

# Evaluation of the spatial patterns and risk factors, including backyard pigs, for classical swine fever occurrence in Bulgaria using a Bayesian model

Beatriz Martínez-López<sup>1,2</sup>, Tsviatko Alexandrov<sup>3</sup>, Lina Mur<sup>2</sup>, Fernando Sánchez-Vizcaíno<sup>2</sup>, José M. Sánchez-Vizcaíno<sup>2</sup>

<sup>1</sup>Center for Animal Disease Modeling and Surveillance, VM: Medicine and Epidemiology, UC Davis, Davis, USA; <sup>2</sup>VISAVET and Animal Health Department, Veterinary School, University Complutense of Madrid, Madrid, Spain; <sup>3</sup>Bulgarian Food Safety Agency, Sofia, Bulgaria

**Abstract.** The spatial pattern and epidemiology of backyard pig farming and other low bio-security pig production systems and their role in the occurrence of classical swine fever (CSF) is described and evaluated. A spatial Bayesian model was used to explore the risk factors, including human demographics, socioeconomic and environmental factors. The analyses were performed for Bulgaria, which has a large number of backyard farms (96% of all pig farms in the country are classified as backyard farms), and it is one of the countries for which both backyard pig and farm counts were available. Results reveal that the high-risk areas are typically concentrated in areas with small family farms, high numbers of outgoing pig shipments and low levels of personal consumption (i.e. economically deprived areas). Identification of risk factors and high-risk areas for CSF will allow to targeting risk-based surveillance strategies leading to prevention, control and, ultimately, elimination of the disease in Bulgaria and other countries with similar socio-epidemiological conditions.

**Keywords:** backyard pig farming, low biosecurity premises, classical swine fever, risk factors, Bayesian modelling, Bulgaria.

---

## Introduction

Classical swine fever (CSF), a widely spread viral disease associated with poor sanitary conditions (Meuwissen et al., 1999), has a high economic impact on pig production. The European Union (EU) is the second most important pig-producing area in the world after China (FAOSTAT, 2012), so prevention is already well regulated with surveillance, control and vaccination programmes ([http://ec.europa.eu/food/animal/diseases/controlmeasures/csf\\_en.htm](http://ec.europa.eu/food/animal/diseases/controlmeasures/csf_en.htm)). Reducing the prevalence of CSF and, ultimately, elimination of the infection are chief EU objectives, and instituted preventive measures have led to a CSF-free status in almost ¾ of the 27 EU member states. However, from 2007 to 2011, 380 outbreaks have been reported in domestic pigs or wild boar by the World Organisation for Animal Health (OIE) in the seven remaining, endemic EU countries (Table 1). These outbreaks were

associated with either the movement of domestic pigs or contact with infected wild boar populations (Artois et al., 2002; Kramer-Schadt et al., 2007; WAHID, 2012).

There is a need to better understand the factors that contribute to CSF outbreaks despite the costly preventive and control programmes that are currently in place in many EU countries. Concerns regarding backyard pig-production have increased after the recent incorporation of new EU member countries (e.g. Romania and Bulgaria) with a predominance of backyard breeding and low biosecurity pig (BLB) premises. In these countries more than 96% of the farms are classified as backyard pig farms. The presence of BLB farms, with high risk practices (e.g. swill feeding), and where farmed pigs may come into contact with infected wild boar populations or contaminated fomites, are hypothesised as factors negatively affecting the control and elimination of CSF in the EU. A challenge to evaluate the role of BLB farms in CSF transmission in the EU is the lack of data, such as pig census, disease presence/absence, location of infected farms, pig movements and sanitary condition. Evaluation of these risk factors, specifically those related to backyard pigs that may be associated with the occurrence of CSF, are needed to better control and eventually eliminate the disease by improved cost-effective and risk-based surveillance and control strategies in

---

Corresponding author:  
Beatriz Martínez-López  
Center for Animal Disease Modeling and Surveillance  
VM: Medicine and Epidemiology  
One Shields Avenue, 1112 Tupper Hall  
UC Davis, Davis, CA 95616, USA  
Tel. +1 530 752-1618  
E-mail: beamartinezlopez@ucdavis.edu

the countries where backyard production dominates.

CSF control measures in Bulgaria have been carried out for many years but the implementation of the national control programme for CSF was enhanced after joining the EU in 2007. Since then, activities have evolved and led to a significant decrease in the number of CSF outbreaks with only sporadic outbreaks, primarily in BLB farms as well as wild boar. Currently, the Bulgarian Food Safety Agency is the authority responsible for control and elimination of CSF in Bulgaria. The programme currently implemented is based on (i) the categorisation of farms based on bio-security level and trade patterns; (ii) active surveillance based on clinical signs, submission of a check list and blood sampling for detection of antibodies in domestic pigs; (iii) vaccination campaigns in wild boar with distribution within a 40 km zone covering the west and north borders of the country; and (iv) surveillance of hunted wild boar by testing organ and blood samples.

The aims of this study were to: (i) evaluate the role that BLB pig farms may have in the occurrence of CSF; (ii) identify other risk factors that may be contributing to CSF occurrence in Bulgaria; and (iii) describe the spatial patterns and regions that may be at highest risk for CSF occurrence in Bulgaria and where surveillance and control measures should be targeted.

## Materials and methods

### *Study area and farm types*

Bulgaria is located in Eastern Europe. It encompasses 110,994 km<sup>2</sup> and is bordered by Romania to the north, the Black Sea to the east, Turkey and Greece to the south and the Republic of Macedonia and Serbia

Table 1. CSF outbreaks and years of positivity in various European countries between 2007 and 2011.

Country	Number of outbreaks	Number of years positive
Bulgaria	8	3
France	1	1
Germany	13	2
Hungary	191	2
Lithuania	5	2
Romania	159	1
Slovakia	3	2
Total	380	13*

\*Mean number of years positive by country = 1.86; standard deviation = 0.69.

The table is based on WAHID (<http://web.oie.int/wahis/public.php?page=home> (accessed on February 2012)).

to the west. Bulgaria was selected for this study because it (i) is a EU country with endemic CSF; (ii) has a high proportion of backyard pig farms (96%); (iii) has implemented a compulsory registration and control of BLB farms, and (iv) made the epidemiologic and demographic data required for this analysis available.

Bulgarian pig farms are categorised into five types: (i) industrial farms with a high level of biosecurity (HLB); (ii) family farms type A with a medium level of biosecurity (MLB); (iii) family farms type B at BLB level; (iv) backyard farms at BLB level; and (v) East-Balkan pig herds (traditionally managed herds where pigs are fed in open grassy areas and only trade with other East-Balkan pig farms is allowed) (Alexandrov et al., 2011a). Full pig trading rights are only allowed at the type (i) and (ii), while restricted to other non-industrial pig farms for (iii) and not allowed at all for the remaining two.

### *Data*

CSF outbreaks, pig demographics and pig movements were provided by the Bulgarian Food Safety Agency. Data included the number of outbreaks per municipality from 1999 to 2010; the number of farms; the number of pigs per farm type; number of pigs per municipality in 2010; and pig movement records per farm from January to October 2010. Human demography during 2010 was obtained from the Bulgarian National Statistical Institute (National Statistical Institute, 2011) and included population by municipality, place of residence (urban, rural) and gender. Socioeconomic factors such as household consumption, overall poverty including level of poverty were obtained from a report based on the Bulgarian National Statistics (Ivaschenko, 2004). Maps for environmental and anthropogenic factors (e.g. water sources, altitude and roads) were obtained from DIVA-GIS, a free computer programme for mapping and geographical data analysis (<http://www.diva-gis.org> - accessed on June 2011). Finally, the CORINE land cover for 2006 (<http://www.eea.europa.eu/> - accessed on May 2011) was obtained from the European Environmental Agency and used to estimate the potential for wild boar presence.

### *Farms and CSF occurrence in Bulgaria*

Detailed information on BLB pig farms and animal movement patterns was sparse, so we described and characterised the pig demographic and movement data

by evaluating the spatial and temporal patterns of available pig movements among different types of premises, and calculated statistics and metrics including mean and standard deviation (SD) for the number of shipments, animals shipped by type of premise and shipping distance (i.e. Euclidean distance). We computed the probability of a shipment that originated in farm type  $o$  being received in farm type  $r$  ( $PS_{o \rightarrow r}$ ) as the number of shipments sent from farm type  $o$  to farm type  $r$  over the total number of shipments sent by farm type  $o$ . Similarly, we computed the probability of shipments sent in the opposite direction ( $PR_{o \rightarrow r}$ ). Maps and graphs describing the distribution of pig farms and pig movements per municipality as well as the spatial and temporal distribution of CSF outbreaks were produced using ArcGIS version 9.3 (ESRI; Redlands, USA) and R version 2.12.1 (R Development Core Team, 2010). These descriptive analyses were used to generate hypotheses about the best predictors to include in the Bayesian model.

*Model specification*

A Bayesian spatial mixed, multivariable logistic regression model was used to evaluate the association between the hypothesised risk factors and the probability of CSF occurrence in Bulgaria per municipality  $i$  ( $p_i$ ). We used a Bayesian framework because we wanted to explicitly account for the spatial correlation structure, for the potential heterogeneity or overdispersion of the data, and to present the uncertainty in model outputs. The incorporation of a spatial structure is fundamental when the independence assumption is violated, usually a concern with regard to acute infectious diseases, and this type of model reduces the potential ecological bias by indirectly accounting for the effect of the unmeasured risk factors (Richardson and Monfort, 2000).

The model presented used an integrated, two-level structure. The first level consisted of local CSF status by municipality  $Y_i$  (value of 1 if the municipality reported a CSF outbreak between 1999 and 2010; otherwise 0). We assumed that the data was Bernoulli distributed,  $Y_i \sim Ber(p_i)$ . The second level of the model consisted of a logistic mixed model of the form:

$$\begin{aligned} \text{logit}[p_i (Y = 1)] &= \log[\text{odds} (Y = 1)] = \\ \log\left(\frac{p_i (Y = 1)}{1 - p_i (Y = 1)}\right) &= \beta_0 + \sum_{k=1}^m \beta_k X_{k,i} + U_i + S_i \end{aligned} \tag{equation 1}$$

where  $\beta_0$  is the intercept;  $\sum_{k=1}^m \beta_k X_{k,i}$  the sum of  $m$  selected covariates ( $X$ ) measured in each municipality  $i$  and multiplied by the respective coefficients  $\beta_k$ ; and where  $U_i$  and  $S_i$  represent unstructured ( $U$ ) and structured ( $S$ ) (spatial) random effects, respectively. Vague or non-informative normal priors of the form  $N(0, 4)$  were specified for intercept and the regression coefficients. The unstructured random effect  $U_i$  was defined using a non-informative normal prior  $U_i \sim N(0, \tau)$ , where  $\tau$  was the precision term modelled with a gamma distribution of the form  $(0.5, 0.0005)$  (Kelsall and Wakefield, 1999). The structured random effect  $S_i$ , which provides spatial smoothing and accounts for correlation between neighbouring areas, was modelled using an intrinsic conditional autoregressive (CAR) prior structure, in which  $S_i$  was Gaussian distributed and centred on the mean neighbouring values for  $S$  (Besag et al., 1991). Specifically,

$$S_i | S_{i \neq j} \sim N\left(\bar{S}_i \frac{\sigma_s^2}{n_i}\right) \tag{equation 2}$$

where  $\bar{S}_i$  is the mean of the structured random effects of the neighbours of municipality  $i$ ,  $\sigma_s^2$  the conditional variance, which characterised the amount of spatial variation, and  $n_i$  the number of neighbours for each municipality  $i$ . The value  $\bar{S}_i$  is computed as:

$$\bar{S}_i = \frac{\sum_{i \neq j} w_{ij} S_j}{\sum_{i \neq j} w_{ij}} \tag{equation 3}$$

where  $w_{ij}$  are neighbourhood weights based on the adjacency rule in which municipalities sharing a common boundary are considered neighbours. A gamma distribution of the form  $(0.5, 0.0005)$  was used as the hyper-prior for  $\sigma_s^2$  (Kelsall and Wakefield, 1999).

Finally, we used the logistic regression formula (equation 1) to compute odds ratios (OR) to compare two or more groups with different sets of predictors. The general formula for the OR for group A and B is as follows:

$$\begin{aligned} \text{Odds ratio}_{X_A \text{ vs. } X_B} &= \frac{\text{odds } X_A}{\text{odds } X_B} = \\ \frac{e^{(\beta_0 + \sum_{k=1}^m \beta_k X_{Ak,i} + U_i + S_i)}}{e^{(\beta_0 + \sum_{k=1}^m \beta_k X_{Bk,i} + U_i + S_i)}} &= e^{\sum_{k=1}^m \beta_k (X_{Ak,i} - X_{Bk,i})} \end{aligned} \tag{equation 4}$$

where  $X_A$  and  $X_B$  denote the collection of  $X$ 's covariates for groups A and B, respectively, and may be

defined by vectors  $X_A = (X_{A1,i}, X_{A2,i}, \dots, X_{Am,i})$  and  $X_B = (X_{B1,i}, X_{B2,i}, \dots, X_{Bm,i})$ , respectively.

### Strategy for variable selection

A total of 68 covariates (and their second-order interactions) were evaluated for an association with CSF occurrence (see Appendix). Covariates were used in one of two forms: (i) standardised, to reduce correlation between beta coefficients, increase numerical accuracy (i.e. produce equal variances for all predictors) and normalise results; or (ii) binomial using the median as the cut-off point.

To assure the selection of the best predictors for CSF occurrence in Bulgaria, three different strategies for variable selection were used: (i) the all-possible-regression procedure; (ii) backward elimination; and (iii) the forward selection processes. A non-spatial, frequentist saturated (i.e. all main effects) logistic regression model was built in R-language. The best predictors of this saturated model were selected using the previously mentioned strategies for variable selection, i.e. the all-possible-regression procedure was run using the *MuMin* package (Bartoń, 2012) and the backward elimination and forward selection processes were conducted using the *MASS* package (Ripley et al., 2012). Akaike information criterion (AIC), was used as criterion for selection of the best models. The area under the curve (AUC) of the receiver operator curve (ROC) and the Hosmer-Lemeshow goodness of fit test (HL) were also employed to evaluate the predictive ability of the best candidate models, which were computed using the *epicalc* (Chongsuvivatwong, 2012) and *MKmisc* (Kohl, 2012) packages. Once the best main effects models were identified, we added the two-way interaction terms and, applied the all-possible-regression procedure, backward and forward model selection methods, with AIC, AUC and HL as criteria for model selection, to identify any significant interaction terms that should be retained in the final models. Variance inflation factor (VIF) was used to test for multicollinearity in the final models with the *car* package (Fox et al., 2012).

Concurrently, a forward selection process was implemented within a Bayesian spatial logistic framework to evaluate the main effects and the two-way interactions that were significantly associated with CSF occurrence. This process was done using the Gibbs sampling Markov chain Monte Carlo (MCMC) algorithm in WinBUGS (Spiegelhalter et al., 2003) with the *R2WinBUGS* package (Sturtz et al., 2005). Specifically, two simultaneous chains of 110,000 iterations were run for each “new” model, with a thin of 10 to reduce built-in autocorrelation. As a result, a total of 20,000 iterations were stored after each run and used to make posterior inferences. The deviance information criterion (DIC) was used to select the best predictors for the Bayesian model (Spiegelhalter et al., 2003).

Finally, main effects and two-way interaction terms were entered in the Bayesian model as well as those predictors retained in the non-spatial logistic regression model were selected for inclusion in the “final” Bayesian model (Wald’s P-value  $\leq 0.10$ ). The final Bayesian model was run 110,000, with two chains, a thin of 10, and from which a total of 20,000 iterations were stored and used to make posterior inferences. After each run in the forward selection process and the final Bayesian model, convergence of the MCMC chains was monitored using trace and autocorrelation plots, the Brooks-Gelman-Rubin plots, and the scale reduction statistic,  $\hat{R}$  (Brooks and Gelman, 1998). Convergence diagnostics were performed using the *coda* package in R-language (Martyn et al., 2012).

## Results

### *Pig demographics, pig movements and CSF occurrence in Bulgaria*

The pig sector in Bulgaria during 2010 was comprised of 60,654 pig farms and a total of 627,339 pigs. The breakdown by pig farm type and number of pigs in Bulgaria is shown in Table 2. The mean, median and standard deviation number of backyard farms per municipality were 223, 154 and 273, respectively. Those values for the family farms type B were 6, 3 and

Table 2. Numbers of farms, pigs and pig shipment data in Bulgaria 2010.

Farm type	Number of farms	Number of pigs (%)	Pigs per shipment	Shipment distance (km)
Industrial farms	63	70.8	22 ± 28	77.0 ± 103.8
Family farms A	88	3.9	19 ± 35	74.5 ± 110.6
Family farms B	1,736	6.6	4 ± 13	32.9 ± 84.5
Backyard farms	58,669	17.1	-	-
East Balkan herds	98	1.6	40 ± 25	90.2 ± 99.7

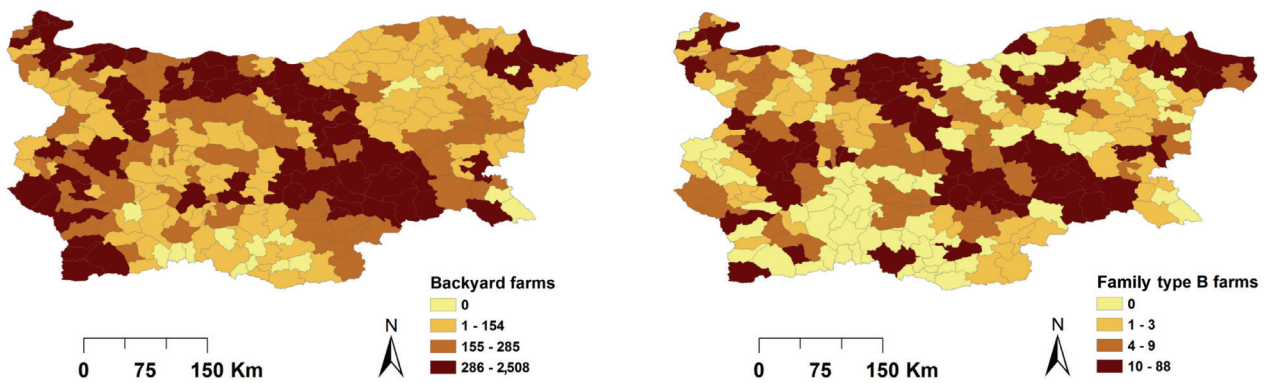


Fig. 1. The number of backyard and family type B farms by municipality during 2010 in Bulgaria.

11, respectively. Spatial distribution of backyard and family farms type B were moderately correlated (Spearman’s rank correlation coefficient  $\rho = 0.42$ ,  $P < 0.001$ ), although backyard farms were much more abundant (Fig. 1). Sixteen municipalities (6% of all municipalities) had no backyard farms and 84 (32%) had no family farms type B.

From January to October 2010, there were 3,804 pig movements involving 64,540 pigs with 17 (SD = 28) animals moved per shipment. Approximately 40% of the shipments occurred within the same municipality with a mean distance covered per shipment equal to 65.6 km (SD = 102.3 km). Overall, most of the shipments and receptions were concentrated in a small number of municipalities (Fig. 2). Municipalities with greater than 100 outgoing movements ( $n = 13$ , 5% of all municipalities) had approximately 60% of outgoing movements occurring from January to October 2010. Similarly, municipalities with more than 100 incoming movements ( $n = 11$ , 5% of all municipalities) received 46.6% of all pig movements. There was a high correlation (Spearman  $\rho = 0.74$ ,  $P < 0.001$ ) between the outgoing and incoming number of movements per municipal-

ity. When looking for the type of farm involved in the movement, we observed that 55.9% of the shipments originated in industrial farms, whereas 26%, 18% and 0.05% originated in type B, type A and East Balkan herds, respectively. Backyard farms had the highest percent (89.7%) of shipments received, whereas industrial farms, type A, type B and East Balkan herds received 7.4%, 1.7%, 1.2% and 0.05%, of the shipments, respectively. Mean and SDs of animals shipped and distances transported are shown in Table 2.

The probability of shipment (PS) and reception (PR) of pig movements for each particular type of farm is detailed in Fig. 3. For example, most of the movements of industrial, type B and type A farms were shipped to backyard farms as indicated by  $PS_{(Industrial \rightarrow Backyard)} = 0.86$ ,  $PS_{(Type\ B \rightarrow Backyard)} = 0.97$  and  $PS_{(Type\ A \rightarrow Backyard)} = 0.91$ , respectively. Conversely, 53% of the receptions of movements in backyard farms were attributed to movements that originated in industrial farms as indicated by  $PR_{(Backyard \leftarrow Industrial)} = 0.53$  and from family farms type B ( $PR_{(Backyard \leftarrow Type\ B)} = 0.29$ ).

Between 1999 to 2010, there were 68 outbreaks registered in Bulgaria with more than 573 infected pigs.

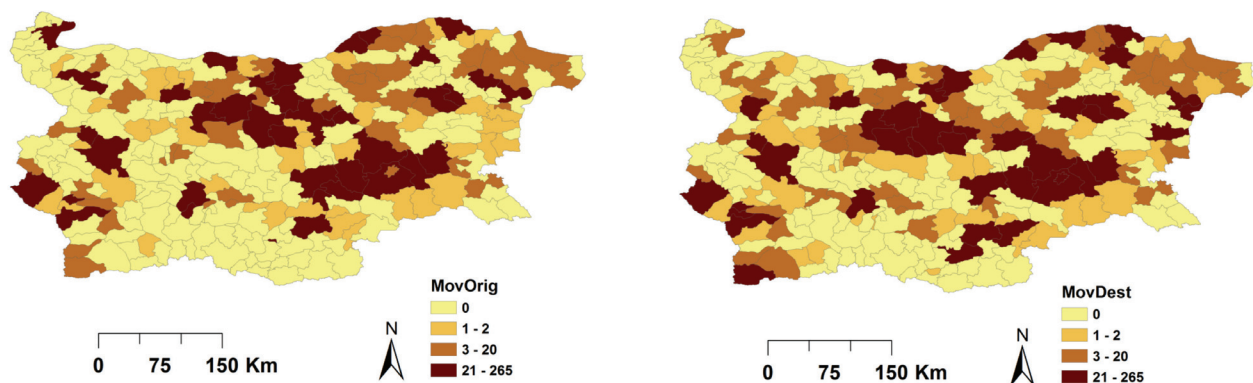


Fig. 2. The number of outgoing and incoming shipments of pigs per municipality from January to October 2010 in Bulgaria. Outgoing (MovOrig) and incoming (MovDest) shipments.

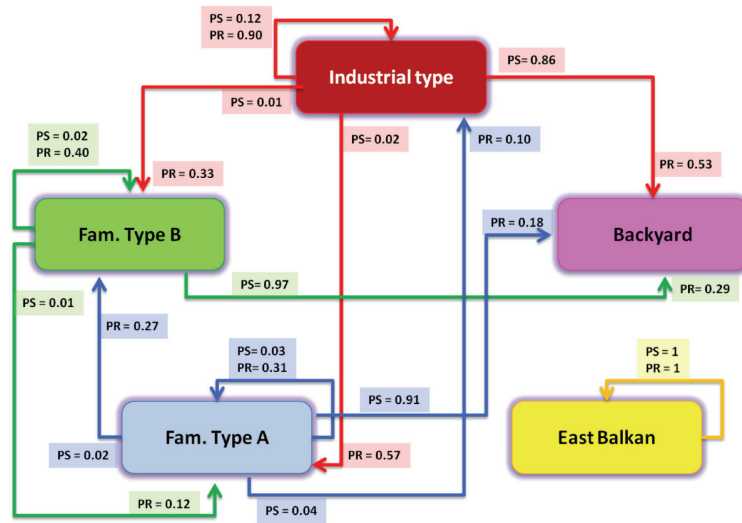


Fig. 3. Probability of shipment and reception of pigs from and to the different categories of Bulgarian pig farms in 2010. PS = probability of shipment of pigs; PR = probability of reception of pigs.

Nearly two-thirds (63.3%) of those outbreaks involved industrial type A and type B farms, whereas East Balkan pig farms, backyard farms and wild boar accounted for 13.2%, 13.2% and 10.3% of the outbreaks, respectively. The temporal distribution of disease showed that most of the outbreaks (92.6%) and of infected pigs (82.0%) occurred before 2007 (Fig. 4).

*Bayesian model*

The predicted probability of CSF occurrence in Bulgaria was in general low with a median value of 0.14 and with most of the risk of CSF occurrence concentrated in a few municipalities (Fig. 5). In fact, only 10% of the municipalities had an estimated probability of CSF occurrence above 0.3. The estimate of the variance ( $s^2$ ) for the structured [S] random effects

( $s^2[S] = 0.009$ , 95% credibility interval (CI\*) = 0.0001-0.463) was higher than the estimate of the variance in the unstructured [U] random effects ( $s^2[U] = 0.002$ , 95% CI = 0.0002-0.108), which indicates that unobserved risk factors tend to be clustered and that there is spatial correlation in the data (Fig. 6). Nevertheless, both variances for S and U were quite low, which may indicate that most of the spatial correlation in the data was captured by the covariates.

Six predictors and one interaction term were retained in the final Bayesian model (Table 3). Of the variables in the model, high number of backyard pigs (BYpigs), high number of outgoing shipments (OMov) and the presence of East Balkan pigs (EBpigs.B) were significant (or borderline significant) risk factors for CSF occurrence in Bulgaria. The interaction between family type B pigs and personal household consump-

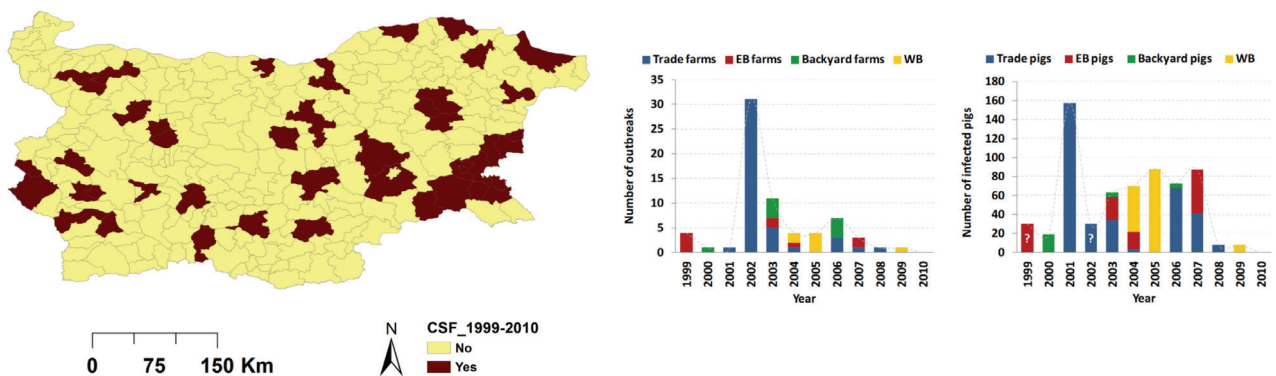


Fig. 4. The municipalities of Bulgaria where number of CSF outbreaks (left) and infected pigs (right) was reported annually from 1999 to 2010. EB = East Balkan; WB = wild Boar; during the 1999 and 2002 CSF outbreaks, the infected pig numbers were unknown as indicated by question marks in the graph.

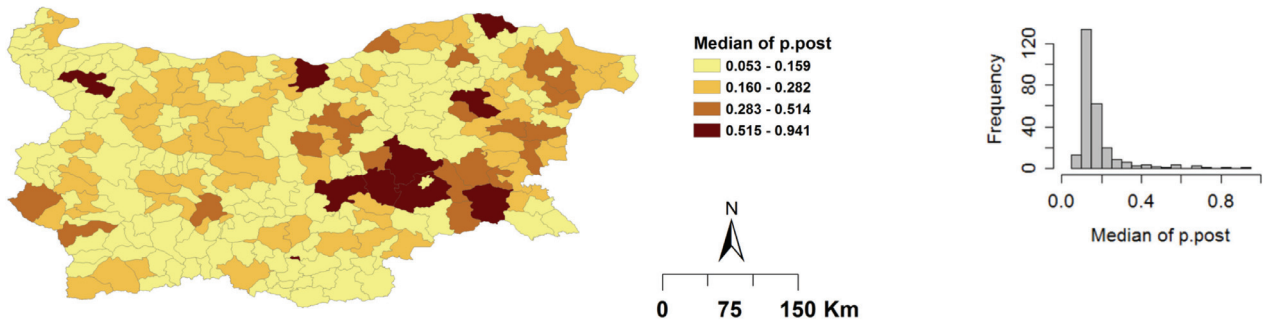


Fig. 5. Histogram and spatial distribution of the median predicted probability of CSF occurrence in Bulgaria from 1999 to 2010. Median of the predicted probability (p.post) calculated for each municipality in Bulgaria.

tion was also significant, as indicated by the 95% CI. Fig. 7, shows that the risk of CSF occurrence increases when there is a combination between high number of family type B pigs and low levels of personal household consumption. The odds of CSF occurrence is 5.2 times greater in a municipality A with a high number of family type B pigs and low personal household consumption when compared with municipality B with low number of family type B pigs and high personal consumption:

$$\begin{aligned}
 \text{Odds ratio}_{X_A \text{ vs. } X_B} &= e^{\beta_1 (4.63-0) + \beta_2 (0-1) + \beta_3 (0-0)} \\
 &\approx 5.21, 95\% \text{ CI}^* [1.41-19.13]
 \end{aligned}$$

(equation 5)

Analysis using Brook-Gelman Rubin plots and the scale reduction statistic ( $R^{\wedge}$ ) showed the model converging with values close to one. There was no evidence of autocorrelation within chains for posterior inferences. Density and trace plots generated for model diagnosis for one of the significant parameters of the model are shown in Fig. 8.

### Discussion

Overall, the study presented here is one of the first to quantitatively describe BLB pig farm demographics and animal movements and to evaluate the association between BLB pig production systems and CSF occurrence. Results revealed that the presence of low biosecurity pig farms (i.e. type B, East Balkan and backyard) and a high number of pig shipments in a municipality were significant contributors to CSF occurrence in Bulgaria. Furthermore, the combination of low levels of personal household consumption with a high number of type B pigs substantially increased the odds of CSF occurrence (Table 3, Fig. 7).

The results at the municipality level indicate that a combination of socioeconomic factors, biosecurity level on farm and trade patterns are the factors that contribute most to CSF occurrence in Bulgaria. The most important factors are the use of contaminated fomites, particularly in poor areas, pig shipments and contact with infected wild boar populations. Previous studies show that vaccination and/or trapping are successful methods to control and prevent CSF virus

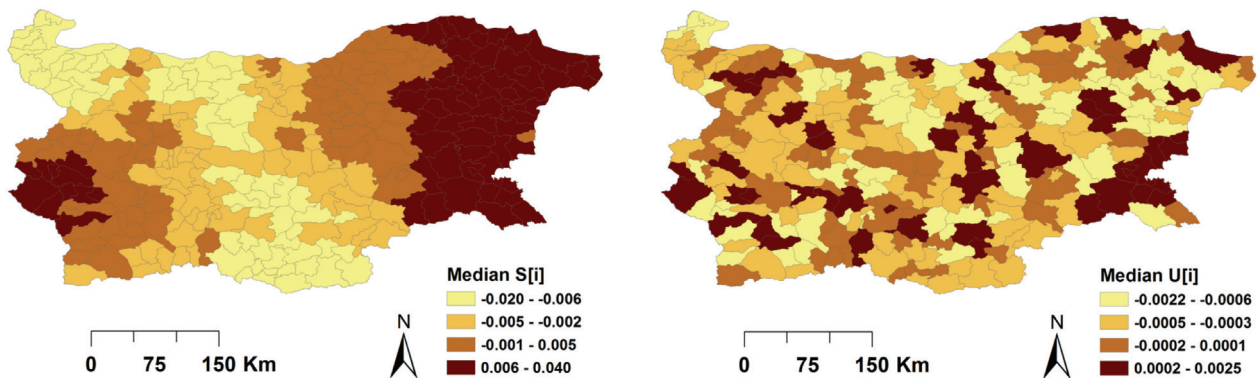


Fig. 6. Maps of the median values for the structured and unstructured random effects obtained in the Bayesian hierarchical model. S = structured random effects; U = unstructured random effects.

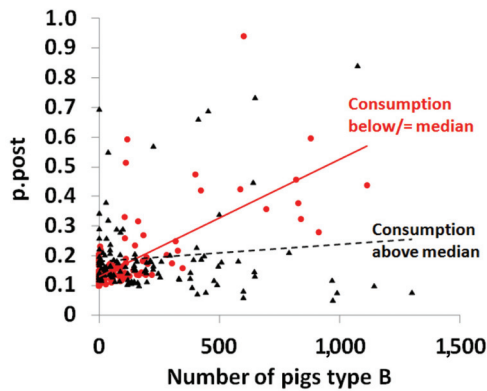


Fig. 7. Effect of the interaction among household consumption and number of pigs in family type B farms per municipality on the predicted probability of CSF occurrence in Bulgaria (p.post).

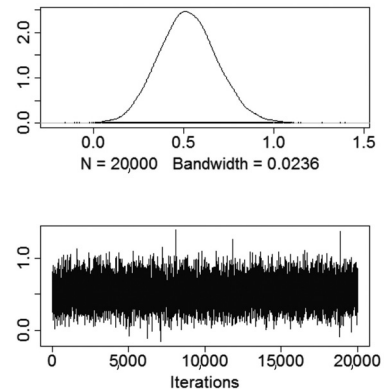


Fig. 8. Example of density and trace plots generated for the significant measure of effect ( $\beta_4$ ) for the predictor of the number of outgoing pig movements per municipality (OMov).

(CSFV) transmission from wild boar to domestic pigs in Bulgaria (Alexandrov et al., 2011b). Although no detailed information on wild boar abundance at the municipality level is available for Bulgaria, we quantified the potential role of wild boar with respect to CSF by including a covariate that indicates the proportion of suitable areas for wild boar in a municipality. Our results show that CSF occurrence was not significantly influenced by the presence or absence of wild boar. However, we hypothesise that East Balkan pigs, because they are in free-ranging herds, are at the highest risk of becoming infected with CSFV from wild boar reservoirs. In our analysis, municipalities with East Balkan pigs had a 2.15 times greater odds of CSF outbreaks than municipalities without East Balkan pigs. Similarly, backyard pigs may also become infected through contact with wild boar or fomites if they are managed under free-ranging conditions. However, after evaluating the trade patterns of pigs in both East

Balkan and backyard pig farms in Bulgaria, we feel that it is unlikely that a direct CSFV-transmission from East Balkan or backyard pigs to other types of domestic pig farms would occur because there is no trade from East Balkan and backyard pig farms to other Bulgarian types of pig farms (Fig. 3).

Contaminated fomites, which we define as any contaminated material that contributes to the indirect transmission of CSFV (e.g. carcasses, swill feeding but also trucks, people and hunting equipment, etc.), seem to play an important role in CSF occurrence in Bulgaria. Based on the results from our model, we hypothesise that swill feeding or other high risk practices can lead to indirect CSFV-transmission, especially in poverty-stricken areas, i.e. neighbourhoods with generally low levels of household consumption, which is inversely and significantly correlated (Spearman  $\rho = -0.89$ ,  $P < 0.001$ ) with the number of poor people per municipality (Fig. 9). The results from the model fur-

Table 3. Association between demographic or socioeconomic factors and CSF occurrence in Bulgaria as a result of the non-spatial logistic regression model and the multilevel Bayesian logistic spatial model.

Variable	Logistic non-spatial model <sup>a</sup>			Bayesian logistic spatial model <sup>b</sup>		
	OR	95% CI	P-value <sup>c</sup>	OR	95% CI	90% CI
Number of type B pigs	1.57	0.92-2.65	0.096	1.36	0.91-2.04	0.97-1.92
Number of backyard pigs	1.20	0.84-1.72	0.317	1.22	0.87-1.72	0.92-1.62
Number of East Balkan pigs <sup>d</sup>	22.99	4.72-111.98	<0.001	2.15	0.92-4.82	1.05-4.27
Number outgoing shipments	1.62	1.32-2.68	<0.001	1.68	1.24-2.36	1.30-2.22
Total urban population	2.81	1.21-6.52	0.016	1.46	0.81-2.60	0.89-2.35
Household consumption <sup>e</sup>	1.09	0.47-2.52	0.841	0.81	0.46-1.42	0.50-1.30
Number of type B pigs x Household consumption	0.31	0.14-0.69	0.004	0.51	0.24-0.88	0.32-0.81

OR = odds ratio; CI = confidence interval; CI\* = credibility interval; <sup>a</sup>AIC=201; AUC (ROC) = 0.80; <sup>b</sup>DIC = 211; spatially structured variance ( $s^2[S]$ ) = 0.009, 95% CI = (0.0001-0.463); non-spatially structured or unstructured variance ( $s^2[U]$ ) = 0.002, 95% CI = (0.0002-0.108); <sup>c</sup>Wald's test; <sup>d</sup>East Balkan pigs above (1) or  $\leq$ (0) the median; <sup>e</sup>per capita consumption above (1) or  $\leq$ (0) the median.



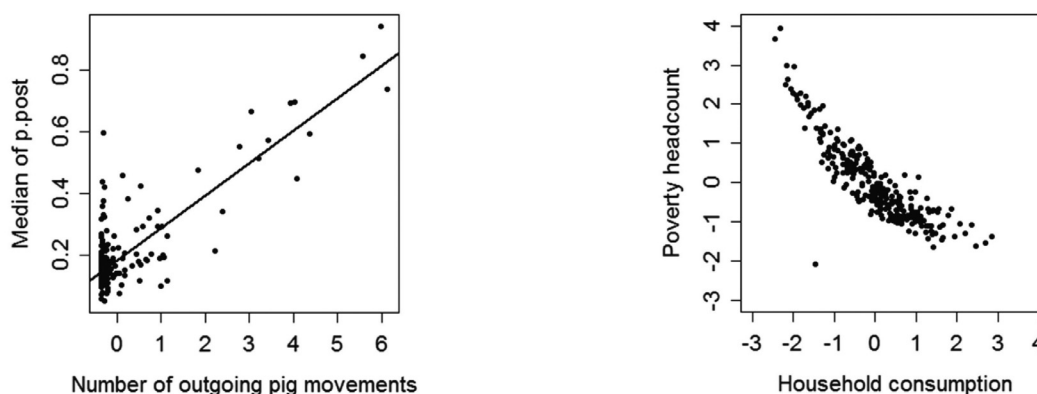


Fig. 9. Relationship between the probability of CSF occurrence in Bulgaria and outgoing pig shipments (above) and between household consumption and the number of poor people (below).

ther suggest that the combination of economically depressed areas and a high number of type B pig farms significantly contribute to CSF occurrence in Bulgaria. This may occur because type B farms in poor regions have poor sanitation, biosecurity and infrastructure and are hypothesised to more likely and frequently use swill feeding which may facilitate the transmission of pathogens in general, CSFV in particular. Moreover, pigs in family farms of type B raised in low biosecurity holdings are allowed to move pigs to other non-industrial pig farms; therefore these holdings are not only at a high risk for CSFV-infection, but may also spread this pathogen to other domestic pig premises. These family farms type B are responsible of the second largest number of pig shipments (26%) to other farms. The substantial trade of type B farms highlights the importance of pig shipments (OR = 1.68, 95% CI\*[1.24-2.36]) with respect to the spread of this disease, either by direct movement of infected pigs or via indirect contact with contaminated trucks. If we consider that most farms receiving pigs are BLB premises (Fig. 3), potential contamination from farm to farm may indeed occur by vehicles used to ship the pigs if these vehicles are not thoroughly disinfected.

A report published by Yoveva et al. (2000) is of interest in this connection since it describes how urban agriculture is widespread and used to overcome food shortages and seasonal fluctuations in food prices, particularly in low-income urban families. These urban agricultural practices include production of vegetables, fruits and animals (in addition to pigs), in private gardens and backyards and this food is generally processed at home for self-consumption. These products contribute to subsistence needs, but if there is an excess of production, the products are sold. This report also highlights that these urban farms are often

managed by retired, non-professional people. In Sofia, the capitol of Bulgaria, for example, urban farms produced 21,133 pigs during 1997, 92% of which were attributed to small private farms (i.e. backyard and type B farms). Similarly, peri-urban municipalities in Bulgaria allocate small farms or backyards mostly for self-consumption or for local sale. Our results, suggesting that the total urban population is a borderline significant predictor for CSF occurrence, may be explained, at least in part, by these characteristic and widespread urban agricultural practices.

Previous studies show that spatial Bayesian modelling is a preferred modelling method that minimised the ecological bias and accounts for unidentified or unmeasured confounders when count-data is analysed (Elliott et al., 2000). We believe that the model used here is at the leading edge of current state-of-art statistical modelling and appropriate to provide reliable estimates for CSF in Bulgaria. The spatial Bayesian logistic regression model explicitly accounts for spatial dependency among municipalities. The inclusion of the structured (i.e. spatial) and unstructured random effects allows for an indirect accounting of the unobserved risk factors or unknown confounders that may be spatially clustered (Elliott et al., 2000). In fact, the spatial random effects map (Fig. 6) suggests the presence of two clustered areas, one in the eastern and a small one in the western part of Bulgaria, where unobserved factors seem to contribute to the odds of CSF occurrence. Studies aimed at identifying and quantifying unobserved risk factors in the eastern and western regions may provide additional information to improve CSF preventive and control measures. Similarly, the inclusion of unstructured random effects accounting for overdispersion or between-area heterogeneity in the model, indirectly incorporates unmea-

sured risk factors that are common to all individuals in Bulgaria and that do not have a spatial pattern (Fig. 6). Structured and unstructured random effect values were low in our analysis, which suggests that most of the spatial correlation was well captured by the covariates and that there was a relatively low influence of unobserved risk factors in the prediction of CSF occurrence. Conversely the frequentist approach ignores the spatial dependence in the statistical analysis, which leads to overestimation of the statistical significance levels and increases the likelihood of making a type I error (i.e. a true null hypothesis is rejected) (Pfeiffer et al., 2008). Estimates and confidence intervals from the frequentist approach are inaccurate for covariates with a small number of observations (e.g. see point estimate and 95% CI of the frequentist model for the number of East Balkan pigs in Table 3). Generally, CI\* of the Bayesian approach are narrower than the confidence intervals (CI) of the frequentist model, which resulted in an increase in precision of the Bayesian estimates. In general, there is agreement in the direction of the estimates in the frequentist and the Bayesian approaches so it is usually a reasonable approach to start fitting a frequentist model to rapidly explore potential good predictors to be included into a subsequent spatial Bayesian model (Table 3) (Gosoni et al., 2010). The intensive variable selection process performed here was considered appropriate to select the best fitting model with a large number of predictors ( $n=68$ ) and second-order interactions; alternatively, a Bayesian variable selection procedure could be implemented to explore all possible models while taking into account residual spatial correlation in the data, which may also save some time (O'Hara and Sillanpää, 2009). In general, there was good agreement among the predictors selected by the different variable selection procedures. Evaluating the two forms (i.e. standardised and binomial) was an extra step to certify that the best fitting predictors were retained in the final model.

We could only incorporate data aggregated at the municipality level the model, because it was the lowest administrative level for which information was available and it is the spatial level used for decision making in CSF eradication and control in Bulgaria. It is true that the aim of the study was, primarily, to provide a smoothed risk map for CSF in Bulgaria with the lowest ecological bias and the highest specificity (Fig. 5). This map accurately indicates areas to target for disease control and allows rapid visualization of the spatial variation in disease risk throughout the country. We also intended to identify factors that contribute to

spatial distribution of CSF occurrence, particularly those related to pig and human demographic, socioeconomic and environmental factors, which are only available at the municipality level. However, an ecological study does not allow identification of the factors or routes of CSF exposure or transmission at a farm or pig level. Data at the farm and pig level could guide control and elimination policies more precisely. We recommend using our study as a baseline to conduct more detailed studies at the farm or pig scale, when (or if) data at these levels become available.

## Conclusion

Methods and results presented here may be useful for risk-based and cost-effective allocation of financial resources for improved prevention and control of CSF in Bulgaria and other countries with similar epidemiological conditions. This study shows an approach to better understand the complexity of CSF epidemiology in BLB production systems in Bulgaria because:

- (i) BLB farms and pig trade have been characterised in detail, providing useful statistics on farm counts, spatial distribution and probability of contact among types of premises, which may be used in future CSF models or to control other pig diseases;
- (ii) a CSF risk map has been provided, which would facilitate targeting of CSF interventions for control and, ultimately, elimination; and
- (iii) the primary factors contributing to the spatial CSF distribution has been identified.

## Acknowledgements

Authors would like to acknowledge the provision of data and support of the Bulgarian Food Safety Agency as well as the support and valuable comments of Frank Koenen and other CSFV\_goDIVA partners. This work has been funded by the European Project CSFV\_goDIVA (KBBE-227003).

## References

- Alexandrov T, Kamenovi P, Depner K, 2011a. Surveillance and control of classical swine fever in Bulgaria, a country with a high proportion of non-professional pig holdings. *Epidemiol Santé Anim* 59/60, 140-142.
- Alexandrov T, Kamenov P, Stefanov D, Depner K, 2011b. Trapping as an alternative method of eradicating classical swine fever in a wild boar population in Bulgaria. *Rev Sci Tech* 30, 911-916.
- Artois M, Depner KR, Guberti V, Hars J, Rossi S, Rutili D,

2002. Classical swine fever (hog cholera) in wild boar in Europe. *Rev Sci Tech* 21, 287-303.
- Bartoń K, 2012. Package “MuMIn”. Available at: <http://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf> (accessed on May 2012).
- Besag J, York J, Mollié A, 1991. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1-59.
- Brooks SP, Gelman A, 1998. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7, 434-455.
- Chongsuvivatwong V, 2012. Package “epicalc”. Available at: <http://cran.r-project.org/web/packages/epicalc/epicalc.pdf> (accessed on May 2012).
- Elliott P, Wakefield J, Best N, Briggs D, 2000. *Spatial epidemiology: methods and applications*. Oxford: Oxford University Press, 494 pp.
- FAOSTAT, 2012. Stocks of live pigs in 2010. Available at: <http://faostat.fao.org> (accessed on March 2012).
- Fox J, Weisberg S, Bates D, Firth D, Friendly M, Gorjanc G, Graves S, Heiberger R, Laboissiere R, Monette G et al., 2012. Package “car”. Available at: <http://cran.r-project.org/web/packages/car/car.pdf> (accessed on May 2012).
- Gosoni L, Mia Veta A, Vounatsou P, 2010. Bayesian geostatistical modeling of malaria indicator survey data in Angola. *PLoS One* 5, e9322.
- Ivaschenko O, 2004. Poverty and inequality mapping in Bulgaria. Available at: <http://siteresources.worldbank.org/INTPGI/Resources/342674-1092157888460/Ivaschenko.MappingBulgaria.pdf> (accessed on March 2012).
- Kelsall JE, Wakefield JC, 1999. Contribution to “Bayesian models for spatially correlated disease and exposure data”. In: *Bayesian statistics*. Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford: Oxford University Press, 151 pp.
- Kohl M, 2012. Package “MKmisc”. Available at: <http://cran.r-project.org/web/packages/MKmisc/MKmisc.pdf> (accessed on May 2012).
- Kramer-Schadt S, Fernández N, Thulke HH, 2007. Potential ecological and epidemiological factors affecting the persistence of classical swine fever in wild boar *Sus scrofa* populations. *Mammal Rev* 37, 1-20.
- Martyn P, Best N, Cowles K, Vines K, 2012. Package “coda”. Available at: <http://cran.r-project.org/web/packages/coda/coda.pdf> (accessed on May 2012).
- Meuwissen MP, Horst SH, Huirne RB, Dijkhuizen AA, 1999. A model to estimate the financial consequences of classical swine fever outbreaks: principles and outcomes. *Prev Vet Med* 42, 249-270.
- National Statistical Institute, 2011. Available at: <http://www.nsi.bg> (accessed on September 2011).
- O’Hara RB, Sillanpää MJ, 2009. A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal* 4, 1-190.
- Pfeiffer DU, Robinson TP, Stevenson M, Stevens KB, Rogers DJ, Clements ACA, 2008. *Spatial analysis in epidemiology*. Oxford: Oxford University Press, 208 pp.
- R Development Core Team, 2010. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Richardson S, Monfort C, 2000. Ecological correlation studies. In: *Spatial epidemiology, methods and applications*. Elliott P, Wakefield JC, Best NG, Briggs DJ (eds). Oxford: Oxford University Press, 205-220 pp.
- Ripley B, Hornik K, Gebhardt A, Firth D, 2012. Package “MASS”. Available at: <http://cran.r-project.org/web/packages/MASS/MASS.pdf> (accessed on May 2012).
- Spiegelhalter DJ, Thomas A, Best N, Lunn D, 2003. *WinBUGS version 1.4: user manual*. Cambridge: Medical Research Council Biostatistics Unit. Available at: <http://www.mrc-bsu.cam.ac.uk/bugs> (accessed on November 2011).
- Sturtz S, Ligges U, Gelman A, 2005. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw* 12, 1-16.
- WAHID, 2012. World Animal Health Information Database (WAHID) Interface. Available at: [http://www.oie.int/wahis\\_2/public/wahid.php/Wahidhome/Home](http://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home) (accessed on February 2012).
- Yoveva A, Gocheva B, Voykova G, Borrisov B, Spassov A, 2000. *Sofia: urban agriculture in an economy in transition*. Available at: <http://www.ruaf.org/sites/default/files/Sofia.pdf> (accessed on May 2012).

## Appendix

Variable description	Standardised	Binomial <sup>a</sup>
Number of industrial farms	1. INDfarms	INDfarms.B
Number of industrial pigs	2. INDpigs	INDpigs.B
Number of type A farms	3. Afarms	Afarms.B
Number of type A pigs	4. Apigs	Apigs.B
Number of type B farms	5. Bfarms	Bfarms.B
Number of type B pigs	6. Bpigs	Bpigs.B
Number of backyard farms	7. Byfarms	Byfarms.B
Number of backyard pigs	8. Bypigs	Bypigs.B
Number of low biosecurity farms (type B + backyard)	9. Lbfarms	Lbfarms.B
Number of low biosecurity pigs (type B + backyard)	10. Lbpigs	Lbpigs.B
Number of East-Balkan herds	11. Ebfarms	Ebfarms.B
Number of East-Balkan pigs	12. EBpigs	EBpigs.B
Total number of pig farms	13. TOTfarms	TOTfarms.B
Total number of pigs	14. TOTpigs	TOTpigs.B
Number of outgoing shipments	15. OMov	OMov.B
Number of incoming shipments	16. IMov	IMov.B
Total number of pigs sent in outgoing shipments	17. OpigsMoved	OpigsMov.B
Total number of pigs received in incoming shipments	18. IpigsMoved	IpigsMov.B
Mean number of pigs moved per outgoing shipment	19. OpigsperMov	OpigsperMov.B
Mean number of pigs received per incoming shipment	20. IpigsperMov	IpigsperMov.B
Total human population	21. Pop	Pop.B
Total male population	22. Male	Male.B
Total female population	23. Fem	Fem.B
Total urban population	24. Urban	Urban.B
Total urban male population	25. UrbMale	UrbMale.B
Total urban female population	26. UrbFem	UrbFem.B
Total rural population	27. Rural	Rural.B
Total rural male population	28. RurMale	RurMale.B
Total rural female population	29. RurFem	RurFem.B
Density of suitable areas for wildboar (km <sup>2</sup> )	30. DenWB	DenWB.B
Density of water areas (km <sup>2</sup> )	31. DenWater	DenWater.B
Density of roads (km <sup>2</sup> )	32. DenRoads	DenRoads.B
Mean altitude (m)	33. Altitude	Altitude.B
Density of industrial farms (km <sup>2</sup> )	34. DenINDfarms	DenINDfarms.B
Density of industrial pigs (km <sup>2</sup> )	35. DenINDpigs	DenINDpigs.B
Density of type A farms (km <sup>2</sup> )	36. DenAfarms	DenAfarms.B
Density of type A pigs (km <sup>2</sup> )	37. DenApigs	DenApigs.B
Density of type B farms (km <sup>2</sup> )	38. DenBfarms	DenBfarms.B
Density of type B pigs (km <sup>2</sup> )	39. DenBpigs	DenBpigs.B
Density of backyard farms (km <sup>2</sup> )	40. DenBYfarms	DenBYfarms.B
Density of backyard pigs (km <sup>2</sup> )	41. DenBYpigs	DenBYpigs.B
Density of low biosecurity (type B + backyard) farms (km <sup>2</sup> )	42. DenLbfarms	DenLbfarms.B
Density of low biosecurity (type B + backyard) pigs (km <sup>2</sup> )	43. DenLbpigs	DenLbpigs.B
Density of East Balkan herds (km <sup>2</sup> )	44. DenEbfarms	DenEbfarms.B
Density of East Balkan pigs (km <sup>2</sup> )	45. DenEBpigs	DenEBpigs.B
Density of pig farms (total) (km <sup>2</sup> )	46. DenTOTfarms	DenTOTfarms.B

*continued*

Variable description	Standardised	Binomial <sup>a</sup>
Density of pigs (total) (km <sup>2</sup> )	47. DenTOTpigs	DenTOTpigs.B
Density of human population (total) (km <sup>2</sup> )	48. DenTOTPop	DenTOTPop.B
Density of male population (km <sup>2</sup> )	49. DenTOTMale	DenTOTMale.B
Density of female population (km <sup>2</sup> )	50. DenTOTFem	DenTOTFem.B
Density of urban population (km <sup>2</sup> )	51. DenTOTUrban	DenTOTUrban.B
Density of urban male population (km <sup>2</sup> )	52. DenUrbMale	DenUrbMale.B
Density of urban female population (km <sup>2</sup> )	53. DenUrbFem	DenUrbFem.B
Density of rural population (km <sup>2</sup> )	54. DenTOTRural	DenTOTRural.B
Density of rural male population (km <sup>2</sup> )	55. DenRurMale	DenRurMale.B
Density of rural female population (km <sup>2</sup> )	56. DenRurFem	DenRurFem.B
Proportion of backyard farms (backyard/total farms)	57. %BYfarms	%BYfarms.B
Proportion of type B farms (type B/total farms)	58. %Bfarms	%Bfarms.B
Proportion of low biosecurity farms (low biosecurity /total farms)	59. %LBfarms	%LBfarms.B
Proportion of backyard pigs (backyard pigs/total pigs)	60. %BYpigs	%BYpigs.B
Proportion of type B pigs (type B/total pigs)	61. %Bpigs	%Bpigs.B
Proportion of low biosecurity pigs (low biosecurity /total pigs)	62. %LBpigs	%LBpigs.B
Household per capita consumption	63. Cons	Cons.B
Poverty headcounts (%)	64. PovH	PovH.B
Poverty depth (%)	65. PovD	PovD.B
Poverty severity (%)	66. PovS	PovS.B
Inequality I: Theil mean log deviation index, GE(0)	67. IneGE0	IneGE0.B
Inequality II: Theil entropy index, GE(1)	68. IneGE1	IneGE1.B

<sup>a</sup>Coded as 0 ( $\leq$ median) or 1 ( $>$ median)