Referee's report, HESS, manuscript no. hess-2013-583

Paper title: Historical land-use induced evapotranspiration changes estimated from present-day observations and reconstructed land-cover maps

Authors: J. P. Boisier, N. de Noblet-Ducoudré and P. Ciais

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.

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?)

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

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.

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 [m^2/m^2] is reliable given the extreme, overall uncertainty?

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?

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?

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)?

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

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.

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

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

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”.

Page 17, lines 5-8: A “dominant signal” of ΔET 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.

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.

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”