Interactive comment on “Modelling fires in the terrestrial carbon balance by incorporating SPITFIRE into the global vegetation model ORCHIDEE – Part 1: Simulating historical global burned area and fire regime” by C. Yue et al.

Anonymous Referee #1

Received and published: 13 May 2014

The manuscript by Yue and coauthors describes the integration of the SPITFIRE fire model into the ORCHIDEE land surface model, and includes detailed comparisons with various aspects of the fire regime. Fire is increasingly recognized as an essential process for simulating ecosystem dynamics, distributions, and interactions with the atmosphere and climate. The authors do an especially nice job of recognizing the various dimensions of fire regimes, which include not only burn area but also the number of fires, their size, and their intensity, and carefully comparing the model to different observational datasets. I also appreciated the ‘quality flags’ presented for the ORCHIDEE comparisons that utilize the variety of observational datasets.

I was somewhat disappointed that the authors did not attempt to address some of the major flaws in the model that were highlighted by their analyses. While I appreciate that some of the comparisons themselves are novel, a few relatively large model biases were discovered, including fire season, fire sizes, fire duration, spread rate, and regional burn area. I realize some of these are beyond the scope of this paper. The major flaw in my view, however, is that the authors did not take advantage of or even seem to be aware of recent work that addressed some of these issues with SPITFIRE in LPJ by Pfeiffer et al. 2013 (Geosci. Model Dev., www.geosci-model-dev.net/6/643/2013/doi:10.5194/gmd-6-643-2013). A revised version of the model was presented here which included improvements for fire duration (including multi-day burning), coalescence of fires, and interannual lightning variability.

[General response] We thank the reviewer for the general positive comments on our study, and the very thoughtful comments on the shortcomings in terms of failure to cite a recent study by Pfeiffer et al. (2013). The manuscript is revised to include an approach to reconstruct the "multi-day fire patches" which functions similarly as the multi-day burning scheme in Pfeiffer et al. (2013). As the approaches to handle lighting ignitions in our model and in Pfeiffer et al. (2013) are very different, we do not include the coalescence of fires in the current version of our model. We also made extensive tests to include the interannual lighting variability following the approach proposed by Pfeiffer et al. (2013), and found that the model-observation agreement degraded for most of the GFED regions for 1997-2009, and thus decided not to include it in our manuscript at the current stage, but leave it for future investigation and improvement. All modifications and tests are either included in our revised manuscript; or described in the responses to the review comments as in the following sections. To make it easy to follow the revised contents in the manuscript, the heavily modified texts are shown in blue in the revised manuscript.

Specific comments

-The authors did not mention a relatively recent paper by Pfeiffer et al. 2013. This describes

a revision of SPITFIRE in LPJ. Although the improvements were focused on better representing burning in preindustrial time, this version contains many model developments advantageous for the simulation of present-day fires. It's unfortunate that the authors of this study did not take advantage of this development, as there are a number of areas for improvement with SPITFIRE that the authors acknowledge which have already been developed.

In my view the major improvement from the Pfeiffer study that is applicable for this paper/analysis is the coalescence of fires within grid cells and the allowance of multiday burning. The authors discuss in numerous places the bias in fire size introduced by the 4 hour restriction on fire duration, as well as the discrepancy between fire patch sizes in reality and in the model. Both of these would seem to be improved with Pfeiffer et al.'s modifications.

The authors use climatological lightning frequency to derive natural lightning ignitions. However, there can be considerable variability in this during the fire season. Pfeiffer et al. 2013 accounted for this by deriving a relationship between lightning frequency and ancillary meteorological data.

[Response] We thank the reviewer for referring to the study by Pfeiffer et al. (2013). Our development was more or less parallel with theirs (though our submission was delayed). We acknowledge that in our original manuscript, the simulated fire size is limited by a fire patch length of only one day, with daily fire size being limited by a maximum fire active burning time of 4 hours (241 minutes). Thus the simulated fire size is not strictly comparable with that derived from the observation data (either government agency statistics or derived from satellite imagery), as also pointed out by the #2 reviewer.

Three major new features were included in the development by Pfeiffer et al. (2013), the multi-day burning, coalescence of fires, and the interannual variability in the input lightning flashes used as potential ignitions of fires. Given the time limit in revising the manuscript, and also considering the limited model calibration in Pfeiffer et al. (2013) (which will be discussed in detail in the following), we decided not to simply incorporate all these features in Pfeiffer et al. (2013) into ORCHIDEE, but tried to redo the comparison between simulated and observed fire size in a more sensible way (by taking into account the multi-day burning), and to acknowledge and discuss what we can do and cannot do.

The multiday-burning

First, the multi-day burning scheme in Pfeiffer et al. (2013) in fact did not drop the 4-hour limit of daily fire active burning time when calculating the daily fire size (Page 653 section 3.2.1, first paragraph of Pfeiffer et al., 2013), and this is a shared feature in our model. However, the "multi-day burning" scheme extended the "fire patch length" from the original single day in the SPITFIRE to allow multiple days of fire span, as long as the climate situation allows the fire to persist (in Pfeiffer et al. 2013 this is done by setting the precipitation threshold).

In order to take the multi-day burning into account when comparing simulated and observed fire patch size, we developed an approach to group fires that are simulated to occur within consecutive days into "multi-day fire patches", with the size of each "multi-day fire patch" being the cumulative daily fire size over its corresponding period of duration. We found this approach improved the comparison of simulated vs. observed fire size. We argue in terms of comparing simulated fire patch size with observation, this approach has a similar function as the "mutli-day" burning scheme in Pfeiffer et al. (2013). This approach is described in section 2.5.2 in the revised manuscript, with the relevant results being updated in section 3.6, followed by discussions in section 4.2.2 and 4.2.3.

The coalescence of fires

The "coalescence" of fires in Pfeiffer et al. (2013), according to our understanding, is that fires starting on a given day were considered as "new fires" to be added on the existing fires during the previous day (so that there are more fires on this day than the previous day). While currently in our approach, they were considered to extend from the fires during the previous day (so that fire patch number remains the same), thus the total fire patch number is the maximum daily fire number during the given consecutive days of burning.

However, there is a significant difference in handling the lightning ignited fires between the two models (Figure C1 and Figure C2). The approach to simulate the lightning ignited fires in Pfeiffer et al. (2013), according to our understanding, will finally allocate either 0 fire, or only 1 fire on a 0.5-degree grid cell on a given day. This single fire is derived by comparing the simulated ignition efficiency with a uniformly distributed random number from [0,1]. The lightning flashes finally lose their quantitative meaning and were used only to provide a 0/1 answer to allow a single fire over the given grid cell on the given day. We are cautious for this approach of simulation. Although Fig.7 in Pfeiffer et al. (2013) shows that the simulated burned area for one ecoregion of Alaska agrees relatively well with the observation data, however burned area in many regions are underestimated (Fig. 12 in Pfeiffer et al., 2013) compared with GFEDv3.1 data (considering also the human ignition was not included). And no information on the simulated fire numbers and fire size was provided, nor were they compared with the observation data.

Because of these considerations, the "coalescence" of fires is currently not included in our model. The discussions above were included in our revised manuscript in section 4.2.2.

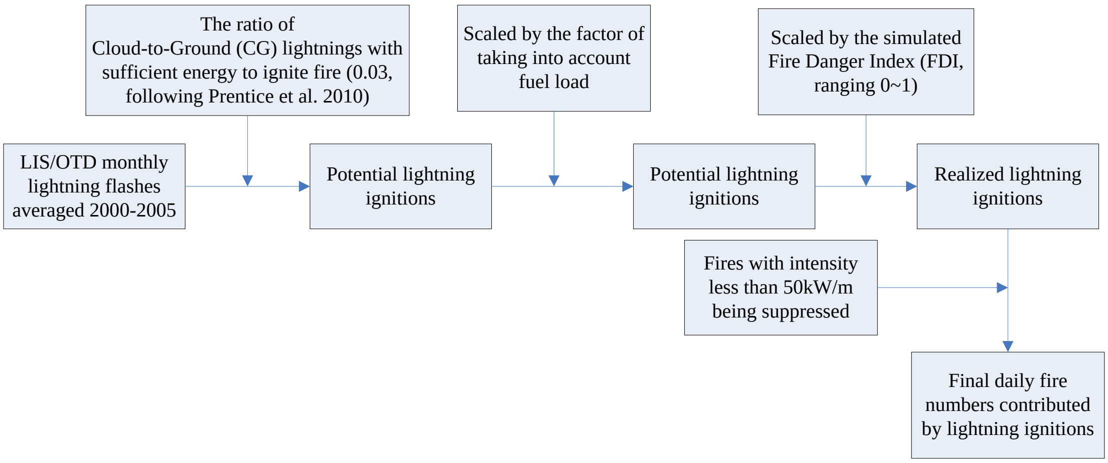

Figure C1 Flow chart of simulation of lightning ignited fire numbers in our study

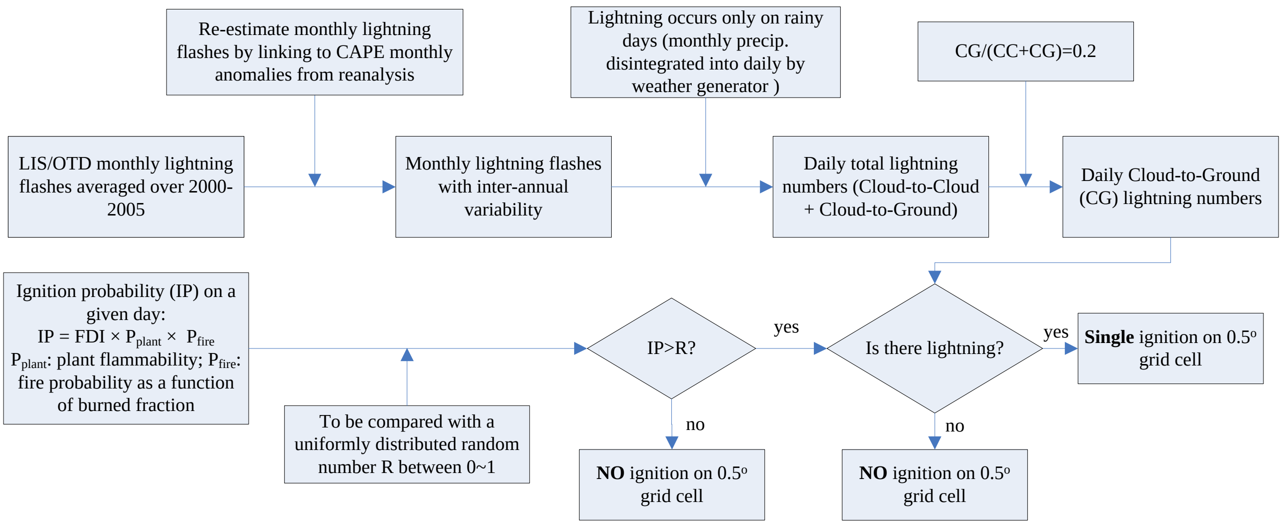

Figure C2 Flow chart of simulation of lightning ignited fire numbers in Pfeiffer et al. (2013)

The interannual variability of lightning flashes

The interannual variability in the lightning flashes is another feature added in the model development by Pfeiffer et al. (2013), in which the variability in lightning activities was linked with the anomaly in the convective potential available energy (CAPE). We noticed that there is a lack of demonstration in Pfeiffer et al. (2013) how the model simulation is improved thanks to the adoption of the new lightning data, especially in terms of the model-observation agreement in the burned area interannual variability. The Fig. 7 in Pfeiffer et al. (Page 663) shows the relatively good agreement in the annual time series of burned area for the "Intermontane Boreal" ecoregion in Alaska between model simulation Alaskan fire agency data. However the model simulation for this Alaskan case study was driven by the lightning data of the Alaskan Lightning Detection System (ALDS) for 1986-2010 rather than the global reconstructed CAPE-derived lighting data (see section 3.4, paragraph 2, Page 658 of Pfeiffer et al. 2013).

We replicated the method as described in Pfeiffer et al. (2013, Equation 1 on Page 649) and produced the CAPE-derived lightning data with interannual variability for 1901-2011, and rerun the whole global simulation by using this new dataset, combined with the spatial a(ND) dataset (Thonicke et al., 2010) which is used in the human ignition equation (Equation 1 in the discussion paper, Page 2382). Besides, we have also done a separate simulation for Alaska by using the local ALDS lighting data, in order to examine the simulation improvement by using this ground-based observation data.

We found the greatest model-observation agreement for 1986-2011 could only be achieved when the model is driven by ALDS lightning data (the Pearson correlation coefficient of annual burned area between the model and Alaskan fire agency data increased from 0.19 to 0.5). And, using the new CAPE-derived lighting data only marginally improved the model-observation agreement for the same period (correlation increased from 0.19 to 0.22). For 1950-2011, the model-observation agreement slightly decreased after shifting to the new CAPE-derived lighting data (correlation coefficient changed from 0.41 to 0.37).

We systematically examined the change in the model-observation agreement for different regions and different time spans when shifting from the mean annual static lighting data to the CAPE-derived data. The agreement of simulated burned area with the observation for 1950-2011 for the boreal North America (i.e., US Alaska + Canada) generally decreased after shifting to the CAPE-derived data, either on annual or decadal basis. Over the 20th century, the shifting of lightning data decreased the agreement of simulated decadal burned area with the Mouillot and Field (2005) reconstruction for half of the 14 regions and increased for the other half. Over 1997-2009 when the observation data by the GFED3.1 is more credible than the 20th century reconstruction, using the new data decreased the agreement of annual simulated and observed burned area for the globe and for most of the regions.

In summary, the CAPE-derived lightning data does not systematically improve the model performance. This could be due to several reasons including the errors in the method to reconstruct the lightning data, the errors in the CAPE data, and model internal uncertainties. We thus finally decide to keep the mean annual lighting data in the present version of the model. However this issue is worth more detailed investigation and will be considered in the future model improvement. For detailed information regarding the comparison of the simulations using the static and CAPE-derived lighting data, please refer to the "***Response supplement material***" (at the end of this document). This is briefly discussed in section 4.2.1 in the revised manuscript.

-In general I would like to see an expanded discussion on what's causing some of the specific model biases. This includes fire season, high burn area and fire intensity in the tundra, big regional biases in North America, the Middle East, southern Africa, Australia, etc.

[Response] The section 4.2.5 is created to accommodate the discussions of regional errors.

-I would advise the authors to be REALLY careful of the long-term burned area observations for Russia. Comparisons are mentioned back to 1920, but the observations are highly uncertain pre-MODIS era, and especially before 1980 (look at the discussion of how Russia data were created in Mouillot and Field (2005), Appendix A, and their uncertainty estimates in Table 2).

[Response] We thank the reviewer for pointing out this. The uncertainty in the historical reconstruction data and the caution needed to interpret this comparison is included in the revised manuscript (section 3.5).

-[Fig. S7] Related to above, I'd like to see these graphs combined for boreal North America, and aggregated to decadal like Fig. 10. The problem is that in the text the authors claim there is good agreement in boreal North America long term. Compared to Mouillet and Field (2005) in Fig. 10, the comparison is decidedly not favorable. It's hard to tell in Fig. S7 what the overall decadal trends are in the national fire databases vs. ORCHIDEE. The authors claim that this reflects the model's ability to capture fire trends driven by climate variation relatively well. I'd also like to see a decadal statistic here, because as it's presented the reader is not convinced, and is hard-pressed to believe that the long-term trends are actually captured.

[Response] The Pearson correlation coefficient between the model and observation are provided for the period after 1950 in the revised manuscript, when the observation is considered to be more reliable. See section 3.5 in the revised manuscript. Both annual and decadal BAs for this region, together with the relevant statistics, are provided in Fig. S5 in Supplement material (also shown as Figure C3 below).

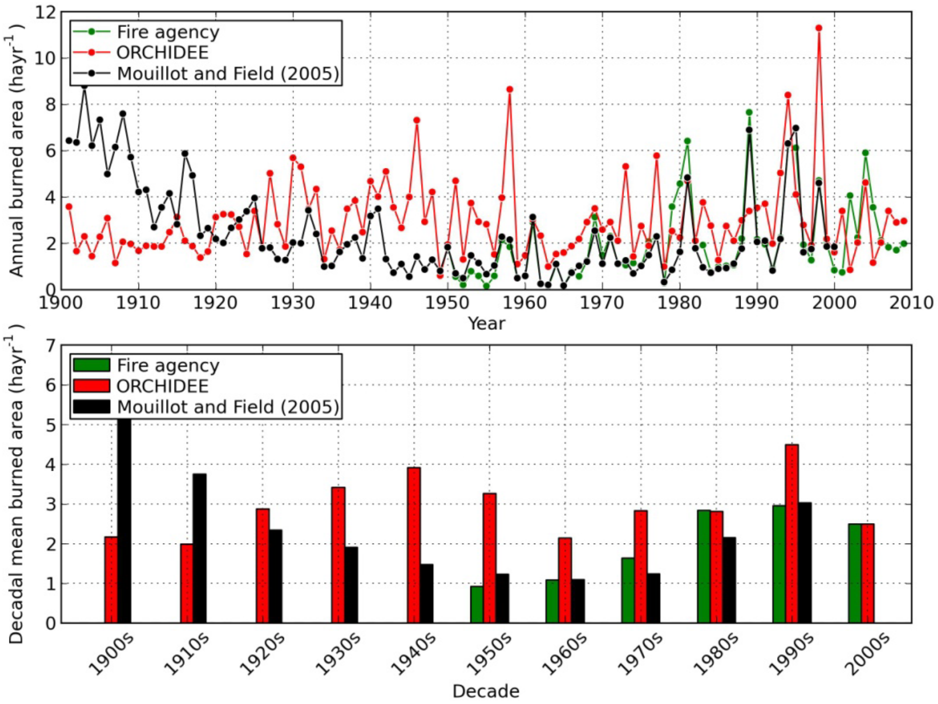

Figure C3 Annual (upper panel) and decadal (lower panel) burned area for the boreal North America, given by the fire agency statistics from Canadian and Alaskan fire agency (green), ORCHIDEE simulation (red), and the historical construction by Mouillot and Field (2005).

-The parameter for ignitions per person per day was spatially explicit in Thonicke et al 2010, and here the authors discarded that and used a global constant. Even though the results are comparable, why move away from something that's arguably more sophisticated? This needs to be justified better I believe. The authors state that the overall average is better when compared with GFED using the spatially-explicit parameter [pg 2392, line 19].

[Response] The revised manuscript used the spatial dataset as in Thonicke et al. (2010).

-Similar to above, why not keep the fire suppression algorithm? It improves the simulation especially in the western US, where fire suppression is known to have decreased burn area by almost an order of magnitude since the mid-20th century.

[Response] The anthropogenic ignitions are implicitly suppressed as contained in the ignition equation (Equation 1 on Page 2382 of the discussion paper). The explicit suppression of lighting ignitions by human alleviated the overestimation of burned in western and central US but also reduced the burned area across the globe. This points to a potential systematic error in the ignition sources (from both lighting and anthropogenic activities) and needs further detailed investigation. For the current version of the model, we decide to not include this, in order to maintain the close agreement of the global burned area with the observation.

-The authors don't discuss how fire intensity is calculated in the model, although it's being compared to data. Is it strictly from the Roethermal equaitons?

[Response] Not exactly. Fire frontline intensity is calculated following Byram (1959), as a product of fuel heat content, fuel consumption, and fire spread rate. And the fire spread rate is further calculated using the Rothermel equation. This is explained in the revised manuscript (section 2.1, within the paragraph of "mean fire size").

-This is somewhat beyond the scope, but if fire intensity and duration are explicitly simulated, why do they not affect combustion completeness? These are crucial drivers, and was an area that was augmented somewhat for the ORCHIDEE integration presented here.

[Response] Physically it could be the reverse. The combustion completeness combined with the fuel load collectively determine the fuel consumption in the fire, which further determine the energy released (fire intensity) and partly determine the fire duration (because the energy available to preheat the adjacent fuel and the fuel load available for burning partly impact where a fire can propagate). The fire duration in the model is currently loosely related with the fire danger index (a general broad indicator for the climate suitability for fire), rather than being mechanistically simulated, and this needs to be improved in the future work.

-Regarding model spinup: a spinup of 200 years seems quite short for aboveground processes. For example, Moorcroft et al. 2001 demonstrated that over 200 years are needed for the accumulation of biomass in tropical rainforests. A second spinup of only 150 years for fire dynamics seems quite short as well. Many fire-prone boreal systems have FRIs of around 150 years, and others are well over 500 years. As such their dynamic equilibrium wouldn't seem to be reached, although I realize the large grid cells burn more frequently than this. Some demonstration of how this spinup was enough, not just for the carbon sink, but also for fire frequency, aboveground biomass, etc., would help.

[Response] A complete spin-up is very computationally expensive in the version of model used here. The intuitive reason to do the initial spinup without fire is to save computation time, as a system without fire would allow faster accumulation in the carbon stocks. Figure C4 shows the evolution of different carbon stocks during the spin-up process. The total live biomass and aboveground litter were found to vary within 0.06% and 0.1% during the last 50 years of the spin-up, respectively (the belowground litter within 0.2%, and the mineral soil carbon stock within 0.08%). They could be considered as in moderate equilibrium for the study purpose here. We agree with the reviewer that the spin-up time is shorter than the fire return interval of some ecosystems (such as some of the boreal forests and the tundra), however the contribution of annual burned area of these ecosystems to the global total is also small. The burned area for the region of 50~70^{o}N takes up ~3% of the global burned area by GFED3.1 data (10.5 Mh yr^{-1} vs. 344 Mha yr^{-1}), and our simulation does not show considerable underestimation for this region (Fig. 5 in the discussion paper).

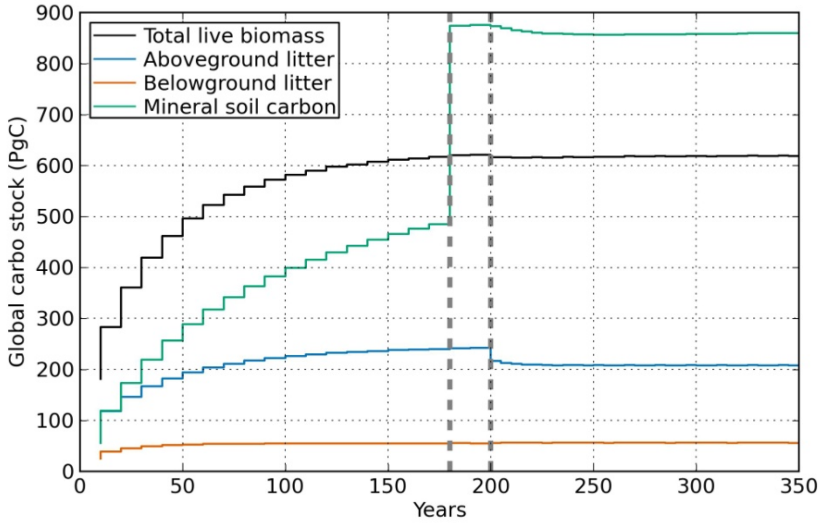

Figure C4 The evolution of global total carbon stock for the live biomass, aboveground litter, belowground litter and mineral soil carbon stock during the spinup simulation. The first vertical dashed gray line indicates that the soil-only processes in ORCHIDEE have been run for 3000 years to speed up the accumulation of mineral soil carbon; and the second vertical dashed gray line indicates the switch-on of the fire module.

-[2386, lines 27 - 29] I wouldn't argue that the model is capable of capturing deforestation fires if the land cover map is static. To support this claim, can the authors somehow generate an estimate of how much 'deforestation; fires there are in their simulations?

[Response] van der Werf et al. (2010) showed that by using the product of forest burned area and the fire persistence time as a proxy for the tropical deforestation rates, 82% of the deforested area by other independent approaches (e.g., Hansen et al., 2008) has been captured for 2000-2005. We replicated this process by using the GFED3.1 forest monthly burned area and the corresponding fire persistence time for the region of 20°S-20°N for 2000-2005. The ORCHIDEE simulated forest burned area for the same region was compared with the GFED3.1 derived deforested area. When making the comparison, only the grid cells with a forest cover >70% by the land cover map used in the simulation were included to make sure that the burned area occurred in relatively closed forest. The mean annual deforestation area for 2000-2005 for the study region by GFED3.1 was 4.0 Mha yr^{-1}, and the forest fire area by ORCHIDEE simulation is 2.7 Mha yr^{-1} (67% of GFED3.1 deforested area), although with rather different spatial distribution due to the fact the land cover map was static in the model (Figure C5). The model could also moderately capture the seasonal variation in the deforestation area as shown by the GFED3.1 data (Figure C6).

This was documented at the end of section 2.3 of the revised manuscript and presented in the Supplement Material.

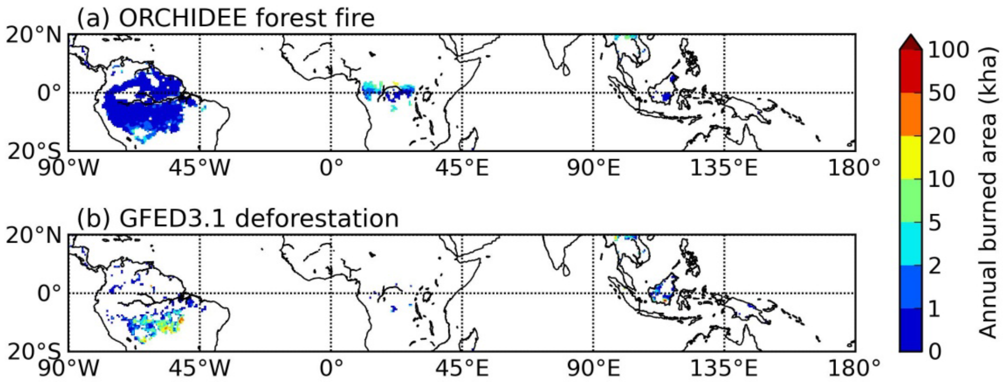

Figure C5 Burned area in the tropical forest (20^{o}S-20^{o}N) given by (a) forest burned area as simulated by ORCHIDEE, and (b) estimated deforestation area by the product of GFED3.1 forest burned area and the fire persistence time as indicated by (van der Werf et al., 2010). Burned area is shown for 2000-2005 for the areas with forest coverage larger than 70% by the land cover map used in the simulation.

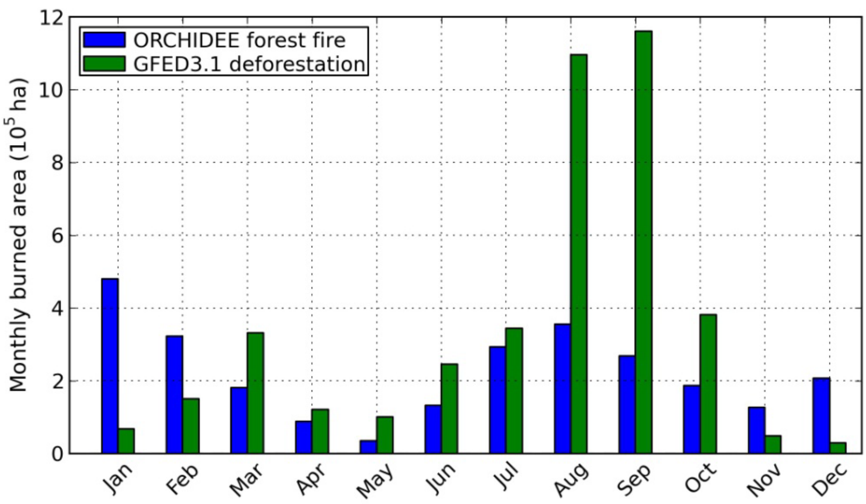

Figure C6 Monthly burned area for the simulated forest fire (blue) and GFED3.1 deforestation area (green) for the spatial extent as in Figure C5 averaged over 2000-2005.

-I believe Archibald et al. 2013 generated 'fire patch' data for the entire globe. Why restrict the patch analysis to boreal North America and southern Africa using the Archibald et al. 2010 data?

[Response] We have compared the simulated 95th quantile fire size with that given by the Archibald et al. (2013). The two case studies have been done to reveal more details in the fire size distribution by model simulation and observation. Note that the observation data for boreal North America are by local fire agencies and the data for the southern Africa are via reconstruction of satellite derived burned area data; and thus they represent different data sources. We think these two case studies are sufficient to demonstrate the model behavior. In addition, the comparison of fire size distribution between model and observation on the global scale should ideally be stratified by different ecosystems or fire types (i.e., derive a *beta* value for each type, or each grid cell and then compare the different *beta* values for their spatial distribution), and is somewhat beyond our scope.

-[2389, lines 22 - 25] This 'pooling of fire patches' requires more explanation here.

[Response] The new approach of grouping fires within consecutive days into "multi-day fire patches" is introduced in section 2.5.2; and the fire size of these "multi-day fire patches" is used in the comparison (section 3.5 in revised manuscript).

-For the comparison with GFED3.1, it might be good to also look at the more recent version that includes small fires (Randerson et al. 2012, JGR-biogeosciences) since burn area and emissions increase by approximately one-third globally.

[Response] We agree with the reviewer that comparing the model simulation with the updated burned area data (including small fires) could bring extra benefits. To our knowledge, the burned area data with "small fires" being included is still not publically available, neither included in the publically available GFED4 dataset (Giglio et al., 2013, http://www.globalfiredata.org/data.html). The recently published modelling studies seem not to include these "small fires" (Li et al., 2013; Yang et al., 2014) and thus the models are calibrated against either GFED3.1 or GFEF4. In the future, if the update in the burned area by the "small fires" is confirmed by the mainstream of observation community, then the model processes need to be further examined and adjusted. This is also part of the reason the current study tries to go beyond the burned area and look into more details on the modelling errors on the global scale especially the potential errors in the ignitions.

-[2390, Seasonal similarity] This statistic makes sense, but it's new to me. Has it been used before? When looking at Table 2 its value means little without context. Is there a way to provide statistical significance or at least more context with good/bad correlations?

[Response] We are not aware of other studies using the same method. We argue this indictor (i.e., the overlapping area of the two normalized monthly burned area time series) might be better than the regression slope or correlation coefficient as it retains the physical meaning (i.e., the fraction of burned area in overlapping months against the total annual burned area). In response to the reviewer's comment, the bootstrapping method is used to associate the derived seasonal similarity (S_{season}, see Equation 3 on page 2390 in the discussion paper) with some statistical significance (i.e., the probability that S_{season} is from a random distribution of seasonality). This is described in section 2.5.1 in revised manuscript.

-[2392, line17] The statement that the model result agrees best with GFED3 is not supported by any number/statistic. Please provide.

[Response] The correlation coefficient was calculated for the annual burned area between different datasets and GFED3.1, see section 3.1, last paragraph in the revised manuscript.

-[Figure 8] SPITFIRE has some peak fire months in February in Alaska, October in Canada, and April/May in the Far East of Siberia. This is very surprising and I think quite unrealistic. What causes this? There are also some December/January grid cells in boreal forests for GFED, which I find somewhat hard to believe. Is this correct?

[Response] After careful examination, we found some error in the mathematical scripts used to treat the monthly burned area data in order to derive the fire peak month, and this error is now corrected. The new fire peak month distribution looks rather reasonable (Fig. 8 in the revised manuscript), with fire peak months in Canada and Alaska mainly in June to September. The fire peak months in Fast East of Siberia are April to June, this is rather reasonable mainly due to the low fuel moisture in spring in the Russian forest, see also discussions in Forkel et al. (2012).

-Regarding fire size distribution in boreal North America, Lehsten et al 2014 (JGR Biogeosciences) show a different function for Canada than the strict power law.

[Response] Lehsten et al. (2014) used log-normal distribution to examine the decadal burned area against fire size, and we used power-law distribution to examine the fire frequency (i.e., fire number over a spatial extent and time span) against fire size, so the purposes are different. We agree the log-normal distribution is feasible for Lehsten et al. (2014), however are not convinced it's superior to the power-law distribution in characterizing the fire frequency. The Lehsten et al. (2014) mentioned that the log-normal distribution "might be questioned" since they "do not develop the decision to choose a log-normal distribution" (Page 8, paragraph 34 of Lehsten et al. 2014) and thus they did not prove the log-normal distribution is more appropriate than power-law distribution. Besides, when the sample size is small, it's very easy to confuse a power-law distribution with a log-normal one (for example, see discussions by Clauset et al., 2009). Finally, we found that the simple comparison of the fire size distribution would suffice our purpose to reveal the model behaviour and introducing the power-law distribution in fact complicated the comparison, so we decided to drop it in the revised manuscript.

-[2397, line 21] Be careful what you call tundra fires. These are still quite rare, and the model greatly overestimates them. In the observations many are in fact forests, just open or sparse in the northern limits.

[Response] We agree with the reviewer that tundra fires are rare and the sparsely forested area is more common. The text is now changed into "boreal forest (and sparsely forested area) or tundra", see the last paragraph of section 3.6 in the revised manuscript. The big fire size for the high-latitude (50^o-70^oN) forest, sparsely forested area or tundra is not overestimated by the model (Fig. 13 in the revised manuscript). The visual outlook of Fig. 4 might be that the burned area for high latitude region (50^o-70^oN) is overestimated compared with the GFED3.1 data but in fact the extra fires are of 0.1%~0.5% of annual burned fraction, which is very small. The simulated burned area for 50^o-70^oN agrees well with the GFED3.1 data. Please refer to see Fig 5 in the revised manuscript and relevant discussions in 2nd paragraph of section 4.2.5 and 4th paragraph of section 4.2.3.

-The high intensity tundra fires jump out at me as a large and somewhat surprising bias. Why is this happening in the model? This may have unfavorable implications for black carbon deposition, etc. They are also spreading incredibly fast, faster than in the tropics. Is this because of grass coverage in the static land cover map?

[Response] This is because the herbaceous plant is simulated as normal C3 grassland in the model, which has a small fuel bulk density and lead to high fire intensity. The result of fire intensity is removed; however the fast spread of these fires and relevant errors are now discussed in the second last paragraph of section 4.2.3 in the revised manuscript.

-[2403] “To fully represent the big fire process in reality, improvements need to be made to the model to allow fire to span multiple days when the climate is favourable: : :” Again, this has been done in Pfeiffer et al. 2013.

[Response] The fire patch size of reconstructed "multi-day fire patch" is used in comparison. The section 4.2.2 and 4.2.3 were re-written to include the discussions relevant with Pfeiffer et al. (2013).

-Regarding the influence of human ignitions. the authors could also cite Knorr et al 2013 (Biogeosciences Discussions) who show that human population seems to have little positive influence on fire occurrence except at very low densities, and even then it’s quite minimal.

[Response] Knorr et al. (2013) shows that it's the fire frequency (i.e., burned fraction) rather than fire numbers (i.e., ignitions) increases with population density when it's lower than 0.1 individual per km^2 but decreases under higher population densities. We now cite this paper in the manuscript.

Technical corrections

-[Title] I know a companion paper will focus on carbon, but as is this paper has no mention of carbon whatsoever, yet the title focuses on fires in the global carbon balance.

Also, I think the following would read better: “Modelling (the role of) fires in the: : :global burned area and fire regime(s)”

We changed the title to "Modelling the role of fires in the terrestrial carbon balance by incorporating SPITFIRE into the global vegetation model ORCHIDEE: Part 1. Simulating historical global burned area and fire regimes", according to the suggestion of the reviewer.

-[2379, line 1] I would suggest stating that fires help determine, or are one of the major determining factors, for the distribution of biomes. They are certainly not the sole determining factor, as this sentence implies.

We agree, the sentence is changed into "Fire is an important global ecological process that influences the distribution of biomes ... "

-[2379, line 7] As with the title, I think the plural ‘regimes’ is more appropriate, as there is no single global fire regime.

The "regime" is changed into "regimes".

-[2379, line 10] It’s quite unclear what this 78 - 92% number actually means. Please be more specific if possible.

We changed into "... 76–92% of the global burned area is simulated as collocated between the model and observation," and hope this is more clear.

-[2380, line 16] Would sound better as “: : :Earth system models is needed to investigate: : :”

We have changed to "Thus fire process and biomass burning emissions need to be included in the Earth system models, which are often used to investigate the role of fire in past, present and future biophysical and biogeochemical processes."

-[2381, line 3 - 5] This statement, particularly the word the ‘infrequent’, is certainly true for many boreal/temperate forests and even chaparral, but not for tropical savannas or grasslands where fire frequency is less than 5 years.

The "infrequent" large fires should be understood in the context of fire size distribution, i.e., in all ecosystems, fire size conforms to a heavy-tailed distribution and large fires are always rare. To avoid the misunderstanding, we changed to "... the magnitude and trend of burned area depend strongly on large fire events that represent only a low fraction in total number of fires ".

-[2381, line 17] Doesn’t ‘fire regime’ here also include intensity, as this is mentioned previously and included in the analyses.

As the fire intensity comparison is dropped in the revised manuscript, we changed into "... This allowed us to simulate global fire activity during the 20th century, and to perform an in-depth model evaluation. In present study, we focus on evaluating the ORCHIDEE-SPITFIRE model performance in simulating fire behaviours and regimes, including ignitions, fire spread rate, fire patch length, fire size distribution, fire season and burned area."

-[2387, line 8] VIRS, not IRS

Changed

-[2387, line 10] This should probably say “: : :by applying a modified version of the CASA model: : :”

Changed.

-[Figure 6] I think an annual mean would be easier to look at here. Perhaps consider adding a panel below. This could also be merged with Fig. 5.

The annual burned area series is shown and merged with Figure 5 to reduce the total number of figures.

-I think Fig. S4 is very helpful for visualizing the regional biases and could be moved to the main text.

it's now moved to the main text as Fig. 6.

-Figures 2 and 3 could potentially be merged to cut down on the total number of figures.

Figure 3 is removed as we find it's not really necessary.

-[2395, line 14] The variability in modeled burned area is much less than the data, which should be stated.

This is stated in section 3.5 first paragraph.

-[Discussion, first paragraph] There is no mention of Pfeiffer et al. 2013, which should be included.

It's now included.

References:

Clauset, A., Shalizi, C. and Newman, M.: Power-Law Distributions in Empirical Data, SIAM Rev., 51(4), 661–703, doi:10.1137/070710111, 2009.

Forkel, M., Thonicke, K., Beer, C., Cramer, W., Bartalev, S. and Schmullius, C.: Extreme fire events are related to previous-year surface moisture conditions in permafrost-underlain larch forests of Siberia, Environ. Res. Lett., 7(4), 044021, doi:10.1088/1748-9326/7/4/044021, 2012.

Giglio, L., Randerson, J. T. and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118(1), 317–328, doi:10.1002/jgrg.20042, 2013.

Hansen, M. C., Stehman, S. V., Potapov, P. V., Loveland, T. R., Townshend, J. R. G., DeFries, R. S., Pittman, K. W., Arunarwati, B., Stolle, F., Steininger, M. K., Carroll, M. and DiMiceli, C.: Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data, PNAS, doi:10.1073/pnas.0804042105, 2008.

Lehsten, V., de Groot, W. J., Flannigan, M., George, C., Harmand, P. and Balzter, H.: Wildfires in boreal ecoregions: Evaluating the power law assumption and intra-annual and interannual variations, J. Geophys. Res. Biogeosci., 119(1), 2012JG002252, doi:10.1002/2012JG002252, 2014.

Li, F., Levis, S. and Ward, D. S.: Quantifying the role of fire in the Earth system – Part 1: Improved global fire modeling in the Community Earth System Model (CESM1), Biogeosciences, 10(4), 2293–2314, doi:10.5194/bg-10-2293-2013, 2013.

Van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10(23), 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

Yang, J., Tian, H., Tao, B., Ren, W., Kush, J., Liu, Y. and Wang, Y.: Spatial and temporal patterns of global burned area in response to anthropogenic and environmental factors: Reconstructing global fire history for the 20th and early 21st centuries, Journal of Geophysical Research: Biogeosciences, n/a–n/a, doi:10.1002/2013JG002532, 2014.

Interactive comment on "Modelling fires in the terrestrial carbon balance by incorporating SPITFIRE into the global vegetation model ORCHIDEE – Part 1: Simulating historical global burned area and fire regime" by C. Yue et al.

Anonymous Referee #2
Received and published: 30 May 2014

The paper of Yue et al. documents the development of including the SPITFIRE algorithm into the global vegetation model ORCHIDEE. The paper is clearly written also the graphs are clear and in good quality. The study includes a number of datasets that were not used before in the evaluation of global fire models, e.g. Fire size and fire radiativ power. Although I think it is a great progress to use such datasets for the evaluation it is my main concern about the manuscript that the datasets and the model output are not comparable. For the fire size the datasets are not comparable because the model does not include multiple day burning, moreover the fire duration is limited to 4 hours.

Another factor is that the dataset based on remote sensing misses all the small fires. In Fig. 13 the study even focusses on the 95th quantile of fire size, these largest fires are likely to be burning for longer than the 4 hour limit in the model, therefore it cannot be expected that the model can reproduce this. Yue et al. emphasize that the fire size is very important, but what difference does it make in the model whether the area was burned in one fire or by two fires? Does it make a difference in ORCHIDEE?

For the fire radiative power even the units are different between the satellite data set and the modelled variable, the FRP is per area, the fireline intensity per m. Therefore the fireline intensity in addition to the energy released per area burnt includes how fast the fire was spreading. This may cause differences not only in the absolute values but also in the spatial patterns. I suggest that either the comparison is removed from the paper or a more quivalent fire radiative power is derived from the model output. I think that this should be possible for instance by reverting the procedure of GFAS, where they derive the carbon emissions based on fire raditative power.

[General response] We thank the reviewer for the general positive comments. In response, we have developed an approach to group the fires occurring within consecutive days into "multi-day fire patches", to make the fire size of these fire patches being able to be compared with the observation data which also contains multiday fires. Although fire size does not make a difference in current model simulation (for example, on the combustion completeness of fuels), however it's important to check whether the model could capture the fire size distribution because of at least two reasons (as described in the introduction of our manuscript). First, this will help to diagnose the model error in burned area simulation. Second, the big fires have more severe social and economic consequences and it's important for the model to be able to predict them.

We agree with the reviewer that FRP and simulated fireline intensity is not strictly comparable although their spatial pattern could be similar for the extremely big, fast-spread fires. To avoid any potential misleading, we decide remove this from the manuscript. All the relevant modifications and changes are included in the revised manuscript. To make it easy to follow the revised contents in the manuscript, the heavily modified texts are shown in blue in the revised manuscript.

p.2383, l. 20: maxmimum fire duration is 240 (actually 240.0937) minutes if the equation was not adjusted.

[Response] We used 241minutes as the maximum fire duration time, following the equation (14) on Page 997 of Thonicke et al. (2010).

p. 2384, l. 25: why an additional parameter? was it not possible to increase the necessary fire intensity?

[Response] The original fireline intensity threshold (50kW m^{-1}) is mainly an expert judgement based on what's been described in Pyne (1996). Ideally, the intensity threshold should depend on fuel load (relating to the amount of energy provided by fire), fuel moisture (relating to the energy needed to heat the adjacent fuel to the ignition temperature), and the fraction of fire-released energy used to heat the adjacent fuel. It's thus difficult to set a single intensity threshold due to the lack of observation data. On the other hand, the use of fuel load to limit the ignition efficiency has been used by other authors (Arora and Boer, 2005; Kloster et al., 2010; Li et al., 2012). The Fig. S2 in our study has also shown that, using the fuel load dependent ignition efficiency has improved the agreement of simulated burned area with the observation for the arid and semi-arid (fuel-limited fire occurrence) regions.

p.2385, l. 8-9: why do you compare the observed "mean" combustion completeness to the "maximum" combustion completeness in the model?

[Response] The combustion completeness (CC) for different fuel types (1h, 10h, 100hr, 1000hr) is simulated as a function of fuel wetness in the model. The fuel wetness is defined as the simulated fuel moisture divided by the prescribed PFT-dependent moisture of extinction. We're calibrating the "maximum" CC as the "mean" value of observed CC, because during the model test we found that the simulated burned area is dominated by very low fuel wetness level (see Figure C1). Given the scheme used to simulate the CC (Fig. 1 in the main text), the mean CC will approach to the maximum CC. So by calibrating the maximum CC as the mean observed, we hope the simulated CC will be realistic. We make the calibration in this way rather than completely drop the CC simulation scheme (and use fixed values of CC), because the scheme used here allows the refinement in the future when more detailed CC observations are available to carefully calibrate the parameters.

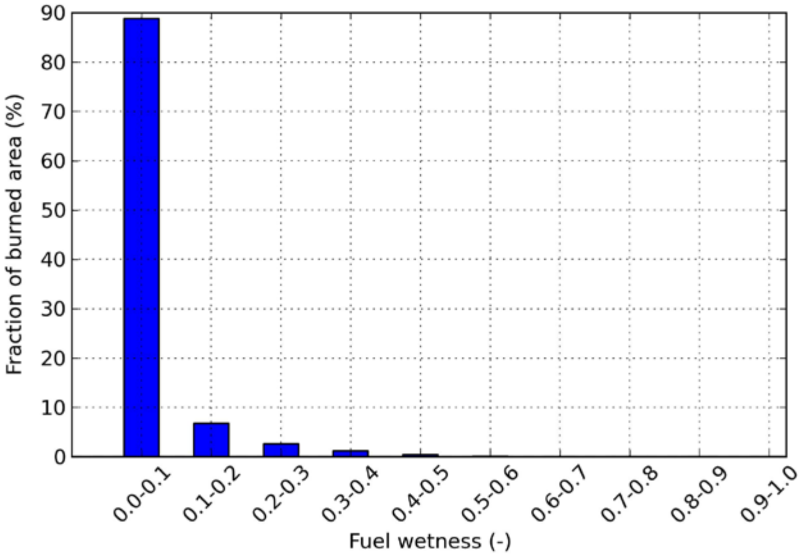

Figure C1. The distribution of global burned area in terms of simulated fuel wetness for 2006. All the 0.5^{o} pixels with fire occurrence across the globe are categorized into ten classes in terms of fuel wetness (shown as the horizontal axis), with the fraction of burned area (in terms of percentage, %) for each fuel wetness class being shown as the vertical axis.

p. 2385, l. 26: what is the reason for the initial spinup without fire? Having the equilibrium of soil pools without the influence of fire should lead to overestimated soil carbon pools and therefore overestimated respiration.

[Response] The intuitive reason to do this initial spinup without fire is to save computation time, as a system without fire would allow fast accumulation in the carbon stocks. All carbon pools (live biomass, aboveground litter and mineral soil) except belowground litter have slightly decreased when the spinup simulation shifted from a fire-free state to a state in which fires are prognostically simulated (Figure C2). The mineral soil carbon stock has been verified to vary within 0.08% during the last 50 years of the spinup. We agree with the reviewer that, even with this decrease in the mineral soil carbon when the fire module was switched on, this carbon stock might still be overestimated given the short time to simulate the fire occurrence before entering the transient simulation. However, we expect the resulting overestimation in respiration would be small because most of these overestimated carbon stocks reside in the passive sub-pool of the mineral soil in the model with a default turnover rate of ~1500 years at 5^{o}C average annual temperature.

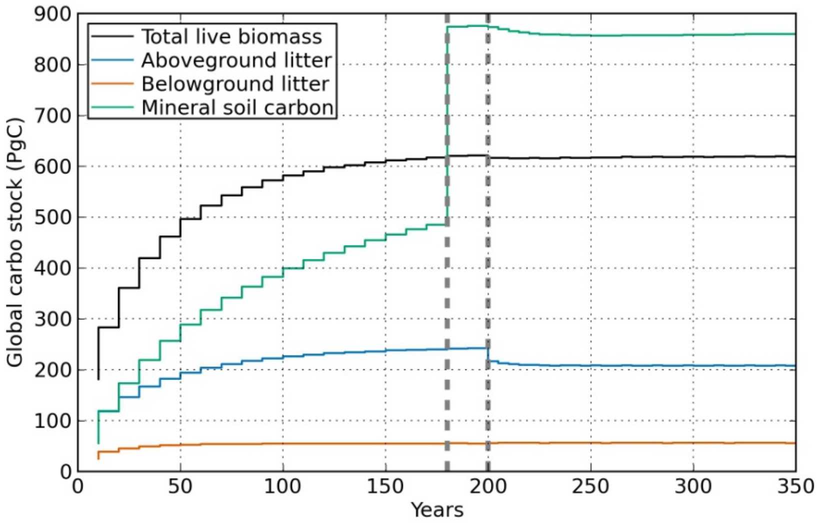

Figure C2 The evolution of global total carbon stock for the live biomass, aboveground litter, belowground litter and mineral soil carbon stock during the spinup simulation. The first vertical dashed gray line indicates that the soil-only processes in ORCHIDEE have been run for 3000 years to speed up the accumulation of mineral soil carbon; and the second vertical dashed gray line indicates the switch-on of the fire module.

p. 2386, l. 15: no land cover change: this may strongly modify the evolution over the 20th century.

[Response] Theoretically, over the long term, the land cover changes (of which the transformation of forest to crops or managed grasslands, and from natural grassland to cropland matter fire the most) will reduce the burned area because the majority of historical land cover change was from forest to managed grassland or croplands; and reduced the forested area available for burning. Over the short term it's rather complex, and might depend on the fire frequency of the land cover before and after transformation. For example, converting a forest with 100-year fire return interval (FRI) by fire into a non-burning cropland will increase the burned area for the year when the forest is burned, but will reduce the burned area for the following 100 years. However, if the cropland is burned each year after the conversion, then the land cover change will increase the burned area. Thus its overall impact on the temporal trend and variation of burned area could be complex.

Very few studies quantified the burned area contributed by land cover change and its net effect on the temporal trend and variability of burned area. Kloster et al., (2010) found that the deforestation and wood harvest for 1850-1990 together reduced the fire carbon emissions in the 1990s by 16% (433 Tg C/year), however the net amount of burned area might be small (assuming a 3000gC m^{-2} of carbon consumption in deforestation fire, the net amount of burned area is 14.4 Mha, or 4% of annual global burned area). We have inserted the following sentence at the end of 3rd paragraph of section 2.3, "This static land cover could affect the model-observation agreement in terms of long-term trend and variation of burned area for regions where land use change fires dominated the burned area.". The regions where model-observation agreement could be affected by the lack of land cover change in the simulation were discussed in section 3.5.

p. 2386: l. 25-29: strange sentence

[Response] We rephrased this part of texts and added Fig. S2 and Fig. S3 for further explanation in the revised manuscript and hope it's more clear, please refer to last paragraph of section 2.3.

p. 2390: l. 10: what happens in grid cells where GFED equals zero? maybe using (GFED+model)/2 could help to be able to include all gridcells in the evaluation (except the ones where both are zero).
[Response] This comparison was done for each GFED region so there is no chance for either the model or GFED burned area to be zero. We add in the revised manuscript that "The evaluation has been done for each GFED3.1 region" (section 2.5.1, first paragraph) to make this clear.

p. 2390, l. 12-14: why do you use the monthly time series for the interannual variation?
[Response] This was originally done to account for the intrinsic seasonality of burned area, but we agree it's better to use the annual burned area time series and now included in the revised manuscript (section 2.5.1).

p. 2390, l.15:eq 3: It took me a while to understand how this is can quantify the similarity in the seasonality. Please explain this is a bit more. for instance explain that the value will be low if the similarity is low, 1 for perfect correlation, what is the value if you compare anticorrelated time series or random time series? what is the advantage compared to a correlation or rank correlation coefficient?
[Response] We argue this indictor (i.e., the overlapping area of the two normalized monthly burned area time series) might be better than the regression slope or correlation coefficient as it retains the physical meaning (i.e., the fraction of burned area in overlapping months against the total annual burned area). In response to the reviewer's comment, a bootstrapping method was used to associate the derived seasonal similarity (S_{season}, see Equation 3 on page 2390 in the discussion paper) with some statistical significance (i.e., the probability that Sseason is from a random distribution of seasonality). This is described in section 2.5.1 and Table 2 in revised manuscript.

p. 2391, l. 28: what are the exact definitions for the categories?
[Response] The exact definitions for the categories are explained in the capital of Fig. 4. Fig. 4 is referred to in the text when the comparison is presented (section 3.1). The definitions are not repeated in the text in order to avoid redundancy.

p.2392, l. 15: Now for the interannual values, the time series is smoothed? please mention this also in the methods section. You could also briefly mention the advantage of a smoothed time series, compared to annual average. In the figure, the seasonality still strongly disctracts from the interannual variability. Where are the correlations mentioned in the methods section? spatial a_nd improvements, possibly also show that the interannual variabitliy is strongly influenced by the african continent?
eq.4: move to methods part. You already have a measure for the seasonality, why are you using another one? Please move to the method section
[Response] The annual series is used in the Fig. 5 in the revised manuscript and correlations are provided at the end of section 3.1. Eq (4) and relevant descriptions are moved to section 2.5.1 in the revised manuscript. The peak fire season, fire season length, and the seasonal similarity are different metrics. The metric in Eq (4) is intended to measure on the global scale the agreement in terms of fire peak month, and can complement the seasonal similarity.

p. 2395, l. 25: mention the model does not include land use change. eq 5. move to methods section.
[Response] The use of static land cover in the simulation is included in section 3.5 (fifth last line) in the revised manuscript. The Eq (5) and the power-law regression analysis is removed because we think the simple comparison already suffice for our purpose (see also the response to the comments by #1 reviewer at Page 11).

p. 2396, l. 17: how big is the minimum fire size in the model and how does this influence the comparison?

[Response] As the power-law regression is removed, the minimum fire size does not matter.

p. 2396, l. 26-27: You state that the simulated fire distribution is skewed towards small fires, big fires are underestimated. The definition for the fire size in the dataset and model is (in my understanding) fundamentally different). The fire size in the datasets used include large fires, that burned over multiple days. In case of the fire model the fire duration is limited to only four hours. but a new fire may start the next day. Therefore this is not surprising. If you can include multiple day burning or estimate from the satellite the size of the fires burned per day (you mention with one dataset that the start and end day are reported) the comparison may therefore be confounded.
[Response] We reconstructed the "multi-day fire patch" to make the simulated fire size being able to be compared with observation. See section 2.5.2 for the method, section 4.2.2 and 4.2.3 for relevant discussions in the revised manuscript. We also use the fire start and end date in the Canadian fire agency data to calculate the fire patch length and compared with the model, see Fig 11c and section 3.6 in the revised manuscript.

p.2397, l. 1-12: This may be strongly influenced by the multiple day burning issue. Does the minimum fire size in the satellite data influence this result?
[Response] The comparison is improved by using the reconstructed "multi-day fire patch" size, the results are updated accordingly, see section 3.6 the third paragraph in the revised manuscript.

p. 2398, l. 1-10: this is the first study that makes use of the FRP global datasets for model evaluation. This is great, but I think the consistency of the comparison can be improved. The units of the two variables are different: FRP is given per m^2, fireline intensity per m. The FRP is the energy of consumed fuel per m-2 burned area. This estimate could be derived as well from the model. The fireline intensity includes the rate of spread as a factor. For the FRP a fast fire that consumes little fuel can have the same FRP value as a slow fire consuming a large amount of fuel. The two datasets are therefore not spatially consistent and it is unclear what the comparison of the two variables with different units can mean. The rate of spread adds spatial patterns to the fireline intensity that may not occur in the FRP datasets. Therefore even when focusing only on the spatial patters the two variables are not comparable. The comparison could be improved by reverting the procedure that is performed when estimating carbon emissions from FRP, this could help to achieve consistency between model and data.
[Response] We agree with the reviewer that these two variables (fireline intensity or FLI, and FRP) are not strictly comparable. The fireline intensity (FLI, in kWm^{-1}) represents the heat transfer per unit length of the fireline, which could be derived as the product of the energy released per square meter (kJm^{-2}) by fire, and the fire spread rate (ms^{-1}) (Byram, 1959). The MODIS FRP measures the radiative energy released in an actively burning fire, by examining the difference in the reflectance of the middle infrared (MIR) band of actively burning pixel and the background pixel.

Because the FRP measures the fire radiative energy on the pixel basis, a pixel with a small fraction being intensively burned will have the same FRP as the pixel with a large fraction being burned but less intensively. This pattern has little relation with the fire spread rate. Besides, the energy loss in fires in forms of conductive and convective energy is not accounted in the FRP (Wooster et al., 2005). However, the FLI accounts for all the energy forms and is closely related with the fire spread rate. Fires with lower fuel consumption (thus less energy release and probably smaller FRP) but with fast spread might have the same FLI as the fires with higher fuel consumption (thus higher energy release and probably larger FRP) but with a low spread rate. Thus their spatial pattern might no be exactly comparable.

Smith and Wooste (2005) reported one approach to derive the FLI from the MODIS FRP. The radiative FLI is derived by dividing the total FRP of the fire front pixels by the fire front length, which is retrieved from the visible imagery. However their derived radiative FLI is one magnitude lower than that from field observations, and their approach does allow large-scale, automatic FLI generation. Before the availability of large-scale FLI observation from satellites, the comparison between simulated and observed FLI will have to be limited on site (or regional) level.

One might think that the total amount of energy (in W or kW) released by fires per square meter simulated by the model (e.g., the product of reaction intensity with fire duration time) could be compared with the MODIS derived fire radiative energy (FRE), which is the FRP being integrated with time. But again, the FRE suffers the same deficiency as FRP that it does not include all the energy forms that a fire releases. The reviewer suggests reverting the GFAS (Kaiser et al., 2012) processes, i.e., if we understand well, probably to use the vegetation-type-dependent conversion factors (Page 533, section 2.3, second paragraph of Kaiser et al., 2012) to adjust the MODIS derived FRE. However, the conversion factors used in GFAS are empirical ratios linking GFAS FRE and GFEDv3.1 emission data. By doing this we are finally in fact comparing the model simulated fuel consumption with GFEDv3.1 data, which is not we want.

So ultimately, the most reliable way is to either have large-scale FLI observations, or compile field-based FLI database which allow direct model-measurement comparison. This could be left for another targeted study and goes beyond the scope of our current one. Finally, to avoid any misleading information by presenting this FLI-FRP comparison, we decide to remove this comparison from our results.

p. 2399, l.11-15: Please mention here that land use was not included in the model simulation.
[Response] The static land cover is included at the end of section 4.1.

References:

Arora, V. K. and Boer, G. J.: Fire as an interactive component of dynamic vegetation models, Journal of Geophysical Research: Biogeosciences, 110(G2), G02008, doi:10.1029/2005JG000042, 2005.

Byram GM (1959) Combustion of forest fuels. In 'Forest Fire: Control and Use'. (Ed. KP Davis) pp. 61–89. (McGraw-Hill: NewYork)

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M. and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9(1), 527–554, doi:10.5194/bg-9-527-2012, 2012.

Kloster, S., Mahowald, N. M., Randerson, J. T., Thornton, P. E., Hoffman, F. M., Levis, S., Lawrence, P. J., Feddema, J. J., Oleson, K. W. and Lawrence, D. M.: Fire dynamics during the 20th century simulated by the Community Land Model, Biogeosciences, 7(6), 1877–1902, doi:10.5194/bg-7-1877-2010, 2010.

Li, F., Zeng, X. D. and Levis, S.: A process-based fire parameterization of intermediate complexity in a Dynamic Global Vegetation Model, Biogeosciences, 9(7), 2761–2780, doi:10.5194/bg-9-2761-2012, 2012.

Pyne, S. J., Andrews, P. L., and Laven, R. D.: Introduction to wildland fire, 2 edition, Wiley, New York, 769 pp., 1996.

Smith, A. M. S. and Wooster, M. J.: Remote classification of head and backfire types from MODIS fire radiative power and smoke plume observations, Int. J. Wildland Fire, 14(3), 249–254, 2005.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L. and Carmona-Moreno, C.: The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model, Biogeosciences, 7(6), 1991–2011, 2010.

Wooster, M. J., Roberts, G., Perry, G. L. W. and Kaufman, Y. J.: Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release, J. Geophys. Res., 110(D24), D24311, doi:10.1029/2005JD006318, 2005.

Response supplement material

1. Reconstructed lightning flashes with interannual variability

The interannual variability of lightning flashes is interpolated form the average monthly satellite observed lightning flashes of LIS/OTD data (http://gcmd.nasa.gov/records/ GCMD_lohrmc.html), by using the interannual variability of the Convective Potential Available Energy (CAPE) during the 20th century as simulated from by the 20th Century Reanalysis Project. The interpolation is done by following the method of Pfeiffer et al. (2013, Equation 1 on Page 649).

$$l_{\mathrm{m}} = \begin{cases} \mathrm{LISOTD_m}\,(1+9\mathrm{CAPE_{anom}}), & \mathrm{CAPE_{anom}} \geq 0 \\ \mathrm{LISOTD_m}\,(1+0.99\mathrm{CAPE_{anom}}), & \mathrm{CAPE_{anom}} < 0 \end{cases}.$$

where l_m the monthly lightning flash numbers for a given month, $CAPE_{anom}$ is CAPE anomaly for the concerned month being normalized to (-1,1) for 1901-2011.

We first compared the reconstructed lightning flashes with the observation by the Alaskan Lightning Detection System for 1986-2011 (Figure 1). Their correlation coefficient is 0.48 (data not detrended).

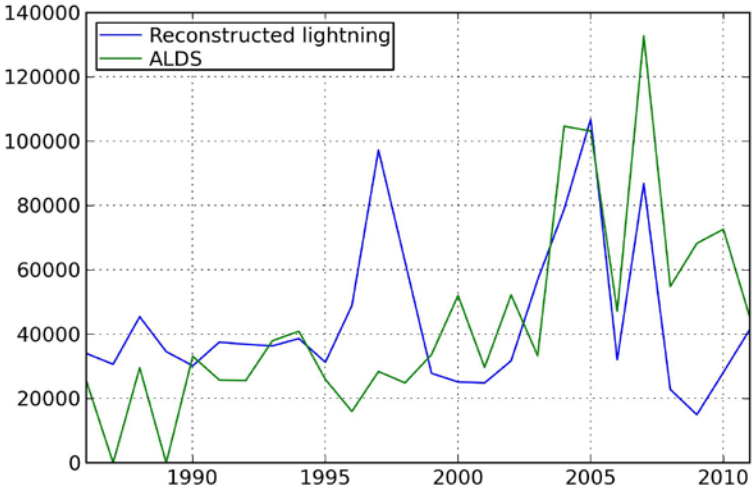

Figure 1. The reconstructed lightning flashes compared with the lightning flashes observed by the Alaskan Lightning Detection System (ALDS) for 1986-2011. To facilitate the comparison of interannual variability, the mean annual lightning numbers of reconstructed CAPE-derived data are adjusted to have the same mean annual lightning flashes as observed by ALDS.

2. Compare the simulated burned area with observation data by using different lightning input data

After the reconstruction of the interannual lightning flashes, we launched a global simulation for 1901-2011 by using the new lightning data with the human ignition parameters of a(ND) (Equation 1 in the discussion paper, Page 2382) as the spatial dataset used in Thonicke et al. (2010). This simulation is denoted as "ORCHIDEE - IAVLightn", and another simulation with mean annual lighting data and the spatial a(ND) dataset is denoted as "ORCHIDEE - CONLightn". Note that the reconstruction of interannual lightning data changed the total amount of flashes, so a constant scaling factor (0.53) has been applied in the "ORCHIDEE - CONLightn" simulation, to ensure on the global scale, the same lighting ignition efficiency factor (0.03) in the original simulation to be maintained (i.e., on the global scale, the mean annual potential lighting flashes available for ignition do not change) over 1901-2011.

Furthermore, we launched a third simulation for Alaska for 1986-2011, using the observed ALDS lightning flashes as input data, and this simulated is denoted as "ORCHIDEE - ALDS". The third simulation allows investigating the simulation improvement by using the ground-based observation of lightning flashes.

2.1 Compare burned area over Alaska

The simulated burned area over 1986-2011 is compared with GFED3.1 burned area data and the burned area by Alaskan fire agency, by using the Pearson correlation coefficient (r-value). The results are shown in Table 1. The increase in r-value (with the Alaskan fire agency data) by shifting from "CONLightn" to "IAVLightn" is very small (0.19 to 0.22). The r-value between simulated BA with the fire agency BA is the highest for the simulation using the ALDS input (0.5), though still lower than that of 0.66 by Pfeiffer et al. (2013) for the "Intermontane Boreal" ecoregion of Alaska who used the same lightning input (the r-value is derived by picking up the data from the Fig. 7 on Page 663 of Pfeiffer et al., 2013). Over 1950-2011, the r-value decreased from 0.41 for "ORCHIDEE - CONLightn" simulation to 0.37 for "ORCHIDEE - IAVLightn" simulation.

We found that using the CAPE-derived interannual lightning data only marginally improved the BA simulation for Alaska for 1986-2011, but using the ground-based observation of lightning data did greatly improved the simulation.

Table 1 Pearson correlation coefficient (r-value) for different annual simulated burned area data with the observation data by the Alaskan fire agency; and the r-value for different data with the ALDS observed flashes.

	1950-2011	1986-2011	Correlation with Alaskan ALDS lightning flashes (1986-2011)
Alaskan Fire Agency	1.00	1.00	0.55
GFED3.1		0.98	0.58
ORCHIDEE - CONLightn	0.41	**0.19**	0.20
ORCHIDEE - ALDS		0.50	0.62
ORCHIDEE - IAVLightn	0.37	**0.22**	0.50

2.2 Compare the simulated burned area with the observation for boreal North America (Alaska, US + Canada)

We examined the agreement between the simulated and observed BA for the two global ORCHIDEE simulations (with CONLightn and IAVLightn) for the boreal North America (Alaska, US + Canada). Burned area in this region is known to be dominated by lightning sources, and thus we expect the improvement in the simulation is expected to occur for this region. We used both the annual fire agency burned area data and the decadal Mouillot and Field (2005) as the observation data. The r-value between different data are shown in Table 2. Surprisingly, for all r-values, the ones by "ORCHIDEE- IAVLightn" is lower than that by "ORCHIDEE - CONLightn", suggesting that ***shifting from mean annual lighting data to CAPE-derived lightning data has generally decreased the model-observation agreement in this region.***

Table 2 The Pearson correlation coefficient (r-value) for the period after 1950 in terms of BA by different data (because after 1950 the fire agency data began to exist). The ***bold italic numbers*** indicate that the agreement with fire agency data deteriorated after shifting from "CONLightn" to "IAVLightn".

	ORCHIDEE - CONLightn	ORCHIDEE - IAVLightn
Annual correlation (n=61)		
ORCHIDEE ~ Fire Agency	***0.44***	***0.41***
ORCHIDEE ~ Mouillot & Field (2005)	***0.57***	***0.44***
Mouillot & Field (2005) ~ Fire Agency	0.92	0.92
Decade correlation (n=6)		
ORCHIDEE ~ Fire Agency	***0.42***	***0.27***
ORCHIDEE ~ Mouillot & Field (2005)	***0.81***	***0.62***
Mouillot & Field (2005) ~ Fire Agency	0.91	0.91

2.3 Compare the simulated burned area with the observation over the 20th century for different Mouillot & Field (2005) regions

We compared the decadal r-value over the 20th century with the Mouillot and Field (2005) reconstructed BA data as shown in Table 3. When examining the r-value for different regions, for some regions the BA are rather poorly simulated by the model with negative r-values (indicating anti-phase between model and observation). Over the whole globe, the r-value after shifting to IAVLightn slightly decreased (by 0.1). Of the 14 region, the r-values decreased after shifting to IAVLightn for 6 regions, with 2 regions showing no change in r-value, and 6 regions with increase in r-value. ***On the global scale, the model-observation agreement decreased after shifting to the CAPE-derived lightning data, and for half the regions the agreement increased and the other half decreased.***

Table 3 The Pearson correlation coefficient between simulated decadal BA and Mouillot and Field (2005) reconstructed BA over the 20th century (n=11). The negative r-values (poor simulation and anti-phase between model and data) and the decrease in r-value after shifting to IAVLightn are shown in red.

	CONLightn (r1)	IAVLightn (r2)	Improvement (r2-r1)
Global	0.6	0.5	-0.1
Australia	-0.4	-0.5	-0.1
BONA	-0.4	-0.5	-0.1
BOAS	-0.1	0.3	0.4
India	0.8	0.6	-0.2
SouthEastAsia	0.0	0.4	0.4
CentralAsia	0.4	0.3	-0.1

WestUS	-0.6	-0.9	-0.3
EastUS	0.1	0.4	0.3
EastAsia	-0.6	-0.7	-0.1
MiddleEastNorthAfrica	-0.6	-0.5	0.1
Africa	-0.5	0.0	0.5
CentralSouthAmerica	0.8	0.8	0.0
SouthAmerica	-0.6	-0.2	0.4
Europe	0.1	0.1	0.0

2.4 Compare the annual simulated burned area with GFED3.1 data for 1997-2009 for the 14 GFED regions

The Pearson correlation coefficients between annual simulated BA with GFED3.1 BA have been calculated for different GFED regions and the globe for simulations with CONLightn and IAVLightn (Table 4). The annual time series of burned area are shown in Figure 2. ***Over the globe, the model-observation agreement decreased, and for only two out of the 14 regions, the r-value increased after shifting to IAVLightn.***

Table 4 The Person correlation coefficient (r-value) between annual simulated BA with the GFED3.1 data for different GFED regions. The negative r-values (i.e., poor simulation of model) and the decrease in r-value after shifting to IAVLightn are shown in red.

	CONLightn (r1)	IAVLightn (r2)	Improvement (r2-r1)
Global	0.5	0.3	-0.2
BONA*	0.5	0.7	0.2
TENA	0.3	0.1	-0.2
CEAM	0.2	-0.1	-0.3
NHSA	-0.1	0.0	0.2
SHSA	0.3	-0.5	-0.9
EURO	-0.1	-0.1	0.0
MIDE	0.3	0.1	-0.2
NHAF	0.2	-0.2	-0.3
SHAF	0.0	0.0	0.0
BOAS	0.4	0.0	-0.4
SEAS	-0.1	-0.4	-0.3
CEAS	0.2	0.0	-0.2
EQAS	1.0	1.0	0.0
AUST	0.2	-0.1	-0.3

* This is not in contradiction with results presented in Section 2.2 as the spatial extend of boreal North America and the BONA here are slightly different. The BONA includes part of the western US where the model overestimated BA.

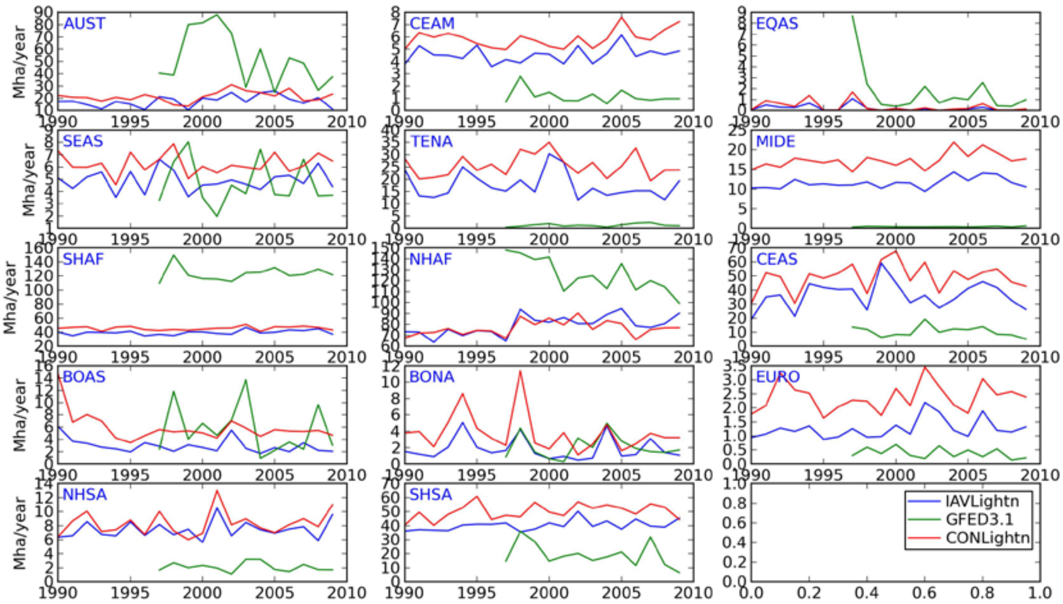

Figure 2 The annual BA time series for different GFED regions for 1997-2009 by GFED3.1 data, and the two model simulations ("ORCHIDEE - CONLightn" and "ORCHIDEE - IAVLightn").

2.5 Compare simulated global BA with GFED3.1 data

The total global BA is 273 Mha yr^{-1} according to "ORCHIDEE - IAVLightn" simulation for 1997-2009 (compared with 342 Mha yr^{-1} for "ORCHIDEE - CONLightn" and 349 Mha yr^{-1} for GFED3.1). Figure 3 shows the annual BA time series of ORCHIDEE and GFED3.1, with the r-value of linearly detrended annual time series between "ORCHIDEE - IAVLightn" and GFED3.1 is 0.46 (compared with 0.57 between "ORCHIDEE - CONLightn" and GFED3.1). There is no significant change in the spatial distribution of fires (pixel-to-pixel correlation between "ORCHIDEE - IAVLightn" and GFED3.1 is 0.481, and 0.475 between "ORCHIDEE - CONLightn" and GFED3.1). ***Thus if the global total potential available lightning ignitions over 1901-2011 were conserved in the simulation, the simulated global burned area decreased from 342 to 273 Mha yr^{-1} for 1997-2009 when shifting to the CAPE-derived lighting data, and the model-GFED3.1 agreement in the global burned area interannual variability decreased.***

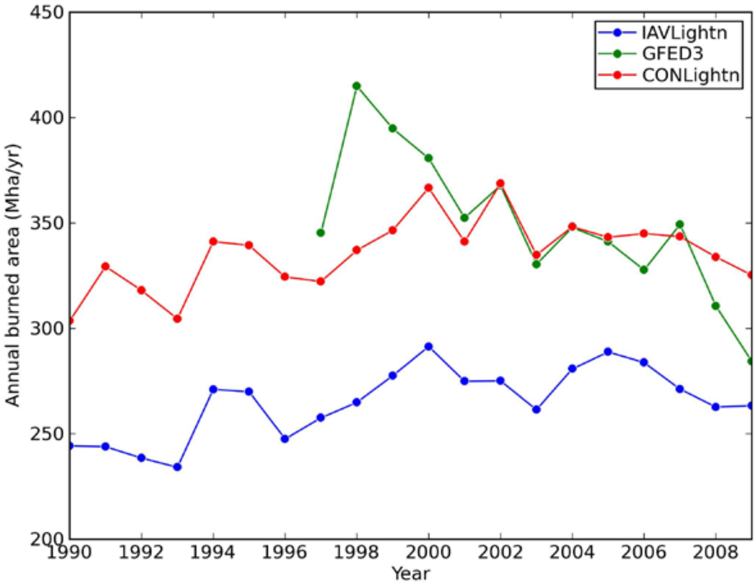

Figure 3 Annual global burned area by model simulation and as given by GFED3.1 data for 1990-2009.

3. Summary

We have followed the method proposed by Pfeiffer et al. (2013) and reconstructed the total lighting flashes with interannual variability for 1901-2011 by using the CAPE data. The new CAPE-derived lightning data moderately agreed with the ground observations of lightning flashes for Alaska for 1986-2011. However, the model-observation agreement for the burned area in Alaska for 1986-2011 has only been marginally improved by using the new CAPE-derived lighting data, compared with repeating the mean annual lightning data without interannual variability being included. For 1950-2011, the model-observation agreement slightly decreased after shifting to the new CAPE-derived lighting data. Large improvement in the simulation was found when the model was directly driven by the locally observed lighting data.

The agreement of simulated burned area with the observation data for 1950-2011 for the boreal North America (i.e., US Alaska + Canada) generally decreased after shifting to the CAPE-derived lightning data, either on annual or decadal basis. Over the 20th century, the shifting of lightning data decreased the agreement of simulated decadal burned area with the Mouillot and Field (2005) reconstruction for half of the 14 regions and increased for the other half. Especially, over 1997-2009 when the observation data by the GFED3.1 is more credible than the 20th century reconstruction, shifting of the lightning data decreased the agreement of annual simulated and observed burned area for the globe and for most of the regions.

The fact that the CAPE-derived lightning data does not systematically improve the model performance could be linked with several explanations. First, despite the physical linkage between the CAPE (atmospheric instability) and the lightning activity, the approach (equation) used here might not apply for all the regions of the globe, as it's mainly derived by the lightning observation in Alaska. Second, the errors in the CAPE data provided by the 20th Century Reanalysis Project might also contribute. Third, the uncertainties of internal model processes might have counteracted some of the expected improvement gains. For example, in Alaska, the complete replacement by local lightning observations only increased the model-observation correlation from 0.19 to 0.5, while less improvement in the lighting input data (the correlation of 0.48 between

ALDS and CAPE lightning data could be considered as an improvement in the input data compared with the otherwise mean annual lighting data) leads to nearly negligible improvement in the simulation result (r-value 0.19 to 0.22).