

## Article

# Wild Bird Densities and Landscape Variables Predict Spatial Patterns in HPAI Outbreak Risk across The Netherlands

Janneke Schreuder <sup>1,2,†</sup> , Henrik J. de Knecht <sup>2</sup> , Francisca C. Velkers <sup>1,\*</sup> , Armin R. W. Elbers <sup>3</sup> , Julia Stahl <sup>4</sup>, Roy Slaterus <sup>4</sup>, J. Arjan Stegeman <sup>1</sup>  and Willem F. de Boer <sup>2</sup> 

<sup>1</sup> Department of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands; janneke.schreuder@wur.nl (J.S.); j.a.stegeman@uu.nl (J.A.S.)

<sup>2</sup> Wildlife Ecology and Conservation Group, Wageningen University & Research, 6708 PB Wageningen, The Netherlands; henjo.deknecht@wur.nl (H.J.d.K.); fred.deboer@wur.nl (W.F.d.B.)

<sup>3</sup> Department of Epidemiology, Bioinformatics and Animal Models, Wageningen Bioveterinary Research, 8221 RA Lelystad, The Netherlands; armin.elbers@wur.nl

<sup>4</sup> Sovon, Dutch Centre for Field Ornithology, 6525 ED Nijmegen, The Netherlands; julia.stahl@sovon.nl (J.S.); roy.slaterus@sovon.nl (R.S.)

\* Correspondence: f.c.velkers@uu.nl; Tel.: +31-30-253-1248

† Current affiliation: Wageningen Livestock Research, Wageningen University & Research, 6708 WD Wageningen, The Netherlands.

**Abstract:** Highly pathogenic avian influenza viruses' (HPAIVs) transmission from wild birds to poultry occurs globally, threatening animal and public health. To predict the HPAI outbreak risk in relation to wild bird densities and land cover variables, we performed a case-control study of 26 HPAI outbreaks (cases) on Dutch poultry farms, each matched with four comparable controls. We trained machine learning classifiers to predict outbreak risk with predictors analyzed at different spatial scales. Of the 20 best explaining predictors, 17 consisted of densities of water-associated bird species, 2 of birds of prey, and 1 represented the surrounding landscape, i.e., agricultural cover. The spatial distribution of mallard (*Anas platyrhynchos*) contributed most to risk prediction, followed by mute swan (*Cygnus olor*), common kestrel (*Falco tinnunculus*) and brant goose (*Branta bernicla*). The model successfully distinguished cases from controls, with an area under the receiver operating characteristic curve of 0.92, indicating accurate prediction of HPAI outbreak risk despite the limited numbers of cases. Different classification algorithms led to similar predictions, demonstrating robustness of the risk maps. These analyses and risk maps facilitate insights into the role of wild bird species and support prioritization of areas for surveillance, biosecurity measures and establishments of new poultry farms to reduce HPAI outbreak risks.

**Keywords:** avian influenza; influenza A virus; highly pathogenic avian influenza; poultry; disease outbreaks; surveillance; wild-domestic interface; spatial modelling; random forest



**Citation:** Schreuder, J.; de Knecht, H.J.; Velkers, F.C.; Elbers, A.R.W.; Stahl, J.; Slaterus, R.; Stegeman, J.A.; de Boer, W.F. Wild Bird Densities and Landscape Variables Predict Spatial Patterns in HPAI Outbreak Risk across The Netherlands. *Pathogens* **2022**, *11*, 549. <https://doi.org/10.3390/pathogens11050549>

Academic Editor: El-Sayed Mohammed Abdel-Whab

Received: 30 March 2022

Accepted: 4 May 2022

Published: 6 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Highly pathogenic avian influenza A viruses (HPAIVs) of clade 2.3.4.4 have spread globally, causing massive outbreaks in commercial poultry farms [1,2], especially in recent years. The migratory movements of wild birds have been shown to play an important role in the global spread of HPAIVs, as spatial and temporal patterns of outbreaks coincided with migratory flyways and the timing of autumn migration [1,3]. Some HPAIV subtypes, e.g., H5N8 (in 2014 and 2016) and H5N6 (in 2017), have caused outbreaks in Europe on commercial poultry farms, as well as massive mortality in wild birds [4–9]. HPAI outbreaks were reported in more than 30 EU countries, mostly with subtypes H5N8 and H5N5, H5N1 and H5N3 in 2020–2021, and dominated by H5N1 from 2021–2022, with detections in four geographical regions, i.e., Europe, Africa, the Americas and Asia [2,10–14]. Most of these cases were reported in wild birds, primarily in waterbirds such as the barnacle goose

(*Branta leucopsis*), graylag goose (*Anser anser*), mute swan (*Cygnus olor*), common buzzard (*Buteo buteo*) and several species of ducks (*Anatidae*) and gulls (*Laridae*) [15,16].

These recurrent and massive outbreaks of HPAI underline the need for better prediction of HPAI risk areas to reduce outbreak risk and take appropriate measures. HPAI outbreaks on poultry farms are spatially associated with the proximity of waterbodies or the presence of wild birds [17–20]. For example, the magnitude of the increase in densities of HPAI high risk bird species nearby HPAIV H5N8 infected poultry farms in the Dutch wetlands in the fall and winter of 2016–2017 was significantly higher compared to the non-infected farms in non-water-rich areas [20]. This was especially true for the Eurasian wigeon (*Mareca penelope*), which was one of the species with massive mortality due to HPAIV in 2016–2017 [6]. In addition, dead wild birds found at sites in the vicinity of the HPAIV infected poultry farms had phylogenetically related viruses, which may suggest that HPAIV on these farms originated from the infected wild birds [4,21]. HPAIV introduction into poultry farms most likely results from indirect contact with wild birds, and it is hypothesized that the virus enters the poultry farm via vectors or fomites contaminated with wild bird feces [21–25]. This suggests that wild bird presence and density can be used as predictor in identifying HPAI high risk areas.

Previously, disease distribution models showed that both land cover, particularly the presence of wetlands, and environmental variables could be used to successfully predict HPAI outbreak risk [17–19]. However, landscape variables are likely to be merely a proxy for the presence and density of wild birds. We hypothesize that including wild bird densities in HPAI outbreak risk modelling, rather than only landscape variables, increases the accuracy of the prediction of HPAI outbreaks.

Several factors complicate the prediction and understanding of the spatial patterns in HPAI outbreak risk. First, associations between outbreak risk and spatial predictors may be spatial scale dependent. Second, machine learning algorithms can be used as a ‘black box’, where either the underlying patterns are not considered, or associations can even be unrealistic and non-interpretable, such as when predicted HPAI risk decreases with increasing wild bird density. Lastly, spatial predictors often show high levels of multicollinearity, for instance, the densities of many waterbird species are often positively correlated, which complicates statistical routines.

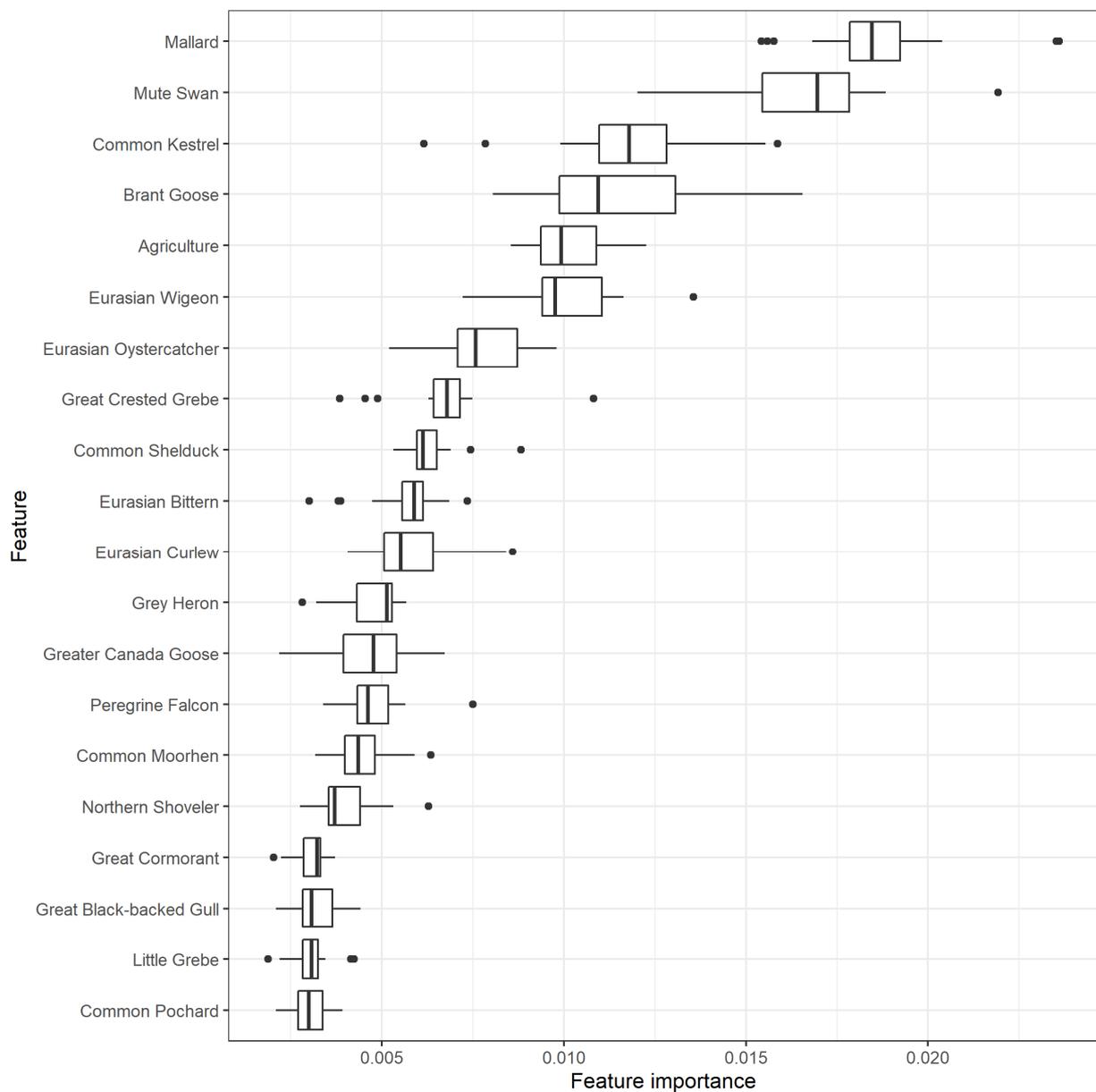
For a water(fowl)-rich country, such as the Netherlands, it is especially challenging to determine which areas are more at risk than others. The aim of this study was to identify wild bird species and land cover variables that are associated with HPAI outbreak risk on poultry farms to study the spatial pattern of HPAI outbreak risk across the Netherlands and generate HPAI risk maps. For this purpose, we used a highly detailed dataset on wild bird densities, obtained from a structured and elaborate system of bird counts across the whole country. Furthermore, we used machine learning classifiers to predict outbreak risk at different spatial scales. In the model building process and analyses we explicitly considered the scale dependency between predictors and HPAI outbreak risk, the correlation structure between predictors, and only allowed monotonously increasing relationships between the wild bird densities and HPAI outbreak risk.

This study contributes to a better understanding of the factors determining HPAI outbreak risk, and a better prediction of spatial patterns. This will help support prioritization of areas for surveillance, biosecurity measures and establishment of new poultry farms, to reduce HPAI outbreak risks.

## 2. Results

A total of 56 features (all 5 land cover features and 51 bird species), from the initial 5 land cover and 58 high risk HPAI bird species, passed the filtering stage. Twenty features had higher-than-average feature importance in the random forest classifier (Figure 1). All but one of these were bird densities, with agricultural cover being the only landscape variable with higher-than-average importance. All of the most important features showed

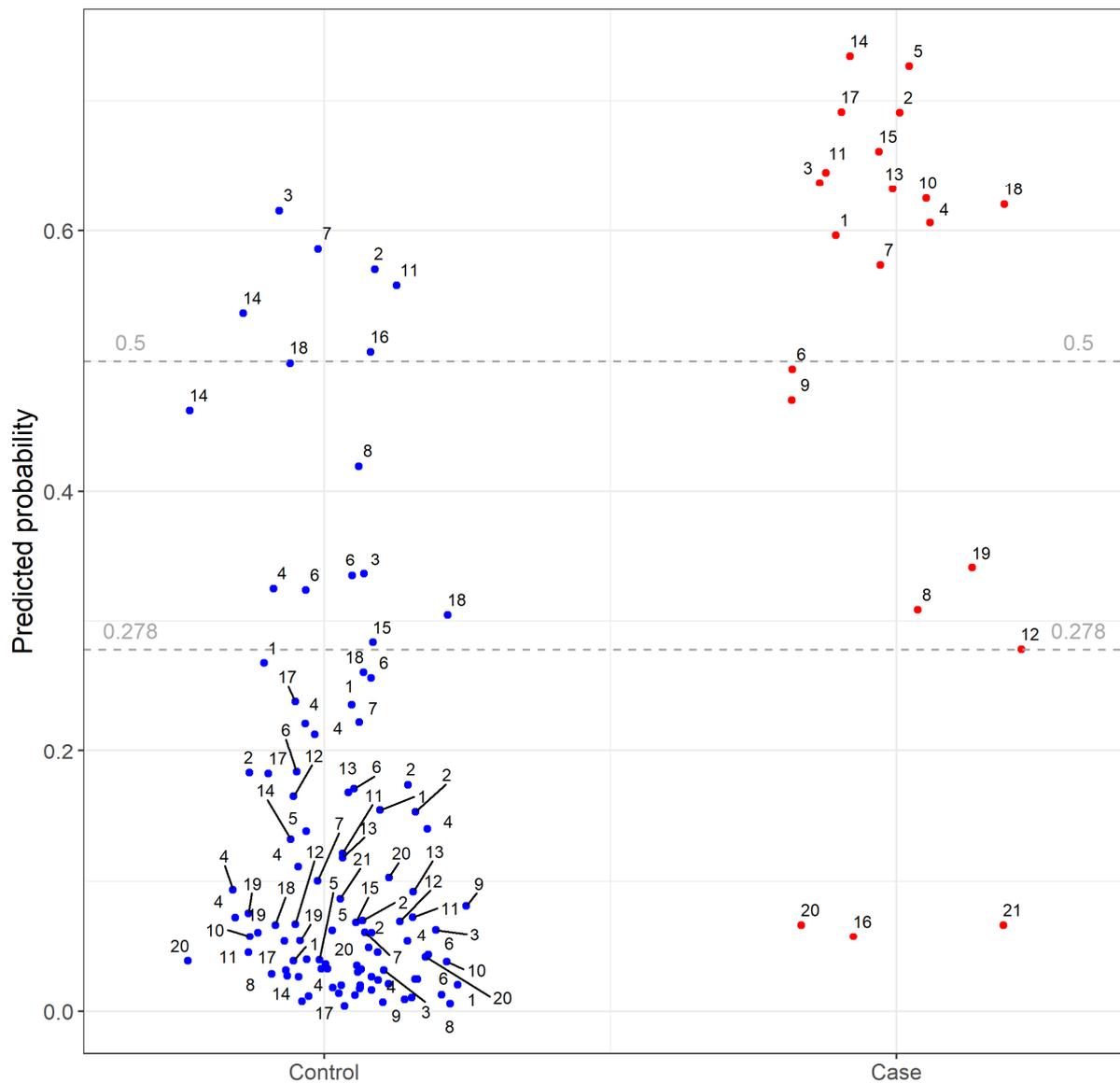
the strongest univariate associations to HPAI outbreak risk at larger scales of spatial smoothing (2.5 and 5 km).



**Figure 1.** Feature importance of the 20 most important scale-aggregated predictors (Table S1) in the random forest using leave-one-group-out cross validation (LOGO-CV). The boxplots indicate the variation in feature importance across the different LOGO-CV iterations.

According to the random forest classifier, the spatial distribution of mallard (*Anas platyrhynchos*) contributed most to the prediction of HPAI risk, followed by mute swan (*Cygnus olor*), common kestrel (*Falco tinnunculus*) and brant goose (*Branta bernicla*).

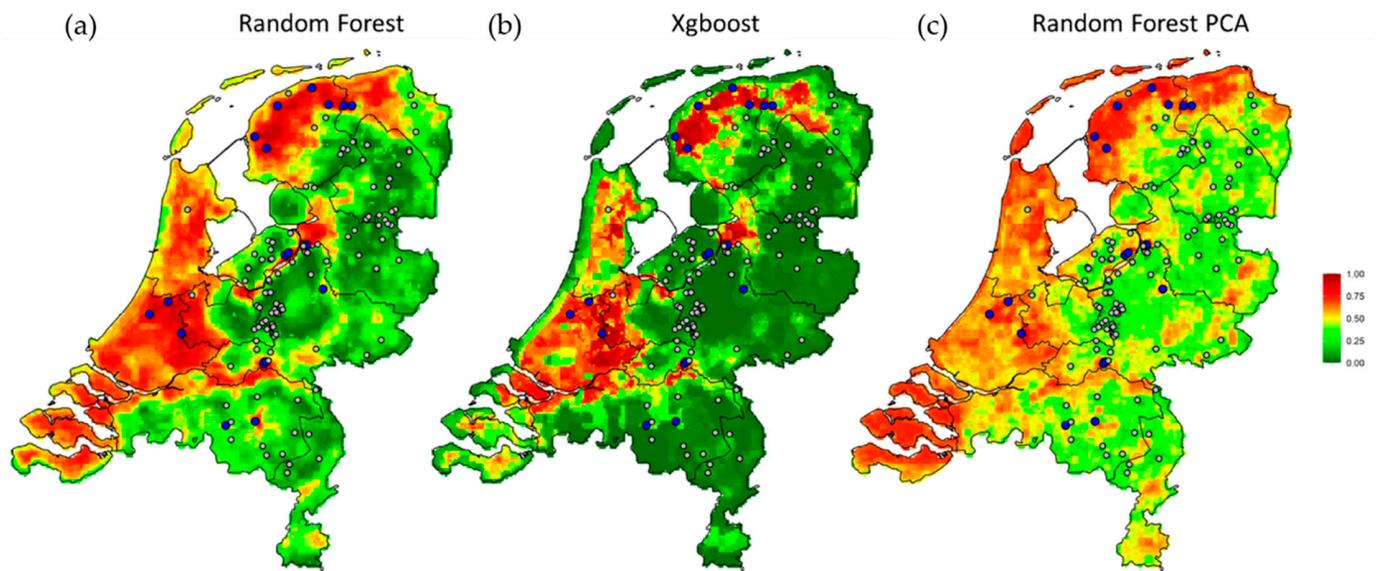
The random forest with all multiscale predictors showed high predictability: an overall accuracy of 0.86, sensitivity (recall) of 0.88 and an area under the receiver operating characteristic curve (AUC-ROC) of 0.92 (Figure 2; Table 1). The gradient boosted decision tree generated even better predictions (accuracy = 0.94; sensitivity = 0.88, AUC-ROC = 0.94), with the distribution of the mallard as, by far, the most important predictor of HPAI outbreak risk (Figure S1).



**Figure 2.** Results of final leave-one-group-out random forest (LoGo random forest). Each dot represents an individual highly pathogenic avian influenza case farm (red) or control poultry farm (blue). The number labels indicate case-ID of each farm. The predicted probability is given for each case and control farms within a set, after training of the LoGo random forest on the remaining cases and controls. The horizontal lines represent different cut-off values for test performance analyses of which 0.278 is the weighted F1 score.

**Table 1.** Classification metrics for predicted highly pathogenic avian influenza risks (Figure 3) for the different algorithms: random forest (RF), gradient boosted decision trees (GBT), and random forest on a PCA transformed feature set (RF-PCA). All classification metrics except for the AUC-ROC are for a cut-off threshold that maximizes the weighted F1 score, which was 0.278 for this model. AUC-ROC is the area under the receiver operating characteristic curve, a threshold-independent performance measure.

Algorithm	Accuracy	Recall/Sensitivity	AUC-ROC
RF	0.86	0.88	0.92
GBT	0.94	0.88	0.94
RF-PCA	0.87	0.81	0.88



**Figure 3.** Mean farm outbreak probability of highly pathogenic avian influenza (HPAI) across all 1 km × 1 km grid-cells in the Netherlands for the three different algorithms; (a) the final leave-one-group-out random forest model (Random Forest); (b) the gradient boosted decision trees (Xgboost); and (c) the random forest on a PCA transformed feature set (Random Forest PCA). The prediction of HPAI risk ranges between 0 (low, dark green) and 1 (high, dark red). Locations of poultry farms with HPAI outbreaks (i.e., cases, blue) and control farms (grey) are shown.

The predicted spatial pattern of HPAI outbreak risk across the Netherlands (Figure 3) shows generally a high risk in the western and northern part of the Netherlands, as well as around the larger rivers in the center of the country (e.g., the Rhine, Waal, IJssel and Meuse rivers). The eastern and southern parts of the Netherlands generally have a lower risk of HPAI outbreak. A negative correlation ( $-53.9\%$ ) between the standard deviation of the predicted infection probabilities (on a logit scale) and the average predicted farm outbreak probability was found, indicating that variation in the areas with high predicted HPAI risk was lower than in the areas where the infection risk was predicted to be low.

The correspondence between the predicted risk maps (Figure 3) produced by different algorithms (the two random forests algorithms, with and without PCA-transformed feature set and the boosted decision trees with monotonicity constraints) indicates that the patterns captured in the risk maps are robust. Namely, removing the possible confounding effects of multicollinearity between predictors or enforcing positive monotonicity in the association between bird densities and HPAI outbreak risk did not lead to major changes in the predicted risk maps (correlation between random forest and gradient boosted decision trees:  $84.3\%$ ; correlation between a random forest with all multiscale predictors and after PCA transformation:  $79.5\%$ ).

### 3. Discussion

We showed that a model using wild bird species' densities can accurately predict HPAI risk areas for poultry farms in the Netherlands. Seventeen waterbird and two raptor species were most strongly associated with the HPAI outbreak risks.

The mallard and mute swan had the highest feature importance, and have both been found infected with HPAIVs, as described in several studies [4,5,9,21,26]. Other species selected by the model, e.g., Eurasian wigeon, peregrine falcon (*Falco peregrinus*) and great black-backed gulls (*Larus marinus*), were found to have high mortality rates during the H5N8 epidemic in 2016 [6], and wild birds of the family of Anatidae and of the order of Charadriiformes are often affected and may also play a role as important reservoir species for HPAIVs [27]. All bird species of which the densities contributed most to the prediction of HPAI outbreak risk for poultry farms were also among the confirmed HPAI wild bird

cases in the EU between 2020–2022 [15,16], with the mute swan being one of the most frequently reported species in 2020–2021 [15] and 2021–2022 [16]. In addition to the mute swan, the top three affected species included the barnacle goose (*Branta leucopsis*) and the graylag goose (*Anser anser*) [15,16]. These were not among the species mostly associated with HPAI outbreak risk for farms, suggesting that the species with the highest feature importance are not always similar to the wild bird species with the highest mortality rates from HPAIV infection (Table S1).

Our model outputs indicate which wild bird species' densities were associated with the risk of infection for poultry farms, based on prior cases and a detailed wild bird count dataset covering the whole of the Netherlands, and was not based on reported numbers of infected or dead wild birds. The wild bird species indicated by the model may be relevant for HPAIV transmission, and, therefore, it may be important to consider these species as well when taking preventive measures around farms and for future studies. However, we have to be careful when drawing conclusions on the exact roles of specific wild bird species in the epidemiological processes at the wild bird/domestic bird interface, and on that of other bird species not included in this study. HPAIV infections among populations of wild bird species depend on a complex multi-species system, influenced by ecosystem properties, bird species diversity and community structure, the specific circulating HPAIV strain(s), and the clinical impact it has among the different hosts species [27–29]. This study does not indicate how many birds of the identified wild bird species were infected, or to what extent their presence contributed to disseminating the virus in the farms' surroundings. For example, not all of the birds that were found often during passive surveillance between 2020–2022, such as the barnacle goose (*Branta leucopsis*) and graylag goose (*Anser anser*) [15,16], were among the birds with the highest feature importance in the model. These species may have made a limited additional contribution to the model because of collinearity with other species, although such effects were limited by constraining the hyperparameters in the random forest model. Another explanation could be that birds that die quickly from HPAIV infection have a more limited role in disseminating the virus in the environment. In addition, it can suggest that the dynamics of species contributing most to HPAI outbreak risks may change over time, making it important to keep updating such models frequently with the newest data. Furthermore, the species selected by the model can include both migratory bird species (e.g., Eurasian wigeon), which likely play a role in long distance dispersal, and species that are less migratory (e.g., mute swan and mallard), which could act as local amplifiers, or bridge species [29–31]. The relative roles of different migratory, sedentary and synanthropic bird species in HPAIV's introduction into poultry farms therefore remains to be elucidated.

The predicted higher HPAIV infection risk in the western and northern part of the Netherlands, and around larger rivers, is in the lowest parts of the Netherlands with soils containing more clay and peat, i.e., characteristic habitats for the more water-dependent bird species. In contrast, the higher and dryer eastern and southern parts of the Netherlands with generally sandy soils, have a lower risk of HPAI outbreak. Nevertheless, in contrast to previous studies [17–19,32,33], only one land cover variable, i.e., cover by agriculture, was selected in the final model, indicating that densities of wild bird species were better predictors of the HPAIV infection risk in our study. Previous studies mostly tried to explain the variation in the occurrence of HPAIV introductions with environmental variables, such as distance to waterways and vegetation index [18,19,33]. Others also included surveillance or telemetry data to track the movements of a selection of wild bird species in a certain time period [17,29,32,34–36], but most studies did not include detailed count data on wild bird species distribution to analyze HPAIV introductions on poultry farms. Environmental variables can be considered as a proxy for habitat suitability for wild birds, and were, in our study, less suitable predictors than the densities of the actual bird species. This does not mean that land cover data could not be of great value. Such data have proven valuable in many previous spatiotemporal analyses. For example, it has been shown in many studies that the presence of wetlands is vital as they are an important habitat for many waterfowl

species [32,33,35]. In the current study, the original land cover classes were aggregated into five major classes, which decreased the resolution of land cover classification in the analyses, and may have reduced sensitivity of these variables for the prediction of farm outbreak probability. Therefore, especially in regions where detailed quantitative wild bird density data are not available, land cover data, and potentially also other ecological or climatic, bird habitat-related features that can be more easily gathered across large areas, could still be a suitable proxy for predicting HPAI risk.

Besides the spatial distribution of wild birds, seasonality and the arrival of migratory birds also play a role in the prediction of HPAI outbreak risk [20,30,36–38]. For example, Velkers et al. (2020) found that the timing of peak densities of Anatidae species observed around Dutch farms coincided with the timing of outbreaks [20]. In addition, reservoir host movement and behavioral states (e.g., resting, foraging, long-distance migration) can affect spatiotemporal overlap with poultry facilities and hence probabilities of HPAIV spillover, as demonstrated by the telemetry data of the blue-winged teal (*Spatula discors*) in the United States [36]. Hence, the spatiotemporal relationships between outbreaks on poultry farms and HPAI wild bird detections represent complex dynamics. In the current study, we only used long-term averages of bird count data, collected at set moments each winter between 2012/2013–2014/2015. Our model could be further improved by including temporal variation in local differences in densities of bird species, and by adding new HPAI cases on poultry farms, provided that these data are available, to further train and validate the model.

The scale dependency of the different predictors was incorporated in all fitted algorithms. This is in line with ecological studies assessing the influence of environmental context, varied over different scales, on the analysis and prediction of habitat selection [39]. This is important, as local dispersal patterns and ecology of bird species differ. Diving ducks, e.g., tufted ducks, are mainly found on large open waterbodies, often at considerable distance from farms, and have relatively few movements over land between foraging and roosting sites. In contrast, Eurasian wigeons and mallards forage on grass- and agricultural lands, and are found more closely to farms [6,20], which is in line with our results.

We realize that our dataset was limited, with 26 confirmed HPAI cases on poultry farms. Although some farms had multiple outbreaks over the years, these were all new introductions, and, thus, independent of one another. Furthermore, we tuned the random forest analysis in such a way as to minimize the risk of overfitting and used a cross-validation approach for the random forest, testing its robustness despite the limited dataset. Moreover, the potential problem of multicollinear predictors was successfully addressed by using the orthogonally transformed predictor-set through a PCA, and the gradient boosted decision tree algorithm successfully constrained the model to only monotonously increasing associations between bird densities and outbreak risk. The results of the latter two algorithms supported the random forest analysis, and the boosted decision tree even slightly outperformed the random forest analyses (Table 1), indicating that the spatial differences in mallard densities was the most important predictor. The mallard is among the most studied species with regard to LPAIV and HPAIV infection prevalence, proximity around farms and outdoor ranges [22] and movement patterns in relation to wetland and agricultural fields, as previously reviewed [40]. Combined with the current study, such knowledge can contribute to improving risk predictions and taking measures to reduce the risk of AIV introduction on poultry farms.

In conclusion, we show that spatial variation in HPAI outbreak risk in the Netherlands was accurately predicted based on wild bird density data, rather than on only land cover variables. The spatial distributions of several waterbird species were important contributors to model the HPAI outbreak risk. New HPAI outbreaks can be used to validate and improve the risk map, but already in its current form, areas classified as high risk for HPAIV introduction on poultry farms should be considered as important targets for surveillance, preventive measures against HPAIV introduction, and may assist in decision making on the locations for new poultry farms. The described modelling approach allows for inclusion of

the best predictors based on the available data, which may include land cover variables in addition to bird data, depending on the local situation. Identification of high risk areas for development of country- or region-specific control programs would be a proactive strategy to combat the global threat of these recurring HPAI outbreaks.

#### 4. Materials and Methods

##### 4.1. Study Design

###### 4.1.1. Case Farms

A case control study was performed retrospectively, using all 26 diagnosed HPAIV H5N8, H5N6 and H5N1 infections on poultry farms in the Netherlands between the autumn–winter periods of 2014/2015, 2016/2017, 2017/2018 and 2020/2021. Ten outbreaks were on layer farms, eight on Pekin (meat production) duck farms, five on broiler breeder farms, two on broiler farms and one on a turkey (meat production) farm. Some farms were affected repeatedly in different years (Table 2). These multiple outbreaks occurred in different years and were all primary introductions, i.e., assumed to be as a result of wild bird to poultry transmission and not secondary between-farm spread, based on the genetic analyses of the viruses and epidemiological investigations [4,5,20,21,26].

**Table 2.** Overview of highly pathogenic avian influenza (HPAI) cases in the Netherlands on individual farms (ID 1 to 21) with confirmed HPAIV infection between 2014–2021. Poultry type indicates the type of farm that was affected. On poultry farms 1, 4 and 6 multiple HPAI outbreaks were diagnosed between 2014–2021.

Case-ID	Poultry Type	2014-H5N8	2016-H5N8	2017-H5N6	2018-H5N6	2020-H5N8 <sup>1</sup>
1	Layer	x	x			x
2	Layer	x				
3	Layer	x				
4	Pekin Duck	x	x		x	
5	Broiler Breeder	x				
6	Pekin Duck		x	x		
7	Pekin Duck		x			
8	Pekin Duck		x			
9	Layer		x			
10	Layer		x			
11	Broiler Breeder		x			
12	Broiler Breeder				x	
13	Broiler Breeder					x <sup>1</sup>
14	Layer					x
15	Layer					x
16	Pekin Duck					x
17	Broiler					x
18	Broiler					x
19	Broiler Breeder					x
20	Turkey					x
21	Layer					x

<sup>1</sup> The broiler breeder case in 2020–2021 was diagnosed with HPAIV H5N1. All other HPAI farms between October 2020 and February 2021 were confirmed with HPAIV H5N8.

###### 4.1.2. Control Farms

All farms that were not confirmed HPAI cases can be considered free of HPAIV infection, as a strict reporting system obligates farmers to notify any signs of HPAIV infection, and regular serological testing is executed on all poultry farms in the Netherlands to confirm absence of subclinical infection [4,5,20,21,26]. We randomly selected four unique uninfected poultry farms from the Netherlands Food and Consumer Product Safety Authority (NVWA) database for each HPAI case farm in every year. Control farms were selected based on similar poultry type to the infected case farm (i.e., layer, Pekin duck, broiler breeder, broiler or turkey farm), registration as an active poultry farm in the same

year as the outbreak year in the case farm and located within a 100 km radius of the case farm. Farms that had been a case in any of the years were excluded for control selection and controls could only be selected for one case farm.

#### 4.2. Selection of Wild Bird Density Data

We reviewed the literature for wild bird associations with HPAIV infection from the Netherlands, but also from other countries [1,5,6,41–43], to compile a list with species that had a known association with HPAIV infection (Table S1). Bird species taxonomically close to the species that had been associated with HPAIV infection were also included. Bird species that were rare (<500 individuals in the winter counts across the Netherlands), had a small geographical range (e.g., only present in the Wadden Sea) or only present in summer months, were excluded to prevent spurious negative associations.

#### 4.3. Bird Count Data

Bird count data collected by Sovon, the Dutch Centre for Field Ornithology (Nijmegen, the Netherlands), and published in the *Bird atlas for the Netherlands* [44], were used for the analyses. The *Bird atlas for the Netherlands* was compiled through, among other things, structured bird counts across the whole of the Netherlands by (largely voluntary) observers in three winter seasons from December–February 2012/2013–2014/2015. We only used winter bird density data, as HPAI outbreaks only occurred between November and March. In short, for organizing the field work and processing the data the whole of the Netherlands was divided into 5 km × 5 km squares. Each square was assigned to an observer who performed bird counts in pre-defined months. Sovon used the obtained data to construct maps with an estimated numbers of wild birds per species per square (Appendix A). For 54 of the 58 bird species of interest the estimates per square were available. For four bird species (i.e., mallard, Eurasian magpie, carrion crow and western jackdaw) the maps with estimates per square did not pass the internal review process, and we therefore transformed the relative abundance (density) maps with 1 km × 1 km resolution, into maps with the estimated number of the particular species per 5 km × 5 km square.

#### 4.4. Land Cover Data

Land cover data (LGN7) were available as a raster layer with 25 m resolution, from which we computed the fractional cover per 5 km × 5 km square, which aligned with the bird count data. We selected land cover classes based on relevance for the distribution of high risk HPAI bird species and HPAI risk on poultry farms [17,19]. These land cover classes were aggregated into five major classes (agriculture, freshwater systems, grasslands, swamps and peat area and saltwater systems) to reduce the number of classes to be used for further analyses (Table S2).

#### 4.5. Including Spatial Context

All downstream analyses were performed in R (version 4.1.0) [45]. We resampled all 5 km × 5 km squares to 1 km × 1 km resolution using nearest-neighbor allocation, so that all grid cells within each 5 km × 5 km square contained the same value. To include the influence of landscape context (i.e., bird density and land cover predictors) beyond the properties of the grid cell, we applied isotropic bivariate Gaussian smoothers to the grid layers, with bandwidths of 2.5, 5 and 10 km [46,47]. The unsmoothed raster layers contained essentially no information on the environmental context beyond the 5 km × 5 km square, and thus only contained site-specific information.

#### 4.6. Cross-Scale Analyses and Dimension Reduction

We thus had a large number of features (252: 5 land cover classes; densities of 58 bird species, each at four spatial scales, including the raw unsmoothed rasters), compared to a limited number of cases ( $n = 26$ ). To reduce the number of potential predictors, we first performed a series of univariate conditional logistic regressions (CLR, using the survival

package, [48]), with cases and matching controls stratified by case farm identifier. All features were standardized to zero-mean and unit-variance prior to analyses. We then applied feature selection by omitting features from further analyses when either (a) there was insufficient variation in the feature values over the case-control points, or (b) when the CLR indicated that feature values were negatively associated with HPAI outbreak risk. This latter filtering step only applied to the bird density features, in order to avoid spurious and uninterpretable negative association. All land cover classes and 51 bird species passed this step of filtering (Table S1).

Then, in order to achieve further dimension reduction, we merged the rasters measured at different scales within a bird species or land cover class into a single raster layer by computing the CLR slope-weighted sum of the individual scales. In addition, these scale-aggregated predictors were standardized to zero-mean and unit-variance. Moreover, CLR analyses showed that these predictors were all positively associated with HPAI outbreak risk.

#### 4.7. Model Training and Evaluation

We performed predictive analyses using a leave-one-group-out cross validation approach, where a group consists of a unique poultry farm on which HPAI occurred (i.e., the cases) together with its matched controls. We used the following machine learning algorithms and data preparation steps. We trained a random forest classifier (using the ranger package, [49]) on all multiscale predictors. Random forests generally deal well with large numbers of input variables, and when tuned well they can be robust against overfitting [50]. We tuned the following hyperparameters in a grid search: number of variables per tree (tuning range: 2–9; tuned hyperparameter value: 8); minimum node size (1–4; 2); and maximum tree depth (2–9; 8), all given a fixed forest size of 100,000 trees. In order to put more emphasis on high recall (i.e., sensitivity or true positive rate) at the expense of having lower precision, we evaluated predictive performance using a weighted F1 score, where recall was weighed four times higher than precision. We then trained the algorithm using the optimal hyperparameters, and quantified feature importance using the permutation variable importance approach [51]. Feature importance values were scaled by dividing them by the mean importance value, so that the average feature importance value was 1 (Table S1). Predictive performance was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve (a threshold independent metric that captures the trade-off between sensitivity and 1-specificity), as well as threshold-dependent metrics such as accuracy and sensitivity (recall). A prediction of the HPAI risk across the Netherlands was made by averaging the cross-validated predicted HPAI probability surfaces.

Furthermore, to account for the potential problem of the correlated feature set, we trained another random forest classifier following the procedure as outlined above, yet with a principal components analysis (PCA) as an intermediate step. The PCA orthogonally transformed the predictor feature set into a new feature set with statistically independent principal components, where we kept all dimensions, and thus kept all variation present in the original predictors.

Moreover, to prevent uninterpretable associations between bird densities and HPAI outbreak risk, we also trained a gradient boosted decision tree algorithm (XGBoost package), where we only allowed monotonously increasing associations between bird densities and HPAI outbreak risk. By comparing the predictions of these three different algorithms, we assessed the robustness of the random forest classifier on the original multiscale predictors by comparing its predictions to the ones generated by the other two algorithms.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pathogens11050549/s1>, Figure S1: Feature importance of 20 most important scale-aggregated predictors according to the gradient boosted tree with positive monotonicity constraints; Table S1: Overview of candidate variables in order of (Rank) decreasing feature importance, including a selection of high risk bird species for HPAIV infection and land cover

variables; Table S2: Aggregation of land cover classes to five major classes used in the random forest algorithms.

**Author Contributions:** Conceptualization, J.S. (Janneke Schreuder), F.C.V., J.A.S. and W.F.d.B.; Data curation, J.S. (Janneke Schreuder), H.J.d.K., F.C.V. and R.S.; Formal analysis, J.S. (Janneke Schreuder) and H.J.d.K.; Funding acquisition, A.R.W.E., J.S. (Julia Stahl), J.A.S. and W.F.d.B.; Investigation, J.S. (Janneke Schreuder), F.C.V. and R.S.; Methodology, J.S. (Janneke Schreuder), H.J.d.K., F.C.V., J.A.S. and W.F.d.B.; Project administration, F.C.V.; Resources, J.S. (Julia Stahl); Software, H.J.d.K.; Supervision, F.C.V., J.A.S. and W.F.d.B.; Validation, J.S. (Janneke Schreuder), F.C.V., J.S. (Julia Stahl), R.S. and W.F.d.B.; Visualization, J.S. (Janneke Schreuder) and H.J.d.K.; Writing—original draft, J.S. (Janneke Schreuder) and H.J.d.K.; Writing—review and editing, J.S. (Janneke Schreuder), H.J.d.K., F.C.V., A.R.W.E., J.S. (Julia Stahl), R.S., J.A.S. and W.F.d.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the 1Health4Food Public Private Partnership project: Fight Flu, grant number TKI-AF-15225.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank Jose Gonzales and Ronald Petie (Wageningen Bioveterinary Research, Lelystad, The Netherlands) for the valuable discussions and input on the manuscript. We are also grateful to Christian Kampichler (Sovon, Dutch Centre for Field Ornithology, Nijmegen, The Netherlands) for the help on the processing of the bird density data and to all bird watchers for collection of the bird counts.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A. Bird Count Data Collection and Analyses

Dutch bird count data collected by Sovon, the Dutch Centre for Field Ornithology (Nijmegen, The Netherlands), and published in the *Bird atlas for The Netherlands* [44] were used for the analyses. The *Bird atlas for The Netherlands* was compiled through a citizen science approach of structured bird counts across the whole of The Netherlands by (largely voluntary) observers in three winter seasons from December–February 2012/2013–2014/2015. We only used winter bird density data, as HPAI outbreaks only occurred between November and March. For organizing the field work and processing the data the whole of The Netherlands was divided in 5 km × 5 km squares (1769 squares in total, including the open water areas of lake IJsselmeer, the Wadden Sea and the large river deltas). Each square was assigned to an observer who performed bird counts in pre-defined spring/summer and winter months. Each observer was instructed to perform a minimum of three bird counts in the assigned square between December and February, visiting all main habitats within the square. As a final result after counts were finished, the observer made an estimate of the number of individuals per bird species present in the square in the winter according to the following classes: 1–3; 4–10; 11–25; 26–50; 51–100; 101–250; 251–500; 501–1000; >1000. Sovon used the obtained data to construct maps with numbers per species per square based on the estimated numbers per square. An internal reviewing process checked the constructed maps for each bird species. Quality checks entailed checking if the pattern of distribution of a bird species over The Netherlands was accurate compared to what is known from the distribution of the species from other projects lead by Sovon.

## References

1. Verhagen, J.H.; van der Jeugd, H.P.; Nolet, B.A.; Slaterus, R.; Kharitonov, S.P.; de Vries, P.P.; Vuong, O.; Majoor, F.; Kuiken, T.; Fouchier, R.A. Wild bird surveillance around outbreaks of highly pathogenic avian influenza A(H5N8) virus in the Netherlands, 2014, within the context of global flyways. *Euro Surveill.* **2015**, *20*, 21–32. [CrossRef] [PubMed]
2. EFSA. Avian influenza overview August–December 2020. *EFSA J.* **2020**, *18*, e06379. [CrossRef]
3. Lycett, S.J.; Bodewes, R.; Pohlmann, A.; Banks, J.; Bányai, K.; Boni, M.F.; Bouwstra, R.; Breed, A.C.; Brown, I.H.; Chen, H.; et al. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science* **2016**, *354*, 213–217. [CrossRef]
4. Beerens, N.; Heutink, R.; Bergervoet, S.A.; Harders, F.; Bossers, A.; Koch, G. Multiple reassorted viruses as cause of highly pathogenic avian influenza A(H5N8) virus epidemic, the Netherlands, 2016. *Emerg. Infect. Dis.* **2017**, *23*, 1966–1973. [CrossRef]
5. Beerens, N.; Koch, G.; Heutink, R.; Harders, F.; Vries, D.P.E.E.; Ho, C.; Bossers, A.; Elbers, A.; Nancy, B.; Guus, K.; et al. Novel Highly Pathogenic Avian Influenza A(H5N6) Virus in the Netherlands, December 2017. *Emerg. Infect. Dis.* **2018**, *24*, 770. [CrossRef]
6. Kleyheeg, E.; Slaterus, R.; Bodewes, R.; Rijks, J.M.; Spierenburg, M.A.H.; Beerens, N.; Kelder, L.; Poen, M.J.; Stegeman, J.A.; Fouchier, R.A.M.; et al. Deaths among wild birds during highly pathogenic avian influenza A(H5N8) virus outbreak, the Netherlands. *Emerg. Infect. Dis.* **2017**, *23*, 2050–2054. [CrossRef]
7. Verhagen, J.H.; Herfst, S.; Fouchier, R.A. Infectious disease. How a virus travels the world. *Science* **2015**, *347*, 616–617. [CrossRef]
8. Bouwstra, R.; Heutink, R.; Bossers, A.; Harders, F.; Koch, G.; Elbers, A. Full-genome sequence of influenza A(H5N8) virus in poultry linked to sequences of strains from Asia, the Netherlands, 2014. *Emerg. Infect. Dis.* **2015**, *21*, 872–874. [CrossRef]
9. Napp, S.; Majó, N.; Sánchez-González, R.; Vergara-Alert, J. Emergence and spread of highly pathogenic avian influenza A(H5N8) in Europe in 2016–2017. *Transbound. Emerg. Dis.* **2018**, *65*, 1217–1226. [CrossRef]
10. EFSA. Avian influenza overview September–December 2021. *EFSA J.* **2021**, *19*, e07108. [CrossRef]
11. EFSA. Avian influenza overview February–May 2021. *EFSA J.* **2021**, *19*, e06951. [CrossRef]
12. FAO. Global AIV with Zoonotic Potential Situation Update 23 February 2022. Available online: [https://www.fao.org/ag/againfo/programmes/en/empres/Global\\_AIV\\_Zoonotic\\_Update/situation\\_update.html](https://www.fao.org/ag/againfo/programmes/en/empres/Global_AIV_Zoonotic_Update/situation_update.html) (accessed on 18 March 2022).
13. ADIS. European Commission’s Animal Disease Information System-HPAI Positive Events by Country and Poultry Category (2020–2021). Available online: <https://www.izsvenezie.com/documents/reference-laboratories/avian-influenza/europe-updates/HPAI/2020-1/total-events.pdf> (accessed on 18 March 2022).
14. ADIS. European Commission’s Animal Disease Information System-HPAI Positive Events by Country and Poultry Category (2021–2022). Available online: <https://www.izsvenezie.com/documents/reference-laboratories/avian-influenza/europe-updates/HPAI/2021-1/total-events.pdf> (accessed on 18 March 2022).
15. ADIS. European Commission’s Animal Disease Information System-Wild Birds: Species Affected 2020–2021. Available online: <https://izsvenezie.com/documents/reference-laboratories/avian-influenza/europe-updates/HPAI/2020-1/wild-birds.pdf> (accessed on 18 March 2022).
16. ADIS. European Commission’s Animal Disease Information System-Wild Birds: Species Affected 2021–2022. Available online: <https://izsvenezie.com/documents/reference-laboratories/avian-influenza/europe-updates/HPAI/2021-1/wild-birds.pdf> (accessed on 18 March 2022).
17. Belkhiria, J.; Hijmans, R.J.; Boyce, W.; Crossley, B.M.; Martínez-López, B. Identification of high risk areas for avian influenza outbreaks in California using disease distribution models. *PLoS ONE* **2018**, *13*, e0190824. [CrossRef] [PubMed]
18. Gilbert, M.; Pfeiffer, D.U. Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: A review. *Spat. Spatiotemporal Epidemiol.* **2012**, *3*, 173–183. [CrossRef] [PubMed]
19. Si, Y.; de Boer, W.F.; Gong, P. Different Environmental Drivers of Highly Pathogenic Avian Influenza H5N1 Outbreaks in Poultry and Wild Birds. *PLoS ONE* **2013**, *8*, e53362. [CrossRef]
20. Velkers, F.C.; Manders, T.T.M.; Vernooij, J.C.M.; Stahl, J.; Slaterus, R.; Stegeman, J.A. Association of wild bird densities around poultry farms with the risk of highly pathogenic avian influenza virus subtype H5N8 outbreaks in the Netherlands, 2016. *Transbound. Emerg. Dis.* **2020**, *68*, 76–87. [CrossRef]
21. Beerens, N.; Heutink, R.; Pritz-Verschuren, S.; Germeraad, E.A.; Bergervoet, S.A.; Harders, F.; Bossers, A.; Koch, G. Genetic relationship between poultry and wild bird viruses during the highly pathogenic avian influenza H5N6 epidemic in the Netherlands, 2017–2018. *Transbound. Emerg. Dis.* **2019**, *66*, 1370–1378. [CrossRef]
22. Elbers, A.R.W.; Gonzales, J.L. Quantification of visits of wild fauna to a commercial free-range layer farm in the Netherlands located in an avian influenza hot-spot area assessed by video-camera monitoring. *Transbound. Emerg. Dis.* **2020**, *67*, 661–677. [CrossRef]
23. Velkers, F.C.; Blokhuis, S.J.; Veldhuis Kroeze, E.J.B.; Burt, S.A. The role of rodents in avian influenza outbreaks in poultry farms: A review. *Vet. Q.* **2017**, *37*, 182–194. [CrossRef]
24. Root, J.; Shriner, S.A. Avian influenza A virus associations in wild, terrestrial mammals: A review of potential synanthropic vectors to poultry facilities. *Viruses* **2021**, *12*, 1352. [CrossRef]
25. Yoo, D.S.; Lee, K.N.; Chun, B.C.; Lee, H.S.; Park, H.; Kim, J.K. Preventive effect of on-farm biosecurity practices against highly pathogenic avian influenza (HPAI) H5N6 infection on commercial layer farms in the Republic of Korea during the 2016–17 epidemic: A case-control study. *Prev. Vet. Med.* **2022**, *199*, 105556. [CrossRef]

26. Beerens, N.; Heutink, R.; Harders, F.; Roose, M.; Pritz-Verschuren, S.B.E.; Germeraad, E.A.; Engelsma, M. Incursion of novel highly pathogenic avian influenza A(H5N8) virus, the Netherlands, October 2020. *Emerg. Infect. Dis.* **2021**, *27*, 1750–1753. [[CrossRef](#)] [[PubMed](#)]
27. Caron, A.; Cappelle, J.; Gaidet, N. Challenging the conceptual framework of maintenance hosts for influenza A viruses in wild birds. *J. Appl. Ecol.* **2017**, *54*, 681–690. [[CrossRef](#)]
28. Lisovski, S.; van Dijk, J.G.B.; Klinkenberg, D.; Nolet, B.A.; Fouchier, R.A.M.; Klaassen, M. The roles of migratory and resident birds in local avian influenza infection dynamics. *J. Appl. Ecol.* **2018**, *55*, 2963–2975. [[CrossRef](#)] [[PubMed](#)]
29. Yoo, D.S.; Kang, S.I.; Lee, Y.N.; Lee, E.K.; Kim, W.Y.; Lee, Y.J. Bridging the Local Persistence and Long-Range Dispersal of Highly Pathogenic Avian Influenza Virus (HPAIV): A Case Study of HPAIV-Infected Sedentary and Migratory Wildfowls Inhabiting Infected Premises. *Viruses* **2022**, *14*, 116. [[CrossRef](#)]
30. Alarcon, P.; Brouwer, A.; Venkatesh, D.; Duncan, D.; Dovas, C.I.; Georgiades, G.; Monne, I.; Fusaro, A.; Dan, A.; Śmietanka, K.; et al. Comparison of 2016–17 and Previous Epizootics of Highly Pathogenic Avian Influenza H5 Guangdong Lineage in Europe. *Emerg. Infect. Dis.* **2018**, *24*, 2270–2283. [[CrossRef](#)]
31. Hill, N.J.; Takekawa, J.Y.; Ackerman, J.T.; Hobson, K.A.; Herring, G.; Cardona, C.J.; Runstadler, J.A.; Boyce, W.M. Migration strategy affects avian influenza dynamics in mallards (*Anas platyrhynchos*). *Mol. Ecol.* **2012**, *21*, 5986–5999. [[CrossRef](#)]
32. Yoo, D.S.; Lee, K.; Beatriz, M.L.; Chun, B.C.; Belkhiria, J.; Lee, K.N. Spatiotemporal risk assessment for avian influenza outbreak based on the dynamics of habitat suitability for wild birds. *Transbound. Emerg. Dis.* **2021**, 1–15. [[CrossRef](#)]
33. Scolamacchia, F.; Mulatti, P.; Mazzucato, M.; Barbujani, M.; Harvey, W.T.; Fusaro, A.; Monne, I.; Marangon, S. Different environmental gradients associated to the spatiotemporal and genetic pattern of the H5N8 highly pathogenic avian influenza outbreaks in poultry in Italy. *Transbound. Emerg. Dis.* **2021**, *68*, 152–167. [[CrossRef](#)]
34. Huang, Z.Y.X.; Xu, C.; van Langevelde, F.; Ma, Y.; Langendoen, T.; Mundkur, T.; Si, Y.; Tian, H.; Kraus, R.H.S.; Gilbert, M.; 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*, 1044–1053. [[CrossRef](#)]
35. McDuie, F.; Matchett, E.L.; Prosser, D.J.; Takekawa, J.Y.; Pitesky, M.E.; Lorenz, A.A.; McCuen, M.M.; Overton Cory, T.; Ackerman, J.T.; De La Cruz, S.E.W.; et al. Pathways for avian influenza virus spread: GPS reveals wild waterfowl in commercial livestock facilities and connectivity with the natural wetland landscape. *Transbound. Emerg. Dis.* **2022**, 1–15. [[CrossRef](#)]
36. Humphreys, J.M.; Douglas, D.C.; Ramey, A.M.; Mullinax, J.M.; Soos, C.; Link, P.; Walther, P.; Prosser, D.J. The spatial–temporal relationship of blue-winged teal to domestic poultry: Movement state modelling of a highly mobile avian influenza host. *J. Appl. Ecol.* **2021**, *58*, 2040–2052. [[CrossRef](#)]
37. Gonzales, J.L.; Pritz-Verschuren, S.; Bouwstra, R.; Wiegel, J.; Elbers, A.R.W.; Beerens, N. Seasonal risk of low pathogenic avian influenza virus introductions into free-range layer farms in the Netherlands. *Transbound. Emerg. Dis.* **2020**, *68*, 127–136. [[CrossRef](#)] [[PubMed](#)]
38. Hill, S.C.; Lee, Y.J.; Song, B.M.; Kang, H.M.; Lee, E.K.; Hanna, A.; Gilbert, M.; Brown, I.H.; Pybus, O.G. Wild waterfowl migration and domestic duck density shape the epidemiology of highly pathogenic H5N8 influenza in the Republic of Korea. *Infect. Genet. Evol.* **2015**, *34*, 267–277. [[CrossRef](#)] [[PubMed](#)]
39. de Knecht, H.J.; van Langevelde, F.; Skidmore, A.K.; Delsink, A.; Slotow, R.; Henley, S.; Bucini, G.; de Boer, W.F.; Coughenour, M.B.; Grant, C.C.; et al. The spatial scaling of habitat selection by African elephants. *J. Anim. Ecol.* **2011**, *80*, 270–281. [[CrossRef](#)] [[PubMed](#)]
40. Verhagen, J.H.; Fouchier, R.A.M.; Lewis, N. Highly pathogenic avian influenza viruses at the wild–domestic bird interface in Europe: Future directions for research and surveillance. *Viruses* **2021**, *13*, 212. [[CrossRef](#)]
41. Animal and Plant Health Agency (UK); Erasmus Medical Centre (NL); Friedrich Loeffler Institute (DE); Istituto Zooprofilattico Sperimentale delle Venezie (IT); Linnaeus University (SE); Wageningen University (NL). Report about HPAI introduction into Europe, HPAI detection in wild birds and HPAI spread between European holdings in the period 2005–2015. *EFSA Supporting Publ.* **2017**, *14*, 1284E. [[CrossRef](#)]
42. Bouwstra, R.J.; Koch, G.; Heutink, R.; Harders, F.; Van Der Spek, A.; Elbers, A.R.; Bossers, A. Phylogenetic analysis of highly pathogenic avian influenza A(H5N8) virus outbreak strains provides evidence for four separate introductions and one between-poultry farm transmission in the Netherlands, November 2014. *Euro Surveill.* **2015**, *20*, 1–12. [[CrossRef](#)]
43. Poen, M.J.; Bestebroer, T.M.; Vuong, O.; Scheuer, R.D.; van der Jeugd, H.P.; Kleyheeg, E.; Eggink, D.; Lexmond, P.; van den Brand, J.M.A.; Begeman, L.; et al. Local amplification of highly pathogenic avian influenza H5N8 viruses in wild birds in the Netherlands, 2016 to 2017. *Euro Surveill.* **2018**, *23*, 17-00449. [[CrossRef](#)]
44. Sovon Dutch Center for Field Ornithology. *Bird Atlas of the Netherlands*, 1st ed.; Kosmos Publishers: Utrecht, The Netherlands; Antwerp, Belgium, 2018; pp. 1–640.
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <https://www.R-project.org/> (accessed on 10 August 2021).
46. de Knecht, H.J.; van Langevelde, F.; Coughenour, M.B.; Skidmore, A.K.; de Boer, W.F.; Heitkönig, I.M.A.; Knox, N.M.; Slotow, R.; van der Waal, C.; Prins, H.H.T. Spatial autocorrelation and the scaling of species–environment relationships. *Ecology* **2010**, *91*, 2455–2465. [[CrossRef](#)]
47. Holland, J.D.; Bert, D.G.; Fahrig, L. Determining the Spatial Scale of Species’ Response to Habitat. *Bioscience* **2004**, *54*, 227–233. [[CrossRef](#)]

48. Therneau, T. A Package for Survival Analysis in R, 2020. R Package Version 3.2-7. 2020. Available online: <https://cran.r-project.org/package=survival> (accessed on 10 August 2021).
49. Wright, M.N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [[CrossRef](#)]
50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
51. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [[CrossRef](#)] [[PubMed](#)]