Ecological Modelling 291 (2014) 152–174

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Spatial modeling of agricultural land use change at global scale

Prasanth Meiyappan[a,*], Michael Dalton[b], Brian C. O'Neill[c], Atul K. Jain[a,**]

[a] Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61801, USA
[b] Alaska Fisheries Science Center, National Ocean and Atmospheric Administration, National Marine Fisheries Service, Seattle, WA 98115, USA
[c] Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80307, USA

ARTICLE INFO

Article history:
Received 28 April 2014
Received in revised form 23 July 2014
Accepted 25 July 2014

Keywords:
Prediction
Drivers
Integrated Assessment
Spatially explicit
Validation
Land change

ABSTRACT

Long-term modeling of agricultural land use is central in global scale assessments of climate change, food security, biodiversity, and climate adaptation and mitigation policies. We present a global-scale dynamic land use allocation model and show that it can reproduce the broad spatial features of the past 100 years of evolution of cropland and pastureland patterns. The modeling approach integrates economic theory, observed land use history, and data on both socioeconomic and biophysical determinants of land use change, and estimates relationships using long-term historical data, thereby making it suitable for long-term projections. The underlying economic motivation is maximization of expected profits by hypothesized landowners within each grid cell. The model predicts fractional land use for cropland and pastureland within each grid cell based on socioeconomic and biophysical driving factors that change with time. The model explicitly incorporates the following key features: (1) land use competition, (2) spatial heterogeneity in the nature of driving factors across geographic regions, (3) spatial heterogeneity in the relative importance of driving factors and previous land use patterns in determining land use allocation, and (4) spatial and temporal autocorrelation in land use patterns.

We show that land use allocation approaches based solely on previous land use history (but disregarding the impact of driving factors), or those accounting for both land use history and driving factors by mechanistically fitting models for the spatial processes of land use change do not reproduce well long-term historical land use patterns. With an example application to the terrestrial carbon cycle, we show that such inaccuracies in land use allocation can translate into significant implications for global environmental assessments. The modeling approach and its evaluation provide an example that can be useful to the land use, Integrated Assessment, and the Earth system modeling communities.

1. Introduction

Changes in land use are driven by non-linear interactions between socioeconomic conditions (e.g. population, technology, and economy), biophysical characteristics of the land (e.g. soil, topography, and climate), and land use history (Lambin et al., 2001, 2003). The spatial heterogeneity in driving factors has led to spatially distinct land use patterns. Land use change models exploit techniques to understand the spatial relationship between historical changes in land use and its driving factors (or proxies for them). Such models are also used to project spatial changes in land use based on scenarios of changes in its drivers. The importance of land use change models is evident from the wide range of existing modeling approaches and applications (see reviews by NRC, 2014; Heistermann et al., 2006; Verburg et al., 2004; Parker et al., 2003; Agarwal et al., 2002; Irwin and Geoghegan, 2001; Briassoulis, 2000; U.S. EPA, 2000). However, most land use change models are designed for local to regional scale studies (typically sub-national to national level); global-scale modeling approaches are scarce (Rounsevell and Arneth, 2011; Heistermann et al., 2006).

Global-scale land use modeling is challenging compared to smaller-scale approaches for three main reasons. First, the set of driving factors and their spatial characteristics of change are diverse across the globe, and models need to represent this variability (van Asselen and Verburg, 2012). Second, the various factors that affect land use decisions operate at different spatial scales. For example, landowners make decisions at local scale, whereas factors like governance, institutions, and enforcement of property rights operate at much larger scales. Ideally, global-scale models should incorporate the effects of driving factors at multiple scales (Rounsevell et al.,

* Corresponding author at: 105 South Gregory Street, Atmospheric Sciences Building, Urbana, IL 61801, USA. Tel.: +1 217 898 1947.
** Corresponding author. Tel.: +1 217 333 2128.
E-mail addresses: meiyapp2@illinois.edu, prasanthnitt89@gmail.com (P. Meiyappan), jain1@illinois.edu (A.K. Jain).

http://dx.doi.org/10.1016/j.ecolmodel.2014.07.027

2014; Heistermann et al., 2006). However, an integrated understanding of how the multi-scale drivers combine to cause land use change is far from complete (Lambin et al., 2001; Meyfroidt, 2013). Third, spatially and temporally consistent data for many important driving factors (e.g. market influence) are not readily available at a global scale and at the required spatial resolution (Verburg et al., 2011, 2013).

Despite these challenges, there are three reasons for modeling land use at a global scale. First, several key drivers of land use (e.g. climate) and their impacts on land use have no regional demarcations and substantial feedback exists between them (Rounsevell et al., 2014). Addressing the feedback between land use and socioecological systems requires a globally consistent framework. Second, regions across the world are interconnected through global markets and trade that can shift supply responses to demands for land across geopolitical regions (Meyfroidt et al., 2013). Modeling such complex interactions among economies demands a global scale approach. Third, the aggregate consequences of land use at the global scale have significant consequences for climate change (Pielke et al., 2011), global biogeochemical cycles (Jain et al., 2013), water resources (Bennett et al., 2001) and biodiversity (Phalan et al., 2011), making global land use modeling a useful component of analyses of these issues.

These reasons have motivated global scale assessments using Integrated Assessment Models (IAMs) that seek to treat the interactions between land and other socioecological systems in a fully coupled manner (Sarofim and Reilly, 2011). In IAMs, socioeconomic models are coupled with biophysical models (process-based vegetation models and/or climate models) to translate socioeconomic scenarios into changes in land cover and its impacts on environmental variables of interest (van Vuuren et al., 2012). IAMs typically disaggregate the world into 14–24 regions (van Vuuren et al., 2011), and land use decisions are made at this regional scale. Some IAMs have spatially explicit biophysical components, and in these cases land use information on geographic grids at a much higher spatial resolution is required (typically $0.5^\circ \times 0.5^\circ$ lat/long). To provide this information, spatial land use allocation approaches are employed to downscale aggregate land demands for large world regions to individual grid cells. Examples of such global scale land use allocation approaches can be found in the Global Forest Model (Rokityanskiy et al., 2007), IMAGE (Bouwman et al., 2006), MagPie (Lotze-Campen et al., 2010), KLUM (Ronneberger et al., 2005, 2009), MIT-IGSM (Reilly et al., 2012; Wang, 2008), GLOBIO3 (Alkemade et al., 2009), GLOBIOM (Havlik et al., 2011), Nexus land use model (Souty et al., 2012, 2013), and the Global Land use Model (GLM) (Hurtt et al., 2011).

In this article, we develop a new global land use allocation model specifically to downscale agricultural (cropland and pastureland) land use from large world regions to the grid cell level. Agricultural land use merits special attention because it is associated with the majority of land use-related environmental consequences (Green et al., 2005), currently occupying ~40% of Earth's land area (Foley et al., 2005). There are two novel features of our approach that distinguish it from previous approaches.

First, our model predicts fractional land use within each grid cell (continuous field approach) driven by time-varying socioeconomic and biophysical factors. In contrast, most existing models do one or the other but not both. For example, many downscaling methods represent land use in each grid cell ($0.5^\circ \times 0.5^\circ$ lat/long or coarser) by the dominant land cover category (e.g. MagPie, IMAGE, GLOBIOM, and the Nexus land use model). This simplified representation in land cover underestimates land cover heterogeneity and is a major source of uncertainty in impact assessments (Verburg et al., 2013). Some recent efforts (e.g. Letourneau et al., 2012; Schaldach et al., 2011) have addressed this problem by increasing spatial resolution, for example using 5-min grid cells that represent dominant land cover types. While such approaches are an improvement, they are also much more computationally intensive and do not escape the problem that for many variables representing land use drivers, high resolution data at the global scale are unavailable (Verburg et al., 2013). In other approaches (e.g., GLOBIO3 and GLM) land cover is represented as fractional units within each grid cell (again $0.5^\circ \times 0.5^\circ$ lat/long), but the approach to allocation is overly simplified, proportionally allocating land use projections for aggregate regions to grid cells as closely as possible to existing land use patterns. Such an approach does not account for the effect of changes over time in land use drivers, which can lead to land use projections that are inconsistent with those drivers (as will be shown later).

Second, we carry out the first global scale evaluation of a spatial land use allocation model over a long historical period (>100 years), reproducing the broad spatial features of the long-term evolution of agricultural land use patterns. Evaluation of global-scale spatial land use models is important because they are used to generate scenarios for 50–100 years into the future, for example, to explore issues related to greenhouse gas emissions and mitigation possibilities (Moss et al., 2010; Kindermann et al., 2008), climate change impacts on ecosystems (MEA, 2005; UNEP, 2012), biodiversity (TEEB, 2010; Pereira et al., 2010), or adaptation options involving land use (OECD, 2012; Phalan et al., 2011). While evaluation of model performance over the past 100 years is no guarantee of good performance over the next 100 years, demonstrating the ability of a model to reproduce long-term historical patterns increases confidence in its suitability for application to long-term scenarios of future change. The model evaluation presented here could serve as an example for how evaluation of other downscaling methodologies could be carried out (O'Neill and Verburg, 2012; Hibbard et al., 2010).

2. Methods and data

2.1. Overview of the approach

Our land use allocation model simulates the spatial and temporal development of cropland and pastureland at a spatial resolution of $0.5^\circ \times 0.5^\circ$ lat/long and at an annual time-step. The model operates at two different spatial levels. On the regional level, the aggregate regional demand for cropland and pastureland is provided as input to the model. The model then allocates this demand to individual grid cells within that region. We use a constrained optimization technique to allocate a fraction of each grid cell to cropland and pastureland while meeting the aggregate regional demand for each type of land. The optimization technique selects the most profitable land to grow crops and pasture based on (1) the suitability of each grid cell for crop or pasture production, determined by a set of 46 biophysical and socioeconomic factors (Table 1), (2) historical land use patterns (temporal autocorrelation) and (3) the land use predicted for neighboring grid cells (spatial autocorrelation).

A primary intended application of this model is as one component of a larger modeling framework that includes a global, regionally resolved economic model that generates scenarios of future demand for land at the regional level, similar to the approach taken in other IAMs or land use models as discussed above. However, the main aim of this paper is to present and evaluate our model in a historical simulation against 20th century gridded data of cropland and pastureland. Ideally, the model should be evaluated against observational data. However, purely observational data for global, spatially resolved land use data do not exist. Rather, existing gridded land use reconstructions are modeled estimates that draw on national and sub-national data to the extent possible (see Appendix A). For practical purposes, we assume existing land use

Table 1
List of the 46 potential explanatory factors used in the regression analysis. The explanatory factors cover the time period 1901–2005 at annual resolution. The spatial resolution is 0.5° × 0.5° lat/long. Each seasonally averaged explanatory factor translates into four explanatory variables in our analysis (one for each season: spring, summer, fall and winter).

Broad category	Explanatory factor	Unit
Climate	Seasonally averaged temperature	K
	Seasonally averaged precipitation	mm/day
	Seasonally averaged potential evapotranspiration (PET)	mm/day
	Squared seasonally averaged temperature	K^2
	Squared seasonally averaged precipitation	mm^2/day^2
	Squared seasonally averaged PET	mm^2/day^2
	Seasonal Temperature Humidity Index (THI)	°C
Climate variability	Seasonal Palmer Drought Severity Index (PDSI)	(–)
	Heat wave duration index	No of days
	Simple Daily Precipitation Intensity Index	mm/day
Soil characteristics	Rooting conditions and nutrient retention capacity Nutrient availability Oxygen availability Chemical composition (indicates toxicities, salinity and sodicity) Workability (indicates texture, clay mineralogy and soil bulk-density)	(–)
Terrain characteristics	Elevation, altitude and slope combined	
Socioeconomic	Built-up/urban land area	Fraction of grid area (m^2/m^2)
	Urban population density	Inhabitants/km^2
	Rural population density	
	Rate of change in rural population density	Inhabitants/km^2/yr
	Rate of change in urban population density	
	Market Influence Index	International dollars/person

reconstructions represent the "truth" for the purpose of our model calibration and evaluation. In this respect, we face the same limitations as all global, spatially resolved models of land use, which typically use such reconstructions as maps representing current land use as a basis for future projections (e.g., Hurtt et al., 2011). However, it must be kept in mind that such reconstructions are estimates, and that estimates can and do differ from one another based on different uses of the underlying data and methods for estimation of gridded outcomes.

We break down the overall approach discussed above into three components for explanatory purpose. First, we formulate a land use allocation model based on profit maximization using mathematical programming methods for constrained optimization with respect to spatial and temporal distribution of land use types. Second, we derive an estimation procedure for the unknown parameters in the land use allocation model. The estimation procedure accounts for the heterogeneous nature and importance of driving factors across geographic regions. Finally, we evaluate the land use allocation model and estimation procedure using a historical global cropland and pastureland dataset.

2.2. The land use allocation model

2.2.1. Theoretical framework

The economic motivation for our land use allocation model is maximization of expected profit by hypothesized landowners within each grid cell (Lubowski et al., 2008). This motivation is consistent with the structure of most IAMs at the regional scale, which generally assume some form of optimization for economic sectors (e.g. profit maximization, cost minimization) that generate aggregate demand for land. We formulate the land use allocation model as a dynamic profit maximization function that consists of two components: a static profit maximization function and a dynamic adjustment cost model. The static profit maximization function maximizes the achievable profit within each grid cell by selecting the most productive land for growing crops and pastures. The dynamic adjustment cost model accounts for the adjustment cost associated with changes in land use patterns over time (Golub et al., 2008). For example, expanding cropland into unmanaged ecosystems would entail some cost to clear unmanaged land and build roads and other infrastructure. This adjustment cost tends to create inertia in land use patterns over time.

For ease of understanding, we introduce the model component-wise. The components are then eventually combined to form the final land use allocation model. In this description of the theoretical basis of the model, we differentiate two broad categories of land: managed and unmanaged. Managed land is the sum of cropland and pastureland, and we define all other land types as unmanaged. Because our focus is on cropland and pastureland, all equations we describe in this (and the next) section refer to managed land area. We account for unmanaged land later in the estimation procedure.

The static profit function for each grid cell 'g' is expressed as:

$$\underset{\{Y_{lg}^t\}}{Maximize} \sum_{l=1}^{2} (P_{lg}^t - W_{lg}^t) Y_{lg}^t - R_{lg} (Y_{lg}^t)^2 \quad (1)$$

In Eq. (1), Y_{lg}^t represents the area to be estimated of land use type 'l' (=1 for cropland, and 2 for pastureland) in grid cell 'g' at time step 't'. P_{lg}^t denotes the price per unit area for commodities produced by land use activity 'l' and W_{lg}^t represents the cost per unit area for producing those commodities. The linear term $(P_{lg}^t - W_{lg}^t)Y_{lg}^t$ represents the 'net profit' for each grid cell, which can be thought of as a measure of land suitability for land use activity 'l'. The second term $-R_{lg}(Y_{lg}^t)^2$ represents the non-linear cost associated with decreasing returns to scale; i.e. output increases less than proportionately to an increase in inputs (land use area) and the rate of increase in output decreases progressively with additional inputs. The non-linear cost term is included because land profitability is assumed to vary within each grid cell, and the most profitable land is used first. Therefore, in the long run, the profitability of each additional hectare of land brought under production within a grid cell declines (Gouel and Hertel, 2006). $R_{lg}(>0)$ is a productivity/returns constant and is a function of land use type 'l' and location 'g'.

Eq. (1) is considered a 'static' profit function because all variables are based on the current time step 't' with no reference to history.

Eq. (1) can be simplified as a quadratic function (see Appendix E.1 for detailed steps):

$$\underset{\left\{Y_{lg}^{t}\right\}}{Maximize}\sum_{l=1}^{2}(-R_{lg})(Y_{lg}^{t}-d_{lg}S_{lg}^{t})^{2} \tag{2}$$

In Eq. (2), $d_{lg}=(1/2R_{lg})$ is a constant for decreasing returns to scale and S_{lg}^{t} (suitability) equals the net price term $P_{lg}^{t}-W_{lg}^{t}$. Eq. (2) is subject to two constraints:

$$Y_{lg}^{t}\geq 0 \tag{3}$$

$$\sum_{l=1}^{2}Y_{lg}^{t}\leq GA_{g} \tag{4}$$

Eq. (3) avoids negative allocations. Eq. (4) implies the total of cropland and pastureland area allocated within each grid cell 'g' should not exceed the grid cell area GA_g.

We formulate a dynamic adjustment cost model for each grid cell 'g' as follows:

$$\underset{\left\{Y_{lg}^{t}\right\}}{Minimize}\sum_{l=1}^{2}Q_{lg}(Y_{lg}^{t}-Y_{lg}^{\bar{t}})^{2} \tag{5}$$

Eq. (5) is also constrained by Eqs. (3) and (4). Eq. (5) represents a constrained least-squares optimization that tends to minimize the adjustment cost by minimizing the changes in land use allocation between the current and a previous time step $\bar{t}(\bar{t}<t)$. Criteria for selecting the value of $\bar{t}$ are explained in Section 2.5. $Q_{lg}(>0)$ is a constant that indicates the adjustment cost per unit area and is a function of land use type 'l' and location 'g'. In Eq. (5) we assume an exponent of 2 because: (1) our land use allocation method is based on quadratic programming, and (2) our model parameter estimation involves differentiating the quadratic program (Section 2.3), which results in linear equations that are convenient to solve.

We combine Eqs. (2) and (5), to write the overall objective function as a minimization problem for each grid cell 'g':

$$\underset{\left\{Y_{lg}^{t}\right\}}{Minimize}\sum_{l=1}^{2}\left[Q_{lg}\left(\frac{Y_{lg}^{t}}{GA_{g}}-\frac{Y_{lg}^{\bar{t}}}{GA_{g}}\right)^{2}+R_{lg}\left(\frac{Y_{lg}^{t}}{GA_{g}}-d_{lg}\frac{S_{lg}^{t}}{GA_{g}}\right)^{2}\right] \tag{6}$$

For convenience we have divided Eq. (6) by a constant that equals the square of grid cell area. The minimization problem is unaffected by this modification. Later, it will become evident that treating variables as fractions instead of areas is convenient in the parameter estimation procedure.

Our aim is to allocate the aggregate land demand among the grid cells within a given region such that the total profits are maximized. Therefore, we stack the individual grid cell level optimizations (Eq. (6)) over the aggregate region and write in matrix notation:

$$\underset{\{Y^{t}\}}{Minimize}\left(Y^{t}-Y^{\bar{t}}\right)'A\left(Y^{t}-Y^{\bar{t}}\right)+\left(Y^{t}-DS^{t}\right)'\left(Y^{t}-DS^{t}\right) \tag{7}$$

In Eq. (7), primes denote the matrix transpose operator. Y^t is a column vector of size '$2N\times 1$' with elements (Y_{lg}^{t}/GA_{g}), i.e. $Y^{t}=\begin{bmatrix}Y_{11}^{t}/GA_{1} & Y_{21}^{t}/GA_{1} & \cdots & Y_{1N}^{t}/GA_{N} & Y_{2N}^{t}/GA_{N}\end{bmatrix}'$, where 'N' represents the total number of grid cells within the aggregate region. Therefore, elements in vector Y^t are normalized by the grid cell area and will therefore range from zero to one. Similarly $Y^{\bar{t}}$, d, and S^t are vectors of $Y_{lg}^{\bar{t}}/GA_{g}$, d_{lg}, and S_{lg}^{t}/GA_{g}, respectively. The term **D** is a diagonal matrix of size '$2N\times 2N$' given by

$$\mathbf{D}=d*\mathbf{I}_{2N\times 2N}$$

where **I** is an identity matrix. The term **A** represents a constant diagonal matrix.

$$A=\begin{bmatrix}\frac{Q_{11}}{R_{11}} & 0 & \cdots & \cdots & 0\\ 0 & \frac{Q_{21}}{R_{21}} & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \frac{Q_{1N}}{R_{1N}} & 0\\ 0 & \cdots & \cdots & 0 & \frac{Q_{2N}}{R_{2N}}\end{bmatrix}_{2N\times 2N}.$$

The matrix **A** can be usefully interpreted as representing the balance between the importance of adjustment costs and of land suitability in determining land allocation across grid cells. Eq. (7) can be regarded as a balance between a dynamic (time-series aspect) and a static (cross-sectional aspect) term. The dynamic term is implicitly minimizing adjustment costs by trying to keep land use similar to the historical (already existing) land use patterns, whereas the static term selects the most suitable land to maximize the net profit regardless of history. The balance between the static and dynamic term is determined by the values of the matrix **A**. In extreme cases, when R_{lg} is zero, no explicit account is taken of the relative suitability of land across grid cells and outcomes are determined entirely by land use history; when Q_{lg} is zero, no account is taken of past land use patterns and outcomes are determined entirely by suitability across grid cells.

We simplify the matrix **A** by assuming the diagonal elements are equal (i.e. Q_{lg}/R_{lg} is same for all 'l' and 'g'). Hence $\mathbf{A}=a\mathbf{I}$, where 'a' is a positive scalar $(=(Q_{lg}/R_{lg}))$ and **I** is an identity matrix of size '$2N\times 2N$'. Substantively, this implies that while adjustment costs and land suitability can vary across grid cells, their relative importance to land allocation decisions is held fixed across grid cells within a given region. This is not an unreasonable assumption and a minor concession given its practical benefits: it is both unrealistic and undesirable to estimate Q_{lg}/R_{lg} for each 'l' and 'g'. It is unrealistic because the number of unknown parameters will increase with the number of grid cells resulting in the incidental parameters problem (see Lancaster, 2000). It is undesirable because the historical data for land use and its driving factors available to constrain the model is limited by both availability and grid level accuracy (see Appendix A).

Eq. (7) is quadratic in the Y^t vector and is subject to two grid cell area constraints (Eqs. (8) and (9)) that are the vector forms of Eqs. (3) and (4), respectively.

$$Y^{t}\geq 0 \tag{8}$$

$$(Y_{2g-1}^{t}+Y_{2g}^{t})\leq 1\quad \forall g\in[1,N] \tag{9}$$

$$\sum_{g=1}^{N}GA_{g}Y_{2g-1}^{t}=\text{regional area demand for cropland} \tag{10}$$

$$\sum_{g=1}^{N}GA_{g}Y_{2g}^{t}=\text{regional area demand for pastureland} \tag{11}$$

Eq. (7) is also subject to regional scale constraints (Eqs. (10) and (11)) that ensure that aggregate regional demand for each land use activity is equal to the total grid cell allocations of that land use activity within that region. Therefore, Eq. (7) is a quadratic program. Competition between land use types is accounted for in Eq. (7) because the profits within each grid cell are maximized by simultaneously weighing both the cropland and pastureland benefits. The first term (adjustment cost) in Eq. (7) accounts for temporal

autocorrelation in land use datasets. In the following section, we update Eq. (7) to account for spatial autocorrelation.

2.2.2. Accounting for spatial autocorrelation

Land use area in a grid cell tends to be more similar to the values at surrounding grid cells than to those farther away, a feature known as spatial autocorrelation (Overmars et al., 2003). When spatial autocorrelation is not accounted for, we violate a key assumption in statistical analysis that the residuals are independent and identically distributed (Dormann et al., 2007). We account for spatial autocorrelation by introducing a spatial weight matrix with the neighborhood size and weighting scheme selected based on trial and error (Augustin et al., 1996). We chose the neighborhood region to be the surrounding eight grid cells (first-order Moore's neighborhood) all with equal weight.

The land use allocation model (Eq. (7)) with a spatial weights matrix **B** for spatial autocorrelation is:

$$\underset{\{Y^t\}}{Minimize} \quad \left(Y^t - Y^{\bar{t}}\right)'(A+B)(Y^t - Y^{\bar{t}}) + \left(Y^t - DS^t\right)'(Y^t - DS^t) \tag{12}$$

The **B** matrix is proportional to a **W** matrix of spatial weights that is assumed to be symmetric and have zeros along its main diagonal. Note the **A** matrix is diagonal. We represent the spatial weights in **W** matrix by a constant scalar '*b*' that can be positive (if positively correlated), negative (negatively correlated), or zero (uncorrelated). We assume zero spatial autocorrelation between the two land use activities for a practical benefit: the resulting $(A+B)$ matrix structure allows us to use specialized techniques to perform matrix inversions quickly that are required for estimating model parameters (described in Section 2.3), which otherwise is computationally expensive. We provide some examples in Appendix E.2 to help illustrate the structure of matrices **A** and **B**. For grid cells lying along political boundaries, slight deviations in averaging could arise due to edge effects.

Eq. (12) is our final land use allocation model and is subject to two grid level constraints (Eqs. (8) and (9)) and two regional constrains (Eqs. (10) and (11)). There are four unknown components in Eq. (12) that need to be estimated from historical data: the potential land suitability vector S^t, the scalar constants '*a*' and '*b*', and the constant vector for decreasing returns '*d*'. For consistency, we estimate all the unknown parameters simultaneously using the following procedure.

2.3. Estimation method for unknown parameters

Consistent estimates for the parameters in Eq. (12) can be obtained with historical data for land use and its driving factors by treating Eq. (12) as a least-squares problem that combines first-order autoregressive stochastic processes (for first-order spatial autocorrelation and dynamic adjustment costs) and a logit function (with explanatory factors) for the term S^t. A restriction on the error process for each grid cell ensures that the sum of Y^t elements for each grid cell ($i.e.(Y^t_{1g}/GA_g) + (Y^t_{2g}/GA_g)$) is bounded between zero and one, or they may take a value of zero or one.

2.3.1. Logit function

We assume S^t (dependent variable) to be a function of a matrix X_{gt} of potential driving factors (exogenous explanatory variables; see Table 1 and discussed in Section 2.6) that is specific to grid cell '*g*' and time '*t*'. The matrix X_{g0} (i.e. for $t=0$) refers to potential driving factors that are time-stationary (e.g. soil and terrain conditions). We model the relationship between the dependent and explanatory variables as a binomial logistic regression (see Lesschen et al. (2005) for regression approaches used in spatial land use models). For each grid cell '*g*' and time '*t*', the logit functions for cropland ($l=1$) and pastureland ($l=2$) are given by Eqs. (13) and (14).

$$S^t_{1g} = \frac{1}{1+e^{\beta_0 + X'_{gt}\beta}} \tag{13}$$

$$S^t_{2g} = \frac{e^{\beta_0 + X'_{gt}\beta}}{1+e^{\beta_0 + X'_{gt}\beta}} \tag{14}$$

In Eqs. (13) and (14), β_0 is a constant coefficient and β is a vector of coefficients with a component for each explanatory variable. β_0 and β need to be estimated. The sum of Eqs. (13) and (14) implies the index of land suitability summed for cropland and pastureland for each grid cell equals one (recall that these equations apply only to managed land). Therefore, Eqs. (13) and (14) can be interpreted to partition the total land use area in grid cell '*g*' as proportions of cropland and pastureland.

2.3.2. Error process

Formally, the unconstrained version of the minimization problem in Eq. (12) implies a set of first-order necessary conditions. Therefore, we differentiate Eq. (12) and equate to zero:

$$(A+B)(Y^t - Y^{\bar{t}}) + \left(Y^t - DS^t\right)' = 0 \Leftrightarrow Y^t$$
$$= (I+A+B)^{-1}DS^t + (I+A+B)^{-1}(A+B)Y^{\bar{t}} \tag{15}$$

In Eq. (15), **I** is an identity matrix. Let $\mathbf{\Psi} = (\mathbf{I}+\mathbf{A}+\mathbf{B})^{-1}$, $\mathbf{\Omega} = \mathbf{\Psi}(\mathbf{A}+\mathbf{B})$, and ε^t_{lg} denote random variables, each with a mean zero that satisfy $\varepsilon^t_{1g} + \varepsilon^t_{2g} = 0$. Non-linear regression equations associated with Eq. (15) are

$$Y^t_{1g} = \sum_{k=1}^{N} \Psi_{2g-1,2k-1} d_{2k-1} \frac{1}{1+e^{\beta_0 + X'_{2k-1,t}\beta}} + \sum_{k=1}^{N} \Omega_{2g-1,2k-1} Y^{\bar{t}}_{1k} + \varepsilon^t_{1g} \tag{16}$$

$$Y^t_{2g} = \sum_{k=1}^{N} \Psi_{2g,2k} d_{2k-1} \frac{e^{\beta_0 + X'_{2k-1,t}\beta}}{1+e^{\beta_0 + X'_{2k-1,t}\beta}} + \sum_{k=1}^{N} \Omega_{2g,2k} Y^{\bar{t}}_{2k} + \varepsilon^t_{2g} \tag{17}$$

However in the above formulation, estimation of the '*d*' parameters (which, as noted above, reflect returns to scale) for each grid cell is inconsistent because of incidental parameter bias (Lancaster, 2000). For consistency, we treat the '*d*' parameters as random effects. For random effects, a logit function that differentiates managed (crop + pasture) from unmanaged land (e.g. forests, grasslands, and bare land that occupy rest of the grid cell area) is a natural specification that builds on a nested logit structure.

The '*d*' parameter corresponding to managed land fraction is:

$$d_g = \frac{1}{1+e^{\gamma_0 + X'_{g0}\gamma}} \tag{18}$$

In Eq. (18), γ_0 is a constant coefficient and γ is a vector of coefficients to be estimated for the set of explanatory variables specified by X_{g0}.

Substituting Eq. (18) into Eqs. (16) and (17) gives Eqs. (19) and (20), respectively.

$$Y^t_{1g} = \sum_{k=1}^{N} \Psi_{2g-1,2k-1} \frac{1}{1+e^{\gamma_0 + X'_{2k-1,0}\gamma}} \frac{1}{1+e^{\beta_0 + X'_{2k-1,t}\beta}} + \sum_{k=1}^{N} \Omega_{2g-1,2k-1} Y^{\bar{t}}_{1k} + \varepsilon^t_{1g} \tag{19}$$

$$Y^t_{2g} = \sum_{k=1}^{N} \Psi_{2g,2k} \frac{1}{1+e^{\gamma_0+X'_{2k-1,0}\gamma}} \frac{e^{\beta_0+X'_{2k-1,t}\beta}}{1+e^{\beta_0+X'_{2k-1,t}\beta}} + \sum_{k=1}^{N} \Omega_{2g,2k} Y^{\bar{t}}_{2k} + \varepsilon^t_{2g} \quad (20)$$

An important remark is that β and γ are not identified in Eqs. (19) and (20) unless $X_{g0} \neq X_{gt}$ for some 't'. We therefore specify the explanatory variables that are time-stationary (factors such as soil and terrain conditions) within X_{g0} and the transient explanatory variables (e.g. climate and socioeconomics) within $\mathbf{X}_{gt}$.

In Eqs. (19) and (20), for each grid cell 'g' the logit function for the 'd' parameter accounts for the fraction of managed land area, whereas the logits for 'S^t_{1g}' and 'S^t_{2g}' further splits the managed land fraction into proportions of cropland and pastureland, respectively.

A third equation in this system applies to unmanaged land fractions, which are the random effects:

$$1 - (Y^t_{1g} + Y^t_{2g}) = \frac{e^{\gamma_0+X'_{g,0}\gamma}}{1+e^{\gamma_0+X'_{g,0}\gamma}} + \eta_g \quad (21)$$

Eq. (21) implies that we can deduce the unmanaged land fraction using Eqs. (19) and (20). This is implied from the assumption built into the model's error process such that the sum of managed and unmanaged land fractions adds up to one for each grid cell. Therefore, Eq. (21) is redundant and dropped from the estimation.

In general for a spatial-weights matrix $\mathbf{W}$, a large number of grid cells implies the components of both $\Psi(a,b)$ and $\Omega(a,b)$ are high-order rational polynomials in powers of parameters 'a' and 'b'. These are derived from the functions $\Psi(a,b) = ((a+1)\mathbf{I} + b\mathbf{W})^{-1}$, and $\Omega(a,b) = \Psi(a,b)(a\mathbf{I} + b\mathbf{W})$, which are applied to form a stacked system of regression equations where 'S^t' and 'd' parameters are logit functions of a vector of explanatory variables as discussed above. Eqs. (19) and (20) can be combined as:

$$Y^t = \Psi(a,b)\mathbf{D}(\mathbf{X}'_0\gamma)S^t(\mathbf{X}'_t\beta) + \Omega(a,b)Y^{\bar{t}} + \varepsilon^t \quad (22)$$

In Eq. (22), $\mathbf{D} = d(\mathbf{X}'_\mathbf{0}\gamma) * \mathbf{I}_{2N\times 2N}$ where $\mathbf{I}$ is an identity matrix.

2.3.3. Least-square estimation

The least-squares problem for the nonlinear regression to estimate the parameters 'β_0', 'β', 'γ_0', 'γ', 'a', and 'b' is obtained by minimizing ε^t in Eq. (22).

$$\underset{\{\beta,\gamma,a,b\}}{Minimize} \quad \sum_t \left(Y^t - (\Psi(a,b)\mathbf{D}(X'_0\gamma)S^t(X'_t\beta) + \Omega(a,b)Y^{\bar{t}})\right)' \left(Y^t - (\Psi(a,b)\mathbf{D}(X'_0\gamma)S^t(X'_t\beta) + \Omega(a,b)Y^{\bar{t}})\right) \quad (23)$$

The summation over 't' in Eq. (23) implies multiple years of data can be used to estimate parameters. There are no grid level or regional demand constraints imposed on Eq. (23). The only constraint imposed on Eq. (23) is that $a > 0$.

Solving for estimates of 'β_0', 'β', 'γ_0', 'γ', 'a', and 'b' using Eq. (23) can be difficult when there are many grid cells, due to the numerical costs of inverting a large matrix to compute $\Psi(a,b)$ for each iteration in the estimation procedure. Therefore, we use specialized techniques to invert the matrix efficiently, the details of which are spelled out in Appendix E.3 with examples.

2.4. Accounting for spatial heterogeneity in driving factors

The set of driving factors and their relative importance (i.e. values of 'β' and 'γ' for a given explanatory variable) often differ between geographic regions. Further, the strength of temporal and spatial autocorrelation (i.e. 'a' and 'b' parameters) may vary between geographic regions. To account for this spatial heterogeneity, we disaggregate the world into 127 distinct sub-regions (Fig. 1) based on administrative boundaries (see Appendix B for methods and rationale) and solve Eq. (23) separately for each sub-region. Though we had earlier assumed that 'a' and 'b' parameters do not vary across grid cells within a region, estimating Eq. (23) for the 127 sub-regions imply 'a' and 'b' parameters can vary across sub-regions.

2.5. Selecting lag-year ($\bar{t}$) associated with the dynamic adjustment cost term

There are two main considerations for selecting a value for $\bar{t}$:

1. In general, the value of $Y^t - Y^{\bar{t}}$ should not be negligible relative to the value $Y^{\bar{t}}$ for most grid cells. If such were the case, the least-square optimization (Eq. (23)) would tend to be biased toward the dynamic adjustment cost term ($R_{lg} \to 0$). An exception applies to grid cells that are completely unsuitable for both cropland and pastureland where $Y^t_{lg} = Y^{\bar{t}}_{lg} = 0$ for any 't'. Typically, for global land use change datasets at $0.5^\circ \times 0.5^\circ$ lat/long, the grid cell level net changes in land use fractions between consecutive years is less than 10^{-3} for both cropland and pastureland when averaged globally over the 20th century (excluding grid cells unsuitable for agriculture and computed based on the land use change data described in Section 2.6). Therefore, a lag on the order of one year ($\bar{t} = t-1$) is not an appropriate choice.
2. Our specification of random errors in Eqs. (19) and (20) imply that they are uncorrelated with the explanatory variables. For this assumption to be valid, the value of the lag should be sufficiently large.

Based on experimentation, we found a lag of 10 years ($\bar{t} = t - 10$) satisfies the above requirements and also roughly matches the temporal autocorrelation present in the historical land use change dataset (Section 2.6).

2.6. Explanatory variables and land use change data

We include a total of 46 variables as potential explanatory variables (or proxies for them) in our regression analysis (Table 1). These variables are restricted to those expected to determine the spatial (as opposed to aggregate) determinants of land use patterns and they broadly align with our existing knowledge of land use dynamics (Lambin et al., 2001, 2003). At the global scale, the factors listed in Table 1 are adequate to describe the major spatial patterns of agricultural land use (Ramankutty et al., 2002). However, this list is not exhaustive. For example, policies that would influence spatial land use patterns within a region are likely relevant but are not explicitly included here. Rather, their effect (present in historical land use patterns) would be captured only implicitly through proxy variables. In cases like this we also rely on the fact that such factors are incorporated at least at the level of aggregate regions in scenarios generated by IAMs.

We synthesize the information for the 46 explanatory variables from a wide range of sources (Table 2). The data for each of the explanatory variables was either available for the time period 1900–2005 (annually) at a spatial resolution of $0.5^\circ \times 0.5^\circ$ lat/long, or was available for shorter time periods and/or coarser resolutions and we extended/refined it to a common time period and resolution. The rationale for selecting these variables and methodologies applied to extend/refine the raw data are detailed in Appendix C.

Historical reconstructions of cropland and pastureland were obtained from Ramankutty (2012) (hereafter referred as RF because

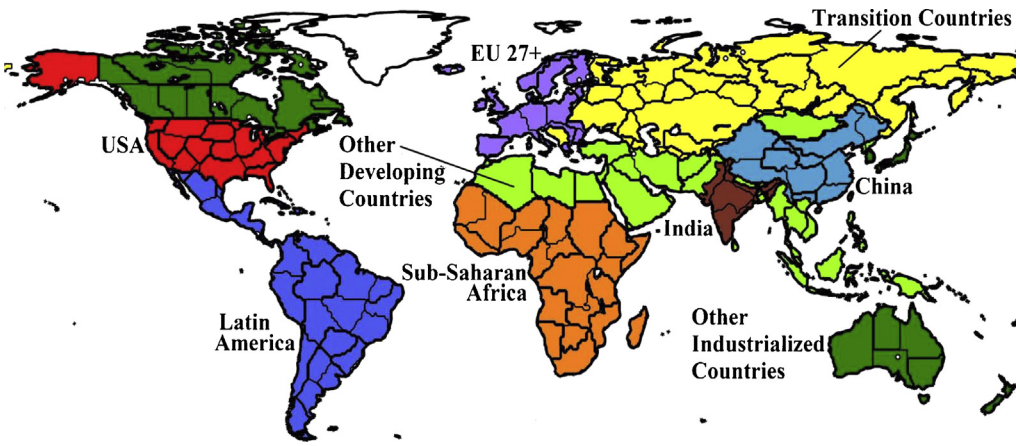

Fig. 1. The nine aggregate world regions used in this study (indicated by different colors). Of these, three are individual countries (US, China, and India) and the other six are aggregated regions. The spatial data for annual cropland and pastureland from historical reconstruction (RF data), were aggregated to these nine regions and used as regional demand constraint to the land use allocation model. The partitioning of the globe into 127 sub-regions is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

it is an updated version of Ramankutty and Foley (1999) data). The reconstruction is available yearly (1700–2007) at 0.5° × 0.5° spatial resolution; we utilize data for the period 1891–2005 for historical model simulation, of which ~20% subset is used for model estimation as explained below. In principle, the model can be estimated and evaluated with any historical reconstruction available in literature. The rationale for selecting RF data for our study and further background information on the land use change dataset is provided in Appendix A.

To estimate the parameters 'β_0', 'β', 'γ_0', 'γ', 'a', and 'b' we select 20 years of RF data for land use (Y^t) and its explanatory variables (X_{gt}) over the 1895–2005 period; i.e. about 20% of the available (annual) data. In selecting particular years to use in estimating parameters, we balance two goals: capturing recent patterns of

Table 2
Data sources used to derive the explanatory variables for model evaluation.

Category	Data variable	Description/units	Spatial characteristics	Period of availability	Source
Climate	Temperature (T_a) Daily average maximum temperature (T_{max})	°C	0.5° (lat/long)	1901–2012 (monthly)	Harris et al. (2013)
	Potential evapotranspiration Precipitation	Millimeters			
	Wet day frequency	Days			
	Palmer Drought Severity Index (PDSI)	No units	2.5°[a] (lat/long)	1870–2010 (monthly)	Dai (2011a,b)
Soil constraints	Rooting conditions and nutrient retention capacity Nutrient availability Oxygen availability Chemical composition (indicates toxicities, salinity and sodicity) Workability (indicates texture, clay mineralogy and soil bulk-density)	Categorical Data classified into 7 gradient classes of land suitability for agriculture	5 min[b] (lat/long)	Constant with time	Fischer et al. (2012)
Terrain constraints	Elevation, slope and inclination combined	Categorical Data classified into 9 gradient classes			
Socioeconomic factors	Urban/built-up land	% of grid cell area	5 min[b] (lat/long)	10,000 BC–2005 AD (decadal)[c]	Goldewijk et al. (2010)
	Urban population Rural population	Inhabitants/km^2			
	Gross Domestic Product (GDP) per capita	Constant 1990 international (Geary-Khamis) dollars/person	National level	1 AD-2010 (annually between 1800–2010)[d]	Bolt and van Zanden (2013)
	Market accessibility	No units	1 km[b] (lat/long)	~2005	Verburg et al. (2011)

[a] Was linearly interpolated to 0.5° × 0.5° spatial resolution.
[b] We aggregated the data to 0.5° × 0.5° spatial resolution by area-weighed averaging.
[c] We calculated annual estimates by linear interpolation of decadal data.
[d] Missing values for countries were gap filled using nearest values.

land use from which future projections will begin, and capturing larger, longer-term changes in land use and explanatory variables to better support use of the model in long-term future projections. We therefore choose two 10-year sets of data. The first set, to capture longer-term changes, consists of 10 years drawn between 1905 and 1995 at 10-year time steps (i.e. 1905, 1915,. . ., 1995). For each year, a corresponding 10-year lag data point (Y^{t-10}) is used in the estimation procedure (Section 2.5). For example, for year 1905, the lag year data corresponds to 1895, and for 1995 the lag year data corresponds to 1985. The second set, to capture contemporary relationships, includes 10 years of data covering the period 1996–2005 at 1-year time steps. For 1996, the lag year data corresponds to 1986, and for 2005 the lag year data corresponds to 1995.

The explanatory variables (Table 1) used in the analysis have different units and scales. Hence, the estimated regression coefficients (β and γ vectors) are of different scale and cannot be directly interpreted to infer the relative importance of explanatory variables on the dependent variable. To address this problem, we standardize all explanatory variables covering the period 1901–2005 before the parameter estimation and model simulation procedure. The standardization also prevents numerical difficulties that could arise due to scaling problems in the least-squares estimation (Eq. (23)). A standardized coefficient indicates how many standard deviations a dependent variable will change, per standard deviation increase in the explanatory variable (Hunter and Hamilton, 2002). For each explanatory variable associated with the vectors β and γ, we calculate its mean and standard deviation using 5 years of data (2001–2005) separately for each of the 127 sub-region. For a given explanatory variable and grid cell, we standardize the variable using the z-score which is computed as the difference between the value of the variable at that grid cell and its mean value for the corresponding sub-region, divided by the standard deviation corresponding to that sub-region.

See Appendix D for a discussion on how we handle multicollinearity among explanatory variables and excess-zeros problem. Appendix E.4 provides details on the solvers used to implement the land use allocation model (Eq. (12)) and the least-squares optimization (Eq. (23)).

2.7. Simulation procedure to evaluate the land use allocation model

To test the land use allocation model, we compared results from model simulations for the historical period (1901–2005) to the historical reconstruction (RF data) over that period. For this test, we first divided the world into nine regions (Fig. 1), consistent with the regions used in a general equilibrium model of the global economy, the PET (Population-Environment-Technology) model (O'Neill et al., 2010). This regional mask will allow us to subsequently link the land use allocation model with the PET model for exploring future scenarios. For each year over the period 1901–2005, we aggregated the $0.5^\circ \times 0.5^\circ$ lat/long reconstruction data for cropland and pastureland to these nine regions. This regionally aggregated land use information was then used as input to the land use allocation model (Eq. (12)) to form the annual regional-scale constraint on the total area demand for each land use type (through Eqs. (10) and (11)). Next, the land use allocation model (Eq. (12)) allocated the regionally aggregated land use information back to $0.5^\circ \times 0.5^\circ$ spatial resolution by applying time-dependent regional demand constraints and two local constraints (Eqs. (8) and (9)). The model-downscaled land use maps were finally compared to the original $0.5^\circ \times 0.5^\circ$ lat/long RF data to evaluate model performance.

Our evaluation test is rigorous given that most IAMs disaggregate the world into a larger number of smaller regions (14–24 regions; van Vuuren et al., 2011). Fig. 2 depicts our evaluation

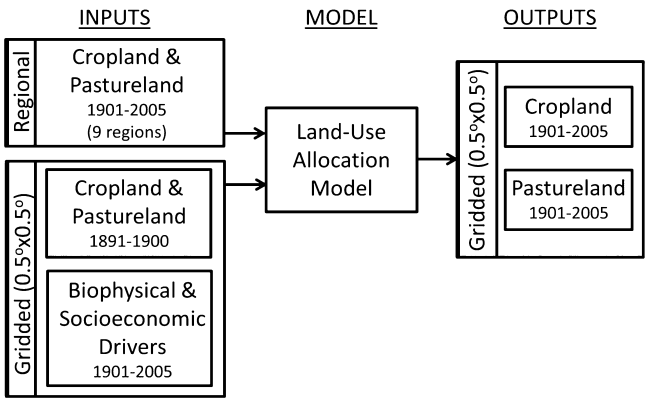

Fig. 2. Schematic representation of the flow of model evaluation experiment.

strategy, which we implement with the algorithm discussed next. This algorithm is repeated separately for each of the nine aggregate world regions.

1. Form the matrices **A** and **B** to use in Eq. (12). Each of the nine aggregate regions has several sub-regions (Fig. 1) with corresponding scalar constants 'a' and 'b'. Therefore, each grid cell in matrices **A** and **B** is weighted based on the parameters 'a' and 'b' within the sub-region the grid cell belongs to. Matrices **A** and **B** are static and need to be computed just once at the start of the model simulation.
2. For each grid cell 'g' within the aggregate region, calculate decreasing returns constant d_g from Eq. (18). For this calculation, spatial data on the standardized explanatory variables X^0_{lg} are used. Note that for each sub-region within the aggregate region (Fig. 1), values for 'γ_0' and 'γ' corresponding to that sub-region are used. This information is used to form the **D** matrix in Eq. (12) for the aggregate region. The term d_g in Eq. (18) is independent of time-step 't', and needs to be calculated only the first time.
3. Set model reference year '$t = 1901$'.
4. Use the RF spatial data for cropland and pastureland for year 1891 to form the $Y^{\bar{t}}$ terms in Eq. (12). The size of vectors $Y^{\bar{t}}$ and $Y^{\bar{t}}$ in Eq. (12) is two times the total number of grid cells within an aggregate region.
5. For time-step 't', and for each grid cell 'g' within the aggregate region, calculate land suitability S^t_{lg} from Eqs. (13) and (14). For this calculation, spatial data on the standardized explanatory variables X^t_{lg} are used. For each sub-region within the aggregate region (Fig. 1), values for 'β_0' and 'β' corresponding to that sub-region are used. This information is used to form the vector S^t in Eq. (12) for the aggregate region.
6. For time-step 't', the land use allocation model computes Y^t (from Eq. (12)) using five variables: (1) matrices **A** and **B** from step 1, (2) matrix **D** from step 2, (3) S^t from step 5, (4) $Y^{\bar{t}}$ where $\bar{t} = t - 10$, and (5) the regional total area demand for each land use type in time-step 't' which is used as input for the regional constraints through Eqs. (10) and (11). This step, when carried out separately for each of the nine world regions, results in a global map of cropland and pastureland for 't'.
7. Increment model time-step by one year ('$t = t + 1$').
8. Repeat steps 5–7 until '$t = 2005$'. For the first ten years of model simulation ($1901 \leq t \leq 1910$), RF data for the period 1891–1900 are used to form the $Y^{\bar{t}}$ term. For $t \geq 1911$, the model predicted maps are utilized to form the $Y^{\bar{t}}$ term.

In summary, the land use allocation model requires three inputs: (1) maps of cropland and pastureland from 1891–1900 to form the lagged $Y^{\bar{t}}$ term for the first 10 years of model simulation (cropland

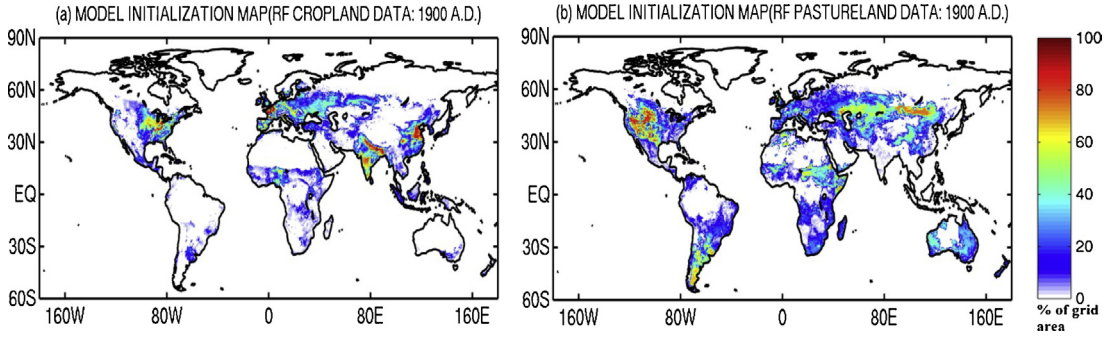

Fig. 3. The cropland and pastureland maps (RF data) for the year 1900 used to form the lagged land use term in the land use allocation model. Units are in percentage of land area within each grid cell.

Table 3
The adjusted kappa coefficient estimated by comparing the model allocated land use maps from the historical simulation (Section 2.7) with RF data for different years. The values are given for cropland and pastureland (brackets).

Year	Land use allocation model developed in this study	Proportional downscaling approach	Mechanistic downscaling approach
1920	0.90 (0.81)	0.73 (0.76)	0.76 (0.72)
1940	0.88 (0.81)	0.71 (0.73)	0.73 (0.69)
1960	0.85 (0.80)	0.63 (0.60)	0.69 (0.60)
1980	0.82 (0.80)	0.61 (0.53)	0.64 (0.58)
2005	0.87 (0.83)	0.59 (0.58)	0.66 (0.61)

and pastureland maps for the 1900 RF data are presented in Fig. 3), (2) annual maps (1901–2005) of explanatory variables, and (3) the annual (1901–2005) aggregate demands for cropland and pastureland for each of the nine world regions. With these inputs the model dynamically allocates aggregate land use information at $0.5^\circ \times 0.5^\circ$ spatial resolution for each year starting from 1901 until 2005. The dependence of our allocation model on previous and neighboring land use through $Y^{\bar{t}}$ and matrix **B** respectively, result in high path dependence of the simulated land use patterns.

3. Results

3.1. Land use allocation model simulation and historical land use patterns

Figs. 4–8 show the model predicted maps for cropland and pastureland from the model simulation at 20-year time intervals. Table 3 summarizes the comparison in terms of adjusted Kappa coefficients. Kappa coefficient is a statistical measure of inter-rater agreement, and range from zero to one (unit less quantities). Greater magnitude of kappa indicates better agreement between the simulated and the actual values (RF data). Adjusted kappa coefficient (Mertens et al., 2003) is same as kappa coefficient, but is intended for sub-grid mapping and ignores grid cells where both predicted and actual values are zero (including zero grid cells would inflate the kappa values without adding much information about models prediction abilities).

Overall, the model predicted fractional areas for cropland and pastureland for the entire 20th century are broadly consistent with RF data (Figs. 4–8; Table 3). There are two points that stand out from the temporal trend in kappa coefficients (Table 3). First, the land use patterns predicted by the model better match the RF data toward the start and end of the simulation period. Second, the prediction of historical pastureland patterns is consistently worse than that for cropland.

The reason for the better performance at the start and end of the simulation is that the model begins in 1901 with the observed land use pattern and projected values deviate over time, so that outcomes early in the century, all else equal, are likely to be more accurate than those later in the century. Toward the end of model simulation, the predicted maps converge toward the actual RF data because our estimated parameters are weighted more toward the contemporary relationships.

The reason for the limited accuracy in predicting pastureland is likely because the spatial reconstructions of pastureland are highly uncertain, so that the explanatory factors used in our study have limited capacity in explaining the historical pastureland patterns. For example, the RF data used here estimates global pastureland area at 26.3 million km^2 during 2005. In comparison, the other well-known HYDE 3.1 reconstruction (see Appendix A) estimates pastureland area at 33.0 million km^2 during 2005, 26% higher than RF data. Therefore, at grid-level, the relative uncertainties in pastureland estimates are even higher (Fig. 9), and increase as we go further back in time from 2005 (Meiyappan and Jain, 2012).

A key feature of the model is its ability to replicate the timing and magnitude of spatial shifts in land use patterns that occur in the RF data, even within an aggregate region. For example, Fig. 10 shows the model predicted net transitions in cropland over the US for the period 1900–1960 (calculated as the difference between 1960 model predictions and the 1900 reference map divided by the number of years). The model is able to reproduce the decline in cropland in the eastern US and subsequent expansion to the mid-western US that occurred over this period (Fig. 10, compare top left and top right panels). The model is able to reproduce the shift in these patterns mainly because we account for the heterogeneous nature of the driving factors within each aggregate region and their changes over time. Similarly, we show the model is able to replicate the key spatial patterns of land-use change (as indicated by RF data) for other world regions: (1) Europe and the western portion of the Former Soviet Union (FSU) between 1935–1960, during which period Europe experienced a gradual decline in cropland, and FSU experienced sharp cropland expansion associated with the opening up of "New Lands" (compare top two left panels in Fig. 11), (2) in the same region, but for the period 1960–2005, when cropland abandonment was common to both Europe and FSU (top two right panels in Fig. 11), and (3) widespread net cropland expansion in the tropics between 1920 and 1980 that resulted in significant deforestation (top two panels in Fig. 12). Overall, results indicate that the general patterns of cropland expansion and abandonment are replicated well compared to RF data, with some exceptions (e.g. in Fig. 12, we simulate cropland expansion in the Caribbean and in parts of India where RF shows abandonment).

4. Estimated parameters

The '*a*' parameters indicate the relative importance of the dynamic adjustment cost model (dependence of land use

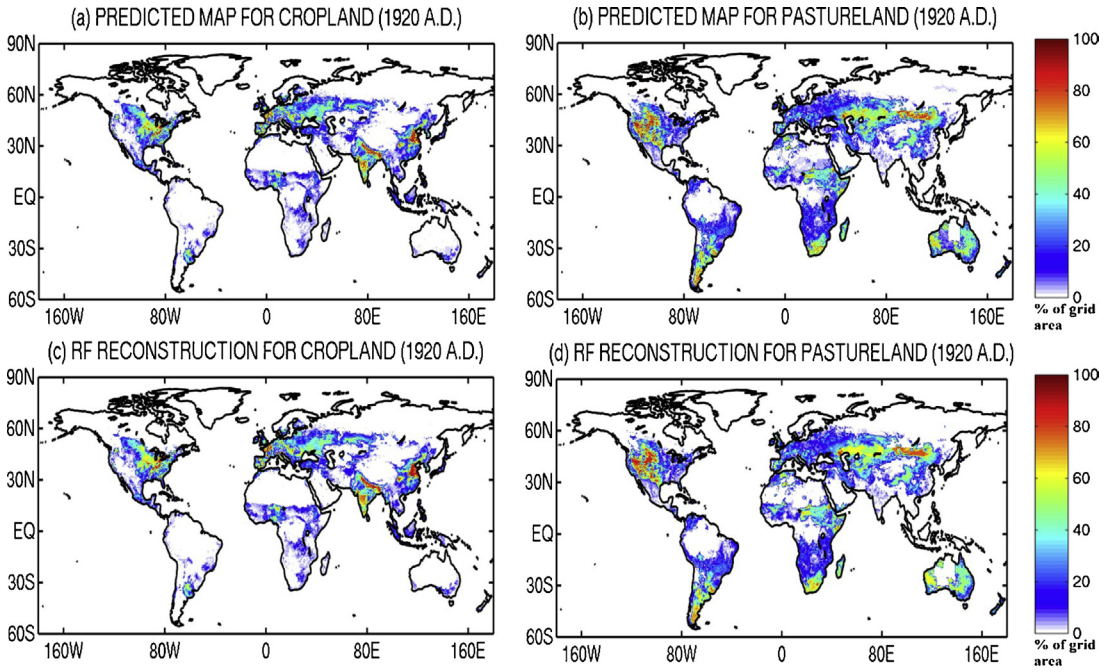

Fig. 4. Model predicted map for cropland and pastureland (top panels) after 20 years of model simulation (i.e. 1920 A.D.). The RF data for 1920 (bottom panels) is shown for comparison purpose. Units are in percentage of land area within each grid cell.

allocation on the previous land use patterns) compared to the static profit maximization function (dependence of land use allocation on potential land suitability). The '*b*' parameters indicate the nature and magnitude of spatial autocorrelation in land use patterns.

Three key results stand out from the estimated '*a*' and '*b*' parameters (Fig. 13). First, the values for both the parameters are spatially heterogeneous across the globe. This heterogeneity would be left unaccounted for if models were not parameterized at sub-global scales.

Second, the '*b*' parameters are non-zero and significant for most regions across the globe indicating that global land use change datasets have significant spatial autocorrelation (note that $b = 0$ indicates no autocorrelation). Therefore, disregarding the presence of spatial autocorrelation from estimation procedure will result in biased parameter estimates. The negative values for '*b*' across most sub-regions indicate the bias would tend to inflate the importance of driving factors in these regions because the estimated '*a*' parameters would be smaller compared to that in Fig. 13a (smaller because when '*b*' is disregarded, the '*a*' parameter would reflect the net effect of the '*a*' and '*b*' parameters).

Third, the '*a*' parameters indicate that temporal autocorrelation is strong for most regions; i.e., the dynamic adjustment cost term

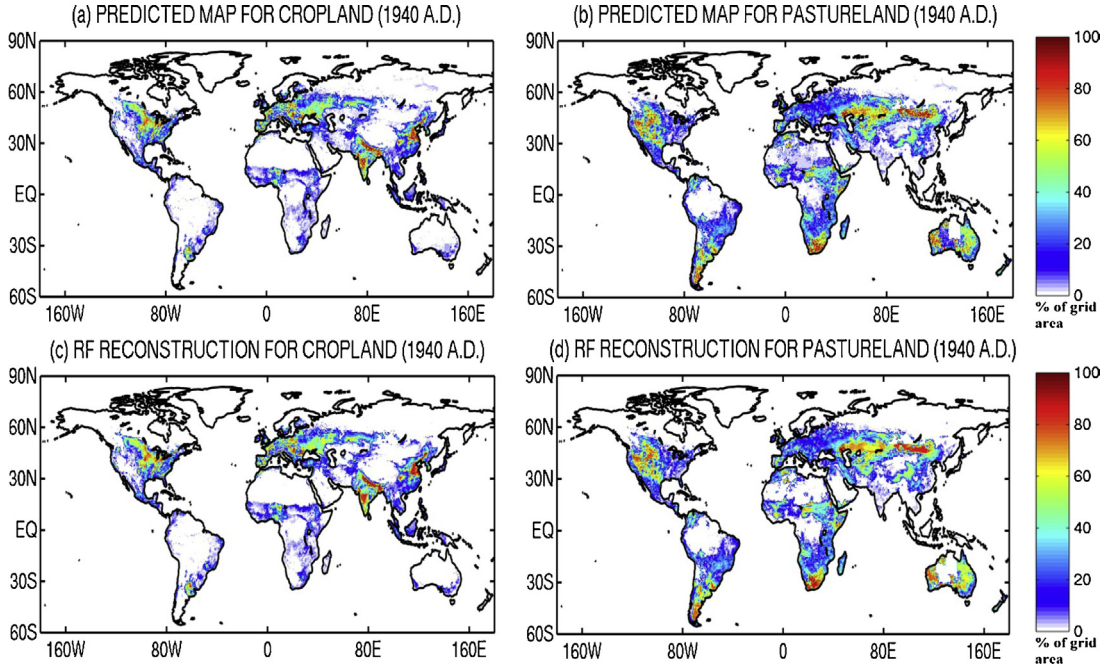

Fig. 5. Same as Fig. 4, but for the year 1940.

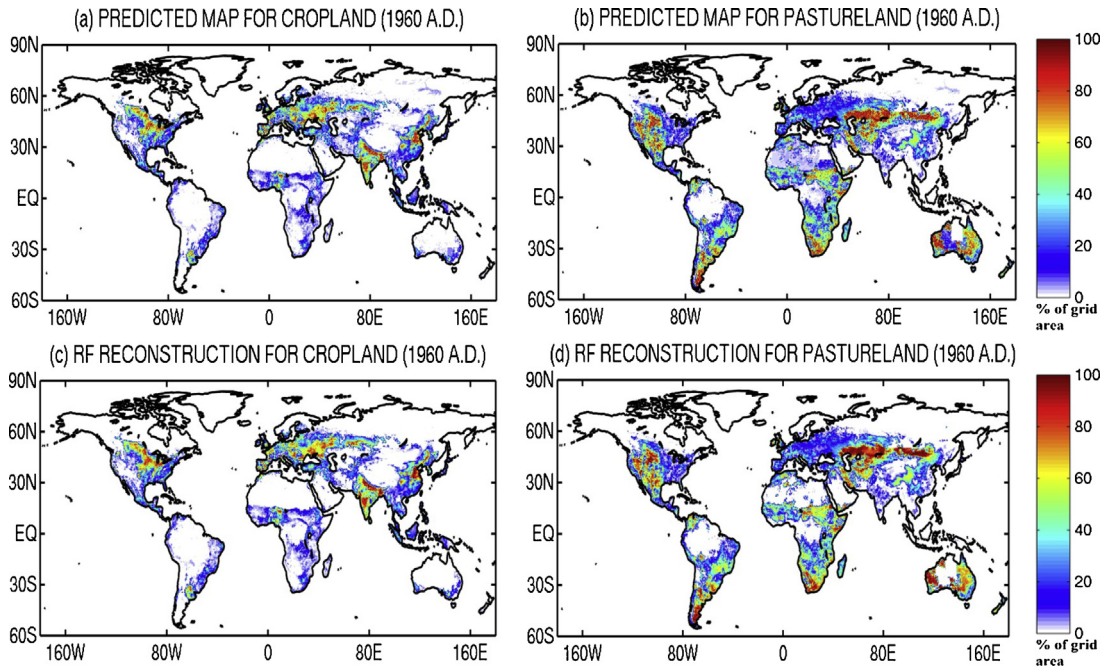

Fig. 6. Same as Fig. 4, but for the year 1960.

dominates land suitability in determining land allocation, leading to a highly path dependent process. Higher '*a*' values generally coincide with regions where extensive agriculture is found as early as 1900 (e.g. cropland in Europe, India and China in Fig. 3a, and pastureland in USA, west of the Mississippi river in Fig. 3b). The spatial patterns of land use in these regions reflect a long history of changes in land use in response to socioeconomic and biophysical factors. The model therefore tends to rely more on previous land use patterns to explain subsequent changes in land use patterns in these regions.

The importance of the adjustment cost term does not imply that driving factors that determine land use suitability are insignificant in a dynamic allocation procedure. High path dependency implies inaccuracies in predicting land use allocations in one year will reduce the accuracy of predictions for subsequent years. As will be shown in the next section, it is this path dependency behavior that makes inclusion of driving factors important. If we exclude driving factors from the land use allocation procedure, the inaccuracy in land use allocations for initial years of simulation would be negligible, but over time they would accumulate to produce land use maps that are substantially different from the historical reconstruction.

4.1. Comparison to other models

To evaluate the land use allocation model, we repeat the historical simulation (Section 2.7) with two other common land use

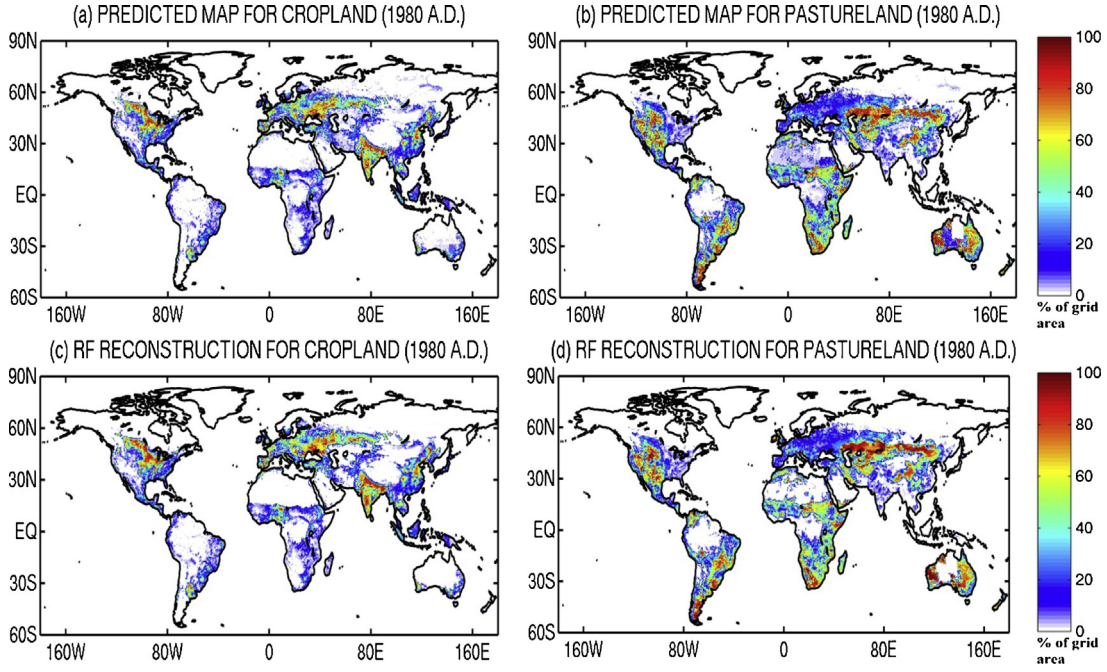

Fig. 7. Same as Fig. 4, but for the year 1980.

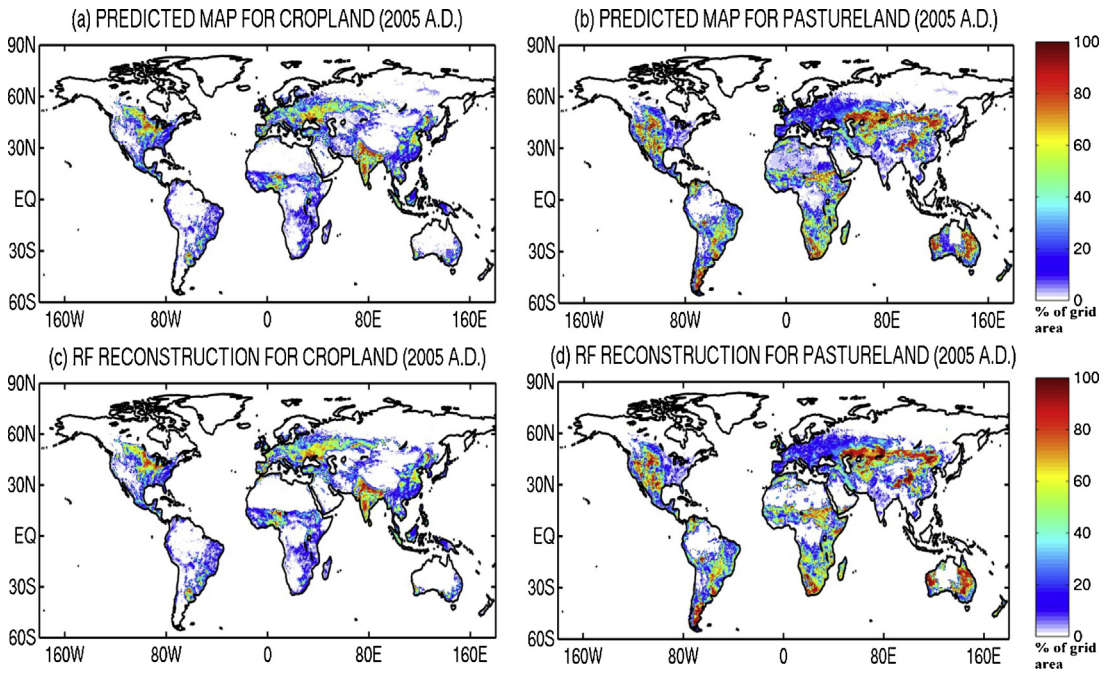

Fig. 8. Same as Fig. 4, but for the year 2005.

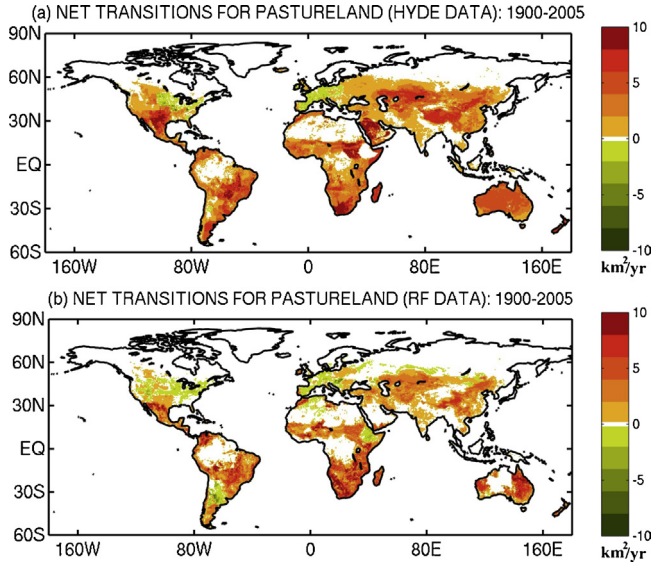

Fig. 9. Comparison of annual net transitions (1900–2005) for pastureland between two widely used spatial reconstructions: (a) HYDE 3.1 database (Klein Goldewijk et al., 2011), and (b) RF data used in our historical simulation. Net transitions for 1900–2005 are calculated as the difference between 2005 map and the 1900 map divided by the number of years. Positive values indicate a net pastureland expansion over the period 1900–2005, and vice versa for negative values. Units are in km^2/yr.

allocation approaches and compare them with our results. We designed both these approaches as representative of general allocation procedures; they do not replicate specific existing models. Full methodologies are provided in Appendix E.5. Here, we highlight the key features of these approaches.

(1) Proportional downscaling approach – The aggregate land use projections are allocated to grid cells as closely as possible to previous year land use patterns. No account is taken of the impact of driving factors. Models following this general approach include GLOBIO3, and GLM used to downscale land use projections from the Global Change Assessment Model (GCAM) model corresponding to the RCP4.5 scenario of the IPCC (van Vuuren et al., 2011).

(2) Mechanistic downscaling approach – The aggregate land use projections are allocated to grid cells as closely as possible to previous year land use patterns (similar to proportional downscaling approach), but the direction of change and the maximum allowable magnitude of change is constrained by the change in a measure of land suitability. Land suitability is determined with regression relationships driven by explanatory variables, and the constraint implies that grid cells in which suitability decreases (increases) must have a decrease (increase) in land use (or no change). This approach mechanistically fits models to explain the spatial relationships between driving factors and land use change and subsequently uses this information to update the allocation maps. Models following this general approach include MIT-IGSM (Wang, 2008).

4.1.1. *Performance of previously published land use allocation approaches*

Results show that the final predicted land use map (2005) using both allocation approaches to downscale the aggregate land demands derived from the RF data are less accurate than our model (compare Figs. 14 and 15 with bottom panels in Fig. 8). The adjusted kappa coefficients indicate that at the end of model simulation (2005 A.D.), both the allocation approaches have 59–66% accuracy in simulating cropland patterns, compared to 87% accuracy by our model (Table 3). The accuracy in simulating pastureland patterns also differ by similar magnitudes between our model and the other two approaches. Inaccuracies using the proportional downscaling approach are driven by the fact that the allocation across grid cells within an aggregate region is homogeneous (Fig. 14), and does not capture major shifts in agricultural patterns caused by changes in the spatial patterns of driving forces (Fig. 10 compare mid-left panel with top-right panel; Figs. 11 and 12 compare middle panels with second-row panels). This leads to severe overestimation of land use within some regions (e.g. ~40% in eastern US for cropland) and a corresponding underestimation in other parts of the same region (e.g. >50% in Great Plains) (Fig. 10). In contrast, the mechanistic downscaling approach tends to reproduce shifts in agricultural

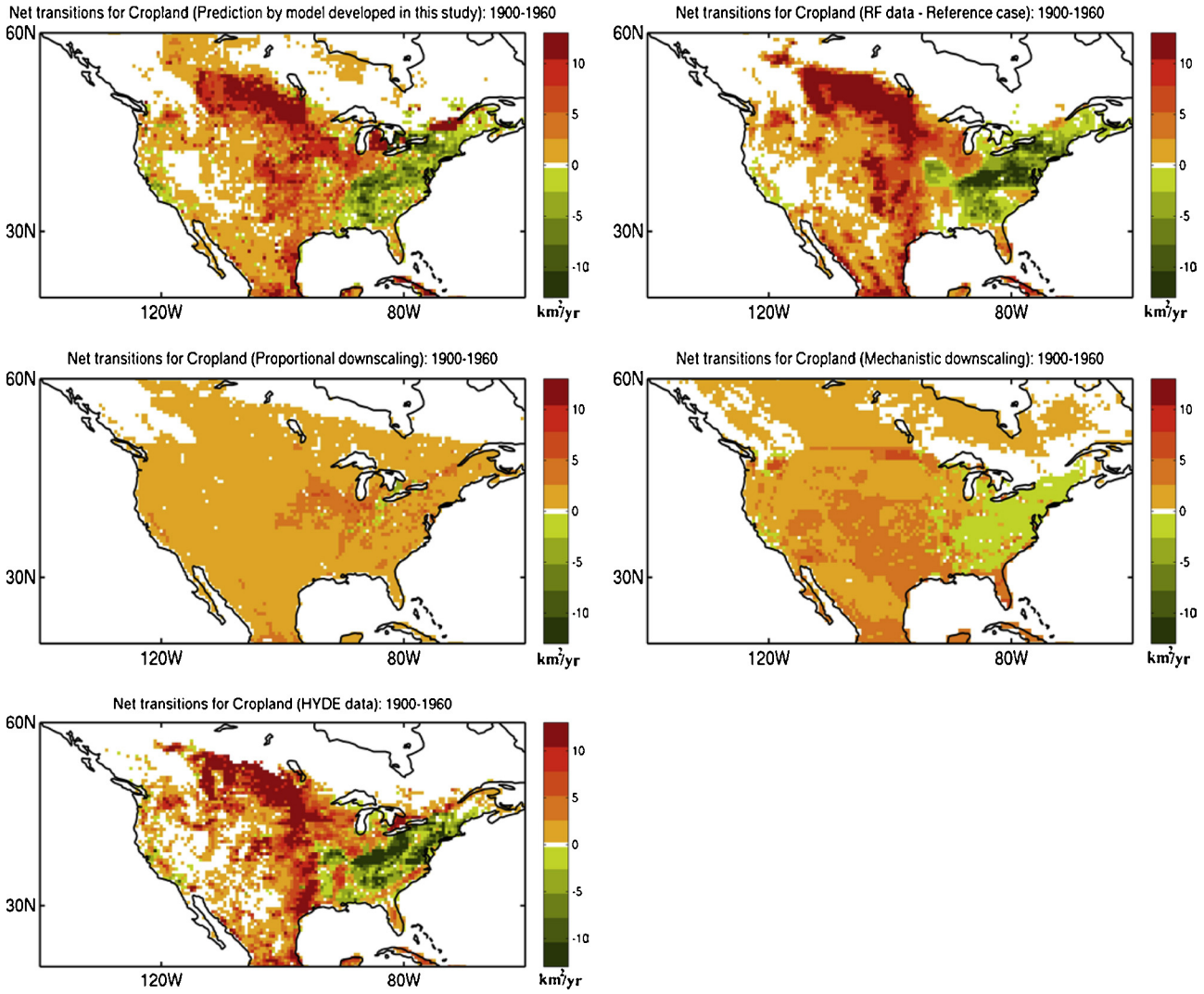

Fig. 10. Net transitions for cropland over the US, averaged over the period 1900–1960 based on: our land use allocation model (top-left), RF data (top-right), proportional downscaling approach (mid-left), mechanistic downscaling approach (mid-right), and HYDE 3.1 data (bottom). Net transitions are calculated as explained in Fig. 9 captions. Units are in km^2/yr.

patterns in some regions, but not accurately. For example, the abandonment of cropland in the eastern US is reproduced to some extent (Fig. 10; mid-right panel), but cropland expansion occurs not only in the Great Plains but also in the western US (Fig. 10; mid-right panel) where pastureland hotspots are located (Fig. 8d). In the case of Europe (Fig. 11) and the tropics (Fig. 12), we find the mechanistic (and proportional) downscaling approach capture the general regions of cropland abandonment and expansion (in reference to RF); however, the hotspots are severely underestimated because the allocation across grid cells within an aggregate region is homogeneous. As evident from our analysis this pattern of allocation is explained by two reasons. First, the importance of driving factors is underrepresented in a mechanistic downscaling approach for most regions, because the driving factors are represented through constraints rather than as an explicit term in the land-use allocation model. Second, this approach does not model the spatial heterogeneity in the relative importance between previous land use patterns and driving factors in determining land use allocation. Therefore, how we represent the role of driving factors within a land use allocation procedure is as important as including the driving factors itself.

We note that the cropland transitions seen in the alpine tundra of the Himalayas (in both proportional and mechanistic allocation approach) is an artifact of our model reproduction methodology (Appendix E.5). In principle, we can force the model not to allocate croplands in such biophysically unfavorable regions through grid cell constraints. However, we did not impose such restrictions, as our aim was only to elucidate the general allocation behavior of both the approaches.

4.1.2. Coupled land use allocation model and historical carbon emissions

To investigate the sensitivity of environmental impacts to alternative land use allocation models, we apply the historical downscaled land use data from our land use allocation model as input to an important type of study land use change models are used for: projecting CO_2 emissions from land use change. The accuracy of the simulated emissions depends on getting the spatial patterns of land use correct where it most matters (i.e. where the carbon consequences are highest). We compare the simulated emissions with those obtained using the two other land use allocation approaches, and with other existing land use reconstructions available in the literature.

We use a land-surface model, the Integrated Science Assessment Model (ISAM) to estimate net CO_2 emissions from land use change at $0.5^\circ \times 0.5^\circ$ lat/long resolution annually for the period 1900–2005.

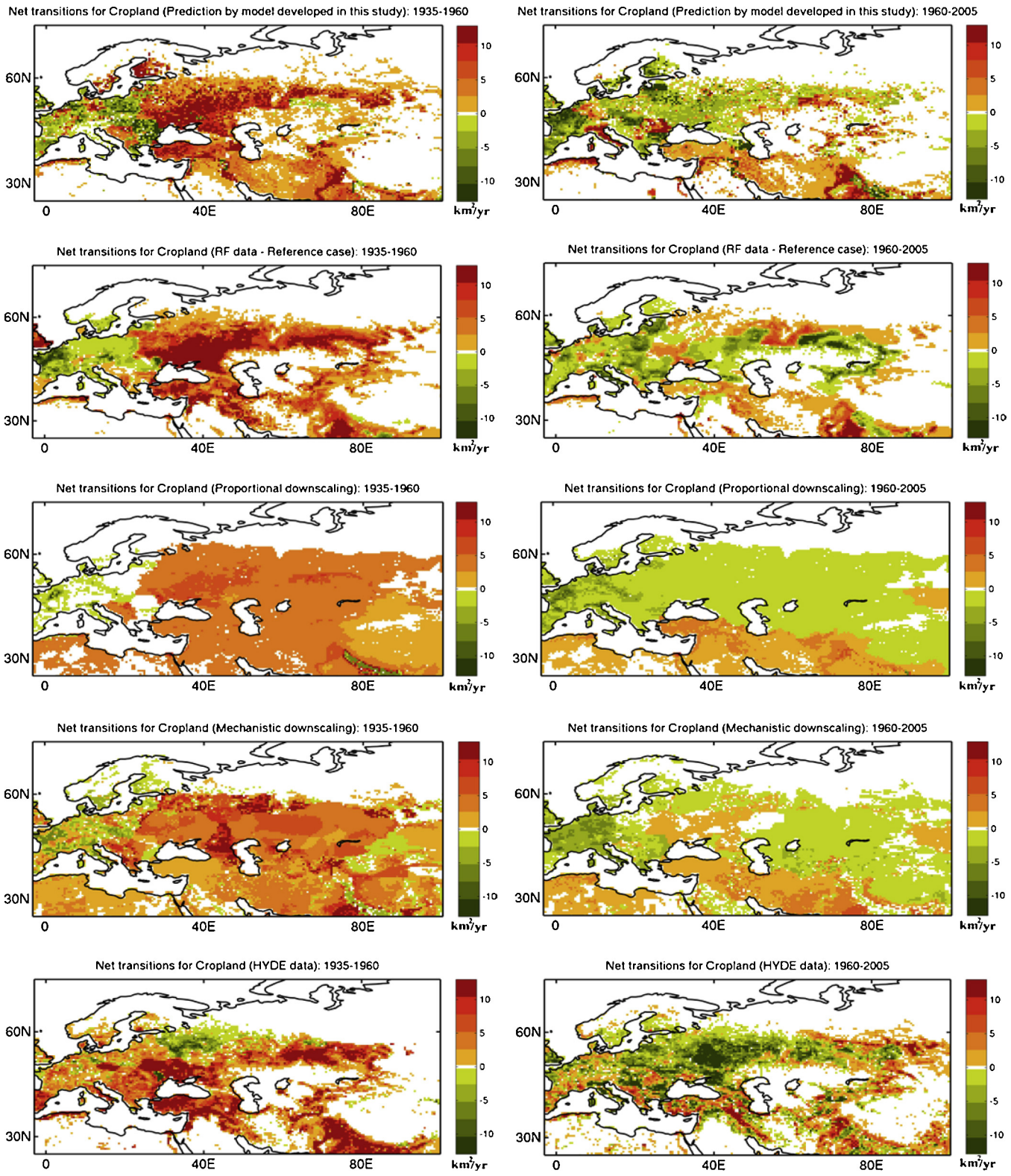

Fig. 11. Net transitions for cropland as in Fig. 10, but shown for the European region. Net transitions shown are averaged for the period 1935–1960 (left panels) and 1960–2005 (right panels). Units are in km^2/yr.

Further background information on ISAM and the simulation protocol is detailed in Appendix E.6. As six separate experiments, we calculate the CO_2 emissions due to changes in the areas of cropland and pastureland from six different land use change datasets. Three of the six land use change datasets are downscaled land use information: (1) from our land use allocation model, (2) the proportional downscaling approach, and (3) the mechanistic downscaling approach. The fourth is the RF data for cropland and pastureland (the reference case because we prescribed the aggregate land demands in datasets (1)–(3) from RF). The other two datasets are independent reconstructions of historical land use change summarized in Meiyappan and Jain (2012): HYDE 3.1 (Klein

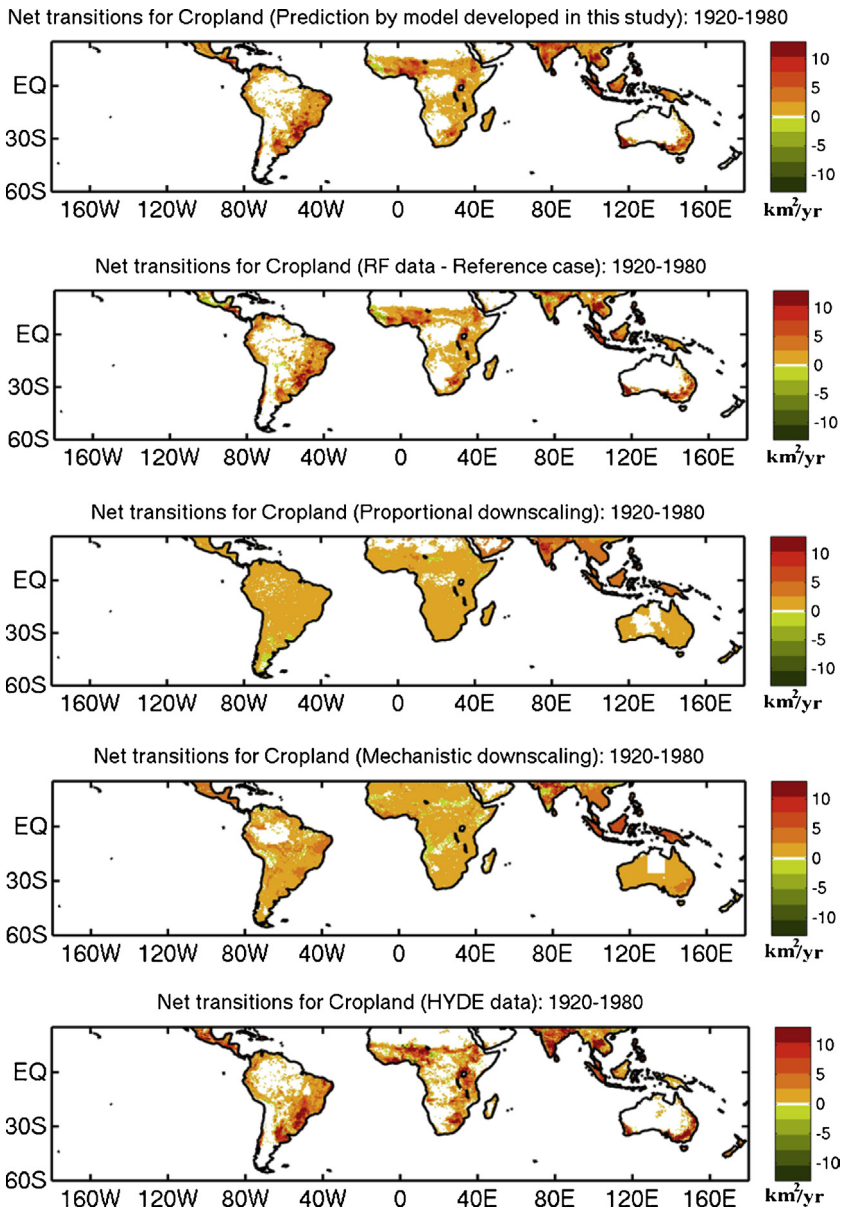

Fig. 12. Net transitions for cropland as in Fig. 10, but shown for the tropics and rest of southern latitudes. Net transitions shown are averaged for the period 1920–1980. Units are in km^2/yr.

Goldewijk et al., 2011), and Houghton (Houghton, 2008). The RF data used in our study is just one realization of what could have happened in the past (Appendix A). Therefore, estimates based on multiple reconstructions are particularly helpful to understand the range of uncertainties among available historical reconstructions of agricultural land use. Our reported net emission excludes emissions from indirect environmental effects (e.g. changes in climate, CO_2 fertilization, and nitrogen deposition).

Fig. 16 provides a comparison of the estimated carbon emissions across the six land use change datasets at aggregate regional scale and cumulated over the period 1900–2005. Four key points are evident. First, estimates based on our land use allocation model compare well with that from RF data, as expected since our modeled spatial land use history also compares well with the RF data, and are within the uncertainty range of three reconstructions. Second, at an aggregate global scale, the proportional and mechanistic downscaling approach overestimate carbon emissions on average by ~0.17 PgC/yr (26%) and ~0.14 PgC/yr (23%), respectively, compared to RF data, even though these two approaches used the same aggregate regional land use change as the RF data. This overestimate is significant given that the total uncertainty (from agricultural land use change, other land disturbance activities, and knowledge gaps in process understanding and modeling) in estimating historical carbon emissions from land-use and land-use change is ~0.5 PgC/yr (Le Quéré et al., 2014). Third, both proportional and mechanistic allocation approaches result in much higher disagreement at a regional scale, compared to RF data. A striking example is North America, where estimates based on the proportional and mechanistic downscaling approaches are ~4.4 and ~2.5 times higher than RF data, respectively. This is a consequence of the higher inaccuracy in reproducing the changes in agricultural hotspots by both downscaling approaches. In the case of RF data and our land use allocation model, abandonment of cropland over the eastern US (Fig. 10; top panels) causes a larger carbon sink (hence, smaller net emissions) due to subsequent forest regrowth (see Fig. E.1 in Appendix E). This important feature is reproduced only to some limited extent in the mechanistic downscaling approach, and is nonexistent for the proportional downscaling approach (Fig. 10;

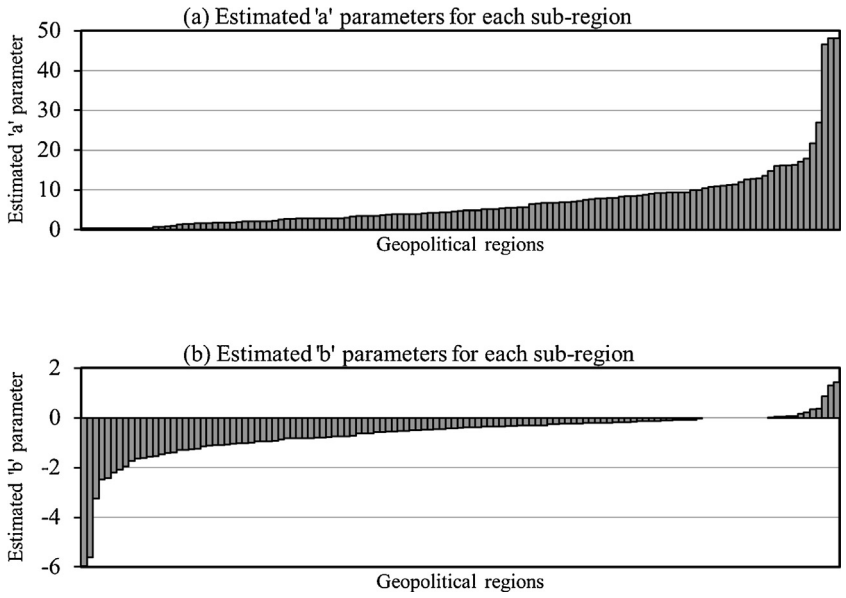

Fig. 13. Estimated '*a*' and '*b*' parameters for each of the 127 sub-regions. The data for both the plots has been independently sorted in ascending order for visualization purpose. The parameters are unit less.

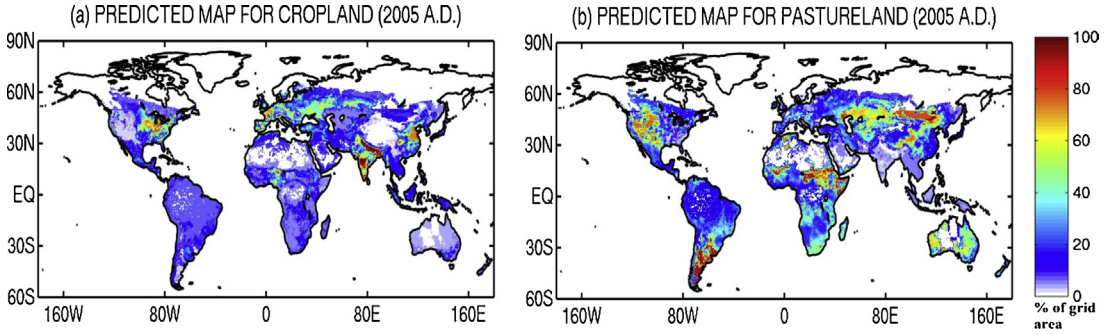

Fig. 14. Cropland and pastureland maps predicted using proportional downscaling approach at the end of model simulation (2005 A.D). Units are in percentage of land area within each grid cell.

mid panels). A consequence of overestimating cropland in the eastern US is a corresponding underestimation of cropland expansion in other parts of the US (consequently lower carbon emissions in these regions compared to RF data). Therefore, at grid level the disagreement in estimated net emissions is much higher for proportional and mechanistic allocation approaches compared to RF data (not shown). Fourth, net emission estimates based on the three existing historical reconstructions of agricultural land use show significant disagreement, especially at the regional scale. This disagreement is largely explained by the difference in agricultural inventory data used by these reconstructions (Jain et al., 2013; Meiyappan and Jain, 2012). No reconstruction is clearly better than another, as is evident from the uncertainties in net transitions estimated between the two reconstructions (see Fig. 9 for pastureland, and for cropland compare the top-right and bottom panel in Fig. 10, and compare the top and bottom panels in Figs. 11 and 12). Therefore, land use allocation approaches need not closely emulate any one reconstruction. However, despite the large uncertainty range

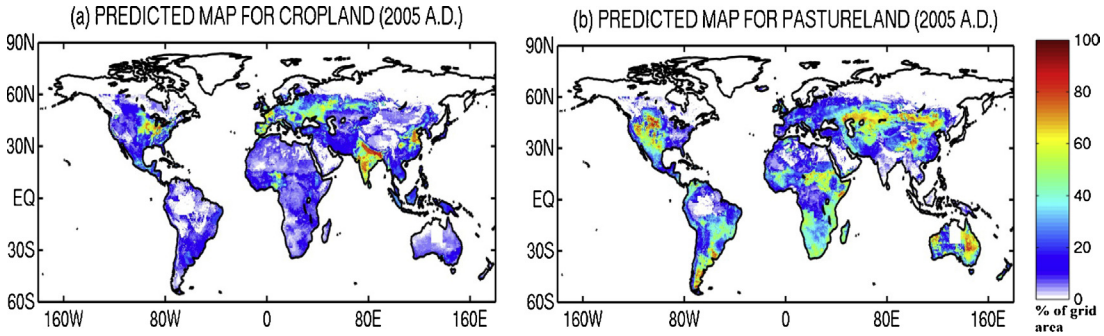

Fig. 15. Cropland and pastureland maps predicted using mechanistic downscaling approach at the end of model simulation (2005 A.D). Units are in percentage of land area within each grid cell.

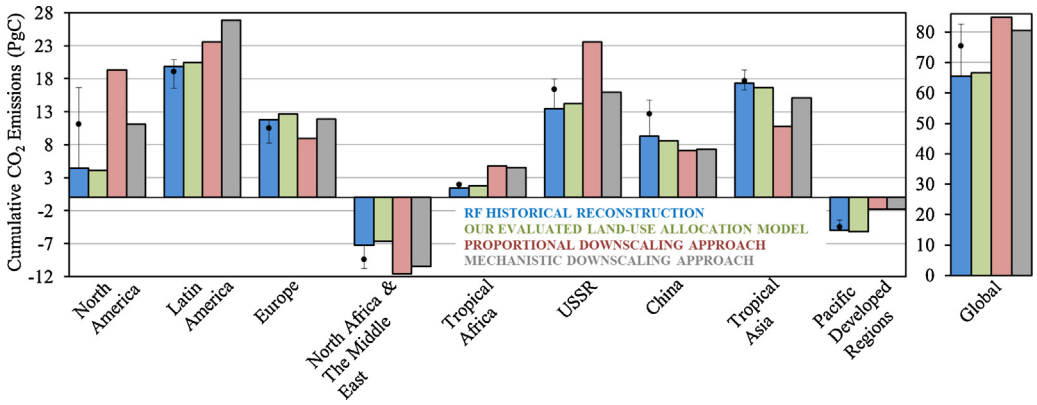

Fig. 16. Estimated carbon emissions from changes in the cropland and pastureland areas calculated using six different land use change datasets. The data presented are aggregated to regional scales and cumulated over the period 1900–2005. The nine regions shown here are consistent with Jain et al. (2013) and different from the nine aggregate regions used in our model evaluation. The black dots represent the estimates averaged over the three historical reconstructions of land use (RF, HYDE 3.1 and Houghton), and the uncertainty bars indicate the maximum range across the three reconstructions. Units are in PgC (1 PgC = 10^{15} gC).

among historical reconstructions, the emission estimates based on both the proportional and mechanistic downscaling approaches fall outside this range for many regions. This underscores the importance of a reliable approach to modeling land use allocation. It is also important to continuously improve the quality of historical land use data to improve models and to more accurately predict future land use change.

5. Discussion

5.1. The land use allocation model

We present a statistical model for land use allocation with an econometric interpretation of land suitability that is based on profit maximization (or cost minimization). The approach integrates economic theory, observed land use, and data on both socioeconomic and biophysical determinants of land use change. It is global in scope and is estimated using long-term historical data, thereby making it suitable for long-term projections, such as in IAMs. The method accounts for spatial heterogeneity in the nature of driving factors across geographic regions. The allocation is modified by autonomous development (previous and neighboring land use patterns, thereby accounting for temporal and spatial autocorrelation), competition between land use types, and exogenous drivers that are treated as explanatory variables. The spatial and temporal resolution at which the model operates is flexible.

Given that the biophysical components in most global-scale land use models operate at a spatial resolution of $0.5^\circ \times 0.5^\circ$ lat/long (or coarser), representing landscape heterogeneity at the sub-grid scale level, at least to some extent, is important. The method of fractional land use prediction developed here is a first step toward representing landscape heterogeneity in global scale land use models. Ultimately, a highly detailed representation of the composition of landscapes is necessary for certain environmental assessments, such as biodiversity (van Asselen and Verburg, 2012, 2013). Two other equally important aspects in land use allocation procedure are to account for the: (1) transient impacts of socioeconomic and biophysical driving factors, and (2) spatial heterogeneity in the relative importance between driving factors and past land use patterns in determining land use allocation. As demonstrated in this study, other downscaling approaches that disregard the first aspect, but predict fractional land use fail to reproduce the hotspots of historical agricultural patterns. A mechanistic land use allocation approach that accounts for the first aspect, but disregards the second, also fails to reproduce the historical land use patterns. We also show that such downscaling approaches could have significant implications for global environmental assessments. Our approach is novel because we predict fractional land use within each grid cell and simultaneously account for the spatial heterogeneity in the relative importance between past land use patterns, and driving factors that change with time.

As a first step, we apply the model framework to evaluate the land use patterns for two generic land use types: cropland and pastureland. However, the framework is extendable to account for individual crop types, for instance to study the global land use implications of large-scale biofuels (Melillo et al., 2009; Havlik et al., 2011; Hallgren et al., 2013). For our model evaluation, we attempt to compile a database of the most important explanatory variables available at the global scale (Table 1). Though the list is incomplete, we show that these variables are adequate to reproduce the changes in the hotspots of historical land use change. To use the allocation framework within an IAM, the stationary explanatory factors (soil and terrain conditions) can be retained, whereas data on transient explanatory factors (e.g. climate and socioeconomics) should be replaced with that simulated by the IAMs. This would allow for studying the two-way interactions between land use and the environment (e.g. climate, hydrology). Further, not all transient explanatory variables used in our historical simulation are projected by current IAMs. An alternative is to replace the explanatory variable with other equivalent proxy indicators simulated by IAMs (e.g. Net Primary Productivity). In such cases, the parameter estimation procedure and evaluation should be repeated using the method discussed in Section 2, following which the evaluated model can directly be used within IAMs to explore future scenarios.

5.2. Land use competition

Competition for land in itself does not drive land use, but is an emergent property of other drivers and pressures (Smith et al., 2010). Our approach accounts for land use competition by simultaneously optimizing the area of cropland and pastureland within each grid cell with the aim of maximizing the overall achievable profits. Hence, the approach prevents inconsistency and approximations in the allocation procedure that would otherwise arise when the spatial patterns for each land use type are determined independently (e.g. see Wang, 2008). Incorporating the effects of competition into spatial allocation models is an important feature, given the potential for growth in demand for agricultural land, particularly for pasture to support projected increase in meat-intensive diets, especially in developing countries (Stehfest et al., 2009).

5.3. Caveats and concluding remarks

Expansion of cropland and pastureland is accompanied by conversion of native vegetation (except when one land use is converted to another). Therefore, one of the important factors that determine the allocation of cropland and pastureland is the type of native vegetation to be replaced. Each native (and managed) vegetation offers different resistance (cost) to conversion depending on the land use type. For example, historically most of the pastureland (managed grassland) has been derived from natural grasslands (with exceptions, notably Latin America where forests are cleared for cattle ranching). However, we chose not to include the type of native vegetation to be converted as a factor in determining the land use allocation patterns. There are three reasons for this choice.

First, our land use allocation is carried out at $0.5^\circ \times 0.5^\circ$ lat/long ($\sim$55 km $\times$ 55 km) resolution, consistent with most global scale land use models. A mix of native vegetation usually occupies such large grid areas. Therefore, the difference in the resistance offered by native vegetation across grid cells becomes less important compared to an approach in which land use data are downscaled to a much higher spatial resolution.

Second, available global scale reconstructions of native vegetation for the 20th century are highly uncertain, especially before the 1960s when remote-sensing observations were unavailable (Meiyappan and Jain, 2012). Therefore, it is undesirable to constrain and evaluate a land use allocation model based on highly uncertain data.

Third, the representation of natural landscapes is fundamentally different among current generation biophysical models. Therefore, land cover maps produced for one model cannot be implemented directly within other models (Pitman et al., 2009). As a result, even in the most recent Coupled Model Intercomparison Project phase 5 (CMIP5) (Taylor et al., 2012), land use changes and the resulting changes in native vegetation are estimated sequentially. First, maps of land use from historical reconstructions are harmonized to connect smoothly with the land use maps for the future scenarios produced by the IAMs (Hurtt et al., 2011). The climate modeling teams combine the land use information with different techniques (e.g. Hurtt et al., 2011; Meiyappan and Jain, 2012; Lawrence et al., 2012; Pitman et al., 2009) to estimate changes in the area of native vegetation, so as to be consistent with the land surface components of their climate or Earth system models. Therefore, our approach is consistent with current approaches.

Despite irrigation exerting a positive influence on agricultural suitability (Lambin et al., 2001, 2003), data limitations precluded us from including irrigation as an explanatory factor in our analysis. While contemporary maps of irrigated areas are available (e.g. Portmann et al., 2010; Thenkabail et al., 2009; Siebert et al., 2005), we could not find historical maps on irrigation to account for the transient impacts. In principle, the presented framework could factor irrigation into the analysis, provided the data are available (e.g. for exploring scenarios that vary assumptions about irrigation).

We also do not explicitly account for the effect of policies on land use allocation patterns, despite their importance. Policy effects are explicitly accounted for in global land use models or IAMs that would provide aggregate regional demands for land that would drive our allocation model. However ideally it would still be useful to reflect policies that operate at smaller spatial scales as well, such as land protection or national planning schemes. Globally and temporally consistent data on policies are not readily available for the 20th century. Although we did not account for policy effects explicitly, they are implicitly reflected in historical land use outcomes and therefore in the effects of other explanatory variables that could be considered proxies. In addition, for exploring policy effects in future scenarios, such assumptions can be accounted for through grid cell constraints. For example, we can force grid cells within protected areas to not allow any land use allocation.

We predict only the net changes in land use areas within each grid cell because our model relies on land use reconstructions that provide only net change information. Available global scale reconstructions of land use (e.g. RF, HYDE, and Hurtt et al. (2011) data which use cropland and pastureland transitions from HYDE) are estimated based on the difference in (sub-)national statistics between two time steps, and therefore estimate net changes. Recent regional studies have shown that relying on net changes rather than gross changes (all area gains and losses) could lead to severe underestimation of land use change (Fuchs et al., 2014), and consequently have significant implications for environmental assessments, especially on the terrestrial carbon cycle (Wilkenskjeld et al., 2014).

Another important limitation is that our land use allocation model does not provide any information on land use intensity. Since 1960, a tripling of crop production has been achieved mainly through intensification, with only a 14% extensification (Bruinsma, 2009). Intensification is expected to become even more decisive in the future in the light of growing population, biofuel consumption, and mandates to protect world forests (Foley et al., 2011; Tilman et al., 2011; Phalan et al., 2011). Several global scale grid level indicators of agricultural land use intensification have recently become available to foster modeling efforts (Kuemmerle et al., 2013), although substantial data gaps, uncertainties, and conceptual challenges exist (Keys and McConnell, 2005; Erb et al., 2013; Kuemmerle et al., 2013; Lambin et al., 2000). If land use intensity is measured by yields (i.e. output per unit area of land use activity) then the product term $S_{lg}^t \times d_g$ (derived from Eq. (13), Eq. (14) and Eq. (18)) itself is a measure of land use intensity. In principle the capital-related inputs (e.g. fertilizer, irrigation, pesticides, or mechanization) that increase agricultural yields can be accommodated as explanatory variables in the logit functions, assuming data are available. If land use intensity is measured by the frequency of cultivation (multiple cropping), then the total harvested area within a grid cell in a year could exceed the grid cell area. Our land use allocation procedure is versatile to accommodate such datasets (e.g. Ray et al., 2012; Portmann et al., 2010) as well. The RF data used in this study does not account for multiple cropping (i.e. $Y_{lg}^t < GA_g$). Therefore, for feasibility, we have assumed the decreasing returns-to-scale is strong enough to deter full use of grid cell area (i.e. $d_g < 1$; see Eq. (18)). However, this restriction when relaxed can handle data on multiple cropping. Detailed representation of management characteristics such as land use intensity is important for IAMs to better capture human–environment interactions and to further improve our prediction capacity.

In a complementary study, we will extend the analysis to examine the role of different explanatory factors in shaping the 20th century patterns of agriculture. This will be carried out using two methods: (1) examining the values of the standardized 'β' and 'γ' parameters, and (2) simulating the land use allocation model (Section 2.7) in the absence of historical changes observed for one or more explanatory variables by keeping the explanatory factors of interest constant at initial values, with all other factors varying with time. We will quantify the effects of an explanatory variable by calculating the grid cell level differences between the final predicted map (2005) without changes in this variable, and the map (2005) obtained by varying all variables (as in Fig. 8).

Our ability to model land use change on longer time scales is crucial for exploring policy alternatives, especially because adaptation and mitigation of climate change requires long-term commitment. Notwithstanding the aforementioned caveats, the framework presented here and the approach to evaluation provides an example that can be useful to the IAM, land use, and the Earth system modeling communities.

Acknowledgements

The National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change Program and the National Science Foundation (NSF) Regional Earth System modeling program supported this work. The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA.

Appendix A. Background information on historical land use change datasets

With regard to historical land use, most spatially explicit agriculture data sets are based on merging of remote sensing and ground-based observations (e.g. Ramankutty and Foley, 1999; HYDE 3.1 from Klein Goldewijk et al., 2011). Ramankutty and Foley (1999) reconstructed spatially explicit (5 min × 5 min lat/long resolution) annual maps of historical (1700–1992) cropland by combining agricultural inventory data with satellite-derived land cover data. More recently, Ramankutty et al. (2008) developed a new global data set of crops and pastures circa 2000 by using agricultural inventory data with much greater spatial detail to train a land cover classification data set obtained by merging two different satellite-derived products. Subsequently, they combined the methodologies from their two previous studies to produce revised annual (1700–2007) maps of historical cropland and pastureland at 0.5° × 0.5° lat/long resolution, which we refer to as RF data (Ramankutty, 2012). The new reconstructions use historical inventory data sets at a higher level of spatial detail than their precursor.

We chose to use the Ramankutty (2012) data in this study due to two reasons: (1) the reconstructions were annual in contrast to decadal in HYDE data, and (2) it represents the most recent and up-to-date estimates available. Their differing time scale of analysis explains the difference in temporal resolution between the two reconstructions. The Ramankutty (2012) study focused on the recent past (three centuries), whereas the reconstruction by Klein Goldewijk et al. (2011) focused on a much longer time scale (greater than 12,000 years).

A key term in our model is the dynamic adjustment cost that accounts for temporal autocorrelation. As described in section 2.5, we had to experiment to select a lag-year ($\tilde{\mathbf{t}}$) that roughly matches the temporal autocorrelation in global historical land use datasets. During the model development stage, the RF data was convenient for experimentation as it provides yearly information by default. For HYDE, we had to linearly interpolate the data between decades that introduces additional uncertainties. In principle, our model could also be applied to HYDE data, and with other reconstructions.

It is important to note that both RF and HYDE reconstructions are themselves results of downscaling (sub-) national agricultural statistics using data on spatial indicators of agricultural locations (e.g. satellite land cover, population density and soil conditions). Therefore, these reconstructions are subject to uncertainties from both inventory data (Meiyappan and Jain, 2012) and methods used (Klein Goldewijk and Verburg, 2013). Further, the agricultural inventory data sets used in the existing reconstructions are collected at 5–10 year gaps (especially before 1960s) and linearly interpolated to yield annual estimates (e.g. Ramankutty and Foley, 1999). Hence, these data sets also lack well-defined annual variations in agricultural land that is an additional source of uncertainty. As a result, both these reconstructions show significant disagreement with each other at various spatial and temporal scales (Meiyappan and Jain, 2012).

Appendix B. Procedure for breaking down the globe into smaller sub-regions

We followed the following five steps to breakdown the globe into 127 sub-regions (Fig. 1).

1. We obtained a spatial database of world administrative boundaries from the Global Administrative Areas database v2 (http://www.gadm.org/). The administrative areas we use are countries and the next lower-level subdivision available such as provinces, states, etc.
2. We converted the administrative maps from the original spatial resolution of 0.315° × 0.315° lat/long to a resolution of 0.5° × 0.5° lat/long. This conversion was necessary to be consistent with the spatial resolution used for the historical simulation.
3. We created a single administrative map by using the country administrative map as a base layer, and overlaying the state level administrative boundaries for the following countries: US, Mexico, Argentina, Brazil, China, Australia and India. These are large countries where the relationships between land use patterns and the explanatory variables could vary substantially within different parts of the country.
4. We then check if each distinct sub-region is composed of at least 200 grid cells. If the sub-region has lesser grid cells, we than eliminate that region by distributing grid cell within that region to proximate regions. We chose the 200-grid threshold based on experimentation, and it represents a compromise between two factors. First, it ensures that each of the sub-regions is large enough to have enough observations to obtain robust parameter estimates (note that we use 20 years of data to estimate the parameters; therefore, the total number of observations is 20 times the number of grid cells within the sub-region). Second, while we want to resolve the spatial heterogeneity in the nature and importance of explanatory variables at a global scale by keeping the regions small, we still want to retain sufficient range in the value of explanatory variables within each small region. An added benefit of breaking down the globe into smaller sub-regions is a reduction in the computational cost associated with inverting the matrix $((a+1)\mathbf{I} + b\mathbf{W})$ to compute $\boldsymbol{\Psi}$ (see Appendix E.3). Ideally, each sub-region should be selected such that its grid cells have similar agricultural characteristics. Selecting regions based on similar biophysical characteristics is possible (e.g. using agro-ecological zones - see Fischer et al., 2012). However, socioeconomic factors do not exhibit patterns as uniform as biophysical factors. Given that a number of socioeconomic factors are included in our analysis (Table 1) we found administrative boundaries to be both a convenient and consistent method to select sub-regions.
5. The above steps resulted in breaking down the globe into 127 distinct sub-regions as shown in Fig. 1. A histogram of the

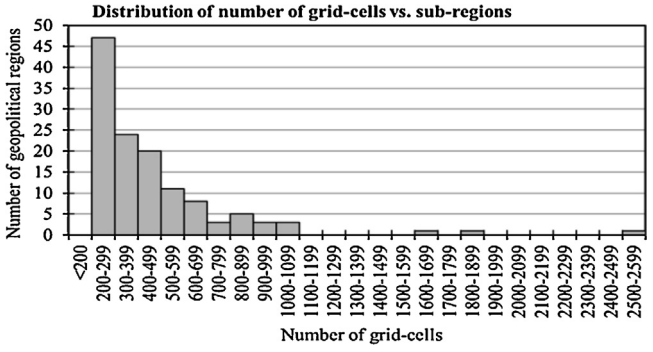

Fig. B1. Histogram of the number of sub-regions versus the total number of grid cells with the region.

number of sub-regions versus the total number of grid cells within the region is shown in Fig. B1.

Appendix C. Explanatory variables for land use allocation model

Table 1 lists the 46 explanatory variables used in our regression analysis. Table 2 lists the source data used to compile the explanatory factors for model evaluation. Table 2 also includes (as footnotes) information on processing of raw data sources to match the spatial and temporal resolution required for our model evaluation. Here we provide details on: (1) the rationale for selecting these explanatory factors, and (2) any additional refinements applied to the raw data to construct the explanatory variables. All the 46 explanatory variables compiled for model evaluation cover the time period 1900-2005 (annual time steps) at a uniform spatial resolution of $0.5^\circ \times 0.5^\circ$ lat/long.

C.1. Climate

To determine the climate suitability to cropland and pastureland for each year, we use seasonal (spring, summer, fall and winter) mean temperature, precipitation, and potential evapotranspiration (PET) for the 3 preceding years to allow for adaptive land use adjustments. Inclusion of time-lagged climate variables as explanatory factors is motivated by the fact that climatic conditions over the past few years are taken into account to evaluate the potential land suitability for agriculture, prior to making investment decisions (Mu et al., 2013). In other words, land use decisions are based on expected climate rather than current weather conditions. This effect is not implicitly reflected in our dynamic cost adjustment term (first term in Eq. (12)) for two reasons. First, the potential land suitability for the current year is evaluated independent of previous land use patterns. Second, the dynamic cost adjustment term includes a ten-year time-lag effect, whereas land use adjustments in response to climate change occur over a relatively smaller time scale (Mu et al., 2013).

According to agronomic studies, plant growth is partly a non-linear function of weather conditions, especially temperature (Schlenker and Roberts, 2006). To account for non-linearity, we also include squared terms for each of the three seasonal climate variables as explanatory factors in our analysis. We use temperature in Kelvins to avoid squaring of negative terms. While historical temperature and precipitation data were measured values (Harris et al., 2013), PET is estimated using the grass reference evapotranspiration method (Ekström et al., 2007). PET indicates the capacity of the atmosphere to remove water from the surface through processes such as winds and radiative energy transfer.

C.2. Climate variability

Given the significant evidence on the impacts of climate variability on land use decisions (e.g. Parry et al., 1999; Olesen and Bindi, 2002), we use three indices reflective of climate variability: drought, temperature, and precipitation extremes. For incidence of drought, we use the seasonal Palmer Drought Severity Index (PDSI) (Dai, 2011a,b). PDSI is a measurement of dryness with values ranging between -10 (dry) and +10 (wet). For temperature extremes, we use the heat wave duration index, calculated by counting the number of days in a year with maximum temperature higher by at least 5 °C than the climatological norm (1961–1990) for the same calendar day (Alexander et al., 2006). For precipitation extremes, we use the simple daily intensity index defined as the annual total precipitation divided by the number of wet days (Alexander et al., 2006). Both the temperature and precipitation extreme index used here are qualitatively representative of other indices measuring temperature and precipitation extremes, respectively (Tebaldi et al., 2006). Similar indices representative of climate extremes were used by Mu et al. (2013) to understand the impacts of climate adaptation on the US land use trends.

Pastureland extent and livestock comfort are correlated. Temperature-humidity index (THI) is an indicator of the degree of discomfort experienced by cattle due to seasonal conditions. Previous studies have shown that higher THI in summer results in lower feed intake of livestock, thus inducing a smaller number of animals per hectare, and a heavier stocking rate in the spring or winter (Hubbard et al., 1999; Mu et al., 2013). We estimated seasonal THI following Mu et al. (2013).

The PDSI data was at $2.5^\circ \times 2.5^\circ$ lat/long resolution which we linearly interpolated to $0.5^\circ \times 0.5^\circ$ lat/long (Table 2). This adjustment will affect the allocation patterns, and we did not conduct a systematic sensitivity analysis to test the results against different interpolation techniques. In general, the interpolation technique for an explanatory variable should be given more focus, if that variable is sensitive in determining the allocation patterns (i.e. its 'β' value estimated from Eq. (23) is non-zero and significant).

C.3. Soil and terrain characteristics

We use five different factors indicating soil suitability for agriculture (Table 1) as explanatory variables in our analysis. Each 5 min grid cell in the original data was classified into one of the seven categories, each indicating a different percentage range of soil suitability (Table 2). For each category, we take the median suitability percentage and aggregate the data to $0.5^\circ \times 0.5^\circ$ lat/long resolution by area-weighted averaging. We process the map for terrain constraints similarly to the aggregation procedure followed for soil constraints.

We assume that soil and terrain conditions do not change within the time frame of analysis (~100 years), because soil development processes are very slow. However, this assumption might not hold in regions where organic content could build up rapidly. Moreover, consistent and temporally varying global maps of soil and terrain condition are unavailable historically and are unlikely to be simulated by Earth System Models for future projections.

C.4. Urban area and population

We include urban land and urban and rural population density variables as potential explanatory factors because these are considered important drivers of land use change (Geist et al., 2006). In addition to population density, we also include four-year averaged rates of change in urban and rural population density to account for the impact of changes in population pressure on land use allocation patterns.

C.5. Market influence index

Markets play a prominent role in shaping agriculture, deforestation, and other land use patterns (Lambin et al., 2001; Lambin and Meyfroidt, 2011). However, global scale studies have largely ignored the influence of markets in shaping land use patterns due primarily to lack of global scale data on markets. Here, we describe our approach to derive a single variable called market influence index to use as an explanatory variable in the allocation procedure. The market influence index is a combination of two variables: market accessibility and market importance.

Verburg et al. (2011) developed the first high-resolution (1 km^2) global map (for 2005) indicating accessibility to markets. In simple terms, their market accessibility index reflects the travel time to nearest cities, taking into account travel impedance due to several terrain characteristics (e.g. road networks). The index varies from 0 (low accessibility) to 1 (high accessibility). For our analysis, we require market accessibility maps for the entire 20th century,

whereas a number of terrain characteristics used by Verburg et al. (2011) (e.g. road networks, wetlands) were unavailable globally for the historical period and are also unlikely to be simulated by IAMs for the future. Instead, we model the market accessibility index to be a function of urban land area alone, neglecting other measures of infrastructural development. The rationale behind our assumption is that a grid cell with a higher fraction of urban area would likely have more accessibility (e.g. road networks and other infrastructure) as well. A reciprocal function is commonly used in spatial interaction and gravity models to manipulate distance or accessibility. We follow a similar approach and model the relationship between the year 2005 urban land area (Goldewijk et al., 2010) and the corresponding market accessibility index (Verburg et al., 2011) of the form:

$$A_g = c_1 b_g^{-c_2}$$

where 'A_g' is the market accessibility index in grid cell 'g', 'b_g' is the urban land fraction specific to each grid cell, and 'c_1' and 'c_2' are the parameters to be estimated. We fit the model separately for each continent. We assume the same relationships to be true at any other point of time and we combine them with historical (1900–2005) urban land data (Goldewijk et al., 2010) to produce annual maps of market accessibility. One can think of market influence as reflective of both market importance and its accessibility. Gross Domestic product (GDP) is a well-accepted measure of market importance. We multiplied the national per capita GDP values (Bolt and van Zanden, 2013) with spatially explicit market access data to produce a data set of market influence for the period 1900-2005. In other words, we downscale GDP per capita to grid cells using the market accessibility index.

Appendix D. On multicollinearity and excess zeros

Multicollinearity is a common problem in land use change modeling where one or more explanatory variables are dependent on each other. A high degree of multicollinearity results in high standard errors and spurious parameter estimates. We used *a priori* correlation analysis and did not find any pair of the 46 standardized explanatory variables that exceeded a correlation of 0.7 for each of the 127 sub-regions. We therefore did not discard any explanatory factor from our model. For comparison, a correlation of 0.8 is used as a threshold in land-change studies to discard explanatory variables (Lesschen et al., 2005). Even in the presence of multicollinearity, the specification of error terms in our estimation procedure implies that our point estimates are unbiased. Further, multicollinearity will not affect the efficiency of the fitted model to new data provided that the explanatory variables follow the same pattern of multicollinearity in the new data as in the data based on which the parameters were estimated. A problem would occur only if two or more explanatory variables are a perfect linear combination of each other.

There are three approaches commonly followed in land use change models to eliminate explanatory variables due to multicollinearity: *a priori* correlation analysis, factor analysis and stepwise regression (Lesschen et al., 2005). All of them have a high risk of throwing out important explanatory variables and causing omitted variable bias, among many others (James and McCulloch, 1990). These methods are viable if used for smaller scale studies, where additional care can be taken on the explanatory variables being eliminated from the model. However, this task is taxing at a global scale. Additionally, given the limited number of explanatory factors available for global scale analysis, it is crucial to utilize them to the maximum potential possible. In the initial stages of our model development, we experimented with a state-of-the-art method to deal with multicollinearity, elastic-net regularization (Zou and Hastie, 2005). We found the elastic-net method to be more effective than the traditional approaches used in land use change models. The elastic-net method tries to capitalize on the strengths of correlated variables by converging their coefficients towards each other rather than eliminating them. Unfortunately, we could not build elastic-net regularization into our final estimation procedure because the current packages (Friedman et al., 2007, 2010) available for elastic-net cannot accommodate nonlinear restrictions among coefficients.

As evident from the historical land use patterns (Figs. 3–8), there are many locations around the world that are unsuitable for agriculture (i.e.) grid cells where cropland and pastureland areas are zero consistently. When estimating Eq. (23), having excessive zero-valued observation could result in biased parameter estimates. This is known as the excess-zeros problem (Breen, 1996). To determine whether there is bias in our estimates, we estimated Eq. (23) with a "trimmed" sample where the grid cells unsuitable for agriculture were removed (see censored regressions – Breen, 1996). We then compared the estimated parameters for each of the 127 subregions with those obtained without "trimming" the sample. With the exception of four regions, we found no significant problem with excess-zeros. Within these four regions (e.g. Alaska), more than 95% of the total grid cells had zero land use areas for the entire 20th century. For the grid cells within these four regions, we force the model not to allocate any cropland and pastureland through Eq. (9) by replacing one on the right-hand side of the equation with zero. If excess-zeros had been a problem in some regions of the globe, then maximum likelihood estimation with a censored regression model is a consistent estimation method (Breen, 1996).

Appendix E. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2014.07.027.

References

Agarwal, C., Green, G.M., Grove, J.M., Evans, T.P., Schweik, C.M., 2002. A Review and Assessment of Land Use Change Models: Dynamics of Space, Time, and Human Choice. US Department of Agriculture, Forest Service, Northeastern Research Station, Newton Square, PA, pp. 12–27.

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., Vazquez-Aguirre, J.L., 2006. Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res. Atmos. (1984–2012) 111 (D5).

Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M., ten Brink, B., 2009. GLOBIO3: a framework to investigate options for reducing global terrestrial biodiversity loss. Ecosystems 12 (3), 374–390.

Augustin, N.H., Mugglestone, M.A., Buckland, S.T., 1996. An autologistic model for the spatial distribution of wildlife. J. Appl. Ecol., 339–347.

Bennett, E.M., Carpenter, S.R., Caraco, N.F., 2001. Human impact on erodable phosphorus and eutrophication: a global perspective. BioScience 51 (3), 227–234.

Bolt, J., van Zanden, J.L., 2013. The Maddison Project: The First Update of the Maddison project; Re-Estimating Growth Before 1820. Maddison-Project Working Paper WP-4, Available from http://www.ggdc.net/maddison/maddison-project/home.htm

Bouwman, A.F., Kram, T., Klein Goldewijk, K., 2006. Integrated Modeling of Global Environmental Change. An overview of IMAGE 2.4. Netherlands Environmental Assessment Agency, Bilthoven.

Breen, B., 1996. Regression Models: Censored, Sample Selected or Truncated Data. Sage Publications, London/Thousand Oaks, CA.

Briassoulis, H., 2000. Analysis of land use change: theoretical and modeling approaches. In: Loveridge, S. (Ed.), The Web Book of Regional Science. West Virginia University, Morgantown.

Bruinsma, J., 2009. The resource outlook to 2050: by how much do land, water and crop yields need to increase by 2050? In: How to feed the World in 2050. Proceedings of a Technical Meeting of Experts, Rome, Italy, 24–26 June 2009. Food and Agriculture Organization of the United Nations (FAO), pp. 1–33.

Dai, A., 2011a. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. J. Geophys. Res. Atmos. (1984–2012) 116 (D12.).

Dai, A., 2011b. Drought under global warming: a review. Wiley Interdiscip. Rev. Climate Change 2 (1), 45–65.

Dormann, C., M McPherson, J., B Araújo, M., Bivand, R., Bolliger, J., Carl, G., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30 (5), 609–628.
Ekström, M., Jones, P.D., Fowler, H.J., Lenderink, G., Buishand, T.A., Conway, D., 2007. Regional climate model data used within the SWURVE project – 1: projected changes in seasonal patterns and estimation of PET. Hydrol. Earth Syst. Sci. 11 (3), 1069–1083.
Erb, K.H., Haberl, H., Jepsen, M.R., Kuemmerle, T., Lindner, M., Müller, D., Verburg, P.H., Reenberg, A., 2013. A conceptual framework for analysing and measuring land use intensity. Curr. Opin. Environ. Sust., http://dx.doi.org/10.1016/j.cosust.2013.07.010.
Fischer, G., Nachtergaele, F.O., Prieler, S., Teixeira, E., Tóth, G., van Velthuizen, H., Wiberg, D., 2012. Global Agro-Ecological Zones (GAEZ v3. 0): Model Documentation. International Institute for Applied systems Analysis (IIASA), Laxenburg. Austria and the Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, Data accessible from http://www.fao.org/nr/gaez/en/
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Snyder, P.K., 2005. Global consequences of land use. Science 309 (5734), 570–574.
Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Zaks, D.P., 2011. Solutions for a cultivated planet. Nature 478 (7369), 337–342.
Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., 2007. Pathwise coordinate optimization. Ann. Appl. Stat. 1 (2), 302–332.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1.
Fuchs, R., Herold, M., Verburg, P., Clevers, J., Eberle, J., 2014. Gross changes in reconstructions of historic land use for Europe during the 20th century. Global Change Biol. (in review).
Geist, H.J., McConnell, W.J., Lambin, E.F., Moran, E., Alves, D., Rudel, T.K., 2006. Causes and trajectories of land use/cover change. In: Lambin, E.F., Geist HJ (Eds.), Land Use and Land-Cover Change. Local Processes and Global Impacts. Springer-Verlag, Berlin Heidelberg, pp. 41–70.
Goldewijk, K.K., Beusen, A., Janssen, P., 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. Holocene 20 (4), 565–573.
Golub, A., Hertel, T., Sohngen, B., 2008. Land Use Modeling in Recursively-dynamic GTAP Framework (No. 2609). GTAP Working paper No. 48. Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.
Gouel, C., Hertel, T.W., 2006. Introducing Forest Access Cost Functions into a General Equilibrium Model. GTAP Research Memorandum 8.
Green, R.E., Cornell, S.J., Scharlemann, J.P., Balmford, A., 2005. Farming and the fate of wild nature. Science 307 (5709), 550–555.
Hallgren, W., Schlosser, C.A., Monier, E., Kicklighter, D., Sokolov, A., Melillo, J., 2013. Climate impacts of a large-scale biofuels expansion. Geophys. Res. Lett. 40 (8), 1624–1630.
Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2013. Updated high-resolution grids of monthly climatic observations. Int. J. Climatol., http://dx.doi.org/10.1002/joc.3711.
Havlik, P., Schneider, U.A., Schmid, E., Böttcher, H., Fritz, S., Skalský, R., Obersteiner, M., 2011. Global land use implications of first and second generation biofuel targets. Energy Policy 39 (10), 5690–5702.
Heistermann, M., Müller, C., Ronneberger, K., 2006. Land in sight? Achievements, deficits and potentials of continental to global scale land use modeling. Agric. Ecosyst. Environ. 114 (2), 141–158.
Hibbard, K., Janetos, A., van Vuuren, D.P., Pongratz, J., Rose, S.K., Betts, R., Feddema, J.J., 2010. Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling. Int. J. Climatol. 30 (13), 2118–2128.
Houghton, R.A., 2008. Carbon flux to the atmosphere from land use changes: 1850–2005. In: TRENDS: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, USA.
Hubbard, K.G., Stooksbury, D.E., Hahn, G.L., Mader, T.L., 1999. A climatological perspective on feedlot cattle performance and mortality related to the temperature-humidity index. J. Prod. Agric. 12 (4), 650–653.
Hunter, J.E., Hamilton, M.A., 2002. The advantages of using standardized scores in causal analysis. Hum. Commun. Res. 28 (4), 552–561.
Hurtt, G.C., Chini, L.P., Frolking, S., Betts, R.A., Feddema, J., Fischer, G., Wang, Y.P., 2011. Harmonization of land use scenarios for the period 1500–2100: 600 years of global gridded annual land use transitions, wood harvest, and resulting secondary lands. Climatic Change 109 (1–2), 117–161.
Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agric. Ecosyst. Environ. 85 (1), 7–24.
Jain, A.K., Meiyappan, P., Song, Y., House, J.I., 2013. CO_2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land cover data. Global Change Biol. 19, 2893–2906.
James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box? Ann. Rev. Ecol. Syst. 21, 129–166.
Keys, E., McConnell, W.J., 2005. Global change and the intensification of agriculture in the tropics. Global Environ. Change 15 (4), 320–337.
Kindermann, G., Obersteiner, M., Sohngen, B., Sathaye, J., Andrasko, K., Rametsteiner, E., Beach, R., 2008. Global cost estimates of reducing carbon emissions through avoided deforestation. Proc. Natl. Acad. Sci. U.S.A. 105 (30), 10302–10307.
Klein Goldewijk, K., Verburg, P.H., 2013. Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set. Landsc. Ecol. 28 (5), 861–877.
Klein Goldewijk, K., Beusen, A., Van Drecht, G., De Vos, M., 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecol. Biogeogr. 20 (1), 73–86.
Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. Curr. Opin. Environ. Sustain. 5, 484–493.
Lambin, E.F., Meyfroidt, P., 2011. Global land use change, economic globalization, and the looming land scarcity. Proc. Natl. Acad. Sci. U.S.A. 108 (9), 3465–3472.
Lambin, E.F., Rounsevell, M.D.A., Geist, H.J., 2000. Are agricultural land use models able to predict changes in land use intensity? Agric. Ecosyst. Environ. 82 (1), 321–331.
Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Xu, J., 2001. The causes of land use and land-cover change: moving beyond the myths. Global Environ. Change 11 (4), 261–269.
Lambin, E.F., Geist, H.J., Lepers, E., 2003. Dynamics of land use and land-cover change in tropical regions. Ann. Rev. Environ. Resour. 28 (1), 205–241.
Lancaster, T., 2000. The incidental parameter problem since 1948. J. Econometrics 95 (2), 391–413.
Lawrence, P.J., Feddema, J.J., Bonan, G.B., Meehl, G.A., O'Neill, B.C., Oleson, K.W., Thornton, P.E., 2012. Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the community climate system model (CCSM4) from 1850 to 2100. J. Clim. 25 (9.).
Le Quéré, C., Peters, G.P., Andres, R.J., Andrew, R.M., Boden, T.A., Ciais, P., Zaehle, S., 2014. Global carbon budget 2013. Earth Syst. Sci. Data 6 (1), 235–263.
Lesschen, J.P., Verburg, P.H., Staal, S.J., 2005. Statistical Methods for Analysing the Spatial Dimension of Changes in Land Use and Farming Systems. International Livestock Research Institute.
Letourneau, A., Verburg, P.H., Stehfest, E., 2012. A land use systems approach to represent land use dynamics at continental and global scales. Environ. Model. Softw. 33, 61–79.
Lotze-Campen, H., Popp, A., Beringer, T., Müller, C., Bondeau, A., Rost, S., Lucht, W., 2010. Scenarios of global bioenergy production: the trade-offs between agricultural expansion, intensification and trade. Ecol. Model. 221 (18), 2188–2196.
Lubowski, R.N., Plantinga, A.J., Stavins, R.N., 2008. What drives land use change in the United States? A national analysis of landowner decisions. Land Econ. 84 (4), 529–550.
MEA, 2005. Ecosystems and Human Well-being, vol. 5. Island Press, Washington, DC.
Meiyappan, P., Jain, A.K., 2012. Three distinct global estimates of historical land-cover change and land use conversions for over 200 years. Front. Earth Sci. 6 (2), 122–139.
Melillo, J.M., Reilly, J.M., Kicklighter, D.W., Gurgel, A.C., Cronin, T.W., Paltsev, S., Schlosser, C.A., 2009. Indirect emissions from biofuels: how important? Science 326 (5958), 1397–1399.
Mertens, K.C., Verbeke, L.P.C., Ducheyne, E.I., De Wulf, R.R., 2003. Using genetic algorithms in sub-pixel mapping. Int. J. Remote Sens. 24 (21), 4241–4247.
Meyfroidt, P., 2013. Environmental cognitions, land change, and social–ecological feedbacks: an overview. J. Land Use Sci. 8 (3), 314–367.
Meyfroidt, P., Lambin, E.F., Erb, K.H., Hertel, T.W., 2013. Globalization of land use: distant drivers of land change and geographic displacement of land use. Curr. Opin. Environ. Sustain. 5 (5), 438–444.
Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Wilbanks, T.J., 2010. The next generation of scenarios for climate change research and assessment. Nature 463 (7282), 747–756.
Mu, J.E., McCarl, B.A., Wein, A.M., 2013. Adaptation to climate change: changes in farmland use and stocking rate in the U.S. Mitigation Adapt. Strat. Global Change 18 (6.), http://dx.doi.org/10.1007/s11027-012-9384-4.
NRC, 2014. Advancing Land Change Modeling: Opportunities and Research Requirements. National Academy Press, Washington, DC.
O'Neill, B., Verburg, P., 2012. Spatial land use modeling: simulating decisions and their consequences for climate, carbon, and water, Global Land Project (GLP) News Letter, Issue No 9.
OECD, 2012. OECD Environmental Outlook to 2050: The Consequences of Inaction. Organisation for Economic Co-operation Development, Paris, http://dx.doi.org/10.1787/9789264122246-en.
Olesen, J.E., Bindi, M., 2002. Consequences of climate change for European agricultural productivity, land use and policy. Eur. J. Agron. 16 (4), 239–262.
O'Neill, B.C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., Zigova, K., 2010. Global demographic trends and future carbon emissions. Proc. Natl. Acad. Sci. U.S.A. 107 (41), 17521–17526.
Overmars, K.P., De Koning, G.H.J., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. Ecol. Model. 164 (2), 257–270.
Parker, D.C., Manson, S.M., Janssen, M.A., Hoffmann, M.J., Deadman, P., 2003. Multi-agent systems for the simulation of land use and land-cover change: a review. Ann. Assoc. Am. Geogr. 93 (2), 314–337.
Parry, M., Rosenzweig, C., Iglesias, A., Fischer, G., Livermore, M., 1999. Climate change and world food security: a new assessment. Global Environ. Change 9, S51–S67.
Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P., Fernandez-Manjarrés, J.F., Walpole, M., 2010. Scenarios for global biodiversity in the 21st century. Science 330 (6010), 1496–1501.
Phalan, B., Onial, M., Balmford, A., Green, R.E., 2011. Reconciling food production and biodiversity conservation: land sharing and land sparing compared. Science 333 (6047), 1289–1291.

Pielke, R.A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., de Noblet, N., 2011. Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdiscip. Rev. Climate Change 2 (6), 828–850.
Pitman, A.J., de Noblet-Ducoudré, N., Cruz, F.T., Davin, E.L., Bonan, G.B., Brovkin, V., Voldoire, A., 2009. Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. Geophys. Res. Lett. 36 (14).
Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24 (1).
Ramankutty, N., 2012. Global Cropland and Pasture Data from 1700–2007, Available online at [http://www.geog.mcgill.ca/~nramankutty/Datasets/Datasets.html] from the LUGE (Land Use and the Global Environment) Laboratory, Department of Geography, McGill University, Montreal, Quebec, Canada (accessed on 25.09.12).
Ramankutty, N., Foley, J.A., 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global Biogeochem. Cycles 13 (4), 997–1027.
Ramankutty, N., Foley, J.A., Norman, J., McSweeney, K., 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. Global Ecol. Biogeogr. 11 (5), 377–392.
Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochem. Cycles 22 (1.).
Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. Nat. Commun. 3, 1293.
Reilly, J., Melillo, J., Cai, Y., Kicklighter, D., Gurgel, A., Paltsev, S., Schlosser, A., 2012. Using land to mitigate climate change: hitting the target, recognizing the trade-offs. Environ. Sci. Technol. 46 (11), 5672–5679.
Rokityanskiy, D., Benítez, P.C., Kraxner, F., McCallum, I., Obersteiner, M., Rametsteiner, E., Yamagata, Y., 2007. Geographically explicit global modeling of land use change, carbon sequestration, and biomass supply. Technol. Forecast. Soc. Change 74 (7), 1057–1082.
Ronneberger, K., Tol, R.S.J., Schneider, U.A., 2005. KLUM: A Simple Model of Global Agricultural Land use a Coupling Tool of Economy and Vegetation. Working Paper FNU-65. Hamburg University, Hamburg, Germany.
Ronneberger, K., Berrittella, M., Bosello, F., Tol, R.S., 2009. KLUM@ GTAP: introducing biophysical aspects of land use decisions into a computable general equilibrium model a coupling experiment. Environ. Model. Assess. 14 (2), 149–168.
Rounsevell, M.D., Arneth, A., 2011. Representing human behaviour and decisional processes in land system models as an integral component of the Earth system. Global Environ. Change 21 (3), 840–843.
Rounsevell, M.D.A., Arneth, A., Alexander, P., Brown, D.G., de Noblet-Ducoudré, N., Ellis, E., Finnigan, J., Galvin, K., Grigg, N., Harman, I., Lennox, J., Magliocca, N., Parker, D., O'Neill, B.C., Verburg, P.H., Young, O., 2014. Towards decision-based global land use models for improved understanding of the Earth system. Earth Syst. Dyn. 5, 117–137, http://dx.doi.org/10.5194/esd-5-117-2014.
Sarofim, M.C., Reilly, J.M., 2011. Applications of integrated assessment modeling to climate change. Wiley Interdiscip. Rev. Climate Change 2 (1), 27–44.
Schaldach, R., Priess, J.A., Alcamo, J., 2011. Simulating the impact of biofuel development on country-wide land use change in India. Biomass Bioenergy 35 (6), 2401–2410.
Schlenker, W., Roberts, M.J., 2006. Nonlinear effects of weather on corn yields. Appl. Econ. Perspect. Policy 28 (3), 391–398.
Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., Feick, S., 2005. Development and validation of the global map of irrigation areas. Hydrol. Earth Syst. Sci. 9, 535–547, http://dx.doi.org/10.5194/hess-9-535-2005.
Smith, P., Gregory, P.J., Van Vuuren, D., Obersteiner, M., Havlík, P., Rounsevell, M., Bellarby, J., 2010. Competition for land. Philos. Trans. R. Soc. B: Biol. Sci. 365 (1554), 2941–2957.
Souty, F., Brunelle, T., Dumas, P., Dorin, B., Ciais, P., Crassous, R., Bondeau, A., 2012. The Nexus Land-Use model version 1.0, an approach articulating biophysical potentials and economic dynamics to model competition for land-use. Geosci. Model Develop. 1, 1297–1322.
Souty, F., Dorin, B., Brunelle, T., Dumas, P., Ciais, P., 2013. Modelling economic and biophysical drivers of agricultural land use change. Calibration and evaluation of the Nexus Land use model over 1961–2006. Geosci. Model Develop. Discuss. 6 (4), 6975–7046.
Stehfest, E., Bouwman, L., van Vuuren, D.P., den Elzen, M.G., Eickhout, B., Kabat, P., 2009. Climate benefits of changing diet. Clim. Change 95 (1–2), 83–102.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498.
Tebaldi, C., Hayhoe, K., Arblaster, J.M., Meehl, G.A., 2006. Going to the extremes. Climatic Change 79 (3-4), 185–211.
TEEB, 2010. The Economic of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: a Synthesis of the Approach, Conclusions and Recommendations to the TEEB.
Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Dutta, R., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. Int. J. Remote Sens. 30 (14), 3679–3733.
Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U.S.A. 108 (50), 20260–20264.
UNEP, 2012. Global Environmental Outlook: Environment for the Future We Want, GEO 5. United Nations Environmental Programme.
US Environmental Protection Agency (EPA), 2000. Projecting Land use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land Use Patterns. U.S. Environmental Protection Agency, Cincinnati, OH, Office of Research and Development Publication EPA/600/R-00/098.
van Asselen, S., Verburg, P.H., 2012. A Land System representation for global assessments and land use modeling. Global Change Biol. 18 (10), 3125–3148.
van Asselen, S., Verburg, P.H., 2013. Land cover change or land use intensification: simulating land system change with a global-scale land change model. Global Change Biol., http://dx.doi.org/10.1111/gcb.12331.
van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Rose, S.K., 2011. The representative concentration pathways: an overview. Climatic Change 109 (1–2), 5–31.
van Vuuren, D.P., Bayer, L.B., Chuwah, C., Ganzeveld, L., Hazeleger, W., van den Hurk, B., van Noije, T., O'Neill, B., Strengers, B.J., 2012. A comprehensive view on climate change: coupling of Earth system and Integrated Assessment Models. Environ. Res. Lett. 7 (2), 024012.
Verburg, P.H., Schot, P.P., Dijst, M.J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61 (4), 309–324.
Verburg, P.H., Ellis, E.C., Letourneau, A., 2011. A global assessment of market accessibility and market influence for global environmental change studies. Environ. Res. Lett. 6 (3), 034019.
Verburg, P.H., van Asselen, S., van der Zanden, E.H., Stehfest, E., 2013. The representation of landscapes in global scale assessments of environmental change. Landsc. Ecol., http://dx.doi.org/10.1007/s10980-012-9745-0.
Wang, X., (Ph.D. Thesis) 2008. Impacts of Greenhouse Gas Mitigation Policies on Agricultural Land. Massachusetts Institute of Technology, Cambridge, MA, Available at globalchange.mit.edu/files/document/Wang_PhD_08.pdf (accessed 28.07.13).
Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T., Reick, C., 2014. Comparing the influence of net and gross anthropogenic land use and land cover changes on the carbon cycle in the MPI-ESM. Biogeosci. Discuss. 11 (4), 5443–5469.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 67 (2), 301–320.