Assessment of the spatio-temporal infection dynamics model of Salmonella in low prevalence regions

JRP6 - NOVA - FBZ1 - 1st Call

Responsible Partner: SVA

GENERAL INFORMATION

European Joint Programme full title	Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards
European Joint Programme acronym	One Health EJP
Funding	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.
Grant Agreement	Grant agreement n° 773830
Start Date	01/01/2018
Duration	60 Months

DOCUMENT MANAGEMENT

JIP/JRP Deliverable	D-4.3
Join Integrative/Research Project	JRP6 - NOVA - FBZ1 - 1st Call
JIP/JRP Leader	Jenny FROSSLING (SVA)
Other contributors	Arianna Comin (SVA), Hyeyoung Kim (SVA), Stefan Widgren (SVA), Estelle Ågren (SVA)
Due month of the deliverable	12
Actual submission month	12
Type *R: Document, report* *DEC: Websites, patent filings, videos, etc.* *OTHER*	R
Dissemination level *PU: Public* *CO: confidential, only for members of the consortium (including the Commission Services)*	CO

Contents

Deliverable 4.3.

Assessment of the spatio-temporal infection dynamics model of Salmonella in low prevalence regions

Salmonella has been selected as example of FBD as it is one of the most common public health problems, causing significant human morbidity and even mortality and consequently high economic losses in both developing and developed countries. Foods of animal origin are still one of the major sources of infection for the general public, with eggs, broiler chickens and pigs being consistently identified among the top attributed food sources. Whereas control programmes for Salmonella in poultry have been applied in the whole UE with high success, only few European countries have implemented eradication or control programmes of Salmonella in swine, beef, or dairy production. Results from efforts made in Denmark, Sweden, Finland, Norway, Ireland, Germany, Great Britain and Holland, are somewhat inconsistent. So far, there are no national control programmes established in any Mediterranean country, where >40% of the pig farms were positive to Salmonella (EFSA baseline study, 2009). In consequence, further efforts to implement control programs for reduction of the prevalence of Salmonella infection in swine in the near future are envisioned. Prerequisites to the implementation of such an approach; scientific efforts directed to improve our preparedness and develop an effective risk-based surveillance system should be carried out.

The deliverable D.4.3. "Assessment of the spatio-temporal infection dynamics model in Salmonella in low prevalence regions", belongs to Task-4.1 whose general objective is to understand the spatio-temporal patterns of infection distribution in livestock and slaughterhouses and its association with human cases to optimize sampling strategies and the implementation of risk based surveillance strategies under two different conditions: low prevalence regions (Subtask 4.1.1) and high prevalence regions (Subtask 4.1.2).

The task 4.1 includes a total of four activities and four deliverables:

T-4.1. Identification of spatial relationships and patterns in Salmonella prevalence (M1-M24)		
TASK/SUBTASK	**ACTIVITIES**	**DELIVERABLES**
Sub-Task 4.1.1: Surveillance in high prevalence regions to detect introduction and changes in prevalence.	A. Intensive pig farm location, industry surveys (slaughterhouse and feed co-operatives) and human cases will be investigated in conjunction with routinely recorded surveillance information using spatial techniques (e.g. smoothing technique, cluster analyses) [M1-M12].	D-4.1. Maps for Salmonella prevalence geographical patterns in intensive livestock and slaughterhouses completed in high prevalence regions (**M12**)
	B. Geographical areas (broad spatial trend and local spatial correlation) and periods with higher probability of detection of infection will be identified by temporal and spatial autocorrelation analyses, which will allow reallocating efforts on sampling strategies. In addition, temporal trends on serotype distribution and antimicrobial resistance profiles in isolates from clinical human cases and those found in swine will be compared [M13-M24].	D-4.2. Identification of periods with higher probability of detection of infection identified in high prevalence regions and temporal evidences for an association with human cases (M24)
Sub-Task 4.1.2: Surveillance in low prevalence regions to reduce prevalence.	C. A detailed model of the spatio-temporal infection dynamics will be applied based on data-driven simulations incorporating the complete population demographic, the time-varying contact animal network and the local spread among proximal holdings (Bauer 2016,	D-4.3. Assessment of the spatio-temporal infection dynamics model in Salmonella in low prevalence regions (**M12**)

		Widgren 2016a, Widgren 2016b). The model parameters will be calibrated against observed data from historical and ongoing monitoring [M1-M12].	
	D.	Optimal surveillance strategies will be explored to detect introduction and an increasing prevalence [M13-M24].	D-4.4. Evaluation of optimal surveillance strategies (M24)

Specifically, the objective of the deliverable D.4.3 was to assess the spatio-temporal infection dynamics model in Salmonella in low prevalence regions.

1. Background

Salmonella Dublin (*S.* Dublin) is a cattle-adapted *Salmonella* serotype. Infections usually cause clinical disease, increased calf mortality and decreased milk production (Nielsen, 2012a), resulting in substantial economic losses for the farmer (Nielsen, 2013a). *S.* Dublin also has a zoonotic potential, but infection in humans is uncommon. However, when it occurs, consequences are often more severe than those caused by other serotypes (Jones, 2008).

S. Dublin is one of the most common *Salmonella* serotypes infecting cattle populations in Europe, and some European countries have started control programs to reduce the prevalence, or even eradicate it (Bergevoet, 2009; Nielsen 2009). In Sweden, legislated *Salmonella* control was initiated already in the 1960s, including all serotypes and all animal species along the entire chain from feed to food. The program has resulted in a progressive reduction of *Salmonella* prevalence in Swedish cattle, but since the mid-1990s, the number of cattle herds detected within the control program has remained at a low but steady level (SVA, 2017). A national bulk milk screening in 2013 showed that only 1% of dairy herds were positive for *S.* Dublin antibodies, but the occurrence of the infection was strongly clustered. Most of the positive herds were in fact located on an island in the south-east, where the proportion of bulk milk positive dairy herds was as high as 15% (Ågren, 2016).

Species-adapted *Salmonella* serotypes have been shown to be possible to eradicate. However, the sensitivity of the Swedish surveillance system in cattle, which consists mainly of clinical passive surveillance, is considered too low to achieve eradication of *S*. Dublin. One possible way to increase the sensitivity of the surveillance for *S*. Dublin detection could be to perform repeated bulk milk samplings in dairy cattle (Warnick , 2006; Ågren, 2016)

In order to evaluate potential surveillance strategies, both the within- and between-herd transmission dynamics of *S.* Dublin infection need to be taken into account. Within-herd transmission of *Salmonella* spp. and *S.* Dublin in dairy herds have been quite extensively studied. However, approaches to model the between-herd spread of this infection are still scarce in the literature.

2. Objective

The aims of this first part of the project were:

a) to develop a two-stage transmission model for *S.* Dublin infection in dairy herds, suitable for evaluation of the effectiveness of different surveillance strategies in a hypoendemic context,

which combines the within-herd transmission via environmental contamination and the between-herd transmission through real animal movements.

b) to detect hidden variables that could affect the transmission of Salmonella by investigating spatial differences of environmental and social characteristics.

3. Progress of the activities: main results.

Data

Bulk milk screening

In these studies, results from bulk milk samples from Swedish dairy herds, collected at two occasions (2007 and 2013) were used.

In the autumn of 2007, 1069 samples for evaluation of a serological test were selected from a yearly national bulk milk screening for other reasons (approximately 7100 dairy herds in 2007 [10]). Selection of samples were performed by first omitting herds under restriction and thereafter every sixth sample was selected (Nyman et al., 2013).

In April 2013, a bulk milk screening for Salmonella was performed, including all Swedish dairy herds (in total 4 683 herds in 2013). Samples were retrieved from Eurofins laboratory in Jönköping, originally collected for routine milk quality testing.

The unfrozen samples with added preservative (bronopol) were sent to the Swedish National Veterinary Institute (NVI) where serological analyses were performed. The diagnostic test used was Prionics PrioCHECK® Salmonella Ab bovine Dublin ELISA (including O-antigens 1, 9 and 12; hereafter referred to as Dublin ELISA) (Thermo Fischer Scientific, Waltham, Massachusetts, United States). Information about the diagnostic sensitivity of this test has not been published, but the sensitivity of tests similar to the Dublin ELISA has been estimated to be within the range of 0.54 to 0.88 when used on a single bulk milk sample (Wedderkopp et al., 2001a; Veling et al., 2001) which is in agreement with Swedish experiences. The diagnostic specificity of both tests is considered to be close to 100% (Nyman et al., 2013). The recommendation by the producer is to use an optical density of 35 as a cut-off for a positive result. However, in order to increase sensitivity, the cut-off was lowered to 20. It has been shown that this change only causes a small decrease (~1%) in specificity when the test is used on Swedish bulk-milk samples. (Nyman et al., 2013).

Movement data

We obtained the Swedish livestock data from 37 221 holdings for the period 1 July 2005 to 31 December 2013 from the Swedish national cattle database managed by the Swedish Board of Agriculture. The dataset, described in detail in (Widgren, 2016), contained 18 649 921 events reported on an individual animal level. Briefly, the livestock data included the following information about each cow: (i) the date and the holding for its birth, (ii) the date and the source and destination holding for any movements, and (iii) the date for slaughter or death. The data was incorporated in the simulations to handle the population demographics and the time-varying contact network from livestock movements.

Geospatial Environmental data

We obtained topography data (elevation), hydrology data (water and wetness) and land cover data (imperviousness and vegetation index) from the pan-European component of Copernicus Land Monitoring Services, managed by the European Environment Agency (European Environment Agency, 2018). They are in raster format and the resolution is between 20m and 100m. We also acquired climate data (annual rainfall, monthly rainfall, annual temperature, and monthly temperature) from SMHI (Swedish Meteorological and Hydrological Institute). A detailed description of the processing of these data can be found in the section Data processing.

Social data

We hypothesized that non-environmental factors such as education and type of sewage system could also affect the disease transmission and collected the related statistical data from Statistics Sweden (http://www.scb.se). The data is provided in a tabular form as statistics aggregated to the municipality level.

Data processing

The data collected for the spatial analysis have different data formats: the topography and the hydrology, and the land cover data have raster format, the climate data have text format with coordinates, and the social data have text format aggregated to municipality level. To use these data for the analysis, a process of resampling into the same spatial unit is required. The finest Swedish spatial division that can be published publicly with statistical data is a municipality, and the areas of municipalities range from 8.67km^2 to 19,140.33km^2. A large difference in area among municipalities can cause distorted results when a spatial analysis is performed using aggregated data in this spatial unit. This issue is usually referred to as the modifiable areal unit problem (MAUP) (Openshaw and Taylor, 1979). To remove the zone effect, we used a grid unit with 25km by 25km resolution that provides the homogeneity of space units than the administrative spatial unit.

Table 1. The characteristics of the data and data processing

Variable	Data type	Resolution, m	Data conversion	Final resolution, km	Data source
Elevation, m	Raster	25	Resampling	25	Copernicus
Slope, °	Raster	25	Extraction from DEM	25	Copernicus
Vegetation Index	Raster	20	Resampling	25	Copernicus
Distance to major and permanent water, m	Raster	100	Resampling, Proximity	25	Copernicus
Imperviousness, %	Raster	100	Resampling	25	Copernicus
Mean monthly temperature (2013-04), °C	Text		Vectorizing (point), Interpolation	25	SMHI
Mean annual temperature (2013), °C	Text		Vectorizing (point), Interpolation	25	SMHI
Mean monthly Rainfall (2013-04), mm	Text		Vectorizing (point), Interpolation	25	SMHI
Mean annual rainfall (2013), mm	Text		Vectorizing (point), Interpolation	25	SMHI
Level of education, %, Municipality	Text		Vectorizing (polygon), Rasterizing	25	Statistics Sweden
Sewage types, %, Municipality	Text		Vectorizing (polygon) Rasterizing	25	Statistics Sweden

Table 1 lists the variables and describes the characteristics of the data and data processing. We resampled all the geospatial data to line them up correctly, consequently, all the data have the same grid size 25k by 25k. The slope of the study area was obtained from digital elevation model (DEM) by using QGIS 2.18 (QGIS, 2018). To retrieve the distance to the major and permanent water resources, we used water and wetness index data and calculated the Euclidean distance. We vectorized the climate data provided with spatial coordinates to point objects, then rasterized them using an interpolation method. Social data provided as an aggregated form into the administrative unit were transformed into polygonal vectors and then rasterized.

Disease spread modelling

Methods

SimInf framework

The disease spread modeling was performed in the R package SimInf (Widgren, 2018b). The SimInf framework is designed to efficiently simulate stochastic disease spread models in a large network of interconnected farms. The framework integrates infection dynamics in each farm as continuous-time Markov chains (CTMC) using the Gillespie stochastic simulation algorithm (Gillespie, 1977) and incorporates available data such as births, deaths or movements as scheduled events.

Within-herd spread

The spread of infection among the animals in a herd was modelled as a compartmental Susceptible-Infected-Carrier-Recovered model based on environmental transmission ($SICR_E$). These compartments stratify the population into four health states: healthy individuals capable of acquiring the infection (Susceptible, S); infected individuals capable of transmitting the infection (Infected, I); long term persistently infected individuals capable of transmitting the infection at a low rate (Carrier, C); healthy recovered individuals who developed immunity against the infection (Recovered, R). The four health states were further subdivided into three age groups: calves (< 6 months), young stock (6 - 30 months) and adults (> 30 months), to capture age related differences in infection dynamics within the host (Nielsen, 2012b; Nielsen 2013a; Nielsen 2013b).

The faecal-oral route of transmission of *Salmonella*, where susceptible individuals get the infection from the environment contaminated by faeces of infectious and carrier animals, was modelled by means of a time dependent environmental infectious pressure variable, uniformly distributed within each herd. It represents the number of bacterial cells per m^2 at any point in time and depends on the bacterial load shed by infected and carrier animals per surface area, and on the rate of bacterial decay (i.e. decimal reduction time) in the environment. Based on the results of Ågren at al. (2016) on the risk factors for a dairy herd to be infected by *S. Dublin*, the environmental infectious pressure was further set to increase proportionally to the number of infected holdings in the neighborhood.

Disease seasonality (Lewerin, 2011; Nielsen, 2012a) was modelled by letting the rate of bacterial decay vary according to the year quarters (Widgren, 2016).

Between-herd spread

The spread of infection between the cattle herds was modelled deterministically, based on the recorded event data described above. The raw data was converted into four categories of events − exit, enter, ageing and transfer − as described elsewhere (Widgren, 2016). The exit event happens when animals leave a holding due to slaughter, euthanasia or export. The enter event includes births

and imports. The ageing event happens the day animals change age group (i.e. from calf to young stock or young stock to adult). The transfer event occurs when animals are moved from one holding to another.

Simulations started the first day of recorded data (01/07/2005) and ended on the last one (31/12/2013). After every day of simulation, the number of animals in each health compartment of each age group and herd was updated according to the event database. In case of enter events, the reported number of animals were added to the susceptible compartment of the relevant age group. In case of ageing events, calves or young animals were removed from their current health compartment and added to the same health state in the next age group (i.e. young stock or adult, respectively) of the same herd. In case of exit events, animals were removed from the herd according to the reported age of the animal in the exit event, and randomly selected health states. In case of transfer events, animals were removed from the sending herd according to the same procedure as an exit event, and added to the receiving herd in the corresponding age-specific health compartments.

Model parametrization

Most of the parameters were considered age-specific and were mainly derived from literature. Given that our model did not assume direct transmission, it was not possible to translate the value of the *infection rate* parameter found in literature into the indirect transmission rate of the contaminated environment (i.e. *uptake rate*) used in our model, which was therefore derived by trial and error, assuming herd level endemic equilibrium (SVA, 2017). Demography within each herd was modelled based on recorded data, as described in the previous paragraph.

Model input

The simulation was initiated by supplying the initial state for each age-specific health compartment in every holding. Initial herd sizes were those reported on the first day of the recorded events. All the animals were considered to be susceptible. Five percent of the initially active dairy holdings (n=420) were artificially seeded as infected. They were randomly selected, proportionally to the county-level prevalence estimated from the national screening performed in 2013 (Ågren, 2016). Of the animals present in the infected herds, 8.2% of calves, 0.7% of young stock and 2.4% of the adults were allocated to the infectious state (SVA, 2013), and the environmental infectious pressure was calculated accordingly.

Model output

Each simulation produced one sample trajectory consisting of the number of individuals in the age-specific health compartments in each holding at weekly intervals. It was based on the whole cattle population and all the recorded events, but inference has been restricted to dairy holdings only. Because cattle holdings can change their production type over time, being a dairy holding was considered to be a dynamic status. Therefore, one holding contributes to the dairy sector only for the period of time it was an active dairy holding. A dairy holding was considered to be infected when it housed at least one infectious animal for at least twelve consecutive weeks. The reported results are based on 100 simulations.

Results and discussion

The modelled proportion of infected dairy herds in Sweden over the study period – as the result of the between-herd spread driven by the real animal movement data – stabilized around 1% after an initial burn-in period (Figure 1) , in accordance with the current knowledge (Ågren, 2016). The model seems to be quite robust to the choice of the starting values used for initialization, as the herd level prevalence eventually reaches the steady state anyway. Unfortunately, the range of available data does not allow to assess whether the final proportion of infected herds would have converged to 1% in all of the cases.

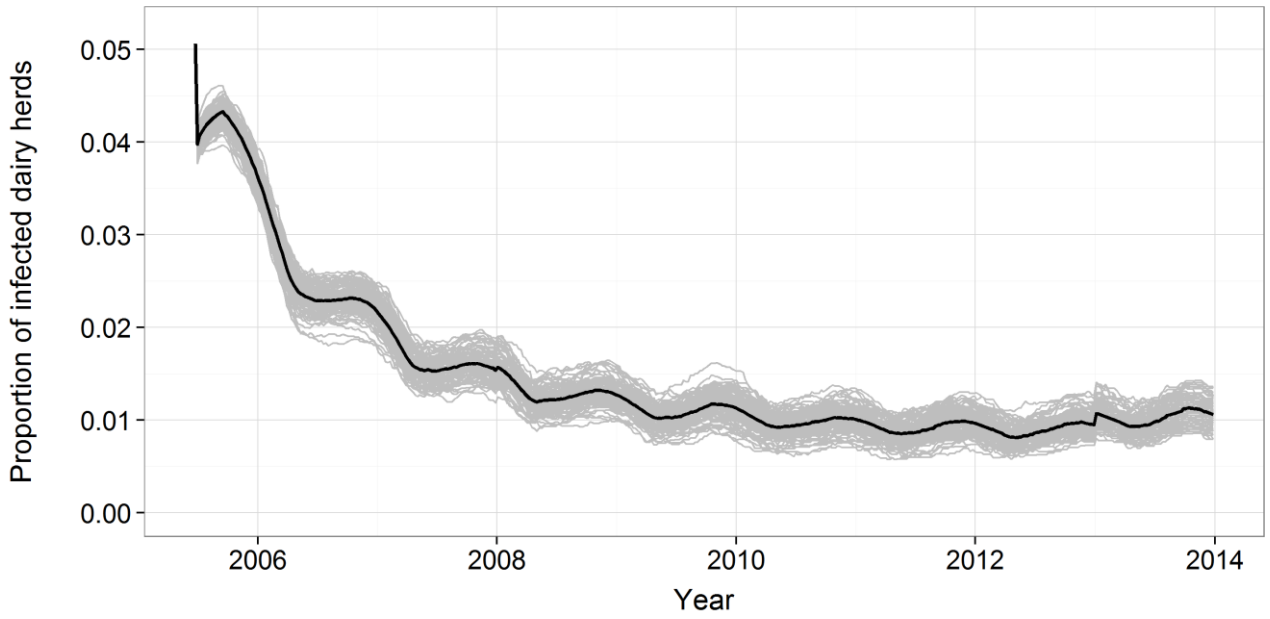

Figure 1. Simulated proportion of infected dairy herds over the years. Grey lines represent each simulation (n=100). The black line represents their average.

The within-herd infection dynamics by age group is summarized in Figure 2. It clearly shows a seasonal trend, as a consequence of the different rate of bacterial decay in different seasons. In particular, the within-herd prevalence peaks in the third quarter of the year and is higher for calves than for young and adult animals (Figure 3). On the other hand, the seroprevalence peak occurs in the last quarter, with young and adult cattle having the highest proportion of recovered animals (Figure 3).

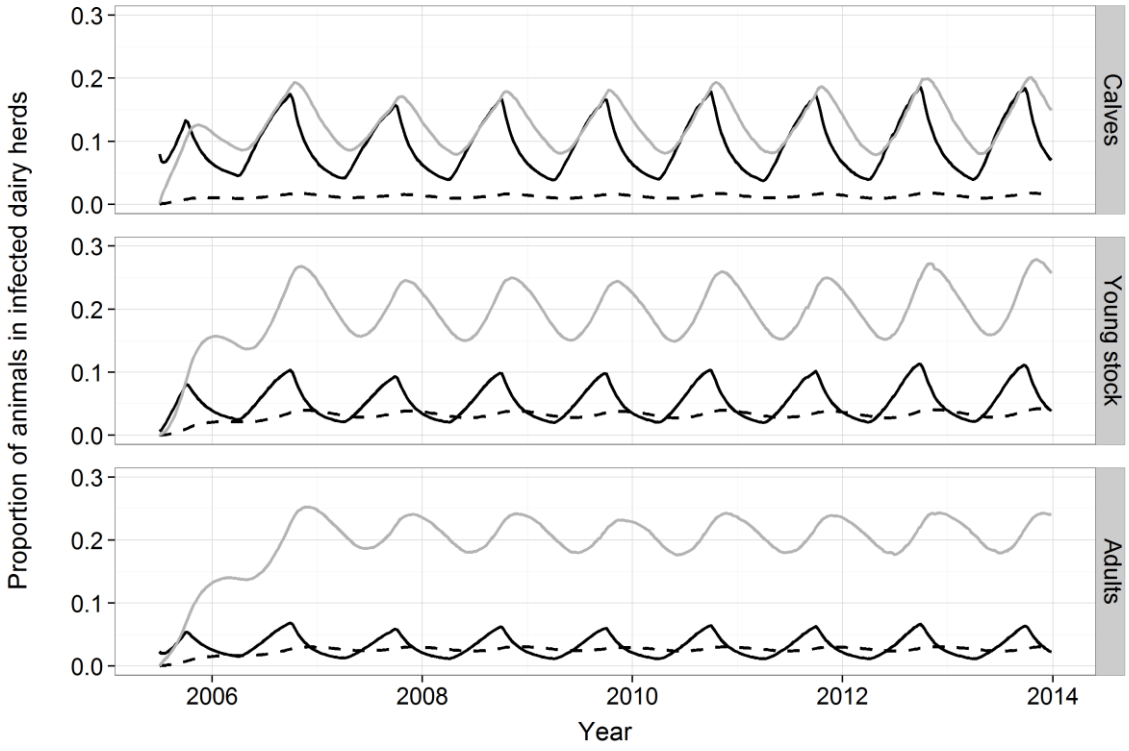

Figure 2. Average simulated within-herd infection dynamics by age group in infected dairy herds. Solid black lines (——) represent the proportion of infected animals. Dashed black lines (- - -) represent the proportion of carrier animals. Solid grey lines (——) represent the proportion of recovered animals. Result of 100 simulations.

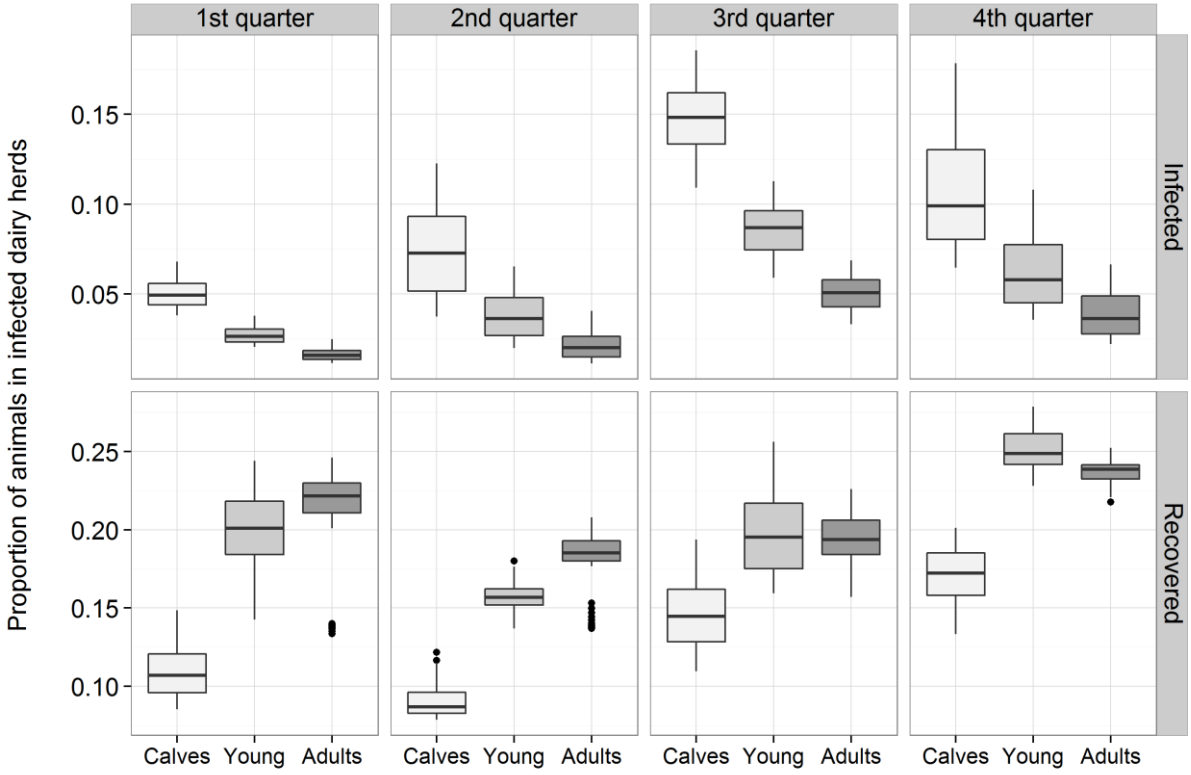

Figure 3. Proportion of infected animals (i.e. prevalence) and recovered animals (i.e. seroprevalence) in infected dairy herds by age group and quarter of the year. Result of 100 simulations using data from 01-01-2006 to 31-12-2013.

Although the model was able to reproduce the expected within-herd dynamics and the overall prevalence, it failed to entirely capture the disease clustering (Figure 4 and Table 2). Clustering is probably driven by some latent process that goes beyond animal movements and proportion of infected herds in the neighborhood and needs to be further addressed in order for the model to be useful to simulate surveillance strategies.

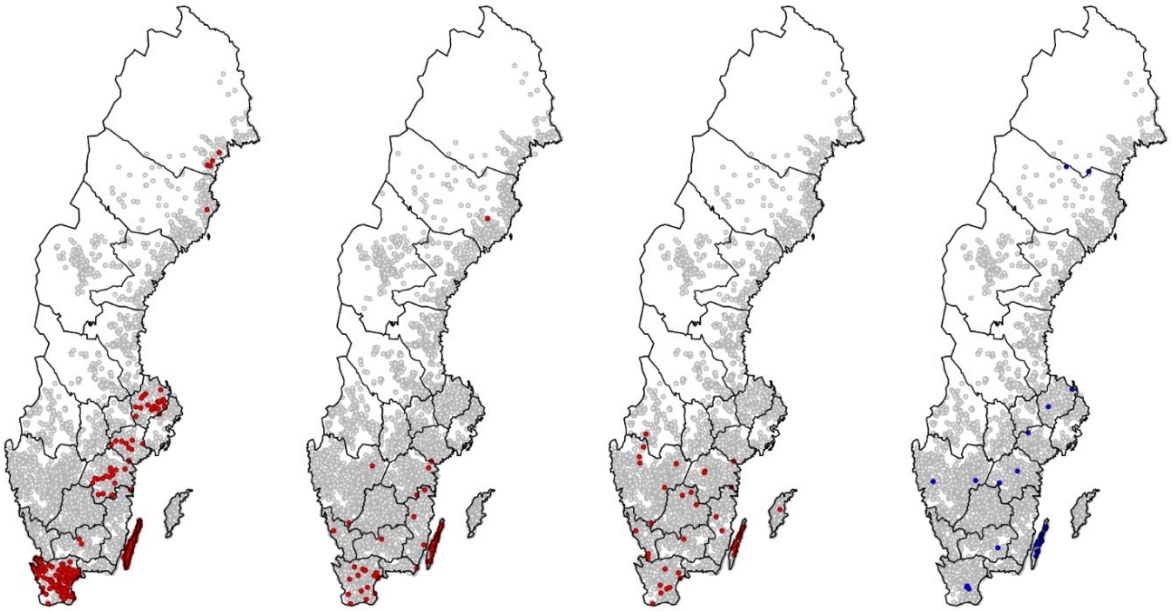

Figure 4. Geographical distribution of most frequently infected herds at begin (artificially seeded), middle and end of 100 simulations in comparison with results of bulk milk screening in 2013.

Table 2. Proportion of infected dairy herds (average of 100 simulations) in the county of Kalmar (where the island of Öland belongs) at different time points and in comparison to the actual results of the national bulk milk screening in 2013.

	median	**5th percentile**	**95th percentile**
Seeded infected herds (July 2005)	27.11% (= 5.42% scaled*)	25.42%	28.51%
Middle (January 2010)	5.51%	4.23%	6.80%
End (December 2013)	3.72%	2.71%	4.74%
Bulk milk screening (October 2013)	5.76%		

* At the beginning of simulation, the overall proportion of herds seeded as infected was intentionally inflated (i.e. 5% infected = 420 herds) in order to get the infection process taking place.

Spatial analyses

Methods

Spatial autocorrelation

We used the spatial autocorrelation (Global Moran's *I*) tool in GeoDa 1.12 (Anselin, Syabri and Kho, 2006) to identify statistically significant spatial clusters of *S.* Dublin prevalence. Global Moran's *I* (Moran, 1948) is given by

$$I = \frac{n}{W}\frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

where

n = the number of spatial units indexed by i and j

x = the variable of interest

$\bar{x}$ = the mean of x

$w_{ij} = \begin{cases} 1, if\ spatial\ units\ i\ and\ j\ are\ neighbours \\ 0, otherwise \end{cases}$ = a matrix of spatial weights

W = the sum of all w_{ij}.

The expected value of Moran's *I* is

$$E(I) = -\frac{1}{n-1}.$$

Moran's *I* value usually ranges from -1 to 1 and is compared with the expected value. When the value *I* exceeds -1/(*N*-1), it indicates positive spatial autocorrelation, and when the value *I* belows -1/(*N*-1), it indicates negative spatial autocorrelation. Moran's *I* value, however, does not identify the local cluster because it indicates global spatial autocorrelation.

To find clusters at a local level, we used local indicators of spatial association (LISA) (Anselin, 1995). For each spatial unit, LISA calculates Moran's *I* and evaluates the clustering in the individual units. Local Moran's *I* is defined by:

$$I_i = \frac{(x_i - \bar{x})}{m_2}\sum w_{ij}\left(x_j - \bar{x}\right),$$

where

$m_2 = \frac{\sum_i (x_i - \bar{x})^2}{N}$, and

N = the number of spatial units

then,

$$I = \sum_i \frac{I_i}{N}.$$

Spatial differences using ANOVA analysis

As we identified that the cluster of positive herds on the island in the south-east is very high (Ågren, 2016), we conducted the analysis of variance (ANOVA) to detect differences between the mainland and the island. ANOVA is used to examine potential differences among groups in data. In this analysis, we used one-way ANOVA, which allowed us to identify if there are differences in independent variables between the two regions.

Spatial regression

We assumed that the spatial pattern of *S*. Dublin prevalence is affected by the other variables in nearby places. To capture the effect of unmeasured independent variables on the spatial pattern of *S*. Dublin prevalence, we used the spatial lag model (Anselin, 2001) which is defined as

$$Y = \beta x_i + \varphi w_i y + e_i$$

where

wy = a vector of spatial lags for the dependent variables,

φ = spatial autoregressive coefficient,

βx = an $N \times K$ matrix of observations on the exogenous explanatory variables multiplied by a $K \times 1$ vector of regression coefficients β for each x,

e = a $N \times 1$ vector of normally distributed random error term.

Results and discussion

Spatial pattern of salmonella outbreaks in 2013

Through Moran's I test, we got the Moran's I value of 0.043 with p-value 0.04 (Figure 5). This result shows a very weak spatial autocorrelation of prevalence in Sweden. We then conducted LISA to investigate further and test for regional clustering. Figure 6 shows that there is a significant clustering in the south-east island. This result means that the pattern of the prevalence between the mainland and the island is completely different, and further investigation is necessary to understand the differences between the two regions.

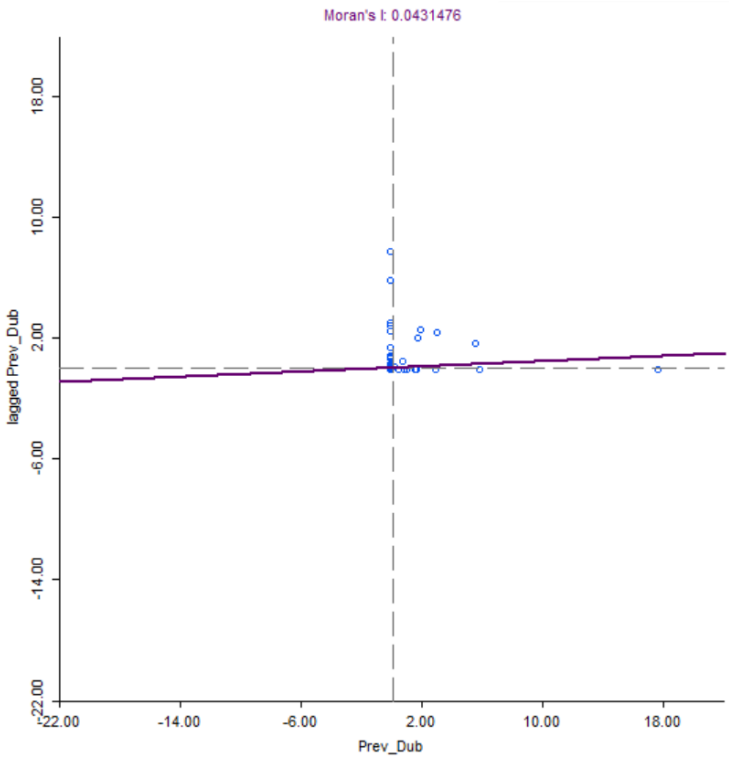

Figure 5. Moran scatter plot. The prevalence of S. Dublin is on the horizontal axis and the vertical axis shows the spatially lagged counterparts of the prevalence.

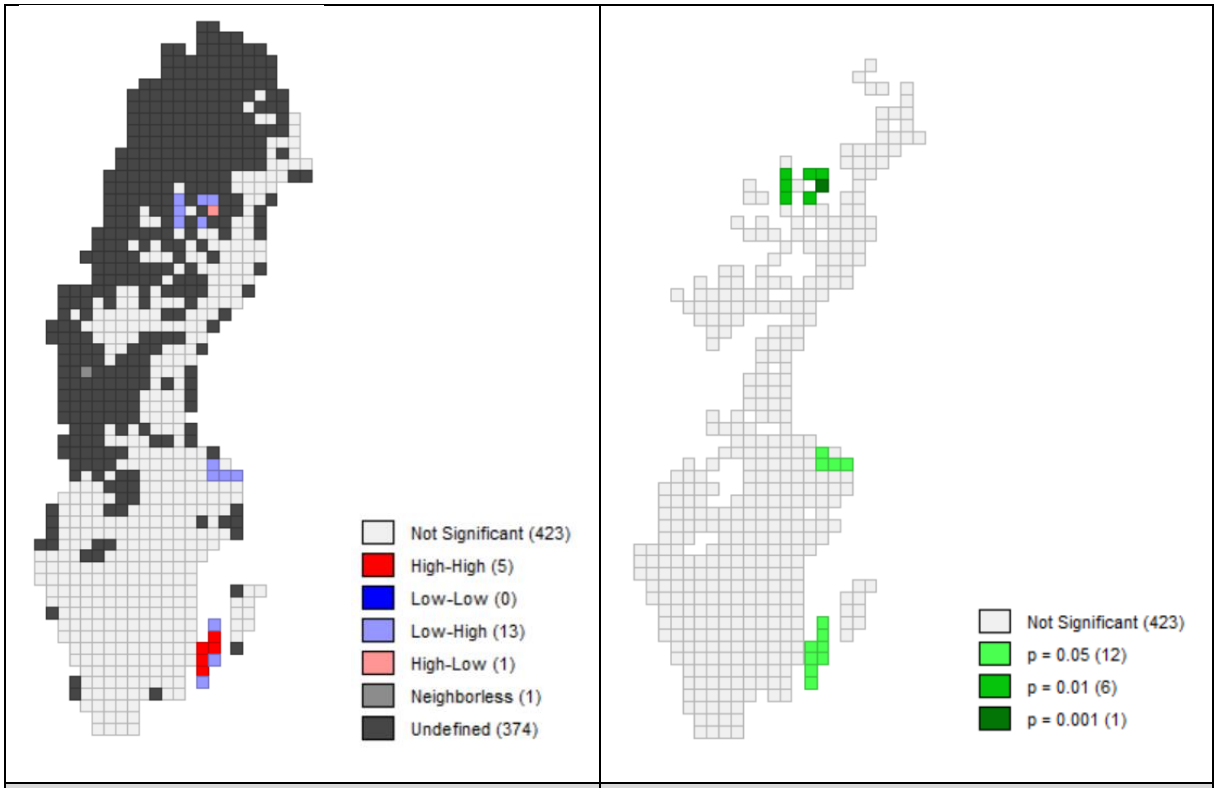

Figure 6. The LISA cluster map (left) shows how the prevalence clusters. The red color represents a strong cluster with high prevalence and blue represents a weak cluster with low prevalence. The LISA significance map (right) shows significant results by each grid.

Spatial heterogeneity

Table 3 shows the output of the ANOVA analysis and whether there is a statistically significant difference between the mainland and the island. The results revealed that there is a significant difference (i.e., p-value < 0.05) in the mean of elevation, slope, temperature, annual rainfall, and level of education between the two regions. On the other hand, vegetation index, distance to water resource, imperviousness, monthly precipitation, and sewage types showed no statistically significant difference between the two regions.

Spatial dependence

The univariable spatial lag model identified that an environmental variable (vegetation index) has a significant association with the prevalence of *S*. Dublin in the mainland (p-value<0.05) (Table 4). The island Öland, on the other hand, showed that an environmental variable (annual temperature) and non-environmental variables (level of education) have a significant association with the prevalence of *S*. Dublin.

Table 3. Results of ANOVA (t and p values)

Variable	t-Statistic	p-value
Elevation	-3.308	0.001
Slope	-4.042	0.000
Vegetation Index	-0.068	0.946
Distance to major and permanent water	-1.899	0.057
Imperviousness	0.187	0.852
Mean monthly temperature (2013-04)	3.512	0.000
Mean annual temperature (2013)	4.511	0.000
Mean monthly Rainfall (2013-04)	0.354	0.723
Mean annual rainfall (2013)	-2.780	0.006
Level of education: Primary+Secondary	0.634	0.523
Level of education: Upper secondary	-3.251	0.001
Level of education: Post-secondary	2.523	0.012
Level of education: Post graduate	9.861	0.000
Sewage types: Public disposal	-1.080	0.281
Sewage types: No sewage	2.561	0.011
Sewage types: Individual disposal	1.138	0.256

Table 4. Spatial regression analysis results with independent variables for the mainland and the island.

Region	Variable	Coefficient	Std.Error	z-value	p-value
Main land	Elevation	0.000	0.000	1.821	0.069
	Slope	-0.001	0.002	-0.405	0.686
	Vegetation Index	-0.066	0.029	-2.269	0.023
	Distance to major water	0.000	0.000	0.018	0.986
	Imperviousness	-0.001	0.002	-0.649	0.517
	Mean monthly temperature (2013-04)	-0.003	0.002	-1.864	0.062
	Mean annual temperature (2013)	-0.002	0.001	-1.781	0.075
	Mean monthly Rainfall (2013-04)	-0.001	0.000	-0.485	0.628
	Mean annual rainfall (2013)	-0.001	0.000	-0.337	0.736
	Level of education: Primary+Secondary	0.038	0.091	0.425	0.671
	Level of education: Upper secondary	0.106	0.083	1.273	0.203
	Level of education: Post-secondary	-0.064	0.062	-1.039	0.299
	Level of education: Post graduate	-0.202	0.554	-0.364	0.719
	Sewage types: Public disposal	-0.026	0.030	-0.866	0.386
	Sewage types: No sewage	0.745	0.421	1.771	0.077
	Sewage types: Individual disposal	0.023	0.030	0.743	0.457
Öland	Elevation	0.001	0.007	0.256	0.798
	Slope	0.073	0.135	0.541	0.588
	Vegetation Index	-0.436	0.563	-0.775	0.438
	Distance to major water	0.000	0.000	0.801	0.423
	Imperviousness	0.007	0.060	0.122	0.902
	Mean monthly temperature (2013-04)	0.155	0.119	1.296	0.195
	Mean annual temperature (2013)	-0.352	0.174	-2.027	0.043
	Mean monthly Rainfall (2013-04)	-0.001	0.017	-0.085	0.933
	Mean annual rainfall (2013)	0.004	0.008	0.501	0.616
	Level of education: primary+Secondary	3.256	1.492	2.182	0.029
	Level of education: Upper secondary	7.309	3.610	2.025	0.043
	Level of education: Post-secondary	-2.539	1.230	-2.064	0.039
	Level of education: Post graduate	-1.597	0.743	-2.151	0.031
	Sewage types: Public disposal	-0.690	0.359	-1.922	0.055
	Sewage types: No sewage	10.65	7.131	1.494	0.135
	Sewage types: Individual disposal	0.690	0.359	1.922	0.055

4. Conclusion and future direction

Degree of achievement: Deliverable D.4.3. has been achieved.

Main conclusions: The developed disease spread model enabled a better understanding of *S*. Dublin infection spread in Swedish dairy herds, as it related to herd population dynamics and time-varying trade patterns between farms.

This model will be improved to further simulate surveillance strategies, as it fails – in its current form – to fully capture disease clustering.

The next step will therefore be to explore different options to improve the model, such as:

(i) Using more sophisticated approaches for the parameterization of the disease spread model, for example, a Bayesian framework with suitable prior information on the parameter values.
(ii) Reformulate the spatial coupling among herds to better capture the between-herd transmission unrelated to cattle movements.
(iii) Include spatial and environmental factors affecting disease spread in the model parameters by region.

References

Anselin, L (1995). Local indicators of spatial association—LISA. Geographical analysis, 27(2), 93-115.

Anselin, L (2001). Spatial econometrics. A companion to theoretical econometrics, 310330.

Anselin L, Syabri I and Kho Y (2006). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1): 5-22.

Bergevoet, RH, van Schaik G, Veling J, Backus GB, Franken P(2009) Economic and epidemiological evaluation of salmonella control in Dutch dairy herds. Prev Vet Med 89(1-2): 1-7.

Gillespie DT (1977) Exact Stochastic Simulation of Coupled Chemical Reactions. The Journal of Physical Chemistry, 81(25), 2340–2361. DOI:10.1021/j100540a008.

Jones TF, Ingram LA, Cieslak PR, Vugia DJ, Tobin-D'Angelo M, Hurd S, Medus C, Cronquist A, Angulo FJ (2008). Salmonellosis outcomes differ substantially by serotype. Journal of Infectious Diseases 198(1): 109-114.

Lewerin SS, Skog L, Frössling J, Wahlström H (2011). Geographical distribution of salmonella infected pig, cattle and sheep herds in sweden 1993-2010. Acta Vet Scand 53: 51.

Nielsen LR, Danish Veterinary and Food Administration (2009). Current salmonella-control in cattle in denmark.Available at: http://www.fodevarestyrelsen.dk/NR/rdonlyres/24AEB101-ABAC-405C-A0EF-693B8E4B36F2/15348/MicrosoftWordBilag7FINALCurrentSalmonellacontrolin.pdf (Accessed 28/01/2016).

Nielsen TD, Green LE, Kudahl AB, Ostergaard S, Nielsen LR (2012a). Evaluation of milk yield losses associated with Salmonella antibodies in bulk tank milk in bovine dairy herds. J Dairy Sci 95(9):4873-85.

Nielsen LR, Kudahl AB, Ostergaard S (2012b). Age-structured dynamic, stochastic and mechanistic simulation model of salmonella dublin infection within dairy herds. Prev Vet Med 105(1-2): 59-74.

Nielsen TD, Kudahl AB, Ostergaard S, Nielsen LR (2013a). Gross margin losses due to Salmonella Dublin infection in Danish dairy cattle herds estimated by simulation modelling. Prev Vet Med 111(1-2):51-62.

Nielsen LR (2013a). Within-herd prevalence of *Salmonella* Dublin in endemically infected dairy herds. Epidemiol Infect 141(10): 2074-2082.

Nielsen LR (2013b). Review of pathogenesis and diagnostic methods of immediate relevance for epidemiology and control of salmonella dublin in cattle. Vet Microbiol 162(1): 1-9.

Nyman AKJ, Ågren ECC, Bergström K, Wahlström H (2013). Evaluation of the specificity of three enzyme-linked immunosorbent assays for detection of antibodies against Salmonella in bovine bulk milk. Acta Veterinaria Scandinavica 55.

Openshaw S, and Taylor P (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Statistical Applications in the Spatial Sciences. Pion, London. 127-144.

QGIS Development Team (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

SVA - Swedish National Veterinary Institute (2013). "Yttrande angående slakt av nötkreatur från besättningar med serologisk reaktion avseende salmonella i tankmjölk". Accession number: 6.3.17-5622/13, pp 7 (In Swedish).

SVA - Swedish National Veterinary Institute (2017). Surveillance of infectious diseases in animals and and humans in sweden 2017. Available at: https://www.sva.se/globalassets/redesign2011/pdf/om_sva/publikationer/surveillance-2017-w.pdf (Accessed 19/11/2018)

Warnick LD, Nielsen LR, Nielsen J, and Greiner M (2006). Simulation model estimates of test accuracy and predictive values for the Danish salmonella surveillance program in dairy herds. Prev Vet Med 77(3-4): 284-303.

Widgren S., Engblom S., Bauer P., Frössling J., Emanuelson U., Lindberg A. (2016) Data-driven network modeling of disease transmission using complete population movement data: Spread of VTEC O157 in Swedish cattle. Vet Res 47:81. DOI: 10.1186/s13567-016-0366-5

Widgren S., Engblom S., Emanuelson U., Lindberg A. (2018a) Spatio-temporal modelling of verotoxigenic Escherichia coli O157 in cattle in Sweden: Exploring options for control. Vet Res 49:78. DOI: 10.1186/s13567-018-0574-2

Widgren S., Bauer P., Eriksson R., Engblom S. (2018b) SimInf: An R package for Data-driven Stochastic Disease Spread Simulations. arXiv preprint arXiv:1605.01421 [q-bio.PE], URL http://arxiv.org/abs/1605.01421

Ågren EC, Lewerin SS, Wahlström H, Emanuelson U, Frössling J (2016). Low prevalence of Salmonella in Swedish dairy herds highlight differences between serotypes. Prev Vet Med 1(125):38-45.