Response to reviewer 2

General evaluation: The paper presents a framework for evaluation of fire models, particularly in the context of analyzing the potential impacts of climate changes in altering fire regimes. The paper is well written and generally well documented (few minor issues are commented later). It does provide a good overview of existing global DGVM including a fire component, with tables summarizing the main assumptions and drivers. I think the paper would benefit from extending the specific benchmark test that will be used to compare the model performance, as the current version only gives insights on potential approaches. I believe the authors should be more specific on what are the planning to do to actually compare model performance, which indicators will they use, on which period and area (including target resolution).

Response: We thank the reviewer for these positive and constructive comments. Our intention in this paper was to describe the current status of global fire modelling, including the current status of model evaluation, and indicate the challenges which need to be addressed in a modelling intercomparison, rather than providing a detailed protocol for FIREMIP itself and we attempt to make our aims clearer in the revised manuscript. We agree that our description of the different steps involved in developing the FIREMIP evaluation scheme was perhaps not spelled out clearly enough, and that there are details the could have been added. As similar types of requests were also asked by reviewer #1 we have therefore:

modified the paragraph describing the Kelley et al. benchmarking system to indicate that this is what we will use initially in FIREMIP. The data sets involved, which provide benchmarks for multiple aspects of both vegetation and fire, were already listed in the paragraph. However we have now also indicated that the comparisons will be made at a 0.5° *resolution – which is the common grid of the data sets in the Kelley et al system – but that spatial resolution does not have any impact on the metrics. If the comparisons were made at the model resolution, the metric score would be identical with two significant figures. However, given that the models all have different resolutions, and that the benchmark data sets are already at 0.5 ° resolution, the most convenient approach is to refer everything to a common framework. The end of this paragraph now reads (lines 393-396):*

"The Kelley et al. (2013) scheme will be used for model evaluation and benchmarking in FireMIP. It has been shown that spatial resolution has no significant impact on the metric scores for any of the targets (Harrison and Kelley, unpublished data); nevertheless, model outputs will be interpolated to the 0.5° common grid of the data sets for convenience."

We have modified the second paragraph to make it clearer that our intention in FireMIP is to expand the Kelley et al benchmarking system, and in particular to take the opportunity to include new data sets as they appear. This paragraph now reads (lines 397-412):

"The Kelley et al. (2013) scheme does not address key aspects of the coupled vegetation-fire system including the amount of above-ground biomass and/or carbon, fuel load, soil moisture, fuel moisture, the number of fire starts, fire intensity, the amount of biomass consumed in individual fires, and fire-related emissions. Global datasets describing some of these properties are now available, and will be included in the FireMIP benchmarking scheme. These data sets include above-ground biomass both derived from vegetation optical depth (Liu et al., 2015) and ICESAT-GLAS LiDAR data (Saatchi et al., 2011), the European Space Agency Climate Change Initiative Soil Moisture product (Dorigo et al., 2010), the Global Fire Assimilation System biomass burning fuel consumption product, fire radiative power, and biomass-burning emissions (Kaiser et al., 2012), and fuel consumption (van Leeuwen et al., 2014). The goal is to provide a sufficient and robust benchmarking scheme for evaluation of fire while ensuring that other aspects of the vegetation model can also be evaluated, and to this end new data sets will be incorporated into the FireMIP benchmarking scheme as they become available during the project."

Finally, we have tried to make it clearer that the development of other aspects of the FireMIP evaluation and benchmarking exercise are research questions that will need to be addressed during the project. We are convinced that new approaches are required to deal with uncertainties caused by the fact that *different, and apparently equally robust, data sets show substantially different patterns. But the techniques for propagating these uncertainties into metrics are in their infancy. We also believe that process evaluation and palaeo evaluation are necessary steps in model evaluation, but have not been used in any systematic way for fire modelling. Therefore, we have also added the following text before the final three paragraphs in this section (lines 413-418 (see response to reviewer no. 1)) as follows:*

"The FireMIP benchmarking system will represent a substantial step forward in model evaluation. Nevertheless there are a number of issues that will need to be addressed as the project develops, specifically how to deal with the existence of multiple data sets for the same variable, how to exploit process understanding in model evaluation, and how to ensure that models which are tuned for modern conditions can respond to large changes in forcing. The answers to these questions remain unclear, but here we provide insights into the nature of the problem and suggest some potential ways forward."

Minor comments:

Response: All suggestions have been taken into account in the revised version.

Line 65: a comma is missed after seasonality: "frequency, intensity, seasonality etc".

Response: corrected

Line 71: What a significant fraction means? Please quantify

Response: The current assumption in carbon budgeting is that all of the carbon lost in fires will be taken up as vegetation regrowth within a decade. This does not hold under a changing climate or if people use the post-fire opportunity to convert the area to e.g. crops. Thus, it is difficult to know how to quantify this accurately and indeed no one has done this. We agree that this perhaps deserves fuller treatment as so we have modified the sentence to read (lines 71-74):

"This is equivalent to ca 25% of those from fossil fuel combustion (Ciais et al., 2013; Boden et al., 2013), although in the absence of climate and/or land use change, nearly all of these emissions are taken up during vegetation regrowth after fire."

Lines 70- 86: No reference to N_2O emissions from fires is made. Why?

Response: We had lumped N_2O in "many other atmospheric constituents" (Line 75), but agree that it is sufficiently important to mention explicitly and we have now done so.

Line 92: Johnston et al., 2012 is missed in the references section.

Response: Added

Lines 108-109: "Fire risk is not quantitatively related to area burnt, fuel consumption, or fire emissions". It is not clear what you mean here. Most fire risk systems are assessed with fire statistics (Chuvieco et al. 2014; Chuvieco et al. 2010; Padilla and Vega-Garcia 2011; Paz et al. 2011), and some are associated to burned area and fuel consumption (Consume, for instance: see Pettinari and Chuvieco 2015).

Response: We agree that this text is not clear. It is true that fire risk (or so-called fire danger) systems are developed and assessed based on fire statistics, and often using burned area or fuel consumption as a target. Our point here is that these systems are calibrated under current conditions, but have been used to assess what might happen in the future. There are many papers that do this, and we have focused on the Moritz et al. (2012) paper as an example because this formed the basis of the statements about future fire *in the last IPCC assessment. But statistical fire risk/danger models cannot account for a number of factors that could influence future fire regimes, such as the impact of CO2 fertilization on in situ productivity or changes in vegetation type. They also cannot account for the possibility that future climates may not have analogues in the modern day, e.g. because of changes in temperature seasonality. Our point here was to make the case for process-based fire modelling if the goal is to project potential changes in fire regimes in the future. We have rewritten this paragraph to make the argument more explicit as follows (lines 110-120) :*

"Statistical models have been used to examine the potential trajectory of changes in fire during the 21st century (e.g. Moritz et al., 2012; Settele et al., 2014). Such models essentially assess the possibility of fire occurring given climate conditions and fuel availability (fire risk or fire danger) based on modern day relationships between climate, fuel and some aspect of the fire regime such as burnt area. However, changes in fire risk/danger will not necessarily be closely coupled to changes in fire regime in the future given the direct impacts of CO2 on water-use efficiency, productivity, vegetation density and ultimately vegetation distribution. This limits the utility of statistically-based models for the investigation of feedbacks to climate through fire-driven changes of land-surface properties, vegetation structure or atmospheric composition – feedbacks which have the potential to exacerbate or ameliorate the effects of future climate change on ecosystems, as well as influence the security and well-being of people."

Line 143: "outcrops can act as natural barriers to fire fronts". Natural barriers is duplicated from previous line.

Response: We have deleted the duplicated words.

Line 147: "and highest in areas of intermediate water availability", assuming a dry period exists.

Response: This statement is globally true regardless of seasonality of precipitation. The highest burnt areas are found in areas with sufficient rainfall to produce good vegetation cover and hence fuel to burn but where rainfall is not so high as to ensure that the fuel is permanently wet. Most such areas do have a marked seasonal cycle of precipitation but this is not necessary to the argument although the timing and length of the dry season affects the quantitative level of what is meant by "intermediate" water availability. We feel that the term "water availability" is somewhat confusing here, and so we have modified the text to make it clearer what we mean as follows:

"Burnt area tends to be lowest in very wet or very dry environments, and highest where the water balance is intermediate between these two states."

Line 159: "purpose, for example for forest clearance, agricultural waste burning or fire", please add pasture management, which is the most common factor in many areas of the world.

Response: We have added this.

Line 170: "gross domestic product, GDP, that are linked to population density) results from the co-variance of population density with vegetation production and moisture". This sentence may be tinged, as those relations depend on other factors, such as the importance of agricultural sector in regional economy. For a global analysis, you may be interested to read Chuvieco and Justice 2010. Bowman et al. 2011 has also an interesting analysis of human-fire relations.

Response: There have indeed been many papers analysing the relationships between postulated drivers of fire and burnt area, both at a regional scale and at a global scale, but many have not taken into account the co-variance between different explanatory variables. Here, our statement is based on the comprehensive global analysis by Bistinas et al. which used GLM to identify the independent relationships between burnt area and specific driver, and showed that spatial and temporal trends in burnt area could *be predicted with a simple model based primarily on vegetation productivity and moisture. Bistinas et al. also showed that the relationship between burnt area and population density was significant but negative, and that the unimodal relationships with population and GDP can be reproduced by this simple model. We have added a further reference to the Bistinas et al. (2014) paper to clarify that this is the source for our assertion. We have now included a reference to Bowman et al., 2011 in the introduction.*

Line 198: the JSBACH acronym is not defined.

We have added the full name of JSBACH, which is the "Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg" at line 198. Thank you for pointing out that this was not defined; the full name is so rarely used that we suspect only the originator of the acronym remembers what it stands for.

Lines 368: When citing alternative sources of model assessment, you do not include reference to the GFED dataset (Giglio et al. 2013), which is widely used for fire –emissions analysis. A reference to the synthesis analysis of Mouillot et al (2014) may be relevant in this point. Please, also note that soil moisture is not equivalent to fuel moisture. The CCI Soil moisture product does not really estimate vegetation wetness.

Response: We agree that GFED is by far the most widely used "reference" data set when analysing fire emissions under present day conditions. However, the emissions are calculated using the CASA vegetation model and are therefore not an independent reference data set. As we have stressed in describing the benchmarking system, we have chosen data sets that are not dependent on a model driven by the same drivers as the models we are seeking to test (line 387-389). Thus while we will use GFED burnt area, we will not use GFED emissions for model evaluation. We do not claim that soil moisture is equivalent to fuel moisture (we now separate them in the text to make this clearer). In describing the benchmarking system we have made the point that it is important to evaluate the simulation of both vegetation properties and fire regimes – it may well be that the failure to capture fire regimes is related to under or over production of woody vegetation, for e.g., which is directly related to the simulated soil moisture. However, clearly these points are worth stressing and we have modified the text describing the alternative data sets to emphasise these two points, as follows (lines 406-412):

"The selection of new data sets is partly opportunistic, but reflects the need both to evaluate all aspects of the coupled vegetation-fire system and the importance of using data sets that are derived independently of any vegetation model that uses the same driving variables as the coupled vegetation-fire models being benchmarked. The goal is to provide a sufficient and robust benchmarking scheme for evaluation of fire while ensuring that other aspects of the vegetation model can also be evaluated, and to this end new data sets will be incorporated into the FireMIP benchmarking scheme as they become available during the project."

Lines 383-388: When analyzing different global burned area products, you may refer to the intercomparison analysis published by Chang and Song 2009 or the most recent validation effort by Padilla et al. 2015. Line 381 and 814: Please not that ESA MERIS burned area product is officially named Fire_cci (see Chuvieco et al. 2016, which also includes an assessment of fire emissions derived from this product). The temporal resolution of the Fire_cci product is Burn Date for the pixel product at aprox. 300 m resolution. However, the burned area is accumulated in 15 day periods for a gridded version of product, which has 0.5 d resolution.

Response: We included the Padilla reference and changed the name of the ESA MERIS product to Fire_cci as indicated. We now indicate that the spatial resolution of MERIS is ± 300m in table 3 and changed the name of the product there, as well as the temporal resolution.

Line 459: please, include the updated reference to Alonso-Canas and Chuvieco 2015.

Response: Changed

Line 819. In fig. 1 you may add to Fuel load, Fuel continuity, which is related to fragmentation.

Response: We have tried to indicate the role of fuel continuity with the arrow going from fragmentation to fuel load. We now mention this in the header of figure 1.