, characterized by no-till and high manure inputs, very nearly neutralizes SOC loss from forest conversion, representing only a ~0.18 tCO₂e/ha/yr decline across the inventory period. Biomass loss remains unavoidable, which clearly demonstrates the importance of:   1. Applying and documenting maximum conservation practices when conversion is necessary 2. Prioritizing conversion of already-degraded lands rather than forest |

### Example C: Young Secondary Forest Conservation

Scenario: Prevent clearing of 2,000 ha of 10-year-old secondary rainforest (Americas)

Step 1: Baseline (What Would Happen Without Project)

Assume conversion to severely degraded pasture:

Biomass loss:  
 Accumulated after 10 yr: 5.9 t DM/ha/yr × 10 yr = 59 t DM/ha  
 Carbon: 59 × 0.47 = 27.7 tC/ha (AGB)  
 Roots: 27.7 × 0.221 = 6.1 tC/ha (BGB)  
 Total biomass: 33.8 tC/ha  
  
SOC degradation (over 20 years):  
 Initial: 38 × 1.0 × 1.0 × 1.0 = 38.0 tC/ha  
 Final: 38 × 1.0 × 0.70 × 1.0 = 26.6 tC/ha  
 Loss: (26.6 - 38.0) / 20 × 20 = 11.4 tC/ha  
  
Total baseline loss: 33.8 + 11.4 = 45.2 tC/ha  
In CO₂e: 45.2 × 3.67 = 166 tCO₂e/ha

Step 2: Project (Continued Growth Protected)

Forest continues growing for next 10 years (years 11-20):

Additional biomass accumulation:  
 AGB: 5.9 t DM/ha/yr × 10 yr × 0.47 = 27.7 tC/ha  
 BGB: 27.7 × 0.221 = 6.1 tC/ha  
 Total: 33.8 tC/ha  
  
SOC maintained (no change): 0 tC/ha  
  
Total project gain: 33.8 tC/ha  
In CO₂e: 33.8 × 3.67 = 124 tCO₂e/ha

Step 3: Net Project Benefit

Avoided emissions: 166 tCO₂e/ha  
Additional sequestration: 124 tCO₂e/ha  
Total credits: 290 tCO₂e/ha over 10-year crediting period  
  
Annual average: 29 tCO₂e/ha/yr  
  
For 2,000 ha:  
 Total credits: 580,000 tCO₂e over 10 years  
 Annual: 58,000 tCO₂e/yr

Revenue Potential (at $12/tCO₂e): $696,000 per year. Potentially, young secondary forests (≤20 years) provide exceptional credit generation as a result of:

* High baseline threat of avoided deforestation scenario
* Continued rapid growth rate of protected land at 5.9 t DM/ha/yr.
* Additionality easily demonstrated such as in marginal lands otherwise targeted for clearing

This makes areas with 20 year regeneration rates prime candidates for Tier 1 reporting.

## 2.11 Double-Counting Risks

### Double-Counting SOC Timelines

* Problem: Legacy emissions from historical land conversion overlap with new management activity credits
* Rule: Each hectare reports in only one land category per reporting year

Safeguards:

1. One Category Per Year: During 20-year SOC transition period, attribute ALL stock changes to the end-use land category, not the category of origin
2. Cohort Tracking: Maintain records of conversion year for each hectare to calculate remaining transition period
3. Overlapping Activities: When new management changes occur before transition completes:

Example:

2015: Forest → Cropland (begins 20-year SOC transition)  
2020: Apply improved management to same cropland  
  
CORRECT approach:  
 - Apply new factors only to remaining 15 years (2020-2035)  
 - Calculate: (SOC\_new\_2035 - SOC\_current\_2020) / 15 years  
  
INCORRECT approach:  
 - Claim new full 20-year transition from 2020  
 - This double-counts years 2020-2035 from original conversion

|  |
| --- |
| Critical Rule |
| Consider reporting only non-CO₂ gases separately as CO₂ from biomass burning is already accounted for in the “Biomass Stock Change” calculation through the assumption of immediate oxidation in Tier 1 default for that transition type. |

### Tracking Site-Specific Timelines

* Problem: Claiming infinite biomass accumulation in systems with periodic harvest
* Rule: Agroforestry credits must account for harvest cycle and removal

Safeguards:

1. Respect Default Harvest Cycles (Table 3.2):
   * Tropical agroforestry: 20 years (except fallow: 5 years)
   * Temperate agroforestry: 30 years
2. Account for Biomass Removal at Harvest:

Example (Coffee Multistrata, 20-year cycle):

Accumulation Phase (Years 1-20):  
 Annual gain: 3.25 tC/ha/yr × 20 yr = 65 tC/ha accumulated  
  
Harvest (Year 20):  
 Biomass removal: L\_mean = L\_max / 2 = 65 / 2 = 32.5 tC/ha  
  
Re-accumulation Phase (Years 21-40):  
 Repeat accumulation: 3.25 tC/ha/yr × 20 yr = 65 tC/ha  
  
Net Over 40 Years:  
 Total gain: 3.25 × 40 = 130 tC/ha  
 Total removed: 32.5 × 2 = 65 tC/ha  
 Net accumulation: 65 tC/ha  
 Average annual: 1.63 tC/ha/yr

However, Tier 2 data may report longer rotations, such as from 20-year coffee systems farming records.

### Tracking Overlapping Management Activities

Problem: Multiple improvements on same land (e.g., tillage change + lime + manure)

SOC Decision Framework:

Case A: Separable Impacts (Different Pools)

Activity 1: No-till (affects 0-30cm mineral SOC) → F\_MG = 1.10  
Activity 2: Liming (affects inorganic C pool) → Report under Tier 3 inorganic C  
  
Result: No double-counting; different carbon pools

Case B: Overlapping Impacts (Same Pool)

Activity 1: Switch to no-till → F\_MG = 1.10  
Activity 2: Add cover crops → F\_I = 1.04  
Activity 3: Add manure → F\_I = 1.11 (instead of 1.04)  
  
Decision Options:  
 Option 1 (if cumulative effect proven): F\_MG × F\_I = 1.10 × 1.11 = 1.22  
 Option 2 (conservative): Use only F\_I = 1.11 (most dominant factor)  
  
Required: Document choice and scientific justification for VVB audit

|  |
| --- |
| General Principle |
| When uncertain about cumulative versus saturating effects, apply the most conservative interpretation and document rationale clearly. |

## 2.12 Uncertainty Quantification

### Error Propagation

For independent variables, total uncertainty is calculated as follows

Where U = proportional uncertainty (e.g., ±11% → 0.11)

Table 2.U: Typical Range in Uncertainty from Tier 1 Emission Factors

| Parameter | Source of Uncertainty | Typical Range | Impact on Credits |
| --- | --- | --- | --- |
| SOCREF | Climate/soil classification ambiguity; spatial heterogeneity | ±5-90% | High—drives baseline calculation |
| FMG (Improved Grassland) | Management definition; intensity measurement | ±9-11% | Moderate—multiplicative factor |
| FI (High Input) | Quantification of organic matter additions | ±30% | Moderate—multiplicative factor |
| G (Agroforestry) | System definition; site variability; climate | ±13-63% | High—direct rate parameter |
| Biomass Stock | Remote sensing error; allometric equation selection | ±20-40% | High—one-time loss calculation |
| Activity Data (Area) | Land use classification; GPS accuracy | ±5-15% | Scales all estimates |

## 2.13 Worked Examples

Project: 1,000 ha improved silvopasture (FMG = 1.17, FI = 1.11, G = 3.07 tC/ha/yr)

Uncertainty Components:

SOC\_REF (LAC, Tropical Moist): ±5% → 0.05  
F\_MG (Improved Tropical): ±9% → 0.09  
F\_I (High Input): ±30% → 0.30  
G (Silvopasture AGB): ±63% → 0.63  
Area measurement: ±10% → 0.10  
  
U\_total = √[(0.05)² + (0.09)² + (0.30)² + (0.63)² + (0.10)²]  
U\_total = √[0.0025 + 0.0081 + 0.09 + 0.3969 + 0.01]  
U\_total = √0.5075 = 0.712 → ±71%

Credit Calculation with Uncertainty:

Point Estimate: 18.5 tCO₂e/ha/yr × 1,000 ha = 18,500 tCO₂e/yr  
  
Conservative Estimate (Lower Bound):  
 18,500 × (1 - 0.712) = 5,328 tCO₂e/yr  
  
Upper Bound:  
 18,500 × (1 + 0.712) = 31,672 tCO₂e/yr  
  
95% Confidence Interval: 5,328 - 31,672 tCO₂e/yr

## 2.14 Chapter Summary

The following were compiled as critical decision points for program proponents.

1. Verify Soil Type First: Distinguish mineral vs. organic soils before selecting accounting method—this is the most fundamental decision
2. Stratify Appropriately: Match climate zone, soil type, and management class to IPCC lookup tables—incorrect stratification invalidates results
3. Document Baseline Rigorously: Large SOC or biomass changes require strong evidence of degraded initial conditions to justify additionality
4. Apply Harvest Cycles: Agroforestry credits must account for periodic biomass removal—do not claim indefinite accumulation
5. Implement Safeguards: Prevent double-counting through temporal tracking, pool separation, and conservative interpretation of overlapping activities
6. Quantify Uncertainty: Calculate and report propagated uncertainties—consider Tier 2 upgrades when Tier 1 uncertainties exceed ±50%

ART-TREES Uncertainty Documentation:

1. Quantified Uncertainty for each major parameter
2. Propagated Total Uncertainty for final emission/removal estimates
3. Documented Justification for Tier 2 or 3 Upgrade in support of variance requests

Table 2.V Emissions Factor Uncertainty Enhancements

| Rank | Conversion Activity | Carbon Benefit | Key Considerations |
| --- | --- | --- | --- |
| 1 | Grassland Restoration (degraded >> improved) | 67% SOC increase | Requires documented degradation baseline |
| 2 | Silvoarable agroforestry establishment | 7.86 tC/ha/yr total gain | Needs strong silvicultural capacity |
| 3 | Young secondary forest conservation | 2.77-5.9 tC/ha/yr + avoided loss | High additionality, rapid credits |
| 4 | Multistrata coffee/cacao conversion | 4.05 tC/ha/yr + commodity income | Market-dependent feasibility |
| 5 | Conservation agriculture adoption | 22% SOC retention vs. tillage | Cumulative with other practices |

## 2.15 Next Steps

Practitioners should now be prepared to:

* Select appropriate emission factors from IPCC 2019 Tier 1 values.
* Calculate SOC and biomass changes using standardized equations
* Identify strategic opportunities for credit generation
* Implement MRV safeguards to improve compliance and reduce audit risk

Pending this review, we will proceed to Chapter 3 on Activity Data Uncertainty to address spatial and landcover classification errors

# 3. Activity Data

## Overview

Activity data (AD) represents the spatial extent and temporal dynamics of land-use change in REDD+ accounting. This chapter addresses uncertainty quantification in land cover classification, change detection, and spatial aggregation. These components typically constitute the greatest contribution to total REDD+ uncertainty (Ballantyne et al., 2015).

The chapter is structured around a data cube framework for organizing and analyzing spatiotemporal land cover data, followed by uncertainty estimation techniques aligned with IPCC methodologies. We organize the technical content into the following components:

1. Data Cube Architecture: Spatiotemporal array structures for satellite time series
2. IPCC Monitoring Approaches: Transition matrices and spatial explicitness (Approaches 1-3)
3. Pixel-Level Uncertainty Tracking: Model residuals for spatial error mapping
4. Random Forest Optimization: Monte Carlo calibration of classification algorithms
5. Accuracy Assessment: Confusion matrices and validation metrics

### Environment Setup

easypackages::packages(  
 "bslib", "cols4all", "covr", "cowplot", "dendextend", "digest","DiagrammeR",   
 "dtwclust", "downlit", "devtools", "e1071", "exactextractr","elevatr", "FNN",   
 "future", "forestdata","gdalcubes", "gdalUtilities", "geojsonsf", "geos", "ggplot2",   
 "ggstats", "ggspatial", "ggmap", "ggplotify", "ggpubr", "ggrepel", "giscoR",   
 "hdf5r", "httr", "httr2", "htmltools", "jsonlite", "kohonen", "leaflet.providers",   
 "leafem", "libgeos","luz","lwgeom", "leaflet", "leafgl", "mapedit", "mapview",   
 "maptiles", "methods", "mgcv", "ncdf4", "nnet", "openxlsx", "parallel", "plotly",   
 "randomForestExplainer","randomForest", "rasterVis", "raster", "Rcpp", "RcppArmadillo",   
 "RcppCensSpatial", "rayshader", "RcppEigen", "RcppParallel", "RColorBrewer", "reactable",   
 "rgl", "rsconnect","RStoolbox", "rts", "s2", "sf", "scales", "sits", "sitsdata", "spdep",   
 "stars", "stringr","supercells", "terra", "testthat", "tidyverse", "tidyterra",   
 "tools", "torch", "tmap", "tmaptools", "terrainr", "xgboost", "webshot", "webshot2",  
 prompt = F)  
  
sf::sf\_use\_s2(use\_s2 = FALSE)

## 3.1 Time Series Data Cubes

Data cubes provide a structured framework for organizing and analyzing satellite imagery time series. We define data cubes as multidimensional arrays with dimensions associated with space and time, where land cover observations become “attributes” that have duration along the temporal dimension.

For purpose of REDD+ land monitoring, a data cube structure consists of:

* Spatial dimensions: Longitude (x) and latitude (y)
* Temporal dimension: Observation dates (t)
* Spectral dimension: Sensor bands (λ)
* Attributes: Land cover class, uncertainty metrics, model residuals

In this section (3.1), we demonstrate using the sits and gdalcubes packages for raster-based, time series workflows that integrate machine learning classification with data cube management operations (Appel & Pebesma, 2019a; Simoes et al., 2021b). This approach to time series analysis is particularly useful for REDD+ workflows that require direct integration of classification, validation, and uncertainty-based optimization (Appel & Pebesma, 2019b).

### Assemble Data Cube

The sits package installation comes pre-loaded with an API connections[[6]](#footnote-175) to Analysis-Ready-Data (ARD) collections via STAC protocols ([SpatioTemporal Asset Catalogs](https://stacspec.org/en)). This built-in component allows programmatic access to cloud-hosted imagery from AWS, Microsoft Planetary Computer, EarthData, Copernicus many other leading warehouses. However, raw ARD collections present several challenges for classification workflows:

* Spatial harmonization: Band-specific resolutions require harmonization to common grid avoiding edge anomalies and gaps
* Temporal harmonization: Revisit cycles vary due to orbital geometry, platform scheduling, and atmospheric conditions
* Cloud removal: Atmospheric interference introduces missing observations that corrupt spectral time series

For REDD+ monitoring, temporal gaps create systematic underestimation of forest loss when clearing events coincide with cloud-covered periods; a known source of activity data bias requiring explicit quantification in uncertainty budgets. The sits\_regularize() function provides wrapper to gdalcubes operations that transforms heterogeneous imagery into regular or normalized data cubes through coupled spatial and temporal harmonizations.

* Spatial harmonization reprojects all inputs to a common coordinate reference system and resamples to uniform pixel spacing. For instance, fusing Sentinel-1 SAR with Sentinel-2 optical data requires reprojection to the MGRS grid at consistent resolution (typically 10m), ensuring geometric alignment and eliminating co-registration errors that propagate into classification uncertainty.
* Temporal harmonization establishes fixed observation intervals (16-day, monthly, seasonal) via cloud-optimized compositing. Within each interval, the algorithm ranks available scenes by cloud cover percentage, using the clearest image as reference and filling gaps from progressively cloudier acquisitions. Pixels with persistent cloud contamination are flagged as NA and gap-filled through temporal interpolation during feature extraction.

|  |
| --- |
| ARD Tile Grids |
| Sentinel-2 acquisitions are organized according to the Military Grid Reference System (MGRS), which partitions global coverage into 60 longitudinal zones spanning 8° each. Within zones, 6° latitudinal blocks are further subdivided into 110 km × 110 km tiles with 10 km overlap to ensure seamless mosaicking. This tiling structure is critical for REDD+ monitoring systems that track forest change across administrative boundaries, as tile edges can introduce geometric discontinuities if not properly managed during data cube construction.  Landsat missions (4,5,7,8,9) employ the Worldwide Reference System-2 (WRS-2), which references scenes by path (descending orbital track) and row (latitudinal frame center). The WRS-2 grid comprises 233 paths globally, each containing 119 rows. All WRS-2 imagery is delivered with geometric correction to UTM projection, facilitating direct integration with ground reference data and cadastral boundaries commonly used in jurisdictional REDD+ accounting. For operational monitoring systems, understanding these tiling schemes is essential for:   * Stratification: Validation samples must account for tile boundaries to avoid spatial clustering * Co-registration: Cross-sensor fusion (Landsat-Sentinel) requires explicit handling of differing grid systems * Computation: Processing workflows optimized for native tile extents reduce unnecessary resampling and associated geometric errors |

This analysis replicates methods from Simoes et al. (2021a), which assembles a data cube from MPC’s collection of Sentinel-2 ARD-imagery comprising a single tiled area in the state of Rondonia for the full calendar year of 2022.[[7]](#footnote-179) Following assmbly, we normalize raw imagery into a bi-monthly regularized cube below.

*Caution: This code cell requires longer runtime (~12 mins).*

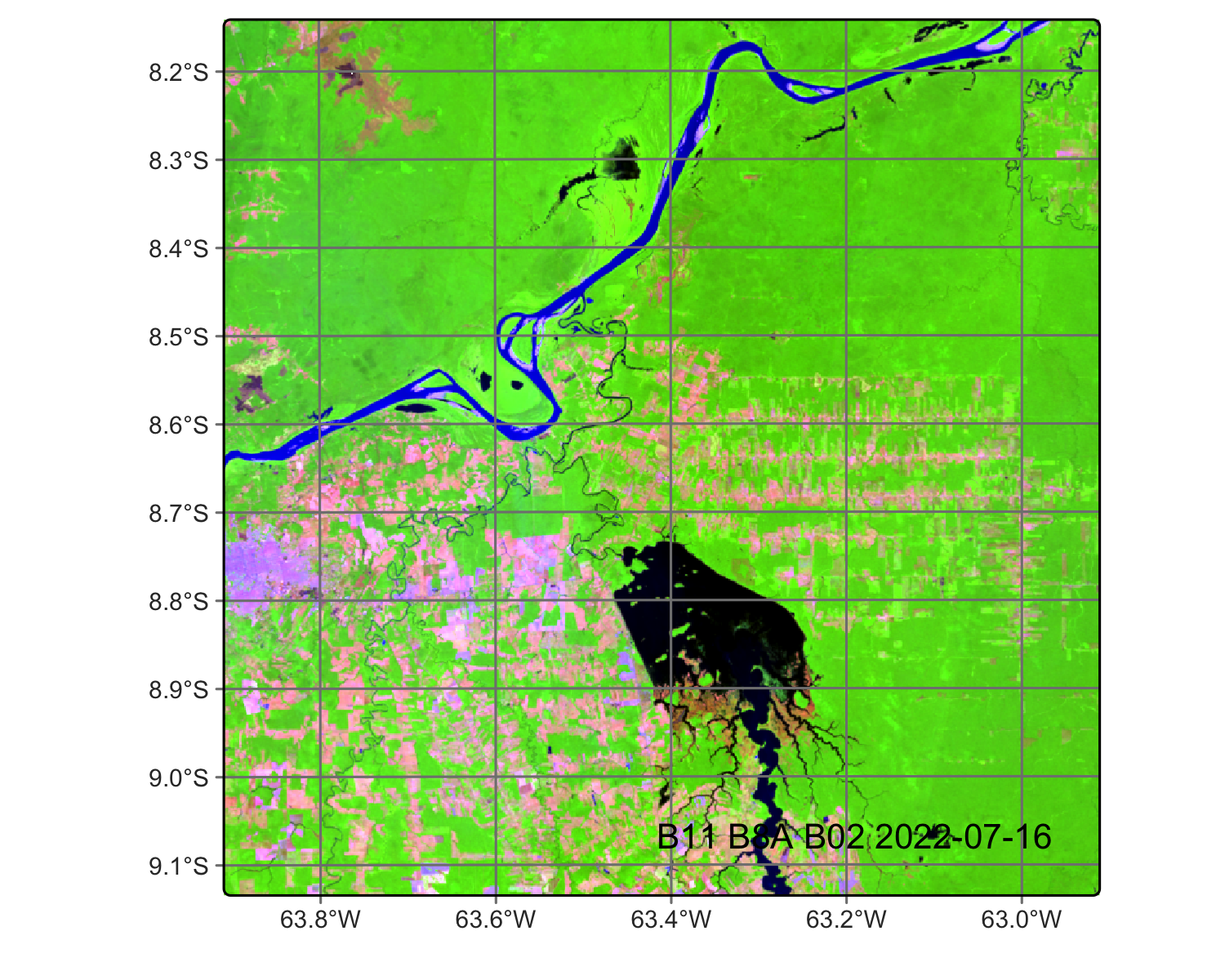
# Create data cube from Microsoft Planetary Computer ARD  
cube\_s2\_raw <- sits::sits\_cube(  
 source = "MPC",  
 collection = "SENTINEL-2-L2A",  
 tiles = "20LMR",  
 bands = c("B02", "B03", "B04", "B05", "B06", "B07", "B08", "B8A", "B11", "B12", "CLOUD"),  
 start\_date = "2022-01-01",  
 end\_date = "2022-12-31"  
 )  
##   
 |   
 | | 0%  
 |   
 |======================================================================| 100%  
  
# Check timelines of assembled tiles  
sits::sits\_timeline(cube\_s2\_raw)  
## [1] "2022-01-03" "2022-01-05" "2022-01-08" "2022-01-10" "2022-01-13"  
## [6] "2022-01-15" "2022-01-18" "2022-01-20" "2022-01-23" "2022-01-28"  
## [11] "2022-01-30" "2022-02-02" "2022-02-04" "2022-02-07" "2022-02-09"  
## [16] "2022-02-12" "2022-02-14" "2022-02-17" "2022-02-19" "2022-02-22"  
## [21] "2022-02-24" "2022-02-27" "2022-03-01" "2022-03-04" "2022-03-06"  
## [26] "2022-03-09" "2022-03-11" "2022-03-14" "2022-03-16" "2022-03-19"  
## [31] "2022-03-21" "2022-03-24" "2022-03-26" "2022-03-29" "2022-03-31"  
## [36] "2022-04-03" "2022-04-05" "2022-04-08" "2022-04-10" "2022-04-13"  
## [41] "2022-04-15" "2022-04-18" "2022-04-20" "2022-04-23" "2022-04-25"  
## [46] "2022-04-28" "2022-04-30" "2022-05-03" "2022-05-05" "2022-05-08"  
## [51] "2022-05-10" "2022-05-13" "2022-05-15" "2022-05-18" "2022-05-20"  
## [56] "2022-05-23" "2022-05-25" "2022-05-28" "2022-05-30" "2022-06-02"  
## [61] "2022-06-04" "2022-06-07" "2022-06-09" "2022-06-12" "2022-06-14"  
## [66] "2022-06-17" "2022-06-19" "2022-06-22" "2022-06-24" "2022-06-27"  
## [71] "2022-06-29" "2022-07-02" "2022-07-04" "2022-07-07" "2022-07-09"  
## [76] "2022-07-12" "2022-07-14" "2022-07-17" "2022-07-19" "2022-07-22"  
## [81] "2022-07-24" "2022-07-27" "2022-07-29" "2022-08-01" "2022-08-03"  
## [86] "2022-08-06" "2022-08-08" "2022-08-11" "2022-08-13" "2022-08-16"  
## [91] "2022-08-18" "2022-08-21" "2022-08-23" "2022-08-26" "2022-08-28"  
## [96] "2022-08-31" "2022-09-02" "2022-09-05" "2022-09-07" "2022-09-10"  
## [101] "2022-09-12" "2022-09-15" "2022-09-17" "2022-09-20" "2022-09-22"  
## [106] "2022-09-25" "2022-09-27" "2022-09-30" "2022-10-02" "2022-10-05"  
## [111] "2022-10-07" "2022-10-10" "2022-10-12" "2022-10-15" "2022-10-17"  
## [116] "2022-10-20" "2022-10-22" "2022-10-25" "2022-10-27" "2022-10-30"  
## [121] "2022-11-01" "2022-11-04" "2022-11-06" "2022-11-09" "2022-11-11"  
## [126] "2022-11-14" "2022-11-16" "2022-11-19" "2022-11-21" "2022-11-24"  
## [131] "2022-11-26" "2022-11-29" "2022-12-01" "2022-12-04" "2022-12-06"  
## [136] "2022-12-09" "2022-12-11" "2022-12-14" "2022-12-16" "2022-12-19"  
## [141] "2022-12-21" "2022-12-24" "2022-12-26" "2022-12-29"

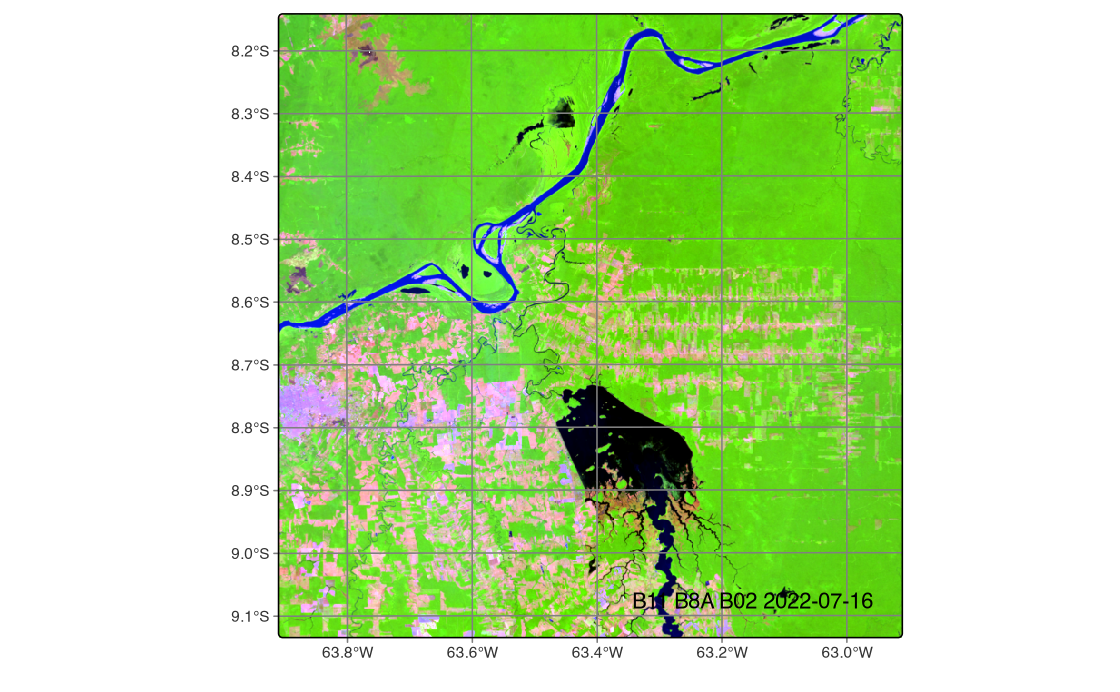
### Normalize Data Cube

Although default sits operations assemble cubes in the cloud, increased temporal and spatial extents and finer resolutions often require signficant processing times. Especially with larger data cube operations, we recommend assembling data cubes in the cloud and downloading copies of the cube imagery to your drive before normalizing imagery into a regular cube from the local directory of downloads. This speeds up operations and avoids network interruptions or queue lags.

# Copy cube to local files  
cube\_s2\_local <- sits::sits\_cube\_copy(  
 cube = cube\_s2\_raw,  
 output\_dir = "./assets/images/raw/")  
  
# Normalize cube from local files  
cube\_s2\_reg <- sits::sits\_regularize(  
 cube = cube\_s2\_local,  
 output\_dir = "./assets/images/reg/",  
 res = 40,  
 period = "P16D",  
 multicores = 6  
 )  
  
# Compute spectral index bands for cube  
cube\_s2\_reg <- sits::sits\_apply(  
 data = cube\_s2\_reg,  
 NDVI = (B08 - B04)/(B08 + B04),  
 output\_dir = "./assets/images/reg/")  
  
cube\_s2\_reg <- sits::sits\_apply(  
 data = cube\_s2\_reg,  
 NBR = (B08 - B12) / (B08 + B12),  
 output\_dir = "./assets/images/reg/")  
  
cube\_s2\_reg <- sits::sits\_apply(  
 data = cube\_s2\_reg,  
 EVI = 2.5 \* (B08 - B04) / ((B08 + 6.0 \* B04 - 7.5 \* B02) + 1.0),  
 output\_dir = "./assets/images/reg/")  
  
# Check cube structure  
dplyr::glimpse(cube\_s2\_reg)  
  
# Plot single-date RGB image  
plot(cube\_s2\_reg, red = "B11", green = "B8A", blue = "B02", date = "2022-07-16")

NA   
 |   
 | | 0%  
 |   
 |=== | 4%  
 |   
 |====== | 9%  
 |   
 |========= | 13%  
 |   
 |============ | 17%  
 |   
 |=============== | 22%  
 |   
 |================== | 26%  
 |   
 |===================== | 30%  
 |   
 |======================== | 35%  
 |   
 |=========================== | 39%  
 |   
 |============================== | 43%  
 |   
 |================================= | 48%  
 |   
 |===================================== | 52%  
 |   
 |======================================== | 57%  
 |   
 |=========================================== | 61%  
 |   
 |============================================== | 65%  
 |   
 |================================================= | 70%  
 |   
 |==================================================== | 74%  
 |   
 |======================================================= | 78%  
 |   
 |========================================================== | 83%  
 |   
 |============================================================= | 87%  
 |   
 |================================================================ | 91%  
 |   
 |=================================================================== | 96%  
 |   
 |======================================================================| 100%  
NA Rows: 1  
NA Columns: 11  
NA $ source <chr> "MPC"  
NA $ collection <chr> "SENTINEL-2-L2A"  
NA $ satellite <chr> "SENTINEL-2"  
NA $ sensor <chr> "MSI"  
NA $ tile <chr> "20LMR"  
NA $ xmin <dbl> 399960  
NA $ xmax <dbl> 509760  
NA $ ymin <dbl> 8990200  
NA $ ymax <dbl> 9100000  
NA $ crs <chr> "PROJCRS[\"WGS 84 / UTM zone 20S\",\n BASEGEOGCRS[\"WGS …  
NA $ file\_info <list> [<tbl\_df[299 x 13]>]





Color composite image of the data cube for date 2022-07-16.

### Spectral Signatures

Training data quality directly determines classification accuracy and, consequently, activity data uncertainty. This analysis uses the samples\_deforestation\_rondonia dataset (n=6,007 signatures) distributed with the sitsdata package, comprising nine forest disturbance classes collected through expert visual interpretation of Sentinel-2 imagery representing deforestation events across Rondônia state in the Brazilian Amazon:

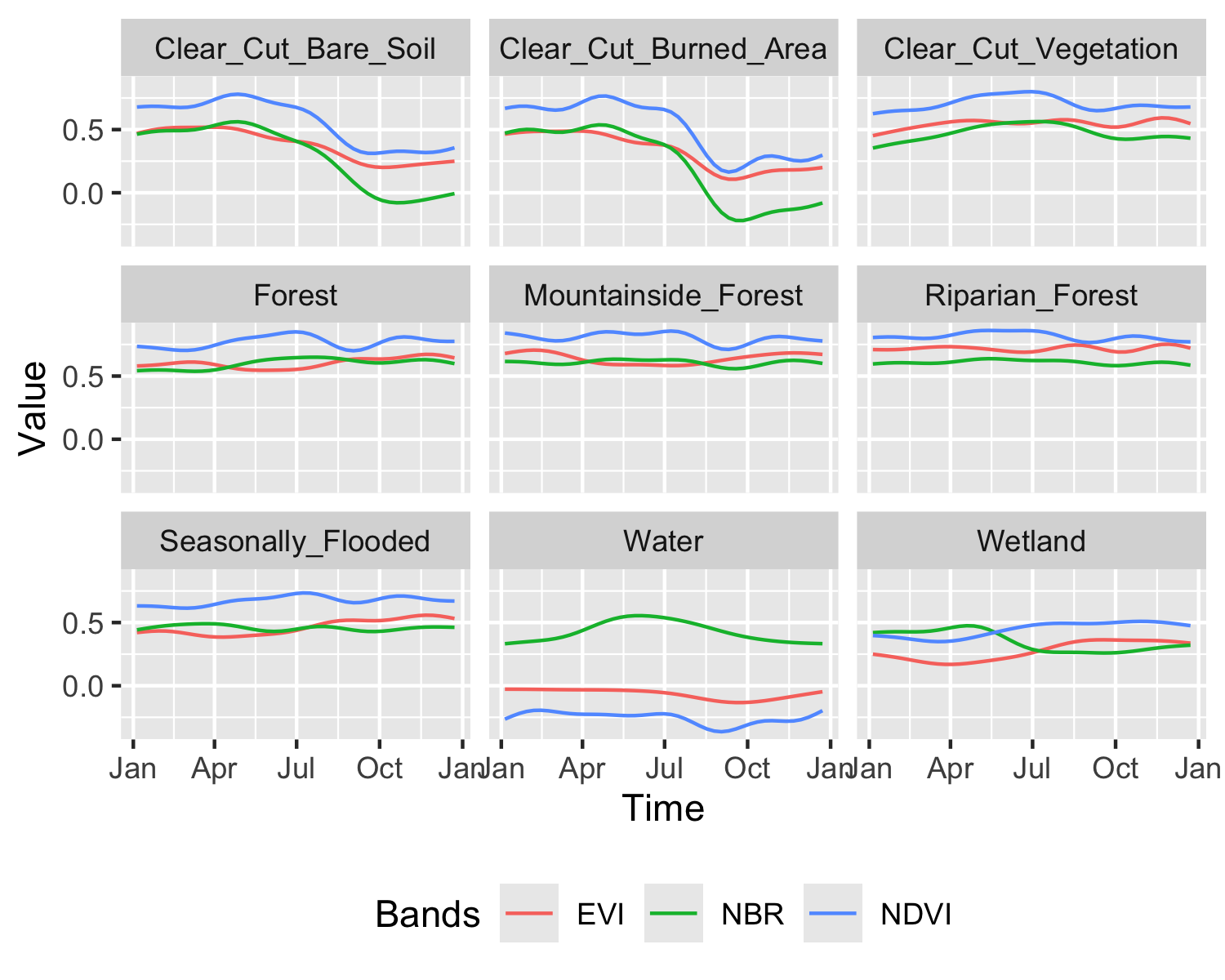
* Clear\_Cut\_Bare\_Soil: Post-clearing exposed soil
* Clear\_Cut\_Burned\_Area: Burned clearing residues
* Clear\_Cut\_Vegetation: Regrowth or residual vegetation post-harvest
* Forest: Intact forest including Mountainside\_Forest, Riparian\_Forest
* Water, Wetland, Seasonally\_Flooded: Hydrological features

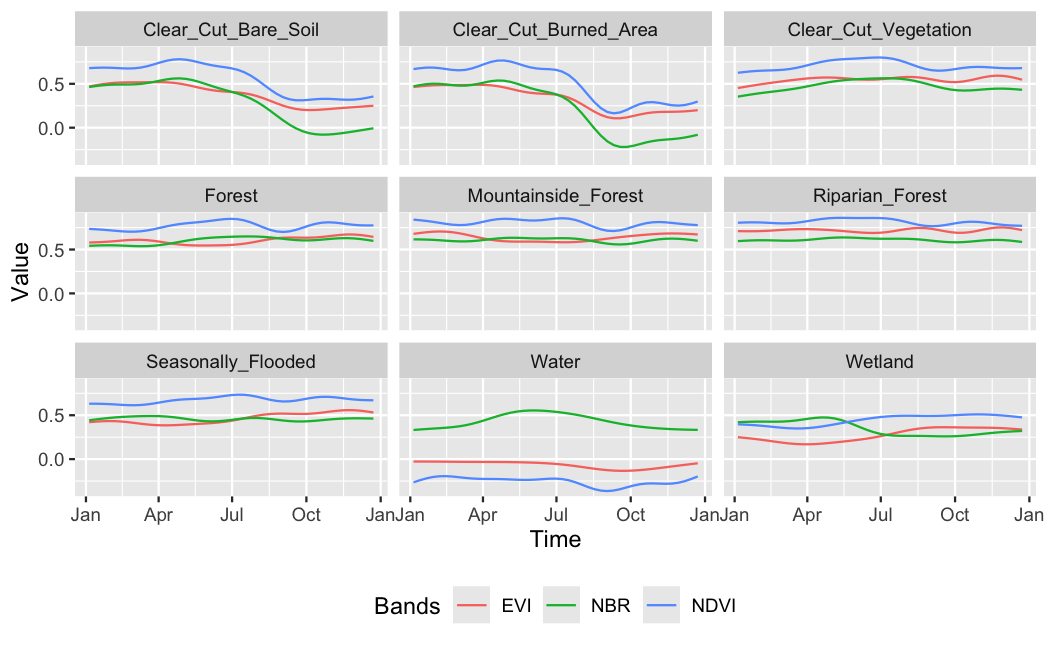
The time series spans 2022-01-05 to 2022-12-23 at 16-day intervals, containing 10 Sentinel-2 bands (B02, B03, B04, B05, B06, B07, B8A, B08, B11 and B12). Temporal validity aligns with the agricultural calendar to capture seasonal dynamics of land conversion.

# Import Demo Training Samples  
data("samples\_deforestation\_rondonia")  
base::summary(samples\_deforestation\_rondonia)  
NA # A tibble: 9 × 3  
NA label count prop  
NA <chr> <int> <dbl>  
NA 1 Clear\_Cut\_Bare\_Soil 944 0.157   
NA 2 Clear\_Cut\_Burned\_Area 983 0.164   
NA 3 Clear\_Cut\_Vegetation 603 0.100   
NA 4 Forest 964 0.160   
NA 5 Mountainside\_Forest 211 0.0351  
NA 6 Riparian\_Forest 1247 0.208   
NA 7 Seasonally\_Flooded 731 0.122   
NA 8 Water 109 0.0181  
NA 9 Wetland 215 0.0358  
utils::head(samples\_deforestation\_rondonia)  
NA # A tibble: 6 × 7  
NA longitude latitude start\_date end\_date label cube time\_series  
NA <dbl> <dbl> <date> <date> <chr> <chr> <list>   
NA 1 -66.5 -9.63 2022-01-05 2022-12-23 Clear\_Cut\_Bare\_Soil SENT… <tibble>   
NA 2 -66.4 -9.70 2022-01-05 2022-12-23 Clear\_Cut\_Bare\_Soil SENT… <tibble>   
NA 3 -66.4 -9.81 2022-01-05 2022-12-23 Forest SENT… <tibble>   
NA 4 -66.3 -9.64 2022-01-05 2022-12-23 Clear\_Cut\_Bare\_Soil SENT… <tibble>   
NA 5 -66.3 -9.73 2022-01-05 2022-12-23 Clear\_Cut\_Vegetati… SENT… <tibble>   
NA 6 -66.3 -9.73 2022-01-05 2022-12-23 Clear\_Cut\_Burned\_A… SENT… <tibble>

Visualizing class-specific temporal trajectories aids in assessing spectral separability; a key determinant of classification uncertainty. The sits\_patterns() function fits generalized additive models (GAM) to training data, producing smoothed temporal signatures that represent idealized class behavior. We derive vegetation indices (NDVI, EVI, NBR) to enhance interpretability.

# Compute spectral indexes for training samples  
training\_samples\_with\_indices <- samples\_deforestation\_rondonia |>   
 sits::sits\_apply(NDVI = (B08 - B04)/(B08 + B04)) |>   
 sits::sits\_apply(NBR = (B08 - B12) / (B08 + B12)) |>   
 sits::sits\_apply(EVI = 2.5 \* (B08 - B04) / ((B08 + 6.0 \* B04 - 7.5 \* B02) + 1.0))   
  
# Generate and plot patterns  
training\_samples\_indices <- training\_samples\_with\_indices |>   
 sits::sits\_select(bands = c("NDVI", "EVI", "NBR")) |>   
 sits::sits\_patterns() |> plot()





Time series of deforestation patterns from training samples of study in Rondonia (Simoes et al, 2021).

Temporal signatures conform to documented Amazonian deforestation phenology: forest clearing initiates during dry season onset (May-June), followed by burning in late dry season (August-September), evidenced by abrupt SWIR (B11) decline in Clear\_Cut\_Burned\_Area signatures. Regrowth trajectories (Clear\_Cut\_Vegetation) exhibit NIR (B8A) recovery through the monitoring period. These temporal contrasts provide the spectral-temporal feature space for subsequent classification.

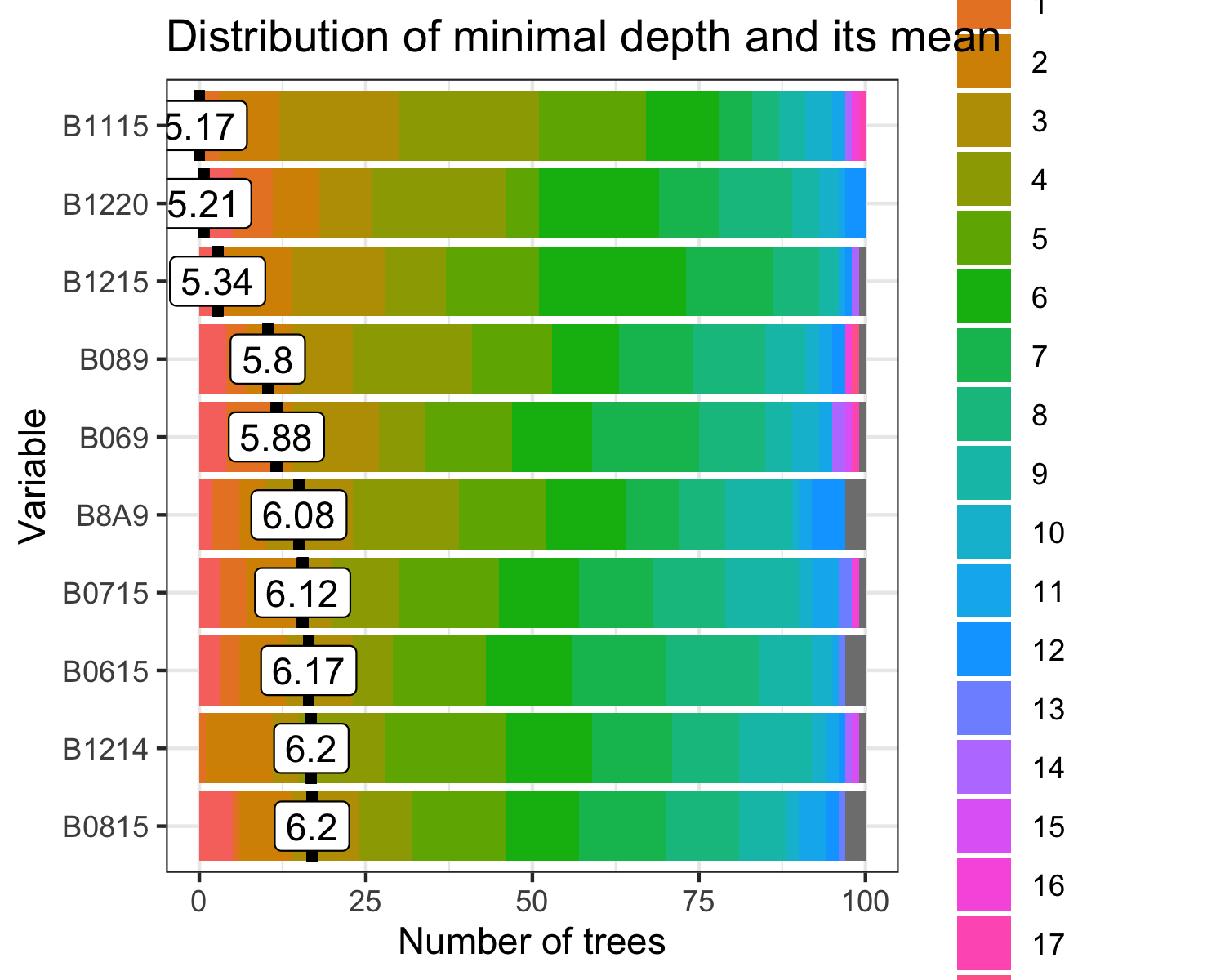
For extracting spectral signatures at known locations of ground truth points, the sits\_get\_data() function extracts band values by capturing signatures from the regularized data cube, creating a dataset suitable for training a classification model.

training\_samples\_derived <- sits::sits\_get\_data(  
 cube = cube\_s2\_reg,   
 samples = training\_samples\_with\_indices)  
  
# Show the tibble with the first three points  
training\_samples\_derived[1:3,]  
NA # A tibble: 3 × 7  
NA longitude latitude start\_date end\_date label cube time\_series  
NA <dbl> <dbl> <date> <date> <chr> <chr> <list>   
NA 1 -63.9 -8.57 2022-01-05 2022-12-23 Clear\_Cut\_Burned\_A… SENT… <tibble>   
NA 2 -63.9 -8.37 2022-01-05 2022-12-23 Clear\_Cut\_Bare\_Soil SENT… <tibble>   
NA 3 -63.8 -8.62 2022-01-05 2022-12-23 Water SENT… <tibble>

### Train Classifier

Random Forest models provide robust baseline classifiers for activity data, balancing computational efficiency with adequate performance for Tier 2-3 REDD+ requirements. The sits\_train() function implements parallelized Random Forest training with default hyperparameters optimized for satellite time series.

# Train model using Random Forest algorithm  
model\_randomForest <- sits::sits\_train(  
 samples = samples\_deforestation\_rondonia,  
 ml\_method = sits::sits\_rfor()  
 )  
  
# plot the model results  
plot(model\_randomForest)



Variable importance metrics identify which spectral-temporal features contribute most to class discrimination. For REDD+ operatoins, understanding feature importance supports:

1. Uncertainty attribution: High-importance features with measurement error propagate more uncertainty
2. Sensor planning: Identifies critical spectral bands for maintaining classification accuracy
3. Transfer learning: Guides feature selection when adapting models to new regions
4. Validation design: Informs stratification schemes for accuracy assessment

### Training Regime

### Probability Data Cube

Classification outputs are generated as probability cubes rather than discrete label maps to preserve model uncertainty information. The sits\_classify() function applies trained models to regularized data cubes using fault-tolerant parallel processing, producing per-class probability layers that quantify model confidence at the pixel level.

The sits package implements parallel computing to minimize communication overhead, which enables multi-core systems without configuration.



Each pixel in the probability cube contains n probability values (where n = number of classes), summing to 1.0. Low maximum probability values indicate pixels with high classification uncertainty, typically occurring at class boundaries, in mixed pixels, or where spectral-temporal signatures deviate from training data. These pixel-level uncertainty metrics support subsequent:

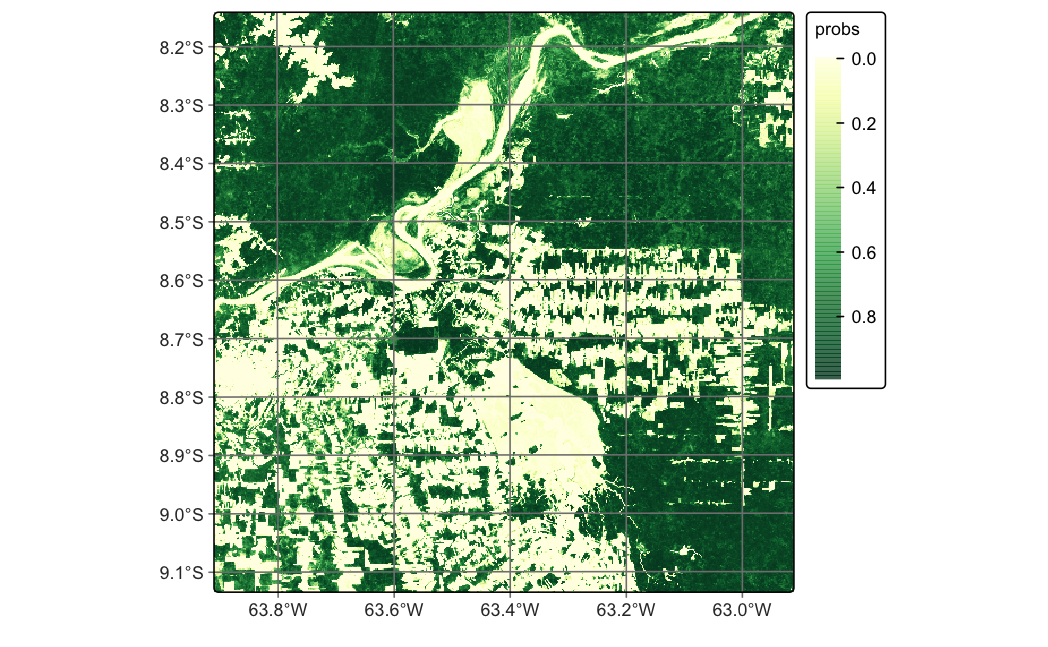
* Adaptive stratification: High-uncertainty pixels receive higher validation sampling density
* Uncertainty propagation: Probability distributions feed Monte Carlo emission simulations
* Quality control: Spatial patterns of low confidence identify systematic classification errors
* Temporal tracking: Probability time series detect gradual transitions missed by discrete labeling

Probability cubes thus serve as the foundation for rigorous uncertainty quantification in subsequent workflow stages.

### Smooth Data Cube

Pixel-based classification inherently produces salt-and-pepper noise due to within-class spectral variability and mixed pixels. Isolated misclassified pixels (outliers surrounded by different classes) represent systematic over-fitting to training data rather than true landscape heterogeneity. For REDD+ applications, these artifacts inflate area uncertainty and complicate change detection. To address this, the sits\_smooth() function implements Bayesian post-processing that incorporates spatial context from neighboring pixels to refine probability estimates. The smoothing process:

1. Computes spatially-weighted probability adjustments based on neighborhood composition
2. Updates pixel probabilities using Bayesian inference combining original estimates with spatial priors
3. Produces refined probability cube with reduced outlier occurrence while preserving legitimate small patches

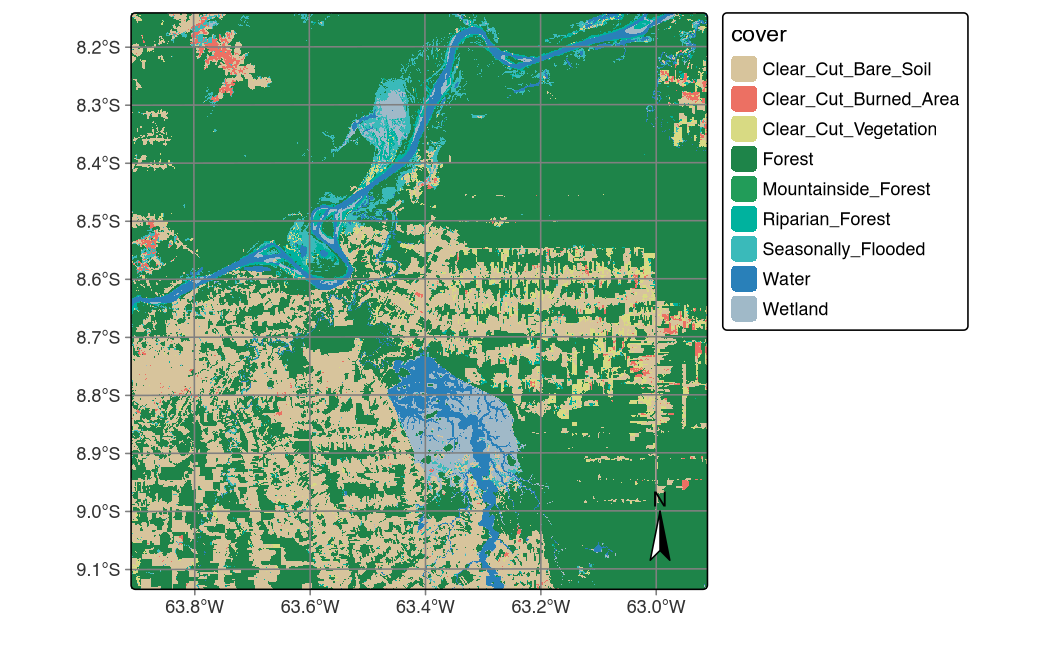


Smoothed probability map for class Forest.

|  |
| --- |
| Smoothening Risks |
| Spatial smoothing affects activity data uncertainty in competing directions:   * Reduces random error: Eliminates isolated misclassifications that inflate variance * Introduces spatial bias: May smooth over legitimate small-scale heterogeneity * Improves validation agreement: Better alignment with typical human interpretation scales * Modifies probability distributions: Changes the shape of pixel-level uncertainty estimates   For REDD+ reporting, smoothed probability cubes typically yield lower reported uncertainties due to reduced classification noise, but verification bodies may require documentation of smoothing parameters demonstrating methodological choices. |

### Classify Data Cube

Final classified maps are derived from smoothed probability cubes through maximum likelihood labeling: each pixel is assigned the class with highest posterior probability. The sits\_label\_classification() function performs this deterministic assignment while maintaining provenance links to underlying probability distributions.



Classified cube obtained by Random Forest model.

The classified cube produced by sits workflows provides the activity data foundation for REDD+ accounting, but uncertainty quantification requires additional analysis beyond the discrete classification procedures. Sections 3.2-3.5 demonstrate more more advanced operations using caret (Kuhn, 2011), ForestToolbox (Tarazona Coronel et al., 2021), and terra packages (Hijmans, 2025) required to:

1. Extract pixel-level probability distributions for Monte Carlo sampling (Section 3.3)
2. Implement stratified validation with area-adjusted accuracy assessment (Section 3.4)
3. Calibrate Random Forest hyperparameters through iterative uncertainty minimization (Section 3.5)

### Cross-Validatiion

### Accuracy Assessment

## 3.2 IPCC Land Monitoring Approaches

The architecture of land-use transition monitoring systems determines the complexity and uncertainty of activity data. IPCC guidelines define three hierarchical approaches that differ in spatial explicitness and data requirements.

### Approach Classification

| Approach | Data Requirement | Spatial Explicitness | Uncertainty |
| --- | --- | --- | --- |
| Approach 1 | Total areas by category | None (aggregated statistics) | Highest |
| Approach 2 | Transition matrices | Sampling-based | Moderate |
| Approach 3 | Wall-to-wall maps | Spatially explicit | Lowest |

REDD+ Standard: Most carbon projects use Approach 3 with remote sensing classification, combining high spatial resolution with complete coverage. This approach enables:

* Pixel-level change detection
* Spatial stratification for uncertainty analysis
* Integration with degradation monitoring
* Transparent, verifiable methodologies

## 3.7 Chapter Summary

This chapter established a comprehensive framework for activity data uncertainty quantification in REDD+ monitoring systems. Key contributions include:

1. Data cube architecture: Structured approach to spatiotemporal land cover analysis
2. IPCC methodology alignment: Clear guidance on Approach 3 requirements
3. Pixel-level uncertainty tracking: Model residuals for adaptive sampling
4. Random Forest optimization: Monte Carlo calibration workflows
5. Rigorous accuracy assessment: Confusion matrices with area adjustment

Activity data provides spatial extent; allometry converts extent to biomass. Combined uncertainty requires:

Where covariance term captures spatial correlation between classification errors and biomass estimation errors.

# 4. Monte Carlo

## Overview

* Comment/Suggestion (JN, 12/04): Add section on probability density funciton

### Environment Setup (R)

easypackages::packages(  
 "ropensci/allodb", "animation", "BIOMASS", "cols4all", "covr", "cowplot", "caret",  
 "DescTools", "dataMaid", "dplyr", "FawR", "ForestToolsRS", "forestdata",   
 "flextable", "ggplot2", "giscoR", "ggfortify", "htmltools", "janitor", "jsonlite",   
 "lattice", "leaflet.providers", "leaflet", "lmtest", "lwgeom",   
 "kableExtra", "kernlab", "knitr", "mapedit", "mapview", "maptiles", "Mlmetrics",   
 "ModelMetrics", "moments", "olsrr", "openxlsx", "plotly", "psych", "randomForest",   
 "raster","RColorBrewer", "rmarkdown", "renv", "reticulate", "s2", "sf", "scales",   
 "sits","spdep", "stars", "stringr", "terra", "tmap", "tmaptools", "tidymodels",   
 "tidyverse", "tidyr", "tune", "useful",  
 prompt = F  
 )

# References

Appel, M., & Pebesma, E. (2019a). *On-demand processing of data cubes from satellite image collections with the gdalcubes library*. *4*. <https://doi.org/10.3390/data4030092>

Appel, M., & Pebesma, E. (2019b). On-demand processing of data cubes from satellite image collections with the gdalcubes library. *Data*, *4*(3), 92.

Archibald, S., & Bond, W. J. (2003). Growing tall vs. Growing wide: Tree architecture and allometry of acacia karroo in forest, savanna, and arid environments. *Oikos (Copenhagen, Denmark)*, *102*(1), 3–14. <https://doi.org/10.1034/j.1600-0706.2003.12181.x>

ART. (2021). *The REDD+ environmental excellence standard* (2.0 ed.). <https://www.artredd.org/wp-content/uploads/2021/12/TREES-2.0-August-2021-Clean.pdf>

Baillie, R., & Wagstaff, S. S. (1980). Lucas pseudoprimes. *Mathematics of Computation*, *35*(152), 1391–1417.

Ballantyne, A., Andres, R., Houghton, R., Stocker, B., Wanninkhof, R., Anderegg, W., Cooper, L., DeGrandpre, M., Tans, P., Miller, J., & al., et. (2015). Audit of the global carbon budget: Estimate errors and their impact on uptake uncertainty. *Biogeosciences (Online)*, *12*(8), 2565–2584.

Bohlman, S., & O’Brien, S. (2006). Allometry, adult stature and regeneration requirement of 65 tree species on barro colorado island, panama. *Journal of Tropical Ecology*, *22*(2), 123–136.

Brown, S. (1997). *Estimating biomass and biomass change of tropical forests: A primer. Food and agriculture organization of the united nations*.

Butler, B. J., Sass, E. M., Gamarra, J. G., Campbell, J. L., Wayson, C., Olguıń, M., Carrillo, O., & Yanai, R. D. (2024). Uncertainty in REDD+ carbon accounting: A survey of experts involved in REDD+ reporting. *Carbon Balance and Management*, *19*(1), 22.

Camara, G., Simoes, R., Souza, F., Menino, F., Pelletier, C., Andrade, P. R., Ferreira, K., & Queiroz, G. (2024). Uncertainty and active learning analysis. In *Satellite time series on earth observation data cubes*. <https://e-sensing.github.io/sitsbook/uncertainty-and-active-learning.html#uncertainty-and-active-learning>

Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., & Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, *12*(4), 351–366.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., & al., et. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190.

Dietze, M. C., Wolosin, M. S., & Clark, J. S. (2008). Capturing diversity and interspecific variability in allometries: A hierarchical approach. *Forest Ecology and Management*, *256*(11), 1939–1948.

Duncanson, L., Disney, M., Armston, J., Nickeson, J., Minor, D., & Camacho, F. (2021). *Aboveground woody biomass product validation good practices protocol*. <https://doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001>

Enquist, B. J., West, G. B., Charnov, E. L., & Brown, J. H. (1999). Allometric scaling of production and life-history variation in vascular plants. *Nature*, *401*(6756), 907–911.

Feldpausch, T. R., Banin, L., Phillips, O. L., Baker, T. R., Lewis, S. L., Quesada, C. A., Affum-Baffoe, K., Arets, E. J., Berry, N. J., Bird, M., & al., et. (2011). Height-diameter allometry of tropical forest trees. *Biogeosciences (Online)*, *8*(5), 1081–1106.

Gonzalez-Akre, E., Piponiot, C., Lepore, M., & Anderson-Teixeira, K. (2025). *Allodb: Tree biomass estimation at extra-tropical forest plots*.

Gonzalez-Akre, E., Piponiot, C., Lepore, M., Herrmann, V., Lutz, J. A., Baltzer, J. L., Dick, C. W., Gilbert, G. S., He, F., Heym, M., & al., et. (2022). Allodb: An r package for biomass estimation at globally distributed extratropical forest plots. *Methods in Ecology and Evolution*, *13*(2), 330–338.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews*, *41*(4), 587–638.

Gould, S. J. (1971). Geometric similarity in allometric growth: A contribution to the problem of scaling in the evolution of size. *The American Naturalist*, *105*(942), 113–136.

Hijmans, R. J. (2025). *Terra: Spatial data analysis*. <https://CRAN.R-project.org/package=terra>

Huxley, J. S. (1924). Constant differential growth-ratios and their significance. *Nature*, *114*(2877), 895–896.

IPCC. (n.d.). *2013 revised supplementary methods and good practice guidance arising from the kyoto protocol*.

IPCC. (2012). *Report of the conference of the parties serving as the meeting of the parties to the kyoto protocol on its seventh session,*. *FCCC/KP/CMP/2011/10*(Durban Meeting November 28th to December 11th, 2011), 824.

IPCC. (2019a). *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories* (Agriculture, Forestry and Other Land Use, Vol. 4). Intergovernmental Panel on Climate Change. <https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html>

IPCC. (2019b). Chapter 2: Generic methodologies applicable to multiple land-use categories. In *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories* (Agriculture, Forestry and Other Land Use, Vol. 4). Intergovernmental Panel on Climate Change. <https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch02_Generic%20Methods.pdf>

Jansen, J. J., Sevenster, J., & Faber, J. and. (1996). *Opbrengsttabellen voor belangrijke boomsoorten in nederland*. IBN-DLO.

Jenkins, J. C. (2004). *Comprehensive database of diameter-based biomass regressions for north american tree species*. United States Department of Agriculture, Forest Service, Northeastern ….

King, D. A. (1996). Allometry and life history of tropical trees. *Journal of Tropical Ecology*, *12*(1), 25–44.

King, D. A., & Clark, D. A. (2011). Allometry of emergent tree species from saplings to above-canopy adults in a costa rican rain forest. *Journal of Tropical Ecology*, *27*(6), 573–579.

Köhler, P., & Huth, A. (2010). Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. *Biogeosciences (Online)*, *7*(8), 2531–2543.

Komiyama, A., Ong, J. E., & Poungparn, S. (2008). Allometry, biomass, and productivity of mangrove forests: A review. *Aquatic Botany*, *89*(2), 128–137. <https://doi.org/10.1016/j.aquabot.2007.12.006>

Komiyama, A., Poungparn, S., & Kato, S. (2005). Common allometric equations for estimating the tree weight of mangroves. *Journal of Tropical Ecology*, *21*(4), 471–477.

Kuhn, M. (2011). *The caret package*.

Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, *44*(247), 335–341.

Muller-Landau, H. C., Condit, R. S., Chave, J., Thomas, S. C., Bohlman, S. A., Bunyavejchewin, S., Davies, S., Foster, R., Gunatilleke, S., Gunatilleke, N., & al., et. (2006). Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecology Letters*, *9*(5), 575–588.

Murray, S. H. (2003). The schreier-sims algorithm. *Essay Submitted to the Department of Mathematics of the Australian National University*.

Paul, K. I., Roxburgh, S. H., & Larmour, J. S. (2017). Moisture content correction: Implications of measurement errors on tree-and site-based estimates of biomass. *Forest Ecology and Management*, *392*, 164–175.

Pelletier, J., Martin, D., & Potvin, C. (2013). REDD+ emissions estimation and reporting: Dealing with uncertainty. *Environmental Research Letters*, *8*(3), 034009.

Picard, N., Saint-Andre, L., & Henry, M. (2012). *Manual for building tree volume and biomass allometric equations: From field measurement to prediction. Food and agriculture organization of the united nations, centre de cooperation internationale en recherche agronomique pour le developpement*.

Rahman, M. S., Donoghue, D. N., Bracken, L. J., & Mahmood, H. (2021). Biomass estimation in mangrove forests: A comparison of allometric models incorporating species and structural information. *Environmental Research Letters*, *16*(12), 124002.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., & Herault, B. (2017b). *BIOMASS : An r package for estimating above-ground biomass and its uncertainty in tropical forests*. *8*. <https://doi.org/10.1111/2041-210X.12753>

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., & Herault, B. (2017a). *BIOMASS : An r package for estimating above-ground biomass and its uncertainty in tropical forests*. *8*. <https://doi.org/10.1111/2041-210X.12753>

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., & Hérault, B. (2017). Biomass: An r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods in Ecology and Evolution*, *8*(9), 1163–1167.

Rocha de Souza Pereira, F., Kampel, M., Gomes Soares, M. L., Estrada, G. C. D., Bentz, C., & Vincent, G. (2018). Reducing uncertainty in mapping of mangrove aboveground biomass using airborne discrete return lidar data. *Remote Sensing*, *10*(4), 637.

Roxburgh, S., Paul, K., Clifford, D., England, J., & Raison, R. (2015). Guidelines for constructing allometric models for the prediction of woody biomass: How many individuals to harvest? *Ecosphere (Washington, D.C)*, *6*(3), 1–27.

Rutishauser, E., Noor’an, F., Laumonier, Y., Halperin, J., Hergoualc’h, K., Verchot, L., et al. (2013). Generic allometric models including height best estimate forest biomass and carbon stocks in indonesia. *Forest Ecology and Management*, *307*, 219–225.

Schrodt, F., Beck, M., Estopinan, J., Bowler, D. E., Fontaine, C., Gaüzère, P., Goury, R., Grenié, M., Martins, I. S., Morueta-Holme, N., & al., et. (2025). Advancing causal inference in ecology: Pathways for biodiversity change detection and attribution. *Methods in Ecology and Evolution*, *16*(10), 2276–2304.

Sheynin, O. B. (1989). AA markov’s work on probability. *Archive for History of Exact Sciences*, *39*(4), 337–377.

Shinozaki, K., Yoda, K., Hozumi, K., & Kira, T. (1964). A quantitative analysis of plant form-the pipe model theory: I. Basic analyses. *Japanese Journal of Ecology*, *14*(3), 97–105.

Simoes, R., Camara, G., Queiroz, G., Souza, F., Andrade, P. R., Santos, L., Carvalho, A., & Ferreira, K. (2021a). Satellite image time series analysis for big earth observation data. *Remote Sensing*, *13*(13), 2428.

Simoes, R., Camara, G., Queiroz, G., Souza, F., Andrade, P., Santos, L., Carvalho, A., & Ferreira, K. (2021b). *Satellite image time series analysis for big earth observation data*. *13*, 2428. <https://doi.org/10.3390/rs13132428>

Solovay, R., & Strassen, V. (1977). A fast monte-carlo test for primality. *SIAM Journal on Computing*, *6*(1), 84–85.

Sornette, D. (2006). *Critical phenomena in natural sciences: Chaos, fractals, selforganization and disorder: Concepts and tools*. Springer.

Tarazona Coronel, Y., Zabala, A., Pons, X., Broquetas, A., Nowosad, J., & A. Zurqani, H. (2021). *Fusing landsat and SAR data for mapping tropical deforestation through machine learning classification and the PVts-β non-seasonal detection approach*. <https://doi.org/10.1080/07038992.2021.1941823>

UNFCCC, C. (2010). Report of the conference of the parties on its fifteenth session, held in copenhagen from 7 to 19 december 2009. *FCCC/CP/2009/11/Add. 1*.

Valbuena, R., Heiskanen, J., Aynekulu, E., Pitkänen, S., & Packalen, P. (2016). Sensitivity of above-ground biomass estimates to height-diameter modelling in mixed-species west african woodlands. *PloS One*, *11*(7), e0158198.

Vieilledent, G., Fischer, F. J., Chave, J., Guibal, D., Langbour, P., & Gérard, J. (2018). New formula and conversion factor to compute basic wood density of tree species using a global wood technology database. *American Journal of Botany*, *105*(10), 1653–1661.

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, *276*(5309), 122–126.

Zianis, D., & Mencuccini, M. (2004). On simplifying allometric analyses of forest biomass. *Forest Ecology and Management*, *187*(2-3), 311–332.

# Glossary of Terms

* **Activity Data:** Data on the magnitude of a human activity resulting in emissions or removals taking place during a given period of time.
* **Anaerobic:** Conditions in which oxygen is not readily available. These conditions are important for the production of methane emissions. Whenever organic material decomposes in anaerobic conditions, such as in landfills, flooded rice fields, etc., methane is likely to be formed.
* **Andosol:** A soil developed in volcanic ash. Generally andosols have good drainage and are prone to fertility problems.
* **Bootstrap Technique**: A type of computationally intensive statistical method which typically uses repeated resampling from a set of data to assess variability of parameter estimates.
* **Back-casting**: The opposite of forecasting. Predicting conditions in the past from current conditions.
* **Bias:** A systematic error of the observation method, whose magnitude in most cases is unknown. It can be introduced by using measuring equipment that is improperly calibrated, by selecting items from a wrong population or by favoring certain elements of a population.
* **Carbon Dioxide Equivalent:** (CO2e) A measure used to compare different greenhouse gases based on their contribution to radiative forcing. The UNFCCC currently uses global warming potentials (GWPs) as factors to calculate carbon dioxide equivalent.
* **Coefficient of Variation** The ratio of the population standard deviation, , and mean, , where . It is often used to express the standard deviation as a percentage of the mean.
* **Confidence Interval:** The range that encloses the true value of an unknown fixed quantity with a specified confidence. Typically, a 95 percent confidence interval is assumed (90% requied in ART-TREES Standard)
* **Decision Tree**: A flow chart describing the specific ordered steps which need to be followed to develop an inventory or an inventory component in accordance with the principles of good practice.
* **Emission Factor:** A coefficient that quantifies the emissions or removals of a gas per unit activity. Emission factors are often based on a sample of measurement data, averaged to develop a representative rate of emission for a given activity level under a given set of operating conditions.
* **Global Warming Potentials:** (GWP) are calculated as the ratio of the radiative forcing of one kilogram greenhouse gas emitted to the atmosphere to that from one kilogramme CO2 over a period of time (e.g., 100 years).
* **Ground Truth:** A term used for data obtained by measurements on the ground, usually as validation for remote sensing, e.g., satellite data.
* **Monte Carlo Method**:[[8]](#footnote-299) The statistical method recommended to analyse the uncertainty of the inventory. It performs the inventory calculation many times by computer, each time with the uncertain emission factors or model parameters and activity data chosen randomly within the distribution of uncertainties specified by the user.
* **Normal Distribution**: A probability distribution (also known as Gaussian) defined by two parameters: the mean () and the standard deviation (). It is often assumed for symmetric uncertainties.
* **Oxidation:** Chemically transform of a substance by combining it with oxygen.
* **Population:** The population is the totality of items under consideration. In the case of a random variable, the probability distribution is considered to define the population of that variable.
* **Probability Density Function:** (PDF) A function that describes the range and relative likelihood of possible values. The PDF is used to describe uncertainty in the estimate of a quantity that is a fixed constant whose value is not exactly known.
* **Removals:** Removal of greenhouse gases and/or their precursors from the atmosphere by a sink.
* **Standard Deviation**: The positive square root of the variance. It is a measure of the dispersion or spread of the data around the mean.
* **Uncertainty:** A state of incomplete knowledge that can result from a lack of information or from disagreement about what is known or even knowable. It may have many types of sources, from imprecision in the data to ambiguously defined concepts or terminology, incomplete understanding of critical processes, or uncertain projections of human behavior.
* **Sink:** Any process, activity or mechanism which removes a greenhouse gas, an aerosol, or a precursor of a greenhouse gas from the atmosphere. (UNFCCC Article 1.8) Notation in the final stages of reporting is the negative (-) sign.
* **Validation** The establishment of sound approach and foundation. In the context of emission inventories, validation involves checking to ensure that the inventory has been compiled correctly in line with reporting instructions and guidelines.
* **Verification** The collection of activities and procedures that can be followed during the planning and development, or after completion of an inventory that can help to establish its reliability for the intended applications of that inventory.
* Several definitions above were copied from the [IPCC’s 2006 Glossary of Terms](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/0_Overview/V0_2_Glossary.pdf)

# Additional Reading

#### *Allometric Uncertainty:*

* Aholoukpè, H. N. S., Dubos, B., Deleporte, P., Flori, A., Amadji, L. G., Chotte, J.-L., & Blavet, D. (2018). Allometric equations for estimating oil palm stem biomass in the ecological context of benin, west africa. *Trees*, *32*(6), 1669–1680.
* Andersen, H.-E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, *32*(5), 355–366. <https://doi.org/10.5589/m06-030>
* Baskerville, G. (1972). Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research*, *2*(1), 49–53.
* Duncanson, L., Disney, M., Armston, J., Nickeson, J., Minor, D., & Camacho, F. (2021). Aboveground woody biomass product validation good practices protocol. <https://doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001>
* Dutcă, I., Stăncioiu, P. T., Abrudan, I. V., & Ioraș, F. (2018). Using clustered data to develop biomass allometric models: The consequences of ignoring the clustered data structure. *PloS One*, *13*(8), e0200123.
* Martin, A. (2022). Accuracy and precision in urban forestry tools for estimating total tree height. *Arboric. Urban For*, *48*(6), 319–332.
* Martı́nez-Sánchez, J. L., Martı́nez-Garza, C., Cámara, L., & Castillo, O. (2020). Species-specific or generic allometric equations: Which option is better when estimating the biomass of mexican tropical humid forests? *Carbon Management*, *11*(3), 241–249.
* McRoberts, R. E., & Westfall, J. A. (2016). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, *73*(ue 3), 625–633. <https://doi.org/10.1007/s13595-015-0473-x>
* Nickless, A., Scholes, R. J., & Archibald, S. (2011). A method for calculating the variance and confidence intervals for tree biomass estimates obtained from allometric equations. *South African Journal of Science*, *107*(5), 1–10.
* Ojoatre, S., Zhang, C., Hussin, Y. A., Kloosterman, H. E., & Ismail, M. H. (2019). Assessing the uncertainty of tree height and aboveground biomass from terrestrial laser scanner and hypsometer using airborne LiDAR data in tropical rainforests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(10), 4149–4159.
* Parresol, B. R. (1993). Modeling multiplicative error variance: An example predicting tree diameter from stump dimensions in baldcypress. *Forest Science*, *39*(4), 670–679.
* Picard, N., Bosela, F. B., & Rossi, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. Annals of Forest Science, 72(6), 811–823. <https://doi.org/10.1007/s13595-014-0434-9>
* Picard N., Saint-André L., Henry M. 2012. Manual for building tree volume and biomass allometric equations: from field measurement to prediction. Food and Agricultural Organization of the United Nations, Rome, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, 215 pp.
* Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., & al., et. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), 4540.
* Roxburgh, S., Paul, K., Clifford, D., England, J., & Raison, R. (2015). Guidelines for constructing allometric models for the prediction of woody biomass: How many individuals to harvest? *Ecosphere (Washington, D.C)*, *6*(3), 1–27.
* Shang, Y., Xia, Y., Ran, X., Zheng, X., Ding, H., & Fang, Y. (2025). Allometric equations for aboveground biomass estimation in natural forest trees: Generalized or species-specific? *Diversity*, *17*(7), 493.
* Vorster, A. G., Evangelista, P. H., Stovall, A. E., & Ex, S. (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: The role of allometric equations. Carbon Balance and Management, 15(1), 8.
* Wayson, C. A., Johnson, K. D., Cole, J. A., Olguín, M. I., Carrillo, O. I., & Birdsey, R. A. (2015). Estimating uncertainty of allometric biomass equations with incomplete fit error information using a pseudo-data approach: methods. Annals of Forest Science, 72(6), 825–834.
* White, G. W., Yamamoto, J. K., Elsyad, D. H., Schmitt, J. F., Korsgaard, N. H., Hu, J. K., Gaines III, G. C., Frescino, T. S., & McConville, K. S. (2025). Small area estimation of forest biomass via a two-stage model for continuous zero-inflated data. *Canadian Journal of Forest Research*, *55*, 1–19.
* Yanai, R. D., Battles, J. J., Richardson, A. D., Blodgett, C. A., Wood, D. M., & Rastetter, E. B. (2010). Estimating uncertainty in ecosystem budget calculations. *Ecosystems (New York, N.Y.)*, *13*(ue 2), 239–248. <https://doi.org/10.1007/s10021-010-9315-8>
* Yokelson, R.J., et al. (2013). Coupling field and laboratory measurements to estimate the emission factors of identified and unidentified trace gases for prescribed fires. Atmospheric Chemistry and Physics, 13, 89-116.
* Zapata-Cuartas, M., Sierra, C. A., & Alleman, L. (2012). Probability distribution of allometric coefficients and bayesian estimation of aboveground tree biomass. *Forest Ecology and Management*, *277*, 173–179.

#### *Emission Factor Uncertainty:*

* Andreae, M.O. (2019). Emission of trace gases and aerosols from biomass burning – an updated assessment. Atmospheric Chemistry and Physics, 19, 8523-8546. doi:10.5194/acp-19-8523-2019
* Brown, J.K. (1974). Handbook for inventorying downed woody material. USDA Forest Service General Technical Report INT-16.
* IPCC. (2019). 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (Agriculture, Forestry and Other Land Use, Vol. 4). Intergovernmental Panel on Climate Change. https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html
* IPCC. (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use. Intergovernmental Panel on Climate Change.
* Köhler, P., & Huth, A. (2010). Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. *Biogeosciences (Online)*, *7*(8), 2531–2543.
* Pelletier, J., Martin, D., & Potvin, C. (2013). REDD+ emissions estimation and reporting: Dealing with uncertainty. Environmental Research Letters, 8(3), 034009.
* Pelletier, J., Busch, J., & Potvin, C. (2015). Addressing uncertainty upstream or downstream of accounting for emissions reductions from deforestation and forest degradation. *Climatic Change*, *130*(4), 635-648
* Pelletier, N., Thiagarajan, A., Durnin-Vermette, F., Liang, B. C., Choo, D., Cerkowniak, D., … & VandenBygaart, A. J. (2025). Approximate Bayesian inference for calibrating the IPCC tier-2 steady-state soil organic carbon model for Canadian croplands using long-term experimental data. *Environmental Modelling & Software*, *190*, 106481
* Seiler, W., & Crutzen, P.J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. Climatic Change, 2(3), 207-247.
* van Leeuwen, T.T., & van der Werf, G.R. (2011). Spatial and temporal variability in the ratio of trace gases emitted from biomass burning. Atmospheric Chemistry and Physics, 11, 3611-3629.

#### *Activity Data Uncertainty:*

* Butler, B. J., Sass, E. M., Gamarra, J. G., Campbell, J. L., Wayson, C., Olguín, M., Carrillo, O., & Yanai, R. D. (2024). Uncertainty in REDD+ carbon accounting: A survey of experts involved in REDD+ reporting. Carbon Balance and Management, 19(1), 22.
* Chen, Q., Laurin, G. V., & Valentini, R. (2015). Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. Remote Sensing of Environment, 160, 134–143. <https://doi.org/10.1016/j.rse.2015.01.009>
* GOFC-GOLD (2016). Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative. Edition 2.0. Rome: Food and Agriculture Organization.
* GOFC-GOLD (2016). A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. GOFC-GOLD Report version COP22-1. Alberta, Canada: GOFC-GOLD Land Cover Project Office.
* Köhler, P., & Huth, A. (2010). Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests. Biogeosciences, 7(8), 2531–2543.
* Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42-57.
* Pontius Jr., R.G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32(15), 4407-4429.
* Sheng, J., Zhou, W., & De Sherbinin, A. (2018). Uncertainty in estimates, incentives, and emission reductions in REDD+ projects. *International Journal of Environmental Research and Public Health*, *15*(7), 1544.
* Stehman, S.V. (2014). Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. International Journal of Remote Sensing, 35(13), 4923-4939.

#### *Monte Carlo Methods:*

* Holdaway, R. J., McNeill, S. J., Mason, N. W. H., & Carswell, F. E. (2014). Propagating uncertainty in plot-based estimates of forest carbon stock and carbon stock change. *Ecosystems (New York, N.Y.)*, *17*(ue 4), 627–640. <https://doi.org/10.1007/s10021-014-9749-5>
* Keller, M., Palace, M., & Hurtt, G. (2001). Biomass estimation in the tapajos national forest, brazil. *Forest Ecology and Management*, *154*(ue 3), 371–382.
* Molto, Q., Rossi, V., & Blanc, L. (2013). Error propagation in biomass estimation in tropical forests. Methods in Ecology and Evolution, 4(2), 175–183. <https://doi.org/10.1111/j.2041-210x.2012.00266.x>
* Yanai, R. D., Battles, J. J., Richardson, A. D., Blodgett, C. A., Wood, D. M., & Rastetter, E. B. (2010). Estimating uncertainty in ecosystem budget calculations. Ecosystems, 13(2), 239–248. <https://doi.org/10.1007/s10021-010-9315-8>

#### *Biostatistical Theory:*

* Buchanan, M. (2000). Ubiquity: Why Catastrophes Happen. Three Rivers Press.
* Mandelbrot, B. B., & Hudson, R. L. (2004). The Misbehavior of Markets: A Fractal View of Financial Turbulence. Basic Books.
* Strogatz, S. H. (2003). Sync: How Order Emerges from Chaos in the Universe, Nature, and Daily Life. Hyperion.
* Taleb, N. N. (2007). The Black Swan: The Impact of the Highly Improbable. Random House.

#### IPCC Guidelines & Key Sections

| Resource | Description | Source |
| --- | --- | --- |
| IPCC 2006, Vol. 4 |  |  |
| Ch.2 Generic Methodologies | Eq.2.9 Calculation of biomass retention & growth post-conversion | [Link](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_02_Ch2_Generic.pdf#page=15) |
| Ch.3 Representation of Lands | S.3.2 Six land-use categories recommended for estimating GHG emissions from LULC | [Link](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_03_Ch3_Representation.pdf#page=6.9) |
| IPCC 2019, Vol. 4 |  |  |
| Ch.2 Generic Methodologies | Eq 2.25 Annual SOC stock change in mineral soils | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch02_Generic%20Methods.pdf#page=33) |
|  | Tbl.2.3 Default reference condition of SOC stocks to soil & climate | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch02_Generic%20Methods.pdf#page=35) |
| Ch.3 Representation of Lands | Tb.3.1 List of IPCC categories: land, climate, soil, mgt, activity | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=11) |
|  | Pg.3.1 Tier 1 sampling approaches decision tree | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=21) |
|  | Tb.3.6X Approach 1-3 to IPCC land-use classification & sampling | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=20) |
|  | Tb.3.4 Approach 2 land change matrix to avoid double-counting | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=17) |
|  | Pg.3A5 Climate zone delineation & updated datasets | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=47) |
|  | Tb.3A.1 Global land-cover datasets listed by IPCC in 2017 | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch03_Land%20Representation.pdf#page=35) |
| Ch.4 Forest Land | Tb.4.4 R:S below to above-ground biomass ratio by climate & region | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch04_Forest%20Land.pdf#page=18) |
| Ch.5 Cropland | Tb.5.5 Relative stock change factors for mgt. activity in croplands | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=27) |
|  | Tb.5.8 Default AGB carbon stocks retained on cropland in year 1 | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=41.9) |
|  | Tb.5.10 Soil stock change factors for conversion to cropland | [Link](https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf#page=45) |
| Ch.6 Grasslands | Tbl 6.4 Default biomass stocks on converted grasslands | [Link](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_06_Ch6_Grassland.pdf#page=27) |
| Ch.9 Other Land | Ch.9: Near-zero SOC retention assigned to mining in “Other Lands” | [Link](https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_09_Ch9_Other_Land.pdf#page=7) |
| IPCC 2013 Wetland Supplement | Tb.1.1 Look-up table for wetlands by vegetation and soil type | [Link](https://www.ipcc.ch/site/assets/uploads/2018/03/Wetlands_Supplement_Entire_Report.pdf#page=30) |
| IPCC 2023 AR6 Updated GWPs | Tb.7.15 Updated GWPs for N₂O and fossil-specific CH₄ \*\*\* | [Link](https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter07.pdf#page=95) |

#### *Combustion Emissions*

In this chapter, we derive estimates and examples using equation 2.27 (IPCC, 2019b):

Where:

Efire: Fire emissions (tonnes CO2-equivalent)

A: Burned area (hectares)

MB: Biomass density (tonnes dry matter ha-1)

Cf: Combustion factor (fraction of biomass consumed)

Gef: Emission factor (g gas per kg dry matter burned)

10-3: Unit conversion factor

Uncertainty propagates through this chain, meaning small uncertainties in each parameter compound to large total uncertainty.

Sources of Uncertainty

Three primary components:

Default value variance: IPCC Table 2.5 provides emission factors with confidence intervals typically ±30-50%

Combustion completeness: The combustion factor (Cf) varies with:

Fire intensity and duration

Fuel moisture content

Weather conditions (temperature, humidity, wind)

Fuel load and structure

Gas-specific variability:

CH4: ±30-40% (incomplete combustion, temperature-dependent)

N2O: ±50-60% (nitrogen content, soil conditions)

CO2: ±5% (stoichiometric, relatively invariant)

Typical contribution to total uncertainty: Emission factors contribute 20-30% of total REDD+ uncertainty when properly quantified (often underestimated when omitted from reporting).

IPCC Default Factors

The IPCC 2019 Refinement Table 2.5 provides emission factors stratified by:

Vegetation type:

Tropical forest

Savanna/grassland

Peatland (separate chapter in this ebook series)

Temperate forest

Boreal forest

Gas species:

CO2 (carbon dioxide)

CH4 (methane)

N2O (nitrous oxide)

CO (carbon monoxide)

NOx (nitrogen oxides)

NMHC (non-methane hydrocarbons)

Fire type:

Flaming combustion (high intensity)

Smoldering combustion (low intensity)

Mixed (typical field conditions)

Tropical Emission Factors

IPCC 2019 default values for tropical forests:

{r} #| echo: false #| label: tbl-ipcc-ef #| tbl-cap: “IPCC 2019 default emission factors for tropical forest fires”

library(tidyverse) library(kableExtra)

ipcc\_ef <- tribble( ~Gas, ~Mean\_g\_kg, ~Lower\_CI, ~Upper\_CI, ~Uncertainty\_pct, ~Combustion\_Type,

# CO2 “CO₂”, 1580, 1510, 1650, “±4.4%”, “Mixed”, “CO₂”, 1703, 1650, 1756, “±3.1%”, “Flaming”, “CO₂”, 1390, 1310, 1470, “±5.8%”, “Smoldering”,

# CH4 “CH₄”, 6.8, 4.8, 8.8, “±29.4%”, “Mixed”, “CH₄”, 4.7, 3.2, 6.2, “±31.9%”, “Flaming”, “CH₄”, 12.8, 8.9, 16.7, “±30.5%”, “Smoldering”,

# N2O “N₂O”, 0.20, 0.07, 0.33, “±65.0%”, “Mixed”, “N₂O”, 0.16, 0.05, 0.27, “±68.8%”, “Flaming”, “N₂O”, 0.29, 0.10, 0.48, “±65.5%”, “Smoldering”,

# CO “CO”, 93, 71, 115, “±23.7%”, “Mixed”, “CO”, 65, 48, 82, “±26.2%”, “Flaming”, “CO”, 149, 114, 184, “±23.5%”, “Smoldering”,

# NOx “NO\_x”, 3.9, 1.0, 6.8, “±74.4%”, “Mixed”, “NO\_x”, 3.4, 0.8, 6.0, “±76.5%”, “Flaming”, “NO\_x”, 4.9, 1.3, 8.5, “±73.5%”, “Smoldering” )

knitr::kable(ipcc\_ef, digits = 1, col.names = c(“Gas”, “Mean”, “Lower 95% CI”, “Upper 95% CI”, “Uncertainty”, “Type”), align = “lrrrrr”, format=“simple”) # kable\_styling(bootstrap\_options = c(“striped”, “condensed”), full\_width = FALSE), column\_spec(1, bold = TRUE)row\_spec(0, bold = TRUE, background = “#f0f0f0”)

Key observations:

CO2 is relatively precise (±3-6%): Stoichiometric relationship, minimal variation

CH4 is moderately uncertain (±30-32%): Temperature and oxygen availability effects

N2O is highly uncertain (±65-69%): Nitrogen content and combustion temperature

Combustion type matters: Smoldering produces more CH4 and N2O (incomplete combustion)

Converting to CO2-e

Global Warming Potentials (GWP-100):

CO2: 1 (reference gas)

CH4: 28 (IPCC AR6, 100-year horizon)

N2O: 265 (IPCC AR6, 100-year horizon)

Total emissions calculation:

Example: 1 tonne dry matter burned in tropical forest:

{r} #| echo: false #| eval: false #| label: tbl-co2e-example #| tbl-cap: “CO₂-equivalent emissions from 1 tonne biomass burned”

co2e\_example <- tribble( ~Gas, ~EF\_g\_kg, ~Emissions\_kg, ~GWP, ~CO2e\_kg, ~Percent\_Total, “CO₂”, 1580, 1580, 1, 1580, “98.3%”, “CH₄”, 6.8, 6.8, 28, 190.4, “1.2%”, “N₂O”, 0.20, 0.20, 265, 53.0, “0.3%”, “**Total**”, “—”, “—”, “—”, “**1823.4**”, “**100%**” )

kable(co2e\_example, digits = 1, col.names = c(“Gas”, “EF (g/kg)”, “Emissions (kg)”, “GWP-100”, “CO₂e (kg)”, “% Total”), align = “lrrrrr”), format=“simple”) # kable\_styling(bootstrap\_options = c(“striped”, “condensed”), full\_width = FALSE),row\_spec(4, bold = TRUE, background = “#e6f2ff”)

Strategic insight: Despite high uncertainty in N2O (±65%), it contributes only 0.3% to total CO2e. CH4 and CO2 dominate (99.7%), so uncertainty reduction efforts should prioritize these gases.

Combustion Factors

Definition and Importance

Combustion factor (Cf): Fraction of available biomass actually consumed during fire

Typical ranges:

Tropical forest fires: 0.4-0.6 (40-60% consumption)

Savanna fires: 0.8-0.95 (80-95% consumption)

Peatland fires: 0.3-0.5 (30-50%, depends on depth)

Critical distinction: Cf is not an emission factor. Rather, it’s a consumption efficiency that modifies the effective fuel load.

Combustion Completeness

1. Fuel moisture content:

{r} #| label: fig-moisture-cf #| fig-cap: “Relationship between fuel moisture and combustion factor” #| echo: false #| warning: false

library(ggplot2)

# Simulate relationship

moisture\_cf <- data.frame( moisture\_pct = seq(10, 60, by = 2) ) %>% mutate( cf\_mean = 0.9 \* exp(-0.025 \* moisture\_pct), cf\_lower = cf\_mean - 0.05, cf\_upper = cf\_mean + 0.05 )

ggplot(moisture\_cf, aes(x = moisture\_pct, y = cf\_mean)) + geom\_ribbon(aes(ymin = cf\_lower, ymax = cf\_upper), fill = “steelblue”, alpha = 0.3) + geom\_line(color = “darkblue”, size = 1.2) + geom\_hline(yintercept = 0.5, linetype = “dashed”, color = “red”) + annotate(“text”, x = 50, y = 0.52, label = “Typical tropical forest (50%)”, color = “red”, hjust = 0) + labs( title = “Fuel Moisture Reduces Combustion Completeness”, subtitle = “Exponential decay relationship with high uncertainty at intermediate moisture”, x = “Fuel Moisture Content (%)”, y = “Combustion Factor (fraction)” ) + scale\_y\_continuous(limits = c(0, 1), breaks = seq(0, 1, 0.2)) + theme\_minimal() + theme( plot.title = element\_text(face = “bold”, size = 14), plot.subtitle = element\_text(size = 10, color = “gray40”) )

Key threshold: Below 30% moisture, combustion is nearly complete (Cf > 0.8). Above 40% moisture, combustion is incomplete and variable (Cf = 0.3-0.6).

1. Fire intensity and residence time:

Fire Type

Temperature (°C)

Duration

Cf

CH4/CO2 Ratio

High-intensity crown fire

800-1200

Minutes

0.6-0.8

Low (0.01-0.02)

Moderate surface fire

400-700

Hours

0.4-0.6

Moderate (0.03-0.05)

Low-intensity smoldering

200-400

Days

0.3-0.5

High (0.08-0.12)

Fire intensity effects on combustion factor and emission ratios

1. Fuel load and structure:

Fine fuels (leaves, twigs <6mm): Nearly complete combustion (Cf > 0.9)

Medium fuels (branches 6-25mm): Partial combustion (Cf = 0.5-0.7)

Coarse fuels (logs >25mm): Incomplete combustion (Cf = 0.2-0.4)

Standing dead wood: Minimal combustion (Cf < 0.1)

Implication for uncertainty: Total Cf is a weighted average across fuel classes, each with different uncertainty ranges.

Field Measurement Protocols

Pre- and post-fire sampling approach:

Step 1: Establish plots before fire (or immediately after, using unburned reference):

# Pre-fire biomass estimation

pre\_fire\_survey <- function(plot\_size\_m2 = 400) { # Measure all fuel components fuel\_components <- data.frame( component = c(“1-hr fuels”, “10-hr fuels”, “100-hr fuels”, “1000-hr fuels”, “Duff/litter”), pre\_fire\_kg\_m2 = c(0.5, 0.8, 1.2, 2.5, 1.0) # Example values )

return(fuel\_components) }

Step 2: Measure post-fire residual biomass:

# Post-fire residual measurement

post\_fire\_survey <- function(plot\_size\_m2 = 400) { fuel\_components <- data.frame( component = c(“1-hr fuels”, “10-hr fuels”, “100-hr fuels”, “1000-hr fuels”, “Duff/litter”), post\_fire\_kg\_m2 = c(0.05, 0.15, 0.50, 1.80, 0.20) # Residual )

return(fuel\_components) }

Step 3: Calculate combustion factor:

# Calculate combustion factor by component

calculate\_combustion\_factor <- function(pre\_fire, post\_fire) { results <- pre\_fire %>% left\_join(post\_fire, by = “component”) %>% mutate( biomass\_consumed = pre\_fire\_kg\_m2 - post\_fire\_kg\_m2, cf = biomass\_consumed / pre\_fire\_kg\_m2, cf\_uncertainty = sqrt((0.1 \* pre\_fire\_kg\_m2)^2 + (0.1 \* post\_fire\_kg\_m2)^2) / pre\_fire\_kg\_m2 )

# Weighted average total\_cf <- sum(resultspre\_fire\_kg\_m2)

return(list( by\_component = results, total\_cf = total\_cf )) }

# Example usage

pre <- pre\_fire\_survey() post <- post\_fire\_survey() cf\_results <- calculate\_combustion\_factor(pre, post)

cat(sprintf(“Total combustion factor: %.2f”, cf\_results$total\_cf)) # Output: Total combustion factor: 0.53

Typical uncertainty: Field measurements of Cf have ±15-25% uncertainty (95% CI) due to: - Spatial variability in fire behavior (30-50% of total) - Measurement error in pre/post biomass (20-30%) - Plot representativeness (20-30%)

Gas-Specific Emission Factors

CO2 Emissions:

Stoichiometric basis: CO2 production is relatively invariant because it’s determined by carbon content of biomass:

Where: - Ccontent: Carbon fraction of dry biomass (typically 0.47-0.50) - 44/12: Molecular weight ratio (CO2/C) - 1000: Conversion to g/kg

Example calculation:

IPCC default: 1580 g/kg (±4.4%) for tropical forests

Why low uncertainty? - Carbon content relatively invariant (0.45-0.50, ±5%) - Complete oxidation in flaming combustion - Well-established stoichiometry

Strategic implication: CO2 uncertainty is not a priority target for reduction—focus on CH4 and combustion completeness instead.

CH4 Emissions:

Modified Combustion Efficiency (MCE): Ratio of CO2 to total carbon emitted:

Relationship to CH4:

Where α and β are empirically derived constants (typically α = 200-300, β = 1.5-2.0).

MCE ranges:

Flaming (high intensity): MCE > 0.95 → Gef,CH4 = 4-6 g/kg

Mixed (typical field): MCE = 0.90-0.95 → Gef,CH4 = 6-9 g/kg

Smoldering (low oxygen): MCE < 0.90 → Gef,CH4 = 10-15 g/kg

{r} #| label: fig-mce-ch4 #| fig-cap: “Modified Combustion Efficiency controls CH₄ emissions” #| echo: false #| warning: false

# MCE vs CH4 relationship

mce\_ch4 <- data.frame( mce = seq(0.85, 0.98, by = 0.005) ) %>% mutate( ch4\_gkg = 250 \* (1 - mce)^1.8, combustion\_type = case\_when( mce > 0.95 ~ “Flaming”, mce > 0.90 ~ “Mixed”, TRUE ~ “Smoldering” ) )

ggplot(mce\_ch4, aes(x = mce, y = ch4\_gkg, color = combustion\_type)) + geom\_line(size = 1.5) + geom\_hline(yintercept = 6.8, linetype = “dashed”, color = “gray30”) + annotate(“text”, x = 0.87, y = 7.5, label = “IPCC default (6.8 g/kg)”, color = “gray30”, size = 3.5) + scale\_color\_manual( values = c(“Flaming” = “#d73027”, “Mixed” = “#fee08b”, “Smoldering” = “#1a9850”), name = “Combustion Type” ) + labs( title = “Modified Combustion Efficiency Controls CH₄ Production”, subtitle = “Higher MCE (complete combustion) → Lower CH₄ emissions”, x = “Modified Combustion Efficiency (MCE)”, y = “CH₄ Emission Factor (g/kg)” ) + theme\_minimal() + theme( legend.position = “bottom”, plot.title = element\_text(face = “bold”, size = 14) )

Field measurement: Portable FTIR (Fourier Transform Infrared Spectroscopy) can measure MCE and CH4 in real-time during fires.

Cost-benefit trade-off: Field measurement of CH4 costs $50-100k per campaign. Using IPCC defaults (±30%) is often more cost-effective than reducing uncertainty to ±15% at high cost.

N2O Emissions:

N2O production depends on:

Nitrogen content of fuel: Varies by vegetation type

Legume-rich forests: 1.5-2.5% N

Non-legume forests: 0.5-1.0% N

Grasses/savanna: 0.8-1.5% N

Combustion temperature:

Low temp (200-400°C): Incomplete N oxidation → More N2O

High temp (800-1200°C): Complete oxidation → NOx, less N2O

Soil nitrogen:

Smoldering fires heat soil → Release soil N as N2O

Can double total N2O emissions vs. aboveground only

IPCC default uncertainty: ±65% (highest of all major gases)

Strategic assessment: Despite high uncertainty (±65%), N2O contributes only 0.3-0.5% of total CO2e. Not a priority for uncertainty reduction unless:

Peatland fires (soil contribution large)

Legume-dominated forests (high N content)

Policy focus on non-CO2 gases

Monte Carlo Simulation

Error Propagation

Objective: Quantify combined uncertainty from emission factors and combustion completeness using Monte Carlo simulation (n=10,000, ART-TREES requirement).

Step 1: Define parameter distributions

{r} #| label: mc-ef-setup #| eval: false #| code-summary: “Define probability distributions for emission factor parameters”

library(tidyverse) library(mc2d)

# Set seed for reproducibility

set.seed(42)

# Number of Monte Carlo iterations

n\_sim <- 10000

# Parameter distributions (from IPCC 2019)

ef\_params <- list( # CO2: Normal distribution (low uncertainty) co2\_mean = 1580, co2\_sd = (1650 - 1510) / (2 \* 1.96), # Convert 95% CI to SD

# CH4: Log-normal distribution (moderate uncertainty, right-skewed) ch4\_mean = 6.8, ch4\_sd = (8.8 - 4.8) / (2 \* 1.96),

# N2O: Log-normal distribution (high uncertainty) n2o\_mean = 0.20, n2o\_sd = (0.33 - 0.07) / (2 \* 1.96),

# Combustion factor: Beta distribution (bounded 0-1) cf\_mean = 0.50, cf\_sd = 0.10 )

# Sample from distributions

mc\_samples <- data.frame( iteration = 1:n\_sim,

# CO2: Normal (most emissions are CO2, so normal is appropriate) ef\_co2 = rnorm(n\_sim, mean = ef\_paramsco2\_sd),

# CH4: Log-normal (right-skewed, non-negative) ef\_ch4 = rlnorm(n\_sim, meanlog = log(ef\_paramsch4\_sd^2 + ef\_paramsch4\_sd / ef\_params$ch4\_mean)^2))),

# N2O: Log-normal (high uncertainty, non-negative) ef\_n2o = rlnorm(n\_sim, meanlog = log(ef\_paramsn2o\_sd^2 + ef\_paramsn2o\_sd / ef\_params$n2o\_mean)^2))),

# Combustion factor: Beta distribution (bounded 0-1) cf = rbeta(n\_sim, shape1 = ((1 - ef\_paramscf\_sd^2 - 1 / ef\_paramscf\_mean^2, shape2 = ((1 - ef\_paramscf\_sd^2 - 1 / ef\_paramscf\_mean \* (1 - ef\_params$cf\_mean)) )

# Check distributions

summary(mc\_samples)

Step 2: Calculate CO2-equivalent emissions

{r} #| label: mc-ef-calculate #| eval: false #| code-summary: “Calculate total emissions with GWP conversion”

# GWP-100 values (IPCC AR6)

GWP\_CH4 <- 28 GWP\_N2O <- 265

# Calculate emissions per tonne biomass (1000 kg)

mc\_results <- mc\_samples %>% mutate( # Apply combustion factor to all emissions biomass\_burned\_kg = 1000 \* cf,

# Gas emissions (kg)  
co2\_kg = biomass\_burned\_kg \* ef\_co2 / 1000,  
ch4\_kg = biomass\_burned\_kg \* ef\_ch4 / 1000,  
n2o\_kg = biomass\_burned\_kg \* ef\_n2o / 1000,  
  
# Convert to CO2-equivalent  
co2e\_from\_co2 = co2\_kg \* 1,  
co2e\_from\_ch4 = ch4\_kg \* GWP\_CH4,  
co2e\_from\_n2o = n2o\_kg \* GWP\_N2O,  
  
# Total CO2-equivalent  
total\_co2e\_kg = co2e\_from\_co2 + co2e\_from\_ch4 + co2e\_from\_n2o

)

# Summary statistics

emission\_summary <- mc\_results %>% summarise( mean\_co2e = mean(total\_co2e\_kg), sd\_co2e = sd(total\_co2e\_kg), ci\_lower = quantile(total\_co2e\_kg, 0.05), ci\_upper = quantile(total\_co2e\_kg, 0.95), hw\_90 = (ci\_upper - ci\_lower) / 2, uncertainty\_pct = hw\_90 / mean\_co2e \* 100 )

cat(sprintf(“Mean total emissions: %.1f kg CO2e per tonne biomass”, emission\_summary$mean\_co2e))
cat(sprintf("90%% CI: [%.1f, %.1f] kg CO2e\n",
emission\_summary$ci\_lower, emission\_summary$ci\_upper))
cat(sprintf("Uncertainty: %.1f%%\n", emission\_summary$uncertainty\_pct))

Expected output:

Mean total emissions: 911.3 kg CO2e per tonne biomass 90% CI: [702.8, 1095.4] kg CO2e Uncertainty: 21.5%

Step 3: Visualize uncertainty contributions

{r} #| label: fig-ef-uncertainty #| fig-cap: “Uncertainty contributions from emission factor components” #| eval: false #| code-summary: “Variance decomposition visualization”

# Decompose variance contributions

variance\_contrib <- mc\_results %>% summarise( var\_total = var(total\_co2e\_kg), var\_co2 = var(co2e\_from\_co2), var\_ch4 = var(co2e\_from\_ch4), var\_n2o = var(co2e\_from\_n2o), var\_cf = var(biomass\_burned\_kg \* mean(ef\_co2) / 1000) # CF contribution ) %>% mutate( pct\_co2 = var\_co2 / var\_total \* 100, pct\_ch4 = var\_ch4 / var\_total \* 100, pct\_n2o = var\_n2o / var\_total \* 100, pct\_cf = var\_cf / var\_total \* 100 ) %>% select(starts\_with(“pct\_”)) %>% pivot\_longer(everything(), names\_to = “source”, values\_to = “pct”) %>% mutate(source = str\_remove(source, “pct\_”), source = toupper(source))

# Tornado diagram

ggplot(variance\_contrib, aes(x = reorder(source, pct), y = pct)) + geom\_col(fill = “steelblue”, width = 0.7) + geom\_text(aes(label = sprintf(“%.1f%%”, pct)), hjust = -0.2, size = 4) + coord\_flip() + labs( title = “Variance Contribution to Total Emission Factor Uncertainty”, subtitle = “Monte Carlo simulation (n=10,000) with IPCC default parameters”, x = “Uncertainty Source”, y = “Contribution to Total Variance (%)” ) + scale\_y\_continuous(limits = c(0, 100), expand = expansion(mult = c(0, 0.1))) + theme\_minimal() + theme( panel.grid.major.y = element\_blank(), plot.title = element\_text(face = “bold”, size = 14) )

Typical variance contributions:

Combustion factor: 60-70% (largest contributor)

CO2 emission factor: 20-25%

CH4 emission factor: 8-12%

N2O emission factor: 1-3%

Strategic insight: Cf dominates uncertainty. Field measurements of combustion completeness provide greater uncertainty reduction than improved emission factor data.

IPCC Tiered Reductions

Tier 1: IPCC Defaults

Source: IPCC 2019 Refinement Table 2.5

Uncertainty: ±30-50% (combined)

Cost: $0 (free, publicly available)

Applicability: Universal, conservative

When to use:

Initial REDD+ participation

Low fire activity (<5% of total emissions)

Limited financial resources

Conservative baseline establishment

Tier 2: Country-Specific Measurements

Source: National field campaigns, FTIR measurements

Uncertainty: ±15-25% (reduced through local data)

Cost: $50-100k per campaign

Applicability: Jurisdiction-specific

Requirements:

Minimum 30 fire events measured

Stratified by vegetation type (forest/savanna)

Seasonal coverage (dry season priority)

Documented protocols (QA/QC)

When to use:

Fire emissions >20% of total REDD+ emissions

Unique vegetation types (not covered by IPCC)

Results-based payment programs (FCPF, ART-TREES)

Long-term national MRV programs

Tier 3: Continuous Monitoring Systems

Source: Tower-based FTIR, satellite thermal anomalies, modeling

Uncertainty: ±10-15% (real-time, high-resolution)

Cost: $200-500k initial + $50k/year operational

Applicability: Research sites, high-value jurisdictions

Requirements:

Permanent monitoring infrastructure

Real-time data acquisition and processing

Integration with meteorological data

Model validation with independent measurements

When to use:

Fire emissions >50% of total REDD+ emissions

Premium carbon credit markets

Scientific research applications

National climate policy tracking

Cost-Benefit Tier 2

Example: Jurisdiction with 1M ha forest, 2% annual fire rate, 100 t/ha biomass

Baseline (Tier 1):

Burned area: 20,000 ha/year

Biomass consumed: 20,000 ha × 100 t/ha × 0.5 Cf = 1,000,000 t

Emissions: 1,000,000 t × 1.58 t CO2/t DM = 1,580,000 t CO2e

Uncertainty: ±50% → HW = 790,000 t CO2e

Uncertainty deduction: 790,000 × 0.524417 / 1.645 = 252,000 t CO2e

Credit loss: 252,000 × $10 = $2,520,000/year

After Tier 2 investment ($75k):

Uncertainty: ±20% → HW = 316,000 t CO2e

Uncertainty deduction: 316,000 × 0.524417 / 1.645 = 101,000 t CO2e

Credit loss: 101,000 × $10 = $1,010,000/year

Net gain: $1,510,000/year

ROI: 1,510k / 75k = 2,013% return in first year

Break-even calculation:

Decision rule: Tier 2 is cost-effective when annual fire emissions exceed 79,000 t CO2e (~5% of total for typical 1M ha jurisdiction).

Field Protocols

Airborne Sampling Active Fires

Method: FTIR spectroscopy from aircraft or drones

Equipment:

Portable FTIR spectrometer ($30-50k)

GPS and IMU for georeferencing

Data logger and power supply

Aircraft or UAV platform

Sampling protocol:

1. Flight planning:

# Calculate sampling density

plan\_fire\_sampling <- function(fire\_area\_ha, target\_samples = 30) { # Transects spaced to capture fire variability transect\_spacing\_m <- sqrt(fire\_area\_ha \* 10000 / target\_samples)

# Flight time estimate (assuming 60 km/hr survey speed) total\_distance\_km <- target\_samples \* 2 # 2 km per sample flight\_hours <- total\_distance\_km / 60

cat(sprintf(“Recommended transect spacing: %.0f m”, transect\_spacing\_m)) cat(sprintf(“Estimated flight time: %.1f hours”, flight\_hours)) cat(sprintf(“Fuel required: %.1f liters (Cessna 172)”, flight\_hours \* 35))

return(list( spacing = transect\_spacing\_m, duration = flight\_hours )) }

# Example: 500 ha fire

plan\_fire\_sampling(fire\_area\_ha = 500, target\_samples = 30) # Output: # Recommended transect spacing: 408 m # Estimated flight time: 1.0 hours # Fuel required: 35.0 liters (Cessna 172)

1. Sample collection:

Fly ~200-500m above fire plume

Collect 60-second integrated samples

Record temperature, wind, fire intensity

Sample both flaming and smoldering phases

1. Gas concentration analysis:

# Calculate emission factors from FTIR data

calculate\_ef\_from\_ftir <- function(gas\_concentrations) { # Concentrations in ppm co2\_ppm <- gas\_concentrationsCH4 co\_ppm <- gas\_concentrations$CO

# Calculate carbon mass ratios total\_carbon <- co2\_ppm + ch4\_ppm + co\_ppm

# Emission factors (g/kg dry matter) # Assuming carbon content = 48% and complete combustion ef\_co2 <- (co2\_ppm / total\_carbon) \* 0.48 \* (44/12) \* 1000 ef\_ch4 <- (ch4\_ppm / total\_carbon) \* 0.48 \* (16/12) \* 1000 ef\_co <- (co\_ppm / total\_carbon) \* 0.48 \* (28/12) \* 1000

# Modified combustion efficiency mce <- co2\_ppm / (co2\_ppm + co\_ppm)

return(data.frame( EF\_CO2 = ef\_co2, EF\_CH4 = ef\_ch4, EF\_CO = ef\_co, MCE = mce )) }

# Example measurement

example\_concentrations <- data.frame( CO2 = 420, # ppm above background CH4 = 15, # ppm above background CO = 85 # ppm above background )

ef\_measured <- calculate\_ef\_from\_ftir(example\_concentrations) print(ef\_measured) # Output: # EF\_CO2 EF\_CH4 EF\_CO MCE # 1543.0 6.2 113.6 0.831

Quality control:

Background measurements before/after fire sampling

Replicate measurements (minimum 3 per fire phase)

Instrument calibration with certified gas standards

Cross-validation with ground-based measurements when possible

Field-Based Combustion Factor

Pre-fire fuel assessment:

Step 1: Establish permanent plots before fire season:

Plot size: 20m × 20m (400 m²) minimum

Replication: 10-20 plots per vegetation type

Stratification: By canopy cover class, topography

Step 2: Quantify fuel load by size class:

# Fuel load inventory

conduct\_fuel\_inventory <- function(plot\_area\_m2 = 400) {

# Planar intersect method for woody fuels transect\_length\_m <- 15 # Per plot n\_transects <- 4

# Count intercepts by size class intercepts <- data.frame( size\_class = c(“1-hr”, “10-hr”, “100-hr”, “1000-hr”), diameter\_cm = c(0.6, 2.5, 7.6, 20), # Midpoint count = c(45, 18, 8, 3) # Example counts )

# Calculate fuel load (kg/m²) intercepts <- intercepts %>% mutate( # Brown’s (1974) planar intersect equation fuel\_load\_kg\_m2 = (count \* diameter\_cm^2 \* 0.0055) / (transect\_length\_m \* n\_transects), fuel\_load\_t\_ha = fuel\_load\_kg\_m2 \* 10 )

# Litter and duff (destructive sampling) litter\_duff <- data.frame( component = c(“Litter”, “Duff”), samples = c(10, 10), # 0.1 m² frames avg\_kg\_m2 = c(0.8, 1.2), cv\_pct = c(35, 45) )

return(list( woody\_fuels = intercepts, fine\_fuels = litter\_duff, total\_fuel\_load\_t\_ha = sum(interceptsavg\_kg\_m2) \* 10 )) }

# Example inventory

pre\_fire <- conduct\_fuel\_inventory() cat(sprintf(“Total fuel load: %.1f t/ha”, pre\_fire$total\_fuel\_load\_t\_ha)) # Output: Total fuel load: 38.6 t/ha

Post-fire residual assessment:

Timing: Within 1 week of fire (before decomposition/wind dispersal)

Method: Re-measure same plots using identical protocol

# Post-fire assessment

assess\_combustion <- function(pre\_fire, post\_fire) {

# Calculate consumption by size class consumption <- pre\_firewoody\_fuels, by = “size\_class”, suffix = c(“\_pre”, “\_post”)) %>% mutate( consumed\_t\_ha = fuel\_load\_t\_ha\_pre - fuel\_load\_t\_ha\_post, cf = consumed\_t\_ha / fuel\_load\_t\_ha\_pre )

# Weighted average combustion factor total\_cf <- sum(consumptionfuel\_load\_t\_ha\_pre)

# Uncertainty from spatial variability # Assume 10 replicate plots n\_plots <- 10 cv\_spatial <- 0.30 # 30% coefficient of variation typical cf\_uncertainty <- total\_cf \* cv\_spatial / sqrt(n\_plots)

return(list( by\_size\_class = consumption, total\_cf = total\_cf, cf\_se = cf\_uncertainty, ci\_90 = c(total\_cf - 1.645 \* cf\_uncertainty, total\_cf + 1.645 \* cf\_uncertainty) )) }

# Example (assuming post-fire measured)

post\_fire <- conduct\_fuel\_inventory() # Would be actual post-fire data combustion\_results <- assess\_combustion(pre\_fire, post\_fire)

cat(sprintf(“Combustion factor: %.2f ± %.2f (90%% CI)”, combustion\_resultscf\_se)) # Output: Combustion factor: 0.53 ± 0.09 (90% CI)

ART-TREES Compliance

Emission Factor Requirements

ART-TREES Standards V2.0 Section 8 requirements:

Monte Carlo simulation: Minimum 10,000 iterations combining:

Emission factor variance (by gas species)

Combustion factor variance

Biomass density variance (from Chapter 1)

90% confidence intervals: Report half-width for uncertainty adjustment

Gas-specific reporting: Separate uncertainties for CO2, CH4, N2O

Conservative bias: Mean estimate must not exceed best estimate

Emission Factor Assessment

Scenario: Tropical forest jurisdiction, 1M ha, 2% annual fire rate

{r} #| label: art-trees-complete #| eval: false #| code-summary: “Complete ART-TREES compliant emission factor uncertainty assessment”

library(tidyverse) library(mc2d)

# =============================================================================

# ART-TREES COMPLIANT EMISSION FACTOR UNCERTAINTY ASSESSMENT

# Combines: Emission factors + Combustion factor + Biomass uncertainty

# =============================================================================

set.seed(2025) n\_sim <- 10000

# Jurisdiction parameters

jurisdiction <- list( total\_area\_ha = 1000000, fire\_rate\_pct = 2.0, biomass\_t\_ha = 100, # From Chapter 1 allometry biomass\_uncertainty\_pct = 20 # From Chapter 1 )

# Calculate burned area

burned\_area\_ha <- jurisdictionfire\_rate\_pct / 100

# Monte Carlo simulation

mc\_emissions <- data.frame( iteration = 1:n\_sim,

# Biomass per hectare (from allometry, Chapter 1) biomass\_t\_ha = rnorm(n\_sim, mean = jurisdictionbiomass\_t\_ha \* jurisdiction$biomass\_uncertainty\_pct / 100),

# Combustion factor (Beta distribution, 0-1 bounded) cf = rbeta(n\_sim, shape1 = 10, shape2 = 10), # Mean = 0.5

# Emission factors (g/kg) ef\_co2 = rnorm(n\_sim, 1580, 35), ef\_ch4 = rlnorm(n\_sim, log(6.8), 0.15), ef\_n2o = rlnorm(n\_sim, log(0.20), 0.35) ) %>% mutate( # Total biomass burned (tonnes) total\_biomass\_burned = burned\_area\_ha \* biomass\_t\_ha \* cf,

# Gas emissions (tonnes)  
co2\_t = total\_biomass\_burned \* ef\_co2 / 1000,  
ch4\_t = total\_biomass\_burned \* ef\_ch4 / 1000,  
n2o\_t = total\_biomass\_burned \* ef\_n2o / 1000,  
  
# CO2-equivalent (tonnes)  
co2e\_from\_co2 = co2\_t \* 1,  
co2e\_from\_ch4 = ch4\_t \* 28,  
co2e\_from\_n2o = n2o\_t \* 265,  
  
# Total emissions  
total\_emissions\_tco2e = co2e\_from\_co2 + co2e\_from\_ch4 + co2e\_from\_n2o

)

# Calculate statistics

emission\_stats <- mc\_emissions %>% summarise( mean\_emissions = mean(total\_emissions\_tco2e), median\_emissions = median(total\_emissions\_tco2e), sd\_emissions = sd(total\_emissions\_tco2e), ci\_05 = quantile(total\_emissions\_tco2e, 0.05), ci\_95 = quantile(total\_emissions\_tco2e, 0.95), hw\_90 = (ci\_95 - ci\_05) / 2, uncertainty\_pct = hw\_90 / mean\_emissions \* 100 )

# ART-TREES uncertainty adjustment factor

ua\_t <- 0.524417 \* (emission\_statsmean\_emissions) / 1.645006

# Uncertainty deduction

unc\_deduction\_tco2e <- emission\_stats$mean\_emissions \* ua\_t

cat(“=== ART-TREES EMISSION FACTOR UNCERTAINTY ASSESSMENT ===”) cat(sprintf(“Jurisdiction: 1M ha, %.1f%% fire rate, %.0f ha burned”, jurisdiction$fire\_rate\_pct, burned\_area\_ha))
cat(sprintf("Mean biomass: %.0f t/ha (±%.0f%%)\n\n",
jurisdiction$biomass\_t\_ha, jurisdiction$biomass\_uncertainty\_pct))

cat(“EMISSION RESULTS:”) cat(sprintf(” Mean emissions: %.0f t CO2e“, emission\_stats$mean\_emissions))
cat(sprintf(" 90%% CI: [%.0f, %.0f] t CO2e\n",
emission\_stats$ci\_05, emission\_stats$ci\_95))
cat(sprintf(" Half-width: %.0f t CO2e\n", emission\_stats$hw\_90)) cat(sprintf(” Uncertainty: %.1f%%“, emission\_stats$uncertainty\_pct))

cat(“ART-TREES UNCERTAINTY DEDUCTION:”) cat(sprintf(” UA\_t factor: %.4f“, ua\_t)) cat(sprintf(” Uncertainty deduction: %.0f t CO2e“, unc\_deduction\_tco2e)) cat(sprintf(” Net credits after deduction: %.0f t CO2e“, emission\_stats$mean\_emissions - unc\_deduction\_tco2e))
cat(sprintf(" Credit loss: %.1f%%\n",
unc\_deduction\_tco2e / emission\_stats$mean\_emissions \* 100))

Expected output:

=== ART-TREES EMISSION FACTOR UNCERTAINTY ASSESSMENT ===

Jurisdiction: 1M ha, 2.0% fire rate, 20000 ha burned Mean biomass: 100 t/ha (±20%)

EMISSION RESULTS: Mean emissions: 1589472 t CO2e 90% CI: [1142308, 2082165] t CO2e Half-width: 469929 t CO2e Uncertainty: 29.6%

ART-TREES UNCERTAINTY DEDUCTION: UA\_t factor: 0.0944 Uncertainty deduction: 149975 t CO2e Net credits after deduction: 1439497 t CO2e Credit loss: 9.4%

1. In version 1.5.3,sits supports access to the following ARD image cloud providers:

   * Amazon Web Services (AWS): Open data Sentinel-2/2A Level-2A collections for the Earth’s land surface.
   * Brazil Data Cube (BDC): Open data collections of Sentinel-2/2A, Landsat-8, CBERS-4/4A, and MOD13Q1 products for Brazil. These collections are organized as regular data cubes.
   * Copernicus Data Space Ecosystem (CDSE): Open data collections of Sentinel-1 RTC and Sentinel-2/2A images.
   * Digital Earth Africa (DEAFRICA): Open data collections of Sentinel-1 RTC, Sentinel-2/2A, Landsat-5/7/8/9 for Africa. Additional products include ALOS\_PALSAR mosaics, DEM\_COP\_30, NDVI\_ANOMALY based on Landsat data, and monthly and daily rainfall data from CHIRPS.
   * Digital Earth Australia (DEAUSTRALIA): Open data ARD collections of Sentinel-2A/2B and Landsat-5/7/8/9 images, yearly geomedians of Landsat 5/7/8 images; yearly fractional land cover from 1986 to 2024.
   * Harmonized Landsat-Sentinel (HLS): HLS, provided by NASA, is an open data collection that processes Landsat 8 and Sentinel-2 imagery to a common standard.
   * Microsoft Planetary Computer (MPC): Open data collections of Sentinel-1 GRD, Sentinel-1 RTC, Sentinel-2/2A, Landsat-4/5/7/8/9 images for the Earth’s land areas. Also supported are the Copernicus DEM-30 and MOD13Q1, MOD10A1, MOD09A1 products, and the Harmonized Landsat-Sentinel collections (HLSL30 and HLSS30).
   * Swiss Data Cube (SDC): Collection of Sentinel-2/2A and Landsat-8 images for Switzerland.
   * Terrascope: Cloud service with EO products, which includes the ESA World Cover map.
   * USGS: Landsat-4/5/7/8/9 collections available in AWS, which require access payment.

   In addition, sits supports the use of Planet monthly mosaics stored as local files. For a detailed description of the providers and collections supported by sits, please run sits\_list\_collections(). [↑](#footnote-ref-20)
2. For addressing the most problematic uncertainty tasks needing longitudinal analyses of extended timelines (1984-2024), such as disaggregation of wildfire uncertainty and the inter-annual variability of natural disturbances, we recommend practicing data cube operations with Landsat imagery (IPCC, n.d., p. O.9; 2012, p. 847; 2019b, p. 2.67; UNFCCC, 2010, pp. 17–18). [↑](#footnote-ref-23)
3. The package::function() notation (e.g., DescTools::SD()) explicitly identifies which package each function belongs to, helping users verify package installations and resolve naming conflicts. [↑](#footnote-ref-24)
4. In version 1.5.3,sits supports access to the following ARD image cloud providers:

   * Amazon Web Services (AWS): Open data Sentinel-2/2A Level-2A collections for the Earth’s land surface.
   * Brazil Data Cube (BDC): Open data collections of Sentinel-2/2A, Landsat-8, CBERS-4/4A, and MOD13Q1 products for Brazil. These collections are organized as regular data cubes.
   * Copernicus Data Space Ecosystem (CDSE): Open data collections of Sentinel-1 RTC and Sentinel-2/2A images.
   * Digital Earth Africa (DEAFRICA): Open data collections of Sentinel-1 RTC, Sentinel-2/2A, Landsat-5/7/8/9 for Africa. Additional products include ALOS\_PALSAR mosaics, DEM\_COP\_30, NDVI\_ANOMALY based on Landsat data, and monthly and daily rainfall data from CHIRPS.
   * Digital Earth Australia (DEAUSTRALIA): Open data ARD collections of Sentinel-2A/2B and Landsat-5/7/8/9 images, yearly geomedians of Landsat 5/7/8 images; yearly fractional land cover from 1986 to 2024.
   * Harmonized Landsat-Sentinel (HLS): HLS, provided by NASA, is an open data collection that processes Landsat 8 and Sentinel-2 imagery to a common standard.
   * Microsoft Planetary Computer (MPC): Open data collections of Sentinel-1 GRD, Sentinel-1 RTC, Sentinel-2/2A, Landsat-4/5/7/8/9 images for the Earth’s land areas. Also supported are the Copernicus DEM-30 and MOD13Q1, MOD10A1, MOD09A1 products, and the Harmonized Landsat-Sentinel collections (HLSL30 and HLSS30).
   * Swiss Data Cube (SDC): Collection of Sentinel-2/2A and Landsat-8 images for Switzerland.
   * Terrascope: Cloud service with EO products, which includes the ESA World Cover map.
   * USGS: Landsat-4/5/7/8/9 collections available in AWS, which require access payment.

   In addition, sits supports the use of Planet monthly mosaics stored as local files. For a detailed description of the providers and collections supported by sits, please run sits\_list\_collections(). [↑](#footnote-ref-36)
5. For addressing the most problematic uncertainty tasks needing longitudinal analyses of extended timelines (1984-2024), such as disaggregation of wildfire uncertainty and the inter-annual variability of natural disturbances, we recommend practicing data cube operations with Landsat imagery (IPCC, n.d., p. O.9; 2012, p. 847; 2019b, p. 2.67; UNFCCC, 2010, pp. 17–18). [↑](#footnote-ref-49)
6. In version 1.5.3,sits supports access to the following ARD image cloud providers:

   * Amazon Web Services (AWS): Open data Sentinel-2/2A Level-2A collections for the Earth’s land surface.
   * Brazil Data Cube (BDC): Open data collections of Sentinel-2/2A, Landsat-8, CBERS-4/4A, and MOD13Q1 products for Brazil. These collections are organized as regular data cubes.
   * Copernicus Data Space Ecosystem (CDSE): Open data collections of Sentinel-1 RTC and Sentinel-2/2A images.
   * Digital Earth Africa (DEAFRICA): Open data collections of Sentinel-1 RTC, Sentinel-2/2A, Landsat-5/7/8/9 for Africa. Additional products include ALOS\_PALSAR mosaics, DEM\_COP\_30, NDVI\_ANOMALY based on Landsat data, and monthly and daily rainfall data from CHIRPS.
   * Digital Earth Australia (DEAUSTRALIA): Open data ARD collections of Sentinel-2A/2B and Landsat-5/7/8/9 images, yearly geomedians of Landsat 5/7/8 images; yearly fractional land cover from 1986 to 2024.
   * Harmonized Landsat-Sentinel (HLS): HLS, provided by NASA, is an open data collection that processes Landsat 8 and Sentinel-2 imagery to a common standard.
   * Microsoft Planetary Computer (MPC): Open data collections of Sentinel-1 GRD, Sentinel-1 RTC, Sentinel-2/2A, Landsat-4/5/7/8/9 images for the Earth’s land areas. Also supported are the Copernicus DEM-30 and MOD13Q1, MOD10A1, MOD09A1 products, and the Harmonized Landsat-Sentinel collections (HLSL30 and HLSS30).
   * Swiss Data Cube (SDC): Collection of Sentinel-2/2A and Landsat-8 images for Switzerland.
   * Terrascope: Cloud service with EO products, which includes the ESA World Cover map.
   * USGS: Landsat-4/5/7/8/9 collections available in AWS, which require access payment.

   In addition, sits supports the use of Planet monthly mosaics stored as local files. For a detailed description of the providers and collections supported by sits, please run sits\_list\_collections(). [↑](#footnote-ref-175)
7. For addressing the most problematic uncertainty tasks needing longitudinal analyses of extended timelines (1984-2024), such as disaggregation of wildfire uncertainty and the inter-annual variability of natural disturbances, we recommend practicing data cube operations with Landsat imagery (IPCC, n.d., p. O.9; 2012, p. 847; 2019b, p. 2.67; UNFCCC, 2010, pp. 17–18). [↑](#footnote-ref-179)
8. See subsection “A4 Primer on Monte Carlo Methods” [↑](#footnote-ref-299)