

Asynchronous crop-record expansions reveal spatial opportunity structures in Holocene China

Yu Li

liyul@lzu.edu.cn

Lanzhou University <https://orcid.org/0000-0003-3381-5372>

Mingjun Gao

Lanzhou University

Zhansen Zhang

Lanzhou University

Simin Peng

Yuxin Zhang

Shiyi Yan

Hao Shang

Lanzhou University

Article

Keywords: Holocene agriculture, crop dispersal, non-local diffusion model, spatial opportunity structures, archaeological crop records

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11184906/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Asynchronous crop-record expansions reveal spatial opportunity structures in Holocene China

Mingjun Gao^{1,2}, Yu Li^{1,2*}, Zhansen Zhang^{1,2}, Simin Peng³, Yuxin Zhang⁴, Shiyi Yan², Hao Shang²

1 Key Laboratory of Western China's Environmental Systems (Ministry of Education), Lanzhou University, Lanzhou 730000, China

2 College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China

3 Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China

4 School of Geographic Sciences, Xinyang Normal University, Xinyang, Henan, 464000, China

*Corresponding author. Email: liyu@lzu.edu.cn

Abstract: Holocene agriculture in China emerged through the asynchronous expansion, convergence and reorganization of multiple crop systems, yet the mechanisms underlying these dispersal processes remain difficult to compare at the national scale. Here we integrate Chinese archaeological crop evidence with environmental data and develop a crop-specific dynamic non-local diffusion model to estimate the relative probability of new crop-record occurrence during 6000-2000 BP. Millet, rice and wheat/barley followed distinct spatial trajectories. Millet records expanded broadly across northern and transitional regions, consistent with inherited dryland agricultural foundations and regional continuity. Rice records remained concentrated in humid lowland environments, consistent with stronger environmental filtering and diffusion-environment coupling in the model. Wheat/barley records intensified later in regions with high corridor-mediated accessibility, consistent with the redistribution of introduced crops into existing agricultural and exchange networks. Our results therefore suggest that archaeologically visible Holocene crop dispersal in China was characterized by a multicentred, corridor-mediated and mechanism-specific process of spatial reorganization, shaped by crop-specific spatial opportunity structures rather than by the uniform expansion of a single agricultural frontier.

Keywords: Holocene agriculture; crop dispersal; non-local diffusion model; spatial opportunity structures; archaeological crop records

Introduction

Agricultural diffusion was a central process of social and environmental transformation during the Holocene [1]. It changed not only food production, but also settlement organization, demographic capacity, land-use intensity and regional interaction networks [2]. Comparative archaeobotanical studies show that the spread of crops often involved expansion, ecological extension, delay and regional reorganization rather than uniform outward movement [3-5]. As crops encountered ecological boundaries, terrain constraints, river systems and human preferences, their movement could be accelerated in some regions and delayed or redirected in others [6,7]. Therefore, agricultural expansion was unlikely to have been driven by a single factor, but instead reflected the combined influence of ecological conditions, historical contexts and human-mediated spatial processes [7].

Holocene agriculture in China provides an important case for examining these processes because it was assembled through the asynchronous expansion and convergence of multiple crop systems rather than the spread of a single agricultural package [8-10]. Millets formed the early

44 foundation of dryland agriculture in northern China and were closely associated with Neolithic
45 settlement expansion in the Yellow River basin, the North China Plain and adjacent northern regions
46 [11,12]. Rice was domesticated mainly in the Yangtze River basin and its dispersal was linked to
47 humid lowlands, lake-marsh plains, alluvial environments and manageable hydrological settings
48 [13,14]. Its later movement beyond the Yangtze region also involved ecological constraints and
49 changing dispersal pathways [8,14]. Wheat and barley, in contrast, were introduced crops that
50 entered China during the late Holocene through Central Asia and the Hexi Corridor before becoming
51 incorporated into northern agricultural exchange networks [15,16]. Previous archaeological and
52 archaeobotanical studies have established important chronological and regional frameworks for
53 these crop systems [8,10]. However, the mechanisms that produced different expansion patterns
54 remain difficult to compare across crop systems.

55 One limitation is that many studies focus on a single crop, a specific region or a particular
56 cultural sequence, making it difficult to evaluate whether different crops followed similar diffusion
57 processes or were shaped by distinct combinations of ecological, historical and spatial constraints
58 [8-10]. Another limitation is that explanations of crop dispersal often remain at the level of
59 qualitative route descriptions, such as the development and spread of millet farming in North China
60 river valleys and foothills [17,18], expansion through the Yangtze lowlands and associated rice-
61 agricultural landscapes [19,20], or movement via the Hexi Corridor, northeastern Qinghai-Xizang
62 Plateau margin and broader interregional exchange routes [21,22]. A further challenge is that
63 agricultural diffusion involves both historical continuity and spatial expansion. New crop records
64 may also reflect local archaeologically visible continuity, as long-sequence archaeobotanical records
65 show that earlier crop use in a region can condition later evidence of continued or reorganized crop
66 use [23]. They are also shaped by non-local accessibility, because crops can appear in new regions
67 through movement, exchange and interaction with already occupied agricultural areas [21,22]. At
68 the same time, environmental suitability may act as a filter that determines where a crop can be
69 adopted, maintained or archaeologically visible [19,20]. These mechanisms can produce similar
70 spatial patterns and distinguishing among them requires models that test the relative contribution of
71 environmental filtering, local persistence and non-local spatial connectivity within the same
72 analytical framework.

73 Here we ask whether the spatial expansion of millets, rice and wheat/barley in Holocene China
74 can be explained by a common diffusion mechanism, or whether each crop system was shaped by a
75 distinct combination of environmental filtering, inherited agricultural distributions and regional
76 accessibility. If these crops differed in ecological requirements, prior agricultural contexts and
77 transmission networks, their expansion should show different mechanistic sensitivities in spatial
78 prediction. To address this question, we compiled georeferenced archaeobotanical macroremain
79 records and crop-associated radiocarbon dates from two published databases [24,25], and integrated
80 them with gridded palaeoclimate, topographic and river-network data to develop a dynamic non-
81 local diffusion framework with crop-specific parameterizations for millets, rice and
82 wheat/barley[26-30]. The model estimates the probability of new crop-record occurrence for millets,
83 rice and wheat/barley during 6000-2000 BP using one-step conditional simulations and continuous
84 rolling simulations. It incorporates same-crop persistence, non-local diffusion potential,
85 environmental suitability and their interaction. We then use counterfactual ablation, spatial-null
86 comparisons, area-constrained evaluation, temporal-transfer validation and sampling-bias controls
87 to test whether the observed distributions can be explained by the proposed mechanisms rather than

by spatial clustering or archaeological sampling alone. Our aim is not to reconstruct a single historical route of crop transmission, nor to reduce agricultural dispersal to environmental determinism. Instead, we use a comparative modelling framework to identify how different crop systems occupied, reused and reorganized spatial opportunity structures during the Holocene.

Results

Archaeological crop records show asynchronous expansion of three crop systems

The compiled archaeological crop records show a multicentric, asynchronous and crop-specific spatiotemporal pattern (Fig. 1). Millet records appeared early and widely across northern China and the Yellow River basin, consistent with the extensive early development of dryland agriculture in the north (Fig. 1a,b). Rice records were concentrated mainly in the Yangtze River basin and adjacent humid lowlands, but also appeared in transitional and mixed-farming regions such as Henan, Shandong and Jiangsu, showing that archaeologically visible rice records extended beyond its core domestication zone (Fig. 1a,b). By contrast, the earliest provincial evidence for wheat/barley was generally later and first concentrated in northwestern China, the Gansu-Qinghai region, Guanzhong and the Central Plains, consistent with the introduction of these crops through Eurasian exchange corridors followed by regional redistribution (Fig. 1a,b).

The temporal record further shows that crop diversity and record intensity increased markedly after 6000 BP (Fig. 1c). Within the archaeologically well-represented 6000–2000 BP interval, millet records reached their highest compiled frequency around 4500–4000 BP and became less frequent toward the younger end of the modelling interval. Because the underlying source datasets were designed primarily for prehistoric and early historical archaeology, the temporal record density is not interpreted as a direct measure of crop abundance, cultivated area or agricultural decline. Rice records followed a broadly similar temporal pattern, although with fewer records overall. Wheat/barley appeared later, around 5000 BP, and increased substantially between 4000 and 2500 BP. The interval 6000–2000 BP therefore captures a major period of archaeologically visible expansion of millet and rice records, together with the principal increase in wheat/barley evidence, making it an appropriate core period for dynamic diffusion modelling.

Dynamic simulations reflect crop-specific expansion pathways

To simulate the dynamic expansion of the three crop systems between 6000 and 2000 BP, we constructed a discrete-time gridded crop-record probability model. The model estimates the relative probability that a new archaeological crop record of crop system c occurs in candidate grid cell x during time window t . This probability is decomposed into four components: same-crop local persistence, discrete non-local diffusion, dynamic environmental suitability and diffusion-environment interaction. The model therefore does not predict the absolute boundary of actual crop cultivation. Instead, it estimates where new archaeological crop evidence is more likely to appear, given previous crop records, environmental context and spatial transmission structures.

One-step conditional simulations show that the three crop systems followed distinct probability-surface trajectories (Fig. 2). Millet high-probability areas were initially concentrated in northern China and the Yellow River basin, and subsequently expanded towards northern, northeastern and parts of northwestern China. This pattern is consistent with the inherited foundation of dryland agriculture and regional continuity, suggesting that the predicted millet-record distribution was strongly associated with earlier crop distributions and local path dependence.

Rice showed a more spatially constrained pattern. Its high-probability areas remained

concentrated in the Yangtze River basin, southern humid regions and adjacent expansion zones. Compared with millet, rice was less likely to expand widely into arid or semi-arid regions, consistent with stronger environmental filtering by humid conditions and lowland settings. This suggests that the spatial pattern of rice-record probability was associated not only with previous rice distributions and dispersal opportunities, but also with environmental suitability.

Wheat/barley followed a different trajectory from the two native crop systems. Its high-probability areas intensified later along the northwest, Hexi Corridor-Central Plains connecting belts, and showed cross-regional expansion across multiple time windows. This pattern is consistent with wheat/barley as an introduced crop system whose archaeologically visible expansion was not well explained by same-crop local persistence alone. However, the model does not directly reconstruct migration routes, exchange networks or cultural transmission pathways. Rather, it identifies corridor-like accessibility patterns inferred from distance decay, topographic costs, river-related accessibility and fixed crop-specific regional accessibility weights (Supplementary Fig. S1). These weights were implemented as weak, non-exclusive spatial modifiers rather than as predefined route masks or historical pathways learned from the crop-record response variable. Wheat/barley records were therefore more consistent with non-local accessibility and possible corridor-mediated redistribution than with a strictly local continuation process.

Continuous rolling simulations had lower predictive performance than one-step conditional simulations, as expected, because early errors accumulate through time. Nevertheless, they retained the main spatial tendencies of the three crop systems. This indicates that the model was not only able to predict window-level changes from observed previous crop distributions, but could also generate long-term crop-specific expansion patterns to some extent. We therefore use one-step conditional simulations as the main basis for mechanistic interpretation and treat continuous rolling simulations as supplementary evidence for long-term process stability.

Model validation supports spatially enriched crop-record prediction

The full dynamic model showed consistent spatial-enrichment ability for new crop records in all three crop systems under one-step conditional simulation. The area under the receiver operating characteristic curve (AUC) values for millet, rice and wheat/barley were approximately 0.94, 0.93 and 0.93, respectively, and the average precision (AP) values were about 0.17, 0.20 and 0.15, all higher than the random AP baseline. These results indicate that the model ranked observed new crop-record cells above background candidate cells more effectively than random expectation in a sparse archaeological prediction setting. However, these metrics should be interpreted as evidence of relative spatial ranking and enrichment, not as evidence that the model precisely reconstructs the true extent of crop cultivation.

Area-constrained evaluation further showed that the model did not identify true records simply by assigning high probabilities to overly broad regions. When only the top 5% of candidate grid cells with the highest predicted probabilities were retained, the one-step forecasts captured a large proportion of observed new crop records, with top-5% recall values of approximately 0.73, 0.68 and 0.74 for millet, rice and wheat/barley, respectively. This result indicates that observed new records were enriched within a limited high-probability area. It does not imply that areas outside the top-ranked cells lacked crop cultivation, because the model predicts archaeologically visible new crop records rather than complete agricultural presence.

Comparisons with spatial null models and sampling-bias controls showed that the full model captured spatial structure beyond several simple alternatives. The AP of the full model was higher

than that of the spatially uniform null model, the preceding same-crop kernel density estimation (KDE) null model and the total-crop sampling-intensity KDE null model. This suggests that the predictions were not reducible to random expectation, previous same-crop clustering or general archaeological sampling intensity. Nevertheless, these comparisons should be understood as controls against major sources of spatial bias rather than as proof that all preservation, publication and excavation biases have been eliminated.

Nested-model comparisons further showed that model improvement was not simply a consequence of increasing model complexity. Adding non-local diffusion improved spatial ranking beyond same-crop local persistence, indicating that spatial transmission pressure contributed additional predictive information. Adding environmental suitability and diffusion-environment interaction terms improved some evaluation metrics, including AUC, AIC, Bayesian information criterion (BIC) or Brier score, but did not always increase AP. The full model should therefore not be interpreted as a universally superior model that automatically explains all crop-record patterns. Instead, the validation results indicate that different mechanism terms contributed differently to the spatial ordering, calibration and extent of predicted crop-record probability surfaces.

Parameter-stability and pseudo-posterior analyses further supported the selected crop-specific environmental-suitability responses, whereas several spatial-accessibility parameters showed more moderate support and were therefore interpreted cautiously as predictive spatial weights rather than precise historical-route parameters (Supplementary Fig. S2).

Probability-calibration tests also showed that predicted probabilities had relative explanatory value. Reliability diagrams indicated that observed crop-record frequencies generally increased with mean predicted probability across quantile bins, but deviations from perfect calibration confirmed that these values should be interpreted as probabilities of newly recorded archaeological crop evidence rather than absolute probabilities of real crop occupation (Supplementary Fig. S3). Temporal-transfer validation provided additional support for cross-stage generalization. After dividing 6000-2000 BP into early and late periods, both early-training-late-testing and late-training-early-testing experiments produced AP values above the random AP baseline. This indicates that the model did not only fit the static spatial structure of the full dataset, but retained some ability to enrich new crop records across different temporal stages.

Additional robustness analyses supported the reliability of these results. Environmental-variable diagnostics showed that VIF values remained below the common threshold of 5, indicating that environmental terms should be interpreted as composite suitability rather than as independent causal effects of single variables (Supplementary Fig. S4a). Residual Moran's I was substantially lower than the spatial autocorrelation of the observed records and predicted probability surfaces, and regional validation retained predictive performance across the Yellow River-North China region, the middle-lower Yangtze region and the Hexi-Gansu-Qinghai corridor (Supplementary Fig. S4b,c). Chronological-uncertainty Monte Carlo experiments based on 100 iterations showed that AUC remained stable under age perturbations, although AP was more sensitive in low-sample windows, especially for sparse wheat/barley records (Supplementary Fig. S5). Spatiotemporal sensitivity tests using 500-year temporal aggregation and 1° spatial aggregation showed that absolute performance metrics varied with scale, but model performance remained above random expectation and the main crop-specific interpretations were not reversed (Supplementary Fig. S6).

Mechanistic and paleoclimate diagnostics reveal heterogeneous expansion processes

Temporal-transfer experiments already suggested that the three crop systems differed in their transferable mechanisms. Millet prediction was mainly supported by same-crop local persistence and non-local diffusion, rice gained more from environmental and interaction terms, whereas wheat/barley improved more clearly after the inclusion of non-local diffusion. We therefore used counterfactual ablation, environmental gating and frontier-mobility diagnostics to identify the dominant expansion mechanisms of each crop system.

Counterfactual ablation showed that removing same-crop local persistence caused the largest decline in AP for all three crop systems, with average AP decreases of about 0.151, 0.186 and 0.147 for millet, rice and wheat/barley, respectively. This indicates that previous same-crop evidence was the strongest predictor of later crop-record occurrence. We interpret this effect as archaeologically visible continuity: regions with earlier crop evidence were more likely to generate additional crop records in later periods, a pattern that is consistent with regional agricultural continuity but does not demonstrate uninterrupted cultivation. However, this interpretation does not assume continuous cultivation within the same grid cell, nor does it ignore the possibility of uneven archaeological visibility. Its mechanistic meaning is supported only in comparison with the sampling and spatial-null models: the full model outperformed both the total-crop sampling-intensity KDE and the same-crop KDE null model, indicating that its predictive ability was not reducible to general research intensity or simple spatial clustering of previous records.

Despite this common dependence on same-crop local persistence, the three crop systems differed in their sensitivity to non-local diffusion and dynamic environmental suitability. Removing non-local diffusion reduced AP by about 0.059 for rice and 0.038 for wheat/barley, whereas the decline for millet was smaller. This suggests that non-local transmission contributed more clearly to the spatial ordering of new rice and wheat/barley records. Wheat/barley records were more consistent with non-local accessibility and archaeologically visible expansion along Hexi Corridor-Central Plains connecting belts.

Dynamic environmental suitability made different contributions to the three crop systems, but its role should be interpreted as a composite environmental filter rather than as the independent causal effect of any single climate variable. Removing the environmental-suitability term reduced rice AP by about 0.064, indicating that rice records were most strongly associated with the spatial ordering provided by humid and lowland environmental suitability. This is consistent with the ecological dependence of rice cultivation on moisture availability and suitable hydrological settings. By contrast, removing environmental suitability did not reduce AP for millet and had only a minor effect on wheat/barley. For these two crop systems, the environmental term therefore appears to have adjusted the intensity, extent or boundary of predicted probability surfaces rather than independently determining where new records were ranked highest. In other words, environmental suitability acted as a filtering context whose importance varied among crops, rather than as a universal driver of crop-record expansion.

Environmental-gating and frontier-mobility diagnostics further clarified how environmental suitability interacted with spatial transmission pressure (Fig. 4b). Rice showed the highest environmental-gating ratio, indicating that high diffusion pressure was more likely to translate into high new-record probability where humid-environment suitability was also high. This does not mean that moisture alone determined rice expansion. Rather, it suggests that dispersal opportunities for rice became archaeologically visible mainly when they overlapped with suitable humid and lowland settings. Millet showed a lower gating ratio and more moderate frontier mobility, consistent

with broad environmental tolerance within northern dryland and transitional subhumid regions, where previous agricultural foundations and regional continuity played a stronger role in structuring new records. Wheat/barley showed the highest frontier mobility, with Q75 frontier velocities substantially higher than those of millet and rice. New wheat/barley records were therefore more likely to appear farther from previous same-crop records. This pattern is consistent with the archaeologically visible spread of introduced crops through non-local or leapfrog-style expansion. However, corridor-mediated accessibility should be interpreted as an inferred spatial condition rather than a directly observed transmission route. The model supports the view that wheat/barley records were more strongly associated with broad Hexi Corridor-Central Plains accessibility structures, but it does not distinguish whether these patterns resulted from migration, exchange, technological adoption, elite consumption, local experimentation or other social processes.

Together, these results indicate that the archaeological distributions of the three crop systems were associated with different combinations of modelled processes. Millet crop-record expansion was most strongly associated with path dependence, same-crop local persistence and regional frontier advancement within broadly suitable dryland environments. Rice crop-record expansion showed the strongest association with humid-environment filtering and diffusion-environment gating, with high diffusion opportunities more likely to coincide with new records under suitable moisture conditions. Wheat/barley records were more consistent with archaeologically visible non-local redistribution along northwest and northern accessibility belts. Overall, crop-record expansion in China during 6000–2000 BP was characterized by crop-specific combinations of path dependence, environmental filtering and non-local dispersal.

Regional paleoclimate comparisons provide additional contextual support for this interpretation (Supplementary Fig. S7). Crop-record intensity was not synchronous across regions and did not follow a single hydroclimatic trajectory. In Central China, crop-record intensity remained high or increased during phases of weakening moisture availability before later declining. In southern China, rice and millet records reached relatively high values under generally humid conditions, whereas wheat/barley strengthened later. In northwestern China, the transition from millet-dominated records to increasing wheat/barley records was not accompanied by a clearly synchronous hydroclimatic shift. These comparisons are not intended as formal causal attribution, but they indicate that hydroclimate acted as a regional boundary condition whose effects were mediated by crop ecology, previous agricultural distributions, corridor accessibility and the timing of crop adoption.

Discussion

From diffusion routes to spatial opportunity structures

Our results suggest that crop dispersal in Holocene China is better understood not simply as the outward movement of crops from several origin centres, but as a crop-specific spatial reorganization process (Figs. 2 and 4). We use the concept of spatial opportunity structures to describe crop-specific spatial conditions produced by the overlap of ecological suitability, inherited agricultural landscapes and corridor-like accessibility. This concept builds on a broader spatial-process perspective that emphasizes the role of spatial structure, landscape connectivity and expansion-frontier dynamics in shaping long-term human and agricultural change [31,32]. In this perspective, environmental change is not treated as a direct and spatially uniform trigger of human expansion or decline, but as a boundary condition whose effects are

mediated by inherited landscapes and regional connectivity [33,34]. Here, we extend this perspective to crop-record dispersal in Holocene China, where different crop systems encountered different ecological constraints, agricultural histories and transmission opportunities [35,36]. The modelled crop-record patterns were therefore associated with different spatial opportunity structures, produced by varying combinations of environmental filtering, previous crop distributions and corridor-like accessibility. Thus, the different expansion directions of these crop systems were not determined only by their points of origin. Rather, new crop records were more likely to occur where ecological conditions, previous crop distributions and regional connections overlapped.

For millet, high-probability areas were associated with northern China, the Yellow River basin, the North China Plain and northeastern expansion zones. This pattern is consistent with the broad ecological adaptability of dryland agriculture across northern and transitional regions [37]. It is also consistent with regional continuity associated with earlier millet cultivation, especially in areas where previous agricultural landscapes provided favourable conditions for later crop-record persistence [38,39]. The spatial overlap among millet records, river systems, plains, settlement density and previous agricultural cores suggests that rivers, foothills and lowland corridors may have facilitated settlement, mobility and agricultural expansion [17,18]. However, explicit river terms did not emerge as independent dominant predictors at the national grid scale. This does not imply that rivers were unimportant. Instead, it indicates that river effects were difficult to separate from plains, existing agricultural centres, sampling intensity and broader regional continuity. The millet crop-record pattern can therefore be interpreted as being associated with river- and plain-related continuity zones, rather than as being directly controlled by river distance or river direction alone.

Rice crop-record expansion shows a different type of opportunity structure. The model predicted high probabilities mainly in the Yangtze River basin, southern humid regions and adjacent transitional zones, indicating that humid environments and low-relief landscapes consistent with humid environments and low-relief landscapes acting as important filters on archaeologically visible rice dispersal [40,41]. However, rice did not appear simply wherever moisture conditions were suitable. Its expansion also associated with previous rice distributions, local adoption, regional cultural networks and the formation of mixed agricultural systems. The observed rice-record pattern is therefore consistent with the combined influence of environmental constraints and historical pathways. This interpretation helps explain why rice records appeared in regions such as the Central Plains, the Yangtze River area and parts of southwestern China. These occurrences do not necessarily indicate a simple northward or westward shift of the rice ecological boundary, but may instead reflect the incorporation of rice into new multi-crop systems under locally suitable environmental and cultural conditions [40-42].

Wheat and barley represent a further type of spatial reorganization. As introduced crops rather than crops independently domesticated within China, their crop-record expansion differed from that of millet and rice [15,16,43]. Their archaeobotanical appearance has often been linked to west-east transmission through northwest China and adjacent Inner Asian interaction zones [44,45]. In our model, wheat/barley probabilities increased later along the northwest, Hexi Corridor, Guanzhong and Central Plains connecting belts. We interpret this pattern as being consistent with corridor-like accessibility rather than as a direct reconstruction

of a single transmission route. The model combines distance attenuation, terrain costs, river-related accessibility and crop-specific regional weights, and therefore captures spatial conditions under which introduced crops were more likely to become archaeologically visible. These spatial conditions are broadly consistent with previous evidence for wheat dispersal and agricultural innovation across northwest China and wider Asian exchange networks [46,47]. However, the model does not by itself distinguish among population movement, exchange, cultural transmission, local adoption or local experimentation [48-50].

Overall, the concept of spatial opportunity structure allows us to move beyond a descriptive route-based narrative while avoiding the assumption that modelled corridors are exact historical pathways. The Yellow River, Yangtze River, Hexi Corridor, Gansu-Qinghai region and northern transmission belts should not be treated as fixed channels through which crops moved mechanically. They are better understood as regions where environmental suitability, previous agricultural activity, settlement opportunities, mobility, exchange and archaeological visibility may have overlapped. Crop-record dispersal in Holocene China was therefore not a single-frontier expansion, but a multi-centred and multi-mechanism reorganization of agricultural space. The model identifies the spatial conditions associated with this reorganization, rather than directly observing the social processes that produced it.

Climate as a boundary condition rather than a direct trigger

The model-based environmental diagnostics suggest that climate and hydroclimate are better understood as boundary conditions and filtering contexts for crop-record expansion, rather than as direct short-term triggers of agricultural rise or decline (Fig. 4; Supplementary Fig. S7). The environmental term in our model integrates precipitation, water balance, productivity, temperature and lake-related hydrological suitability, and should therefore be interpreted as composite environmental suitability. It does not identify the causal effect of a single variable, nor does it define physiological thresholds for crop cultivation. Instead, it estimates whether the environmental background of a grid cell made the archaeological appearance of a given crop system more or less likely, conditional on previous crop distributions and spatial transmission opportunities [34,35].

This distinction is important because the three crop systems responded differently to environmental filtering. Rice records were most strongly associated with humid-environment suitability and showed the clearest diffusion-environment gating, suggesting that rice dispersal opportunities were more likely to produce new archaeological records where moisture and lowland conditions were favourable [40,41]. Millet records, by contrast, were less dependent on the environmental-suitability term in spatial-ranking performance, consistent with broader tolerance across northern dryland and transitional subhumid regions and stronger dependence on inherited dryland agricultural landscapes [37,38]. Wheat/barley records showed weaker environmental gating and higher frontier mobility, indicating that the spread of these introduced crops was more closely related to non-local accessibility, exchange corridors and incorporation into existing farming systems [44,48]. Climate therefore shaped the potential space of crop adoption, but the archaeological realization of crop-record expansion depended on the overlap between environmental suitability, historical archaeologically visible continuity and regional connectivity [31,33,36].

These patterns make it difficult to explain Holocene crop changes in China through a simple model in which a single climatic event caused simultaneous agricultural decline or expansion.

A more plausible interpretation is that climate reshaped the environmental space in which crops could be adopted and maintained, while the actual appearance of crop records depended on whether suitable environments overlapped with existing agricultural distributions, settlement systems, technological transmission and regional exchange networks [34,36,47].

This interpretation is especially important for understanding the 4.2 ka event and related mid- to late-Holocene transitions. The decline of some crop records around this period does not necessarily imply a direct and uniform climatic collapse of agriculture. Instead, the effects of hydroclimatic change were mediated by crop ecology and regional social-environmental contexts (Supplementary Fig. S7) [34,47]. Rice was more strongly constrained by humid environmental suitability; millet had broader adaptability within northern subhumid and semi-humid regions; and wheat/barley expanded later through exchange networks and mixed farming systems [37,40,45]. Climatic stress may therefore have altered the relative advantages, risks and spatial limits of different crops, but it did not automatically determine where agricultural activity expanded or declined [35,36].

In this sense, climate acted as one component of a broader spatial opportunity structure. Precipitation, water balance, productivity, temperature and local hydrological conditions shaped the potential space for crop-record expansion, but they became archaeologically visible only when combined with human settlement, crop adoption, technological practice and regional interaction [29,30]. This perspective helps explain why similar climatic changes could produce different agricultural outcomes in different regions, and why apparent lags or resilience in previous studies may reflect the buffering and filtering effects of regional agricultural systems [33,34,47].

Environmental filtering, archaeologically visible continuity and corridor-like redistribution

The crop-specific patterns identified by the dynamic diffusion model can be interpreted in terms of three linked dimensions: environmental filtering, archaeologically visible continuity and corridor-like redistribution. These processes did not act independently. Instead, they combined in different ways to create the spatial opportunity structures through which different crops expanded [31,32].

Same-crop persistence provided the strongest common predictive signal across the three crop systems, but it should be distinguished from two non-mechanistic sources of spatial predictability. First, archaeological sampling intensity may concentrate records in regions with more intensive excavation, better preservation or stronger publication traditions [34,35]. Second, spatial clustering may arise because archaeological sites and crop records are geographically aggregated even without a specific crop-expansion process [51,52]. Our validation framework was designed to separate these effects from historical persistence (Fig. 3). The sampling-intensity KDE null model represents the general tendency for any crop record to occur in well-sampled areas, whereas the same-crop KDE null model represents spatial clustering around previous records of the same crop [52]. The fact that the full model exceeded these null expectations suggests that same-crop persistence should not be interpreted merely as sampling bias or spatial autocorrelation. Instead, it captures a crop-specific historical signal in which earlier crop-record distributions were associated with later archaeological visibility. Nevertheless, this signal remains an inference from archaeological records and should be understood as archaeologically visible continuity rather than direct proof of uninterrupted

cultivation [34,36].

Environmental filtering and corridor-mediated redistribution then differentiated the three crop systems. Rice showed the strongest sensitivity to environmental suitability and the highest environmental-gating ratio, indicating that diffusion pressure was more likely to generate new records where humid conditions were suitable [40,41]. Wheat/barley showed stronger non-local expansion and higher frontier mobility, consistent with the spread of introduced crops through northwest and northern exchange corridors [44,45]. Their later incorporation into existing agricultural and exchange networks further suggests that corridor-mediated redistribution operated through already structured interaction systems rather than through a single linear route [48,49]. Millet, by contrast, was more closely associated with inherited dryland agricultural landscapes and regional continuity [37,38]. Thus, the same broad process of crop dispersal was realized through different mechanism combinations in different crop systems.

Together, these model results suggest that crop-record expansion in Holocene China was neither environmentally deterministic nor purely culturally transmitted. Environmental suitability was associated with broad spatial limits and opportunities; same-crop persistence captured archaeologically visible continuity; and corridor-like accessibility was associated with the non-local redistribution of crop records [31,33]. The archaeological record of China's Holocene agricultural landscape can therefore be interpreted as reflecting a multi-crop, multi-pathway spatial reorganization, involving crop-specific combinations of ecological filtering, archaeologically visible continuity and regional connectivity [35,38,43].

Limitations

This study has several limitations. The model estimates the probability of observing archaeological crop evidence, rather than the actual extent of cultivation, crop yield, dietary contribution or population movement. Archaeobotanical records are affected by excavation intensity, publication bias, preservation conditions, dating uncertainty and regional research traditions [34,35]. Although area-constrained evaluation, temporal-transfer validation, spatial null models and sampling-intensity controls were used to reduce over-optimistic inference in spatially and temporally structured data, archaeological data unevenness cannot be fully eliminated [53]. The absence of crop records in a grid cell or time window should therefore not be interpreted as the absence of cultivation, but as the absence of archaeologically visible or currently compiled evidence [34-36,51]. In addition, the environmental and diffusion terms should be interpreted as simplified mechanistic representations. Environmental-suitability responses should not be interpreted as physiological crop thresholds. They represent relative, model-based environmental filters that best explain crop-record occurrence within the present modelling framework. Likewise, non-local diffusion, river, topographic and regional-accessibility terms approximate spatial transmission opportunities rather than reconstruct exact historical routes or distinguish among migration, exchange, technological adoption and cultural transmission. The results should therefore be interpreted as identifying spatial conditions associated with crop-record expansion, rather than as demonstrating deterministic causal pathways.

Conclusion

By integrating archaeological crop records, topographic and riverine information, and

paleoclimate reconstructions, this study developed a dynamic multi-crop diffusion model to examine crop-record expansion in China during 6000–2000 BP. The results indicate that the archaeological distributions of the three crop systems were associated with different combinations of modelled processes. Millet crop-record expansion was mainly associated with same-crop persistence, archaeologically visible continuity and regional expansion across northern China and adjacent transitional zones. Rice crop-record expansion showed a stronger association with humid environmental suitability and diffusion-environment coupling, with new records more likely to occur where dispersal opportunities overlapped with favourable moisture and lowland conditions. Wheat/barley crop-record expansion showed stronger non-local diffusion and higher frontier mobility, consistent with archaeologically visible redistribution along northwest and Central Plains accessibility belts. Model validation and robustness tests indicate that these patterns cannot be fully explained by archaeological sampling hotspots, spatial clustering, simple distance decay or increasing model complexity alone. These archaeological patterns are therefore more consistent with a multi-centred, multi-corridor and crop-specific process of spatial reorganization than with the uniform outward expansion of a single agricultural frontier. This framework provides a quantitative way to compare how crop-specific environmental conditions, previous crop-record distributions and regional accessibility were associated with the archaeologically visible reorganization of crop systems in Holocene China.

Method

1 Archaeological crop records, environmental variables and gridding

We compiled archaeological crop evidence from two complementary published datasets. The Chronology of Early China database contains 7,083 radiocarbon determinations assembled from five regional datasets [24]. Crop-associated records were identified from the reported taxonomic and contextual information, and the calibrated median age and, where available, the 95.4% calibrated age range were retained. We additionally used a published archaeobotanical macroremain dataset containing 538 site or assemblage records from northern China [25]. A crop occurrence was retained when either the reported absolute count or absolute weight was greater than zero. These procedures yielded 944 crop-associated radiocarbon records and 1,426 positive macroremain observations, giving 2,370 source observations before subsequent spatial, chronological and taxonomic harmonization. Our database is restricted to archaeologically documented crop evidence and does not incorporate crop occurrences reported only in historical texts, local gazetteers, agricultural treatises or taxation records. These documentary sources follow a fundamentally different observation and preservation process from archaeobotanical and radiocarbon evidence and therefore were not combined with the archaeological occurrence model. Importantly, both source datasets are weighted toward prehistoric and early historical periods. Consequently, the decline in compiled archaeological crop records toward and after approximately 2000 BP reflects the temporal scope of the source datasets and a transition in the dominant evidence regime, rather than a decline in crop cultivation. We therefore restrict quantitative interpretation and dynamic modelling to 6000–2000 BP.

Taxonomic labels were harmonized into three crop systems: foxtail millet and broomcorn millet were grouped as millets, rice was treated as a separate crop system, and wheat and barley were combined as wheat/barley. Soybean and records identified only as unspecified millet were excluded from the crop-specific analyses. These three crop systems represent contrasting agricultural trajectories in Holocene China, including indigenous millet agriculture, rice cultivation,

and the introduction and subsequent spread of West Eurasian wheat and barley. After spatial, chronological and taxonomic screening and removal of duplicate records, the final model dataset comprised 2,100 unique archaeological crop records, including 1,223 millet, 450 rice and 427 wheat/barley records. Of these, 1,945 occurred within the principal modelling interval of 6000–2000 BP. Records were assigned to consecutive 300-year time windows according to their median calibrated ages. The 6000–2000 BP interval was used for the main modelling analysis because it encompasses the major spatial expansion of millet and rice agriculture and the principal period of wheat/barley introduction and subsequent spread in China. Earlier records extending to 9000 BP were retained only to establish the preceding observed crop state required by the dynamic model. The effects of chronological uncertainty were evaluated separately through Monte Carlo robustness analyses using the reported calibrated age ranges where available.

All archaeological and environmental data were represented on a common $0.5^\circ \times 0.5^\circ$ longitude–latitude grid covering the study region. Candidate cells were defined as land grid cells with valid core terrain and environmental information. Each archaeological coordinate was assigned to the grid cell containing the observation rather than to the nearest grid-cell centre.

For crop system c , candidate grid cell i , and time window t , we defined a binary archaeological occurrence variable as

$$y_{c,i,t} = \begin{cases} 1, & \text{if crop system } c \text{ is observed in cell } i \text{ during window } t, \\ 0, & \text{otherwise.} \end{cases}$$

Multiple archaeological observations within the same crop–cell–window combination therefore contributed a single positive occurrence, whereas the same grid cell could be positive in more than one time window. Here, an observed crop occurrence refers to the archaeological presence of a crop within a particular grid–time window. It does not imply the first introduction or first cultivation of that crop, continuous cultivation through time, or the complete spatial extent of crop production. Previous occurrence was represented separately through the same-crop persistence term in the dynamic model.

Static landscape conditions were characterized using elevation and slope derived from GMTED2010 [26], together with river proximity and river-direction consistency calculated from the global river hydrography and network-routing dataset [27]. These variables were used to describe terrain constraints, accessibility and potential movement corridors. Dynamic environmental conditions were represented by reconstructed precipitation, temperature, precipitation minus evaporation, net primary productivity, and virtual lake level or related hydrological suitability [28–30]. These predictors were selected to characterize variation in water availability, thermal conditions, ecosystem productivity and surface hydrology, rather than to impose deterministic physiological thresholds for crop cultivation. Crop-specific regional-accessibility variables were additionally included in the dynamic model to characterize spatial differences in access to previously occupied or potential source regions.

Dynamic environmental fields were spatially interpolated to candidate grid cells and averaged within each 300-year time window. Remaining missing values in precipitation, temperature, precipitation minus evaporation and net primary productivity were replaced by the median value for the corresponding time window, whereas missing lake values were assigned a value of zero. Continuous predictors were subsequently standardized or normalized before model fitting to reduce differences in numerical scale among variables and facilitate comparison among predictor effects.

2 Dynamic crop-record occurrence probability models

To simulate the spatial expansion of crop systems under changing environmental and accessibility conditions, we developed a discrete-time, grid-based dynamic crop-record occurrence model. The model estimates the relative probability that a new archaeological crop record of crop system c occurs in candidate grid cell x during time window t . The model does not attempt to reconstruct the actual boundary of crop cultivation, crop yield, population density or dietary contribution. Instead, archaeological crop records are treated as observable evidence of agricultural activity, and the model focuses on the relative spatial probability of new crop-record occurrence.

For crop system c , candidate grid x , and time window t , the probability of new crop records occurring is defined as:

$$\text{logit}\{p_c(x, t)\} = \alpha_c + \beta_{\text{persistence}} U_c(x, t) + \beta_{\text{diffusion}} D_c(x, t) + \beta_{\text{environment}} E_c(x, t) + \beta_{\text{diffusion} \times \text{environment}} D_c(x, t) E_c(x, t).$$

Where α_c is the intercept term of crop system c , β represents the weight; $U_c(x, t)$ is the same-crop local persistence term, indicating whether the grid x has records of similar crops before the current window; $D_c(x, t)$ is the discrete non-local diffusion term, indicating the distance attenuation propagation pressure exerted by previously recorded crop grids on candidate grids x ; $E_c(x, t)$ is the dynamic environmental suitability term, composed of climate changes affecting crop; $D_c(x, t) E_c(x, t)$ are diffusion-environment interaction terms, used to represent the relationship between diffusion opportunities and environmental suitability. Therefore, the model represents crop-record expansion using four interconnected components: same-crop local persistence associated with previously observed crop records, non-local influence from other previously recorded grids, crop-specific environmental filtering, and diffusion-environment interaction. The model does not predict the absolute boundary of the actual crop planting range, but rather the relative probability of a candidate grid generating new crop archaeological evidence.

2.1 Same-crop local persistence

The same-crop local persistence term $U_c(x, t)$ indicates whether crop system c has already appeared in candidate grid x before the current time window t . If grid x already has a crop system record before the current window, then

$$U_c(x, t) = 1,$$

Otherwise

$$U_c(x, t) = 0.$$

For archaeological crop data, the same-crop local persistence term does not mean that a crop is continuously planted within the same grid, but rather that the area already has archaeological visible evidence of the crop system, making it more likely that similar crop records will continue to be produced in subsequent time windows. In one-step simulation (see Section 2.6), $U_c(x, t)$ comes from the actual observed crop record state prior to the current window. That is, each time window's prediction is based on the distribution of previous real archaeological records to test whether the model can predict the location of new records in the next stage given the agricultural distribution in the previous stage. In the continuous simulation (see Section 2.6), $U_c(x, t)$ comes from the model's updated state in the previous time window, used to test whether the model can maintain a long-term expansion pattern without real-world follow-up observation corrections. It should be noted that $U_c(x, t)$ more accurately represents "same-crop local persistence", rather than strictly defined "local diffusion." Diffusion usually refers to propagation from one spatial location to another, while $U_c(x, t)$ indicates the impact of an existing recorded state on the probability of new records in the same grid.

2.2 Discrete non-local diffusion

The non-local diffusion term $D_c(x, t)$ is used to represent the spatial propagation pressure exerted by previously recorded or simulated crop cells on candidate grid cell x . Unlike considering only local diffusion from adjacent grids, this study uses a discrete non-local diffusion kernel defined on candidate grids. In other words, the diffusion pressure of candidate grid cell x is not determined solely by its neighbouring grid cells, but is obtained from the weighted influence of previous source cells.

The diffusion pressure of candidate grid cell x was calculated as the normalized weighted influence of previous source cells:

$$D_c(x, t) = \text{norm} \left[\sum_{y \in S_c(<t)} U_c(y, t - \Delta t) K_c(x, y) \right]$$

where $S_c(<t)$ is the set of previous source cells for crop system c , $U_c(y, t - \Delta t)$ indicates whether source cell y had previous evidence or simulated state intensity of crop system c , and $K_c(x, y)$ is the crop-specific source-destination kernel from source cell y to candidate cell x . The operator $\text{norm}[\cdot]$ denotes normalization of the resulting diffusion pressure before it enters the logistic occurrence model.

The propagation kernel was defined as:

$$K_c(x, y) = R_c(x) \exp \left[\frac{-d(x, y)}{\sigma_c} - \theta_{\text{slope}, c} S(x) - \theta_{\text{elev}, c} H(x) + \theta_{\text{river}, c} B(x) + \theta_{\text{align}, c} A(x, y) + \theta_{\text{dir}, c} Q_c(x, y) \right]$$

Here, $d(x, y)$ is the distance between source cell y and candidate cell x , and σ_c is the crop-specific distance-decay scale. $S(x)$ is the normalized slope term, and $H(x)$ is the normalized elevation-related cost term. $B(x)$ represents river-proximity accessibility, $A(x, y)$ represents the consistency between the source-destination direction and the local river-alignment direction, and $Q_c(x, y)$ is the crop-specific directional prior. $R_c(x)$ is the fixed crop-specific regional accessibility weight of the candidate cell.

The parameters $\theta_{\text{slope}, c}$, $\theta_{\text{elev}, c}$, $\theta_{\text{river}, c}$, $\theta_{\text{align}, c}$ and $\theta_{\text{dir}, c}$ control how slope, elevation, river proximity, river alignment and the directional prior modify the non-local diffusion term. These parameters correspond to the diffusion-kernel parameters listed in Supplementary Table S1. By contrast, $R_c(x)$ was fixed before model fitting and used only as a weak destination accessibility modifier. It did not exclude candidate grid cells or force diffusion along predefined routes.

2.3 Dynamic environmental suitability

Dynamic environmental suitability term $E_c(x, t)$ represents the comprehensive environmental suitability of crop system c under candidate grid x and time window t . This term consists of precipitation (P), precipitation-evaporation (P-E), net primary productivity (NPP), temperature (T), and virtual lake level (LL), and is used to represent the combined constraints on the probability of crop records due to local hydrological background, ecological productivity and heat conditions.

$$E_c(x, t) = w_{P, c} G(P(x, t); \mu_{P, c}, \sigma_{P, c}) + w_{PE, c} G(P - E(x, t); \mu_{P-E, c}, \sigma_{P-E, c}) + w_{N, c} NPP(x, t) + w_{T, c} T(x, t) + w_{LL, c} LL(x, t).$$

Where $w_{P, c}$, $w_{P-E, c}$, $w_{NPP, c}$, $w_{T, c}$ and $w_{LL, c}$ are the weights of each environmental component. The suitability of P and P-E is expressed using the Gaussian response function:

$$G(z; \mu, \sigma) = \exp \left[-\frac{1}{2} \left(\frac{z - \mu}{\sigma} \right)^2 \right]$$

Here, z is the environmental variable value in the current grid and time window, μ is the candidate optimum, and σ is the response width. This function indicates that the closer the environmental variable is to the candidate optimal value of the crop system, the higher the suitability; The farther it deviates from the optimal value, the lower the suitability.

NPP, T, and LL were normalized to a 0-1 range before entering the model. For millet and rice, the temperature term indicates suitability for relatively warm conditions; For wheat/barley, the temperature term is introduced as relative cool suitability, reflecting its adaptation background in cooler or dry environments. It should be emphasized that $E_c(x, t)$ is an environmental suitability indicator used to predict the probability of new archaeological crop records, rather than a complete physiological model of crop ecological needs.

2.4 Diffusion-environment interaction

The diffusion-environmental interaction term $D_c(x, t)E_c(x, t)$ is used to represent the coupling relationship between transmission opportunity and environmental suitability. If a region has high environmental suitability but lacks spatial correlation with existing crop distribution, the probability of new crop records may remain low; Conversely, if a region is close to existing agricultural areas or under high transmission pressure but environmental suitability is insufficient, it may be difficult to form new crop records. Therefore, diffusion and environment are represented as interacting model components that jointly contribute to the predicted probability of new crop-record occurrence.

To avoid interpreting the interaction term solely as a statistical item that increases model flexibility, this study further calculates the environmental gating effect and the coupling accessibility effect. Specifically, we compare the prediction probabilities for three types of grids: a grid with high diffusion pressure and high environmental suitability, a grid cell with high diffusion pressure but low environmental suitability, and a grid with high environmental suitability but low diffusion pressure. If the prediction probability for the first type of grid is significantly higher, it indicates a coupling relationship between diffusion opportunity and environmental suitability. This interaction term is especially important for distinguishing the expansion mechanisms of different crop systems. For rice, if diffusion pressure mainly translates into a new record probability in areas with high humidity suitability, it indicates that rice expansion has strong environmental filtering and moist environment gating characteristics. For millet, if the interaction term contributes weakly, it suggests that expansion is more likely to depend on existing dryland agricultural foundations and regional frontier continuation. For wheat/barley, if the non-local diffusion term contributes more prominently and the moist environment gating is weaker, it indicates that expansion tends more toward regional diffusion and spatial reachability processes.

2.5 Temporal intensity and probability correction

The expansion rates of these three systems also differ during simulated time periods. Millet and rice recorded higher intensity in the middle stage, while wheat/barley increased rapidly in the later stage. To ensure the continuous simulation (see Section 2.6) reflects both the spatial pattern and temporal rhythm of crop-record expansion instead of uncontrolled growth, this study fits the time intensity term $I_c(t)$ for each crop system.

$$I_c(t) = \exp(\gamma_{0,c} + \gamma_{1,c}a(t) + \gamma_{2,c}a(t)^2),$$

Here, $a(t)$ is the standardized age, and $\gamma_{0,c}$, $\gamma_{1,c}$ and $\gamma_{2,c}$ are the time intensity parameters

of crop system c . This term is obtained by fitting the intensity of new crop records observed within each time window and is used to adjust the overall probability of new crops in each window. It should be emphasized that the time intensity term is not an independent causal mechanism nor is it used to explain the specific driving force behind the expansion of a particular crop. Its function is to control differences in recording density between different crop systems and periods, preventing the model from being systematically shifted by a few high-recording or low-recording periods. In other words, the spatial probability term $p_c(x, t)$ represents where new records are more likely to appear, while the time intensity term $I_c(t)$ corrects for differences in overall new record strength across different time windows.

After correcting for time intensity and window length, the final window probability used for simulation is expressed as:

$$\tilde{p}_c(x, t) = \min\{p_c(x, t)I_c(t)Q_c(t), 0.95\},$$

Where $Q_c(t)$ is the window length and window-level intensity correction factor, and 0.95 is the probability limit, used to prevent unrealistic deterministic occupancy of a few high-probability grids in continuous rolling simulations.

2.6 One step and continuous simulation

We use $U_c(x, t)$ to denote the observed or input persistence state used in one-step prediction, and $S_c(x, t)$ to denote the simulated state intensity updated through continuous rolling simulations. In one-step simulation, the model uses real pre-crop records as state input for each time window, so it is mainly used to test window-level forecasting capability. To further test whether the model can maintain a long-term expansion pattern without relying on corrections from real previous observations, this study conducted continuous rolling simulations. The continuous rolling simulation starts with the initial crop state before 6000 BP, followed by updates over time windows.

$$S_c(x, t + \Delta t) = (1 - \lambda_c(t))S_c(x, t) + [1 - S_c(x, t)]\tilde{p}_c(x, t),$$

Where $S_c(x, t)$ is the simulated state before the current window starts, $\tilde{p}_c(x, t)$ is the probability of new records after time intensity and window length correction, and $\lambda_c(t)$ is the state loss rate. This update equation contains two processes. The first term $(1 - \lambda_c(t))S_c(x, t)$ represents the retention of existing crop states, meaning that grids with previously recorded crop or simulated states can continue to maintain a certain state intensity in subsequent windows. The second term $[1 - S_c(x, t)]\tilde{p}_c(x, t)$ represents a grid that previously had no crop state and now gains a new state in the current window based on the probability of new crops. The loss rate $\lambda_c(t)$ is used to represent state decline caused by factors such as decreased crop record intensity, reduced archaeological visibility, regional agricultural system transformation, or weakened local crop use. Continuous simulation is not intended to replace one-step simulation as the most accurate window-level forecast, but rather to test whether the model can produce spatial processes consistent with long-term crop-record expansion patterns even as errors gradually accumulate.

3 Model fitting and parameter selection

For each crop system, model coefficients were fitted using logistic regression with new crop-record occurrence as the binary response variable. The fitted coefficients included the intercept, same-crop local persistence, non-local diffusion, environmental suitability and diffusion-environment interaction terms. We distinguish between diffusion-kernel parameters and logistic occurrence coefficients. Diffusion-kernel parameters, denoted by θ , control how slope, elevation, river proximity, river alignment and directional priors modify the non-local diffusion term $D_c(x, t)$. Logistic coefficients, denoted by β , quantify the effects of same-crop local persistence, non-local

diffusion, environmental suitability and their interaction on crop-record occurrence probability.

Spatial diffusion parameters, environmental suitability profiles and temporal-intensity parameters were selected from predefined candidate sets. These included the distance-decay scale, terrain-cost weights, river-related weights, regional-accessibility weights and crop-specific environmental response profiles. For each candidate parameter combination, the model generated probability surfaces across time windows, and predictive performance was evaluated using average precision, AUC and probability-calibration metrics.

The modelling framework distinguished among fixed design choices, candidate parameters and fitted statistical coefficients. Fixed design choices included the study period, grid system, time-window length, candidate-cell mask and probability ceiling. Candidate parameters included spatial diffusion-kernel parameters and environmental suitability profiles. Logistic coefficients were fitted separately for each crop system. The final parameter configurations for millet, rice and wheat/barley are summarized in Supplementary Table S1.

To avoid over-interpreting single best-fit parameters, we evaluated near-optimal parameter combinations and pseudo-posterior support based on likelihood differences. Parameters with stable support across near-optimal combinations and bootstrap resampling were interpreted as more reliable indicators of crop-specific expansion mechanisms. Parameters with weak or unstable support were interpreted cautiously as predictive weights rather than precise historical-route parameters.

4 Model evaluation and robustness tests

Model performance was evaluated using AUC and average precision. AUC measures whether true new-record cells are ranked higher than non-record cells. Average precision was used alongside AUC because positive crop-record cells were sparse relative to the full candidate grid. For each crop system and time window, the random AP baseline was calculated from the proportion of positive grid cells. Model AP was then compared with this random baseline to assess whether the model enriched true new records in high-probability areas.

Area-constrained evaluation was used to test whether the model predicted true records merely by assigning high probabilities to overly broad areas. Candidate grid cells were ranked by predicted probability, and we calculated the proportion of observed new records captured within the top 1%, 5%, 10% and 20% of predicted areas. Top-k precision was also calculated for fixed numbers of high-probability grid cells.

Spatial null models were used to control for spatial aggregation and archaeological sampling bias. The full model was compared with three null models: a spatially uniform null model, a same-crop KDE null model based on previous records of the same crop system, and a total-crop sampling KDE null model based on all crop records. These comparisons tested whether the full model performed better than random expectation, previous same-crop clustering and general archaeological sampling intensity.

Nested-model comparisons were used to evaluate whether model improvement resulted simply from increasing model complexity. Four model levels were compared:

$$M0: u_c(x, t)$$

$$M1: u_c(x, t) + D_c(x, t)$$

$$M2: u_c(x, t) + D_c(x, t) + E_c(x, t)$$

$$M3: u_c(x, t) + D_c(x, t) + E_c(x, t) + D_c(x, t)E_c(x, t)$$

Here, $M0$ includes only same-crop local persistence, $M1$ adds non-local diffusion, $M2$

adds dynamic environmental suitability, and *M3* is the full interaction model. Models were compared using AUC, AP, AIC, BIC and Brier score.

Probability calibration was evaluated using Brier score and reliability diagrams. Because new crop records are sparse, calibration curves were constructed using predicted-probability quantile bins rather than fixed probability intervals. Predicted probabilities were interpreted as relative probabilities of new archaeological crop-record occurrence, not as absolute probabilities of real crop cultivation.

Temporal-transfer validation was used to test whether model mechanisms generalized across different parts of the study period. The 6000-2000 BP interval was divided into early and late stages. In one experiment, the model was fitted using early windows and tested on late windows. In the reverse experiment, the model was fitted using late windows and tested on early windows. Performance was evaluated against the random AP baseline. If AP remained higher than random expectation in the test period, the model was considered to have cross-stage predictive ability.

Additional robustness tests evaluated environmental collinearity, residual spatial structure, regional validation, chronological uncertainty and spatiotemporal resolution sensitivity. Environmental collinearity was assessed using correlation matrices, variance inflation factors and principal-component structure, and residual spatial autocorrelation was evaluated using Moran's *I* for observed records, predicted probabilities and residuals. Regional validation was also conducted in the Yellow River-North China region, the middle-lower Yangtze region and the Hexi corridor. Chronological-uncertainty Monte Carlo tests randomly perturbed record ages within their uncertainty ranges and recalculated model metrics. Spatiotemporal resolution sensitivity was evaluated by comparing the baseline 300-year, 0.5° setting with 500-year temporal aggregation and 1° spatial aggregation.

5 Mechanistic diagnostics

To identify crop-specific expansion mechanisms, we conducted counterfactual ablation experiments, environmental-gating diagnostics and frontier-mobility analysis.

In the ablation experiments, one mechanism term was removed from the full model at a time while other parameters were kept unchanged. The ablation scenarios included removing same-crop local persistence, non-local diffusion, environmental suitability, terrain cost and river-related terms. For each scenario, we recalculated predicted probabilities and compared AP with the full model:

$$\Delta AP = AP_{\text{ablation}} - AP_{\text{full}}$$

A large negative ΔAP indicates that the removed mechanism contributed strongly to the spatial ranking of new crop records. If AP changed little but the total expected probability changed substantially, the mechanism was interpreted as affecting expansion intensity or probability-surface extent rather than directly controlling the ranking of new record locations.

Environmental filtering was evaluated by grouping candidate grid cells according to environmental suitability quantiles and comparing predicted probabilities and observed new-record frequencies among quantiles. Strong concentration of new records in high-suitability quantiles indicates stronger environmental filtering.

Environmental gating was assessed by comparing predicted probabilities among high-diffusion-high-suitability cells, high-diffusion-low-suitability cells and low-diffusion-high-suitability cells. A strong increase in probability under high diffusion and high suitability indicates that diffusion pressure was more likely to generate new records only when environmental conditions were favourable.

Frontier mobility was assessed by calculating the distance from each new crop-record cell to the nearest previous same-crop record cell. Frontier velocity was defined as:

$$V_c(x, t) = \frac{d_{\min, c}(x, t)}{\Delta t}$$

and converted to kilometres per 100 years. This index does not represent the actual speed of population movement or crop cultivation. Instead, it measures the apparent spatial advance of archaeologically visible crop records. Higher frontier mobility indicates a greater tendency for non-local or leapfrog-style crop-record expansion.

Together, these diagnostics were used to distinguish whether each crop system was primarily associated with same-crop persistence, environmental filtering, non-local diffusion, corridor-mediated accessibility or diffusion-environment coupling.

References

- [1] Gignoux CR, Henn BM, Mountain JL. Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci USA* 2011;108:6044-6049.
- [2] Shennan S, Downey SS, Timpson A, et al. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nat Commun* 2013;4:2486.
- [3] Ammerman AJ, Cavalli-Sforza LL. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton: Princeton University Press; 1984.
- [4] Diamond J, Bellwood P. *Farmers and their languages: the first expansions*. Science 2003;300:597-603.
- [5] Bellwood P. *First Farmers: The Origins of Agricultural Societies*. Malden, MA: Blackwell Publishing; 2005.
- [6] Purugganan MD, Fuller DQ. The nature of selection during plant domestication. *Nature* 2009;457:843-848.
- [7] Jones MK, Liu X. Origins of agriculture in East Asia. *Science* 2009;324:730-731.
- [8] Stevens CJ, Fuller DQ. The spread of agriculture in eastern Asia: archaeological bases for hypothetical farmer/language dispersals. *Lang Dyn Chang* 2017;7:152-186.
- [9] Liu X, Jones PJ, Motuzaite Matuzeviciute G, et al. From ecological opportunism to multi-cropping: mapping food globalisation in prehistory. *Quat Sci Rev* 2019;206:21-28.
- [10] Fuller DQ. Expansion, extension, or retreating? A meta-archaeobotany approach to patterns in crop dispersal. *Veget Hist Archaeobot* 2026. doi:10.1007/s00334-026-01114-6.
- [11] Zhao Z. New archaeobotanic data for the study of the origins of agriculture in China. *Curr Anthropol* 2011;52:S295-S306.
- [12] Yang X, Wan Z, Perry L, et al. Early millet use in northern China. *Proc Natl Acad Sci USA* 2012;109:3726-3730.
- [13] Fuller DQ, Qin L, Zheng Y, et al. The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* 2009;323:1607-1610.
- [14] Zong Y, Wang Z, Innes JB, et al. Holocene environmental change and Neolithic rice agriculture in the lower Yangtze region of China: a review. *Holocene* 2012;22:623-635.
- [15] Stevens CJ, Murphy C, Roberts R, et al. Between China and South Asia: a Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *Holocene* 2016;26:1541-1555.
- [16] Liu X, Lister DL, Zhao Z, et al. Journey to the east: diverse routes and variable flowering times

- for wheat and barley en route to prehistoric China. PLoS One 2017;12:e0187405.
- [17] Liu X, Hunt HV, Jones MK. River valleys and foothills: changing archaeological perceptions of North China's earliest farms. *Antiquity* 2009;83:82-95.
- [18] Liu X, Jones MK, Zhao Z, et al. The earliest evidence of millet as a staple crop: new light on Neolithic foodways in North China. *Am J Phys Anthropol* 2012;149:283-290.
- [19] Zhang C, Hung H-C. The emergence of agriculture in southern China. *Antiquity* 2010;84:11-25.
- [20] Zhuang Y, Ding P, French C. Water management and agricultural intensification of rice farming at the late-Neolithic site of Maoshan, Lower Yangtze River, China. *Holocene* 2014;24:531-545.
- [21] Flad R, Li S, Wu X, et al. Early wheat in China: results from new studies at Donghuishan in the Hexi Corridor. *Holocene* 2010;20:955-965.
- [22] Dong G, Yang Y, Liu X, et al. Prehistoric trans-continental cultural exchange in the Hexi Corridor, northwest China. *Holocene* 2018;28:621-628.
- [23] Deng Z, Qin L, Gao Y, Weisskopf AR, Zhang C, Fuller DQ. From early domesticated rice of the Middle Yangtze Basin to millet, rice and wheat agriculture: archaeobotanical macroremains from Baligang, Nanyang Basin, Central China (6700-500 BC). *PLoS One* 2015;10:e0139885.
- [24] Qiu MH, Du LY, Staff R, et al. Chronology of early China: A radiocarbon databank for Chinese archaeology. *Sci Data* 2025;12:1665.
- [25] He K, Lu H, Zhang J, Wang C. Holocene spatiotemporal millet agricultural patterns in northern China: a dataset of archaeobotanical macroremains. *Earth System Science Data*. 2022;14:4777–4791.
- [26] Danielson JJ, Gesch DB. Global multi-resolution terrain elevation data 2010 (GMTED2010). *US Geol Surv Open-File Rep* 2011;2011-1073:1-26.
- [27] Lehner B, Grill G. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol Process* 2013;27:2171-2186.
- [28] Erb MP, McKay NP, Steiger N, et al. Reconstructing Holocene temperatures in time and space using paleoclimate data assimilation. *Clim Past* 2022;18:2599-2629.
- [29] He, F. & Clark, P. U. Freshwater forcing of the Atlantic Meridional Overturning Circulation revisited. *Nat. Clim. Change* **12**, 449–454 (2022).
- [30] Li Y, Zhang YX, Zhang XZ, et al. A continuous simulation of Holocene effective moisture change represented by variability of virtual lake level in East and Central Asia. *Sci China Earth Sci* 2020;63:1161-1175.
- [31] Smith BD. Niche construction and the behavioral context of plant and animal domestication. *Evol Anthropol* 2007;16:188-199.
- [32] Boivin N, Zeder MA, Fuller DQ, et al. Ecological consequences of human niche construction and domestication on Neolithic and global scales. *Proc Natl Acad Sci USA* 2016;113:6388-6396.
- [33] Fuller DQ, Denham T, Arroyo-Kalin M, et al. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc Natl Acad Sci USA* 2014;111:6147-6152.
- [34] Jaffe YY, Hein A. Considering change with archaeological data: reevaluating local variation in the role of the ~4.2k BP event in Northwest China. *Holocene* 2021;31:169-182.

- [35] Jaffe YY, Castellano L, Shelach-Lavi G, et al. Mismatches of scale in the application of paleoclimatic research to Chinese archaeology. *Quat Res* 2021;99:1-20.
- [36] Kintigh KW, Ingram SE. Was the drought really responsible? Assessing statistical relationships between climate extremes and cultural transitions. *J Archaeol Sci* 2018;89:25-31.
- [37] Bettinger RL, Barton L, Morgan C. The origins of food production in north China: a different kind of agricultural revolution. *Evol Anthropol* 2010;19:9-21.
- [38] Lu H, Zhang J, Liu KB, et al. Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proc Natl Acad Sci USA* 2009;106:7367-7372.
- [39] Barton L, Newsome SD, Chen FH, et al. Agricultural origins and the isotopic identity of domestication in northern China. *Proc Natl Acad Sci USA* 2009;106:5523-5528.
- [40] Fuller DQ, Qin L. Water management and labour in the origins and dispersal of Asian rice. *World Archaeol* 2009;41:88-111.
- [41] Crawford GW. Early rice exploitation in the lower Yangzi valley: what are we missing? *Holocene* 2012;22:613-621.
- [42] Spengler RN III, Stark S, Zhou X, Fuks D, Tang L, et al. A journey to the west: the ancient dispersal of rice out of East Asia. *Rice* 2021;14:83.
- [43] d'Alpoim Guedes JA, Lu H, Hein AM, et al. Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan Plateau. *Proc Natl Acad Sci USA* 2015;112:5625-5630.
- [44] Chen FH, Dong GH, Zhang DJ, et al. Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science* 2015;347:248-250.
- [45] Li X, Dodson J, Zhou X, Zhang H, Masutomoto R. Early cultivated wheat and broadening of agriculture in Neolithic China. *Holocene* 2007;17:555-560.
- [46] Dodson JR, Li X, Zhou X, Zhao K, Sun N, Atahan P. Origin and spread of wheat in China. *Quat Sci Rev* 2013;72:108-111.
- [47] d'Alpoim Guedes JA, Bocinsky RK. Climate change stimulated agricultural innovation and exchange across Asia. *Sci Adv* 2018;4:eaar4491.
- [48] Spengler RN III, Frachetti MD, Doumani PN, et al. Early agriculture and crop transmission among Bronze Age mobile pastoralists of Central Eurasia. *Proc R Soc B* 2014;281:20133382.
- [49] Frachetti MD. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr Anthropol* 2012;53:2-38.
- [50] Hermes TR, Frachetti MD, Doumani Dupuy PN, et al. Early integration of pastoralism and millet cultivation in Bronze Age Eurasia. *Proc R Soc B* 2019;286:20191273.
- [51] Dormann CF, McPherson JM, Araújo MB, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 2007;30:609-628.
- [52] Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17-23.
- [53] Roberts DR, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 2017;40:913-928.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was funded by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan

of the Ministry of Education of China (Grant JYB2025XDXM910), and the National Natural Science Foundation of China (Grant 42371159).

Author contributions

Mingjun Gao: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft.

Yu Li: Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Zhansen Zhang: Data curation, Methodology, Writing – review & editing.

Simin Peng: Data curation, Paleoclimate analysis, Writing – review & editing.

Yuxin Zhang: Data curation, Paleoclimate analysis, Writing – review & editing.

Shiyi Yan: Data curation, Validation.

Hao Shang: Formal analysis, Visualization.

Data and code availability

The original archaeological crop records used in this study are publicly available from Chronology of early China: A radiocarbon databank for Chinese archaeology (<https://doi.org/10.5281/zenodo.17107786>) and A dataset of direct observations of plant macroremains in northern China (<https://doi.org/10.5281/zenodo.6669730>). The processed data, model outputs and analysis code supporting the findings of this study, including the source data underlying Figs. 2–4, are publicly available in Zenodo at <https://doi.org/10.5281/zenodo.23333862>.

Figure

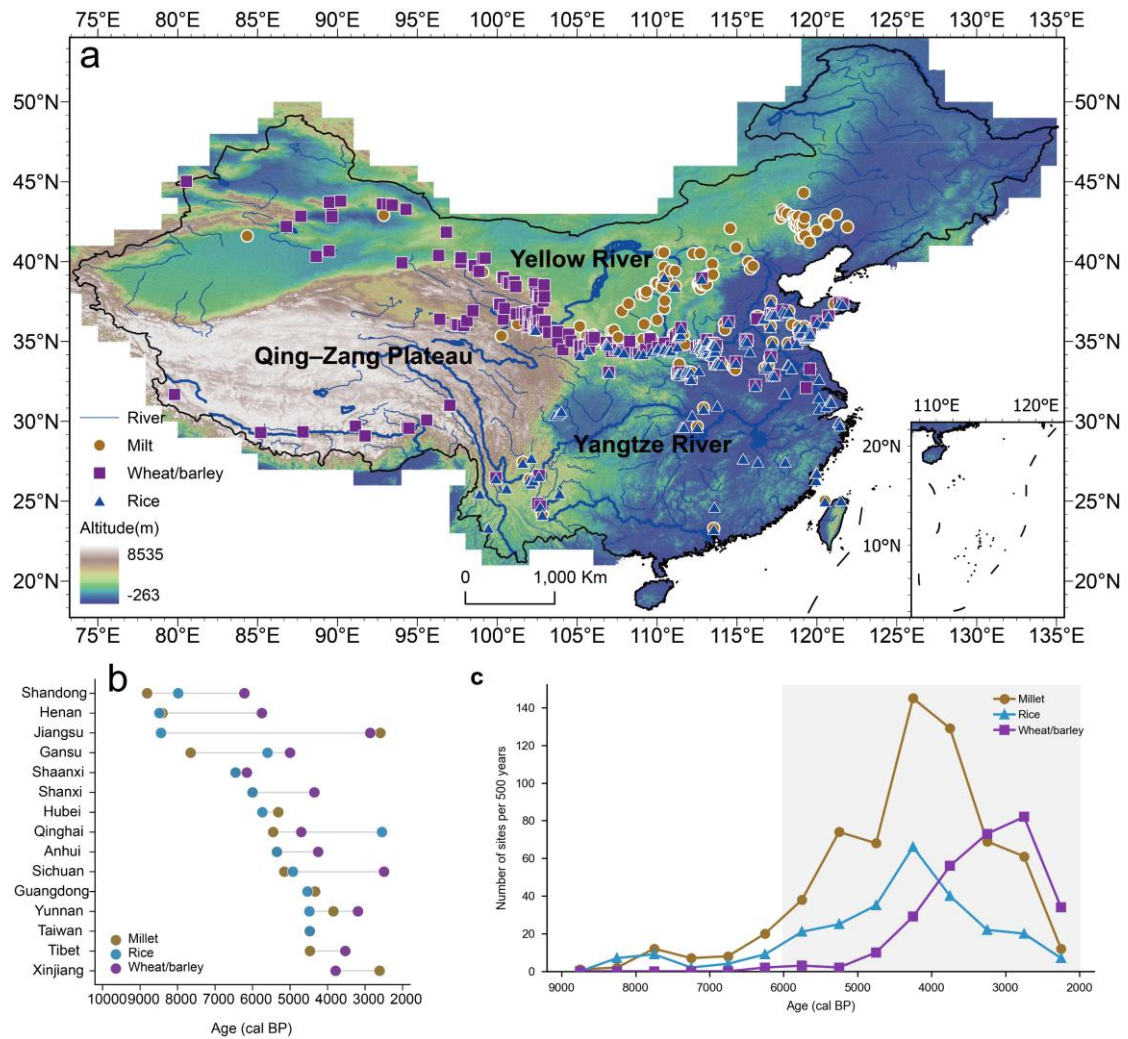


Fig 1 | The spatiotemporal pattern of crop evidence in China during the Holocene. (a) Spatial distribution of millet, rice, and wheat/barley records within China. (b) The earliest emergence time of the three crop systems in different provinces in China. (c) Temporal distribution of the compiled archaeological crop records in 500-year bins. The source datasets primarily cover prehistoric to early historical periods and become incomplete toward and after approximately 2000 BP; therefore, record frequencies younger than 2000 BP are not interpreted as changes in agricultural abundance or cultivation extent.

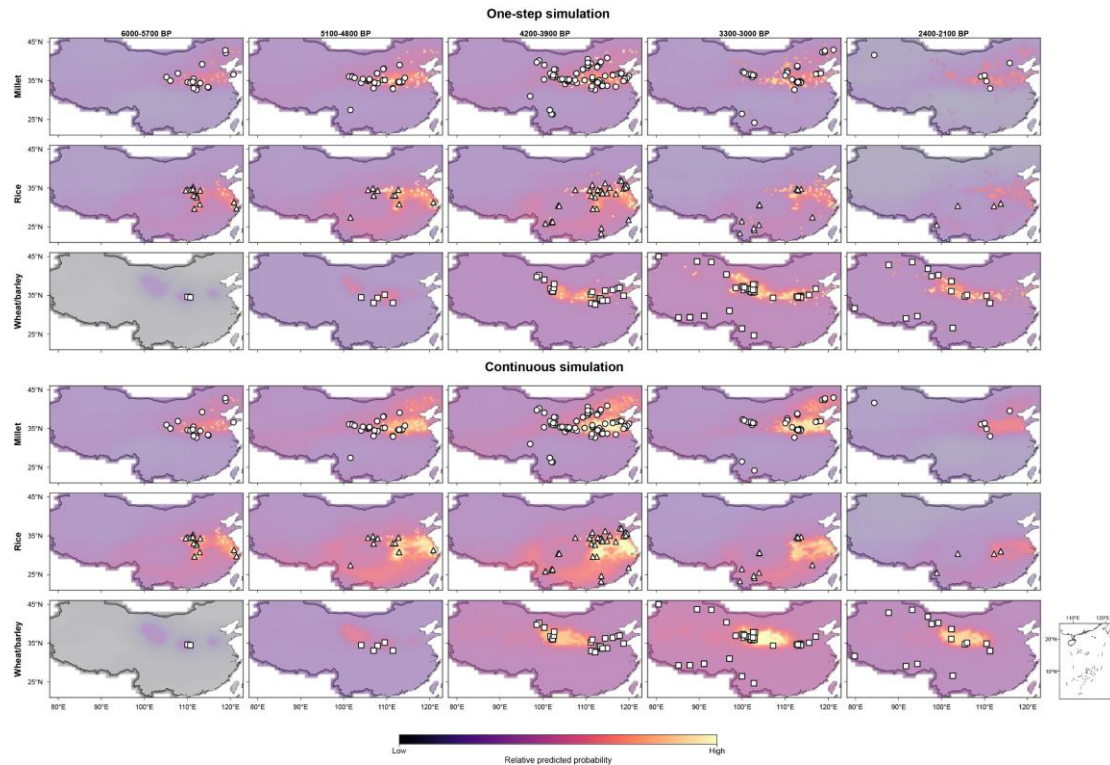


Fig 2 | Dynamic probability surfaces of crop-record expansion. One-step conditional simulations show the predicted probability of new crop-record occurrence based on observed previous crop states. Continuous rolling simulations show simulated crop-record state intensity generated through successive time-window updating. The two simulation modes are used to evaluate short-term spatial prediction and long-term process stability, respectively.

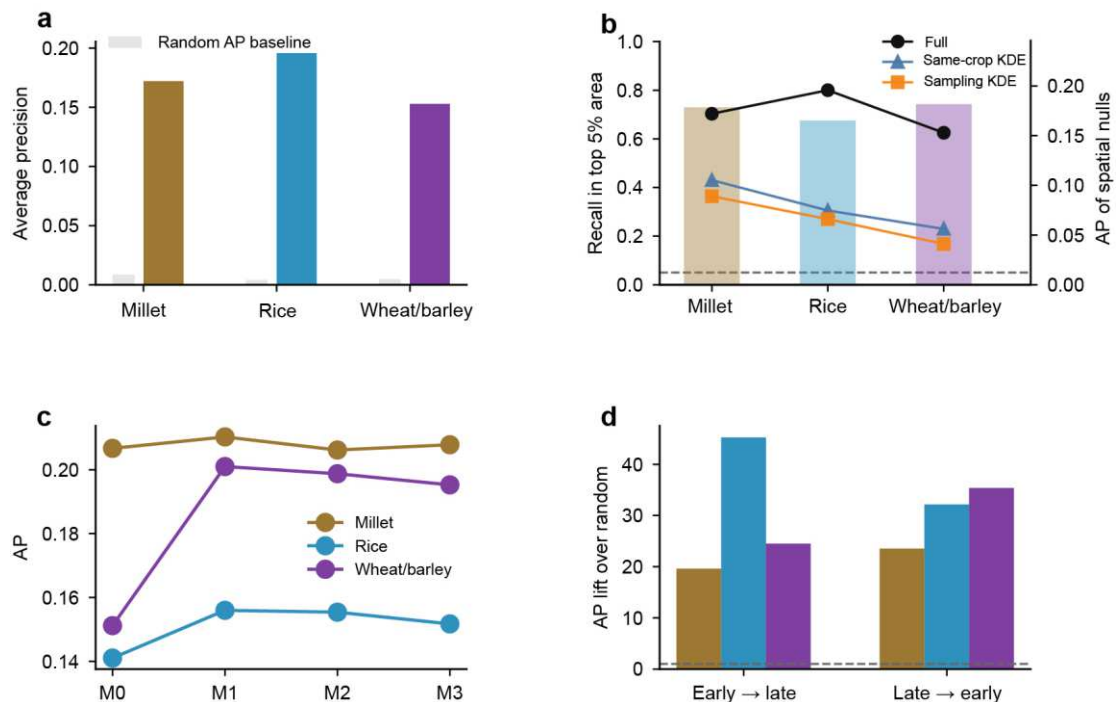


Fig. 3 | Core validation of the dynamic crop-record model. **a**, Average precision (AP) of the full model compared with the random AP baseline for the three crop systems. **b**, Area-constrained recall

and spatial-null comparisons. Bars show the fraction of observed new crop records captured within the top 5% predicted area, and lines show AP values for the full model, same-crop KDE null model and sampling KDE null model. The dashed horizontal line indicates the 5% area expectation. **c**, Nested model hierarchy. M0 includes same-crop local persistence only; M1 adds non-local diffusion; M2 adds dynamic environmental suitability; and M3 adds the diffusion-environment interaction. **d**, Temporal transfer validation, expressed as AP lift over the random baseline. The dashed line indicates random performance(AP lift = 1).

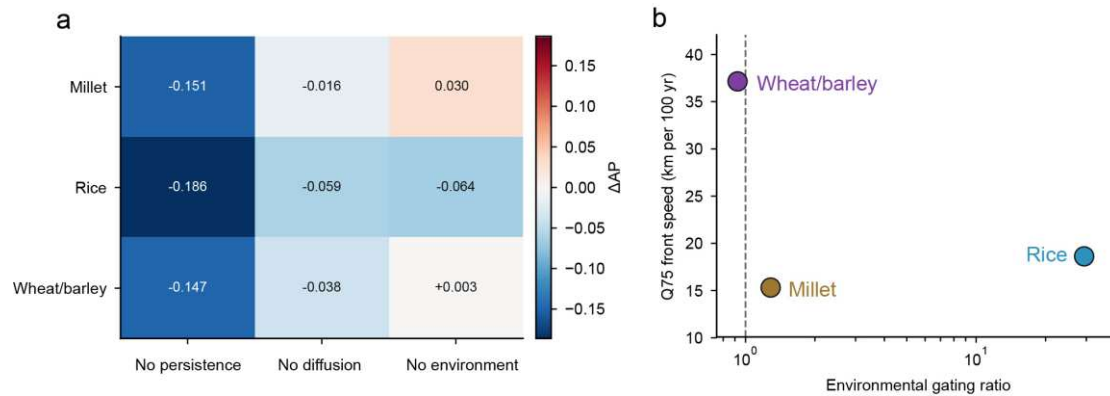


Fig. 4 | Mechanistic diagnostics of crop-specific expansion. a, Core counterfactual ablations showing the change in average precision (ΔAP) relative to the full model after removing same-crop local persistence, non-local diffusion or environmental suitability. Negative values indicate reduced predictive performance relative to the full model. **b**, Crop-specific mechanism space defined by environmental gating and frontier mobility. The x axis shows the environmental gating ratio (logarithmic scale), comparing predicted probabilities in cells with high diffusion pressure under high versus low environmental suitability. The y axis shows the 75th percentile of archaeological frontier velocity. The dashed vertical line marks a gating ratio of 1, indicating no environmental gating effect.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement.docx](#)