# Supplementary Material for 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)

## 1 Additional information about gcamland

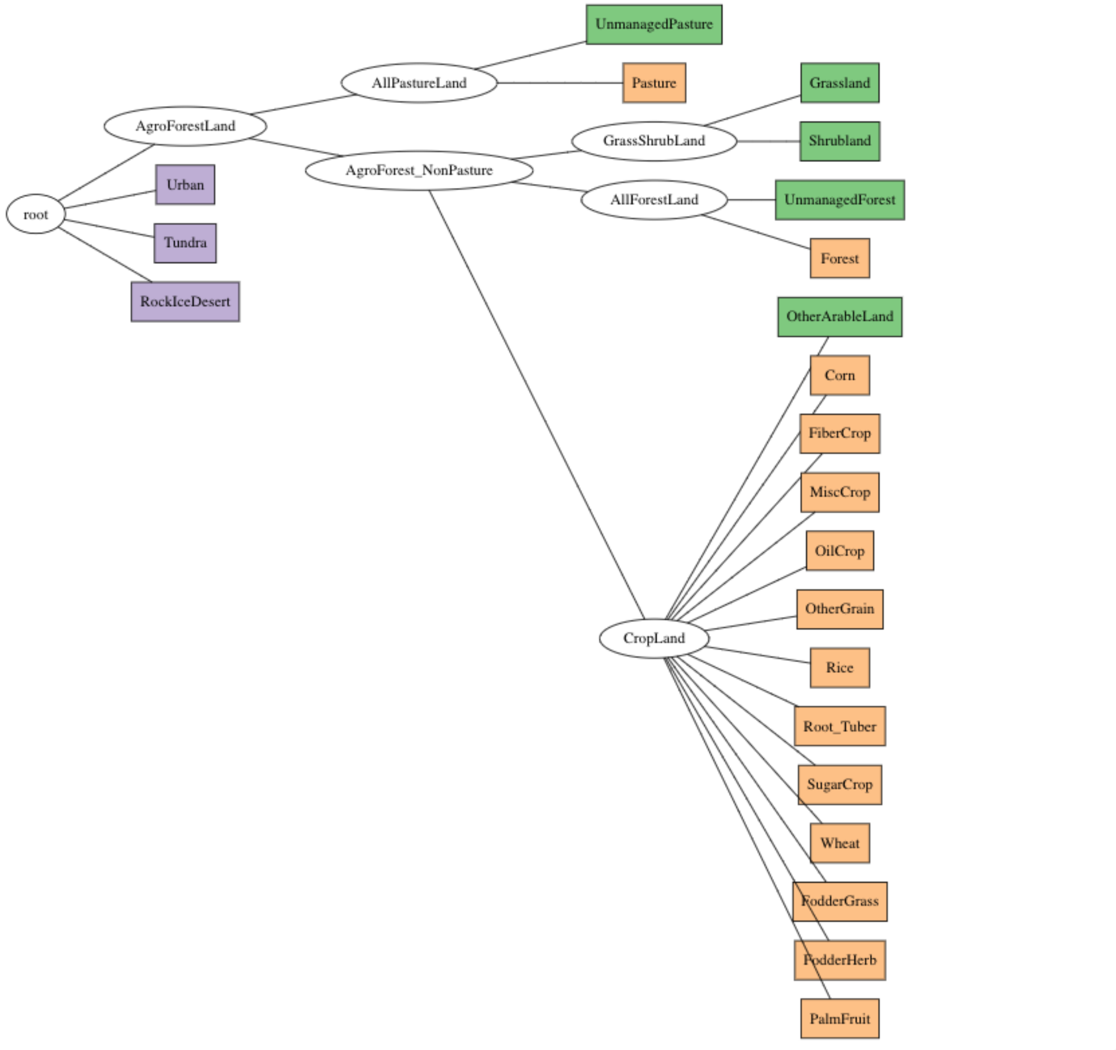



**Figure S1: Nesting diagram for gcamland, using the names of land types and nodes that appear in the model. Boxes indicate different land types, with purple indicating those where land area is held constant over time, green indicates non-commercial land types, and orange indicates land types that produce a product. Ovals indicate nodes, which are used to group land types. See Table S2 for a mapping between the labels for nodes used in gcamland (and this diagram) and the labels used in this paper. Note that gcamland includes on additional cropland type ("biomass") that represents lignocellulosic bioenergy crops (e.g., switchgrass and miscanthus). However, since these were not grown at commercial scale in the historical period, its land area is zero in the simulations described in this paper and thus we have removed it from this figure.**

**Table S1: Land types, data sources, and 1990 values for all gcamland land types.**

| gcamland land type | Commodities and/or land categories included | Land area data for initialization[1] | Yield data for simulation[2] | Price data for simulation[2] | Land area data for comparison[3] | Physical area in 1990 (thous $km^2$) | Harvested area in 1990 (thous $km^2$) | Expected profit in 1990 using Best Adaptive Expectations Model (1975$/thous $km^2$)[9] | Calibrated profit in 1990 using Best Adaptive Expectations Model[10] |
|---|---|---|---|---|---|---|---|---|---|
| Corn | Maize; Maize, green; Popcorn | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 267 | 274 | 15754777 | 14705 |
| FiberCrop | Agave Fibres Nes; Cotton; Fibre Crops Nes; Flax fibre and tow; Hemp Tow Waste; Jute; Manila Fibre (Abaca); Other Bastfibres; Ramie; Sisal | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 47 | 47 | 41093927 | 723 |
| FodderGrass | Forage Products; Grasses Nes for forage; Rye grass for forage and silage; Sorghum for forage and silage | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area for the base year. Held constant in subsequent years.[4] | 70% of the FodderHerb price | N/A[6] | 64 | 66 | 50839308 | 1251 |
| FodderHerb | Alfalfa for forage and silage; Beets for Fodder; Cabbage for Fodder; Carrots for Fodder; | FAO harvested area, converted to planted area | Calculated from FAO production and FAO | Uses alfalfa prices from USDA | N/A[6] | 131 | 135 | 45852685 | 4255 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Clover for forage and silage; Green Oilseeds for Silage; Leguminous for Silage; Maize for forage and silage; Swedes for Fodder; Turnips for Fodder; Vegetables Roots Fodder; Vetches | using a fixed harvested-to-cropped ratio | harvested area for the base year. Held constant in subsequent years.[4] | (USDA, 2011) | | | | | |
| MiscCrop | Almonds, with shell; Anise, badian, fennel, corian.; Apples; Apricots; Arecanuts; Artichokes; Asparagus; Avocados; Bambara beans; Bananas; Beans, dry; Beans, green; Berries Nes; Blueberries; Brazil nuts, with shell; Broad beans, horse beans, dry; Cabbages and other brassicas; Carobs; Carrots and turnips; Cashew nuts, with shell; Cashewapple; Cauliflowers and broccoli; Cherries; Chestnuts; Chick peas; Chicory roots; Chillies and peppers, dry; Chillies and peppers, green; Cinnamon (canella); Citrus fruit, nes; Cloves; Cocoa beans; Coffee, green; Cow peas, dry; Cranberries; Cucumbers and gherkins; Currants; Dates; Eggplants (aubergines); Figs; Fruit Fresh Nes; Fruit, | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 37 | 39 | 80898080 | 481 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | tropical fresh nes; Garlic; Ginger; Gooseberries; Grapefruit (inc. pomelos); Grapes; Hazelnuts, with shell; Hops; Kiwi fruit; Kolanuts; Leeks, other alliaceous veg; Leguminous vegetables, nes; Lemons and limes; Lentils; Lettuce and chicory; Lupins; Mangoes, mangosteens, guavas; Mate; Mushrooms and truffles; Nutmeg, mace and cardamoms; Nuts, nes; Okra; Onions (inc. shallots), green; Onions, dry; Oranges; Other melons (inc.cantaloupes); Papayas; Peaches and nectarines; Pears; Peas, dry; Peas, green; Pepper (Piper spp.); Peppermint; Persimmons; Pigeon peas; Pineapples; Pistachios; Plantains; Plums and sloes; Pulses, nes; Pumpkins, squash and gourds; Pyrethrum,Dried; Quinces; Raspberries; Sour cherries; Spices, nes; Spinach; Stone fruit, nes; Strawberries; String beans; Tangerines, mandarins, clem.; Tea; Tea Nes; | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tobacco, unmanufactured; Tomatoes; Vanilla; Vegetables fresh nes; Walnuts, with shell; Watermelons | | | | | | | | |
| OilCrop | Castor oil seed; Groundnuts, with shell; Hempseed; Jojoba Seeds; Karite Nuts (Sheanuts); Linseed; Melonseed; Mustard seed; Oilseeds, Nes; Olives; Poppy seed; Rapeseed; Safflower seed; Sesame seed; Soybeans; Sunflower seed; Tung Nuts | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 240 | 246 | 18255102 | 12168 |
| OtherGrain | Barley; Buckwheat; Canary seed; Cereals, nes; Fonio; Millet; Mixed grain; Oats; Quinoa; Rye; Sorghum; Triticale | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 91 | 95 | 10077226 | 2274 |
| PalmFruit | Coconuts; Oil palm fruit | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 0 | 0 | 0 | 0 |
| Rice | Rice, paddy | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 11 | 11 | 52170549 | 62 |
| Root_Tuber | Cassava; Potatoes; Roots and Tubers, nes; Sweet potatoes; Taro | FAO harvested area, converted to | Calculated from FAO production | FAO producer prices, | FAO harvested area | 6 | 6 | 108292840 | 19 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (cocoyam); Yams; Yautia (cocoyam) | planted area using a fixed harvested-to-cropped ratio | and FAO harvested area | weighted by production quantity | (FAO, 2020b) | | | | |
| SugarCrop | Sugar beet; Sugar cane; Sugar crops, nes | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices, weighted by production quantity | FAO harvested area (FAO, 2020b) | 9 | 9 | 78293544 | 38 |
| Wheat | Wheat | FAO harvested area, converted to planted area using a fixed harvested-to-cropped ratio | Calculated from FAO production and FAO harvested area | FAO producer prices | FAO harvested area (FAO, 2020b) | 267 | 280 | 11874038 | 14664 |
| OtherArableLand | Fallow or idle land | Calculated as the difference between cropland area from Hyde (Klein Goldewijk et al., 2011) and the sum of all crop planted area | N/A[5] | N/A[5] | N/A | 685 | | 15853409 | 75094 |
| Pasture | Pastureland that is intensively grazed | Calculated from the amount of pasture consumed by livestock and the yield | Yield is set to the yield of hay from GTAP (2009) | Equivalent to the FodderGrass price | N/A | 68 | | 47481863 | 190039 |
| UnmanagedPasture | Pasture | Calculated as the difference between pastureland area from Hyde (Klein | N/A[5] | N/A[5] | N/A | 2313 | | 15853409 | 702264 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Goldewijk et al., 2011) and managed pasture | | | | | | | |
| Forest | Forestland that is actively logged | Calculated from the yield, the rotation length, and the production of wood products from FAO (FAO, 2011)[7] | Yield is derived from its vegetation carbon density, which is from Houghton et al. (1999). | Derived from FAO export volume and export value (FAO, 2011) | | 656 | | 9844587 | 493642 |
| UnmanagedForest | Tropical Evergreen Forest/Woodland, Tropical Deciduous Forest/Woodland, Temperate Broadleaf Evergreen Forest/Woodland, Temperate Needleleaf Evergreen Forest/Woodland, Temperate Deciduous Forest/Woodland, Boreal Evergreen Forest/Woodland, Boreal Deciduous Forest/Woodland, Evergreen/Deciduous Mixed Forest/Woodland | Total forest is calculated as the difference between potential forest from SAGE (Ramankutty and Foley 1999) and the amount of land used for cropland, pastureland, or urban areas on areas that were potential forest. UnmanagedForest is the difference between total forest and Forest | N/A[5] | N/A[5] | Total forest is compared to satellite data from the European Spatial Agency (ESA) Climate Change Initiative (CCI), as reported by the FAO (FAO, 2020a)[8] | 2328 | | 15853409 | 1103659 |
| Grassland | Savanna, Grassland/Steppe | Calculated as the difference between potential grassland from | N/A[5] | N/A[5] | Satellite data from the European Spatial | 589 | NA | 15853409 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | SAGE (Ramankutty and Foley 1999) and the amount of land used for cropland, pastureland, or urban areas on areas that were potential grassland | | | Agency (ESA) Climate Change Initiative (CCI), as reported by the FAO (FAO, 2020a)[8] | | | | |
| Shrubland | Dense Shrubland, Open Shrubland | Calculated as the difference between potential shrubland from SAGE (Ramankutty and Foley 1999) and the amount of land used for cropland, pastureland, or urban areas on areas that were potential shrubland | N/A[5] | N/A[5] | Satellite data from the European Spatial Agency (ESA) Climate Change Initiative (CCI), as reported by the FAO (FAO, 2020a)[8] | 743 | NA | 15853409 | 2 |
| Urban | Built-up land | Urbanland from Hyde (Klein Goldewijk et al., 2011) | N/A[5] | N/A[5] | N/A | 116 | | 15853409 | 15853409 |
| RockIceDesert | Desert, Polar Desert/Rock/Ice | Calculated as the difference between potential rock/ice/desert from SAGE (Ramankutty and Foley | N/A[5] | N/A[5] | N/A | 34 | | 15853409 | 15853409 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1999) and the amount of land used for cropland, pastureland, or urban areas on areas that were potential rock/ice/desert | | | | | | | |
| Tundra | Tundra | Calculated as the difference between potential tundra from SAGE (Ramankutty and Foley 1999) and the amount of land used for cropland, pastureland, or urban areas on areas that were potential tundra | N/A[5] | N/A[5] | N/A | 241 | NA | 15853409 | 15853409 |

[1] gcamland only uses land area in a single base year (either 1975, 1990, or 2005 depending on the configuration in this paper). This land area is used to initialize the model. Land area in all subsequent periods is endogenously estimated in the model. Note that the original source data is listed in this column, but as described in Section 2.4.1 we use reconciled data from GCAMv4.3 for initialization (https://github.com/JGCRI/gcam-core/releases/tag/gcam-v4.3).

[2] gcamland uses yield and price data throughout the simulation period to calculate profits for managed lands. Unless otherwise specified, price information is from FAO 2018a and FAO 2018b; yields are derived from production and harvested area from FAO 2020b.

[3] In calculating model performance (e.g., NRMSE), we used land area for all time periods available as a point of comparison. For some land types, not all modeled years are available.

[4] FAO only includes fodder data through 2011. We use this information to initialize the model, but do not use it beyond initialization since it is unavailable for all simulation years.

[5] For non-commercial land types (those that do not produce a product), we do not use price or yield to calculate profit. The calibration process in gcamland infers an effective profit rate based on base year land allocation. This effective profit rate is calculated such that the land allocation equations return the initial land area. The effective profit rate is held constant in future periods.

[6] FAO only includes fodder crops prior to 2011. Due to this data limitation, we do not compare fodder crops to observations in this paper. See also footnote 4 above.

[7] Land area is equal to forest product production divided by yield multiplied by the rotation length. That is, the managed forest area is the amount of land required to meet current and future production needs in continuous harvest rotations.

[8] The CCI satellite data (FAO 2020a) is only available starting in 1992. For this reason, we cannot use it as initialization data since we need that data for 1975, 1990, and 2005.

[9] Expected profit depends on the expectation type used. For this table, we are using adaptive expectations with the parameters that minimize NRMSE (see Figure 3).

[10] Calibrated profit depends on expectation type and logit exponents. For this table, we are using adaptive expectations with the parameters that minimize NRMSE (see Figure 3).

Table S2 provides information the gcamland nodes, the total land area for each node in 1990, and logit exponents used in this study. As noted in the main text, three of the logit exponents used in gcamland are varied as part of the analysis in this paper. For the remaining logit exponents (root, Pastureland, Grass/shrubs, Forestland), we use the default values used in GCAM. These values were chosen based on heuristics, where larger values are used for land types that are more substitutable. For the root, this is set to zero, as we do not allow conversion into or out of urban, tundra, or rock/ice/desert. For grass/shrubs, the decision to shift between grassland and shrubland is unlikely to be an economic choice; for this reason, we set the logit exponent to a very low value, effectively preserving the shares of grassland vs shrubland in the initial model year. Both the Forestland and Pastureland logit exponents govern substitution between commercial and non-commercial land types; a shift between these land types is not a land conversion (i.e., it does not require re-planting) but a shift in use (i.e., either moving livestock or engaging in logging activities). For this reason, higher logit exponents are chosen. A higher logit exponent governs Pastureland than Forestland as the change in use of pastureland is likely to be easier than the change in use of Forestland. We have chosen not to vary these in this exercise as they do not directly impact the amount of cropland area, which was the output we focused on in the main text. We have done an additional sensitivity analysis quantifying the impact of these logits on cropland area. In this analysis, we doubled each logit one at a time. The area allocated to each crop changes by less than 1% (the largest change in magnitude is -0.12%). These logits do have a larger impact on other land types. For example, doubling the ForestLand logit results in a shift in the distribution of commercial and non-commercial forest, with commercial forest increasing by as much as 27%. However, total forest is largely unchanged (maximum change of 0.22%).

**Table S2: Node mapping, logit specification, and 1990 land areas, including a mapping from names of nodes used in the gcamland code and Figure S1 to descriptors used in this paper where applicable.**

| **gcamland node name** | **Descriptor used in this paper** | **Logit** | **Physical area in 1990 (thous km$^2$)** |
|---|---|---|---|
| root | N/A | Exponent fixed at 0, meaning the total area of the nodes immediately beneath it are held constant in time. | 8942 |
| AgroForestLand | Dynamic land | Varied as part of the analysis in this paper, indicated with the superscript “R” in equation 2 | 8551 |
| AgroForest_NonPasture | Ag, Forest, and Other | Varied as part of the analysis in this paper, indicated with the superscript “A” in equation 2 | 6170 |
| AllPastureLand | Pastureland | Exponent set to 2.7 in all simulations | 2380 |
| GrassShrubLand | Grass/shrubs | Exponent set to 0.05 in all simulations | 1331 |
| AllForestLand | Forestland | Exponent set to 1.575 in all simulations | 2983 |

| | | | |
|---|---|---|---|
| CropLand | Cropland | Varied as part of the analysis in this paper, indicated with the superscript “C” in equation 2 | 1170 |

# 2 Historical Context

## 2.1 Price, yield, and profit

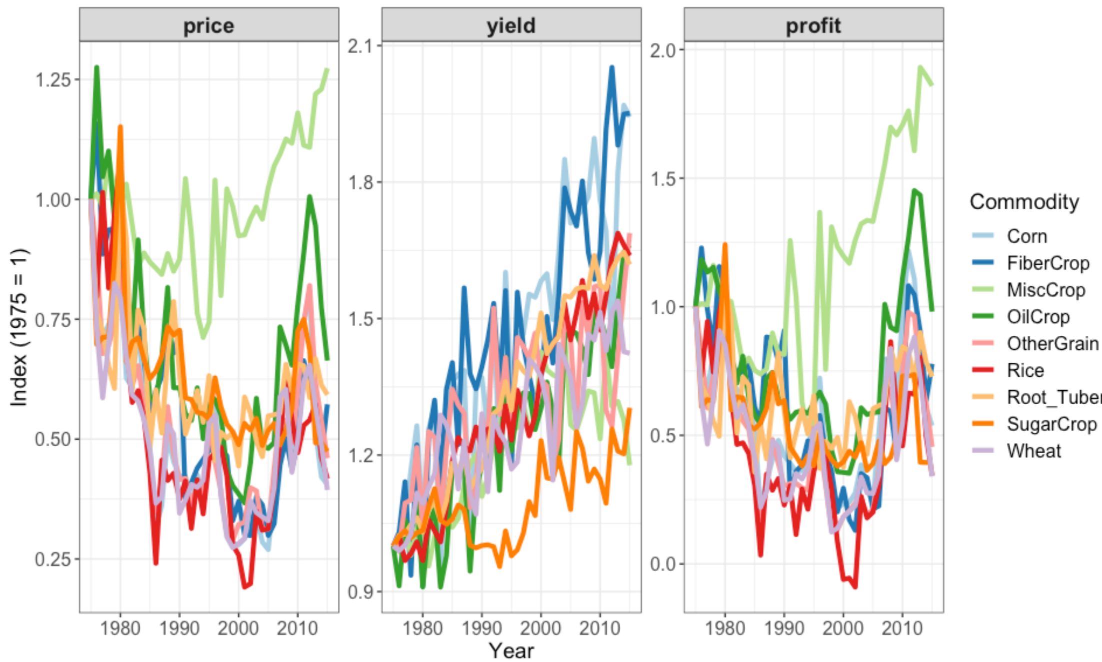



**Figure S2: Price, Yield, and Profit by commodity from 1975 to 2015. Note that we use information from gcamland inputs to ensure consistent time series and units. The original inputs to gcamland are from FAO (for prices and yields) and USDA (for costs). Using only USDA information would not qualitatively change this figure.**

## 2.2 Correlations between profit and yield

Figure S3 shows the change in area by crop as a function of the change in observed profit (i.e., perfect expectations). The correlation between expected profit and cropland area varies by crop, expectation scheme, and the time horizon (Figure S4). All crops have strong correlation between profit and area prior to 1990. However, in recent years, the correlation between profit and land area has changed. For crops where the market has changed dramatically (e.g., corn and soybeans), relying more heavily on recent information provides a better predictor of land area. This suggests that farmers growing these crops are weighting recent information about price and yield more heavily.

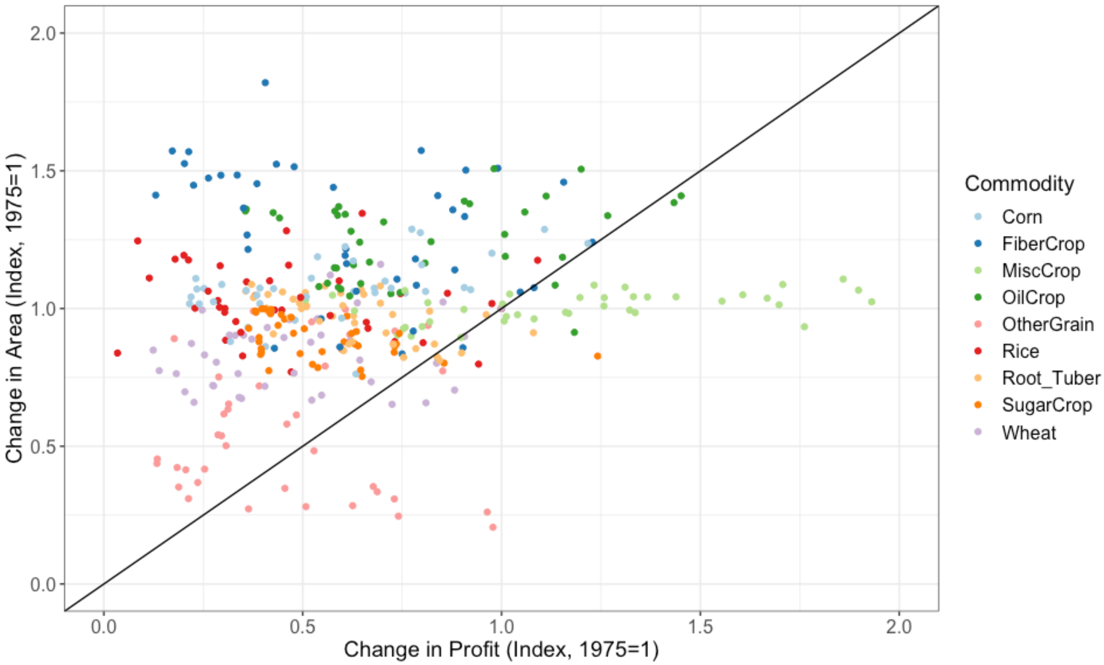



**Figure S3: Correlation between change in observed profit (i.e., perfect expectations) as calculated in gcamland and change in cropland area from FAO (1975-2015). Each point is a crop-year combination. Black line is a 1-to-1 line indicating equal relative changes in profit and area.**

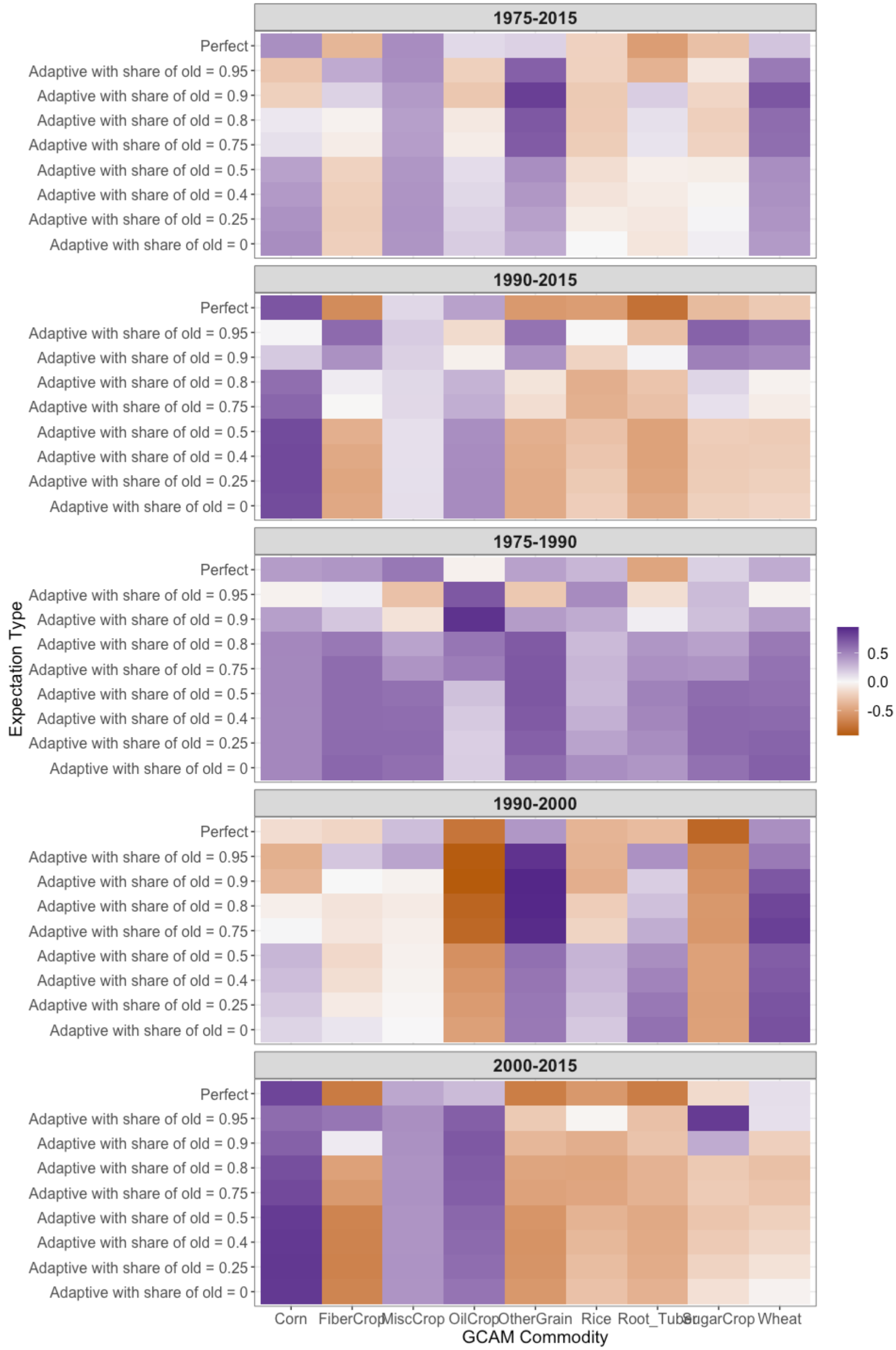



**Figure S4: Pearson correlation coefficient between cropland area and expected profit, using FAO land and GCAM profits. Purple indicates correlation; Orange indicates anti-correlation.**

## 3 Additional Results

In this section, we provide additional information on parameter sets and model results. The main text of this paper focuses on four commodity groups (Corn, Wheat, OtherGrain, and OilCrop), as these four commodities represent the largest land area in the United States. However, gcamland includes twelve commodity groups in total, representing all crops reported by the FAO, and fallow or idled cropland (referred to as other arable land in gcamland). In addition, gcamland includes commercial forest and pasture, as well as several other land cover types, including forest, grassland, shrubland, tundra, rock/ice/desert, and urbanland. We include results for other agricultural commodities and the land cover types in this section.

### 3.1 Parameter sets that minimize NRMSE

**Table S3: Parameter Sets that Minimize NRMSE in the Default model**

| | Adaptive | Linear | Hybrid Linear Adaptive | Perfect |
|---|---|---|---|---|
| NRMSE | 1.40 | 1.87 | 1.58 | 1.67 |
| Logit (Dynamic Land) | 0.41 | 0.46 | 0.41 | 0.25 |
| Logit (Ag, Forest, and Other) | 0.42 | 0.23 | 0.54 | 0.01 |
| Logit (Cropland) | 0.58 | 0.03 | 0.07 | 0.05 |
| Share (Corn, OilCrop) | 0.36 | NA | 0.71 | NA |
| Share (Wheat, OtherGrain) | 0.93 | NA | 0.86 | NA |
| Share (All Other Crops) | 0.99 | NA | 0.94 | NA |
| Number of Years (Corn, OilCrop) | NA | 16.00 | 21.00 | NA |
| Number of Years (Wheat, OtherGrain) | NA | 18.00 | 13.00 | NA |
| Number of Years (All Other Crops) | NA | 7.00 | 10.00 | NA |

### 3.2 Expected price, yield, and profit

Figure S5 shows the expected price, yield, and profit for the different expectation types, using the parameters that minimize NRMSE for those expectation types.

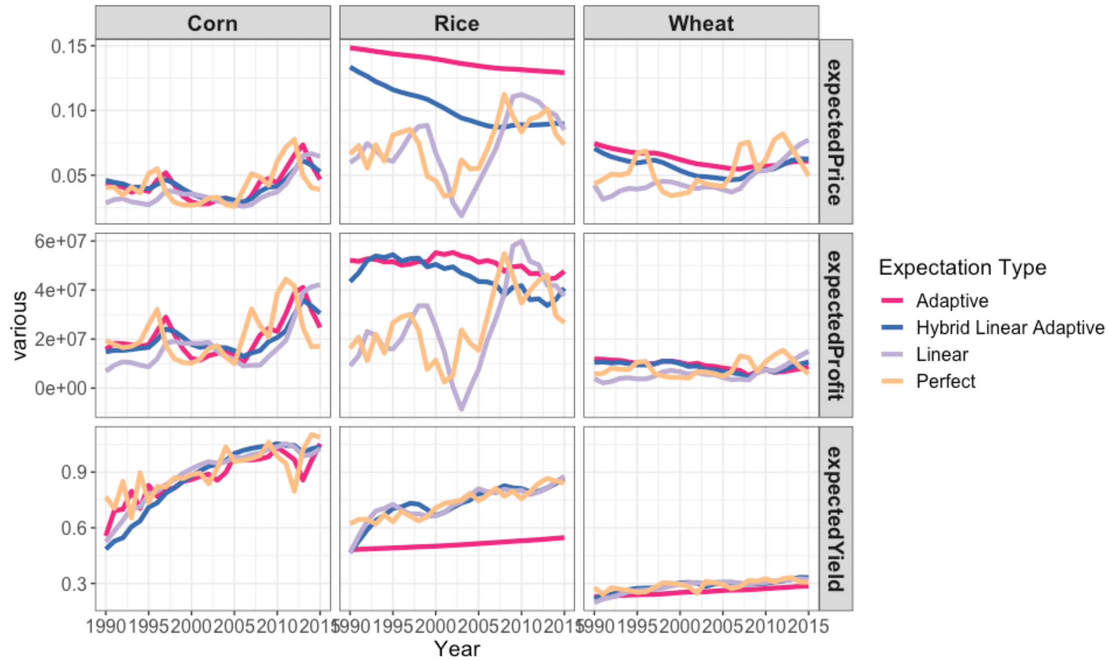
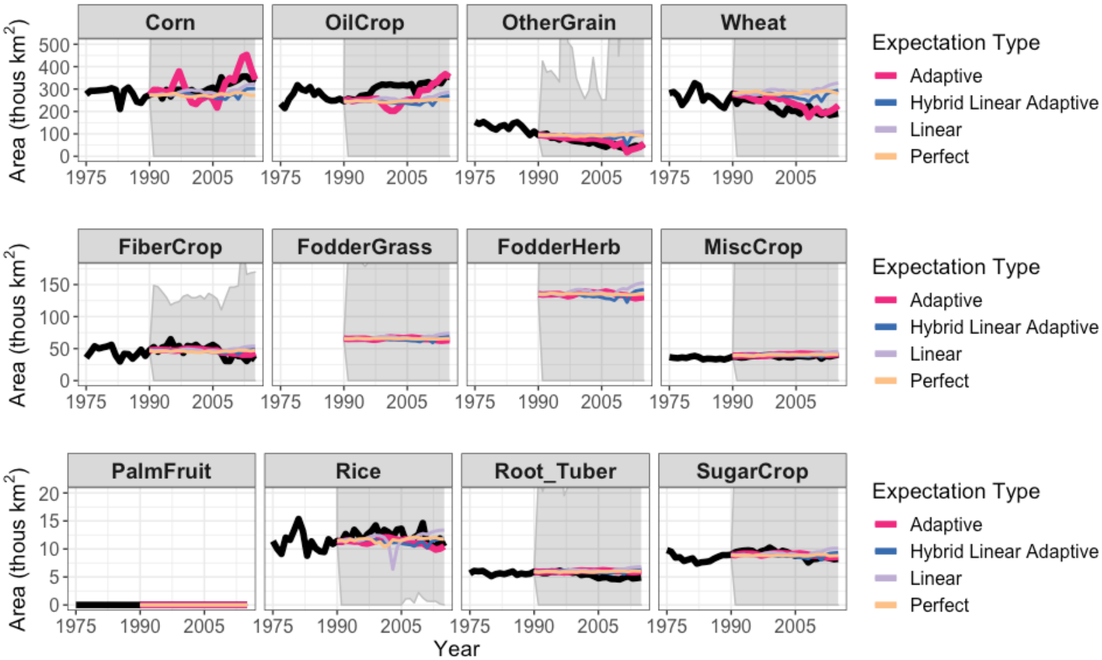



**Figure S5: Expected price, profit, and yield over time by expectation type and crop for the Default model.**

### 3.3 Comparing modelled land to observations in the Default model

#### 3.3.1 Crops

Figure S6, Figure S7, and Table S4 show three different ways of displaying the difference between observations and simulation results. Figure S6 shows time series of all crops for both gcamland simulation results (colored lines) and observations (black lines). Figure S7 shows simulation versus observation as a scatter plot to show correlations. Finally, Table S4 summarizes the error (simulation minus observation) in both absolute (million km$^2$) and percentage terms, as well as including NRMSE for each expectation type and crop. We include all three metrics in this table; however, in this study, we primarily use NRMSE. Normalized measures of error are key for interpreting whether a simulated data set acceptably replicates available observational data. While normalizing to present errors in terms of percentages is common, this can result in large magnitude percentage errors when dealing with multiple variables (land types) with a wide range of magnitudes. Given the significant difference in land historically allocated to different uses in the United States (e.g, the PalmFruit vs Corn commodities in gcamland) and the fact that we are seeking parameter sets to minimize error measures across these commodities, this can lead to misleading results. Rather, we follow the literature normalizing by the standard deviations of observations (Nash and Suttcliff 1970; Willmott 1981; Legates and McCabe 1999; Willmott et al 2012; Tebaldi et al 2020), captured in our NRMSE. This allows a benchmark of whether the discrepancies between simulated and observed data fall within the natural variability of the observed data, giving a statistically justifiable benchmark to determine whether those discrepancies are acceptably sized.




120 **Figure S6: Harvested cropland area (total and by crop) over time by expectation type in the Default Model. 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 observations are missing from this figure for fodder crops (FodderGrass and FodderHerb) due to data limitations.**

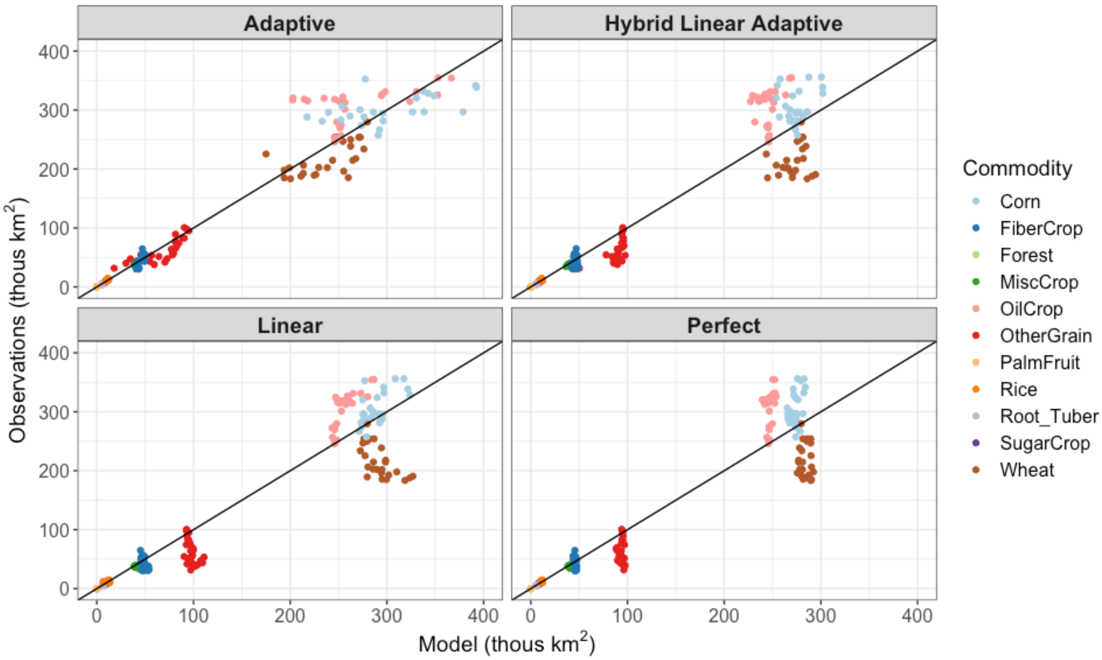



**Figure S7: Model vs. Observations by Crop and Expectation Type (Harvested Area) in the Default Model. Each point is a crop-year combination; each panel shows a different expectation type.**

**Table S4: Relative and absolute error by crop and expectation type in the Default Model. Values indicate the median and range across years for each commodity and expectation type.**

| Expectation type | gcamland land type | Error (million $km^2$) | Error (%) | NRMSE |
|---|---|---|---|---|
| Adaptive | Corn | 8 (-75 to 96) | 2 (-25 to 28) | 1.67 |
| Adaptive | FiberCrop | -1 (-18 to 13) | -2 (-27 to 42) | 0.77 |
| Adaptive | MiscCrop | 3 (-1 to 7) | 7 (-2 to 21) | 2.21 |
| Adaptive | OilCrop | -31 (-118 to 27) | -10 (-37 to 8) | 1.78 |
| Adaptive | OtherGrain | 9 (-14 to 28) | 12 (-43 to 68) | 0.72 |
| Adaptive | PalmFruit | 0 (0 to 0) | -100 (-100 to -100) | 2.01 |
| Adaptive | Rice | -1 (-4 to 0) | -8 (-25 to 3) | 1.20 |
| Adaptive | Root_Tuber | 1 (0 to 1) | 14 (-3 to 32) | 1.67 |
| Adaptive | SugarCrop | 0 (-1 to 1) | 0 (-11 to 19) | 0.83 |
| Adaptive | Wheat | 20 (-50 to 75) | 10 (-22 to 40) | 1.14 |
| Hybrid Linear Adaptive | Corn | -23 (-95 to 18) | -8 (-27 to 7) | 1.47 |
| Hybrid Linear Adaptive | FiberCrop | -4 (-18 to 17) | -8 (-28 to 57) | 0.98 |
| Hybrid Linear Adaptive | MiscCrop | 2 (0 to 4) | 5 (0 to 12) | 1.59 |
| Hybrid Linear Adaptive | OilCrop | -77 (-95 to 0) | -24 (-29 to 0) | 2.16 |
| Hybrid Linear Adaptive | OtherGrain | 27 (-6 to 52) | 42 (-5 to 137) | 1.58 |
| Hybrid Linear Adaptive | PalmFruit | 0 (0 to 0) | -100 (-100 to -100) | 2.01 |
| Hybrid Linear Adaptive | Rice | -1 (-3 to 2) | -8 (-23 to 16) | 1.26 |
| Hybrid Linear Adaptive | Root_Tuber | 1 (0 to 2) | 10 (0 to 33) | 1.58 |
| Hybrid Linear Adaptive | SugarCrop | 0 (-1 to 1) | -5 (-13 to 14) | 1.02 |
| Hybrid Linear Adaptive | Wheat | 59 (0 to 104) | 28 (0 to 56) | 2.17 |
| Linear | Corn | -13 (-76 to 21) | -5 (-21 to 8) | 0.98 |
| Linear | FiberCrop | -3 (-20 to 23) | -6 (-30 to 74) | 1.18 |
| Linear | MiscCrop | 3 (0 to 8) | 8 (1 to 24) | 2.70 |
| Linear | OilCrop | -58 (-78 to 0) | -18 (-24 to 0) | 1.71 |
| Linear | OtherGrain | 35 (-8 to 66) | 59 (-8 to 207) | 2.10 |
| Linear | PalmFruit | 0 (0 to 0) | -100 (-100 to -100) | 2.01 |
| Linear | Rice | 0 (-6 to 3) | -1 (-47 to 32) | 1.62 |
| Linear | Root_Tuber | 1 (0 to 2) | 15 (-3 to 42) | 2.12 |
| Linear | SugarCrop | 0 (-1 to 2) | 0 (-10 to 24) | 1.36 |
| Linear | Wheat | 82 (0 to 136) | 38 (0 to 74) | 2.97 |
| Perfect | Corn | -27 (-80 to 17) | -9 (-23 to 6) | 1.47 |
| Perfect | FiberCrop | -4 (-19 to 17) | -8 (-30 to 54) | 1.05 |
| Perfect | MiscCrop | 2 (0 to 6) | 5 (0 to 18) | 1.77 |
| Perfect | OilCrop | -71 (-104 to 0) | -22 (-29 to 0) | 2.10 |
| Perfect | OtherGrain | 33 (-7 to 64) | 55 (-6 to 203) | 1.84 |
| Perfect | PalmFruit | 0 (0 to 0) | -100 (-100 to -100) | 2.01 |
| Perfect | Rice | -1 (-3 to 2) | -4 (-23 to 20) | 1.29 |
| Perfect | Root_Tuber | 1 (0 to 1) | 10 (-5 to 32) | 1.59 |
| Perfect | SugarCrop | 0 (-1 to 1) | -3 (-14 to 16) | 1.03 |
| Perfect | Wheat | 71 (0 to 106) | 34 (0 to 58) | 2.55 |

We see similar results when comparing shares of harvested area by crop to observations (Figure S8) as shown when comparing absolute land area.

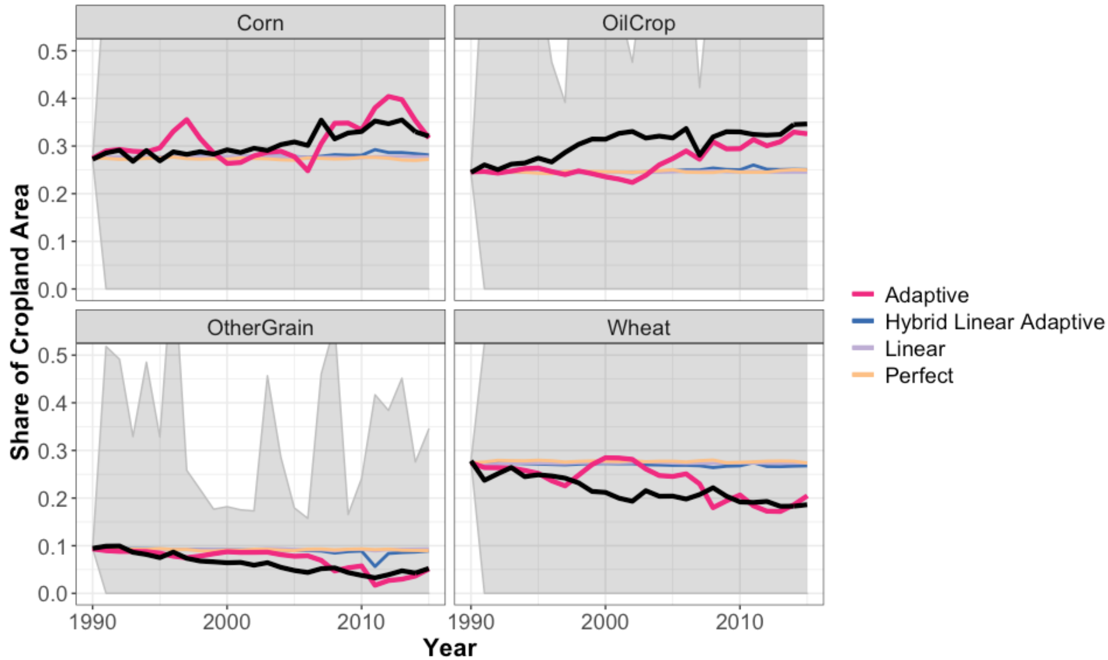



**Figure S8: Cropland Share over Time by Crop and Expectation Type in the Default Model. Black lines are observations (FAO). Gray shading is the range across all gcamland simulations.**

### 3.3.2 Other land types

Due to differences in definitions of land cover between gcamland and the CCI land cover product, Grassland and Shrubland do not match in absolute value between gcamland and the observation data (Figure S9); however, the trends are reasonably similar (Figure S10). Forest is much more consistent, both in terms of magnitude and trends.

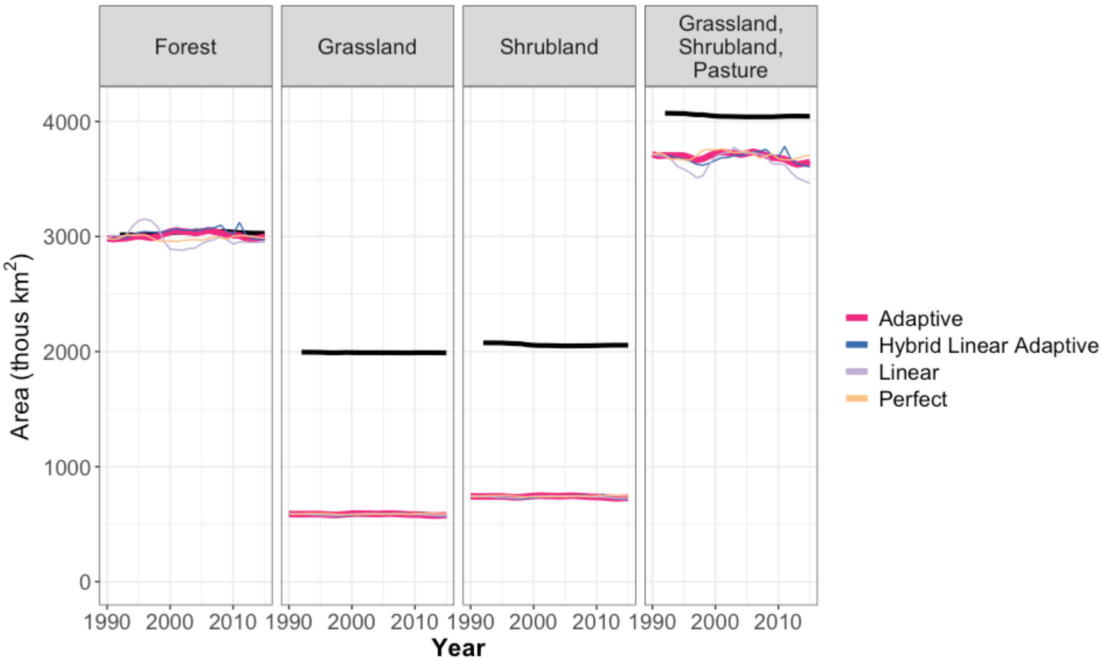



**Figure S9: Land Cover over Time by Expectation Type in the Default Model. Black lines are data from the CCI satellite (downloaded from FAOSTAT). Forest is the sum of commercial and non-commercial forest in gcamland, since CCI does not distinguish these categories. The right panel shows the sum of Grassland, Shrubland, and Pasture to address issues of inconsistent land type definitions between gcamland and CCI. Note that all four expectation types produce similar values for Grassland and Shrubland, so these lines are overlapping.**

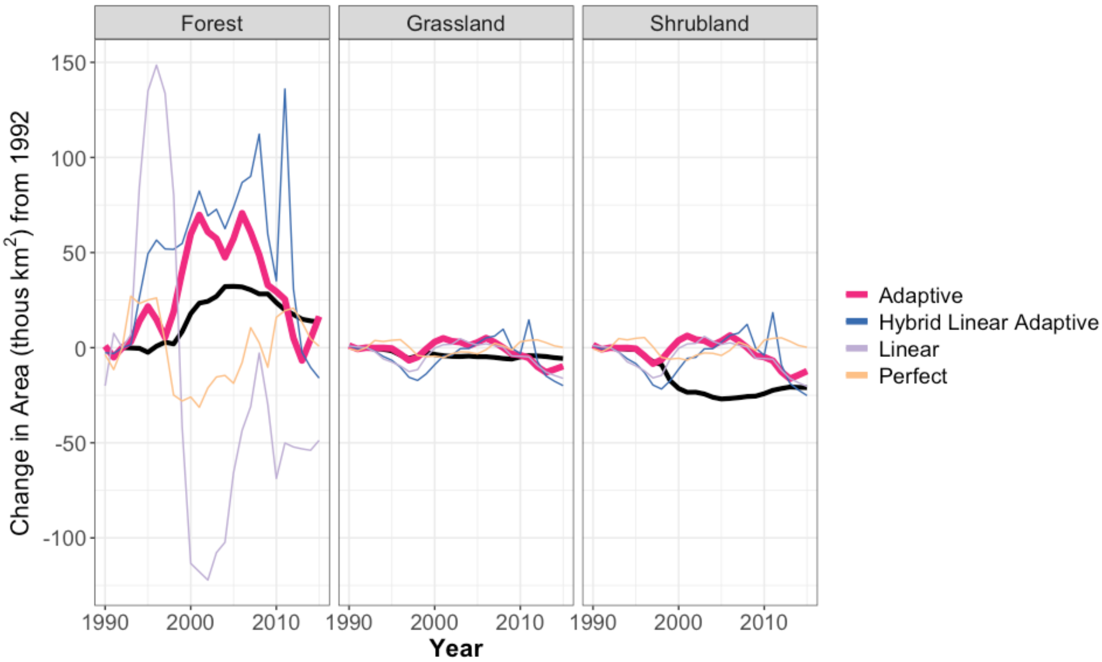



**Figure S10: Land Cover Change from 1992 over Time by Expectation Type in the Default Model. Changes in area are relative to 1992 since that is the first year of CCI satellite data. Forest is the sum of commercial and non-commercial forest in gcamland, since CCI does not distinguish these categories. Black lines are data from the CCI satellite (downloaded from FAOSTAT).**

165 Finally, Figures S6-S10 and Table S4 focus on comparing gcamland simulation results to observations for categories or sums of categories where observation data is present. However, there are other land types included in gcamland (see Figure S1 and Table S1). Figure S11 shows the evolution of all individual land categories in gcamland for the default simulations, with observation data plotted when it is available for the individual category.

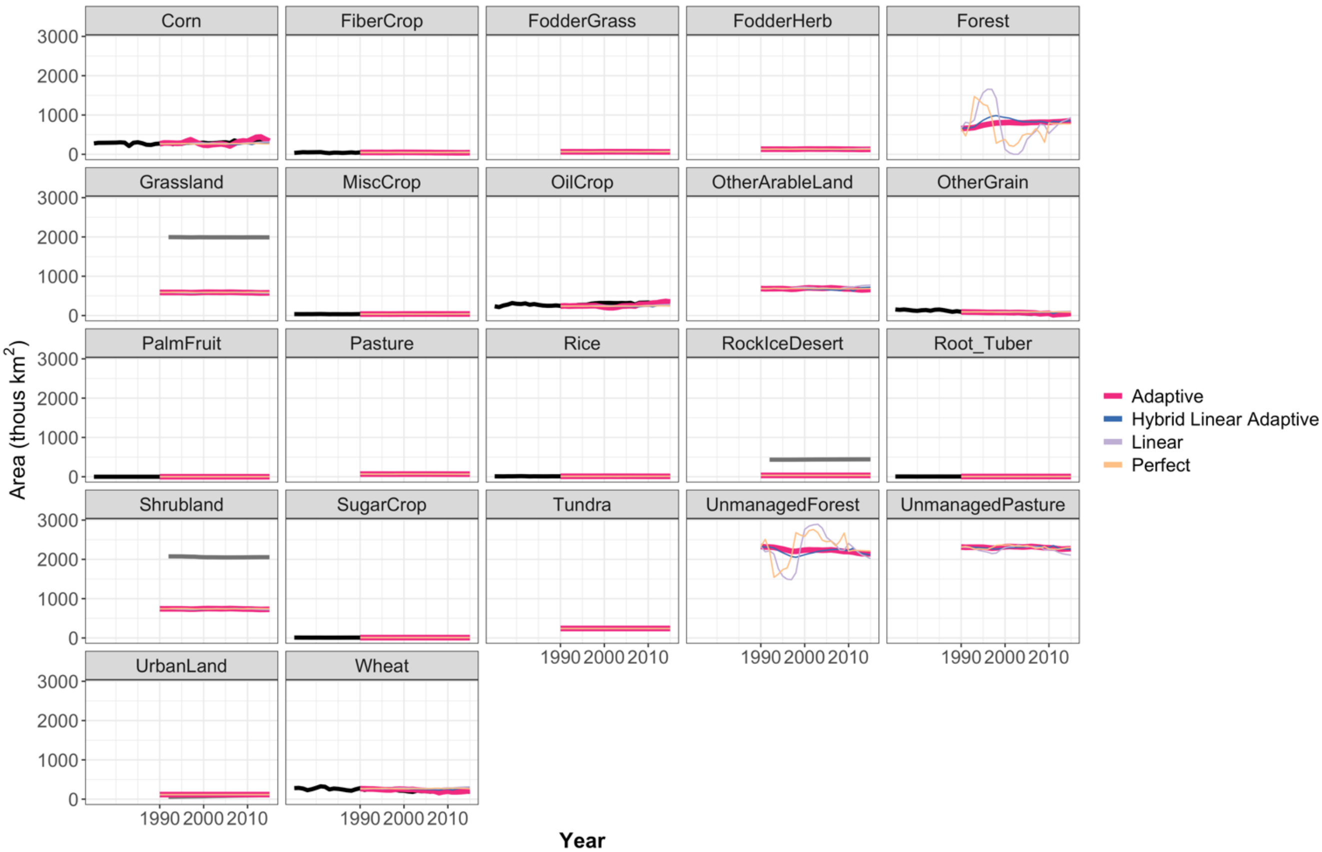



**Figure S11: Area of all land types over time by Expectation Type in the Default Model. Black lines are observation data either from the CCI satellite (downloaded from FAOSTAT) or the FAO. Observations are only included for categories where we have data for that specific category (i.e., forest is excluded because our observation data is the sum of Forest and UnmanagedForest). Note that observations are included for UrbanLand; however, they are overlapping with the simulation lines and thus difficult to see in this figure. The gcamland outputs shown in this figure are also provided as a supplementary Excel spreadsheet.**

Figure S12 shows the NRMSE from the numerically optimal model in this paper to the NRMSE from the previous GCAM hindcast efforts (Snyder et al., 2017). We see improvements in all crops compared to the GCAM default in that paper ("Snyder2017_AY"). However, the variants from Snyder et al. (2017) that forecast yields and explicitly include biofuels policies ("Snyder2017_FYB") outperform the default assumptions in this paper for Corn and OilCrop when NRMSE is minimized across all crops. The parameter sets that explicitly minimize NRMSE for Corn only or for OilCrop only (labeled "ThisPaper_SingleCrop") in this paper have lower NRMSE than all variants in the Snyder et al. (2017) paper.

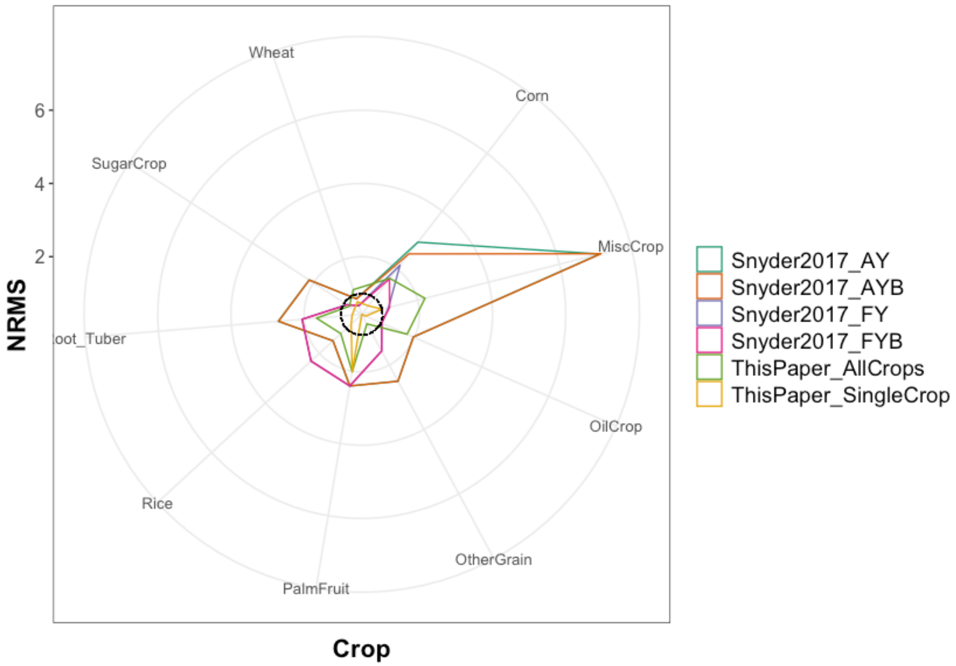



**Figure S12: Comparing NRMSE from this paper to previous GCAM hindcast results from Snyder et al. (2017). Colored lines are different simulations from either the Snyder et al. (2017) paper or this paper.**

## 4 Additional results from the sensitivity analyses

### 4.1 Sensitivity to model assumptions

**Table S5: Parameter sets that minimize NRMSE for different modeling assumptions. Only the best model per assumption set is shown.**

| | Default | Same Parameters | With Subsidy |
|---|---|---|---|
| Expectation Type | Adaptive | Adaptive | Adaptive |
| NRMSE | 1.399 | 1.531 | 1.4560 |
| Logit (Dynamic Land) | 0.412 | 2.048 | 0.412 |
| Logit (Ag, Forest, and Other) | 0.424 | 0.531 | 0.424 |
| Logit (Cropland) | 0.577 | 0.134 | 0.577 |
| Share (Corn, OilCrop) | 0.356 | 0.989 | 0.356 |
| Share (Wheat, OtherGrain) | 0.934 | 0.989 | 0.934 |
| Share (All Other Crops) | 0.988 | 0.989 | 0.988 |

### 4.2 Sensitivity to objective function

#### 4.2.1 Optimizing for different objective functions

**Table S6: Parameter sets that minimize different objective functions. Only the best model per assumption set is shown.**

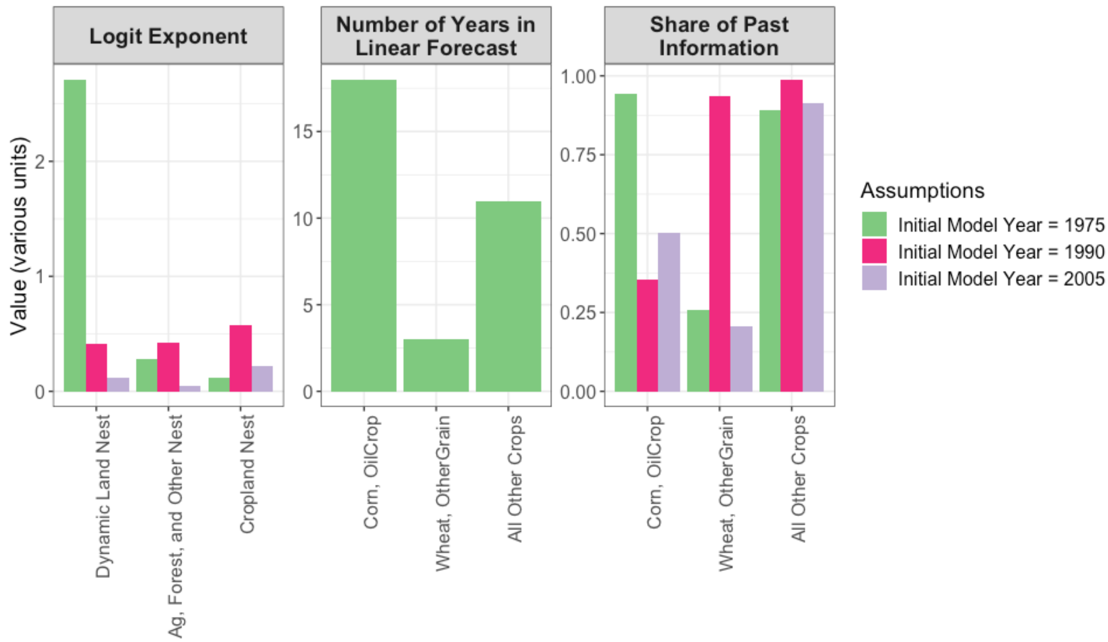
| | NRMSE | RMSE | Bias | KGE |
|---|---|---|---|---|
| Expectation Type | Adaptive | Adaptive | Linear | Hybrid Linear Adaptive |
| Objective Value | 1.399 | 16.14 | 5.51 | 0.761 |
| Logit (Dynamic Land) | 0.41 | 0.41 | 2.18 | 0.53 |
| Logit (Ag, Forest, and Other) | 0.42 | 0.42 | 1.38 | 0.05 |
| Logit (Cropland) | 0.58 | 0.58 | 0.28 | 0.37 |
| Share (Corn, OilCrop) | 0.36 | 0.36 | NA | 0.61 |
| Share (Wheat, OtherGrain) | 0.93 | 0.93 | NA | 0.95 |
| Share (All Other Crops) | 0.99 | 0.99 | NA | 0.93 |
| Number of Years (Corn, OilCrop) | NA | NA | 2 | 19 |
| Number of Years (Wheat, OtherGrain) | NA | NA | 25 | 15 |
| Number of Years (All Other Crops) | NA | NA | 5 | 12 |

#### 4.2.2 Optimizing for different land types

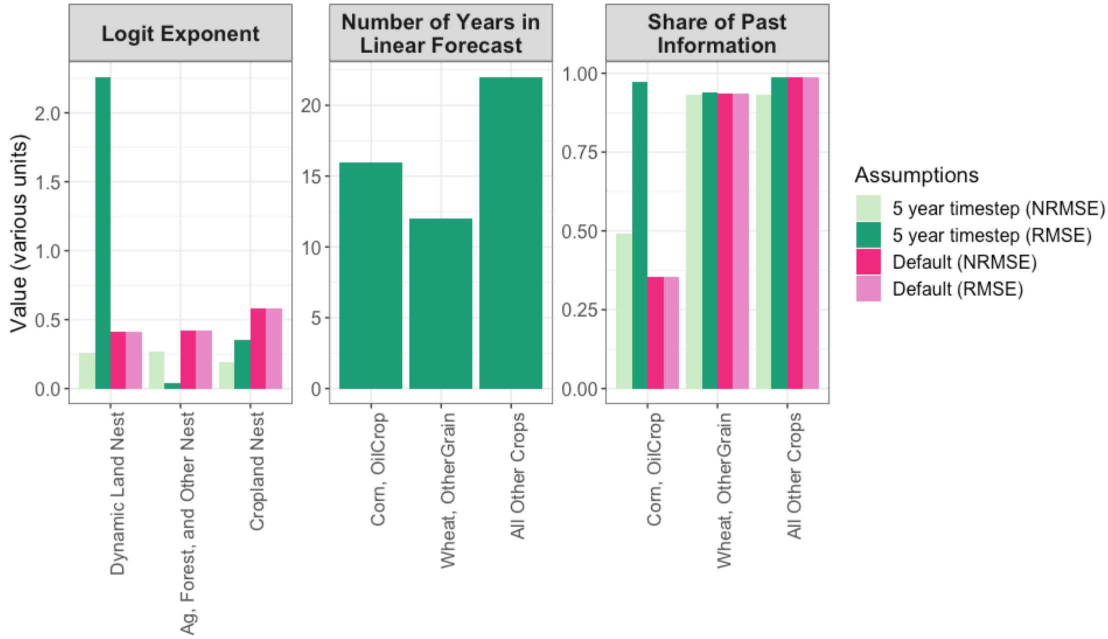
**Table S7: Parameter sets that minimize NRMSE over different land types. Only the best model per assumption set is shown.**

| | All Crops | Corn | OilCrop | Wheat | OtherGrain |
|---|---|---|---|---|---|
| Expectation Type | Adaptive | Hybrid Linear Adaptive | Linear | Adaptive | Adaptive |
| NRMSE | 1.399 | 0.732 | 0.545 | 0.926 | 0.496 |
| Logit (Dynamic Land) | 0.412 | 0.224 | 2.459 | 2.454 | 0.815 |
| Logit (Ag, Forest, and Other) | 0.424 | 1.383 | 0.279 | 1.976 | 2.340 |
| Logit (Cropland) | 0.577 | 0.222 | 0.093 | 0.811 | 0.603 |
| Share (Corn, OilCrop) | 0.356 | 0.988 | NA | 0.294 | 0.789 |
| Share (Wheat, OtherGrain) | 0.934 | 0.497 | NA | 0.953 | 0.928 |
| Share (All Other Crops) | 0.988 | 0.876 | NA | 0.929 | 0.957 |
| Number of Years (Corn, OilCrop) | NA | 21 | 18 | NA | NA |
| Number of Years (Wheat, OtherGrain) | NA | 4 | 16 | NA | NA |
| Number of Years (All Other Crops) | NA | 20 | 18 | NA | NA |

### 4.3 Sensitivity to initial model year and/or time step




200 **Figure S13: Parameter sets that minimize NRMSE for different initial model years.**




205 **Figure S14: Parameter sets that minimize NRMSE or RMSE for different model time steps. Note we use RMSE to compare across model configurations with different timesteps because NRMSE normalizes by standard deviation which varies depending on the timestep (see Section 5.4).**

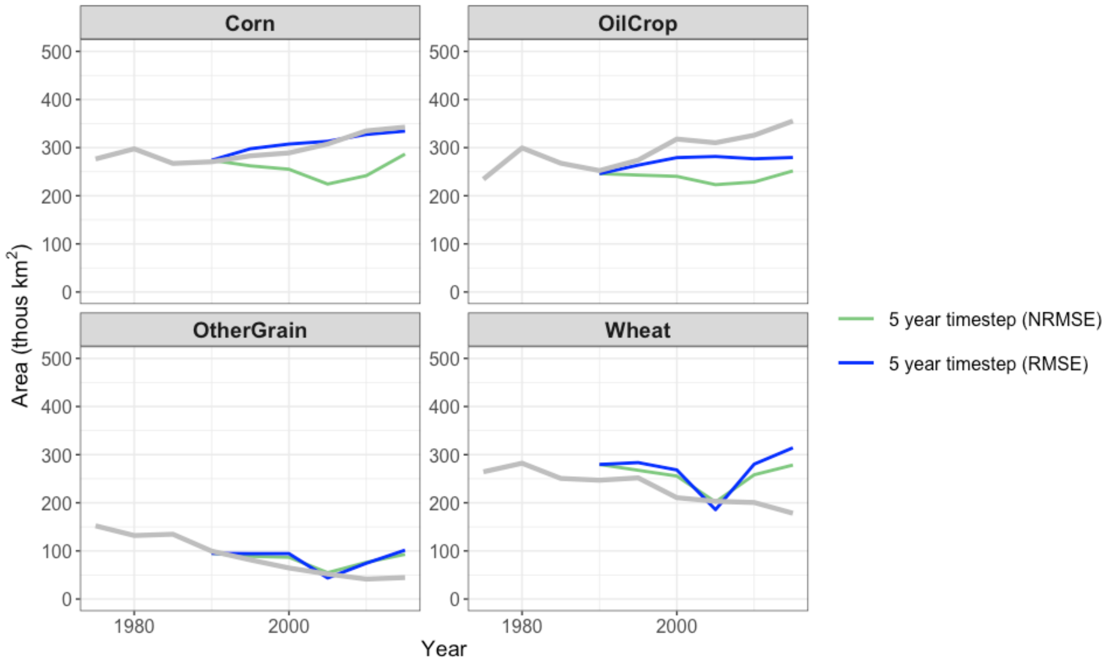



**Figure S15: Harvested area by crop with five year time steps, including parameter sets that minimize RMSE and parameter sets that minimize NRMSE. Gray line is five year average of observations (FAO). Green line is NRMSE; blue line is RMSE.**

210

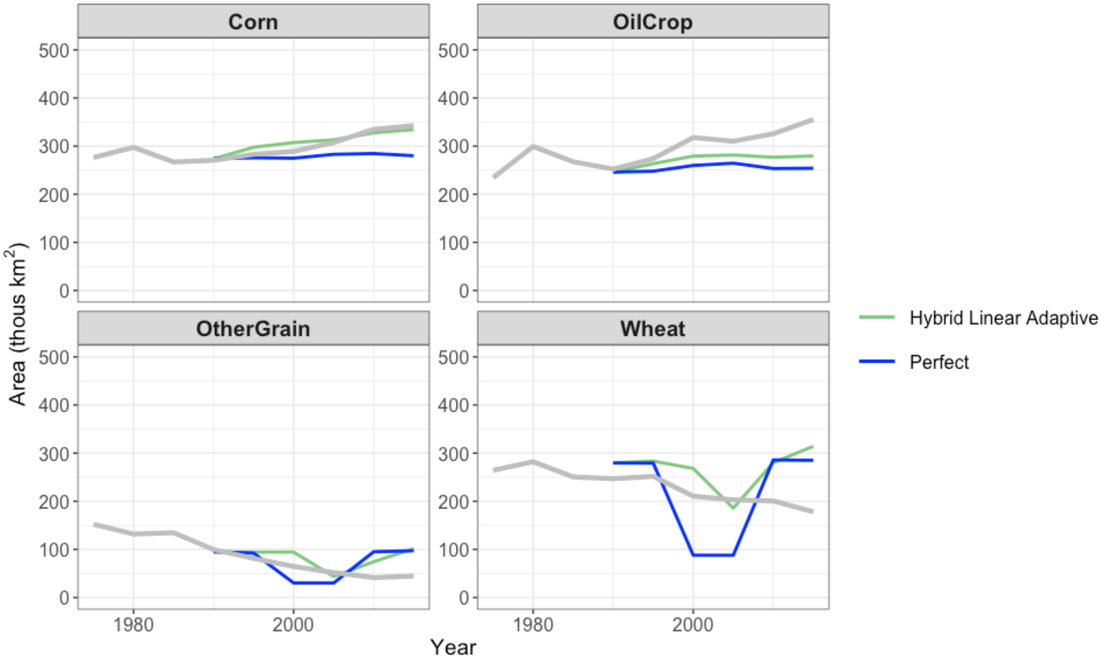



**Figure S16: Harvested area by crop with five-year time steps, using parameter sets that minimize RMSE. Gray line is five-year average of observations (FAO). Green line is adaptive expectations; blue line is perfect expectations.**

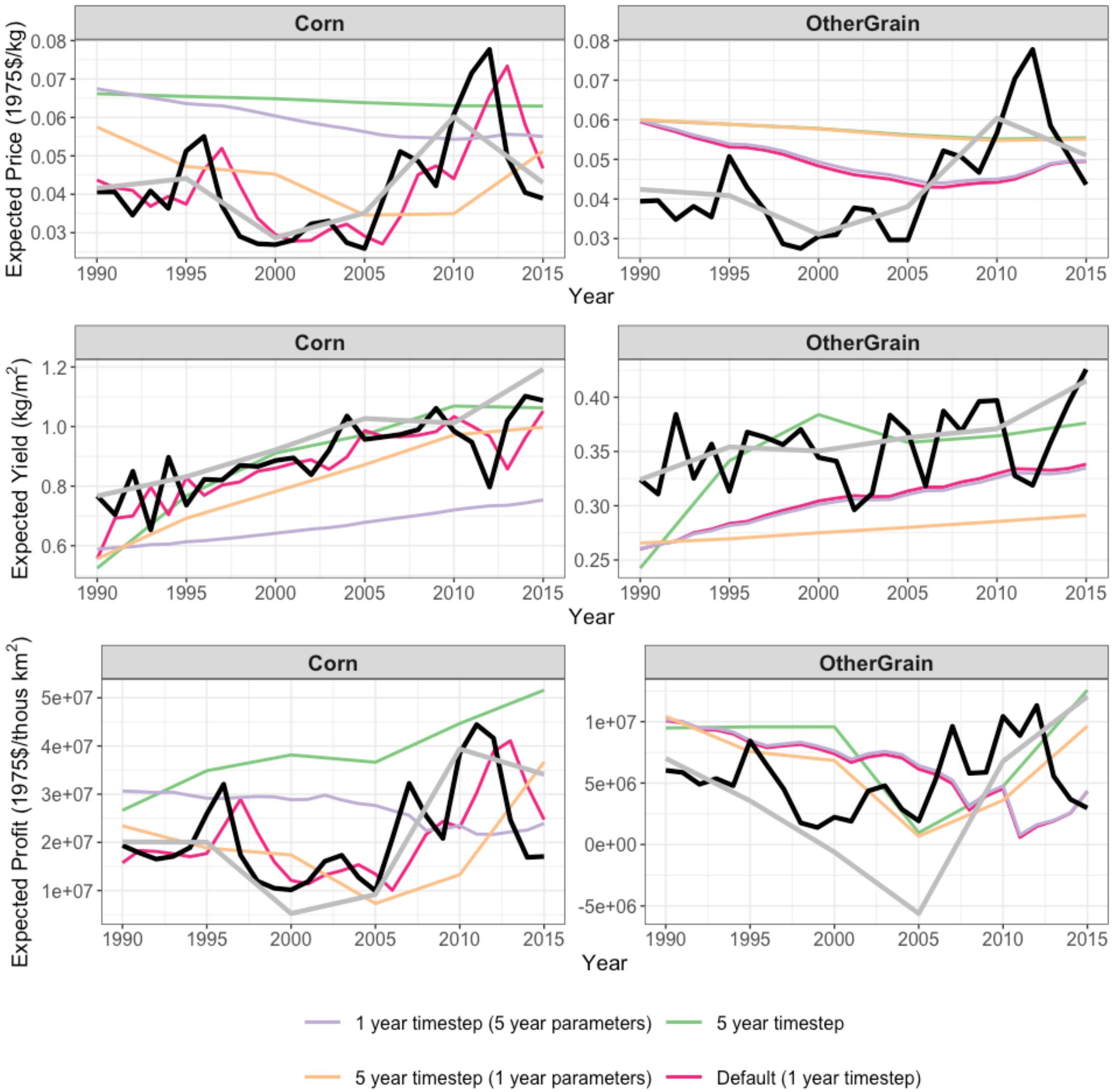



**Figure S17: Expected price, yield, and profit by crop under different model time steps and parameter sets, using parameter sets that minimize RMSE. Black line is annual observations; gray line is five-year averages of observations (FAO). Colored lines are different combinations of time steps and parameter sets. Note that the parameter set that minimizes RMSE also minimizes NRMSE for the Default model.**

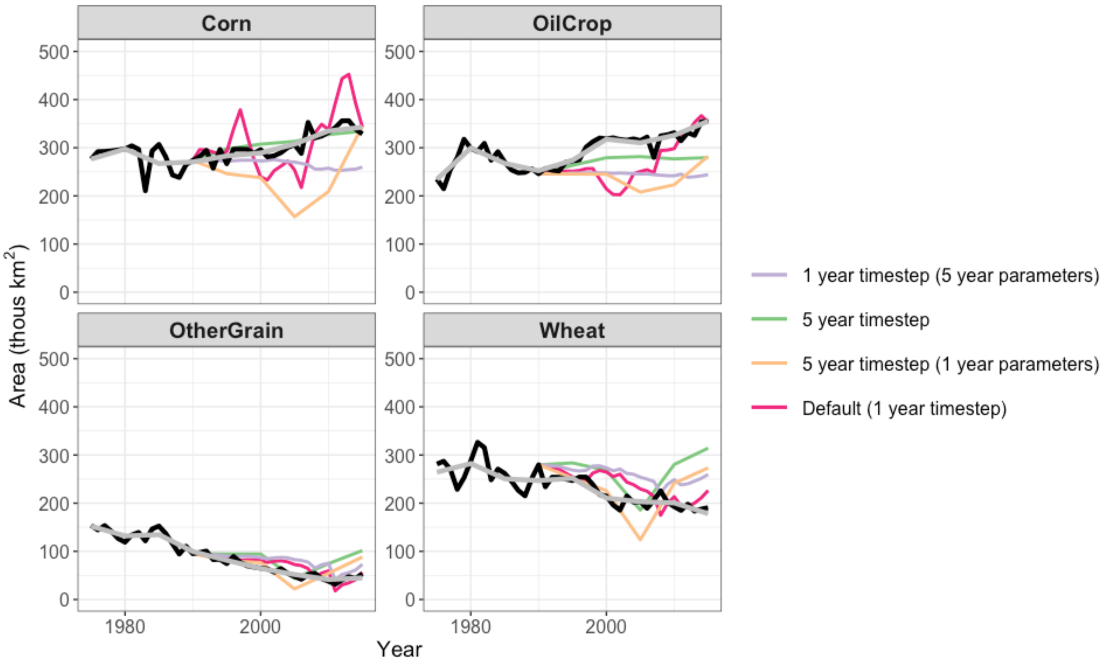



**Figure S18: Harvested area by crop under different model time steps and parameter sets, using parameter sets that minimize RMSE. Black line is annual observations; gray line is five-year averages of observations (FAO). Colored lines are different combinations of time steps and parameter sets. Note that the parameter set that minimizes RMSE also minimizes NRMSE for the Default model.**

## References

FAO: Forestry Production and Trade, FAOSTAT [online] Available from: http://www.fao.org/faostat/en/#data/FO, 2011.

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

GTAP. 2009. GTAP Land Use Data Base, Release 2.1, 2009. Global Trade and Analysis Project. Available at: https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=1900

Houghton, R.A. 1999. The annual net flux of carbon to the atmosphere from changes in land use 1850-1990. Tellus 51B: 298-313.

Klein Goldewijk, K., Beusen, A., van Drecht, G. and de Vos, M. (2011), The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography, 20: 73-86. https://doi.org/10.1111/j.1466-8238.2010.00587.x

Legates, D. R. and McCabe, G. J.: Evaluating the use of "goodnessof-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233–241, 1999

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, 1970.

Ramankutty, N., and J.A. Foley. 1999. Estimating historical changes in global land cover: croplands from 1700 to 1992. Global Biogeochemical Cycles 13(4): 997-1027.

Tebaldi, C., A. Armbruster, H. P. Engler, and R. Link, 2020: Emulating climate extreme indices. *Environ. Res. Lett.*, **15**, 74006, https://doi.org/10.1088/1748-9326/ab8332.

USDA 2011. Prices Received for Alfalfa Hay, Baled, Washington. National Agricultural Statistics Service, U.S. Department of Agriculture.

Willmott, C. J.: On the validation of models, Phys. Geogr., 2, 184– 194, 1981.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, Int. J. Climatol., 32, 2088–2094, 2012.