We thank Reviewer 1 for his/her comments. In this reply we address the suggestions for revisions of the manuscript point by point.

General comments:

> *Wang et al. use simple stand-alone primary productivity models to attempt to improve understanding of the processes responsible for the widely ranging DGVM model response to both CO2fertilization and the combined effects of climate change. Their idea is that the use of these simple stand-alone models will allow for a reduction in the uncertainty associated with the differing responses of the more complex models. A similar approach has seen great success with high complexity climate models, so it seems to be have good possibility of success. I generally find their analysis to be sound, however I have a few points I would like to see addressed.*
>
> *First, the language used is often overly vague. For example, the authors often refer to 'the data' or 'forest NPP data ', but it can be difficult to determine if they are referring to 'observed data' or 'model output data' or even regressions on observed data rather than their simple LUE/WUE models. I believe I generally understood which dataset was referred to in each instance, but I had to re-read sections repeatedly due to vague language. For clarity the authors should amend their language to be more specific.*

We have systematically revised the text for clarity.

> *Second, two of the conclusions of the study relate to the likely incorrect CO2 response of the LUE and WUE models, which the authors attribute to an ability to capture an increase in runoff (for WUE) and an increase in vegetation cover (for LUE). Both of these proposed reasons for the models' likely-incorrect response are plausible, but they are just educated guesses that do not come directly from the model results. I don't feel then that these conclusions merit such a prominent role in the study's conclusion (they are mentioned in the abstract). I would recommend leaving these interpretations in the main text, or at least provide further support for the interpretation.*

As discussed at the end of this document, the inferences about runoff changes and vegetation cover changes are valid – they are not just educated guesses – but (we suspect) they are not clearly enough justified in the original text. In the revision we have provided a more explicit statement of our reasoning. We have also removed these conclusions from the Abstract as they are really extensions rather than primary results of our analysis.

> *Lastly, I find the discussion on ratios of ANPP to total NPP for grassland to be unconvincing. Given the very large range from the Hui and Jackson (2006) paper, it is actually surprising that the value from the WUE model falls outside of it. I am not sure then if these model results can be used in sparse ecosystems as the authors contend. I would like to see further evidence supporting the validity of the models in grasslands.*

Unfortunately the data available for this purpose are very limited. But the reviewer's point is inescapable – we cannot make a very strong case for the extension of the models into non-forest vegetation in the absence of more data. We have added a caveat to this effect.

Detailed review:

> *Throughout the manuscript, the authors occasionally neglect to list the units of variables. Please ensure units are consistently labeled.*

In the revision, we have carefully checked this point and supplied units wherever they were missing.

> *p.4289 l.5-10: The assumed NPP/GPP ratio is never given that I could find. Please state the value used outright. This is especially confusing as the authors later derive a NPP/GPP ratio for their LUE and WUE models (section 4.1), but it is unclear if these values are used elsewhere.*

When developing the semi-empirical models, we assumed that NPP is an approximately constant fraction of GPP. Therefore, when we fit our WUE and LUE models to NPP data, the NPP/GPP ratio is implicit in the estimated slope. We did not pre-define it, so there is no 'value used' to report.

Later, however, we also fit the models to GPP data (from the global gridded data set of Beer *et al.*) By comparing the slope of the GPP relationship (from the global gridded data) in each model to the slope of the NPP relationship (from the forest site data), we can infer a value for the ratio NPP/GPP.

We agree that this two-step approach could bring some confusion. We have therefore provided a clearer wording in the "model" section of our revised manuscript.

> *l. 12: I think the authors should also examine/discuss the implications of their use of a managed forest as opposed to an old-growth forest.*

This is a very good suggestion and we included more discussion of this issue in the revised manuscript.

> *p. 4290 l. 5: Were the grasslands also managed or pastured? What are the implications of this dataset's values if they were?*

The grasslands were managed (as grazing land), but heavily disturbed sites were avoided. This was already stated in the text.

> *p. 4295 l. 15: Remove one instance of 'performed separate'. There are also other instances of typos and grammar problems that should be carefully checked for in other parts of the MS.*

We have carefully checked for typos and corrected the grammar throughout the revised version.

> *p. 4295 l. 20-25: Please better describe what was done here, it is difficult to understand at present how this performs the independent check that the authors describe. The Beer et al. (2009) dataset is also based upon a WUE model so perhaps it is not truly an independent check.*

In the paper of Beer et al. (2010), they "estimate terrestrial GPP and its spatial details by diagnostic models". Site-level GPP data derived from eddy covariance flux data "was used to calibrate five highly diverse diagnostic models, which relate GPP to meteorology, vegetation type, or remote sensing indices at daily, monthly, or annual time scales. Two of these approaches are machine learning techniques: a model tree ensemble (MTE), and an artificial neural network (ANN)". "MTE is either driven by fAPAR only (MTE1) or by both fAPAR and climate data(MTE2)". "The Koppen-Geiger cross Biome (KGB) approach is a look-up table of mean GPP per ecoregion. GPP of the whole river catchment areas is estimated by the water use efficiency approach (WUE), which combines recently derived global WUE fields with the long-term averaged evapotranspiration at the watershed scale. This is an important constraint at the global scale, but the spatial resolution is too coarse to use the WUE approach for estimating the spatial distribution of GPP. The light-use efficiency approach (LUE) was applied by combining in situ Bayesian calibration with an uncertainty propagation per vegetation and climate class."

The data we were used as an independent check are the median valuers of annual GPP ($gC/m^2/a$) from the spatially explicit approaches (MTE1, MET2, ANN, LUE, and KGB) at a resolution of 0.5 degree. It is neither based on a WUE model, nor just based on a LUE model. It can be considered as an independent source because the underlying observations are eddy covariance flux measurements, not forest mensuration measurements as we have used. But the reviewer is right to mention that the data set of Beer et al. (2010) is to some extent a modelled product. Our text already made this point and we have further amplified it in the revision.

> *p.4297 l.1-22. This whole section is very opaque on what model/regression/observation data the authors are using. Please re-write this section to enhance clarity.*

We have re-written this section to clarify which data are used for what purpose.

> *p. 4297 l. 11-19: A maximal slope of 21% strikes me as a large value. In the discussion, a lot is made of the hypothesized influence of runoff and vegetation cover, but little to the nutrients. Even though the influence of the nutrients is likely not heavily important, more discussion of their influence should be given as the influence is not insignificant. While I agree that nutrient availability is not the primary control of forest NPP, I don't see much support for the authors' contention (p. 4300 l.20) that 'the data provide no support ... that nutrient availability is the primary control on forest NPP'(again, specify which data!). I think the authors need to back this statement up with further evidence.*

We appreciate this comment and have included more discussion about the influence of nutrients, and also provided a more detailed justification for why our results contradict the conclusions of Huston and Wolverton.

> *p 4297 l.19-21: This difference for the oldest age class is one reason I would like to see more discussion on the implications of comparing the model results to managed as opposed to old-growth forests. I am also puzzled as why later on (p.4299 l. 25) the authors compare their NPP/GPP ratio against forest stands > 100 yrs old, but give no information about the proportion of their modelled forests that are of that age.*

We agree that the difference for the oldest age class is an important finding. We have amplified our discussion of this point. We also now provide information on the proportion of forests in the oldest age class.

> *p. 4299 l.18: The Zhang et al. (2009) value noted is the global average, not really comparable to a China only value. The authors should get a China specific value or at least give the range for China that are shown in the Zhang et al. (2009) paper.*

In the revision we have provided the range of NPP/GPP ratios for China from Zhang et al. (2009).

> *p. 4300 l.19: Besides my earlier objection to this statement around the nutrient avail- ability, I also don't understand the statement regarding NPP in the tropics than in temperate regions. This whole paragraph needs to be rewritten as it does not presently make sense on which data are apparently contradicting Huston and Wolverton (2009).*

The revision includes a rewriting of this paragraph.

> *p. 4300 l.18-30: The range in Hui and Jackson (2006) is so large that it is surprising to be outside of it. Also since they are fractions, to not be of 'similar magnitude'would be exceptional! I find this to be weak proof that the models are application to sparse vegetation types. More evidence should be provided to justify the models use in sparse ecosystems.*

This is indeed a limitation of our study. We cannot fully address this issue here, however, because of the paucity of data on the productivity of sparse ecosystems in China. We have added a caveat to this effect.

> *p. 4303 l. 12-16: Where in the models does runoff appear? I can understand why the authors hypothesize about the influence of both runoff and vegetation cover, but I can't see how it is backed up by their model results. As a result I don't think the discussion surrounding runoff or vegetation cover should feature so prominently in their conclusions (e.g. abstract l. 11 -13).*

We have delete the runoff/vegetation cover implication from our Abstract. However, it is a valid point, which arises as an indirect consequence of the models. We gather that the logic of this point was not completely clear in our original text. Therefore, we have expanded on it in the revision.