Constraining a land cover map with satellite-based aboveground biomass estimates over Africa

Guillaume Marie[1], B. Sebastiaan Luyssaert[2], Cecile Dardel[3], Thuy Le Toan[4], Alexandre Bouvet[4], Stéphane Mermoz[4,6], Ludovic Villard[4], Vladislav Bastrikov[5], Philippe Peylin[1].

[1]Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[2]Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3]Laboratoire Géosciences Environnement, Paul Sabatier Université, Toulouse III, Toulouse, France
[4]Centre d'Etudes Spatiales de la Biosphère (CESBIO), Toulouse, France
[5]Science Partner, Paris, France
[6]GlobEO, Toulouse, France

Correspondance to : Guillaume Marie (Guillaume.Marie@uantwerpen.be)

Abstract. Most land surface models can, depending on the simulation experiment, calculate the vegetation distribution and dynamics internally by making use of biogeographical principles or use vegetation maps to prescribe spatial and temporal changes in vegetation distribution. Irrespective of whether vegetation dynamics are simulated or prescribed, it is not practical to represent vegetation across the globe at the species level because of its daunting diversity. This issue can be circumvented by making use of 5 to 20 plant functional types (PFT) by assuming that all species within a single functional type show identical land–atmosphere interactions irrespective of their geographical location. In this study, we hypothesize that remote-sensing based assessments of above-ground biomass can be used to constrain the process in which real-world vegetation is discretized in PFT maps. Remotely sensed biomass estimates for Africa were used in a Bayesian framework to estimate the probability density distributions of woody, herbaceous, and bare soil fractions for the 15 land cover classes, according to the UN-LCCS typology, present in Africa. Subsequently, the 2,5th and 97,5th percentile of the probability density distributions were used to create 2,5% and 97,5% credible interval PFT maps. Finally, the original and constrained PFT maps were used to drive biomass and albedo simulations with the ORCHIDEE model. This study demonstrates that remotely sensed biomass data can be used to better constrain the share of dense forest PFTs but that additional information on bare soil fraction is required to constrain the share of herbaceous PFTs. Even though considerable uncertainties remain, using remotely sensed biomass data enhances the objectivity and reproducibility of the process by reducing the dependency on expert knowledge and allows assessing and reporting the credible interval of the PFT maps which could be used to benchmark future developments.

1 Introduction

Degradation, fires and deforestation of tropical forests are responsible for two thirds of the global net deforestation emissions (Houghton et al., 2012; Le Quéré et al., 2015; Friedlingstein et al., 2020). Although African tropical rainforests represent only one third of the global tropical rainforests (Lewis et al., 2009), they were responsible for almost all, i.e., 1,48 PgC in 2015 and 1,65 PgC in 2016, of the net carbon (C) emissions of pan-tropical regions, but substantial uncertainty is associated with these estimates, i.e., 1,15 for 2015 and 1,0 PgC for 2016, mainly driven by fire and land use changes (Palmer et al., 2019). The uncertainty of model estimates, such as mentioned above, broadly comes from three sources: (1) the vegetation distribution in the model, (2) the ability of the model to simulate biomass accumulation of undisturbed vegetation, and (3) the ability of the model to simulate natural and anthropogenic disturbances of the standing biomass. As this study will focus on improving the description of the vegetation distribution, the first question that needs to be answered is why vegetation distribution remains so uncertain?

Most land surface models can either calculate the vegetation distribution internally by making use of biogeographical principles (Sitch et al., 2003; Krinner et al., 2005; Clark et al., 2011) or use vegetation maps to prescribe spatial and temporal changes in vegetation distribution. Where the first approach results in a description of the potential vegetation, the second approach is more suitable when actual vegetation is to be studied. Irrespective of whether potential or actual vegetation is studied, it is not practical to represent vegetation across the globe at the species level because there are already over 60,000 tree species (Beech et al., 2017), not to mention the diversity in herbs, forbs and mosses. Land surface models represent this daunting diversity by making use of 5 to 20 plant functional types (PFT) (Huete et al., 2016). The underlying assumption of plant functional types is that all species within a single functional type show identical land–atmosphere interactions irrespective of their geographical location (Huete et al., 2016; Bonan et al., 2002; Brovkin et al., 1997; Chapin et al., 1996). Discretizing real-world vegetation in PFTs is a first source of uncertainty.

When actual vegetation is the focus of a modelling study, the vegetation distribution will have to be prescribed. The construction of vegetation maps first requires real-world observations, typically through satellite-based remote sensing. Current remote sensing technology does not enable distinguishing individual tree species; hence, vegetation is observed as land cover types (Defourny, P., 2019) which group vegetation with similar sensory characteristics. The remote sensing observations themselves as well as classifying them in land cover types are the second and third source of uncertainties (Hansen et al., 2013, Mitchard et al., 2014, Hurtt et al., 2004). Because the land surface models require the vegetation to be discretized in PFTs, which may differ between different land surface models, the land cover types will have to be remapped on PFT maps. The rules applied in remapping satellite-based land cover types in PFT maps is formalized in so-called “cross-walking tables” (CWT) (Poulter et al., 2011; Poulter et al., 2015) which are a fourth source of uncertainty (Hartley et al., 2017).

Although CWTs have been extensively used to create PFT maps (Wei et al., 2018; Wei et al., 2016; Poulter et al., 2011; Krinner et al 2005), the process of associating land cover types with specific PFTs remains difficult to reproduce since several iterations of expert knowledge are required (Poulter et al., 2011; Poulter et al., 2015). Various land cover classifications exist, the commonly used FAO (Food and Agriculture Organization) Land Cover Classification System (LCCS; Di Gregorio and Jansen, 2000). Most classes of the LCCS correspond to a mix of PFTs, which fractions are difficult to assess and likely variable across regions. For example, several classes are labelled as a mosaic of vegetation types (i.e., “Mosaic of natural vegetation (tree, shrubs, herbs)”; see Table 2 in Poulter et al., 2015). Not surprisingly, efforts have been made to decrease the need of expert knowledge in favour of more objective and reproducible approaches, e.g., classification rules based on a suite of improved and standard MODIS products (Wanxiao et al., 2008). Moreover, producing PFT maps from satellite-based land cover maps needs to become fully automated when the temporal frequency of satellite-based land cover and biomass maps will increase thanks to the recent GEDI Lidar data (Dubayah et al., 2020) or future SAR missions like the NASA-ISRO Synthetic Aperture Radar (NiSAR) or BIOMASS missions (Le Toan et al., 2011; Quegan et al., 2019).

In this study, we hypothesize that remote sensing-based assessments of above ground biomass (ABG) can decrease the dependency on expert knowledge when setting up CWTs and as such contribute to the automation of the land cover class mapping into PFTs for land surface models. The main rationale is that the above-ground biomass content of an ecosystem provides information on the fraction of tree PFTs of that ecosystem. In this context, the objective of this study are: (1) construct a framework of data assimilation in which biomass remote sensing products can be routinely used to update an existing or create a new CWT, (2) constrained a cross-walking table used to convert the ESA-CCI Global Land Cover map into a PFT map, and (3) propagate the credible interval from using a CWT in the production of PFT maps, to the simulation results of biomass and albedo maps derived from a land surface model. Such a framework will be applied and tested over Africa using the above ground biomass product derived by (Bouvet et al., 2018) for that continent with the ORCHIDEE land surface model (Krinner et al., 2005) more specifically tag 2.0 revision 6592 close tag 2.2 used for the recent Climate Modelling Intercomparison Project - phase 6 (CMIP6) (Boucher et al., 2020).

2 Materials and methods

2.1 Overview

Cross-walking tables (CWT) (Poulter et al., 2015) are used to convert the 43 land cover types distinguished on the ESA-CCI land cover product into generic plant functional types (13 PFTs in (Poulter et al., 2015)) distinguished by large-scale land surface models such as the ORCHIDEE model (Krinner et al., 2005) used in this study. These generic PFTs are further grouped and/or divided to match each model-specific PFT classification, using additional grid-cell information to separate grassland and crop C3 versus C4 photosynthetic pathway (Still et al., 2003) and to split generic PFT according to bioclimatic zones (i.e., Koppen Geiger climate classification map) (see more details for the ORCHIDEE model in (Lurton et al., 2020)). In this study, we provide a proof of concept by creating a new ORCHIDEE PFT map by combining information from the ESA-CCI land cover product and the AGB product for Africa (Bouvet et al., 2018) to estimate woody, herbaceous and bare soil cover fractions within each land cover type of the ESA-CCI product. Subsequently, the estimated cover fractions are used to constrain the existing CWT and create a new ORCHIDEE PFT map applicable primarily for Africa (Fig. 1). Finally, the impact of using AGB maps to constrain the PFT maps on the skill of the ORCHIDEE model to simulate the contemporary biomass and its spatial distribution over Africa is quantified. Note that the approach is tested over Africa but is generic enough to be applied everywhere.

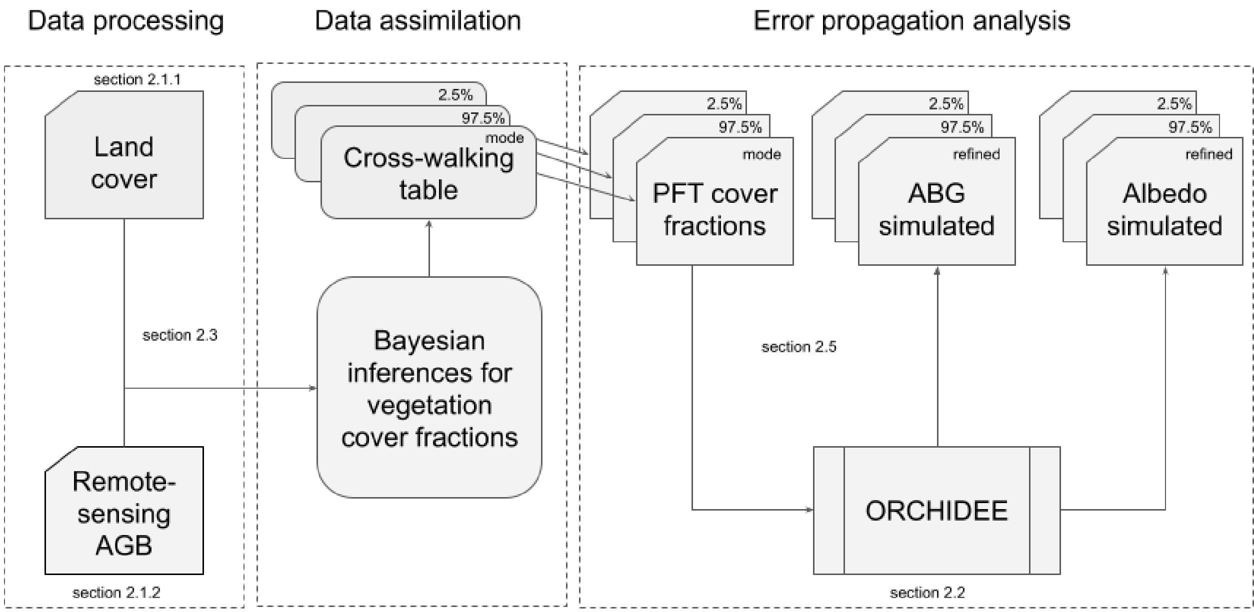

Figure 1: Approach to assimilate the information held by aboveground biomass (AGB) maps into plant functional type (PFT) maps. Remote sensing AGB and land cover products are jointly assimilated to obtain cross-walking tables that can be used to make PFT maps. Owing to the uncertainty analysis in the data assimilation approach, an ensemble of cross-walking tables and PFT cover fraction maps can be produced. Subsequently, the land surface model ORCHIDEE can be run for different PFT maps to quantify the uncertainty from propagation of the uncertainty from remote sensing products into a model simulation.

2.2 Dataset products

2.2.1 Land cover map

ESA's Climate Change Initiative for Land Cover (CCI-LC) produced consistent global LC maps at 300 m spatial resolution on an annual basis for the year 2015 (Defourny, P., 2019). Only one year (2015) has been used to estimate the new vegetation cover cross-walking table. The typology of CCI-LC maps follows the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agriculture Organization (FAO), to enhance compatibility with similar products such as GLC2000, and GlobCover 2005 and 2009. The UN-LCCS typology was designed as a hierarchical classification, which allows adjusting the thematic detail of the legend. The "level 1" legend, also called "global" legend, counts 22 classes and is globally consistent and thus suitable for global applications such as creating PFT maps for land surface models. The "level 2" or "regional" legend counts 43 classes which are not present all over the world and could be used in this study given its focus on a single continent, i.e., Africa (see section 2.2.3). In addition, the UN-LCCS partly overlaps with the PFTs used in climate models.

2.2.2 Aboveground biomass map

This study also makes use of a continental map of AGB of African savannas and woodlands for the year 2010 (Bouvet et al., 2018). The map has a 25 m resolution and is built from the 2010 L-band data of the Phased Array L-band Synthetic Aperture Radar (PALSAR) on the Advanced Land Observing Satellite (ALOS) satellite. Covering the African continent required about 180 data strips of which 91% were acquired between May and November 2010. The remaining 9% of the domain was filled with imagery from 2009 and 2008. The data have been processed by the Japan Aerospace Exploration Agency (JAXA) using the large-scale mosaicking algorithm described in Shimada and Ohtaki (2010), including ortho-rectification, slope correction and radiometric calibration between neighbouring strips, and multi-image filtering described in Bouvet et al., 2018.

The continental AGB map was derived as follows: (1) stratification into wet/dry season areas in order to account for seasonal effects in the relationship between PALSAR backscatter and AGB, (2) the development of a statistical model relating the PALSAR backscatter to observed AGB, (3) Bayesian inversion of the direct model, to obtain AGB and its credible interval for pixels where no observations are available, and (4) masking out non-vegetated areas using the ESA-CCI Land Cover dataset (but see section 2.1.1). The resulting AGB map was visually compared with existing AGB maps (Saatchi et al., 2011; Baccini et al., 2012; Avitabile et al., 2016) and cross-validated with AGB estimates obtained from field measurements and LiDAR datasets (Naidoo et al., 2015). Cross-validation revealed a good accuracy of the dataset, with an RMSD between 8 and 17 t/ha. For more details on the creation and evaluation of the AGB maps see Bouvet et al., 2018.

2.2.3 Pre-processing

One known limitation of the original AGB map (Bouvet et al., 2018) is the signal saturation and in some cases the decrease of the signal (Mermoz et al., 2015) occurring in L-band SAR for AGB values higher than 85 t/ha. In order to overcome this issue, a second AGB map was created, based on two other ancillary datasets: a map of tree cover (Hansen et al., 2013) and a map of tree height (Simard et al., 2011). The AGB was estimated by deriving an empirical relationship between biomass, available from airborne Lidar estimates, and the product of tree cover and tree height. The second version targets dense forest areas such as in the Congo basin and is used to adjust the AGB values at locations where signal saturation occurred. Because of a coarser resolution from the tree height map (0,01°x0,01°, 100 ha) than the original AGB map (0,00025°x0,00025°, 0,0625 ha), the new biomass map has been rescaled to 0,01° resolution. The rescaling drastically reduced the noise produced by PALSAR measurement artefacts (personal communication Thuy Le Toan). The original AGB map was downscaled by an average resampling method, i.e., computing the weighted average of all contributing pixels. To do so, we used the Gdalwarp function from GDAL (GDAL/OGR). The map used in this study is a composite of the two versions of the biomass map by using the following rules:

- For broadleaved evergreen forests (UN-LCCS land cover type 50), flood forests (UN-LCCS 160), and closed broadleaved deciduous forests (UN-LCCS 61), the map based on tree cover and tree height was used because there is no AGB estimates in the map based on PALSAR.

- For broadleaved deciduous forests (UN-LCCS 60) the maximum between the two maps was used because its biomass ranged around the threshold of 85 t ha^{-1} and may create truncated distribution.
- For the other land cover types, which typically have a biomass well below 85t ha^{-1} the AGB value from the PALSAR map was used because it is considered more reliable than the statistical relationship between biomass, vegetation cover and vegetation height especially for the lower biomass.

Given the spatial domain of this study, only the 31 land cover types defined on the ESA CCI-LC map and present in Africa were retained. The complexity of the study was further reduced by removing all land types that cover less than 1,0% (304,158 km2) of the African surface or that contain less than 1% (i.e., 1,1 Gt) of the total AGB of Africa. Filtering retrained 15 out of the 31 land cover types including bare land. These 15 land cover types (Table 1) represent 96% of the surface of Africa and 98% of its AGB.

One additional issue had to be dealt with: the spatial resolution of the land cover map (9 ha) largely differed from the resolution of the AGB map (0,01°x0,01°, 100 ha). Therefore, each observational point on the AGB map is represented by 11x11 pixels on the land cover map. To simplify the overall data assimilation methodology (see section 3.2), we chose to use only AGB pixels (100 ha) which have a unique land cover type (i.e., pure pixels, in terms of land cover type). To this aim, the variety of land cover types across the 11x11 pixels within each AGB pixel (i.e., the number of present, *Vlct*) was calculated and only pixels where *Vlct=1* was retained. Although this criterion resulted in discarding 99% of the pixels, each of the 15 land cover types considered could be represented by at least 2000 pixels. To remove outlier pixels, we choose to pick up the 2000 pixels strictly below the biomass value representing the 97.5th percentile of each LCT biomass distribution shown is the figure 2.

2.3 Data assimilation

2.3.1 Linking land cover fractions and AGB

A linear model was used to relate the satellite-based AGB of a 100-ha pixel to the cover fraction of the satellite-based vegetation types present at the same location. This relationship can be written as:

$$B_p = \sum_{i=1}^{nV} F_{p,i} \cdot Bref_i \qquad (1)$$

where B_p, is the AGB at a given pixel p, $F_{p,\,i}$ is the cover fraction of the vegetation type i (i.e., the generic plant functional type (PFT) used for land surface models, see section 2.1 - overview), $Bref_i$ is the reference AGB for the vegetation type i and nV is the number of vegetation types (i.e., number of PFTs) present in the pixel p. Given the number of unknowns (nV being usually above 1), equation 1 has many solutions; many of which have no biological meaning. The equifinality of this model can be reduced by arguing that the large difference in biomass between woody, herbaceous and non-vegetated ecosystems combined by their respective cover fraction explains most of the biomass at pixel level. Following this assumption, equation 1 can be simplified as:

$$B_p = F_{p,w} \cdot Bref_w + [\![(1 + F]\!]_{p,h} - F_{p,b}\big) \cdot Bref_h \tag{2}$$

where $F_{p,w}$, $F_{p,h}$ and $F_{p,b}$ are the fractions cover for woody vegetation (i.e., woody PFTs), herbaceous vegetation (i.e., grassland and cropland) and non-vegetated areas, respectively. $Bref_w$ and $Bref_h$ are the reference AGB of woody and herbaceous vegetation, respectively. Equation 2.1 is constrained by equation 2.2 (i.e., the total area coverage of each pixel), hence, $F_{p,h}$ in equation 2 can be substituted according to $F_{p,w} + F_{p,h} + F_{p,b} = 1$ to obtain:

$$B_p = F_{p,w} \cdot Bref_w + F_{p,h} \cdot Bref_h \tag{3}$$

Although equation 3 no longer details which vegetation types i (i.e., PFTs) are present on each pixel p, it still has four unknowns and can, therefore, not be solved analytically. Nevertheless, a statistical solution is within reach if $F_{p,w}$, $F_{p,h}$, $Bref_w$ and $Bref_h$ are estimated from a population of AGB observations containing several independent repetitions that largely exceeds the number of unknowns. In this study, over 2000 repetitions were available for each of the 15 land cover types that were retained following filtering (section 2.2.3). The statistical solution will thus consist of four mean parameter values (i.e., $F_{p,w}$, $F_{p,h}$, $Bref_w$ and $Bref_h$) for each of these 15 land cover types.

As described in section 2.2.3, the selection of homogeneous AGB pixels, i.e., which have a unique land cover class across the 11,11 underlying land cover sub-pixels allow us to rewrite the equation 3 as follow:

$$\underline{Bp_p} = F_{lc,w} \cdot Bref_{lc,w} + F_{lc,h} \cdot Bref_{lc,h} \tag{4}$$

where $\underline{Bp_p}$ is the average AGB of a specific land cover type lc and $F_{lc,w}$, $F_{lc,h}$, $Bref_{lc,w}$, $Bref_{lc,h}$ are the unknowns. The unknown parameters of the regression model (eq. 4) were estimated by using a Bayesian inference method. This approach has been chosen because it helps to synthesize various sources of information as well as to propagate credible intervals in the result of our land surface model (Ellison 2004). Bayesian inference requires, however, setting prior probability distributions for each of the unknowns, i.e., the biomasses and land cover fractions for each of the 16 land cover types. Given these prior probability distributions, Bayesian inference retrieves the posterior probability distribution for each of the unknown parameters.

2.3.2 Prior value distributions for $Bref_{lc,w}$, $Bref_{lc,h}$ and Bp_p

The AGB pixels were stratified according to their land cover type and for each land cover type the information contained in the distribution of the satellite based AGB served to estimate the mean and standard deviation of the prior values of $Bref_{lc,w.}$

To avoid negative Bref values we used a normal truncated distribution with $0 < a, b < +\infty$ where (a, b) are the truncated range:

$$Bref_{lc,w} \sim N\left(\mu_{lc,w}, \sigma_{lc,w}, a, b\right) \quad (5)$$

where, $\mu_{lc,\,w}$ is calculated as follow:

$$\mu_{lc,w} = Xth\,per.\left(\underline{Bp_{lc}}\right) \quad (6)$$

Where Bp_{lc} is a vector containing Bp_p values that belong to the land cover type lc and $X^{th}per$ denotes the 97,5th percentile for the woody cover types. This choice assumes that with the 97,5th percentile we select the AGB value of a pixel covered only by woody vegetation (i.e., woody PFT) for the selected land cover type. In contrast to using a few in-situ observations to define $\mu_{lc,\,w}$, our approach offers the advantage to rely on a large ensemble of satellite-derived AGB observations and to be coherent with the following optimization.

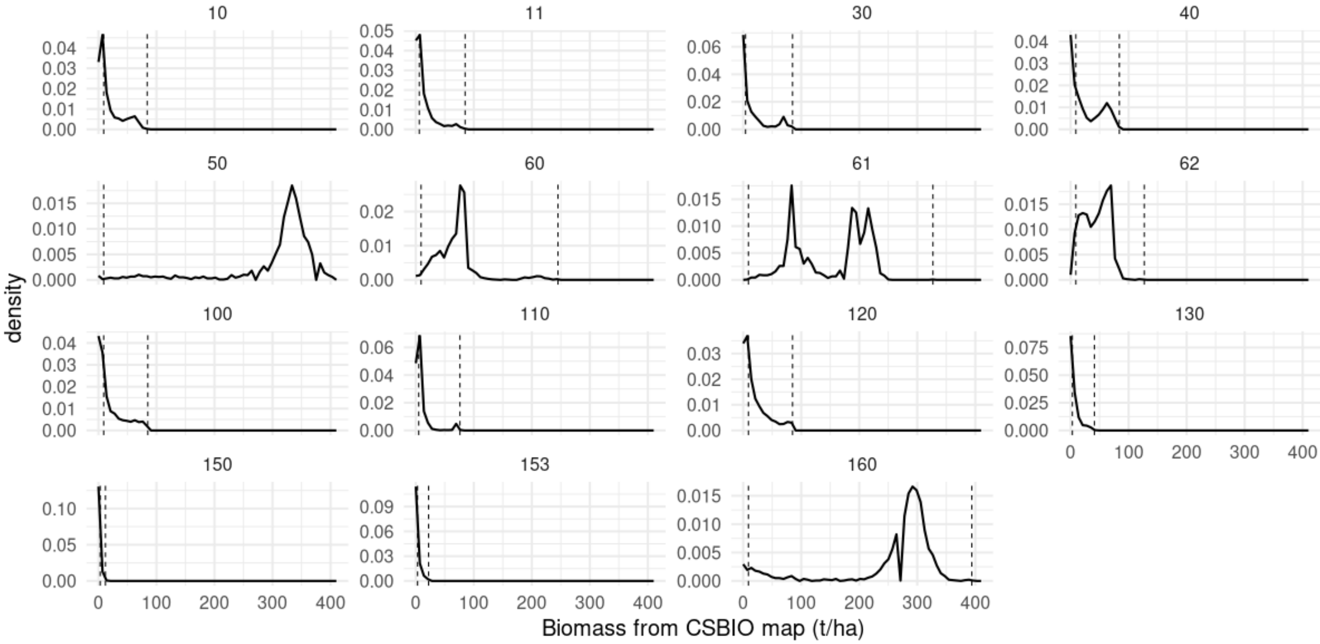

Figure 2: Probability density distribution of the pure land cover pixel for biomass concentration Bp_p for 15 selected land cover types. The dashed line represents the 97,5th percentile used as the prior estimate for the reference biomass concentration for trees $Bref_{lc,\,w}$. The dashed line represents the 50th percentile also used as the prior estimate for the reference biomass for herbaceous cover

Without any information about the variability of $Bref_{lc,\,w}$, we choose to represent $\sigma_{lc,\,w}$ as:

$$\sigma_{lc,w} = \mu_{lc,w}.0{,}0375 \qquad (7)$$

Where 0,0375 accounts for a 30% uncertainty encompassed between the interquartile range of the normally distributed $Bref_{lc,\,w}$. Compared to $Bref_{lc,\,w}$, $Bref_{lc,\,h}$ is more difficult to assess from the satellite-derived data because it often shows bimodal distributions which may stem from biomass degradation or the presence of shrubs which biomass better resembles that of a grassland than a woody ecosystem (Fig. 2). We found that while the 2,5th percentile is representing the lowest biomass for herbaceous ecosystem, the 50th percentile seems to better describe $Bref_{lc,\,h}$, following equation 6. Having no information about the variability of $Bref_{lc,\,h}$, $\sigma_{lc,\,w}$ followed equation 7.

Finally, Bp_p which was the 97,5th for woody cover types or the 50th percentile for herbaceous cover types, comes with a measurement uncertainty that was thought to follow a normal truncated distribution with $0 < a, b < Bref_{lc,w}$ where (a, b) are the truncated range. Given that this uncertainty is not known at the pixel level, an uninformative prior was set for the standard deviation σb_{lc} which can vary between 0 and 200 t/ha. We deliberately took a large uncertainty to cover the observation that considerable uncertainty remains in satellite-based biomass estimates (bouvet et al., 2018):

$$Bp_p \sim N(\mu, \sigma b_{lc}^2, a, b) \text{ with } \sigma b_{lc} \sim U(0{,}200) \text{ and } \mu = Bp_p \quad (8)$$

2.3.3 Prior value distributions for $F_{lc,w}$, $F_{lc,b}$ and $F_{lc,h}$

$F_{lc,w}$, $F_{lc,b}$ and $F_{lc,h}$ were defined as fractions of respectively woody vegetation, bare soil and herbaceous vegetation within a given land cover type, their values thus range between zero and one and their sum is equal to 1. For this reason, a Dirichlet distribution was used to describe the probability distribution of the woody, bare soil and herbaceous cover fractions:

$$(F_{lc,w}, F_{lc,b}, F_{lc,h}) \sim Di(\theta_{lc,t}, \theta_{lc,b}, \theta_{lc,h}) \quad (9)$$

OpenBUGS (Thomas, 2010), the software that was used in this study, cannot use a Dirichlet distribution as a stochastic node. This constraint can be overcome by making the cover fractions dependent on each other:

$$F_{lc,w} = q_{lc,1} \quad (10)$$
$$F_{lc,b} = q_{lc,2} \cdot (1 - q_{lc,1}) \quad (11)$$
$$F_{lc,h} = (1 - q_{lc,1}) \cdot (1 - q_{lc,2}) \quad (12)$$

Let $q_{lc,i}$ with $i = 1, \ldots, K - 1$ and K the number of fractions, be a series of independent beta distributions, Be (α_i , β_i).

$$q_{lc,i} \sim Be(\alpha_{lc,i}, \beta_{lc,i}) \quad (13)$$

The parameters of the beta distribution of the cover fraction of bare soil, woody vegetation and herbaceous vegetation (eq. 9) can then be estimated as follows:

$$\alpha_{lc,i} = \theta_{lc,i} \cdot (\omega_{lc,i} - 2) + 1 \quad (14)$$

$$\omega_{lc,i} \sim U(0{,}1000) \quad (15)$$

where $\theta_{lc,i}$ which represents the fraction of each land cover type taken from expert knowledge used to define the so-called cross walking table (CWT) and taken from a recent update of the CWT. $\omega_{lc,i}$ was described by an uninformative uniform distribution and thus reflects the relatively low trust we have in the current CWT. The dependencies between the beta distributions comes from $\beta_{lc,\, i}$ that is estimated as:

$$\beta_{lc,i} = \sum_{u=i+1}^{K} \alpha_{lc,u} \tag{16}$$

2.4 Confident interval propagation

2.4.1 Propagating the credible interval from the CWT into the PFT map

The posterior estimates of the cover fractions ($F_{lc,\, w}$, $F_{lc,\, b}$, $F_{lc,\, h}$) will be directly used to make up a new cross-walking table. The posterior estimates of the cover fractions values are then used to recalculate woody and herbaceous fraction of each generic PFT of the CWT. In other words, we keep the original split of the different woody PFT defined in the prior CWT and only rescale the total woody fraction to $F_{lc,\, w}$. Then we rescale the bare soil fraction based on $F_{lc,\, b}$ to finally rescale short vegetation PFTs (grass and crop).

Given that these posterior estimates come with a probability distribution, a probability distribution of the CWT could be made. In this study, the 2,5 and 97,5 percentiles and the mode, i.e., the most common value, of the posterior estimates were used to create three cross-walking tables that were then applied on the ESA-CCI-LC product to create two PFT maps that represent the 95% interval confidence of the ESA-CC-LC product and one PFT map which represents the one that is used in an ORCHIDEE simulation. The impact of the various PFT map was quantified for simulated above ground biomass and simulated surface albedo by running three simulations that only differed by the PFT map used to initialize the ORCHIDEE land surface model.

In the study, the uncertainty propagation index aimed to identify the ecoregions where the AGB and surface albedo estimates are most sensitive to uncertainties from the PFT map. This sensitivity was calculated as:

$$S_{eco,b} = \frac{ABS\left(X^{97,5} - X^{2,5}\right)}{ABS\left(F_{eco,b}^{97,5} \;\; - F_{eco,h}^{2,5}\right) \times 100} \tag{17}$$

Where X stands for AGB (t/ha) or surface albedo (unitless), $S_{eco,\, b}$ is expressed in the unit of X for a 1% change in bare soil fraction.

2.4.2 Description of the ORCHIDEE land surface model

ORCHIDEE (Krinner et al., 2005; Boucher et al 2020) is the land surface model of the IPSL (Institut Pierre Simon Laplace) Earth system model. Hence, by conception, it can be coupled to a global circulation model. In a coupled setup, the atmospheric conditions affect the land surface and the land surface, in turn, affects the atmospheric conditions. However, when a study focuses just on changes in the land surface ORCHIDEE rather than on the interaction with the atmosphere, it also can be run as a stand-alone land surface model. The stand-alone configuration receives atmospheric conditions such as temperature, humidity, and wind, to mention a few, from the so-called meteorological forcing. The resolution of the meteorological forcing determines the spatial resolution and can cover any area ranging from a single grid point to the entire globe.

Although ORCHIDEE does not enforce a spatial or temporal resolution, the model does use a spatial grid and equidistant time steps. The spatial resolution is an implicit user setting that is determined by the resolution of the meteorological data. ORCHIDEE can run on any temporal resolution; however, this apparent flexibility is restricted as the processes are nested and formalised at given time steps: half-hourly (i.e., photosynthesis and energy budget), daily (i.e., net primary production), and annual (i.e., vegetation dynamics). Hence, meaningful simulations have a temporal resolution of 1 min to 1 h for the energy balance, water balance, and photosynthesis calculations. In the land-only configuration used in this study, the default time step for these processes is 30 minutes.

When an application requires the land surface to be characterised by its actual vegetation, the vegetation will have to be prescribed by annual land cover maps. These maps must follow specific rules for the land surface models to be able to read them. In the case of ORCHIDEE the share of each of the 15 possible plant functional types needs to range between 0 and 1 and be specified for each pixel. When satellite-based land cover maps are used as the basis for an ORCHIDEE-specific PFT map, the satellite-based land cover classification will need to be converted to match the ORCHIDEE specifications. As mentioned already above, this involves two steps: i) the derivation of generic PFTs from the satellite land cover classes (in our case the ESA-CCI-LC product) through the CWT discussed in this paper and ii) the final mapping of the generic PFTs into the 15 ORCHIDEE-specific PFTs using additional information on the bioclimatic zones and the partition of grassland/crops into C3 versus C4 photosynthetic pathway (Lurton et al., 2020).

In this study, AGB was defined as the sum of leaf biomass, fruit biomass, aboveground sapwood biomass, and aboveground heartwood biomass which are default output variables of ORCHIDEE. Surface albedo was defined as the albedo in the visible wavelengths and is a default output variable of ORCHIDEE.

Table 1: Description of the 15 plant functional types (PFT) used in ORCHIDEE to represent global vegetation.

PFT	Climate	Vegetation type	Phenology class
1	global	NA	Bare soil
2	Tropical	Woody	Broadleaf evergreen
3	Tropical	Woody	Broadleaf deciduous
4	Temperate	Woody	Needleleaf Evergreen
5	Temperate	Woody	Broadleaf Evergreen
6	Temperate	Woody	Broadleaf Summer green
7	Boreal	Woody	Needleleaf Evergreen

8	Boreal	Woody	Broadleaf Summer green
9	Boreal	Woody	Needleleaf Deciduous
10	Temperate	Herbaceous	Natural (C3)
11	global	Herbaceous	Natural (C4)
12	global	Herbaceous	Managed (C3)
13	global	Herbaceous	Managed (C4)
14	Tropical	Herbaceous	Natural (C3)
15	Boreal	Herbaceous	Natural (C3)

2.4.3 Experimental setup

ORCHIDEE tag 2.0 (rev 6592) was used to run tree simulations that only differed by the PFT map used. Following a 340-yearlong spinup to initialise the carbon pools in the model, each simulation consisted of a 110years long simulation between 1901 to 2010 with the CRU-NCEP v8 climate reconstruction (Viovy, 2017) that matched the simulation years. CO2 concentration was fixed to 299,16 ppm and thus corresponds to the 2010 concentration.

2.4.4 Ecoregions

Results related to the land surface model simulation were presented by subdividing the African continent into ecologically homogeneous regions, so-called ecoregions, as defined by Olson et al., (2001).

3 Results

3.1 Prior and posterior distributions estimates

3.1.1 Vegetation cover fraction: prior and reference biomass distributions

Prior distributions for the cover fractions and reference biomasses were determined for all 15 land cover classes separately, nevertheless, four broadly different groups could be distinguished: (1) The 97,5th percentile of biomass distribution for each land cover belonging in the first group was so high, i.e., from 245 to 416 t/ha, that the land cover types in this group must correspond to a substantial tree cover., i.e., a woody cover fractions of 0,58 to 0,75. Examples of this group are land cover types UN-LCCS 50, 61, and 160 (tree cover broadleaf types in Table 2). (2) Contrary to the first group, the 97,5th percentile of biomass distribution for each land cover type of the second group is so low, i.e., from <12 to 42 t/ha, that these land cover types must be dominated by grasses or bare soil, i.e., a woody cover fraction of 0.1 or less and a substantial bare soil cover fraction up to 0,71. Examples of this group are UN-LCCS 130, 150 and 153 (grassland and sparse vegetation in Table 2). (3)

The biomass of the third group falls in between these extremes representing mosaic land cover types like the UN-LCCS 10, 11, 30, 40, 100, 110 and 120 (mosaic landscape in Table 2). When taken over the African continent, the biomass distribution of these land cover types shows bimodal biomass distributions indicating considerable variability within these land cover types (Fig. 2). (4) The bimodal biomass distribution of the fourth group is backed by a rather high woody reference biomass associated with a low woody cover fraction which may represent an ecosystem highly disturbed by either silvicultural practice or a fire regime. UN-LCCS 60, 62 fall into this group which represents the woodland to dry savanna continuum.

3.1.2 Vegetation cover fraction: posterior distributions

Owing to the Bayesian approach, the woody and herbaceous fraction within each land cover type is no longer deterministic (as was the case with the previous generation of cross-walking table such as in Poulter et al., 2015) but now comes with a distribution. This distribution is the outcome of propagating the credible interval on the retrieved parameters obtained from the Bayesian approach into the final product, i.e., the PFT cover fraction map. The 95% credible interval was studied by comparing the 2,5 and 97,5 percentiles of the distribution of woody, herbaceous and bare soil fractions ($F_{lc,\,w}$, $F_{lc,\,h}$, $F_{lc,\,b}$).

The mean change in forest cover fraction between the 2,5 and 97,5 percentiles of the distribution of constrained PFT maps over Africa was 1,6±2,6%. At the ecoregion scale (when averaging the cover fraction over the ecoregion), the largest uncertainty in forest cover fraction was found in the Congo basin with an average of -6,3±0,5% for the six ecoregions where LCT 50 is dominant (Fig 3A).

The 95% uncertainty interval for bare soil cover fraction is 13±8% mainly due to the large uncertainty of the cropland and mosaic cropland (UN-LCCS 10, 11, 30,40). In ecoregions where these LCTs are dominant, this credible interval increases to 24±7% (Fig. 3B). Moreover, dense forest land cover type i.e., LCT 50, 160 also come with 15±4% uncertainty in their bare soil fraction estimates (Fig. 3B).

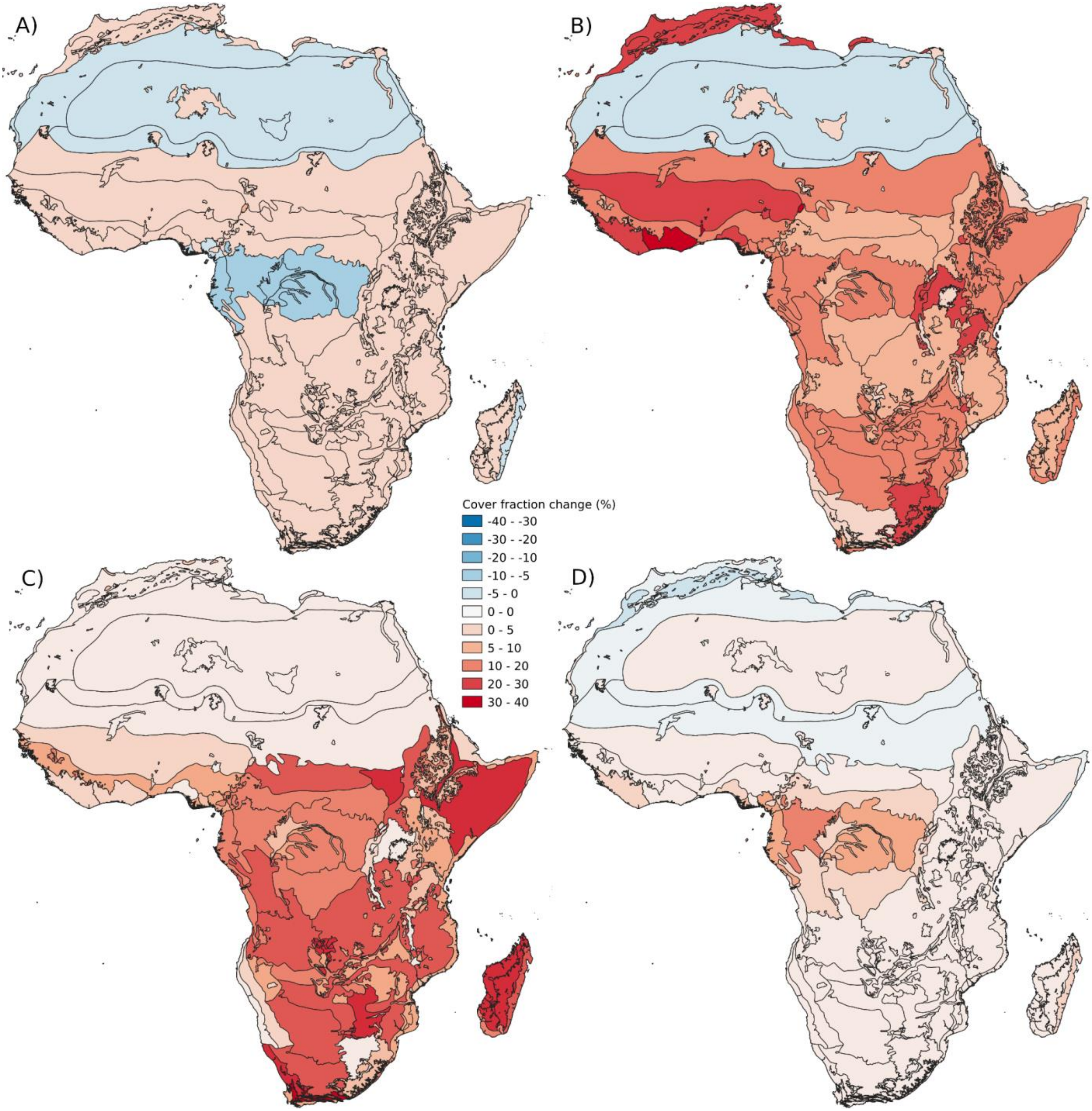

Figure 3: Uncertainty in CWT constrained by an AGB map. Absolute change in forest (A) and bare soil (B) cover fraction (%) between the 2,5 and 97,5 percentile PFT maps. High values represent a large uncertainty in the estimation of the true cover fraction. C and D represent disagreement estimated as the difference between the CWT based on expert knowledge and the CWT constrained by an AGB map. Disagreement in forest (C) and bare soil (D) is expressed as absolute change (%). High values represent a strong disagreement between the two methods. Black lines delimit the different ecoregion according to Olson et al., 2001.

Table 2: Surface area (%), share in the continental biomass (%), prior parameters, and posterior median and credible interval values for each of the 15 land cover types considered in this study. The numbering, description and surface area of each land cover type is based on the ESA-CCI product (Defourny, P., 2019), where its share in the continental biomass is based on a compilation of Bouvet et al 2018. θ_{lc}, μ_{lc} and σ_{lc} represent the parameters describing the prior distributions of F_{lc} and $Bref_{lc}$. Estimation of these parameters is detailed in section 2.3. For each land cover type and each parameter, the 2,5, the constrained and the 97,5 percentiles are computed. We use the mode for the constrained CWT as an approximation of the posterior $\theta_{lc,w}$, since the posterior distributions of $F_{lc,i}$ may be asymmetric.

Informations				Priors					Posterior fractions used for the CWTs								
id	UN-LCCS short description	Surface area (%)	Biomass (%)	$F_{lc,w}$	$F_{lc,h}$	$F_{lc,b}$	$Bref_{lc,w}$	$Bref_{lc,h}$	CWT 2,5 %			CWT 97,5 %			Refined CWT		
				$\theta_{lc,w}$	$\theta_{lc,h}$	$\theta_{lc,b}$	$\mu_{lc,w} \pm \sigma_{lc,w}$	$\mu_{lc,h} \pm \sigma_{lc,h}$	$F_{lc,w}$	$F_{lc,h}$	$F_{lc,b}$	$F_{lc,w}$	$F_{lc,h}$	$F_{lc,b}$	$F_{lc,w}$	$F_{lc,h}$	$F_{lc,b}$
10	Cropland rainfed	7,6	5	0,01	0,98	0,01	83±3,1	9±0,3	0,19	0,34	0,47	0,13	0,86	0,01	0,14	0,83	0,03
11	Cropland rainfed - Herbaceous cover	3,2	3,3	0,01	0,98	0,01	84±3,1	6±0,2	0,13	0,49	0,38	0,09	0,89	0,02	0,11	0,85	0,04
30	Mosaic cropland (>50%) / natural vegetation (tree/shrub/herbaceous cover) (<50%)	2,3	3,1	0,25	0,74	0,01	85±3,2	4±0,2	0,2	0,77	0,03	0,17	0,83	0,00	0,18	0,81	0,01
40	Mosaic natural vegetation (tree/shrub/herbaceous cover) (>50%) / cropland (<50%)	2,2	1,9	0,5	0,49	0,01	84±3,1	9±0,3	0,26	0,69	0,05	0,22	0,78	0,00	0,24	0,75	0,01
50	Tree cover broadleaved evergreen closed to open (>15%)	6,7	45,1	0,99	0	0,01	416±15,6	9±0,3	0,71	0,02	0,27	0,79	0,1	0,11	0,76	0,01	0,23
60	Tree cover broadleaved deciduous closed to open (>15%)	4,2	8,7	0,7	0,29	0,01	245±9,2	9±0,3	0,27	0,66	0,07	0,23	0,76	0,01	0,25	0,73	0,02
61	Tree cover broadleaved deciduous closed (>40%)	0,4	1,8	0,85	0,14	0,01	252±9,5	9±0,3	0,54	0,44	0,02	0,61	0,3	0,09	0,58	0,4	0,02
62	Tree cover broadleaved deciduous open (15-40%)	10,6	13,1	0,55	0,44	0,01	111±4,2	9±0,3	0,35	0,61	0,04	0,3	0,69	0,01	0,32	0,66	0,02
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	1,8	1,5	0,6	0,39	0,01	85±3,2	9±0,3	0,16	0,77	0,07	0,13	0,86	0,01	0,15	0,84	0,01
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	1,6	1,2	0,4	0,59	0,01	75±2,8	5±0,2	0,07	0,85	0,08	0,05	0,94	0,01	0,06	0,93	0,01
120	Shrubland	13,3	7,7	0,6	0,39	0,01	85±3,2	9±0,3	0,16	0,74	0,1	0,12	0,87	0,01	0,14	0,85	0,01
130	Grassland	6,5	1,5	0,01	0,98	0,01	42±1,6	3±0,1	0,11	0,56	0,33	0,07	0,92	0,01	0,09	0,9	0,01
150	Sparse vegetation (tree/shrub/herbaceous cover) (<15%)	1,6	0,2	0,1	0,2	0,7	12±0,5	3±0,1	0,04	0,45	0,51	0,15	0,01	0,84	0,1	0,19	0,71
153	Sparse herbaceous cover (<15%)	1,1	0,1	0,01	0,29	0,7	22±0,8	3±0,1	0,1	0,29	0,61	0,01	0,97	0,02	0,02	0,81	0,17
160	Tree cover flooded fresh or brackish water	0,7	3,5	0,75	0,24	0,01	386±14,5	9±0,3	0,6	0,39	0,01	0,69	0,27	0,04	0,65	0,34	0,01

Nonetheless, in a classic simulation experiment the most common values of $F_{lc,\,w}$, $F_{lc,\,h}$, $F_{lc,\,b}$ will be used. The most common values of $F_{lc,\,w}$, $F_{lc,\,h}$, $F_{lc,\,b}$ are given by the mode of the posterior distribution ("constrained CWT" in Table 2.a). The mode was used to show the difference between the original and the constrained PFT maps (Fig. 3C-D). The mean difference in forest cover fraction between the prior (original) and the constrained PFT maps is -15±12% (Fig. 3C). Largest disagreement between was observed over the Somali Acacia-Commiphora Bushlands and Thickets and the Kalahari Xeric Savanna where forest cover fraction was found to be on average 32±1% lower in the constrained PFT maps (Fig. 3C). The Bare soil cover fraction changes on average by 3,1±0,5% (Fig. 3D). The constrained PFT map has on average 16±4% more bare soil cover fraction over the Congo Basin than the original map (Fig. 3C).

3.3 Uncertainty propagation of the PFT maps on the aboveground biomass and visible albedo estimates from ORCHIDEE simulations

PFT maps are essential boundary conditions of land surface models because they condition the spatial distribution of various ecosystem states-properties (i.e., carbon content, albedo, water-carbon-energy fluxes, etc). When tested with ORCHIDEE tags 2.0 (rev 6592), the absolute difference in biomass stock between the 2,5 and 97,5 percentile maps was 0,5±5,7 t/ha (Fig. 4A) representing 0.2 t/ha/% of cover faction (Fig 4C). Notable exception is the Congo basin where different PFT maps could result in AGB estimates that differ by 18 t/ha (Fig. 4A) for a 6,5% difference in the forest cover (Fig 3A). Different PFT maps make the average visible albedo range from 0,081±0,055 to 0,083±0,055. The largest uncertainty for the visible albedo simulated with ORCHIDEE was found over the Nigerian lowland forest (0,158) and West Sudanian Savanna (0,107) (Fig. 4B) which represent a 24% to 11% change in forest cover respectively. The sensitivity is the highest in the western Congo basin with 1,4% of albedo/% of cover faction. In contrast, West Sudanian Savanna possesses a low sensitivity with 0.5%. To summarise, we found that a smaller forest to bare soil transition uncertainty can drastically change the albedo of an ecoregion than a larger uncertainty in the grassland/cropland to bare soil transition.

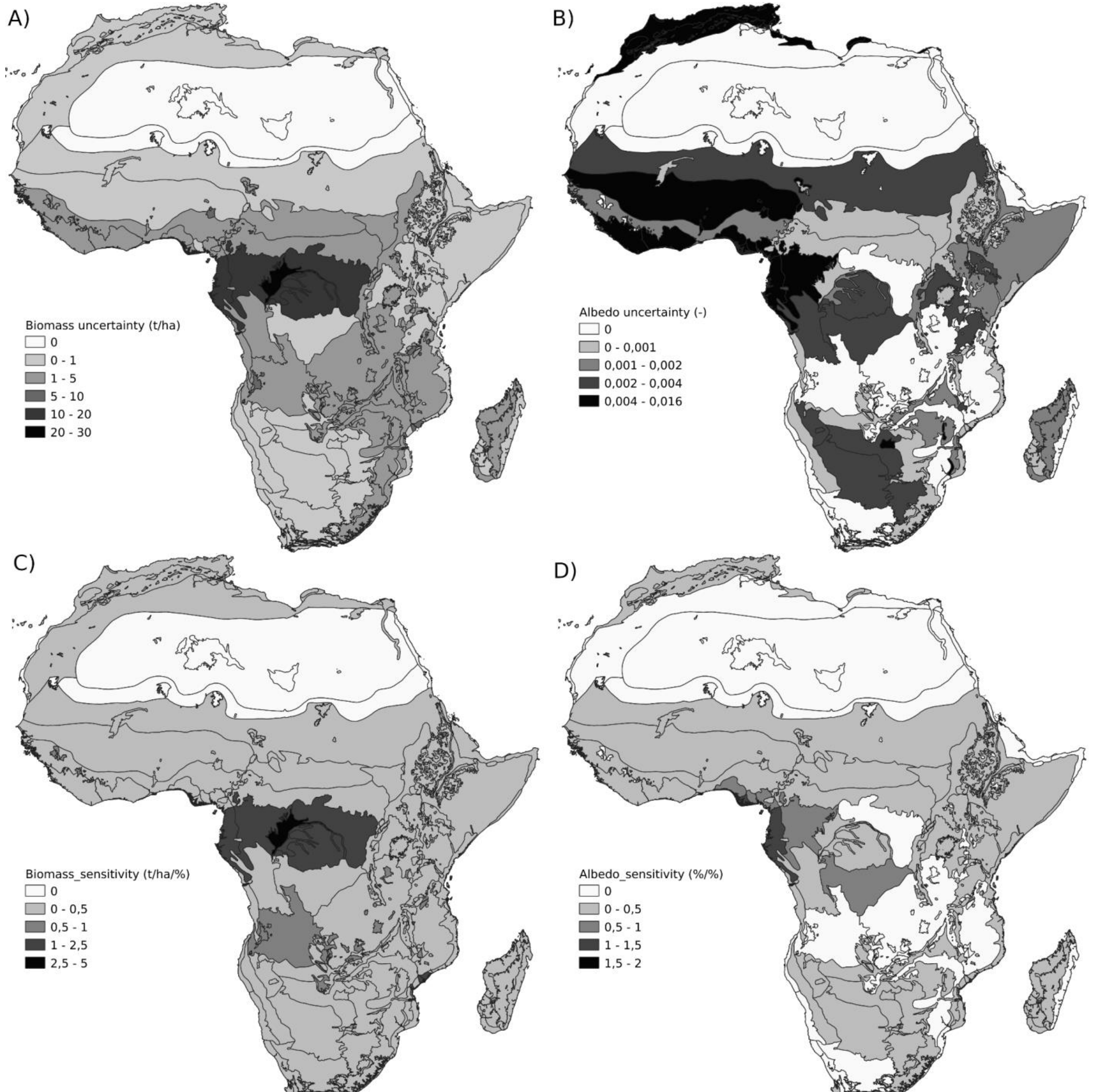

Figure 4: Confident interval propagation of the PFTs maps into AGB and visible albedo simulated by ORCHIDEE. (A) uncertainty propagation into AGB and (B) uncertainty propagation into visible albedo from the difference between the 2,5% and the 97,5% PFT map defined by the optimisation procedure. uncertainty propagation index (eq. 16) for AGB (C) and visible albedo (D).

4 Discussion

4.1 Discretizing vegetation

Irrespective of the data products, the methods, and the model used, discretizing vegetation comes with its own challenges. Discretizing transitions of ecosystems into land cover type classes (Sankaran et al., 2005), for example, can lead to systematic uncertainties since all pixels that belong to the same land cover class will get the same vegetation cover fractions in the cross-walking table (see 4.1.3). This approach articulates a key assumption underlying the PFT-approach, i.e., that only one life form survives and thus dominates the vegetation due to competition for nutrients, light and water (Hutchinson et al., 1961). However, the Savanna ecosystem, for example, is characterised by the coexistence of trees, shrubs and grasses which has been explained by interactions between vegetation, rainfall, fire, and browsing regimes (Eigentler and Sherratt., 2020). This makes savannas one of the most difficult ecosystems to classify in a land cover type and subsequently convert it into a PFT map.

Over Africa, land cover classes such as shrubland (UN-LCCS 120) represent a wide range of ecosystems, from sparse xeric shrubland composed of small bushes, e.g., *Penzia incana* (Thunb.) Kuntze, grasses, e.g., *Sip agrostis* spp. such as found in Karoo desert, to dense thicket composed by succulent, e.g., *Portulacaria afra* Jacq. and spinescent shrubs (~3m tall) (Mills, et al., 2005). Combining land cover types and biomass maps showed that the shrubland pixels in Africa more often resemble sparse xeric shrubland than dense thickets. Improving the ability to simulate land surface properties of shrublands in a changing world, especially in Africa where shrub encroachment is an important land cover dynamic (Wigley et al., 2010, Buitenwerf et al., 2012, O'Connor et al., 2014), is likely to benefit from a more detailed representation of shrublands in land surface models. A first step could be to represent shrubs as small trees, as was tested with the ORCHIDEE model for arctic ecosystems (Druel et al., 2017), but ultimately the control of precipitation on plant density (Rietkerk et al., 2002) should also be modelled.

Another major challenge with discretizing vegetation is how degraded ecosystems should be classified. From a modelling point of view, they should be classified as the land cover type that occurred prior to the degradation and the cause of the degradation. e.g., fire, grazing, erosion, should be explicitly accounted for in the land surface model. This ideal strongly differs from the current approach in which the degraded vegetation is classified as if it is in its natural state. Even when having the correct PFTs, the current approach would fail to simulate the observed biomass if degradation occurred. As an alternative, the PFT map could duplicate all PFTs to distinguish between a PFT in its natural state and in its degraded state. This approach in which degradation is accounted for in the PFT maps would, however, reduce degradation to a binary problem rather than addressing its continuous nature.

4.2 Knowledge gain from using the AGB map

In the absence of an AGB map, previous efforts to build cross-walking tables (Poulter et al., 2015) had to rely in part on expert knowledge. That generation of cross-walking tables can be considered as the best-available-knowledge in the absence of AGB data or other information on the land surface cover. The method developed and demonstrated in this study mostly relies on data but comes with its own assumptions and statistical complexities. The key assumptions are that: (1) previous cross-walking tables (Poulter et al., 2015) are a reliable source to set the prior distribution for PFT cover, (2) the biomass map (Bouvet et al., 2018) is a reliable source to set the prior distribution of the reference biomasses, and (3) the land cover classification contains homogeneous land cover types (Defourny, P. et al., 2019). A key question is thus whether the added complexity justifies the knowledge gained by jointly assimilating a land cover and a biomass map when producing a CWT?

Ideally this question should be addressed by assessing the reduction of the credible interval associated to the posterior distribution of the PFT map when using the AGB map to constrain the CWT (in comparison to a prior when no AGB is used). However, the present generation of CWT without AGB information, does not come with a distribution (except the attempt in Hartley et al., (2017)), calling for an alternative approach to assess the knowledge gain. Given that the prior distribution of the cover fraction was based on the previous CWT, the difference between the prior and the posterior distributions can be considered as the knowledge gained from using AGB information. Following this reason, the question we seek to answer is: "Is the cover fraction used by the original cross walking table falling outside the 95% credible interval of our posterior estimate?"

If the answer is no, the biomass map is more likely in agreement with the previous effort to estimate the original cross walking table. If the answer is yes, adding the information contained in the satellite-based biomass maps is most likely in strong disagreement with the previous effort to estimate the original cross walking table. The original CWT has a global extent, and the constrained CWT is only valid for Africa. Therefore, knowledge gains should be carefully interpreted as they may reflect trade-offs that had to be made previously to construct a global rather than regional CWT. Knowledge gains were assessed for: "croplands", "dense evergreen forests", "woodlands and savannas", and "xeric shrublands and grasslands" separately.

4.2.1 Croplands (UN-LCCS 10, 11, 30, 40).

Despite the cover fraction of woody vegetation on croplands being close to none in the original CWT, this study found that the four land cover types associated with croplands, UN-LCCS 10, 11, 30, 40 are in fact covered with 11% to 24% woody vegetation (Table 2). This large difference in the presence of woody vegetation on croplands is also reflected in the biomass data, which suggest two distinct but co-existing agricultural systems in Africa, i.e., one system with a low biomass and one around with a higher biomass.

The agricultural system with the low biomasses likely represents annually replanted crops such as millet, sorghum, wheat, sweet potatoes or cassava (FAO), with a maximum reported biomass between 10 and 15 t/ha for high-input cropping associated with commercial production of cassava and sweet potatoes. These values are in line with values estimated as reference biomass (see 2.3.2). Nonetheless, 97% of total cropland area Africa is rainfed (Calzadilla et al., 2009) and most of Africa's agricultural land is used for subsistence or small-scale farming associated with low-input cropping which explains why the actual average biomass estimate from the CESBIO map for cropland is between 2,0± 0,7 t/ha (Fig. 2) and thus considerably lower than the potential production.

The high biomass agricultural system which is estimated at 83±3 t/ha in the CESBIO map (Fig. 2) likely includes plantations for coffee, rubber, fruits as well as shelter trees and forest remnants (FAO). Permanent croplands do not have their own land cover type in the UN-LCCS or in ORCHIDEE. The mixture of bare soil, herbaceous vegetation and woody vegetation, makes it challenging to discretize African croplands into the current PFTs (Table 1). Moreover, small changes in the woody reference biomass for high biomass agricultural systems lead to large changes in cover fractions of herbaceous vegetation and bare soil ratio. Without constraint reference biomass estimates, total biomass alone does not sufficiently constrain the share of woody vegetation. For the time being, high biomass agricultural systems could be with a woodland fraction ranging from 9,0% to 26% (Table 2). Although this could be an acceptable solution for biomass and albedo simulations, it will underestimate the agricultural production in the region.

4.2.2 Tropical rainforest (UN-LCCS 50, 160).

The woody cover fraction of tropical rainforest in the original CWT is close to 90% and falls outside the credible interval of the posterior estimates, i.e., 71 to 79%. This lower cover fraction from many pixels classified as tropical rainforest that do not reach the reference biomass of 416±16 t/ha (Fig. 2). The reference derived from the biomass map matches the AGB observed at field plots of intact forests in the Congo basin (Lewis et al., 2013) but the large value in bare soil cover fraction for these land cover types may thus reflect wide-spread degradation of the forests in the region (Tyukavina et al., 2018) or a too high reference biomass (Kearsley et al., 2013).

4.2.3 Tropical deciduous forest, woodland, and savanna (UN-LCCS 61, 60 and 62).

The woody cover fraction of the tropical deciduous forest ranged between 45% and 75% in the original CWT. Refining the CWT using AGB information shifts this range to between 27% and 58%. For savanna (UN-LCCs 62) the original cover fractions are within the constrained 95% CI. For woody cover, the fraction of deciduous forest (UN-LCCS 61) decreased from 85% to 58%. We observe an overall decrease for the woody cover fraction since the reference biomass is much higher than the actual biomass of most of the pixels.

Although the reference biomasses used in this study are in line with previously reported values (Carreira et al., 2013), disagreement between the original and the constrained CWT is considerable. The original CWT starts from the view that all ecosystems (except croplands) are in their natural state. The AGB map, however, does not contain any evidence in support of this view but rather suggests that 50% of the savanna (UN-LCCs 62) are 65% below their reference biomass. Likewise, 50% for dry woodland (UN-LCCs 60) are 71% below their reference biomass (Fig. 2). The AGB map thus suggests wide-spread degradation of these ecosystems which are in a highly anthropized region (Mitchard et al., 2013). Uncertainty coming from the reference biomasses could be reduced by field observations at the ecoregion or finer spatial scales.

For deciduous forest, however, the difference in cover fraction of woody vegetation between the original CWT and the constrained CWT could also be explained by an inaccurate estimation of the reference biomass due to a too coarse definition of the deciduous woody vegetation ranging from deciduous forest, over woodlands to savannas which are composed by different dominant tree species, with different biomasses (Sawadogo et al., 2010).

4.2.4 Xeric shrubland (UN-LCCS 100, 110, 120).

The woody cover fraction of xeric shrublands and grasslands ranged between 40 and 60% in the original CWT. Accounting for the information contained in the AGB map significantly decreased the woody cover fraction range toward 5,0 and 16%. Indeed, shrubs which represent a large part of the xeric shrublands were originally classified as woody vegetation for the ORCHIDEE model (i.e., when moving from the generic PFTs to the ORCHIDEE-specific PFTs; see section 2). This assumption is true from an ecological point of view but in a simplified world like in land surface models, xeric shrubland has an aboveground biomass that resembles cropland and grassland (Fig. 2). By overlaying the land cover type and aboveground biomass maps, 37% of the African shrublands were found to be degraded with a biomass of 2,7±1,5 t/ha, 54% were found to be intact with a biomass of 22±19 t/ha and 9% of the shrublands are thickets with a biomass of 68±11 t/ha. This is in line with other aboveground biomass estimates from remote sensing products (Saatchi et al., 2011; Mitchard et al., 2013; Avitabile et al., 2016) and in situ measurements where shrublands, degraded thicket, and intact thicket in south Africa accumulated 3, 24 and 102 t/ha of biomass respectively (Mills, et al., 2005). These findings suggest that in the model world, xeric shrubland is best represented by a large fraction of herbaceous plant functional groups, when the overall objective is to model AGB.

4.2.4 Sparse vegetation (UN-LCCS 150, 153).

The constrained cover fraction estimates are in line with the original CWT for UN-LCCS 150 which represent the most common class of sparse vegetation. The constrained cover fraction for UN-LCCS 153 has a larger herbaceous i.e., 29 to 97%, then the bare soil cover fraction, i.e., 2,0 to 61% contrary to the original CWT. The herbaceous cover fraction could be overestimated if a too low reference biomass was used. A reference biomass of 3,0 t/ha was used and is acceptable compared to the reported biomass for the Succulent and Nama Karoo Biomes ranging from 0.5 to 7,6 t/ha (Rutherford, 1978; Rutherford and Westfall, 1986). Given the current lack of reference biomass observations, disagreement between the original and constrained CWT could be resolved by using an independent estimate of bare soil fraction.

4.3 Consequences for land surface modelling

4.3.1 Which land cover types affect the biomass estimate?

The large disagreement in cover fraction estimates (30 to 40%) resulted in small disagreement in biomass, i.e., <1,0 t/ha in regions with little precipitation like Somali Acacia-Commiphora Bushlands and Thickets and the Kalahari Xeric Savanna. This counter-intuitive result is explained by the growth processes simulated in ORCHIDEE. Under xeric climate conditions

ORCHIDEE simulates low tree biomasses (< 2,0 t/ha) because the low precipitation and subsequent plant water availability results in a continuous high tree mortality. Nonetheless, in forest ecoregions like the eastern Guinean forests or in the Congo basin, where the sensitivity to a change in the cover fractions ranged from 1,0 to 5,0 t/ha/% and had a considerable impact on the simulation since a 15% uncertainty in the bare soil fraction may lead to a 75 t/ha uncertainty of the biomass in the tropical forest of the Congo basin. Underestimating the forest cover in humid ecoregions will have a much larger consequence on the simulated AGB than overestimating the forest cover in xeric ecoregions. The uncertainty surrounding the land cover fractions should thus be further reduced for the land cover types that already come with the lowest uncertainty, i.e., the forests.

4.3.2 Which land cover types affect the albedo estimate?

As for AGB, uncertainties in land cover fractions are only partly reflected in the uncertainties of the visible albedo. Dampening is caused by the fact that the reflectivity of grassland (0.06), cropland (0.06) are close to the leaf reflectivity of a forest (0.03 to 0.04) compared to bare soils reflectivity (0.1 to 0.25 depending on the colour of the soil) in ORCHIDEE. By increasing the bare soil cover fraction, the albedo will increase accordingly but changing forest into grassland will not drastically change albedo. The most sensitive area is the western tropical forest in the Congo basin for which a 15% change in bare soil cover fraction may trigger a 15% change in the visible albedo (fig 3C). Similar as for AGB, the uncertainty surrounding the land cover fractions of the forested land cover types should be further reduced to reduce the uncertainty of the model simulations.

4.4 Outlook

In this study a single biomass map was used as this enabled keeping the focus on the method itself. Nevertheless, other biomass products are available (Saatchi et al., 2011; Baccini et al., 2012; Avitabile et al., 2016; Santoro et al., 2021) and could have been used. Repeating this study for each of these biomass products would add another source of uncertainty to the cross-walking table. Owing to the method presented in this study, this uncertainty could then be propagated into the PFT map and all the way up to the simulated biomass, albedo -as done in this study for one biomass product- and other land surface properties. Considering different biomass products would give an insight of the impact of satellite-based biomass estimates on the discretisation of the vegetation and by extension surface properties as estimated by land surface models. Likewise, a single land cover map has been used in our analysis, but other products are available as well (Copernicus, UN-spider, Li et al., 2020). By using different land cover maps, one could quantify the uncertainty in the land cover classification and propagate it to evaluate its impact on the simulated land surface properties.

Compared to other continents, the Africa vegetation has been documented by relatively few quantitative observations (Mills, et al., 2005; Saatchi et al., 2011; Asner et al., 2012; Réjou-Méchain et al., 2015). Hence, it is the continent where remote sensing data could largely enhance our knowledge on the issue. Recent high-resolution satellite observations bear the promise to significantly reduce the credible interval around the aboveground carbon stock to estimate the CO_2 emissions from tropical forests (Hansen et al., 2013; Bouvet et al., 2018; Defourny et al., 2019; Buchhorn et al., 2020) but land surface models will need to be ready to routinely assimilate these data to fully benefit from the information contained in biomass maps. This study demonstrated one way of how satellite-based biomass data can help modellers to constraint the initialization process by means of refining the cross-walking tables that are used to map land cover classes derived from satellite observations into PFT maps. Nevertheless, biomass maps could be used for applications other than model initialization (this study), including model parameterisation and model evaluation.

The biomass map could be used to optimise model parameters related to growth, turnover and mortality to better simulate the vegetation biomass for the different PFTs. The evaluation stage could benefit from the biomass maps by benchmarking the model results against observed relationships between biomass-climate and biomass-land-use to better distinguish and simulate the difference between actual and potential biomass (Sankaran et al., 2005). Although the availability of several biomass products makes it possible to use one product to inform the cross-walking tables and another product to evaluate the simulated surface properties, the magnitude of present-day differences between biomass products (Mitchard et al., 2013) is expected to result in major inconsistencies when different biomass products are used for different purposes (e.g., assimilation, parameterization, evaluation) into a single analysis. In this study, less than 0,01% (see 2.3.1) of the information contained in the biomass map was used to constraint the cross-walking table and none was used to optimise model parameters. The simulated biomass remains, therefore, largely independent from the biomass map which implies that a single biomass map can be used for land cover optimisation (as in this study), and in a second step for parameter optimization or model evaluation.

With an increase in resolution of the land cover map comes a decrease in the reliance on the cross-walking tables. Cross walking tables will no longer be required once the resolution will be high enough (around 10 x 10 m) such that each pixel contains a single vegetation type equivalent to a single PFT classification used by land surface models (Li et al., 2020). No longer having to rely on cross walking tables would likely reduce the width of the credible intervals of the PFT map. As there would no longer be a need to estimate woody and herbaceous fractions, there would no longer be a need for the information contained in the biomass map. It will then be feasible to solely use biomass maps to better parameterize the processes that contribute to simulating the reference biomass. It should be noted, however, that higher resolutions will not solve the basic challenge of discretizing vegetation. High resolution land cover maps would split structurally complex ecosystems, for example savannas, into a pure forest fraction and a pure grassland fraction. This would overlook the interactions between the grasses and the trees which are among the defining ecological characteristics of a savanna.

Finally, we should note that other satellite-derived products than the AGB could be used to constrain the mapping of the land cover classes into model PFTs (i.e., CWT). For instance, the global tree cover fraction map, at 30-meter resolution, from Hansen et al., (2013) could also be used to constraint the fraction of bare soil within each land cover class like what was done in this study with the AGB map.

4.5 Conclusion

This study demonstrates how an aboveground biomass map could be used to constrain a cross-walking table that enables remapping land cover types derived from satellite-observations into plant functional types used as a boundary condition in land surface models. Given that previous cross-walking tables did not report uncertainties as they were mostly based on expert-knowledge, it remains unclear how much the use of an additional constraint really improved the cross-walking tables. Nevertheless, the considerable uncertainties remaining in the cross-walking table that made use of the aboveground biomass map suggests that total biomass map should be complemented with a bare soil map to better constrain the cross-walking table. Likewise, the reference biomass for both herbaceous and woody vegetation need to be constrained to at least the ecoregion scale to avoid underestimating or overestimating bare soil fractions. The method developed in this study helped to estimate the uncertainty of cross-walking tables which can now be used to benchmark further methodological developments. Moreover, the method identified bare soil cover fraction would be required to reduce the uncertainty of future cross-walking tables and the plant functional type maps they generate.

5 Acknowledgements

This study was primarily financed by the French space agency, Centre National d'Etude Spatiale (CNES), through the "BIOMASS-Valorisation" project from the TOSCA research program which contributed to the funding of Guillaume Marie and Cécile Dardel. The Marie Sklodowska Curie Fellowship CLIMPRO (MSCA-Fellowship EU 895455) partly funded Guillaume Marie.

6 Data availability

- CESBIO African AGB map. Biomass map of Africa created by CESBIO can be downloaded at https://www.theia-land.fr/en/product/african-biomass-map. It consists of a GIF file in which Africa is spatially discretized in pixels of 1x1km. The unit is a ton of dry mass per hectare (t/ha). Contact person: alexandre.bouvet@cesbio.cnes.fr
- The land cover map is freely available from: http://www.esa-landcover-cci.org.
- The ecoregion map used in this study is freely available from: https://databasin.org/datasets/68635d7c77f1475f9b6c1d1dbe0a4c4c/

7 Code availability

- All R scripts and ORCHIDEE tags 2.0 (rev 6592) source code is available at: https://zenodo.org/badge/latestdoi/345907299 or DOI: 10.5281/zenodo.4785328

- ORCHIDEE tags 2.0 (rev 6592) code is also available from: https://forge.ipsl.jussieu.fr/orchidee/wiki/GroupActivities/CodeAvalaibilityPublication/ORCHIDEE_tags_2.0_gmd_2021_Africabrowser/tags/ORCHIDEE_2_1

8 Author contribution:

G. Marie, S. Luyssaert and P. Peylin designed the experiments and G. Marie carried them out. G. Marie developed the OPENBUGS model code and performed the simulations. G. Marie and S. Luyssaert prepared the manuscript with contributions from all co-authors.

9 References

ALOS - EoPortal Directory - Satellite Missions, earth.esa.int/web/eoportal/satellite-missions/a/alos. *Tools*, orchidas.lsce.ipsl.fr/dev/lccci/tools.php.

Asner, Gregory P., et al., "A Universal Airborne LiDAR Approach for Tropical Forest Carbon Mapping." *Oecologia*, vol. 168, no. 4, 2011, pp. 1147–1160., doi:10.1007/s00442-011-2165-z.

Avitabile, Valerio, et al., "An Integrated Pan-Tropical Biomass Map Using Multiple Reference Datasets." *Global Change Biology*, vol. 22, no. 4, 2016, pp. 1406–1420., doi:10.1111/gcb.13139.

Baccini, A., et al., "Estimated Carbon Dioxide Emissions from Tropical Deforestation Improved by Carbon-Density Maps." *Nature Climate Change*, vol. 2, no. 3, 2012, pp. 182–185., doi:10.1038/nclimate1354.

Bauer, Peter, et al., "The Quiet Revolution of Numerical Weather Prediction." *Nature*, vol. 525, no. 7567, 2015, pp. 47–55., doi:10.1038/nature14956.

Beech, E., et al., "GlobalTreeSearch: The First Complete Global Database of Tree Species and Country Distributions." *Journal of Sustainable Forestry*, vol. 36, no. 5, 2017, pp. 454–489., doi:10.1080/10549811.2017.1310049.

Bonan, Gordon B., et al., "Landscapes as Patches of Plant Functional Types: An Integrating Concept for Climate and Ecosystem Models." *Global Biogeochemical Cycles*, vol. 16, no. 2, 2002, doi:10.1029/2000gb001360.

Boucher, Olivier, et al., "Presentation and Evaluation of the IPSL-CM6A-LR Climate Model." Journal of Advances in Modeling Earth Systems, vol. 12, no. 7, 2020, p. e2019MS002010, doi:https://doi.org/10.1029/2019MS002010.

Bouvet, Alexandre, et al., "An Above-Ground Biomass Map of African Savannahs and Woodlands at 25 m Resolution Derived from ALOS PALSAR." *Remote Sensing of Environment*, vol. 206, 2018, pp. 156–173., doi:10.1016/j.rse.2017.12.030.

Brovkin, Victor, et al., "A Continuous Climate-Vegetation Classification for Use in Climate-Biosphere Studies." *Ecological Modelling*, vol. 101, no. 2-3, 1997, pp. 251–261., doi:10.1016/s0304-3800(97)00049-5.

Buchhorn, Marcel, et al., "Copernicus Global Land Cover Layers—Collection 2." *Remote Sensing*, vol. 12, no. 6, 2020, p. 1044., doi:10.3390/rs12061044.

Buitenwerf, R., et al., "Increased Tree Densities in South African Savannas: >50 Years of Data Suggests CO2 as a Driver." *Global Change Biology*, vol. 18, no. 2, 2011, pp. 675–684., doi:10.1111/j.1365-2486.2011.02561.x.

Calzadilla, Alvaro, et al., “Climate Change and Agriculture: Impacts and Adaptation Options in South Africa.” *Water Resources and Economics*, vol. 5, 2014, pp. 24–48., doi:10.1016/j.wre.2014.03.001.

Carreira, Valeria Paula, et al., “Gene-by-Temperature Interactions and Candidate Plasticity Genes for Morphological Traits in Drosophila Melanogaster.” *PLoS ONE*, vol. 8, no. 7, 2013, doi:10.1371/journal.pone.0070851.

Chapin, F. Stuart, et al., “Plant Functional Types as Predictors of Transient Responses of Arctic Vegetation to Global Change.” Journal of Vegetation Science, vol. 7, no. 3, 1996, pp. 347–58, doi:10.2307/3236278.

Clark, D. B., et al., “The Joint UK Land Environment Simulator (JULES), Model Description – Part 2: Carbon Fluxes and Vegetation Dynamics.” *Geoscientific Model Development*, vol. 4, no. 3, 2011, pp. 701–722., doi:10.5194/gmd-4-701-2011.

Defourny, Pierre, and ESA Land Cover CCI project team. *Dataset Record: ESA Land Cover Climate Change Initiative (Land_Cover_cci): Global Land Cover Maps, Version 2.0.7*, 28 Nov. 2019, catalogue.ceda.ac.uk/uuid/b382ebe6679d44b8b0e68ea4ef4b701c.

“Development Activities.” *Documentation/Forcings – ORCHIDEE*, forge.ipsl.jussieu.fr/orchidee/wiki/Documentation/Forcings.

Dietze, Michael C., et al., “Iterative Near-Term Ecological Forecasting: Needs, Opportunities, and Challenges.” *Proceedings of the National Academy of Sciences*, vol. 115, no. 7, 2018, pp. 1424–1432., doi:10.1073/pnas.1710231115.

Di Gregorio, A. (2005). Land cover classification system: classification concepts and user manual: LCCS (Vol. 2). Food & Agriculture Org.

Druel, A., Peylin, P., Krinner, G., Ciais, P., Viovy, N., Peregon, A., ... & Mironycheva-Tokareva, N. (2017). Towards a more detailed representation of high-latitude vegetation in the global land surface model ORCHIDEE (ORC-HL-VEGv1. 0). *Geoscientific Model Development*, *10*(12), 4693-4722.

Dubayah, Ralph, et al., “The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth’s Forests and Topography.” Science of Remote Sensing, vol. 1, June 2020, p. 100002, doi:10.1016/j.srs.2020.100002.

Eamus, Derek, et al., “Savannas.” *Vegetation Dynamics*, pp. 383–414., doi:10.1017/cbo9781107286221.017.

Eigentler, L., and J.a. Sherratt. “Spatial Self-Organisation Enables Species Coexistence in a Model for Savanna Ecosystems.” *Journal of Theoretical Biology*, vol. 487, 2020, p. 110122., doi:10.1016/j.jtbi.2019.110122.

Ellison, Aaron M. “Bayesian Inference in Ecology.” *Ecology Letters*, vol. 7, no. 6, 2004, pp. 509–520., doi:10.1111/j.1461-0248.2004.00603.x.

“FAO.org.” *Land Cover Classification System (LCCS), Food and Agriculture Organization of the United Nations*, www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1036361/.

Friedlingstein, Pierre, et al., “Global Carbon Budget 2020.” *Earth System Science Data*, vol. 12, no. 4, Dec. 2020, pp. 3269–340, doi:10.5194/essd-12-3269-2020.

GDAL/OGR contributors (2022). GDAL/OGR Geospatial Data Abstraction Software Library. Open-Source Geospatial Foundation. URL https://gdal.org

Hansen, M. C., et al., "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science*, vol. 342, no. 6160, 2013, pp. 850–853., doi:10.1126/science.1244693.

Hardin, G. "The Competitive Exclusion Principle." *Science*, vol. 131, no. 3409, 1960, pp. 1292–1297., doi:10.1126/science.131.3409.1292.

Hartley, A. J., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant functional type distributions and its impact on land surface models. *Remote Sensing of Environment*, *203*, 71-89.

"Home." *ESA Climate Office*, climate.esa.int/en/.

"Home." *ORCHIDEE*, orchidee.ipsl.fr/.

Houghton, R. A., et al., "Carbon Emissions from Land Use and Land-Cover Change." *Biogeosciences*, vol. 9, no. 12, 2012, pp. 5125–5142., doi:10.5194/bg-9-5125-2012.

Huete, Alfredo. "Vegetation's Responses to Climate Variability." Nature, vol. 531, no. 7593, Mar. 2016, pp. 181–82, doi:10.1038/nature17301.

Hurtt, G. C., et al., "Linking Models and Data on Vegetation Structure." *Journal of Geophysical Research: Biogeosciences*, vol. 115, no. G2, 2010, doi:10.1029/2009jg000937.

"Japan Aerospace Exploration Agency." *JAXA*, global.jaxa.jp/.

Kearsley, Elizabeth, et al., "Conventional Tree Height–Diameter Relationships Significantly Overestimate Aboveground Carbon Stocks in the Central Congo Basin." Nature Communications, vol. 4, no. 1, Aug. 2013, p. 2269, doi:10.1038/ncomms3269.

Krinner, G., et al., "A Dynamic Global Vegetation Model for Studies of the Coupled Atmosphere-Biosphere System." *Global Biogeochemical Cycles*, vol. 19, no. 1, 2005, doi:10.1029/2003gb002199.

Le Toan, T., et al., "The BIOMASS Mission: Mapping Global Forest Biomass to Better Understand the Terrestrial Carbon Cycle." Remote Sensing of Environment, vol. 115, no. 11, Nov. 2011, pp. 2850–60, doi:10.1016/j.rse.2011.03.020.

Lewis, Simon L., et al., "Increasing Carbon Storage in Intact African Tropical Forests." *Nature*, vol. 457, no. 7232, 2009, pp. 1003–1006., doi:10.1038/nature07771.

Lewis, Sophie C., et al., "Modeling Insights into Deuterium Excess as an Indicator of Water Vapor Source Conditions." *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 2, 2013, pp. 243–262., doi:10.1029/2012jd017804.

Li, Qingyu, et al., "Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine." *Remote Sensing*, vol. 12, no. 4, 2020, p. 602., doi:10.3390/rs12040602.

Li, Wei, et al., "Major Forest Changes and Land Cover Transitions Based on Plant Functional Types Derived from the ESA CCI Land Cover Product." International Journal of Applied Earth Observation and Geoinformation, vol. 47, May 2016, pp. 30–39, doi:10.1016/j.jag.2015.12.006.

Li, Wei, et al., “Gross and Net Land Cover Changes in the Main Plant Functional Types Derived from the Annual ESA CCI Land Cover Maps (1992–2015).” Earth System Science Data, vol. 10, no. 1, Jan. 2018, pp. 219–34, doi:https://doi.org/10.5194/essd-10-219-2018.

Liang, Shunlin. *Advances in Land Remote Sensing: System, Modelling, Inversion and Application*. Springer, 2008.

Lurton, Thibaut, et al., “Implementation of the CMIP6 Forcing Data in the IPSL-CM6A-LR Model.” Journal of Advances in Modeling Earth Systems, vol. 12, no. 4, 2020, p. e2019MS001940, doi:https://doi.org/10.1029/2019MS001940.

Mermoz, Stéphane, et al., “Decrease of L-Band SAR Backscatter with Biomass of Dense Forests.” Remote Sensing of Environment, vol. 159, Mar. 2015, pp. 307–17, doi:10.1016/j.rse.2014.12.019.

Mills, A. J., et al., “Ecosystem Carbon Storage under Different Land Uses in Three Semi-Arid Shrublands and a Mesic Grassland in South Africa.” *South African Journal of Plant and Soil*, vol. 22, no. 3, 2005, pp. 183–190., doi:10.1080/02571862.2005.10634705.

Mills, A. J., et al., “Ecosystem Carbon Storage under Different Land Uses in Three Semi-Arid Shrublands and a Mesic Grassland in South Africa.” *South African Journal of Plant and Soil*, vol. 22, no. 3, 2005, pp. 183–190., doi:10.1080/02571862.2005.10634705.

Mitchard, Edward T. A., et al., “Markedly Divergent Estimates of Amazon Forest Carbon Density from Ground Plots and Satellites.” *Global Ecology and Biogeography*, vol. 23, no. 8, 2014, pp. 935–946., doi:10.1111/geb.12168.

Mitchard, Edward Ta, et al., “Uncertainty in the Spatial Distribution of Tropical Forest Biomass: A Comparison of Pan-Tropical Maps.” *Carbon Balance and Management*, vol. 8, no. 1, 2013, doi:10.1186/1750-0680-8-10.

Naidoo, Laven, et al., “Savannah Woody Structure Modelling and Mapping Using Multi-Frequency (X-, C- and L-Band) Synthetic Aperture Radar Data.” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 105, July 2015, pp. 234–50, doi:10.1016/j.isprsjprs.2015.04.007.

NiSAR : First global SAR mission, in partnership with ISRO, https://nisar.jpl.nasa.gov .

O'connor, Tim G, et al., “Bush Encroachment in Southern Africa: Changes and Causes.” *African Journal of Range & Forage Science*, vol. 31, no. 2, 2014, pp. 67–88., doi:10.2989/10220119.2014.939996.

Oke, T. R. Boundary Layer Climates. Routledge, 2002.

Organisation Des Nations Unies Pour L'alimentation Et L'agriculture. *Food and Agriculture Organization of the United Nations*, www.fao.org/home/fr/.

Palmer, Paul I., et al., “Net Carbon Emissions from African Biosphere Dominate Pan-Tropical Atmospheric CO2 Signal.” *Nature Communications*, vol. 10, no. 1, 2019, doi:10.1038/s41467-019-11097-w.

Poulter, B., et al., “Plant Functional Type Mapping for Earth System Models.” *Geoscientific Model Development*, vol. 4, no. 4, 2011, pp. 993–1010., doi:10.5194/gmd-4-993-2011.

Pütz, Sandro, et al., “Long-Term Carbon Loss in Fragmented Neotropical Forests.” *Nature Communications*, vol. 5, no. 1, 2014, doi:10.1038/ncomms6037.

Quegan, Shaun, et al., “The European Space Agency BIOMASS Mission: Measuring Forest Above-Ground Biomass from Space.” Remote Sensing of Environment, vol. 227, June 2019, pp. 44–60, doi:10.1016/j.rse.2019.03.032.

Quéré, C. Le, et al., “Global Carbon Budget 2015.” *Earth System Science Data*, vol. 7, no. 2, 2015, pp. 349–396., doi:10.5194/essd-7-349-2015.

Réjou-Méchain, Maxime, et al., “Using Repeated Small-Footprint LiDAR Acquisitions to Infer Spatial and Temporal Variations of a High-Biomass Neotropical Forest.” *Remote Sensing of Environment*, vol. 169, 2015, pp. 93–101., doi:10.1016/j.rse.2015.08.001.

Rietkerk, M., Boerlijst, M. C., van Langevelde, F., HilleRisLambers, R., de Koppel, J. V., Kumar, L., ... & de Roos, A. M. (2002). Self-organization of vegetation in arid ecosystems. *The American Naturalist*, *160*(4), 524-530.

Saatchi, S. S., et al., “Benchmark Map of Forest Carbon Stocks in Tropical Regions across Three Continents.” *Proceedings of the National Academy of Sciences*, vol. 108, no. 24, 2011, pp. 9899–9904., doi:10.1073/pnas.1019576108.

Sankaran, Mahesh, et al., “Determinants of Woody Cover in African Savannas.” *Nature*, vol. 438, no. 7069, 2005, pp. 846–849., doi:10.1038/nature04070.

Santoro, Maurizio, et al., “The Global Forest Above-Ground Biomass Pool for 2010 Estimated from High-Resolution Satellite Observations.” Earth System Science Data Discussions, August 2021, pp. 1–38, doi:10.5194/essd-13-3927-2021

Schaaf: MODIS Albedo and Reflectance Anisotropy Products. Accessed 24 Mar. 2021.

Shimada, Masanobu, and Takahiro Ohtaki. “Generating Large-Scale High-Quality SAR Mosaic Datasets: Application to PALSAR Data for Global Monitoring.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, no. 4, 2010, pp. 637–656., doi:10.1109/jstars.2010.2077619.

Simard, Marc, et al., “Mapping Forest Canopy Height Globally with Spaceborne Lidar.” Journal of Geophysical Research: Biogeosciences, vol. 116, no. G4, 2011, doi:https://doi.org/10.1029/2011JG001708.

Sitch, S., et al., “Evaluation of Ecosystem Dynamics, Plant Geography and Terrestrial Carbon Cycling in the LPJ Dynamic Global Vegetation Model.” *Global Change Biology*, vol. 9, no. 2, 2003, pp. 161–185., doi:10.1046/j.1365-2486.2003.00569.x.

Still, Christopher J., et al., “Global Distribution of C3 and C4 Vegetation: Carbon Cycle Implications.” Global Biogeochemical Cycles, vol. 17, no. 1, 2003, pp. 6-1-6–14, doi:https://doi.org/10.1029/2001GB001807.

Sun, Wanxiao, et al., “Mapping Plant Functional Types from MODIS Data Using Multisource Evidential Reasoning.” Remote Sensing of Environment, vol. 112, no. 3, Mar. 2008, pp. 1010–24, doi:10.1016/j.rse.2007.07.022.

Thomas, Neal, 2010. "Overview". OpenBUGS website. Retrieved 9 October 2010.

Tyukavina, Alexandra, et al., “Congo Basin Forest Loss Dominated by Increasing Smallholder Clearing.” *Science Advances*, vol. 4, no. 11, 2018, doi:10.1126/sciadv.aat2993.

UN-spider: "CCI LAND COVER - S2 Prototype Land Cover 20m Map of Africa." *UN*, un-spider.org/links-and-resources/data-sources/cci-land-cover-s2-prototype-land-cover-20m-map-africa.

Viovy, Nicolas, 2017, CRUNCEP data set: https://vesg.ipsl.upmc.fr/thredds/fileServer/work/p529viov/cruncep/readme.html

Wigley, B. J., et al., "Bush Encroachment under Three Contrasting Land-Use Practices in a Mesic South African Savanna." *African Journal of Ecology*, vol. 47, 2009, pp. 62–70., doi:10.1111/j.1365-2028.2008.01051.x.