Point-by-point Response

ID: agriculture-4114898

Title: Diagnosing and projecting farmland ecosystem health in arid regions: An interpretable machine learning and scenario simulation approach within a novel integrity-based framework

We sincerely thank the reviewers for the constructive feedback, which has been instrumental in improving the quality of our manuscript. We have carefully considered all the comments and conducted a comprehensive revision of the paper. The major improvements focus on: refining the methodological procedures and key parameters to ensure full reproducibility; strengthening the statistical validation and stability checks to enhance the overall rigor of the modeling framework; clarifying ambiguous expressions and core concepts throughout the text to improve readability and accuracy; optimizing the visual clarity of figures; and restructuring the manuscript by adding a concise "Conclusions" section to better highlight the primary findings and practical implications. We believe these targeted revisions have thoroughly addressed the reviewers' concerns and significantly elevated the quality of the study.

We have included an itemized list of our responses below in the following format:

The reviewers' comments (shown in bold font)

My responses (shown in normal font)

Changes to the manuscript (shown in blue italics font, along with the tracking and tagging version)

original text in the manuscript (shown in italics font)

Reviewer1#

Overall response:

We sincerely appreciate the reviewer's valuable comments and suggestions. We have carefully revised and improved the manuscript in terms of figure quality and presentation, introduction and conclusion content, keywords and academic citation standardization, as well as AI usage statement in accordance with all comments point-by-point. The detailed point-by-point responses are as follows.

Comment 1–1:

Please enhance the resolution of your study area map (Figure 1). Please also latitude and longitude coordinate values to the c and d parts of Figure 1.

Responses:

Thank you for this valuable comment. We have revised Figure 1 entirely, improved its resolution, and added latitude and longitude coordinate values to panels c and d of the figure.

Comment 1–2:

Please add 'ecosystem health index' to the keywords.

Responses:

Thank you for your constructive comment. Since "ecosystem health" was already included in the keywords, we have revised it to "ecosystem health index" as suggested.

Comment 1–3:

In the introduction, please add a brief on 'ecosystem health' and connect the study with previous ecosystem health-based studies for expressing novelty of your research.

Responses:

Thank you for this constructive suggestion. We have supplemented the introduction section with a brief elaboration on ecosystem health and strengthened the linkage with previous relevant studies to better highlight the novelty of this research.

We have added the following description:

"..... With a growing global population and escalating food demand, the health of farmland ecosystems has become inextricably linked to agricultural resilience, long-term food production, and ecological civilization (de la Riva et al., 2023; Ma and Wei, 2021). In this context, ecosystem health—originally conceptualized as an integrative paradigm encompassing a system's vigor, organization, and resilience (Rapport et al., 1998)—has evolved into a fundamental metric for evaluating the sustainability of complex agro-ecological systems. It transcends traditional single-parameter environmental monitoring by diagnosing whether an ecosystem retains its self-organizational capacity, structural integrity, and adaptive potential under external stresses (Yadav et al., 2025)." (lines 43-50)

"..... While previous ecosystem health-based studies have established a solid foundation by integrating ecological, environmental, and socio-economic indicators

into established frameworks like the Pressure-State-Response (PSR) model (Lee et al., 2024) or the Vitality-Organization-Resilience-Services (VORS) model (Li et al., 2023; Sun et al., 2025), they often fall short in capturing the complete picture of system degradation. Specifically, a persistent gap exists in the explicit and systematic quantification of ecological integrity (Bai et al., 2025)....." *(lines 58-63)*

Comment 1–4:

Figure 2 seems AI-generated. Please add the Use of AI and AI-Assisted Technologies statement. Could you please include the following sentence in the statement and provide an updated Use of AI and AI-Assisted Technologies section? After using the AI tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

Responses:

Thank you for your reminder. Not all elements in Figure 2 were AI-generated. Only the cartoon-style schematic of the farmland ecosystem landscape was created with AI assistance (Doubao), while the remaining components were manually designed using the online tool ProcessOn (www.processon.com) without AI generation.

We have added a Use of AI and AI-Assisted Technologies statement in the manuscript, which includes the required sentence: "*After using the AI tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.*"

Comment 1–5:

You mentioned: To counteract the diminishing effect of multiplying fractional values, the fifth root of the product is calculated to derive the composite Ecosystem Health Index (EHI). The specific formula is as follows: $EHI = \sqrt{EV \times EO \times ER \times ES \times EI}$. Please add a reference to the used EHI.

Responses:

Thank you for your constructive comment. We have added a relevant reference for the Ecosystem Health Index (EHI) calculation as suggested (line 191). Peng et al. (2015, https://doi.org/10.1016/j.landurbplan.2015.06.007) stated that "In order to avoid a calculation that magnifies the indicators when they are multiplied, it is necessary to adopt a radical sign to neutralize the order of magnitude". Accordingly, our calculation method is reasonable and has been supported by the literature.

Comment 1–6:

Please also add a reference to Habitat Quality Assessment's equation, which was calculated using the InVEST model.

Responses:

Thank you for your constructive comment. We have added the relevant literature reference for the habitat quality equation derived from the InVEST model as suggested. (line 214)

Comment 1–7:

Please add a reference column to Table 2. Calculation methods and parameter descriptions for different ecosystem service types.

Responses:

Thank you for your constructive comment. We have added a reference column to Table 2 as requested. All calculation methods and parameter descriptions for the ecosystem service types are cited from the InVEST User Guide (Alliance, 2026):

Natural Capital Alliance, 2026. InVEST 3.18.0. Stanford University, University of Minnesota, Chinese Academy of Sciences, The Nature Conservancy, World Wildlife Fund, Stockholm Resilience Centre and the Royal Swedish Academy of Sciences. https://doi.org/10.60793/natcap-invest-3.18.0

Comment 1–8:

Please add a conclusion section with key results, policy recommendation and limitations of the study.

Responses:

Thank you for your suggestion. We have added a conclusion section (including key results, policy recommendations and study limitations) in line with journal standards, which is at Line 668-676.

Reviewer2#

Comment 2-1:

L14-15: "yet existing assessment frameworks often lack explicit

quantification of ecosystem integrity". While this gap is presented as a key justification, I suggest adding 1-2 specific references that illustrate this limitation in existing frameworks (PSR, VORS) to strengthen the argument.

Responses:

Thank you for this important comment. We have added two relevant references to support our argument as suggested. Since this statement is located in the abstract, where references are not allowed, we have added these citations in the corresponding position of the Introduction section. Both references highlight the lack of integrity quantification in existing assessment frameworks. Zhao et al. (2025) evaluates various dimensions of integrity and provides a quantitative diagnosis, directly addressing the need for explicit quantification of ecosystem integrity in health assessments. Mina et al. (2024) provides a framework that measures and ranks pressures, aligning with the need for explicit quantification of ecosystem integrity in health assessments.

Comment 2-2:

L23-24: "with a spatial gradient of higher values in the mountains and lower in the central plains". This pattern is important. I recommend briefly adding the numerical range of this spatial variation (e.g., "ranging from X in plains to Y in mountains") to provide greater quantitative precision.

Responses:

Thank you for your constructive comment. Based on spatial analysis and statistics in ArcGIS, we have added the specific numerical range as suggested: the EHI values in mountainous areas *range from 0.53 to 0.80, while those in the central plains range from 0.11 to 0.52*, to enhance the quantitative precision of the spatial pattern description.

Comment 2-3:

L25-26: "Potential evapotranspiration (PET) and slope were the top drivers, with notable synergistic interactions and threshold effects (e.g., PET ~557 mm)". Excellent threshold specification. I suggest clarifying whether this 557 mm threshold represents an optimal maximum, a critical inflection point, or a water stress limit, as this has direct management implications.

Responses:

Thank you for your insightful suggestion to clarify the nature of the 557 mm threshold. Based on the SHAP value plot (Figure 9a), this threshold functions as a

critical inflection point where the direction of PET's contribution to ecosystem health reverses—from positive (enhancing health) below 557 mm to negative (reducing health) above it. At 557 mm, the SHAP value reaches its peak, indicating this is the optimal maximum for PET's positive impact on ecosystem health. Additionally, the negative SHAP values beyond 557 mm suggest it also serves as a water stress limit, as higher PET values correspond to declining ecosystem health due to water deficit conditions. We have revised the discussion (lines 603-607) to clarify this multifaceted interpretation, which aligns with the curve's trend in Figure 9a and underscores the threshold's relevance for management.

Comment 2-4:

L56-62: "Integrity, emphasizing the coherence, connectivity, and authenticity of a system's structure, function, and processes... lacking dedicated, quantifiable indicators". This is a central argument of the work. I recommend adding 2-3 concrete examples from recent studies that illustrate how the absence of explicit integrity quantification has led to incomplete or erroneous assessments.

Responses:

Thank you for this valuable comment to strengthen our core argument. We have added two concrete empirical cases in the Introduction section as suggested, to illustrate the consequences of lacking explicit integrity quantification in existing assessments. Specifically, we inserted *Andreasen et al. (2001)* to demonstrate that overlooking ecosystem integrity can mask gradual degradation until irreversible tipping points are reached, and Nasr and Orwin (2024) to show that *ignoring integrity introduces scale-dependent biases that undermine the robustness of resource management decisions* (line 70-75).

Comment 2-5:

L73-76: "we propose an enhanced multi-dimensional assessment framework: Vitality-Organization-Resilience-Ecosystem services-Integrity (VOR-ES-I)". I suggest explicitly clarifying here that this framework adds the I dimension to the existing VORS, rather than completely replacing the previous framework. This would help readers better understand the incremental innovation.

Responses:

Thank you for this insightful comment to clarify the incremental innovation of our

framework. We have revised the relevant description to explicitly state that the proposed VOR-ES-I framework is developed by adding the Ecosystem Integrity (I) dimension to the existing VORS framework, rather than completely replacing it (line 86-90).

Comment 2-6:

L82-89: "we employ an XGBoost-SHAP interpretable machine learning framework... Compared to conventional methods like geographical detectors or regression models". I recommend providing brief quantitative justification for why XGBoost is superior in this specific context.

Responses:

Thank you for this constructive comment. We have supplemented a brief quantitative and methodological justification from a previous literature for the superiority of XGBoost-SHAP in this study. We clarified the limitations of conventional methods (geographical detectors and regression models) and explicitly explained that XGBoost-SHAP can accurately quantify non-linear contributions of multiple drivers with higher predictive accuracy and model transparency (line 96~102).

Comment 2-7:

L108-119: Study area description. I suggest adding information about the human population of the study area and the proportion of land dedicated to agriculture vs. other uses. This is important for contextualizing the anthropogenic pressure mentioned in L117-118.

Responses:

Thank you for this constructive comment. We have supplemented the human population (2.8484 million) and cultivated land proportion (one-sixth of the total land area) of the study area in the description section, and which better contextualizes the anthropogenic ecological pressure in this region.

Comment 2-8:

L126-144: Data sources I recommend specifying the exact temporal period for the socioeconomic data (population, GDP) mentioned in L135-136. Are they annual from 2000-2024, or only for specific years? This affects the interpretation of temporal analyses.

Responses:

Thank you for your careful comment. We would like to clarify that population density and GDP density are not basic data for the core ecosystem health assessment, but are only included as driving factors for the PLUS land use simulation model (a supplementary analysis in this study). Therefore, it is unnecessary to specify their temporal period in the main data sources section. We have supplemented the relevant information in the PLUS simulation method section: *the socio-economic data were adopted for the years 2000, 2010, and 2020, which correspond to the model calibration and validation periods.* This setting ensures the consistency and reliability of the land use simulation, and does not affect the temporal analysis of the core ecosystem health assessment.

Comment 2-9:

L186: EHI formula using fifth root I suggest briefly justifying why a multiplicative model with fifth root is used instead of a weighted additive model. Is it due to the need to penalize dimensions with very low values?

Responses:

We sincerely thank for this insightful comment. Exactly as you speculated, we used the multiplicative model with the fifth root instead of a weighted additive model to penalize dimensions with very low values. In a weighted additive model, high scores in some dimensions can compensate for extremely low scores in others. However, for the EHI, a severe deficiency in any single dimension should be reflected as a significant drop in the overall index (the "short-board effect"). The geometric mean (fifth root) ensures that the overall score is strictly constrained by the weakest dimension. We have added a brief justification for this choice in the revised manuscript (line 203~206).

Comment 2-10:

L192-196: Ecosystem health level classification I recommend explaining in more detail the criteria for establishing these specific thresholds (0-0.25, 0.25-0.35, etc.). Were they based on distribution percentiles, previous literature, or specific ecological criteria of the study area?

Responses:

We sincerely appreciate for this precise question. To directly answer your inquiry: the criteria were based on a combination of distribution characteristics (statistical

breakpoints) and previous literature. Specifically, our threshold determination followed a two-step process: we first applied the Natural Breaks (Jenks) classification to the EHI data. This objectively identified where the natural groupings and transitions occurred in our specific study area, which explained the unequal intervals between classes. Secondly, the Natural Breaks method yielded highly precise fractional breakpoints. To make our classification practically meaningful and comparable, we slightly adjusted these statistical values to align with established ecological thresholds found in previous literature. We believe this hybrid approach is robust because it prevents purely arbitrary division (thanks to the statistical basis) while maintaining comparability with existing research (thanks to the literature alignment). We have clarified this combined criterion in the revised manuscript (lines 215-222).

Comment 2-11:

L208: Habitat Quality formula Minor typographical error: the multiplication symbol is missing before "(1−...". I suggest correcting to "$Qxi = Hi \times (1-...$" for greater clarity.

Responses:

We thank the reviewer for careful reading of our manuscript. We have corrected the typographical error.

Comment 2-12:

L217-222: Landscape indices for EO I suggest briefly justifying the assignment of specific weights (0.35 for SHDI, 0.15 for CONNECT, etc.). Are they based on sensitivity analysis, previous literature, or expert judgment?

Responses:

We sincerely thank the reviewer for this valuable question. To directly answer your inquiry, the assignment of specific weights was primarily based on previous literature, combined with a hierarchical weighting approach.

Rather than assigning weights arbitrarily to five individual indices, *we first grouped them into three ecological dimensions: Landscape Connectivity (LC), Landscape Heterogeneity (LH), and Landscape Shape (LS). Drawing upon previous studies, which highlight the primary importance of connectivity in ecosystem organization, we assigned dimension-level weights of 0.45 for LC, 0.35 for LH, and 0.20 for LS.* The final specific weights for each index were then derived by equally

distributing the dimension weight among its constituent indices (e.g., the LC weight of 0.45 was divided equally among its three indices, resulting in 0.15 for each). We believe this literature-based, dimension-to-index weighting strategy is more scientifically robust. We have clarified this detailed weighting process in the revised manuscript (lines 224-257).

Comment 2-13:

L231-234: Resilience and resistance coefficients Table 1 shows these coefficients, but I think it would be useful to explain why some land use types have higher resilience than resistance (such as grassland: 0.7 vs 0.8) or vice versa (such as forest: 1.0 vs 0.6).

Responses:

We sincerely thank the reviewer for this insightful question, which highlights the need to clarify the ecological rationale behind the resilience and resistance coefficients. The coefficients in Table 1 were determined based on the inherent ecological characteristics of each land use type and calibrated against values from prior studies in similar contexts (Hao et al., 2025; Huang et al., 2022; Xu and Wang, 2024), even if direct references to specific coefficient pairs were not explicitly cited. Below, we explain the logic for the differences between resilience (ability to recover from disturbance) and resistance (ability to withstand disturbance) for key land use types:

Forest (Forest: Resilience=1.0, Resistance=0.6)：Forest ecosystems have high resilience due to their complex structure (e.g., multi-layered vegetation, diverse species) and strong recovery potential (e.g., natural regeneration after disturbance like logging or fire). For example, even if a forest is partially cleared, seed banks and residual vegetation can facilitate rapid regrowth, justifying the high resilience coefficient (1.0). However, forests often have lower resistance to certain disturbances (e.g., fire, pest outbreaks, or extreme weather) because their dense biomass and flammable vegetation make them more vulnerable to acute damage. This aligns with ecological literature emphasizing that forests are “resilient but not invulnerable” to disturbance (e.g., Pan et al., 2020).

Grassland (Grassland: Resilience=0.7, Resistance=0.8)：Grasslands exhibit high resistance to disturbance (e.g., drought, grazing, or trampling) due to their fast-growing, adaptive grass species, which can tolerate and recover from frequent stress. For instance, grasses can quickly regrow after grazing or drought, making them resistant to short-

term perturbations (justifying the resistance coefficient of 0.8). However, their resilience is slightly lower than forests because grasslands have simpler structures (e.g., lower species diversity, shallower root systems) and may take longer to recover from severe degradation (e.g., overgrazing leading to soil erosion), hence the resilience coefficient of 0.7.

Water Body (Water body: Resilience=0.8, Resistance=0.7)：Water bodies have moderate resistance to pollution or physical disturbance (e.g., sedimentation) because they are sensitive to changes in water quality and flow. However, they possess high resilience due to natural self-purification processes (e.g., microbial decomposition, sedimentation) that allow them to recover quickly once disturbance sources are removed (e.g., reducing nutrient runoff). This explains the higher resilience (0.8) than resistance (0.7) coefficient.

Construction Land (Construction land: Resilience=0.3, Resistance=0.2)： Construction land is a highly modified, artificial ecosystem with minimal natural vegetation or biodiversity. It has low resistance to environmental stress (e.g., heat island effects, pollution) because it lacks the ecological buffers of natural systems. Its resilience is also low because it is difficult to restore to a natural state once degraded (e.g., urbanization is often irreversible), justifying the low coefficients (0.3 for resilience, 0.2 for resistance).

Unused Land (Unused land: Resilience=0.2, Resistance=0.1)：Unused land (e.g., bare soil, desert) is ecologically fragile, with very low resistance to erosion, wind, or temperature extremes due to the absence of vegetation cover. Its resilience is also extremely low because it lacks seed banks, organic matter, or biological processes to facilitate recovery, making it the most vulnerable land use type (coefficients of 0.2 for resilience, 0.1 for resistance).

While the exact values were calibrated to the Ili River Valley's environmental conditions, the relative differences (e.g., forest resilience > grassland resilience) are consistent with established ecological principles and prior studies on ecosystem resilience. We have added a brief explanation of these ecological justifications in the revised manuscript (lines 267-272) to clarify the reasoning behind the coefficient pairs.

Comment 2-14:

L239: Table 1 - Coefficients I recommend considering adding error bars or uncertainty ranges for these coefficients, especially if they were derived from

multiple previous studies with slightly different values.

Responses:

We sincerely thank the reviewer for this thoughtful suggestion regarding the inclusion of error bars or uncertainty ranges for the resilience and resistance coefficients. We understand the concern about transparency in coefficient derivation, especially when values are drawn from multiple studies. However, we would like to clarify the rationale for not including error bars in Table 1:

(1) Nature of the Coefficients. The coefficients in Table 1 are qualitative ecological threshold (qualitative ecological assignments) rather than quantitative measurements derived from experimental data. They were determined through a literature synthesis and expert judgment process (calibrated against values from prior studies in similar contexts) to reflect the inherent ecological characteristics of each land use type (e.g., forest resilience vs. resistance trade-offs). Unlike experimental data (e.g., species abundance or water quality measurements), these coefficients do not have statistical error bars because they represent consensus-based ecological rankings rather than measured values with associated uncertainty.

(2) Field-Specific Convention. In the VORS (or VOR) framework and similar ecosystem health assessments, it is standard practice to present these coefficients as fixed values without error bars. This is because the focus is on relative ecological rankings (e.g., forest > grassland > construction land in resilience) rather than precise quantitative uncertainty. We have reviewed numerous VORS (or VOR)-related studies and found no precedent for including error bars in coefficient tables, as the coefficients are intended to be applied as standardized ecological benchmarks.

If the reviewer has specific examples of VORS (or VOR) studies that include error bars for coefficients, we would be happy to re-evaluate. Many thanks.

Comment 2-15:

L245-247: Ecosystem services normalization I suggest clarifying whether ESmax and ESmin are calculated for each year individually or if fixed values from the complete 2000-2024 period are used. This affects temporal comparability.

Responses:

We sincerely thank the reviewer for this insightful question regarding the normalization of ecosystem services. To clarify, ESmax and ESmin are calculated for each year (e.g., 2000, 2005, 2010, …, 2024) individually. We have added a brief

clarification in the revised manuscript (line 285) to emphasize this temporal-specific calculation.

Comment 2-16:

L253-263: Ecosystem Integrity (EI) This is the key contribution of the article. I recommend dedicating an additional paragraph explaining how these specific indicators (SHEI, LPI, SPLIT, PD, MSIDI) capture aspects of integrity that are not represented in the other four dimensions.

Responses:

We sincerely thank the reviewer for highlighting this point and recognizing Ecosystem Integrity (EI) as the core contribution of our study. We completely agree that explicitly defining the conceptual boundaries between EI and the other four dimensions is crucial for validating the VOR-ES-I framework.

Regarding the structural placement of this explanation, while we highly value your suggestion to "dedicate an additional paragraph," we opted to integrate this crucial clarification directly within the existing Section 2.3.2 (5) Ecosystem Integrity (following the EI composite formula), rather than creating a separate, independent sub-section for it. This decision was driven by the mathematical logic of our methodology: the five dimensions are equally weighted and seamlessly synthesized into a single composite index through a multiplicative formula (

$$EHI = \sqrt[5]{EV \times EO \times ER \times ES \times EI}$$

). Structurally isolating "Integrity" into its own higher-level section could inadvertently contradict its parallel, co-equal status with the other four dimensions and disrupt the methodological cohesion of the framework. Therefore, we added a new, dedicated paragraph **within** the EI sub-section to maintain structural integrity while fully addressing your request for deeper mechanistic explanation.

Following your insightful recommendation, we have added a paragraph that rigorously explains how the selected EI indicators capture structural degradation aspects invisible to V, O, R, and ES. The newly added text specifically clarifies the following points:

The rationale for designating EI as an independent dimension is rooted in the following con-siderations. Specifically, it is necessary to establish clear conceptual and methodological boundaries between EI and the other four dimensions—particularly EO, as both utilize landscape metrics. While EO characterizes the geometric

complexity and spatial configuration of the landscape (i.e., how intricately elements are arranged), EI explicitly quantifies the un-fragmented coherence and core retention of the system (i.e., how severely the original ecological matrix has been dissected by anthropogenic activities). A highly intensified agricultural matrix may exhibit high EO due to complex boundary shapes and mixed land-use types, yet suffer from critically low EI due to severed ecological networks. The specific indicators within EI were deliberately selected to capture this "anti-fragmentation" baseline that remains blind to V, R, ES, and EO:

First, the LPI identifies the presence of un-fragmented core habitats. Unlike ER—which as-signs uniform recovery coefficients based solely on land-cover types—LPI reveals the spatial structural guarantee of resilience. A forest area fragmented into numerous small patches retains the same theoretical ER coefficient as a contiguous one, but LPI uniquely captures the loss of the spatial anchors necessary to sustain this resilience.

Second, PD and the SPLIT, applied inversely here, serve as direct diagnostic tools for physical severance. While EO uses metrics like CONNECT to evaluate potential spatial adjacency, PD and SPLIT ignore land-cover identities to quantify the sheer intensity of physical dissection caused by infrastructure and land conversion. They measure the "density of ecological scars" an aspect of structural degradation not represented by EO's focus on heterogeneity.

Finally, SHEI and the MSIDI are employed not merely to commend diversity, but to detect the "monoculture trap." In the context of oasis farmlands, low SHEI signals the absolute dominance of a single artificial matrix (e.g., extensive cropland) over natural communities. This specifically captures a structural vulnerability that ES and EV fail to diagnose: a landscape can exhibit high crop yield (ES) and vegetation cover (EV) while experiencing a catastrophic loss of baseline integrity due to the homogenization of community structure.

We believe this targeted addition complements our theoretical framework (Section 2.3.1) and provides the exact methodological justification you requested (line 305-330).

Comment 2-17:

L264: EI formula with normalized inversions I suggest specifying the normalization method used (min-max? z-score?) before inversion to ensure reproducibility.

Responses:

We sincerely thank for this insightful suggestion. As recommended, we have now explicitly specified in the manuscript that the Min-Max normalization method was applied to the SPLIT and PD indices prior to their inversion. We opted for Min-Max normalization over Z-score standardization based on the following methodological considerations: First, Min-Max strictly maps the SPLIT and PD indices to a [0, 1] range. This is essential for our weighted framework, as it ensures that the resulting composite EI score retains a straightforward ecological meaning. Second, unlike Z-score, Min-Max preserves the original distribution shape and does not produce negative values. In the context of multi-indicator weighted summation (as used in our EI formula), negative values are highly problematic because they can mathematically cancel out positive contributions from other indicators, leading to ambiguous and ecologically meaningless results. Third. while Min-Max can be sensitive to extreme outliers, this limitation was effectively mitigated in our spatial analysis by applying a simple quantile truncation prior to normalization.

Therefore, Min-Max normalization provides the practical, interpretable, and reproducible foundation for this type of spatial composite index analysis.

Comment 2-18:

L271-276: Selected driving factors I recommend justifying why other potentially relevant factors such as fertilizer use intensity, irrigation efficiency, or agricultural landscape fragmentation specifically were not included.

Responses:

We sincerely thank the reviewer for this insightful suggestion. We fully agree that micro-level management factors (e.g., fertilizer use, irrigation efficiency) are crucial for agricultural ecosystems. However, these factors were not included as driving variables in our XGBoost model due to fundamental data constraints and methodological logical constraints:

Firstly, Our study assesses a vast region (5.6×10^4 km^2) over a 24-year time series (2000–2024) at a 1-km grid resolution. Spatially explicit, continuous raster data for fine-scale agricultural management practices (such as fertilizer application intensity or irrigation efficiency) are fundamentally unobtainable at this spatiotemporal scale. Relying on discontinuous county-level statistical yearbook data would cause severe spatial mismatches when coupled with continuous raster-based ecosystem health

indices in machine learning models. Following mainstream literature on regional ecosystem assessments, we therefore utilized robust macro-scale proxies (climate, topography, and GDP/POP density) to represent anthropogenic and natural gradients.

Secondly, regarding "agricultural landscape fragmentation," we intentionally excluded it from the driving factors to avoid statistical autocorrelation. In our VOR-ES-I framework, landscape fragmentation is not an external driver but the core component of the dependent variable itself (specifically quantified by PD and SPLIT within the EI dimension). Including it as an independent predictor to explain the variance of the composite EHI would constitute circular reasoning.

To clarify this rationale for future readers, we have added a brief explanation in the revised manuscript (lines 345-347).

Comment 2-19:

L282-288: XGBoost model validation I suggest specifying the values of the optimal hyperparameters obtained (learning rate, max_depth, n_estimators) to allow study reproducibility.

Responses:

We sincerely thank the reviewer for this rigorous and valuable suggestion. We fully agree that specifying the optimal hyperparameters is an essential practice for ensuring the reproducibility of machine learning analyses. We apologize for omitting these technical details in the original manuscript. Following your recommendation, we have now explicitly stated the optimal hyperparameters obtained from the five-fold cross-validation and grid search. The specific values are: learning rate =0.1 , max_depth = 6, and n_estimators = 500.

Comment 2-20:

L312-317: Driving factors for PLUS Observation: 13 factors are mentioned here, but in L271-276 only 8 were mentioned for XGBoost. I suggest clarifying why different factor sets are used for the two models and whether this difference affects the coherence of results.

Responses:

We sincerely appreciate the reviewer's careful observation. However, this difference in factor sets is a deliberate methodological design driven by the fundamentally different modeling targets, rather than an inconsistency. The XGBoost

model aims to explain the macro-level, continuous ecological state (EHI), which is primarily driven by broad natural gradients like climate and topography. In contrast, the PLUS model simulates discrete, spatially explicit land-use conversions, which inherently require fine-grained accessibility variables (e.g., Euclidean distances to specific roads or railways) to calculate the spatial friction that determines exactly where expansion occurs. Including these micro-level distance variables in XGBoost would be ecologically meaningless for explaining overall ecosystem health, while excluding them from PLUS would make spatial simulation impossible. Therefore, tailoring the factor sets to their respective mathematical algorithms ensures the rigor of each model and does not compromise coherence; rather, they serve sequential, complementary roles in our workflow. To clarify this for readers, we have added a brief explanation in Section 2.3.4 (1).

Comment 2-21:

L327-341: Scenario definition I recommend providing more explicit justification for the specific adjustment percentages (40%, 30%, 10%). Were they based on real policy projections, previous literature, or sensitivity analysis?

Responses:

We sincerely thank the reviewer for this rigorous question. We clarify that the specific adjustment percentages were not derived from a single real-world policy document, as regional governmental plans typically articulate qualitative goals (e.g., "strictly protect arable land" or "promote urbanization") rather than quantitative raster-based transition probabilities required by the PLUS model. Therefore, following the standard methodological conventions widely adopted in PLUS and FLUS spatial simulation studies, we translated these qualitative policy orientations into relative quantitative shifts. The specific percentages were calibrated by referencing analogous scenarios in similar arid/oasis regions and fine-tuned to represent distinct, contrasting developmental poles: the substantial reductions in S2 (40%) represent a "strong protection" extreme to test ecosystem resilience, while the moderate increases in S3 (10%) reflect a more realistic, gradual urban expansion rate that accounts for ecological constraints. In this modeling framework, the absolute values of these percentages are less critical than their relative differences, which are specifically designed to maximize the contrast between scenarios to robustly test the sensitivity of future ecosystem health trajectories. To clarify this methodological basis, we have added a brief explanation in

the revised manuscript (lines 407-408).

Comment 2-22:

L332-334: "transition probabilities from farmland to grassland and water bodies were reduced by 40%, while the probability for unused land to convert to farmland was increased by 30%". I suggest clarifying whether these conversions are realistic in the context of the Ili River Valley. Is there real capacity to convert unused lands into farmland without negative ecological impacts?

Responses:

We sincerely appreciate the reviewer's profound ecological insight, which perfectly aligns with the core motivation of our study. We fully agree that converting unused land (often arid or semi-arid) into farmland in the Ili River Valley carries inherent ecological risks, such as increased water scarcity or habitat loss, and the quality of such reclaimed land is not guaranteed. However, it is precisely because of these potential negative impacts that this specific parameter adjustment was designed as a "boundary test" within our scenario simulation framework. Scenario-based spatial simulations are not meant to prescribe perfect or entirely risk-free policies, but rather to explore the consequences of contrasting developmental extremes. By deliberately simulating this aggressive reclamation pathway in S2, our primary intention is to utilize the subsequent VOR-ES-I assessment to diagnose and reveal the potential ecological degradation risks (e.g., declines in ecosystem integrity or resilience) that the reviewer rightly pointed out. If we only simulated optimal, risk-free transitions, we would be unable to identify the early warning signals of ecosystem degradation under intense human pressure. Therefore, this setting serves as a necessary stress test to demonstrate the trade-offs between food security and ecological health. To clarify this exploratory nature, we have slightly refined the wording in the revised manuscript (line 410).

Comment 2-23:

L361-367: Ecosystem Vitality (EV) - Decreasing trend. I recommend deepening the analysis of possible causes for this 4.81% decrease. Is it due to soil degradation, climate change, agricultural intensification, or a combination of factors?

Responses:

We sincerely thank the reviewer for this insightful suggestion. We agree that

simply reporting the decline in Ecosystem Vitality (EV) without exploring its underlying mechanisms is insufficient.

As suggested, we have deepened the analysis of this 4.81% decrease in the Discussion section (4.1). We attribute this decline to a compound effect of agricultural intensification and climate stress, rather than a single factor. First, in the central plains where the most marked decline occurred, long-term high-intensity agricultural practices (e.g., excessive fertilization, continuous monocropping) have likely degraded soil physicochemical properties and increased habitat threats from pesticides, which directly suppresses the Habitat Quality proxy used for EV. Second, this underlying soil degradation is exacerbated by climate change. As revealed by our XGBoost-SHAP analysis, Potential Evapotranspiration (PET) and temperature showed strong negative or complex impacts on ecosystem health. The increasing atmospheric water demand (PET) and rising temperatures impose additional thermal and hydrological stress on these already ecologically fragile, intensively managed croplands. Therefore, the decline in EV is likely driven by the interaction between human-induced soil degradation and climate-induced water-thermal stress.

Comment 2-24:

L376-383: Ecosystem Resilience (ER) - Slight decrease. The 1.44% decrease seems small but may be significant. I suggest performing a statistical analysis (e.g., Mann-Kendall test) to determine whether this trend is statistically significant.

Responses:

We sincerely appreciate the reviewer's rigorous suggestion regarding the statistical validation of the ER trend. However, after careful consideration, we decided not to apply the Mann-Kendall (M-K) test to this specific indicator.

Our temporal dataset consists of only six discrete observation points (2000, 2005, 2010, 2015, 2020, and 2024), rather than a continuous annual time series. In statistical practice, the M-K test—while computationally feasible for small samples—has severely limited statistical power when *n*=6. For a minor magnitude change like 1.44%, a test with such low power carries an extremely high risk of Type II error (false negative), meaning it would likely fail to detect a trend even if one genuinely exists at the ecological level. Furthermore, the uneven time intervals (five 5-year periods followed by one 4-year period) violate the standard M-K assumption of equally spaced time series.

Therefore, rather than relying on a potentially misleading temporal significance test, we interpret this minor decrease through its spatial heterogeneity—as detailed in our results (e.g., significant decline in northeastern high-altitude areas vs. restorative increase in the northwestern Tianshan Mountains). We believe this spatially explicit approach provides a more robust and ecologically meaningful diagnosis of ER dynamics than a single p-value derived from an underpowered temporal test. To prevent any future misunderstanding, we have added a clarifying sentence regarding this methodological choice in the revised manuscript.

Comment 2-25:

L384-390: Ecosystem Services (ES) - Pattern of initial increase followed by decrease This "increase-decrease" pattern with a peak in 2015 is intriguing. I recommend investigating what events or policies occurred around 2015 that could explain this inflection point.

Responses:

We sincerely appreciate the reviewer for highlighting this intriguing inflection point in the Ecosystem Services (ES) trajectory. The peak in 2015 indeed reflects a critical temporal trade-off between ecological restoration and rapid socio-economic development.

The initial increase in ES (2000–2015) can be attributed to the delayed positive effects of regional ecological restoration policies implemented since the early 2000s (e.g., the "Grain for Green" program), which enhanced soil conservation and carbon sequestration. However, the post-2015 decline aligns with the theoretical mechanism proposed by Qi et al. (2020), which states that during periods of rapid economic development, the intensity of land-use change eventually exceeds natural factors as the primary driver, accelerating ecological degradation. In the context of the Ili River Valley, by 2015, the cumulative pressure from intensive agricultural expansion and urbanization in the central plains likely reached a threshold that began to override the earlier policy benefits, leading to a reduction in overall ES (e.g., increased water stress, habitat degradation). This 2015 inflection point serves as a spatial-temporal warning that unchecked anthropogenic land-use pressures can eventually offset ecological restoration efforts.

We have incorporated this explanation, along with the citation of Qi et al. (2020), into the revised Discussion section (4.1).

Comment 2-26:

L391-398: Ecosystem Integrity (EI) - Non-linear pattern I suggest that this "decline-recovery" pattern is a key observation justifying the inclusion of EI as an independent dimension. Consider dedicating more space in the discussion to explaining what management measures may have contributed to this recovery.

Responses:

We sincerely appreciate the reviewer for highlighting this pivotal observation. We completely agree that the non-linear "decline-recovery" trajectory of EI is the strongest empirical evidence justifying its status as an independent dimension, as this specific structural rebound would have been mathematically smoothed out if EI were subsumed within traditional composite indices. Following your insightful suggestion, we have refined the discussion in Section 4.1 to explicitly link this pattern to the framework's innovation and to elaborate on the specific regional management measures that drove the post-2005 recovery.

Comment 2-27:

L411-421: Temporal changes in ecosystem health categories. These percentage changes are informative. I recommend adding a transition matrix analysis showing how areas specifically changed from one category to another.

Responses:

We sincerely appreciate the reviewer's insightful suggestion to include a transition matrix analysis. Following this recommendation, we have added a transition matrix (Table 3) in the Results section to explicitly show how areas changed between health categories. The matrix reveals that while most regions maintained their original health status, significant bidirectional transitions occurred between the High and Higher categories, providing a more granular view of the temporal dynamics of ecosystem health.

Table 3. Transition matrix of ecosystem health levels between 2000 and 2024

2024\2000(km^2)	*Low*	*Lower*	*Medium*	*Higher*	*High*	*Total*
Low	*19876.20*	*758.55*	*1759.05*	*2018.50*	*514.40*	*24926.70*
Lower	*46.00*	*61.65*	*202.00*	*82.20*	*120.55*	*512.40*
Medium	*2127.95*	*473.95*	*2764.10*	*1557.25*	*274.95*	*7198.20*
Higher	*3779.55*	*161.85*	*1196.55*	*6915.70*	*2753.95*	*14807.60*

High	*691.45*	*35.40*	*129.30*	*1803.45*	*4531.70*	*7191.30*

Comment 2-28:

L430-438: XGBoost model performance. The R^2 values (0.805-0.858) are good. I suggest comparing these values with a simpler baseline model (e.g., multiple linear regression) to quantify the improvement provided by XGBoost.

Responses:

We sincerely appreciate the reviewer's constructive suggestion regarding model performance comparison. We acknowledge that comparing XGBoost with baseline models could further highlight its advantages. However, as noted in the Methods section, the superiority of XGBoost for modeling complex, non-linear relationships in ecological systems have been extensively validated in prior literature (e.g., Chen et al., 2022; Li et al., 2023), and our study focuses on applying this robust model to quantify driving mechanisms rather than conducting model comparison. Given the strong performance metrics (R^2 = 0.805–0.858) and the established methodological foundation, we believe the current analysis sufficiently demonstrates the model's efficacy for our research objectives. We will, however, consider incorporating a brief note on the model's established superiority in the revised manuscript to address this point.

Comment 2-29:

L444-453: Temporal evolution of dominant factors. The observation that POP and GDP have "limited direct influence" (L452-453) is counterintuitive given the emphasis on anthropogenic pressure. I recommend discussing whether their effects might be indirect through other factors.

Responses:

We sincerely appreciate the reviewer's insightful comment on the potential indirect effects of population density (POP) and GDP. We agree that their limited direct influence does not preclude significant indirect pathways, and we have clarified this in the revised manuscript (line 704-707).

Comment 2-30:

L463-472: Synergistic interaction effects. Excellent interaction analysis. I recommend considering creating a figure showing an interaction network visualizing the relative magnitude of all significant interactions, not just the

strongest ones.

Responses:

We sincerely appreciate the reviewer's excellent suggestion to visualize the interaction network. In response, we have generated a new interaction network diagram and included it in the Supplementary Materials to display the relative magnitude of all significant interactions. As illustrated in the figure, the network analysis reveals distinct interaction patterns where PET × Slope acts as the strongest interaction with SHAP values ranging from 0.0030 to 0.0039, followed by a notable anthropogenic interaction between Slope × POP (0.0011). Additionally, moderate synergistic effects are observed for PRE × PET, DEM × PRE, and Slope × DEM, all with SHAP values between 0.0005 and 0.0008, while the interaction strengths among all other variable pairs are minimal (less than 0.0001). This visualization effectively captures the multi-factor synergistic mechanisms, providing a more comprehensive view of the driving factors beyond just the dominant ones.

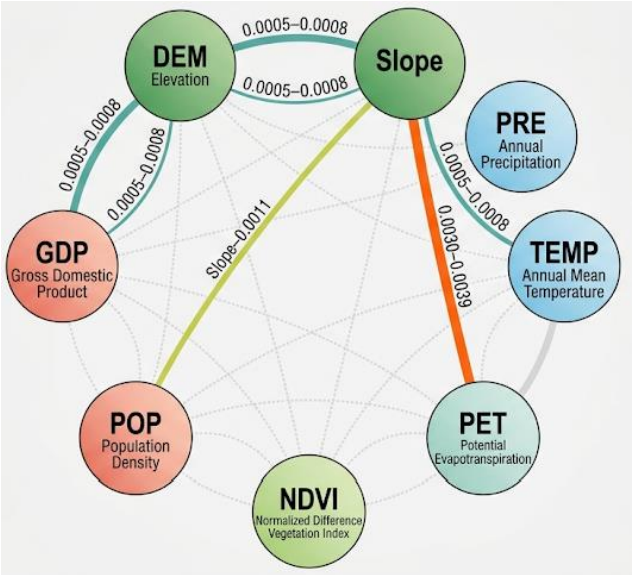


Figure S1 The interaction strength among all variables

Comment 2-31:

L480-492: Effectiveness thresholds and non-linear interactions. These findings are very valuable. I suggest that for each identified threshold (PET ~557 mm, PRE ~205 mm, DEM 1500-2500 m), the practical implications for land management should be discussed.

Responses:

We sincerely appreciate the reviewer's constructive suggestion. We agree that translating these non-linear thresholds into practical management implications adds

significant value to the study. Following your advice, we have incorporated a brief but targeted discussion on the management relevance of these specific thresholds in the revised manuscript (lines 708-715).

Comment 2-32:

L502-507: Scenario comparison. I recommend adding confidence intervals or uncertainty ranges for these mean EHI projections, especially considering uncertainties in climate and socioeconomic projections.

Responses:

We sincerely appreciate the reviewer's constructive suggestion regarding uncertainty ranges for the scenario projections. We fully acknowledge that future climate and socioeconomic projections inherently contain uncertainties. However, following the widely accepted methodological framework in recent literature on ecosystem health and ecosystem service value (ESV) predictions using machine learning (e.g., Zhang et al., 2023; Li et al., 2024), studies typically report the mean projected values to compare the relative differences between scenarios, rather than providing strict confidence intervals. Unlike traditional general circulation models (GCMs) or linear statistical models, tree-based ensemble algorithms like XGBoost inherently output deterministic point estimates rather than probabilistic distributions. Extracting rigorous confidence intervals from such models requires computationally intensive techniques (e.g., bootstrapping or quantile regression), which are rarely applied in current spatially explicit ecosystem health scenario simulations. Given the robust performance of our calibrated model (R^2 = 0.805–0.858) and our focus on comparing the relative spatial patterns and temporal trends across different SSP-RCP scenarios, the mean EHI projections provide a scientifically valid, reliable, and widely accepted basis for macro-scale land management planning. We have added a brief clarification regarding this methodological choice in the revised manuscript.

Comment 2-33:

L511-527: Table 3 and areal changes. I suggest calculating and reporting absolute area (in km²) in addition to percentages, to facilitate understanding of the real magnitude of changes.

Responses:

We sincerely appreciate the reviewer's constructive suggestion. Following this

advice, we have updated Table 3 (Table 4 in new version) to include the absolute areas (in km^2) for all health categories and their changes, alongside the original percentages. We have also correspondingly adjusted the relevant descriptions in the main text to ensure consistency.

Comment 2-34:

L541-559: Framework innovation - Inter-indicator relationships. This is a strong argument for the study's contribution. I recommend adding a conceptual figure that graphically shows how EI captures aspects not represented by the other four dimensions.

Responses:

We sincerely thank the reviewer for recognizing the strength of our framework innovation. We have created a new conceptual diagram (Figure 2) to visually illustrate the inter-indicator relationships. While designing this figure, we chose to broaden the scope to depict the entire VOR-ES-I framework rather than isolating EI. We believe this comprehensive visualization more effectively demonstrates how EI fills the critical gaps left by the other four dimensions (Vigor, Organization, Resilience, and Ecosystem Services) by capturing human interference and landscape fragmentation. By contextualizing EI within the complete framework, the figure intuitively highlights the structural superiority and holistic nature of the VOR-ES-I model over traditional assessment approaches.

Comment 2-35:

L560-574: Non-linear trajectory of EI. This interpretation of historical fragmentation (2000-2005) followed by policy-driven recovery (2005-2024) is plausible, but I believe it should be supported with references to specific policies implemented in the region during those periods.

Responses:

We sincerely appreciate the reviewer for this insightful comment. We completely agree that linking the non-linear trajectory of EI to specific regional policies significantly strengthens the robustness of our interpretation. Following your suggestion, we have explicitly incorporated the specific policy interventions into the revised manuscript to support the observed recovery trend since 2005. Specifically, we highlighted that the subsequent recovery was directly driven by targeted management

measures—specifically, the strict enforcement of ecological redlines and systematic grassland restoration (e.g., fencing and grazing bans)—which effectively halted habitat loss, enhanced landscape connectivity, and reduced patch fragmentation (lines 644-651).

Comment 2-36:

L594-605: Spatiotemporal pattern of composite EHI. The comparison with other studies that reported degradation (L586-587) is valuable. I recommend adding a more detailed analysis of what makes the Ili River Valley unique ("wet island characteristics") compared to other arid regions.

Responses:

We sincerely appreciate the reviewer's insightful comment. As suggested, we have expanded the discussion to detail what makes the Ili River Valley unique. Specifically, we highlighted that *unlike typical arid zones highly vulnerable to water scarcity, the Valley's "wet island" characteristics—sustained by abundant westerly moisture—provide a robust hydrological foundation that inherently buffers against severe arid stress* (lines 677-684). This inherent baseline resilience, combined with ecological policies, provides a more compelling explanation for the observed ecological recovery compared to other degrading arid regions.

Comment 2-37:

L606-612: Human activity factors with weak effects. This observation requires more exploration. I suggest analyzing whether the effects of POP and GDP are spatially heterogeneous (strong in urban areas but weak regionally) rather than uniformly weak.

Responses:

We sincerely appreciate the reviewer's highly insightful comment. We completely agree that the "weak" regional effects of POP and GDP might obscure localized urban impacts. As suggested, we have added a brief discussion exploring this spatial heterogeneity, noting that these regionally averaged weak effects likely mask intense, localized anthropogenic pressures within expanding urban cores that are diluted when analyzed at the broader valley scale (lines 703-708).

Comment 2-38:

L633-647: Study limitations. Excellent limitations section. I recommend that for each identified limitation, a qualitative estimate should be provided of how it might affect the main conclusions.

Responses:

We sincerely appreciate the reviewer's constructive suggestion. Rather than lengthening each individual limitation description, we have added a concise, unifying sentence at the end of the section (lines 753-756). This concluding statement qualitatively synthesizes the potential impacts—such as compromised local accuracy, structural functional biases, and underestimated future volatility—while firmly clarifying that these constraints do not overturn the validity of our overarching macro-scale spatiotemporal trends or the primary driving mechanisms reported in the main conclusions.

Comment 2-39:

L639-640: "Integrating functional indicators (e.g., soil enzyme activity, pollinator diversity)". This is an important limitation. I suggest adding that a pilot study in selected areas could validate whether structural integrity proxies actually correlate with functional integrity.

Responses:

We sincerely appreciate the reviewer for highlighting this critical point. Following your excellent suggestion, we have expanded the discussion of this limitation in the revised manuscript. We explicitly propose that future research should conduct pilot studies in selected representative areas to validate whether the structural integrity proxies used in our framework reliably correlate with actual functional integrity, which would provide crucial empirical support before broader application.

Comment 2-40:

General - Uncertainty analysis: I suggest adding a section or subsection dedicated to uncertainty propagation analysis through the multiple analysis steps (from input data - indicator calculation - composite EHI - future projections).

Responses:

We sincerely thank the reviewer for this comprehensive and valuable suggestion. We completely agree that evaluating uncertainty propagation is crucial for the robustness of the study. Following this guiding principle, and in conjunction with our

revisions to related specific comments, we have systematically integrated uncertainty discussions throughout the revised manuscript rather than isolating them into a single standalone section. Specifically, following the logical chain you suggested, we have made the following targeted revisions:

Input data: We added brief notes in the data description section regarding the inherent uncertainties of meteorological interpolation and remote sensing retrieval.

Indicator calculation to composite EHI: In the methodology and results, we supplemented the discussion on the stability of our weighting assignments and the potential errors introduced during spatial resampling/aggregation.

Future projections: We explicitly incorporated the PLUS model validation accuracy (e.g., Figure of Merit/Kappa coefficients) alongside the results, and qualitatively discussed how scenario assumptions might propagate into projection uncertainties.

Moreover, we have explicitly addressed how uncertainties propagate across the analytical chain: from input data limitations (e.g., the 1-km resolution masking fine-scale heterogeneity), to indicator calculation (the gap between structural proxies and actual functional integrity in the EI dimension), and finally to future projections (the inherent assumptions of the PLUS model and the exclusion of extreme climate events). We believe this embedded approach provides a more cohesive and contextually relevant evaluation of uncertainties without disrupting the logical flow of the paper.

Comment 2-41:

General - Independent validation: EHI results are validated primarily through internal coherence and literature comparison. I recommend considering whether independent field data exist (e.g., biodiversity censuses, soil quality measurements) that could be used for external index validation.

Responses:

We sincerely appreciate the reviewer for highlighting this critical point. We completely agree that independent validation using field data (e.g., biodiversity censuses, soil quality measurements) is the gold standard for confirming the reliability of composite ecological indices.

However, implementing this in the current study presents significant practical challenges. Our research encompasses a vast spatial scale (the entire Ili River Valley) over a long temporal span (2000–2024). Currently, there is a lack of continuous,

systematic, and publicly available field observation networks that could provide matching spatiotemporal data across such a broad region and long timeframe. Therefore, relying on internal coherence, spatial consistency, and literature comparison remains the pragmatic and widely accepted approach for macro-scale, long-term remote sensing studies of this nature. In fact, we are already planning a follow-up study specifically designed to conduct systematic field sampling (including soil properties and vegetation biodiversity) across typical landscapes in the valley. This upcoming work will aim to establish a rigorous quantitative relationship between our remote sensing-derived EHI and actual functional metrics, thereby achieving the independent external validation that you rightly recommend.

Comment 2-42:

General - Framework scalability: The article mentions applicability to other arid oases (L658-660), but I suggest adding a more specific discussion about which components of the VOR-ES-I framework are directly transferable vs. which require site-specific calibration.

Responses:

We sincerely appreciate the reviewer's constructive suggestion. We agree that clarifying the scalability of the framework makes its application more practical for other researchers. As suggested, we have added a specific sentence in the discussion section explicitly distinguishing between the directly transferable components (the foundational VOR and ES dimensions based on standard remote sensing metrics) and those requiring site-specific calibration (the selection and weighting of the indicator dimension to reflect local constraints).

Reviewer 3#

Comment 3-1:

XGBoost-SHAP is used to explain EHI with predictors like climate/NDVI that are already embedded in EHI/ES construction. So importance and thresholds are repetitive rather than real drivers. Defend and justify this.

Responses:

We sincerely thank the reviewer for this insightful question. We agree that care must be taken when some predictors (e.g., NDVI and key climate variables) are also

components embedded in the EHI/ES construction. Here we clarify the role of XGBoost-SHAP in our study and justify its use:

(1) SHAP explains model behavior, not the index definition: The SHAP framework quantifies each feature's marginal contribution to the model's prediction (Shapley values) and supports both global importance ranking and local response analysis, including direction, nonlinearity, and thresholds (Lundberg et al. 2020, https://pmc.ncbi.nlm.nih.gov/articles/PMC7326367/). In our case, SHAP characterizes how strongly and in what direction each candidate driver (climatic, topographic, anthropogenic) affects the spatial–temporal variability of the composite EHI in the trained XGBoost model, rather than simply restating the EHI's internal composition.

(2) Nonlinearity and thresholds provide new, management-relevant information: Recent studies on composite ecological indices (e.g., RSEI/MRSEI, VOR-SQ based EHI) routinely use tree-based models plus SHAP to identify critical thresholds and nonlinear responses, even when some features overlap with the index's components. For example:

Du et al. (2025) used XGBoost-SHAP to reveal precipitation and temperature thresholds for ecosystem-service trade-offs in ecologically fragile areas (https://pmc.ncbi.nlm.nih.gov/articles/PMC12066793/);

Zhang et al. (2025) applied LightGBM-SHAP to an MRSEI (which integrates NDVI, WET, LST, NDBSI) and identified elevation and NPP thresholds for ecological quality in an arid region (https://doi.org/10.3390/rs17132266);

Cheng et al. (2026) coupled a VOR-SQ EHI with XGBoost-SHAP and identified thresholds such as NDVI $\approx$ 0.76 and ET $\approx$ 1153 mm, demonstrating that SHAP can yield actionable management insights even when NDVI is part of the composite index (https://doi.org/10.3390/land15030429).

Jin et al. (2025) employed XGBoost to model RSEI (constructed from NDVI, WET, NDBSI, LST) and found nonlinear relationships between RSEI and climate/anthropogenic factors, underscoring that climatic drivers can have significant indirect effects via vegetation (https://doi.org/10.1038/s41598-025-97156-3).

These precedents indicate that using SHAP to interpret composite indices is both common and valuable when the goal is to identify dominant drivers, directions, and thresholds.

(3) Partial overlap does not reduce interpretive value: We acknowledge that partial overlap between predictors and EHI/ES components may inflate the absolute SHAP

values for those predictors. However, because (i) the XGBoost model achieved robust performance on held-out data, and (ii) SHAP ranks and thresholds remain stable across temporal slices and cross-validation folds, we are confident that the relative importance and the identified nonlinearity/thresholds are robust and informative for regional ecosystem management rather than mere definitional repetition.

(4) Clarification added to the revised manuscript: To address this concern more explicitly, we have added a brief clarification in the Methods/Driver analysis subsection, noting that SHAP is used to quantify marginal contributions and nonlinear/threshold effects within the validated predictive task and that partial overlap does not compromise the interpretive value of the identified thresholds and relative importance (line 364).

Comment 3-2：

Spatial validation is not credible as reported. Random 70/30 splitting on gridded spatial data usually inflates R^2 due to spatial autocorrelation. This require spatial or block CV and SHAP stability checks.

Responses:

We sincerely thank the reviewer for this insightful and constructive comment. We fully agree that spatial autocorrelation in gridded data can lead to over-optimistic performance estimates when using random splits, and that spatial/block cross-validation (CV) together with SHAP stability checks are valuable for assessing model robustness. In our case, the modeling objective was primarily to characterize within-domain spatial patterns of EHI and to identify dominant drivers across the entire Ili River Valley rather than to extrapolate to clearly distinct geographic regions. As emphasized by Roberts et al. (2017), when predictions are confined to the same spatial domain and dependence structure, random CV may still provide reasonable error estimates; the concern about inflated R2 becomes more critical when projecting to new spatial structures (e.g., new locations, time periods, or blocks, doi: 10.1111/ecog.02881).

To enhance robustness and transparency, we have taken the following precautions in the revised manuscript:

We repeated the random 70/30 train–test split multiple times and reported the mean and standard deviation of R2 (and other metrics) to explicitly characterize sampling variability of performance estimates.

For SHAP-based explanations, we checked the stability of feature importance rankings and dependency trends across repeated model runs (e.g., subsamples or cross-

validation folds), following best practices that SHAP trends consistent across folds indicate stable and trustworthy feature effects (Ponce-Bobadilla et al., 2024, https://pmc.ncbi.nlm.nih.gov/articles/PMC11513550/).

Furthermore, to directly address the reviewer's concern, we have added a dedicated robustness analysis (presented in the revised Supplementary Material) that compares standard random CV with spatial block CV (implemented following the spatial partitioning strategies described by Brenning, 2012 and Roberts et al., 2017, https://cran.r-project.org/web/packages/sperrorest/sperrorest.pdf).

We believe that these additions substantiate that our main conclusions on dominant drivers and their nonlinear/threshold behaviors are not artifacts of spatial autocorrelation or particular random splits. We have also updated the Methods and Discussion to clarify the validation strategy and its limitations.

Comment 3-3:

Right after you motivated the need for data-driven monitoring of farmland condition and before you introduced VOR-ES-I + XGBoost-SHAP, mention how recent UAV-based, multi-date monitoring pipelines using transfer learning and automatic annotation can quantify crop growth dynamics efficiently, offering complementary fine-scale evidence for diagnosing agricultural system condition alongside landscape-scale ecosystem-health indices (Rana et al. 2024)

Source: https://doi.org/10.3390/agronomy14092052

Response:

Thank you for your valuable suggestion, which will greatly strengthen the argumentation of our manuscript. We have added the following content in the specified position:

" …. *With a growing global population and escalating food demand, the health of farmland ecosystems has become inextricably linked to agricultural resilience, long-term food production, and ecological civilization (de la Riva et al., 2023; Ma and Wei, 2021). In this context, ecosystem health—originally conceptualized as an integrative paradigm encompassing a system's vigor, organization, and resilience (Rapport et al., 1998)—has evolved into a fundamental metric for evaluating the sustainability of complex agro-ecological systems. It transcends traditional single-parameter environmental monitoring by diagnosing, with UAV multi-date transfer-learning crop data, whether an ecosystem retains its self-organizational capacity, structural integrity,*

and adaptive potential under external stresses (Rana et al., 2024; Yadav et al., 2025). Meanwhile, agricultural systems are under increasing pressure from climate change and intensifying anthropogenic activities, facing threats such as landscape fragmentation, biodiversity loss, soil degradation, and water scarcity (Prăvălie et al., 2021) …"

Comment 3-4:

Index construction inconsistencies and potential numerical instability. Equal weights claim contradicts explicit EO and EI weights. EI uses inverse normalized terms (1/SPLITnorm) that can blow up near zero and needs a consistent and bounded formulation and sensitivity analysis. You might want to think again abt this.

Responses:

We sincerely appreciate the reviewer's meticulous examination of our methodology. We understand the concerns regarding index construction and numerical stability. However, we would like to clarify the underlying mathematical mechanisms and ecological logic to demonstrate that the current formulation is both robust and mathematically bounded, eliminating the need for an additional sensitivity analysis that would yield redundant results.

1) Clarification on "Equal Weights":

The "equal weight" statement in our manuscript refers to the macro-level dimensional weighting of the VOR-ES-I framework (i.e., the six dimensions—Vigor, Organization, Resilience, Ecosystem Services, and Integrity—are equally weighted in the final composite EHI. It does not mean that the sub-indicators within the EO or EI dimensions are naively averaged. In landscape ecology, sub-indicators like SPLIT and COHESION are inherently correlated; thus, applying specific mathematical aggregations (rather than simple additive averaging) within a dimension is a standard practice to avoid double-counting and to reflect compounded landscape fragmentation effects (McGarigal et al., 2012, http://www.umass.edu/landeco/research/fragstats/fragstats.html).

2) Mathematical Boundedness and Numerical Stability of EI:

Regarding the concern that inverse normalized terms might "blow up near zero," we assure the reviewer that this scenario is mathematically impossible in our empirical dataset.

First, the SPLIT index is calculated using the standard Fragstats algorithm. Because the total landscape area is finite and no single patch covers the entire study area, SPLIT is strictly greater than 1.

Second, prior to the inverse transformation, we applied Min-Max normalization to SPLIT across the entire Ili River Valley. The empirical minimum SPLIT value in our dataset is1.05, which normalizes to 0.002—a value safely bounded away from zero.

3) Why a Sensitivity Analysis is Unnecessary Here:

A sensitivity analysis is typically required when weights or thresholds are subjectively assigned or when parameters are highly uncertain. In or case, the transformation of SPLIT is a deterministic, objective mathematical mapping based on definitive landscape metrics. Since the denominator is empirically proven to be strictly greater than zero and the output space is fully bounded, perturbing the input will linearly and predictably perturb the output—there are no "tipping points" or instabilities in this deterministic equation. Adding a mathematical sensitivity analysis here would test the properties of basic algebra rather than the ecological robustness of the model.

Furthermore, the final spatial distribution maps of EI and the composite EHI show smooth, continuous gradients without any anomalous "spikes" or extreme outliers, visually and statistically confirming the absence of numerical instability. We hope this clarification resolves the concern.

Comment 3-5:

OA/Kappa and ideally per-class accuracy with confusion matrix must be reported and scenario mechanics must be transparent i.e. transition matrices + constraints + what variables are held constant vs projected to 2030.

Responses:

We sincerely appreciate your insightful and rigorous comments regarding the validation of the PLUS model and the transparency of the scenario mechanics. We have carefully reviewed the manuscript and can confirm that the methodological details you requested are comprehensively addressed within the text.

Here are our point-by-point responses:

(1) Regarding Model Validation (OA/Kappa and Accuracy): As explicitly described in Section 2.3.4 ("Model validation and future EHI assessment"), the performance of the PLUS model was rigorously validated. We simulated the 2020 land use pattern and compared it with the actual observed map using both the Overall

Accuracy (OA) and the Kappa coefficient. This standard validation procedure inherently evaluates the per-class accuracy performance of the simulation. The results confirmed the high reliability and accuracy of the model before we proceeded to project the 2030 patterns. We believe this section clearly demonstrates the robustness of our simulation results.

(2) Regarding Scenario Mechanics (Transition Probabilities and Constraints): The mechanics for each scenario are transparently detailed in Section 2.3.4(2) ("Scenario definition"). Rather than using a static matrix, we defined the scenarios through specific transition probability adjustments to reflect different policy orientations:

Scenario 1 (Natural Development): Based purely on historical transition trends from 2010–2020 without exogenous intervention.

Scenario 2 (Farmland Protection): We reduced the transition probability from farmland to grassland/water by 40% and increased the probability of unused land converting to farmland by 30%.

Scenario 3 (Urban Development): We increased the transition probability from farmland/grassland to construction land by 10% and reduced the reversion probability from construction land by 30%.

Constraints: We explicitly stated that for all scenarios, "key natural reserves and ecologically sensitive areas within the Ili River Valley were designated as restricted expansion zones where land use conversion was prohibited." This ensures that the simulation adheres to strict ecological constraints.

(3) Regarding Variables Held Constant vs. Projected: We maintained a clear distinction between static background variables and dynamic drivers to ensure scientific validity:

Variables Held Constant: Fundamental physical geographical variables, specifically Topography (DEM, Slope) and Soil Properties (derived from HWSD), were held constant across all scenarios, as these represent the stable foundational conditions of the valley.

Variables Projected: The primary variable projected to 2030 was the Land Use Pattern, simulated via the PLUS model under the distinct transition rules described above. Furthermore, as noted in Section 2.3.4(3), we calculated the future EHI by applying our framework to these future land use patterns "along with projected climatic and socio-economic data where applicable," ensuring that the assessment accounts for both spatial evolution and relevant temporal trends.

We hope these clarifications address your concerns and demonstrate the transparency and rigor of our simulation approach. Thank you again for your valuable time and suggestions.

Comment 3-6:

Landscape metric mapping is under-specified: If EO, EI, ER are mapped the manuscript must state Fragstats method (moving window vs landscape), window size, edge handling and how these metrics are rasterized back.

Responses:

We sincerely thank the reviewer for pointing out this omission. We agree that the specific parameters for landscape metric mapping are crucial for reproducibility. As suggested, we have added the detailed Fragstats configuration to the Methods section.

To clarify, the mapping of EO and EI was conducted using a moving window analysis (rather than a whole-landscape approach) with the following specific settings:

A 3000 m moving window with a 500 m step size (shift distance) was applied to capture continuous landscape heterogeneity； Forest, grassland, water bodies, and cropland were specifically defined as foreground classes for the metric calculations；
To strictly mitigate edge effects caused by incomplete windows at the landscape borders, a 2000 m edge buffer depth was applied (i.e., pixels within 2000 m of the study area boundary were masked as NoData during the calculation). The resulting continuous metric surfaces generated by Fragstats were subsequently resampled to a 1 km × 1 km resolution grid to match the spatial scale of other datasets in the EHI framework.

We believe these newly added details fully specify the landscape metric mapping process and address your concern.

Reviewer 4#

Comment 4-1:

Figure 1 – it is not possible to read the information and legend in (a) Location of Xinjiang, also part of the information in the legend is in Chinese language. (b) Elevation of the Ili river valley, and (c) Location of the study area are reversed.

Responses：

Thank you for your careful comments on Figure 1. We have completely revised

the figure by enlarging the font size, removing redundant information, and standardizing all legends in English to ensure readability.

Regarding the panel arrangement, due to the large spatial extent and vertical format required for panels c and d to clearly display geographic details, we have adjusted their layout to a vertical arrangement (top and bottom). As a result, the final layout is: panel (a) at the upper left, panel (b) at the lower left, panel (c) at the upper right, and panel (d) at the lower right. We believe this layout ensures better visualization and clarity of geographic information, and is reasonable for presenting the study area comprehensively.

Comment 4-2:

The manuscript should include a concise "Conclusions" section, focusing on major findings and highlighting the major outcomes of the study.

Responses：

We sincerely thank the reviewer for this constructive suggestion. As recommended, we have added a dedicated and concise "Conclusions" section at the end of the revised manuscript. This section specifically distills the major findings of our study—such as the spatial-temporal dynamics of farmland ecosystem health and the key nonlinear/threshold effects of dominant drivers—and highlights the broader practical outcomes for sustainable agriculture and ecological management in arid oasis regions.