Biogeosciences Discuss., 6, S1221–S1226, 2009
www.biogeosciences-discuss.net/6/S1221/2009/

Biogeosciences Discussions

Interactive comment on "Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia" *by* S. Saatchi et al.

Anonymous Referee #3

Received and published: 14 July 2009

GENERAL COMMENT

This study represents a very interesting and useful study into the distribution of various biologically important variables across the Amazon basin, using a wide range of plots. The maps of fraction of large trees and fraction of palms are novel and interesting, and the maps of productivity, basal area and AGB also useful. The subject matter is relevant to Biogeosciences, and the methodology used and some of the maps produced are sufficiently novel to warrant publication. I recommend acceptance of the paper following some revisions. These need not necessarily need to lead to re-calculations and analysis of the data if the steps used can be justified more completely, but if they can’t then it is possible that some re-analysis will be necessary to produce better maps.

SPECIFIC COMMENTS

1. My major query relates to the use of the Maxent model to produce these maps from the (unquestionably good) ground data and remote sensing layers. This model was not designed for continuous distributions of the same variable, and the paper as written does not justify the assumption that the most accurate maps can be produced by a categorical technique designed for modelling species distributions: all these variables are continuous, and it thus to me it makes more sense that a continuous model will produce the best map. Dividing these continuous ranges into 3-4 broad classes, as the authors did, is necessary for Maxent, as it is necessary to have sufficient data points in each class. However doing so removes a huge quantity of information for the pain-stakingly collected RAINFOR data; to illustrate this with a random example a plot with a basal area of 22.9 will be seen by the model as identical to a plot with a basal area of 10, but a plot with a basal area of 23.1 will be put into an entirely different class, and thus it’s presence modelled using an entirely different set of criteria. A continuous model would see that the two are very close, and there would thus be no information lost due to arbitrary classes. I therefore believe the authors need to justify the use of a Maxent model much better in the introduction, and discuss it in relation to alternative, continuous models (for example Random Forest (as used successfully for biomass estimation across Africa in Bacinni et al 2008), generalized linear modelling, multivariate adaptive regression splines, alternating conditional expectations, neural networks etc). It is possible the authors may be able to justify the use of a categorical, class-based technique, by saying that the aim of the study is to produce approximate maps showing the trends. If so, then it is fairly easy to see why Maxent is used – though it is a black-box and therefore in some ways unideal, and largely untested for this kind of study, it is an exciting new technique that appears to use the available data layers very successfully. However, if the authors are unable to justify the use of this technique, I would recommend that recalculation of the data using a continuous model is necessary before this paper can be accepted for publication.

BGD

6, S1221–S1226, 2009

2. A smaller point, but still important, is that it would be good to have more information on how the class divisions were chosen before entering them into the Maxent model, and also on how the thresholding process was done. As this is a new application of this methodology it is important to be very clear on the process that was gone through so that these methods can be reproduced. In the case of the thresholds particularly more information is needed in order to reproduce these results: we are not told the criteria used to choose these thresholds, nor if the same threshold was used for each class within the same dataset.

3. A further issue with the paper is the accuracy assessment. It is not made quite clear in the paper exactly how this is achieved, but mention of made of a 60/40 split between input and training data. While this is useful for accuracy assessments (though it will tend to overestimate accuracy due to the clumped nature of the RAINFOR plots, a fact that should be mentioned in the paper), I would have ideally liked all the plots to contribute to the final models. I’m not sure how long these Maxent models take to run, but at a 5 km resolution I’d think it would be possible to run the models say ten times (or preferably even more) with a different set of data removed to perform the modelling, and then to average the resulting classes. This would give both a more robust accuracy assessment, and a more accurate map. The proportion of times which the eventually-chosen class was picked for each pixel between the different models would also produce a map showing an estimate of the robustness of the model in space, an excellent idea which the first author used very successfully in his Saatchi et al. 2007 GCB paper.

4. I also question the relationship between a 1 ha plot and their surrounding 5 km pixel that is assumed by this analysis. This is always going to be a problem with this kind of analysis, but forest are very heterogeneous and so this is a very real problem, that is likely to introduce a number of errors. It would be interesting if the authors could include information on the differences in these values between plots that fall within 5 km of each other, as this could give some information on the scale of this problem. It could

BGD

6, S1221–S1226, 2009

Interactive Comment

also be explored in relationship to the BCI or CTFS plots, which are large enough to be able to assess the variance of these variables for 1 ha plots within a larger area. I do not think this problem negates the worthiness and usefulness of the analysis produced, but I do think it should be discussed, and if possible the potential errors assessed using the methodologies described above.

5. I think that the AGB map produced is potentially very interesting, but discussion resulting from it is lacking. Why is it not compared to other AGB maps, for example those produced by the authors: Saatchi et al. 2007, Malhi et al. 2006, as well as others. As the biomass distribution of the Amazon is better studied than some of these other methods, this might provide a useful justification and validation of the modelling technique used, and if it is significantly different from these that needs to be discussed and explained.

6. I think it would be good to have another set of tables in the supplementary information that gave the variable contributions to each class and each data type, based on the jacknife analyses. These are included in snippets in the text in the case of the most important, but for the interested reader to be able to make a full assessment of the importance of these different variables across different classes it would be useful to include these.

TECHNICAL COMMENTS

P5463 line 14: should be ‘analyses were’ not ‘analysis are’

P5463 line 17: needs a ‘the’ before ‘dry season’

Final sentence of abstract: surely ‘areas of low elevation and [with a] high density of small rivers’ could be replaced by ‘flood plains’, or some other more ecological description – the description given of the areas appears to be derived merely from looking at remotely sensed imagery.

BGD

6, S1221–S1226, 2009

Interactive Comment

P5464 Lines 21-24: this sentence is not well written. I suggest this replacement ‘…change may have long-term effects on the forests structure and function by changing the morality and growth rates of trees and lianas, and by increasing the frequency of disturbance.’

Following sentence – surely one of the main reasons these changes will not happen uniformly is that anthropogenic disturbance is not uniform.

P5465 line 18: I think here there should be some discussion of what Maxent is and why you choose to use it, as well as of the alternatives – see the general comments above. Also some discussion of comparing 1 ha plots to 5 km pixels (i.e. the relationship between one semi-randomly chosen hectare plot to its surrounding 2500 hectares that is just assumed to be identical by the study as written).

P5466 Section 2.2: You say you choose LAI over vegetation indices, but it is not precisely clear why you could not also include something such as EVI or NDVI. Such vegetation indices would produce useful alternative information, particularly in terms of seasonality. However it is not a major worry, LAI probably contains a similar signal, but if these models were run again it would be interesting to include it, or at least test its correlation with LAI, in a similar manner to the VV and HH polarisations of Quickscat being compared.

P5471 Section 3.1: If you have not covered it in the introduction by this stage, there needs to be some discussion here of why you choose Maxent, and why you rejected alternatives. This is a two-stage process – first why you choose to use a categorical not continuous technique (this is the stage I am still quite unsure about), and second having chosen a categorical technique why you chose Maxent (much easier to justify).

P5473 Lines 3-7: More information is needed here on how this thresholding was done. We need to know that a rigorous procedure was gone through in order to balance ‘under and over prediction’. This is partially covered in Appendix A, but

BGD

6, S1221–S1226, 2009

not thoroughly (just giving an example rather than a methodology). Also this Appendix is not referred to in the text at the relevant place.

P5473 Section 3.2: More information needed here on how you split up the continuous distribution of your variables into classes. What was the reasoning behind these splits? Was there a minimum number of points you felt that could be used for each class? If was what was it? Did you try runs with more classes and find the results inaccurate? If other people are to use these methodologies we need to have this kind of information published.

P5479 Section 4.5: As discussed earlier, there needs to be a comparison to other published maps of Amazonian AGB here.

P5483 Line 8: ‘many’ should say ‘may’.

P5485 Line 11: WD should be ‘higher’ not ‘larger’.

Interactive comment on Biogeosciences Discuss., 6, 5461, 2009.