This paper uses Random Forest machine learning trained on CCI-Biomass AGB data with spaceborne K- and C-band scatterometer data and mean climate variables to produce a dataset of annual biomass maps running from 1993 to 2020 at a mapping scale of 8.9 km. This is used to estimate above ground carbon and total vegetation carbon globally and within biomes, and to examine the trends present in the dataset. It is potentially valuable because it is the longest such dataset currently available, and it is unlikely that any longer comparable datasets will be produced as satellite data to do so are not available. Weaknesses of the paper include: (1) The fact that it is just regression-based and no physical reasoning is given as to why such short wavelength radar is able to provide meaningful AGB estimates when it scatters mainly in the upper canopy; (2) The training data from CCI-Biomass are known to contain biases that will presumably transfer into this dataset. This needs a fuller discussion, including indication that the authors know something about these biases; (3) The reference inventory data indicate significant biases in the dataset. The implications of these biases for their estimates of stocks and stock change need to be clarified.

It is generally well written and clear, though with some small English errors, and there is systematic misuse of words like “dynamic” (see below). The figures and tables are generally useful and clear.

Detailed comments follow

Th abstract asserts that the estimate of global AGC aligns with other estimates, but it is 20% less than several. It also asserts that this dataset is a benchmark, which implies it can be taken as a valid reference. However, this is not supported by Fig. 1 which shows significant and systematic biases compared to inventory data and by the fact that the training data (CCI-Biomass) is known to exhibit biases, particularly for high and low AGB. It also says the dataset is temporally continuous; it is not, it is annual and this should be stated.

l. 47-48 The sentence makes no English sense; in addition, and it talks about 10 years when the data period quoted is 16 years.

l. 58-59 this sentence has something missing or is miswritten.

l. 64 (& 361) What is meant by space for time substitution?

l. 80. Where does the 8.9 km resolution come from? L. 358 says that the resolution of the scatterometer data is ~25 km so how is the gain in resolution achieved? Or is it not resolution but instead grid spacing (which implies high correlation between grid cells).

84. It says 7 biomes but only lists 6.

91. Many of 100 m pixels in the CCI dataset in the tropics have CV > 80%, so this exclusion is likely to distort the tropical estimates.

96. What is the meaning of biomass_ag_OM, AGB_ini, AGB_fin, Initial AGB? How can one aggregate inventory data to 9.9 km? And how is “sufficient” defined for sampling density.

108. Longer wavelength radar backscatter (L-band and longer wavelengths) is considered useful for estimating AGB, up to a point, due to penetration to the woody parts of the forests, but C- and K-band are not since they scatter in the upper canopy. So what is the rationale for estimating AGB from these shorter wavelengths as is is done in this paper?

110 & 132 Soil moisture, not surface water, is the more problematic confusing effect. There is little direct evidence relating vegetation water to backscatter at this scale. Also important is that the large scale soil moisture products are based on the same scatterometer data that are here being used to estimate AGC. This is likely to be the reason behind the correlations shown in Fig. S3.

111. Generally agreed? By whom? More importantly, why should this be true? And what effect is being referred to in l. 112?

135 Here we start to talk about pixels instead of grid cells – are these the same? What is the population being talked and for which the SD is calculated?

157. The CCI dataset has known biases: how are these accounted for?

Para 3.1. How well does a model perform that is based solely on climate? How much does the scatterometer data add?

Table 1 It's naïve to include the BIC to 7 significant figures.

216. This para is the wrong way round: the dataset underestimates high AGB values and overestimates low ones. This probably reflects biases introduced form CCI which also shows such effects. Name the 2 datasets used for comparison.

224. How is it possible for there to be significant IAV in forests in arid and semi-arid regions? This is effect is highly likely to be because the scatterometer data are particularly sensitive to soil moisture in such regions, and this is a soil moisture, not AGB, effect.

Fig. 1 The inventory data are different between the 3 plots. Why is this and how does this affect the results. The keys quote a bias, but in what quantity? What are the solid lines (I assume best linear fits)?

Fig.2 What does the SD represent in 2C? The error bars shown are not credible as indictors of the true uncertainties in these estimates.

260 What is meant by temporally consistent?

272. I don't know if the authors are looking at a different plot, but 100-300 is not consistent with Fig. 1A.

275. What are these structural differences and why is bias inevitable? What would be the sources of any bias, and would it be in the inventory data or the estimator used in this paper?

280 the paper should remind the reader that CCI is used to train the algorithm used. The para. mentions their estimates being "slightly" smaller than some others, but in several cases they are significantly lower (20-25%). The last sentence is not very sensible; inventories don't take place in wetlands, barren land etc.

Table 2. Where does the value for 1990 come from? It's not in the period covered by the dataset.

295 delete dynamics

333. I what sense is the dataset robust?

337. Earlier the paper says Ku-band scatt data are used, not just C-band. Which is correct?

344. What is meant by “spatial heterogeneity in calibration data” for CCI -Biomass? Which calibration data are being referred to, since CCI-Biomass does not use in situ data for calibration?

342. What is meant by uncertainty throughout this paragraph? For example, what actual quantity is displayed in Fig. 5? If this is just precision then it tells us little about error or bias and is misleading as regards the overall uncertainty in the dataset.

347. should say backscatter. What the forest structure predictors referred to here?

356. It the resolution of the data is 25 km, then it is wrong to say the spatial resolution of the dataset is 8.9 km. This is the interpolated grid spacing but the underlying resolution is unchanged.

362. What is meant by annually consistent?

368 get rid of dynamical in 4.3; it is just annual

369. these are not “direct” observations of biomass changes but very indirect estimates using data that are fairly insensitive to woody biomass.

372 these are not continuous temporal observations but just annual estimates

Some general remarks on terminology:

The word “dynamic” is misused: most of its occurrences in the paper should be replaced by “annual”.

The word “continuous” is used in several places and it is often unclear whether this means spatially or temporally continuous, e.g. l.39, l.46, but many others. But it mainly serves no purpose: by definition a map will be spatially continuous, and temporally continuous really means annual in this paper.

The paper sometimes talks about grid cells and sometimes about pixels and it’s not clear if these are synonymous; use of a single term would be better (unless these indicate different things).

In many places the paper needs to be specific and say “scatterometer” rather than just “radar”, e.g., l. 89, 178, but many others.