Supporting Information for
Analysis of uncertainty in emissions from biofuels-induced land use change

Richard J. Plevin*[1], Jayson Beckman[2], Alla Golub[3], Julie Witcover[1], and Michael O'Hare[4]

[1]Institute of Transportation Studies, University of California–Davis
[2]Economic Research Service, US Department of Agriculture
[3]Center for Global Trade Analysis, Purdue University
[4]Goldman School of Public Policy, University of California–Berkeley

November 9, 2014

Contents

*Corresponding author (plevin@ucdavis.edu)

List of Figures

List of Tables

S1 Model components

Modeling system comprises GTAP-BIO-ADV and the Agro-ecological zone Emission Factor (AEZ-EF) model.

S1.1 GTAP-BIO-ADV

The computable general equilibrium (CGE) framework used in this study[1] is an extension of GTAP-BIO-ADV documented in Tyner et al. (2010).[2]. First, the data base employed in this study is further disaggregated to explicitly introduce various vegetable oils (soy, palm, rape and other vegetable oils), and to include vegetable oil-specific biodiesels; sorghum is separated from coarse grains and ethanol from sorghum is a separate biofuel (Taheripour and Tyner, 2014). Second, non-CO_2 emissions are incorporated into the model (Golub, 2013).

Each model's region land endowment is disaggregated into agro-ecological zones (AEZs; figure S1) in the effort to reduce land heterogeneity. In each region of the model, there may be as many as 18 AEZs which differ along two dimensions: growing period (6 categories of 60-day growing period intervals), and climatic zones (3 categories: tropical, temperate and boreal). Even after introduction of AEZs, there is still considerable heterogeneity within these units, and this, in turn, is likely to limit the mobility of land across uses within an AEZ.

To further limit land mobility within each AEZ, in the model land mobility across uses is further restricted by a Constant Elasticity of Transformation (CET) frontier. The elasticity of land transformation parameter is meant to reflect how easy or difficult to transform land from one use to another (e.g. from pasture to cropland) due to: biophysical land heterogeneity within AEZ; region-specific infrastructure, socioeconomic factors, ownership of land; costs of conversion, managerial inertia, unmeasured benefits from crop rotation, etc. The parameter, together with land rents share of a given land use in total AEZ land rents, determines the land supply elasticity to the given land use.

S1.1.1 Land-cover changes

The GTAP TABLO code (gtap.tab) was modified to write out land cover changes (in hectares) for forestry, pastureland, and cropland, as well as breaking out separately changes in cropland-pasture, sugar crops, and oil palm. These data are required by the AEZ-EF model, which computes the total greenhouse gas (GHG) emissions associated with these land-cover changes. The AEZ-EF model is described further in section 1.2, and elsewhere (Plevin et al., 2014).

S1.1.2 Armington elasticities

The default values for the Armington elasticities follow the "rule of 2" which states that the substitution elasticities among domestic and imported goods (ESBD) are half of their corresponding substitution elasticities among imported goods (ESBM) (Keeney and Hertel, 2005). To maintain this relationship in the Monte Carlo simulation, we assign a random variable to ESBM and compute ESBD = 0.5 * ESBM.

S1.1.3 Land net displacement factor

The land net displacement factor (NDF) is the ratio of the total increase in cropland area globally resulting from a biofuel shock to the land area required to produce the increased amount of biofuel, at nominal yields,

[1]The model, data, and and parameters are available at https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4347

[2]See http://www.transportation.anl.gov/pdfs/MC/625.PDF

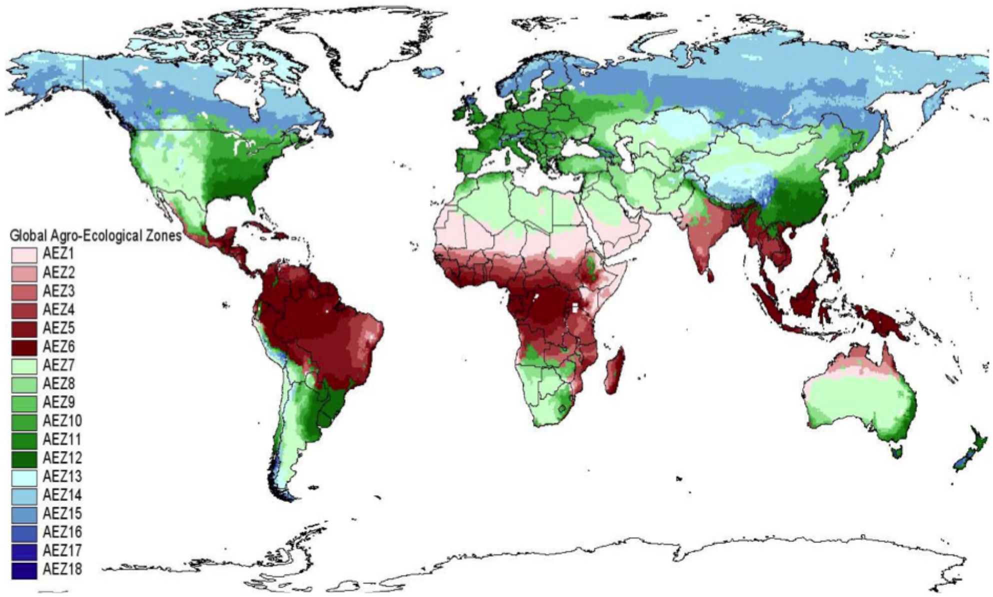


Figure S1: Distribution of agro-ecological zones (AEZs 1-18) and regions used in GTAP (Monfreda et al., 2009)

before accounting for co-products. This metric was first defined in the reduced-form model of ILUC (RFMI) by (Plevin, 2010) and subsequently computed by Laborde and Valin (2012) using the MIRAGE model to estimate LUC for EU biofuel policies.

We modified the GTAP model code (gtap.tab) to compute areal fuel yield (gal/ha). The AEZ-EF model retrieves this result and computes the nominal land requirement (ha) as areal fuel yield divided by shock gallons. The total change in cropland area (ha) is divided by nominal land requirement to produce NDF for each trial. A distribution for NDF is then produced from the MCS results.

S1.2 Agro-ecological zone emission factor (AEZ-EF) model

The agro-ecological zone emission factor (AEZ-EF) model estimates the total CO_2-equivalent emissions from land use changes, e.g., from an analysis of biofuels impacts or policy analyses such as estimating the effect of changes in agricultural productivity on emissions from land use. The model combines matrices of carbon fluxes (Mg CO_2 ha^{-1} y^{-1}) with matrices of changes in land use (ha) according to land-use category as projected by GTAP or similar AEZ-oriented models.

The AEZ-EF model contains separate carbon stock estimates (Mg C ha^{-1}) for biomass and soil carbon, indexed by GTAP region and AEZ (Gibbs et al., 2014). The model combines these carbon stock data with assumptions about carbon loss from soils and biomass, mode of conversion (i.e., whether fire is used), quantity and species of carbonaceous and other GHG emissions resulting from conversion, carbon remaining in harvested wood products and char, and foregone sequestration. The model relies heavily on IPCC greenhouse gas inventory methods and default values (IPCC, 2006), augmented with more detailed and recent data where available. We refer the reader to the model documentation for complete details (Plevin et al., 2014).

S2 Model Parameters

S2.1 GTAP-BIO-ADV behavioral parameters

The GTAP-BIO-ADV behavioral parameters file contains 44 named parameters, 38 of which are appropriate for manipulation in an MCS.[3] The parameters relevant to this study are listed in table S1. Most parameters are vectors or matrices; together they represent 18,319 values that can be manipulated independently. Owing to a paucity of supporting data, however, several of these named parameters consist of one or a few values that are repeated across all regions, AEZs, or sectors. In effect, the model implements these parameters as a single (or few) parameter(s) applied to a range of circumstances. We take advantage of this to simplify the definition of parameter distributions.

The general approach taken here is to allow vectors or matrices of parameters to be manipulated using a single random variable, e.g., by selecting a value from a range and adding this to, or multiplying this by, each element of the matrix. In a few cases, parameters are manipulated by row, column, or specific cell. These approaches can be used in combination: it is possible to define a random factor that is applied to all values except for those singled out for different treatment. The syntax for specifying parameter distributions is described in section S3.3.

We did not include distributions for consumer demand behavioral parameters. These are calibrated to mimic certain income and price elasticities (Hertel et al., 2008), so estimating distributions for them would ideally be done in the calibration stage by varying inputs to that process (target income and price elasticities).

S2.1.1 Additional issues

The rapid increase in corn-based ethanol led to an equally large increase in dried distillers grains with solubles (DDGS), however estimates of the substitution between this feed and coarse grains (corn) were nonexistent before Taheripour et al. (2010). We utilize their estimate as the upper bound and the estimate from Beckman et al. (2011) as the lower bound. The two estimates result from different calibration procedures.

The GTAP-BIO-ADV model allocates land among the three land use categories crops, pasture land, and forest (with cropland-pasture considered part of cropland). The ease of transformation of land from one use to another is modeled by the elasticity of transformation (ETL1). Among crops (including cropland-pasture), the ease of land transformation is determined by a second elasticity of transformation (ETL2). For these parameters, we use of default values and ranges suggested in modeling done for the CARB Low Carbon Fuel Standard (LCFS) (Taheripour and Tyner, 2014). CARB modeling includes regionalized default values for ETL1 and ETL2 found in Taheripour and Tyner (2013), updating an earlier model with a single value for each of these parameters, both applied in every Region-AEZ. Though the parameter is same across regions and AEZs, the actual region and AEZ land supply elasticities in the model are not uniform because they are endogenous to the model and determined not only by the chosen transformation elasticity but also by land rent share of the given land use in total land rents (Golub and Hertel, 2012). Default values were determined by examining recent historical evidence on land use change, categorizing regions based on their patterns of deforestation and expansion of agricultural area or crop switching (based on maize and oilseeds vis--vis other crops), and assigning elasticity values ranked according to land use change category. The approach differs from the prior approach, which derived a global ETL1 from an econometric determination of land supply elasticity in the US controlling for factors such as exogenous time trend, and used expert judgment about crop switching, again grounded in US data, for a global ETL2. Lack of sufficient data prevents broad replication of US-based econometrics in other model regions and for all relevant land use categories. The trend analysis applied to derive these new parameters should be subject to additional analysis to assess how

[3] Other parameters contain metadata, are not referred to in the model code, are structural parameters (e.g., identifying sluggish commodities), or are not relevant to our study (e.g., emission factors).

uncontrolled-for elements inherent in the empirical data trends used affect assigned elasticity values (and model results). While we maintained the regionalized values as in the model for consistency with existing regulatory modeling, we emphasize a need to undertake an exercise along these lines (outside our scope here), and note that applying uncertainty analysis along the lines undertaken here (MCS) can investigate the importance of the choice of the parameter to ILUC emission intensity (within bounds set by the parameter ranges and distributions). Only ETL1 emerged as an important parameter in the primary analysis, and only for sugarcane ethanol, food not fixed, with less than 10% contribution to ILUC emission intensity variance). Lacking additional analysis, we state a modeling preference for using as defaults the prior global values, or implementing strategies employed in other CGE-based ILUC emission studies that have regionalized land transformation elasticities in empirical data, more isolated from trends exogenous to the model.

For example, Golub et al. (2012) constructed a heterogeneity index based on biophysical characteristics and derived regionalized ETL1 values for GTAP based on this index, reasoning that heterogeneity of biophysical characteristics critical to productivity would capture much of the factors that hinder land transformation in a given Region-AEZ. Laborde and Valin (2012) derived model parameters roughly analogous to ETL2 in GTAP-BIO-ADV that is, capturing ease of crop switching – for the MIRAGE model (using a modified GTAP database), calibrated to approximate crop-based land supply elasticities determined for key region/crop combinations in the international agricultural commodity and trade model FAPRI.

GTAP-BIO-ADV model parameter specifications lacking empirical justification (likely because few or no estimates of values were available in the literature) might be important in our MCS and have to be reviewed. In particular, the elasticity of substitution in vegetable oils sub-consumption is specified in our analysis (and the CARB model) as less elastic for developed countries due to an assumption (expert judgment) that consumers base their decisions on nutrition, while in developing countries it is based on price (and likely to be more elastic). Because this parameter has emerged as particularly important in the assessment of biodiesel ILUC emissions from policy especially for the EU (Laborde and Valin, 2012), it is worthy of additional investigation.

Table S1: GTAP model parameters. The first column shows the unique header name used to reference the parameter; the second column shows the parameters dimensions; the third column shows the number of distinct values for each parameter.

Name	Dimensions	Values	Description
CDDG	ALL_INDS*REG	817	Elasticity of substitution in CDDGC and CDDGS feed subproduction
CDGC	ALL_INDS*REG	817	Elasticity of substitution in Oth_CrGr and DDGS feed subproduction
CDGS	ALL_INDS*REG	817	Elasticity of substitution in Sorghum and DDGSS feed subproduction
CRFD	ALL_INDS*REG	817	Elasticity of substitution in crop-based feed subproduction
EAEZ	ALL_INDS	43	Elasticity of substitution in AEZ nest
EFED	ALL_INDS*REG	817	Elasticity of substitution in feed subproduction
ELBO	ALL_INDS*REG	817	Elasticity of subst. in bio-oil subproduction
ELEG	REG	19	Elasticity of substitution in energy consumption
ELEN	ALL_INDS*REG	817	Substitution elasticity in energy sub-production
ELHB	REG	19	Elasticity of substitution in biofuel subconsumption
ELHL	REG	19	Elasticity of substitution in veg. oils subconsumption
ELKE	ALL_INDS*REG	817	Elasticity of substitution in capital-energy subproduction
ELNC	ALL_INDS*REG	817	Substitution elasticity in non-coal energy subproduction
ELNE	ALL_INDS*REG	817	Substitution elasticity in non-electr.energy subproduction
ELVL	ALL_INDS*REG	817	Elasticity of substitution between oils in production
EPSR	ALL_INDS*REG	817	Elasticity of substitution in pasturecrop and pasturecover
ESBD	TRAD_COMM	48	Armington CES for domestic/imported allocation
ESBM	TRAD_COMM	48	Armington CES for regional allocation of imports
ESBT	ALL_INDS	43	Elasticity of intermediate input substitution

Table S1 GTAP model parameters (cont.)

Name	Dimensions	Values	Description
ESBV	ALL_INDS*REG	817	Elasticity of substitution in value-added-en. subproduction
ETA	AEZ18*REG	342	Elasticity of effective hectares with respect to harvested area
ETBD	6	6	Elasticity of transformation among outputs
ETL1	REG	19	Elasticity of transformation among land cover categories
ETL2	REG	19	Elasticity of transformation for crop land in supply tree
ETL3	1	1	Elasticity of transformation for land between beef and milk
ETRE	ENDW_COMM	22	CET between sectors for sluggish primary factors
INCP	CDE_COMM*REG	646	CDE expansion parameter
LVFD	ALL_INDS*REG	817	Elasticity of substitution in livestock-based feed subproduction
OBCD	ALL_INDS*REG	817	Elasticity of substitution between soy-based and corn-based feed
OBDB	ALL_INDS*REG	817	Elasticity of subst. in OBDBS, OBDBO, OBDBO in feed subproduction
OBDO	ALL_INDS*REG	817	Elasticity of subst. in Oth_Oilseed and OBDBO feed subproduction
OBDP	ALL_INDS*REG	817	Elasticity of substitution in palmf and OBDBP feed subproduction
OBDR	ALL_INDS*REG	817	Elasticity of substitution in Rapeseed and OBDBR feed subproduction
OBDS	ALL_INDS*REG	817	Elasticity of substitution in soybeans and OBDBS feed subproduction
PAEL	REG	19	Scalar yield elasticity target for cropland pasture
SUBP	CDE_COMM*REG	646	CDE substitution parameter
YDEL	1	1	Scalar yield elasticity target
YDRS	REG	19	Scale of yield elasticity target relative to base value for given region

Table S2 lists the 19 regions used in GTAP-BIO-ADV.

Table S2: Region definitions used in the GTAP-BIO-ADV model

Region ID	Description
USA	United States
EU27	European Union 27
Brazil	Brazil
Canada	Canada
Japan	Japan
ChiHkg	China and Hong Kong
India	India
C_C_Amer	Central and Caribbean Americas
S_O_Amer	South and Other Americas
E_Asia	East Asia
Mala_Indo	Malaysia and Indonesia
R_SE_Asia	Rest of South East Asia
R_S_Asia	Rest of South Asia
Russia	Russia
Oth_CEE_CIS	East Europe and Rest of Former Soviet Union
Oth_Europe	Rest of European Countries
ME_N_Afr	Middle Eastern and North Africa
S_S_Afr	Sub Saharan Africa
Oceania	Oceania

S2.2 AEZ-EF model parameters

Table S3: AEZ-EF model parameters. The first column shows the unique header name used to reference the parameter; the second column shows the parameters dimensions; the third column shows the number of distinct values for each parameter. In the “dimensions” column, AEZ refers to the 18 agro-ecological zones shown in figure S1; REG refers to the 19 regions used in GTAP-BIO-ADV, described in table S2; LATITUDE refers to the 3 major climate zones: boreal, temperate, and tropical; SPECIES refers to 5 combustion emissions: CO_2, CO, CH_4, N_2O, and non-methane hydrocarbons.

Name	Dimensions	Values	Description
GWP_CO2	scalar	1	CO_2 global warming potential
GWP_CH4	scalar	1	CH_4 global warming potential
GWP_N2O	scalar	1	N_2O global warming potential
MalaIndoPeatEF	scalar	1	Emissions from peatland conversion (Mg C ha^{-1})
MalaIndoPeatFraction	scalar	1	The faction of conversion to oil palm occurring on peatland
N2O_N_EF	scalar	1	Fraction of N in applied fertilizer that is released as N_2O
carbonNitrogenRatio	scalar	1	Represents the mass ratio of carbon to nitrogen loss
cropCarbonAnnualizationFactor	scalar	1	Ratio of annual average to maximum crop carbon
croplandLandUseFactor	AEZ	18	IPCC land use factor for cropland
croplandPastureEmissionRatio	scalar	1	The ratio of emissions from converting cropland-pasture to cropland to those from converting pasture to cropland
croplandSoil_C	AEZ*REG	342	Cropland soil carbon density (Mg C ha^{-1}) to 30 cm depth
croplandSubsoil_C	AEZ*REG	342	Cropland soil carbon density (Mg C ha^{-1}) from 30 to 100 cm depth
deadwoodByLatitude_C	LATITUDE	3	Deadwood carbon density (Mg C ha^{-1}) by latitude
deadwoodByRegion_C	REG	19	Deadwood carbon density (Mg C ha^{-1}) by region
deforestedFraction	REG	19	Fraction of forest cover change that is deforestation rather than afforestation
excludedLitterFraction	scalar	1	Litter fraction not included in regrowth
fireClearingFraction	REG	19	Fraction of land-cover cleared using fire
foregoneGrowthRate	AEZ*REG	342	Foregone sequestration rate (Mg C ha^{-1} y^{-1})
forestBurningEF	LATITUDE*SPECIES	15	Emissions of 5 species for forest burning (kg per Mg dry matter)
forestCombustionFactor	LATITUDE	3	Fraction of fuel biomass combusted when clearing forests with fire
forestDefaultRootShootRatio	scalar	1	Default ratio of live root biomass to above-ground live biomass for forests
forestLandUseFactor	AEZ	18	IPCC land use factor for forest land
forestLitter_C	AEZ	18	Forest litter carbon density (Mg C ha^{-1})
forestRegrowthRate	AEZ*LATITUDE	54	Forest regrowth rate (Mg C ha^{-1} y^{-1})
forestRootShootRatio	AEZ*REG	342	Ratio of live root biomass to above-ground live biomass for forests
forestSoilLossFraction	LATITUDE	3	Fraction of forest soil (to 30 cm) lost during conversion to cropland

Table S3 AEZ-EF model parameters (cont.)

Name	Dimensions	Values	Description
forestSoil_C	AEZ*REG	342	Forest soil carbon density (Mg C ha^{-1}) to 30 cm depth
forestSubsoilLossFraction	LATITUDE	3	Fraction of forest soil (30 to 100 cm) lost during conversion to cropland
forestSubsoil_C	AEZ*REG	342	Forest soil carbon density (Mg C ha^{-1}) from 30 to 100 cm depth
grassCarbonFraction	scalar	1	Fraction of herbaceous biomass composed of carbon
hwpFraction	REG	19	Fraction of above-ground biomass removed in harvested wood products
oilPalmBiomass_C	scalar	1	Carbon density (Mg C ha^{-1}) of oil palm trees
pastureAgb	AEZ	18	Carbon density (Mg C ha^{-1}) of above-ground pasture biomass
pastureBgb	AEZ	18	Carbon density (Mg C ha^{-1}) of below-ground pasture biomass
pastureBurningEF	LATITUDE*SPECIES	15	Emissions of 5 species for pasture burning (kg per Mg dry matter)
pastureCombustionFactor	LATITUDE	3	Fraction of fuel biomass combusted when clearing pastures with fire
pastureLitter_C	scalar	1	Pasture litter carbon density (Mg C ha^{-1})
pastureSoil_C	AEZ*REG	342	Pasture soil carbon density (Mg C ha^{-1}) to 30 cm depth
pastureSoilLossFraction	LATITUDE	3	Fraction of forest soil (to 30 cm) lost during conversion to cropland
pastureSubsoilLossFraction	LATITUDE	3	Fraction of forest soil (30 to 100 cm) lost during conversion to cropland
pastureSubsoil_C	AEZ*REG	342	Pasture soil carbon density (Mg C ha^{-1}) from 30 to 100 cm depth
regrowth_C	AEZ*REG	342	Estimated C stored in afforestation over 30 years (Mg C ha^{-1})
totalTree_C	AEZ*REG	342	Carbon density (Mg C ha^{-1}) of total tree (above- plus below-ground)
tropicalForestRootShootRatio	scalar	1	Ratio of live root biomass to above-ground live biomass for tropical forests
understory_C	LATITUDE	3	Understory soil carbon density (Mg C ha^{-1}) to 30 cm depth
woodyCarbonFraction	scalar	1	Fraction of woody biomass composed of carbon

S3 Monte Carlo framework

One challenge to performing Monte Carlo simulation with models such as GTAP is the length of time the models require to complete a single simulation. The GTAP model used in this analysis typically required 5-15 minutes per solution on a desktop computer. Thus, a Monte Carlo analysis using 1000 trials could require over a week of continuous computation. To address this, our analysis was performed on a high-performance parallel computing cluster at the National Energy Research Scientific Computing (www.nersc.gov) facility, managed by the US Department of Energys Lawrence Berkeley National Lab. The ILUC-MCS model is based on a new software framework called *Distributed MCS*, or DMCS, developed in the Python language. DMCS will be made freely available as under an open-source license. Please contact the authors for more information.

S3.1 What is represented by parameter distributions

The Monte Carlo approach produces a joint output frequency distribution by executing the model numerous times with alternative parameter values drawn from defined input distributions. Although this correctly represents the joint probability, the semantics of this distribution depends critically on what exactly is represented by the input parameter distributions.

In many cases, there are inadequate data to draw parameter distributions, or in many cases, to distinguish values by region or industry. We treat values that are constant across regions or industries in the GTAP database as single parameters rather than individual parameters.

There is disagreement among experts as to the best value for many parameters. Elliott et al. (2011) compared values from GTAP and MITs EPPA model and fail to distinguish any pattern to the disagreement, with GTAP values sometimes the highest and sometimes the lowest. So one setting for sensitivity analysis is to examine the sensitivity to different expert opinion, which can vary quite widely for any single parameter. In our analysis, we have examples of DDGS substitution elasticities of Taheripour versus those of Beckman, and disagreement over the most appropriate value for YDEL.

Another setting examines the sensitivity to the uncertainty around a value that is treated as reasonably well-characterized, i.e., there is strong data or expert agreement supporting the approximate value, but there is still measurement or approximation uncertainty. Yet another setting—the one we focus on here—is to understand the sensitivity of the model and the range of plausible output values that result from all of the above, regardless of the source.

Its important to note that our output distributions indicate the uncertainty in the final result based on the described uncertainty in parameters—treating the model structure and underlying base data as certain. Thus our results should not be treated as characterizing probabilities of any real-world outcome; rather, they represent the distribution of results for this model, as implemented, given our choice of distributions.

Tables S5 and S6 list the parameter distributions used in the Monte Carlo simulation. Table S4 shows the values from the IPCC's Guidelines for National Greenhouse Gas Inventories (IPCC, 2006) which were used to define distributions for the AEZ-EF model parameter "croplandUseFactor".

S3.2 Correlations

Frey et al. (2006, p 3.25) note that dependencies among inputs matter only if the parameters are important contributors to variance and the dependency (correlation or covariance) is strong. Otherwise, modeling this dependence is unimportant to the resulting uncertainty.

The modeling framework used in this study supports the implementation of rank correlations among random variables based on the method of Iman and Davenport (1982). We used this to impose rank correlations in only one situation: we assigned a rank correlation coefficient of 0.9 to the elasticities of substitution between (i) other coarse grains (predominantly corn) and distillers dried grains with solubles (DDGS) and (ii) sorghum and DDGS given the similarity of these products.

S3.3 Distribution definition file format

Many of the parameters to these models are matrices of values with a common purpose, e.g., the carbon density of soil in each region represented in the model, or the elasticities of substitution among a set of industrial sectors, by region. These parameter groups can be manipulated stochastically in various ways: using a single random variable (RV) assigned to the entire group, or by RVs for each row, column, cell in a matrix. One goal of the analysis is to identify important parameter groups affected by one or more RV, as well as individual RVs.

- Values drawn from the distributions can be used either to substitute directly for the default value for that parameter, or as a factor multiplied by the default value to produce a value to which is then substituted for the default value.
- Distributions can be applied to scalar values, entire matrices, or individual rows, columns, or cells of a matrix.
- The distributions currently supported include: Uniform, Normal, Lognormal, Triangle, Binary, and Discrete. The system is design to allow additional distributions to be added fairly easily.
- Scalar parameters are equivalent to $1 * 1$ matrices, so whenever matrix parameters are mentioned, this includes scalar and vector $(1 * N)$ parameters.
- It is possible to declare (rank) correlations between pairs of model parameters, as long as they have the identical dimensions, and the distribution must be assigned to the same dimensions for both parameters (see more on this below). In the matrix case, the cells at the same location within the matrix are treated as correlated.
- Also possible to declare that random variables associated with a matrix are correlated, but use with caution: this can generate hundreds of random variables. Most practical to use with small distribution dimensions.

S3.3.1 Defining distributions

Here we describe the assignment of distributions to model parameters using the Distributed-MCS framework. Note that this framework will be released as an open-source project; please contact the authors for more information.

- Blank lines are ignored
- Text after # are treated as comments and ignored
- Two types of entries are processed: distribution declarations and correlation declarations
- A single distribution declaration can produce multiple random variables (RV).

The general format for a distribution declaration is:

```
parameter target distro arg1=value arg2=value ...
```

`Parameter` Specifies the model parameter to which this distribution applies.

`Target` One of: None, Single, Rows, Cols, Cells, [*row*] or [*row, col*]

`None` The parameter is treated as constant using the value given in the parameter file.

`Single` A single random variable (RV) is created, the value of which is applied to all non-zero elements of the matrix (unless the modifier `_updateZero=1` is specified, in which case all matrix elements are updated). Note that "applying" a value can mean either assigning the value directly, or using it as a multiplier, in which case the parameter value used in the trial is the product of the RV value and the default parameter value.

`Rows` Similar to "single" except that an RV is created for each row of the matrix.

`Cols` Similar to "single" except that an RV is created for each column of the matrix.

`Cells` Similar to "single" except that an RV is created for each individual cell of the matrix.

[***row***] The *row* value specifies a row index by name or numeric value. A separate RV is generated for this row. If declared after a matrix that includes this row, the subsequent definition overrides the prior one.

[***row, col***] The *row* and *col* values specify the row and column indices by name (e.g., regions, industries) or numeric value. or are an asterisk (*) to indicate an entire row or column. For example, to specify column 3 only, you would write [*,3], meaning "all rows of column 3".

`Distro` One of: UniformFactor, LogFactor, TriangleFactor, Normal, Lognormal, Uniform, Triangle, Binary.

`UniformFactor` Args: min=*arg1*, max=*arg2* or factor=*arg1*
Adds random noise factor of [*arg1, arg2*] by multiplying all values indicated by the noise factor selected from this range. If only one value is given, it must be a fraction between 0 and 1, which defines a range of ±*arg1*, i.e., the range is [1-*arg1*, 1+*arg1*]

`LogFactor` Args: factor=*arg1*
Similar to `UniformFactor` but multipliers are chosen from lognormal with a 95% confidence interval of [1/*arg1*, *arg1*]. For example, a value of 3 means that multiplier values are selected from a lognormal distribution with 95% CI = [1/3, 3].

`TriangleFactor` Args: width=*arg1* or min=*arg1*, mode=*arg2*, max=*arg3*
Similar to `UniformFactor` but multipliers are chosen from a triangular distribution with the given *min*, *mode*, and *max*, or centered on zero with *min* and *max* set to ±*width*.

`Normal` Args: mean=*arg1*, std=*arg2*
Set values to a random choice from normal distribution with mean of *arg1* and standard deviation of *arg2*.

`Lognormal` Args: mean=*arg1*, std=*arg2* *or* low95=*arg1*, high95=*arg2*
Set values to a random choice from a lognormal distribution, either with (i) mean value (of the lognormal, not the underlying normal distribution) of *arg1* and standard deviation (of the lognormal) of *arg2*, or (ii) 95% confidence interval of [*arg1, arg2*]

`Uniform` Args: min=arg1, max=arg2
Set values to a random choice from interval [arg1, arg2]

`Triangle` Args: min=*arg1*, mode=*arg2*, max=*arg3*
Set values from triangular distribution with minimum value *arg1*, mode *arg2*, and maximum value of *arg3*.

`Binary` Args: none.
Choose randomly from the set {0, 1}

```
parameter[row,col] value1:prob1 val2=prob2 ...
```

S3.3.2 Modifiers

Some distributions accept modifiers, which are like arguments but the names must being with an underscore).

- Random values are directly assigned by default. Alternative specifications are `_apply=mult` and `_apply=add`, which cause the random value to be multiplied by, or added to (respectively) the default value for the parameter.
- In conjunction with `_apply=mult`, distributions can specify `_lowBound=`*arg1* and/or `_highBound=`*arg2*, in which case after multiplying the default value by the value drawn from the random variable, the new value is set to *arg1* if `_lowBound` is specified and the value is less than *arg1*, and the value is set to *arg2* if `_highBound` is specified and the value is greater than *arg2*. This is useful when dealing with parameters representing values that must be between 0 and 1.
- Any distribution can also specify `_updateZero=1` to indicate that zero default values should be updated; otherwise zero values are left unchanged. That is, by default a value of zero will not be replaced by the random value, nor (in the case of `_apply=add`) have the random value added to it.
- Discrete distributions can specify `_tolerance=`*arg1* and/or `_precision=`*arg1*. The `_tolerance` modifier set the amount by which the sum of the probabilities can differ from 1. The default tolerance is 0.01. The `_precision` modifier sets the number of bins into which the discrete values are sorted, thus the resulting probability values will be accurate within 1/precision. The default precision is 100.

S3.3.3 Correlations

Correlations within a single matrix parameter, or between two parameters can be specified as:

```
Correlation Param1 [Param2] value
```

If only 1 parameter is named, this defines a correlation among the RVs for the vector or matrix defined for the named parameter. If 2 parameters are named, both must matrices with identical dimensions, in which case each cell of the first matrix is correlated with the cell at the same `address` in the second matrix.

S3.4 Parameter distributions used in simulations

For the Cropland Land Use Factor, the IPCC suggests the following uncertainty values shown in table S4.

Table S4: IPCC uncertainty ($\pm 2\sigma$) ranges for cropland land-use factors.

Regime	Factor	Error (95% CI)
Dry temp/boreal	0.80	±9%
Moist temp/boreal	0.69	±12%
Dry tropical	0.58	±61%
Moist tropical	0.48	±46%
Tropical montane	0.64	±50%

Table S5: Parameter distributions for the AEZ-EF model.

Parameter name	Target	Distribution	Parameters
croplandLandUseFactor	[AEZ-1]	UniformFactor	factor=0.61
croplandLandUseFactor	[AEZ-2]	UniformFactor	factor=0.61

Table S5 Continued:

Parameter name	Target	Distribution	Parameters
croplandLandUseFactor	[AEZ-3]	UniformFactor	factor=0.61
croplandLandUseFactor	[AEZ-4]	UniformFactor	factor=0.46
croplandLandUseFactor	[AEZ-5]	UniformFactor	factor=0.46
croplandLandUseFactor	[AEZ-6]	UniformFactor	factor=0.46
croplandLandUseFactor	[AEZ-7]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-8]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-9]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-10]	UniformFactor	factor=0.12
croplandLandUseFactor	[AEZ-11]	UniformFactor	factor=0.12
croplandLandUseFactor	[AEZ-12]	UniformFactor	factor=0.12
croplandLandUseFactor	[AEZ-13]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-14]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-15]	UniformFactor	factor=0.09
croplandLandUseFactor	[AEZ-16]	UniformFactor	factor=0.12
croplandLandUseFactor	[AEZ-17]	UniformFactor	factor=0.12
croplandLandUseFactor	[AEZ-18]	UniformFactor	factor=0.12
grassCarbonFraction	Single	UniformFactor	factor=0.05
woodyCarbonFraction	Single	UniformFactor	factor=0.05
oilPalmBiomassC	Single	Normal	mean=35 std=5.5
N2O-N-EF	Single	Lognormal	low95=0.004 high95=0.04
carbonNitrogenRatio	Single	Lognormal	mean=15 std=5.8
cropCarbonAnnualizationFactor	Single	Triangle	min=0.45 mode=0.5 max=0.55
croplandPastureEmissionRatio	Single	Triangle	min=0.0 mode=0.5 max=1.0
croplandSoil-C	Single	UniformFactor	factor=0.50
croplandSubsoil-C	Single	UniformFactor	factor=0.50
deadwoodByLatitude-C	Single	UniformFactor	factor=0.75
deadwoodByRegion-C	Single	UniformFactor	factor=0.75
deforestedFraction	Single	UniformFactor	factor=0.50 highBound=1
deforestedFraction	[Mala-Indo]	UniformFactor	min=0.55 max=1.00
excludedLitterFraction	Single	UniformFactor	factor=0.25 highBound=1
ipccCroplandLandUseFactor	Single	UniformFactor	factor=0.25
ipccForestLandUseFactor	Single	UniformFactor	factor=0.25
fireClearingFraction	Single	UniformFactor	factor=0.50 highBound=1
foregoneGrowthRate	Single	UniformFactor	factor=0.50
forestBurningEF	Single	UniformFactor	factor=0.25
forestCombustionFactor	Single	UniformFactor	factor=0.50 highBound=1
forestDefaultRootShootRatio	Single	Triangle	min=0.20 mode=0.25 max=0.30
forestLandUseFactor	Single	UniformFactor	factor=0.25
forestLitter-C	Single	UniformFactor	factor=0.50
forestRootShootRatio	Single	UniformFactor	factor=0.23
forestSoilLossFraction	Single	UniformFactor	factor=0.25 highBound=1
forestSubsoilLossFraction	Single	UniformFactor	factor=0.50 highBound=1
forestSoil-C	Single	UniformFactor	factor=0.50
forestSubsoil-C	Single	UniformFactor	factor=0.50
GWP-CH4	Single	Normal	mean=25 std=4.35
GWP-N2O	Single	Normal	mean=298 std=52.15
hwpFraction	Single	UniformFactor	factor=0.25
MalaIndoPeatEF	Single	UniformFactor	factor=0.25
MalaIndoPeatFraction	Single	UniformFactor	factor=0.25 highBound=1
pastureAgb	Single	UniformFactor	factor=0.80

Table S5 Continued:

Parameter name	Target	Distribution	Parameters
pastureLitter-C	Single	Triangle	min=0.05 mode=0.40 max=0.50
pastureBurningEF	Single	UniformFactor	factor=0.25
pastureCombustionFactor	Single	UniformFactor	factor=0.75 highBound=1
pastureSubsoilLossFraction	Single	UniformFactor	factor=0.25 highBound=1
pastureSoil-C	Single	UniformFactor	factor=0.25
pastureSubsoil-C	Single	UniformFactor	factor=0.50
totalTree-C	Single	UniformFactor	factor=0.25
totalTree-C	[*,Canada]	UniformFactor	factor=0.80
totalTree-C	[*,ME-N-Afr]	UniformFactor	factor=0.80
totalTree-C	[*,EU27]	UniformFactor	factor=0.80
totalTree-C	[*,ChiHkg]	UniformFactor	factor=0.80
tropicalForestRootShootRatio	Single	UniformFactor	factor=0.25
regrowth-C	Single	UniformFactor	factor=0.50
Correlation	foregoneGrowthRate regrowth-C		0.75

Table S6: Parameter distributions for the GTAP model.

Parameter name	Target	Distribution	Parameters
CDDG	Single	Uniform	min=10 max=20
CDGC	Single	Uniform	min=10 max=30
CDGS	Single	Uniform	min=10 max=30
CRFD	Single	LogFactor	factor=1.5
EFED	Single	Triangle	min=0.15 mode=0.50 max=0.85
ELEG	Rows	UniformFactor	factor=0.5
ELEN	Rows	LogFactor	factor=2
ELHB	Single	UniformFactor	factor=0.50
ELHL	Single	LogFactor	factor=2
ELKE	Rows	LogFactor	factor=1.5
ELNC	Rows	LogFactor	factor=1.5
ELNE	Rows	LogFactor	factor=1.5
ELVL	Single	LogFactor	factor=1.5
EPSR	Single	UniformFactor	factor=0.5
ESBM	Single	LogFactor	factor=2
ESBV	Rows	LogFactor	factor=1.5
ETA	Single	UniformFactor	factor=0.20 highBound=1.0
ETL1	Single	TriangleFactor	width=0.2
ETL2	Single	TriangleFactor	width=0.2
LVFD	Single	LogFactor	factor=1.5
OBCD	Single	Uniform	min=0.14 max=0.3
OBDO	Single	Uniform	min=10 max=20
OBDP	Single	Uniform	min=10 max=20
OBDR	Single	Uniform	min=10 max=20
OBDS	Single	Uniform	min=10 max=20
PAEL	[USA]	Uniform	min=0.1 max=0.6
PAEL	[Brazil]	Uniform	min=0.1 max=0.3
YDEL	Single	Uniform	min=0.03 max=0.25

S4 Model Results

Figure S2 shows the frequency distributions for 3 model outputs: ILUC Emission Factor, Non-CO_2 Emission Factor, and Total Emission Factor, which is the sum of the prior two quantities on a trial-by-trial basis. For each model output, 2 distributions are shown for each of the three fuel pathways examined. The items labeled "FF" (food fixed) were simulated with food consumption fixed in non-Annex-I countries; those labeled "FNF" (food not fixed) were run without this constraint. Constraining food consumption removes a degree of freedom for the model, causing other modeled behavior (e.g., extensification) to take up the slack, and resulting in ILUC emissions that were consistently about 10 g CO_2 MJ^{-1} higher than without the constraint.

S4.1 Non-CO_2 emissions

The GTAP non-CO_2 version 7 database (Rose et al., 2010) includes nitrous oxide (N_2O), methane (CH_4) and fourteen fluorinated gases (F-gases). In each region, non-CO_2 emissions are provided for each economic sector and driver, and regional household. To track changes in non-CO_2 emissions within the GTAP-BIO-ADV model, emissions are tied to specific drivers within each sector: factor inputs, intermediate inputs, or output. For example, emissions from fertilizer application in crop production are proportional to fertilizer use in crops. In livestock sectors, emissions from enteric fermentation and manure management are proportional to livestock capital. Household non-CO_2 emissions are tied to energy use.

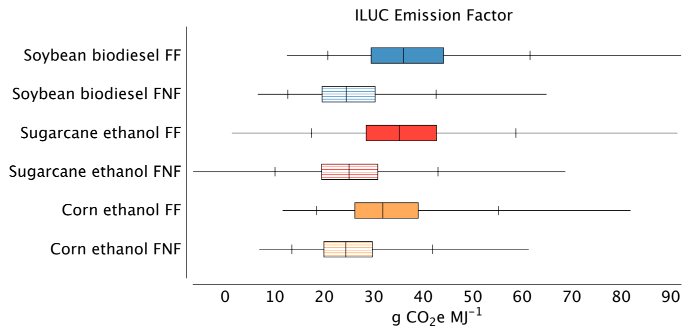
Tables S7 and S8 show the mean and bounds of the central 95% of the distribution for simulations with the three fuels, with food consumption fixed in non-Annex I countries (FF) and not fixed (FNF), both excluding (table S7) and including (table S8) NonCO_2 emissions. Note that for cane ethanol, including non-CO_2 emissions *reduces* the total emissions.

Table S7: Summary of results for ILUC emissions (g CO_2e MJ^{-1}), not including changes in emissions of methane (CH_4) or nitrous oxide (N_2O). FNF=food consumption not fixed anywhere; FF=food consumption is fixed in non-Annex I countries.

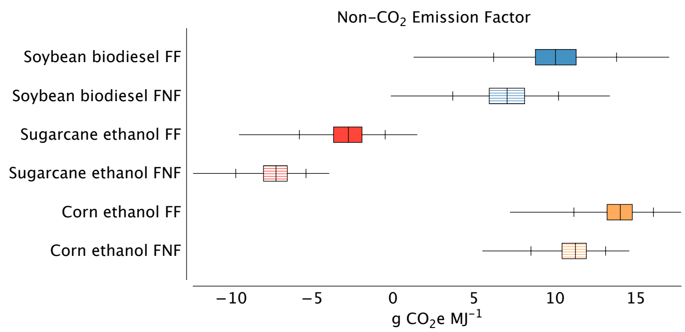
Experiment	Mean	2.5 percentile	97.5 percentile
Corn ethanol, FNF	25	13	42
Corn ethanol, FF	33	18	55
Cane ethanol, FNF	25	10	43
Cane ethanol, FF	36	17	59
Soybean biodiesel, FNF	25	13	43
Soybean biodiesel, FF	38	21	62

Table S8: As described in Table S7, except *including* changes in emissions of methane (CH_4) and nitrous oxide (N_2O).

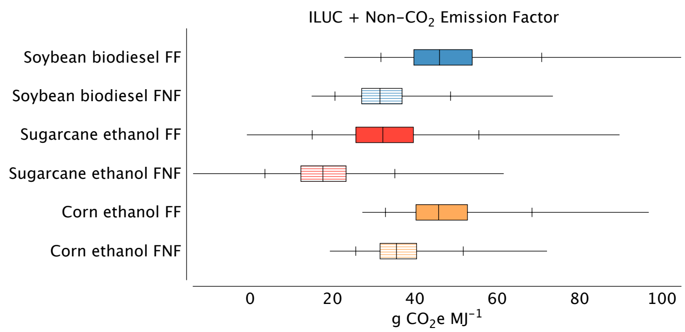
Experiment	Mean	2.5 percentile	97.5 percentile
Corn ethanol, FNF	36	26	52
Corn ethanol, FF	46	33	68
Cane ethanol, FNF	18	4	35
Cane ethanol, FF	33	15	56
Soybean biodiesel, FNF	32	21	49
Soybean biodiesel, FF	48	32	71



(a) ILUC emission factor



(b) Non-CO_2 emission factor, i.e., N_2O emissions from changes in fertilizer and manure application and in CH_4 emissions from changes in rice cultivation and livestock production

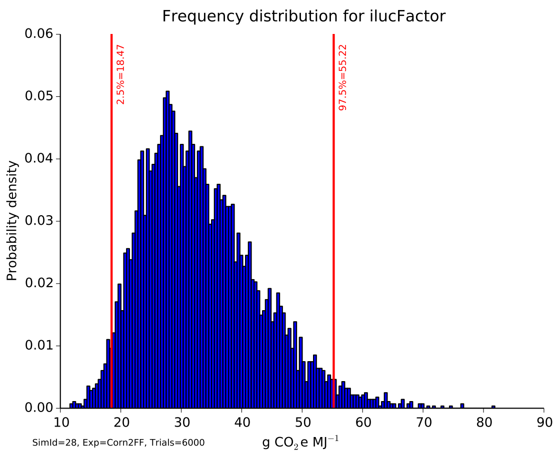


(c) Total emission factor (ILUC emissions + Non-CO_2 emissions)

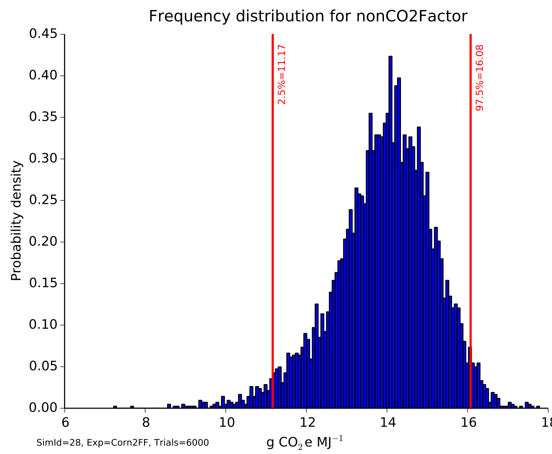
Figure S2: Comparison of ILUC, nonCO_2 and total emission factors for three fuel systems, both with food fixed (FF) and food not fixed (FNF).

Figures S3 through S5 show the frequency distributions for corn ethanol, sugarcane ethanol, and soybean biodiesel, for 3 model outputs: ILUC emissions, non-CO_2 emissions, and total emissions, which is simply the sum of the first two. In these model runs, food consumption has been held fixed in non-Annex I countries.

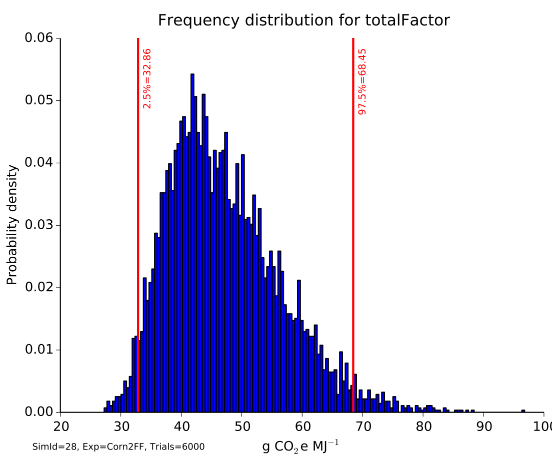
Figures S6 through S8 show the same results but for model runs in which food consumption was not held fixed anywhere.



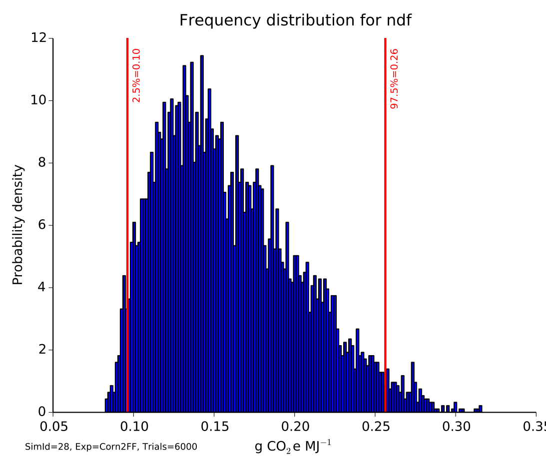
(a) Corn ethanol ILUC factor (food fixed)



(b) Corn ethanol Non-CO_2 Factor (food fixed)

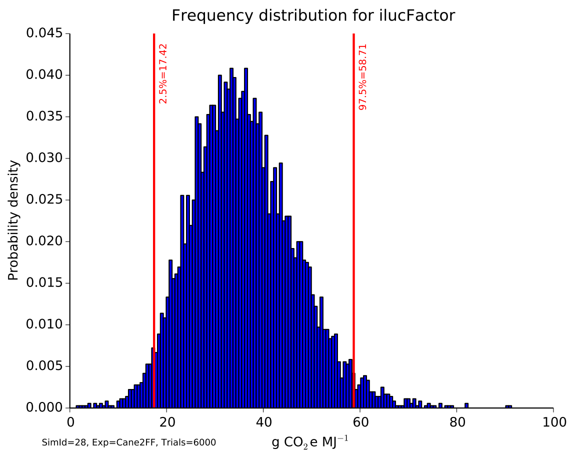


(c) Corn ethanol Total (ILUC+Non-CO_2) Factor, (food fixed)

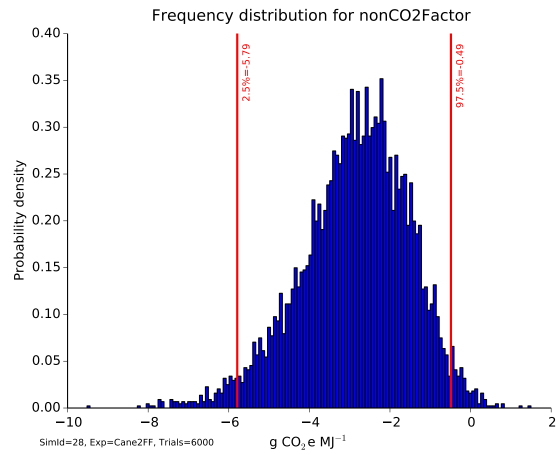


(d) Corn ethanol land Net Displacement Factor (food fixed)

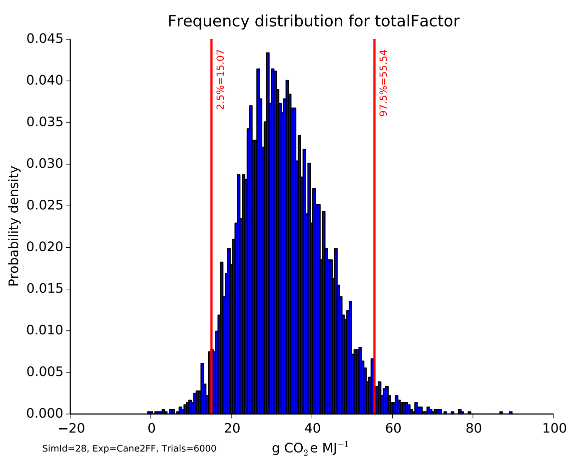
Figure S3: Key model output distributions for corn ethanol, holding food consumption fixed in developing countries.



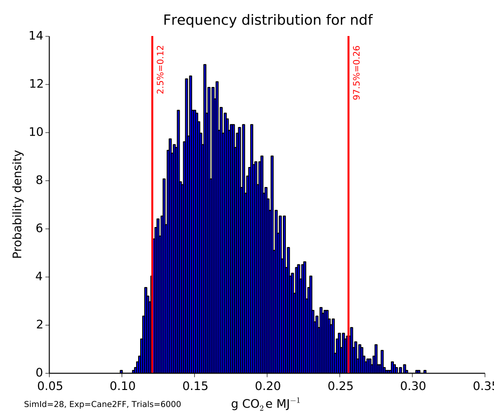
(a) Sugarcane ethanol ILUC factor (food fixed)



(b) Sugarcane ethanol Non-CO_2 Factor (food fixed)

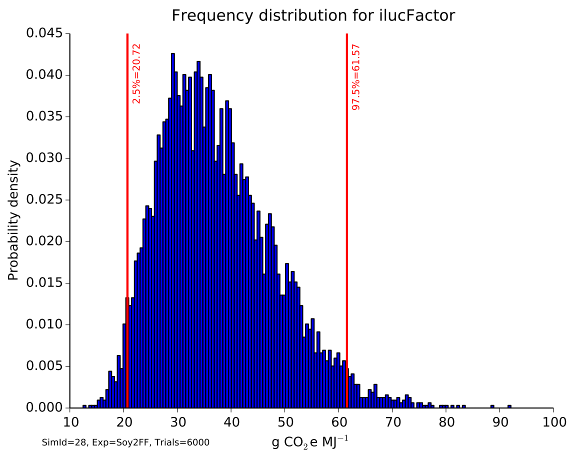


(c) Sugarcane ethanol Total (ILUC+Non-CO_2) Factor, (food fixed)

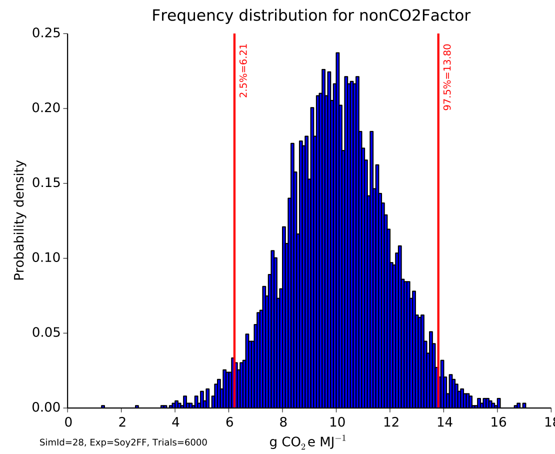


(d) Sugarcane ethanol land Net Displacement Factor (food fixed)

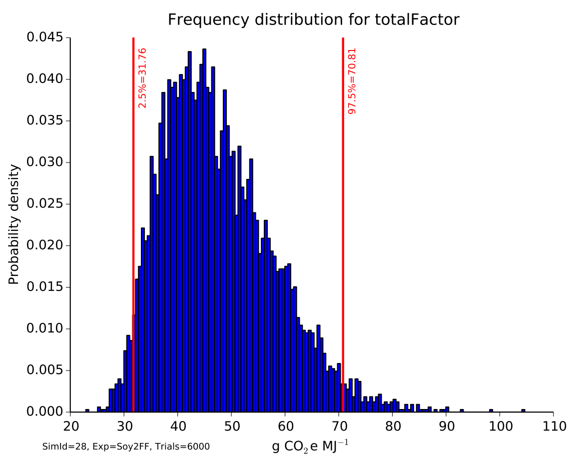
Figure S4: Key model output distributions for sugarcane ethanol, holding food consumption fixed in developing countries.



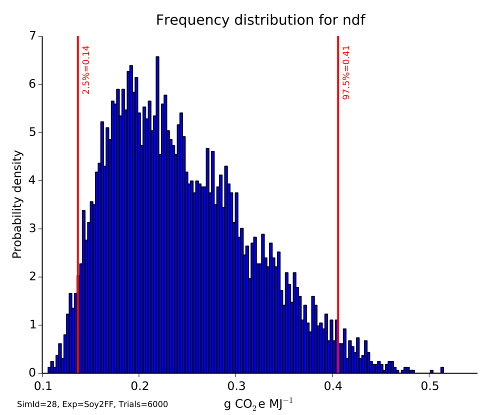
(a) Soybean biodiesel ILUC factor (food fixed)



(b) Soybean biodiesel Non-CO_2 Factor (food fixed)

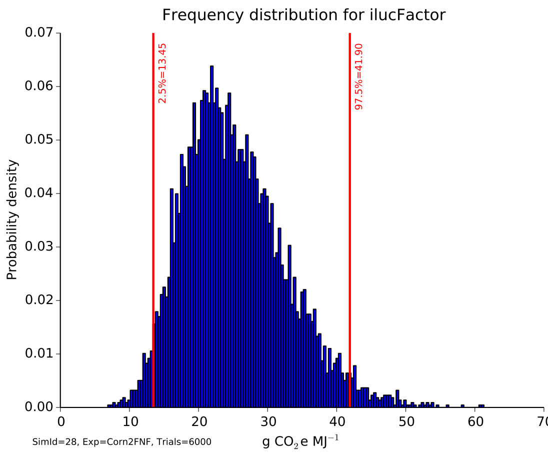


(c) Soybean biodiesel Total (ILUC+Non-CO_2) Factor, (food fixed)

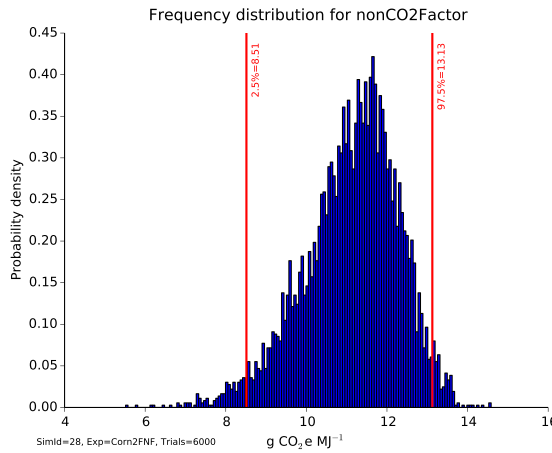


(d) Soybean biodiesel land Net Displacement Factor (food fixed)

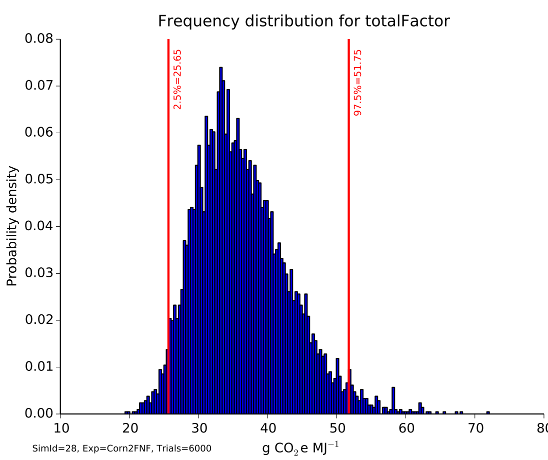
Figure S5: Key model output distributions for soybean biodiesel, holding food consumption fixed in developing countries.



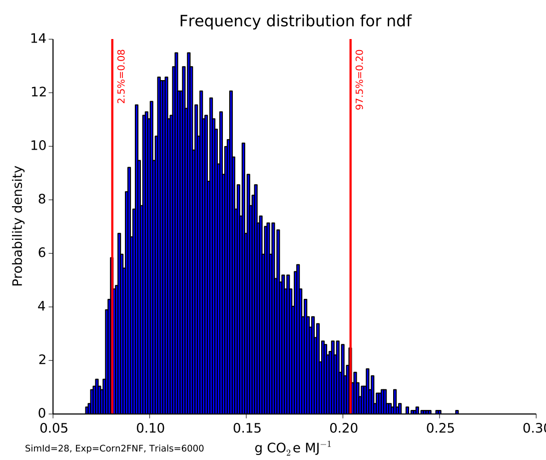
(a) Corn ethanol ILUC factor (food not fixed)



(b) Corn ethanol Non-CO_2 Factor (food not fixed)

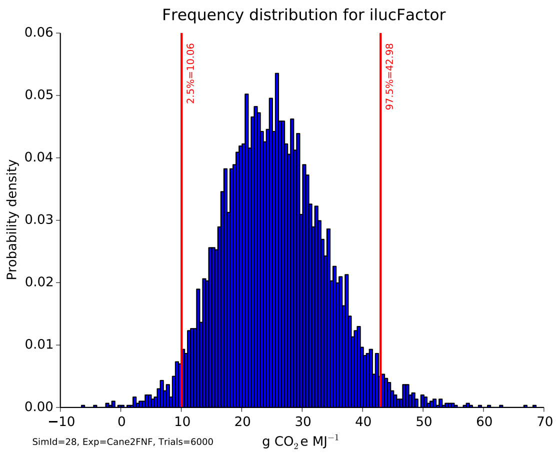


(c) Corn ethanol Total (ILUC+Non-CO_2) Factor, (food not fixed)

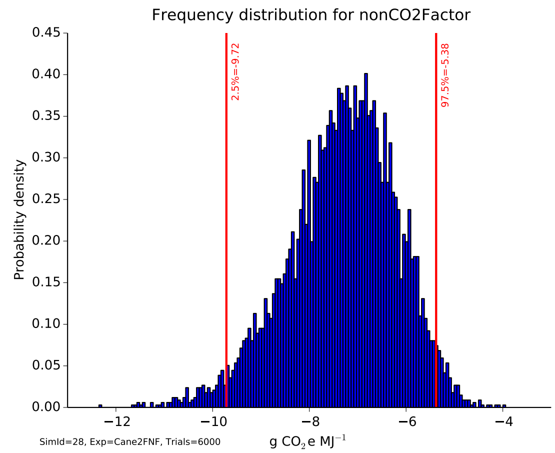


(d) Corn ethanol land Net Displacement Factor (food not fixed)

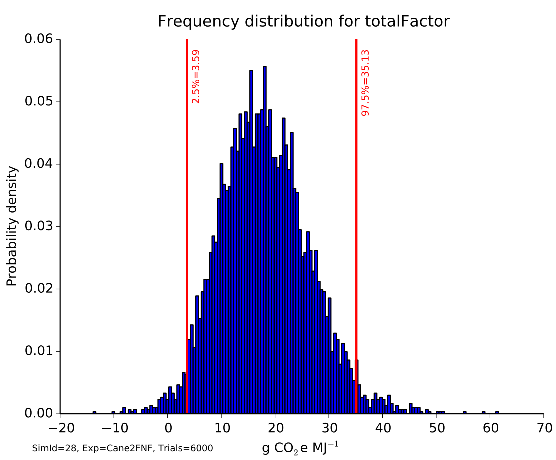
Figure S6: Key model output distributions for corn ethanol, without holding food consumption fixed.



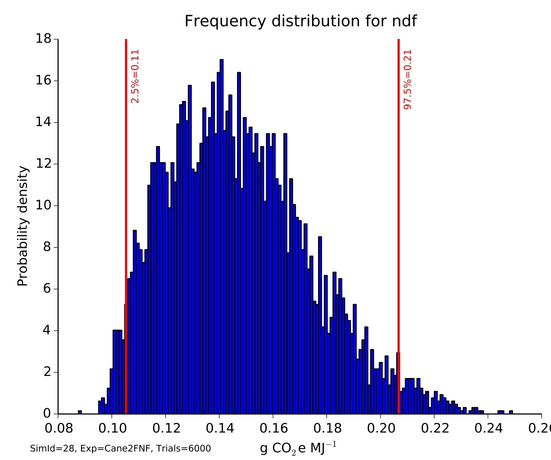
(a) Sugarcane ethanol ILUC factor (food not fixed)



(b) Sugarcane ethanol Non-CO_2 Factor (food not fixed)

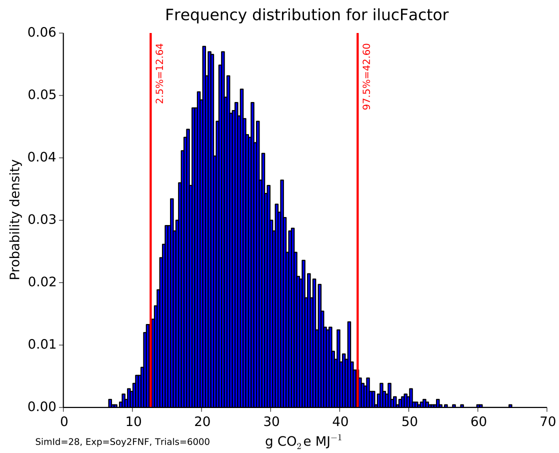


(c) Sugarcane ethanol Total (ILUC+Non-CO_2) Factor, (food not fixed)

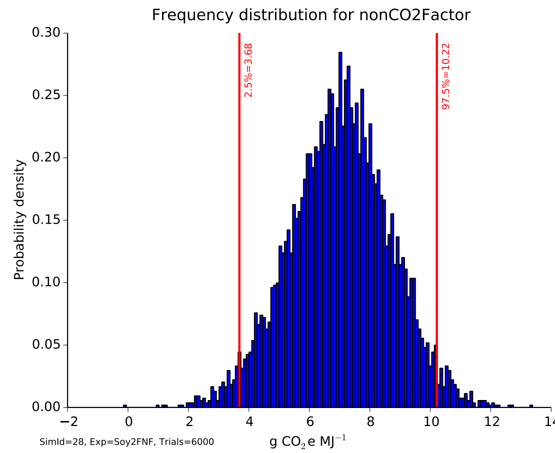


(d) Sugarcane ethanol land Net Displacement Factor (food not fixed)

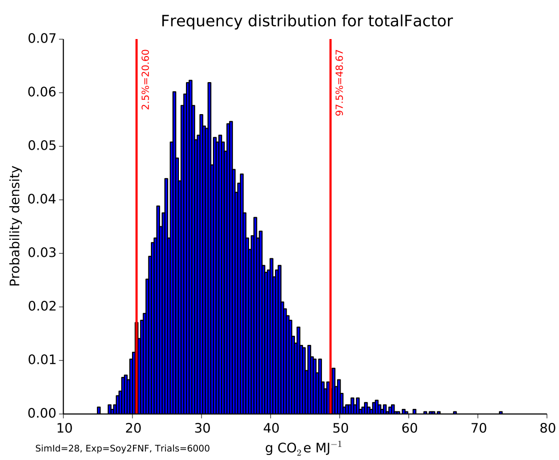
Figure S7: Key model output distributions for sugarcane ethanol, without holding food consumption fixed.



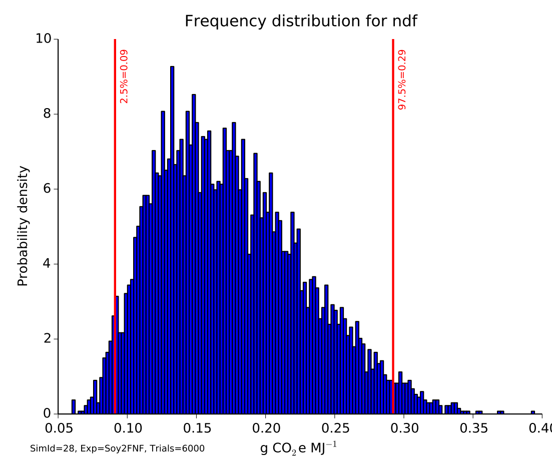
(a) Soybean biodiesel ILUC factor (food not fixed)



(b) Soybean biodiesel Non-CO_2 Factor (food not fixed)



(c) Soybean biodiesel Total (ILUC+Non-CO_2) Factor, (food not fixed)

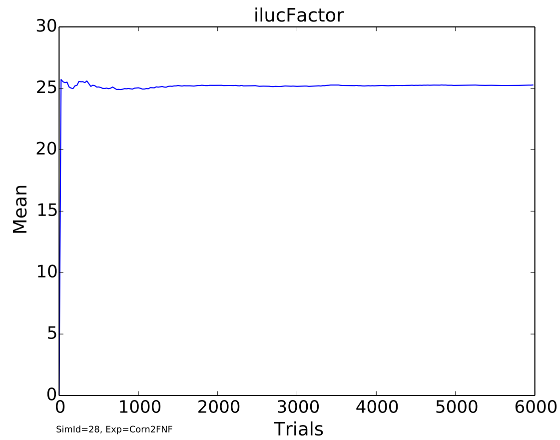


(d) Soybean biodiesel land Net Displacement Factor (food not fixed)

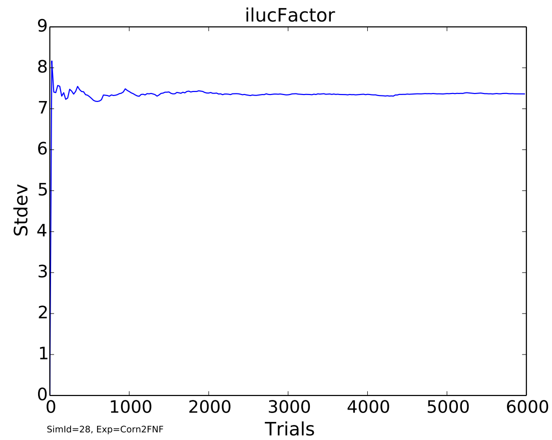
Figure S8: Key model output distributions for soybean biodiesel, without holding food consumption fixed.

S4.2 Statistical Convergence

Figures S9a and S9b show that the mean value for ILUC emissions intensity for corn ethanol converges within about 500 trials, and standard deviation by about 1,500 trials. The corresponding plots for other output variables and biofuel pathways are quite similar.

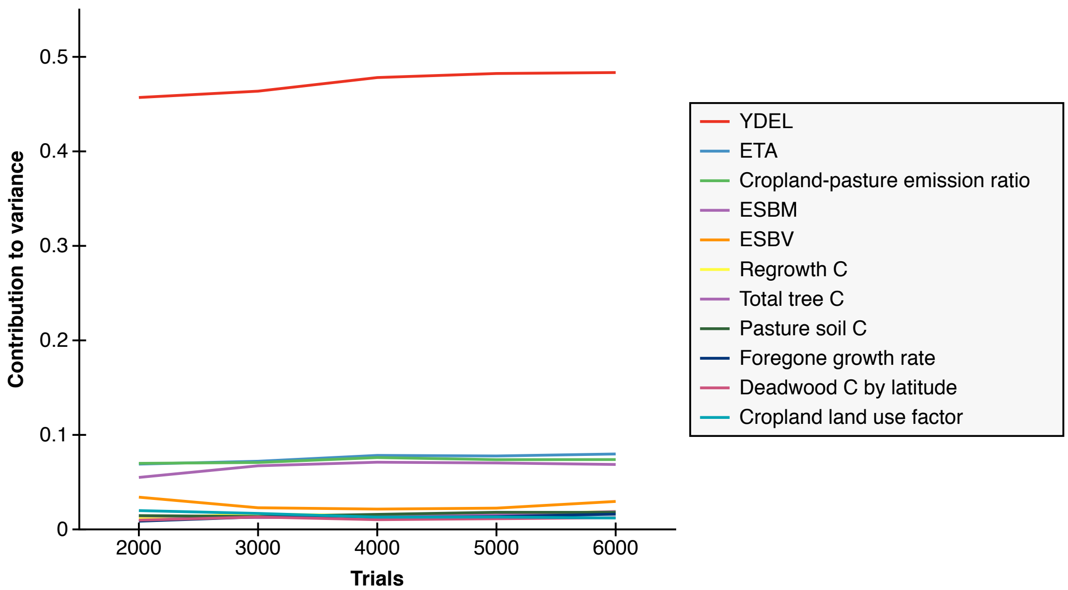


(a) Convergence of the mean value for ILUC emissions, corn ethanol.

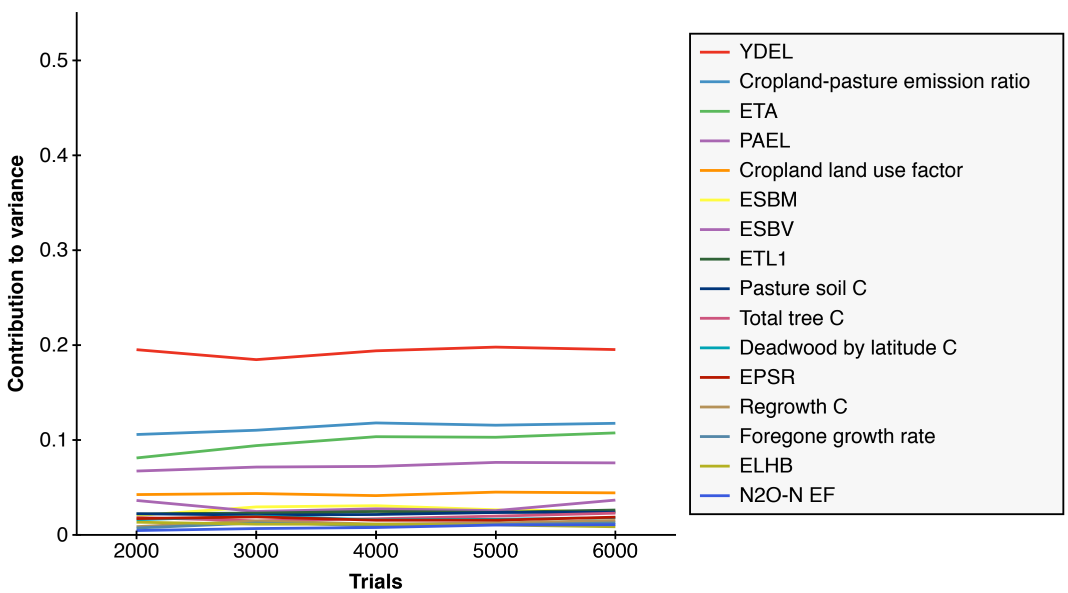


(b) Convergence of the standard deviation for ILUC emissions, corn ethanol.

Figure S9: Convergence plots for the mean and standard deviation for the ILUC emissions associated with corn ethanol (food not fixed).



(a) Convergence of contribution to variance for corn ethanol over 6,000 trials (food not fixed).



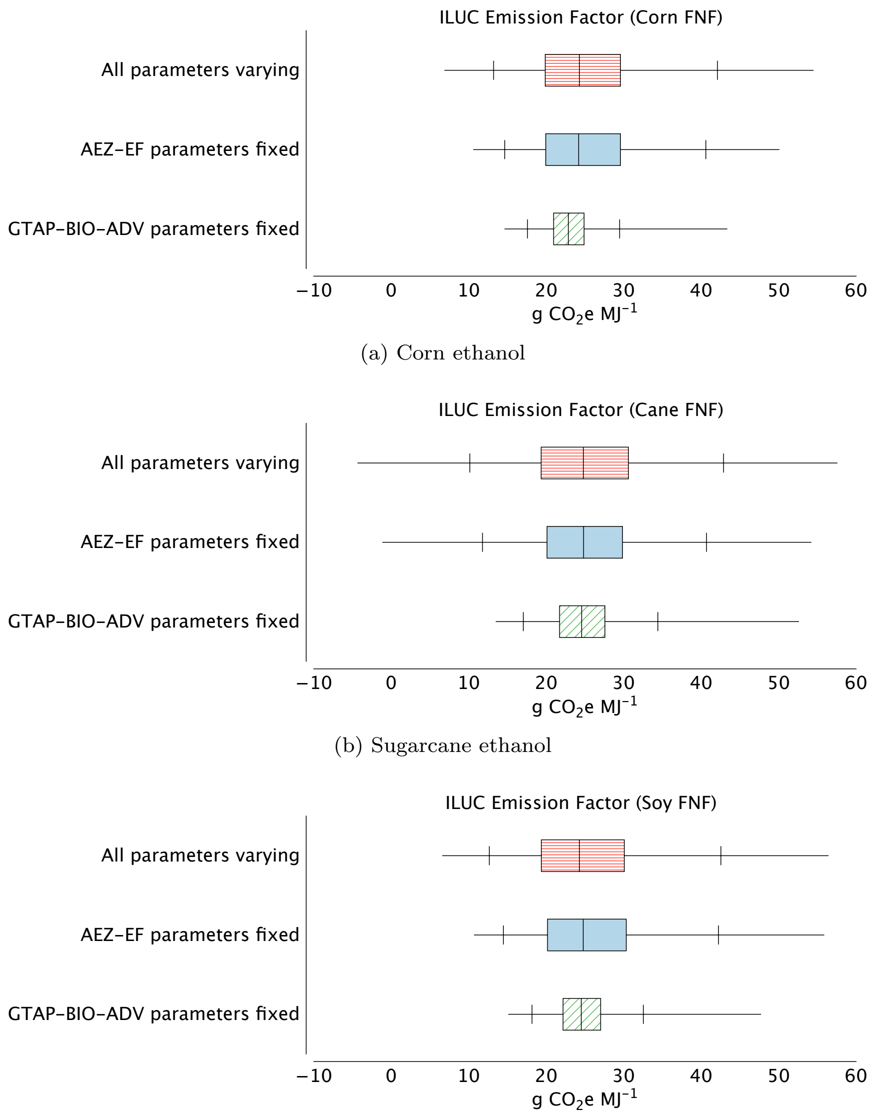
(b) Convergence of contribution to variance for sugarcane ethanol over 6,000 trials (food not fixed).

Figure S10: Convergence of contribution to variance as a function of the number of trials examined.

S4.3 Contribution to variance

We estimate contribution to variance using normalized rank (Spearman) correlations. For each input parameter, we compute the rank correlation with various output parameters across all trials. The rank correlations are squared and normalized to a percentage by dividing each by the sum of the squared correlation values. We restore the original sign to indicate directionality. Figures S12 through S17 show the percentage contribution to variance of the most influential input parameters to ILUC emissions.

Parameters contributing 1% or more to total variance (Table S9) were included; others were considered unimportant contributors individually, though they together accounted for 20% of the total variance. Reducing the number of parameters to the "most important" 16 results in a reduction in the number of random variables from 538 in the "broader" stochastic scenarios to 37. (Several matrix or vector parameters have individual random value for rows and/or columns, thus the larger number of random variables than model parameters.) Simulation with this reduced number of parameters does not change results much. Predictably, the right tail is slightly less extreme, but for the purposes of, say, identifying parameters to include in an SSA, or for running numerous alternative MC simulations with fewer trials, this is a good approximation.



(a) Corn ethanol

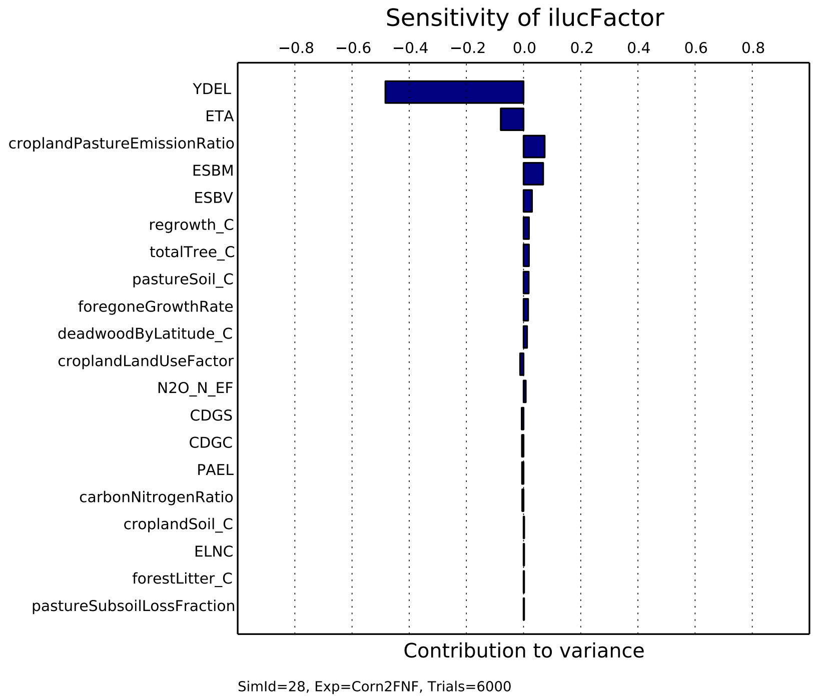
(b) Sugarcane ethanol

(c) Soybean biodiesel

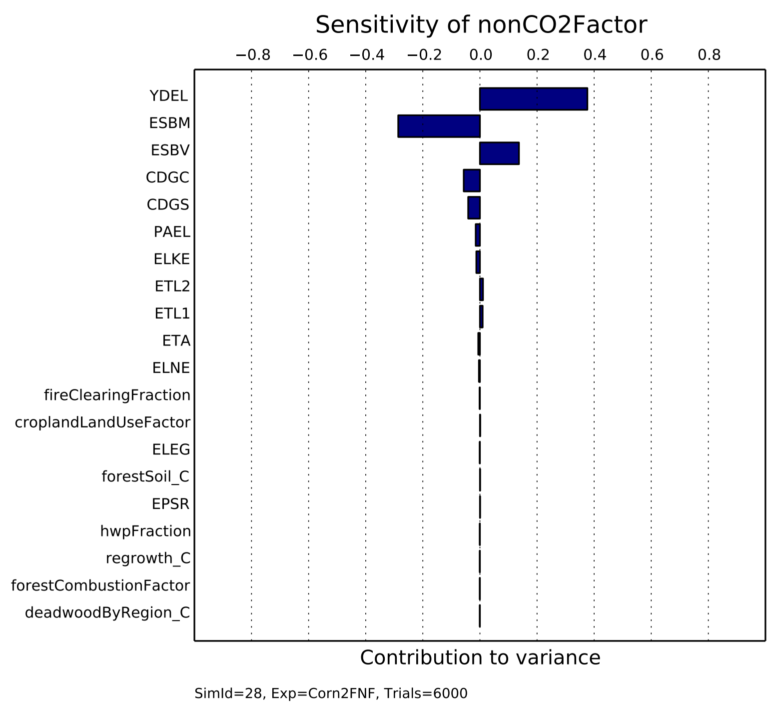
Figure S11: Frequency distributions for ILUC emission intensity showing the relative contributions to variance of the GTAP-BIO-ADV and AEZ-EF models for corn ethanol (a), sugarcane ethanol (b), and soybean biodiesel (c), in all cases with food consumption not fixed. In each plot, the bottom boxplot shows the results when GTAP-BIO-ADV parameters were fixed and AEZ-EF parameters varying. The middle boxplot shows the results with parameters from GTAP-BIO-ADV varying and those from AEZ-EF fixed. The top boxplot shows the results with all parameters varying. The simulations each used 3000 trials.

Table S9: Parameters (in alphabetical order) contributing at least 1% of the variance in ILUC emissions for the CornFNF or CaneFNF experiments.

Parameter Name	Description
	GTAP-BIO-ADV model
ELHB	Elasticity of substitution in biofuel subconsumption
EPSR	Elasticity of substitution in pasturecrop and pasturecover
ESBM	Armington elasticity of substitution within composite import bundle
ESBV	Elasticity of substitution in value-added-energy sub-production
ETA	Relative productivity of newly converted cropland, by Region-AEZ
ETL1	Elasticity of transformation between forest, cropland, and pasture
PAEL	Scalar yield elasticity target for cropland pasture
YDEL	Elasticity of yield with respect to price
	AEZ-EF model
Cropland land use factor	A parameter used in IPCCs method to compute soil carbon change
Cropland-pasture emission ratio	The fraction of emissions from pasture conversion assumed to be emitted upon conversion of cropland-pasture
Deadwood by latitude C	Carbon content of deadwood in boreal, temperate, and tropical AEZs
Foregone growth rate	The rate of tree growth (Mg C ha^{-1} y^{-1}) that would have occurred absent land use change
N_2O-N emission rate	The fraction of applied N released in the form of N_2O
Pasture soil C	The carbon content of pasture soil to 30 cm
Regrowth C	Tree growth for reforestation over 30 years (correlation with Foregone growth rate=.75)
Total tree C	The total amount of carbon in trees

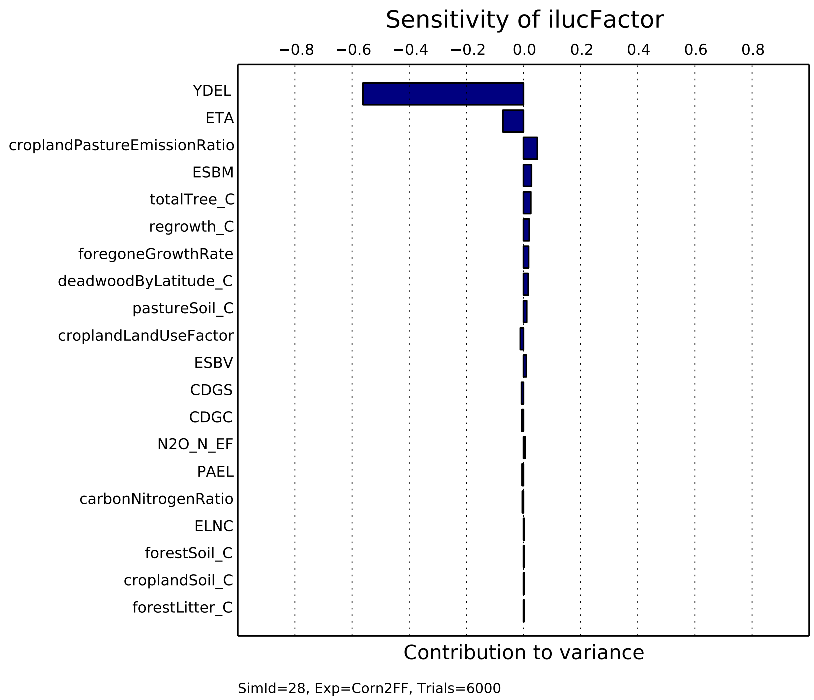


(a) Uncertainty importance for ILUC emissions, corn ethanol (food not fixed).

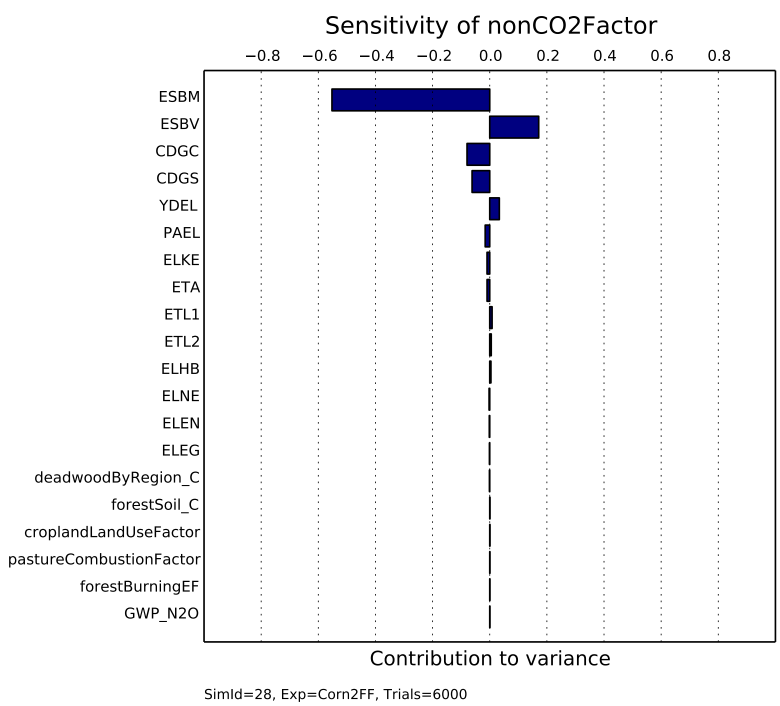


(b) Uncertainty importance for non-CO_2 emissions, corn ethanol (food not fixed).

Figure S12: Contribution to variance in ILUC factor and non-CO_2 emissions (corn ethanol; food not fixed).

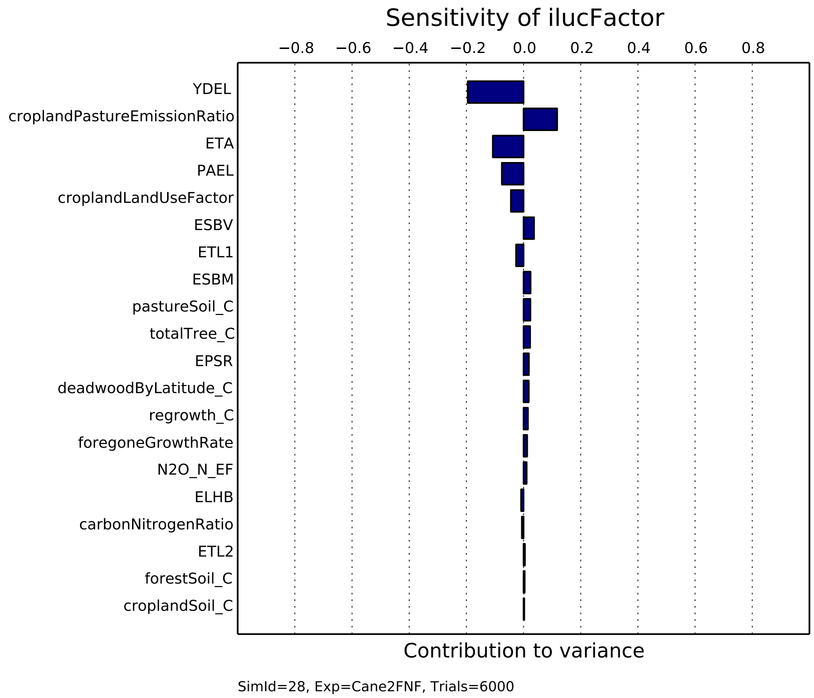


(a) Uncertainty importance for ILUC emissions, corn ethanol (food fixed).

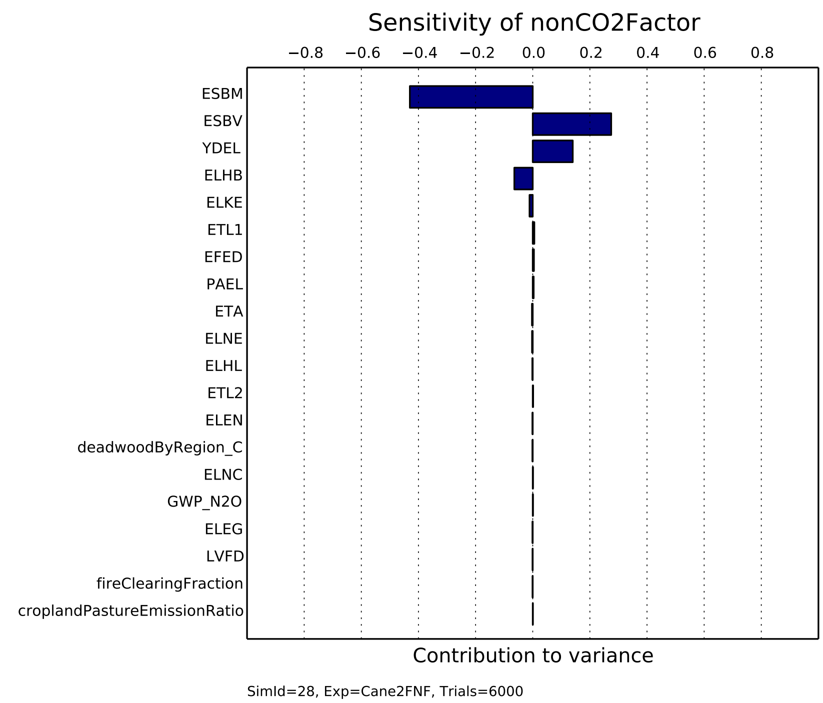


(b) Uncertainty importance for non-CO_2 emissions, corn ethanol (food fixed).

Figure S13: Contribution to variance in ILUC factor and non-CO_2 emissions (corn ethanol; food fixed).

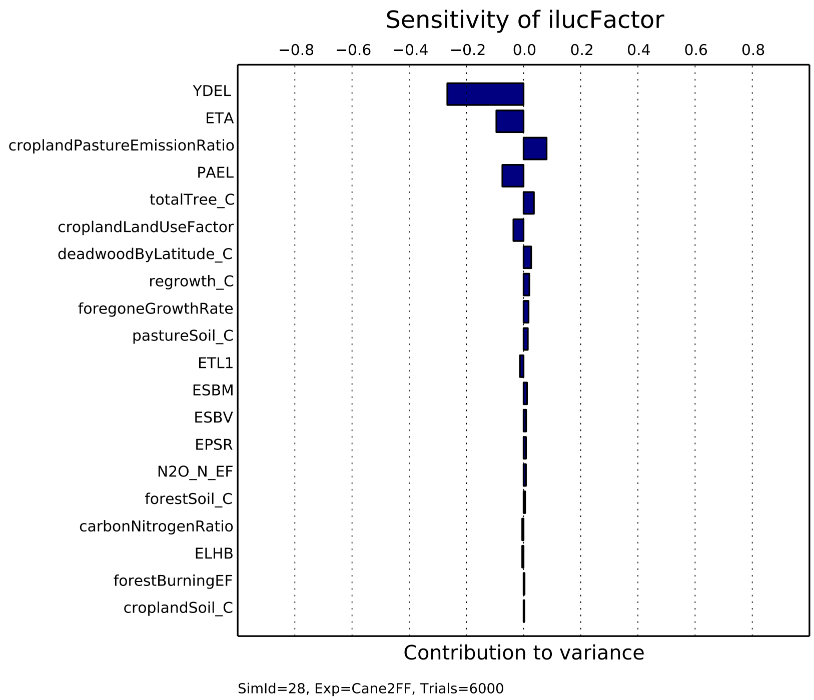


(a) Uncertainty importance for ILUC emissions, sugarcane ethanol (food not fixed).

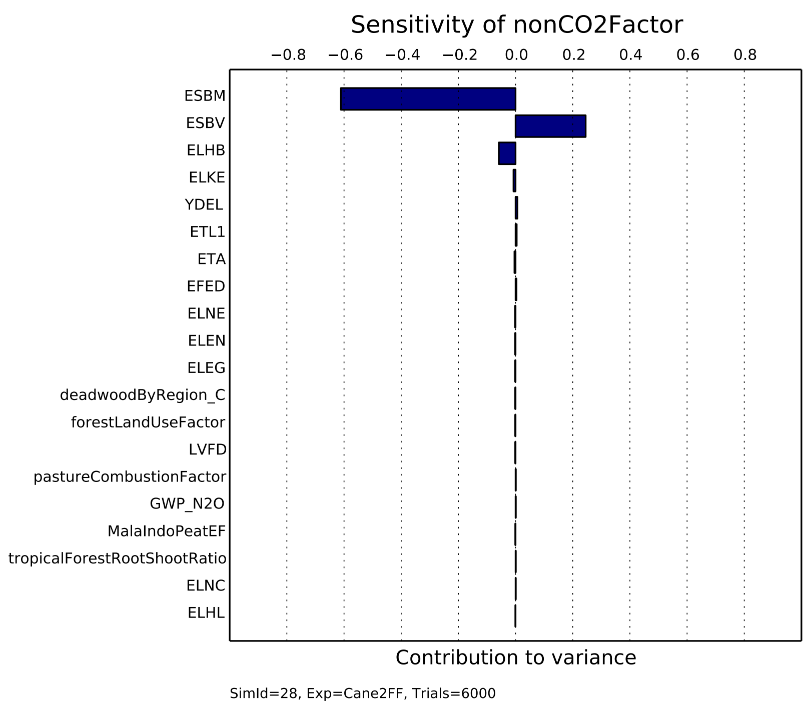


(b) Uncertainty importance for non-CO_2 emissions, sugarcane ethanol (food not fixed).

Figure S14: Contribution to variance in ILUC factor and non-CO_2 emissions (sugarcane ethanol; food not fixed).

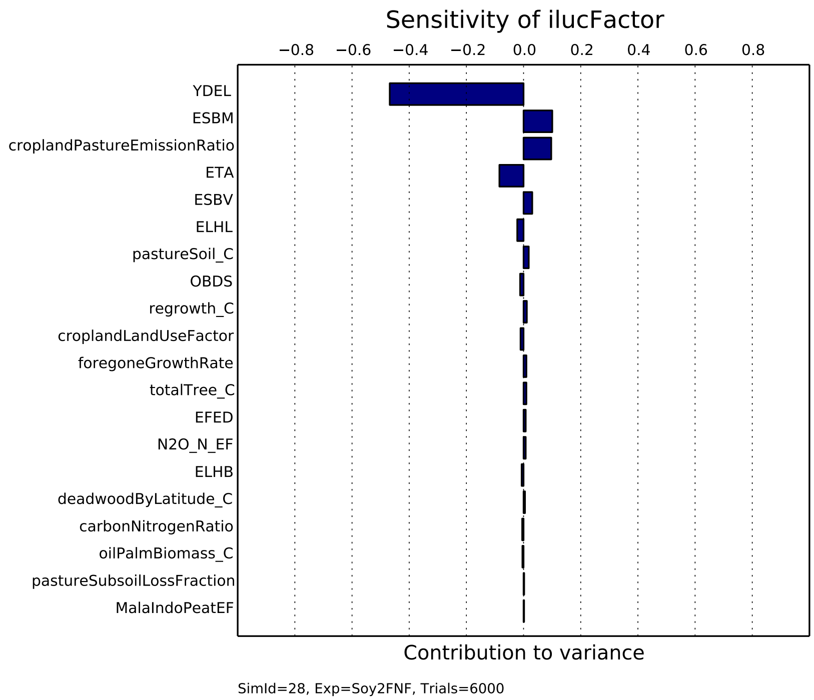


(a) Uncertainty importance for ILUC emissions, sugarcane ethanol (food fixed).

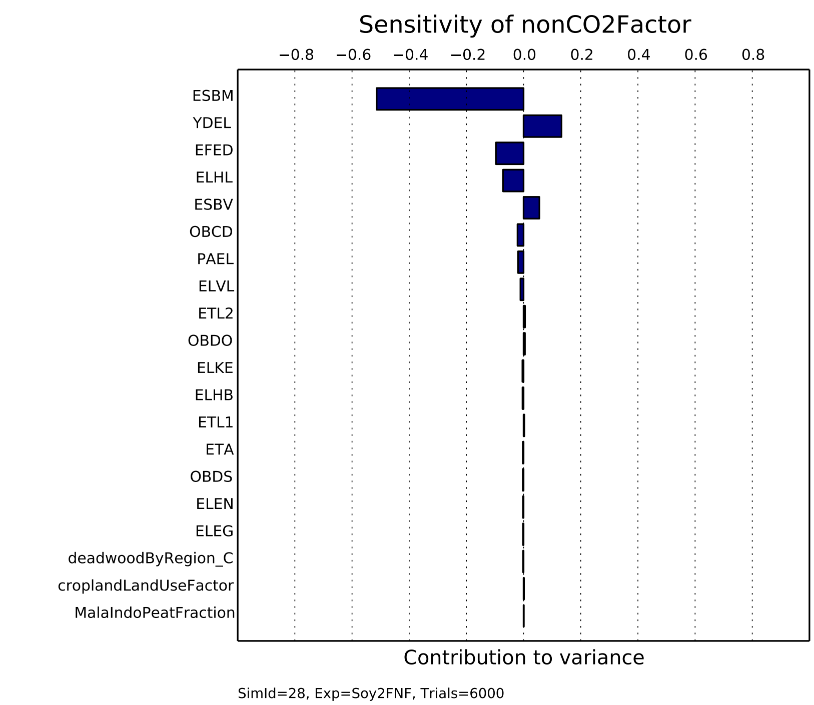


(b) Uncertainty importance for non-CO_2 emissions, sugarcane ethanol (food fixed).

Figure S15: Contribution to variance in ILUC factor and non-CO_2 emissions (sugarcane ethanol; food fixed).

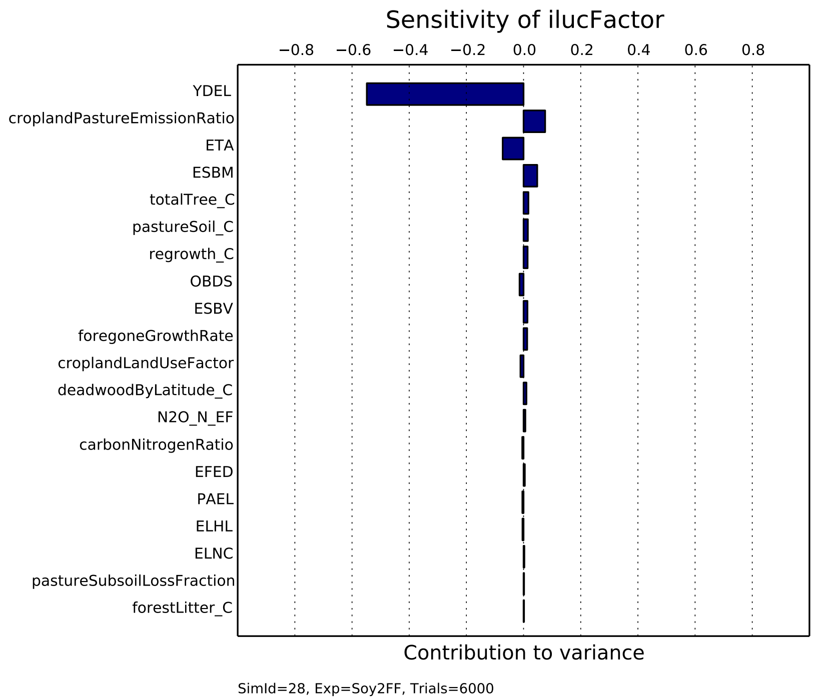


(a) Uncertainty importance for ILUC emissions, soybean biodiesel (food not fixed).

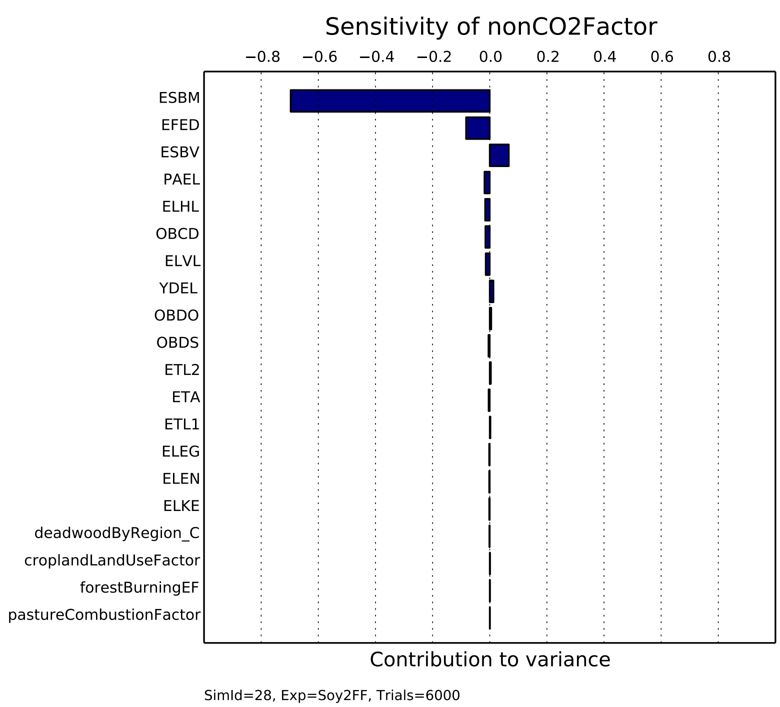


(b) Uncertainty importance for non-CO_2 emissions, soybean biodiesel (food not fixed).

Figure S16: Contribution to variance in ILUC factor and non-CO_2 emissions (soybean biodiesel; food not fixed).



(a) Uncertainty importance for ILUC emissions, soybean biodiesel (food fixed).



(b) Uncertainty importance for non-CO_2 emissions, soybean biodiesel (food fixed).

Figure S17: Contribution to variance in ILUC factor and non-CO_2 emissions (soybean biodiesel; food fixed).

S5 Comparison with other studies

Table 1 compares results for corn ethanol uncertainty analysis from this paper and with those from other studies, indicating which type of model parameters (economic or GHG accounting) were incorporated in the uncertainty analysis.

Below, we briefly discuss uncertainty analysis for the two studies less described in the main text (USEPA, 2010; Laborde and Valin, 2011). For the Plevin (2010) entry, the values presented are those generated using uniform distributions on all parameters. The values shown for Hertel et al. (2010) were generated using the RunGTAP Systematic Sensitivity Analysis feature, which is based on a Gaussian Quadrature approach. The remaining analyses used Monte Carlo simulation. The values from Laborde and Valin (2011) were multiplied by 2/3 to convert from 20-year to 30-year amortization as was used in the other studies listed. ILUC-MCS (A) shows results without fixing food consumption in developing countries, while (B) shows results with fixing food consumption. ILUC-MCS (C) shows results varying only GTAP parameters, while (D) shows results varying only AEZ-EF parameters. Note that the evaluation years for USEPA and Laborde studies were hypothetical worlds in 2022 and 2020, respectively, whereas the Hertel et al. study and ILUC-MCS assumed medium-term adjustment to a shock imposed on the global economy in 2001 and 2004, respectively. All studies targeted US corn ethanol shocks, except for Laborde and Valin (2011), where the amount and location of corn production to satisfy the EU RED policy was endogenously determined (alongside the rest of the biofuel mix, given an exogenously determined ethanol/biodiesel ratio).

Table S10: Uncertainty ranges estimated for indirect land-use change emission intensity from expanding corn ethanol production. See text for detailed explanation.

	Parameters varied		ILUC emission factor (g CO_2e MJ^{-1})		
Model	Economic	GHG Accounting	2.5% value	Mean	97.5% value
ILUC-MCS (A)	✓	✓	13	25	42
ILUC-MCS (B)	✓	✓	18	33	55
Hertel et al. (2010)[a]	✓	✓	2	27	52
Plevin (2010)[b]	✓	✓	21	62	142
ILUC-MCS (C)	✓	–	15	25	41
Laborde and Valin (2011)	✓	–	4[d]	7	8.8[e]
ILUC-MCS (D)	–	✓	18	23	29
USEPA (2010)[c]	–	✓	22	30	40

[a] Examining the variation of the most controversial yield parameters yielded a range of 15 to 90 g CO_2e MJ^{-1}.
[b] Based on the results using uniform distributions.
[c] International (outside US) LUC emissions only, for year 2022.
[d] 5% value
[e] 95% value

S5.1 Laborde (2011)

This study included a Monte Carlo analysis on seven economic parameters, over 1000 trials, assuming a log-uniform distribution centered on the default value. The approach assumes perfect correlation across sectors/countries or regions for a given parameter, but independent draws for each parameter. To establish parameter ranges, they extended initial parameter ranges drawn from studies in the literature, dividing the lower bound by 2 and multiplying the upper bound by 2, except as noted below.

Economic factors incorporated in this analysis overlapped partially with our analysis, although the parameter list differed somewhat because of the differences between the GTAP-BIO-ADV and MIRAGE-Biof model structures and assumptions. Overlapping factors included (default values and ranges for developed and developing countries —sometimes different—are in the report):

- Endogenous yield response (YDEL in GTAP). In MIRAGE-BIof, this effect is incorporated in two separate parameters, due to that models splitting out of key production inputs (feedstuff, fertilizer) from primary factors (e.g., capital).
- Land substitution among crops (ETL2 in GTAP). In MIRAGE-Biof, the land use nesting structure includes a different nest (and elasticity) for substitution among highly substitutable crops (e.g., grains and oilseeds), and between other crops.
- Land expansion into other covers (ETL1 in GTAP). In MIRAGE-Biof, this parameter is used to characterize ease of land transformation between used and available land for cropping. Thus, new pasture and forest land are both considered potentially available for cropping. Although they were not distinguished from one another in the MCS analysis, a shifter to a variable indicating the share of new land coming from primary forest (based on work by Winrock International for the US EPA), was also included in the Monte Carlo analysis.[4]
- Marginal yield return on new cultivated land (ETA in GTAP). For this parameter, the lower bound and upper bound were set (at 0.5 and 1, respectively), rather than determined by dividing and multiplying identified ranges by a factor of 2.
- Shifter in demand elasticity for use of agricultural commodities as intermediate inputs (changes how easily processing sectors substitute away from biofuel crops and oils, in response to the modeled shock) (not in our MCS analysis using GTAP).

Several findings are worthy of note. The MCS found that the amount of newly converted land due to the biofuel shock was more uncertain (widely dispersed) than changes in land already in production, highlighting the impact of uncertainty about competition among forest, pasture, and cropland. The study also used MCS results to analyze correlations of LUC effects across feedstocks, with a preliminary recommendation to consider diversification of feedstock portfolio between tropical and non-tropical sources (as feedstocks within these geographical areas were correlated with each other).

S5.2 US EPA (2010)

The study applied Monte Carlo analysis only to international (not domestic) land use change emissions, based on output from the FAPRI-CARD international trade economic model on changes in area in particular crops by region. The analysis first applied heuristic rules to translate economic model output into changes in several land cover types "agriculture"—annual, perennial, and pasture—or "natural (unmanaged) land. The CI analysis required estimating emission consequences of conversion of various types of unmanaged land—savanna, forest, shrubland, wetlands, and mixed (cropland and natural vegetation). In contrast to the GTAP treatment of forest land (accessible for economic activity, in particular forestry), all forest for the EPA analysis was considered "natural." The method was to apply proportional conversion rates of natural land estimated from historical satellite imagery.

The uncertainty analysis then focused on estimates of types of unmanaged land that would be likely to undergo conversion, and their associated emissions, using:

[4]Other assumptions on competition between productively used noncrop land were examined in sensitivity analysis supplemental to the MCS analysis.

(a) classification of land covers from satellite imagery that were used to derive proportional allocations for expansion of cropland into unmanaged territory (MODISv5 images for 2001and 2007); and

(b) errors in parameters used to estimate emission factors for conversion from one land cover type to another.

For (a), a stochastic model based on estimated standard errors for the process used to adjust land cover classifications relying on number of sampled validation training sites and the aggregation of land cover classes to categories used by the EPA, and generated 95% confidence intervals for each of 54 international regions. The stochastic model assumed normally distributed classification errors, and generated 300 alternative realities of land covers based on the standard errors, each of which was used to calculate the pool of land conversion shares for each region.

For (b), the uncertainty in emission factors, Monte Carlo analysis was performed to generate 95% confidence intervals for administrative units/countries, combining uncertainty in sources of emissions (input parameters) derived from data or expert opinion, assuming normality in parameter distributions, and imposing perfect cross-variable correlation in cases using common data source, geographic location, or data interpolation process.

A Monte Carlo analysis was run combining uncertainty from (a) and (b), generating 300 land use change trials by the method used in (a), and 50 emission factor draws by the method used in (b) for each of these 300 trials, for a total of 15,000 iterations. Mean and 95% CI of emissions were calculated for each region and across regions using weighted average emissions across the iterations. The EPA emphasized that the uncertainty surrounding the use of historical patterns to predict future land use change share allocations for unmanaged land was not part of the MCS.

S6 Model limitations

This analysis examined only a small subset of the uncertainties associated with the GTAP-BIO-ADV and AEZ-EF models. Out of necessity, most model parameter probability distributions were based on our subjective judgment. In addition to parameter uncertainty, both GTAP-BIO-ADV and AEZ-EF involve model uncertainties that are difficult to quantify, such as choices of functional forms for production and demand functions and for soil carbon loss, uncertainties in base year data, biases or inaccuracies introduced by aggregating sectors and regions, and omissions such as irrigation constraints on agricultural expansion and the inability to convert non-commercial land into commercial use.

Like all models, the two models used here gain tractability by simplifying a more complex reality. Although the models are useful for illuminating relationships among key parameters and sub-systems, as noted at the outset of the paper, their results should not be interpreted as a prediction of real world outcomes. We discuss model uncertainties and limitations in more detail below.

S6.1 GTAP

Some GTAP parameters are econometrically estimated, however many are based on expert opinion, i.e., educated guesses. Even where parameters are estimated, owing to a paucity of data, many regions and sectors are assigned a single value based on literature describing a small number of regions. Given the lack of data for point estimates, it is not surprising that there is generally inadequate information from which to develop distributions for model inputs. For this reason, some studies assign a simple stylized distribution to all or most parameters such as normal distributions with a coefficient of variation of 20% (Elliott et al., 2011), or uniform distributions from 50% to 200% of the parameters point estimate (Laborde and Valin, 2012).

The GTAP-BIO-ADV model employed in this paper is a static rather than dynamic model; the experiments presented in this paper do not capture changes in population, labor force, preferences and technology over time. Thus, the model results therefore portray the effects of an instantaneous increase in biofuel production in the assumed circumstances and year. This version of the GTAP model represents only land in current economic use for forestry, livestock grazing, and cropping. Unlike some other CGE models (e.g., MITs EPPA and IFPRIs MIRAGE) and partial equilibrium models (e.g., GCAM), this version of GTAP cannot project conversion to economic use of land not currently in economic use. For the purposes of estimating ILUC emissions, it would be helpful if GTAP were modified to include this capability.

In addition to parameter uncertainty, CGE models involve model uncertainties stemming from modeling choices that are difficult to quantify, such as the effects of the choice of functional forms for production and demand functions, the level of sectoral and regional aggregation, the choice of baseline year, and the calibration of model parameters to that year. In addition, the national input-output data used to produce the core social accounting matrix (SAM) is of varying accuracy and the SAM must be manipulated into an initial equilibrium state that didn't exist in the real world. The procedure for adjusting the SAM is inevitably somewhat subjective. These uncertainties are not generally included in uncertainty analyses of CGE models, yet these choices can substantively affect model outcomes (Jansen, 1994; Roberts, 1994; McKitrick, 1998; Abler et al., 1999). Model uncertainty is particularly challenging to quantify because we cannot compare results to the real world to gauge the models accuracy. Though, with respect to the GTAP model, several studies did validate the model (Liu et al., 2004; Valenzuela et al., 2007; Beckman et al., 2011). More generally, all complex, open systems (including CGE and ecosystem carbon accounting models) in which processes are incompletely understood, and input data incompletely known, are fundamentally unverifiable (Oreskes et al., 1994).

With time, as more data become available, additional aspects of the modeling exercise (behavioral parameters, expected correlations, outcomes) may be verifiable empirically using statistical analyses.

Similar uncertainties arise with respect to ecosystem carbon accounting models: functional forms, choice of the parameters, and assumptions required to fill in missing values. These are discussed further in section S6.2.

Beyond uncertainties associated with model construction and data manipulation, estimating ILUC emissions requires several additional assumptions or modeling choices that substantively affect model results, including:

- Assumption that ILUC can be estimated through economic modeling alone, i.e., that economic model parameters—primarily elasticities—and constraints can be adjusted to reflect important economic or non-economic factors, e.g., if land that is not used productively is subject to expropriation
- Choice of model type: CGE vs PE (trading off the strength of capturing complex economy-wide feedback effects (in CGE) against important sectoral bottom-up detail)
- Choice of, and consistency in timeframe examined
- Policy choices about which features to include in the model (e.g., whether reduction in food consumption should receive GHG reduction credit), and which real-world factors are important enough to include (e.g., irrigation constraints)

Other limitations particular to the version of GTAP we've used include:

- Disagreement among experts about correct values for model parameters
- Forest that is not currently in economic use cannot be brought into production.

S6.1.1 Estimating ETA: the relative productivity of newly converted cropland

The GTAP-BIO-ADV model estimates the relative productivity of land converted to cropland using the ratio of (i) the average net primary productivity (NPP) of land not in crop production at in the initial equilibrium state, to (ii) the average NPP of land in crop production in the initial equilibrium (Taheripour and Tyner, 2012). These values are computed per Region-AEZ combination using the Terrestrial Ecosystem Model by modeling the NPP of a C4 crop (calibrated to corn grown in Iowa in 1996).

To account for the increased productivity resulting from irrigation of currently cropped land, the ratios are reduced by 10%. Based on the assumption that new cropland would not be more productive than existing cropland, on average, the ratios are then truncated to 1. While this approach is a conceptual improvement over applying a single value for ETA globally (e.g., Hertel et al., 2010), the method used requires several critical assumptions:

1. **TEMs estimates of NPP are reasonably accurate.**
   Pan et al. (1996) performed a sensitivity analysis on the TEM model (version 4.0), showing that estimated NPP is sensitive to different assumptions about soil texture, temperature, precipitation, and radiation. These factors vary over space and time.

2. **The ratio of average NPP of non-cropped land to cropped land is a good proxy for the relative yield of land actually brought into production.**
   For this to be true, either the variance around yield values in a given region must be small, or the selection of land must be random. If land selection is based on assumed yield potential, using the average would underestimate the yield, whereas if land is selected by proximity to existing cropland, its unclear whether the average under- or overestimates the yield. As the authors indicate, a single Region-AEZ can contain land with widely varying productivity (Taheripour and Tyner, 2012). In addition, differences in management practices can produce large differences is yield, regardless of potential NPP. Note that if land selection were indeed random, there should be no difference in productivity between cropped and non-cropped landand this analysis wouldnt be necessary.

3. **Reducing the NPP ratios by 10% is a good proxy for the effects of irrigation.**
   Irrigation is only one of the management practices that affects actual productivity, and its unclear that a 10% correction accounts for these differences.

4. **Truncating the ratios to 1 produces a more accurate result.**
   If the basic approach of using the NPP ratio to estimate relative productivity is correct, its unclear why a correction should be required. Nor is it clear why 1 is the best value: why not 0.9, 1.1, or some other value? Should truncation to 1 precede reduction by 10% for irrigation?

   We hasten to note that these factors do not invalidate the method of estimating ETA, but the cascade of uncertainties represented by these assumptions does suggest treating the resulting ETA values as coarse approximations. In the end, its difficult to judge whether this approach produces a more accurate result than was achieved using a single global value for ETA.

S6.2 AEZ-EF

The report on the AEZ-EF model documents numerous uncertainties and limitations associated with that model (Plevin et al., 2014). We briefly summarize them here.

- Forest carbon stocks represent the area-weighted average for all forested land in each Region-AEZ, while GTAP represents only (economically) accessible forest.

- Wetlands are assumed to not be cropped; a carbon density threshold is applied to identify and filter out wetlands.
- Estimates of forgone sequestration depend on the unknowable future state of a forest.
- Estimates of carbon in long-lived wood products are very coarse.
- Estimates of below-ground carbon are estimated based on allometric equations.
- Estimates of the carbon stored in litter, understory plants, and harvested wood products are based on coarse estimates.
- CO2-equivalence is summed for only CO2, CH4, and N2O. Other known climate forcing effects (e.g., albedo change, emissions of aerosols and GHG-precursors) are excluded from the analysis.
- GTAP does not project specific land cover transitions; it provides only projected net changes in area for each land cover class. Heuristics are applied to identify plausible land transition sequences.
- Up-front emissions resulting from LUC are linearly amortized over 30 years; the atmospheric residence time and decay of GHGs is not accounted for.
- All emissions from above and below-ground carbon are assumed to be released immediately.

S6.2.1 Cropland-pasture emission ratio

To represent the emission from conversion of cropland-pasture (C-P), the AEZ-EF model multiplies the emissions computed (by region and AEZ) for pasture conversion by 50% to estimate the emissions from cropland-pasture conversion. We note that in the Monte Carlo analyses, we represented this value with a triangular distribution bounded by 0 and 1 with a mode of 0.5.

S6.2.1.1 Treatment of Cropland-Pasture in the CCLUB model

Here we compare the approach taken in the CCLUB model, released by Argonne National Laboratory (Dunn et al., 2013).

To assess changes in soil carbon in the U.S., CCLUB uses a "surrogate" Century model (results of saved Century model runs), at the county level, using each counties dominant soil textures, as well as yield and weather history.

CCLUB estimates emissions only for the conversion of cropland-pasture, forest, and pasture to biofuel feedstock production. That is, when modeling biofuels from corn, miscanthus, switchgrass or corn with stover removal, CCLUB assumes these lands are converted to the modeled biofuel feedstock. CCLUB averages emissions factors across counties in an AEZ to produce a single, average AEZ value, which is then applied to area projected to change by the GTAP model. We note that the emission factors are not weighted to account for the different land area in each county. For land-use changes outside, the model incorporates emission factors developed by Winrock International for the U.S. Renewable Fuel Standard (Harris et al., 2008; Harris, 2011).

The CCLUB model authors created a young forest-shrub category within 'accessible' forest to reconcile GTAP forest data with other data sources. This assumption, however, is applied to land-use change estimates for which GTAP-BIO-ADV considered this land to be commercial forest.

The accuracy of results using the CCLUB method depends on these assumptions:

1. **The dominant (majority or plurality) soil type in each county is assumed to be a good proxy for the average soil type (and C stock or conversion emissions) in the county.**

2. **The simple (i.e., not area-weighted) average C stock by county in an AEZ is assumed to be a good proxy for the average C stock in the AEZ.**

   One value is used to represent the C stock in each county in an AEZ, and the resulting values are averaged without regard to relative land area: an area-weighted average would be more appropriate since the size of counties is highly variable. According to data downloaded from the US Census website, the maximum county area in the US is 145,505 sq mi, with an average of 1,124 sq. mi. and a standard deviation of 3,611 sq. mi. From our carbon database, its clear that C stocks are also highly variable spatially. It's unclear how the use of simple averaging biases the results: if the extremes of area line up with the extremes of C, using the average could be highly distorting. Of course, we might be lucky and it all just averages out despite the high variance.

3. **The Century model is assumed to accurately represent actual land conversion emissions.**

   Century represents changes in soil C to a depth of about 20 cm, while recent research suggests that studies that sampling to this level misses important changes in soil C occurring in the full top meter of soil as a result of different tillage practices. The point here is that a model such as Century is only as accurate as the data used to calibrate it, and if methods used to produce that data introduce biases, the modeled results will be similarly biased.

   To produce the data used in the surrogate model, the authors first spin up Century to represent current soil C stocks, based on an assumed land-use history. Specifically, the authors assumed that all C-P was in crops from 1880-1950, in pasture/hay/grasslands from 1950-2010, and then in corn-corn or miscanthus/switchgrass from 2011-2040. Even if Century models this land-use history perfectly, if the land-use history of the land converted deviates from these assumptions, the Century projection will misrepresent the actual state of the land.

   The actual land-use history of the converted cropland-pasture strongly determines conversion emissions: land recently in crops will have very low emissions, while lands taken out of crop production long ago will have high emissions. Simply assuming a single land-use history across all land does not address this key information gap.

4. **All converted C-P is assumed to be replaced by the biofuel feedstock being examined in GTAP-BIO-ADV.**

   GTAP-BIO-ADV, however, offers no indication of which specific crop is grown on converted C-P, nor does it indicate how many ha were converted from cropland-pasture to cropland overall; it merely provides the net change in each land use type, by region and AEZ.

5. **Treating GTAP-BIO-ADV results as applying to a young forest-shrub category is assumed to not bias the results.**

   Even if the new category represents land more accurately, this assumption is at odds with assumptions underlying the economic logic. For example, if the GTAP model projects conversion of forestry land to cropland, the supply of timber is reduced, which increases price and induces afforestation in other regions. If this land is actually young forest-shrub, timber supply would be unaffected and thus the afforestation would not be induced.

6. **The average carbon stock in an AEZ is assumed to be a good proxy for carbon on the land actually converted.**

   This assumption is reasonable if one further assumes either low variance of C stocks across the AEZ, or that the land converted is randomly selected across the AEZ. The point is that if there is wide variability and some non-random approach is used to select C-P land for conversion, the C stock on land actually converted could be biased toward one extreme or the other. Of course it also requires that all the preceding assumptions are reasonable.

Given all of the assumptions required in this more complex procedure, there is little basis for calling the result obtained more accurate than a simple assumption with a wide variance, and the apparently greater precision imparted by excluding uncertainty is specious. We agree that the approach of assigning cropland-pasture conversion half the emissions of pasture conversion is a coarse simplification.

However, since the real value is unknown, we assigned this parameter a broad distribution (uniform with minimum of 0 and maximum of 1) in the Monte Carlo simulation. The USDA definition of cropland-pasture admits everything from cropland to pasture[5].

In addition to not knowing the actual carbon stock on actual cropland-pasture, the modeling framework cannot identify which plot of cropland-pasture is converted, other than placing it within a given Region-AEZ. High variance in C stock, and probably land-use history, therefore translates into broad uncertainty about actual emissions.

The primary difference between the CCLUB approach and the AEZ-EF approach is that CCLUB employs more finely-resolved data, not that it requires fewer assumptions. The CCLUB result appears more scientific, but given the reliance on a number of assumptions without a strong empirical basis, e.g, computing average AEZ carbon stocks using unweighted county-based averages, assuming that C-P transitions exclusively to biofuel feedstocks, and assuming that all C-P has a single, known, land-use history—there's no basis for concluding that CCLUBs approach is more accurate. Both approaches are limited by the assumptions required to bridge low-resolution CGE model results—which are not spatially-explicit and do not identify specific land transitions—to the spatial resolution and land-transition specificity required to estimate carbon changes.

S6.3 Limitations of the Monte Carlo simulation and analysis

In their analysis of ILUC using a CGE model of Brazil, Ferreira Filho and Horridge (2014) note:

> A CGE model like that used here builds on a host of assumptions; about functional forms; about assumed elasticity values; and about initial data. Rarely do we have a probability distribution which measures the uncertainty of estimates that are fed in—so we cannot in general compute probability distributions for model outputs. We can however merely report how results depend on input values.

For example, in the GTAP-BIO-ADV model used in the present analysis, the land nesting structure is very simple (cropland, pasture and forests compete in the same nest), potentially leading to overestimation of conversion from forests to cropland (Babcock and Carriquiry, 2010); other CGE models employ a more complex structure (Ahammad and Mi, 2005; Golub and Hertel, 2008). However, a more complex nested structure requires additional transformation parameters, by AEZ and region, which should be calibrated to land supply elasticities for which econometric estimates are not currently available. Given this issue, Laborde and Valin (2012) conducted sensitivity analysis on the nesting structure of non-cropland.

For these and other reasons, a Monte Carlo simulation with the GTAP-BIO-ADV and AEZ-EF models should likewise not be interpreted as a prediction. The MCS does, however, allow us to interrogate the relationships among input and output parameters given the present model structure and data.

[5]See http://www.ers.usda.gov/data/majorlanduses/glossary.htm#cropforpasture

References

Abler, D. G., A. G. Rodrguez, and J. S. Shortle (1999). Parameter uncertainty in CGE modeling of the environmental impacts of economic policies. *Environmental and Resource Economics 14*(1), 75–94.

Ahammad, H. and R. Mi (2005). Land use change modeling in GTEM: Accounting for forest sinks.

Babcock, B. and M. Carriquiry (2010). An exploration of certain aspects of carbs approach to modeling indirect land use from expanded biodiesel production. Technical report, Center for Agricultural and Rural Development, Iowa State University.

Beckman, J., T. Hertel, and W. Tyner (2011). Validating energy-oriented CGE models. *Energy Economics 33*(5), 799–806.

Beckman, J., R. Keeney, and W. Tyner (2011). Feed demands and coproduct substitution in the biofuel era. *Agribusiness 27*(1), 1–18.

Dunn, J. B., S. Mueller, H. Y. Kwon, and M. Q. Wang (2013). Land-use change and greenhouse gas emissions from corn and cellulosic ethanol. *Biotechnol Biofuels 6*(1), 51.

Elliott, J., M. Franklin, I. Foster, T. Munson, and M. Loudermilk (2011). Propagation of data error and parametric sensitivity in computable general equilibrium models. *Computational Economics 39*(3), 219–241.

Ferreira Filho, J. B. d. S. and M. Horridge (2014). Ethanol expansion and indirect land use change in Brazil. *Land Use Policy 36*, 595–604.

Frey, H. C., J. Penman, L. Hanle, S. Monni, and S. M. Ogle (2006). *Volume 1, Chapter 3: Uncertainties.*

Gibbs, H., S. Yui, and R. Plevin (2014). New estimates of soil and biomass carbon stocks for global economic models. Global Trade Analysis Project (GTAP) Technical Paper No. 33. Technical report, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Golub, A. (2013). Impact of expanded production of biofuels on non-CO_2 GHG emissions. GTAP Resource #4326. Technical report, Purdue University, prepared for the California Air Resources Board.

Golub, A. and T. Hertel (2008). Global economic integration and land use change. *Journal of Economic Integration 23*(3), 463–488.

Golub, A. A., B. B. Henderson, T. W. Hertel, P. J. Gerber, S. K. Rose, and B. Sohngen (2012). Global climate policy impacts on livestock, land use, livelihoods, and food security. *Proc Natl Acad Sci U S A*.

Golub, A. A. and T. W. Hertel (2012). Modeling land-use change impacts of biofuels in the GTAP-BIO framework. *Climate Change Economics 03*(03), 1250015.

Harris, N. (2011). Revisions to land conversion emission factors since the RFS2 final rule. Technical report, Winrock International report to EPA.

Harris, N., S. Grimland, and S. Brown (2008, Oct). GHG emission factors for different land-use transitions in selected countries of the world. Report submitted to US EPA. Technical report, Winrock International.

Hertel, T., A. Golub, A. D. Jones, M. O'Hare, R. J. Plevin, and D. M. Kammen (2010). Effects of US maize ethanol on global land use and greenhouse gas emissions: Estimating market-mediated responses. *BioScience 60*(3), 223–231.

Hertel, T., W. Tyner, and D. Birur (2008). Biofuels for all? understanding the global impacts of multinational mandates. Technical report, Center for Global Trade Analysis, Purdue University.

Iman, R. L. and J. M. Davenport (1982). Rank correlation plots for use with correlated input variables. *Communications in Statistics - Simulation and Computation 11*(3), 335–360.

IPCC (2006). 2006 IPCC Guidelines for national greenhouse gas inventories, Volume 4: Agriculture, Forestry and Other Land Use.

Jansen, P. S. M. K. (1994). Analysis of multipliers in stochastic input-output models. *Regional Science and Urban Economics 24*(1), 55–74.

Keeney, R. and T. Hertel (2005). GTAP-AGR: A framework for assessing the implications of multilateral changes in agricultural policies. Technical report, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Laborde, D. and H. Valin (2011). Modelling land use changes in a global CGE: Assessing the EU biofuel mandates with the MIRAGE-BioF model.

Laborde, D. and H. Valin (2012). Modeling land-use changes in a global CGE: assessing the EU biofuel mandates with the MIRAGE-BioF model. *Climate Change Economics 03*(03), 1250017.

Liu, J., C. Arndt, and T. W. Hertel (2004). Parameter estimation and measures of fit in a global, general equilibrium model. *Journal of Economic Integration 19*(3), 626–649.

McKitrick, R. R. (1998). The econometric critique of computable general equilibrium modeling: the role of functional forms. *Economic Modelling 15*(4), 543–573.

Monfreda, C., N. Ramankutty, and T. Hertel (2009). *Global Agricultural Land Use Data for Climate Change Analysis*. Routledge Press.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science 263*(5147), 641–646.

Pan, Y., A. D. McGuire, D. W. Kicklighter, and J. M. Melillo (1996). The importance of climate and soils for estimates of net primary production: a sensitivity analysis with the terrestrial ecosystem model. *Global Change Biology 2*(1), 5–23.

Plevin, R., M. Delucchi, and F. Creutzig (2014). Letter to the editor. *Journal of Industrial Ecology 8*(3), 468–470.

Plevin, R. J. (2010). *Life Cycle Regulation of Transportation Fuels: Uncertainty and its Policy Implications.* Ph. D. thesis.

Roberts, B. M. (1994). Calibration procedure and the robustness of CGE models: Simulations with a model for Poland. *Economics of Planning 27*(3), 189–210.

Rose, S., M. Avetsiyan, and T. Hertel (2010). Development of the preliminary version 7 non-co2 ghg emissions dataset. gtap research memorandum no. 17. Technical report, Center for Global Trade Analysis, Purdue University.

Taheripour, F. and W. Tyner (2012). Induced land use emissions due to first and second generation biofuels and uncertainty in land use emissions factors.

Taheripour, F. and W. Tyner (2013). Biofuels and land use change: Applying recent evidence to model estimates. *Applied Sciences 3*(1), 14–38.

Taheripour, F. and W. Tyner (2014). CARB 2014 model. Technical report, Purdue University for the California Air Resources Board.

Tyner, W. E., F. Taheripour, Q. Zhuang, D. Birur, and U. Baldos (2010, Apr). Land use changes and consequent CO_2 emissions due to US corn ethanol production: A comprehensive analysis. Technical report, Dept. of Agricultural Economics, Purdue University.

USEPA (2010, Feb 3). Renewable fuel standard program (RFS2) regulatory impact analysis. Technical report, US Environmental Protection Agency.

Valenzuela, E., K. Anderson, and T. W. Hertel (2007). Impacts of trade reform: Sensitivity of model results to key assumptions. Technical report, University of Adelaide, Centre for International Economic Studies.