This is an excellent and interesting study, presenting a fascinating comparison of LiDAR data separated by 10 years over a very large (50 ha) and exceedingly well studied field plot. I am not aware of any analysis of lidar data in the tropics having been performed over such a long time period (10 years), nor with field data available at a similar time point to both LiDAR acquisitions. This represents an excellent opportunity to assess the abilities of LiDAR to see changes in biomass in old growth forest.

In theory this should be a relatively simple study: the ability of LiDAR to detect forest structure, and thus AGB, is well known, so this should enable easy comparison of two biomass maps at a variety of resolutions, with an accuracy assessment based on the field data. However, the data analysis is complicated due to the very different sensor characteristics of the two LiDAR datasets: LVIS is a full waveform system with 20 m wide footprints, whereas the DRL system is a discreet return LiDAR with a footprint size two orders of magnitude smaller. There are further difficulties due to the geographic uncertainties in the location of both LiDAR datasets and ground control points, all of which are at best likely to only be known with an accuracy of a few meters. However, despite these difficulties, I believe the authors have done an excellent job in processing these data layers, and draw important and useful conclusions.

I believe the paper should be published following some modifications, mostly related to the way the data has been presented, and some analyses with minimal validity.

Major comments:

1. I worry that dividing the data into 20 x 20 m subplots is of questionable value. This is for two reasons:
    a) LVIS footprints are already 20 m. The authors have used the central point of the footprints to decide whether or not the LVIS footprint should be included in a subplot or not, which is sensible, but in reality much of the LVIS footprint at this resolution will include data from outside the 20 x 20 m subplot. I do not think it is appropriate to compare 20 x 20 subplots to LVIS data. This is different for the DRL system – as long as the geolocation is good then going to small subplots for DRL is fine.
    b) Geolocation errors will dominate at the 20 x 20 m resolution. There will be errors in location of ground, DRL and LVIS systems; these are hard to quantify, but will be on the order of a few meters at best. These are relatively negligible at a 1 ha scale, and less dominant at the 0.25 ha scale, but at the 0.04 (20 m) scale they will dominate.

    The authors are aware of these limitations, and discuss them – but I think it would be better if the analysis was never done at this resolution as it doesn't make sense. There are additional reasons for not analysing at this scale to do with the scaling of tropical forests and the canopy size of larger trees (Chave et al. 2003), which are amply discussed by the authors both in the paper and Section 4 of the supplement – but the authors focus on these scaling problems while largely ignoring a) and b) above. I would therefore remove the left hand panels from Fig. 2 and 3 and not perform regressions at these resolutions.

    Alternatively these could be left in the paper if the authors feel they make an important point about the difficulties of using LiDAR data at this resolution, showing that there is not much improvement in AGB-detection between LVIS and DRL despite the higher resolution of the second system, but it should be made explicit in the Introduction and Methods that no good relationship would ever be expected at this resolution.

2. RS100max is not defined in the main paper. It is in the supplement, but I feel strongly that it should be explained in the paper as it is used in the LVIS regression equation (equation 1).

3. In this Discussion, page 1969 line 23, you state that "Differences in one height metric derived from Lidar waveforms … had no significant relationship with differences in forest biomass derived from the analysis of field surveys." This is a critically important finding, but it is hidden away in the discussion and no calculations, p-values or results are shown. I believe this analysis should be specifically presented in the methods and results, as it suggests that the link with biomass, and the use of multiple metrics from lidar, is essential for mapping change. This is not what one would expect necessarily from first principles: one might think that mapping changes in height metrics, especially at a coarser resolution, would correlate closely with biomass change: given this is not the case here, this is a very important result and should be justified and highlighted.

4. On page 1966 lines 16-20 and elsewhere the authors make much of the ability of LiDAR to see very similar change results at the scale of the whole plot as using the field data. However, this is not in the remotest surprising, and, unlike the results at a smaller scale, would be entirely expected given the inputs. The relationships between LiDAR and ground data were developed separately for 2000 and 2010, so any changes in the ground data would automatically be seen in the LiDAR data when the full plot is considered. These numbers can still be reported, but the text should clearly state that given the methodology (field data calibration available for both dates) an exact match it expected. This would be different if the calibration had been done using ground data from one time point only, or only a small subset of the plot data, but that is not the case.

Minor comments

Abstract line 16 (and later) it would be good to put a p-value on this sign of 36 out of 50 plots detected correctly. I calculate that using a binomial distribution, if you had a 0.5 chance of correctly guessing the direction of change of each hectare, you would get >= 36 right only 0.13 % of the time – in other words this suggests there's less than 1 % chance that you would get the sign right for 36/50 plots by chance alone. This (or other statistics if seen as more appropriate) should be included here.

Page 1959 line 13, should be 'wood specific gravity' or 'wood density', not a mixture of both.

Page 1966 line 3 – it's not stated but presumably the relative contributions of each metric is for the DRL data only? Could we see the figures for LVIS too?

Page 1967 line 17 – I do not understand the meaning of 'amplitude of differences'.

Page 1969 line 1-5 – there is no scientific validity in comparing one plot to the Baker et al. 2004 paper, which compared very many plots across the basin, a subset of which showed negative biomass change, but on average showed an increase in biomass. I do not think this sentence should be included – the comparison can be made, but this does not 'run against recently published results' because a single data point cannot possibly negate a study involving many plots. Pan et al. 2011 (Science) should probably also be cited here.

Page 1969 lines 18-21 – this paragraph could be extended to add in the evidence and theory that drought is associated with the death of large trees, so this is consistent with that hypothesis.

Supplementary information

Page 1: what altitude was LVIS flown at? Would be useful to include the incidence angle range for the two lidar datasets as well.

Page 6: Why is Figure S2 only shown for DRL data? Would be good and interesting to see this for LVIS too. Similarly it would be interesting to see the relationships between say RH75 or RH100 with AGB: all we have is the full model plotted against AGB. There is unlikely to be much relationship at the 0.04 ha, and this feeds into my major point about why the analysis was done at this resolution.