sustainability

MDPI

Article

Quantifying Policy-Induced Cropland Dynamics: A Probabilistic and Spatial Analysis of RFS-Driven Expansion and Abandonment on Marginal Lands in the U.S. Corn Belt

Shuai Li and Xuzhen He *

School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo 2007, Australia; shuai.li-1@student.uts.edu.au

* Correspondence: xuzhen.he@uts.edu.au

Abstract

Rapid biofuel expansion has significantly reshaped agricultural land use in the United States, raising concerns about the conversion and long-term sustainability of marginal croplands. Understanding how policy incentives influence these land-use changes remains a key challenge in sustainable land management. This study aims to quantify the effects of the Renewable Fuel Standard on cropland expansion and subsequent abandonment in the U.S. Midwest using a probabilistic and spatially explicit framework. The analysis integrates geospatial datasets from USDA, USGS, gridMET, and the U.S. Energy Information Administration, combining indicators of soil productivity, slope, precipitation, temperature, and market accessibility. Bayesian logistic regression models were developed to estimate pre-policy baseline probabilities of corn cultivation and to generate counterfactual scenarios—hypothetical conditions representing land-use patterns in the absence of policy incentives. Results show that over one-quarter of marginal land cultivated in 2016 would likely not have been planted without biopower policy-related incentives, indicating that policy-driven expansion extended into less suitable areas. A second-stage analysis identified regions where such lands were later abandoned, revealing the role of climatic and economic constraints in shaping long-term sustainability. These findings demonstrate the effectiveness of integrating probabilistic modelling with high-resolution spatial data to evaluate causal policy effects and quantify counterfactual impacts—that is, the measurable differences between observed and simulated land-use outcomes.

Keywords: RFS; biofuel policy; marginal cropland; spatial modelling; probabilistic analysis; land-use change

check for updates

Academic Editor: Mehdi Seraj

Received: 18 September 2025
Revised: 23 October 2025
Accepted: 25 October 2025
Published: 28 October 2025

Citation: Li, S.; He, X. Quantifying Policy-Induced Cropland Dynamics: A Probabilistic and Spatial Analysis of RFS-Driven Expansion and Abandonment on Marginal Lands in the U.S. Corn Belt. *Sustainability* **2025**, *17*, 9568. https://doi.org/10.3390/su17219568

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Renewable Fuel Standard (RFS) stands as a landmark policy intervention in the United States, designed to reduce transportation-related greenhouse gas (GHG) emissions and enhance national energy security through mandated increases in renewable fuel use [1]. Established with the Energy Policy Act of 2005 and substantially expanded in 2007, the RFS marked a new era for the U.S. biofuel sector, particularly by driving large-scale production of corn ethanol and soybean-based biodiesel [1]. As a result, the policy has fundamentally reshaped agricultural land use across the Midwest and especially within the U.S. Corn Belt, placing the RFS at the centre of ongoing debates about the environmental and economic sustainability of biofuel mandates [2–4].

While the RFS was conceived to promote climate mitigation and rural economic development, its implementation has raised significant controversy, particularly concerning unintended land-use change [5]. Critics have argued that increased demand for biofuel feedstocks, especially corn, has accelerated the conversion of grasslands, wetlands, and other natural ecosystems into cropland [6]. Such land conversion can release substantial amounts of previously sequestered carbon, thereby offsetting the intended GHG reductions and raising questions about the net environmental benefits of the RFS [2]. The debate is further complicated by the difficulty of assigning direct causality: land-use decisions are influenced by an array of overlapping policies, fluctuating commodity prices, local agronomic conditions, and broader socio-economic trends [7,8]. Thus, isolating the precise contribution of the RFS to observed cropland expansion—particularly on land that is suboptimal for corn production—remains a persistent challenge [9].

Over the past decade, scholars have employed a variety of economic models and empirical analyses to evaluate the impacts of the RFS [10–12]. Simulation-based approaches, such as computable general equilibrium (CGE) and partial equilibrium models, have yielded valuable insights into potential policy-driven land-use change [13]. However, these models often rely on aggregate parameters and scenario assumptions that limit their spatial and causal specificity [1]. Empirical studies using remote sensing data and econometric methods have provided spatially explicit observations of cropland expansion in the RFS era, revealing patterns of increased corn and soybean acreage [14]. One study quantified the consequences of grassland conversion to cropland between 2008 and 2016, revealing that the net change led to a 7.9% increase in annual soil erosion and a 3.7% increase in nitrogen loss, despite cropland area expanding by only 2.5%. This confirms the substantial impact of land conversion on soil and water degradation [15]. Other reviews indicate that for every billion gallons of biofuel produced, cropland area has expanded by approximately 0.01 to 2.45 million acres. Such expansions frequently involve the conversion of grasslands, wetlands, and other marginal lands, raising growing concerns over biodiversity loss, increased carbon emissions, and soil degradation [16]. Empirical evidence further suggests that between 2008 and 2016, corn prices rose by about 31%, the corn-planted area expanded by 8.7%, and the total cropland area increased by 2.4%, accompanied by deterioration in water quality and related environmental impacts. Elevated corn prices made the cultivation of low-productivity lands economically viable, thereby accelerating the conversion and utilisation of marginal lands [17].

However, these analyses often grapple with disentangling the unique effect of the RFS from other confounding factors, such as tax credits, state-level policies, market dynamics, and climatic variability [7,16,18,19]. Moreover, most studies generate deterministic or average estimates, overlooking the spatial heterogeneity and uncertainty that are intrinsic to land-use decision-making under policy intervention [20–26].

A key methodological bottleneck in this field arises from the inherent uncertainty in attributing cropland expansion—especially the cultivation of corn on lands with low agronomic suitability—directly to the RFS policy [16]. Because land-use change is the product of multiple interacting drivers, and because not all new corn fields on marginal or previously uncultivated land can be unequivocally linked to RFS incentives, a purely deterministic attribution may misrepresent both the scale and the character of policy impacts [7,12,14,27,28]. Consequently, there is a growing need for probabilistic approaches that can estimate the likelihood of policy-driven land-use transitions rather than assigning binary or absolute causal relationships. In this study, the analysis is restricted to lands with a National Commodity Crop Productivity Index (NCCPI) of 0.42 or less, consistent with prior literature defining marginal land suitability [29].

To address these challenges, this study develops an integrated framework that utilises probabilistic machine learning and multi-source remote sensing for spatial counterfactual attribution of land-use change. Using pre-policy land-use and biophysical, economic, and climatic variables, we construct probabilistic "no-policy" scenarios. Comparing them with observed outcomes reveals RFS-driven land conversion, including on marginal land unlikely to host corn cultivation otherwise. This probabilistic framework not only improves the granularity and credibility of policy impact assessment but also explicitly quantifies uncertainty, providing policymakers and stakeholders with a more detailed understanding of both the extent and the confidence of RFS-related land-use change.

This study introduces a probabilistic and spatially explicit counterfactual framework that jointly evaluates policy-driven cropland expansion and subsequent abandonment. Unlike conventional econometric or equilibrium models that produce deterministic outcomes, our approach estimates the probability and uncertainty of land-use transitions using high-resolution spatial data. The two-stage design enables assessment of both the initiation and sustainability of RFS-induced land-use change. By integrating Bayesian modelling, multi-source geospatial data, and uncertainty quantification, this research offers a replicable and flexible method for causal attribution in large-scale sustainability policy evaluation.

2. Methodology and Materials

2.1. Study Design

This study adopts a two-stage sequential design to investigate the causal impacts and long-term consequences of the Renewable Fuel Standard (RFS) policy on marginal croplands within the U.S. Corn Belt. The first phase seeks to identify marginal lands that were cultivated with corn in 2016 but would likely not have been planted under pre-policy market conditions. To achieve this, counterfactual planting probabilities for 2016 were estimated using a Bayesian logistic regression model trained on land-use data from 2000 to 2006, with 2005 excluded due to missing or unreliable data. The year 2006 thus serves as the final pre-policy benchmark.

The year 2016 is selected as the post-policy reference point, as it marks the culmination of corn expansion within the region, with several studies identifying it as the year of peak growth before subsequent stagnation or decline. This choice allows for a robust assessment of policy-driven land conversion at its height [30].

The second phase evaluates the long-term sustainability of these newly cultivated lands by determining which were subsequently abandoned. Complete abandonment is operationally defined as the absence of corn cultivation during any year from 2022 to 2024, the most recent period for which data are available. This definition accounts for crop rotation and identifies lands that ultimately proved unsustainable or economically unviable for continued production. Figure 1 presents the flowchart of the study. A total of 12 states were included in the investigation (Figure 2).

2.2. Model Selection

In this study, several alternative models were initially considered, including traditional logistic regression, random forest, and neural networks. Traditional logistic regression offers simplicity and interpretability but is limited in its ability to fully quantify parameter uncertainty [31]. Random forest and neural networks deliver strong predictive performance; however, they are more complex, less interpretable, and do not naturally provide posterior distributions for uncertainty analysis [32]. Although Bayesian neural networks can provide uncertainty estimation, they are less suitable for this study due to their high computational cost and limited interpretability in linking posterior parameters to specific predictors [33].

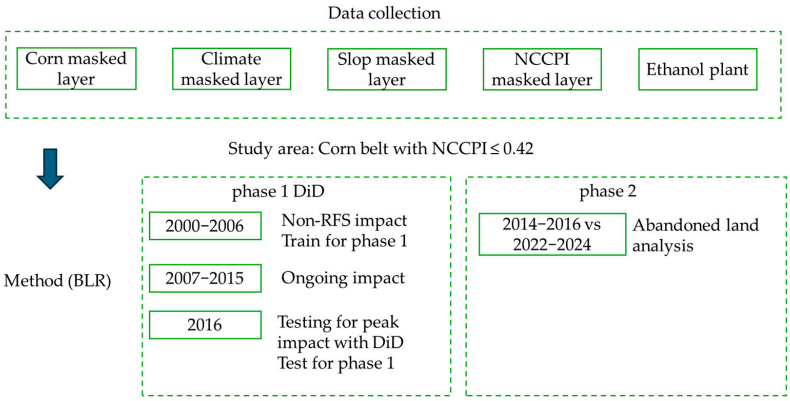


Figure 1. Flowchart illustrating the two-phase analytical framework. Phase 1 estimates pre-policy baseline probabilities (2000–2006) and predicts counterfactual scenarios for 2016 to identify RFS-induced expansion. Phase 2 assesses post-expansion abandonment (2022–2024) to evaluate the long-term sustainability of policy-driven land-use change.

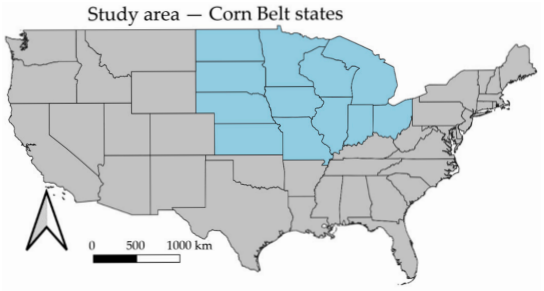


Figure 2. Geographic extent of the U.S. Corn Belt used in this study. The highlighted (shown in blue) states include Iowa, Illinois, Indiana, Nebraska, Minnesota, Missouri, Kansas, South Dakota, Wisconsin, Michigan, Ohio, and North Dakota. The other U.S. regions are coloured in light grey.

Bayesian logistic regression was selected as the core framework for both phases due to its methodological advantages. The aim is not to maximise predictive accuracy but to generate probabilistic estimates of corn planting that reflect biophysical and economic conditions. By modelling the posterior probability distribution, this approach provides an uncertainty-aware understanding of planting tendencies [34,35].

In the context of counterfactual inference, Bayesian methods quantify uncertainty in both parameters and outcomes, addressing partial observability and label noise in land-use data through credible intervals [36–38]. Compared to complex machine learning models, they also offer greater interpretability, with transparent posterior distributions linking covariates such as NCCPI, slope, and precipitation to land-use decisions [39,40]. Finally, the Bayesian framework aligns naturally with the two-stage design, where the focus is on estimating policy-free baselines and long-term land-use sustainability under uncertainty.

2.3. Phase 1: Counterfactual Modelling of Cropland Expansion

In the first phase, a Bayesian logistic regression model is implemented using a standardised input set comprising biophysical and climatic inputs, following standard practices [34]. The binary response variable indicates whether a given pixel of land was planted with corn. Land parcels are mapped to pixels using the corn mask layer available on Google Earth Engine (GEE), which provides annual coverage of corn cultivation. Model fitting is conducted using the No-U-Turn sampler, a Hamiltonian Monte Carlo algorithm, with 2000 posterior draws following 1000 tuning steps to ensure robust convergence. The expression can be specified as Equations (1)–(3):

$$y_i \sim \text{Bernoulli}(p_i) \quad (1)$$

$$\text{logit}(p_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} \tag{2}$$

$$\beta_j \sim \mathcal{N}\left(0, \sigma^2\right) \tag{3}$$

where y_i is a binary indicator of corn planting for pixel i, $x_{i,j}$ denotes the jth standardised predictor (e.g., NCCPI, precipitation, slope, temperature metrics) for pixel i, β_0 is the intercept, β_j are regression coefficients, p_i is the probability of corn planting, σ^2 is the prior variance. In this study, non-informative (weakly informative) priors are adopted for all regression coefficients, σ is specified as 5.

To ensure the reliability of posterior inference, Markov chain convergence was assessed using the potential scale reduction factor $\hat{R}$ [41,42]. Values of $\hat{R}$ close to 1 indicate satisfactory convergence of the sampling chains; all estimated parameters in this study exhibited $\hat{R} < 1.01$, confirming robust posterior estimation.

After model training, the fitted model is applied to the 2016 data to generate counterfactual predictions. For each pixel of land, the mean predicted probability across all posterior samples is used to represent the expected likelihood of planting under pre-policy conditions. Pixels with actual planting but predicted counterfactual probability below 0.5 are identified as likely cases of policy-induced expansion. The posterior distributions of model parameters are visualised and summarised to enable uncertainty-aware interpretation of input effects.

2.4. Phase 2: Modelling Long-Term Abandonment

In the second phase, we develop a separate Bayesian logistic regression model focused on those lands that were newly brought into production during the expansion period (any year from 2014 to 2016, with NCCPI ≤ 0.42). The target variable is defined as complete abandonment, operationalised as a lack of corn cultivation in all years from 2022 to 2024. Negative cases are land pixels that were planted at least once during the same period. This definition excludes short-term crop rotations by requiring continuous non-cultivation of corn for three consecutive years, thereby distinguishing true abandonment from rotational fallowing.

Posterior inference again uses the No-U-Turn sampler, with 2000 draws following 1000 tuning steps. Model performance is evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) on a held-out test set (30% of samples), and the distribution of predicted outcomes is assessed through posterior predictive sampling. The posterior coefficient distributions are interrogated to determine which inputs are most strongly associated with abandonment, enabling an uncertainty-aware analysis of the conditions under which policy-driven expansion failed to sustain long-term cultivation.

2.5. Model Evaluation

Model performance in both phases is first assessed using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) (Equation (4)), quantifying the ability to distinguish between positive and negative cases [43,44].

$$\text{AUC} = \frac{1}{N_1 N_0} \sum_{i:y_i=1} \sum_{j:y_j=0} I\left(\hat{p}_i > \hat{p}_j\right) \tag{4}$$

where N_1 and N_0 are the numbers of positive and negative cases, respectively, $\hat{p}_i$ and $\hat{p}_j$ denotes the predicted probability for sample i. An AUC value closer to 1 indicates better model performance in distinguishing between positive and negative cases. An AUC of 1.0 represents perfect discrimination, while an AUC of 0.5 suggests no discriminative ability (equivalent to random guessing). However, in this context, AUC should be inter-

preted with caution, as observed land-use labels are only a partial manifestation of true suitability—many "false negatives" may represent land that could have been cultivated under different circumstances, introducing label noise and partial observability.

To complement AUC, the calibration of predicted probabilities using reliability curves is evaluated. These curves compare the predicted probabilities to the empirical frequency of positive cases across uniformly spaced bins. A well-calibrated model should yield probability estimates that closely match observed outcome frequencies. In addition, the Brier Score is computed to quantitatively assess the accuracy of probabilistic predictions [45]. The Brier Score is defined as Equation (5):

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2 \tag{5}$$

where N is the number of samples, $\hat{p}_i$ is the predicted probability for sample i, and y_i is the observed outcome for sample i. The Brier Score takes values between 0 and 1, with lower values indicating better calibrated and more accurate probabilistic predictions.

Beyond point predictions, the Bayesian framework provides posterior distributions over model parameters and predicted outcomes, formally quantifying epistemic uncertainty. This is particularly valuable in the context of counterfactual modelling, where many cases are ambiguous or unobserved. For each pixel of land, both a predicted probability and a credible interval are reported, allowing for detailed, uncertainty-aware interpretation. Posterior distributions of model coefficients provide insight into the relative certainty with which each input influences planting or abandonment likelihood. Notably, predictions with high uncertainty can be flagged for further investigation, especially in policy contexts where land-use sustainability is in question.

2.6. Input Data and Variable Preparation

To prepare the modelling dataset, a balanced sample of 3000 positive cases (pixels classified as corn) and 3000 negative cases (uncultivated pixels) was randomly drawn from marginal lands (NCCPI $\leq$ 0.42). All continuous predictors were standardised to zero mean and unit variance prior to model fitting to ensure comparability of regression coefficients. Soil productivity was captured by the National Commodity Crop Productivity Index (NCCPI), a composite indicator derived from the SSURGO database, which integrates key soil properties including texture, organic matter, pH, and water-holding capacity [46]. Due to the comprehensive nature of NCCPI, additional soil chemical variables (e.g., nitrogen, carbon, EC, CEC) were excluded as either redundant or unavailable at adequate spatial resolution. The spatial distribution of key inputs is illustrated in Figure 3.

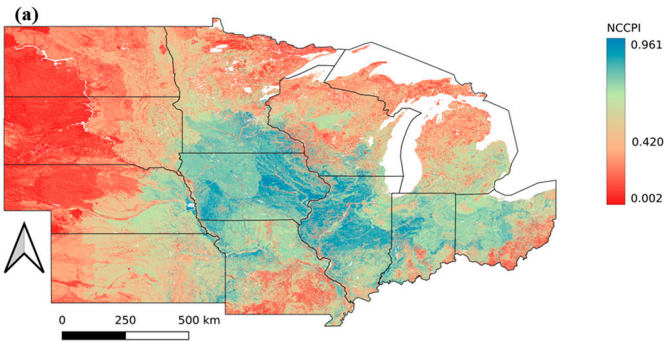


Figure 2. *Cont.*

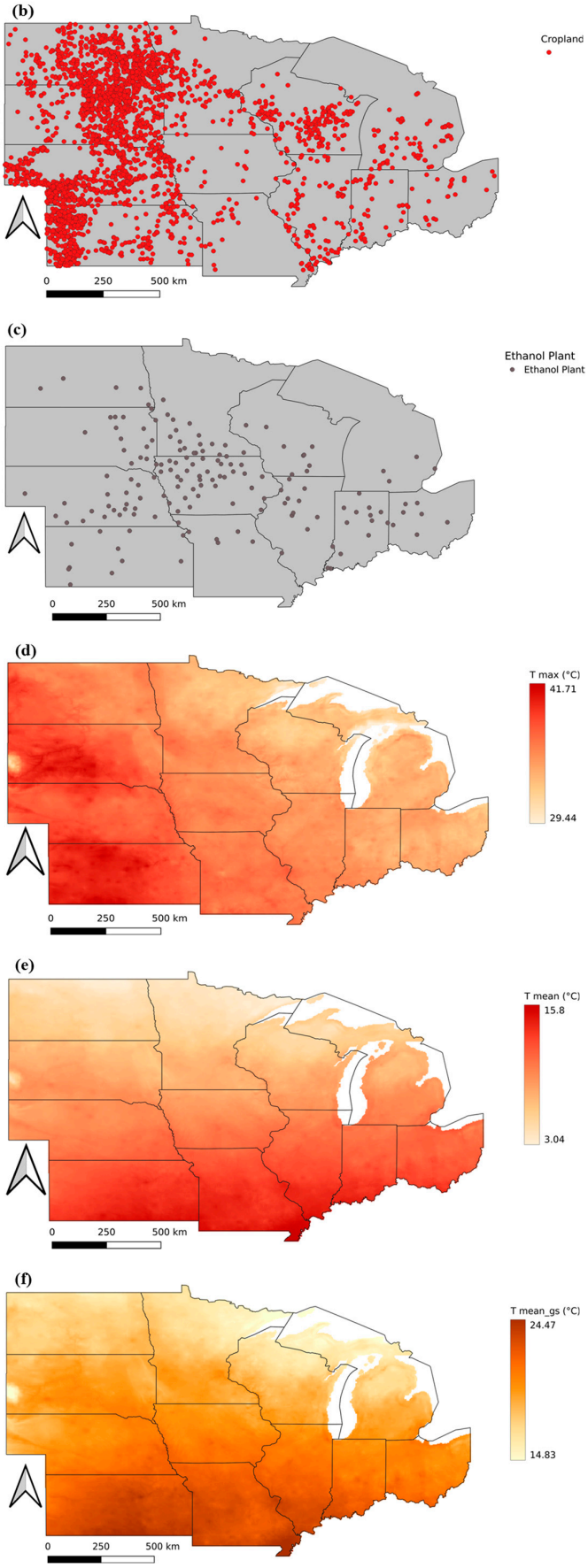


Figure 2. *Cont.*

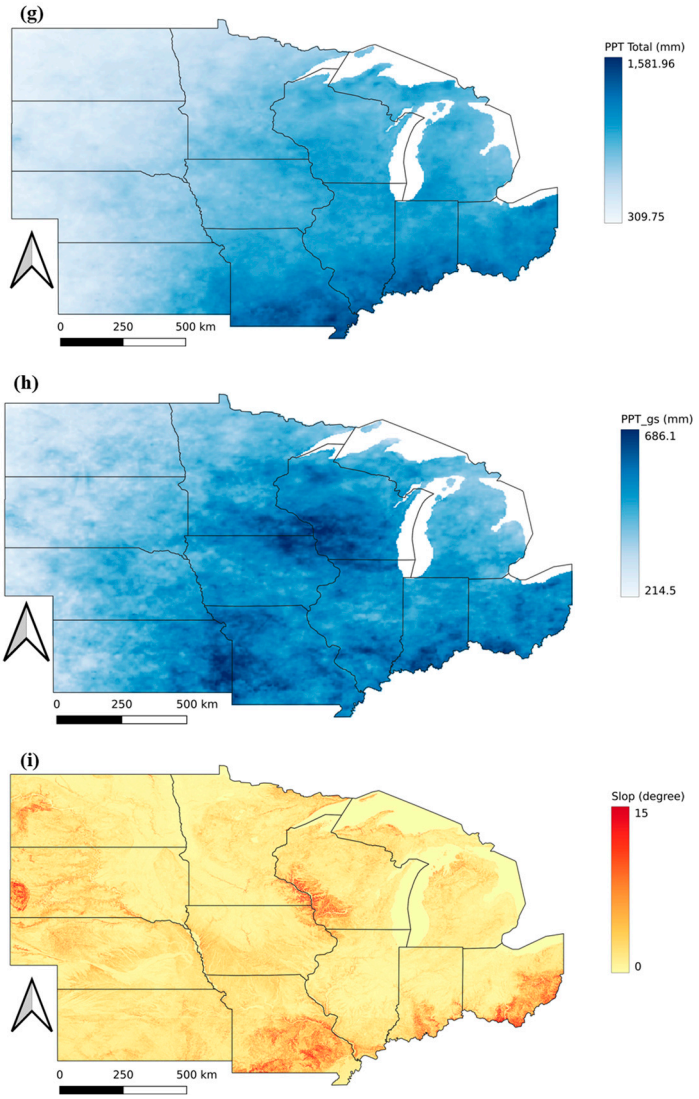


Figure 3. The spatial distribution of inputs: (**a**) NCCIP, (**b**) cropland with 3000 random samples in 2016, (**c**) ethanol plant in 2024, (**d**) max temperature in 2016, (**e**) mean temperature in 2016, (**f**) mean temperature during growth season in 2016, (**g**) total annual precipitation in 2016, (**h**) precipitation during growth season in 2016, and (**i**) slope.

The point-biserial correlation coefficient (r_{pb}) measures the strength and direction of the association between a binary variable and a continuous variable [47] (Equation (6)). In this study, this method is used to conduct a preliminary assessment of the correlation between each input variable and the binary response prior to model fitting (Table 1). This simple correlation analysis helps to identify potentially informative predictors for subsequent modelling.

$$r_{pb} = \frac{\overline{X}_1 - \overline{X}_0}{s_X} \sqrt{\frac{n_1 n_0}{n^2}} \tag{6}$$

where $\overline{X}_1$ and $\overline{X}_0$ are the means of the continuous variable for the binary groups (1 and 0), s_X is the standard deviation of the continuous variable, n_1 and n_0 are the sample sizes for each group, and n = n_1+n_0. r_{pb} indicates the strength and direction of association between the binary and continuous variables. A positive value suggests that the continuous variable is, on average, higher when the binary variable equals one, while a negative value indicates the opposite. The closer the absolute value of r_{pb} is to 1, the stronger the association; values near zero indicate weak or no association.

Table 1. Input selection.

Category	Period	Source	Inputs	$r_p b$
Corn mask	2000–2004, 2006, 2014–2016, 2022–2024	USDA-NASS Cropland Data Layer	Cropland	N/A
Climate	1999–2003, 2005, 2015–2024	gridMET	PPT	−0.0777
			PPT_gs	−0.0346
			T max	0.1565
			T mean	0.0823
			T mean_gs	0.1430
GIS	N/A	USGS DEM	Slop	−0.2174
	N/A	SSURGO	NCCPI	0.2700
	2024	EIA	Ethanol	−0.2298

3. Results

3.1. Model Validation

The performance of the Bayesian logistic regression model was evaluated using data from the pre-RFS period (2000–2006, excluding 2005). The model exhibited a moderate level of discriminatory ability, with an AUC of 0.75 (Figure 4a). This suggests that the model was able to distinguish between cultivated and uncultivated marginal lands with reasonable accuracy, though not without limitations.

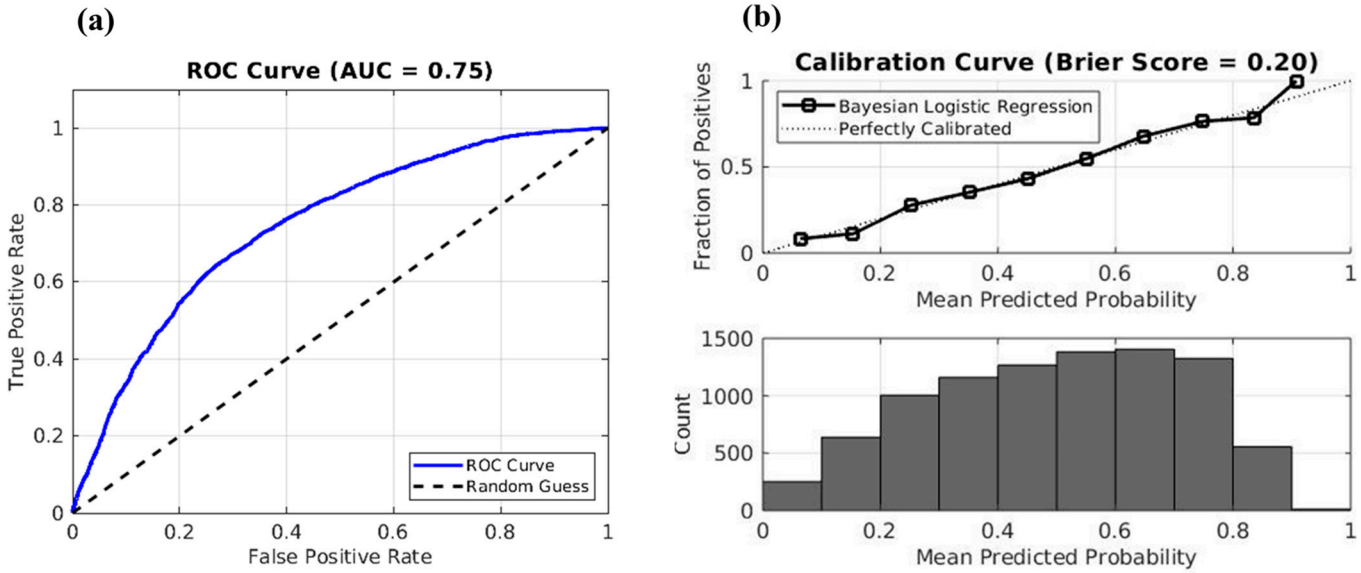


Figure 4. Results of evaluation. (**a**) ROC curve with AUC = 0.75, (**b**) calibrated curve with Brier Score = 0.20. They indicate a moderate positive correlation between simulated probabilities and observed land-use outcomes.

Model calibration was further assessed through a reliability curve and Brier Score. The calibration plot (Figure 4b, upper panel) indicated a close alignment between predicted probabilities and observed frequencies, particularly in the mid-to-high probability range, while some underestimation was observed at lower predicted values. The Brier Score of 0.20 reflects moderate overall probabilistic accuracy. The histogram of predicted probabilities (Figure 4b, lower panel) showed a uniform distribution, suggesting the model did not over-concentrate predictions in extreme probability bins.

3.2. Counterfactual Prediction of Cultivation Probability in 2016

The application of the Bayesian logistic regression model to the 2016 data yielded an AUC of 0.79, reflecting strong discriminatory ability in distinguishing between cultivated and uncultivated pixels on marginal land (Figure 5). The calibration curve (Brier Score = 0.19) indicates that predicted probabilities are generally well aligned with observed frequencies, and the distribution of predicted values is broadly uniform across probability bins.

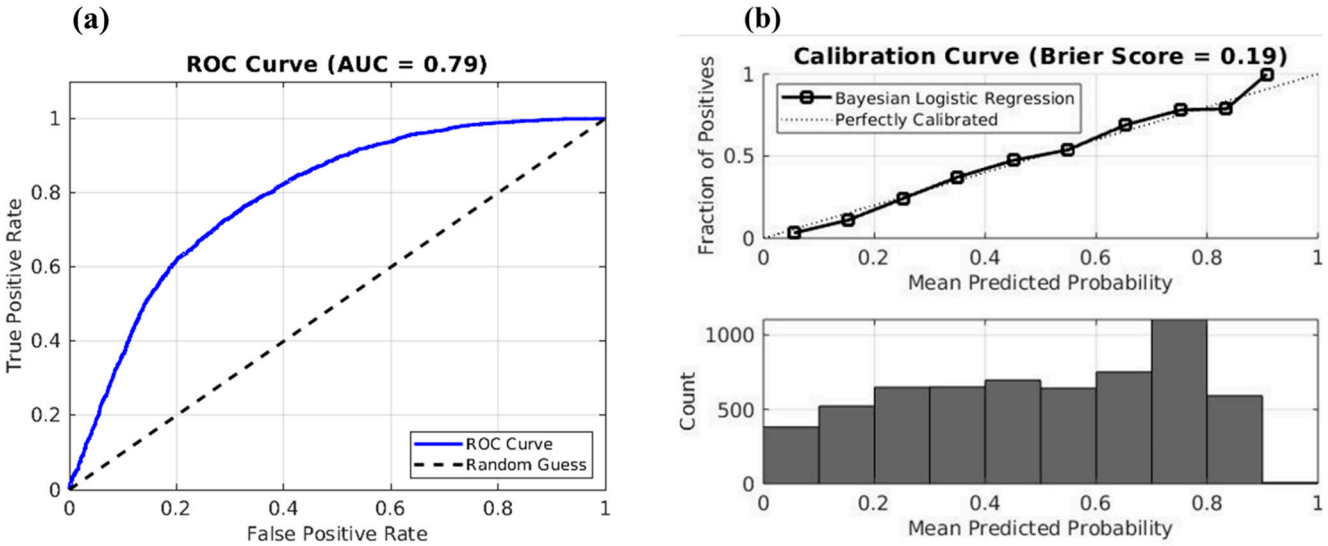


Figure 5. Model performance in 2016. (**a**) ROC curve with AUC = 0.79, (**b**) calibration curve with Brier Score = 0.19.

The counterfactual predictions of corn cultivation in 2016 without the influence of the RFS policy are illustrated in Figure 6. These predictions represent the likelihood that marginal and sub-marginal land areas would have been cultivated purely due to their biophysical suitability and climatic conditions, independent of RFS-induced market incentives.

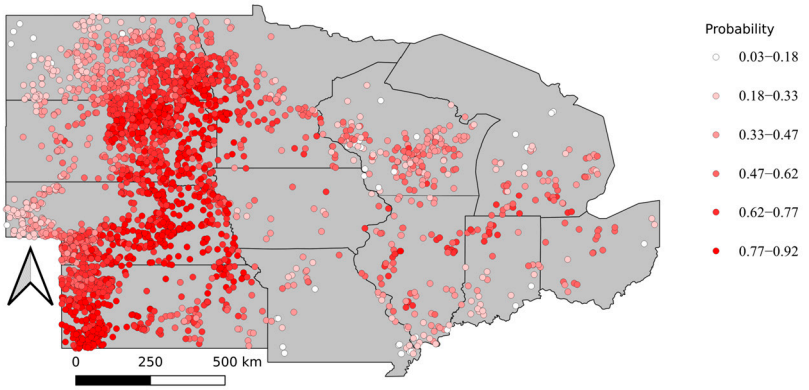


Figure 6. The probabilities of planting in 2016 without RFS.

As shown in Figure 6, the spatial distribution of the predicted probabilities exhibits considerable variability across the U.S. Corn Belt. Regions with the highest predicted probabilities (0.77–0.92) primarily cluster in the central and southern portions of the Corn Belt, particularly concentrated around Nebraska, Iowa, and southern Minnesota. In contrast, lower predicted probabilities (below 0.5) are predominantly found at the peripheries, including northern and western fringe regions characterised by relatively poor soil productivity (low NCCPI), higher slopes, or less favourable climatic conditions.

Quantitative analysis indicates that 26.67% of pixels cultivated in 2016 were assigned predicted probabilities below the 0.5 threshold.

Table 2 presents the posterior distributions of the regression coefficients, providing further insights into the drivers of cultivation decisions under baseline conditions. Ethanol plant proximity exhibited a strong negative association with predicted cultivation likelihood (mean = −0.643; 95% CrI: −0.688 to −0.597), consistent with expectations that greater distance from ethanol plants would reduce cultivation incentives. Climatic variables demonstrated complex relationships with baseline cultivation probabilities. Total annual precipitation (ppt_total) showed a strong negative association (mean = −1.291; 95% CrI: −1.408 to −1.174), suggesting excessive precipitation may limit corn cultivation due to drainage and operational constraints. Conversely, growing season precipitation (ppt_gs) exhibited a positive relationship (mean = 0.708; 95% CrI: 0.613 to 0.801), underlining the critical role of adequate rainfall during crop development periods.

Table 2. Posterior distribution of coefficients for the phase 1 model.

Inputs	Mean	SD	2.5% CrI	97.5% CrI	$\hat{R}$
NCCPI	0.376	0.017	0.341	0.408	1.0
ppt_gs	0.708	0.048	0.613	0.801	1.0
ppt_total	−1.291	0.059	−1.408	−1.174	1.0
tmax	−0.048	0.027	−0.104	0.003	1.0
tmean	0.667	0.069	0.539	0.802	1.0
tmean_gs	−0.244	0.065	−0.370	−0.117	1.0
slope	−0.442	0.019	−0.479	−0.404	1.0
ethanol_distance	−0.643	0.023	−0.688	−0.597	1.0

Temperature variables presented similarly complex patterns. Annual mean temperature (tmean) had a clearly positive effect on the likelihood of corn cultivation (mean = 0.667; 95% CrI: 0.539 to 0.802), reflecting the climatic suitability within optimal corn-growing temperature ranges. However, mean growing season temperature (tmean_gs) had a negative impact (mean = −0.244; 95% CrI: −0.370 to −0.117), indicating that overly high temperatures during key developmental stages may adversely affect corn productivity, thus reducing cultivation probabilities on marginal lands.

Furthermore, soil productivity as measured by the NCCPI was positively related to cultivation probability (mean = 0.376; 95% CrI: 0.341 to 0.408), reaffirming that, even within marginal areas, relative improvements in soil quality significantly influence cultivation likelihood. Slope showed a negative relationship (mean = −0.442; 95% CrI: −0.479 to −0.404), confirming that steeper terrain remains a critical deterrent to corn expansion under baseline conditions.

3.3. Phase 2: Analysis of Complete Cropland Abandonment

In the second phase of the analysis, attention was directed toward the fate of marginal lands that had been newly cultivated during the RFS-induced expansion but subsequently experienced complete abandonment of corn cultivation. Land pixels were classified as abandoned if they were cultivated in any year between 2014 and 2016 but remained uncultivated in each year from 2022 to 2024. These abandoned pixels were designated as the positive class, while those with continuous cultivation throughout the entire period formed the negative class.

The performance of the Bayesian logistic regression model in predicting abandonment risk is illustrated in Figure 7. The ROC curve for the test set yields an AUC of 0.87. This is mainly due to the strong contrast in key variables—particularly ethanol plant distance, temperature, and precipitation—between abandoned and retained cropland. These re-

lationships allow the Bayesian logistic regression model to correctly classify most cases, demonstrating genuine predictive skill rather than random separation. The calibration curve (Brier Score = 0.14) further demonstrates that predicted probabilities closely match observed frequencies, supporting the reliability of the model for probabilistic risk assessment.

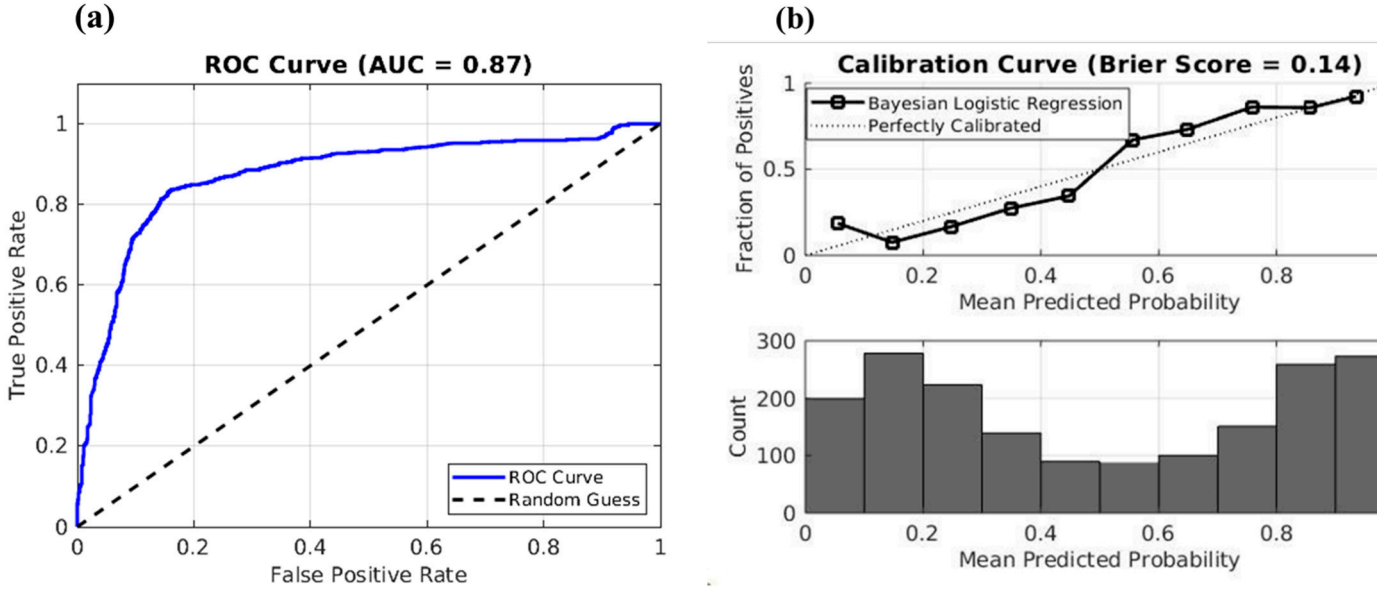


Figure 7. Model performance of phase 2. (**a**) ROC curve with AUC = 0.87, (**b**) calibration curve with Brier Score = 0.14. The results show strong separation from the random-guess line, indicating excellent discrimination between abandoned and retained cropland.

Spatial analysis revealed that the distribution of abandoned cropland was highly uneven across the study area. As depicted in Figure 8a, most abandonment events occurred in the northwestern and peripheral regions of the Corn Belt, where marginal soils, steeper slopes, and less favourable climate conditions predominate. In contrast, retained cultivation sites were more prevalent in central zones with higher baseline suitability. The mapped distribution of predicted abandonment probabilities (Figure 8b) corroborates these findings, highlighting a concentration of high-risk areas in locations with persistent biophysical and economic constraints.

The posterior parameter estimates (Table 3) provide further insight into the drivers of abandonment risk. Higher mean annual temperatures (tmean) and greater distances from ethanol plants (ethanol_distance) were strongly associated with increased probability of abandonment, reflecting both climatic vulnerability and reduced market access. Conversely, greater annual precipitation (ppt_total) and higher soil productivity (NCCPI) reduced the risk of abandonment, consistent with the fundamental role of these factors in supporting sustainable corn production. The effects of growing season climate variables were more complex, with evidence that excessive temperatures or inadequate precipitation during the crop development window increased vulnerability to abandonment.

Overall, these results demonstrate that while the RFS policy initially drove corn expansion onto marginal lands, the long-term retention of such cultivation is highly contingent on local environmental and market conditions.

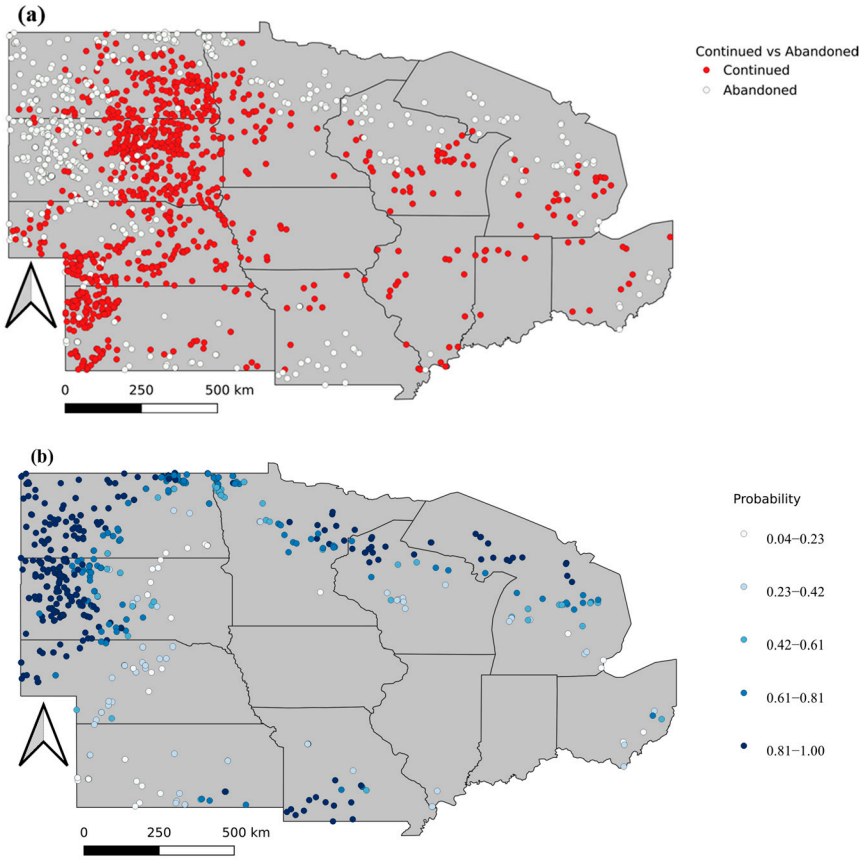


Figure 8. (**a**) Continued vs. abandoned land, (**b**) probabilities of abandoned land map.

Table 3. Posterior distribution of coefficients for the phase 2 model.

Inputs	Mean	SD	2.5% CrI	97.5% CrI	$\hat{R}$
NCCPI	−0.398	0.046	−0.488	−0.313	1.0
ppt_gs	−0.326	0.111	−0.529	−0.113	1.0
ppt_total	1.324	0.148	1.041	1.593	1.0
slope	0.112	0.044	0.031	0.198	1.0
tmax	0.700	0.119	0.486	0.938	1.0
tmean	−1.982	0.199	−2.363	−1.623	1.0
tmean_gs	0.734	0.224	0.338	1.177	1.0
ethanol_distance	1.663	0.078	1.509	1.798	1.0

4. Discussion and Limitation

The results highlight both the promise and inherent limitations of probabilistic modelling for quantifying policy-driven land-use change. While the Bayesian logistic regression framework effectively distinguishes between cultivated and uncultivated land pixels and provides interpretable probability estimates, several factors constrain the precision of policy attribution. The choice of Bayesian logistic regression reflects a deliberate balance between model interpretability, computational feasibility, and the need for formal uncertainty quantification. While non-Bayesian approaches such as random forest or conventional neural networks could achieve higher predictive accuracy, they lack explicit posterior inference, which is crucial for probabilistic counterfactual evaluation. Future work could explore Bayesian neural networks or ensemble probabilistic models to further improve predictive performance while maintaining transparency and uncertainty awareness.

While previous studies have estimated cropland expansion due to biofuel production, ranging from 0.01 to 2.45 million acres per billion gallons [16], our probabilistic counterfac-

tual analysis provides a spatially explicit estimate of marginal land conversion under RFS incentives. Unlike deterministic models, our approach quantifies uncertainty and reveals that over 26% of marginal land cultivated in 2016 would likely not have been planted without policy intervention. These findings align with life-cycle assessments showing elevated carbon intensity and environmental degradation [15,17], but offer a more granular understanding of where and why such impacts occur. These environmental outcomes align with the well-documented impacts of agricultural expansion on marginal or ecologically sensitive lands, thereby offering indirect empirical support for the finding that the RFS policy contributed to the conversion of substantial areas of marginal land into corn cultivation.

A key limitation arises from the composition of the training dataset. The inclusion of agronomically suitable but historically uncultivated marginal lands in both positive and negative samples introduces ambiguity in the conceptualisation of baseline suitability. Therefore, the predicted probabilities for the 2016 test set may be systematically deflated, particularly for pixels at the suitability threshold. This ambiguity is reflected in the finding that only 26% of the 2016 cultivated marginal land pixels were classified as unlikely to be planted under no-policy conditions when a threshold of 0.5 was used. While this offers a conservative estimate of RFS-induced expansion, it likely underestimates the true policy impact. Adopting a higher probability threshold (e.g., 0.7 or 0.8) may provide a more realistic benchmark for identifying lands that would otherwise have remained uncultivated. However, there is currently no universally accepted criterion for selecting an optimal threshold, and such decisions inherently involve normative assumptions rather than purely statistical evidence. For this reason, the present study reports results based on the conventional 0.5 threshold while acknowledging that future work should incorporate sensitivity analysis or decision-theoretic approaches to formalise threshold selection.

Another important limitation concerns the challenge of endogeneity and omitted variable bias, particularly regarding factors such as ethanol plant location and evolving local market conditions. While spatial and biophysical controls were included, unobserved variables may still influence cultivation decisions, complicating strict causal interpretation.

Furthermore, the transferability of the model to other regions or time periods may be limited by changing agronomic practices, technological advances, and evolving climate conditions, all of which could shift the underlying suitability frontier. Finally, the use of land pixel-level probability thresholds for policy attribution, while methodologically transparent, inevitably involves value judgements about what constitutes a "likely" versus "unlikely" planting event, which may vary depending on context and policy objectives.

Overall, while the modelling approach provides a robust, data-driven framework for estimating policy-driven land-use change, results should be interpreted as lower-bound estimates and with appropriate caution regarding threshold sensitivity and data limitations.

5. Conclusions

This study makes both theoretical and methodological contributions. Theoretically, it advances understanding of how biofuel policies influence land-use dynamics by quantifying the probability—rather than the certainty—of policy-induced cropland expansion and abandonment. This probabilistic perspective challenges deterministic assumptions and enriches the conceptual framework for analysing policy–environment interactions. Methodologically, the integration of Bayesian logistic regression with spatial counterfactual analysis offers a robust and transparent framework for causal attribution under uncertainty. This approach can be extended to other sustainability policies where spatial heterogeneity and uncertainty are critical. Empirical findings show strong model discrimination and calibration, confirming the robustness of the analytical framework. Counterfactual results

indicate that a substantial share of marginal lands cultivated after RFS implementation would likely not have been planted without policy incentives, underscoring the pivotal role of the RFS in driving cropland expansion onto less suitable areas. Sensitivity analysis further reveals that higher probability thresholds lead to greater estimated policy impacts. In the second phase, systematic patterns of post-expansion abandonment were identified, mainly associated with climatic disadvantage, poor soil productivity, and distance from ethanol markets. These results highlight that the sustainability of policy-driven land-use changes is highly contingent on local environmental and economic conditions.

From a policy perspective, the findings underscore the need for more spatially targeted and environmentally differentiated implementation of the RFS program. Policymakers could incorporate explicit land suitability thresholds into renewable fuel mandates to discourage cultivation on low-productivity or ecologically sensitive soils. Incentive mechanisms—such as RIN credits or biofuel subsidies—should be linked to environmental performance indicators, including soil carbon retention, erosion control, and nutrient management. In addition, promoting crop diversification and rotational practices on marginal lands can reduce long-term abandonment and maintain soil resilience. Finally, the development of spatially explicit monitoring systems based on remote sensing can help evaluate compliance and detect unintended land-use changes in real time.

Despite its strengths, the framework remains subject to limitations such as potential endogeneity, threshold sensitivity, and the evolving nature of agronomic and market dynamics. Nevertheless, the study contributes new empirical evidence to the debate on biofuel policy impacts and demonstrates the value of combining probabilistic counterfactual modelling with high-resolution spatial data for land-use policy assessment.

Author Contributions: Conceptualisation, S.L. and X.H.; methodology, S.L.; software, S.L.; validation, S.L. and X.H.; formal analysis, S.L.; investigation, S.L.; resources, S.L.; data curation, S.L.; writing—original draft preparation, S.L.; writing—review and editing, X.H.; visualisation, X.H.; supervision, X.H.; project administration, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AUC	Area Under the ROC Curve
DEM	Digital Elevation Model
EIA	The U.S. Energy Information Administration
GIS	Geographic Information System
gridMET	Gridded Surface Meteorological Dataset
ppt_gs	Precipitation during the growing season
ppt_total	Total annual precipitation
NASA	National Aeronautics and Space Administration
ROC	Receiver Operating Characteristic
$r_p b$	Point-biserial correlation coefficient

SSURGO	Soil Survey Geographic Database
tmean	Mean temperatures
tmean_gs	Mean growing season temperature
tmax	Maximum temperature
USGS	United States Geological Survey

References

1. Miller, J.; Clark, C.; Peterson, S.; Newes, E. Estimated attribution of the RFS Program on soybean biodiesel in the U.S. using the Bioenergy Scenario Model. *Energy Policy* **2024**, *192*, 114250. [CrossRef]
2. Aui, A.; Wang, Y. Post-RFS supports for cellulosic ethanol: Evaluation of economic and environmental impacts of alternative policies. *Energy Policy* **2022**, *170*, 113221. [CrossRef]
3. Kim, S.; Dale, B.E.; Zhang, X.; Jones, C.D.; Reddy, A.D.; Izaurralde, R.C. The Renewable Fuel Standard May Limit Overall Greenhouse Gas Savings by Corn Stover-Based Cellulosic Biofuels in the U.S. Midwest: Effects of the Regulatory Approach on Projected Emissions. *Environ. Sci. Technol.* **2019**, *53*, 2288–2294. [CrossRef] [PubMed]
4. Abulbasher, A.; Ulrich-Schad, J.D.; Kolady, D.; Wang, T.; Clay, D. Entrepreneurial Aspirations of South Dakota Commodity Crop Producers. *Sustainability* **2024**, *16*, 6839. [CrossRef]
5. Sesmero, J.; Sun, X. The influence of feedstock supply risk on location of stover-based bio-gasoline plants. *GCB Bioenergy* **2015**, *8*, 495–508. [CrossRef]
6. Green, T.R.; Kipka, H.; David, O.; McMaster, G.S. Where is the USA Corn Belt, and how is it changing? *Sci. Total Environ.* **2018**, *618*, 1613–1618. [CrossRef]
7. Traldi, R.; Asprooth, L.; Usher, E.M.; Floress, K.; Arbuckle, J.G.; Baskerville, M.; Church, S.P.; Genskow, K.; Harden, S.; Maynard, E.T.; et al. "Safer to plant corn and beans"? Navigating the challenges and opportunities of agricultural diversification in the U.S. Corn Belt. *Agric. Hum. Values* **2024**, *41*, 1687–1706. [CrossRef]
8. Atwell, R.C.; Schulte, L.A.; Westphal, L.M. How to build multifunctional agricultural landscapes in the U.S. Corn Belt: Add perennials and partnerships. *Land Use Policy* **2010**, *27*, 1082–1090. [CrossRef]
9. Shao, Y.; Taff, G.N.; Ren, J.; Campbell, J.B. Characterizing major agricultural land change trends in the Western Corn Belt. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 116–125. [CrossRef]
10. Korting, C.; Just, D.R. Demystifying RINs: A partial equilibrium model of U.S. biofuel markets. *Energy Econ.* **2017**, *64*, 353–362. [CrossRef]
11. Ebadian, M.; van Dyk, S.; McMillan, J.D.; Saddler, J. Biofuels policies that have encouraged their production and use: An international perspective. *Energy Policy* **2020**, *147*, 111906. [CrossRef]
12. Taheripour, F.; Baumes, H.; Tyner, W.E. Economic Impacts of the U.S. Renewable Fuel Standard: An Ex-Post Evaluation. *Front. Energy Res.* **2022**, *10*, 749738. [CrossRef]
13. Moschini, G.; Lapan, H.; Kim, H. The Renewable Fuel Standard in Competitive Equilibrium: Market and Welfare Effects. *Am. J. Agric. Econ.* **2017**, *99*, 1117–1142. [CrossRef]
14. Roesch-McNally, G.E.; Arbuckle, J.G.; Tyndall, J.C. Barriers to implementing climate resilient agricultural strategies: The case of crop diversification in the U.S. Corn Belt. *Glob. Environ. Change* **2018**, *48*, 206–215. [CrossRef]
15. Zhang, X.; Lark, T.J.; Clark, C.; Yuan, Y.; LeDuc, S.D. Grassland-to-cropland conversion increased soil, nutrient, and carbon losses in the US Midwest between 2008 and 2016. *Environ. Res. Lett.* **2021**, *16*, 054018. [CrossRef]
16. Austin, K.G.; Jones, J.P.H.; Clark, C.M. A review of domestic land use change attributable to U.S. biofuel policy. *Renew. Sustain. Energy Rev.* **2022**, *159*, 112181. [CrossRef] [PubMed]
17. Lark, T.J.; Hendricks, N.P.; Smith, A.; Pates, N.; Spawn-Lee, S.A.; Bougie, M.; Booth, E.G.; Kucharik, C.J.; Gibbs, H.K. Environmental outcomes of the US Renewable Fuel Standard. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2101084119. [CrossRef]
18. Copenhaver, K.; Mueller, S. Considering Historical Land Use When Estimating Soil Carbon Stock Changes of Transitional Croplands. *Sustainability* **2024**, *16*, 734. [CrossRef]
19. Brock, C.; Jackson-Smith, D.; Kumarappan, S.; Culman, S.; Herms, C.; Doohan, D. Organic Corn Production Practices and Profitability in the Eastern U.S. Corn Belt. *Sustainability* **2021**, *13*, 8682. [CrossRef]
20. Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [CrossRef]
21. Deines, J.M.; Swatantran, A.; Ye, D.; Myers, B.; Archontoulis, S.; Lobell, D.B. Field-scale dynamics of planting dates in the US Corn Belt from 2000 to 2020. *Remote Sens. Environ.* **2023**, *291*, 113551. [CrossRef]

22. Hartman, T.; Cirone, R.; Togliatti, K.; Hornbuckle, B.K.; VanLoocke, A. A spatial and temporal evaluation of the SMAP cropland b-parameter across the U.S. Corn Belt. *Remote Sens. Environ.* **2023**, *297*, 113752. [CrossRef]
23. Walker, V.A.; Hornbuckle, B.K.; Cosh, M.H.; Prueger, J.H. Seasonal Evaluation of SMAP Soil Moisture in the U.S. Corn Belt. *Remote Sens.* **2019**, *11*, 2488. [CrossRef]
24. Rodriguez-Alvarez, N.; Misra, S.; Morris, M. The Polarimetric Sensitivity of SMAP-Reflectometry Signals to Crop Growth in the U.S. Corn Belt. *Remote Sens.* **2020**, *12*, 1007. [CrossRef]
25. Wimberly, M.C.; Janssen, L.L.; Hennessy, D.A.; Luri, M.; Chowdhury, N.M.; Feng, H. Cropland expansion and grassland loss in the eastern Dakotas: New insights from a farm-level survey. *Land Use Policy* **2017**, *63*, 160–173. [CrossRef]
26. Valjarević, A.; Morar, C.; Brasanac-Bosanac, L.; Cirkovic-Mitrovic, T.; Djekic, T.; Mihajlović, M.; Milevski, I.; Culafic, G.; Luković, M.; Niemets, L.; et al. Sustainable land use in Moldova: GIS & remote sensing of forests and crops. *Land Use Policy* **2025**, *152*, 107515. [CrossRef]
27. Togliatti, K.; Hartman, T.; Walker, V.A.; Arkebauer, T.J.; Suyker, A.E.; VanLoocke, A.; Hornbuckle, B.K. Satellite L–band vegetation optical depth is directly proportional to crop water in the US Corn Belt. *Remote Sens. Environ.* **2019**, *233*, 111378. [CrossRef]
28. Zhang, H. Effects of Soybean–Corn Rotation on Crop Yield, Economic Benefits, and Water Productivity in the Corn Belt of Northeast China. *Sustainability* **2023**, *15*, 11362. [CrossRef]
29. Bandaru, V.; Izaurralde, R.C.; Manowitz, D.; Link, R.; Zhang, X.; Post, W.M. Soil carbon change and net energy associated with biofuel production on marginal lands: A regional modeling perspective. *J. Environ. Qual.* **2013**, *42*, 1802–1814. [CrossRef]
30. Lark, T.J.; Spawn, S.A.; Bougie, M.; Gibbs, H.K. Cropland expansion in the United States produces marginal yields at high costs to wildlife. *Nat. Commun.* **2020**, *11*, 4295. [CrossRef]
31. Mosia, M. A Bayesian State-Space Approach to Dynamic Hierarchical Logistic Regression for Evolving Student Risk in Educational Analytics. *Data* **2025**, *10*, 23. [CrossRef]
32. Allen, G.I.; Gan, L.; Zheng, L. Interpretable Machine Learning for Discovery: Statistical Challenges and Opportunities. *Annu. Rev. Stat. Its Appl.* **2024**, *11*, 97–121. [CrossRef]
33. Yu, C.-H.; Wang, S. A Comparative Study of Bayesian Neural Networks and Machine Learning Based on COVID-19 Image Classification. *Stat. Data Sci. Imaging* **2025**, *2*, 2497555. [CrossRef]
34. Gelman, A.; Jakulin, A.; Pittau, M.G.; Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2008**, *2*, 1360–1383. [CrossRef]
35. Rahman, E.; Rao, P.; Nassif, A.D.; Webb, W.R.; Wu, W.T.L.; Carruthers, J.D.A. Beyond binary endpoints: Modernizing aesthetic trial analytics: Simulation-Based reappraisal of phase 2/3 BoNT-A studies using advanced statistical models. *Eur. J. Plast. Surg.* **2025**, *48*, 70. [CrossRef]
36. Cunha, M.C.; Zeferino, J.A.; Simões, N.E.; Saldarriaga, J.G. Optimal location and sizing of storage units in a drainage system. *Environ. Model. Softw.* **2016**, *83*, 155–166. [CrossRef]
37. Miao, J.; Pan, Y.; Zhao, K. Analysis of the Spatiotemporal Effects on the Severity of Motorcycle Accidents Without Helmets and Strategies for Building Sustainable Traffic Safety. *Sustainability* **2025**, *17*, 3280. [CrossRef]
38. Lee, C.-Y.; Lee, M.-K. Demand Forecasting in the Early Stage of the Technology's Life Cycle Using a Bayesian Update. *Sustainability* **2017**, *9*, 1378. [CrossRef]
39. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
40. Prishchepov, A.V.; Ponkina, E.; Sun, Z.; Müller, D. Revealing the determinants of wheat yields in the Siberian breadbasket of Russia with Bayesian networks. *Land. Use Policy* **2019**, *80*, 21–31. [CrossRef]
41. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* **2021**, *16*, 667–718. [CrossRef]
42. Bertoni, D.; Aletti, G.; Ferrandi, G.; Micheletti, A.; Cavicchioli, D.; Pretolani, R. Farmland Use Transitions After the CAP Greening: A Preliminary Analysis Using Markov Chains Approach. *Land Use Policy* **2018**, *79*, 789–800. [CrossRef]
43. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
44. Saeidi, S.; Mohammadzadeh, M.; Salmanmahiny, A.; Mirkarimi, S.H. Performance evaluation of multiple methods for landscape aesthetic suitability mapping: A comparative study between Multi-Criteria Evaluation, Logistic Regression and Multi-Layer Perceptron neural network. *Land Use Policy* **2017**, *67*, 1–12. [CrossRef]
45. Tenenbaum, J.; Williams, P.D.; Turp, D.; Buchanan, P.; Coulson, R.; Gill, P.G.; Lunnon, R.W.; Oztunali, M.G.; Rankin, J.; Rukhovets, L. Aircraft observations and reanalysis depictions of trends in the North Atlantic winter jet stream wind speeds and turbulence. *Q. J. R. Meteorol. Soc.* **2022**, *148*, 2927–2941. [CrossRef]

46. Kang, S.; Post, W.; Wang, D.; Nichols, J.; Bandaru, V.; West, T. Hierarchical marginal land assessment for land use planning. *Land Use Policy* **2013**, *30*, 106–113. [CrossRef]
47. Pushkareva, E.; Baumann, K.; Van, A.T.; Mikhailyuk, T.; Baum, C.; Hrynkiewicz, K.; Demchenko, E.; Thiem, D.; Köpcke, T.; Karsten, U.; et al. Diversity of microbial phototrophs and heterotrophs in Icelandic biocrusts and their role in phosphorus-rich Andosols. *Geoderma* **2021**, *386*, 114905. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.