

1 Supplementary material for “Modelling forest management within a global vegetation
2 model – Part 2: model validation”

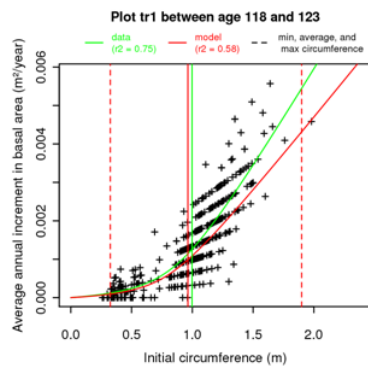
3 **1 Estimation method for variables not directly available** 4 **from the raw data**

5 Volume, standing biomass, and tree height are estimated from circumference
6 measurements using the default allometric rules of ORCHIDEE-FM. For volume
7 increment, we use the increment-specific biomass expansion factors of IPCC (2003). All
8 equations and parameters values are fully described in Bellassen et al. (this issue).
9 Average yearly increase in aboveground woody biomass between two measurements n
10 and $n+1$ can be estimated from Eq. S1:

$$11 \text{ wood}_{inc} = \frac{\sum_{i=1}^{ntrees} (\text{biomass}_i(n+1) - \text{biomass}_i(n))}{(\text{year}(n+1) - \text{year}(n)) \times \text{plot_size}} + \varepsilon \quad (\text{S1})$$

12 where wood_{inc} is the yearly increase in aboveground woody biomass between
13 measurement n at $\text{year}(n)$ and measurement $n+1$ at $\text{year}(n+1)$ in $\text{gC}\cdot\text{m}^{-2}\cdot\text{yr}^{-1}$; $ntrees$ is the
14 number of trees still alive at measurement $n+1$; $\text{biomass}_i(n+1)$ is the aboveground woody
15 biomass of tree i at $\text{year}(n+1)$ in gC ; and plot_size is the size of the plot in m^2 . ε is the
16 increase in aboveground woody biomass of the trees that died between the two
17 measurements. A few percent (at most 15%) of trees commonly die between two
18 measurements. They are usually smaller trees and not they do not all die right before the
19 second measurement. Therefore, ε is much smaller than the wood increment of the trees
20 that survived, and it is neglected in the calculations.

21 Finally, the average individual tree growth indicators σ and γ between two measurements
 22 can be estimated by fitting Equation 1 to the data using the Gauss-Newton non-linear
 23 algorithm. In other words, this estimation of σ and γ corresponds to a calibration of the
 24 growth individuation part of the FMM (distribution of NPP to each tree) on the
 25 permanent plot data.
 26 This estimation of the individual tree growth parameters σ and γ is illustrated by



27
 28 Figure S 1. The fit of Equation 1 on the data is generally satisfactory ($r^2 = 0.75$ in this
 29 example) and confirms the relevance of this equation which was established on similar
 30 data by Deleuze et al. (2004). Many observed data points are lined up, forming upward
 31 diagonal lines that reflect the limits in the tree circumference measuring accuracy (0.01
 32 m): these data points with different initial circumference have the same measured
 33 increase in circumference. Their difference in basal area increase is, therefore,
 34 proportional to their initial circumference (see Eq. S 2).

$$35 \quad \delta ba_i = \frac{\delta circ_i^2 + 2 \times circ_i \times \delta circ_i}{4\pi} \quad (S 2)$$

36 where δba_i and $\delta circ_i$ are, respectively, the increment in basal area and circumference of
 37 tree i.

38 **2 Uncertainty associated with validation data**

39 Three sources of uncertainty undermine these validation datasets: measurement
40 uncertainty, allometric uncertainty and sampling uncertainty (Phillips et al., 2000). All of
41 the following estimates for uncertainty relate to a 90% or greater confidence interval.

42 ✓ The measurement uncertainty comes from errors in the *in situ* measurement of
43 tree characteristics such as diameter or height, with 0.5% and 10% uncertainty
44 at tree level, respectively (USDA, 2007). This source of uncertainty applies to all
45 datasets. It is intrinsically negligible for diameter and becomes negligible for
46 height because we only use averages at plot or regional scales for permanent
47 plots and yield tables, respectively.

48 ✓ Allometric uncertainty relates to the discrepancy between a variable such as
49 volume predicted by an allometric model and the actual value. This discrepancy
50 can typically vary between 15% and 30% at tree level depending on the local
51 relevance of the allometric model (Dhote et al., 2000; Zianis and Mencuccini,
52 2004; Newton and Amponsah, 2007). This source of uncertainty does not apply
53 to permanent plots because the allometric model is the same for the *in situ*
54 estimate and ORCHIDEE-FM. Assuming that the allometric model is not
55 consistently biased, this uncertainty is reduced to below 5% when averaging at a
56 plot or regional scale for the NFI and yield tables, respectively. This uncertainty
57 estimate does not take into account the error arising from the use of a biomass
58 expansion factor (BEF), which is necessary to compare the commercial volume
59 of the NFI dataset to the total aboveground volume simulated by ORCHIDEE-FM.

60 ✓ Sampling uncertainty arises at two levels. First, the sampled trees may not be
61 representative of their plot, and second, the sample plot may not be
62 representative of its region. The first level can be neglected due to the elaborate
63 design of forest inventory sampling schemes (IFN, 2006). However, the second
64 level can be quite high: although the regional average of aboveground biomass
65 for 400,000 hectares of forested land is estimated to have only a 6% uncertainty
66 (EPA, 2010), the sampling error of each plot is closer to 60% (Waggoner, 2009;
67 Bellassen et al., submitted). Assuming a Gaussian structure of these sampling
68 errors, uncertainty is reduced to 20% when using averages of 10 NFI plots.

69 Overall, the uncertainty related to datasets is low and is hereafter neglected in this
70 study. The only exception is the maps of standing volume and volume increment
71 interpolated from the NFI dataset, which carry a 20% uncertainty.

72 **3 Stand-scale vs. tree-scale: which type of variable for** 73 **validation?**

74 ***3.1 Stand-scale volume increment vs. tree-scale basal area*** 75 ***increase***

76 NFI campaigns provide measured tree-scale variables such as tree basal area growth over
77 the past five years, which can be directly compared to their counterparts simulated by
78 ORCHIDEE-FM. The campaigns also estimate stand-scale variable such as volume
79 increment by aggregating and extrapolating the tree-scale measures. These estimates can
80 also be compared to their simulated counterparts. For validating a model representing the

81 “average stand” over a 0.25° resolution, the most suitable variables would be
 82 homogeneous at this spatial scale.
 83 To determine the most homogeneous type of variables between the tree-scale tree
 84 circumference growth and the stand-scale volume increment, we applied a 0.25°
 85 resolution grid to the French territory and selected the pixels containing more than 20
 86 measured plots. A perfectly homogeneous pixel would have a single volume increment
 87 value \bar{V} . The tree basal area growth would still be variable because it is a function of
 88 both the plot conditions and the initial tree circumference (see Equation S 3 from Deleuze
 89 (2004)). However, the plot-specific σ , γ , and m parameters of Equation S 3 would be
 90 fixed, yielding one theoretical value per tree circumference \overline{ba}_i .

$$91 \quad \delta ba_i = \frac{\gamma}{2} \times \left(circ_i - m\sigma + \sqrt{(m\sigma + circ_i)^2 - 4m\sigma \times circ_i} \right) \quad (S\ 3)$$

92 where δba_i is the annual increase in the basal area of tree i in square meters; $circ_i$ is the
 93 circumference of tree i in meters; and σ , γ , and m are parameters.
 94 \bar{V} is estimated by the pixel average for the volume increment and \overline{ba}_i by fitting Equation
 95 S 3 for each pixel, using the assumption that all of the trees of one pixel belong to a
 96 single homogeneous plot. The average relative distances S and T to the theoretical
 97 variable value is then determined by Equations S 4 and S 5 for the stand-scale volume
 98 increment and the tree-scale basal area growth, respectively.

$$99 \quad S = \frac{1}{n-1} \sum_j \frac{V_j - \bar{V}}{\bar{V}} \quad (S\ 4)$$

100 where S is the pixel average relative distance to the theoretical value of the volume
 101 increment, n is the number of plots in the pixel; V_j is the volume increment of plot j and
 102 \bar{V} is the theoretical value of the volume increment under the homogeneity assumption.

$$103 \quad T = \frac{1}{n-1} \sum_j \left[\frac{1}{m} \sum_i \frac{ba_{ij} - \overline{ba_{ij}}}{\overline{ba_{ij}}} \right] \quad (S 5)$$

104 where T is the pixel average relative distance to the theoretical value of the volume
 105 increment; n is the number of plots in the pixel, ba_{ij} is the basal area growth of tree i in
 106 plot j and $\overline{ba_{ij}}$ is the theoretical value of the basal area growth of tree i in plot j under the
 107 homogeneity assumption.

108 For each pixel, the $\frac{S}{T}$ ratio thus determines which variable is closest to the homogeneity
 109 assumption. If $\frac{S}{T} < 1$, then the stand-scale variable is more homogeneous than the tree-
 110 scale variable.

111 **3.2 Stand-scale variables: volume or basal area?**

112 Having decided on stand-scale variables (because $\frac{S}{T} < 1$), the options available for
 113 validation are: the volume increment and the basal area increment. In contrast to the basal
 114 area increment, the volume increment is not a directly measured variable on NFI plots: it
 115 requires the use of NFI allometric equations. However, the volume increment is a more
 116 direct output of ORCHIDEE-FM than the basal area increment, which rests heavily on
 117 the model's ability to simulate tree circumference distribution. Indeed, we find that the
 118 model simulates distributions that are more even than real distributions, often to the
 119 benefit of smaller circumference classes. For a given stand volume increment, the

120 increment in stand basal area simulated by the model is therefore consistently
121 overestimated (data not shown). Thus, we chose to use stand volume increment for
122 validation.

123 Interestingly enough, the bias between the stand volume increment and the stand basal
124 area increment is systematic. This systematic bias means that when there is doubt about
125 the allometries used in forest inventories, it is still of interest to compare them to
126 ORCHIDEE-FM. For these comparisons however relative regional trends will be more
127 trustworthy than absolute simulated values.

128 **4 “Shadow models”**

129 **4.1 PP_f simulation**

130 **4.1.1 Stand-scale variables**

131 For stand-scale variables (tree density, basal area, standing volume, etc.), the “shadow
132 model” uses three input variables:

$$133 \quad mes(i) = a + b \times vol_{tot}(i) + c \times med_{circ}(i) + d \times rdi_{target}(i) + \varepsilon(i) \quad (S\ 6)$$

134 where mes is the measured variable of interest; vol_{tot} is the total volume of the stand at
135 the last measurement; med_{circ} is the median circumference of the stand at the first
136 measurement; rdi_{target} is the post-thinning relative density index; a , b , c , and d are
137 regression coefficients; and i is the permanent plot number.

138 **4.1.2 Tree-scale variables**

139 For the two tree-scale variables σ and γ , the current volume increment is added to the
140 “shadow model”:

141
$$mes(i) = a + b \times vol_{tot}(i) + c \times med_{circ}(i) + d \times rdi_{target}(i) + e \times vol_{inc}(i) + \varepsilon(i) \quad (S\ 7)$$

142 where *mes* is the measured variable of interest; vol_{tot} is the total volume of the stand at the
143 last measurement; med_{circ} is the median circumference of the stand at the first
144 measurement; rdi_{target} is the post-thinning relative density index; vol_{inc} is the volume
145 increment during the last measurement period; a, b, c, d, and e are regression coefficients;
146 and i is the permanent plot number.

147 **4.2 PP_{ic} simulation**

148 For the PP_{ic} simulation, the “shadow model” uses two input variables:

149
$$mes(i) = a + b \times vol_{tot}(i) + c \times rdi_{target}(i) + \varepsilon(i) \quad (S\ 8)$$

150 where *mes* is the measured variable of interest; vol_{tot} is the total volume of the stand at the
151 last measurement; rdi_{target} is the post-thinning relative density index; a, b, and c are
152 regression coefficients; and i is the permanent plot number.

153 **4.3 YT_f simulation**

154 The “shadow model” of YT_f uses the only input variable of ORCHIDEE-FM, the total
155 volume:

156
$$mes(i) = a + b \times vol_{tot}(i) + \varepsilon(i) \quad (S\ 9)$$

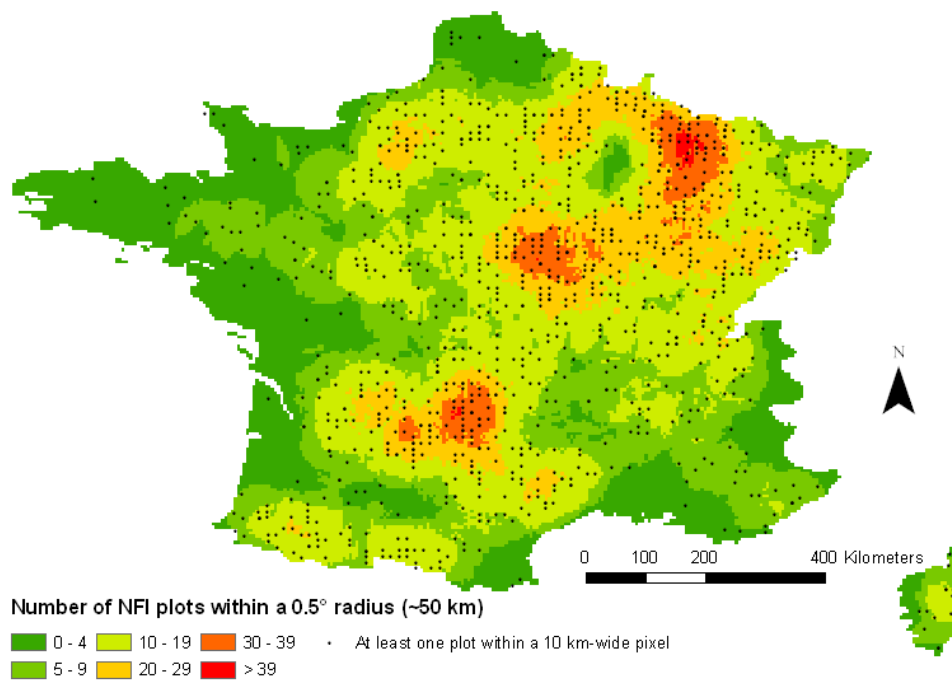
157 where *mes* is the measured variable of interest; vol_{tot} is the total volume of the stand at the
158 last measurement; a and b are regression coefficients; and i is the permanent plot number.

159 **5 PP_{ic} simulation – Detailed model-data comparison**

160 See Figure S 2.

161 **6 Datasets and their use**

162 See Figure S 3, Figure S 4, Figure S 5,



163

164 Figure S 6, Figure S 7, Figure S 8 and Figure S 9.

165 **7 R outputs**

166 **7.1 Linear mixed model of equation 4**

167 $vol_{tot}(i, j, k) = \alpha + \beta_i + \gamma_j + \varepsilon(i, j, k)$ (1)

168 where α is the intercept, β_i and γ_j are respectively the coefficients associated with plant
 169 functional type (PFT) i and country j , and $vol_{tot}(i,j,k)$, and $\varepsilon(i,j,k)$ are respectively the
 170 total volume produced at year 80 and the error term associated with yield table k of PFT i
 171 and country j .

```
172 lm_voltot<-lme(vol_tot~Country+PFT,data=yt100a,random=~0|old_ID)
173 anova(lm_voltot)
174           numDF denDF F-value p-value
175 (Intercept)  1  697 1359.4204 <.0001
176 Country      21  162  6.1517 <.0001
177 PFT          1  162 47.6980 <.0001
178 r2<-1-sum((yt100a$vol_tot-fitted(lm_voltot))^2)/sum((yt100a$vol_tot-mean(yt100a$vol_tot,na.rm=T))^2)
179 r2
180 [1] 0.6446316
```

181 **7.2 Linear mixed model of equation 5**

$$182 \text{dens}(i, j, k) = \alpha + \beta_i + \gamma_j + \delta \times vol_{tot}(i, j, k) + \varepsilon(i, j, k) \quad (2)$$

183 where α is the intercept, β_i , γ_j and δ are respectively the coefficients associated with plant
 184 functional type (PFT) i , country j , and total volume produced at year 80, $dens(i,j,k)$,
 185 $vol_{tot}(i,j,k)$ and $\varepsilon(i,j,k)$ are respectively the density, the total volume produced at year 80
 186 and the error term associated with yield table k of PFT i and country j .

```
187 lm_dens<-lme(dens~Country+PFT+vol_tot,data=yt100a2,random=~0|old_ID)
188 anova.lme(lm_dens)
189           numDF denDF F-value p-value
190 (Intercept)  1  676 456.7740 <.0001
191 Country      21  145  5.0263 <.0001
192 PFT          1  145  6.7864 0.0101
193 vol_tot      1  676 336.1248 <.0001
194 r2<-1-sum((yt100a2$dens-fitted(lm_dens))^2)/sum((yt100a2$vol_tot-mean(yt100a2$vol_tot,na.rm=T))^2)
195 r2
196 [1] 0.4674039
```

197 **7.3 Example of shadow model (equation 8)**

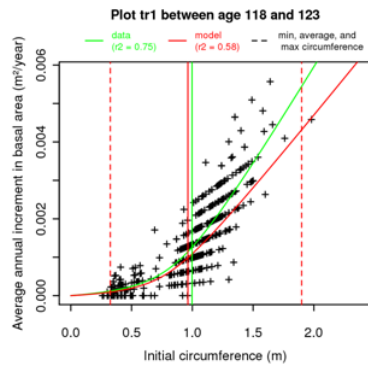
198 $mes(i) = a \times vol_{tot}(i) + b \times med_{circ}(i) + c \times rdi_{target}(i) + \varepsilon(i)$ (S 10)

199 where *mes* is the measured variable of interest (eg. tree density, standing volume, ...),
200 *vol_{tot}* is the total volume of the stand at the last measurement, *med_{circ}* is the median
201 circumference of the stand at the first measurement, *rdi_{target}* is the post-thinning relative
202 density index, *a*, *b*, *c*, and *d* are regression coefficients, and *i* is the permanent plot
203 number.

```
204 # Dataset split in half: sub2 for calibration, sub3 for validation
205 lm_data_sub2<-subset(stats_all,stats_all$ID<=30)
206 lm_data_sub3<-subset(stats_all,stats_all$ID>30)
207 simple_lm<-
208 lm(lm_data_sub2[,var_no]~lm_data_sub2$vol_totth+lm_data_sub2$med_circ_init+lm_data_sub2$rdi)
209 fix_coef<-coefficients(simple_lm)
210 gamma_fitted<-fix_coef[1]+fix_coef[2]*lm_data_sub3$vol_totth+
211 fix_coef[3]*lm_data_sub3$med_circ_init+fix_coef[4]*lm_data_sub3$rdi
212 r2_stat[which(y_ab==var)]<-1-sum((gamma_fitted-
213 lm_data_sub3[,var_no])^2)/sum((lm_data_sub3[,var_no]-mean(lm_data_sub3[,var_no]))^2)
214 ab_stat[which(y_ab==var)]<-mean((gamma_fitted-lm_data_sub3[,var_no])/
215 abs(lm_data_sub3[,var_no]))
```

217

218 **Figure captions**

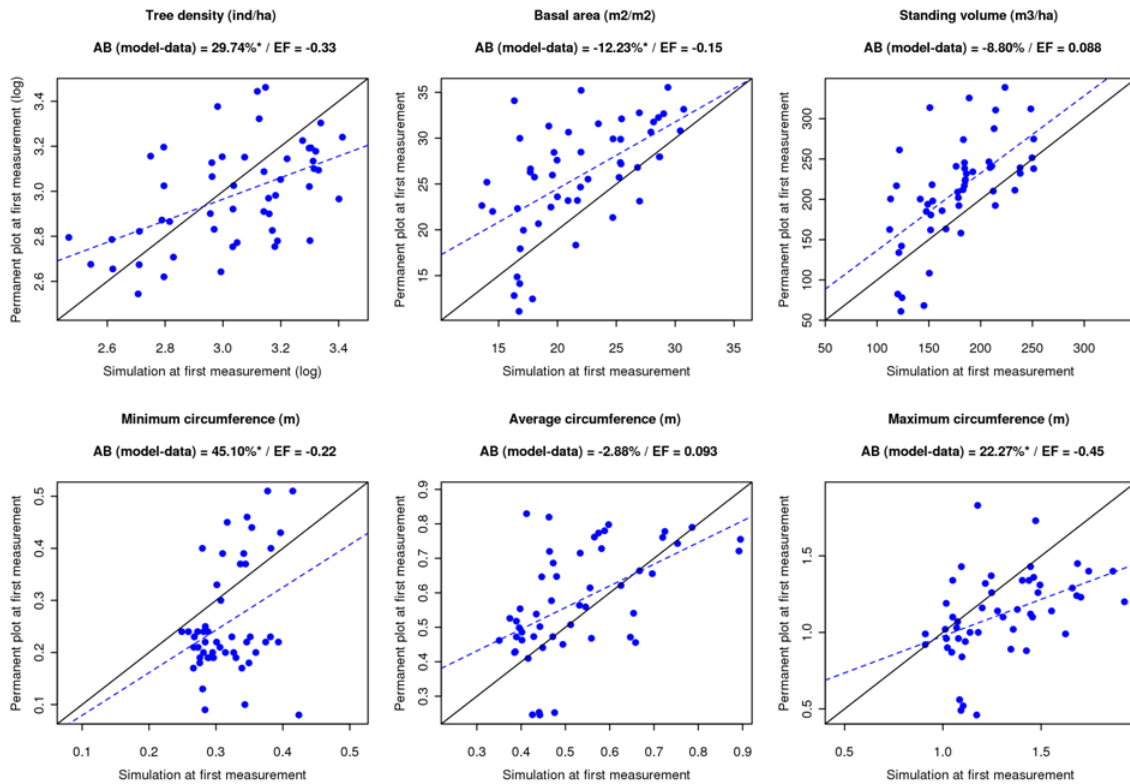


219

220 Figure S 1. Validation method for individual tree growth – example of plot 56 at last
221 measurement

222 Each data point (cross) represents the average annual increment in basal area of one tree
223 between the last two consecutive measurements (age 118 and age 123). Equation 1 is
224 fitted on this data set to estimate σ and γ , respectively the threshold and slope of increase
225 in basal area as a function of circumference (green curve). The simulated tree growth
226 curve (red curve) is drawn from the average simulated σ and γ for the corresponding
227 period of simulation 1. Vertical lines mark the average circumference of the data set (full
228 green), of the simulation (full red), and the minimum and maximum simulated
229 circumferences (dotted red).

230



231

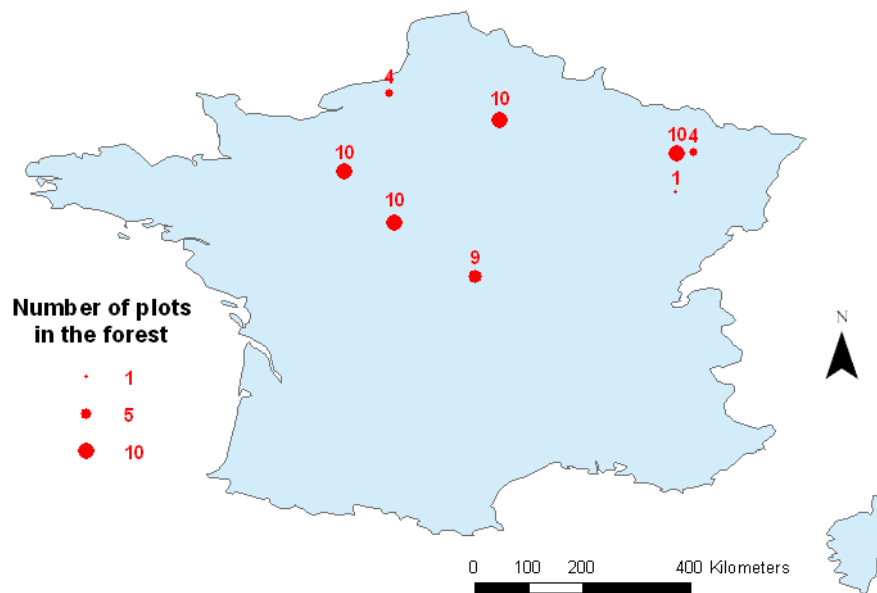
232 Figure S 2. Validation of stand characteristics – PP_{ic} simulation

233 Each blue dot corresponds the state of one permanent plot at its last measurement. The

234 dotted blue line represents their linear regression. AB and EF are respectively average

235 relative bias and model efficiency. A “*” indicates that AB is significantly different from

236 0 with a 95% confidence interval.



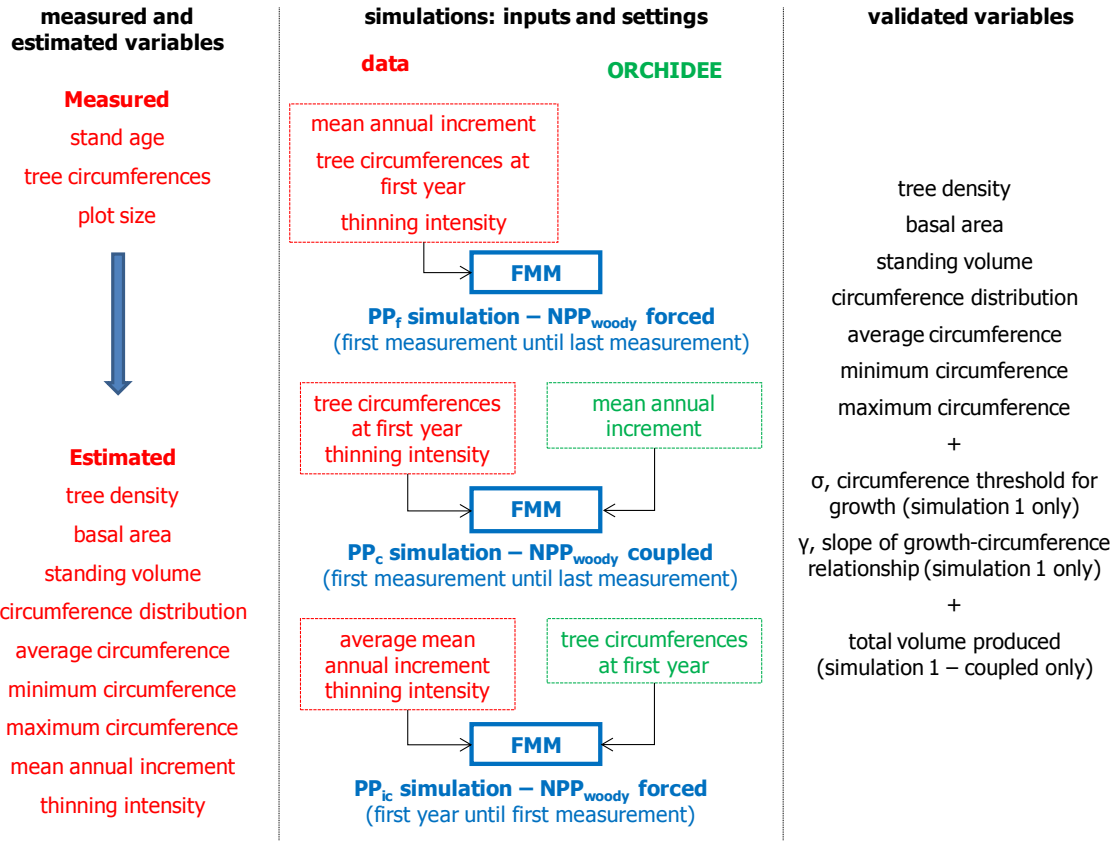
238

239 Figure S 3. Location of broadleaf permanent plots

240 Eight French forests contain 1 to 10 permanent plots with broadleaves, amounting to 58

241 plots over the metropolitan territory.

242

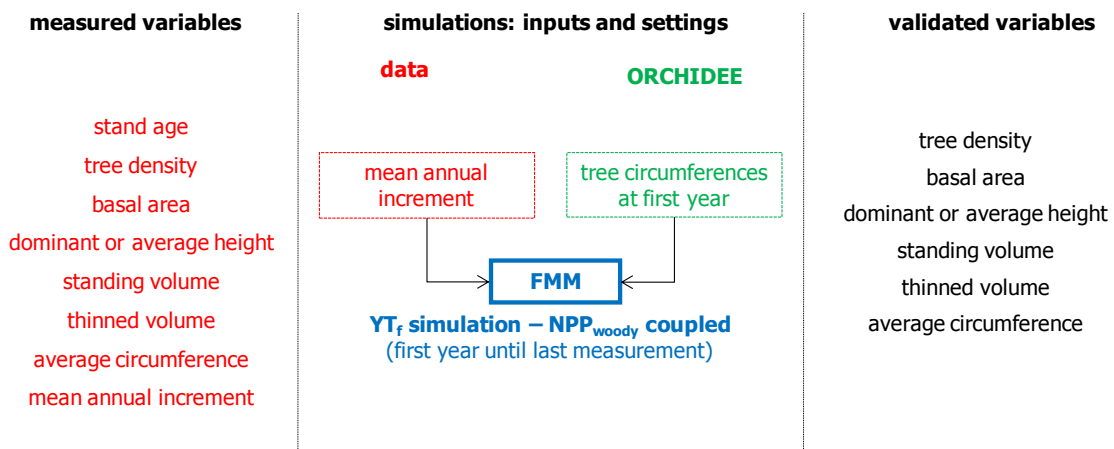


243

244 Figure S 4. Permanent plots: summary of measurements, simulations and validated

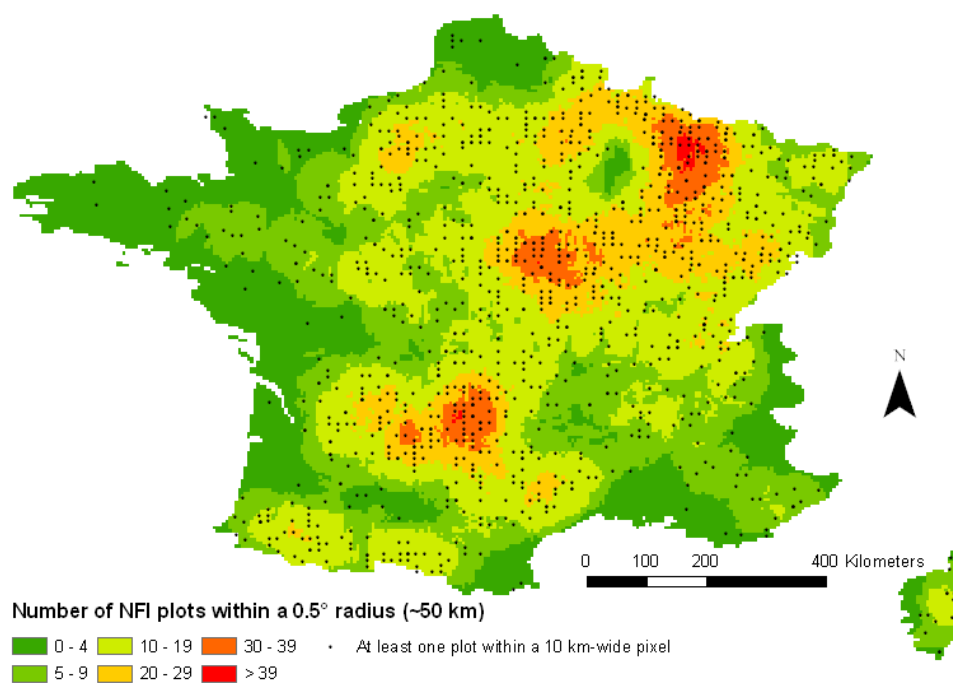
245 variables

246



247

248 Figure S 5. Yield tables: summary of measurements, simulations and validated variables



250

251 Figure S 6. National Forest Inventory (NFI) broadleaf plots of age class 80-100 years

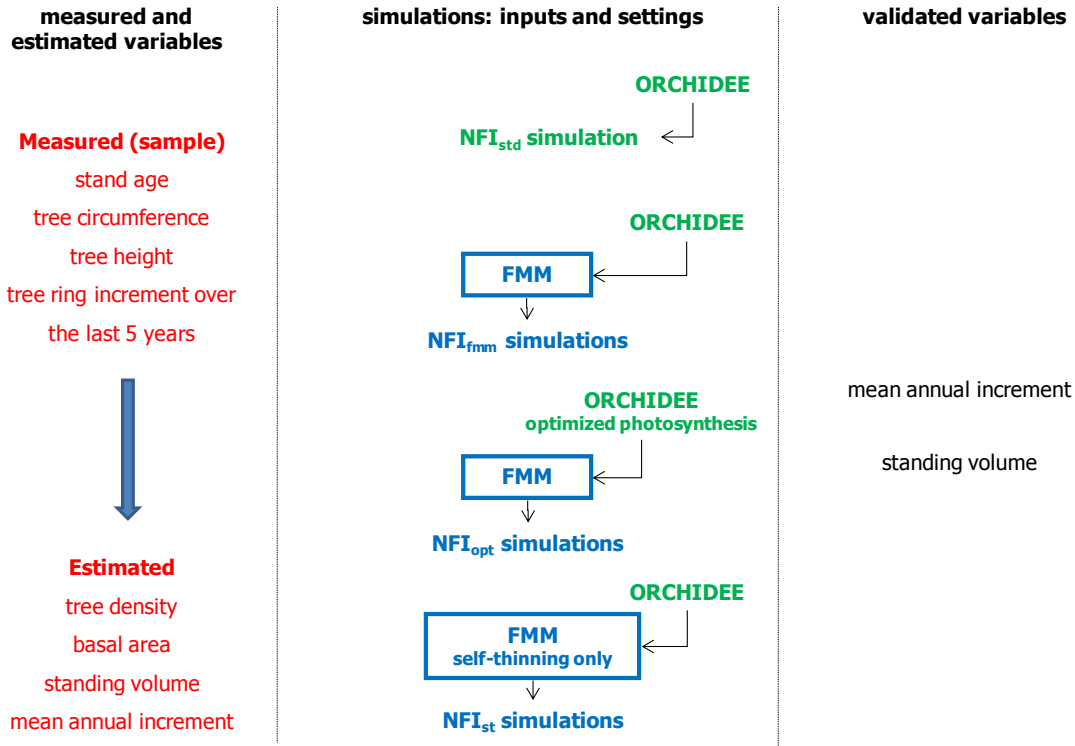
252

253 Figure S 7. French “regions” for which an average is calculated and discussed

254

255 Figure S 8. Relative density index of needleleaf plots of the 40-60 age class

256



257

258 Figure S 9. NFI plots: measurements, simulations and validated variables

259

260 Bellassen, V., Le Maire, G., Dhote, J.F., Viovy, N. and Ciais, P., this issue. Modeling
261 forest management within a global vegetation model – Part 1: model structure and
262 general behaviour. Ecological Modelling.

263 Bellassen, V., Le Maire, G., Luysaert, S., Schelhaas, M.J., Ciais, P. and Viovy, N.,
264 submitted. Reconstruction and attribution of the carbon sink of European forests between
265 1950 and 2000. Global Change Biology.

266 Deleuze, C., Pain, O., Dhote, J.F. and Herve, J.C., 2004. A flexible radial increment
267 model for individual trees in pure even-aged stands. Annals of Forest Science, 61:327-
268 335.

269 Dhote, J.F., Hatsch, E. and Rittie, D., 2000. Stem taper curves, volume tables and volume
270 yield compartments in Sessile Oak. Annals of Forest Science, 57:121-142.

271 EPA, 2010. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2008, United
272 States Department of Energy, Washington DC, USA, 457 p.

273 IFN, 2006. Observer la forêt française : mission première de l'IFN. L'IF:12.

274 IPCC, 2003. Good Practice Guidance for Land-Use, Land-Use Change and Forestry,
275 Intergovernmental Panel on Climate Change, Kanagawa, Japan, 534 p.

276 Newton, R.F. and Amponsah, I.G., 2007. Comparative evaluation of five height-diameter
277 models developed for black spruce and jack pine stand-types in terms of goodness-of-fit,
278 lack-of-fit and predictive ability. Forest Ecology and Management, 247:149-166.

279 Phillips, D.L., Brown, S.L., Schroeder, P.E. and Birdsey, R.A., 2000. Toward error
280 analysis of large-scale forest carbon budgets. *Global Ecology and Biogeography*, 9:305-
281 313.

282 USDA, 2007. *Forest Inventory and Analysis National Core Field Guide, Volume I: Field*
283 *Data Collection Procedures for Phase 2 Plots*, United States Department of Agriculture
284 Forest Service, Washington DC, USA, 224 p.

285 Waggoner, P.E., 2009. *Forest Inventories - Discrepancies and Uncertainties, Resources*
286 *For the Future*, Washington DC, USA, 45 p.

287 Zianis, D. and Mencuccini, M., 2004. On simplifying allometric analyses of forest
288 biomass. *Forest Ecology and Management*, 187:311-332.

289
290