https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions
Open Access

Modeling Land Use and Land Cover Change: Using a Hindcast to Estimate Economic Parameters in gcamland v2.0

Katherine V. Calvin[1], Abigail Snyder[1], Xin Zhao[1], Marshall Wise[1]

[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, 20740, USA

Correspondence to: Katherine V. Calvin (katherine.calvin@pnnl.gov)

Abstract. The world has experienced a vast increase in agricultural production since the middle of the last century. Agricultural land area has also increased at the expense of natural lands over this period, though at a lower rate than production. Future changes in land use and cover have important implications not only for agriculture but for energy, water use, and climate. However, these future changes are driven by a complex combination of uncertain socioeconomic, technological, and other factors. Estimates of future land use and land cover differ significantly across economic models of agricultural production, and efforts to evaluate these economic models over history have been limited. In this study, we use an economic model of land use, gcamland, to systematically explore a large set of model parameter perturbations and alternate methods for forming expectations about uncertain crop yields and prices. We run gcamland simulations with these parameter sets over the historical period in the United States to explore model fitness and to identify combinations that improve fitness. We find that an adaptive expectation approach minimizes the error between simulated outputs and observations, with parameters that suggest that for most crops landowners put a significant weight on previous information. Interestingly, for corn, where ethanol policies have led to a rapid growth in demand, the resulting parameters show that a larger weight is placed on more recent information. We conclude with the observation that historical modeling exercises such as this study are valuable both for understanding real world drivers of land use change and for informing modeling of future land use change.

1 Introduction

Between 1961 and 2015, global agricultural production has increased substantially, including more than a tripling of wheat production, a five-fold increase in maize production, and a twelve-fold increase in soybean production (FAO, 2020b). Agricultural area has increased, but by a smaller amount (10% increase in harvested area for wheat, 180% increase for maize, five-fold increase for soybeans), due to increases in agricultural productivity (FAO, 2020b). Total global cropland area has increased by 15% between 1960 and 2015, from 1377 million hectares (Mha) to 1591 Mha (Goldewijk et al., 2017). These changes have resulted in changes in natural land area, including declines in global forest area (Hurtt et al., 2020).

Similar trends occurred in the United States. Crop production in the United States has increased substantially in the last several decades, but much of that increase in production is due to increases in yields (Babcock, 2015; Fuglie, 2010). Total cropland

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions

area in the United States has remained relatively constant between 1975 and 2015. Instead, there has been a shift in crop distribution, with an increasing share of corn and soybeans and a decreasing share of wheat and other grains (Figure 1, (FAO, 2020a; Taheripour and Tyner, 2013)).

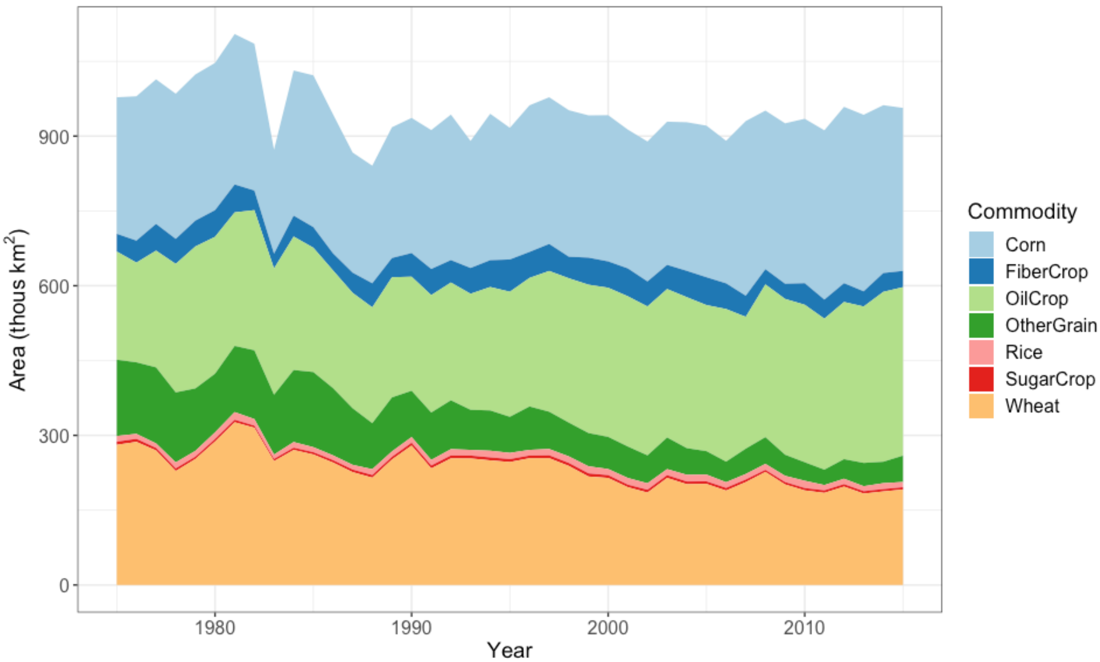

Figure 1: Harvested area by crop for major commodities in the United States (1975-2015). Source: USDA.

Future changes in land use and land cover have implications for agricultural production, energy production, water use, and climate. For example, changes in land cover can alter albedo, resulting in changes in local and global temperature and precipitation (Brovkin et al., 2013; Jones et al., 2013; Manoli et al., 2018). Similarly, changes in land use and land cover have implications for water withdrawals and water scarcity (Bonsch et al., 2016; Chaturvedi et al., 2013; Hejazi et al., 2014a, 2014b; Mouratiadou et al., 2016). However, there is significant uncertainty in the future evolution of land use and land cover, due to uncertainties in future socioeconomic conditions (e.g., population, income, diet) (Popp et al., 2017; Stehfest et al., 2019), technological change (Popp et al., 2017; Tilman et al., 2011; Wise et al., 2014), climate (Calvin et al., 2020; Nelson et al., 2014), and incentives for bioenergy, afforestation, reforestation (Calvin et al., 2014; Hasegawa et al., 2020; Popp et al., 2014, 2017).

Economic models are widely used to estimate future agricultural production and land use, and estimates of future land use and land cover also differ significantly across such models (Alexander et al., 2017; Von Lampe et al., 2014; Popp et al., 2017). These models use economic equilibrium, statistical, agent-based, machine learning, and hybrid approaches (Engström et al., 2016; National Research Council, 2014). Even within each category, there are differences across models, both in terms of

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions

structure and parameters. For example, among economic equilibrium models of land use change (the approach most commonly used in integrated energy-water-land-climate models), some models use constrained optimization (e.g., GLOBIOM), while other models use a non-linear market equilibrium approach (e.g., GCAM) (Wise et al., 2014).

Efforts to evaluate land use models over the historical period are limited. Baldos and Hertel (2013) compare the net change in cropland area, agricultural production, average crop yield, and crop price between 1961 and 2006 simulated by the SIMPLE model to observed changes. Their model matches observations better at the global scale than at the regional scale; additionally, they find that "even knowing yields with certainty does not allow us to predict cropland change accurately over this historical period." Bonsch et al. (2013) compare simulated land-use change CO_2 emissions from MAgPIE to observations, finding that the choice of observation dataset matters for how well the model performs. Calvin et al. (2017) and Snyder et al. (2017) compare agricultural production and land area simulated by the GCAM model to observations, finding that the model does better for trends than annual values and that some region/crop combinations are better than others. The authors test the use of expectations about yield using a linear forecast as a driver of land use change instead of observed yield, finding that simulations using expected yield better match observations than those using observed yield. Engstrom et al. (2016) use a Monte Carlo approach to sample parameters in PLUM, simulating agricultural production and land area over the historical period and comparing results to observations. The authors find the model performs better at larger regional aggregations, but the observed grassland and cereal land area falls outside the full range of their ensemble results. However, most land use models outside of these have not used historical simulations for evaluation/validation.

Only a few studies have attempted to draw land use modeling parameters from econometric estimates of land supply elasticity (Ahmed et al., 2009; Lubowski et al., 2008). However, there is usually no fixed relationship between the land supply elasticities and land use modeling parameters in equilibrium models (Zhao et al., 2020a) and, more importantly, empirically estimated elasticities only provide a limited coverage of regions and land use categories (Barr et al., 2011; Lubowski et al., 2008). Thus, the parameters used in land use models are often based on heuristics (Schmitz et al., 2014). For example, Taheripour and Tyner (2013) group regions into four categories based on historical land use change and assign substitution parameters based on those categories. Wise et al. (2014) choose model parameters to replicate empirically estimated parameters; however, there is no unique mapping between the empirical parameter (constant elasticity of land transformation) and the model parameter (logit exponent). While there are many examples of studies exploring sensitivity to drivers of land use change or sensitivity across models, most studies exclude sensitivity to parameters. The small number of studies that do test alternative parameters find that it could significantly alter land use change (Engström et al., 2016; Taheripour and Tyner, 2013; Zhao et al., 2020b).

In this paper, we use a large perturbed parameter ensemble to determine the model expectation configuration and parameter set that best replicate historical land use and land cover within the United States. Section 2 describes the methodology used in

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

this study. Section 3 introduces the scenarios included in the study. The primary results and sensitivity analyses are discussed in Sections 4 and 5, respectively. Section 6 includes the discussion and conclusions.

2 Methodology

2.1 Land use modelling

2.1.1 gcamland

We use the gcamland v2.0 software package in this study (Calvin et al., 2019a). gcamland separates the land allocation mechanism in GCAM (Calvin et al., 2019b) into an R package. The model calculates land allocation over time; changes in land use and land cover are driven by changes in commodity prices, yields, costs, and subsidies, all of which are inputs into gcamland. gcamland includes all land use and land cover types, with crops aggregated into 12 commodity groups (see (Calvin et al., 2020) for a mapping). gcamland can be run in several different modes, including hindcast and future scenario options and single and multiple ensemble options. For this paper, we utilize the ensemble and hindcast options, generating large ensembles of hindcast simulations. gcamland can be run for any of the 32 geopolitical regions within GCAM, but for this study we focus on the United States.

2.1.2 Economic approach in GCAM and gcamland

Land allocation in gcamland (and GCAM) is determined based on relative profitability, using a nested logit approach (McFadden, 1981; Sands, 2003; Wise et al., 2014). The logit land supply is presented in equation (1). All else equal, an increase in the rental profit rate (r_i) of one land type will result in an increase in the land area (X_i) allocated to that land type. The magnitude of the land supply response is dependent on the positive logit exponent (ρ) and share-weight parameters (λ_i). Y is the total land supply, i.e., $\sum_i X_i = Y$. The logit formulation assumes that there is a distribution of profit rates for each land type, and the resulting land allocation for a given land type is the probability that land type has the highest profit (Zhao et al., 2020b). The logit share-weights (the scale parameters in the distribution) are calibrated to perfectly reproduce the data in a base year. The logit exponent (the shape parameter in the distribution which governs the magnitude of land transformation given relative profit shocks) is one of the parameters of interest in our study.

$$X_i = \frac{(\lambda_i r_i)^\rho}{\sum_j (\lambda_j r_j)^\rho} \cdot Y \ , \tag{1}$$

The logit approach is advantageous compared with the constant elasticity of transformation (CET) approach widely used in GGE models as it can directly provide traceable physical land transformation. But like the CET function, the logit land sharing function is parsimonious and nested structure can be used. In gcamland, all crops are nested under cropland. Cropland is nested with forest and then pasture, see Figure 1 in Calvin et al. (2019a). Thus, there are three logit exponent parameters for governing land transformation in the three-level nesting structure.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions
Open Access

2.1.3 Means of forming expectations

There are multiple means of forming expectations in the literature. With perfect foresight, the expected value of a given variable is equal to its realized value:

$$\mathbb{E}x_t = x_t. \tag{2}$$

In an adaptive expectation approach (Nerlove, 1958), the expected value is a linear combination of the previous expectation and the new information acquired, with α being the coefficient of expectations:

$$\mathbb{E}x_t = (1-\alpha)x_{t-1} + \alpha\mathbb{E}x_{t-1}. \tag{3}$$

Finally, a linear expectation approach uses a linear extrapolation of previous information to form the expectation:

$$\mathbb{E}x_t = \frac{Cov[\boldsymbol{x}(n), \boldsymbol{year}(n)]}{Var[\boldsymbol{year}(n)]} year_t, \tag{4}$$

where n is a fixed number of previous years considered in forming the expectation, $\boldsymbol{x}(n)$ and $\boldsymbol{year}(n)$ are vectors of the variable and year index, respectively, with historical information from year $t-1$ to $t-n$. That is, instead of using all available historical information, forward-looking producers are assumed to rely on only information of the most recent n years.

In our study, we combine these basic approaches into four different expectation types, specifying the means of calculating expected price and expected yield (Table 1).[1] The expected prices and yield would affect farmers' expected rental profits and, thus, land use decisions.

Table 1: Expectation types tested in this study.

Expectation type	Price expectations	Yield expectations	Examples in the literature
Perfect	Perfect expectations	Perfect expectations	All integrated models and most agriculture economic models
Adaptive	Adaptive expectations	Adaptive expectations	(Fémenia and Gohin, 2011; Lundberg et al., 2015; Mitra and Boussard, 2012; Roberts and Schlenker, 2013)
Linear	Linear expectations	Linear expectations	(Calvin et al., 2017; Snyder et al., 2017)
Hybrid Linear Adaptive	Adaptive expectations	Linear expectations	Tested in this paper

[1] Note that other expectation types can be tested within gcamland, e.g., expectation types that are a hybrid of past and perfect information. Such expectations types can be useful for understanding the value of additional information. However, we exclude them in this paper as they are unlikely to explain past behavior and are not covered in the literature on land use decision making.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions

2.2 Perturbed parameter ensemble

Within this study, we vary a total of nine parameters (Table 2), including three logit exponents, the coefficient of expectations (α) for the *Adaptive* expectation and the number of years (n) used in the *Linear* expectation. In addition, we allow α and n to vary across commodity groups, resulting in three separate realizations for each parameter. We group the commodities to minimize the number of free parameters. The first group includes Corn and OilCrop, which are used for biofuels in the United States and have had shifts in the demand over time as a result of biofuels policies. The second group includes the other two large commodities produced in the United States, Wheat and OtherGrain. The third group includes all other crops. The range of values spanned in the ensemble was chosen to cover all plausible values of each parameter but avoid potential numerical instabilities. Those ranges and their justification are described in Table 2. We use a Latin Hypercube Sampling strategy to generate the ensembles, with 10,000 ensemble members per expectation type and model configuration.

Table 2: Parameters perturbed in this study, including the range of values tested.

Type	Parameter	Description	Range	Rationale for Range
Logit	Arable Land	Logit exponent dictating competition between all arable land types within GCAM	0.01-3	The minimum value is chosen to be close to zero (which would result in no shifts in land) but without causing numerical instability. Very large logit exponents result in winner-take-all behavior (Wise et al., 2014; Zhao et al., 2020b). Such behavior may be reasonable at small scale but not for the United States as a whole, so an upper bound of 3 is chosen to prevent this.
	Ag, Forest, and Other	Logit exponent dictating competition between crops, forest, and grass/shrubs		
	Cropland	Logit exponent dictating competition among crops		
Share of Past Information	Corn, OilCrop	Weight on previous expectations (α) for Corn and OilCrop in the *Adaptive* expectations	0.1-0.99	Parameter is restricted to the range [0, 1]. A value of 1 would keep expected profit constant at its initial value, so we choose a value slightly smaller for the upper bound. Very small values of this parameter have been shown to result in divergence of the system (Fémenia and Gohin, 2011).[2] A lower bound of 0.1 is chosen to prevent this.
	Wheat, OtherGrain	Weight on previous expectations (α) for Wheat and OtherGrain		

[2] Note that Femenia and Gohin (2011) define their parameters differently than is done in this paper. Thus, an α value of 1 in their study is equivalent to a value of 0 here.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

		in the *Adaptive* expectations		
	All Other Crops	Weight on previous expectations (α) for all other crops in the *Adaptive* expectations		
Number of years	Corn, OilCrop	Number of previous years (n) used in the linear extrapolation in the *Linear* expectations for Corn and OilCrop	2-25	Linear extrapolation is undefined for values less than 2. Only integer values allowed.
	Wheat, OtherGrain	Number of previous years (n) used in the linear extrapolation in the *Linear* expectations for Wheat and OtherGrain		
	All Other Crops	Number of previous years (n) used in the linear extrapolation in the *Linear* expectations for all other crops		

2.3 Evaluating model performance

Following the analysis of previous hindcasting experiments with GCAM (Snyder et al., 2017), different measures of model performance are used to select parameter sets that optimize different aspects of model performance.

Normalized root mean square error (NRMSE) considers all deviations between simulated and observed values, and places them in the context of the variance seen in the observational data. For crop *i*,

$$NRMSE_i = \frac{\sqrt{mean_i\left(obs_{i,t}-sim_{i,t}\right)^2}}{\sqrt{mean_i\left(obs_{i,t}-\overline{obs_i}\right)^2}}. \tag{5}$$

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

One benefit of this measure is that it includes a natural benchmark of acceptable model performance. While $NRMS = 0$ corresponds to perfect model performance, any $NRMS < 1$, is considered acceptable model performance. This measure puts the deviations between simulation and observation for each crop in the context of that crop's historical variance.

We also consider the root mean square error (RMSE),

$$RMSE_i = \sqrt{mean_i\left(obs_{i,t} - sim_{i,t}\right)^2}\,, \tag{6}$$

and bias,

$$bias_i = (\overline{obs_i} - \overline{sim_i}\,). \tag{7}$$

These un-normalized measures make no distinction between different crops; a bias or RMSE of 200 km^2 means exactly the same for Corn as it does for Rice, despite the fact that Corn represents a larger proportion of harvested area in the United States in the historical period. While RMSE is concerned with all deviations between observation and simulation for a crop, bias simply compares the means between observation and simulation. While these means tend to be determined more by the smoothed trend in a time series than variations about the trend, it is important to note that bias specifically does not penalize volatility the way that RMSE and other measures may.

Finally, the Kling-Gupta Efficiency score (Knoben et al., 2019) is also implemented for each crop,

$$KGE_i = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}\,, \tag{8}$$

for correlation coefficient *r,* standard deviation σ, and mean μ). While a perfect simulation (NRMSE=RMSE=bias=0) would by definition give perfect KGE (KGE=1), KGE is obviously more defined by penalties between different time series summary statistics, as opposed to the penalties based on simple deviations between simulation and observation at each time point in the other error metrics considered here.

For a given error measurement, the metric is calculated for each crop in each ensemble member. The average metric value across crops is then minimized to select the ensemble member with the most optimal parameters for matching observation. For bias, it is the average across crops of the magnitude of bias that is minimized, to avoid cancellation of errors between crops. For KGE, it is the average across crops of the quantity $1 - KGE_i$ that is minimized so that the average across crops of KGE_i is optimized as needed. As an additional sensitivity, the actual land types included in this average metric can be adjusted to include all crops, simply one individual crop, or any combination of crops of interest.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

2.4 Data

2.4.1 Initialization data

To initialize gcamland in this study, we started from the GCAM v4.3 agriculture and land use input data. The GCAM data processing reconciles land use data from FAO with land cover data, ensuring that total areas do not exceed the amount of land in a region. Thus, we chose to use this reconciled data instead of using FAO data directly. We have made two changes to the GCAM v4.3 initialization data.

First, since GCAM has a five-year time step, it uses five-year averages of land use and agricultural production for initialization. For this study, we have updated the input data to remove the averaging since we are primarily focused on annual time steps; that is, the initialization data in gcamland for a particular year is the data for that year only and not a five-year average around that year as it is in GCAM.

Second, GCAM models land use and land cover at the subnational level (v4.3 used Agro-Ecological Zones; v5.1 and subsequent versions use water basins). However, much of the comparison data is provided at national level. For this study, we aggregate the initialization data to the national level, representing the USA as a single region. The qualitative insights in this paper would not change if we disaggregated to subnational level, but the exact quantitative results would.

Third, GCAM uses constant costs over time. For this study, we have updated the costs to use time-evolving cost data (see next section).

2.4.2 Scenario data

We use data for price and yield from the U.N. Food and Agricultural Organization (FAO, 2018a, 2018b, 2020b), with data available for all commodities for 1961-2018. Data was aggregated from individual crops to the GCAM/gcamland commodity groups, weighting crops by their production quantity. In some cases, data prior to 1961 are required to generate expectations for the model years (1975-2015); in these cases, we assume that prices and yields prior to 1961 are held constant at their 1961 values. For cost, we use data provided by the U.S. Department of Agriculture (USDA, 2020a), with data available for major crops from 1975-2018. We only include the variable costs as reported by USDA and exclude the allocated overhead costs. We use a representative crop from USDA for each GCAM/gcamland commodity group, as data does not exist for all crops (i.e., we use soybean cost from USDA as a proxy for the cost of OilCrop in gcamland). For subsidies, we combine two different data sets from USDA: the Federal government direct payments (USDA, 2020c) and the farm business income (USDA, 2020b). We only include direct payments from these two reports; thus, our subsidy data is missing many other forms of payment. Additionally, we only have data for a subset of crops and the categories reported change over time across the two data sets.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Because this data is inconsistent and incomplete, we only use it as a sensitivity in this paper and do not include it in the primary analysis.

2.4.3 Observation data

We compare model outputs to observation data to evaluate the performance of gcamland. Ideally, the observation data would be completely independent of the model. However, due to limited availability of data sets,[3] we use the FAO harvested area for crops as the observational dataset, despite the fact that it is used to calibrate gcamland. For calibration, we only use a single year of data, so the comparison to the FAO time series is still valid. For land cover, an independent data set is available for use in gcamland; specifically, we use satellite data from the European Spatial Agency (ESA) Climate Change Initiative (CCI), as reported by the FAO (FAO, 2020a) and aggregated to the gcamland land cover classes. Due to differences in definitions and classifications, the grassland and shrubland reported by CCI differ substantially from the gcamland areas even in the calibration year.

3. Simulations and sensitivities

The default ensemble analyzed in this paper uses 1990 as a calibration year, runs annually through 2015, excludes subsidies, and differentiates the expectation parameters (α and n) by crop groups. To test the sensitivity of the results to each of these assumptions, we re-run the ensemble for with alternative specifications for each assumption (Table 3).

Table 3: Model specifications used in this study.

Name	Calibration Year	Time Step	Subsidies?	Parameters differentiated by crop group?
Default	1990	Annual	No	Yes
Same parameters	1990	Annual	No	No
With subsidy	1990	Annual	Yes	Yes
1975	1975	Annual	No	Yes
2005	2005	Annual	No	Yes
5 year timestep	1990	Five year	No	Yes

Over the last several decades, yields have increased in the United States; prices and profits are more variable (Figure S1). Changes in the area of a particular crop, however, are not always correlated with in year profit (Figure S2, S3). There are several potential reasons for this:

1. Farmers do not know the profit at the time of planting and instead are basing their planting decisions on expectations.

[3] The only other data set we are aware of the provides a time series of cropland area by crop is the USDA. However, since FAO basis their reporting for the United States on submissions from the USDA, these two datasets are identical.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions

2. The profit calculated here is missing some other factor (e.g., a government subsidy).
3. Profit relative to another commodity may be better predictor (e.g., if two crops have increases in profit, a farmer might shift to the one with faster increases, resulting in a decline in land area for the other despite its increase in profit).
4. Different crops may have undergone very different improvements in yields over time.
5. Other non-economic factors (e.g., distance to markets) might drive land use decisions.

We explicitly test the first two explanations in this paper. The third and fourth are captured in all of our simulations. The fifth is implicitly captured in the calibration routine in gcamland but we do not vary this over time.

4. Results

This section describes the results from the default gcamland ensemble. This ensemble assumes a calibration year of 1990, an annual time step, subsidies are excluded, and the parameter sets are chosen to minimize the average NRMSE across all crops. Sensitivity to each of these assumptions is presented in the next section. Note that throughout the results and sensitivity sections the default configuration, with the numerically optimal parameter set and expectation type are shown in thick magenta lines for consistency.

4.1. Parameter Sets that Minimize NRMSE in gcamland

NRMSE varies across expectation types, ranging from 1.399 with Adaptive expectations to 1.874 with Linear expectations. The parameters that minimize NRMSE vary by expectation type (see Figure 2), including the ordering of the logit exponents. In the Adaptive expectations, the logit exponent dictating substitution among crops is larger than the logit exponents determining substitution between crops and other land types. This rank ordering of logit exponents is consistent with the intuition from historical trends in USA land allocation (Figure 1); specifically, the larger changes in crop mix than total crop area in the observations suggest that the logit for the cropland nest should be larger than the other logits. In all models with imperfect expectations, expected profits are heavily weighted toward previous information, as evidenced by the large values for the share of past information and the number of years in the linear forecast. However, these values vary across crop groups. For example, Corn and OilCrop rely less on past information than other crops in the Adaptive expectations and for prices in the Hybrid Linear Adaptive expectations, likely due to changes in the market due to the introduction of biofuels policies circa 2005.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

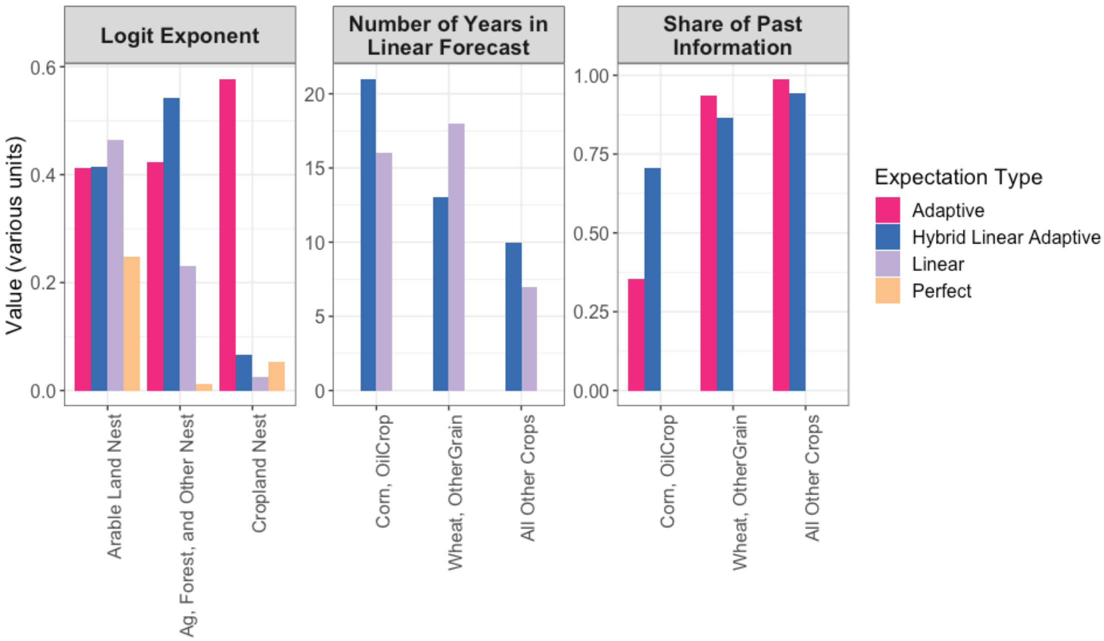

Figure 2: Parameters that minimize NRMSE by Expectation Type.

4.2. Comparing modeled land area to observations

The full ensemble of gcamland simulations results in a large range of land allocated to crops, covering +/-100% of the observed area. The parameter sets that minimize NRMSE in gcamland replicate total harvested cropland area over time in the United States fairly well (Figure 3, left panel). However, gcamland misses some of the transitions in crops shown in Figure 1. In particular, for Adaptive expectations (the numerically optimal expectation type and parameter set), gcamland underestimates the growth in OilCrop in the mid-1990s and overestimates the growth in Corn in recent years (Figure 3). The insights from Figure 3 are confirmed when examining the crop-specific NRMSE in this simulation. The NRMSE for Corn and OilCrop are larger (1.88 and 1.67, respectively) than the NRMSE for other Wheat and OtherGrain (1.16 and 0.7, respectively). A scatter plot of this comparison for all GCAM crop types is included in the supplementary material (Figure S6). Time series of the cropland share over time for these four crop are also included in the supplementary material (Figure S8).

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development
Discussions
Open Access
EGU

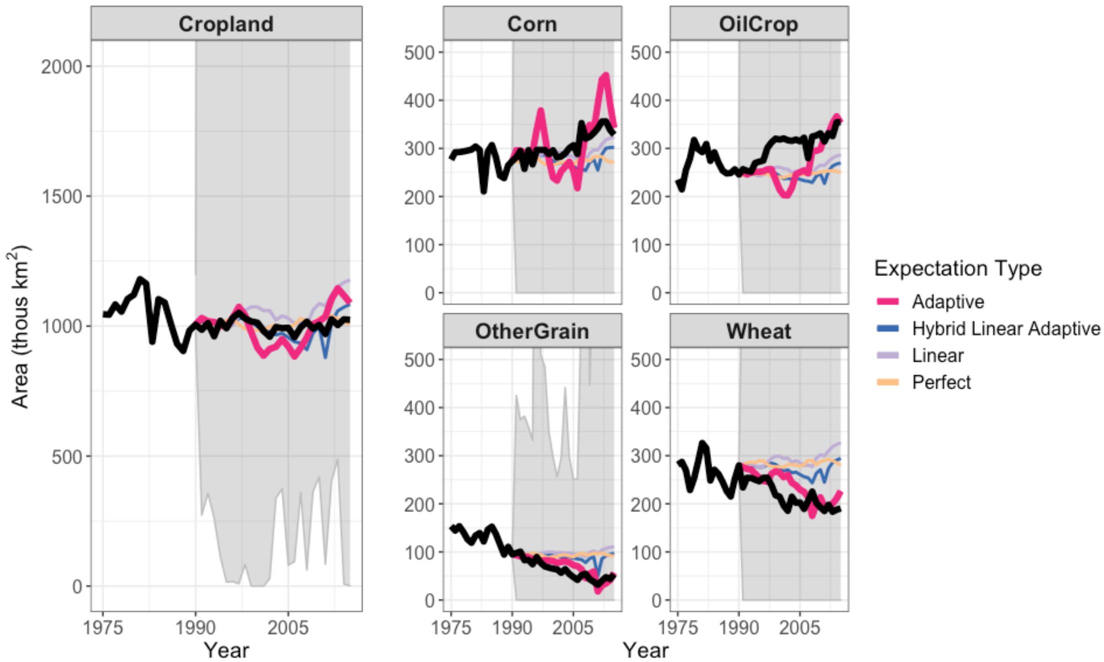

Figure 3: Harvested cropland area (total and by crop) over time by expectation type. Black line is observations (FAO). Colored lines are gcamland results for the models that minimize NRMSE. The expectation type with the minimum NRMSE (Adaptive) is shown with a thicker line. Gray area is the range of all gcamland simulations. Note that fodder crops are excluded total cropland area in this figure due to data limitations.

5. Sensitivity Analysis

In this section, we describe the sensitivity of the results above to several different assumptions, including those related to the configuration of the model, the calibration year, the model timestep, and the objective function used. For the model configuration, calibration year, and timestep sensitivities, we generate new ensembles of gcamland results with the appropriate assumption altered. For the sensitivity to objective function, we filter the original ensemble using different criteria to determine the numerically optimal parameter sets.

5.1. Sensitivity to Model Assumptions

First, we test the sensitivity of the analysis to two different assumptions: (1) whether subsidies are included in the expected profit for crops, and (2) whether the expectation-related parameters differ across crops. For all three sets of assumptions, Adaptive expectations minimizes NRMSE. Varying these assumptions results in differences in parameters (Table S2) and cropland area (Figure 4). Including subsidies increases the NRMSE (from 1.399 in the Default case to 1.46 with subsidies). This is due to the quality of the subsidy data. Including all factors that affect profit should improve the model; however, the subsidy data is incomplete (only direct payments were included for crops where these were reported) and inconsistent (reporting changed over time). However, previous studies have shown that direct payments have little effect on crop production

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

or land area in the United States (Weber and Key, 2012), suggesting that better subsidy data may not change land allocation decisions substantially.

Using the same expectation parameters across commodity groups increases NRMSE (from 1.399 in the Default case to 1.531 with uniform parameters). There are several reasons why different crops could require different parameters. First, one would expect differences between annual and perennial crops due to the lag between planting and harvesting and the multi-year investment required by perennial crops. Second, some crops (e.g., Corn and OilCrop) have had shifts in policy or demand over time (e.g., for biofuels). Such shifts may lead landowners to prioritize newer information. Finally, there could be differences in how markets are structured (e.g., futures contracts) or region-specific differences. These effects are difficult to disentangle in gcamland. Perennial crops are all included in the “All other crops” group. This group is a mix of both perennial and annual, but we do see higher shares of past information in this group than in the other commodity groups in the Default model. Corn and OilCrop rely more heavily on new information when parameters vary, which is consistent with the market shifting hypothesis.

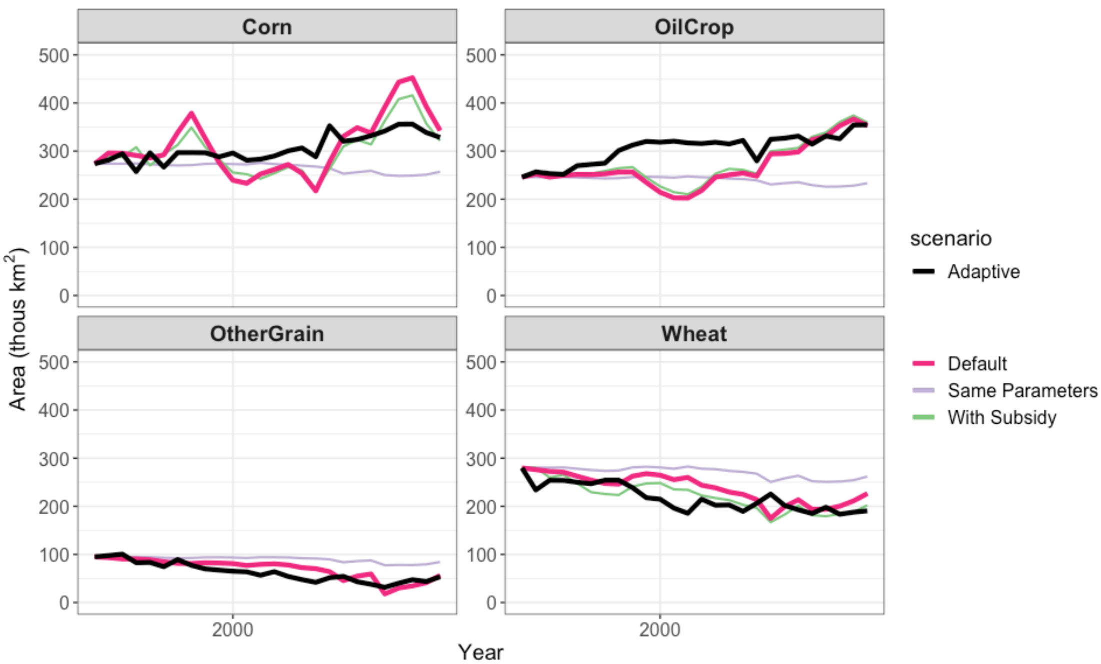

Figure 4: Harvested area by crop under different model assumptions. Black line is observations (FAO). Colored lines are gcamland results for the models that minimize NRMSE.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development Discussions

5.2. Sensitivity to the Objective Function

The analysis above uses the average NRMSE across all crops as an indicator of "goodness of fit", but other objective functions are possible. In this section, we discuss alternative measures of "goodness of fit", including bias, RMS, and KGE. Additionally, we examine the implications of minimizing NRMSE for an individual crop as opposed to the full set of crops.

5.2.1. Optimizing for different objective functions

The parameter sets (Table S3) and cropland time series (Figure 5) that are numerically optimal for KGE is somewhat similar to those of NRMSE and the parameter set that minimizes RMSE is identical to that of NRMSE.[4] The NRMSE and RMSE minimize objective function values with the Adaptive expectation, while the KGE minimizes values with the Hybrid Linear Adaptive expectation. All three rely less on past price information for Corn and OilCrop (share ranges from 0.36 with NRMSE and RMSE to 0.6 with KGE) than for all other crops (share of past information > 0.925). The logit exponents are relatively small (0.05 to 0.58 across all three objective functions and all three nests), with modest substitution allowed in the cropland nest (logit exponent of 0.36 in KGE and 0.58 in NRMSE and RMSE).

The parameter set that minimizes bias, however, is fundamentally different. The logit exponents dictating the substitution between crops and other land types are large (2.17 for the Arable Land nest; 1.37 for the Ag, Forest, and Other nest). The parameter set that minimizes bias also includes the lowest Cropland nest logit value of any objective function (0.28). The resulting simulations for bias exhibit large volatility in land area. Given that bias simply compares the model mean across time to the observation mean across time, this volatility is not penalized in the bias metric, whereas it is penalized for KGE, RMSE and NRMSE. For example, the parameter sets that minimize bias result in an average observed Corn area of 306 and an average simulated Corn area of 307, resulting in a bias of less than 1 thous km^2. This bias is much lower than the bias for Corn in the other objective functions (NRMSE and RMSE have a bias of 6 thous km^2; KGE has a bias of 16 thous km^2). Bias is effectively assessing whether the model is correct on average and not whether it captures the trends or volatility; such an objective function is less useful in systems where trends are significant or where the goal is to capture the volatility. From a mechanistic perspective, we hypothesize that the difference in the bias cropland area volatility is due to the differences in the Ag, Forest, and Other logit.

[4] Note that this is not true in general, but is true for the Default model. Other configurations of the model have different parameter sets that minimize NRMSE than those that minimize RMSE.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

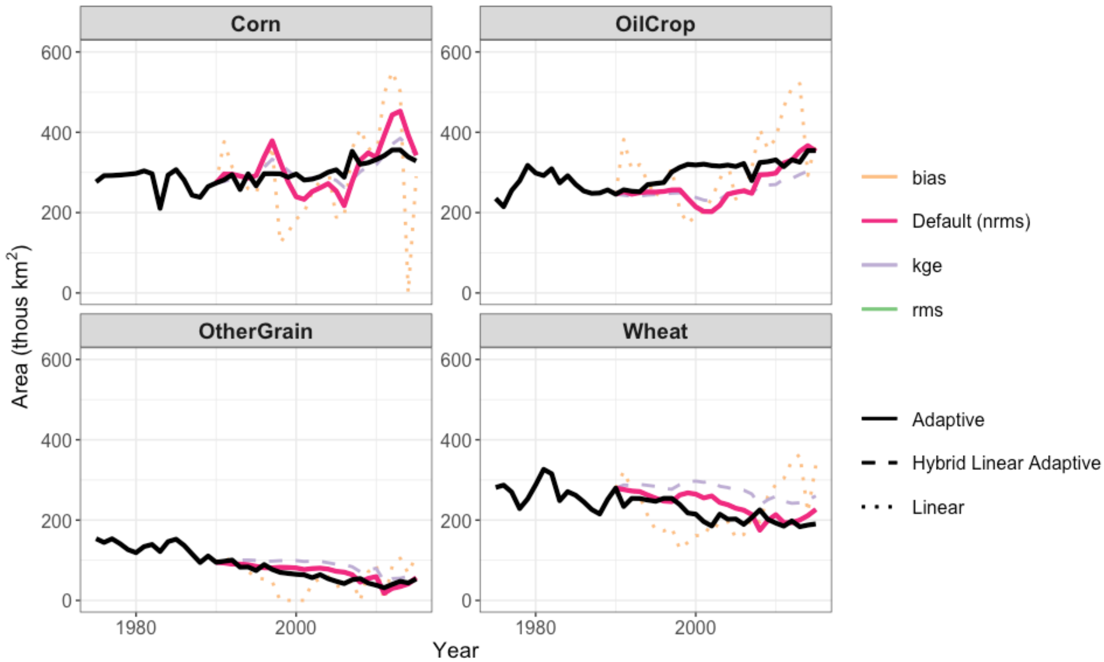

Figure 5: Harvested area by crop when optimizing for different objective functions. Colors indicate objective function. Line type indicates the expectation type that minimizes that objective function. Only the objective function minimizing expectation type is shown. Note that NRMSE and RMSE result in identical parameter sets in the default model and thus have identical land allocation in this figure.

5.2.2. Optimizing for different land types

Figure 6 shows the difference in the best models when you optimize for a particular set of land types or crops. As seen in this figure, gcamland can track land area for any given crop very well when the ensemble with optimal parameters is chosen specifically for that crop. However, matching all crops at once is more challenging. For example, the parameter sets that minimize NRMSE for Corn result in an excellent match between observations and model output for Corn; however, those parameters result in an overestimation of Wheat land by 250 thous km^2 in 2015 (or ~1/2 of the actual area). The insights from this figure are also confirmed numerically. The NRMSE for Corn is reduced from 1.88 to 0.72 when we go from minimizing NRMSE across all crops to minimizing NRMSE for Corn only. Similarly, the NRMSE for OilCrop is reduced from 1.67 to 0.54 when we go from minimizing NRMSE across all crops to minimizing NRMSE for OilCrop only. Optimizing for a single crop has less effect on the NRMSE for Wheat and OtherGrain (from 1.16 to 0.79 for Wheat, and from 0.7 to 0.43 for OtherGrain).

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

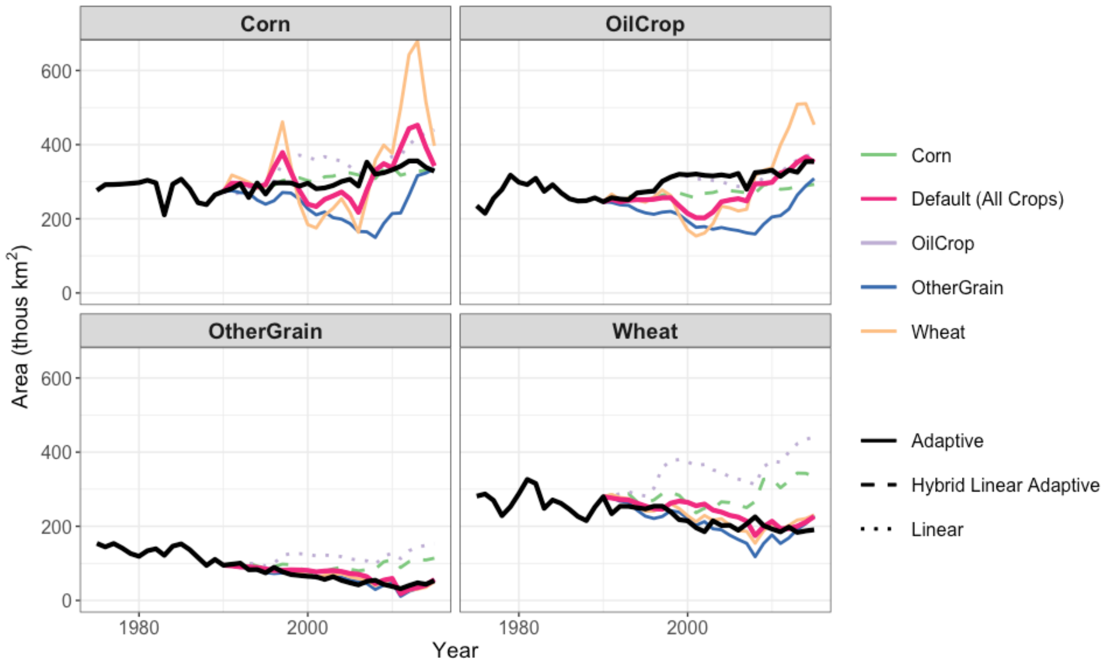

Figure 6: Harvested area by crop when optimizing for different land types. Colors indicate crops included in the objective function. Line type indicates the expectation type that minimizes NRMSE for that set of crops. Only the NRMSE minimizing expectation type for each set of crops is shown.

5.3. Different calibration years

The calibration routine in gcamland calculates share weight parameters (λ_i in equation 1) to ensure that the land area is exactly replicated in the specified base year. Those parameters are held constant in all subsequent periods. Changing the base year could result in different calibration parameters and thus different land allocation, even if all other parameters are the same. In this section, we test this sensitivity, using 1975 and 2005 as alternative base years. Figure 7 shows the difference in cropland area for the parameter sets that minimize average crop NRMSE for each calibration year. Those parameter sets are shown in Table S4. The resulting parameters and land use are relatively similar between variants with base years of 1990 (the default described above) and 2005. The logit exponents are small for all three nests, with the largest value over the cropland nest. Both models use more past information for All Other Crops than for Corn and OilCrop, but they differ in the degree of past information used for Wheat and OtherGrain. The variant with a 1975 calibration year, however, has large differences in parameters and behavior from those with 1990 and 2005 calibration years. We hypothesize two reasons for these differences. First, we have a limited time series prior to 1975, which results in erroneous estimates of expected price and expected yields for parameter sets with large reliance on past information. Second, there is a discrepancy between FAO harvested area and the land cover data sets used in GCAM in 1975 (this discrepancy exists but is much smaller from 1990 onwards). In particular, FAO harvested area is larger than the physical crop area. We correct for this in gcamland by assuming that some areas are planted more than once in a year. However, this results in larger annual yields in gcamland than the harvest yield provided by

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

FAO. This results in higher profit rates that could affect the land allocation. Note that this issue is not a problem in future simulations, like those typically run in GCAM, since the calibration information used in future periods is the information calculated from a more recent year without these data challenges (2010 or 2015 depending on the version of GCAM).

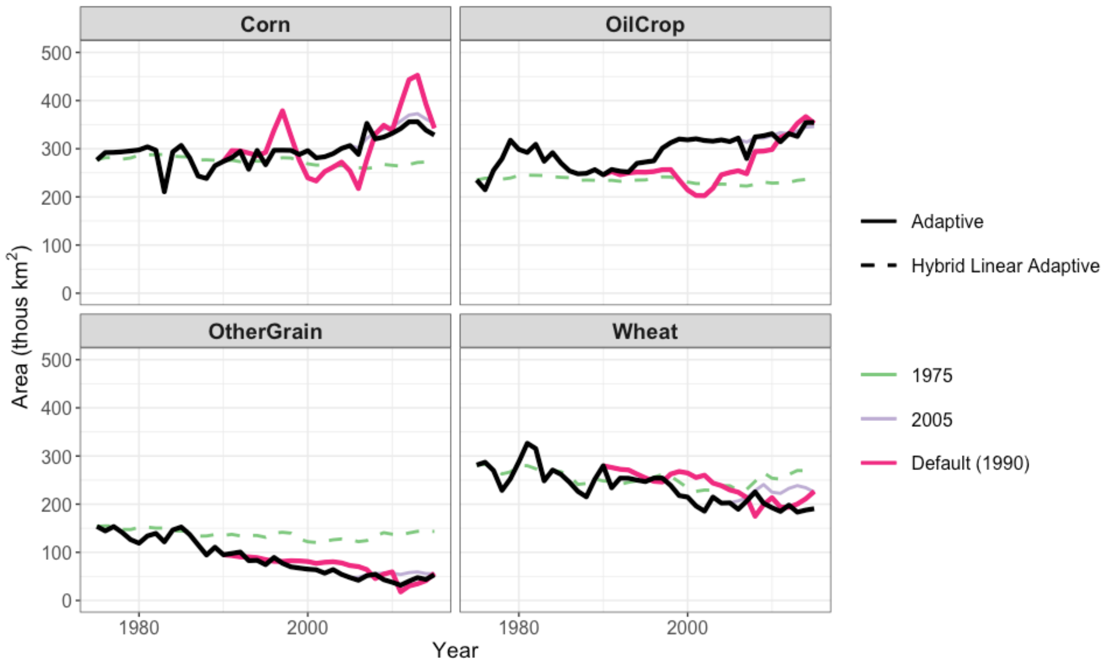

Figure 7: Harvested area by crop when using different calibration years. Colors indicate calibration year. Line type indicates the expectation type that minimizes that NRMSE for that calibration year. Only the NRMSE minimizing expectation type for each calibration year is shown.

5.4. Short-run versus long-run parameters

Finally, we examine the sensitivity of the results to the time step. Most studies using GCAM use a five-year time step with perfect expectations. However, we are increasingly interested in quantifying the implications of climate variability and change on agriculture, land use, and the coupled human-Earth system, which requires higher temporal resolution. For purposes of this comparison, we focus on RMSE instead of NRMSE. NRMSE and RMSE differ in that NRMSE is normalized by standard deviation; however, the inclusion of standard deviation introduces inconsistencies when comparing across time steps. For the Default model, the choice of RMSE or NRMSE has no effect on results, but for the five year time step it does. We note any differences that would emerge from using NRMSE in this discussion.

Our hypothesis was that longer time steps would result in larger logit exponents since farmers would have more time to make adjustments and that expectations would matter less with longer time steps. Using RMSE, the former is true, but the latter is

Geoscientific Model Development Discussions

not.[5] The five-year timestep results in higher logit exponents, particularly in the Arable Land nest and the Cropland nest; the expectation parameters are similar though (Figure S11). However, the Hybrid Linear Adaptive expectations minimizes RMSE in the five-year time step model (Table 4), suggesting that expectations are still important for longer time steps.[6] We find that the one-year time step results in a lower RMSE than the five-year time step model, even when the differences in comparison years are taken into account (Table 4): the RMSE computed over five-year increments in the one-year model is still lower than the RMSE in the five-year model. In the five-year time step model, farmers use five-year averages of price and yield when forming expectations. As a result, the five-year time step model will produce different expectations (Figure S14) and different land allocation results (Figure S15) than the one-year time step model even when the same parameters are used. The fact that annual time steps reduces RMSE suggests that interannual variability may have a noticeable influence on expectations and the resulting land allocation; that is, farmers consider not just the trend in yield and price but also the variability around that trend. This is particularly true for Corn and OilCrop where more recent information has a larger effect on expectations.

Table 4: The effect of time step, expectations, and comparison years on RMSE

Time Step	Expectations	Comparison Years	RMSE
1 year*	Adaptive	Annual, 1990-2015	16.1
1 year	Adaptive	5-year increments, 1990-2015	14.5
5 year	Hybrid Linear Adaptive	5-year increments, 1990-2015	18.7
5 year	Perfect	5-year increments, 1990-2015	25

* This variant is equivalent to the Default shown earlier in the paper.

6 Discussion and Future Work

In this paper, we have used a perturbed parameter ensemble of simulations of land use and land cover over the historical period to guide the selection of parameters for an economic model. We find that adaptive expectations minimize the error between simulated outputs and observations, consistent with empirical evidence (Mitra and Boussard, 2012). The resulting parameters suggest that for most crops, landowners put a significant weight on previous information. For Corn and OilCrop, however, a large weight is placed on more recent information. This is consistent with an observation by Kelley et al. (2005): "In the case of agriculture, anecdotal evidence suggests that some farmers are more myopic, weighing recent information more than is efficient."

[5] Using NRMSE, the logit exponents are slightly smaller in the five-year timestep model than in the one-year timestep model (Figure S10), but expectations reduces error in the five-year timestep model under both RMSE and NRMSE.
[6] With NRMSE, Adaptive expectations minimizes error. Like RMSE, we still find that expectations are important for longer timesteps.

The optimal set of parameters is sensitive to the choice of objective function, with differences emerging either when the mathematical formulation of the error is altered or when the set of land types included in the calculation of error is changed. For the former, we find that using bias as an objective function leads to the largest volatility in annual land allocation. While GCAM has historically performed better at capturing overall trend behavior than annual variations and this has been considered acceptable model behavior (Calvin et al., 2017; Snyder et al., 2017), the results of this study highlight the importance of penalizing variations about the trend as well. For the latter, it is possible to significantly improve the performance for the model for any single crop by optimizing for that crop; however, the resulting parameters may lead to a larger error for a different crop. For example, the parameter sets that minimize NRMSE for Corn result in an excellent match between observations and model output for Corn; however, those parameters result in an overestimation of Wheat land by approximately 250 thous km^2 in 2015.

We hypothesize that limitations of data affect the performance of some variants of the model. For example, the variant that explicitly excludes subsidies outperforms the one with subsidies, likely due to the poor quality of the subsidy data. Similarly, the variant with 1975 as a calibration year is fundamentally different from the variant with 1990 or 2005 as a calibration year, likely due to discrepancies between harvested and physical area in 1975 and limited availability of the data prior to 1975 that is needed to form expectations. Similarly, the land cover data provides little constraint on the model due to the short time series and difference in definitions of land category. Future work could include improvements in the data and the addition of new data sets to constrain the model. Additionally, we have focused on the United States, using national level data; future work could replicate this analysis for subnational regions or for other countries around the world. Our expectation is that we would find different parameters best replicate observations in other countries, similar to what is asserted in Taheripour and Tyner (2013).

Other potential research directions include testing other assumptions in the model (e.g., the nesting structure), new explanatory variables (e.g., crop insurance, speculative storage), alternative decision-making frameworks (e.g., non-logit approaches), or additional behavioral processes (e.g., learning, diffusion). For the nesting structure, we have only tested the default GCAM nesting structure here. Taheripour and Tyner (2013) test an alternative nest and find that it has implications for the share of forest cover (14% vs. 3% depending on the nest). For explanatory variables, studies have indicated that some programs, like crop insurance, are likely to have a direct impact on area planted and production (Young and Westcott, 2000). For alternative decision-making frameworks, Zhao et al. (2020b) demonstrate that the resulting change in land use due to a shock differs depending on the combination of functional form (logit, constant elasticity of transformation, constrained optimization) and parameter value.

In this paper, we have focused on the historical period. However, these models and parameter estimates could be used in a simulation of future land use and land cover change to better understand their implications. Additionally, given the connections between energy, water, land, and climate, using these parameters, and the uncertainty around them, in the fully coupled GCAM would be useful in the future.

References

Ahmed, S. A., Hertel, T. and Lubowski, R.: Calibration of a land cover supply function using transition probabilities, Purdue, Indiana. [online] Available from: https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=2947, 2009.

Alexander, P., Prestele, R., Verburg, P. H., Arneth, A., Baranzelli, C., Batista e Silva, F., Brown, C., Butler, A., Calvin, K., Dendoncker, N., Doelman, J. C., Dunford, R., Engström, K., Eitelberg, D., Fujimori, S., Harrison, P. A., Hasegawa, T., Havlik, P., Holzhauer, S., Humpenöder, F., Jacobs-Crisioni, C., Jain, A. K., Krisztin, T., Kyle, P., Lavalle, C., Lenton, T., Liu, J., Meiyappan, P., Popp, A., Powell, T., Sands, R. D., Schaldach, R., Stehfest, E., Steinbuks, J., Tabeau, A., van Meijl, H., Wise, M. A. and Rounsevell, M. D. A.: Assessing uncertainties in land cover projections, Glob. Chang. Biol., 23(2), doi:10.1111/gcb.13447, 2017.

Babcock, B. A.: Extensive and Intensive Agricultural Supply Response, Annu. Rev. Resour. Econ., 7(1), 333–348, doi:10.1146/annurev-resource-100913-012424, 2015.

Baldos, U. L. C. and Hertel, T. W.: Looking back to move forward on model validation: Insights from a global model of agricultural land use, Environ. Res. Lett., 8(3), doi:10.1088/1748-9326/8/3/034024, 2013.

Barr, K. J., Babcock, B. A., Carriquiry, M. A., Nassar, A. M. and Harfuch, L.: Agricultural Land Elasticities in the United States and Brazil, Appl. Econ. Perspect. Policy, 33(3), 449–462, doi:10.1093/aepp/ppr011, 2011.

Bonsch, M., Dietrich, J. P., Popp, A., Lotze-Campen, H. and Stevanovic, M.: Validation of land use models, in GTAP Conference, Shanghai, China. [online] Available from: https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4064, 2013.

Bonsch, M., Humpenöder, F., Popp, A., Bodirsky, B., Dietrich, J. P., Rolinski, S., Biewald, A., Lotze-Campen, H., Weindl, I., Gerten, D. and Stevanovic, M.: Trade-offs between land and water requirements for large-scale bioenergy production, GCB Bioenergy, doi:10.1111/gcbb.12226, 2016.

Brovkin, V., Boysen, L., Arora, V. K., Boisier, J. P., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., Van den hurk, B. J. J. M., Hurtt, G. C., Jones, C. D., Kato, E., De noblet-ducoudre, N., Pacifico, F., Pongratz, J. and Weiss, M.: Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century, J. Clim., 26(18), 6859–6881, doi:10.1175/JCLI-D-12-00623.1, 2013.

Calvin, K., Wise, M., Kyle, P., Patel, P., Clarke, L. and Edmonds, J.: Trade-offs of different land and bioenergy policies on the path to achieving climate targets, Clim. Change, 123(3–4), 691–704, doi:10.1007/s10584-013-0897-y, 2014.

Calvin, K., Wise, M., Kyle, P., Clarke, L. and Edmonds, J.: A hindcast experiment using the GCAM 3.0 agriculture and land-use module, Clim. Chang. Econ., 8(1), doi:10.1142/S2010007817500051, 2017.

Calvin, K., Link, R. and Wise, M.: gcamland v1.0 – An R Package for Modelling Land Use and Land Cover Change, J. Open

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Res. Softw., 7, 1–6, doi:10.5334/jors.233, 2019a.

Calvin, K., Mignone, B. K., Kheshgi, H. S., Snyder, A. C., Patel, P. L., Wise, M. A., Clarke, L. E. and Edmonds, J. A.: Global Market and Economic Welfare Implications of Changes in Agricultural Yields due to Climate Change, Clim. Chang. Econ., 11(01), 2050005, doi:10.1142/S2010007820500050, 2020.

Calvin, K. V., Patel, P. L., Clarke, L. E., Asrar, G. R., Bond-Lamberty, B., Cui, Y., Di Vittorio, A., Dorheim, K. R., Edmonds, J. A. and Al., E.: GCAM v5.1: Representing the linkages between energy, water, land, climate, and economic systems, Geosci. Model Dev., 12, 1–22, 2019b.

Chaturvedi, V., Hejazi, M., Edmonds, J., Clarke, L., Kyle, P., Davies, E. and Wise, M.: Climate mitigation policy implications for global irrigation water demand, Mitig. Adapt. Strateg. Glob. Chang., 20(3), 389–407, doi:10.1007/s11027-013-9497-4, 2013.

Engström, K., Rounsevell, M. D. A., Murray-Rust, D., Hardacre, C., Alexander, P., Cui, X., Palmer, P. I. and Arneth, A.: Applying Occam's razor to global agricultural land use change, Environ. Model. Softw., 75, 212–229, doi:10.1016/j.envsoft.2015.10.015, 2016.

FAO: Producer Prices, FAOSTAT [online] Available from: http://www.fao.org/faostat/en/#data/PP, 2018a.

FAO: Producer Prices (old series), FAOSTAT, 2018b.

FAO: Land Cover, FAOSTAT [online] Available from: http://www.fao.org/faostat/en/#data/LC (Accessed 4 February 2020a), 2020.

FAO: Production: Crops, FAOSTAT [online] Available from: http://www.fao.org/faostat/en/#data/QC, 2020b.

Féménia, F. and Gohin, A.: Dynamic modelling of agricultural policies: The role of expectation schemes, Econ. Model., 28(4), 1950–1958, doi:10.1016/j.econmod.2011.03.028, 2011.

Fuglie, K. O.: Total factor productivity in the global agricultural economy: Evidence from FAO data, in The shifting patterns of agricultural production and productivity worldwide, pp. 63–95., 2010.

Goldewijk, K. K., Beusen, A., Doelman, J. and Stehfest, E.: Anthropogenic land use estimates for the Holocene - HYDE 3.2, Earth Syst. Sci. Data, 9(2), 927–953, doi:10.5194/essd-9-927-2017, 2017.

Hasegawa, T., Sands, R. D., Brunelle, T., Cui, Y., Frank, S., Fujimori, S. and Popp, A.: Food security under high bioenergy demand, Clim. Change, 2020.

Hejazi, M., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., Wise, M., Patel, P., Eom, J., Calvin, K., Moss, R. and Kim, S.: Long-term global water projections using six socioeconomic scenarios in an integrated assessment modeling framework, Technol. Forecast. Soc. Change, 81(1), 205–226, doi:10.1016/j.techfore.2013.05.006, 2014a.

Hejazi, M. I., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., Wise, M., Patel, P., Eom, J. and Calvin, K.: Integrated assessment of global water scarcity over the 21st century under multiple climate change mitigation policies, Hydrol. Earth Syst. Sci., 18(8), 2859–2883, doi:10.5194/hess-18-2859-2014, 2014b.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B., Calvin, K., Doelman, J., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Kennedy, J., Kristzin, T., Lawrence, D.,

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F., van Vuuren, D. and Zhang, X.: Harmonization of Global Land-Use Change and Management for the Period 850–2100 (LUH2) for CMIP6, Geosci. Model Dev. Discuss., (April), 1–65, doi:10.5194/gmd-2019-360, 2020.

Jones, A. D., Collins, W. D., Edmonds, J., Torn, M. S., Janetos, A., Calvin, K. V., Thomson, A., Chini, L. P., Mao, J., Shi, X., Thornton, P., Hurtt, G. C. and Wise, M.: Greenhouse gas policy influences climate via direct effects of land-use change, J. Clim., 26(11), 3657–3670, doi:10.1175/JCLI-D-12-00377.1, 2013.

Kelly, D. L., Kolstad, C. D. and Mitchell, G. T.: Adjustment costs from environmental change, J. Environ. Econ. Manage., 50(3), 468–495, doi:10.1016/j.jeem.2005.02.003, 2005.

Knoben, W. J. M., Freer, J. E. and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23(10), 4323–4331, doi:10.5194/hess-23-4323-2019, 2019.

Von Lampe, M., Willenbockel, D., Ahammad, H., Blanc, E., Cai, Y., Calvin, K., Fujimori, S., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Lotze-Campen, H., Mason d'Croz, D., Nelson, G. C., Sands, R. D., Schmitz, C., Tabeau, A., Valin, H., van der Mensbrugghe, D. and van Meijl, H.: Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison, Agric. Econ. (United Kingdom), 45(1), doi:10.1111/agec.12086, 2014.

Lubowski, R. N., Plantinga, A. J. and Stavins, R. N.: What Drives Land-Use Change in the United States? A National Analysis of Landowner Decisions, Land Econ., 84(4), 529–550, 2008.

Lundberg, L., Jonson, E., Lindgren, K. and Bryngelsson, D.: A cobweb model of land-use competition between food and bioenergy crops, J. Econ. Dyn. Control, 53, 1–14, doi:10.1016/j.jedc.2015.01.003, 2015.

Manoli, G., Meijide, A., Huth, N., Knohl, A., Kosugi, Y., Burlando, P., Ghazoul, J. and Fatichi, S.: Ecohydrological changes after tropical forest conversion to oil palm, Environ. Res. Lett., 13(6), 64035, doi:10.1088/1748-9326/aac54e, 2018.

McFadden, D.: Econometric models of probabilistic choice, in Structural Analysis of Discrete Data with Econometric Applications, edited by C. Manski and D McFadden, MIT Press, Cambridge MA., 1981.

Mitra, S. and Boussard, J. M.: A simple model of endogenous agricultural commodity price fluctuations with storage, Agric. Econ., 43(1), 1–15, doi:10.1111/j.1574-0862.2011.00561.x, 2012.

Mouratiadou, I., Biewald, A., Pehl, M., Bonsch, M., Baumstark, L., Klein, D., Popp, A., Luderer, G. and Kriegler, E.: The impact of climate change mitigation on water demand for energy and food: An integrated analysis based on the Shared Socioeconomic Pathways, Environ. Sci. Policy, 64, 48–58, doi:10.1016/j.envsci.2016.06.007, 2016.

National Research Council: Advancing Land Change Modeling: Opportunities and Research Requirements, National Academies Press, Washington DC, USA. [online] Available from: http://nap.edu/18385, 2014.

Nelson, G. C., Valin, H., Sands, R. D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., Kyle, P., Von Lampe, M., Lotze-Campen, H., Mason d'Croz, D., van Meijl, H., van der Mensbrugghe, D., Müller, C., Popp, A., Robertson, R., Robinson, S., Schmid, E., Schmitz, C., Tabeau, A. and Willenbockel, D.: Climate change effects on agriculture: Economic responses to biophysical shocks, Proc. Natl. Acad. Sci., 111(9), 3274–3279, doi:10.1073/pnas.1222465110, 2014.

https://doi.org/10.5194/gmd-2020-338
Preprint. Discussion started: 1 December 2020

Geoscientific Model Development
Discussions
Open Access

Nerlove, M.: Adaptive Expectations and Cobweb Phenomena, Q. J. Econ., 72(2), 227–240, doi:10.2307/1880597, 1958.

Popp, A., Rose, S. K., Calvin, K., Van Vuuren, D. P., Dietrich, J. P., Wise, M., Stehfest, E., Humpenöder, F., Kyle, P., Van Vliet, J., Bauer, N., Lotze-Campen, H., Klein, D. and Kriegler, E.: Land-use transition for bioenergy and climate stabilization: Model comparison of drivers, impacts and interactions with other land use based mitigation options, Clim. Change, 123(3–4), 495–509, doi:10.1007/s10584-013-0926-x, 2014.

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., Bodirsky, B. L. B. L., Dietrich, J. P. J. P., Doelmann, J. C. J. C., Gusti, M., Hasegawa, T., Kyle, P., Obersteiner, M., Tabeau, A., Takahashi, K., Valin, H., Waldhoff, S., Weindl, I., Wise, M., Kriegler, E., Lotze-Campen, H., Fricko, O., Riahi, K. and Vuuren, D. P. va. D. P. V.: Land-use futures in the shared socio-economic pathways, Glob. Environ. Chang., 42, 331–345, doi:10.1016/j.gloenvcha.2016.10.002, 2017.

Roberts, M. J. and Schlenker, W.: Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate, Am. Econ. Rev., doi:10.1257/aer.103.6.2265, 2013.

Sands, R. and M. L.: Modeling agriculture and land use in an integrated assessment framework, Clim. Change, 56, 185–210, 2003.

Schmitz, C., van Meijl, H., Kyle, P., Nelson, G. C., Fujimori, S., Gurgel, A., Havlik, P., Heyhoe, E., D'Croz, D. M., Popp, A., Sands, R., Tabeau, A., van der Mensbrugghe, D., von Lampe, M., Wise, M., Blanc, E., Hasegawa, T., Kavallari, A. and Valin, H.: Land-use change trajectories up to 2050: Insights from a global agro-economic model comparison, Agric. Econ. (United Kingdom), 45(1), 69–84, doi:10.1111/agec.12090, 2014.

Snyder, A. C., Link, R. P. and Calvin, K. V.: Evaluation of integrated assessment model hindcast experiments: A case study of the GCAM 3.0 land use module, Geosci. Model Dev., 10(12), doi:10.5194/gmd-10-4307-2017, 2017.

Stehfest, E., Van Zeist, W.-J., Valin, H., Havlik, P., Popp, A., Kyle, P., Tabeau, A., Mason-D'Croz, D., Hasegawa, T. and Bodirsky, B. L.: Key determinants of global land-use projections, Nat. Commun., 10(1), 2166, 2019.

Taheripour, F. and Tyner, W.: Biofuels and Land Use Change: Applying Recent Evidence to Model Estimates, Appl. Sci., 3(1), 14–38, doi:10.3390/app3010014, 2013.

Tilman, D., Balzer, C., Hill, J. and Befort, B. L.: Global food demand and the sustainable intensi fi cation of agriculture, , 108(50), doi:10.1073/pnas.1116437108, 2011.

USDA: Commodity costs and returns, [online] Available from: https://www.ers.usda.gov/data-products/commodity-costs-and-returns.aspx, 2020a.

USDA: Farm Business Income, Farm Sect. Income Financ. [online] Available from: Farm Business Income Statement (Accessed 3 June 2020b), 2020.

USDA: Federal Government direct farm program payments, Farm Income Wealth Stat. [online] Available from: https://data.ers.usda.gov/reports.aspx?ID=17833 (Accessed 3 June 2020c), 2020.

Weber, J. G. and Key, N.: How much do decoupled payments affect production? An instrumental variable approach with panel data, Am. J. Agric. Econ., 94(1), 52–66, 2012.

Wise, M., Calvin, K., Kyle, P., Luckow, P. and Edmonds, J.: Economic and Physical Modeling of Land Use in Gcam 3.0 and an Application To Agricultural Productivity, Land, and Terrestrial Carbon, Clim. Chang. Econ., 05(02), 1450003, doi:10.1142/S2010007814500031, 2014.

Young, C. E. and Westcott, P. C.: How Decoupled Is U.S. Agricultural Support for Major Crops ?, Agric. Appl. Econ. Assoc., 82(3), 762–767, 2000.

Zhao, X., van der Mensbrugghe, D. Y., Keeney, R. M. and Tyner, W. E.: Improving the way land use change is handled in economic models, Econ. Model., 84, 13–26, doi:https://doi.org/10.1016/j.econmod.2019.03.003, 2020a.

Zhao, X. I. N., Calvin, K. V. and Wise, M. A.: The critical role of conversion cost and comparative advantage in modeling agricultural land use change, Clim. Chang. Econ., 11(1), doi:10.1142/S2010007820500049, 2020b.

Author Contributions

KC and MW designed the experiment. KC and AS developed the model code. KC performed the simulations. KC, AS, and XZ analyzed results. KC prepared the manuscript with contributions from all authors.

Acknowledgments

This research was supported by the US Department of Energy, Office of Science, as part of research in Multi-Sector Dynamics, Earth and Environmental System Modeling Program. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

Data Availability

gcamland code and inputs are available at https://github.com/JGCRI/gcamland and https://zenodo.org/record/4071797. All outputs are available at [PNNL DataHub link to be inserted upon acceptance]. The code used to generate the figures in this paper is available at [github link to be inserted upon acceptance].