Dear two anonymous reviewers,

We greatly appreciate your constructive comments to improve our manuscript. A point-by-point reply to the comments is as follows.

[Reviewer 1]

1. Page 16775, line 8 and Table 4: "All of the simulations have a low bias over Africa..." Clearly Mod-new is a major improvement, but the low burned area bias remains substantial (~34%) in southern-hemisphere Africa. Are there reasons this might be expected? Please clarify. A similar burned area bias also seems to occur in northern Australia but is not mentioned in the text.

   Reply: According to your suggestion, in revised Para.5 of Sect. 5.1, we have discussed the reason for simulation bias of Mod-new in southern-hemisphere Africa and described the simulation bias of burned area in northern Australia:

   "In Southern Hemisphere Africa, almost all of the present-day fires are managed by humans for various purposes and the fire management is affected by the culture and socioeconomic status (Saarnak, 2001; Hoffmann et al., 2009). The new fire model only captures the primary statistical relationship between fires and population density/GDP on a global scale, and may miss some useful information about the human dimension of fires in the region. This may be a reason for the underestimation of burned area in Mod-new (Fig.9, Table 4). In addition, all of the CLM4 simulations underestimate annual burned area fraction in savanna region of northern Australia (Fig. 9, GFED3: >10% yr^{-1}, Mod-new: ~2% yr^{-1}, Mod-old: ~1% yr^{-1}, Mod-CTEM: ~0.1% yr^{-1}). The underestimation may be related to the simulation bias of aboveground biomass in CLM4. In savanna region of northern Australia where fuel load is the main limit factor of fire occurrence (van der Werf et al., 2008), simulated aboveground biomass in CLM4 is clearly lower than the reports based on inventory plots and satellite data (William et al., 1998; Saatchi et al. 2011)."

2. Technical Corrections

   (1) Page 16756, line 14: "...the product of the fire counts and average burned area of a fire."

   Reply: It has been changed to "Burned area in a grid cell was estimated by the product of fire counts and average burned area of a fire."

The sentence describes the basic equation of our earlier fire scheme (Li et al., 2012a, Biogeosciences): $A_b = N_f a$, where A_b (km^2 (time step)$^{-1}$) is burned area in a grid cell per time step; N_f (count (time step)$^{-1}$) is fire counts in the grid cell; a (km^2) is average fire spread area of a fire. Our earlier fire scheme (Li et al., 2012a, Biogeosciences) is applied for region C in the present study (see Para.1 of Sect. 3)

(2) Page 16756, line 20: "...the fire counts have MODIS observations..." Might be clearer to say that the fire counts are derived from, or are obtained from, MODIS observations.
?
Reply: the sentence has been changed to "Second, MODIS provides observations for the fire counts…".

(3) Page 16765, line 21: Consider replacing the phrase "run out of" with "spread beyond" for clarity.

Reply: Changed

(4) Page 16770, line 13: Change "First, effect of fire temperature..." to "First, the effect of fire temperature..."

Reply: Changed

(5) Page 16811, Figure 15 - Three-dimensional pie charts are notorious for distorting the data they portray. I suggest using a two-dimensional pie chart or a table instead.

Reply: Changed

(6) Fig. 1: The text "Wire Diagram" is not needed in the diagram.

Reply: Changed

(7) Fig. 3: Correct spelling of indicators in caption.

Reply: Changed

(8) Fig. 16: Please shift horizontal mapping range to be consistent with previous figures and to avoid splitting Africa into two pieces.

Reply: Changed

[Reviewer 2]

1. Why did the historical simulation start in 1850? It seems a bit close to the peak in burned area that seemed to have happened in the 1870ies as discussed p 16779. As of now, it looks like this peak is simply an incomplete spin-up of the model. Linked to this, I don't quite understand what in the drivers of the model can explain this peak. If I understand well, it is not the climate (same 25 years cycled). From the description of the data, the only available data that covers that period is population density from the HYDE database. Is that right ? So, why a peak in 1870 ? This should be further discussed and the time frame of the databases better described (a table may be useful). I also don't understand what is the driver of the continuous decrease in the case without anthropogenic influence. This is very interesting and should be discussed in more detail (may be in another paper ?).

Reply: (1)About 1850

The historical simulation starts in 1850 is set by CESM1's I20TRCN component set. We have clarified this in the revised version (see the reply to your comment 2).

The setting of 1850 in CESM1's I20TRCN is because models that will contribute to scenario analysis for IPCC AR5 report are required to provide an historical run which starts in 1850 (Taylor et al., 2009).

(2) About spin-up

1850-2004 simulation in I20TRCN is a transient simulation rather than an equilibrium simulation to represent a realistic mode of land state change. According to the setting of CESM1, the initial data for I20TRCN component set is the equilibrium state of CLM4 with carbon/nitrogen biogeochemistry driven by cycling 25-yr (1948−1972) of Qian et al. (2006)'s atmospheric observations with CO_2 concentration, nitrogen and aerosol deposition, and land cover in 1850. We have clarified this in the revised version (see the reply to your comment 2).

(3) About time-varying input data
Besides population density from the HYDE database, the transient time-varying CO_2 concentration, nitrogen and aerosol deposition, and land use and land cover change data (wood harvesting included) from 1850 to 2005 used in the I20TRCN also cover the period. We have clarified this in the revised version (see the reply to your comment 2).

(4) About driver for the peak in 1870
The main driver for the peak around 1870 is direct anthropogenic influence on fires according to the comparison between historical burned area simulations with and without direct anthropogenic influence in Fig. 18 (discussed in the Para. 4 of Sect. 6). In the simulation without direct anthropogenic influence (i.e., the deforestation fires, crop fires, anthropogenic ignitions, anthropogenic suppression, and peat fires due to anthropogenic ignitions are turned off), there is no peak around 1870 for burned area.

(5) About the continuous decrease of burned area in the case without direct anthropogenic influence
A reference may provide a clue for this. "Kloster et al. (2010) shows that downward trend of global fire carbon emissions from 1900 to 1960 was mainly caused by reduced fuels as a consequence of land use and wood harvesting (i.e., the indirect anthropogenic influence on fires)". We have mentioned this work when we discussed the historical variation of fire carbon emissions in the Para. 4 of Sect. 6. However, Kloster et al. (2010) haven't analyzed the impact of this indirect anthropogenic influence on global burned area.

The main purpose of the present study is to introduce the improved global fire simulation when an extended version of our earlier fire model is used in CESM1. The extended version is our earlier fire model with more realistic representation of direct anthropogenic influence on fires. Sect.6 only discussed whether the direct anthropogenic influence is the major driver for historical variation of global burned area and fire carbon emissions. We'll prepare a paper to discuss the reasons for global fire history in more detail and add regional analysis. Thanks for your suggestion.

2. p16772: The description of the numerical experiment is a bit vague. The authors say they followed a suggestion of the CESM user guide and used the "I20TRCN component set rather than component sets which use full coupled CESM1". What are component sets ? Are they the modules of the CESM ? (land vs atmosphere vs ocean etc) ? The authors should simply state which model they used (CLM4 with carbon/nitrogen biogeochemistry if I am not mistaken). The authors say that

I20TRCN is a “20th century simulation (1850-2004) . . .. So is I20TRCN a “component set” or a simulation ? I don’t believe they should mention this I20TRCN at all. They also say that the atmospheric observations are obtained by cycling 25 yr (1948-1972). Do they mean that they cycled 1948-1972 to force the model from 1850 to 1947 ?

Reply: In Para. 1 of Sect. 4, we have revised the description about CESM1:

> **“**CESM1 consists of five geophysical models: atmosphere model (atm), ocean model (ocn), land model (lnd), land-ice model (glc), and sea-ice model (ice), plus a coupler that coordinates the five models and passes information between them (Vertenstein et al. 2012). Each model may have "active," "data," "dead," or "stub" component version allowing for a variety of "plug and play" combinations. The active (dynamical) components (CAM5, POP2, CLM4, CISM, or CICE4) are generally fully prognostic, and state-of-the-art climate prediction and analysis tools. Because the active models are expensive to run, data models that cycle input data are included for model parameterization development, testing, and spin-up. The dead components generate scientifically invalid data and exist only to support technical system testing. Stub components exist only to satisfy interface requirements when the component is not needed for the model configuration.”

In the last para. of Sect.4., we have added “**A component set** is a particular mix of six components: one component version ("active," "data," "dead," or "stub") from each model (atm, ocn, lnd, glc, and ice) plus the coupler, along with component-specific simulation setting.”, and revised the description about **I20TRCN** as “It uses the “data” version for atmosphere model, “active” version for land model (CLM4 with carbon/nitrogen biogeochemistry), and “stub” version for ocean, sea-ice, and land-ice models, plus the coupler; and provides a 1850-2004 transient simulation forced by 1850-2004 transient time-varying CO_2 concentration, nitrogen and aerosol deposition, and land use and land cover changes (wood harvesting included). The initial data of I20TRCN component set is the equilibrium state of CLM4 with carbon/nitrogen biogeochemistry forced by cycling 25-yr (1948−1972) of Qian et al. (2006)’s atmospheric observations, with CO_2 concentration, nitrogen and aerosol deposition and land cover in 1850.” Also, we have clarified the **atmospheric forcing** used in I20TRCN as “In I20TRCN, the “data” atmosphere model is obtained by cycling 25-yr (1948–1972) Qian et al. (2006) and bias-corrected relative humidity for 1850–1949 followed by the full time series for the years 1950–2004…”. That is, 25-yr (1948–1972) atmosphere observations are cycled for 4 times to force CLM4 from 1850 to 1949 (100 years).

3. p 16775 last paragraph: "Mod-new and Mod-CTEM underestimate . . .. Fire season (summer)". This sentence is awkward, rephrase

Reply: we have changed this sentences to **"**For boreal forests in Eastern Siberia and North America, Figs. 9, 10 in Bonan et al. (2011) showed that CLM4 driven by atmospheric observations obviously underestimated the latent heat flux, especially during the fire season (i.e. summer). The latent heat flux is directly proportional to the water flux output from the land to the atmosphere, so the underestimation of latent heat flux suggests overestimation of the water retained by the land. The wet simulation bias of land surface in CLM4 partly explains the underestimation of burned area for Mod-new and Mod-CTEM in the boreal forests which is the moisture-limited fire regime region"

4. Just a comment about the underestimation in Africa. Beside socio-economic factors, there might also be cultural ones. In Gabon for instance, people burn grasslands every year mainly because they've always done it. I believe there might be some interesting sociological/anthropological study to be done here but that is obviously not the purpose of this paper.

Reply: Thanks for the suggestion. The information is used when we reply to the first comment of reviewer 1.