Poultry farm distribution models developed along a gradient of intensification

Celia Chaiban [a,b], Daniele Da Re [a,b], Timothy P. Robinson [c], Marius Gilbert [b,d,**], Sophie O. Vanwambeke [a,*]

[a] *Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium*
[b] *Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, 1050 Brussels, Belgium*
[c] *Livestock Information, Sector Analysis and Policy Branch (AGAL), Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, 00153 Rome, Italy*
[d] *Fonds National de la Recherche Scientifique (FNRS), 1000 Brussels, Belgium*

ARTICLE INFO

ABSTRACT

Efficient planning of measures limiting epidemic spread requires information on farm locations and sizes (number of animals per farm). However, such data are rarely available. The intensification process which is operating in most low- and middle-income countries (LMICs), comes together with a spatial clustering of farms, a characteristic epidemiological models are sensitive to. We developed farm distribution models predicting both the location and the number of animals per farm, while accounting for the spatial clustering of farms in data-poor countries, using poultry production as an example. We selected four countries, Nigeria, Thailand, Argentina and Belgium, along a gradient of intensification expressed by the per capita Gross Domestic Product (GDP). First, we investigated the distribution of chicken farms along the spectrum of intensification. Second, we built farm distribution models (FDM) based on censuses of commercial farms of each of the four countries, using point pattern and random forest models. As an external validation, we predicted farm locations and sizes in Bangladesh. The number of chicken per farm increased gradually in line with the gradient of GDP per capita in the following order: Nigeria, Thailand, Argentina and Belgium. Interestingly, we did not find such a gradient for farm clustering. Our modelling procedure could only partly reproduce the observed datasets in each of the four sample countries in internal validation. However, in the external validation, the clustering of farms could not be reproduced and the spatial predictors poorly explained the number and location of farms and farm sizes in Bangladesh. Further improvements of the methodology should explore other covariates of the intensity of farms and farm sizes, as well as improvements of the methodology. Structural transformation, economic development and environmental conditions are essential characteristics to consider for an extrapolation of our FDM procedure, as generalisation appeared challenging. We believe the FDM procedure could ultimately be used as a predictive tool in data-poor countries.

1. Introduction

Population and income growth, urbanisation and technological advances have led to the intensification of livestock production systems over the past decades. Although this trend in intensification is observable globally, the level of intensification varies between countries and it mainly occurs today in low- and middle-income countries (LMICs). Intensification leads to large-scale and specialized production units, market-oriented and capital- and input-intensive. Such production units make heavy use of feed concentrates, antibiotics and vaccination. The increase in production meets the increasing demand for animal source food, but it has tremendous effects on livelihoods, animal and human health. Concentration of monogastric species have led to pollution of land and water due to the overload of improperly managed manure (Gerber et al., 2005). In terms of public health, intensive livestock production fosters the emergence of infectious diseases (Jones et al., 2013) and often leads to increased use of antibiotics, which is likely to increase antimicrobial resistance (Aarestrup, 2005; Vieira et al., 2011;

---

* Corresponding author.
** Corresponding author at: Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, 1050 Brussels, Belgium.
E-mail addresses: mgilbert@ulb.ac.be (M. Gilbert), sophie.vanwambeke@uclouvain.be (S.O. Vanwambeke).

Chantziaras et al., 2014; Van Boeckel et al., 2015). Livestock also traditionally represents an important source of livelihood for rural populations in LMICs but the intensification of livestock production seldom benefits the local population (FAO, 2018).

As production intensifies, the geographical distribution of monogastric animal farming changes. At a pre-industrial stage, chickens and pigs are mainly associated with human settlements for their role of waste converters (Perry et al., 1999; Steinfeld et al., 2006), with farms homogeneously distributed among rural populations. This situation is encountered in traditional production systems of LMICs (Steinfeld et al., 2006). Production intensification, in parallel to urbanisation, leads larger farms to locate within the peri-urban belt of major consumption centres (Steinfeld et al., 2006). Livestock products being perishable, the production profits from being located close to demand centres. However, as cities expand, land value in the periphery increases and, with improving transport infrastructures, farms move further from peri-urban belts, with the advantage of being closer to feed production or transportation areas (Steinfeld et al., 2006), and benefits from agglomeration economies (Roe et al., 2002a; Herath et al., 2005; Larue et al., 2011). This evolution is particularly pronounced for monogastric species, which are raised in landless production systems (Naylor, 2005; Steinfeld et al., 2006).

As the distribution of farms changes as it intensifies, so do the effects of livestock raising activities on their surroundings. Highly-detailed maps of livestock production allow the assessment of the local effects of intensive production (Robinson et al., 2011). In the case of disease outbreaks for instance, adequate control is planned based on mathematical modelling of infectious disease which require farm distribution and sizes (number of animals) (Hill et al., 2018; Tildesley and Ryan, 2012). Accurate data on the distribution and stock of farms rarely exist as agricultural censuses are time and resource consuming. In high-income countries, censuses are usually conducted but data access may be restricted due to confidentiality and privacy reasons. In the United States for instance, livestock data are only provided as a number of animal per county (Tildesley et al., 2010). In LMICs, censuses are not systematically carried out and data from occasional censuses are available as total numbers of animals per administrative units. However, the resolution at which these are provided vary greatly across countries (Robinson et al., 2011).

Downscaling procedures that increase the spatial resolution of existing livestock data, such as the Gridded Livestock of the World (GLW) are available (Wint et al., 2007; Robinson et al., 2014; Gilbert et al., 2018). However, these models predict livestock as a continuous, gradually varying, density of animals per pixel in a raster. Such data do not provide information at farm level, and densities of animal per pixel do not specify any information on the production systems. As production intensifies, livestock are fed on imported feed concentrates rather than on crops produced locally. Livestock production is hence less determined by environmental characteristics of the surrounding lands (Naylor, 2005; Steinfeld et al., 2006). Moreover, the spatial clustering of intensive farms is poorly explained by spatially continuous surfaces (Van Boeckel et al., 2012; Robinson et al., 2014). As suggested by Gilbert et al. (2015), a method accounting for the spatial clustering of intensive farms would be required to predict their distribution. In a previous study, we developed farm location models in Thailand, based on point pattern analysis methods (Chaiban et al., 2019). Other authors have predicted cattle population on individual farms in New Zealand with a zero-inflated Poisson regression, but only carried out an internal validation (van Andel et al., 2017). Burdett et al. (2015) modelled farm locations based on detailed geographical information as well as total animal and number of farms at county level. In this study, we present a methodology to predict both the location and the size of farms even where spatial information is scarce and which would account for the spatial clustering of farms.

We selected four countries to develop our farm distribution models, i. e. Nigeria, Thailand, Argentina and Belgium (hereafter referred to as the

"sample countries"). These sample countries were assumed to have different levels of intensification as they have increasing levels of Gross Domestic Product (GDP) in the following order: Nigeria, Thailand, Argentina and Belgium. The GDP per capita (in purchasing power parity) is strongly associated with the level of intensification of livestock production of a country (Robinson et al., 2011; Gilbert et al., 2015). As farms tend to be more clustered as production intensifies (Abdalla et al., 1995; Roe et al., 2002a), we hypothesized that intensive farms would be spatially more clustered with increasing GDP. In addition to being at different stages of the intensification process, these countries present different ecological, demographic and socio-economic conditions. Developing farm distribution models (FDMs) along a wide range of conditions allows to assess how these model compare to each other in very different settings.

We first aimed to investigate the spatial distribution of commercial farms along the spectrum of intensification. Second, we built FDMs for intensive farms while accounting for the spatial clustering of farms. We then assessed the capacity of our modelling procedure to reproduce the farm distribution observed in Bangladesh, where farm locations and sizes were available.

2. Material and methods

2.1. Response variable and predictors

The modelling procedure used data on size (number of chickens per farm) and location of all intensive chicken farms in each country. As these data arose from different agricultural censuses, the criteria used to consider a farm as "intensive" may differ. Besides, the number of chickens recorded for each farm differed from one country to another. It can represent the number of chickens present at the time of the census or the maximum capacity. Details on the different datasets are presented in Table 1. All geographic coordinates originated from an agricultural census except in the dataset from Thailand where the farms lacked coordinates but were linked to each village they belong to. Coordinates were generated by randomly distributing farms within each village as done previously (for more details see Chaiban et al. (2019)). In our models, we considered farm size, i.e. the number of chickens per farm, as the maximum capacity of the farm. Farms with no chicken recorded were thus removed. Different farms having identical geographic coordinates were kept in the datasets (the dataset from Nigeria includes 440 farms having common coordinates with another farm, and there are 2226 and 6 farms in the datasets from Argentina and Belgium respectively).

Among the main factors known to influence poultry distribution and population (Van Boeckel et al., 2012), we selected spatial predictors available at global extent (Table 2): (i) log10-transformed human population density (human population density values were incremented by one to avoid infinite values of the logarithm in the case of null human population density), (ii) accessibility covariates (iii) percentage of tree cover and (iv) percentage of cropland. Three accessibility layers were computed based on the global map of travel time from Weiss et al. (2018), i.e. accessibilities to the capital city, to main cities and to harbours. The geographical coordinates of capital cities, main cities and harbours were used to calculate a travel time (at pixel level) for the optimal path (the shortest in time) between each pixel and these points of interest. This travel time is computed based on a friction surface which uses variables such as road properties, railroads, bodies of water, rivers, topography, land cover and national borders (Weiss et al., 2018). As intensive chicken production is mainly based on imported feed (Steinfeld et al., 2006; Neumann et al., 2009; Mulatu and Wossink, 2014), the proximity to harbours was considered as a favourable factor for farm location. We obtained harbour locations from the World Port Index database. Accessibilities to the capital city and main cities were included as they represent the demand centres. The main cities within each country were defined as all cities with a population higher than a

Table 1

Farm data information and averaged gross domestic product per capita per country in purchasing power parity (GDP ppp 2010–2017).

Country	Census year	Number of chicken	Location source	"Intensive farm" definition criteria	Number of farms	GDP ppp
Nigeria	2010	Number of chickens - unspecified	Agricultural census	Not communicated	4958	5406
Thailand	2010	Number of chickens - unspecified	Random distribution within villages	$\geq$ 500 chickens / farm	6587	14792
Argentina	2017	Maximum capacity	Agricultural census	Not communicated	7587	19060
Belgium	2017	Maximum capacity	Agricultural census	Not communicated	1042	41578
Bangladesh	2010	Number of chickens - unspecified	Agricultural census	Not communicated	53911	2938

Table 2

Spatial predictors included in the modelling procedure.

	Initial resolution (m)	Units	Source
Human population density	500	Log10 people per hectare	WorldPop (www.worldpop.org), 2015 (WorldPop, 2017, 2016, 2015, 2014) (unavailable in Belgium, Landscan was used instead)
Accessibility	1000	Minute	Weiss et al. (2018)
Cropland	1000	Pixel % covered by crops	Fritz et al. (2015)
Tree cover	30	Pixel % covered by forest	Hansen et al. (2013)
Gross domestic product (GDP) per capita in purchasing power parity	One value/ country	–	World Bank

threshold based on the estimated population by city in 2000 provided by the CIESIN (Balk et al., 2006). We selected as threshold the population corresponding to the 97th percentiles of the distribution of the estimated population by city. It resulted in a number of 6, 8, 7, 34 and 18 main cities in Nigeria, Thailand, Bangladesh, Argentina and Belgium, respectively. The corresponding population thresholds were 1 043 211, 166 989, 274 495, 227 115, and 73 994, respectively. All predictors were resampled to 1 km using a bilinear interpolation (*resample()* function from *raster* R package).

The Gross Domestic Product (GDP) per capita in purchasing power parity (ppp) at country level was added in the farm size model along with the other spatial predictors since GDP was shown to be related to the proportion of animals raised in intensive systems within a country (Gilbert et al., 2015). We used the average of the GDP per capita over the period from 2010 to 2017 in ppp for 2011 (Table 2). The GDP ppp was used to allow comparability between countries.

All analyses were conducted with the projections World Geodetic System (WGS) 84 – Universal Transverse Mercator (UTM). This system of projection divides the world into 60 zones, used to project the co-ordinates in individual zones. We selected the appropriate zones for each country: 32 N, 47 N, 45 N, 20S and 31 N respectively for Nigeria, Thailand, Bangladesh, Argentina and Belgium. The Global Administrative Unit Layer (GAUL) database from 2013 was used to define country boundaries. To avoid computational problems with the point pattern analysis method, islands were excluded and we considered only the continental landmass. Country boundaries were simplified with the function *simplify.owin(dmin = 1000)* from *spatstat* package in R, in which *dmin* indicates the smallest permissible length of a polygon's edge in meters.

2.2. Characterisation of the spatial pattern of intensive farms – descriptive analysis

We used the L-function to characterize and compare the distribution of intensive farms in our sample countries. The Besag's L-function

(Besag, 1977) is a summary statistic allowing the characterisation of the spatial distribution of point locations (here, farm locations) at various scales (Baddeley et al., 2015). In order to characterise a distribution of farms, the empirical L-function has to be compared with the corresponding L-function in the case of a random distribution of farms. As farms tend to cluster with intensification, we would expect the intensity of points to vary across the space, and should therefore be described with the inhomogeneous L-function. The distribution is considered as clustered, regular or random if the empirical L-function is above, below, or close, respectively, to the expected L-function of a random distribution. A random distribution refers in this paper to a complete spatial randomness (CSR).

2.3. Statistical analysis

The modelling procedure to predict location and farm size was separated in two successive steps. The comprehensive procedure is summarized in Fig. 1.

2.3.1. Farm location modelling procedure

The procedure for modelling farm locations followed a methodology previously developed based on the point pattern analysis method (Chaiban et al., 2019). We used a Log-Gaussian Cox Processes (LGCP) model with the Palm maximum likelihood method of parameters optimisation, which models clustered distribution of points. Spatial predictors were included in the LGCP model so that the intensity of points could be modelled according to landscape and anthropogenic characteristics. The intensity of points is *inhomogeneous* and modelled through an intensity function defined as

$$\lambda(u) = \exp\big(\beta_0 + \beta_1 pred_1(u) + \beta_2 pred_1{}^2(u) + \beta_3 pred_2(u) + \beta_4 pred_2{}^2(u) + \ldots$$
$$+ \beta_{k-1} pred_n(u) + \beta_k pred_n{}^2(u)\big)$$

where u is any location in the study, β_0, β_1…, β_k are the parameters to be estimated, *pred* the predictors and n the number of predictors. The forward stepwise selection of the covariates, was based on the Akaike Information Criterion (AIC). The model with the lowest AIC was selected for each country. Each predictor was added with its quadratic term to account for non-linear effects. The relative importance of each covariate within a LGCP model is obtained by taking the exponential of the estimated parameter of a covariate multiplied by the range of values of that covariate (Baddeley et al., 2015). Our model was hereafter referred to as "iLGCP", for inhomogenous LGCP.

Point pattern models require distribution of points from a contiguous area to fit the model and capture the spatial structure of the observed point pattern. Such models can therefore not be fitted on non-adjacent countries. One iLGCP model was hence fitted separately for each country, with the *kppm (clusters = "LGCP", method = "palm")* function from the *spatstat* package in R (Baddeley et al., 2015). Based on each respective fitted iLGCP model, the distributions of farms could be simulated in each country, using *simulate.kppm()* function form *spatstat* package. As a different distribution is produced each time a simulation is run, a fitted model can therefore be considered as a simulator of point patterns.

The iLGCP models per country were first internally assessed in their ability to reproduce realistically the distributions of farms they were
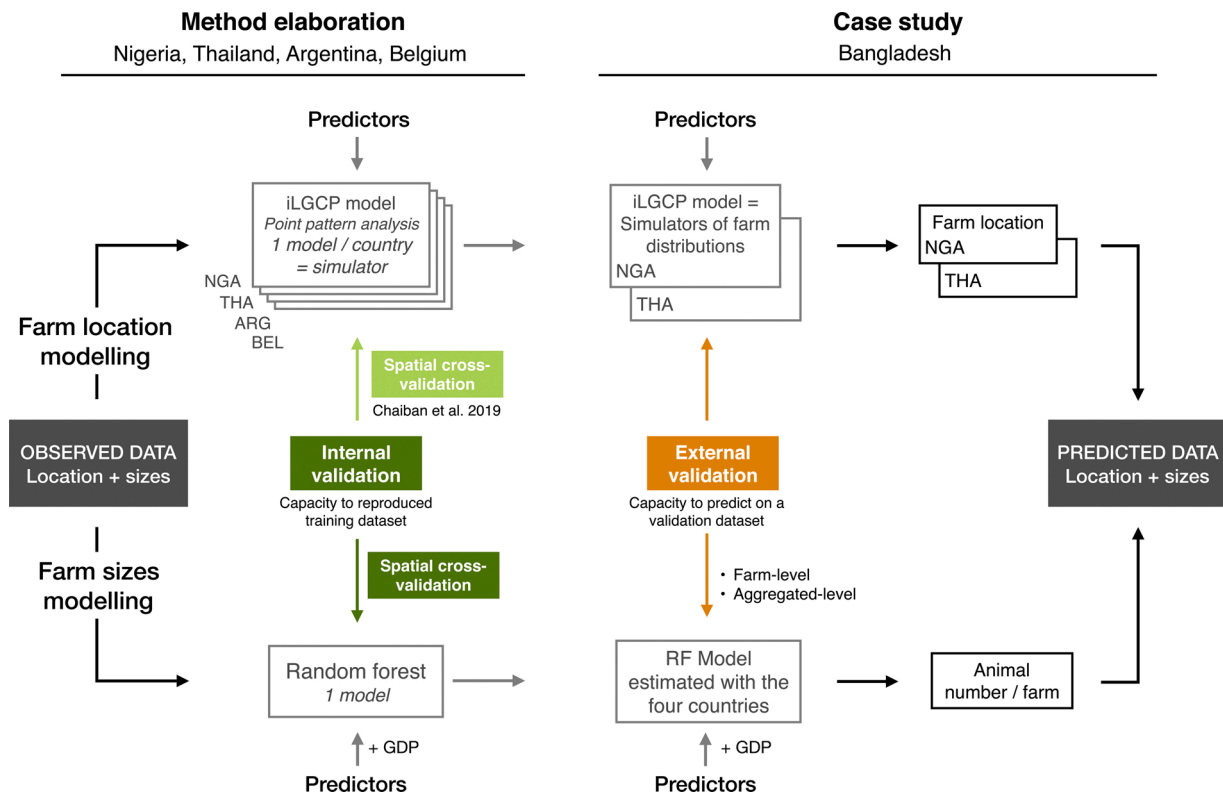
Fig. 1. Modelling procedure, which includes two successive steps: farm location modelling (based on point pattern model (an inhomogeneous Log-Gaussian Cox Processes – iLGCP)) and farm size modelling (based on a random forest model – RF).

fitted on. To do so, the iLGCP models were compared to three other point pattern models: (i) "CSR", a spatially random model, without covariates (i.e. homogeneous intensity of points), which distributes the points randomly; (ii) "iCSR", an inhomogeneous random model in which the average density of points is spatially varying and explained with covariates (the same set of predictors as the ones used in the iLGCP); and (iii) "LGCP", a clustered model, Log-Gaussian Cox Process, with a homogeneous intensity, without any covariate. The functions *ppm()* and *kppm()* from the *spatstat* package were used to generate CSR models and LGCP models, respectively. We used two goodness-of-fit indices (GoF) to compare the models. As a first GoF, we selected the two-sided global rank envelope (GRE) test to assess the ability of a model to reproduce distributions similar to the observed one. The null hypothesis tests whether the model predicts a distribution similar to the observed distribution. It builds an envelope around the different simulations provided to the test, to which a p-value and graphical interpretation are associated. The p-value decreases as the empirical L-function goes out of the envelope. The global rank envelope test (GET) was implemented with 8000 simulations and the extreme rank lengths (erl), which allowed to run fewer simulations (Mrkvička et al., 2016; Myllymäki et al., 2017). It was carried out with the *global_envelope_test(x, type= 'erl')* function from the *GET* package in R (Mrkvička et al., 2017; Myllymäki et al., 2017). The number of simulations was selected according to preliminary analyses performed on Belgium (Supplementary Material Fig. S1), where 8000 simulations could balance the time required to run the simulations while obtaining a stable p-value. As a second GoF, we used the Pearson correlation coefficient to assess how accurately the model could locate the farms. Each country was divided into 16 quadrats, and the coefficient of correlation was computed between the observed and the simulated number of farms per quadrat.

2.3.2. Farm size modelling procedure

In order to predict farm sizes, we used a random forest model, shown to produce good levels of accuracy to predict population in the GLW

(Nicolas et al., 2016; Gilbert et al., 2018) and in WorldPop (Gaughan et al., 2013). The two parameters that drive the performance of random forest are the number of trees and the minimum number of variables randomly selected at each splitting point. We used 500 trees, considered as a good rule of thumb (Lawrence et al., 2006) and a minimum number of variables (*mtry*) of two. The *mtry* value was chosen according to the optimising *tuneRF()* function from the *randomForest* package. It was trained on the entire dataset, made of the datasets of the four sample countries. The dependent variable was the log10-transformed number of chickens per farm as it showed a distribution close to normal. We used the entire set of predictors tested in the point pattern model and the GDP per capita. As goodness-of-fit indices, we considered the variance explained and the Pearson correlation coefficient between observed and predicted farm sizes on the entire dataset.

We implemented an internal spatial cross-validation procedure to assess the predictive capacity of the model on a validation dataset (i.e. a dataset not used in training the model). The following procedure was implemented 100 times for each country to create a training and a validation dataset. Ten points were randomly distributed within a country, based on which Voronoi polygons were constructed. All farms within four randomly selected Voronoi polygons constituted the validation dataset. The rest of the farms within the country were included along with the data of the other countries into the training dataset. We considered the average of the correlation coefficient between observed and predicted log10 number of chickens per farm, on the training dataset and on the validation dataset.

2.4. External validation - case study of Bangladesh

2.4.1. Farm location modelling

We validated externally the capacity of our protocol to predict farm locations and sizes by applying it to Bangladesh. Since point-pattern models can only be trained on a continuous territory, only one of our sample countries could be selected to subsequently simulate farm

distribution in Bangladesh. We assessed our procedure using two models, fitting one on the Thai data and the other on the Nigerian data, as they were the closest to Bangladesh in GDP per capita among the four countries (Table 1). We used the same GoF indices, the global rank envelope test (comparing the iLGCP model with its corresponding random model, the iCSR), and the correlation coefficient by quadrats.

2.4.2. Farm size modelling

We trained the random forest on the entire four countries dataset, to predict farm size based on the set of predictors in Bangladesh. To assess the ability of the model to predict farm sizes in Bangladesh, we predicted farm sizes on the real farm locations in Bangladesh, and computed a correlation coefficient between observed and predicted sizes. The model was also evaluated at a coarser resolution, aggregating the farms at district level (second administrative level) and at an artificial finer level. The artificial level was created with Voronoi polygons constructed around 3000 points randomly distributed within Bangladesh. The average area of the Voronoi polygons was 46 km², and the average area of the second administrative level was 307 km². The Spearman and Pearson correlation coefficient and RMSE were computed on the log-10 transformed of the number of chickens.

3. Results

3.1. Characterisation of spatial distribution of intensive farms

Intensive chicken farms were clustered in all countries. However, the level of clustering differed across countries, as highlighted by the L-function values (Fig. 2). Chicken farms in Argentina had the most clustered distribution. In Nigeria and in Thailand, chicken farms were also clustered but less so than in Argentina. While still clustered, chicken farms in Belgium were closer to a random distribution.

3.2. Internal validation - Farm location modelling

The combinations of predictors showing the lowest AIC in the iLGCP model (SM Fig. S2) by country are presented with the parameters
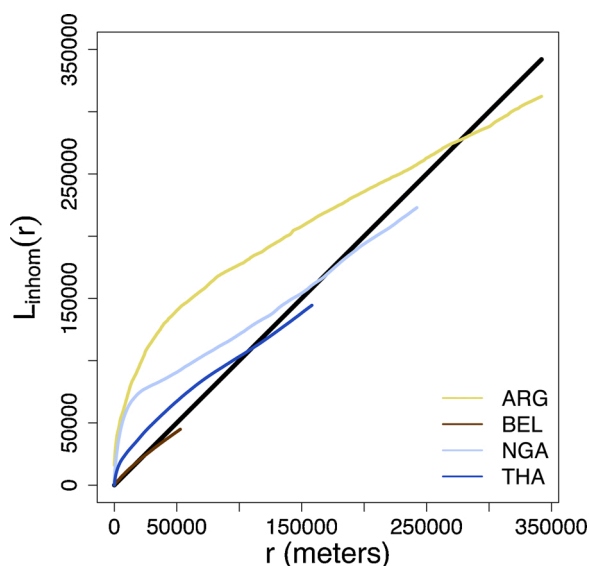
Fig. 2. Descriptive analysis of the observed farm datasets using the homogeneous and inhomogeneous L-functions. Dashed lines represent the stationary L-function (Lest) and color lines the non-stationary L-function. Both L-functions were estimated by country. The black line represents the theoretical L-function of a the completely spatial randomness (CSR) process, i.e. a random distribution of the points. A L-function above the random case translate a cluster distribution while a L-function below shows a regular pattern.

estimated for each predictor in SM Table S1. The model for Argentina did not include the accessibility to the capital city and the tree cover was absent from the Thailand model.

All models were rejected by the global rank envelope test (p < 0.05) (Fig. 3 a). The iLGCP model performed better than any other model tested in reproducing the level of clustering and farm locations although it could not explain the observed data. The iLGCP model predicted the level of clustering better than the other models in Nigeria, Thailand and Belgium as indicated by the higher p-values (Fig. 3 a), the better envelope shapes and the lower variability (i.e. narrower envelopes) than LGCP models (Fig. 4, SM Figs. S3, S4 and S5). The p-value of the LGCP model was higher in Argentina, but similar to the iLGCP model. The iLGCP model seemed characterising the level of clustering better also in Argentina, as the global rank envelope followed the shape of the observed L-function well, with a smaller variability than the LGCP model (Fig. 4). iLGCP models performed poorly at short range distances (below 10 km), where farms were more clustered than predicted, especially in Argentina (Fig. 4) and Thailand (SM Fig. S5). This was, however, less so in the models for Belgium and Nigeria (SM Figs. S3, S4), which had the highest p-values (Fig. 3 a). In all countries, both inhomogeneous models (i.e. with covariates), iCSR and iLGCP, located the farms better, as shown by the highest correlation coefficients (Fig. 5a). The iCSR models had higher correlation coefficients with a lower variability than iLGCP models. However, accounting for both the level of clustering and the location of clusters, the iLGCP model reproduced the distribution of farms better in each country.

The relative importance of each predictor and its shape varied between countries (Fig. 6 and SM Fig. S6), but human population density and accessibility to harbours and capital city had higher weights everywhere. Human population density had a high weight in the model for Nigeria and was the second or the third variable in the other countries. The accessibility to harbours had a strong importance in the models for Thailand and Argentina. Accessibility to the capital had the highest weight in Belgium and had a high weight in Nigeria, following closely human population density. The accessibility to main cities had a lower weight than the other accessibilities but had still a higher weight than cropland and tree cover in all countries except Belgium. Cropland and tree covers had low weights in all country models, but in Belgium, cropland was the second main predictor.

3.3. Internal validation – farm size modelling

The log10-transformed distributions of the number of chicken per farm by country were almost unimodal (Fig. 7). The mode increased following the gradient of intensification, from Nigeria, through Thailand and Argentina to Belgium.

The random forest model trained on the four sample countries explained the observed data well, with an explained variance of 64 % and a correlation coefficient of 0.80 between predicted and observed farm sizes. However, in the spatial cross-validation, the correlation coefficient between observed and predicted had a much lower average of 0.20. The importance of each predictor in the random forest model trained on all commercial farm size of the sample countries is provided in SM Fig. S7.

3.4. External validation – prediction on Bangladesh

All models underestimated the number of farms observed in Bangladesh (53 911 farms). The models fitted on Nigeria predicted on average 7922 and 8093 farms with the iLGCP and iCSR models, respectively. The models fitted on Thailand predicted an average of 2518 and 2569 farms with iLGCP and iCSR models. Our models reproduced poorly the level of clustering and the clusters locations. Both iCSR and iLGCP models had a non-significant p-value of the global rank envelope test (Fig. 3b), and lower correlation coefficients by quadrats than in the internal validation (Fig. 5b). Despite these poor performances, the
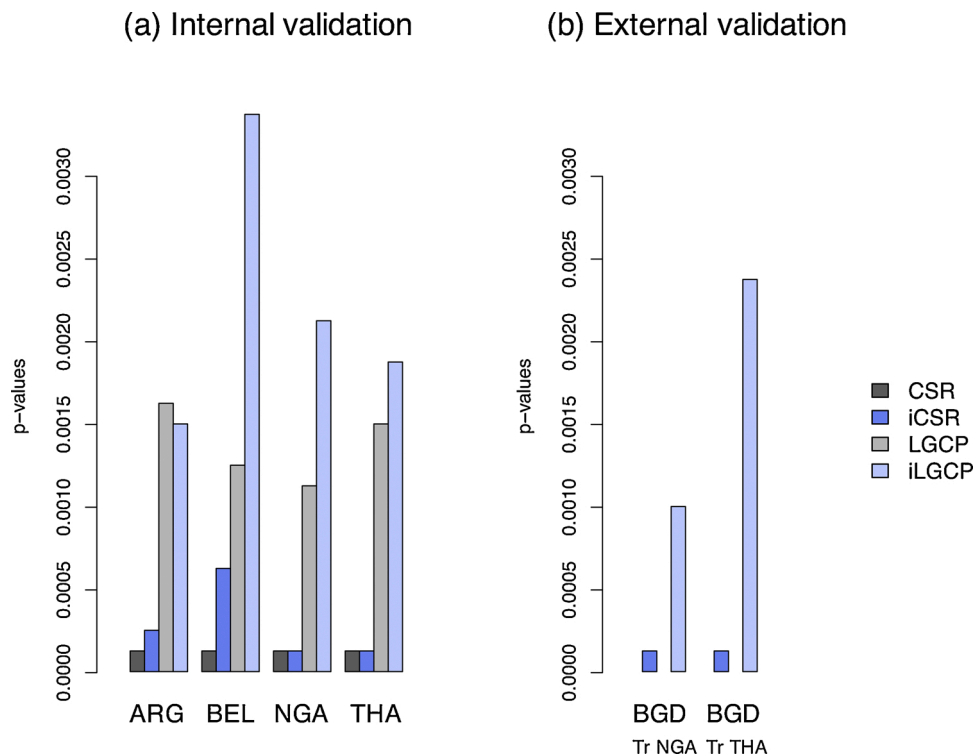
(a) Internal validation   (b) External validation

Fig. 3. Global rank envelope test with extreme rank lengths results. (a) Internal validation p-values of the global rank envelope test for the different models (CSR, iCSR, LGCP iLGCP) by country (Nigeria, Thailand, Argentina and Belgium). (b) External validation p-values of the global rank envelope test the iLGCP and iCSR models fitted on Nigeria (BGD Tr NGA) or Thailand (BGD Tr THA) to simulate farms in Bangladesh.

iLGCP model could still produce simulations with a spatial clustering closer to the observed distribution compared to the random model. This is suggested by the higher p-value of the iLGCP model (Fig. 3b) and the graphical interpretation of the global rank envelope (SM Fig. S8). Moreover, the simulations were also closer to the observed distribution of farms in Bangladesh than to the observed distribution of farms in Nigeria or Thailand (SM Fig. S8). Although Bangladesh is closer to Nigeria in terms of GDP, the iLGCP model reproduced better the spatial clustering when fitted on Thailand (SM Fig. S9).

The random forest model trained on the entire dataset (Nigeria, Thailand, Argentina and Belgium) predicted poorly farm sizes, as indicated by the low correlation of 0.14 between the observed and the predicted farm size in Bangladesh. Our model could reproduce a distribution of farm sizes similar to the observed distribution in Bangladesh despite a shrinkage of the value range (SM Fig. S9). Predictions were more accurate when aggregated, with Pearson correlation coefficients of 0.90 and 0.93, at the Voronoi level and administrative level 2, respectively (SM – Table S2). RMSE also showed a better performance of the model at the aggregated levels, with decreasing values from farm level to administrative level 2 (SM – Table S2).

4. Discussion

The paper aimed to (i) study the distribution of intensive farms along a gradient of intensification and (ii) investigate farm distribution models that would predict both location and size of chicken farms, while accounting for spatial clustering of farms in data-poor countries. Our model could reproduce only partly the location and size of farms in the observed data and poorly in external validation on Bangladesh. Despite some disappointing results, we believe our results identify potential for this methodology to bring a valuable contribution in countries with scarce data upon further improvements.

Intensive farms were clustered in all sample countries, likely as a result of the intensification process. The intensification of monogastric

production induces a spatial reorganisation of farms since they largely rely on manufactured feeds. As production intensifies, a segmentation of production stages occurs and each stage can be located to minimise operation costs (Steinfeld et al., 2006). This leads to a clustering of industrial farms close to processing plants benefiting from economies of scale (Roe et al., 2002a; Herath et al., 2005; Steinfeld et al., 2006; Larue et al., 2011). In Thailand, the process of relocation of intensive production toward a clustered distribution around the peri-urban belt of Bangkok was already observed in relation to the poultry sector between 1992 and 2000 (Steinfeld et al., 2006). Moreover, its intensive production in 2010 was much more clustered than backyard production (Chaiban et al., 2019).

The gradient of intensification the countries were spread along, as described by the GDP per capita, was observed in the number of chickens per farm, which is increasing from Nigeria to Belgium. However, we did not identify a corresponding gradient in clustering. Although Belgium has the highest and Nigeria the lowest GDP, Belgium had the least clustered distribution among all sample countries and Nigeria a higher level of clustering than Belgium and Thailand. The distribution of farms in Belgium, although clustered, and better explained by a cluster model than a random model, was closer to a homogeneous distribution of farms. The small size of the country and the quality of the road infrastructure, which allows rapid transportation of chicken products from production to consumption centres, may explain this. By contrast, the high clustering of intensive farms in Nigeria could result from a poor quality of the road network, which could lead to farms concentrating around the main cities and demand centres to reduce transport time and cost. In Argentina, farms were mainly concentrated in provinces with relatively high human population densities (Buenos Aires, Entre Rios, Santa Fe and to a lower extent Cordoba). The location of the main consumption centres may influence the location of farms, whether it is multiple large cities as in Nigeria and Argentina or the capital city as in Belgium and Thailand. Comparing the distribution of intensive farms in different countries revealed that factors other than the
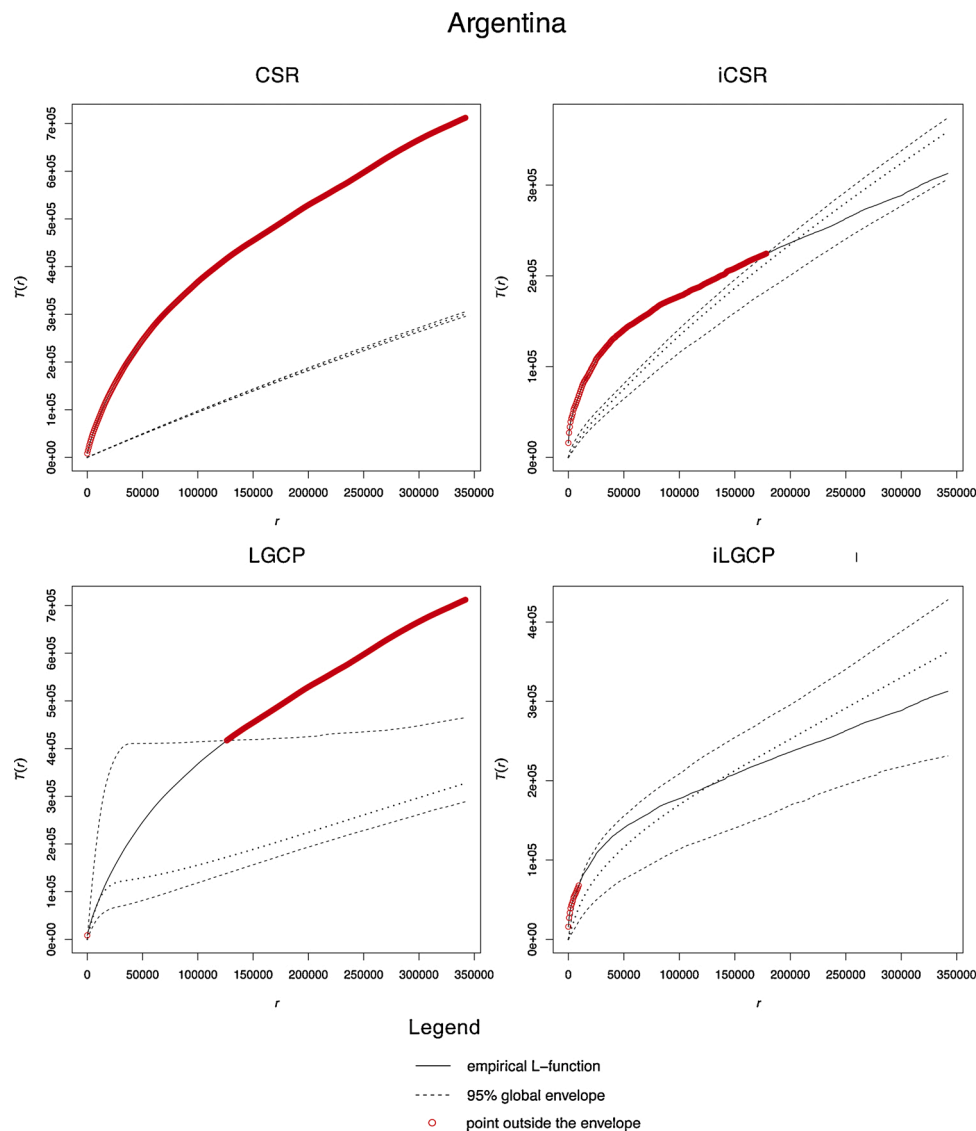
Fig. 4. Graphical interpretation of the Global Rank Envelope (GRE) test with extreme rank lengths in Argentina for the four point pattern models (CSR, iCSR, LGCP, iLGCP). Dashed lines represent the 95 % global envelope with 8000 simulations; the black line shows the empirical L-function estimated from the observed distribution of farms and all red points represent the points of the empirical L-function which are outside the envelope.

level of intensification could affect the level of farm clustering. Our observations suggest that while intensive chicken farms are clustered, the degree of clustering, the number of clusters and their position are affected by the geographic specificities of individual countries, such as size, transport infrastructure, population and economic activity distribution. However, the extent of the study areas may affect our results as we compared countries of various sizes. Should we consider the extent of Europe, farm clusters comparable to those in Argentina may exist, as there are known concentration of farms in Europe (e.g. around harbours in Belgium and the Netherlands).

Our modelling procedure had somewhat unconvincing results in internal and external validations. First, the predictions of our best point pattern models (iLGCP models) differed significantly from the observed spatial distribution of farms in internal validation. This contrasts with our expectations and previous results in Thailand, where iLGCP models explained the observed distribution of farms better at finer scales (Chaiban et al., 2019). The complex shape of Thailand could have lowered the result quality when considering here the entire country. More specifically, iLGCP models could not simulate farm distributions as clustered as the observed distributions at short ranges, especially in Thailand and Argentina. This may be an issue when interactions at close

range matter, such as epidemic transmission of infectious diseases between neighbouring farms. This poor performance at short distances could be owing to an incomplete set of predictors as this deviance appeared with iLGCP models but was not present in LGCP models in Argentina, or less strongly in Thailand. The use of a single model at the scale of Argentina may not capture the variability of factors explaining farm clustering across the country. Fitting different models on extents smaller than the country (i.e. by dividing Argentina into four to ten subareas) could potentially improve model performance. This suggest that our procedure may perform better at finer scales in large countries, as we observed in Thailand (Chaiban et al., 2019).

The point pattern models poorly predicted farm locations, numbers and clustering for Bangladesh. The underestimation of the farms could be owing to the high population density of Bangladesh. It peaks close to 10 persons per hectare, but it is lower than three in Nigeria and Thailand. The bell-shape curve of the human population density with farm intensity (SM Fig. S6) may prevent the model from predicting a sufficient number of farms in Bangladesh as it did not increase linearly. This issue may be addressed by including the density of rural population, a relative human population density (human population density/ average human density per country), or the size of the closest city. The
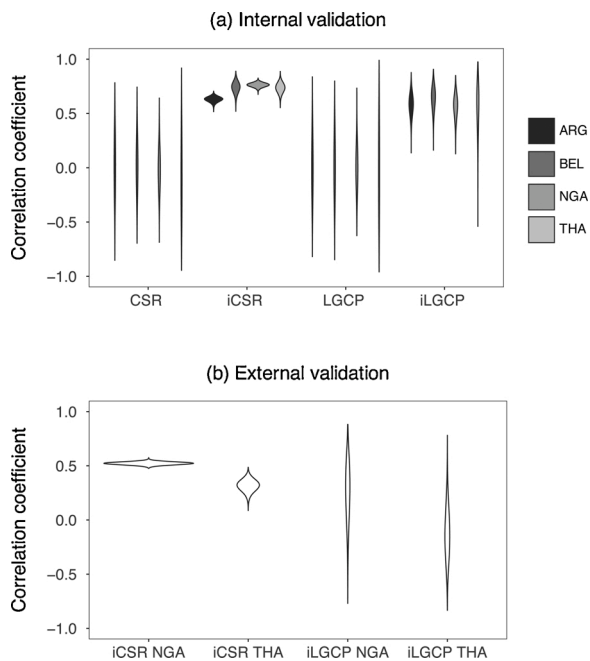
Fig. 5. Correlation coefficient between the numbers of points per quadrat in observed and each simulated pattern. (a) Internal validation. The distribution of correlation coefficient values for all 8000 simulations is plotted for the four models (completely spatial randomness (CSR), the CSR with covariates (iCSR), the Log-Gaussian Cox process (LGCP) and the LGCP with covariates (iLGCP)). (b) External validation. The distribution of correlation coefficient values for all 8000 simulations is plotted for the two models fitted on Nigeria or Thailand to simulate farms in Bangladesh, for the inhomogeneous completely spatial randomness, a CSR with covariates (iCSR) and the Log-Gaussian Cox process (LGCP) with covariates (iLGCP).

low correlation coefficient by quadrat reflects an underestimation of farms that was not constant across space. The failure to predict a realistic number of farms and to locate them correctly in Bangladesh, with models based on Nigeria or Thailand, could also be owing to an incomplete or poorly suited set of predictors, or poorly estimated coefficients. Second, when modelling farm sizes, the observed data could only be predicted with a high level of accuracy in internal validation, with the random forest model trained on the data of the four sample countries (Nigeria, Thailand, Argentina and Belgium). In cross-validation and in external validation, the correlation coefficient computed on the validation datasets showed much lower values than on the training datasets. This reflected a low capacity of the set of predictors to explain the number of chicken at farm-level.

The challenge to predict farm locations and sizes was expected as production intensification - especially in the case of monogastric species – leads to an integration of production into landless systems. Those systems are more complex to predict as farms are no longer determined by the quality of land but rather by socio-economic factors (Roe et al., 2002b; Herath et al., 2005; Larue et al., 2011; Neumann et al., 2009; Van Boeckel et al., 2012; Robinson et al., 2014). Moreover, the locations of farms may sometimes be explained by an opportunity for a business arising in an area without any factors favourable to the implementation of such a business, in terms of investment capacity or favourable policies. This was the case of broiler production in the US, which emerged as an alternative source of income following entrepreneurship and promotional efforts (Lord, 1971). The challenge of explaining farm size with environmental variables has been documented previously (Van Boeckel et al., 2012; Robinson et al., 2014). This low accuracy in predicting the number of animals per farm was also observed when predicting cattle number, even while predicting on known farm locations (van Andel et al., 2017). Predicting the exact distribution of farms

represents a challenging task when considering fine-scale location options and choices available to individual farms. We therefore attempted to provide realistic rather than exact distribution of farms with this FDM procedure. Our FDM procedure output should therefore be considered as a set of potential farm distributions in a given country.

We believe our results should be considered in the broader context of livestock distribution mapping. In internal and external validation, p-values of the global rank envelope test were not far from the rejection threshold. Despite the strong underestimation of the number of farms in Bangladesh (external validation), the p-values were in the same range of values as in the internal validation. Moreover, when predicting farm sizes, the correlation increased greatly at both aggregated levels. Similarly, the lower predictability at farm level compared with an aggregated level (3 km) was shown in a previous study (van Andel et al., 2017). This result, subject to the Modifiable areal unit problem (MAUP) effect was expected (Da Re et al., 2020), but it still highlights a higher correlation than existing modelling procedure in raster-based model, the GLW (Gilbert et al., 2018; Da Re et al., 2020). This confirmed that farm size at the individual level can be predicted without major loss in accuracy compared to aggregated distribution. It also confirmed that our procedure did not underperform the GLW, an important but expected result as the model was trained on finer data (farm locations).

The models may be improved in both their ability to reproduce farm distribution (clustering and location – point pattern model) and farm sizes (random forest model) by adding predictors. Our set appeared incomplete at several modelling steps. As farms tend to cluster with intensification, farm density could be an interesting factor to predict farm sizes. We could therefore use a farm density layer to predict farm sizes, or alternatively include the output of the farm size modelling step as a covariate within the point pattern model. Moreover, in countries where subnational livestock census data are available, including them as a covariate when modelling either location or size could refine the predictions. Intensification being driven, among others, by income growth, it would be worth considering the economic status at a subnational level with for instance the gridded GDP per capita (Kummu et al., 2018). The recently developed Subnational Human Development Database (Smits and Permanyer, 2019), could also be assessed as the economic development affects the level of intensification (Gilbert et al., 2015). Another possibility would be to include the Global Human Settlement Built-up layer (GHS-BUILT) (Corbane et al., 2018). The GHS-BUILT may better represent the demand centres than human population density, as urban areas account for the majority of the demand (Tefft et al., 2017). Finally, the accessibility layer we used, was recently refined to account for different farm sizes. Cities with a human population below 50 000 were not considered rural as in the previous versions (Nelson et al., 2019).

Further technical improvements of the FDM approach could be the integration of farm location and farm sizes steps into a common statistical framework. Farm location and size modelling are two distinct steps in our FDM procedure. The marked point pattern analysis method would be an option, but, it has, to our knowledge, not yet been developed for Log-Gaussian Cox-Processes (Baddeley et al., 2015). Another possibility would be to consider the Gaussian process stacked generalisation (Bhatt et al., 2017) to predict farm locations or sizes, accounting for non-linear relationship between the dependant variable and the covariates. Two final steps that may improve our model performance would be to first add a mask on known unsuitable areas such as national parks, water bodies or middle of urban centres, as is developed in the GLW (Gilbert et al., 2018). A second step could be to adjust the flock size so that at aggregated level it matches the totals recorded in agricultural census, since subnational data exist in many countries.

Bangladesh, chosen due to data availability, could be a suboptimal choice to externally validate our FDM approach. Bangladesh stands outside of the range of GDP covered by the four sample countries, which could explain the low farm size prediction capacity. In a further study, testing whether the low predictability in terms of farm numbers, farm
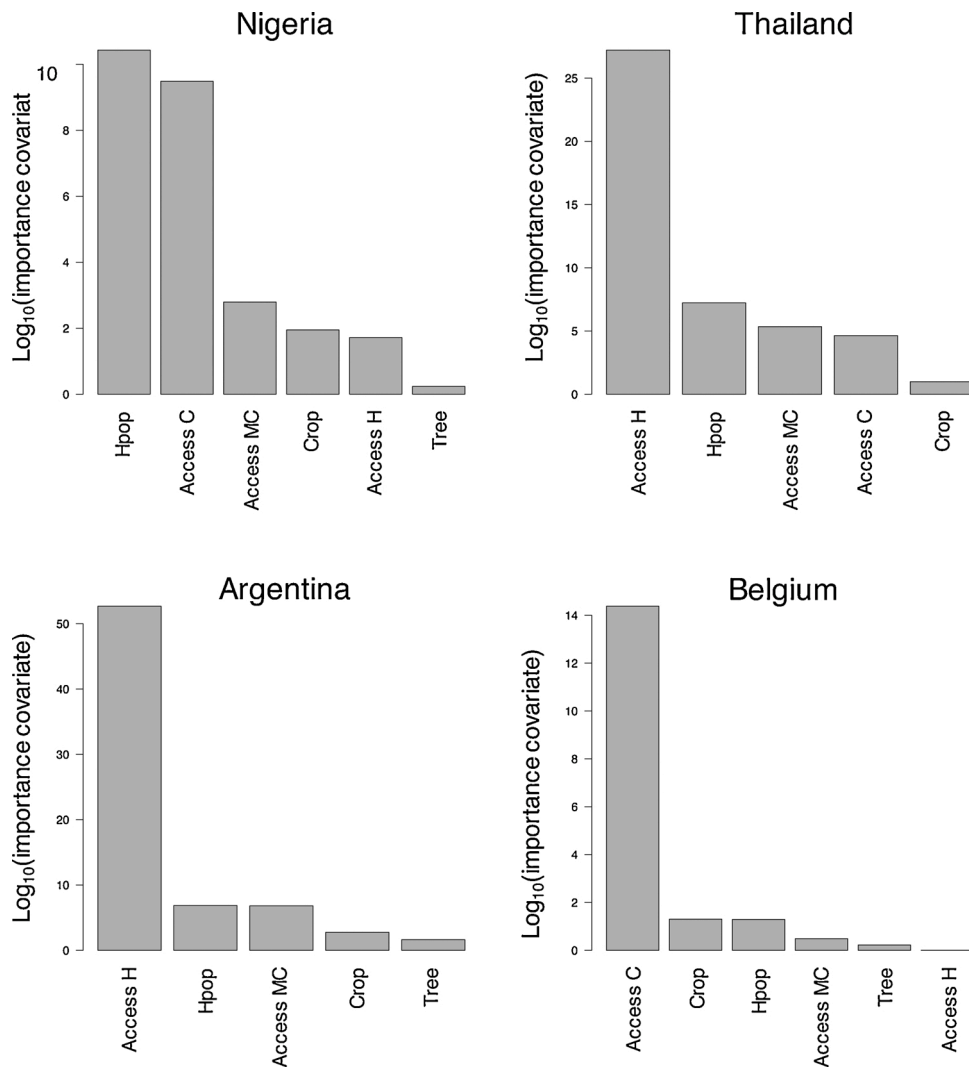
Fig. 6. Log10 of the predictors importance in the LGCP model. All covariates importance includes the sum of the importance of each covariate of its quadratic term.
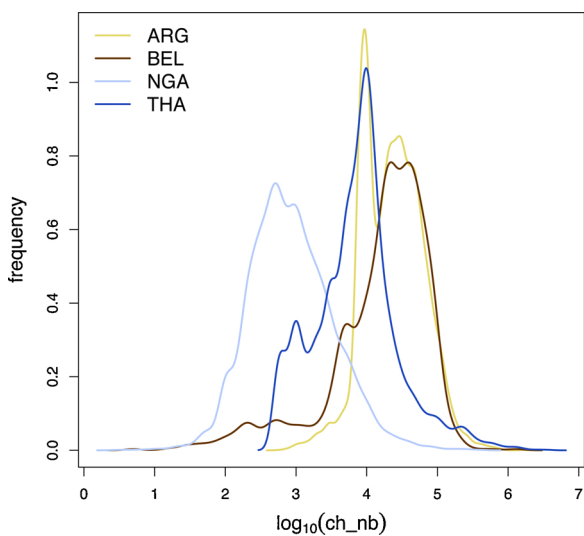

Fig. 7. Log-transformed distribution of chicken number per farm in Nigeria, Thailand, Argentina and Belgium.

locations and farm sizes comes from the set of predictors or the approach itself by assessing other swaps between training and test countries would be valuable (e.g. whether Belgian data could provide good results in predicting Argentinean farm data, Argentina to Thailand, Thailand to Nigeria, Bangladesh to Thailand).

Nigeria and Thailand have socio-economic and environmental conditions different from those of Bangladesh. We assumed in our procedure that a country whose socio-economic conditions are more comparable with those of Bangladesh should provide better results. However, Thailand - closer geographically only - provided better results than Nigeria, which is closer from Bangladesh in terms of GDP per capita, thereby suggesting that both socio-economic and environmental conditions are important to take into account. The GDP per capita was a good proxy of the intensification level (Gilbert et al., 2015), but less so for the level of clustering. A farm distribution model would be better trained on a country with similar socio-economic and environmental conditions, and of a comparable extent and population density. Nevertheless, our tests on Bangladesh already highlighted some limits of generalisability of our approach. The path of structural transformation and economic development taking place in Sub Saharan Africa may differ from what occurred in Europe and North America, then in Asia and Latin America (Loison, 2015). These differences in path may pose a limit in the extrapolation of our model using data from Asia or South America to predict farm distribution in Africa.

5. Conclusion

We attempted to develop a FDM approach predicting both the locations of individual farms and the flock sizes in the shape of a set of potential farm distributions in a given country. This methodology still has major issues, especially when extrapolated to a new area, which was the main purpose of the procedure. Further improvements should include further socio-economic and demographic covariates, and methodological modifications. Extrapolation should likely only be considered between countries of highly similar socio-economic and environmental conditions, but it still has to be assessed. Moreover, further assessments of this procedure may consider finer scale than the country scale, as it could potentially improve the results. Despite the poor performance of our models, we set bases for a methodology and identified challenges to the method, which can be improved and eventually used as a predictive tool thereby providing realistic distribution maps at a farm-level.

Acknowledgement

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prevetmed.2020.105206.

References

Aarestrup, F.M., 2005. Veterinary drug usage and antimicrobial resistance in bacteria of animal origin. Basic Clin. Pharmacol. Toxicol. 96, 271–281. https://doi.org/10.1111/j.1742-7843.2005.pto960401.x.

Abdalla, C.W., Lanyon, L.E., Hallberg, M.C., 1995. What we know about historical trends in firm location decisions and regional shifts: policy issues for an industrializing animal sector. Am. J. Agric. Econ. 77, 1229. https://doi.org/10.2307/1243353.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications With R. Chapman and Hall/CRC Press.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: methods, applications and data. In: Hay, Simon I., Graham, A., Rogers, D.J. (Eds.), Advances in Parasitology, Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications. Academic Press, pp. 119–156. https://doi.org/10.1016/S0065-308X(05)62004-0.

Besag, J., 1977. Discussion on Dr Ripley's paper. J. R. Stat. Soc. Ser. B 39, 193–195. https://doi.org/10.1111/j.2517-6161.1977.tb01616.x.

Bhatt, S., Cameron, E., Flaxman, S.R., Weiss, D.J., Smith, D.L., Gething, P.W., 2017. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. J. R. Soc. Interface 14, 20170520. https://doi.org/10.1098/rsif.2017.0520.

Burdett, C.L., Kraus, B.R., Garza, S.J., Miller, R.S., Bjork, K.E., 2015. Simulating the distribution of individual livestock farms and their populations in the United States: an example using domestic swine (Sus scrofa domesticus) farms. PLoS One 10, e0140338. https://doi.org/10.1371/journal.pone.0140338.

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. Agric. Syst. 173, 335–344. https://doi.org/10.1016/j.agsy.2019.03.004.

Chantziaras, I., Boyen, F., Callens, B., Dewulf, J., 2014. Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: a report

on seven countries. J. Antimicrob. Chemother. 69, 827–834. https://doi.org/10.1093/jac/dkt443.

Corbane, C., Florczyk, A., Pesaresi, Martino, Politis, P., Syrris, V., 2018. GHS Built-Up Grid, Derived From Landsat, Multitemporal (1975-1990-2000-2014), R2018A. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/jrc-ghsl-10007.

Da Re, D., Gilbert, M., Chaiban, C., Bourguignon, P., Thanapongtharm, W., Robinson, T.P., Vanwambeke, S.O., 2020. Downscaling livestock census data using multivariate predictive models: sensitivity to modifiable areal unit problem. PLoS One 15, e0221070. https://doi.org/10.1371/journal.pone.0221070.

FAO, 2018. World Livestock: Transforming the Livestock Sector Through the Sustainable Development Goals, p. 222. Licence: CC B Y-NC-SA 3.0 IGO. ed. Rome, Italy.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, R., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Change Biol 21, 1980–1992. https://doi.org/10.1111/gcb.12838.

Gaughan, A.E., Stevens, F.R., Linard, C., Jia, P., Tatem, A.J., 2013. High resolution population distribution maps for Southeast Asia in 2010 and 2015. PLoS One 8, e55882. https://doi.org/10.1371/journal.pone.0055882.

Gerber, P., Chilonda, P., Franceschini, G., Menzi, H., 2005. Geographical determinants and environmental implications of livestock production intensification in Asia. In: Bioresour. Technol., The 10th International Conference on Recycling of Agricultural, Municipal and Industrial Residues in Agriculture, 96, pp. 263–276. https://doi.org/10.1016/j.biortech.2004.05.016.

Gilbert, M., Conchedda, G., Van Boeckel, T.P., Cinardi, G., Linard, C., Nicolas, G., Thanapongtharm, W., D'Aietti, L., Wint, W., Newman, S.H., Robinson, T.P., 2015. Income disparities and the global distribution of intensively farmed chicken and pigs. PLoS One 10, e0133381. https://doi.org/10.1371/journal.pone.0133381.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T.P., Vanwambeke, S.O., Wint, G.R.W., Robinson, T.P., 2018. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Sci. Data 5, 180227. https://doi.org/10.1038/sdata.2018.227.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-Century forest cover change. Science 342, 850–853. https://doi.org/10.1126/science.1244693.

Herath, D.P.B., Weersink, A., Carpentier, C.L., 2005. Spatial dynamics of the livestock sector in the United States: do environmental regulations matter? J. Agric. Resour. Econ. 30.

Hill, E.M., House, T., Dhingra, M.S., Kalpravidh, W., Morzaria, S., Osmani, M.G., Brum, E., Yamage, M., Kalam, Md.A., Prosser, D.J., Takekawa, J.Y., Xiao, X., Gilbert, M., Tildesley, M.J., 2018. The impact of surveillance and control on highly pathogenic avian influenza outbreaks in poultry in Dhaka division, Bangladesh. PLOS Comput. Biol. 14, e1006439 https://doi.org/10.1371/journal.pcbi.1006439.

Jones, B.A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M.Y., McKeever, D., Mutua, F., Young, J., McDermott, J., Pfeiffer, D.U., 2013. Zoonosis emergence linked to agricultural intensification and environmental change. Proc. Natl. Acad. Sci. 110, 8399–8404. https://doi.org/10.1073/pnas.1208059110.

Kummu, M., Taka, M., Guillaume, J.H.A., 2018. Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. Sci. Data 5, 180004. https://doi.org/10.1038/sdata.2018.4.

Larue, S., Abildtrup, J., Schmitt, B., 2011. Positive and negative agglomeration externalities: arbitration in the pig sector. Spat. Econ. Anal. 6, 167–183. https://doi.org/10.1080/17421772.2011.557773.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362. https://doi.org/10.1016/j.rse.2005.10.014.

Loison, S.A., 2015. Rural livelihood diversification in Sub-Saharan Africa: A literature review. J. Dev. Stud. 51, 1125–1138. https://doi.org/10.1080/00220388.2015.1046445.

Lord, J.D., 1971. The growth and localization of the United States broiler chicken industry. Southeast. Geogr. 11, 29–42. https://doi.org/10.1353/sgo.1971.0000.

Mrkvička, T., Hahn, U., Myllymaki, M., 2016. A One-way ANOVA Test for Functional Data with Graphical Interpretation. ArXiv161203608 Stat.

Mrkvička, T., Myllymäki, M., Hahn, U., 2017. Multiple Monte Carlo testing, with applications in spatial point processes. Stat. Comput. 27, 1239–1255. https://doi.org/10.1007/s11222-016-9683-9.

Mulatu, A., Wossink, A., 2014. Environmental regulation and location of industrialized agricultural production in Europe. Land Econ. 90, 509–537. https://doi.org/10.3368/le.90.3.509.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., Hahn, U., 2017. Global envelope tests for spatial processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 79, 381–404. https://doi.org/10.1111/rssb.12172.

Naylor, R., 2005. Losing the links between livestock and land. Science 310, 1621–1622. https://doi.org/10.1126/science.1117856.

Nelson, A., Weiss, D.J., van Etten, J., Cattaneo, A., McMenomy, T.S., Koo, J., 2019. A suite of global accessibility indicators. Sci. Data 6, 266. https://doi.org/10.1038/s41597-019-0265-5.

Neumann, K., Elbersen, B.S., Verburg, P.H., Staritsky, I., Pérez-Soba, M., de Vries, W., Rienks, W.A., 2009. Modelling the spatial distribution of livestock in Europe. Landsc. Ecol. 24, 1207–1222. https://doi.org/10.1007/s10980-009-9357-5.

Nicolas, G., Robinson, T.P., Wint, G.R.W., Conchedda, G., Cinardi, G., Gilbert, M., 2016. Using random forest to improve the downscaling of global livestock census data. PLoS One 11, e0150424. https://doi.org/10.1371/journal.pone.0150424.

Perry, J., Banker, D., Green, R., 1999. Poultry Production in the United States, p. 9.

Robinson, T.P., Thornton, P., Franceschini, G., Kruska, R., Chiozza, F., Notenbaert, A., Cecchi, G., Herrero, M., Epprecht, M., Fritz, S., You, L., Conchedda, G., Lee, L., 2011. Global Livestock Production Systems. Food and Agriculture Organization of the United Nations, Rome.

Robinson, T.P., Wint, G.R.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. https://doi.org/10.1371/journal.pone.0096084.

Roe, B., Irwin, E.G., Sharp, J.S., 2002a. Pigs in space: modeling the spatial structure of hog production in traditional and nontraditional production regions. Am. J. Agric. Econ. 84, 259–278.

Roe, B., Irwin, E.G., Sharp, J.S., 2002b. Pigs in space: modeling the spatial structure of hog production in traditional and nontraditional production regions. Am. J. Agric. Econ. 84, 259–278. https://doi.org/10.1111/1467-8276.00296.

Smits, J., Permanyer, I., 2019. The subnational human development database. Sci. Data 6, 190038. https://doi.org/10.1038/sdata.2019.38.

Steinfeld, H., Gerber, P., Wassenaar, T.D., Castel, V., de Haan, C., 2006. Livestock's long shadow: environmental issues and options. Food Agric. Org.

Tefft, J., Jonasova, M., Adjao, R., Morgan, A., 2017. Food Systems for an Urbanizing World. World Bank. https://doi.org/10.1596/32502.

Tildesley, M.J., Ryan, S.J., 2012. Disease prevention versus data privacy: using landcover maps to inform spatial epidemic models. PLoS Comput. Biol. 8 https://doi.org/10.1371/journal.pcbi.1002723.

Tildesley, M.J., House, T.A., Bruhn, M.C., Curry, R.J., O'Neil, M., Allpress, J.L.E., Smith, A., Keeling, M.J., 2010. Impact of spatial clustering on disease transmission and optimal control. Proc. Natl. Acad. Sci. 107, 1041–1046. https://doi.org/10.1073/pnas.0909047107.

van Andel, M., Jewell, C., McKenzie, J., Hollings, T., Robinson, A., Burgman, M., Bingham, P., Carpenter, T., 2017. Predicting farm-level animal populations using environmental and socioeconomic variables. Prev. Vet. Med. 145, 121–132. https://doi.org/10.1016/j.prevetmed.2017.07.005.

Van Boeckel, T.P., Thanapongtharm, W., Robinson, T., D'Aietti, L., Gilbert, M., 2012. Predicting the distribution of intensive poultry farming in Thailand. Agric. Ecosyst. Environ. 149, 144–153. https://doi.org/10.1016/j.agee.2011.12.019.

Van Boeckel, T.P., Brower, C., Gilbert, M., Grenfell, B.T., Levin, S.A., Robinson, T.P., Teillant, A., Laxminarayan, R., 2015. Global trends in antimicrobial use in food animals. Proc. Natl. Acad. Sci. 112, 5649–5654. https://doi.org/10.1073/pnas.1503141112.

Vieira, A.R., Collignon, P., Aarestrup, F.M., McEwen, S.A., Hendriksen, R.S., Hald, T., Wegener, H.C., 2011. Association between antimicrobial resistance in Escherichia coli isolates from food animals and blood stream isolates from humans in Europe: an ecological study. Foodborne Pathog. Dis. 8, 1295–1301. https://doi.org/10.1089/fpd.2011.0950.

Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C.D., Howes, R.E., Tusting, L.S., Kang, S.Y., Cameron, E., Bisanzio, D., Battle, K.E., Bhatt, S., Gething, P.W., 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 553, 333–336. https://doi.org/10.1038/nature25181.

Wint, W., Robinson, T., FAO, 2007. Gridded Livestock of the World. Food and Agriculture Organization of the United Nations, Rome.

WorldPop, 2014. Thailand 100m Population. Alpha Version 2010, 2015 and 2020 Estimates of Numbers of People Per Pixel (ppp) and People Per Hectare (pph), With National Totals Adjusted to Match UN Population Division Estimates (http://esa.un.org/wpp/) and Remaining Unadjusted. School of Geography and Environmental Science, University of Southampton. https://doi.org/10.5258/SOTON/WP00267.

WorldPop, 2015. Nigeria 100m Population. Alpha Version 2010, 2015 and 2020 Estimates of Numbers of People Per Pixel (ppp) With National Totals Adjusted to Match UN Population Division Estimates (http://esa.un.org/wpp/). School of Geography and Environmental Science, University of Southampton. https://doi.org/10.5258/SOTON/WP00648.

WorldPop, 2016. Argentina 100m Population. Alpha Version 2010, 2015 and 2020 Estimates of Numbers of People Per Pixel (ppp) and People Per Hectare (pph), With National Totals Adjusted to Match UN Population Division Estimates (http://esa.un.org/wpp/) and Remaining Unadjusted. School of Geography and Environmental Science, University of Southampton.

WorldPop, 2017. Bangladesh 100m Population, Version 2. University of Southampton. School of Geography and Environmental Science, University of Southampton. https://doi.org/10.5258/SOTON/WP00533.