

Manuscript ID: sustainability-3892916

Title: Comprehensive Benefit Evaluation of Technological Models for Fertile Topsoil Restoration in Thin-Layer Black Soil Region: Evidence from Farmer Survey Data in the Southern Songnen Plain, China

Journal: *Sustainability*

Dear Reviewer

We sincerely thank you for the valuable comments and suggestions. Following yours comments, we have carefully revised our manuscript. We hope that the revised manuscript can satisfactorily address your comments and meet the publication standards of *Sustainability*. Below please find our detailed responses to you. If there are any remaining questions, please let us know.

Sincerely,

1. Theoretical framework: The article justifies the use of the entropy weight–fuzzy comprehensive evaluation method. Why was this methodology chosen over others (e.g., AHP, DEA, or traditional multi-criteria analysis)? Are the limitations of the chosen method sufficiently discussed?

Response 1: We sincerely thank the reviewer for raising this important methodological question. In response to your comment, we have added a dedicated subsection "3.1.3. Methodological Selection Justification" to the manuscript to provide a clearer rationale for our choice of the entropy weight–fuzzy comprehensive evaluation method. We believe this addition significantly strengthens the methodological rigor of our study, with the specific content detailed below:

3.1.3. Methodological Selection Justification

The entropy weight–fuzzy comprehensive evaluation method was selected over alternative multi-criteria approaches due to its unique capacity to address both objective data variability and subjective ambiguity inherent in agricultural technology assessments. Unlike the Analytic Hierarchy Process (AHP), which relies on subjective expert judgments for weighting, the entropy weight method determines indicator weights objectively based on the dispersion of empirical survey data, thereby minimizing arbitrariness. In contrast to Data Envelopment Analysis (DEA), which focuses on efficiency evaluation, our method

accommodates multi-dimensional non-efficiency criteria (economic, social, and ecological) essential for a holistic sustainability assessment. Furthermore, the fuzzy comprehensive evaluation component explicitly handles qualitative uncertainties—such as farmers' perceptions of satisfaction or soil fertility—through membership functions, a feature lacking in traditional multi-criteria analysis. By integrating objective entropy weighting with fuzzy set theory, this method provides a robust framework for evaluating complex technological trade-offs in heterogeneous agricultural systems.

2. Sample: The study relies on 263 households. Is the statistical representativeness adequately demonstrated with respect to the whole farming population of the Songnen Plain?

Response 2: We thank the reviewer for this important question. Our study's target population is explicitly the thin-layer black soil region in the southern Songnen Plain, not the entire vast plain. This precise scope is defined in the title, abstract, and Section 2.1.

To ensure the representativeness of the research sample, we employed a multi-stage stratified random sampling method. First, four representative counties (Nong'an, Gongzhuling, Changling, and Qian'an) in the thin-layer black soil region of the southern Songnen Plain were selected. Then, within these counties, townships, villages, and households were randomly chosen. This method ensures that our sample is not concentrated in a few unique locations but probabilistically represents the diversity of conditions within our study area. The sample characteristics (Table 1) further confirm this by reflecting the region's typical farm size distribution and farmer demographics.

3. Sampling bias: Could the most innovative or open-minded households have been overrepresented in the sample, thereby introducing bias in the results?

Response 3: We thank the reviewer for raising this crucial point. We fully acknowledge that the "innovator bias" or "self-selection bias" is a classic and valid concern in technology adoption studies. In our study, we were aware of this potential risk and implemented the following strategies to mitigate its impact:

First, the core of our methodology was a multi-stage stratified random sampling approach, wherein respondents at the final village and household level were selected randomly. This probabilistic method serves as the primary safeguard against systematically overrepresenting any particular type of household. Second, our sample's demographic profile further mitigates this concern, as it includes a significant proportion of groups typically associated with slower technology adoption—such as older farmers (49.8% over 51 years) and those with lower education levels (53.6% with secondary school or below). Their presence effectively counterbalances any potential overrepresentation of younger, highly educated "innovators."

4. Data temporality: The data were collected in November 2024. Does a single year of data collection truly allow the authors to assess the long-term sustainability (ecological and social)

of the soil restoration models?

Response 4: We sincerely thank the reviewer for this insightful and critical question. We completely agree that a comprehensive assessment of long-term sustainability, by its very nature, requires longitudinal data to capture dynamic trends in soil properties, economic resilience, and community acceptance. The reviewer rightly points out a fundamental limitation of cross-sectional studies like ours. In response, we would like to clarify the objectives of our study and explain how, within the constraints of a single-year survey, we have sought to provide a robust and meaningful evaluation of the potential for sustainability.

(1). Clarification of Study Scope. Our primary aim was not to document long-term trends but to conduct a systematic comparative evaluation of the comprehensive benefits of different restoration models at a critical point in time. We conceptualize these benefits—particularly the ecological and social dimensions—as the essential foundation upon which long-term sustainability is built. A model that performs poorly economically, is rejected by farmers, or shows weak ecological benefits in the short-to-medium term is unlikely to be sustainable in the long run. Therefore, our study provides a crucial first-order screening of which models hold the most promise for sustainability.

(2). Use of Proxy Indicators for Sustainability. While our data is from a single year, the indicators we selected are well-established in the literature as robust proxies for assessing sustainability potential. The key indicators in the ecological and social dimensions—such as soil fertility (X_{33}), moisture retention effect (X_{34}), and pesticide use (X_{31})—are not short-term snapshots but rather reflect the cumulative effects of agricultural practices. They are directly linked to the long-term foundation of soil health and ecosystem stability. Similarly, farmer satisfaction (X_{24}) and income-increasing effect (X_{21}) serve as critical early signals for predicting whether a technology can be sustainably adopted. Thus, these indicators collectively function as effective proxies for assessing the long-term sustainability potential of technological models. Our study utilizes them to identify viable solutions that not only establish a sound ecological foundation but also overcome the "acceptance gap."

We acknowledge that our study design is a limitation for making definitive claims about absolute long-term sustainability. In the revised manuscript, we will strengthen the "Research Limitations and Future Directions" section (Section 5.3) to explicitly state this and outline the necessary next steps. The proposed addition will read:

"Furthermore, this study relies on cross-sectional data from a single year, which limits our ability to capture the dynamic evolution of the technologies' impacts. True long-term sustainability, encompassing aspects like the resilience of soil organic matter increases, the stability of yield gains, and the persistence of farmer adoption, can only be confirmed through long-term positioning experiments and panel data studies. Future research should establish such long-term monitoring sites in the Songnen Plain to validate and extend the findings of this initial comparative assessment."

5. Weight robustness: The article mentions robustness testing of weights by removing indicators (<5% variation). But what about a full sensitivity analysis, for example by randomly adjusting weightings?

Response 5: Thank you for this insightful comment regarding the robustness of our weight assignments. We agree that a comprehensive sensitivity analysis is crucial. In response to your suggestion, we have now conducted a comprehensive sensitivity analysis and included the results as a new subsection (4.2. Robustness and Sensitivity Analysis of Weights) in the revised manuscript. The added content is as follows:

4.2. Robustness and Sensitivity Analysis of Weights

To ensure the reliability of the evaluation results, a comprehensive sensitivity analysis was conducted to examine the robustness of the entropy-based weights before proceeding with the final comprehensive evaluation.

4.2.1. Ranking Stability Analysis.

A Monte Carlo simulation with 10,000 iterations was performed by randomly perturbing each indicator's weight within a $\pm 20\%$ range while preserving the unity constraint of the weight vector. The ranking probability distribution (Figure 3) demonstrates high stability across all technology models. The strip-tillage model maintained the top ranking position with a probability exceeding 0.95, confirming that its superior performance is robust against weight uncertainties. The no-tillage, deep-tillage, and indirect-tillage models also showed consistent ranking patterns, each preserving their respective positions in over 90% of simulations. This high degree of ranking stability provides strong evidence for the reliability of the evaluation outcomes.

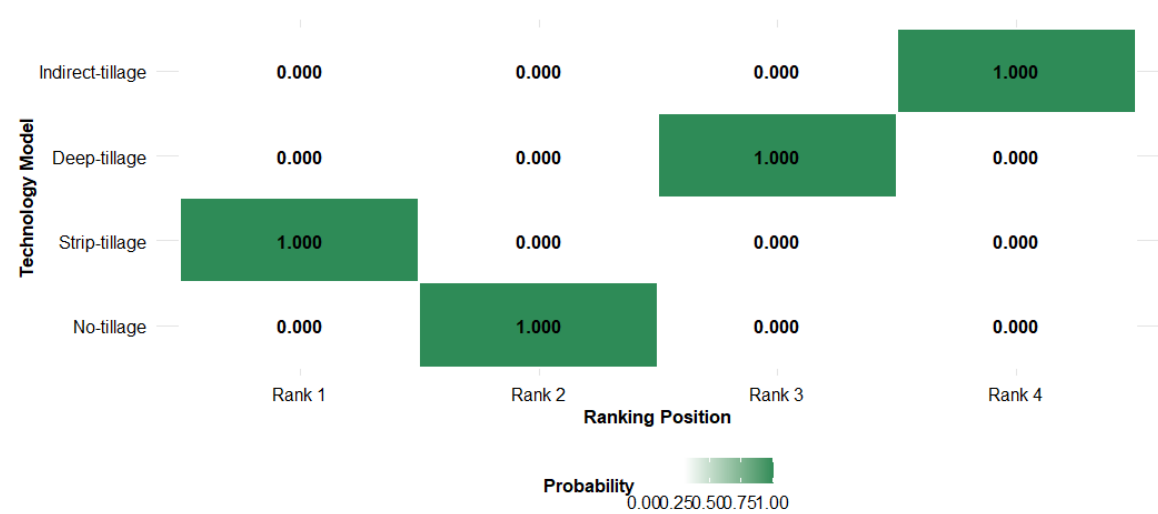


Figure 3. Probability distribution of technology model rankings under weight perturbations (10,000 Monte Carlo simulations).

4.2.2. Global Sensitivity Analysis.

A variance-based global sensitivity analysis using the Sobol' method was conducted to identify the weights that contribute most significantly to output uncertainty. The results (Figure 4) reveal that ranking uncertainty is primarily driven by a small subset of key indicators. The weight for 'Profit per unit area (X_{12})' emerged as the most influential factor, accounting for the largest proportion of output variance. This was followed by 'Farmer satisfaction (X_{24})' and 'Soil fertility (X_{33})', which together explained a substantial portion of the remaining variance. Most other weights showed negligible sensitivity indices, with individual contributions typically below 5% of the total uncertainty.

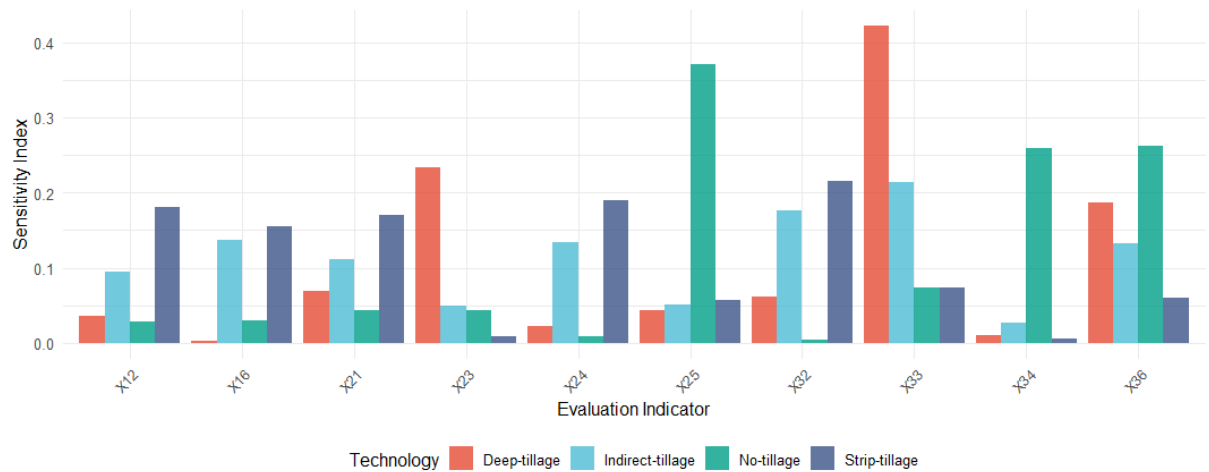


Figure 4. Contribution of each weight to output uncertainty (Top10 most influential indicators).

In summary, these analyses collectively verify that the subsequent evaluation results are not sensitive to minor fluctuations in the majority of weights, thereby underscoring the robustness of the ranking derived from the entropy weight method.

6. Measurement of social benefits: Social indicators (satisfaction, income, training, etc.) partly rely on self-reported data. How can the authors ensure these results are not biased by social desirability effects?

Response 6: We thank the reviewer for raising this crucial methodological point. We fully acknowledge that social desirability bias is a pervasive and legitimate concern in survey-based research, where respondents may overreport positive behaviors or underreport negative ones. In our study, we were acutely aware of this risk and implemented a multi-faceted strategy throughout the research process to mitigate its impact and enhance the validity of the self-reported data. In response, we have added a new subsection, "3.1.4. Mitigation Strategies for Social Desirability Bias".

We believe these additions significantly strengthen the methodological rigor of our manuscript and thank the reviewer for the insightful suggestion. The specific additions are as follows:

3.1.4. Mitigation Strategies for Social Desirability Bias

Recognizing the potential for social desirability bias in farmer-reported data—where respondents may overreport positive behaviors or underreport negative ones—this study implemented a multi-faceted strategy throughout the research process to enhance data validity. The approach was informed by established methodologies for mitigating respondent bias in survey-based environmental and agricultural studies [29].

(1). Assurance of Anonymity and Confidentiality. Prior to each interview, respondents were explicitly and repeatedly assured of the complete anonymity and confidentiality of their responses. They were informed that their answers were solely for academic research purposes, would not be shared with local authorities or any third party, and could not be

linked to their personal identity. This fundamental measure aimed to reduce respondents' motivation to provide socially desirable answers.

(2). **Neutral Questionnaire Design.** The questionnaire was carefully designed with neutral, non-leading wording to minimize interviewer influence. For instance, rather than asking, "Don't you think this technology is excellent?" the survey used neutral phrasing such as, "Please rate your overall satisfaction with this technology on a scale of 1 to 5." This approach allowed farmers to express genuine opinions without perceived pressure to conform to a specific response.

(3). **Data Triangulation via Multi-dimensional Indicators.** The evaluation framework incorporated multiple indicator types to cross-validate subjective responses. As presented in Table 2, the social benefit dimension integrated subjective perceptions (e.g., farmer satisfaction) with more objective behavioral measures (e.g., commodification rate, crop yield per unit area, frequency of technical training). The consistency observed across these different indicator types reinforces the validity of the results and mitigates the risk of perception bias unduly influencing the overall evaluation.

7. Spatial heterogeneity: The Songnen Plain is vast. Do the results sufficiently account for intra-regional agro-climatic variability (microclimate, soil texture, straw availability)?

Response 7: We thank the reviewer for this insightful comment regarding spatial heterogeneity, which is a critical consideration in any regional-scale agricultural study. We agree that the broader Songnen Plain exhibits significant variability. In response, we would like to clarify how our research scope and design specifically address this issue to ensure the validity of our findings within the defined study area.

We would like to clarify that this study focuses on the "thin-layer black soil region in the southern Songnen Plain" (as indicated in the title and Section 2.1). This is not the entire, vast Songnen Plain but a specific agro-ecological sub-zone characterized by: (1) **Similar Soil Conditions:** The dominant soils are black soil and chernozem with a cultivated topsoil thickness generally between 20 and 30 cm, which defines the "thin-layer" characteristic central to our research question. (2) **Consistent Climate Zone:** The entire study area falls within a temperate semi-humid continental monsoon climate zone, with minimal variation in annual temperature (4-6°C) and precipitation (400-600 mm) across the sampled counties. By delimiting our study to this specific sub-region, we intentionally bound the analysis to a context where key agro-climatic factors are relatively consistent, thereby reducing the confounding effects of extreme spatial heterogeneity.

Additionally, following your suggestion, we will explicitly state in the revised discussion that extending the study to the entire Songnen Plain to investigate inter-regional differences represents a critical direction for future work.

8. Adoption by farm size: The study mentions differences between small- and large-scale farmers. Were results stratified by farm size to derive more targeted recommendations?

Response 8: We thank the reviewer for this insightful comment. To address it, we performed a stratified analysis which revealed that the comprehensive benefit score for Strip-tillage was

significantly higher on large-scale farms (>5 hm²), whereas No-tillage performed best on small-scale farms (<1 hm²). These findings have been incorporated into a new paragraph in Section 4.3 and serve as the direct evidence for the differentiated promotion system proposed in the conclusion.

4.3. Stratified Analysis by Farm Management Scale

To identify the optimal technology models for different types of farming entities, this study further stratified the sample households into three groups based on management scale: small-scale (< 1 hm²), medium-scale (1-5 hm²), and large-scale (> 5 hm²). The entropy weight-fuzzy comprehensive evaluation method was then applied separately to each subgroup to determine the optimal technology model for each scale-based cohort. The results of this stratified evaluation are presented in Table 5.

Table 5. Stratified Analysis of Technology Models by Farm Size

Farm Size Category	Sample Proportion	Preferred Technology Model	Comprehensive Benefit Score
Small-scale Farmers	17.50%	No-tillage	7.65
Medium-scale Farmers	69.20%	Strip-tillage	8.18
Large-scale Farmers	13.30%	Strip-tillage	8.41

Analysis of the results indicates significant differences in optimal technology choices among farmers of different management scales. For small-scale farmers, the No-tillage model demonstrated the most favorable comprehensive benefits. This is primarily attributable to the model's minimal machinery investment costs and lower labor requirements, allowing it to exhibit optimal adaptability within the smallholder economy characterized by dual constraints of capital and labor.

For both medium-scale and large-scale farmers, the Strip-tillage model proved to be the optimal choice. As the dominant group in this study (comprising 69.2% of the sample), the performance of the Strip-tillage model among medium-scale farmers stems from its superior output-input ratio and profit per unit area. It successfully integrates the ecological advantages of no-tillage with the agronomic convenience of localized tillage, perfectly aligning with this cohort's core demand for "stable yield and increased income". For large-scale farmers, the advantages of the Strip-tillage model are further amplified by economies of scale, resulting in the highest comprehensive benefit score (8.41) across all groups. This is particularly evident in the economic benefit and farmer satisfaction indicators. This finding confirms that the Strip-tillage model is a key technology for achieving the goal of "reducing costs and increasing efficiency" in large-scale operations.

9. Ecological assessment: Key indicators such as long-term carbon sequestration, soil biodiversity, or biological soil quality are absent. Why were these not included, given their importance in ecological evaluation?

Response 9: We sincerely thank the reviewer for this valuable insight. We completely agree with the reviewer regarding the importance of indicators such as long-term carbon sequestration and soil biodiversity in a comprehensive ecological evaluation.

The main reason these specific indicators were not included stems from the

methodological design of our study. As our research is primarily based on a large-scale socioeconomic survey, obtaining direct measurements of these parameters would have required specialized soil sampling and laboratory analysis, which was unfortunately beyond the practical scope and resources of this particular project.

In designing our evaluation framework, we therefore selected established proxy indicators that are widely recognized in the literature as being closely related to these deeper ecological processes. For instance, our indicator "Soil fertility (X_{33})" is directly correlated with soil organic matter content, which serves as a key indicator for carbon sequestration potential. Similarly, "Reduced pesticide use (X_{31})" and "Improved moisture retention (X_{34})" are well-established as creating favorable conditions for the development and maintenance of soil biodiversity.

We acknowledge that the inclusion of the direct measurements suggested by the reviewer would have further enriched our ecological assessment. In response to this helpful suggestion, we will explicitly acknowledge this point in the revised manuscript's discussion of limitations and future research directions, highlighting the importance of incorporating such advanced measurements in subsequent studies.

10. International comparison: The study focuses only on China. Would a comparison with other chernozem regions (Ukraine, Russia, Canada) not enrich the discussion and strengthen the global relevance of the study?

Response 10: We sincerely thank the reviewer for this excellent suggestion. We completely agree that a comparative discussion with other major global chernozem regions would significantly enhance the global relevance and scholarly contribution of our work. In response to this valuable suggestion, we have incorporated a dedicated paragraph on international comparison within the "Discussion of Research Findings" (Section 5.1). This addition significantly enhances the global context of our research discussion by situating our findings within the broader landscape of chernozem management practices worldwide.

Furthermore, situating our findings within a global context enriches the discussion and underscores the importance of region-specific technology adaptation. The challenges of black soil degradation are not unique to the Songnen Plain but are a global concern across the world's major chernozem belts. However, the optimal solution has significant regional differences. For instance, in the warmer and more extensive steppes of Ukraine, the classic no-tillage model is often more widely applicable and effective [48]. In the North American prairies, while conservation tillage systems including no-till and strip-till are also widely implemented, their adoption patterns and implementation scales are significantly influenced by different farm structures, market mechanisms, and policy frameworks [49]. These international comparisons underscore a fundamental principle: while the core conservation objective of minimizing soil disturbance remains universal across chernozem regions, its practical implementation must be carefully calibrated to local conditions. The specific agro-climatic constraints of the Songnen Plain—characterized by its thin soil layers and cold springs—necessitate technological adaptations that may differ from solutions optimal in other global chernozem regions.

11. No-tillage limitations: The manuscript highlights high pest and weed management costs. Were alternative agronomic solutions (e.g., crop rotations, cover crops) considered in the discussion?

Response 11: We thank the reviewer for raising this insightful point. We agree that integrated pest and weed management strategies, such as the use of cover crops and diversified crop rotations, are crucial components of sustainable agricultural systems and could potentially mitigate the operational challenges associated with the no-tillage model.

The primary focus of the current manuscript, however, is on the comparative evaluation of the core fertile topsoil restoration models themselves (no-tillage, strip-tillage, deep-tillage, and indirect-tillage) as standalone, widely recognized technological packages. Delving into the optimization of each model through complementary agronomic practices, while highly valuable, would substantially broaden the scope beyond the central comparative framework of this study.

In response to the reviewer valuable suggestion, we will explicitly acknowledge this limitation and incorporate a discussion on the potential of integrating these alternative agronomic solutions into the "Research Limitations and Future Directions" section. We will add a sentence to this effect:

"Furthermore, while this study evaluated the core technological models, it did not explore the potential of integrated agronomic strategies—such as cover cropping or diversified crop rotations—to mitigate specific limitations like the high weed management costs in no-tillage systems. Investigating such synergistic combinations represents a promising avenue for future research."

12. Indirect tillage relevance: The low adoption of indirect tillage is attributed to technical complexity. Could this be more a question of institutional limitations (e.g., access to composting or biogas infrastructure)?

Response 12: We sincerely thank the reviewer for this exceptionally insightful comment, which rightly directs our attention to the critical institutional and infrastructural dimensions that underpin technology adoption.

In direct response, we have revised our discussion in the manuscript to explicitly reframe this point. We have expanded the relevant paragraph in Section 5.1 to state that the promotion of the indirect tillage model is constrained not merely by its technical demands but more critically by "a lack of accessible institutional frameworks and supporting infrastructure, such as centralized composting facilities, biogas plants, and efficient straw collection and redistribution networks." The specific amendments are as follows:

The indirect model exhibits the best ecological sustainability (7.781 points), but its widespread adoption is hindered not only by high technical complexity and cost but also by a lack of accessible institutional frameworks and supporting infrastructure, such as centralized composting facilities, biogas plants, and efficient straw collection and redistribution networks. This aligns with the "high willingness–low adoption" phenomenon reported by Farahbakhsh

et al. [47], reflecting the compound challenges of technical, capital, and institutional constraints faced by smallholder farmers.

13. Public policy dimension: The paper recommends a differentiated promotion system. Have the economic impacts (subsidies, investments) of such a policy been evaluated to test its feasibility at scale?

Response 13: We thank the reviewer for this crucial question regarding policy feasibility. We agree that a detailed economic impact assessment, including the costs of subsidies, investments in machinery, and infrastructure, is essential for the large-scale implementation of a differentiated promotion system.

The primary objective of our study was to provide a scientific basis and a strategic framework for such policies by first identifying which technologies are most suitable for which types of farmers. A full-scale economic and fiscal evaluation, while critically important, represents a distinct and subsequent phase of policy analysis that was beyond the scope of this initial techno-economic comparison.

In response to this comment, we have strengthened the conclusion to explicitly state this as a necessary next step. We have clarified in the revised manuscript that: "Future research should conduct a comprehensive cost-benefit analysis of the proposed promotion system, quantifying the necessary public investment, potential subsidy schemes, and long-term economic returns to verify its fiscal feasibility and optimize its design."

14. Duration of adoption: Did the study account for how long farmers had been using each technology, since short- and long-term effects may differ significantly?

Response 14: We thank the reviewer for this insightful question regarding a crucial aspect of technology impact assessment. We acknowledge that our cross-sectional survey design captures a snapshot in time and does not systematically account for the varying lengths of time that farmers have been using each technology. While we collected basic adoption information, our analysis primarily focused on comparing the current perceived performance of the technologies as they are implemented in the field, regardless of their adoption history. This approach is effective for identifying the immediate, comparative benefits and constraints that drive farmer satisfaction and initial adoption decisions, which was a primary goal of our research.

However, we fully agree with the reviewer that understanding the dynamic effects of these technologies—such as how soil properties and economic returns evolve over years of continuous use—is essential for a complete assessment of their long-term sustainability. This important dimension could not be robustly addressed with our current data. In response to this valuable comment, we have explicitly acknowledged this limitation in the revised manuscript's "Research Limitations and Future Directions" section. We have added a statement emphasizing that:

"Future longitudinal studies or panel data analyses are needed to trace the dynamic effects of these restoration models over time, thereby capturing the divergence between

short-term adaptations and long-term sustainability outcomes."

15. External validity: To what extent can these findings be generalized to other black soil regions in China or abroad, which may have different climatic and socio-economic conditions?

Response 15: We sincerely thank the reviewer for raising this important question regarding the external validity of our findings. While our evaluation was conducted specifically in the thin-layer chernozem region of the Southern Songnen Plain, we believe both the methodological framework and practical implications offer valuable transferability to other black soil regions.

The integrated evaluation methodology developed in this study, which combines entropy weighting with fuzzy comprehensive evaluation, establishes an adaptable analytical framework suitable for assessing technology benefits across diverse geographical and socio-economic contexts. While specific technology rankings may vary under different climatic conditions or farming structures, the core principles of our approach maintain broad relevance for technology assessment in various degraded soil regions. This relevance stems particularly from our multidimensional evaluation framework that systematically incorporates economic, social, and ecological dimensions of technology performance. Building on this foundation, our systematic approach enables the identification of context-specific advantages and constraints of different technology models, thereby providing valuable guidance for designing tailored promotion strategies in other black soil areas.

16. Strengthen methodological justification by comparing the chosen approach with alternative multi-criteria methods.

Response 16: We sincerely thank the reviewer for raising this important methodological question. In response to your comment, we have added a dedicated subsection "3.1.3. Methodological Selection Justification" to the manuscript to provide a clearer rationale for our choice of the entropy weight–fuzzy comprehensive evaluation method. We believe this addition significantly strengthens the methodological rigor of our study, with the specific content detailed below:

3.1.3. Methodological Selection Justification

The entropy weight–fuzzy comprehensive evaluation method was selected over alternative multi-criteria approaches due to its unique capacity to address both objective data variability and subjective ambiguity inherent in agricultural technology assessments. Unlike the Analytic Hierarchy Process (AHP), which relies on subjective expert judgments for weighting, the entropy weight method determines indicator weights objectively based on the dispersion of empirical survey data, thereby minimizing arbitrariness. In contrast to Data Envelopment Analysis (DEA), which focuses on efficiency evaluation, our method accommodates multi-dimensional non-efficiency criteria (economic, social, and ecological) essential for a holistic sustainability assessment. Furthermore, the fuzzy comprehensive

evaluation component explicitly handles qualitative uncertainties—such as farmers' perceptions of satisfaction or soil fertility—through membership functions, a feature lacking in traditional multi-criteria analysis. By integrating objective entropy weighting with fuzzy set theory, this method provides a robust framework for evaluating complex technological trade-offs in heterogeneous agricultural systems.

17. **Conduct a more in-depth sensitivity analysis** to test the robustness of results.

Response 17: Thank you for this insightful comment regarding the robustness of our weight assignments. We agree that a comprehensive sensitivity analysis is crucial. In response to your suggestion, we have now conducted a comprehensive sensitivity analysis and included the results as a new subsection (4.2. Robustness and Sensitivity Analysis of Weights) in the revised manuscript. The added content is as follows:

4.2. *Robustness and Sensitivity Analysis of Weights*

To ensure the reliability of the evaluation results, a comprehensive sensitivity analysis was conducted to examine the robustness of the entropy-based weights before proceeding with the final comprehensive evaluation.

4.2.1. Ranking Stability Analysis.

A Monte Carlo simulation with 10,000 iterations was performed by randomly perturbing each indicator's weight within a $\pm 20\%$ range while preserving the unity constraint of the weight vector. The ranking probability distribution (Figure 3) demonstrates high stability across all technology models. The strip-tillage model maintained the top ranking position with a probability exceeding 0.95, confirming that its superior performance is robust against weight uncertainties. The no-tillage, deep-tillage, and indirect-tillage models also showed consistent ranking patterns, each preserving their respective positions in over 90% of simulations. This high degree of ranking stability provides strong evidence for the reliability of the evaluation outcomes.

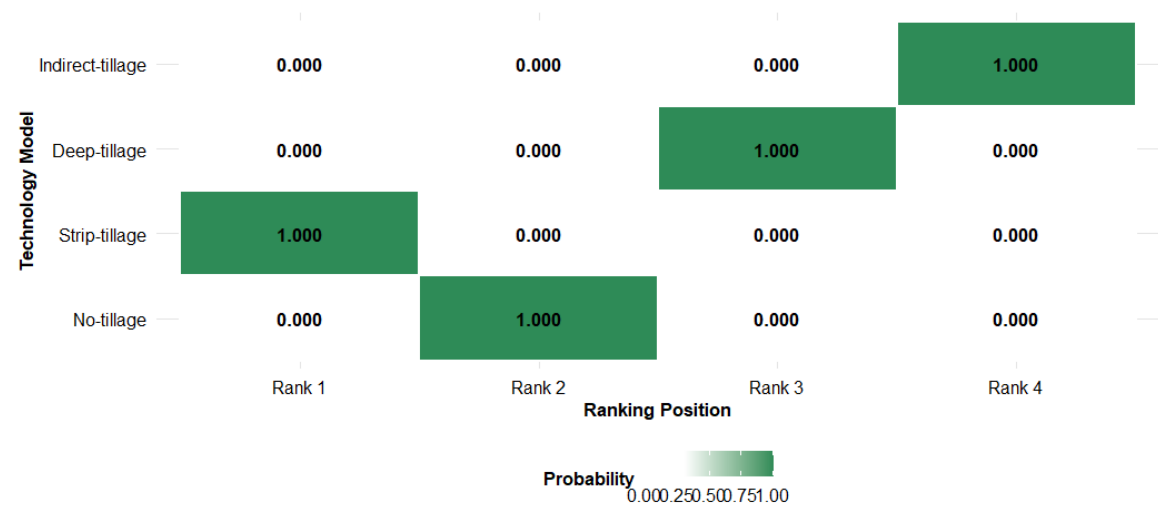


Figure 3. Probability distribution of technology model rankings under weight perturbations (10,000 Monte Carlo simulations).

4.2.2. Global Sensitivity Analysis.

A variance-based global sensitivity analysis using the Sobol' method was conducted to identify the weights that contribute most significantly to output uncertainty. The results (Figure 4) reveal that ranking uncertainty is primarily driven by a small subset of key indicators. The weight for 'Profit per unit area (X_{12})' emerged as the most influential factor, accounting for the largest proportion of output variance. This was followed by 'Farmer satisfaction (X_{24})' and 'Soil fertility (X_{33})', which together explained a substantial portion of the remaining variance. Most other weights showed negligible sensitivity indices, with individual contributions typically below 5% of the total uncertainty.

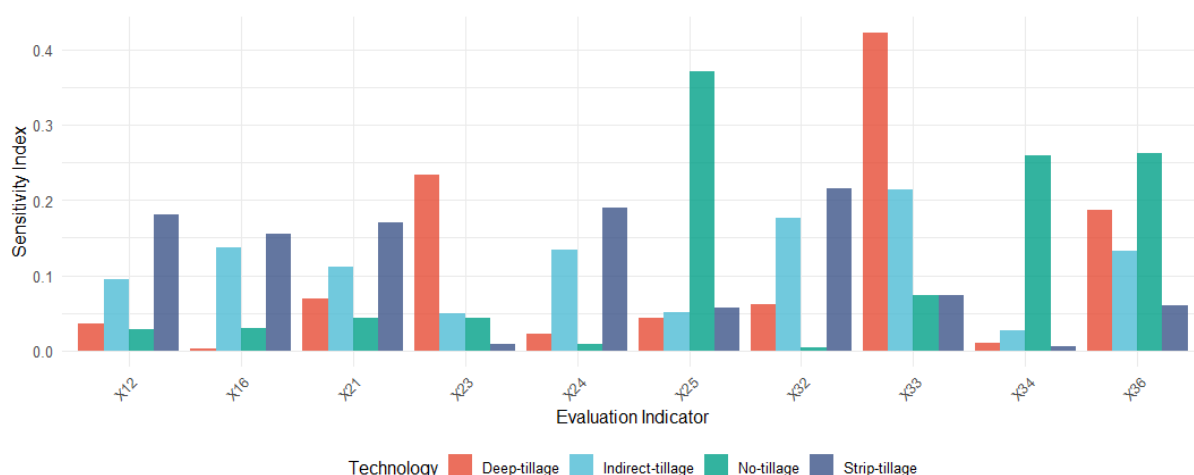


Figure 4. Contribution of each weight to output uncertainty (Top10 most influential indicators).

In summary, these analyses collectively verify that the subsequent evaluation results are not sensitive to minor fluctuations in the majority of weights, thereby underscoring the robustness of the ranking derived from the entropy weight method.

18. Expand ecological evaluation to include fundamental indicators (carbon, biodiversity, biological soil quality).

Response 18: We sincerely thank the reviewer for this valuable insight. We completely agree with the reviewer regarding the importance of indicators such as long-term carbon sequestration and soil biodiversity in a comprehensive ecological evaluation.

However, as our research is primarily based on a large-scale socioeconomic survey, obtaining direct measurements of these parameters would have required specialized soil sampling and laboratory analysis, which was unfortunately beyond the practical scope and resources of this particular project. In designing our evaluation framework, we therefore selected established proxy indicators that are widely recognized in the literature as being closely related to these deeper ecological processes. For instance, our indicator "Soil fertility (X_{33})" is directly correlated with soil organic matter content, which serves as a key indicator for carbon sequestration potential. Similarly, "Reduced pesticide use (X_{31})" and "Improved moisture retention (X_{34})" are well-established as creating favorable conditions for the development and maintenance of soil biodiversity.

We acknowledge that the inclusion of the direct measurements suggested by the

reviewer would have further enriched our ecological assessment. In response to this helpful suggestion, we will explicitly acknowledge this point in the revised manuscript's discussion of limitations and future research directions, highlighting the importance of incorporating such advanced measurements in subsequent studies.

19. **Stratify results by farm size** to better inform policy recommendations.

Response 19: We thank the reviewer for this insightful comment. We have performed a stratified analysis by farm size, and the results have been incorporated into a new subsection (Section 4.3 Stratified Analysis by Farm Management Scale) and Table 5 in the revised manuscript. This analysis provides the direct empirical evidence for the differentiated promotion system proposed in the conclusion. The following new content has been added to the manuscript:

4.3. Stratified Analysis by Farm Management Scale

To identify the optimal technology models for different types of farming entities, this study further stratified the sample households into three groups based on management scale: small-scale (< 1 hm²), medium-scale (1-5 hm²), and large-scale (> 5 hm²). The entropy weight-fuzzy comprehensive evaluation method was then applied separately to each subgroup to determine the optimal technology model for each scale-based cohort. The results of this stratified evaluation are presented in Table 5.

Table 5. Stratified Analysis of Technology Models by Farm Size

Farm Size Category	Sample Proportion	Preferred Technology Model	Comprehensive Benefit Score
Small-scale Farmers	17.50%	No-tillage	7.65
Medium-scale Farmers	69.20%	Strip-tillage	8.18
Large-scale Farmers	13.30%	Strip-tillage	8.41

Analysis of the results indicates significant differences in optimal technology choices among farmers of different management scales. For small-scale farmers, the No-tillage model demonstrated the most favorable comprehensive benefits. This is primarily attributable to the model's minimal machinery investment costs and lower labor requirements, allowing it to exhibit optimal adaptability within the smallholder economy characterized by dual constraints of capital and labor.

For both medium-scale and large-scale farmers, the Strip-tillage model proved to be the optimal choice. As the dominant group in this study (comprising 69.2% of the sample), the performance of the Strip-tillage model among medium-scale farmers stems from its superior output-input ratio and profit per unit area. It successfully integrates the ecological advantages of no-tillage with the agronomic convenience of localized tillage, perfectly aligning with this cohort's core demand for "stable yield and increased income". For large-scale farmers, the advantages of the Strip-tillage model are further amplified by economies of scale, resulting in the highest comprehensive benefit score (8.41) across all groups. This is particularly evident in the economic benefit and farmer satisfaction indicators. This finding confirms that the Strip-tillage model is a key technology for achieving the goal of "reducing costs and increasing efficiency" in large-scale operations.

20. Incorporate a temporal dimension (longitudinal or multi-year studies) to evaluate true sustainability.

Response 20: We thank the reviewer for this critical insight. We fully agree that longitudinal data tracking the dynamic evolution of soil properties, economic outcomes, and farmer adoption would provide a more robust assessment of long-term sustainability.

However, as a cross-sectional study designed to provide a comparative snapshot of different technologies' performance, our current methodology could not capture these temporal dynamics. The primary goal of this research was to establish a foundational multi-dimensional evaluation framework and identify the most promising technologies for the region, which can now serve as a baseline for future long-term monitoring.

In response to this valuable comment, we have explicitly acknowledged this limitation in the revised manuscript and emphasized the necessity of longitudinal studies in the "Research Limitations and Future Directions" section. We have added the following statement:

"Future longitudinal studies or panel data analyses are needed to trace the dynamic effects of these restoration models over time, thereby capturing the divergence between short-term adaptations and long-term sustainability outcomes."

21. Consider institutional and logistical barriers (straw collection, processing, redistribution).

Response 21: We thank the reviewer for this highly relevant and constructive suggestion. We agree that a deeper analysis of the institutional and logistical barriers is crucial for understanding the real-world feasibility of the indirect tillage model.

In direct response to this comment, we have substantially enriched our discussion in the manuscript (Section 5.1) to move beyond a purely technical explanation. We now explicitly analyze the systemic constraints, stating that the low adoption of indirect tillage is not merely a question of on-farm technical complexity but is fundamentally constrained by "the absence of robust institutional frameworks and logistical infrastructure for straw collection, cost-effective processing (e.g., composting, biogas production), and the redistribution of organic amendments to farmland."

22. Develop policy scenarios with quantified costs (subsidies, mechanization, training) to check economic feasibility.

Response 22: We thank the reviewer for this crucial suggestion regarding policy feasibility. We agree that developing detailed policy scenarios with quantified costs for subsidies, mechanization services, and training programs is an essential next step for implementing the recommended differentiated promotion system.

The primary objective of our study was to establish the scientific foundation for such

policies by first identifying which technologies are most suitable for which types of farmers. A full-scale economic and fiscal evaluation, including the simulation of different subsidy levels and their impacts, represents a distinct and subsequent phase of policy analysis that was beyond the scope of this initial techno-economic comparison.

In direct response to this comment, we have strengthened the "Research Limitations and Future Directions" section to explicitly state this as a necessary and logical continuation of our work. We have added a clear recommendation that:

"Future research should build upon these findings to develop and simulate specific policy scenarios. This involves quantifying the required public investment in machinery subsidies, technical training, and infrastructure support, and conducting a comprehensive cost-benefit analysis to validate the economic feasibility and optimize the design of the promotion system at scale."

23. Connect findings to the international literature for broader global relevance.

Response 23: We sincerely thank the reviewer for this excellent suggestion. In response to this valuable suggestion, we have incorporated a dedicated paragraph on international comparison within the "Discussion of Research Findings" (Section 5.1). This addition significantly enhances the global context of our research discussion by situating our findings within the broader landscape of chernozem management practices worldwide.

Furthermore, situating our findings within a global context enriches the discussion and underscores the importance of region-specific technology adaptation. The challenges of black soil degradation are not unique to the Songnen Plain but are a global concern across the world's major chernozem belts. However, the optimal solution has significant regional differences. For instance, in the warmer and more extensive steppes of Ukraine, the classic no-tillage model is often more widely applicable and effective [48]. In the North American prairies, while conservation tillage systems including no-till and strip-till are also widely implemented, their adoption patterns and implementation scales are significantly influenced by different farm structures, market mechanisms, and policy frameworks [49]. These international comparisons underscore a fundamental principle: while the core conservation objective of minimizing soil disturbance remains universal across chernozem regions, its practical implementation must be carefully calibrated to local conditions. The specific agro-climatic constraints of the Songnen Plain—characterized by its thin soil layers and cold springs—necessitate technological adaptations that may differ from solutions optimal in other global chernozem regions.

24. Explore technological synergies (e.g., strip tillage + organic fertilization) in future perspectives.

Response 24: We sincerely thank the reviewer for this forward-looking and insightful suggestion. We completely agree that exploring synergistic technological combinations, represents a highly promising avenue to maximize the comprehensive benefits of fertile topsoil restoration.

While the current study focused on evaluating the core models as standalone systems to

establish a clear baseline for comparison, we recognize that the integration of their strengths—for instance, leveraging the precise soil loosening of strip tillage alongside the nutrient recycling benefits of organic amendments—could potentially create a system superior to any single model.

In direct response to this valuable comment, we have incorporated a specific discussion on this potential in the "Research Limitations and Future Directions" section of the revised manuscript. We now explicitly state:

"Future research should prioritize exploring synergistic technology integrations, such as combining strip tillage with organic fertilizer application. Investigating these combinations for their ability to simultaneously enhance soil structure, nutrient use efficiency, and long-term carbon sequestration will be crucial for developing the next generation of sustainable soil management practices in the region."

25. Clarify biases in subjective data by explaining how farmers' perception bias was addressed.

Please cite this article : <https://doi.org/10.1016/j.ecoleng.2006.09.021>

Response 25: We sincerely thank the reviewer for this valuable comment. We fully agree that addressing potential perception bias is crucial for ensuring the validity of subjective data. In response, we have added a new subsection, "3.1.4. Mitigation Strategies for Social Desirability Bias".

Furthermore, as suggested by the reviewer, we have cited the recommended literature (Ádám et al., 2007, Ecological Engineering) as reference [29] within this new subsection to anchor our methodological approach in established practices.

We believe these additions significantly strengthen the methodological rigor of our manuscript and thank the reviewer for the insightful suggestion. The specific additions are as follows:

3.1.4. Mitigation Strategies for Social Desirability Bias

Recognizing the potential for social desirability bias in farmer-reported data—where respondents may overreport positive behaviors or underreport negative ones—this study implemented a multi-faceted strategy throughout the research process to enhance data validity. The approach was informed by established methodologies for mitigating respondent bias in survey-based environmental and agricultural studies [29].

(1). Assurance of Anonymity and Confidentiality. Prior to each interview, respondents were explicitly and repeatedly assured of the complete anonymity and confidentiality of their responses. They were informed that their answers were solely for academic research purposes, would not be shared with local authorities or any third party, and could not be linked to their personal identity. This fundamental measure aimed to reduce respondents' motivation to provide socially desirable answers.

(2). Neutral Questionnaire Design. The questionnaire was carefully designed with neutral, non-leading wording to minimize interviewer influence. For instance, rather than asking, "Don't you think this technology is excellent?" the survey used neutral phrasing such as, "Please rate your overall satisfaction with this technology on a scale of 1 to 5." This approach allowed farmers to express genuine opinions without perceived pressure to conform to a specific response.

(3). Data Triangulation via Multi-dimensional Indicators. The evaluation framework incorporated multiple indicator types to cross-validate subjective responses. As presented in Table 2, the social benefit dimension integrated subjective perceptions (e.g., farmer satisfaction) with more objective behavioral measures (e.g., commodification rate, crop yield per unit area, frequency of technical training). The consistency observed across these different indicator types reinforces the validity of the results and mitigates the risk of perception bias unduly influencing the overall evaluation.