Authors responses to referee #2

General comments:

The authors are dealing with an extremely interesting and relevant question, namely how historical land-use induced land-cover changes (LULCC) modified large scale evapotranspiration (ET). There is still high uncertainty regarding a reliable reproduction of worldwide ET, thus studies to figure out the global water and heat balance still show high discrepancies. Historical, large scale LULCC had undoubtedly an impact on global and regional ET, but appropriate, historic land-cover maps in combination with reliable ET estimates have been lacking so far. It is therefore commendable that the authors are dealing with this challenge and attempt to overcome the existing uncertainties. From the reviewer's opinion however, the author's main problem was to tackle with these uncertainties, and they tried hard to identify any results which are utilizable among this uncertainty. It is therefore highly respectable that the authors always point out the uncertainties while the amount of "reliable" results is quite low.

The article is probably welcome by the global modelling community, and therefore it might be useful to be published in HESS. There is also no doubt that the methods and tools presented in the article are novel and of high scientific quality. However, not only that scientists working on smaller scales (both field experimentalists and modellers) can hardly profit from the findings in the article, but also they have to realize that uncertainties which are unacceptable for the regional (or smaller) scale are seen as an advance on the global scale. For example, discrepancies in LAI as depicted in Figure 3 would lead to such an enormous variation of simulated ET in a SVAT model that the regional water balance would differ over magnitudes.

Some of the findings of this global analysis sound quite trivial since they present facts that are well known even without such an extensive study (see specific comments for some examples). To conclude, this is just a small (and hardly reliable) step towards a better estimate of ET on a global scale. It is questionable if historical ET changes can be reproduced giving the enormous uncertainties reported in this article. The paper is in general well written and structured, but when the reader works through the article, he is getting more and more confused with the different datasets, models and methodologies applied for the study. Table 1 doesn’t help too much to keep an overview. Therefore, it is recommended that the authors should think about a special section in which the different working steps and required data, models, methods are explained step-by-step, maybe supported by a flow chart. Several parts of the article read like a technical document, and the reader has a hard job to keep attention until the end.

We thank referee #2 for the in-depth review of our study.

The major concern of the reviewer regards the uncertainties reported in this global-scale study, compared to arguably smaller uncertainties at site-scale. The main motivation of this study is indeed to contrast -in a global and historical context- the past LULCC-induced ET changes obtained previously from free-running coupled AGCM-LSM models (within the LUCID project), with the independent empirical estimations (observations-based) carried out here. Obviously, this is a difficult quest, and as is raised by the reviewer, it is logical that uncertainties on global ET changes since pre-industrial times are larger than uncertainties on site or regional budgets.

Despite large remaining uncertainties, our attempt to use data for reconstructing ET changes is a significant new step forward towards understanding model biases and reducing structural uncertainty. In fact, better data-products would be directly helpful to falsify model results and improve them, so there is a direct co-benefit from accurate site-level data (used for budgeting or process studies) into the global and historical upscaling that is the purpose of our study. This was clarified in the revised manuscript.

We use our data-driven ET reconstructions consistently with the different LULCC vegetation maps used by each LUCID model, instead of adopting a unique reference pair of land-cover maps for the preindustrial times and present-day –an option that would have widely reduced the range of the results, but given a too optimistic view of current between model uncertainties.

We think that a given Earth-system variable (in our case, the large-scale ET change driven by LULCC) is more robust when derived from multiple (an ensemble of) cases, i.e., when the spread from a known source of uncertainty is quantified. Previous studies, such as those referred in the manuscript (e.g., Gordon et al. 2005, see complete list of refs. at the end), haven't managed to reduce the uncertainties associated to the LULCC reconstruction. They just didn't include it.

We then respectfully disagree with the referee's main opinion, namely that the results described here do not contribute significantly to improve the estimation of past ET change because of the large uncertainties associated with the reconstruction method. As mentioned, the reported uncertainties represent in themselves a key result of the manuscript. Our study also provides seasonal and spatial details of ET changes that haven't been reported in any other earlier work. Despite the different sources of ET global “data” and the different land-cover maps used (a structural model uncertainty in LUCID), we found some systematic responses to LULCC across all the models that emerge out of these uncertainties, notably across the temperate regions of the northern hemisphere. This emerging behavior has not been shown with clarity in previous modelling experiments nor in observation-based studies. We offer the reviewer to clarify this point in the conclusion of the revised manuscript.

We use a large set of data and a methodology that involves various steps. We agree with referees 1 and 2 regarding the need of a general picture of the method used (and an easy-to-follow description of all the steps and datasets used). We will therefore improve the method section and include a flowchart figure synthesizing the methodology and the input / output datasets.

* All the comments that follow indicated pages/lines numbers which did not correspond to the version of the manuscript published online by HESSD. Also, some comments did not match with the information described in the paper published. Perhaps the reviewer had access and used an earlier version (unpublished) that did not include a number of typesetting corrections. We did our best anyway to address all the comments from the reviewers, which are very helpful.

Specific comments:

Page 5, lines 19/20: what are “large-scale observations”? Do you mean observation-based, interpolated data products? Which data that was used for this article are real observations (i.e., station-based data?)

Yes, we refer to observation-based gridded data-products. “Large-scale observations” is not a proper term, so we will use “observation-based, interpolated data products” in the first instance and “data-products” in the rest of the revised text. We also point out in the revised method section that none of the “data-product” represents direct observation, and that there is obviously no global gridded direct observation (so we used some of the best available global knowledge derived from observations).

Page 5, line 29: what is the difference between “climate” and “meteorology” within the context of this sentence?

The analysis carried by Jung et al. in their ET computation, as well as the one performed in our study, use both monthly varying data and on long-term monthly means of a number of environmental variables. We refer to “meteorology” and “climate” in respect to the former and the later, respectively. Meteorology isn't indeed a proper term for monthly means. We have reformulated the terminology in the revised manuscript.

Page 7, lines 5-8: It sounds a bit trivial that “the water-limited and energy-limited regions, which roughly correspond to global dry and wet areas respectively, are captured by all three datasets”. I think this should at least be expected by a global dataset, otherwise it would be useless.

That sentence was redundant and will be removed. However, given the very large differences in interannual variability shown by various ET products, we could not assume initially that all the products reproduce coherently the spatial distribution of the two regimes (water-limited and energy-limited), and we therefore estimate its evaluation necessary.

Figure 3: There is such a high spread of the mean LAI (Figure 3a) that any SVAT or hydrological model considering LAI in its computations would give ET and water balance data within a wide range of magnitude. The LAI differences given in Figure 3b show even a worse picture. Do you think that a relative change of 0.2 [m2 /m2] is reliable given the extreme, overall uncertainty?

LAI is a key variable controlling ET, given that it defines the effective foliage area and therefore the ET partitioning (transpiration/interception/evaporation). There is indeed a large spread in the simulated present-day LAI and the LAI changes (1992-1870). Such discrepancies between the models represent then a main source of uncertainties in modelled ET. The simulated changes in LAI also differ widely from the values of LAI changes reconstructed from the MODIS data-product. Despite the uncertainties, the diagnosed LAI show a clear signal of change (decrease) from 1870 to 1992. Hence, the main message from Fig. 3 is that one cannot expect “a priori” a concordance between the modeled ET response to LULCC if models do not simulate first LAI properly. This kind of result motivates the analysis that is further developed in this study, based on empirical (data-driven) models of ET instead of LSMs. This has been clarified in the revised text.

Figure 4: This is a multi-product (i.e., the average) from 18 reconstructed ET climatologies. Is it possible to see the range of the results from the different realizations in an additional graph? What means a change of say 40 mm/year in regions with today’s mean annual ET of 800 mm given the uncertainty from the individual realizations?

We agree. We will upgrade this figure in order to include the spread between the multiple estimations. Yes, changes in ET on annual scales are moderate compared to net ET (~ 5-10 %) and compared to the spread of the single estimations (of the same order actually).

Indeed, an important conclusion is that the changes of ET on annual time-scale is not a good metric to evaluated the impacts of LULCC, which masks seasonal-scale changes of larger amplitude and different sign, as is shown in figures 5 and 6. This has been highlighted in the revised text.

Figure 7: The map showing the distribution of crops, grassland and forest has a very coarse spatial resolution (1.0 degree). This is probably immutable since no better dataset is available, but this leads to the problem that information about vegetation distribution gets quite fuzzy and thus less reliable. Each 1.0 x 1.0◦ grid cell includes certain heterogeneity, but it is scaled out through the coarse resolution. The number of pixels for which the given criterion is valid (> 75% coverage of the corresponding class) is getting quite small then. The ET values from this comparatively small number of pixels are then averaged for the regions above 20◦ N. Another problem is that averaging of ET from the three vegetation classes has been done without taking individual climate zones into consideration. Thus, this averaging includes strong heterogeneity from the individual driving forces of ET. How reliable is this data?

Fig. 7 shows a result from a very simple exercise allowing the comparison of the mean ET across various products. As you mentioned, the comparison across the groups of vegetation is not really suitable since the selected regions (pixels) have different climate regimes. As we explain in page 2063, this is the reason that motivates the analysis of the sensitivity depicted in Fig 8. The results we finally obtain with those two approaches (mean values and sensitivities comparisons) are quite coherent.

Figure 7: Could you please let us know the number of pixels for the three vegetation classes? What mean the red dots / the red surfaces above the boreal zone in Northern America and Eurasia? Is it right that they show grassland? Is this possible for 70◦ N (the tundra ecotone)?

Number of pixels used in Fig 7: 529 (crops), 1051 (grass), 1092 (forest). Note that we fixed a rendering problem in this figure (the former version shows an artificial poor resolution).

The PFT classification adopted in most LSMs, as well as in the MODIS land-cover product used here, synthesizes some biomes (such as Tundra) into larger groups of PFTs (e.g. tundra corresponds to grass PFTs). In the MODIS case, the boreal regions you mention are actually flagged as grass.

Page 16, equation (4) and line 25: where is Delta Fv in equation (4)?

The subscript v represents a land cover type, it represents any of the five types used in this study. This has been clarified in the text.

Page 17, lines 1-2 and Figure 8a: Why did you select results from the MPI-based reconstructions, and why just for July (why, for example not for June-August?)? Why this selective analysis? Please indicate the number of pixels analysed.

Fig. 8a shows one example from one ET product that illustrates how the results depicted in Fig. 8b (that includes all seasons and products) are derived. We chose the month of July because ET is maximized in the boreal summer in the regions assessed, and a clearer signal can therefore be observed. The MPI case wasn't chosen for any particular reason.

Page 17, line 4: why just (at least) four times larger?

The ideal situation for this analysis would be to have a large number of pixels with simple transitions (e.g. a pixel totally deforested to allocate crops). This is not the case given the resolution used, so we finally applied this rule. The factor of 4 value was defined by trials and errors. A larger value would have reduced the ambiguity of the transition assessed but, at the same time, would have reduced substantially the numbers of pixels for the analysis. This has been clarified in the revised text.

Figure 8a: Please explain the meaning of Delta LC in the figure caption

Which Delta LC? See the above-mentioned comment (*).

Page 17, lines 9 and 13, as well as Figure 8b: I have a problem reading the terms “sensitivity” or “sensitivity analysis” under such an uncertainty. The solution could be to combine an uncertainty analysis with the “sensitivity analysis”, but that would probably show that the uncertainty clearly exceeds “sensitivity”.

We use the term sensitivity because we present an objective analysis that compares, across the various products assessed, normalized ET responses to specific LU transitions. It is true that there are uncertainties. This is expected given that the different rates of ET between two types of land cover is not unique, but depends on a number of other factors (climate, LAI, etc). However, some preferred signals clearly emerge from this uncertainty, allowing to quantify inherent (to each product) ET responses (sensitivity) to the selected transitions.

Page 17, lines 5-8: A “dominant signal” of Delta ET for land-use change has been identified. It is however not surprising that ET decreases when forests are replaced by grassland or that ET increases when grassland is converted into intensive cropland (that has been proven by numerous experimental studies and simulations). An extensive study would therefore not be necessary for this conclusion.

We have doubts regarding the suggested “wide consensus” of these ET signals between grasslands, croplands and forests (see e.g. Teuling et al., 2010). In particular, croplands can have higher or equivalent ET rates than forests given management and alleviation of limitations.

Anyway, we are not trying to prove here that forest or cropland must have higher ET rates than grasses in the real world. This is not a field study, but an attempt to estimate large-scale changes in ET driven by LULCC. In that sense, we estimate reasonable to compare the ET response to specific land-use transitions in order to identify key differences between the "data" sources used (ET data-products in this case). We actually found that the main difference between them regards the crops ET rates.

Page 21, lines 15/16 and Figure A1: The finding that "evergreen trees (in EA and NA) show moderate LAI, comparable to that of short vegetation" conflicts with numerous (field) studies documenting LAI of coniferous trees and non-forest vegetation. The clearly higher LAI of coniferous trees with respect to deciduous trees and short vegetation is clearly visible in their higher ET, for example through increased interception evaporation. Why shows Figure A1c such a strong seasonal behaviour of LAI for evergreen trees in EA and NA? It is also surprising that LAI of deciduous trees in EA and NA during summer shows such a high value with respect to LAI of evergreen trees. Any references which are able to prove this behaviour? It is also questionable why summertime LAI of evergreen forest in NA should be (clearly) higher than that in EA.

We were also surprised with some LAI patterns shown in Fig A1. These are the values that result from the MODIS LAI fields projected on the corresponding land-cover classes and region. The seasonal pattern of evergreen trees in temperate/boreal areas are actually dominated by those of needleleaf forest (coniferous). Broadleaf forest doesn't show such seasonal cycle (not shown). This behaviour is however consistent with what the ref. paper of the used dataset (Yuan et al. 2011) and other studies based on MODIS have reported (e.g. Tian et al. 2004, Garrigues et al., 2008, De Kauwe et al., 2011, Fang et al. 2013). It is possibly related to the differences between LAI of a conifer tree and LAI of a pixel covered by conifers in boreal regions, where canopy closure is not often reached. Uncertainties in the land-cover classification and associated errors on the attributed LAI values have also been reported (Fang et al. 2013). The high LAI of deciduous trees at pixel level in summer is also consistent with earlier findings (see e.g. the above-mentioned papers of Fang et al. or Tian et al.).

In-depth assessments of remote-sensed LAI products and the identification of systematic discrepancies with field measurements are tasks of large interest, notably regarding the use of these datasets for model evaluations. However, this type of evaluation is beyond the scope of our study.

Technical corrections:

Page 2, line 30: Pongratz and Caldeira

Page 8, line 26: minimize

Table 2: footnote "b" is applied twice

Figure A1, b and c: change "EU" in "EA"

All changes made

References cited in this response

De Kauwe et al (2011). An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest. Remote Sensing of Environment, 115, 767–780.

Fang, H., et al. (2013), Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties, J. Geophys. Res. Biogeosci., 118, 529–548, doi:10.1002/jgrg.20051.

Garrigues et al.: Validation and intercomparison of global Leaf Area Index products derived from remote sensing data, J. Geophys. Res., 113, G02028, doi:10.1029/2007JG000635, 2008.

Gordon et al: Human modification of global water vapor flows from the land surface, P. Natl. Acad. Sci. USA, 102, 7612--7617, doi:10.1073/pnas.0500208102, 2005.

Jung et al.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010.

Teuling et al.: Contrasting response of European forest and grassland energy exchange to heatwaves, Nat. Geosci., 3, 722–727, doi:10.1038/ngeo950, 2010.

Tian, Y., et al. (2004), Comparison of seasonal and spatial variations of leaf area index and fraction of absorbed photosynthetically active radiation from Moderate Resolution Imaging Spectroradiometer (MODIS) and Common Land Model, J. Geophys. Res., 109, D01103, doi:10.1029/2003JD003777.

Yuan et al.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sens. Environ., 115, 1171–1187, doi:10.1016/j.rse.2011.01.001, 2011.