Author's response

We highly appreciate the constructive comments and literature advice to further improve the paper. We considered all suggestions as prescribed below. The different colors indicate: (grey:) editors comment and (black:) author's response.

As the authors correctly mention, regional-scale uncertainty analyses of land surface models are rare, but the various points they raise in the results sections have already been addressed in various studies. For example, the sensitivity of net carbon fluxes to Q10 value, or the various environmental response functions that affect turnover rates, has been addressed in detail by Davidson and Janssens (2006), Todd-Brown et al. (2013, 2014), Exbrayat et al. (2014a,b). Furthermore, the transferrability of PFT parameters has been studied by Kuppel et al. (2014).

It is correct that other studies like these investigated aspects that we investigated too in this paper. However, these studies did not estimate the uncertainty of modelled region scale NEE as we did herein, taking into account the uncertainty of forcings data, initial states and parameter values. We will address this comment and point out the differences including references in the revised version of the manuscript.

p3 l25: see also Xia et al. (2012) / Exbrayat et al. (2014a);

We will add the suggested literature references.

p3 l27: please consider adding a reference to the Global Carbon Budget / TRENDY intercomparison project here

We will add the suggested reference.

p5 l2: please indicate time series length here

Time series lengths are summarized in Table 1. We will indicate this in this sentence: "Eddy covariance (EC) data in the Rur catchment were measured four C3-crop sites for the evaluation periods summarized in Table 1."

p5 l10-12: although this has been reported in a previous paper, can you please develop a bit more on how parameters were optimized for these sites?

In section 2.3.1 we explained how parameters were optimized and refer to the previous paper for further details. As the focus of this paper is not on parameter estimation but on model uncertainty (which we will now better highlight in the abstract), we would not extent the outline of the parameter estimation methods. If there is some particular information you are missing, we can add it.

p6 l13: can you indicate the respective fraction of each PFT here instead of l25

The sentence "The percentage PFT coverage of the vegetated land in Rur catchment was ~34% C3-crops, ~32% grassland, ~17% broadleaf deciduous forest and ~14% coniferous forest." was moved upward as suggested.

p7 l10: how representative of mean climate were the three years used for spin-up?

According to http://www.dwd.de/DE/klimaumwelt/klimaatlas/klimaatlas_node.html , the annual average temperature in the years 2008 and 2009 was ~ 0.5-1.0 °C higher the long term average (1961-1990), the mean annual temperate in 2010 was ~ 0.5-1.0 °C lower. Mean precipitation amounts and freezing days where representative for the long term average. Before conducting this study, we extensively studied the effect of the forcing data used for the model spin-up. For example, we tested using a longer time series (1998-2004) of global climate data. However, we found that using more recent, regional forcing data during the spin-up resulted in carbon- and energy fluxes in better agreement with the observations. This will be clarified in the modified version of the manuscript.

p7 l11: what is an "exit" spin-up? does that simply mean that the spin-up was 1203 years repeated meteorological drivers for 2008-2010

Principally, yes. We reformulated this sentence. "Exit spin-up" is a technical term for CLM. Before conducting the actual model runs with the restart files from the model spin-up, an intermediate run (the exit spin-up") is

necessary with specific parameter settings. This is e.g. because CLM uses an accelerated model spin-up mode. We will clarify in the paper: "The model states obtained after the 1200-year spin-up were then used as input for a second three years "exit spin-up" also using the meteorological data for the years 2008-2010. The "exit spin-up" in CLM is necessary for technical reasons and switches the CLM settings from the (accelerated) spin-up mode to the "normal" mode in terms of the calculated carbon-nitrogen cycling."

p7 l24: what value of the convergence diagnostic did you use?

We used a threshold of 1.2 to declare convergence as suggested e.g. in Vrugt (2015).

p8 l9: please add reference for Fontainebleau's site

We can add a reference. Is there a particular reference you can suggest?

p8 l26: compared

This will be corrected

p10 l1: please add full name and references for the two sites. Are there particular references you can suggest?

We will add it.

p11 l19: are 15 years enough for this perturbed spin-up?

We think so. We also tested longer perturbed spin-up periods up to 30 years and found there were no significant differences of the ensemble spread. We found that after about 8 years, the ensemble spread reaches some kind of maximum and than does not increase considerably further.

p12 l8: I am not clear what "deterministic initial states" mean here

We added "(non-perturbed)" in the sentence (p. ll.): CLM-EnsP: ensemble model runs for 1 Dec 2012 – 30 Nov 2013 with deterministic (non-perturbed) initial states and non-perturbed input data from COSMO-DE (Sect.2.2)." to clarify that we mean the default, non perturbed initial states obtained with the restart files from the 1203 years non perturbed spin-up.

p13 l1: why are results for ENSp outside the range from ENSpai? from the experiment description it feels that ENSpai's uncertainty should envelop that of ENSp (and ENSpa)

In case of ENS_{PAI} perturbed atmospheric forcing AND perturbed initial states were applied in contrast to ENS_P. For ENS_P, the non-perturbed forcing and initial state data sets were used. The large discrepancy between ENS_{PAI} and ENS_P is one of the major findings of this paper and highlights the overly strong sensitivity of the modeled carbon pools and fluxes to the climate forcings.

We will extend the discussion on that.

p13 l21: there is a problem with figure numbering.

Which problem? We checked the sentence and it seems correct/ as intended.

p13 l23: daytime NEE is not GPP, please replace/revise in the rest of the paragraph.

This of course is correct. We revised the paragraph (p. , ll.): "The delay of the plant emergence indicated by these LAI courses is related to the strong underestimation of daytime NEE in spring 2013 (Figure 3, Figure 4). This underestimation of NEE can mainly be attributed to an underestimation of GPP, which is too low in spring because the simulated plant onset was about 2 weeks later than observed. For CLM-Ref and for most of the CLM-EnsPAI ensemble members, leaf onset started in May. For those model realizations, the underestimation of daytime NEE in spring was highest. In contrast, for CLM-EnsP and a small proportion of CLM-EnsPAI, leaf onset already started in March. For those cases, the underestimation of daytime NEE in spring was notably lower. This elucidates the close link of modelled NEE and LAI and highlights that errors in the timing of leaf onset can lead to substantial errors in simulated NEE. The evaluation of simulated LAI showed that modelled

and observed LAI are closer for simulations with estimated CLM parameters than for simulations with default parameters and that in case of C3-crop, simulated leaf onset was better represented."

p 15 l17: the discussion is very short in regards of the amount of results that are described.

We agree and will extend the discussion, including the references you suggested.

Figures 3, 4 and 5: please correct the legend

The legend has been corrected.