Raport științific anual
privind implementarea proiectului:

„Mobilizarea și monitorizarea efortului cu impact climatic pozitiv din sectorul forestier”

(cod ERANET-FACCE ERAGAS - FORCLIMIT)
Contract 82/2017

Etapa 3: Ajustarea modelelor la nivelul zonelor test locale

Perioada de implementare: 01.01.2019-31.12.2019

1. Introducere. Contextul științific.

Consorțiul FORCLIMIT susține ca potențialul de reducere de emisii asociat pădurilor din Europa este semnificativ, dar insuficient utilizat în cadrul politicilor UE privind clima, și dăm ca exemplu Regulemantul LULUCF (Regulation (EU) 2018/841 of the European Parliament and of the Council of 30 May 2018 on the inclusion of greenhouse gas emissions and removals from land use, land use change and forestry in the 2030 climate and energy framework, and amending Regulation (EU) No 525/2013 and Decision No 529/2013/EU (Text with EEA relevance). Luand nota de insuficienta efortului global de reduceri de emisii, Parlementul European recomanda printr-o rezolutie din Noiembrie 2018 ca UE sa devina neutrala climatic la jumatatea acestui secol, ceea ce justifica si mai mult nevoia de clarificare a contributiei padurii si sectorului folosintei terenului.

Până în prezent, resursele forestiere și sectorul forestier european au compensat aproximativ 13% din emisiile cauzate de utilizarea combustibililor fosili în Europa, reprezentând aproximativ 569 Mt CO2/an (Nabuurs et al., 2015), rezultate din sechestrarea carbonului în păduri și din activități de evitare a reducerilor de emisii. În această propunere, ne concentrăm în mod special pe potențialul de reducere de emisii al pădurilor și al resurselor forestiere (o parte semnificativă a așa numitului sector LULUCF ce include folosinta terenurilor) în cadrul mai larg al sectorului AFOLU (IPCC, 2006), care include Agricultura pe langa LULUCF. Potențialul suplimentar disponibil de reducere de emisii al pădurilor, al solurilor și al resurselor forestiere este ridicat, însă acest potențial este incert, pe de o parte, din cauza lipsei de stimulente din partea politicilor existente și, pe de altă parte, din cauza incertitudinii privind aplicarea și efectele activităților desfășurate în acest sens de proprietarii de păduri și utilizatorii de resurse lemnoase. Noi abordăm aceste două aspecte împreună, deoarece numai astfel pot fi făcute progrese evidente.

FORCLIMIT are trei **obiective principale**:

(1) să analizeze și să propună îmbunătățiri ale cadrului de contabilizare reduceri de emisii intr-un cadru de politici unificate internațional, care să faciliteze o contabilizare consistentă a emisiilor din păduri din diferite țări;

(2) să analizeze strategiile economice și ale politicilor existente în motivarea proprietarilor de terenuri ca aceștia să depună eforturi pentru reducerile de emisii din păduri și lanțul de custodie al lemnului;

(3) să adauge la sistemul MRV actual, care vizează doar estimarea nationala a emisiilor, posibilitatea de estimare îmbunătățită la scara mica, ex. Arboret, unitate de administrare, precum și evaluarea măsurilor economice și a politicilor existente. Acest lucru este demonstrat prin trei studii de caz în trei țări diferite: Olanda, Romania si Suedia.

2. Metode si rezultate

Ca și în anii precedenți, activitățile realizate in 2019 sunt prezentate in format de publicare in anexe. Anumite secțiuni includ informațiile suplimentare din anul curent. Următoarele titluri prezinta succint aspectul stiintific abordat si legatura cu pachetele angajate prin contract. Pentru fiecare titlu sunt mentionati contribuitorii principali. De asemenea se face referire si la articolele publicate (inregistrate si pe platforma la raportarea pentru anul 2019):

a) Rezultate curente privind experimentul privind "cuantificarea descompunerii litierei prin metoda litter bag" (C. Petritan, M. Miclaus, I. Dutca, V. Blujdea)

Rezultatele cumulate obținute de la inceputul priectului sunt prezentate in Anexa 1. Activitatea face parte din WP4. Metodologia initiala a fost descrisa in Raportul anual din primul an de implementare 2017 (Raport 1), aici fiind repetata pentru transparenta si continuitate cu ajustari minime in urma aplicarii in teren. Experimentul asociat a constat în amplasarea a 640 plicuri cu litieră și a 448 plicuri cu lemn mort în 4 tipuri de pădure de pe raza O.S. Pădurile Șincii (jud. Brașov). Experimentul va fi urmărit pentru o perioadă de 3 ani prin prelevare de probe potrivit calendarului din metodologie.

În anul 2019 au fost prelevate câte 5 plicuri de litieră în lunile Aprilie, Iunie, Iulie și Septembrie conform agendei prestabilite in anul 2017 și modificată în 2018. În anul 2020 se vor efectua ultimele 2 recoltări în Mai și Septembrie. De asemenea în lunile Aprilie, Iulie și Octombrie 2019 au fost recoltate câte 6 probe pentru fiecare variantă de studiu în cazul experimentului de descompunere a lemnului de mici dimensiuni (sub 5.6 cm diametru). Și în acest caz am redus de la 8 la 6 numărul de probe pentru fiecare recoltare ceea ce ne oferă avantajul unei prelevări suplimentare dedicată anului 4 (2021) și anului 5 (2022).

Probele au fost recoltate și transportate în laborator unde după câteva zile de uscare la temperatura camerei au fost scoase din plicuri, curățate de orice impuritate externă și uscate 5 zile la 80 grade în etuvă. În urma recântăririi după uscare, am putut determina care a fost procentul de pierdere în biomasă prin raportarea la masa inițială (masa avută la momentul instalării în teren).

În figura 1 (**anexa 1**) este redată dinamica descompunerii frunzelor și acelor în primele 24 luni ale experimentului pentru toate cele 8 variante de studiu, cu punerea accentului pe scoaterea în evidență a variabilității în cadrul fiecărei etape de recoltare. În figura 2 și tabelul 1 din anexa 1 sunt redate modelele și coeficienții aferenți acestora, modele ce descriu relația dintre cantitățile de masă rămasă exprimate ca și procent din masa inițială și timpul de descompunere (exprimat în luni). Cele mai mari rate de descompunere, dar și cele mai mari valori ale coeficienților de determinare ale modelelor au fost înregistrate la specia brad, cu o ușoară tendință de superioritate pentru arboretul virgin comparativ cu cel parcurs cu lucrări. Contrar așteptărilor, fagul, singura specie de foioase din cele trei studiate, prezintă ratele de descompunere cele mai mici, având de asemenea și cele mai mici valori ale coeficientului de determinare pentru modelul exponențial negativ folosit la ajustarea dinamicii descompunerii. Molidul prezintă valori intermediare celorlalte două specii. La speciile de rășinoase, descompunerea în pădurea virgină a fost mai intensă comparativ cu pădurea parcursă, în timp ce la fag a fost depistat un comportament contrar.

Așa cum se poate vedea din Figura 3 a anexei 1, la categoria de lemn foarte subțire (d=0.1-2cm) cea mai mare rată de descompunere s-a înregistrat la specia bradul din pădurea virgină (pierdere în biomasă de 20% în Iulie și 25% în Octombrie), urmată îndeaproape de fagul din arboretul pur (21% Iulie și 23 % Octombrie). La polul opus, cea mai mică rată a fost semnalată la molidișul pur (aproximativ 5% în Iulie 13% în Octombrie). În cadrul clasei de mărime lemn mijlociu (d=2.1-4.0) (figura 3 b), fagul a prezentat cele mai ridicate rate de descompunere, în timp ce molidul cele mai mici. Ratele de descompunere ale lemnului de la cea mai mare categorie de grosime (d=4.1-5.6 cm) (figura 3 c) au fost similare pentru toate cele 8 variante structurate (cu variații între 5 și 10%). În ceea ce privește influența managementului asupra gradelor de descompunere, prin comparația ratelor de descompunere a celor 3 specii din pădurea parcursă cu intervenții silviculturale cu ratele înregistrate în pădurea virgină, s-au găsit diferențe semnificative doar la specia brad și clasa de lemn foarte subțire (o rată mai mare în pădurea neparcursă). Din punct de vedere al influenței amestecului asupra descompunerii, fagul a prezentat în arboretul pur rate de descompunere aproape duble comparativ cu pădurea de amestec atât pentru clasa de diametre mici, precum și pentru clasa de dimensiuni mijlocii.

Conform modelului carbonului din sol și de descompunere a litierei (Yasso15) litiera se descompune în 4 grupuri de componente, așa-numitul AWEN(A-substanțe hidrosolubile în acid, W-substanțe solubile în apă, E-solvenți (ex. etanol sau diclorometan), W-substanțe care nu sunt nici solubile nici hidrosolubile). Am proiectat ca un total de 144 de probe (3 specii x 2 tipuri de material –litieră și lemn de dimensiuni mici x 3 perioade de recoltare – la început de experiment –Noiembrie 207, la mijloc de experiment Noiembrie 2018 si la sfârșit de proiect Noiembrie 2019 x 8 replicații) să fie trimise în Finlanda, la partenerul finlandez, care pe baza protocolului aferent să fie determinate aceste 4 grupuri de componente. Au fost obținute rezultatele analizelor primului set de date și se află în lucru în laboratorul finlandez setul al doilea corespunzător momentului 2 (noiembrie 2018), urmând ca în cel mai scurt timp să fie trimisă și a treia rundă de probe la doi ani după începerea experimentului.

Conform acestor prime analize obținute, componenta A (substanțe hidrosolubile în acid) este semnificativ mai mare la lemnul de fag comparativ cu cel de brad și molid, în timp ce

componentele E și N sunt semnificativ mai mici la lemnul de fag comparativ cu cel de molid și brad (între conifere neexistând diferențe semnificative). Componenta W deși este mai redusă în lemnul de fag decât în lemnul rășinoaselor, diferențele între cele trei specii nu sunt semnificative (Anexa I, figura 4). În ceea ce privește procentul de participare al fiecărei componente AWEN în cazul descompunerii frunzelor/acelor (Anexa I, figura 5), se poate observa cum componentele N și A, la frunzele de fag sunt semnificativ mai mari decât cele ale rășinoaselor, în timp ce componenta W prezintă un comportament opus (valoarea minimă fiind întâlnită la fag). Referitor la componenta E, bradul prezintă valoarea cea mai indicată și molidul pe cea mai scăzută, fagul posedând o valoare intermediară (totuși cu diferențe semnificative între toate cele 3 specii).

În ceea ce privește relația dintre componentele lemnului de mici dimensiuni și diametrul pieselor eșantionate putem concluziona următoarele:

-la specia fag nu a existat nici o legătură semnificativă între variația diametrului și cele 4 componente structurale.

-la ambele specii de rășinoase s-a identificat o corelație pozitivă între componenta A și diametru (crește procentul componentei A cu creșterea diametrului), precum și o corelație negativă între componenta W și diametru (pe măsură ce crește diametrul scade proporția componentei W în compoziția /structura lemnului de brad și molid).

În plus, la mijlocul anului 2019 au fost descărcate informațiile înregistrate în ultimele 12 lunii către senzorii de temperatură din aer și din sol, în figura 6 din Anexa I fiind redate spre exemplificare locația de molidiș pur parcurs cu lucrări silvotehnice. Aceste date climatice vor fi folosite în final la parametrizarea și validarea modelului Yasso 15 pentru zonele test ale proiectului.

b) Calibrarea si validarea stocurilor de C din sol cu CBM-CFS3 si Yasso15 si validarea cu măsurătorile din Inventarul Forestier National (V. Blujdea, Gh. Marin)

Progresul principal in anul 2019 îl reprezintă parametrizarea, calibrarea și validarea modelelor de simulare a stocurilor de C din solurile forestiere la nivel local, respectiv asociate combinației celor 10 tipuri de pădure cu 5 climate (definite de parametrii climatici) angajat prin pachetul "Combinarea măsurătorilor de carbon din sol cu modelarea pentru inventariere eficientă din punct de vedere al costurilor (WP4)". Este un progres pentru că în general, modelele sunt creditate ca furnizori de estimări precise la nivel agregat la scara mare (i.e. național), dar nu și la nivel local (Didion et al., 2015).

Pentru simularea și validarea dinamicii stocului de masă organică moartă/ carbon asociate solului mineral in 10 tipuri de pădure combinat cu 5 tipuri de climate se au in vedere următoarele depozite de carbon: lemn mort (dead wood - DW), litieră (LT- Litter) și materie organică specifică solului (soil organic carbon - SOC) C din sol (**Anexa 2c**). Următorii pași sunt efectuați:

a) parametrizarea CBM-CFS3 cu date din IFN2, respectiv Procesarea cu metode statistice state-of-the-art de curbe actualizate ale creșterii nete curente anuale si ale volumului de lemn măsurate de Inventarul Forestier National al Romaniei (sprijin acordat de Oregon State University, Corvallis, USA) rata mortalității anuale si input de biomasa potrivit Anexa 1c;
b) calibrarea CBM-CFS3 cu stocurile de lemn mort raportate de IFN1 si IFN2;
c) alimentarea Yasso15 cu input de biomasa armonizat ca output din CBM-CFS;
d) validarea cu valorile stocurilor de C măsurate în IFN (2010) cumulate pentru cele trei depozite (lemn mort, litieră si sol mineral).

Modelul de descompunere asociat materiei organice moarte incorporat de **CBM-CFS3** include nouă depozite de C (vezi figura următoare) care sunt grupate in coloana „GPG pool" in cele trei depozite de carbon: litiera, lemn mort si sol mineral.

Table 2 – Correspondence between pools in the Carbon Budget Model of the Canadian Forest Sector 3—version 1.1 (CBM-CFS3) and recommended pools by the Intergovernmental Panel on Climate Change Good Practice Guidance (GPG) (IPCC, 2003). SW = softwood, HW = hardwood, DOM = dead organic matter.

CBM-CFS3 pool	Description	GPG pool
Merchantable + bark (SW or HW)	Live stemwood of merchantable size[a] plus bark	Aboveground biomass
Other wood + bark (SW or HW)	Live branches, stumps and small trees including bark	Aboveground biomass
Foliage (SW or HW)	Live foliage	Aboveground biomass
Fine roots (SW or HW)	Live roots, approximately <5 mm diameter	Belowground biomass
Coarse roots (SW or HW)	Live roots, approximately ≥5 mm diameter	Belowground biomass
Snag stems DOM (SW or HW)	Dead standing stemwood of merchantable size including bark	Dead wood
Snag branches DOM (SW or HW)	Dead branches, stumps and small trees including bark	Dead wood
Medium DOM	Coarse woody debris on the ground	Dead wood
Aboveground fast DOM	Fine and small woody debris plus dead coarse roots in the forest floor, approximately ≥5 and <75 mm diameter	Litter
Aboveground very fast DOM	The L horizon[b] comprised of foliar litter plus dead fine roots, approximately <5 mm diameter	Litter
Aboveground slow DOM	F, H and O horizons[b]	Litter
Belowground fast DOM	Dead coarse roots in the mineral soil, approximately ≥5 diameter	Dead wood
Belowground very fast DOM	Dead fine roots in the mineral soil, approximately <5 mm diameter	Soil organic matter
Belowground slow DOM	Humified organic matter in the mineral soil	Soil organic matter

Modelul și baza de date Yasso15 sunt adecvate pentru generarea de estimări îmbunătățite pentru carbonul din soluri (Liski et al., 2005, Tuomi et al., 2011, Järvenpää et al., 2017). Actuala bază de date cu care este parametrizat Yasso15 include deja aproximativ 20000 de măsurători legate de sol din diferite părți ale lumii (stocurile de carbon din sol și schimbarea lor, descompunerea litierei, etc.). Rezultatele sunt prezentate in anexa 2c. In general se constata ca ambele modele supraestimează stocul de C din sol, iar CBM simulează sistematic valori ale stocului mai mari decât Yasso15. Urmează a fi inițiate procedurile de QA/QC pentru bazele de date utilizate, si de calibrare si armonizare a parametrilor de descompunere utilizați de celle doua modele.

Prin sprijinul acordat de Oregon State University, Corvallis, USA s-au realizat de curbe actualizate de creștere netă curentă anuală și actualizarea volumului de lemn din păduri. Curbele obținute reprezintă elementele de baza in realizarea scenariilor de reduceri de emisii de gaze cu efect de sera asociate gospodăririi pădurii prin utilizare de modele empirice prevăzuta in sarcinile 3, 5 si 6 din proiect. Rezultatele pre-analizei validității la scara

locala/regionala si parte din procesarea statistica a modelarii creșterii arboretelor si stocurilor de volum pe picior sunt prezentate in Anexa 7a,7b,7c si 7d. Acestea au constat in: au fost efectuate cercetări folosind informațiile de pe carotele de creștere recoltate pentru principalele 9 specii forestiere din România, respectiv fag, molid, gorun, brad, cer, garniță, stejar, carpen și salcâm, care acoperă împreună aproximativ 60% din pădurile țării. Modele de creștere folosite au diferit de la o specie forestieră la alta. Au fost folosite coduri SAS și s-au obținut corelații de 84% între parametrii luați în considerare, pentru 8 din cele 9 specii studiate, la nivel de țară. Pentru una dintre specii (fag) trebuie continuate cercetările, deoarece corelația obținută a fost mai slabă. Deoarece s-a studiat și variabilitatea spațială a celor 9 specii, pentru 5 dintre ele s-au elaborat modele de creștere, cu același grad de corelație, și la nivel de ecoregiuni omogene din punct de vedere al condițiilor naturale de creștere.

c) Estimarea dinamicii stocului de carbon folosind modelul Yasso 15, simulare și parametrizare locală în condiții de schimbare a folosinței terenului la/de la pădure (M. Miclaus).

Pentru a intelege contribuția schimbării folosinței terenului la bilanțul emisiilor gazelor cu efect de seră (GES) – în special a dioxidului de carbon (CO_2) , asociate conversiilor simetrice *la* și *de la* terenurile forestiere de la si la alte folosințe, este necesara implementarea unor metode robuste care sa surprinda, pe de o parte, absorbtia de CO_2 extrem de lenta în cazul conversiilor de la alte folosințe la pădure (e.g. împăduriri) și pe de altă parte emisiile accelerate de (CO_2) aferente conversiilor de la pădure la alte folosințe (ex: despadurir).

Cea mai noua versiune a modelului Yasso, Yasso 15, care descrie ciclul C organic în sol (Järvenpää et al 2015), reprezintă o îmbunătățire a unei versiuni anterioare Yasso (Liski et al. 2005) precum și a Yasso07 (Tuomi și al. 2009, Tuomi et al. 2011b). Acesta in plus cuantificasi respirația heterotrofică a solului. Aplicațiile sale se extend la simularea dinamicii stocurilor de C din schimbarea folosinței terenului, gestionarea ecosistemelor, schimbările climatice. Sintaxa modelui Yasso15 este relativ simplă, datele de intrare necesită doar informații cu privire la cantitatea de C plus parametrii climatici (temperature si precipitatii).

Versiunea anterioară Yasso07 a fost bazată pe un număr substanțial mai mare de măsurători și metode matematice mai avansate decât versiunea Yasso precedenta, estimările privind incertitudinea reprezentând partea fixă a rezultatelor din Yasso07.

Versiunea curenta Yasso15 utilizează un set de date mai diversificate, punându-se mai mult accent pe ipotezele de modelare și unele detalii matematice care au condus la o calitate mai bună a modelarii, respectiv o mai bună reprezentare a metodelor și proceselor ecologice fundamentale. În plus, estimările de incertitudine sunt parte importantă a acestei versiuni,

facilitând și simulări ce pot implica transferuri ale carbonului organic între diferite tipuri de folosințe a terenului.

Definiții: În acest experiment s-au ales trei suprafete de proba (SP) care să reflecte secventa conversiei la pădure de la pajiște.

Design experimental: Tipul de pădure Gospodărită, conform planului amenajistic SP-urile se poziționează în U.A. 7A , a Ocolului Pădurile Șincii (vezi figura urmatoare cu locația suprafețelor de probă).

Distribuția altitudinală: trei suprafețe de monitorizare corespunzând altitudinii la care se găsește tipul de pădure cu compoziția, Fag și Carpen (cod FA, CA, vârstă 80 ani) 600-700 m, tranziția (amestec FA, CA, vârsta 21-21 ani), urmată de pajiste.

Recoltare probe sol și preprocesare: Pentru recoltarea probelor de sol din fiecare secvență s-a folosit o sonda tip Edelman și Riverside/ Eijelkamp (vezi figura), s-au efectuat câte 5 repetiții din 10 în 10 cm, din care s-au prelevat probe până la adâncimea de aproximativ 1m, locația fiecărei repetiții fiind înregistrată în GPS, totatul probelor fiind de 82.

Ulterior au fost aduse în laborator în pungi de plastic etichetate corespunzator, urmând a fi procesate pentru determinarea conținutului de C organic, azot N, determinarea texturii, și a densității aparente, rezultatele fiind necesare pentru parametrizarea modelului.

Caracteristicile suprafeței de probă culese sunt : Locația, Secvență, Specia, Tipul de sol, Textură, Densitate aparentă, C/N.

d) Analiza incertitudinii metodelor utilizate pentru detectarea schimbarii folosintei terenului prin metode diferite (M. Miclaus, V. Blujdea)

Acesta raspunde obligatiilor asociate sarcinilor din pachetul 3 si 5. Imbunatatirea consistentei metodelor de estimare a schimbarii stocurilor de carbon cu suprafata terenurilor este una din marile provocari legate de implemetaea inventarelor de gaze cu efect de sera si a reducerior de emisii asociate obligatiilor internationale (Protocolul de la Kyoto, Acordul de la Pari/legislatia uniunii europene). Activitatea face parte din cadrul WP5. Compararea a trei metode utilizate in diverse sisteme de raportare este descrisa in versiunea avansata de articol inclusa in **Anexa 3.**

e) Studiu privind specificitatea modelelor alometrice (I. Dutca, V. Blujdea)

Acesta raspunde obligatiilor asociate sarcinii 5.2. Este general recunoscut ca modelele alometrice necesare pentru estimarea biomasei in păduri sunt specifice zonei din care au fost eșantionați arborii. Asta pentru ca forma arborilor este influențată de genotip dar si de factorii de mediu cum ar fi solul, clima dar si competiția dintre arbori. Plecând de la premisa ca aceste caracteristici au o variabilitate spațială, concluzionam ca si alometria arborilor are o variabilitate spațială. Folosind modele ierarhice cu interceptul variabil, am putut arata cat de mult sunt afectate aceste modele de variabilitatea spațială. Coeficientul de corelație intraclasa este des folosit in sociologie pentru a arata proporția variantei modelului, cauzata de diferențele dintre grupuri. In mod similar, noi am arătat ca diferențele dintre plantațiile tinere de molid (Picea abies) in Romania produc proporții foarte mari din varianta totala a modelului alometric. Aceasta proporție a variat intre 33 si 86% din varianta totala a modelului, in funcție de variabila independenta folosita si componenta arborilor estimata. Am mai arătat ca, folosind diametrul ca variabila independenta in model efectele produse de gruparea arborilor eșantionați in plantații este mai mic decât atunci când folosim înălțimea arborilor. Atunci când sunt folosite ambele variabile (diametrul si înălțimea) este mai bine sa fie folosita o variabila combinata (D2H) deoarece efectul plantației asupra modelului este mai mic. Dintre componentele arborilor, biomasa fusului are o specificitate mai mare decât biomasa frunzelor sau ramurilor.

Rezultatele obținute sunt foarte importante pentru domeniul estimării carbonului in păduri, pentru ca in acest fel se poate decide daca modelele elaborate pentru un arboret pot fi folosite si in alte arborete. Deși se vorbește foarte des despre specificitatea modelelor alometrice, acest studiu este primul studiu care arata intr-un mod cantitativ ca specificitatea modelelor alometrice are foarte ridicata. Acest studiu a fost publicat in revista Biomass & Bioenergy nr. 116 din Septembrie 2018. Varianta publicata a articolului este disponibila la: https://www.sciencedirect.com/science/article/pii/S0961953418301259?via%3Dihub or https://doi.org/10.1016/j.biombioe.2018.05.013.

f) Dezvoltarea unui indicator pentru justificarea deciziei de a combina variabilele in modelele alometrice (I. Dutca, V. Blujdea)

Modele alometrice folosesc variabile ușor de măsurat (e.g. diametrul D si/sau înălțimea H) pentru a estima caracteristici greu de măsurat ale arborilor (e.g. biomasa). Însă diametrul si înălțimea arborilor sunt caracteristici care sunt corelate, deoarece arborii cu diametrul mai mare au de regula si o înălțime mai mare. Pentru a limita efectele coliniarității dintre variabile adesea se folosește o variabila combinata D^2H, plecând de la premisa ca biomasa supraterana este proporțională cu volumul unui cilindru cu diametrul=D si înălțimea=H. Însă variabila combinata constrânge modelul sa producă un raport fix al parametrilor pentru D si H, respectiv 2.0. Așadar, ipoteza studiului este ca pierderea de acuratețe a modelului este in funcție de raportul Q (raportul dintre parametrul lui D si parametrul lui H). Cu cat raportul Q este mai diferit de 2.0 cu atât pierderea de acuratețe atunci când folosesc D^2H este mai mare. Folosind cinci seturi de date cu observații de biomasa am confirmat ipoteza studiului.

Studiul a fost publicat in revista Forestry (online first) https://doi.org/10.1093/forestry/cpz041

g) Studiu privind efectul covarianței neliniare (intre parametrii modelelor alometrice) si a combinării variabilelor independente in modelele alometrice asupra estimărilor de biomasa la nivel de suprafață întinsă (I. Dutca, V. Blujdea).

Datorita faptului ca modele alometrice sunt modele neliniare, atunci când modele neliniare sunt folosite in locul celor liniare aplicate datelor logaritmate, covariantele dintre parametrii modelelor alometrice sunt curbate (nu sunt liniare). Covariantele modelelor alometrice sunt necesare pentru a putea propaga incertitudinea din modelele alometrice in incertitudinea estimărilor de biomasa la nivel de suprafață întinsă (e.g. in suprafețe de proba gen IFN). Uzual, pentru estimarea covarianței parametrilor modelelor alometrice sunt folosite aproximări bazate pe serii Taylor, care însă omite neliniaritatea dintre parametri. Folosind un set de date de biomasa pentru specia molid, precum si suprafețele de proba din IFN (Romania) unde toți arborii sunt din specia molid (243 suprafețe) am aplicat o metoda Monte-Carlo de propagare a erorilor din modelele alometrice dar si din diferențele dintre suprafețe, in estimarea biomasei la nivel de suprafață întinsă (i.e. tone/ha). In cadrul acestei metode, covariantele parametrilor au fost estimate folosind doua metode: serii Taylor si Bootstrap aplicat valorilor reziduale.

Pe de alta parte, combinarea variabilelor in modelele alomterice, deși am arătat ca pot produce o pierdere a acurateței estimărilor, aceasta pierdere depinde de raportul Q (Q = raportul parametrilor pentru D si H intr-un model cu variabile separate). Deci daca este folosita o variabila combinata D^QH in loc de D^2H, acesta pierdere de acuratețe nu mai exista. Însă combinarea variabilelor poate avea efecte benefice, deoarece numărul covariantelor de estimat se reduce de la 3 (in cazul modelelor bazate de doua variabile) la 1 (pentru modelul combinat).

O varianta de lucru a acestui articol este prezentata in **Anexa 4**.

h) Studiu privind impactul caracteristicelor eșantionului in modelele alometrice asupra acurateței si preciziei estimărilor de biomasa (I. Dutca)

Este bine știut ca acuratețea si precizia estimărilor de biomasa depind într-o oarecare măsură de modelele alometrice pentru estimarea biomasei la nivel de individ. Deși este cunoscut faptul ca variabilitatea intrinseca a relației intre biomasa si predictor(i) precum si mărimea eșantionului (numărul de observații) influențează acuratețea si precizia (acuratetea a fost definita ca diferenta dintre valoarea estimata si valoarea reala), acest studiu aduce in discuție si alte caracteristici ale eșantionului cum ar fi mărimea intervalului de diametre al eșantionului, poziția acestui interval (data de valoarea de start a intervalului) si distribuția diametrelor in eșantion. Folosind simulări Monte-Carlo am generat seturi de date cu diferite caracteristici. Mai departe, cu aceste date am elaborat modele alometrice care au fost folosite pentru a estima biomasa unei suprafețe de proba. Concluziile studiului au fost:

- Variabilitatea relației Biomasa-Diametru a fost cel mai important factor care influențează acuratețea si precizia estimărilor de biomasa;
- Mărimea eșantionului (numărul de observații) deși a influențat semnificativ acuratețea estimării a avut un efect nesemnificativ asupra preciziei estimării;
- Distribuția diametrelor in eșantion a avut un efect similar mărimii eșantionului; a influențat semnificativ acuratețea estimării, însă nesemnificativ precizia estimării;
- Am demonstrat ca arborii mici aduc o cantitate mai mare de informație in modelul alometric, deci modelele alometrice care includ arbori mici vor avea o ajustare mai buna (o valoare a coeficientului de determinare R^2 mai mare) si valori mai mici ale erorilor standard ale parametrilor. Acest lucru este datorita faptului ca varianta in modele alometrice (care sunt neliniare) nu este constanta si creste cu diameterul (=heteroscedasticitate). Insa in modele alometrice heteroscedasticitatea este controlata prin ajustarea observațiilor cu un factor care se calculează ca inversa variantei. Cum varianta este mica la arborii mici, acest factor este mai mare la arborii mici, deci cantitatea de informație (sau importanta) arborilor mici in model este mai mare. Cu toate acestea, deși modelul in care sunt incluși arborii mici este mai bun (din punct de vedere al coeficientului de determinare), impactul asupra acurateței si preciziei estimărilor de biomasa este nesemnificativ.
- Am arătat ca eșantionând un număr constant de arbori pentru fiecare categorie de diametre rezulta modele care produc estimări cu acuratețe si precizie mai ridicata.
- De asemenea, deși R^2 este frecvent folosit pentru alegerea modelelor (sub ipoteza ca un model cu R^2 mai mare este mai bun) am arătat ca R^2 este dependent de mărimea intervalului de diametre folosit pentru elaborarea modelului iar acuratețea si precizia modelelor nu depind de intervalul de diametre folosit. Așadar, am arătat ca precizia si acuratețea estimărilor de biomasa nu depind de valoarea R^2 a modelului.

Manuscrisul este prezentat in **Anexa 5**.

i) Studiu asupra variabilității generate de diferențele dintre specii si de diferențele dintre locații in modelele alometrice (I. Dutca)

Modele alometrice sunt vitale pentru estimările de biomasă si pentru buna implementare a programelor de reduceri de emisii din păduri. Aceste modele folosesc variabile independente ușor de măsurat (e.g. diametrul de baza, înălțimea arborilor) pentru a estima biomasa arborilor in picioare. Însă ele au doua mari limitări: (1) faptul ca modele sunt specifice speciei si locației si (2) faptul ca măsurătorile de biomasă pentru elaborarea de noi modele sunt dificile si implica logistica si costuri mari. Cunoașterea nivelului de specificitate in raport cu specia si in raport cu locația a acestor modele nu este bine cunoscuta. De aceea scopul acestui studiu a fost de a arata gradul de specificitate al modelelor in raport cu specia si locația, folosind doua seturi de date din Eurasia si Canada. Aplicând un model ierarhic ANOVA valorilor reziduale ale modelelor alometrice, am separat varianta totala in (i) varianta explicata de diferențele dintre specii, (ii) varianta explicata de diferențele dintre locații si (ii) varianta reziduala. Mai departe am folosit proporția variantei explicata de fiecare din cele doua nivele (specie si locație) pentru a evalua cat de specifice sunt modele alometrice speciei respectiv locației. Pentru determinarea erorilor standard ale acestor proporții am aplicat o analiza Bootstrap. Rezultatele au arătat ca specia explica o proporție a variantei totale mult mai mare decât locația. Proporția variantei explicate de diferențele dintre specii a fost de 42.56% (SE = 6.10%) pentru Eurasia si 47.54% (SE=6.07%) pentru Canada, pe când proporția explicata de diferențele dintre locații a fost de 20.08% (SE=3.35%) pentru Eurasia si 8.27% (SE=1.38%) pentru Canada. Așadar diferențele dintre specii generează o variabilitate mult mai mare in modele alometrice in comparație cu diferențele dintre locații. Folosind diametrul si înălțimea arborilor ca variabile independente in modelul apometric (comparativ cu situația in care doar diametrul este folosit ca variabila independenta), a condus la o scădere a proporției variantei explicata de diferențele dintre locații de cca 24-44%, pe când proporția variantei explicata de diferențele dintre specii a rămas neschimbata.

Aceste informații sunt extrem de valoroase deoarece ele arata cat de mari pot fi diferențele dintre modelele alometrice ale diferitelor specii, precum si cat de mari pot fi diferențele intre modelele alometrice specifice locațiilor. De asemenea, mai indica riscul de erori sistematice atunci când modele specifice unei specii este folosit pentru o alta specie si cat de mari sunt riscurile atunci când un model dezvoltat pentru o locație este folosit într-o alta locație. Deoarece proporția variantei explicata de diferențele dintre specii a fost mai mare decât cea explicata de diferențele dintre locații, riscul de erori sistematice este mai mare când modele sunt mutate de la o specie la alta decât atunci când ele sunt mutate de la o locație la alta. De asemenea, am arătat ca, deoarece variația condițiilor climatice este mai mare in setul de date din Eurasia, proporția variantei explicata de diferențele dintre locații este mai mare decât pentru setul de date din Canada.

Studiul a fost publicat in revista Forests 2019, 10(11), 976;
https://doi.org/10.3390/f10110976

j) Calibrarea modelului PREBAS pentru Romania (I. Dutca, V. Blujdea)

Modelul PREBAS este un model care simulează dinamica pădurii la nivel de arboret (sau strat din arboret) si a luat naștere prin combinarea modelelor CROBAS si PRELES. CROBAS este un model pentru estimarea creșterii individuale a arborilor. Creșterea se

bazează pe acumularea si alocarea carbonului, așadar creșterea este egala cu producția neta. PRELES este un model folosit pentru estimarea capacitații de fotosinteza a unei păduri, input care este esențial in CROBAS. Fotosinteza bruta este calculata ca produs intre masa frunzelor si rata specifica a fotosintezei.

Pentru calibrarea modelului PREBAS am folosit datele IFN referitoare la caracteristicile arboirlor masurati, dar si o serie de date climatice specifice fiecarei suprafete de proba IFN.

```
nYears = 100
siteInfo <- read.csv("inputs/siteInfo.csv",header = T)
thinning <- read.csv("inputs/Thinning.csv",header = T)
initVar <- read.csv("inputs/initVar.csv",header = T, row.names = 1)
obsData <- read.csv("inputs/obsData.csv",header = T)

weather <- read.csv("inputs/weather.csv",header = T)
PAR = c(weather$PAR,weather$PAR,weather$PAR)
TAir = c(weather$TAir,weather$TAir,weather$TAir)
Precip = c(weather$Precip,weather$Precip,weather$Precip)
VPD = c(weather$VPD,weather$VPD,weather$VPD)
CO2 = c(weather$CO2,weather$CO2,weather$CO2)
DOY = c(weather$DOY,weather$DOY,weather$DOY)

PREBASout <- prebas(
                    nYears=nYears,
                    pCROBAS = pCROB,
                    pPRELES = pPREL,
                    siteInfo = siteInfo,
                    thinning = thinning,
                    PAR=PAR,TAir=TAir,VPD=VPD,Precip=Precip,CO2=CO2,
                    P0=NA,
                    initVar = as.matrix(initVar),
                    defaultThin = 0.,
                    ClCut = 1.,
                    inDclct = NA,
                    inAclct = NA)
```

Figura 1. Un exemplu din scriptul R al modelului PREBAS, cu funcția „prebas”.

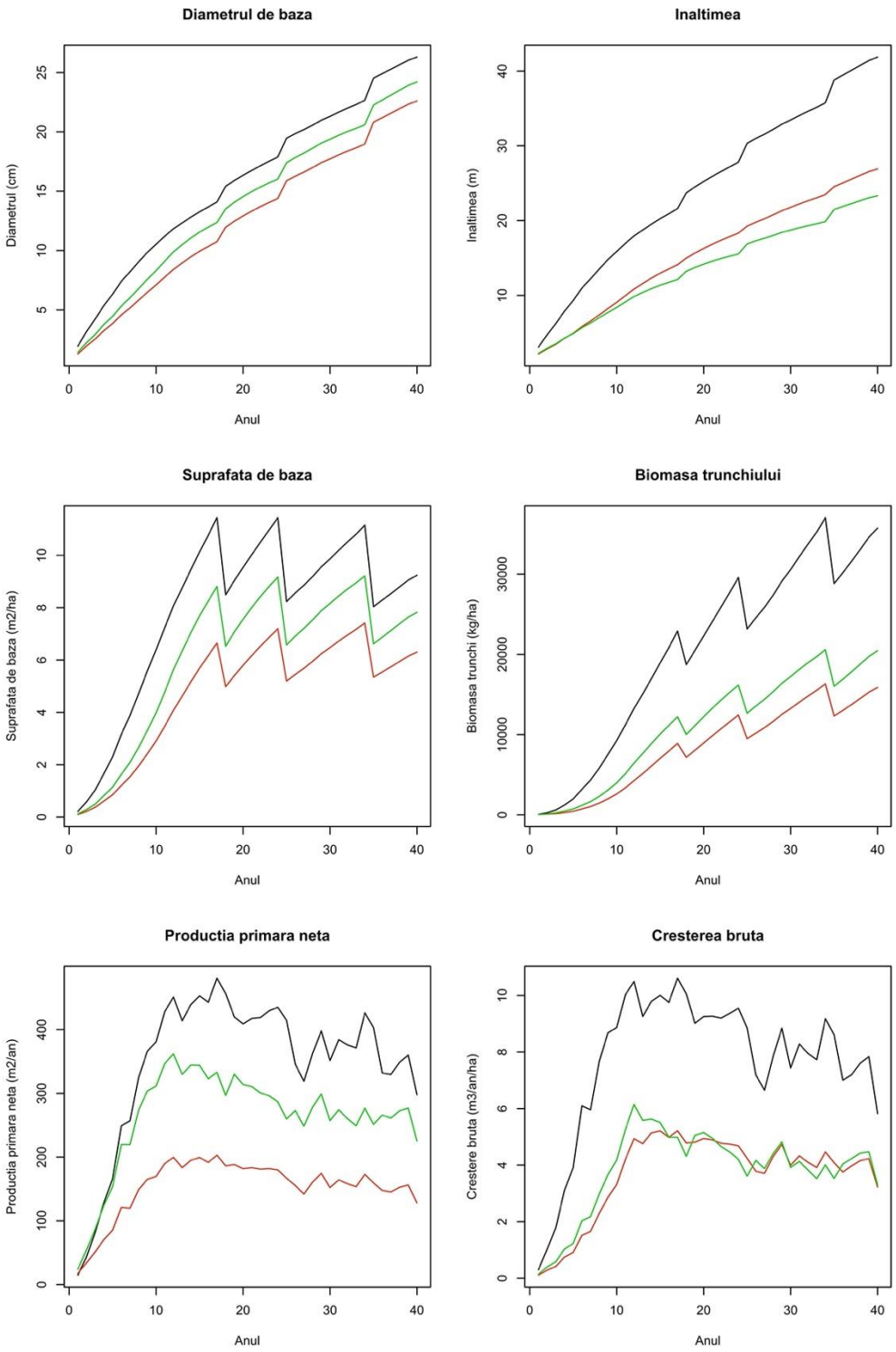


Figura 2. Un exemplu de rezultat obținut pentru o perioada de simulare de 40 de ani

Mai multe detalii despre calibrarea modelului PREBAS, in **Anexa 6.**

k) Administrarea bazei de date a proiectului

- procesarea statistică se face cu prioritate în R (open source): https://cran.r-project.org/bin/windows/base/
- modul de stocare și actualizare a bazei de date: fișiere Excel pentru EFISCEN si Microsoft Access pentru CBM-CFS.

l) Sprijin activități incluse in alte pachete de lucru din FORCLIMIT

- informare continua cu privire la regulile de contabilizare a reducerilor de emisii din sectorul folosinței terenurilor incluse în Pachetul energie clima 2030 (https://ec.europa.eu/clima/policies/strategies/2030_en), in sprijinul Pachetelor de lucru 1 si 2 ale FORCLIMIT.

3. Managementul proiectului

Toate sarcinile asumate prin contract sunt antamate si in stadiu corespunzator ultimei jumatati a perioadei de implementare a proiectului. Cele mai dificile aspecte, respectiv adaptarea bazelor de date si definitivarea scripturilor pentru pre-procesare si armonizare date sunt rezolvate pentru EFISCEN (varianta clasica) si CBM-CFS3, si în ce privește integrarea CBM-CFS3 cu Yasso15. Calibrarea modelul Yasso15 pentru Romania este incetinita de ritmul experimentului de descompunere (Anexa 1) si de analiza bio-chimica (de Institutul Meteorologic partener din Finlanda), insa rezultatele sunt conforme cu teoria, si reprezinta primele rezulate de acest fel din Europa de est.

Sunt organizate intalniri periodice pentru o zi de lucru in comun, cel putin odata la 3 saptamani. Calendarul de colectare probe de sol (pentru validarea modelului) si descompunere litiera este mentinut cu strictete.

4. Vizibilitate nationala si internațională a proiectului FORCLIMIT

Referitor la **comunicarea excelentă avută cu partenerii europeni implicați în proiect** amintim

- includerea lui V. Blujdea ca expert Type A (independent, din partea com,unitatii stiintifice) in LULUCF Expert Group of the European Comision (https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3638&NewSearch=1&NewSearch=1).

Brașov, 4.12.2019 **Dr. ing. Viorel Blujdea**

Anexa 1. Decomposition of needle/leaf and small wood litter from European beech, Norway spruce and Silver fir: influence of mixture, climate (temperature x altitude) and forest management

1.Introduction

Litter decomposition is a fundamental process of forest ecosystems for the carbon and nutrients cycles (dead organic matter is transfered from the above-ground part of trees to the forest floor, where under the action of microorganisms and soil fauna is decomposed gradually depending on climate factors (temperature, precipitation) (Gholz et al. 2000), substrate availability/soil properties (Vesterdal 1999) and litter quality (Cornwell et al. 2008). In a meta-analysis, grouping data for 818 species from 66 decomposition experiments on six continents, Cornwell et al. (2008) found that plant functional traits as litter quality is more important than climate factors affecting litter decomposition rate (the species driven differences control predominantly the litter decomposition rate worldwidely).

2.Material and methods

Site

The study site is located in Transilvanian side of Southern Carpathians (Fagaras Mountains), Padurile Sincii forest district.

The study was carried out on European beech (Fagus sylvatica) leaves, and Norway spruce (Picea abies) and Silver fir (Abies alba) needle litter, but also on small wood litter of all three species. In October 2017, fresh fallen brown leaves were collected beneath several randomly distributed trees, while the needles were collected from branches cut from several trees selected at random. In November 2017, small wood pieces were collected from branches cut from trees recently harvested during the thinning interventions. Both litter types (leaves/needles and small wood) were stored 2 weeks in laboratory at air temperature. The litterbags were made of ??nets (20 x 30 cm; 1 mm mesh size? for beech and 10 x 10 cm; ..mm mesh size for needles) and filled with 10-20 g leaves and 5-15 g needles, respectively and labelled properly. In 6th of November 2017, on each of the four study sites 80 litterbags per species (8 replicates of 10 samplings campaigns over three years) were placed on the soil. Subsequent samples were taken according to a preplanned schedule (every month starting with 24th of April till 24 October). We dried a first set of bags at 80 C for five days and weighed and calculated for each species an average correction factor as ratio between oven dry mass and air-dry mass. We applied this correction factor to all litterbags in order to obtain the initial oven-dry mass of each leaf amount of every litterbag (we multiplied air-dry mass of all leaf bags for humidity by the average correction factor).

Data analysis

The relationship of the mass loss of leaves and needles and decomposition time is often modelled by a negative exponential decay model:

$$Mt=M0 \times \exp(-k \times t),$$

where:
-Mt is the mass at time t,
-M0 is the initial mass (mass at time 0),
-t is time in months
-k is the exponential decay coefficient or decomposition rate.

In our case, we used mass remaining as % from initial mass account (consequently, M0 = 100).

3.Results

3.1. Leaf/needle litter decomposition

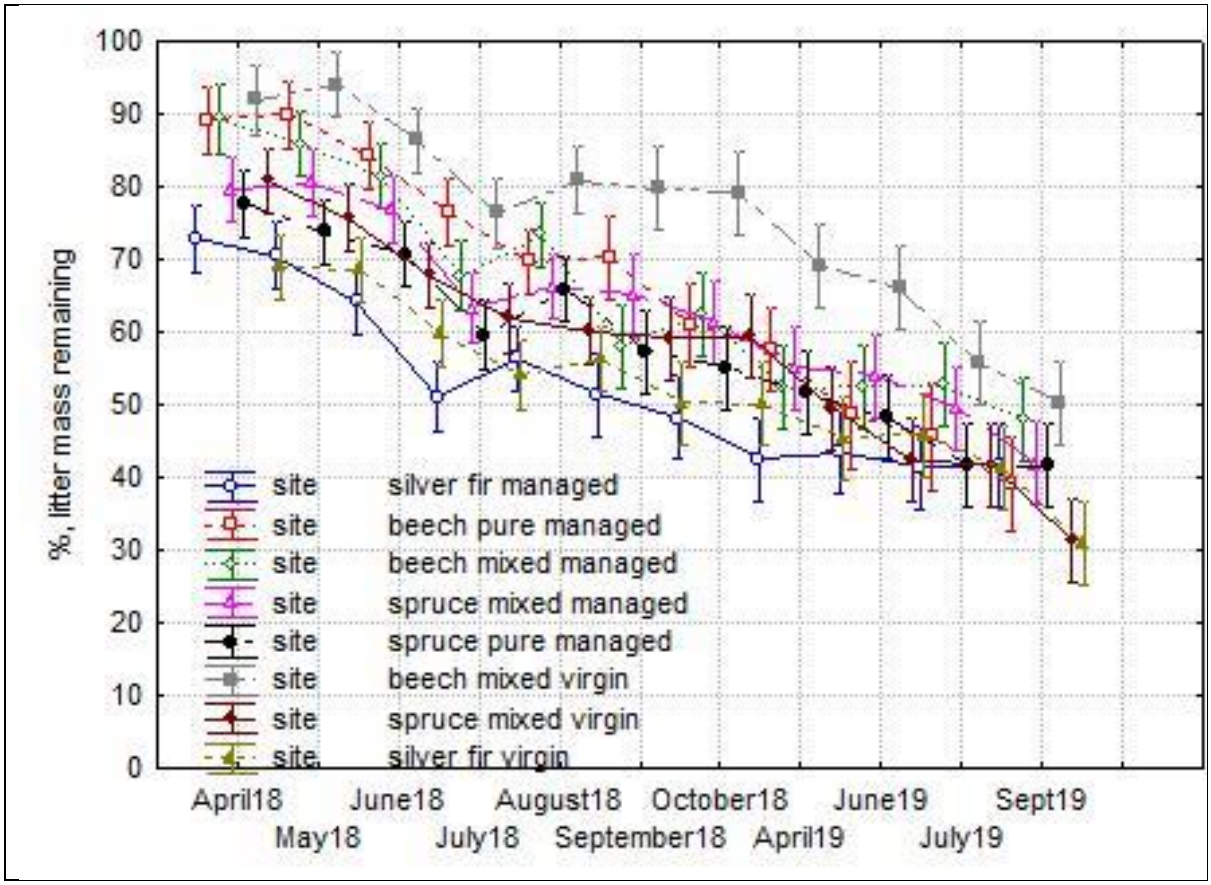
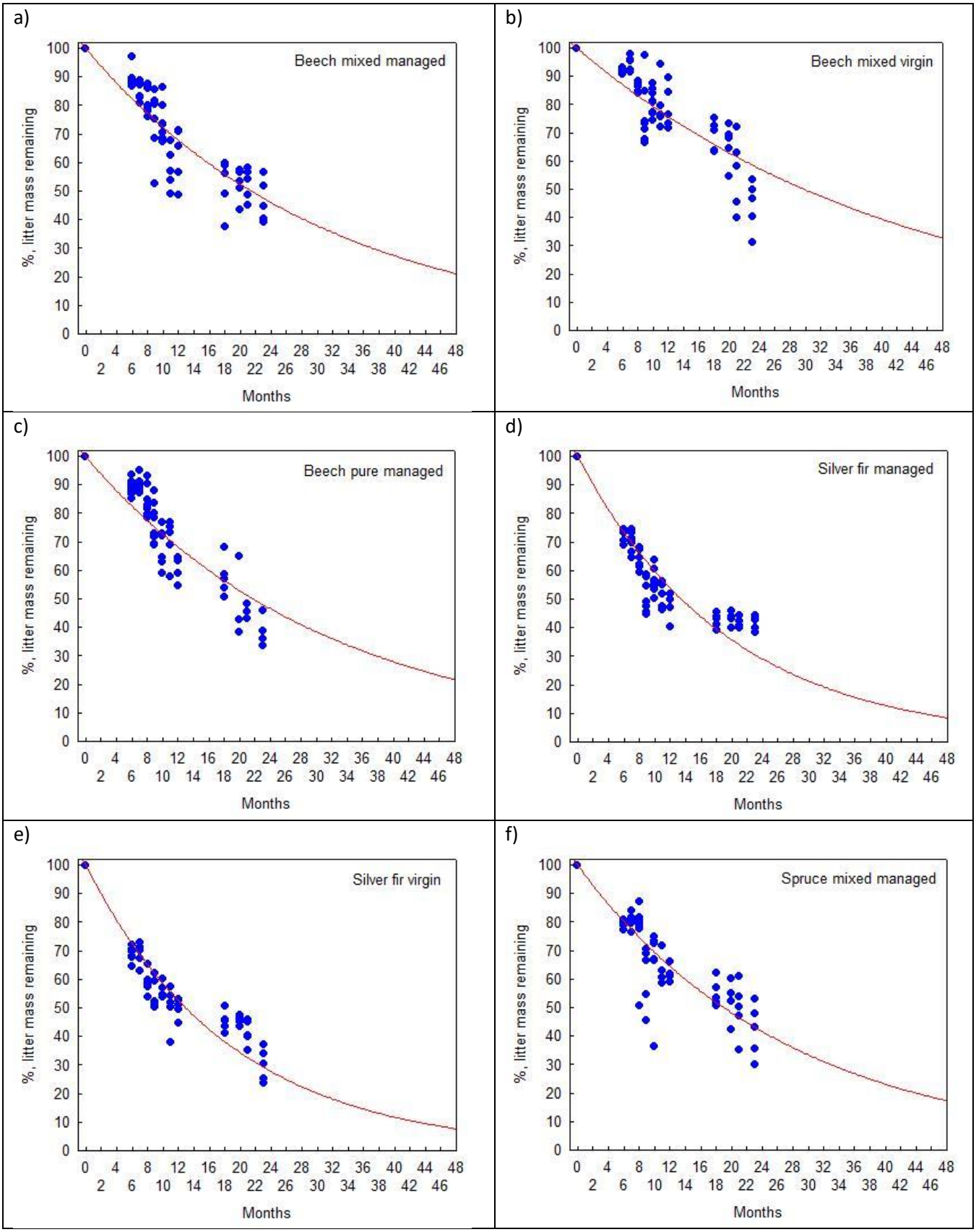


Fig.1. Leaf/needle mass remaining (as % from initial amount) for all investigated variants after each bags collection.

During the first year of experiment the mass loss


a)
Beech mixed managed
%, litter mass remaining
Months
b)
Beech mixed virgin
%, litter mass remaining
Months
c)
Beech pure managed
%, litter mass remaining
Months
d)
Silver fir managed
%, litter mass remaining
Months
e)
Silver fir virgin
%, litter mass remaining
Months
f)
Spruce mixed managed
%, litter mass remaining
Months

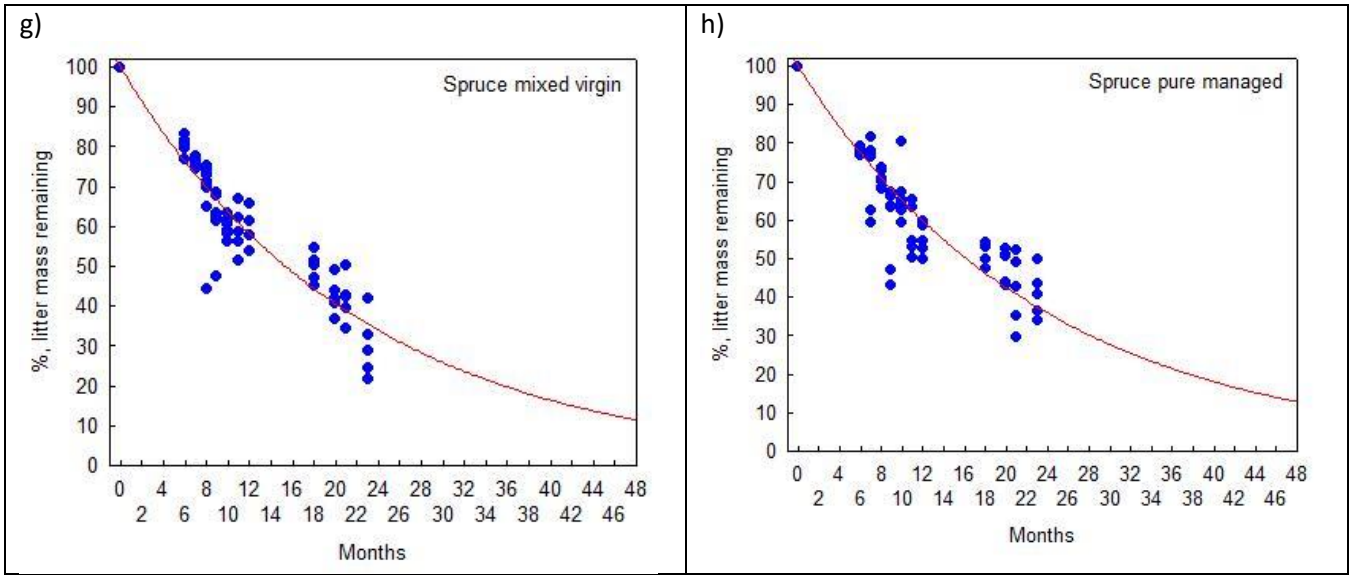
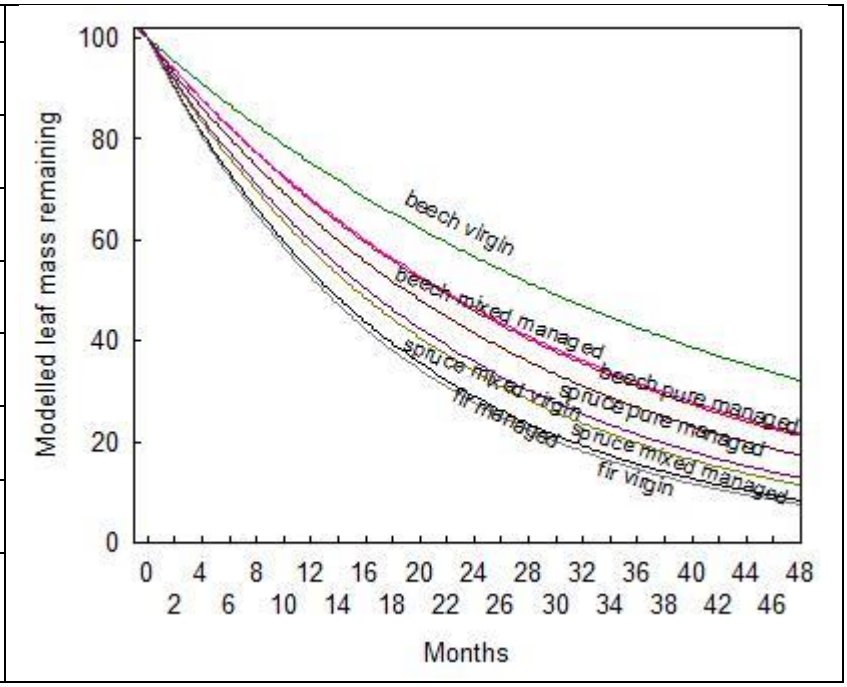


Fig.2. Litter mass remaining (as % from initial amount) modelled as a function of decomposition time (months) (y=100 x exp(-k x)) for each studied variant (a-beech mixed managed, b-beech mixed virgin, c-silver fir mixed managed, d-silver fir mixed virgin, e-spruce mixed managed, f-spruce mixed virgin, g-beech pure managed, h-spruce pure managed).

Table 1. Regression analysis (%mass remaining=100 x exp(-k x time)).

	K	p_k	R^2
Beech mixed managed	0,0324	0,000	0,74
Beech mixed virgin	0,0237	0,000	0,68
Silver fir mixed managed	0,0516	0,000	0,71
Silver fir mixed virgin	0,0535	0,000	0,72
Norway spruce mixed managed	0,0366	0,000	0,67
Norway spruce mixed virgin	0,0459	0,000	0,85
Norway spruce pure managed	0,0428	0,000	0,74
Beech pure managed	0,0237	0,000	0,77



3.2. Small wood litter decomposition

a)	b)

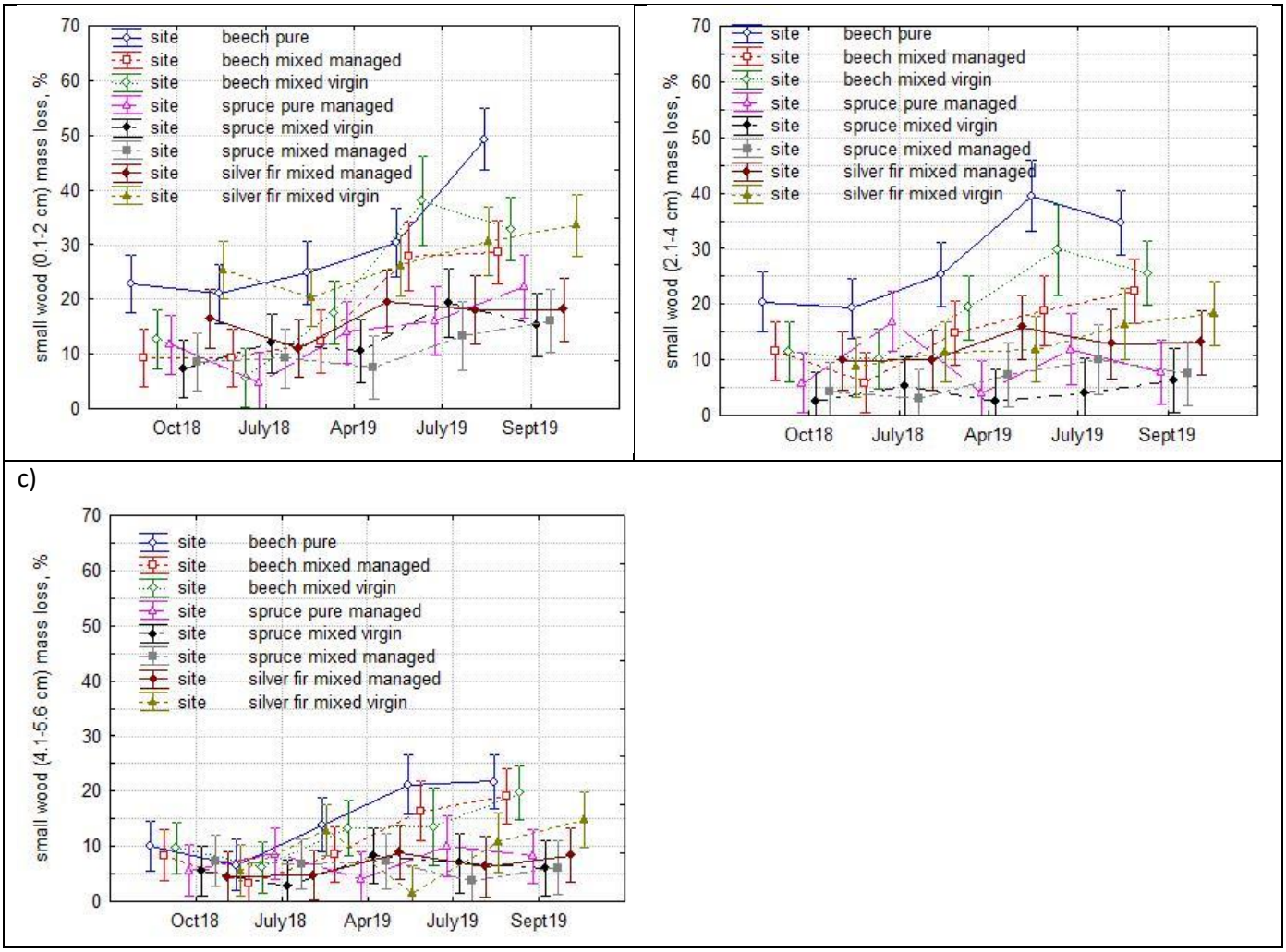
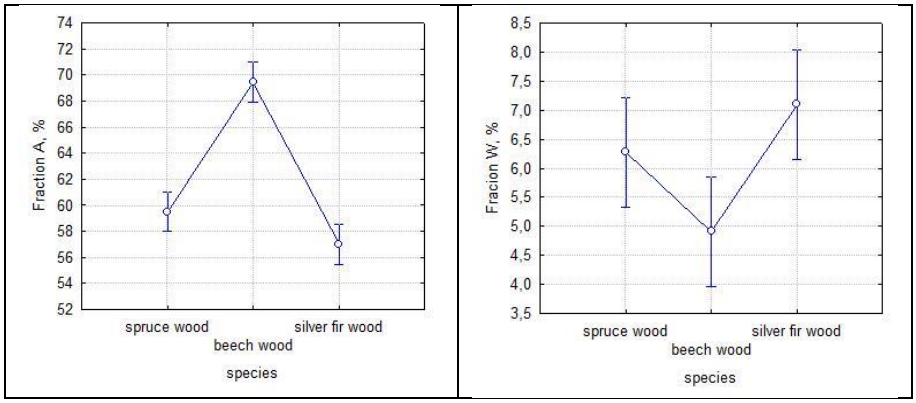


Fig.3. Small wood litter mass loss (as % from initial amount) for all forest/species variants (a) wood with d=0.1-2 cm, b) d=2.1-4.0 cm, c) d=4.1-5.6 cm

3.3. Species-specific preliminary AWEN values (intially time)

3.3.1. AWEN values for small wood



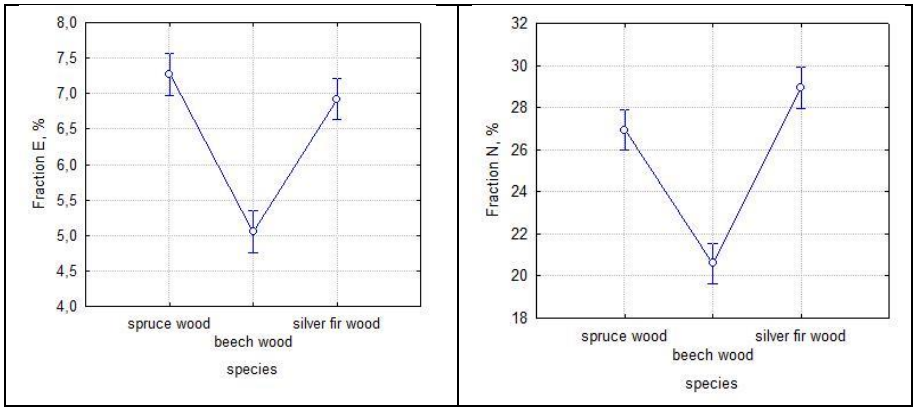


Fig. 4. Species specific variation of AWEN values of small wood.

3.3.2. AWEN values for leaves/needles

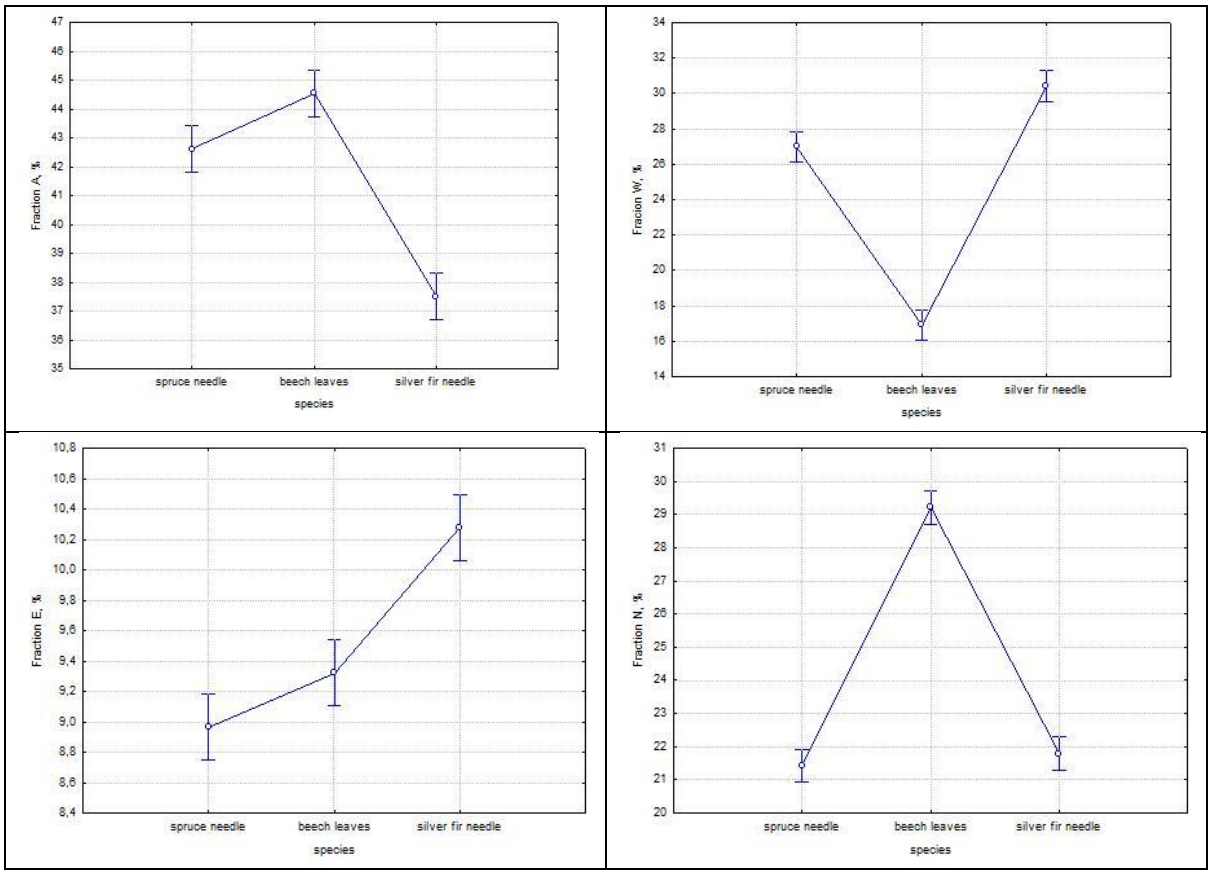


Fig. 5. Species specific variation of AWEN values of leaves/needles

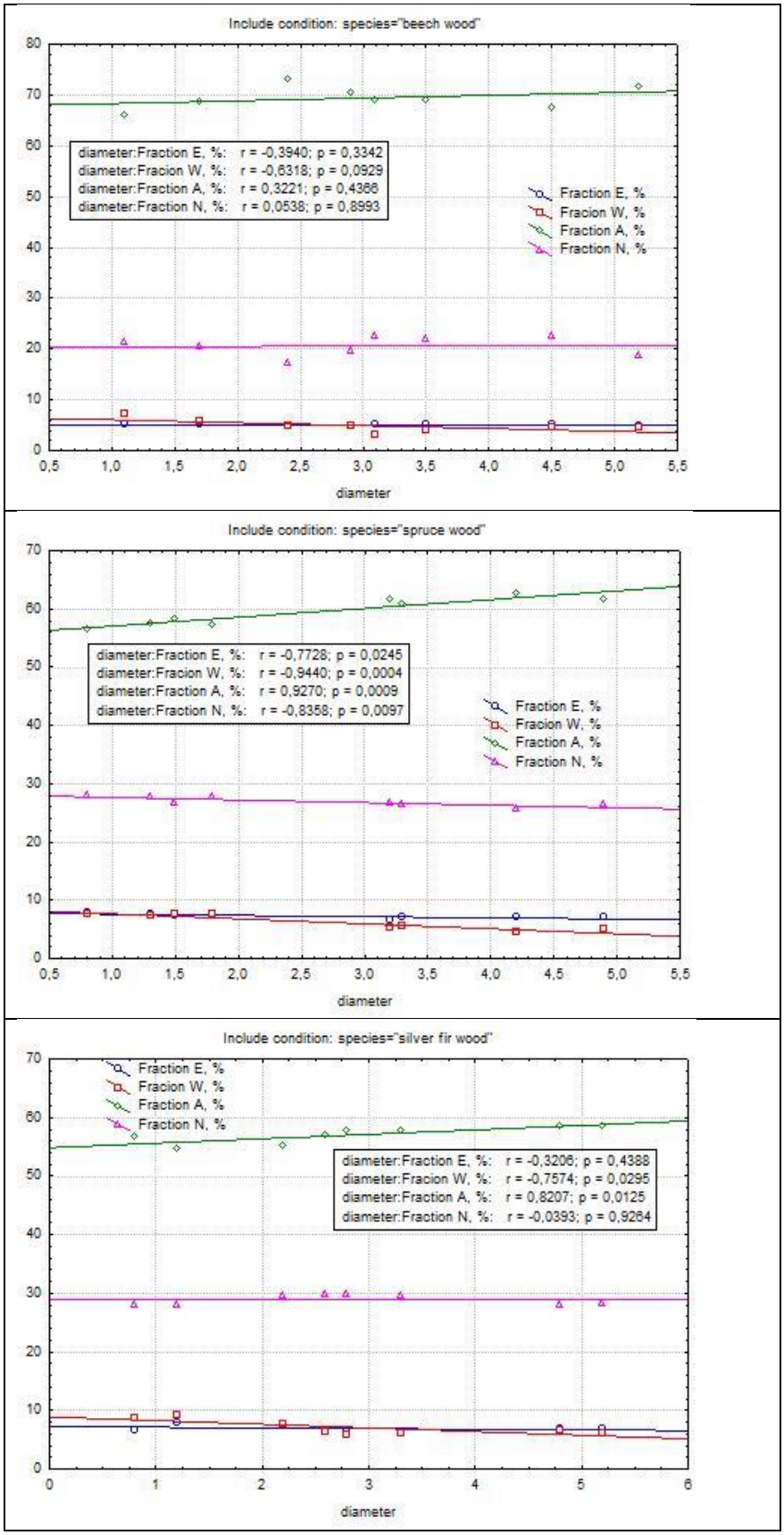


Fig. 6. Relations between AWEN values of small wood and diameter

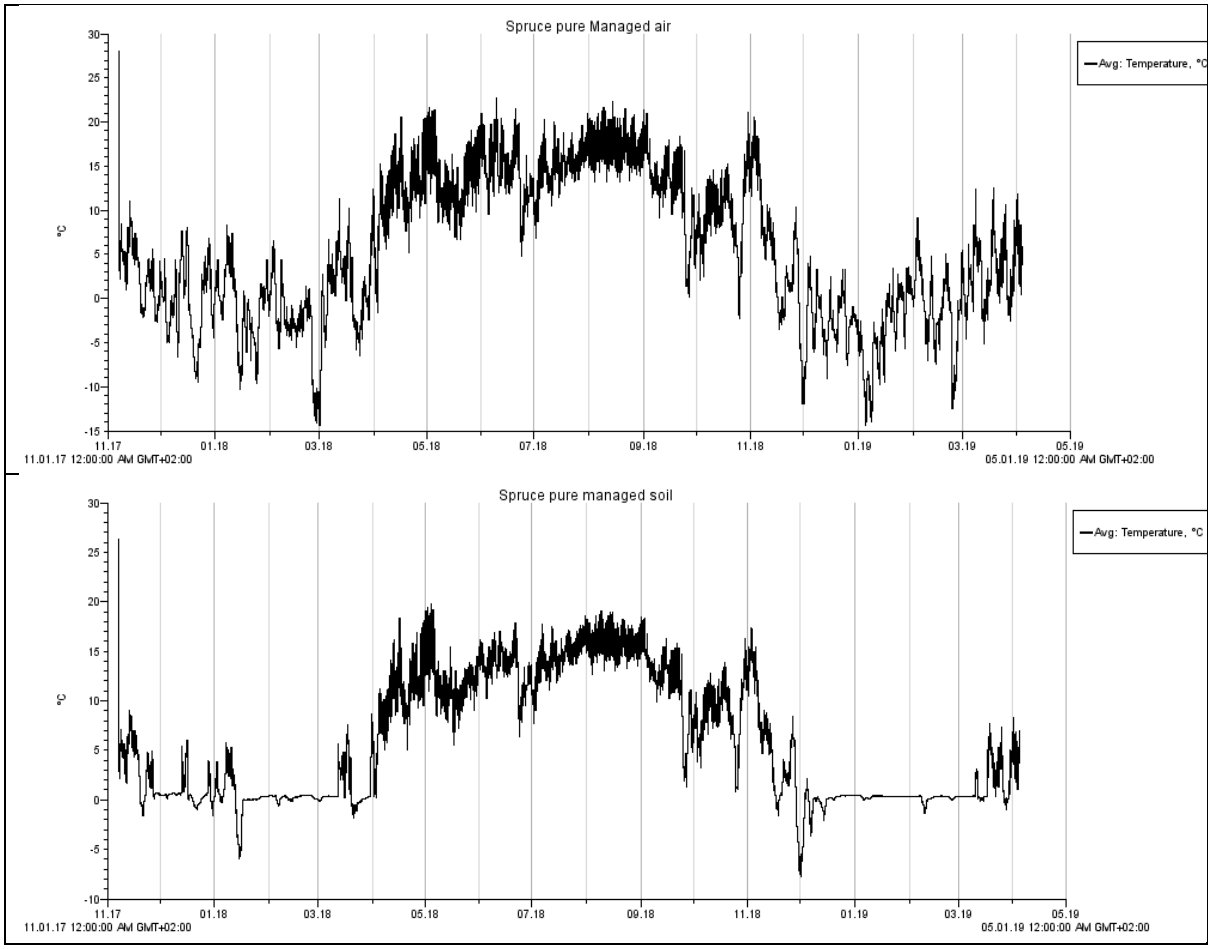


Fig. 7. Variation of air and soil temperature in sampled sites.

4. **Discussion.**
5. **References**

Anexa 2. Informatii privind armonizarea bazelor de date in vederea validarii reciproce a medelelor CBM-CFS si EFISCEN

Anexa 2a. Criteriile de clasificare si parametrii agregați regional pentru baza de date națională din Inventarul Forestier National

Criterii	Specificatii
Tip de padure/ specii	Rasinoase (OC), **Molid (PA), Brad (AA), Predom rasinoase (PredCon), Amestecuri (ConBroad), Predom foioase (PredCon),** Foioase (OB), Fag (FS), **Cvercinee (QR), Salcam(RP)** – pentru tipurile de padure ingrosate parametrii modelului sunt actualizati prin ajustare la nivel de regional (clima si regiune)
Clase de varsta	1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-110, 111-120, 121-130, 131-140, 141-150, 151-160, >160, Unevenaged
Regiuni administrative (NUTS-2)	RO11, RO12, RO21, RO22, RO31, RO32, RO41, RO42
Volum pe picior	Volume annual, m3 y-1
Recolta de masa lemnoasa	Volume annual, m3 y-1
Suprafata	Area, ha
Creserea neta anuala	Net annual growth, m3 y-1 ha-1
Eroarea de eșantionare (in %) pentru toți parametrii de mai sus	Estimation error, %

Parametrii ecuatiilor utilizati la modelare

$$V=a*e^{(-b*A)}*(1-e^{(-b*A)})^{(c-1)}$$, *unde*

V- volumul comercial,
A – clasa de varsta de 10 ani,
a,b,c – parametrii ecuatiei specifici ficarei tip de padure

Parametrii ecuatiei pentru estimarea volumului lemnului comercial pe picior

Tip de padure	ConBroad	AA	FS	OB	OC	PA	PredBroad	PredCon	QR	RP
a	2291.41	136381.7553	2019.821	976.8087	3787.497176	2777.876	3696.275	2841.894	1607.577	3541.647
b	0.009851	3.81253E-05	0.005134	0.006911	0.015951353	0.016171	0.01238	0.008661	0.011314	0.002407

c	2.598057	1.949198118	2.137377	2.012281	4.180130563	3.50011	3.635651	2.8985 9	2.956918	2.413442

Parameteii ecuatiei pentru estimarea cresterii curente cumulate a volumului lemnului comercial pe picior

Tip de padure	ConBroad	AA	FS	OB	OC	PA	PredBroad	PredCon	QR	RP
a	46.67395443	30.53049718	44.82908538	12.60159597	44.91925629	32.29905709	16.71558839	25.99785093	18.19606152	32.28165566
b	0.014718484	0.003007487	3.28696E-05	0.003763308	0.018643759	0.010442337	0.00294835	0.005746935	0.010859768	0.044339613
c	2.33569566	1.542279681	1.349733947	1.264787544	2.574587006	2.109134766	1.388390928	1.474466432	1.659962736	2.806735827

Parametrii ecutiei Boudewyn privind modelarea alocarii de biomasa in compartimetele arborelui functie de volumul lemnului comercial. *P* reprezinta proportia componentei de biomasa din biomasa supraterana integrala (potrivit Boudewyn, P., Song, X., Magnussen, S., Gillis, M.D., 2007. Model-based, Volume-to-Biomass Conversion for Forested and Vegetated Land in Canada. Canadian Forest Service, Victoria, Canada (Inf. Rep. BC-X-411).).

$$(4)\quad p_{stemwood} = \frac{1}{1 + e^{a1+a2\times vol+a3\times lvol} + e^{b1+b2\times vol+b3\times lvol} + e^{c1+c2\times vol+c3\times lvol}}$$

$$(5)\quad p_{bark} = \frac{e^{a1+a2\times vol+a3\times lvol}}{1 + e^{a1+a2\times vol+a3\times lvol} + e^{b1+b2\times vol+b3\times lvol} + e^{c1+c2\times vol+c3\times lvol}}$$

$$(6)\quad p_{branches} = \frac{e^{b1+b2\times vol+b3\times lvol}}{1 + e^{a1+a2\times vol+a3\times lvol} + e^{b1+b2\times vol+b3\times lvol} + e^{c1+c2\times vol+c3\times lvol}}$$

$$(7)\quad p_{foliage} = \frac{e^{c1+c2\times vol+c3\times lvol}}{1 + e^{a1+a2\times vol+a3\times lvol} + e^{b1+b2\times vol+b3\times lvol} + e^{c1+c2\times vol+c3\times lvol}}$$

Valorile parametrilor pentru cele zece tipuri de padure

Tip de padure	a1	a2	a3	b1	b2	b3	c1	c2	c3
ROU_PC	-1.573653143	-0.001653423	0.043681989	-1.917251538	-0.001318462	0.067893453	-0.753406708	0.005322017	-0.854548877
ROU_CB	-1.688343	0.001696	-0.255443	-2.022535	-0.001800	0.128927	-0.722283	0.005140	-1.059489
ROU_AA	-1.426523	-0.000687	-0.083774	-1.822640	-0.000141	-0.056877	-0.522418	-0.000518	-0.500000
ROU_OC	-1.195958	-0.000340	0.044504	-1.588882	-0.002690	-0.172668	-0.888850	-0.004805	-0.407255
ROU_PA	-1.573125306	-0.000498028	-0.022566376	-1.926269813	-0.00016829	-0.011293606	-0.870537754	-0.002046936	-0.443987026

ROU_FS	-1.675509	0.000425	-0.153451	-1.988408	-0.001124	0.070280	-0.796988	0.005713	-1.132685
ROU_PB	-1.716351128	0.000573495	-0.139975714	-2.052043708	-0.001049959	0.055252471	-0.951411 23	0.003589983	-0.968666404
ROU_OB	-1.677640	0.000431	-0.104280	-1.990934	-0.002655	0.119850	-0.890889	0.008447	-1.127068
ROU_QR	-1.578718567	-0.002813506	0.057617124	-1.918073416	-0.001676584	0.076810471	-0.756820282	0.008479747	-0.862874224
ROU_RP	-1.631169997	-0.008240 22	0.295419876	-1.940141497	-0.015736249	0.303245098	-1.100035 8	0.018019029	-0.720251145

Parametrii pentru conversia volmului comercial in biomasa lemnoasa supraterana

Ecuatia $B=A*Vol^B$, unde Vol – volumul comercial pe picior

Tip de padure	A	B
ROU_PC	0.453425409	1.002847289
ROU_CB	0.488376	1.011117
ROU_AA	0.401728	0.997698
ROU_OC	0.414060	0.995031
ROU_PA	0.364690872	1.016230027
ROU_FS	0.649242	0.997663
ROU_PB	0.567652516	1.00460649
ROU_OB	0.638217	0.989001
ROU_QR	0.708919191	0.982355399
ROU_RP	0.605874314	1.014093923

Anexa 2b. Versiune articol asociat sarcina 5.6. (transmis la Carbon Balance and ManagementȘ https://cbmjournal.biomedcentral.com/)

Anexa 2c. Article title: *Downscaling dead organic matter and soil simulations and validations*

Introduction

.....

Methods

- Running Yasso15 with CBM output on sub national scale (intersection of climates and 10 forest types)
- Climate description for each climate unit (CLU)

CLU	Tma	Tmaxa	Tmina	Tamp	Precipa
44	8.6	20.1	-3.7	23.8	550.1
43	8.7	19.4	-4	23.4	663
36	10	21.3	-3.5	24.8	477

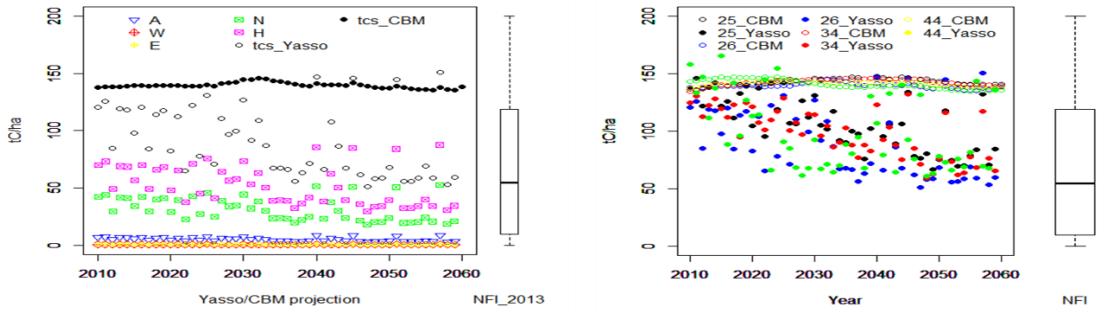
35	10.5	21	-2	23	600
34	10.6	21.1	-1.6	22.7	592
26	11.1	22.9	-1.2	24.1	587.2
25	11.6	20.8	1.7	19.1	540
45	8.6	20.1	-3.7	23.8	550.1

- Validation against total soil C stock measured by NFI in 2013

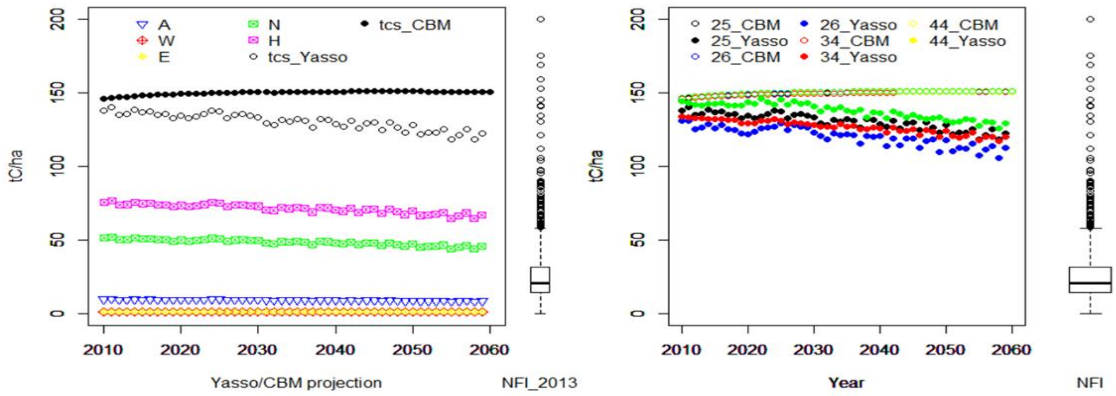
Results and discussion

Two analysis have been carried out: a) representationn of the AWEN and tocal C stock simulated by Yasso15 compared to output from CBM-CFS – as "dot" graph compared to NFI box-and-plot on the righ side (average value represented by a vertical bar), and b) representation of the total C stock for each climatic unit vs. Totalc C stock from CBM and NFI2013.

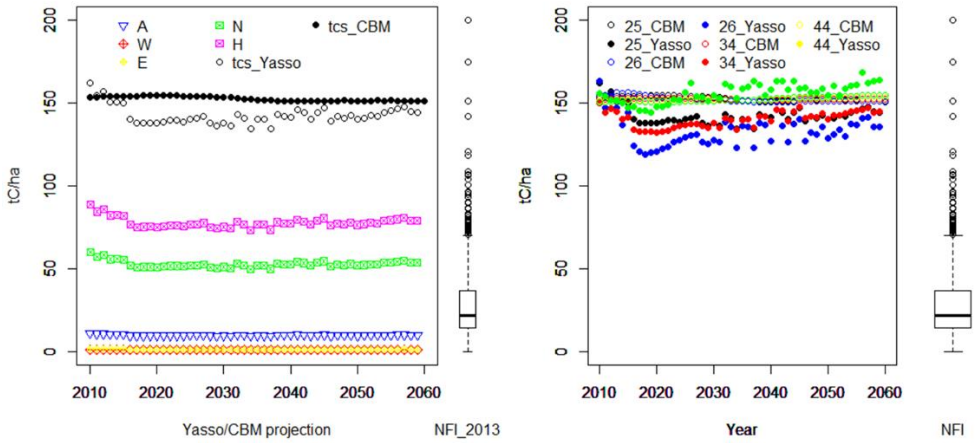
Graphical representation and comparison for AA (Abies alba) and CLU (climate unit). Left is AWENH vs. Total C stock from NFI. Right is comparison between Yasso15 and CBM for various CLU. Both graphs compare against measured NFI(=NFI2010).



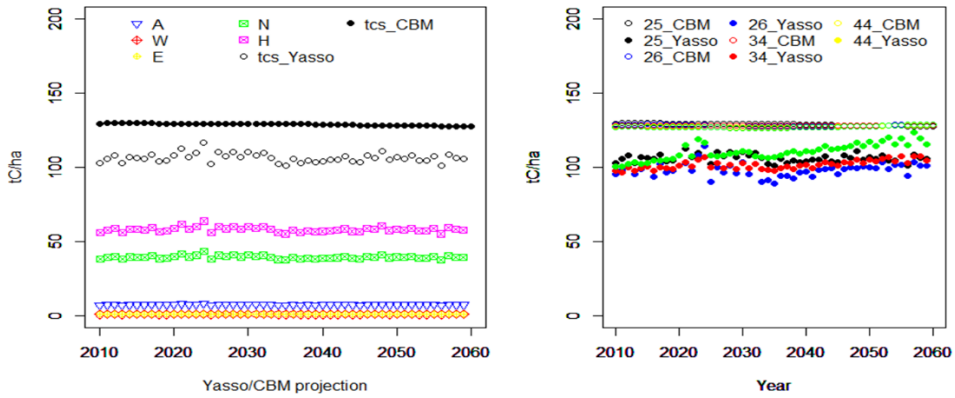
Graphical representation and comparison for RP (Fagus sylvatica) and CLU (climate unit)



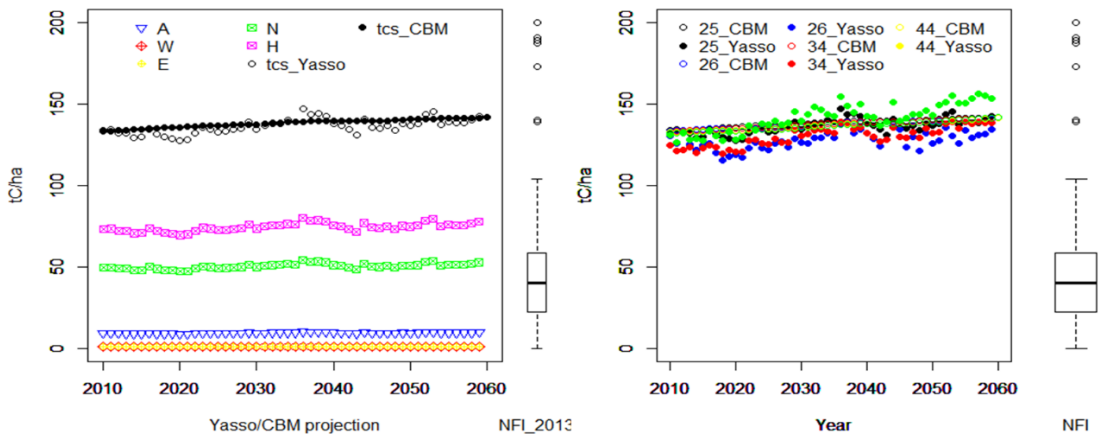
Graphical representation and comparison for ConBroad (mixed Coniferous and Broadleaves) and CLU (climate unit)



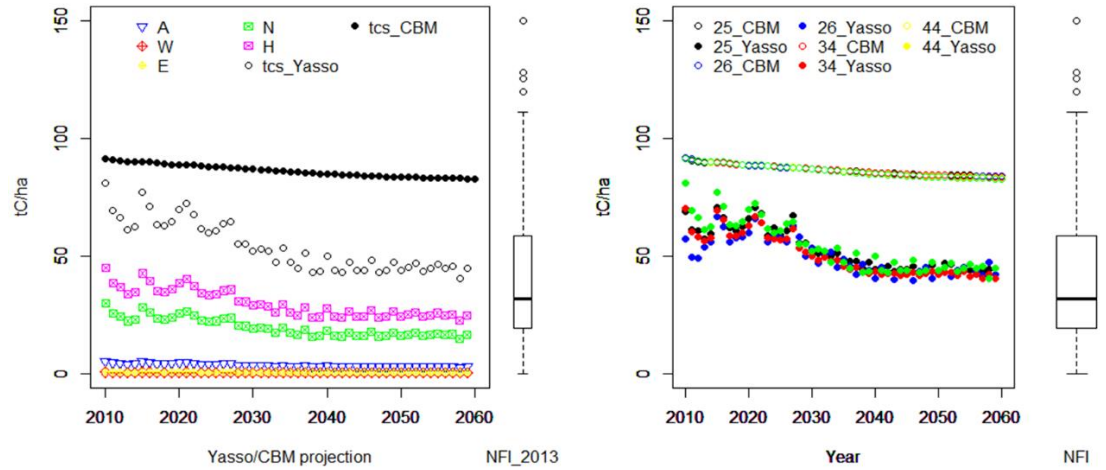
Graphical representation and comparison for PredBroad (predominanlty broadleaves) and CLU (climate unit)



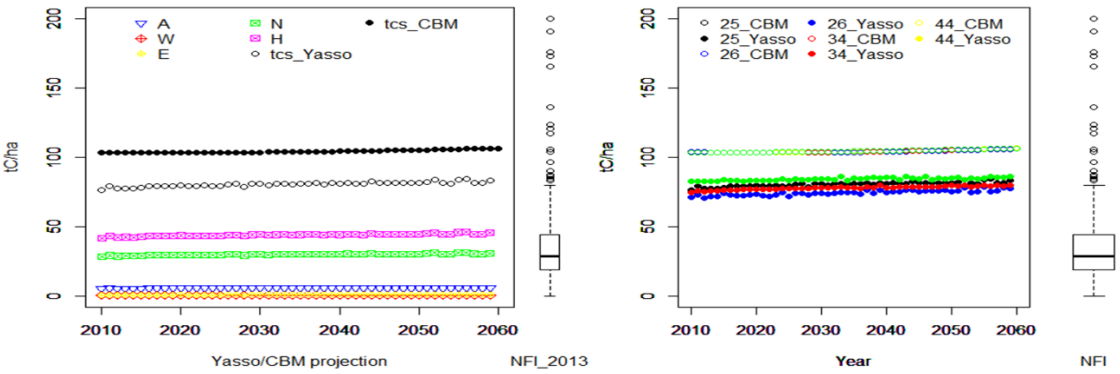
Graphical representation and comparison for PredCon (predominantly coniferous) and CLU (climate unit)



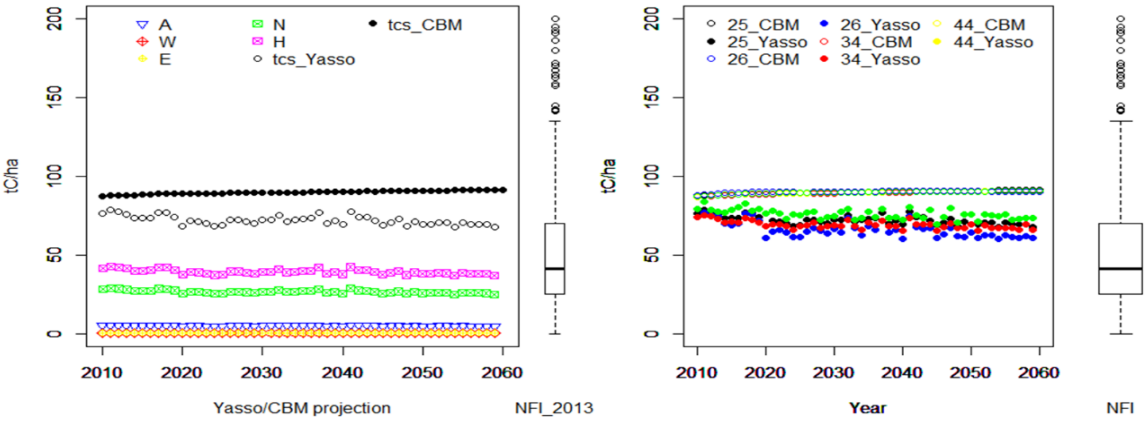
Graphical representation and comparison for RP (Robinia sp.) and CLU (climate unit)



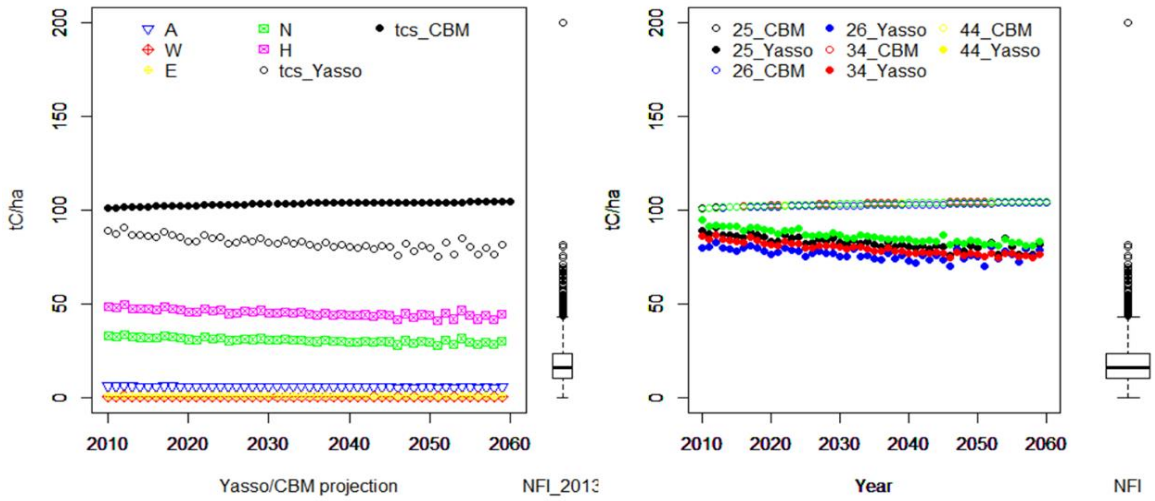
Graphical representation and comparison for OB (other broadleaved) and CLU (climate unit)



Graphical representation and comparison for PA (picea abies) and CLU (climate unit)



Graphical representation and comparison for PA (picea abies) and CLU (climate unit)



Anexa 4.

Propagation of uncertainty from allometric biomass model parameters: an investigation on parameter covariances

Ioan Dutca, Ronald McRoberts, Erik Naesset and Viorel NB Blujdea

Abstract
Forest biomass estimations at large scale typically rely on plot data from forest inventories. The uncertainty of such estimations is essential for successful implementation of mitigation policies such as REDD+. Allometric biomass models are routinely used to predict tree biomass which is further aggregated at plot level. Two approaches are common for fitting biomass allometric models: (i) linear model on log-log transformed data and (ii) nonlinear model. Tree diameter at breast height (D) and tree height (H) are used as predictors of aboveground biomass in these allometric models, either as separate variables, or as a combined variable (D^QH). Using a biomass dataset to develop generic allometric models and an inventory dataset (for a forest district in Romania), we investigated the effect of fitting approach (i.e. logarithmic transformation vs. nonlinear), and of predictor type (i.e. separate vs. combined) on biomass prediction precision and accuracy. Although the uncertainty sourced in model parameters was negligible, the uncertainty produced by residual variance was not, being the main source of uncertainty. Furthermore, when nonlinear models were weighted for heteroscedasticity, the differences between nonlinear and log-transformation were minor, in terms of both mean predicted biomass and its standard error. Combining the predictors (i.e. D^2H), the model uncertainty reduction due to avoidance of collinearity between D and H was modest and depended on fitting approach.

Keywords: biomass prediction uncertainty, error propagation, Monte Carlo, allometric models, combined predictor, logarithmic transformation

1. Introduction

It is widely acknowledged that world's forests represent a key player in the reduction of atmospheric CO_2 concentration (Bonan 2008, Bellassen and Luyssaert 2014, Grassi et al. 2017), forests alone absorbing roughly 30% of anthropogenic CO_2 emissions in the last few decades (Pan et al. 2011). The rate of CO_2 absorption by forests could be improved by conserving and enhancing the capacity for the carbon uptake in the current forests, by increasing forest area and by reducing/avoiding deforestation (Grassi et al. 2017). Because of large uncertainties associated with forest carbon estimations, it is often more difficult to implement effective mitigation policies (Gren and Zeleke Aklilu 2016). These limitations hamper the success of so called 'forest carbon activities' such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries), AR (Afforestation/Reforestation mechanism, part of Clean Development Mechanism) or IFR (Improved Forest Management), with possible negative consequences on climate change mitigation efforts.

Estimating biomass with great accuracy and precision is challenging. Uncertainties in forest biomass/carbon estimations arises from several sources that relate to sampling, measurement, model selection and model uncertainty (McRoberts et al. 2015, Yanai et al. 2018). Model uncertainty is often ignored when aggregating tree predictions at plot level in forest inventory programs (McRoberts and Westfall 2014). When ignoring model uncertainty, it is assumed that model parameters are fixed and that model residuals are always zero. In other words, it is assumed that trees of similar diameter and height exhibit always the same aboveground biomass, which in reality is never true.

Diameter at breast height (D) and tree height (H) are common predictors of biomass. Since D and H are related, the models based on both D and H will always show some degree of collinearity (Sileshi 2014, Dutcă et al. 2018b) with possible negative consequences on parameter non-identifiability and model uncertainty (Dormann et al. 2013). To avoid collinearity, a combined independent variable (i.e. D^2H) is frequently used in forestry studies, being argued that aboveground tree biomass/volume is proportional to a cylinder of diameter, D and height, H (Burkhart and Tomé 2012).

Combining the variables D and H into D^2H was shown to produce a loss in prediction accuracy (Dutcă et al. 2019), which is proportional to the Q-ratio (i.e. the ratio between parameter estimate of D and parameter estimate of H, see Dutcă et al. (2019)). However, the authors showed that a combined predictor can still be used (without adverse effect of prediction accuracy loss) if the ratio between parameter of D and parameter of H in this new combined predictor equals the Q-ratio. Therefore, in this study we use the combined predictor D^QH, where Q is the Q-ratio, as proposed by (Dutcă et al. 2019). The advantage of this combined predictor consists in the reduction of number of parameters (and therefore of variances) to be estimated and also in the reduction of number of covariances, from 3 (for a model where D and H are separate predictors of AGB) to just one, for the combined predictor based model (AGB predicted by D^QH). Therefore, in this paper we investigate whether reducing the number of variances and covariances to be estimated (i.e. by using the combined variable D^QH instead of D and H separately) has any effect on large scale biomass estimation accuracy and precision.

2. Material and methods

2.1. Data

2.1.1. Biomass data

Spruce data from (Schepaschenko et al. 2017).

To develop the allometric models, we used a publicly available dataset (Schepaschenko et al. 2017) consisting of 6308 trees sampled from Europe and Asia. Since the dataset contained many species, we removed those species not existing in the inventory dataset (to which the

model was applied to predict biomass). Therefore, we developed the models based on 1983 trees, called further 'model dataset'. For this reduced dataset, the diameter at breast height (D) varied between 5 and 72.9 cm, tree height (H), between 3.5 and 42.8 m, and aboveground biomass (AGB), between 2.2. and 4291.3 kg.

2.1.2. Inventory data

Pure spruce plots from Romanian NFI (=243 plots and 4946 trees).
The models developed based on biomass dataset were further applied to estimate biomass in 243 sample plots of pure Norway spruce. The 243 plots were selected from Romanian NFI, based on two conditions, simultaneously: (1) the plots contain only Norway spruce trees; (2) the H-D ratio of trees within plots ranges only between 0.5 and 1.7. This decision is justified by the fact that plot data should be within the same range as biomass data (with regard to H-D ratio), and by the fact that trees outside this H-D ratio range may be trees that are damaged (e.g. broken trunks) or trees with inaccurate measurements of either D or H. Because the Romanian NFI grid for mountain area (where pure Norway spruce occur) is 4 by 4 km, the 243 plots correspond to a forest area of 388.8 thousand hectares.

Height-diameter (H-D) ratio is one of the main drivers of variance in biomass allometric models (Feldpausch et al. 2010, Dutcă et al. 2018a). We checked whether the distribution of H-D ratio in model dataset matches the inventory dataset. Figure 1 shows a good agreement between histogram (model dataset) and density line (inventory dataset). H-D ratio ranged between 0.36 and 2.56 for model dataset and between 0.42 and 2.11 for the inventory dataset. The ranges of D and H were also similar to model dataset, varying between 5 to 72.8 cm and respectively between 3.8 and 44.8 m.

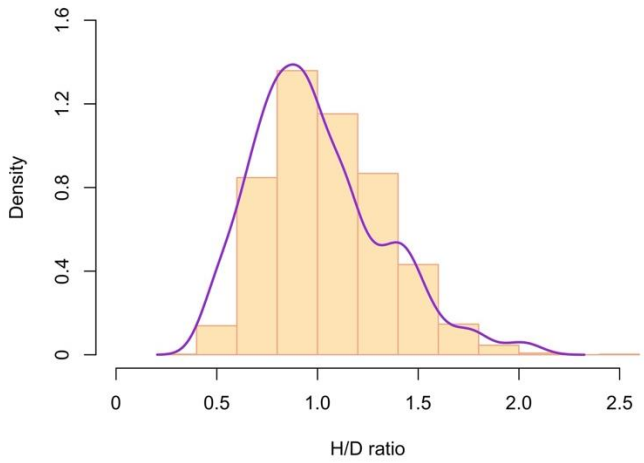


Fig. 1. The H-D ratio for dataset used to develop the models (histogram) and for inventory dataset (violet line)

2.2. Modelling framework

2.2.1. Weighted nonlinear model with separate variables

(1) $AGB = \beta_0 \cdot D^{\beta_1} \cdot H^{\beta_2} + \varepsilon$

Based on Eq. 6 we further determined the Q-ratio, $Q = \frac{\hat{\beta}_1}{\hat{\beta}_2}$, (Dutcă et al. 2019) and then calculated the combined variable $K = D^Q H$ which was used further into the modelling process.

2.2.2. Weighted nonlinear model with combined variable

(2) $AGB = \beta_0 \cdot K^{\beta_3} + \varepsilon$

Mathematically, Equations (1) and (2) are equivalent under $\beta_2 = \beta_3$, since equation (2) can be re-written as:

(3) $$AGB = \beta_0 \cdot (D^{\frac{\beta_1}{\beta_2}} \cdot H)^{\beta_3} + \varepsilon$$

However, equation (2) has a reduced number of parameters (2 parameters instead of 3 parameters in Eq. (1)) therefore a reduced number of variances and covariances to be estimated. The number of covariances is reduced from 3 covariances in Eq. (1) to just one covariance in Eq. (2).

2.3. The weights in nonlinear approach

Since in their nonlinear power-law form the allometric models usually exhibit heteroscedasticity (increase of variance with predictor) a weighted nonlinear model was used. We predicted the variance as a function of predicted AGB, and then calculated the weights as inverse of predicted variance, following the procedure used by (Dutcă et al. 2019).

2.4. Adjustment of heteroscedastic residual standard error

Because variance is heteroscedastic in allometric models, the residual standard error in not constant across the predictor range. To propagate the error from residual variance, we sampled from a normal distribution $\mathcal{N}(0, 1)$, which was truncated to the interval [-3, 3], and then adjusted the sampled value with the predicted standard deviation ($\widehat{\sigma_i}$), as a function of predicted biomass ($\widehat{AGB_i}$). We used a procedure in 8 steps, similar to calculating the weights (Dutcă et al. 2019): (i) calculate $\widehat{AGB_i}$ and residuals (ε_i) from weighted nonlinear allometric models (Equations 1 and 2); (ii) the pairs $\widehat{AGB_i}$ and ε_i were ordered ascending with respect to $\widehat{AGB_i}$; (iii) the pairs $\widehat{AGB_i}$ and ε_i were aggregated into groups of size 25; (iv) for each group, the mean $\widehat{AGB_g}$ ($\overline{\widehat{AGB_g}}$) and the standard deviation of ε_i (σ_g) were calculated; v) the resulted values ($\overline{\widehat{AGB_g}}$ and σ_g) were log-transformed (using natural logarithm); (vi) a linear model was fitted, predicting standard deviation [$\ln(\sigma_g)$] as a function of [$\ln(\overline{\widehat{AGB_g}})$]; (vii) the models were back transformed, using a correction factor as in Eq. 3; (viii) the residual standard error ($\widehat{\sigma_i}$) was predicted further as a function of $\widehat{AGB_i}$:

(4) $$\widehat{w} = 1/(0.2240 \times \widehat{AGB_i}^{0.9407})$$ for Equation (1), and

(5) $$\widehat{w} = 1/(0.2234 \times \widehat{AGB_i}^{0.9414})$$ for Equation (2).

2.5. Propagation of errors in AGB prediction

The AGB prediction and its uncertainty over 243 plots was assessed, following a seven-step Monte Carlo simulation procedure, adapted from McRoberts et al. (2015, 2016):

2.5.1. Using Taylor series based covariances

Step 1. For the *k*th replication, a random vector containing two or three parameters (i.e. β_0 and β_1 for models based on combined predictor; β_0, β_1 and β_2 for models based on separate predictors) is drawn at a time from a multivariate normal distribution; the multivariate normal distribution was created based on parameters resulted from Eqs. 1, 2, 6 and 7 and their variance-covariance matrices.

Step 2. Sample the residual. From a normal distribution N(0, 1) one value at a time was drawn. This value was adjusted by multiplication with the predicted standard deviation ($\widehat{\sigma_i}$) described earlier in section 2.4.

Step 3. For the i^{th} tree on the j^{th} plot the AGB was calculated. For nonlinear approach, the AGB was calculated as in Eqs. 6 and 7. The model parameters are those sampled in step 1 and the residual is from step 2.

Step 4. The biomass of each plot was calculated as: $AGB_j = 20 \times \sum_{i=1}^{n_j} AGB_i$, where n_j is the number of trees in plot j and AGB_i is the aboveground biomass of tree i, calculated in step 3.

Step 5. For the *k*th replication, the mean and its variance were calculated as:

(8) $\overline{AGB_k} = \frac{1}{m}\sum_{j=1}^{m} AGB_j$

(9) $\mathrm{var}(\overline{AGB_k}) = \frac{1}{m(m-1)}\sum_{j=1}^{m}(AGB_j - \overline{AGB_k})^2$

where m is the number of plots (i.e. m = 38).

Step 6. The steps 1-5 were replicated n_{rep} = 1000 times and the mean and variance over replications were calculated as:

(10) $\hat{\mu} = \frac{1}{n_{rep}}\sum_{k=1}^{n_{rep}} \overline{AGB_k}$

(11) $\mathrm{var}(\hat{\mu}) = (1 + \frac{1}{n_{rep}}) \cdot \frac{1}{n_{rep}-1}\sum_{k=1}^{n_{rep}}(\overline{AGB_k} - \hat{\mu})^2 + \frac{1}{n_{rep}}\sum_{k=1}^{n_{rep}} \mathrm{var}(\overline{AGB_k})$

(12) $\mathrm{SE}(\hat{\mu}) = \sqrt{\mathrm{var}(\hat{\mu})}$

Step 7. The steps 1-6 were repeated for 5000 times, until $\hat{\mu}$ and $\mathrm{SE}(\hat{\mu})$ stabilized. The last values in the chain were further reported (i.e. the mean $\hat{\mu}$ and $\mathrm{SE}(\hat{\mu})$ over 5000 repetitions).

5.2.2. Including a bootstrap procedure in the error propagation process

2.7. Data processing

Statistical analysis was performed in R (R Core Team 2017) with the RStudio interface (RStudio Team 2016) and using the packages ‘nlme’ (Pinheiro et al. 2018), ‘MASS’ (Venables and Ripley 2002).

3. Results

3.1. The curvature in the model parameters covariance

3.2. The effects on large scale biomass prediction accuracy and precision

The differences between logarithmic transformation and weighted nonlinear approach were minor, of up to 0.34% for the mean predicted biomass and of up to 4.3% for the standard error of the mean (Table 1). However, when nonlinear models were not weighted for heteroscedasticity, that resulted in larger differences, of up to 6.1% with regard to mean biomass and of up to 10% with regard to standard error. Comparable differences were observed also between weighted nonlinear and non-weighted nonlinear approach (5.8% and 9.3% with regard to mean biomass and standard error respectively). The mentioned differences between fitting approaches were slightly smaller for the separate variables than for combined variable models.

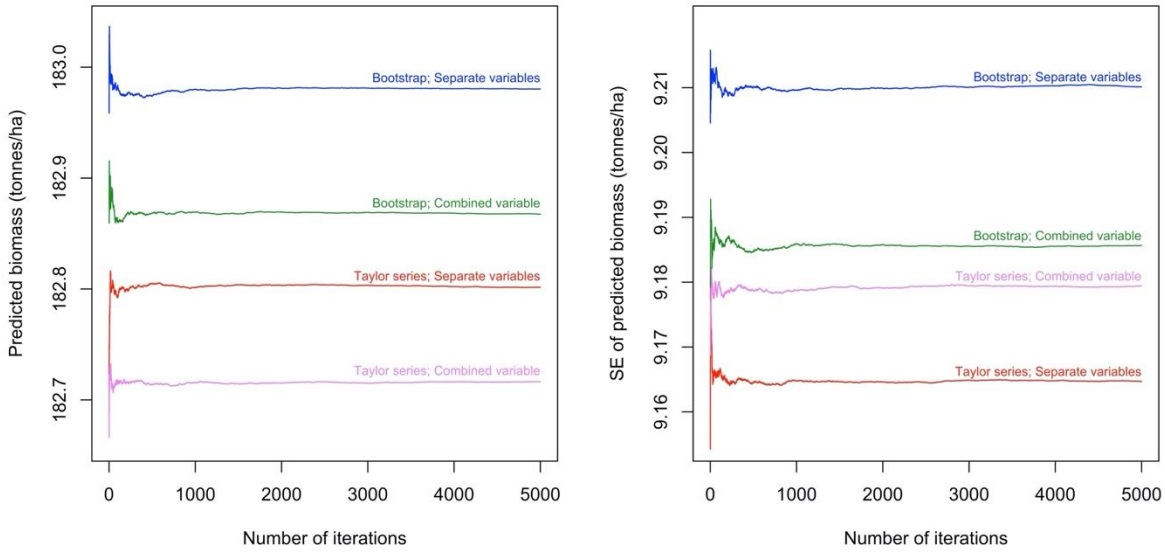


Fig. 2.
Mean biomass and standard error of the mean as resulted from different fitting approaches (i.e. logarithmic transformation, weighted nonlinear and nonlinear without weights), different predictor types (i.e. separate and combined) and different sources of uncertainty. *Note: Uncertainty sources: (1) uncertainty due to differences between plots only; (2) uncertainty due to differences between plots and model parameters; (3) uncertainty due to differences between plots and residual variance; (4) uncertainty due to differences between plots, model parameters and residual variance (I.a.5., I.a.6, I.b.5 and I.b.6 from Table 1, were not plotted).*

3.2. Predictor type

Contrary to our expectation that combined predictor model, due to avoidance of collinearity between D and H, would reduce the prediction uncertainty (when model uncertainty is included), the reduction was modest and depended on fitting approach. For non-weighted nonlinear models, the standard error of the mean, reduced indeed by up to 9%, for weighted nonlinear model the reduction was by up to 2%, while for the log-transformation approach the reduction was very small (i.e. 0.09%).

When ignoring model uncertainty, the standard error of the mean was lower for the separate variable models, due to introducing more degrees of freedom into the modelling process. The difference (between separate and combined predictor models when ignoring model uncertainty) was largest for non-weighted nonlinear model (i.e. 7.6%) and smallest for log-transformation approach (i.e. 0.7%). However, when including model uncertainty, that reversed in favour of combined predictor model. Standard error of the mean became smaller for the combined predictor model. But contrary to our expectation, this recovery was not produced by improvement of standard errors of the model parameters (due to avoidance of collinearity), but by residual variance.

The decrease of standard errors of the predicted mean biomass (due to avoidance of collinearity, therefore from separate variables to combined variable model) was largest for non-weighted nonlinear model (i.e. from -7.6% to +9%) and smallest for log-transformed

approach (i.e. from -0.7% to +0.09%). Therefore, the effect on prediction uncertainty produced by avoiding collinearity between D and H depended on fitting approach, being largest for non-weighted nonlinear models and lowest for logarithmic transformation approach.

With regard to mean predicted biomass, the differences between separate and combined variable models, were negligible for log-transformation and weighted nonlinear approaches (differences of up to 0.5%) and slightly larger for the non-weighted nonlinear approach (of up to 1.3%). In our analysis, mean predicted biomass per hectare was larger for separate variables model when using log-transformation and weighted nonlinear approaches, and larger for combined predictor models when using non-weighted nonlinear fitting approach.

3.3. Uncertainty sources

Regardless of fitting approach and predictor type, the effect of model parameters uncertainty was minor. The standard errors of the mean biomass per hectare increased only by up to 1.8% compared to when model uncertainty was completely ignored. However, the largest model uncertainty was sourced in residual variance. The residual variance of the models increased the standard error of the mean biomass per hectare by between 268% and 336%, depending on fitting approach and predictor type.
Very small differences between models including different uncertainty sources were observed with regard to mean predicted biomass (differences of up to 0.3%).

3.4. Multiplicative vs. additive error propagation

The residuals were attached to the back-transformed model (i.e. log-transformation approach) by multiplication (Eq. 4 and 5), rather than addition, as it was the case of weighted and non-weighted nonlinear models. The results showed that differences between the two propagation methods did not differ considerably. The standard error of the mean predicted biomass resulted from additive model differed from multiplicative model only by 4.3% for the combined predictor and by 2.3% for the separate variables models.

4. Discussions

4.1. Model uncertainty

Our results showed that when using generic allometric models to predict biomass, model uncertainty has a great impact on biomass prediction uncertainty and therefore should not be ignored. The greatest source of model uncertainty was in the residual variance. Although McRoberts and Westfall (2014) found residual variance to make little contribution to overall uncertainty in volume estimates, that was only valid when using species-specific allometric models. However, our results confirm the hypothesis of McRoberts et al. (2016) that the residual variance of a generic allometric model may produce uncertainty in the estimates that cannot be ignored. Our results also confirm that uncertainty produced by model parameter is small and therefore may be ignored in biomass estimations (McRoberts and Westfall 2014, Ståhl et al. 2014, McRoberts et al. 2015). This is an important aspect for practice, since the variance-covariance matrix needed to propagate errors from model parameters is often missing anyway in published research.

The residual variance was the main driver of model uncertainty. For our dataset the standard error of residuals was 0.2. Nevertheless, we expect larger effects of model uncertainty on total prediction uncertainty when the standard error of residuals is larger. For example, Chave et al. (2014) reported a standard error of residuals in logarithmic scale of 0.357, expecting therefore a larger effect on biomass prediction uncertainty in tropical forests.

We showed that using a correct weighting in nonlinear models, while the logarithmic transformation removes all heteroscedasticity, the results obtained by these two approaches are very similar. However, ignoring the weighting of heteroscedastic variance, the two approaches produce slightly different estimates. Logarithmic transformation is easy to implement and very straightforward. The error term (multiplicative), is normally distributed in logarithmic scale, therefore, model uncertainty can be propagated directly from a normal distribution. However, logarithmic transformation makes sense as long as heteroscedasticity is removed. If heteroscedasticity is not removed entirely, then a weighted nonlinear approach would be much more flexible. Although the multiplicative error approach can be implemented also for nonlinear models (Parresol 1999), the nonlinear approach, assumes an additive error. Because of heteroscedasticity, the distribution of additive residuals is assumed normal for every value of independent variable, but the resulting distribution of all residuals although symmetric, is not normal. Therefore, the limitation of additive error propagation in allometric models, is represented by the need to construct a tool to propagate errors from a normal distribution while the distribution of residuals is not normal (see section 2.4.).

For our analysis, the differences between separate predictors models and combined predictor models were dependent on fitting approach. The differences between separate and combined predictor models were larger for non-weighted nonlinear models with regard to both mean predicted biomass and standard error of the mean. However, Dutca et al. (in preparation) showed that efficiency of D^2H predictor depends on the ratio between parameter of D and parameter of H in separate variables model (Q ratio). For our dataset, Q-ratios were closer to the neutral value of 2.0 (which show equivalence between separate and combined predictor models) for log-transformation approach (Q = 2.31) and weighted nonlinear approach (Q = 2.45) whereas for non-weighted nonlinear approach the Q-ratio was much more distant from 2.0 (Q = 4.88). Therefore, our results confirm the hypothesis of Dutca et al. (in preparation) that the difference between separate and combined variable models becomes larger as Q-ratio increases. Nonetheless, our results should be interpreted with care, and make always reference to Q-ratio when making inferences about separate and combined variable models for log-transformation and weighted nonlinear approaches.

5. Conclusions

The conclusions of the study can be summarized as: (i) the uncertainty sourced in model parameters was small and therefore negligible, however, the uncertainty produced by residual variance was important and therefore should be always included when using generic allometric models; (ii) when nonlinear models were weighted for heteroscedasticity, the differences between nonlinear and log-transformation were minor, in terms of both means and standard errors, however, when not weighted, the differences were larger, of up to 10%; (iii) combining the predictors, the uncertainty reduction due to avoidance of collinearity between D and H was modest and depended on fitting approach; (iv) when propagating errors from residual variance, the correction factor for back transformation in allometric models is not necessary since the simple random sampling estimates directly the mean of lognormal distribution of residuals; (v) mean predicted biomass is based on mean of the lognormal

distribution of the intercept (which is different from the back-transformed mean of normal distribution) and therefore is biased, being proposed a correction factor.

References

Bellassen, V., and Luyssaert, S. 2014. Carbon sequestration: Managing forests in uncertain times. Nature **506**(7487): 153–155. doi:10.1038/506153a.

Bonan, G.B. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science **320**(5882): 1444–9. American Association for the Advancement of Science. doi:10.1126/science.1155121.

Burkhart, H.E., and Tomé, M. 2012. Modeling forest trees and stands. Springer.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., and Vieilledent, G. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. **20**(10): 3177–3190. doi:10.1111/gcb.12629.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., and Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.). **36**(1): 27–46. Blackwell Publishing Ltd. doi:10.1111/j.1600-0587.2012.07348.x.

Dutcă, I., Mather, R., Blujdea, V.N.B., Ioraș, F., Olari, M., and Abrudan, I.V. 2018a. Site-effects on biomass allometric models for early growth plantations of Norway spruce (Picea abies (L.) Karst.). Biomass and Bioenergy **116**. doi:10.1016/j.biombioe.2018.05.013.

Dutcă, I., Mather, R., Blujdea, V.N.B., Ioraș, F., Olari, M., and Abrudan, I.V. 2018b. Site-effects on biomass allometric models for early growth plantations of Norway spruce (Picea abies (L.) Karst.). Biomass and Bioenergy **116**: 8–17. doi:10.1016/j.biombioe.2018.05.013.

Dutcă, I., McRoberts, R.E., Næsset, E., and Blujdea, V.N.B. 2019. A practical measure for determining if diameter (D) and height (H) should be combined into D2H in allometric biomass models. For. An Int. J. For. Res. **In press**: 1–8. doi:10.1093/forestry/cpz044.

Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., Quesada, C.A., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bird, M., Brondizio, E.S., de Camargo, P., Chave, J., Djagbletey, G., Domingues, T.F., Drescher, M., Fearnside, P.M., França, M.B., Fyllas, N.M., Lopez-Gonzalez, G., Hladik, A., Higuchi, N., Hunter, M.O., Iida, Y., Abu Silam, K., Kassim, A.R., Keller, M., Kemp, J., King, D.A., Lovett, J.C., Marimon, B.S., Marimon-Junior, B.H., Lenza, E., Marshall, A.R., Metcalfe, D.J., Mitchard, E.T.A., Moran, E.F., Nelson, B.W., Nilus, R., Nogueira, E.M., Palace, M., Patiño, S., Peh, K.S.-H., Raventos, M.T., Reitsma, J.M., Saiz, G., Schrodt, F., Sonké, B., Taedoumg, H.E., Tan, S., White, L., Wöll, H., and Lloyd, J. 2010. Height-diameter allometry of tropical forest trees. Biogeosciences Discuss. **7**(5): 7727–7793. doi:10.5194/bgd-7-7727-2010.

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., and Penman, J. 2017. The key role of forests in meeting climate targets requires science for credible mitigation. Nat. Clim. Chang. **7**(3): 220–226. Nature Publishing Group. doi:10.1038/nclimate3227.

Gren, I.-M., and Zeleke Aklilu, A. 2016. Policy design for forest carbon sequestration: A

review of the literature. doi:10.1016/j.forpol.2016.06.008.
McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., and Westfall, J.A. 2016. Hybrid estimators for mean aboveground carbon per unit area. For. Ecol. Manage. **378**: 44–56. Elsevier. doi:10.1016/J.FORECO.2016.07.007.
McRoberts, R.E., Moser, P., Zimermann Oliveira, L., and Vibrans, A.C. 2015. A general method for assessing the effects of uncertainty in individual-tree volume model predictions on large-area volume estimates with a subtropical forest illustration. Can. J. For. Res. **45**(1): 44–51. NRC Research Press. doi:10.1139/cjfr-2014-0266.
McRoberts, R.E., and Westfall, J.A. 2014. Effects of Uncertainty in Model Predictions of Individual Tree Volume on Large Area Volume Estimates. For. Sci. **60**(1): 34–42. doi:10.5849/forsci.12-141.
Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., and Hayes, D. 2011. A large and persistent carbon sink in the world's forests. Science **333**(6045): 988–93. doi:10.1126/science.1201609.
Parresol, B.R. 1999. Assessing Tree and Stand Biomass: A Review with Examples and Critical Comparisons. For. Sci. **45**(4): 573–593.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team. 2018. nlme: Linear and Nonlinear Mixed Effects Models.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
RStudio Team. 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
Schepaschenko, D., Shvidenko, A., Usoltsev, V., Lakyda, P., Luo, Y., Vasylyshyn, R., Lakyda, I., Myklush, Y., See, L., McCallum, I., Fritz, S., Kraxner, F., and Obersteiner, M. 2017. A dataset of forest biomass structure for Eurasia. Sci. Data **4**: 170070. Nature Publishing Group. doi:10.1038/sdata.2017.70.
Sileshi, G.W. 2014. A critical review of forest biomass estimation models, common mistakes and corrective measures. For. Ecol. Manage. **329**: 237–254. Elsevier B.V. doi:10.1016/j.foreco.2014.06.026.
Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., and Holm, S. 2014. Sample-Based Estimation of Greenhouse Gas Emissions From Forests—A New Approach to Account for Both Sampling and Model Errors. For. Sci. **60**(1): 3–13. Oxford University Press. doi:10.5849/forsci.13-005.
Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. *In* 4th edition. Springer, New York.
Yanai, R.D., See, C.R., and Campbell, J.L. 2018. Current Practices in Reporting Uncertainty in Ecosystem Ecology. Ecosystems **21**(21): 971–981. doi:10.1007/s10021-017-0197-x.

Anexa 5.

Sampling trees to develop allometric biomass models: How does tree selection affect model's prediction accuracy and precision

Ioan Dutcă [a,b,*], Richard Mather [b] and Florin Iorași [b]

[a] Department of Silviculture, Transilvania University of Brasov, 1 Șirul Beethoven, Brașov 500123, Romania
[b] Buckinghamshire New University, Queen Alexandra Rd, High Wycombe HP11 2JZ, UK
[*] Corresponding author E-mail: idutca@unitbv.ro

Abstract

Developing allometric biomass models is a particularly important process, because the accuracy and precision of forest biomass and carbon estimations depend, to a degree, on these models. The effects of tree sampling on tree aboveground biomass (AGB) prediction accuracy and precision are intricate and, therefore, can often be misleading. In this paper, using a Monte Carlo simulation, we investigated how the model prediction accuracy and precision are affected by the tree sampling characteristics. Because diameter at breast height (D) is the most common predictor of AGB, we focused our analysis on AGB-D relationship. The following sampling characteristics were investigated: (i) sample size, (ii) length of D-range (difference between the largest and the smallest D value), (iii) position of D-range and (iv) distribution of sample trees. We found that, although the natural variability of AGB-D relationship was a key driver of both prediction accuracy and precision, the sampling characteristics were important mainly for improving prediction accuracy. Both, the sample size and distribution of sample trees, although having a negligible effect on precision, had a substantial role in improving prediction accuracy. Despite that small trees were more informative in allometric models, influencing goodness of fit and standard errors of model parameters, we showed that selecting a constant number of trees for each D class (i.e. uniform distribution of the sample trees on D-range) produced models with improved capability regarding prediction accuracy. The length and position of D-range, although considerably affecting the goodness of fit and the standard errors of allometric model parameters, had only a marginal effect on AGB prediction accuracy and precision. Furthermore, we showed that R^2 was a poor indicator of model prediction accuracy and precision, being sensitive to changes in D-range. Following these results, we develop practical recommendations to improve biomass prediction accuracy and precision.

Keywords: tree sampling, prediction, uncertainty, allometric equation, aboveground biomass, diameter at breast height

1. Introduction

It is widely accepted that forests play a critical role in the fight against climate change (Grassi et al., 2017), therefore accumulation of carbon in tree biomass is regarded as an important service provided to the society. However, to create sustainable mitigation measures and programmes such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation) requires that accumulation of carbon in forests is accurately and precisely estimated. Estimating carbon accumulation in forests is typically done based on forest inventory data to which allometric models are applied (Brown, 2002; Chave et al., 2004;

Clark et al., 2001; Stephenson et al., 2014). In a first step the forest biomass is estimated, then, using a constant proportionality ratio, e.g., 0.47 (IPCC, 2006), the biomass is transformed in carbon, which can be further converted into equivalent CO_2. Therefore, since the ratio between biomass and carbon is a constant, the terms 'carbon accumulation' and 'biomass accumulation' have roughly the same meaning.

Producing accurate and precise predictions of biomass is challenging for several reasons. First, it needs an unbiased forest inventory design with accurate measurements on tree attributes. Second, it requires that allometric biomass models are representative for the forest inventory data to which the model is applied. Selection of the allometric model was shown to be an important step in the reduction of biomass prediction uncertainty (Picard et al., 2015). Allometric biomass models are nonlinear regression models that typically use tree diameter at breast height (D) and/or tree height (H) to predict tree aboveground biomass (AGB), based on a sample of trees for which biomass was measured. Representativeness of the model to the forest inventory data requires that the trees were selected from the population (to which applies the forest inventory). Allometric biomass models were shown to depend largely on site conditions (Dutcă et al., 2018a) which would increase the complexity of tree sampling and would limit the transferability of these models.

The range of tree sizes and the distribution across the range are important prerequisites for sampling. The range represents the difference between largest and the smallest value of predictor (e.g. D) of sample trees used to build the model. Distribution of sample trees (on D-range) is often referred to as 'D class distribution' (Chave et al., 2004; Roxburgh et al., 2015) because D is usually measured in forest inventories with an increment of 2 cm (or other values) and therefore, a variable which would be expected to be continuous, is in fact discrete. However, when developing allometric biomass models, diameter at breast height (D) is measured as accurately as possible, therefore, D is usually a continuous variable.

Because allometric models are site-specific (Dutcă et al., 2018a), there are numerous examples of published allometric models based on trees sampled from one or few forest stands limiting therefore their D-range (Chojnacky et al., 2014; Jia et al., 2015; Marziliano et al., 2015; Morhart et al., 2016, 2013; Mosseler et al., 2014; Zianis et al., 2005) or deliberately developed allometric models only for small trees (e.g. Pajtík et al. 2008; Dutcă et al. 2010; Blujdea et al. 2012; Ciuvat et al. 2013). Nevertheless, the tree size is naturally limited, the maximum tree height being influenced by the stress and resource abundancy, while being limited by hydraulic constraints (Koch et al., 2004). Although maximum tree height is physically limited, trees continue to accumulate biomass by increasing their diameter (Stephenson et al., 2014). Generic allometric models and biomass databases often include very large trees, for example, D of up to 212 cm (Chave et al., 2014), up to 293 cm (Jucker et al., 2017) or even up to 648 cm (Falster et al., 2015).

The process of biomass measurement is very laborious; therefore, it is very important that sampling (for model development) is done optimally. The result of an optimized sampling should be an allometric model that predicts biomass as accurately and precisely as possible. In this paper, using a Monte Carlo analysis, we investigate which characteristics of tree selection affect biomass prediction accuracy and precision, and how. The sampling characteristics that were investigated are: (i) sample size, (ii) length of D-range (i.e. difference between largest and the smallest sample tree), (iii) position of D-range (i.e. when two D-ranges have the same length, the position is given by the starting or ending point of the range) and (iv) distribution of sample trees (i.e. frequency distribution of selected trees across D-range). Our study aims to provide key information for improvement of biomass prediction in forests.

2. Material and methods

To demonstrate the effects of sampling characteristics on biomass prediction accuracy and precision we performed a simulation study, which entails the following steps: (1) bivariate sets of AGB-D data were simulated, featuring various characteristics of the sample trees (e.g. AGB-D variability, sample size, D-range); (2) allometric biomass models were fitted to simulated data; (3) the allometric biomass models were then applied to predict the biomass in a plot, and the errors from model parameters and residual variability were propagated into plot AGB prediction; (4) the AGB prediction accuracy and precision (at plot level) were assessed; (5) examined which characteristics of the sample trees considered in the first step (i.e. AGB-D variability, sample size, D-range) affected the model's prediction accuracy and precision, and how.

2.1. Some rationale on the simulation design

Logarithmic transformation is widely regarded as a standard procedure in the development of allometric biomass models, although this is the subject of great debate (Kerkhoff and Enquist, 2009; Packard, 2012; Packard and Boardman, 2008; Xiao et al., 2011). The standard assumptions of this type of transformation are (i) the heteroscedasticity, which is common in allometric models, is entirely removed by transformation and (ii) the errors are lognormally distributed in original scale, so they will be normally distributed in log-log scale. If these two assumptions hold true, then the errors in original scale can be assumed multiplicative (the residuals in original scale are expressed as a ratio between observed and predicted biomass and therefore show the percent variation of observed biomass relative to predicted biomass). However, if the two assumptions do not hold true, then the logarithmic transformation would not be recommended, as the general assumptions of a linear model (e.g. normality of residuals, homogeneity of variance) would not be met. Xiao et al. (2011) showed that although both the multiplicative and the additive error-type relationships occur in nature, the relationships with multiplicative errors were much more frequent. Also, because diameter at breast height (D) is the most common predictor of tree aboveground biomass (AGB), we focused our simulation on ABG-D relationship, starting from a log-log linear model:

(1) $\ln(\mathrm{AGB}) = \beta_0 + \beta_1 \cdot \ln(\mathrm{D}) + \varepsilon$

Where AGB is the aboveground biomass (in kg); D is the diameter at breast height (in cm); 'ln' is the natural logarithm, β_0 and β_1 are the model parameters in logarithmic scale and ε is the additive error term (additive for the log-log scale), normally distributed with the mean zero. We then defined some true parameters of a hypothetical population. Because the population is hypothetical, to make the values of parameters credible, we derived the parameters from a real biomass dataset reported by Schepaschenko et al. (2017). Therefore, the true model parameters for our hypothetical population were:

(2) $\ln(\mathrm{AGB}) = -2.11 + 2.33 \cdot \ln(\mathrm{D}) + \varepsilon; \qquad \varepsilon \sim \mathrm{N}(0, \mathrm{RSE})$

Starting from these true parameters, we generated random sets of ln(AGB) – ln(D) data which were further fitted and then the resulting model applied to a plot dataset to estimate the biomass. Each generated dataset had specified characteristics such as RSE (residual standard error) of log-log model, number of observations, D-range length, position and distribution. A Monte Carlo approach (which will be described in greater detail) was used.

2.2. Natural variability of AGB-D relationship

The natural variability of AGB-D relationship is that intrinsic variability in the population that should be captured by the sampling design. This natural variability can be reduced by including additional AGB predictors in the model. However, the natural variability should not be influenced by the sampling characteristics such as the ones investigated in this study.

Because we assumed that heteroscedasticity is removed by logarithmic transformation and that errors are lognormally distributed in original scale, the natural (or intrinsic) variability of AGB-D relationship can be expressed as the standard deviation of residuals of the log-log linear model, or, commonly reported, as the residual standard error (RSE) (see Eq. 2). Since the residuals of a back-transformed log-log linear model show relative variation of AGB (relative to predicted AGB), the RSE can be interpreted, for original scale, as a form of coefficient of variation (Cole and Altman, 2017). We tested two values of RSE in this study, 0.2 and 0.3, which can be interpreted as 20% and 30% coefficient of variation. These two values are within a range that is common in allometric biomass models. For the dataset used to derive the parameters (Schepaschenko et al., 2017), the RSE was 0.28.

2.3. Sampling characteristics

2.3.1. Number of observations (sample size)

The number of sample trees necessary to develop an allometric model depends on desired precision, on the level of intrinsic variability (of AGB-D) relationship and some other factors. Roxburgh et al. (2015) performed a simulation study to find the number of sampled trees necessary to develop allometric models. They concluded that, given the intrinsic variability of trees and the differences between distribution of tree diameters used to construct the model and the distribution of tree diameters of the inventory data, a number of 17 to 166 trees were required to obtain prediction with a standard deviation within 5% from the mean. However, Picard et al. (2012) suggested that approximately a minimum number of 100 trees was needed to construct volume models. In our simulation design we tested three values of sample size, n = 100, n = 150 and n = 1000 trees. The first two values (n = 100 and n = 150) were intended to determine the effect of a 50% increase in sample size, as to compare it to a 50% increase in RSE (from RSE = 0.2 to RSE = 0.3). The third value (n = 1000) was intended to see how increasing the sample size influences the model prediction performance.

2.3.2. The length of D-range

The range of diameter at breast height (D) used in allometric biomass models varies greatly. In a compilation of allometric models, Zianis et al. (2005) showed that the largest sampled tree included in that compilation of studies was 90 cm. However, most studies included in the compilation were based on a relatively narrow D-range, with no consistent starting point for the range. For example, the largest tree of 90 cm was recorded in an allometric model for *Quercus ilex* in Italy and the minimum recorded diameter was 20 cm. Comparable maximum limits of D-range has been reported in some recent biomass datasets for boreal and temperate forests (Schepaschenko et al., 2017; Ung et al., 2017), although larger D-range of sampled trees were reported for tropical regions (Chave et al., 2014; Falster et al., 2015; Jucker et al., 2017). For our simulation study, assuming a maximum D-range in allometric biomass models between 0 and 90 cm (Zianis et al., 2005), we divided this interval in three, resulting three diameter intervals of 30 cm. Starting from the second interval (i.e. I_{min} = [30, 60]), we gradually expanded I_{min} in two directions (i.e. towards small diameter and towards large diameters) until reaching the limits of the maximum D-range, resulting seven D intervals. However, besides these seven intervals we examined another one for the entire D-range (i.e. I_{max} = [0, 90]), testing therefore a total of eight D intervals (Table 1).

Table 1. D-range intervals used for simulation

Code	D-range (cm)	Description
S_3	[0, 60]	I_{min} + 30 cm towards small diameters
S_2	[10, 60]	I_{min} + 20 cm towards small diameters

S_1	[20, 60]	I_{min} + 10 cm towards small diameters
I_{min}	[30, 60]	The minimum D-range length
B_1	[30, 70]	I_{min} + 10 cm towards large diameters
B_2	[30, 80]	I_{min} + 20 cm towards large diameters
B_3	[30, 90]	I_{min} + 30 cm towards large diameters
I_{max}	[0, 90]	The maximum D-range length

2.3.3. The position of D-range

The position of D-range is characterized by the starting point of D-range. Each pair of similar length of D-range, began at two different positions (Table 1). For example, the intervals S_1 and B_1 have similar length (i.e. 40 cm) but their starting position differ by 10 cm. This difference increases to 20 cm for S_2 vs. B_2 and to 30 cm for S_3 vs. B_3 (Table 1).

2.3.4. Distribution of sample trees

The desired frequency distribution of the sample trees is a very important prerequisite for tree selection because it dictates the level of physical effort and logistics required for measuring biomass. If the selection of trees would be done in a completely random manner, the distribution of the sample trees would match that of population. However, the trees are not selected in a completely random manner but they are randomly sampled for each D-class (McRoberts et al., 2015). A 'D class' represents a grouping of tree diameters into classes of a specified range. For example, for a 2 cm D class, the entire D-range is divided into intervals ("classes") of 2 cm (e.g. D = 10 to 12 cm represents a D class). The worker has therefore the possibility to control the shape of resulting frequency distribution over the entire D-range. Nevertheless, the distribution of sample trees will influence how the model will be informed across the range of D, with consequences for prediction. In our simulation, we explored four types of distribution (Fig. 1):

(a) Uniform distribution on D-range (Fig. 1, a), where a constant number of sample trees is selected for each D class. However, for our simulation, we randomly selected the trees from a continuous, uniform distribution, because D was assumed being accurately measured.
(b) Normal distribution on D-range (Fig. 1, b), where the largest number of sample trees is from the middle of D-range and decreases towards the margins of the range;
(c) Uniform distribution on ln(D)-range (Fig. 1, c1), which, for the original scale (D-range) is equivalent to inverse of uniform distribution (Fig. 1, c2, resulted by exponentiation of observations sampled from a uniform distribution on ln(D)-range).
(d) Normal distribution on ln(D)-range (Fig. 1, d1), which is equivalent to lognormal distribution on D-range (Fig. 1, d2). For both, the uniform and normal distribution on ln(D)-range, a larger number of small trees is sampled compared to large trees (Fig. 1, c2 and d2).

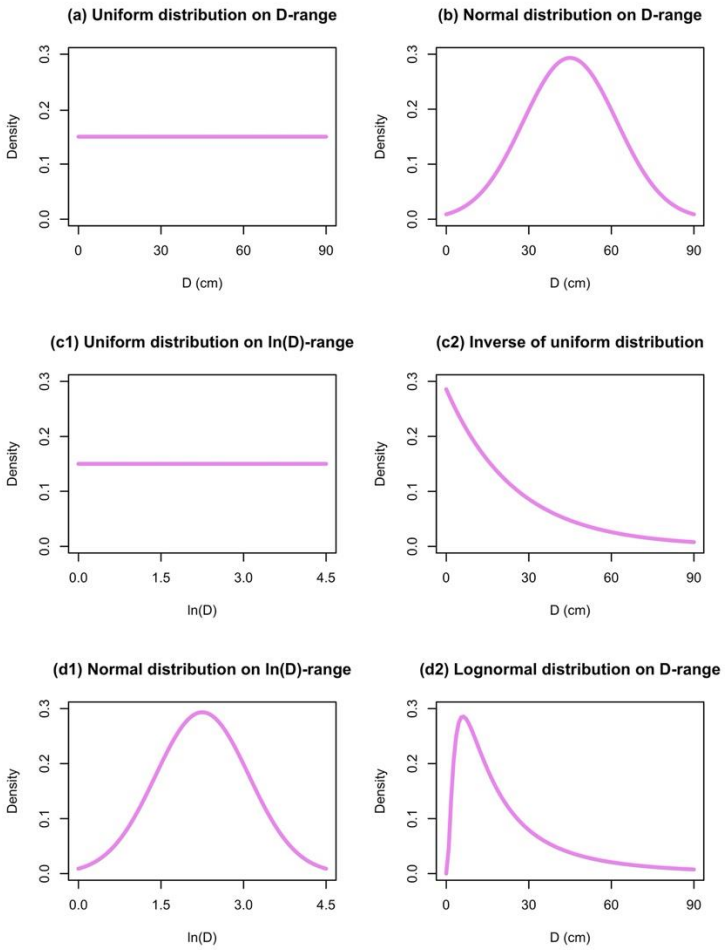


Fig. 1. The illustration of distributions used for simulation. (a) Uniform distribution on D-range; (b) Normal distribution on D-range; (c1) Uniform distribution on ln(D)-range, which, on D-range, is equivalent to inverse of uniform distribution (c2); (d1) Normal distribution on ln(D)-range, for which, the equivalent in original scale (therefore on D-range) is the lognormal distribution (d2).

For uniform distributions, the limits of D interval are easily implemented. However, the normal distribution has no limited range, which, theoretically, extends to infinity. For our simulation we therefore sampled from a truncated normal distribution, in which lower and upper bounds of D-range were established. This was done using 'rtruncnorm' function from 'truncnorm' package in R (Mersmann et al., 2018). We have set the D-range to correspond to $\pm$ two standard deviations, therefore representing an interval in which 95% of observations from a normal distribution should occur. The mean of the normal distribution (μ_d) was the mean of D interval of interest:

(3) $$\mu_d = D_{min} + \frac{(D_{max} - D_{min})}{2}$$

whereas the standard deviation (σ_d) was calculated as:

(4) $$\sigma_d = \frac{\mu_d - D_{min}}{2}$$

where D_{min} and D_{max} the minimum and respectively maximum limits of the D-range of interest (Table 1). For example, the normal distribution for I_{min} = [30, 60] cm was defined by the mean (for this example μ_d = 45) and standard deviation of the mean, which in this case σ_d = 7.5.

2.4. Plot data

We compared the simulated models, by assessing their accuracy and precision, when estimating the biomass in a plot. Therefore, each allometric model developed based on simulated data was further applied to estimate the biomass in a 500 m^2 plot. The plot contains 21 trees for which biomass was predicted as a function of D, by all simulated models. Because the only D interval that is common to all tested D-ranges is I_{min} = [30, 60] cm, we selected a plot that contained only tree diameters that fell within the I_{min} interval (Fig. 2). The aim of this plot is to be used as reference for prediction, for all simulated models in this study (in total, 4.8 billion allometric models). Therefore, the predicted AGB of this plot should be of no value unless compared to predicted AGB resulted from another model with different sampling characteristics.

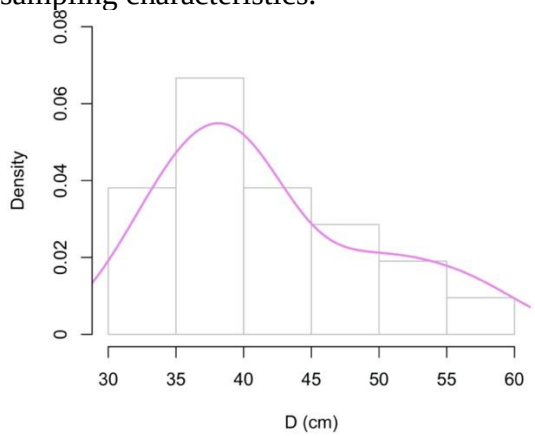


Fig. 2. The distribution of the 21 sample trees in the plot.

It is known that models have a poorer prediction performance on the extremes of covariate range. For example, a biomass model that was developed based on sample trees with D = 1 to 90 cm, would perform best when predicting biomass of trees from the centre of D-range (D = 45.5 cm) and worst when predicting the biomass of a tree with either D = 1 cm or D = 90 cm. Therefore, one of the aims of the study was to check how models perform across D-range. Consequently, another reason for selecting just one plot with D-range limited to I_{min} was to check the prediction performance when just part of model's D-range was used for prediction. A third reason for considering just a single plot was linked to the sources of uncertainty. In this study we aimed to account only the uncertainty produced by the model (from model parameters and residuals) and intentionally avoided including the uncertainty produced by differences between plots.

2.5. Monte Carlo simulation

A Monte Carlo analysis, including an error propagation approach adapted from McRoberts et al. (2015, 2016), was used to assess the effects of sampling characteristics on biomass prediction. We followed the next steps:

1. For the *k*th replication (K = 5000), an allometric model was developed and applied to predict biomass in the plot. The allometric model was developed based on simulated ln(AGB)-ln(D) data selected from the hypothetical population:
    1.1. defined a vector containing the errors of log-log linear model. The length of this vector equals the sample size (3 values of sample size were used in this analysis, *n* =100, *n* = 150 and *n* = 1000, see section 2.3.1). The values of the vector were randomly selected from a normal distribution with the mean zero and standard deviation either 0.2 or 0.3 (later on in the simulation design, the standard deviation of

this distribution will become the residual standard error, RSE, of the allometric model; two values of RSE were used, RSE = 0.2 and RSE = 0.3, see section 2.2).

1.2. defined a vector containing sample ln(D) values, which were randomly selected from a specific distribution (4 types of distribution were used, see section 2.3.4) and a specific D-range (a total of 8 intervals were used, Table 1). Because we fit the models in log-log scale, for uniform and normal distributions on D-range (Fig. 1, a and b), we randomly selected the sample D values from a uniform and respectively normal distribution on D-range and then log-transformed the sampled values (in order to get the ln(D) values); for uniform and normal distributions on ln(D)-range, we sampled the ln(D) values directly in log-log scale, from a uniform and respectively normal distribution on ln(D)-range (Fig. 1, see c1 and d1);

1.3. defined a vector (length of the vector equals the sample size, see section 2.3.1) containing the sample ln(AGB) values. Using the ln(D) values (obtained at step 1.2) and the error term (obtained at step 1.1) in Eq. 2, we predicted the ln(AGB) values.

1.4. fitted a linear model with the sample ln(AGB) (obtained at step 1.3) and ln(D) values (obtained from step 1.2):

(5) $$\ln(\mathrm{AGB}) = \beta_0 + \beta_1 \cdot \ln(\mathrm{D}) + \varepsilon$$

1.5. We retained the standard errors of model parameters, SE(β_0) and SE(β_1), and the coefficient of determination (R^2):

(6) $$R^2 = 1 - \frac{\sum(\ln(\mathrm{AGB})_i - \ln(\widehat{\mathrm{AGB}})_i)^2}{\sum(\ln(\mathrm{AGB})_i - \overline{\ln(\mathrm{AGB})}_i)^2}$$

Where $\ln(\mathrm{AGB})_i$ is the observed ln(AGB); $\ln(\widehat{\mathrm{AGB}})_i$ is the predicted ln(AGB) and $\overline{\ln(\mathrm{AGB})}_i$ is the mean of ln(AGB).

1.6. defined the variance-covariance matrix to account for the covariance between β_0 and β_1 in the next steps;

2. In this step we used the allometric model developed at step 1 (one model for each *k*th repetition) to estimate the biomass in a plot. Therefore, for the *j*th repetition (J = 5000), we:

2.1. defined a vector containing two values (β_0 and β_1) sampled at a time from a bivariate normal distribution (based on variance-covariance matrix of the allometric model developed at step 1.6, and model parameters);

2.2. defined a vector containing one error term (ε_i) sampled at a time from a normal distribution with the standard deviation equal to the residual standard error of the allometric model (Eq. 5) developed for each *k*th replication.

2.3. calculate the prediction of each tree ($\widehat{\mathrm{AGB}}_i$) in the plot based on the sampled parameters (from step 2.1) and error (from step 2.2):

(7) $$\widehat{\mathrm{AGB}}_i = \exp(\beta_0 + \beta_1 \cdot \mathrm{D}_i + \varepsilon_i)$$

2.4. calculate the predicted plot biomass ($\widehat{\mathrm{AGB}}_j$) by aggregation of individual tree predictions:

(8) $$\widehat{\mathrm{AGB}}_j = \sum_{i=1}^{m} \widehat{\mathrm{AGB}}_i$$

Where m = 21 is the total number of trees in the plot.

3. Step 2 was repeated for J times. For each *k*th replication we calculated:

3.1. the mean plot AGB over J repetitions:

(9) $$\overline{\widehat{\mathrm{AGB}}_k} = \frac{1}{J}\sum_{i=1}^{J} \widehat{\mathrm{AGB}}_j$$

3.2. standard error of the mean:

(10) $$\hat{\sigma}_k = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(\widehat{\mathrm{AGB}}_j - \overline{\widehat{\mathrm{AGB}}_k})^2}$$

3.3. relative bias:

(11) $$\mathrm{Bias_k}(\%) = \frac{\left(\overline{\widehat{\mathrm{AGB}}_\mathrm{k}} - \mu\right)}{\mu} \cdot 100$$

where μ is the plot AGB, based on true population parameters (plot true AGB); was calculated by applying the model based on true parameters (see Eq. 2) with a correction factor (Baskerville, 1972; Goldberger, 1968). The model was applied to all m = 21 trees in the plot and then their biomasses were aggregated. RSE is the residual standard error and can take two values, 0.2 and 0.3 (see section 2.2):

(12) $$\mu = \sum_{\mathrm{i}=1}^{\mathrm{m}} (\exp(2.11 + \frac{\mathrm{RSE}^2}{2}) \cdot \mathrm{D_i}^{2.33})$$

4. Steps 1-3 were repeated for K times (K = 5000) and we reported further:
   4.1. Standard deviation of relative bias, which is reported here as a measure of prediction accuracy (P_A):

(13) $$\mathrm{P_A} = \sqrt{\frac{1}{\mathrm{K}-1} \sum_{\mathrm{k}=1}^{\mathrm{K}} (\mathrm{Bias_k} - \overline{\mathrm{Bias}})^2}$$

Where $\overline{\mathrm{Bias}} = \frac{1}{\mathrm{K}} \sum_{\mathrm{k}=1}^{\mathrm{K}} (\mathrm{Bias_k})$

   4.2. The mean coefficient of variation of predicted biomass, which is reported as a measure of prediction precision (P_P):

(14) $$\mathrm{P_P} = \frac{1}{\mathrm{K}} \sum_{\mathrm{k}=1}^{\mathrm{K}} \frac{\widehat{\sigma}_\mathrm{k}}{\overline{\widehat{\mathrm{AGB}}_\mathrm{k}}} \cdot 100$$

Where $\widehat{\sigma}_\mathrm{k}$ is the standard error of predicted biomass (Eq. 10); $\overline{\widehat{\mathrm{AGB}}_\mathrm{k}}$ is the mean predicted plot biomass (Eq. 9).

2.6. Prediction accuracy and precision

The prediction accuracy and precision represent two qualitative concepts that are used to describe the performance of an estimator (Walther and Moore, 2005). Prediction accuracy, as used in this paper, is defined as the overall distance between predicted value and the true value (Walther and Moore, 2005). Since in our simulation design we determined not one, but 5000 values showing the distance between predicted AGB (at plot level) and true AGB (at the plot level), the accuracy was reported as standard deviation of these 5000 values (Standard deviation of relative bias, P_A, Eq. 13). Furthermore, the prediction precision is a measure of 'the statistical variance of an estimation procedure' (Walther and Moore, 2005) which is a type of uncertainty caused by random variation. In this study, the precision was reported as the mean coefficient of variation of predicted biomass at plot level (P_P) in Eq. 14.

2.7. Data processing

The simulation analysis was performed in R (R Core Team, 2017) with the RStudio interface (RStudio Team, 2016) and using the packages "MASS" (Venables and Ripley, 2002) and "rtruncnorm" (Mersmann et al., 2018).

3. Results

3.1. The effects on standard errors of model parameters and on goodness of fit

The simulation results demonstrate that with increasing D-range, the standard errors of model parameters (SE(β_0) and SE(β_1) in Eq. 5) decreased while the R^2 values (Eq. 6) increased (Fig. 3 and Annex 1). The effects were stronger when the length of D-range increased towards small trees (Fig. 3, $S_1 - S_3$) compared to large diameters (Fig. 3, $B_1 - B_3$). Increasing the length of D-range, the largest reduction of SE(β_0) and SE(β_1) and the largest increase of R^2 occurred for normal distribution on ln(D)-range (Fig. 3, d1-d3). Although in Fig. 3 the results are shown only for *n* = 100 and RSE = 0.3, the results for the other values of sample size and RSE (Annex 1) show a similar pattern.

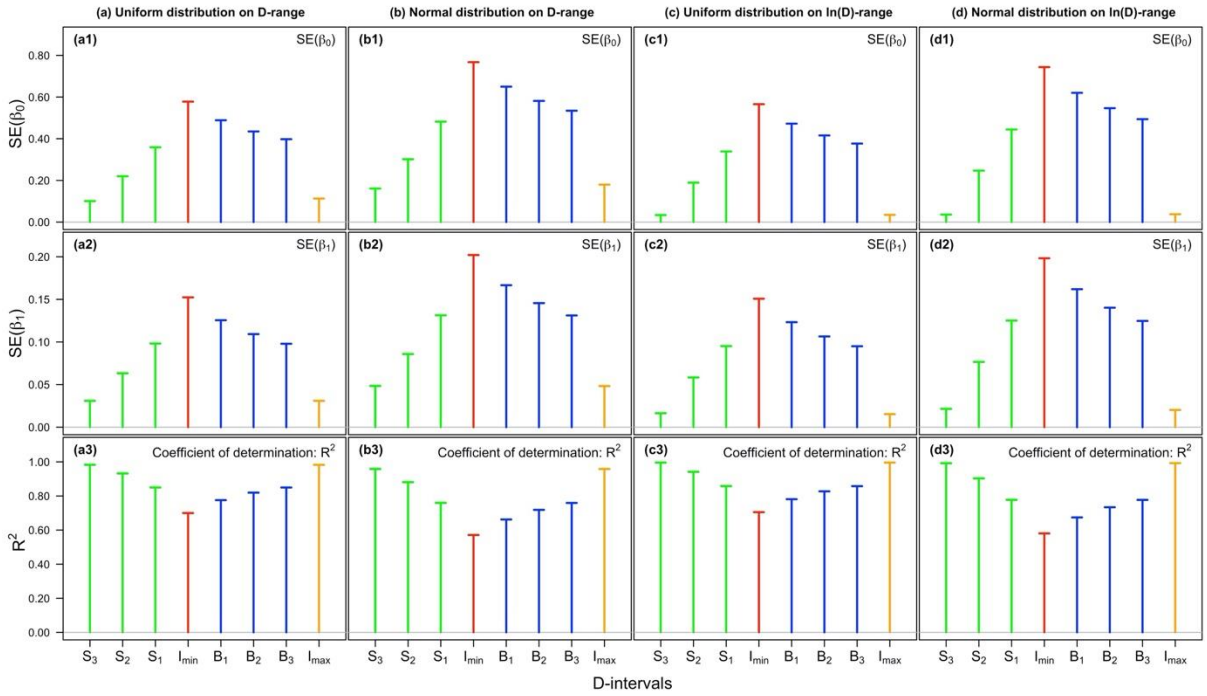


Fig. 3. The standard errors of model parameters SE(β_0) and SE(β_1), and the model goodness of fit (R^2) of log-log transformed allometric biomass model (Eq. 5), shown for different types of distribution of the sample trees and different D-intervals. Note: Each column (a-d) represents a different type of distribution of the sample trees (for more information see section 2.3.4); SE(β_0) is the standard error of the intercept in Eq. 5 (see steps 1.4 and 1.5 in section 2.5); SE(β_1) is the standard error of the slope in Eq. 5 (see steps 1.4 and 1.5 in section 2.5); R^2 is the coefficient of determination (Eq. 6). Each of the values presented in this plot (for SE(β_0), SE(β_1) and R^2) was calculated as the mean over 5000 replications (see section 2.5). In this graph is presented the data only for models based on one value of sample size (*n* = 100) and one value of residual standard error (RSE = 0.3).

The standard errors of model parameters were affected by both RSE and sample size. However, the model goodness of fit (R^2) was affected mainly by the RSE with only a very minor effect from the sample size.
Increasing the RSE by 50% (from 0.2 to 0.3) the standard errors of model parameters (intercept and slope) increased by the same 50% rate (SD = 0.31%; calculated based on values presented in Table A1, and Table A2 in Annex1) whereas the effect on R^2 was dependent on the length of D-range and on the type of distribution (Fig. 3). For models based on smaller D-range lengths and on trees sampled over a normal distribution (on either D or ln(D)), the effects of increasing RSE on R^2 were stronger.
Increasing the sample size by 50% (from 100 to 150 trees), the standard errors of model parameters decreased, on average, by 18.7% (SD = 0.36%) while increasing the sample size by 1000% (from 100 to 1000) the standard errors decreased by 68.7% (SD = 0.33%). Nevertheless, increasing the sample size by 50% (from 100 to 150) the R^2 values increased by an average of 0.07% while increasing the sample size 10-fold (from 100 to 1000), the R^2 values increased only by an average of 0.18% (Annex 1).

3.2. The effects on biomass prediction accuracy

As expected, residual standard error (RSE) was an important driver of prediction accuracy (expressed as standard deviation of relative bias, P_A, Eq. 13). A low P_A value means that the

distance between predicted AGB and the true AGB is small, and therefore the model is more accurate. Increasing RSE from 0.2 to 0.3 (therefore, by 50%), P_A increased by approximately the same ratio (i.e. by an average of 51.4%, SD = 2.3%; the mean and SD were calculated on 96 P_A values presented in Table A4, Annex1, resulted from 8 D-intervals, 3 values of sample size and 4 types of distribution). The effect was stronger for models based on shorter D-range lengths (Fig. 4 and Table A4 in Annex 1). Sample size was also an important factor affecting biomass prediction accuracy, although its effect was weaker compared to that of RSE. Increasing the sample size by 50% (from 100 to 150), P_A decreased by an average of 18.4% (SD = 1.2%; calculated on 96 values in Table A4). Increasing the sample size by 1000% (from 100 to 1000) the average decrease of P_A was 67% (SD = 0.8%; calculated on 96 values in Table A4). These effects were very similar to those found for standard errors of model parameters.

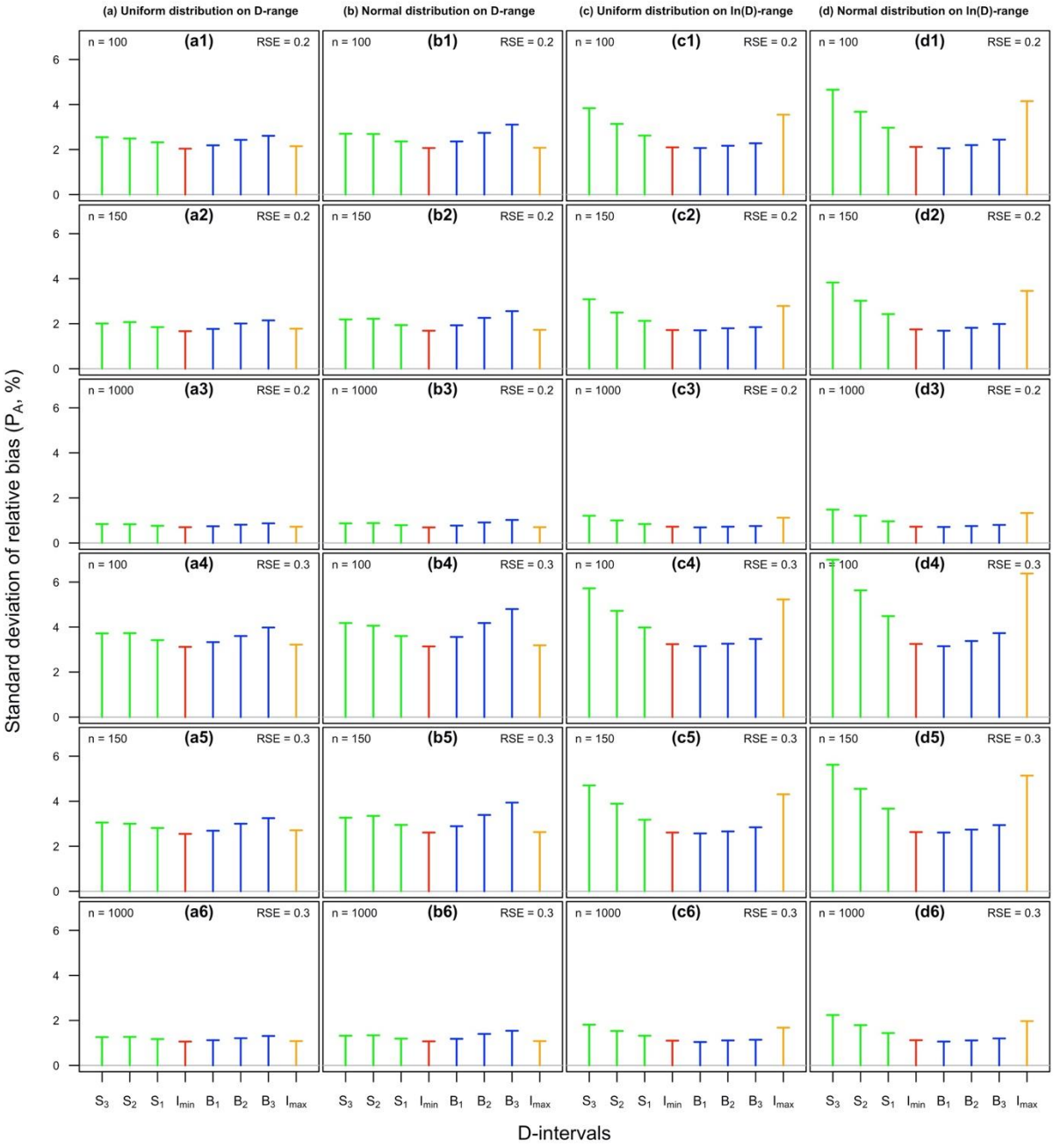


Fig. 4. The standard deviation of relative bias, describing prediction accuracy (P_A, Eq. 13, see section 2.5), for different sampling characteristics. Note: Each column (a-d) represents a

different type of distribution of the sample trees (for more information see section 2.3.4); The rows (1 to 6) represent a combination of sample size (n = 100, n = 150 and n = 1000, see section 2.3.1) and residual standard error values (RSE = 0.2 and RSE = 0.3, see section 2.2); Inside each plot are presented the P_A values for each D-interval (in cm, S_1 = [0, 60]; S_2 = [10, 60]; S_3 = [20, 60]; I_{min} = [30, 60]; B_1 = [30, 70]; B_2 = [30, 80]; B_3 = [30, 90]; I_{max} = [0, 90], see Table 1).

The variability of P_A values was lowest for uniform distribution on D-range (Fig. 4, a1-a6). That means the models which were constructed based on trees which were selected following a uniform distribution on D-range produced more stable prediction accuracies across model's D-range. In other words, sampling a constant number of trees for each D-class makes the allometric models less vulnerable to accuracy loss, when just part of model D-range is used for prediction.

However, models that were based on trees selected over uniform or normal distributions on ln(D) range (Fig. 4, c1-c6 and d1-d6), produced larger P_A values for $S_1 – S_3$ intervals compared to $B_1 – B_3$. The cause of these differences is how well the model was informed for D = 30 to 60 cm. We mentioned already (section 2.3.4) that the uniform or normal distribution on ln(D) range (see Fig. 1, c1, c2, d1 and d2) assume that a larger number of small trees are selected compared to large ones. Therefore, the models based on uniform and normal distribution on ln(D)-range (Fig. 4, c1-c6 and d1-d6) are better informed on their left side of D-range (towards small trees). However, the models based on $S_1 – S_3$ (in Fig. 4, c1-c6 and d1-d6) used their right-side of D-rage for prediction, which was not so well informed (e.g. models based on S_3 were developed for D = 0 to 60 cm and were used to predict biomass of trees with D = 30 to 60 cm), producing less accurate predictions. This is opposite to models based on $B_1 – B_3$ which used their well-informed part of D-range (e.g. models based on B_3 were developed for D = 30 to 90 cm, and, used to predict biomass of trees with D = 30 to 60 cm), and, therefore, produced more accurate AGB predictions.

Because the models based on $S_1 – S_3$ and $B_1 – B_3$ intervals used just part of their D-range for prediction (e.g. the model based on S_3 although being developed for D = 0 to 60 cm, was used to predict the biomass of trees with D = 30 to 60 cm), would be biased to compare the prediction accuracy of I_{min} with $S_1 – S_3$ (or $B_1 – B_3$). Since the prediction accuracy is poorer at the margins of D-range (of the model) it is to be expected that P_A values increase slightly (for models based on $S_1 – S_3$ and $B_1 – B_3$). However, both I_{min} and I_{max} based models, use the central portion of D-range for prediction and therefore these two can be compared to assess how increasing the length of D-range affects prediction accuracy. Increasing the range from I_{min} to I_{max} the prediction accuracy did not improve, but had the opposite effect, especially for distributions on ln(D)-range (Fig. 4, c1-c6 and d1-d6) for which the P_A value increased by up to 98%. For models based on uniform and normal distribution on D-range (Fig. 4, a1-a6 and b1-b6) the increase was much smaller, of up to 6.6%.

We showed the effects of increasing D-range length from I_{min} to I_{max} when the number of observations remained constant. Therefore, although the models based on I_{max} show greater R^2 and smaller standard errors of model parameters (Fig. 3), their prediction accuracy was poorer compared to models based on I_{min} (Fig. 4, see I_{min} vs. I_{max}). This suggests that the absolute number or density of observations for each part of D-range (or for each diameter class) is important. For the specific D-range of the plot data (i.e. D = 30 to 60 cm), the models based on I_{max} had a lower density of observations, compared to models based on I_{min}, since the same number of observations had to be distributed over a wider D-range (in case of I_{max} based models). These results are important, because they show that is not the model fitting and the standard errors of model parameters that matters for prediction accuracy, but the RSE (in log-log scale) and the absolute number of trees across the D-range.

3.3. The effects on biomass prediction precision

Although increasing the length of D-range produced a decrease of standard errors of model parameters and an increase of R^2 (Fig. 3), this was not reflected in the precision of biomass prediction (here, expressed as mean coefficient of variation of predicted biomass, P_P, in Eq. 14). The P_P did not decrease with the increasing D-range and in some cases even increased slightly (Table 2). When comparing the P_P of models based on I_{min} vs. I_{max}, the results were found to be highly consistent when trees were sampled on D-range (differences of up to 0.1%). However, when the trees were sampled on ln(D)-range the differences increased slightly up to 1.9% (Table 2).

Table 2
The mean coefficient of variation of predicted biomass (P_P, Eq. 14)

D interval	Uniform distribution on D-range			Normal distribution on D-range			Uniform distribution on ln(D)-range			Normal distribution on ln(D)-range		
	n=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000	*n*=100	*n*=150	*n*=1000
RSE = 0.2												
S_3	20.32	20.29	20.22	20.35	20.30	20.22	20.53	20.41	20.23	20.78	20.54	20.26
S_2	20.31	20.24	20.21	20.32	20.32	20.20	20.44	20.33	20.22	20.54	20.38	20.22
S_1	20.30	20.26	20.20	20.30	20.30	20.22	20.39	20.29	20.20	20.44	20.34	20.22
I_{min}	20.27	20.25	20.20	20.30	20.25	20.21	20.28	20.28	20.20	20.26	20.23	20.20
B_1	20.28	20.27	20.21	20.28	20.25	20.21	20.25	20.26	20.21	20.26	20.27	20.21
B_2	20.29	20.31	20.21	20.39	20.31	20.22	20.29	20.26	20.20	20.25	20.25	20.21
B_3	20.35	20.32	20.21	20.45	20.35	20.23	20.27	20.27	20.21	20.30	20.27	20.21
I_{max}	20.28	20.25	20.21	20.31	20.27	20.21	20.49	20.36	20.20	20.64	20.49	20.25
RSE = 0.3												
S_3	30.94	30.82	30.72	31.07	30.82	30.73	31.30	31.02	30.74	31.68	31.36	30.78
S_2	30.91	30.82	30.71	30.96	30.88	30.70	31.18	30.98	30.73	31.28	31.01	30.75
S_1	30.85	30.81	30.72	30.90	30.88	30.71	30.98	30.89	30.72	31.08	30.94	30.72
I_{min}	30.80	30.78	30.71	30.81	30.78	30.69	30.85	30.80	30.71	30.89	30.83	30.70
B_1	30.86	30.75	30.70	30.90	30.85	30.70	30.89	30.81	30.72	30.85	30.80	30.68
B_2	30.95	30.83	30.69	31.04	30.83	30.73	30.94	30.77	30.69	30.93	30.83	30.70

B_3	30.93	30.88	30.72	31.09	30.98	30.73	30.92	30.81	30.70	30.96	30.88	30.72
I_{max}	30.82	30.78	30.71	30.79	30.76	30.70	31.22	30.99	30.73	31.44	31.14	30.76

In Table 2 can be observed that P_P is highly related to residual standard error (RSE). We mentioned (section 2.2) that RSE in log-log scale can be interpreted as a form of coefficient of variation for original scale. The slightly larger P_P values than 20% and respectively 30% were caused by the uncertainty in model parameters, since the P_P values contain the errors propagated from both model parameters and residual variance. Therefore, RSE was the main driver of model prediction precision, with a very small share produced by uncertainty in model parameters (up to 5.3%). Increasing the RSE by 50% (from 0.2 to 0.3), P_P increased on average, by 52.1% (SD = 0.2%; the mean and SD were calculated on 96 P_P values, presented in Table 2, for each value of RSE), regardless of sample size, D-range and distribution type. However, sample size, although greatly influencing prediction accuracy, had only a minor effect on prediction precision. Since increasing the sample size has direct effect on the standard errors of model parameters (producing a decrease in standard errors) and since the propagated errors from model parameters represent only a very small share in P_P value (up to 5.3%), it is to be expected that sample size will have a insignificant effect on prediction precision. Increasing the number of observations by 50% (from 100 to 150), the P_P decreased by 0.33% (SD = 0.29%) while increasing by 1000% (from 100 to 1000) the P_P decreased by 0.81% (SD = 0.56%). However, both these effects were not significantly different from zero ($p = 0.26$ and $p = 0.16$ respectively).

4. Discussion

4.1. The drivers of biomass prediction accuracy and precision

The effects of tree sampling characteristics on biomass prediction accuracy and precision are intricate and can sometimes be misleading. In this paper we show which characteristics of the sampling strategies are important for improving model's prediction accuracy and precision. We found that the natural variability of AGB-D relationship (expressed by RSE) was the main driver of model's prediction accuracy and precision (an increase of RSE by 50% produced a roughly similar effect on prediction accuracy and precision). Increasing the sample size was important for improving the accuracy but not so important for improving precision. The effect of sample size on prediction accuracy depended on RSE and D-range, and was a function of $1/\sqrt{n}$, confirming previous work (Chave et al., 2004; Picard et al., 2012).

We showed that a wider D-range would improve the model fit and standard errors of model parameters (Fig. 3). This may be helpful for any statistical test associated with allometric models, since reduction of standard errors may produce a stronger effect against null hypotheses (e.g. t-test, F test) (Dutcă et al., 2018b). However, we also showed that, although the model based on a wider D-range had a better fit, the prediction accuracy was poorer (Fig. 4, see I_{min} vs. I_{max}). This result, which may be surprising, can be explained by another important factor which is influenced by the distribution of trees across D-range, which is the density of the observations across the D-range. If the number of observations remain constant, increasing the D-range, requires that the density of observations across D-range is reduced, with negative consequences on AGB (aboveground biomass) prediction accuracy. Often, increasing the range of D is achieved by merging different datasets for different D-ranges. In this case, the density of observations across the D-range is not reduced and the resulting increase of sample size increases prediction accuracy.

Furthermore, Roxburgh et al. (2015) suggested that the optimal distribution of sampled trees to develop allometric models is the one that matches the distribution of trees to which the model is applied. Although our plot data appears to be lognormally distributed (Fig. 2), the greatest accuracy (lowest P_A value) was obtained for models based on uniform distribution of D-range and not for models based on lognormal distribution on D-range (Fig. 4, a1-a6 vs. d1-d6), as suggested by the authors. Since our plot D data only appeared to be lognormal, we checked this hypothesis by generating a new D dataset of 1000 observations lognormally distributed on I_{max} interval. We investigated whether the model based on uniform distribution (developed for the same I_{max} range) produced lower P_A and P_P values (when predicting AGB of this new D dataset of 1000 observations) than model based on lognormal distribution. The results confirmed that uniform distribution on D-range produced lower P_A and P_P values (model based on uniform distribution: P_A = 3.2% and P_P = 30.8%; model based on lognormal distribution: P_A = 6.3% and P_P = 31.4%). We repeated the comparison, for models based on uniform vs. normal distribution on D-range, when predicting AGB of 1000 trees normally distributed. Again, the model based on uniform distribution produced lower P_A and P_P values compared to model based on normally distributed sample trees (model based on uniform distribution: P_A = 3.5% and P_P =30.8%; model based on normal distribution: P_A = 3.6% and P_P = 30.9%). Therefore, our results show that models based on uniform distribution of the sample trees on D-range perform better (produce more accurate and precise predictions), regardless of distribution of the trees to which the model is applied.

4.2. Small trees are more informative in allometric models

We showed that, for models based on similar number of observations and similar length of D-range (and similar residual standard errors in logarithmic scale), if the models included smaller diameter trees, the standard errors of model parameters were reduced while R^2 values were greater (e.g. see S_3 vs. B_3 in Fig. 3). Therefore, it is suggested that small trees are more informative in allometric models, compared to large trees. However, this seemingly anomalous finding can be explained by (or represents the indirect effect of) a commonly heteroscedastic nature of relationship between biomass and tree diameter. The variance in allometric models is not constant and increases with D (Zianis, 2008). As a result, to fit a nonlinear model the observations are usually weighted inversely to residual variance (the lower the residual variance, the larger the weight and vice-versa) (Dutcă et al., 2019). Logarithmic transformation on the other hand, performs a similar function: it re-scales data so that units are stretched for small values of variables (D and AGB) and compressed for large ones. Therefore, log-log transformation more heavily weights the influence of small trees over large ones, to ensure that residuals are comparable residuals across predictor range (i.e. homoscedasticity). As the lowest residual variance usually occurs for the smallest D values (Zianis and Mencuccini, 2004), small trees are more heavily weighted and have a greater influence on regression models than larger trees. Therefore, small trees impart more information to models, and exert greater overall influence over the standard errors of model parameters and goodness of fit. Given the fact that small trees require less effort for biomass measurement, they are highly cost-effective to sample. Nevertheless, we showed that, although the models that included small trees produced smaller standard errors of model parameters and larger R^2 values, those effects were not significantly reflected in biomass prediction performance (as indicated by the accuracy and precision of models, Fig. 4 and Table 2).

4.3. Selection criteria of allometric models

Goodness of fit (R^2 of linear model in log-log scale) is often reported with allometric biomass models, being recommended as criterion for model selection (Sanquetta et al., 2018). The assumption is that a model with the best fit will reasonably predict the biomass of other trees. Our results confirm that R^2 was not affected by sample size (Sanquetta et al., 2018). However, we showed that R^2 was a poor indicator of model prediction performance with respect to both accuracy and precision. Plotting the R^2 against P_A (Fig. 5, a) and P_P (Fig. 5, b) we observed no clear relationship between R^2 and model prediction accuracy or precision. Although not sensitive to changes in sample size, R^2 was sensitive to variations of D-range (Fig. 3 and Annex 1). The maximum R^2 values occurred when D-range length was also maximum (i.e. I_{max}, see Fig. 3) and when distribution of sampled trees was uniform on ln(D)-range (R^2 = 0.998, Table A3, Annex 1). However, we showed that the length of D-range did not affect the prediction accuracy nor precision, and that actually the models based on trees sampled along a ln(D)-range produced poorer prediction accuracies. Therefore, these findings suggest that R^2 may not be a good indicator of model prediction performance.

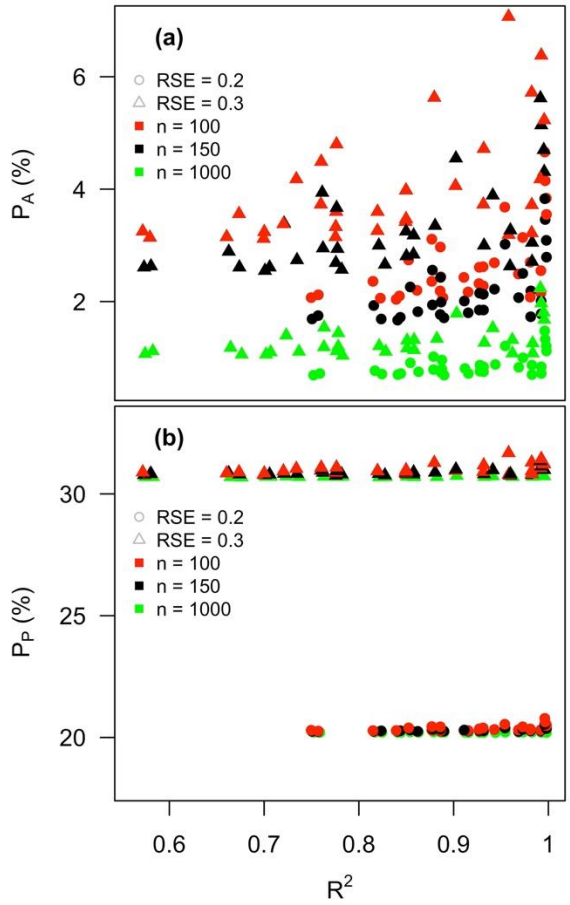


Fig. 5. The relationship between model's goodness of fit (R^2) and prediction accuracy (P_A, %) and precision (P_P, %). Note: (a) The relationship between standard deviation of relative bias (P_A, Eq. 13) and model's R^2 (Eq. 6); (b) The relationship between mean coefficient of variation of predicted biomass (P_P, Eq. 14) and model's R^2 (Eq. 6); The values plotted are from Table A3, Table A4 (Annex 1) and Table 2.

4.4. Limitations of the study

Our study has the following limitations. First, the conclusions are only valid if the assumptions hold that heteroscedasticity is removed by logarithmic transformation and that errors are normally distributed in log-log scale. Second, the study was limited to AGB-D relationship; therefore, the conclusions should not be inferred to other types of relationships.

Third, this study did not consider the uncertainty produced by differences between plots. Fourth, we assumed that the diameters of trees in the inventory (plot) dataset were always within the D-range used to construct the model. We did not investigate the consequences of predicting AGB of trees outside the range of diameters used to construct the models.

4.5. Recommendations:

(1) Select a constant number of trees for each D class (use a uniform distribution of sample trees). We showed that the models based on uniformly distributed sample trees over D-range produced more accurate AGB predictions, regardless of distribution of inventory dataset. Also, the variations in prediction accuracy across D-range were minimal.

(2) Avoid using R^2 as criterion for model selection. We showed that R^2 was a poor indicator of model prediction performance.

(3) Use strategies to avoid unnecessary large levels of RSE in allometric models. Because RSE is indicative of the intrinsic AGB variability for any given D, it cannot be naturally reduced. However, because RSE was a key driver of both prediction accuracy and precision, we recommend adopting strategies that can help reducing unnecessary large levels of AGB variability, such as: (i) avoiding using generic allometric models, where species effect is ignored and use species-specific allometric models instead; (ii) test and include additional predictors in the models that may explain part of the residual variance, such as tree height, crown diameter and wood density.

(4) Because the residuals represent the most important factor influencing model performance, we recommend that residuals should always be checked, paying special attention to normality and homogeneity of variance.

5. Conclusions

The key conclusions drawn from this study are as follow: (i) residual variance was the most important driver of model's prediction accuracy and precision; (ii) increasing the sample size improved prediction accuracy (although its effect was weaker than that of residual standard error), but had negligible effect on prediction precision; (iii) increasing the length of D-range, although improving both the goodness of fit and standard errors of model parameters, did not affect prediction accuracy nor precision; (iv) distribution of sampled trees was important for prediction accuracy; we found that uniform distribution of D-range was optimal, regardless of the distribution of the inventory dataset, (v) R^2 was not a good indicator of prediction performance of the allometric model, and (vi) small trees were more informative in allometric models, because of inherently heteroscedastic variance. However, the effect on the overall prediction performance of the model was negligible.

Acknowledgements

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CCCDI – UEFISCDI, project number ERANET-FACCE ERAGAS – FORCLIMIT (82/2017), within PNCDI II.

References

Baskerville, G.L., 1972. Use of Logarithmic Regression in the Estimation of Plant Biomass. Can. J. For. Res. 2, 49–53. doi:10.1139/x72-009

Blujdea, V.N.B., Pilli, R., Dutca, I., Ciuvat, L., Abrudan, I.V., 2012. Allometric biomass equations for young broadleaved trees in plantations in Romania. For. Ecol. Manage. 264, 172–184. doi:10.1016/j.foreco.2011.09.042

Brown, S., 2002. Measuring carbon in forests: current status and future challenges. Environ. Pollut. 116, 363–372. doi:10.1016/S0269-7491(01)00212-3

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for tropical forest biomass estimates. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 409–20. doi:10.1098/rstb.2003.1425

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. 20, 3177–3190. doi:10.1111/gcb.12629

Chojnacky, D.C., Heath, L.S., Jenkins, J.C., 2014. Updated generalized biomass equations for North American tree species. Forestry 87, 129–151. doi:10.1093/forestry/cpt053

Ciuvat, A.L., Abrudan, I.V., Blujdea, V., Dutca, I., Nuta, I.S., Edu, E., 2013. Biomass Equations and Carbon Content of Young Black Locust (Robinia pseudoacacia L.) Trees from Plantations and Coppices on Sandy Soils in South-Western Romanian Plain. Not. Bot. Horti Agrobot. Cluj-Napoca 41, 590–592. doi:10.15835/NBHA4129355

Clark, D.A., Brown, S., Kicklighter, D.W., Chambers, J.Q., Thomlinson, J.R., Ni, J., 2001. Measuring Net Primary Production in Forests: Concepts and Field Methods. Ecol. Appl. 11, 356–370. doi:10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2

Cole, T.J., Altman, D.G., 2017. Statistics Notes: Percentage differences, symmetry, and natural logarithms. BMJ 358, j3683. doi:10.1136/bmj.j3683

Dutcă, I., Abrudan, I.V., Stăncioiu, P.T., Blujdea, V.N.B., 2010. Biomass conversion and expansion factors for young Norway Spruce (Picea abies (L.) Karst.) trees planted on non-forest lands in Eastern Carpathians. Not. Bot. Horti Agrobot. Cluj-Napoca 38. doi:http://dx.doi.org/10.15835/nbha3835450

Dutcă, I., Mather, R., Blujdea, V.N.B., Ioraș, F., Olari, M., Abrudan, I.V., 2018a. Site-effects on biomass allometric models for early growth plantations of Norway spruce (Picea abies (L.) Karst.). Biomass and Bioenergy 116, 8–17. doi:10.1016/j.biombioe.2018.05.013

Dutcă, I., McRoberts, R.E., Næsset, E., Blujdea, V.N.B., 2019. A practical measure for determining if diameter (D) and height (H) should be combined into D2H in allometric biomass models. For. An Int. J. For. Res. In press, 1–8. doi:10.1093/forestry/cpz044

Dutcă, I., Stăncioiu, P.T., Abrudan, I.V., Ioraș, F., 2018b. Using clustered data to develop biomass allometric models: The consequences of ignoring the clustered data structure. PLoS One 13, e0200123. doi:10.1371/journal.pone.0200123

Falster, D.S., Duursma, R.A., Ishihara, M.I., Barneche, D.R., FitzJohn, R.G., Vårhammar, A., Aiba, M., Ando, M., Anten, N., Aspinwall, M.J., Baltzer, J.L., Baraloto, C., Battaglia, M., Battles, J.J., Bond-Lamberty, B., van Breugel, M., Camac, J., Claveau, Y., Coll, L., Dannoura, M., Delagrange, S., Domec, J.-C., Fatemi, F., Feng, W., Gargaglione, V., Goto, Y., Hagihara, A., Hall, J.S., Hamilton, S., Harja, D., Hiura, T., Holdaway, R., Hutley, L.S., Ichie, T., Jokela, E.J., Kantola, A., Kelly, J.W.G., Kenzo, T., King, D., Kloeppel, B.D., Kohyama, T., Komiyama, A., Laclau, J.-P., Lusk, C.H., Maguire, D.A., le Maire, G., Mäkelä, A., Markesteijn, L., Marshall, J., McCulloh, K., Miyata, I., Mokany, K., Mori, S., Myster, R.W., Nagano, M., Naidu, S.L., Nouvellon, Y., O'Grady, A.P., O'Hara, K.L., Ohtsuka, T.,

Osada, N., Osunkoya, O.O., Peri, P.L., Petritan, A.M., Poorter, L., Portsmuth, A., Potvin, C., Ransijn, J., Reid, D., Ribeiro, S.C., Roberts, S.D., Rodríguez, R., Saldaña-Acosta, A., Santa-Regina, I., Sasa, K., Selaya, N.G., Sillett, S.C., Sterck, F., Takagi, K., Tange, T., Tanouchi, H., Tissue, D., Umehara, T., Utsugi, H., Vadeboncoeur, M.A., Valladares, F., Vanninen, P., Wang, J.R., Wenk, E., Williams, R., de Aquino Ximenes, F., Yamaba, A., Yamada, T., Yamakura, T., Yanai, R.D., York, R.A., 2015. BAAD: a Biomass And Allometry Database for woody plants. Ecology 96, 1445–1445. doi:10.1890/14-1889.1

Goldberger, A.S., 1968. The Interpretation and Estimation of Cobb-Douglas Functions. Econometrica 36, 464–472. doi:10.2307/1909517

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., Penman, J., 2017. The key role of forests in meeting climate targets requires science for credible mitigation. Nat. Clim. Chang. 7, 220–226. doi:10.1038/nclimate3227

IPCC, 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Institute for Global Environmental Strategies.

Jia, Q., Liu, Q., Li, J., 2015. Individual-based fine root biomass and its functional relationship with leaf for Pinus tabuliformis in northern China. Eur. J. For. Res. 134, 705–714. doi:10.1007/s10342-015-0884-0

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A., 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. Glob. Chang. Biol. 23, 177–190. doi:10.1111/gcb.13388

Kerkhoff, A.J., Enquist, B.J., 2009. Multiplicative by nature: Why logarithmic transformation is necessary in allometry. J. Theor. Biol. 257, 519–521. doi:10.1016/j.jtbi.2008.12.026

Koch, G.W., Sillett, S.C., Jennings, G.M., Davis, S.D., 2004. The limits to tree height. Nature 428, 851–854. doi:10.1038/nature02417

Marziliano, P.A., Lafortezza, R., Medicamento, U., Lorusso, L., Giannico, V., Colangelo, G., Sanesi, G., 2015. Estimating belowground biomass and root/shoot ratio of Phillyrea latifolia L. in the Mediterranean forest landscapes. Ann. For. Sci. 72, 585–593. doi:10.1007/s13595-015-0486-5

McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., Westfall, J.A., 2016. Hybrid estimators for mean aboveground carbon per unit area. For. Ecol. Manage. 378, 44–56. doi:10.1016/J.FORECO.2016.07.007

McRoberts, R.E., Moser, P., Zimermann Oliveira, L., Vibrans, A.C., 2015. A general method for assessing the effects of uncertainty in individual-tree volume model predictions on large-area volume estimates with a subtropical forest illustration. Can. J. For. Res. 45, 44–51. doi:10.1139/cjfr-2014-0266

Mersmann, O., Trautmann, H., Steuer, D., Bornkamp, B., 2018. truncnorm: Truncated Normal Distribution.

Morhart, C., Sheppard, J., Spiecker, H., 2013. Above Ground Leafless Woody Biomass and Nutrient Content within Different Compartments of a P. maximowicii × P. trichocarpa Poplar Clone. Forests 4, 471–487. doi:10.3390/f4020471

Morhart, C., Sheppard, J.P., Schuler, J.K., Spiecker, H., 2016. Above-ground woody biomass allocation and within tree carbon and nutrient distribution of wild cherry (Prunus avium L.) – a case study. For. Ecosyst. 3. doi:10.1186/s40663-016-0063-x

Mosseler, A., Major, J.E., Labrecque, M., Larocque, G.R., 2014. Allometric relationships in coppice biomass production for two North American willows (Salix spp.) across three different sites. For. Ecol. Manage. 320, 190–196. doi:10.1016/j.foreco.2014.02.027

Packard, G.C., 2012. Is non-loglinear allometry a statistical artifact? Biol. J. Linn. Soc. 107, 764–773. doi:10.1111/j.1095-8312.2012.01995.x
Packard, G.C., Boardman, T.J., 2008. Model selection and logarithmic transformation in allometric analysis. Physiol. Biochem. Zool. 81, 496–507. doi:10.1086/589110
Pajtík, J., Konôpka, B., Lukac, M., 2008. Biomass functions and expansion factors in young Norway spruce (Picea abies [L.] Karst) trees. For. Ecol. Manage. 256, 1096–1103. doi:10.1016/j.foreco.2008.06.013
Picard, N., Boyemba Bosela, F., Rossi, V., 2015. Reducing the error in biomass estimates strongly depends on model selection. Ann. For. Sci. 72, 811–823. doi: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. FAO and CIRAD, Rome, Italy, and Montpellier, France.
R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Roxburgh, S.H., Paul, K.I., Clifford, D., England, J.R., Raison, R.J., 2015. Guidelines for constructing allometric models for the prediction of woody biomass: How many individuals to harvest? Ecosphere 6, 1–27. doi:10.1890/ES14-00251.1
RStudio Team, 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.
Sanquetta, C.R., Dalla Corte, A.P., Behling, A., de Oliveira Piva, L.R., Péllico Netto, S., Rodrigues, A.L., Sanquetta, M.N.I., 2018. Selection criteria for linear regression models to estimate individual tree biomasses in the Atlantic Rain Forest, Brazil. Carbon Balance Manag. 13, 25. doi:10.1186/s13021-018-0112-6
Schepaschenko, D., Shvidenko, A., Usoltsev, V., Lakyda, P., Luo, Y., Vasylyshyn, R., Lakyda, I., Myklush, Y., See, L., McCallum, I., Fritz, S., Kraxner, F., Obersteiner, M., 2017. A dataset of forest biomass structure for Eurasia. Sci. Data 4, 170070. doi:10.1038/sdata.2017.70
Stephenson, N.L., Das, A.J., Condit, R., Russo, S.E., Baker, P.J., Beckman, N.G., Coomes, D.A., Lines, E.R., Morris, W.K., Rüger, N., Álvarez, E., Blundo, C., Bunyavejchewin, S., Chuyong, G., Davies, S.J., Duque, Á., Ewango, C.N., Flores, O., Franklin, J.F., Grau, H.R., Hao, Z., Harmon, M.E., Hubbell, S.P., Kenfack, D., Lin, Y., Makana, J.-R., Malizia, A., Malizia, L.R., Pabst, R.J., Pongpattananurak, N., Su, S.-H., Sun, I.-F., Tan, S., Thomas, D., van Mantgem, P.J., Wang, X., Wiser, S.K., Zavala, M.A., 2014. Rate of tree carbon accumulation increases continuously with tree size. Nature 507, 90–93. doi:10.1038/nature12914
Ung, C.H., Lambert, M.C., Raulier, F., Guo, X.J., Bernier, P.Y., 2017. Biomass of trees sampled across Canada as part of the Energy from the Forest Biomass (ENFOR) Program. doi:https://doi.org/10.23687/fbad665e-8ac9-4635-9f84-e4fd53a6253c
Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S, 4th ed. Springer, New York.
Walther, B.A., Moore, J.L., 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography (Cop.). 28, 815–829. doi:10.1111/j.2005.0906-7590.04112.x
Xiao, X., White, E.P., Hooten, M.B., Durham, S.L., 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology 92, 1887–1894. doi:10.1890/11-0538.1
Zianis, D., 2008. Predicting mean aboveground forest biomass and its associated variance. For. Ecol. Manage. 256, 1400–1407. doi:10.1016/J.FORECO.2008.07.002
Zianis, D., Mencuccini, M., 2004. On simplifying allometric analyses of forest biomass. For. Ecol. Manage. doi:10.1016/j.foreco.2003.07.007

Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and stem volume equations for tree species in Europe. Finnish Society of Forest Science, Finnish Forest Research Institute.

Anexa 6. A template of data for PREBAS calibration and application

1 Site description data

1.1 Required variables

Table 1.1 Variables of the site summary information

Names	Unit	Description
SiteID	-	SiteID is for identifying the plot of the forest.
ClimateID	-	ClimateID is for identifying the regions. Several sites might belongs to a same ClimateID, which means that they share the same weather condition.
Latitude	degree	Latitude of the plot in decimal unit, WGS84 (World Geodetic System 1984).
Longitude	degree	Longitude of the plot in decimal unit, WGS84 (World Geodetic System 1984).
Elevation	m	The elevation of the site. This variable is optional. If possible, providing the aspect and slope of the site will also be helpful in checking data and model output.
SoilType	-	Classification based on soil textures. For instance, sand, loam, light clay, etc.
SoilDepth	mm	Thickness of soil or ecosystem rooting depth.
FieldCapacity	mm	Soil property. Field Capacity is the amount of soil moisture or water content held in the soil after excess water has drained away and the rate of downward movement has decreased. The value range is 0 to1000.
WiltingPoint	mm	Soil property. Permanent wilting point or wilting point is defined as the minimal amount of water in the soil that the plant requires not to wilt. The value range is 0 to 1000.
SiteType	-	Classification based on site fertility. This column can be replaced by site index, site class, site form, or any other phytocentric and geocentric indicators of forest site productivity. If using site index, please indicate the reference age by changing the name of the variable. For instant, 'Hdom_100' means the dominant height at age 100.

PREBAS do not require Longitude and Latitude as inputs. However, the location information is essential in collecting useful data and information from other databases in both model calibration and application.

SoilType is used for gap-filling and validating the FieldCapacity and WiltingPoint records. FieldCapacity and WiltingPoint can be estimated based on the soil texture (SoilType).

1.2 Data format

Site description data should be provided in format of tables like csv files (comma delimited). Below an example of the site description table ：

SiteID	ClimateID	Latitude	Longitude	SoilType	Soil Depth	FieldCapacity	Wilting Point	Site Type
1	1	39.33902	-9.21183	Loamy Sand	1275	0.25	0.15	2
2	1	39.33902	-9.21183	Loamy Sand	1275	0.25	0.15	2
3	1	39.33891	-9.22342	Sand Loam	1275	0.305	0.18	2
4	1	39.33891	-9.22342	Sand Loam	1275	0.305	0.18	3
5	1	39.33891	-9.22342	Sand Loam	1275	0.305	0.18	2
…	…	…	…	…	…	…	…	…
19	14	39.314407	-8.909976	Sand Loam	1275	0.305	0.18	1
20	14	39.314329	-8.92157	Sand Loam	1087.5	0.305	0.18	2

2 Weather data

2.1 Required variables

Table 2.1 Variables of the weather input for PREBAS

Names	**Unit**	**Description**
ClimateID	-	ClimateID is for identifying the regions. Several sites might belongs to a same ClimateID, which means that they share the same weather condition. (Same with Table 1.1)
Year	-	Date was separated into Year, Month, Day because the data format for different operation systems could largely differ.
Month	-	-

Day	-	-
PAR	mol PPFD m-2 d-1	Daily sum of photosynthetic photon flux density above the canopy.
TAir	°C	Daily mean air temperature
VPD	kPa	Daily mean vapour-pressure deficit
Precip	mm d-1	Daily sum of precipitation
CO2	ppm	Daily mean CO2 concentration. If this column is missing, PREBAS will use the global average daily value.

PAR (daily sums of photosynthetically active radiation) is seldom provided in global climate databases. However, it can be easily calculated from solar radiation (shortwave radiation) from established empirical relationships. The ratio of PAR to broad-band solar radiation varies from 0.4 to 0.6, and is nominally taken to be 0.44 or 0.5 when no local data for validation. Most meteorological datasets include solar radiation measurements.

2.2 Data format

Weather inputs should be provided in format of tables like csv files (comma delimited) or data.table objects in R. If many regions include long duration of the record and the combined file has millions of rows, we suggest to make each climate ID as an independent file. Then name the files in a uniform and explicit form. For instance, “ClimateID_1_1970_2005.csv” means that the climateID is 1, and observations include the years from 1970 to 2005. Below an example of the weather input table :

ClimateID	Year	Month	Day	PAR	TAir	VPD	Precip	CO2
1	1970	1	1	28.23	19.83	1.04	0	325.04
1	1970	1	2	28.77	19.41	1.12	10	325.04
1	1970	1	3	28.81	16.99	1.01	0	325.04
1	1970	1	4	16.95	17.40	0.97	0,2	325.04
...	…	…	…	…	…	…	…	…
1	2005	12	30	28.77383	19.52	1.14	0	380.9
1	2005	12	31	29.14447	21.015	1.28	0	380.9

3 Forest inventory data

3.1 Required variables

Based on the stand structure, PREBAS simulates forest dynamics at stand-level or layer-level (size-class) level. Thus, simulations of pure even-aged forest require stand average information. For the forest with mixed tree species or multiple layers, the average information for each layer or species is required.

Table 3.1 Variables of forest inventory data for PREBAS

Names	Unit	Description
SiteID	-	Identifying the plot. (Same with Table 1.1).
Year	-	The year when the forest inventory was implemented.
Rotation	-	Identifying coppice by Indicating which rotation the record belongs. 1 = first rotation, 2 = the second rotation.
Thinning	-	NoThin = No thinning was implemented this year; BeforeThin = Thinning was implemented this year and this record is the measurement before thinning; AfterThin= Thinning was implemented this year and this record is the measurement after thinning
nLayers	-	Number of layers in the plot. (Same with Table 1.1)
Layer	-	Identifying which layer this row belong. 1 = the 1st layer, 2 = the 2nd layer, etc. For even-aged pure forest, both nLayers and Layer equal 1.
Species	-	Tree species of this layer.
Age	yr	Average age of trees in this layer.
Height	m	Average height of trees in this layer.
DBH	cm	Average DBH (Diameter at Breast Height) of trees in this layer.
BasalArea	m^2 ha^{-1}	Total basal area of trees in this layer.
Density	ha^{-1}	Number of trees in this layer.
CrownBaseH	m	Average height of the crown base in this layer.
CrownWidth	m	Average crown width in this layer.

CrownLength	m	Average crown length in this layer.
Volume	m^3 ha^{-1}	Layer volume in this layer.
W_Stem	kg DM ha^{-1}	Stem biomass in this layer.
W_Foliage	kg DM ha^{-1}	Foliage biomass in this layer.
W_Branch	kg DM ha^{-1}	Branch biomass in this layer.
W_FineRoot	kg DM ha^{-1}	Fine root biomass in this layer.
W_CoarseRoot	kg DM ha^{-1}	Coarse root biomass in this layer.

Age, Height, DBH, and all the others variables concerned in the table are the average of the layer or size-class. For even-aged forests, the whole stand is referred as one layer. The variable can be estimated by choosing the medium tree of the layer, or by taking the basal-area-weighted average of all the trees in the layer. For natural uneven-aged forests with mixed species and complex structures, individual-tree level measurements are also needed.

Biomass information are only used in PREBAS calibration. After the model being calibrated, the application requires only basic inventory variables such as Height, DBH, and Density.

Forest inventory might exclude biomass investigation. Thus, destructive sample data are needed as described in section 4.1.

3.2 Data format

Forest inventory data should be provided in format of table like csv files (comma delimited). Below an example of the forest inventory table：

SiteID	Year	Rotation	Thinning	nLayers	Layer	Species	Age	Height	DBH	Basal Area	…	WneR	W_CoarseRoot
1	1970	1	NoThin	1	1	Eucalyptus globulus	4	10.4	7.8	5.174	…	776	1708
1	1971	1	NoThin	1	1	Eucalyptus globulus	5	12.5	9.4	7.457	…	1150	2532
1	1972	1	NoThin	1	1	Eucalyptus globulus	6	14.1	11.1	10.402	…	1686	3708
1	1973	1	NoThin	1	1	Eucalyptus globulus	7	15.4	13	13.987	…	2424	5332
1	1974	1	NoThin	1	1	Eucalyptus globulus	8	16.5	14	16.168	…	3120	6866
1	1975	1	NoThin	1	1	Eucalyptus globulus	9	17.1	15.2	18.968	…	4374	9622
1	1976	1	NoThin	1	1	Eucalyptus globulus	10	18.6	16	21.213	…	5890	12956

…	…	…	…	…	…	…	…	…	…	…	…	…	…
20	2002	1	NoThin	1	1	Eucalyptus globulus	35	31.6	25.4	55.016	...	38328	84322

4 Additional useful data

4.1 Destructive sample data

Destructive sample data here means individual-tree level biomass data. The information are essential for PREBAS calibration. Destructive sample data can be an independent dataset, but basic site information are still needed.

Table 4.1 Variables of destructive sample data. The default unit of the biomasses is kg dry matter (DM) per tree (kg DM).

Names	**Unit**	**Description**
D	cm	Diameter at breast height.
H	m	Tree height.
Hc	m	Height of the Crow base
Cw	m	Crown width
$\mathbf{A_c}$	m^2	Cross-sectional area at crown base.
$\mathbf{W_{Stem}}$	kg DM	Stem biomass
$\mathbf{W_{Foliage}}$	kg DM	Foliage biomass
$\mathbf{W_{Branch}}$	kg DM	Live branch biomass
$\mathbf{W_{FineRoot}}$	kg DM	Fine root biomass
$\mathbf{W_{CoarseRoot}}$	kg DM	Coarse root biomass

4.2 Classification of site fertility

The suitable method of site evaluation varies with tree species and regions. When the phytocentric and geocentric indicators of forest site productivity is missing in Table 1.1. Please provide Age and Height information of the dominant trees for each plot. Then the site index can be calculated.

4.3 Eddy covariance data

Eddy covariance data are required for the calibration of PREBAS. Although many global eddy covariance network are providing open access data, those free datasets only cover limited tree species and regions. Thus, eddy covariance data could be considered as optional depending on the tree species and regions.

Table 4.2 Data requirement for the eddy covariance site. (Shading means same variables with previous tables)

Variable	**Abbreviation**	**unit**	**time step**	**Data type**
Soil Data				
soil depth	SoilDepth	mm	-	Site-specific
field capacity	FieldCapacity	mm	-	Site-specific
wilting point	WiltingPoint	mm	-	Site-specific
Soil water content	-	mm	Daily	Measurement
Canopy Information				
Fraction of Absorbed Photosynthetically Active Radiation	f_{APAR}	-	Daily or Yearly	light interception
Meteorological Data				
photosynthetic photon flux density	PAR	mol PPFD m^{-2} d^{-1}	Daily	weather
Air Temperature	TAir	°C	Daily	weather
Vapour pressure deficit	VPD	kPa	Daily	weather
Precipitation	Precip	mm	Daily	weather
Flux Data				
Gross primary production	GPP	g C m^{-2}	Daily	Eddy Tower
Evapotranspiration	ET	mm	Daily	Eddy Tower
Quality Flag	-	%	Daily	Eddy Tower

Extra Information could be useful, including 1) Forest inventory data of the site (remeasurements of DBH, basal area, height, etc), 2) Soil or canopy nitrogen information, e.g. C/N , 3) Shrubs and ground vegetation, e.g. LAI, chamber measurements.

f_{APAR} is either measured or calculated based on LAI (leaf area index). It changes with canopy growth or thinnings. Quality Flag is assigned to each day to indicate percentage of measured (non-gapfilled) and good quality gap-filled half-hourly data used to calculate the daily value. We prefer the nighttime partitioning method for GPP records.

4.4 Soil carbon storage

PREBAS can link the soil carbon model Yasso to simulate the dynamics of soil carbon, and also the ecosystem carbon fluxes. In this case, the information about soil carbon storage of the stand is needed. The data are optional because it’s difficult to obtain.