1 Reviewer 2

In this manuscript y et al. report on an interesting analysis on the environmental drivers behind gap size distribution in tropical forests, using Lidar data. Their methodology builds on previous work by Lobo and Dalling (2014), but they introduce an interesting new method to determine height thresholds and minimum gap sizes.

They rightfully criticize the empirical cutoffs that have been used, and come up with an innovative and interesting alternative. The authors further introduce Lidar-derived environmental parameters to include in the Bayesian model, which enables them to assess the effect of physical environment on gap size distribution.

Although I've enjoyed reading this interesting work, I think the authors could im prove/clarify their manuscript in some parts ; both for science and form/language

A first point would be the height cutoff ; the whole study is dependent on the value you set here. You set up a nice probabilistic model, but then you still go back to an empirical threshold (being 0,001percentile). If you have no theoretical considerations to justify this cutoff, than your method is as such not better than any other empirical threshold used in previous studies. Hence, this would need bit more clarification. I wonder whether you could not use the average of the second (lower) normal distribution as a cutoff. This would, I believe, also increase the applicability of your method in other forest types/regions, since the relative difference between canopy tree height/gap height will shift in other forest types. Even if you have a sound reason to use 0,001perc in this case, you would need to reselect a threshold in other forest types.

Thank you for this comment. If the threshold is chosen from the first distribution, then the expected value of the first law will be too dependant on the frequency of gap formation in time and this dependence is a problem. In other words, with a similar gap generation process (the same drop in height) a frequency of 1% of the forest area that will be transformed in gaps each year will generate a target value (the average of the first law as suggested) mechanically lower than for a landscape where the frequency is 0.5%. In the last case, the lower frequency of young gaps will produce on average a higher target value. We still do not want to depend on this gap frequency at the landscape scale.
We choose this threshold value to keep the maximum of information from the first distribution (gap height) while minmizing biases due to including too much information from the 2nd one (canopy height). Because the whole process needs several hours to be performed, from the GIS works to the model inference, with thousands of gaps, we cannot run a well-performed sensitivity analysis. However, we provide in supplementary informations , the parameter values for 2 additional thresholds (0.0001th and 0.01th). The posterior values of almost all variables are quite similar and thus do not change interpretation : Slope, TRI are always positive whatever the threshold TOPEX, HAND are always negative whatever the threshold DA and HAlt always include zero in the credibility interval

Secondly, I have not been working with Bayesian statistics myself, but I think the manuscript should be clear for the broad readership of Biogeosciences. Some questions related to the rest of your methodology :

— I wonder why you use equation 8 to constrain lambda. What is the reasoning behind an exponential model ?
  We have constrained the value of lambda in equation 8 because the linear combination, without the exponential constraint, may have result, during the inference process, in negative lambda values. And Riemann's Zeta function only admits values > 1.

— You explain the interpretation of Lambda on p 5 L 146-150. I think the authors are confused here (or maybe I am...) ; "lambda is not defined when lambda>=1 ? ? ", and "A value close to 1 means there are a large number of small gaps". Both of these statements seem absolutely wrong to me (I would expect the contrary with both), and they are actually vital to the interpretation of this very manuscript. Either I am wrong, but if not, I am a bit worried for the misinterpre-

tation of the results by the authors. Please have a look at this. Maybe this is also at the basis for the contradiction in some of the statements through the manuscript ;
I apologize for this mistake in the manuscrit. Thank you for pointing it out. The error has been corrected at line : "In forests dominated by small canopy openings, values of λ are larger, whereas smaller values of λ increase the frequency of large events (Fisher et al. 2008)". Line : 147

— Discussion (L296-297) : “”We found similar results to Lobo and Dalling in BCI ; i.e., large gaps are more frequent on gentle slopes”. Conclusion : “We expected that slope would also play an important role, with steeper slopes leading to larger gap sizes, but found the opposite effect.” Please go through the MS again and make sure the interpretations are the same, and are right, everywhere. If not, you fail to give the reader a clear take home message
the text has been modified for each section. Line 10, 305, 345

Thirdly, I have listed some other comments. The list is not complete ; some of these are clear typos or sloppiness. This would need to be avoided for your resubmission ... In general ; make sure results, M and M, and discussion are in the appropriate section, redo your subheadings, avoid typos, avoid repetition...

Specific comments

— P1 L10 : “we plan to scale up”
done line 10
— P2 L23 : a large quantity of leaf and wood litter becomes available. But please rephrase this anyway. It’s not a good sentence. Mineralisation and decomposition makes the nutrients available, not the wood and leaf litter available as such.
Rewritten. Line 24
— p4 L90 : the buffer you applied to anthropogenic tracks : is the Approuage an anthropogenic track ? For sure not masking out natural rivers out of your algorithm would hugely affect your results. I think (hope) you did include natural rivers in your buffers, but you would need to rephrase, since these are not anthropogenic...
Line 88 : Approuague is a large natural river. Indeed, we have applied a buffer in order to only account for areas (i) not affected by forest logging and (ii) natural rivers : *In order to remove areas close to natural rivers : a 20 m buffer was first applied to all shorelines. Then a 25 m buffer was applied to anthropogenic tracks.*
— P4 : Sloppiness ; your subtitles have the same rank on this page, while they all fall within the first subtitle ‘Environmental data’ Page 4 : Paragraphs are now correctly ranked. See the document Line 96
— P5 L 141-142 : what do you define as contiguous ? Diagonal pixels would be contiguous ? You know from the field that some trees may be left standing in certain gaps, so this could be important for your results . Line 144 : contiguous is defined as a pixel that has any contact with another i.e., a contact by edges or by vertices. In our framework, diagonal pixels are indeed considered as contiguous pixels.
— P6 l 153-154 : Please cite both R packages properly. Done. Line 157 : Most of the analysis was performed under R and making use of poweRlaw and VGAM packages.
— P6 L172 : Here you use X as the vector of covariates ; while on the next page in equation 8 you use varig. Please be consistent to make your MS more comprehensible. The vector of covariates as been properly replaced as X in equation 8.
— P7 L191-193 : For clarity I would rename the transformed variables. Also in eq 6 you use the HA as the new variable, and the hydraulic altitude in full as the old, while in 2.1.6. on p 4 you already use the HA abbreviation. Please correct these small errors for your future readers.
Done. Transformed variables have a new name. Line 195 : Halt, Topex
— P7 l 204 : investigated ; Material and methods should be in past tense. Please correct
Done. Line : 207
— Figure 4 : This figure does not have a lot of information. I would leave it out and describe in

text instead. We prefer maintaining this fig in the text but remains open to suggests from the editor.

— P14 L285 : why don't you show the values from the Kellner studies in brackets, like you do for Lobo and Asner ?Line 295, Done

— P14 L300 : Please add reference to your 75% statement. Line 306 : We changed this sentence

— P 14 l 309-311 : "Together with : : :" I don't get this sentence. Please rephrase : : : Line 320 : In agreement with Asner et al., (2013), our results suggest that we can effectively extend these results to bottomlands, where we already know that aboveground biomass and mean wood density are 10% lower than on hilltops (Ferry et al., 2010).

— P 15 l 330-331 : Really ? The first study ? And what about Lob and Dallin (a study which served as a basis for your study) Please review the reference list. Line 340 : Rewritten. "To our knowledge, this is the first study where the precise environmental descriptors associated to each canopy gap were explicitly taken into account in the general model likelihood. We were able to do so because we wrote general model likelihood as the product of all the single likelihoods (*i.e.* each gap had its own likelihood depending on the environmental covariate values). Doing so, we were able to predict gap size distribution from the fine environmental covariates, an impractical task when the scale exponent is estimated once at the forest level (*i.e.* mixing all the found gaps together) and compared between forests *a posteriori*"

2 Supplementary information

TABLE 1 – List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the 0.0001*th* percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95%)
Slope	Slope	°	0.119	[0.0416 ; 0.208]
Terrain Ruggedness Index	TRI	-	0.119	[0.083 ; 0.157]
TOPographic EXposure	TOPEX	-	-0.128	[-0.188 ; 0.00202]
Drained Area	DA	m^2	0.0843	[-0.0574 ; 0.179]
The Hydraulic Altitude	HAlt	m	-0.0135	[-0.04 ; 0.042]
HAND	HAND	-	-0.0615	[-0.152 ; 0.0162]

TABLE 2 – List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the 0.001*th* percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95%)
Slope	Slope	°	0.0735	[-0.02 ; 0.15]
Terrain Ruggedness Index	TRI	-	0.0718	[0.04 ; 0.10]
TOPographic EXposure	TOPEX	-	-0.082	[-0.12 ; -0.05]
Drained Area	DA	m^2	-0.0176	[-0.09 ; 0.05]
The Hydraulic Altitude	HAlt	m	-0.0177	[-0.05 ; 0.02]
HAND	HAND	-	-0.003	[-0.08 ; 0.09]

TABLE 3 – List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the $0.01th$ percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95%)
Slope	Slope	°	0.0975	[-0.02 ; 0.17]
Terrain Ruggedness Index	TRI	-	0.089	[0.05 ; 0.12]
TOPographic EXposure	TOPEX	-	-0.012	[-0.03 ; -0.32]
Drained Area	DA	m^2	-0.004	[-0.08 ; 0.05]
The Hydraulic Altitude	HAlt	m	0.063	[-0.04 ; 0.08]
HAND	HAND	-	-0.01	[-0.09 ; 0.06]