Highlighting continued uncertainty in global land cover maps for the user community

2011 Environ. Res. Lett. 6 044005

(http://iopscience.iop.org/1748-9326/6/4/044005)

Environ. Res. Lett. 6 (2011) 044005 (6pp) doi:10.1088/1748-9326/6/4/044005

Highlighting continued uncertainty in global land cover maps for the user community

Steffen Fritz[1], Linda See[1], Ian McCallum[1], Christian Schill[1,2], Michael Obersteiner[1], Marijn van der Velde[1], Hannes Boettcher[1], Petr Havlík[1] and Frédéric Achard[3]

[1] International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria
[2] Department of Remote Sensing and Landscape Information Systems (FeLis), University of Freiburg, Tennenbacherstrasse 4, D-79106 Freiburg, Germany
[3] Institute for Environment and Sustainability (IES), Joint Research Centre of the European Commission, Via Enrico Fermi, I-21027 Ispra, Italy

E-mail: fritz@iiasa.ac.at

Received 16 August 2011
Accepted for publication 3 October 2011
Published 21 October 2011
Online at stacks.iop.org/ERL/6/044005

Abstract
In the last 10 years a number of new global datasets have been created and new, more sophisticated algorithms have been designed to classify land cover. GlobCover and MODIS v.5 are the most recent global land cover products available, where GlobCover (300 m) has the finest spatial resolution of other comparable products such as MODIS v.5 (500 m) and GLC-2000 (1 km). This letter shows that the thematic accuracy in the cropland domain has decreased when comparing these two latest products. This disagreement is also evident spatially when examining maps of cropland and forest disagreement between GLC-2000, MODIS and GlobCover. The analysis highlights the continued uncertainty surrounding these products, with a combined forest and cropland disagreement of 893 Mha (GlobCover versus MODIS v.5). This letter suggests that data sharing efforts and the provision of more *in situ* data for training, calibration and validation are very important conditions for improving future global land cover products.

Keywords: land cover validation, uncertainty, cropland, forest cover

S Online supplementary data available from stacks.iop.org/ERL/6/044005/mmedia

1. Introduction

Global maps of land cover derived from satellite-based earth observations have existed for almost two decades and represent one of the most important sources of baseline terrestrial information for a wide variety of applications, e.g. as inputs to global models of land use and land use change (Foley *et al* 2005, Verburg *et al* 2010), climate modelling (Pielke 2005), assessment of available land for biofuels (Cai *et al* 2011), food security (Liu *et al* 2008) and as the basis for crop distribution modelling (You *et al* 2009). Applications in other areas such as biodiversity and population are presented in Giri (2005), who acknowledges that we are still a considerable way from producing global land cover products that are of a high enough quality for many applications.

One avenue for research has involved the comparison of global land cover datasets, either against one another or with higher resolution regional products. Many of the studies found similar results, i.e. good overall agreement but disagreement in either the individual land classes or in the spatial distribution of the land cover (McCallum *et al* 2006, Giri *et al* 2005, Fritz and See 2005, See and Fritz 2006, Neumann *et al* 2007, Herold *et al* 2008, Fritz *et al* 2010, Seebach *et al* 2011). This general trend continues to be the case when comparing the

Table 1. Classes in each of the land cover maps that fall in the cropland and forest domains.

Domain	Classes in GLC-2000	Classes in GlobCover	Classes in MODIS v.5 (IGBP)
Cropland	Cultivated and managed areas	Post-flooding or irrigated croplands Rainfed croplands	Croplands
	Mosaic: cropland/tree cover/natural vegetation	Mosaic cropland/grass or shrub or forest	Cropland/natural vegetation mosaic
	Mosaic: cropland/shrub or grass cover	Mosaic grass or shrub or forest/cropland	
Forest	Tree cover, broadleaved, evergreen	Closed–open broadleaved evergreen/semi-deciduous forest	Evergreen needleleaf forests
	Tree cover, broadleaved, deciduous, closed	Closed broadleaved deciduous forest	Evergreen broadleaf forests
	Tree cover, broadleaved, deciduous, open	Open broadleaved deciduous forest	Deciduous needleleaf forests
	Tree cover, needleleaved, evergreen Tree cover, needleleaved, deciduous	Closed needleleaved evergreen forest	Deciduous broadleaf forests
	Tree cover, mixed leaf type	Open needleleaved deciduous or evergreen forest Closed–open mixed broadleaved–needleleaved forest	Mixed forests
	Mosaic: tree cover/other natural vegetation	Mosaic forest or shrub/grassland	
	Tree cover, burnt mosaic: cropland/tree cover/natural vegetation	Mosaic grassland/forest or shrub	
	Tree cover, regularly flooded, fresh tree cover, regularly flooded, saline		

overall accuracy of the most recent global land cover products, i.e. the GLC-2000 (Bartholomé and Belward 2005), MODIS version 5 (Friedl *et al* 2010) and GlobCover (Bicheron *et al* 2008, Bontemps *et al* 2011). The overall accuracy of these maps is reported as 68.5% (±5%) for GLC-2000 (Mayaux *et al* 2006), 74.8% (±1.3%) for MODIS and 67.1% for GlobCover 2005 (Bicheron *et al* 2008), which shows similar results when taking the 95% confidence bands into account. However, this paper will show that the overall spatial disagreement in both the forest and particularly the cropland domain continues to be very high. The purpose of this paper is therefore to make the user community aware of the continued uncertainties in these products, which could potentially impact the outcomes of any assessment or modelling exercises undertaken (see e.g. Quaife *et al* 2008, Feddema *et al* 2005, Havlík *et al* 2011). The paper concludes with a discussion of what is needed to improve the accuracy of land cover information.

2. Data and methods

2.1. Global land cover and agricultural datasets

The GLC-2000 was developed by the JRC to provide baseline information for the year 2000 and has the coarsest resolution at 1 km (Bartholomé and Belward 2005). It was first developed regionally using experts in the field and then integrated into a single global product. In contrast, Boston University's MODIS land cover product was developed using a top-down approach where a classification algorithm was applied to create a global product at a resolution of 500 m (MODIS v.5) for the year 2005 (Friedl *et al* 2010). At 300 m, GlobCover is the finest resolution product available for the year 2005–2006 (Bicheron *et al* 2008). This new product is intended to update and complement other existing comparable global products but the higher spatial resolution was also expected to provide improvements in thematic accuracy as the overall number of mixed classes found in a pixel decreased. A supervised and unsupervised classification algorithm was used to classify pixels into similar spectral and temporal classes. An automated labelling procedure using the Land Cover Classification System (Di Gregorio and Jansen 2000) and a global reference dataset (including the GLC-2000) was then used to create the final product. Compared to the GLC-2000, the more automated nature of the classification algorithms used to produce MODIS and GlobCover means that they can be easily repeated and produce updated products on a more regular basis. Further details of these datasets can be found in Fritz *et al* (2010). The focus of the disagreement between land cover products in this paper is in the cropland and forest classes. Table 1 provides the legend classes that fall in the cropland and forest domains in each land cover class while the full legend definitions for these classes are given in supplementary tables 1 and 2 (available at stacks.iop.org/ERL/6/044005/mmedia).

2.2. Aggregation to a common spatial resolution

In order to compare the three main global land cover products, the difference in their spatial resolutions was first reconciled. In the Plate-Carrée projection with a Geographic Lat/Lon representation, the spatial resolution of the GLC-2000 is $1/112° \times 1/112°$ or approximately 1 km × 1 km at the equator; MODIS v.5 is $1/240° \times 1/240°$; while GlobCover is $1/360° \times 1/360°$ or approximately 300 m × 300 m at the equator. A common grid of $0.125° \times 0.125°$ was chosen in which all land cover products could then be aggregated. This equates to an aggregation of 14 pixels for GLC-2000, 30 pixels for MODIS v.5 and 45 pixels for GlobCover. The percentage of forest

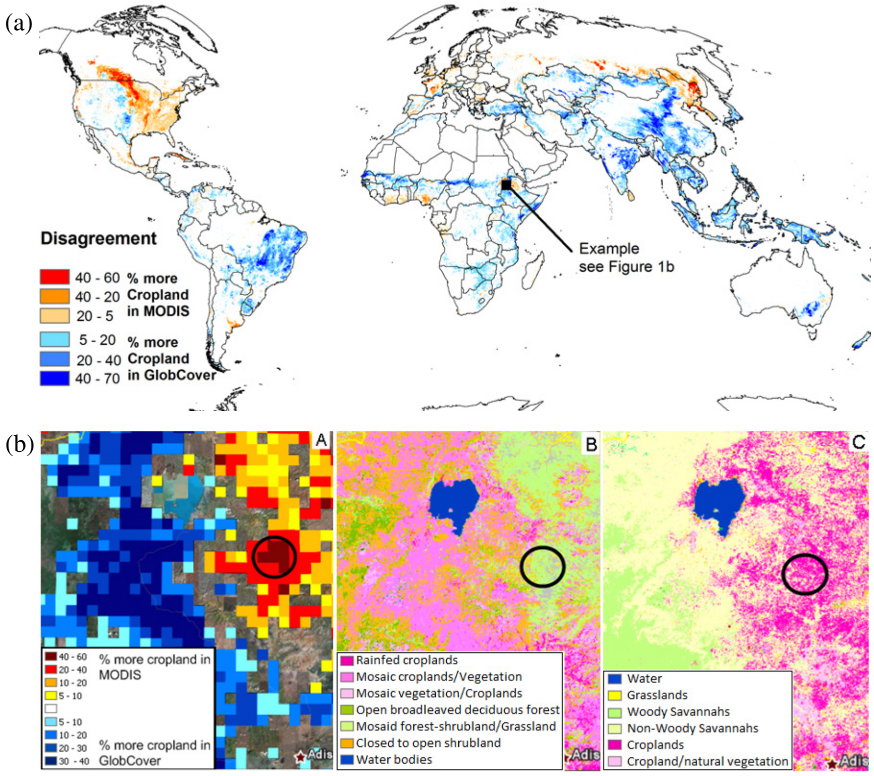

Figure 1. (a) Global disagreement between MODIS v.5 and GlobCover 2005 in the cropland domain. (b) Disagreement for an area in Ethiopia: (A) GlobCover and MODIS cropland disagreement overlaid on Google Earth; (B) GlobCover land cover map; and (C) MODIS land cover map. In (A), blue tones show pixels of $0.0125° \times 0.0125°$ size where GlobCover records cultivated/managed and MODIS records woody savannah/non-woody savannah and red tones show the opposite. For example, (see the black circle) in (A), disagreement is recorded (with a higher proportion of cropland in MODIS), whereas GlobCover (B) shows a mosaic of forest/shrubland/grassland and MODIS shows croplands (C).

and cropland of each aggregated grid cell was then determined using the minimum and maximum percentages from the class definitions (as provided in supplementary tables 1 and 2, available at stacks.iop.org/ERL/6/044005/mmedia). For example, if 98 out of 196 pixels for the GLC-2000 were of forest type (where 80%–100% in that pixel would be forest according to the definition of this land cover type), then the new aggregated grid cell would contain 50% forest for the maximum forest class, i.e. 100% divided by 2, and 40% forest for the minimum forest class, i.e. 80% divided by 2. A similar approach was used in Fritz *et al* (2010) to aggregate the pixels.

2.3. Creation of cropland and forest disagreement maps

The disagreement in cropland and forest cover between a given pair of land cover maps is derived using a concept termed in this paper as the Minimum Measurable Disagreement (MMD). This is a modified version of the approach used in Fritz and See (2008), which was originally based on the concepts proposed in Ahlqvist (2005) but is applied to the aggregated grid cell. Each aggregated grid cell in each of the land cover datasets will have a minimum and maximum cropland or forest cover. To compare a pair of land cover datasets and calculate the disagreement at each pixel, the range of cropland/forest cover is compared by examining the amount of definitional overlap. The calculation of disagreement is illustrated in supplementary figure 1 (available at stacks.iop.org/ERL/6/044005/mmedia). If there is any overlap in the definitions, then the disagreement or MMD is 0. Where there is no overlap, the MMD is calculated. For example, if the aggregated pixel for GlobCover has 0–40% cropland and MODIS has 60–100%, then the MMD or disagreement is 20%. The MMD takes the most conservative approach to assessing disagreement. This process was applied to three pairs of land cover comparisons: GLC-2000 and MODIS v.5, GLC-2000 and GlobCover, and MODIS v.5 and GlobCover to create three maps of disagreement for forest and three for cropland. The three cropland disagreement maps were then further summed to create a single map of cropland disagreement; this was repeated to create a single forest disagreement map. Finally, a combined cropland/forest disagreement map was created by summing together the per cent disagreement within all the disagreement maps in the cropland as well as in the forest domain. The following thresholds were then applied to the map: (i) 0 to less than 5%—no disagreement (i) from 5% to 40%—'disagreement' and (ii) greater than 40%—'high disagreement'.

2.4. Quantification of the thematic accuracy and the disagreement

The accuracy of the cropland and forest classes was calculated using the validation data and the confusion matrices published

Table 2. The disagreement between different land cover products in hectares as a percentage of the average of all land cover maps in the forest and cropland domains.

Disagreement between land cover products	Cropland (Mha)	% relative to average across all products	Forest (Mha)	% relative to average across all products
Overall disagreement GLC-2000 versus MODIS v.5	325.8	23.4	730.8	18.5
Present in MODIS v.5 Absent in GLC-2000	94.8	6.8	517.9	13.1
Present in GLC-2000 Absent in MODIS v.5	231.1	16.6	212.9	5.4
Overall disagreement GlobCover versus MODIS v.5	505.9	36.4	387.2	9.8
Present in GlobCover, absent in MODIS v.5	360.0	25.9	285.6	7.2
Present in MODIS v.5, absent in GlobCover	145.8	10.5	101.7	2.6
Overall disagreement GLC-2000 versus GlobCover	395.2	28.4	314.3	8.0
Present in GLC-2000, absent in GlobCover	162.3	11.7	167.8	4.2
Present in GlobCover, absent in GLC-2000	232.9	16.8	146.5	3.7

in Mayaux *et al* (2006), Friedl *et al* (2010) and Bicheron *et al* (2008). The validation data for individual cropland and forest classes were first aggregated using table 1 and then the overall accuracy of these classes was calculated.

The total area of disagreement between each pair of land cover products was calculated separately for cropland and forest as well as the areas of commission and omission for each pair. To provide a reference figure for comparison across pairs of land cover products, the average cropland and forest areas across all products was calculated and used as a denominator to quantify the amount of disagreement.

3. Results

The total areas under cropland based on GLC-2000, GlobCover and MODIS v.5 are 1621 Mha, 1589 Mha and 1311 Mha, respectively while the corresponding figures for forest cover are 3876 Mha, 4029 Mha and 4241 Mha. The overall accuracy for the cropland/cultivated/cultivated and managed class for the GLC-2000 (classes 16–18) is 76%, MODIS v.5 (class 14 and 16) is 77% while GlobCover 2005 (class 11, 14, 20 and 30) decreases to 57.6%. Similarly, for the aggregated forest classes, the GLC-2000 (classes 1–9) has an accuracy of 81%, for MODIS v.5 (classes 1–5) it is 80%, and for GlobCover 2005 (classes 40–120) it drops to 60%.

Table 2 provides the overall differences in Mha and as a percentage of the average cropland and forest cover from the GLC-2000, MODIS v.5 and GlobCover 2005. The disagreement between the GLC-2000 and MODIS v.5 is 731 Mha and 326 Mha respectively for forest and cropland. Comparing the two newer products (i.e. GlobCover and MODIS v.5), the overall forest disagreement decreases to 387 Mha but the cropland disagreement increases to 506 MHa or 36% of the average area of the three land cover products.

The differences between the different land cover products highlighted in table 2 are even more significant when viewed spatially. Figure 1(a) provides a map of global disagreement between the two most recent and highest resolution products: MODIS v.5 and GlobCover 2005 for the cropland domain. Areas of high disagreement are visible across North America, Russia and across the tropical world. Large cropland disagreements are evident in China, North America and many countries in Africa. A more detailed example is provided in figure 1(b), which highlights an example in Ethiopia. Cropland disagreement with more cropland in MODIS is displayed in yellow and red shades, and disagreement with more cropland in GlobCover is displayed in blue shades (figure 1(b-A)). Individual land cover products are shown in figure 1(b-B) (GlobCover) and figure 1(b-C) (MODIS), which clearly demonstrates that they differ both thematically and spatially regarding the distribution of identical or similar land cover types. For example, large cultivated and managed areas on the GlobCover map are labelled forest/woody savannah or savannah on the MODIS map, and likewise, mosaic forest or shrubland on the GlobCover map is labelled as cropland on the MODIS map.

This phenomenon of large differences in both cropland and forest cover occurs in many regions of the world. The full set of maps showing the disagreement between each pair of land cover products as well as the combined disagreement for cropland and forest can be found on geo-wiki.org (Fritz *et al* 2009). A version showing disagreements in the urban domain can be found on urban.geo-wiki.org (Fritz *et al* 2011).

4. Discussion and conclusions

This paper has shown that there are critical differences between the land cover products as expressed by the spatial disagreement (particularly for cropland). For example, 360 Mha are identified as cropland in GlobCover but as non-cropland in MODIS, which is a discrepancy that equates to approximately 20% of the global cropland area. The thematic accuracy of GlobCover for the aggregated cropland and forest classes was also shown to be worse than that of the GLC-2000 and MODIS v.5, despite being a newer product that has already been downloaded more than 50 000 times by users (GEO 2011). These disagreements can be caused by differences in classification methodology, differences in training and ground reference data, the type of satellite sensors used and georeferencing errors (McCallum *et al* 2006, Fritz and See 2005). A small portion of the differences might also be attributed to differences in the date of the satellite data acquisition used in creating the land cover maps. Therefore, as a minimum guideline, these maps cannot be used for land cover change detection since the error in the original map is higher than the change detected (e.g. GLC-2000 versus GlobCover). Due to the large disagreements between these land cover products, we recommend that the user community does not, by default, use the latest product with the highest resolution, but carefully examines the sensitivity of these products within a specific application. In the situation where the maps are used for national and regional applications, we would recommend examining the disagreement of the products in the areas of interest and also to compare them with high resolution ground data or aerial photography. One way to do this would be to use geo-wiki.org (Fritz *et al* 2009), a global land cover validation tool, which can be used to visualize the global land cover products and the disagreement directly on top of Google Earth. By exploring the discrepancies at the level of an individual country in combination with local knowledge, the user can gain insight into which product is better in a specific region and which product is better suited for a particular application.

More research efforts should be directed towards finding ways to improve global land cover. There are promising developments on the horizon such as open access to the Landsat archive (Woodcock *et al* 2008, Roy *et al* 2010), the development of a new 30 m global land cover product being developed by the USA and China (US Department of the Interior 2010), and the launch of the Sentinel satellites over the next decade with a finer temporal and spatial resolution (ESA 2011). However, validation of future products will remain a crucial issue. This will require greater involvement of experts on the ground and the collection of a larger quantity of *in situ* data. The task of improving validation data for land cover datasets is now being increasingly discussed by the CEOS Cal Val land cover validation subgroup, which advocates the collection of more 'authorized' validation data. Visualization tools such as those provided by Google Earth represent a promising platform for validation and for the collection of increasing amounts of citizen volunteered information on land cover through crowd-sourcing for validation of land cover products. Crowd-sourced data in the form of geo-tagged photos and information collected through Web 2.0 applications like geo-wiki and smart phones could also be harnessed as a rich source of training and calibration data for global land cover algorithms. Moreover, serious gaming has a potential role to play in developing applications that engage a wider community in data collection as part of contributing to environmental solutions.

The benefits from these more promising solutions will only be reaped in the medium to long term. Bottom up initiatives such as geo-wiki.org may provide, at the very least, a short-term solution (Macauley and Sedjo 2010), particularly in the development of an integrated global land cover product. Such an integrated product would use existing regional and national land cover products where available, the best global product in situations where higher quality national maps are not available and crowd-sourced data provided by citizens. The design of appropriate land use policies with global dimensions requires reliable and accurate land cover data. Although the results of an integrated product will need to be assessed against fitness for purpose, the integrated assessment community in particular is eager to work with the best available products now. In developing countries in particular, good baseline data and monitoring techniques for ecosystem services are needed immediately (Andelman 2011). Waiting for future developments, however promising, is not an option if we are to tackle the burning issues surrounding sustainable provision of ecosystem services and providing sufficient food for a growing global population.

Acknowledgments

We would like to acknowledge the following EC FP7 Projects (EuroGEOSS no. 226487, CC-TAME no. 212535 and EnerGEO no. 226364) and the Austrian (FFG) (Landspotting no. 2008999) project, which have supported this work.

References

Andelman S A 2011 Conservation science outside the comfort zone *Nature* **475** 290–1

Ahlqvist O 2005 Using uncertain conceptual spaces to translate between land cover categories *Int. J. Geograph. Inform. Sci.* **19** 831–57

Bartholomé E and Belward A S 2005 GLC2000: a new approach to global land cover mapping from earth observation data *Int. J. Remote Sensing* **26** 1959–77

Bicheron P *et al* 2008 *GLOBCOVER: Products Description and Validation Report* (Toulouse: Medias France) (available online at http://ionia1.esrin.esa.int/docs/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf)

Bontemps S *et al* 2011 *GLOBCOVER 2009: Products Description and Validation Report* (ESA and UCLouvain) (available online at http://ionia1.esrin.esa.int/docs/GLOBCOVER2009_Validation_Report_2.2.pdf)

Cai X *et al* 2011 Land availability for biofuel production *Environ. Sci. Technol.* **45** 334–9

Di Gregorio A and Jansen L 2000 *Land Cover Classification System: Classification Concepts and User Manual* (Rome: Food and Agriculture Organization of the United Nations)

ESA (European Space Agency) 2011 *GMES Sentinels* (available online at www.esa.int/esaLP/SEM097EH1TF_LPgmes_0.html)

Feddema J *et al* 2005 A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present day land cover representations *Clim. Dyn.* **26** 581–609
Foley J *et al* 2005 Global consequences of land use *Science* **309** 570–4
Friedl M *et al* 2010 MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets *Remote Sensing Environ.* **114** 168–82
Fritz S and See L 2005 Comparison of land cover maps using fuzzy agreement *Int. J. GIS* **19** 787–807
Fritz S and See L 2008 Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications *Glob. Change Biol.* **14** 1057–75
Fritz S *et al* 2009 Geo-Wiki.Org: The use of crowdsourcing to improve global land cover *Remote Sensing* **1** 345–54
Fritz S *et al* 2010 Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa *Int. J. Remote Sensing* **25** 1527–32
Fritz S *et al* 2011 Building a crowd-sourcing tool for the validation of urban extent and gridded population *Lect. Not. Comput. Sci.* **6783** 39–50
GEO (Group on Earth Observations) 2011 Extracts from GEO 2012-2015 Work Plan V0 *GEO Work Plan Symp. (Geneva, May 2011)* (available online at ftp://ftp.earthobservations.org/201105_Work_Plan_Symposium/Outcomes%20of%20Discussions%20on%20Work%20Plan%20Tasks_V0/DS-10_LandCover.ppt)
Giri C 2005 Global land cover mapping and characterization: present situation and future research priorities *Geocarto Int.* **20** 35–42
Giri C *et al* 2005 A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets *Remote Sensing Environ.* **94** 123–32
Havlík *et al* 2011 Global land-use implications of first and second generation biofuel targets *Energy Policy* **39** 5690–702
Herold M *et al* 2008 Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets *Remote Sensing Environ.* **112** 2538–56
Liu L *et al* 2008 A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change *Glob. Planet. Change* **64** 222–35
Macauley M and Sedjo R 2010 Forests in climate policy: technical, institutional and economic issues in measuring and monitoring *Mitig. Adapt. Strat. Glob. Change* **16** 499–513
Mayaux P *et al* 2006 Validation of the global land cover 2000 map *IEEE Trans. Geosci. Remote Sensing* **44** 1728–37
McCallum I *et al* 2006 A spatial comparison of four satellite derived 1 km global land cover datasets *Int. J. Appl. Earth Observ. Geoinform.* **8** 246–55
Neumann K *et al* 2007 Comparative assessment of CORINE2000 and GLC2000: spatial analysis of land cover data for Europe *Int. J. Appl. Earth Observ. Geoinform.* **9** 425–37
Pielke R A 2005 Land use and climate change *Science* **310** 1625–6
Quaife T *et al* 2008 Impact of land cover uncertainties on estimates of biospheric carbon fluxes *Glob. Biogeochem. Cycles* **22** GB4016
Roy D P *et al* 2010 Accessing free Landsat data via the internet: Africa's challenge *Remote Sensing Lett.* **1** 111–7
See L and Fritz S 2006 Towards a global hybrid land cover map for the year 2000 *IEEE Trans. Geosci. Remote Sensing* **44** 1740–6
Seebach L *et al* 2011 Comparative analysis of harmonized forest area estimates for European countries *Forestry* **84** 285–99
US Department of Interior 2010 *United States Launches New Global Initiative to Track Changes in Land Cover and Use* (available online at www.doi.gov/news/pressreleases/United-States-Launches-New-Global-Initiative-to-Track-Changes-in-Land-Cover-and-Use-Data-Sharing-Will-Assist-Land-Managers-Worldwide.cfm)
Verburg P H *et al* 2010 Challenges in using land use and land cover data for global change studies *Glob. Change Biol.* **17** 974–89
Woodcock C *et al* 2008 Free access to Landsat imagery *Science* **23** 1011
You L *et al* 2009 Generating plausible crop distribution and performance maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach *Agricult. Syst.* **99** 126–40