Reviewer 1

Generally, this study provides new and interesting insights in key environmental drivers of gap formation at landscape scale. While this study certainly deserves to be published, I have pointed a few issues that should be addressed before publication. The authors have developed an innovative analytic method to define gaps, but the entire analysis relies upon an a priori threshold equal to the 0.001th percentile of the estimated "natural variation of canopy height". While this choice may be well grounded, the rationale beyond it remains unexposed. How sensitive is the definition of gap and all subsequent results to this threshold? What if the authors had chosen the 0.01th percentile? Some kind of sensitivity analysis would make their choice more reliable.

We easily admit that our initial choice of the 0.001th percentile for the height threshold may have been seen as arbitrary. 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

More importantly, while a landscape scale approach seems meaningful to infer gap size distribution, this study highlights the importance of environmental factors on both gap frequency and size. Thus, I wonder if a fixed definition of gap remains meaningful, or if that definition should not adapt to the different forest types and/or main topographical features found at large scale. Doing so would point towards a more "ecological" definition of gaps, instead of a pure remote-sensing approach, and ultimately raises the question of the aim of detecting gaps. For instance, does a 100m2 gap in waterlogged areas dominated by Euterpe oleracea has the same ecological meaning than on hilltops? Certainly not in term of number of trees killed, biomass loss and forest turnover . Depending on the variable of interest (e.g. carbon emission), a fit-them-all definition is questionable. This issue, if not formally addressed, should at least be discussed. The following recent publications may provide additional information (Chambers et al., 2013 ; Lobo and Dalling, 2014 ; Schliemann and Bockheim, 2011).

The choice of the values of height and threshold may be adapted to different forest types and topographic characteristics. In our case, the choice was fully data-driven using the DCM and DEM and no ecological knowledge. Within our framework it is likely that in waterlogged areas, areas covered with mature trees that do not exceed the height thresholds may appear in our analysis as forest gaps. In order to clarify this question, an approach using time-series would allow to identify these 'false' gaps that never get filled and thus are not part of the forest endogeneous dynamics. These are not gaps in the ecological meaning.

Finally, the manuscript requires additional efforts in editing (loads of typo citations errors, unclear headers and acronyms) and reviewing recent literature (lots of relevant publications is missing, comparing lambda with other studies). A proof-reading by a native English would also greatly help. This has been done. The manuscript has been edited by a professional science editing service. We do believe that directly comparing lambda values between studies is difficult because it may depend on the assumed (or inferred) height and size thresholds.

Specific comments

— l. 99 : For clarity, please define explicitly all the acronyms used, i.e. topographic exposure (TOPEX).
See line 99

— l.119 : Sub-header should be : "Height above the nearest drainage" to be consistent with previous sub-headers
See line 120 . Section Methods 2.1.2.

— l.193 : What is the resolution of the TOPEX variable ? Do you have several indices by 5m2 ? Please clarify how you can get 2 values (min max), or did you standardize TOPEX as : abs((TOPEX − min(TOPEX))/(max(TOPEX)-min(TOPEX))).
The native pixel resolution is 5 m× 5m. Original values for TOPEX are hardly interpretable because they are counter-intuitive. In order to simplify the interpretation, we thus modified the TOPEX values to get the highest for the highest exposure. See the variable trasnformation section.

— l. 212 : I suggest to change the header here, as Kuo-Mallik refers to a method, but you used it to select the variables. "Variables selection" looks more appropriate. There is also an issue in the way the reference is quoted.
Line 212 : Agreed

— l. 216 : there is an missing (or extra) parenthesis in your expression
Line 217 : done

— l. 226 : "Given this height, we retained the surface xmin = 104 m2". What is the link between the height threshold and the minimal gap area, here ? I thought both minimal height and gap size were defined separately.
Yes they are. Firstly, we define a height threshold from which we observe a gap. Secondly we used this height threshold in order to determine the minimum gap size area using the Kolmogorov-Smirnov (KS) distance criterion

— l.254 : "Environmental covariates with posterior KM values close to 1 , NAMELY Slope, TO-PEX, and HAND ...
Line 254 : Done. "Environmental covariates with posterior KM values close to 1, namely Slope, Topex, and HAND"

— l.260 : "Defining the height threshold at which forest gaps may be delineated is a major difficulty faced by foresters. Many times, canopy gaps have been defined in the field, adopting Brokaw's definition" Is it only the minimal height that is at stake here, or also the minimal area ? Many studies define gaps regarding to their size (e.g. Denslow et al., 1998 ; Hérault et al., 2010 ; Lima, 2004). This sentence sounds odd, I suggest rephrasing as follow : "Delineating forest gaps is a persistent challenge for foresters and ecologists, among whom Brokaw's gap definition (1982) has remained very popular/extensively used."
We thank you for your suggestion. The changes were made at line 263

— l. 265-269 : There are several studies that do not use this 2m-threshold definition of gaps, but 10m (e.g. Hubbell et al., 1999 ; Meer and Bongers, 1996 ; Welden et al., 1991). While the authors are extensively referring to the seminal paper of Brokaw, there are way more references defining gaps in complex tropical forests that are lacking here.
We have updated the bibliography

— l.300 (onwards) : This paragraph is very confusing. Where does the 75% comes from ? What is the remain 25% then ? In sloppy areas, does it make a big difference if a tree falls due to breakage, or being uprooted ? I don't think so, and tree size seems to be a more important factor in the cascading effect than mode of death. Yet, the turnover may be more rapid on slopes than bottomland, resulting in fewer large trees (and tree fall gaps).
Rewritten

— l.338 : but WE found
Done, line 345