SOCCJ Soil & Water Research Summary

Xunhang Gao

October 2025

1 Introduction & Data

The survey is taken once a year from 2020 to 2025. And there are 1858 observations, individuals, and each observation contains 1402 variables. The individuals are consistent across all years, while certain individuals are not observed in some years. And we focus on the variable recording whether using a method of production, Cover crops, for each individual in each year, leveled from 1 to 4, with 1 as not used last year and will not plan to use it, 2 as not used last year and not planning to use it within 3 years but open to use in future, 3 as not used last year but intend to use within 3 years, and 4 as already used last year. The data is blocked into 3 groups (time period), (2020, 2021), (2022, 2023), (2024, 2025). And we may assume data from different year blocks share some of the variables, that is part of the variables from year 2020, 2022, and 2024 are shared, but some are different. This is similar for year 2021, 2023, and 2025. The “cover crops”, individual choices of method of production, is recorded in year 2021, 2023, 2025. And there is a group of other substitution methods of production also recorded as same 4 level variables in year 2021, 2023, 2025 together with “cover crops”.

1.1 Scientific Questions

For this project, we focus on answering following questions:

(1) Which factors contributed to the individuals’ choice of using “cover crops” from year (2020, 2021) to (2022, 2023);

(2) Which factors contributed to the individuals’ choice of using “cover crops” from year (2022, 2023) to (2024, 2025);

After the investigation of the scientific questions (1) and (2) and based on the result we have from question (1) from year 2021 to 2023, we have an expectation, prediction, of the expected proportion of usage of “cover crops” in year 2025 given information in year block (2022, 2023), but the actual proportion of usage of “cover crops” in year 2025 is different from our expectation, prediction. Thus, the second part of scientific questions is

(3) what factors contribute to the differences between our expected, predicted, proportion of usage of “cover crops” in year 2025 and the actual proportion of usage of “cover crops” in year 2025.

2 Notation & Data Processing

- Individuals (i=1,. . . ,N) (about 1000).
- Years ($t \in 2020, \ldots, 2025$).
- Blocks:
  - Block 1: (2020, 2021)
  - Block 2: (2022, 2023)

– Block 2: (2024, 2025)

- Ordered adoption state of cover crops (4 levels) in year ($t \in 2021, 2023, 2025$):

$$Y_{it} \in 1, 2, 3, 4$$

(1 = no use & no plan, …, 4 = already used last year).

- Let

$$S_i^{(1)} = Y_{i,2021}, \quad Y_i^{(1)} = Y_{i,2023}, \quad S_i^{(2)} = Y_{i,2023}, \quad Y_i^{(2)} = Y_{i,2025}$$

be the "start" and "end" states for blocks 1→2 and 2→3.

- Covariates:
  – (X_{it}^{even}) for even years (2020, 2022, 2024) – "production / structural" covariates.
  – (X_{it}^{odd}) for odd years (2021, 2023, 2025) – attitudes, plans, etc.

  Some of these are shared across blocks (same conceptual variable re-asked).

- Other methods of production (substitutes) are also 4-level variables ($Z_{it}^{(k)}$).

- Panel attrition & refreshment → nonresponse indicators (R_{it}) and design weights (d_{it}). Ultimately we'll work with analysis weights (w_{it}) that combine design + nonresponse adjustments (IPCW + imputation based on PMM).

- PCA was applied to a selected subset of survey variables to reduce dimensionality and summarize their joint variation into a smaller number of orthogonal components. The number of components retained was determined using parallel analysis, ensuring that only components explaining more variance than expected under random noise were kept. Then, we use summation scales of variables grouped based on loadings as low-dimensional representations of the original variables for further analysis.

3 Analytical Framework

3.1 Step A. Construct transition datasets for the two periods

- Dataset A (2021 → 2023) Covariates:

$$\mathcal{X}_i^{(1)} = \big(X_{i,2020}, X_{i,2021}, X_{i,2022}, X_{i,2023}, Z_{\cdot,2021}, Z_{\cdot,2023}\big),$$

possibly with engineered lags/changes (e.g., ($X_{2023} - X_{2021}$)).

- Dataset B (2023 → 2025) Covariates:

$$\mathcal{X}_i^{(2)} = \big(X_{i,2022}, X_{i,2023}, X_{i,2024}, X_{i,2025}, Z_{\cdot,2023}, Z_{\cdot,2025}\big),$$

remembering that some 2025 questions refer to 2024.

Keep weights ($w_i^{(1)}$) and ($w_i^{(2)}$) to account for design + attrition.

3.2 Step B. Q(1) & (2): model "transition behavior" and drivers

Objective: for each block (b=1,2), estimate

- Transition kernel

$$p^{(b)}_{k\to m}(x) = \Pr\big(Y^{(b)} = m, \big|, S^{(b)} = k,\ \mathcal{X}^{(b)} = x\big),$$

- upward-movement risk

$$u^{(b)}_{k}(x) = \Pr\big(Y^{(b)} > S^{(b)}, \big|, S^{(b)} = k,\ \mathcal{X}^{(b)} = x\big).$$

From these we can:

- describe transitions (Sankey-style) conditional on covariates;
- identify factors associated with more upward movement (from SHAP, partial dependence, or regression coefficients);
- optionally embed this in a causal estimator later (DML/TMLE) for selected modifiable factors.

3.3 Step C. Q(3): expected vs realized 2025 adoption

The analytical framework of this section is currently under development and has not yet been decided.

We want to compare:

- Expected 2025 adoption using only information up to block 2, what we would have predicted for 2025 if the world in 2024–2025 behaved like 2020–2023.
- Actual 2025 adoption with all later changes, including new covariate values in 2024–2025.

Possible set up:

- Fit a model for transition structure using block 1→2 (dataset A):

$$g^{(1)}(s, x_{\text{pre}}) \approx \mathbb{E}\big[Y^{(2)} \mid S^{(2)} = s,\ \mathcal{X}_{\le 2023} = x_{\text{pre}}\big]$$

  trained only on variables available by 2023.

- Apply $(g^{(1)})$ to the 2023 population (with 2022 and 2023 covariates) to obtain a predicted distribution $(\hat{Y}^{\text{pred}}_{2025})$ and its proportion $(\hat{\pi}^{\text{pred}}_{2025})$.
- Compare with realized 2025 adoption proportion $(\hat{\pi}^{\text{obs}}_{2025})$ (using directly 2025 data).

Define the gap

$$\Delta = \hat{\pi}^{\text{obs}}_{2025} - \hat{\pi}^{\text{pred}}_{2025}.$$

Fit a full model $(g^{(2)})$ on block 2→3 with more covariates $(\mathcal{X}^{(2)})$ (including 2024/25 covariates), then investigate

$$d_i = g^{(2)}(S^{(2)}_i, \mathcal{X}^{(2)}_i) - g^{(1)}(S^{(2)}_i, \mathcal{X}^{\text{pre}}_i)$$

explained by new covariates.

4 Theoretical Models & Estimands

4.1 Transition model for Q(1) & Q(2)

4.1.1 Multiclass transition model (focusing on transition kernel)

For each block $b \in \{1,2\}$ and individual $i \in \{1,\dots,N\}$, we observe an origin state $S_i^{(b)} \in \{1,2,3,4\}$ and a destination state $Y_i^{(b)} \in \{1,2,3,4\}$, together with a covariate vector $\mathcal{X}_i^{(b)}$. We model the transition probabilities $p_{k\to m}^{(b)}(x) = \Pr(Y^{(b)} = m \mid S^{(b)} = k, \mathcal{X}^{(b)} = x)$ using a multiclass gradient-boosted tree model.

We form a feature vector

$$x_i^{(b)} \equiv \big(S_i^{(b)},\ \mathcal{X}_i^{(b)}\big), \tag{1}$$

where $S_i^{(b)}$ is included as a categorical predictor (implemented via one-hot indicators in the model matrix; e.g., `Y_from1`, ..., `Y_from4`).

The model produces one real-valued *class margin* (score) for each destination level m:

$$f_m^{(b)}\big(x_i^{(b)}\big) \in \mathbb{R}, \qquad m = 1,2,3,4. \tag{2}$$

The predicted transition probabilities are obtained through the softmax link:

$$\hat{p}_{im}^{(b)} = \Pr\Big(Y_i^{(b)} = m \,\Big|\, x_i^{(b)}\Big) = \frac{\exp\big(f_m^{(b)}(x_i^{(b)})\big)}{\sum_{r=1}^4 \exp\big(f_r^{(b)}(x_i^{(b)})\big)}, \qquad m = 1,2,3,4. \tag{3}$$

4.1.2 From leveled responses to multiclass training targets

For training, the original state and observed destination state $Y_i^{(b)} \in \{1,2,3,4\}$ is represented by indicator variables $y_{im}^{(b)}$:

$$y_{im}^{(b)} = \mathbf{1}\{Y_i^{(b)} = m\}, \qquad m = 1,2,3,4. \tag{4}$$

Given the softmax probabilities $\hat{p}_{im}^{(b)}$, the likelihood contribution of observation i can be written as

$$\Pr\Big(Y_i^{(b)} \,\Big|\, x_i^{(b)}\Big) = \prod_{m=1}^{4} \Big(\hat{p}_{im}^{(b)}\Big)^{y_{im}^{(b)}}. \tag{5}$$

We use analysis weights $w_i^{(b)}$ (design weights with attrition/nonresponse adjustments) in model training and in all summary estimands.

4.1.3 Loss function for training (weighted multiclass cross-entropy)

Let $\hat{p}_{im}^{(b)}$ be the softmax probabilities implied by margins $f_m^{(b)}(x_i^{(b)})$. The weighted multiclass negative log-likelihood (cross-entropy) optimized by XGBoost is:

$$\mathcal{L}^{(b)} = -\sum_{i=1}^{n_b} w_i^{(b)} \sum_{m=1}^{4} y_{im}^{(b)} \log \hat{p}_{im}^{(b)}, \tag{6}$$

equivalently

$$\mathcal{L}^{(b)} = -\sum_{i=1}^{n_b} w_i^{(b)} \log \hat{p}_{i,Y_i^{(b)}}^{(b)}. \tag{7}$$

4.1.4 Transition matrix estimation from fitted probabilities

After fitting the Transition Boosting model for block b, we compute fitted probabilities $\hat{p}_{im}^{(b)}$ for each observation. The weighted (model-based) transition matrix estimator is obtained by averaging predicted probabilities within each origin level k:

$$\widehat{T}_{k\to m}^{(b)} = \frac{\sum_{i:\, S_i^{(b)}=k} w_i^{(b)}\, \hat{p}_{im}^{(b)}}{\sum_{i:\, S_i^{(b)}=k} w_i^{(b)}}, \qquad k, m \in \{1, 2, 3, 4\}. \tag{8}$$

This provides an estimate of the transition distribution in the weighted target population, as presented in Table 1 and Table 2.

4.1.5 Covariate ranking by gain (loss reduction attributed to splits)

We rank covariates using the *gain* importance from XGBoost. Conceptually, gain measures how much a covariate contributes to reducing the training objective through splits in the fitted tree ensemble.

For block b, the model is trained by minimizing the weighted multiclass cross-entropy loss

$$\mathcal{L}^{(b)} = -\sum_{i=1}^{n_b} w_i^{(b)} \log \hat{p}_{i,Y_i^{(b)}}^{(b)}, \tag{9}$$

where $\hat{p}_{i,Y_i^{(b)}}^{(b)}$ is the fitted probability assigned to the observed destination level $Y_i^{(b)}$ under the softmax model.

During training, XGBoost grows trees by selecting splits that decrease the loss. For a candidate split s, define its (training) loss reduction as

$$\Delta\mathcal{L}^{(b)}(s) \equiv \mathcal{L}_{\text{before split } s}^{(b)} - \mathcal{L}_{\text{after split } s}^{(b)}. \tag{10}$$

A split is beneficial when $\Delta\mathcal{L}^{(b)}(s) > 0$.

Let j denote a feature (a column) in the model matrix. The gain importance for feature j is defined as the sum of loss reductions over all splits in the fitted ensemble that use j:

$$\text{Gain}^{(b)}(j) = \sum_{s:\,\text{split } s \text{ uses feature } j} \Delta\mathcal{L}^{(b)}(s). \tag{11}$$

Because categorical covariates and year-suffixed covariates can expand into multiple model-matrix columns, we aggregate gain to a base covariate (denoted B). Let $j \in B$ indicate that column j belongs to base covariate B. The aggregated gain is

$$\text{Gain}^{(b)}(B) = \sum_{j\in B} \text{Gain}^{(b)}(j). \tag{12}$$

We rank covariates by $\text{Gain}^{(b)}(B)$ to identify which base covariates most strongly improve prediction of transition outcomes in the fitted Transition Boosting model, as presented in Table 3 and Table 4.

4.1.6 SHAP decomposition of margins and weighted signed summaries (shap_dir)

For a fixed block b (time block) and a fixed destination class m, TreeSHAP provides an additive decomposition of the fitted margin:

$$f_m^{(b)}(x_i^{(b)}) = \phi_{0m}^{(b)} + \sum_{j=1}^{p_b} \phi_{ijm}^{(b)}, \tag{13}$$

where $\phi_{0m}^{(b)}$ is a baseline margin and $\phi_{ijm}^{(b)}$ is the SHAP contribution of feature j for observation i and class m.

To summarize SHAP at the base-covariate level, we aggregate SHAP values across model-matrix columns belonging to the same base covariate c:

$$\Phi_{icm}^{(b)} = \sum_{j\in c} \phi_{ijm}^{(b)}. \tag{14}$$

To align interpretation with transition rows, we compute origin-stratified, weighted signed averages of SHAP contributions (denoted `shap_dir` in outputs):

$$\text{shap\_dir}^{(b)}_{k,m}(c) = \frac{\sum_{i:\, S_i^{(b)}=k} w_i^{(b)}\, \Phi^{(b)}_{icm}}{\sum_{i:\, S_i^{(b)}=k} w_i^{(b)}}, \qquad k, m \in \{1, 2, 3, 4\}. \tag{15}$$

The SHAP direction (signed SHAP) summaries are presented from Table 5 to Table 12.

4.1.7 Permutation impact tables on the probability scale

To present probability-scale directional effects aligned with the transition matrix, we compute permutation impact for each base covariate $c \in \mathcal{C}_b$. For a fixed origin level s, we permute the covariate block c within the stratum $\{i : S_i^{(b)} = s\}$ to break the association between c and the remaining covariates. Let π denote a random permutation over indices in this stratum, and define the permuted feature vector

$$x_i^{(-c)} \;=\; \big(x^{(b)}_{i,-c},\; x^{(b)}_{\pi(i),c}\big), \tag{16}$$

where $x^{(b)}_{i,-c}$ denotes all features except group c, and $x^{(b)}_{\pi(i),c}$ denotes the permuted value of group c.

For destination level m, define the (weighted) probability-impact estimand

$$\Delta^{\text{PI},(b)}_{s\to m}(c) \;=\; \mathbb{E}_w\Big[\hat{p}^{(b)}_{im}(x_i^{(b)}) - \hat{p}^{(b)}_{im}(x_i^{(-c)}) \;\Big|\; S_i^{(b)} = s\Big]. \tag{17}$$

Its empirical weighted estimator is

$$\widehat{\Delta}^{\text{PI},(b)}_{s\to m}(c) \;=\; \frac{\sum_{i:\, S_i^{(b)}=s} w_i^{(b)} \Big(\hat{p}^{(b)}_{im}(x_i^{(b)}) - \hat{p}^{(b)}_{im}(x_i^{(-c)})\Big)}{\sum_{i:\, S_i^{(b)}=s} w_i^{(b)}}. \tag{18}$$

Because permutation is random, we average over R independent permutations to obtain a stable estimate:

$$\overline{\Delta}^{\text{PI},(b)}_{s\to m}(c) \;=\; \frac{1}{R}\sum_{r=1}^{R} \widehat{\Delta}^{\text{PI},(b,r)}_{s\to m}(c). \tag{19}$$

The permutation impact tables are presented from Table 13 to Table 20.

4.2 Results

4.2.1 Transition Matrices

from\to	1	2	3	4
1	0.62	0.25	0.08	0.06
2	0.35	0.44	0.15	0.06
3	0.14	0.35	0.29	0.22
4	0.11	0.08	0.10	0.71

Table 1: The estimated transition distribution from year 2021 to year 2023. Rows correspond to the initial (origin) cover-crop level $S = s$ and columns correspond to the destination level $Y = k$; the entry $\widehat{T}_{s \to k}$ is the estimated (weighted) probability of transitioning from level s to level k, with each row summing to 1.

from\to	1	2	3	4
1	0.66	0.17	0.08	0.09
2	0.24	0.57	0.11	0.08
3	0.15	0.21	0.29	0.35
4	0.04	0.03	0.08	0.85

Table 2: The estimated transition distribution from year 2023 to year 2025.

4.2.2 Rank of Covariates using the Gain Importance

	Covariates	Gain
1	Y_from4	0.16
2	Y23_20s	0.06
3	Y23_20v	0.06
4	Y23_20t	0.05
5	Y23_20_N_mgmt_prac	0.04
6	Y_from1	0.03
7	Y23_21_Econ_capacity	0.03
8	Y21_24i	0.02
9	Y21_24_CC_barriers	0.02
10	Y21_11_4R_Info_ag	0.02
⋮		

Table 3: Rank of covariates using the gain importance from XGBoost from year block (2020, 2021) to year block (2022, 2023). Each row corresponds to a (year-suffixed) base covariate; the reported gain is the total loss reduction attributed to splits using that covariate (aggregated across expanded model-matrix columns), with larger gain indicating greater predictive contribution to the transition outcome.

	Covariates	Gain
1	Y25_25b	0.20
2	Y_from1	0.09
3	Y25_27r	0.05
4	Y25_25c	0.05
5	Y25_27o	0.05
6	Y_from2	0.05
7	Y25_27h	0.03
8	Y25_27p	0.03
9	Y25_26_Compaction_mgmt_prac	0.02
10	Y24_3_Satisfaction	0.02
⋮		

Table 4: Rank of covariates using the gain importance from XGBoost from year block (2022, 2023) to year block (2024, 2025).

4.2.3 SHAP Direction (signed SHAP) Summary

Destination level: 1

	Covariates	1	2	3	4
1	ϕ_0 (baseline)	0.75	0.75	0.75	0.75
2	Y23_20s	0.09	-0.07	-0.09	-0.05
3	Y23_20v	0.22	0.00	-0.26	-0.26
4	Y23_20t	0.13	-0.07	-0.16	-0.10
5	Y23_20_N_mgmt_prac	0.08	-0.04	-0.04	-0.06
6	Y23_21_Econ_capacity	0.01	-0.02	-0.02	-0.02
7	Y21_24i	0.00	0.01	0.00	-0.02
8	Y21_24_CC_barriers	0.00	-0.00	-0.01	-0.03
9	Y21_11_4R_Info_ag	0.00	0.00	-0.02	-0.01
10	Y20_14	0.01	0.00	-0.01	-0.02
⋮					

Table 5: SHAP direction (signed SHAP) summary for destination level 1, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023). Columns 1–4 correspond to the initial (origin) cover-crop level $S = s$; each cell reports the weighted mean signed SHAP contribution of that covariate to the destination-1 margin $f_1(x)$ among individuals with $S = s$. The first row ϕ_0 is the SHAP baseline margin for class 1 (the intercept/reference score); each subsequent row is a covariate (year-suffixed base variable), with positive values indicating the covariate tends to increase the class-1 margin (and thus, all else equal, increase $\Pr(Y = 1)$) within that origin stratum, and negative values indicating it tends to decrease the class-1 margin. For example, for `Y23_20s`, a positive value in column $s = 1$ means this covariate pushes predictions toward destination level 1 for those starting at level 1, whereas negative values in columns $s = 2, 3, 4$ mean it pushes predictions away from destination level 1 for those starting at levels 2–4 relative to the baseline ϕ_0. And similar interpretations apply to the rest SHAP direction summaries.

Destination level: 2

	base	1	2	3	4
1	ϕ_0 (baseline)	0.43	0.43	0.43	0.43
2	Y23_20s	-0.37	-0.00	-0.01	-0.10
3	Y23_20v	0.00	0.02	-0.01	-0.06
4	Y23_20t	0.01	0.01	-0.00	-0.02
5	Y23_20_N_mgmt_prac	-0.15	0.01	0.03	0.00
6	Y23_21_Econ_capacity	-0.01	0.00	-0.01	0.01
7	Y21_24i	0.00	0.02	-0.01	-0.08
8	Y21_24_CC_barriers	-0.01	-0.00	0.00	-0.02
9	Y21_11_4R_Info_ag	-0.02	-0.01	0.01	0.00
10	Y20_14	0.00	0.00	-0.01	-0.02
⋮					

Table 6: SHAP direction (signed SHAP) summary for destination level 2, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 3

	base	1	2	3	4
1	ϕ_0 (baseline)	0.10	0.10	0.10	0.10
2	Y23_20s	-0.04	-0.02	0.00	-0.00
3	Y23_20v	-0.07	-0.04	0.01	0.03
4	Y23_20t	-0.12	-0.03	0.02	0.03
5	Y23_20_N_mgmt_prac	-0.12	-0.03	-0.04	-0.00
6	Y23_21_Econ_capacity	-0.01	-0.00	-0.00	-0.00
7	Y21_24i	0.00	0.00	0.00	-0.01
8	Y21_24_CC_barriers	-0.00	0.00	-0.00	-0.02
9	Y21_11_4R_Info_ag	-0.01	-0.01	-0.02	-0.01
10	Y20_14	-0.00	-0.00	-0.00	-0.00
⋮					

Table 7: SHAP direction (signed SHAP) summary for destination level 3, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 4

	base	1	2	3	4
1	ϕ_0 (baseline)	0.58	0.58	0.58	0.58
2	Y23_20s	-0.01	-0.00	-0.00	0.00
3	Y23_20v	-0.01	-0.01	-0.00	0.01
4	Y23_20t	-0.01	-0.00	0.00	0.00
5	Y23_20_N_mgmt_prac	-0.01	0.00	0.00	0.01
6	Y23_21_Econ_capacity	-0.03	-0.03	-0.01	-0.02
7	Y21_24i	-0.10	-0.13	-0.08	0.09
8	Y21_24_CC_barriers	-0.04	-0.06	-0.03	0.04
9	Y21_11_4R_Info_ag	-0.01	-0.01	-0.00	-0.00
10	Y20_14	-0.07	-0.06	-0.01	0.03
⋮					

Table 8: SHAP direction (signed SHAP) summary for destination level 4, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 1

	base	1	2	3	4
1	ϕ_0 (baseline)	0.79	0.79	0.79	0.79
2	Y25_25b	0.01	0.02	-0.02	-0.13
3	Y25_27r	0.03	-0.12	-0.15	-0.15
4	Y25_25c	0.05	0.05	-0.04	-0.26
5	Y25_27o	0.04	-0.07	-0.08	-0.11
6	Y25_27h	-0.09	-0.08	-0.12	-0.13
7	Y25_27p	0.05	0.07	-0.12	-0.13
8	Y25_26_Compaction_mgmt_prac	0.06	0.01	-0.14	-0.22
9	Y24_3_Satisfaction	0.01	-0.01	-0.01	-0.02
10	Y25_27l	0.03	-0.04	-0.05	-0.06
⋮					

Table 9: SHAP direction (signed SHAP) summary for destination level 1, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 2

	base	1	2	3	4
1	ϕ_0 (baseline)	0.42	0.42	0.42	0.42
2	Y25_25b	0.13	0.20	-0.03	-1.22
3	Y25_27r	-0.04	0.00	0.02	0.01
4	Y25_25c	0.01	0.01	-0.01	-0.06
5	Y25_27o	-0.15	0.04	-0.01	-0.00
6	Y25_27h	-0.04	-0.03	0.00	-0.01
7	Y25_27p	-0.01	0.00	0.00	-0.01
8	Y25_26_Compaction_mgmt_prac	-0.00	0.00	-0.00	-0.00
9	Y24_3_Satisfaction	-0.04	0.01	-0.04	-0.01
10	Y25_27l	-0.11	0.01	0.04	0.02
⋮					

Table 10: SHAP direction (signed SHAP) summary for destination level 2, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 3

	base	1	2	3	4
1	ϕ_0 (baseline)	-0.18	-0.18	-0.18	-0.18
2	Y25_25b	0.02	0.03	0.02	-0.06
3	Y25_27r	-0.23	-0.07	0.07	-0.01
4	Y25_25c	-0.02	-0.02	0.01	0.05
5	Y25_27o	-0.03	-0.05	-0.02	-0.01
6	Y25_27h	-0.01	-0.00	0.00	-0.00
7	Y25_27p	-0.03	-0.04	-0.00	0.01
8	Y25_26_Compaction_mgmt_prac	0.01	0.00	-0.00	-0.04
9	Y24_3_Satisfaction	-0.00	-0.01	0.01	-0.00
10	Y25_27l	-0.01	-0.00	0.00	0.00
⋮					

Table 11: SHAP direction (signed SHAP) summary for destination level 3, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 4

	base	1	2	3	4
1	ϕ_0 (baseline)	0.86	0.86	0.86	0.86
2	Y25_25b	-0.78	-0.95	-0.46	1.19
3	Y25_27r	-0.01	-0.00	0.00	-0.00
4	Y25_25c	-0.35	-0.40	-0.15	0.47
5	Y25_27o	-0.03	-0.03	-0.01	0.01
6	Y25_27h	-0.01	-0.00	0.00	0.01
7	Y25_27p	-0.05	-0.10	0.03	0.05
8	Y25_26_Compaction_mgmt_prac	-0.01	-0.01	0.01	0.01
9	Y24_3_Satisfaction	-0.00	-0.00	0.00	-0.00
10	Y25_27l	-0.01	0.00	-0.01	-0.00
⋮					

Table 12: SHAP direction (signed SHAP) summary for destination level 4, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

4.2.4 Permutation Impact Tables

Destination level: 1

	base	1	2	3	4
1	transition prob	0.715	0.361	0.161	0.113
2	Y23_20s	-0.012	0.004	0.003	0.001
3	Y23_20v	0.006	0.015	0.008	0.021
4	Y23_20t	0.003	0.005	-0.002	0.001
5	Y23_20_N_mgmt_prac	-0.003	0.009	-0.003	0.007
6	Y23_21_Econ_capacity	0.001	0.002	-0.002	0.000
7	Y21_24i	-0.001	-0.000	-0.002	-0.000
8	Y21_24_CC_barriers	0.002	0.000	-0.001	-0.000
9	Y21_11_4R_Info_ag	0.002	0.001	-0.000	-0.000
10	Y20_14	-0.001	-0.002	-0.001	0.005
⋮					

Table 13: Permutation impact (probability-scale) summary for destination level 1, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023). Columns 1–4 correspond to the initial (origin) cover-crop level $S = s$; the first row gives the baseline weighted transition probability $\widehat{T}_{s\to 1} = \mathbb{E}_w[\hat{p}_{i1} \mid S = s]$ from the fitted Transition Boosting model. Each subsequent row is a covariate (year-suffixed base variable), and each cell reports the estimated permutation impact $\Delta_{s,1}(b) = \mathbb{E}_w[\hat{p}_{i1} - \hat{p}_{i1}^{(-b)} \mid S = s]$, where $\hat{p}_{i1}^{(-b)}$ is computed after permuting covariate block b within the origin stratum. A positive value means the covariate helps predict destination level 1 in that origin group (permuting it reduces $\Pr(Y = 1)$), whereas a negative value means the covariate suppresses destination level 1 on average (permuting it increases $\Pr(Y = 1)$). For example, for `Y23_20s`, the negative entry in column $s = 1$ indicates that shuffling this covariate slightly increases the predicted probability of ending in level 1 among those starting in level 1, so its observed alignment acts (on average) against destination level 1 relative to the baseline transition probability. And similar interpretations apply to the rest permutation impact tables.

Destination level: 2

	base	1	2	3	4
1	transition prob	0.183	0.413	0.366	0.046
2	Y23_20s	0.021	0.005	0.001	-0.001
3	Y23_20v	-0.002	-0.011	-0.002	-0.005
4	Y23_20t	-0.001	-0.005	0.000	-0.001
5	Y23_20_N_mgmt_prac	0.005	-0.011	0.002	-0.002
6	Y23_21_Econ_capacity	-0.001	-0.002	-0.004	-0.000
7	Y21_24i	0.001	0.000	-0.000	0.002
8	Y21_24_CC_barriers	0.000	0.000	0.000	-0.002
9	Y21_11_4R_Info_ag	-0.001	-0.003	-0.001	-0.001
10	Y20_14	0.000	0.001	0.001	0.001
⋮					

Table 14: Permutation impact (probability-scale) summary for destination level 2, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 3

	base	1	2	3	4
1	transition prob	0.060	0.172	0.274	0.120
2	Y23_20s	-0.006	-0.004	0.000	0.003
3	Y23_20v	-0.002	0.001	-0.002	-0.004
4	Y23_20t	-0.000	0.005	0.003	0.002
5	Y23_20_N_mgmt_prac	-0.001	0.005	0.003	-0.002
6	Y23_21_Econ_capacity	-0.002	-0.000	-0.001	0.001
7	Y21_24i	-0.000	-0.000	0.000	0.002
8	Y21_24_CC_barriers	-0.001	-0.000	-0.001	0.003
9	Y21_11_4R_Info_ag	-0.001	0.002	-0.000	0.002
10	Y20_14	0.000	-0.000	0.000	0.000
⋮					

Table 15: Permutation impact (probability-scale) summary for destination level 3, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 4

	base	1	2	3	4
1	transition prob	0.042	0.053	0.199	0.720
2	Y23_20s	-0.004	-0.005	-0.004	-0.003
3	Y23_20v	-0.002	-0.005	-0.005	-0.012
4	Y23_20t	-0.002	-0.005	-0.002	-0.003
5	Y23_20_N_mgmt_prac	-0.001	-0.003	-0.002	-0.004
6	Y23_21_Econ_capacity	0.001	-0.000	0.007	-0.001
7	Y21_24i	0.000	0.000	0.001	-0.004
8	Y21_24_CC_barriers	-0.001	-0.001	0.002	-0.001
9	Y21_11_4R_Info_ag	-0.001	-0.000	0.001	-0.001
10	Y20_14	0.000	0.001	-0.000	-0.006
⋮					

Table 16: Permutation impact (probability-scale) summary for destination level 4, stratified by origin state, from year block (2020, 2021) to year block (2022, 2023).

Destination level: 1

	base	1	2	3	4
1	transition prob	0.664	0.239	0.155	0.039
2	Y25_25b	-0.007	-0.001	-0.010	-0.005
3	Y25_27r	-0.019	-0.002	0.007	0.002
4	Y25_25c	-0.002	0.002	-0.004	-0.001
5	Y25_27o	-0.021	0.009	0.014	-0.001
6	Y25_27h	-0.010	0.006	-0.000	-0.000
7	Y25_27p	-0.006	0.014	-0.004	-0.001
8	Y25_26_Compaction_mgmt_prac	0.005	0.000	0.014	0.002
9	Y24_3_Satisfaction	0.007	0.000	0.000	0.000
10	Y25_27l	-0.005	-0.001	0.001	-0.000
⋮					

Table 17: Permutation impact (probability-scale) summary for destination level 1, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 2

	base	1	2	3	4
1	transition prob	0.170	0.570	0.209	0.035
2	Y25_25b	-0.004	0.003	-0.005	-0.016
3	Y25_27r	0.005	-0.005	-0.009	-0.003
4	Y25_25c	-0.001	-0.000	-0.002	-0.002
5	Y25_27o	0.017	-0.007	-0.006	-0.001
6	Y25_27h	0.003	0.000	-0.000	-0.000
7	Y25_27p	0.002	-0.003	-0.003	-0.000
8	Y25_26_Compaction_mgmt_prac	-0.001	0.001	-0.009	-0.001
9	Y24_3_Satisfaction	-0.004	0.001	-0.001	-0.000
10	Y25_27l	0.006	0.004	0.000	-0.001
⋮					

Table 18: Permutation impact (probability-scale) summary for destination level 2, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 3

	base	1	2	3	4
1	transition prob	0.080	0.114	0.286	0.076
2	Y25_25b	-0.003	-0.000	-0.036	-0.013
3	Y25_27r	0.017	0.008	0.003	0.005
4	Y25_25c	-0.002	-0.001	-0.010	-0.004
5	Y25_27o	0.006	0.001	-0.006	0.003
6	Y25_27h	0.005	-0.004	-0.000	0.000
7	Y25_27p	0.002	-0.005	-0.005	-0.000
8	Y25_26_Compaction_mgmt_prac	-0.001	-0.000	-0.004	-0.000
9	Y24_3_Satisfaction	-0.002	-0.000	-0.000	0.001
10	Y25_27l	-0.000	-0.003	0.001	0.000
⋮					

Table 19: Permutation impact (probability-scale) summary for destination level 3, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).

Destination level: 4

	base	1	2	3	4
1	transition prob	0.086	0.078	0.350	0.851
2	Y25_25b	0.015	-0.001	0.051	0.034
3	Y25_27r	-0.003	-0.000	-0.000	-0.003
4	Y25_25c	0.005	-0.001	0.016	0.007
5	Y25_27o	-0.001	-0.003	-0.001	-0.001
6	Y25_27h	0.003	-0.002	0.000	-0.000
7	Y25_27p	0.002	-0.007	0.012	0.002
8	Y25_26_Compaction_mgmt_prac	-0.002	-0.001	-0.002	-0.001
9	Y24_3_Satisfaction	-0.002	-0.001	0.001	-0.000
10	Y25_27l	-0.001	-0.000	-0.002	0.001
⋮					

Table 20: Permutation impact (probability-scale) summary for destination level 4, stratified by origin state, from year block (2022, 2023) to year block (2024, 2025).