Timothee Vergne, Fedor Korennoy, Lisa Combelles, Andrey Gogin, Dirk U. Pfeiffe, Modelling African swine fever presence and reported abundance in the Russian Federation using national surveillance data from 2007 to 2014, *Spatial and Spatio-temporal Epidemiology* (2016), doi: 10.1016/j.sste.2016.06.002

TITLE: Modelling African swine fever presence and reported abundance in the Russian Federation using national surveillance data from 2007 to 2014

AUTHORS: **Timothee Vergne**, Fedor Korennoy, Lisa Combelles, Andrey Gogin, Dirk U. Pfeiffe

JOURNAL TITLE: Spatial and Spatio-temporal Epidemiology

PUBLISHER: Elsevier

PUBLICATION DATE: 16 June 2016 (online)

DOI: 10.1016/j.sste.2016.06.002

Modelling African swine fever presence and reported abundance in the Russian Federation using national surveillance data from 2007 to 2014

Timothée Vergne[a*], Fedor Korennoy[b], Lisa Combelles[a], Andrey Gogin[c], Dirk U. Pfeiffer[a]

[a] *Veterinary Epidemiology, Economics and Public Health group, Royal Veterinary College, London, United Kingdom*

[b] *FGBI Federal Center for Animal Health (FGBI ARRIAH), Vladimir, Russia*

[c] *National Research Institute for Veterinary Virology and Microbiology Russian Academy of Agricultural Science, Pokrov, Russia*

* Corresponding author

Timothée Vergne, tvergne@rvc.ac.uk

Abstract

African swine fever (ASF) is a viral disease of swine that has been present in the Russian Federation since 2007. Counts of ASF outbreaks reported in the Southern regions of the country (2007-2014) were aggregated to a grid of hexagons, and a zero-inflated Poisson model accounting for spatial dependence between hexagons was used to identify factors associated with the presence of ASF outbreaks and factors associated with the number of ASF reports in affected hexagons. Increasing density of pigs raised on low biosecurity farms was found to be positively associated with the probability of occurrence of at least one ASF outbreak in a hexagon and with the average number of reported ASF outbreaks amongst affected hexagons. Increasing human population density and increasing distance from the closest diagnostic laboratory were additional variables associated with number of reported ASF outbreaks amongst affected hexagons. The model was shown to have good predictive ability.

Keywords:

African swine fever, zero-inflated Poisson, spatial, Russian Federation, risk, mapping

INTRODUCTION

African swine fever (ASF) is a highly contagious viral disease that affects all classes of domestic and non-domestic swine including domestic pigs, wild boar, warthogs and bush pigs (Costard et al., 2012). In domestic pigs and wild boar, some ASF virus (ASFv) strains result in mortality rates of almost 100% (Blome et al., 2012; Sanchez-Vizcaino et al., 2012; Guinat et al., 2014). There is currently no vaccine available. Because of its large economic impact in affected countries, ASF is listed as a notifiable disease by the World Organization for Animal Health (OIE).

In 2007, ASFv was introduced into Georgia from Africa, probably through the transport of infected swill fed to pigs in the region of the sea port Poti. In just a few months, ASFv spread across the country and entered the Russian Federation (RF) in late 2007, probably as a result of the movement of infected wild boar (FAO, 2013; Sanchez-Vizcaino et al., 2013). Initially restricted to the southern regions of the RF, ASFv expanded its geographical distribution in 2011 when it was introduced into central and northern regions where it started to spread locally (Sanchez-Vizcaino et al., 2013). In 2014, ASFv was identified in both domestic pigs and wild boars in several countries of the eastern European Union, including Estonia, Latvia, Lithuania and Poland. Due to a relatively high transmission rate combined with a case fatality rate of almost 100% in naïve pigs and wild boar (Blome et al., 2012; Guinat et al., 2015), the fact that the virus is able to survive for lengthy periods in both the environment and pork products (EFSA, 2014; Davies et al., 2015) and the absence of both vaccine and treatment, understanding the epidemiology of ASF and identifying risk factors for disease occurrence is paramount to allow for the timely prevention and control of ASFv.

On the African continent, ASF was shown to be associated with many different risk factors, including density of free-ranging pigs, movement of pigs and pig products, low biosecurity measures, the practice of swill-feeding, purchase of pigs, proximity of slaughter houses and human behaviour (Randriamparany et al., 2005; Penrith and Vosloo, 2009; Fasina et al., 2012; Chenais et al., 2015; Nantima et al., 2015). In the RF, two studies have used national surveillance data to investigate risk factors for ASF occurrence in the RF. In their early paper, Gulenkin et al. (2011) used linear regression to model the density of outbreaks reported in domestic pigs in the Southern regions between 2007 and 2010, and showed that risk of ASF outbreaks increased with increasing density of domestic pigs, rivers, main highways and secondary roads. Korennoy et al. (2014) used the maximum entropy method to model the distribution of ASF outbreaks reported between 2007 and 2012 and found that increasing disease risk was associated mostly with increasing rural population density and increasing density of pigs raised on low-biosecurity farms.

While these two studies provide important insights into the risk of ASF, their main limitation is that they do not account for a potential spatially heterogeneous reporting rate of ASF outbreaks, which has been suggested to be one of the most important challenges of ASF control in the Russian Federation as a consequence of inefficient control measures and inadequate compensation strategies (FAO, 2013). As in ecological studies focused on the modelling of site occupancy by cryptic wildlife populations (MacKenzie et al., 2002), the observed spatial distribution of a disease is a combination of its true distribution and of a reporting bias. The methods used in these two studies cannot explicitly differentiate between variables associated with the true ASF distribution and variables that potentially influence the reporting rate or local abundance of outbreaks. As a result, some of the retained explanatory variables could be associated with an increased reporting rate rather than an increased ASF risk, and the measure of the identified associations with the ASF risk could be confounded by variables influencing the reporting rate.

Extensively used in ecology to model the distribution of wildlife species that cannot be observed with certainty (Martin et al., 2005), zero-inflated count models could be a useful approach in epidemiology to model the distribution of an infectious disease observed through the lens of an imperfect surveillance system, i.e. when the probability of reporting outbreaks is less than one (Vergne et al., 2015b). These models allow one to distinguish between variables influencing the likelihood of occurrence of at least one outbreak and those influencing the number of reported outbreaks in affected areas. In animal health, this approach has been used to model bovine abortions in France (Bronner et al., 2013) and highly pathogenic avian influenza in Vietnam (Lockhart, 2008) and Thailand (Vergne et al., 2014).

In this study, we used a conditional autoregressive zero-inflated Poisson regression approach to model the distribution of ASF outbreaks reported in the Southern regions of the RF between 2007 and 2014, in order to identify variables associated with the distribution of reported outbreaks.

MATERIALS AND METHODS

The analysis was restricted to the territories of the North Caucasus and of the South of the European region of the RF as this zone has one of the highest pig densities in the country and was most affected since the start of the epidemic. The surveillance system is based on the reporting of suspicions by farmers to regional veterinary services. These suspicions are subsequently confirmed using laboratory diagnostic techniques by the State Research Institute of Veterinary Virology and Microbiology which maintains the national dataset of ASF outbreaks. This dataset includes several characteristics of the infected farms as well as the location of the outbreaks recorded as the longitude and latitude of the centroid of the town to which the infected farms belong (Korennoy et al., 2014). All outbreaks reported to the veterinary authorities between 2007 and 2014, involving domestic pigs from this region, were included in this study. The data were sourced from the official website of the Russian federal service for veterinary and phytosanitary surveillance (http://fsvps.ru/fsvps/asf; last update: August 2014). Overall, 104 outbreaks in domestic pigs were reported during the period of interest. The study area was partitioned into a regular grid of 389 hexagons of 60-kilometer diameter (hexagons are more similar to circles than squares which allows for a more efficient aggregation of data around their centroids), and counts of ASF outbreaks were aggregated to the hexagon level. The outcome variable was therefore the number of outbreaks reported in the domestic pig population in each hexagon during the study period. Figure 1A presents the spatial distribution of the outcome variable.

The putative explanatory variables used in this analysis focused on anthropogenic, swine-related and environmental variables. The anthropogenic variables included human population density, distance to the nearest regional capital, distance to the nearest diagnostic laboratory, and road density. Human population data were obtained for the year 2013 from the Federal State Statistics Services. The location of regional capitals and information about the road network were extracted from the database distributed by ESRI-CIS (http://www.esri-cis.ru/). The location of regional veterinary diagnostic laboratories was obtained from the Federal Service for Veterinary and Phytosanitary Surveillance. Similar to Korennoy et al. (2014), distribution of pigs kept in high and low biosecurity farms was based on estimates generated by the Federal State Statistics Service in 2011. High biosecurity farms were defined as large holdings, often comprising of more than a thousand pigs, typically owned by the State or private companies, regularly visited by a veterinarian (they often have their own) and associated with several biosecurity measures, including access restrictions and disinfection procedures; low biosecurity farms were defined as backyard or small-scale holdings characterised by a small number of pigs, very low veterinary support and low level or absence of biosecurity measures (Oganesyan et al., 2013). As data on wild boar density was not readily available, forest coverage (also extracted from the database distributed by ESRI-CIS) was used as a proxy variable (Boitani et al., 1994; Fernandez et al., 2006). All these variables had either been identified as risk factors in previous studies (Gulenkin et al., 2011; Korennoy et al., 2014) or were suspected to be associated with the reporting rate. All variables were summarised at hexagon level either by calculating the Euclidian distance between the centroid of the hexagons and any location of interest (i.e. nearest regional capital and nearest diagnostic laboratory) or by averaging the value of the continuous variables for each hexagon using the ArcGIS 10.2 software (ESRI, 2011). To be able to accommodate non-linear associations, all variables were categorised into three categories based on their 33rd and the 66th percentiles, except for the density of high biosecurity farms which was included as a binary variable due to the small number of hexagons with more than one high biosecurity farm.

A conditional autoregressive zero-inflated Poisson (CAR-ZIP) regression model was developed to quantify the association between the putative explanatory variables and the spatial distribution of the number of outbreaks reported in domestic pigs. Zero-inflated Poisson models (ZIP) assume that the outcome variable Y (here, the number of outbreaks reported in a hexagon) follows a mixture of a zero-point mass distribution and a Poisson distribution, and can be expressed using the following probability function:

$$P(Y_i = y) = \begin{cases} (1-\Phi_i) + \Phi_i \exp(-\lambda_i) & if\ y = 0 \\ \Phi_i \frac{\lambda_i{}^y \exp(-\lambda_i)}{y!} & if\ y \geq 1 \end{cases} \qquad (1)$$

with Φ_i being the parameter of the mass distribution and λ_i the parameter of the Poisson distribution (Lambert, 1992; Cameron and Trivedi, 1998). In the context of this study, Φ_i and λ_i can be seen as the probability that hexagon *i* experienced at least one ASF outbreak in the domestic pig population (i.e. the probability that ASF virus circulated in domestic pigs in the hexagon *i*) and as the average number of reported outbreaks in hexagon *i* had it been affected, respectively. Both Φ_i and λ_i can be expressed as a function of explanatory variables using a logit and a log transformation, respectively:

$$logit(\Phi_i) = \alpha_0 + \sum \alpha x_{1i} \quad (2)$$

and

$$\log(\lambda_i) = \beta_0 + \sum \beta x_{2i} \quad (3)$$

with α_0 and β_0 being two intercepts, x_{1i} and x_{2i} being vectors of explanatory variables (they can be two different sets of covariates) and α and β being the vectors of their associated coefficients.

As ASFv is highly contagious, it is likely that the probability Φ_i that hexagon *i* experienced at least one outbreak of ASF was dependent on the probability that its neighbouring hexagons experienced at least one outbreak. This spatial autocorrelation was incorporated in the models by extending the logistic expression of Φ_i as follows:

$$logit(\phi_i) = \alpha_0 + \sum \alpha x_{1i} + u_i \quad (4)$$

with u_i being the spatially structured random effect. This was assumed to have a conditional autoregressive structure using the first-order spatial interaction neighbourhood based on contiguity between the hexagons (Besag et al., 1991). This type of CAR-ZIP regression has been used in various applications in ecology (Agarwal et al., 2002; Rathbun and Fei, 2006; Flores et al., 2009) and epidemiology (Musal and Aktekin, 2013; Vergne et al., 2014).

First, the collinearity between each pair of variables was assessed using the Kendall rank correlation coefficient (Abdi, 2006). Pair-wise collinearity was considered significant if the absolute value of the coefficient exceeded 0.7. In case of collinearity, the set of variables to be removed from the analysis was the smallest set of variables leading to the elimination of all significant pair-wise collinearity. Univariable analyses were then conducted and variables associated with at least one coefficient for which zero was not included in the 80% credible interval of its posterior distribution were retained for multivariable analyses. Next, all selected variables were added one by one following a stepwise forward selection procedure using the Deviance Information Criterion (DIC) as the selection criterion (Ntzoufras, 2009). The DIC is based on a trade-off between the fit of the data to the model and the complexity of the model. It is generally accepted that models with smaller DIC are better supported by the data. The best model was considered to be the most parsimonious model whose DIC was less than two points greater than that of the model associated with the smallest DIC (Spiegelhalter et al., 2002). Retained variables associated with the parameter Φ were considered as risk factors for the occurrence of at least one outbreak in hexagons. Retained variables associated with the average number of reported outbreaks in an affected hexagon (λ) may either be associated with the true number of ASF outbreaks in an affected hexagon (and therefore with the local spread of the virus) or with the probability that an ASF outbreak is reported (and therefore with the sensitivity of the reporting). Note that such a model does not allow for differentiating explicitly between the two.

All analyses were performed in a Bayesian framework using the WinBUGS software (Lunn et al., 2000) embedded in the R software (R-Development-Core-Team, 2008) using the R2WinBUGS package (Sturtz et al., 2005). As we did not have any information regarding the value of the variable coefficients, we assumed normal prior distributions of mean 0 and variance 10 for each of the fixed effects included in the model. For the variance of the spatially structured random effect, we used a gamma prior distribution (which has density only for positive real numbers) of mean 5 and variance 5. In the exploratory phase, gamma priors of mean 5 and variance 10 and 20 were also tested for this latter parameter but they led to unstable models unable to converge properly. Two simulation chains of 100,000 iterations were run, with the first 5,000 iterations discarded to allow for burn-in of the chain. The chains were then thinned, taking every hundredth sample to reduce autocorrelation amongst the samples. Convergence was assessed by checking the trace plots for all monitored parameters (Gelman et al., 2004). The posterior distribution of each parameter was summarised using the median and the 95% credible interval (95%CI).

Choropleth maps were produced to show the spatial distributions of the median predicted probability of occurrence of at least one outbreak (Φ_i), of the median sensitivity of detection at the hexagon level (i.e. the median predicted probability that at least one outbreak would have been reported had the hexagons been affected: $1 - \exp(-\lambda_i)$), and of the median probability that at least one outbreak was reported ($\Phi_i[1 - \exp(-\lambda_i)]$).

The ability of the model to correctly predict presence or absence of at least one reported outbreak at the hexagon level was assessed by using the posterior distribution of the probability that at least one outbreak was reported in each hexagon ($\Phi_i[1 - \exp(-\lambda_i)]$) to estimate the area under the curve (AUC) of the receiver operating characteristic (ROC) plots.

RESULTS

No correlation was detected between any pair of investigated variables. The best-fit model and the estimated effects of the associated variables are presented in Table 1.

The density of pigs kept in low biosecurity farms was found to be the only explanatory variable associated with the risk of occurrence of at least one ASF outbreak in a hexagon. Compared with hexagons with a low density of pigs kept on low biosecurity farms, the odds of ASF outbreak occurrence was increased by a factor of 7.37 (95%CI: 1.93; 30.05) and 5.53 (95%CI: 1.15; 26.92) for hexagons with medium (between 0.44 and 2.37 head/km2) and high (greater than 2.37 head/km2) densities of pigs kept on low biosecurity farms, respectively. The variance of the spatially structured random effect, which can be seen as the amount of variation in logit(Φ) due to spatial proximity between neighbouring hexagons, was estimated to be 2.56 (95%CI: 0.78; 9.14).

The average number of reported outbreaks in ASF-affected hexagons was positively associated with medium and high human population densities (incidence rate ratios (IRR) = 3.57 (95%CI: 2.03; 6.45) and 4.83 (95%CI: 2.78; 8.60), respectively), with a medium density of pigs kept on low biosecurity farms (IRR = 2.45 (95%CI: 1.41; 4.59)) and with medium and long distances to the closest diagnostic laboratory (IRR = 1.90 (95%CI: 1.30; 2.75) and 2.73 (95%CI: 1.87; 3.97), respectively). Density of pigs kept on high-biosecurity farms, forest coverage, road density and distance to the closest regional capital were neither found to be associated with the risk of occurrence of at least one ASF outbreak in a hexagon nor with the average number of reported outbreaks amongst ASF affected hexagons. As shown in Figure 2, the discriminatory power of the model to predict whether at least one outbreak was reported was relatively good with an AUC value of 0.86 (95%CI: 0.84; 0.91).

Using the spatial distribution of the variables retained in the model, the predicted probability of occurrence of ASF in domestic pigs, the sensitivity of ASF detection at hexagon level and the probability that at least one outbreak was reported were mapped (Figure 1B, 1C and 1D, respectively).

Table 1: Results of the conditional autoregressive zero-inflated Poisson regression

	Variable	Category	Number of hexagons	Posterior mean (standard deviation) of the regression coefficients	Monte Carlo error	Odds ratio (95% credible interval) related to the probability of occurrence of at least one outbreak	Incidence rate ratio (95% credible interval) related to the average number of outbreaks reported in affected hexagons
Probability of occurrence of at least one ASF outbreak in hexagons							
	Density of pigs kept on low biosecurity farms	[0; 0.44[	129	*Ref*	*Ref*	*Ref*	-
		[0.44; 2.37[	128	2.01 (0.70)	0.010	7.37 (1.93; 30.05)	-
		≥ 2.37	132	1.71 (0.80)	0.011	5.53 (1.15; 26.92)	-
Average number of reported outbreaks amongst ASF affected hexagons							
	Human population density	[0; 18.2[	129	*Ref*	*Ref*	-	*Ref*
		[18.2; 40.8[	128	1.28 (0.30)	0.006	-	3.57 (2.03; 6.45)
		≥ 40.8	132	1.58 (0.29)	0.006	-	4.83 (2.78; 8.60)
	Density of pigs kept on low biosecurity farms	[0; 0.44[	129	*Ref*	*Ref*	-	*Ref*
		[0.44; 2.37[	128	0.91 (0.30)	0.007	-	2.45 (1.41; 4.59)
		≥ 2.37	132	0.51 (0.31)	0.006	-	1.65 (0.94; 3.14)
	Distance to the nearest diagnostic laboratory	[0; 177[	129	*Ref*	*Ref*	-	*Ref*
		[177; 302[	128	0.64 (0.19)	0.002	-	1.90 (1.30; 2.75)
		≥ 302	132	1.00 (0.19)	0.002	-	2.73 (1.87; 3.97)

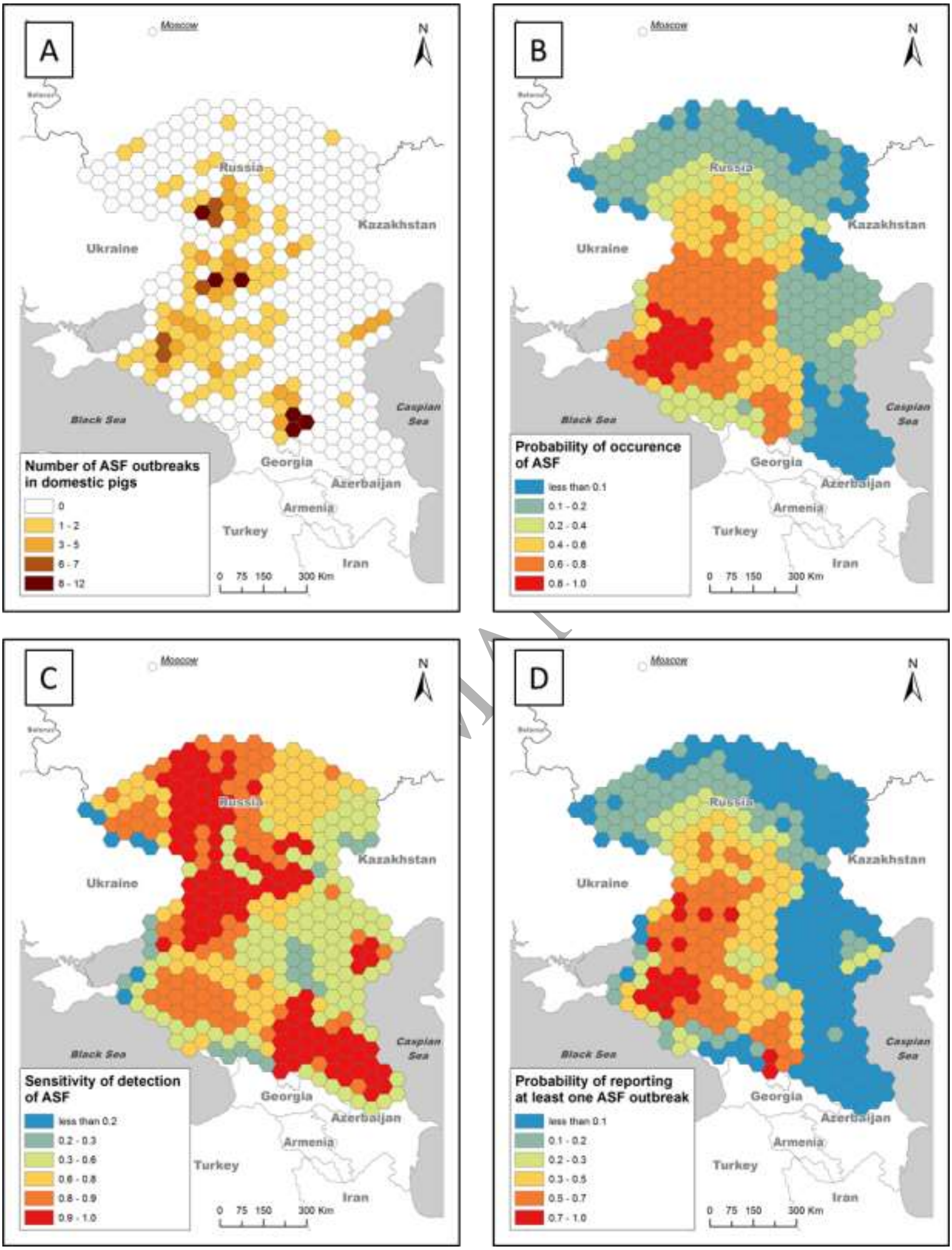

Figure 1: Spatial distribution of the number of African swine fever (ASF) outbreaks reported in the southern regions of the Russian Federation between 2007 and 2014 (A), of the median predicted probability of occurrence of at least one ASF outbreak (B), of the median predicted probability that at least one outbreak would have been

reported had the hexagons been affected (sensitivity of the detection at hexagon level), (C) and of the probability of reporting at least one ASF outbreak (D) at hexagon level between 2007 and 2014. Note the cut-points are different between panels.

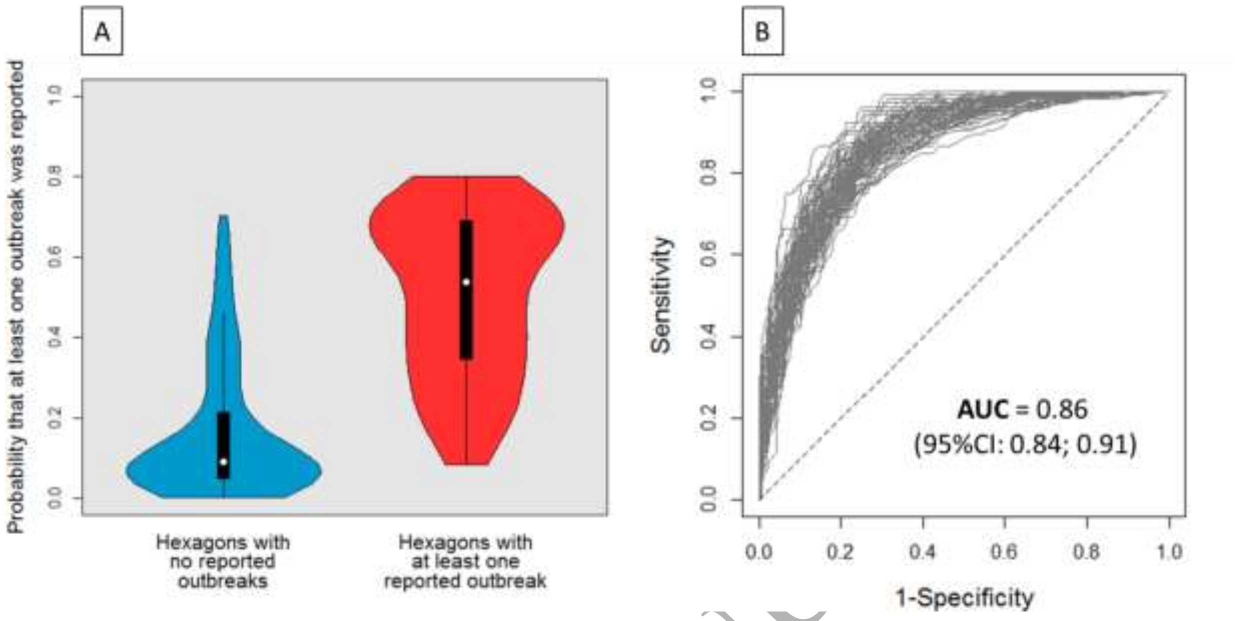

Figure 2: Discriminatory power of the zero-inflated Poisson model: violin plots showing the predicted probability that at least one ASF outbreak was reported for hexagons with (red) and without (blue) actually reported outbreaks (A); receiver operating characteristic curves quantifying the discriminatory power of the model to predict whether at least one outbreak of ASF was reported (B).

DISCUSSION

At the time this research was conducted, i.e. eight years after ASFv had started circulating in the Russian Federation causing significant economic losses to the national pig industry, very few studies have tried to identify factors associated with the spatial pattern of the epidemic (Gulenkin et al., 2011; Korennoy et al., 2014). The present paper contributes to filling this gap by analysing the distribution of the outbreaks reported from 2007 to 2014 in South Russia using spatial zero-inflated Poisson (ZIP) regression. The main findings of this study are threefold: 1) the risk of presence of ASF outbreaks was strongly associated with the density of pigs kept on low biosecurity farms (Table 1); 2) the human population density and, to a lesser extent, the density of pigs kept on low biosecurity farms and the distance to the nearest laboratory, were shown to increase the average number of ASF reports in ASF affected hexagons (Table 1); 3) the ability of the car-ZIP model to predict whether at least one ASF outbreak was reported was found to be good with an AUC larger than 0.8 (Figure 2).

Consistent with Korennoy et al. (2014), the density of pigs kept on low biosecurity farms was found to be strongly associated with the distribution of ASF outbreak reports in the Russian Federation. Backyard and free-range pigs played a critical role in the transmission of ASFv in the RF as they were associated with an increase in both the probability of occurrence of at least one outbreak in hexagons and the average number of reported outbreaks in affected hexagons. Indeed, at the time of the epidemic, extensive outdoor breeding was very common in the South of the Russian Federation (Gogin et al., 2013), and complicated the control of the epidemic as this practice was poorly regulated and the location of the farms unknown. In addition, initially delayed intervention strategies and inappropriate financial compensation schemes encouraged some affected small-scale farmers to dispose of dead or affected pigs inappropriately or slaughter and sell illegally their apparently healthy pigs (FAO, 2013). This contributed significantly to the release of the virus (Costard et al., 2015) and consequently to the observed spatial pattern of the epidemic.

The results of the CAR-ZIP regression show that human population density was positively associated with the average number of reported outbreaks in ASF affected hexagons (meaning that if at least one outbreak was to occur in a highly populated area, there would be more reported outbreaks than in an area of low human density). This suggests that highly populated areas either increased the risk of local spread of ASFv, potentially through movement of pigs or people entering farms, therefore increasing the actual number of outbreaks, or resulted in pig farmers being more likely to report ASF outbreaks because of enhanced awareness, better compliance with the rules or better trust in the state veterinary services. These results provide useful complimentary information to that produced by Korennoy et al. (2014) who showed that the density of rural human population was associated with an increase in the general risk of ASF occurrence. Indeed, the structure of the zero-inflated model allows more detailed inference by indicating that human population density is associated with an increase in the average number of ASF reports if ASFv was circulating in the area (i.e. with an increase in the probability that at least one outbreak would be reported in affected areas), but not with an increase in the risk that ASFv circulates in the area.

We hypothesized that the average number of reported outbreaks in affected hexagons would decrease with increasing distance to the diagnostic laboratory because submitting a biological sample in remote areas was expected to be more difficult, making the probability of reporting suspicions in affected hexagons less likely. In fact, model outputs showed that this variable had an effect in the opposite direction: the further the affected hexagons from diagnostic laboratories, the larger the average number of reported ASF outbreaks. This result may not be as counterintuitive as it might seem: one possible explanation for these findings is that intervention measures may be implemented in a more effective and timely manner close to laboratories compared with those in remote areas, thereby curbing the epidemic at an earlier stage and therefore limiting the total number of outbreaks in affected hexagons.

From 2007 to 2014, disease awareness amongst farmers and veterinarians, as well as surveillance and intervention effectiveness, varied markedly. At the beginning of the epidemic, most efforts were focused on intervention due to the overwhelming need for an emergency response, with a focus on culling and depopulation. In the early stages of the epidemic, efforts for improving the early reporting of suspected cases were neglected, leading to substantial periods during which ASFv was able to spread without veterinary services being notified (FAO, 2013; Gogin et al., 2013). Over the following years, timeliness of reporting improved and interventions became more effective as veterinary services became more experienced with handling the situation, farmers became more aware of the disease and its consequences, and cooperation between local and federal authorities improved (Gogin et al., 2013). As a consequence, the drivers of the spatial distribution of the reported ASF outbreaks were likely to vary both temporally and spatially. Because the current analysis used aggregated data collected over almost eight years in a relatively large region (corresponding to almost twice the size of France), the approach was only able to identify general trends and may not have been able to detect more local (in time or in space) associations. Despite this limitation, it is worth noting that the model generates good predictions for whether at least one outbreak was reported in a hexagon (Figure 2).

Potential non-linear relationships between explanatory variables and the outcome were accounted for by categorising each of the explanatory variables included in the model. It is acknowledged that other modelling methods such as the boosted-regression tree approach would have allowed a more explicit investigation of non-linear effects (Martin et al., 2011; Fang et al., 2013; Dhingra et al., 2014) but it would not have allowed characterising the role of the risk factors in the same way as the zero-inflated approach did. The usefulness of the zero-inflated Poisson approach for gaining more detailed epidemiological insights from the distribution of reported outbreaks could be further enhanced by replacing the Poisson process with a hierarchical process combining a zero-truncated count distribution and a binomial distribution to model the true number of outbreaks in affected hexagons and the number of reported outbreaks given the true number of outbreaks, respectively. This development would allow estimating the prevalence of outbreaks in affected hexagons and the probability of reporting an outbreak. However, because of the complex distribution of the latent state (true number of outbreaks in affected hexagons), it is expected that such models will be challenging to fit and will need to be parameterized using input data that truly reflects the field situation.

Although the statistical model for the probability that at least one outbreak was reported in a hexagon had a good fit to the data (Figure 2), its performance and interpretation would have benefited from being able to use direct measurements of risk factors rather than proxy variables. For example, although wild boar are considered not to have played a major role in the transmission of ASFv to domestic pigs (Iglesias et al., 2015; Vergne et al., 2015a), the use of wild boar density instead of the forest coverage would have allowed a more precise examination of wild boar’s role in the epidemiology of ASF.

CONCLUSION

This study assessed the spatial risk of ASF while accounting for a potential heterogeneous reporting rate of ASF outbreaks. The observed pattern of the most important epidemic of ASF that occurred in the RF seems to have been influenced by several risk factors, such as the density of pigs kept in low biosecurity farms, human population density and the distance to the nearest diagnostic laboratory. Importantly, these results could assist the

Russian veterinary services in defining high risk areas to be targeted for improved early detection and early reporting in order to minimise the impact of ASF on pig farmers' livelihoods.

COMPETING INTERESTS

We have no competing interests

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 311931 (ASFORCE). We are very grateful to Dr Martha Betson and Dr Kim Stevens for improving the readability of this paper and to Dr Chang Yu-Mei for the very useful discussions we had on the zero-inflated methodology.

REFERENCES

Abdi, H., 2006. The Kendall Rank Correlation Coefficient. In: Salkind, N.J. (Ed.) Encyclopedia of Measurment and Statistics Sage Publication Inc.

Agarwal, D.K., Gelfand, A.E., Citron-Pousty, S., 2002. Zero-inflated models with application to spatial count data. Environ. Ecol. Stat. 9, 341-355.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restauration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 1-59.

Blome, S., Gabriel, C., Dietze, K., Breithaupt, A., Beer, M., 2012. High virulence of African swine fever virus caucasus isolate in European wild boars of all ages. Emerging infectious diseases 18, 708.

Boitani, L., Mattei, L., Nonis, D., Corsi, F., 1994. Spatial and activity patterns of wild boars in Tuscany, Italy. Journal of Mammalogy 75, 600-612.

Bronner, A., Henaux, V., Vergne, T., Vinard, J.L., Morignat, E., Hendrikx, P., Calavas, D., Gay, E., 2013. Assessing the mandatory bovine abortion notification system in France using unilist capture-recapture approach. PLoS One 8, e63246.

Cameron, A., Trivedi, P., 1998. Regression analysis of count data. Economtric Society Monographs N°30. Cambridge University Press.

Chenais, E., Boqvist, S., Sternberg-Lewerin, S., Emanuelson, U., Ouma, E., Dione, M., Aliro, T., Crafoord, F., Masembe, C., Stahl, K., 2015. Knowledge, Attitudes and Practices Related to African Swine Fever Within Smallholder Pig Production in Northern Uganda. Transboundary and emerging diseases.

Costard, S., Mur, L., Lubroth, J., Sanchez-Vizcaino, J.M., Pfeiffer, D.U., 2012. Epidemiology of African swine fever virus. Virus research.

Costard, S., Zagmutt, F.J., Porphyre, T., Pfeiffer, D.U., 2015. Small-scale pig farmers' behavior, silent release of African swine fever virus and consequences for disease spread. Sci Rep 5, 17074.

Davies, K., Goatley, L.C., Guinat, C., Netherton, C.L., Gubbins, S., Dixon, L.K., Reis, A.L., 2015. Survival of African Swine Fever Virus in Excretions from Pigs Experimentally Infected with the Georgia 2007/1 Isolate. Transboundary and emerging diseases.

Dhingra, M.S., Dissanayake, R., Negi, A.B., Oberoi, M., Castellan, D., Thrusfield, M., Linard, C., Gilbert, M., 2014. Spatio-temporal epidemiology of highly pathogenic avian influenza (subtype H5N1) in poultry in eastern India. Spatial and spatio-temporal epidemiology 11, 45-57.

EFSA, 2014. Scientific opinion on African swine fever. EFSA Journal 12, 1-77.

ESRI, 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Fang, L.Q., Li, X.L., Liu, K., Li, Y.J., Yao, H.W., Liang, S., Yang, Y., Feng, Z.J., Gray, G.C., Cao, W.C., 2013. Mapping spread and risk of avian influenza A (H7N9) in China. Sci Rep 3, 2722.

FAO, 2013. African Swine Fever in the Russian Federation: Risk factors for Europe and beyond. Empress Watch 28.

Fasina, F.O., Agbaje, M., Ajani, F.L., Talabi, O.A., Lazarus, D.D., Gallardo, C., Thompson, P.N., Bastos, A.D., 2012. Risk factors for farm-level African swine fever infection in major pig-producing areas in Nigeria, 1997-2011. Preventive veterinary medicine 107, 65-75.

Fernandez, N., Kramer-Schadt, S., Thulke, H.H., 2006. Viability and risk assessment in species restoration: Planning reintroductions for the wild boar, a potential disease reservoir. Ecol Soc 11.

Flores, O., Rossi, V., Mortier, F., 2009. Autocorrelation offsets zero-inflation in models of tropical saplings density. Ecological modelling 220, 1797-1809.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis, second edition. In: Chatfield, C., Tanner, M., Zideck, J. (Eds.) Boca Raton: CRC Press.

Gogin, A., Gerasimov, V., Malogolovkin, A., Kolbasov, D., 2013. African swine fever in the North Caucasus region and the Russian Federation in years 2007-2012. Virus research 173, 198-203.

Guinat, C., Gubbins, S., Vergne, T., Gonzales, J.L., Dixon, L., Pfeiffer, D.U., 2015. Experimental pig-to-pig transmission dynamics for African swine fever virus, Georgia 2007/1 strain. Epidemiology and infection doi:10.1017/S0950268815000862.

Guinat, C., Reis, A., Netherton, C.L., Goatley, L., Pfeiffer, D.U., Dixon, L., 2014. Dynamics of African swine fever virus shedding and excretion in domestic pigs infected by intramuscular inoculation and contact transmission. Veterinary research 45, 93.

Gulenkin, V.M., Korennoy, F.I., Karaulov, A.K., Dudnikov, S.A., 2011. Cartographical analysis of African swine fever outbreaks in the territory of the Russian Federation and computer modeling of the basic reproduction ratio. Preventive veterinary medicine 102, 167-174.

Iglesias, I., Munoz, M.J., Montes, F., Perez, A., Gogin, A., Kolbasov, D., de la Torre, A., 2015. Reproductive Ratio for the Local Spread of African Swine Fever in Wild Boars in the Russian Federation. Transboundary and emerging diseases.

Korennoy, F.I., Gulenkin, V.M., Malone, J.B., Mores, C.N., Dudnikov, S.A., Stevenson, M.A., 2014. Spatio-temporal modeling of the African swine fever epidemic in the Russian Federation, 2007-2012. Spatial and spatio-temporal epidemiology 11, 135-141.

Lambert, D., 1992. Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics 34, 1-14.

Lockhart, C.Y., 2008. Surveillance for diseases of poultry with specific reference to avian influenza. PhD Thesis at Massey University, New Zealand.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput 10, 325-337.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248-2255.

Martin, G.T., Wintle, A.B., Rhodes, J.R., Field, A.S., Low-Choy, J.S., Tyre, A.J., Possingham, P.H., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology letters 8, 1235-1246.

Martin, V., Pfeiffer, D.U., Zhou, X., Xiao, X., Prosser, D.J., Guo, F., Gilbert, M., 2011. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. PLoS Pathog 7, e1001308.

Musal, M., Aktekin, T., 2013. Bayesian spatial modeling of HIV mortality via zero-inflated Poisson models. Statistics in medicine 32, 267-281.

Nantima, N., Ocaido, M., Ouma, E., Davies, J., Dione, M., Okoth, E., Mugisha, A., Bishop, R., 2015. Risk factors associated with occurrence of African swine fever outbreaks in smallholder pig farms in four districts along the Uganda-Kenya border. Tropical animal health and production.

Ntzoufras, I., 2009. Bayesian Modeling using WinBUGS. New York: John Wiley and Sons.

Oganesyan, A.S., Petrova, O.N., Korennoy, F.I., Bardina, N.S., Gogin, A.E., Dudnikov, S.A., 2013. African swine fever in the Russian Federation: spatio-temporal analysis and epidemiological overview. Virus research 173, 204-211.

Penrith, M.L., Vosloo, W., 2009. Review of African swine fever: transmission, spread and control. Journal of the South African Veterinary Association 80, 58-62.

R-Development-Core-Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org.

Randriamparany, T., Grenier, A., Tourette, I., Maharavo Rahantamalala, C.Y., Rousset, D., Lancelot, R., 2005. Epidemiological situation of African swine fever in Lake Alaotra region (Madagascar) and possible consequences on the organization of disease control and surveillance. Revue d'elevage et de medecine veterinaire des pays tropicaux 58, 15-20.

Rathbun, S.L., Fei, S.L., 2006. A spatial zero-inflated poisson regression model for oak regeneration. Environ Ecol Stat 13, 409-426.

Sanchez-Vizcaino, J.M., Mur, L., Martinez-Lopez, B., 2012. African swine fever (ASF): Five years around Europe. Veterinary microbiology.

Sanchez-Vizcaino, J.M., Mur, L., Martinez-Lopez, B., 2013. African swine fever (ASF): Five years around Europe. Veterinary microbiology 165, 45-50.

Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Statist. Soc. B 64, 583-639.

Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: A package for running WinBUGS from R. Journal of statistical software 12, 1-16.

Vergne, T., Gogin, A., Pfeiffer, D.U., 2015a. Statistical Exploration of Local Transmission Routes for African Swine Fever in Pigs in the Russian Federation, 2007-2014. Transboundary and emerging diseases DOI: 10.1111/tbed.12391.

Vergne, T., Paul, M.C., Chaengprachak, W., Durand, B., Gilbert, M., Dufour, B., Roger, F., Kasemsuwan, S., Grosbois, V., 2014. Zero-inflated models for identifying disease risk factors when case detection is imperfect: application to highly pathogenic avian influenza H5N1 in Thailand. Preventive veterinary medicine 114, 28-36.

Vergne, T., Vilas, V.J.D., Cameron, A., Dufour, B., Grosbois, V., 2015b. Capture-recapture approaches and the surveillance of livestock diseases: A review. Preventive veterinary medicine 120, 253-264.