1   **Point pattern simulation modelling of extensive and intensive chicken farming**

2   **in Thailand: accounting for clustering and landscape characteristics.**

3

4   Celia Chaiban[1,2], Weerapong Thanapongtharm[3], Michael Tildesley[4], Xiangming Xiao[5],

5   Timothy P Robinson[6], Sophie O Vanwambeke[1,*], Marius Gilbert[2,7,*].

6

7   [1] Georges Lemaître Centre for Earth and Climate research, Earth and Life Institute, Université

8   catholique de Louvain, Louvain-la-Neuve, Belgium.

9   [2] Spatial Epidemiology Lab. (SpELL), Université Libre de Bruxelles, Brussels, Belgium.

10   [3] Department of Livestock Development (DLD), Bangkok 10400, Thailand.

11   [4] School of Life Sciences and Mathematics Institute, University of Warwick, Warwick, UK

12   [5] Department of Microbiology and Plant Biology, Center for Spatial Analysis, University of Oklahoma,

13   Norman, Oklahoma, USA.

14   [6] Livestock Information, Sector Analysis and Policy Branch (AGAL), Food and Agriculture Organization

15   of the United Nations (FAO), Viale delle Terme di Caracalla, 00153 Rome, Italy.

16   [7] Fonds National de la Recherche Scientifique (FNRS), Brussels, Belgium.

17

18   * corresponding authors SVW (sophie.vanwambeke@uclouvain.be) and MG (mgilbert@ulb.ac.be)

19

20

21 **Abstract**

22 In recent decades, intensification of animal production has been occurring rapidly in

23 transition economies to meet the growing demands of increasingly urban populations. This

24 comes with significant environmental, health and social impacts. To assess these impacts,

25 detailed maps of livestock distributions have been developed by downscaling census data at

26 the pixel level (10km or 1km), providing estimates of the density of animals in each pixel.

27 However, these data remain at fairly coarse scale and many epidemiological or

28 environmental science applications would make better use of data where the distribution and

29 size of farms are predicted rather than the number of animals per pixel. Based on detailed

30 2010 census data, we investigated the spatial point pattern distribution of extensive and

31 intensive chicken farms in Thailand. We parameterized point pattern simulation models for

32 extensive and intensive chicken farms and evaluated these models in different parts of

33 Thailand for their capacity to reproduce the correct level of spatial clustering and the most

34 likely locations of the farm clusters. We found that both the level of clustering and location of

35 clusters could be simulated with reasonable accuracy by our farm distribution models.

36 Furthermore, intensive chicken farms tended to be much more clustered than extensive

37 farms, and their locations less easily predicted using simple spatial factors such as human

38 populations. These point-pattern simulation models could be used to downscale coarse

39 administrative level livestock census data into farm locations. This could be of particular

40 value in countries where farm location data are unavailable.

41

45

1. Introduction

Following demographic and economic development, the per capita consumption of animal-source food has increased continuously over the past few decades, with significant consequences for livestock production (Delgado, 1999; Food and Agriculture Organization of the United Nations, 2013; Steinfeld, 2004). The growth in demand for animal products, mainly meat, eggs and milk, was met primarily through intensification of livestock production, which was particularly marked for monogastric species such as poultry and pigs (Gilbert et al., 2015; Smil, 2002). Today, production from monogastric species continues to grow, while production from ruminants is relatively stable (FAO, 2014).

The level of intensification of livestock production varies considerably across countries. Intensive systems represent the large majority of pig and poultry production in high-income countries (Gilbert et al., 2015), where extensive systems remains marginal in terms of output volumes, despite growing interest in small-scale and organic production (Willer, 2011). In low-income countries, pig and poultry production is mostly extensive, and in transition economies, both extensive backyard production and intensive farming systems coexist with a gradient of intensification that can be correlated to per capita gross domestic product (Gilbert et al., 2015). Demand for animal products in high-income countries has mostly levelled off and in some countries is decreasing, while it is still increasing in developing countries (Alexandratos and Bruinsma, 2012). Therefore, changes in livestock production system conferring productivity increases are much more pronounced in developing and emerging countries.

Intensification of pig and poultry production comes with significant environmental, health and societal impacts (Leibler et al., 2009; Mennerat et al., 2010; Steinfeld et al., 2010; Pulliam et al., 2012; Jones et al., 2013a; P.J. Gerber et al., 2013; Slingenbergh et al., 2013; Van Boeckel et al., 2014). Health impacts, notably through pathogen emergence and re-emergence, has a potential global relevance, as illustrated by the threat of pandemic

74   influenza (Leibler et al., 2009; Li et al., 2004; Monne et al., 2014). Intensified systems

75   promote high densities of genetically similar individuals, which promotes pathogen

76   amplification, selection of more virulent pathogens and risk of pathogen spill-over (Jones et

77   al., 2013b). Owing to their close interactions with humans, particularly in peri-urban

78   environments, and possible contacts with wild animals, intensive production systems can

79   also serve as an intermediate between wildlife and human populations and as amplifier

80   (Childs et al., 2007). Transmission of pathogens through livestock to humans and virulence

81   shifts of pathogens are thus very real risks.

82

83   Detailed spatial distribution data on livestock numbers and farms are an important element in

84   understanding livestock-environment interactions, the spread of epidemics and in assessing

85   and preventing zoonotic transmissions (Burdett et al., 2015; Martin et al., 2015; Steinfeld et

86   al., 2006). In most high-income countries, like Europe or the USA, detailed farm registers

87   exist. However, public access to these detailed data sets is not straightforward, in order to

88   preserve the confidentiality and privacy of farmers' data. Access to anonymous data can be

89   granted for research following specific applications. In some instances, privacy is protected

90   by providing only aggregated data. This is the case in the USA where data on livestock are

91   only made available at county level. In low and middle-income countries, registers rarely

92   exist and the most accurate data sets are produced through agricultural censuses. However,

93   when such censuses are performed, the level of detail of the data released by the authorities

94   varies considerably from one country to another (Robinson et al., 2014; Wint et al., 2007).

95   Both situations, from data-rich or -poor countries, may lead to livestock statistics being only

96   available at coarse spatial scales, such as a province or a district.

97

98   In addition to a lack of spatial detail in available data, a distinction between intensive and

99   extensive production systems is rarely made, though this is an important distinction in terms

100 of their health and environmental impacts (Van Boeckel et al., 2012; P. J. Gerber et al., 2013;

101 Jones et al., 2013b; Gilbert et al., 2015). Differentiating between extensive and intensive

102 systems, or simply knowing where the largest farms are, is particularly important in regions

103 where production is currently undergoing intensification, as the relative distributions of

104 extensive and intensive production may have different spatial patterns and are likely to

105 change rapidly over time.

106

107 In order to increase the spatial detail of coarse livestock data, previous studies on livestock

108 distribution mapping developed spatial statistical algorithms linking densities to pixel-level

109 environmental variables to downscale census data from administrative census boundaries to

110 finer resolution estimates at the pixel level, where each pixel contains the estimated density

111 for a particular species. This leads to a representation of livestock densities as continuous

112 surface, pixel-level variables changing in space, and this is typically the output of databases

113 such as the Gridded Livestock of the World (GLW) version 1 (Wint et al., 2007) and version 2

114 (Robinson et al., 2014). Other authors have also applied similar approaches to map different

115 livestock species at country or continental scale (Neumann et al., 2009; Prosser et al., 2011;

116 Van Boeckel et al., 2011). Thus far, few attempts have been made to distinguish extensive

117 from intensive production systems. Gilbert et al. (2015) developed an approach to separate

118 extensive from intensively raised animals in global chicken and pig maps. At the country

119 scale, Van Boeckel *et al*. (2012) observed a distinct bimodal distribution in poultry farms in

120 Thailand that could be used to distinguish extensive from intensive farms. Based on this

121 distinction they modelled extensive and intensive poultry separately using a methodology

122 similar to that of GLW, and noted a relatively poor predictive accuracy for intensively-raised

123 chickens compared to extensive chickens using that approach.

124 A continuous surface, pixel-based model may not be the best way to represents intensive

125 farms. Indeed, intensification of poultry production is such that a very large number of birds

126 can be present in a single location (e.g. typically more than 100 000 birds can be found in a

127 farm or site), with perhaps none, or very few in an adjacent pixel. A discrete spatial

128 representation of individual farms as single point locations, with the number of birds as an

129 attribute, may thus be a more appropriate representation of intensive farms than a

130 continuous surface image. Another issue with regards to modelling farm locations instead of

131 animal densities is that such models would better fit the needs of mathematical models of

132 livestock diseases (Martin et al., 2015). Epidemic mathematical transmission models may be

133 sensitive to the spatial clustering, distribution, type and overall density of farms (Reeves,

134 2012; Tildesley and Ryan, 2012), and mitigation measures of disease transmission are in

135 part based on the distance between farms. Fine-scale maps of farm distribution, including

136 farm position and level of clustering, could thus make an important contribution to models

137 that can inform control strategies (Bruhn et al., 2012). While broad-scale clusters of farms

138 may be captured by aggregated data, the factors influencing farm distribution are poorly

139 known at finer scales (Burdett et al., 2015). In the presence of aggregated census data, the

140 distribution of individual farm locations have tended to be based on random allocation of

141 points, regardless of other geographic information (Tildesley et al., 2010) or, in some cases,

142 constrained by geographical information contained in probability surfaces (Bruhn et al., 2012;

143 Burdett et al., 2015; Emelyanova et al., 2009; Tildesley and Ryan, 2012). However, none of

144 these methods have captured both first and second order characteristics to predict the

145 spatial clustering of farms as well as differences in their broader distributions.

146

147 In this paper, we investigated the use of point-pattern models as a way to predict the

148 distribution of individual farms both in terms of dependency on external variables influencing

149 their presence and in terms of spatial clustering. This approach may provide more realistic

150 representations of animal distribution at fine spatial scales than continuous pixel-based

151 distributions, especially for species such as poultry and pigs that may be raised in high

152 numbers in single premises. Our analyses focused on Thailand chicken farms, as an

153 example of a middle-income country where extensive production systems (backyard poultry

154 farms) coexist with intensive ones (large-scale chicken farms) (Van Boeckel et al., 2012).

155

156 **2. Methods**

157 *2.1. Data*

158   A detailed census was conducted in 2010 at the household level by the Department of

159   Livestock Development (DLD), Bangkok, Thailand. The census included the number of

160   chickens per owner for all farms in Thailand. The coordinates of the village were

161   subsequently linked to each farm. 1,936,590 chicken owners, or farms, were recorded in a

162   total of 62,091 villages. Henceforth, we will use the term 'farm' to represent both

163   smallholders, who may be a single family with a few chickens, and large-scale farms having

164   several thousand birds. Farms with no chickens were removed from the dataset. The precise

165   locations of individual farms were assigned randomly within a set of Voronoi polygons

166   (Okabe et al., 2000) built from the village coordinates (the median area of the Voronoi

167   polygons was 4 km^2, the mean area was 8 km^2). A mask excluding permanent water bodies

168   and the province and city of Bangkok (1,569 km^2, which is by far the largest city of Thailand,

169   populated by more than 20 times the second largest city of Thailand) was used prior to

170   distributing the farms within the Voronoi polygons. So, our input data set did not include the

171   exact locations of individual farms, but an approximate location within a 2 x 2 km area for

172   50% of the 62,091 villages. However, given the extent (all of Thailand) and resolution (1km)

173   of our analyses, this loss of accuracy was considered to be of a negligible impact on our

174   potential results.

175

176   The distribution of chickens per farm showed a clear bimodal pattern (Van Boeckel et al.,

177   2012) and a threshold of 500 chickens per farm was used in order to separate extensive

178   small-scale producers from intensive large-scale systems. This threshold was used as it

179   maximized the correlation coefficient between the quantiles of the respective intensive and

180   extensive distribution of animals per farm in the two groups and the quantiles of two normal

181   distributions of same mean and standard deviation. This resulted in two datasets of

182   1,930,003 extensive farms with a median number of 20 chickens per farm, and 6,587

183   intensive farms with a median number of 8,000 chickens per farm. Although extensive and

184   intensive farming should in principle be based on the quantification of inputs and outputs,

185   farm size was considered a good proxy in the context of Thailand poultry production systems

186   as large farms necessarily use high amounts of inputs such as specialized feeds and breeds

187   as well as significant infrastructure (such as buildings for protective housing).

188

189   Spatial predictor variables were selected to be both generic and available in databases with

190   a global extent (**Erreur ! Nous n'avons pas trouvé la source du renvoi.**, Supplementary

191   material) so that the models and approaches followed in this study could be transferred to

192   data-poor countries. The predictor variables were chosen amongst those previously identified

193   as having strong predictive capacity by Van Boeckel (2012) and included: (i) the logarithm

194   (base 10) of human population density from the Worldpop database

195   (http://www.worldpop.org.uk), (ii) the "remoteness" or travel time to Bangkok and to the

196   closest provincial capital, which were recomputed from (Nelson, 2008) friction surfaces to

197   include provincial capitals, (iii) the tree cover or percentage of land covered by forest (Ellison

198   and Bachtrog, 2013) and (iv) the cropland or percentage of land covered by crops (Fritz et

199   al., 2015). Human population density was assumed to be an important predictor variable of

200   farm location, since farms are unlikely to be located either in city centres or in completely

201   remote areas. Remoteness accounted for differences in accessibility to provincial or national

202   markets through the road and railway networks. Since remoteness also captures friction to

203   movement due to water bodies, high elevation or slope, it also helps in identifying areas

204   where chicken farms would be unlikely to be placed. Areas covered by dense and permanent

205   forest may also be exclusive to poultry farming, which is why forest cover was included as a

206   spatial variable. Finally, the distribution of croplands may be correlated with areas favourable

207   to chicken farming because they provide access to grain for feed.

208

209   **Table 1. Predictor variables tested in our models**

	Resolution (m)	Units
Human population density	1000	People per km^2
Remoteness	1000	Minute
Cropland	1000	Pixel % covered by crops
Tree cover	30	Pixel % covered by forest

210

211 *2.2. Analysis*

212 The distribution of extensive and intensive farm locations was investigated using point

213 pattern analysis and modelling.

214

215 First, we used the Besag's L-function, a transformation of Ripley's K-function, to describe

216 different point patterns. The K-function is a summary statistic of a point process, defined as

217 the expected number of *r*-neighbours of a point of **X** divided by the intensity λ:

218 $$K(r) = \frac{1}{\lambda} \mathbb{E}[\text{number of neighbours of } u \mid \mathbf{X} \text{ has a point at location } u]$$

219 for any $r \geq 0$ at any location *u*, where *r* is the radius, λ is the homogeneous intensity of

220 points, **X** is the point process and *u* is any location. This assumes that the intensity is

221 constant and does not depend on the location, so the process is considered stationary

222 (Baddeley et al., 2015). The empirical K-function is a summary of the pairwise distances of a

223 point pattern, which allows point patterns with different intensities to be compared, and the

224 analysis of a pattern at different scales, since the function is normalized by the intensity. The

225 empirical K-function is defined as

226 $$\widehat{K}(r) = \left(\frac{a}{n(n-1)}\right) \sum_{i,j=1} I\left(d[i,j] \leq r\right) e[i,j]$$

227 where *a* is the study area, *n* is the total number of points in *a*, the sum is taken over all

228 ordered pairs of points *i* and *j*, $d[i,j]$ is the distance between two points and $I(d[i,j] \leq r)$ is

229 the indicator that equals 1 if the distance is less than or equal to *r*. The term $e[i,j]$ is the

230 edge correction weight, which is a border method or "reduced sample" estimator (Ripley,

231 1988). By using $\frac{a}{n(n-1)}$, the K-function assumes that the process is a stationary process.

232 Comparing the empirical and theoretical K-functions of a point pattern enables us to

233 determine if a pattern is clustered, random or regular, with K-functions higher than, close to

234 or lower than the random case, respectively. Besag's L-function $L(r) = \sqrt{\frac{K(r)}{\pi}}$ is a

235 transformation of the K-function for which a random point pattern is a straight line

236 $L_{random}(r) = r$ when $L(r)$ is plotted against r.

237

238   *2.3.   Point pattern simulation*

239   In order to predict the spatial distribution of intensive and extensive farms as points, the Log

240   Gaussian Cox Processes (LGCP) model was used (Møller et al., 1998), with the Palm

241   maximum likelihood method of optimizing the parameters (Tanaka et al., 2008). Preliminary

242   investigation involved visual comparison of the pattern produced by the five processes

243   designed to model clustered point patterns. These five processes are the Matérn cluster

244   process, the Thomas process, the Cauchy cluster process, the Variance gamma cluster

245   process and the LGCP (Fig. 1) (Baddeley et al., 2015). In addition, we quantified how these

246   different models were able to reproduce the clustering of the observed point pattern by

247   estimating the root mean squared error (RMSE) between the L-function of the observed point

248   pattern and the L-function of 100 simulations of each model (Fig. 2), with the RMSE defined

249   as:

250
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{u}}$$

251   where y_i is the value of le L-function of the observed pattern, $\hat{y}_i$ is the value of the L-function

252   of the ith simulated point pattern and u is the total number of *r* values of the L-function being

253   compared. The conclusion of this preliminary analysis was that the LGCP performed best so

254   it was used for all subsequent modelling.

255

256   In order to address potential processing problems linked to the very high number of points of

257   all chicken farms in Thailand, we parameterised a series of models based on subsets of

258   squares of equal area (see Supplementary material). For the intensive farms dataset,

259   Thailand was divided into squares of 200 x 200 km, and we retained only the 12 squares with

260   over 250 farms in the analyses. For the more numerous extensive farms, 43 squares of 112

261   x 112 km (which included more than 50 percent of Thailand) were used.

262

263   Three different types of model were built and compared: (i) a random (null) model, which

264   randomly distributed farms within the squares, (ii) a LGCP model with a homogeneous

265   intensity (without any covariates), which distributed clusters of farms randomly within the

266   squares and (iii) a LGCP model with covariates predicting an inhomogeneous intensity and

267   identifying highly probable locations for clusters. For the later model, the Akaike Information

268   Criterion (AIC) was used to select the best combination of predictor variables,

269
$$AIC = 2\log(PL) + k(edf)$$

270   where *PL* is the maximised Palm likelihood of the fitted model, and *edf* is the effective

271   degrees of freedom of the model (Tanaka et al., 2008). The AIC values of the models with

272   different combination of covariates were compared on the 12 areas for the intensive farms

273   dataset using the standardized difference with null model AIC,

274
$$\frac{AIC_{null} - AIC_{model_i}}{AIC_{null}}$$

275   where AIC_{null} is the AIC of a LGCP model without covariates and AIC_{model_i} is the AIC of ith

276   *LGCP* models with a certain combination of variables. The model showing the greatest

277   (positive) difference with the AIC_{null} model was selected.

278

279   We aimed to evaluate the goodness-of-fit of our simulated patterns in their capacity to

280   reproduce both the level of clustering and the location of clusters in comparison to the

281   observed patterns. For each square and type of model, and using the best-fit parameters, we

282   simulated the distribution of 100 point patterns. For each point pattern, in order to quantify

283   the similarities in the level of clustering, the RMSE between the simulated and observed

284   point-pattern L-function was estimated, as explained above. This function allows a point

285   pattern to be characterised independently from the density of points, which enabled us to

286   compare RMSEs across simulations and areas. In order to evaluate the goodness-of-fit of

287   the 100 simulated patterns in terms of location of the clusters, each square study area was

288   further divided into 64 square quadrats. The correlation coefficient between the observed and

289   modelled number of farms per quadrat for each simulation was computed. Quadrats

290   intersecting the Thai border were removed when less than 95% of their area was in Thailand.

291   Quadrat size was chosen to have a sufficient number of quadrats and of points per quadrat

292   to produce a meaningful correlation coefficient. In addition to goodness-of-fit metrics

293   estimated for each model type (random, LGCP and LGCP with covariates) on the calibration

294   area, we also estimated goodness-of-fit metrics (RMSE and correlation coefficient) on a

295   different square from the model calibration area, henceforth referred to as the validation

296   area. The relative importance of each predictor variable was estimated as the exponential of

297   the coefficient value of a covariate multiplied by the range of values of the covariate

298   (Baddeley et al., 2015).

299

300   **3. Results**

301   The distributions of both extensive and intensive farms were found to be clustered, but the

302   intensive farms appeared more clustered than the extensive ones, as assessed by the L-

303   function (Fig. 3). All four spatial predictors and their quadratic terms were included in the

304   LGCP model with covariates following the comparison of AIC on the intensive farms dataset

305   (Fig. 4).

306

307   Fig. 5 shows examples of simulations produced by the three types of models (random, LGCP

308   and LGCP with covariates) applied to intensive and extensive farms datasets and a plot of

309   the observed farm pattern. In both cases, the LGCP model with covariates produced the

310   pattern the most similar to the observed pattern. The model captured the clustering of the

311   observed farm locations and located them better than the two other models.

312

313   Goodness-of-fit indices showed that the LGCP models with covariates gave better results

314   than the two other models (Fig. 6). In terms of indicators of level of clustering (Fig. 6a and

315   6b), LGCP with and without covariates reproduced the observed level of clustering better

316   than the random model, but did not differ from each other. When the models were evaluated

317   into the validation area, they showed a higher variability in their goodness-of-fit metrics

318 compared to the calibration area, but their average indicators remained better than the

319 random model. For the extensive farms, the RMSE of models in the validation area were only

320 slightly lower than the random model and their variability was larger. RMSE values were

321 always higher for the intensive than for the extensive datasets, but the difference between

322 RMSE values of the random model and clustered models was higher for the intensive than

323 the extensive dataset. The median of the RMSE values of the different models (random,

324 LGCP and LGCP with covariates (calibration and validation)) were 55.3, 27.5, 21.9 and 24.5

325 for the intensive dataset while there were 29.5, 7.9, 10.0 and 27.7 for the extensive dataset.

326 In terms of location of clusters (Fig. 6c and 6d), the LGCP model with covariates model

327 performed better than the two other models. The two sets of metrics of the LGCP models

328 with covariates in the calibration and validation areas had significantly higher correlation

329 coefficients than the other models (random model and LGCP model without covariates). This

330 difference was apparent for both intensive and extensive farm point patterns. The medians of

331 the correlation coefficients of LGCP models with covariates were generally higher for the

332 extensive than for the intensive dataset. The medians of the correlation coefficients of the

333 different models (random, LGCP and LGCP with covariates (calibration and validation)) were

334 0.01, 0.00, 0.41 and 0.31 for the intensive dataset and 0.01, 0.01, 0.67 and 0.60 for the

335 extensive dataset.

336 Table 2 presents the averaged coefficients of the different model parameters for both

337 intensive and extensive datasets. Human population density was by far the most important

338 predictor of intensive and extensive models on average, followed by tree cover, cropland and

339 remoteness (Fig. 7), and the relative importance of predictor variables were similar for the

340 intensive and extensive farms.

341

342 **Table 2. Averaged coefficients of the different model parameters**

	$\sigma2$	α	Intercept	Hpop	Crop	Tree	Remot	$Hpop^2$	$Crop^2$	$Tree^2$	$Remot^2$
Extensif	0,749	3,17E-03	-16,8	3,31	3,18E-3	-0,012	-9,07E-05	-0,519	-6,83E-06	1,50E-04	-6,84E-07
Intensif	2,707	7,82E-03	-24,8	6,00	7,57E-03	0,013	-190E-05	-1,16	-9,84E-05	-12,8E-04	-4,95E-06

343

4. Discussion

Our LGCP models with covariates were able to produce simulated point patterns reproducing both the level of clustering of farms and the local density of farms better than could the random models. The observed distribution of extensive farms was closer to a random model than that of the intensive farms. The benefit of the LGCP extensive models over the random models was therefore more limited, but this was partially balanced by a better predictability of the local density as quantified by the correlation coefficients. Conversely, intensive farms were more clustered, so the LGCP models reproduced these point patterns much better than the random model, but the quality of the prediction of local densities was lower.

The higher clustering of intensive farms and the higher difficulty in predicting their location was already noted by Van Boeckel *et al.* (2012), and may be linked to two causes. First, the initial establishment of an intensive farm is probably influenced both by fine-scale spatial factors (land availability, location suitability and access to inputs and markets, for example) and to individual farmer characteristics (where they live, and the locations of their other investments, for example). It would be difficult to obtain such information from a census-based approach. At the scale of the variables used in our models, several sites may seem equally suitable for setting up a farm, for example, by having an easy access to markets and inputs such as feed. However, many factors occurring at finer scale, such as land availability, may make areas more suitable than others. The factors making those sites suitable will result in a higher overall density of farms than in less suitable areas, where other type of economy could then be developed. Therefore, there is an inherent amount of variability in intensive farm location that our models are unable to capture. Second, the establishment of a farm at a location may induce other farms to establish in close proximity, which may lead to geographical clustering of farms. Such clustering of farms may enable farmers to benefit from economies of scale (Van Boeckel et al., 2012). Many farms in Thailand are owned by contract farmers who work for large consolidator companies such as Charoen Pokphand (CP) who provide them with inputs and take the outputs. Farms directly owned by CP may

372 also be clustered for these reasons. Imitation effects may also operate in the establishment

373 of farm clusters. As described by (Feder et al., 1985), the adoption of agricultural innovations

374 in developing countries is also affected by group influences on individual behaviour. The

375 presence of a well-established, successful, intensive poultry farm may stimulate similar

376 economic activity within a neighbourhood. The improved prediction of intensive farm

377 locations by including clustering thus makes sense. More surprising was the dominance of

378 human population density as a predictor variable compared to others since broiler production

379 in Thailand is mainly located around hatcheries, feed mills and processing plants (Costales,

380 2004; NaRanong, 2007), but these may themselves correlate to human population too. The

381 association with high human population density can be explained through market access,

382 and the model parameters for the intensive farms typically placed them in areas with

383 intermediate human population density, i.e. peri-urban areas. The establishment of a chicken

384 farm is thus constrained by a trade-off between market access (for the easy delivery of

385 outputs) and the cost of land, which may become prohibitive in more urbanized areas. Our

386 results contrasted with the results of Van Boeckel et al. (2012), where they used a cropping

387 intensity (number of cycle of crops/year), irrigated areas, human population density, travel

388 time to main citites and rural population as predictor variables. It showed cropping factor with

389 a stronger effect than human population in their logistic regression models of

390 presence/absence of intensively raised chickens. Despite, this cropping factor, i.e. number of

391 crop cycles per year, was previously found to be correlated with chicken distributions. This

392 variable was not included in our model since it was not available globally. Another difference

393 between the protocols may also explain the lower effect of some factors. Van Boeckel et al.

394 (2012) analysed the entire extent of Thailand, whereas our models were trained within much

395 smaller spatial units. Cropping patterns influencing the distribution of chicken farms at broad

396 spatial scales, but not at fine scales, and explain why the cropping variable had a lower

397 influence in our models.

398

399 The distribution of extensive farms was less clustered, and more readily predicted, with the

400 main predictor variable being human population density. This fitted our expectations because

401 extensively raised chickens are typically owned as backyard poultry by rural populations.

402 Therefore, with the exception of urban centres, where people would be less likely to raise

403 chickens, a strong correlation was expected as it was previously shown (Van Boeckel et al.,

404 2012).

405

406 The geography of chicken farming thus appears to be structured by the level of

407 intensification. During the pre-industrial period, chickens and pigs were associated with the

408 human settlements in which they were raised to provide meat and eggs. At this time, they

409 were probably homogeneously distributed within rural populations. This still occurs today in

410 low and middle-income countries, where traditional production systems are still in practice,

411 as is the case for extensive farms in Thailand. Intensification of production most likely

412 occurred as people began to move into cities. Larger farms started to be located in the peri-

413 urban belt around major consumption centres. However, as the cities expand, the land value

414 in the periphery increases, and thanks to improvement in transport infrastructures, farms can

415 be move further from peri-urban belts, with the advantage of being closer to feed production

416 areas. This was also driven by factors such as labour prices, availability, costs and quality of

417 feeds, and the risk of disease spread from livestock to cities (Steinfeld et al., 2006). As

418 production intensifies, a segmentation of production steps occurs and each stage is located

419 so that operating costs are minimised. In Thailand, this process of relocation of intensive

420 production away from the most immediate peri-urban belt of Bangkok, was already observed

421 for poultry sector between 1992 and 2000 (Thanapongtharm et al., 2016) for intensive pig

422 sector.

423

424 Other global predictors may be considered in future developments of this model. For

425 example, the SPAM database (You and Wood, 2005) contains global data on the distribution

426 of several crop commodities at a 10 km resolution that could possibly allow considering only

427   cropland relevant to animal feed. In monsoon Asia, the model may also include rice cropping

428   intensity (Xiao et al., 2006), i.e. number of rice production cycles as this would be an

429   important feed resource that was previously shown to be correlated with chicken (Van

430   Boeckel et al., 2012) and duck (Gilbert et al., 2006) distributions. Another possible

431   improvement may be gained from the inclusion of other accessibility predictors, such as

432   travel distance to ports where feed could be imported, or where outputs could be exported.

433   However, we did not view this as a major concern because to some extent, the factors

434   influencing the macro-scale distribution of chickens may be already captured in coarse level

435   censuses used as input in the models. What matters is that those data could be downscaled

436   into a spatial distribution of farms that has the same spatial point pattern characteristics as

437   the observed ones. Finally, settlement locations could provide valuable information on

438   access to service and markets.

439

440   An important next step would be to apply and validate the models developed here in a

441   country with a similar level of intensification, such as Vietnam for example, where detailed

442   census data exist. At a later stage, it would be interesting to investigate how the extensive

443   and intensive models could predict the distribution of farms according to different situations.

444   One could imagine high-income countries where 99% of the production is intensive to be

445   best predicted by the intensive model alone, and, conversely, that the extensive model could

446   be tested in low-income countries. In intermediate situations, one could apply both models

447   according to the proportion of extensively raised poultry predicted at the national level by

448   Gilbert et al. (2015). This study helped understanding the patterns and the underlying causes

449   of farm locations. Further extension of this work will lead to the development of entire farm

450   allocation models, where the total number of animals of an administrative unit could be

451   allocated to farms at locations predicted by the LGCP simulation model in such a way to

452   reproduce a given distribution of animals per farm.

453

454   The problem that this paper addressed is not limited to chicken production, and other types

455   of livestock farming may benefit from similar approaches. Pig farming, for example, is also

456   disconnected from the land and could be expected to be subject to similar spatial constraints

457   linked to feed availability and market access. In contrast, the distribution of grazing ruminant

458   farms may have very different spatial determinants. Except for feedlot cattle, dependence on

459   large areas for grazing may result in a more homogenous spatial distribution. Land-use

460   predictor variables such as rangeland or pastures may thus become more important factors.

461   Approaches may become further complicated in the case of mixed farms that produce a

462   combination of crops and a variety of livestock types.

463

464   The extensions of this type of farm distribution model, upon validation, will be relevant in

465   diverse situations. In high income countries, access to detailed farm distribution data can be

466   impossible for confidentiality reasons. The lack of individual farm information may pose

467   difficulties to the management of human and animal health risk, for example in the US

468   (Burdett et al., 2015). However, high-income countries also have more economic resource to

469   prevent and control infectious diseases (Perry et al., 2013). In contrast, low-income countries

470   have few resource for agricultural censuses, often have coarse-resolution and outdated

471   livestock statistics, and these are precisely where the impact of livestock diseases on

472   livelihoods, animal and human health are greatest (Childs et al., 2007) and where good

473   quality data may help with disease prevention. Such areas are also precisely where farming

474   is mostly extensive, and where farm distribution models show a better predictability, and may

475   prove particularly useful. In middle-income countries, the situation is more complex. While in

476   Brazil livestock data are available at fine scale, in some other large livestock producing

477   countries, such as China and India, livestock data are only available at coarse resolution.

478   Combining models of extensive and intensive production may be particularly useful in

479   middle-income counties since both production systems usually coexist in high number in

480   these countries.

481

5. Conclusions

We developed farm distribution models using a point pattern modelling technique, which allowed the simulation of chicken farm distributions both in terms of spatial clustering and location of clusters. The methods developed here no longer predict livestock distribution as a continuous variable but as a discrete variable (i.e. point locations), which is better suited for situations in which animals are raised in very large numbers in a single premises. Upon validation in other countries, this may facilitate several applications in epidemiology or environmental science in countries where such detailed data are lacking, or where livestock data are aggregated to protect privacy.

Acknowledgments

References

Alexandratos, N., Bruinsma, J., 2012. World agriculture towards 2030/2050: The 2012 revision. ESA Working paper Rome, FAO, Rome.

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

Bruhn, M., Munoz, B., Cajka, J., Smith, G., Curry, R., Wagener, D., Wheaton, W., 2012. Synthesized Population Databases: A Geospatial Database of US Poultry Farms. RTI Press, Research Triangle Park, NC.

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. doi:10.1371/journal.pone.0140338

Childs, J.E., Richt, J.A., Mackenzie, J.S., 2007. Introduction: conceptualizing and partitioning the emergence process of zoonotic viruses from wildlife to humans. Curr. Top. Microbiol. Immunol. 315, 1–31.

Costales, A., 2004. A review of the Thailand poultry sector. Livestock Sector Report - Thailand elaborated for the FAO-AGAL.

Delgado, C.L., 1999. Livestock to 2020: the next food revolution. International Food Policy Research Institute ; Food and Agriculture Organization of the United Nations ; International Livestock Research Institute, Washington, D.C.; Rome, Italy; Nairobi, Kenya.

Ellison, C.E., Bachtrog, D., 2013. Dosage Compensation via Transposable Element Mediated Rewiring of a Regulatory Network. Science 342, 846–850. doi:10.1126/science.1239552

Emelyanova, I.V., Donald, G.E., Miron, D.J., Henry, D.A., Garner, M.G., 2009. Probabilistic Modelling of Cattle Farm Distribution in Australia. Environ. Model. Assess. 14, 449–465. doi:10.1007/s10666-008-9140-z

FAO, 2014. FAOSTAT [WWW Document]. URL http://www.fao.org/faostat/en/#home (accessed 6.27.17).

Feder, G., Just, R., Zilberman, D., 1985. Adoption of Agricultural Innovations in Developing Countries: A Survey. Econ. Dev. Cult. Change 33, 255–98.

Food and Agriculture Organization of the United Nations, 2013. World livestock 2013: changing disease landscapes.

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, A., 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.

538   doi:10.1111/gcb.12838

539   Gerber, P.J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., Falcucci, A.,
540   Tempio, G., 2013. Tackling climate change through livestock – A global assessment of
541   emissions and mitigation opportunities. Food and Agriculture Organization of the United
542   Nations, Rome.

543   Gerber, P.J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., Falcucci, A.,
544   Tempio, G., 2013. Tackling climate change through livestock: a global assessment of
545   emissions and mitigation opportunities. xxi + 115 pp.

546   Gilbert, M., Chaitaweesub, P., Parakamawongsa, T., Premashthira, S., Tiensin, T.,
547   Kalpravidh, W., Wagner, H., Slingenbergh, J., 2006. Free-grazing Ducks and Highly
548   Pathogenic Avian Influenza, Thailand. Emerg. Infect. Dis. 12, 227–234.
549   doi:10.3201/eid1202.050640

550   Gilbert, M., Conchedda, G., Van Boeckel, T.P., Cinardi, G., Linard, C., Nicolas, G.,
551   Thanapongtharm, W., D'Aietti, L., Wint, W., Newman, S.H., Robinson, T.P., 2015. Income
552   Disparities and the Global Distribution of Intensively Farmed Chicken and Pigs. PLOS ONE
553   10, e0133381. doi:10.1371/journal.pone.0133381

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

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

562   Leibler, J.H., Otte, J., Roland-Holst, D., Pfeiffer, D.U., Soares Magalhaes, R., Rushton, J.,
563   Graham, J.P., Silbergeld, E.K., 2009. Industrial Food Animal Production and Global Health
564   Risks: Exploring the Ecosystems and Economics of Avian Influenza. EcoHealth 6, 58–70.
565   doi:10.1007/s10393-009-0226-0

566   Li, K.S., Guan, Y., Wang, J., Smith, G.J.D., Xu, K.M., Duan, L., Rahardjo, A.P.,
567   Puthavathana, P., Buranathai, C., Nguyen, T.D., Estoepangestie, A.T.S., Chaisingh, A.,
568   Auewarakul, P., Long, H.T., Hanh, N.T.H., Webby, R.J., Poon, L.L.M., Chen, H., Shortridge,
569   K.F., Yuen, K.Y., Webster, R.G., Peiris, J.S.M., 2004. Genesis of a highly pathogenic and
570   potentially pandemic H5N1 influenza virus in eastern Asia. Nature 430, 209–213.
571   doi:10.1038/nature02746

572   Martin, M.K., Helm, J., Patyk, K.A., 2015. An approach for de-identification of point locations
573   of livestock premises for further use in disease spread modeling. Prev. Vet. Med. 120, 131–
574   140. doi:10.1016/j.prevetmed.2015.04.010

575   Mennerat, A., Nilsen, F., Ebert, D., Skorping, A., 2010. Intensive Farming: Evolutionary
576   Implications for Parasites and Pathogens. Evol. Biol. 37, 59–67. doi:10.1007/s11692-010-
577   9089-0

578   Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox Processes.
579   Scand. J. Stat. 25, 451–482. doi:10.1111/1467-9469.00115

580   Monne, I., Fusaro, A., Nelson, M.I., Bonfanti, L., Mulatti, P., Hughes, J., Murcia, P.R., Schivo,
581   A., Valastro, V., Moreno, A., Holmes, E.C., Cattoli, G., 2014. Emergence of a Highly
582   Pathogenic Avian Influenza Virus from a Low-Pathogenic Progenitor. J. Virol. 88, 4375–
583   4388. doi:10.1128/JVI.03181-13

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

586   Nelson, A., 2008. Travel time to major cities: A global map of Accessibility. Global
587   Environment Monitoring Unit—Joint Research Centre of the European Commission, Ispra,
588   Italy.

589   Neumann, K., Elbersen, B.S., Verburg, P.H., Staritsky, I., Pérez-Soba, M., Vries, W. de,
590   Rienks, W.A., 2009. Modelling the spatial distribution of livestock in Europe. Landsc. Ecol.
591   24, 1207. doi:10.1007/s10980-009-9357-5

592   Okabe, A., Sugihara, K., Kendall, D.G., Boots, B., 2000. Spatial tessellations: concepts and

593   applications of Voronoi diagrams, 2nd ed. ed, Wiley series in probability and statistics. Wiley,
594   New York.
595   Perry, B.D., Grace, D., Sones, K., 2013. Current drivers and future directions of global
596   livestock disease dynamics. Proc. Natl. Acad. Sci. 110, 20871–20877.
597   doi:10.1073/pnas.1012953108
598   Prosser, D.J., Wu, J., Ellis, E.C., Gale, F., Van Boeckel, T.P., Wint, W., Robinson, T., Xiao,
599   X., Gilbert, M., 2011. Modelling the distribution of chickens, ducks, and geese in China.
600   Agric. Ecosyst. Environ. 141, 381–389. doi:10.1016/j.agee.2011.04.002
601   Pulliam, J.R.C., Epstein, J.H., Dushoff, J., Rahman, S.A., Bunning, M., Jamaluddin, A.A.,
602   Hyatt, A.D., Field, H.E., Dobson, A.P., Daszak, P., the Henipavirus Ecology Research Group
603   (HERG), 2012. Agricultural intensification, priming for persistence and the emergence of
604   Nipah virus: a lethal bat-borne zoonosis. J. R. Soc. Interface 9, 89–101.
605   doi:10.1098/rsif.2011.0223
606   Reeves, A., 2012. Construction and evaluation of epidemiologic simulation models for the
607   within- and among-unit spread and control of infectious diseases of livestock and poultry
608   (Thesis). Colorado State University. Libraries.
609   Ripley, B.D., 1988. Statistical Inference for Spatial Processes. Cambridge University Press.
610   Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E.,
611   Cinardi, G., D'Aietti, L., Hay, S.I., Gilbert, M., 2014. Mapping the Global Distribution of
612   Livestock. PLoS ONE 9, e96084. doi:10.1371/journal.pone.0096084
613   Smil, V., 2002. Worldwide transformation of diets, burdens of meat production and
614   opportunities for novel food proteins. Enzyme Microb. Technol. 30, 305–311.
615   Steinfeld, H., 2004. The livestock revolution—a global veterinary mission. Vet. Parasitol. 125,
616   19–41. doi:10.1016/j.vetpar.2004.05.003
617   Steinfeld, H., Gerber, P., Wassenaar, T.D., Castel, V., de Haan, C., 2006. Livestock's long
618   shadow: environmental issues and options. Food & Agriculture Org.
619   Steinfeld, H., Mooney, H.A., Schneider, F., Neville, L.E., 2010. Livestock in a Changing
620   Landscape, Volume 1: Drivers, Consequences, and Responses. Island Press.
621   Tanaka, U., Ogata, Y., Stoyan, D., 2008. Parameter Estimation and Model Selection for
622   Neyman-Scott Point Processes. Biom. J. 50, 43–57. doi:10.1002/bimj.200610339
623   Thanapongtharm, W., Linard, C., Chinson, P., Kasemsuwan, S., Visser, M., Gaughan, A.E.,
624   Epprech, M., Robinson, T.P., Gilbert, M., 2016. Spatial analysis and characteristics of pig
625   farming in Thailand. BMC Vet. Res. 12. doi:10.1186/s12917-016-0849-7
626   Tildesley, M.J., House, T.A., Bruhn, M.C., Curry, R.J., O'Neil, M., Allpress, J.L.E., Smith, G.,
627   Keeling, M.J., 2010. Impact of spatial clustering on disease transmission and optimal control.
628   Proc. Natl. Acad. Sci. 107, 1041–1046. doi:10.1073/pnas.0909047107
629   Tildesley, M.J., Ryan, S.J., 2012. Disease Prevention versus Data Privacy: Using Landcover
630   Maps to Inform Spatial Epidemic Models. PLoS Comput. Biol. 8.
631   doi:10.1371/journal.pcbi.1002723
632   Van Boeckel, T.P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B.T., Levin, S.A.,
633   Laxminarayan, R., 2014. Global antibiotic consumption 2000 to 2010: an analysis of national
634   pharmaceutical sales data. Lancet Infect. Dis. 14, 742–750.
635   Van Boeckel, T.P., Prosser, D., Franceschini, G., Biradar, C., Wint, W., Robinson, T., Gilbert,
636   M., 2011. Modelling the distribution of domestic ducks in Monsoon Asia. Agric. Ecosyst.
637   Environ. 141, 373–380. doi:10.1016/j.agee.2011.04.013
638   Van Boeckel, T.P., Thanapongtharm, W., Robinson, T., D'Aietti, L., Gilbert, M., 2012.
639   Predicting the distribution of intensive poultry farming in Thailand. Agric. Ecosyst. Environ.
640   149, 144–153. doi:10.1016/j.agee.2011.12.019
641   Willer, H., 2011. The world of organic agriculture 2011: summary. The world of organic
642   agriculture.
643   Wint, W., Robinson, T., FAO, 2007. Gridded livestock of the world. Food and Agriculture
644   Organization of the United Nations, Rome.
645   Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J.Y., Salas, W., Moore, B., 2006. Mapping
646   paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images.
647   Remote Sens. Environ. 100, 95–113. doi:10.1016/j.rse.2005.10.004

648 You, L., Wood, S., 2005. Assessing the spatial distribution of crop areas using a cross-
649 entropy method. Int. J. Appl. Earth Obs. Geoinformation, Bridging Scales and Epistemologies
650 7, 310–323. doi:10.1016/j.jag.2005.06.010
651

Figure captions

Fig. 1. Observed point pattern of a sample area from Thailand (a) compared to simulations with (b) a Matérn process model (c) a Thomas process model (d) a Cauchy process model (d) a Variance Gamma process model (e) a Log-Gaussian Cox Processes model.

Fig. 2. Comparaison of the different processes. Boxplot of the RMSE of L-function values of each simulations comparing the five processes: Matérn, Thomas, Cauchy, Variance Gamma and Log-Gaussian Cox Processes (LGCP).

Fig. 3 Descriptive analysis of intensive and extensive farms datasets using L-function. Color lines: L estimates of the observed point pattern from each square area; black line: random distribution.

Fig. 4. Comparison of models with different combination of covariates (human population density (Hpop), remoteness (Remot), cropland (Crop) and tree cover (Tree)) with AIC standardized difference. The first model is fitted with Hpop, the second model is fitted with Hpop + Remot, the third model is fitted with Hpop + Remot + Crop, the fourth is fitted with Hpop + Remot + Crop + Tree, for the four variables de square term is also added. Grey lines represent values for each square area and the black line the average line.

Fig. 5. Examples of simulations of intensive and extensive chicken farm point pattern using 3 different models and the observed point pattern. (a) Observed point pattern (b) Random simulated point pattern with a random model (c) Clustered simulated point pattern using a LGCP model (d) Clustered simulated point pattern with spatial predictors using a LGCP model with covariates.

679   **Fig. 6. Intensive and extensive systems goodness-of-fit indexes** RMSE between L-

680   function value of the observed and each simulated pattern for a) extensive and b) intensive

681   farms. The distribution of RSME values for all 100 simulations on each area is plotted with a

682   violin plot for the random model, the LGCP model, the LGCP model with covariates and the

683   Validation model (LGCP with covariates). Correlation coefficient between the numbers of

684   points per quadrat between all quadrats in observed and each simulated pattern for c)

685   extensive and d) intensive farms. The distribution of correlation coefficient values for all 100

686   simulations on each area is plotted for the four models, random, LGCP, LGCP with

687   Covariates and Validation models.

688

689   **Fig. 7. Relative covariates importance of LGCP models with covariates.** Logarithm of

690   the relative importance of each covariate and its quadratic term: human population density

691   (Hpop + Hpop^2), tree cover (Tree + Tree^2), cropland (Crop + Crop^2) and the remoteness or

692   accessibility (Remot + Remot^2).

693