

# Mapping Global Avian Influenza Risk Patterns Through Waterbird Activity Entropy

Corresponding Author: Professor Jinwei Dong

**This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.**

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The paper uses multiple machine learning approaches to model monthly distributions of birds species with known AIV infections using a large-scale biodiversity dataset. Using this approach the authors developed a waterbird activity entropy (WAE) value. The WAE is proposed as a metric to inform avian influenza risk at a global scale. The results indicate these models are effective at determining AIV presence based on available surveillance data.

A key strength of this study is its ambitious global scale, allowing for broad comparisons and trend assessments that would be impossible with smaller datasets. However, this wide-reaching approach is also a limitation. It is not clear how the findings can be effectively translated into actionable surveillance or mitigation actions at finer spatial scales. The reliance on a citizen science-based bird dataset raises additional concerns, as user observations introduce significant biases in data collection, including geographical, seasonal, and species-specific reporting inconsistencies. While the authors attempt to mitigate these biases through statistical adjustments, the extent to which these corrections improve data reliability remains uncertain.

My major concern is the role of missing AIV surveillance data. I think at least additional discussion on how this could influence results is needed. The discussion highlights the need for increased surveillance in many data limited countries but the model suggests these are not high risk areas. Is the model suggesting that because they are data limited? In addition, I worry the limits in the citizen science bird data correlate with limit AIV surveillance. I am unsure if this is an effective model or if it has identified areas with strong datasets. I would like to see an exploration of the numbers of bird reports with the number of AIV tests done (or some other way of exploring this potential large bias).

In general, I find this to be a well-executed study at a broad scale with impressive congruence of the WAE to observed AIV. I have some additional minor comments below.

Minor:

Line 66: Over 50 million chickens have been culled in the USA. Unless I misunderstood what this 1.8 million is referencing.

Line 100 -102: The lack of data in certain regions seems to confound your ability to relate your model to influenza. How is this addressed?

Line 507: Please provide at least one additional citation to corroborate these being "widely used in species distribution models" and not only the original proof of concept paper.

Line 514: How were these five environmental variables selected?

The entire additional data section would seem to fit in the environmental data category. I am unclear how this additional data was incorporated into the model as well.

Reviewer #2

(Remarks to the Author)

Overall, this is a very interesting and valuable study. Personally, I really like the author's innovation and exploration in avian influenza risk assessment methods. I also have high expectations for the future use of WAE, and the manuscript is well written. While, I still have some questions that the author needs to consider and answer, which may mainly focus on the discussion section:

1. The author mentioned some limitations of WAE at the end of the discussion, especially the impact of data input on the results. I completely agree, but when it comes to how to improve the data source of WAE in the future, I suggest considering

also the data on poultry or livestock farming patterns (exposure models), and environmental surveillance data (various locations, subtypes of AVI).

2. The author specifically discussed waterfowl and proposed that waterfowl and seabird monitoring should be given priority in the discussion. In fact, over the past few decades, the key position of waterfowl in surveillance, prevention, and research of avian influenza has already been a consensus in the international community. Therefore, can the author propose a new risk indicator, WAE, to accurately identify one or several species of waterfowl and seabirds that require special attention based on WAE, in order to provide more targeted guidance for surveillance and research? For example, in the discussion section, the author mentioned Ciconiiformes (e.g. *Bubulcus ibis*) and Gruiformes. Can we prioritize different species of waterfowl and seabirds based on WAE? Alternatively, based on the WAE, identify areas with higher risk of poultry-waterfowl/seabirds interface, in order to determine the key poultry surveillance areas (as poultry surveillance is easier to implement than waterfowl and seabirds surveillance), which is also an efficient way to improve surveillance system.

3. I think the more likely reason for the difference in prediction accuracy between HPAI and LPAI is that LPAI is less sensitive to identification and reporting than HPAI, resulting in insufficient breadth and depth of data available for WAE. This is also a specific manifestation of the limitation of WAE.

4. The author did not discuss the risks exposed by the H5N1 outbreak in cattle in the United States. I think the current H5N1 outbreak in the United States is related to the local cattle farming patterns, involving the risk of contact between livestock-wild birds interface. I wonder if author got any findings in this regard.

5. The author specifically mentioned surveillance issues in the discussion section, especially in African regions where WAE indicating higher risks. The weak surveillance capabilities of African countries are not only limited to avian influenza, but also pose significant challenges to Africa if surveillance is conducted simultaneously for both humans and birds. Therefore, based on the WAE, can more specialized suggestions or solutions be proposed for Africa, such as prioritizing the establishment or improvement of monitoring for specific species?

6. Finally, it is suggested that the discussion should not repeat too much the content already presented in the results.

#### Reviewer #3

##### (Remarks to the Author)

This manuscript assesses the relevance of a potentially valuable metric, based on Shannon's entropy index, which estimates the spatial distribution of wild waterbird activity and proposes its use as an indicator of avian influenza (AIV) spillover risk. However, the conclusion that WAE predicts AIV case occurrence is weakened by the indiscriminate use of all outbreak data, leading to a likely conflation of initial introductions from wild birds with subsequent amplification events, particularly within poultry populations. With a more focused analytical approach and clearer positioning of the metric's intended application, WAE could become a useful tool for guiding targeted surveillance and mitigation strategies.

The predictive performance of WAE is evaluated using all reported AIV cases, regardless of whether these were infections in wild birds, direct spillovers into other species, or lateral spread among domestic poultry (or other species). This is problematic, as many outbreaks, especially in poultry, result from secondary transmission that is not linked to wild bird exposure. Furthermore, poultry-to-poultry transmission may occur over large geographical distances, well beyond the sites of initial viral incursion from wild birds. As the manuscript states that "over 81% of AIV-positive cases occurred in these high WAE areas", this statistic is potentially misleading if most cases are due to transmission occurring within domestic species. The use of all outbreak data conflates distinct epidemiological processes and may overestimate the utility of WAE as a predictor of true spillover risk. More broadly, the manuscript does not clearly distinguish whether WAE is intended as an indicator of AIV occurrence in wild birds or of spillover from wild birds into other species. This ambiguity weakens the conceptual framework and interpretation of the results.

#### Minor Comments and Clarifications

##### Introduction

Several typographical errors are present throughout the introduction. Proofreading is recommended, particularly to clarify logical flow in some paragraphs.

Line 51: Consider replacing "reassortants" with "reassortments," which may be more appropriate in this context.

Line 53: Referring to AIV emergence as a "progressive" process may not be accurate; many reassortment and host adaptation events are likely punctuated.

Line 55–56: "H5N1 have broadened rapidly in the host range in waterbirds" this sentence contains a grammatical error.

Lines 60–73: The manuscript should more clearly distinguish between spillover from wild birds and subsequent lateral spread among farms.

Line 78: "thus enhancing the risky reassortants" this sentence may be missing a word.

Line 93–94: "it has also been used to involve residency time to quantify the intensity of human activity" the phrasing needs correction.

Line 104–105: "global efforts against to AIV spreading" contains a typographical error.

##### Methods

Line 188: Clarify the type of model used for the correlation analysis and specify whether spatial autocorrelation was accounted for.

Please provide more detail on the cross-validation procedure used for model evaluation.

While code is available, it is important that sufficient documentation is included to ensure reproducibility. For instance, the

described analytical workflow in the README file does not appear to match the structure of the scripts provided.

#### Discussion

Several paragraphs in the discussion section reiterate results without providing further interpretative insight. These sections could be streamlined to focus on implications and limitations.

Line 227–228: “Waterbirds have been verified as the main reservoirs of HPAI H5 and may act as dominant transmission vectors” Poultry have been the primary reservoir of H5 since its emergence in 1996.

Lines 245–249: The discussion on surveillance capacity in less developed countries is vague; this section would benefit from engagement with the extensive literature on surveillance disparities and data quality.

Lines 325–328: “Our study confirms that waterfowl pose a greater transmission risk...” Demonstrating such a link may require exploring the association between the distribution of this specific species group and relevant outbreaks.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors provided a detailed and clear response with corresponding edits to all concerns I have raised. The ability to accurately conduct a study at this scale remains challenging but this work represents a significant advancement on previous approaches.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The author has already responded to and revised all of my questions/comments/suggests in the revision.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I thank the authors for their thoughtful response and the substantial efforts made to revise their manuscript. The improvements are appreciated. However, several core concerns remain insufficiently addressed, particularly regarding the interpretation of outbreak data in relation to wild bird spillover, and the evaluation of the predictive performance of the indicator. These issues continue to affect the robustness and interpretability of the proposed indicator.

A central concern remains the conflation of wild bird-to-poultry transmission events with poultry-to-poultry transmission. In the current analytical pipeline, Step 2 of the outbreak filtering process aims to retain all reported outbreaks not known to result from lateral spread. However, this approach relies heavily on the quality of the underlying surveillance data and on assumptions regarding reporting completeness. I remain concerned that this step retains a large number of outbreaks caused by secondary transmission among poultry. This is particularly problematic in many low- and middle-income countries, where surveillance systems are incomplete and highly pathogenic avian influenza is endemic in poultry populations. In such contexts, reported outbreaks often do not reflect the true epidemiological origin of infection, and lateral spread is frequently under-ascertained. As such, the authors' argument regarding spatial consistency in the proportion of outbreaks attributable to wild bird introductions is not tenable. It highlights the limitations of using globally aggregated outbreak data without adequately addressing spatial and temporal heterogeneity in surveillance sensitivity. If the predictive performance of the metric is evaluated using outcomes that likely include a high proportion of within-poultry transmission, it cannot be interpreted as an indicator of spillover risk. An illustration of this issue is the large number of H7N9 outbreaks retained in the analysis (Fig. S8): to my knowledge, both epidemiological and phylogenetic evidence suggest that this virus is maintained within poultry populations and not in wild birds. It is therefore likely that the majority, if not all, of these outbreaks retained in this analysis reflect lateral spread in poultry, rather than wild bird-to-poultry transmission.

Another major concern is that the reported Area Under the Curve (AUC) values may be misleading. As currently computed, these metrics appear to be heavily influenced by regions with little/no poultry and low wild bird activity. In such areas, low WAE values are trivially associated with the absence of outbreaks, which can inflate apparent model performance. This issue is particularly evident in the case of H7N9: despite its very restricted geographic distribution (China), the WAE model still achieves an AUC of 0.67. This suggests that the performance metric is capturing broad-scale absence patterns rather than fine-scale spillover dynamics. To more accurately assess the predictive relevance of WAE, the evaluation should be restricted to areas with substantial poultry presence, i.e. areas where spillover is plausible, and focus on distinguishing presence from absence within the poultry production landscape. This is where WAE is intended to identify gradients in exposure risk among susceptible host populations. Without such refinement, the current evaluation framework risks overstating the model's capacity to detect meaningful ecological interactions.

In conclusion, while the manuscript presents an innovative metric, the current evaluation framework does not convincingly establish WAE as a predictor of avian influenza spillover from wild birds. The conflation of outbreak types and reliance on broad-scale absence data for validation all weaken the strength of the conclusions. I encourage the authors to refine their validation approach in order to more effectively support the central claims.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #4

(Remarks to the Author)

The new analyses of poultry density gradients clearly demonstrate the robustness of WAE, and the justifications provided for the data filtering process are convincing. With the revisions made, the manuscript is now solid compared to previous versions.

(Remarks on code availability)

**Open Access** This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

## Response to Reviewer #1's Comments

The paper uses multiple machine learning approaches to model monthly distributions of bird species with known AIV infections using a large-scale biodiversity dataset. Using this approach the authors developed a waterbird activity entropy (WAE) value. The WAE is proposed as a metric to inform avian influenza risk at a global scale. The results indicate these models are effective at determining AIV presence based on available surveillance data.

**Response:** Thank you for your constructive feedback and insightful comments. We appreciate your recognition of our study's significance in assessing avian influenza risk at a global scale. In response to your suggestions, we have carefully revised the manuscript to address your major concerns in a comprehensive manner, as detailed below. All line numbers mentioned throughout this response refer to our revised manuscript (clean version).

A key strength of this study is its ambitious global scale, allowing for broad comparisons and trend assessments that would be impossible with smaller datasets. However, this wide-reaching approach is also a limitation. It is not clear how the findings can be effectively translated into actionable surveillance or mitigation actions at finer spatial scales.

**Response:** We appreciate your recognition of the global scope of our study. At the same time, we fully agree that translating our modeling results into actionable surveillance and mitigation strategies at finer spatial scales is facing a great critical challenge. While global-scale analyses have inherent limitations, they still could offer some unique advantages for assessing AIV exposure risk.

First, our study focuses on the natural hosts of AIVs, waterbirds, which are highly mobile and exhibit pronounced long-distance seasonal migration patterns; some of these host species even migrate intercontinentally or wander globally. Therefore, **a global-scale modeling approach is particularly valuable for capturing the overarching dynamics of host activity and exposure risk.** It could provide a strategic overview that complements finer-scale analyses by identifying high-risk countries and regions.

Second, **to better mitigate the influence of data sparsity at finer scales, we employed species distribution models (SDMs) based on observation data and environmental variables rather than relying solely on surveillance data.** This approach allows us to infer exposure risk even in regions with limited surveillance infrastructure, thereby enhancing the robustness and assessment accuracy of our global assessment in finer resolution.

To better discuss the actionable surveillance or mitigation actions at finer spatial scales, we have revised the discussion section to provide suggestions on actionable surveillance and mitigation actions.

(1) Potential implications for actionable surveillance: Suggestions on strengthen regional surveillance in Africa (Lines 248-250 on page 12): “Specifically, Tanzania, Ethiopia, and Zambia each contained hotspots exceeding 34 Mha, yet no AIV cases were recorded, which constrained early detection and epidemic response <sup>1</sup>.” Suggestions on timely and accurate surveillance according host immigration (Lines 218-220 on page 11): “Considering the temporal dimension is crucial, seasonal peaks in waterbird species richness can inform strategically focused surveillance efforts (Supplementary Fig. 5), thereby maximizing the efficiency of limited resources <sup>2</sup>”.

(2) Suggestions on mitigation actions: (Lines 295-298 on page 15): “In addition, in the EU and China, seabird activity poses significant AIV exposure risks, especially in coastal areas <sup>3</sup>, such as Shanghai, Guangzhou, and Rotterdam, where numerous urban water bodies in these regions harbor large populations of waterbirds <sup>4</sup>.” (Lines 303-306 on page 15): “The cattle exposure hotspot areas in the USA (226 Mha) suggest a need for stronger surveillance during this bovine flu outbreak, with targeted mitigation efforts recommended in dairy-intensive states such as Texas, California, and Colorado <sup>5,6</sup>.”

The reliance on a citizen science-based bird dataset raises additional concerns, as user observations introduce significant biases in data collection, including geographical, seasonal, and species-specific reporting inconsistencies. While the authors attempt to mitigate these biases through statistical adjustments, the extent to which these corrections improve data reliability remains uncertain.

**Response:** Thank you for your professional concern. The limitations associated with the reliance on citizen science-based datasets were widely recognized and discussed in current research. Despite inherent biases, databases such as GBIF and eBird are widely employed in species distribution modeling, providing extensive observational records that support large-scale ecological analyses <sup>7,8</sup>. Given their comprehensiveness, these datasets remain the most viable option for global assessments.

In this study, we implemented multiple strategies to enhance the reliability and accuracy of the monthly species distribution results.

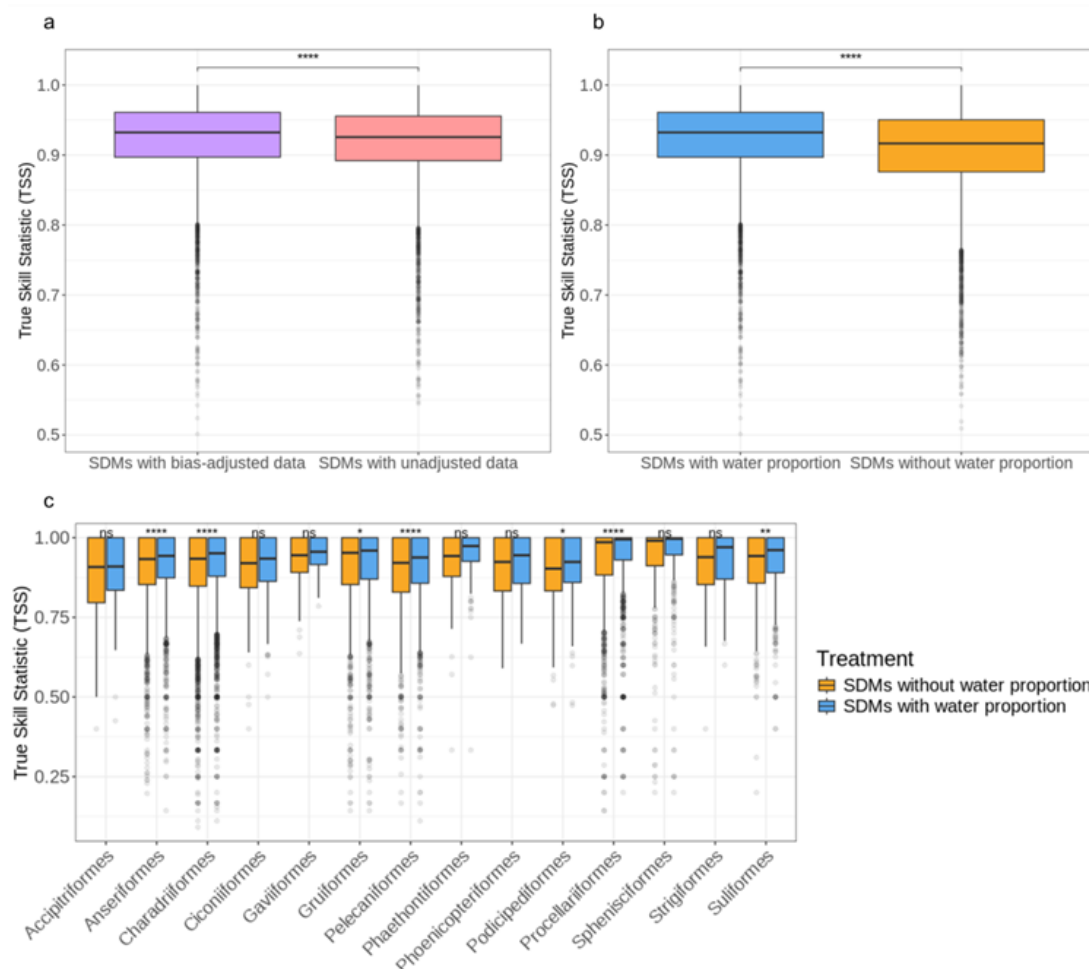
**(1) Added comparison of the True Skill Statistic (TSS) before and after adjusting for observational biases:** To mitigate spatial bias in citizen science data, we applied a sparsification process to reduce clustering and overrepresentation in well-sampled areas. Specifically, we retained only one unique point for each waterbird species per 0.5° grid cell <sup>9</sup>. Additionally, we realize that certain cryptic or human-avoidant species may be underrepresented in observation datasets due to their behavior. However, their human-avoidant characteristics mean they are less likely to appear in human-dominated ecosystems, resulting in lower exposure risks. Therefore, we believe it is acceptable not to include these concealed species. To clarify the extent of these corrections, we compared the True Skill Statistic (TSS) of SDMs developed using bias-adjusted observation data versus unadjusted observation data. As shown in

Supplementary Fig. 15a, SDMs based on bias-adjusted observations result in significantly higher TSS values (mean = 0.923,  $P < 0.001$ ) compared to models using unadjusted data (mean = 0.918). Therefore, the results suggest that the sparsification process improved model accuracy by mitigating observational bias.

In the method section, we reorganized and added the sentences (Lines 526-531 on page 30): “Fourth, to reduce spatial observation bias arising from uneven observation efforts across geographic regions, we applied sparsification procedure by retaining only one unique point for each waterbird species per 0.5° grid cell<sup>9</sup>. The comparison results (Supplementary Fig. 15a) showed that SDMs produced significantly higher TSS values after adjusting biases (mean = 0.923,  $P < 0.001$ ), compared to the unadjusted results (mean = 0.918).”

**(2) Involving more suitable environment variables to improve the accuracies of SDMs:** Due to seasonal variability of environment factors, monthly environment data were used in our study, to produce more robust, temporally explicit models<sup>10</sup>. Furthermore, we have considered the land use data to improve the precision<sup>11,12</sup>, specifically, given the importance of water to the waterbird species, we introduced the proportion of open water as a key driver variable into waterbird monthly SDM construction (Lines 539-541 on page 30-31). We compared SDMs with water proportion (Supplementary Fig. 15b, c), which generally demonstrate higher accuracy than SDMs without water proportion.

In conclusion, by addressing observational biases through statistical controls and incorporating ecologically relevant environmental variables, we improved the robustness and performance of our monthly distribution models.



**Supplementary Fig. 15 Comparison of the True Skill Statistic (TSS) before and after adjusting for observational biases.** (a) Comparison of the TSS of conducting SDMs based on bias-adjusted observation data and unadjusted observation data. Purple boxplots represent SDMs without bias-adjusted observation data, whereas red boxplots represent SDMs with unadjusted observation data. (b) Comparison of the TSS of conducting SDMs with and without water proportion. Blue boxplots represent SDMs with water proportion, whereas orange boxplot represent SDMs without water proportion. (c) Comparison of the TSS of conducting SDMs with and without water proportion across different orders. We combined the outcomes of various modeling approaches to simulate the distribution patterns of waterbird species, enabling a comprehensive evaluation of model performance and predictive accuracy.

My major concern is the role of missing AIV surveillance data. I think at least additional discussion on how this could influence results is needed. The discussion highlights the need for increased surveillance in many data limited countries but the model suggests these are not high risk areas. Is the model suggesting that because they are data limited?

**Response:** We appreciate your thoughtful comment. We would like to clarify the role of missing AIV surveillance data and its potential impact on our results from several

perspectives as follows, and accordingly more discussion were added in the revised manuscript.

**(1) The WAE is calculated from monthly host species distribution simulations, and does not incorporate the AIV surveillance data during its formulation** (Lines 600-606 on page 33-34, Supplementary Fig. 4). Instead, the AIV surveillance data is used to evaluate the reliability and predictive capability of WAE in indicating AIV infection occurrences, and subsequently, to determine the optimal threshold for delineating high-risk areas of waterbird host exposure (Fig. 2). Therefore, while missing AIV surveillance data does not directly affect WAE itself, it can influence the determination of the optimal threshold used to classify regions as high-risk or low-risk. For instance, we identified imbalanced surveillance data for infected animals in Southeast Asia, where only 4% of reported cases in China were wild bird infections, compared to 52% in the EU. This discrepancy impacts risk classification. In Southeast Asia, the issue can be mitigated by incorporating environmental sample data and human-affected records, and selecting poultry cases more carefully (Supplementary Fig. 7). However, the missing data in African countries remains unresolved due to the factors affecting surveillance coverage. Although our AUC results demonstrate that WAE maintains strong predictive power (Line 137-138 on page 7), substantial gaps in disease monitoring can still affect the calibration of risk classification thresholds, which may lead to mischaracterization of high-risk zones, such as overgeneralization or underestimation. These limitations highlight the urgent need to strengthen AIV surveillance systems in under-monitored regions, particularly across eastern sub-Saharan Africa, parts of Southeast Asia. By bridging surveillance gaps, especially where WAE indicates elevated host exposure but surveillance data remain sparse, local authorities can deploy timely interventions to mitigate spread at the onset of epizootics.

**(2) Limited or missing regional surveillance data does not necessarily lead to low-risk prediction, as the WAE robustly identifies areas of elevated AIV host exposure based on heightened waterbird activity.** This structural and systematic approach allows WAE to avoid biases introduced by uneven surveillance coverage. In our results, the identification of high-risk areas based on WAE does not strictly align with regions that have more AIV surveillance data, and vice versa. In Africa, 90% of reported cases are concentrated in Egypt, Nigeria, and South Africa, yet these countries account for only 19% of Africa's high-risk areas. This suggests that many African nations with high waterbird activity remain under-monitored for AIV. As reflected in Fig. 2a, large high-risk regions in Eastern Africa have almost no surveillance records (Lines 265-267 on page 13). Conversely, in the west-central USA, a high volume of reported cases falls within low-risk regions, suggesting it may be less critical for targeted surveillance efforts compared to eastern USA high-risk zones. These patterns suggest that we should not misestimate the avian influenza risks by relying solely on current surveillance distributions or intensity. More importantly, it should not be used as a guide for allocating future surveillance resources.

While missing AIV surveillance data influences the threshold of high-risk WAE,

our work still robustly emphasizes the importance of highlighting AIV surveillance data-limited countries/regions with intensive waterbird host activity. Many of these regions are less developed, reinforcing the need to allocate surveillance resources equitably to ensure fairness in influenza epidemic prevention and risk management.

To more clearly emphasize these traits, **we added more statements in the discussion, specifically, the “Uncertainty and Implications” section** (Lines 369–377 on page 18): “In this study, AIV surveillance data was used to assess the reliability and predictive capability of the WAE in identifying AIV cases likely resulting from spillover events involving waterbirds, and subsequently to determine the optimal threshold for delineating high-risk areas of waterbird host exposure (see Methods). However, missing surveillance data and substantial gaps in disease monitoring may still affect threshold calibration, potentially resulting in either overgeneralization or underestimation of high-risk zones. To address this, enhanced surveillance efforts are urgently needed in less developed regions, particularly in eastern sub-Saharan Africa, where WAE indicates high-risk areas yet recorded AIV cases remain scarce.”

In addition, I worry the limits in the citizen science bird data correlate with limit AIV surveillance. I am unsure if this is an effective model or if it has identified areas with strong datasets. I would like to see an exploration of the numbers of bird reports with the number of AIV tests done (or some other way of exploring this potential large bias).

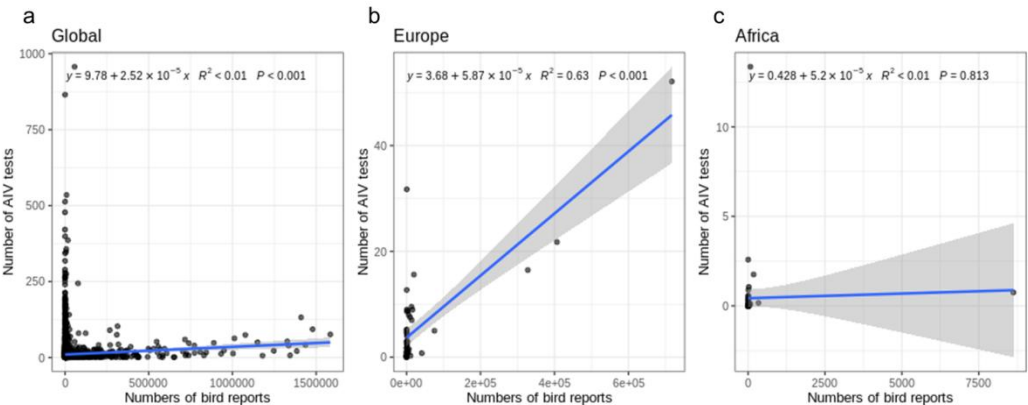
**Response:** Thank you for your thoughtful comments. **To address your concern, in the revision we compared the number of bird observation reports with the number of AIV tests conducted per country across Africa and Europe** (Fig. R1). In Europe, we observed a positive correlation between citizen science bird data availability and AIV surveillance efforts, suggesting a degree of co-distribution in data richness ( $R^2=0.63$ ,  $P<0.001$ ). In contrast, this relationship was negligible and not statistically significant in Africa.

Furthermore, it’s important to note that **the high-risk areas identified by our model do not always align with regions of intensive AIV testing**. This disconnect is particularly striking in Africa. As shown in Fig. R2, we present the proportion of high-risk area alongside the proportion of AIV testing across African countries. Notably, the top ten countries identified by our model account for over 60% of the total high-risk areas (approximately 348 million ha in total), yet these same countries report only 11% of AIV cases (885 cases). Among them, seven countries, including Tanzania, Ethiopia, Zambia, Zimbabwe, Madagascar, Angola, and Kenya, each contain more than 20 million hectares of high-risk area yet have no recorded AIV detections.

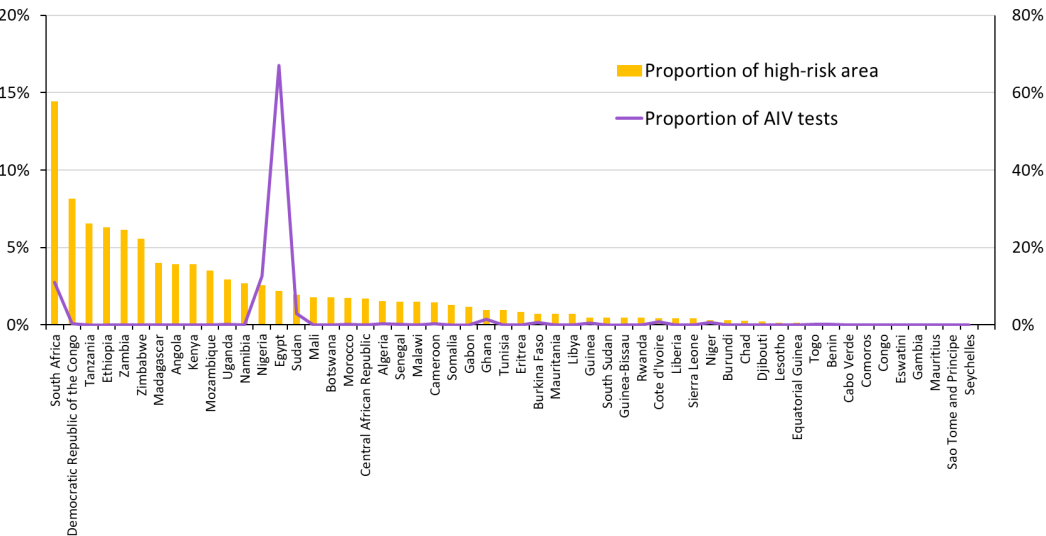
**The analysis reveals significant regional disparities in the relationship between citizen science bird observation data and avian influenza surveillance**

**data globally.** Some countries/regions prioritize disease surveillance but lack host animal monitoring, while some others maintain abundant wilderness tourism and wildlife observation records, yet allocate limited resources to disease surveillance. Thus, based on citizen science bird data, our finding underscores a significant mismatch between AIV potential risk and current surveillance coverage, suggesting that data-poor (both disease surveillance and host species observation) regions may warrant greater attention in future monitoring strategies.

**To highlight this concern, we have added more statements in discussion** (Lines 390–394 on page 19): “Overall, disparities in citizen science observation effort and AIV surveillance intensity across countries and regions may partially affect the delineation of high-risk areas. Our analysis still provides meaningful guidance, particularly in identifying data-deficient high-risk regions where enhanced host observation and pathogen monitoring are urgently needed.”



**Fig. R1** Linear relationships between bird observation reports and AIV test counts per grid in (a) Global, (b) Europe, and (c) Africa, based on linear regression analysis.



**Fig. R2** Proportions of high-risk area and AIV tests for each African country,

calculated as each country's share of the total high-risk area and AIV reports across Africa.

In general, I find this to be a well-executed study at a broad scale with impressive congruence of the WAE to observed AIV. I have some additional minor comments below.

**Response:** Thank you for your positive evaluation of our study. We appreciate your recognition of the broad-scale execution and the strong alignment of WAE with observed AIV patterns. We have carefully reviewed and incorporated revisions based on your comments below.

Minor:

Line 66: Over 50 million chickens have been culled in the USA. Unless I misunderstood what this 1.8 million is referencing.

**Response:** Thanks for pointing out this issue. We found that 1.8 million chickens specifically pertain to Colorado's situation (*Colorado Faces Major Bird Flu Outbreak: 1.8 Million Chickens to Be Culled – CareDoctor*). And we include 50 million chickens data in the revised sentence (Lines 61-64 on page 3-4): "In late August 2024, 50% of herds in Colorado were reported to be affected, poultry infections led to the culling of nearly 1.8 million chickens, while nationwide, the situation was more severe, with over 50 million chickens culled<sup>13</sup>".

Line 100 -102: The lack of data in certain regions seems to confound your ability to relate your model to influenza. How is this addressed?

**Response:** Thanks for your comment. We acknowledge the uneven availability of surveillance data across regions. To address this imbalance, we incorporated diversified surveillance data, including environmental sample data and human infection records. Besides, we carefully curate poultry case data (Supplementary Fig. 7), which results in scarification surveillance data in intensive monitoring areas. By spatial sparsity of the surveillance data (see Supplementary Text 2: Statistical analysis on page 4-5), we have partially mitigated the potential impacts caused by imbalances or deficiencies in specific types of surveillance data records.

However, missing surveillance data in specific African countries remains unresolved due to insurmountable limitations. Despite these gaps, our AUC results demonstrate that WAE maintains strong predictive power, serving as a reliable indicator for assessing AIV exposure risk and guiding surveillance priorities (Line 137-138 on page 7). Additionally, we have explicitly discussed these impacts and underlying causes within discussion sections (Lines 245-254 on page 12-13).

Line 507: Please provide at least one additional citation to corroborate these being "widely used in species distribution models" and not only the original proof of concept paper.

**Response:** We have incorporated additional citations at the relevant section to further substantiate the claim that WorldClim environmental variables are widely used in species distribution modeling. The sentence has been updated to (Line 537-539 on page 30): “WorldClim data are widely used in species distribution models (SDMs) and include monthly climatic variables that were averaged from 1970-2000<sup>14-17</sup>.” The additional citations are:

- 1) Carlson, C. J. et al. Climate change increases cross-species viral transmission risk. *Nature* 607, 555-562 (2022).
- 2) Liu, X. et al. Animal invaders threaten protected areas worldwide. *Nature Communications* 11, 2892, doi:10.1038/s41467-020-16719-2 (2020).
- 3) Crego, R. D., Stabach, J. A. & Connette, G. Implementation of species distribution models in Google Earth Engine. *Diversity and Distributions* 28, 904-916 (2022).

Line 514: How were these five environmental variables selected?

**Response:** Our study aims to model the monthly species distribution of waterbirds, and therefore, we primarily utilized WorldClim’s historical monthly weather data, which only includes three core environmental variables: minimum temperature, maximum temperature, and precipitation. Therefore, these variables were directly incorporated into the model rather than selected through a filtering process. However, relying on a limited number of environmental predictors in SDMs can reduce model flexibility and may fail to capture important ecological gradients, potentially leading to underfitting or oversimplification of species-environment relationships<sup>18</sup>.

Therefore, we incorporated two supplementary variables: elevation (m) and water proportion (%). Elevation was included as it plays a crucial role in waterbird habitat suitability, migratory patterns, and climate adaptation. Land cover or land type has also been widely used as a key driver in SDMs<sup>11,12</sup>. Therefore, water proportion (%) was selected due to its essential role in defining suitable habitats for waterbirds, as water availability directly impacts their foraging and breeding. Finally, we chose five variables for monthly species distribution predictions.

To ensure clarity, we have explicitly stated in the manuscript that our environmental dataset consists of monthly weather data. Additionally, we have revised the term “selected” to “used” to better reflect our approach.

The entire additional data section would seem to fit in the environmental data category. I am unclear how this additional data was incorporated into the model as well.

**Response:** Sorry for the unclear statement. The additional data mentioned in the manuscript were primarily utilized in subsequent Mantel tests (Supplementary Fig. 16) to explore drivers of WAE in high exposure risk regions and did not contribute to the species distribution modeling process. We have clarified this distinction to enhance

314 transparency in data utilization (Lines 551-554 on page 31): “To analyze the drivers  
315 of WAE, we selected 16 potential variables and classified them into three groups  
316 (geographic and climate factors, vegetation factors, and human activity factors) to  
317 conduct Mantel tests (Supplementary Table 2).”  
318

## Response to Reviewer #2's Comments

Overall, this is a very interesting and valuable study. Personally, I really like the author's innovation and exploration in avian influenza risk assessment methods. I also have high expectations for the future use of WAE, and the manuscript is well written. While, I still have some questions that the author needs to consider and answer, which may mainly focus on the discussion section:

**Response:** We greatly appreciate your recognition of our study's innovative approach to assessing avian influenza risk and your enthusiasm for the future applications of WAE. All your comments and questions were carefully considered, and hopefully, our subsequent one-by-one responses will satisfy for your concerns. All line numbers mentioned throughout this response refer to our revised manuscript.

1. The author mentioned some limitations of WAE at the end of the discussion, especially the impact of data input on the results. I completely agree, but when it comes to how to improve the data source of WAE in the future, I suggest considering also the data on poultry or livestock farming patterns (exposure models), and environmental surveillance data (various locations, subtypes of AVI).

**Response:** Thanks for your valuable comments. We appreciate your insights on enhancing the data sources for future WAE analysis.

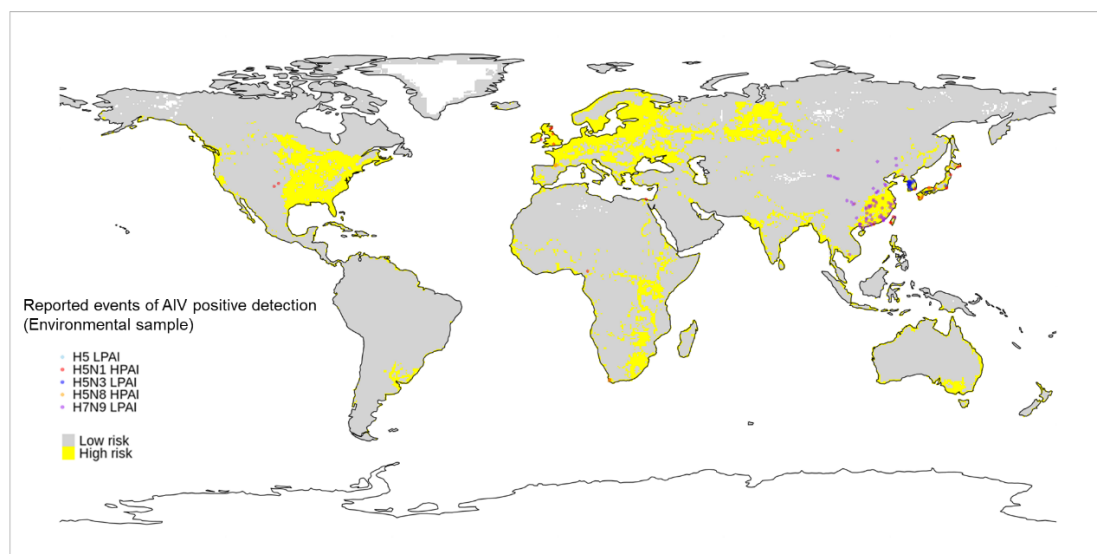
**(1) Poultry and livestock farming patterns play a critical role in shaping the consequences of waterbird host exposure.** While this study primarily examines the spatial density of cattle and poultry farming (Fig. 3), we also recognized and discussed the distinct risk differences between backyard, smallholder, and industrial farming systems across various management settings (Lines 329-341 on page 16-17). However, due to a lack of accurate and fine-resolution data on poultry or livestock farming patterns, we look forward to being able to incorporate such farm-scale exposure models and explore them in the future to improve our understanding of poultry or livestock exposure risk, as well as the deeper utility of WAE.

**(2) Environmental surveillance data can indeed play a critical role in enhancing our understanding of virus persistence and transmission dynamics in ecological contexts, particularly in the absence of confirmed host detections.** However, current environmental surveillance data remain limited. For example, of the 56,300 records collected from FAO, only 380 pertain to environmental samples, and these are primarily concentrated in southeastern China (Fig. R3). Given this spatial bias and low sample size, we believe that evaluating WAE performance based solely on purely environmental surveillance data may be premature. Expanding and standardizing the global collection and sharing of environmental AIV surveillance data will be essential to fully realize its potential in future assessments.

While conducting corresponding analyses remains challenging due to current limitations in data quantity and quality, we still fully endorse your professional

insights that considering the data on poultry or livestock farming patterns, and environmental surveillance data in the future could provide a more comprehensive representation of avian influenza risk and refine future predictions.

**Following your valuable insights, we have further improved our discussion section** (Lines 377-382 on page 18-19): “Enhancing environmental surveillance data, particularly in terms of geographic coverage and subtype resolution, is also needed to improve the accuracy of high-risk region identification. In addition, for exposure models, accounting for specific farming practices in future research is essential, as neglecting production mode differences may lead to overestimation of risk.”



**Fig. R3** Spatial distribution of high and low binary exposure risk areas, overlaid with the locations of reported environmental sample detections.

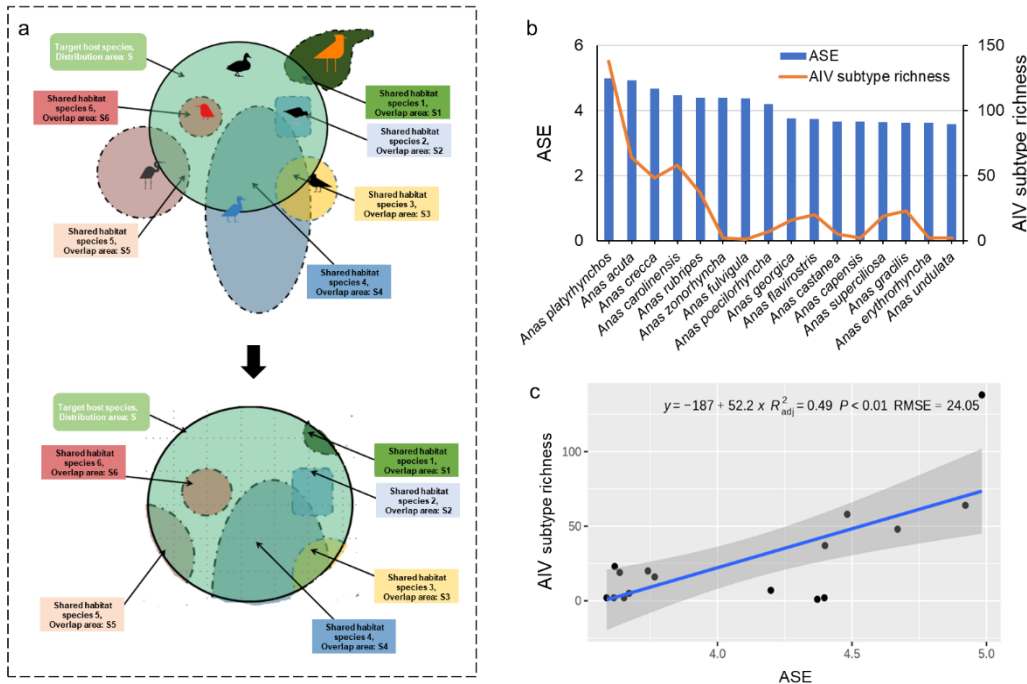
2. The author specifically discussed waterfowl and proposed that waterfowl and seabird monitoring should be given priority in the discussion. In fact, over the past few decades, the key position of waterfowl in surveillance, prevention, and research of avian influenza has already been a consensus in the international community. Therefore, can the author propose a new risk indicator, WAE, to accurately identify one or several species of waterfowl and seabirds that require special attention based on WAE, in order to provide more targeted guidance for surveillance and research? For example, in the discussion section, the author mentioned Ciconiiformes (e.g. *Bubulcus ibis*) and Gruiformes. Can we prioritize different species of waterfowl and seabirds based on WAE?

**Response:** Thank you for your insightful suggestion. We greatly appreciate your suggestion to refine the WAE framework for identifying key waterbird and seabird species that warrant enhanced surveillance. **While WAE effectively captures aggregated waterbird activity intensity, it could not inherently rank individual species based on their relative importance in virus transmission.**

**For clarification, we have revised discussion to better convey this limitation**

(Lines 383-390 on page 19): “While WAE effectively captures the aggregated activity intensity of waterbird groups, it does not inherently rank individual species by their relative importance in virus transmission. In this study, all species contribute relatively equally to WAE calculations and are affected primarily by residence time. Future improvements could integrate more host competence factors<sup>19</sup> for evaluating host exposure and spillover risk importance and contributions, such as population density, host suitability, or susceptibility, to reflect better contributions of host species and enhance WAE’s predictive accuracy in avian influenza risk.”

To support species-level prioritization, we are developing a new approach based on Avian Influenza Virus Shared Entropy (ASE)—a landscape overlap index that quantifies spatial co-occurrence and distributional overlap between target host species, in our coming efforts. This method enhances the traditional Shannon index to assess interspecies interaction strength (Fig. R4a). Applying ASE to several *Anas* species (Fig. R4b), *Anas platyrhynchos* exhibited the highest ASE value and also harbored over 100 AIV subtypes. Linear regression analysis (Fig. R4c) confirmed that ASE is positively associated with subtype richness of specific host species, supporting its potential to identify key transmission species. However, identifying and evaluating the roles and epidemiological significance of specific host species in avian influenza transmission constitutes a complex and relatively independent research endeavor, incorporating this new dimension into the present study focused on risk zone delineation would be premature. The detailed methods are addressed in another manuscript in preparation, which concentrated on high-risk species assessment.



**Fig. R4** Conceptual framework of Avian Influenza Virus Shared Entropy (ASE). (a) Schematic illustration of ASE calculation. (b) Comparison of ASE and AIV subtype richness. (c) Linear regression analysis of ASE vs AIV subtype richness.

richness among *Anas* species. (c) Linear regression analysis of ASE and subtype richness across *Anas* species.

Alternatively, based on the WAE, identify areas with higher risk of poultry-waterfowl/seabirds interface, in order to determine the key poultry surveillance areas (as poultry surveillance is easier to implement than waterfowl and seabirds surveillance), which is also an efficient way to improve surveillance system.

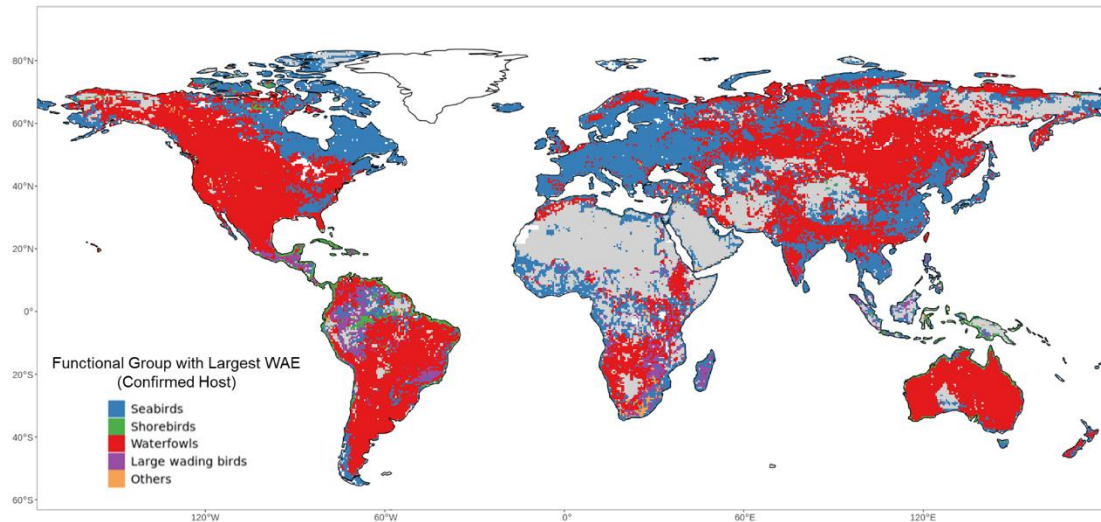
**Response:** We appreciate your insightful suggestion. WAE offers several options for identifying high-risk poultry surveillance areas and functional host groups of concern:

**(1) Poultry-waterfowl/seabirds interface risk: in the revision, we overlaid waterfowl and seabird WAE with poultry density to delineate interface zones with elevated spillover potential** (Supplementary Fig. 12). The top three regions with extensive high-risk overlap are the EU, southeastern China, and the eastern USA. These areas can benefit from enhanced poultry-focused surveillance efforts, and targeted strategies could be developed to reduce interactions between poultry and wild waterfowl or seabirds, thereby mitigating potential transmission risks.

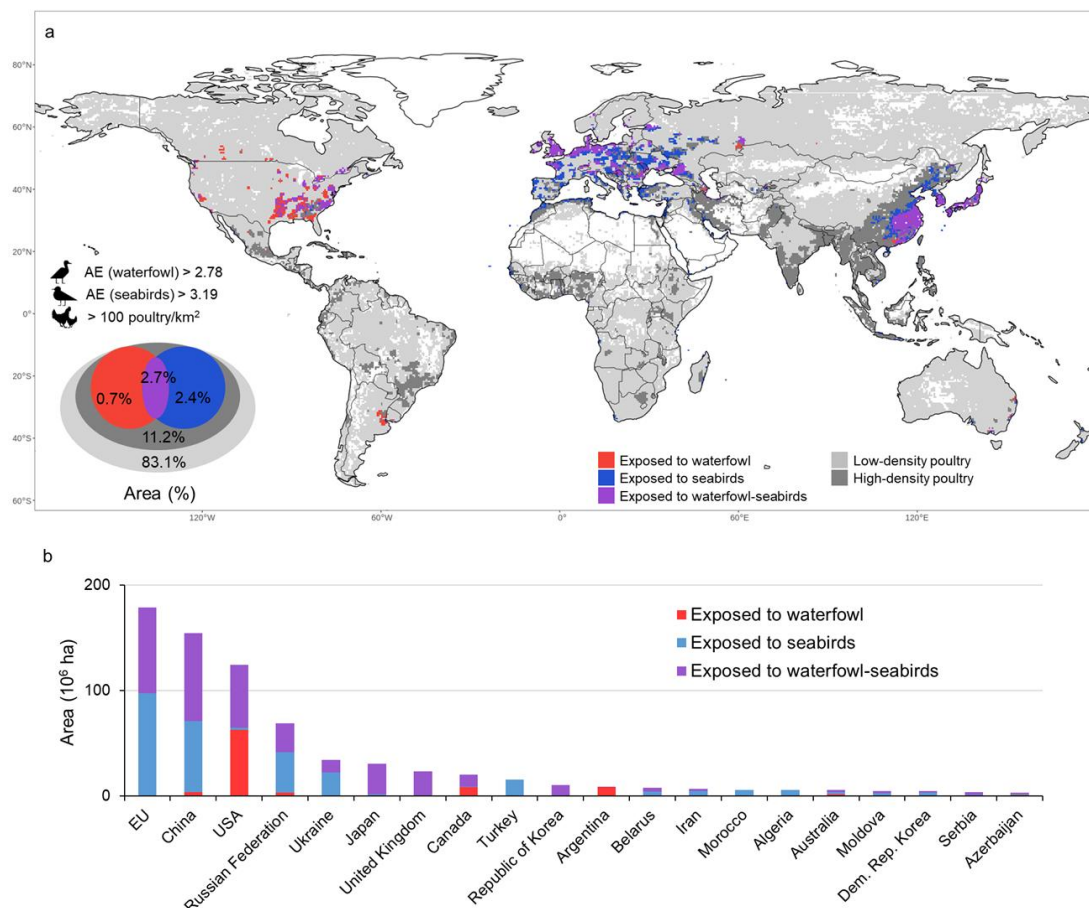
**(2) Functional group – poultry correlations:** in the revision, we identified distinct poultry–functional group associations across four key exposure countries/regions, based on linear correlation analyses (Fig. 4, Supplementary Table 3). In the USA, seabirds exhibited the strongest positive correlation with poultry, indicating they may warrant surveillance priority. In China, both waterfowl and seabirds showed high correlations, while in the EU and India, large wading birds were more strongly linked to poultry. These findings highlight region-specific exposure interfaces between poultry and particular avian functional groups.

**(3) Prioritization by WAE values:** By identifying functional groups with the highest WAE values, we found that seabirds and waterfowl consistently represent higher-priority hosts compared to other confirmed groups, both within key exposure countries/regions (Supplementary Fig. 11) and globally (Fig. R5).

**To reflect the importance of poultry-waterfowl/seabird interface risk, we added more statements in discussion** (Lines 336-340 on page 17): “Given the recent emergence of seabirds as HPAI hosts across multiple continents<sup>20,21</sup>, marked by their high activity entropy with waterfowl in regions such as the central EU and southeastern China (Supplementary Fig. 12), there is also a parallel need to strengthen biosecurity in industrial farms with high-density environments.” Although waterfowl and seabirds are central to most high-risk regions, avian influenza-infected waterbird species beyond waterfowl and seabirds are rapidly increasing; therefore, it remains essential to consider evaluating the risk of the full spectrum of waterbird functional groups.



**Fig. R5** Global distribution of waterbird functional groups exhibiting the highest waterbird activity entropy (WAE). This map highlights the dominant functional group in each area, defined as the group with higher WAE values relative to others.



**Supplementary Fig. 12** Global hotspots of high-density poultry exposed to activity entropy (AE) of waterfowl and seabirds. (a) Overlay map illustrating the spatial convergence of high-density poultry, waterfowl AE, and seabird AE. Thresholds for each variable were defined as the top 15% quantile of their respective distributions. (b)

Statistics on the risk areas stratified by exposure to waterfowl AE, seabird AE, or both, across the top twenty countries/regions.

3. I think the more likely reason for the difference in prediction accuracy between HPAI and LPAI is that LPAI is less sensitive to identification and reporting than HPAI, resulting in insufficient breadth and depth of data available for WAE. This is also a specific manifestation of the limitation of WAE.

**Response:** Agreed. Including LPAI in the analysis is important given its potential role in viral evolution through genetic reassortment. Despite being underreported, LPAI viruses can act as precursors to highly pathogenic strains, as demonstrated by past emergence events such as the evolution of H5N8 and other reassorted subtypes<sup>22</sup>. Indeed, detection bias plays a significant role in shaping the observed differences in prediction accuracy between HPAI and LPAI. LPAI is generally less sensitive (Line 143-144 on page 7), and its data availability and coverage are comparatively limited, which may affect the accuracy of WAE-based risk assessments.

**We appreciate your valuable insight and have further refined our discussion to ensure that this limitation is clearly articulated** (Lines 230-232 on page 11-12): “Although LPAI viruses are often underreported, they may serve as evolutionary precursors to highly pathogenic strains, as illustrated by past emergence events such as the evolution of H5N8 and other reassorted subtypes<sup>22</sup>.”

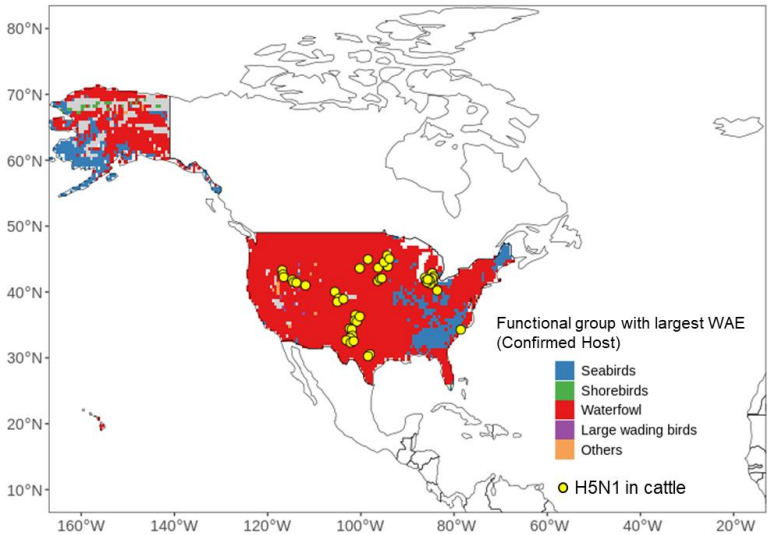
4. The author did not discuss the risks exposed by the H5N1 outbreak in cattle in the United States. I think the current H5N1 outbreak in the United States is related to the local cattle farming patterns, involving the risk of contact between livestock-wild birds interface. I wonder if author got any findings in this regard.

**Response:** Yes, the outbreak of H5N1 in cattle in the United States is a significant concern, particularly regarding its relationship with local cattle farming patterns and the potential livestock – wild bird interface risks. Our analysis suggests (Supplementary Fig. 13) that, based on waterbird activity dominance and intensity, **the waterfowl are the most likely original source of AIV exposure in affected regions**, although some granivorous passerines perhaps also play significant roles in the final spillover stages. **Besides, among confirmed host records, waterfowl dominate large portions of the USA, including areas where H5N1 outbreaks in cattle have been reported.**

**These spatial patterns and supporting evidence imply that waterfowl may play a central role in the initial spillover of H5N1 to cattle<sup>23</sup>.** However, the mechanisms of inter-state transmission remain uncertain, whether through wild bird migration, human-mediated livestock movement, or other pathways<sup>24,25</sup>. Further investigation is needed to assess potential risks to neighboring countries along migratory flyways and to inform cross-border surveillance strategies, which is out of the scope of this specific study.

**We have added this figure to the supplementary materials and highlighted**

the association between the WAE of functional groups and H5N1 outbreaks (Lines 325-329 on page 16): “Our study confirms that waterfowl pose a greater transmission risk than other waterbird groups do, as their activity entropy is more correlated with human, cattle, and poultry density (Supplementary Table 3), and strongly spatially associated with the past avian influenzas outbreak records (Supplementary Fig. 13), indicating closer interactions with susceptible populations in anthropogenic ecosystems.”



**Supplementary Fig. 13** Spatial distribution of dominant waterbird functional groups in relation to H5N1 outbreaks in U.S. cattle. (a) Functional group with the highest WAE among confirmed host. (b) Functional group with the highest WAE among all suspected host.

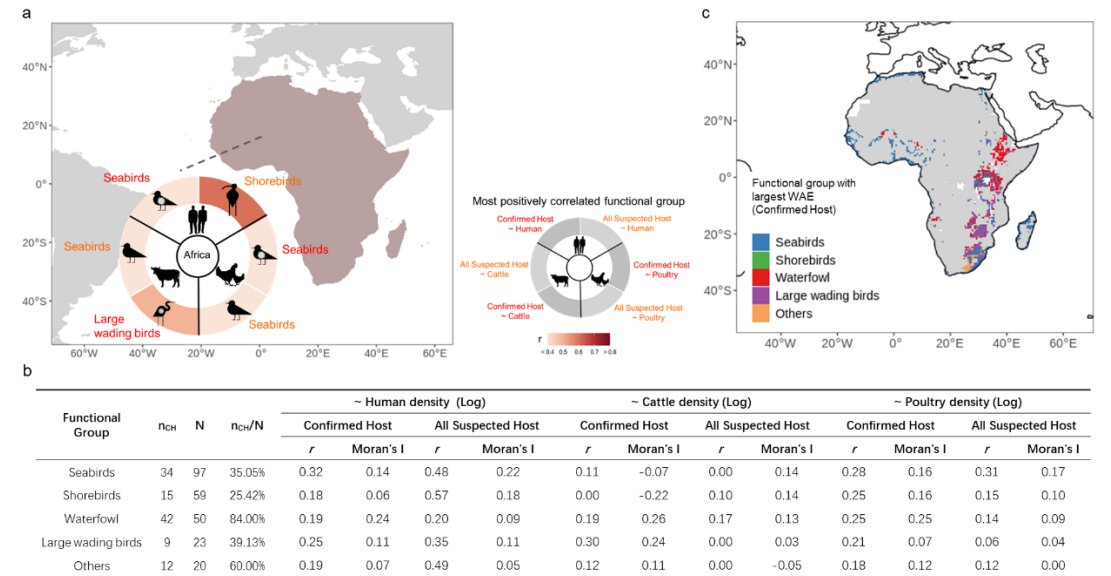
5. The author specifically mentioned surveillance issues in the discussion section, especially in African regions where WAE indicating higher risks. The weak surveillance capabilities of African countries are not only limited to avian influenza, but also pose significant challenges to Africa if surveillance is conducted simultaneously for both humans and birds. Therefore, based on the WAE, can more specialized suggestions or solutions be proposed for Africa, such as prioritizing the establishment or improvement of monitoring for specific species?

**Response:** Thank you for your valuable suggestion. As previously noted, while WAE effectively captures aggregated activity intensity across functional groups, it does not directly rank individual species by their relative roles in virus transmission. **To refine surveillance strategies in high-risk African regions, we added a functional group-level analysis of waterbird activity in the supplementary section (Supplementary Fig. 14) to provide more specific species group-targeted guidance for Africa.**

Among confirmed host groups (Supplementary Fig. 14a, b), seabirds showed the strongest positive correlation with both human ( $r = 0.32$ ) and poultry ( $r = 0.28$ ), while

large wading birds showed the highest correlation with cattle ( $r = 0.30$ ), suggesting both groups should be prioritized for surveillance in Africa. Regarding WAE-based hotspot patterns (Supplementary Fig. 14c), seabirds dominated southern countries/regions, large wading birds were prominent in central areas, and waterfowl were more active in northern Africa.

Based on these findings, we have added and proposed specialized suggestions for Africa in the discussion section (Lines 343-351 on page 17) that: “Although our analysis included all waterbird species and not just confirmed hosts, we found that large wading birds (mainly Ciconiiformes and Gruiformes) strongly overlapped with humans and cattle not only in four key exposure countries/regions (Fig. 4) but also in African countries (Supplementary Fig. 14). Africa’s surveillance system development should prioritize both seabirds and large wading birds <sup>26</sup>.”



**Supplementary Fig. 14 Analysis of waterbird functional groups in Africa.** (a) The waterbird functional groups most positively correlated with human density, cattle density, and poultry density in Africa. (b) Relationships between waterbird activity entropy (WAE) and human or poultry density among different waterbird functional groups in Africa. ‘n<sub>CH</sub>’ represents the number of confirmed hosts, and ‘N’ represents the number of all suspected waterbirds. ‘r’ represents the Pearson correlation coefficient, and ‘Moran’s I’ represents the Global Bivariate Moran’s I statistic. Waterbird functional groups with the highest water activity entropy (WAE) in Africa. (c) The identification is based on confirmed avian influenza virus-carrying hosts. Each pixel is identified as one of the following functional groups with the highest WAE: seabirds, shorebirds, waterbirds, large wading birds, and others.

6. Finally, it is suggested that the discussion should not repeat too much the content already presented in the results.

**Response:** Sorry for the previous redundancy between the discussion and results

sections. To avoid repetition between the results and discussion sections, we streamlined several parts of the text (Lines 226-227, 255-257, 265-267, 285-287, and 325-329), placing greater emphasis on interpretation, implications, and key limitations rather than restating findings.

Specifically, we streamlined (i) interpretive statements overlapping with predictive performance (Lines 226-227, 285-287), (ii) regional observations already covered in the results (Lines 265-267, 255-257), and (iii) host-specific risk comparisons (Lines 325-329).

## Response to Reviewer #3's Comments

This manuscript assesses the relevance of a potentially valuable metric, based on Shannon's entropy index, which estimates the spatial distribution of wild waterbird activity and proposes its use as an indicator of avian influenza (AIV) spillover risk. However, the conclusion that WAE predicts AIV case occurrence is weakened by the indiscriminate use of all outbreak data, leading to a likely conflation of initial introductions from wild birds with subsequent amplification events, particularly within poultry populations. With a more focused analytical approach and clearer positioning of the metric's intended application, WAE could become a useful tool for guiding targeted surveillance and mitigation strategies.

**Response:** We sincerely appreciate your insightful comments. We are grateful for your acknowledgment of WAE's potential value in evaluating avian influenza spillover risk, as well as your valuable recommendations concerning the discrimination and selection of outbreak data.

We have implemented a rigorous multi-criteria filtering protocol in our revised analysis to systematically exclude probable poultry-amplified cases. The specific methodological approach and corresponding updated results are presented in detail in our responses below. All line numbers mentioned in this response refer to our revised manuscript.

The predictive performance of WAE is evaluated using all reported AIV cases, regardless of whether these were infections in wild birds, direct spillovers into other species, or lateral spread among domestic poultry (or other species). This is problematic, as many outbreaks, especially in poultry, result from secondary transmission that is not linked to wild bird exposure. Furthermore, poultry-to-poultry transmission may occur over large geographical distances, well beyond the sites of initial viral incursion from wild birds. As the manuscript states that "over 81% of AIV-positive cases occurred in these high WAE areas", this statistic is potentially misleading if most cases are due to transmission occurring within domestic species. The use of all outbreak data conflates distinct epidemiological processes and may overestimate the utility of WAE as a predictor of true spillover risk.

**Response:** Thank you for your constructive comments. We agree that the inclusion of all outbreak data may conflate initial introductions from wild birds with secondary amplification events within poultry populations. Given that poultry-to-poultry transmission often originates from an initial avian spillover event, the reasonable data screening process might focus on retaining early-stage cases during seasonal poultry avian influenza outbreaks through specific criteria and methodologies. **To address this concern, referring to previous research <sup>27,28</sup>, in the revision we developed a multi-criteria filtering process to progressively exclude poultry-to-poultry transmission cases from the dataset.**

**Multi-criteria filtering process for restricting poultry-to-poultry transmission cases:**

To minimize the interference caused by secondary AIV infections within poultry flocks, we applied a systematic multi-step filtering process to exclude likely poultry-to-poultry transmission cases from the global surveillance dataset compiled by FAO (Fig. R6). The initial dataset included 56300 cases, of which 39374 involved poultry. Each step of the framework was designed to progressively eliminate recorded poultry cases that were probably transmitted within poultry flocks, or not directly infected from wild waterbird exposure.

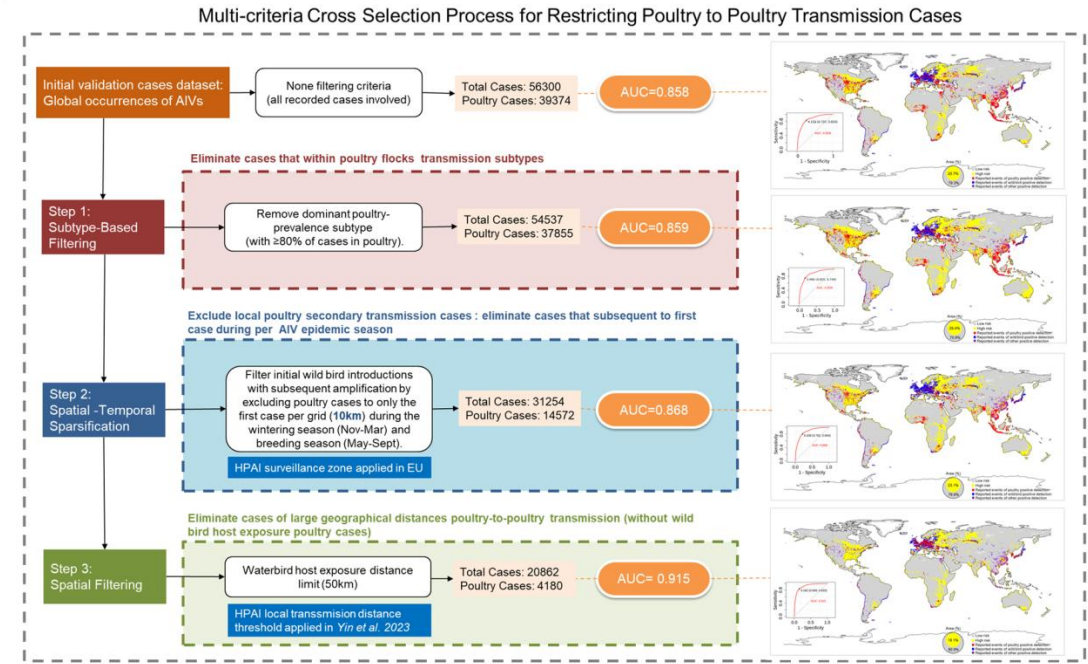
**Step 1, Exclusion of poultry-dominant AIV subtypes:** We excluded infection subtypes that primarily spread within poultry. Specifically, subtypes with  $\geq 80\%$  of cases reported in poultry were removed (1519 cases removed). This filter resulted in minimal change in the predictive performance of WAE (AUC = 0.859).

**Step 2, Spatio-temporal filter based on local transmission assumptions:** We filtered poultry-to-poultry transmission cases using a spatio-temporal criterion. Based on the EU's High Pathogenicity Avian Influenza (HPAI) Surveillance Zone<sup>27</sup>, we assumed a 10 km radius as the spatial range for local transmission within poultry populations. Accordingly, we segmented the global AIV surveillance dataset into 10×10 km grids during epidemic seasons, defined as November to March for wintering and May to September for breeding. Within each grid cell, only the earliest poultry case was retained; any subsequent poultry cases reported during the same season were considered intra-poultry transmissions and excluded. This filtering step removed 23284 poultry cases and substantially improved the predictive power of WAE (AUC = 0.868).

**Step 3, Wild waterbird encounter proximity filter:** We applied a spatial filtering criterion based on the local activity range of wild waterbird hosts (~50 km)<sup>28</sup>. We defined valid poultry cases that were infected with AIV from wild waterbird hosts through direct encounters or contacts, requiring at least one confirmed wild bird AIV detection within a 50 km radius of the poultry case. Under this criterion, 10392 poultry cases were excluded from the dataset. The resulting dataset comprised 20862 total cases, including 4,180 poultry cases. The predictive power of WAE continually increased (AUC = 0.915).

**Following the application of the above filtering procedures, the WAE framework continually demonstrated robust predictive performance in delineating high-risk areas of AIV exposure.** Although Step 3 performed well in regions with more established disease surveillance systems, such as Europe, where wild bird monitoring is relatively robust. This step resulted in the near-complete exclusion of poultry cases in Southeast Asia and China. We are concerned that this outcome reflects substantial over-exclusion of valid cases, primarily attributable to insufficient surveillance or limited public availability of wild bird AIV surveillance data in certain regions (e.g., Southeast Asia). For example, although China reported 6,329 AIV cases and India reported 404, the two adjacent countries exhibited great inconsistency in wild bird cases proportion—only 4% in China and 19% in India, and both were largely lower than European country data. Under the Step 3 criterion,

regions lacking sufficient wild bird AIV surveillance data would inevitably suffer over-deletion of poultry cases that may have genuinely resulted from wild bird exposure.

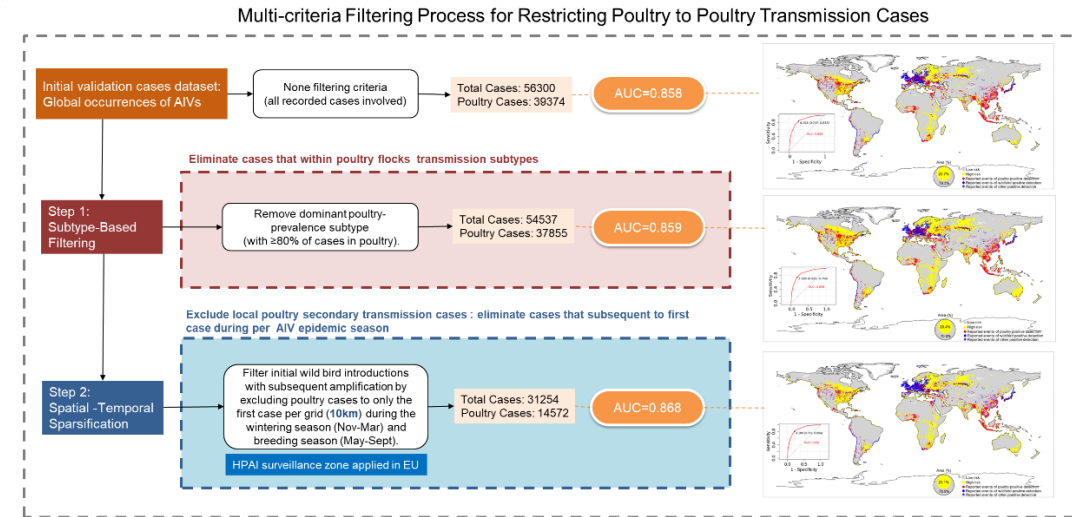


**Fig. R6 Multi-step Filtering Process for excluding Poultry-to-Poultry Transmission Cases.** Each panel displays the final output of one filtering step, highlighting the spatial distribution of reported avian influenza (AIV) detections. Red dots represent reported events of poultry positive detection, blue dots indicate wild bird positive detection, and purple dots mark other positive detections. Yellow-shaded areas denote high-risk areas, derived using the optimal WAE threshold identified via the Youden index at each corresponding step.

Despite its higher prediction accuracy ( $AUC = 0.915$ ), Step 3 led to the exclusion of valid case records. Therefore, we chose to retain Step 1 and Step 2 (Supplementary Fig. 7) as the final filtering criteria in the main analysis. While it is recognized that many poultry outbreaks, particularly the initial case of a poultry outbreak, commonly originate from such spillover events<sup>23</sup>, even if followed by secondary transmission within poultry populations. Therefore, we retained results based on both unfiltered and filtered poultry-to-poultry transmission cases in the Supplementary Materials, allowing readers to assess the WAE's predictive performance under different assumptions. The specific methodology and filtering steps are illustrated in Supplementary Fig. 7 and Supplementary Text 2: Statistical analysis.

**In light of your valuable suggestion, we have updated the main figures, corresponding results, methods, and supplementary figures accordingly.** Specifically, Fig. 2 now presents the revised results of risky areas after the multi-criteria filtering process for restricting poultry-to-poultry transmission cases. The WAE predictive performance slightly improved, with the AUC increasing from

0.86 to 0.87, which demonstrates the robustness of the WAE framework. In addition, Fig. 3 shows the updated hotspot distribution using the recalibrated WAE threshold of 4.108 (rounded to 4.11 in the text). This threshold adjustment led to minor refinements in the predicted high-risk areas: while the top 20 high-risk countries remained consistent, the relative extent of high-risk regions shifted, most notably, Turkey now ranks above Brazil in terms of predicted hotspot size. All analyses and visualizations related to risk and exposure zone (including Supplementary Figs. 8 to 12) have been revised accordingly to ensure consistency across the manuscript.



**Supplementary Fig. 7 Multi-step Filtering Process for Restricting Poultry-to-Poultry Transmission Cases.** Each panel displays the final output of one filtering step, highlighting the spatial distribution of reported avian influenza (AIV) detections. Red dots represent reported events of poultry positive detection, blue dots indicate wild bird positive detection, and purple dots mark other positive detections. Yellow-shaded areas denote high-risk areas, derived using the optimal WAE threshold identified via the Youden index at each corresponding step.

More broadly, the manuscript does not clearly distinguish whether WAE is intended as an indicator of AIV occurrence in wild birds or of spillover from wild birds into other species. This ambiguity weakens the conceptual framework and interpretation of the results.

**Response:** We appreciate your insightful comment. In this study, we constructed and applied the Waterbird Activity Exposure (WAE) to combine the indicators, including richness, duration, and evenness of dominant AIV natural host species, which emphasizes waterbirds as the original natural source for most AIV spillover instances. The human, poultry, and cattle populations are treated as the exposed susceptible population. Thus, WAE was constructed and designed to do more.

Thus, WAE was designed to more precisely assess the exposure risk of anthropogenic susceptible groups (including humans, poultry, and livestock) to wild

AIV host/carrier like waterbirds. For clarity, we added more precise expressions in the introduction, results and discussion sections.

**To avoid confusion, we have clarified the definition and application of WAE in the revised manuscript:**

In the introduction section (Lines 91-95 on page 5): “Here, we used waterbird activity entropy (WAE) to evaluate the richness, duration, and evenness of the dominant AIV natural host species. By integrating WAE with domestic animal and human densities, we infer contact possibility and estimate **their AIV exposure risk to waterbird hosts** at cross-interfaces (i.e., among waterbirds, domestic animals, and humans)<sup>29,30</sup>.”

In the results section (Lines 135-137 on page 7): “We evaluated the spatial predictive ability of WAE **for the waterbird-associated AIV case** distribution via filtered global surveillance data and a confusion matrix (Methods, Fig. 2a).”

In the discussion section (Lines 369-372 on page 18): “In this study, AIV surveillance data was used to assess the reliability and predictive capability of the **WAE in identifying AIV cases likely resulting from spillover events involving waterbirds**, and subsequently to determine the optimal threshold for delineating high-risk areas of waterbird host exposure (see Methods).”<sup>31</sup>

#### Minor Comments

##### Introduction

Several typographical errors are present throughout the introduction. Proofreading is recommended, particularly to clarify logical flow in some paragraphs.

**Response:** We are sorry for the typographical and structural issues in the original introduction. In response to your helpful suggestion, we carefully revised the relevant sentences to improve clarity and strengthen the logical flow. Specifically, we corrected errors and clarified phrasing in key paragraphs covering waterbird activity, spillover risks, and regional disparities in surveillance (Lines 54-55, 58-60, 61-64, 72-74, 79-81, 93-95, 96-98, 98-99, 101-102, and 110-112). These revisions enhance the accessibility and interpretability of the introductory section, and we appreciate your guidance in prompting this refinement.

Additional revisions were also made in line with your helpful suggestions, as follows.

Line 51: Consider replacing “reassortants” with “reassortments,” which may be more appropriate in this context.

**Response:** We have replaced “reassortants” with “reassortments”.

Line 53: Referring to AIV emergence as a “progressive” process may not be accurate; many reassortment and host adaptation events are likely punctuated.

**Response:** Following your comment, we have revised the wording to better reflect this pattern. Revised sentence: “These various and recurring reassortments, coupled with frequent interactions between wild waterbirds and domestic animals, facilitate the virus’s ability to cross between diverse host species <sup>32</sup>, including wild and domestic birds, mammals <sup>33-35</sup>, and occasionally humans <sup>19,36,37</sup>.”

Line 55 – 56: “H5N1 have broadened rapidly in the host range in waterbirds” this sentence contains a grammatical error.

**Response:** Thank you for your careful reading. In the revised manuscript, we refined the sentence to improve grammatical accuracy and clarity. The revised sentence now reads (Lines 50-54 on page 3): “For example, viruses of the highly pathogenic avian influenza (HPAI) subtype H5N1 have rapidly expanded their host range among waterbirds (especially in long-distance migrated seabird and shorebird species) in recent years, increasing the risk of cross-continental spreading of emerging AIVs <sup>38,39</sup>.”

Lines 60 – 73: The manuscript should more clearly distinguish between spillover from wild birds and subsequent lateral spread among farms.

**Response:** We agree that distinguishing between wild bird spillover and subsequent lateral spread among farms is important for clarity. Accordingly, we have revised the paragraph to better reflect this distinction. Specifically, we now state (Lines 64-68 on page 4): “Notably, phylogenomic analyses revealed that birds obtained from infected farms carried related viral sequences, suggesting future spillover risk could originate from a shared wild bird reservoir <sup>6</sup>. Concurrently, intensive poultry or cattle farms lacking adequate biosecurity measures may also facilitate subsequent lateral spread among farms and increase the risk of zoonotic spillover to humans <sup>40</sup>.”

Line 78: “thus enhancing the risky reassortants” this sentence may be missing a word.

**Response:** Right, we have revised the phrase to “thus enhancing the risk of reassortants” to ensure clarity and grammatical accuracy.

Line 93 – 94: “it has also been used to involve residency time to quantify the intensity of human activity” the phrasing needs correction.

**Response:** The corrected version has been modified to: “it has also been used to incorporate residency time as a metric to quantify the intensity of human activity intensity.”

Line 104 – 105: “global efforts against to AIV spreading” contains a typographical error.

**Response:** Thank you for your careful reading and the corrected version has been updated to “global efforts against AIV spread” for improved clarity and grammatical accuracy.

Methods

Line 188: Clarify the type of model used for the correlation analysis and specify whether spatial autocorrelation was accounted for.

**Response:** We have clarified that our correlation analysis was conducted using a Pearson correlation model to examine the relationship between the activity entropy of each functional group and human/cattle/poultry density. Additionally, to address spatial dependency, we incorporated global Bivariate Moran's I to evaluate spatial autocorrelation patterns between these variables. The following revisions were made:

In Fig. 4, we added global Bivariate Moran's I statistics alongside Pearson correlation results. The related sentence in results has been updated to (Lines 186-188 on page 9): "We conducted Pearson correlation and global Bivariate Moran's I analyses using human density, cattle density, and poultry density as predictors to identify waterbird groups that pose the highest exposure risk to susceptible populations." We also replaced  $R^2$  values with the Pearson correlation coefficient ( $r$ ) and Bivariate Moran's I statistics to more accurately characterize both linear associations and spatial dependency patterns (Lines 194-211 on page 10-11). Methods section has been updated to (Lines 667-672 on page 37): "We explored spatial relationship between human/cattle/poultry density and WAE of different waterbird functional groups using both Pearson correlation and global Bivariate Moran's I, including confirmed hosts and all suspected hosts (all waterbirds), in the key exposure countries/regions (Fig. 4, Supplementary Table 3, Supplementary Table 4)."

Please provide more detail on the cross-validation procedure used for model evaluation.

**Response:** Thank you for your suggestion. Here, we have expanded our explanation of the cross-validation procedure to clarify its role in model evaluation. Specifically, we employed fivefold cross-validation, where the dataset was divided into five equal subsets. Each fold was used as a validation set once, while the remaining four folds served as the training set. This process was repeated five times, ensuring that each subset was used for validation exactly once. During each iteration, we determined the optimal regularization parameter by evaluating model performance on the validation set, helping to mitigate overfitting while improving generalizability. After completing the cross-validation process, we selected the best-performing model for each machine learning method based on the averaged evaluation metrics across all five folds.

We include this detailed information in methods (Lines 582-584 on page 33): "Specifically, the dataset was partitioned into five equal subsets; each fold served once as the validation set, while the remaining four folds were used for training."

While code is available, it is important that sufficient documentation is included to ensure reproducibility. For instance, the described analytical workflow in the README file does not appear to match the structure of the scripts provided.

**Response:** We have improved the clarity and consistency of the provided documentation, ensuring that the described analytical workflow in the README file accurately reflects the structure of the scripts. The new version is:

“Code of WAE (Version 202507).R” ([birdsAIV/Code of WAE \(Version 202507\).R at master · bobilong/birdsAIV](#))

“README file” ([birdsAIV/README.md at master · bobilong/birdsAIV](#))

#### Discussion

Several paragraphs in the discussion section reiterate results without providing further interpretative insight. These sections could be streamlined to focus on implications and limitations.

**Response:** To avoid repetition between the results and discussion sections, we streamlined several parts of the text (Lines 226-227, 255-257, 265-267, 285-287, and 325-329), placing greater emphasis on interpretation, implications, and key limitations rather than restating findings.

Specifically, we streamlined (i) interpretive statements overlapping with predictive performance (Lines 226-227, 285-287), (ii) regional observations already covered in the results (Lines 265-267, 255-257), and (iii) host-specific risk comparisons (Lines 325-329).

Line 227 – 228: “Waterbirds have been verified as the main reservoirs of HPAI H5 and may act as dominant transmission vectors” Poultry have been the primary reservoir of H5 since its emergence in 1996.

**Response:** We appreciate your clarification and professional comments. While poultry has served as the primary amplification host of H5 since its emergence in 1996, the reservoir dynamics of highly pathogenic avian influenza (HPAI) H5 have become increasingly complex following the widespread circulation of H5N1 clade 2.3.4.4b since 2020. Notably, epidemiological evidence indicates that this particular viral lineage has now been maintained in wild birds without the need for incursion from poultry. In this study, waterbirds, owing to their high mobility and migratory behavior, are considered to play a major role in the long-distance transmission of HPAI H5.

To more accurately reflect current understanding, we have revised the phrasing to (Lines 228-230 on page 11): “Waterbirds have been verified as the natural main reservoirs of HPAI H5<sup>26,38,39,41</sup> and may act as dominant transmission vectors after the emergence of 2.3.4.4.b”.

Lines 245 – 249: The discussion on surveillance capacity in less developed countries is vague; this section would benefit from engagement with the extensive literature on surveillance disparities and data quality.

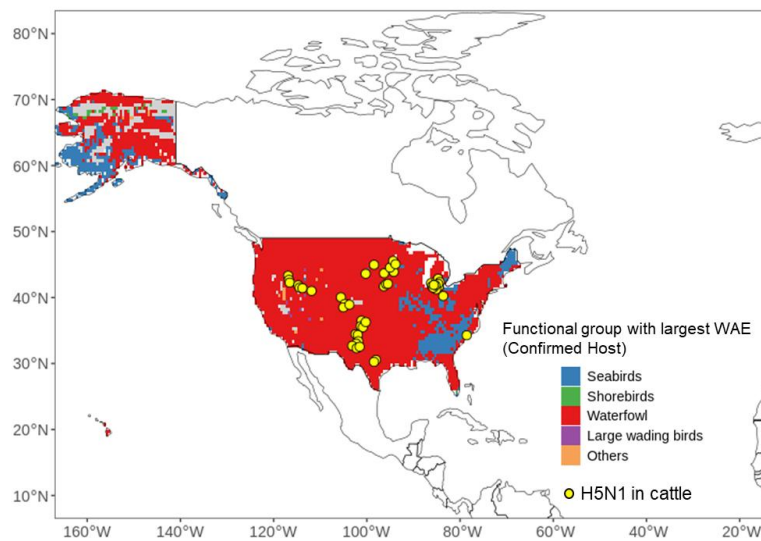
**Response:** We have revised this section to systematically integrate relevant literature. The key references about surveillance disparities and data quality in Africa that we have found include:

- 1) Rohr, J. R. et al. A planetary health innovation for disease, food and water challenges in Africa. *Nature* 619, 782-787, doi:10.1038/s41586-023-06313-z (2023).
- 2) Branda, F., Mohapatra, R. K., Tuglo, L. S., Ciccozzi, M. & Scarpa, F. J. B. R. N. Real-time epidemiological surveillance data: tracking the occurrences of avian influenza outbreaks around the world. 18, 95 (2025).
- 3) Abolnik, C. et al. The Molecular Epidemiology of Clade 2.3.4.4B H5N1 High Pathogenicity Avian Influenza in Southern Africa, 2021-2022. *Viruses-Basel* 15, doi:10.3390/v15061383 (2023).
- 4) Fusaro, A. et al. Disentangling the role of Africa in the global spread of H5 highly pathogenic avian influenza. *Nature Communications* 10, 5310, doi:10.1038/s41467-019-13287-y (2019).

The discussion section was supplemented with the above citations. The revised content and corresponding sentences are presented below (Lines 250-254 on page 12-13). “Therefore, increased funding and resources for enhanced surveillance in developing and less developed countries are crucial to improving surveillance data quality<sup>42,43</sup>, which would not only strengthen the prediction capacity of WAE but also help prevent HPAI from developing silently in under monitored hotspots<sup>26,44</sup>.”

Lines 325 – 328: “Our study confirms that waterfowl pose a greater transmission risk...” Demonstrating such a link may require exploring the association between the distribution of this specific species group and relevant outbreaks.

**Response:** Based on your suggestion, here we analyzed the spatial relationship between the WAE of waterfowl group and historical records of a regional avian influenza outbreak (H5N1 in dairy cattle in the United States, 2024 epidemic), which further confirmed the spatially consistent and strong association between waterfowl activity entropy and case records. Accordingly, we have revised the relevant expressions and restated the findings (Lines 325-329 on page 16): “Our study confirms that waterfowl pose a greater transmission risk than other waterbird groups do, as their activity entropy is more correlated with human, cattle, and poultry density (Supplementary Table 3), and strongly spatially associated with the past avian influenza outbreak records (Supplementary Fig. 13), indicating closer interactions with susceptible populations in anthropogenic ecosystems.”



**Supplementary Fig. 13** Spatial distribution of dominant waterbird functional groups in relation to H5N1 outbreaks in U.S. cattle. (a) Functional group with the highest WAE among confirmed host. (b) Functional group with the highest WAE among all suspected host.

## References:

- 1 Mremi, I. R. *et al.* Twenty years of integrated disease surveillance and response in Sub-Saharan Africa: challenges and opportunities for effective management of infectious disease epidemics. *One Health Outlook* **3**, 1-15 (2021).
- 2 Van Dijk, J. G. *et al.* Juveniles and migrants as drivers for seasonal epizootics of avian influenza virus. *J. Anim. Ecol.* **83**, 266-275 (2014).
- 3 Sun, L., Chen, J., Li, Q. & Huang, D. Dramatic uneven urbanization of large cities throughout the world in recent decades. *Nature Communications* **11**, 5366, doi:10.1038/s41467-020-19158-1 (2020).
- 4 Chyb, A., Jedlikowski, J., Włodarczyk, R. & Minias, P. Consistent choice of landscape urbanization level across the annual cycle in a migratory waterbird species. *Sci Rep* **11**, 836, doi:10.1038/s41598-020-80872-3 (2021).
- 5 Mellis, A. M. & Report, M. W. Serologic Evidence of Recent Infection with Highly Pathogenic Avian Influenza A (H5) Virus Among Dairy Workers—Michigan and Colorado, June–August 2024. *MMWR. Morbidity* **73** (2024).
- 6 Caserta, L. C. *et al.* Spillover of highly pathogenic avian influenza H5N1 virus to dairy cattle. *Nature*, doi:10.1038/s41586-024-07849-4 (2024).
- 7 Chandler, M. *et al.* Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation* **213**, 280-294, doi:10.1016/j.biocon.2016.09.004 (2017).
- 8 Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* **19**, 10-15, doi:10.1016/j.ecoinf.2013.11.002 (2014).
- 9 Steen, V. A., Tingley, M. W., Paton, P. W. & Elphick, C. S. Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution* **12**, 216-226 (2021).
- 10 Ingenloff, K. & Peterson, A. T. Incorporating time into the traditional correlational distributional modelling framework: A proof-of-concept using the Wood Thrush *Hylocichla mustelina*. *Methods in Ecology Evolution* **12**, 311-321 (2021).
- 11 Pequeno, D. N. L. *et al.* Production vulnerability to wheat blast disease under climate change. *Nature Climate Change* **14**, 178-183, doi:10.1038/s41558-023-01902-2 (2024).
- 12 Stanton, J. C., Pearson, R. G., Horning, N., Ersts, P. & Reşit Akçakaya, H. Combining static and dynamic variables in species distribution models under climate change. *Methods in Ecology Evolution* **3**, 349-357 (2012).
- 13 Ramos, S., MacLachlan, M. & Melton, A. Impacts of the 2014-2015 highly pathogenic avian influenza outbreak on the US poultry sector. *Washington, DC, USA: USDA* (2017).
- 14 Ingenloff, K. & Peterson, A. T. Incorporating time into the traditional correlational distributional modelling framework: A proof-of-concept using the Wood Thrush *Hylocichla mustelina*. *Methods in Ecology* **12**, 311-321 (2021).
- 15 Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk.

966 *Nature* **607**, 555-562, doi:10.1038/s41586-022-04788-w (2022).

967 16 Liu, X. *et al.* Animal invaders threaten protected areas worldwide. *Nature*  
968 *Communications* **11**, 2892, doi:10.1038/s41467-020-16719-2 (2020).

969 17 Crego, R. D., Stabach, J. A. & Connette, G. Implementation of species distribution  
970 models in Google Earth Engine. *Diversity and Distributions* **28**, 904-916 (2022).

971 18 Austin, M. P. & Van Niel, K. P. J. J. o. b. Vol. 38 1-8 (Wiley Online Library,  
972 2011).

973 19 Becker, D. J., Seifert, S. N. & Carlson, C. J. Beyond Infection: Integrating  
974 Competence into Reservoir Host Prediction. *Trends in Ecology & Evolution* **35**,  
975 1062-1065, doi:<https://doi.org/10.1016/j.tree.2020.08.014>. (2020).

976 20 Clark, B. L. *et al.* Global assessment of marine plastic exposure risk for oceanic birds.  
977 *Nature Communications* **14**, 3665, doi:10.1038/s41467-023-38900-z (2023).

978 21 Leguia, M. *et al.* Highly pathogenic avian influenza A (H5N1) in marine mammals  
979 and seabirds in Peru. *Nature Communications* **14**, 5489,  
980 doi:10.1038/s41467-023-41182-0 (2023).

981 22 Engelsma, M., Heutink, R., Harders, F., Germeraad, E. A. & Beerens, N. Multiple  
982 introductions of reassorted highly pathogenic avian influenza H5Nx viruses clade 2.3.  
983 4.4 b causing outbreaks in wild birds and poultry in The Netherlands, 2020-2021.  
984 *Microbiology Spectrum* **10**, e02499-02421 (2022).

985 23 Krammer, F. *et al.* Influenza. *Nature Reviews Disease Primers* **4**, 3,  
986 doi:10.1038/s41572-018-0002-y (2018).

987 24 Uyeki, T. M. *et al.* Highly Pathogenic Avian Influenza A(H5N1) Virus Infection in a  
988 Dairy Farm Worker. *N Engl J Med*, doi:10.1056/NEJMc2405371 (2024).

989 25 Abdelwhab, E. M. & Beer, M. Panzootic HPAIV H5 and risks to novel mammalian  
990 hosts. *npj Viruses* **2**, 22, doi:10.1038/s44298-024-00039-z (2024).

991 26 Fusaro, A. *et al.* Disentangling the role of Africa in the global spread of H5 highly  
992 pathogenic avian influenza. *Nature Communications* **10**, 5310,  
993 doi:10.1038/s41467-019-13287-y (2019).

994 27 Pittman, M. & Laddomada, A. Legislation for the control of avian influenza in the  
995 European Union. *Zoonoses and Public Health* **55**, 29-36 (2008).

996 28 Yin, S. *et al.* Functional traits explain waterbirds' host status, subtype richness, and  
997 community-level infection risk for avian influenza. *Ecology letters* **26**, 1780-1791,  
998 doi:10.1111/ele.14294 (2023).

999 29 Gibb, R. *et al.* Zoonotic host diversity increases in human-dominated ecosystems.  
1000 *Nature* **584**, 398-402, doi:10.1038/s41586-020-2562-8 (2020).

1001 30 Sun, B. *et al.* Urbanization affects spatial variation and species similarity of bird  
1002 diversity distribution. *SCIENCE ADVANCES* **8**, eade3061,  
1003 doi:doi:10.1126/sciadv.ade3061 (2022).

1004 31 (FAO), F. a. A. O. o. t. U. N. FAOSTAT Statistical Database. (2024).

1005 32 Bao, P. *et al.* Human infection with a reassortment avian influenza A H3N8 virus: an  
1006 epidemiological investigation study. *Nature Communications*.

1007 33 Murawski, A. *et al.* Highly pathogenic avian influenza A(H5N1) virus in a common  
1008 bottlenose dolphin (*Tursiops truncatus*) in Florida. *Commun Biol* **7**, 476,  
1009 doi:10.1038/s42003-024-06173-x (2024).

1010 34 Leguia, M. *et al.* Highly pathogenic avian influenza A (H5N1) in marine mammals  
1011 and seabirds in Peru. *Nat Commun* **14**, 5489, doi:10.1038/s41467-023-41182-0  
1012 (2023).

1013 35 Kim, Y. *et al.* Lessons for cross-species viral transmission surveillance from highly  
1014 pathogenic avian influenza Korean cat shelter outbreaks. *Nature Communications* **14**,  
1015 6958, doi:10.1038/s41467-023-42738-w (2023).

1016 36 Taylor, L. H., Latham, S. M. & Woolhouse, M. E. J. Risk factors for human disease  
1017 emergence. *Philosophical Transactions of the Royal Society of London* **356**, 983-989  
1018 (2001).

1019 37 Plowright, R. K., Parrish, C. R. & McCallum, H., Peter J.Ko, Albert I.Graham,  
1020 Andrea L.Lloyd-Smith, James O. Pathways to zoonotic spillover. *Limnology and*  
1021 *oceanography, methods* **15**, 502-510 (2017).

1022 38 Yin, S. *et al.* Functional traits explain waterbirds' host status, subtype richness, and  
1023 community - level infection risk for avian influenza. *Ecology Letters*,  
1024 doi:10.1111/ele.14294 (2023).

1025 39 Xie, R. *et al.* The episodic resurgence of highly pathogenic avian influenza H5 virus.  
1026 *Nature* **622**, 810-817, doi:10.1038/s41586-023-06631-2 (2023).

1027 40 Willyard, C. Flu on the farm. *Nature* **573**, S62-S63 (2019).

1028 41 Yang, Q. *et al.* Synchrony of bird migration with global dispersal of avian influenza  
1029 reveals exposed bird orders. *Nature Communications* **15**, 1126,  
1030 doi:10.1038/s41467-024-45462-1 (2024).

1031 42 Branda, F., Mohapatra, R. K., Tuglo, L. S., Ciccozzi, M. & Scarpa, F. J. B. R. N.  
1032 Real-time epidemiological surveillance data: tracking the occurrences of avian  
1033 influenza outbreaks around the world. **18**, 95 (2025).

1034 43 Rohr, J. R. *et al.* A planetary health innovation for disease, food and water challenges  
1035 in Africa. *Nature* **619**, 782-787, doi:10.1038/s41586-023-06313-z (2023).

1036 44 Abolnik, C. *et al.* The Molecular Epidemiology of Clade 2.3.4.4B H5N1 High  
1037 Pathogenicity Avian Influenza in Southern Africa, 2021-2022. *Viruses-Basel* **15**,  
1038 doi:10.3390/v15061383 (2023).

1039

## Response to Reviewer #1's Comments

Reviewer #1 (Remarks to the Author):

The authors provided a detailed and clear response with corresponding edits to all concerns I have raised. The ability to accurately conduct a study at this scale remains challenging but this work represents a significant advancement on previous approaches.

**Response:** We sincerely appreciate your thoughtful comments and encouraging remarks. Thank you for your constructive feedback and valuable suggestions again.

## Response to Reviewer #2's Comments

Reviewer #2 (Remarks to the Author):

The author has already responded to and revised all of my questions/comments/suggests in the revision.

**Response:** Thank you for your positive conclusion. We sincerely appreciate your careful review and are pleased that our revisions have successfully addressed your questions, comments, and suggestions.

## Response to Reviewer #3's Comments

Comment #1. I thank the authors for their thoughtful response and the substantial efforts made to revise their manuscript. The improvements are appreciated. However, several core concerns remain insufficiently addressed, particularly regarding the interpretation of outbreak data in relation to wild bird spillover, and the evaluation of the predictive performance of the indicator. These issues continue to affect the robustness and interpretability of the proposed indicator.

**Response:** We sincerely appreciate your constructive suggestions and thoughtful comments. We agree that a reliable evaluation of the predictive performance of the WAE framework is essential.

Regarding the interpretation of outbreak case data in relation to wild bird spillover, as you rightly pointed out, we acknowledge the discrepancies between the recorded outbreak case data in global public databases (FAO, <https://empres-i.apps.fao.org/>) and actual spillover events. In the previous revision, we introduced a Multi-criteria Filtering Process, trying to screen out poultry-to-poultry transmission cases when assessing spillover risk. However, given the global scope of this study, there is considerable variation in the completeness of surveillance systems across regions, compounded by limitations in the data available from global public databases. Although we made every effort to filter out the poultry-to-poultry transmission cases using available spatiotemporal attributes of recorded cases, it remains challenging to fully and accurately eliminate all such cases.

To further enhance the reliability of our approach, **we have added assessment on the effectiveness of our screening process and the sensitivity of excluding poultry-to-poultry transmission cases through a series of comparative assessments. Furthermore, we have conducted more systematic validations across regions with a series of poultry density gradients.** We hope you will agree that despite the remaining limitations in spillover risk precision based on these efforts, our approach nonetheless provides new and valuable insights into the exposure risks faced by various susceptible groups (including poultry, livestock, and humans) in relation to waterbird activities.

The results from newly added comprehensive assessment are provided in the Results section (Lines 137-146 on page 7). Additionally, to offer a comprehensive explanation of the data limitations and potential directions for future improvement, we have explicitly addressed the relevant constraints and included a discussion on the possibility that some filtered cases may still represent partial poultry-to-poultry transmission events rather than true spillover events (Lines 234-238 on page 12 and Lines 407-414 on page 20). These detailed supplementary analyses and discussions are presented in the following responses.

Comment #2. A central concern remains the conflation of wild bird-to-poultry

transmission events with poultry-to-poultry transmission. In the current analytical pipeline, Step 2 of the outbreak filtering process aims to retain all reported outbreaks not known to result from lateral spread. However, this approach relies heavily on the quality of the underlying surveillance data and on assumptions regarding reporting completeness. I remain concerned that this step retains a large number of outbreaks caused by secondary transmission among poultry. This is particularly problematic in many low- and middle-income countries, where surveillance systems are incomplete and highly pathogenic avian influenza is endemic in poultry populations. In such contexts, reported outbreaks often do not reflect the true epidemiological origin of infection, and lateral spread is frequently under-ascertained. As such, the authors' argument regarding spatial consistency in the proportion of outbreaks attributable to wild bird introductions is not tenable. It highlights the limitations of using globally aggregated outbreak data without adequately addressing spatial and temporal heterogeneity in surveillance sensitivity. If the predictive performance of the metric is evaluated using outcomes that likely include a high proportion of within-poultry transmission, it cannot be interpreted as an indicator of spillover risk. An illustration of this issue is the large number of H7N9 outbreaks retained in the analysis (Fig. S8): to my knowledge, both epidemiological and phylogenetic evidence suggest that this virus is maintained within poultry populations and not in wild birds. It is therefore likely that the majority, if not all, of these outbreaks retained in this analysis reflect lateral spread in poultry, rather than wild bird-to-poultry transmission.

**Response:** We sincerely appreciate your detailed explanation regarding the interpretation of outbreak records and the limitations of surveillance data. As you pointed out, distinguishing wild bird-to-poultry spillover events from subsequent poultry-to-poultry transmission is particularly challenging, especially when relying solely on global-scale public databases. The epidemiological signatures of these transmission routes often overlap, making precise differentiation challenging.

Given the global scope of this study, there is considerable variation in the completeness of surveillance systems across regions, as well as limitations in the data available from global public databases. Although we made every effort to screen poultry-to-poultry transmission cases using spatiotemporal attributes, we acknowledge that it remains difficult to fully and accurately eliminate all such cases. To better evaluate the effectiveness of our screening process and the sensitivity of excluding poultry-to-poultry transmission cases through the introduced Multi-criteria Filtering Process, we added more analyses as stated from the following three aspects:

### **1. Sensitivity analysis of the Multi-criteria Filtering Process based on the different proportion of poultry-to-poultry transmission cases in the AIV transmission process, classified by AIV subtypes.**

Although an avian influenza outbreak or epidemic process entirely free of lateral transmission is rare, global public databases (FAO, <https://empres-i.apps.fao.org/>) lack detailed genealogical information on recorded AIV events, making it difficult to precisely isolate those events that are credible and do not result from lateral spread.

However, we can categorize the AIV subtypes recorded in the database into **poultry-adapted subtypes** and those primarily circulating in wild waterbirds with **sporadic spillover into poultry**. This categorization enables the identification and extraction of a smaller proportion of poultry-to-poultry transmission cases with specific subtypes of AIV case records. Based on the AIV subtype, its recorded case-type composition (poultry case proportion or ratio of poultry and wild bird cases), and the verification with expert experience, we selected and classified the AIV subtypes into two representative groups. One group comprises **poultry-adapted subtypes**, including H7N6, H7N8, H7N9, H7N3, H5N2, H9N2, and H5N9, which presented a larger proportion of poultry cases and likely a larger proportion of poultry-to-poultry transmission among poultry cases. The other group is **wild-bird-host-dominant subtypes**, which presented a smaller proportion of poultry cases and likely a larger proportion of wild bird-to-poultry transmission among their poultry cases.

In the previous revision, we developed a spatial-temporal filter protocol to remove the poultry cases that are more likely to result from poultry-to-poultry transmission. As step 2 described, we filtered poultry-to-poultry transmission cases using a spatial-temporal criterion. Based on the EU's High Pathogenic Avian Influenza (HPAI) Surveillance Zone <sup>1</sup>, we assumed a 10 km radius as the spatial range for local transmission within poultry populations. Accordingly, we segmented the global AIV surveillance dataset into 10×10 km grids during epidemic seasons, defined as November to March for wintering season and May to September for breeding season. Within each grid cell, only the earliest poultry case was retained; any subsequent poultry cases reported during the same season were considered poultry-to-poultry transmissions and excluded.

As indicated by the following sensitivity analysis (Table R1), due to the limited data attributes (mainly spatial and temporal characteristics of recorded cases) of global public databases (FAO, <https://empres-i.apps.fao.org/>), it is difficult to precisely remove all the poultry-to-poultry transmission cases. **Step 2 of our Multi-criteria Filtering Process still demonstrated a strong ability to screen out poultry-to-poultry transmission.** This is supported by the data presented in Table R1, which shows a significantly higher mean poultry case reduction rate (75.7%) for AIV poultry-adapted subtypes after the filtering process. In contrast, wild bird-host-dominant subtypes, which should contain fewer poultry-to-poultry transmission events among their poultry cases, exhibited a much lower mean poultry case reduction rate (41.8%). **Notably, for H7N9 outbreaks, 91.9% of poultry cases were removed.** The remaining cases were strictly spatially adjacent to wild bird host records, suggesting that complete elimination could lead to the incorrect removal of actual spillover events.

The detailed analysis procedure is as follows: To evaluate the effectiveness of our first two-step filtering process (Supplementary Fig. 7), we conducted an analysis exploring classifications based on AIV subtypes that contain varying proportions of poultry-to-poultry transmission cases in their poultry cases. The results (Table R1) showed that poultry-adapted subtypes (with a higher poultry case proportion) exhibited substantial reductions in poultry cases after filtering, with mean reduction rates of

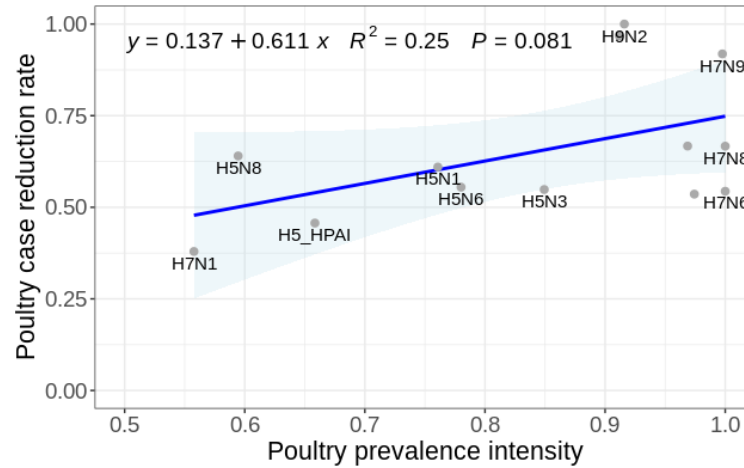
143 75.7%, which is much higher than wild bird-host-dominant subtypes (41.8%). This  
144 suggests a larger number and higher proportion of poultry cases in poultry-adapted  
145 subtypes, which are more likely to result from poultry-to-poultry transmission, were  
146 effectively excluded through our multi-criteria filtering process. In contrast, wild bird-  
147 dominant subtypes showed relatively lower reduction rates, which suggests the filtering  
148 process reasonably retained the poultry cases of wild bird-host-dominant subtypes,  
149 since a larger proportion of their poultry cases are more likely attributed to wild bird-  
150 to-poultry transmission. Furthermore, linear regression analysis across AIV subtypes  
151 (Fig. R1) revealed a positive correlation ( $P < 0.1$ ) between the prevalence level in  
152 poultry (reflected by poultry case proportion) and the subsequent poultry case reduction  
153 rate after filtering. These results confirm the effectiveness and sensitivity of our filtering  
154 workflow, supporting the conclusion that the first two-step filter process can effectively  
155 target and remove poultry cases that are more likely to result from poultry-to-poultry  
156 transmission, and retain those wild bird-to-poultry transmission cases.

157 In addition to case reduction rates, we compared the predictive performance of the  
158 WAE framework across subtypes using three key metrics (Table R1): AUC, accuracy,  
159 and sensitivity. For poultry-adapted subtypes after 2-step filtering, AUC values  
160 remained consistently high (all  $AUC \geq 0.79$ ) and even improved in some subtypes (e.g.,  
161 H5N9: 0.82 raw, 0.89 filtered). At the same time, the accuracy and sensitivity of  
162 poultry-adapted subtypes also remained above 80% after filtering. For all wild-bird-  
163 host-dominant subtypes, they exhibited improved AUCs after filtering the poultry-to-  
164 poultry transmission cases, as well as sensitivity improvements (e.g., H7N1: 93% raw,  
165 95% filtered). By removing poultry-to-poultry transmission cases, the applied filtering  
166 process tends to improve the predictive performance of the WAE framework. **These**  
167 **results indicate strong reliability and robustness of WAE in predicting AIV cases**  
168 **when excluding a large number of poultry-to-poultry transmission cases,**  
169 **supporting the applicability and utility of our two-step filter process in the risk**  
170 **assessment workflow of the WAE framework.**

171 In summary, as illustrated by the results (Table R1), our filtering method  
172 effectively filters out a substantial proportion of poultry-to-poultry transmission  
173 cases. While this approach is constrained by data availability, it provides a practical  
174 way to reduce the inclusion of poultry-to-poultry transmission cases in our analysis.  
175 This contributes to a more refined validation dataset and increases the relative  
176 representation of wild bird-to-poultry spillover events. **The improvement in AUC also**  
177 **suggests that excluding cases from the later stages of the epidemic season helps**  
178 **remove infections likely resulting from local poultry-to-poultry transmission**  
179 **subsequent to earlier spillover events. The remaining early-season poultry**  
180 **transmission cases, which are more difficult to exclude, tend to be geographically**  
181 **closer to the initial spillover events.** Thus, those retained poultry-to-poultry  
182 transmission cases that occurred in the early stages of the epidemic season are more  
183 likely to preserve the spatial characteristics associated with earlier wild bird-to-poultry  
184 spillover events.

**Table R1 Subtype-specific performance of WAE in predicting poultry AIV cases under alternative transmission classifications.** This table presents an analysis evaluating the predictive performance of waterbird activity entropy (WAE) across AIV subtypes under two transmission classification scenarios: the raw dataset and a filtered dataset after two-step filtering. Subtypes with poultry-to-(poultry + wild bird) ratios >90% are classified as poultry-dominant; those <70% as wild bird-dominant. For each subtype, total and poultry case counts, poultry case proportions, and poultry-to- (poultry + wild bird) case proportions are reported. Total cases include poultry, wild birds, wild and domestic mammals, human infections and environmental samples. Predictive metrics including AUC, accuracy, and sensitivity are shown for both datasets. The final two columns indicate the reduction rate in poultry cases following two-step filtering.

| Classification                                                                                                | Subtype   | Raw Dataset |               |                         |                                          |      |          |             | Filtered Dataset |               |                         |                                         |      |          |             | poultry case reduction rate after filter process | Mean poultry case reduction rate |
|---------------------------------------------------------------------------------------------------------------|-----------|-------------|---------------|-------------------------|------------------------------------------|------|----------|-------------|------------------|---------------|-------------------------|-----------------------------------------|------|----------|-------------|--------------------------------------------------|----------------------------------|
|                                                                                                               |           | Total cases | Poultry cases | Poultry case proportion | Poultry cases/(poultry+ wild bird cases) | AUC  | Accuracy | Sensitivity | Total cases      | Poultry cases | Poultry case proportion | Poultry cases/(poultry+ wild bird case) | AUC  | Accuracy | Sensitivity |                                                  |                                  |
| Poultry adapted subtypes (higher proportion of poultry-to-poultry transmission among poultry cases)           | H7N6      | 105         | 103           | 98.1%                   | 100.0%                                   | 0.89 | 0.89     | 0.88        | 49               | 47            | 95.9%                   | 100.0%                                  | 0.87 | 0.86     | 0.80        | 54.4%                                            | 75.7%                            |
|                                                                                                               | H7N8      | 18          | 18            | 100.0%                  | 100.0%                                   | 0.92 | 0.89     | 0.89        | 6                | 6             | 100.0%                  | 100.0%                                  | 1.00 | 1.00     | 1.00        | 66.7%                                            |                                  |
|                                                                                                               | H7N9      | 2324        | 847           | 36.4%                   | 99.8%                                    | 0.90 | 0.84     | 0.94        | 1546             | 69            | 4.5%                    | 97.2%                                   | 0.89 | 0.82     | 0.84        | 91.9%                                            |                                  |
|                                                                                                               | H5N2      | 1876        | 1793          | 95.6%                   | 96.9%                                    | 0.92 | 0.84     | 0.77        | 680              | 597           | 87.8%                   | 91.1%                                   | 0.87 | 0.80     | 0.74        | 66.7%                                            |                                  |
|                                                                                                               | H9N2      | 1724        | 1483          | 86.0%                   | 91.6%                                    | 0.79 | 0.73     | 0.78        | 0                | 0             | \                       | \                                       | \    | \        | \           | 100.0%                                           |                                  |
|                                                                                                               | H5N9      | 36          | 31            | 86.1%                   | 91.2%                                    | 0.82 | 0.86     | 0.75        | 3                | 1             | 33.3%                   | 100.0%                                  | 0.89 | 0.83     | 1.00        | 96.8%                                            |                                  |
| Wild bird host dominant subtypes (higher proportion of wild bird-to-poultry transmission among poultry cases) | H5(HP AI) | 1223        | 742           | 60.7%                   | 65.8%                                    | 0.77 | 0.72     | 0.69        | 884              | 403           | 45.6%                   | 51.1%                                   | 0.79 | 0.73     | 0.70        | 45.7%                                            | 41.8%                            |
|                                                                                                               | H7N1      | 57          | 29            | 50.9%                   | 55.8%                                    | 0.53 | 0.65     | 0.93        | 46               | 18            | 39.1%                   | 43.9%                                   | 0.56 | 0.74     | 0.95        | 37.9%                                            |                                  |



**Fig. R1 Relationship of poultry prevalence intensity and poultry case reduction rate across AIV subtypes after the 2-step filter process.** Linear regression analysis was performed only on verified and sufficiently sampled subtypes (total cases > 15); Poultry prevalence intensity is defined as the proportion of the initial poultry case relative to the combined total of poultry and wild bird cases; The scatter plot includes subtype-specific labels for key strains to highlight notable patterns. The regression line is shown with a blue shaded area representing the 95% confidence interval. The coefficient of determination ( $R^2$ ) and p-value ( $P$ ) are displayed to indicate the strength and significance of the association.

## 2. Sensitivity analysis of the long-distance lateral transmission filter step (Step 3) of the Multi-criteria Filtering Process across countries/regions with different proportions of poultry cases.

In the Multi-criteria Filtering Process proposed in the previous revision, we also developed a spatial-temporal filter protocol “Step 3”, which aimed to remove the poultry cases that are more likely to involve long-distance lateral transmission within poultry populations. As step 3 described, based on the encounter probability between recorded poultry cases and adjacent wild bird host, we applied a spatial filtering criterion reflecting the local activity range of wild bird hosts ( $\sim 50$  km)<sup>2</sup>. We defined valid poultry cases that were infected with AIV from wild bird hosts through potential direct encounters or contacts, requiring at least one confirmed wild bird AIV detection within a 50 km radius of the poultry case. However, although this step substantially improved the global predictive performance of the WAE framework (AUC increase to 0.92), we ultimately had to shelve this filtering step, due to distortions of wild bird host surveillance across countries/regions. The specific rationale for this decision is analyzed as follows.

To more specifically explain the reason for not applying Step 3 of the multi-criteria filtering process to exclude long-distance lateral transmission for poultry cases in the last version of our manuscript, we conducted a comparative analysis across countries with varying wild bird surveillance quality. For the reliability of the analysis, countries with more than 300 total cases were retained (Table R2), resulting in a subset of AIV

surveillance data from 29 countries. Among these, only 8 countries had wild bird cases accounting for over 50% of total records, suggesting that wild bird surveillance is well-established in only a few regions/countries. For detailed assessment of WAE performance before and after the effectiveness and sensitivity of Step 3 in filtering long-distance lateral transmission cases within poultry, we selected and compared the case composition and predictive performance of the WAE framework before and after the Step 3 filtering between **countries with robust wild bird host surveillance** (e.g., United Kingdom and Germany) and **those with inadequate wild bird host surveillance** (e.g., Vietnam, Thailand, Indonesia and Philippines).

In the United Kingdom (Fig. R2a), with more than 1,000 wild bird cases, approximately 41% of poultry cases were filtered out after the Step 3 filtering process, resulting in an AUC improvement from 0.88 to 0.90 (s.d. = 0.001), demonstrating the excellent effectiveness of Step 3 filtering. A similar improvement was observed in Germany (Fig. R2b), where the AUC increased from 0.734 to 0.753. However, in countries with poor wild bird surveillance, such as Vietnam and Indonesia (Fig. R3a, b), wild bird cases accounted for less than 0.3% of total cases, with over 98% being poultry cases. Applying the 3-step filter in these countries led to the removal of nearly all poultry cases (98%) and resulted in a decline in AUC. The situation was even more extreme in the Philippines (Fig. R3c, Table R2), where no cases remained after 3-step filtering. Although the 3-step filtering occasionally improves AUC in regions with poor wild bird surveillance (e.g., Thailand, Fig. 4d), its reliability is questionable. This improvement stems from an excessive reduction in poultry cases (98% in Thailand, Fig. R3d), where the few remaining cases coincidentally fall within high-risk areas, artificially inflating the AUC.

Step 3 establishes an indirect verification mechanism for wild bird-to-poultry transmission events by applying a filtering criterion that requires the presence of at least one wild bird case within a 50 km radius of a poultry case. While this step effectively identifies transmission events with clear evidence of potential wild bird host contact, its stringent spatial criteria introduce systematic bias in regions with limited wild bird surveillance. In developing countries such as Vietnam and the Philippines, an insufficient density of wild bird surveillance networks may lead to the erroneous exclusion of genuine spillover cases due to the absence of recorded wild bird cases within a 50 km radius. This selection bias, rooted in disparities in surveillance completeness, results in a filtering process (Step 3) that is only applicable to well-monitored regions and fails to accurately represent transmission patterns on a global scale.

Therefore, while the strict 3-step filtering method proves effective for countries or regions with comprehensive surveillance systems (e.g., the United Kingdom, Germany, and other Western European countries), its application in less-developed regions is challenged by limited wild bird surveillance, which may lead to inaccurate filtering due to data constraints. Over-screening potential spillover cases in such contexts highlights the need to strengthen surveillance efforts in high-risk but under-resourced countries. On a global scale, in the assessment presented in our main manuscript, and considering

the widespread data limitations across developing and undeveloped countries, we ultimately adopted the 2-step filtering process that focuses more on removing the local poultry-to-poultry transmission cases occurring after initial spillover events. **Accordingly, due to the imbalance in wild bird host surveillance records, omitting the long-distance lateral filtering of Step 3 in the global assessment helps prevent the mistaken exclusion of poultry spillover cases in developing and undeveloped countries with insufficient wild bird surveillance.**

Due to the inevitable co-existence of spillover cases and poultry-to-poultry transmission cases in the spatiotemporally filtered validation datasets, we revised the manuscript to emphasize that **our approach was aimed at assessing the exposure risk posed by waterbird activity, which could potentially lead to spillover events and subsequent regional epidemics in multiple host groups (humans, cattle, and poultry) rather than directly predicting spillover events. We have refined corresponding sentences (Lines 579-581 on page 32):** “To analyze exposure risk posed by waterbird activity among anthropogenic ecosystems, we use human population density, poultry (chicken and duck) and cattle density to conduct the risk analysis.” **(Lines 641-644 on page 35):** “To assess exposure risk relevant to potential spillover events sourced from waterbird activity, we developed a multi-criteria filtering process for restricting poultry-to-poultry transmission cases from the original poultry cases data (Supplementary Fig. 7).”

**Furthermore, to provide a comprehensive explanation of the data limitations and potential avenues for future improvement, we have explicitly addressed the relevant constraints and included a discussion on the possibility that some filtered cases may still represent lateral spread events rather than true spillover events (Lines 407-414 on page 20):** “In addition, our 2-step filtering process was designed to exclude poultry-to-poultry transmission occurring after initial spillover from wild to domestic hosts. It should be noted that even after spatio-temporal filtering, the remaining AIV cases may still include instances of lateral spread cases within poultry populations, rather than solely representing true spillover events from wild waterbirds. This ambiguity could influence the interpretation of exposure risks and the calibration of predictive models, particularly in regions with limited surveillance data or imbalanced surveillance efforts between wild and domestic animal populations. ”

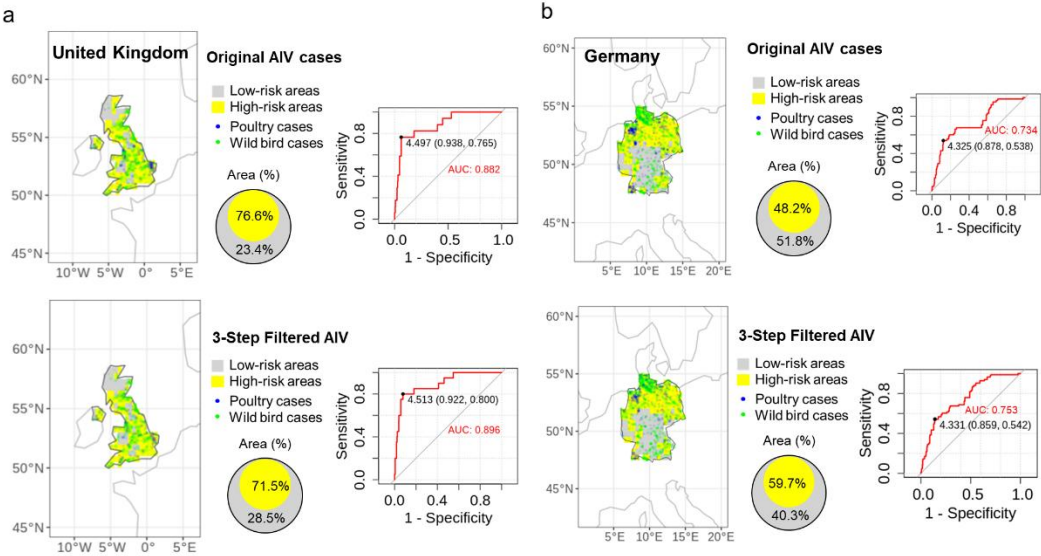
**Table R2 Country-level classification of wild bird surveillance adequacy based on AIV case composition.** This table presents a comparative summary of avian influenza virus (AIV) case statistics across countries, with at least 300 AIV reported cases. For each country, total AIV case counts, poultry and wild bird case counts, and wild bird case proportions are reported. Total cases include poultry, wild birds, wild and domestic mammals, human infections, and environmental samples. This table highlights disparities in surveillance efforts and attribution bias, which may influence risk assessment.

| Country               | Raw Dataset |               |                 |                           | Filtered Dataset (3-step filtered) |               |                 | Poultry case reduction rate after filter process Step 3 |
|-----------------------|-------------|---------------|-----------------|---------------------------|------------------------------------|---------------|-----------------|---------------------------------------------------------|
|                       | Total cases | Poultry cases | Wild bird cases | Wild bird case proportion | Total cases                        | Poultry cases | Wild bird cases |                                                         |
| Sweden                | 345         | 37            | 305             | 88.4%                     | 333                                | 25            | 305             | 32.4%                                                   |
| Denmark               | 626         | 76            | 550             | 87.9%                     | 597                                | 47            | 550             | 38.2%                                                   |
| Netherlands           | 1025        | 130           | 895             | 87.3%                     | 982                                | 87            | 895             | 33.1%                                                   |
| <b>United Kingdom</b> | <b>1406</b> | <b>266</b>    | <b>1126</b>     | <b>80.1%</b>              | <b>1297</b>                        | <b>157</b>    | <b>1126</b>     | <b>41.0%</b>                                            |
| <b>Germany</b>        | <b>3229</b> | <b>683</b>    | <b>2526</b>     | <b>78.2%</b>              | <b>2896</b>                        | <b>350</b>    | <b>2526</b>     | <b>48.8%</b>                                            |
| Belgium               | 440         | 73            | 343             | 78.0%                     | 418                                | 51            | 343             | 30.1%                                                   |
| Japan                 | 1108        | 275           | 784             | 70.8%                     | 948                                | 120           | 779             | 56.4%                                                   |
| Russian Federation    | 849         | 450           | 395             | 46.5%                     | 511                                | 112           | 395             | 75.1%                                                   |
| Poland                | 723         | 396           | 325             | 45.0%                     | 473                                | 146           | 325             | 63.1%                                                   |
| Czech Republic        | 333         | 188           | 145             | 43.5%                     | 240                                | 95            | 145             | 49.5%                                                   |
| Canada                | 1043        | 552           | 436             | 41.8%                     | 549                                | 59            | 435             | 89.3%                                                   |
| Italy                 | 763         | 463           | 293             | 38.4%                     | 425                                | 125           | 293             | 73.0%                                                   |
| Romania               | 540         | 344           | 194             | 35.9%                     | 283                                | 87            | 194             | 74.7%                                                   |
| South Africa          | 950         | 677           | 267             | 28.1%                     | 484                                | 211           | 267             | 68.8%                                                   |
| Peru                  | 392         | 301           | 87              | 22.2%                     | 154                                | 63            | 87              | 79.1%                                                   |

|                    |             |             |           |             |            |           |          |               |
|--------------------|-------------|-------------|-----------|-------------|------------|-----------|----------|---------------|
| France             | 2800        | 2153        | 607       | 21.7%       | 1116       | 469       | 607      | 78.2%         |
| India              | 426         | 347         | 77        | 18.1%       | 102        | 25        | 77       | 92.8%         |
| Hungary            | 1415        | 1191        | 224       | 15.8%       | 432        | 208       | 224      | 82.5%         |
| Republic of Korea  | 1678        | 1265        | 262       | 15.6%       | 846        | 435       | 262      | 65.6%         |
| USA                | 2258        | 1197        | 281       | 12.4%       | 1091       | 57        | 254      | 95.2%         |
| Mexico             | 308         | 278         | 28        | 9.1%        | 69         | 39        | 28       | 86.0%         |
| China              | 6297        | 4400        | 196       | 3.1%        | 2112       | 382       | 114      | 91.3%         |
| Bangladesh         | 650         | 632         | 7         | 1.1%        | 60         | 46        | 6        | 92.7%         |
| Nigeria            | 982         | 975         | 5         | 0.5%        | 105        | 98        | 5        | 89.9%         |
| <b>Viet Nam</b>    | <b>5839</b> | <b>5715</b> | <b>16</b> | <b>0.3%</b> | <b>125</b> | <b>17</b> | <b>1</b> | <b>99.7%</b>  |
| <b>Thailand</b>    | <b>2297</b> | <b>2261</b> | <b>2</b>  | <b>0.1%</b> | <b>78</b>  | <b>42</b> | <b>2</b> | <b>98.1%</b>  |
| <b>Indonesia</b>   | <b>5652</b> | <b>5454</b> | <b>2</b>  | <b>0.0%</b> | <b>295</b> | <b>97</b> | <b>2</b> | <b>98.2%</b>  |
| Egypt              | 5307        | 4939        | 1         | 0.0%        | 441        | 77        | 1        | 98.4%         |
| <b>Philippines</b> | <b>328</b>  | <b>328</b>  | <b>0</b>  | <b>0.0%</b> | <b>0</b>   | <b>0</b>  | <b>0</b> | <b>100.0%</b> |

307

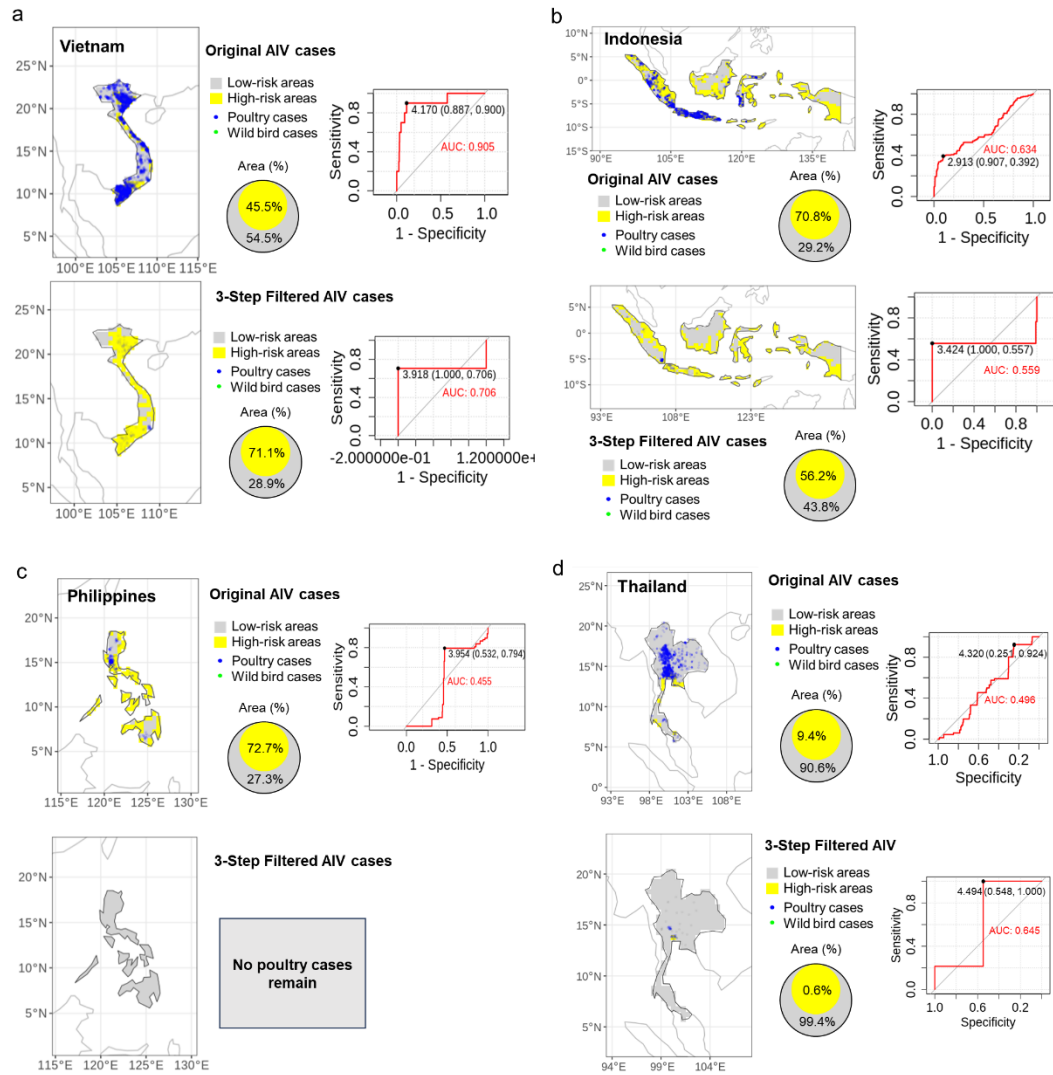
308



310

311 **Fig. R2 Spatial distribution and WAE performance as a predictor of poultry**  
312 **AIV cases in representative robust wild bird surveillance countries: (a) United**  
313 **Kingdom, (b) Germany.** Maps illustrate original and 3-step filtered AIV data,  
314 highlighting high-risk areas (yellow), low-risk areas (green), poultry cases (blue dots),  
315 and wild bird cases (green dots). A pie chart indicating the proportion of high-risk  
316 areas and low-risk areas. The ROC curve on the right evaluates the classification  
317 performance of the WAE.

318



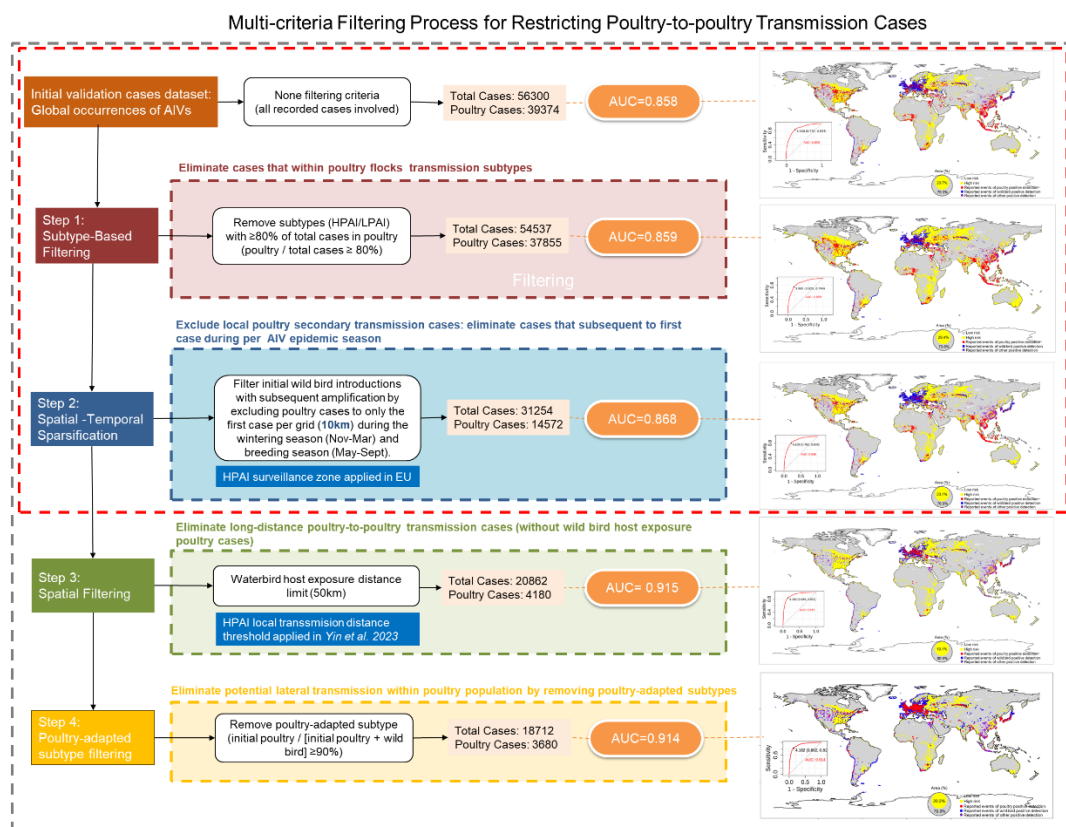
**Fig. R3 Spatial distribution and WAE performance as a predictor of poultry AIV cases in representative minimal wild bird surveillance countries: (a) Vietnam, (b) Indonesia, (c) Philippines, (d) Thailand.** Maps illustrate original and 3-step filtered AIV data, highlighting high-risk areas (yellow), low-risk areas (green), poultry cases (blue dots), and wild bird cases (green dots). A pie chart indicating the proportion of high-risk areas and low-risk areas. The ROC curve on the right evaluates the classification performance of the WAE.

### 3. A stricter elimination of potential lateral transmission within poultry population (Step 4) by removing poultry-adapted subtypes.

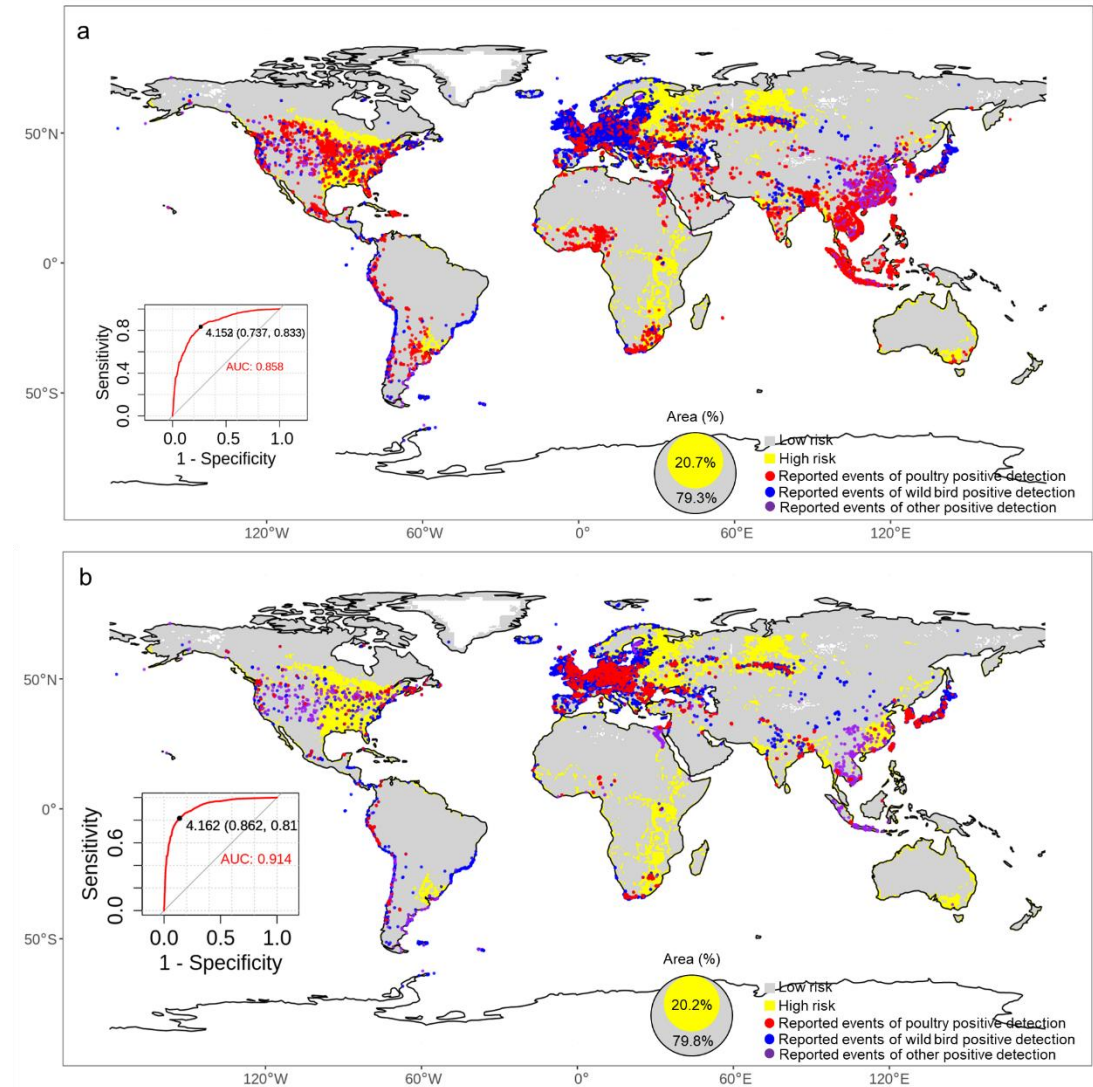
We acknowledge the concern regarding the inclusion of H7N9 outbreaks, which are widely recognized to be maintained within poultry populations rather than wild birds. In response to concerns and eliminate potential lateral transmission cases within poultry population, we introduced **Step 4**, specifically designed to exclude poultry-adapted subtypes (Supplementary Fig. 7). **Subtypes were excluded if poultry cases accounted for  $\geq 90\%$  of bird detections (i.e., initial poultry cases / [initial poultry**

+ wild bird cases]  $\geq 0.9$ ). The excluded subtypes included H7N6, H7N8, H7N9, H5N2, H9N2, H5N9. The final AUC of Step 4 was 0.914.

Consequently, due to the imbalance in wild bird surveillance records, Steps 3 and 4 were omitted from the global assessment to avoid inevitable excessive removal of spillover events, which may occur in areas with limited wild bird surveillance or among poultry-adapted subtypes. To ensure methodological transparency and enable readers to fully evaluate the filtering logic, we added the following predictive performance results obtained from various validation datasets to the Results section (Lines 137-142 on page 7): “WAE demonstrated high explanative power (AUC = 0.87, s.d. 0.001) for global avian influenza cases after applying spatiotemporal filtering to exclude local poultry-to-poultry transmission events (Supplementary Text 2). The framework also exhibited robustness when compared with the unfiltered original avian influenza case data (Supplementary Fig. 17a, AUC = 0.86) and validated with the outbreak data excluding poultry-adapted subtypes (Supplementary Fig. 17b, AUC = 0.91).” We have also provided the results of all four validation case data filtering steps in the Supplementary Materials. Detailed descriptions are provided in the section “Multi-criteria filtering process for restricting poultry-to-poultry transmission cases.” Additionally, we retained validation results based on both unfiltered and complete filtered (after the 4-step process) poultry-to-poultry transmission cases in the Supplementary Fig. 17, allowing readers to assess the WAE’s predictive performance under different assumptions.



**Supplementary Fig. 7 Multi-criteria Filtering Process for Restricting Poultry-to-Poultry Transmission Cases.** Each panel displays the final output of one filtering step, highlighting the spatial distribution of reported avian influenza (AIV) detections. Red dots represent reported events of poultry positive detection, blue dots indicate wild bird positive detection, and purple dots mark other positive detections. Yellow-shaded areas denote high-risk area, derived using the optimal WAE threshold identified via the Youden index at each corresponding step. Due to widespread limitations in wild bird surveillance, only the first two filtering steps were applied in the main analysis to target local poultry-to-poultry transmission following initial spillover events.



**Supplementary Fig. 17 Validation of waterbird activity entropy (WAE) for predicting avian influenza before and after complete filtering poultry-to-poultry transmission. (a)** Spatial distribution of high and low binary exposure risk areas and

locations of original cases from the raw dataset. The high- and low-exposure risk areas were divided by the optimal threshold of WAE (4.15), which was determined from the receiver operating characteristic (ROC) curve via the Youden index. **(b)** Spatial distribution of high and low binary exposure risk areas and locations of filtered cases from a 4-step filtering process (see Supplementary Fig. 7). The high- and low-exposure risk areas were divided by the optimal threshold of WAE (4.16), which was determined from the receiver operating characteristic (ROC) curve via the Youden index. The red points represent poultry cases; the blue points denote wild bird cases; the purple points indicate other types of cases.

Comment #3. Another major concern is that the reported Area Under the Curve (AUC) values may be misleading. As currently computed, these metrics appear to be heavily influenced by regions with little/no poultry and low wild bird activity. In such areas, low WAE values are trivially associated with the absence of outbreaks, which can inflate apparent model performance. This issue is particularly evident in the case of H7N9: despite its very restricted geographic distribution (China), the WAE model still achieves an AUC of 0.67. This suggests that the performance metric is capturing broad-scale absence patterns rather than fine-scale spillover dynamics. To more accurately assess the predictive relevance of WAE, the evaluation should be restricted to areas with substantial poultry presence, i.e. areas where spillover is plausible, and focus on distinguishing presence from absence within the poultry production landscape. This is where WAE is intended to identify gradients in exposure risk among susceptible host populations. Without such refinement, the current evaluation framework risks overstating the model's capacity to detect meaningful ecological interactions.

**Response:** We agree that a reliable evaluation of the predictive performance of the WAE framework is essential. To address this concern, we have conducted more systematic validations and sensitivity analysis of WAE framework across regions with a series of poultry density gradients (Supplementary Fig. 18). Although the predictive ability tended to decline with an increase in poultry density, **the results demonstrated that our framework has strong robustness under these conditions, maintaining an  $AUC \geq 0.7$  across all ten tested density gradients.**

Specifically, when the analysis was restricted to areas above the median poultry density (excluding the lower 50%), the framework exhibited strong predictive performance ( $AUC > 0.8$ ). The identified high-risk zones, which accounted for 13% of the assessed area, encompassed nearly 80% of all positive records, with both accuracy and sensitivity surpassing 0.75. This indicates that the WAE assessment framework remains robust in the higher half of poultry density area.

Moreover, although sensitivity decreased to approximately 0.7 in areas within the top 10% of global poultry density ( $\geq 195$  poultry/100 ha), the predictive ability of the WAE framework remained acceptable (Accuracy  $\geq 0.7$ ,  $AUC = 0.76$ ). We suppose that this pattern is primarily attributable to the extremely high poultry density (over 200/100ha) typically found in densely populated towns and intensive farming operations. In such areas, elevated human activity and the absence of natural waterbird habitats result in poultry cases being reported more frequently in human-dominated, low-waterbird-activity zones (e.g., poultry plants or intensive farms) rather than in high-WAE natural areas. Nevertheless, under these extreme conditions, the WAE assessment framework could still serve as a valuable tool for identifying farms adjacent to high-WAE areas that require enhanced vigilance against AIV spillover.

The details of this sensitivity analysis are provided as follows.

**Assessing the Sensitivity of the WAE Framework for Explaining AIV Occurrence in Areas of High Poultry Density:** To evaluate the predictive

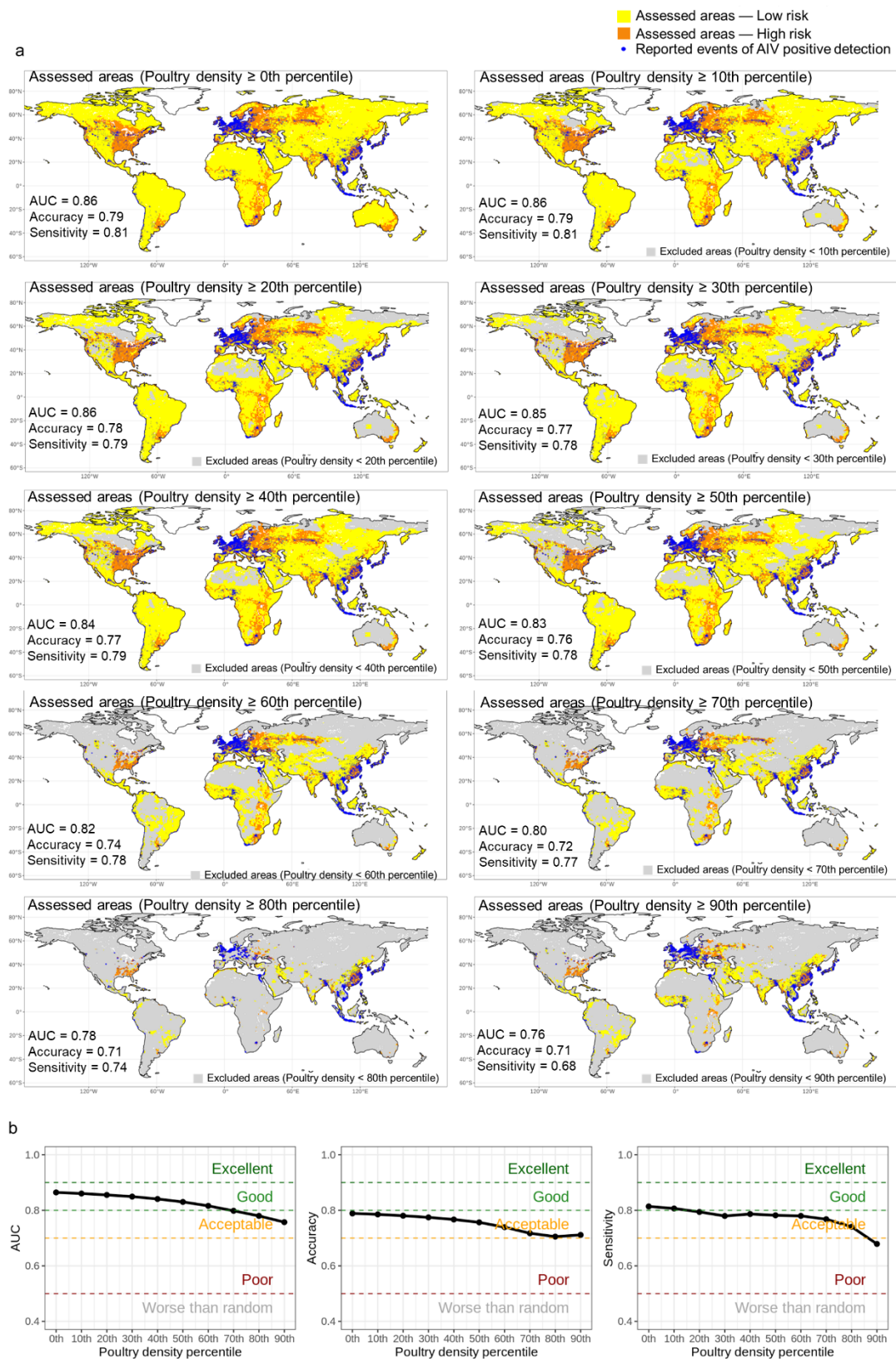
performance of WAE in high poultry density areas, we progressively narrowed the assessed areas by applying thresholds based on poultry density percentiles. These thresholds ranged from the 0th to the 90th percentile in 10-percentile increments (Supplementary Fig. 18) using global poultry density data. Within each assessed range, we performed Receiver Operating Characteristic (ROC) analysis using AIV case data derived from the filtering process (first two steps) (Supplementary Fig. 7), in order to validate the predictive performance and robustness of the WAE framework. The optimal high-risk WAE threshold was identified by maximizing Youden's index, the same method applied in global assessment. To more comprehensively evaluate the robustness and sensitivity of the WAE framework, changes in key performance metrics (e.g., AUC, accuracy, and sensitivity) across this poultry density gradient were presented as analyzed results (Supplementary Fig. 18b).

As illustrated in Supplementary Fig. 18b, the WAE framework demonstrated strong predictive performance up to the 80th percentile of global poultry density (54 poultry/km<sup>2</sup>), with the AUC consistently exceeding 0.8 (s.d. = 0.001). **Across all density thresholds, accuracy remained above 70%. Notably, even beyond the 90th percentile, where sensitivity declined slightly to below 0.7, both AUC and accuracy remained acceptable (AUC > 0.70, accuracy > 70%).** The series of spatial distribution results of risk areas mapped under the WAE framework (Supplementary Fig. 18a) shows that although the extent of relatively high-risk zones contracted considerably as poultry density thresholds increased, most predictive performance metrics remained robustly above acceptable levels. These results underscore the reliability of the WAE framework in predicting AIV occurrence and affirm its value as a useful indicator for risk assessment, even in regions with high poultry density.

The results of the sensitivity analysis of the robustness of WAE framework was added and mentioned to the results and discussion section (**Lines 143-146 on page 7-8**) as follows, “Furthermore, WAE maintained strong predictive robustness (mean AUC = 0.83 across ten poultry density gradient regions) for high-risk regions, even as regions with lower poultry densities were progressively excluded from the validation (Supplementary Fig. 18).” **In the discussion section (Lines 234-238 on pages 12):** “Notably, the WAE demonstrated consistently acceptable predictive capacity (AUC > 0.7) across areas with varying poultry density levels, further underscoring its robustness (Supplementary Fig. 18), although its predictive performance declined as the assessment area narrowed to regions of higher poultry density.”

**Furthermore, we would like to clarify that our study is aimed at assessing exposure risk originating from waterbird activity for various susceptible host populations, including humans, poultry, and cattle.** The WAE metric is designed to quantify waterbird activity patterns to identify regions with elevated exposure potential. This framework could help regions with few or no outbreak records more accurately evaluate their exposure risk. We believe this approach could be valuable for highlighting under-surveilled developing and undeveloped regions, where reported AIV event data may be sparse or incomplete.

469  
470



471

472

Supplementary Fig. 18 Validation of waterbird activity entropy (WAE) as a

**predictor of poultry avian influenza detections across varying poultry density levels. (a)** Spatial distribution and validation of risk areas mapped using the WAE framework across ten poultry density gradient ranges. Assessed areas in each poultry density gradient were derived from regions where poultry density exceeded each percentile threshold (0th to 90th percentile, in 10-percentile increments). These percentiles were calculated using poultry density raster data (chickens and ducks, from the GLW4 dataset <sup>3</sup>). Within each map, the blue points represent the validation cases, which were poultry AIV detections after two-step filtering and were restricted to their corresponding assessed areas. The orange area represents high-risk regions and the yellow area indicates relatively low-risk regions, based on consistent methods as global assessment. **(b)** Robustness evaluation of the WAE framework across a poultry density gradient. Line plots depict changes in the area under the curve (AUC), accuracy, and sensitivity across ten poultry density percentile gradients. AUC represents overall predictive performance, accuracy indicates the proportion of correct predictions, and sensitivity reflects the true positive detection rate. These metrics, derived from ROC analysis (pROC package, R), were used to comprehensively evaluate the robustness of the WAE framework.

Comment #4. In conclusion, while the manuscript presents an innovative metric, the current evaluation framework does not convincingly establish WAE as a predictor of avian influenza spillover from wild birds. The conflation of outbreak types and reliance on broad-scale absence data for validation all weaken the strength of the conclusions. I encourage the authors to refine their validation approach in order to more effectively support the central claims.

**Response:** We sincerely appreciate your critical suggestion, which has helped us strengthen our validation methods.

A more systematic validation and sensitivity analysis of the WAE framework across regions with a series of poultry density gradients (gradually exclude the effect of absence data) was conducted to verify the robustness and reliability of the WAE framework in regions with high poultry density.

We have assessed the effectiveness of our two-step filtering process and clarified our reason for not adopting Step 3 in the results of main text of manuscript through: 1) Sensitivity analysis of the Multi-criteria Filtering Process based on the proportion of poultry-to-poultry transmission cases in the AIV transmission process, classified by AIV subtypes; and 2) Sensitivity analysis of the long-distance lateral transmission filter step (Step 3) of the Multi-criteria Filtering Process across countries/regions with varying proportions of poultry cases. Furthermore, we introduced a stricter filtering process (Step 4) to exclude poultry-adapted subtypes and provide readers with additional perspectives for evaluating WAE under alternative assumptions. While our global-scale analysis cannot fully disentangle spillover from lateral spread, future localized studies with finer-resolution surveillance data may help validate this distinction more precisely (Fig. R2). However, it is outside the scope of this global scale

study.

Additionally, we have added a limitation analysis in the discussion regarding the challenge of fully excluding poultry-to-poultry transmission. " Yet, it is important to recognize even after spatio-temporal filtering, the remaining AIV cases may still contain localized lateral spread cases within poultry populations rather than just true spillover events from wild waterbirds. Such ambiguity may affect the interpretation of exposure risk and the calibration of predictive models, particularly in regions with limited surveillance data or unbalanced surveillance strategies between wild and susceptible domestic animals." We would like to clarify the role of WAE, due to the limitation of global scale data availability, under current applied cases filtering assumptions, **it is more appropriate to consider WAE as an indicator for assessing exposure risk for various susceptible host populations (including humans, poultry, and cattle) posed by waterbird activity, rather than a direct index of spillover risk for poultry.** We recognize that the AIV cross-species spillover process involves other more complex factors beyond exposure to virus carrier, such as viral characteristics and host susceptibility. We hope these revisions adequately respond to your concerns and improve the interpretability of our findings.

**References:**

1. Pittman M, Laddomada A. Legislation for the control of avian influenza in the European Union. *Zoonoses and Public Health* **55**, 29-36 (2008).
2. Yin S, *et al.* Functional traits explain waterbirds' host status, subtype richness, and community-level infection risk for avian influenza. *Ecology letters* **26**, 1780-1791 (2023).
3. FAO. GLW 4: Gridded Livestock Density (Global - 2015 - 10 km).  
<https://data.apps.fao.org/catalog/dataset/15f8c56c-5499-45d5-bd89-59ef6c026704>  
(2015).