Social structure defines spatial transmission of African swine fever in wild boar

Kim M. Pepin[1,*], Andrew Golnar[1], and Tomasz Podgórski[2,3]

1. National Wildlife Research Center, USDA, APHIS, Wildlife Services, 4101 Laporte Ave., Fort Collins, CO, 80526.
2. Mammal Research Institute, Polish Academy of Sciences, Stoczek 1, 17-230 Białowieża, Poland
3. Department of Game Management and Wildlife Biology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Kamýcká 129, 165 00 Praha 6, Czech Republic

*Corresponding author: Kim Pepin, National Wildlife Research Center, United States Department of Agriculture, 4101 Laporte Ave., Fort Collins, CO, 80526; kim.m.pepin@aphis.usda.gov

RH: Pepin et al. ˙Spatial spread of African swine fever in wild boar

Author E-mails:
Kim Pepin: kim.m.pepin@aphis.usda.gov
Tomasz Podgórski:t_podgorski@ibs.bialowieza.pl

Abstract

1. African swine fever virus (ASFv) is endemic in wild boar in Eastern Europe, challenging elimination in domestic swine. Estimates of the distances between transmission events are crucial for predicting rates of disease spread to guide allocation of surveillance and

control resources. Transmission distances are mainly defined by spatial and social processes in hosts, but effects of these processes on spread are poorly understood, and inferences often include only one process.

2. To understand effects of spatial and social processes on disease dynamics we developed spatially-explicit transmission models with different assumptions about social and/or spatial contact processes. We fit the models to ASFv surveillance data from Eastern Poland from 2014-2015 and evaluated how inclusion of social structure affected inference.
3. The model that accounted for social along with spatial processes provided better inference of spatial spread and predicted that ~80% of transmission events are within the same family group.
4. The models predicted dramatically different effective reproductive numbers, both in magnitude and variation.
5. Specifying contact structure with spatial but not social processes can lead to very different disease dynamics and inference of epidemiological parameters. Uncertainty in these processes should be accounted for in predicting spatial spread in social species.

KEYWORDS: African swine fever, Effective reproduction number, Spatial transmission kernel, Surveillance, Wild boar,

INTRODUCTION

Spatial transmission kernels (STKs) are probability distribution functions of transmission distance between sequential cases of infection. They describe the variation and limits of spatial

disease spread per transmission event and inform how surveillance, containment, or mitigation strategies should be deployed [1]. For example, information on where cases may arise can inform what spatial radius should be used for ring culling, ring vaccination, spatial quarantine or intensive surveillance [2]. Without detailed genetic data or contact tracing data to reconstruct transmission history, STKs are predominantly estimated indirectly by fitting disease transmission models to available case data [1-3] providing valuable insight for developing intervention strategies [4-7]. However, models often make simplifying assumptions based on the available information that could have negative impacts on policy decisions if the models are not robust to violation of these assumptions. Common assumptions in models for estimating STKs include assuming a single introduction event and assuming observation of all transmission events [8]. Methods that account for these processes are important for providing more realistic predictions of spatial spread in systems where these assumptions are violated.

Another common issue is that the potential scope of contact heterogeneities is often simplified or lacking, such that uncertainty in model specification cannot be considered despite its potential importance [9]. In non-vector-borne disease systems, key drivers of contact heterogeneities include social [10] and spatial processes [2], yet our understanding of the relative role of these processes in driving spatial disease dynamics is weak in most systems [11, 12] and analyses that consider contact heterogeneities tend to focus on one or the other. Understanding the potential effects of these different elements of contact heterogeneity on inference of disease dynamics will provide insight on the appropriate scope of model uncertainty that should be considered in practical applications.

African swine fever virus (ASFv), a virulent virus of swine, emerged in domestic swine in Georgia in 2007 following a single introduction event from Africa [13]. After its initial

emergence, the virus spread quickly to Eastern Europe becoming endemic in wild boar, which has challenged elimination. With no effective vaccine or treatment options, control strategies are focused on reducing swine movement, decontamination, and culling [14]. Effectiveness of these strategies depends on being able to rapidly find new cases and target high-risk areas, thus models that can predict spatial spread are crucial tools. However, it has become clear that modeling efforts need to include realistic ecological details to improve predictions of spatial spread and our understanding of factors that drive it [15]. Wild boar have limited spatial movement and cluster into family groups suggesting that both social and spatial processes likely need to be considered for estimating STKs and thus spatial spread [16-18]. Also surveillance is mostly passive with only a small proportion of cases likely being observed, and genetic [19] and other analyses [16] suggest that re-introductions are common. Thus estimates of STKs that account for these realities are likely important for providing more accurate guidance for disease control policies.

In previous work we used ASFv surveillance data from wild boar in Poland during 2014-2015 to estimate the frequency of carcass-based transmission and re-introduction in outbreak dynamics while accounting for partially observed data [16]. Here, we extended the modeling approach to three different assumptions about the form of spatial transmission processes: 1) neighborhood (local transmission only), 2) exponential decay (distance distribution that includes long-distance processes), and 3) distance distribution with social structure (Fig. 1). We then predicted the STKs under each set of assumptions using the fitted models and evaluated the effects of spatial and social processes on STKs and a key epidemiological parameter – the effective reproduction number. Our results show striking differences in disease dynamics when social structure modifies spatially-determined contact heterogeneities, showing that both

processes need to be considered for accurate predictions of spatial spread in socially-structured species.

METHODS

Surveillance data and study site

The index case of ASFv in wild boar was detected in February 2014 in the north-eastern part of the country (53°19'33"N, 23°45'31"E), less than 1 km from the border with Belarus. Subsequent cases occurred close to the Belarusian border [20, 21]. By the end of 2015, 139 wild boar tested positive for ASFv in the area, with maximum distance of 27.4km west of the border and a 100km range along the border. The affected area is dominated by a mosaic of woodlands and agricultural land (crop fields, pastures, meadows) with several large (several hundred square kilometers), continuous forests. On average, forest covers 53% of the area. In 2014, average wild boar densities were estimated at 1.5 - 2.5 boar/km^2, locally ranging from 0.5-1 boar /km^2 to 3-5 boar/km^2 (Regional Directorate of State Forests, Białystok, Poland). ASFv surveillance used a combination of active and passive mechanisms, with samples being obtained through hunter recovery (active) or reports of road kills or carcasses found on the landscape opportunistically (passive). A total of 4625 samples were from hunters, while 271 were from road kill or non-road kill carcasses. Samples, collected by veterinary services and hunters, were submitted to the National Reference Laboratory for ASFv at the National Veterinary Research Institute in Puławy, Poland. Detailed description of laboratory procedures and tests can be found in Woźniakowski et al. 2015 and Śmietanka et al. 2016. We used surveillance data from the area of 8 administrative districts where ASFv occurred during 2014-2015 (‘infected zone’, 2224 samples tested) to fit the model and define spatio-temporal intensity of sampling in our model.

114

115 **Process model**

116 We used a spatially-explicit, individual-based modeling framework fitted to the ASFv
117 surveillance data using Approximate Bayesian Computation (ABC) as in Pepin et al. 2020. We
118 evaluated three models that differed by social and spatial transmission process assumptions
119 (described below) in host populations that were structured by family groups and dispersal as
120 defined by field data (Fig. 1). We estimated transmission parameters and some other
121 epidemiological and demographic parameters as described below. With parameters from the
122 fitted models we then predicted cases over time, spatial spread over time, spatial transmission
123 kernels, effective reproductive numbers over time, and age- and sex-structure of infected
124 individuals. All analyses were implemented in Matlab (Version R2016b, The MathWorks, Inc.,
125 Natick, Massachusetts, United States). A full description of the individual-based model is given
126 in Pepin et al. (2020). Below is an overview of the approach with emphasis on differences from
127 our previous work.

128 We used a 5 x 5 km (25 km^2) gridded landscape to map spatial movement. The total
129 landscape size was 120 x 50 km (6000 km^2), similar to the 'infected' zone. Grid cells each had a
130 carrying capacity of 0.5 or 2 boars/km^2, which controlled heterogeneity in population density
131 across the landscape through density-dependent reproduction. In previous work we found that
132 this level of heterogeneous boar density fit the surveillance data better than homogenous
133 densities of 1, 2 or 4 wild boar / km^2 [16].

134 Individual-boar attributes were monitored and updated at a daily time step. These
135 included age, unique group identification, X and Y coordinates of the home range centroid, grid-
136 cell ID; and status of life, reproduction, disease. Thus, the distribution of wild boar locations was

continuous but density was controlled at the grid cell level. The variable attributes changed based on time, age, group size, grid-cell density, natal dispersal timing, and the disease transmission process. Attributes that were fixed at birth included sex, dispersal distance, dispersal age, and age at natural death. Thus, natal dispersal age and distance, and natural death, occurred at pre-set ages and distance (for dispersal).

Individual-boar status was updated by the following order of processes: daily movement (defined by the contact processes described below) and disease transmission, natural mortality (occurring according to the pre-set age), natal dispersal (occurring according to the pre-set age), dispersal due to other factors (i.e., family groups becoming too large, single females searching for groups; occurring based on current family group size), surveillance sampling (permanent removal of individuals from the landscape), conception (rates dependent on current grid cell density), and new births (occurring with gestating females reach the end of their gestation period). Fixed parameters included longevity (a data-based distribution), litter size (6), age at reproductive maturity (180 days), minimum time between conception and farrowing (90 days), gestation time (115 days), age of natal dispersal (~Poisson(13 months) truncated between 10-24 months), dispersal distance (~Weibull(2.5,0.5)), maximum size of family groups (10), incubation period for ASFv (~Poisson(4 days) truncated at 1), infectious period for ASFv (~Poisson(5 days) truncated at 1), and disease-induced mortality (100%). There were also fixed seasonal trends that varied monthly for conception probability and carcass persistence that were based on data [22-25]. Rationale and sources for these parameters are derived from ecological studies of wild boar and are described in Pepin et al. 2020.

Based on the attributes and processes described above, epidemiological states for individual boar included: susceptible, exposed, infectious, infectious carcass, uninfectious

carcass, and removed from the landscape. Mortality only occurred from the disease (leading to an infectious carcass) or reaching the age of longevity (leading to an uninfectious carcass). The hunting process of alive individuals caused direct removal from the landscape (no carcass). Our model also included multiple spatio-temporal scales of spatial processes because the dispersal process (~Weibull(2.5,0.5) allowed for longer-distance movements and occurred less frequently relative to the contact process that occurred daily and mostly at shorter distances.

We compared three different forms of contact structure: 1) neighborhood (local transmission only), 2) exponential decay (distance distribution that includes long-distance processes), and 3) distance distribution with social structure (Fig. 1). Forms 1 and 2 are common ways of considering contact in space at population-level scales [26], whereas 3 incorporates heterogeneity due to social groups. For 1, infectious individuals could transmit to all susceptible individuals within a fixed radius with equal probability. The radius of the local neighborhood was constant across individuals and time – thus similar to a queen's neighbor effect (Eq. 1). For 2, infectious individuals could transmit to all susceptible individuals on the landscape, but the probability of transmission decayed with distance (Eq. 2). Model structure 3 was the same as 2, except that transmission rates varied due to both group membership and space - individuals in the same family group had higher transmission rates with each other relative to those among family groups (Eq. 3). In general the daily force of infection (λ) for each contact structure was defined as follows:

$$\lambda = \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,d}\begin{cases} S_j\beta_d, & x_{k,j} < \zeta \\ 0, & otherwise \end{cases} + \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,c}\begin{cases} S_j\beta_c, & x_{k,j} < \zeta \\ 0, & otherwise \end{cases} + \beta_{0,\{j\}} \quad (1)$$

$$\lambda = \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,d} S_j \beta_d e^{-\alpha x_{k,j}} + \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,c} S_j \beta_c e^{-\alpha x_{k,j}} + \beta_{0,\{j\}} \quad (2)$$

$$\lambda = \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,d}(S_j\beta_d\, e^{-\alpha x_{k,j}} + S_{j,w}\beta_{w,d}) + \sum_{k=1}^{K}\sum_{j=1}^{J} I_{k,c}(S_j\beta_c e^{-\alpha x_{k,j}} + S_{j,w}\beta_{w,c}) + \beta_{0,\{Sj\}} \quad (3)$$

Where ζ is a fixed local neighborhood (Eq. 1) delimiting the contact radius, $x_{i,j}$ is the distance between infectious individual k (I_k) and susceptible individual j (S_j), α is the rate at which transmission decays with distance (Eq. 2 and 3), d denotes alive individuals or direct transmission, c denotes infectious carcasses or carcass-based transmission, β is the transmission rate that is specific to the transmission mechanism (d or c), $\beta_{0,\{j\}}$ is the baseline rate at which re-introduction occurs to susceptible individuals near the Eastern border ($\{Sj\}$), and family group structure (w denotes contacts within the same family group, absence of w denotes among family groups; Eq. 3).

Observation model

Because surveillance sampling was small compared to the full population it was important to calibrate the process model with an observation model. Thus, we sampled the true disease dynamics according to the surveillance process that was used in Poland, i.e., alive individuals were available to be harvested by hunters, and carcasses were available to be found for carcass sampling. As negative samples could not be georeferenced to the grid cell level (they were only available at the district level), we were not able to account for the spatial distribution

of sampling accurately. However, there were strong temporal trends in the number of samples collected thus we used those trends to describe sampling heterogeneity. First, we calculated the relative number of boar sampled by hunters and carcass-sampling from the data (number sampled on day *t*/maximum ever sampled separately for each method) to produce seasonal trends in the proportion of the population sampled. Then we multiplied the seasonal trend data for each method by the scaling factors (ρ_h and ρ_c) to determine the daily proportion of boar that would be sampled (detection probability) by hunter harvesting or dead carcasses across the landscape at random. We assumed that boar < 6 months of age would not be hunted (typically not targeted by hunters) and that boar < 3 months of age would not be sampled by the dead carcass method (because they are unlikely to be found).

Model fitting and evaluation

Unknown parameters were estimated based on Approximate Bayesian Computation (ABC) with rejection sampling as described in Pepin et al. 2020. For all models, estimated parameters included: frequency of introduction at the eastern border ($\beta_{0,\{j\}}$), β_d, β_c, scaling parameters on seasonal trends of hunted hosts (ρ_h) and carcass sampling (ρ_c), a scaling parameter on seasonal trends in the length of carcass persistence on the landscape (π), and a scaling parameter on seasonal patterns of host birth probabilities (θ). In addition, we estimated spatial parameters that describe three different contact structures: 1) ξ (nearest-neighbor), 2) α (the decay of contact probability with distance, and 3) $\beta_{w,d}$ and $\beta_{w,c}$ (direct and carcass-based transmission rates for within-group contacts). Prior distributions are listed in Table S1 (with restrictions: $\beta_d > \beta_c$, $\beta_{w,d} > \beta_d$, $\beta_{w,c} > \beta_c$) and were informed by movement and contact data [17, 18, 27, 28].

To sample across parameter space efficiently we used a Latin hypercube algorithm to generate 979,592 parameter sets and then ran the model twice on each parameter set (for a total of 1,959,184 iterations; or 2 chains of 979,592). β_d, β_c, and ρ_c were sampled on a log scale. A two-tiered approach was applied for evaluating parameter sets to improve efficiency. Simulations were terminated early if they were unrealistic, specifically: 1) when landscape-wide host density < 20% of the initial density, 2) > 150 new cases per day; 3) no new cases sampled for 6 months, or 4) > 300 total cases (more than double the actual number). We then only considered parameter sets for which the simulation reached the end of the two-year time frame. The posterior distributions consisted of all unique parameter sets (considering both chains) that were within the absolute distance of three metrics: the sum of absolute differences between observed and simulated surveillance data for monthly cases from live and dead animals (considered separately), and the maximum monthly Euclidian distance of cases from the eastern border. Distance metric tolerance values were 48 for monthly cases from carcasses, 24 for monthly cases from hunter-harvest samples, and 120 for maximum distance from the border. This allowed average error rates of 2 (carcass) and 1 (hunter harvest) cases, and 5 km from the border per month on average. These error rates represent levels of uncertainty that we expected from the data sources in our system, sensitivity analyses revealed that less stringent error rates would affect the posterior distribution estimates (data not shown), and more stringent error rates would require restrictively large computational resources unless prior distributions are more informed.

Average distance metrics for parameter sets from the posterior distribution were used to evaluate goodness of fit along with R^2 values (squared correlation of observed and predicted case and spatial distance trajectories, Table S1) and mean absolute error (MAE, Fig. 2). For each fitted model we predicted outbreak dynamics using 1000 random samples from the posterior

246 distribution. The average of the 1000 predictions was used to calculate R^2 and MAE. We also
247 tested the ability of our models to forecast ASF dynamics by using the parameters estimated
248 from fits to the 2014-2015 data to predict the first 7 months of 2016 (Jan.-Jul.). We predicted
249 underlying spatial transmission kernels, effective reproductive number over time, and age-sex
250 structure of cases by simulating from the fitted models.
251
252 **RESULTS**
253 **Parameter inference and model fit**
254 The model with both social and spatial processes (Model 3) qualitatively captured spatial
255 spread better than Models 1 (Local neighborhood only) and 2 (Distance distribution only), and
256 Model 1 largely overestimated cases during the largest peak (Fig. 2). Also, the posterior
257 distribution of transmission probabilities were much lower and more realistic for Models 2 and 3
258 relative to Model 1 (see β_d and β_c in Table S1). The inferred STK for each model revealed two
259 distinct peaks symbolic of within- and between-group transmission (Fig. 3), but predicted
260 different amounts of within group transmission when within and between-group transmission
261 probabilities were allowed to vary (Model 3). Model 3 predicted the highest proportion of within
262 group transmission (0.8), followed by model 2 (0.6) and model 1 (0.3) (Fig. 3). For both Models
263 2 and 3, between-group transmission peaked between 0.5 km and 1 km, with a peak proportion
264 of transmission events reaching 0.05 and 0.02 for Model 2 and Model 3, respectively (Fig. 3).
265 For Model 1, between-group transmission events plateaued between 1-1.5 km at a frequency of
266 0.2 before dropping rapidly to a frequency of 0 around 1.5km (Fig. 3). The realized STKs for
267 Models 2 and 3 had long tails that indicated a low frequency of long-distance pathogen dispersal
268 (Fig. 3).

Impacts of model structure on epidemiological processes

The specification of spatial and social transmission processes in the model structure resulted in different inference of R_e. Model 1 (mean: 2.5 with 95% confidence interval: [1.8, 3.2]) predicted higher average R_e over time, followed by Model 3 (1.5 [1.1-2.0]), and then Model 2 (1.1 [1.0-1.3]), including both direct and carcass-based transmission (Fig. 4). However, predictions from Model 2 suggest R_e is relatively homogenous over time, while Models 1 and 3 predicted much more variability, with R_e values reaching above a value of 4 on multiple occasions, and above a value of 8 at least once (Fig. 4). Model 3 predicted higher R_e during annual birth pulses (Fig. 4). Models 1 and 2 predicted lower contributions of carcass-based transmission in overall R_e whereas Model 3 predicted more similar levels of each transmission mechanism (with carcass-based transmission being slightly lower on average). All models predicted that the infected class is predominantly composed by juveniles (<6 months of age; Fig. S1-S3), reflecting the age-structure in the population. However, Models 1 and 2 predicted a slight male-bias in infected individuals while Model 3 predicted a slight female bias (Fig. S1-S3).

DISCUSSION

Understanding how infectious diseases spread in space and time is key for developing effective surveillance and intervention strategies [6]. Statistical inference of wildlife disease systems is often challenged by partially observed data, multiple pathogen introductions, and a limited understanding of host contact processes, all of which create uncertainty in our ability to characterize patterns of infectious disease spread. Here we demonstrate an approach for

292 estimating patterns of ASFv spread that accounts for these complexities while evaluating how
293 different assumptions about contact structure influence the inference of spatial disease dynamics.
294 Results show that both spatial and social sources of contact heterogeneity are important for
295 capturing the spatial dynamics of ASFv in wild boar of Eastern Poland, and that not accounting
296 for both processes can lead to very different inferences of a key epidemiological quantity, R_e.
297 Our results highlight the importance of considering both spatial and social processes in
298 estimating STKs. There are numerous approaches for modeling contact heterogeneities due to
299 spatial [2, 26, 29, 30] and social processes [10]. Network models are a useful approach for
300 accounting for contact heterogeneity but these are often focused only on heterogeneities due to
301 social structure [10], while neglecting spatial processes. Including both of these processes in a
302 single network makes it difficult to disentangle the role of each process, which can be important
303 for designing optimal control strategies. On the flip side, the field of movement ecology has
304 developed new strategies for accounting for heterogeneities due to animal movement [11, 12],
305 but these methods remain underdeveloped for application in disease ecology, and also have not
306 been used to disentangle spatial and social processes. In order to provide practical inference for
307 control of important animal diseases it will be important to develop methods that can account for
308 heterogeneities due to both social and spatial processes, in a manner that the role of each process
309 can be inferred. This will allow control to be targeted to the appropriate process while improving
310 predictions of spatial spread for optimizing risk-based surveillance and control.
311 Our models suggested that between 30-80% of transmission events were within the same
312 family group and almost all transmission events were within 1.5 km, with some rare events at
313 longer distances. STK estimates can be used to establish control and surveillance zones. For
314 example, depopulation efforts could be intensified (or abandoned to allow natural fade out)

315 within 1.5 km where most transmission is occurring, with surveillance intensified out to further
316 distances (i.e., the tail of the STK) to capture the rare long-distance spreading events. A useful
317 approach could be to employ an adaptive radius that focuses intervention efforts within 99% (or
318 more – this should be validated with modeling) of the STK, but adapts surveillance based on
319 real-time surveillance. However, the precise recommendations will depend on how soon a
320 detection is made relative to where the infection front is currently. With ASFv travelling at 1-2
321 km per month [18], the radii for high-intensity culling and surveillance would need to be
322 increased by 1-2 km for each month that detection has lagged behind the infection front,
323 highlighting the importance of accurate predictions of spatial spread. A longer lag time for
324 detection will also amplify challenges that arise from long-distance jumps highlighting that this
325 process is especially important to understand.

326 Long-distance jumps were observed on multiple occasions in Poland after 2015 [19] and are
327 thought to be due to human-mediated activities, [31, 32]. Thus, developing estimates of STKs
328 that account for mechanisms of long-distance dispersal will be important for appropriately
329 targeting disease control going forward. Our approach allowed for some long-distance events but
330 we assumed a monotonic functional form for contact distances, such that we did not infer the
331 effects of spatial contact processes occurring on multiple spatial scales. In order to infer spatial
332 spread with later surveillance data (i.e., 2016-present when longer-distance events occurred
333 multiple times), it will be important to incorporate other spatial mechanisms in the inference of
334 the STK, perhaps using covariate data that can inform these long-distance processes. Such an
335 approach would provide refined recommendations for surveillance and control targets at longer
336 distances.

Our estimates of R_e (ranging from 1.1-2.5 on average across models) were similar to an estimate of R_0 of ASFv in wild boar (1.13-3.77) in Russia [33]. However, although our estimates were similar in magnitude, they revealed substantial temporal variation in R_e that could impact policy decisions. For example, all models estimated higher R_e during seasonal birth pulses indicating that at this time of year we may expect to see higher rates of spatial spread. Additionally, estimates of the R_e over time suggested a different role for carcass-based transmission in driving ASF transmission, with Models 1 and 2 suggesting that carcass-based transmission is lower than direct transmission while Model 3 suggested that the two types of transmission occurred at similar frequencies.

In addition to differences in R_e, the different contact structures predict differences in the role of sex in transmission over time. Model 3 predicted a slight female bias because this model predicted that 80% of transmission events occurred within family groups, which are female-biased. In contrast, Models 1 and 2 predicted a slight male bias because they predicted more between-group transmission and with a 50:50 sex ratio there are more independent males relative to family groups. Although it is known that males will travel longer distances than females [34], especially during mating season to seek out females, we did not account for this temporal heterogeneity in dispersal. Considering these types of movement heterogeneities in future work could be important for improving our understanding of which sex might present a higher risk of ASFv transmission and persistence.

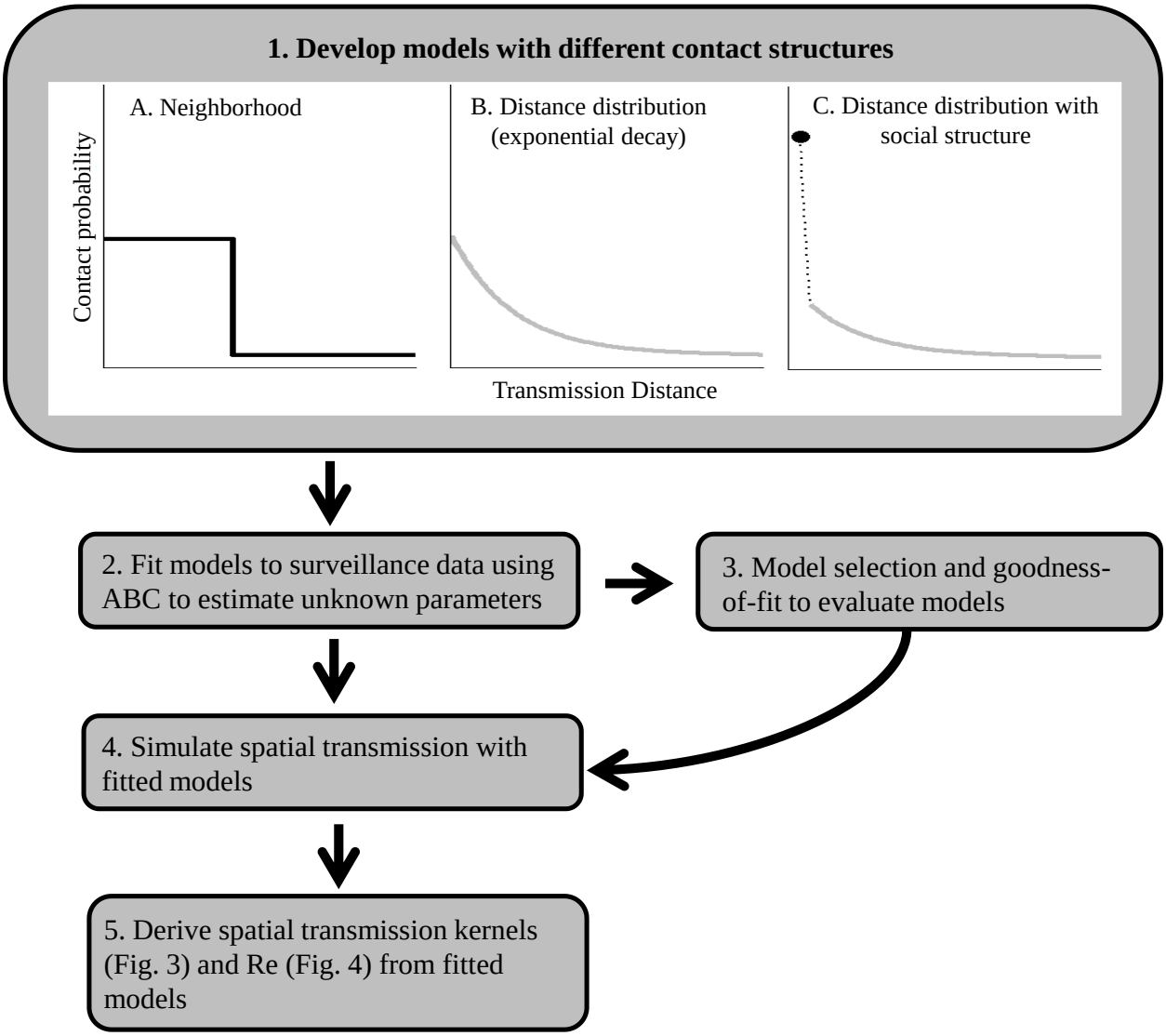
Considered separately, spatial and social processes can have similar impacts on disease dynamics. For example, social aggregation and spatial structuring can both reduce epidemic potential by fragmented populations or by restricting the spread of pathogens [10, 35]. However, our results highlight that spatial and social processes can also have quite different impacts on

epidemiological quantities, especially estimates of STKs, R_e, the frequency of different transmission mechanisms, and potential risk factors such as sex. Being able to appropriately infer the role of these quantities is crucial for optimizing disease control strategies. When these contact heterogeneities are inappropriately accounted for it can bias inference (e.g., [36]) and potentially misguide policy decisions [9]. Moving forward, the field of disease ecology needs to develop mainstream methods that account for multiple sources of contact heterogeneities in a manner that their relative roles can be inferred. This will allow uncertainties in contact processes to be appropriately evaluated and incorporated into predictions of spatial spread [9] and control to be targeted to the most important risk factors.

Acknowledgements

KMP was supported by the United States Department of Agriculture, Animal and Plant Health Inspection Service's National Feral Swine Damage Management Program. AG was supported by the United States Department of Agriculture, Animal and Plant Health Inspection Service's APHIS Science Fellowship. TP was supported by the National Science Centre, Poland (grant number 2014/15/B/NZ9/01933). We thank M. Łyjak, A. Kowalczyk, K. Śmietanka, and G. Woźniakowski from Department of Swine Diseases, National Veterinary Research Institute in Pulawy, Poland, for surveillance data. N. Selva provided valuable information on carcass persistence time.

382 **Figures**



383

384

385 **Fig 1.** Schematic of methods and contact structures.

386

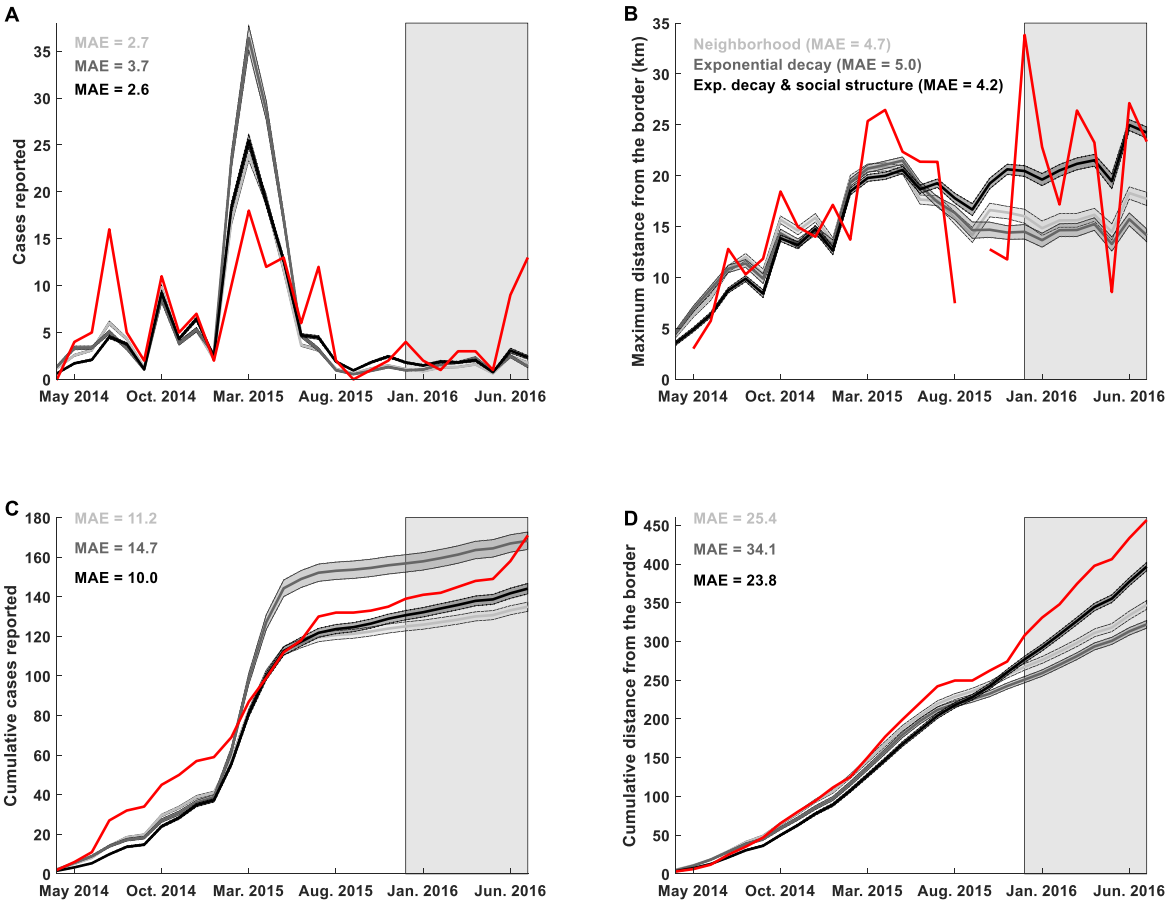


Fig. 2. Model fits to the observed surveillance data (red). Lines are the mean predictions from 1000 simulations with each fitted model (see legend in B for color code). Shading around the lines areas are 95% prediction intervals of the means. Shading on the right side of plots indicates the time frame for out-of-sample predictions.

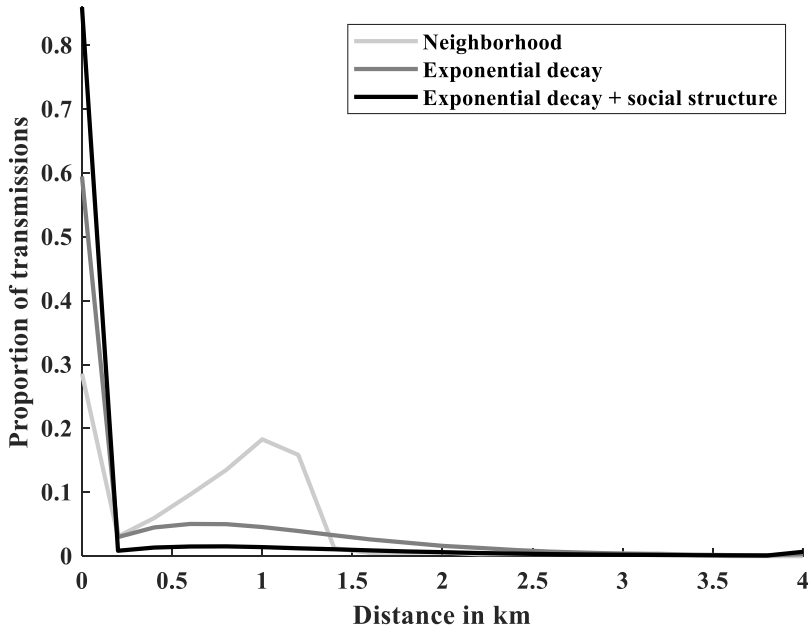
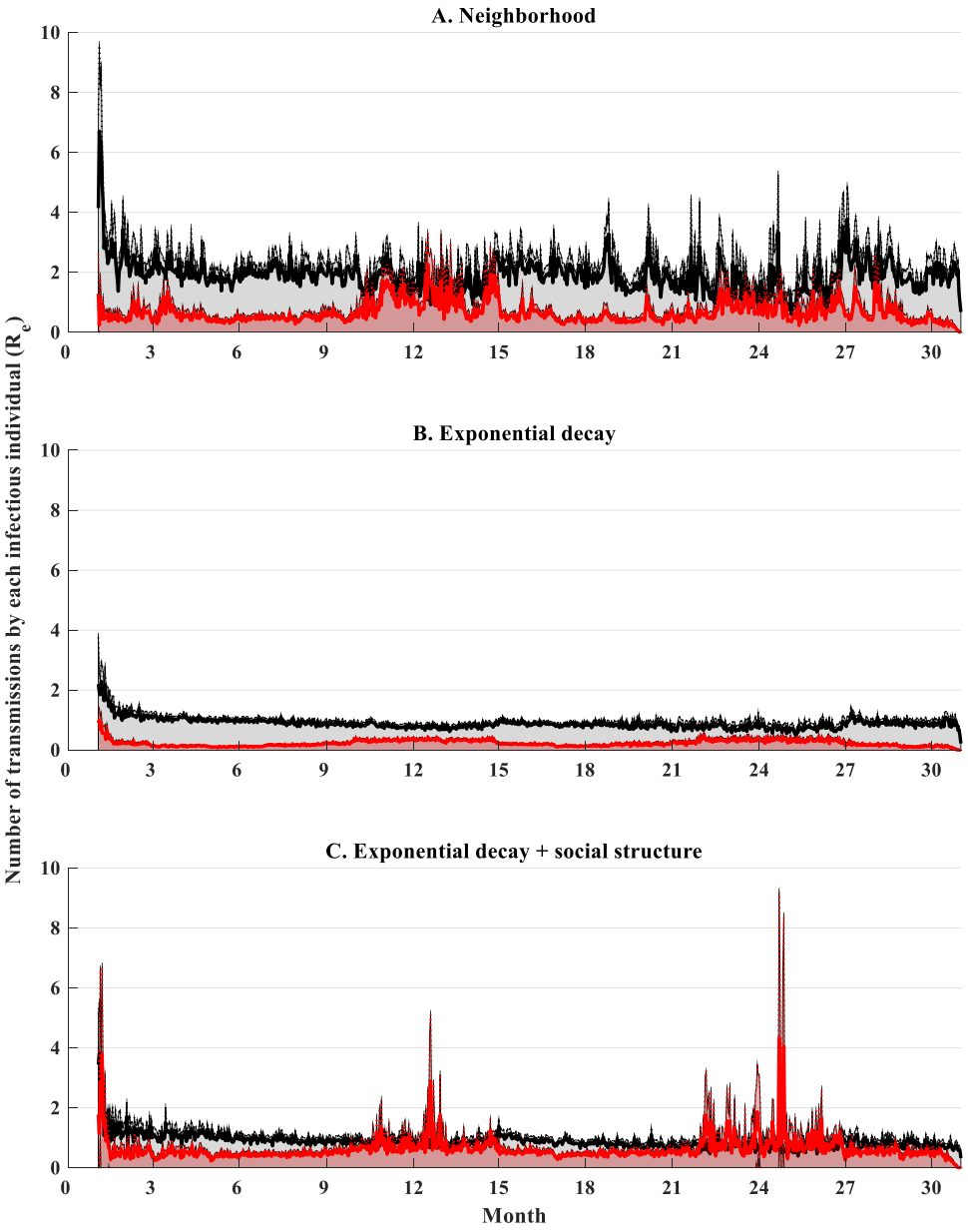


Fig. 3. Realized spatial transmission kernels for each model (see legend). The X-axis is the distance between home range centroids of infectious and susceptible individuals for which transmission occurred. Y-axis is the proportion of all transmission events. Lines are the means of 100 simulations using random samples from the posterior distributions of the fitted models.



401

402 **Fig. 4.** Effective reproduction number (R_e) over time for each model. Effective reproduction
403 number at a given time point was calculated as the average number of transmissions made
404 throughout the infectious period for individuals that initially became infectious on day *t* (where *t*
405 is a day on the X-axis). Dark lines are the means of 100 simulations using random samples from
406 the posterior distributions of the fitted models; shading indicates 95% prediction intervals of the
407 means. Overall means with 95% prediction intervals for each model and each transmission

mechanism (black: direct, red: carcass-based) were: A) 1.9 [1.4, 2.3], 0.6 [0.4, 0.9], B) 0.9 [0.8 1.0], 0.2 [0.2, 0.3], C) 0.9 [0.7, 1.1], 0.6 [0.4, 0.9].

References

[1] Salje, H, Cummings, DA & Lessler, J. 2016 Estimating infectious disease transmission distances using the overall distribution of cases. *Epidemics* **17**, 10-18.

[2] Riley, S. 2007 Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298-1301.

[3] Pepin, KM, Davis, AJ, Streicker, DG, Fischer, JW, VerCauteren, KC & Gilbert, AT. 2017 Predicting spatial spread of rabies in skunk populations using surveillance data reported by the public. *PLoS neglected tropical diseases* **11**, e0005822.

[4] Ferguson, NM, Donnelly, CA & Anderson, RM. 2001 The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155-1160.

[5] Ferguson, NM, Donnelly, CA & Anderson, RM. 2001 Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542-548.

[6] Keeling, MJ, Woolhouse, ME, Shaw, DJ, Matthews, L, Chase-Topping, M, Haydon, DT, Cornell, SJ, Kappey, J, Wilesmith, J & Grenfell, BT. 2001 Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813-817.

[7] Woolhouse, M. 2011 How to make predictions about future infectious disease risks. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2045-2054.

[8] Chipperfield, JD, Holland, EP, Dytham, C, Thomas, CD & Hovestadt, T. 2011 On the approximation of continuous dispersal kernels in discrete-space models. *Methods in Ecology and Evolution* **2**, 668-681.

[9] Shea, K, Tildesley, MJ, Runge, MC, Fonnesbeck, CJ & Ferrari, MJ. 2014 Adaptive management and the value of information: learning via intervention in epidemiology. *PLoS biology* **12**, e1001970.

436 [10] Sah, P, Mann, J & Bansal, S. 2018 Disease implications of animal social network structure:
437 a synthesis across social systems. *Journal of Animal Ecology* **87**, 546-558.
438 [11] Dougherty, ER, Seidel, DP, Carlson, CJ, Spiegel, O & Getz, WM. 2018 Going through the
439 motions: incorporating movement analyses into disease research. *Ecology letters* **21**, 588-604.
440 [12] Emch, M, Root, ED, Giebultowicz, S, Ali, M, Perez-Heydrich, C & Yunus, M. 2012
441 Integration of spatial and social network analysis in disease transmission studies. *Annals of the*
442 *Association of American Geographers* **102**, 1004-1015.
443 [13] Chapman, DA, Darby, AC, Da Silva, M, Upton, C, Radford, AD & Dixon, LK. 2011
444 Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus.
445 *Emerging infectious diseases* **17**, 599.
446 [14] EFSA, EFS, Miteva, A, Papanikolaou, A, Gogin, A, Boklund, A, Bøtner, A, Linden, A,
447 Viltrop, A, Schmidt, CG & Ivanciu, C. 2020 Epidemiological analyses of African swine fever in
448 the European Union (November 2018 to October 2019). *EFSA Journal* **18**, e05996.
449 [15] Arias, M, Jurado, C, Gallardo, C, Fernández-Pinero, J & Sánchez-Vizcaíno, J. 2018 Gaps in
450 African swine fever: analysis and priorities. *Transboundary and emerging diseases* **65**, 235-247.
451 [16] Pepin, KM, Golnar, A, Abdo, Z & Podgorski, T. 2020 Ecological drivers of African swine
452 fever virus persistence in wild boar populations: insight for control. *Ecol Evolv.* **10**, 2846–2859.
453 (doi:10.1002/ece3.6100).
454 [17] Pepin, KM, Davis, AJ, Beasley, J, Boughton, R, Campbell, T, Cooper, SM, Gaston, W,
455 Hartley, S, Kilgo, JC & Wisely, SM. 2016 Contact heterogeneities in feral swine: implications
456 for disease management and future research. *Ecosphere* **7**, e01230.
457 [18] Podgórski, T, Apollonio, M & Keuling, O. 2018 Contact rates in wild boar populations:
458 Implications for disease transmission. *The Journal of Wildlife Management* **82**, 1210-1218.

459 [19] Mazur-Panasiuk, N & Woźniakowski, G. 2019 The unique genetic variation within the
460 O174L gene of Polish strains of African swine fever virus facilitates tracking virus origin.
461 *Archives of virology* **164**, 1667-1672.
462 [20] Śmietanka, K, Woźniakowski, G, Kozak, E, Niemczuk, K, Frączyk, M, Bocian, Ł,
463 Kowalczyk, A & Pejsak, Z. 2016 African swine fever epidemic, Poland, 2014–2015. *Emerging*
464 *infectious diseases* **22**, 1201.
465 [21] Woźniakowski, G, Kozak, E, Kowalczyk, A, Łyjak, M, Pomorska-Mól, M, Niemczuk, K &
466 Pejsak, Z. 2016 Current status of African swine fever virus in a population of wild boar in
467 eastern Poland (2014-2015). *Archives of virology* **161**, 189-195.
468 [22] Rosell, C, Navas, F & Romero, S. 2012 Reproduction of wild boar in a cropland and coastal
469 wetland area: implications for management. *Animal Biodiversity and Conservation* **35**, 209-217.
470 [23] Selva, N, Jędrzejewska, B, Jędrzejewski, W & Wajrak, A. 2005 Factors affecting carcass
471 use by a guild of scavengers in European temperate woodland. *Canadian Journal of Zoology* **83**,
472 1590-1601.
473 [24] Ježek, M, Štípek, K, Kušta, T, Červený, J & Vícha, J. 2011 Reproductive and morphometric
474 characteristics of wild boar (Sus scrofa) in the Czech Republic. *Journal of forest science* **57**, 285-
475 292.
476 [25] Bieber, C & Ruf, T. 2005 Population dynamics in wild boar Sus scrofa: ecology, elasticity
477 of growth rate and implications for the management of pulsed resource consumers. *Journal of*
478 *Applied Ecology* **42**, 1203-1213.
479 [26] Keeling, MJ & Rohani, P. 2008 *Modeling infectious diseases in humans and animals*. NJ,
480 USA, Princeton University Press.

481 [27] Podgórski, T, Baś, G, Jędrzejewska, B, Sönnichsen, L, Śnieżko, S, Jędrzejewski, W &
482 Okarma, H. 2013 Spatiotemporal behavioral plasticity of wild boar (Sus scrofa) under
483 contrasting conditions of human pressure: primeval forest and metropolitan area. *Journal of*
484 *Mammalogy* **94**, 109-119.
485 [28] Kay, SL, Fischer, JW, Monaghan, AJ, Beasley, JC, Boughton, R, Campbell, TA, Cooper,
486 SM, Ditchkoff, SS, Hartley, SB & Kilgo, JC. 2017 Quantifying drivers of wild pig movement
487 across multiple spatial and temporal scales. *Movement ecology* **5**, 14.
488 [29] White, LA, Forester, JD & Craft, ME. 2018 Dynamic, spatial models of parasite
489 transmission in wildlife: Their structure, applications and remaining challenges. *journal of*
490 *animal ecology* **87**, 559-580.
491 [30] White, LA, Forester, JD & Craft, ME. 2018 Disease outbreak thresholds emerge from
492 interactions between movement behavior, landscape structure, and epidemiology. *Proceedings of*
493 *the National Academy of Sciences* **115**, 7374-7379.
494 [31] Chenais, E, Depner, K, Guberti, V, Dietze, K, Viltrop, A & Ståhl, K. 2019 Epidemiological
495 considerations on African swine fever in Europe 2014–2018. *Porcine health management* **5**, 1-6.
496 [32] EFSA, EFS, Depner, K, Gortazar, C, Guberti, V, Masiulis, M, More, S, Oļševskis, E,
497 Thulke, H, Viltrop, A & Woźniakowski, G. 2017 Epidemiological analyses of African swine
498 fever in the Baltic States and Poland: (Update September 2016–September 2017). *EFSA Journal*
499 **15**, e05068.
500 [33] Iglesias, I, Muñoz, M, Montes, F, Perez, A, Gogin, A, Kolbasov, D & De la Torre, A. 2016
501 Reproductive ratio for the local spread of African swine fever in wild boars in the Russian
502 Federation. *Transboundary and emerging diseases* **63**, e237-e245.

503 [34] Podgórski, T, Scandura, M & Jędrzejewska, B. 2014 Next of kin next door–philopatry and
504 socio-genetic population structure in wild boar. *Journal of Zoology* **294**, 190-197.
505 [35] Keeling, MJ. 1999 The effects of local spatial structure on epidemiological invasions.
506 *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 859-867.
507 [36] Ames, GM, George, DB, Hampson, CP, Kanarek, AR, McBee, CD, Lockwood, DR, Achter,
508 JD & Webb, CT. 2011 Using network properties to predict disease dynamics on human contact
509 networks. *P Roy Soc B-Biol Sci* **278**, 3544-3550. (doi:10.1098/rspb.2011.0290).

510

511

Supplemental Material

Social structure defines spatial transmission of African swine fever in wild boar

Kim M. Pepin[1,*], Andrew Golnar[1], and Tomasz Podgórski[2,3]

4. National Wildlife Research Center, USDA, APHIS, Wildlife Services, 4101 Laporte Ave., Fort Collins, CO, 80526.
5. Mammal Research Institute, Polish Academy of Sciences, Stoczek 1, 17-230 Białowieża, Poland
6. Department of Game Management and Wildlife Biology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Kamýcká 129, 165 00 Praha 6, Czech Republic

*Corresponding author: Kim Pepin, National Wildlife Research Center, United States Department of Agriculture, 4101 Laporte Ave., Fort Collins, CO, 80526; kim.m.pepin@aphis.usda.gov

Contents

Table S1	**Contact structure:**	**Neighborhood**	**Exp. Decay (ED)**	**ED & social structure**
	R^2			
Monthly cases (Median of the R^2's for 1000 individual time series ± 95% confidence interval)	in sample	0.57 ± 0.0059	0.51 ± 0.0054	0.55 ± 0.0060
	all	0.48 ± 0.0054	0.43 ± 0.0046	0.47 ± 0.0049
Monthly distance from border (Median of the R^2's for 1000 individual time series ± 95% confidence interval)	in sample	0.28 ± 0.018	0.29 ± 0.017	0.31 ± 0.018
	all	0.19 ± 0.016	0.14 ± 0.011	0.28 ± 0.011
Monthly cases (R^2 of median values from 1000 points at each month)	in sample	0.65	0.57	0.63
	all	0.55	0.48	0.53
Monthly distance from border (R^2 of median values from 1000 points at each month)	in sample	0.53	0.49	0.53
	all	0.49	0.38	0.55
	Distance metrics[a]			
Median absolute error in monthly live cases ± 95% confidence interval	in sample	20 ± 0.8	20 ± 0.9	20 ± 0.5
Median absolute error in monthly carcass cases ± 95% confidence interval	in sample	44 ± 1.4	43 ± 2.8	44 ± 0.8
Median absolute error in monthly distance from border ± 95% confidence interval	in sample	109.5 ± 5.0	111.7 ± 5.7	106.6 ± 3.0
Number of values in posterior distribution		16 / 1,959,184 = 0.00082%	7 / 1,959,184 = 0.00036%	53 / 1,959,184 = 0.0027%
	Uniform Priors	95% credible intervals of posterior distributions		
Probability of direct transmission given proximity (β_d)	[0.0001,1][b]	[0.0046, 1.0]	[0.039, 0.15]	[0.0003, 0.18]
Probability of carcass-based transmission given proximity (β_c)	[0.0001,1][b]	[0.0002, 0.53]	[0.0038, 0.029]	[0.0001, 0.029]
Annual frequency of spillover from neighboring country ($\beta_{0,\{\varphi\}}$)	[0,60][c]	[13, 59]	[9, 60]	[6, 59]
Detection probability in hunted boar samples (ρ_h)	[0.0005,0.1][b]	[0.0006, 0.012]	[0.0007, 0.021]	[0.0007, 0.017]
Detection probability in carcass samples (ρ_c)	[0.0005,0.8][b]	[0.012, 0.080]	[0.017, 0.045]	[0.014, 0.30]
Scaling parameter on seasonal trends in carcass persistence (π)	[0.1,1.5]	[0.58, 1.47]	[0.77, 1.41]	[0.17, 1.47]
Scaling parameter on seasonal trends in conception probability (θ)	[0.5,6]	[0.93, 5.73]	[1.31, 5.87]	[0.67, 5.88]
Constant contact radius (ζ)	[0.5,5]	[1.09, 3.35]	NA	NA
Rate parameter for decay of contact probability with distance (α)	[0.1,2.5]	NA	[1.62, 2.49]	[0.05, 2.49]
Probability of direct transmission given contact is in the same group ($\beta_{w,d}$)	[0.01,1]	NA	NA	[0.077, 0.98]
Probability of carcass-based transmission given contact is in the same group ($\beta_{w,c}$)	[0.001,1]	NA	NA	[0.14, 0.97]

[a]Median distance metrics ± 95% confidence intervals for 1000 simulations from the posterior distribution.

[b]These prior distributions were sampled on a natural log scale.

[c]0 indicates one introduction ever whereas values > 0 indicate the number of introductions / year.

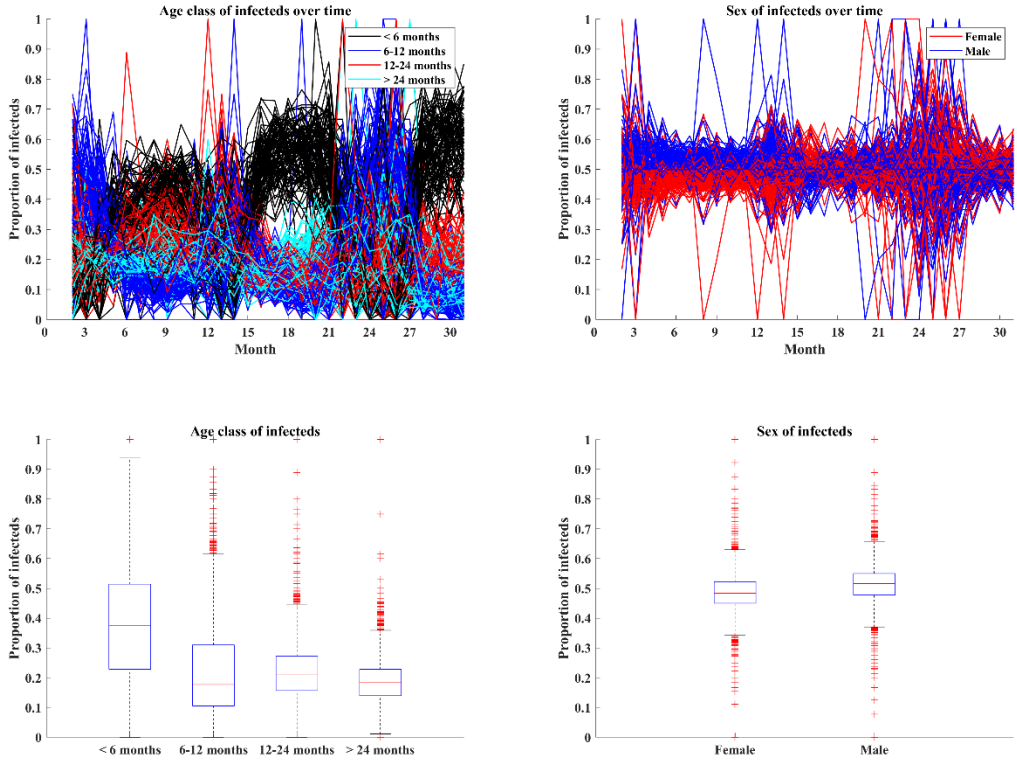


Figure S1. The demographic dynamics of African swine fever virus based on Model 1 output (Neighborhood). Top plots show the frequency of infection in wild boar of different age classes (left) and sexes (right). Each line is a prediction from a separate sample of the posterior distribution of the fitted exponential decay & social structure model (100 trajectories in total). Bottom plots show the corresponding age class and sex distributions of infection over all time.

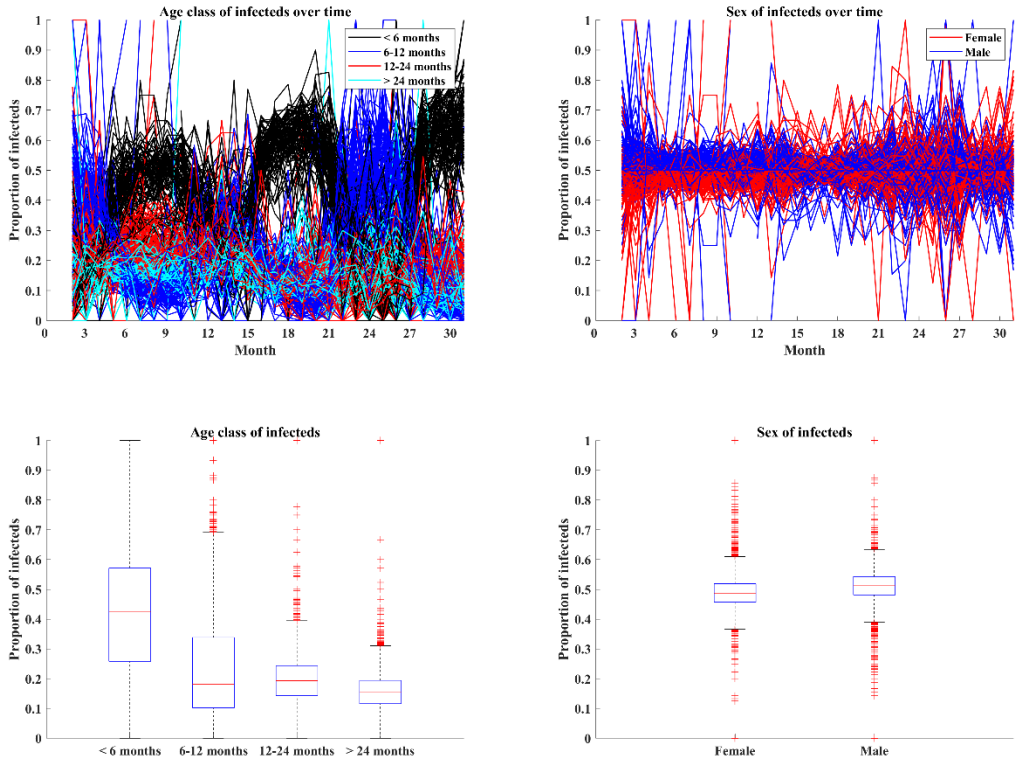
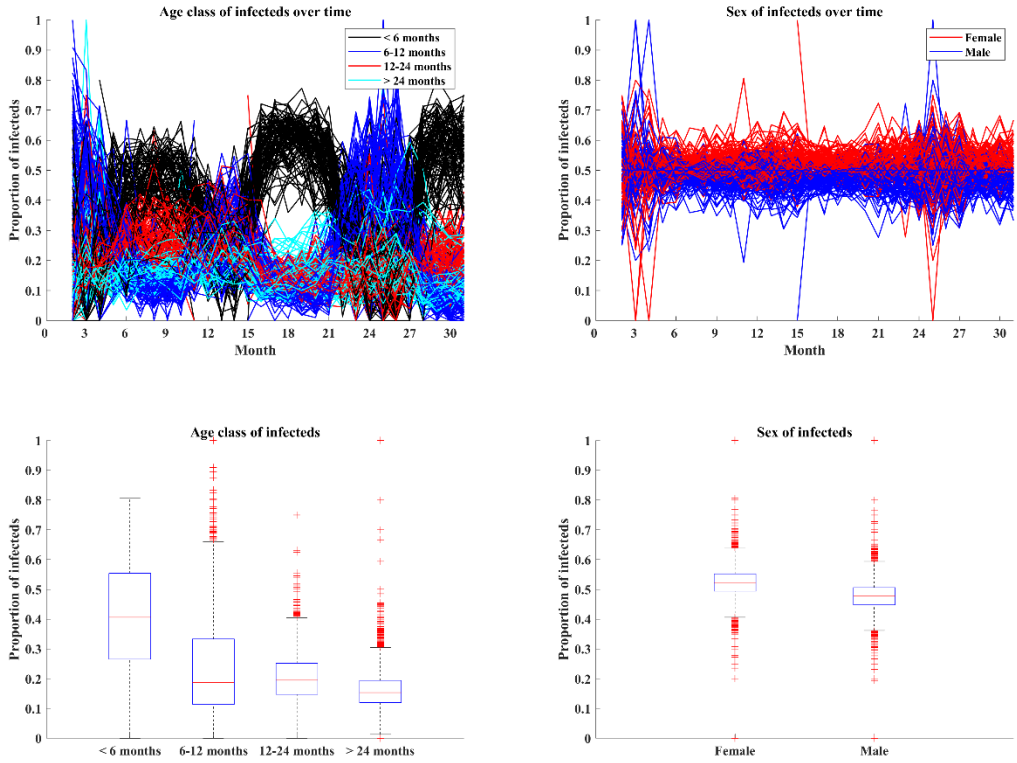


Figure S2. The demographic dynamics of African swine fever virus based on Model 2 output (exponential decay). Top plots show the frequency of infection in wild boar of different age classes (left) and sexes (right). Each line is a prediction from a separate sample of the posterior distribution of the fitted exponential decay & social structure model (100 trajectories in total). Bottom plots show the corresponding age class and sex distributions of infection over all time.



566

567 **Figure S3. The demographic dynamics of African swine fever virus based on Model 3**
568 **output (social and exponential).** Top plots show the frequency of infection in wild boar of
569 different age classes (left) and sexes (right). Each line is a prediction from a separate sample of
570 the posterior distribution of the fitted exponential decay & social structure model (100
571 trajectories in total). Bottom plots show the corresponding age class and sex distributions of
572 infection over all time.

573

574