

RESEARCH ARTICLE

Open Access



Data-driven network modelling of disease transmission using complete population movement data: spread of VTEC O157 in Swedish cattle

Stefan Widgren^{1,2*} , Stefan Engblom³, Pavol Bauer³, Jenny Frössling^{2,4}, Ulf Emanuelson¹ and Ann Lindberg²

Abstract

European Union legislation requires member states to keep national databases of all bovine animals. This allows for disease spread models that includes the time-varying contact network and population demographic. However, performing data-driven simulations with a high degree of detail are computationally challenging. We have developed an efficient and flexible discrete-event simulator *SimInf* for stochastic disease spread modelling that divides work among multiple processors to accelerate the computations. The model integrates disease dynamics as continuous-time Markov chains and livestock data as events. In this study, all Swedish livestock data (births, movements and slaughter) from July 1st 2005 to December 31st 2013 were included in the simulations. Verotoxigenic *Escherichia coli* O157:H7 (VTEC O157) are capable of causing serious illness in humans. Cattle are considered to be the main reservoir of the bacteria. A better understanding of the epidemiology in the cattle population is necessary to be able to design and deploy targeted measures to reduce the VTEC O157 prevalence and, subsequently, human exposure. To explore the spread of VTEC O157 in the entire Swedish cattle population during the period under study, a within- and between-herd disease spread model was used. Real livestock data was incorporated to model demographics of the population. Cattle were moved between herds according to real movement data. The results showed that the spatial pattern in prevalence may be due to regional differences in livestock movements. However, the movements, births and slaughter of cattle could not explain the temporal pattern of VTEC O157 prevalence in cattle, despite their inherently distinct seasonality.

Introduction

European Union legislation requires member states to keep a registry of all bovine animals in national databases [1, 2]. The registry must contain the location and date of birth of each animal, the date and source and destination holding when an animal is moved and date of death or slaughter of an animal [1, 2]. The use of real livestock data allows for disease spread models with data-driven introduction of population demographics and the time-varying contact network. Mathematical models and computer

simulations can be used to study the spread of infectious diseases and to evaluate intervention strategies [3–5].

Performing detailed data-driven stochastic simulation is computationally challenging and requires efficient algorithms. Moreover, both model selection and parameter inference are challenging when exploiting rich livestock data in infectious diseases modelling [6]. Computational tools need improvements to allow network models to include epidemiologically relevant data [7]. We recently presented an efficient computational and flexible modelling framework for fast event-based epidemiological simulations of infectious diseases [8, 9]. The framework integrates within-herd infection dynamics as continuous-time Markov chains and livestock data as scheduled events.

*Correspondence: stefan.widgren@sva.se

² Department of Disease Control and Epidemiology, National Veterinary Institute, 751 89 Uppsala, Sweden

Full list of author information is available at the end of the article

Verotoxigenic *Escherichia coli* O157:H7 (VTEC O157) is a zoonotic bacterial pathogen infecting through the faecal-oral route. Infected humans, notably children, often develop bloody diarrhoea [10, 11]. Moreover, a severe complication, the haemolytic-uremic syndrome (HUS) is observed in about 10% of the cases [12, 13]. Cattle are considered to be the main reservoir of the pathogen [14]. Infected cattle excrete the bacteria in their faeces, which can contaminate hides, the environment, water and subsequently food and recreational areas [15]. Implementing targeted intervention strategies to reduce the incidence and prevalence of VTEC O157 infections in the cattle population could potentially reduce the number of human cases.

Identifying risk factors is the basis for understanding how to prevent and control disease. For VTEC O157 infection in cattle, several risk factors are known and may be attributed to (i) individual factors, for example age [16], (ii) herd level factors, such as animal group size [17], introduction of animals [18], and (iii) external factors, like season [19–21], or presence of an VTEC O157 positive farm in the proximity [22]. However, a better understanding of the interaction between the various risk factors is necessary to design efficient intervention strategies for VTEC O157.

The aims of the current study were to: (i) incorporate real livestock data in VTEC O157 spread simulations, (ii) estimate parameters in a VTEC O157 spread model, and (iii) explore the spatio-temporal spread of VTEC O157 on a national scale, when real livestock data are incorporated in the simulations.

Materials and methods

Disease spread model

The VTEC O157 infection dynamics was modelled in each holding with a stochastic within-holding model, coupled to other holdings through animal movements. The within-holding spread model was a SIS_E compartment model with the two disease states: susceptible (S) and infected (I) and E representing the environmental compartment contaminated with VTEC O157 by infected animals. We assumed that susceptible animals may become infected indirectly through contact with pathogens in the environment and that infected animals recover and return to the susceptible state. To capture age related differences in the infection dynamics within the host [16, 23] and in the likelihood of being moved [24], the two disease states were further subdivided into three age categories indexed with j ; calves 0–119 days, young stock 120–364 days and adults older than 364 days. The specific cut points for the age categories was chosen to match the age categories in the longitudinal observational study [25] that was used for the

parameter estimation of the model (see below). The six disease compartments and the environmental compartment within each holding i was represented by $S_{i,j}$, $I_{i,j}$ and E_i (Figure 1). The state transitions between the susceptible and infected compartments within each holding were modelled as a continuous-time discrete-state Markov process with the Gillespie's Direct Method [26, 27]. We used the implementation of the method as described in [8].

The environmental compartment E_i was modelled as a time dependent environmental infectious pressure $\varphi_i(t)$ within each holding i . The infectious pressure $\varphi_i(t)$ was assumed to be uniformly distributed within each holding and to depend on the amount of pathogens shed by infected animals. The constant α was the average shedding rate per day per infected individual that contributed to the environmental infectious pressure. For simplicity, and in absence of more detailed information, the floor surface area was assumed to be proportional to the number of individuals in each holding. We let $\beta(t)$ capture the rate per day of the bacterial decay and therefore reduction in the environmental infectious pressure $\varphi_i(t)$. We have also chosen to include a small background infectious pressure ε to allow for other indirect sources of environmental contamination (e.g. birds, rodents). The differential equation for the environmental infectious pressure in each holding was

$$\frac{d\varphi_i}{dt} = \frac{\alpha \sum_j I_{i,j}(t)}{N_i(t)} - \beta(t)\varphi_i(t) + \varepsilon \quad (1)$$

where $N_i(t)$ is the total number of animals in holding i at time t . State transitions from the susceptible to the infected state depend on the age dependent indirect transmission rate v_j [28, 29] and exposure to the environmental infectious pressure $\varphi_i(t)$, i.e. the susceptible individual's response to the environmental infectious pressure,

$$S_{i,j} \xrightarrow{v_j \varphi_i} I_{i,j}. \quad (2)$$

Infected individuals return to the susceptible compartment after the infection ceases. The expected time an infected individual stays in the infected state before returning to the susceptible state again depends on the age dependent recovery rate γ_j

$$I_{i,j} \xrightarrow{\gamma_j} S_{i,j}. \quad (3)$$

Specification of events

The following four types of events were defined; enter, internal transfer, external transfer and exit. The enter event handles births and imports from abroad. The internal transfer event happens the day an animal changes

age category from calf to young stock or young stock to adult. The external transfer event occurs when an animal moves from one holding to another. The exit event implies slaughter, euthanasia or export of the animal to another country. From that day, the animal will no longer be included in the simulation. The scheduled events are executed when the simulation in continuous time reaches the time for any of the events. The individuals are randomly sampled from the compartments affected by the event. For example, for an external transfer event of n calves from holding 1 to holding 2, n calves are randomly selected from all susceptible and infected calves in holding 1 and placed in the same compartments in holding 2.

Individuals entering the model, i.e. born or imported, are assumed to be susceptible in their respective age category. Since the aim was to explore spread in Sweden and not introduction from abroad, imported individuals were assumed to be susceptible. On average 16 cattle (range 0–45) were imported per year during the study period according to information from TRACES (the Trade Control and Expert System of the European Commission). When an individual changes age category from calf to

young stock or young stock to adult, it remains in its current disease state. Moved animals will keep the same disease state in the new holding as in the previous holding. A flow diagram of the state transitions of the described model is shown in Figure 1.

Input data

The present study was based on all reports to the national cattle database at the Swedish Board of Agriculture covering the period from 2005-07-01 to 2013-12-31. The data contained a total of 18 649 921 reports with information about the identifier of the reporting holding, the animal identification and birth date, and the date of the report [30]. If the report concerned a movement, then there should be one report from both the sending and receiving holding [30]. Each unique holding identifier in our data corresponds to a single geographical location where animals are kept, and could e.g. correspond to a farm building or pasture. Hereafter these are jointly referred to as a holding. A holding was considered as “active” if one or more animals were registered at it.

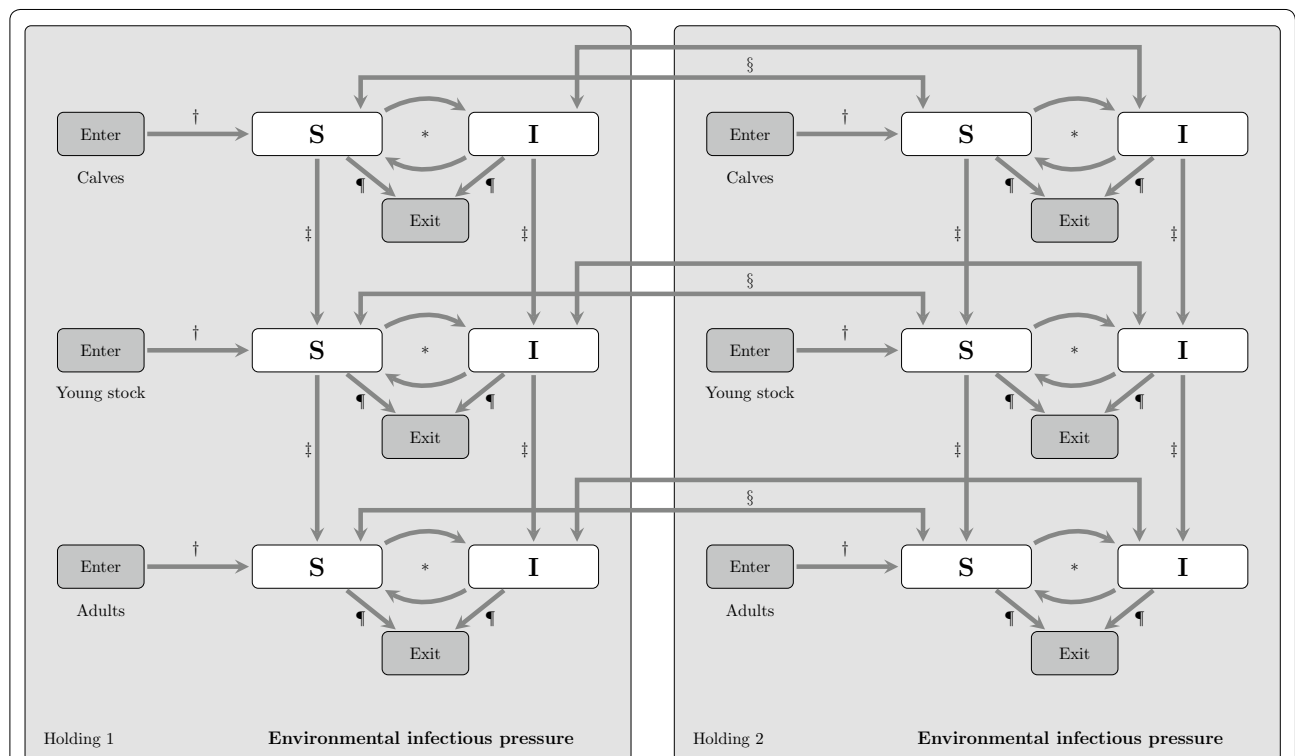


Figure 1 Conceptual SIS_E compartment model for VTEC O157 in cattle. Schematic representation of a Verotoxigenic *Escherichia coli* O157:H7 disease spread model in cattle with indirect transmission via the environment and with animal movements between holdings. The model is a SIS_E compartment model with the environmental infectious pressure compartment (E) and the two disease states susceptible (S) and infected (I). The population is divided in three age categories. *State transitions between the S and I disease states are modelled as a continuous-time discrete state Markov process (Gillespie’s direct method). The other arrows represent state transitions due to scheduled events from livestock data: (†) enter, (‡) ageing, (S) movement between holdings, and (¶) exit.

The raw data were processed to generate events for the simulation as follows. First each animal was checked for a valid and unique birth date. Each animal was then followed from birth through all reports, where each report was classified into one of the four event categories. Based on the animal birth date, internal transfer events, i.e. moving from calf to young stock or from young stock to adult, were inserted as the simulation reached the relevant time. For movement reports with conflicting dates of when the movement occurred, the animal was considered moved to the next location at the first reported date. For movements with non-conflicting reports or only one report from either the sending or receiving holding, the animal was moved at the reported date. Finally, all individual animal events were aggregated by holding, day and age category. The final data set used for the simulation contained 37 221 unique holdings and the following number of events and the mean number of individuals affected by each event: enter ($n = 3\,479\,000$, mean = 1.3), internal transfer ($n = 6\,593\,921$, mean = 1.2), external transfer ($n = 732\,292$, mean = 4.2) and exit ($n = 1\,438\,506$, mean = 3.2).

To enable spatial analysis, holdings were classified according to their NUTS level 2 region (Nomenclature of territorial units for statistics) [31] based on their coordinates. There are eight NUTS level 2 regions in Sweden, see Figure 6 for the location of each region. Exact coordinates were found for 83.8% ($n = 31\,187$) of the holdings. Other holdings were checked for a valid 5-digit postal code, and then randomly sampled for a coordinate within the postal code ($n = 4748$). Finally, remaining holdings were checked for a valid postal area (contains several postal codes), and randomly sampled within the postal area ($n = 1283$). For two holding identifiers, no coordinate could be assigned. These two holdings were kept in the simulations since the event data was based on the holding identifier and not on the coordinate.

To explore seasonality in the input data, time series with the number of events, the number of holdings with at least one animal and number of animals per age category were produced. A smoother, using local polynomial regression fitting (loess) [32], was added to the time series. Moreover, a time series with the proportion of holdings, per day of the year, connected to at least one other holding was determined. A generalised additive model [33] (GAM) of the proportion against the day of the year was fitted. A smooth term with cyclic cubic regression splines was used for the GAM model.

Computational simulation framework

The disease spread model was implemented in *SimInf* [9, 34]. *SimInf* is an R [35] package for data-driven stochastic

disease spread simulations, developed by us. The overall design was inspired and partly adapted from the Unstructured Mesh Reaction–Diffusion Master Equation (URDME) framework [36, 37]. The *SimInf* package uses object oriented programming in R. We defined objects with logical layers connected by well-defined interfaces for different modelling scenarios. The package also uses the ability to interface high performance compiled code from R. We implemented the core algorithm of the simulator in the compiled language C [38]. To improve performance further, we used OpenMP in the core simulation algorithm to divide work over multiple processors and perform computations in parallel. A detailed description of the implementation and data structures of the simulation algorithm are presented in [8, 9].

The disease spread simulations in the present paper were performed with the model named *SISe3* in the *SimInf* package version 2.0.0 using R version 3.2.3. The simulation was initiated by supplying the initial state in every holding (see below) together with all events.

Initialisation

The geographical distribution of the VTEC O157 herd prevalence in Sweden was reported by [39]. In southern Sweden, the herd prevalence varied between 3.3 and 23.3% while no farms were found positive in northern Sweden. The expected number of infected holdings in each county was estimated from the reported geographical distribution of the herd prevalence. Initially infected holdings were identified through random sampling of all herds, without replacement, where each holding had a probability weight equal to its county prevalence.

A Swedish nationwide abattoir survey was conducted during 2005–2006 [40] to determine the prevalence of cattle carrying VTEC O157. In initially infected holdings, the within holding prevalence was calculated to meet three requirements from that study; (i) an overall prevalence of 3.4%, (ii) a prevalence of 15.7% in cattle younger than 1 year and, (iii) a prevalence of 2.6% in cattle older than one year.

The initial environmental infectious pressure, $\varphi_i(0)$, was set to

$$\varphi_i(0) = \frac{\alpha \sum_j I_{i,j}(0)}{N_i(0)}. \quad (4)$$

To evaluate if the outcome of the spread on national scale (see below) was a result of the initial state, initialisation was also performed with a uniform geographical distribution. In these simulations, initially infected holdings were identified through random sampling of all herds, without replacement, where each holding had a probability weight equal to a national prevalence of 20%. The within holding prevalence was 100% in initially infected

holdings and the initial environmental infectious pressure, $\varphi_i(0)$, was calculated from Equation. 4.

Parameter estimation

A previous longitudinal observational study over 38 months in 126 cattle holdings located in southern Sweden in the two NUTS 2 regions “SE21: Småland med öarna” and “SE23: Västsverige” [25] was used to estimate the parameters in the model. To determine the status that could have been found if simulated holdings had been sampled, the sampling strategy was replicated. In this previous study, environmental samples were repeatedly collected from 126 Swedish cattle holdings during the period October 2009 to December 2012 to determine the VTEC O157 herd status at multiple time points ($n = 2009$). The environmental sampling strategy used in the longitudinal study [25] has previously been evaluated against pooled individual faecal samples (pool size = 3), showing that the sensitivity depended upon the prevalence of positive pools [41]. Moreover, the sensitivity to detect VTEC O157 in pooled samples has been shown to depend upon the proportion of positive samples in the pool [42].

The environmental sampling was simulated at each sample point as follows. First, pools (pool size = 3) were randomly created within each age category from the number of susceptible and infected individuals at the time for the sample point in the simulation. Then each pool was randomly classified as positive or negative, with P (positive) equal to the test sensitivity from [42], given the proportion of infected individuals in the pool. Similarly, using the estimated pool prevalence, the holding

status was randomly classified as positive or negative given the sensitivity of the environmental sampling protocol from [41].

In total, there were 12 parameters in the SIS_E model (Table 1). To maintain model parsimony, the recovery rate γ_j was assigned equal values in all age categories and the shed rate α , was fixed at 1.0 per day, thus defining the unit of the environmental infectious pressure variable $\varphi_i(t)$. The duration of infection in cattle excreting VTEC O157 was studied in a longitudinal study [43] and the recovery rate γ_j was estimated from the mean duration.

The parameters β , v_j and ε were estimated by evaluating the agreement between observed statuses, obtained from results of the longitudinal study [25], and the simulated statuses. The two time-series with observed and simulated statuses was defined as $Y(t)$ and $Y^*(t, \theta)$ where θ was the vector of model parameters in the simulation. A GAM model of the status against the day of the year was fitted to $Y(t)$ and $Y^*(t, \theta)$ with binomial distribution, logit link function and day of the year as a smoothing term with cyclic cubic regression splines. We let η_k and $\eta_k^*(\theta)$ be the coefficients from the fitted GAM model of $Y(t)$ and $Y^*(t, \theta)$, respectively, with $k \in \{1, 2, \dots, 9\}$.

The parameter estimation was approached as an optimisation problem to find the values of θ that minimised the difference between the coefficients η_k and $\eta_k^*(\theta)$ under the constraint that all parameters $\theta \geq 0$. Using the stochastic simulator *SimInf*, each outcome provided a measurement of the system with process noise. Accordingly, the average coefficient $\bar{\eta}_k^*(\theta)$ was estimated from $N = 40$ trajectories

Table 1 Parameters in a SIS_E VTEC O157 model

| Parameter | Description (unit) | Value |
|---------------|---|-----------------------|
| α | Rate of shedding from infected individuals (units per day) | 1.00×10^0 |
| β_{q1} | Decay of environmental infectious pressure in quarter 1 (per day) | 1.75×10^{-1} |
| β_{q2} | Decay of environmental infectious pressure in quarter 2 (per day) | 1.19×10^{-1} |
| β_{q3} | Decay of environmental infectious pressure in quarter 3 (per day) | 0.80×10^{-1} |
| β_{q4} | Decay of environmental infectious pressure in quarter 4 (per day) | 1.58×10^{-1} |
| v_c | Indirect transmission rate of the environmental infectious pressure in calves (per animal per day) | 3.83×10^{-2} |
| v_y | Indirect transmission rate of the environmental infectious pressure in young stock (per animal per day) | 3.83×10^{-2} |
| v_a | Indirect transmission rate of the environmental infectious pressure in adults (per animal per day) | 5.88×10^{-3} |
| γ_c | The recovery rate of infection in calves (per day) | 1.00×10^{-1} |
| γ_y | The recovery rate of infection in young stock (per day) | 1.00×10^{-1} |
| γ_a | The recovery rate of infection in adults (per day) | 1.00×10^{-1} |
| ε | Background environmental infectious pressure (units per day) | 6.89×10^{-5} |

Parameters in a stochastic simulation to explore the spread of Verotoxigenic *Escherichia coli* O157:H7 (VTEC O157) in the entire Swedish cattle population based on data reported to the Swedish Board of Agriculture during the period 2005-07-01 to 2013-12-31. The within-herd disease spread was modelled with a SIS_E compartment model the two disease states: susceptible (S) and infected (I) and E representing the environmental compartment contaminated with VTEC O157 by infected animals. The decay of the environmental infectious pressure was varied in each of the four quarters of the year. Individuals were divided into the following three age categories; calves 0–119 days, young stock 120–364 days and adults older than 364 days.

$$\bar{\eta}_k^*(\theta) = \frac{1}{N} \sum_N \eta_k^*(\theta). \quad (5)$$

The objective function to measure the agreement was defined as

$$G(\theta) = \sum_k (\bar{\eta}_k^*(\theta) - \eta_k)^2. \quad (6)$$

The parameter combination θ that minimized $G(\theta)$ was obtained with the Nelder-Mead algorithm [44] using a linearly constrained optimisation method in R. The parameter estimation was first performed with the decay of the environmental infectious pressure, β , fixed at a decimal reduction rate (the time required at a given temperature to kill 90% of the exposed microorganisms) of 16 days for VTEC O157 [45–47]. The state transition rate v_j from susceptible to infected state was assumed to be equal in the calves and young stock age categories. To improve model fit, the parameter estimation was refitted with the decay of the environmental infectious pressure, β , allowed to vary in each quarter of the year.

Exploring spread on a national scale

The following simulation experiment was conducted to explore the VTEC O157 spread model on a national scale. In each of the eight NUTS level 2 regions in Sweden, a set of 126 holdings were randomly selected. In each region, each selected holding was mapped to represent one herd in the longitudinal observational study [25]. These eight new sets of holdings represent what may have been found if the observational study [25] had been conducted in each of the regions. One thousand trajectories were simulated over the time period 2005-07-01 to 2013-12-31 with the initial prevalence according to [39, 40] (see above). For every trajectory, simulated environmental samples were generated for all selected holdings in each region at the time points each herd was sampled in the longitudinal observational study [25] during the time period October 2009 to December 2012. A GAM model, as previously described, was fitted to the sample points in each region for every trajectory. Using the GAM model, the predicted proportion selected holdings that were infected in each NUTS 2 region the first day in each quarter of the year were calculated for each trajectory and visualized with a boxplot. A multivariable linear regression model was used to assess the relationship between the proportion infected holdings and the NUTS 2 region and the quarter of the year. The simulation experiment was repeated with an initial holding prevalence of 20% and all individuals infected in the infected holdings (see above).

Sensitivity analysis

Sensitivity analysis was performed to explore how variation in the model parameters would influence the outcome of the simulation experiment on national scale. The variation was done with a scaling factor for α , β_{q1} , β_{q2} , β_{q3} , β_{q4} , γ_j and v_j that varied from 0.95 to 1.05 in steps of 0.01 and a scaling factor for ϵ that varied from 0.0 to 2.0 in steps of 0.2. The simulation experiment was repeated for each combination of the scaled values of β against the scaled values of ϵ , for each combination of the scaled values of β against the scaled values of v_j and for each combination of the scaled values of α against the scaled values of γ_j . For every combination, the average proportion positive holdings in each NUTS 2 region at the first day in quarter 1, 2, 3 and 4 was determined from 100 trajectories.

Results

Cattle population and events

The number of holdings decreased in the population during the 8.5 year study period with an evident seasonal pattern with more active holdings during the pasture season (Figure 2). The total cattle population in Sweden was about 1.6 million individuals with a slightly decreasing trend during the study period (Figure 2). There was a seasonal pattern in the number of animals within each age category. The externally scheduled events based on register data had an evident seasonal variation (Figure 3). The enter events, representing births and imports, peaked during spring each year. Both the movement data, external transfer events, (Figure 3) and the proportion of connected holdings (Figure 4) had one peak during spring and one peak during autumn. Slaughter and export events, represented by exit, had a bimodal shape with a sharp decline at the end of each year.

Parameter estimation

The parameters estimated for the SIS_E model are shown in Table 1. The simulated outcome showed no seasonal variation in the proportion positive holdings unless the decay of the environmental infectious pressure β was allowed to vary in each quarter of the year. A comparison of the results from the longitudinal observational study [25] and the simulated outcome from the constant and the time-varying β is presented in Figure 5.

Exploring spread on a national scale

Figure 6 shows the result from the simulation experiment where 126 holdings were randomly sampled in each of the eight NUTS level 2 regions in Sweden to explore the spread on a national scale. The coefficients for the multivariable linear regression model to assess the relationship between the proportion infected holdings based on

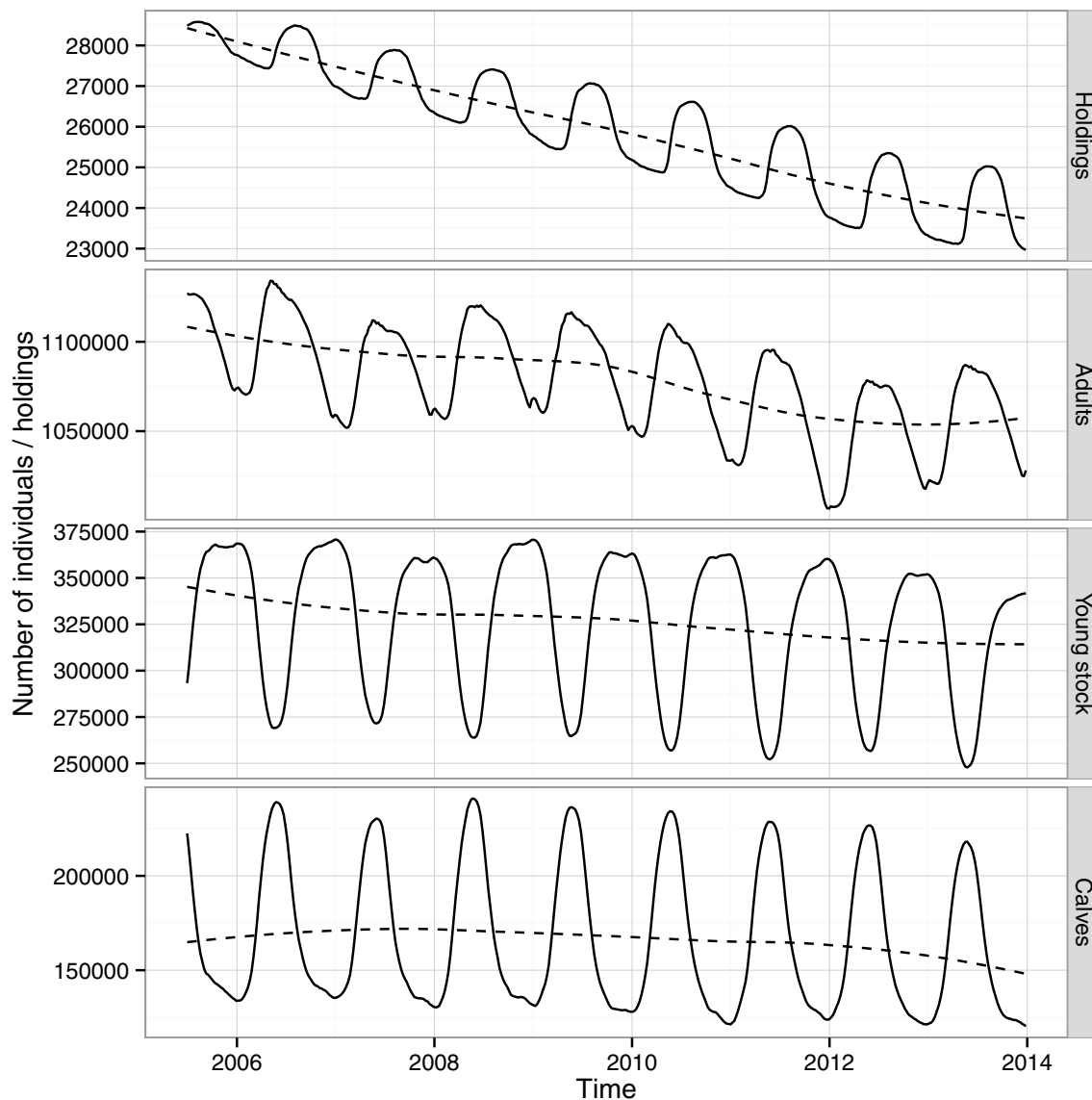


Figure 2 The number of holdings and cattle. The number of active holdings (having at least one animal) and the number of cattle in Sweden during the period 2005-07-01 to 2013-12-31. Data is from the national cattle database managed by the Swedish Board of Agriculture. The number of cattle is grouped by age category; Calves 0–119 days, Young stock 120–364 days and Adults older than 364 days. The dashed line is a loess smoother. The data was used in a simulation to explore the spread of VTEC O157 in the complete Swedish cattle population.

the NUTS 2 region and the quarter of the year are shown in Table 2. The proportion infected holdings was significantly higher in the southern region SE22 (Sydsverige) compared to the other regions. Furthermore, it was significantly lower in quarter two and three and higher in quarter four compared to quarter one. The highest proportion of positive holdings were observed in SE22 (Sydsverige), on average between 8–10%. In contrast, the lowest proportion of positive holdings, on average between 2–3% was observed in SE32 (Mellersta Norrland). The same pattern was found when the simulation

was initialised at a 20% holding prevalence and 100% infected individuals in infected holdings.

Sensitivity analysis

Figures 7 and 8 show the result from the sensitivity analysis in quarter 4 of the simulation experiment on national scale when varying β_{q1} , β_{q2} , β_{q3} , β_{q4} against ϵ and β_{q1} , β_{q2} , β_{q3} , β_{q4} against v_i , respectively. The overall pattern from the sensitivity analysis in quarter 1, 2 and 3 is similar, however, with a lower proportion positive holdings (data not shown). The proportion of positive holdings

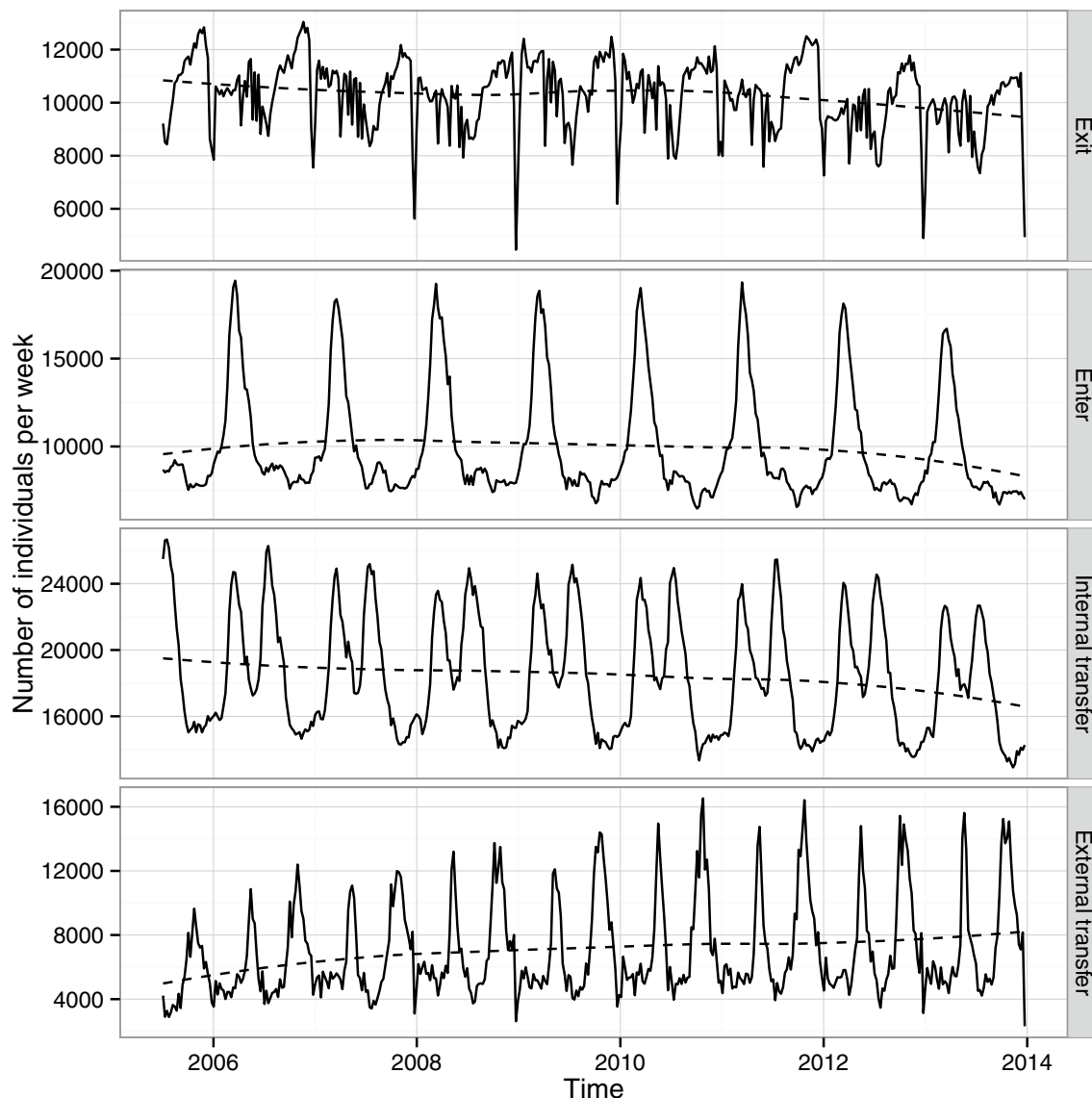


Figure 3 Events in the Swedish cattle population. The number of cattle affected per week by four event types in a simulation of VTEC O157 in the complete Swedish cattle population based on data reported to the Swedish Board of Agriculture during the period 2005-07-01 to 2013-12-31. The exit event implies slaughter, euthanasia or export, and defines the day an animal is excluded from the simulation. The enter event handles births and imports. The internal transfer event occurs on the day an animal changes age category from calf (0–119 days) to young stock (120–364 days) or young stock to adult (older than 364 days). The external transfer event occurs when animal moves from one holding to another. The dashed line is a loess smoother.

decreased in all regions and all quarters of the year when ϵ and v_j was decreased and when β was increased. In all regions except SE22 (Sydsverige) the average proportion of positive holdings was below 0.03 in quarter 1, 2, 3 and 4 when ϵ was zero, regardless of β in the investigated range of values. In contrast, in SE22 (Sydsverige) the average proportion of positive holdings was above 0.04. When varying β against v_j , the average proportion positive holdings was above 0.01 in all regions and quarters

of the year. Figure 9 shows the result from the sensitivity analysis in quarter 4 of the simulation experiment on national scale when varying α against γ_j . The overall pattern from the sensitivity analysis in quarter 1, 2 and 3 was similar, although with a lower proportion positive holdings (data not shown). The proportion of positive holdings decreased in all regions and all quarters of the year when α was decreased and when γ_j was increased. When varying α against γ_j , the average proportion positive

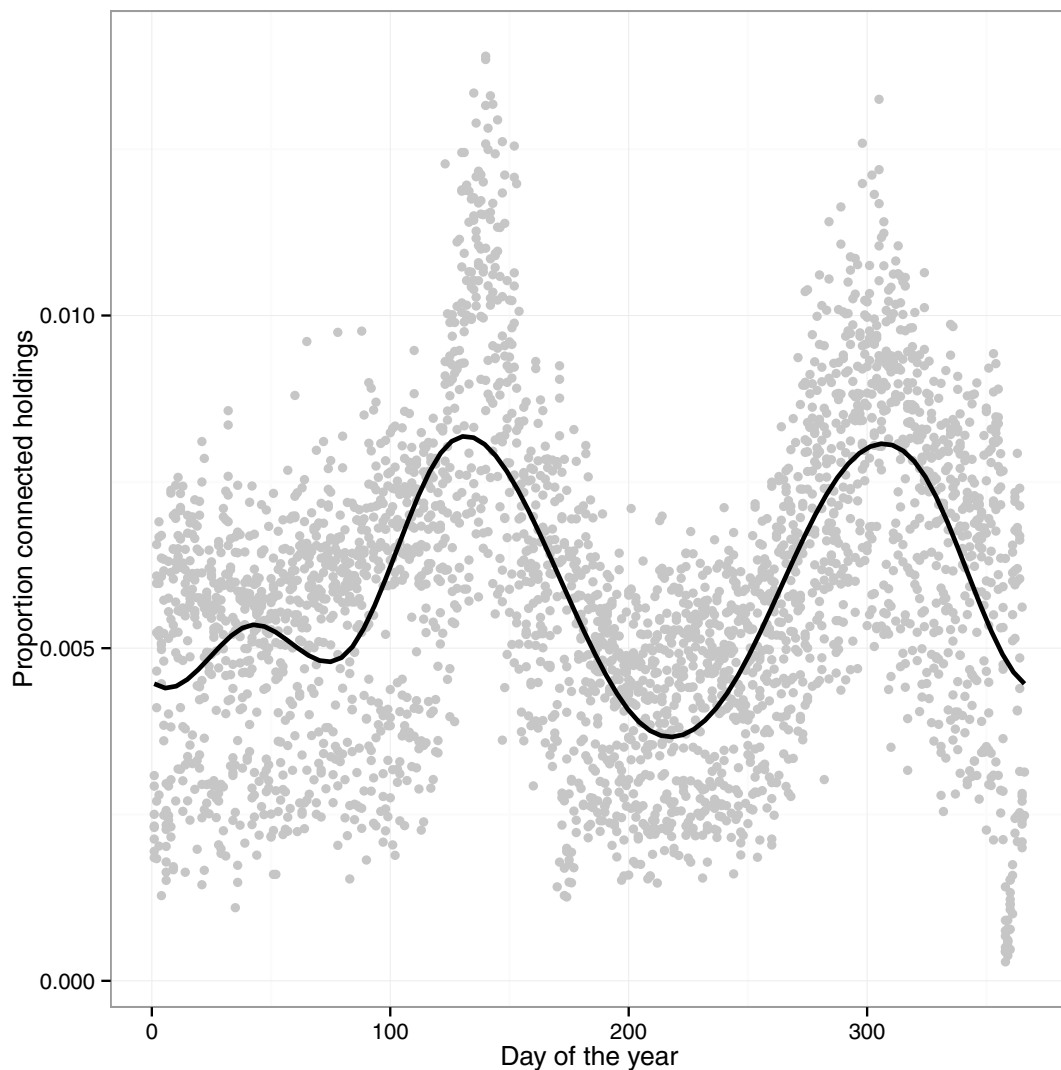


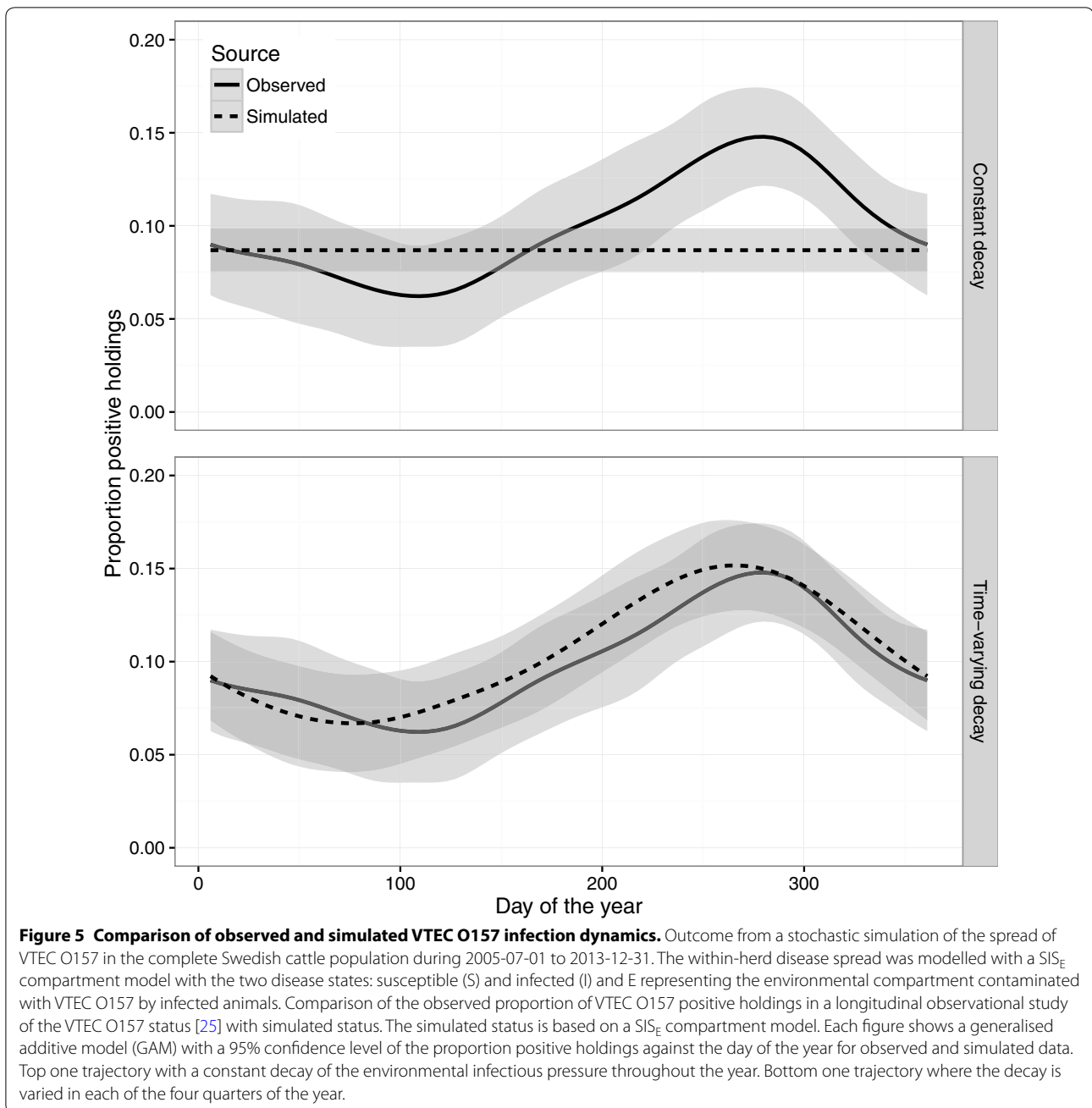
Figure 4 Proportion of holdings connected with cattle movements. Scatter plot with the proportion of holdings per day of the year with at least one connection to another holding in a simulation of VTEC O157 in the complete Swedish cattle population. The graph is based on cattle movements reported to the Swedish Board of Agriculture during the period 2005-07-01 to 2013-12-31. The solid line is a smoother based on a generalised additive model (GAM) with cyclic cubic regression splines.

holdings was above 0.01 in all regions and quarters of the year.

Discussion

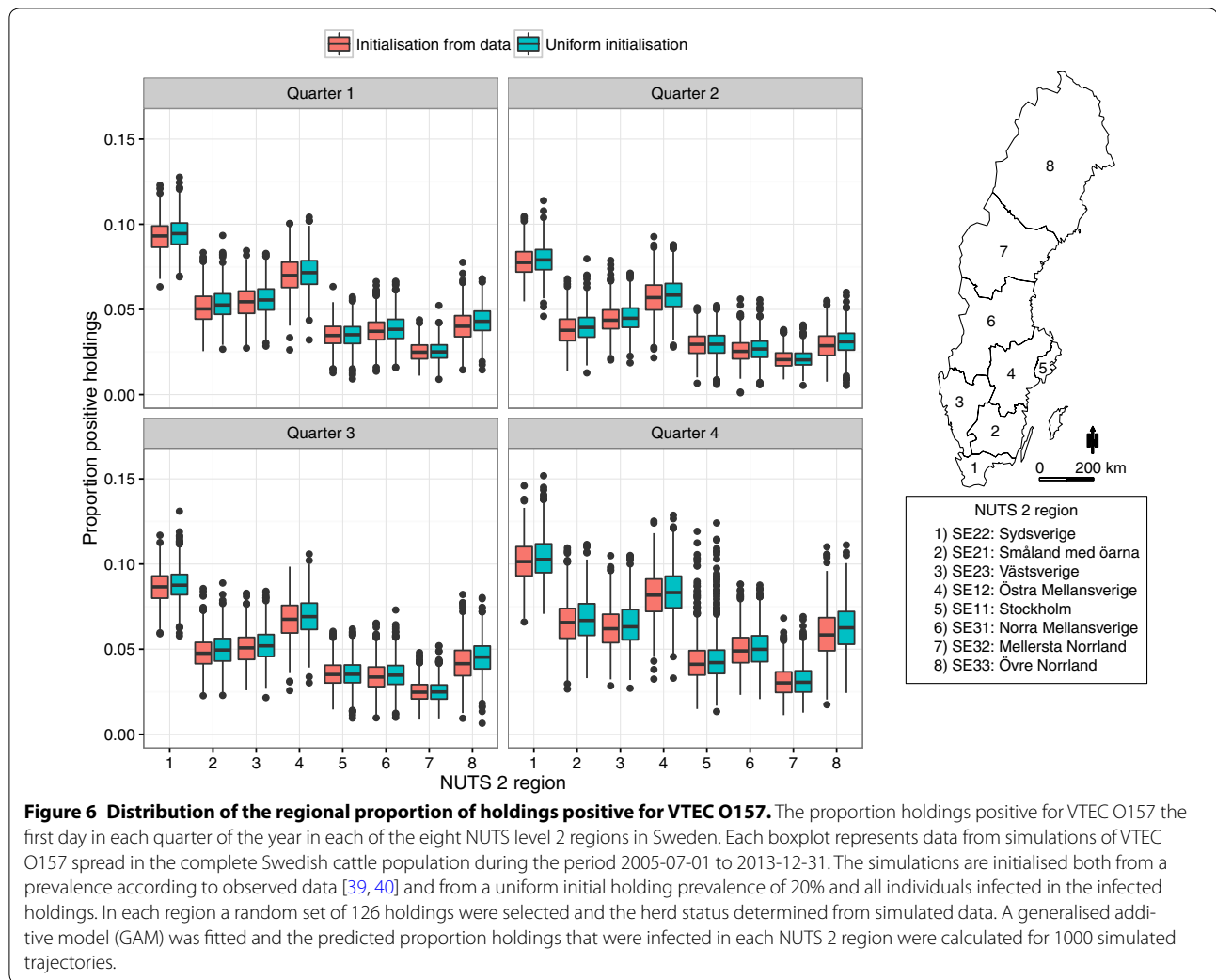
The present study used real livestock data (births, movements and slaughter) over 8.5 years (2005-07-01–2013-12-31) to explore the spatio-temporal spread of VTEC O157 in the entire Swedish cattle population. This approach allows for disease spread modelling that naturally incorporates the time-varying contact network and the population demographic. The results show that the

data-driven simulation captures previously observed spatial trends in disease occurrence, with higher prevalence of VTEC O157 in southern Sweden [48, 49]. This was regardless of whether the simulations were initialised from observed data [39, 40] or from a uniform distribution. In this study, the parameters of the stochastic within-herd disease spread model were the same for all herds, irrespective of geographic location, and therefore the regional differences from the simulation depends solely on properties intrinsic to the livestock data. Characteristics of a time-varying contact network are



important for the spread of disease [50–52] and has been reported to influence the spread of other cattle diseases, for example foot-and-mouth disease [53, 54], bovine viral diarrhoea virus [55] and paratuberculosis [56]. This highlights the importance of including available and detailed network data in simulation studies since it has implications for the spread of several infectious diseases in a population of interacting farms.

The seasonality of the livestock movements in Sweden has been demonstrated in previous studies [24, 30] and has also been reported from other European countries, such as Italy [57] and France [50]. We observed a slightly increasing trend in the number of moved animals during the extended study period, compared to the earlier Swedish studies. In addition, there was a seasonal pattern in the number of active holdings,



the number of animals and the proportion of holdings connected with movements. Furthermore, the average herd size has increased over time, which could have implications for local disease transmission as herd size has been identified as a risk factor for VTEC O157 [17, 22, 25].

There was a better agreement between the observed seasonality in the longitudinal study of the VTEC O157 herd status [25] and simulated results, when the decay and removal of the environmental infectious pressure β was allowed to vary by quarters of the year. This suggests that the seasonality in population demographic, movement data and contact network in itself are not enough

to explain the observed seasonal pattern. A similar result may be achieved by introducing an influence of seasonality on other model parameters. It has been hypothesised that day length may explain the seasonal shedding of VTEC O157 [58] which could be represented in the model with a seasonal shedding parameter α . Alternatively, the background infectious pressure ε could vary over season. Another approach could be to model the growth rate of the bacteria by ambient temperature [59, 60]. However, to keep the model parsimonious and because good data on α and ε are difficult to find, we chose to only vary the decay of the environmental infectious pressure β .

Table 2 The results of the multivariable linear regression

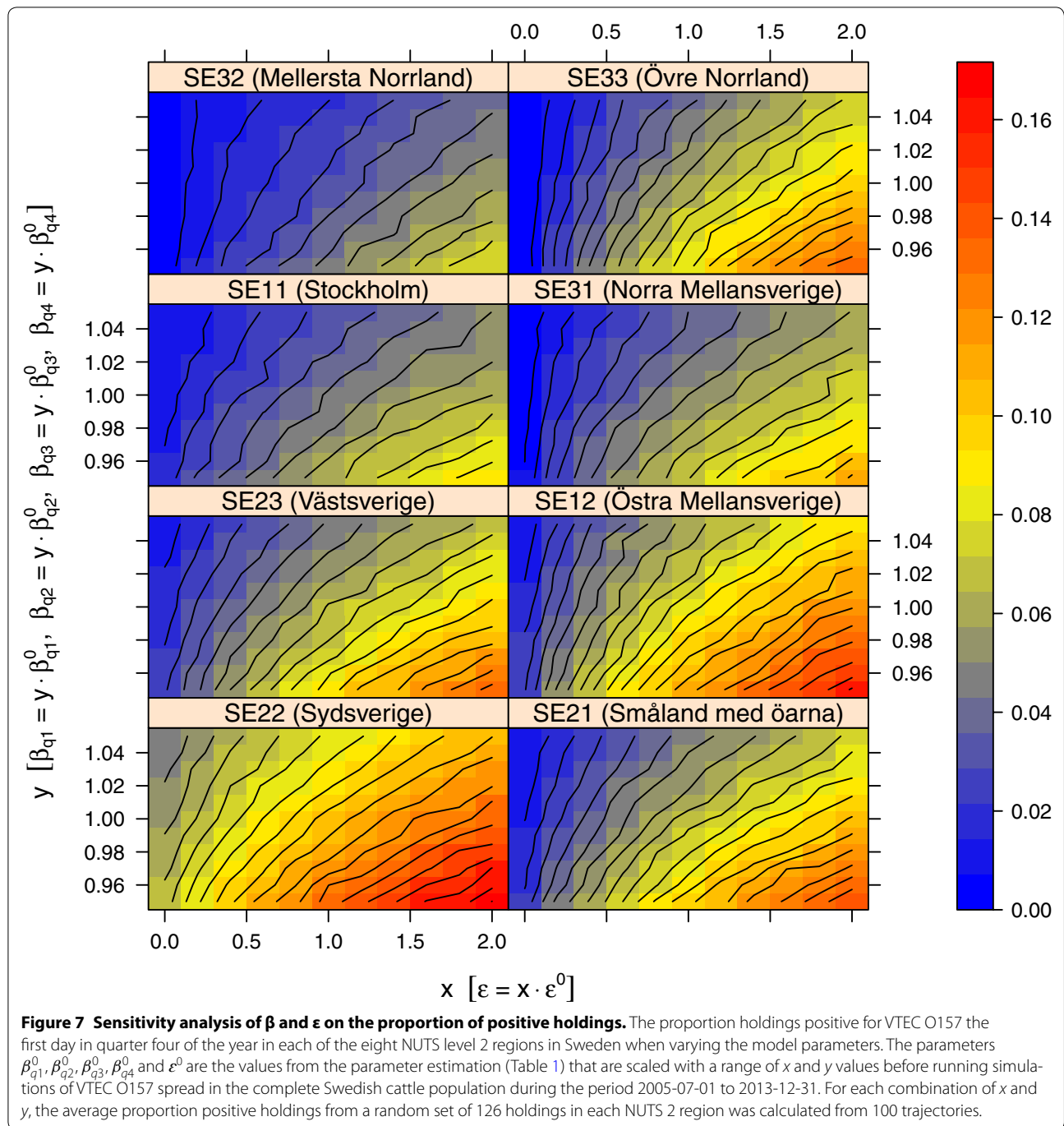
| Covariate | Estimate | Standard error | P |
|----------------------------|----------|----------------------|--------|
| Intercept | 0.090 | 1.9×10^{-4} | <0.001 |
| SE22 (Sydsverige) | Baseline | | |
| SE21 (Småland med öarna) | -0.039 | 2.3×10^{-4} | <0.001 |
| SE23 (Västsverige) | -0.037 | 2.3×10^{-4} | <0.001 |
| SE12 (Östra Mellansverige) | -0.021 | 2.3×10^{-4} | <0.001 |
| SE11 (Stockholm) | -0.054 | 2.3×10^{-4} | <0.001 |
| SE31 (Norra Mellansverige) | -0.053 | 2.3×10^{-4} | <0.001 |
| SE32 (Mellersta Norrland) | -0.064 | 2.3×10^{-4} | <0.001 |
| SE33 (Övre Norrland) | -0.047 | 2.3×10^{-4} | <0.001 |
| Quarter 1 | Baseline | | |
| Quarter 2 | -0.011 | 1.6×10^{-4} | <0.001 |
| Quarter 3 | -0.002 | 1.6×10^{-4} | <0.001 |
| Quarter 4 | 0.011 | 1.6×10^{-4} | <0.001 |

The coefficients to assess the relationship between the proportion infected holdings and the NUTS 2 region and quarter of the year. The multivariable regression is calculated from the outcome in a stochastic simulation to explore the spread of Verotoxigenic *Escherichia coli* O157:H7 (VTEC O157) in the entire Swedish cattle population based on data reported to the Swedish Board of Agriculture during the period 2005-07-01 to 2013-12-31.

The model seemingly overestimated the prevalence of VTEC O157 in the two most northern regions in Sweden, SE32 (Mellersta Norrland) and SE33 (Övre Norrland). There are several plausible reasons for this outcome and the results from the sensitivity analysis suggests directions for improvements. The length of the four seasons in Sweden differs by region, and the model might have a better fit if the decay of the environmental infectious pressure β is also varied by region. A limitation of the disease spread model used in the present paper is that ϵ is stationary and that the model does not allow the environmental infectious pressure to influence neighbouring herds. Recent research has highlighted the relevance of local spread of VTEC O157 between cattle farms [22, 25] and ϵ could be split into two parts, one part for random introduction and one part that is local spread. We suggest future research to explore the spatial dependence in model parameters and local spread between neighbouring cattle farms. Our conclusion is that the spatial components in the population demographic and the contact network gives a partial explanation of the observed distribution of VTEC O157 in the Swedish cattle population, but that there are other factors that likely should be included in order to reach a more comprehensive understanding of the observed pattern.

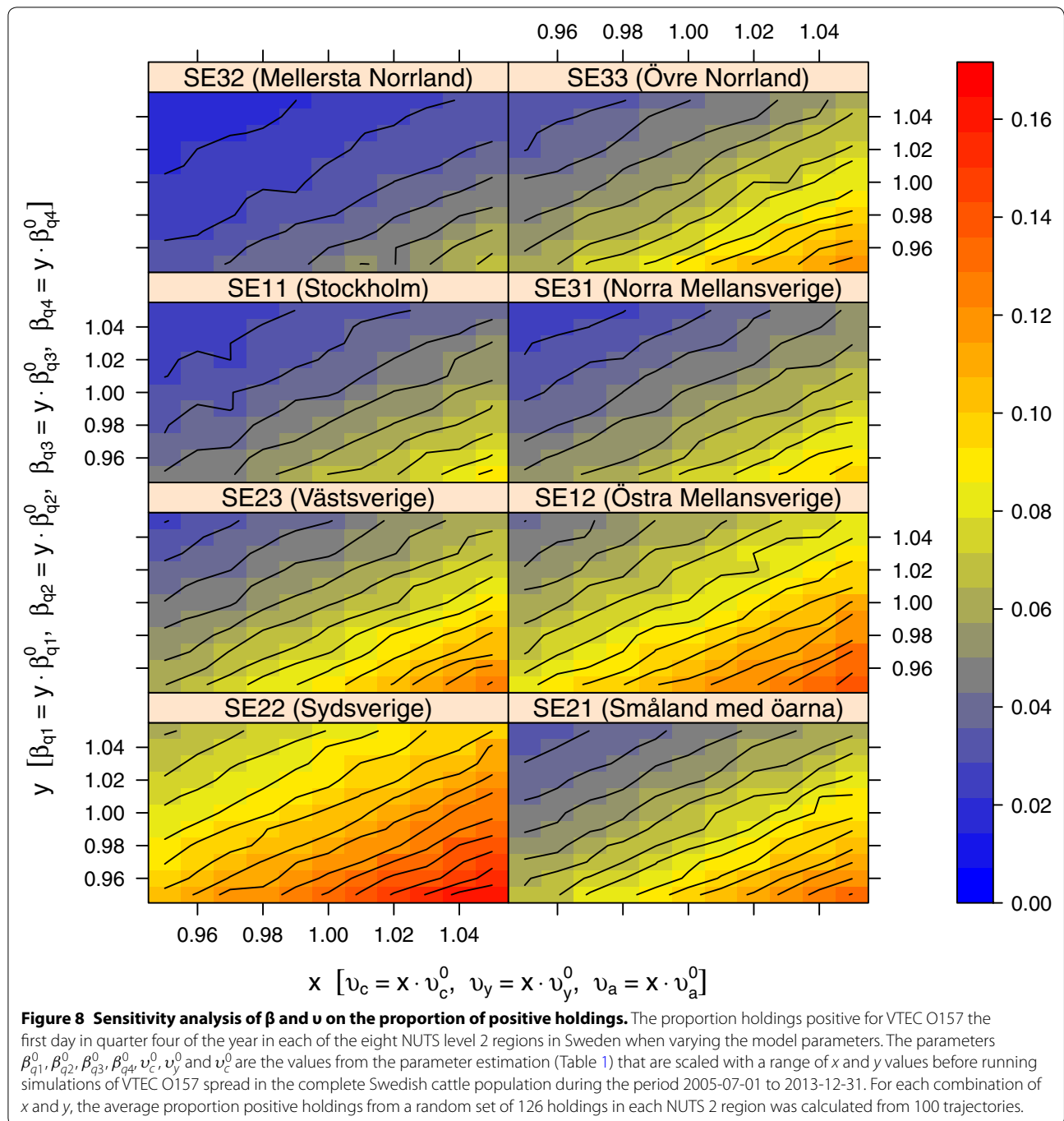
The disease dynamics in the most southern region SE22 (Sydsverige) appear unique among the investigated regions in the study. After removing introduction of infection from other sources than animal movements, by assigning the background infectious pressure ϵ to zero, the proportion infected holdings was higher than in the other regions. It is unrealistic to assume that ϵ could be zero, even though appropriate biosecurity measures most likely could reduce ϵ . To maintain model parsimony, both the shedding rate α and the recovery rate γ were kept at equal values between the age categories. However, it has been reported that the magnitude of shedding is greater and the duration longer in calves compared to adult cattle [29] and that a small proportion of infected animals shed at much higher levels [61]. The results show that varying these two model parameters influences the proportion of infected holdings. Increasing the number of parameters in the model would also increase the complexity and the computational challenge of the parameter estimation. Nevertheless, a more complex disease spread model could further enhance the understanding of the transmission dynamics among interacting farms.

It has been suggested that characteristics of the spatio-temporal network of livestock movements can be used to improve surveillance and control spread of disease on a regional and national scale [50, 57, 62]. There are several challenges to develop models that make use of detailed livestock data, e.g. combining it with disease data from surveillance programmes and field studies [6]. Moreover, it is computationally challenging to model disease transmission over large dynamic networks [7]. Using the design of URDM [36], as a starting point in the development of *SimInf*, provided a logical separation between the core simulator and the model specification. The use of sophisticated parallel algorithms enabled computationally efficient national scale simulations of within-herd infection dynamics combined with realistic data-driven modelling of the time varying contact network and the population demographic. Furthermore, *SimInf*, being an R package, provided an environment for pre- and post-processing of data, statistical analysis and visualisation, all of which are important components in simulation. It also provides an infrastructure to share and extend knowledge on various operating systems through CRAN (The Comprehensive R Archive Network). Our goal is that *SimInf* [9] will grow by including new model specifications through contributions from the scientific community.



This paper explored the spread of VTEC O157 in the complete Swedish cattle population. Real livestock data was used to incorporate the time varying contact

network and the population demographic. The results showed that regional differences in prevalence may be due to regional differences in livestock movements.



Furthermore, although movements, births and slaughter of cattle also have distinct seasonal patterns, these factors could not by themselves explain the seasonal pattern

of VTEC O157 prevalence in cattle. With this work, we describe how an efficient and freely available software can contribute to development of realistic large-scale

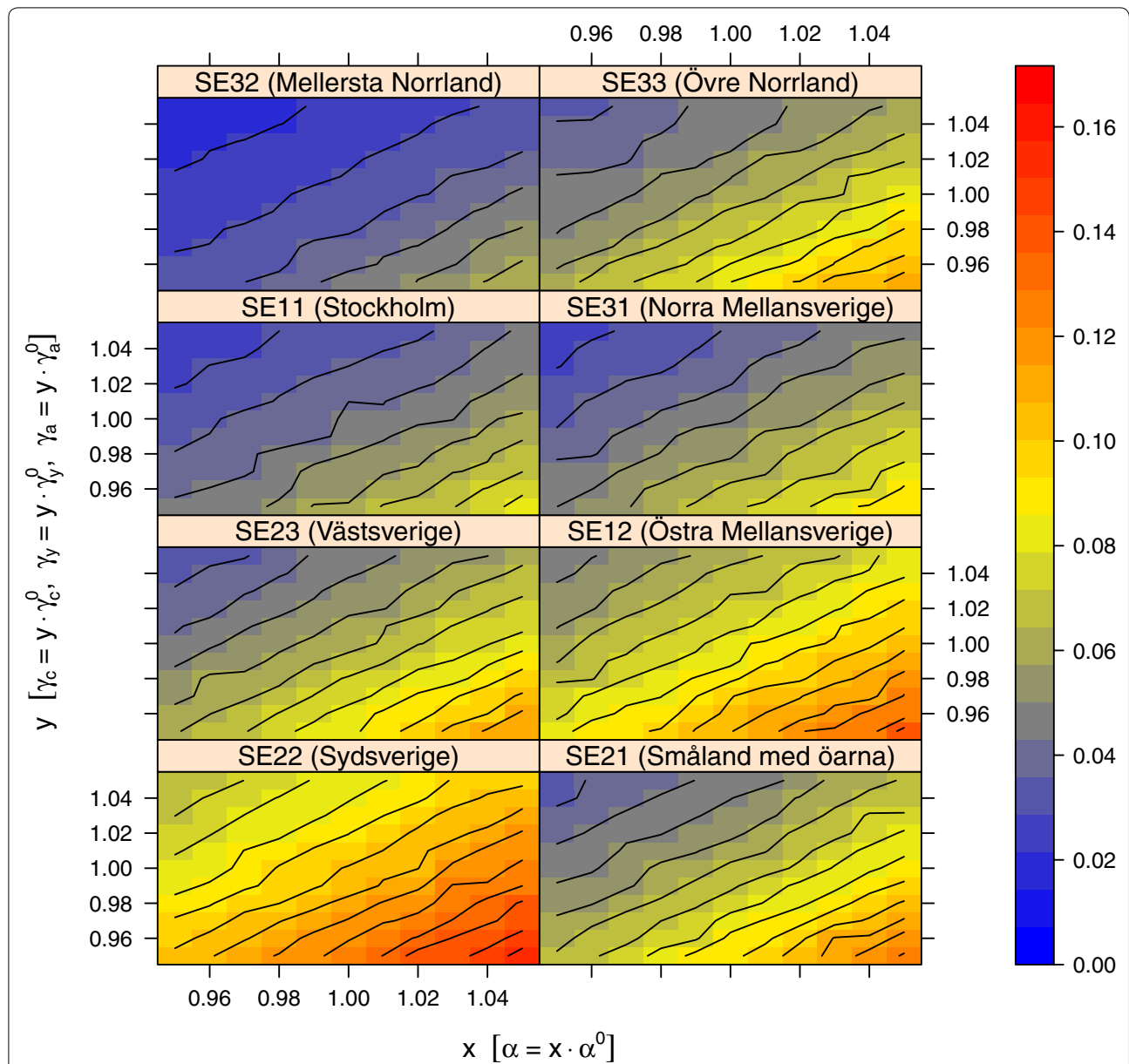


Figure 9 Sensitivity analysis of γ and α on the proportion of positive holdings. The proportion holdings positive for VTEC O157 the first day in quarter four of the year in each of the eight NUTS level 2 regions in Sweden when varying the model parameters. The parameters $\gamma_c^0, \gamma_y^0, \gamma_a^0$ and α^0 are the values from the parameter estimation (Table 1) that are scaled with a range of x and y values before running simulations of VTEC O157 spread in the complete Swedish cattle population during the period 2005-07-01 to 2013-12-31. For each combination of x and y , the average proportion positive holdings from a random set of 126 holdings in each NUTS 2 region was calculated from 100 trajectories.

epidemiological models. Future research will encompass studies of more complex disease spread models and the effects of various intervention strategies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JF and SW designed the simulation study. SW performed the simulations and drafted the manuscript. PB, SE and SW designed the mathematical model

and designed and implemented the SimInf R package. AL, JF, SE, SW and UE participated in the epidemiological analyses and critically reviewing the paper. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank (in alphabetical order) Dr. Arianna Comin, Dr. Kaisa Sörén and Dr. Thomas Rosendal for valuable comments on the manuscript. The simulations were partly performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX.

Funding

This work was financially supported by the Swedish Research Council within the UPMARC Linnaeus centre of Excellence (Pavol Bauer and Stefan Engblom).

Author details

¹ Department of Clinical Sciences, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden. ² Department of Disease Control and Epidemiology, National Veterinary Institute, 751 89 Uppsala, Sweden. ³ Division of Scientific Computing, Department of Information Technology, Uppsala University, 751 05 Uppsala, Sweden. ⁴ Department of Animal Environment and Health, Swedish University of Agricultural Sciences, Box 234, 532 23 Skara, Sweden.

Received: 13 January 2016 Accepted: 18 July 2016

Published online: 11 August 2016

References

- Anonymous (2000) Regulation (EC) No 1760/2000 of the European Parliament and of the Council of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products. *Off J Eur Union L* 204:1–10
- Anonymous (2004) Commission Regulation (EC) No 911/2004 of 29 April 2004 implementing Regulation (EC) No 1760/2000 of the European Parliament and of the Council as regards eartags, passports and holding registers. *Off J Eur Union L* 163:65–70
- Garner MG, Beckett SD (2005) Modelling the spread of foot-and-mouth disease in Australia. *Aust Vet J* 83:758–766
- Harvey N, Reeves A, Schoenbaum MA, Zagmutt-Vergara FJ, Dube C, Hill AE, Corso BA, McNab WB, Cartwright CI, Salman MD (2007) The North American Animal Disease Spread Model: a simulation model to assist decision making in evaluating animal disease incursions. *Prev Vet Med* 82:176–197
- Stevenson MA, Sanson RL, Stern MW, O'Leary BD, Sujau M, Moles-Benfell N, Morris RS (2013) InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. *Prev Vet Med* 109:10–24
- Brooks-Pollock E, de Jong MC, Keeling MJ, Klinkenberg D, Wood JL (2015) Eight challenges in modelling infectious livestock diseases. *Epidemics* 10:1–5
- Pellis L, Ball F, Bansal S, Eames K, House T, Isham V, Trapman P (2015) Eight challenges for network epidemic models. *Epidemics* 10:58–62
- Bauer P, Engblom S, Widgren S (2016) Fast event-based epidemiological simulations on national scales. *Int J High Perform Comput Appl*. doi:10.1177/1094342016635723. Accessed 7 Aug 2016
- Widgren S, Bauer P, Engblom S (2016) SimInf: an R package for data-driven stochastic disease spread simulations. *ArXiv e-prints*, arXiv:1605.01421 [q-bio.PE]. <http://www.arxiv.org/abs/1605.01421>. Accessed 7 Aug 2016
- Karmali MA, Petric M, Lim C, Fleming PC, Steele BT (1983) *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. *Lancet* 322:1299–1300
- Riley LW, Remis RS, Helgerson SD, McGee HB, Wells JG, Davis BR, Hebert RJ, Olcott ES, Johnson LM, Hargrett NT, Blake PA, Cohen ML (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* 308:681–685
- Mead PS, Griffin PM (1998) *Escherichia coli* O157:H7. *Lancet* 352:1207–1212
- Tarr PI, Gordon CA, Chandler WL (2005) Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* 365:1073–1086
- Hancock D, Besser T, Lejeune J, Davis M, Rice D (2001) The control of VTEC in the animal reservoir. *Int J Food Microbiol* 66:71–78
- Strachan NJ, Dunn GM, Locking ME, Reid TM, Ogden ID (2006) *Escherichia coli* O157: burger bug or environmental pathogen? *Int J Food Microbiol* 112:129–137
- Wilson JB, Renwick SA, Clarke RC, Rahn K, Alves D, Johnson RP, Ellis AG, McEwen SA, Karmali MA, Lior H, Spika J (1998) Risk factors for infection with verocytotoxigenic *Escherichia coli* in cattle on Ontario dairy farms. *Prev Vet Med* 34:227–236
- Ellis-Iversen J, Smith RP, Snow LC, Watson E, Millar MF, Pritchard GC, Sayers AR, Cook AJ, Evans SJ, Paiba GA (2007) Identification of management risk factors for VTEC O157 in young-stock in England and Wales. *Prev Vet Med* 82:29–41
- Nielsen EM, Tegtmeier C, Andersen HJ, Gronbaek C, Andersen JS (2002) Influence of age, sex and herd characteristics on the occurrence of Verocytotoxin-producing *Escherichia coli* O157 in Danish dairy farms. *Vet Microbiol* 88:245–257
- Chapman PA, Siddons CA, Gerdan Malo AT, Harkin MA (1997) A 1-year study of *Escherichia coli* O157 in cattle, sheep, pigs and poultry. *Epidemiol Infect* 119:245–250
- Mechie SC, Chapman PA, Siddons CA (1997) A fifteen month study of *Escherichia coli* O157:H7 in a dairy herd. *Epidemiol Infect* 118:17–25
- Milnes AS, Sayers AR, Stewart I, Clifton-Hadley FA, Davies RH, Newell DG, Cook AJ, Evans SJ, Smith RP, Paiba GA (2009) Factors related to the carriage of Verocytotoxigenic *E. coli*, *Salmonella*, thermophilic *Campylobacter* and *Yersinia enterocolitica* in cattle, sheep and pigs at slaughter. *Epidemiol Infect* 137:1135–1148
- Herbert LJ, Vali L, Hoyle DV, Innocent G, McKendrick IJ, Pearce MC, Mellor D, Porphyre T, Locking M, Allison L, Hanson L, Matthews L, Gunn GJ, Woolhouse MEJ, Chase-Topping ME (2014) *E. coli* O157 on Scottish cattle farms: evidence of local spread and persistence using repeat cross-sectional data. *BMC Vet Res* 10:95
- Heuvelink AE, Zwartkruis-Nahuis JT, Beumer RR, de Boer E (1999) Occurrence and survival of verocytotoxin-producing *Escherichia coli* O157 in meats obtained from retail outlets in The Netherlands. *J Food Prot* 62:1115–1122
- Widgren S, Frössling J (2010) Spatio-temporal evaluation of cattle trade in Sweden: description of a grid network visualization technique. *Geospat Health* 5:119–130
- Widgren S, Söderlund R, Eriksson E, Fasth C, Aspan A, Emanuelson U, Alenius S, Lindberg A (2015) Longitudinal observational study over 38 months of verotoxigenic *Escherichia coli* O157:H7 status in 126 cattle herds. *Prev Vet Med* 121:343–352
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361
- Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals. Princeton University Press, New Jersey
- Besser TE, Richards BL, Rice DH, Hancock DD (2001) *Escherichia coli* O157:H7 infection of calves: infectious dose and direct contact transmission. *Epidemiol Infect* 127:555–560
- Cray WC Jr, Moon HW (1995) Experimental infection of calves and adult cattle with *Escherichia coli* O157:H7. *Appl Environ Microbiol* 61:1586–1590
- Nöremark M, Håkansson N, Lindström T, Wennergren U, Lewerin SS (2009) Spatial and temporal investigations of reported movements, births and deaths of cattle and pigs in Sweden. *Acta Vet Scand* 51:37
- Anonymous (2003) Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS). *Off J Eur Union L* 154:1–41
- Cleveland WS, Devlin SJ, Grosse E (1988) Regression by local fitting: methods, properties, and computational algorithms. *J Econom* 37:87–114
- Hastie T, Tibshirani R (1986) Generalized additive models. *Stat Sci* 1:297–310
- Widgren S, Bauer P, Engblom S (2016) SimInf: a framework for data-driven stochastic disease spread simulations. R package version 2.0.0. <https://www.CRAN.R-project.org/package=SimInf>. Accessed 28 Jul 2016
- R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Drawert B, Engblom S, Hellander A (2012) URDM: a modular framework for stochastic simulation of reaction-transport processes in complex geometries. *BMC Syst Biol* 6:76
- Engblom S, Ferm L, Hellander A, Lötstedt P (2009) Simulation of stochastic reaction-diffusion processes on unstructured meshes. *SIAM J Sci Comput* 31:1774–1797
- Kernighan BW, Ritchie DM (1988) The C programming language, 2nd edn. Prentice-Hall, New Jersey
- Eriksson E, Aspan A, Gunnarsson A, Vågsholm I (2005) Prevalence of verotoxin-producing *Escherichia coli* (VTEC) O157 in Swedish dairy herds. *Epidemiol Infect* 133:349–358
- Boqvist S, Aspan A, Eriksson E (2009) Prevalence of verotoxigenic *Escherichia coli* O157:H7 in fecal and ear samples from slaughtered cattle in Sweden. *J Food Prot* 72:1709–1712

41. Widgren S, Eriksson E, Aspan A, Emanuelson U, Alenius S, Lindberg A (2013) Environmental sampling for evaluating verotoxigenic *Escherichia coli* O157:H7 status in dairy cattle herds. *J Vet Diagn Invest* 25:189–198
42. Arnold ME, Ellis-Iversen J, Cook AJ, Davies RH, McLaren IM, Kay AC, Pritchard GC (2008) Investigation into the effectiveness of pooled fecal samples for detection of verocytotoxin-producing *Escherichia coli* O157 in cattle. *J Vet Diagn Invest* 20:21–27
43. Davis MA, Rice DH, Sheng H, Hancock DD, Besser TE, Cobbold R, Hovde CJ (2006) Comparison of cultures from rectoanal-junction mucosal swabs and feces for detection of *Escherichia coli* O157 in dairy heifers. *Appl Environ Microbiol* 72:3766–3770
44. Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7:308–313
45. Fenlon DR, Ogden ID, Vinten A, Svoboda I (2000) The fate of *Escherichia coli* and *E. coli* O157 in cattle slurry after application to land. *Symp Ser Soc Appl Microbiol* 88:149S–156S
46. Ogden LD, Fenlon DR, Vinten AJ, Lewis D (2001) The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. *Int J Food Microbiol* 66:111–117
47. Strachan NJ, Fenlon DR, Ogden ID (2001) Modelling the vector pathway and infection of humans in an environmental outbreak of *Escherichia coli* O157. *FEMS Microbiol Lett* 203:69–73
48. Anonymous (2009) Surveillance of zoonotic and other animal disease agents in Sweden 2009. National Veterinary Institute (SVA), Uppsala. http://www.sva.se/globalassets/redesign2011/pdf/om_sva/publikationer/trycksaker/1/surveillance2009.pdf. Accessed 31 Jul 2016
49. Anonymous (2012) Surveillance of infectious diseases in animals and humans in Sweden 2012. National Veterinary Institute (SVA), Uppsala. http://www.sva.se/globalassets/redesign2011/pdf/om_sva/publikationer/surveillance2012.pdf. Accessed 31 Jul 2016
50. Dutta BL, Ezanno P, Vergu E (2014) Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Prev Vet Med* 117:79–94
51. Masuda N, Holme P (2013) Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Rep* 5:6
52. Fefferman NH, Ng KL (2007) How disease models in static networks can fail to approximate disease in dynamic networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 76:031919
53. Green DM, Kiss IZ, Kao RR (2006) Modelling the initial spread of foot-and-mouth disease through animal movements. *Proc Biol Sci* 273:2729–2735
54. Vernon MC, Keeling MJ (2009) Representing the UK's cattle herd as static and dynamic networks. *Proc Biol Sci* 276:469–476
55. Gates MC, Humphry RW, Gunn GJ, Woolhouse ME (2014) Not all cows are epidemiologically equal: quantifying the risks of bovine viral diarrhoea virus (BVDV) transmission through cattle movements. *Vet Res* 45:110
56. Beaunee G, Vergu E, Ezanno P (2015) Modelling of paratuberculosis spread between dairy cattle farms at a regional scale. *Vet Res* 46:111
57. Bajardi P, Barrat A, Natale F, Savini L, Colizza V (2011) Dynamical patterns of cattle trade movements. *PLoS One* 6:e19869
58. Edrington TS, Callaway TR, Ives SE, Engler MJ, Loofer ML, Anderson RC, Nisbet DJ (2006) Seasonal shedding of *Escherichia coli* O157:H7 in ruminants: a new hypothesis. *Foodborne Pathog Dis* 3:413–421
59. Gautam R, Bani-Yaghoub M, Neill WH, Dopfer D, Kaspar C, Ivanek R (2011) Modeling the effect of seasonal variation in ambient temperature on the transmission dynamics of a pathogen with a free-living stage: example of *Escherichia coli* O157:H7 in a dairy herd. *Prev Vet Med* 102:10–21
60. Wang X, Gautam R, Pinedo PJ, Allen LJ, Ivanek R (2014) A stochastic model for transmission, extinction and outbreak of *Escherichia coli* O157:H7 in cattle as affected by ambient temperature and cleaning practices. *J Math Biol* 69:501–532
61. Matthews L, McKendrick IJ, Ternent H, Gunn GJ, Synge B, Woolhouse ME (2006) Super-shedding cattle and the transmission dynamics of *Escherichia coli* O157. *Epidemiol Infect* 134:131–142
62. Buttner K, Krieter J, Traulsen A, Traulsen I (2016) Epidemic spreading in an animal trade network—comparison of distance-based and network-based control measures. *Transbound Emerg Dis* 63:e122–e134

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

