Point pattern simulation modelling of extensive and intensive chicken farming in Thailand

Accounting for clustering and landscape characteristics

Chaiban, Celia; Biscio, Christophe; Thanapongtharm, Weerapong; Tildesley, Michael J. ; Xiao, Xiangming ; Robinson, Timothy P. ; Vanwambeke, Sophie O.; Gilbert, Marius

Published in:
Agricultural Systems

DOI (link to publication from Publisher):
10.1016/j.agsy.2019.03.004

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Point pattern simulation modelling of extensive and intensive chicken farming in Thailand: accounting for clustering and landscape characteristics.

Celia Chaiban[1,2], Christophe Biscio[3], Weerapong Thanapongtharm[4], Michael Tildesley[5], Xiangming Xiao[6], Timothy P Robinson[7], Sophie O Vanwambeke[1,*], Marius Gilbert[2,8,*].

[1] Georges Lemaître Centre for Earth and Climate research, Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

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

[3] Department of Mathematical Sciences, Aalborg University, Denmark

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

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

[6] Department of Microbiology and Plant Biology, Center for Spatial Analysis, University of Oklahoma, Norman, Oklahoma, USA.

[7] 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.

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

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

Abstract

In recent decades, intensification of animal production has been occurring rapidly in transition economies to meet the growing demands of increasingly urban populations. This comes with significant environmental, health and social impacts. To assess these impacts, detailed maps of livestock distributions have been developed by downscaling census data at the pixel level (10km or 1km), providing estimates of the density of animals in each pixel. However, these data remain at fairly coarse scale and many epidemiological or environmental science applications would make better use of data where the distribution and size of farms are predicted rather than the number of animals per pixel. Based on detailed 2010 census data, we investigated the spatial point pattern distribution of extensive and intensive chicken farms in Thailand. We parameterized point pattern simulation models for extensive and intensive chicken farms and evaluated these models in different parts of Thailand for their capacity to reproduce the correct level of spatial clustering and the most likely locations of the farm clusters. We found that both the level of clustering and location of clusters could be simulated with reasonable accuracy by our farm distribution models. Furthermore, intensive chicken farms tended to be much more clustered than extensive farms, and their locations less easily predicted using simple spatial factors such as human populations. These point-pattern simulation models could be used to downscale coarse administrative level livestock census data into farm locations. This methodology could be of particular value in countries where farm location data are unavailable.

Keywords

Agricultural intensification, Point pattern analysis, Farm distribution model, Livestock production systems

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; Slingenbergh et al., 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). Interest in good spatial data on livestock distribution has grown along intensification and the growing importance of livestock as a food and income source, as well as a source of environmental and sanitary issues (Burdett et al., 2015; Martin et al., 2015; Steinfeld et al., 2006). Several challenges exist in relation to the production of such maps, among which the level of intensification and the available source data stand out.

In most high-income countries, detailed farm registers exist, but are often distributed in aggregated form to protect privacy. In low and middle-income countries, registers rarely exist and the most accurate data sets are produced through agricultural censuses, the detail of which varies considerably across countries (Robinson et al., 2014; Wint et al., 2007). Both situations, from data-rich or -poor countries, may lead to livestock statistics being only available at coarse spatial scales insufficient for detailed analyses. To increase the spatial detail of coarse livestock data, previous studies on livestock distribution mapping developed spatial statistical algorithms linking densities to environmental variables to downscale census data from administrative boundaries to density estimates at the pixel level. This represents livestock densities varying gradually across pixels, as in databases such as the Gridded Livestock of the World (GLW) version 1 (Wint et al., 2007), version 2 (Robinson et al., 2014) and version 3 (Gilbert et al., In press). Other authors have applied similar approaches to map livestock at country or continental scale (Neumann et al., 2009; Prosser et al., 2011; Van Boeckel et al., 2011).

In addition to a lack of spatial detail, a distinction between intensive and extensive production systems, is rarely made. Intensive systems were defined as large-scale commercial, market-oriented and high-input farms and extensive systems as small-scale, low-input backyard production systems (Van Boeckel et al., 2012). However, this is an important distinction in terms of their health and environmental impacts (Van Boeckel et al., 2012; Gerber et al., 2013; Jones et al., 2013a; Gilbert et al., 2015). More specifically, intensification of pig and poultry production comes with significant, among others, health impacts (Leibler et al., 2009; Mennerat et al., 2010; Pulliam et al., 2012; Jones et al., 2013b; 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 influenza (Leibler et al., 2009; Li et al., 2004; Monne et al., 2014). Intensified systems promote high densities of genetically similar individuals, which promotes pathogen amplification, selection of more virulent pathogens and risk of pathogen spill-over (Jones et al., 2013a). Owing to their close interactions with humans, particularly in peri-urban environments, and possible contacts with wild animals, intensive production systems can also serve as an intermediate between wildlife and human populations and as amplifier (Childs et al., 2007). Differentiating between extensive and intensive systems, or simply knowing where the largest farms are, is therefore particularly important in regions where production is currently undergoing intensification, as the distributions of extensive and intensive farms may have different spatial patterns and may change rapidly through time. Thus far, few attempts have been made to distinguish extensive from intensive production systems. Gilbert et al. (2015) developed an approach to separate extensive from intensively raised animals in global chicken and pig maps based on a simple mode using GDP per capita. At the country scale, Van Boeckel *et al*. (2012) observed a distinct bimodal distribution in poultry farms in Thailand that could be used to distinguish extensive from intensive farms. They modelled extensive and intensive poultry separately using a methodology similar to that of GLW, and noted a relatively poor predictive accuracy for intensively-raised chickens compared to extensive chickens using that approach.

Finally, a continuous surface, pixel-based model may not be the best way to represent intensive farms. Indeed, intensification of poultry production is such that a very large number of birds can be present in a single location (e.g. typically more than 100 000 birds can be found in a farm or site), with very few in an adjacent pixel. A discrete spatial representation of individual farms as single point locations, with the number of birds as an attribute, may thus represent intensive farms better than a continuous surface image. Another issue with regards to modelling farm locations instead of animal densities is that such models would better fit the needs of mathematical models of livestock diseases (Martin et al., 2015). Epidemic mathematical transmission models may be sensitive to the spatial clustering, distribution, type and overall density of farms (Reeves, 2012; Tildesley and Ryan, 2012), and mitigation measures of disease transmission are in part based on the distance between farms. Fine-scale maps of farm distribution, including farm position and level of clustering, could thus make an important contribution to models that can inform control strategies (Bruhn et al., 2012). While broad-scale clusters of farms may be captured by aggregated data, the factors influencing farm distribution are poorly known at finer scales (Burdett et al., 2015). In the presence of aggregated census data, the distribution of individual farm locations have tended to be based on random allocation of points, regardless of other geographic information (Tildesley et al., 2010) or, in some cases, constrained by geographical information contained in probability surfaces (Bruhn et al., 2012; Burdett et al., 2015; Emelyanova et al., 2009; Tildesley and Ryan, 2012). However, none of these methods have captured both the number of points and the pairwise interaction between points (first and second order characteristics) to predict the spatial clustering of farms as well as differences in their broader distributions.

In this paper, we investigated the use of point-pattern models as a way to predict the distribution of individual farms both in terms of spatial clustering and in terms of dependency on external variables influencing their presence. This approach may provide more realistic representations of animal distribution at fine spatial scales than continuous pixel-based distributions, especially for species such as poultry and pigs that may be raised in high numbers in single premises. Our analyses focused on Thailand chicken farms, as an example of a middle-income country where extensive production systems (backyard poultry farms) coexist with intensive ones (large-scale chicken farms) (Van Boeckel et al., 2012).

2. Methods

2.1. Data

A detailed census of poultry holders was conducted in 2010 by the Department of Livestock Development (DLD), Bangkok, Thailand. The census included the number of chickens per owner for all farms in Thailand. The administrative levels in Thailand are province, district, sub-district and village, the latter being the smallest. The three first levels have defined boundaries, while villages are recorded by coordinates, usually at the center of the main cluster of houses. During the census, the coordinates of each poultry holder were not collected. The coordinates of the village were subsequently linked to each poultry holder. The census recorded 1,936,590 chicken owners in a total of 62,091 villages. Henceforth, we will use the term ‘farm’ to represent both smallholders, who may be a single family with a few chickens, and large-scale farms having several thousand birds. Farms with no chickens were removed from the dataset. A set of Voronoi polygons (Okabe et al., 2000) was built from the village coordinates. The median area of the Voronoi polygons was 4 km^2, the mean area was 8 km^2 (Supplementary Material (SM) – Figure S1). A mask excluding permanent water bodies and the province and city of Bangkok was applied. Individual farms were assigned a random coordinate within their polygon excluding of masked areas. Our input data set thus did not include the exact locations of farms, but an approximate location. However, given the extent (whole of Thailand) and the resolution of our predictors (1km), we considered this loss of accuracy to have a negligible effect on our results.

The distribution of chickens per farm showed a clear bimodal pattern (Van Boeckel et al., 2012) and a threshold of 500 chickens per farm was used to separate extensive small-scale producers from intensive large-scale systems. This threshold maximized the correlation between the quantiles of the intensive and extensive distributions of animals per farm in the two groups and the quantiles of two normal distributions of same mean and standard deviation. This resulted in two datasets of 1,930,003 extensive farms with a median number of 20 chickens per farm, and 6,587 intensive farms with a median number of 8,000 chickens per farm. In the absence of other information on the farm (size, inputs, outputs, practices), we assumed flock size to be an acceptable proxy for the classification in ‘extensive’ or ‘intensive’ holdings.

Spatial predictor variables were selected to be both generic and available in databases with a global extent (Table 1,Fig. 1) so that the models and approaches followed in this study could be transferred to data-poor countries. The predictor variables were previously identified as having strong predictive capacity by Van Boeckel (2012). The logarithm (base 10) of human population density (Worldpop database, http://www.worldpop.org.uk was included as farms are unlikely to be located either in city centres or in completely remote areas. “Remoteness”, defined as the travel time to Bangkok and to the closest provincial capital, accounted for differences in accessibility to provincial or national markets through the road and railway networks. This was computed from Nelson’s accessibility which is based on a cost-distance algorithm in unit of time. The weighted surface accounts for transport networks, environment and political factors affecting travel times (Nelson, 2008). Thus, it also helps identifying areas less suitable for chicken farms. Tree cover or percentage of land covered by forest was included as areas covered by dense and permanent forest may also exclude poultry farming (Hansen et al., 2013). Cropland or percentage of land covered by crops accounted for areas providing access to grain for feed (Fritz et al., 2015).

Table 1. Predictor variables tested in our models

	Resolution (m)	*Units*	*Reference*
Human population density	1000	People per km^2	Worldpop database
Remoteness	1000	Minute	Nelson et al. 2008

Cropland	1000	Pixel % covered by crops	Fritz et al. 2015
Tree cover	1000	Pixel % covered by forest	Hansen et al. 2013

2.2. Sample areas

The analysis was applied on squares samples of equal area sampling the Thai territory (Fig. 2). This allowed keeping processing time reasonable by dealing with a fraction of the very numerous chicken farms in Thailand and also avoided computational difficulties at the complex edges of the country. Creating sample areas also allowed to cross-validate model results. The size and location of the sample areas were chosen to cover most of Thailand completely, to cover a sufficient number of farms, and to include a diversity of predictor values and farm densities. For intensive farms, Thailand was divided into square areas of 200 x 200 km, and we analysed only the 11 sample areas with over 250 farms (Fig. 2a). For extensive farms, 38 sample areas of 112 x 112 km, each having at least half over Thailand, were used (Fig. 2b).

2.3. Descriptive analysis

The distribution of extensive and intensive farm locations was investigated using point pattern analysis. We used the stationary and non-stationary Besag's L-function, a transformation of Ripley's K-function, to define the spatial pattern of intensive and extensive farms between three different broad types of point pattern: random, clustered and regular. The random case referred to the completely spatial randomness (CSR) or homogenous Poisson process model. The L-functions were estimated by sample areas with *Lest()* and *Linhom()* functions from the *spastat* package in R.

Ripley's K-function is a summary statistic of a point process, defined as the expected number of *r*-neighbours of a point of **X** divided by the intensity λ i.e.:

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

for any $r \geq 0$ at any location *u*, where *r* is the radius, λ is the homogeneous intensity of points, **X** is the point process and *u* is any location. This definition assumes that the process is stationary, which imply that the intensity is constant and does not depend on the location (Baddeley et al., 2015). The empirical K-function is a summary of the pairwise distances of a point pattern, which allows point patterns with different intensities to be compared, and the analysis of a pattern at different scales, since the function is normalized by the intensity. The empirical K-function is defined as

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

where *a* is the study area, *n* is the total number of points in *a*, the sum is taken over all ordered pairs of distinct points *i* and *j*, $d[i,j]$ is the distance between two points and $I(d[i,j] \leq r)$ is the indicator that equals 1 if the distance is less than or equal to *r*. The term $e[i,j]$ is the edge correction weight, which was discarded as the number of points considered in both datasets was very large. By using $\frac{a}{n(n-1)}$, it assumes that the process is stationary. An observed point pattern is considered as clustered, random or regular depending on whether its empirical K-function is respectively higher than, close to or lower than the K-function of a CSR, i.e. the curve of equation $y = \pi r^2$. In the case of a non-stationary process, a generalisation of the later should be used, the inhomogeneous K-function. This generalisation assumes that **X** is a point process with a non-constant intensity $\lambda(u)$ at each location u, i.e.

$$\hat{K}_{inhom}(r) = \left(\frac{1}{A}\right)\sum_{i}\sum_{j,\,i\,\neq\,j}\frac{1(d[i,j] \leq r)}{(\lambda(x_i)\lambda(x_j))}$$

where A is a constant denominator, and $d[i,j]$ is the distance between points x_i and x_j (Baddeley et al., 2000). Besag's L-function $L(r) = \sqrt{\frac{K(r)}{\pi}}$ is a transformation of the K-function for which a CSR is a straight line $L_{random}(r) = r$ when $L(r)$ is plotted against r.

2.4. Point pattern simulation

2.4.1. Model choice

To predict the spatial distribution of intensive and extensive farms as points, the Log-Gaussian Cox Processes (LGCP) model was used (Møller et al., 1998), with the Palm maximum likelihood method of parameter optimisation (Baddeley et al., 2015; Tanaka et al., 2008). The Palm maximum likelihood method provides almost the same results as the minimum contrast method and our study may be done with both of these algorithm (Baddeley et al., 2015).

We compared the five processes modelling clustered point patterns; the Matérn cluster process, the Thomas process, the Cauchy cluster process, the Variance gamma cluster process and the LGCP with exponential covariance function (SM-Figure S 2) (Baddeley et al., 2015). These models were fitted on one sample area of 200 km length in Thailand using the intensive dataset, including covariates with the command line *kppm(X, ~ Hpop + Crop + Tree + Remot + I(Hpop^2)+ I(Crop^2) + I(Tree^2) + I(Remot^2), clusters = c("Thomas","MatClust","Cauchy","VarGamma","LGCP"), method = "palm")* using the *kppm()* function from *spatstat* package in R (all other arguments had default settings). The covariates were selected based on the Akaike Information Criterion (AIC) as below. We assessed how these different models were able to reproduce the clustering of the observed point pattern by using the two-sided global rank envelope test. The hypothesis tested by the rank envelope test is that the model tested can explain the process from which the observed point pattern originates. The test provides a p-value and a graphical representation of the envelope. The p-value decreases when the empirical L-function goes out of the global rank envelope. It was implemented based on extreme rank lengths with the *global_rank_envelope()* function from *GET* package in R (Mrkvička et al., 2017; Myllymäki et al., 2017) for 100,000 simulations of each model. The extreme rank lengths type was selected because it allowed to run fewer simulations (Mrkvička et al., 2016; Myllymäki et al., 2017). The conclusion of the extreme rank envelope test was that the LGCP performed best. It had by far the highest p-value, 5.40e-02, compared to the other models with a p-value of 2.80e-04, 2.40e-04, 1.08e-03, 6.48e-03, for the Matern, Thomas, Variance Gamma, Cauchy models, respectively (SM – Figure S 3). Hence, LGCP was used for all subsequent modelling of clustered point patterns.

2.4.2. Model fitting and validation

Four different types of model were built and compared: (i) “CSR”: a completely spatial randomness (CSR) or homogenous Poisson process model, which randomly distributed farms; (ii) “iCSR”: inhomogeneous Poisson process model, a CSR in which the average density of points is spatially varying. The average density is an intensity function $\lambda(u)$ of spatial location u. In our model, the intensity was modelled as $\lambda = \exp{(covariates)}$; (iii) “LGCP”: a LGCP model with a homogeneous intensity (without any covariates) with an exponential covariance function (Baddeley et al., 2015); and (iiv) “iLGCP”: a LGCP model with covariates predicting an inhomogeneous intensity and identifying highly probable locations for clusters. iLGCP was defined with a covariate exponential function and a random intensity modelled as $\lambda = \exp$ $(covariates)$. For the later model, the AIC was used to select the best combination of predictor variables:

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

where *PL* is the maximised Palm likelihood of the fitted model, and *edf* the effective degrees of freedom of the model (Baddeley et al., 2015- section 12.6.4; Tanaka et al., 2008). The AIC values of the models with different combination of covariates were compared on the 11 areas for the intensive farms dataset using the standardized difference with null model AIC,

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

where AIC_{null} is the AIC of a LGCP model without covariates and AIC_{model_i} is the AIC of i^{th} *LGCP* models with a set of variables. The model showing the greatest (positive) difference with the AIC_{null} model was selected for both non-stationary models, the iCSR and the iLGCP. This was implemented with the functions *ppm()* and *kppm()* from the R package *spatstat* when the model was the CSR and LGCP, respectively. The relative importance of each predictor variable was estimated as the exponential of the coefficient value of a covariate multiplied by the range of values of the covariate (Baddeley et al., 2015).

We aimed to evaluate the goodness-of-fit of our simulated patterns in their capacity to reproduce both the level of clustering and the location of clusters in comparison to the observed patterns. For each sample area and type of model, and using the best-fit parameters, we simulated 1500 and 8000 point patterns for extensive and intensive datasets, respectively. The number of simulations was chosen to balance the stability of the p-value and computing time (SM – Figure S 4). We implemented the global rank envelope test again to quantify the similarities in the level of clustering. This function allows a point pattern to be characterised independently from the density of points, which enabled the comparison of the p-values across simulations and areas. We then looked at the proportion of sample areas with significant p-values. To evaluate the goodness-of-fit of the simulated patterns in terms of location of the clusters, each sample area was further divided into 64 square quadrats. The correlation coefficient between the observed and modelled number of farms per quadrat for each simulation was computed. Quadrats intersecting the Thai border were removed when less than 95% of their area was in Thailand. Quadrat size was chosen to have a sufficient number of quadrats and of points per quadrat to produce a meaningful correlation coefficient (SM - Figure S 5). In addition to goodness-of-fit methods estimated for each model type (CSR, iCSR LGCP and iLGCP) on the calibration area, we also estimated goodness-of-fit methods (global rank envelope test and correlation coefficient) on a different sample area from the model calibration area, henceforth referred to as the validation area.

3. Results

Intensive farms were clustered, as assessed by the L-functions (Fig. 3). Extensive farms were randomly distributed, L-function being around the CSR case L-function. Empirical non-stationary L-functions (L-inhom on Fig. 3) were closer to CSR case than the stationary L-functions (L-hom on Fig. 3). All four spatial predictors and their quadratic terms were included in the non-stationary models (iCSR and iLGCP), following the comparison of AIC on the intensive farms dataset (Fig. 4). This intensity function was defined as

$$\lambda(u) = \exp\left(\beta_0 + \beta_1 Hpop(u) + \beta_2 Remot(u) + \beta_3 Crop(u) + \beta_4 Tree(u) + \beta_5 Hpop^2(u) + Remot^2(u) + \beta_7 Crop^2(u) + \beta_8\right.$$

,

with $\beta_0, \beta_1 \ldots, \beta_8$ to parameters to be estimates, *Hpop* the human population density, *Remote* the remoteness, *Crop* the cropland and *Tree* the tree cover.

(a) Intensive dataset

CALIBRATION

Significance threshold	0,001	0,01	0,05	0,1
CSR	100	100	100	100
iCSR	91	100	100	100
LGCP	0	27	64	73
iLGCP	27	36	73	82

VALIDATION

Significance threshold	0,001	0,01	0,05	0,1
CSR	100	100	100	100
iCSR	82	91	91	100
LGCP	0	27	64	73
iLGCP	45	64	73	73

(b) Extensive dataset

TRAINING

Significance threshold	0,001	0,01	0,05	0,1
CSR	100	100	100	100
iCSR	100	100	100	100
LGCP	0	16	34	45
iLGCP	24	53	68	76

VALIDATION

Significance threshold	0,001	0,01	0,05	0,1
CSR	100	100	100	100
iCSR	100	100	100	100
LGCP	50	68	79	79
iLGCP	79	89	92	95

Table. 2. Proportions of sample areas with a significant p-value at different significance thresholds.

In terms of indicators of level of clustering (Table. 2a and b), measured with the global rank envelope test, LGCP and iLGCP reproduced the observed level of clustering better than the random models (CSR and iCSR), having higher p-values in almost all sample areas from both datasets. CSR and iCSR models were almost always highly significant ($p<0.05$), thus neither models explained the observed point patterns. LGCP was more often the best model but did not explain the data in all sample areas since their p-values were significant in some areas. In sample areas where a model was not rejected, both LGCP and iLGCP performed well for the intensive dataset. LGCP and iLGCP were significant at 0.05 in 64% and 73% of cases for the calibration and the validation. However, LGCP performed better than iLGCP in extensive dataset. LGCP and iLGCP were significant at the $p<0.05$ level on extensive dataset, in 34% and 68% of cases for calibration and 79% and 92% of cases validation. However, the variance of LGCP models was higher than iLGCP models. iLGCP models were then more easily rejected by the global rank envelope as seen with the width of the envelopes (Fig. 5).

In terms of location of clusters (Fig. 6a and b), the models with covariates (iCSR and iLGCP) performed better than the models without (CSR and LCGP). The two sets of metrics of the iCSR and iLGCP models in the calibration and validation areas had significantly higher correlation coefficients than the other models (CSR and LGCP), for both intensive and extensive farm point patterns. This result was expected since these models are inhomogeneous, having an intensity explained by covariates. The medians of the correlation coefficients of iCSR and iLGCP were generally higher for the extensive than for the intensive dataset. However, correlation coefficients were slightly higher for iCSR models compared to iLGCP models in both calibration and validation area. The medians of the correlation coefficients of the different models (CSR, iCSR, LGCP and iLGCP (calibration and validation)) were 0.008, 0.565, 0.004, 0.411, -0.006, 0.521, -0.002 and 0.356 for the intensive dataset and 0.006, 0.752, 0.003, 0.631, 0.000, 0.711, 0.007 and 0.576 for the extensive dataset. Taking into account both indices, of the level of clustering and the location of clusters, iLGCP performed the best. We provided as an illustration a simulation produced by the four model types (CSR, iCSR, LGCP and iLGCP) applied to a sample are from intensive and extensive farms datasets and a plot of the observed farm patterns (Fig. 5), along with the plot of the global rank envelope test.

The coefficients of the different iLGCP model parameters for both intensive and extensive datasets are presented in Fig. 8. Human population density was by far the most important predictor of intensive and extensive models on average, followed by tree cover, cropland and remoteness (Fig. 7), and the relative importance of predictor variables were similar for the intensive and extensive farms.

4. Discussion

In this paper, we explored the potential of point pattern simulation models to reproduce real-world distribution of intensive and extensive chicken farms. The implementation of these models allowed to produce a set of discrete and realistic point locations. Our iLGCP models were able to reproduce the level of clustering and the local density of farms better than the other models. LGCP models reproduced the level of clustering, but not the cluster location well, whereas iCSR located the clusters well, but did not capture the level of clustering. Extensive farm distribution was closer to a random distribution than intensive farms, and these simulations benefitted less from using a LGCP. Conversely, intensive farms were more clustered, so the LGCP models reproduced these patterns much better than the random model, but the quality of the prediction of local densities was lower.

Our result indicated clearly the need to account for clustering in the distribution of intensive farms. Such clustering of farms may enable farmers to benefit from economies of scale (Van Boeckel et al., 2012), or facilitate operations for contract farming. Many farmers in Thailand operate as contractors for large consolidator companies such as Charoen Pokphand (CP). Farms directly owned by CP may also be clustered. Also, as described by (Feder et al., 1985), the adoption of agricultural innovations in developing countries is affected by group influences on individual behaviour. The presence of a well-established, successful, intensive poultry farm may stimulate similar economic activity nearby. The improved prediction of intensive farm locations by including clustering thus makes sense.

More surprising was the dominance of human population density as a predictor of intensive farms since broiler production in Thailand was previously described as being mainly located around hatcheries, feed mills and processing plants (Costales, 2004; NaRanong, 2007), but these may themselves correlate to human population. The association with human population density could relate to market access, and the model typically placed intensive farms in areas with intermediate human population density, such as in peri-urban areas. The establishment of a chicken farm is thus constrained by a trade-off between market access and the cost of land, which may become prohibitive in more urbanized areas. Our results contrasted with the results of Van Boeckel et al. (2012), who showed cropping factor had a stronger effect than human population in their logistic regression models of presence/absence of intensively raised chickens. This variable was not included in our model since it was not available globally. Methodological differences may also explain the lower effect of some factors. Van Boeckel et al. (2012) analysed the entire extent of Thailand, whereas our models were trained within much smaller spatial units. Further predictors may be worth including, if available at the global level. Other accessibility predictors, such as travel distance to ports where feed could be imported, or where outputs could be exported may improve our predictions. Another global predictor which could provide valuable information on access to service and markets is the location of settlement.

At the local scale, a degree of "random noise" in the location of intensive farms is inevitable, which we did not expect to capture. The initial establishment of an intensive farm may be influenced both by fine-scale spatial factors (i.e. land availability, location suitability and access to inputs and markets) and by individual farmer characteristics (i.e. where they live, the locations of their other investments, their family history and land ownership). Such factors would be difficult to account for in models at the scale considered here. At the scale of the variables used in our models, several sites may then seem equally suitable for setting up a farm, for example, by having an easy access to markets and inputs such as feed. This does not interfere with our objective to depict a realistic distribution of farms.

The distribution of extensive farms was less clustered, and more readily predicted by human population density. This fitted our expectations because extensively raised chickens are typically owned as backyard poultry by rural populations (Van Boeckel et al., 2012).

The resolution of the sample areas should not influence our results, since sample areas were chosen to optimise the variability of situations encountered within Thailand in terms of predictor values and density of farms. The reason why the variability of model performed in the different sample areas could be due to the range of predictor value which differ from one area to another. An analysis on whole Thailand would only deal with its geometry and the number of points in the extensive dataset (leading to computation problems).

Our results indicate that a producing point-based distribution maps of intensive and extensive flocks is feasible. To use this approach in data-poor countries with a comparable farming system, an important next step will be to validate the model in a different country, but with similar environmental conditions, such as Vietnam, where detailed census data exist. Eventually, it would be interesting to investigate how the extensive and intensive models could predict the distribution of farms according to different levels of intensification. One could imagine high-income countries where 99% of the production is intensive to be best predicted by the intensive model alone, and, conversely, that the extensive model could be tested in low-income countries. In intermediate situations, one could apply both models according to the proportion of extensively raised poultry predicted at the national level by Gilbert et al. (2015). To predict farm locations into these different situations, LGCP models would be applied with the same parameters in neighboring countries or in countries with similar agro-ecologies. Several datasets would thus be required to predict farm distribution into countries from different regions or environments. Further extension of this work will lead to the development of entire farm allocation models, where the total number of animals of an administrative unit could be allocated to farms at locations predicted by the LGCP simulation model in such a way to reproduce a given distribution of animals per farm. While artificial-intelligence-based image processing may soon allow to detect most intensified livestock raising infrastructure automatically, it would not detect middle-size commercial poultry farms, which still exists in large numbers in Thailand, and that can look like a normal building. We believe that statistical approaches such as these still hold value for different settings but also for hind- and forecasting of the farming distribution.

Other types of livestock production may benefit from similar approaches. Pig farming, for example, is also disconnected from the land and could be subject to similar spatial constraints linked to feed availability and market access. In contrast, the distribution of grazing ruminant farms may have very different spatial determinants. The dependence on large areas for grazing may result in a more homogenous spatial distribution (except for feedlot cattle). Land-use predictor variables such as rangeland or pastures may thus become more important factors.

Middle- and low-income countries are those where this approach bears the greatest value, in relation to the data scarcity some face, and the co-existence, to varying degrees, of extensive and intensive production. While in Brazil livestock data are available at fine scale, in some other large livestock producing countries, such as China and India, livestock data are only available at coarse resolution. These are precisely where the impact of livestock diseases on livelihoods, animal and human health are greatest (Childs et al., 2007) and where good quality data may help with disease prevention. In high-income countries, where intensive production dominates, results like ours offer an interesting substitute to the original data protected by privacy laws.

5. Conclusions

We developed farm distribution models using a point pattern modelling method, which allowed the simulation of chicken farm distributions both in terms of spatial clustering and location of clusters. The methods used 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 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

The authors would like to thank the “Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture” (FRIA) and the FNRS PDR “Mapping people and livestock” (PDR T.0073.13) who supported this project. We also thank the Department of Livestock Development, Thailand for chicken census data. Christophe A.N. Biscio is supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 “Statistics for point processes in space and beyond”, and by the “Centre for Stochastic Geometry and Advanced Bioimaging”, funded by grant 8721 from the Villum Foundation.

References

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

Baddeley, A.J., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 54, 329–350. https://doi.org/10.1111/1467-9574.00144

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. https://doi.org/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.

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

Feder, G., Just, R., Zilberman, D., 1985. Adoption of Agricultural Innovations in Developing Countries: A Survey. Economic Development and Cultural Change 33, 255–98.

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. Global Change Biology 21, 1980–1992. https://doi.org/10.1111/gcb.12838

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

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., Vanwambeke, S., Wint, G., In press. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Nature Scientific Data.

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

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., 2013a. Zoonosis emergence linked to agricultural intensification and environmental change. Proceedings of the National Academy of Sciences 110, 8399–8404. https://doi.org/10.1073/pnas.1208059110

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., 2013b. Zoonosis emergence linked to agricultural intensification and environmental change. Proceedings of the National Academy of Sciences 110, 8399–8404. https://doi.org/10.1073/pnas.1208059110

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

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

Martin, M.K., Helm, J., Patyk, K.A., 2015. An approach for de-identification of point locations of livestock premises for further use in disease spread modeling. Preventive Veterinary Medicine 120, 131–140. https://doi.org/10.1016/j.prevetmed.2015.04.010

Mennerat, A., Nilsen, F., Ebert, D., Skorping, A., 2010. Intensive Farming: Evolutionary Implications for Parasites and Pathogens. Evolutionary Biology 37, 59–67. https://doi.org/10.1007/s11692-010-9089-0

Møller, J., Syversveen, A.R., Waagepetersen, R.P., 1998. Log Gaussian Cox Processes. Scandinavian Journal of Statistics 25, 451–482. https://doi.org/10.1111/1467-9469.00115

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

Mrkvička, T., Hahn, U., Myllymaki, M., 2016. A one-way ANOVA test for functional data with graphical interpretation. arXiv:1612.03608 [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

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., Hahn, U., 2017. Global envelope tests for spatial processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 381–404. https://doi.org/10.1111/rssb.12172

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

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

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

Okabe, A., Sugihara, K., Kendall, D.G., Boots, B., 2000. Spatial tessellations: concepts and applications of Voronoi diagrams, 2nd ed. ed, Wiley series in probability and statistics. Wiley, New York.

Prosser, D.J., Wu, J., Ellis, E.C., Gale, F., Van Boeckel, T.P., Wint, W., Robinson, T., Xiao, X., Gilbert, M., 2011. Modelling the distribution of chickens, ducks, and geese in China. Agriculture, Ecosystems & Environment 141, 381–389. https://doi.org/10.1016/j.agee.2011.04.002

Pulliam, J.R.C., Epstein, J.H., Dushoff, J., Rahman, S.A., Bunning, M., Jamaluddin, A.A., Hyatt, A.D., Field, H.E., Dobson, A.P., Daszak, P., the Henipavirus Ecology Research Group (HERG), 2012. Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. Journal of The Royal Society Interface 9, 89–101. https://doi.org/10.1098/rsif.2011.0223

Reeves, A., 2012. Construction and evaluation of epidemiologic simulation models for the within- and among-unit spread and control of infectious diseases of livestock and poultry (Thesis). Colorado State University. Libraries.

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

Slingenbergh, J., Cecchi, G., Engering, A., Hogerwerf, L., 2013. World livestock 2013: changing disease landscapes.

Smil, V., 2002. Worldwide transformation of diets, burdens of meat production and opportunities for novel food proteins. Enzyme and Microbial Technology 30, 305–311.

Steinfeld, H., 2004. The livestock revolution—a global veterinary mission. Veterinary Parasitology 125, 19–41. https://doi.org/10.1016/j.vetpar.2004.05.003

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

Tanaka, U., Ogata, Y., Stoyan, D., 2008. Parameter Estimation and Model Selection for Neyman-Scott Point Processes. Biometrical Journal 50, 43–57. https://doi.org/10.1002/bimj.200610339

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

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

Van Boeckel, T.P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B.T., Levin, S.A., Laxminarayan, R., 2014. Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. The Lancet infectious diseases 14, 742–750.

Van Boeckel, T.P., Prosser, D., Franceschini, G., Biradar, C., Wint, W., Robinson, T., Gilbert, M., 2011. Modelling the distribution of domestic ducks in Monsoon Asia. Agric Ecosyst Environ 141, 373–380. https://doi.org/10.1016/j.agee.2011.04.013

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

Figure captions

Fig. 1. Predictors values. Human population density (logarithm of human population density in heads per km^2), Remoteness (travel time to province capital cities in minute), Cropland (percent of pixel covered by crops) and Tree cover (percent of pixel covered by forest).

Fig. 2. Sample areas defined for the study (a) 11 sample areas of 200 km length side defined for the intensive dataset (b) 38 sample areas of 112 km length side defined for the extensive dataset.

Fig. 3. **Descriptive analysis of intensive and extensive farms datasets using stationary and non-stationary L-functions.** Each dashed line represents the empirical L-function, L(r), estimated from the observed point pattern from each sample area, and *r* is the radius in meters. Comparing the empirical L-functions of a point pattern with the theoretical L-function of a completely spatial randomness (CSR) enables to determine if a pattern is clustered, random or regular, with L-functions higher than, close to or lower than the CSR case, respectively. Dashed grey line: stationary empirical L-function, $L_{hom}(r)$, for each sample area; dashed blue lines: non-stationary empirical L-function, $L_{inhom}(r)$, for each sample area; black line: theoretical L-function, $L_{poisson}(r)$, of a CSR.

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 sample area of the intensive dataset and the black line the average line.

Fig. 5. The observed point pattern and examples of simulations produced by the four model types along with the global rank envelope test, for a sample area for both intensive and extensive datasets. The four models were the completely spatial randomness (CSR), the CSR with covariates (iCSR), the Log-Gaussian Cox process (LGCP) and the LGCP with covariates (iLGCP). In the global rank envelope test, with extreme rank lengths: dashed lines represent the 95% global envelope with 8,000 and 1,500 simulations for intensive and extensive datasets, respectively; black line: the empirical L-function estimated from the observed point pattern; and red points: the points of the empirical L-function which are outside the envelope.

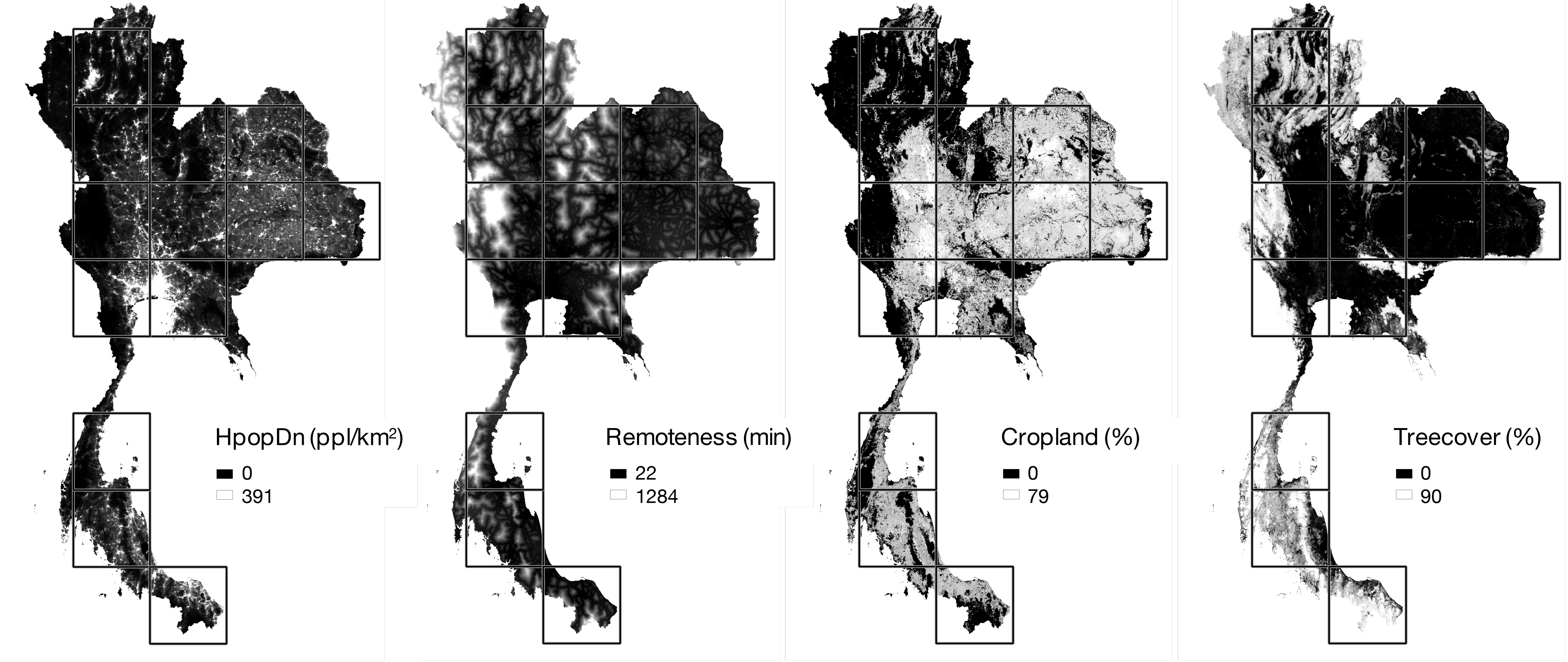
Fig. 6. Correlation coefficient between the numbers of points per quadrat for all quadrats in observed and each simulated pattern for a) extensive and b) intensive farms. The distribution of correlation coefficient values for all simulations (1500 and 8000 simulations for extensive and intensive datasets, respectively) on each area 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)), for calibration and validations areas.

Fig. 7. Relative covariates importance of iLGCP models by sample area with covariates for a) intensive and b) extensive dataset. Logarithm of the relative importance of each covariate and its quadratic term: human population density (Hpop + $Hpop^2$), tree cover (Tree + $Tree^2$), cropland (Crop + $Crop^2$) and the remoteness or accessibility (Remot + $Remot^2$).

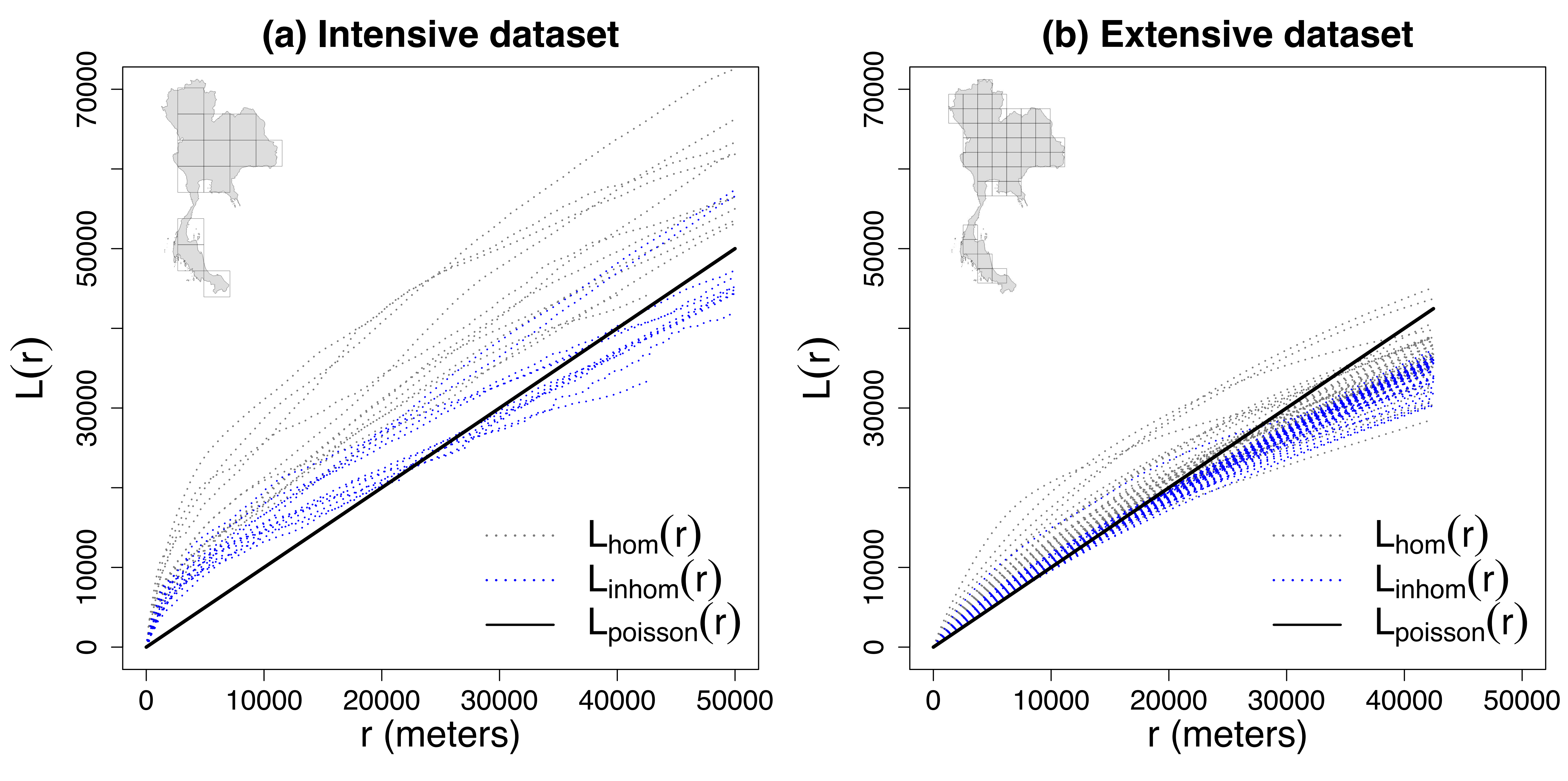
Fig. 8. Boxplots of the coefficients from the different iLGCP model parameters fitted on each sample area (α, σ^2 and $\beta_0, \beta_1 ... \beta_8$)**.** In LGCP models, the covariance is defined as $C_0(r) = \sigma^2 \exp\left(r/\alpha\right)$ where σ^2 is the variance and α the scale parameter and the intensity function was defined as $\boldsymbol{\lambda}(\boldsymbol{u}) =$

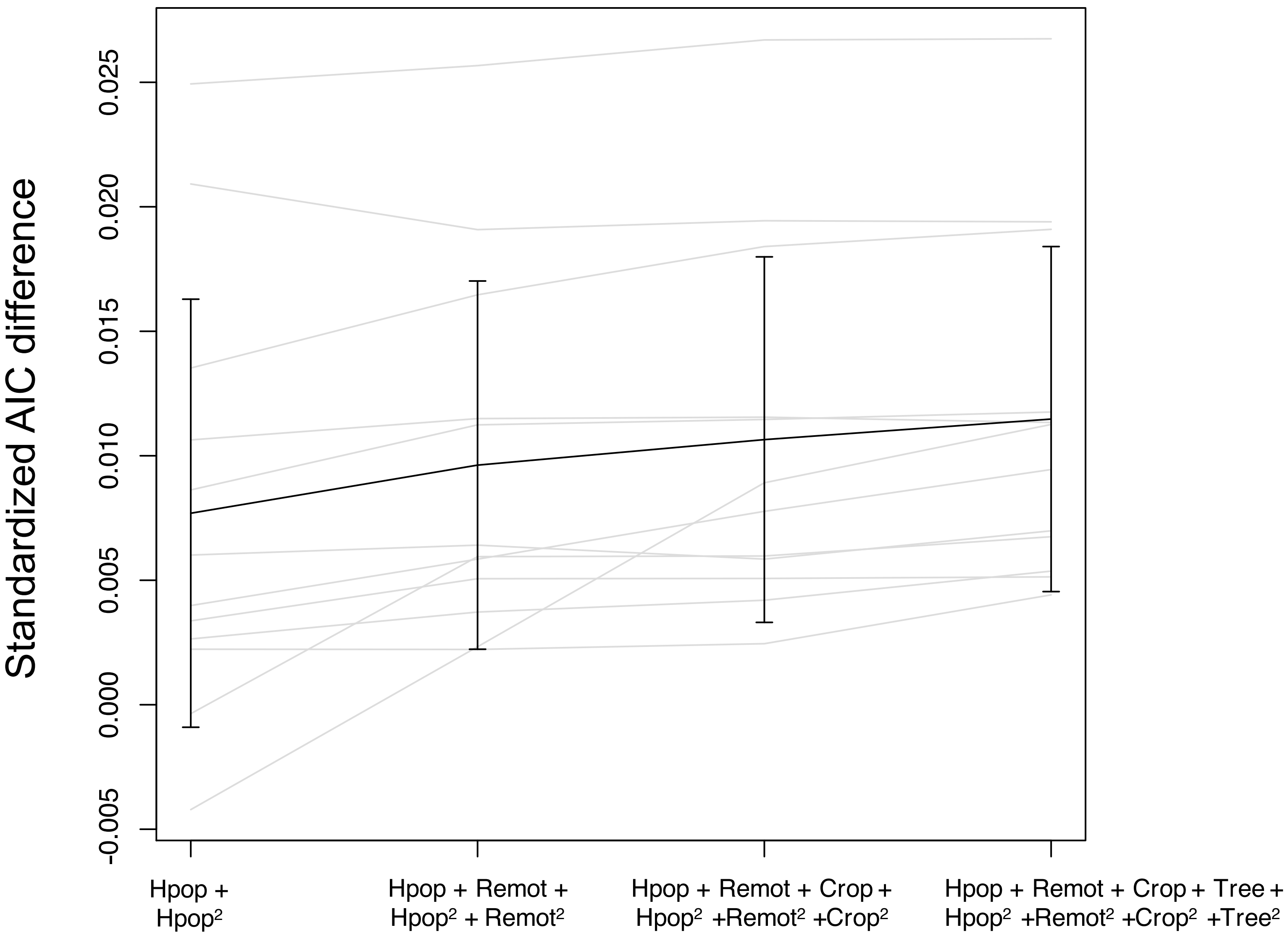
$$\mathbf{exp}\left(\boldsymbol{\beta_0} + \boldsymbol{\beta_1 Hpop}(\boldsymbol{u}) + \boldsymbol{\beta_2 Crop}(\boldsymbol{u}) + \boldsymbol{\beta_3} Tree(\boldsymbol{u}) + \boldsymbol{\beta_4 Remot}(\boldsymbol{u}) + \boldsymbol{\beta_5 Hpop^2}(\boldsymbol{u}) + \boldsymbol{\beta_6 Crop^2}(\boldsymbol{u}) + \boldsymbol{\beta_7 Tree^2}(\right.$$

.

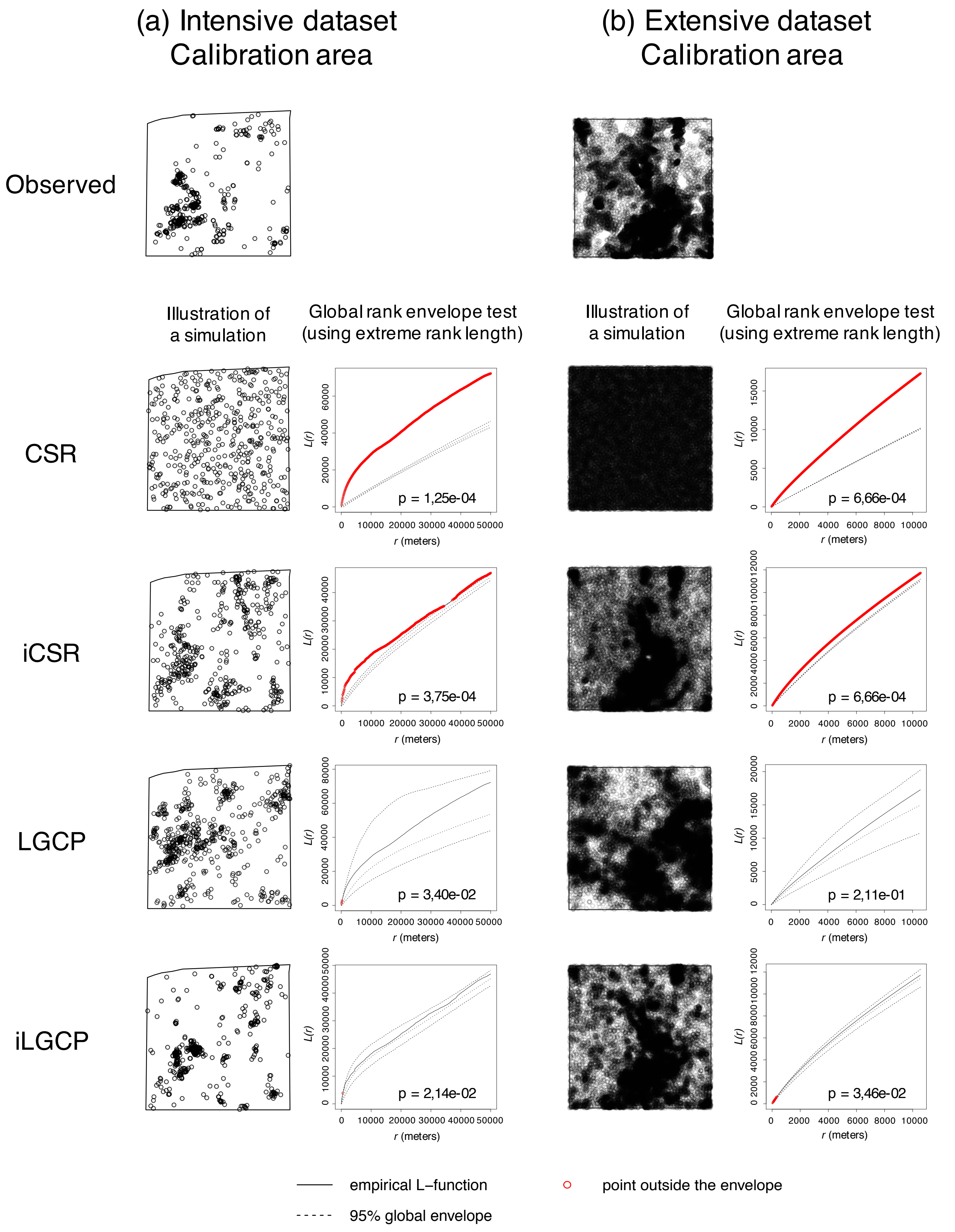

HpopDn (ppl/km2)
0
391
Remoteness (min)
22
1284
Cropland (%)
0
79
Treecover (%)
0
90


(a) Intensive farms
(b) Extensive farms
200 km
112 km

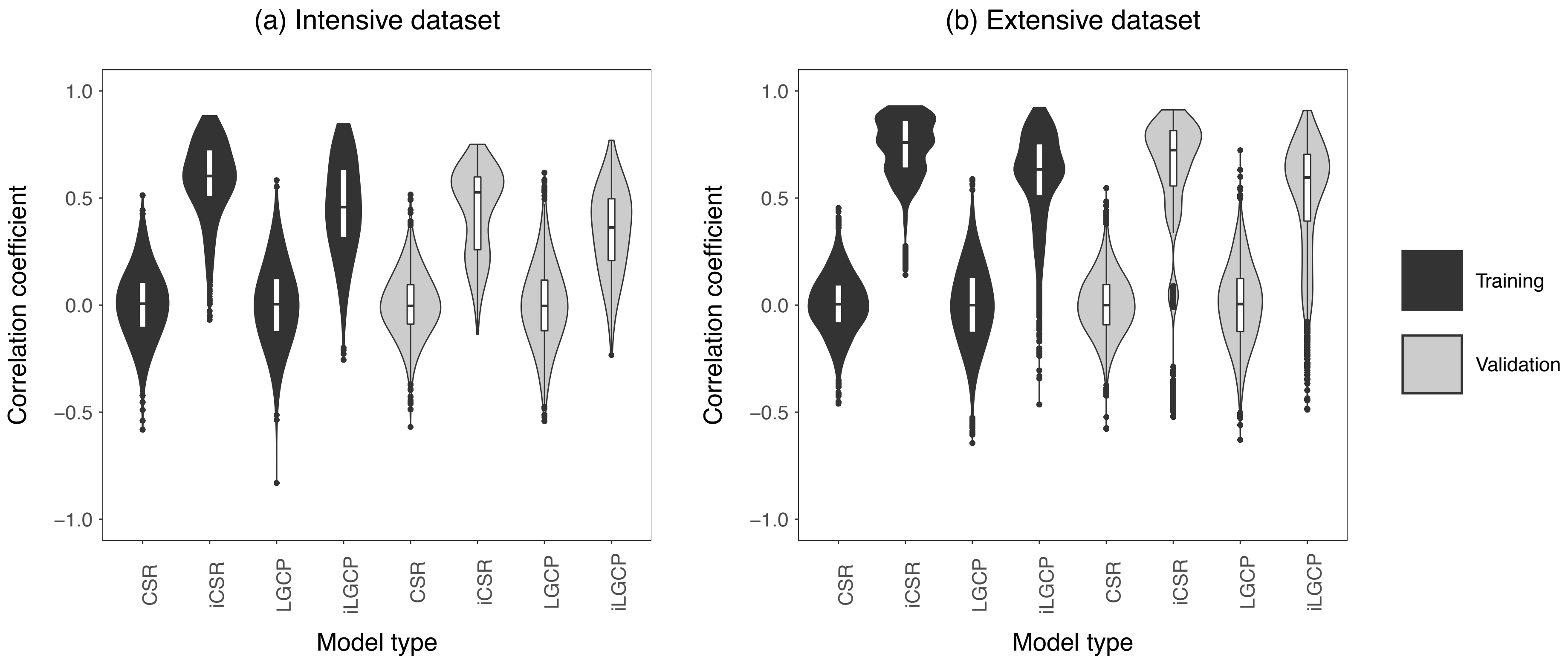

(a) Intensive dataset
(b) Extensive dataset
L(r)
r (meters)
0
10000
20000
30000
40000
50000
70000
$L_{hom}(r)$
$L_{inhom}(r)$
$L_{poisson}(r)$



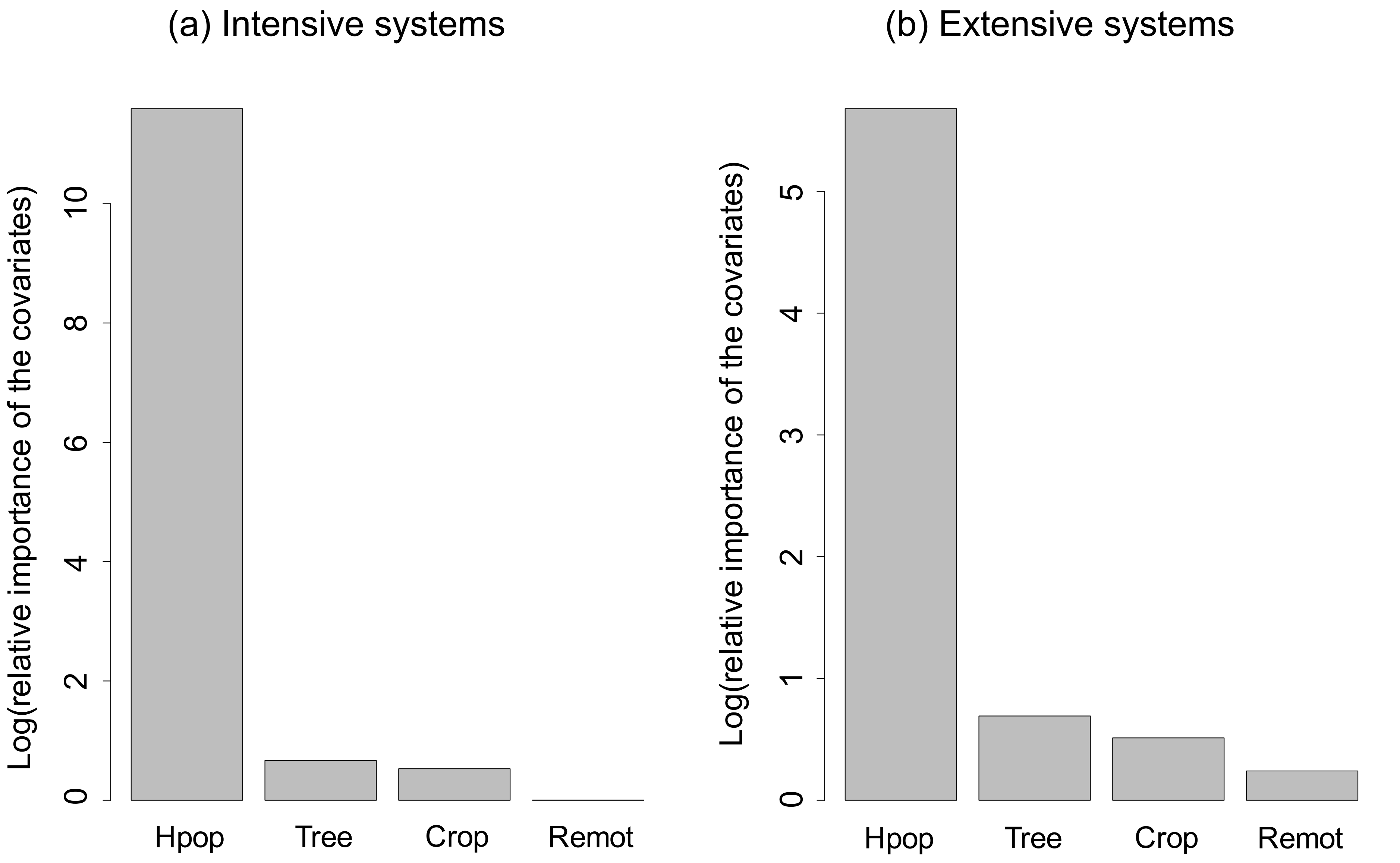
Standardized AIC difference
-0.005
0.000
0.005
0.010
0.015
0.020
0.025
Hpop + Hpop²
Hpop + Remot + Hpop² + Remot²
Hpop + Remot + Crop + Hpop² +Remot² +Crop²
Hpop + Remot + Crop + Tree + Hpop² +Remot² +Crop² +Tree²

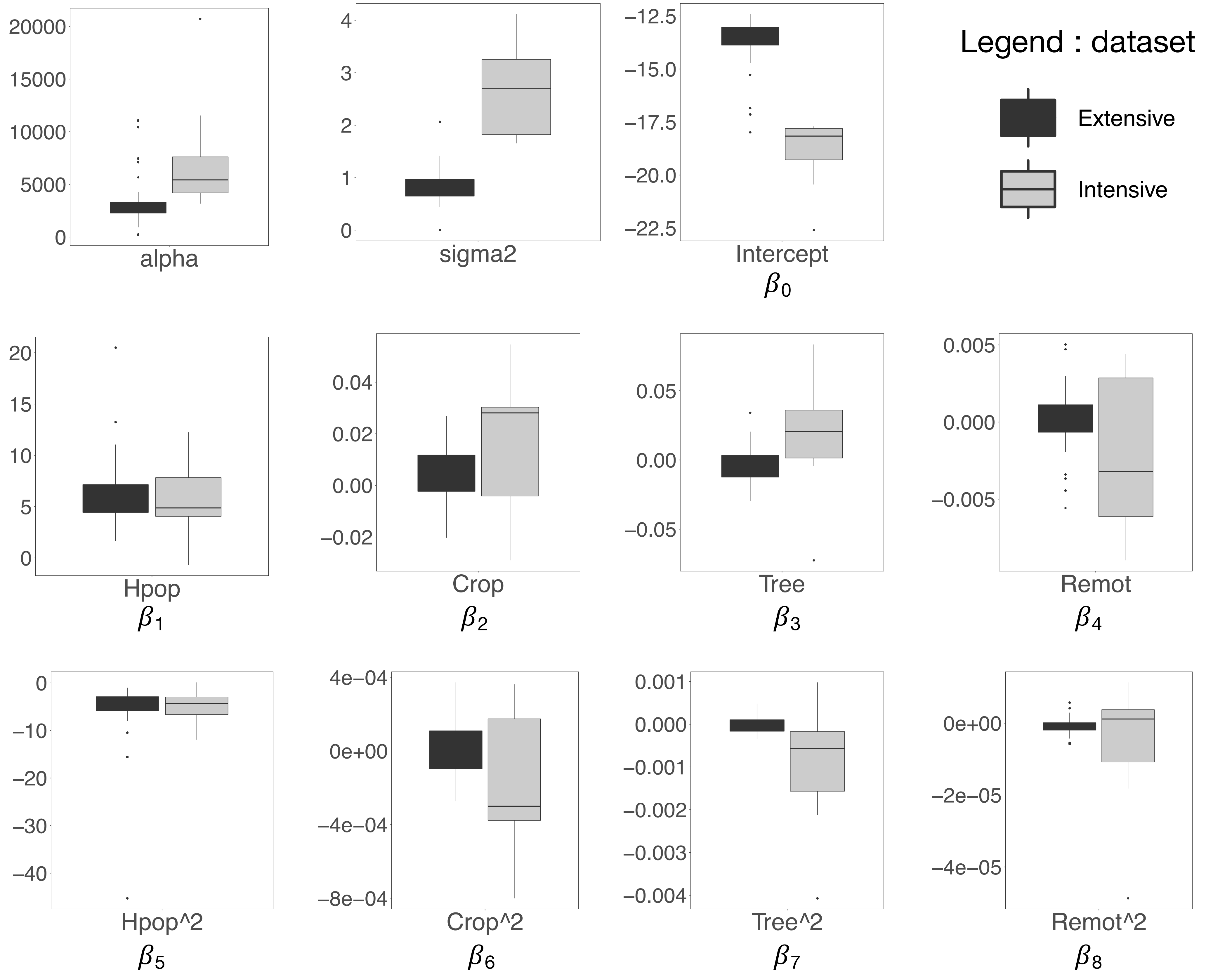


(a) Intensive dataset
Calibration area
(b) Extensive dataset
Calibration area
Observed
Illustration of
a simulation
Global rank envelope test
(using extreme rank length)
CSR
iCSR
LGCP
iLGCP
L(r)
r (meters)
p = 1,25e-04
p = 6,66e-04
p = 3,75e-04
p = 6,66e-04
p = 3,40e-02
p = 2,11e-01
p = 2,14e-02
p = 3,46e-02
empirical L–function
95% global envelope
point outside the envelope



(a) Intensive dataset
(b) Extensive dataset
Correlation coefficient
1.0
0.5
0.0
−0.5
−1.0
CSR
iCSR
LGCP
iLGCP
CSR
iCSR
LGCP
iLGCP
Model type
Training
Validation


(a) Intensive systems
Log(relative importance of the covariates)
0
2
4
6
8
10
Hpop
Tree
Crop
Remot
(b) Extensive systems
Log(relative importance of the covariates)
0
1
2
3
4
5
Hpop
Tree
Crop
Remot



Legend : dataset
Extensive
Intensive
alpha
sigma2
Intercept
β_0
Hpop
β_1
Crop
β_2
Tree
β_3
Remot
β_4
Hpop^2
β_5
Crop^2
β_6
Tree^2
β_7
Remot^2
β_8

Supplementary materials to « Point pattern simulation modelling of extensive and intensive chicken farming in Thailand: accounting for clustering and landscape characteristics »

Celia Chaiban, Christophe Biscio, Weerapong Thanapongtharm, Michael Tildesley, Xiangming Xiao, Timothy P Robinson, Sophie O Vanwambeke, Marius Gilbert

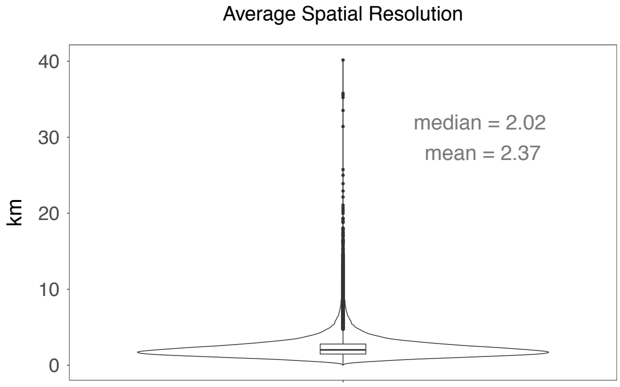
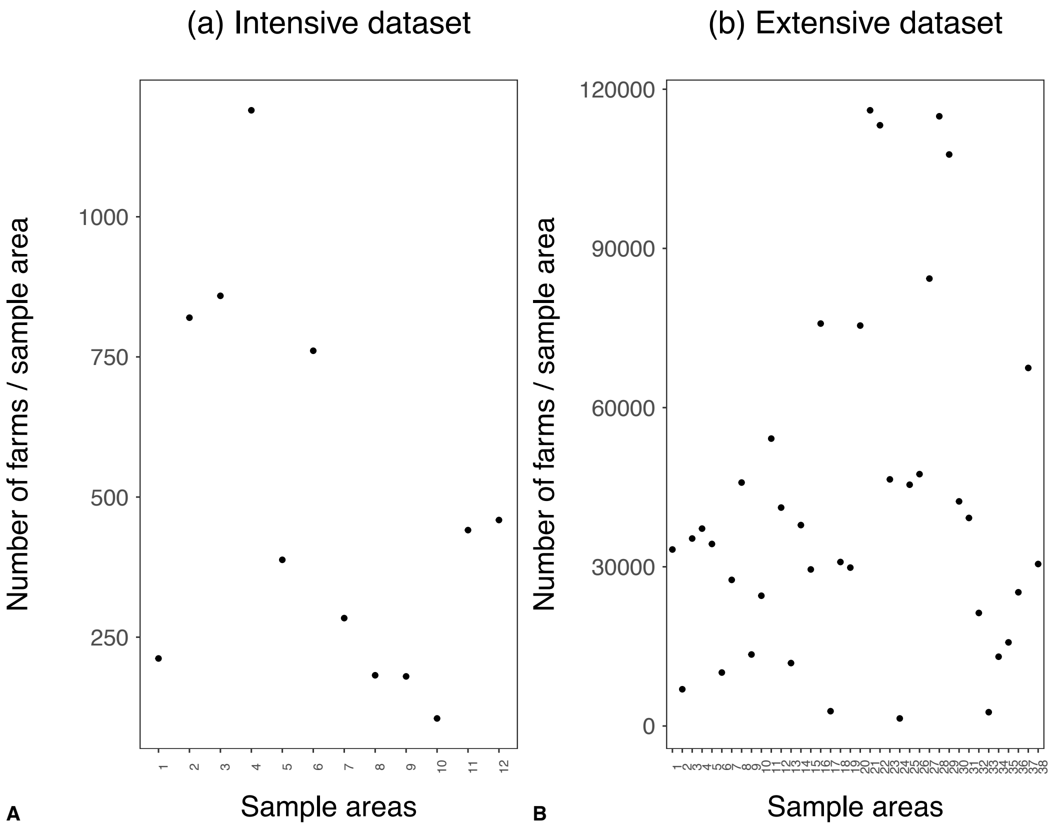


Figure S 1. Violin plot of the distribution of the average spatial resolution (root square of the area) of Voronoi polygons in kilometers



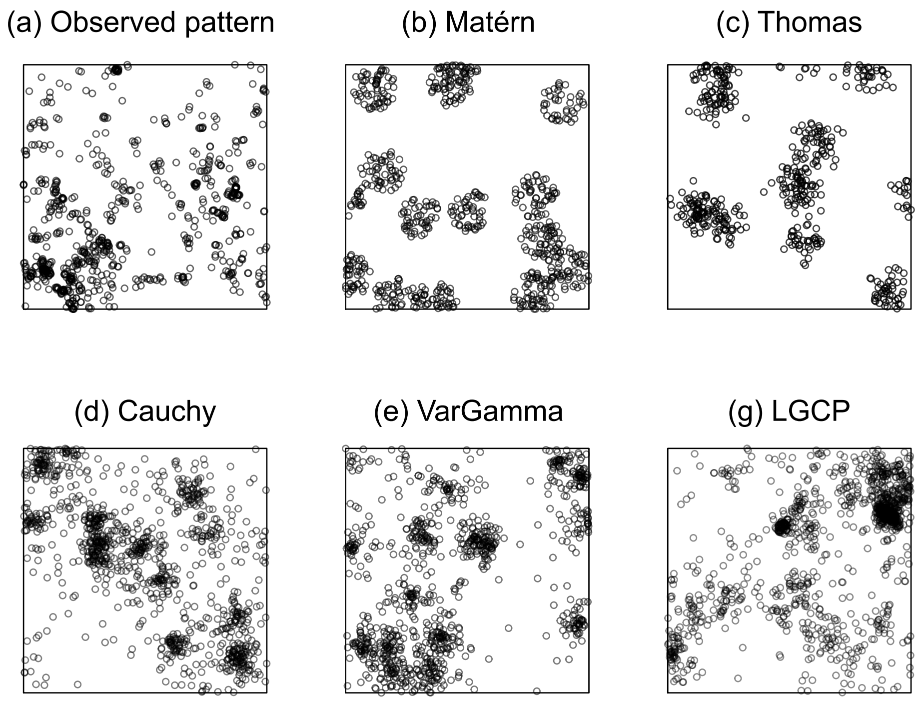


Figure S 2. Observed point pattern of a sample area from Thailand (a) compared to a simulation obtained by the different cox processes. (b) a Matérn process model (c) a Thomas process model (d) a Cauchy process model (d) a Variance Gamma process model (e) a Log-Gaussian Cox Processes model.

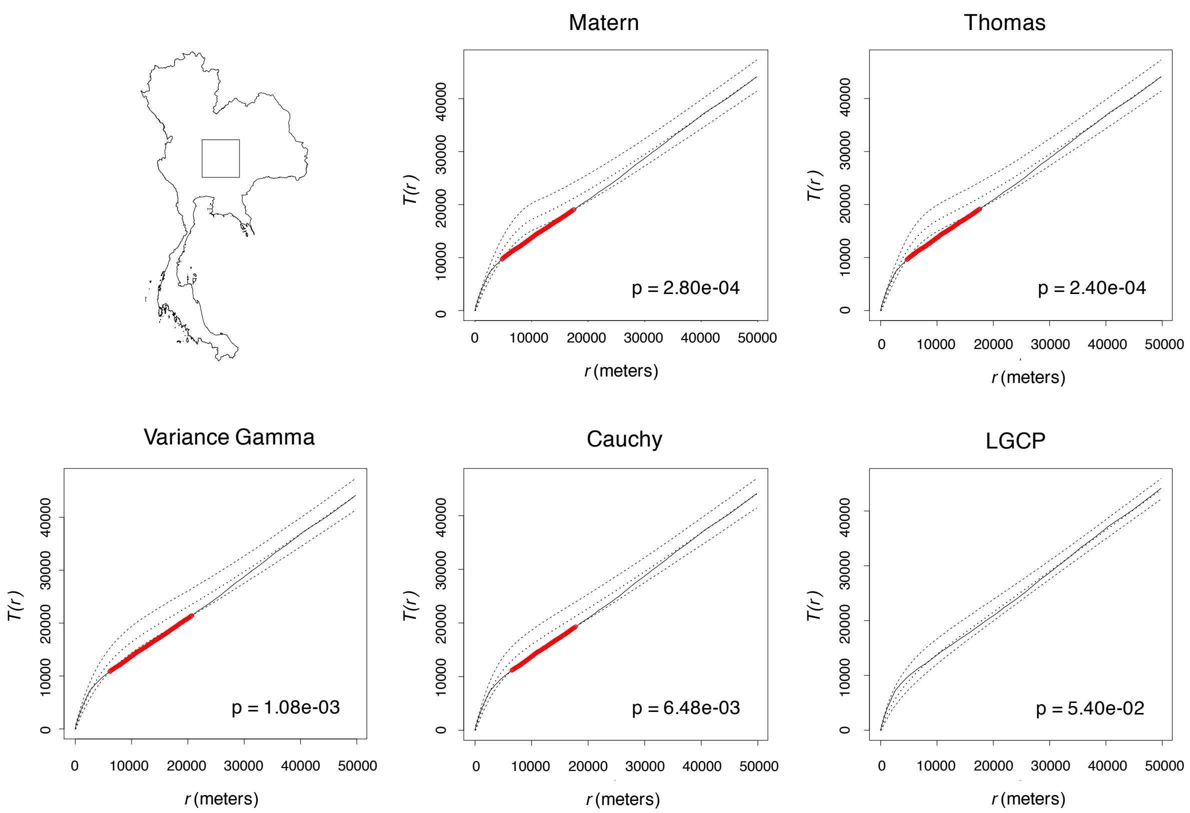


Figure S 3. Global rank envelope test on the five different processes, Matérn, Thomas, Cauchy, Variance Gamma and Log-Gaussian Cox Processes (LGCP), based on the L-function.

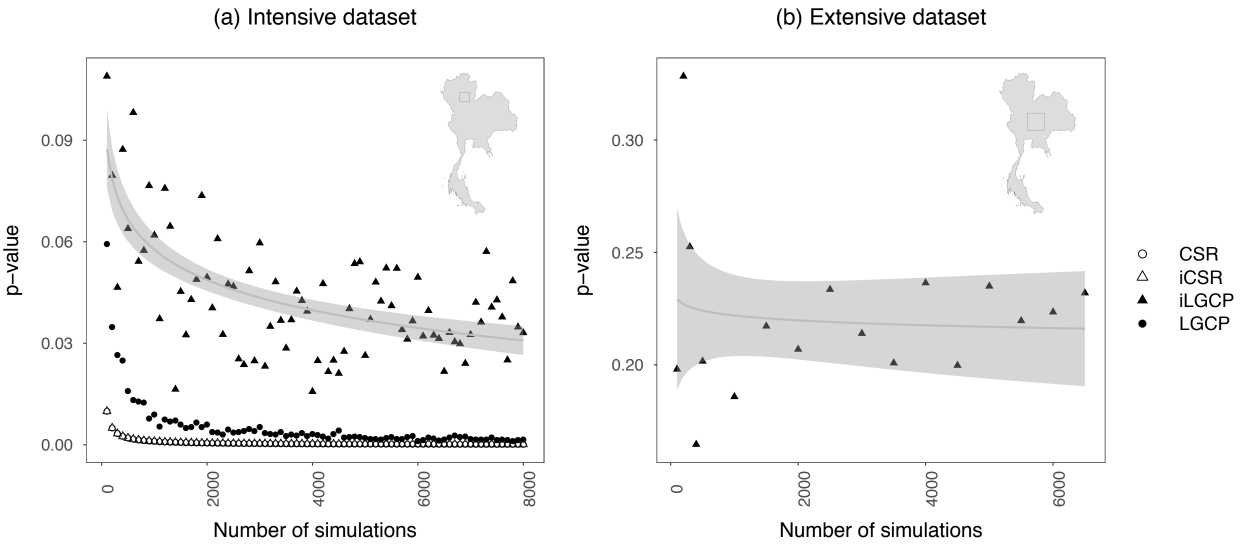


Figure S 4. Extreme rank envelope test p-values with different number of simulations for extensive (a) and intensive (b) datasets, on a sample area of Thailand.

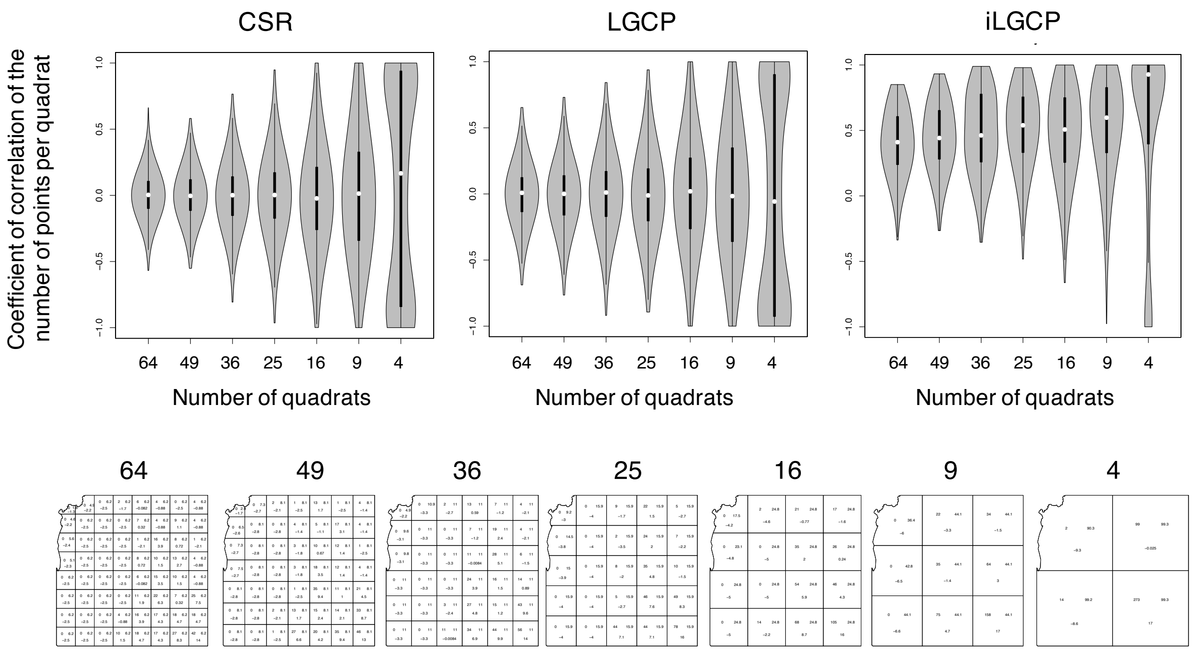


Figure S 5. Coefficient of correlation on the number of points per quadrats for different quadrat sizes for the Complete Spatial Randomness (CSR), the log-Gaussian cox Processes (LGCP) and the LGCP with covariates (iLGCP).

INTENSIVE DATASET

CALIBRATION

	B		D		K		L		N		O	
CSR	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***
iCSR	0,000375	***	0,000125	***	0,000125	***	0,000125	***	0,000250	***	0,000125	***
LGCP	0,502937		**0,091739**	.	0,001125	**	0,026747	*	0,004749	**	0,030746	*
iLGCP	**0,533558**		0,012373	*	0,000125	***	0,001250	**	0,000250	***	0,034746	*

	P		S		t		U		X	
CSR	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***
iCSR	0,000125	***	0,000250	***	0,004499	**	0,000125	***	0,000125	***
LGCP	0,002500	**	**0,238720**		0,036995	*	**0,147357**		0,033996	*
iLGCP	0,000750	***	0,026997	*	**0,107737**		0,093113	.	0,021372	*

VALIDATION

	B		D		K		L		N		O	
CSR	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***
iCSR	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000250	***
LGCP	**0,521435**		**0,066742**	.	0,008624	**	0,033621	*	0,004499	**	0,043745	*
iLGCP	0,000125	***	0,009749	**	0,000125	***	0,001500	**	0,000125	***	**0,245719**	

	P		S		t		U		X	
CSR	0,000125	***	0,000125	***	0,000125	***	0,000125	***	0,000125	***
iCSR	0,066742	.	0,000125	***	0,000125	***	0,002000	**	0,000125	***
LGCP	0,002500	**	**0,115486**		0,032871	*	0,146732		0,021997	*
iLGCP	**0,237970**		0,000125	***	0,041870	*	**0,316585**		0,000125	***

Table S 1 Extreme rank envelope test p-values per sample area of the intensive dataset in calibration and validation, for the different models: **completely spatial randomness (CSR); CSR with an inhomogeneous intensity (iCSR), log-Gaussian cox-processes (LGCP), LGCP with an inhomogeneous intensity (iLGCP).** Significance codes: '***' for 0.001, '**' for 0.01, '*' for 0.05 and '.' for 0.1. Grey highlighting: highest p-value; bold blue: highest and non-significate p-value.

EXTENSIVE DATASET

CALIBRATION

	A		B		C		D		E		F		G		H	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	2,94E-01		6,22E-01		**2,11E-01**		**2,36E-01**		**2,74E-01**		3,34E-01		**3,11E-01**		**2,06E-01**	
iLGCP	**4,28E-01**		**7,44E-01**		3,46E-02	*	3,66E-02	*	9,73E-02	.	**9,45E-01**		1,65E-01		1,83E-01	

	I		J		K		L		M		N		O		R	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	**3,42E-01**		**1,04E-01**		2,66E-03	**	1,27E-02	*	**3,75E-01**		**5,46E-02**	.	1,49E-01		3,46E-02	*
iLGCP	3,86E-02	*	8,59E-02	.	5,33E-03	**	7,33E-03	**	5,66E-02	.	6,00E-03	**	**2,81E-01**		7,24E-04	***

	S		t		U		V		W		X		Y		Z	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	1,07E-01		1,27E-02	*	2,03E-01		2,13E-02	*	2,66E-03	**	3,33E-03	**	3,13E-02	*	**2,96E-01**	
iLGCP	6,66E-04	***	3,33E-03	**	**9,65E-01**		6,66E-04	***	6,66E-04	***	6,66E-04	***	4,66E-03	**	2,00E-02	*

	AA		AB		AC		AD		AE		AG		AH		AI	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	**5,46E-02**	.	2,66E-03	**	1,13E-02	*	2,66E-03	**	4,66E-03	**	2,13E-02	*	**8,39E-02**	.	**2,71E-01**	
iLGCP	3,33E-03	**	6,00E-03	**	6,66E-04	***	6,66E-04	***	6,66E-04	***	5,33E-03	**	3,33E-03	**	4,66E-03	**

	AJ		AK		AM		AN		AO		AQ	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	**2,93E-01**		**5,04E-01**		**6,46E-01**		**4,29E-01**		**5,26E-02**	.	**3,98E-01**	
iLGCP	6,66E-04	***	5,66E-01		2,60E-02	*	3,60E-02	*	2,66E-03	**	1,41E-01	

VALIDATION

	A		B		C		D		E		F		G		H	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	4,46E-02	*	5,19E-01		**6,66E-01**		1,67E-02	*	**9,07E-01**		**4,60E-01**		**7,83E-01**		2,00E-03	**
iLGCP	6,66E-04	***	**7,57E-01**		6,66E-04	***	6,66E-04	***	6,66E-04	***	5,46E-02	.	4,13E-02	*	6,66E-04	***

	I		J		K		L		M		N		O		R	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	2,40E-02	*	1,33E-03	**	6,66E-04	***	6,66E-04	***	**1,83E-01**		6,66E-04	***	**1,13E-01**		6,66E-03	**
iLGCP	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	2,00E-03	**	6,66E-04	***	6,66E-04	***	6,66E-04	***

	S		t		U		V		W		X		Y		Z	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	1,33E-03	**	1,33E-03	**
iLGCP	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***

	AA		AB		AC		AD		AE		AG		AH		AI	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	3,33E-03	**	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	1,13E-02	*	6,66E-04	***
iLGCP	5,33E-03	**	6,66E-04	***	6,66E-04	***	6,66E-04	***	9,25E-04	***	6,66E-04	***	2,66E-03	**	2,00E-03	**

	AJ		AK		AM		AN		AO		AQ	
CSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
iCSR	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***
LGCP	6,66E-04	***	1,33E-03	**	**4,53E-01**		6,66E-04	***	6,66E-04	***	6,66E-04	***
iLGCP	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	6,66E-04	***	**2,04E-01**	

Table S 2 Extreme rank envelope test p-values per sample area of the extensive dataset in calibration and validation, for the different models: completely spatial randomness (CSR); CSR with an inhomogeneous intensity (iCSR), log-Gaussian cox-processes (LGCP), LGCP with an inhomogeneous intensity (iLGCP). Significance codes: '***' for 0.001, '**' for 0.01, '*' for 0.05 and '.' for 0.1. Grey highlighting: highest p-value; bold blue: highest and non-significate p-value.

Intensive dataset

	$\sigma 2$	α	Intercept (β_0)	Hpop (β_1)	Crop (β_2)	Tree (β_3)	Remot (β_4)	Hpop² (β_5)	Crop² (β_6)	Tree² (β_7)	Remot² (β_8)
B	3,25E+00	4,19E+03	-1,78E+01	-6,71E-01	2,85E-02	2,36E-02	4,41E-03	9,16E-02	-3,00E-04	-5,66E-04	-1,82E-05
D	1,78E+00	7,61E+03	-1,87E+01	7,83E+00	3,03E-02	1,40E-03	-3,56E-03	-6,70E+00	-4,01E-04	-1,69E-05	1,16E-06
K	2,70E+00	5,24E+03	-1,77E+01	4,57E+00	-4,16E-03	3,60E-02	-6,13E-03	-4,32E+00	1,74E-04	-1,57E-03	3,77E-06
L	1,65E+00	1,15E+04	-1,81E+01	4,05E+00	2,81E-02	2,16E-03	2,86E-03	-3,01E+00	-3,75E-04	-1,74E-04	-1,08E-05
N	2,90E+00	3,17E+03	-2,04E+01	4,87E+00	4,90E-02	5,06E-02	2,82E-04	-2,96E+00	-3,77E-04	-1,22E-03	-7,26E-07
O	2,33E+00	6,28E+03	-1,79E+01	5,62E+00	-2,90E-02	-3,54E-03	-3,20E-03	-4,61E+00	3,62E-04	-3,25E-04	1,23E-06
P	2,80E+00	7,31E+03	-1,95E+01	7,82E+00	1,61E-04	8,34E-02	-7,52E-03	-6,66E+00	-1,82E-05	-4,07E-03	6,12E-06
S	4,11E+00	3,74E+03	-1,93E+01	5,48E+00	5,46E-02	-4,55E-03	-1,60E-03	-4,51E+00	-8,01E-04	-5,45E-05	-4,75E-06
t	1,93E+00	2,07E+04	-2,26E+01	1,23E+01	2,54E-02	3,16E-02	9,53E-04	-9,14E+00	-2,24E-04	-3,89E-04	-1,52E-07
U	3,50E+00	8,94E+03	-1,92E+01	1,15E+01	-4,02E-03	-7,25E-02	3,13E-03	-1,20E+01	-9,72E-05	9,78E-04	-4,87E-05
X	2,83E+00	5,61E+03	-1,94E+01	1,13E+01	-1,33E-02	1,76E-02	-4,84E-03	-1,01E+01	2,41E-04	-2,67E-04	2,77E-06

Extensive dataset

	$\sigma 2$	α	Intercept (β_0)	Hpop (β_1)	Crop (β_2)	Tree (β_3)	Remot (β_4)	Hpop² (β_5)	Crop² (β_6)	Tree² (β_7)	Remot² (β_8)
A	8,44E-01	3,24E+03	-1,33E+01	3,40E+00	-2,06E-03	-2,54E-02	1,12E-03	-2,50E+00	1,12E-04	2,68E-04	-3,30E-06
B	7,50E-01	7,11E+03	-1,47E+01	9,56E+00	-2,03E-02	2,04E-02	-1,42E-04	-8,04E+00	3,72E-04	-2,44E-04	2,37E-07
C	6,63E-01	3,08E+03	-1,38E+01	5,84E+00	1,22E-02	-1,04E-02	1,34E-03	-3,71E+00	-9,64E-05	3,23E-05	-1,93E-06
D	9,86E-01	3,31E+03	-1,38E+01	7,70E+00	-2,32E-03	1,39E-02	-8,73E-05	-5,83E+00	7,87E-05	-1,65E-04	-3,17E-07
E	8,53E-01	1,59E+03	-1,27E+01	6,89E+00	-2,55E-05	-7,87E-03	-2,03E-05	-4,51E+00	3,74E-05	2,39E-05	-6,32E-07
F	6,48E-01	1,04E+04	-1,68E+01	9,80E+00	1,19E-02	1,88E-02	4,72E-03	-6,64E+00	-6,25E-05	-1,42E-04	-2,59E-06
G	1,02E+00	1,82E+03	-1,29E+01	5,60E+00	7,97E-03	-2,37E-02	-1,33E-03	-3,23E+00	-1,03E-04	1,89E-04	1,38E-07
H	9,66E-01	2,32E+03	-1,35E+01	6,12E+00	1,17E-02	-5,79E-03	7,73E-04	-3,19E+00	-1,77E-04	-5,23E-06	-9,98E-07
I	1,24E+00	2,38E+03	-1,46E+01	1,32E+01	1,97E-02	1,78E-02	7,73E-04	-1,56E+01	-1,92E-04	-2,17E-04	-7,50E-07
J	8,36E-01	2,54E+03	-1,36E+01	6,85E+00	-6,93E-04	-1,66E-03	-1,42E-03	-5,52E+00	6,35E-05	8,67E-05	8,66E-07
K	6,13E-01	2,01E+03	-1,26E+01	2,46E+00	-1,63E-04	-1,19E-02	-4,45E-04	-1,34E+00	2,84E-05	1,19E-04	-1,77E-06
L	7,60E-01	3,26E+03	-1,24E+01	4,43E+00	-1,03E-03	-1,07E-02	-4,45E-03	-2,85E+00	3,09E-05	7,72E-05	5,71E-06
M	1,42E+00	2,90E+03	-1,41E+01	6,71E+00	-4,96E-03	-6,01E-03	-1,14E-04	-6,36E+00	1,21E-04	-6,34E-05	-2,42E-07
N	6,92E-01	2,33E+03	-1,35E+01	4,20E+00	3,65E-03	-1,55E-02	-4,72E-04	-2,90E+00	-2,09E-05	1,81E-04	-9,92E-07
O	7,86E-01	3,90E+03	-1,37E+01	7,14E+00	-1,19E-02	-7,09E-03	3,57E-04	-4,67E+00	1,83E-04	7,44E-06	-7,16E-07
R	5,38E-01	2,33E+03	-1,31E+01	6,86E+00	1,52E-02	-5,01E-03	-1,82E-03	-5,17E+00	-1,67E-04	4,15E-05	3,49E-07
S	1,01E+00	2,90E+03	-1,53E+01	1,11E+01	-1,02E-02	3,25E-03	9,36E-04	-1,05E+01	1,46E-04	-2,09E-04	-9,11E-07
t	7,30E-01	3,95E+03	-1,39E+01	4,16E+00	9,50E-03	-2,05E-03	-2,61E-04	-3,33E+00	-5,65E-05	-2,32E-04	-7,23E-07
U	8,74E-01	2,28E+03	-1,35E+01	6,44E+00	-9,06E-03	-1,63E-02	-3,24E-04	-5,25E+00	1,17E-04	1,05E-04	-5,52E-07
V	6,20E-01	1,94E+03	-1,30E+01	5,24E+00	-2,31E-03	-2,48E-02	-6,63E-04	-2,91E+00	9,26E-05	3,03E-04	-8,49E-07
W	7,42E-05	2,11E+02	-1,26E+01	5,70E+00	8,37E-04	-1,24E-02	1,63E-03	-4,86E+00	1,69E-05	2,27E-04	-5,76E-06
X	1,02E-05	2,47E+02	-1,27E+01	4,47E+00	-2,63E-03	-2,13E-03	2,40E-03	-2,84E+00	8,64E-05	-4,48E-05	-5,78E-06
Y	4,43E-01	4,20E+03	-1,38E+01	7,07E+00	-3,82E-03	-2,95E-02	1,33E-03	-5,55E+00	1,36E-04	4,81E-04	-4,27E-06
Z	5,67E-01	1,11E+04	-1,71E+01	2,05E+01	1,85E-02	-2,57E-03	2,26E-03	-4,53E+01	-1,52E-04	-1,03E-04	-1,42E-06
AA	5,20E-01	3,12E+03	-1,34E+01	4,67E+00	-6,11E-03	9,22E-03	-1,15E-03	-3,46E+00	1,09E-04	-1,95E-04	-7,13E-08
AB	9,92E-01	2,51E+03	-1,34E+01	3,31E+00	3,08E-03	4,36E-03	1,11E-03	-1,85E+00	1,53E-05	-1,99E-04	-2,43E-06
AC	8,31E-01	9,28E+02	-1,26E+01	4,25E+00	1,14E-02	-6,59E-03	-1,75E-03	-2,50E+00	-1,13E-04	-6,98E-05	2,85E-06
AD	3,82E-06	2,30E+02	-1,25E+01	4,68E+00	1,40E-02	8,39E-04	-5,35E-04	-3,05E+00	-1,46E-04	-1,00E-04	1,23E-06
AE	1,69E-06	2,30E+02	-1,30E+01	5,04E+00	3,69E-03	-1,33E-02	1,88E-03	-3,54E+00	3,84E-05	3,97E-05	-3,61E-06
AG	5,58E-01	7,47E+03	-1,37E+01	3,43E+00	2,17E-02	-2,16E-03	-3,67E-03	-2,22E+00	-2,01E-04	-7,06E-05	4,17E-06
AH	9,21E-01	7,44E+03	-1,35E+01	1,64E+00	1,39E-02	1,06E-02	2,99E-03	-1,00E+00	-1,19E-04	-2,26E-04	-5,47E-06
AI	8,54E-01	4,27E+03	-1,33E+01	4,47E+00	-1,06E-03	-1,41E-02	-1,61E-03	-2,67E+00	1,42E-05	4,22E-05	3,69E-07
AJ	2,06E+00	1,10E+04	-1,80E+01	6,92E+00	2,40E-02	3,41E-02	5,03E-03	-3,84E+00	-1,08E-05	-3,48E-04	-3,89E-06
AK	1,25E+00	3,17E+03	-1,36E+01	2,69E+00	1,66E-02	-1,72E-02	-1,04E-03	-1,43E+00	-2,15E-04	1,71E-04	1,50E-06
AM	8,22E-01	1,99E+03	-1,30E+01	4,99E+00	1,79E-02	8,50E-04	-3,40E-03	-3,28E+00	-2,73E-04	9,29E-06	1,47E-06
AN	9,44E-01	2,79E+03	-1,35E+01	4,48E+00	6,27E-03	8,65E-03	-1,92E-03	-2,19E+00	-9,27E-05	-1,08E-04	1,60E-07
AO	5,93E-01	3,63E+03	-1,32E+01	4,67E+00	2,69E-02	-1,13E-02	-1,33E-03	-3,11E+00	-2,37E-04	3,19E-05	-4,95E-07
AQ	1,17E+00	5,66E+03	-1,29E+01	3,27E+00	1,43E-02	-1,18E-02	-5,58E-03	-2,17E+00	-1,79E-04	2,07E-04	2,94E-06

Table S 3 Coefficients of the different model parameters (α, σ^2 and $\beta_0, \beta_1 \dots \beta_8$)**.** In LGCP models, the covariance is defined as $C_0(r) = \sigma^2 \exp\,(r/\alpha)$ where σ^2 is the variance and α the scale parameter and the intensity function was defined as $\lambda(u) = \exp\big(\beta_0 + \beta_1 Hpop(u) + \beta_2 Crop(u) + \beta_3 Tree(u) + \beta_4 Remot(u) + \beta_5 Hpop^2(u) + \beta_6 Crop^2(u) + \beta_7 Tree^2(u) + \beta_8 Remot^2\big)$.

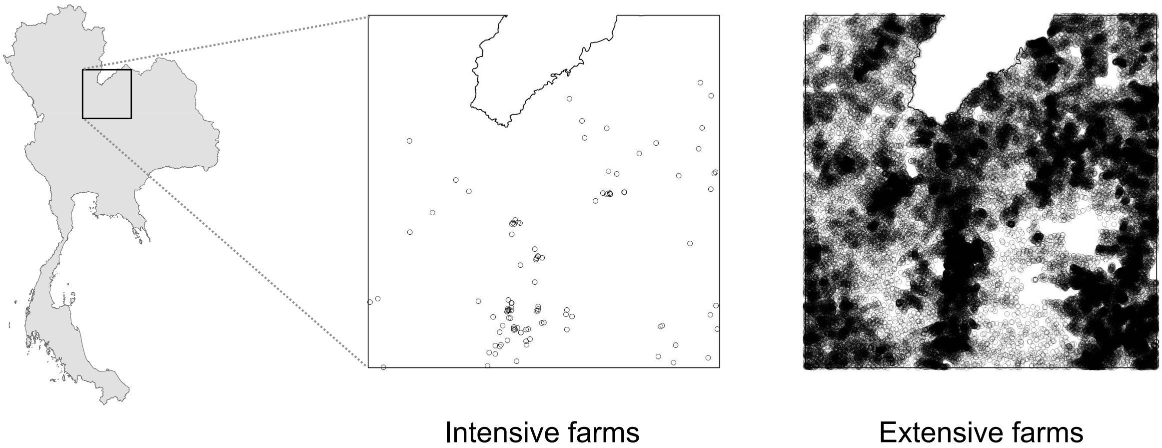


Figure S 6. Intensive and extensive observed distribution of farms in a sample area.