

JASON ONELL ARDILA GALVIS

**Spatial analysis of the bovine tuberculosis and animal
movement network in the state of Espírito Santo, Brazil**

São Paulo

2019

JASON ONELL ARDILA GALVIS

Spatial analysis of the bovine tuberculosis and animal movement network in the state of Espírito Santo, Brazil

Thesis presented to the Postgraduate Studies Program in Experimental Epidemiology Applied to Zoonoses of the School of Veterinary Medicine and Animal Science of the University of São Paulo to obtain the title of Doctor in Science.

Department:

Preventive Veterinary Medicine and Animal Health

Concentration area:

Experimental Epidemiology Applied to Zoonoses

Advisor:

Prof. Dr. José Henrique Hildebrand Grisi Filho

In agreement: _____

Advisor

São Paulo
2019

Total or partial reproduction of this work is permitted for academic purposes with the proper attribution of authorship and ownership of the rights.

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO

(Biblioteca Virgínie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo)

T. 3835
FMVZ

Galvis, Jason Onell Ardila
Spatial analysis of the bovine tuberculosis and animal movement network in the state of Espírito Santo, Brazil / Jason Onell Ardila Galvis. – 2019.
100 f. : il.

Título traduzido: Análise espacial da Tuberculose Bovina e redes de movimentação animal no Estado do Espírito Santo, Brasil.

Tese (Doutorado) – Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Medicina Veterinária Preventiva e Saúde Animal, São Paulo, 2019.

Programa de Pós-Graduação: Epidemiologia Experimental Aplicada às Zoonoses.
Área de concentração: Epidemiologia Experimental Aplicada às Zoonoses.
Orientador: Prof. Dr. José Henrique Hildebrand Grisi Filho.

1. Análise de redes. 2. Análise espacial. 3. Zoonoses. 4. Fazendas de leite. 5. Pecuária.
I. Título.



São Paulo, 13 de agosto de 2019
CEUAX N 8713070316

Ilmo(a). Sr(a).

Responsável: José Henrique De Hildebrand E Grisi Filho

Área: Medicina Veterinária Preventiva E Saúde Animal

Equipe envolvida: Jason Onell Ardila Galvis - executor (universidade De São Paulo); José Henrique De Hildebrand E Grisi Filho (orientador)

Título do projeto: "Análise espacial da Tuberculose Bovina e redes de movimentação animal no Estado do Espírito Santo, Brasil".

Parecer Consubstanciado da CEUA FMVZ

o estudo objetiva avaliar a distribuição espacial das propriedades positivas para a tuberculose e os fatores de risco associados aos parâmetros da rede de movimentação bovina no Estado de Espírito Santo, Brasil.

A Comissão de Ética no Uso de Animais da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, na reunião de 03/08/2016, **ANALISOU** e **APROVOU** o protocolo de estudo acima referenciado. A partir desta data, é dever do pesquisador:

1. Comunicar toda e qualquer alteração do protocolo.
2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do protocolo.
3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.
4. **Relatórios parciais** de andamento deverão ser enviados **anualmente** à CEUA até a conclusão do protocolo.

Profa. Dra. Anneliese de Souza Traldi
Presidente da Comissão de Ética no Uso de Animais
Faculdade de Medicina Veterinária e Zootecnia da Universidade
de São Paulo

Roseli da Costa Gomes
Secretária
Faculdade de Medicina Veterinária e Zootecnia da Universidade
de São Paulo

EVALUATION SHEET

Author: ARDILA GALVIS, Jason Onell

Title: **Spatial analysis of the bovine tuberculosis and animal movement network in the state of Espírito Santo, Brazil**

Thesis presented to the Postgraduate Studies Program in Experimental Epidemiology Applied to Zoonoses of the School of Veterinary Medicine and Animal Science of the University of São Paulo to obtain the title of Doctor in Science.

Date: ____/____/____

Examining Committee

Prof. Dr. _____

Institution: _____ Judgment: _____

Dedication

Dedico este estudio a mi querida familia y a todas las personas que me apoyaron a lo largo de mis estudios de Maestría y Doctorado.

Acknowledgment

I want to express my sincere gratitude to my supervisor the Prof. Dr. José Henrique Hildebrand Grisi Filho, who took care of me since my internship, and supported me in each decision I made along my master and doctoral studies.

Thanks to my committee members to help me improve my research

Thanks to the professors from LEB: Fernando Ferreira, Marcos Amaku, Ricardo Dias and Oswaldo Baquero for the advices given to me in the moments I needed.

Thanks to my friends and labmates Nicolas (el niko) Cardenas and Daniel Magalhães, thanks for your patient.

Thanks to my friends and previous students from LEB that I met in my first years in Brazil: Alexandre Guerra, Alice de Carvalho, Aline Alves, Amanda Oliveira, Ana Pérola Drulla, Ana Rita Pinheiro, Camila Martins, Dejelia Gomez, Fernando Marques, Francisco Ulloa, Gina Polo, Guilherme Braga, Isabella Fontana, Mariana Queiroz, Raul (Raulinho Gaucho) Ossada, Tatiana Villegas, Anaiá da Paixão, Rafael Cipullo. You are amazing and without you my first years in Brazil would not have been the same.

Thanks to my new friends from LEB that I met along my PhD: Alfredo Acosta, Aluisio Filho, Germana Vizzotto, Guilherme Camargo, João Revolta, Juliana Tozzi, Clebiano da Costa, Maíra Marques, Winni Alves, Yeison Caicedo.

Thanks to the department' secretary, workers and professors from VPS for all the help given to me in these years.

Thanks to the University of São Paulo (USP) and the School of Veterinary Medicine and Animal Science (FMVZ) for giving me the opportunity to reach this goal of my life.

Thanks to my friends: André Pegoraro, Carlos Cabrera, Claudia Carranza, Eliana Cubillos, Joyce Lopes, Juan Valencia, Nelson Santana, Sebastian Muñoz, Vasco Túlio.

Thanks to the people from Massey Institute, University of Melbourne and Roslin Institute, the experiences lived there helped me to grow as a better researcher.

Thanks to the grant nº 2016/03441-6, São Paulo Research Foundation (FAPESP) 2016/03441-6).

Thanks to all the people I probably forgot to put here, you all know me, I am always on the clouds, but I am sincerely grateful with all of you.

Thanks to everyone, it was an amazing experience.

Abstract

ARDILA GALVIS, J. O. **Spatial analysis of the bovine tuberculosis and animal movement network in the state of Espírito Santo, Brazil.** [Análise espacial da Tuberculose Bovina e redes de movimentação animal no Estado do Espírito Santo, Brasil]. 2019. 100 p. Tese (Doutorado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2019.

Brazil is one of the countries that has the strongest bovine production worldwide, with 2.5 thousand million farms and 178 thousand million bovines distributed in its national territory. The number of farms and animals in the country present a major challenge to animal disease control since have high surveillance coverage in this scenario is expensive and inefficient process. Therefore, it is necessary to identify and focus the resources on groups of farms with an increased risk of infection and which strongly contribute to disseminate a disease widely in the national territory. The main objective of the present study is to provide information and tools which may help decision makers to build efficient animal disease control programs. In general, in the present study we focused on the surveillance of bovine tuberculosis (bTB) and used the state of Espírito Santo (ES) in southeast Brazil to apply our methodologies. We divided this study into four sections: (1) provide information on the situation of bTB in ES and farms' characteristics associated with this infectious disease; (2) complement the previous study with the identification of risk factors associated with the characteristic of animal transit and the distribution of cases of the disease that were reported in ES; (3) describe the characteristic of the cattle transit in ES and propose a methodology for targeted surveillance of farms based on the structure of the trade network; (4) develop a methodology to predict farm locations based on the spatial information of trade partners. In conclusion, this study provided information and tools which can help to the decision makers to complement animal disease surveillance programs in ES and Brazil.

Keywords: network analysis, spatial analysis, zoonotic, dairy farms, livestock.

Resumo

ARDILA GALVIS, J. O. **Análise espacial da Tuberculose Bovina e redes de movimentação animal no Estado do Espírito Santo, Brasil.** [Spatial analysis of the bovine tuberculosis and animal movement network in the state of Espírito Santo, Brazil.]. 2019. 100 f. Tese (Doutorado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2019.

O Brasil tem uma das maiores produções bovinas no mundo, composta por 2,5 milhões de fazendas e 178 milhões de bovinos distribuídos em seu território. Este grande número de fazendas e animais representam um desafio importante no controle de doenças animais, principalmente porque manter uma alta cobertura e sensibilidade de vigilância neste cenário se torna um processo caro e ineficiente. Assim, é necessário identificar e focar os recursos no grupo de fazendas com maior risco de serem infectadas e que possam contribuir de forma mais significativa na dinâmica de uma doença. O principal objetivo deste estudo foi prover informação e ferramentas que possam auxiliar os responsáveis pela vigilância sanitária a construir planos mais eficientes no controle de doenças. Em geral, neste estudo focamos na vigilância da Tuberculose Bovina (TB) e aplicamos nossas metodologias no estado de Espírito Santo (ES). Dividimos o estudo em 4 seções: (1) geração de informação sobre a situação da tuberculose bovina no ES e identificação de características das fazendas associadas com esta doença; (2) complementação do estudo anterior com a identificação de riscos associados com medidas de centralidades da rede de comércio de bovinos e distribuição espacial dos casos positivos a TB no E; (3) descrição das características da rede de trânsito de bovinos no ES e avaliação de uma metodologia de vigilância para identificar fazendas baseada na estrutura da rede de comércio de bovinos; (4) desenvolvimento de uma metodologia para prever coordenadas das fazendas baseada na localização espacial dos parceiros comerciais. Em conclusão, este estudo fornece informações e ferramentas que podem ajudar a complementar os programas de vigilância de doenças animais no ES e no Brasil.

Palavras chave: análise de redes, análise espacial, zoonoses, fazendas de leite, pecuária

Index

General introduction.....	12
Chapter 1: Epidemiologic characterization of bovine tuberculosis in the state of Espírito Santo, Brazil.....	13
Introduction	14
Materials and Methods.....	15
Sample design	15
Test protocol	16
Prevalence estimation.....	17
Risk factor analysis	17
Results	18
Discussion.....	21
Conclusion	22
References.....	23
Chapter 2: Risk Analysis for Bovine Tuberculosis in Espírito Santo, Brazil: an approach based on spatial and network analysis.	28
Introduction	29
Materials and methods	31
Study area	31
Bovine Tuberculosis database.....	31
Animal movement database	32
Network construction	32
Risk analysis based on the centrality metrics from the network.....	32
Risk analysis based on spatial dispersion.....	33
Results	35
Risk analysis based on the centrality metrics from the network.....	35
Risk analysis based on spatial dispersion.....	37
Discussion.....	38
References.....	41
Supplementary Material	44
Chapter 3: Epidemiological characterization and target surveillance of the cattle trade network in Espírito Santo, Brazil.....	51
Introduction	52
Methodology.....	53
Dataset	53
Network Analysis.....	53

Classification of the herds in groups	55
Network evolution over the years	56
Fragmentation of the GSCC.....	56
Comparison of herd ranking between different periods of time.....	57
Disease spread and control simulation	58
Results	60
Characteristics of the cattle network and evolution over the years.....	60
Fragmentation of the GSCC and GWCC.....	63
Disease Simulation	66
Discussion.....	67
References.....	71
Supplementary material.....	76
Chapter 4: Cattle farm spatial location recovery through animal trade movements.....	80
Introduction	81
Methodology.....	82
Data set	82
Recovery of missing farms	82
Evaluating the disease spread model with the predicted coordinates.....	84
Results	86
Farm location prediction accuracy.....	88
Disease spread model comparison	90
Discussion.....	90
References.....	96
Supplementary Material	98
General conclusion.....	99
General references.....	100

General introduction

Brazil is one of the countries that has the strongest cattle production worldwide, with 2.5 thousand million farms and 178 thousand million heads distributed throughout its national territory (IBGE, 2017). The number of farms and animals in the country present a major challenge to animal disease control since it is not easy to establish a strong surveillance program in each of these farms. Therefore, it is necessary to identify and focus the resources on groups of farms with an increased risk of infection and which strongly contribute to disseminate a disease widely in the national territory. Thus, it will be possible to create more efficient control measures and promote advances in disease prevention (LeBlanc, Lissemore, Kelton, Duffield, & Leslie, 2006). In general, in the present study we focused on the surveillance of bovine tuberculosis (bTB) which is an endemic bacterial disease widely distributed in Brazil that has a negative impact on the economy of country (Ferreira Neto et al., 2016; LAGE et al., 2006). We used the state of Espírito Santo (ES) in southeast Brazil to apply our methodologies. Throughout the study we discuss how the methodologies can be applied to other diseases in other Brazilian states.

We divided this study into four sections. In **Chapter 1**, titled “Epidemiologic characterization of bovine tuberculosis in the state of Espírito Santo, Brazil”, we provide information on the situation of bTB in ES and farms’ characteristics associated with this infectious disease. In **Chapter 2**, titled “Risk analysis for Bovine Tuberculosis in Espírito Santo, Brazil: an approach based on spatial and network analyses”, we complemented the previous study with the identification of risk factors associated with the characteristic of animal transit and the distribution of cases of the disease that were reported in ES. In **Chapter 3**, titled “Epidemiological characterization and target surveillance of cattle trade network in Espírito Santo, Brazil”, we describe the characteristic of the cattle transit in ES and propose a methodology for targeted surveillance of farms based on the structure of the trade network and the role of each farm to connect it. In **Chapter 4**, titled “Cattle farm spatial location recovery through animal trade movements”, we describe a methodology to predict farm locations based on the spatial information of trade partners. The main objective of the present study is to provide information and tools which may help decision makers to build efficient animal disease control programs.

Chapter 1: Epidemiologic characterization of bovine tuberculosis in the state of Espírito Santo, Brazil

Galvis, J. O. A., Grisi-Filho, J. H. H., Costa, D. Da, Said, A. L. P. R., Amaku, M., Dias, R. A., ... Ferreira Neto, J. S. (2016). Epidemiologic characterization of bovine tuberculosis in the State of Espírito Santo, Brazil. *Semina: Ciências Agrárias*, 37(5Supl2), 3567. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3567>

Abstract

A cross-sectional study was carried out between January 2012 and May 2014 to investigate the status of bovine tuberculosis in the state of Espírito Santo. The state was divided into two regions, and in each of them, 300 farms with reproductive activity were randomly selected and considered as the primary sampling units. In the selected farms, a fixed number of female bovines aged over 2 years were randomly selected to undergo a comparative cervical tuberculin test; an epidemiologic questionnaire was also applied. In the state of Espírito Santo, the apparent prevalence of tuberculosis-positive farms was 7.6% (95% confidence interval [CI] = 5.7-9.9). Prevalence at the herd level varied from 4.6% (95% CI = 2.6-7.3) in region 1 to 11.1% (95% CI = 7.7-15.3) in region 2. The apparent prevalence of tuberculosis-positive animals was 0.7% (95% CI = 0.3-1.1) in the state, and the prevalence varied from 0.3% (95% CI = 0.2-0.6) in region 1 to 1.2% (95% CI = 0.3-2.9) in region 2. The risk factors associated with tuberculosis prevalence in Espírito Santo were the number of adult females ≥ 10 (odds ratio [OR] = 2.40; 95% CI = 1.17-5.31) and milking type (milking machine/milking parlor) (OR = 2.88; 95% CI = 1.36-5.86)]. The state of Espírito Santo should set up a surveillance system to detect and control bovine tuberculosis, taking into account the importance of dairy farms and animal trade in the state.

Keywords: Bovine; Brazil; Espírito Santo; Prevalence; Risk factor; Tuberculosis.

Introduction

Bovine tuberculosis (bTB) is a chronic zoonotic disease caused by *Mycobacterium bovis*, which in addition to cattle, also infects small ruminants, humans, and other domestic and wild animals. The bacterium initiates a chronic, granulomatous, caseous-necrotising, inflammatory process in many organs depending on its route of entry. The main route of entry is inhalation (causing primary lesions in the lungs and accessory lymph nodes), but it can also be ingested (causing primary lesions in the mesenteric lymph nodes) (DOMINGO et al., 2014). Apart from the risk to public health, the disease causes economic losses to the cattle industry (ZINSSTAG et al., 2006). Usually, bTB is endemic and widespread in large territories connected by commercial relations, especially animal trade (GILBERT et al., 2005).

In 2001, the Brazilian Ministry of Agriculture, Livestock and Food Supply (MAPA) launched the National Program for the Control and Eradication of Animal Brucellosis and Tuberculosis (PNCEBT), with an aim to reduce the negative impacts of bovine brucellosis and tuberculosis on human health, as well as to promote the competitiveness of the national livestock industry (LAGE et al., 2006). The control measures recommended by the PNCEBT included movement control of reproducing animals and certification of tuberculosis-free farms (LAGE et al., 2006). This includes diagnostic tests and veterinary services, which are paid for by the farmers. The official veterinary service is only responsible for auditing the entire system.

Recent studies carried out in 12 Brazilian states, which hold 74% of the Brazilian cattle population, showed that the prevalence of tuberculosis-infected herds ranged from 0.36% in the Federal District to 9.0% in São Paulo (BAHIENSE et al., 2016; BARBIERI et al., 2016; DIAS et al., 2016; GUEDES et al., 2016; LIMA et al., 2016; NÉSPOLI et al., 2016; QUEIROZ et al., 2016; RIBEIRO et al., 2016; ROCHA et al., 2016; SILVA et al., 2016; VELOSO et al., 2016; VENDRAME et al., 2016). The state of Espírito Santo has a herd of 2,216,144 heads, of which 43.9% is concentrated in the north of the state-a region in which beef cattle farms are predominant. Dairy farming is practiced in the southern region of the state-a region which has a more traditional organization, with a predominance of small farms (BRASIL, 2015).

To date, no survey has tried to accurately characterize bTB in the entire territory of the state of Espírito Santo. Lavagnoli et al. (2010) collated data on routine tests for tuberculosis conducted throughout Espírito Santo between January 2005 and June 2007, and found a bTB prevalence of 0.54% at the individual animal level. However, that study used non-random sampling and could have yielded biased results.

From January 2009 to December 2012, the percentage of bovines that had to be slaughtered under the provisions of state Inspection because of the presence of macroscopic lesions suspected of being induced by bTB was 0.23%. Among the 1260 suspected bTB lesions detected during the abattoir inspection, 1026 (81.4%) were in bovines from the southern region of the state, which has a predominantly dairy and dual-purpose livestock, and which regularly sends discarded animals for slaughter (PAOLI, 2013).

The present study aimed to estimate bTB prevalence and risk factors in the state of Espírito Santo in order to better inform the management of the animal health policy. As stated by Skuce et al. (2012), "Risk factors will vary across regions due to factors such as differing farm structures, farm management practices, local TB control, and the relative importance of specific risk factors within individual areas." Therefore, local epidemiological studies like the one reported here are extremely important to understand tuberculosis dynamics, spread, and maintenance within a region, and, consequently, to better inform decision makers and control agencies on how to deal with this important disease.

Materials and Methods

Sample design

The present study was designed by the MAPA; the Collaborator Centre in Animal Health, School of Veterinary Medicine, University of São Paulo (FMVZ-USP); and the Animal Health Service of the state of Espírito Santo (Instituto de Defesa Agropecuária e Florestal do Espírito Santo-IDAF). The field work was performed by IDAF staff from January 2012 to May 2014, after being trained to standardize the procedures.

In order to characterize regional differences in epidemiological parameters of bTB, the state of Espírito Santo was divided into two regions, taking into account the livestock production systems, management practice, herd size, and trade systems according to Azevedo et al. (2009). This division also considered the operational and logistic capacity of the IDAF to perform the field work based around its 40 regional offices. A map of the livestock regions was plotted using the software ArcGIS 10.0.

In each region, a cross-sectional study was performed to estimate the herd- and animal-level prevalence of bTB by using a two-stage sampling method. In the first stage, a predetermined number of farms with reproductive activity were randomly

selected (primary sampling units). In the second stage, a predetermined number of female bovines aged over 2 years were randomly selected (secondary sampling units).

In farms with more than one individual herd, we chose the biggest, more economically important one, in which the animals were submitted to the same management procedures, i.e., animals exposed to identical risk factors. The choice of the primary sampling units was based on the official farm registry database. If a selected farm could not be visited by IDAF staff, a new one was randomly selected to replace it. The number of selected farms per livestock region was estimated by the simple random sample formula proposed by Thrusfield (2007) and Noordhuizen et al. (2001):

$$N = \frac{Z_{\alpha}^2 * P * (1 - P)}{d^2}$$

where N denotes the sample size, Z_{α} is the normal distribution value for the 95% confidence level, P is the expected prevalence (20%), and d is the absolute error (5%).

The choice of the secondary sampling units aimed at the appropriate classification of a farm as positive or negative. To achieve this, we used the herd sensitivity and specificity concept (DOHOO et al., 2003). The values for sensitivity and specificity for the test protocol were 80% and 99.5%, respectively (FLETCHER et al., 1998), and the expected prevalence was considered to be 20%. The calculations were made using Herdacc version 3, and the selected sample size was the one that allowed herd sensitivity and specificity above 90%. Thus, in farms with 99 or less females aged over 2 years, 20 animals were sampled, and in farms with more than 99 females aged over 2 years, 40 animals were sampled. If the selected herd was smaller than the required sample size, all animals were sampled. The sampling method for the selection of females was systematic.

If up to 20 animals in a herd were tested, and if there was at least one bTB-positive animal, the herd was classified as an infected herd. If 40 animals in a herd were tested, at least two animals had to test positive for the herd to be classified as an infected herd.

Test protocol

The comparative cervical tuberculin test was used according to the guidelines of the Brazilian Brucellosis and Tuberculosis Control and Eradication National Program (LAGE et al., 2006). Tested animals were separated from the herd for 72 h, when the results were being analysed. Animals with inconclusive results were retested 60 days

after the first test. If the second test result was inconclusive or positive, the animal was considered positive according to the guidelines of the national program (LAGE et al., 2006).

Prevalence estimation

The sample design allowed us to estimate the herd- and animal-level prevalence of bTB in Espírito Santo and in the livestock regions as well. The apparent prevalence and the respective confidence intervals (CIs) were estimated according to Lumley (2004). All calculations were weighted (DOHOO et al., 2003). The weight of each farm (W_f) in the calculation of the positive herd prevalence in the whole state was derived as follows:

$$W_f = \frac{\text{number of farms in the region}}{\text{number of sampled farms in the region}}$$

The weight of each female bovine aged over 2 years (W_a) in the calculation of animal prevalence in the whole state was derived as follows:

$$W_a = \frac{\text{females} \geq 2 \text{ years in the farm}}{\text{sampled females} \geq 2 \text{ years in the farm}} \\ * \frac{\text{females} \geq 2 \text{ years in the region}}{\text{females} \geq 2 \text{ years in the sampled farms of the region}}$$

In the above expression, the first term refers to the weight of each sampled animal in the farm and the second term refers to the weight of each farm in the livestock region. The prevalence estimates and respective 95% CIs were calculated using R (R DEVELOPMENT CORE TEAM, 2012) along with the “survey” package (LUMLEY, 2004).

Risk factor analysis

In each sampled farm, a questionnaire was applied in order to generate data about its management practices. All information generated was archived in a database.

In this cross-sectional study, risk factors such as production system (meat, milk, or mixed), raising system (extensive or any degree of confinement), cattle breeds, number of cows, total herd size, presence of other domesticated species, presence of wild animals, animal trade, routine of tuberculosis testing, slaughter in the farm, pasture sharing, feeding calves with whey, indirect contact between farms, flooded pastures, and veterinary assistance were assessed according to the literature (MARANGON et al., 1998; SKUCE et al., 2012). These variables were organized on an increasing risk

scale. When necessary, a recategorization was performed. The least-risk category was always considered the baseline for comparisons with the other categories. Quantitative variables were categorized using the median as cutoff.

An exploratory univariate analysis was performed using the chi-square (χ^2) or Fisher exact test with all variables, taking into account all of the data obtained from the entire state. Variables with a significance level of less than 0.20 were selected for use in a multivariate analysis using logistic regression, performed according to Hosmer and Lemeshow (2000). All calculations were conducted using R (R DEVELOPMENT CORE TEAM, 2012).

Results

The state of Espírito Santo was divided into two livestock regions (Figure 1), according to Azevedo et al. (AZEVEDO et al., 2009). The apparent prevalence of bTB-positive herds in the state was estimated to be 7.6% (95% CI = 5.7-9.9), ranging from 4.6% (95% CI = 2.6-7.3) in region 1 to 11.1% (95% CI = 7.7-15.3) in region 2. The results showed that the prevalence of bTB-positive farms was significantly different between the two regions (Table 1).

Figure 1. Map of the state of Espírito Santo showing the division of the two livestock regions, 2013.

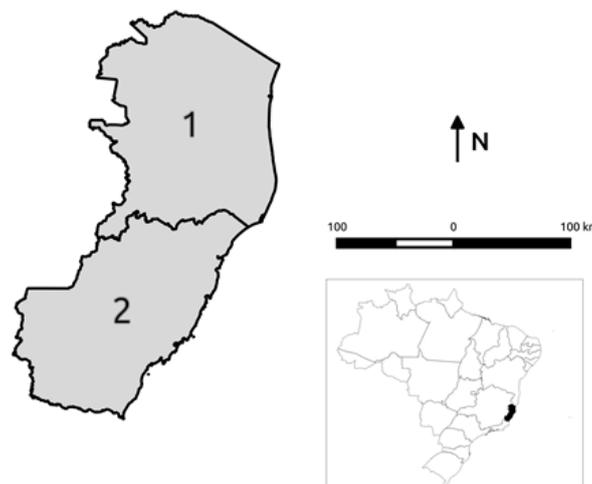


Table 1. Apparent prevalence of bovine tuberculosis at the herd level in the state of Espírito Santo, 2013.

Region	Farms with reproductive activities	Sampled farms	Positive farms	Prevalence (%)	95% CI (%)
1	17,496	349	16	4.6	2.6-7.3
2	15,100	288	32	11.1	7.7-15.3
Total	32,596	637	48	7.6	5.7-9.9

Moreover, no significant differences in the herd-level apparent prevalence were observed between farm enterprises (beef, mixed, and dairy) and for each of the farm enterprises between the regions (Table 2). The apparent prevalence of bTB-positive females aged over 2 years in the state of Espírito Santo was estimated to be 0.7% (95% CI = 0.3-1.1), ranging from 0.3% (95% CI = 0.2-0.6) in region 1 to 1.2% (95% CI = 0.3-2.9) in region 2 (Table 3).

Table 2. Apparent prevalence of bovine tuberculosis at the herd level, categorized by farm enterprise type, in the state of Espírito Santo, 2013.

Region	Beef cattle			Dairy cattle			Mixed cattle		
	n	%	95% CI	n	%	95% CI	n	%	95% CI
1	0/38	0.0	0.0-7.4*	12/201	6.0	3.1-10.2	3/106	2.8	0.6-8.1
2	2/30	6.7	0.8-22.1	27/195	13.9	9.3-19.5	3/62	4.8	1.0-13.5

*Estimated with the Beta distribution, according to Vose (2008).

Table 3. Apparent prevalence of tuberculosis-positive female bovines aged over 2 years in the state of Espírito Santo, 2013.

Region	Females > 2 years of age	Sampled females > 2 years of age	Positive animals	Prevalence (%)	95% CI (%)
1	660,512	4,589	20	0.3	0.2-0.6
2	383,789	3,110	52	1.2	0.3-2.9
Total	1,044,301	7,699	72	0.7	0.3-1.1

The results of the univariate analysis, showing risk factors with $p < 0.20$ are presented in Table 4, and the results of the multivariate analysis are presented in Table 5. The final multivariate model indicated two risk factors: number of adult females ≥ 10 (odds ratio [OR] = 2.4; 95% CI = 1.17-5.31) and milking type (having a milking parlor or portable milking machine; OR = 2.88; 95% CI = 1.36-5.86).

Table 4. Univariate analysis of the risk factors ($p \leq 0.20$) for bovine tuberculosis in the state of Espírito Santo, 2013.

Variable	Proportion of infected herds	%	<i>p</i>
Milking type			0.001
No milking/Manual milking	31/547	5.7	
Milking machine/Milking parlor	13/71	18.3	
Number of females > 2 years			0.004
1-10*	12/285	4.2	
≥ 10	36/352	10.2	
Number of milkings per day			0.004
No milking	5/95	5.3	
1	27/445	6.1	
2-3	15/89	16.9	
Milk refrigeration			0.011
No	19/366	5.2	
Yes	26/240	10.8	
Farm enterprise			0.011
Beef	2/68	2.9	
Dairy	39/396	9.8	
Mixed	6/168	3.6	
Raising system			0.015
Extensive	37/559	6.6	
Any degree of confinement	11/71	15.5	
Acquisition of breeding animals**			0.021
No	22/386	5.7	
Yes	26/240	10.8	
Veterinary assistance			0.033
No	27/439	6.2	
Yes	21/187	11.2	
Breed			0.046
Beef	1/48	2.1	
European dairy	3/26	11.5	
European beef	3/10	30	
Mixed-breed	38/513	7.4	
Others	2/21	9.5	
Acquisition of animals**			0.073
No	19/328	5.8	
Yes	29/301	9.6	
Bulk tank			0.14
Owned	15/102	14.7	
Shared	11/138	8	
Presence of wild animals			0.18
Yes	30/338	8.9	
No	18/299	6	
Equipment, feedstock, or personal sharing			0.183
Yes	3/83	3.6	
No	45/554	8.1	

*Median. **During the previous 12 months.

Table 5. Final multivariate model of the risk factors for bovine tuberculosis in the state of Espírito Santo, 2013.

Variable	Odds ratio	95% CI	<i>p</i>
Number of females > 2 years of age			
1-9 (reference category)	-	-	-
≥10	2.40	1.17 – 5.31	0.022
Milking type			
No milking/Manual (reference category)	-	-	-
Milking machine/Milking parlor	2.88	1.36 – 5.86	0.004

$r^2 = 6.9\%$.

Discussion

The prevalence of bovine herds infected with bTB in Espírito Santo was estimated to be 7.6% (95% CI = 5.7-9.9) (Table 2), which was statistically equal to the prevalence in the state of São Paulo and higher than that in the states of Bahia, Pernambuco, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Mato Grosso, Rondônia, Goiás, Minas Gerais, and the Federal District (BAHIENSE et al., 2016; BARBIERI et al., 2016; DIAS et al., 2016; GUEDES et al., 2016; LIMA et al., 2016; NÉSPOLI et al., 2016; QUEIROZ et al., 2016; RIBEIRO et al., 2016; ROCHA et al., 2016; SILVA et al., 2016; VELOSO et al., 2016; VENDRAME et al., 2016). Moreover, in the state of Espírito Santo, the prevalence of bTB-infected herds was higher in dairy farms in both livestock regions, but it was significantly higher in region 2, which has more dairy cattle than does region 1 (Tables 1 and 2) (LAVAGNOLI et al., 2010). The prevalence of bTB-positive animals was also higher in region 2 than in region 1, but the confidence intervals for these proportions overlapped greatly (Table 3).

Among the different farm enterprises, dairy farms have been reported previously as risk factors for bTB (ALVAREZ et al., 2012; GRISI-FILHO et al., 2011; KAROLEMEAS et al., 2011; PORPHYRE et al., 2008; RAMÍREZ-VILLAESCUSA et al., 2010). This relationship may be attributed to the dairy cattle's longer life expectancy and higher production stress, as well as to their gathering in large numbers during milking (HUMBLET et al., 2009). We also found this relationship (Table 4), but the variable did not make it to the final logistic model (Table 5), mainly because of its association with another more important variable-milking type. This latter variable is interesting, since it indicates two risk factors. First, a beef farm typically does not milk its cows, and thus its herds are in the baseline category of this variable. Second, dairy farms with milking machines or milking parlors have better technological infrastructure,

and as such are more likely to have a higher animal density than its counterparts; this increases the risk of bTB transmission and maintenance in a farm, as previously identified in Belgium (HUMBLET et al., 2010). Moreover, enterprises with milking machines or milking parlors have shown a higher probability of purchasing breeding animals in the past 12 months (Table 6). Animal purchase has been found to be a risk factor in previous studies (BESSELL et al., 2012; CARRIQUE-MAS et al., 2008; CLEGG et al., 2012; OLOYA et al., 2007; RAMÍREZ-VILLAESCUSA et al., 2010; REILLY; COURTENAY, 2007; SKUCE et al., 2012; TSCHOPP et al., 2009); therefore, this variable is probably increasing the statistical significance of the milking type variable even further.

Table 6. Proportion of farms that purchased breeding animals in the past 12 months, according to its milking type, in the state of Espírito Santo, 2013.

Purchase of breeding animals	No milking/Manual milking	Milking machine/ Milking parlor	Total
No	64%	41%	62%
Yes	36%	59%	38%

The number of adult females in a herd was also found as a risk factor in the final model (Table 5). The relationship between herd size and bTB infection has already been found in several studies (BESSELL et al., 2012; CLEGG et al., 2012; GREEN; CORNELL, 2005; GRIFFIN et al., 1996; PORPHYRE et al., 2008; RAMÍREZ-VILLAESCUSA et al., 2010; REILLY; COURTENAY, 2007; SKUCE et al., 2012). The two main mechanisms underlying this finding are the increased risk induced by purchasing animals, since bigger herds are typically more likely to purchase animals (CIPULLO et al., 2016), and the greater persistence of the disease (BROOKS-POLLOCK; KEELING, 2009).

Conclusion

The southern region of Espírito Santo has a higher bTB prevalence at the herd level (11%) probably because of its higher concentration of dairy farms, whereas bTB prevalence in the northern region is moderate (5%). Herd size and milking type (having a milking machine or milking parlor) were found to be risk factors for bTB infection in herds. The latter probably indicates the influence of other variables, such as enterprise type (dairy), animal density, and animal introduction. On the basis of these findings, we recommend that the abovementioned risk factors should be incorporated into the design of risk-based surveillance and control programs.

References

- Alvarez, J.; Perez, A. M.; Bezos, J.; Casal, C.; Romero, B.; Rodriguez-Campos, S.; Aezlloriente, J. L.; Diaz, R.; Carpintero, J.; Juan, L. de; Domínguez, L. Eradication of bovine tuberculosis at a herd-level in Madrid, Spain: study of within-herd transmission dynamics over a 12 year period. *BMC Veterinary Research*, London, v. 8, n. 1, p. 100-107, 2012.
- Azevedo, S. S.; Ferreira Neto, J. S.; Dias, R. A.; Ferreira, F.; Amaku, M.; Figueiredo, V. C. F.; Lôbo, J. R.; Gonçalves, V. S. P.; Souza, A. C.; Vasconcellos, S. A. Situação epidemiológica da brucelose bovina no Estado do Espírito Santo. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, Belo Horizonte, v. 61, n. 1, p. 19-26, 2009.
- Bahiense, L.; Bavia, M. E.; Amaku, M.; Dias, R. A.; Grisi-Filho, J. H. H.; Ferreira, F.; Telles, E. O.; Gonçalves, V. S. P.; Heinemann, M. B.; Ferreira Neto, J. S. Prevalence and risk factors for bovine tuberculosis in the State of Bahia, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3549-3560, 2016. Suplemento 2.
- Barbieri, J. M.; Oliveira, L. F.; Dorneles, E. M. S.; Mota, A. L. A. A.; Gonçalves, V. S. P.; Maluf, P. P.; Ferreira Neto, J. S.; Ferreira, F.; Dias, R. A.; Telles, E. O.; Grisi-Filho, J. H. H.; Heinemann, M. B.; Amaku, M.; Lage, A. P. Epidemiological status of bovine tuberculosis in the state of Minas Gerais, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3531-3548, 2016. Suplemento 2.
- Bessell, P. R.; Orton, R.; White, P. C. L.; Hutchings, M. R.; Kao, R. R. Risk factors for bovine Tuberculosis at the national level in Great Britain. *BMC Veterinary Research*, London, v. 8, n. 1, p. 51-60, 2012.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Dados de rebanho bovino e bubalino do Brasil. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2015. 1 p. Disponível em: . Acesso em: 24 mar. 2015.
- Brooks-Pollock, E.; Keeling, M. Herd size and bovine tuberculosis persistence in cattle farms in Great Britain. *Preventive Veterinary Medicine*, Amsterdam, v. 92, n. 4, p. 360-365, 2009.
- Carrique-Mas, J. J.; Medley, G. F.; Green, L. E. Risks for bovine tuberculosis in British cattle farms restocked after the foot and mouth disease epidemic of 2001. *Preventive Veterinary Medicine*, Amsterdam, v. 84, n. 1-2, p. 85-93, 2008.
- Cipullo, R. I.; Grisi-Filho, J. H. H.; Dias, R. A.; Ferreira, F.; Ferreira Neto, J. S.; Gonçalves, V. S. P.; Marques, F. S.; Negreiros, R. L.; Ossada, R.; Amaku, M. Cattle movement network, herd size and bovine brucellosis in the state of Mato Grosso, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3777-3792, 2016. Suplemento 2.
- Clegg, T. A.; Blake, M.; Healy, R.; Good, M.; Higgins, I. M.; More, S. J. The impact of animal introductions during herd restrictions on future herdlevel bovine tuberculosis risk. *Preventive Veterinary Medicine*, Amsterdam, v. 109, n. 3-4, p. 246-257, 2012.
- Dias, R. A.; Stanojlovic, F. M. U.; Belchior, A. P. C.; Ferreira, R. S.; Gonçalves, R. C.; Barão, R. S. C.; Sousa, P. R.; Santos, A. M. A.; Amaku, M.; Ferreira, F.; Telles,

- E. O.; Grisi-Filho, J. H. H.; Gonçalves, V. S. P.; Heinemann, M. B.; Ferreira Neto, J. S. Prevalence and risk factors for bovine tuberculosis in the state of São Paulo, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3673-3684, 2016. Suplemento 2.
- Dohoo, I.; Martin, W.; Stryhn, H. *Veterinary epidemiologic research*. Charlottetown: Atlantic Veterinary College, 2003. 706 p.
- Domingo, M.; Vidal, E.; Marco, A. Pathology of bovine tuberculosis. *Research in Veterinary Science, London*, v. 97, p. S20-S29, 2014. Supplement. Available at: . Accessed at: 05 July 2016.
- Fletcher, R. H.; Fletcher, S. W.; Wagner, E. H. *Clinical epidemiology: the essentials*. 2th ed. Baltimore: Williams & Wilkins, 1998. 246 p.
- Gilbert, M.; Mitchell, A.; Bourn, D.; Mawdsley, J.; Clifton-Hadley, R.; Wint, W. Cattle movements and bovine tuberculosis in Great Britain. *Nature, London*, v. 435, n. 7041, p. 491-496, 2005.
- Green, L. E.; Cornell, S. J. Investigations of cattle herd breakdowns with bovine tuberculosis in four counties of England and Wales using VETNET data. *Preventive Veterinary Medicine, Amsterdam*, v. 70, n. 3-4, p. 293-311, 2005.
- Griffin, J. M.; Martin, S. W.; Thorburn, M. A.; Eves, J. A.; Hammond, R. F. A case-control study on the association of selected risk factors with the occurrence of bovine tuberculosis in the Republic of Ireland. *Preventive Veterinary Medicine, Amsterdam*, v. 27, n. 3-4, p. 217-229, 1996.
- Grisi-Filho, J. H. H.; Rosales, C. A. R.; Ferreira, F.; Amaku, M.; Dias, R. A.; Ferreira Neto, J. S. Análise epidemiológica das condenações de bovinos por tuberculose em abatedouros do Estado de São Paulo. *Arquivos do Instituto Biológico, São Paulo*, v. 78, n. 2, p. 175-181, 2011.
- Guedes, I. B.; Bottene, I. F. N.; Monteiro, L. A. R. C.; Leal Filho, J. M.; Heinemann, M. B.; Amaku, M.; Grisi-Filho, J. H. H.; Dias, R. A.; Ferreira, F.; Telles, E. O.; Gonçalves, V. S. P.; Ferreira Neto, J. S. Prevalence and risk factors for bovine tuberculosis in the state of Mato Grosso do Sul, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3579-3588, 2016. Suplemento 2.
- Hosmer, D. W.; Lemeshow, S. *Applied logistic regression*. Hoboken: Wiley, 2000. v. 2, 375 p. Humblet, M. F.; Boschiroli, M. L.; Saegerman, C. Classification of worldwide bovine tuberculosis risk factors in cattle: a stratified approach. *Veterinary Research, Paris*, v. 40, n. 5, p. 40-50, 2009.
- Humblet, M. F.; Gilbert, M.; Govaerts, M.; Fauville-Dufaux, M.; Walravens, K.; Saegerman, C. New assessment of bovine tuberculosis risk factors in Belgium based on nationwide molecular epidemiology. *Journal of Clinical Microbiology, Washington*, v. 48, n. 8, p. 2802-2808, 2010.
- Karolemeas, K.; Mckinley, T. J.; Cliftonhadley, R. S.; Goodchild, A. V.; Mitchell, A.; Johnston, W. T.; Conlan, A. J. K.; Donnelly, C. A.; Wood, J. L. N. Recurrence of bovine tuberculosis breakdowns in Great Britain: Risk factors and prediction. *Preventive Veterinary Medicine, Amsterdam*, v. 102, n. 1, p. 22-29, 2011.
- Lage, A. P.; Roxo, E.; Müller, E.; Poester, F.; Cavalléro, J. C. M.; Ferreira Neto, J. S.; Mota, P. M. P. C.; Gonçalves, V. S. P. Programa nacional de controle e

erradicação da brucelose e da tuberculose animal (PNCEBT). Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2006. 188 p. (Manual técnico).

- Lavagnoli, M. R.; Amorim, B. M. de; Machado, G. P.; Demoner, L. de C.; Zanini, M. S.; Antunes, J. M. A. de P. Tuberculosis in bovines from Espírito Santo State. *Veterinária e Zootecnia, Botucatu*, v. 17, n. 1, p. 71-78, 2010.
- Lima, P. B.; Nascimento, D. L.; Almeida, E. C.; Pontual, K. A. Q.; Amaku, M.; Dias, R. A.; Ferreira, F.; Gonçalves, V. S. P.; Telles, E. O.; Grisi-Filho, J. H. H.; Heinemann, M. B.; Silva, J. C. R.; Ferreira Neto, J. S. Epidemiological situation of bovine tuberculosis in the state of Pernambuco, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3601-3610, 2016. Suplemento 2.
- Lumley, T. Analysis of complex survey samples. *Journal of Statistical Software, Los Angeles*, v. 9, n. 8, p. 1-19, 2004.
- Marangon, S.; Martini, M.; Dalla Pozza, M.; Ferreira Neto, J. A case-control study on bovine tuberculosis in the Veneto Region (Italy). *Preventive Veterinary Medicine, Amsterdam*, v. 34, n. 2-3, p. 87-95, 1998.
- Néspoli, J. M. B.; Negreiros, R. L.; Amaku, M.; Dias, R. A.; Ferreira, F.; Telles, E. O.; Heinemann, M. B.; Grisi-Filho, J. H. H.; Gonçalves, V. S. P.; Ferreira Neto, J. S. Epidemiological situation of bovine tuberculosis in the state of Mato Grosso, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3589-3600, 2016. Suplemento 2.
- Noordhuizen, J. P. T. M.; Thrusfield, M. V.; Frankena, K.; Graat, E. A. M. Application of quantitative methods in veterinary epidemiology. Wageningen: Wageningen Academic Publishers, 2001. 429 p.
- Oloya, J.; Muma, J. B.; Opuda-Asibo, J.; Djønne, B.; Kazwala, R.; Skjerve, E. Risk factors for herd-level bovine-tuberculosis seropositivity in transhumant cattle in Uganda. *Preventive Veterinary Medicine, Amsterdam*, v. 80, n. 4, p. 318-329, 2007.
- Paoli, T. M. P. de. Avaliação de risco relativo para tuberculose bovina em municípios do Estado do Espírito Santo a partir de dados de matadouros. 2013. Dissertação (Mestrado em Ciências Veterinárias) – Centro de Ciências Agrárias, Universidade Federal do Espírito Santo, Alegre.
- Porphyre, T.; Stevenson, M. A.; Mckenzie, J. Risk factors for bovine tuberculosis in New Zealand cattle farms and their relationship with possum control strategies. *Preventive Veterinary Medicine, Amsterdam*, v. 86, n. 1-2, p. 93-106, 2008.
- Queiroz, M. R.; Groff, A. C. M.; Silva, N. S.; Grisi-Filho, J. H. H.; Amaku, M.; Dias, R. A.; Telles, E. O.; Heinemann, M. B.; Ferreira Neto, J. S.; Gonçalves, V. S. P.; Ferreira, F. Epidemiological status of bovine tuberculosis in the state of Rio Grande do Sul, Brazil. *Semina: Ciências Agrárias, Londrina*, v. 37, n. 5, p. 3647-3658, 2016. Suplemento 2.
- R Core Team – R: a language and environment for statistical computing. Austria Ed. Vienna: R Foundation for Statistical Computing, 2012.
- Ramírez-Villaescusa, A. M.; Medley, G. F.; Mason, S.; Green, L. E. Risk factors for herd breakdown with bovine tuberculosis in 148 cattle herds in the south west of

- England. *Preventive Veterinary Medicine*, Amsterdam, v. 95, n. 3-4, p. 224-230, 2010.
- Reilly, L. A.; Courtenay, O. Husbandry practices, badger sett density and habitat composition as risk factors for transient and persistent bovine tuberculosis on UK cattle farms. *Preventive Veterinary Medicine*, Amsterdam, v. 80, n. 2-3, p. 129-142, 2007.
- Ribeiro, L. A.; Gonçalves, V. S. P.; Francisco, P. F. C.; Mota, A. L. A. A.; Nascimento, G. T.; Licurgo, J. B.; Ferreira, F.; Grisi-Filho, J. H. H.; Ferreira Neto, J. S.; Amaku, M.; Dias, R. A.; Telles, E. O.; Heinemann, M. B.; Borges, J. R. J. Epidemiological status of bovine tuberculosis in the Federal District of Brazil. *Semina: Ciências Agrárias*, Londrina, v. 37, n. 5, p. 3561-3566, 2016. Suplemento 2.
- Rocha, W. V.; Jayme, V. S.; Mota, A. L. A. A.; Brito, W. M. E. D.; Pires, G. R. C.; Grisi-Filho, J. H. H.; Dias, R. A.; Amaku, M.; Telles, E. O.; Heinemann, M. B.; Ferreira, f.; Ferreira Neto, J. S.; Gonçalves, V. S. P. Prevalence and herd-level risk factors of bovine tuberculosis in the State of Goiás, Brazil. *Semina: Ciências Agrárias*, Londrina, v. 37, n. 5, p. 3625-3628, 2016. Suplemento 2.
- Silva, M. C. P.; Gonçalves, V. S. P.; Mota, A. L. A. A.; Koloda, M.; Ferreira Neto, J. S.; Grisi-Filho, J. H. H.; Dias, R. A.; Amaku, M.; Telles, E. O.; Ferreira, F.; Heinemann, M. B.; Alfieri, A. A.; Muller, E. E. Prevalence and herd-level risk factors for bovine tuberculosis in the state of Paraná, Brazil. *Semina: Ciências Agrárias*, Londrina, v. 37, n. 5, p. 3611-3624, 2016. Suplemento 2.
- Skuce, R. A.; Allen, A. R.; Mcdowell, S. W. J. Herd-level risk factors for bovine tuberculosis: a literature review. *Veterinary Medicine International*, New York, v. 2012, Article ID 621210, p. 1-10, 2012. Available at: . Accessed at: 07 oct. 2016.
- THRUSFIELD, M. *Veterinary epidemiology*. 3th ed. Hoboken: Wiley-Blackwell, 2007. 624 p.
- Tschopp, R.; Schelling, E.; Hattendorf, J.; Aseffa, A.; Zinsstag, J. Risk factors of bovine tuberculosis in cattle in rural livestock production systems of Ethiopia. *Preventive Veterinary Medicine*, Amsterdam, v. 89, n. 3-4, p. 205-211, 2009.
- Veloso, F. P.; Baumgarten, K. D.; Mota, A. L. A. A.; Ferreira, F.; Ferreira Neto, J. S.; Grisi-Filho, J. H. H.; Dias, R. A.; Amaku, M.; Telles, E. O.; Heinemann, M. B.; Gonçalves, V. S. P. Prevalence and herd-level risk factors of bovine tuberculosis in the State of Santa Catarina, Brazil. *Semina: Ciências Agrárias*, Londrina, v. 37, n. 5, p. 3659-3672, 2016. Suplemento 2.
- Vendrame, F. B.; Amaku, M.; Ferreira, F.; Telles, E. O.; Grisi-Filho, J. H. H.; Gonçalves, V. S. P.; Heinemann, M. B.; Ferreira Neto, J. S.; Dias, R. A. Epidemiologic characterization of bovine tuberculosis in the State of Rondônia, Brazil. *Semina: Ciências Agrárias*, Londrina, v. 37, n. 5, p. 3639-3646, 2016. Suplemento 2.
- Vose, D. *Risk analysis: a quantitative guide*. Hoboken: John Wiley & Sons, 2008. 752 p.

Zinsstag, J.; Schelling, E. L.; Roth, F.; Kazwala, R. Economics of bovine tuberculosis.
In: Thoen, C. O.; Steele, J. H.; Gilsdorf, M. J. *Mycobacterium bovis* infection in
animals and humans. Hoboken: Wiley-Blackwell, 2006. p. 68-83.

Chapter 2: Risk Analysis for Bovine Tuberculosis in Espírito Santo, Brazil: an approach based on spatial and network analysis.

Abstract

The Brazilian program for the control and eradication of Bovine Tuberculosis (bTB) in cattle populations was launched in 2001 with the aim to reduce the negative impact of this disease in the country. One of the main long-term challenges that the Brazilian program faces is to detect farms with bTB.

This study is focused in Espírito Santo (ES), a Brazilian state with one of the highest rates of bTB prevalence in the country. Our objective is evaluate the spatial distribution of the farms that tested positive to bTB and the level of risk associated with metrics of the bovine trade network. We analysed the bovine trade network from ES between 2014 and 2015 using network metrics (Degree, Weighted Degree, Betweenness, PageRank, Closeness, Contact Chain, Kleinberg's authority and Network components). We categorized the farms' network metrics into high and low values, and associated it with the farms' bTB status (positive and negative) and calculate the Odds Ratio from those associations in both univariate and multivariate analyses.

We evaluated the variation in spatial risk of bTB in ES by calculating the relative risk with the localization of cases and controls in the region. The results from the univariate analysis showed that the variables Weighted All-Degree, Weighted In-Degree, Weighted Out-Degree, Betweenness, PageRank and Authority had the strongest association ($p \leq 0.05$) with the farms' bTB status. In addition, the multivariate analyses showed that Weighted Out-Degree (OR = 2.3, IC = 0.98-2.9) and Betweenness (OR = 2.2, IC = 0.93-5.7) were the two variable that could explain better the bTB farms' status. The south of ES was the area with higher risk for bTB and those areas were located close to the border with Minas Gerais, another Brazilian state with high bTB prevalence. In conclusion, the risk factors associated with bTB in this study can help to complement the surveillance program to control this disease in ES.

Keywords: network analysis, spatial analysis, zoonotic, dairy farms, livestock.

Introduction

Bovine Tuberculosis (bTB) is an infectious zoonotic disease with chronic evolution caused by *Mycobacterium bovis*. It is distributed worldwide and localized mainly in developing countries (Cosivi, 1998; Müller et al., 2013). In animal production, this disease has high economic importance due to the associated animal deaths (natural or by sanitary slaughter), reduction of weight and milk production, and loss of credibility with its trade partners (LAGE et al., 2006).

The Brazilian program for the control and eradication of bTB in cattle populations was launched in 2001 with the aim to reduce the negative impact of zoonosis in human and animal health, and also to increase the competitiveness of the animal production of the nation (LAGE et al., 2006). The program has been efficient in improving data quality and characterizing the state of this disease across the country and create an anti-tuberculosis culture (Ferreira Neto et al., 2016).

One of the main long-term challenges that the Brazilian program to control bTB faces is to detect bTB positive animals in farms; although the instituted programs have been training veterinarians from both private and public sectors to perform diagnostic tests for bTB in livestock productions, testing is not routinely performed in the country (Ferreira Neto et al., 2016; LAGE et al., 2006). As a consequence, the lack of a continuous surveillance allows the disease to spread to other animals and other herds, increasing the cost of control and eradication (VanderWaal, Enns, Picasso, Packer, & Craft, 2016). On the other hand, some farms characteristics can be associated as risk factor with bTB (Humblett, Boschiroli, & Saegerman, 2009). Therefore, it is possible to identify set of farms with higher risk for this disease than others. Hence, focus the surveillance program in this group can increase the odds to find more bTB positive farms. Thus, in order to detect bTB cases earlier and improve the sensitivity of the surveillance system, it is necessary to focus on the subset of farms with higher risk to be infected with bTB.

This study is focused in Espírito Santo (ES), a Brazilian state with one of the highest levels of bTB prevalence in the country (Galvis et al., 2016). Previously, we studied the bTB risk factors in this state considering variables such as production system (meat, milk, or mixed), raising system (extensive or any degree of confinement), cattle breeds, number of cows, total herd size, presence of other domesticated species, presence of wild animals, animal trade, routine of tuberculosis testing, slaughter in the farm, pasture sharing, feeding calves with whey, indirect contact between farms, flooded pastures, and veterinary assistance (Galvis et al.,

2016). The results indicated that the high number (>9) of female bovines > 24 months and the milking type (milking machine/milking parlor) were associated with bTB infection (Galvis et al., 2016). This study widely explored variables associated with bTB which had been mentioned in the literature as risk factors (Humblet, Boschioli, & Saegerman, 2009). However, some other variables, such as the characteristics of the whole bovine transit and spatial agglomerations, were not studied and can provide additional information to understand the behavior of this disease in ES.

Animal transit is an important variable to study bTB because the movement of cattle has been identified as a risk factor, especially in areas with low prevalence, to introduce infected animals in bTB-free zones (M. Carolyn Gates, Volkova, & Woolhouse, 2013; Gilbert et al., 2005). The methodology used to analyse this transit is normally called Network Analysis, derived from social sciences; this can help to understand the structure of contacts between farms and the role that they play in global trade (Dubé, Ribble, Kelton, & McNab, 2009; Martínez-López, Perez, & Sánchez-Vizcaíno, 2009). Its applications have risen in recent years, especially in public health, because it can be useful to simulate disease spread and control measures in animal populations (Kao, Danon, Green, & Kiss, 2006; Natale et al., 2009; Rossi, Smith, Pongolini, & Bolzoni, 2017).

In network analysis, there are usually two studied entities, nodes and edges. In this study the nodes are the farms and the edges are the bovine movements between those farms. There are many metrics that can be evaluated from the farms in the network, such as In-degree, Out-degree, Betweenness, Closeness, Clustering Coefficient, Page Rank, Contact Chain, and many more (Martínez-López et al., 2009). Each one of those gives us individual information about the trade characteristics and its centrality, i.e. how well connected a farm is. One of our hypotheses is that highly connected farms are more likely to be infected with bTB than those who are poorly connected, thus some network metrics should be associated with the presence of the disease. This hypothesis has been evaluated in other studies with bTB and other diseases around the world (Cárdenas et al., 2018; Frössling, Ohlson, Björkman, Håkansson, & Nöremark, 2012; M. Carolyn Gates et al., 2013; Palisson, Courcoul, & Durand, 2016).

Another important way to improve the control and eradication of diseases is to know the spatial risk variation of the cases in a geographical area, avoiding the confusion caused by the underlying population dispersion (Bithell, 1991). Thus, in prevalence studies, it is necessary to not only study the location of the cases, but also

study the location of the controls. Hence, it is possible to find an estimated ratio from the case and control densities and, as consequence, evaluate the spatial variation of the disease risk and the existence of statistically significant sub-regions (Davies & Hazelton, 2010; Hazelton & Davies, 2009). In ES, the south of the state has higher bTB prevalence (Galvis et al., 2016), but there is not information about the existence of delimited areas with higher risk. One of the goals of this study is identify the existence of any bTB risk areas in ES.

This study's aim is to support and complement the study published by Galvis, JOA et al. in 2016, about risk factors associated with bTB in ES. We have used the same bTB outbreak data, which were the bTB positive farms detected by the Animal Health Service of the state of Espírito Santo (Instituto de Defesa Agropecuária e Florestal do Espírito Santo-IDAF) in ES, between 2012 and 2014. That study has the final goal to develop a risk-based surveillance for bTB, and minimize the cost of disease control and eradication. The objective of this study is to evaluate the spatial distribution of the farms that tested positive to bTB and the level of risk associated with metrics of the bovine trade network.

Materials and methods

Study area

The study was developed in Espírito Santo (ES), a Brazilian state localized in the southeast region of the country. This state has 1.34 million hectares of pasture available for an estimated 1.6 million bovines and over 3,000 bubaline distributed in approximate 30,000 farms, which can be stratified in 18 thousand milk type farms and 14 thousand beef type farms (IBGE, 2017). The milking and the beef production represented 1.4% and 1% of the Brazilian production, respectively, and together they generated close to 1.178 Billion Reais (around 286 Million Dollars) in 2014 (GOVERNO DO ESTADO DO ESPÍRITO SANTO, 2016).

Bovine Tuberculosis database

This database has the information about 637 farms sampled to detect Bovine Tuberculosis (bTB) in ES. The data was collected from January 2012 to May 2014 by the Animal Health Service of the state of Espírito Santo (Instituto de Defesa Agropecuária e Florestal do Espírito Santo-IDAF). The database contains each of these farms geographic location (spatial coordinates, latitude and longitude) and the results of the comparative cervical tuberculin test, classifying the farms as either

positive or negative according to the guidelines of the national program (LAGE et al., 2006). At the end of that study 48 farms were found to be positive and 589 were negative (SM Figure 1). Additional information about the methodology design and results of this study can be found in the study published by Galvis et al. (2016).

Animal movement database

In Brazil, it is necessary that the farmers report animal movements to its state Inspection Service (Decree No. 5,741, March 30, 2006, Brazil). In ES, the government entity in charge to receive and store these information is the IDAF, which provided us with the complete database of bovine and bubaline movements from January 2014 to December 2015 made within ES and between ES and other Brazilian states. This dataset included: movement ID; date; type of transport; specie; reason for movement: Slaughter, event and animal reposition; source and destination; and the number of animals moved stratified by sex and age (a short description of the data are available in SM Table 1). In this study, we specifically selected the movements involving farms and livestock events (we excluded movements to abattoirs).

Network construction

There were two administrative levels in the source and destination of the trade movements. These were the farm level (the trade link between two farms) and the municipality level (the trade link between two municipalities). This study worked with the farm level data to get more detailed information about the network metrics associated with bTB in ES. The network was considered as static, meaning that the movements were aggregated over the two study years and multiple movements between two farms were considered as one unique edge. However, the network conserved its temporal structure (daily movement records) in the evaluation of the Outgoing and Ingoing contact chain (Table 1).

Risk analysis based on the centrality metrics from the network

We evaluated a set of network centrality metrics (Table 1) used widely in veterinary epidemiology to identify the role of the farms to connect other farms in the network either directly or indirectly (Dubé et al., 2009; Martínez-López et al., 2009); farms with higher centrality values are more strongly connected with other farms than those with lower values. We calculated the network parameters for the farms for which at least one animal movement was recorded in ES between 2014 and 2015, and then selected those with available status for bTB. From the 637 farms sampled for bTB, only 409 farms had bovine movement records. Thus, this was our initial number of farms to

evaluate the risk associated with the network parameters. The analysis was stratified by the type of farm productions (dairy, beef and mixed). However, beef and mixed productions had a lower number of farms positive to bTB (1 and 6 positive farms respectively). Thus, we decided to apply this methodology only to dairy premises, which had 297 active farms (farms with animal movement records in the years studied) in the bovine trade network, 28 of them being positive and 269 negative.

The network metrics calculated for the dairy farms were categorized in two levels, high and low, accordingly to the optimal cutoff to separate its distribution by farms' bTB status (Thiele, 2019) (SM Figure 2 and SM Table 2). Additionally, we evaluated the group of farms which belong to the Giant Strong Connected Component (GSCC) (Table 1), and categorized them into two groups: farms which belong to the GSCC (in) and those which did not belong (out). The association of these metrics with the farms' bTB status were evaluated with a univariate analysis; this consisted of a chi-square test and calculation of the Odds Ratio (OR), a measure of association between an exposure and an outcome. Next, those variables with strong association (chi-square p value < 0.2) were included in a multivariate logistic regression. The final variables in the model were chosen accordingly to the lower Akaike Information Criterion (AIC) from the regression model.

Risk analysis based on spatial dispersion

The variation in spatial risk of bTB in ES was analysed with a methodology proposed by Hazelton and Davies (2009). This method provides functions to perform fixed and adaptive bivariate kernel density estimation and calculate the relative risk with the spatial location of cases and controls in a study region. Also, it is able to identify statistically significant fluctuations in the risk variation by calculating p values over an adaptive log relative risk function (Davies & Hazelton, 2010; Hazelton & Davies, 2009). This analysis was made with the support of the R package's "sparr" (Davies, Hazelton, & Marshall, 2011).

Table 1. Description of the network metrics used

Network metrics applied to individual farms	Description
In-Degree	Number of farms which have sent animals to a specific farm (Wasserman & Faust, 1994).
Out-Degree	Number of farms which have obtained animals from a specific farm (Wasserman & Faust, 1994).
All-Degree	It is the sum of the in-degree and the out-degree (Wasserman & Faust, 1994).
Weighted In-Degree	Total number of animals sent to a specific farm
Weighted Out-Degree	Total number of animals sent from a specific farm
Weighted All-Degree	The sum of the weighted in-degree and the weighted out-degree
Betweenness	The number of times a farm is in the shortest path between pairs farms in the network (Wasserman & Faust, 1994).
Closeness	A metric of how closely connected a farm is to all other farms of the network, calculated as the sum of the length of the shortest paths between the farm and all other farms (Linton C. Freeman, 1978).
PageRank	The importance of a farm measured by the importance of the farms it is connected to in the network (Brin & Page, 1998).
Outgoing contact chain (OCC)	The number of farms reached through direct movements, and indirect contacts through further movements, where the sequence of the movements is taken into account (Nöremark, Frössling, & Lewerin, 2013).
Ingoing contact chain (ICC)	The number of farms can reach an specific farm through direct movements, and indirect contacts through further movements, where the sequence of the movements is taken into account (Nöremark et al., 2013).
Cluster Coefficient	Measure how the farms in the network tend to cluster together (Watts & Strogatz, 1998).
Kleinberg's authority	Calculate an authority score by the connections of farms to hubs in the network (Kleinberg, 1999)
Giant strong connected Component (GSCC)	The largest subsets of farms with directed paths between all pairs (Dorogovtsev, Mendes, & Samukhin, 2001).

Results

Risk analysis based on the centrality metrics from the network

The number of active farms in Espírito Santo (ES) during the studied period was 28,334, and those were accounted for 124,531 bovine movements between 2014 and 2015. As mentioned earlier, we focused this section in 297 dairy farms sampled to bTB in ES and that were also active in the bovine trade network. Those farms were responsible for 2,030 bovine movements and were distributed across the state (SM Figure 3). Most of the dairy farms sampled had low in-degree and out-degree values (median degree <2) and <25% of them had high values (>50) (Table 2). Similar patterns were found in the weighted in-degree and weighted out-degree metrics. On the other hand, the ICC had a maximum value 2 times higher than the OCC, but the median and quantiles from the OCC were larger. Thus, most of the farms can reach many other farms in the trade network (75% can reach ≥ 1667 farms), and be reached by few of them (75% can be reached by ≥ 11 farms) (Table 2). Since the bovine trade network in ES was not well connected, and the Closeness centrality is not effective in this kind of scenarios (Wasserman & Faust, 1994), we did not calculate it.

Table 2. Summary of network centrality metrics from the dairy farms sampled to Bovine Tuberculosis in ES between 2012 and 2014. (Q25 = quantile 25% and Q75 = quantile 75%)

Network metrics	Q25	Median	Q75	Max
In-Degree	0	1	2	100
Out-Degree	1	2	4	56
All-Degree	2	3	6	156
Weighted In-Degree	0	3	20	1159
Weighted Out-Degree	8	21	45	882
Weighted All-Degree	10	30	63	2041
ICC	0	1	11	10994
OCC	8	218	1667	4808
Betweenness	0	0	3e-05	0.0086
PageRank	8e-06	1e-05	2.3e-05	0.00066
Cluster Coefficient	0	0	0.1	1
Kleinberg's authority	0	4e-07	0.0001	0.007

In general, the dairy farms that were positive to bTB had network metrics values higher than those that were negative (SM Figure 2). Thus, it looks like farms more

strongly connected in the network are more likely to have bTB. However, this difference was clearer in some parameters than in others. The results indicated that the Weighted All-Degree, Weighted In-Degree, Weighted Out-Degree, Betweenness, PageRank and Authority had the strongest association ($p \leq 0.05$) with the farms' bTB status (Table 3). Also, the category "high" (farms with higher centrality network metrics) was a risk factor in all the associations evaluated, and the Weighted Out-Degree was the variable with highest odds ratio value (3.16).

Table 3. Univariate and Multivariate analysis between network metrics and bTB positive and negative dairy farms in ES.

Variable	Cat	-	+	Univariate		Multivariate	
				OR (IC)	<i>p</i>	OR(IC)	<i>p</i>
All-Degree	Low	150	11	1		-	-
	High	119	17	1.93 (0.9-4.4)	0.1	-	-
In-Degree	Low	124	8	1		-	-
	High	145	20	2.11 (0.9-5.3)	0.08	-	-
Out-Degree	Low	100	7	1		-	-
	High	169	21	1.75 (0.7-4.6)	0.2	-	-
Weighted All-Degree	Low	147	8	1		-	-
	High	122	20	2.97 (1.3-7.5)	0.01	-	-
Weighted In-Degree	Low	123	7	1		-	-
	High	146	21	2.48 (1.1-6.6)	0.04	-	-
Weighted Out-Degree	Low	162	9	1		1	-
	High	107	19	3.16 (1.4-7.6)	0.01	2.3(0.98-2.9)	0.06
Betweenness	Low	160	9	1		1	-
	High	109	19	3.06 (1.4-7.4)	0.01	2.2(0.93-5.7)	0.07
PageRank	Low	128	8	1		-	-
	High	141	20	2.24 (1-5.6)	0.05	-	-
Clust. Coefficient	Low	116	13	1		-	-
	High	84	10	1.07 (0.4-2.6)	0.89	-	-
Authority	Low	144	9	1		-	-
	High	125	19	2.4 (1.1-5.8)	0.03	-	-
ICC	Low	244	23	1		-	-
	High	25	5	2.16 (0.7-5.9)	0.15	-	-
OCC	Low	185	16	1		-	-
	High	84	12	2.65 (0.7-3.7)	0.21	-	-
GSCC	Out	170	14	1		-	-
	In	99	14	1.7 (0.7-3.8)	0.17	-	-

The interaction between the variables in the multivariate logistic regression analysis made it impossible to include more than 1 variable in the model considering the *p* values less than 0.05. However, we found that the best fit with two variables was the Weighted Out-Degree and the Betweenness centrality. Therefore, farms positive for bTB had 2.2 times greater odds ratio for farms that sold a high number of animals (Weighted Out-Degree ≥ 28) than for farms that sold a low number of animals (Weighted Out-Degree < 28) and 2.3 greater odds for farm which had high

Betweenness meaning that they were an important bridge to connect other farms in the network (Betweenness value $\geq 2.8 \times 10^{-06}$) than those which had less importance (Betweenness value $< 2.8 \times 10^{-06}$).

Additionally, we associated those variables with a risk factor previously identified in ES, the number of bovine females in the farms > 2 years old. This variable was categorized in two groups, farms with < 10 such animals and farms with ≥ 10 , we referred this variable as farm size. As a result, the number of bovine females was stronger associated with the Weighted Out-Degree than the Betweenness (Table 4). Thus, farms with ≥ 10 animals had 2.8 times greater odds ratio for farms with high Weighted Out-Degree (≥ 28) than farms with a low Weighted Out-Degree (< 28). Although the association between those two variables was not strong ($p \geq 0.05$), it makes sense that farms with a large number of animals would tend to move more animals than those which do not.

Table 4. Univariate analysis between network metrics and number of female bovine > two years in Espírito Santo.

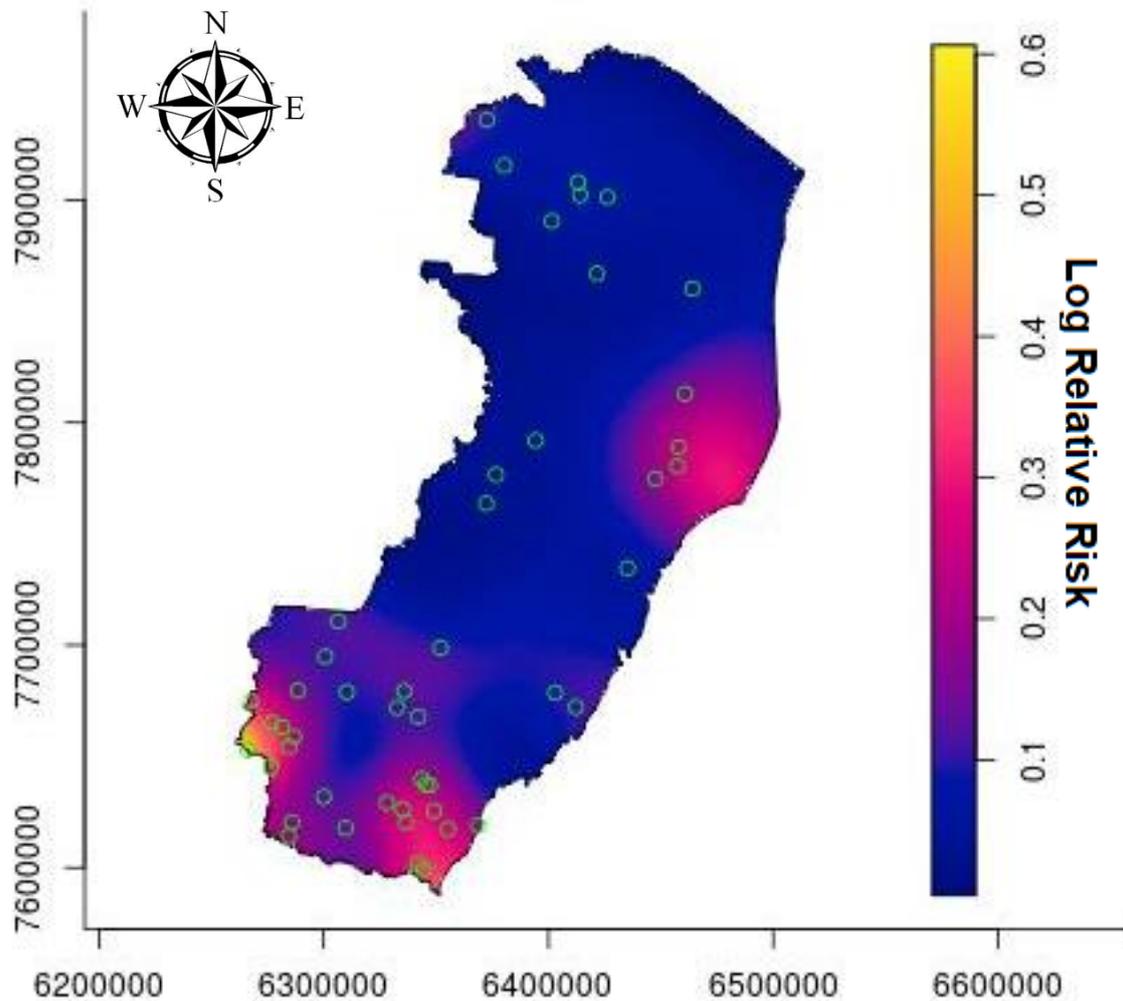
Variable	Cat	Farm size		Univariate	
		<10	≥ 10	OR (IC)	<i>p</i>
Weighted Out-Degree	Low	166	5	1	0.05
	High	116	10	2.8 (0.96-9.4)	
Betweenness	Low	163	6	1	0.17
	High	119	9	2 (0.7-6.3)	

Risk analysis based on spatial dispersion

For our analysis we used the 637 farms sampled for bTB in ES and used their disease status as cases and controls. The results are visualized on a map, with the superposition of the adaptive log relative risk function and the addition of one or more tolerance contour if there is sub-regions with significant risk ($p < 0.05$). The existence of a tolerant contour determines whether or not a given peak in an estimated surface reflects truly heightened risk or if the distribution of cases is simply a product of random variation.

The south of ES was the area with higher relative risk intensity for bTB (Figure 1) and those areas located close to the border of the state. The region with higher risk are located in the south-west, followed by another in the south-east and other close to the middle-east. However, in spite of identifying those areas with higher relative risk, we did not find a sub-region with significant risk (i.e with $p < 0.05$).

Figure 1. Spatial distribution of Bovine Tuberculosis risk of the farms sampled in Espírito Santo between 2012 and 2014. Positive farms for Bovine Tuberculosis represented with green circles on the map.



Discussion

In this study we identified that dairy farms stronger connected in the bovine trade network were more likely to have Bovine Tuberculosis (bTB). Also, it was possible to observe that the south of the state, an area with higher number of dairy farms in ES (LAVAGNOLI et al., 2010), had higher relative risk for bTB than other areas in the state.

In general, all the network metrics categorized as high were a risk factor for bTB in this study, a relationship identified also in other studies in Africa and Europe for this disease (Mekonnen et al., 2019; Palisson et al., 2016; Pozo et al., 2019; Sintayehu, Prins, Heitkönig, & de Boer, 2017). Thus, farms strongly connected in the network have more risk to be infected with bTB. However, not all network metrics studied had strong associations with the disease, the most important variables studied were the Weighted

Out-Degree (number of animals sold by the farms) and Betweenness (Farms which act as a bridge to connect other farms in the network). With respect to the Weighted Degree, the high Weighted In-Degree and Weighted All- Degree were also risk factor associated to bTB. Therefore, a possible explanation may be that a high number of animals traded can make a farm more vulnerable to receiving infected animals, especially if those are purchased from areas with high bTB prevalence (M.C. Gates, Volkova, & Woolhouse, 2013), as is the case of ES, which has one of the highest bTB prevalences in Brazil (Ferreira Neto et al., 2016; Galvis et al., 2016).

In addition, the Weighted Out-Degree is closely associated with farm size. Thus, this association can mean that the number of animals traded is a variable to identify those farms with a large number of animals, which in turn have higher risk to be infected with bTB (Galvis et al., 2016). Additionally, those farms with high Betweenness can facilitate the disease spread between different groups of farms in the network (Marquetoux, Stevenson, Wilson, Ridler, & Heuer, 2016; Natale et al., 2009). As a consequence, some studies have indicated this metric as a way to target farms and decrease the probability of large epidemics (Marquetoux et al., 2016; Natale et al., 2009). Implementation of this metric in the surveillance programs can aid in the identification and control of bTB.

In the risk factor analysis between network metrics and bTB in ES we focused on the dairy productions, and excluded beef and mixed productions due to the lack of bTB positive farms active in the trade network. This meant that we lost information to evaluate better the risk factors associated with the disease in the state. Nevertheless, it is possible to extrapolate the risk factors found here in dairy farms to beef and mixed farms. However, those productions can have different trade patterns, for example, in France, the beef farms had a higher number and reciprocity of movements (Dutta, Ezanno, & Vergu, 2014), or, in United Kingdom, beef has stronger seasonal trading behavior than dairy farms (M. C. Gates, 2014). Thus, those characteristics can influence the vulnerability of the beef farms to purchase bTB infected animals and potentially show different results to those found in the dairy farms. Therefore, using our outcomes exclusively for dairy farms may bring better results in the surveillance of bTB in ES.

In the analysis of the network metrics, there is not a specific cutoff reference to split the values in high or low. Thus, we decided to split those metrics by the bTB farm status. The network metrics values were divided in the best way to be associated with the bTB farm status, using maximization of sensitivity and specificity. As a

consequence, one strong limitation in the study is the overfitting in the univariate and multivariate association analysis.

The fact we did not find any delimited sub-regions within ES with significant relative risk did not mean that the cases were distributed uniform along the state. The south of the state clearly showed higher relative risk (Figure 1), and that must be considered in the control of bTB in ES. Also, that area was located next to the Minas Gerais (MG) border, another Brazilian state with higher bTB prevalence, particularly in the area close to the ES border (Ferreira Neto et al., 2016). Thus, the bovine trade between those two regions may be significant in the transmission of the disease. We recommend the identification of areas in MG, close to ES, with higher relative risk and evaluate the commercial relationships between those possible areas and those already found with high risk in ES. Additionally, areas found with higher relative risk can act as a hub to spread the pathogen into the state and into neighboring states. Thus, a stronger surveillance in this particular area may help to decrease the prevalence of bTB in ES in the future.

In this study, we had an important limitation with the association between the network metrics and the farm bTB status due to the temporal gap between network and identification of bTB cases. We used two years of the bovine trade network, 2014 and 2015, but the cases of bTB in ES were collected between 2012 and 2014. Thus, the network metrics calculated were posterior to the identification of the bTB cases. We therefore cannot say that the network metrics evaluated were a probable cause of bTB in those positive farms. However If we suppose that the network doesn't change from year to year, then our result can be similar to the one using a set of bovine movements from previous years to the farms sampled to bTB. Nevertheless, the risk factors presented in this study must be taken with caution if they are to be considered as part of the surveillance program in ES.

In conclusion, the risk factors associated with bTB in this study can help to complement the surveillance program to control this disease in ES. The farms with a higher number of animals commercialized, and strongly connected in the network, can be more vulnerable to infection from bTB. Also, the south of ES, border with MG, it is an area with high relative risk and probably requires a stronger surveillance in order to avoid the continual spread of the disease into the state and to the others close to it.

References

- Bithell, J. F. (1991). Estimation of relative risk functions. *Statistics in Medicine*, 10(11), 1745–1751. <https://doi.org/10.1002/sim.4780101112>
- Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine The Anatomy of a Search Engine. *Computer Networks and ISDN Systems*, 30(June 2000), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Cárdenas, N. C., Galvis, J. O. A., Farinati, A. A., Grisi-Filho, J. H. H., Diehl, G. N., & Machado, G. (2018). Burkholderia mallei: The dynamics of networks and disease transmission. *Transboundary and Emerging Diseases*, (July 2018), 715–728. <https://doi.org/10.1111/tbed.13071>
- Cosivi, O. (1998). Zoonotic Tuberculosis due to Mycobacterium bovis in Developing Countries. *Emerging Infectious Diseases*, 4(1), 59–70. <https://doi.org/10.3201/eid0401.980108>
- Davies, T. M., & Hazelton, M. L. (2010). Adaptive kernel estimation of spatial relative risk. *Statistics in Medicine*, n/a-n/a. <https://doi.org/10.1002/sim.3995>
- Davies, T. M., Hazelton, M. L., & Marshall, J. C. (2011). sparr : Analyzing Spatial Relative Risk Using Fixed and Adaptive Kernel Density Estimation in R. *Journal of Statistical Software*, 39(1). <https://doi.org/10.18637/jss.v039.i01>
- Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2001). Giant strongly connected component of directed networks. *Physical Review E*, 64(2), 025101. <https://doi.org/10.1103/PhysRevE.64.025101>
- Dubé, C., Ribble, C., Kelton, D., & McNab, B. (2009). A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transboundary and Emerging Diseases*, 56(3), 73–85. <https://doi.org/10.1111/j.1865-1682.2008.01064.x>
- Dutta, B. L., Ezanno, P., & Vergu, E. (2014). Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Preventive Veterinary Medicine*, 117(1), 79–94. <https://doi.org/10.1016/j.prevetmed.2014.09.005>
- Ferreira Neto, J. S., Silveira, G. B. da, Rosa, B. M., Gonçalves, V. S. P., Grisi-Filho, J. H. H., Amaku, M., ... Lage, A. P. (2016). Analysis of 15 years of the National Program for the Control and Eradication of Animal Brucellosis and Tuberculosis, Brazil. *Semina: Ciências Agrárias*, 37(5Supl2), 3385. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3385>
- Frössling, J., Ohlson, A., Björkman, C., Håkansson, N., & Nöremark, M. (2012). Application of network analysis parameters in risk-based surveillance - Examples based on cattle trade data and bovine infections in Sweden. *Preventive Veterinary Medicine*, 105(3), 202–208. <https://doi.org/10.1016/j.prevetmed.2011.12.011>
- Galvis, J. O. A., Grisi-Filho, J. H. H., Costa, D. Da, Said, A. L. P. R., Amaku, M., Dias, R. A., ... Ferreira Neto, J. S. (2016). Epidemiologic characterization of bovine tuberculosis in the State of Espírito Santo, Brazil. *Semina: Ciências Agrárias*, 37(5Supl2), 3567. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3567>
- Gates, M. C. (2014). Characteristics of replacement breeding cattle trade in Great

- Britain. *Veterinary Record*, 175(3), 67–67. <https://doi.org/10.1136/vr.102351>
- Gates, M. Carolyn, Volkova, V. V., & Woolhouse, M. E. J. (2013). Risk factors for bovine tuberculosis in low incidence regions related to the movements of cattle. *BMC Veterinary Research*, 9. <https://doi.org/10.1186/1746-6148-9-225>
- Gates, M.C., Volkova, V. V., & Woolhouse, M. E. J. (2013). Impact of changes in cattle movement regulations on the risks of bovine tuberculosis for Scottish farms. *Preventive Veterinary Medicine*, 108(2–3), 125–136. <https://doi.org/10.1016/j.prevetmed.2012.07.016>
- Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R., & Wint, W. (2005). Cattle movements and bovine tuberculosis in Great Britain. *Nature*, 435(7041), 491–496. <https://doi.org/10.1038/nature03548>
- GOVERNO DO ESTADO DO ESPÍRITO SANTO. (2016). *Pedeag 3*. Retrieved from <https://seag.es.gov.br/pedeag3>
- Hazelton, M. L., & Davies, T. M. (2009). Inference Based on Kernel Estimates of the Relative Risk Function in Geographical Epidemiology. *Biometrical Journal*, 51(1), 98–109. <https://doi.org/10.1002/bimj.200810495>
- Humblet, M. F., Boschioli, M. L., & Saegerman, C. (2009). Classification of worldwide bovine tuberculosis risk factors in cattle: A stratified approach. *Veterinary Research*, 40(5). <https://doi.org/10.1051/vetres/2009033>
- IBGE. (2017). Censo Agropecuário. Retrieved from <https://cidades.ibge.gov.br/brasil/es/pesquisa/24/76693>
- Kao, R. . R., Danon, L., Green, D. . M., & Kiss, I. . Z. (2006). Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597), 1999–2007. <https://doi.org/10.1098/rspb.2006.3505>
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632. <https://doi.org/10.1145/324133.324140>
- Lage, A. P. ., Roxo, E. ., Müller, E. ., Poester, F. ., Cavalléro, J. C. M. ., Ferreira Neto, J. S. ., ... Gonçalves, V. S. P. (2006). *Programa nacional de controle e erradicação da brucelose e da tuberculose animal (PNCEBT)*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Lavagnoli, M. R.; Amorim, B. M. de; Machado, G. P.; Demoner, L. de C.; Zanini, M. S.; Antunes, J. M. A. de P. Tuberculosis in bovines from Espírito Santo State. *Veterinária e Zootecnia, Botucatu*, v. 17, n. 1, p. 71-78, 2010.
- Linton C. Freeman. (1978). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 179(1968), 215–239.
- Marquetoux, N., Stevenson, M. A., Wilson, P., Ridler, A., & Heuer, C. (2016). Using social network analysis to inform disease control interventions. *Preventive Veterinary Medicine*, 126, 94–104. <https://doi.org/10.1016/j.prevetmed.2016.01.022>
- Martínez-López, B., Perez, A. M., & Sánchez-Vizcaíno, J. M. (2009). Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transboundary and Emerging Diseases*, 56(4), 109–120. <https://doi.org/10.1111/j.1865-1682.2009.01073.x>

- Mekonnen, G. A., Ameni, G., Wood, J. L. N., The ETHICOBOTS Consortium, Berg, S., & Conlan, A. J. K. (2019). Network analysis of dairy cattle movement and its implication on Bovine Tuberculosis spread and control in emerging dairy belts of Ethiopia. *BMC Veterinary Research*, 1, 1–14.
- Müller, B., Dürr, S., Alonso, S., Hattendorf, J., Laisse, C. J. M., Parsons, S. D. C., ... Zinsstag, J. (2013). Zoonotic Mycobacterium bovis –induced Tuberculosis in Humans. *Emerging Infectious Diseases*, 19(6), 899–908. <https://doi.org/10.3201/eid1906.120543>
- Natale, F., Giovannini, A., Savini, L., Palma, D., Possenti, L., Fiore, G., & Calistri, P. (2009). Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine*, 92(4), 341–350. <https://doi.org/10.1016/j.prevetmed.2009.08.026>
- Nöremark, M., Frössling, J., & Lewerin, S. S. (2013). A survey of visitors on swedish livestock farms with reference to the spread of animal diseases. *BMC Veterinary Research*, 9. <https://doi.org/10.1186/1746-6148-9-184>
- Palisson, A., Courcoul, A., & Durand, B. (2016). Role of cattle movements in bovine Tuberculosis spread in France between 2005 and 2014. *PLoS ONE*, 11(3), 1–19. <https://doi.org/10.1371/journal.pone.0152578>
- Pozo, P., VanderWaal, K., Grau, A., de la Cruz, M. L., Nacar, J., Bezos, J., ... Alvarez, J. (2019). Analysis of the cattle movement network and its association with the risk of bovine tuberculosis at the farm level in Castilla y Leon, Spain. *Transboundary and Emerging Diseases*, 66(1), 327–340. <https://doi.org/10.1111/tbed.13025>
- Rossi, G., Smith, R. L., Pongolini, S., & Bolzoni, L. (2017). Modelling farm-to-farm disease transmission through personnel movements: From visits to contacts, and back. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-02567-6>
- Sintayehu, D. W., Prins, H. H. T., Heitkönig, I. M. A., & de Boer, W. F. (2017). Disease transmission in animal transfer networks. *Preventive Veterinary Medicine*, 137, 36–42. <https://doi.org/10.1016/j.prevetmed.2016.12.017>
- Thiele, C. (2019). *cutpointr: Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks*. Retrieved from <https://cran.r-project.org/package=cutpointr>
- VanderWaal, K., Enns, E. A., Picasso, C., Packer, C., & Craft, M. E. (2016). Evaluating empirical contact networks as potential transmission pathways for infectious diseases. *Journal of The Royal Society Interface*, 13(121), 20160166. <https://doi.org/10.1098/rsif.2016.0166>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Watts, D., & Strogatz, S. (1998). Collective dynamics of networks. *Nature*, 393(June), 440–442.

Supplementary Material

Table 1. Number of movements by animal cause

Causes	Number of Farms	Number of Movements	Number of Animals
Reposition	27,350	119,459	1'911,288
Events	647	1,532	11,187
Reposition + Events	27,393	120,991	1'922,475
All	28,925	190,491	2'750,158

Table 2. Cutoff values to divide the centrality network metrics into the high and low levels.

Network parameter	Cutoff value
All-Degree	4
In-Degree	1
Out-Degree	2
Weighted All-Degree	33
Weighted In-Degree	1
Weighted Out-Degree	28
Betweenness	2.838115×10^{-06}
PageRank	9.983922×10^{-06}
Clust. Coefficient	0.02777778
Authority	9.553375×10^{-07}
ICC	1
OCC	1033

Figure 1. Spatial distribution of the farms positive and negative to Bovine Tuberculosis in Espírito Santo, Brazil.

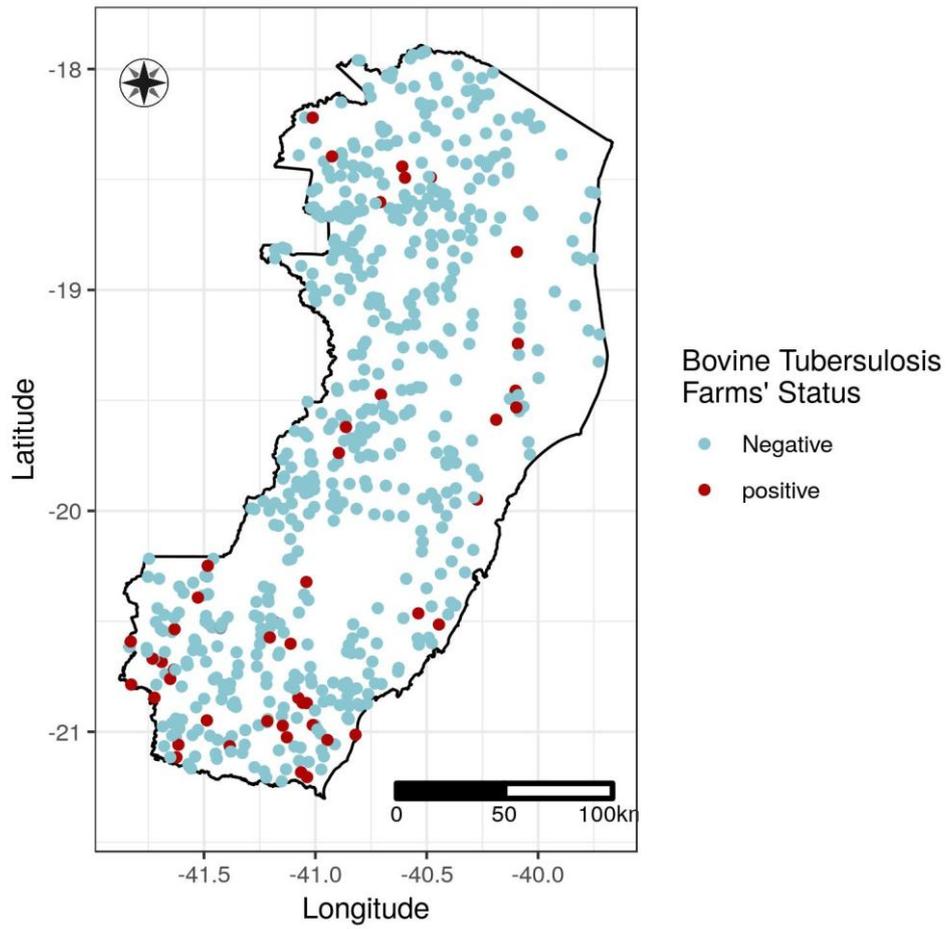


Figure 2. Distribution of network centrality metrics calculated from dairy farms positive and negative to Bovine Tuberculosis in ES.

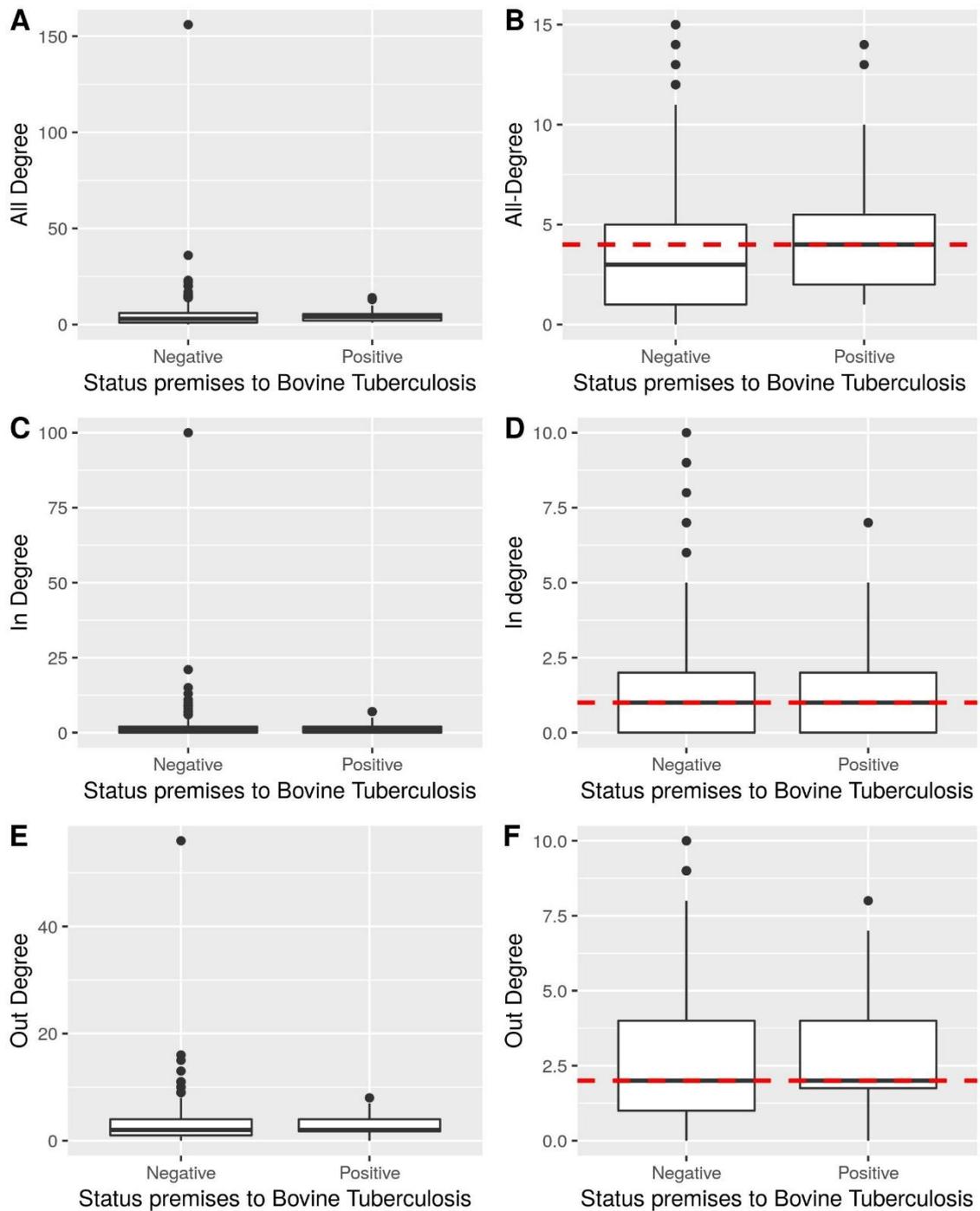


Figure 2. Distribution of network centrality metrics calculated from dairy farms positive and negative to Bovine Tuberculosis in ES (continuation).

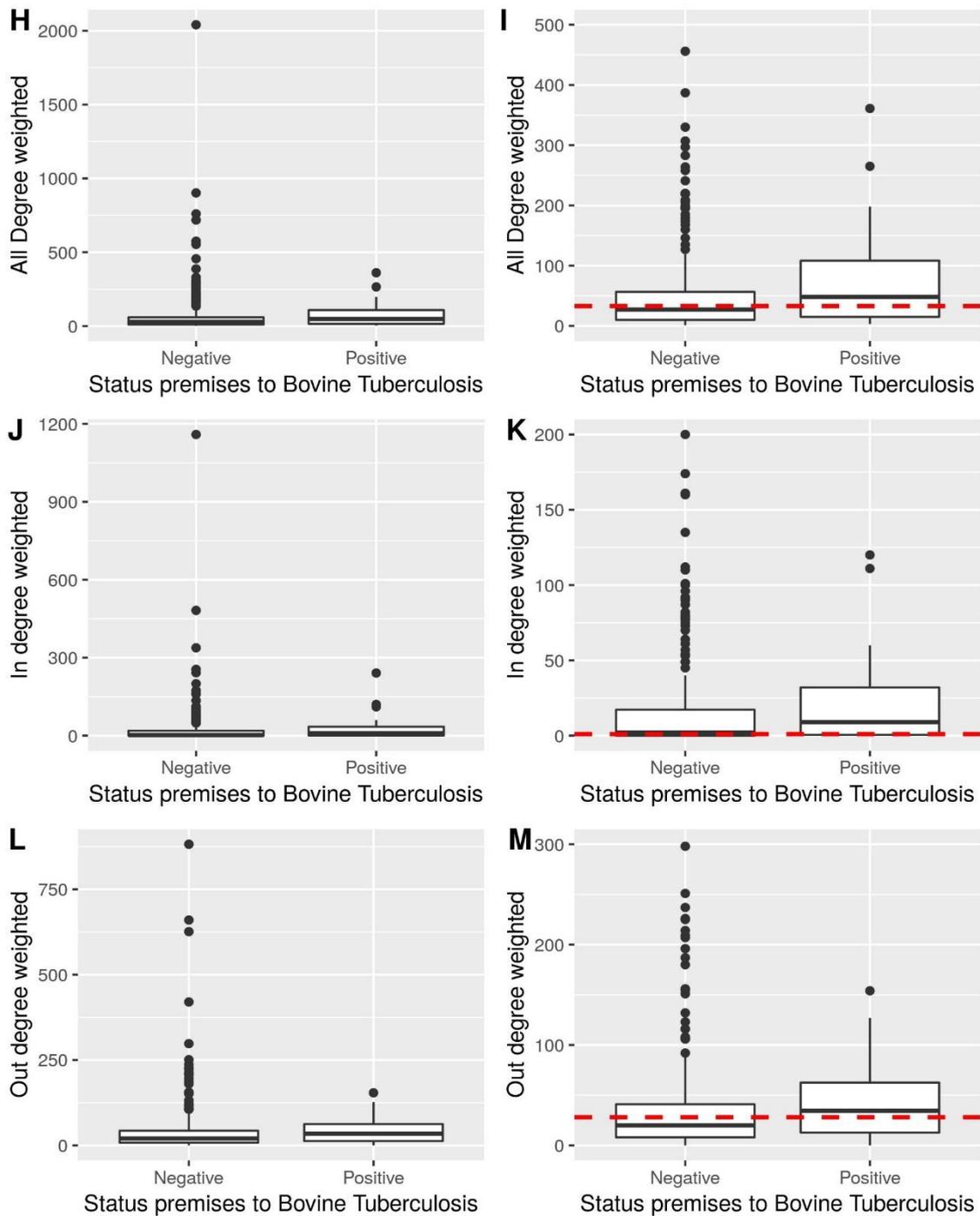


Figure 2. Distribution of network centrality metrics calculated from dairy farms positive and negative to Bovine Tuberculosis in ES (continuation).

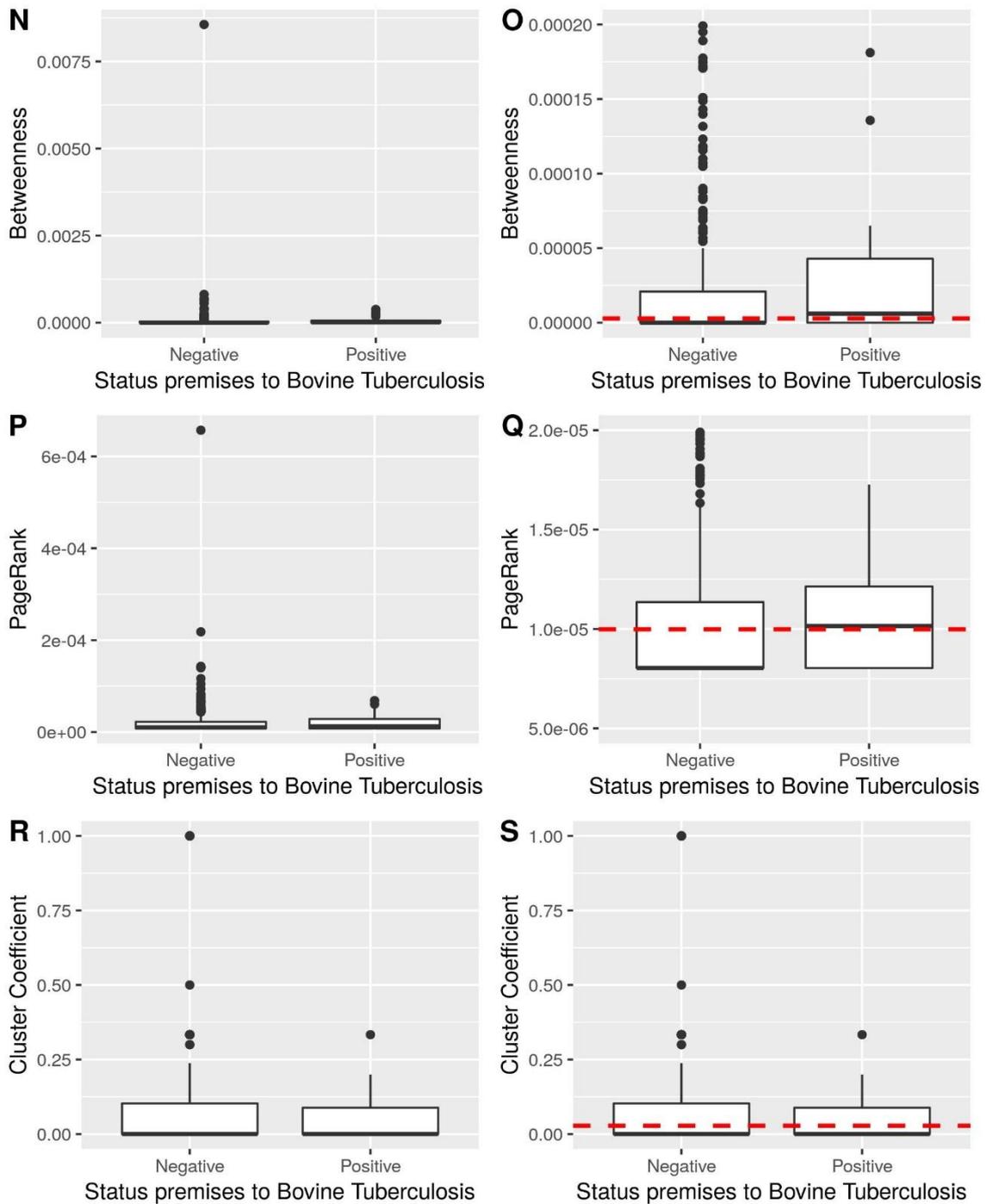
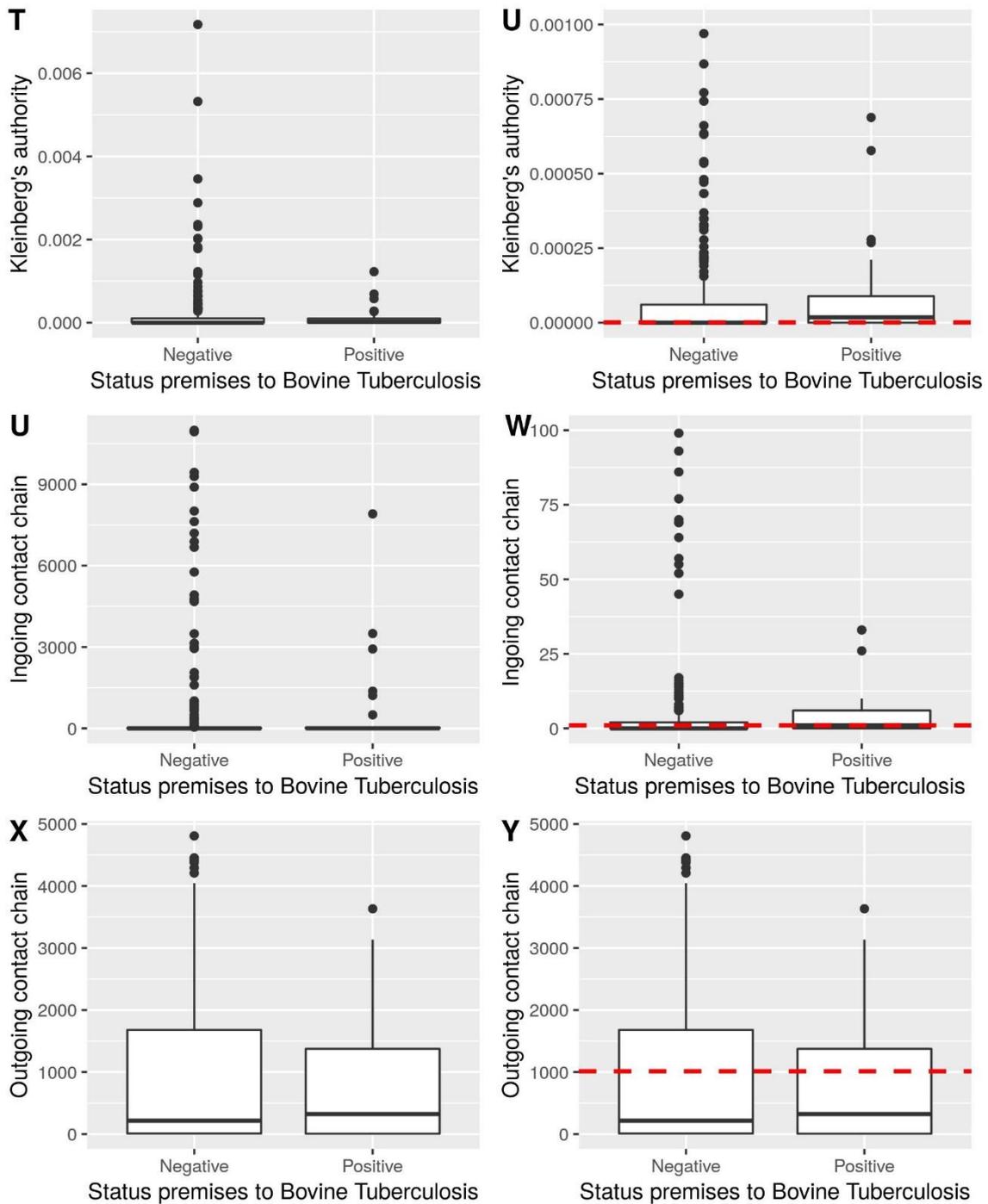


Figure 2. Distribution of network centrality metrics calculated from dairy farms positive and negative to Bovine Tuberculosis in ES (continuation).



Chapter 3: Epidemiological characterization and target surveillance of the cattle trade network in Espírito Santo, Brazil

Abstract

Livestock infections have a large impact on the economy of a country and one of the main ways to spread those diseases is by the trade movements between herds, where a pathogen can spread fast and for long distances. Therefore, it is necessary to understand the potential of the network to spread disease to propose surveillance program in the future. Our objective was describe the cattle trade network from Espírito Santo and provide a methodology to target herds presented in the trade network that contribute significantly to the risk of an epidemic and evaluate if the information from 2014 to 2016 can be useful to avoid a disease spread in 2017. We used a set of network centrality metrics to identify those livestock productions with high risk. In addition, we evaluated how the absence of those influence the connectivity of the network and the formation of groups of herds strongly connected. Finally, we simulate various different methods for targeting farms to control a hypothetical endemic disease in 2017 based in the information of the aggregated network from the previous years. We found that there were two distinct groups of farms that we have assessed to have increased significance in the spread of the disease. Also, there were many herds trading actively and with similar relevance to connect the network over the years. Finally, the use of centrality metrics, like Betweenness or Degree centrality, based in the information of a static network from previous years can reduce the transmission of the disease spread by the network, and consequently decrease the incidence of an endemic disease.

Key words: livestock productions, animal movements, animal health,

Introduction

Livestock infections have a large impact on the economy of a country. The foot-and-mouth disease outbreak in the UK in 2001 final cost was estimated up to £3.1 billion (THOMPSON et al., 2002), while Bovine Tuberculosis has been associated with substantial, persistent economic losses in livestock productions around the globe (BENNETT, 2017; Cosivi et al., 1998; Ejeh et al., 2014). Those Infectious diseases can be transmitted by several pathways between livestock productions. These pathways include contacts with wildlife populations (Fitzgerald & Kaneene, 2012; Mayen, 2003), contact between adjacent farms (Boender, Van Den Hengel, Van Roermund, & Hagens, 2014; K. VanderWaal, Enns, Picasso, Packer, & Craft, 2016) or the transmission of fomites through the movement of vehicles or humans (Mansley, Donaldson, Thrusfield, & Honhold, 2011; Nöremark, Frössling, & Lewerin, 2013; Rossi, Smith, Pongolini, & Bolzoni, 2017). The most important pathway for longer distance spread of disease is the movement of infected animals between herds for trade network, which allows pathogens to potentially reach many farms in a short period of time while also travelling long distances (Fèvre, Bronsvoort, Hamilton, & Cleaveland, 2006; Gibbens et al., 2001; Gilbert et al., 2005).

A clear example was the foot-and-mouth disease outbreak in United Kingdom in 2001, where the epidemic reached many farms along the contact chain of the trade network, including farms located at long distances from the original source of the outbreak (Gibbens et al., 2001; Rowland R. Kao, 2002). Other studies have found the existence of links between infected farms by the trade network; for example, the case of Bovine Tuberculosis (BTb) in Uruguay (K. VanderWaal et al., 2017) and the Glanders in Brazil (Cárdenas et al., 2018). Therefore, the study of these network structures is an opportunity to understand the role of farm-to-farm contact in the transmission and dynamics of diseases. In Brazil, there have been relatively few studies analysing cattle transit in the whole country, or even individually in any of the 27 states (Aragão et al., 2017; Cipullo et al., 2016; de Sá et al., 2018; Grisi-Filho et al., 2013; Silva Júnior et al., 2017). Therefore, there is an information deficiency about the impact of the animal trade network on the spread of diseases in the country.

In the analysis of animal trade networks, several metrics are available to evaluate the disease spreading potential of the network and from each individual herd within it (Dubé, Ribble, Kelton, & McNab, 2009; Martínez-López, Perez, & Sánchez-Vizcaíno, 2009). Components classification analysis (classification of the herds in groups accordingly to its trade connection) has been used to evaluate the disease

spread potential of cattle networks, and the hypothetical epidemic size if any control measure is implemented in the control of diseases (Rowland R. Kao, Green, Johnson, & Kiss, 2007; Lentz et al., 2016). Similarly, centrality metrics have been proposed to build target surveillance programs, being the herds with higher values the most relevant in some respect (Dubé et al., 2009). On the other hand, the study of animal movements with disease spread models has increased the understanding of the evolution of infectious diseases, providing tools to predict infected areas or simulate the impact of control programs (Gilbert et al., 2005; Martínez-López, Perez, & Sánchez-Vizcaíno, 2010). In the first step of this study, we described the characteristic of the cattle trade network in the Brazilian state of Espírito Santo (ES) by the use of network metrics. Besides, considering that livestock markets can change the structure of connection between herds over the years (Valdano et al., 2015), we evaluated the characteristic of the network in different time periods and the sets of herds and movements presented in each period.

Our main goal is to provide a methodology to target herds presented in the trade network that contribute significantly to the risk of an epidemic and evaluate if the information from 2014 to 2016 can be useful to reducing or eliminating farm-to-farm disease spread in the following year (2017). We used a set of centrality metrics to identify those livestock productions with high risk. In addition, we evaluated how the absence of those influence the connectivity of the network and the formation of groups of herds strongly connected. Finally, we simulate various different methods for targeting farms to control a hypothetical endemic disease in 2017 based in the information of the aggregated network from the previous years.

Methodology

Dataset

Information related to the movements of animals is in the Instituto de Defesa Agropecuária e Florestal do Espírito Santo (IDAF/ES) database from 2014 to 2017. Each movement registered has a date, source location, destination, trade purpose and number of animals moved (by sex and age). In this study, we specifically selected the movements whose trade purpose was animal replacement or participation in livestock events.

Network Analysis

The trade network was analysed as a static and directed network, taking into account the trade direction and aggregating the connections yearly for each of 2014 to 2016. This means that we also considered multiple movements between two herds as one unique movement. The general structure of the network was described using a collection of network metrics described in table 1 and calculated with the “igraph” package (Csardi & Nepusz, 2006) in R (R Core Team, 2018).

Table 1. Network metrics and definitions.

Network Parameter	Description
Density ¹	The proportion of pair of nodes (i.e. herds) that have a links (i.e. trade movement) between them (Wasserman & Faust, 1994).
Path and Shortest path ²	There is a path when a node i can reach a node j directly or by travelling along a sequence of links in the network. The length of that path is given by the number of contacts necessary to go from i to j. The shortest path among two nodes is the smallest number of contacts required to go from i to j (Wasserman & Faust, 1994).
Average Shortest path ¹	Averaged (mean) of shortest path over all pairs of nodes in the network (Watts & Strogatz, 1998).
Diameter ¹	The longest path between any pair of nodes in the network (Wasserman & Faust, 1994).
Degree ²	It is the sum of the in-degree (number of nodes source which have sent animals to a specific node) and the out-degree (number of nodes receipt which have obtain animals from a specific node) (Wasserman & Faust, 1994).
Median and percentiles of the Degree ¹	Median and percentiles of the degree over all the nodes in the network
Betweenness ²	The frequency a node is in the shortest path between pairs nodes in the network (Wasserman & Faust, 1994).
Closeness ²	Estimate of how closely connected a node is to all other nodes of the network, calculated as the sum of the length of the shortest paths between the node and all other nodes (Linton C. Freeman, 1978).
PageRank ²	The importance of a node measured by the importance of the nodes it is connected to in the network (Brin & Page, 1998).
Outgoing contact chain ²	The number of nodes reached through direct movements, and indirect contacts through further movements, and the sequence of the movements is taken into account (Nöremark et al., 2013).
Range ²	We consider the range of a node to be the number of nodes it can reach in the network within a given interval of time. This range can be evaluated by the existence of a path between two nodes in static approach (Newman, 2010) or by the existence of a path by sequential order of contacts between farms along the time (Nöremark & Widgren, 2014).

Network level¹, Node level².

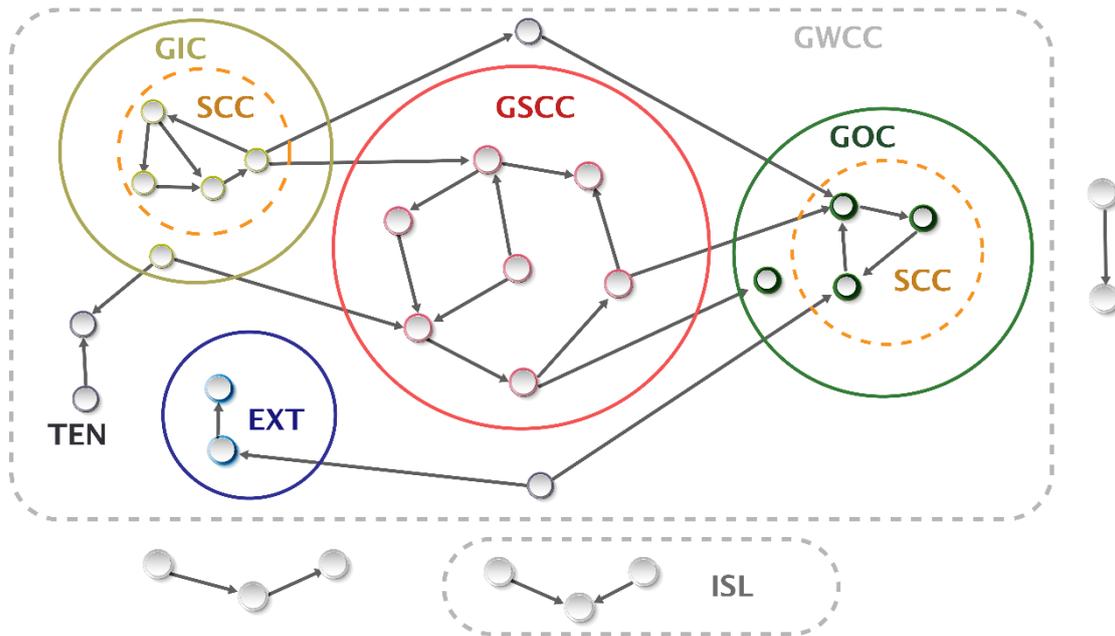
Classification of the herds in groups

The herds were classified according to the following groups:

- **Strongly Connected Component (SCC):** these are the subsets of herds with a directed paths between all pairs in the SCC (Newman, 2010).
- **Giant Strongly Connected Component (GSCC):** it is the largest SCC (Dorogovtsev, Mendes, & Samukhin, 2001; Kenah & Robins, 2008).
- **Giant In-Component (GIC):** these are set of herds that can reach any node in the GSCC by the existence of a path in the network (Dorogovtsev et al., 2001; Kenah & Robins, 2008).
- **Giant Out-Component (GOC):** these are set of herds that can be reached from any starting point in the GSCC by the existence of a path in the network (Dorogovtsev et al., 2001; Kenah & Robins, 2008).
- **Tendrils (TEN):** these are set of herds that do not belong to the GSCC, but can be reached from the GIC or can reach the GOC by the existence of paths in the network (Dorogovtsev et al., 2001; Kenah & Robins, 2008).
- **External (EXT):** these are set of herds that are part of the GWCC, but there is not a path to have access to the GOC (Lentz et al., 2016).
- **Giant Weakly Connected Component (GWCC):** it is the maximum subset of herds connected by at least one path and the direction of the edge is not important. In general, the GSCC, GIS, GOS and TEN are included in this component and the herds which are not in this component form small islands in the network (Dorogovtsev et al., 2001; Kenah & Robins, 2008).
- **Islands (ISL):** or disconnect component, these are set of herds that are not part of the GWCC and form small groups in the network (Dorogovtsev et al., 2001; Kenah & Robins, 2008).

A graphical representation of the network components groups is available in the Figure 1.

Figure 1. Classification of the herds in network components accordingly its trade connections



(Adapted from Lentz et.al. (2016)).

Network evolution over the years

We compared the sets of herds and movements between the years with a similar methodology to that described by Valdano et. al., (2015), which use the Jaccard index to evaluate the loyalty of the network and herds. We modified the formula as follows:

$$S_{i,j} = \frac{|Y_i \cap Y_j|}{|Y_i|},$$

where Y_i is the set of herds from year i and Y_j is the set of herds from the year j to compare. This method measure the proportion of herds from year i also present in year j considering only the number of herds in the year i . We applied the same method to measure the similarity of movements between the years and to compare the aggregated network from 2014 to 2016 with the network from 2017. Additionally, we calculated the how many movements in 2017 were registered individually in each one of the years of the aggregated network.

Fragmentation of the GSCC

The two questions that we want to answer in this section are: what is the best centrality metric to rank the importance of the herds in the aggregated network from 2014 to 2016, and what is the appropriate number of them that should be removed from the network in order to fragment the GSCC.

The importance of the farms within the network is assessed by the centrality metric, being those with higher centrality values the most important to connect the network in some respects. The herds are ranked from highest to lowest, according to the metrics of degree, Closeness, Betweenness and PageRank (Table 1). Thus, there is a farms importance ranking associated to each of these metrics. If two or more herds had the same value of degree centrality, the rank order was first decided by out-degree value between those, and in the case where this was equal as well, PageRank values were given precedence (the Betweenness and the Closeness had similar results). The exclusion of the herds happens gradually. First, the node of greater importance (lower ranking) for each one of the centrality metrics is removed. Then, the centrality metrics are recalculated and generate a new ranking for the remaining herds of the network. Subsequently, a new node will be chosen and removed, repeating the process of elimination, a similar methodology was used by Lentz, et al (2016).

The centrality metrics and the ideal number of herds to be removed to fragment the GSCC were evaluated by the size of the GWCC and GSCC after the removal of each single farm, up to 2000.

Comparison of herd ranking between different periods of time

We would like to know if the herds with low rank calculated from the centrality metrics with the fragmentation methodology retain similar rankings across different time periods. To measure this, we evaluated how many of those herds with low rank in each one of the years studied and the aggregated from 2014 to 2016 also had a low rank between those periods. Thus, we compared two sets of herds with the lowest rank from two different time periods using the Jaccard Index,

$$S = \frac{|N^i \cap N^j|}{|N^i \cup N^j|},$$

where N are the set of herds ranked with a centrality metrics and i and j are the time periods to be compared. In addition, we performed a deeper comparison between the aggregated period from 2014 to 2016 with 2017, since this aggregated period was the one used to propose the target surveillance in 2017. we compared the lowest ranked herds between 2014 and 2016 with the lowest ranked herds from 2017, with the difference that we compared the ranked herds starting from the one with the lowest ranking and then groping with the next herds with the follow ranking, up to group the first 2000 herds lowest ranked, using the same index

$$S_i^{2014-2016,2017} = \frac{|N_i^{2014-2016} \cap N_i^{2017}|}{|N_i^{2014-2016} \cup N_i^{2017}|},$$

where N are the herds ranked and i is the number of ranks agglomerated by sequentially order in each evaluation. This method allowed us observe if the proportion of herd with the same relevance in 2017 increase or not with more herds used in the target surveillance based in the information from 2014 to 2016.

Disease spread and control simulation

We simulate the spread a theoretical infectious disease with *SimInf*, an R package for stochastic epidemic simulations (Widgren, Bauer, & Engblom, 2016; Widgren, Bauer, Eriksson, & Engblom, 2016), which has been used in similar context (Widgren, Engblom, et al., 2016). This model works at two different transmission scales, within-herd and between-herd. The simulation incorporates empirical data, specifically the set of farms which bought or sold animals in 2017 and the movements between those farms. Transmission between animals on the same farm follows a SI (susceptible (S), infected (I)) model, with one transition in each node i ,

$$S_i \xrightarrow{\beta S_i I_i / (S_i + I_i)} I_i,$$

where β is the transmission rate, which can also be considered as the product of the contact rates and transmission probability between animals into the farms (Keeling & Rohani, 2008). In this study we assume homogenous mixing population within the farms, which correspond to all the animals in contact with equal probability and neglect heterogeneities arising from age, space or behavioral aspects.

The simulation model took into count the demography and the number of animals moved in each farm in 2017. However, we did not have information about the number of cattle in each farm (i.e. farm size) in 2017 in ES, only for 2015. Therefore, we used the information from 2015 and in case where the number of outgoing animals was greater than the farm size we set the lower limit of farm size to be zero. This allowed us to conserve the size of the farms to transmit a disease. Additionally, we included the movements of animals to slaughterhouse. These animals were randomly sampled from the farm population and once they were selected, they did not have more influence in the disease transmission.

To create the SI model, we defined an initial number of infected herds when the simulation starts (t_0). Those herds were randomly sampled from the total of herds present in the network in 2017. In this study, we wanted to evaluate the spread and control from an endemic disease, for that reason we used information to approximate

Bovine Tuberculosis (bTB), an endemic disease in ES, which was previously studied in the state, with a herd prevalence of 7% (Galvis et al., 2016). However, we only took that value to simulate the herd prevalence at t_0 in the disease transmission model and did not try to emulate all the characteristics of the bTB transmission, for that reason we simulated the disease with other herd prevalences at t_0 , considering 1%, 5% and 10%. In addition, we evaluated how different set of β , number of infected animals at t_0 , and sets of farms infected at t_0 can influence the disease incidence in 2017 without any control measure.

The farms nominated as infected at t_0 were sampled from the farms which had at least one trade movement (in or out-going) in 2017. From those which were sampled, the number of animal to be infected at t_0 in each farm was calculated in base in the total number of bovines in each one. We assumed a within-herd prevalence of 1% at t_0 , or at least 1 animal of the selected farm had a size lower than 100 animals. In addition, we tested different within-herd prevalences: 5%, 10% and 20%.

The control was introduced by the sequential elimination of the ranked herds in the network of 2017 accordingly to the centrality metrics used to fragment the GSCC between 2014 and 2016. Additionally, we included a new ranking by the Out-going contact chain (Table 1) created with the temporal network from 2014 to 2016, using the daily trade as the sequential order. We chose an arbitrary number of 2,000 herds to be removed and at the end of each simulation scenario the incidence of new cases for 2017 was calculated

$$Incidence = \frac{Number\ of\ new\ infected\ herds\ in\ 2017}{Population\ in\ risk = \frac{Number\ of\ herds\ that\ traded\ with\ other\ herds\ in\ 2017}{Initial\ number\ of\ infected\ herds\ in\ 2017}} .$$

Also, we used different sets of farms infected at t_0 and performed 10 simulations for each scenario, calculating the median incidence for each one.

Despite of the spatial proximity between the herds was thought to be included in the model as path to transmit the disease, the lack of geographical coordinates from the farms in ES made that impossible. Therefore, the simulations just have in consideration the temporal component of the network. A deeper description of the methodology used in SimInf to simulate the dynamic of a disease can be found in (Bauer, Engblom, & Widgren, 2016; Widgren, Bauer, Eriksson, et al., 2016).

All the analyses in this study were developed with R (R Core Team, 2018).

Results

Characteristics of the cattle network and evolution over the years

We found that since 2014, the number of herds and movements has decreased over the years in Espírito Santo (ES), with the largest gap between 2015 and 2016 (Table 2). On the other hand, the proportion of same herds present across the different years varied from 55 % to 72% and the movements from 7% to 15% (Figure 2A and 2B). These results suggest that each year there were many new movements between the herds in ES, but the herds which were active trading animals were similar between some years. Also, the number and proportion of herds in the GSCC and GWCC followed the same decreasing behavior (Table 2), being the maximum value 5.609 (25% of the herd in the network) in 2014. The movement density between the years was relatively similar, with low values which varied between 8.33×10^5 and 9.19×10^5 , with the higher value in 2017, but this result can be influenced by the lower number of active herds in 2017 (Table 2). Furthermore, the median of the degree distribution was the same in all the years, but percentile 95% was superior by 1 in 2014 and 2015 (Table 2).

From the aggregated network from 2014 to 2016 there were 31,490 active herds in the cattle trade network of ES and the number of movements was 101,231 along the 3-years. As expected, this network was more strongly connected than each one of the years evaluated individually, with higher density and degree values and lower values in the average shortest path and diameter.

In this study we wanted to know if it is possible to target surveillance in 2017 with the data from the aggregated network from the last three years (2016, 2015 and 2014), and for that reason we compared the similarity of herds and movements between those two periods of time. The result indicated that 23% of the movements between herds which happened in 2017 also had happened in 2014-2016 (in the static network; i.e., not considering dates of movement) and 87% of the herds which traded in 2017 had also traded in 2014-2016 (Figure 2C). These results meant that 77% of movements and 13% of herds in 2017 were new compared with the aggregated of the three previous years. Additionally, not all the 23% of movements from 2017 were registered individually in each one of the 3 years of the aggregated network, only 4% of them were registered each year, and 13% were registered only in one year (Figure 2D). Contrary, most of the 87% of the herds from 2017 were active individually each year, being that 48% of them traded each year, 23% traded in two years and 16% only one year (Figure 2D).

The components classification analysis from the aggregated network revealed that 97.4% of the herds belonged to the GWCC, while the remaining 2,6% made small groups or islands in the network (Figure 3A). Most of the herds were part of either the GSCC or GIC components, with 45.9% and 36,7% respectively (Figure 3B). Also, those two components had the longer range to reach other herds in the static (between 17,810 and 17,870 herds both groups) and temporal network (between 1 and 10,000 herds both groups) (Figure 3C and 3D). Otherwise, the GOC, the TEN, EXT and ISL had lower frequency and shorter range (< 10 herds all those groups) (Figure 3).

Table 2. Network metrics in different time periods from the cattle trade network of Espírito Santo, Brazil.

Metric/Period	2014	2015	2016	2017	2014-2016
Herds	22,161	21,706	19,446	17,804	31,490
Movements	40,954	39,997	32,736	29,122	101,231
Density	8.33x10 ⁵	8.49x10 ⁵	8.66x10 ⁵	9.19x10 ⁵	1.02x10 ⁴
Median Degree	2	2	2	2	3
Percentile 95% Degree	11	11	10	10	20
GSCC size	5,609	4,656	3,281	3,292	14,450
GSCC %	25%	21%	17%	18%	45.9%
GWCC size	20,403	20,181	17,467	15,778	30,662
GWCC %	92%	93%	90%	89%	97.4%
Av. shortest path	8.6	8.4	8.4	9.3	6,4
Diameter	27	29	31	29	21

Figure 2. Comparison of the movements and herds over time in Espírito Santo. , A) Comparison of equal movements between years, B) Comparison of equal herds between years, C) Proportion of herds and movements in 2017 that were present in the aggregated network from 2014 to 2016 and D) Frequency of movements and herds from 2017 presented in one, two or three years in the network from 2014 to 2016.

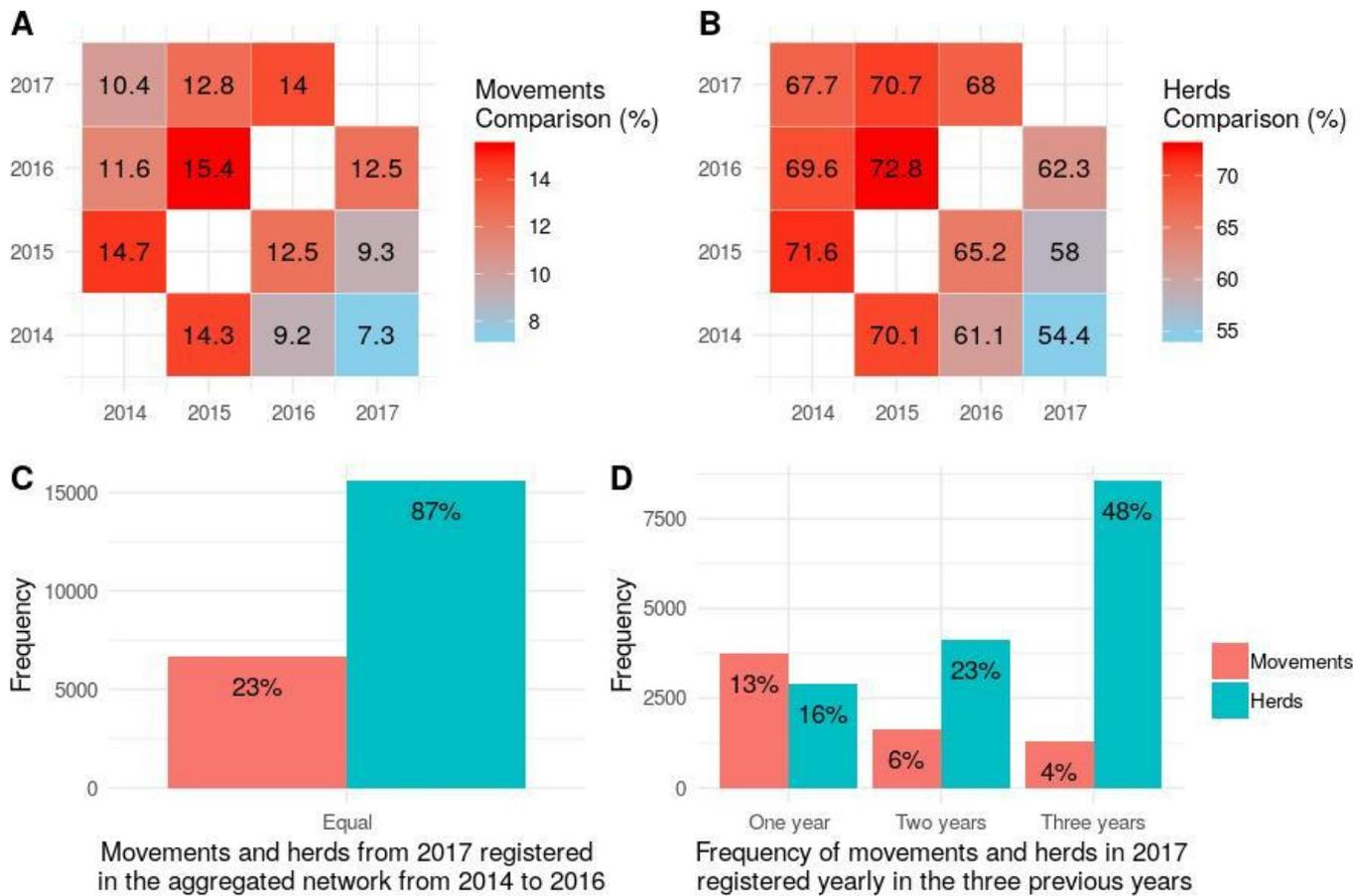
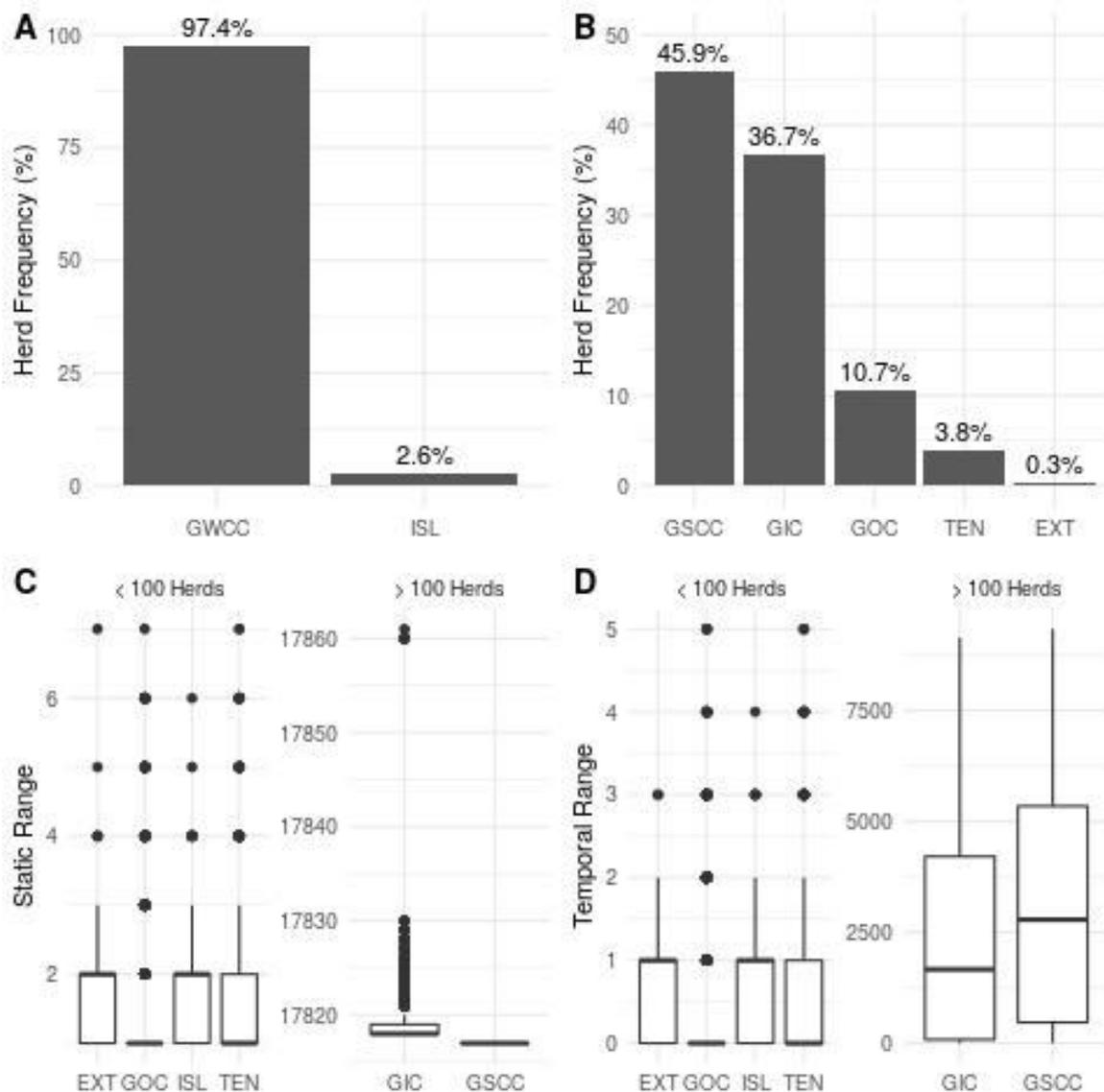


Figure 3. Components characterization and range distribution of the cattle trade network from Espírito Santo, Brazil, between 2014 and 2016. A) Proportion of herds in the GWCC and ISL. B) Proportion of herds in the GSCC, GIC, GOC, TEN and EXT. C) Distribution of the static ranges of the herds classified by the components. D) Distribution of the temporal ranges of the herds classified by the component.

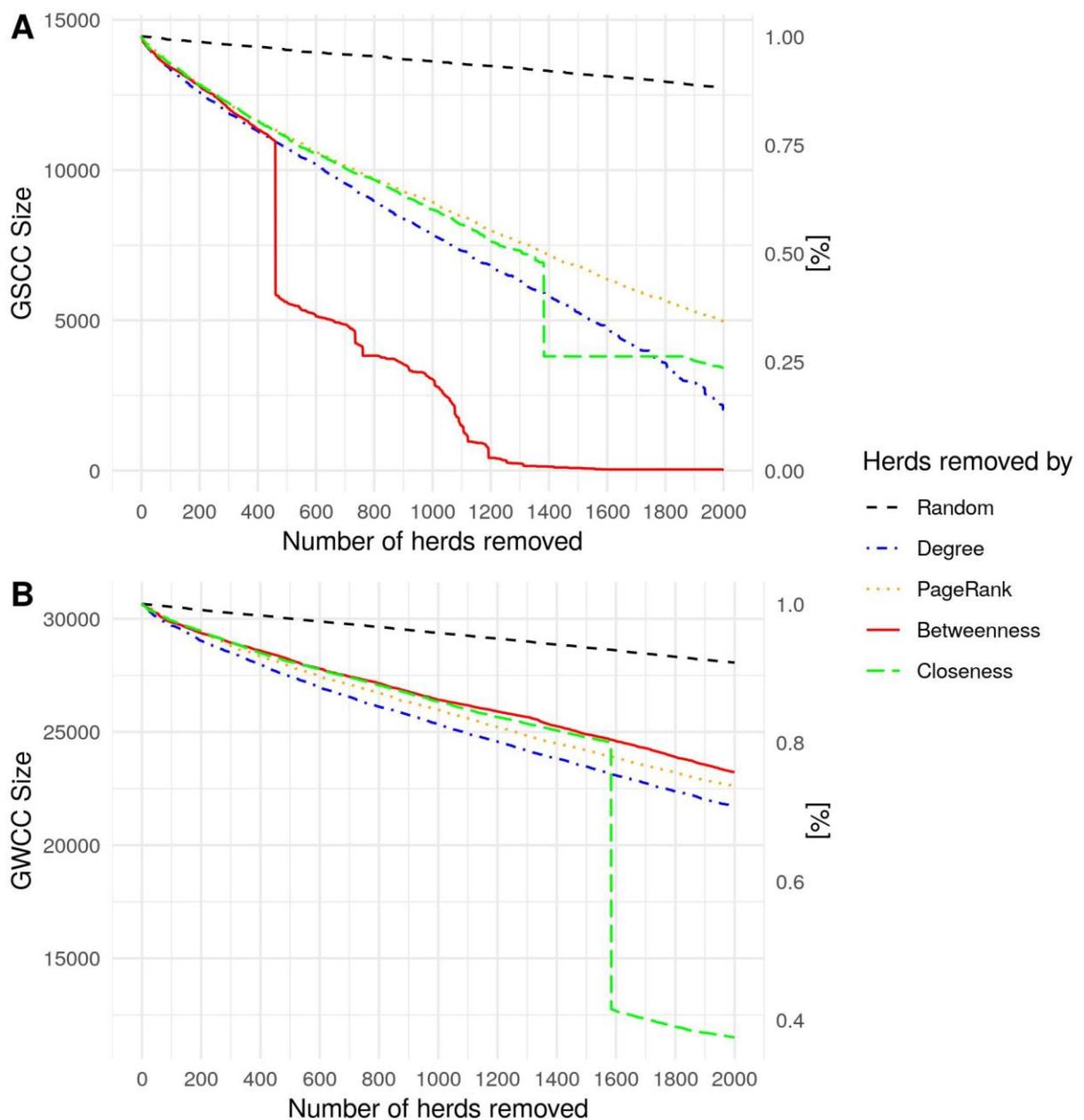


Fragmentation of the GSCC and GWCC

Evaluating the best centrality metric to fragment the GSCC, the Betweenness centrality had the best performance; this is because the GSCC size decreased 60% after 461 herds with higher values were removed, contrary to the other metrics which needed to remove more herds to get similar results (Figure 4A). Although the Betweenness got to break the GSCC after 461 herds (1,4% from the total herds between 2014 and 2016), that did not mean that the network was broken in many small

groups; in that point the network was broken in two large GSCC components with a similar number of herds, the first one with 5,089 and the second with 4,217 herds (SM table 2). In addition, the Closeness could also break the GSCC, but it required 1,400 herds to be removed, approximately 1,000 herds more than when using Betweenness (Figure 4A). On the other hand, the Closeness had a good performance to break the GWCC, but it just happened after almost 1,600 herds were removed from the network (Figure 4B).

Figure 4. Size of the GSCC and GWCC by a sequential set of herds removed in the cattle network from ES between 2014 and 2016. A) Size of the GSCC and B) Size of the GWCC.



The previous result indicated that the surveillance of herds with low ranking by the centrality metrics could reduce the size of the GSCC. However, the structure of the paths in the network changed each year and consequently, the role of the herds to connect it can have varied too. Therefore, to consider targeting surveillance with the centrality metrics it is necessary to know how many of these herds keep a constant role in in different time periods. We compared the first 500 lowest ranked herd for the comparison because with this number the GSCC was break in the fragmentation process. We found that the Degree centrality was the metric with higher proportion of herds ranked with the same relevance in the comparison of different time periods (Figure 5). Also, as expected, the aggregated network from 2014 to 2016 had high index value in the comparison with 2014, 2015 and 2016 due that those share the same information. On the other hand, for each year and all the metrics evaluated, the previous and next year had higher similarity. However, we had a short temporal data; therefore, it is not possible to infer that this is a normal pattern of the network in ES.

Focusing in the comparison between the aggregated network and 2017, the Degree had higher similarity values than the other centrality metrics in almost all the comparison (Figure 6A). This metric had a maximum of 45% of similarity when were compared the first 12 ranked herds (Figure 6B), and then varied between 44% and 32% until were compared the 2,000 first herds ranked. Otherwise, the other centrality metrics had lower proportions, being close to the Degree in some points of the first comparisons (Figure 6).

Figure 5. Comparison of the 500 lowest ranked herds between different periods of time for the centrality network metrics Degree, Betweenness, Closeness and PageRank in ES.

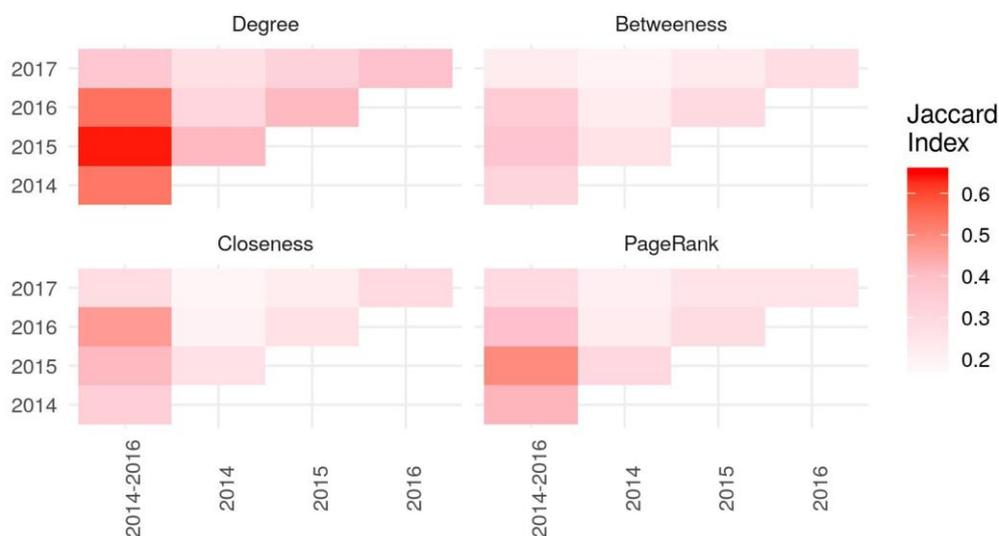
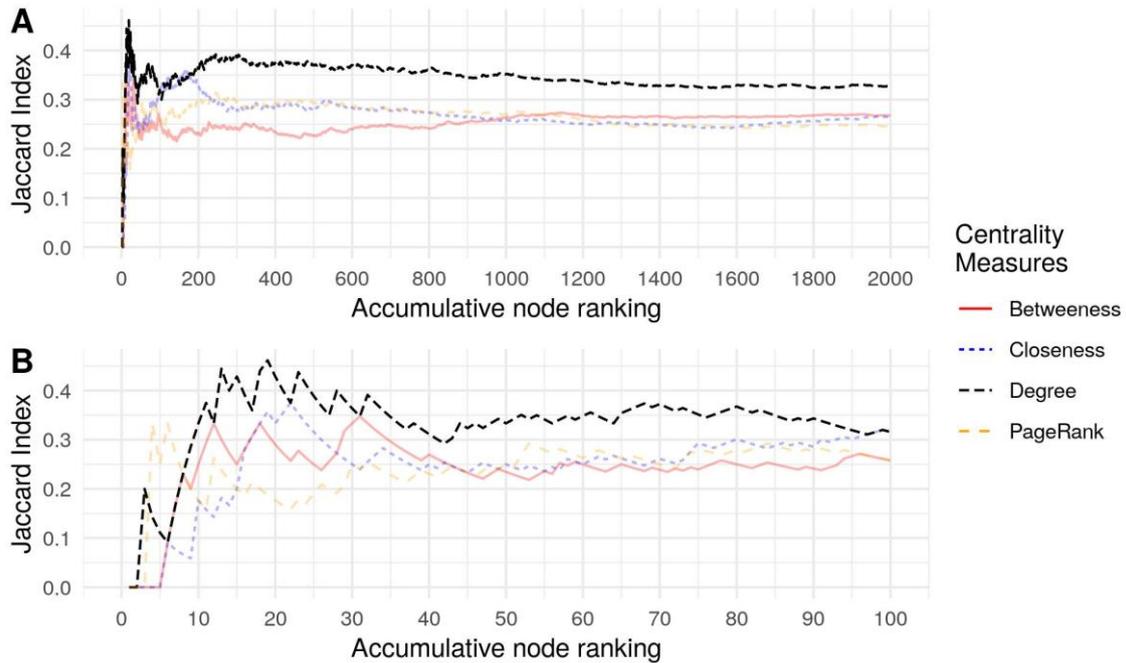


Figure 6. Sequential comparison of the ranking herds with the centrality metrics from the aggregated network between 2014 and 2016 with the network from 2017. A) Comparison until 2,000 herds and B) a subset of the comparison with the first 100 herds.



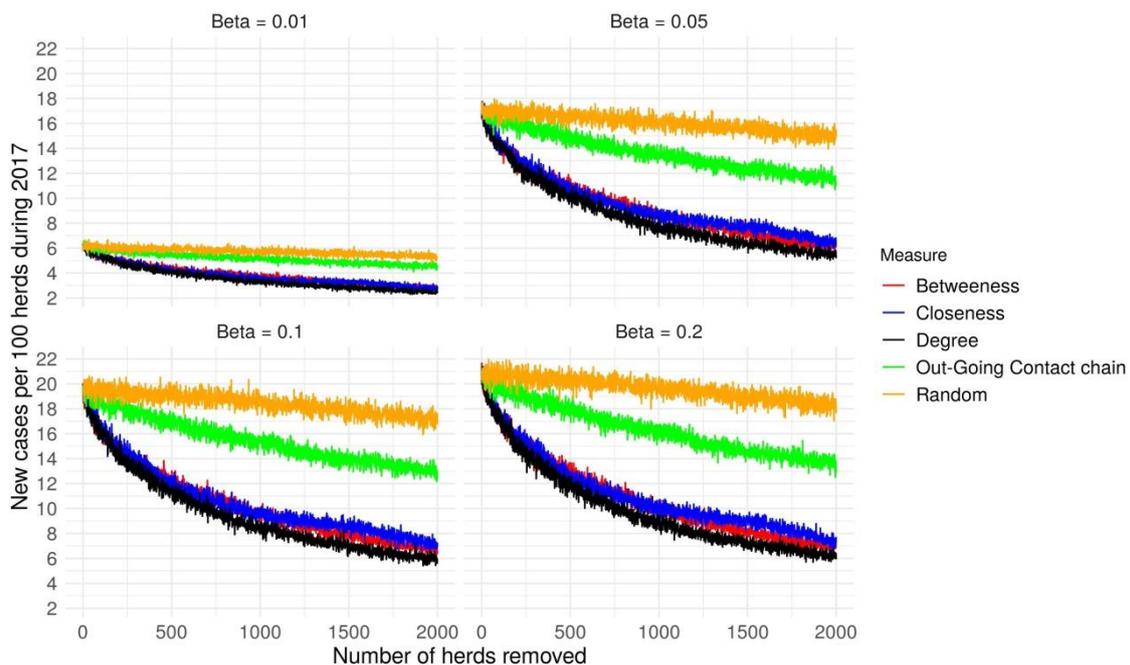
Disease Simulation

The simulation indicated that within-herd prevalence at t_0 had a small influence in the incidence with low β values ($<0,1$), and had no impact for β higher than 0,2 (SM Figure 1). However, the herd prevalence at t_0 had higher influence in the incidence values. As expected, the incidence increased for increasing of β values. Also, in most of the scenarios evaluated, the incidence kept a maximum constant value for β equal or higher than 0,2 (SM Figure 1).

Our earlier result indicated that Betweenness was good metric to break the GSCC (Figure 4A). For that reason we used that metric as control procedure to avoid a disease spread along the network. Additionally, we used the degree, closeness, the outgoing contact chain range and a random sample to establish a baseline (We exclude the PageRank because showed similar results than the others metrics). The results indicated that control targeted using the degree had a better performance to reduce the incidence than the other parameters; however Betweenness and Closeness had a similar behavior (Figure 7). On the other hand, the control with the herds which had a higher Out-going contact chain did not have a good performance compared with the other centrality metrics, with the exception of the random sample that had a

minimum impact to control the spread of the disease (Figure 7). In addition, the incidence decreased as increased of the number of herds removed from the network in all scenarios studied. Additionally, similar results were obtained when tested the simulation with 1% and 10% herd prevalence (SM Figure 2 and 3).

Figure 7. Simulation of disease spread in the cattle network from ES in 2017 considering the removal of sequential herds as control measure. It was considered 7% herd prevalence and 1% within-herd prevalence at t_0



Discussion

In this study we have shown that characteristic of the network can bring valuable information to ES to control the spread of livestock disease. We found that there were two distinct groups of farms that we have assessed to have increased significance in the spread of the disease. Also, there were many farms trading actively and with similar relevance to connect the network over the years. Finally, the use of centrality metrics, like Betweenness or Degree centrality, based in the information of a static network from previous years can reduce the transmission of the disease spread by the network, and consequently decrease the incidence of an endemic disease.

Our findings show that the cattle network between 2014 and 2017 in ES was poorly connected due to a low movement density. Other studies in cattle networks have found similar results, as 1.2×10^4 movement density in Great Britain (Fielding, McKinley, Silk, Delahay, & McDonald, 2019) or 1.2×10^5 in Sweden (Nöremark,

Håkansson, Lewerin, Lindberg, & Jonsson, 2011). Also, as seen in previous studies, cattle trade networks can follow a power law distribution, where most of the herds have few partners and only few of them have many (Christley, Robinson, Lysons, & French, 2005; Natale et al., 2009; Pozo et al., 2019; Rautureau, Dufour, & Durand, 2011). Thus, only a small fraction of possible connection between herds happens, resulting in a low network density. Therefore, low densities seem to be a consistent pattern for these networks.

The number of movements and herds was found to be decreasing over time; this could be due to changes in the cattle production every year in ES, where the effect of price of a replacement heifer, the carcass price for culled cows, price of the milk, feed, growth rate on reproductive outcomes or longer lifespan in cows could have reduced the necessity of animal replacement from other farms (Van Arendonk, 1985). Other possible explanations are the possibility of increasing unrecorded and illegal movements every year (although we do not have proof of this in ES, it is a problem discussed previously in research about cattle trade in Brazil) (Avila, Gonçalves, & Perez, 2018) or the influence of the weather in the cattle population (Lunde & Lindtjørn, 2013; Morignat et al., 2014). However, we did not research this issue in depth because were not in our mains objectives, and leave this question open for future studies in ES. We use the term "remove" to indicate selected herds were excluded from the trade network. However, in the real world, it does not mean necessarily that the selected herds are forbidden to move animals. It may mean that these will have a more rigorous surveillance control, for example vaccination and diagnostic tests for entry and exit animals, to reduce their ability to transmit a disease.

Between the network components, the Giant Strong Connected Component (GSCC), is typically one of the most significant to the spread of a pathogen (R.R Kao, Danon, Green, & Kiss, 2006; Rowland R. Kao et al., 2007; Kiss, Green, & Kao, 2006). Consequently, surveillance of herds in the GSCC can be a good strategy to protect against infectious disease (Rowland R. Kao et al., 2007). However, this group can have a large number of herds, and for logistic reasons, it is almost impossible to provide surveillance to all of them. It is thus necessary to target a smaller number of herds that play a more significant role in the spread of the disease. Then, if we monitor a large enough number, the infectious chain of the network will not have the potential to spread a disease widely, because the GSCC is fragmented or broken in small groups of herds and will not be able to reach many other herds by the network (Lentz et al., 2016). As consequence, it avoids the rapid spread of a disease in a region. We found that Betweenness centrality had the best performance to break the GSCC after 461 (1,4%)

herds were removed and the difference with other centrality metrics increased as more herds were excluded from the network. This is similar to results found in others studies indicating the Betweenness as the most effective strategy to fragment the GSCC in livestock trade networks (Lentz et al., 2016; Motta et al., 2017). In addition, the use of other centrality metrics like degree, closeness and PageRank had similar effect in the GSCC size with fewer than 461 herds removed. Therefore, if it is intended to perform a surveillance program based in the fragmentation of the GSCC in ES, the Degree, Closeness, PageRank or Betweenness can be useful when a low number of herds (<460) need to be monitored and the Betweenness can be a better option if a high number of herds are necessary in the surveillance program.

The incidence decreased as the number of herds removed from the network was increased with any control measure here tested. However, the herds ranked with the centrality metrics, Degree, Closeness and Betweenness worked better than the others, with a small difference between them. In spite that, the Degree centrality had a better performance to control the disease in all the scenarios studied. This outcome can be the result of the higher frequency of herds ranked similarly with this measure between the different periods of time studied (Figure 5 and Figure 6). Thus, the herds ranked with the Degree in one year (i.e. 2016) can have similar rank with other years (i.e. 2017), and this important to perform a target surveillance based in the rank of herd from previous years. Therefore, the surveillance of herds ranked with Degree can be a better option to control the disease spread in a network considering information from previous years.

In the disease spread simulation we showed several scenarios to spread an endemic disease in ES. However, some of them had high values of transmission rate that could not represent a real transmission value for an endemic disease. Therefore, those results must be taken with caution, particularly those with high values of beta as they represent some of the worst possible outbreak scenarios. In addition, due to the lack of information, it was not possible evaluate the distance between the farms to transmit a disease, or how the characteristic of the farms, as type of production, can influence the potential disease spreading in the network as have been identified in other studies (K. L. VanderWaal et al., 2016). Thus, the future assessment of those variables can bring valuable information to complement the epidemiological role of the farms to transmit a disease in the state.

To significantly reduce the incidence of an endemic disease with the methods presented in this study, a large proportion of herds would need to be targeted. Therefore, provide those farms with information about biosecurity plans might be easier

to implement than apply a strong surveillance program in all of them. With this methodology the most important herds in the network will be monitored first to suppress disease spreading. We propose that decision takers should, at each stage of the process, evaluate if there is a considerable cost- benefits in increasing the number of herds monitored to interrupt the transmission of a disease. The successful control of a disease depends of many other variables like pathogen characteristics, frequency of contact between infected and susceptible individuals, risk factor associated to the disease, accuracy of diagnostic test, economic impact and topography, among some others (THRUSFIELD, 2007). Therefore, the target surveillance of the herds based only on network analysis can be not enough to control a disease. The methodology presented here can, however, bring additional information to be used in conjunction with other variables to create a more efficient plan against a specific infectious disease.

References

- Aragão, S. C., Ito, P. K. R. K., Paulan, S. C., Utsunomyia, Y. T., Filho, J. H. H. G., & Nunes, C. M. (2017). Animal movement network analysis as a tool to map farms serving as contamination source in cattle cysticercosis. *Pesquisa Veterinaria Brasileira*, 37(4), 319–324. <https://doi.org/10.1590/S0100-736X2017000400004>
- Avila, L. N., Gonçalves, V. S. P., & Perez, A. M. (2018). Risk of Introduction of Bovine Tuberculosis (TB) Into TB-Free Herds in Southern Bahia, Brazil, Associated With Movement of Live Cattle. *Frontiers in Veterinary Science*, 5(October), 1–4. <https://doi.org/10.3389/fvets.2018.00230>
- Bauer, P., Engblom, S., & Widgren, S. (2016). Fast event-based epidemiological simulations on national scales. *International Journal of High Performance Computing Applications*, 30(4), 438–453. <https://doi.org/10.1177/1094342016635723>
- BENNETT, R. M. (2017). The political economy of bovine tuberculosis in Great Britain. *Revue Scientifique et Technique de l'OIE*, 36(1), 105–114. <https://doi.org/10.20506/rst.36.1.2614>
- Boender, G. J., Van Den Hengel, R., Van Roermund, H. J. W., & Hagenaars, T. J. (2014). The influence of between-farm distance and farm size on the spread of classical swine fever during the 1997-1998 epidemic in the Netherlands. *PLoS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0095278>
- Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine The Anatomy of a Search Engine. *Computer Networks and ISDN Systems*, 30(June 2000), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Cárdenas, N. C., Galvis, J. O. A., Farinati, A. A., Grisi-Filho, J. H. H., Diehl, G. N., & Machado, G. (2018). Burkholderia mallei: The dynamics of networks and disease transmission. *Transboundary and Emerging Diseases*, (July 2018), 715–728. <https://doi.org/10.1111/tbed.13071>
- Christley, R. M., Robinson, S. E., Lysons, R., & French, N. P. (2005). Network analysis of cattle movement in Great Britain. *Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a Meeting Held at Nairn, Inverness, Scotland, 30th March-1st April 2005*, (April), 234–244.
- Cipullo, R. I., Grisi-Filho, J. H. H., Dias, R. A., Ferreira, F., Neto, J. S. F., Gonçalves, V. S. P., ... Amaku, M. (2016). Cattle movement network, herd size, and bovine brucellosis in the State of Mato Grosso, Brazil. *Semina:Ciencias Agrarias*, 37(5), 3777–3792. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3777>
- Cosivi, O., Grange, J. M., Daborn, C. J., Raviglione, M. C., Fujikura, T., Cousins, D., ... Meslin, F.-X. (1998). Zoonotic Tuberculosis due to Mycobacterium bovis in Developing Countries - Volume 4, Number 1—March 1998. *Emerging Infectious Diseases*, 4(1), 59–70. <https://doi.org/10.3201/eid0401.980108>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Sy*, 1695. Retrieved from <http://igraph.org>
- de Sá, M. E. P., de Sá, C. V. G. C., Nicolino, R. R., Haddad, J. P. A., McManus, C., Seixas, L., & de Melo, C. B. (2018). Data on network of live cattle exports from Brazil. *Data in Brief*, 19, 1963–1969. <https://doi.org/10.1016/j.dib.2018.06.059>
- Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2001). Giant strongly connected component of directed networks. *PHYSICAL REVIEW E*, 64(1), 1–4. <https://doi.org/10.1103/PhysRevE.64.025101>

- Dubé, C., Ribble, C., Kelton, D., & McNab, B. (2009). A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transboundary and Emerging Diseases*, 56(3), 73–85. <https://doi.org/10.1111/j.1865-1682.2008.01064.x>
- Ejeh, E. F., Raji, M. A., Bello, M., Lawan, F. A., Francis, M. I., Kudi, A. C., & Cadmus, S. I. B. (2014). Prevalence and Direct Economic Losses from Bovine Tuberculosis in Makurdi, Nigeria. *Veterinary Medicine International*, 2014, 1–6. <https://doi.org/10.1155/2014/904861>
- Fèvre, E. M., Bronsvoort, B. M. D. C., Hamilton, K. A., & Cleaveland, S. (2006). Animal movements and the spread of infectious diseases. *Trends in Microbiology*, 14(3), 125–131. <https://doi.org/10.1016/j.tim.2006.01.004>
- Fielding, H. R., McKinley, T. J., Silk, M. J., Delahay, R. J., & McDonald, R. A. (2019). Contact chains of cattle farms in Great Britain. *Royal Society Open Science*, 6(2). <https://doi.org/10.1098/rsos.180719>
- Fitzgerald, S. D., & Kaneene, J. B. (2012). *Wildlife Reservoirs of Bovine Tuberculosis Worldwide: Hosts, Pathology, Surveillance, and Control*. 50(3), 488–499. <https://doi.org/10.1177/0300985812467472>
- Galvis, J. O. A., Grisi-Filho, J. H. H., Costa, D. Da, Said, A. L. P. R., Amaku, M., Dias, R. A., ... Ferreira Neto, J. S. (2016). Epidemiologic characterization of bovine tuberculosis in the State of Espírito Santo, Brazil. *Semina: Ciências Agrárias*, 37(5Supl2), 3567. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3567>
- Gibbens, J. C., Sharpe, C. E., Wilesmith, J. W., Mansley, L. M., Michalopoulou, E., Ryan, J. B., & Hudson, M. (2001). Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *The Veterinary Record*, 149(24), 729–743. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11808655>
- Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R., & Wint, W. (2005). Cattle movements and bovine tuberculosis in Great Britain. *Nature*, 435(7041), 491–496. <https://doi.org/10.1038/nature03548>
- Grisi-Filho, J. H. H., Amaku, M., Ferreira, F., Dias, R. A., Ferreira Neto, J. S. F., Negreiros, R. L., & Ossada, R. (2013). Detecting livestock production zones. *Preventive Veterinary Medicine*, 110(3–4), 304–311. <https://doi.org/10.1016/j.prevetmed.2012.12.013>
- Kao, R.R, Danon, L., Green, D. ., & Kiss, I. . (2006). Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597), 1999–2007. <https://doi.org/10.1098/rspb.2006.3505>
- Kao, Rowland R. (2002). The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends in Microbiology*, 10(6), 279–286. [https://doi.org/10.1016/S0966-842X\(02\)02371-5](https://doi.org/10.1016/S0966-842X(02)02371-5)
- Kao, Rowland R., Green, D. M., Johnson, J., & Kiss, I. Z. (2007). Disease dynamics over very different time-scales: Foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of the Royal Society Interface*, 4(16), 907–916. <https://doi.org/10.1098/rsif.2007.1129>
- Keeling, M. J., & Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. New Jersey: Princeton University Press.
- Kenah, E., & Robins, J. M. (2008). Second look at the spread of epidemics on networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 76(iid).

- Kiss, I. Z., Green, D. M., & Kao, R. R. (2006). The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *Journal of the Royal Society Interface*, 3(10), 669–677. <https://doi.org/10.1098/rsif.2006.0129>
- Lentz, H. H. K., Koher, A., Hövel, P., Gethmann, J., Sauter-Louis, C., Selhorst, T., & Conraths, F. J. (2016). Disease Spread through Animal Movements: A Static and Temporal Network Analysis of Pig Trade in Germany. *PLOS ONE*, 11(5), e0155196. <https://doi.org/10.1371/journal.pone.0155196>
- Linton C. Freeman. (1978). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 179(1968), 215–239.
- Lunde, T. M., & Lindtjørn, B. (2013). Cattle and climate in Africa: How climate variability has influenced national cattle holdings from 1961–2008. *PeerJ*, 1, e55. <https://doi.org/10.7717/peerj.55>
- Mansley, L. M., Donaldson, A. I., Thrusfield, M. V, & Honhold, N. (2011). Destructive tension : mathematics versus experience – the progress and control of the 2001 foot and mouth disease epidemic The index case and. *Rev. Sci. Tech. Off. Int. Epiz*, 30(2), 483–498.
- Martínez-López, B., Perez, A. M., & Sánchez-Vizcaíno, J. M. (2009). Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine. *Transboundary and Emerging Diseases*, 56(4), 109–120. <https://doi.org/10.1111/j.1865-1682.2009.01073.x>
- Martínez-López, B., Perez, A. M., & Sánchez-Vizcaíno, J. M. (2010). A simulation model for the potential spread of foot-and-mouth disease in the Castile and Leon region of Spain. *Preventive Veterinary Medicine*, 96(1–2), 19–29. <https://doi.org/10.1016/j.prevetmed.2010.05.015>
- Mayen, F. (2003). *Haematophagous Bats in Brazil , Their Role in Rabies Transmission , Impact on Public Health , Livestock Industry and Alternatives to an Indiscriminate Reduction of Bat Population*. 472, 469–472.
- Morignat, E., Perrin, J. B., Gay, E., Vinard, J. L., Calavas, D., & Hénaux, V. (2014). Assessment of the impact of the 2003 and 2006 heat waves on cattle mortality in France. *PLoS ONE*, 9(3). <https://doi.org/10.1371/journal.pone.0093176>
- Motta, P., Porphyre, T., Handel, I., Hamman, S. M., Ngu Ngwa, V., Tanya, V., ... Bronsvort, B. M. D. (2017). Implications of the cattle trade network in Cameroon for regional disease prevention and control. *Scientific Reports*, 7(August 2016), 1–13. <https://doi.org/10.1038/srep43932>
- Natale, F., Giovannini, A., Savini, L., Palma, D., Possenti, L., Fiore, G., & Calistri, P. (2009). Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine*, 92(4), 341–350. <https://doi.org/10.1016/j.prevetmed.2009.08.026>
- Newman, M. (2010). *Networks: An Introduction*. New York: Oxford University Press.
- Nöremark, M., Frössling, J., & Lewerin, S. S. (2013). A survey of visitors on swedish livestock farms with reference to the spread of animal diseases. *BMC Veterinary Research*, 9. <https://doi.org/10.1186/1746-6148-9-184>
- Nöremark, M., Håkansson, N., Lewerin, S. S., Lindberg, A., & Jonsson, A. (2011). Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance. *Preventive Veterinary Medicine*, 99(2–4), 78–90. <https://doi.org/10.1016/j.prevetmed.2010.12.009>
- Nöremark, M., & Widgren, S. (2014). EpiContactTrace: an R-package for contact

- tracing during livestock disease outbreaks and for risk-based surveillance. *BMC Veterinary Research*, 10(1), 71. <https://doi.org/10.1186/1746-6148-10-71>
- Pozo, P., VanderWaal, K., Grau, A., de la Cruz, M. L., Nacar, J., Bezos, J., ... Alvarez, J. (2019). Analysis of the cattle movement network and its association with the risk of bovine tuberculosis at the farm level in Castilla y Leon, Spain. *Transboundary and Emerging Diseases*, 66(1), 327–340. <https://doi.org/10.1111/tbed.13025>
- R Core Team. (2018). *R: a language and environment for statistical computing*. Retrieved from <https://www.r-project.org/>.
- Rautureau, S., Dufour, B., & Durand, B. (2011). Vulnerability of Animal Trade Networks to The Spread of Infectious Diseases: A Methodological Approach Applied to Evaluation and Emergency Control Strategies in Cattle, France, 2005. *Transboundary and Emerging Diseases*, 58(2), 110–120. <https://doi.org/10.1111/j.1865-1682.2010.01187.x>
- Rossi, G., Smith, R. L., Pongolini, S., & Bolzoni, L. (2017). Modelling farm-to-farm disease transmission through personnel movements: From visits to contacts, and back. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-02567-6>
- Silva Júnior, J. L., Almeida, E. C., Corrêa, F. N., Lima, P. R. B., Ossada, R., Marques, F. S., ... Silva, J. C. R. (2017). Livestock markets play an important role in the cattle movement network in Pernambuco, Brazil. *Brazilian Journal of Veterinary Research and Animal Science*, 54(3), 225–237. <https://doi.org/10.11606/issn.1678-4456.bjvras.2017.124303>
- THOMPSON, D. K., MURIEL, P., RUSSELL, D., OSBORNE, P., BROMLEY, A., ROWLAND, M., ... BROWN, C. (2002). Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. *Revue Scientifique et Technique de l'OIE*, 21(3), 675–687. <https://doi.org/10.20506/rst.21.3.1353>
- THRUSFIELD, M. (2007). *Veterinary epidemiology*. Hoboken: Wiley-Blackwell.
- Valdano, E., Poletto, C., Giovannini, A., Palma, D., Savini, L., & Colizza, V. (2015). Predicting Epidemic Risk from Past Temporal Contact Data. *PLoS Computational Biology*, 11(3), 1–19. <https://doi.org/10.1371/journal.pcbi.1004152>
- Van Arendonk, J. A. M. (1985). Studies on the replacement policies in dairy cattle. II. Optimum policy and influence of changes in production and prices. *Livestock Production Science*, 13(2), 101–121. [https://doi.org/10.1016/0301-6226\(85\)90014-4](https://doi.org/10.1016/0301-6226(85)90014-4)
- VanderWaal, K., Enns, E. A., Picasso, C., Alvarez, J., Perez, A., Fernandez, F., ... Wells, S. (2017). Optimal surveillance strategies for bovine tuberculosis in a low-prevalence country. *Scientific Reports*, 7(1), 1–12. <https://doi.org/10.1038/s41598-017-04466-2>
- VanderWaal, K., Enns, E. A., Picasso, C., Packer, C., & Craft, M. E. (2016). Evaluating empirical contact networks as potential transmission pathways for infectious diseases. *Journal of The Royal Society Interface*, 13(121), 20160166. <https://doi.org/10.1098/rsif.2016.0166>
- VanderWaal, K. L., Picasso, C., Enns, E. A., Craft, M. E., Alvarez, J., Fernandez, F., ... Wells, S. (2016). Network analysis of cattle movements in Uruguay: Quantifying heterogeneity for risk-based disease surveillance and control. *Preventive Veterinary Medicine*, 123, 12–22. <https://doi.org/10.1016/j.prevetmed.2015.12.003>
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Watts, D., & Strogatz, S. (1998). Collective dynamics of networks. *Nature*, 393(June),

440–442.

- Widgren, S., Bauer, P., & Engblom, S. (2016). *SimInf: a framework for data-driven stochastic disease spread simulations*. <https://www.CRAN.R-project.org/package=SimInf>.
- Widgren, S., Bauer, P., Eriksson, R., & Engblom, S. (2016). *SimInf: An R package for Data-driven Stochastic Disease Spread Simulations*. Retrieved from <http://arxiv.org/abs/1605.01421>
- Widgren, S., Engblom, S., Bauer, P., Frössling, J., Emanuelson, U., & Lindberg, A. (2016). Data-driven network modelling of disease transmission using complete population movement data: spread of VTEC O157 in Swedish cattle. *Veterinary Research*, *47*(1), 1–17. <https://doi.org/10.1186/s13567-016-0366-5>

Supplementary material

Table 1. GSCC groups found in the bovine cattle network from Espírito Santo between 2014 and 2016

Number of herds by group	Frequency of the groups	Total herds in the groups	Proportion of the herds by group
1	15,900	15,900	0.00
2	425	850	0.01
3	73	219	0.01
4	12	48	0.01
5	2	10	0.02
6	1	6	0.02
7	1	7	0.02
14.450	1	14,450	45.89

Table 2. GSCC groups found in the bovine cattle network from Espírito Santo between 2014 and 2016 after 461 herds were removed.

Number of herds by group	Frequency of the groups	Total herds in the groups	Proportion of the herds by group
1	16,839	16,839	0.00
2	650	1,300	0.01
3	131	393	0.01
4	32	128	0.01
5	12	60	0.02
6	6	36	0.02
7	2	14	0.02
40	1	40	0.14
4.908	1	4,908	16.60
5.852	1	5,852	19.79

Figure 1. Simulation of the dynamic of a disease in the cattle network from ES with different β values and different herd prevalence, within herd prevalence and sets of herds infected at t_0 .

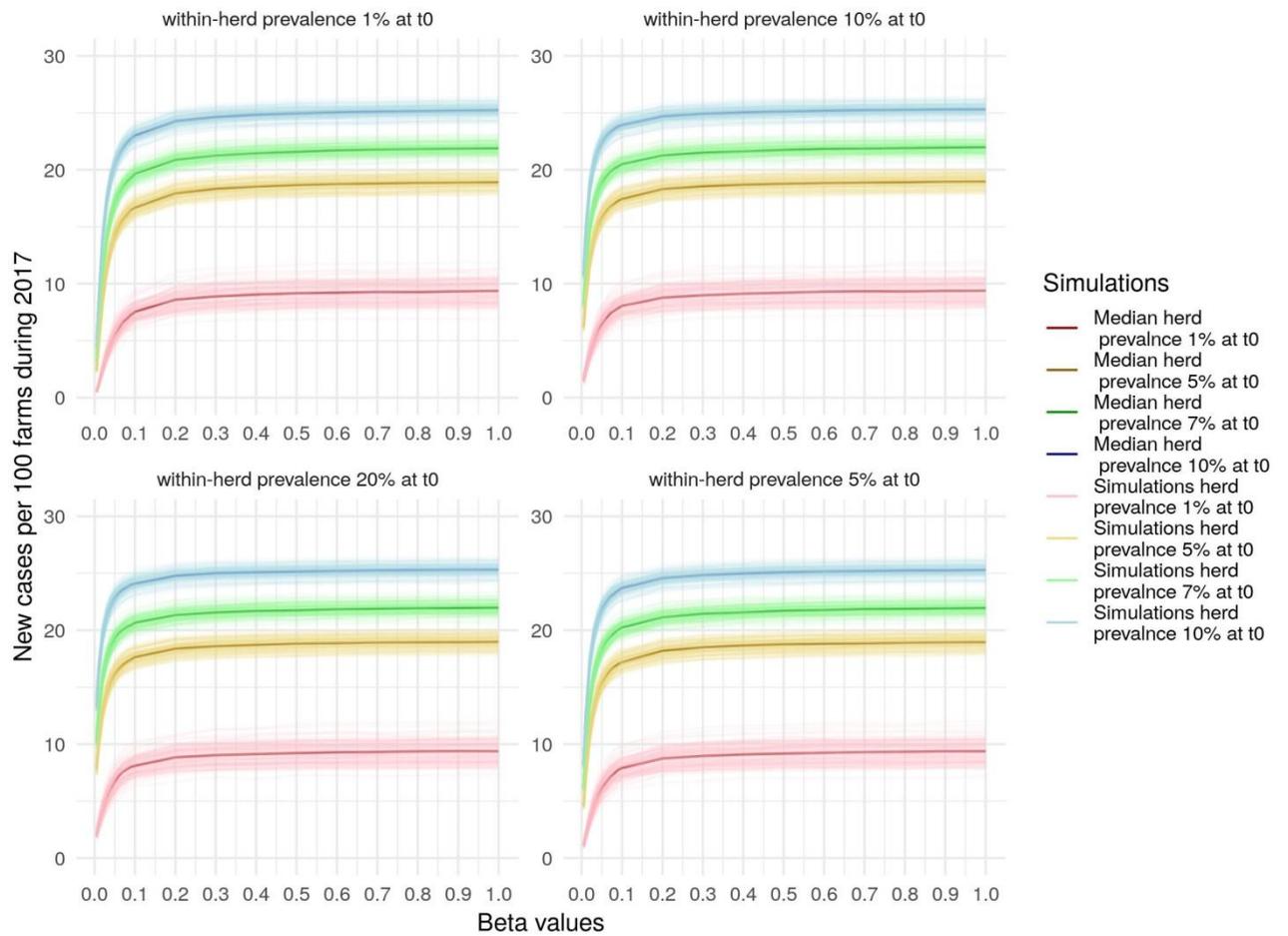


Figure 2. Simulation of disease spread in the cattle network from ES in 2017 considering the removal of sequential herds as control measure. It was considered 1% herd prevalence and 1% within-herd prevalence at t_0

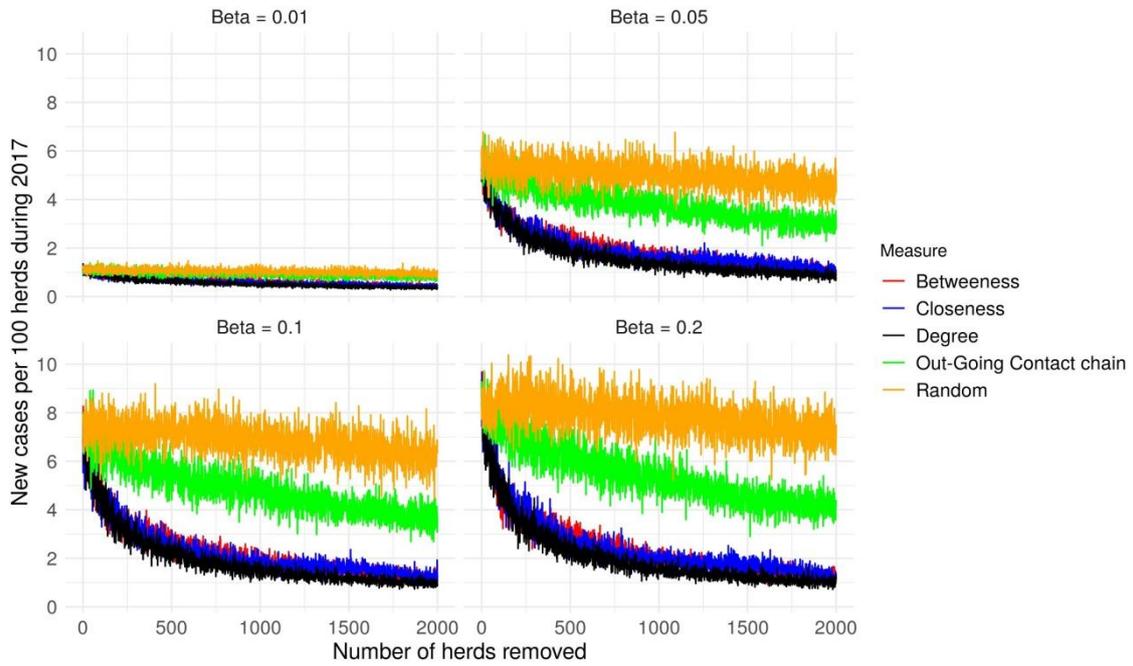
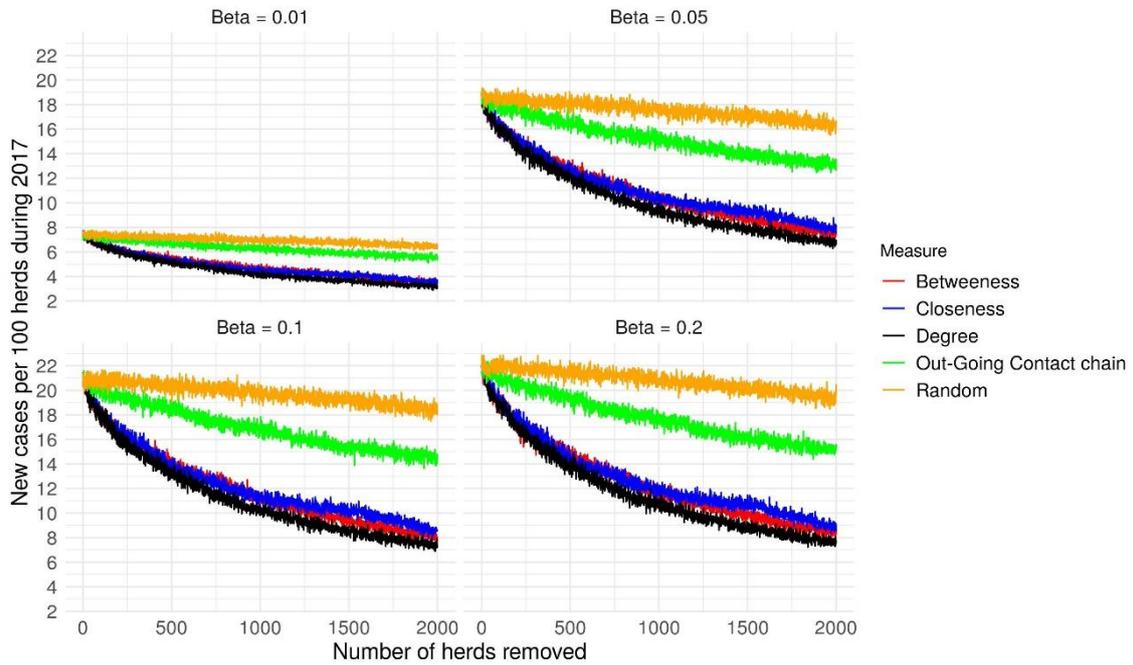


Figure 3. Simulation of disease spread in the cattle network from ES in 2017 considering the removal of sequential herds as control measure. It was considered 10% herd prevalence and 1% within-herd prevalence at t_0



Chapter 4: Cattle farm spatial location recovery through animal trade movements

Abstract

Information on spatial farm distribution is crucial to effectively design animal disease surveillance programs. Therefore, methods to compensate the lack of information on farm spatial is of paramount important to successfully implement such programs. The present study aims to provide a methodology to predict farm spatial location based on animal trading exchanges from or to other farms, and evaluate if this methodology can be applied in spread disease modeling in a scenario in which there is a set of farms without spatial coordinates. Our main hypothesis is that the distance between a target farm and its trade partner may be a good predictor to find its possible location within the specified municipality. For this purpose, we used farms which had incoming cattle or outgoing cattle from two Brazilian states. To predict the location of a target farm in its municipality, we divided the municipality into cells and calculated the weight for each one based on the distance to the trade partners. Then, cells with higher values where the target farm was more likely to be located were used. To assess if our methodology works better than a random coordinates sample in its municipalities, we compare them evaluating the distance to the actual farm location. Also, we compared the prevalence calculated from a disease spread model with the actual, predicted, and random coordinates. The results of this study show that our methodology was more efficient than a random coordinates sampling. Thus, the methodology presented herein may be useful to aid spatial analysis in a group of farms without spatial coordinates.

Keywords: missing coordinates, cattle herds, livestock production, spatial, disease spread coordinates

Introduction

In order to create an efficient surveillance program to prevent epidemics, knowledge of the system in which such program will be implemented is fundamental since there are several variables associated with the host, pathogen, and environment which can influence disease spread in one particular region (Thrusfield, 2007). As a result, decision makers must consider all the information available in order to develop an accurate and adequate strategy to suppress a pathogen preventing the spread of an infectious disease within herds and between herds. However, poor data quality or lack of information may have a negative impact on the efficacy of these programs.

On top of characteristics such as rearing practices in place and the network of animal movements, farm spatial locations are also crucial information to effectively design such programs as spatial proximity is a useful proxy for many disease risks. (M. J. Tildesley et al., 2010). Therefore, knowledge on the spatial information may improve evaluation of risks for potential disease spread, especially for livestock diseases where spread has strong spatial characteristics, either through direct transmission or due to common risk factors including wildlife reservoirs or insect vectors (e.g. foot-and-mouth disease, bovine tuberculosis, bluetongue virus, African swine fever, and avian influenza) (Gibbens et al., 2001; Gulenkin, Korennoy, Karaulov, & Dudnikov, 2011; Jacquot, Nomikou, Palmarini, Mertens, & Biek, 2017; Palisson, Courcoul, & Durand, 2016; Ssematimba et al., 2018). Moreover, it may affect epidemiological studies that evaluate risk factors associated with topology. It may also slow down or delay the response of decision makers in the case of a new epidemic outbreak. As a result, methods to compensate the lack of farm spatial information are of paramount importance during the implementation of disease control measures.

Previous studies have addressed the lack of farms coordinates with the use of land cover data to create synthetic locations to model disease spread but have not inferred the predicted locations to a specific farm (Michael J. Tildesley & Ryan, 2012). In contrast, methodologies such as triangulation and trilateration have been used with the aim to determine the position of unknown spatial points (Feigl, 2002). These methods use the angles or distances calculated from the closest benchmarks with a known position in space to find a probable area where the missing point can be located. Usually, these benchmarks are identified by signals (visual signals or satellite signals) to link the target points to the closest benchmarks around these components (Feigl, 2002). Regarding prediction of farm locations, we may use the location of other farms as benchmarks. Unfortunately, signal records are not usually available to identify farms

close to a target farm (i.e. a farm that we are interested in predicting its location). However, animal movements between a farm and its trade partners may solve this problem, because these movements due to logistical and economic reasons tend to occur in close distances (Ensoy, Faes, Welby, Van der Stede, & Aerts, 2014; Robinson & Christley, 2007). As a result, a target farm possibly will be close to its trade partner. The main hypothesis is that the distance between a target farm and its trade partners may be a good predictor to find its possible location within the specified municipality. The objective of the present study was to provide a methodology to predict farm spatial location based upon animal trading exchanges from or to other farms, and evaluate if this methodology can be applied in spread disease modeling in a scenario which there is a set of farms without spatial coordinates.

Methodology

This methodology was developed and applied to cattle rearing system from two Brazilian states (state A and state B).

Data set

Information regarding to the animal movements (between 2016 and 2017) and farms were provided by the local veterinary authority in Brazil. Each one of these movement records has date, species, farm and municipality source and destination code, trade purpose and number of animals moved (by gender and age).

In this study, we selected movements involving cattle with only farm-to-farm movements (thus excluding movements to and from events and slaughterhouses) and made only within the states (movements involving other states were excluded). We also considered multiple movements between two herds as one single movement. In addition, we used the farm census database for information retrieval for each farm registered in the area in 2017. Information available included municipality (e.g. county) and spatial coordinates (latitude and longitude). We evaluated which farms had a good data quality on coordinates by checking if there were available coordinates and if these had spatial correspondence with their municipality.

Recovery of missing farms

Information used to predict farm location was the distance measured in kilometers (km) between a farm and its trade partners according to cattle contacts between 2016 and 2017. Inter-farm distance was measured between farms which had good data quality

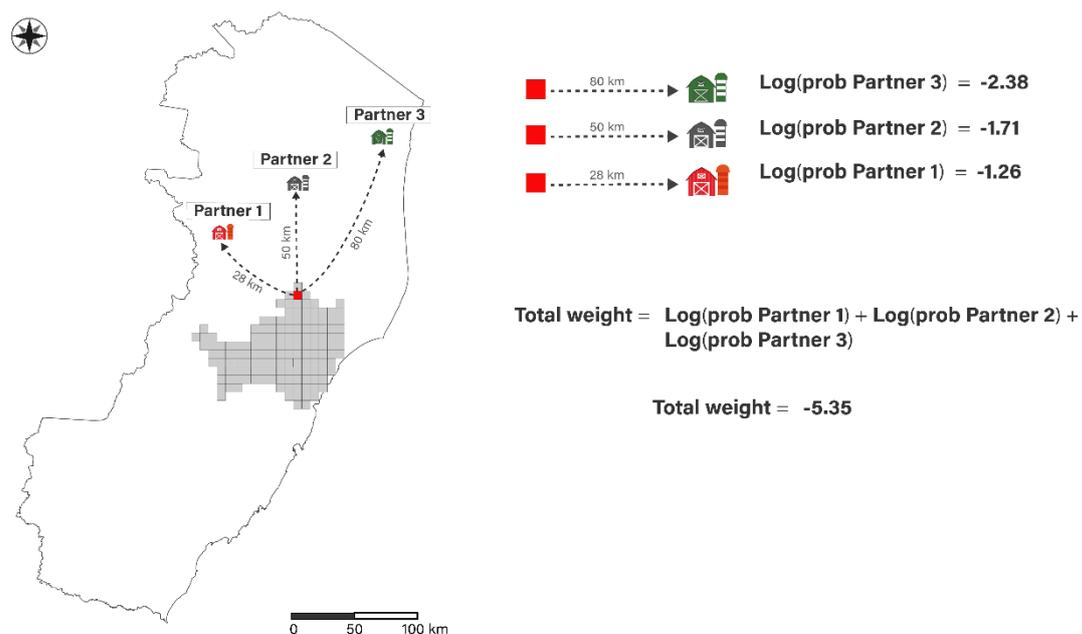
on its coordinates. We calculated a geodesic distance in which two farms are connected by trade links using a method that considers the curvature of the earth (Karney, 2013). Then, we evaluated the distribution of all the distances calculated and fitted it with a theoretical distribution, with the purpose of getting an events probability distribution. We checked visually distributions including Normal, Log-Normal, Gamma, uniform Weibull, and Exponential on our data, and chose the one that most suited the distances.

We attempted to predict a possible zone where a target farm can be located in its municipality. To find this zone, we divided that municipality into a grid of cells of 1km² and calculated the geodesic distance (km) of each one of those cells to the trade partners from a target farm. Thus, each cell has a set of distances stored. Then, these distances were matched with the theoretical distribution (previously calculated) to obtain the probabilities of the events and transformed these to a logarithmic scale to facilitate the analysis. Finally, we summed the logarithmic probabilities in each cell in order to obtain the total weight by cell,

$$\text{Total weight cell} = \sum_{i=1}^n a_i = \log(a_1) + \log(a_2) + \log(a_3) + \dots + \log(a_n),$$

In which a is the probability value of the partner i be located in that cell. An example of the methodology is provided in the Figure 1. We considered that the cells with higher total weight were the ones where the target farm was more likely to be located.

Figure 1. Example of the calculation of the total weight of one cell (red square cell)



To assess prediction accuracy, we calculated the distribution of distances from the actual farm location to the one predicted by our model. As a second accuracy measurement, for each location predicted we assigned a ranking value (from 1 to the total number of cells in the municipality) to all municipality cells based on the total weight values. Then, we identified the cell where the farm (with actual coordinates) was located (SM Figure 1). As the number of cell in the municipality is roughly equal to the municipality area (km^2), this procedure allowed us to estimate the percentage of municipality area that was necessary to find the actual location of each farm. Finally, we fitted an empirical cumulative distribution in both of these measures.

To test if our method was better than a random model, we sampled 100 random cells in the municipality for each farm predicted and calculated the median distances distribution from the actual farm location to the random locations. Next, we fitted an empirical cumulative distribution and compared with the results obtained previously from distance between the farm actual location and our predicted location.

In addition, we calculated some parameters that might be associated with the prediction accuracy of the farm locations. These parameters consisted in municipality and network information from the target farms (see Table 1). Similarly, to understand which variables mostly affected the model ability to predict farm location, we built a scatterplot with each one of these variables, in function of the log distance of the predicted farms location to the actual locations and fitted this relationship with a smooth function from a Generalized Additive Model (GAM).

Evaluating the disease spread model with the predicted coordinates

We simulate the spread of a theoretical infectious disease considering animal movements and the distance (geodesic) between the farms. This model works at two different transmission scales - within herd and between herd - and the transmission between animals in the same farm follows the SI (susceptible (S), infected (I)) model. Also, the model took into account the demography and the number of animals moved in each farm. However, information on the number of cattle in each farm (i.e. herd size) was not available. Therefore, we specified an arbitrary number of 10,000 animals in each farm.

Table 1. Variables calculated from the target farms.

	Variables	Description
Cattle trade network	Partners	Number of trade partners from a target farm
	Movements	Number of movements made by a target farm
	Partners same municipality	Number of trade partners located in the same municipality of a target farm
	Partners different municipality	Number of trade partners located in a different municipality of a target farm
	Proportion Partners same municipality	Proportion of trade partners from the same municipality of a target farm (partners from the same municipality/total partners)
	Distance between partners	Median distance (km) between trade partners from a target farm
	Municipality distance	Median distance (km) from the centroid of the target farm municipality to the centroids of all the partners' municipalities.
	Municipality border distance	Distance (km) from the predicted area to the closest municipality border.
	Distance between predicted area and trade partners	Median distance (km) from the predicted area of a target farm to all its trade partners.
	Partner spatial distribution Up or Down	Spatial distribution of the trade partners around the cell with highest total weight, when most of the partners (>75%) from a target farm are located up or down from the cell with highest total weight, it means that the prediction has an unbalanced partner spatial distribution, if not it means that have balanced partner spatial distribution.
Partner spatial distribution Right or Left	Spatial distribution of the trade partners around the cell with highest total weight, when most of the partners (>75%) from a target farm are located to the right or left from the cell with highest total weight, it means that the prediction has an unbalanced partner spatial distribution, if not it means that have balanced partner spatial distribution.	
Partner spatial distribution Total	If the farm location prediction has unbalanced partner spatial distribution Up or Down and/or Right or Left, it means that the prediction has an unbalanced partner spatial distribution, if not it means that have balanced partner spatial distribution.	
Municipality	Farms	Number of farms registered in the corresponding municipality from a target farm.
	Area Density	Target farm municipality area (km ²) Density of farms in the corresponding municipality from a target farm (number of farms by km ²)

We tried to simulate scenarios where our methodology might be used. We chose scenarios removing randomly 5%, 10%, 20%, 30%, 40%, and 50% farms coordinates. Subsequently, we predicted new coordinates using our methodology. However, our procedure worked properly in farms which had at least two trade partners during the time period being studied. Some of these premises did not have enough number partners. Therefore, for these farms, we sampled a random point into its corresponding municipality. Finally, we compared the prevalence from 2016 to 2017 between three scenarios as follows: 1) using the actual coordinates for all farms, 2) applying our methodology and recovering a fraction of the farms coordinates, and 3) sampling random coordinates in the municipality for all the missing farms coordinates.

The disease spread model was constructed using *SimInf*, an R packaged software for stochastic epidemic simulations (Widgren, Bauer, & Engblom, 2016; Widgren, Bauer, Eriksson, & Engblom, 2016). We built the model with the following parameter: herd prevalence of 1%, within-herd prevalence of 5%, decay of the environmental infectious pressure of 0.005, distance cutoff to transmit the disease within 1km, initial environmental infectious pressure of zero, indirect transmission rate of the environmental infectious pressure of 0.01, coupling between neighboring nodes (contamination among proximal nodes to capture between-node spread) of 0.1 and a shedding rate of a pathogen to the environment of 0.0003. These parameters were chosen to have a prevalence of less than 0.1 at the end of the simulations. Also, we performed 100 simulations and calculated the median value.

This methodology was applied only to state B, since this state had a higher proportion of farms with good coordinate quality than ES (Table 2). Therefore, the result of the model would be more robust.

Results

State A had 13,049 farms with good coordinate data quality, representing 29% of the farms in the state. State B had 316,730 farms with good coordinate data quality, 71% from the total farms in the state (Table 2). Also, not all the farms with good data quality in both states traded cattle between 2016 and 2017, there were 7,995 active farms (farms with animal movement records in the years studied) in state A and 144,033 active farms in state B (Table 2), those farms were responsible for 32,331 movements in state A and 440,596 movements in state B.

Table 2. Summary of the data quality of the farms registered and cattle movements from state A and state B.

	Variables	State A		State B	
		Number	%	Number	%
Farms (registered in 2017)	Registered	45,419	100	448,065	100
	Available coordinates	19,525	43	328,222	73
	Good coordinate quality	13,049	29	316,730	71
	Active in the trade network	7,995	18	144,033	32
Bovine trade network (2016-2017)	Movements	57,769	100	484,757	100
	Movements made by active farms with good coordinates quality	32,331	56	432,996	89

The distance distribution indicated that farms traded animals in short distances more frequently in both states analysed, 90% of the movements were made in less than 30 km (Figures 2 and 3) in both regions. These distributions were fitted with a Log-Normal distribution which visually had the best fit to our data compared with the others.

Figure 2. Distance distribution between trade partners between 2016 and 2017 fitted with a Log-Normal distribution (red line) in state A.

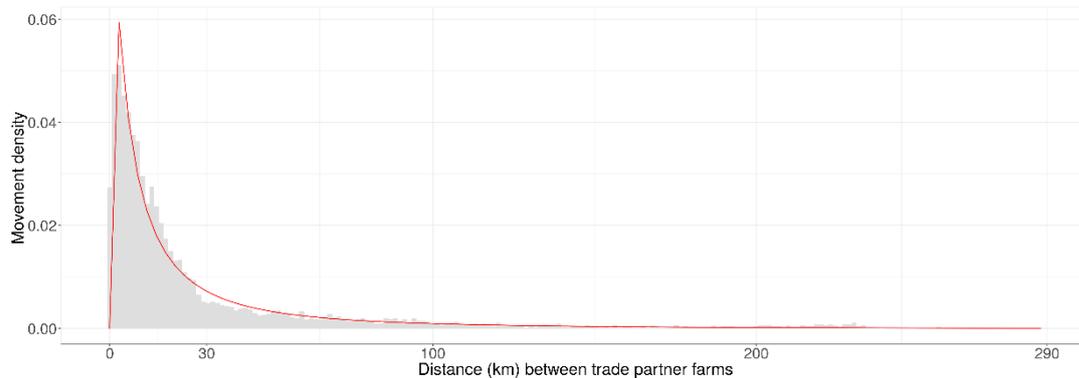
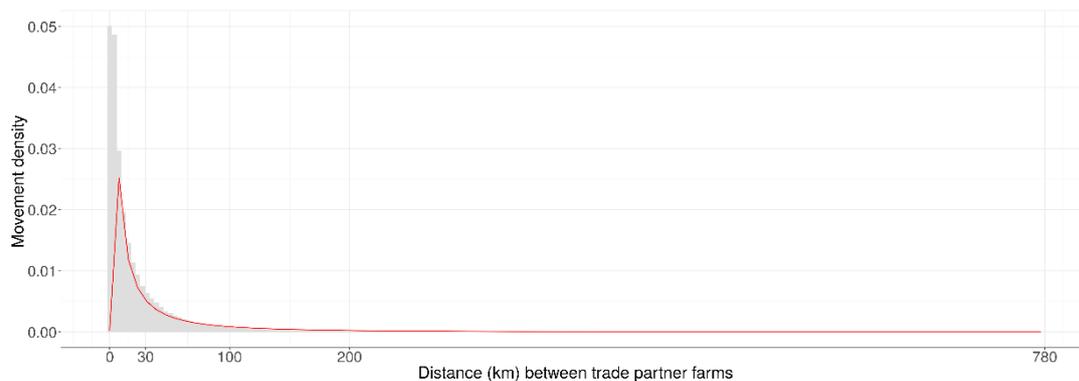


Figure 3. Distance distribution between trade partner farms between 2016 and 2017 fitted with a Log-Normal distribution (red line) in state B.



Farm location prediction accuracy

The methodology used to predict farm locations was tested on 3,393 farms with known coordinates, and active in the trade network with at least two trade partners in the years studied from state A, and an arbitrary number of 24,328 random sampled farms with known coordinates and active in the trade network from state B. Our results showed that in state A 50% of the farms were located less than 6 km radius from the areas predicted but 9 km less than using a random sample in the municipalities (Figure 4A). Also, it was possible to find 50% of the actual farms location with less than 11% of the municipality area (Figure 4B). The methodology used in state B had a better performance, 50% of the farms were located at less than 3.3 km radius from the predicted areas, 11 km less than using a random sample in the municipalities (Figure 5A), and it was possible to find 50% of the actual farms location with less than 5% of the municipality area (Figure 5B).

Figure 4. Results for state A data. Empirical cumulative distribution of A) distance distribution from the actual farm to the predicted farm location (red solid line) and from the actual farm to the random farm location (blue dashed line), and B) distribution of the percentage of municipality area needed to find the actual location of each farm for the farms prediction.

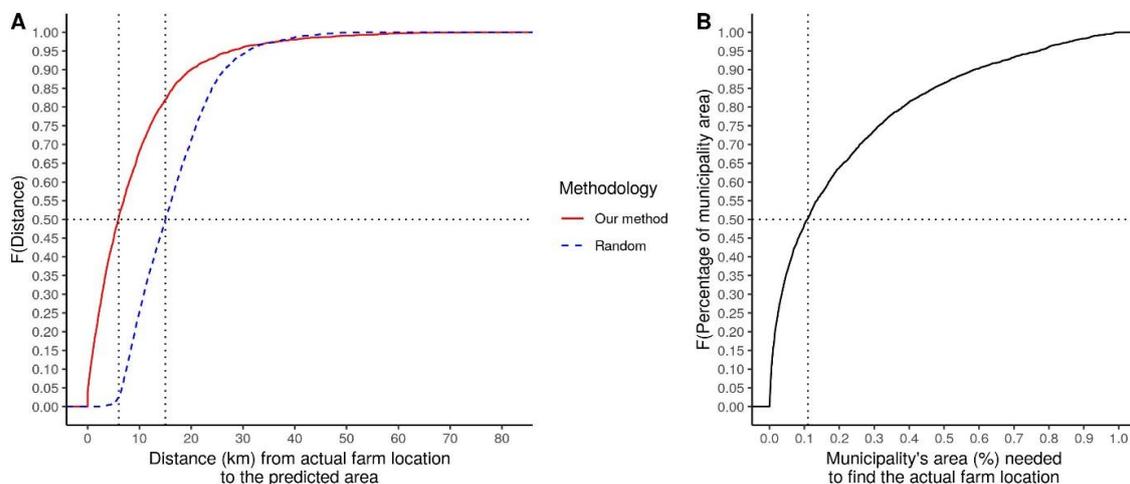
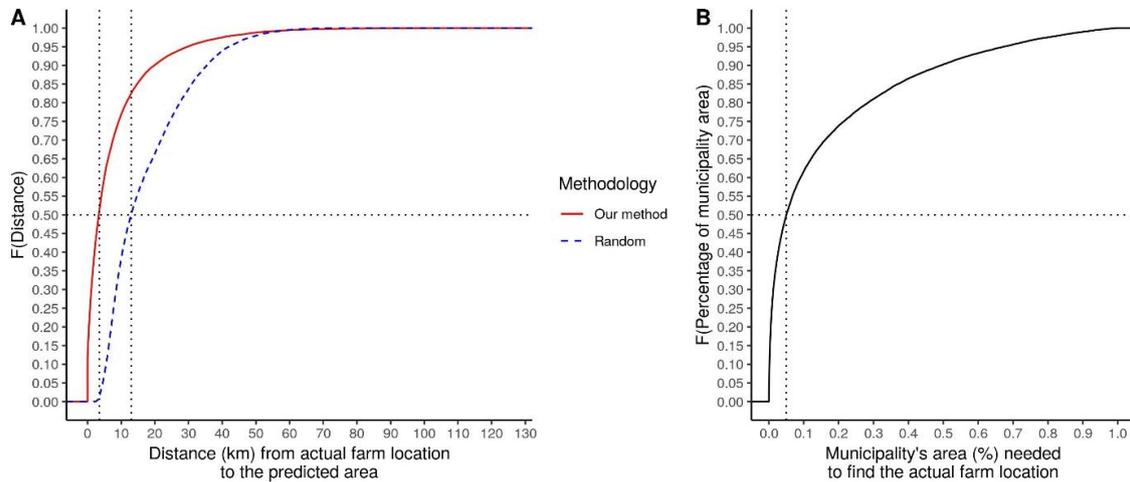


Figure 5. Results for state B data. Empirical cumulative distribution of A) distribution of distances from the actual farm to the predicted farms location (red solid line), from the actual farm to the random farm location (blue dashed line), and B) distribution of the percentage of municipality area needed to find the true location of each farm for the farms prediction.



To understand which variable may be correlated with the accuracy of our prediction method, we assessed the relationships between municipality and network variables (Table 1) with the distance from the predicted to the actual farm location. Visually, the correlations evaluated in states A and B are similar (Figures 6 and 7). Our results suggest that high values in the municipalities' areas (Figures 6N and 7N), the distance between partners (Figures 6F and 7F) and the distance between the predicted area and the trade partners (Figures 6I and 7I) increase the distance between the predicted and actual farm location. Also, the number of partners from the target farm also increase the distance. However, this relationship was stronger in state A (Figure 6A), and not so clear in state B (Figure 7A). On the contrary, high values in farm density (Figures 6O and 7O) and proportion of farm from the same municipality (Figure 6E and 7E) decrease the distance. These were the variables with clearer relationship. In contrast, the number of farms (Figures 6M and 7M) and the distance to the polygon border (Figures 6H and 7H) appears to have two distributions. The distance decreases as the values of these variables increase in the first steps. However, in later stages the distance increases as the other variables increase. In addition, the partner spatial distribution up or down (Figures 6J and 7J) and right or left (Figures 6K and 7K) around the cell with highest total weight showed that the median distance of the predictions with unbalanced partner spatial distribution was slightly larger than the predictions with balanced partner spatial distribution.

Disease spread model comparison

Our results show that the prevalence between the actual coordinates and the predicted coordinates was more similar than the prevalence between actual coordinates and random coordinates (Figure 8). Moreover, the difference between the actual, predicted and random increased as there was an increase in the proportion of missing farm coordinates. However, the difference between actual and random was always higher than between the actual and predicted coordinates. In addition, the disease prevalence with the random coordinates was always inferior to the actual and predicted coordinates. Therefore, to simulate a disease considering the space with missing farms coordinates our methodology to predict farm location worked better than a random sample of the coordinates into its municipalities.

Discussion

In the present study, we tested a methodology to predict farm coordinates based on the location of their trade partners. The results evaluating the distance and the prevalence from disease simulations models between actual and predicted locations showed that our methodology worked better than a random sample of coordinates to replace the farms actual locations.

Cattle trade network in state A and state B, as expected, followed a pattern with short distance (less than 30 km) between trade partners. Similar results were found in previous studies carried out in Belgium and United Kingdom (Ensoy et al., 2014; Robinson & Christley, 2007; Vernon, 2011). In spite of these findings, more in depth studies to identify the characteristics of the farms which made long distance movements may provide valuable information to the surveillance system (M. J. Tildesley et al., 2010), mainly because those farms can play an important role to spread a pathogen widely in a region, as it was the case of the outbreak of foot-and-mouth disease in United Kingdom in 2001 (Gibbens et al., 2001).

Figure 6. Correlation between network and municipality characteristics, and the log distance from the predicted area to the actual farm location calculated from 3,393 target farms in state A.

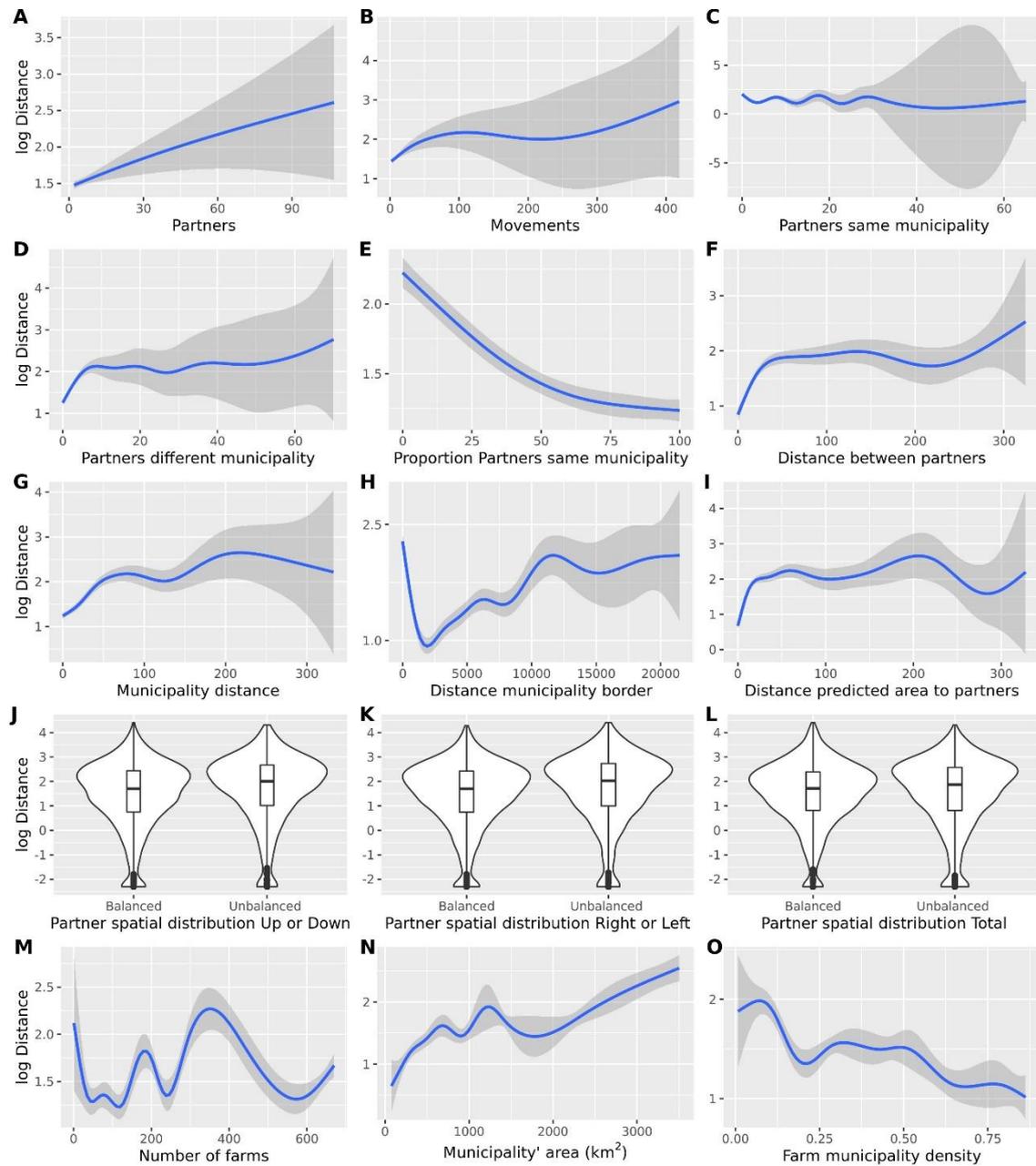


Figure 7. Correlation between network and municipality characteristics and the log distance from the predicted area to the real farms location calculated from 24,328 target farms in state B.

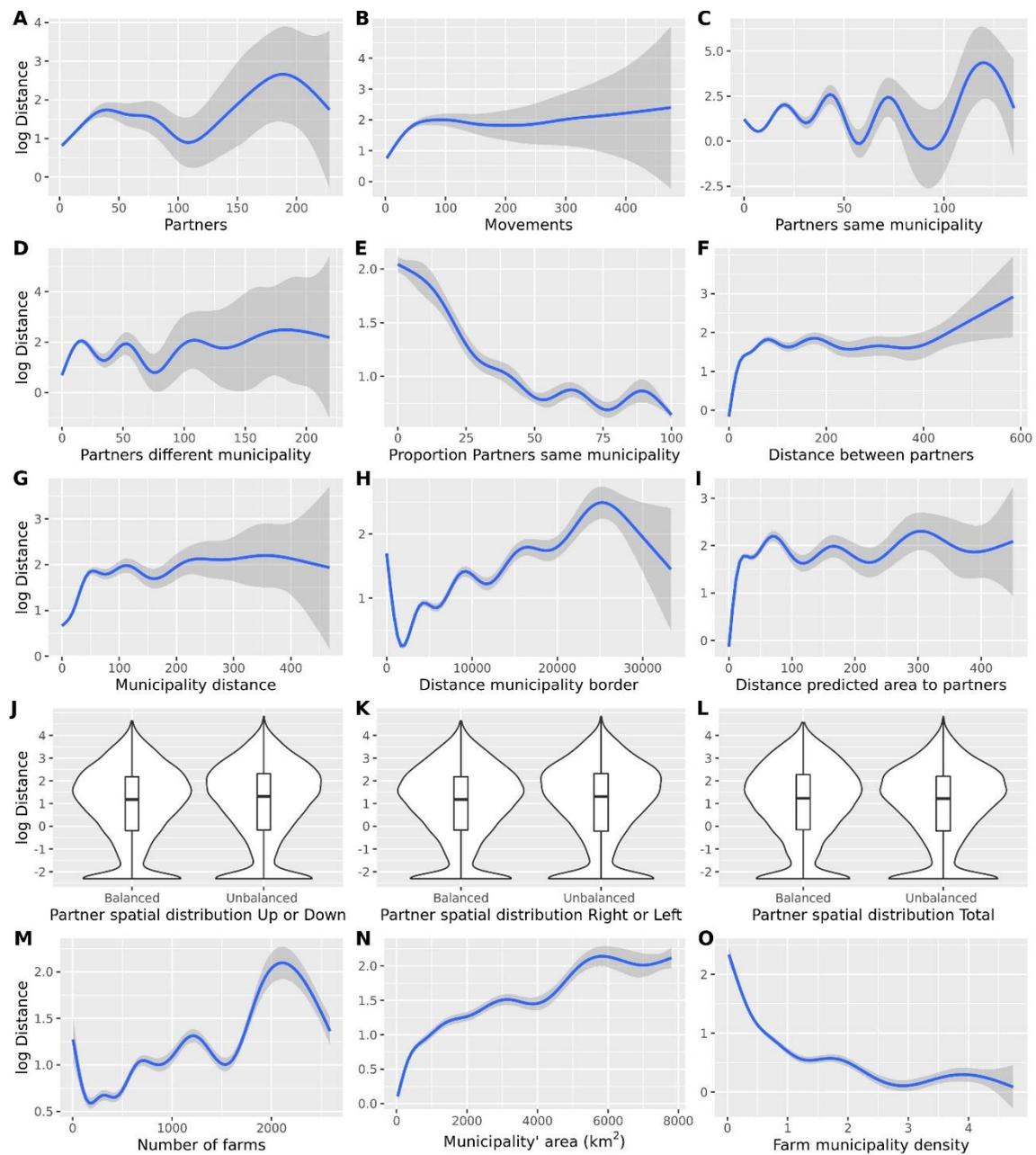
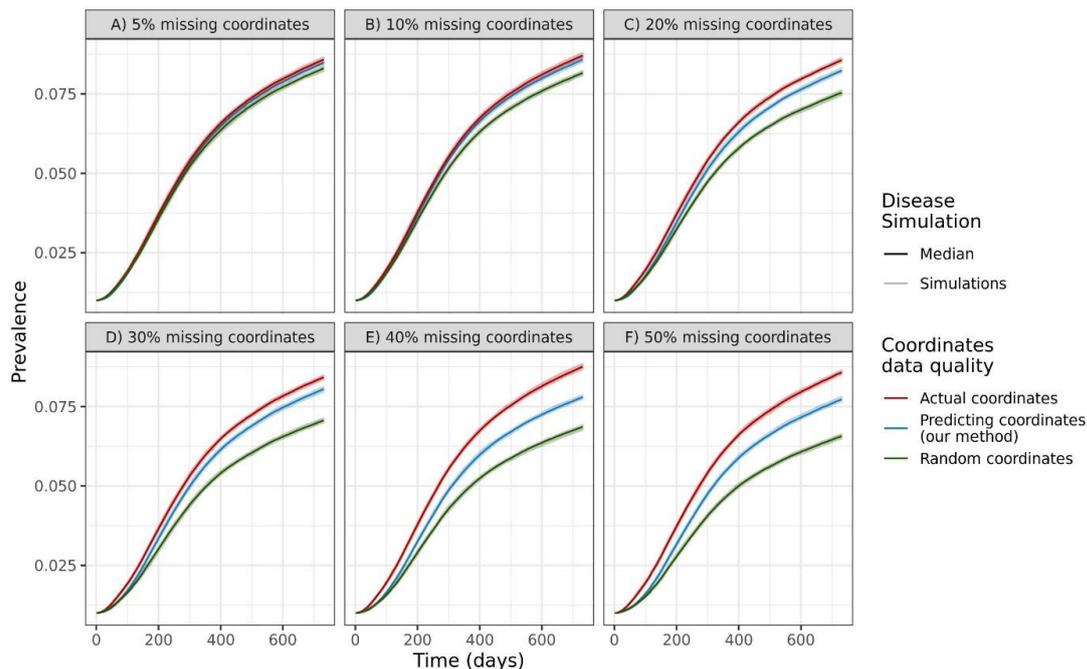


Figure 8. Disease spread simulation between 2016 and 2017 in the state B A) Scenario with 7,202 (5% of the farms active in the network and with good coordinate quality) farms coordinates removed and 4,334 predicted locations, B) Scenario with 14,403 (10%) farms coordinates removed and 8,275 predicted locations, C) Scenario with 28,807 (20%) farms coordinates removed and 15,194 predicted locations, D) Scenario with 43,310 (30%) farms coordinates removed, and 20,642 predicted locations, E) Scenario with 57,613 (40%) farms coordinates removed and 24,495 predicted locations, F) Scenario with 72,016 (50%) farms coordinates removed, and 26,320 predicted locations.



The correlations studied including the distance between the predicted and actual farm location and the municipality and network variables may help improve this methodology in the future. An unexpected finding from our study was that a higher number of partners did not result in the decrease of distance between the actual and predicted. We thought that a higher number of trade partners would increase the accuracy of the predictions. However, this was not the case in the present study. Despite these results, the partner spatial distribution and distance to the predicted area had better results. Thus, these variables seems to be more important to predict a farm location correctly.

The random sample of the coordinates within the correspondence municipalities underestimate the distance between farms in both states. This means that the distance between farms was higher with the random sample of coordinates than the distances between farms with the actual locations. As a result, the disease spread simulation had

lower prevalence with the random coordinates than the scenarios using the actual and predicted coordinates. We believe that this distance between farms with the random sample of coordinates may increase when applied to larger municipality areas. In contrast, the prevalence calculated with our methodology to predict farm locations was more similar to the one found with the actual coordinates. However, the gap between these prevalence distributions increase as there is an increase in the proportion of farms with missing coordinates. Additionally, the prevalence found with the predicted locations was lower than the one using actual locations. A similar result was obtained with the random methodology. Thus, the methodology to predict the farm location also underestimate the distance between farms.

We used a cutoff distance between farms of 1 km to spread a disease between these premises. However, for some contagious infectious diseases affecting livestock, this distance may be shorter or greater for a pathogen be transmitted between herds. In our study, we did not evaluate many distance cutoffs to transmit a disease due to the time processing data. We believe that the prevalence calculated using the random sample coordinates and the actual coordinates may be similar if a higher cutoff distance to transmit the disease between the farms is implemented.

This methodology can be improved or complemented with additional analysis to improve the prediction of the farm locations. In this study we used a two-year period of cattle transit to predict the locations. However, the use of a higher interval of time can improve the predictions by increasing the number of partners distributed around and near to a target farm. Also, the use of satellite imagery, for example, may help to identify actual coordinates of farms in the region as in studies conducted in the United Kingdom (Michael J. Tildesley & Ryan, 2012). Thus, it would be possible to change the cell grid used in our methodology for more realistic locations. An additional method may be to increase or decrease the probability from some cells so that a target farm can be located. This can done according to the farms production characteristics in the municipality, for example, within a municipality the dairy farms are clustered in one specific area (dairy zone), and the target farm we aim to predict is classified as dairy as well, then it makes sense that the target farm by its characteristics can be located in that particular dairy zone. Thus, increasing the cell weight located in one municipality region can improve our location prediction. Similarly, we may apply this idea to increase or decrease cells weight with other municipality characteristics, as environmental, for example, presence of deserts, lakes or cities may indicate a low probability of a farm location. Thus, decreasing the cells weight located in these areas may also improve the location prediction. As mentioned in this section, the

methodology herein described to predict farm locations has the potential to upgrade and become a useful tool to recover farm locations.

States with good quality spatial data may perform a number of exploratory analyses considering the farm spatial location and distance between these premises. As a result, these states would be able to act promptly and more efficiently to control infectious disease where the distance between farms plays an important role in the transmission (Boender, Meester, Gies, & De Jong, 2007). On the contrary, those states with poor spatial information may have problems to control an infectious disease outbreak. For example, the farm spatial information plays a major role in the control of an outbreak of foot-and-mouth disease (Gibbens et al., 2001). Thus, the states with poor spatial information would need to plan control programs without this information or delay the implementations of these control programs in order to collect the missing spatial information necessary to build these programs. Such delay may allow a pathogen to spread widely throughout one particular region. Finally, it would result in great economic losses to these states. The methodology described herein to predict farm locations may assist these states with poor spatial information quality to assess the potential spread of a disease in its territory. However, this methodology still has many limitations to date. Therefore, we recommend to evaluate the quality of the most updated information from the states and identify its deficiencies in order to be able to use the data available in the disease surveillance programs.

In conclusion, the methodology here described in the present study to predict a farm locations within a municipality which minimized the average geodesic distance to known contacts was found to be a significantly better predictor of farm location than random locations within the municipalities. However, other variables need to be considered in the methodology in order to improve farm location prediction since some of the predictions were not close from its actual location. Therefore, this methodology may be a good option in scenarios where states aim to develop a disease spread model considering the distance between farms. However, a set of its farms do not have good coordinates quality to be used.

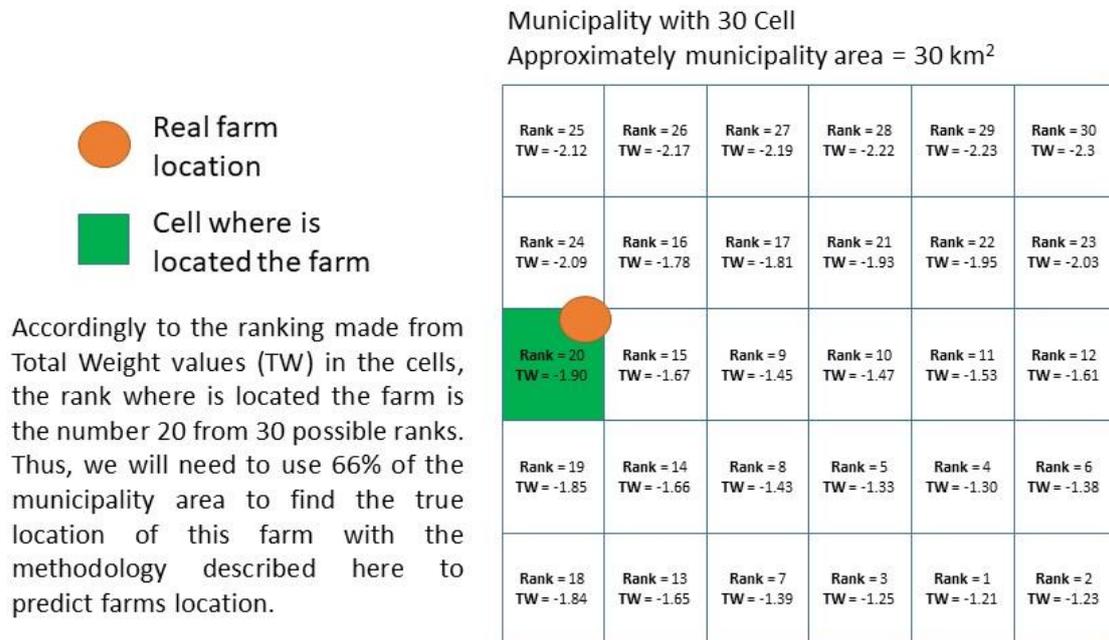
References

- Boender, G. J., Meester, R., Gies, E., & De Jong, M. C. M. (2007). The local threshold for geographical spread of infectious diseases between farms. *Preventive Veterinary Medicine*, *82*(1–2), 90–101. <https://doi.org/10.1016/j.prevetmed.2007.05.016>
- Ensoy, C., Faes, C., Welby, S., Van der Stede, Y., & Aerts, M. (2014). Exploring cattle movements in Belgium. *Preventive Veterinary Medicine*, *116*(1–2), 89–101. <https://doi.org/10.1016/j.prevetmed.2014.05.003>
- Feigl, K. L. (2002). 37 Estimating earthquake source parameters from geodetic measurements. *International Geophysics*, *81*(PART A), 607–621. [https://doi.org/10.1016/S0074-6142\(02\)80240-6](https://doi.org/10.1016/S0074-6142(02)80240-6)
- Gibbens, J. C., Sharpe, C. E., Wilesmith, J. W., Mansley, L. M., Michalopoulou, E., Ryan, J. B., & Hudson, M. (2001). Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *The Veterinary Record*, *149*(24), 729–743. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11808655>
- Gulenkin, V. M., Korennoy, F. I., Karaulov, A. K., & Dudnikov, S. A. (2011). Cartographical analysis of African swine fever outbreaks in the territory of the Russian Federation and computer modeling of the basic reproduction ratio. *Preventive Veterinary Medicine*, *102*(3), 167–174. <https://doi.org/10.1016/j.prevetmed.2011.07.004>
- Jacquot, M., Nomikou, K., Palmarini, M., Mertens, P., & Biek, R. (2017). Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1864). <https://doi.org/10.1098/rspb.2017.0919>
- Karney, C. F. F. (2013). Algorithms for geodesics. *Journal of Geodesy*, *87*(1), 43–55. <https://doi.org/10.1007/s00190-012-0578-z>
- Palisson, A., Courcoul, A., & Durand, B. (2016). Role of cattle movements in bovine Tuberculosis spread in France between 2005 and 2014. *PLoS ONE*, *11*(3), 1–19. <https://doi.org/10.1371/journal.pone.0152578>
- Robinson, S. E., & Christley, R. M. (2007). Exploring the role of auction markets in cattle movements within Great Britain. *Preventive Veterinary Medicine*, *81*(1-3 SPEC. ISS.), 21–37. <https://doi.org/10.1016/j.prevetmed.2007.04.011>
- Ssematimba, A., Boender, G. J., Bonney, P. J., Weaver, J. T., Malladi, S., Cardona, C. J., & Halvorson, D. A. (2018). Spatial transmission of H5N2 highly pathogenic avian influenza between Minnesota poultry premises during the 2015 outbreak. *Plos One*, *13*(9), e0204262. <https://doi.org/10.1371/journal.pone.0204262>
- Thrusfield, M. V. (2007). *Veterinary Epidemiology* (3rd editio). Hoboken: Wiley-Blackwell.
- Tildesley, M. J., House, T. A., Bruhn, M. C., Curry, R. J., O’Neil, M., Allpress, J. L. E., ... Keeling, M. J. (2010). Impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences*, *107*(3), 1041–1046. <https://doi.org/10.1073/pnas.0909047107>
- Tildesley, Michael J., & Ryan, S. J. (2012). Disease Prevention versus Data Privacy: Using Landcover Maps to Inform Spatial Epidemic Models. *PLoS Computational Biology*, *8*(11), 8–9. <https://doi.org/10.1371/journal.pcbi.1002723>

- Vernon, M. C. (2011). Demographics of cattle movements in the United Kingdom. *BMC Veterinary Research*, 7(1), 31. <https://doi.org/10.1186/1746-6148-7-31>
- Widgren, S., Bauer, P., & Engblom, S. (2016). *SimInf: a framework for data-driven stochastic disease spread simulations*. <https://www.CRAN.R-project.org/package=SimInf>.
- Widgren, S., Bauer, P., Eriksson, R., & Engblom, S. (2016). *SimInf: An R package for Data-driven Stochastic Disease Spread Simulations*. Retrieved from <http://arxiv.org/abs/1605.01421>.

Supplementary Material

Figure 1. Farms location prediction accuracy to estimate the percentage of municipality area needed to find the real farm location.



General conclusion

In this study we provided information and methodologies to control the Bovine Tuberculosis (bTB) in Espírito Santo (ES). It was found that ES had one the highest bTB prevalence in Brazil, and some farm production characteristic (Herd size and milking type) and metrics of the bovine trade network (All-Degree, Weighted In-Degree, Weighted Out-Degree, Betweenness, PageRank and Authority) were associated with this disease in the state. In addition, the southern region of ES had higher bTB prevalence, and areas with higher risk associated with the disease.

We identified that some networks metrics of the farms in ES (like Degree and Betweenness) may be useful to complement a target surveillance for diseases transmitted by the cattle trade. Finally, in the absence of farms spatial coordinates to build simulations models to spread diseases, we provided a methodology that can recover some of those missing farm locations. These results can provide additional information and tools which can help to the decision makers to complement animal disease surveillance programs in ES and Brazil.

General references

- Ferreira Neto, J. S., Silveira, G. B. da, Rosa, B. M., Gonçalves, V. S. P., Grisi-Filho, J. H. H., Amaku, M., ... Lage, A. P. (2016). Analysis of 15 years of the National Program for the Control and Eradication of Animal Brucellosis and Tuberculosis, Brazil. *Semina: Ciências Agrárias*, 37(5Supl2), 3385. <https://doi.org/10.5433/1679-0359.2016v37n5Supl2p3385>
- IBGE. (2017). Censo Agro 2017. Retrieved July 30, 2019, from https://censos.ibge.gov.br/agro/2017/templates/censo_agro/resultadosagro/pecuaria.html
- LAGE, A. P. ., ROXO, E. ., MÜLLER, E. ., POESTER, F. ., CAVALLÉRO, J. C. M. ., FERREIRA NETO, J. S. ., ... GONÇALVES, V. S. P. (2006). *PROGRAMA NACIONAL DE CONTROLE E ERRADICAÇÃO DA BRUCELOSE E DA TUBERCULOSE ANIMAL (PNCEBT)*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- LeBlanc, S. J., Lissemore, K. D., Kelton, D. F., Duffield, T. F., & Leslie, K. E. (2006). Major Advances in Disease Prevention in Dairy Cattle. *Journal of Dairy Science*, 89(4), 1267–1279. [https://doi.org/10.3168/jds.S0022-0302\(06\)72195-6](https://doi.org/10.3168/jds.S0022-0302(06)72195-6)