Spatial distribution of poultry farms using point pattern modelling: a method to address livestock environmental impacts and disease transmission risks

Marie-Cécile Dupas, Francesco Pinotti, Chaitanya Joshi, Madhvi Joshi, Damer Blake, Fiona Tomley, Marius Gilbert, Guillaume Fournié

Spatial distribution of poultry farms using point pattern modelling: a method to address livestock environmental impacts and disease transmission risks

Marie-Cécile Dupas[a,b], Francesco Pinotti[c], Chaitanya Joshi[d], Madhvi Joshi[d], Damer Blake[e], Fiona Tomley[e], Marius Gilbert[a], Guillaume Fournié[e,f,1]

[a] *Université Libre de Bruxelles, Spatial Epidemiology Lab, Belgium*
[b] *Hasselt University, Data Science Institute, Belgium*
[c] *University of Oxford, UK*
[d] *Gujarat Biotechnology Research Centre, India*
[e] *Department of Pathobiology and Population Sciences, Royal Veterinary College, London, UK,*
[f] *INRAE, VetAgro Sup, UMR EPIA, Université de Lyon, Marcy l'Etoile,69280, France,*
[g] *INRAE, VetAgro Sup, UMR EPIA, Université Clermont Auvergne, Saint Genes Champanelle, 63122, France,*

Abstract

The distribution of farm locations and sizes is paramount to characterize patterns of disease spread. With some regions undergoing rapid intensification of livestock production, resulting in increased clustering of farms in peri-urban areas, measuring changes in the spatial distribution of farms is crucial to design effective interventions. However, those data are not available in many countries, their generation being resource-intensive. Here, we develop a farm distribution model (FDM), which allows the prediction of locations and sizes of poultry farms in countries with scarce data. The model combines (i) a Log-Gaussian Cox process model to simulate the farm distribution as a spatial Poisson point process, and (ii) a random forest model to simulate farm sizes (i.e. the number of animals per farm). Spatial predictors were used to calibrate the FDM on intensive broiler and layer farm distributions in Bangladesh, Gujarat (Indian state) and Thailand. The FDM yielded realistic farm distributions in terms of spatial clustering, farm locations and sizes, while providing insights on the factors influencing these distributions. Finally, we illustrate the relevance of modelling realistic farm distributions in the context of epidemic spread by simulating pathogen transmission on an

array of spatial distributions of farms. We found that farm distributions generated from the FDM yielded spreading patterns consistent with simulations using observed data, while random point patterns underestimated the probability of large outbreaks. Indeed, spatial clustering increases vulnerability to epidemics, highlighting the need to account for it in epidemiological modelling studies. As the FDM maintains a realistic distribution of farm location and sizes, its use to inform mathematical models of disease transmission is particularly relevant for regions where these data are not available.

1. Introduction

Livestock contribute to food security as a main source of animal protein (dairy, meat and eggs) and provide 17% of the world population's dietary energy intake (FAO, 2017) whilst occupying 80% of global agricultural land (Ritchie and Roser, 2019). The consumption of animal source food is increasing most rapidly in low- and middle-income countries (LMICs) (Kastner et al., 2012). In Asia, chicken meat production has quadrupled in the last two decades (FAOSTAT); this rapid intensification has been characterised by geographic displacement of farms, especially pig and poultry farms which have smaller land requirements than ruminants, from rural to peri-urban areas (Steinfeld, 2006).

Accurate and up-to-date maps of livestock farms are crucial to assess their environmental impacts and the risk of diseases spreading through livestock populations (Keeling et al., 2001). However, many LMICs do not have the resources needed to keep track of exact farm locations. For this reason, some studies have relied on modelling frameworks to estimate high-resolution farm distribution at various administrative levels, and to improve on coarse census data. Such frameworks have mostly employed linear regression models (Prosser et al., 2011; Robinson et al., 2014; Van Boeckel et al., 2011). Van Boeckel et al. (2012) mapped the distribution of intensive poultry farms in Thailand using a simultaneous autoregression model (SAR) that explicitly accounted for spatial autocorrelation. However, as reported by the authors, the model failed to capture the high levels of spatial clustering that are observed among intensive farms in that country. Random forest models have been shown to outperform linear regression models when used to downscale census data at the global scale for several livestock animal species (Gilbert et al., 2018) or at national scale for pig populations in Thailand

(Thanapongtharm et al., 2016) and China (Zhao et al., 2022). However, the distribution of animals at farm level and the process of generating clustered point distributions are yet to be embedded in these models. For instance, the Gridded Livestock of the World (GLW) predicts livestock as a continuous, gradually varying, density of animals per pixel at 10 km or 1 km resolution (Wint and Robinson, 2007; Robinson et al., 2014; Gilbert et al., 2018). Thus, these models do not provide information about how animals are distributed across farms, and how farms are distributed across space, despite these parameters having major influence on both environmental impacts and disease risk associated with intensification of livestock production.

This paper develops a novel modelling framework that alleviates limitations of previous models and can be used to predict both farm locations and sizes. The framework builds on a previous point pattern model introduced by Chaiban et al. (2019, 2021) that was used to predict clustered farm distributions. We show that this Farm Distribution Model (FDM) successfully predicts spatial farm locations and sizes of poultry farms in three 'test' geographic regions (Bangladesh, Gujarat and Thailand) that are characterised by different levels of intensification and for which farm data were already available.

We trained a Log Gaussian Cox Process (LGCP) and Random Forest model (RF) and assessed their external and internal validity using three observed point patterns. This allowed us to test the robustness of the method to reproduce farm distribution in data-scarce countries. We further illustrate the relevance of our approach to inform models of disease spread in livestock by comparing epidemic simulations on empirical and synthetic farm distributions generated with different methods.

2. Methods

2.1. Training data sets

The modelling procedure was based on farm size and location data from three regions in Asia: Thailand and Bangladesh (whole country) and Gujarat (state in India). The Gujarat Biotechnology Research Centre collected data on the distribution of farms in Gujarat, India. The data represented are based on the information acquired from the Department of Animal Husbandry, Dairying & Fisheries, Ministry of Agriculture & Framers Welfare, Government of India and Directorate of Animal Husbandry, Government of Gujarat, Gandhinagar, Gujarat, India in the year 2020.

Area	Area Code	GDP[a] (billion US$)	Surface area (10^3 km^2)	Intensity (pts/m^2)	Intensity of broiler farms (pts/m^2)	Intensity of layer farms (pts/m^2)	Number of farms (broilers - layers)
Gujarat	IN.GJ	230	196	$1.77\ 10^{-8}$	$9.88\ 10^{-9}$	$1.39\ 10^{-9}$	2,611 - 311
Thailand	THA	543.5	514	$4.81\ 10^{-9}$	$3.14\ 10^{-9}$	$1.03\ 10^{-9}$	3,717 - 1,439
Bangladesh	BGD	302.6	135	$2.27\ 10^{-7}$	$9.80\ 10^{-8}$	$3.75\ 10^{-8}$	22,159 - 9,074

Table 1: Characteristics of data sets in terms of area surface, intensity of points distribution (points/m^2) and economic features.

[a]worldbank 2019

Around 59% of farm locations in Gujarat corresponded to village centroids coordinates. Farms with overlapping locations were assigned to random points within the area of the corresponding villages. A similar procedure was adopted for farms in Thailand for which only the village location is known (data collected in 2010 by the Department of Livestock Development (Chaiban et al., 2019)). Finally, the geographic coordinates of the farms in the Bangladesh dataset were obtained from an agricultural census collected by the Food and Agriculture Organization of the United Nations, ensuring accuracy and reliability of the location data.

Bangladesh and Gujarat data sets cover areas of similar size, while farms in Thailand are scattered over a region that is around three times larger (Table 1). Data sets consisted of the coordinates and capacity of farms differentiated according to their type of production into broiler and layer farms (Figure S1). We assumed that the size of a farm coincides with its capacity (i.e. maximum number of animals that can be raised on a farm) and hence ignored yearly stock variations. We kept only farms with more than 500 chickens since the original FDM was developed for intensive farms (Chaiban et al., 2019).

2.2. Spatial predictors

Table 2 lists the spatial predictors used for the LGCP and RF models accross 4 categories of covariates: anthropogenic, topographical, vegetation and livestock characteristics. The distribution of chicken density was derived from the most recent version of the Gridded Livestock of the World (GLW, (Gilbert et al., 2018)). Proximity predictors were the inverse of time travel to major cities, ports and roads ($x = \frac{1}{timetravel+1}$), so that the maximal values were associated to the closest locations. These predictors allowed us to assess if farm locations were affected by infrastructure density. Other predictors were used as originally published (references in the Table 2).

Type	Variable	Units	Source	Abbreviation
Anthropogenic	Human population density	Log10 people per hectare	Tatem, 2017	Hpop
	Proximity to cities with 5,000,000<x<50,000,000 inhabitants	$Minute^{-1}$	Nelson et al, 2019	Access_MC1
	Proximity to cities with 50,000<x<50,000,000 inhabitants	$Minute^{-1}$	Nelson et al, 2019	Access_MC11
	Proximity to cities with with 1,000,000<x<5,000,000	$Minute^{-1}$	Nelson et al, 2019	Access_MC2
	Proximity to large and medium ports	$Minute^{-1}$	Nelson et al, 2019	Access_Port12
	Proximity to roads	$Minute^{-1}$	Meijer et al, 2019	Proxim_roads
Topography	Slope		Amatulli et al, 2018	Slope
Vegetation	Crop cover	Pixel % covered by crops	Fritz et al, 2015	Crop
	Tree cover	Pixel % covered by forest	Hansen et al, 2013	Tree
Livestock	Chicken population density	Log10 animals per hectare	Gilbert et al, 2018	nChicken

Table 2: List of spatial predictors.

2.3. Point pattern modelling

The procedure for modelling farm locations is based on the point pattern analysis method described in Chaiban et al. (2021). We modelled spatial point patterns using LGCP associated with spatial predictors and the Palm maximum likelihood method of parameters optimisation (Tanaka et al., 2008). This approach was found to outperform other types of point pattern models at reproducing clustered farm distributions (Chaiban et al., 2021, 2019). This method is suitable to deal with highly inhomogeneous point intensity and spatial autocorrelation. Point distributions are generated in space stochastically according to a Poisson process with intensity $\lambda(u)$:

$$\lambda(u) = exp(\theta_0 + \theta_1 pred_1 + \theta_2 pred_2 + ... + \theta_n pred_n) \, . \tag{1}$$

where u denotes a location of the area, θ_i are the model weights associated to spatial predictor $pred_i$.

We applied a LGCP model to each pairing of study region (Bangladesh, Thailand, and Gujarat) with a poultry production type (broiler or layer), resulting in a total of six models. The validity of these models was then evaluated both within their respective training regions (internal validation) and by application to regions where they were not originally trained (external validation).

The importance of each spatial predictor $pred_i$ was computed as the product of its maximum value across space and its estimated weight θ_i.

2.4. Point pattern characterisation and model validation

2.4.1. Spatial correlation analysis of points pattern

Ripley's K-function measures the clustering behaviour in a spatial point pattern (SPP), and is defined as the cumulative average number of data points found within a distance r of a typical data point (Ripley, 1976; Baddeley et al., 2015). The inhomogeneous K-function, $K_{inhom}(r)$, is a generalization of Ripley's K-function designed to analyze point patterns with varying intensity across space. The inhomogeneous K-function is defined as follows:

$$K_{inhom}(r) = \frac{|W|}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathcal{I}\{d_{ij} \leq r\} e_{ij} \,, \tag{2}$$

where N is the number of points, $|W|$ denotes total study area, $\mathcal{I}\{d_{ij} \leq r\}$ equals 1 if the euclidian distance d_{ij} between points i,j is less than r and is 0 otherwise, and e_{ij} is an edge correction weight to avoid sampling biases. Given the location of the first point x and the distance $d = ||x - x'||$, the second point x' must lie in the circle b of radius d and centred at x. However, the circle b is generally only partly inside the study area W for large d. Then, the Ripley's isotropic correction e_{ij} uses the fraction of the length of the circle, ℓ, that is within the study area and considers that the point pattern is isotropic (statistically invariant under rotation). We calculated the probability of the second point x' being inside the window W as:

$$p(x, d) = \frac{\ell(W \cap \delta b(x, d))}{2\pi d}, \tag{3}$$

Finally, the edge correction is:

$$e_{ij} = \frac{1}{p(x_i, d_{ij})}. \tag{4}$$

We used the Besag's transform of the in-homogeneous K-function ($K(r)$) by using the function *linhom* in the package *spastat* which is:

$$L_{inhom}(r) = \sqrt{\frac{K_{inhom}(r)}{\pi}}. \tag{5}$$

2.4.2. Global Rank Envelope Test for Model Validation

The global rank envelope test is a robust statistical method used to evaluate the goodness-of-fit of a model by comparing the observed data to a collection of simulated data generated from the model. It provides a comprehensive approach for assessing whether simulated and observed patterns are consistent.

The global rank envelope test is based on a chosen test statistic, in this study the $L_{inhom}(r)$ function, a transformed version of the inhomogeneous K-function. For each simulation and at each distance r, we computed the test statistic and ranked the observed value of the statistic among the simulated values. This created an envelope of expected values under the model. If the observed test statistic lay within this envelope for all distances, it suggested that the model was an adequate fit to the observed data. The test procedure is described in the Appendix A.2.

2.4.3. Quadrat counting tests

We divided study areas into quadrats, and computed counts of points within each quadrat (n = 23 - 44 depending on the study area) (De Cola, 1991). The patterns of quadrats are shown in Figure S2. We did not consider quadrats that occupy less than 80% of the complete theoretical polygon to avoid edge effects. The performance of a model was evaluated by computing the correlation coefficient between the log-transformed of the number of points per quadrat between the observed and simulated pattern patterns.

2.5. Farm size modelling

2.5.1. Random Forest model with spatial predictors

The second-step of the algorithm consists of training a RF regression model to predict farm sizes. First, we averaged spatial predictors within a radius of 5,000 m around each farm. We tested different buffer zone sizes, of 2,500 m, 5,000 m and 7,500 m, and we selected the 5,000 m buffer zone in the final analysis as it performed slightly better than others. Secondly, we transformed farm sizes X using a power function to reduce the skewness of their distribution:

$$X_{transform} = \frac{X^a - 1}{a}\,, \tag{6}$$

and used the function *PowerTransformer* from the *sklearn* package in Python to fit the parameter a. We used the function *RandomForestRegressor* of the *sklearn* package in Python, with 500 decision trees.

The goodness of fit (GOF) metrics of the predictions were all established through cross-validation, i.e. by measuring the correlation between observed and predicted animal numbers in farms that were not used to train the Random Forest models. The total data set was divided into a training data set (75% of the data) and a validating data set (25%). This process was repeated 5 times, each time selecting a different random set of farms to train the RF models. We then calculated GOF measures, i.e. the correlation coefficient and the root-mean-square error (RMSE) between predicted and observed farm sizes for each fold. Both GOF measures are calculated using log transformed and absolute values of farm sizes.

2.6. Mathematical modelling of disease transmission

2.6.1. ***Simulations***

We simulated the spread of a pathogen over M poultry farms with spatial coordinates (x_i, y_i) and sizes $X_i, i = 1, ..., M$. Simulations were stochastic and farms' infection statuses were updated synchronously, with each time step being 1 day long. Each farm was either susceptible to infection (S), infectious (I) or removed (R). Removed farms do not contribute to transmission and cannot be reinfected. An infectious farm i transmits the pathogen to a susceptible farm j with daily probability:

$$p_{ij}(S \to I) = 1 - exp(-\gamma_{ij}) \,, \tag{7}$$

where the force of infection exerted by i on j is given by:

$$\gamma_{ij} = \beta \cdot X_i^{Q_I} \cdot X_j^{Q_S} \cdot K(d_{ij}) \,, \tag{8}$$

β denoting transmissibility and $K(d_{ij})$ representing a spatial transmission kernel depending solely on the (euclidean) distance between i and j. The exponents Q_I and Q_S allow for different scalings of the force of infection with the sizes of infectious and susceptible farms, respectively.

Infectious farms recover with daily probability:

$$p(I \to R) = 1 - exp(-\mu), \tag{9}$$

where μ is the recovery rate.

We implemented our simulations in C++ using the *Conditional subsample* algorithm (Sellman et al., 2018). Briefly, the algorithm overlays a grid over the study area, so that transmission attempts involving farms belonging

to different grid cells can be checked only after resolving whether any transmission occurs between those cells. In order to ensure an efficient implementation, we used a heuristic, adaptive routine to identify an optimal gridding. Both simulation and cell-construction routines are detailed in (Sellman et al., 2018).

2.6.2. ***Transmission kernels***

We considered a power-law transmission kernel employed by Hill et al. (2017) to study the spread of H5N1 avian influenza virus in Bangladesh:

$$K(d) = 1 \quad \text{if} \quad 0 \leq d < d_{min}, \tag{10}$$

and

$$K(d) = ({}^{d_{min}}/_{d})^{\alpha} \quad \text{if} \quad d \geqslant d_{min}, \tag{11}$$

with $d_{min} = 0.1km$.

We considered long-ranged and short-ranged transmission kernels corresponding respectively to $\alpha = 0.643$ and to α=3 (Figure S3).

2.6.3. ***Simulation scenarios***

For a given study area, we simulated pathogen transmission using six different spatial distributions of farms. First, we considered the empirical distribution of farms, which we used as a reference. Results from this scenario were then compared with SPP generated from the LGCP, and a random distribution of points pattern generated according to a Complete Spatial Randomness process. We also compared a homogeneous farm size scenario, where all farm sizes were set to the average farm size (Constant Size; CS), with a heterogeneous farm size scenario, where RF model was used to assign a size to each farm (Random Forest Size; RFS). All scenarios are summarized in table 3 and displayed in Figure S4 in the Supp Material.

We ran 2000 independent simulations for each scenario. In the case of simulated farm distributions, we generated 40 independent farm distributions and ran 50 disease spreading simulations for each model formulation. In each simulation, we initialised the infection by selecting and infecting a random farm at $t = 0$; a simulation stopped when no infectious farms remained.

Name	Description
Empirical	Uses observed data
Empirical (CS)	Uses empirical locations but farm size is set to the average farm size for all farms.
Random+CS	Farms are scattered uniformly at random over the study area; farm size is set to a constant value (Constant Size; CS), namely average farm size.
LGCP+CS	Farm locations are generated from a LGCP; farm sizes are set to a constant value (Constant Size; CS), namely average farm size.
Random+RFS	Farms are scattered uniformly at random over the study area; farm sizes are generated from a RF model (Random Forest Size; RFS).
LGCP+RFS	Farm locations are generated from a LGCP; farm sizes are generated from a Random Forest model (Random Forest Size; RFS).

Table 3: Farm distributions considered in the disease transmission modelling.

2.6.4. ***Spatial epidemic risk.***

In order to compare spatial predictions of epidemic risk using different farm distributions, we implemented the following methodology. We defined the risk V_i for farm i as the proportion of 100 simulations in which an epidemic starting from this farm reaches at least 100 farms. For each distribution of farms, we first calculated the risk V_i for each farm. In the case of BGD we considered only 4000 random farms as initial seeds due to long computation times. Then, in order to compare different farm distributions, we defined a common spatial grid covering the study area and averaged risk V_i in each cell. This procedure yielded an average risk V_a for each cell in the cell $a = 1, ..., N_{cel}$. Using the same spatial grid for all point models, we then performed a quantitative comparison between the different models on the basis of V_a. We used a rectangular grid of 40x40 km cells to the empirical distribution of farms while allowing for an additional margin of 20 km in each direction.

To assess the extent to which the maps of V_a obtained using LGCP models trained on different sites match the estimate of $V_a^{empirical}$ obtained by using the empirical distribution, we calculated the Spearman's rank correlation coefficient between $V_a^{empirical}$ and the maps $V_a(j)$, where $j = 1, 2, ..., 40$ extends over all the realisations generated from the same model (we omit any pair of cells where at least one does not contain farms). We thus obtain a collection of 40 correlation coefficients for each point model.

3. Results

3.1. Characterisation of spatial homogeneity of farms distribution

Farm density is higher in Bangladesh than in Gujarat and Thailand (Table 1). In all three study areas, the density of broiler farms is higher than for layer farms. According to the L-function, all empirical SPPs are more clustered than a random SPP for distances under 100 km (Figure 1A). The maximal level of clustering occurs under 20 and 25 km for all SPPs, except for layer farms in Bangladesh where it occurs at around 9 km. For all three study areas, layer farms are more clustered than broiler farms for around $r < \frac{r_{max}}{2}$, with r_{max} being the maximum radius for each area.

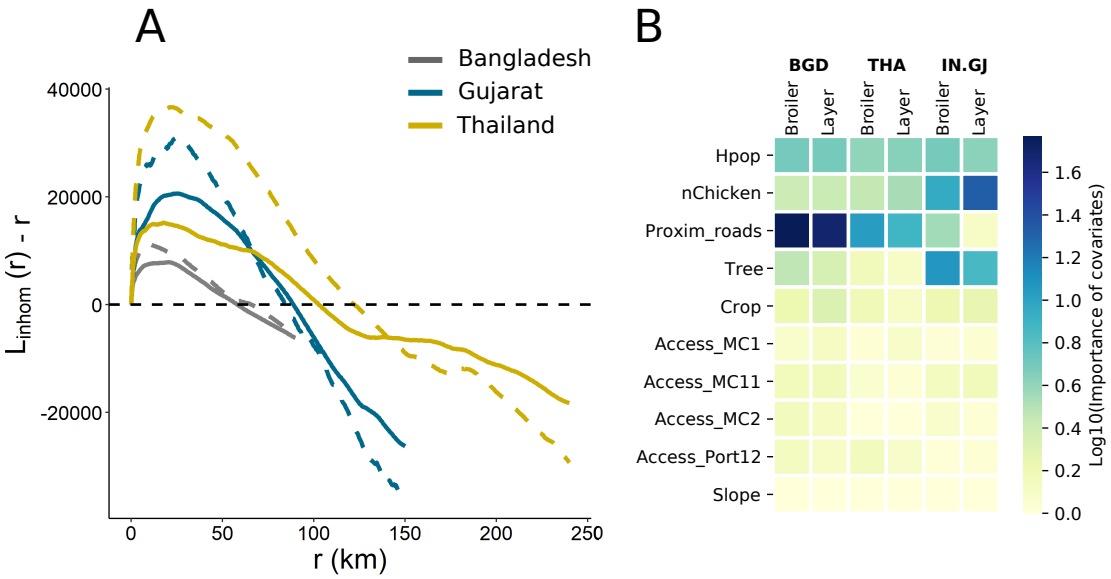

Figure 1: **A. L-function as a function of r for each training data set.** The L-function for broiler (solid) and layer (dashed) farms are presented. Subtracting r from $L_{inhom}(r)$ aids in interpreting the plots; when $L_{inhom}(r) - r$ equals zero, it signifies complete spatial randomness. Values of $L_{inhom}(r) - r$ greater than zero suggest a clustering pattern, whereas negative values indicate a dispersed or regular pattern relative to a random spatial point pattern at the scale of r. The black dashed line represents the L-function of a completely random point pattern. Points above this line denote more clustering, whereas points below indicate greater dispersion than would be expected under spatial randomness. **B. Importance of covariates for LGCP models on a logarithmic scale.** One model was trained per production type and study area (Bangladesh, Thailand, and Gujarat).

We trained a LGCP model for each area and production type (6 models).

Among all predictors, proximity to roads had the highest influence on the locations of farms, except for layer farms in Gujarat for which the distribution is affected by chicken density and tree cover (Figure 1B). Crop and human density are important for all models. Finally, other accessibility predictors and slope were the least important covariates.

3.2. Performance of the farms location model (LGCP)

LGCP model generates a different simulated SPP at each simulation (Figure S5). We evaluated the goodness of fit of the farm location models by two procedures. First, we assessed if the simulated and observed SPPs display similar inhomogeneous patterns by calculating the L-function (section 2.4.1). Second, the quadrat count test allowed us to assess if clusters of farms in the observed and simulated SPPs were similarly located across a study area (section 2.4.2). As an L-function was computed for each simulated SPP, we plotted the envelope of all L-functions generated by 8000 simulations, and compared these to the L-function of the observed SPPs. Figure 2 shows the results for broiler farms.

3.2.1. Broiler farms

For broiler farms, the model trained using Bangladesh data offers the best prediction in terms of both internal and external validation. Indeed, the envelopes generated with the model trained on Bangladesh data and applied to Bangladesh and Gujarat include, or are near, the respective observed L-function (Figure 2). Although the Bangladesh model underestimates the clustering level of the Thailand SPP (2B), it reproduces the L-function for low radii of the Gujarat SPPs (Figure 2C) even though it is different from the Bangladesh L-function. Moreover, the Bangladesh-trained model locates the cluster better with high correlation coefficient between observed and simulated SPPs (Figure 3C).

Although the Gujarat model fails the global rank envelope test for internal validation (Figure S6A & B), the observed L-function remains close to the global envelope. We also note that the latter is particularly thin, indicating consistency between simulated SPPs. While the Gujarat model has a high p-value when applied to Bangladesh (Figure S6A & B), the global envelope is wide (Figure 2H), implying high variability in clustering between simulated SPPs. This suggests the need to interpret the p-value of the global envelope test in combination with the visualisation of the observed L-function and simulated envelope. In addition, the model locates clusters of farms in

Thailand and Gujarat, but not in Bangladesh (Figure 3C). However, for simulated SPPs in Thailand, the global envelope test indicates a higher level of clustering for distances above 50 km than observed.

Finally, the model trained in Thailand reproduces only its own spatial point patterns (Figure 2E), with a thin global envelope of simulations, indicating consistency between simulated SPPs.

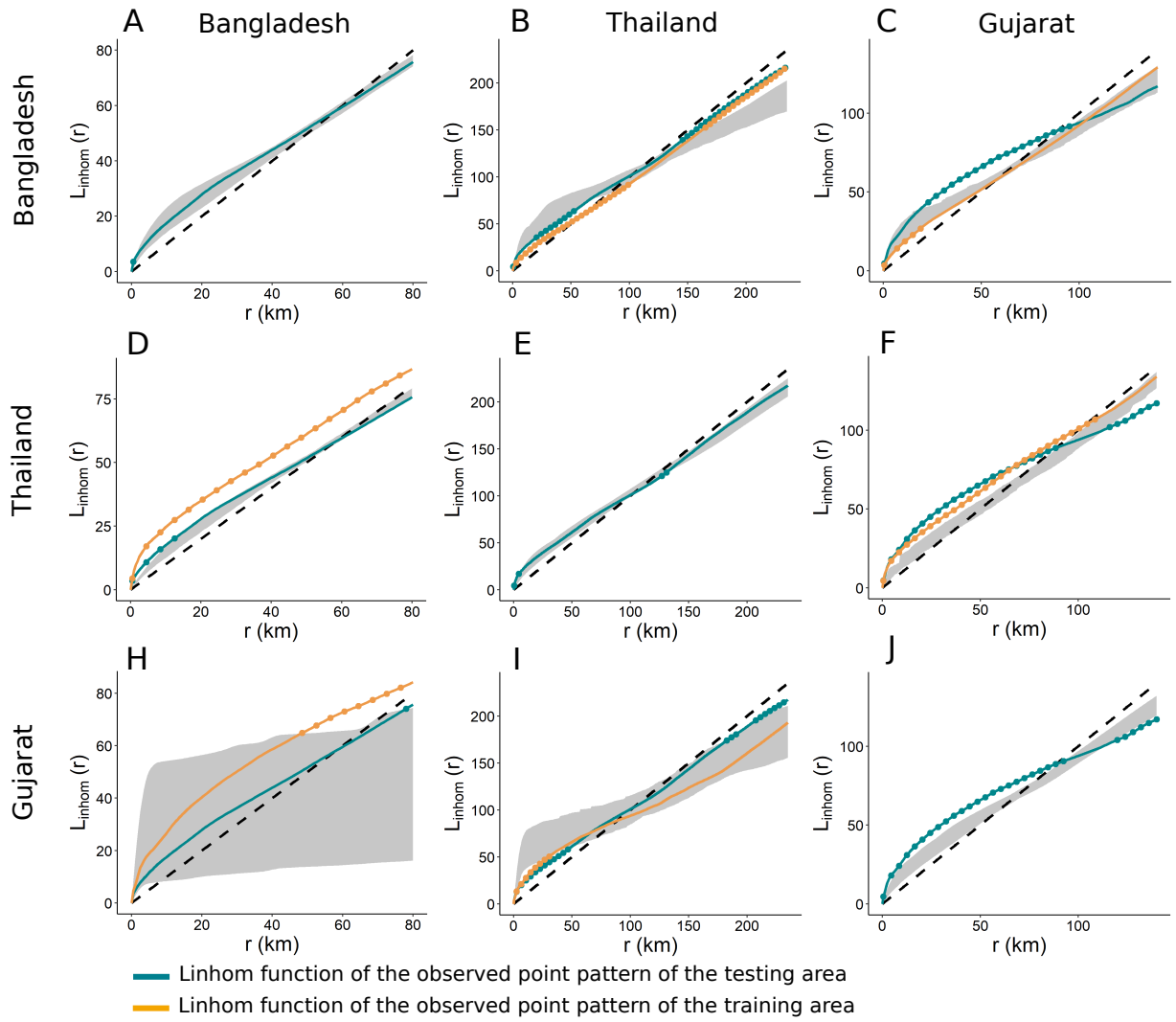

Figure 2: **Global envelope of the L-function of simulated points patterns with the LGCP model for broiler farms.** The L-function of the training point pattern is in orange, and the L-function of the observed points pattern of the testing area is in blue. Points outside the envelope are highlighted with dots. **A & B & C.** Envelope test of simulated SPPs generated with the model trained in Bangladesh, in respectively Bangladesh, Thailand and Gujarat. **D & E & F.** Envelope test of simulated SPPs generated with the model trained in Thailand, in respectively Bangladesh, Thailand and Gujarat. **H & I & J.** Envelope test of simulated SPPs generated with the model trained in Gujarat, in respectively Bangladesh, Thailand and Gujarat.

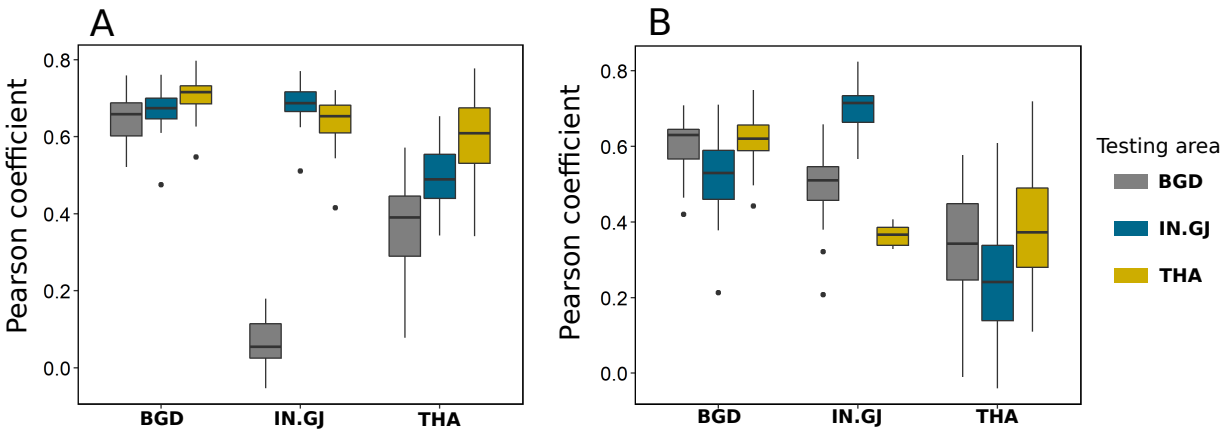

Figure 3: **Boxplot of the correlation coefficient between the numbers of points per quadrat in observed and each simulated SPP for each type of production: broiler (A) and layer (B).** Model names are indicated in the abscissa labels and refer to the area where the model was trained. The color of the boxes indicate where the model is tested (grey for Bangladesh, blue for Gujarat and yellow for Thailand).

3.2.2. Layer farms

All three models for layer farms satisfy the internal validation test of the global envelope (Figure S6B & D). Again, the model trained in Bangladesh performs best, with high global envelope p-value when applied to Bangladesh and Gujarat, and high quadrat correlation coefficient for the three areas. The model underestimates the level of clustering in Thailand, even though the envelope remains close to the observed L-function and followed the same trend.

Although the Gujarat model is associated with high p-values when applied to Bangladesh and Gujarat, the envelopes are wide (Figure 4H). Also, the model does not reproduce the level of clustering and the locations of clusters in Thailand (Figure 3D).

Finally, the Thailand model reproduces the Bangladesh L-function, despite SPPs differing widely across countries. However, prediction of cluster locations is poor (Figure 3D).

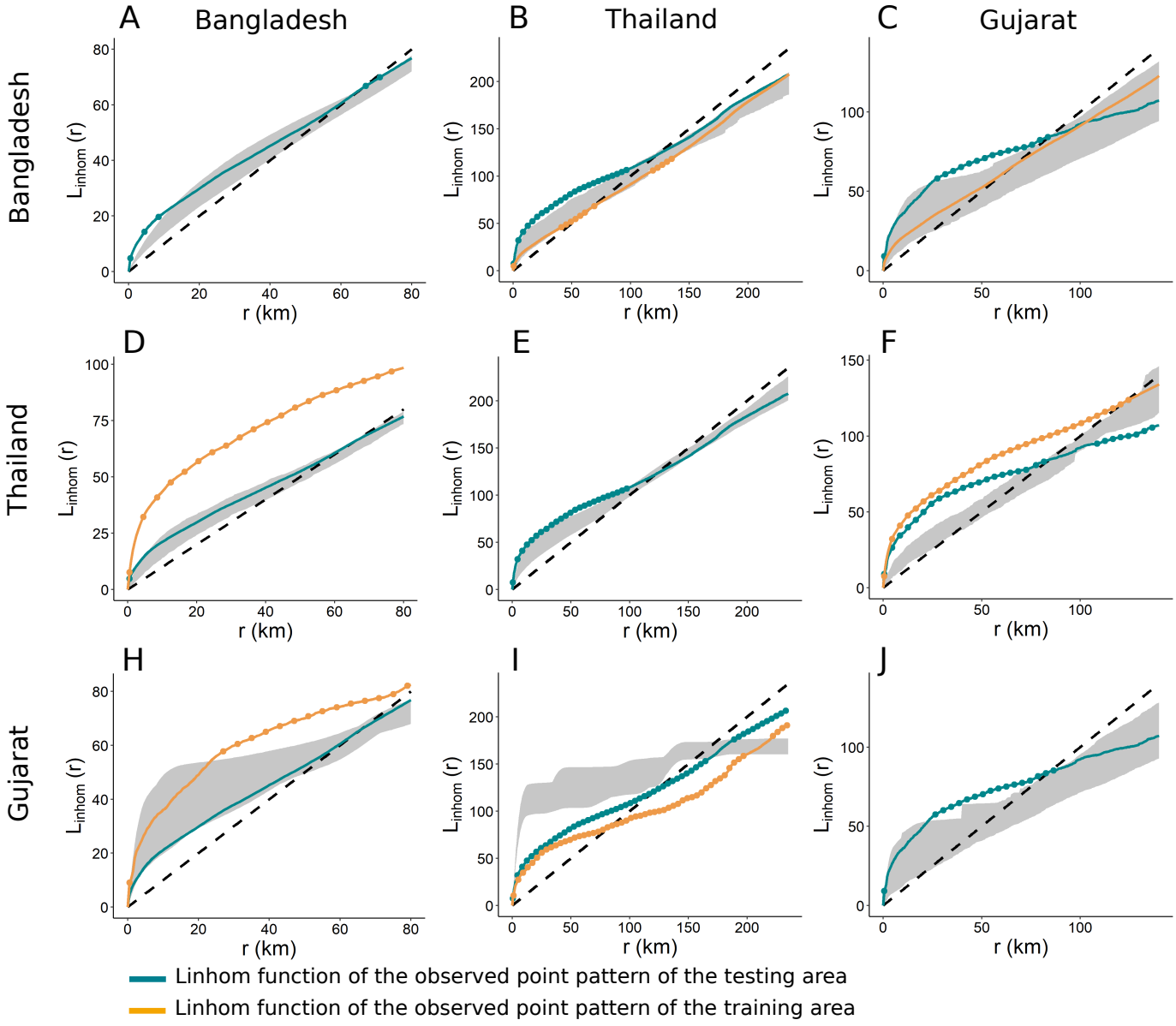

Figure 4: **Global envelope of the L-function of simulated points patterns for layer farms.** The L-function of the training point pattern is in orange, and the L-function of the observed points pattern of the testing area is in blue. Points outside the envelope are highlighted with dots. **A & B & C.** Envelope test of simulated SPPs generated with the model trained in Bangladesh, in respectively Bangladesh, Thailand and Gujarat. **D & E & F.** Envelope test of simulated SPPs generated with the model trained in Thailand, in respectively Bangladesh, Thailand and Gujarat. **H & I & J.** Envelope test of simulated SPPs generated with the model trained in Gujarat, in respectively Bangladesh, Thailand and Gujarat.

3.3. Farm size predictions with random forest model

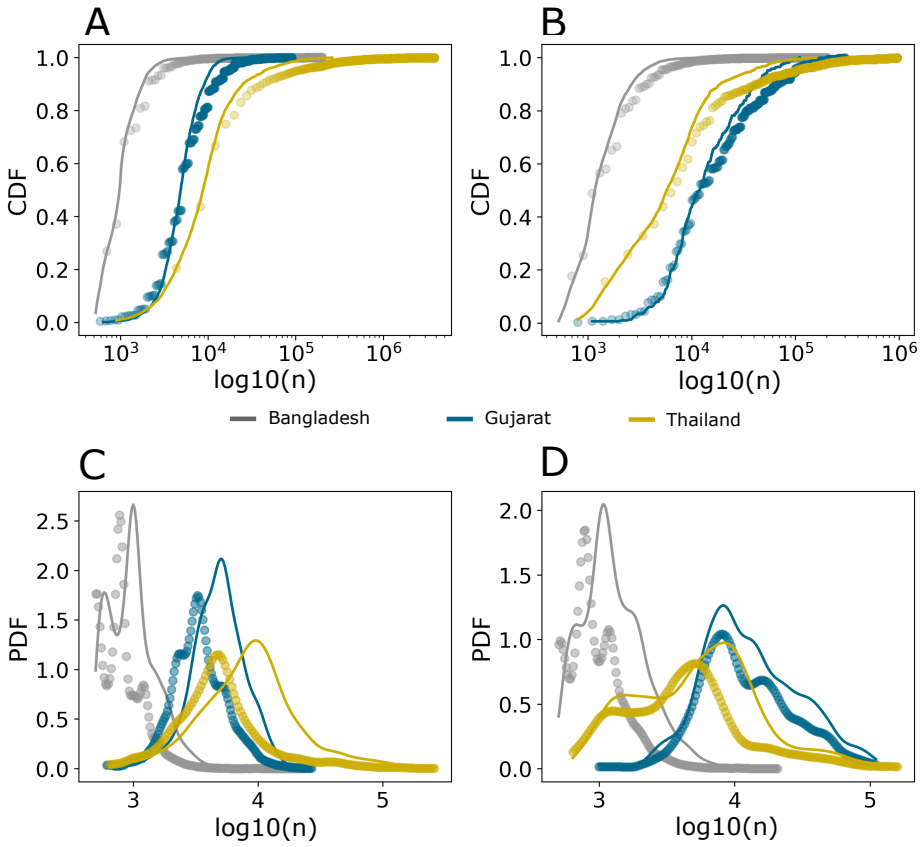

Figure 5: **Distribution of observed and predicted farm size.** Cumulative distribution function ($P(N > n)$) of observed (dots) and predicted (solid line) farm size for broiler farms (A) and layer farms (B). Probability density function of observed (dots) and predicted (solid line) farm size smoothed by a Kernel Density Estimation (KDE) for broiler farms (C) and layer farms (D).

The second step of the FDM consists of predicting the farm sizes using a RF model conditioned by farm locations (generated during the first step). Two RF models were trained, one for broiler farms and another for layer farms. The training data set of these two models covered the three study areas. The most important predictors for farm size include proximity to major cities, tree cover, human and chicken population densities (Fig S7). The distributions of the log size of farms are close to an unimodal distribution

for all three data sets. Bangladesh, which has the largest number of farms of all three countries (table 1), exhibits the lowest farm size peak with a median of 1,000 broiler chickens and 1,100 layer hens per farm. In contrast, Gujarat and Thailand, which are characterised by more intensive livestock production systems, have a median of respectively 5,000 and 10,000 chickens per broiler farm; and a median of 12,000 and 6,500 chickens per layer farm.

The RF model predicts log transformed size of farms with an average correlation coefficient of 0.83 and 0.70 over the five bootstraps for respectively broiler and layer farms (tables S4 and S5). The RMSE between log observed and predicted values is also weaker for broiler farms with around 0.275 against 0.343 for layer farms. These two GOF measures indicate a significant predictability of farm sizes through RF model. Moreover, the distribution of observed and predicted farm sizes shows that the RF model allows us to reproduce the high heterogeneity of the farm size range thanks to the log transformation (Fig. 5). However, heterogeneity is not maintained when the RF model is applied to the distribution of farms generated with LGCP (Fig S8).

3.4. Epidemic transmission modelling.

We now present the results of disease transmission simulations. As detailed in the Methods section, we considered an array of 6 spatial farm distribution models with farm locations corresponding to either observed data or random samples from the LGCP and random point pattern models, and either homogeneous (CS) or heterogeneous (RFS) farm sizes (table 3).

Clustering or random distributions yielded substantial differences in terms of predicting the probability of large disease outbreaks under two transmission kernels with different spatial ranges. Figure 6 shows that epidemic simulations with LGCP-generated point patterns matched those in the empirical networks more closely than simulations performed in fully random farm distributions.

In the context of short-ranged transmission, it is noteworthy that the empirical and simulated farm distributions exhibited substantial discrepancies in terms of epidemic potential. Simulations using random farm distributions significantly underestimated the probability of large outbreaks. In contrast, despite also underestimating the probability of large outbreaks, LGCP models more accurately determined the critical threshold of the transmission parameter (β), beyond which the risk of an epidemic substantially increases from zero (Figure 6). This suggests that LGCP models are capable of capturing the fundamental dynamics and conditions necessary for disease transmission to occur, even if they somewhat underestimate the overall risk.

Using RF-generated farm sizes or employing a constant (average) value for all farms had minimal impact on the simulations, except in the specific case of the empirical distribution in Thailand (compared using black and grey markers). This discrepancy arose due to the inherent heterogeneity of farm sizes across Thailand (Figure 5C), which was not accurately captured by the RF algorithm (Figure S8). Consequently, when the heterogeneities were mitigated by homogenizing farm sizes, the agreement between LGCP and Empirical+CS (grey) improved significantly.

LGCP models trained in Thailand, Gujarat, and Bangladesh exhibited similar results in Gujarat and Bangladesh, but not in Thailand. Specifically, the model trained in Thailand demonstrated superior performance in the context of the short-ranged kernel, whereas the Gujarat model produced more realistic epidemics when considering the long-ranged kernel. Therefore, it appeared that a single best-performing model cannot be identified based on this analysis.

Spatial risk maps were generated by evaluating the epidemic potential of each farm based on its location (Figure S9). Boxplots depicting correlation coefficients between risk maps generated with observed and simulated farm distributions are shown in Figure 7. The performance of LGCP models in relation to the long-ranged kernel was robust, as evidenced by their capability to accurately predict the risk maps. Correlation coefficients exhibited some variability with some simulated configurations displaying risk patterns that were quite different from the observed one. Nevertheless, the median correlation coefficients across all sites are relatively high, suggesting a robust alignment between the predicted risk maps and the actual observed data.

Conversely, correlation coefficients were generally lower for short-ranged transmission, suggesting a decreased predictive accuracy in terms of risk maps. However, it is worth noting that correlation coefficients tended to increase on average with the transmission parameter β, albeit up to a certain threshold beyond which they reached a plateau.

Remarkably, the model trained in Thailand consistently demonstrated the poorest performance at recovering spatial patterns of epidemic risk across all instances, even when applied to the same country. On the other hand, the model trained in Bangladesh appeared to outperform the other models. Not only did it exhibit higher average correlation coefficients, but it also displayed a narrower range of values, indicating a more consistent and reliable predictive performance compared to the other models, which exhibited greater variability in this regard. These findings highlight the importance of carefully selecting and training LGCP models for specific sites and transmission scenarios, as the choice of training data can significantly impact their predictive capabilities.

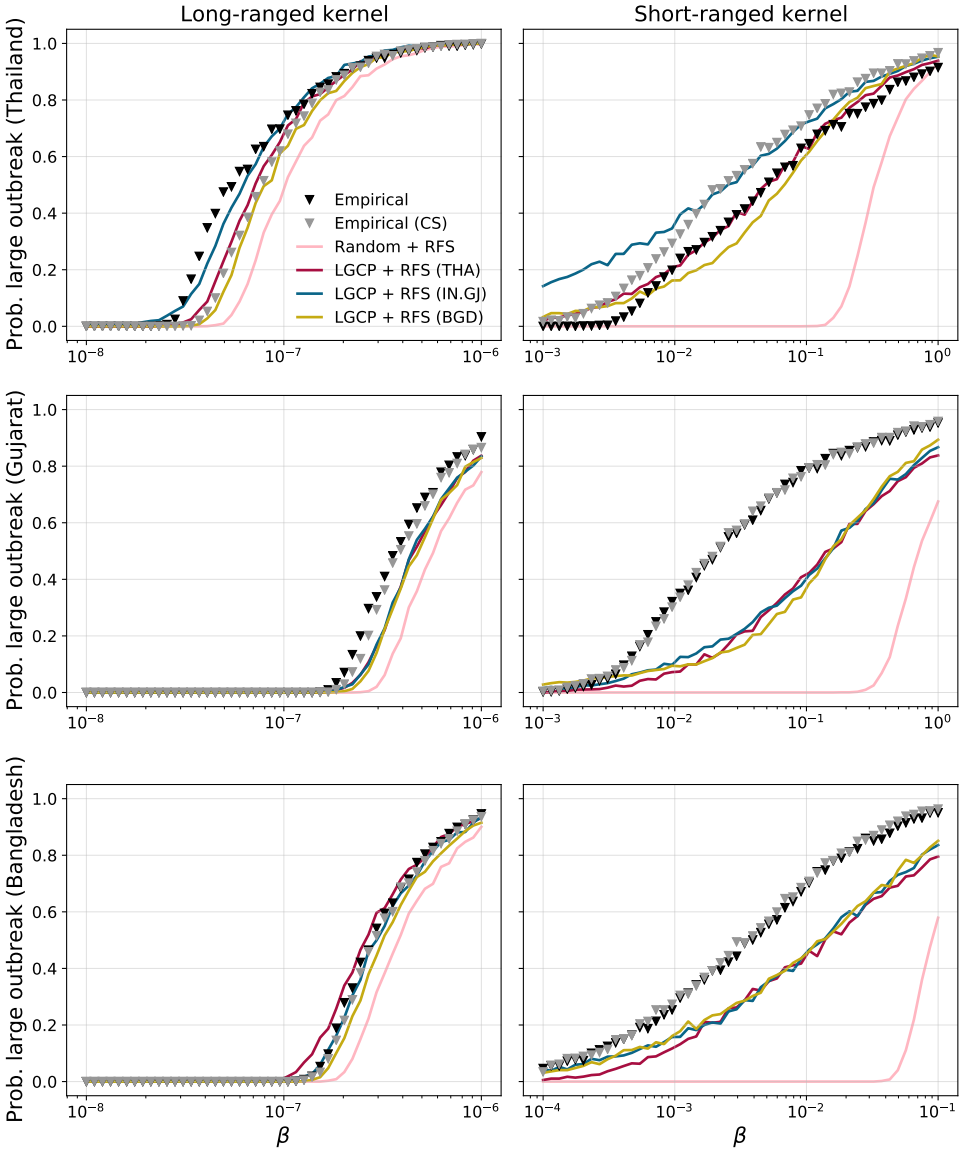

Figure 6: **Probability of large outbreak.** Average probability of large outbreak (*i.e.* the proportion of simulations where the attack rate exceeds 100 farms) as a function of transmissibility for long- and short-range kernels calculated for Thailand (first row), Gujarat (second row) and Bangladesh (last row). The curves shown include LGCP + RFS models trained in Thailand (red), Bangladesh (yellow) and Gujarat (blue). The markers denote simulations using empirical farm locations with original (black) and homogeneous (grey) farm sizes. The pink line corresponds to random farm locations with RFS-generated farm sizes. We set $\alpha = 0.643$ for long-range kernel and $\alpha = 3$ for short-range kernel. Other parameters are: $\mu = 0.143d^{-1}$, $Q_S = 1.06$, $Q_I = 0.057$, $d_{min} = 0.1$ km.

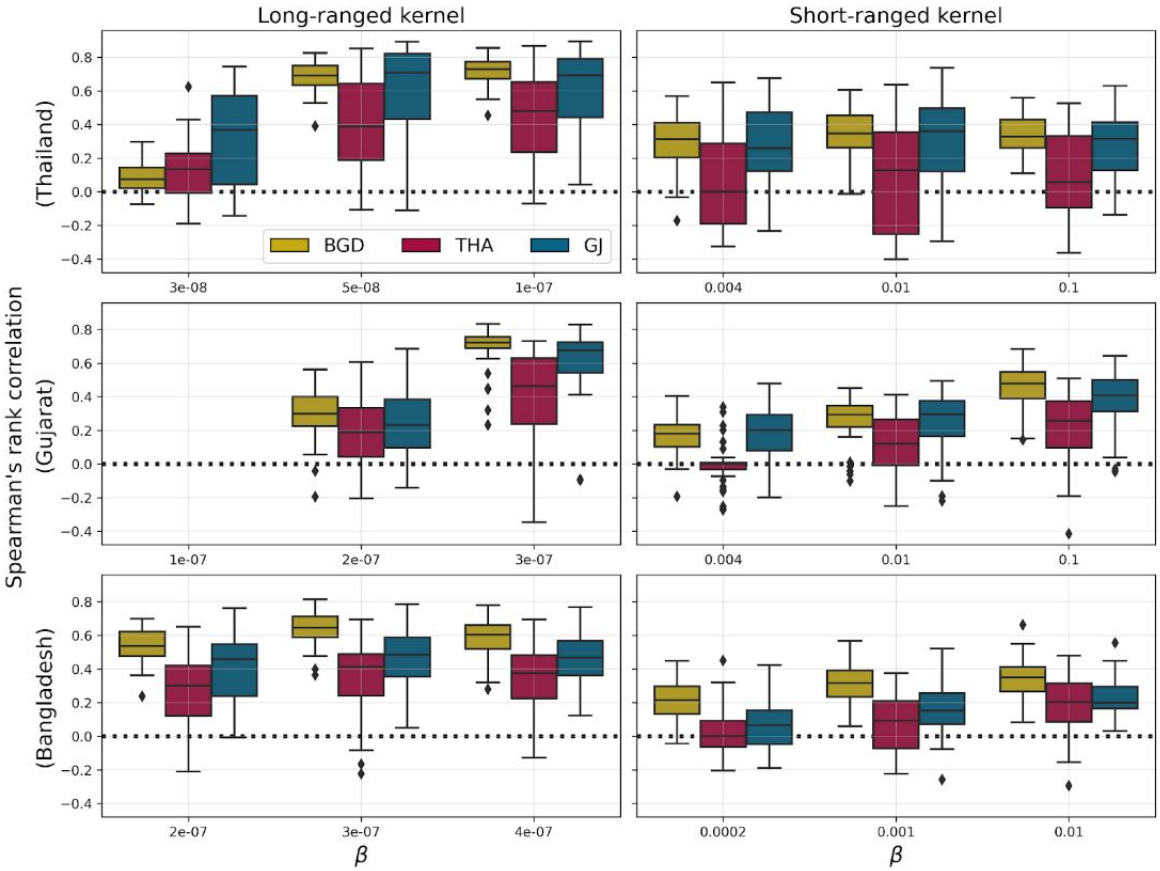

Figure 7: **Spatial risk analysis.** Boxplots show Spearman's rank correlation coefficients between gridded risk distributions for the empirical farm distributions of Thailand (top row), Gujarat (middle row) and Bangladesh (bottom row) and the LGCP+RFS models trained on each area. A single correlation coefficient is calculated for each of 40 realisations from every point pattern model. The first and second columns correspond to long- and short-ranged transmission kernels, respectively. The missing boxplots in the middle-left panel are the due to the fact that no farms became infected when $\beta = 10^{-7}$ and Spearman's correlation coefficient is not defined when all variables in one input set are the same (all equal to 0 in this case).

4. Discussion

Producing accurate spatial maps of livestock farm distributions is paramount for assessing the risk of future epidemics. This study aimed to develop farm distribution models that simulate the locations and sizes of chicken farms, while accounting for production types and spatial clustering of farms in data-scarce countries. The FDM enables the partial reproduction of clustering patterns, the locations of clusters, and farm size in external areas (i.e. when trained in an area and applied to another). In addition, the FDM outputs were used as an input in a disease transmission model to assess whether epidemic patterns are consistent with simulations using observed data.

The LGCP model produces simulated SPPs with farms clustered within specific distances (around <50 to 100 kms for all global envelope tests). The LGCP model, which includes a Gaussian Random Field to induce additional spatial correlation between points (Møller et al., 1998), allows us to maintain the level of clustering of farms distribution. In comparison, linear regression models and random forest modelling poorly reproduce levels of clustering for intensive livestock production (Van Boeckel et al., 2012). Indeed, model outputs such as density of animals per pixel do not allow for significant heterogeneity of values, despite employing log transformation of values for model calibration.

Chaiban et al. (2021) were able to reproduce the spatial clustering of farms in the same areas where their model was trained, but failed when considering further regions. In other words, their model failed external validation tests. Indeed, the areas considered in their study were highly heterogeneous in terms of geographical location and level of intensification. Here, we focus instead on areas within South and Southeast Asia with similar degrees of intensification.

We found that the same spatial predictors were able to explain all farm distributions in our study, and lead to acceptable external validation of farm locations compared to the study by Chaiban et al. (2021), where study areas were located in different continents. Similarity in production systems, driven by climatic and economic features, is therefore a crucial factor for choosing appropriate areas for model training. However, the model trained in Thailand provided the worst external validation. This could be caused by the country's geographical characteristics and the specific configuration of the production system, with the country presenting the highest GDP of our selected study areas. Indeed, economies of scale have shifted the structure of Thai poul-

try production towards industrialized systems, with fewer producers owning larger holdings (NaRanong, 2007). In addition, these structural changes have been supported by the shift from agricultural subcontracting to vertical integration, which involved the centralization of production steps by a few companies and have contributed to the clustering of poultry farms within the peri-urban belt of Bangkok. In contrast, most chickens in Bangladesh are produced in smaller units, which, for most, are not contracted by integrators but rely on credit provided by local production input suppliers to operate (Hennessey et al., 2021). In contrast, in Gujarat, farm sizes are more comparable to Bangladesh but mostly contracted, and not owned, by integrators.

Our study emphasises the importance of accounting for production types when modelling farm distributions, as the level of clustering differs between layer and broiler farms. However, the transition point in the L-function, where the clustering of farms shifts to dispersion, occurs at a similar distance for both broiler and layer farms. This reflects influence of specific country characteristics on these patterns. Indeed, for a given area, the most important predictive variables are the same for both production types.

Epidemic patterns simulated by the disease transmission model align more closely with those obtained using empirical farm distributions under long-range than short-range transmission, particularly in Gujarat. Short-range transmission models are more sensitive to the farms distribution at short distance, therefore their lower efficiency at reproducing similar epidemic patterns might be due to the lower clustering level at short distances in LGCP farm distributions. In Thailand, the discrepancy between observed and simulated farm size distributions likely impacted simulated epidemic patterns. This observation parallels findings from the 1997–1998 Classical Swine Fever epidemic in The Netherlands (Boender et al., 2014) and epidemic simulations conducted in New Zealand (Van Andel et al., 2018), where farms' sizes appeared to affect their susceptibility to infection and infectivity. Nonetheless, the LGCP model outperforms a random distribution and accurately predicts the transmissibility threshold above which a major outbreak becomes probable.

Indeed, spatial clustering increases epidemic risk by lowering said threshold (Tildesley et al., 2010; Brown and Bolker, 2004). In highly clustered point distributions, the dynamic of an epidemic strongly depends on the probability of transmission between clusters (Benincà et al., 2020). Our modelling framework can thus be used to gain insights into the vulnerability of livestock

production systems to disease outbreaks under scenarios assuming various levels of clustering and variations in cluster locations. As previously highlighted, when simulating farm distributions in a specific targeted area, the selection of the training area should be based on the similarity of production systems. In cases where there is limited evidence to guide this choice, employing models trained across diverse areas becomes beneficial. This approach generates a spectrum of epidemic patterns capturing uncertainties in the actual underlying distributions of farm locations and sizes.

A limitation of the FDM is that the number of farms cannot be fixed as a parameter, but varies with each simulation. Knowledge about the expected density (i.e. intensity) of farms in the area where the model will be used to simulate farm distributions may be a relevant indicator to select the region used to calibrate the model. Another limitation is that a large number of predictors may lead to an overestimation of the number of farms. Therefore, we tested all combinations of predictors to select the model associated with the largest AUC as an input for the disease transmission model. Assessing the performance of SPP models is non-trivial and still debated in the literature (Baddeley et al., 2014). Our results raise questions about the reliability of the envelope test p-value. Indeed, LGCP models associated with large variations in the level of clustering across simulations will be associated with large p-values as the wide enveloped would include the L-function. We argue that envelope tests should not only be interpreted based on the p-value, but also graphically.

In conclusion, the FDM, by simulating farm locations and sizes, enhances our understanding of the way these spatial patterns, especially farm clustering, influence disease spread. Such understanding is crucial for designing more effective disease control and prevention strategies tailored to local characteristics of production systems. This modelling framework is particularly relevant in resource-limited countries where the intensification of poultry production may often outpace the availability of reliable data. In such contexts, the FDM offers a forward-looking tool, enabling stakeholders to proactively assess epidemic risks associated with various intensification scenarios, and can thus serve as a pivotal tool for informing agricultural planning.

Direct applications of our study include the testing of different levels of clustering and the assessment of their impacts on epidemic spread (Benincà et al., 2020). By manipulating the degree of clustering in farm distributions and the clusters locations, policy-makers and stakeholders can gain insights into the vulnerability of livestock systems to disease outbreaks. Understand-

ing the relationship between clustering and epidemic spread is crucial for transitioning towards safer production systems.

References

Baddeley, A., Diggle, P.J., Hardegen, A., Lawrence, T., Milne, R.K., Nair, G., 2014. On tests of spatial pattern based on simulation envelopes. Ecological Monographs 84, 477–489.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial point patterns: methodology and applications with R. CRC press.

Benincà, E., Hagenaars, T., Boender, G.J., van de Kassteele, J., van Boven, M., 2020. Trade-off between local transmission and long-range dispersal drives infectious disease outbreak size in spatially structured populations. PLOS Computational Biology 16, e1008009.

Boender, G.J., Hengel, R.v.d., Roermund, H.J.v., Hagenaars, T.J., 2014. The influence of between-farm distance and farm size on the spread of classical swine fever during the 1997–1998 epidemic in the netherlands. PLoS One 9, e95278.

Brown, D.H., Bolker, B.M., 2004. The effects of disease dispersal and host clustering on the epidemic threshold in plants. Bulletin of mathematical biology 66, 341–371.

Chaiban, C., Biscio, C., Thanapongtharm, W., Tildesley, M., Xiao, X., Robinson, T.P., Vanwambeke, S.O., Gilbert, M., 2019. Point pattern simulation modelling of extensive and intensive chicken farming in thailand: Accounting for clustering and landscape characteristics. Agricultural systems 173, 335–344.

Chaiban, C., Da Re, D., Robinson, T.P., Gilbert, M., Vanwambeke, S.O., 2021. Poultry farm distribution models developed along a gradient of intensification. Preventive Veterinary Medicine 186, 105206.

De Cola, L., 1991. Fractal analysis of multiscale spatial autocorrelation among point data. Environment and Planning A 23, 545–556.

FAO, 2017. The future of food and agriculture–trends and challenges. Annual Report 296.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T.P., Vanwambeke, S.O., Wint, G., Robinson, T.P., 2018. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Scientific data 5, 1–11.

Hennessey, M., Fournié, G., Hoque, M.A., Biswas, P.K., Alarcon, P., Ebata, A., Mahmud, R., Hasan, M., Barnett, T., 2021. Intensification of fragility: Poultry production and distribution in bangladesh and its implications for disease risk. Preventive Veterinary Medicine 191, 105367.

Hill, E.M., House, T., Dhingra, M.S., Kalpravidh, W., Morzaria, S., Osmani, M.G., Yamage, M., Xiao, X., Gilbert, M., Tildesley, M.J., 2017. Modelling h5n1 in bangladesh across spatial scales: Model complexity and zoonotic transmission risk. Epidemics 20, 37–55.

Kastner, T., Rivas, M.J.I., Koch, W., Nonhebel, S., 2012. Global changes in diets and the consequences for land requirements for food. Proceedings of the National Academy of Sciences 109, 6868–6872.

Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 uk foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294, 813–817.

Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log gaussian cox processes. Scandinavian journal of statistics 25, 451–482.

NaRanong, V., 2007. Structural changes in thailand's poultry sector and its social implications. Thailand Development Research Institute. Bangkok, Thailand .

Prosser, D.J., Wu, J., Ellis, E.C., Gale, F., Van Boeckel, T.P., Wint, W., Robinson, T., Xiao, X., Gilbert, M., 2011. Modelling the distribution of chickens, ducks, and geese in china. Agriculture, ecosystems & environment 141, 381–389.

Ripley, B.D., 1976. The second-order analysis of stationary point processes. Journal of applied probability 13, 255–266.

Ritchie, H., Roser, M., 2019. Half of the world's habitable land is used for agriculture. Our World in Data Https://ourworldindata.org/global-land-for-agriculture.

Robinson, T.P., Wint, G.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S.I., Gilbert, M., 2014. Mapping the global distribution of livestock. PloS one 9, e96084.

Sellman, S., Tsao, K., Tildesley, M.J., Brommesson, P., Webb, C.T., Wennergren, U., Keeling, M.J., Lindström, T., 2018. Need for speed: An optimized gridding approach for spatially explicit disease simulations. PLoS computational biology 14, e1006086.

Steinfeld, H., 2006. Livestock's long shadow: environmental issues and options. Food & Agriculture Org.

Tanaka, U., Ogata, Y., Stoyan, D., 2008. Parameter estimation and model selection for neyman-scott point processes. Biometrical Journal: Journal of Mathematical Methods in Biosciences 50, 43–57.

Thanapongtharm, W., Linard, C., Chinson, P., Kasemsuwan, S., Visser, M., Gaughan, A.E., Epprech, M., Robinson, T.P., Gilbert, M., 2016. Spatial analysis and characteristics of pig farming in thailand. BMC veterinary research 12, 1–15.

Tildesley, M.J., House, T.A., Bruhn, M.C., Curry, R.J., O'Neil, M., Allpress, J.L., Smith, G., Keeling, M.J., 2010. Impact of spatial clustering on disease transmission and optimal control. Proceedings of the National Academy of Sciences 107, 1041–1046.

Van Andel, M., Hollings, T., Bradhurst, R., Robinson, A., Burgman, M., Gates, M.C., Bingham, P., Carpenter, T., 2018. Does size matter to models? exploring the effect of herd size on outputs of a herd-level disease spread simulator. Frontiers in veterinary science 5, 78.

Van Boeckel, T.P., Prosser, D., Franceschini, G., Biradar, C., Wint, W., Robinson, T., Gilbert, M., 2011. Modelling the distribution of domestic ducks in monsoon asia. Agriculture, ecosystems & environment 141, 373–380.

Van Boeckel, T.P., Thanapongtharm, W., Robinson, T., D'Aietti, L., Gilbert, M., 2012. Predicting the distribution of intensive poultry farming in thailand. Agriculture, ecosystems & environment 149, 144–153.

Wint, W., Robinson, T., 2007. Gridded livestock of the world 2007. FAO 636.2 W784 2007, FAO, Roma (Italia).

Zhao, Q., Dupas, M., Axelsson, C., Artois, J., Robinson, T., Gilbert, M., 2022. Distribution and intensification of pig production in china 2007–2017. Environmental Research Letters 17, 124001.

5. Acknowledgments

This study was funded by the UKRI GCRF One Health Poultry Hub (Grant No. B/S011269/1), one of twelve interdisciplinary research hubs funded under the UK government's Global Challenge Research Fund Interdisciplinary Research Hub initiative. G.F. is supported by the French National Research Agency and the French Ministry of Higher Education and Research.

A. Appendix Section: Material and Methods

A.1. Farm distribution in Bangladesh, Gujarat (state of India) and Thailand

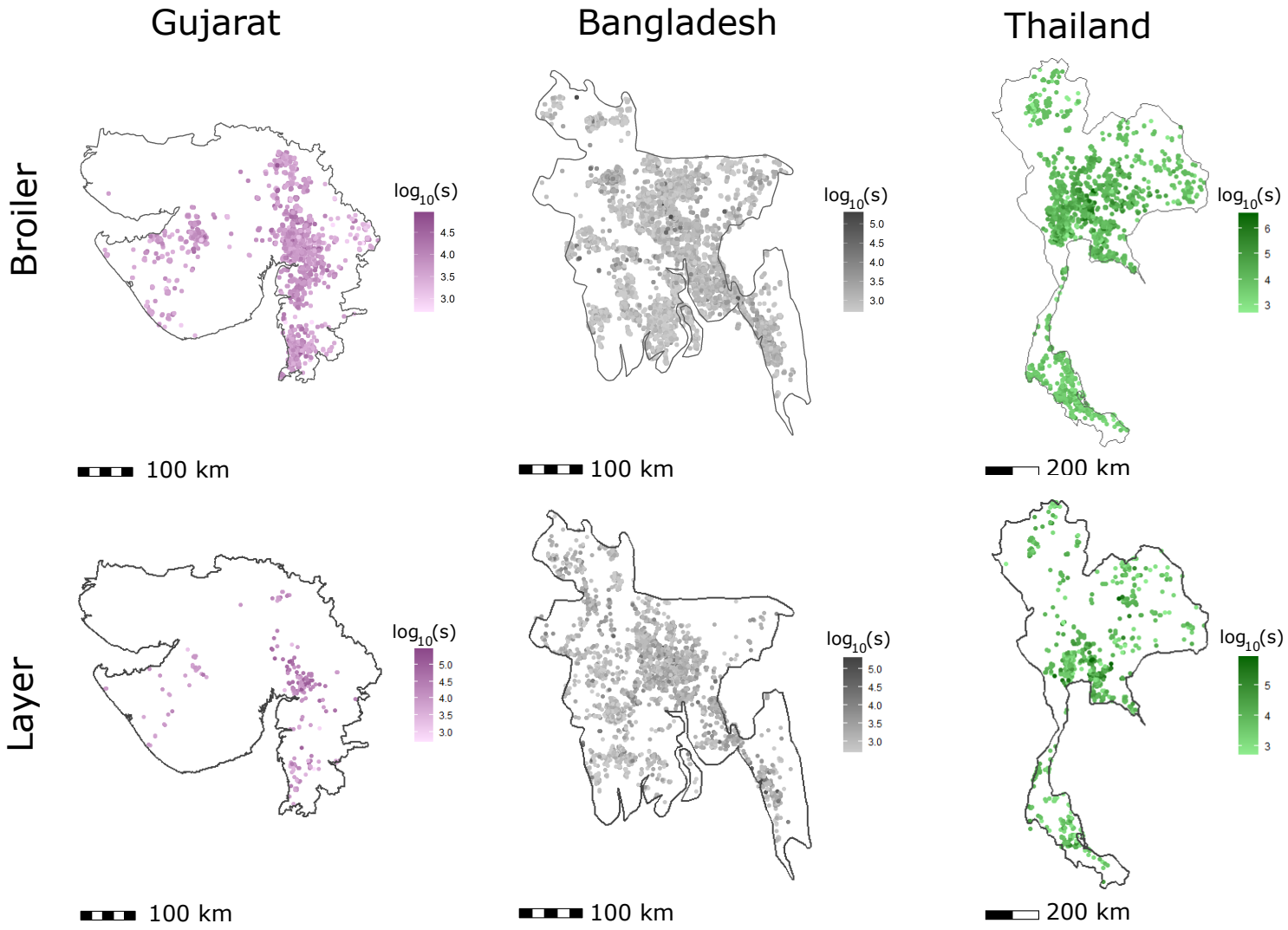

Figure S1: Spatial distribution of layers and broilers farm in Gujarat (2020), Bangladesh and Thailand (2010).

A.2. Algorithm of Global Envelope Test

The test procedure is as follows:

1. Generate a large number of simulated point patterns (1000 simulations) based on the fitted model parameters.
2. For each simulation and each distance r, calculate the test statistic ($L_{inhom}(r)$).
3. Rank the observed test statistic among the simulated values at each r.
4. Construct the global envelope based on the ranks. In this case, the envelope is constructed using the most extreme ranks from the simulations, i.e. the minimum and maximum (the 1st rank from the bottom and top, respectively) simulated values at each distance r.

A.3. Quadrat count test

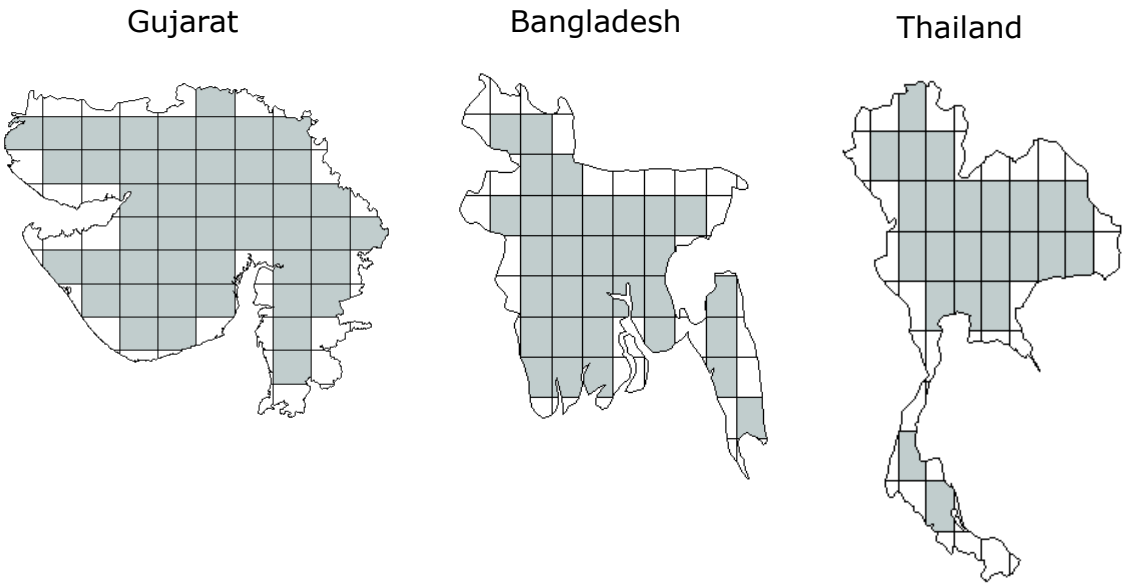

Figure S2: Window division and selection for quadrat count test. Grey tiles were used to calculate the coefficient correlation between observed and simulated points pattern. We did not consider quadrats that occupy less than 80% of the complete theoretical polygon to avoid edge effects.

A.4. Epidemic transmission modelling

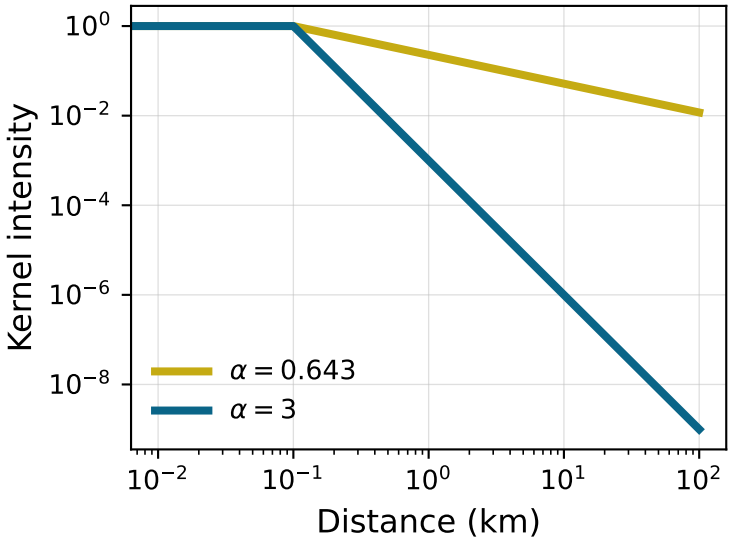

Figure S3: Short-ranged (blue) and long-ranged kernels (yellow) in a log-log scale.

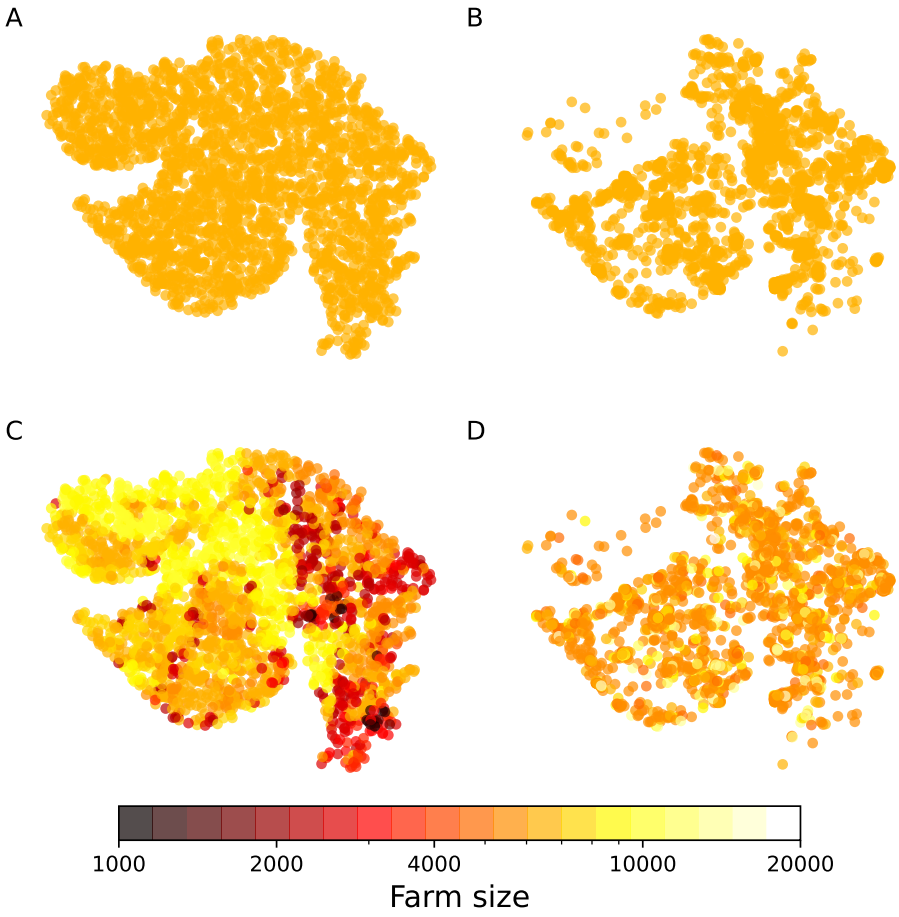

Figure S4: Scenarios for the epidemiological model in Gujarat. A. Random farms distribution with constant farm size. B. Farms distribution generated with the LGCP model with constant farm size. C. Random farms distribution with farm sizes predicted with RF model. D. Farms distribution generated with the LGCP model with farm sizes predicted with RF model.

B. Appendix Section: Results

B.1. Simulations of point patterns in Gujarat

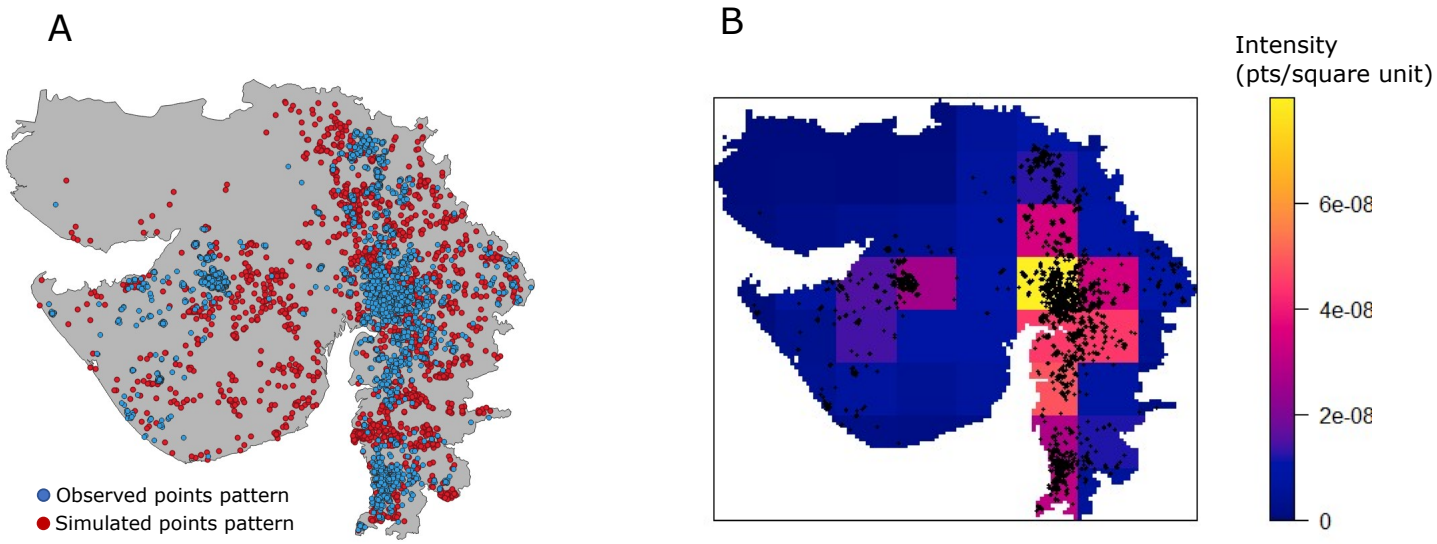

Figure S5: A. Observed points patterns of broiler farms in Gujarat and one simulated point pattern with the model trained with Bangladesh broiler farms. B. Mean intensity of points of 8000 simulations from the model trained in Bangladesh and the observed point pattern is represented with black dots.

B.2. Performance of the LGCP

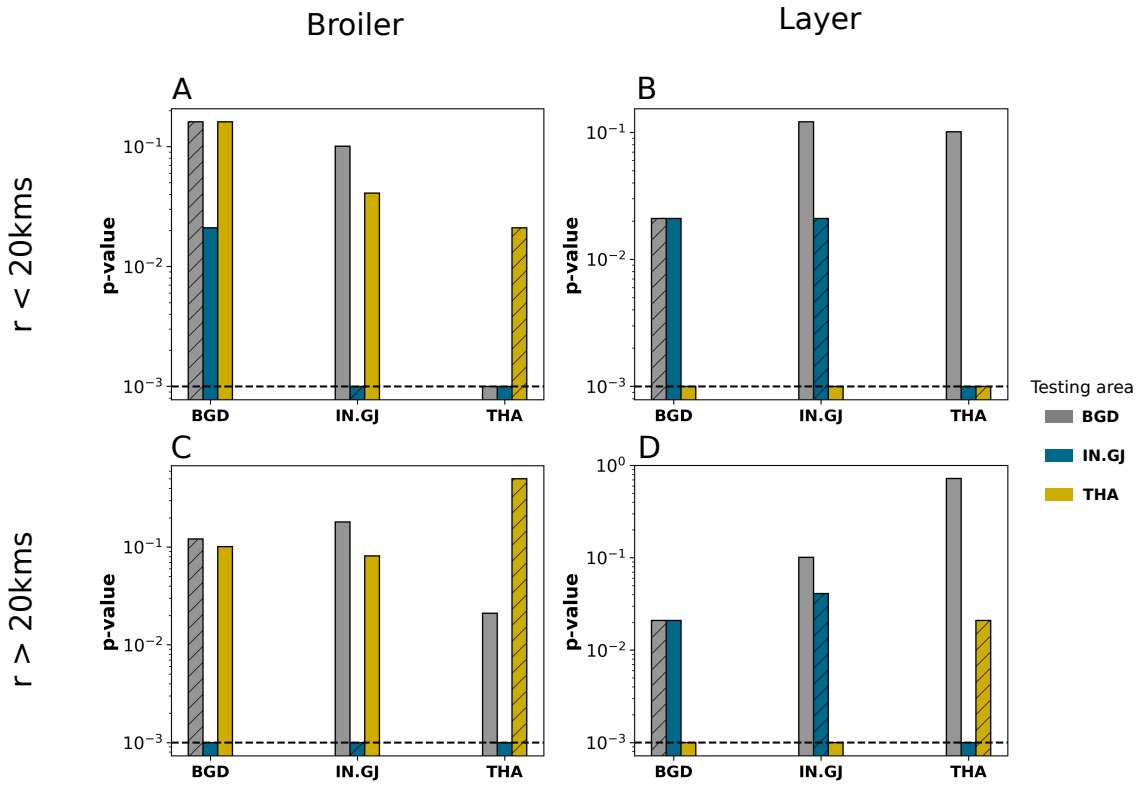

Figure S6: **Internal and external validation p-values of the global rank envelope test for the different models (Bangladesh: BGD, Gujarat: IN.GJ and Thailand: THA), for the two types of farm: broilers (A & C) and layers (B & D) and for radii under 20kms (A & B) and for radii above 20kms (C & D).** Labels on the x axis denote the training area. Hatched bars distinguish p-values for internal validation from those for external validation. The color of the bar charts indicate where the model is tested (grey for Bangladesh, blue for Gujarat and yellow for Thailand). The horizontal dashed line indicate the threshold of significance of the p-values for the envelope of 1000 simulations.

B.3. Farm size modelling

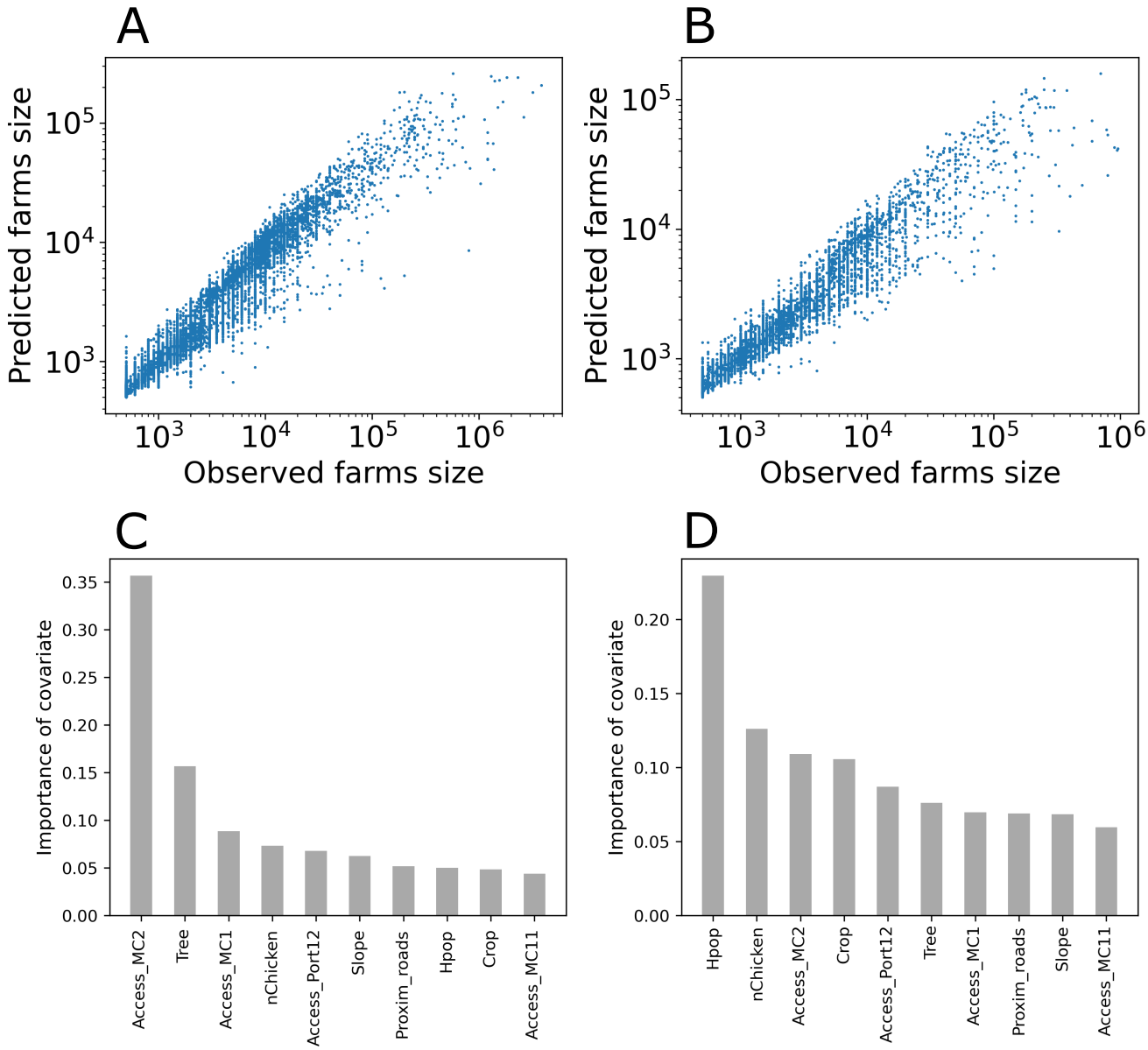

Figure S7: **A & B.** Predicted farm size in function of observed farm size (A. Broiler and B. Layer). **C & D.** Importance of each covariate for broiler farm RF model (A) and layer farm RF moodel (B).

Bootstrap	RMSE	Pearson coefficient
1	0.281	0.823
2	0.276	0.824
3	0.273	0.830
4	0.275	0.824
5	0.272	0.830

Table S4: Broiler farms.

Bootstrap	RMSE	Pearson coefficient
1	0.335	0.708
2	0.350	0.701
3	0.348	0.695
4	0.341	0.701
5	0.343	0.683

Table S5: Layer farms.

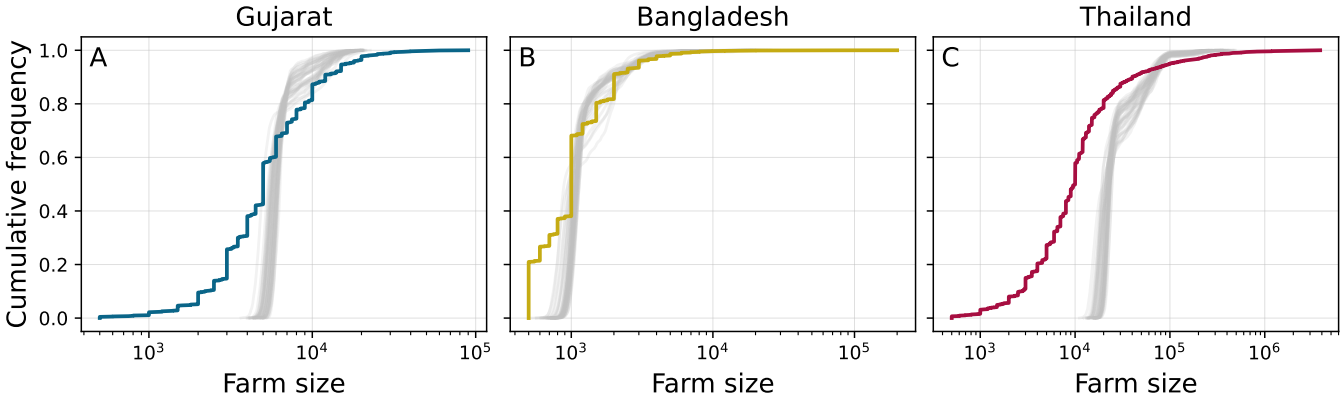

Figure S8: Cumulative empirical size farm distribution (thick line) and farm size distribution from 40 individual realisations of the LGCP+RF model (grey lines) used for the epidemic transmission modellings.

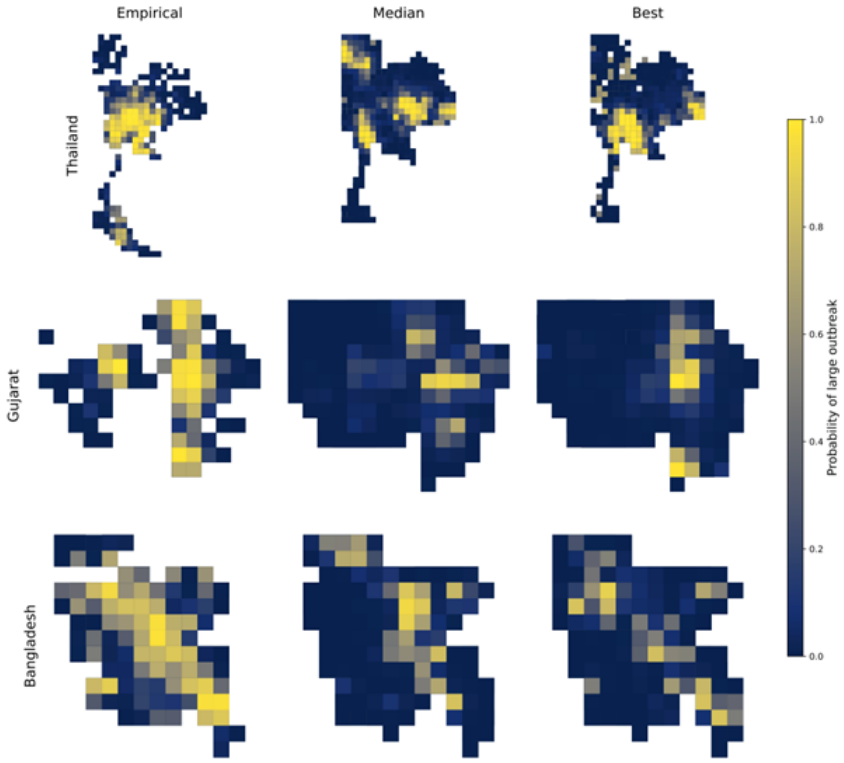

Figure S9: **Examples of epidemic risk maps.** Shown for Thailand (top row), Gujarat (middle row) and Bangladesh (bottom row). All simulations are based on a long-ranged transmission kernel and the middle β values in Figure 7. The first column shows risk calculated from using the empirical farm distribution. The second and third columns use the average- and best-performing point pattern distributions sampled from the iLGCP+RFS model trained on the same area. Performance is based on Spearman's rank correlation coefficient between gridded risk distributions. White cells contain no farms.