Chapter 9
Model Validation: Comparing Gridded and Regional Simulations to Observations

Iman Haqiqi, Zhan Wang, and Uris Lantz C. Baldos

Model validation is important to ensure that a model is accurate and reliable. The general goal of validation is to compare a model's predictions or projections to actual data to check whether the model can simulate the conditions that are observed independently of the model and its input data. Since it is generally not possible to validate all aspects of a high-dimensional, multi-scale model, such validation exercises generally have a focus area that depends on the purpose for which it will be used. This varies by application.

Each discipline has a different approach to validation depending on the research purpose and complexity of the processes being modeled (Hansen and Heckman 1996; Rykiel 1996; Oreskes 1998; Biondi et al. 2012; Ngo and See 2012; Kersebaum et al. 2015; van Vliet et al. 2016). For example, climate models seek to explain underlying physical processes and make predictions and projections about a changing climate. In agronomic models, validation often involves estimating out-of-sample yields and biophysical processes. In hydrological models, validation may focus on replicating observed streamflow and river discharge at another location or at another point in time. Economic models aim to explain economic decisions and provide policy insights and evaluate the impacts of future changes. They typically exhibit greater uncertainty due to the confounding effects of human behavior, which does not follow the laws of physics, biology, or chemistry. These economic models generally include many parameters and lean heavily on theoretical structures obtained from microeconomic theoretical foundations.

I. Haqiqi (✉) · Z. Wang · U. L. C. Baldos
Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA
e-mail: ihaqiqi@purdue.edu

I. Haqiqi, T. W. Hertel (eds.), *SIMPLE-G*,
https://doi.org/10.1007/978-3-031-68054-0_9

1 Validation Challenge When Multiple Drivers Interact Across Multiple Scales

The SIMPLE-G model involves economic decisions about land use and water withdrawals at the grid-cell level. Here, the focus is on geospatial validation of economic decisions about cropland and water withdrawals. However, land-use changes are the result of many mutually influential local, regional, and global drivers that together shape the global cropland patterns. Thus, the model must be able to simulate complex processes in order to represent the richness in observed changes in land-use patterns. Unfortunately, few of these models are validated: Advances in model validation techniques have been much slower than new model developments (van Vliet et al. 2016). Accordingly, there is no agreed-upon set of methods for assessing the results of integrated agricultural, economic, water, and land-use models (Razavi and Gupta 2015; Razavi et al. 2021). Here, we draw on validation methods from economics, sociohydrology, and land-use modeling.

There are several possible ways to validate a geospatial economic model like SIMPLE-G. However, the initial step is benchmark replication or calibration. Comparing the results of a model to a set of observed data will ensure that the model can replicate a base reference condition. To perform benchmark replication, the model is first calibrated to the benchmark data (e.g., 2010). The model's predictions should match the benchmark data. Backcasting. another validation method, employs the model to make predictions about observed past events and then compares the model's predictions to actual observed outcomes from independent data sources not used in the calibration process. If the predictions are accurate according to predefined criteria, then it is likely that the model will be accurate in the future or for unobserved conditions. This ensures that the structural processes that can explain changes in the system are modeled correctly. Please note that we use the terms 'prediction' and 'projection' interchangeably.

A complementary approach to economic model validation is to use a sensitivity analysis, which involves varying the model's assumptions, parameters, and drivers to investigate how the predictions change. This helps identify the assumptions and parameters of the model that are most important in determining outcomes. Sensitivity analyses are also important for ensuring the robustness of the model findings (Haqiqi et al. 2023).

While economists try to explain the observations, there are inevitably prediction errors and unobservable variables. This challenge is even more daunting for a geospatial economic model with numerous unobservable local variables. Uncertainty quantification and characterization can help uncover sources of uncertainty in the results and reveal their implications for the major conclusions of each study.

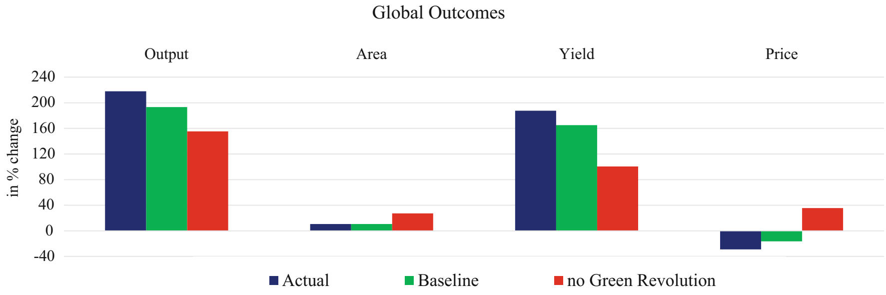

Fig. 9.1 SIMPLE model global validation. SIMPLE is capable of partitioning the actual tripling of crop output (*first blue bar*) into area and yield changes (*second and third blue bars*) while also reproducing the historical change in crop price over the period 1961–2006 (Hertel et al. 2014) The *green bars* are model predictions, and the *red bars* represent a counterfactual experiment exploring the impact of eliminating gains from the Green Revolution, which was a focal point of this chapter (Hertel et al. 2014, p. 13800)

2 Validation at the Global Level

SIMPLE-G is nested within the broader SIMPLE model, which is a global model of crop production, land use, and the environment. In general, the process of model validation requires four steps. First step involves selecting a period wherein the model outcomes will be compared to actual data. In this case, previous work has validated the SIMPLE model against observations from 1961—the start of the FAOSTAT data series for global agriculture—to 2006, just prior to the global food and energy crisis that began in 2007. The second step in general model validation is to project the model backward in time to generate a historical database using actual changes in key model drivers (e.g., 2006–1961). At this point it is possible to compare the model's performance to observed changes. However, forward-looking simulations are often easier to apply; therefore, it is common to follow with the third and the fourth steps, projecting the model forward to the current period using the historical database and then comparing projected with observed changes over the historical period. The key model drivers include population, per capita income growth, and productivity growth for the crop, livestock, and processed food sectors. Figures 9.1 and 9.2 are taken from Hertel et al. (2014) and Hertel and Baldos (2016) to compare the performance of the SIMPLE model in predicting changes in global production, prices, and yields as well as regional crop production over this 45-year period.

At the global level, the SIMPLE model is capable of capturing the direction of historical changes in crop production, cropland use, and crop prices. However, regional historical changes in crop output and cropland use are more difficult to replicate due to region-specific drivers, which are omitted from this historical simulation. Explicit government policies could shape regional patterns of agricultural production, as was in the case for Brazil, where agriculture was initially heavily

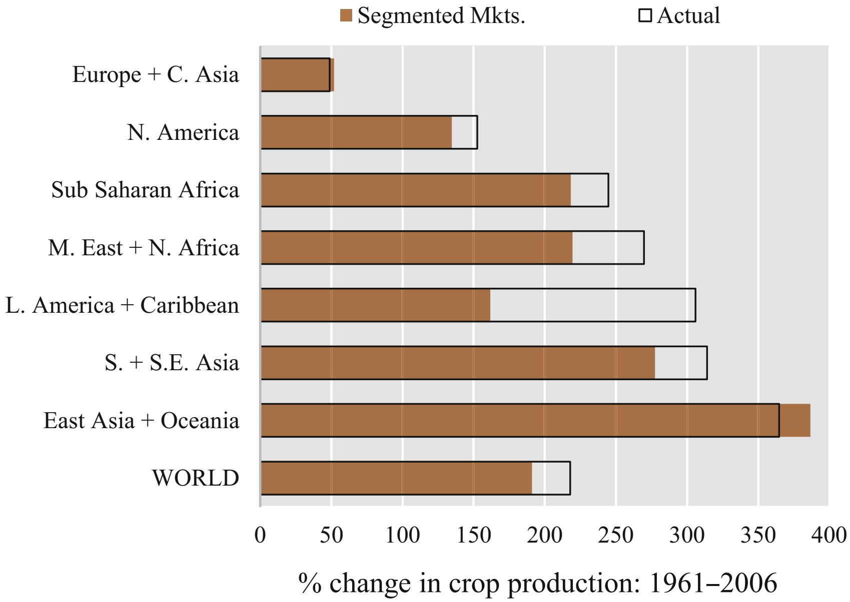

Fig. 9.2 SIMPLE model regional validation. The model is largely capable of reproducing the regional pattern of output growth over the period 1961–2006 (Hertel and Baldos 2016, p. 198) Market segmentation follows Armington (1969) such that there is imperfect substitution between foreign and domestic goods in each region. The most problematic region is Latin America and the Caribbean, where the model does not capture Brazil's dramatic output growth

subsidized through rural credit and price support mechanisms. However, market reforms introduced during the early 1990s reduced trade barriers in commodity markets (Chaddad and Jank 2006). There are also barriers to international trade in agricultural products, including poor quality domestic transport infrastructure, burdensome customs procedures, and poorly developed port facilities. These barriers to trade loom particularly large in Sub-Saharan Africa (Wilson et al. 2004) and have limited that region's engagement in the global trading system.

3 Validation at the Regional Level: The Case of Brazil

By focusing the validation exercise on a single region, it is possible to dig more deeply into region-specific data sources and policies. This section reports on the work of validating a region-specific model focused on gridded agriculture, land use, and the environment against region-level observations in Brazil (SIMPLE-G-Brazil, see Chap. 16 for additional details). After developing the initial version of SIMPLE-G-Brazil, we begin by hindcasting the model from its baseline (2017) to the year 2000 to compare the simulated crop output (converted to crop-equivalent and

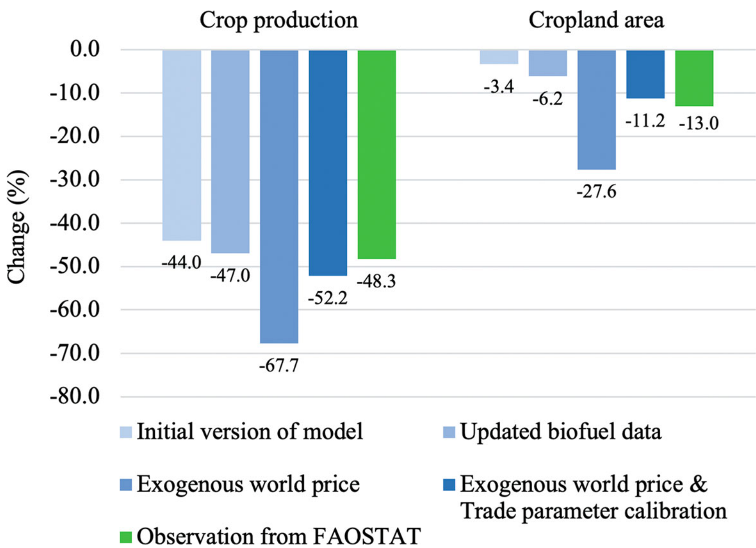

Fig. 9.3 Validation of crop output and cropland for Brazil
Model simulations (*blue*) are compared with observations reported by the United Nations Food and Agriculture Organization (*green*)

aggregated) and cropland area with observed data from FAOSTAT for Brazil. (Note that this validation exercise is backward-looking, in contrast to the forward-looking validation reported in Figs. 9.1 and 9.2.) Socioeconomic drivers used in the initial step of hindcasting include population and per capita GDP, calculated from population and GDP (in constant dollars) data from the World Bank (3-year averages are used to remove short-term volatility); total factor productivity (TFP), calculated from TFP growth rates (Fuglie 2022) for the crop, livestock (Ludena et al. 2007), and processed food (Griffith et al. 2004) sectors. The percentage change in demand for crops by the biofuel sector is calculated with the ratio of global biofuel demand in the transportation sector (representing biodiesel and ethanol from crops) between 2017 and 2000 from the World Energy Outlook 2018 (IEA 2018).

As is shown in the first bar of each grouping in Fig. 9.3, the initial version of the model underestimates the reduction in crop output (−44.0% in simulation versus −48.3% observed) and cropland area (−3.4% in simulation versus −13.0% observed) in Brazil. To identify potential issues that affect the performance of SIMPLE-G-Brazil, we apply a stepwise strategy to check dimensions of the model that may cause mismatches between simulations and observations.

According to the structure of SIMPLE-G (Fig. 4.1), demand for crop output can be divided into three categories: exogenous demand for biofuel production, endogenous demand for domestic consumption, and endogenous demand for foreign countries. First, we checked data sources on biofuel demand in the initial version and found that these data are based on the crop sales share for biofuel from GTAP-Bio database Version 6, with 2006 as the baseline. These data did not capture the

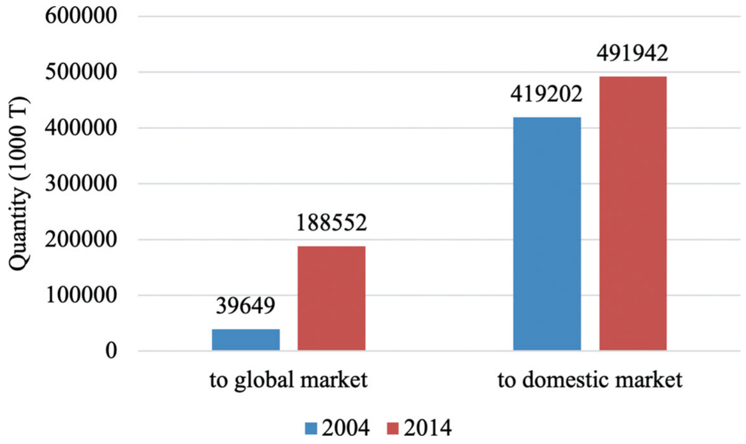

Fig. 9.4 Crop supply by market in Brazil, calculated using market share from GTAP database

rapid growth of biofuel production in Brazil during the 2000–2017 hindcasting period. To fix this problem, we update the biofuel demand data with the sales share from GTAP-Bio database Version 9 (with 2011 serving as the baseline). As a result, the validation outcomes (i.e., the "updated biofuel data" category in Fig. 9.3) show that the simulated change in crop output from 2017 to 2000 (−47.0%) becomes closer to the observed change (−48.3%), but the model still underestimates the (backward-looking) reduction of cropland area (−6.2% in simulation versus −13.0% in observation). A smaller backward-looking reduction in area means that the model is not picking up all of the growth in cropland area from 2000 to 2017.

Second, we compared the simulated crop supply pattern (sales to domestic versus international markets) with the crop supply patterns calculated using the domestic and international market shares of crop supply from the GTAP version 10 database, with the baseline years of 2004 and 2014, respectively (Fig. 9.4).[1] We found that, compared with 2014, the GTAP database indicated that the crop supply in 2004 should have had a much greater reduction for the global market (−79%) than in the domestic market (−15%), indicating strong export growth over the 2004–2014 period. However, under the "updated biofuel data" category, the simulated reduction in crop supply (variable QSCROPr) was −60% for the global market and − 42% for the domestic market, which appears to greatly underestimate the role of export growth over this historical period. Given that the global crop price is simulated endogenously with socioeconomic drivers in this step, these findings indicate that there are some uncontrolled impacts on the global crop price, possibly from other regions, that cause mismatches in crop supply patterns.

[1] Although our simulation is between 2017 and 2000, GTAP version 10 database does not provide data on these two years, so we used the year closest to these two years to show the general patten of crop supply change.

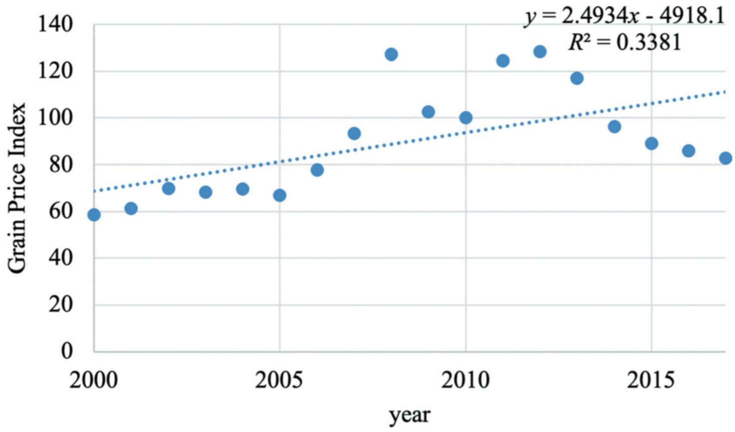

Fig. 9.5 Grain price index and its linear regression model

To control for potential bias originating from our modeling of the global crop market, we added another socioeconomic driver to the hindcast—the observed global crop price change over this historical period—to isolate Brazil from uncontrolled impacts emanating from non-Brazil regions via the global crop market. This allows us to focus more sharply on the response of crop production and land use within Brazil. The shock to the change in global crop prices was calculated with the annual indices for grains based on World Bank commodity price data (Fig. 9.5). We used the grain price index from 2000–2017 to fit a linear model between the grain price index and time and use the fitted value from the model between 2017 and 2000 to calculate the percentage shock for global grain price in order to smooth short-term variation in the data.

The simulation results with both socioeconomic drivers and additional global price shocks are reported under the "exogenous global price" category in Fig. 9.3. Although this simulation overestimates the reduction in both cropland and crop output, the simulated change in global crop supply is close to observations. The results suggest a 91% reduction of supply to the global market and a 59% decline in supply to the domestic market, which more closely mimics the observed crop supply pattern reported in the GTAP database. This result indicates that the global crop price is a partial source of the mismatch and that Brazil is overresponsive to global crop price changes in the current model.

To address the problem of overresponsiveness to global price changes, we further calibrated two key parameters in the Armington trade structure: the elasticity of substitution between crops demanded from domestic and international markets and the elasticity of transformation between crops supplied to domestic and international markets. The calibration was conducted by gradually changing these two parameters by the same amount and finding the parameter value (0.7) that produces the simulated results (both cropland and crop output) that most closely match the observations as reported in FAOSTAT. After calibration, the results are shown in

Fig. 9.3 as the group "exogenous global price & trade parameter calibration"; these results are not only more accurate than the initial validation results, but also exhibit more similar changes in the crop supply pattern (−73% to global market and −44% to domestic market) with the data from the GTAP database. This exercise highlights several useful strategies in model validation. It is important to break down the problem based on the theoretical foundation of this model, which can lead to fruitful scrutiny of the data and parameters used in the model. Proceeding in a stepwise fashion is important, as shown in Fig. 9.3.

4 Validation of Gridded Model for the United States

As above, we use backcasting as the main validation strategy for SIMPLE-G-US. This approach also provides insights into the strengths and weaknesses of a model in a specific application. In the case of the gridded model, the main challenge is obtaining accurate geospatial data as a reference for comparison and accurate geospatial data for drivers of the model. Here, we ask whether the model can replicate a movement to a new state of the agricultural economic systems from the period 2001–2002 to 2016–2017, an analysis that requires controlling many social, political, and economic variables.

4.1 Precision Assessment, 2001–2002 to 2016–2017

Here, we study the performance of the SIMPLE-G-US model in predicting cropland area and its changes in the continental United States over the long run using independently collected reference data. The validation results are reported first for the entire continental United States followed by tables for the USDA Farm Resource Regions. Given data availability, we draw on cropland reference data for 2001 and 2016 from the USGS National Land Cover Database (NLCD) and economic drivers for 2002 and 2017 from the USDA Census of Agriculture. Figure 9.6 illustrates the validation steps of SIMPLE-G-US.

Reference for Comparison Advances in satellite imagery over the past 2 decades have allowed researchers to compile new datasets with improved detail and geospatial accuracy. We draw on the National Land Cover Database (NLCD) for 2001–2016, a valuable resource for US land cover, as the reference dataset for this validation exercise. The NLCD is a collection of datasets created by the United States Geological Survey (USGS) that provides information on US land cover at a resolution of 30 meter. The maps are created by classifying USGS/NASA Landsat satellite imagery into different land cover categories. Here, we take cropland area from class #82 "Cultivated Crops," which includes annual crops and perennial woody crops. We aggregate these data to the SIMPLE-G model spatial resolution (5 arcmin) and then to the US county level, the level at which key USDA census data

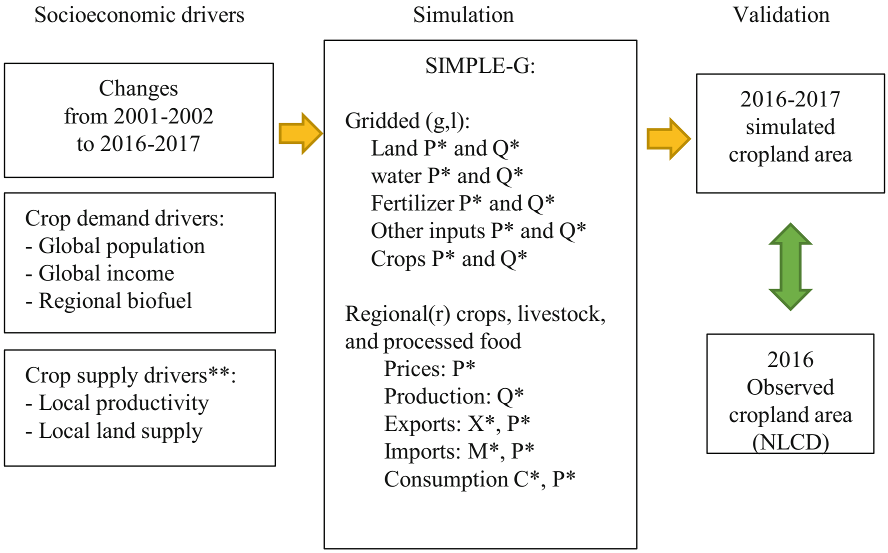

Fig. 9.6 Steps in the validation of the SIMPLE-G-US-Allcrops model, 2001–2002 to 2016–2017 When deployed in SIMPLE-G-US, the socioeconomic drivers of cropland change (*lefthand column*) result in gridded and regional outcomes (*center column*). For model validation, we compare simulated cropland area in 2016–2017 to observed area in 2016 (*righthand column*)

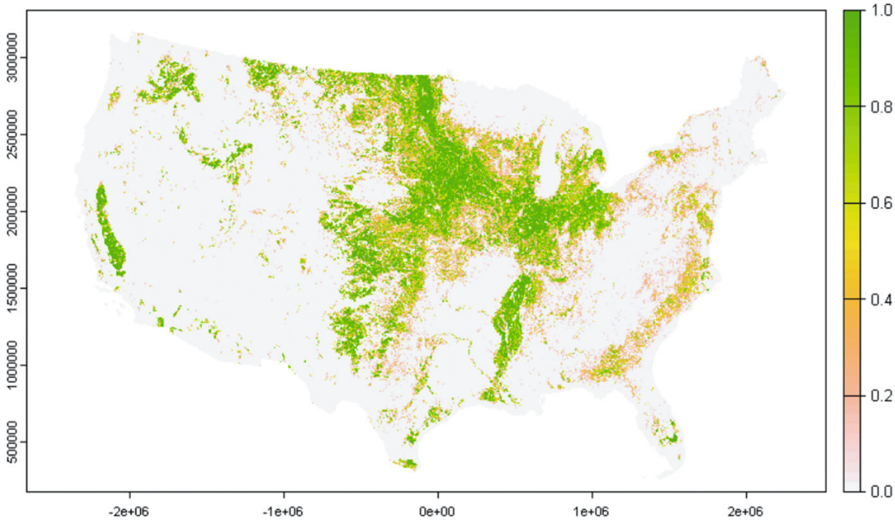

Fig. 9.7 Share of cropland in total grid cell area calculated from the National Land Cover Database, 2016

are reported. Fig. 9.7 illustrates the cropland share in total grid cell area calculated based on NLCD for the year 2016.

Global and Local Drivers of Land-Use Change from 2001–2002 to 2016–2017 - Historical validation is complicated when multiple biophysical and socioeconomic drivers interact at multiple scales. Hydroclimatic, agricultural, and economic systems are interconnected, and it would be impossible to factor in all the drivers of change in these systems. However, we find that including the most significant drivers in a well-structured model can give us a good representation of reality. The first step is the identification of these critical drivers. During the study period, four major changes are observed. First, population and income increased. We track these changes at the regional level. These increases lead to more demand for food and more demand for agricultural inputs, along with demand for land and water usage. Second, the United States mandated a sharp increase in ethanol production over this period, causing an increase in the production of corn. Third, enrollment in Conservation Reserve Programs declined significantly, such that more land became available for reversion to cropland. In addition, there has been a change in the rate of fallow/idle cropland, with significant implications for cropland supply curves. Finally, growth in agricultural productivity continued to be a dominant factor contributing to an increase in production in the United States (Clancy et al. 2016; Wang et al. 2020; Fuglie et al. 2022). We consider these drivers in the historical validation exercise. The drivers of the model are obtained from the USDA Census of Agriculture and FAO annual statistics, described below. The supply-side economic drivers are changes in county-level Hicks-neutral agricultural productivity for irrigated and rainfed crop production and county-level changes in the economic supply of cropland. The demand-side economic drivers are county-level changes in crop demand for biofuels and changes in global population and income.

Demand Drivers: Regional Population and Income The population and per capita income in 2001–2002 and 2016–2017 are obtained from the World Bank for each country and then aggregated to 16 SIMPLE regions (see Table 9.1). In this period, global population increased by 19.8% while per capita income increased by around 57.7%.

Demand Drivers: Crop Demand for Biofuels The biofuel boom has had important implications for corn-producing regions. However, the current SIMPLE-G model includes an aggregated measure of all crops produced. To accurately model the biofuel boom, we incorporate the increase in crop demand for biofuel for corn-producing subregions according to their share of corn in total crop production. For each county, the increase in demand is calculated using the following equation:

$$b_j = \theta_{corn,j} b_r, \tag{9.1}$$

where b_j is the percentage change in the demand for all crops produced in county j, $\theta_{corn,\,j}$ is the share of corn in county j's crop output, and b_r is the regional percentage

Table 9.1 Global drivers of land-use change, 2002–2017

	Population			Income		
	2002	2017	Growth (%)	2002	2017	Growth (%)
Eastern Europe	284,796	282,095	−0.9	4,678	7,842	66.0
North Africa	148,197	191,427	29.2	3,103	3,647	51.8
Sub Saharan Africa	648,302	985,597	52.0	929	1,378	125.5
South America	179,870	213,451	18.7	6,379	8,942	66.3
Australia, New Zealand	23,365	29,287	25.3	42,100	51,418	53.1
Europe	472,201	498,457	5.6	31,031	36,235	23.3
South Asia	1,419,749	1,755,775	23.7	832	1,778	164.2
Central America	354,091	420,080	18.6	6,958	8,400	43.2
South Africa	53,958	66,098	22.5	4,462	5,593	53.5
Southeast Asia	542,834	652,564	20.2	2,042	3,689	117.2
Canada	31,178	36,732	17.8	39,117	44,109	32.8
United States	287,279	325,085	13.2	49,184	58,330	34.2
China	1,336,765	1,452,625	8.7	2,600	8,901	272.0
Middle East	243,072	329,399	35.5	7,390	10,566	93.7
Japan, Korea	199,232	204,029	2.4	24,384	29,772	25.0
Central Asia	66,504	92,197	38.6	933	1,988	195.4
World	6,293,395	7,536,915	19.8	7,899	10,405	57.7

change in corn demand due to biofuel demand observed in the 2002–2017 period reported by the USDA (Chen et al. 2011).

Supply Drivers: Economic Supply of Cropland The cropland supply curve may shift with changes in the regulatory environment (e.g., conservation policy) or the relative return to alternative land use (e.g., pastureland rents decline). Here, the supply shifters (called "slack variables" in our model) are calculated based on changes in the average share of harvested cropland to total cropland (i.e., harvest rate) for each location:

$$\begin{aligned} H_g &= \Psi_g L_g \\ h_g &= \psi_g + l_g, \end{aligned} \tag{9.2}$$

where H is the harvested area, Ψ is the harvest rate, and L is the cropland area defined for each grid cell g. Percentage changes are shown with lowercase letters. These data are taken from the USDA Census of Agriculture at the county level. Total cropland area includes harvested cropland, pastured cropland, idle cropland, and fallowed land.

Supply Drivers: Hicks-Neutral Agricultural Productivity TFP growth has been the major driver of US crop production in recent decades (Wang et al. 2020). Here, the productivity variables are represented by capital letter A (lowercase letter a for the percentage change). Hicks-neutral agricultural productivity is an economic term reflecting an increase in crop production using the same levels of land and nonland inputs. If we show the production volume by Q and the composite inputs by X, the percentage change in TFP can be shown as follows:

$$a = q - x. \tag{9.3}$$

Assuming no change in the production technology (input mixes), TFP growth can be approximated by changes in average yields. We estimate the changes in average yields using the following relationship:

$$\begin{aligned} V &= L.Y.P \\ v &= l + y + p \\ y &= v - l - p, \end{aligned} \tag{9.4}$$

where V is the value of crop sales; L is the cropland harvested area; Y is the average corn-equivalent yield; and P is the average crop price. For the validation exercise, P is calculated based on the average crop price index from the FAO for Farm Resource Regions. We obtain V and L from the USDA Census of Agriculture for 2002 and 2017 at the county level. Then, y is calculated for all counties using Eq. 9.4. Uniform values for productivity growth are assigned to all grid cells within a county.

Supply Drivers: Economic Supply of Water and Crop Water Demand Crop production is sensitive to water availability. Thus, changes in water conditions can affect the extent of cropland. In SIMPLE-G, the irrigation water supply curve may shift with changes in weather conditions (e.g., lower surface water availability) and changes in irrigation costs and expenditures (e.g., groundwater pumping costs). Ideally, we would calculate the shift in surface water supply for each grid cell. While the USGS reports the water conditions for each 5-year interval, it is not possible to decompose the demand drivers and supply drivers of changes in water withdrawal from those reports. As no other observational dataset is available, data collection can rely on outputs of a hydrologic model (e.g., the Water Balance Model). An additional complication is that measuring changes in groundwater supply requires converting changes in the groundwater table to equivalent changes in the supply shifter at each 5 arcmin grid cell in SIMPLE-G. However, as running a hydrology model requires multidisciplinary collaborations, we validate the model while excluding the supply shifters related to water. This may lead to less accurate cropland projections in highly irrigated areas where significant changes in water conditions have been observed (e.g., 2012 and 2015 drought and heat stress).

4.2 Agreement Between Observation (Satellite) and Simulation (SIMPLE-G)

At the national level, SIMPLE-G projects 128.68 million ha of cropland in the continental United States for 2016–2017. According to the NLCD, this number was actually 129.30 million ha for 2016. But how is this area distributed? For that, we focus on observations and simulation outcomes at the administrative units given by counties and states within the continental United States as well as the USDA Farm Resource Regions.

Agreement at the County Level Figure 9.8 illustrates projected cropland area (SIMPLE-G) and reference cropland area (NLCD) for 2016–2017 by county. Overall, the projected area is close to the reference data ($R = 0.98$).

Agreement at the State Level Figure 9.9 illustrates the cropland area projected by SIMPLE-G versus NLCD cropland area for major farming states around 2016–2017. Overall, SIMPLE-G can simulate total cropland area with only minor differences. However, in the case of Montana the difference is noticeable. Further investigation is required to explore the possible causes of this discrepancy.

Agreement at Farm Resource Regions Figure 9.10 illustrates the cropland area projected by SIMPLE-G versus NLCD cropland area for USDA Farm Resource Regions around 2016–2017. Visual inspection indicates that SIMPLE-G accurately simulates the total cropland area.

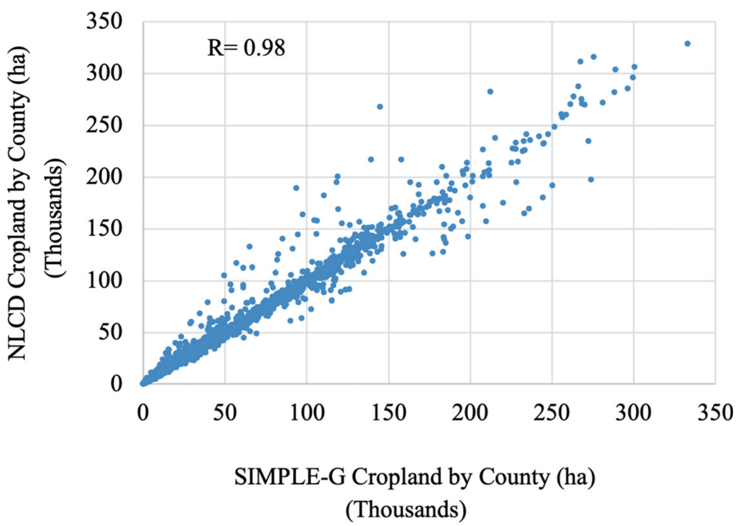

Fig. 9.8 Observed and simulated cropland area in the United States in 2016–2017 based external drivers, 2001–2002 to 2016–2017

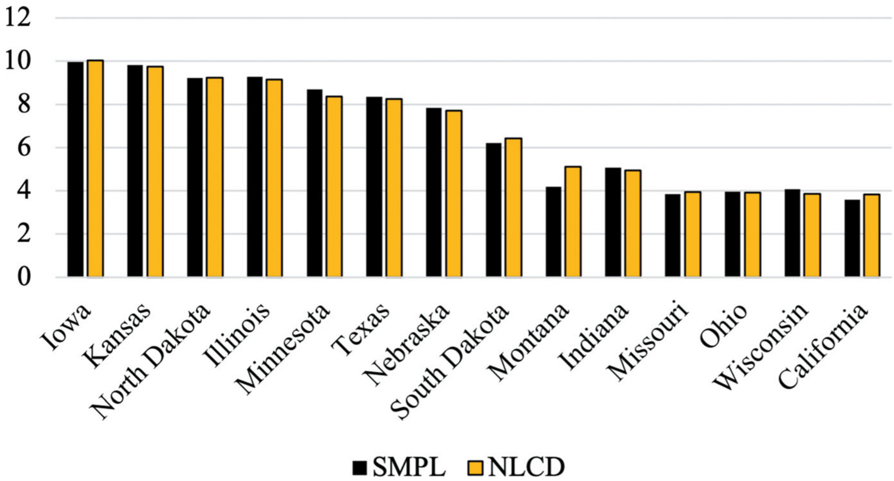

Fig. 9.9 Observed and simulated cropland area by major farming states, 2016–2017

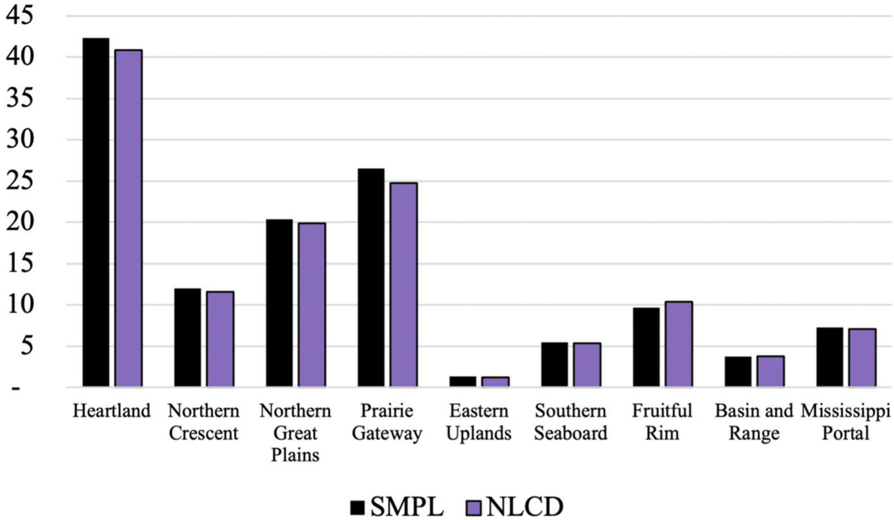

Fig. 9.10 Observed and simulated cropland area by USDA Farm Resource Regions in 2016–2017

To evaluate the significance of different drivers of cropland changes, Fig. 9.11 illustrates the role of supply (i.e., productivity and land supply) and demand drivers (i.e., global population, global income, local biofuels) by USDA Farm Resource Region. The results show that the supply and demand drivers are working in opposite directions in the Fruitful Rim and Basin and Range regions. This means that the final net change is sensitive to the magnitude of these drivers.

Kolmogorov–Smirnov Test The Kolmogorov–Smirnov test is a nonparametric test used to compare two continuous distributions. It is calculated as the maximum

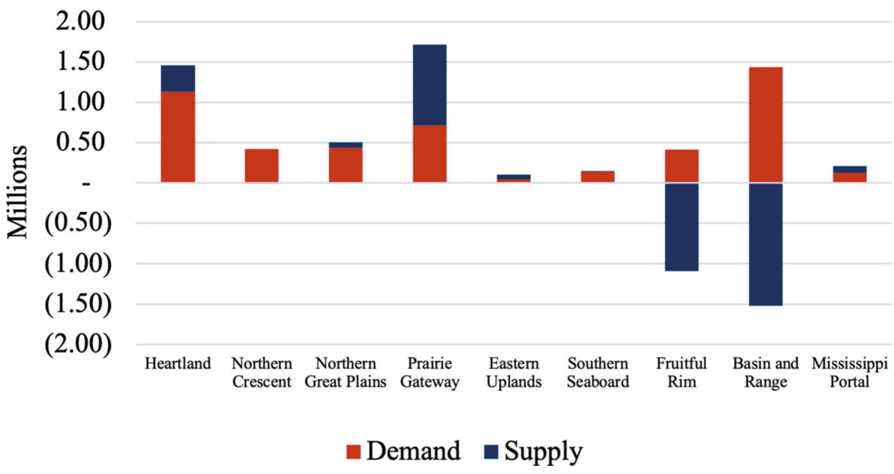

Fig. 9.11 Supply and demand drivers of change in cropland area by USDA Farm Resource Regions, 2016–2017

absolute difference between the cumulative distribution function of the reference distribution and the cumulative distribution function of a sample, estimated parameters, or simulated output:

$$D = \max_x |F_{SMPL}(x) - F_{NLCD}(x)|. \tag{9.5}$$

Figure 9.12 illustrates the cumulative distributions of cropland changes in the NLCD (reference) and SIMPLE-G. Overall, the distributions of changes are in agreement.

4.3 Discussion: Causes of Disagreement

Because the underlying human decisions that drive local changes in land use are rarely deterministic, these results can never be expected to be perfect when compared to empirical data. Additionally, many land-use changes are not the result of one simple process but rather a combination of biophysical and socioeconomic drivers that are mutually influential. Here, we review some of the most important confounding factors.

Government Policies Capturing changes in agricultural-related policies in the model can potentially improve the agreement between SIMPLE-G projections and NLCD data. Overall, government policies can introduce a significant amount of disagreement into coupled economic and environmental models, making it difficult to predict future outcomes. Local and national policies can change farmers'

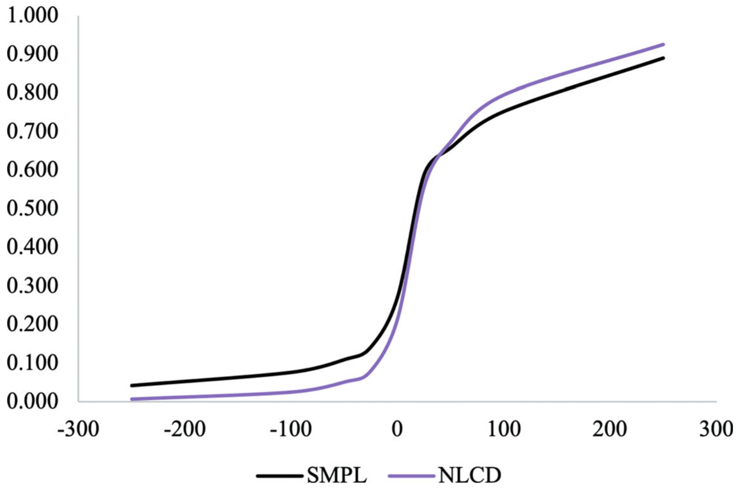

Fig. 9.12 Cumulative distribution of change in cropland (ha) by grid cell, 2001–2002 to 2016–2017

incentives and their decisions on the extent and intensity of farming. For example, policies that promote wetlands and improve water quality may increase the opportunity costs of cropland use and may increase the application of fertilizer on cropland, which will lead to changes in production patterns and application rates. Capturing these changes makes it difficult to accurately predict local conditions. This is important as some of these policies may have a significant local impact on economic and environmental outcomes and thus are necessary to be included in validation. In addition, the decision environment or policies can change over time. Farmers account for uncertainty about the future direction of policies in their land and water use decisions. However, government policies can be complex and difficult to implement. For example, many of the Conservation Reserve Programs (CRP) have multiple components and complicated details that may require different approaches for each location and program. Overall, acreage enrolled in CRPs declined from around 33 million acres in 2002 to around 24 million acres in 2017. Another example of policies with implications for land use is the Sustainable Groundwater Management Act, which requires local agencies to "form groundwater sustainability agencies (GSAs) for the high and medium priority basins." Prior expectations about such a policy might have implications for farmers' decisions about irrigation even before the act is passed.

Local Water Conditions We suspect that a major cause of disagreement between SIMPLE-G projections and NLCD data is related to changes in weather and water conditions between 2002 and 2017. One major determinant of the extent of irrigated cropland is hydroclimatic conditions. Higher temperatures increase irrigation requirements, and lower precipitation may reduce the availability of water resources. While SIMPLE-G does not include a hydrological model, it does contain an economic model for long-run economic demand and supply for water. Any changes to hydroclimatic conditions should be translated to economic demand and supply before entering the model as input. Obtaining gridded information on changes in water availability and water requirements is computationally expensive but can be done in the future with improved availability of these data.

Unobserved Local, Regional, and Global Drivers Many other factors related to agriculture can be sources of disagreement between SIMPLE-G and NLCD. These include, for example, the US–China trade war, which had significant implications for farmers in 2017 (Marchant and Wang 2018) and can be a negative demand driver for soy-producing grid cells. Additionally, the increase in energy prices observed in 2007–2008 led to increases in fertilizer prices and reductions in demand (Beckman and Riche 2015). This increase was a supply-side driver that affected production expenses. The grid cells with high-cost shares of commercial fertilizer will be more affected than other grid cells. Finally, the increasing trend in nonfarm investment in farmland has caused increases in land values and cropland rental rates (Burns et al. 2018). Overall, it is important to note that numerous external factors like trade wars, energy prices, and farmland investment create discrepancies between SIMPLE-G and NLCD's agricultural predictions. While it is impossible to consider all the changes, users need to consider the most important drivers.

Remote Sensing Accuracy While the overall accuracy of the NLCD 2019 data product is relatively high ($90.3\% \pm 0.7\%$), the user accuracy and producer accuracy for forest loss and grass gain were $> 70\%$ and generally $<50\%$ for all other change themes (Wickham et al. 2021, 2023a, b), which means that we may be comparing model results to inaccurate observations. Therefore, caution should be exercised when directly comparing our model results to these observations. While this highlights the challenges of comparing model outputs to imperfect observations, it also underlines the importance of ongoing efforts to improve land cover change data collection and analysis methods.

5 Summary

Understanding the connections between food, trade, agriculture, land use, and water resources requires reliable quantitative insights. Gridded quantitative economic models can provide better insights into these relationships. However, validating these models is challenging because of the complex interactions, unprecedented changes, and uncertainties in human decisions. This chapter is about validating such gridded economic models by comparing their predictions to real-world data. It employs SIMPLE-G, a new economic model that can analyze land-use change, water management, and agricultural production. The goal is to demonstrate the potential of gridded economic models for understanding food systems, land use, and water management in sustainable development frameworks.

In this chapter, we draw on methods from different disciplines to validate the SIMPLE-G model. We first perform benchmark replication or calibration by comparing the results of the model to a set of observed data to ensure that the model can replicate a base reference condition. We then use backcasting to employ the model to make predictions about observed past events and compare them to actual observed outcomes from independent data sources not used in the calibration process. This helps ensure that the structural processes that can explain changes in the system are modeled correctly. We also call to use sensitivity analysis to vary the model's assumptions, parameters, and drivers to investigate how the predictions change (Haqiqi et al. 2023). This helps identify the assumptions and parameters of the model that are most important in determining outcomes and ensures the robustness of the model findings. Finally, we use uncertainty quantification and characterization to uncover sources of uncertainty in the results and reveal their implications for the major conclusions of each study. This approach to model validation can be applied to other geospatial economic models and can help uncover sources of uncertainty in the results and reveal their implications for the major conclusions of each study.

Acknowledgment and Competing Interests The authors acknowledge support from the U.-S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreement DE-SC0022141; the National Science Foundation award #2118329: "NSF Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE)," the United States Department of Agriculture AFRI grant #2019-67023-29679, "Economic Foundations of Long Run Agricultural Sustainability," and the National Science Foundation INFEWS award #1855937, "Identifying Sustainability Solutions Through Global-Local-Global Analysis of a Coupled Water-Agriculture-Bioenergy System."

The findings and conclusions presented in this chapter are those of the authors and should not be construed to represent any official determination or policy of the US Department of Agriculture (USDA), the US government, the Department of Energy (DOE), or the National Science Foundation (NSF). Furthermore, we declare that there is no conflict of interest related to this work.

References

Armington, Paul S. 1969. A theory of demand for products distinguished by place of production. *Staff Papers* 16: 159–178. https://doi.org/10.2307/3866403.

Beckman, Jayson, and Stephanie Riche. 2015. Changes to the natural gas, corn, and fertilizer price relationships from the biofuels era. *Journal of Agricultural and Applied Economics* 47: 494–509. https://doi.org/10.1017/aae.2015.22.

Biondi, Daniela, Gabriele Freni, Vito Iacobellis, Giuseppe Mascaro, and Alberto Montanari. 2012. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the Earth, Parts A/B/C* 42–44: 70–76. https://doi.org/10.1016/j.pce.2011.07.037.

Burns, Christopher, Nigel Key, Sarah Tulman, Allison Borchers, and Jeremy Weber. 2018. *Farmland values, land ownership, and returns to farmland, 2000–2016*. Washington, DC: Economic Research Report 245. US Department of Agriculture, Economic Research Service. https://doi.org/10.22004/ag.econ.276249.

Chaddad, Fabio R., and Marcos S. Jank. 2006. The evolution of agricultural policies and agribusiness development in Brazil. *Choices* 21: 85–90. https://doi.org/10.22004/ag.econ.94415.

Chen, Xiaoguang, Haixiao Huang, Madhu Khanna, and Hayri Önal. 2011. Meeting the mandate for biofuels: Implications for land use, food, and fuel prices. In *The intended and unintended effects of US agricultural and biotechnology policies*, ed. Joshua S. Graff Zivin and Jeffrey M. Perloff, 223–267. Chicago: University of Chicago Press.

Clancy, Matthew, Keith Fuglie, and Paul Heisey. 2016. US agricultural R&D in an era of falling public funding. *Amber Waves*. https://doi.org/10.22004/ag.econ.249840.

Fuglie, Keith. 2022. International agricultural productivity. In *Economic research service, USDA*. Washington, DC: US Department of Agriculture, Economic Research Service. https://www.ers.usda.gov/data-products/international-agricultural-productivity/. Accessed 1 Oct 2023.

Fuglie, Keith, Srabashi Ray, Uris Lantz C. Baldos, and Thomas W. Hertel. 2022. The R&D cost of climate mitigation in agriculture. *Applied Economic Perspectives and Policy* 44: 1955–1974. https://doi.org/10.1002/aepp.13245.

Griffith, Rachel, Stephen Redding, and John Van Reenen. 2004. Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *The Review of Economics and Statistics* 86: 883–895. https://doi.org/10.1162/0034653043125194.

Haqiqi, I., Bowling, L., Jame, S., Baldos, U., Liu, J. and Hertel, T., 2023. Global drivers of local water stresses and global responses to local water policies in the United States. *Environmental Research Letters, 18*(6): 065007. https://doi.org/10.1088/1748-9326/acd269

Hansen, Lars Peter, and James J. Heckman. 1996. The Empirical foundations of calibration. *Journal of Economic Perspectives* 10: 87–104. https://doi.org/10.1257/jep.10.1.87.

Hertel, Thomas W., Navin Ramankutty, and Uris Lantz C. Baldos. 2014. Global market integration increases likelihood that a future African Green Revolution could increase crop land use and CO2 emissions. *Proceedings of the National Academy of Sciences* 111: 13799–13804. https://doi.org/10.1073/pnas.1403543111.

Hertel, Thomas W., Lantz C. Uris, and Baldos. 2016. Attaining food and environmental security in an era of globalization. *Global Environmental Change* 41: 195–205. https://doi.org/10.1016/j.gloenvcha.2016.10.006.

IEA. 2018. *World Energy Outlook 2018*. Paris: IEA.

Kersebaum, K.C., K.J. Boote, J.S. Jorgenson, C. Nendel, M. Bindi, C. Frühauf, T. Gaiser, et al. 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. *Environmental Modelling & Software* 72: 402–417. https://doi.org/10.1016/j.envsoft.2015.05.009.

Ludena, Carlos E., Thomas W. Hertel, Paul V. Preckel, Kenneth Foster, and Alejandro Nin. 2007. Productivity growth and convergence in crop, ruminant, and nonruminant production: Measurement and forecasts. *Agricultural Economics* 37: 1–17. https://doi.org/10.1111/j.1574-0862.2007.00218.x.

Marchant, Mary A., and H. Holly Wang. 2018. Theme overview: US–China trade dispute and potential impacts on agriculture. *Choices* 33: 1–3. https://doi.org/10.22004/ag.econ.273328.

Ngo, The An, and Linda See. 2012. Calibration and validation of agent-based models of land cover change. In *Agent-based models of geographical systems*, ed. Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 181–197. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8927-4_10.

Oreskes, N. 1998. Evaluation (not validation) of quantitative models. *Environmental Health Perspectives* 106: 1453–1460. https://doi.org/10.1289/ehp.98106s61453.

Razavi, Saman, and Hoshin V. Gupta. 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and environmental systems models. *Water Resources Research* 51: 3070–3092. https://doi.org/10.1002/2014WR016527.

Razavi, Saman, Anthony Jakeman, Andrea Saltelli, Clémentine Prieur, Bertrand Iooss, Emanuele Borgonovo, Elmar Plischke, et al. 2021. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software* 137: 104954. https://doi.org/10.1016/j.envsoft.2020.104954.

Rykiel, Edward J. 1996. Testing ecological models: The meaning of validation. *Ecological Modelling* 90: 229–244. https://doi.org/10.1016/0304-3800(95)00152-2.

van Vliet, Jasper, Arnold K. Bregt, Daniel G. Brown, Hedwig van Delden, Scott Heckbert, and Peter H. Verburg. 2016. A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software* 82: 174–182. https://doi.org/10.1016/j.envsoft.2016.04.017.

Wang, Sun Ling, Roberto Mosheim, Richard Nehring, and Eric Njuki. 2020. *Productivity is the major driver of US farm sector's economic growth*. Amber Waves. https://doi.org/10.22004/ag.econ.303975.

Wickham, James, Stephen V. Stehman, Daniel G. Sorenson, Leila Gass, and Jon A. Dewitz. 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. *Remote Sensing of Environment* 257: 112357. https://doi.org/10.1016/j.rse.2021.112357.

Wickham, James, K. Anne Neale, Maliha Nash Riitters, Jon Dewitz, Suming Jin, Megan van Fossen, and D. Rosenbaum. 2023a. Where forest may not return in the western United States. *Ecological Indicators* 146: 109756. https://doi.org/10.1016/j.ecolind.2022.109756.

Wickham, James, Stephen V. Stehman, Daniel G. Sorenson, Leila Gass, and Jon A. Dewitz. 2023b. Thematic accuracy assessment of the NLCD 2019 land cover for the conterminous United States. *GIScience & Remote Sensing* 60: 2181143. https://doi.org/10.1080/15481603.2023.2181143.

Wilson, John S., Catherine L. Mann, and Tsunehiro Otsuki. 2004. *Assessing the potential benefit of trade facilitation: A global perspective*. Washington, DC: World Bank.