BOREAL ENVIRONMENT RESEARCH 20: 243–260
ISSN 1239-6095 (print) ISSN 1797-2469 (online) Helsinki 30 April 2015

Assessment of land-cover data for land-surface modelling in regional climate studies

Markus Törmä[1)]*, Tiina Markkanen[2)], Suvi Hatunen[1)], Pekka Härmä[1)], Olli-Pekka Mattila[1)] and Ali Nadir Arslan[3)]

[1)] *Finnish Environment Institute, Data and Information Centre, P.O. Box 140, FI-00251 Helsinki, Finland (*corresponding author's e-mail: Markus Törmä, markus.torma@ymparisto.fi)*
[2)] *Finnish Meteorological Institute, Global Change Research, P.O. Box 503, FI-00101 Helsinki, Finland*
[3)] *Finnish Meteorological Institute, Arctic Research, P.O. Box 503, FI-00101 Helsinki, Finland*

Received 3 Dec. 2013, final version received 3 Dec. 2014, accepted 26 Nov. 2014

Törmä M., Markkanen T., Hatunen S., Härmä P., Mattila O.-P. & Arslan A.N. 2015: Assessment of land-cover data for land-surface modelling in regional climate studies. *Boreal Env. Res.* 20: 243–260.

We studied the land-cover data used by the regional climate model REMO and the land surface model JSBACH, for Finland and surrounding areas. To date, the land-cover data determining REMO's surface parameterisations have originated from the Global Ecosystem classification of the Global Land Cover Characteristics (GLCC-GEC) database. The same database has also been used as basis for prescribed plant functional type distribution with JSBACH. We showed that the GLCC-GEC does not represent the Finnish landscape particularly well, and there are large errors in the land cover type distributions. Furthermore, we have inspected the values of the land surface parameters forest ratio and leaf area index, which were assigned to land-cover types, and found them to typically be too large for Finland. Different revised land-cover data sets were created using GlobCover and different versions of Corine Land Cover (CLC) classifications. The benefits of the new land-cover data sets were much more spatial detail and thematic content which corresponded better to the Finnish environment, unlike in the GLCC-GEC. For example, there are wetlands and they are correctly located. Although no definite reference exists to assess the qualification of the land-cover data in the light of the model results, modelling benefits from the use of land-cover data that is more spatially accurate and recent. Even though regionally the differences are not great, at a more local level they become substantial.

Introduction

Land cover is the observed (bio)physical cover of the Earth's surface, which should include directly observable vegetation and man-made structures, but quite often bare rock, soil and water are also included. A companion to land cover is land use, which is defined as the human arrangements, activities and inputs on a certain land cover type to produce, change or maintain it. Changes in land cover and land use either affect global systems (e.g. the atmosphere and oceans) or occur in a localised fashion, impacting microclimates as well as air and water quality. Hence, land cover is a geographical feature which may form a reference base for applications including forest and rangeland monitoring, production of statistics, planning, investment,

Editor in charge of this article: Eero Nikinmaa

biodiversity, climate change, and desertification control. Nowadays, it has been acknowledged that it is highly important to know how land cover has changed over time, in order to make assessments of the changes to be expected in the future, and the impact these changes will have on peoples' lives (Di Gregorio and Jansen 2000).

An accurate representation of the land surface is needed for describing the interactions between the surface and the atmosphere in climate modelling. The evapotranspiration of vegetated surfaces affects both the hydrological cycle and the energy balance. Surface albedo is an important determinant in the radiation balance, and surface roughness regulates the surface friction extracting energy from the atmospheric circulation. The above-mentioned interactions are key to general circulation models, as well as in their limited area counterparts (Hagemann *et al*. 1999). In the context of Earth system models targeted at the global change research, ecophysiological feedback between surface vegetation and the climate are also accounted for (Bonan 2008). For evaluation of the surface processes of such models, it is essential to assess their performance at the current state of the land cover (Jung *et al*. 2006).

There are several alternatives to land-cover data for global modelling purposes. Some of the alternatives are ordinary classifications with global or regional classification nomenclature, whilst others provide land surface parameters in addition to classification. One of the most widely used global land-cover classification is the Global Land Cover Characteristics Database, developed by the U.S. Geological Survey (USGS), the University of Nebraska-Lincoln (UNL), and the European Commission's Joint Research Centre (JRC). It has 1 km spatial resolution and is based on the interpretation of NOAA AVHRR-images from April 1992 to March 1993. There are seven global data sets, each representing a landscape based on a different classification legend, using 8 to 96 classes (U.S. Geological Survey 2001, Loveland *et al*. 2000).

A slightly newer global data set is Global Land Cover 2000 (GLC2000), which has been coordinated and implemented by the Global Vegetation Monitoring unit of the JRC. Its general objective is to provide a harmonised land-cover database with 1 km spatial resolution over the entire globe for the year 2000. Regional partners have used their own classification methodology to classify Spot Vegetation image time series from 1999. The classification system is based on the FAO Land Cover Classification System, which allows the regionally defined legends to be translated into more generalised global land-cover classes for the GLC2000 global product. The global classification legend has 22 classes, but regional classifications can have more, for example the classification of northern Eurasia has 26 classes (Bartalev *et al*. 2003, GLC2000 2010).

The MODIS Land Cover Type product contains five classification schemes, which describe land cover properties with 500 m spatial resolution, derived from observations spanning a year's input of MODIS data from Terra and Aqua satellites. There are five land-cover classification schemes with 9 to 17 classes that are produced for each calendar year, derived from a supervised decision-tree classification method. There is also a product describing seasonal phenology (Friedl *et al*. 2010). The positive characteristics of this product are better spatial resolution than in earlier products and annual updates giving information about land cover changes. The main drawback is the relatively poor thematic resolution.

The GlobCover land cover map produced by the European Space Agency and University of Louvain contains 22 land cover classes based on the FAO Land Cover Classification System. The classification is based on a time series of Envisat MERIS images with 300 metre spatial resolution. Original GlobCover 2005 was based on images from 2005, and the second version, GlobCover 2009, on images from 2009. The positive characteristics of this product are even better spatial resolution, and the inclusion of classifications from two different times, the drawback again is the relatively poor thematic resolution (Bontemps *et al*. 2011).

The drawback of the previously presented data sets is that they are nothing but classifications, and the number of classes can be quite low. The overall classification accuracy, in other words the probability that the pixel has been classified correctly, can be rather moderate at 70%. Although the classification algorithms are more automatic, and the quality of remote sensing data has increased, the repeatability of classifications can be surprisingly poor; when different versions of MODIS land-cover classification were compared, significant year-to-year variations were noted, 10%–30% of pixels in land-cover labels were not associated with land-cover change. In the case of GlobCover 2005 and 2009, this variability affects about 25% of the pixels (Jung *et al*. 2006, Bontemps *et al*. 2012).

The climate modelling community is interested in the role of land-use and land-cover change in assessing the impacts and vulnerabilities of land cover and its change. However, these modellers experience problems in that the classes of land-cover classification are not usually easily translated into what models actually need, and this is one extra source of uncertainty in modelling. What these modellers need is stable land-cover products, including vegetation dynamics using the land surface parameters that models use; flexibility, so that the same product can be used with different models; and reasonable spatial resolution, being about 300 m for global modelling (Jung *et al*. 2006, Bontemps *et al*, 2012). However regional modelling requires greater detail, meaning a smaller pixel size. It should also be pointed out that landscape has an effect on the pixel size requirement; if the landscape consists of many small patches then a smaller pixel size is needed.

One way to enhance land-cover classification is to provide classes with land-surface parameters, in other words continuous variables which describe vegetation and its annual changes as well as soil properties. Land-cover data used in the REMO and JSBACH models is based on the Global Ecosystem Classification of Global Land Cover Characteristics (GLCC) database, where Hagemann *et al*. (1999) has assigned land-surface parameters for classes. Another similar land-cover dataset is Ecoclimap (Champeaux *et al*. 2005) which is a combination of existing global land-cover map, climate regions, and land-surface parameters of classes determined using various sources. Another way to improve the situation is to combine several different land-cover classifications in one classification, taking into account the individual advantages and limitations of different classifications (Jung *et al*. 2006).

The objectives of this study were to determine how well the land-cover data used by the modelling system represents northern European land cover, how those land-cover data could be improved using other land-cover classifications, and what the effect of different land-cover data is on modelling results. We assumed that the modelling results would improve as better land-cover data would give more realistic distributions of vegetation and plant functional types. As the land-cover data consists of discrete classification and the assigned land-surface parameters, both are validated by comparing them to local, more detailed land-cover data or surface parameters from Finland. Finnish high resolution Corine land-cover classification (Törmä *et al*. 2004, Härmä *et al*. 2005, Törmä *et al*. 2011) is the most accurate land-cover classification of Finland, and we consider it as ground truth when compared with other land-cover classifications.

Material and methods

Modelling environment

The limited area climate model REMO (Jacob 2001, Jacob and Podzun 1997) whose land-cover data and related parameter allocations were analysed in this work, is based on the operational weather forecast model of the German weather service. The resolved variables (air temperature, specific humidity, wind speed, etc.) are represented vertically by a hybrid vertical coordinate system throughout the atmosphere (Simmons and Burridge 1981). Horizontally REMO is operated in a rotated spherical grid, the resolution of which was 0.167° in our application. REMO has an implicit land-surface model (LSM) that accounts for the surface–climate exchange of energy and water.

REMO has a fractional surface coverage, i.e., each grid box can contain land, water and sea ice fractions. While topography, surface roughness length, land–sea mask, forest fraction and soil field capacity are constant in time (Hagemann 2002), surface background albedo, vegetation fraction and leaf area index (LAI) have a prescribed yearly cycle (Rechid 2008, Rechid e*t al*. 2009). The above-mentioned surface parameter values are allocated to the land-cover types and aggregated to grid cell-specific values according to the fractions of land-cover types within each grid cell.

In this work, JSBACH (Reick *et al*. 2013) model, which is the LSM of the Max Planck Institute-Earth System Model (MPI-ESM) (Giorgetta *et al*. 2013), was used to produce regional gross primary production (GPP) by photosynthesis of land ecosystems. Its primary function is to provide the climate model ECHAM6 (Stevens *et al*. 2013) with energy and matter balance terms related to land-surface processes. These interactive terms include radiation and sensible and latent heat, as well as water and carbon exchange in terms of CO_2, between the surface and the atmosphere.

In JSBACH vegetation-related surface parameters are linked to a fractional presentation of plant functional types (PFT), which occupy a model grid cell corresponding their areal portions in land-cover data (Reick *et al*. 2013). In our application, the four most prominent PFTs were accounted for and their fractions were prescribed and fixed in time. A new five-layer soil-moisture scheme (Hagemann and Stacke 2013) was applied, whose soil information was based on soil-texture data (FAO 1971–1981).

Here we ran JSBACH for a regional domain centred on Finland, including Scandinavia and the Baltic countries with a resolution of 0.167°. As in its primarily applications the domain is global with typical resolution of 1.9° (Reick *et al*. 2013), we needed to refine the boundary and initial data on surface characteristics for our application.

In order to assess the impact of land-cover data on the surface carbon-balance we ran JSBACH with three PFT distributions based on the GLCC-GEC, GlobCover and a combination of Finnish high resolution (HR) Corine Land Cover (CLC), European CLC and GlobCover classifications (*see* Tables 1–3 for class names). To produce grid cell-specific PFT fractions for JSBACH, the translation rules given in Table 4 were applied between the GLCC-GEC and PFT. In this conversion, only the grid cells with sea or lake fractions of less than 50% were taken into account. All the other surface boundary fields, such as soil moisture parameters, were left unchanged.

The hourly climatic forcing for JSBACH was produced with REMO run that used ERA-Interim (Dee *et al*. 2011) from 1980 to 2011 as boundary data. Preceding the 31-year forward run, a spin-up of 5 years was performed in order to stabilise the deep soil moisture and temperature values at realistic levels. The surface parameterisation for this run was based on the CLC2000, and a detailed description of the generation of the parameter fields, as well as the assessment of its impact on the regional climate simulations are given in Gao *et al*. (2014).

Land-cover data

Both REMO and former versions of ECHAM use a land-surface parameter (LSP) data set (Hagemann *et al*. 1999, Hagemann 2002) based on the USGS Global Land Cover Characteristics database ver. 2.0 (Loveland *et al*. 2000, U.S. Geological Survey 2001). The database contains several land-cover classifications based on the interpretation of NOAA AVHRR 10-day mosaics. One of these classifications is the Global Ecosystem classification, hereafter known as the GLCC-GEC, which represents the Earth's seasonal land cover in a consistent global framework by identifying 94 ecosystem types (Table 3), using the definitions given by Olson (Olson 1994a, 1994b). For climate modelling purposes, the following surface parameters were allocated to each class of the GLCC-GEC: background surface albedo α_s, surface roughness length due to vegetation $z_{0,veg}$, fractional vegetation cover c_v and leaf area index LAI (the ratio of the leaf area to the projection) for the growing and dormancy seasons, forest ratio c_f, plant-available soil water holding capacity W_{ava}, and volumetric wilting point f_{pwp} (Claussen *et al*. 1994, Hagemann *et al*. 1999, Hagemann 2002).

The surface data were aggregated into surface maps of the size and resolution of the prospective modelling domain. For most of the parameters the pre-processing steps consist of weighed areal averaging, but for some, such as roughness length, the processing is more complicated (Hagemann 2002). In the Nordic areas, the aggregation of two parameters, soil field capacity and fractional vegetation, which is related to forest ratio, has been modified to account for region specific vegetation and soil features. Field capacity is typically high in areas of forested wetlands in Finnish and Swedish Lapland, where

Table 1. Corine land-cover classes on level 2 and their proportions of Finnish Territory. Level 3 classes are also listed.

CLC level 2 class	Proportion (%) of Finnish territory	Level 3 classes
11 Urban fabric	1.5	111 Continuous urban fabric, 112 Discontinuous urban fabric
12 Industrial, commercial and transport units	0.7	121 Industrial or commercial, 122 Road and rail networks, 123 Port areas, 124 Airport
13 Mine, dump and construction sites	0.1	131 Mineral extraction, 132 Dump sites, 133 Construction sites
14 Artificial non- agricultural vegetated area	0.6	141 Green urban areas, 142 Sport and leisure facilities
21 Arable land	5.8	211 Non-irrigated arable land
22 Permanent crop	0.0	222 Fruit trees and berry plantations
23 Pasture	0.2	231 Pastures
24 Heterogeneous agricultural area	0.6	242 Complex cultivation, 243 Land principally occupied by agriculture
31 Forest	43.7	311 Broadleaf, 312 Coniferous, 313 Mixed
32 Shrubs/herbaceous vegetation	17.4	321 Natural grassland, 322 Moors and heathland, 324 Transitional woodland/shrub
33 Open spaces with little or no vegetation	0.5	331 Beaches, dunes, sand, 332 Bare rock, 333 Sparsely vegetated areas
41 Inland wetland	7.0	411 Inland marshes, 412 Peatbogs
42 Coastal wetland	0.1	421 Salt marshes
51 Inland water	8.5	511 Water courses, 512 Water bodies
52 Sea water	13.4	523 Sea and ocean

Table 2. Classes of GlobCover classification within Finnish territory, their proportions and the Finnish HR CLC classes (*see* Table 1 for class names) within those GlobCover classes.

GlobCover Class	GlobCover proportion in Finland (%)	Most common CLC classes
14 Rainfed croplands	0.003	21, 51, 313, 312
20 Mosaic cropland (50%–70%)/vegetation (20%–50%)	0.03	52, 51, 312, 313
30 Mosaic vegetation (50%–70%)/cropland (20%–50%)	0.00005	52
50 Closed (> 40%) broadleaf deciduous forest ($h > 5$ m)	11.9	313, 312, 32, 21
70 Closed (> 40%) needleleaf evergreen forest ($h > 5$ m)	0.03	312, 52, 51
90 Open (15–40%) needleleaf deciduous or evergreen forest ($h > 5$ m)	27.1	312, 313, 32
100 Closed to open (> 15%) mixed broadleaf and needleleaf forest ($h > 5$ m)	21.8	312, 313, 32
110 Mosaic forest or shrubland (50%–70%)/grassland (20%–50%)	3.2	312, 313, 21, 32
140 Closed to open (> 15%) herbaceous vegetation	0.003	23, 52
150 Sparse (< 15%) vegetation	12.9	32, 312, 21, 41
180 Closed to open (> 15%) grassland or woody vegetation, regularly flooded or waterlogged	3.0	41, 312, 32
190 Artificial surfaces and associated areas (Urban areas > 50%)	0.4	11, 12, 312
200 Bare areas	0.05	33
210 Water bodies	19.4	52, 51

the land cover is boreal coniferous forest. The allocated soil field capacity is 0.21, which is too low for these soils. Thus the value was overwritten with a constant value of 0.71 according to the distribution of class 15 of the FAO soil-type data set (FAO 1971–1981, Hagemann 2002). The fractional vegetation cover of the land-cover class boreal coniferous forest was in Fennoscandia increased from 0.52 to 0.91.

There are different alternatives for land-

Table 3. Classes of the GLCC-GEC ecosystem classification within Finnish territory, their proportions, the most common Finnish HR CLC classes (*see* Table 1 for class names) within those GLCC-GEC classes, and which GlobCover (Table 2) or CLC (Table 1) classes were recoded to which GLCC-GEC classes.

GLCC-GEC class	Proportion (%) of Finnish territory	Most common CLC classes	GlobCover recoded	CLC recoded
1 Urban	0.03	11, 12	190	111, 121, 122, 123, 131, 132, 133
2 Low Sparse Grassland	0.03	312, 313, 51	–	–
4 Deciduous Conifer Forest	3.2	313, 312, 32, 41	91	–
8 Bare Desert	0.01	33, 32	201	–
9 Upland Tundra	0.01	32, 33	–	–
11 Semi Desert	–	–	200	–
12 Glacier Ice	> 0.01	33	220	–
13 Wooded Wet Swamp	–	–	180	–
14 Inland Water	9.8	51, 21, 31	210	511, 512
15 Sea Water	14.2	52	–	523
16 Shrub Evergreen	–	–	131	–
17 Shrub Deciduous	3.2	31, 32, 41	134	–
19 Evergreen Forest and Fields	0.01	313, 312, 21, 32	–	–
21 Conifer Boreal Forest	44.6	312, 313, 32	70, 90, 92	312
22 Cool Conifer Forest	0.1	21, 31	–	–
23 Cool Mixed Forest	10.2	312, 313, 32	101	313
25 Cool Broadleaf Forest	–	–	50, 60	141, 311
30 Cool Crops and Towns	–	–	112, 124, 142	–
31 Crops and Town	0.02	11, 21, 312, 313	–	–
38 Cool Irrigated Cropland	0.01	12, 32, 21, 14	–	–
40 Cool Grasses and Shrubs	–	–	120	321
42 Cold Grassland	0.7	32	150, 151, 152, 140, 141	231
44 Mire, Bog, Fen	–	–	185	412
45 Marsh Wetland	–	–	–	411, 421
47 Dry Woody Shrub	–	–	–	324
50 Sand Desert	–	–	202	331
51 Semi Desert Shrubs	> 0.01	51	–	–
53 Barren Tundra	0.07	33, 32	–	333
55 Cool Fields and Woods	0.3	312, 313, 21, 32	20	–
56 Forest and Field	0.08	312, 313, 32, 21	–	–
57 Cool Forest and Field	0.8	312, 313, 21, 32	32, 110	–
60 Small Leaf Mixed Woods	1.7	32, 311, 41	-	–
61 Deciduous and Mixed Boreal Forest	1.6	312, 313, 32, 21	100	–
62 Narrow Conifers	6.8	312, 313, 32, 21	–	–
63 Wooded Tundra	0.6	32, 33	–	–
64 Heath Scrub	0.1	32, 312, 33, 51	–	322
69 Polar and Alpine Desert	> 0.01	33, 32	–	332
93 Grass Crops	1.7	312, 313, 32, 21	–	211
94 Crops, Grass, Shrubs	0.07	312, 32, 51	14, 21, 30	222, 243

cover data, such as the GLC2000 (Bartalev *et al*. 2003, GLC2000 2010) or MODIS Land Cover Type (Friedl *et al*. 2010), but if there is a need for more spatially-detailed data then the GlobCover or Corine Land Cover classifications could be useful. The global land-cover classification GlobCover ver. 2.2 is based on MERIS images. After detecting cloud and snow, images are mosaicked and the mosaics classified. Its 22 land-cover classes (Table 2) have been defined using the UN Land Cover Classification System. Different geographical regions can have their own more detailed classes (Bicheron *et al*. 2008).

The Corine Land Cover classification, also called the European CLC, is a pan-European project aimed at gathering information related to the environment for the European Union. The classification was produced using satellite images and visual interpretation. The mapping scale is 1:100 000 with a minimum mapping unit of 25 hectares and a minimum width of units of 100 m. Only area elements are classified. The classification nomenclature is hierarchical and contains five classes at the first level, 15 classes at the second level and 44 classes at the third level (Törmä *et al*. 2004).

The Finnish Corine Land Cover classification (Törmä *et al*. 2004, Härmä *et al*. 2005, Törmä *et al*. 2011) is a combination of existing digital map data and land-cover interpretation of satellite images. The land-cover interpretation is performed by estimating various variables describing tree and vegetation cover for each image pixel, and thresholding these to the CLC classes (Table 1). The satellite images were Landsat-7 ETM+ images in the case of the CLC2000, and Spot-4/5 XS and IRS LISS images in the case of the CLC2006. CLC classes related to land use are created by recoding digital map data to the CLC classes, and in some cases updating them with satellite images. The result is a raster database, called the Finnish high resolution (HR) CLC with a resolution of 25 m in the case of the CLC2000 and 20 m in the case of the CLC2006. The European version, with a 25 ha minimum mapping unit, was created using automatic generalisation methods. The Finnish HR CLC is the most accurate land-cover classification of Finland, because it is based on:

— digital map data such as the Topographic database of the National Land Survey, which is produced using photogrammetric interpretation of aerial images,
— registers such as the Building and Dwelling register containing information about every building in Finland, or
— interpretations of moderate resolution satellite images like the forest parameters of the National Forest Inventory by the Finnish Forest Research Institute.

Therefore, we consider it to be ground truth when compared with other land-cover classifications.

The tree-crown cover estimated by the Finnish Forest Research Institute Metla was used as reference when the forest ratio assigned to the GLCC-GEC classes was evaluated. Tree crown cover was estimated using field sample plots measured in the National Forest Inventory and IMAGE2006 satellite data using the *k*-NN classification method (Tomppo 2006) for the CLC2006 classification process. The crown cover of deciduous and coniferous trees has been estimated for each 20 meter grid cell for all of Finland in the range 0%–100%. Open, non-forested land-cover classes were given a crown cover of 0%, including agricultural areas, water bodies, open bogs, and so on. These data are now available from http://kartta.metla.fi/index-en.html.

Table 4. Translation rules between GLCC-GEC classification and PFTs in JSBACH. Only the most prominent land-cover types according to the GLCC-GEC (*see* Table 3) and Cool Broadleaf Forest, which is an important class in other land-cover data sources, are included.

GLCC-GEC class	PFT
25 Cool Broadleaf Forest	Temperate broadleaf deciduous trees
21 Conifer Boreal Forest	Coniferous evergreen trees
23 Cool Mixed Forest	50% temperate broadleaf deciduous trees, 50% coniferous evergreen trees
62 Narrow Conifers	50% coniferous evergreen trees, 50% coniferous deciduous trees

Reference data for the evaluation of the leaf area index (LAI) was obtained from the VALERI project (*see* http://w3.avignon.inra.fr/valeri/) and the University of Helsinki's (UH) LAI map of Finland (Heiskanen *et al*. 2011a, Heiskanen *et al*. 2011b). The VALERI data cover three approximately 3 km × 3 km sites in Finland: Hirsikangas, Rovaniemi and Hyytiälä. LAI was measured in the field and estimated for pixels of satellite images covering the period of ground measurements using an empirical model. Another source for high-resolution LAI information was the UH LAI maps covering entire Finland, which are based on IMAGE2000 and IMAGE2006 satellite image mosaics and *in-situ* data. Effective LAI produced was corrected for shoot-level clumping in coniferous and mixed forests using the Finnish HR CLC2006 classification.

We made comparisons between different classifications in order to get an idea of the thematic content of the GLCC-GEC and GlobCover classes and their possible thematic errors. We transformed the land-cover classifications to common coordinate system and counted the proportions of Finnish HR CLC2000 classes for each GLCC-GEC or GlobCover class. In other words, we took the area of the class from a lower spatial resolution classification and counted the proportions of different Finnish HR CLC classes within that area. We also counted the proportions of different classes within Finnish territory. Due to different classification systems, we had to group classes together to make them comparable. These groups were based on the CLC classification system, and were CLC1 Artificial surfaces, CLC2 agricultural areas, CLC31 Forests, CLC32 Transitional woodlands and open spaces on mineral soil, CLC4 Wetlands and CLC5 Water.

We validated the two land-surface parameters — forest ratio and LAI — by comparing them with tree-crown cover estimates from the Finnish National Forest Inventory and high-resolution LAI maps, respectively. We compared the forest ratio of the selected GLCC-GEC classes with the mean tree-crown cover values computed from tree-crown cover maps. We computed these mean values in two different ways; (1) the class area was defined using the GLCC-GEC with a 1-km pixel size, and (2) using the Finnish HR CLC with a 25-m pixel size. The selected GLCC-GEC classes were 1 Urban, 14 Inland Water, 21 Conifer Boreal Forest and 23 Cool mixed Forest. We compared the LAI values of selected GLCC-GEC classes with the mean LAI values computed from the high resolution LAI maps from VALERI and UH. Finnish HR CLC was recoded into the GLCC-GEC classes and mean LAI was calculated for each class. The selected GLCC-GEC classes were 21 Conifer Boreal Forest, 23 Cool Mixed Forest, 25 Cool Broadleaf Forest, and 47 Dry Woody Scrub.

As the GLCC-GEC is quite outdated, its spatial and thematic resolution is poor and there are newly produced global, regional and local land-cover classifications available, we produced new land-cover classifications for the modelling. We used the following land-cover classifications instead of the GLCC-GEC (Fig. 1a): The regional ver. 2.2 of GlobCover land-cover classification (Fig. 1b) (Bicheron 2008), as well as the CLC in its European (Fig. 1c) and Finnish HR (Fig. 1d) versions (Törmä *et al*. 2011). As the modelling environment uses the GLCC-GEC (Fig. 2a) nomenclature, we recoded the revised land-cover classifications to use the GLCC-GEC nomenclature (Fig. 2b, c and d). We did this by comparing class definitions in different classifications with the aid of surface parameters allocated to each land cover category in the GLCC-GEC. We carried out the following processing:

- We recoded GlobCover classification to the GLCC-GEC classes (Table 3, fourth column). As the pixel size of GlobCover classification is 300 m, majority filtering was done using a 3 × 3 filtering window, and then the data was resampled into a 1000 m grid using the nearest-neighbour interpolation method (Fig. 2b).
- We recoded the CLC to the GLCC-GEC classes. This recoding was carried out by finding the closest GLCC-GEC class for each CLC class (Table 3, fifth column). In its original form, the European CLC classification is vector data with a 25-hectare minimum mapping unit. A rasterised version with a 100-m pixel size was used in this study. Due to differences in pixel sizes (CLC 100 m, GLCC-GEC 1000 m), processing was carried out so that class recoding was done first, then majority-

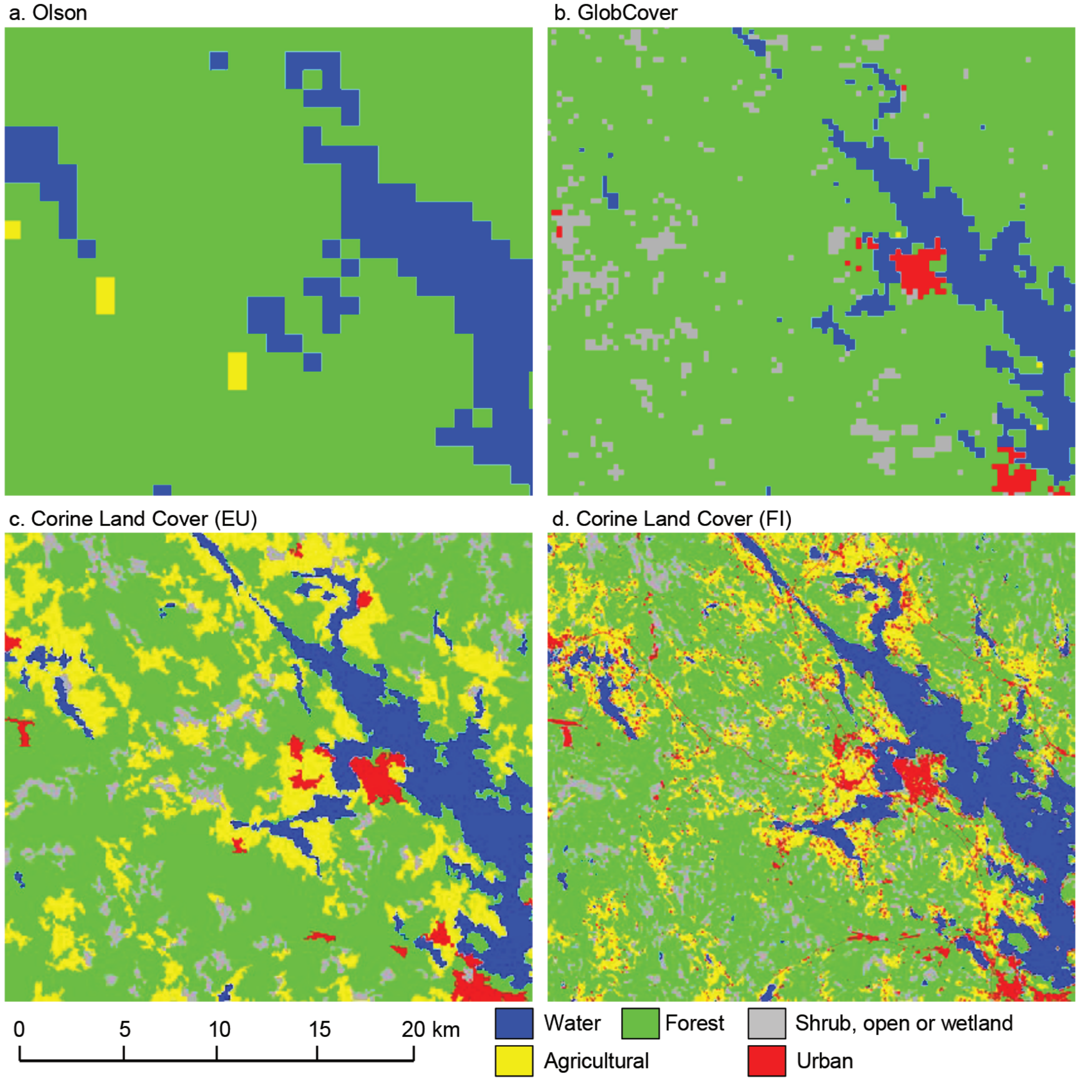

Fig. 1. An example of different land-cover classifications, (**a**) the GLCC-GEC (1-km pixel size), (**b**) GlobCover (300 m), (**c**) the European CLC (100 m), and (**d**) the Finnish HR CLC (25 m). The figures show the town of Ikaalinen and its surroundings.

filtering with a 9 × 9 filtering window, and finally resampling to a 1000 m pixel size using the NN interpolation (Fig. 2c).

— The Finnish HR CLC was used instead of the European CLC in Finland. Processing works just as well as with the European CLC, but the size of the majority filter was 41 × 41 pixels. Outside Finland, the European CLC and GlobCover were used (Fig. 2d).

The geographical coverage of these land-cover classifications varied and not all covered the whole modelling window. Therefore, two different combinations of land-cover classifications were created:

— classification based on GlobCover, and
— classification based on the Finnish HR CLC within Finland, the European CLC where available and Globcover elsewhere.

We compared the statistical significance of

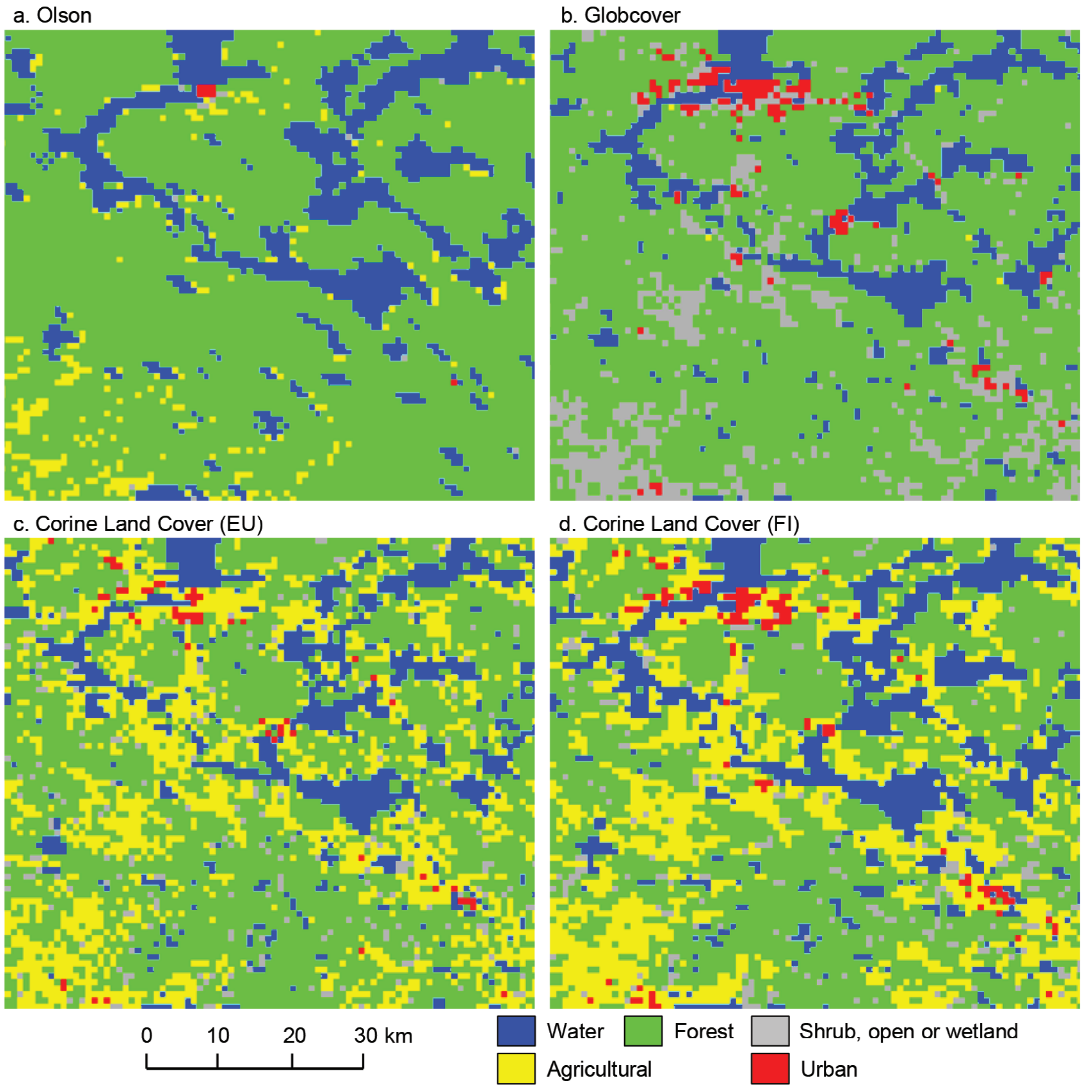

Fig. 2. An example of different land cover classification alternatives for modelling, (**a**) the original GLCC-GEC, (**b**) recoded and resampled GlobCover, (**c**) the recoded and resampled European CLC, and (**d**) the recoded and resampled Finnish CLC. In the figures, the city of Tampere is located near the top.

the differences between the used classifications with unpaired Student's *t*-test.

Results

Validation of land cover classifications and land surface parameters

We compared the GLCC-GEC classes with the Finnish HR CLC2000 classification. In all, there are 29 different GLCC-GEC classes within Finnish territory (Table 3), but just three classes (21 Conifer Boreal Forest, 15 Sea Water, 23 Cool Mixed Forest) cover 69% of this area. There are seven other classes with individual proportions of 1%–10% that cover 28% of Finland. For the rest of the country, 19 classes cover about 3%. According to more accurate and recent global land-cover classification, GlobCover, about 20% of Finnish territory is covered by water, including sea, 13% sparse vegetation, 3% wetlands and the rest is forests (Table 2). The proportion of urban areas is very low, only 0.4%.

We compared how different classifications divided the Finnish territory into the formed general classes based on the Corine Land Cover classification (CLC1-CLC5) (Table 5). The proportion of forests was overestimated in GlobCover and the GLCC-GEC and the proportions of smaller classes underestimated when compared with the Finnish HR CLC. This is typical in the sense that the proportion of the dominant class increases in the coarser-resolution data, and the proportions of less dominant classes decrease. The exception was the CLC33 Open spaces of GlobCover, whose proportion was much greater in GlobCover than in the HR CLC. It seems that CLC33 of GlobCover included areas which are classified as CLC2 or CLC32 in the HR CLC. One major failure is that there are no open bogs, mires and marshes in the GLCC-GEC classification of Finland. Quite often agricultural areas in the HR CLC are in forest classes in the GLCC-GEC. This is because the Finnish landscape is a mosaic of small polygons with different land covers and uses. In addition to this, the open bogs of the HR CLC have usually been classified as forest in the GLCC-GEC. Compared with the GLCC-GEC, the benefit of GlobCover is that wetlands are included, as well as being relatively well located. It can be seen from the examples of different classifications for modelling (Fig. 2) that forested areas and water are over-represented in the GLCC-GEC, and there are too many sparsely vegetated areas and a lack of agricultural areas in GlobCover.

We also compared the average patch size of land cover classifications, in order to evaluate their possibilities for showing the details of a landscape. The average land cover patch size of different classifications on the test area varied a lot (Table 6), from 0.7 ha for the Finnish HR CLC to 1643 ha for the GLCC-GEC. The number of classes in different classifications also varied a lot, from 35 classes in the Finnish HR CLC to 11 classes in GlobCover and the GLCC-GEC. The patch size depends on the properties of an area itself, such as the geometric and thematic variations of the landscape, but also the technical characteristics of the classification such as the minimum mapping unit and the number of classes used to represent the thematic variation. A smaller minimum mapping unit and more classes means that a classification is able to represent the smaller geometric and thematic details more accurately. On the other hand, just decreasing the minimum mapping unit does not necessarily cause a decrease in patch size if thematic detail is poorly represented due to a small number of classes. The minimum mapping units of the European CLC and GlobCover are 25 ha and 9 ha, respectively, indicating that GlobCover should represent the landscape with more details

Table 5. The percentages of land-cover classes of different land-cover classifications within Finnish territory. The land cover classes were based on the CLC classification system, and were CLC1 Artificial surfaces, CLC2 agricultural areas, CLC31 Forests, CLC32 Transitional woodlands and CLC33 open spaces on mineral soil, CLC4 Wetlands and CLC5 Water.

	HR CLC	EU CLC	GlobCover	GLCC-GEC
CLC1	2.8	1.2	0.4	0.03
CLC2	6.1	7.6	0.03	2.8
CLC31	44.1	50.7	61.1	69.1
CLC32	18.1	13.5	3.2	3.7
CLC33	0.6	0.3	12.9	0.1
CLC4	7.4	5.8	3.0	0
CLC5	20.9	20.9	19.4	24.0

Table 6. The characteristics of different land-cover classifications of a test area used to compare the average land-cover patch size of different classifications.

	Fi HR CLC	EU CLC	GlobCover	GLCC-GEC
Number of Classes	35	19	11	11
Minimum mapping unit	25 m pixel	25 ha polygon	300 m pixel	1000 m pixel
Number of samples	1232552	10091	10315	493
Mean ± SD size (ha)	0.7 ± 28.5	80.3 ± 1238.9	78.5 ± 611.1	1643.0 ± 26565.2
Minimum (ha)	0.03	0.5	4.5	50.0
Maximum (ha)	24589.1	104303.8	33123.4	588262.5

resulting smaller land-cover patch size. Overall, they had about the same number of land-cover patches within the test area (Table 6) and the likely reason is that the European CLC uses 19 classes to represent the thematic details, whereas GlobCover uses just 11 classes.

We compared the surface parameter forest ratio of the GLCC-GEC classes (Hagemann 2002) with tree-crown cover estimated by the Finnish Forest Research Institute (Metla). When we defined coniferous forest according to the Finnish HR CLC, the mean value for crown cover was 0.46 for the whole Finland (Table 7, class 21 Coniferous boreal forests). The density of forests varies within Finland in different vegetation zones: in the northern boreal zone the average crown cover was 0.30 (SD = 0.10, n = 94846370), while the corresponding value for the southern boreal zone was 0.50 (SD = 0.13, n = 113109816). The difference between these mean values was statistically significant (t-test: t_{14} = 12528.2, two tailed $p < 0.0001$). When we defined the coverage of coniferous forests using the GLCC-GEC with a 1-km spatial resolution, the mean crown cover was only 0.27. The inaccuracy of the delineation of coniferous forests in the GLCC-GEC and its low spatial resolution causes mixing of sparsely forested areas with dense forests, which decreases the crown cover values of the GLCC-GEC. In mixed forests (class 23), the value given is much higher than the estimated crown cover. The forest ratio can be greater than zero for water (class 14) because the pixel size is large and mixed pixels exist on shorelines. The forest ratio values given for urban areas (class 1) are too low because in reality urban areas also contain some treed or forested areas. In all cases, the difference of class mean tree crown cover estimated using the GLCC-GEC and the Finnish HR CLC was statistically significant (Table 7).

The LAI values given for the growing season of the GLCC-GEC classes (Hagemann 2002) were considerably higher than the corresponding estimates from the VALERI or UH LAI maps (Table 8). Although the LAI values of the VALERI data tend to be underestimated, since foliage clumping is not taken into account, it can be concluded that generally the LAI values of the GLCC-GEC are too high. LAI of the GLCC-GEC should be lower for classes 21 Conifer Boreal Forest and 25 Cool Broadleaf Forest, and could be a little lower for class 47 Dry Woody Scrub. The LAI value for class 23 Cool Mixed Forest seems to be relatively good. The VALERI and UH LAI-estimates for different land-cover classes were significantly different for each class (Table 8). This illustrates that LAI is quite a difficult parameter to estimate using remote sensing.

Effect of land cover data on regional carbon modelling

The impact of replacing the GLCC-GEC with the CLC2000 in regional climate model REMO was studied by Gao *et al*. (2014) and it was found that the CLC2000 slightly improved in the simulated air temperature and precipitation. However, the deviations of model results from measurements were a lot larger than the deviations between the surface parameter sets.

Table 7. The forest ratio values given (FR GEC) and statistics of tree crown cover estimated by the Finnish Forest Research Institute Metla for classes 1 Urban, 14 Inland Water, 21 Conifer Boreal Forest and 23 Cool Mixed Forest. Classes have been defined using the GLCC-GEC (GEC) with 1-km pixel and the Finnish HR CLC (CLC) recoded to the GLCC-GEC classes with 25-m pixel.

	Class 1	Class 14	Class 21	Class 23
FR GEC	0	0	0.46	0.93
Mean ± SD GEC	0.11 ± 0.18	0.16 ± 0.23	0.27 ± 0.21	0.31 ± 0.22
Number of samples	166713	60766821	278102704	63682023
Mean ± SD CLC	0.14 ± 0.19	0	0.41 ± 0.14	0.45 ± 0.15
Number of samples	16744129	53095593	217257263	62818542
t-test's t	67.7	5422.8	8875.9	4186.7
Two-tailed p	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Below, the changes in the carbon balance in Finland predicted by JSBACH are inspected. Thus the results of the run with the Finnish HR CLC, European CLC and GlobCover combination based PFT distribution, are shown solely for the area of influence of the Finnish HR CLC. Average gross primary production (GPP) rates and their standard deviations (SD) for the 15-year period from 1997 to 2011 (Table 9) in Finland were quite similar among the three PFT distributions. However, the seasonal partitioning showed on average higher spring GPPs in the GLCC-GEC and Finnish HR CLC than in GlobCover, which were mostly compensated for by respective lower summer values. Winter values were generally negligible and autumn values were very similar among the three land cover data. The higher spring values can be attributed to higher proportion of coniferous evergreen species in the GLCC-GEC and Finnish HR CLC than in GlobCover. Generally the differences of yearly and seasonal average GPPs were not significant, but GlobCover-based spring GPP was significantly different from GPP of the other land-cover data (GlobCover *vs*. GLCC-GEC *t*-test: $t_{28} = 3.63$, two tailed $p = 0.001$; GlobCover *vs*. Finnish HR CLC *t*-test: $t_{28} = -5.69$, two tailed $p < 0.001$). Peltoniemi *et al*. (2014) showed that JSBACH predicts equally high season GPP rates for both deciduous and evergreen trees while the growing season start of deciduous PFT is delayed on average for approximately a month, resulting in a lower annual GPP. Our results show that coniferous evergreen species account for 86%, 79% and 84% of the average annual land ecosystem GPP, with GLCC-GEC, GlobCover and Finnish HR CLC, respectively.

Regionally, the largest change in 15-year average GPP between the GLCC-GEC and GlobCover took place in northern Finland and close to the northern part of the coast of the Gulf of Bothnia, displaying large areas of increased production (Fig. 3, left). In southern Finland GPP decreased, which is due to the increase of the proportion of deciduous broadleaf trees at the expense of the proportion of coniferous evergreen trees. The Finnish HR CLC run (Fig. 3, right) showed an increase in GPP mostly in the same areas in northern and western Finland as GlobCover runs. In southeastern Finland an increase in GPP took place, which is in accordance with a slight decrease in deciduous broadleaf trees.

The impact of land cover on the differences between the three land cover data sets was further examined by looking at the regional

Table 8. The LAI values given (LAI GEC) and statistics estimated from LAI-maps of VALERI-project and University of Helsinki for classes 21 Conifer Boreal Forest, 23 Cool Mixed Forest, 25 Cool Broadleaf Forest and 47 Dry Woody Shrub.

	Class 21	Class 23	Class 25	Class 47
LAI GEC	6.0	4.3	5.2	4.0
Mean ± SD VALERI	1.9 ± 0.7	1.8 ± 0.7	1.7 ± 0.7	1.2 ± 0.7
Number of samples	35304	10792	1549	4240
Mean ± SD UH	4.4 ± 1.3	4.1 ± 1.2	3.5 ± 1.3	2.7 ± 1.3
Number of samples	217061736	62757565	15719037	59956373
t-test's *t*	670.9	341.2	101.2	139.5
Two-tailed *p*	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 9. Mean ± SD yearly and seasonal [spring (March–May), summer (June–August), and autumn (September–November)] gross primary production (GPP) rates (g C m^{-2} a^{-1}) for a 15-year period from 1997 to 2011 modelled using different land cover (LC) data.

LC data	Whole year	Spring	Summer	Autumn
GLCC-GEC	510 ± 34	78 ± 8.9	371 ± 28	61 ± 5.3
GlobCover	515 ± 35	66 ± 8.7	386 ± 30	63 ± 5.7
Finnish HR CLC	522 ± 33	86 ± 9.7	374 ± 28	62 ± 5.3

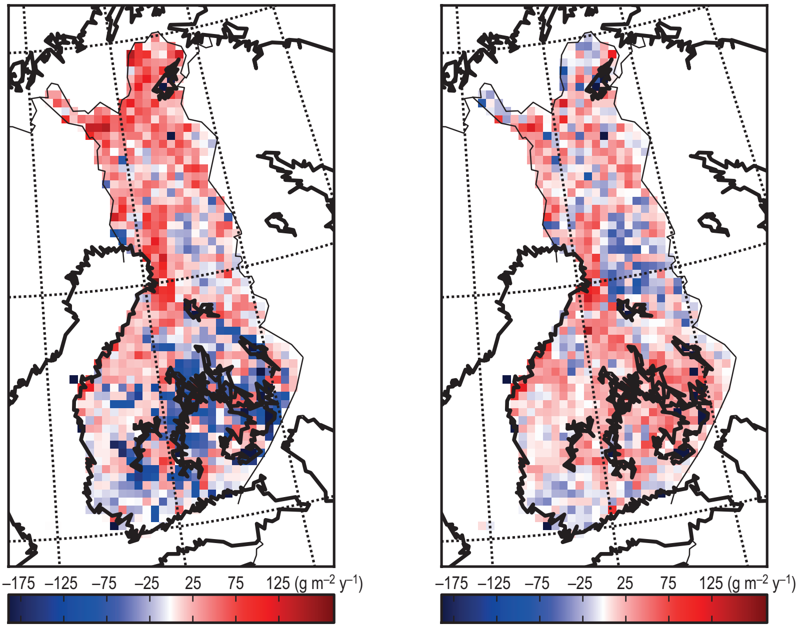

Fig. 3. Deviations of the mean annual GPP from JSBACH for the years 1997–2011 derived with two new PFT distributions from the GLCC-GEC based PFT distribution. Difference between GlobCover and the GLCC-GEC results (left) and difference between the Finnish HR CLC and GLCC-GEC results (right).

distributions of SDs of annual GPPs during for the years 1997–2011 (Fig. 4). The areas dominated by evergreen coniferous forests display the lowest inter-annual variability. Additionally, the fell areas in northern Finland, with little or no vegetation, show a small inter-annual variation. For the GLCC-GEC, the highest interannual variability was located in central and southern Finland, where the classification placed narrow conifers that are translated into coniferous deciduous and deciduous broadleaf trees (Table 4). Even though in the PFT map based on the Finnish HR CLC there are no grid cells with coniferous deciduous trees among dominant PFTs, it produced a very similar pattern of SD of the annual GPPs (Fig. 4, right) to the GLCC-GEC based PFT map (Fig. 4, left). The reason for this is that both deciduous PFTs have similar descriptions of phenology in the model. The same area shows locally the highest variability also with the GlobCover data set (Fig. 4, middle). These high SDs in the areas with relatively high proportions of deciduous broadleaf trees reflect the sensitivity of deciduous PFTs to the climatic conditions controlling the annual development of LAI and consequently the start of growing season. As the length of the growing season plays a major role in the annual carbon balance it is of pivotal importance to have correct proportions of evergreen and deciduous PFTs.

Discussion

So far, the REMO and the JSBACH model runs with the prescribed land-cover have used the GLCC-GEC as base land-cover data, with PFTs and surface parameters assigned to its classes. Unfortunately, the GLCC-GEC does not represent the Finnish landscape particularly well, and there are significant errors in the distribution of land-cover types. Technical weaknesses are its large pixel size and that it is based on the NOAA AVHRR time series from the early 1990s. There are 94 classes, but these do not represent Finland well because just three classes cover 69% of this area. When compared with the Finnish HR CLC, the proportion of coniferous forests is greater in the GLCC-GEC, and the proportion of other land-cover types lower. Due to its large pixel size, the proportion of dominant land-cover increases. The Finnish landscape is composed of small patches of different land cover. The effects of this can be seen in comparison; almost half of inland water in the GLCC-GEC is something other than water in the Finnish HR CLC,

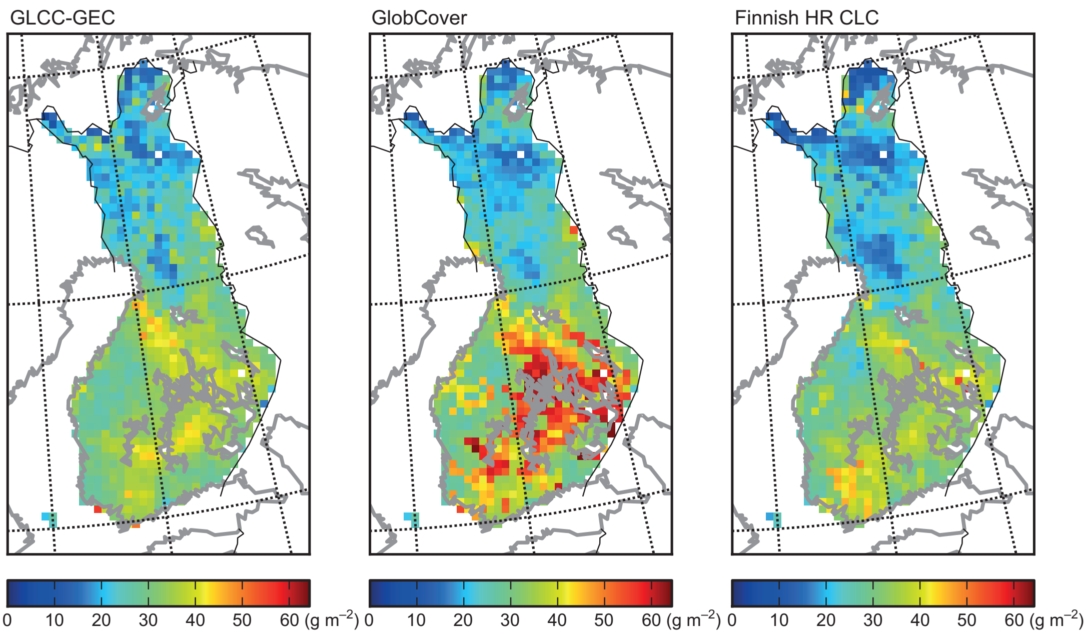

Fig. 4. Standard deviation of GPP for the years 1997–2011 from JSBACH with three different land cover data sets, GLCC-GEC (left), GlobCover (middle) and Finnish HR CLC (right).

and quite often agricultural areas in the Finnish HR CLC are classified as forest classes in the GLCC-GEC. One significant failure is that there are no open bogs, mires and marshes in the GLCC-GEC classification of Finland and these areas are classified mainly as forests.

Land-surface parameters, forest ratio and LAI allocated for the GEC-GLCC classes (Claussen *et al*. 1994, Hagemann *et al*. 1999, Hagemann 2002) were studied further. It was noted that the forest ratio of classes was usually much higher than the tree crown cover estimates calculated from moderate resolution satellite images. LAI of forest classes for the GLCC-GEC were also higher than estimates from high-resolution LAI maps. Allocation of a constant value for surface parameter for a land-cover category does not always correspond properly to the situation in the field, in particular where surface characteristics vary within a single land-cover category. Thus, land-cover categories should be further divided into subcategories, or preferably, whenever possible surface characteristics should be provided for the models as continuous fields.

Ultimately, it is the modelling environment which should be used to decide what is good land-cover data. Two revised land-cover data sets recoded into the GLCC-GEC nomenclature were produced covering the modelling window in Scandinavia and the surrounding areas. The first was based on the GlobCover classification, and the other on a combination of the GlobCover, European CLC and Finnish HR CLC classifications. In comparison with the Finnish HR CLC classification, GlobCover overestimates the proportions of forests and sparsely vegetated areas, whereas agricultural areas and shrubs in particular are heavily underestimated. One way to enhance GlobCover could be to study GlobCorine (Defourny *et al*. 2010) and find land-cover and land-use classes which are represented better in GlobCorine than in GlobCover. For example, agricultural areas, grasslands and heathlands could be taken from GlobCorine if they are classified more accurately than the corresponding areas in GlobCover. Compared with the GLCC-GEC, the benefit of GlobCover is that it includes wetlands, and they are relatively well located. One alternative to create new land-cover data would be to acquire estimates of land-surface parameters that are as good as possible, and find the closest GLCC-GEC class in the land-surface parameter table (Hagemann 2002). This was not done because it was not possible to estimate all the required land-surface parameters.

It was quite straightforward to recode the GlobCover and CLC classifications to the GLCC-GEC classes; the main difficulty was deciding on correspondence between classes when the descriptions of classes were not very precise. Unfortunately it is not possible to say which is the best for modelling purposes, but recently the CLC classification has been used in REMO instead of the GLCC-GEC by Gao *et al*. (2014) and in JSBACH by Peltoniemi *et al*. (2014).

Conclusions

For the JSBACH land-surface model, which uses the PFT presentation of vegetation distribution, it is crucial to have correct proportions of correctly-located land-cover classes. For the REMO climate model, which utilises prescribed parameter fields in its descriptions of surface processes, it is also important to have a suitable set of parameter values allocated to land-cover classes. The standard land-cover classification used with these models is based on old satellite images with poor spatial resolution and it represents the Finnish landscape quite poorly. Therefore, we created new land-cover classifications based on newer and more spatially-accurate land-cover classifications that should be used instead.

Although the differences between modelling results using different land-cover data were quite small, we still think that the land-cover data based on spatially, thematically and temporally more accurate land cover classifications should be used instead of the GLCC-GEC. We think that our lack of positive evidence is due to the lack of spatially explicit reference, or compensatory effects such as the decrease of spatial variability when integrating over larger area, or an erroneous response of the modelling system to land-cover data producing 'correct' results. When we use land-cover data which is as correct as possible, we will be prepared for improvements in modelling systems, such as enhancements in the descriptions and parameterizations of plant functional type specific processes and the increase of the spatial resolution of climate models.

Acknowledgements: The authors would like to thank the referee and editor for their constructive comments. This work was made possible by LIFE+ project SNOWCARBO ENV/FIN/000133. We thank Dr. Christian Reick and Dr. Stefan Hagemann from MPI-Meteorology and Dr. Sönke Zaehle from MPI-Biogeochemistry for providing the code and support for the JSBACH model.

References

Bartalev S., Belward A., Erchov D. & Isaev A. 2003. A new SPOT4-VEGETATION derived land cover map of Northern Eurasia. *Int. J. Remote Sensing* 24: 1977–1982.

Bicheron P., Defourny P., Brockmann C., Schouten L., Vancutsem C., Huc M., Bontemps S., Leroy M., Achard F., Herold M., Ranera F. & Arino O. 2008. *GLOBCOVER Products description and validation report*. MEDIAS-France. [Available at http://due.esrin.esa.int/globcover/LandCover_V2.2/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf].

Bonan G.B. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* 320: 1444–1449.

Bontemps S., Defourny P., Van Bogaert E., Arino O., Kalogirou V. & Perez J. 2011. *GLOBCOVER 2009 — products description and validation report*. University of Louvain and European Space Agency. [Available at http://due.esrin.esa.int/globcover/LandCover2009/GLOBCOVER2009_Validation_Report_2.2.pdf].

Bontemps S., Herold M., Kooistra L., van Groenestijn A., Hartley A., Arino O., Moreau I. & Defourny P. 2012. Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences* 9: 2145–2157.

Champeaux J., Masson V. & Chauvin F. 2005. ECOCLIMAP: a global database of land surface parameters at 1 km resolution. *Meteorol. Appl.* 12: 29–32.

Claussen M., Lohmann U., Roeckner E. & Schulzweida U. 1994. *A global dataset of land surface parameters*. Report 135, Max-Planck-Institute for Meteorology, Hamburg, Germany.

Dee D.P., Uppala S.M., Simmons A.J., Berrisford P., Poli P., Kobayashi S., Andrae U., Balmaseda M.A., Balsamo G., Bauer P., Bechtold P., Beljaars A.C.M., van de Berg L., Bidlot J., Bormann N., Delsol C., Dragani R., Fuentes M., Geer A.J., Haimberger L., Healy S.B., Hersbach H., Hólm E.V., Isaksen L., Kållberg P., Köhler M., Matricardi M., McNally A.P., Monge-Sanz B.M., Morcrette J.-J., Park B.-K., Peubey C., de Rosnay P., Tavolato C., Thépaut J.-N. & Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137: 553–597.

Defourny P., Bontemps S., van Bogaert E., Weber J.-L., Steenmans C. & Brodsky L. 2010. *GLOBCORINE 2009 — description and validation report*. University of

Leuven. [Available at http://dup.esrin.esa.it/files/p114/GLOBCORINE2009_DVR_2.1.pdf].

Di Gregorio A. & Jansen L. 2000. *Land Cover Classification System (LCCS): classification concepts and user manual*. Food and Agriculture Organization of the United Nations. [Available at http://www.fao.org/docrep/003/x0596e/x0596e00.htm].

FAO 1971–1981. *Soil map of the world*, vols. 1–10. Unesco, Paris.

Friedl M.A., Sulla-Menashe D., Tan B., Schneider A., Ramankutty N., Sibley A. & Huang X. 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114: 168–182.

Gao Y., Weiher S., Markkanen T., Pietikäinen J.-P., Gregow H., Henttonen H.M., Jacob D. & Laaksonen A. 2015: Implementation of the CORINE land use classification in the regional climate model REMO. *Boreal Env. Res.* 20: 261–282.

Giorgetta M., Jungclaus J., Reick C., Legutke, S., Bader J., Böttinger M., Brovkin V., Crueger T., Esch M., Fieg K., Glushak K., Gayler V., Haak H., Hollweg H-D., Ilyina T., Kinne S., Kornblueh L., Matei D., Mauritsen T., Mikolajewicz U., Mueller W., Notz D., Pithan F., Raddatz T., Rast S., Redler R., Roeckner E., Schmidt H., Schnur R., Segschneider J., Six K., Stockhause M., Timmreck C., Wegner J., Widmann H., Wieners K-H., Claussen M., Marotzke J. & Stevens B. 2013. Climate and carbon cycle changes from 1850 to 2100 in MPIESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.* 5: 572–597.

GLC2000 2010. *Global land cover 2000*. Joint Research Centre of the European Commission. [Available at http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php].

Hagemann S. 2002. *An improved land surface parameter dataset for global and regional climate models*. Report 336, Max-Planck-Institute for Meteorology, Hamburg, Germany.

Hagemann S. & Stacke T. 2013. Impact of the soil hydrology scheme on simulated soil moisture memory. *Clim. Dyn.*, doi:10.1007/s00382-014-2221-6.

Hagemann S., Botzet M., Dümenil L. & Machenhauer B. 1999. *Derivation of global GCM boundary conditions from 1 km land use satellite data*. Report 289, Max-Planck-Institute for Meteorology, Hamburg, Germany.

Heiskanen J., Rautiainen M., Korhonen L., Mõttus M. & Stenberg P. 2011a. Generating fine resolution leaf area index maps for boreal forests of Finland. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2011*, pp. 2326–2329.

Heiskanen J., Rautiainen M., Korhonen L., Mõttus M. & Stenberg P. 2011b. Retrieval of boreal forest LAI using a forest reflectance model and empirical regressions. *International Journal of Applied Earth Observation and Geoinformation* 13: 595–606.

Härmä P., Teiniranta R., Törmä M., Repo R., Järvenpää E. & Kallio E. 2005. *CLC2000 Finland: final report*. Geoinformatics and Land Use Division, Finnish Environment Institute. [Available at http://www.syke.fi/download/noname/%7BB4231601-439E-4903-B6FC-2F2FE718B465%7D/39571].

Jacob D. 2001. A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorology and Atmospheric Physics* 77: 61–73.

Jacob D. & Podzun R. 1997. Sensitivity studies with the regional climate model REMO. *Meteorology and Atmospheric Physics* 63: 119–129.

Jung M., Henkel K., Herold M. & Churkina G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment* 101: 534–553.

Loveland T.R., Reed B.C., Brown J.F., Ohlen D.O., Zhu Z., Yang L. & Merchant J.W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sensing* 21: 1303–1330.

Olson J.S. 1994a. *Global ecosystem framework 1. Definitions*. EROS Data Center Report 1 for the U.S. Geological Survey, Report GCP94.1, Global Patterns Company, Lenoir City, USA.

Olson J.S. 1994b. *Global Ecosystem Framework 2. Transiation Strategy*. EROS Data Center Report 2 for the U.S. Geological Survey, Report GCP94.2, Global Patterns Company, Lenoir City, USA.

Peltoniemi M., Markkanen T., Härkönen S., Muukkonen P., Thum T., Aalto T. & Mäkelä A. 2015: Consistent estimates of gross primary production of Finnish forests — comparison of estimates of two process models. *Boreal Env. Res.* 20: 196–212.

Rechid D. 2008. *On biogeophysical interactions between vegetation phenology and climate simulated over Europe*. Reports on Earth System Science 61, University of Hamburg. [Available at: http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEB_BzE_61.pdf].

Rechid D., Raddatz T.J. & Jacob D. 2009. Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling. *Theoretical and applied climatology* 95: 245–255.

Reick C.H., Raddatz T., Brovkin V. & Gayler V. 2013. Representation of natural and anthropogenic land cover change in MPI-ESM. *J. Adv. Model. Earth Syst.* 5: 459–482.

Simmons A.J. & Burridge D.M. 1981. An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinate. *Monthly Weather Review* 109: 758–766.

Stevens B., Giorgetta M., Esch M., Mauritsen T., Crueger T., Rast S., Salzmann M., Schmidt H., Bader J., Block K., Brokopf R., Fast I., Kinne S., Kornblueh L., Lohmann U., Pincus R., Reichler T. & Roeckner E. 2013. Atmospheric component of the MPI-M Earth System Model: ECHAM6. *J. Adv. Model. Earth Syst.* 5: 146–172.

Tomppo E. 2006. The Finnish Multisource National Forest Inventory — small area estimation and map production. In: Kangas A. & Maltamo M. (eds.), *Forest inventory — methodology and applications*, Managing Forest Ecosystems vol. 10, Springer, Dordrecht, pp. 195–224.

Törmä M., Härmä P., Teiniranta R., Repo R., Järvenpää E. & Kallio M. 2004. The production of Finnish Corine Land Cover 2000 classification. In: Altan O. (ed.), *XXth ISPRS Congress Technical Commission IV, ISPRS Archives vol. XXXV Part B4*, pp. 1330–1335.

Törmä M., Härmä P., Hatunen S., Teiniranta R., Kallio M. & Järvenpää E. 2011. Change detection for Finnish CORINE land cover classification. In: Michel U. & Civco D. (eds.), *Earth resources and environmental remote sensing/GIS applications II, Proceedings of SPIE* vol. 8181, SPIE, Bellingham, WA 2011: 81810Q.

U.S. Geological Survey 2001. *Global land cover characteristics data base version 2.0*. U.S. Geological Survey. [Available at http://edc2.usgs.gov/glcc/glcc.php].