Ecology and environment predict spatially stratified risk of highly pathogenic avian influenza in wild birds across Europe

Sarah Hayes[1,2*]|Joe Hilton[3*], Joaquin Mould-Quevedo[4], Christl Donnelly[1,2], Matthew Baylis[3], Liam Brierley[3,5]

1. Department of Statistics, University of Oxford, Oxford, UK
2. Pandemic Sciences Institute, University of Oxford, Oxford, UK
3. Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK
4. CSL Seqirus USA, Summit, NJ, USA
5. Institute of Population Health, University of Liverpool, Liverpool, UK

* Joint first authorship

Abstract

Highly pathogenic avian influenza (HPAI) represents a threat to animal health, human health, and economic prosperity, with the ongoing outbreak in wild and domestic animals since 2021 being the largest on record. This outbreak is associated with the 2.3.4.4b clade of influenza A, and it is as yet unclear what factors have contributed to its spread at the continental scale. In this study we use Bayesian additive regression trees, a machine learning method designed for probabilistic modelling of complex nonlinear phenomena, to construct species distribution models for HPAI presence across Europe. Using these models we identify factors driving the geospatial distribution of cases and project the distribution of risk across Europe. Our models are stratified by time to capture both seasonal changes in risk patterns and shifts in HPAI epidemiology associated with the introduction of the 2.3.4.4b clade. While previous studies have aimed to predict HPAI presence from physical geography, here we explicitly consider the impact of wild bird ecology by including in our model estimates of bird species richness, abundance of specific high-risk bird taxa, and "species-trait abundance indices" describing the total abundance of species with high-risk behavioural and/or dietary traits. Our projections point to a shift in concentration of risk towards cold, low-lying regions of coastal northwest Europe associated with 2.3.4.4b, with the margins of uncertainty extending that risk further into central and eastern Europe. In coastal northwest Europe specifically, we predict a persistence of high risk throughout the year. Methodologically, we demonstrate that while the majority of variation in risk can be explained by climate and other aspects of physical geography, the addition of ecological covariates represents a valuable refinement to species distribution models of HPAI.

Introduction

Avian influenza is a highly contagious viral disease caused by avian influenza type A viruses. It can have devastating effects on animal health and livestock economics with losses reaching $billions in the USA alone (1), but also poses a threat to human public health and has pandemic potential. Spillover of avian influenza from birds to mammals is known to occur, but most spillover events are associated with little or no ongoing transmission. However, reports of mass mortality events in marine mammal populations in South America (2–4) and farmed mink in Spain (5) sparked concerns surrounding the potential for mammal-to-mammal transmission. The recent discovery of HPAI infection in cattle in North America (6,7) and multiple instances of onward transmission to farm workers (8) have prompted high-priority surveillance responses considering the threat posed by potential human adaptation of the virus.

Whilst avian influenza viruses may infect both domestic and wild birds, they are classified as either low pathogenic avian influenza (LPAI) viruses or high pathogenic avian influenza (HPAI) viruses according to their ability to cause disease in poultry. HPAI viruses are defined as those that cause severe disease in young chicks intravenously inoculated with the virus in a laboratory setting, or those that possess multiple basic amino acids in the haemagglutinin cleavage site, which has been associated with high virulence. HPAI causes severe disease and high mortality rates in infected poultry, whilst LPAI tends to be much milder both in poultry and in other avian species. Following transmission of LPAI viruses to poultry, viruses may evolve into HPAI viruses and cause severe disease. Avian influenza viruses are

divided into subtypes based on the surface glycoproteins haemagglutinin (H) and neuraminidase (N), with HPAI typically restricted to subtypes containing the H5 or H7 glycoprotein.

Historically, HPAI infections were restricted mainly to domestic birds, but since 2002, reports of symptomatic HPAI in wild birds have become much more common (9–13). The HPAI virus strain A/goose/Guangdong/1/1996 (H5N1) (gs/Gd) first emerged in geese in southeast Asia in 1996 (14). In late 2002, mortality in wild birds due to gs/Gd HPAI was reported in Hong Kong (10). This was the first report of HPAI mortality in wild birds since an isolated outbreak in wild terns in South Africa in 1961 (15). By late 2005, gs/Gd HPAI viruses had reached Europe, with the migration of wild birds implicated in this long-distance transmission (16,17). Since then, there have been multiple outbreaks of gs/Gd origin HPAI in Europe in both domestic and wild birds (13,18). Genetic divergence and reassortment has led to multiple clades of the gs/Gd HPAI virus, but in recent years clade 2.3.4.4b has emerged to be the dominant clade and has expanded both in geography and host range (18), affecting a much wider diversity of wild species, including gannets (*Sulidae*), pelicans (*Pelicanidae*), terns (*Sterninae*), and skuas (*Stercorariidae*). Phylodynamic analyses also show longer duration of circulation and wider spatial dissemination of 2.3.4.4b viruses within wild birds compared to previous H5Nx clades (19,20). This expansion of transmission increases the risk of further reassortment events akin to those reported as preceding the current H5N1 2.3.4.4b outbreak (19). Therefore, there is a critical need to better understand the role of wild bird ecology in the transmission and maintenance of avian influenza viruses.

Since the emergence of the gs/Gd HPAI viruses in wild birds, there has been much interest in whether these viruses can be maintained in wild bird populations. In 2006, six HPAI H5N1 viruses were isolated from apparently healthy migratory ducks in China (21) and recent infection studies have suggested that mallard ducks (*Anas platyrhynchos)* may shed the currently circulating H5N1 virus whilst showing asymptomatic infection (22,23). In addition to risks associated with particular host species, host and virus ecology has also been proposed to be important in the transmission of avian influenza viruses. Behaviours that increase bird density and host-to-host contact rates such as colony breeding and pre-migratory congregation may be associated with an increased risk of exposure to avian influenza viruses via an increase in both direct and indirect contact (24). Migration places a high physiological demand on the birds, which may have a negative impact on the immune system (25) and thus may be expected to increase susceptibility to infection, although evidence for this is somewhat conflicting (24). LPAI viruses are spread predominantly via the faecal-oral route and thus foraging behaviour has been proposed to affect likelihood of transmission. LPAI viruses in water can remain infectious for over 30 days at 0 degrees C and the high prevalence of LPAI viruses in dabbling ducks may be in part due to increased exposure via surface water feeding (26,27). However, virus shedding with HPAI viruses has also been linked to the respiratory route (28–31) which may reduce the importance of foraging behaviours on transmission.

Species distribution models (SDMs) make predictions across geographic space and time using a set of predictor variables. They have traditionally been used in ecology to predict the distribution of plant and animal species using environmental variables, but more recently have been applied to infectious disease (32–36). SDMs have been used for modelling avian influenza viruses in domestic birds only (37–43) , wild

birds only (44–50) and both domestic and wild birds combined (51,52). Models have been produced at both a local level (48–50) and on a much larger scale (44–47), but few have tried to include host ecology at a continental scale, with many using environmental predictors alone. Belkhiria *et al.* (48) included Important Bird and Biodiversity Areas as model predictors to act as a proxy for the presence of migratory birds, whilst Moriguchi *et al.* (49) included population sizes of migratory and dabbling ducks. Hill *et al.*(53) used poultry holding densities, poultry species-specific risk estimates, and wild bird population sizes to inform a spatial mechanistic model of incursion into poultry farms in the UK, but did not include physical geography in their model. Huang *et al.* (46) explored more sophisticated ecological measures of host community composition on HPAI H5N1 across Europe, incorporating features such as species richness and abundance of waterbirds in general and of specific waterbird high-risk hosts, as well as phylogenetic relatedness to those high-risk hosts. None of these previous studies combine a continental scale, ecological detail, and a direct comparison between the epidemiology of H5N1 2.3.4.4b and that of previously circulating HPAI strains.

The aim of this study is to map and understand the geospatial risk of HPAI presence across Europe using an SDM approach. Our modelling builds on previous studies which focus on physical geography by explicitly incorporating ecological covariates including avian species richness, abundance of avian species with specific behavioural risk factors, and abundance of specific high-risk avian taxa. We train a Bayesian additive regression trees (BART) model on HPAI detections since 2005 and use data on the physical geography and ecology of over 600 wild bird species of Europe to estimate the probability of HPAI presence at a 10km resolution. Although we primarily use models to identify specific locations at high risk of acting as epicentres for future HPAI outbreaks, we also calculate variable importance and partial dependence to interrogate which specific risk factors are highly predictive of HPAI and why, including novel contributions from host ecological data. We sense-check our models by comparing their risk projections with data on domestic bird HPAI detections, as introduction of HPAI into domestic poultry may result from contact with infected wild birds or wild bird-derived fomites, and so we expect domestic poultry cases to be more likely to occur in areas where we predict high risk of HPAI in wild birds.

Methods

Table 1. Environmental covariates included in the models.

Topographic Covariates	Details	Resolution of data source	Year of data	Source
Elevation	Minimum elevation (metres above sea level)	30 arc seconds	GTOP030 - 1999 CGIAR-SRTM - 2018	Geodata package R https://cran.r-project.org/web/packages/geodata/geodata.pdf
Elevation	Maximum elevation (metres above sea level)	30 arc seconds	GTOP030 - 1999 CGIAR-SRTM - 2018	Geodata package R https://cran.r-project.org/web/packages/geodata/geodata.pdf
Elevation	Difference between minimum and maximum elevation	30 arc seconds	GTOP030 - 1999 CGIAR-SRTM - 2018	Processed using minimum and maximum elevation listed above
Elevation	Modal elevation (metres above sea level)	30 arc seconds	GTOP030 - 1999 CGIAR-SRTM - 2018	Geodata package R https://cran.r-project.org/web/packages/geodata/geodata.pdf
Normalised Difference Vegetation Index (NDVI)	Mean over the quarter. Normalised to limit values to between -1 and 1	1km	2022	Modis https://modis.gsfc.nasa.gov/data/dataprod/mod13.php
Land cover	One raster layer for each of the 17 land cover categories	0.5km	2022	Modis https://lpdaac.usgs.gov/products/mcd12q1v061/
Distance to coast	Distance from centre of raster cell to nearest coast (km)		N/A	Calculated directly from map data
Distance to inland	Distance from centre of raster	30 arc	2004	Calculated using

water	cell to nearest permanent inland water (m)	seconds		data from Global Lakes and Wetland Database https://www.worldwildlife.org/publications/global-lakes-and-wetlands-database-lakes-and-wetlands-grid-level-3
BioClimatic Covariates				
Relative humidity	Quarterly mean of daily midday (local time) relative humidity measured at 2 metres (%)	0.1 degrees	2022	Copernicus https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-agrometeorological-indicators?tab=overview
Temperature	Mean of the difference in the minimum monthly temperature and maximum monthly temperature for each month within the quarter (degrees centigrade)	2.5 arc minutes	2018	WorldClim https://www.worldclim.org/data/monthlywth.html
Temperature	Quarterly mean of monthly mean temperatures (degrees centigrade) (Mean monthly temperature for each month calculated using: Mean temperature = Minimum temperature + diurnal range/2)	2.5 arc minutes	2018	WorldClim https://www.worldclim.org/data/monthlywth.html
Temperature	Quarterly temperature variation (degrees centigrade) (Difference between the maximum of the 3 monthly maximum temperatures within the quarter and the minimum of the 3 minimum monthly temperatures within the quarter)	2.5 arc minutes	2018	WorldClim https://www.worldclim.org/data/monthlywth.html
Precipitation	Quarterly mean of the	2.5 arc	2018	WorldClim

	monthly total precipitation (mm)	minutes		https://www.worldclim.org/data/monthlywth.html
Zero-degree isotherm	Quarterly mean of the mean monthly zero-degree isotherm (metres)	0.25 degrees	2022	Copernicus https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview
Zero-degree isotherm	Number of days the zero-degree isotherm was below 1 metre at midday at Coordinated Universal Time (UTC)	0.25 degrees	2022	Copernicus https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview
Livestock demography				
Chicken density	Headcount of animals	5 arc minutes	2010	Gridded livestock of the world 3 https://www.fao.org/livestock-systems/global-distributions/chickens/en/
Duck density	Headcount of animals	5 arc minutes	2010	Gridded livestock of the world 3 https://www.fao.org/livestock-systems/global-distributions/ducks/en/

Data assembly

We assembled a set of geospatial covariate rasters at a 10km spatial resolution. This resolution was chosen to smooth potential error in geolocated coordinates of positive HPAI records given the wide-ranging movement of wild bird hosts, and for close alignment with the native resolution of many covariates (Table 1). We used the EPSG:3035 coordinate reference system, which is intended for spatial statistical analysis across continental Europe (54). To account for seasonal changes in climate and bird ecology we assemble separate covariate rasters for each quarter, with Quarter 1 running from January 1st to March 31st, Quarter 2 running from April 1st to June 30th, Quarter 3 running from July 1st to September 30th, and Quarter 4 running from October 1st to December 31st.

Environmental covariates

Environmental covariates previously reported to be important in avian influenza dynamics were selected (44–50,55). Covariates were categorised as topographical, bioclimatic or livestock demography and are listed in Table 1. There was a large amount of variability in the resolution and coordinate reference systems of the different data sets used. Where calculations were required these were done at the original resolution prior to reprojection. Rasters were then reprojected to the EPSG:3035 coordinate reference system and the resolution converted to 10km using bilinear interpolation for continuous covariates and nearest neighbour for categorical covariates.

Topographical

Elevation data were sourced using the R package "geodata" v0.5-8 (56) which combines Shuttle Radar Topography Mission (SRTM) data and Global 30 Arc-Second Elevation (GTOPO30) data. Elevation data were processed to extract the minimum, maximum and modal elevation for each of the 10km resolution cells of our project area. A fourth elevation raster containing the difference between the minimum and maximum values recorded within each of the cells was also produced.

The normalised difference vegetation index (NDVI) is a measure of the vegetation canopy greenness and acts as a marker of vegetation density. NDVI data were sourced from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) data in 16-day periods at a 1km resolution. The data were grouped to reflect the quarters as closely as possible. A mean NDVI for each quarter was calculated and standardised to limit the range between -1 and 1.

Data on land cover were sourced from MODIS Land Cover Type (MCD12Q1) Version 6.1. Land cover was categorised as one of 17 different land cover types numbered 1-17. During reprojection, the nearest neighbour method was used to create a raster at 10km resolution. Separate rasters for each of the land cover types are then created with the value of the cell representing whether or not that land cover type is assigned to that cell.

Inland water bodies were identified using the Global Lakes and Wetlands database provided by the WWF. We excluded intermittent (i.e., non-permanent) lakes and wetlands with the resulting raster layer containing distance to any permanent water body of any type.

Bioclimatic

Temperature and precipitation data were obtained from WorldClim historical monthly data for the most recent year available at initial download (2018) at a 2.5 arc minute resolution. Temperature data are entered as quarterly mean of the monthly differences in minimum and maximum temperatures, quarterly mean of monthly mean temperatures and the variation in temperature over the quarter (which is calculated by taking the difference between the minimum of the minimum values for each of the three months within the quarter and the maximum of the maximum values for each of the three months of the quarter). Precipitation is entered as the mean of the monthly values for total precipitation for each of the three months of the quarter.

Relative humidity data were downloaded from the European Centre for Medium Range Weather Forecasts (ECMWF) Copernicus dataset, obtained by the ERA-Interim reanalysis of monthly observations from 1979 to the present. Data reflect the daily relative humidity at 2 metres above ground level at midday local time. Mean quarterly humidity was calculated as the mean of the daily humidity values for each of the three months of the quarter.

Zero-degree isotherm data were also obtained from Copernicus and reflect the height above the Earth's surface where the temperature in Celsius passes from positive to negative values at the specified time. Quarterly mean of the mean monthly zero-degree isotherm and the number of days within the quarter where the zero-degree isotherm was below 1 metre at midday (UTC) are included in the model.

Livestock demography

Data regarding the density of chickens and ducks were obtained from Gridded Livestock of the World version 3. Version 3 was used for both species due to large amounts of missing data for ducks in version 4.

Table 2. Ecological covariates included in the models.

Covariate	Total abundance or behaviour-weighted?	Species-level source
Anatinae ("dabbling ducks")	Total	eBird https://science.ebird.org/en/status-and-trends
Anserinae (swans and geese)	Total	eBird https://science.ebird.org/en/status-and-trends
Ardeidae (herons)	Total	eBird https://science.ebird.org/en/status-and-trends
Arenaria/Calidris (turnstones and sandpipers)	Total	eBird https://science.ebird.org/en/status-and-trends
Aythyini ("diving ducks", or pochards)	Total	eBird https://science.ebird.org/en/status-and-trends
Laridae (gulls)	Total	eBird https://science.ebird.org/en/status-and-trends
Percentage time spent feeding within 2m of water surface	Weighted	EltonTraits https://doi.org/10.6084/m9.figshare.c.3306933.v1
Percentage time spent feeding >2m below water surface	Weighted	EltonTraits https://doi.org/10.6084/m9.figshare.c.3306933.v1
Percentage diet plants	Weighted	EltonTraits https://doi.org/10.6084/m9.figshare.c.3306933.v1
Percentage diet scavenging	Weighted	EltonTraits https://doi.org/10.6084/m9.figshare.c.3306933.v1
Percentage diet endothermic vertebrates	Weighted	EltonTraits https://doi.org/10.6084/m9.figshare.c.3306933.v1
Congregative	Total	IUCN Red List

		https://www.iucnredlist.org/
Migratory	Total	IUCN Red List https://www.iucnredlist.org/
Below threshold phylogenetic distance to known host species	Total	BirdTree https://birdtree.org/ CLOVER https://github.com/viralemergence/clover
Species richness	N/A	eBird https://science.ebird.org/en/status-and-trends

Ecological covariates

We constructed raster layers of ecological risk factors by combining species-level risk factors,estimates of global species population and relative spatial abundance data into a single index of 'species-trait abundance'. To reflect the role of seasonal migration patterns in wild bird HPAI epidemiology (57,58) we obtained species-level indicators specifying whether a given species was migratory and/or congregative from the IUCN Red List (59). To capture bird diet and foraging behaviour, we took estimates from EltonTraits (60) of the percentage contribution of specific food sources and foraging behaviours to each species' total diet and time spent foraging. Of the food sources considered in EltonTraits, we considered plants and endothermic vertebrates (mammals and birds). Of the foraging behaviours, we considered feeding within 2m of water surfaces, feeding below the water surface at depths exceeding 2m, and scavenging. Evolutionary relatedness between hosts is a well-established predictor of infection potential in various systems (61,62), including avian influenza (63). To capture this effect of phylogenetic structure in susceptibility across avian host species,we downloaded 100 posterior estimates of the bird phylogenetic tree from Birdtree (64) and used the function cophenetic.phylo from the R package ape, v5.8 (65) to calculate a matrix of pairwise phylogenetic distances between bird species. We used the CLOVER database (66) to identify known avian influenza host species, and for each species in the Birdtree phylogeny calculated the minimum phylogenetic distance to a known host species. Based on preliminary analysis we converted this to an indicator variable specifying for each species whether the phylogenetic distance to any known avian influenza host species met a threshold value of 50 million years since lineage divergence. Taxonomic conventions differed between data sources, with disagreements over both species names and distinctions between species and subspecies. To allow for cross-referencing between data sources with differing taxonomic conventions we performed a semi-automated rationalisation process using the AVONET database of bird species names (67), detailed in Supplemental Text S1. A complete list of ecological covariates used in our models is given in Table 2.

We obtained estimates of species abundance in the form of percentage population at a 9km grid resolution from the 2022 data release of eBird Status and Trends (68). These estimates are calculated using species distribution models trained on checklist-based bird observation data submitted to the eBird citizen science platform, and, for a given species, take the form of 52 weekly geospatial raster layers specifying the percentage of the global population of a given species in each grid cell in each week of the year. The species modelled by eBird include only a subset of those known to be present in Europe; a total of 708 species as opposed to 944 species listed as being observed in Europe in Avibase (69). For each species eBird provides quality ratings on a scale of 0 (lowest quality) to 3 (highest quality) stratified by bird life cycle season ("breeding", "nonbreeding", "postbreeding migration", and "prebreeding migration" seasons, or "resident" for non-migratory species), and numerical calendar weeks specifying the timing of these seasons. The exact combination and timing of seasons is unique to each bird species, with bird life cycle seasons lacking an exact correspondence with calendar quarters. We used the seasonal quality metrics and temporal boundaries to generate weekly quality measures for each species, and for each quarterly model included only those species which had a quality rating ≥2 for the majority of the quarter (i.e., for ≥7 of 13 weeks). We defined the estimated abundance of a given

species in a given grid cell for a given quarter to be the mean of the weekly abundance across all weeks in that quarter with quality rating ≥2. To convert from percentage population to absolute abundance we subsequently scaled each species abundance raster by the corresponding estimated global population size provided by Callaghan *et al.*(70), giving us an estimated absolute abundance raster. After discarding species with problematic taxonomic statuses and year-round low quality levels we were left with 617 of the 708 European species modelled by eBird Status and Trends.

We used these estimated absolute abundance layers to develop spatial raster layers representing indices of population-level ecological traits for use as covariates. Species richness was estimated by counting the total number of species in each grid cell that had a total estimated abundance greater than 1. For migratory behaviour, congregative behaviour, and phylogenetic host distance threshold, we created indices by adding together all the abundance layers for species which had a positive indicator for these factors. For the dietary and foraging behaviour risk factors, we summed all of the abundance layers weighted by the percentage of dietary and foraging behaviour allocated to the specific behaviour for each species in EltonTraits, to give a measure of relative population behaviour in each grid cell. We constructed covariate layers representing six specific bird taxa by adding together all the abundance layers for species in each. These taxa were Anatinae, Anserinae, Ardeidae, Aythyini, Laridae, as well as *Arenaria* and *Calidris*, which were combined in a single layer due to their close relatedness and the small number of species in *Calidris* (64). These taxa were chosen based on high numbers of appearances in the CLOVER database (66) or previous identification as high-risk avian influenza species (71).

Outcome data

Data on confirmed avian influenza cases in wild birds were obtained from publicly available data sources - Food and Agricultural Organization of the United Nations (FAO) Empres-i and World Animal Health Information System (WAHIS) database provided by the World Organization for Animal Health (WOAH). Data spanned from October 2005 until April 2024 and were filtered to be HPAI only and to have coordinates within the specified study area. Duplicate entries based on date and geographical coordinates were removed. Data entries that had the same date and geographical coordinates but were listed as different subtypes were retained within the data. Data were mapped to a 10km grid considering cells positive if they contained ≥1 HPAI record.

Based on patterns of epidemic peaks and subtype emergence in Europe (Figure 1), independent models were constructed to capture A) discrete, short-term HPAI outbreaks from 2005 to 2021 and B) the extended H5N1 epizootic outbreak, ongoing since 2021. Each set was divided into a training period, and a subsequent test period held out from training to assess model performance in near-future projections. For A) models were trained on data from 7/10/2005 to 31/12/2019, capturing significant H5N1 and H5N8 outbreaks as well as smaller outbreaks of H5N6 (Figure 1), before testing on a discrete H5N8 outbreak from 1/1/2020 to 31/8/2021. For B) models were trained on H5N1 data from 1/9/2021 to 30/3/2023 before testing on H5N1 data from 1/4/2023 to 30/3/2024. Training/test divisions were chosen to cover a minimum of one example of each calendar quarter (Figure 1C).

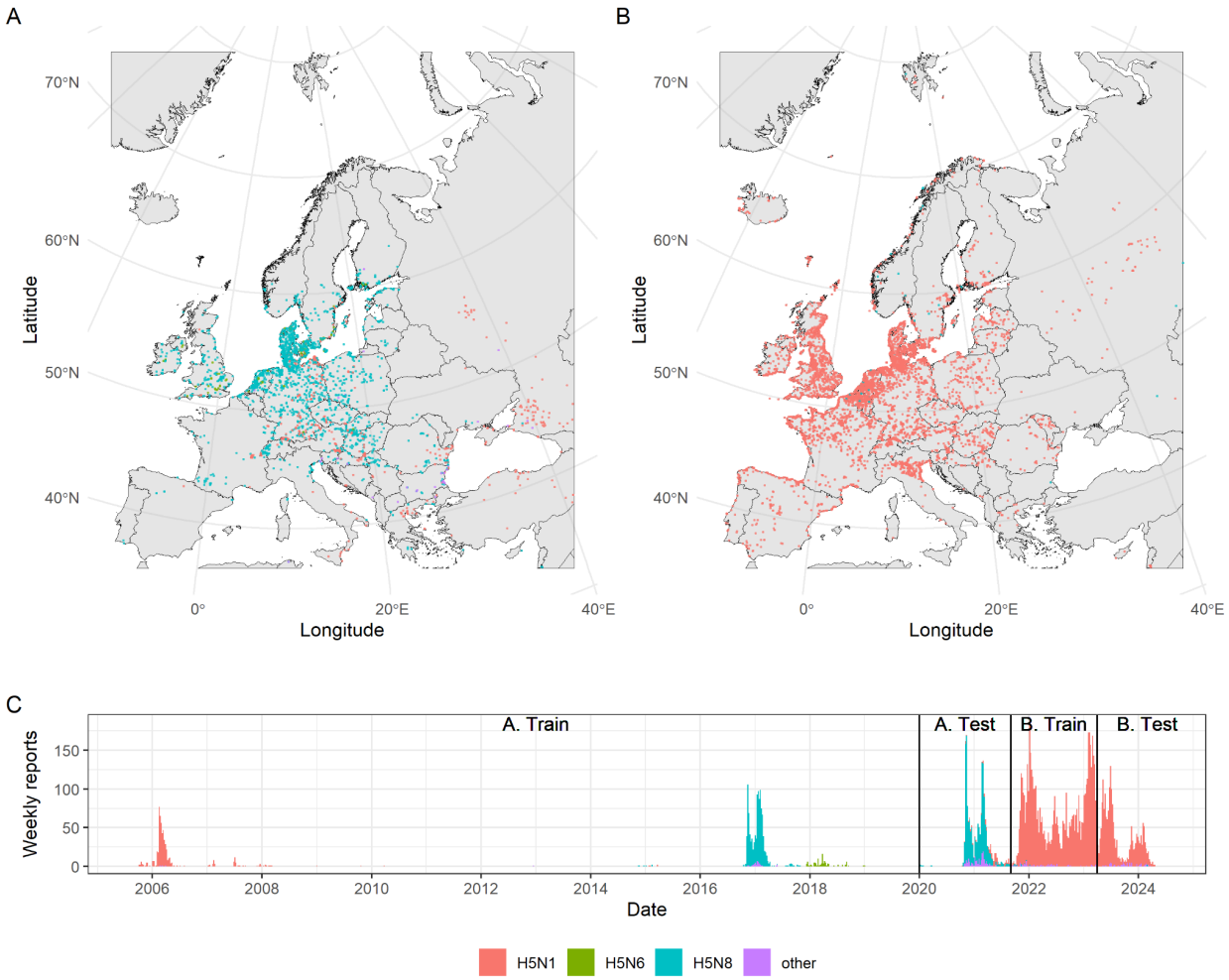


Figure 1. Wild bird HPAI incident data in Europe.
Spatiotemporal distribution of incidents of highly pathogenic avian influenza subtypes reported in wild birds, harmonised from EMPRES-i and WAHIS databases. A) and B) Geolocated incident reports over time periods A and B, respectively. C) Total independent incident reports per week with separate model training and test periods annotated. Colour denotes subtype recorded (H5N1; H5N6; H5N8; or other, including H5N5, H5N3, H5N4, and H7N7).

Pseudo-absence generation

SDM analyses are often challenged by lack of available absence data (i.e., true negatives) to train on alongside presences; for HPAI, few authorities systematically report their negative test results. Several model approaches exist to handle this problem, including presence-only or 'one-class' classification and niche boundary estimation (72,73). We use an established solution in generating 'pseudo-absence' or background points, which fulfil the role of the negative class by representing areas the modelled species could feasibly be present in but are not observed in. It has been strongly emphasised that pseudo-absences should reflect the same biological restrictions and spatial sampling heterogeneities as presences (72,74). Therefore, to approximate avian influenza surveillance processes which depend on

human density and accessibility to sample birds, we obtained bird sighting records sourced via citizen surveillance.

We obtained all public bird observations recorded in eBird by downloading the eBird Basic Dataset (EBD) (75) before filtering to records within Europe using R package `auk`, v0.7.0 (76). We then calculated total sightings (i.e., unique data-user-geolocation records, regardless of species sighted) per 10km grid cell over the study period (7/10/2005 - 30/3/2024) to create a spatial raster layer approximating bird sampling accessibility (Supp Figure S1). Pseudo-absence cells were then sampled at an initial 1:1 ratio with positives for each quarter using function add_pseudoabsence() from R package `ibis.iSDM` v0.1.2 (77), weighted by normalised total eBird sightings with an additional constraint of being ≥ 25km from the nearest positive cell. Pseudo-absences were generated jointly for training/test sets to avoid positive cells from one being considered pseudo-absences in the other.

Data resampling

As a result of sampling processes, occurrence records for species (or diseases) are often spatially concentrated in clusters, which can bias SDM fitting procedures. As in other computational fields, these biases can be mitigated by resampling training data with spatial data thinning being a popular approach (78). However at continental scales, this may retain records that are geographically distant yet still environmentally homogeneous.

Therefore, we reduce training data redundancy by environmental thinning, which typically outperforms spatial thinning (79,80). Considering environmental traits as an n-dimensional space and dividing each trait into a given number of bins, random stratified sampling can be conducted to retain records based on environmental uniqueness (79), i.e., if there are many records from area(s) falling within given ranges of temperature, humidity, precipitation, etc., only one needs to be retained to represent this niche. We thinned using the function `occfilt_env()` from R package `flexsdm`, v1.3.4 (81), based on nine purely environmental covariates (mean relative humidity, mean temperature, temperature range, temperature variation, total rainfall, max elevation, NDVI, distance to coast, distance to inland water), each stratified into six bins. Thinning was conducted independently for positive cells and pseudo-absence cells from each quarter's training data from each dataset, resulting in between 66 and 1531 total cells (Table 3).

Table 3. Training and test set sizes in terms of number of grid cells.

Dataset	Raw training (positive, pseudo-absence)	Thinned training (positive, pseudo-absence)	Test (positive, pseudo-absence)
A, Q1	(808, 613)	(520, 457)	(541, 418)
A, Q2	(94, 90)	(81, 84)	(79, 77)
A, Q3	(37, 37)	(30, 36)	(22, 22)
A, Q4	(314, 270)	(212, 226)	(397, 342)
B, Q1	(1630, 1187)	(703, 828)	(255, 186)
B, Q2	(387, 315)	(192, 275)	(653, 533)
B, Q3	(473, 402)	(220, 338)	(322, 274)
B, Q4	(1066, 846)	(450, 630)	(252, 200)

Modelling approach

Using the above training data, we fitted BART, an ensemble-based nonparametric machine learning method based around constructing large numbers of decision trees (82). The decision tree structure of BART makes it ideal for modelling complex nonlinear relationships such as environment-host-disease interaction. An ensemble of candidate models is constructed via Markov Chain Monte Carlo (MCMC), with the predicted response value for a given covariate datapoint given by the mean predicted response across these candidate models. One particular advantage of BART is that credible intervals can be generated by taking quantiles of the ensemble of projected values. We carried out our model fitting using the R programming language package *embarcadero* (83), which offers an implementation of BART designed to interface with geospatial data and generate spatial projections; based on a fitted BART model *embarcadero* can generate a raster of predicted response variables from a multilayer raster of covariates.

We explore four possible formulations of our basic model, based on the inclusion or exclusion of two model features: cross-seasonal covariates and a random intercept term on country. When modelling without cross-seasonal covariates, we construct models using seasonally-variable environmental and ecological covariates for only the same quarter as the HPAI records being trained on, whereas with cross-seasonal covariates, we construct using seasonally-variable covariates from all four quarters. This is to allow for potential delay effects in the infectious disease dynamics; for instance, winter case numbers in a given location could be influenced by the local climate over summer, or large numbers of high-risk birds congregating in a location at one time of year could drive the circulation of infection lasting throughout the year. When modelling with a random intercept term on country, we label each raster cell by the country its coordinates lie within using the R package rworldmap, v1.3.8

(84), and use the random intercept option provided in *embarcadero* to fit the random intercept term. This is intended to capture country-level heterogeneities in HPAI detection and reporting capacity. However, since our pseudo-absence generation process accounts for spatial heterogeneity in observation processes through its weighting on birdwatching intensity data, there is a possibility that introducing a random intercept term will result in us "double accounting" for spatial heterogeneity, motivating us to compare models with and without random intercept. Along with this double accounting problem, the substantial spatial heterogeneity in our data means that the testing data for the different dataset-quarter combinations consistently contains data from countries which do not appear in the corresponding training data, which is likely to make assessment of model performance difficult.

We performed five-fold cross validation to optimise the parameters of the MCMC, and performed stepwise variable set reduction to optimise our covariate set for predictive ability. We fit a separate instance of BART for each of our eight quarterly datasets, with the variable set reduction allowing models for different quarters to retain distinct covariates, reflecting seasonal variations in epidemiology.

Comparison to domestic cases

We sense-checked by calculating the odds of observing a domestic bird HPAI detection in locations where domestic birds were present, stratified by model-projected risk. We extracted data on domestic bird HPAI detections from the same sources and over the same period as for wild bird cases. To ensure we only looked at cells with non-negligible domestic bird populations we added together chicken and domestic duck density raster layers to produce a raster capturing the combined density of both species. We first selected all 10km grid cells with > 1000 domestic ducks and chickens combined and labelled them as infected or uninfected based on the presence of at least one reported domestic case of HPAI. Each cell was also associated with a model-predicted probability of HPAI presence in wild birds. For each 10% band of predicted probability, we calculated the risk-stratified odds of infection as the number of cells with domestic bird HPAI detection divided by the number of cells without (each restricted to those with > 1000 domestic ducks and chickens combined). For each period (A and B) and for each quarter this gave a set of 10 risk-stratified odds which were visualised to examine potential linkage between HPAI outbreaks in wild and domestic birds.

The code used to generate our training data and perform our analysis is available at https://github.com/sarahhayes/avian_flu_sdm.

Results

Table 4. Posterior means of performance metrics for each model option on each dataset's test set. AUROC denotes Area Under the Receiver-Operating Characteristic curve, CV denotes covariates. Presence data in test set A is from the 2020-2021 H5N8 outbreak, with 95% credible intervals in brackets. Presence data in test set B is from 1/4/2024 to 30/3/2024, during the ongoing (since 2021) H5N1 outbreak.

			Model			
			base model	**+ random intercept**	**+ cross-seasonal CV's**	**+ random intercept + cross-seasonal CV's**
A	**Q1: January – March**	**Sens.**	0.81 (0.73, 0.82)	0.83 (0.58, 0.85)	0.85 (0.78, 0.85)	0.83 (0.74, 0.86)
		Spec.	0.89 (0.80, 0.92)	0.77 (0.53, 0.89)	0.81 (0.77, 0.88)	0.79 (0.72, 0.84)
		AUROC	0.92 (0.87, 0.92)	0.84 (0.68, 0.88)	0.91 (0.87, 0.92)	0.87 (0.80, 0.88)
	Q2: April – June	**Sens.**	0.70 (0.55, 0.81)	0.73 (0.10, 0.95)	0.82 (0.70, 0.86)	0.82 (0.66, 0.86)
		Spec.	0.73 (0.54, 0.86)	0.59 (0.03, 0.99)	0.68 (0.56, 0.82)	0.68 (0.47, 0.83)
		AUROC	0.80 (0.68, 0.83)	0.68 (0.45, 0.78)	0.86 (0.75, 0.88)	0.84 (0.72, 0.86)
	Q3: July – September	**Sens.**	0.77 (0.45, 0.91)	0.59 (0.18, 0.95)	0.95 (0.64, 1.00)	0.82 (0.14, 1.00)
		Spec.	0.25 (0.10, 0.75)	0.25 (0.05, 0.75)	0.50 (0.20, 0.70)	0.50 (0.05, 0.90)
		AUROC	0.50 (0.30, 0.75)	0.38 (0.24, 0.60)	0.75 (0.53, 0.82)	0.63 (0.30, 0.73)
	Q4: October - December	**Sens.**	0.86 (0.77, 0.88)	0.80 (0.53, 0.88)	0.81 (0.74, 0.84)	0.85 (0.68, 0.88)
		Spec.	0.69 (0.66, 0.79)	0.88 (0.59, 0.94)	0.80 (0.71, 0.84)	0.68 (0.50, 0.82)
		AUROC	0.89 (0.83, 0.90)	0.93 (0.77, 0.92)	0.90 (0.84, 0.90)	0.86 (0.76, 0.88)
B	**Q1: January – March**	**Sens.**	0.81 (0.76, 0.83)	0.74 (0.53, 0.77)	0.77 (0.70, 0.79)	0.69 (0.59, 0.75)
		Spec.	0.87 (0.81, 0.90)	0.94 (0.87, 0.96)	0.92 (0.85, 0.94)	0.94 (0.84, 0.96)
		AUROC	0.92 (0.87, 0.93)	0.90 (0.83, 0.92)	0.93 (0.89, 0.93)	0.91 (0.82, 0.91)
	Q2: April – June	**Sens.**	0.84 (0.75, 0.86)	0.80 (0.70, 0.84)	0.80 (0.71, 0.82)	0.80 (0.73, 0.84)
		Spec.	0.39 (0.33, 0.48)	0.48 (0.34, 0.58)	0.55 (0.43, 0.61)	0.51 (0.37, 0.61)
		AUROC	0.69 (0.63, 0.70)	0.70 (0.61, 0.72)	0.73 (0.66, 0.74)	0.74 (0.66, 0.76)
	Q3: July – September	**Sens.**	0.77 (0.71, 0.81)	0.76 (0.71, 0.84)	0.85 (0.79, 0.91)	0.76 (0.72, 0.82)
		Spec.	0.74 (0.62, 0.78)	0.75 (0.59, 0.77)	0.60 (0.49, 0.67)	0.75 (0.56, 0.77)
		AUROC	0.83 (0.77, 0.84)	0.84 (0.78, 0.85)	0.84 (0.77, 0.84)	0.82 (0.77, 0.83)
	Q4: October - December	**Sens.**	0.84 (0.77, 0.86)	0.88 (0.69, 0.90)	0.80 (0.76, 0.85)	0.86 (0.76, 0.89)
		Spec.	0.78 (0.70, 0.84)	0.74 (0.61, 0.86)	0.78 (0.71, 0.81)	0.74 (0.63, 0.80)
		AUROC	0.90 (0.85, 0.90)	0.89 (0.83, 0.89)	0.89 (0.84, 0.89)	0.87 (0.83, 0.88)

Performance metrics for each model option are given in Table 4. When trained on short-term outbreaks of multiple subtypes in the 2005-2021 period (dataset A), models showed good generalisability in capacity to distinguish areas with HPAI (Table 4). Performance was weaker in July-September (Q3) across all models, likely reflecting the smaller amount of training data available for this calendar period. Including a random intercept term stratified by country did not improve performance for dataset A. When a random intercept term was not included, including cross-seasonal covariates substantially improved performance in terms of Area Under the Receiver-Operating Characteristic curve (AUROC) for Quarters 2 and 3 relative to the model without cross-seasonal covariates, while showing comparable performance for Quarters 1 and 4.

For the ongoing (since 2021) H5N1 outbreak (dataset B), there was a more moderate difference in performance between the different models. Fitting to Quarter 2 of dataset B is particularly difficult given the test set here is larger than the training set (192 positive records in the training data compared to 654 positive records in the testing data; Table 3), and neither of the models without cross-seasonal covariates achieve a specificity of more than 50% in Quarter 2.

Based on the performance metrics in Table 4 we selected models with cross-seasonal covariates and no random intercept term to generate risk projections and assess variable importance. We generated 10km resolution projections of HPAI risk across Europe with 95% credible intervals (Figure 2). The larger number of observations available for dataset B allowed us to project with narrower credible intervals during this period.

Considering short-term outbreaks before the ongoing H5N1 outbreak (period A), risk hotspots for HPAI in wild birds are present across central and eastern Europe in colder seasons (Figure 2A, Q1 and Q4), with wider and more northerly persistence predicted in warmer seasons (Q2 and Q3). Restricting to highest-confidence hotspots of HPAI presence (high risk at 2.5th percentiles) showed consistent risk concentrated around central and eastern European shorelines and waterways, while restricting to highest confidence hotspots of HPAI absence (low risk at 97.5th percentiles) suggested low suitability for HPAI in high-elevation, mountainous regions as well as the Iberian peninsula.

Predicted spatial distributions of the ongoing (since 2021) epizootic in wild birds (period B) showed distinct differences. Although uncertainty was wider, projected risk was more concentrated around the British Isles and North Sea coast (particularly in warmer quarters) and less concentrated around inland central and eastern Europe (Figure 2B). Although different environments were observed to be at higher risk compared to previous short-term outbreaks, the lowest risk was again consistently observed for more mountainous and arid regions.

These continental-scale differences can be explained by differing importance of bioclimatic, topographic, livestock demographic, and host ecological factors in dynamics of each HPAI outbreak (Figure 3, Figure 4). During period A, bioclimatic variables were broadly the most heavily represented after variable selection in models of each quarter (Figure 3A). The presence of lagged variables among the bioclimatic variables retained for period A models pointed to an association between colder

quarters (Q1 and Q4; October to March) and year-round risk. Partial dependence curves showed that more variable temperatures over the entire quarter led to greater predicted risk of influenza (Figure 4A.i), and that lower zero-degree isotherms (i.e., freezing temperatures nearer ground level) during Q1 or Q4 were associated with risk in Q1, Q2 and Q4 (Figure 4A.ii). The main exception to the trend of bioclimatic dominance in period A was Q3 (July to September) where species richness across seasons was retained and showed an approximately linear relationship with projected risk (Figure 4A.vi). Even in this quarter, however, the covariate with highest influence was variation in quarterly mean temperature. Besides species richness, the other ecological covariates to be retained in models for period A were *Anatinae*, *Anserinae*, and *Ardeidae*, where estimated abundances of these taxa in colder quarters (Q1 and Q4; October to March) corresponded to increased risk throughout the year (Figure 4A.v, Supplemental Figures S3-S5). Behavioural and dietary species-trait abundance indices were not retained in any models for period A.

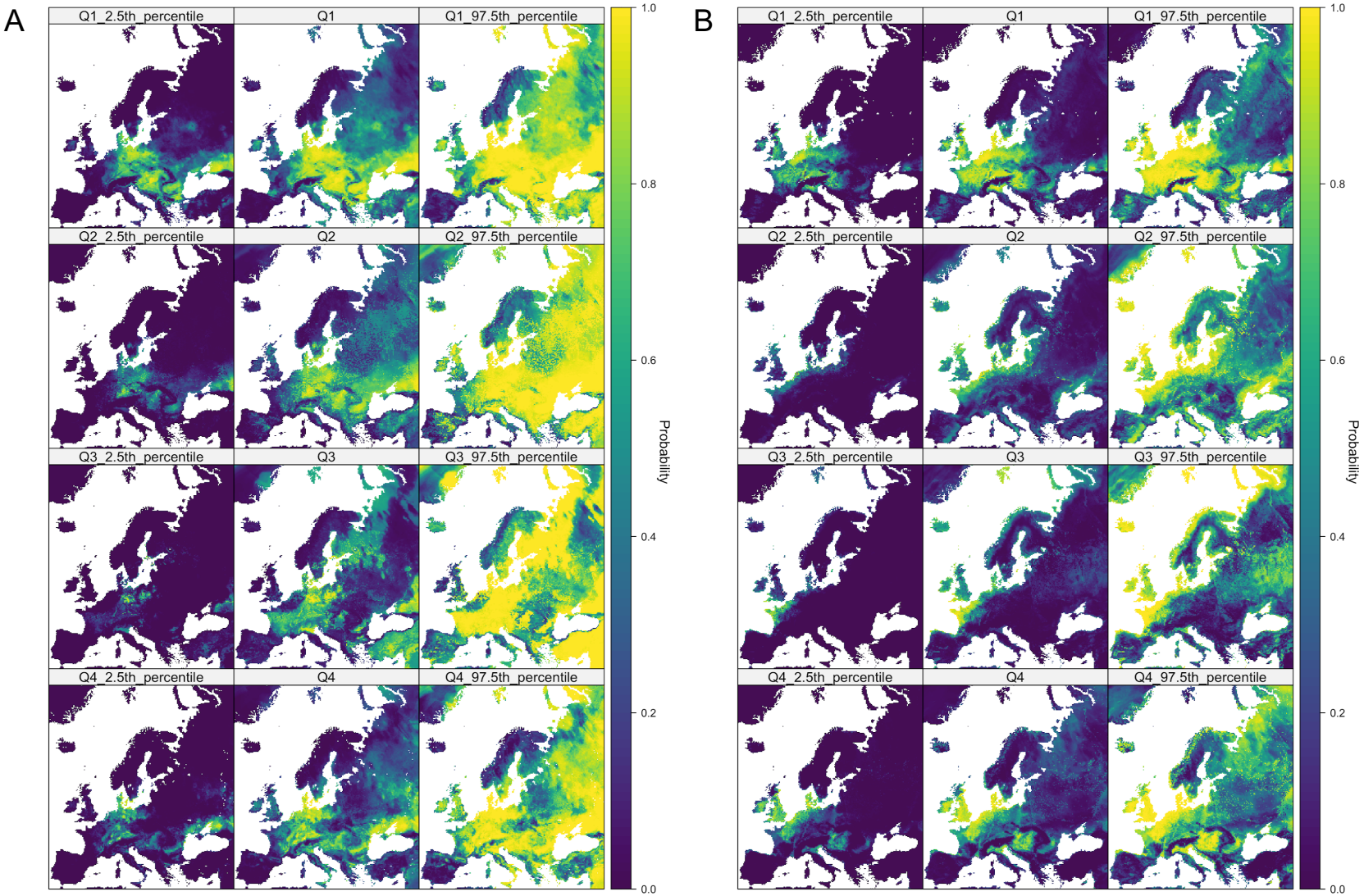


Figure 2. Projected probability of HPAI presence by quarter during period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak) and period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak) from selected final models with cross-seasonal covariates and no random intercept term. Central columns are posterior means from BART ensemble, left and right columns represent 95% credible intervals.

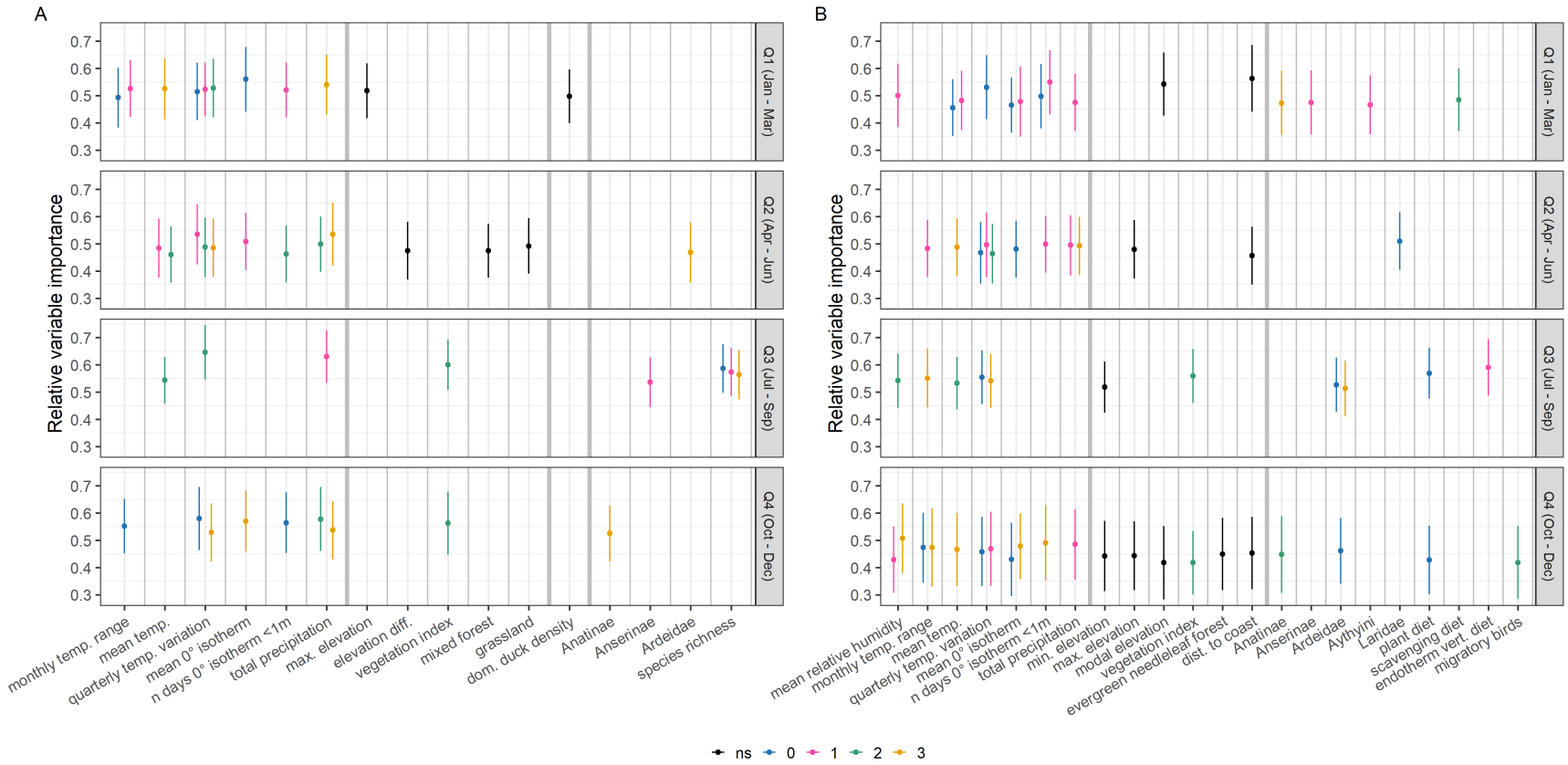


Figure 3. Variable importance of BART models.

Relative variable importance of each covariate kept in optimal fitted BART models for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak) and period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Points denote normalised proportion of times variable was used in a tree decision split out of all splits, averaged over 1000 draws from the posterior tree space. Error bars denote +/- 1 SD. Colour key denotes precedence of calendar quarter of covariate to modelled quarter, either, 0 (same quarter, e.g., Q3 model using Q3 covariate) or 1 (one quarter prior, e.g., Q3 model using Q2 covariate), 2, or 3 quarters prior; black denotes non-seasonally variant covariates ("ns"). Thicker grey lines separate bioclimatic, topographic, livestock demography, and host ecological covariates.

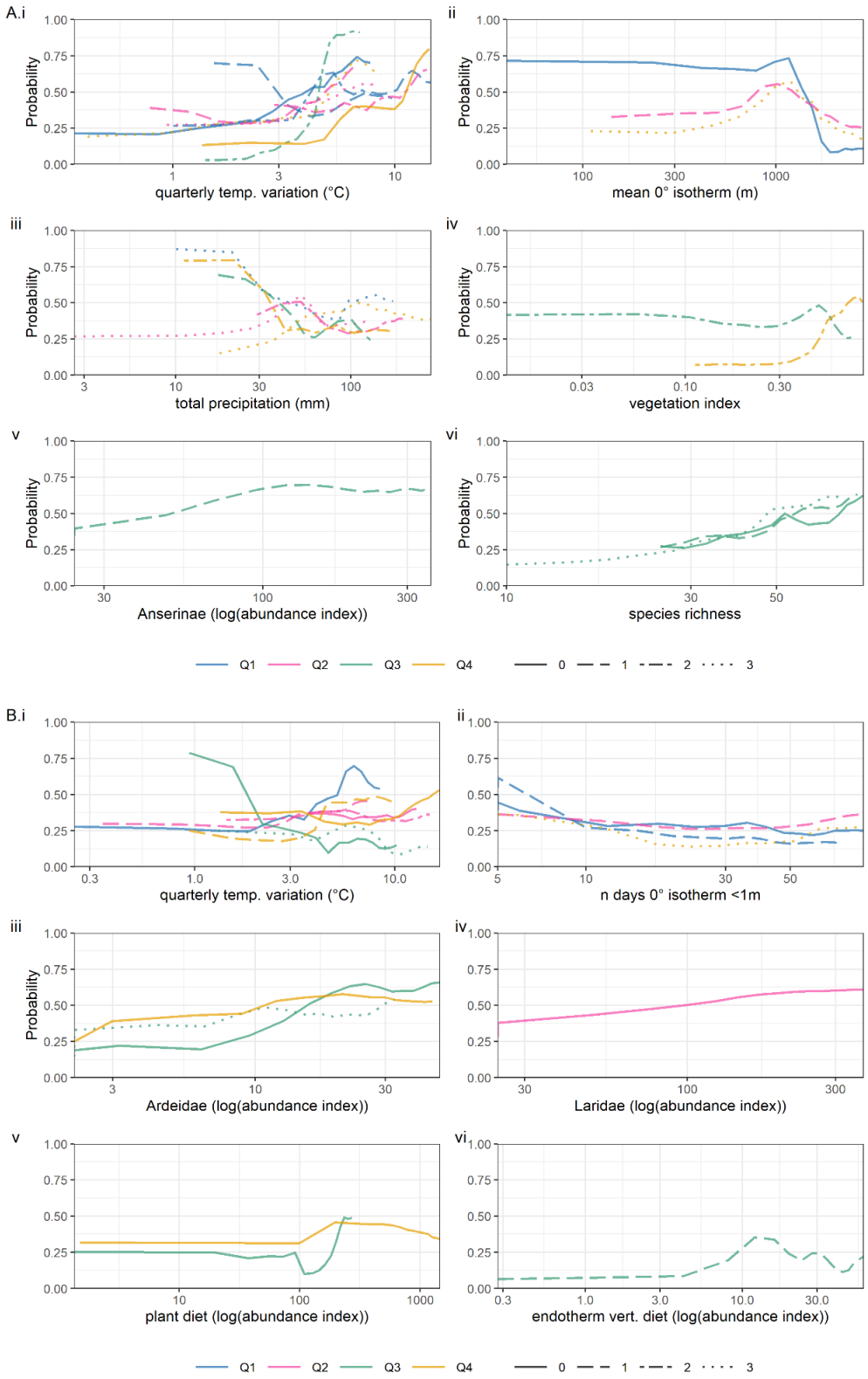


Figure 4. Partial dependence of BART models.

Partial dependence associated with top six covariates by average variable importance in final fitted BART models

for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak) and period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. Solid lines denote median values over 1000 draws from the posterior tree space. Colour denotes separate modelled quarters of HPAI and line type denotes calendar quarter delay of seasonally variant covariates with delay increasing from 0 (solid) to 3 (dotted).

Models trained on geolocated H5N1 data from period B also tended to retain bioclimatic variables, although ecological variables were more consistently featured. As in period A, winter climate, including humidity, temperature range, temperature variation, and height of zero-degree isotherm were associated with HPAI presence throughout the year (Figure 3B, Figure 4B.i, ii). Total rainfall in the previous quarter was also predictive for each quarter except Q3, although the direction of influence was not consistent (Supplemental Figures S6, S7, S9). Topographical covariates were also more consistently retained for period B models, where low elevation was associated with greater risk (Supplemental Figures S6-S9), consistent with topography of the projected risk maps. Abundance indices for *Anatinae*, *Anserinae*, *Ardeidae*, *Aythyini*, and *Laridae* were all retained in at least one period B model (Figure 3B), and all showed positive association with avian influenza risk that frequently plateaued after a threshold abundance (Figure 4B.iii, iv, Supplemental Figures S6-S9), suggesting a stronger impact from presence of specific high-risk host species during the ongoing H5N1 2.3.4.4b outbreak than in previous avian influenza outbreaks. Dietary and behavioural species-trait abundance indices were retained in Q1, Q3, and Q4, although risk profiles tended to be variable and highly nonlinear (Figure 3B.v, vi, Supplemental Figures S6, S8, S9). Along with other partial dependence curves for the period B models, this suggests that localised ecology, i.e., small-scale interactions between different covariates, are important to understanding the distribution of H5N1 2.3.4.4b.

Our comparison of model predictions with domestic HPAI detections indicated a trend towards increased odds of infection in domestic birds with model-projected risk of HPAI presence in wild birds for both model periods (Figure 5). Spatial concordance between model-generated predicted probabilities and reports of HPAI in domestic birds is visualised in Supplemental Figure S10.

Figure 5: Comparison of model predicted probabilities of HPAI risk in wild birds to reports of HPAI in domestic birds.

Odds of infection in domestic birds for each period and quarter by model predicted probabilities of HPAI risk for wild birds.

Discussion

In this study we have demonstrated how inclusion of both environmental and ecological covariates can lead to generalisable models of avian influenza distribution in wild birds, particularly in colder quarters of the year. Our analysis reveals a detectable difference in the epidemiology of the ongoing (since 2021) H5N1 outbreak compared to past outbreaks of avian influenza within Europe. This is reflected both in the projected geospatial risk distributions, and in the differences in which covariates are selected in our period A and period B models. Our projections suggest that the ongoing H5N1 outbreak is associated with an increased risk of avian influenza presence around northwest Europe, in agreement with the observed large numbers of HPAI detections in this region. Underlying risk patterns were more ambiguous in eastern and southern Europe due to less intensive sampling, though our projections suggest that the risk of HPAI disperses across these areas during the warmer spring and summer months.

In line with these observed patterns of detection, we expect risk to be associated with proximity to coastlines, low elevations, low temperatures, and presence of seabirds. We found that environmental factors reflecting physical geography tend to be more informative than wild bird ecology in predicting HPAI risk, possibly reflecting inherent constraints on wild bird ecology imposed by background environmental factors. Our results suggest that while risk is indeed highest in the heavily sampled regions of northwest Europe, there is also a strong potential for HPAI presence in regions where case detections are less frequent. From a theoretical perspective, the incorporation of ecological risk factors into HPAI presence/absence models appears to offer a small but non-negligible improvement in predictive ability.

The minimal differences in performance when adding country-level random intercepts suggest that any spatial heterogeneity in our detection data was effectively captured by weighting for background citizen surveillance of birds in our pseudo-absence generation process. However, the very low numbers of detections in specific eastern European countries (e.g., Albania, Belarus, Moldova, Ukraine) suggest that there are indeed national-level differences in detection and reporting rates which are difficult to estimate due to the absence of information from these countries in many of our training datasets. The improvement in performance we obtained by incorporating cross-seasonal covariates suggests that future mechanistic models may need to incorporate delay effects to fully capture the complex dynamics of avian influenza outbreaks in wild birds. The improvement offered by these cross-seasonal covariates was most pronounced in Q2 and 3 of dataset A; prior to the emergence of the H5N1 2.3.4.4b clade Q1 and Q4 represented the peak winter avian influenza season across Europe, whereas incidents during Q2 and Q3 respectively are likely to represent the tail and lead-in periods respectively of this peak season and thus be driven by factors underlying the peak-season dynamics.

Our risk projections point to an increased risk of H5N1 presence in northwest Europe, particularly around the North Sea coast, associated with the emergence of the 2.3.4.4b strain. The upper confidence

limits for our projections suggest that the Black Sea coast may also be a zone of high H5N1 risk, as it was prior to the ongoing H5N1 outbreak. This shift in geographical burden of risk appears to be driven by an increased importance of rainfall, elevation, and abundance of specific bird taxa (*Anatinae*, *Anserinae*, *Ardeidae*, *Aythyini*, and *Laridae*). This aligns with previous studies which have identified associations between elevation and precipitation with avian influenza risk in wild birds (44–49,63), whilst dabbling duck and diving duck populations were reported to be strongly correlated with the potential distribution of avian influenza in Japan (49).

Previous research has highlighted the association between temperature-driven waterbird movements and congregation of waterbirds at the zero-degree isotherm with the distribution and spread of HPAI (85). In line with this research and SDM studies which have identified associations between temperature and avian influenza in wild birds (44,45,47,48), temperature and zero-degree isotherm variables showed a strong influence in our models. At least one temperature variable was selected for all 8 models and temperature variation had the highest average variable importance across both periods A and B, whilst zero-degree isotherm variables were selected for all quarters except Q3 of periods A and B. The suggested influence of temperature on HPAI risk in wild birds is likely to be multifactorial and may reflect features such as habitat suitability for high-risk hosts, migratory bird movements, congregation of birds near unfrozen water sources during winter months and duration of survival of virus in the environment (26,27,85). The importance of temperature in our models suggests that future climate change might drive changes in the spatiotemporal distribution of AI in Europe. Whether past warming has played a role in the recent spread of AI remains unknown.

By conducting variable selection we demonstrated that incorporating ecology provides a non-negligible improvement in models of avian influenza presence; only period A's Q1 reduced model did not feature any ecological covariates and for both periods A and B, the influence of ecology was most pronounced during Q3. This identifies wild bird ecology as a distinct source of information not present in background environmental data, particularly during warmer months. In contrast to the more consistent set of climate covariates, ecological covariates varied between models and were more specific, i.e., rather than generalised diversity measures, exact indicator taxa and/or behaviours were preferentially kept in models of different calendar times and with different temporal lags. The combination of consistent climate covariates between models with more variable sets of ecological covariates suggests that avian influenza incidence is primarily driven by these patterns of host species-level environmental suitability but also subject to highly localised ecological factors.

We observed an association with bird species richness only for late summer (Q3) in period A, again indicating during the warmer summer months risk is more directly driven by local host ecology than in cooler periods of the year, where extremes of temperature are more important. Whether host diversity increases or decreases disease risks is somewhat contested (86) and depends on scale, pathogen traits, and host species composition. We report a positive relationship between species richness and HPAI presence, potentially capturing greater opportunities for contact and transmission between different populations (24) and/or importance of additional host taxa following Huang *et al.* reporting a similar positive effect with "higher risk host" richness (46).

Among host taxa we explicitly modelled, relative abundance of *Anatinae* (dabbling ducks), *Anserinae* (swans and geese), and *Ardeidae* (herons) were the most consistent predictors of HPAI presence. The former two taxa are generally considered maintenance hosts of avian influenza, and within *Anatinae,* common mallards are effective carriers as they may experience only mild disease, even when infected with H5N1 clade 2.3.4.4b (23). Although the total abundance index of migratory birds was only retained in Q4 for period B, the effects of migration may be reflected in the varying predictive power of these taxa throughout the calendar year. In addition to migrating birds introducing influenza to new areas, migration can also play a role in host susceptibility through stress and reduction in individual immunity (24). Alongside the abundance estimates for high-risk taxa, our construction of species-trait abundance indices represents a novel attempt to quantify the effect of ecological factors on HPAI risk at the continental scale. Although they improved models in period B, abundance indices of birds with specific diets showed variable and highly localised trends in risk (Supplemental Figures S6 - S9). We did not detect effects of specific habitat of foraging at the continental scale. Foraging behaviour has previously been suggested as a major driver of avian influenza (24) as faecal-oral transmission primarily occurs through viral shedding into water, although studies looking at viral shedding of HPAI viruses have reported high levels of respiratory shedding in addition to gastrointestinal shedding (28–31). Equivalent functional ecology and life history measures are well-established as predictors of individual-level infection (46,87), with emerging studies highlighting similar traits for community-level infection such as plant diet (88). The low importance assigned to functional ecological factors in our models is potentially due to a degree of redundancy in the models with both taxon abundance layers and species-trait abundance layers; the derived nature of the species-trait abundance layers means that they carry an inherently higher level of error than the taxon abundance layers, while contributing a limited amount of independent information compared to the taxon abundance layers. For instance, knowing the spatial abundance of specific high-risk waterfowl taxa may contribute the same information as knowing the combined species-trait abundance layers for plant-feeding, feeding around the water surface, and feeding below the water surface.

If we are to better capture host dynamics at continental scales and improve spatial understanding of avian influenza, we need more refined, higher-quality underlying ecological data. The spatial ecology layers we use come with several important caveats. In the absence of comprehensive survey data, we use projected abundance layers from eBird Status and Trends products and global species population size estimates. Our models are therefore effectively downstream from another set of models and thus need to be considered in light of any uncertainties and limitations associated with those models (89). The projections from eBird Status and Trends also only covered a subset of all bird species known to be present in Europe, meaning that our ecological covariates could potentially underestimate abundances in spatially heterogeneous ways. The role of ecological covariates in HPAI risk prediction points to high-confidence models of wild bird abundance for a more comprehensive selection of species and covering a wider geographical area as a key target for future research. Automated “computer vision” systems to recognise presence, abundance, and location of species from digital images and their metadata could prove useful in this context (90).

There are also substantial challenges associated with working with current standards of HPAI case reports. Although HPAI is classified as a listed disease by the WOAH meaning that national authorities are required to report HPAI viruses detected in both poultry and non-poultry species, including wild birds (91), surveillance strategies across countries are not standardised and may differ substantially between countries in the same region (92,93). This results in significant underreporting of wild bird HPAI, particularly in sparsely populated areas (94). Passive surveillance is frequently used and relies on submission of sick or dead wild birds. Rates of data submission are likely to show spatial and temporal variation based on factors such as human population density and public awareness of disease, whilst testing capacity may also vary over time depending on factors such as disease priority and demand. Similar variation also exists across active surveillance strategies. In this study we responded to these data quality issues by considering presence/pseudoabsence at the geographic level rather than finer-grained reported incidence data, which are associated with inherently higher levels of error due to the more complex nature of the signal being measured. Consistency in surveillance both within and between countries, reporting of negative tests and/or submission rates from passive surveillance, and accurate and consistent labelling of affected species would significantly improve data interpretability for end users. Higher quality reporting data should improve the predictive ability of complex mechanistic models, as well as broader presence/absence approaches as in our work.

In this study we have identified a shift in wild bird HPAI risk towards northwest Europe associated with the ongoing H5N1 2.3.4.4b clade, associated with changes to the underlying predictive factors driving HPAI epidemiology. Our work also serves as an assessment of the role of ecological factors in predictive models of HPAI, with our results suggesting that these factors can improve model performance, particularly if more accurate quantitative measures of wild bird ecology become available in the future.

Acknowledgements

We thank Sarah Hill, Stephen Vickers, Kieran Sharkey, Stephen Brough, Michelle Wille, and John Drake for their helpful comments on this work. This work was undertaken on Barkla, part of the High Performance Computing facilities at the University of Liverpool, UK. JH, MB, JM, and LB would like to acknowledge funding support by CSL Seqirus UK Ltd to the University of Liverpool through the 2022 – 2027 Framework Research Collaboration Agreement Influenza Research Partnership, collaboration plan 002. SH and CAD would like to acknowledge funding support from the National Institute for Health and Care Research Unit Emerging and Zoonotic Infections (Grant HPRU200907). CAD would also like to acknowledge funding support from the Medical Research Council (MRC) Centre for Global Infectious Disease Analysis (grant number MR/R015600/1), which is jointly funded by the UK MRC and the UK Foreign, Commonwealth and Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union (EU).

Bibliography

1. Chatsuwan M, Zabolonskaya E, Wright H. Spotlight on the Economic and Public Health Impacts of Avian Flu [Internet]. FAIRR; 2023 Sep [cited 2024 Jun 14]. Available from: https://www.fairr.org/resources/reports/industry-reinfected-avian-flu
2. Ulloa M, Fernández A, Ariyama N, Colom-Rivero A, Rivera C, Nuñez P, et al. Mass mortality event in South American sea lions (Otaria flavescens) correlated to highly pathogenic avian influenza (HPAI) H5N1 outbreak in Chile. Vet Q. 2023 Dec 31;43(1):1–10.
3. Leguia M, Garcia-Glaessner A, Muñoz-Saavedra B, Juarez D, Barrera P, Calvo-Mac C, et al. Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. Nat Commun. 2023 Sep 7;14(1):5489.
4. Puryear W, Sawatzki K, Hill N, Foss A, Stone JJ, Doughty L, et al. Highly Pathogenic Avian Influenza A(H5N1) Virus Outbreak in New England Seals, United States. Emerg Infect Dis. 2023 Apr;29(4):786–91.
5. Agüero M, Monne I, Sánchez A, Zecchin B, Fusaro A, Ruano MJ, et al. Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022. Eurosurveillance. 2023 Jan 19;28(3):2300001.
6. Ly H. Highly pathogenic avian influenza H5N1 virus infections of dairy cattle and livestock handlers in the United States of America. Virulence. 2024 Dec 31;15(1):2343931.
7. Burrough ER, Magstadt DR, Petersen B, Timmermans SJ, Gauger PC, Zhang J, et al. Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4b Virus Infection in Domestic Dairy Cattle and Cats, United States, 2024. Emerg Infect Dis [Internet]. 2024 Apr 1 [cited 2024 May 23];30(7). Available from: https://doi.org/10.3201/eid3007.240508
8. Garg S, Reed C, Davis CT, Uyeki TM, Behravesh CB, Kniss K, et al. Outbreak of Highly Pathogenic Avian Influenza A(H5N1) Viruses in U.S. Dairy Cattle and Detection of Two Human Cases - United States, 2024. MMWR Morb Mortal Wkly Rep. 2024 May 30;73(21):501–5.
9. Brown IH. Summary of Avian Influenza Activity in Europe, Asia, and Africa, 2006–2009. Avian Dis. 2010 Mar 1;54(s1):187–93.
10. Ellis TM, Barry Bousfield R, Bissett LA, Dyrting KC, Luk GSM, Tsim ST, et al. Investigation of outbreaks of highly pathogenic H5N1 avian influenza in waterfowl and wild birds in Hong Kong in late 2002. Avian Pathol. 2004 Oct 1;33(5):492–505.
11. Liu J, Xiao H, Lei F, Zhu Q, Qin K, Zhang X w., et al. Highly Pathogenic H5N1 Influenza Virus Infection in Migratory Birds. Science. 2005 Aug 19;309(5738):1206–1206.
12. Chen H, Smith GJD, Zhang SY, Qin K, Wang J, Li KS, et al. H5N1 virus outbreak in migratory waterfowl. Nature. 2005 Jul;436(7048):191–2.
13. Ramey AM, Hill NJ, DeLiberto TJ, Gibbs SEJ, Camille Hopkins M, Lang AS, et al. Highly pathogenic avian influenza is an emerging disease threat to wild birds in North America. J Wildl Manag. 2022;86(2):e22171.
14. Wan X f. Lessons from Emergence of A/Goose/Guangdong/1996-Like H5N1 Highly Pathogenic Avian Influenza Viruses and Recent Influenza Surveillance Efforts in Southern China. Zoonoses Public Health. 2012;59(s2):32–42.
15. Becker WB. The isolation and classification of Tern virus: Influenza Virus A/Tern/South Africa/1961. Epidemiol Infect. 1966 Sep;64(3):309–20.
16. Gilbert M, Xiao X, Domenech J, Lubroth J, Martin V, Slingenbergh J. Anatidae Migration in the Western Palearctic and Spread of Highly Pathogenic Avian Influenza H5N1 Virus. Emerg Infect Dis. 2006 Nov;12(11):1650–6.
17. Ward MP, Maftei DN, Apostu CL, Suru AR. Association Between Outbreaks of Highly Pathogenic Avian Influenza Subtype H5N1 and Migratory Waterfowl (Family Anatidae) Populations. Zoonoses Public Health. 2009;56(1):1–9.

18. Caliendo V, Lewis NS, Pohlmann A, Baillie SR, Banyard AC, Beer M, et al. Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021. Sci Rep. 2022 Jul 11;12(1):11729.
19. Xie R, Edwards KM, Wille M, Wei X, Wong SS, Zanin M, et al. The episodic resurgence of highly pathogenic avian influenza H5 virus. Nature. 2023 Oct;622(7984):810–7.
20. Zeng J, Du F, Xiao L, Sun H, Lu L, Lei W, et al. Spatiotemporal genotype replacement of H5N8 avian influenza viruses contributed to H5N1 emergence in 2021/2022 panzootic. J Virol. 2024 Feb 15;98(3):e01401-23.
21. Chen H, Smith GJD, Li KS, Wang J, Fan XH, Rayner JM, et al. Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. Proc Natl Acad Sci. 2006 Feb 21;103(8):2845–50.
22. Tarasiuk K, Kycko A, Świętoń E, Bocian Ł, Wyrostek K, Śmietanka K. Homo- and Heterosubtypic Immunity to Low Pathogenic Avian Influenza Virus Mitigates the Clinical Outcome of Infection with Highly Pathogenic Avian Influenza H5N8 Clade 2.3.4.4.b in Captive Mallards (Anas platyrhynchos). Pathogens. 2023 Feb;12(2):217.
23. Spackman E, Pantin-Jackwood MJ, Lee SA, Prosser D. The pathogenesis of a 2022 North American highly pathogenic clade 2.3.4.4b H5N1 avian influenza virus in mallards (Anas platyrhynchos). Avian Pathol. 2023 May 4;52(3):219–28.
25. Weber TP, Stilianakis NI. Ecologic Immunology of Avian Influenza (H5N1) in Migratory Birds. Emerg Infect Dis. 2007 Aug;13(8):1139.
26. Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus ADME, Fouchier RAM. Global Patterns of Influenza A Virus in Wild Birds. Science. 2006 Apr 21;312(5772):384–8.
27. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992 Mar;56(1):152–79.
28. Koethe S, Ulrich L, Ulrich R, Amler S, Graaf A, Harder TC, et al. Modulation of lethal HPAIV H5N8 clade 2.3.4.4B infection in AIV pre-exposed mallards. Emerg Microbes Infect. 2020 Jan 1;9(1):180–93.
29. Pantin-Jackwood MJ, Costa-Hurtado M, Shepherd E, DeJesus E, Smith D, Spackman E, et al. Pathogenicity and Transmission of H5 and H7 Highly Pathogenic Avian Influenza Viruses in Mallards. J Virol. 2016 Oct 14;90(21):9967–82.
30. Pantin-Jackwood MJ, Costa-Hurtado M, Bertran K, DeJesus E, Smith D, Swayne DE. Infectivity, transmission and pathogenicity of H5 highly pathogenic avian influenza clade 2.3.4.4 (H5N8 and H5N2) United States index viruses in Pekin ducks and Chinese geese. Vet Res. 2017 Jun 7;48(1):33.
31. Graziosi G, Lupini C, Catelli E, Carnaccini S. Highly Pathogenic Avian Influenza (HPAI) H5 Clade 2.3.4.4b Virus Infection in Birds and Mammals. Animals. 2024 Jan;14(9):1372.
32. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013 Apr;496(7446):504–7.
33. Simons RRL, Croft S, Rees E, Tearne O, Arnold ME, Johnson N. Using species distribution models to predict potential hot-spots for Rift Valley Fever establishment in the United Kingdom. PLOS ONE. 2019 Dec 23;14(12):e0225250.
34. de Almeida MAB, dos Santos E, Cardoso J da C, da Silva LG, Rabelo RM, Bicca-Marques JC. Predicting Yellow Fever Through Species Distribution Modeling of Virus, Vector, and Monkeys. EcoHealth. 2019 Mar 1;16(1):95–108.
35. Miller RH, Masuoka P, Klein TA, Kim HC, Somer T, Grieco J. Ecological Niche Modeling to Estimate the Distribution of Japanese Encephalitis Virus in Asia. PLoS Negl Trop Dis. 2012 Jun 19;6(6):e1678.
36. Li Y peng, Gao X, An Q, Sun Z, Wang H bin. Ecological niche modeling based on ensemble algorithms to predicting current and future potential distribution of African swine fever virus in China. Sci Rep. 2022 Sep 16;12(1):15614.
37. Alkhamis M, Hijmans RJ, Al-Enezi A, Martínez-López B, Perea AM. The Use of Spatial and

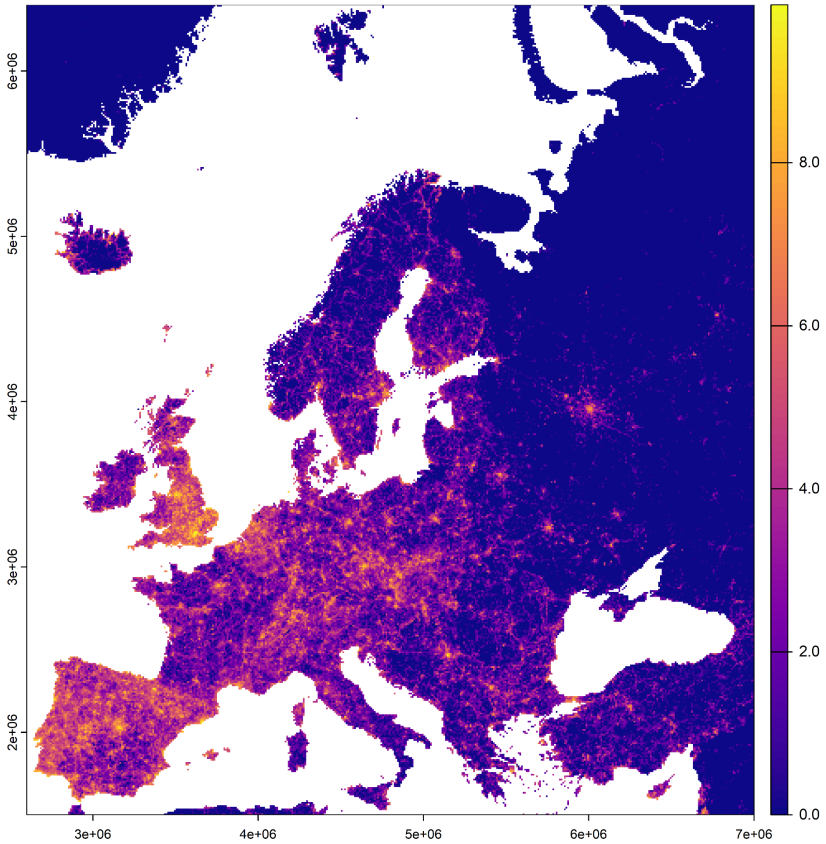
Spatiotemporal Modeling for Surveillance of H5N1 Highly Pathogenic Avian Influenza in Poultry in the Middle East. Avian Dis. 2016 Jan 12;60(1s):146–55.
38. Schreuder J, de Knegt HJ, Velkers FC, Elbers ARW, Stahl J, Slaterus R, et al. Wild Bird Densities and Landscape Variables Predict Spatial Patterns in HPAI Outbreak Risk across The Netherlands. Pathogens. 2022 May;11(5):549.
39. Yoo D sung, Chun BC, Hong K, Kim J. Risk Prediction of Three Different Subtypes of Highly Pathogenic Avian Influenza Outbreaks in Poultry Farms: Based on Spatial Characteristics of Infected Premises in South Korea. Front Vet Sci. 2022 May 31;9:897763.
40. Dhingra MS, Dissanayake R, Negi AB, Oberoi M, Castellan D, Thrusfield M, et al. Spatio-temporal epidemiology of highly pathogenic avian influenza (subtype H5N1) in poultry in eastern India. Spat Spatio-Temporal Epidemiol. 2014 Oct 1;11:45–57.
41. Yoo D sung, Song Y han, Choi D woo, Lim JS, Lee K, Kang T. Machine learning-driven dynamic risk prediction for highly pathogenic avian influenza at poultry farms in Republic of Korea: Daily risk estimation for individual premises. Transbound Emerg Dis. 2022;69(5):2667–81.
42. Gilbert M, Golding N, Zhou H, Wint GRW, Robinson TP, Tatem AJ, et al. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. Nat Commun. 2014 Jun 17;5(1):4116.
43. Stevens KB, Gilbert M, Pfeiffer DU. Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia: A spatial multicriteria decision analysis approach. Spat Spatio-Temporal Epidemiol. 2013 Mar 1;4:1–14.
44. Belkhiria J, Alkhamis MA, Martínez-López B. Application of Species Distribution Modeling for Avian Influenza surveillance in the United States considering the North America Migratory Flyways. Sci Rep. 2016 Sep 14;6(1):33161.
45. Herrick KA, Huettmann F, Lindgren MA. A global model of avian influenza prediction in wild birds: the importance of northern regions. Vet Res. 2013 Jun 13;44(1):42.
46. Huang ZYX, Xu C, van Langevelde F, Ma Y, Langendoen T, Mundkur T, et al. Contrasting effects of host species and phylogenetic diversity on the occurrence of HPAI H5N1 in European wild birds. J Anim Ecol. 2019;88(7):1044–53.
47. Si Y, Wang T, Skidmore AK, de Boer WF, Li L, Prins HHT. Environmental Factors Influencing the Spread of the Highly Pathogenic Avian Influenza H5N1 Virus in wild birds in Europe. Ecol Soc [Internet]. 2010 [cited 2022 Nov 30];15(3). Available from: https://www.jstor.org/stable/26268169
48. Belkhiria J, Hijmans RJ, Boyce W, Crossley BM, Martínez-López B. Identification of high risk areas for avian influenza outbreaks in California using disease distribution models. PLOS ONE. 2018 Jan 31;13(1):e0190824.
49. Moriguchi S, Onuma M, Goka K. Potential risk map for avian influenza A virus invading Japan. Divers Distrib. 2013;19(1):78–85.
50. Yoo DS, Lee K, Beatriz ML, Chun BC, Belkhiria J, Lee KN. Spatiotemporal risk assessment for avian influenza outbreak based on the dynamics of habitat suitability for wild birds. Transbound Emerg Dis. 2022;69(4):e953–67.
51. Bui CM, Gardner L, MacIntyre R, Sarkar S. Influenza A H5N1 and H7N9 in China: A spatial risk analysis. PLOS ONE. 2017 Apr 4;12(4):e0174980.
52. Gulyaeva M, Huettmann F, Shestopalov A, Okamatsu M, Matsuno K, Chu DH, et al. Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Influenza (AI) in the wider pacific rim using big data sets. Sci Rep. 2020 Oct 8;10(1):16817.
53. Hill A, Gillings S, Berriman A, Brouwer A, Breed AC, Snow L, et al. Quantifying the spatial risk of Avian Influenza introduction into British poultry by wild birds. Sci Rep. 2019 Dec 27;9(1):19973.
54. ETRS89-extended / LAEA Europe [Internet]. [cited 2024 Mar 11]. Available from: https://epsg.org/crs_3035/ETRS89-extended-LAEA-Europe.html

56. Hijmans RJ, Barbosa M, Ghosh A, Mandel A. geodata: Download Geographic Data [Internet]. 2023. Available from: https://CRAN.R-project.org/package=geodata
57. Yang Q, Wang B, Lemey P, Dong L, Mu T, Wiebe RA, et al. Synchrony of Bird Migration with Global Dispersal of Avian Influenza Reveals Exposed Bird Orders. Nat Commun. 2024 Feb 6;15(1):1126.
58. van Dijk JG, Verhagen JH, Wille M, Waldenström J. Host and virus ecology as determinants of influenza A virus transmission in wild birds. Curr Opin Virol. 2018 Feb 1;28:26–36.
59. IUCN Red List of Threatened Species [Internet]. [cited 2024 Mar 12]. The IUCN Red List of Threatened Species. Available from: https://www.iucnredlist.org/en
60. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: *Ecological Archives* E095-178. Ecology. 2014 Jul;95(7):2027–2027.
61. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. Host Phylogeny Determines Viral Persistence and Replication in Novel Hosts. Schneider DS, editor. PLoS Pathog. 2011 Sep 22;7(9):e1002260.
62. Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE. Host Phylogeny Constrains Cross-Species Emergence and Establishment of Rabies Virus in Bats. Science. 2010 Aug 6;329(5992):676–9.
63. Wille M, Lisovski S, Roshier D, Ferenczi M, Hoye BJ, Leen T, et al. Strong host phylogenetic and ecological effects on host competency for avian influenza in Australian wild birds. Proc R Soc B Biol Sci. 2023 Jan 25;290(1991):20222237.
64. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. The global diversity of birds in space and time. Nature. 2012 Nov;491(7424):444–8.
65. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Schwartz R, editor. Bioinformatics. 2019 Feb 1;35(3):526–8.
66. Gibb R, Albery GF, Becker DJ, Brierley L, Connor R, Dallas TA, et al. Data Proliferation, Reconciliation, and Synthesis in Viral Ecology. BioScience. 2021 Nov 2;71(11):1148–56.
67. Tobias JA, Sheard C, Pigot AL, Devenish AJM, Yang J, Sayol F, et al. AVONET: morphological, ecological and geographical data for all birds. Coulson T, editor. Ecol Lett. 2022 Mar;25(3):581–97.
68. Fink D, Auer T, Johnston A, Strimas-Mackey M, Ligocki S, Robinson O, et al. eBird Status and Trends [Internet]. 2023 [cited 2024 Mar 13]. Available from: https://ebird.org/science/status-and-trends/data-access
69. Europe bird checklist - Avibase - Bird Checklists of the World [Internet]. [cited 2024 Mar 13]. Available from: https://avibase.bsc-eoc.org/checklist.jsp?lang=EN&p2=1&list=ioc&synlang=&region=EUR&version=text&lifelist=&highlight=0
70. Callaghan CT, Nakagawa S, Cornwell WK. Global abundance estimates for 9,700 bird species. Proc Natl Acad Sci. 2021 May 25;118(21):e2023170118.
71. Ornithological data relevant to the spread of Avian Influenza in Europe (Phase 2) - Wetlands International [Internet]. [cited 2024 Mar 28]. Available from: https://www.wetlands.org/publication/ornithological-data-relevant-to-the-spread-of-avian-influenza-in-europe-phase-2/
72. Drake JM. Range bagging: a new method for ecological niche modelling from presence-only data. J R Soc Interface. 2015 Jun;12(107):20150086.
73. Drake JM, Richards RL. Estimating environmental suitability. Ecosphere. 2018 Sep;9(9):e02373.
74. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl. 2009;19(1):181–97.
75. eBird Basic Dataset. Cornell Lab of Ornithology, Ithaca, New York; 2023.

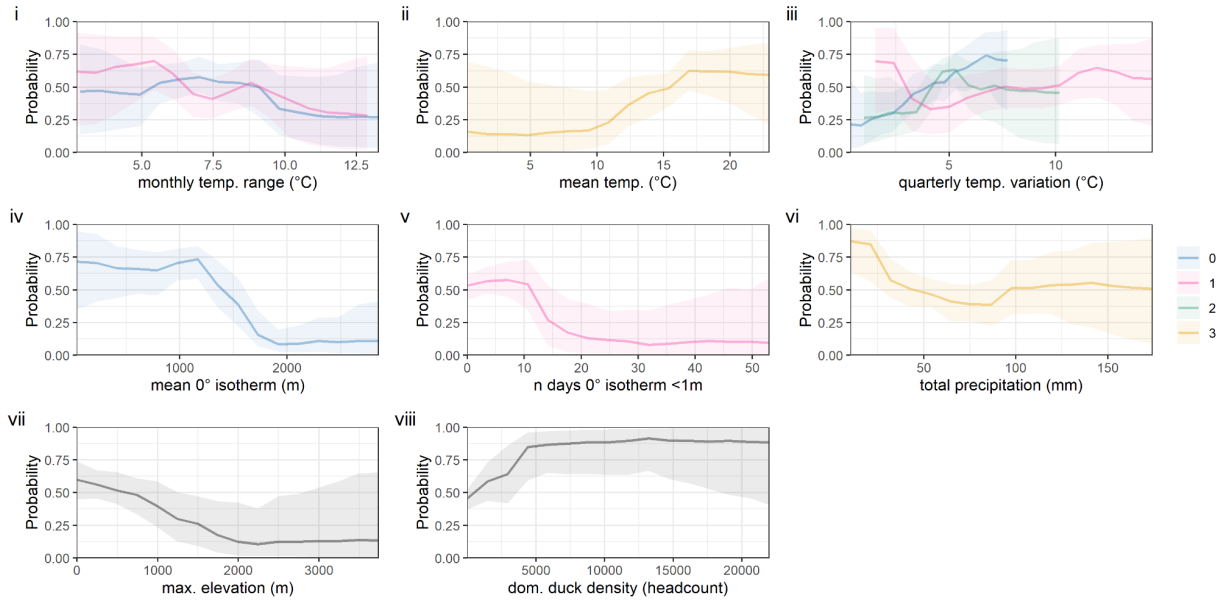
76. Strimas-Mackey M, Miller E, Hochachka W. auk: eBird Data Extraction and Processing with AWK. R package version 0.7.0 [Internet]. 2023. Available from: https://cornelllabofornithology.github.io/auk/
77. Jung M. An integrated species distribution modelling framework for heterogeneous biodiversity data. Ecol Inform. 2023 Sep 1;76:102127.
78. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography. 2015;38(5):541–5.
79. Varela S, Anderson RP, García-Valdés R, Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography. 2014;37(11):1084–91.
80. Castellanos AA, Huntley JW, Voelker G, Lawing AM. Environmental filtering improves ecological niche models across multiple scales. Methods Ecol Evol. 2019;10(4):481–92.
81. Velazco SJE, Rose MB, de Andrade AFA, Minoli I, Franklin J. flexsdm: An r package for supporting a comprehensive and flexible species distribution modelling workflow. Methods Ecol Evol. 2022;13(8):1661–9.
82. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R [Internet]. New York, NY: Springer US; 2021 [cited 2023 May 17]. (Springer Texts in Statistics). Available from: https://link.springer.com/10.1007/978-1-0716-1418-1
83. Carlson CJ. embarcadero: Species distribution modelling with Bayesian additive regression trees in R. Price S, editor. Methods Ecol Evol. 2020 Jul;11(7):850–8.
84. South A. rworldmap : a new R package for mapping global data. R J. 2011;3(1):35.
85. Reperant LA, Fučkar NS, Osterhaus ADME, Dobson AP, Kuiken T. Spatial and Temporal Association of Outbreaks of H5N1 Influenza Virus Infection in Wild Birds with the 0°C Isotherm. PLOS Pathog. 2010 Apr 8;6(4):e1000854.
86. Rohr JR, Civitello DJ, Halliday FW, Hudson PJ, Lafferty KD, Wood CL, et al. Towards common ground in the biodiversity–disease debate. Nat Ecol Evol. 2020;4(1):24–33.
87. Walsh DP, Ma TF, Ip HS, Zhu J. Artificial intelligence and avian influenza: Using machine learning to enhance active surveillance for avian influenza viruses. Transbound Emerg Dis. 2019 Nov;66(6):2537–45.
88. Yin S, Li N, Xu W, Becker DJ, de Boer WF, Xu C, et al. Functional traits explain waterbirds' host status, subtype richness, and community-level infection risk for avian influenza. Ecol Lett [Internet]. 2023 [cited 2023 Aug 21];26(10). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.14294
89. Fink D, Auer T, Johnston A, Ruiz-Gutierrez V, Hochachka WM, Kelling S. Modeling avian full annual cycle distribution and population trends with citizen science data. Ecol Appl [Internet]. 2020 Apr [cited 2023 Jul 28];30(3). Available from: https://onlinelibrary.wiley.com/doi/10.1002/eap.2056
90. Weinstein BG. A computer vision for animal ecology. J Anim Ecol. 2018;87(3):533–45.
91. WOAH - World Organisation for Animal Health [Internet]. [cited 2024 Jun 19]. Avian Influenza. Available from: https://www.woah.org/en/disease/avian-influenza/
92. Pavade G, Weber L, Hamilton K, Dehove A, Zepeda C. OFFLU review of avian influenza surveillance and epidemiological projects in some European, African, and Asian countries.
93. Authority (EFSA) EFS, Aznar I, Kohnle L, Stoicescu A, van Houtum A, Zancanaro G. Annual report on surveillance for avian influenza in poultry and wild birds in Member States of the European Union in 2022. EFSA J. 2023;21(12):e8480.
94. Klaassen M, Wille M. The plight and role of wild birds in the current bird flu panzootic. Nat Ecol Evol. 2023 Oct;7(10):1541–2.
95. Avibase - The World Bird Database [Internet]. [cited 2024 Mar 20]. Available from: https://avibase.bsc-eoc.org/avibase.jsp

96. Grey-headed Swamphen - eBird [Internet]. [cited 2024 May 21]. Available from: https://ebird.org/species/purswa3
97. Milvago chimachima (Yellow-headed Caracara) - Avibase [Internet]. [cited 2024 May 21]. Available from: https://avibase.bsc-eoc.org/species.jsp?lang=EN&avibaseid=BA33DDD5D79EAB89

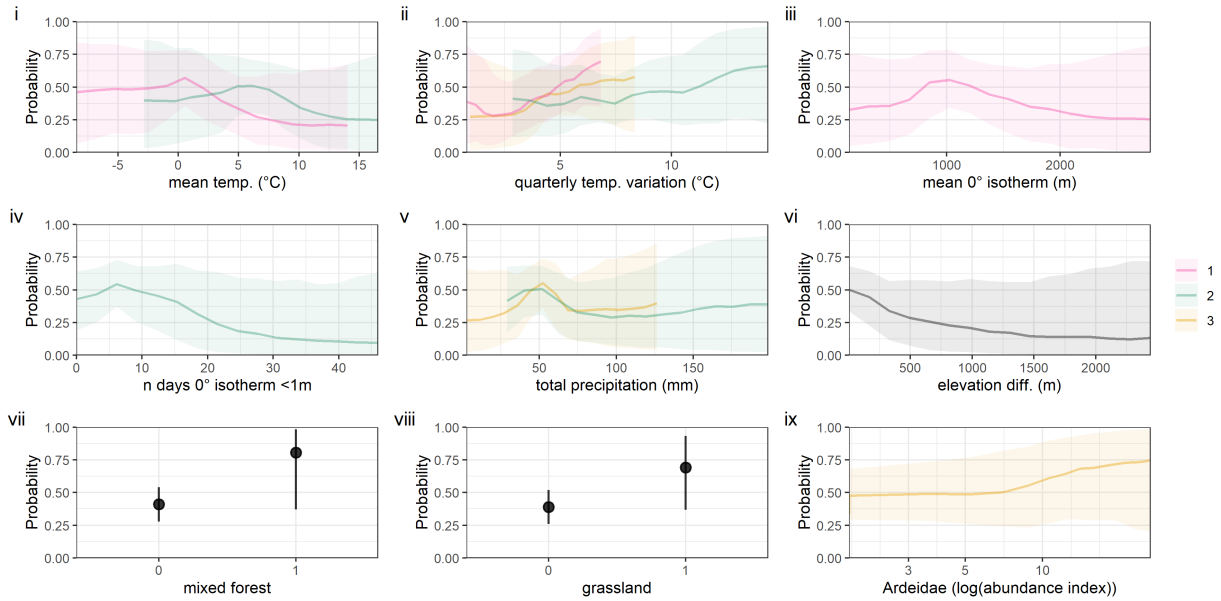
Supplementary information



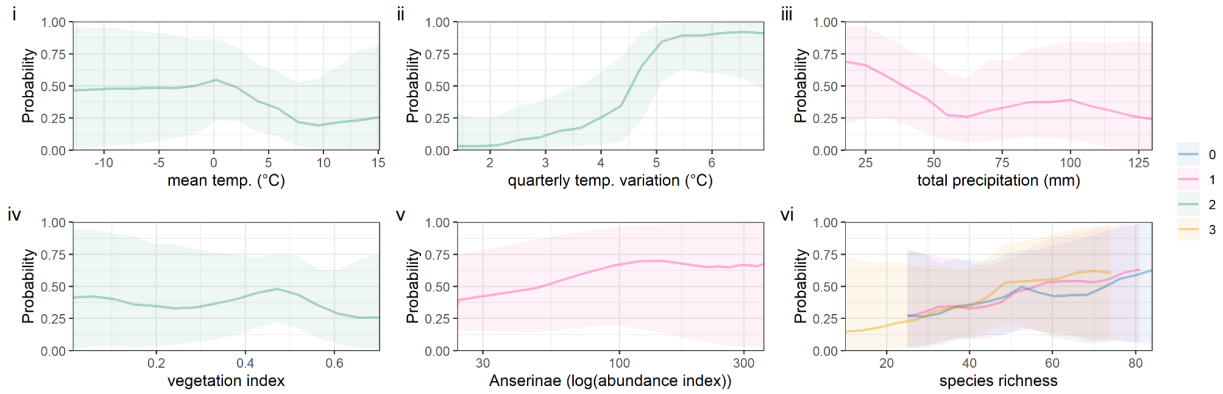
Supplemental Figure S1. Mapped density of eBird sightings.
Mapped density of bird sighting records in Europe from the eBird citizen surveillance platform. Log-transformed total sightings were counted as instances of unique data-user-geolocations over duration of study period (7/10/2005 - 30/3/2024).



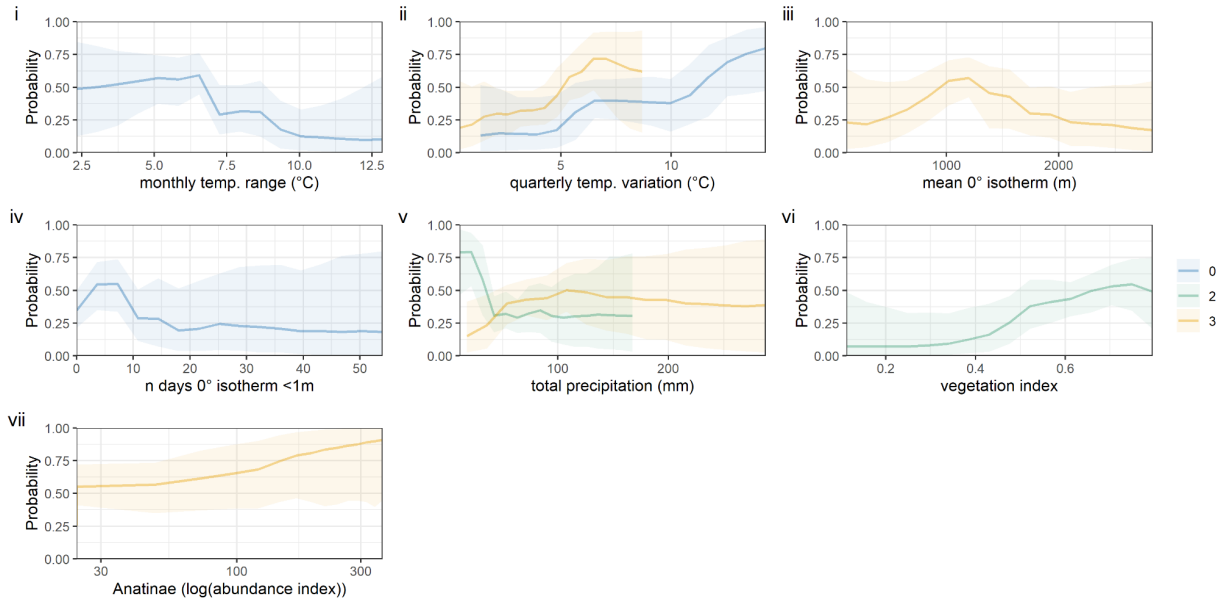
Supplemental Figure S2. Full partial dependence of BART model trained on dataset A, Q1.
Partial dependence associated with all covariates in final BART model of Q1 (Jan - Mar) fitted to geospatial HPAI data for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



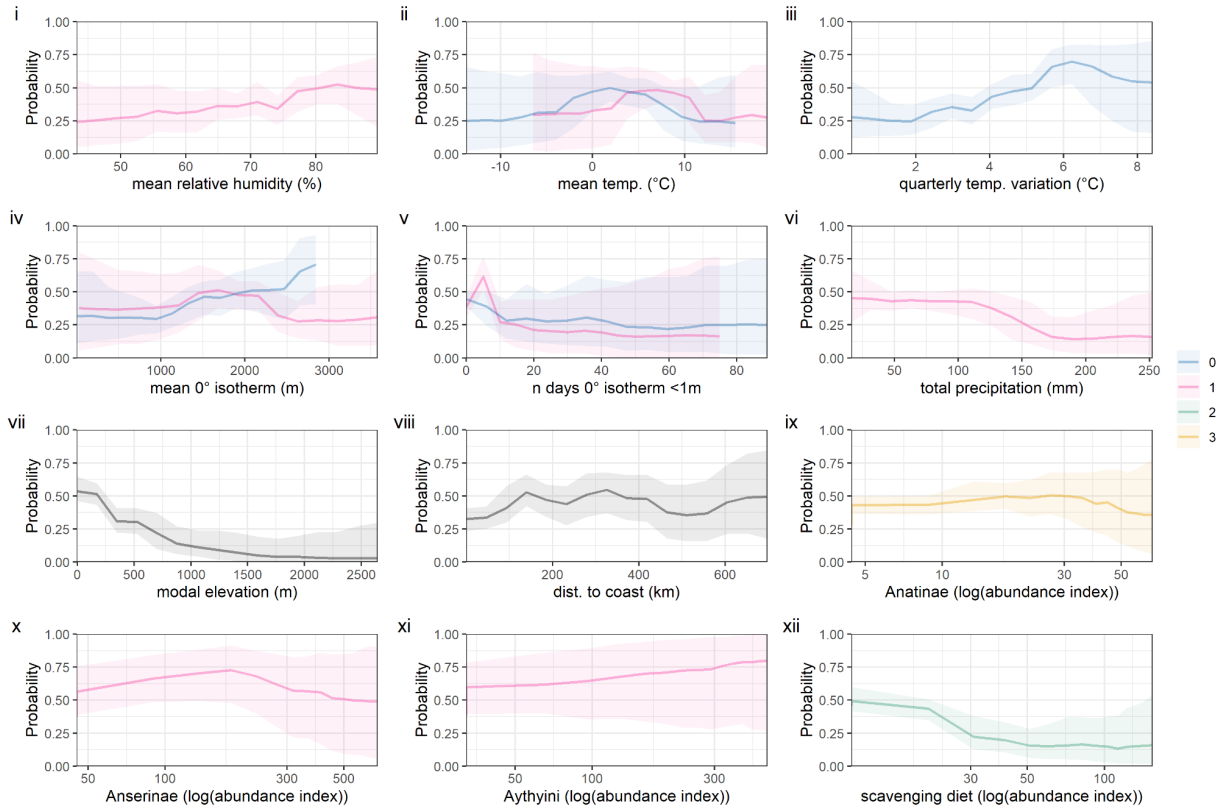
Supplemental Figure S3. Full partial dependence of BART model trained on dataset A, Q2.
Partial dependence associated with all covariates in final BART model of Q2 (Apr - Jun) fitted to geospatial HPAI data for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



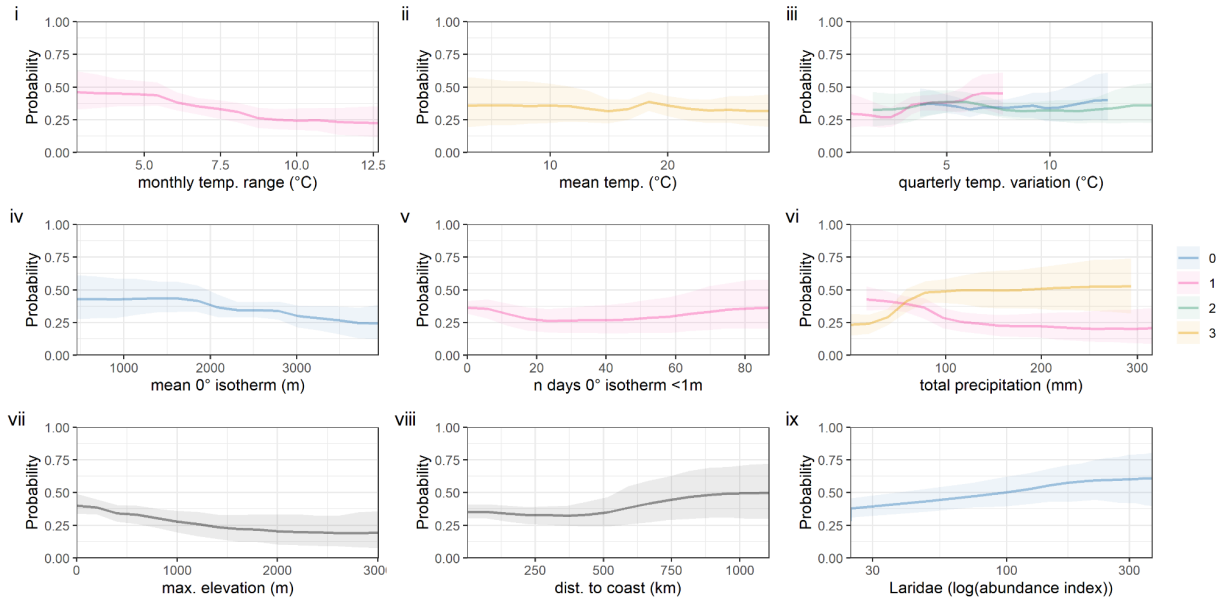
Supplemental Figure S4. Full partial dependence of BART model trained on dataset A, Q3.
Partial dependence associated with all covariates in final BART model of Q3 (Jul - Sep) fitted to geospatial HPAI data for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



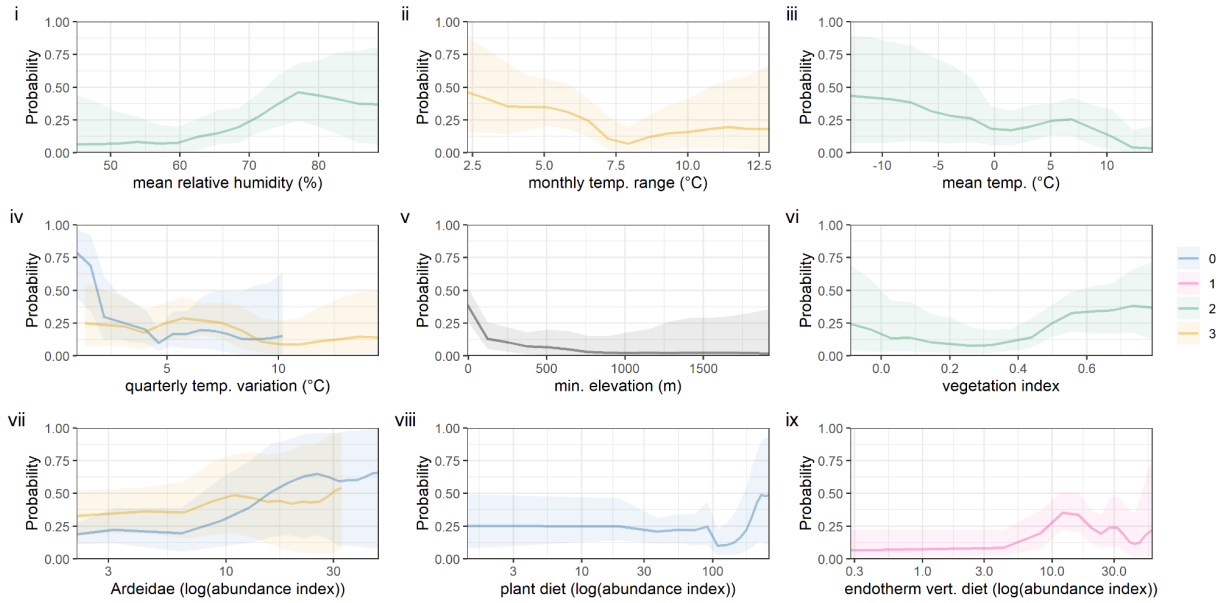
Supplemental Figure S5. Full partial dependence of BART model trained on dataset A, Q4.
Partial dependence associated with all covariates in final BART model of Q4 (Oct - Dec) fitted to geospatial HPAI data for period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



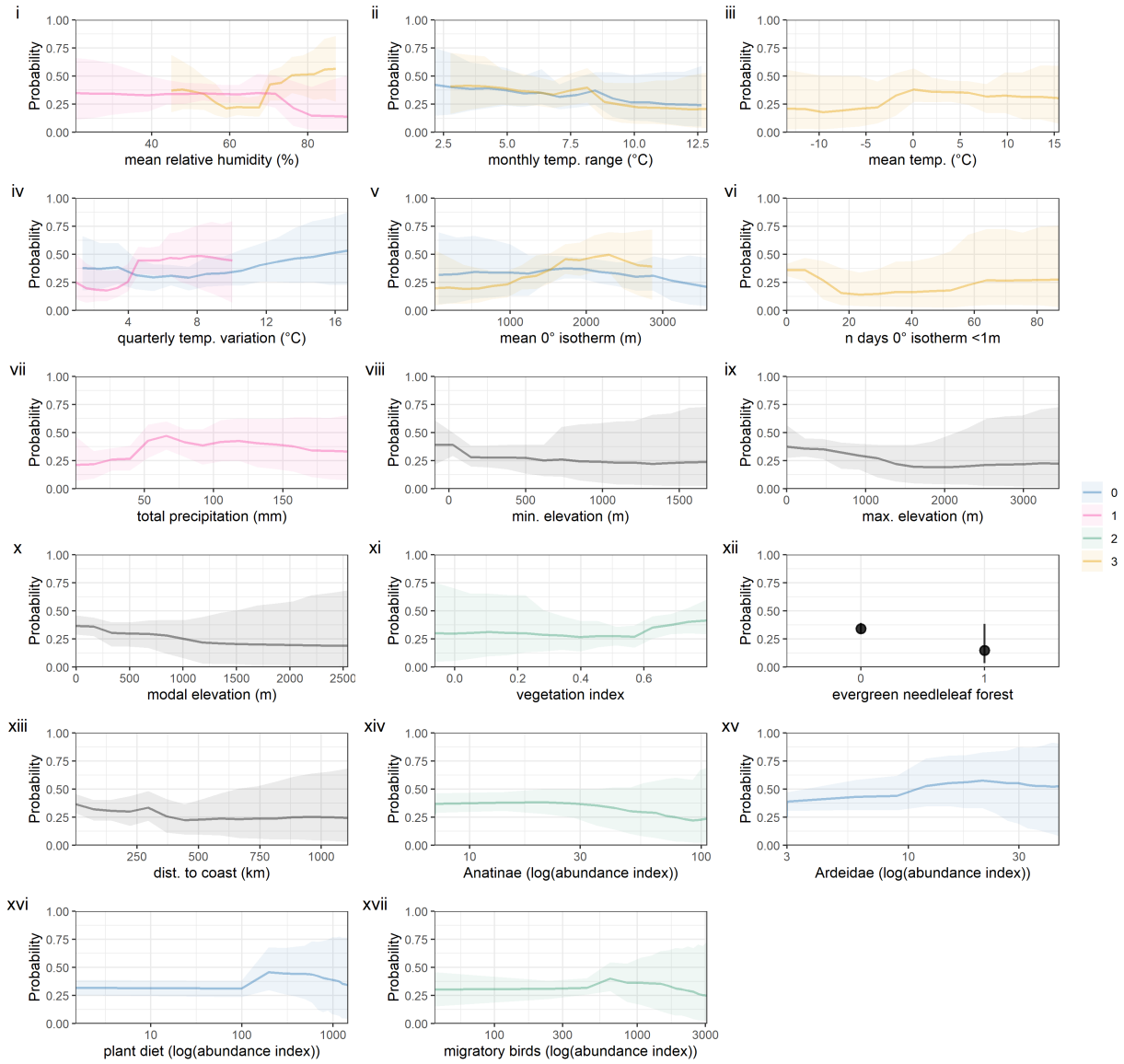
Supplemental Figure S6. Full partial dependence of BART model trained on dataset B, Q1.
Partial dependence associated with all covariates in final BART model of Q2 (Apr - Jun) fitted to geospatial H5N1 HPAI data for period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



Supplemental Figure S7. Full partial dependence of BART model trained on dataset B, Q2.
Partial dependence associated with all covariates in final BART model of Q2 (Apr - Jun) fitted to geospatial H5N1 HPAI data from 2021 onwards. Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



Supplemental Figure S8. Full partial dependence of BART model trained on dataset B, Q3.
Partial dependence associated with all covariates in final BART model of Q3 (Jul - Sep) fitted to geospatial H5N1 HPAI data for period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.



Supplemental Figure S9. Full partial dependence of BART model trained on dataset B, Q4.
Partial dependence associated with all covariates in final BART model of Q4 (Oct - Dec) fitted to geospatial H5N1 HPAI data for period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Y axis denotes marginal probability of HPAI presence, i.e., averaging out effects of all other covariates. For continuous covariates, solid lines denote median values while shaded areas denote uncertainty via 2.5th percentile and 97.5th percentile values over 1000 draws from the posterior tree space. For categorical covariates, points denote median values while error bars denote via 2.5th percentile and 97.5th percentile values. Colours denote calendar quarter delay of seasonally variant covariates increasing from 0 (predicted quarter) to 3 (three quarters prior to predicted quarter); black denotes non-seasonally variant covariates.

Supplemental Text S1. Rationalising taxonomic nomenclature

Taxonomic labelling was not consistent between our species-level data sources, meaning that certain species identified by eBird as being present in Europe did not appear in some of the other data sources, despite these sources theoretically being more comprehensive than the eBird Status and Trends data (which is limited to those species which are confidently modelled by eBird). To remedy this we performed a semi-automated rationalisation process using the AVONET database of bird species names (67). Each datapoint in the AVONET database corresponds to a single species and includes binomial names from three different resources (BirdLife, eBird, and BirdTree) as well as an alphanumeric ID string, allowing for identification of synonyms when comparing data from sources using differing taxonomies. For each species identified by eBird as being present in Europe (obtained by filtering by region within the eBird status and trends data) we used AVONET to generate a set of possible synonyms for that species. We matched eBird species abundance records of a species with species-level trait data from other datasets by first searching for that species' binomial name in that dataset, and then searching for its synonyms if the binomial name was missing. Where the synonyms were absent or where multiple synonyms for a species appeared in the trait data, we checked the data manually and used a combination of renaming and removal of problematic species to ensure that each species had exactly one match in each trait dataset, based on the Avibase online bird taxonomy database (95).

Our species-level trait data construction begins by loading in 708 species records from eBird Status and Trends. We then bring in the list of global species population size estimates from Callaghan *et al.* (70), and for each species in the eBird Status and Trends data attempt to assign a population size. The following 6 species in the eBird list could not be matched to any records in the population size estimate data:

- Northern Hawk Owl, *Surnia ulula*
- Iberian Green Woodpecker, *Picus sharpei*
- Western Subalpine Warbler, *Curruca iberiae*
- Kruper's Nuthatch, *Sitta krueperi*
- Amur Stonechat, *Saxicola stejnegeri*
- Eastern Black-eared Wheatear, *Oenanthe melanoleuca*

We removed *Surnia ulula*, *Sitta krueperi*, and *Saxicola stejnegeri* from our species list as we were unable to find possible matches in the population size estimate data. *Picus sharpei* and *Oenanthe melanoleuca* are classified in some sources as subspecies of other species found in eBird (*Picus viridis*, the European green woodpecker, and *Oenanthe hispanica*, the Western black-eared wheatear, respectively). In these cases it was necessary to remove both species in each pair from our species list because the total *Picus viridis* population size in Callaghan *et al.* effectively included both *Picus viridis* and *Picus hispanica*, meaning that keeping *Picus viridis* in the model would result in us effectively overestimating the numbers of *Picus viridis* present outside of *Picus hispanica*'s range. The same considerations motivated us to remove both *Oenanthe melanoleuca* and *Oenanthe hispanica*. The genus *Curruca* has only recently been recognised and does not appear in some sources and so we remove all of its members. After these removals we were left with 691 of the initial 708 species. Checking the CLOVER database with scientific names and pseudonyms from AVIBASE revealed that none of the species we removed were known avian influenza hosts.

After matching the species from eBird Status and Trends to population size estimates, we used the AVONET database to identify possible species synonyms. The following species could not be found in AVONET:

- Grey-headed Swamphen, *Porphyrio poliocephalus*
- Yellow-headed caracara, *Daptrius chimachima*
- Oriental cuckoo, *Cuculus optatus*

Consulting eBird suggested that *Porphyrio poliocephalus* is found only in small numbers on the outskirts of Europe (96), and so we removed it from our species list. We replaced *Daptrius chimachima* with *Milvago chimachima*, which is synonymous with *Daptrius chimachima*(97). Although *Cuculus optatus* was absent from AVONET, it was present in all of our other databases meaning we were able to match it to species-level traits without using synonyms, and so we kept it in our species list without renaming. This left us with 690 of our initial 708 species.

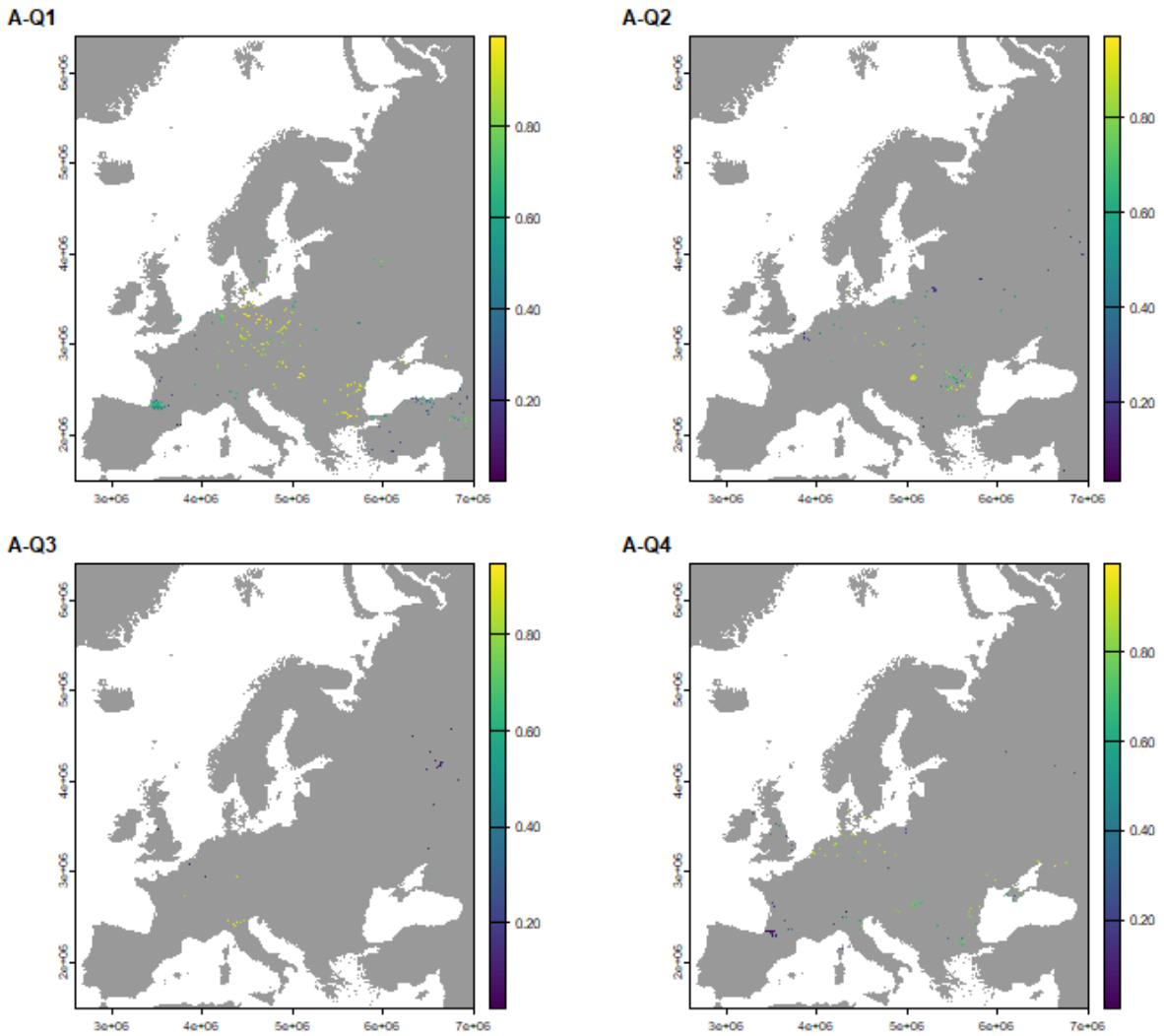
The binomial names for three of the known host species in CLOVER could not be found in AVONET, motivating the following changes to binomial names in CLOVER based on species entries in Avibase (95):

- Eurasian jackdaw, *Coloeus monedula*, replaced with *Corvus monedula;*
- Mongolian gull, *Larus mongolicus*, replaced with *Larus cachinnans;*
- Little cuckoo, *Piaya minuta*, replaced with *Coccycua minuta.*

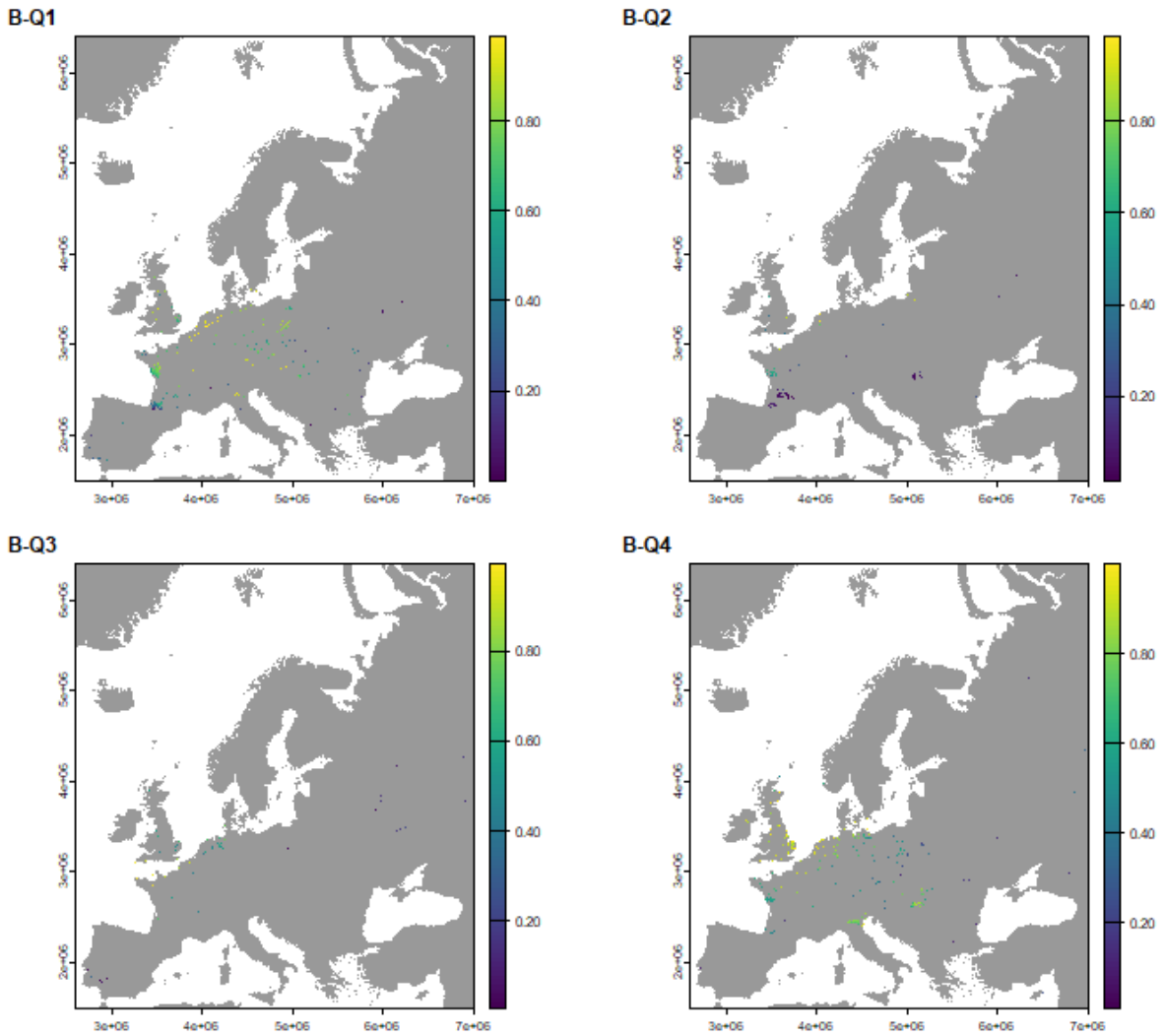
We replaced the binomial names of the following four species in EltonTraits to ensure that every species from EltonTraits could be matched with a single species from eBird Status and Trends:

- Long-tailed pipit, *Anthus longicaudatus*, replaced with *Anthus vaalensis*
- Iquitos gnatcatcher, *Polioptila clementsi*, replaced with *Polioptila guianensis*
- Bluntschli's vanga, *Hypositta perdita*, replaced with *Oxylabes madagascariensis*
- Vietnamese pheasant, *Lophura hatinhensis*, replaced with *Lophura edwardsi*
- Yellow warbler, *Setophaga petechia*, replaced with *Dendroica petechia*

In the IUCN trait data we changed the binomial name of the yellow-throated parrotbill from *Suthora webbiana* to *Sinosuthora webbiana*.



Supplemental Figure S10. Domestic bird cases and model-predicted wild bird risk, period A.
Maps of the study area with cells containing HPAI in domestic birds highlighted and coloured by the predicted probability from BART models over period A (7/10/2005 to 31/8/2021, before ongoing H5N1 outbreak). Panels denote individual quarters.



Supplemental Figure S11. Domestic bird cases and model-predicted wild bird risk, period B.
Maps of the study area with cells containing HPAI in domestic birds highlighted and coloured by the predicted probability from BART models over period B (1/9/2021 to 30/3/2024, ongoing H5N1 outbreak). Panels denote individual quarters.