Interactive comment on 'Deriving a per-field land use and land cover map in an agricultural mosaic catchment'

Bumsuk Seo[1,2], Christina Bogner[2], Patrick Poppenborg[3], Emily Martin[4], Mathias Hoffmeister[5,6], Mansig Jun[7], Thomas Koellner[3], Björn Reineking[1,8], Christopher L. Shope[9], and John Tenhunen[10]

[1]*Biogeographical Modelling, BayCEER, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany*
[2]*Ecological Modelling, BayCEER, University of Bayreuth, Dr.-Hans-Frisch-Str. 1–3, 95448 Bayreuth, Germany*
[3]*Professorship of Ecological Services, BayCEER, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany*
[4]*Department of Animal Ecology and Tropical Biology, Biocentre, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*
[5]*Department of Organismic Biology, University Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria*
[6]*Institute of Sensory Ecology, Department of Biology, Heinrich-Heine-University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany*
[7]*Regional Institute for Gangwon, Junangro 5-GA, Chuncheon, Republic of Korea*
[8]*UR EMGR Écosystèmes Montagnards, Irstea, 2 rue de la Papeterie-BP 76, F-38402, St-Martin-d'Hères, France*
[9]*U.S. Geological Survey, 2329 W. Orton Circle, Salt Lake City, UT 84119-2047, USA*
[10]*Plant Ecology Department, University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany*

July 8, 2014

*Corresponding author: Bumsuk Seo (Bumsuk.Seo@uni-bayreuth.de)

1 Comments by anonymous Referee #1

The manuscript by Seo et al. describes the creation of a vector based land use and land cover (LULC) map for an agricultural landscape with diverse management activities. Based on repeated annual census dates, 2009-2011, the authors create several LULC maps with varying degrees of thematic resolution, and compare these with a global land cover dataset from MODIS, using the IGBP classification. Because the vector land cover map is based on ground-based observations, it far exceeds the thematic accuracy of other available products, and provides an important perspective for mosaic agriculture at high spatial resolution. However, the authors argue in their Introduction that they aim to overcome limitations posed by global land cover datasets, but the proposed solution presented for the Haean catchment is obviously not practical at larger scales. What is a very interesting outcome of their study is the comparison between the survey approach and the MODIS IGBP data. This comparison provides an opportunity to evaluate bias in the MODIS product for land cover and land cover change.

1.1 Please define the meaning of 'Per field' - this term occurs only in the title and needs to be defined for readers

We will add a new paragraph in the introduction. We censused the complete area and recorded field-by-field land use/land cover type. Hence the unit entity of our dataset is a single land parcel and we call it 'per-field'. We borrowed the term from Conrad et al. (2010) in which they defined 'per-field' data as a dataset on the basis of actual agricultural fields.

1.2 Section 2.4.1 - please define what types of 'quality issues' you encountered.

Section 2.4.1 (current) We digitised the field records and converted patches and linear elements into polygons and polylines, respectively. The base map served as a background and complemented the field records. Additionally, we stored LULC type and other descriptive information in an attribute table. Any quality issue, if occurred, was marked in the Quality Assurance (QA) column of the corresponding year. The digitised data was cross-checked with GCPs recorded later during the census period and corrected when needed.

Section 2.4.1 (revised) We digitised the field records into polygons and polylines with LULC type labels. The base map served as a background information to complement the field records. In addition to LULC classes, we stored other descriptive information in an attribute table. In the corresponding columns, Quality Assurance (QA) was recorded as: '?' (Questionable), '*' (Unknown) and '/' (Not valid). For instance, a question mark was assigned if we could not identify the crop reliably. Gap-filled data was also marked by a question mark. A forward slash represents invalid data. For further information, the reader is referred to the legend of the dataset at Pangaea repository.

1.3 P 284, Line 11

> *Here you make the suggestion that MODIS may be more accurate than your surveyed land cover dataset. I find this very unlikely given the census intensity and ground-based approach used to develop your land cover map. In the context of uncertainty and accuracy, please can you add a section on how you can quantify the land cover land use survey map and provide something like a kappa statistic.*

Comparison to MODIS MCD12Q1 product

It is crucial to quantitatively evaluate land use/land cover products. Therefore, we compared our data with the MODIS Land Cover Type product as an attempt to investigate the accuracy of the satellite-borne land cover dataset.

Since the MODIS product includes primary and secondary land cover types in raster form, we also developed primary and secondary cover type raster layers out of the vector data. At each pixel, we assigned the land cover class labels based on the exact area fractions of the classes. Primary and secondary cover types were determined by the highest and the second highest fractions. However we did not assign a label if second highest fraction was less than 10%. The rasterisation was done in `R` (R Core Team, 2013) using the geometry engine `GEOS` (GEOS Development Team, 2014) and the package `rgeos` (Bivand and Rundel, 2014). Note that the secondary land cover of the MODIS product is not based on estimated fractions but on measures of classification confidence (Friedl et al., 2002). Note that MCD12Q1 is based on a decision tree algorithm (C4.5) with boosting (**friedl˙modis˙2010**; Quinlan, 1992).

Eventually we had two comparable land cover type layers; primary types shown in the first rows and secondary types in the second rows of Fig. 1 and Figure 2 for HaeanCover and MCD12Q1, respectively.

First we measured correspondence of the two primary land cover types. We derived a confusion matrix and similarity metrics for year 2009 (Table 1) and year 2010 (Table 2). Note that we skipped year 2011 due to the incomplete ground observation.

We followed Cohen (1960) for calculating Cohen's non-weighted κ as defined in Eq. 1. Therein p_o connotes the proportion of pixels in which the two products agreed and p_c is the proportion of pixels for which agreement is expected by random chance.

$$\kappa = \frac{p_o - p_c}{1 - p_c} \quad (1)$$

The mean κ for two years was 0.3, which indicates a fair but not substantial agreement. Due to the unequal distributions over land cover types, however, the κ values may have been biased towards the major classes. We did not explicitly discuss the imbalance in this paper, however, we shall be aware of the issue since forest dominates the landscape.

Secondly we compared the relative frequency of the IGBP classes. We calculated proportions of primary and secondary types omitting pixels with 'NA' values. The two separately calculated proportions were averaged, thereby we obtained mean relative prevalences.

In Figure 3, we have (a) and (b) from the original vector map and the rasterised observation shown in (c) and (d). The MODIS data is visualised in (e) and (f) in the same figure.

The rasterised HaeanCover corresponds better to the vector map than the MODIS product. Moreover, urban and other minor types are completely missing in (e) and (f) even if we consider the secondary cover type layer.

Table 1: Confusion matrix between MODIS (columns) and HaeanCover (rows) (2009). Cohen's κ is 0.32

	4	5	10	12	14
4	99	36	4	4	26
5	4	1	0	0	0
6	0	0	0	2	0
10	0	0	1	0	0
12	0	0	27	87	7
13	0	0	0	1	0
100	50	17	0	0	0

Table 2: Confusion matrix between MODIS (columns) and HaeanCover (rows) (2010). Cohen's κ is 0.28

	4	5	10	12	14
4	79	72	6	4	10
5	2	3	0	0	0
6	0	0	0	2	0
10	0	0	9	3	0
12	0	0	32	70	6
13	0	0	0	1	0
100	39	28	0	0	0

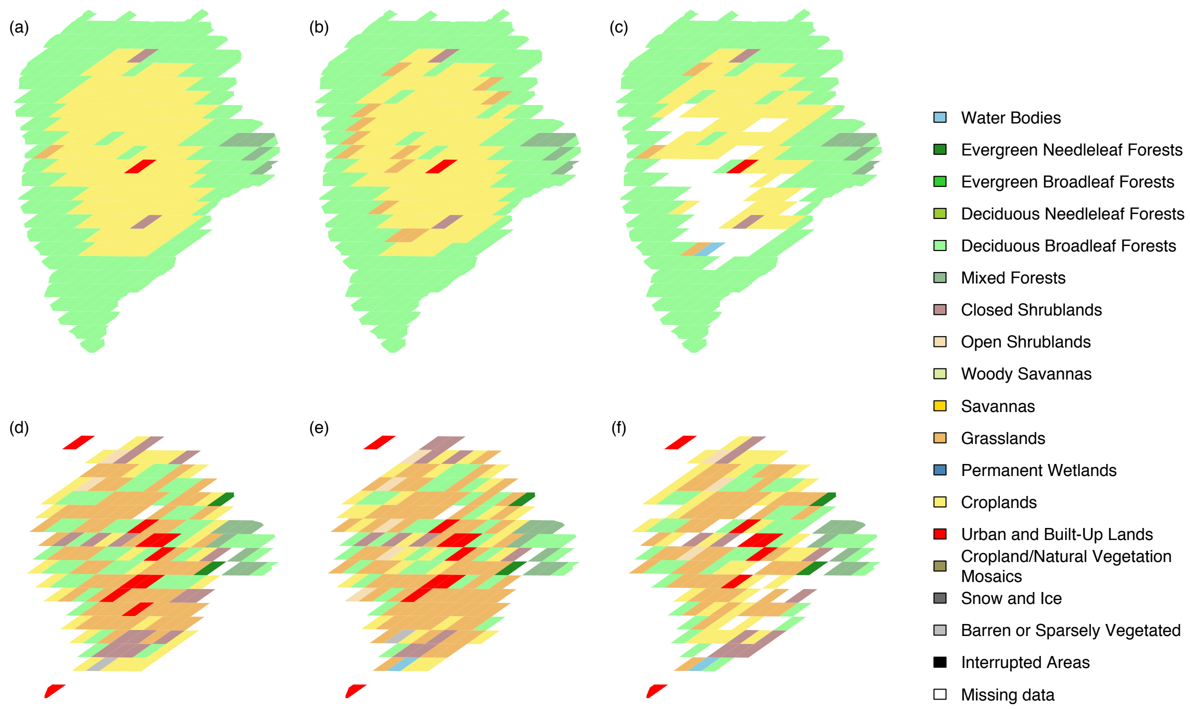

Figure 1: The raster images of primary and secondary land cover types are displayed. Based on the vector map, (a), (b) and (c) represent primary types determined by the 'most prevalent' rule regarding fractions. (d), (e) and (f) represent secondary types by the second dominant cover types.

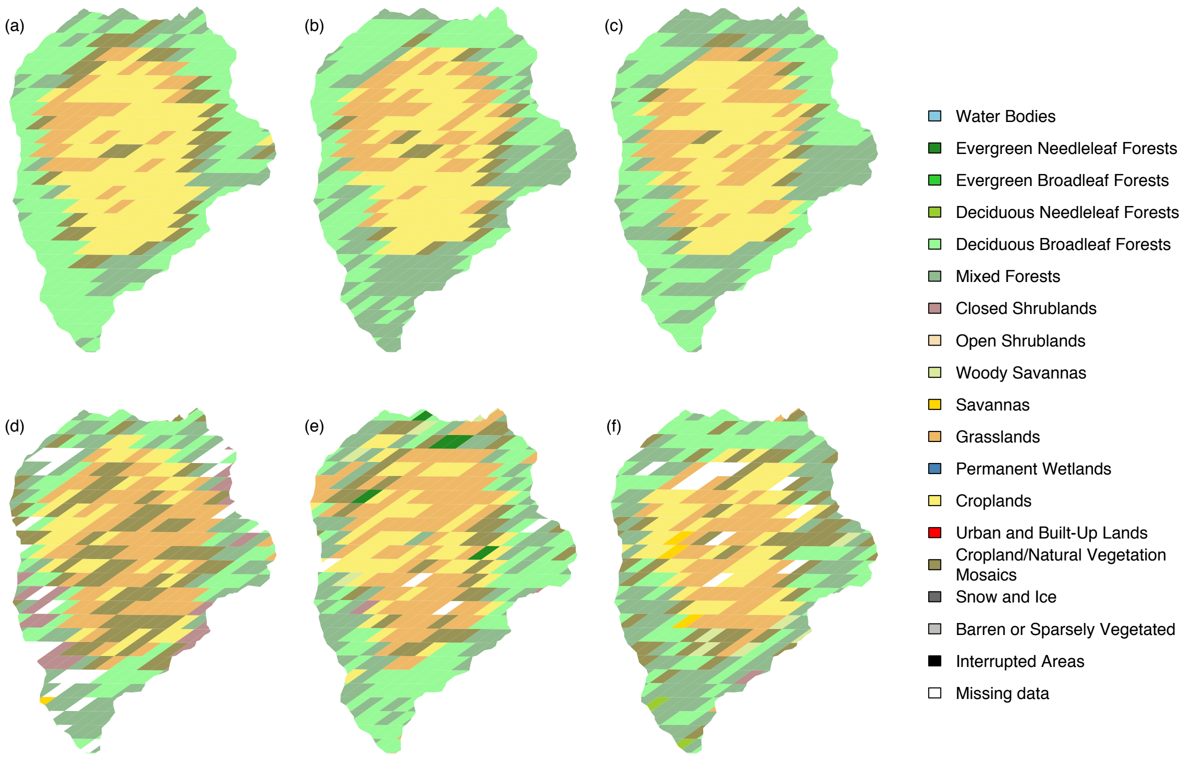

Figure 2: Primary and secondary land cover types of the MODIS Land Cover Type product (MCD12Q1).

Figure 3: Primary and secondary land cover types for the vector HaeanCover data in (a) and (b), rasterised HaeanCover in (c) and (d), and MCD12Q1 in (e) and (f).

1.4 Section 4.0

Please provide more information on the data archived on Pangea – do you provide separate files for each year, are the transition areas included? Providing the data as a raster would be useful for users to make their own transition maps.

Section 4.0 (current) The data set and its description are available at the Pangaea repository (Supplement) under the Creative Commons Attribution-NonCommercial 3.0 Unported license. The data contains LULC observations in ESRI shape file format and ancillary information for the study area. The LULC type, Quality Assurance (QA), Management, Double cropping and Mixed use information are provided on an annual basis. The definition of classes and reclassification table are given separately in plain text format. For each polygon, multiple entries in the LULC type column are available in case the polygon exhibited mixed land uses spatially or temporally.

Section 2.4.1 (revised) The data set and its description are available at the Pangaea repository (Supplement) under the Creative Commons Attribution-NonCommercial 3.0 Unported license. The data contains LULC observations and ancillary information in a single ESRI polygon shape file. The LULC type, Quality Assurance (QA), Management, Double cropping and Mixed use information are provided on an annual basis. The definition of classes and reclassification table are given separately in a legend table. Each polygon attains three years information in separated columns (e.g., LULC2009, LULC2010 and LULC2011). Note that multiple entries in a LULC type column occur in case the polygon exhibited mixed land uses spatially or temporally.

1.5 Section 5.0

Please add a discussion on how your methodology can be used to advance MODIS, GLOBCOVER, GLC2000, land cover type datasets, where you point out in the Introduction their shortcomings.

First of all, we can use our observation to validate the existing global databases. It can be applied for many different classification systems and can be transformed to any raster grid. It could particularly useful to researchers investigating a remotely sensed land cover in complex terrains and/or in agricultural mosaics.

Also we would hope that forth-coming updates of global land cover databases become high-resolution. For natural vegetations, global high-resolution databases are becoming available (e.g., Hansen et al. (2013)). Our vector form data can be useful in developing/validating a high-resolution dataset. Furthermore, we look forward to seeing new global databases providing additional crop types such as paddy crops. There have been ongoing efforts to extend MODIS land cover data bases (Biggs et al., 2006; Potgieter et al., 2007; Wardlow, Egbert, and Kastens, 2007; Wardlow and Egbert, 2008; Pittman et al., 2010; He and Bo, 2011; Gumma, Thenkabail, and Nelson, 2011). Our dataset can be also useful to the development as it includes detailed crop type information.

References

Biggs, T W et al. (2006). "Irrigated area mapping in heterogeneous landscapes with MODIS time series, ground truth and census data, Krishna Basin, India". In: *International Journal Of Remote Sensing* 27.19, pp. 4245–4266. DOI: 10.1080/01431160600851801. URL: http://www.informaworld.com/openurl?genre=article&doi=10.1080/01431160600851801&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3.

Bivand, Roger and Colin Rundel (2014). *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. R package version 0.3-4. URL: http://CRAN.R-project.org/package=rgeos.

Cohen, Jacob (1960). "A Coefficient of Agreement for Nominal Scales". English. In: *Educational and Psychological Measurement* 20.1, pp. 37–46. DOI: 10.1177/001316446002000104. URL: http://epm.sagepub.com/cgi/doi/10.1177/001316446002000104.

Conrad, Christopher et al. (2010). "Per-Field Irrigated Crop Classification in Arid Central Asia Using SPOT and ASTER Data". In: *Remote Sensing* 2.4, pp. 1035–1056.

Friedl, M A et al. (2002). "Global land cover mapping from MODIS: algorithms and early results". In: *Remote Sensing of ...* 83.1-2, pp. 287–302.

GEOS Development Team (2014). *GEOS - Geometry Engine, Open Source*. Open Source Geospatial Foundation. URL: http://trac.osgeo.org/geos/.

Gumma, Murali Krishna, Prasad S Thenkabail, and Andrew Nelson (2011). "Mapping Irrigated Areas Using MODIS 250 Meter Time-Series Data: A Study on Krishna River Basin (India)". English. In: *Water* 3.1, pp. 113–131. DOI: 10.3390/w3010113. URL: http://www.mdpi.com/2073-4441/3/1/113/.

Hansen, M C et al. (2013). "High-Resolution Global Maps of 21st-Century Forest Cover Change". In: *Science* 342.6160, pp. 850–853.

He, Yaqian and Yanchen Bo (2011). "A consistency analysis of MODIS MCD12Q1 and MERIS Globcover land cover datasets over China". English. In: *Geoinformatics, 2011 19th International Conference on*. IEEE, pp. 1–6. ISBN: 978-1-61284-849-5. DOI: 10.1109/GeoInformatics.2011.5980667. URL: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5980667.

Pittman, Kyle et al. (2010). "Estimating Global Cropland Extent with Multi-year MODIS Data". In: *Remote Sensing* 2.7, pp. 1844–1863.

Potgieter, A B et al. (2007). "Estimating crop area using seasonal time series of Enhanced Vegetation Index from MODIS satellite imagery". In: *Crop and Pasture Science* 58.4, pp. 316–325.

Quinlan, Ross J. (1992). *C4.5: Programs for Machine Learning*. 1st ed. Morgan Kaufmann. ISBN: 9781558602380. URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/1558602380.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Wardlow, Egbert, and Kastens (2007). "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains". English. In: *Remote Sensing of Environment* 108.3, pp. 21–21. DOI: 10.1016/j.rse.2006.11.021. URL: http://www.sciencedirect.com/science/article/pii/S0034425706004949.

Wardlow, Brian D and Stephen L Egbert (2008). "Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains".

English. In: *Remote Sensing of Environment* 112.3, pp. 1096–1116. DOI: 10.1016/j.rse.2007.07.019. URL: http://www.sciencedirect.com/science/article/pii/S0034425707003458.