Reviewers comments are in normal type; our responses are in *italics*, and new or revised text that will be included is given in red.

Major comments

Both reviewer 1 and reviewer 3 point out that the inclusion of AVHRR as a benchmark is unnecessary.

Reviewer 1. Page 951, Line 23-28: AVHRR is not originally designed for fire detection. The burned area simply scaled from AVHRR active fires does not provide much additional informa- tion than the GFED burned area. Unless the authors explicitly state the advantages of using AVHRR burned area, I don't think it is necessary to include a comparison to this dataset in this paper

Reviewer 3: Why do you use AVHRR in the benchmark? Convertion factor to derive burnt area from AVHRR number of fires varies among plant functional types [see work by Wooster et al].
We originally included the AVHRR comparison because this data source has been used to document fire regimes in the Australian context. However, we agree with the reviewers that this is unnecessary and adds little to the comparison with GFED. We therefore propose to remove the AVHRR-derived map from the figure and reference to the data set from the text. We have removed Lines 23-29, page 21 in section 4, the reference to this data set in Table 4, and the map from Figure 9. We have changed and renumbered the caption for Figure 9. We have also removed the Craig et al and Maier & Russell-Smith references. We have also changed the score comparisons for fire to reflect the fact that these comparisons no longer include the AVHRR data set, both in the text (Section 5, Section 6) and in Tables 6 and 7.

Both reviewer 1 and 2 point out that our evaluation of the model is based only on Australia, and that we have not established that the improvements would apply to globally. Reviewer 2 suggests that we modify the title of the paper to reflect the fact that the evaluation was only performed for Australia
Reviewer 1: The authors only evaluated the model performance in Australia and claimed the modified version improved the simulation. It would be interesting to see whether this improvement is at a cost of deteriorating model performances in other regions.
Reviewer 2: As reviewer #1 mentioned it would be nice to see how the modified LPX model performs globally and not only for Australia as presented here. As long as this is not shown in the manuscript statements such as "The new model incorporates a more realistic description of fire processes and produces a better simulation of veg etation properties and fire regimes across Australia, and is expected to produce a considerable improvement in the simulation of fire-prone vegetation worldwide." are not justified.
Title: That the evaluation of the model is restricted to Australia should be reflected in the title
Our motivation for focusing on Australia for the evaluation was because this is the region of the world which shows the poorest performance in the original LPX and where fire is important. The new parameterisations were not specifically developed or tuned for Australian conditions, and therefore we believe that the changes will produce an overall improvement in the model performance globally. Given that the parameterisations are generic, we prefer not to change the title of the paper – this is not a model for Australia, tuned for Australia; it is a global model that has been tested for Australia. A comprehensive evaluation would represent a large amount of work and is beyond the scope of the current paper. However, we agree that we have not specifically evaluated the model for other regions, and that until we do so we should be more cautious in our expectations for the model. However, we would like to point out that the claim that the model is expected to perform better globally is only made in the final conclusion – it does not figure in the abstract or anywhere else in the text.
We propose to add text to make it clear that the new parameterisations are generic an not specific for Australian conditions at the beginning of Section 3, as follows:
The improvements are based on analyses of large-scale regional and/or global data sets, and are therefore generic. Although we focus on Australia for model evaluation, we have made no attempt to tune the new parameterisations using Australian observations.
We also propose modifying the final conclusion to read:

The new model incorporates a more realistic generic description of fire processes, and has been shown to produce a better simulation of vegetation properties and fire regimes across Australia. Further tests are required to establish that it produces the expected improvement in the simulation of fire-prone vegetation worldwide.

Reviewer 2 also points out that "The introduction of adaptive bark thickness and resprouting produces more realistic fire regimes in savannas, including simulating biomass recovery rates consistent with observations" – This must not be the case for African savannahs. You only tested for Australia.
It is true that our comment about fire regimes in savannas in the abstract refers to Australian savannas. Our evaluation of the biomass recovery rates however is not specific to Australia, although we chose a site from southeastern Australia in order to have consistent information about vegetation and climate. The observations are derived from several different biomes on several different continents (as is clear from the species involved, see Table 5), and the evaluation we performed can therefore be seen as a generic test of what would happen in wooded biomes containing resprouting or non-resprouting trees. This was not made clear sufficiently in the original text. We propose to modify the abstract to read:
The introduction of adaptive bark thickness and resprouting produces more realistic fire regimes in Australian savannas. We also show that the model simulates biomass recovery rates consistent with observations from several different regions of the world characterized by resprouting vegetation.

We have modified the text in Section 4 describing this experiment, to make our intention clearer as follows:
To assess the response of vegetation to the presence/absence of resprouting, we ran both LPX-Mv1-rs and LPX as described above for a single gridcell with woodland vegetation. For this purpose, we selected a cell in southeastern Australia with <20% wood cover as determined by the International Satellite Land-Surface Climatology Project (ISLSCP) II vegetation continuous field (VCF) remotely-sensed dataset (Hall et al., 2006; DeFries and Hansen, 2009) but the response (Fig. 8) is not dependent on the actual location.

The simulated total foliage projected cover (FPC) in the years post-fire was compared against site-based remotely-sensed observations of inter-annual post-fire greening following fire in fire-prone sites with Mediterranean or humid subtropical vegetation from several different regions of the world (Fig. 7; Table 5), split into sites dominated by either RS and other fire adapted vegetation (normally Obligate Seeders – OS) as defined in Sect. 3.6 based on the dominant species listed in each study (Table 5). (The use of observations from other regions of the world reflects the lack of observations of post-fire recovery in Australia.)

We have also modified the text describing the results of this experiment in Section 5.2 as follows:
The simulated regeneration after fire in RS-dominated communities is fast: NDVIsim reaches 90% of pre-fire values within 7 yr, whereas post-fire regrowth takes 30 yr in the simulations that do not include RS (Fig. 7). Observations show that post-fire recovery in RS-dominated vegetation takes between 4–14 yr with a mean recovery time of 7 yr, whereas the recovery takes 8–16 yr (with a mean of 13 yr) in OS-dominated communities; and 7–22 yr (mean of 19) in boreal ecosystems.

We have also added a column to Table 5 to indicate the geographic location of each study.

Reviewer 3 comments on the fact that, although DGVMs simulated the behavior of plant functional types rather than vegetation types, we refer to fire behavior within specific biomes and specifically savannas.
"The authors made several statements about savanna ecosystems that feel disconnected. I see why the new parameterization may target toward a better reprensentation of (wild)fires in savanna (arid region, mixed woody and grassland interfaces, seasonal fire occurrence). However, DGVM sees the world as a mixture of plant functional types, pe se, and so does the analysis in the manuscript. Please elaborate on your statements."

Biomes are a convenient way of describing vegetation that is a mixture of PFTs. For example, *forests are dominated by woody PFTs whereas savannas are a mixture of woody PFTs and grasses. We use the biome terminology in this paper as a shorthand for this. We think that the reviewer is requesting a more explicit statement about the nature of the mixture of PFTs that characterize savannas in the real and in the model world.*

We have addressed this by modifying the text describing savannas in the introduction as follows:
LPX produces sharp boundaries between areas of high burning and no burning in tropical and temperate regions. These sharp fire boundaries produce sharp boundaries between grasslands and closed-canopy forests. The unrealistically high fire-induced tree mortality prevents the development of vegetation characterized by varying mixtures of tree and grass plant functional types (PFTs) that are characteristic of more open forests, savannas and woodlands.
We have also modified the text in section 5.1, to make the relationship between simulated PFTs and derived biomes clearer, specifically:
The boundaries between vegetation characterized by woody PFTs (i.e. closed forests) and mixtures of woody and grass PFTs (i.e. savanna) in this region are still too sharp (Fig. 8).
And similarly in section 5.2:
Including resprouting in LPX-Mv1 (LPX-Mv1-rs) produces a more accurate representation of the gradual increase in the importance of grass PFTs characteristic of the transition from forest through woodland/savanna to grassland (Fig. 8)

Reviewer 1 suggests that the manuscript (e.g., section 4) can be written more concisely. Some related tables and figures can be moved to the supporting material (see the following specific comments for detail).
We have removed Tables 3 and 4, and Table A1, to the Supplement. We have edited the manuscript to be more concise, paying particular attention to Section 4. We have not cut the text describing the benchmarking system because many of the reviewers' comments suggested that this needed to be explained more fully rather than vice versa. Unfortunately, we have also had to add some text in response to the reviewers' comments.

Reviewer 3 suggests that some restructuring of the manuscript would be useful in order to emphasise which model components we changed. "I would recommend to adjust the overall structure of the manuscript. The audience needs to get the idea, up front, of where you did the work. Perhaps try to rearrange Section 2 (i.e. model basics before re-parameteration), so that you keep a concise description of the fire module e.g. Fig. 1.; and combine equations 1-8 and related texts with Section 3 to have a clear comparation of what was changed and what was not. Table 5 may be more suitble for supplementary information except for site information and recovery times."
We end the introduction by a listing of the areas which were re-examined and re-parameterised in LPX-Mv1, and we then go on to describe the basic model (in Section 2) and the new treatments (in Section 3). We think that combining sections 2 and 3 will make it harder for the reader to see what has been changed. However, we think that it would be useful to emphasise that the model description in Section 2 focuses only one things that are going to be changed and therefore described in Section 3. We therefore propose to add a sentence to the end of the introduction explaining the structure of the paper as follows:
In this paper, we begin by describing the basic fire parameterisations in LPX (Section 2) and then go on to explain how these parameterisations were changed in LPX-Mv1 (Section 3) before evaluating whether these new data-derived parameterisations improve the simulation of vegetation patterns and fire regimes (Section 4).
We will also add text in section 2 to make it clearer that we are only describing parts of the fire modules that will be changed as follows:
LPX incorporates a process-based fire scheme (Fig. 1) run on a daily timestep (Prentice et al., 2011). The LPX fire scheme is modified from the Spread and Ignitions FIRE model (SPITFIRE: Thonicke et al., 2010). In this Section, we describe those aspects of the LPX fire model that appear to contribute to poor simulation of fire regimes in Australia (and likely other semi-arid regions) and which we have re-examined and re-parameterised on the basis of data analyses (see Section 3).
We have now moved Table 5 (and also Tables 3 and A1) to Supplementary (see response to reviewer 1)

Reviewer 3 suggests that some aspects of model performance were over-stretched without support.

1. A performance score 5% better than random but 30% worse than mean null model is still a miss. The conclusion of 65-95% improvement in burnt area in SE Australia was made on the fact that it was 129% worse than mean null model.
2. Additionally, I think that the major contribution of re-parameterizations of lightning, fuel drying times and decomposition is at vegetation fields, not burnt area.

We agree that model scores worse than mean null model is not an absolute success. However, we do not claim that the new model is perfect but that it is better than the old model, which in turn is better than other formulations such as the LPJ model (see Kelley et al., 2013). This is why we quote the difference between LPX and LPX-Mv1 for all of the comparisons. We have consistently reported the values for the new model(s) against the standard benchmarks and the mean and randomly-resampled models and all of this information is given in the tables. In fact, as shown by the metric statistics (Table S5), the lightning, fuel-drying times and decomposition improves both the vegetation fields and aspects of the burn area. In fact we believe that new parameterisations are important for improving the simulation of fire, but that at least in part the remaining issues reflect problems in the vegetation component of the DGVM such as overly slow growth rates and overly slow decomposition rates. Improving the vegetation component of the model was not the focus on the current study, but will be addressed in the future.

Resprouting is an interesting feature and the authors did an admirable amount of work for allocation of resprouting species. It did not, however, improve burnt area reported by LPX. I would love to see an analysis in the future when runoff and NPP data sets are available for the benchmark.

The introduction of resprouting does not significantly improve the simulation of burnt area overall but, as is clear from the metric table, it does improve the simulation of burnt area in SE Australia. And as we point out in the text, it also improves the simulation of burnt area in northern Australia. The main impact of incorporating resprouting is on tree cover – which is simulated more correctly in the new model – and hence on carbon stocks. Although not shown in this paper, the LPX model (and LPX-Mv1-nr) produces estimates of above ground carbon stores that are too low compared to observations. Thus, we feel that it is worthwhile to include the treatment of resprouting in the model because it will improve our ability to predict vegetation changes and carbon cycle changes e.g. in response to future climate changes.

We have used an Australia-specific NPP data set for evaluation and the results are included in Table S1. We agree with the reviewer that it would be useful to benchmark the model for runoff, when appropriate data sets are available.

Specific comments

Reviewer 1.

Page 934, Line 24 - Page 935, Line 2: The beta value given here for the original LPX is extremely small. Given this small beta value, I don't understand the meaning of equation (1) in partitioning the CG lightning. Unless P(wet) is 1 (i.e., every day in this month is a wet day), the CG(dry)/CG fraction derived from Equation (1) is always very close to 1 (e.g., with only one dry day in a month, (1-P(wet))ˆbeta = (1-30./31.)ˆ0.00001 = 0.99996566071). This essentially assigns all CG lightning strikes to dry days in a month, not "removes all strikes in months with more than two wet days" as stated by the author. And the blue line in Figure 2b should be a straight line with value of 1. By the way, the beta value for the original LPX given in Prentice et al (2011) was 0.001, which may still result in unrealistic dry-wet-CG partitioning.

There is a mistake in this equation, which was also present in the formulation given in Prentice et al. (2011). The beta value should not be outside the bracket. We thank the reviewer for identifying this and will correct it in the revised manuscript.

Page 935, Line 7-10: "This problem can be corrected by: : :", but it is not clear to me whether it is actually corrected using this way in LPX or LPX-Mv1?

This correction was not applied in LPX but has been applied in LPX-Mv1. We will amend the text to make this clear as follows:
This problem can be corrected by scaling the fractional projective cover (FPC) and Leaf Area Index (LAI) of each grass PFT by the ratio of total simulated grass leaf mass of both PFTs to the leaf mass expected if only one grass PFT was present (B. Stocker, personal communication, 2012). This was done in LPX-Mv1.

Page 936, Line 10: Please change (1-1/exp) 63% to (1-1/exp) = 63%.
We will change the text as suggested.

Page 939, Line 7-11: The unit for L given here (flash/m2/month) is different from that in Fig 2a (flashes/km2/day). If possible, please use consistent unit.
We will change the labeling of Figure 2 and the caption to ensure that the units given in the text and on the figure are consistent.

Page 940, Line 7-8: Equation 10 seems to provide a more realistic partitioning. But will the removing all lightning in wet days result in an underestimation of the fire ignition? Meteorological conditions in wet days may be unfavorable for fire ignition and spread, but due to their large numbers, the wet day CGs may still be able to start significant amount of fires (especially in partially wet days).
Although there are lightning strikes on wet days, very few of these strikes generate large fires because the fuel is too wet to burn. To verify this, we have run an additional sensitivity experiment in which we allow ignitions on wet days. This experiment shows that the inclusion of wet-day ignitions increases the number of fire starts (particularly in northern Australia), but has little impact on the total area burnt. In fact, this experiment produces a slightly worse match to the observations (GFED4), with a slight degradation in burnt fraction and phase of burning. There is a very slight improvement in seasonal concentration, but the simulation with wet-day ignitions is still worse than the null model. There is no change in the interannual variability of fire. We have added the results of this experiment to the supplementary, including both maps of the different fire components and a table showing the metrics for this experiment. We have added a sentence in Section 3.1 to refer to this sensitivity test as follows:
A sensitivity test including lightning on wet days shows that such ignitions have little impact or degrade the simulation of fire: see Supplement.)

Page 940, Line 15-17: Please update the reference of Pfeiffer and Kaplan (2013) with new title and new model name.
The reference was given correctly in the reference list. We will correct the citation in the text to Pfeiffer et al. (2013) and change the model name to LPJ-LMfire (v1.0).

Page 957, Line 19-22: It is hardly to draw the conclusion from Figure 8 that LPX-Mv1-rs performs better than LPX-Mv1-nr in representing the transition from forest to grassland.
This conclusion is drawn from the metric statistics. The difference between the two models is slight and somewhat difficult to see at the small scale of the plots in Figure 8. These plots will be bigger in the new version of this Figure and thus more easy to see, so we will leave the citation to the Figure but will add a reference to the metric table also at this point.

Page 957, Line 16-28: There is not much fire in Southwestern Queensland. Probably a typo here for 'northwestern Queensland' ?
There is in fact an increase in fire in southwestern Queensland, and this is not a typo.

Page 959, Line 8-9: "Adaptive bark thickness has not been included in any vegetation model before" is duplicate of "Adaptive bark thickness and post-fire aerial resprouting behaviour have not been included in DGVMs until now" in Line 5-6 of the same page.

We agree that there is some repetition in the text, and we will amend these sentences to read
Post-fire aerial resprouting behaviour has not been included in DGVMs until now, although resprouting has been included in forest succession models (e.g. Loehle, 2000) and the BORFIRE stand-level fire response model (Groot et al., 2003). Adaptive bark thickness has

not been included in any vegetation model before, despite the considerable variation in this trait both across and within ecosystems and the fact that the distribution of bark thicknesses within an ecosystem is known to shift with changes in fire regime.

Page 976, Table 1: It would be beneficial to readers if the authors include the sources for each PFT-specific parameters in this table.
This is a useful suggestion and we will implement it

Page 977, Table 2: Some PFTs are not consistent with that in Table 1 (e.g., TN and BN in Table 2 vs. BNE in Table 1). Please make the notations consistent throughout the paper.
We thank the reviewer for pointing out the use of non-consistent abbreviations for the PFTs. We will correct the notations in Table 2 and Table A1 to be consistent with the text.

Page 978, 980, Tables 3 and 5: The contents in these two tables are good summaries of previous studies and have been used in this study to derive the recruitment penalty for resprouting PFTs and post-fire recovery. However, I think it is better to move them to the supporting material for reference only in order to shorten the paper. Also, some information from these studies has already been presented in other place (e.g., Fig 7).
We will move Tables A1, 3 and 5 to Supplementary as suggested by the reviewer, and renumber the remaining tables.
We will add text in the Supplementary to describe the new tables, and re-number the tables that are currently in the Supplementary, as follows:
Table S1 provides information on the allocation of species to plant functional types and to resprouting and non-resprouting classes, as used in the bark thickness analyses. Table S2 provides a summary of the studies about post-fire recruitment rates and Table S3 provides information used to calculate recovery rates.

Page 983, Table 6: The 'Bootstrap mean' and 'Bootstrap SD' have not been mentioned in the main text. If the 'bootstrapping experiment' is equivalent to the 'randomly resampled' null model described in Page 952, please make a note of it.
We have altered the caption and replaced bootstrapping experiments with random-resampled null model:
obtained from random-resampled null model (Bootstrap mean, Bootstrap SD).

Page 984, Table 7: Some items are not clearly defined. For example, what is 'mean ratio'? What are 'NME', 'MPD', 'MM' standing for? Readers may have to resort to Kelley et al (2013) for understanding the whole table.
We will provide definitions of the metrics (NME, MPD and MM) in the caption, as follows:
The metrics used are the normalised mean error (NME), the mean phase difference (MPD) and the Manhattan Metric (MM).

Page 991, Fig. 1: If possible, please highlight the modified modules or parameters in this study. It is not worthwhile to include a same figure that was already published in an early paper.
We will modify this figure in order to identify where we have made changes to the model.

Page 996, Fig. 6: When alpha > 0.5, it seems the modeled (LPX or LP-Mv1-rs) sum of tree fraction and grass fraction is close to 1, while the observed sum fraction is much smaller than 1. Could you explain this discrepancy?
We thank the reviewer for drawing this to our attention because there is an error in the diagram. There is a discrepancy between the data and the mask provided for the VCF, such that the mask includes cells that have no information on the proportion of trees, grass and bare ground. In using the VCF data for benchmarking, we only use cells that correspond with the LPX mask, and this does not include any of the VCF cells without data. Unfortunately, when drawing this figure, we omitted to remove the cells with missing data. Most of these cells are around the coast (i.e. in regions with high moisture values) and this lowers the totals for trees considerably. We have now removed these cells and recalculated the averages for Figure 6, and this increases the total sum fraction and produces values for trees that are

comparable with our simulation. We will include the redrawn version of the Figure in the manuscript.

Page 997, Fig. 7: What does the horizontal dashed line stand for (close to 90%)?
The horizontal dashed line shows the 90% recovery level. It is unnecessary and confusing, and we will remove it from the figure.

Page 998, Fig. 8: The authors did not provide an extensive discussion on the differences between modeled tree covers using LPX-Mv1-rs with crop masking and without crop masking. Panel (e) to (g) can be moved to supplementary material.
We agree with the reviewer that the impact of crop masking on the simulations is not discussed and that panels (e) to (g) in Figure 8 are therefore redundant. We will redraw this figure and remove these panels, and amend the caption. We do not think it necessary to move these plots to Supplementary, because they are not discussed.

Page 1000, Fig. 10: The green and blue colors in a) are very difficult for me to differentiate. Please consider using another pair of colors.
We will replot Figure 10a using shades of brown to indicate reduced tree cover and shades of green to indicate increased tree cover.

Reviewer 2.

The new model (LPX-Mv1) improves Australian vegetation composition by 33 % and burnt area by 19 % compared to LPX. – the improvement in % is hard to understand out of context. Here you should be more specific, e.g. referring to your benchmarking score used. Also the values refer probably just to one metric (annual average) and do not summarize all the metrics applied (including seasonal variation and interannual variability).
The handling editor suggested that we eliminate such detail from the abstract prior to posting on GMD, particularly given that it would be difficult to explain the benchmarking metrics in a simple way in an abstract. We do not think it appropriate to give details of the scores here. However, we agree that the sentence itself could be made more explicit and propose to change it to read:
The new model (LPX-Mv1) produces an improved simulation of observed vegetation composition and mean annual burnt area, by 33 % and19 % respectively compared to LPX.

The introduction is very short and misses, for example, an overview of other global fire model activities.
Our intention here was not to review the state of fire modeling (although we agree that the time is ripe for a new review of the state-of-the-art), but rather to provide a motivation for work to improve the LPX model. We therefore do not think it appropriate to provide an extensive review of previous models, especially given the reviewers' strictures about condensing the manuscript. However, we think that it might be appropriate to briefly summarise the processes that are explicitly simulated in LPX, and to provide references to other models that adopt similar approaches. We also think it might be appropriate to draw attention to the fact that LPX does not include anthropogenic ignitions, though many other models do, and the reason why Prentice et al. (2011) explicitly excluded such igntitions. We have therefore added two sentences to the introduction as follows:
In common with several other fire models (e.g. Arora and Boer, 2005; Kloster et al., 2010; Thonicke et al., 2010; Li et al., 2012; Prentice et al., 2011; Pfeiffer et al., 2013) LPX explicity simulates lightning ignitions, fuel load, susceptibility to burning, fire spread and fire-induced mortality. However, it does not consider anthropogenic ignitions because the dependencies of such ignition with population density, used as a basis for such ignitions in other models, have been shown to be unrealistic (Prentice et al., 2011; Bistinas et al., 2014).
We have added the references cited here to the reference list.

The shortcomings of the LPX model must be demonstrated somewhere. If this is part of Keeley et al. (2013) than it should be properly referenced, otherwise it should be shown in more detail here.

The global shortcomings of LPX have indeed been demonstrated in Kelley et al. (2013) and we cite this paper. We explicitly list the three major problems with the LPX simulations that are related to fire. When we evaluate the new parameterisations, we always show the results for Australia obtained with LPX and with the new model LPX-Mv1 (e.g. original Figs 6, 8, 9). We also give the metric scores for LPX and and LPX-Mv1 based on the comparison over Australia (original Table 7, and Table S2). We believe that this is an adequate demonstration of the LPX model both globally and for Australia.

Page 934/Line 16: Here it should be mentioned that the work by Prentice et al. is based on an earlier version of the SPITFIRE model by Thonicke et al. This than also helps the reader to understand the numerous references to Thonicke et al. in the following paragraphs.
We agree that reference to the SPITFIRE model at this point would make it easier to understand the subsequent references to Thonicke et al. and so we will modify the text by the addition of a sentence as follows:
LPX incorporates a process-based fire scheme (Fig. 1) run on a daily timestep (Prentice et al., 2011). The LPX fire scheme is modified from the Spread and Ignitions FIRE model (SPITFIRE: Thonicke et al., 2010).

Page 938: Lightning ignitions
(i) I have problems to follow your description on lightning ignitions. The total flash rates are taken from the LIS sensor and the CG flashes from NLDN, and you combine those two to derive a relationship between total flashes and CG flashes? How does RL than relate to L? I think I'm missing here something.
We realize that the distinction between RL and L was not clear. In response to comments by reviewer 3, we have altered the text explaining RL as follows:
We scaled the flash count from each overpass for detection efficiency and the ratio of observed to total overpass time. These scaled flash counts were summed for each month, to give monthly recorded total lightning (RL), which includes both cloud to cloud and cloud to ground strikes (i.e. IC+CG).
(ii) How does your relationship compare to the findings of Baldocchi et al., 2000?
This paper was incorrectly cited; it should be Boccippio et al., 2001. We have changed this in the text and in the reference list. Boccippio et al. (2001) find that the strongest determinant of cloud-to-ground strikes is total strikes (as do we). They show that there some influence of topography but this is non-unique relationship. They do not quantify the strength of these relationships, or attempt to develop a predictive model. Neither do they examine the relationship between lightning and precipitation explicitly. Our aim here was to develop quantitative relationships, so although the Boccippio et al. (2001) is an interesting starting point, the results are not directly comparable to our analyses.
(iii) The formula gives flashes in (flash m2 month). The figure (1) in flashes / km*/day. This should be unified.
We will correct the Figure to ensure that the same units are used throughout.

In the LIS product the lightning frequency ranges between 0.1 and 70 flashes/km**2/year. According to your equation this leads to CG ratio way below 1%, which is increasing with increasing total flashes and not decreasing as the figure suggests. Please clarify.
*We think this confusion arises because we have used different units in the text and on the figure (which we will correct, to plot the data in flashes/m2/month on the figure). In fact, the lightning frequency in the LIS product ranges from 0.1 and 70 flashes per km**2 per* ***day*** *(not per year) which produces a range from effectively 0 to 0.0021 when converted to flash/m**2/month (as in the text). When we substitute these values into equation 9, we obtain a result where the percentage of CG is decreasing with the number of total strikes as we show on the figure.*

(iv) Where does CG_dry come from and how is Pwet defined?
CGdry is defined by equation 10, and Pwet is the rainfall on days with rain. Although these definitions are implicit in the text, we realize it would be useful to provide explicit definitions of the terms in equation 10 and will do so as follows:
where CG_{dry} is the number of strikes on days with zero precipitation, and P_{wet} is the amount of precipitation on days with rain.
(v) What does “Discontinuous current lightning“ refer to?

Lightning strikes have discrete discharges called return strokes that consist of current surges that last only a few microseconds. Negative strikes can have one to several return strokes of short duration, while positive strikes usually have only one of longer duration. Observations indicate that the short and discontinuous return strokes associated with negative strikes do not raise fuel temperature sufficiently to start a fire, while the long continuous stroke generally associated with positive strikes is more likely to start a fire. Thus although both negative and positive charges reach the ground and can start fires, positive strokes are very much more likely to do so. The polarity of lightning strikes is not given in global lightning data sets, nor is the duration of individual strikes, but for modeling purposes the cloud-to-ground strikes are generally down-weighted to take account of the propensity for positive strikes to be more persistent. Here we use the term discontinuous current lightning to cover the short strikes associated with negative polarity. We realize that this terminology may be unfamiliar to some readers, so we have expanded the text to provide a definition of what this means in this context as follows:

Polarity affects the duration of lightning pulses, with negative polarity more likely to produce discontinuous pulses that are insufficient to raise the temperature to ignition point. This discontinuous current lightning was removed at the same constant rate as in LPX because there are no data sets that would allow analyses on which to base a re-parameterisation.

Page 940/Line 17: In the revisions Pfeiffer and Kaplan named their model differently, please refer to the new model name.

We will modify both the reference and the model name (see response to reviewer 1)

Page 941: Fuel drying. This new treatment than fully replaces the Nesterov index as described in the LPX model description?

No, the new treatment replaces the fuel-specific drying rate parameter described in Section 2. The Nesterov Index is still used to describe the likelihood of an ignition starting a fire. Since this component of the model has not been changed relative to LPX, we did not mention it. However, we realize that the current description of fuel drying could imply that we have replaced all the functions of the Nesterov Index, and will therefore add a sentence at the end of section 3.2 to clarify the situation as follows:

Although we have replaced the formulation of fuel-drying rate, including the formulation of Tdew, we continue to use the NI to describe the likelihood of an ignition starting a fire in LPX-Mv1.

Page941/Line23: Please define Hr

H_R is relative humidity. We apologise for not defining this term, and will do so in the text.

Page 941/Equation 18: Could you indicate how much this formulations differs from the simple assumption T_dew=Tmin– 4 .

The point of this formulation is to take account both of atmospheric moisture content and temperature effects on potential evapotranspiration on fuel drying. The difference between this formulation and the conventional assumption that Tdew = Tmin-4 has been discussed by Kimball et al. (1997), the source of this formulation. They show that using the conventional Tdew = Tmin-4 results in Tdew temperatures being too high, which would result in slower drying rates. We have added a sentence to the manuscript to describe the implications of using this formulation compared to the more conventional formulation as follows:

The more conventional assumption that Tdew = Tmin-4 would thus result in higher dew-point temperatures and slower fuel-drying rates.

Page 944/Line7: Is it too high or too low?

There is too high a production of fine litter because grass production is too high as a result of having too deep roots. We will amend the text to read:

roughly 250% greater than observations.

Page 950/ Model configuration: Did the spin-up not use lightning ignitions?

We realize that our description of the spin-up procedure was somewhat abbreviated. We have modified this section of text to make it clear that the spin-up was run using a detrended climate data set and a standard lightning climatology (following the protocol given in Prentice et al., 2013) as follows:

Each change in parameterisation was implemented and evaluated separately. For each change, the model was spun-up using detrended climate data from the period 1950-2000 and the standard lightning climatology (following the protocol outlined in Prentice et al., 2011) until the carbon pools were in equilibrium. The length of the spin-up varies but is always more than 5000 yr.

age 951/ Line 2: "in the original model (see Table 7)" in Keeley et al. 2013?
Yes, Kelley et al. (2013) benchmarks the LPX model and the maps in that paper show that Australia is the worst simulated region. We will add a reference to this paper in the text.

Page 951: "We re-gridded the data for the period 1997–2006" Why did you restrict your analysis to this time period? GFEDv3/4 is now available for longer time periods.
Although the GFED data are now available up to 2012, the climate data we use to drive the model are only available until the end of 2006. Thus, this interval represents the common interval between the two. We will add a phrase to the text to make this clearer, as follows:
We re-gridded the data for the period 1997–2006 (i.e. the period for which we have climate data to drive the LPX-Mv1 simulations).

Page 952/Line 12: Table 6 refers to a bootstraping mean model, this should be named similar in the main text and in the table.
We have corrected the caption of the Table in response to the comments of Reviewer 1, so that the text and the table now use consistent terminology.

Page 952/Line 22: Manhattan Metric is MM in the following?
We apologise for not having defined the abbreviation of the Manhattan Metric here in the text, and will add the abbreviation here at the first reference.

Page 953: The description of test of the resprouting treatment on plant recovery should have an own section. It does not really fit unter Model configuration and test.
We agree that the description of the test of the resprouting treatment should be separated from the description of the general approach to benchmarking, so we have created a sub-section of Section 4 to cover this material.
4.1. Testing the formulation of resprouting

Page 956/Line 20: "Despite an improvement of 65–95 %" this refers to what?
This is the improvement shown by LPX-Mv1-nr compared to the original model. We will clarify this in the text as follows:
Despite an improvement of 65–95% relative to the original LPX model,

Page 956/Line 25: "around ca. 1–5 %" this refers to what?
This is the fractional burnt area per year in grid cells from this region. We agree that the text is not clear at this point, and will change it to read:
fractional burnt areas of between 1–5% per year

Page 956: Figure 8/9/10 should be introduced in the main text. Throughout the analysis you refer to these figures. It would be easier for the reader to get an idea in the beginning of these sections what is shown in the single graphs.
We will add a paragraph at the beginning of the section on model performance (Section 5) which explains the evaluation approach and introduces the Figures and metric tables, as follows:
Evaluation of the model simulations focuses on changes in vegetation distribution (expressed through changes in the relative abundance of PFTs) and changes in fire regimes burnt area (both total area burnt each year in each grid cell, i.e. fractional burnt area, and the seasonal distribution and timing of burning). We show the simulated change in tree cover (Fig 8) and in mean annual burnt area (Fig 9) for the original model compared to the simulations with LPX-M-v1 in both the resprouting (LPX-M-v1-rs) and non-resprouting (LPX-Mv1-nr) variants, as well as the differences between the two LPX-M-v1 simulations. We use benchmarking metrics to quantify the differences between the simulations (Table 5, Table S5).
As the NME and MM metric are the sum of the absolute spatial variation between the model and observations, the comparison of scores obtained by two different models shows the

relative magnitude of their biases with respect to the observations, and the improvement can be expressed in percentage terms. Although we focus on vegetation distribution and fire, we have also evaluated model performance in terms of other vegetation characteristics, including fAPAR, net primary production, and height (Table S5), to ensure that changes in the model do not degrade the simulation of these characteristics.

Page 957/Line1: "The simulated distribution of trees in climate space is improved in LPX-Mv1-rs compared to LPX." How to you evaluate this?
This is based on the metrics for treecover, as shown in the metric table (formerly Table 7). We have added a reference to the Table at this point. Given that some of the tables will be moved to the Supplementary material section, at the suggestion of the reviewers, this will now be Table 5.

Page 958/Line14: The discussion of Figure 7 appears here a bit out of context maybe this should be done earlier, when discussing the overall resprouting treatment. Also Figure 7 needs a more descriptive caption.
We have removed the reference to Fig 7 from the methods section as it is inappropriate to cite it how. In response to the reviewer's suggestion that we create a separate category of the discussion of Figure 7, we have done so (see response above). We have also expanded the Figure caption to make it more self-explanatory as follows:
Comparison of the time taken for leaf area (as indexed by total foliage projective cover, FPC), to recover after fire in different ecosystems, as shown in the LPX-Mv1-rs simulations and from observations. For comparison with the observations, which were all made after a significant loss of above-ground biomass through fire, the LPX simulations show recovery after a loss of 60% of the leaf area. Red denotes ecosystems dominated by above-ground resprouting (RS) species; blue denotes ecosystems dominated by other fire-adapted species, mostly obligate seeders (OS); black denotes vegetation which does not display specific fire adaptations (NR). The solid lines show LPX simulations; dotted lines show the mean of the relevant observations; the shaded areas show interquartile ranges of the relevant observations. The plots show that LPX-M-v1 reproduces the observed recovery rate in ecosystems dominated by resprouting species; recovery in ecosystems lacking respouting trees is slower than observed, which could either reflect issues with simulated growth rates or the absence of other forms of fire adaptation.

Page 958/line 23: "LPX-Mv1-rs compared to LPX is 18–19 % for burnt area" Here and throughout the manuscript I do have problems to understand the measure "burnt area". Does this refer to the performance of the model to reproduce the spatial distribution of the annual burnt area.
We have added a paragraph to explain the evaluation procedures at the beginning of Section 5, including defining what is meant by burnt area and explaining the use of the percentage estimates of improvement relative to the observations. In this text we explain that burnt area refers to the percentage burnt in every grid cell over the year. We think that this makes clear our terms, and thus have made no change to the text here.

Page 957/Line9: "fAPAR was already on average 59 % higher in LPX compared to observations (Table 7)" Does this refer to the NME or the absolute value?
We realize that this sentence is somewhat clumsy. We have now explained the use of percentages in the introduction to Section 5. We will rephrase this as:
LPX overestimates fAPAR by 59% compared to observations (Table 5), mostly due to simulating too much tree cover in southeastern Australia (Fig. 8b).

Page 961/Line2:" We have used the conceptual framework of Clarke et al. (2013), which is based on extensive field observations, to evaluate our simulations of RS- dominance in a qualitative way" it is not clear to me to what kind of framework this refers to.
Clarke et al. (2013) present a conceptual model for the distribution of resprouters, which is expressed in terms of abundance along a gradient of moisture. However, they do not provide any quantification of the relationships but merely suggest that the abundance of resprouting will peak at intermediate moisture levels. We realize that the term conceptual framework is not clear and will replace this by conceptual model.

Page 961/Line 26: “and is expected to produce a considerable improvement in the simulation of fire-prone vegetation worldwide“ – without showing any global results this is still a very speculative.
We have modified the text (see response to major comments above).

Table6/7: the step1,2,3 notation is confusing. The table should be extended to include the information directly.
The notation here follows that defined in Kelley et al. (2013) In the caption of Tables 6 and 7 (now renamed Table 4 and 5, because of the removal of some material to Supplementary) we define what is mean by these steps. Step 1 is a straight comparison; 2 is a comparison with the influence of the mean removed; 3 is with mean and variance removed. It would be difficult to use these more extended descriptions in the table itself. However, we have added a sentence in the text to clarify this terminology.

Figure 8: applies a crop mask. The differences are actually hard to distinguish. I can not find a reference to the crop mask in the main text. What fields are used in the benchmarking, the one with or without the crop mask?
The benchmarking is done on the straight simulations, i.e. before the application of the crop mask. In fact we do not make any use of the results post removal of this mask and the inclusion of the results with the mask applied in Fig 8 obviously causes some confusion and is unnecessary. We have removed the plots showing the results with the crop mask rom Fig. 8, as requested by Reviewer 1.

Reviewer 3

[Page 934, Line 18]: “the number of lighting strikes that reach the ground. . .” change to: . . .the number of lightning strikes that reach the ground. . .
It is unclear what the reviewer is requesting changed here (since the two phrases are identical) so we propose to leave this sentence as it is.

Page 938, Line 25]: “RL” What is this?
We do define RL at the bottom of the page, but the difference between RL and L is obviously not clear (see comments of reviewer 2 also). RL is the total number of lightning flashes that should be recorded if the satellite observations were continuous, and is derived by scaling the number of flashes in the observation window on the basis of detection efficiency and the ratio of the observation window to the total overpass time. We will modify the text to make this clearer as follows:
We scaled the flash count from each overpass for detection efficiency and the ratio of observed to total overpass time. These scaled flash counts were summed for each month, to give monthly recorded total lightning (RL), which includes both cloud to cloud and cloud to ground strikes (i.e. IC+CG).

[Page 939, Line 16]: “CPC” [Page 939, Line 21]: “CRU” [Page 943, Line 7]: “TRY” Please spell them out.
We apologise for not spelling out the meaning of CPC and CRU in the text. We will do so at first occurrence. TRY is not an acronym, but as indicated on their webpage “a statement of sentiment. However, in order to make it a little clearer what this refers to we have expanded the description as follows:
TRY plant trait database (Kattge et al., 2011; http://www.try-db.org/TryWeb/About.php)

[Page 941, Line 20]: “. . .fuel size x hours” change to: . . .fuel size at x hours. . . or. . .fuel size in x hours category
We will replace this phrase by:
fuel size in each drying-time class (x)

[Page 941, Line 23]: “HR” Spell out “relative humidity” when it first appears as you did with ET. I suppose HR is not a universal understood term.

We will define this in the text (see response to reviewer 2)

Page 946, Line 17]: “parmid changes each day...” Consider making the sentence shorter, or altering punctuations: “parmid changes daily when there is a fire event, based on bark thickness of surving plants, and also annually from establishment, based on...”
We realize that this sentence is somewhat cumbersome, and so will change it as follows:
parmid changes after a fire event, based on the bark thickness of surviving plants. It will also change with establishment, when the post-establishment value represents the wighted average of the bark thickness of new and existing plants (Fig. 5).

Page 948, Line 14]: ca. Use “approximately” instead. And check through the manuscript and edit the others.
We have used ca. consistently throughout the text, and prefer to keep this as is. We have made no changes

[Page 950, Line 22]: “the effect of this change were small.” Can you put in the numbers
We have not quantified the impact of this change on the original model. This change was made in all the simulations made with LPX-Mv1, and therefore does not affect the differences in performance between the simulations. Given this, it is not necessary to comment on the impact of this change, and so we will remove this phrase from the text and make it clear that this has no impact on the results shown in this paper as follows:
but given that this correction was made for all of LPX-Mv1 runs this change as no impact on the differences caused by the new parameterisations.