l'université
nantes
angers
le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

Oniris
École Nationale Vétérinaire, Agroalimentaire et de l'Alimentation Nantes Atlantique
Santé et alimentation au cœur de la vie

Thèse de Doctorat

Bhagat Lal DUTTA

Mémoire présenté en vue de l'obtention du
grade de Docteur d'Oniris - l'École Nationale Vétérinaire Agroalimentaire et de l'Alimentation Nantes-Atlantique
sous le label de L'Université Nantes Angers Le Mans

École doctorale : **Biologie-Santé (BS)**

Discipline : **Biomathématiques, Systèmes complexes**
Spécialité : **Analyse de réseau, Épidémiologie, Dynamique de population**
Unité de recherche : UMR1300 INRA ONIRIS BioEpAR

Soutenue le 28-11-2014

Modélisation spatio-temporelle de la propagation d'un agent pathogène dans une métapopulation bovine : application au virus de la diarrhée virale bovine (BVDV)

JURY

Rapporteurs :	**Ann LINDBERG**, Associate Professor, National Veterinary Institute (SVA), Uppsala, Sweden **Benoît DURAND**, Directeur de recherche, ANSES ENVA, Maisons-Alfort, France
Examinateurs :	**Renaud LANCELOT,** Chercheur, CIRAD UMR CMAEE, Montpellier, France **François BEAUDEAU**, Professeur, Oniris, Nantes, France
Directrice de Thèse :	**Pauline EZANNO**, Chargée de recherche HDR, INRA Oniris UMR1300 BioEpAR, Nantes, France
Co-Encadrante :	**Elisabeta VERGU**, Chargée de recherche, INRA UR341 MIA-Jouy, Jouy-en-Josas, France

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 General context

Domestic coexistence of human and animals is said to be as old as the beginning of the concept of a Creator, the Almighty and His divinity, if archaeological information based texts, like [Cauvin, 2000], are to be believed. '*Human civilisation started with ploughing crop fields using oxen and milking cows*', is no way an exaggeration. The root of agriculture and animal husbandry germinated with *protocultivation* and *proto-animal raising* practices in the Neolithic age. Origins of most of the domesticated crop and animal species that are still in production today can be traced back to the *Neolithic agricultural revolution* [Mazoyer and Roudart, 2006]. The trajectory the human race had been tracing for these thousands of years had been shaped with the help of domesticated animals, since oldest civilizations. Domestic animals became exchangeable *wealth*, the backbone of what we now call *economy*, for a substantial length of the human history and later a *product*.

Figure 1.1: Agriculture in ancient Egypt as depicted in a burial chamber (circa 1200 BCE). [Source: http://en.wikipedia.org/wiki/Agriculture]

Although human race discovered alternatives to animal labour, still today animal husbandry is an indispensable part of human society, directly or indirectly. We, humans, are still dependent on animals, at least for food and wear. Over the ages humans developed more and more specialised knowledge about the best possible practices for animal breeding and farming ensuring best returns in terms of productivity and profit. And more factors of 'economics' were getting involved in livestock farming practices.

In the present world, reduction in productivity and/or increasing death losses have severe socio-economic impact on livestock farming, at both industrial and individual levels. One of the major causes of such an impact is pathogenic infections of farm animals. These infections are broadly of two types, those which can infect and are transmissible between (*non-human* vertebrate) animals and humans (zoonoses)[Fig. 1.2], and those which infect only animals [World Health Organisation, 2014]. Zoonoses are often given very high attention as they may create directly visible havoc in the human world apart from the non-human victims. There are collaborative strategies partnered by world organisations to fight the zoonoses [FAO-OIE-WHO, 2010]. A few examples of zoonoses are avian influenza, ebola hemorrhagic fever, plague, salmonellosis, anthrax, West Nile virus infection [The Merck veterinary manual, 2014, Centers for Disease Control and Prevention (CDC), 2014]. Non-zoonotic animal diseases had traditionally been given lesser importance, unless there is a sudden widespread outbreak with large number of cases, creating significant (and immediately visible) impact on the farming industry and consequently on the economy (e.g., foot and mouth disease outbreak in the UK in 2001 and 2007 [Thompson et al., 2002, Knight-Jones and Rushton, 2013], bluetongue outbreak in France in 2006-2007 [Wilson and Mellor, 2009, Pioz et al., 2012], classical swine fever outbreak in Germany and the Netherlands in the late 1990s [Saatkamp et al., 2000, Edwards et al., 2000, Moennig et al., 2003]).

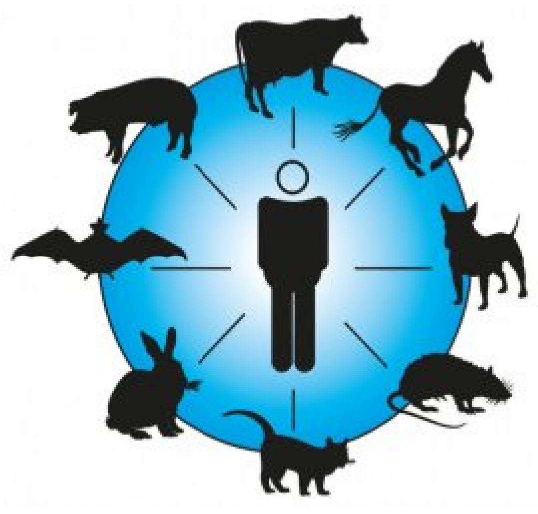


Figure 1.2: Zoonoses are transmissible between animals and humans.
[Source: http://www.publichealthagency.org/]

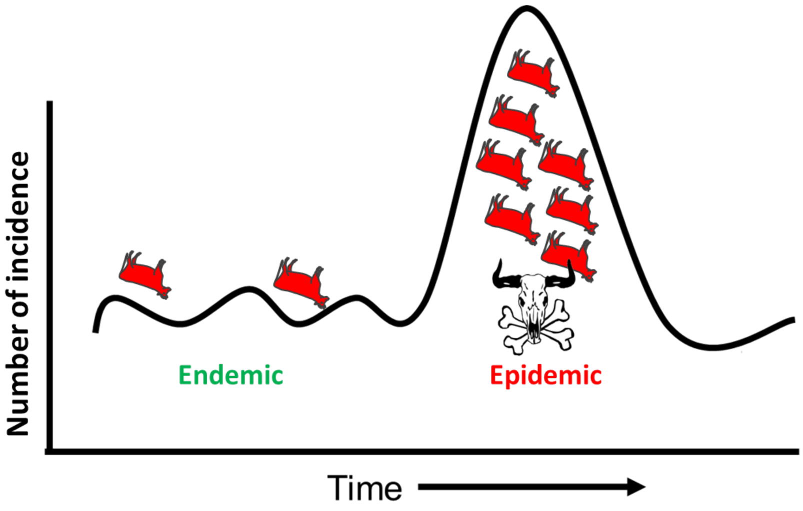


Figure 1.3: Epidemic and endemic situations

An *epidemic* is the condition of incidence of a disease in larger numbers than expected in a given area or in a specific population over a particular period of time. On the contrary, an *endemic* refers to the situation of constant presence (*usual prevalence*) of a disease or an infectious agent (with *normal* rate of occurrence) within a specified population or within an area [Fig. 1.3]. An epidemic is termed as *pandemic* when the pathogen spreads beyond local populations, possibly even to intercontinental scales [Centers for Disease Control and Prevention (CDC) et al., 2006]. In an another classification, from the point of view of geo-political policy making, livestock infectious diseases are classified into *endemic* and *exotic* classes [Carslake et al., 2010]. These definitions are political definitions, relevant to socio-economics, legislation and policy implementation at a country level depending on if a disease is present within its political boundaries (endemic) or not (exotic). Every emergence or re-emergence of exotic pathogens often leads to epidemic situations and an impulsive effort of human hours and money goes into action in control and (re)elimination of the pathogen with large media coverage. In contrast, endemic pathogens are often left low-lying to individual animal farmers, animal health practitioners and animal health organisations, without much attention, as if they lack significant impact. But the truth is that many endemic diseases maintain a stable *low profile* prevalence for very long periods, slowly and silently siphoning out the productivity and performance of farms [Bennett, 2003, Nieuwhof and Bishop, 2005]. Efforts of individual farmers willing to eradicate a particular endemic pathogen goes into vein, as (s)he is surrounded by a high prevalent environment of infected herds [Carslake et al., 2011]. In the long run the loss to the community is significant [Bennett et al., 1999, Bennett, 2003, Nieuwhof and Bishop, 2005], although it may not be readily visible to the public eyes unlike exotic epidemics.

1.2 Specific context and motivation

Among such *low profile silent killers*, bovine viral diarrhoea virus (BVDV) infections are endemic worldwide.

Bovine viral diarrhoea is caused by a Pestivirus belonging to the Flaviviridae family. BVDV infections result in reduced reproductive performance, reduced milk production, retarded growth and increased occurrence of other diseases in cattle [Lindberg, 2003]. BVDV infection is an enhancer of mortality among the young stock and early culling. Performance and economic sustainability of farms may be crippled to a high degree due to high prevalence of BVDV. The economic impact on the *dairy* industry ranges between $10-40 million per million calvings with annual incidence rate of 20-40%, variations attributed to herd structure and farming systems [Houe, 2003]. Estimated losses to *beef* herds range between $3000 to $9000 per 100 cows per year whereas the annual incidence of acute infections stands at nearly 34% [Heuer et al., 2008].

	Production losses in €/cow-year	Treatment expenditure in €/cow-year	Total costs of BVDV infection in €/1000 l of milk
Average case	67	8	10.7
Severe case	121	12	19.0

Table 1.1: Estimated economic consequences of BVDV infection in a dairy herd in the European context. (Source: [Fourichon et al., 2005])

Widespread persistent prevalence of endemic cattle diseases, like BVDV, in cattle farming regions may be a major concern for the sustainability and growth of farms. Their spread and persistence involve several scales: between individuals, between groups within a structured herd, between different herds in a locality to between herds not in geographical vicinity. Persistence of pathogens between connected populations has been linked to the ecological niche theory involving patches connected by exchange of individuals [Mollison and Levin, 1995, Moilanen and Hanski, 1998, Guégan et al., 2005, Ram et al., 2008], a population of populations distributed in space or a *metapopulation* (a metapopulation is a collection of subpopulations connected through migration/dispersal [Hanski, 1999]). In comparison to a single population, such spatially distributed populations may aid in species persistence or lead to extinction level depending on the situation (level of fragmentation, migration rates). Invasive situations may be a boon for ecologically threatened species, but it serves as a harmful factor when comes to pathogens, enhancing persistence and spread of pathogen populations [Grenfell and Harwood, 1997, Jesse et al., 2008].

Persistence of a species in a metapopulation via extinction-recolonization depends on the contact structure connecting the subpopulations. The contacts in a metapopulation can be identified into four types: *migration or dispersal* (member(s) of a subpopulation moving into another subpopulation with no return to the original population), *visits* (a member from a subpopulation temporarily moves into another subpopulation but returns back to home population soon), *neighbourhood* (no member from subpopulations physically leaves its home population but they have contacts, say, meeting at the boundaries or at some kind of common resource), and *indirect contacts* (the members do not physically meet each other, but exchange of pathogenic material takes place via an agent not belonging to the metapopulation, an animal (human or non-human) or an inert vector or through environmental routes (air/water/soil)) [Keeling and Rohani, 2008].

In the context of cattle farming, farms need a constant replacement of breeding cattle to maintain productivity and size. For this replacement, farmers have two alternatives, either raise heifers internally or buy from other farms. Owing to the *additional* cost and involvement required to grow heifers domestically, many farms may prefer the purchase option, exposing themselves to the risk of (re-)introducing pathogens into their herds (or spreading the pathogen to other herds by an infected herd) [Álvarez et al., 2011, Tinsley et al., 2012, Gates and Woolhouse, 2014]. Pathogens may also get introduced into farms via over the fence contacts at the boundaries between two farms, or at common water holes at pastures. Apart from the contact structure among different herds, the internal herd structure also plays an influential role in the within-population spread and persistence [Keeling and Rohani, 2008, Ezanno et al., 2012], especially for moderately spreading endemic pathogens like BVDV [Viet et al., 2004, Ezanno et al., 2007]. Therefore, a full assessment of the contributions of different paths is necessary to understand the endemic diseases, which in turn leads to the design and implementation of long term effective remedies to prevent, control and (possibly) eliminate pathogens from the metapopulation.

The control of BVDV spread has long been left to individual breeders. However, given the socio-economic importance, control programs are being implemented by farmer's organisations in several regions in France (and Europe) [Joly et al., 2005, Lindberg et al., 2006, Ståhl and Alenius, 2012, Gates et al., 2013]. The ex-ante evaluation of the effectiveness of the control strategies to resist BVDV spread (in terms of intra and inter-herd breeding contexts) is a major professional concern keeping in mind the diversity of cattle farms and diversity of contacts among them. In the prevention and control measures front, there are three scenarios available in the literature [Houe et al., 2006, Lindberg et al., 2006, Ståhl and Alenius, 2012, Graham et al., 2013, Lanyon and Reichel, 2014]: (a) BVDV spread without control (do nothing, self clearance), (b) vaccination strategy and (c) test and cull (removal of 'super shedders' identified by testing). A biosecurity and surveillance mechanism is also advised for effective implementation of control and eradication protocols [Lindberg et al., 2006]. Still not much progress has been seen to evaluate these control strategies in a regional level addressing the heterogeneous herd structure and animal exchange management.

1.3 Modelling

When field observation based studies of complex and interacting systems are hard to achieve owing various reasons (inherent complexity, logistic and economic issues or time constraints) often modelling comes as a rescue. Besides theory and experiment, 'modeling and simulation' has now been established as the third branch of science [Edelstein-Keshet, 1987, Murray, 2002, Boccara, 2004, Brauer and Castillo-Chavez, 2011].

A mathematical model is 'a representation of the essential aspects of an existing system (or a system to be constructed) which presents knowledge of that system in usable form' [Eykhoff, 1974]. Models have their own limitations, they cannot (and are not supposed to) represent the whole system. They usually have two distinct roles: understanding and prediction. What they can deliver, up to what extent, is dependent on the design of the model as well as number of inputs (and their accuracy). Predictive models demand high levels of accuracy whereas transparency is more demanded for a model for the purpose of understanding. A 'good' model is usually a trade-off between accuracy and transparency, but this 'goodness' is relative to the context and issue of interest [Diekmann and Heesterbeek, 2000, Keeling and Rohani, 2008, Ezanno et al., 2012]. In epidemiology, models are tools that allow translation between behaviours at various scales. They enable extrapolation from a known set of conditions to another. In other words, models can be used to predict the population-level epidemic dynamics from an individual-level knowledge of epidemiological factors, the long-term behaviour from the early invasion dynamics, or the impact of vaccination on the spread of infection [Keeling and Rohani, 2008]. More specifically, relevance of epidemiological models depends on assumptions of the population structure and dynamics, on the biology of the host-parasite infection, and a mathematical framework to address the problem [Viet et al., 2007, Diekmann et al., 2012].

Several theoretical studies exist on the dynamics of spread of BVDV in a cattle herd [Pasman et al., 1994, Sørensen et al., 1995, Innocent et al., 1997, Cherry et al., 1998, Gunn et al., 2004, Viet et al., 2004, Ezanno et al., 2007]. The importance of herd structure and between-group transmission was shown in [Viet et al., 2004, Ezanno et al., 2007]. [Ezanno et al., 2007] performed an extensive study to identify the key-parameters in a structured dairy herd. Similar work on beef herds was performed in [Damman et al., 2014]. Models are, in general, designed to study effects and efficacy of control measures. The scenario of between-herd contacts and its impact on the persistence of BVDV at a regional scale is not much addressed so far, compared to the intra-herd scenario. Very recently the issue is being addressed with some light on this part of the story by [Courcoul and Ezanno, 2010, Ersbøll et al., 2010, Tinsley et al., 2012] but much has still to be done. In an attempt to represent the spread of BVDV in a fully susceptible metapopulation of cattle herds [Courcoul and Ezanno, 2010] presented a small meta-population composed of 100 dairy herds. They considered random patterns in animal exchanges and uniform neighbourhood without explicit representation of space. [Ersbøll et al., 2010] focused on spatial neighbourhood relations, while [Tinsley et al., 2012] used a network based approach with simple *suscpetible-infected* type two state herd status. Explicit spatio-temporal model of BVDV spread in a regional scale is not known so far, except in [Ersbøll et al., 2010] where the risk of a dairy herd in terms of changing infection status (from not having persistently infected (PI) animals to having PI-animals) in relation to location and infection status of neighbouring cattle herds in the context of Denmark was addressed. Both [Courcoul and Ezanno, 2010] and [Tinsley et al., 2012] had no consideration of explicit regional scale variations in the farming system or contact structure. [Tinsley et al., 2012] and [Ersbøll et al., 2010] did not consider the importance of the complex within-herd dynamics.

Spread and persistence of BVDV have been shown to be dependent on the diversity in cattle breeding/farming systems and diversity of contacts between farms [Ezanno et al., 2006, Ezanno et al., 2007, Ezanno et al., 2008]. Cattle farming system (controlled population, exclusive grouping of animals based on age and physiological status) induces heterogeneity in contact structure among different groups within a herd. Presence of more than one type of transmission and infected animals shedding the virus at different rates along with heterogeneous contact pattern between the groups within a farm and animals between different farms add a high level of complexity to the transmission dynamics of BVDV at all levels. Again, the management of herd is different for beef and dairy herds. In dairy herds the structured grouping (based on age and physiological status) is quite rigid and retained throughout the year. In beef herds most of the age groups mix together, a structure being imposed mainly when bred females are with the bull, i.e., while grazing [Damman et al., 2014]. Therefore, the obvious interest of modelling BVDV is a mean of integrating synthetic hierarchical processes at all levels (individual animal, groups within a farm, all animals within a farm and between farms and types of farms).

1.4 Objectives and outline of the thesis

1.4.1 Objectives

The general objective of this thesis is to model the spread of BVDV, in a metapopulation of cattle at a regional scale relying on the management of between-herd animal movements and herd epidemiological statuses paving the way for the evaluation of control strategies. In the process to achieve the objectives, the thesis addresses three main issues: (i) understanding of the contact structure among herds underlying BVDV spread; (ii) development of a dynamic model of BVDV spread at inter-herd level; and (iii) investigation of BVDV regional dynamics with this model. The final goal is the usability of the model to assess prevention and control measures at a regional scale. The dynamical modelling of BVDV spread integrating several levels of population structures is a useful tool for testing hypotheses on propagation pathways between herds and predicting the spread at different spatiotemporal horizons. A clear understanding of the topological characteristics of the contact structure underlying pathogen spread and proxies for epidemic spread are always handy in designing control strategies in the event of an outbreak.

1.4.2 Methodology and approach

Methodologically the study is divided into two areas: (i) analysis of the spatio-temporal network of contacts between farms and (ii) modelling of propagation of the pathogen at intra- (local) and inter- (global) herd levels on a dynamical network of cattle movements. This involves implementation and adaptation of existing methodological approaches from network analysis and elaboration of multi-scale dynamic stochastic models, from an existing local dynamical model.

First, methods of network analysis were employed to quantify network parameters (e.g., degree distributions, clustering coefficients etc.) with respect to well characterized graph classes. The network was viewed both as a classical static and a dynamical network, where temporal signatures of movements shape the structure. Structural characteristics for both static and temporal networks were investigated and proxies for (nonspecific) epidemic spread on the network were evaluated.

Second, a stochastic model in discrete time has been developed describing the propagation of BVDV between farms in a region. The BVDV spread model between herds, involves three levels of dynamics: (a) the local population dynamics of a herd, (b) infection dynamics within a herd between grouped animals and (c) the interactions between herds. The main novelty of the model presented here lies in the fact that it accounts for heterogeneities which arise at the within-herd scale, such as herd size, locations etc. For movements among farms it relies on actual movement patterns from the cattle movement network. The neighbourhood considerations are also based on actual geo-locations of farms. Owing to the fact that the French cattle farming industry is computationally prohibitively large to be tackled in simulations, we concentrated our studies to the Finistère department of Brittany region of France, an important region both in terms of farming activity and cattle density. The model had been designed to simulate a metapopulation of dairy herds in this region but it is generic enough to be applied to any other specific context. There is an added advantage of focussing on this area, as reliable surveillance data for this region for a reasonable time period was obtained which might be handy to test implications of the model based on herd epidemiological statues, whereas such surveillance information for elsewhere was hardly accessible for this study.

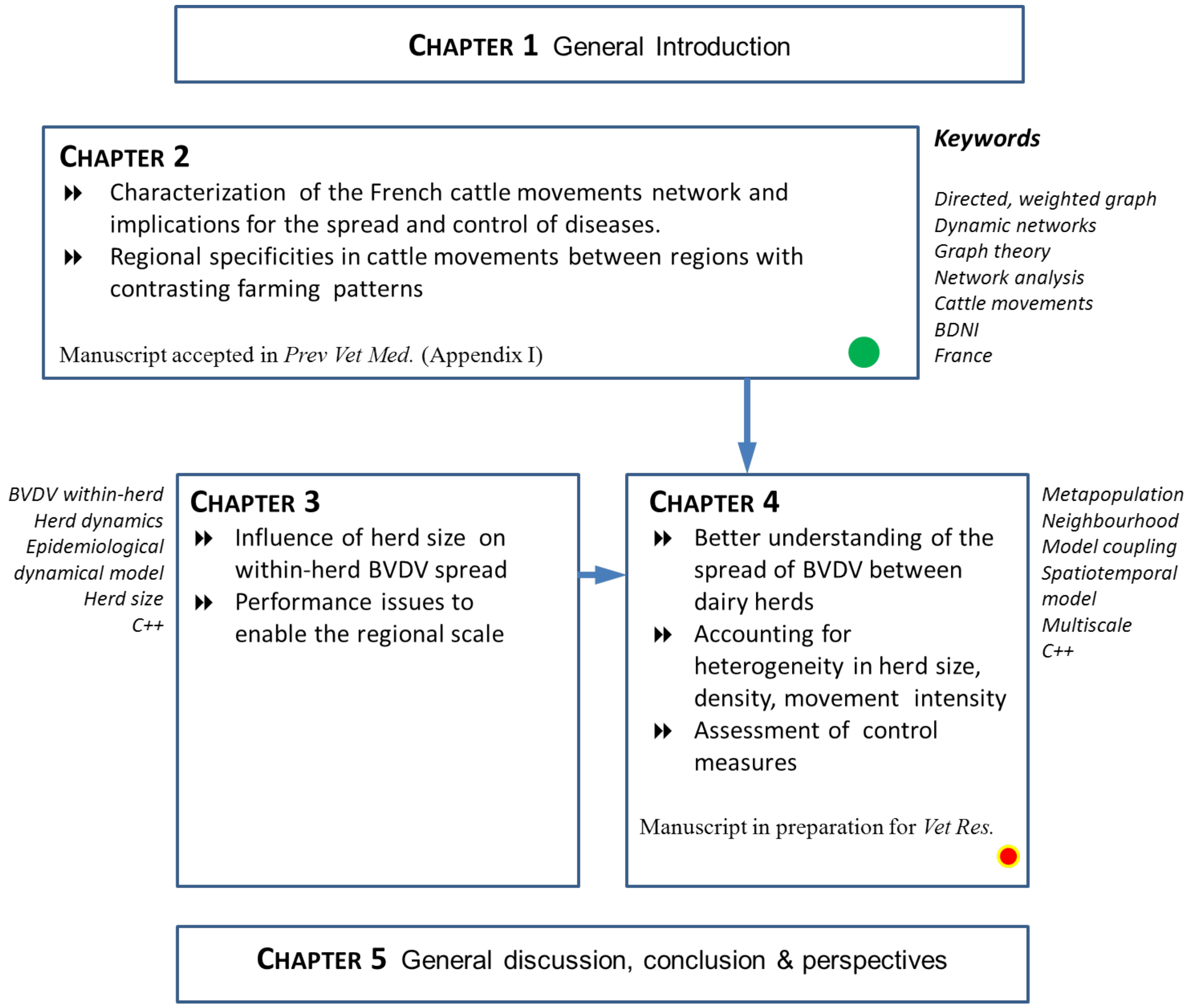


Figure 1.4: The layout of the thesis

1.5 Chapter-wise presentation of the thesis

The plan of presentation of this work is the following, a brief overview of which is provided in Fig. 1.4.

In this first chapter, we introduced the issues addressed in this thesis and their context. Here we highlight the long run impact of the endemic diseases in terms of their the socio-economic relevance. The context of this thesis work is to unwind the multi-scale complexity involved in transmission and persistence of endemic diseases, specifically BVDV, using mathematical modelling and simulations. Specific scientific and research issues are discussed in detail in the following chapters.

In Chapter 2, we study the network of cattle movements in France. We start with a brief literature review relevant network analysis, in general and specifically applied to livestock. We describe the data obtained from the French national bovine database and their filtering and then present a descriptive analysis of these data. Then, we view the animal exchange data as networks and analyse them using methods of social network analysis and graph theory. The network formalism is realised in two ways, firstly in a time-aggregation of movements to achieve a network with frozen links for the whole duration of observation, i.e. static network, and then in a time-stamped sequence of consecutive paths, i.e. dynamic or temporal network. Network descriptors relevant to cattle farming systems and cattle epidemiology are evaluated and their implications for the spread of infectious diseases through the network (especially for BVDV) are discussed based on the key results.

In Chapter 3, dynamics of BVDV within a herd is explored. Literatures show that within-herd BVDV dynamics is very much dependent on the farming methods/herd structure/management. Therefore, the dynamics need to be modelled differently for beef and dairy herds, as their management and farming are fundamentally different. We analyse BVDV spread in a well-structured dairy herd which throws light into the system as close as possible. Out of the different models present, we choose a model already developed in-house, an intra-herd model of BVDV spread in dairy herds [Ezanno et al., 2007], as we restrict our focus to dairy herds. This model was in good agreement with field observations. Optimisation of the model using C++ as the programming language for performance enhancement in terms of computational resources, assumptions added in or removed from the model or any other modifications during optimisation are presented. The outputs of the original model are compared with the ones of the optimised model. We explore the key advantages (along with the limits) and discuss the scopes of the intra-herd model in designing a region wide inter-herd model of BVDV spread.

Chapter 4 is exclusively about modelling the spread of BVDV among herds in contact, through neighbouring relationships or animal exchanges. This chapter starts with discussing what has already been done in this context, explore their pros and cons. Then, we present our model for inter-herd transmission of BVDV. We discuss the approach we take, assumptions, parameters and initial conditions chosen for the model and justify their choices. We also present the details of coding and implementation, how we incorporate data in framing the model. We define the main outputs we are interested in, and justify their choice. We present how we tackle heterogeneities in herd sizes, animal exchanges and neighbouring relationships. We present the results of simulations of the inter-herd BVDV spread model and explore different scenarios. Then, we discuss the key results and their implications.

Chapter 5 is a discussion of the overall findings of the study, the pros and cons with all possible practical issues. This is the concluding chapter of this thesis with notes for future goals.

Chapter 2

Cattle movements and the risks of epidemic spread

In this chapter, we study trade cattle movements in France at different spatial and temporal scales. Starting with a review of the literature relevant to this work, we then present the data obtained from the French national bovine database. Data cleaning process and a descriptive analysis of the data are presented. Then the animal movement data is represented as networks at different levels of spatial and temporal granularities. Cattle movement networks are analysed using methods of social network analysis and graph theory. Two formalisms of network construction were used, (i) static networks, where the time aggregated movements were considered to be frozen for the whole duration of observation and (ii) temporal or dynamic formalism, where time-stamped sequence of consecutive movements constituted the paths. We evaluate the network descriptors relevant to cattle farming systems and cattle epidemiology and their implications for the spread of infectious diseases through the network (especially for BVDV) are discussed based on the key results.

Most of the analyses included in this chapter are the subject of a paper accepted for publication in Preventive Veterinary Medicine (Appendix I).

2.1 General context, state of the art and objectives

Trade contacts among animal holdings are important in the spread of infectious diseases. These contacts pose a significant risk with possible introduction of infected animals into naive herds or coming to close contact to infectious individuals during transportations. Network analysis provides analytical framework to study the trade relationships between the farms and sheds light on the patterns and implications of these relationships exploring the underlying topological features. It can be used to assess the extent to which an epidemic process can propagate through the links and to identify nodes to be used for targeted removal or vaccination as a measure to minimize the spread. Livestock movement datasets can be represented as networks, with farms as nodes and trade relations as directed links (flows of traded animals are most often not symmetrical). Animal trade links, along with directionality, possess weights, the number of animals or batches traded. Moreover, these links are time-varying too. Two farms exchange animals at specific time points, not necessarily continuously at all times, therefore, the links are not permanent. In other words, both network topology and the flow on the network are time dependent.

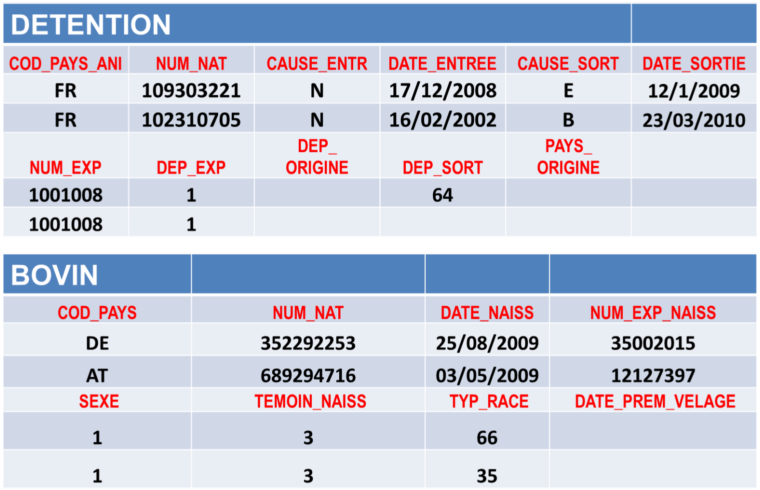
The practice of representing livestock movement datasets as networks and their exploration [Keeling and Eames, 2005, Kao et al., 2006, Kiss et al., 2006, Danon et al., 2011, Büttner et al., 2014] using the methodologies of graph theory and social network analysis [Wasserman, 1994, Dubé et al., 2011, Martínez-López et al., 2009, Newman, 2010] had been around for quite sometime. These analyses provide insights for controlling pathogen spread among livestock herds. Regarding the dynamical characteristics of animal trade data, there are at least two distinct primary approaches to address movement networks: (i) as static networks, by considering temporal aggregated variants of the initial datasets [Kiss et al., 2006, Kao et al., 2006, Rautureau et al., 2011] or (ii) as continuous increments of small width time windows [Vernon and Keeling, 2009, Büttner et al., 2013, Bajardi et al., 2011]. Most of the recent studies use a mix of both approaches. The studies in this field are of very wide range, from investigating demographic structure and pathogen dynamics [Kao et al., 2006, Natale et al., 2009] to analysis of dynamical patterns in longitudinal data [Bajardi et al., 2011] and surveillance optimization [Bajardi et al., 2012] to spatial structures revealing regional risks [Ensoy et al., 2014]. Different animal markets in several countries were explored: cattle in Denmark, Canada, France, Italy, Sweden, UK and Belgium [Bigras-Poulin et al., 2006, Dubé et al., 2010, Rautureau et al., 2011, Bajardi et al., 2011, Nöremark et al., 2011, Frössling et al., 2012, Vernon, 2011,Ensoy et al., 2014], sheep in UK [Kiss et al., 2006,Kao et al., 2006] and pigs in Sweden, France, Germany and Canada [Nöremark et al., 2011,Rautureau et al., 2012,Büttner et al., 2014,Dorjee et al., 2013]. These works investigated various characteristics relating networks representing animal trade and pathogen spread, such as vulnerability to and effects of infectious diseases based on animal exchange data. The network of movements of cattle in France in 2005 was studied to identify its vulnerability to the spread of pathogens [Rautureau et al., 2011]. Demographic patterns in the movement of cattle in the UK were explored and effects of the foot and mouth disease outbreaks in 2001 and 2007 on the flow of animals were detected using data across the years 1999-2009 [Vernon, 2011]. [Buhnerkempe et al., 2013] studied the national network of cattle movements in the US using veterinary inspection data at the state borders and showed that a county level approach is the most appropriate for examining processes influenced by cattle shipments, such as economic analyses and disease outbreaks. The UK cattle movement network was analysed to identify the indicators valid both in dynamic and static network approximations [Vernon and Keeling, 2009]. The longitudinal dataset of cattle movements in Italy for 2007 was studied to identify dynamical patterns at various time scales, from one day to the whole year [Bajardi et al., 2011]. The French cattle movements in 2005 were also studied at monthly and weekly time scales [Rautureau et al., 2011]. Cattle and pig movements in Sweden between 2006 and 2008 were investigated as monthly and yearly networks [Nöremark et al., 2011] and a measure for assessing disease control strategies, accounting for temporally compatible sequences of movements, was proposed. Application of network analysis parameters for risk based surveillance of bovine infections based on temporal network formalism had been demonstrated in [Frössling et al., 2012] for Swedish cattle trade data. The Danish cattle movements were studied for a period of 10 years (2000-2009): the (static) network characteristics were calculated at monthly time-aggregation level and their evolution followed over the whole period [Mweu et al., 2013]. The nodes of a time-varying network relevant from the point of view of pathogen spread were identified and the approach was tested on the German database on pig trade [Konschake et al., 2013]. The literature on network analysis had so far mainly been about analyses with the static representation of networks [Newman et al., 2006,Newman, 2010]. Recently, there have been growing interests on temporal networks across divergent fields of research [Blonder et al., 2012,Holme and Saramäki, 2012]. Researchers had been developing tools to analyse temporal movement network data and a few had already been available in public software repositories [Blonder and Dornhaus, 2011, Nöremark and Widgren, 2014, R Core Team, 2014].

In the context of time-varying networks, a question of particular interest concerns the spatial and temporal granularities. Is it possible to simplify the network of animal movements, in terms of number of nodes and links (by considering broader boundaries for node definition) and temporal dynamics (by considering aggregated snapshots), while preserving the information it encompasses? Another important aspect is related to the potential impact of farm type and animal breed on the topological and temporal characteristics of the network.

In this work, we analysed the French cattle movement network for the period from 2005 to 2009, from both spatial and temporal perspectives. Our objectives were, (i) providing an extensive description of this database for different spatial granularities (holdings and administrative communes as nodes) and temporal windows (from monthly to yearly); (ii) investigating the stability of network characteristics over consecutive and increasing time windows; (iii) computing proxies for the outcome of disease spread on the network representing these data, when accounting for its time-varying properties; and (iv) exploring in what extent (i)-(iii) lead to different interpretations when considering subsets of data based on the main characteristics of farms and animal breeds. Our consideration of node aggregation at different spatial and temporal scales was to capture possible constraints set by econo-politico-geographic layers and to identify their implications in disease spread and design of emergency interventions.

2.2 Data

In France, the Ministry of Agriculture maintains the French Database of Cattle Movements (FDCM), which records the life history of every animal from birth to death including movements between *holdings* (i.e. farms, markets, and assembling centres or whole-sellers), and to slaughterhouses, as well as imports and exports. For the present study, we used the FDCM database from 2005 to 2009 to construct the network of movements. The raw data set had five layers of information. The first layer contained the detention record in various holdings and had 11 fields of information identifying country code of the animal, national animal identification number, holding identification number, cause of entry into the holding (birth, purchase), date of entry, cause of exit from the holding (death, sale), date of exit, administrative department information of the holding, department of origin, destination department and country of origin (if imported). The second layer stated the identity



DETENTION					
COD_PAYS_ANI	NUM_NAT	CAUSE_ENTR	DATE_ENTREE	CAUSE_SORT	DATE_SORTIE
FR	109303221	N	17/12/2008	E	12/1/2009
FR	102310705	N	16/02/2002	B	23/03/2010
NUM_EXP	DEP_EXP	DEP_ ORIGINE	DEP_SORT	PAYS_ ORIGINE	
1001008	1		64		
1001008	1				

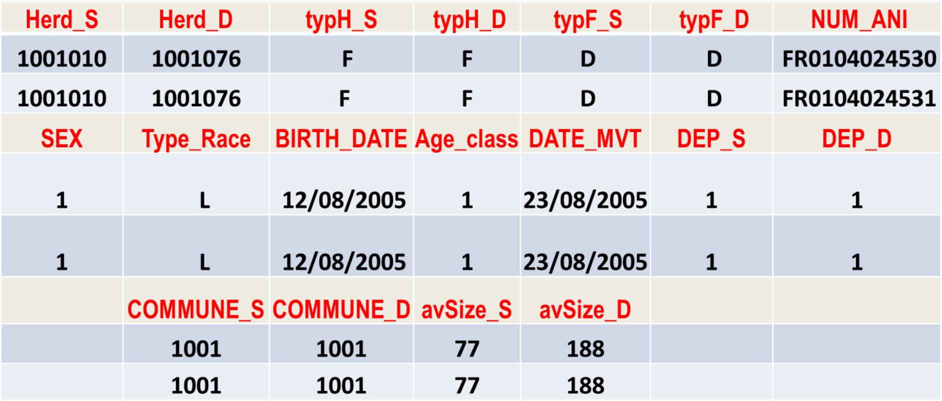
BOVIN			
COD_PAYS	NUM_NAT	DATE_NAISS	NUM_EXP_NAISS
DE	352292253	25/08/2009	35002015
AT	689294716	03/05/2009	12127397
SEXE	TEMOIN_NAISS	TYP_RACE	DATE_PREM_VELAGE
1	3	66	
1	3	35	

Figure 2.1: Examples of the first (detention) and second (bovine identity) layers of *raw* French Database of Cattle Movements (FDCM) data.

of a bovine with country code, national identification number, date of birth, place (farm) of birth, sex, race and date of first calving (for females). The other three layers contained information about markets, assembling centres and slaughterhouses. Out of the five layers, we had not included the information about the slaughterhouses in this study, as in the network they act only as *sinks* [Dubé et al., 2009] and do not play a direct role in the potential spread of pathogens.

A series of cleaning processes were performed on this set of data, based on matching of entries among the layers to obtain a filtered data set comprising the information about origin and destination (holdings, communes and departments of the holdings, and types of holdings), sex, race, date of birth and date of movements of animals identified by their national animal identification number. A dictionary for races and breeds was used to classify different breeds and races into three classes, dairy, beef or crossed (mixed). Each line in this dataset represents one movement made at a particular date. In the final form of the dataset, we do not consider the imports and exports and restrict our study to the movements within France only. We used SAS 9.2 software in the extraction and cleaning process of the FDCM.

Based on the cleaned dataset, several classifications required for further analysis were made. At the animal level, five age classes were defined: class 1 ($0-30$ days), class 2 ($1-8$ months), class 3 ($8-20$ months), class 4 ($20-30$ months) and class 5 (> 30 months), based on practices of cattle exchange among farms [Ezanno et al., 2006]. Average herd sizes were estimated on an



Herd_S	Herd_D	typH_S	typH_D	typF_S	typF_D	NUM_ANI
1001010	1001076	F	F	D	D	FR0104024530
1001010	1001076	F	F	D	D	FR0104024531
SEX	Type_Race	BIRTH_DATE	Age_class	DATE_MVT	DEP_S	DEP_D
1	L	12/08/2005	1	23/08/2005	1	1
1	L	12/08/2005	1	23/08/2005	1	1
	COMMUNE_S	COMMUNE_D	avSize_S	avSize_D		
	1001	1001	77	188		
	1001	1001	77	188		

Figure 2.2: Example of processed data used to construct the French Cattle Movement Network.

yearly scale based on detention records and in-out movements over a calender year. Herd types were defined on a yearly basis, using average number of animals per gender, racial type (beef, dairy, crossed/mixed) and age class present in farms. A farm was labelled as *dairy (or beef)* if it had more than 15 dairy (or beef) animals in age class 5. A minimum of 15 adult animals of a particular type was considered to be a *production unit* of that type [Ezanno et al., 2006]. If it had more than one type of production unit then it was considered *mixed*. Markets and assembling centres had similar properties (e.g. no internal herd structure for breeds or age classes, short stay periods etc.) and therefore were considered as same class for herd-type-wise analyses. Very small farms ($<$ 15 adult animals), insemination centres etc. were grouped into an another class, 'others'. Fig. 2.2 shows a screen-shot of the processed data. In this study, cattle movements were analysed at two spatial (holdings and administrative communes as entities) and three temporal (monthly=4 weeks, quarterly=13 weeks and yearly=52 consecutive weeks) scales of accumulation.

2.3 Network representation of livestock movement data

The movement data extracted from the FDCM and preprocessed contains information of animal exchanges between holdings underlying a network of contacts. Using the network terminology [Dubé et al., 2011,Martínez-López et al., 2009,Newman, 2010], holdings become *nodes* and animals moving from one holding to another connect these nodes through *links*. These links are directed as a consequence of non-symmetrical trade flows. A link can also be *weighted*, where the weight is defined either as the number of transactions (batches) carried out or as the number of movements (animals) exchanged between the source and the destination holdings during the time window under consideration. A node is said *active* with reference to a particular time window if it had at least one *in* or *out* link (i.e., it exchanged at least one animal) during the period considered.

In addition to networks including all active nodes and movements in a given time window, five subnetworks based on animal breed (*beef*, *dairy* and *mixed*) on one hand, and on herd type and animal breed simultaneously (*BEEF* and *DAIRY*), on the other hand, were analysed. A *breed-wise* subnetwork consisted of all the movements of animals of a particular breed, irrespective of the types of source and destination holdings. A *herd-wise* network includes all the movements of animals of a particular breed (e.g. dairy) having as source and destination herds of the same type (e.g. dairy), and markets or assembling centres.

We took two approaches in construction and analyses of livestock movement networks: a classical approach, where links were considered to be constantly present during a given period of observation and a temporal approach, where links were active only at specific time points [Newman, 2010, Blonder et al., 2012, Holme and Saramäki, 2012]. For the static approach, a variety of indicators (means and distributions) were calculated based on holdings or communes as nodes, and for monthly and yearly periods. The temporal analysis of the networks concerned the study of the stability of connexions over time, the influence of time aggregation window on network's main features and their evolution over increasing time snapshots. Proxies for pathogen spread on networks were computed from a static (for quarterly and yearly aggregated networks) perspective for the national networks and subnetworks. The dynamical view on the risk of propagation was implemented for two regional networks (French Brittany - a dairy region, and Limousin - a beef region). All the indicators used are defined in the following subsections.

2.3.1 Static networks: classical descriptors for directed weighted networks and proxies for pathogen spread

A network (or graph) $G = (V, E)$ is an ordered pair of the set of nodes (or vertices) V and the set of links (or edges, if undirected) E. The number of nodes $|V|$ is called the *order* of the network and the number of links $|E|$ represents its *size*. A *directed weighted network* for cattle movements can be represented by the matrix, $W = \{w_{ij}\}$, where w_{ij} is the number of movements from holding i to holding j (0, if no movement) and analysed based on classical indicators of networks adapted to account for directed and weighted links, when these variants were proposed in the literature [Newman, 2002, Barrat et al., 2004, Barthélemy et al., 2005, Fagiolo, 2007, Martínez-López et al., 2009, Opsahl and Panzarasa, 2009, Foster et al., 2010, Squartini et al., 2012]. We have the unweighted counterpart of W, to which we associate the *adjacency matrix* $A = \{a_{ij}\}$, with $a_{ij} = 1$ if there is at least one movement from i to j, and 0 otherwise.

Centrality measures allow ranking the nodes according to their '*importance*' in a network. Degree and strengths are important measures of centrality. *Degree* (k) is the number of holdings a holding is connected to. *Out-degree* (k^{out}) of a holding is the number of holdings to which at least one cattle is sent and *in-degree* (k^{in}) represents the number of holdings from which at least one cattle is received

$$k_i^{out} = \sum_j a_{ij};\ k_i^{in} = \sum_j a_{ji};\ \text{and } k_i = k_i^{in} + k_i^{out}.$$

Degree correlation is measured as Spearman's rank correlation between out-degrees and in-degrees of nodes [here, between selling and purchasing abilities of holdings, in terms of numbers of sellers (from whom an animal is bought) and buyers (to whom an animal is sold)].

Strength (s) is the number of movements per holding (animals sent, *out-strength* and received, *in-strength*)

$$s_i^{out} = \sum_j w_{ij};\ s_i^{in} = \sum_j w_{ji};\ \text{and } s_i = s_i^{in} + s_i^{out}.$$

Average nearest neighbour degree measures the average of the degrees of all the nodes connected to a node. It is related to the correlation between the degrees of linked nodes For weighted networks it is defined as

$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_j w_{ij} k_j\ .$$

The unweighted counterpart is

$$k_{nn,i} = \frac{1}{k_i} \sum_j a_{ij} k_j\ .$$

A *path* is a source destination trajectory without travelling the same vertex more than once. The path between two vertices with minimum number of intermediate vertices (hence links too) in between is called the *shortest path* (λ_{ij} represents its length) between them. For weighted graphs, it is the path for which the sum of the weights of the constituent edges is minimum.

Betweenness centrality (C_B) of a node measures the number of shortest paths passing through it, out of all the shortest paths from all vertices to all others. Formally,

$$C_B(i) = \sum_{j \neq i \neq l} \sigma_{jl}(i)/\sigma_{jl} \ ,$$

where $\sigma_{jl}(i)$ is the number of shortest paths between j and l passing through i, and σ_{jl} is their total number. *Closeness centrality* (C_{Cl}) of a node is the measure of of its ability to connect to all other nodes in fewer steps.

$$C_{Cl}(i) = \sum_j \frac{1}{\lambda_{ij}} \ ,$$

where λ_{ij}, the shortest path, is equal to infinity if there is no path between i and j.

Several indicators represent *network cohesion* in terms of local density, connectivity, assortativity and mixing.

The *average weighted clustering coefficient,* ($C^w(k)$), measures the average cohesiveness based on animal exchange intensities around holdings having k neighbours, where, for holding i,

$$c^{w,i} = \frac{1}{s_i(k_i - 1)} \sum_{j,h} a_{ij} a_{ih} a_{jh} \frac{(w_{ij} + w_{ih})}{2} \ .$$

$C(k) = \frac{1}{k_i(k_i-1)} \sum_{j,h} a_{ij} a_{ih} a_{jh}$ is its unweighted counterpart. The largest shortest path among all the shortest paths is called the *diameter* of the network. In the cattle movement network context, it represents the maximum lowest number of intermediaries between two holdings over all the indirectly connected pairs of holdings. The *giant strongly connected component (GSCC)* is the largest subnetwork where for every pair of nodes i and j, there is a directed path from i to j and a directed path from j to i. Here, it represents the set of holdings that an animal can enter from any node of this set. *Assortativity* is calculated as the Pearson correlation coefficient between linked nodes degrees. A positive correlation indicates the preference of being connected to vertices of similar degrees, whereas a negative correlation indicates preference of being connected to vertices of different degrees (*disassortative*). Assortativity close to zero implies a random distribution of the degree of the neighbours. The assortativity coefficient is formalised as [Newman, 2002],

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2}$$

where $q_k = \frac{(k+1)p_{k+1}}{\sum_j jp_j}$ is the normalised distribution of the remaining degree (the edges leaving the node, other than the one that connects the pair with p_k is the degree distribution). e_{jk} is the joint probability distribution of the remaining degrees of the two vertices and $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k kq_k]^2$ is the variance of the distribution q_k, a scaling term [Newman, 2002]. For undirected networks, e_{jk} is symmetrical and $\sum_{jk} e_{jk} = 1$ and $\sum_j e_{jk} = q_k$. For directed graphs, this coefficient can have four possible flavours, $r(\alpha, \beta)$, with $\alpha, \beta \in \{out, in\}$ and

$$r(\alpha,\beta) = \frac{\sum_i (j_i^\alpha - \bar{j^\alpha})(k_i^\beta - \bar{k^\beta})}{\sqrt{\sum_i (j_i^\alpha - \bar{j^\alpha})^2}\sqrt{\sum_i (k_i^\beta - \bar{k^\beta})^2}} \ ,$$

where for a given link i, j_i^α is the α degree of the source node of the link, and k_i^β is the β degree of the destination node, $\bar{j^\alpha}$ and $\bar{k^\beta}$ are the link averages of α-degree of sources and β-degree of destinations, respectively [Foster et al., 2010]. We calculated $r(out, out)$, the tendency of the nodes to connect with other nodes that have similar out-degrees, to have a feeling of the 'spreading abilities' of the neighbours.

Reciprocity is the ratio of total reciprocated weight (number of animal movements for reciprocated links between holdings) to the total weight [Squartini et al., 2012] (total number of animals exchanged) of the network, given by,

$$\frac{\sum_i \sum_{j \neq i} (w_{ij} + w_{ji}) a_{ij} a_{ji}}{\sum_i \sum_{j \neq i} w_{ij}}.$$

For unweighted directed graphs, reciprocity is the fraction of movements in any direction for which there exist movements in the opposite direction, $\frac{\sum_i \sum_{j \neq i} a_{ij} a_{ji}}{\sum_i \sum_j a_{ij}}$.

The *percolation* is a concept allowing to investigate the effect of node (or link) removal on the existence and the size of the GSCC, quantified in terms of the ratio between GSCC sizes after and before removal. In the context of pathogens spreading through cattle movements, percolation allows evaluating the efficacy of targeted control strategies based on the removal of holdings and of their connections (e.g. direct removal through ban on livestock movements, or indirect action through vaccination). Removal of nodes from a node may be either random or based on network properties. Targeted removal of nodes may be in terms of the centralities of the nodes (degree, strength or betweenness). Percolation study under different modes of node removal may be used as a proxy for the underlying capacity of the network favouring pathogen spread [Kiss et al., 2006].

2.3.2 Time varying characteristics of networks

For time-varying networks, connectivity between the nodes and distributions of network's characteristic measures evolve with time. In such networks, links do not persist over the whole time period, and the measures defined for connectivity driven 'static networks' may be insufficient to describe the evolution of the network in time. We need specific descriptors defined for networks with dynamical (time varying) character of the links (at different time windows).

Activity x_j of a node j is defined as the ratio between the number of interactions the node performs in a particular time window and the total number of interactions performed by all nodes during the same time window [Perra et al., 2012]. Network models driven by activity had been proposed to describe the instantaneous and fluctuating dynamics of time-varying networks [Perra et al., 2012, Karsai et al., 2014]. For cattle movements, the activity of a holding is the number of animals transacted during a time interval divided by the total number of transactions taking place during the same interval over the whole network. It can also be defined in terms of the number of batches transacted if the network links had been defined in terms of batches, but in the present study we have used the previous definition, if not specifically mentioned otherwise.

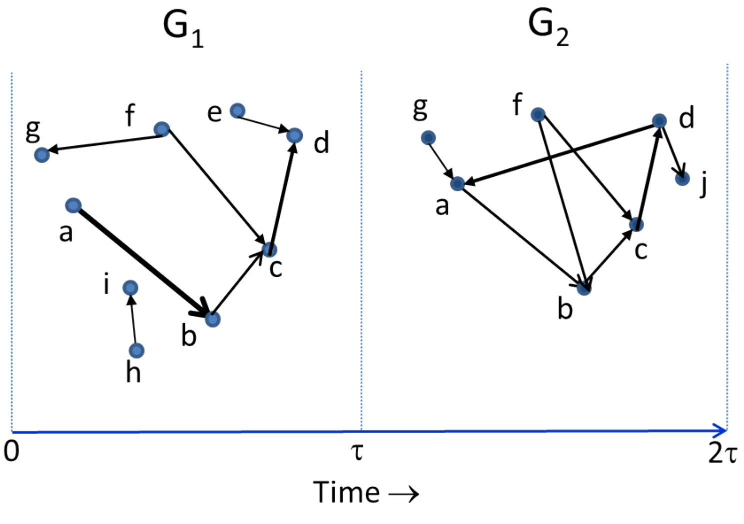


Figure 2.3: Two time-aggregated networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ at consecutive time windows of width τ. G_1 is on $(0, \tau)$; G_2 on $(\tau, 2\tau)$.

Convergence of the rescaled distributions onto a single curve is a signature of presence some stationary property of the system being time aggregated. The convergence of the rescaled distributions of network descriptors are measured by L^2 *distances* [Krings et al., 2012]. This distance is defined as the distance between a given rescaled distribution of a descriptor for networks aggregated over an interval of length Δt and networks aggregated over twice longer interval $2\Delta t$. For binned distributions with same number of bins n, $L^2 = \frac{\sum(x_i - y_i)^2}{n}$, where x and y are distributions for accumulation times Δt and $2\Delta t$ respectively.

Activation and deactivation of links and nodes over continuous windows were measured calculating the parameter *short range similarity* (SRS), which is defined as the average fraction of links (or nodes) common to consecutive aggregation windows of the same time intervals. We calculate the *similarity* in links between two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ [Fig. 2.3] at two consecutive time intervals $0 - \tau$ and $\tau - 2\tau$ as $\xi^l_{12} = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|}$ [Krings et al., 2012]. Similarly, the node similarity is: $\xi^n_{12} = \frac{|V_1 \cap V_2|}{|V_1 \cup V_2|}$. $\xi = 1$ if the networks are identical, and $\xi = 0$ if they share no links (nodes). SRS is the average of all such consecutive $\xi_{i,j}$s, $SRS = \frac{\xi_{12} + \xi_{23} + \ldots + \xi_{n-1,n}}{n-1}$, where n subnetworks of equal windows of aggregation time (τ) are constructed from the full time-stamped network.

2.3.3 Temporal network : main indicators

The temporal network approach exploits the fact that edges for many networks are not continuously active [Fig. 2.4]. In dynamic networks, some of the edges are practically instantaneous (e.g., network of emails, phone calls, over the fence contacts of cattle in farms, etc.) and others are active for non-negligible periods of time [Holme and Saramäki, 2012] (proximity network of in-patients in hospitals, cattle in the pasture etc.). In this approach, the network itself is considered to be a dynamical system.

When the duration of interactions is negligible (instantaneous), a *contact process* can be defined by a set of triplets (i, j, t), where $(i, j) \in V$, the set of vertices and t is the time [Fig. 2.4]. A triplet of a contact is restricted not to occur twice by construction. The same system can can also be represented by a set of (directed) links E and for all $e \in E$, a non empty set of time time of contacts $T_e = \{t_1, t_2, ..., t_n\}$ [Holme and Saramäki, 2012]. Very often, to study the dynamical evolution of a graph, all the contacts happening at the same discrete time step are grouped into a single graph and the dynamics of the network is presented as an evolving set of such graph sequences to reveal the topological dynamics [Blonder et al., 2012]. In this situation, some of the methods we discussed

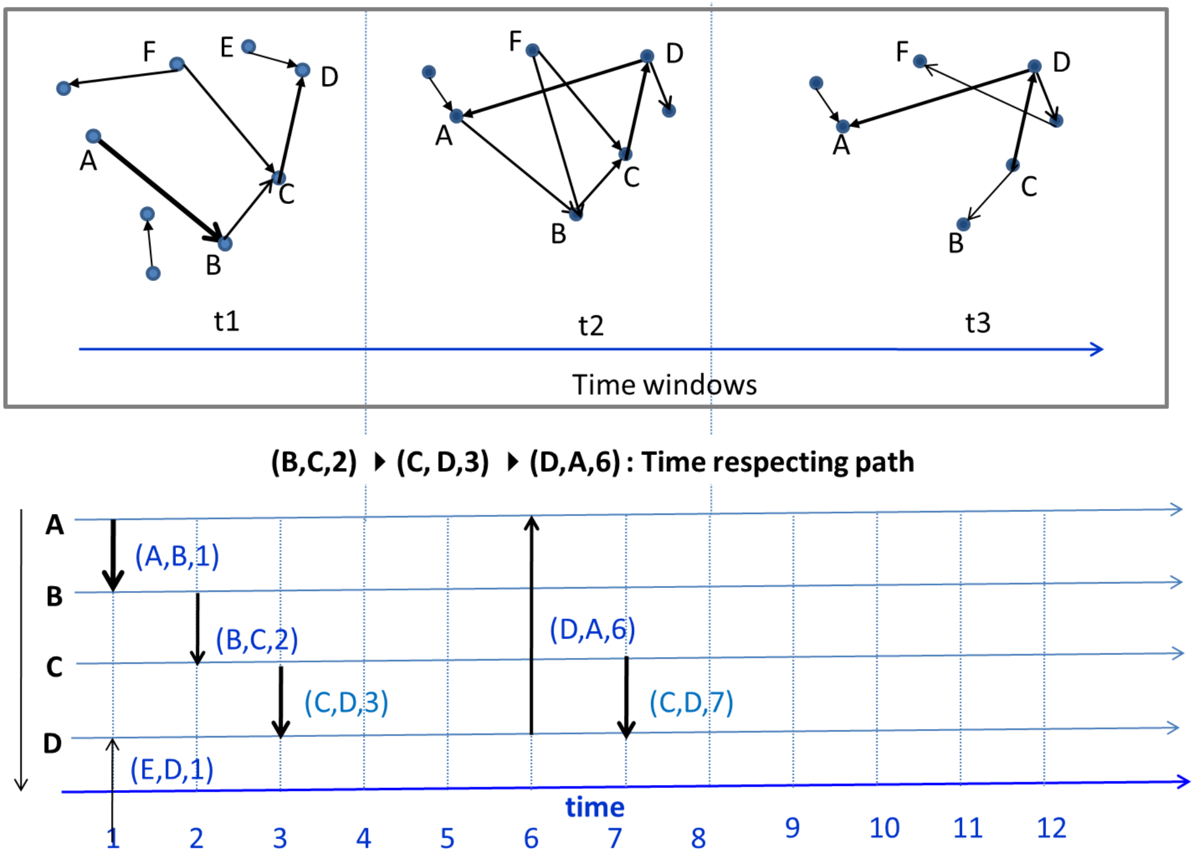


Figure 2.4: Time variation of links, temporal networks and time respecting paths.

in the sections above are loosely applicable.

Measures of temporal networks are based on the concept of *time respecting paths.* Time respecting paths are sequences of contacts with non-decreasing times that connects sets of vertices [Holme and Saramäki, 2012]. In terms of contact sequence triplets, in [Fig. 2.4], $(B, C, 2) \rightarrow (C, D, 3) \rightarrow (D, A, 6)$ is a time respecting path from $B \rightarrow A$. The constraint of having to follow time ordered sequences of contacts produces differences between temporal paths and paths in static networks. Time respecting paths define which vertices can be reached from which other vertices within same observation window $t \in [t_0, T]$ [Holme and Saramäki, 2012].

Set of influence of a node i is the set of vertices that can be reached by time respecting paths from i that begin at time t or later [Holme and Saramäki, 2012]. *Source set* of a node i is the set of vertices that can reach i through time respecting paths. Since the source set is time dependent, one can monitor the source count as a function of time, i.e. study how many other vertices may reach vertex i by time respecting paths by time t , when the paths begin no earlier than $t^{'} < t$ [Holme and Saramäki, 2012]. There are analogues for set of influence and and source set in epidemiological literature as *out-going infection chain* and *in-coming infection chain* [Nöremark et al., 2011,Frössling et al., 2012]. Source set is equivalent to in-coming infection chain and set of influence is equivalent to out-going infection chain [Fig. 2.5].

Reachability ratio (RR) is the average fraction of vertices in the sets of influence of all vertices [Holme and Saramäki, 2012]. RR may be considered as the temporal counterpart of GSCC, to estimate the average reach in a temporal network. We calculate the RRs for two regions in France (temporal network analysis and RR calculation for the national network could not be done due to constraints of computational resources).

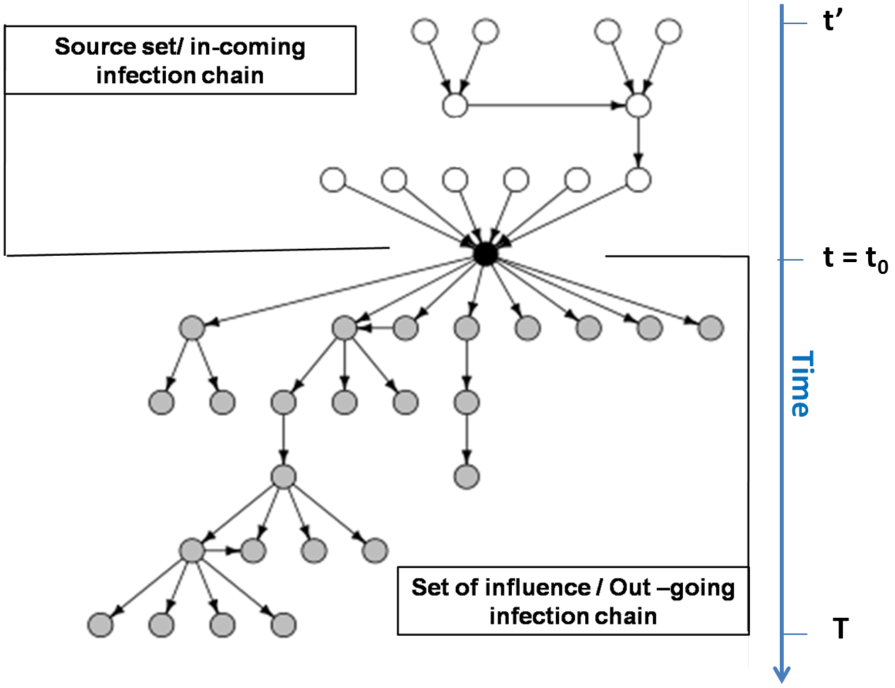


Figure 2.5: Equivalence of temporal network terms with those of epidemiological terms. Source-set is equivalent to in-coming infection chain and set of influence is equivalent to out-going infection chain. [Nöremark et al., 2011, Frössling et al., 2012, Holme and Saramäki, 2012, Nöremark and Widgren, 2014]

The estimator used for maximum epidemic size is the largest out-going chain [Nöremark et al., 2011, Dorjee et al., 2013]. The reachability ratio (equivalently, out-going infection chain averaged over all nodes) estimates the average fraction of the total nodes reachable following time respecting paths, which may be a proxy for the average epidemic size in the temporal context.

2.3.4 Regional networks and local specificities

Cattle trade networks of two administrative regions, distinct with respect to the types of farms, were compared: (a) Brittany, with the highest number of dairy farms in France and (b) Limousin, with more than 60% of total farms being beef farms. The comparisons were to reveal temporal variations of the main network descriptors, proxies for pathogen spread and identification regional specificities in the network descriptors related to the main type of farms (dairy versus beef). These regional networks were subnetworks of the FDCM network, considering the animal exchanges within the administrative boundaries of the respective geo-political regions.

2.3.5 Statistical methods and computing tools

The Kolmogorov-Smirnov (KS) statistical test was used for comparison of empirical distributions. ANOVA and t-test were used to analyse the variance and means. To test the normality, in a few occasions, the Lilliefors version of the KS test was used. Bonferroni correction was applied in the case of multiple testing (e.g. when comparing pairs of annual empirical distributions). The tests used are properly mentioned at due context when discussing the results. To fit distributions and estimate their slopes using power-law fitting, the recipe of [Clauset et al., 2009] was implemented. Statistical analyses were performed using R 3.x.x [R Core Team, 2014]. Network analysis was performed with packages igraph 0.7.1 (R 3.x.x) [Csardi and Nepusz, 2006] and networkx 1.8.1 (Python 2.7.3) [Hagberg et al., 2008].

2.4 Results

2.4.1 Data description

During the period 2005-2009, $274,231$ holdings exchanged $24,485,015$ animals through $40,357,979$ movements in $12,421,180$ batches (a batch is defined as all the movements that took place between two holdings in a single day). The share of different herd types as source (seller) and destination (buyer) in terms of flows (cattle movements), along with yearly (364 days) splitting, are shown in

Time window	**Number of**									
	total holdings(V)	total communes	movements(W)	links(A)	transaction batches	animals exchanged	**holdings** sources	**holdings** destinations	**communes** sources	**communes** destinations
2005 (01Jan05−30Dec05)	243,324	29,515	8,636,018	1,279,576	2,791,261	5,533,854	228,400	141,249	28,995	26,164
2006 (31Dec05−29Dec06)	234,440	29,293	8,624,309	1,161,018	2,660,523	5,489,154	220,219	136,470	28,731	25,875
2007 (30Dec06−28Dec07)	225,003	29,161	7,839,570	1,112,511	2,433,860	5,198,401	210,607	130,318	28,504	25,767
2008 (29Dec07−26Dec08)	214,895	28,857	7,622,453	1,026,096	2,295,869	5,110,712	200,193	123,830	28,242	25,272
2009 (27Dec08−25Dec09)	207,526	28,739	7,579,747	987,840	2,223,164	5,004,990	193,641	118,462	28,050	24,978
01/01/2005 – 31/12/2009	274,231	30,503	40,357,979	3,899,363	12,421,180	24,485,015	260,324	218,061	30,012	29,226

Table 2.1: Summary of cattle movements amongs holdings [farms (F), markets (M) and assembling centres (C)] within France during 2005-2009 based on FDCM

Table 2.1. 55% of these movements were made by male animals. In terms of race, the contributions were, 48%, 35% and 17% by beef, dairy and crossed breeds respectively. The participation of age groups were, 27%, 15%, 27%, 6% and 24% for age classes 1 to 5. [Table 2.2]

	Holding types (% of no. of holdings)			**Race types** (% of no. of animals exchanged)			**Sexes**	
Time window	Centre(C)	Farm(F)	Market(M)	Beef	Dairy	Crossed	male	female
2005	0.54	99.42	0.03	46.15	35.70	18.15	52.28	47.72
2006	0.53	99.44	0.03	47.64	34.77	17.59	52.24	47.76
2007	0.51	99.45	0.03	49.22	33.71	17.07	52.43	47.57
2008	0.48	99.48	0.04	49.98	34.18	15.83	52.84	47.16
2009	0.45	99.51	0.04	50.80	33.98	15.21	52.54	47.46
2005−2009	0.56	99.40	0.03	47.85	34.98	17.16	53.10	46.89

Table 2.2: Description of movement of cattle in terms of type of herds, race of animal and sex for FDCM 2005-2009, yearly and complete dataset

	Type of source-destination pairs (% of no. of movements)								
Time window	**CC**	**CF**	**CM**	**FC**	**FF**	**FM**	**MC**	**MF**	**MM**
2005	8.07	15.18	4.62	38.90	17.64	7.19	4.37	3.78	0.25
2006	8.61	15.24	4.45	38.98	17.18	7.08	4.49	3.74	0.23
2007	8.02	14.62	3.98	40.43	18.90	6.71	4.00	3.14	0.19
2008	7.64	14.89	3.57	41.12	19.24	6.64	3.54	3.15	0.20
2009	8.09	14.09	4.07	40.97	18.57	6.65	4.08	3.31	0.18
2005−2009	8.09	14.81	4.15	40.03	18.29	6.86	4.11	3.43	0.21

Table 2.3: Summary of movement patterns among different types of holdings [Centre (C), Farm (F), Market (M)] for the FDCM 2005-2009.

Over the calendar years, on average $23 - 24$ animals performed $35 - 37$ movements per holding and the average number of movements made by an animal was ~ 1.5 in the national scale. We also notice that there is a decrease in the number of holdings participating (active) as well as there

	Fraction of number of animals exchanged (by age class)				
Time window	Class 1	Class 2	Class 3	Class 4	Class 5
2005	0.29	0.16	0.25	0.06	0.24
2006	0.28	0.16	0.26	0.06	0.24
2007	0.26	0.15	0.27	0.06	0.24
2008	0.26	0.13	0.29	0.07	0.25
2009	0.25	0.14	0.29	0.07	0.25
2005-2009	0.27	0.15	0.27	0.06	0.24

Table 2.4: Movement pattern of cattle in French cattle trace network (2005-2009) by age classes.

	% herds as destination	32.79	20.18	28.60	17.79	0.63	218,061 (total herds as dest)
% herd types as source	Destination → Source ↓	**BEEF**	**DAIRY**	**MIXED**	**OTHER**	**MC**	% of total movements as source
30.92	**BEEF**	3.78	0.31	2.09	0.35	19.29	25.83
21.60	**DAIRY**	0.48	2.83	1.54	0.15	10.05	15.04
27.02	**MIXED**	1.45	1.69	2.42	0.29	16.79	22.62
19.98	**OTHER**	0.30	0.15	0.39	0.08	0.77	1.70
0.49	**MC**	5.62	7.69	4.52	0.42	16.57	34.81
260,324 (total herds as source)	% movements as destination	11.63	12.66	10.95	1.3	63.46	40,357,979 (total movements)

Table 2.5: Share (in %) of 'active' holdings in terms of herd type and corresponding fractions of movements between herd types for cattle movements in France from 2005 to 2009 (based on FCDM).

is a corresponding reduction in number of movements and number of animals. The percentage of decrease in number of animals exchanged between consecutive years is $1-5\%$, the highest being between 2006 (5,489,154 animals) and 2007 (5,198,401 animals). In terms of number of movements the range of variation is $0.1-9\%$ again highest change took place between the years 2006 (8,624,309 movements) and 2007 (7,839,570 movements).

The movement pattern among different types of herds [Centre (C), Farm (F), Market (M)] is not balanced. The largest chunk of movements are from F to C ($\sim 39 - 41\%$ of total movements). For incoming connections to F, direct F to F connections ($17-19\%$) and C to F ($14-15\%$) movements dominate the scenario. [Table 2.3]. Kolmogorov-Smirnov (KS) test shows that the distributions of the movement pattern in terms of holding types are not significantly different ($p >> 0.05$) over the years.

Movements of the young stock (class 1) represent $25 - 29\%$, class 3 animals also fall in the same range, whereas class 5 is slightly behind the previous two with $\sim 23-25\%$ of the movements. Fraction of total movements for the class 2 animals is $\sim 14 - 16\%$, and class 4 movements are the least $\sim$ 6% [Table 2.4].

The respective percentage shares per herd-type were 30.06 (beef),20.63 (dairy), 25.79 (mixed), 23.05 (others) and 0.47% (markets and centres, MC) among all the holdings active in 2005-2009. Among these five subnetworks defined, those based on animal race drive important fractions of the whole network : beef, dairy and mixed breed-wise subsets represent 0.75, 0.60, 0.73, respectively, in terms of nodes, 0.41, 0.49, 0.29 in terms of links and 0.49, 0.33 and 0.18 of global flows, for the period from 2005 to 2009. Beef and dairy herd-wise networks cover 0.36 and 0.26 respectively of the total active nodes, 0.19 and 0.21 of the total links and 0.33 and 0.19 of total flows. It is also observed that the X-MC movements (X: any of the categories defined above) are the dominating type of movements, more than 63% of the total movements are destined to MCs. In terms of out-flows (holdings as source) also MCs dominate with $\sim 35\%$ of the total movements originating from them, followed by BEEF ($\sim 26\%$) and MIXED ($\sim 23\%$) herds [Table 2.5]. This analysis shows the importance of MCs in the cattle trade network in France.

2.4.2 Static networks at different levels of spatial and time aggregation

At the national level, for yearly aggregated networks, irrespective to herd type or animal breed, the average values of indicators are quite stable from one year to the next for both holdings and communes as nodes. For the network based on holdings, average nodes characteristics such as in- and out-strength vary between 60-64 and 37-39, respectively, for 2005 to 2009. Network cohesion is also quite stable over years: the GSCCs contain $0.42, 0.42, 0.41, 0.40$, and 0.39 of the total number of nodes in the annual networks from 2005 to 2009; and other indicators such as assortativity (-0.06 to -0.11), clustering coefficient (0.0051-0.0054), reciprocity (0.17- 0.19), average path length (4.7-4.9), and diameter (21-25) exhibit little variation. When aggregating the transactions of all holdings in a commune and considering each commune as a node, the GSCCs comprise $0.85, 0.85, 0.84, 0.84$, and 0.82 of the total active communes for 2005 to 2009. All distances on networks globally decrease, assortativity and clustering coefficient also are close to zero. The main change compared to the holdings based network, not necessarily directly related to the aggregation effect, is the increase in reciprocity (0.23-0.25) [Table 2.6 and 2.7]. At the annual aggregation level, the network descriptors are not conclusively different from each other.

Time window	in-strength / node	out-strength/node	Av. path length	Assortativity	Clustering coefficient
2005	61.14	37.81	4.70	−0.06	0.0054
2006	63.20	39.16	4.76	−0.08	0.0051
2007	60.16	37.22	4.87	−0.11	0.0053
2008	61.56	38.08	4.92	−0.11	0.0054
2009	63.98	39.14	4.92	−0.11	0.0054

Table 2.6: Network descriptors for yearly aggregated networks with holdings as nodes (2005-2009).

Time window	Reciprocity	Diameter	GSCC size (proportion of total nodes in the time window)	No of SCCs
2005	0.186	25	42.2%	57.1%
2006	0.182	25	42.5%	56.9%
2007	0.178	22	41.5%	57.7%
2008	0.172	22	40.0%	59.2%
2009	0.182	21	39.5%	59.6%

Table 2.6: (... continued) Network descriptors for yearly aggregated networks with holdings as nodes (2005-2009).

For a monthly (four weeks) time window, the number of active holdings varies between 82,301 and 130,238, the number of movements ranges between 416,483 and 864,621. The GSCC sizes are between 4.6% and 8.0% of the number of nodes active during the window. The ranges of the other

Time window	in-strength / node	out-strength/node	Av. path length	Assortativity	Clustering coefficient	Reciprocity	Diameter
2005	297.85	330.07	3.48	-0.06	0.080	0.245	11
2006	300.17	333.31	3.54	-0.08	0.075	0.240	14
2007	275.03	304.25	3.62	-0.09	0.077	0.234	13
2008	269.90	301.62	3.64	-0.10	0.076	0.227	11
2009	270.22	303.48	3.65	-0.10	0.076	0.231	13

Table 2.7: Network descriptors for yearly aggregated networks with communes as nodes (2005-2009).

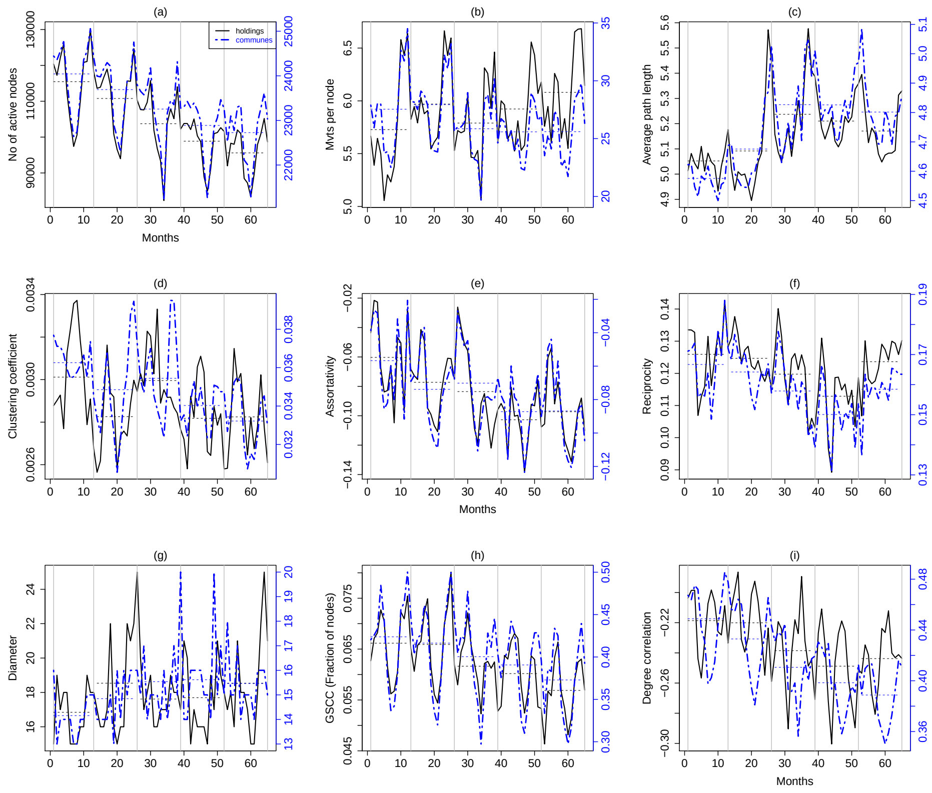


Figure 2.6: Classical network descriptors (labelled on the y-axes of each panel from (a) to (i)) for the national network of cattle movements in monthly (4-week) windows from January 2005 to December 2009. Calculations were performed using the unweighted directed variant of networks. In black: holdings as nodes; in blue: communes as nodes; vertical gray lines: 364 days; horizontal dashed lines: average over the year.

observed average descriptors are: number of SCCs, $91.6 - 94.9\%$ of the total nodes, average path length: $4.89 - 5.57$, assortativity:-0.14− -0.02 , clustering coefficient: $0.0026 - 0.0034$, reciprocity:

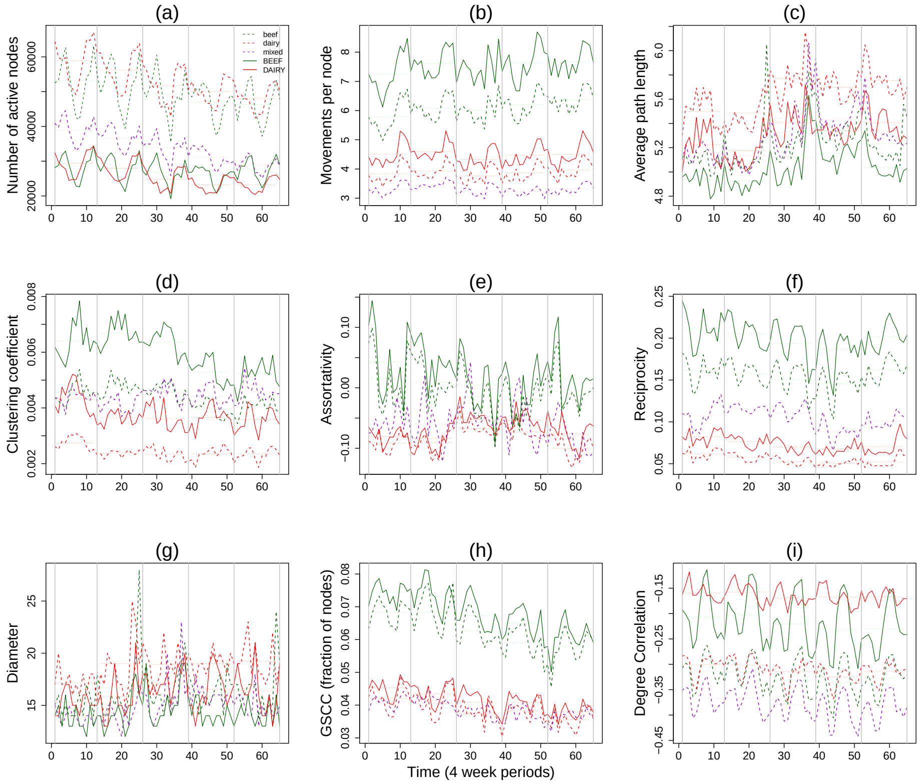


Figure 2.7: Classical network indicators (labelled on the y-axes of each panel from (a) to (i)) for the subnetworks based on animal breed and herd type for the national network of cattle movements in 4-week time windows in France from January 2005 to December 2009. Subnetworks consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). Markets and assembling centres are part of the subnetworks. Calculations are performed using the directed unweighted variant of subnetworks with holdings as nodes. Averages over years for each indicator are represented (horizontal dotted lines). 364-day windows are marked by vertical grey lines.

$0.089 - 0.143$ and diameter: $15 - 25$. Figure 2.6 shows the monthly variations of some these parameters and Fig. 2.7 shows corresponding variations for the subnetworks based on race and herd type respectively. Ranges of variation of the main monthly indicators for the global network, calculated for the directed unweighted variants, over the whole period studied, are provided in Fig. 2.6. Globally, similar fluctuating behaviours are noticed for the majority of indicators, irrespective to the level of aggregation for node definition (holdings or communes). Despite a decreasing trend, a yearly pattern in the number of active nodes is visible, with a ratio between upward (March) and downward (August) peaks of about 20% (Fig. 2.6a). A more severe downward peak is visible in 2007 and in the same year, conversely, the average path length achieves a noticeable increase. There is almost no clustering and the assortativity has low negative values (Fig. 2.6d, e). The proportion of bi-directional links is also low (reciprocity less than 0.13 and 0.19 for holdings and communes, respectively; Fig. 2.6f). The proportion of nodes belonging to the GSCC is, on average, 6 to 7 times greater when communes are the nodes of the network (Fig. 2.6h). The most significant difference with respect to the level of aggregation is illustrated by the correlation of degrees: it is negative when nodes are holdings and positive when nodes are communes (Fig. 2.6i). Here again, the most important variation, especially for the node-based network, is observed in 2007. The analysis of monthly herd-wise and breed-wise subnetworks reveals variations that could be assimilated to annual cycles and suggests more generally characteristics similar to those of the global network (Fig. 2.7). Race-wise subnetworks include almost twice more nodes than herd-wise subnetworks (Fig. 2.7a). A noticeable difference between beef and dairy subnetworks concerns the number of movements per node (Fig. 2.7b), which is uniformly larger for the former over the whole period. Clustering coefficient and assortativity are close to zero (Fig. 2.7d, e). The reciprocity and the cohesion expressed through the GSCC are two to three times more important for beef than for dairy subnetworks (Fig. 2.7f).

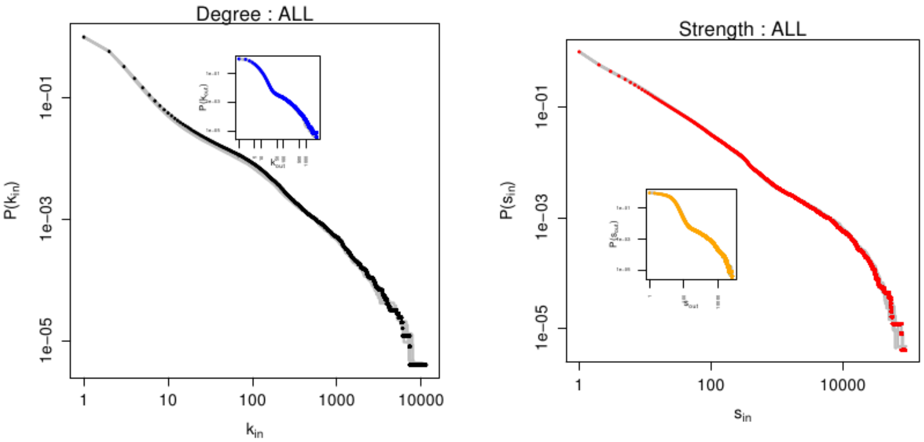


Figure 2.8: Distributions of degree [and strength] for annual time windows with holdings as nodes. The main figures are distributions of in- degree and in-strength. Figures in the inset are the distributions of out-degree and out-strength respectively. The grey lines are for five yearly networks and the coloured lines are for the year 2005.

2.4.3 Distribution of centrality measures and other features for the networks with holdings as nodes

In the annual time window, the degree distributions for the aggregated networks with holdings as nodes are not significantly different from each other (Kolmogorov-Smirnov and Fisher exact test $p >> 0.05$). The slopes of the cumulative degree distributions when fitted to a power-law fitting [Clauset et al., 2009] are 1.95 (in degree) and 1.77 (out degree) respectively and those for the cumulative distributions of strength are 2.04 (in) and 2.08 (out) respectively [Fig. 2.8]. Between year variations for the herd and race type based networks are also not significantly different.

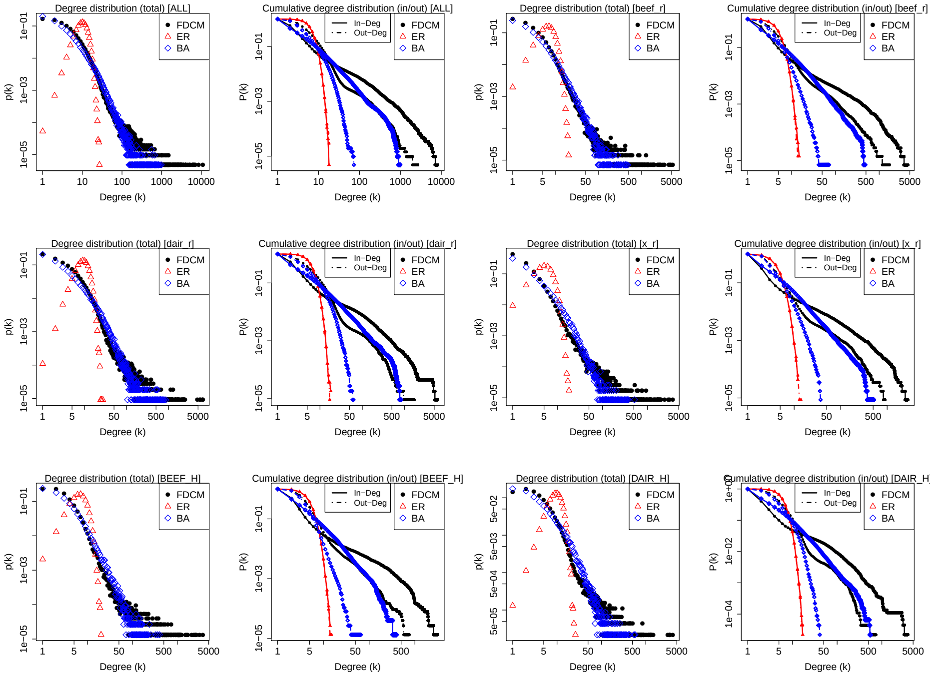


Figure 2.9: Degree distributions of the network based on French Database of Cattle Movements (FDCM) [along with the race and herd type based subnetworks] with theoretical Erdös-Rényi (ER) random network model and Barabási-Albert (BA) scale-free network model of same dimensions.

For a visual comparison of the cattle movement network with some well characterised theoretical networks, we simulated two networks models, Erdös-Rényi (ER) random network model (degree distribution is a Poisson distribution) and Barabási-Albert (BA) scale-free network model (degree distribution is a power law distribution) [Albert and Barabási, 2002] with equal sizes for all the five race and herd type based subnetworks and the full network of active holdings. Degree distributions of the simulated and FDCM networks (for 2005-2009) were plotted together for the six networks in

[Fig. 2.9]. These two model networks were chosen as these two are well characterised examples of random and scale-free networks, respectively. The FDCM network resembles to a BA like network when *total* degree is considered. But there are significant differences when in- and out- degrees are separately considered. Degree distributions of FDCM are nowhere near those of ER random networks.

The distributions of betweenness, closeness centralities show no significant difference among the same classes at different levels of time-aggregation. For instance, the distribution betweenness centrality between two annual networks with holdings as nodes were not significantly different, same for other spatially or temporally aggregated networks (data not shown).

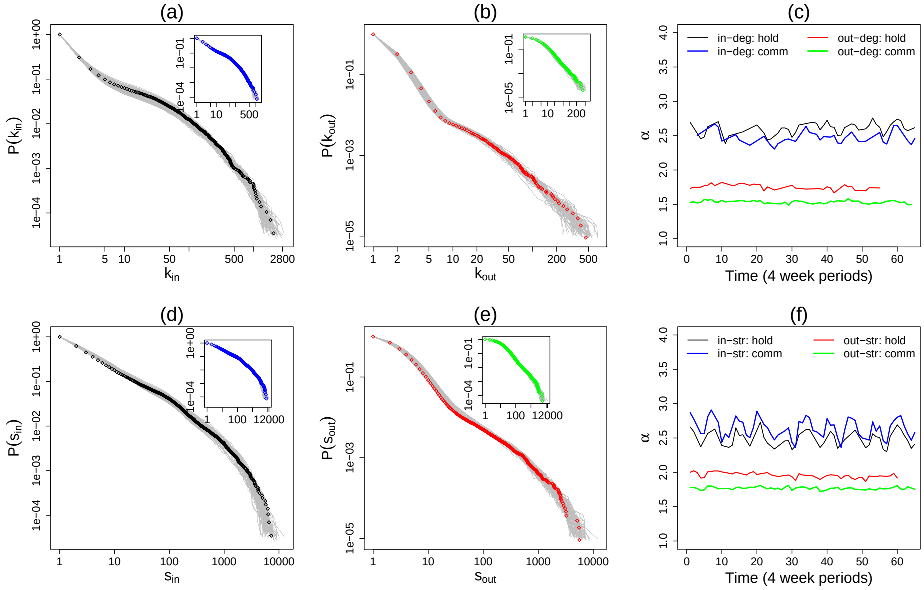


Figure 2.10: Cumulative distribution of degrees (number of holdings in contact with each node); in-degrees (a), out-degrees (b)) and strengths (number of movements per node; (in-strength (d); out-strength (e)) in monthly (4-week time periods) full networks at national scale, with holdings as nodes for major graphs, and communes as nodes for insets. Coloured dots are for January 2005, the first month of the the data set, spanning until December 2009 and including all cattle movements in France. A power law fit of the distributions ((c) and (f)) shows a range of exponents for each indicator (only months with a statistically significant estimate are kept).

For monthly networks, exponents of degree and strength distributions (Fig. 2.10) are rather consistent over time. Monthly exponents range between 2.3 and 2.9 for in-degree and in-strength distributions (Fig. 2.10c, f), and between 1.5 and 2 for out-degree and out-strength distributions (Fig. 2.10c, f), respectively, for both holdings and communes as nodes. When separately analysing monthly degree and strength distributions for beef and dairy networks, they appear quite similar, except for the in-strength for which monthly exponents are around 2.5 for beef and around 3 for dairy herd-based subnetwork (data not shown).

Local features of network cohesion are captured by the distributions of clustering coefficient and average nearest-neighbour degree, both weighted and unweighted variants, when considering the global network with holdings as nodes (Fig. 2.11, data for 2009 as an example, consistent results for the other years). While clustering coefficients exhibit a continuously decreasing trend (Fig. 2.11a), average nearest-neighbour degrees have distributions with a plateau for low values followed by a decreasing phase (Fig. 2.11b). Accounting for weights modifies the shape of these distributions.

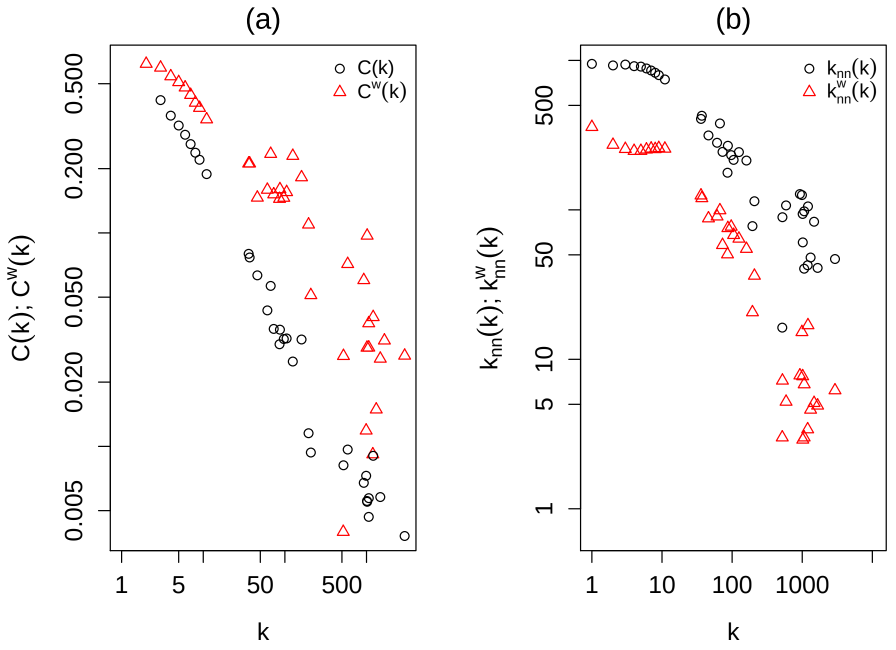


Figure 2.11: Distributions of unweighted (black dots) and weighted (red triangles) clustering coefficients (a) and nearest-neighbour degrees (b), averaged over neighbours of each node (holding) of degree k. All cattle movements at the national level, in France, during 2009 are considered.

The relationship between geographical distance and shortest path length on network (accounting for directed links) was explored. According to data for 2009 (similar results for the other years), distributions of distance for path lengths from 1 to 3 are not significantly different between subnetworks (Fig. 2.12a-c). At least 85% of directly connected farms are at a distance lower than 200 km and they exchange preferentially within the same commune, irrespective to the type of network (peak for very low values in Fig. 2.12a). When connected through one intermediate node on the global network, holdings are more likely to be located at 40 km from each other and 95% of them are at less than 504 km Fig. (2.12b). For all subnetworks, the geographical distance for a shortest path equal to three is in 95% of cases equal at most 633 km (Fig. 2.12c). For direct connections, the mean distance is 77 km for beef, 107 km for dairy herd-wise subnetworks and 94 for the global network, whereas medians are 43 km, 64 km and 50 km, respectively.

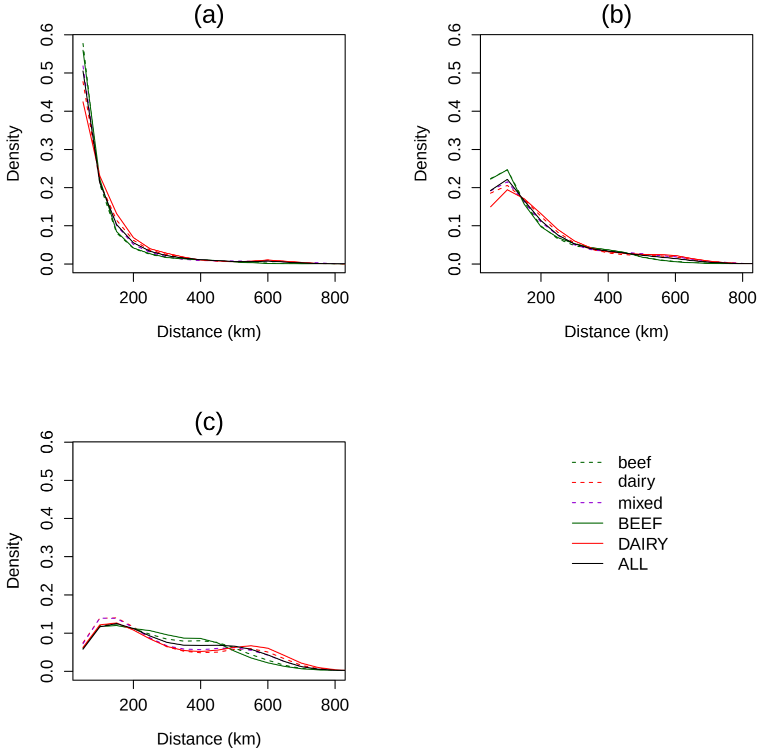


Figure 2.12: Distributions of geographical distances between pairs of nodes separated by shortest paths of length 1 (a), 2 (b) and 3 (c). Subnetworks consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). The full network is also analysed (black). Markets and assembling centres are part of the subnetworks. All cattle movements at the national level, in France, during 2009 are considered.

2.4.4 Proxy for pathogen spread and its control for static networks

Percolation studies investigating variation of the GSCC size with respect to random and three targeted node removal schemes (degree, strength and betweenness centrality) are shown in [Figures 2.13 and 2.14] for the national network and the five race-breed based networks respectively, at the

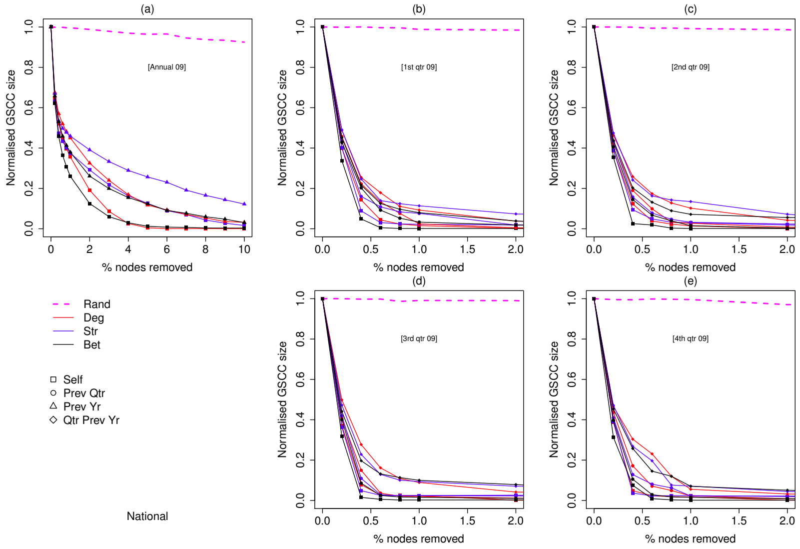


Figure 2.13: Effect on the GSCC of different strategies of node removal for the national network of holdings. (a) Network for the whole year and (b)-(e) for each quarter in 2009. Scenarios tested consist in targeting nodes at random, and based on degree, strength, and betweenness distributions, from the network in the current year, in the previous year, in the previous quarter and in the same quarter the previous year.

national level, yearly (Fig. 2.13a) and quarterly (Fig. 2.13b-e), based on information on the current year and the current trimester (real-time scenario) or the previous year, the previous trimester or the corresponding trimester of the previous year (delayed access to data). The random sampling of nodes for removal has almost no impact on the GSCC size up to a removal of 10% of the total number of active nodes (Fig. 2.13a-e, data for 2009). When the removal of nodes is based on centrality measures (degree, strength, and betweenness) by targeting the top nodes (ordered in descending order for each of theses three measures), the most efficient strategy is the one based on betweenness, whereas the less efficient one is based on strength for the annual network (Fig. 2.13a), and on degree or strength for 3-month networks (Fig. 2.13b-e). Indeed, removing the top 5% of nodes based on betweenness achieves to completely break the GSCC structure for the annual network, this fraction being only 0.5% for 3-month networks. For these networks, there is no significant difference in the results depending on the quarter analysed. Unsurprisingly, the most efficient strategy is based on information from the current network (unlikely scenario, used here only to provide a lower bound). Removing 5% of the nodes based on their betweenness or degree the previous year enables the GSCC size to be reduced by more than 80%. Moreover and unexpectedly, knowing the characteristics of the previous 3-month network is more efficient than basing targeting according to the network in the same quarter the previous year.

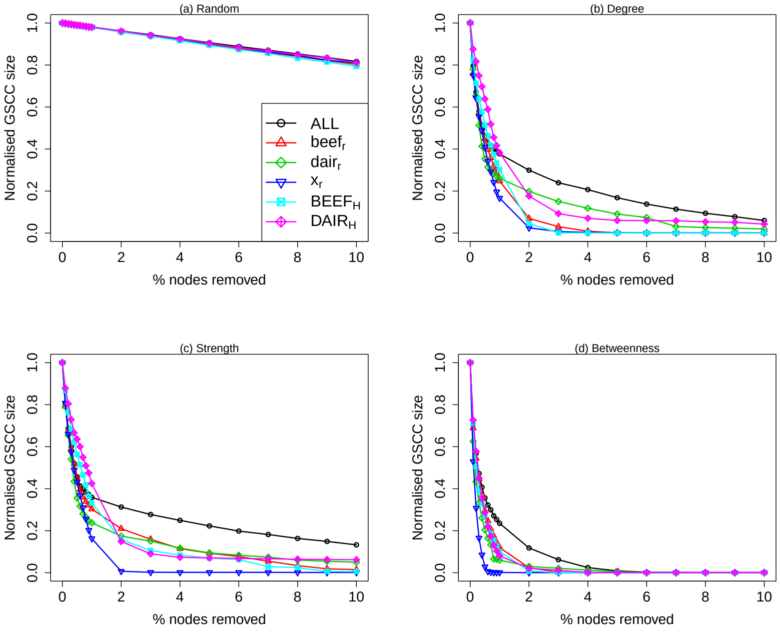


Figure 2.14: Effect on the GSCC of different strategies of node removal for the subnetworks [race based: beef$_r$, dair$_r$, x$_r$ (mixed/crossed); herd based: BEEF$_H$, DAIR$_H$ and the complete network: ALL] of holdings par different strategy of node removal: (a) Random, (b) Degree, (c) Strength, and (d) Betweenness for annual network of 2009.

Percolation study for the five race and herd-based subnetworks are shown in (Fig. 2.14). All the subnetworks were found to be comparatively highly resilient to random removal of nodes, compared to targeted removal (Fig. 2.14a). Subnetworks were found to be more sensitive to node removal than the full network, with the mixed breed subnetwork being most vulnerable. For degree based removal strategy, beef subnetworks (both race and herd-wise) were more vulnerable compared to dairy subnetworks (Fig. 2.14b). Similarly to the the full network, the betweenness centrality centred node removal strategy is the most efficient in breaking down: all subnetworks almost break-down with top 2% of the nodes removed, whereas for the full network it takes about 5% of the nodes (Fig. 2.14d).

2.4.5 Network dynamics: temporal variation of network descriptors

We observe that when the time window of aggregation is continuously increased, the descriptors show variations in their distributions. In Fig. 2.15 (a-d), we show the variation in degree and strength distribution with increasing time window size. We increase the time window of aggregation in doubles of a week and plot the distributions. The distributions at first glance appear different, shifting away in the upward direction as the time window increases. But when rescaled by the mean of the distributions, we observe that they fall on each other. The similarity of the rescaled distributions are tested by calculating L^2 distances among the scaled distributions [Fig 2.16]. Observed L^2 distances are small (range: 0.005 - 0.04). Rescaling and calculating L^2 distances show that the degree distribution of cattle movement network under study is not dependent on window width of aggregation. Similar observations are noted for the other descriptors too.

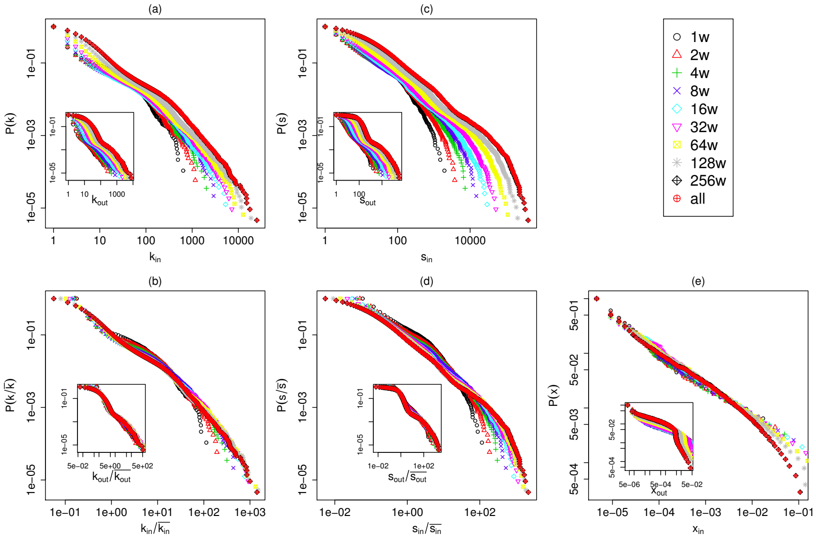


Figure 2.15: Variations in the distributions of the descriptors with increasing time window of aggregation (in multiples of weeks): (a) degree and (c) strength. Here we see increase the time window of aggregation in doubles of a week and plot the in-degree/strength (in the inset out-degree/strength) distribution as these different window. Rescaled distributions, [(b) degree and (d) strength], showing time window independence of the descriptor. Rescaling shows that the degree distribution of cattle movement network under study is not dependent on window width of aggregation. (e) Distributions for activity.

We see that over all the years 65% of the holdings were active whereas the common links are only 3% of total links. To quantify the fraction of links common between consecutive time windows we calculate short range similarity (SRS). It is worth noticing that the range of SRS in case of cattle trade network is quite low (maximum $\sim 20\%$) [Fig. 2.17].

In terms of activity, incoming activity distributions for increasing sizes of time window were found to be statistically similar, while those of outgoing activity do not fit towards the tails (highly active nodes) [Fig. 2.15(e)].

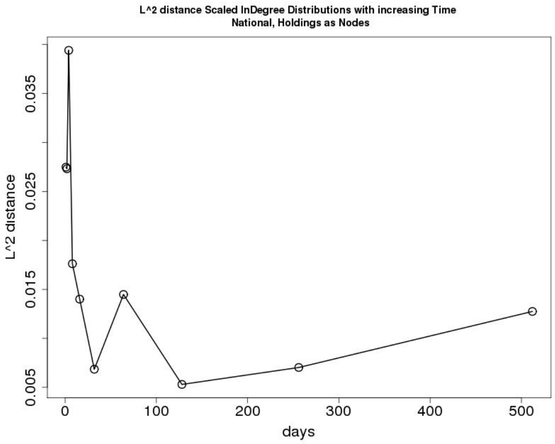


Figure 2.16: L^2 distance of rescaled in-degree distributions of cattle movement network. (after [Krings et al., 2012]).

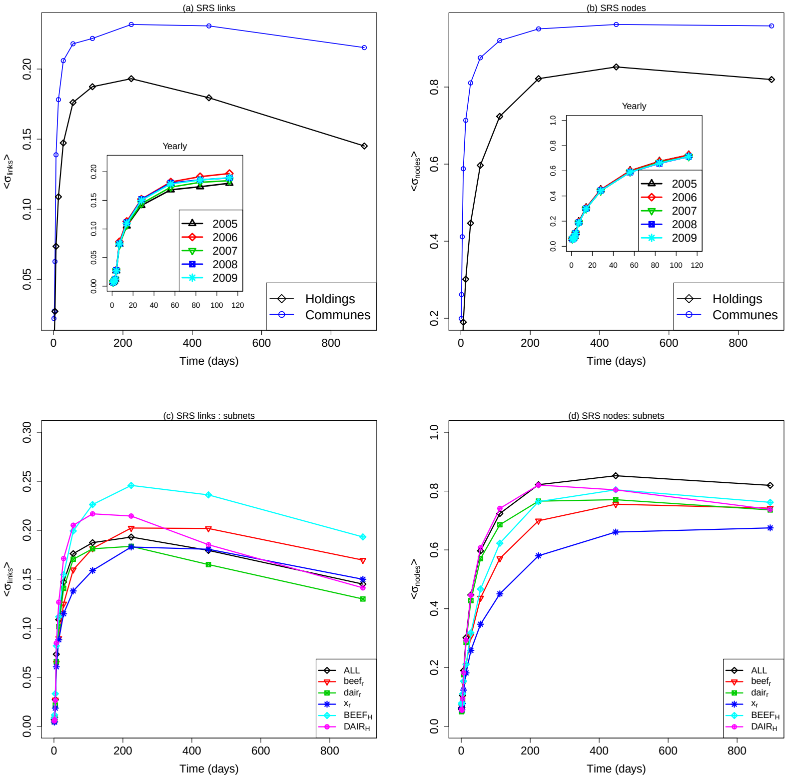


Figure 2.17: Short range similarity (SRS) for the whole network [(a) for links and (b) for nodes] We show the SRS for both holdings as nodes and communes as nodes for the whole period 2005-2009. In the inset are the SRS's for yearly networks. In (c) and (d) we show SRS for links and nodes for the subnetworks based on farm types and breed types.

2.4.6 Regional networks : two regions with contrasting farming systems

Amongst average monthly descriptors of regional networks based on holdings as nodes, noticeable differences between the two regions concern the number of active nodes (10000-14000 for Brittany and 2500-5000 for Limousin), the number of movements per node (3.2-4.2 in Brittany and 5-7.5 in Limousin), and the reciprocity (0.02-0.035 in Brittany and 0.1-0.2 in Limousin). Other characteristics are similar with those of the national networks (no clustering, small GSCC as fraction of total number of active nodes, moderate negative degree correlation). Comparison of the network descriptors for the two regions is shown in Figure 2.18.

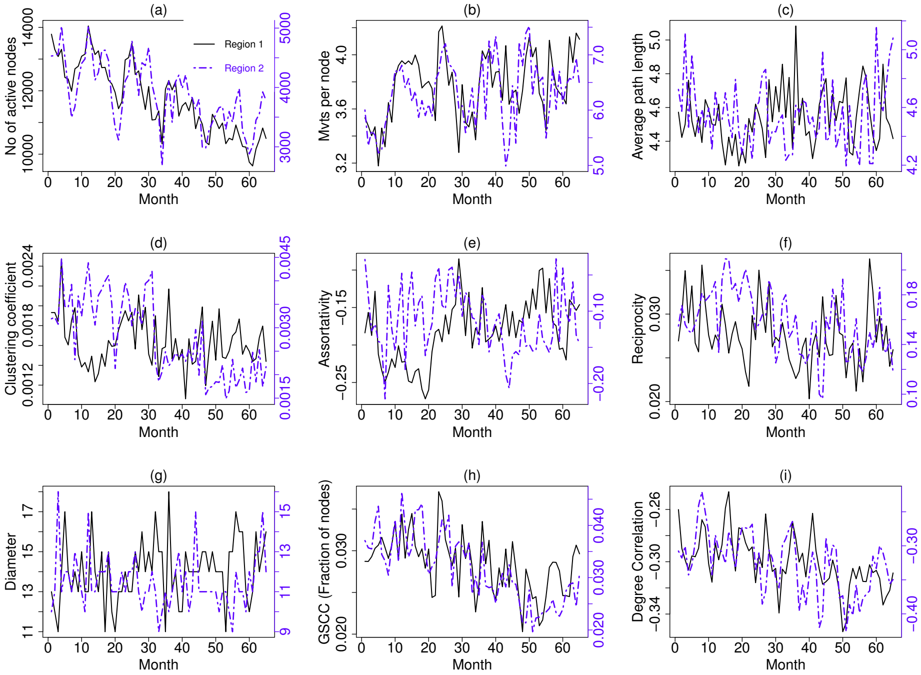


Figure 2.18: Comparison of network descriptors for a mainly dairy region (Region 1: Brittany) and a mainly beef region (Region 2: Limousin) of France, for monthly networks, from 2005 to 2009, with holdings as nodes.

The stability of the backbone (in terms of links) over time, expressed through the SRS, reaches its maximum for a 120-day window: two consecutive networks aggregated over this window have 20% of common links in Brittany and almost 27% in Limousin. [Figure 2.19]

Similarly to the national networks, percolation analysis reveals that the most efficient strategy in terms of the reduction in the GSCC size consists in targeting nodes in descending order of their betweenness. The reachability ratio (average and maximum) was calculated for increasing time

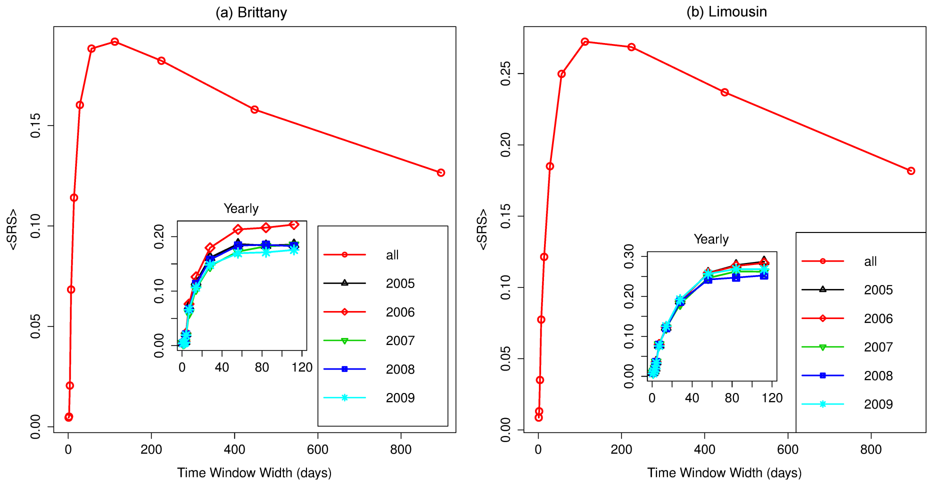


Figure 2.19: Short range similarity (SRS) for the regional networks of (a) Brittany and (b) Limousin region. The main plots show the SRS (links) for the movement network 2005-2009 for the regions. The inset-plots are the SRS (links) for the annual networks.

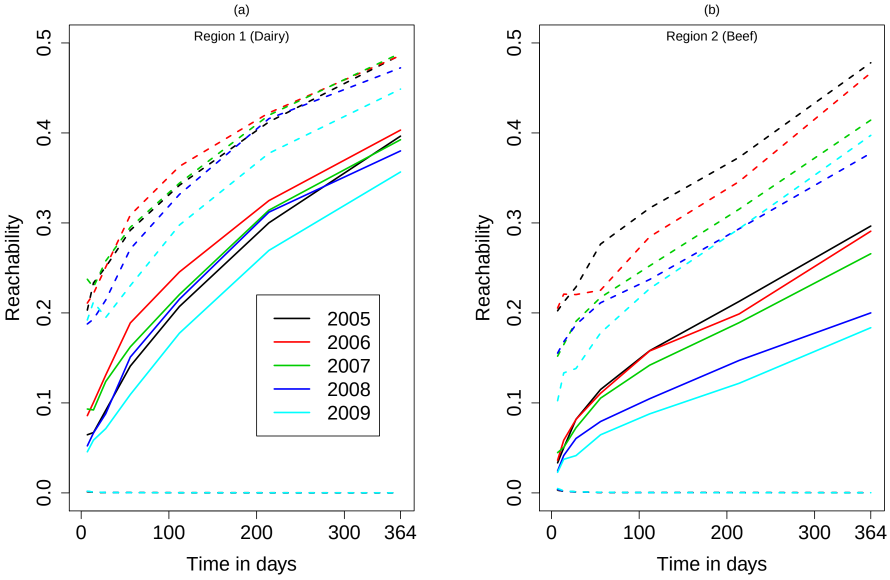


Figure 2.20: Average (plain line) and maximum (dashed lines) reachability ratio calculated for increasing time windows for a mainly dairy region (a: Brittany) and a mainly beef region (b: Limousin) of France, for annual networks, from 2005 to 2009, with holdings as nodes. The denominator is the total number of nodes with at least one in-connection in the time window considered.

windows for each of the five years (Figure 2.20) for both regions. Whereas the maximum RR are comparable between the two regions for one-year windows for all years, this is less true for average RR. In Brittany, the average fraction of nodes to be reached by time respecting paths is between

0.35 and 0.4 for a one-year time window for all years, whereas it is lower (0.18-0.28) in Limousin. The between year variability is more pronounced in the beef region.

2.5 Discussion

French cattle movement network between 2005 and 2009 was studied at different spatial and temporal granularities. The full network is comprised yearly of around 2.40×10^5 of holdings (farms, assembling centres, and markets), involving 8.6×10^6 of movements and 5.5×10^6 of animals (mainly young, i.e. less than 20 months of age). Our study provides a deeper outlook of this cattle exchange network and completes previous studies based on a one-year dataset [Rautureau et al., 2011]. Subnetworks based on race and herd type (beef, dairy and mixed) were thoroughly explored together with the global network. In addition to calculating indicators for static networks, we followed their temporal evolution and investigated the impact of the width of increasing time aggregation windows. Proxies for the outcome of disease spread on this network were analysed with the objective of guiding possible intervention measures to control epidemic outbreaks. Specificities of two regional networks corresponding to dominant populations of beef and dairy herds were separately assessed.

The study highlights a decreasing trend in the network size (number of active nodes) over time, regardless of the aggregation unit and subnetwork type, which may be a sign of merging of smaller individual farms into bigger corporate farms via acquisitions and mergers. In addition to this trend, a yearly pattern in the number of active nodes is visible, with an upward peak in March and a downward peak in August. A more noticeable downward peak for the global network is observed in 2007, along with a more severe increase in the average shortest path length. This latter aspect could imply that the connectivity of the network was somehow affected in 2007. The GSCC also exhibits annual trends, with seasonal minimum and maximum values. This may suggest that the spread of an infection on this network of cattle farms would have different behaviours depending on the onset time. The other indicators do not significantly vary among years for the temporal and spatial granularities studied. Most of the network measures exhibit similar qualitative trends between holdings and communes as nodes, either on a monthly or a yearly basis. The proportion of bi-directional links (reciprocity) is quite low, suggesting that holdings (communes) to which a given holding (commune) sells animals are more probably different from those from each it buys animals. This behaviour could contribute to the spread of a potential infectious agent beyond the local neighbourhood (on the network) of initially infected holdings (communes). However, this effect could be counterbalanced by the fact that direct contacts between holdings mostly occur in an area with a radius smaller than 200 km, with half of movements at less than 50 km and preferentially even within the same commune.

2.5.1 Static networks

We studied the evolution of the network of cattle movement in France at different time windows and at three levels of spatial aggregation. We observed that the distributions of degree and strength are heavy tailed, a sign of resilience to random node removal, like many other types of networks, natural or man-made [Albert and Barabási, 2002, Newman, 2010, Rautureau et al., 2011, Bajardi et al., 2011]. The distributions were found to be similar at the two levels of spatial aggregation (holdings and communes) and also the two levels of time aggregation (monthly and yearly). The results were similar for both national and regional networks. The network descriptors were found to be not significantly different from each other at their corresponding scales of time and space. We could not significantly identify any differences among the sub-networks in terms of topological characteristics over time of observation.

The analysis of monthly herd-wise and breed-wise subnetworks shows variations that resemble to annual cycles, with characteristics similar to those of the global network. Race-wise subnetworks, built with no condition on the type of source and destination herds, include almost twice more nodes than herd-wise subnetworks. This suggests that herds sell and buy any race of animals, regardless of their type, although they may prefer to trade with herds of similar type. Moreover, if we consider the 5-year period, beef, dairy and mixed herds send animals to or receive animals mostly from markets and assembly centres (rather than other herds). Hence, it seems that in France, beef and dairy cattle sectors are interrelated by trade, which points out towards a global management of the cattle system in the case of a crisis. Beef and dairy subnetworks differ in the number of movements per node, uniformly larger for the former over the whole period. The reciprocity and the cohesion expressed through the GSCC are two to three times more important for beef than for dairy subnetworks. These three elements could induce a potential increased vulnerability of beef system to infection.

On the field, interventions are generally implemented at a regional scale, in relation with a local organization of animal health management for many animal diseases, but also because it is more convenient to geographically group interventions. The fact that cattle movements largely occur in the neighbourhood of the source holding also points out towards a regional organization. Moreover, regional networks could have specificities which should be taken into account when implementing interventions, to enhance control effectiveness. Indeed, when comparing a mainly dairy region, as Brittany, with a mainly beef region, as Limousin, we found that not only the number of holdings and the number of movements differ, but also the reciprocity, the backbone stability, and the reachability of nodes do, all the three being higher in the latter. This corroborates the analysis of monthly variations of network indicators for dairy and beef subnetworks.

2.5.2 Network dynamics, temporal networks and epidemic size estimation

There were variations observed in the distributions of the network descriptors with increasing time. These 'virtual differences' among the distributions of the network descriptor with changing time window were shown to be most probably a manifestation of scale, and they were shown to be 'similar' by rescaling the distributions and verified by calculating L^2 distances among the scaled distributions [Krings et al., 2012]. It shows that topologically the evolution of the network over time is robust. This helps modelling the empirical network and sheds light on its intrinsic dynamics. The activity seems not to be influenced by time dependence (as noticed in [Perra et al., 2012, Karsai et al., 2014]). This also corroborates the fact that it is more appropriate to use relative contributions of holdings with respect to global interactions than absolute values, since the former are a priori less sensitive to the period considered. However, increasing the width of time window provides more details on tail distributions. Our findings suggest that a 8-16 week period would suffice for robust estimation of centrality distributions, even if complete datasets would provide more accurate ones. Although the topological properties of the network at various time and space scale are similar, we observe low *short range similarity* (SRS) among networks captured as consecutive time-windows (of equal width). Lower value of SRS indicates that there are not many common links from window to window. In terms of the cattle movement context, this means that two farms exchanging animals in a particular time window (a day/week/month) do not necessarily exchange animals in the next time window nor sell or buy animals in every time window. But they do follow a pattern, the year-wise SRS distributions show similar behaviour over the years (at both regional and national scales). The

SRS over the years were not significantly different from each other.

Since the number of common links (and nodes) over various (consecutive) time windows (week, month, year) is low (both at national and regional scales), any “backbone” of the network capturing the most stable connections not necessarily reflect the overall features of the original network.

For subnetworks, common links vary on average, between 17% for mixed to 24% for beef breed-wise subnetworks. This latter seems to have an increased faithfulness in trade relationships. However, temporal stability being low overall, longer time series data should be preferred, in order to have broader information on exchange structure, when implications of the network topology on pathogen spread are studied. This is especially true as the dynamic nature of such networks has a substantial impact on pathogen spread. This was revealed using time stamped chain of contacts [Nöremark et al., 2011, Büttner et al., 2013, Dorjee et al., 2013, Konschake et al., 2013, Nöremark and Widgren, 2014]. In the absence of time series of animal movements long enough to allow a comprehensive representation of exchanges between herds, statistical and mechanistic modelling studies may prove to be useful. Indeed, by identifying, based on available observations, the determinants of trade between holdings or modelling the mechanisms underlying such relationships, it would be possible to simulate animal exchange networks with realistic characteristics. This is of high importance to accurately predict the possible outcome of a new infection of cattle spreading at large scale and the impact of different possible interventions.

To control pathogen spread through regulations of animal movements, issues remain because available resources are limited (material, human resources) and the information on holding characteristics is only retrospective (on past movements). Identifying which nodes should be targeted to efficiently control disease spread is valuable to focus the management effort on the most relevant nodes. According to our findings following percolation analysis, when only access to delayed data is possible, preventing the totality of animal movements concerning 1 to 5% of the holdings with the highest betweenness in the network based on the previous year reduces the size of the GSCC by 60% to 90 %. This strategy outperforms random removal (completely ineffective) and nodes targeting based on degrees and strengths, for this latter even when real-time data is available. This is directly related to the fact that networks with power law degree distributions, also called scale-free networks [Albert and Barabási, 2002] are robust against random failures, but, when the node removal is done according to nodes ranked by degree, a comparatively small number of removals can lead to break down of the network. This also corroborates up to some extent the results of previous studies [Kiss et al., 2006, Rautureau et al., 2012, Büttner et al., 2013].

2.5.3 Implications for the spread of BVDV

Exchanges of animals between herds are important for any epidemiological studies. Network is just a part of the story, the local within-herd disease dynamics giving the probability of an animal exchanged to be infected being the key starting point of BVDV spread between two herds, rather than the network alone.

Dairy herds need to continuously renew their breeding stock. Therefore, the results of network analyses are important not only for particular disease like BVDV but also from any epidemic process point of view which may crawl on the network between the herds. From a theoretical point of view, the cattle movement network we studied was found to be topologically robust at all time scales and resilient to random node removal. This is very important from the point of view of designing control strategies. Percolation analyses showed that to control the spreading of a generic epidemic process, nodes with highest connectivity or flow through them should be the locations where the control methods are to be pressed upon. The role of highly connected holdings in the spread of BVDV is not clearly known. The information obtained from the network analyses may be used to test hypotheses about the role of nodes with different topological characteristics. For example, how does the BVDV infection spread in a network given the centrality measure of node with initial infection? Is there a linear relationship between the centrality of a node and the reachability of a pathogen in the network? These questions may be tested given the BVDV dynamics within a herd.

2.5.4 Key results

- The cattle movement network studied shows heavy tailed distributions of degree and strength, which implies resilience to random node removal.
- The evolution of the network topology is robust. It preserves the topology described by classical network descriptors at all levels of studied temporal resolution.
- Topological robustness does not guarantee similarity in terms of common links (and nodes) across consecutive time windows.

- Direct animal exchanges mostly take place within local distances, preferably within the same commune.
- Temporal analysis of the cattle movement network allows identifying the most vulnerable time windows (with largest possible epidemic size.)
- The temporal measure of estimating the largest epidemic sizes (reachability ratio) shows that a large chunk of the holdings in the cattle movement network in Brittany and Limousin were not accessible through exchange of animals during the period 2005-2009.

Chapter 3

Spread of Bovine Viral Diarrhoea Virus (BVDV) within a cattle herd: focus on dairy herd

In this chapter dynamics of BVDV within a herd is explored. Literatures show that within-herd BVDV dynamics is very much dependent on the farming methods/herd structure/management. Therefore, the dynamics need to be modelled differently for beef and dairy herds, as their management and farming are fundamentally different. We analyse BVDV spread within a structured dairy herd which throws light onto the system as close as possible. Out of the different models present, we choose a model already developed in-house [Ezanno et al., 2007], as we restrict our focus to dairy herds. This model corroborates very well with field observations. Optimisation of the model using C++ as the programming language for performance enhancement in terms of computational resources, assumptions added in or removed from the model or any other modifications during optimisation is presented. The outputs of the original model with the optimised model are compared. We explore the key advantages (along with the limits) and discuss its implementation in designing a region wide inter-herd model of BVDV spread.

3.1 Introduction to BVDV

Bovine Viral Diarrhoea (BVD) is caused by a Pestivirus belonging to the Flaviviridae family, called the Bovine Viral Diarrhoea Virus (BVDV). At the herd level, BVDV infections typically result in an increased incidence of reproductive disorders affecting conception rates, gestation rates, abortion rates and time to first calving [Baker, 1987, Carlsson et al., 1989, Lindberg, 2003, Muñoz-Zanzi et al., 2004]. It also results in impaired calf health and increase in calf mortality [Houe and Meyling, 1991, Sivula et al., 1996, Wittum et al., 2001, Ersbøll et al., 2003, Svensson et al., 2006]. Presence of BVDV in dairy herds increases the risk of infectious diseases [Chase, 2013]. BVDV infections reduce milk production incurring significant financial losses [Houe, 2003, Fourichon et al., 2005, Heuer et al., 2007]. In beef herds, growth retardation, immuno-suppression and resulting increase in other infectious diseases add to production losses [Booker et al., 2008].

There are two classes of infected animals in the case of BVD, persistently infected (PI) animals and acute or transiently infected animals (TI). PIs are the natural reservoirs for BVDV and shed the virus in large amounts in their secretions and excretions throughout the life, whereas an acutely infected animal shed the virus in lesser amounts, that too for a limited infectious period, ranging from few-days to around two weeks [Brownlie et al., 1987, Tremblay, 1996, Houe, 1999, Lindberg, 2003]. An animal surviving the acute infection is considered immune (R) to BVDV for the rest of its life [Brownlie et al., 1987].

Persistent infections occur strictly through vertical transmissions. A PI animal may appear in a closed herd in two possible ways, (i) transmission of BVDV from a PI cow to its foetus, or (ii) acute infection of a pregnant cow during *mid gestation* (7-22 weeks after conception) passing the virus to the foetus, if not aborted due to infection [Brownlie et al., 1987, Houe, 1999, Fray et al., 2000, Lindberg, 2003]. The rate of appearance of PI animals in a herd is therefore proportional to the number of PI cows present in the herd and the fraction of the pregnant cows suffering from transient infection during mid gestation. Cows infected subclinically during gestation recover after the infectious period [Lindberg and Alenius, 1999, Lindberg and Houe, 2005]. Transient infections result from horizontal transmission. BVDV can be transmitted both by direct *nose-to-nose* contact between an infected animal (PI or TI) and a susceptible animal [Houe, 1999], or by indirect transmission though equipments-tools-persons dealing with an infected animal [Niskanen and Lindberg, 2003] or even through the air on a few meters [Mars et al., 1999]. The most efficient way of acquiring an infection is by direct contact with an animal shedding BVDV [Houe, 1999]. PI bulls shed BVDV in their semen with the risk of transmission to heifers and cows venereally, via artificial insemination [Meyling and Mikel Jensen, 1988] or during embryo transfer [Brock et al., 1991]. BVDV may also spread through indirect means such as re-use of needles, nose tongs and rectal gloves [Niskanen and Lindberg, 2003]. There are reports of spreading of BVDV by blood feeding flies also [Tarry et al., 1991]. However, generally, the impact of such indirect transmissions may be considered to be negligible [Tremblay, 1996].

The clinical outcomes of BVD may vary from subclinical inapparent situations to severe clinical symptoms. Acute infections are generally not fatal. Some of the signs of transient infection are fever, lethargy, anorexia, nasal and ocular secretions and oral abrasion [Brownlie, 1985, Brownlie et al., 1987, Houe, 1999]. These signs may be accompanied by diarrhoea [Brownlie, 1985, Brownlie et al., 1987]. Significant drop in milk production may occur as a result of BVDV infection in cows [Houe, 2003, Fourichon et al., 2005]. Persistently infected animals may develop signs of fatal mucosal disease [Brownlie, 1985]. Most PI animals die prematurely, on average half of the total PIs born in a herd die within the first year of life [Baker, 1987, Houe, 1993], but some may also live for several years showing no symptoms [Houe, 1993, Houe, 1995].

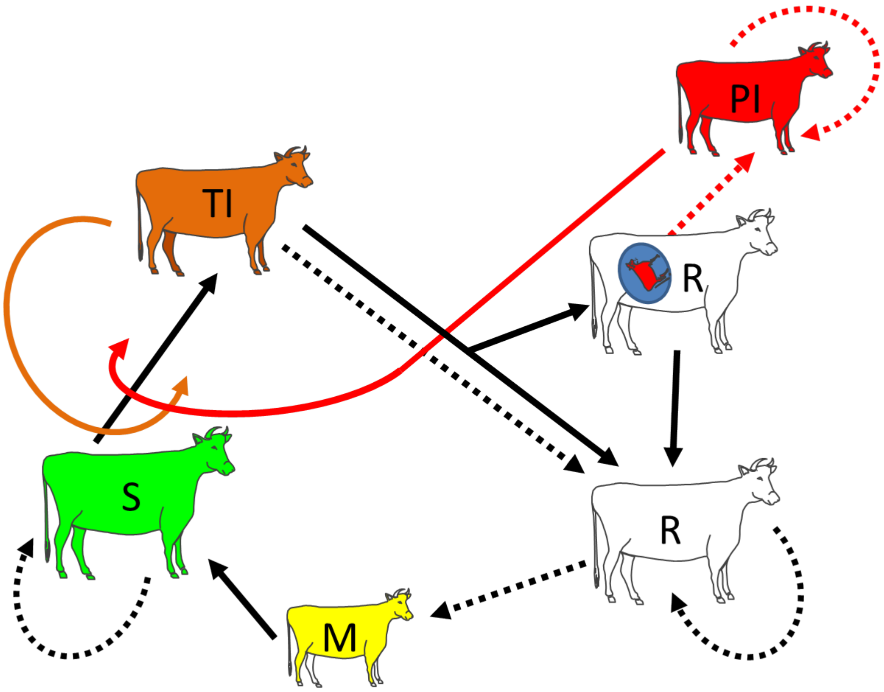


Figure 3.1: Health states of cattle when BVDV spreads and transitions between states. Here, the solid black lines represent change in epidemiological state, the dotted lines represent calving (in red, the birth of PI calves) and the coloured solid lines represent the shedding of the virus by infected animals leading to infection.

Therefore, in terms of epidemiology of BVD, a cattle can belong to one of the four epidemiological classes, susceptible (S), transiently infected (TI), recovered and immune (R) and persistently infected (PI) [Tremblay, 1996,Houe, 1999,Lindberg and Houe, 2005]. Calves born to immune dams are protected by maternal antibodies and considered separately (class M) [Hartley and Richards, 1988,Fray et al., 2000]. They remain protected from BVDV for a few months and then become susceptible. An S animal coming to direct (or indirect) contact with a PI or TI animal becomes TI and then becomes R after recovery [Lindberg and Houe, 2005]. If infection occurs to a pregnant animal, there are three possibilities with different probabilities depending on the stage of gestation. In the first stage of gestation, infections most often (very likely) result in abortion and the dam goes to the R class after recovery. If infected in mid-gestation, then the dam may abort or continue its gestation until calving, giving birth to a PI calf (most probable) [McClurkin et al., 1984,Fray et al., 2000,Lindberg, 2003]. It may give birth to a calf in either M or R state also, but the chances are very little. Lastly, if the infection takes place towards the end of the gestation, the calf is most likely to be immune at birth (R) [Hartley and Richards, 1988,Fray et al., 2000]. The transmission of BVDV in a cattle herd in terms of the epidemiological classes are shown in Fig. 3.1. An outbreak of acute, possibly sub-clinical, BVD in pregnant cattle can later result in a "batch" of PI calves, which poses the greatest risk of possible socio-economic losses and long term prevalence of BVD in the herd, also a risk to any other herd in contact with [Lindberg and Alenius, 1999, Lindberg and Houe, 2005]. The complexity with the presence of two types of infected animals and two ways of transmission of the virus, along with dependence on physiological status (pregnant or not, if yes, what stage) puts BVDV transmission in the genre of complex dynamical systems, a suitable candidate from the modelling and simulation perspective [Viet et al., 2007, Ezanno et al., 2007].

3.2 Modelling BVDV spread in a cattle herd

There have been several models developed to model BVDV spread within a cattle herd. The general aim of these models is to study the effect and efficiency of control measures. They handle both herd dynamics and the corresponding BVDV propagation in the herd. Since herd dynamics is dependent on herd types, the models had to adopt themselves upto the system they were addressing. The models addressing dairy herds are [Pasman et al., 1994, Sørensen et al., 1995, Innocent et al., 1997, Cherry et al., 1998, Viet et al., 2004, Ezanno et al., 2007]. A summarised comparison of these models in terms of modelling options and simulation procedure are presented in Table 3.1 derived form [Viet et al., 2007] and [Ezanno et al., 2007]. There are lots of differences and discrepancies among these existing BVDV models. The main are in the way herd structure was considered, not all of them considering a heterogeneous herd structure explicitly. [Pasman et al., 1994] considered 3 groups within a herd based on age (< 1 year, between 1 and 2 years, > 2 years), whereas [Sørensen et al., 1995] and [Cherry et al., 1998] considered no group structure within a herd at all. [Innocent et al., 1997] considered 3 groups as cows, heifers and calves. [Viet et al., 2004] is an individual based model, where, although the herd structure is inherent, there were 4 groups, cows, heifers ready for breeding, heifers before breeding and calves. [Ezanno et al., 2007] is a compartmental counterpart of [Viet et al., 2004], where a herd is grouped into the following 5 groups, calves, young heifers, older heifers, and cows, lactating and dry.

Herd structure and between group transmission have been shown to significantly influence BVDV spread in a dairy herd [Viet et al., 2004, Ezanno et al., 2008]. Older models do not capture this significant effect. The infection process of [Ezanno et al., 2007] is based on [Viet et al., 2004], which

	[Ezanno et al., 2007]	[Viet et al., 2004]	[Cherry et al., 1998]	[Innocent et al., 1997]	[Sørensen et al., 1995]	[Pasman et al., 1994]
Effect of chance	Stochastic	Stochastic	Deterministic	Stochastic	Stochastic	Deterministic
Treatment of Animals (Variable type)	Discrete (Number)	Discrete (Number)	Continuous (Density)	Discrete (Number)	Discrete (Number)	Continuous (Number)
Time variable type	Discrete (2 weeks)	Event driven	Continuous	Discrete (1 month)	Discrete (1 week)	Discrete (3 months)
Time dependence of transition	Constant transition rates	Time spent	Independent	Time spent	Time spent	Independent
Herd heterogeneity	Y	Inherent	N	Inherent	Inherent	N (3 age classes)

Table 3.1: Comparison of six models of BVDV spread in a dairy cattle herd based on modelling options and simulation procedures (adaptation following [Viet et al., 2007] and [Ezanno et al., 2007]).

had previously been qualitatively validated [Viet et al., 2006]. Results of these two formalisms have a large overlapping between them and differences ranged within an acceptable range [Ezanno et al., 2007]. The compartmental model has comparatively lesser number of variables (523) and parameters (20), whereas the complexity of the individual based model is much higher (number of variables $\propto$ number of animals $\times$ number of all possible states). Moreover, from the point of view of computational resources (processing time and memory requirements) [Ezanno et al., 2007] is far less expensive compared to [Viet et al., 2004]. So far, the model [Ezanno et al., 2007] constitutes the best current option to model BVDV spread in a dairy herd and to be used as a building block in a metapopulation of interacting herds. Therefore, we decided to use this model as the herd-level building block.

3.3 An update of the model from Ezanno et al. 2007

The within-herd BVDV transmission model is a compartmental model, where time is discrete and events are stochastic with constant transition rates among compartments. The time step of simulating the model is 14 days. Clinical observations show that on average recovery of a transiently infected animals is around 2 weeks. Such a time step is therefore to include the TIs into the model [Brownlie et al., 1987, Houe, 1999]. The model couples a model of the herd dynamics with an epidemiological model, and is described in the following subsections.

3.3.1 Herd dynamics

A typical dairy herd is divided into five exclusive groups based on age and physiological statuses: (i) calves, c [with two subgroups: males ($0 \leq$ age < 2 weeks) and females (2 weeks $<$ age $\leq$ 12 weeks)]; (ii) young heifers, h (12 weeks $<$ age $\leq$ 72 weeks); (iii) bred heifers, H(72 weeks $<$ age $\leq$ 112 weeks); (iv) lactating cows, LC (for a duration of 38 weeks after calving); and (v) dry cows, DC (last 10 weeks of gestation of a cow). There is no male group beyond the age of 2 weeks as typically all male calves are removed from dairy herds by the age of 2 weeks. Each of these groups is divided into intervals of 2 week duration. The choice of a constant duration in each of the physiological stages/groups was chosen over an exponential distribution of their duration because in a multigroup livestock population with two transmission levels, spreading via the low transmission path between groups is not well estimated with an exponential distribution [Ezanno et al., 2007,Viet and Jacob, 2008]. Therefore, the number of compartments in different physiological intervals are as following: calves: 7 intervals (1 for male and 6 for females); small heifers: 30 intervals; bred heifers: 22 intervals; lactating cows: 19 intervals and dry cows: 5 intervals. There are additional compartments for the bred heifers (2 intervals) and lactating cows (1 interval) if the first insemination fails (4 weeks after calving), till the next menstrual cycle and their next insemination. In case of abortions, the waiting time for next insemination for bred heifers and cows are 12 and 4 intervals, respectively. The ageing process of the animals is a continuous (deterministic) incremental process. All animals at age class a at time t moves to the next age class $a+1$ at time $t+1$, except (stochastic) outflows from the herd (due to death or replacement of breeding stock). The replacement of the breeding stock is performed from a uniform distribution of the animals in the eligible classes.

3.3.2 Infection dynamics

The epidemiological classes are the ones described in section 3.1: M: calves with maternal immunity, S: susceptible (not infected but without maternal immunity), TI: transiently infected (shades low amount of virus), R: recovered (immune for the rest of the life), and PI: persistently infected (shades high amount of virus). Cows and heifers which had been infected in early ($0 - 41$ days

Parameters	Value	Definition	References
s_r	0.50	Sex ratio	[a]
r_H	0.62	Probability of success in heifer insemination (group H)	[a]
r_{LC}	0.46	Probability of success in cow insemination (group LC)	[a]
$z_{g,I}$		Probability of being sold or culled in group g and physiological interval i [a]	
z_H	0.11	Global probability for heifers (group H)	
z_{LC}	0.25	Global probability for cows (group LC)	
$m_{g,X,i}$		Probability of mortality in group g, infectious state X and physiological interval i	[a]
$m_{g,P,i}$	$m_{g,X,i} + 0.026$	Mortality of P animals in group g and physiological interval i	[Baker, 1987, Houe, 1993]
$m_{c,P,0}$	$m_{c,X,0} + 0.01$	Mortality at birth of P animals	
$e_{g,X,i}$	$z_{g,i} + m_{g,X,i}$	Probability of exit from the herd in group g, infectious state X and physiological interval i	[a]
$e_{c,X,1,f}$	0.18	Exit of female calves (group c) 2 weeks after birth	[a]
$e_{c,X,1,m}$	1	Exit of male calves 2 weeks after birth	[a]
β_w^P	0.50	Within group transmission rate for P animals	[Moerman et al., 1993]
β_b^P	0.10	Between group transmission rate for P animals	[Niskanen and Lindberg, 2003, Mars et al., 1999]
β_w^T	0.03	Within group transmission rate for T animals	[Baker, 1987]
β_b^T	0	Between group transmission rate for T animals	
a_{R^a}	0.80	Abortion rate due to infection at early gestation	[Hartley and Richards, 1988]
a_{R^b}	0.25	Abortion due to infection in mid gestation	[Hartley and Richards, 1988]
z_{R^b}	0.80	Culling rate in case of abortion due to infection in mid-gestation	[a]
n_X		Probability of giving birth to a calf in state X if infection in mid gestation ($X = P, M$ R)	[Hartley and Richards, 1988, Fray et al., 2000]
n_P	0.934		
n_M	0.033		
n_R	0.033		

[a] Parameters for a typical Holstein dairy herd of medium size.

Table 3.2: Definition and value of parameters for BVDV model in a dairy herd (From [Ezanno et al., 2007]).

of conception), mid- ($42 - 150$ days) and late (> 150 days) gestation (and do not abort due to infection) enter R^a, R^b, and R^c classes to keep track of vertical transmission and its consequences for calves to be born. If aborted, they join the R class and wait for next insemination.

A representation of the within dairy herd BVDV model is shown in Fig. 3.2. All epidemiological transitions in the model are stochastic. The transition rates are constant, but their values at different time points are dependent on the sizes of the managed classes of animals during that point of time. The epidemiological transitions take place following binomial distributions. The probability of infections is calculated considering a Poisson process.

The horizontal infection, $S \rightarrow T$ transition, depends on the proportions of shedding animals

(T, P) belonging to the different groups. The transmission rate for group g at time t, $p_{\text{inf}}(g,t)$ is given by:

$$p_{\text{inf}}(g,t) = \beta_w^P \frac{P_g(t)}{N_g(t)} + \beta_w^T \frac{T_g(t)}{N_g(t)} + \sum_{a \neq g} \beta_b^P \frac{P_a(t)}{N_a(t)\ N_g(t)} + \sum_{a \neq g} \beta_b^T \frac{T_a(t)}{N_a(t)\ N_g(t)} \tag{3.1}$$

The probability of infection:

$$\text{prob}_{\text{inf}}(g,t) = 1 - e^{-\Delta p_{\text{inf}}(g,t)} \tag{3.2}$$

The number of newly infected animals in physiological class i in group g at time t is:

$$I_{g,i}(t) = \text{Bin}\left(S_{g,i}\left(t\right); \text{prob}_{\text{inf}}(g,t)\right) \tag{3.3}$$

and which are in state T at time $(t+1)$:

$$T_{g,i}(t+1) = \text{Bin}\left(I_{g,i-1}(t); 1 - e_{g,T,i}\right) \tag{3.4}$$

where a denotes for all groups other than group g; β_w^X and β_b^X are within- and between-group transmission rates per day for animals in infectious state X (P or T; Table 3.2); $P_g(t), T_g(t),$ and $N_g(t)$ are the number of PI animals, TI animals and total number of animals in group g at time t, respectively. $X_{g,i}(t)$ represents the number of animals in state X, physiological interval i and group g at time t; Δ is the length of the time interval in days (= 14 days); and $e_{g,T,i}$ is the probability of exit from the herd for transiently infected animals in group g and physiological interval i.

The parameters of the model are shown in Table 3.2. The infectious period of TI animals was consistently of 1 time interval (2 weeks). Within group infection rates are higher than between two different groups, primarily due to possible direct transmission among animals in the same group. If the groups share common boundaries either in house or on pasture, there may be a possibility of direct between group transmission. In a dairy herd, the group structure is quite strongly defined/maintained and generally animals belonging to two different groups hardly come to direct contact. Therefore the paths of transmission of BVDV in a dairy herd are of the following types: direct transmission within a group, indirect transmission between groups via other means, venereal transmission and rarely by direct transmission between groups. Susceptible gestating cows infected during gestation are identified at recovery and go from state T to states R^a, R^b and R^c

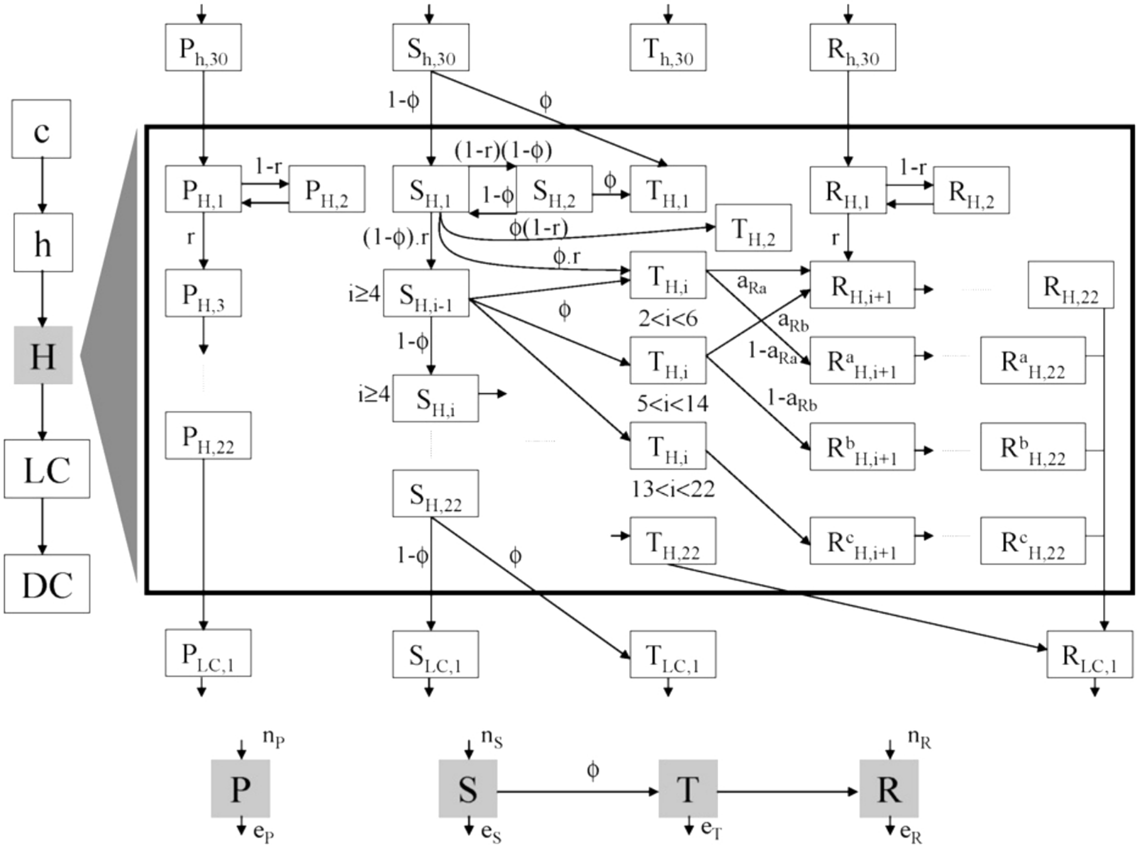


Figure 3.2: Representation of the model (c: calves; h: young heifers; H: older heifers; LC: lactating cows; DC: dry cows): example for group H (S: susceptible; T: transiently infected; P: persistently infected; R: immune; R^a, R^b, R^c: immune which have been infected in early, mid- and late gestation, respectively). The protection due to maternal antibodies (state M) ends before entering group H and hence is not represented here. The parameters of the between-interval transitions are: r the probability of success in heifer insemination (r_H in Table 3.2), ϕ the probability of infection in susceptible heifers ($\text{prob}_{\text{inf}}(H,t)$ in Eq. 3.2, and a_{R^x} the abortion rate (see Table 3.2 for definitions). (Reproduced from [Ezanno et al., 2007])

depending on the stage of gestation. These states remain same till calving, after which cows were counted as simply R. The necessity of these three 'special' states for pregnant animals comes from the fact that depending on the stage of gestation at the time of infection, the outcomes are many. Calves born may be of several infectious states: M if infection occurs in early gestation, P, M, or R if infection occurs in mid-gestation (with probabilities n_P, n_M, n_R respectively) and with R state if infection occurs in late pregnancy. The number of calves born per state from heifers and cows in state Rb is given by the following multinomial distribution:

$$[b_{R^b}(t,1); b_{R^b}(t,2); b_{R^b}(t,3)] = Mult\left(R^b_{H,22}(t) + R^b_{DC,5}(t); (n_P, n_M, n_R)\right) \tag{3.5}$$

with $R^b_{H,22} \equiv R^b$ heifers in their last interval of gestation; $R^b_{DC,5} \equiv R^b$ (dry) cows in the last interval of gestation (5^{th}). PI females give birth to PI calves only. Therefore, the number of calves born at time t per infection state is:

$$P_{c,0}(t+1) = Bin\left(P_{H,22}(t) + P_{DC,5}(t) + b_{R^b}(t,1); 1 - m_{c,P,0}\right) \tag{3.6}$$

$$M_{c,0}(t+1) = R^a_{H,22}(t) + R^a_{DC,5}(t) + b_{R^b}(t,2) + R_{H,22}(t) + R_{DC,5}(t) \tag{3.7}$$

$$S_{c,0}(t+1) = S_{H,22}(t) + S_{DC,5}(t) \tag{3.8}$$

$$R_{c,0}(t+1) = R^c_{H,22}(t) + R^c_{DC,5}(t) + b_{R^b}(t,3) \tag{3.9}$$

where $m_{c,P,0}$ is mortality at birth of PI calves [Baker, 1987, Houe, 1993, Ezanno et al., 2007].

3.3.3 Coding, modifications and optimization

The within-herd model of [Ezanno et al., 2007] was originally implemented in SciLab. Since the goal was to use the within-herd model as a building block of a metapopulation of interacting herds (Chapter 4), the model has been recoded in C++ (using C++11 standard) [Stroustrup, 2013] as this language is recognised to produce highly efficient programs. The code contains purely C++11 standard libraries in GNU Compiler Collection(gcc-4.8.0), no proprietary or external (non-GNU) templates or libraries were used. The herd at a particular time point is represented by a vector array, and vector arrays are passed with reference between objects and functions. The recoding enabled not only portability across platforms, but also speed up the processing time more than ten fold.

In SciLab to C++ translations, no modifications were performed into the existing model. The model in C++ is an *'as it is'* translation of the SciLab program. One dimensional arrays in SciLab are translated as vector data types, whereas for multi-dimensional arrays (matrices) map template had been used.

For analysis and plotting C++ and R are integrated using RCPP package in R (3.x.x) [R Core

Team, 2014, Eddelbuettel, 2013]. It is possible to directly call the C++ object from an R interface, which enables doing the statistical analysis of the outputs in one-go, using the speed of C++ and specialised analyses powers of R.

3.3.4 Initial conditions and scenarios evaluated

A typical average Holstein dairy herd, following [Ezanno et al., 2007, Ezanno et al., 2008] of herd size 85 (45 cows, 20 heifers and rest are young stocks) was the base initial condition. The initial distribution of animals in different groups divided into physiological classes of two weeks was the following: 6 female calves, one each in every class; 14 young heifers, one each in randomly chosen 14 compartments out of 30 available compartments; 20 bred heifers, one each in every physiological class for the bred heifer group, and rest 45 cows randomly distributed among the cow (lactating and dry) group compartments, not allowing more than two animals in a particular class within this group. Two larger herd sizes in multiples of the base herd were also considered (with 170 and 340 animals in the herd). The internal herd structure of these herds was kept the same, and the number of animals in each class was proportional to the herd size. All the scenarios were ran for 10 years of disease free condition and then an infected animal was introduced into that naive herd.

Each naive herd was introduced with a single infected animal from different physiological classes at $t = 0$ and simulated. Twelve different modes of introducing an infection were studied, 8 of which were introduction of a PI animal, namely: female calf (PiC), small heifer (Pih), bred heifer (PiH), gestating heifer (PiH_Gesta), lactating cow (PiL), gestating cow (PiL_Gesta), immune gestating animal with PI foetus: heifer ($Rb2H$) and cow ($Rb2L$); and 4 of introducing a TI animal, calves, small heifer, bred heifers and lactating cows (TiC, Tih, TiH, TiL respectively). No re-infection was allowed. In the larger herds, apart from the introduction of a single infected animal, an additional scenario was studied, where the number of initial infections was proportional to herd size: if the herd size was n-times the base herd size, then we introduced n infected animals into the herd at $t = 0$. This was to observe if there was a relation between herd size and the number of infections entering a herd in terms of disease dynamics. If the relationship was linear, we expected to see normalised graphs collapsing on each other. The parameters of the model were kept constant throughout the simulations (Table 3.2).

3.3.5 Model outputs and simulations

The probability of virus persistence and epidemic size are outputs that concerns all replicates of each scenario / simulation. The *persistence probability* is defined as the fraction of the replicates of the simulation where the herd is still infected at a given time t, i.e. at least one P, T or immune dam carrying a PI foetus exists in the herd. Epidemic size is the cumulative sum of all the infections (P or T) taking place since $t = 0$ till end of simulation, i.e., 10 years after initial virus introduction. This procedure was repeated for 200 repetitions, if not mentioned otherwise, for each scenario.

In addition, the model predicts the number of P, T and immune dams carrying a PI foetus, in an infected herd over time. Average, variance, percentiles are calculated over the replications concerned (with persistence at a given time).

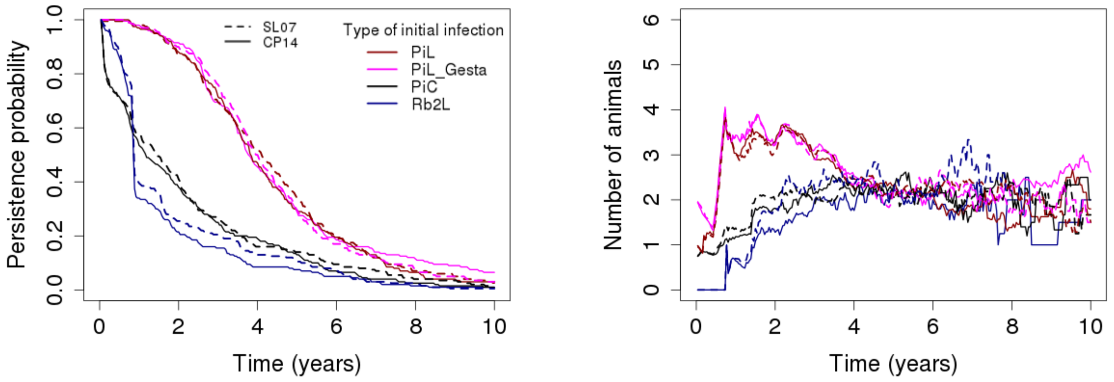
3.3.6 Model comparison

Outputs obtained with the C++ version of the model ($CP14$) were compared with the ones obtained by [Ezanno et al., 2007] ($SL07$) for a typical dairy herd over a period of 10 years and showed them for four different modes of introducing the infection into the herd: a PI female calf, a PI lactating cow not pregnant, a PI lactating cow gestating and carrying a PI foetus and an immune cow carrying a PI foetus, respectively. Although we compared the outputs for all the 12 modes of introduction as described in the previous section, we showed only these four for brevity. Model outputs considered for the purpose are persistence probability, and average numbers of infected animals (PI, TI and recovered pregnant dam carrying a PI foetus in an infected herd. We restricted ourself to a visual comparison between the model outputs as model $SL07$ had already been compared with a qualitatively validated model [Viet et al., 2006] in [Ezanno et al., 2007, Ezanno et al., 2008].

3.4 Results

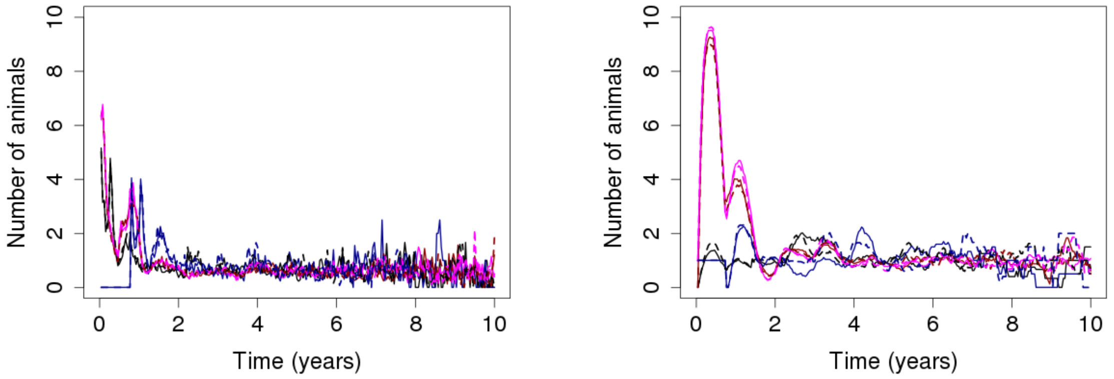
3.4.1 Model comparison

Visual comparison of the model outputs indicated that the recoded version was in very good agreement with the model described in [Ezanno et al., 2007] (Fig. 3.3). Apart from minor fluctuations between the graphs representing the same output from $SL07$ and $CP14$, overall the qualitative overlapping was large enough. The minor fluctuations observed between them were attributed to



(a) Persistence probability

(b) Average number of PI animals



(c) Average number of TI animals

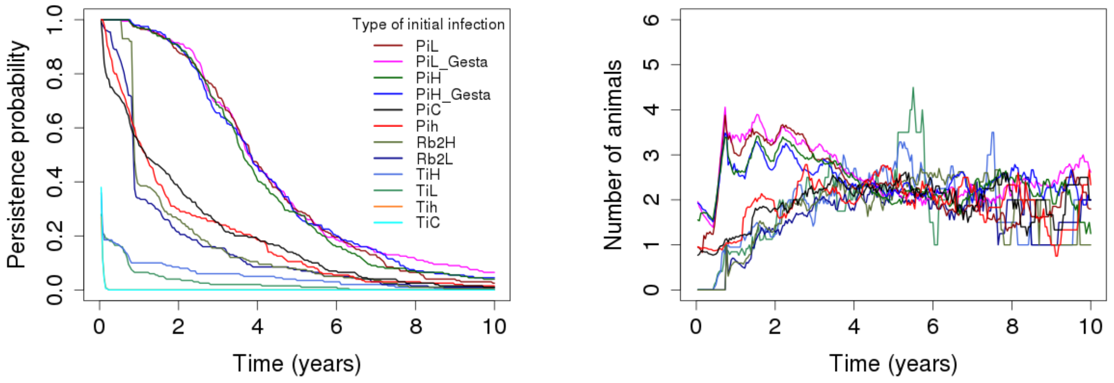
(d) Average number of PI foetus carried by recovered dam

Figure 3.3: Comparison of model outputs between "SL07" ([Ezanno et al., 2007]) and "CP14" (version used in this work) for different types of initial infection introduction in a typical Holstein dairy herd of medium size.

the stochastic nature of the model.

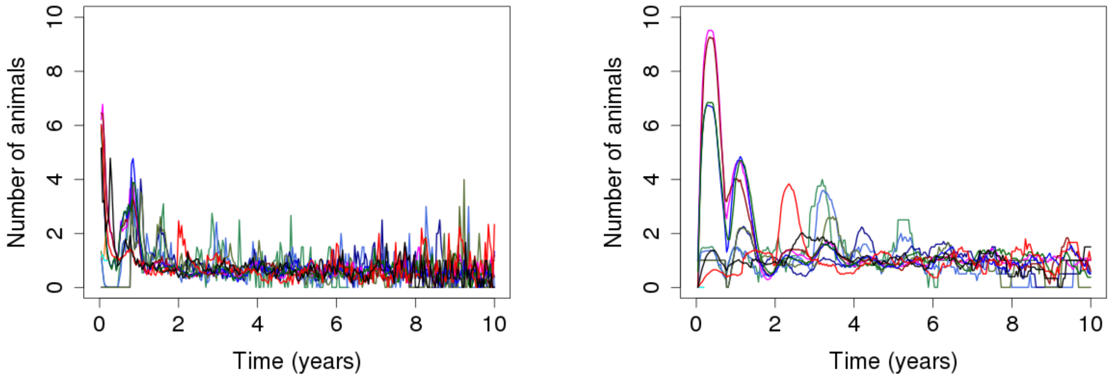
3.4.2 Effect of the type of initial virus introduction in the herd

The type of virus introduction in a naive dairy herd had a significant effect on BVDV persistence [Fig. 3.4 (a)] and on the number of infected animals, at least during the first three years following the virus introduction [Fig. 3.4(b-d)]. Introducing TI animals had barely any effect, especially if they were young TIs (calves or young heifers). After 2 years, BVDV was still present in less than 10% of the runs concerned by such an introduction. Introducing a PI heifer or a PI cow (gestating or not) resulted in the largest persistence and the highest number of infected animals. 2 years after



(a) Persistence probability

(b) Average number of PI animals)

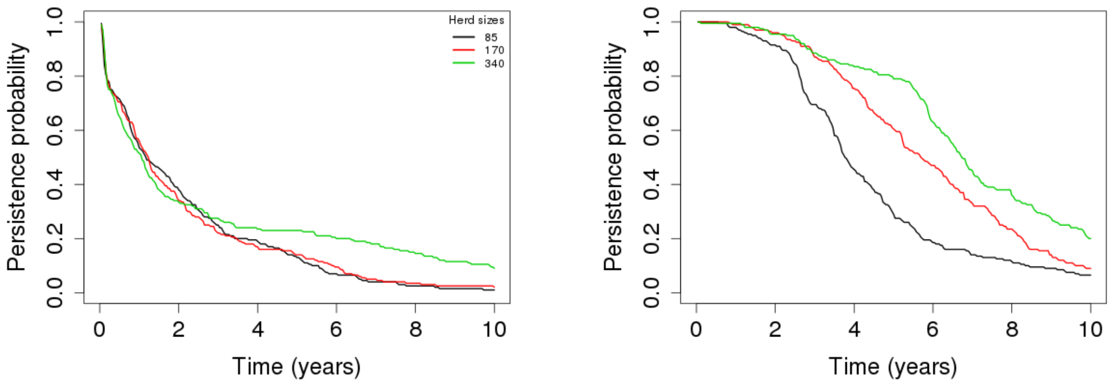


(c) Average number of TI animals

(d) Average number of PI foetuses)

Figure 3.4: Disease dynamics for different types of virus introduction in a typical Holstein dairy herd of medium size

such an introduction, BVDV was still present in more than 80% of the runs. The number of PIs was twice as high in infected herds as for other types of virus introduction. The number of immune dams carrying PI foetuses was also much higher, especially in the first year of herd infection. Four years after BVDV introduction, the number of infected animals in an infected herd reached the same level irrespective of the type of initial virus introduction. On the contrary, the effect of virus introduction on the probability of virus persistence was still visible after more than 6 years, with persistence in around 20% of the runs where PI adults had been introduced against less than 10% of persistence in the other cases. This implies that as long as PI animals remain undetected, and especially as adults, BVDV clearance will hardly occur.



(a) Initial infection: 1 female *PI* calf. (b) Initial infection: 1 gestating *PI* cow

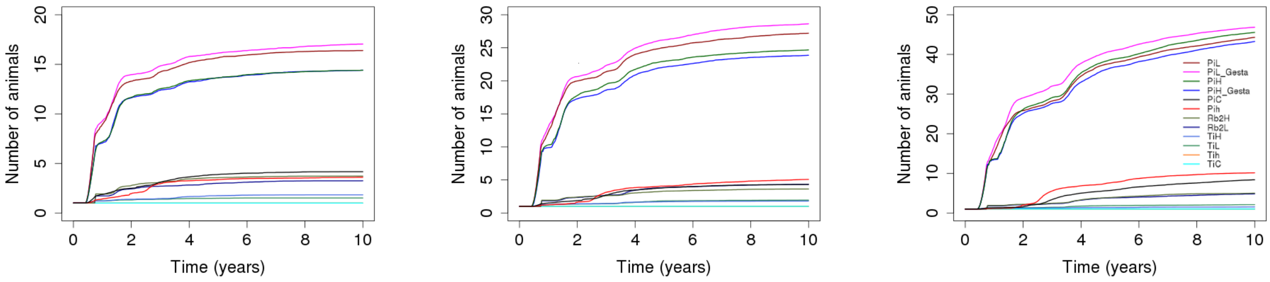
Figure 3.5: Disease persistence for different herd sizes and types of virus introduction.

3.4.3 Effect of herd size on BVDV spread and persistence

Herd size largely impacted BVDV persistence, the virus persisting longer in larger herds [Fig. 3.5], especially once the groups of adults (heifers and cows) were affected. As a result, the difference in persistence was visible only 4 years after virus introduction when a *PI* calf was introduced [Fig. 3.5 (a)] whereas it was sooner when a gestating *PI* cow was introduced [Fig. 3.5 (b)] with in addition a much larger effect. Six years after the introduction of a *PI* calf, BVDV was still present in 10% of the runs for medium herd sizes (85 and 170) vs 20% of the runs for larger herds (340). Six years after the introduction of a gestating *PI* cow, persistence occurred in around 20%, 50% and 65% of the runs in herd sizes 85, 170 and 340, respectively.

Concerning the effect of herd size on epidemic size, as expected the larger the herd size was, the larger was the epidemic size [Fig. 3.6]. If a *PI* calf was initially introduced, *PI*s occurred later in larger herds but were more numerous cumulatively after 10 years, especially in herd size 340 [Fig. 3.6(c)]. When introducing a gestating *PI* cow, the impact of herd size on the cumulative number of *PI*s was immediate and seemed to be directly related to the difference of size.

To evaluate more precisely if the effect of herd size on epidemic size is only a linear effect, we plot the epidemic size normalised by the herd size in Fig. 3.7. When normalised, we observe that the proportional epidemic size was larger in smaller herds when equal number of initial infections were introduced (1 infected animal per herd, irrespective of herd size). This was particularly true for the types of introduction, namely the introduction of *PI* adults [Fig. 3.7 (c-f)] and less when



(a) Herd size 85 (b) Herd size 170 (c) Herd size 340

Figure 3.6: Cumulative number of PIs for different herd sizes and different types of virus introduction

introducing a PI calf or young heifer [Fig. 3.7 (a-b)]

To test a case where the number of initial virus introductions was proportional to herd size, we introduced a single infected animal to the base herd size (85 animals) and a number of introductions proportional to their size in larger herds (2 and 4 infected animals in herds of size 170 and 340, respectively). As mentioned previously, the herd structure was kept the same for all herd sizes. Epidemic sizes in this situation are shown in Fig. 3.8. Almost no effect of herd size can be observed on epidemic size except when introducing a PI young heifer or bred heifer. Therefore we can conclude that - when assuming no difference in the herd structure between herd sizes - the effect of herd size on BVDV spread is mainly a scale effect directly related to our assumption of a frequency-dependent transmission. However, in the specific case of the introduction of PI heifers, larger epidemic sizes can be expected in larger herd size, in addition to a larger persistence.

3.5 Discussions

3.5.1 Key advantages and limits

We have translated the intra-herd BVDV transmission model of [Ezanno et al., 2007] into a different coding environment. A comparison of the model outputs of the two versions of the model showed no significant change in model outputs, a test against any inadvertent error-omissions or issues related to use of programming libraries. This confirmed that we used the ‘same’ model.

We studied disease dynamics for different herd sizes and types of initial infection. We had noticed that there were significant differences in disease persistence probability on either situations.

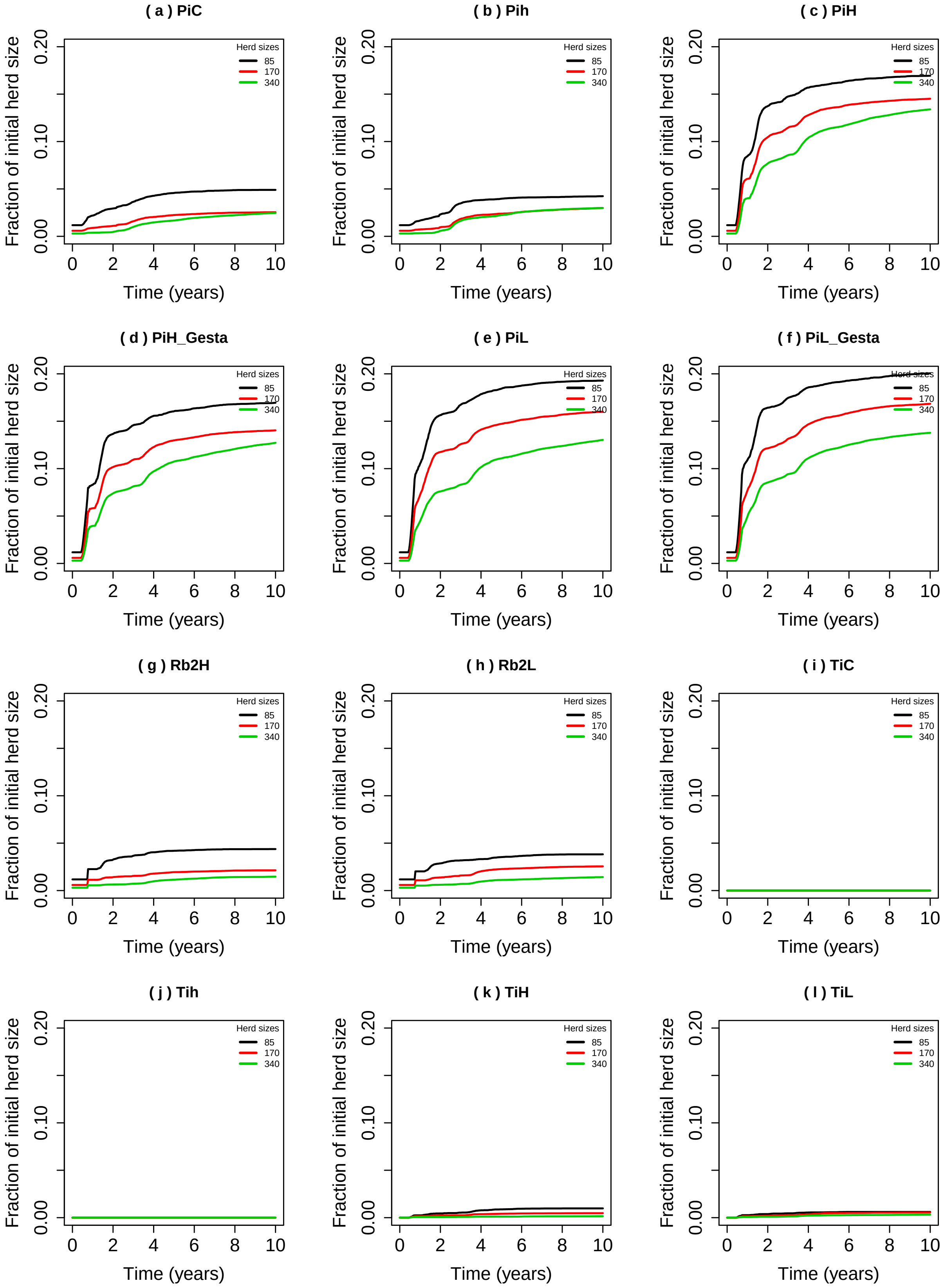


Figure 3.7: Normalised epidemic size in terms of cumulative number of PIs. Here a single infectious animal of a particular type was introduced into a naive herd, irrespective of the herd size.

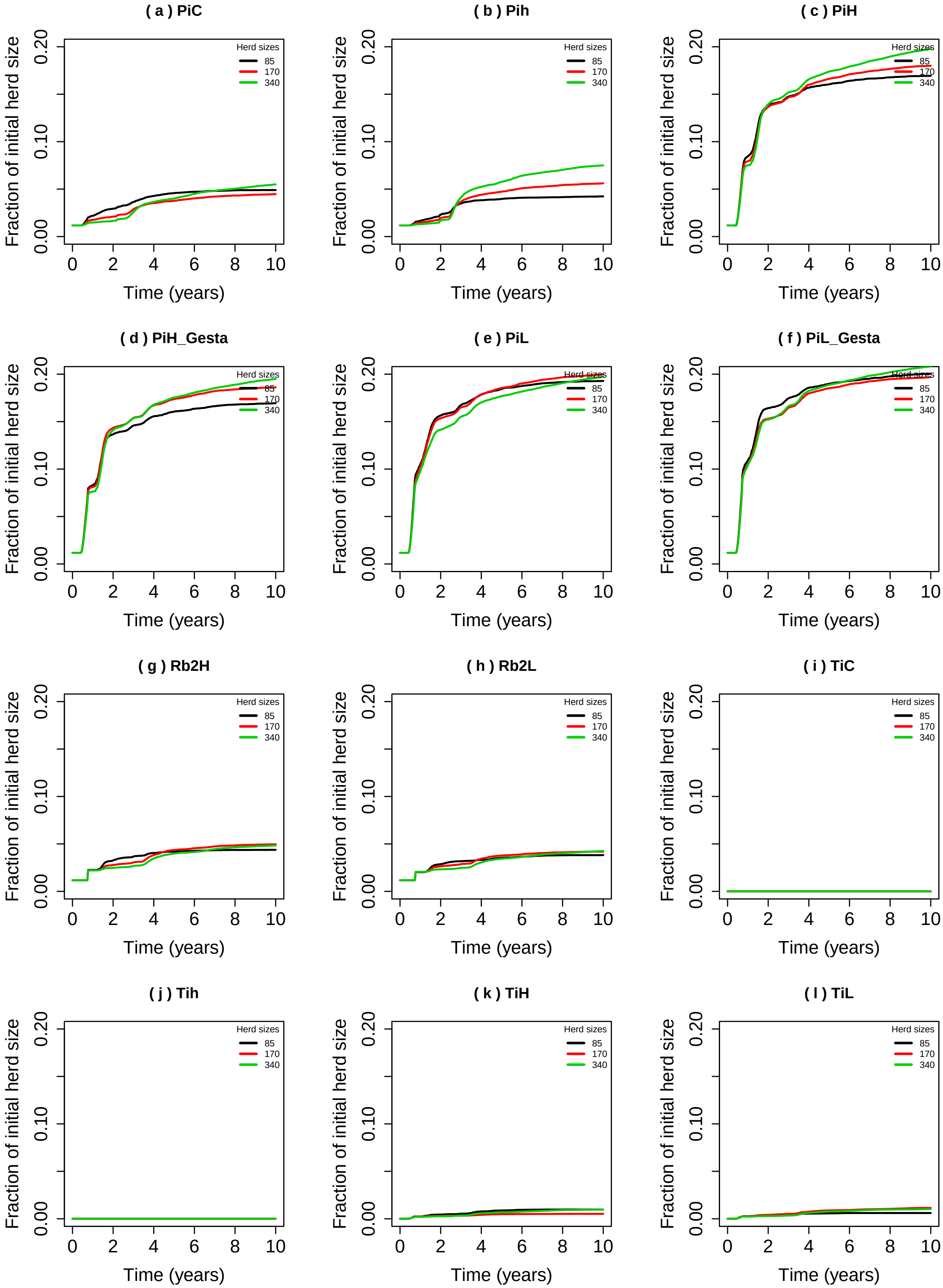


Figure 3.8: Normalised epidemic size in terms of cumulative number of PIs. Here the number of introductory infectious animal (n) was proportional to the herd size (herd size 85: $n = 1$, herd size 170: $n = 2$ and herd size 340: $n = 4$).

The disease persistence probability was found to be dependent not only on the herd-structure and management but also on the size of the herd, if same number (and type of animal) of initial infections enter the herd. Average numbers of infected animals (averaged over only infected herds) differed among herd sizes, as well as among different types of initial infections. An adult infectious animal entering the herd was found induce a larger persistence probability and epidemic size. Since we kept the herd-structure and parameters related to herd management practices the same throughout our study, we could not comment on the effect of these two prime factors upto this point.

Our model incorporates the most important key parameters involved in BVDV spread within a dairy herd [Ezanno et al., 2008, Ezanno et al., 2007, Viet et al., 2007]. Having 20 parameters representing both biological and herd management factors, it gives the free hand to test dependence of the dynamics on each of these parameters as done in [Ezanno et al., 2007]. The number of variables (523) and parameters (20) are not that high given the complexity of the dynamics of BVDV.

From the implementation point of view, the primary advantage of this model is that it is comparatively not hard to implement and computational resource requirements are not that expensive. In our case, simulation for 10 years (and 200 iterations) of a typical herd takes $\sim$ 20 seconds in an ubiquitous 2.8GHz Intel i7 processor, and no significant memory consumption was noticed in a 64Bit Linux platform. The SciLab version takes $\sim$ 3 minutes on the same configuration (insignificant memory consumption).

The most prominent limit of the model is that this model is so specific about dairy herd management that it is hard to extend it to a beef herd or a mixed herd (with both beef and dairy animals). The model in the present form does not consider seasonal adjustments in herd-management, like pasturing. It considers that the within and between group interactions are the same throughout the year. But consideration of these factors needs many locality specific extra parameters, which may force the model to loose its generality. Also there is no consideration of reinfection during renewal of breeding stock. These last points will be specifically addressed for the between herd model.

3.5.2 Implications for the regional scale model

At a regional scale, the situation extends to exchange of animals between farms and neighbouring relationships. At the herd level, the population and infection dynamics remain the same. Therefore, this intra herd model can be used to take care of this local dynamics. The key to this implementation is that at every time-step the population needs to be updated accounting for the number of animals bought or sold during that time-step. This can be done by encapsulating the within-herd dynamics by the exchange of animals. First, we should calculate the exchange of animals, update all the herds, and then update as-per local dynamics. The point to be kept in mind is that the physiological and epidemiological states of out-going and incoming animals should be kept track for successful design and implementation of regional scale model of BVDV spread. In its extension into intra-herd model, it should be taking care of re-infections during renewal of breeding stock and necessary locality specific parameters, which were some of the limitations in its within herd *avatar*.

Chapter 4

Modelling BVDV spread between dairy herds

In this chapter we model BVDV spread between dairy herds in a region. We start by introducing different paths of BVDV spread between herds and discuss what has already been done in this context and explore their pros and cons. Then we present our model for inter-herd transmission of BVDV. Assumptions, parameters and initial conditions chosen for the model are presented with justifications of their choices. Details of coding, implementation and incorporation of data in framing the model are discussed. Chosen model outputs are presented with justifications. Tackling heterogeneities in herd sizes, animal exchanges and neighbouring relationships are discussed. Then we proceed to present different scenarios studied. Next the results of simulations of the inter-herd BVDV spread model are presented. We present the results of explorations for different scenarios, relative contribution of different pathways of infection without control. After the results, we discuss the key results and their implications.

The work presented in this chapter is the subject of a paper in preparation for Veterinary Research.

4.1 Introduction

Cattle farms are not isolated entities, they interact with each other. The prime mode of interaction is buying and selling of animals between farms, either directly or indirectly via markets and assembling centres. Buying poses the risk of introducing a pathogen into a herd and selling poses the same for the destination, depending on the type and prevalence of the infection in the source herd. Animal exchange between herds is primarily necessitated by the requirement of constant renewal of the breeding stock to sustain productivity and herd size [Heinrichs, 1993, Groenendaal et al., 2004, Gates and Woolhouse, 2014]. A safer option from this point of view would have been to raise the heifers *locally*, but farms often prefer to buy the required replacement owing to the extra management and cost requirements to do so [Heinrichs, 1993, Groenendaal et al., 2004, Álvarez et al., 2011, Gates and Woolhouse, 2014] exposing themselves to the potential risk of introducing pathogens into their farms [Ezanno et al., 2006, Tildesley et al., 2011, Frössling et al., 2014]. Concerning dairy herds, a given number of heifers and cows are bought every year [Ezanno et al., 2006] as a part of the breeding stock management. The sources are usually considered to be random for each purchase as no significant purchase preferences were observed so far [Ezanno et al., 2006, Rautureau et al., 2011]. Cattle movement databases like the French National Cattle Database (FDCM) come into picture at this stage. These databases record all of the animal exchanges among holdings. They can be used to design statistical models of livestock movement networks [Ensoy et al., 2014] or networks constructed directly from the database are used to study the spread of pathogens in the network and designing control strategies based on network properties [Keeling and Eames, 2005, Kao et al., 2006, Kiss et al., 2006, Danon et al., 2011, Nöremark et al., 2011, Büttner et al., 2013]. Graph theoretic analyses of the network constructed using FDCM and qualitative characteristics of the database were presented in Chapter 2. The use of actual movement data includes local and regional connectivity among the farms (region specific nature of the cattle movement network was also presented in Chapter 2). It should be noted that using the movement information as a static network with frozen links in time may lead to non-realistic predictions of disease dynamics. Time-stamped (dynamical) form of the network is necessary to account for constraints and specificities arising from time-sequential contacts between farms [Nöremark et al., 2011, Frössling et al., 2012, Büttner et al., 2013, Konschake et al., 2013]. The use of real movement data not only keeps the time-sequential property of the network, but also keeps region specific local seasonal patterns and/or preferences traceable, which might be very handy in the design and implementation of region specific control measures. Although contact between farms through buying and selling had been a major pathway of transporting pathogens from one herd to another, it is not the only possible pathway. Some pathogens might be transported between herds also by live agents (human, other non-human animals, insects) as well as by an inert path (contaminated tools and equipments) or environmental routes (air, water, soil) [Keeling and Rohani, 2008]. In the case of pathogens spreading through contacts, a herd might get exposed to infection in any one of the following modes of contact between herds: an animal moved from one farm to another farm with no return to the source farm (*migration* or *dispersal*); an animal from a farm temporarily moves into another farm and returns back to the home farm in a short duration (*visits*); animals physically remain within their corresponding farm boundaries but might touch each other *over the fence* at the boundaries or at some other common place, such as a water hole (*neighbourhood*); or through indirect contact through live, inert or environmental agents as mentioned above.

In the study of spread and persistence of infectious pathogens in host populations, issues at multiple levels are involved. The questions to be understood are: the type of pathogen (source/origin), transmission routes (direct, indirect, horizontal/vertical etc.), dependence on host and environmental factors and possible control measures. A broader understanding demands these questions put together and investigated under various possible scenarios. In case of livestock populations, infectious disease propagation can be viewed as a complex biological system where structured and managed host populations localised in space interact at various levels: between individual animals within a group, between groups within a herd and between herds at regional level. These interactions may be perturbed by the farmers' decisions from time to time. Often information about many of the (biological) parameters and their degree of influence on the system are also not clearly understood. Given the complexity involved in such complex biological systems, modelling and simulation may be a suitable approach to investigate such systems. Modelling may be complementary to observations for such complex systems when either vital biological information about the systems is not sufficiently available or observational studies are difficult to perform (may be due to socio-economic, logistic or time constraints or lack of reference scenarios). Since interactions leading to possible pathogen spread take place at various levels, the models may need to have multiple layers to account for these multiple interactions taking place at different scales, depending on the context and goals of the investigation. Modelling coupling interactions at multiple scales may be either a *top-down* or a *bottom-up* approach depending on the system involved, choice of situation and objectives at hand [Meier-Schellersheim et al., 2009, Qu et al., 2011]. A top-down approach starts with features on a higher level of a system and then attempts to interpolate into mechanisms at lower fundamental scales. The starting model in a top-down approach is directly backed by data and it has the advantage of sequentially adding levels of details. Possible emergence of ambiguity in increasing details at lower scales as a higher scale phenomena may be a manifestation of multiple phenomena underlying at the lower scales. On the contrary, the bottom-up approach starts with details of the very individual components level of a system. Higher level behaviour is deduced from the dynamic interactions of these fundamental components spanning over space and time. In this approach, the type of collective behaviour responsible for a particular phenomena at the higher level can be identified [Meier-Schellersheim et al., 2009].

Given the spatial separation between the farms and different modes of interaction in multiple levels, a *metapopulation* framework for modelling the inter-herd spread of pathogens is a suitable approach in the line of interacting subpopulations in a patchy ecological environment [Grenfell and Harwood, 1997, Keeling and Rohani, 2008, Ball et al., 2014, Brooks-Pollock et al., 2014]. Farms constitute the subpopulations in this approach. Migration of infected animals or transport of the pathogen via indirect paths are the sources of infection for the farms. With this framework, the persistence of both hosts and pathogens in the metapopulation can be studied. In case of BVDV spread in a region, the contact routes between the farms of a region might be classified into two classes: *movements* (migration and visits) and *neighbourhood*. The contribution of the indirect paths were generally considered negligible [Tremblay, 1996]. At a regional scale, the herds form a metapopulation, a set of farms (local populations) being connected by movements. When farms do not share common boundaries, movement (of animals or involvement of indirect contacts) is the only mode of transport of BVDV between farms. If farms share common boundaries then neighbourhood also comes into picture. Therefore, to study the dynamics of BVDV spread in a region, apart from (local) spreading between groups within individual herds, contributions of migratory (non-local) and neighbouring (semi-local) interactions among herds are also to be accounted for. Considering the farm level dynamics as building blocks for a bottom-up approach, the spread of BVDV at a regional scale can be modelled. It is justified to assume that within-herd pathogen spread varies among infected herds. In a connected subpopulations scenario, an infected animal may quickly disappear from a subpopulation (mortality, recovery or migration) or persist for a long time (dynamic herd status). This heterogeneity in the within-herd force of infection may have a significant effect on the global spread and persistence of the pathogen at the metapopulation level (all subpopulations are at a heterogeneous risk). To account for this within-herd models need to be coupled with between-herd dynamics. In coupling these two levels of dynamics, multiple time scales may be involved, as generally time scales at the microscopic level (local herd dynamics, interaction of groups within a herd) and that at the macroscopic level (global herd interactions, exchange of animals) are different.

There had been a few models developed to study the spread of BVDV between cattle herds [Courcoul and Ezanno, 2010,Ersbøll et al., 2010,Tinsley et al., 2012]. The state of the art concerning BVDV spread between farms are discussed in more details in the following section. Briefly, some understanding of the main routes of transmission at a regional scale have been achieved. However, no mechanistic model is available to our knowledge on real trade movement data and herd localisation till date. Our objective is to model spread and persistence of BVDV at a regional scale, where herds with dynamical infection and prevalence statuses interact according to real animal exchange data and geo-location based neighbourhoods.

4.2 Modelling BVDV spread between farms: state of the art

A theoretical metapopulation of dairy herds was considered by [Courcoul and Ezanno, 2010]. Their objective was to study the spread and persistence of BVDV in a completely susceptible managed metapopulation of cattle herds, considering managed intensity of movements and number of neighbours of the herds. In the model, a metapopulation of 100 typical dairy herds was considered, all subjected to the same structure and management practices. Within herd spread dynamics and risks through exchange of animals and neighbours all had been included into this study. In this model, each farm buys a given number of heifers per year with random choice of the source farm for each purchase. Therefore, the underlying animal exchange network in this model was a random network, although its properties from a graph theoretic point of view was not presented. Similar to the fixed number of animals to be purchased per year, the number of neighbours of a farm was also fixed. The neighbours were chosen randomly from the metapopulation and once chosen kept fixed for all time and all realizations of a particular scenario. They argued that, for a small region containing a few herds, the possibility of having pastures of each farm anywhere within the modelled region implies that any farm in the metapopulation can be neighbour of any of the other. The neighbourhoods were commutative but not-necessarily transitive. Initial herd sizes of all herds were the same with the same demographic parameters. In this model, initially the whole metapopulation was susceptible. The introduction of the virus was through the purchase of a PI heifer in one of the herds randomly chosen. No new infected animal entered the metapopulation later. The dynamics of BVDV spread within the metapopulation was governed by the discrete time stochastic local herd dynamics of [Ezanno et al., 2007] (described in Chapter 3) along with the assumed random animal exchanges and neighbourhoods. Influence of stochastic events on BVDV spread could be seen in this model. It could estimate the global average of the number of PI animals over time in the metapopulation. Probability of persistence of the virus in the metapopulation, distribution of the metapopulation infection duration, probability of pathogen endemicity, the mean number of infected herds over time in the metapopulation still infected and the cumulative epidemic size could be estimated from this model. In this model, the quintessential aspect of herd structure dependence in BVDV spread was well considered, whereas the neighbourhood assumptions were valid only under 'small' region assumption. Given the small number of herds in a locality, the neighbourhood assumptions were valid, but could not be extrapolated to larger area such as a region. The assumptions of fixed number of neighbours and number of annual purchases were good assumptions to have insights about effects of animal purchase and neighbouring relations in BVDV spread in the context of small model metapopulation presented. At a regional scale neither of them nor equal size for all the herds might hold good.

[Ersbøll et al., 2010] evaluated the risk of a dairy herd changing infection status (from not having persistently infected (PI) animals to having PI-animals) in relation to location and infection status of neighbouring cattle herds in Denmark. This study was exclusively dedicated to the contribution of the neighbourhood, they did not consider the movement of animals. In this study they used geographic coordinates of the farms to decide the neighbours of a farm. For the purpose they used a *Delauney triangularization* [O'Sullivan and Unwin, 2003,Ersbøll et al., 2010] method limited to a distance of 5 km. The study involved spatial statistical model of risks to farms based on a priori herd statuses from a milk sample recording herd status database. This setup was good enough for investigating the sole risk of having infected neighbours and their contributions in BVDV spread.

This was indeed an essential part of the spatial BVDV spread but not the whole story. When a complete picture of BVDV spread between herds was the demand, one had to account for the contribution of animal exchanges too. They categorised the risk factors into spatial and non-spatial risk factors, and the only *non-spatial* risk factor they consider in this study was herd size, defined as the number of lactating cows in a herd. The *spatial factors* of this model is based on cattle herds in the first order neighbourhood defined by the triangularization method. The herd status of a herd in the neighbourhood is considered positive, depending on the status of the herd in the database at the starting time. They consider three groups of spatial risk factors (1) regional, (2) herd density and (3) infection in the neighbourhood. Herd density wise estimations are performed with two criteria (i) number of herds per unit area in the triangularized area and (ii) mean distance to neighbouring herds. When herd density and mean distance to neighbouring herds in the region were the same, the risk of infection was considered to be dependent on four factors: (i) occurrence of PI herds (yes or no); (ii) number of PI herds; (iii) proportion of PI herds; and (iv) distance from the nearest PI herd in the neighbourhood. The prime outcome of this model was the estimated risk of becoming a PI infected herd through neighbouring relationships, without explicitly considering the contact mechanism between the neighbours nor exchange of animals between farms. The model showed that occurrence of PI herds in the neighbourhood had a significant influence on the risk of becoming PI-herd for a herd belonging to that neighbourhood. Increasing herd size was positively correlated, while mean distance to the neighbours correlated negatively with this risk.

Contrary to the purely neighbourhood contributing study, discussed above, a purely animal exchange network based study of BVDV spread was presented in [Tinsley et al., 2012]. This study used a dynamic network of cattle exchange constructed from real movement data between beef-herds in Scotland. The model relies on the assumption that primarily BVDV spreads through transportation of PIs between herds. The model made no distinction between a PI animal physically roaming in the herd and a PI foetus carried by an immune dam (which resulted from the infection of the dam in the second stage of gestation, Chapter 3). The model considered an open system where introduction of (possibly infected) animals from outside the metapopulation (through animal exchanges) was allowed in any time of the simulation (depending on real movement data), unlike the one time introduction procedure followed in [Courcoul and Ezanno, 2010]. The probability of importing an infected animal from a source outside the metapopulation was taken care by considering the ‘outside world’ as a single infected pool. Both direct farm to farm and indirect (via markets) movements were considered. Instead of individual animals, in this work, herds were identified with two epidemiological statuses either susceptible (S) or infected (I), depending on absence or presence of a PI animal in the herd. Moreover, herd structure was not considered in this study, instead the process of infection is modelled in a SIS formalism, where a susceptible herd may change its epidemiological status to an infected herd with probability p when a movement from an infected source herd occurred. p was associated with the probability that the transferred animal was a PI. A susceptible herd could not change its epidemiological status with movements from a S herd. An I herd might clear the infection and become a S farm in two ways, (i) self clearance (death of PI, all susceptible animal becomes immune after transient infections or non-BVDV specific reasons) (ii) detection and removal due to control surveillance and control policy in force, if any. The model addressed three situations of detecting and removing PIs from the system: (i) restrict all PI movements; (ii) restrict PI movements through markets and (iii) restrict all PI movements from farms of a defined size based upon the volume of outflows from the farm. Since this model considers purely the animal exchange (network) sides contribution of BVDV spreading, it did not consider any possibility of getting an infection via neighbourhood or any other indirect path. In this model explicit heterogeneity in within herd prevalence among I-herds was absent. Persistence of PI animals in a herd is a complex phenomena and without explicit herd structuring it might be unrealistic to define the existence of PI animals on a purely probabilistic ground in a herd and its contribution in the within-herd dynamics [Ezanno et al., 2007, Courcoul and Ezanno, 2010].

Given the pros and cons of the previous works, discussed above, addressing the spread of BVDV between herds in a regional scenario, in this thesis we propose a model which takes care of some of the limitations of the previous works. In this model, we assemble all the ‘goodness’ of the previous works and apply to a metapopulation of dairy herds in the Finistère department of Brittany region in Western France. This model is aimed to mould all the ingredients mentioned so far into a monolith with a generalised approach: (i) heterogeneous herd structure and size; (ii) dynamical movement network; (iii) heterogeneous neighbourhood; and (iv) dynamical herd status towards infection and prevalence. The model should also be of open type, able to handle both direct (farm to farm) and indirect (farm - market - farm) animal movements in and out of the metapopulation. Such a model should be valuable to test various control strategies applicable to reduce prevalence as well as eradication of the pathogen from a metapopulation.

4.3 An explicit model of between herd BVDV spread in region

4.3.1 The approach: from intra to inter-herd model

To extend the intra-herd BVDV model to an inter-herd model, we need to make suitable adjustments keeping in mind the two main routes of between-herd BVDV spread. We take a bottom-up approach and propose a two step model. At every time step, we first account for the contacts, and then updating takes place according to the local dynamics.

Spread of BVDV through movements is determined by a stochastic process related to the prevalence of the pathogen in the source herd. A movement between two herds is determined by the movement network built from the FDCM. Therefore, the model works in the following way. All the sub-populations (herds) are updated as per animal movement information. In executing a movement of an animal between two farms (as recorded in the FDCM network), first the age information is used to determine the age-group of the animal. Then an animal belonging to that particular age group across the spectrum of the epidemiological states present in the group is selected.The probability of selecting an animal with a particular epidemiological state is proportional to the number of animals of that state in that group. Once selected, an animal belonging to that age class and epidemiological state is deducted from the source population and the same is added to the destination. When an animal sought after by the movement data is not available in the source herd (this situation may sometimes happen, depending on herd size and frequency of outflows), an animal belonging to either higher or lower age-group is selected with equal probability. The above stated step is applied for direct movements between farms. In case of an indirect movement, say farm to market or assembling centre then to farm ($F - M/C - F$) then the situation is slightly different. To take care of such situations there are two possibilities. First, we can rearrange the movement network beforehand. $F - M/C - F$ *time respecting paths* [Holme and Saramäki, 2012] then are redrawn as $F - F$ paths and the date of out movement from the source farm and of in-movement into the destination farm is flagged to account for possible change in the epidemic status. This method of tracing time respecting paths may be cumbersome, if time difference between out (from F) and in (to F) movements are long, especially when we concatenate the movement information multiple times to simulate the system for time periods longer than the available period [Tinsley et al., 2012] (we have the movement data for 5 years, if we simulate the system for 10 years, we repeat the movements for another 5 years). To avoid such cumbersomeness, the movements through markets and assembling centres can be considered to be into a common pool with a certain possibility of buying an animal of a particular epidemic state from that pool. We choose this last option if $F-M/C-F$ duration is longer than the simulation time step. The same is done for movements into or from farms outside the metapopulation [Tinsley et al., 2012]. This probability may be varied to see the effect of the external prevalence on BVDV spread in the metapopulation.

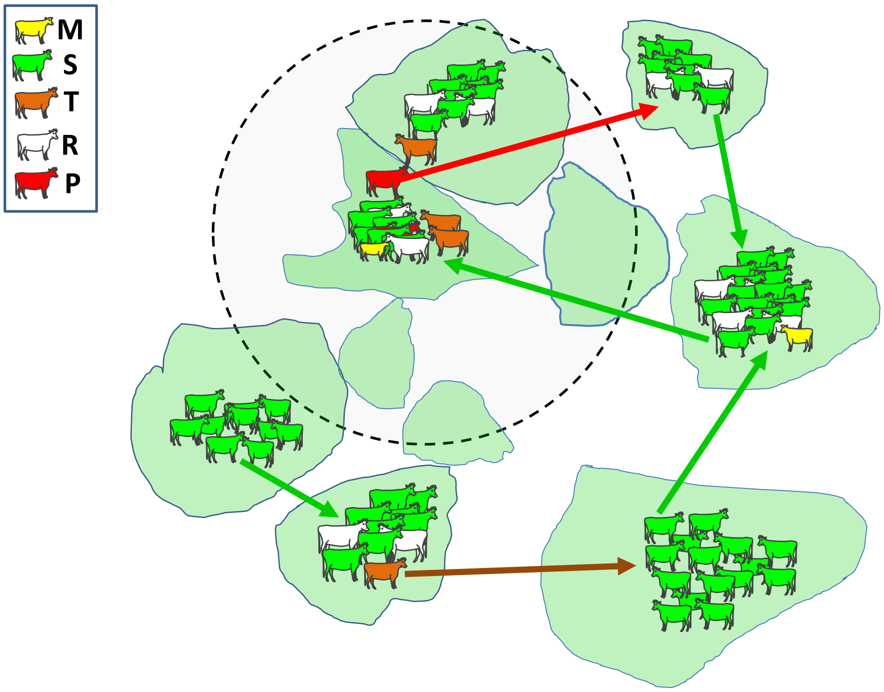


Figure 4.1: Outline of the inter-herd BVDV spread model. The model considers all the main perspectives of BVDV spread in a region: (i) heterogeneity in size of the herds; (ii) heterogeneous herd density; (iii) non-uniform (heterogeneous) number of neighbours and (iv) heterogeneous movement pattern between the herds. Herds touched by the dotted circle are considered to be neighbours of the herd at centre. The herd heterogeneity (in terms of herd density as well as herd sizes) is also reflected in the diagram. Animals of different epidemiological states in different herds are also shown. The exchange of animals between the herds may lead to transport of the pathogen depending on the prevalence in the source herd. During pasturing season, the infection may pass into the neighbouring farms too.

Since the geographical coordinates of a farm pointed to the location of the 'registered' housing unit of the concerned farm, it did not necessarily represent the locations of the pastures. During field visits the author had found that the pastures could be anywhere within a few kilometres around the geo-location of the farm and were not necessarily contiguous, similar to the description in [Courcoul and Ezanno, 2010]. Precise boundaries of the farms might not be obtainable due to various difficulties including privacy concerns. Therefore, a simple approximation for deciding the neighbours was used. Since the pastures belonging to a farm could be anywhere within its 'vicinity' of a few kilometres and not necessarily contiguous, we considered that all farms whose geo-locations fall within a *radius of vicinity* around a farm possibly contain all the bordering pastures [Fig. 4.1]. Although it posed the risk of overestimating the number of neighbours, it had two advantages: simpler to implement once the coordinates of the farms were available, no triangularization nor any complex tessellation procedural approximation was needed to define the neighbours; the commutative (if farm A is neighbour of farm B, then B is neighbour of A) and partial-non-transitivity (farms B and C are neighbours of A not necessarily imply that B and C themselves are neighbours) properties of neighbourhood were also satisfied with this definition of neighbourhood upto some extent, although not full proof. BVDV between-herd spread through neighbourhood is a stochastic process incorporated into the within-herd dynamics.

At the herd level the building block is the intra herd model we discussed in Chapter 3. The model is modified to take account of contributions of neighbouring relationships during the pasturing season. The difference between the 'original' intra-herd model without considering the contribution of neighbours and the one considering the neighbours' contributions is as following. Neighbouring contacts were assumed to occur only during the pasturing season (spring-summer, mid-March to mid-November). During this season, dry cows (DC), heifers (H) and young heifers older than 6 months ($h*$) were grazing and exposed to possible neighbouring contacts. It was assumed that only persistently infected animals in a group contribute to infection via neighbouring relations. The probability of infection by neighbouring contacts at pasture was assumed to be dependent on the prevalence of the infection in the neighbouring populations and was considered to be frequency dependent. The rate of infection (per day) due to neighbouring farms for group g at time t was given by

$$p_{inf}^{nnb}\left(g,t\right) = b_{nnb}^{P} \frac{\sum_{nnb} P_{h*}^{nnb} + P_{H}^{nnb} + P_{DC}^{nnb}}{\sum_{nnb} N(g,t)\left(N_{h*}^{nnb} + N_{H}^{nnb} + N_{DC}^{nnb}\right)} \tag{4.1}$$

where b_{nnb}^{P} : transmission rate between an infectious animal and a susceptible animal belonging to two neighbouring herds; $N(g,t)$: number of animals in group g at time t; N_x^{nnb} : number of animals in group x in the neighbouring herd; and P_x^{nnb} : number of persistently infected animals in group

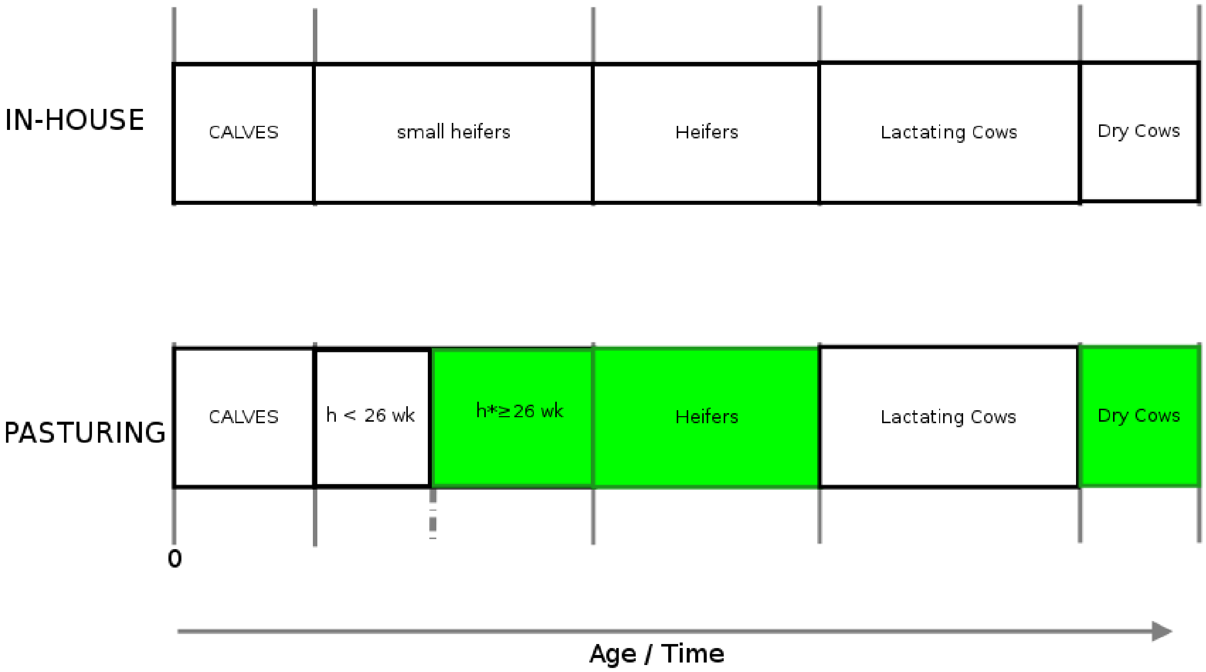


Figure 4.2: Herd structure depending on the season. During the pasturing season, the young heifer class is split into two subgroups, animals older than 6 months going to pasture. Similarly, bred heifers and dry cows are sent for grazing in pasture. The groups that go to pasture are shown in *green*. Susceptible animals in these groups are at risk of infection on pasture, depending on the prevalence of the pathogen in these three groups in the neighbouring herds.

x at the same time. The probability of infection due to within-herd prevalence and neighbouring relations of group g at time t was given by

$$\text{prob}_{inf}(g,t) = 1 - e^{-\Delta[p_{inf}^{herd}(g,t)+p_{inf}^{nnb}(g,t)]} \tag{4.2}$$

with p_{inf}^{herd} the infection rate due to the within herd prevalence; and Δ : the time step of the model. If time t did not belong to the pasturing season, $p_{inf}^{nnb}(g,t) = 0$ and the model reduced to the model of Chapter 3. This model took account of the local population dynamics and a two state epidemiological dynamics: local (within-herd) and semi-local (between neighbours) at the herd level. Essentially the herd structure during the pasturing season is slightly modified than that during the in-house season [Fig. 4.2].

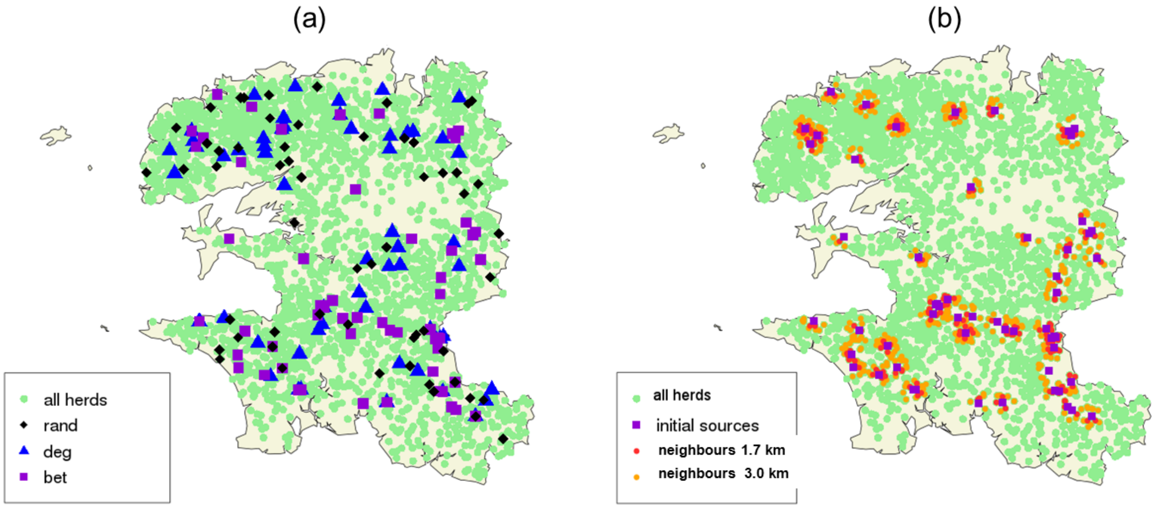


Figure 4.3: Different scenarios of introducing initial infections into the metapopulation. (a) Spatial locations chosen randomly and top 2% herds in terms of degree and betweenness centrality; (b) Locations top 2% herds in terms of betweenness centrality and their neighbours with two different radius of *vicinity* [1.7 km and 3.0 km]

4.3.2 Parameters and initial conditions

The time step of the intra-herd BVDV model had been set at 14 days. Therefore, the network of exchange of animals among farms was aggregated at a two week interval, making it a semi-temporal network with real time increment of two weeks per simulation step.

The metapopulation we considered consists of $2,846$ dairy farms from the Finistère department. These farms were chosen as their geo-coordinates were available. Moreover the prevalence of the pathogen in this region is also known for last 10 years. All other farms exchanging animals with these farms consisted of the external pool of farms. In the present investigations, for brevity, markets and assembling centres were included in the same pool, if the concerned animal did not make $F - M/C - F$ move within one simulation time step. They will be considered as separate entities in a near future and movements through will be treated separately.

Within group transmission parameters were kept the same as those mentioned in Table 3.2. They were similar irrespective of herd size as the herd structure was also kept the same. Initial herd sizes were described as per the FDCM based estimation of the subpopulations and distributed among the age classes proportional to a base herd (of 85 animals) used for the within herd BVDV spread model [Ezanno et al., 2007]. The metapopulation was first simulated for 10 years in a disease free scenario. The output distributions for different herds were then used as initial populations. The

10 year disease free simulation of the metapopulation dynamics leads to stable populations with animals in all age and physiological classes determined by the herd management parameters.

No biological information was available for the rate of transmission between farms on pasture. We assumed that the value of parameter b^P_{nnb} was equal to the rate of between group transmissions β^P_b due to PI animals, within a herd [Table 3.2]. This choice is justified because both events are closely related. To observe the effect of this parameter on BVDV spread in the metapopulation we compared simulations for two more values of this parameter, 0 (no contribution of neighbourhood) and 0.1 (contribution less than between group transmission); with $b^P_{nnb} = 0.5$ $(= \beta^P_b)$. We also studied two scenarios with $b^P_{nnb} = 0.5$ at three neighbourhood radii $r = \{1.7, 3.0.5.0\}$ km without movements (spreading through neighbourhood relationships only).

Neighbours were defined before the start of the simulation as the farms within or touched by a circle drawn with particular radius around the geo-coordinates of a farm [Fig. 4.1]. We used three values of this radius, to be able to compare with the results of existing models: (i) radius of 1.7 km around farms led to neighbourhood distribution of 5 neighbours on average (one of the scenarios in [Courcoul and Ezanno, 2010] used 5 neighbours per herd); (ii) 5 km (distance for Delauney triangularization in [Ersbøll et al., 2010]; and (iii) 3 km, an intermediate between the previously mentioned two. All scenarios were simulated with a value from this set.

In the presented work, we did not consider the risk of buying an infected animal from a source outside the metapopulation. But we accounted for indirect movements through farms and assembling centres (not part of the metapopulation) if the leaving and entering dates are within a time step of the model (2 weeks), by replacing the indirect movements with direct movements in the aggregated network.

To introduce the disease into the initially naive metapopulation, a single infected animal of specific infectious state was introduced to each of a certain fraction of the members of the metapopulation: (i) randomly chosen; and (ii) selected from descending ordered list of their *centralities* [degree (out): number of customer farms, strength (out): volume of sell of a farm, and betweenness centrality: number of interlinking transfers between all other farms passing through a farm] from the FDCM network. We chose top 2% of the farms for each criteria. The type of infected animal to be introduced was chosen to be a PI cow carrying a PI foetus. This choice was taken based on the observations that this type of animal introduction resulted in the longest persistence probability within primary case herds [Chapter 3]. The locations of the farms chosen for initial introduction of the pathogen into the metapopulation is shown in Fig. 4.3.

To observe possible dependence on the type of animal carrying the infection initially introduced into the completely naive metapopulation, we introduced three types of animals PI calves, PI heifers, and PI cows carrying PI foetuses [in ascending order of persistence, observed for intra-herd dynamics of BVDV spread, Fig. 3.4(a)]

Total number of scenarios studied here is 30. Every scenario is repeated 200 times for simulations running for 10 years after the introduction of the pathogen ($t = 0$, starting with movements on 01/01/2005 from FDCM).

4.3.3 Outputs of the model

Outputs of this model are: (i) persistence probability of the pathogen in the metapopulation (fraction of simulations where the virus is not cleared from the metapopulation); (ii) percentage of infected herds in the metapopulation still-infected; (iii) number of infected animals (acute, chronic and in-utero) in herds still infected; (iv) infection entering naive herds per infection route (movement and/or neighbourhood); (v) epidemic size in terms of cumulative number of infected animals (acute, chronic, in-utero and all); and (vi) number of seroconverted (previously infected, presently recovered) lactating cows over time. The choice of these outputs are considered to be able to compare the model behaviour with those of existing models in the same context. The last item in the list of regional scale BVDV model outputs, namely, the number of seroconverted lactating cows per herd, might be used to correlate the model outputs with the on-field BVDV surveillance and estimation of BVDV prevalence in dairy herds using the bull milk tank sample screening in the same region [Beaudeau et al., 2001, Beaudeau et al., 2005]. This would enable one to validate the presented between herd BVDV spread model at a regional scale.

4.3.4 Coding and implementation

The BVDV spreading model between dairy herds on regional scale was coded in C++ (C++11 standard) [Stroustrup, 2013]. The scheme of implementation is shown in Fig. 4.4. In the program, a herd was represented as an array of physiological classes and epidemiological states with number of animals in corresponding states as array elements. The program ran in three loops. The outer

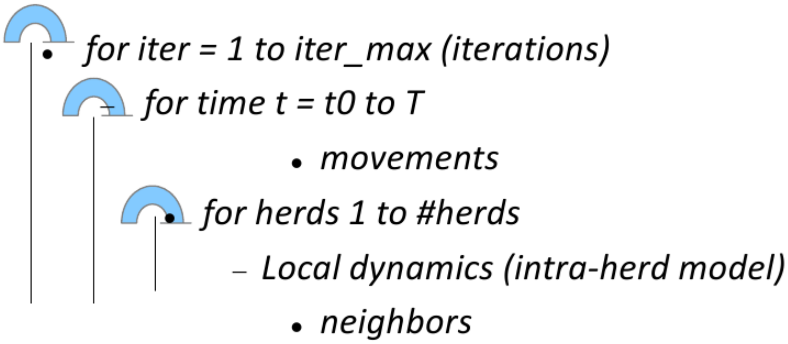


Figure 4.4: Coding and implementation scheme of the regional scale inter-herd BVDV spread model.

most loop was for repeating each scenario given number of times (200 repetitions, if not mentioned otherwise). In the second loop, for every time-step the program reads the movement information (FDCM) from an indexed data structure and moved animals between all source and destination pairs (identified by age class converted into physiological class and epidemiological state), by deducting from the source and adding to the destination (same physiological class and epidemiological state as the source). The exchanges (subtractions and additions) were implemented in terms of vector arrays (of same length and structure as that of a herd) and could be directly added or subtracted when an incoming or outgoing movement took place. Once all the movements in a time-step were taken cared of, it was the time for the innermost loop. In this loop, for each herd synchronously updated the population dynamics and BVDV spread within herd (including neighbourhood relationships, if the time-step concerned was flagged as pasturing season, using a neighbour information data structure).

4.4 Results

Spatiotemporal propagation of the pathogen is shown in Fig. 4.5. This shows spreading of the disease over the metapopulation landscape with time. Visually we can see that the disease spreads locally (neighbourhood) and through animal movements: infection persisted in initially infected herds and newly infected herds appeared in their vicinity as well as in new locations. Thereafter, we give more detailed results on the relative contributions of both transmission routes.

We observed that irrespective to the type of infected animal introduced initially into the metapopulation or the neighbouring radii, the virus persisted in the metapopulation very long compared to

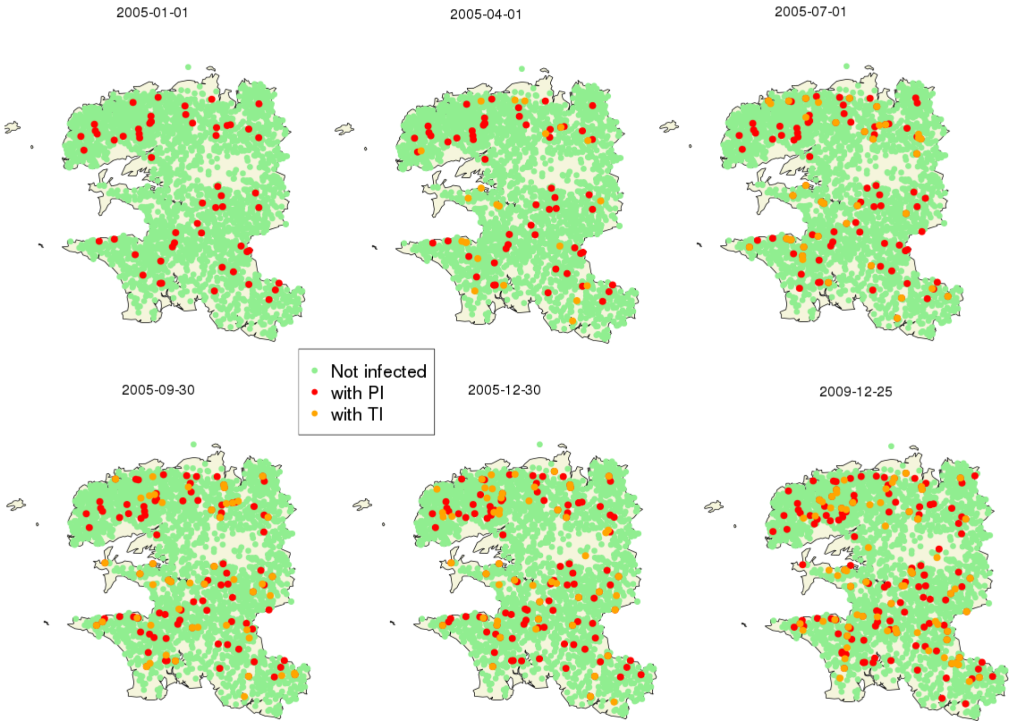


Figure 4.5: Snapshots showing spatial propagation of BVDV in the the metapopualtion. The snap shots were taken at day 0, 1 month, 3 months, 9 months, 1 year and 5 year of initial infections. If a farm had at least one PI or PI foetus, its location is shown as a red filled circle. Farms infected only with TI animals at the time of the snapshot were shown as an orange filled circle. Here external risk was set to zero. Initial herds were chosen based on degree centrality. Each initial infection was a PI cow carrying a PI foetus. The neighbourhood used was 3 km.

individual intra-herd scenario. This in terms of number herds still infected over time is shown in Fig. 4.6. Moreover, the average number of infected animals show no descending trend upto 10 years after introduction of infection into the metapopulation [Fig. 4.7].

The number of still infected herds in the metapopulation was found to be $\sim 6-11\%$ which was consistent with a previous theoretical model and field observations said that around 10% of the dairy herds in the region we considered had at least one persistently infected animal as reported in [Courcoul and Ezanno, 2010]. This was well observed when persistently infected adult animals were introduced as initial infections. Although the differences were not very high among them, a PI cow with a PI foetus induced the largest epidemic size and the largest fraction of the metatpopulation was infected [Fig. 4.6].

For the neighbourhood parameters and neighbourhood radii, we observed that the larger was the neighbourhood transition parameter b_{nnb}, the higher was the chance of getting a neighbouring

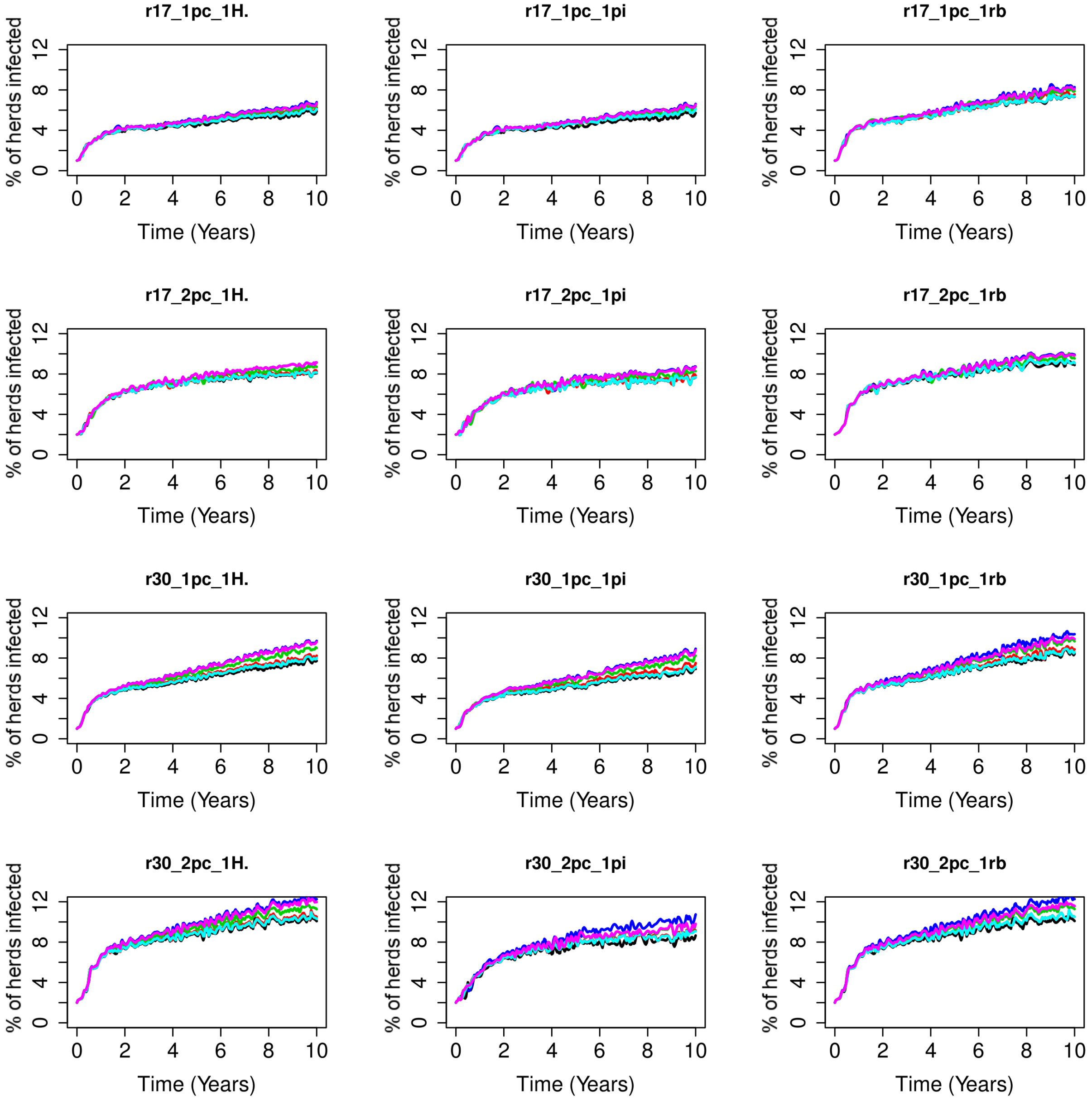


Figure 4.6: Percentage of herds still infected after 10 years depending on initial infection introduced to the metapopulation with different selection criteria. 1% and 2% of the total herds in the metapopulation were chosen *randomly* to introduce infections at $t = 0$. Each initial infection were PI cow carrying a PI foetus (rb), PI female calf (pi) or PI heifer (H). The neighbourhood shown here were 1.7 and 3 km. No external risk was considered.

herd infected through this path [Fig. 4.8]. The relationship between neighbourhood infection rate and the increase in the fraction of herds getting infected through this path was not linear.

The dependence on the choice of initial location of infection (based on movement network descriptors) for different radii of neighbourhood and neighbouring infection rates are shown in [Fig.

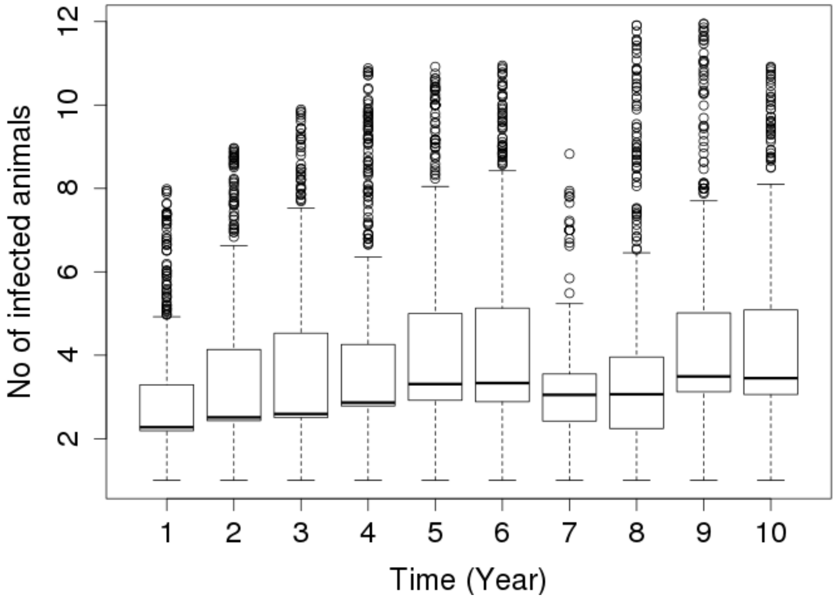


Figure 4.7: Boxplot of the number of infected animals per infected herd over time. The initial herds were chosen based on their betweenness centrality. Each initial infection was the introduction of a PI cow carrying a PI fetus (2%). (b=0.5, r=3 km and no external risk).

4.9]. We observed an increase in the mean fraction of herds infected with increasing parameter and radii, but the relations were not linear. Among the initial location choices, betweenness centrality based choices led to comparatively large fraction of the metapopulation still infected 10 years after the initial infection.

Persistence probability of BVDV in the metapopulation for different choices of initial herds, neighbourhood radii and neighbourhood infection parameters are shown in Fig. 4.10. We observed that in the metapopulation BVDV persistence with exchange of animals was higher than 60%, 10 years after introduction, whereas that when BVDV spreads through neighbourhood relationships only was below 20% 10 years after initial infection. When there were both movements and neighbouring relationships together, there was a slightly higher persistence probability with increasing neighbourhood radii and neighbouring transmission rate. For the same radius, the differences between probabilities were very small.

The distribution of the number of PI animals per infected herd over the years is shown in Fig. 4.11. We noticed that the median value is 1, which is constant over the years although the spread may be from $0 - 5$ animals.

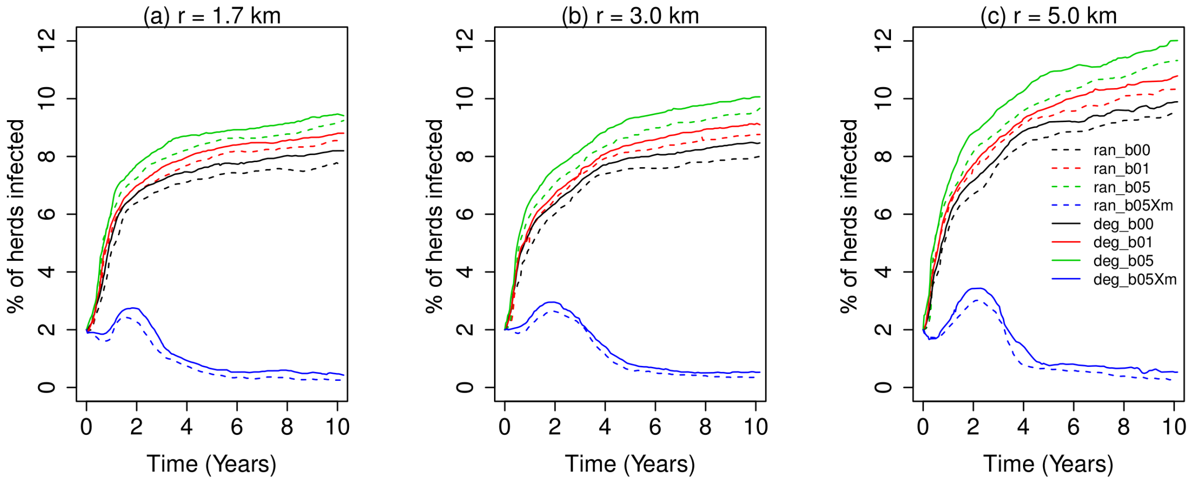


Figure 4.8: Fraction of herds still infected for different neighbourhood infection rates and different neighbourhood radii $r = 1.7, 3.0$ and 5.0 km [(a), (b) and (c)] respectively. A PI cow carrying a PI foetus was initially introduced into 2% of the total herds chosen either randomly (*ran*) or based on high degree centrality (*deg*). Three choices of transmission rate due to neighbourhood were considered 0 (no neighbourhood, only movements), 0.1 and 0.5 (denoted by *b00, b01* and *b05* respectively in the figure legends). Two scenarios with high neighbourhood infection rate *b05* were also shown when movement of only susceptible animals were assumed between farms (*b05Xm*, a scenario when the pathogen spreads only through neighbouring relations). The risk from external sources (outside the metapopulation) was considered to be zero.

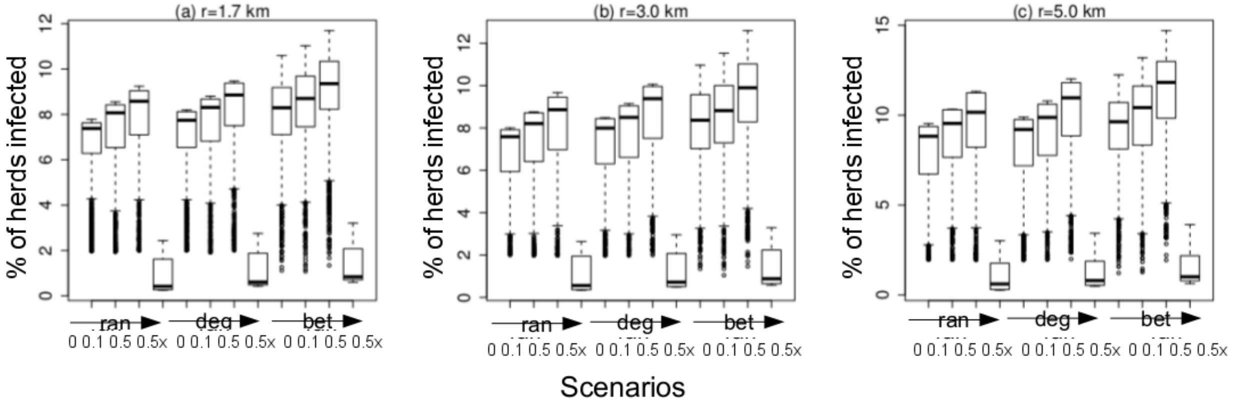


Figure 4.9: Variation in the fraction of herds still infected after 10 years depending on initial infection introduced to the metapopulation with different selection criteria [random (*ran*), degree centrality (*deg*) and betweenness centrality (*bet*)] of initial herds and different neighbourhood infection rates [0, 0.1, 0.5 (with movements), and 0.5× (only neighbourhood) respectively] for each selection criteria. (2% of the total herds were initially introduced a PI cow with a PI foetus each and no risk from outside the metapopulation was considered).

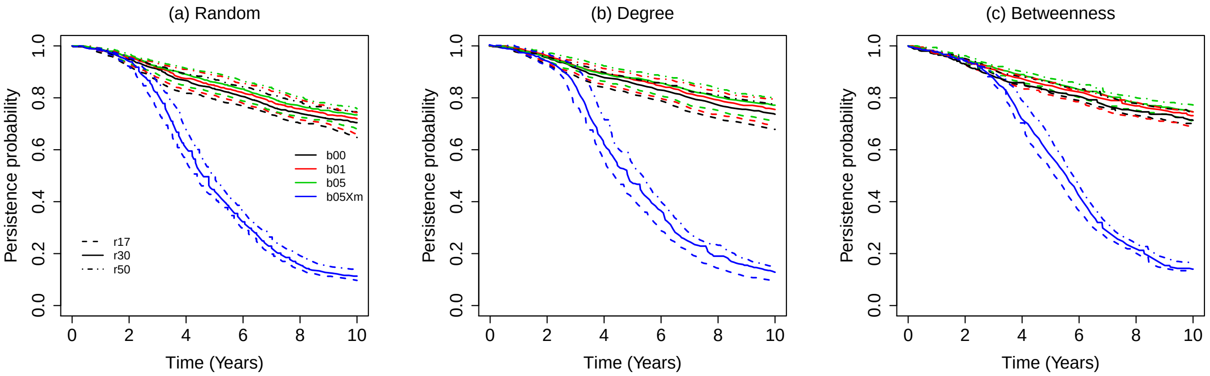


Figure 4.10: Persistence probability of BVDV in the metapopulation for different neighbourhoods (radii $= \{1.7\ (r17), 3.0\ (r30), 5.0\ (r50)\}$ km) and neighbouring infection rate ($b = 0\ (b00), 0.1\ (b01), 0.5\ (b05)$. Neighbourhood relation ship scenarios without infection via movements for the largest value of neighbourhood infection rate ($b05Xm$) are also shown for different neighbourhood radii. 2% of the total farms in the metapopulation were infected with a PI cow carrying a PI foetus each. They were chosen (a) randomly; and based on (b) degree and (c) betweenness centrality.

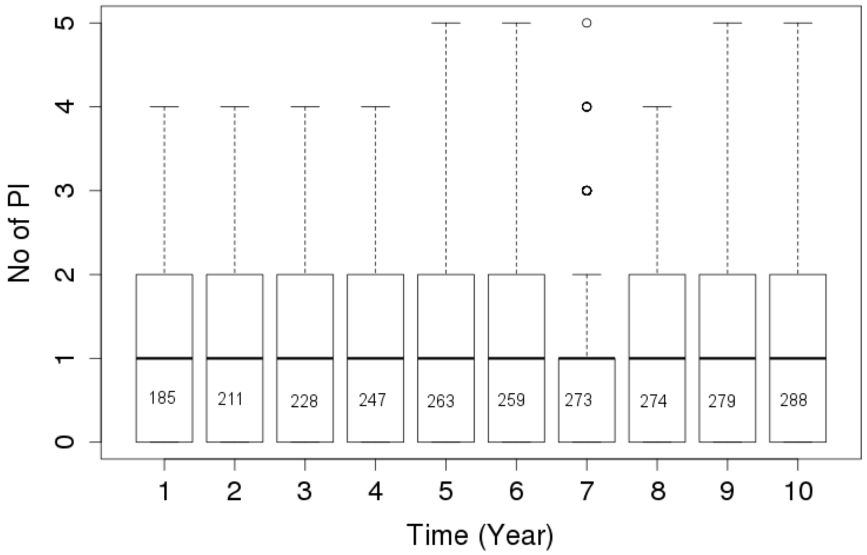


Figure 4.11: Boxplot number of PI animals in the infected herds over time. Initial herds were chosen based on betweenness centrality. Each initial infection was a PI cow carrying a PI foetus. The neighbourhood used was 3 km. (The numbers within the boxes represent the number of herds still infected, over which the average was taken.)

In terms of contribution of pathways of introducing a new infection into naive herds, we plotted the fraction of new infections (in a trimester basis) caused through neighbouring relationships in Fig. 4.12. We noted that both radii of neighbourhood and the neighbourhood transmission

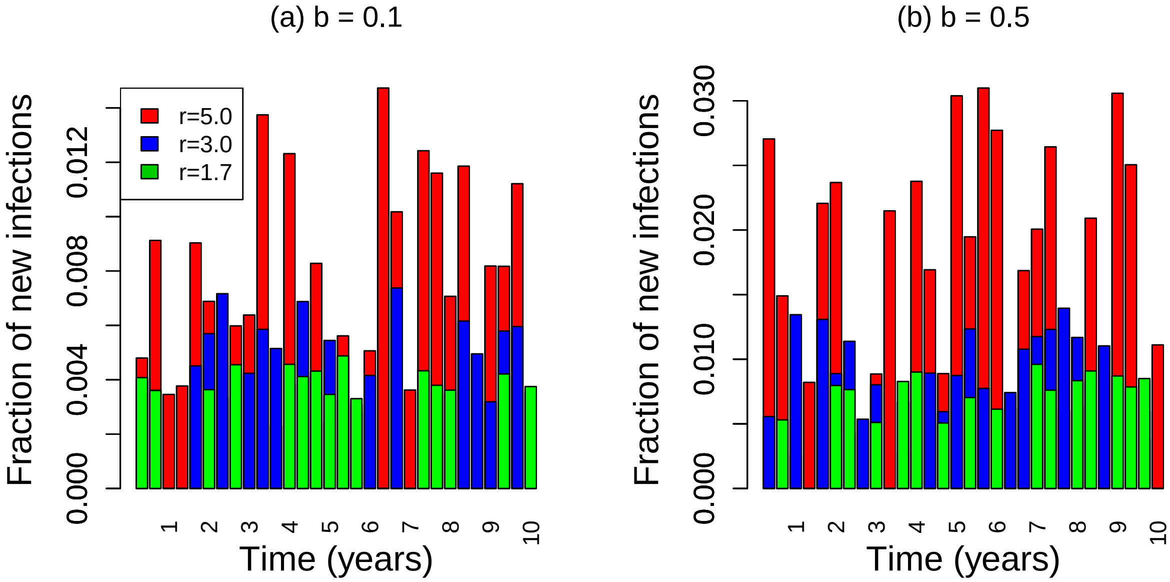


Figure 4.12: Fraction of new infections (cumulative heights of the bars) coming through neighbouring relationships for different neighbourhood distances ($r = \{1.7, 3.0, 5.0\}$ km) and transmission rate due to neighbourhood relationships (a) $b = 0.1$ and (b) $b = 0.5$. 2% of the total herds in the metapopulation were chosen to introduce infections (PI cow carrying a PI foetus). No external risk was considered.

rate influenced the number of infections through the pathway. The contribution of neighbouring relationships was between $1 - 3\%$ of the total number of new infections for the largest values of neighbourhood radius (5km) and the corresponding transmission parameter (0.5).

The model can be used to evaluate the proportion of seroconverted lactating cows in each herd at every time step. As an example, we show a snapshot of the percentage of seroconverted lactating cows per herd [Fig. 4.13]. The results obtained at herd scale may be compared with bulk tank milk samples obtained over several years in most of the dairy herds due to a BVDV surveillance scheme active in the region under consideration [Beaudeau et al., 2001, Beaudeau et al., 2005].

4.5 Discussions

A model of BVDV spreading between interacting herds on a regional scale was presented. The model used realistic animal movement network between farms. It accounted for the heterogeneity in herd size and localization. The model was specifically applied to a metapopulation of dairy herds in Finistère département in Western France. The outputs of the model could be used for both

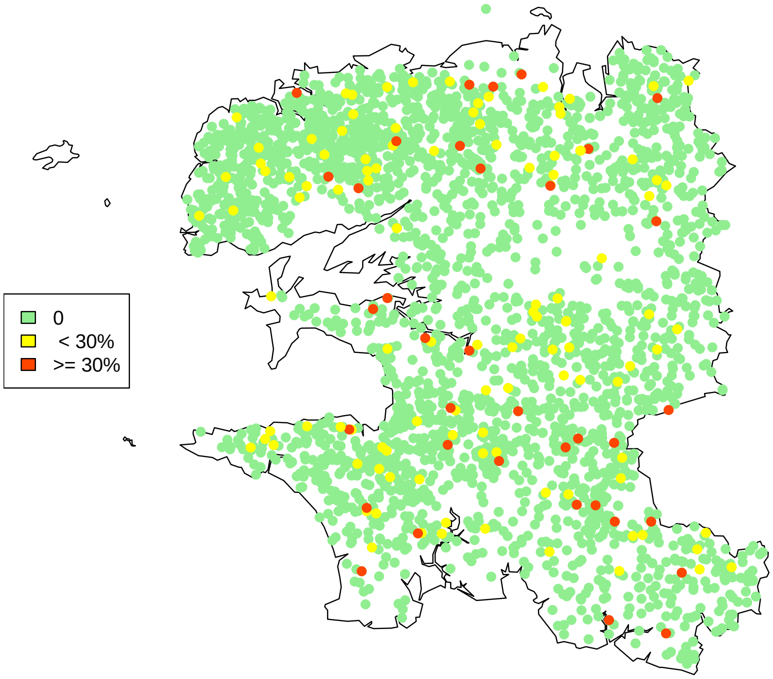


Figure 4.13: Snapshot of the percentage of seroconverted lactating cows per herd. The snapshot was taken 1 year after the initial infection. [Simulation parameters: initial infection locations (2%) were chosen on the basis of the (network) degree, type of infected animal: PI cow carrying a PI foetus, neighbourhood radius = 3 kms, neighbour transmission rate = 0.5 and no external risk]

understanding propagation of the pathogen and testing efficacy of in-practice and hypothetical control strategies.

The model considered two routes of BVDV transmission between herds, movement of animals and neighbouring relationships. Information about herds (size, type) and movement of animals among them were taken from the French National Bovine Database (BDNI or French Database of Cattle Movements, FDCM). Geo-locations of the farms were used to choose the neighbours. The major route had been through movement of animals between farms, in agreement with [Courcoul and Ezanno, 2010, Tinsley et al., 2012]. We found that neighbourhood had an impact on spread and persistence of BVDV between herds, which is higher than the theoretical estimation in [Courcoul and Ezanno, 2010]. Although the contribution of neighbouring relationships were qualitatively in agreement with [Ersbøll et al., 2010] about contributions uniquely from neighbouring relationships (no movements), quantitative comparisons were not possible, as the present model counted the number of animals in different epidemic states or number of herds with infected animals unlike estimation of risk in [Ersbøll et al., 2010]. The differences between the current model and that of [Courcoul and Ezanno, 2010] were heterogeneity involved in movements (defined by the actual trade network), neighbourhood (defined by geo-locations) and herd sizes (defined by BDNI). We observed the contribution of accounting for these factors (using real data) in the model outputs. Although qualitatively similar, quantitative outputs of the present model were higher than those of [Courcoul and Ezanno, 2010].

To evaluate control strategies at a regional scale, the approach presented in this model is the most relevant one among the available options to date. Explicit consideration of movements between farms in the model enables one testing consequences of controlling movements of infected animals [Tinsley et al., 2012], for various levels of efficacy for different diagnosis methods [Lindberg and Alenius, 1999, Hilbe et al., 2007]. Within-herd hygiene or biosecurity measures [Kelling et al., 2000, Smith and Grotelueschen, 2004, Lindberg et al., 2006] can also be implemented as the model explicitly considers both infection and herd dynamics. Number of seroconverted animals can be counted in each physiological group within a herd, which may be used to define herd statuses based on the percentage of seroconverted lactating cows in a herd, which may be correlated with bulk tank milk sampling method of BVDV surveillance [Beaudeau et al., 2001, Beaudeau et al., 2005]. Test and cull strategy for new born persistently infected calves, subject to efficacy of diagnostic tests [Hilbe et al., 2007] can also be implemented in the current model. The versatility of the model in implementing the control strategies in practice is an added advantage of the model presented, which may potentially be developed into an in-silico diagnostic tool for BVDV spread on a regional scale.

Over all, taking cue of all the goodness of the available models of BVDV spread between herds, the present model may potentially be the most relevant model considering the complexity of the dynamics (at herd and regional levels) and testing of control strategy features concerned. Since the inter-herd contacts are independent of the local herd and infection dynamics, the model can be extended to account for beef farms present in the region. Within beef herd dynamics of BVDV spread model may be completely independent of the within dairy herd model, but to account neighbourhood relationships between neighbouring beef and dairy herds, slight modifications may be necessary in the neighbourhood contribution term of the model, given the different herd structures and grazing practices for beef and dairy herds. A beef herd BVDV spread model with that purpose in mind has been developed recently [Damman et al., 2014]. Incorporating these possible additions will lead to development of a comprehensive tool for animal health managers.

Chapter 5

General discussion

5.1 A short summary of the thesis

This thesis began with cleaning data extracted from the *Base de Données Nationale d'Identification* (BDNI, equivalently the French Database of Cattle Movements, FDCM) for a period of five years (2005-2009). Network analysis of the FDCM network led to identification of key parameters which might influence spread or control of a pathogen via animal movements. The network descriptors were studied for different levels of temporal (from weekly to yearly scales) and spatial resolutions (from holdings to communes and regions to the national level). The analyses were of two types: (i) classical static network analysis and (ii) temporal analysis. The most central nodes which might take important role in spreading a pathogen were identified. Comparison of geographical and network distances reveal the relationship between distances between farms exchanging animals and number of network 'hops'. The network analyses had a three fold objective: exploring temporal variations of the main network characteristics, computing proxies for pathogen spread on the network, which accounts for its time varying properties and identifying specificities related to main types of animals and farms (dairy and beef). This thesis provided extensive analysis of the properties of the network of cattle movements in France. It was the first study to do network analysis of the French cattle movement data revealing similarities and dissimilarities between networks constructed at different levels of spatial and temporal aggregation over a multi-annual dataset.

A model concerning spread and persistence of bovine viral diarrhoea virus (BVDV) in a region was presented in this thesis. The model focused on the dairy industry in the Finistère administrative unit (*département français*) in the French Brittany region. The subpopulations of the metapopulation model presented here are dairy cattle farms of various sizes. The model involved four levels of modelling: (i) local herd population dynamics dependent on farm management practices, which might be specific to regions; (ii) the virus spreading dynamics within the herd, once the virus enters a naive herd or already present in the herd; (iii) possible transportation of the virus between herds by exchange of animals (depending on prevalence of the virus in the source herds), mostly necessitated by the requirement of replacing the breeding stock to keep the production and herd size stable in a dairy herd or any other reason decided by the farmer; and (iv) possibility of acquiring the virus at over-the-fence nose-to-nose contacts or at common water holes during pasturing, between animals belonging to neighbouring farms. The local herd population and infection spread within a herd in the model were taken care by the within herd BVDV spread model of [Ezanno et al., 2007]. Sizes of the herds and movements of the animals, possibly carrying the virus from one farm to another, were also based on real movement data from FDCM.

5.2 Main results

5.2.1 Network analysis

Network properties were not found to be qualitatively different among different temporal and spatial scales considered in the study. About 40% of the holdings and 80% of the communes were found to be directly interconnected. The width of temporal aggregation window had shown no significant impact on normalised distribution of network descriptors. A time window of 8 to 16 weeks would suffice for robust estimation of the main trends of the descriptors. However longer windows of observation would be required to shed lights on the extreme ends of the distributions. The small overlap between consecutive networks of equal time windows revealed the dynamic nature of the network. We noticed that there were 65% of common holdings while only 3% of the links between them were common over the observed period. Percolation under various schemes of node removal were performed to estimate the response of the network to possible control strategies based on blocking animal flows in case of pathogen spreading on the network. We showed that a random removal of nodes was not an effective measure to resist the connectivity (hence spread of an epidemic process crawling in the network). But to our relief, centrality measure (degree, strength and betweenness) based removal of nodes based on previous year movement data proved effective in reducing the size of the largest connected component of the network in the running year. We showed that restricting movements from $1-5\%$ of the holdings with the highest centrality would result in a reduction of the possible maximum epidemic size by 80% in terms of the largest connected component of the network. Similar trends were observed in temporal variation of average indicators and their distributions in herd-type as well as breed-type based subnetworks of the FDCM network. By analysing the regional networks with contrasting dominance of farming systems, we noted that consideration of local specificities and temporal dynamics of animal trade networks were important to evaluate control measures of a possible pathogen spread on the network. We showed that even though the classical and static network parameters were not significantly different, epidemic burden proxy for temporal (time-stamped) formalism might be significantly different. Most importantly, aggregated static network analysis predicts large connected component (proxy of epidemic size in static formalism) in beef cow farming subnetwork whereas in time-stamped temporal formalism dairy cow subnetwork was found to have higher reachability (proxy for epidemic size in temporal formalism). We utilized the cattle trade network data and some of its analyses in modelling possible transportation of BVDV between dairy herds.

5.2.2 Intra-herd dynamics

We observed that persistence of BVDV in a dairy herd was dependent on the types of infected animals present/introduced (in case of naive herds) in/to the herd. Presence/introduction of infected adults in a herd was found to be more influential than the one of infected calves and young heifers. In terms of infectious states, chronic infections resulted in longer persistence and larger epidemic size compared to acute infections. When one infected animal was introduced into naive herds of different sizes (of same herd structure), we observed expected larger epidemic sizes in larger herds. When normalising epidemic size to herd size, it was found to be larger for smaller herds. When the number of introduced infected animals was proportional to herd size, then in most occasions normalised epidemic sizes were comparably of the same levels irrespective of herd sizes. This linear relationship might be an indication of the importance of herd structure in the spread and persistence of BVDV in a dairy herd.

5.2.3 Inter-herd dynamics

The virus persisted in the metapopulation for longer time compared to individual unconnected herds. The number of still infected herds in the metapopulation was found to be $\sim 6-11\%$ which was consistent with a previous model [Courcoul and Ezanno, 2010]. Field observations showed that around 10% of the dairy herds in the region we considered to have at least one persistently infected animal before implementation of control scheme in the region as reported in [Courcoul and Ezanno, 2010]. This was well observed when persistently infected adult animals were introduced as initial infections. Half of the infected herds had atleast one persistently infected animal as predicted by the model.

We observed that persistence probability of BVDV in the metapopulation was dominated by movement of animals compared to the contribution of neighbouring relationships. When there was only neighbouring relationships but no movement, persistence probability after 10 years was $\sim 10-20\%$, whereas with animal exchanges it was $\sim 60-80\%$. Expectedly, the longest persistence, largest number of infected herds and epidemic sizes were observed when persistently infected cows were introduced at $t = 0$. The shortest persistence and low epidemic size were observed when transiently infected animals were introduced into the metapopulation. These observations were similar to those observed for spread of the virus within a herd.

The contribution of the network characteristics was observed when the initial infections were introduced based on centrality measures of the herds. Initially infected herds chosen according to high centrality measures resulted in slightly higher number of infected herds compared to randomly chosen initial herds. Among the centrality measures we have used, betweenness centrality based choice had the lead in terms of number of herds infected.

In spreading the virus between herds, the relative contribution of the movements was found to be higher compared to those due to neighbouring relationships. Contributions of neighbouring relationship was found to be dependent on the number of neighbours (based on different neighbourhood radii). The contribution of neighbours was $\sim 1-3\%$ of the new infections for a typical moderate neighbourhood radius of 5 km (average number of neighbours = 41) whereas that for a radius of 3 km was about $40-60\%$ less. As expected, neighbourhood contributions were higher for higher transmission rates when the radii was kept constant, but the relation was not proportional to the value of the transmission rate. The contributions were $30-40\%$ less when $b = 0.1$ compared to $b = 0.5$.

5.3 Comments on data, methods, models and findings

5.3.1 Data on animal movements and network analysis

Representing livestock movement data as networks and their graph theoretic analyses had been done to gain insights for controlling pathogen spread through movements of infected animals. The methodologies were common over different fields in academia and industry, ranging from food chain networks to online social networks. Livestock movement networks are not only directed and weighted, they are time varying too. There had been substantial work on network analysis of any kind of data representable as networks [Wasserman, 1994, Newman et al., 2006, Barrat et al., 2008, Newman, 2010, Easley and Kleinberg, 2010, Holme and Saramäki, 2012, Blonder et al., 2012, Dorogovtsev and Mendes, 2013] and it had been a hot subject among recent multi-disciplinary areas of study. The development of network descriptors for weighted and directed network, the most general type of networks, and their interpretation are still expanding topics in the literature. We used descriptors and parameters already available in the network analysis literature. We evaluated only the parameters and descriptors which could be related or at least interpretable in relation to livestock farming, trade and epidemic spread between elements of livestock industry [Christley et al., 2005, Bigras-Poulin et al., 2006, Kiss et al., 2006, Ortiz-Pelaez et al., 2006b, Dubé et al., 2009, Rautureau et al., 2011, Nöremark et al., 2011, Bajardi et al., 2012, Frössling et al., 2012, Mweu et al., 2013, Büttner et al., 2014].

Integrity and accuracy of movement database system was a must for constructing networks of livestock movements. Inaccurate recording of the data or incomplete information might lead to serious differences between what had been estimated from the analysis and what actually happened in the field. In case of FDCM, the volume of data we had to discard due mostly to incomplete or inconsistent information was about 0.018% ($\sim 12,000$ out of ~ 60 million entries) of the raw volume of the data extracted from the BDNI. The main inconsistency were incoherence in dates of entry (including birth) and exit, same animal simultaneously present in more than one herd, multiple movements of the same animal on the same day and recording in the causes of entry and exit. Keeping in mind the volume of the data, the quality of the data was appreciable.

The results of the network analysis might be very useful to design control strategies in case of epidemic emergencies. The network may also be analysed in the line of statistical models of livestock trade networks [Ensoy et al., 2014] to estimate risks at regional, local or individual farm levels for various temporal windows [Nöremark et al., 2011, Frössling et al., 2014]. Only a few outcomes of the network analysis performed in this thesis was actually used in the model specific to BVDV spread. To our knowledge, this is the first instance of analyses of the French cattle movement data to show the robustness of the network in various spatio-temporal scales, exploring evolution and dynamic similarities and dissimilarities over multiple years. Given the temporal characteristics of the dataset the observations can be pushed beyond to frame a temporal network model of the French cattle trade, which might be very useful in designing big data epidemic simulators (can in fact be the backbone of such a simulator in the French perspective). It might also help in understanding global trends of evolving livestock movements and their implications in transporting zoonoses, a future big data epidemiological model. We observed that for FDCM distributions of network parameters becomes similar when the time aggregation window were 16 weeks or more, in the static framework. While in the dynamical framework, the similarities (in terms of repetition of links and nodes, measured by short range similarity) between consecutive time windows were comparatively small. This brings out a question, what is the suitable time length of observation good enough for a network study of epidemic risk. In this work we used data containing information for 5 years, was that sufficient enough? Given the observed stability of the descriptors, the answer is yes from one side. The static network descriptors may be used to simulate networks to study consequences of epidemic spread [Keeling and Eames, 2005]. Whereas from a dynamic network point of view, this dataset does not appear sufficient, the longer the period, the better may be the estimates based on temporal networks, as short time window estimates of temporal network based descriptors may not be a good representative of the frequency and sequence of temporal events. Static and temporal approach of network formalisms may lead to different conclusions, which may be compared using any epidemiological model as local dynamics of the nodes, including the BVDV model.

Over time, even though the network topology was being retained (distributions of network descriptors were not significantly different), the animal exchanges were mostly in non-repeating links. This is important from the point of view of designing control measures as the links of the network were dynamically switching nodes, although the topological properties remained rather stable. The idea of a percolation study was to estimate the minimum number of farms to be disconnected from the trade network to make the epidemic proxy as small as possible. Temporal analysis identified a significant difference in epidemic burden between two regions with regional specificities in terms of dominating herd types, whereas static network based epidemic proxies failed to show any significant effect of the regional specificities. In designing restrictive control strategies the previous years data could be used to identify the key nodes (and links) to be restricted [Kiss et al., 2006, Büttner et al., 2013, Büttner et al., 2014, Frössling et al., 2014].

5.3.2 BVDV spread over time and space

The choice of the within herd BVDV spread model of [Ezanno et al., 2007] was justified by the fact that it constituted the best option as it had comparatively lesser number of variables and parameters (less complexity) still considering the very important aspect of managed structure of the dairy herds. Implementation and computational resources also were comparatively less expensive than its immediate neighbour in the line [Viet et al., 2004]. The model was improved in terms of programming aiming at efficiency in computational resources. Given the socio-economic importance of the BVDV [Houe, 2003, Fourichon et al., 2005, Heuer et al., 2007] and risks posed by animal exchanges between farms [Álvarez et al., 2011, Tinsley et al., 2012, Gates and Woolhouse, 2014] and neighbouring relationships [Ersbøll et al., 2010] necessitates to study the spread of the pathogen in a metapopulation of interacting subpopulations [Courcoul and Ezanno, 2010].

From the pathogen transportation point of view, contact networks of movement of individuals greatly influence the spread and persistence of epidemics [Keeling, 2005, Keeling et al., 2010, Jesse and Heesterbeek, 2011, Altizer et al., 2011, Danon et al., 2011]. This is equally valid for human as well as non-human populations may the modes and scales of transportation of hosts and pathogens be different [Colizza et al., 2006, Kao et al., 2007, Balcan et al., 2009, Keeling et al., 2010, Altizer et al., 2011]. Incorporation of different modes and scales in a model may be very challenging and computationally intensive depending on the complexity of the disease dynamics and modes of spreading. But there exist approaches where information from real data are used in understanding and predicting epidemic processes, even in the big data scenario of global pandemic simulator like [Broeck et al., 2011]. Similar simulators are in fact possible for livestock diseases over different scales of space, time and modes. Unification of contact patterns and disease dynamics over multiple scales of space, time and modes is the requirement here. We observe some developments in this direction recently for livestock diseases [Ortiz-Pelaez et al., 2006a, Kiss et al., 2006, Kao et al., 2007, Courcoul and Ezanno, 2010, Ersbøll et al., 2010, Álvarez et al., 2011, Natale et al., 2011, Polansky and Wittemyer, 2011, Tinsley et al., 2012, Ensoy et al., 2014, Brooks-Pollock et al., 2014]. Although unification of contact networks and multiscale heterogeneous disease dynamics is rarely seen, these recent trends are cues for unified models. Following the trend, in this thesis we present a model of pathogen spread moulding four levels of dynamics in space and time. Although the application here is limited to a specific pathogen in bovine populations in a particular type of farming system (bovine viral diarrhoea virus in dairy herds, to be specific), its formalism is quite generic, and it could be applied to study other pathogen spread on a network and multiple levels of interactions, local, global and different time scales.

We explored spread of BVDV in a metapopulation of dairy herds in a region, applied to the 'Finistère département' in Western France. The metapopulation we studied contained 2,846 dairy farms. We had to incorporate two levels of interactions between herds: (i) exchange of animals via animal trade and (ii) semi-local neighbourhood relations at pastures. For the first level we used the real animal exchange data from FDCM network of cattle trade. Using this dataset enables the realisation of consequences of actual movements. But it does not allow to see the consequences if movements were other than those observed or if there were some restrictions of movements applied. To assess these consequences, a model of animal movements is desirable, which may be built in the same footage of the topological properties of the observed network. It will allow experimentation about the consequences of movements taking place in different possible situations/scenarios. For this, one may consider studying the propagation of the epidemic process over simulated networks [Keeling and Eames, 2005], providing options to dig the alternatives within limits of simulation of networks [Ammar, 2005]. It is not that only animals exchanged between herds can spread a pathogen. Animals of neighbouring farms may exchange pathogen at common boundaries or resources. To include neighbourhood into the model, we identified the neighbours of each farm based on the geographical coordinates of the farms. Since the coordinates were of the location of the farm building, it could not represent the distribution of the pastures belonging to the farm, as the pastures may be anywhere within a few kilometres of the locality and not necessarily contiguous. Therefore we had to assume that the pastures were randomly distributed within a cut-off radius around a farm. Tessellation methods [Ersbøll et al., 2010] is sometimes a good method to use in defining the neighbourhoods, which allows alternative methods of defining the neighbours including randomised and preferential approaches of choosing neighbours. Given the high density of farms in the region we studied (average 41 neighbours for a radius of 5km), it was a good approximation to consider the radial distance approach over complex tessellation methods. Neighbourhood relationships were considered to be semi-local interactions and incorporated into the local dynamics. Therefore, the within-herd model was modified to take care of the possible risk of introducing BVDV posed by neighbours during the pasturing season (mid-March to mid-November). The extension from BVDV spread within a herd to a metapopulation of herds involved the encapsulation of the local (and semi-local neighbourhood) dynamics by the cattle movement network. Another possibility here would have been to use all the data from the database (deaths, births, calving) along with the movements and model only the spread of the pathogen stochastically. This approach would restrict the study into a specific scenario instead of vast possibilities of outcomes when a dynamic model of herd is used. Similarly, instead of considering the animal level epidemiological statuses, herd level statuses and movements either from the data [Tinsley et al., 2012] or simulated networks could have been modelled. Since the persistence and spread of slow spreading pathogens like BVDV are very much dependent on the herd structure and contacts between groups within a herd [Ezanno et al., 2008], herd level statuses are difficult to define. On the other hand, the techniques of simulating real networks are still semi-open problems among the scientific community [Ammar, 2005]. Real metapopulations are not isolated from the 'rest of the world', there is always a risk of multiple entry of a pathogen into any of its subpopulations. Therefore, it is utmost necessary to account for the extrenal risks. In the presented metapopulation model multiple introduction/exchange of infectious animals were possible depending on prevalence of BVDV in the metapopulation through both animal exchange and neighbourhood transmission routes.

The presented model incorporates individual specialities of its predecessors [Courcoul and Ezanno, 2010, Ersbøll et al., 2010, Tinsley et al., 2012], in modelling BVDV spread among dairy herds in a region relying on real movements and geo-location information to simulate the dynamics of the disease, which is unique so far to our knowledge. The beauty of the model lies in its ability to perform efficient evaluation of the field measurable outputs which will be used to examine efficacy of control/eradication strategies [Lindberg et al., 2006, Presi et al., 2011] of the pathogen, considering it as an open-system. Some fine tuning of the assumptions in the model might raise this model to be an efficient tool of computational evaluation of BVDV dynamics and control strategies on a regional scale. So far there is no such tool available (to our knowledge, till date) to ex-ante evaluation of the control strategies at large spatio-temporal scales. This could be the first integrated multiscale model to estimate and validate BVDV control strategies in dairy herds. The construction of the model is such that, given the availability of a specific intra-herd dynamics specific to breed of cattle or regional specificity and movement data, it should be equally applicable across breeds and regions.

From the computation point of view, the structure of the model is such that simulations could be arranged to fractionally run in parallel. It can be split into any number of available computing threads and could efficiently be used even for very large metapopulations with necessary optimization in coding and data input.

5.4 Perspectives

The model may very well be used to test current and innovative control strategies. The generic nature of accounting for between herd interactions makes the model open to include beef herds with very little modification given a relevant within herd model is available. Indeed, the local dynamics in beef herds will have to be taken care by a within-herd BVDV spread model accounting for beef herd management practices. This is important as dairy and beef herds may share a given region. Given sufficient computational resources are available, the model can be extended to larger geographical scales. Keeping these additional gear ups in mind the model is a step towards developing a tool for animal health managers.

References

[Albert and Barabási, 2002] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97.

[Altizer et al., 2011] Altizer, S., Bartel, R., and Han, B. A. (2011). Animal migration and infectious disease risk. *science*, 331(6015):296–302.

[Álvarez et al., 2011] Álvarez, L. G., Webb, C. R., and Holmes, M. A. (2011). A novel field-based approach to validate the use of network models for disease spread between dairy herds. *Epidemiol Infect*, 139(12):1863–1874.

[Ammar, 2005] Ammar, M. (2005). Why we still don't know how to simulate networks. In *Simulation Symposium, 2005. Proceedings. 38th Annual*, page 3. IEEE.

[Bajardi et al., 2011] Bajardi, P., Barrat, A., Natale, F., Savini, L., and Colizza, V. (2011). Dynamical patterns of cattle trade movements. *PLoS One*, 6(5):e19869.

[Bajardi et al., 2012] Bajardi, P., Barrat, A., Savini, L., and Colizza, V. (2012). Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface*, 9(76):2814–2825.

[Baker, 1987] Baker, J. (1987). Bovine viral diarrhea virus: a review. *Journal of the American Veterinary Medical Association*, 190(11):1449.

[Balcan et al., 2009] Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489.

[Ball et al., 2014] Ball, F., Britton, T., House, T., Isham, V., Mollison, D., Pellis, L., and Tomba, G. S. (2014). Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, (0):–.

[Barrat et al., 2008] Barrat, A., Barthelemy, M., and Vespignani, A. (2008). *Dynamical processes on complex networks*, volume 1. Cambridge University Press Cambridge.

[Barrat et al., 2004] Barrat, A., Barthlemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proc Natl Acad Sci U S A*, 101(11):3747–3752.

[Barthélemy et al., 2005] Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. (2005). Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications*, 346(1):34–43.

[Beaudeau et al., 2001] Beaudeau, F., Assi, S., Seegers, H., Belloc, C., Sellal, E., and Joly, A. (2001). Assessing the within-herd prevalence of cows antibody-positive to bovine viral diarrhoea virus with a blocking elisa on bulk tank milk. *Vet Rec*, 149(8):236–240.

[Beaudeau et al., 2005] Beaudeau, F., Fourichon, C., Robert, A., Joly, A., and Seegers, H. (2005). Bulk milk somatic cell counts and bovine viral diarrhoea virus (bvdv) infection in 7252 dairy herds in brittany (western france). *Preventive veterinary medicine*, 72(1):163–167.

[Bennett, 2003] Bennett, R. (2003). The direct costs of livestock disease: the development of a system of models for the analysis of 30 endemic livestock diseases in great britain. *Journal of Agricultural Economics*, 54(1):55–71.

[Bennett et al., 1999] Bennett, R. M., Christiansen, K., and Clifton-Hadley, R. S. (1999). Estimating the costs associated with endemic diseases of dairy cattle. *Journal of Dairy Research*, 66(03):455–459.

[Bigras-Poulin et al., 2006] Bigras-Poulin, M., Thompson, R. A., Chriel, M., Mortensen, S., and Greiner, M. (2006). Network analysis of danish cattle industry trade patterns as an evaluation of risk potential for disease spread. *Prev Vet Med*, 76(1-2):11–39.

[Blonder and Dornhaus, 2011] Blonder, B. and Dornhaus, A. (2011). Time-ordered networks reveal limitations to information flow in ant colonies. *PLoS ONE*, 6(5):e20298.

[Blonder et al., 2012] Blonder, B., Wey, T. W., Dornhaus, A., James, R., and Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, 3(6):958–972.

[Boccara, 2004] Boccara, N. (2004). *Modeling complex systems*, volume 1. Springer.

[Booker et al., 2008] Booker, C. W., Abutarbush, S. M., Morley, P. S., Guichon, P. T., Wildman, B. K., Jim, G. K., Schunicht, O. C., Pittman, T. J., Perrett, T., Ellis, J. A., et al. (2008). The effect of bovine viral diarrhea virus infections on health and performance of feedlot cattle. *The Canadian Veterinary Journal*, 49(3):253.

[Brauer and Castillo-Chavez, 2011] Brauer, F. and Castillo-Chavez, C. (2011). *Mathematical models in population biology and epidemiology.* Springer.

[Brock et al., 1991] Brock, K. V., Redman, D. R., Vickers, M. L., and Irvine, N. E. (1991). Quantitation of bovine viral diarrhea virus in embryo transfer flush fluids collected from a persistently infected heifer. *Journal of Veterinary Diagnostic Investigation*, 3(1):99–100.

[Broeck et al., 2011] Broeck, W. V., Gioannini, C., Gonçalves, B., Quaggiotto, M., Colizza, V., and Vespignani, A. (2011). The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases*, 11(1):37.

[Brooks-Pollock et al., 2014] Brooks-Pollock, E., de Jong, M., Keeling, M., Klinkenberg, D., and Wood, J. (2014). Eight challenges in modelling infectious livestock diseases. *Epidemics*, (0):–.

[Brownlie, 1985] Brownlie, J. (1985). Clinical aspects of the bovine virus diarrhoea. *In practice.*

[Brownlie et al., 1987] Brownlie, J., Clarke, M., Howard, C., and Pocock, D. (1987). Pathogenesis and epidemiology of bovine virus diarrhoea virus infection of cattle. *Ann. Rech. Vet*, 18(2):157–166.

[Buhnerkempe et al., 2013] Buhnerkempe, M. G., Grear, D. A., Portacci, K., Miller, R. S., Lombard, J. E., and Webb, C. T. (2013). A national-scale picture of u.s. cattle movements obtained from interstate certificate of veterinary inspection data. *Preventive Veterinary Medicine*, 112(3):318–329.

[Büttner et al., 2013] Büttner, K., Krieter, J., Traulsen, A., and Traulsen, I. (2013). Efficient interruption of infection chains by targeted removal of central holdings in an animal trade network. *PLoS ONE*, 8(9):e74292.

[Büttner et al., 2014] Büttner, K., Krieter, J., Traulsen, A., and Traulsen, I. (2014). Epidemic spreading in an animal trade network–comparison of distance-based and network-based control measures. *Transboundary and Emerging Diseases*.

[Carlsson et al., 1989] Carlsson, U., Fredriksson, G., Alenius, S., and Kindahl, H. (1989). Bovine virus diarrhoea virus, a cause of early pregnancy failure in the cow. *Journal of Veterinary Medicine Series A*, 36(1-10):15–23.

[Carslake et al., 2011] Carslake, D., Grant, W., Green, L. E., Cave, J., Greaves, J., Keeling, M., McEldowney, J., Weldegebriel, H., and Medley, G. F. (2011). Endemic cattle diseases: comparative epidemiology and governance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1573):1975–1986.

[Carslake et al., 2010] Carslake, D., Jonathan A. K., .-C., Grant, W., Greaves, J., Green, L. E., Keeling, M. J., McEldowney, J. F., Medley, G., and Weldegebriel, H. T. (2010). Animal health and welfare : a case study of science, law and policy in a regulatory environment. *Law, Society and Policy*, Vol.3:227–255.

[Cauvin, 2000] Cauvin, J. (2000). *The Birth of the Gods and the Origins of Agriculture*. Cambridge University Press.

[Centers for Disease Control and Prevention (CDC), 2014] Centers for Disease Control and Prevention (CDC) (2014). Zoonotic disease: When humans and animals intersect (online).

[Centers for Disease Control and Prevention (CDC) et al., 2006] Centers for Disease Control and Prevention (CDC) et al. (2006). Principles of epidemiology in public health practice. *An introduction to applied epidemiology and biostatistics. Self-study course SS1000. 3e éd.*

[Chase, 2013] Chase, C. C. (2013). The impact of bvdv infection on adaptive immunity. *Biologicals*, 41(1):52–60.

[Cherry et al., 1998] Cherry, B., Reeves, M., and Smith, G. (1998). Evaluation of bovine viral diarrhea virus control using a mathematical model of infection dynamics. *Preventive veterinary medicine*, 33(1):91–108.

[Christley et al., 2005] Christley, R., Robinson, S., Lysons, R., French, N., Mellor, D., Russell, A., Wood, J., et al. (2005). Network analysis of cattle movement in great britain. In *Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a meeting held at Nairn, Inverness, Scotland, 30th March-1st April 2005.*, pages 234–244. Society for Veterinary Epidemiology and Preventive Medicine.

[Clauset et al., 2009] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

[Colizza et al., 2006] Colizza, V., Barrat, A., Barthlemy, M., and Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci U S A*, 103(7):2015–2020.

[Courcoul and Ezanno, 2010] Courcoul, A. and Ezanno, P. (2010). Modelling the spread of bovine viral diarrhoea virus (bvdv) in a managed metapopulation of cattle herds. *Veterinary microbiology*, 142(1):119–128.

[Csardi and Nepusz, 2006] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

[Damman et al., 2014] Damman, A., Viet, A.-F., Arnoux, S., Guerrier-Chatellet, M.-C., Petit, E., and Ezanno, P. (2014). Modeling the spread of bovine viral diarrhoea virus (bvdv) in a beef cattle herd and its impact on herd productivity. *Under review*.

[Danon et al., 2011] Danon, L., Ford, A., House, T., Jewell, C., Keeling, M., Roberts, G., Ross, J., and Vernon, M. (2011). Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011:284909:28.

[Diekmann et al., 2012] Diekmann, O., Heesterbeek, H., and Britton, T. (2012). *Mathematical Tools for Understanding Infectious Diseases Dynamics*. Princeton University Press.

[Diekmann and Heesterbeek, 2000] Diekmann, O. and Heesterbeek, J. (2000). *Mathematical epidemiology of infectious diseases*, volume 146. Wiley, Chichester.

[Dorjee et al., 2013] Dorjee, S., Revie, C., Poljak, Z., McNab, W., and Sanchez, J. (2013). Network analysis of swine shipments in ontario, canada, to support disease spread modelling and risk-based disease management. *Preventive Veterinary Medicine*, 112(12):118 – 127.

[Dorogovtsev and Mendes, 2013] Dorogovtsev, S. N. and Mendes, J. F. (2013). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press.

[Dubé et al., 2010] Dubé, C., Ribble, C., and Kelton, D. (2010). An analysis of the movement of dairy cattle through 2 large livestock markets in the province of ontario, canada. *Can Vet J*, 51(11):1254–1260.

[Dubé et al., 2009] Dubé, C., Ribble, C., Kelton, D., and McNab, B. (2009). A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound Emerg Dis*, 56(3):73–85.

[Dubé et al., 2011] Dubé, C., Ribble, C., Kelton, D., and McNab, B. (2011). Introduction to network analysis and its implications for animal disease modelling. *Revue scientifique et technique (International Office of Epizootics)*, 30(2):425–436.

[Easley and Kleinberg, 2010] Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.

[Eddelbuettel, 2013] Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer.

[Edelstein-Keshet, 1987] Edelstein-Keshet, L. (1987). *Mathematical models in biology*, volume 46. Siam.

[Edwards et al., 2000] Edwards, S., Fukusho, A., Lefevre, P.-C., Lipowski, A., Pejsak, Z., Roehe, P., and Westergaard, J. (2000). Classical swine fever: the global situation. *Veterinary microbiology*, 73(2):103–119.

[Ensoy et al., 2014] Ensoy, C., Faes, C., Welby, S., Van der Stede, Y., and Aerts, M. (2014). Exploring cattle movements in belgium. *Preventive veterinary medicine*.

[Ersbøll et al., 2010] Ersbøll, A. K., Ersbøll, B. K., Houe, H., Alban, L., and Kjeldsen, A. (2010). Spatial modelling of the between-herd infection dynamics of bovine virus diarrhoea virus (bvdv) in dairy herds in denmark. *Preventive veterinary medicine*, 97(2):83–89.

[Ersbøll et al., 2003] Ersbøll, A. K., Rugbjerg, H., and Stryhn, H. (2003). Increased mortality among calves in danish cattle herds during bovine virus diarrhoea infection. *ACTA VETERINARIA SCANDINAVICA SUPPLEMENTUM*, pages 224–224.

[Eykhoff, 1974] Eykhoff, P. (1974). *System identification: parameter and state estimation.* Wiley-Interscience London.

[Ezanno et al., 2006] Ezanno, P., Fourichon, C., Beaudeau, F., and Seegers, H. (2006). Between-herd movements of cattle as a tool for evaluating the risk of introducing infected animals. *Animal Research*, 55(3):189–208.

[Ezanno et al., 2008] Ezanno, P., Fourichon, C., and Seegers, H. (2008). Influence of herd structure and type of virus introduction on the spread of bovine viral diarrhoea virus (bvdv) on the spread of bovine viral diarrhoea virus (bvdv) within a dairy herd. *Veterinary research*, 39(5):1.

[Ezanno et al., 2007] Ezanno, P., Fourichon, C., Viet, A.-F., and Seegers, H. (2007). Sensitivity analysis to identify key-parameters in modelling the spread of bovine viral diarrhoea virus in a dairy herd. *Preventive veterinary medicine*, 80(1):49–64.

[Ezanno et al., 2012] Ezanno, P., Vergu, E., Langlais, M., and Gilot-Fromont, E. (2012). Modelling the dynamics of host-parasite interactions: basic principles. In *New Frontiers of Molecular Epidemiology of Infectious Diseases*, pages 79–101. Springer.

[Fagiolo, 2007] Fagiolo, G. (2007). Clustering in complex directed networks. *Phys. Rev. E*, 76:026107.

[FAO-OIE-WHO, 2010] FAO-OIE-WHO (2010). *The FAO-OIE-WHO Collaboration Sharing responsibilities and coordinating global activities to address health risks at the animal-human-ecosystems interfaces.*

[Foster et al., 2010] Foster, J. G., Foster, D. V., Grassberger, P., and Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24):10815–10820.

[Fourichon et al., 2005] Fourichon, C., Beaudeau, F., Bareille, N., and Seegers, H. (2005). Quantification of economic losses consecutive to infection of a dairy herd with bovine viral diarrhoea virus. *Prev Vet Med*, 72(1-2):177–81; discussion 215–9.

[Fray et al., 2000] Fray, M., Paton, D., and Alenius, S. (2000). The effects of bovine viral diarrhoea virus on cattle reproduction in relation to disease control. *Animal Reproduction Science*, 60:615–627.

[Frössling et al., 2014] Frössling, J., Nusinovici, S., Nöremark, M., Widgren, S., and Lindberg, A. (2014). A novel method to identify herds with an increased probability of disease introduction due to animal trade. *Preventive veterinary medicine.*

[Frössling et al., 2012] Frössling, J., Ohlson, A., Björkman, C., Håkansson, N., and Nöremark, M. (2012). Application of network analysis parameters in risk-based surveillance - examples based on cattle trade data and bovine infections in sweden. *Prev Vet Med*, 105(3):202–208.

[Gates et al., 2013] Gates, M., Woolhouse, M., Gunn, G., and Humphry, R. (2013). Relative associations of cattle movements, local spread, and biosecurity with bovine viral diarrhoea virus (bvdv) seropositivity in beef and dairy herds. *Preventive veterinary medicine*.

[Gates and Woolhouse, 2014] Gates, M. C. and Woolhouse, M. E. J. (2014). Suboptimal herd performance amplifies the spread of infectious disease in the cattle industry. *PLoS One*, 9(3):e93410.

[Graham et al., 2013] Graham, D., Clegg, T., Lynch, M., and More, S. (2013). Herd-level factors associated with the presence of bovine viral diarrhoea virus in herds participating in the voluntary phase of the irish national eradication programme. *Preventive veterinary medicine*, 112(1):99–108.

[Grenfell and Harwood, 1997] Grenfell, B. and Harwood, J. (1997). (meta) population dynamics of infectious diseases. *Trends in Ecology & Evolution*, 12(10):395–399.

[Groenendaal et al., 2004] Groenendaal, H., Galligan, D., and Mulder, H. (2004). An economic spreadsheet model to determine optimal breeding and replacement decisions for dairy cattle. *Journal of dairy science*, 87(7):2146–2157.

[Guégan et al., 2005] Guégan, J.-F., Morand, S., and Poulin, R. (2005). Are there general laws in parasite community ecology? the emergence of spatial parasitology and epidemiology. *Parasitism and ecosystems*, pages 22–42.

[Gunn et al., 2004] Gunn, G., Stott, A., and Humphry, R. (2004). Modelling and costing bvd outbreaks in beef herds. *The Veterinary Journal*, 167(2):143–149.

[Hagberg et al., 2008] Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA.

[Hanski, 1999] Hanski, I. (1999). *Metapopulation ecology*, volume 232. Oxford University Press Oxford.

[Hartley and Richards, 1988] Hartley, P. and Richards, M. (1988). A study of the transmission of bovine virus diarrhoea virus between and within cattle herds. *Acta Vet. Scand. Suppl.*, 84:164166.

[Heinrichs, 1993] Heinrichs, A. (1993). Raising dairy replacements to meet the needs of the 21st century. *Journal of dairy science*, 76(10):3179–3187.

[Heuer et al., 2007] Heuer, C., Healy, A., and Zerbini, C. (2007). Economic effects of exposure to bovine viral diarrhea virus on dairy herds in new zealand. *Journal of dairy science*, 90(12):5428–5438.

[Heuer et al., 2008] Heuer, C., Tattersfield, G., West, D., and Olson, W. (2008). Effect of reproductive pathogens on pregnancy rates in beef herds (neospora caninum, bvd-virus & leptospira sv. hardjo & pomona). In *Proceedings of the Society of Sheep and Beef Cattle Veterinarians of the NZVA*, pages 141 – 147.

[Hilbe et al., 2007] Hilbe, M., Stalder, H., Peterhans, E., Haessig, M., Nussbaumer, M., Egli, C., Schelp, C., Zlinszky, K., and Ehrensperger, F. (2007). Comparison of five diagnostic methods for detecting bovine viral diarrhea virus infection in calves. *Journal of Veterinary Diagnostic Investigation*, 19(1):28–34.

[Holme and Saramäki, 2012] Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports*.

[Houe, 1993] Houe, H. (1993). Survivorship of animals persistently infected with bovine virus diarrhoea virus (bvdv). *Preventive Veterinary Medicine*, 15(4):275–283.

[Houe, 1995] Houe, H. (1995). Epidemiology of bovine viral diarrhea virus. *The Veterinary clinics of North America. Food animal practice*, 11(3):521–547.

[Houe, 1999] Houe, H. (1999). Epidemiological features and economical importance of bovine virus diarrhoea virus (bvdv) infections. *Veterinary microbiology*, 64(2):89–107.

[Houe, 2003] Houe, H. (2003). Economic impact of bvdv infection in dairies. *Biologicals*, 31(2):137–143.

[Houe et al., 2006] Houe, H., Lindberg, A., and Moennig, V. (2006). Test strategies in bovine viral diarrhea virus control and eradication campaigns in europe. *Journal of Veterinary Diagnostic Investigation*, 18(5):427–436.

[Houe and Meyling, 1991] Houe, H. and Meyling, A. (1991). Surveillance of cattle herds for bovine virus diarrhoea virus (bvdv)-infection using data on reproduction and calf mortality. In *Ruminant Pestivirus Infections*, pages 157–164. Springer.

[Innocent et al., 1997] Innocent, G., Morrison, I., Brownlie, J., and Gettinby, G. (1997). A computer simulation of the transmission dynamics and the effects of duration of immunity and survival of persistently infected animals on the spread of bovine viral diarrhoea virus in dairy cattle. *Epidemiology and infection*, 119(1):91–100.

[Jesse et al., 2008] Jesse, M., Ezanno, P., Davis, S., and Heesterbeek, J. (2008). A fully coupled, mechanistic model for infectious disease dynamics in a metapopulation: movement and epidemic duration. *Journal of theoretical biology*, 254(2):331–338.

[Jesse and Heesterbeek, 2011] Jesse, M. and Heesterbeek, H. (2011). Divide and conquer? persistence of infectious agents in spatial metapopulations of hosts. *Journal of theoretical biology*, 275(1):12–20.

[Joly et al., 2005] Joly, A., Fourichon, C., and Beaudeau, F. (2005). Description and first results of a bvdv control scheme in brittany (western france). *Preventive veterinary medicine*, 72(1):209–213.

[Kao et al., 2006] Kao, R. R., Danon, L., Green, D. M., and Kiss, I. Z. (2006). Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proc Biol Sci*, 273(1597):1999–2007.

[Kao et al., 2007] Kao, R. R., Green, D. M., Johnson, J., and 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.

[Karsai et al., 2014] Karsai, M., Perra, N., and Vespignani, A. (2014). Time varying networks and the weakness of strong ties. *Scientific reports*, 4.

[Keeling, 2005] Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theor Popul Biol*, 67(1):1–8.

[Keeling et al., 2010] Keeling, M. J., Danon, L., Vernon, M. C., and House, T. A. (2010). Individual identity and movement networks for disease metapopulations. *Proc Natl Acad Sci U S A*, 107(19):8866–8870.

[Keeling and Eames, 2005] Keeling, M. J. and Eames, K. T. D. (2005). Networks and epidemic models. *J R Soc Interface*, 2(4):295–307.

[Keeling and Rohani, 2008] Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press.

[Kelling et al., 2000] Kelling, C., Grotelueschen, D., Smith, D., Brodersen, B., et al. (2000). Testing and management strategies for effective beef and dairy herd bvdv biosecurity programs. *Bovine Practitioner*, 34(1):13–22.

[Kiss et al., 2006] Kiss, I. Z., Green, D. M., and Kao, R. R. (2006). The network of sheep movements within great britain: Network properties and their implications for infectious disease spread. *J R Soc Interface*, 3(10):669–677.

[Knight-Jones and Rushton, 2013] Knight-Jones, T. J. D. and Rushton, J. (2013). The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev Vet Med*, 112(3-4):161–173.

[Konschake et al., 2013] Konschake, M., Lentz, H. H., Conraths, F. J., Hövel, P., and Selhorst, T. (2013). On the robustness of in-and out-components in a temporal network. *PloS one*, 8(2):e55223.

[Krings et al., 2012] Krings, G., Karsai, M., Bernhardsson, S., Blondel, V. D., and Saramäki, J. (2012). Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):1–16.

[Lanyon and Reichel, 2014] Lanyon, S. and Reichel, M. (2014). Bovine viral diarrhoea virus ('pestivirus') in australia: to control or not to control? *Aust Vet J*.

[Lindberg, 2003] Lindberg, A. (2003). Bovine viral diarrhoea virus infections and its control. a review. *Veterinary Quarterly*, 25(1):1–16.

[Lindberg et al., 2006] Lindberg, A., Brownlie, J., Gunn, G. J., Houe, H., Moennig, V., Saatkamp, H. W., Sandvik, T., and Valle, P. S. (2006). The control of bovine viral diarrhoea virus in europe: today and in the future. *Rev Sci Tech*, 25(3):961–979.

[Lindberg and Houe, 2005] Lindberg, A. and Houe, H. (2005). Characteristics in the epidemiology of bovine viral diarrhea virus (bvdv) of relevance to control. *Preventive veterinary medicine*, 72(1):55–73.

[Lindberg and Alenius, 1999] Lindberg, A. L. and Alenius, S. (1999). Principles for eradication of bovine viral diarrhoea virus (bvdv) infections in cattle populations. *Veterinary Microbiology*, 64(23):197 – 222.

[Mars et al., 1999] Mars, M., Bruschke, C., and Van Oirschot, J. (1999). Airborne transmission of bhv1, brsv, and bvdv among cattle is possible under experimental conditions. *Veterinary microbiology*, 66(3):197–207.

[Martínez-López et al., 2009] Martínez-López, B., Perez, A., and Sánchez-Vizcaíno, J. (2009). Social network analysis. review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis*, 56(4):109–120.

[Mazoyer and Roudart, 2006] Mazoyer, M. and Roudart, L. (2006). *A history of world agriculture: from the neolithic age to the current crisis*. Earthscan.

[McClurkin et al., 1984] McClurkin, A., Littledike, E., Cutlip, R., Frank, G., Coria, M., and Bolin, S. (1984). Production of cattle immunotolerant to bovine viral diarrhea virus. *Canadian Journal of Comparative Medicine*, 48(2):156.

[Meier-Schellersheim et al., 2009] Meier-Schellersheim, M., Fraser, I. D., and Klauschen, F. (2009). Multiscale modeling for biologists. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(1):4–14.

[Meyling and Mikel Jensen, 1988] Meyling, A. and Mikel Jensen, A. (1988). Transmission of bovine virus diarrhoea virus (bvdv) by artificial insemination (ai) with semen from a persistently-infected bull. *Veterinary microbiology*, 17(2):97–105.

[Moennig et al., 2003] Moennig, V., Floegel-Niesmann, G., and Greiser-Wilke, I. (2003). Clinical signs and epidemiology of classical swine fever: a review of new knowledge. *The Veterinary Journal*, 165(1):11–20.

[Moerman et al., 1993] Moerman, A., Straver, P., De Jong, M., Quak, J., Baanvinger, T., and Van Oirschot, J. (1993). A long term epidemiological study of bovine viral diarrhoea infections in a large herd of dairy cattle. *Veterinary Record*, 132(25):622–626.

[Moilanen and Hanski, 1998] Moilanen, A. and Hanski, I. (1998). Metapopulation dynamics: effects of habitat quality and landscape structure. *Ecology*, 79(7):2503–2515.

[Mollison and Levin, 1995] Mollison, D. and Levin, S. (1995). Spatial dynamics of parasitism. *Ecology of infectious diseases in natural populations*, pages 384–398.

[Muñoz-Zanzi et al., 2004] Muñoz-Zanzi, C. A., Thurmond, M. C., and Hietala, S. K. (2004). Effect of bovine viral diarrhea virus infection on fertility of dairy heifers. *Theriogenology*, 61(6):1085–1099.

[Murray, 2002] Murray, J. D. (2002). *Mathematical Biology I and II.* Springer, New York, NY, USA,.

[Mweu et al., 2013] Mweu, M. M., Fourni, G., Halasa, T., Toft, N., and Nielsen, S. S. (2013). Temporal characterisation of the network of danish cattle movements and its implication for disease control: 2000-2009. *Prev Vet Med.*

[Natale et al., 2009] Natale, F., Giovannini, A., Savini, L., Palma, D., Possenti, L., Fiore, G., and 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.

[Natale et al., 2011] Natale, F., Savini, L., Giovannini, A., Calistri, P., Candeloro, L., and Fiore, G. (2011). Evaluation of risk and vulnerability using a disease flow centrality measure in dynamic cattle trade networks. *Prev Vet Med*, 98(2-3):111–118.

[Newman, 2010] Newman, M. (2010). *Networks: An Introduction.* Oxford University Press, Inc., New York, NY, USA.

[Newman et al., 2006] Newman, M., Barabasi, A.-L., and Watts, D. J. (2006). *The structure and dynamics of networks.* Princeton University Press.

[Newman, 2002] Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, 89(20):208701.

[Nieuwhof and Bishop, 2005] Nieuwhof, G. and Bishop, S. (2005). Costs of the major endemic diseases of sheep in great britain and the potential benefits of reduction in disease impact. *Animal Science*, 81(01):23–29.

[Niskanen and Lindberg, 2003] Niskanen, R. and Lindberg, A. (2003). Transmission of bovine viral diarrhoea virus by unhygienic vaccination procedures, ambient air, and from contaminated pens. *The veterinary journal*, 165(2):125–130.

[Nöremark et al., 2011] Nöremark, M., Håkansson, N., Lewerin, S. S., Lindberg, A., and 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):78–90.

[Nöremark and Widgren, 2014] Nöremark, M. and Widgren, S. (2014). Epicontacttrace: an r-package for contact tracing during livestock disease outbreaks and for risk-based surveillance. *BMC Vet Res*, 10(1):71.

[Opsahl and Panzarasa, 2009] Opsahl, T. and Panzarasa, P. (2009). Clustering in weighted networks. *Social networks*, 31(2):155–163.

[Ortiz-Pelaez et al., 2006a] Ortiz-Pelaez, A., Pfeiffer, D., Soares-Magalhaes, R., and Guitian, F. (2006a). Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (fmd) epidemic in the uk. *Preventive veterinary medicine*, 76(1):40–55.

[Ortiz-Pelaez et al., 2006b] Ortiz-Pelaez, A., Pfeiffer, D. U., Soares-Magalhes, R. J., and Guitian, F. J. (2006b). Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (fmd) epidemic in the uk. *Prev Vet Med*, 76(1-2):40–55.

[O'Sullivan and Unwin, 2003] O'Sullivan, D. and Unwin, D. J. (2003). *Geographic information analysis*. John Wiley & Sons.

[Pasman et al., 1994] Pasman, E., Dijkhuizen, A., and Wentink, G. (1994). A state-transition model to stimulate the economics of bovine virus diarrhoea control. *Preventive Veterinary Medicine*, 20(4):269–277.

[Perra et al., 2012] Perra, N., Gonalves, B., Pastor-Satorras, R., and Vespignani, A. (2012). Activity driven modeling of time varying networks. *Sci Rep*, 2:469.

[Pioz et al., 2012] Pioz, M., Guis, H., Crespin, L., Gay, E., Calavas, D., Durand, B., Abrial, D., and Ducrot, C. (2012). Why did bluetongue spread the way it did? environmental factors influencing the velocity of bluetongue virus serotype 8 epizootic wave in france. *PLoS One*, 7(8):e43360.

[Polansky and Wittemyer, 2011] Polansky, L. and Wittemyer, G. (2011). A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data. *J R Soc Interface*, 8(56):322–333.

[Presi et al., 2011] Presi, P., Struchen, R., Knight-Jones, T., Scholl, S., and Heim, D. (2011). Bovine viral diarrhea (bvd) eradication in switzerlandexperiences of the first two years. *Preventive veterinary medicine*, 99(2):112–121.

[Qu et al., 2011] Qu, Z., Garfinkel, A., Weiss, J. N., and Nivala, M. (2011). Multi-scale modeling in biology: how to bridge the gaps between scales? *Progress in biophysics and molecular biology*, 107(1):21–31.

[R Core Team, 2014] R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

[Ram et al., 2008] Ram, K., Preisser, E. L., Gruner, D. S., and Strong, D. R. (2008). Metapopulation dynamics override local limits on long-term parasite persistence. *Ecology*, 89(12):3290–3297.

[Rautureau et al., 2011] Rautureau, S., Dufour, B., and 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.

[Rautureau et al., 2012] Rautureau, S., Dufour, B., and Durand, B. (2012). Structural vulnerability of the french swine industry trade network to the spread of infectious diseases. *Animal*, 6(7):1152–1162.

[Saatkamp et al., 2000] Saatkamp, H., Berentsen, P., and Horst, H. (2000). Economic aspects of the control of classical swine fever outbreaks in the european union. *Veterinary microbiology*, 73(2):221–237.

[Sivula et al., 1996] Sivula, N., Ames, T., Marsh, W., and Werdin, R. (1996). Descriptive epidemiology of morbidity and mortality in minnesota dairy heifer calves. *Preventive Veterinary Medicine*, 27(3):155–171.

[Smith and Grotelueschen, 2004] Smith, D. R. and Grotelueschen, D. M. (2004). Biosecurity and biocontainment of bovine viral diarrhea virus. *Veterinary Clinics of North America: Food Animal Practice*, 20(1):131–149.

[Sørensen et al., 1995] Sørensen, J. T., Enevoldsen, C., and Houe, H. (1995). A stochastic model for simulation of the economic consequences of bovine virus diarrhoea virus infection in a dairy herd. *Preventive Veterinary Medicine*, 23(3):215–227.

[Squartini et al., 2012] Squartini, T., Picciolo, F., Ruzzenenti, F., and Garlaschelli, D. (2012). Reciprocity of weighted networks. *arXiv preprint arXiv:1208.4208.*

[Ståhl and Alenius, 2012] Ståhl, K. and Alenius, S. (2012). Bvdv control and eradication in europe–an update. *Jpn J Vet Res*, 60 Suppl:S31–S39.

[Stroustrup, 2013] Stroustrup, B. (2013). *The C++ Programming Language (4th Edition).* Addison-Wesley.

[Svensson et al., 2006] Svensson, C., Linder, A., and Olsson, S.-O. (2006). Mortality in swedish dairy calves and replacement heifers. *Journal of dairy science*, 89(12):4769–4777.

[Tarry et al., 1991] Tarry, D., Bernal, L., and Edwards, S. (1991). Transmission of bovine virus diarrhoea virus by blood feeding flies. *The Veterinary Record*, 128(4):82–84.

[The Merck veterinary manual, 2014] The Merck veterinary manual (2014). Zoonotic Diseases, The Merck veterinary manual (online), Merck & Company.

[Thompson et al., 2002] Thompson, D., Muriel, P., Russell, D., Osborne, P., Bromley, A., Rowland, M., Creigh-Tyte, S., and Brown, C. (2002). Economic costs of the foot and mouth disease outbreak in the united kingdom in 2001. *Rev Sci Tech*, 21(3):675–687.

[Tildesley et al., 2011] Tildesley, M. J., Volkova, V. V., and Woolhouse, M. E. (2011). Potential for epidemic take-off from the primary outbreak farm via livestock movements. *BMC Vet Res*, 7:76.

[Tinsley et al., 2012] Tinsley, M., Lewis, F. I., and Brülisauer, F. (2012). Network modeling of bvd transmission. *Veterinary research*, 43(1):11.

[Tremblay, 1996] Tremblay, R. (1996). Transmission of bovine viral diarrhea virus. *Veterinary medicine.*

[Vernon, 2011] Vernon, M. C. (2011). Demographics of cattle movements in the united kingdom. *BMC Vet Res*, 7:31.

[Vernon and Keeling, 2009] Vernon, M. C. and Keeling, M. J. (2009). Representing the uk's cattle herd as static and dynamic networks. *Proc Biol Sci*, 276(1656):469–476.

[Viet et al., 2007] Viet, A., Fourichon, C., and Seegers, H. (2007). Review and critical discussion of assumptions and modelling options to study the spread of the bovine viral diarrhoea virus (bvdv) within a cattle herd. *Epidemiology and infection*, 135(5):706–721.

[Viet et al., 2006] Viet, A.-F., Fourichon, C., and Seegers, H. (2006). Simulation study to assess the efficiency of a test-and-cull scheme to control the spread of the bovine viral-diarrhoea virus in a dairy herd. *Preventive veterinary medicine*, 76(3):151–166.

[Viet et al., 2004] Viet, A.-F., Fourichon, C., Seegers, H., Jacob, C., and Guihenneuc-Jouyaux, C. (2004). A model of the spread of the bovine viral-diarrhoea virus within a dairy herd. *Preventive veterinary medicine*, 63(3):211–236.

[Viet and Jacob, 2008] Viet, A.-F. and Jacob, C. (2008). Illustration of some limits of the markov assumption for transition between groups in models of spread of an infectious pathogen in a structured herd. *Theoretical population biology*, 74(1):93–103.

[Wasserman, 1994] Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

[Wilson and Mellor, 2009] Wilson, A. J. and Mellor, P. S. (2009). Bluetongue in europe: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530):2669–2681.

[Wittum et al., 2001] Wittum, T., Grotelueschen, D. M., Brock, K., Kvasnicka, W. G., Floyd, J., Kelling, C. L., and Odde, K. (2001). Persistent bovine viral diarrhoea virus infection in us beef herds. *Preventive veterinary medicine*, 49(1):83–94.

[World Health Organisation, 2014] World Health Organisation (2014). *Zoonoses and the Human-Animal-Ecosystems Interface.*

Appendix I: Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period

Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period

Bhagat Lal Dutta[a,b,c], Pauline Ezanno[a,b], Elisabeta Vergu[c]

[a]*INRA, UMR1300 Biologie, Epidémiologie et Analyse de Risque en santé animale, CS 40706, F-44307 Nantes, France*
[b]*LUNAM Université, Oniris, Ecole nationale vétérinaire, agroalimentaire et de l'alimentation Nantes Atlantique, UMR BioEpAR, F-44307 Nantes, France*
[c]*INRA, UR341 Unité Mathématiques et Informatique Appliquées, Domaine de Vilvert, 78352 Jouy-en-Josas cedex, France*

Abstract

A good knowledge of the specificities of the animal trade network is highly valuable to better control pathogen spread on a large regional to transnational scale. Because of their temporal dynamical nature, studying multi-annual datasets is particularly needed to investigate whether structural patterns are stable over the years. In this study, we analysed the French cattle movement network from 2005 to 2009 for different spatial granularities and temporal windows, with the three-fold objective of exploring temporal variations of the main network characteristics, computing proxies for pathogen spread on this network, which accounts for its time-varying properties and identifying specificities related to the main types of animals and farms (dairy versus beef). Network properties did not qualitatively vary among different temporal and spatial granularities. About 40% of the holdings and 80% of the communes were directly interconnected. The width of the aggregation time window barely impacted normalised distributions of indicators. A period of 8 to 16 weeks would suffice for robust estimation of their main trends, whereas longer periods would provide more details on tails. The dynamic nature of the network could be seen through the small overlap between consecutive networks with 65% of common active nodes for only 3% of common links over 2005-2009. To control pathogen spread on such a network, by reducing the largest strongly connected component by more than 80%, movements should be prevented from 1 to 5% of the holdings with the highest centrality in the previous year network. The analysis of breed-wise and herd-wise subnetworks, dairy, beef and mixed, reveals similar trends in temporal variation of average indicators and their distributions. The link-based backbones of beef subnetworks seem to be more stable over time than those of other subnetworks. At a regional scale, node reachability accounting for time-respecting paths, as proxy of epidemic burden, is greater for a dairy region than for a beef region. This highlights the importance of considering local specificities and temporal dynamics of animal trade networks when evaluating control measures of pathogen spread.

Keywords: Network analysis, livestock movements, epidemiological modelling, surveillance, dynamic networks, disease control

---

1. Introduction

Uncovering of network structures has been of utmost importance from the point of view of many applications (Newman, 2010), specifically by providing insights for controlling pathogen spread (Keeling and Eames, 2005; Danon et al., 2011). Usually, the livestock movement datasets are represented as networks (with farms as nodes and trade relationships as links) and analysed using the methodologies of graph theory and social network analysis (Dubé et al., 2009). In addition to the fact that they are directed (flows of traded animals are most often not symmetrical) and weighted (to each trade link between two farms, a number of animals or batches is associated), these networks are also time-varying (i.e. two active farms exchange animals at specific time points, connections are not permanent).

The studies in this field range from investigation of demographic structure and pathogen dynamics (Kao et al., 2006; Natale et al., 2009) to analysis of dynamical patterns in longitudinal data (Bajardi et al., 2011) and surveillance optimization (Bajardi et al., 2012) based on farm animal trade network data. Different animal markets in several countries were explored: cattle in Denmark, Canada, France, Italy, Sweden and UK (Bigras-Poulin et al., 2006; Dubé et al., 2010; Rautureau et al., 2011; Bajardi et al., 2011; Nöremark et al., 2011; Vernon, 2011), sheep in UK (Kiss et al., 2006; Kao et al., 2006) and pigs in Sweden, France, Germany and Canada (Nöremark et al., 2011; Rautureau et al., 2012; Büttner et al., 2013; Dorjee et al., 2013). Specifically, to cite only a few examples, various characteristics relating, on the one hand, networks representing animal trade and, on the other hand, pathogen spread, such as vulnerability to and effects of infectious diseases, were investigated based on these data. The network of movements of cattle in France in 2005 was studied to identify its vulnerability to the spread of pathogens (Rautureau et al., 2011). Demographic patterns in the movement of cattle in the UK were explored and effects of the foot and mouth disease outbreaks in 2001 and 2007 on the flow of animals were detected using data across the years 1999-2009 (Vernon, 2011). A more recent work studied the national network of cattle movements in the US using veterinary inspection data at the state borders (Buhnerkempe et al., 2013) and showed that a county level approach is the most appropriate for examining processes influenced by cattle shipments, such as economic analyses and disease outbreaks.

Concerning the temporal dynamical characteristics of animal trade data, there are at least two distinct primary approaches to address movement networks: (i) as static networks, by considering temporal aggregated variants of the initial datasets (Kiss et al., 2006; Kao et al., 2006; Rautureau et al., 2011) or (ii) as continuous increments of small width time windows (Vernon and Keeling, 2009; Büttner et al., 2013; Bajardi et al., 2011). Most of the recent studies use a mix of both approaches. The UK cattle movement network was analysed to identify the indicators valid both in dynamic and static network approximations (Vernon and Keeling, 2009). The longitudinal dataset of cattle movements in Italy for 2007 was studied to identify dynamical patterns at various time scales, from one day to the whole year (Bajardi et al., 2011). The French cattle movements in 2005 were also studied at monthly and weekly time scales (Rautureau et al., 2011). Cattle and pig movements in Sweden between 2006 and 2008 were investigated as monthly and yearly networks (Nöremark et al., 2011) and a measure for assessing disease control strategies, accounting for temporally compatible sequences of movements, was proposed. The Danish cattle movements were studied for a period of 10 years (2000-2009): the (static) network characteristics were calculated at monthly time-aggregation level and their evolution followed over the whole period (Mweu et al., 2013). The nodes of a time-varying network relevant from the point of view of pathogen spread were identified and the approach was tested on the German database on pig trade (Konschake et al., 2013).

In the context of time-varying networks, a question of particular interest concerns the spatial and temporal granularities. Is it possible to simplify the network of animal movements, in terms of number of nodes and links (by considering broader boundaries for node definition) and temporal dynamics (by considering aggregated snapshots), while preserving the information it encompasses? Another important aspect is related to the potential impact of farm type and animal breed on the topological and temporal characteristics of the network.

Our objective was to analyse the French cattle movement network on a period ranging from 2005 to 2009, from both spatial and temporal perspectives. More specifically, we aimed at (i) providing an extensive description of this database for different spatial granularities (holdings and administrative communes as nodes) and temporal windows (from monthly to yearly); (ii) investigating the stability of network characteristics over consecutive and increasing time windows; (iii) computing proxies for the outcome of disease spread on the network representing these data, when accounting for its time-varying properties; and (iv) exploring in what extent (i)-(iii) lead to different interpretations when considering subsets of data based on the main characteristics of farms and animal breeds.

2. Materials and methods

2.1. Database and data processing

In France, the Ministry of Agriculture maintains the French Database of Cattle Movements (FDCM), which records the life history of every animal from birth to death including movements between holdings (i.e. farms, markets, and assembling centres), and to slaughterhouses, as well as imports and exports. For the present study we use the FDCM database from 2005 to 2009 to construct the network of movements. The raw data set has five layers of information. The first layer contains the detention record in various holdings and has 11 fields of information identifying country code of the animal, national animal identification number, holding identification number, cause of entry into the holding (birth, purchase), date of entry, cause of exit from the holding (death, sale), date of exit, administrative department information of the holding, department of origin, destination department and country of origin (if imported). The second layer states the identity of a bovine with country code, national identification number, date of birth, place (farm) of birth, sex, race and date of first calving (for females). The other three layers contain information about markets, assembling centres and slaughterhouses. Out of the five layers, we have not included in the present study the information about the slaughterhouses as in the network they act only as sinks and do not play a direct role in the potential spread of pathogens. A series of cleaning processes were performed on this set of data, based on matching of entries among the layers to obtain a filtered data set comprising the information about origin and destination (holdings, communes and departments of the holdings, and types of holdings), sex, race, date of birth and date of movements of animals identified by their national animal identification number. Each line in this dataset represents one movement made at a particular date. In the final form of the dataset we do not consider the imports and exports and restrict our study to the movements within France only. We used SAS 9.2 software in the extraction and cleaning process of the FDCM.

Based on the cleaned dataset, several classifications required in further analysis were made. At the animal level, five age classes were defined: class 1 ($0-30$ days), class 2 ($1-8$ months), class 3 ($8-20$ months), class 4 ($20-30$ months) and class 5 (> 30 months), based on practices of cattle exchange among farms (Ezanno et al., 2006). Herd types were defined on a yearly basis, using average number of animals per gender, racial type (beef, dairy, crossed/mixed) and age class present in farms. A herd is labelled as *dairy (beef)* if it has more than 15 dairy (beef) animals in age class 5 and no animal of the other type. If it has both types of animals the herd is considered *mixed* (Ezanno et al., 2006). Markets and assembling centres are not handled separately as they have similar properties (e.g. no internal herd structure for breeds or age classes, short stay periods etc.). Cattle movements were analysed at two spatial (holdings and administrative communes as entities) and three temporal (monthly=4 weeks, quarterly=13 weeks and yearly=52 consecutive weeks) scales of accumulation.

2.2. *Network analysis*

The movement data extracted from the FDCM and preprocessed contains the information on the animal exchanges between holdings underlying a *network* of contacts. Using network terminology, holdings become *nodes* and animals moving from one holding to another connect these nodes through *links*. These links are *directed* as a consequence of non symmetrical trade flows. A link can also be *weighted*, where the weight is defined either as the number of transactions (batches) carried out or as the number of movements (animals) exchanged between the two holdings connected by the link during the time window under consideration. A node is said *active* with reference to a particular time window if it had at least one in or out link (i.e. it exchanged at least one animal) during the period considered. In addition to networks including all active nodes and movements in a given time window, five subnetworks based on animal breed (*beef*, *dairy* and *mixed*) on one hand, and on herd type and animal breed simultaneously (*BEEF* and *DAIRY*), on the other hand, were analysed. A *breed-wise* subnetwork consists in all the movements of animals of a particular breed, irrespective of the types of source and destination holdings A *herd-wise* network includes all the movements of animals of a particular breed (e.g. dairy) having as source and destination herds of the same type (e.g. dairy), markets or assembly centres.

Two approaches were taken: a classical static approach where the links are considered to be present during a given period of observation, and a temporal approach, where the fact that the links are active only at specific time points is taken into account (Newman, 2010; Holme and Saramäki, 2012). The investigations were performed for the global network and for the five breed-wise and herd-wise subnetworks. For the static approach, a variety of indicators (means and distributions) were calculated based on holdings or communes as nodes, and for monthly and yearly periods. The temporal analysis of the networks concerned the study of the stability of connexions over time, the influence of time aggregation window on network main features and their evolution over increasing time snapshots. Proxies for pathogen spread on networks were computed from a static (for quarterly and yearly aggregated networks) perspective for the national networks and subnetworks. The dynamical view on the risk of propagation was implemented for two regional networks (French Brittany - a dairy region, and Limousin - a beef region). All the indicators used are defined in the next subsection.

The Kolmogorov-Smirnov (K-S) statistical test was used for comparison of empirical distributions and the Bonferroni correction was applied in the case of multiple testing (e.g. when comparing pairs of annual empirical distributions). To fit distributions and estimate their slopes using power-law fitting, the recipe of Clauset et al. (2009) was implemented. Network analysis was performed with packages igraph 0.7.1 (R 3.1)(R Core Team, 2014; Csardi and Nepusz, 2006) and networkx 1.8.1 (Python 2.7) (Hagberg et al., 2008).

2.2.1. *Static networks: classical indicators for directed weighted networks*

The *directed weighted networks* $W = \{w_{ij}\}$, where w_{ij} is the number of movements from holding i to holding j (0 if no movement), were analysed based on classical indicators of networks adapted to account for directed and weighted links, when these variants were proposed in the literature (Newman, 2002; Barrat et al., 2004; Barthélemy et al., 2005; Martínez-López et al., 2009). The unweighted counterpart of W is $A = \{a_{ij}\}$, with $a_{ij} = 1$ if there is at least one movement from i to j and 0 otherwise.

Node and link characteristics were investigated using the following indicators. The *degree (k)* is the number of holdings a holding is connected to (the *out-degree* k^{out} is the number of holdings to which at least one cattle is sent; the *in-degree* k^{in} represents the number of holdings from which at least one animal is received). This writes as $k_i^{out} = \sum_j a_{ij}$, $k_i^{in} = \sum_j a_{ji}$, $k_i = k_i^{in} + k_i^{out}$. The *degree correlation* is measured as Spearman's rank correlation between out-degrees and in-degrees of nodes (here between selling and buying abilities of holdings). The *strength (s)* is the number of movements per holding (animals sent and received): *out-strength* $s_i^{out} = \sum_j w_{ij}$; *in-strength* $s_i^{in} = \sum_j w_{ji}$; and $s_i = s_i^{in} + s_i^{out}$. The *average degree of nearest neighbours*, $k_{nn}^w(k)$ if weighted and $k_{nn}(k)$ if unweighted, writes as the average over all nodes of degree k of $k_{nn,i}^w = 1/s_i \sum_j w_{ij} k_j$ and $k_{nn,i} = 1/k_i \sum_j a_{ij} k_j$, respectively. It is related to the correlation between the degrees of linked nodes (here between the number of partners for connected holdings). The *centrality measures* allow ranking the nodes according to their 'importance' in a network. Besides the degree and strength as measures of centrality, already mentioned, the *closeness centrality* is formally defined as $C_{Cl}(i) = \sum_j \frac{1}{\lambda_{ij}}$ (where λ_{ij} is equal to infinity if no path between i and j). The *betweenness centrality* is defined as $C_B(i) = \sum_{j \neq i \neq l} \sigma_{jl}(i)/\sigma_{jl}$. where $\sigma_{jl}(i)$ is the number of shortest paths between j and l passing through i, and σ_{jl} is their total number.

Network cohesion, in terms of local density, connectivity, assortativity and mixing was explored by calculating several indicators. The *average weighted clustering coefficient,* $C^w(k)$, measures the average cohesiveness based on animal exchange intensities around holdings having k neighbours, where, for holding i, $c^{w,i} = 1/(s_i(k_i - 1)) \sum_{j,h} a_{ij} a_{ih} a_{jh} (w_{ij} + w_{ih})/2$. $C(k)$ is its unweighted counterpart. The *shortest path* (λ_{ij} represents its length) is the shortest trajectory (without travelling the same node more than once) between two nodes i and j. For weighted graphs, it is the path for which the sum of the weights of the constituent links is minimum. The *diameter* is the largest shortest path among all the shortest paths in the network. In our context it represents the maximum lowest number of intermediaries between two holdings over all the indirectly connected pairs of holdings. The *giant strongly connected component (GSCC)* is the largest subnetwork where for every pair of nodes i and j, there is a directed path from i to j and a directed path from j to i. Here, it represents the set of holdings that an animal can enter from any node of this set. The *assortativity* is calculated as the Pearson correlation coefficient between linked nodes degrees. The *reciprocity* is the ratio of total reciprocated weight (number of animal movements for reciprocated links between holdings) to the total weight (total number of animals exchanged) of the network (Squartini et al., 2013).

2.2.2. *Measures for assessing time-varying characteristics of networks*

In order to account for dynamical aspects of the network, the stability over time of indicators rescaled distributions was investigated (Krings et al., 2012). Their similarity was measured by L^2 *distances*, defined as the distance between the rescaled distributions of networks aggregated over an interval Δt and networks aggregated over $2\Delta t$. Activation and deactivation of nodes and links over continuous windows were measured by *short range similarity (SRS)*, which is defined as the fraction of links (similarly calculated for nodes also) common to two consecutive aggregation windows, averaged over all the equally large windows. From the perspective of time-varying networks, we also looked at an extension of the concept of *activity* x_i of node i (Perra et al., 2012), which we defined here as the ratio between the number of animal movements concerning holding i in a particular time window and the total number of movements during the same period.

2.2.3. *Proxies for pathogen spread and its control for static and time-stamped networks*

To evaluate the underlying capacity of the network of favouring pathogen spread, two tools were used: the iterative calculation of GSCCs for embedded networks through a *percolation* analysis and the *reachability ratio (RR)*, the second being appropriate for time-varying networks. The percolation allows to investigate the effect of node (or link) removal on the existence and the size of the GSCC, quantified in terms of the ratio between GSCC sizes after and before removal. In the context of pathogens spreading through cattle movements, percolation allows evaluating the efficacy of targeted control strategies based on the removal of holdings and of their connections (e.g. direct removal through ban on livestock movements, or indirect action through vaccination). The RR can be calculated as the average fraction of nodes in the sets of influence of all nodes. The set of influence of node i comprises nodes that can be reached from i by time respecting paths (i.e. sequences of contacts with non-decreasing times that connect sets of nodes) during the time window considered (Holme and Saramäki, 2012). It is also possible to calculate the maximum RR over all nodes (as starting points for time respecting paths). The reachability ratio may be considered as the temporal counterpart of the GSCC.

3. Results

3.1. Dataset description for the period 2005-2009

We counted $274,231$ holdings exchanging $24,485,015$ animals through $40,357,979$ movements in $12,421,180$ batches. 55% of these movements were made by male animals. In terms of race, the contributions were 49%, 33% and 18% by beef, dairy and crossed breeds respectively. The participation of age groups were, 27%, 15%, 27%, 6% and 24% for age classes 1 to 5. The share of different herd types as source (seller) and destination (buyer) in terms of links (transactions) and flows (cattle movements) is shown in Table 1.

Table 1: Share of the active holdings in terms of herd type as source and destination and corresponding fractions of cattle movements, in France from 2005 to 2009. In addition to BEEF and DAIRY herds, MIXED farms (containing both dairy and beef production units) are considered. The category OTHER includes very small farms (< 15 animals in age class 5), insemination centres etc. Markets and assembling centres are grouped into the same category M/C.

	Herd type as destination (%)	32.79	20.18	28.60	17.79	0.63	218,061 (total number of herds as destination)
Herd type as source (%)	Destination → Source ↓	**BEEF**	**DAIRY**	**MIXED**	**OTHER**	**M/C**	Movements per herd type as source (%)
30.92	**BEEF**	3.78	0.31	2.09	0.35	19.29	25.83
21.60	**DAIRY**	0.48	2.83	1.54	0.15	10.05	15.04
27.02	**MIXED**	1.45	1.69	2.42	0.29	16.79	22.62
19.98	**OTHER**	0.30	0.15	0.39	0.08	0.77	1.70
0.49	**M/C**	5.62	7.69	4.52	0.42	16.57	34.81
260,324 (total number of herds as source)	Movements per herd type as destination (%)	11.63	12.66	10.95	1.3	63.46	40,357,979 (total number of movements)

Among the five subnetworks defined, those based on animal race drive important fractions of the whole network : beef, dairy and mixed breed-wise subsets represent 0.75, 0.60, 0.73, respectively, in terms of nodes, 0.41, 0.49, 0.29 in terms of links and 0.49, 0.33 and 0.18 of global flows, for the period from 2005 to 2009. Beef and dairy herd-wise networks cover 0.36 and 0.26 respectively of the total active nodes, 0.19 and 0.21 of the total links and 0.33 and 0.19 of total flows.

3.2. Description of yearly aggregated data

The basic characteristics of the FDCM after year wise (364 days) splitting of the movements are shown in Table 2. A decrease in the number of active holdings between consecutive years can be noticed, as well as a corresponding reduction in the number of movements $(0.1 - 9\%)$ and animals $(0.8 - 5.3\%)$. However, the ratios between movements and animals traded are relatively stable over the calendar years: annually on average $23 - 24$ animals moved per holding with each animal performing ~ 1.54 movements. The number of connections undergoes the most important drop, of almost 23% in four years. The last four lines of Table 2 show that the cattle trade network is not balanced, in the sense that there are more sellers than buyers. Concerning the flows for different types of origin-destination pairs of

Table 2: Summary of cattle movements among holdings in France from 2005 to 2009, based on the FDCM. The yearly values are shown as percentage in change with respect to observations in 2005.

	01Jan05-30Dec05 (2005)	**31Dec05-29Dec06 (2006)**	**30Dec06-28Dec07 (2007)**	**29Dec07-26Dec08 (2008)**	**27Dec08-25Dec09 (2009)**
Total number of holdings	243,324	-3.7	-7.5	-11.7	-14.7
No of communes	29,515	-0.8	-1.2	-2.2	-2.6
No of movements (W)	8,636,018	-0.1	-9.2	-11.7	-12.2
No of links (A)	1,279,576	-9.3	-13.1	-19.8	-22.8
No of batches	2,791,261	-4.7	-12.8	-17.7	-20.4
No of cattle	5,533,854	-0.8	-6.1	-7.6	-9.6
No of holdings as origin	228,400	-3.6	-7.8	-12.3	-15.2
No of holdings as destination	141,249	-3.4	-7.7	-12.3	-16.1
No of communes as origin	28,995	-0.9	-1.7	-2.6	-3.3
No of communes as destination	26,164	-1.2	-1.5	-3.4	-4.5

holdings, the three largest parts of animal trade go from farms to assembling centres ($\sim 39-41\%$ of total movements), followed by direct farm to farm connections ($17-19\%$), and assembling centres to farms movements ($14-15\%$). The proportion of farms, assembling centres, and markets among holdings is very stable between years (99.46%, 0.5%, and 0.04%, respectively). Distributions in terms of source-destination pairs of holdings, sex and age-classes of exchanged animals, are not significantly different over the years ($p > 0.05$). Movements of young-stock (age class 1) and of animals of age class 3 each represent on average 27% of the total flows.

3.3. Average indicators for static networks based on holdings and communes as nodes, and for yearly and monthly aggregation periods

At the national level, for yearly aggregated networks, irrespective to herd type or animal breed, the average values of indicators are quite stable from one year to the next for both holdings and communes as nodes. For the network based on holdings, average nodes characteristics such as in- and out-strength vary between 60-64 and 37-39, respectively, for 2005 to 2009. Network cohesion is also quite stable over years: the GSCCs contain $0.42, 0.42, 0.41, 0.40$, and 0.39 of the total number of nodes in the annual networks from 2005 to 2009; and other indicators such as assortativity (-0.06 to -0.11), clustering coefficient (0.0051-0.0054), reciprocity (0.17- 0.19), average path length (4.7-4.9), and diameter (21-25) exhibit little variation. When aggregating the transactions of all holdings in a commune and considering each commune as a node, the GSCCs comprise $0.85, 0.85, 0.84, 0.84$, and 0.82 of the total active communes for 2005 to 2009. All distances on networks globally decrease, assortativity and clustering coefficient also are close to zero. The main change compared to the holdings based network, not necessarily directly related to the aggregation effect, is the increase in reciprocity (0.23-0.25).

Ranges of variation of the main monthly indicators for the global network, calculated for the directed unweighted variants, over the whole period studied, are provided in Fig. 1. Globally, similar fluctuating behaviours are noticed for the majority of indicators, irrespective to the level of aggregation for node definition (holdings or communes). Despite a decreasing trend, a yearly pattern in the number of active nodes is visible, with a ratio between upward (March) and downward (August) peaks of about 20% (Fig. 1a). A more severe downward peak is visible in 2007 and in the same year, conversely, the average path length achieves a noticeable increase. There is almost no clustering and the assortativity has low negative values (Fig. 1d, e). The proportion of bi-directional links is also low (reciprocity less than 0.13 and 0.19 for holdings and communes, respectively; Fig. 1f). The proportion of nodes belonging to the GSCC is, on average, 6 to 7 times greater when communes are the nodes of the network (Fig. 1h). The most significant difference with respect to the level of aggregation is illustrated by the correlation of degrees: it is negative when nodes are holdings and positive when nodes are communes (Fig. 1i). Here again, the most important variation, especially for the node-based network, is observed in 2007. The analysis of monthly herd-wise and breed-wise subnetworks reveals variations that could be assimilated to annual cycles and suggests more generally characteristics similar to those of

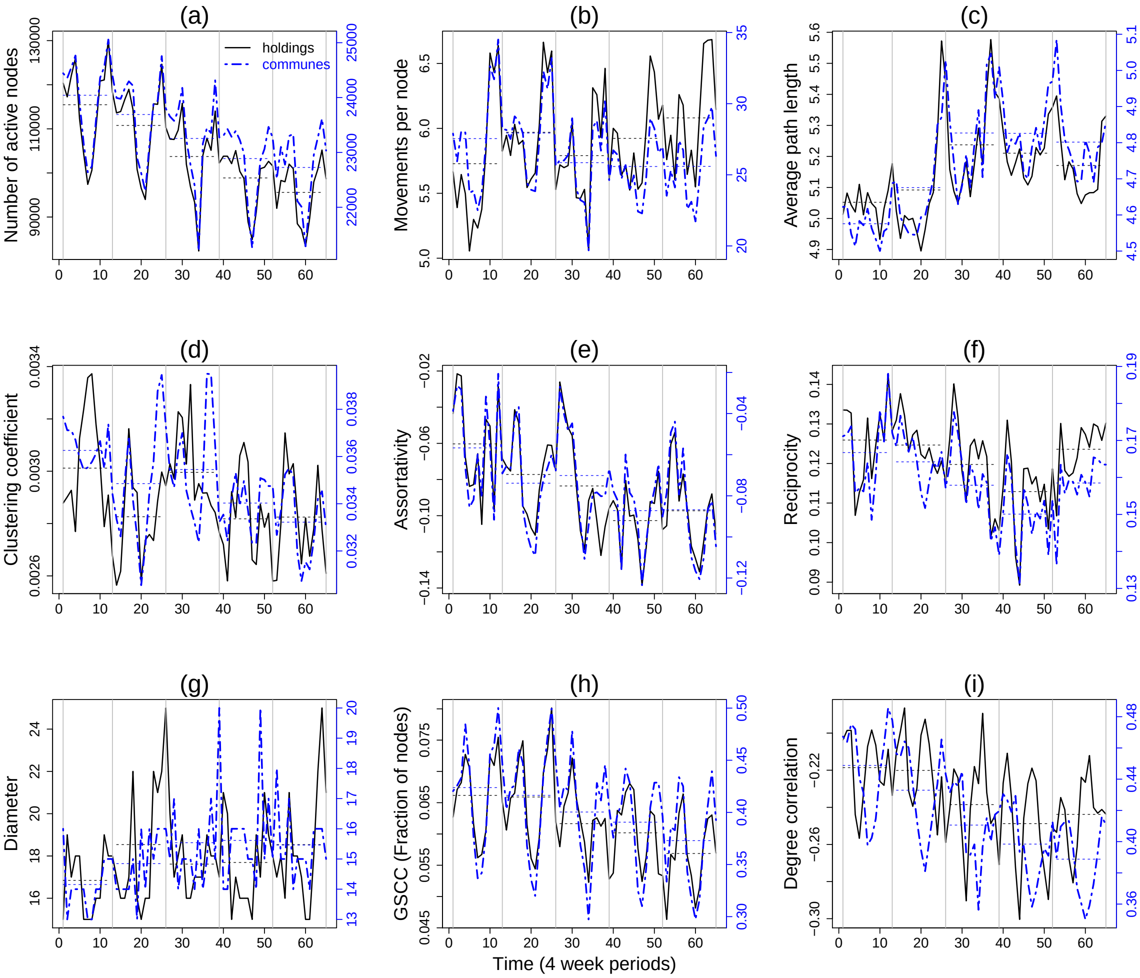


Figure 1: Classical network indicators (labelled on the y-axes of each panel from (a) to (i)) for the national full network of cattle movements in 4-week time windows in France from January 2005 to December 2009. Calculations are performed using the directed unweighted variant of networks. Two types of entities are considered as nodes: holdings - herds, markets and assembly centres (black) and administrative communes (dashed blue). Averages over years for each indicator are represented (horizontal dotted lines). 364-day windows are marked by vertical grey lines.

the global network (Fig. 2). Race-wise subnetworks include almost twice more nodes than herd-wise subnetworks (Fig. 2a). A noticeable difference between beef and dairy subnetworks concerns the number of movements per node (Fig. 2b), which is uniformly larger over the whole period. Clustering coefficient and assortativity are close to zero (Fig. 2d, e). The reciprocity and the cohesion expressed through the GSCC are two to three times more important for beef than for diary subnetworks (Fig. 2f).

3.4. *Distribution of centrality measures and other features for the networks with holdings as nodes, and for monthly and yearly aggregation periods*

In the annual time window, the degree distributions for the aggregated networks with holdings as nodes and all movements are not significantly different from each other ($p > 0.05$). The slopes of the degree distributions when fitted to a power-law are 1.8 (in-degree) and 1.5 (out-degree), and those for the cumulative distributions of strength

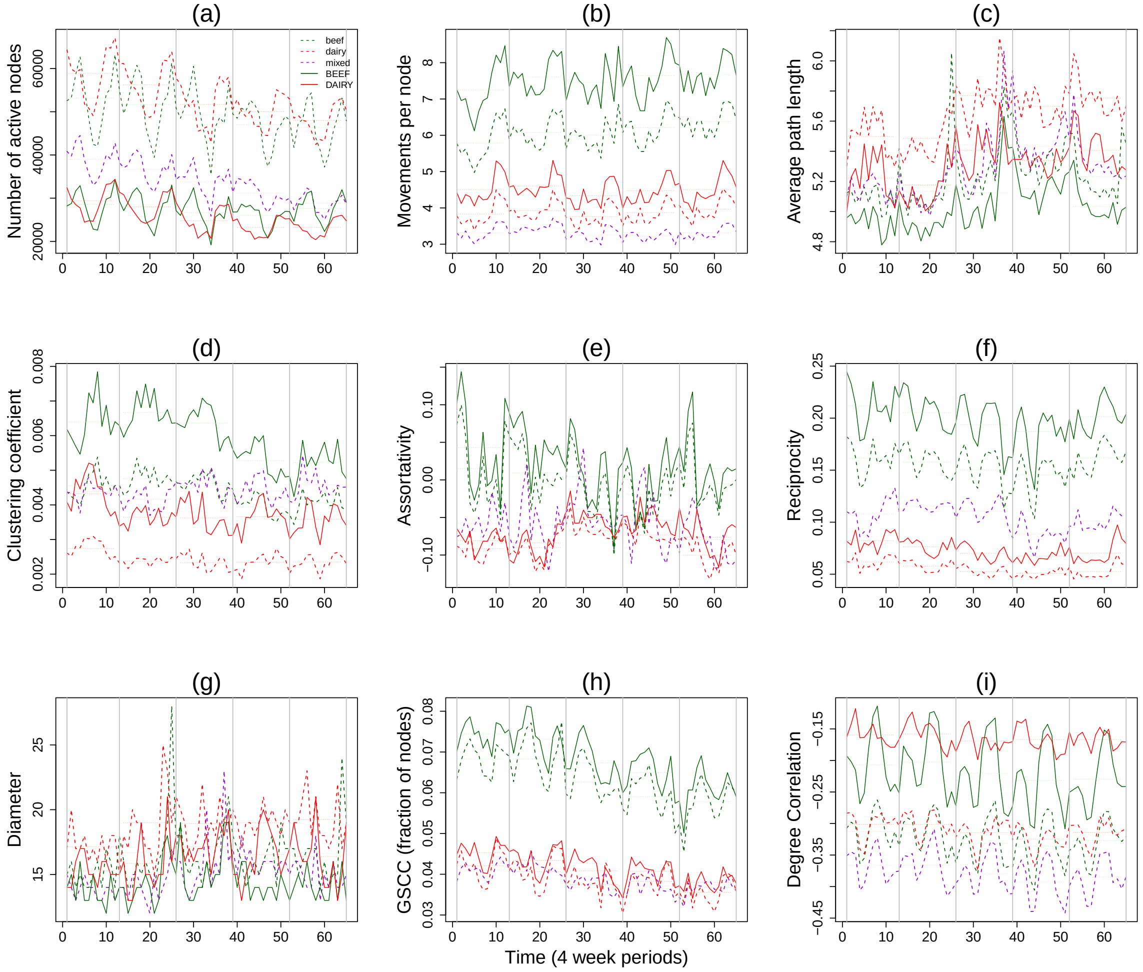


Figure 2: Classical network indicators (labelled on the y-axes of each panel from (a) to (i)) for the subnetworks based on animal breed and herd type for the national network of cattle movements in 4-week time windows in France from January 2005 to December 2009. Subnetworks consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). Markets and assembling centres are part of the subnetworks. Calculations are performed using the directed unweighted variant of subnetworks with holdings as nodes. Averages over years for each indicator are represented (horizontal dotted lines). 364-day windows are marked by vertical grey lines.

are 1.7 (in-strength) and 1.5 (out-strength). The distributions of betweenness and closeness also show no significant difference at different levels of time aggregation (data not shown). For monthly networks, exponents of degree and strength distributions (Fig. 3) are rather consistent over time. Monthly exponents range between 2.3 and 2.9 for in-degree and in-strength distributions (Fig. 3c, f), and between 1.5 and 2 for out-degree and out-strength distributions (Fig. 3c, f), respectively, for both holdings and communes as nodes. When separately analysing monthly degree and strength distributions for beef and dairy networks, they appear quite similar, except for the in-strength for which monthly exponents are around 2.5 for beef and around 3 for dairy herd-based subnetwork (data not shown).

Local features of network cohesion are captured by the distributions of clustering coefficient and average nearest-neighbour degree, both weighted and unweighted variants, when considering the global network with holdings as nodes (Fig. 4, data for 2009 as an example, consistent results for the other years). While clustering coefficients exhibit

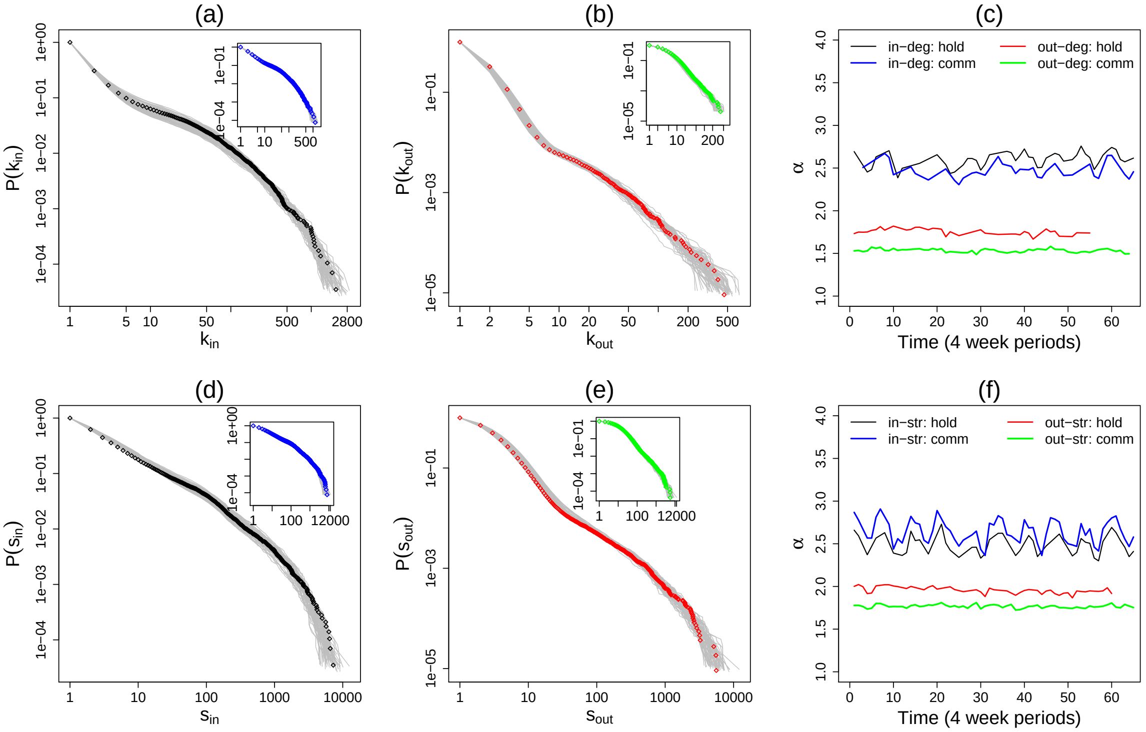


Figure 3: Cumulative distribution of degrees (number of holdings in contact with each node); in-degrees (a), out-degrees (b)) and strengths (number of movements per node; (in-strength (d); out-strength (e)) in monthly (4-week time periods) full networks at national scale, with holdings as nodes for major graphs, and communes as nodes for insets. Coloured dots are for January 2005, the first month of the the data set, spanning until December 2009 and including all cattle movements in France. A power law fit of the distributions ((c) and (f)) shows a range of exponents for each indicator (only months with a statistically significant estimate are kept).

a continuously decreasing trend (Fig. 4a), average nearest-neighbour degrees have distributions with a plateau for low values followed by a decreasing phase (Fig. 4b). Accounting for weights modifies the shape of these distributions.

The relationship between geographical distance and shortest path length on network (accounting for directed links) was explored. According to data for 2009 (similar results for the other years), distributions of distance for path lengths from 1 to 3 are not significantly different between subnetworks (Fig. 5a-c). At least 85% of directly connected farms are at a distance lower than 200 kms and they exchange preferentially within the same commune, irrespective to the type of network (peak for very low values in Fig. 5a). When connected through one intermediate node on the global network, holdings are more likely to be located at 40 km from each other and 95% of them are at less than 504 km Fig. (5b). For all subnetworks, the geographical distance for a shortest path equal to three is in 95% of cases equal at most 633 km (Fig. 5c). For direct connections, the mean distance is 77 km for beef, 107 km for dairy herd-wise subnetworks and 94 for the global network, whereas medians are 43 km, 64 km and 50 km, respectively.

3.5. Network dynamics: influence of increasing aggregation time window on network indicators

The impact of increasing the aggregation time window is investigated through in- and out- degree and strength distributions for the global network, both unscaled and scaled (normalized by the corresponding average value). Rescaling leads to a collapse of the different distributions into a single pattern (Fig. 6), which is noticed through the stabilisation of L^2 distances between distributions for doubling time windows (data not shown). Differences are mostly located at right tails (for large values). Distributions of activity also shows independence of time window of accumulation (Fig. 6e).

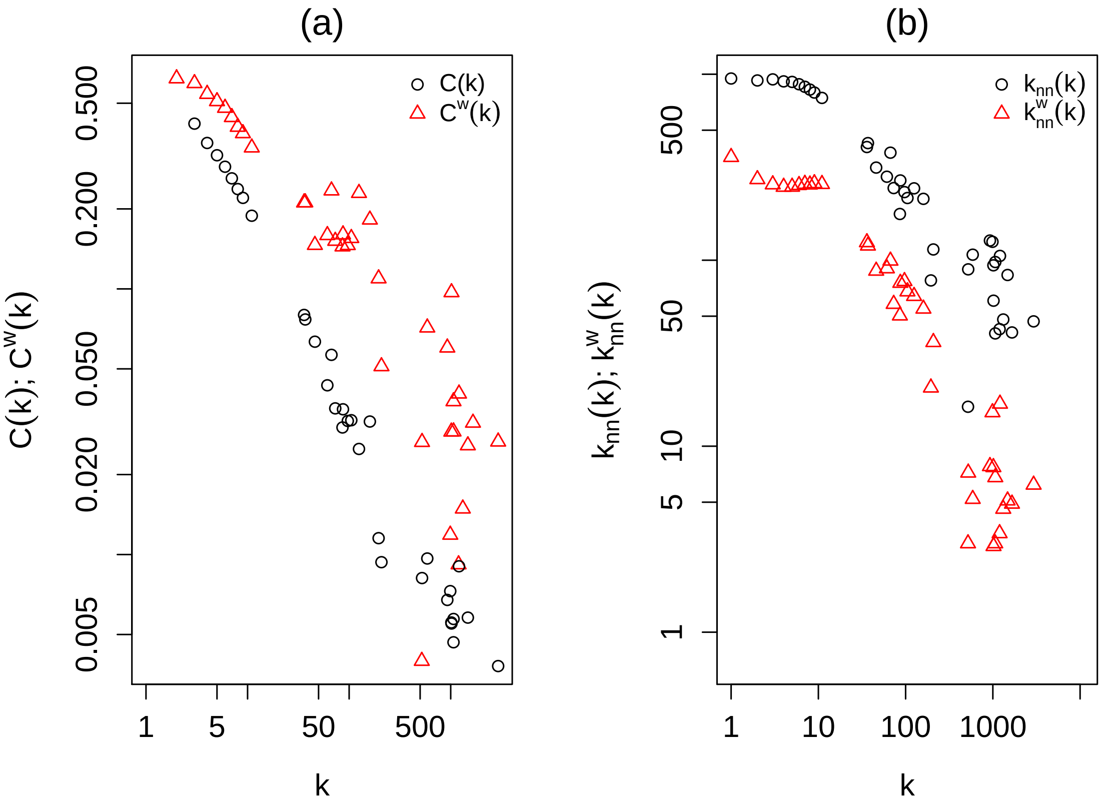


Figure 4: Distributions of unweighted (black dots) and weighted (red triangles) clustering coefficients (a) and nearest-neighbour degrees (b), averaged over neighbours of each node (holding) of degree k. All cattle movements at the national level, in France, during 2009 are considered.

The SRS also provides a proxy for the importance of the time period of observation in connection with the stability of the backbone. The SRS exhibits a saturation behaviour, with the maximum average value almost equal to 0.2 (Fig. 7a; between 0.1 and 0.31 for consecutive years comparison) and 0.8 (Fig. 7b; from 0.75 to 0.82) for link-based and node-based backbone, respectively, for annual global networks, but also for the whole period for which data are available. The year-wise SRS plots show similar behaviour over the years (Fig. 7a, b, insets). The proportion of nodes active every year over all nodes is 0.65, while that of connections occurring each year is only 0.03 of the total number of links during the period 2005-2009. For subnetworks, beef related ones are in general more stable over time than diary-based subnetworks and than the global network in terms of links (Fig. 7c), whereas they are closer for nodes conservation over time (Fig. 7d).

3.6. *Proxies for outbreaks and their control on networks*

Percolation was investigated at the national level, yearly (Fig. 8a) and quarterly (Fig. 8b-e), based on information on the current year and the current trimester (real-time scenario) or the previous year, the previous trimester or the corresponding trimester of the previous year (delayed access to data). The random sampling of nodes for removal has almost no impact on the GSCC size up to a removal of 10% of the total number of active nodes (Fig. 8a-e, data for 2009). When the removal of nodes is based on centrality measures (degree, strength, and betweenness) by targeting the top nodes (ordered in descending order for each of theses three measures), the most efficient strategy is the one

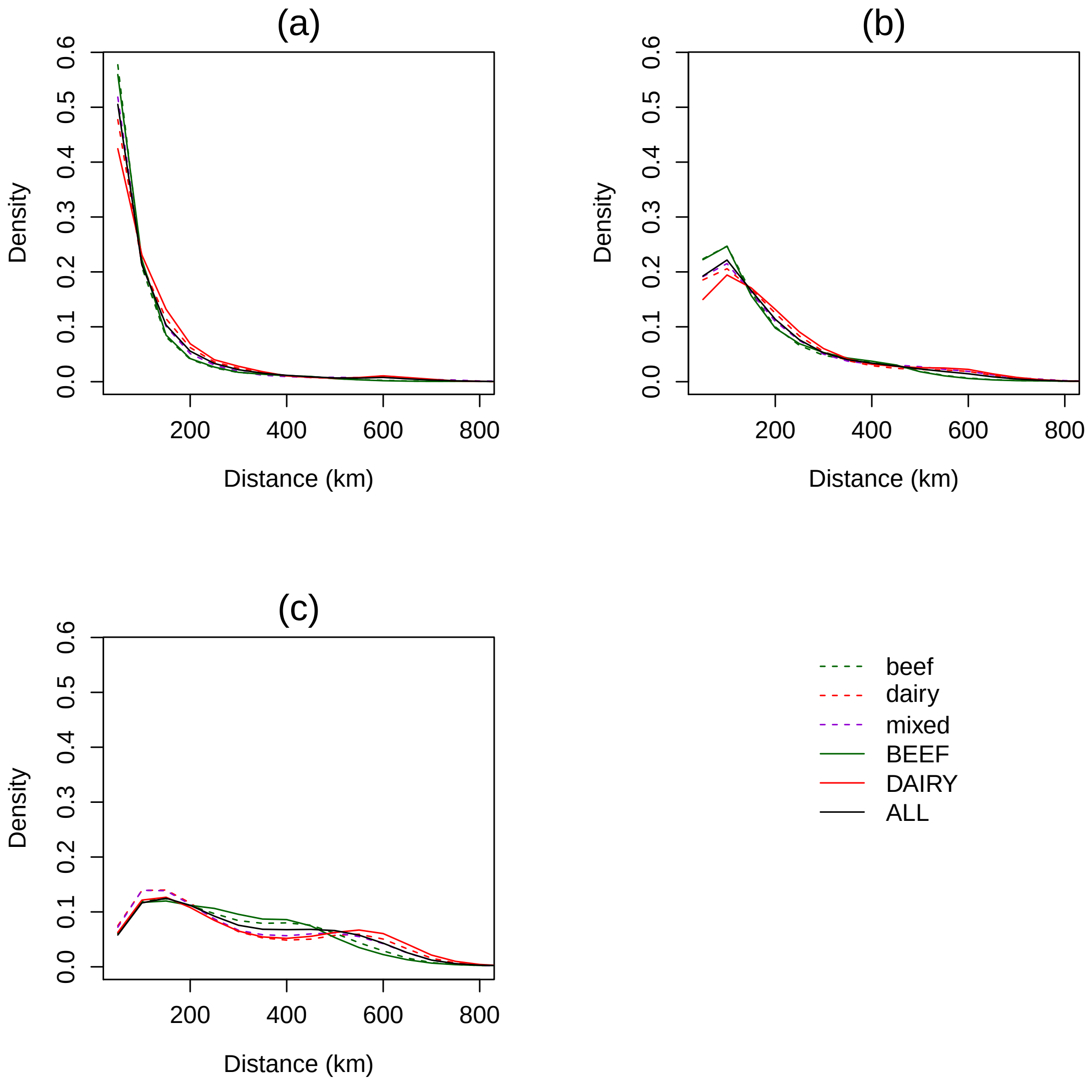


Figure 5: Distributions of geographical distances between pairs of nodes separated by shortest paths of length 1 (a), 2 (b) and 3 (c). Subnetworks consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). The full network is also analysed (black). Markets and assembling centres are part of the subnetworks. All cattle movements at the national level, in France, during 2009 are considered.

based on betweenness, whereas the less efficient one is based on strength for the annual network (Fig. 8a), and on degree or strength for 3-month networks (Fig. 8b-e). Indeed, removing the top 5% of nodes based on betweenness achieves to completely break the GSCC structure for the annual network, this fraction being only 0.5% for 3-month networks. For these networks, there is no significant difference in the results depending on the quarter analysed. Unsurprisingly, the most efficient strategy is based on information from the current network (unlikely scenario, used

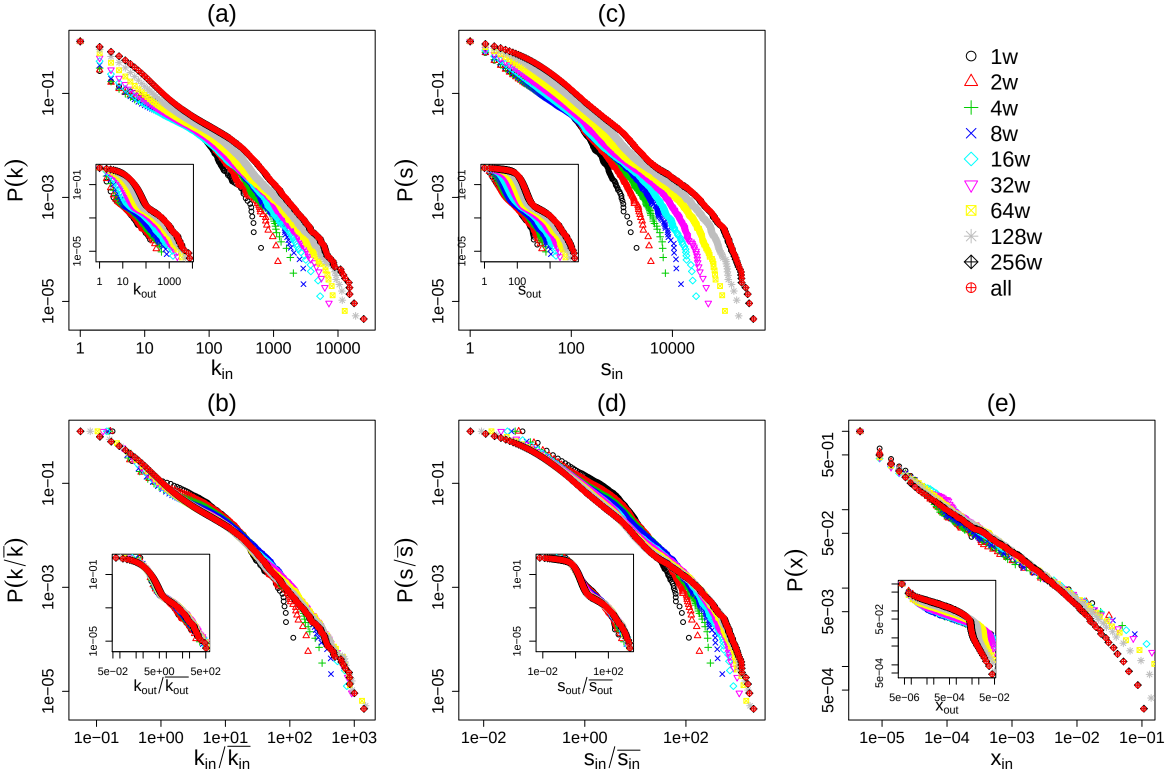


Figure 6: Distributions of in-degrees ((a), rescaled (b)) and out-degrees (inset (a), rescaled inset (b)), in-strength ((c), rescaled (d)) and out-strength (inset (c), rescaled inset (d)) and activity (e), computed on increasing time windows (in number of weeks (w)). All cattle movements from 2005 to 2009 in France are considered, networks are built with holdings as nodes.

here only to provide a lower bound). Removing 5% of the nodes based on their betweenness or degree the previous year enables the GSCC size to be reduced by more than 80%. Moreover and unexpectedly, knowing the characteristics of the previous 3-month network is more efficient than basing targeting according to the network in the same quarter the previous year.

Percolation study was also done for the five race and herd-based subnetworks (Fig. 9). All the subnetworks were found to be comparatively highly resilient to random removal of nodes, compared to targeted removal (Fig. 9a). Subnetworks were found to be more sensitive to node removal than the full network, with the mixed breed subnetwork being most vulnerable. For degree based removal strategy, beef subnetworks (both race and herd-wise) were more vulnerable compared to dairy subnetworks (Fig. 9b). Similarly to the the full network, the betweenness centrality centred node removal strategy is the most efficient in breaking down: all subnetworks almost break-down with top 2% of the nodes removed, whereas for the full network it takes about 5% of the nodes (Fig. 9d).

3.7. *Regional networks: contrasting dairy and beef herds*

Two administrative regions, distinct with respect to the type of farms, were compared: Brittany, with almost half the farms being dairy farms, and Limousin, with more than 57% of beef farms. In terms of animal breed, more than 66% of animals in Brittany are dairy and more than 88% of cattle are beef. Amongst average monthly indicators of regional networks based on holdings as nodes, noticeable differences between the two regions concern the number of active nodes (10000-14000 for Brittany and 2500-5000 for Limousin), the number of movements per node (3.2-4.2 in Brittany and 5-7.5 in Limousin), and the reciprocity (0.02-0.035 in Brittany and 0.1-0.2 in Limousin). Other characteristics are similar with those of the national networks (very weak clustering, small GSCC as fraction of total

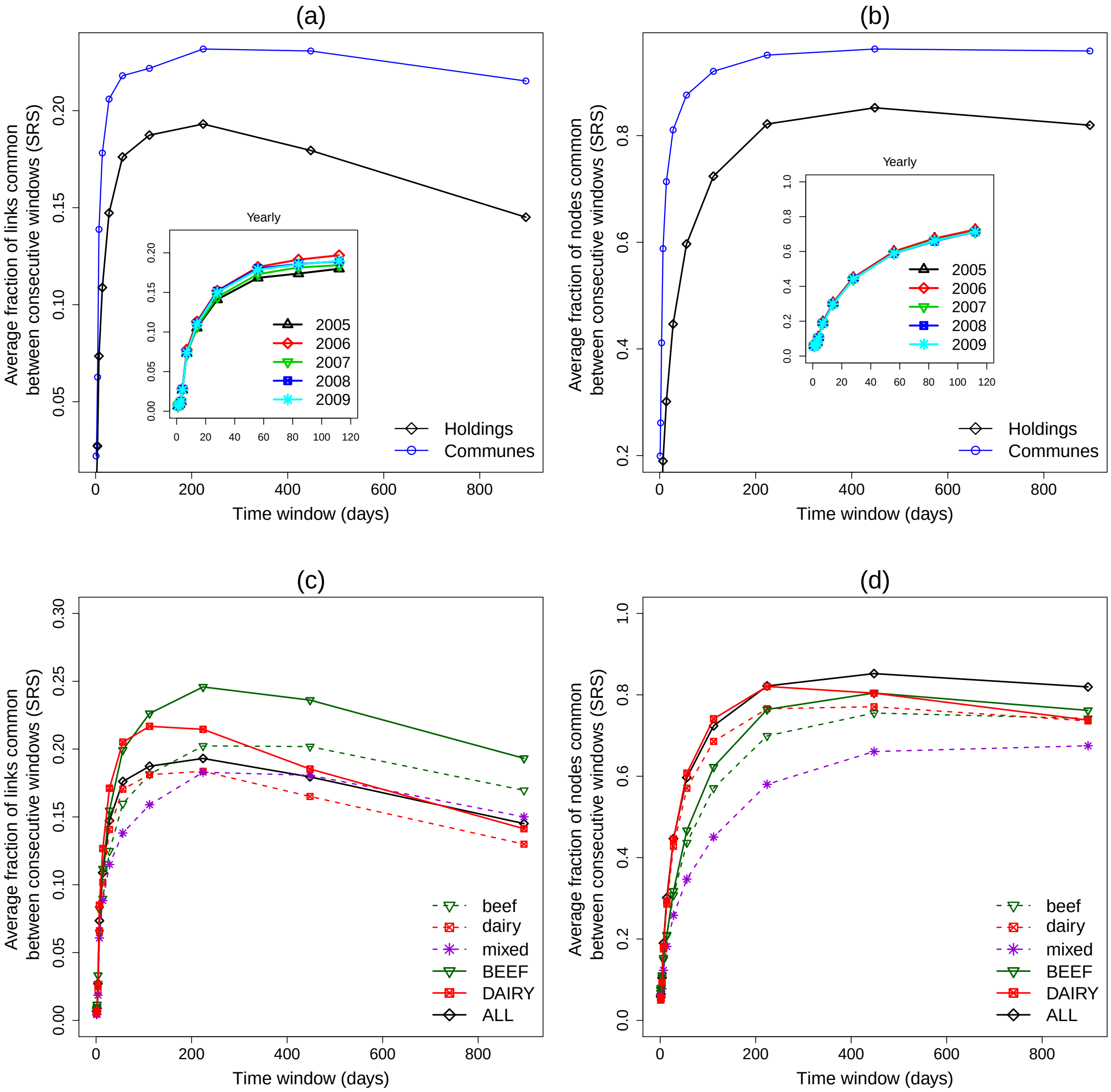


Figure 7: Short range similarity (SRS) between networks on consecutive time-windows where links ((a), (c)) and nodes ((b), (d)) are compared. The national network of holdings based on cattle movements in France over the whole period (2005-2009) and for yearly networks (insets) is studied ((a), (b)), based on holdings as nodes (black) and on communes as nodes (blue). Subnetworks ((c), (d)) consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). Markets and assembling centres are part of the subnetworks.

number of active nodes, moderate negative degree correlation). The stability of the backbone (in terms of links) over time, expressed through the SRS, reaches its maximum for a 120-day window: two consecutive networks aggregated over this window have 20% of common links in Brittany and almost 27% in Limousin. Similarly to the national networks, percolation analysis reveals that the most efficient strategy in terms of the reduction in the GSCC size

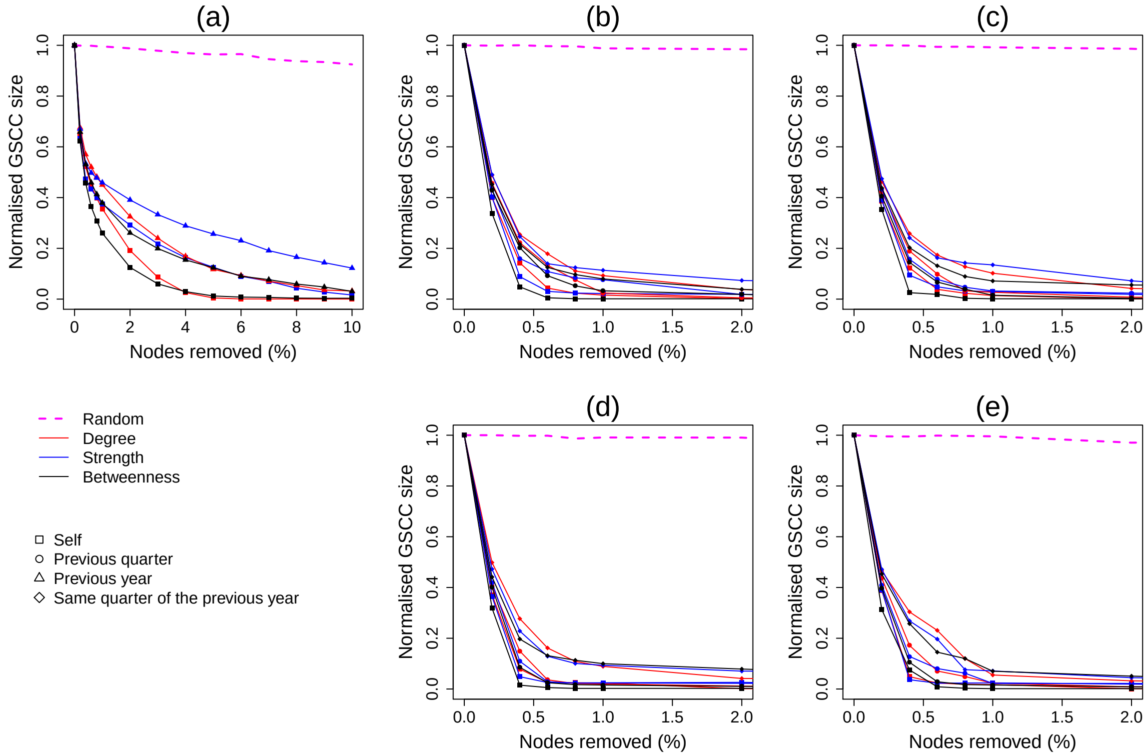


Figure 8: Effect on the GSCC of different strategies of node removal for national networks of holdings, based on cattle movements in France during 2009: for the whole year (a) and for each quarter (b)-(e) . Scenarios tested consist in targeting nodes at random (dashed pink), and based on degree (red), strength (blue), and betweenness (black) distributions, from the network in the current year (real time data) and from the network in the previous year, in the previous quarter and in the similar quarter of the previous year (delayed access to data).

consists in targeting nodes in descending order of their betweenness. The reachability ratio (average and maximum) was calculated for increasing time windows for each of the five years (Fig. 10) for both regions. Whereas the maximum RR are comparable between the two regions for one-year windows for all years, this is less true for average RR. In Brittany, the average fraction of nodes to be reached by time respecting paths is between 0.35 and 0.4 for an one-year time window for all years (Fig. 10a), whereas it is lower (0.18-0.28) in Limousin (Fig. 10b). The between year variability is more pronounced in the beef region.

4. Discussion

In this work, we have studied the French cattle movement network between 2005 and 2009 at different spatial and temporal granularities. The full network is comprised yearly of around 240 thousands of holdings (farms, assembling centres, and markets), involving 8.6 millions of movements and 5.5 millions of animals (mainly young, i.e. less than 20 months of age). Our study provides a deeper outlook of this cattle exchange network and completes previous studies based on a one-year dataset (Rautureau et al., 2011). Subnetworks based on race and herd type (beef, dairy and mixed) were thoroughly explored together with the global network. In addition to calculating indicators for static networks, we followed their temporal evolution and investigated the impact of the width of increasing time aggregation windows. Proxies for the outcome of disease spread on this network were analysed with the objective of guiding possible intervention measures to control epidemic outbreaks. Specificities of two regional networks corresponding to dominant populations of beef and dairy herds were separately assessed.

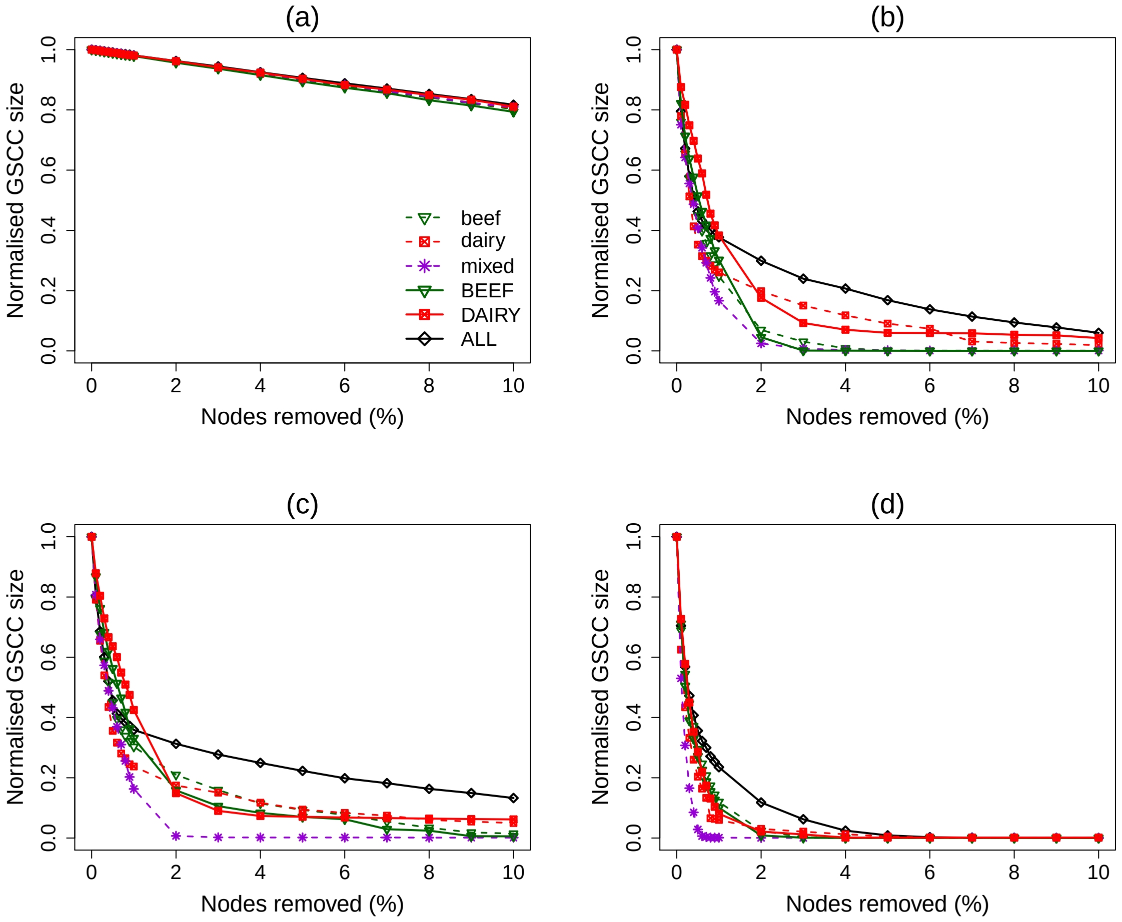


Figure 9: Effect on the GSCC of different strategies of node removal. Subnetworks, based on cattle movements in France during 2009, consider either all movements of a particular breed, irrespective of source and destination herds (beef - dashed green, dairy - dashed red, mixed - dashed violet) or all movements of a particular breed involving only a particular herd type (BEEF - plain green, DAIRY - plain red). Full network is represented in black. Different strategy of node removal are tested: random (a), degree (b), strength (c), and betweenness (d) for cattle movements data in France during 2009.

The study highlights a decreasing trend in the network size (number of active nodes) over time, regardless of the aggregation unit and subnetwork type, which may be a sign of merging of smaller individual farms into bigger corporate farms via acquisitions and mergers. In addition to this trend, a yearly pattern in the number of active nodes is visible, with an upward peak in March and a downward peak in August. A more noticeable downward peak for the global network is observed in 2007, along with a more severe increase in the average shortest path length. This latter aspect could imply that the connectivity of the network was somehow affected in 2007. The GSCC also exhibits annual trends, with seasonal minimum and maximum values. This may suggest that the spread of an infection on this network of cattle farms would have different behaviours depending on the onset time. The other indicators do not vary among years for the temporal and spatial granularities studied. Most of the network measures exhibit similar qualitative trends between holdings and communes as nodes, either on a monthly or an yearly basis. The proportion of bi-directional links (reciprocity) is quite low, suggesting that holdings (communes) to which a given holding (commune) sells animals are more probably different from those from each it buys animals. This behaviour

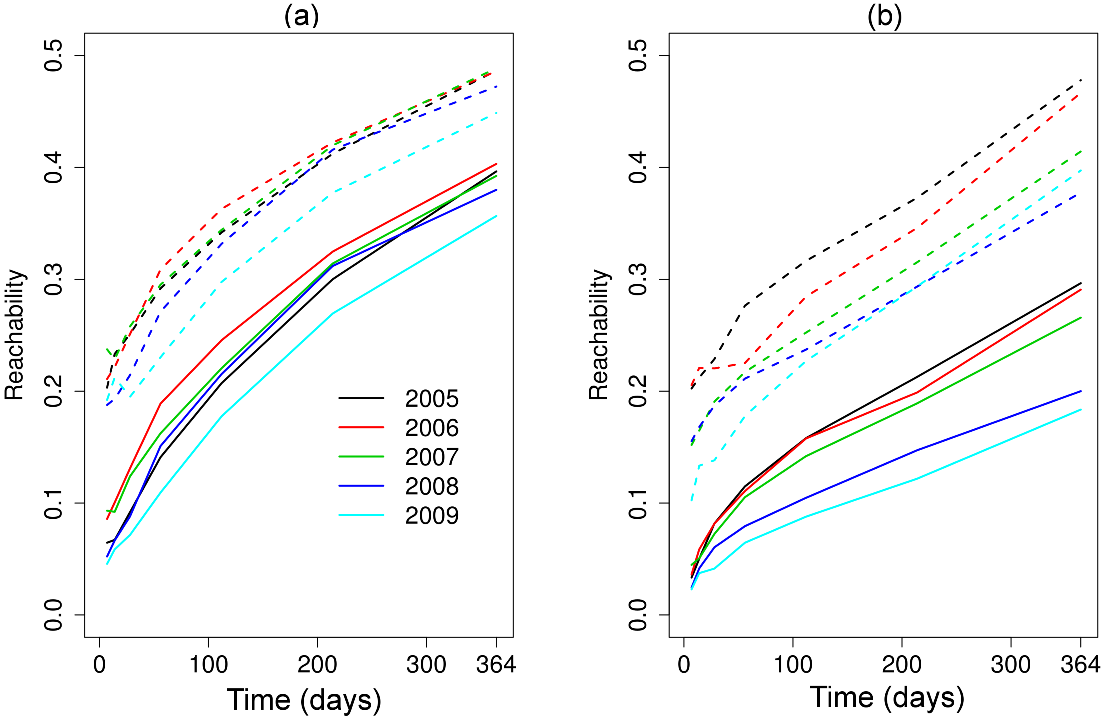


Figure 10: Average (plain line) and maximum (dashed lines) reachability ratio calculated for increasing time windows for a mainly dairy region (Brittany (a)) and a mainly beef region (Limousin (b)) in France, for annual networks of cattle movements, from 2005 to 2009, with holdings as nodes. The denominator is the total number of nodes with at least one in-connection in the time window considered.

could contribute to the spread of a potential infectious agent beyond the local neighbourhood (on the network) of initially infected holdings (communes). However, this effect could be counterbalanced by the fact that direct contacts between holdings mostly occur in an area with a radius smaller than 200 km, with half of movements at less than 50 km and preferentially even within the same commune.

The analysis of monthly herd-wise and breed-wise subnetworks shows variations that resemble to annual cycles, with characteristics similar to those of the global network. Race-wise subnetworks, built with no condition on the type of source and destination herds, include almost twice more nodes than herd-wise subnetworks. This suggests that herds sell and buy any race of animals, regardless of their type, although they may prefer to trade with herds of similar type. Moreover, if we consider the 5-year period, beef, dairy and mixed herds send animals to or receive animals mostly from markets and assembly centres (rather than other herds). Hence, it seems that in France, beef and diary cattle sectors are interrelated by trade, which points out towards a global management of the cattle system in the case of a crisis. Beef and dairy subnetworks differ in the number of movements per node, uniformly larger for the former over the whole period. The reciprocity and the cohesion expressed through the GSCC are two to three times more important for beef than for diary subnetworks. These three elements could induce a potential increased vulnerability of beef system to infection.

The spatial granularity (holdings versus communes as nodes) barely impacts the qualitative trends in indicators variations, while a commune-based network is much simpler to analyse, mainly because of smaller size. However, two main differences hold. First, the network of communes is largely more connected than the one of holdings: around 40% of the holdings and 80% of the communes are interconnected (and belong to the yearly GSCC), whereas for monthly time windows the proportion of nodes belonging to the GSCC is, on average, 6 to 7 times greater for the commune based network. Hence, the network of communes is much more vulnerable to pathogen spread. However, from the point of view of epidemic dynamics, the commune based network is particularly suitable for pathogens that both spread on long distances through animal movements and locally on short distances through neighbouring relationships (such as for porcine reproductive and respiratory syndrome virus (Dee et al., 2009)). In such a case, local transmission due to within commune interactions could be considered as a homogeneous mixing process and the commune level network becomes valuable. For diseases mainly spreading through animal movements (such as paratuberculosis (Marcé et al., 2011), and bovine viral diarrhoea (Courcoul and Ezanno, 2010)), holding based networks should be preferred. Second, the degree correlation is negative at the holding level, whereas it is positive at the commune level. This means that holdings have more often a (moderately) dominant activity (e.g. buyers are farms with a fattening activity and a large in-degree and sellers are breeding farms with a large out-degree and a low in-degree). In contrast, communes mostly differ by their volume of trade (e.g. large buyers are also large sellers and small buyers have also a reduced activity of selling).

The assessment of distributions of key characteristics (degree, strength) of the national network for increasing time windows shows that, when normalised (by the mean), they do not exhibit significant differences between two consecutive windows, which is in agreement with the literature (Krings et al., 2012). This helps modelling the empirical network and sheds light on its intrinsic dynamics. The activity seems not to be influenced by time dependence (as noticed in Perra et al. (2012); Karsai et al. (2014)). This suggests that it is more appropriate to use relative contributions of holdings with respect to global interactions than absolute values, since the former are a priori less sensitive to the period considered. However, increasing the width of time window provides more details on tails distributions. Our findings suggest that a 8-16 week period would suffice for robust estimation of centrality distributions, even if complete datasets would provide more accurate ones.

The dynamic nature of the network is evidenced by the low overlapping between temporally successive networks: if 65% of the active nodes are common over 2005-2009, this proportion drastically decreases when calculated for links (3% only). Common backbone is small even between two consecutive years: 20% of common links on average, varying between 17% for mixed to 24% for beef breed-wise subnetworks. This latter seems to have an increased faithfulness in trade relationships. However, temporal stability being low overall, longer time series data should be preferred, in order to have broader information on exchange structure, when implications of the network topology on pathogen spread are studied. This is especially as the dynamic nature of such networks has a substantial impact on pathogen spread. This was revealed using time stamped chain of contacts (Nöremark et al., 2011; Büttner et al., 2013; Dorjee et al., 2013; Konschake et al., 2013; Nöremark and Widgren, 2014). In the absence of time series of animal movements long enough to allow a comprehensive representation of exchanges between herds, statistical and mechanistic modelling studies may prove to be useful. Indeed, by identifying, based on available observations, the determinants of trade between holdings or modelling the mechanisms underlying such relationships, it would be possible to simulate animal exchange networks with realistic characteristics. This is of high importance to accurately predict the possible outcome of a new infection of cattle spreading at large scale and the impact of different possible interventions.

To control pathogen spread through regulations of animal movements, issues remain because available resources are limited (material, human resources) and the information on holding characteristics is only retrospective (on past movements). Identifying which nodes should be targeted to efficiently control disease spread is valuable to focus the management effort on the most relevant nodes. According to our findings following percolation analysis, when only access to delayed data is possible, preventing the totality of animal movements concerning 1 to 5% of the holdings with the highest betweenness in the network based on the previous year reduces the size of the GSCC by 60% to 90 %. This strategy outperforms random removal (completely ineffective) and nodes targeting based on degrees and strengths, for this latter even when real-time data is available. This is directly related to the fact that networks with power law degree distributions, also called scale-free networks (Albert and Barabási, 2002) are robust against random failures, but, when the node removal is done according to nodes ranked by degree, a comparatively small number of removals can lead to break down of the network. This also corroborates up to some extent the results of previous studies (Kiss et al., 2006; Rautureau et al., 2012; Büttner et al., 2013; Iyer et al., 2013). We also investigated 3-month networks as this duration is close to the one allowing the maximum short-range similarity and because monthly networks were too sparse and hence less meaningful for the assessment of network vulnerability or, conversely, resilience. For quarterly networks, the GSCC is reduced by more than 90% for only 1% of nodes removed proportional to their betweenness in the network of previous quarter. This is the most efficient strategy for quarterly networks and suggests that, despite a certain seasonality in network indicators, the most faithful information on a given snapshot of animal movements comes from the temporally closest network. The rapid decline in GSCC size after targeted nodes removal is also related to the disassortativity, as illustrated by the average nearest-neighbour degrees (Barrat et al., 2004), weighted and unweighted. In our data, both are decreasing functions of the degree, which implies that nodes exchanging few animals, which are the most numerous, are pointing to neighbours with large interactions. So, removing even a low proportion of highly connected nodes leads to the removal of many links. For the five race and herd-based subnetworks, the percolation analysis revealed that, similarly to the global network, they are highly resilient to random removal of nodes compared to targeted removal. For these latter strategies, subnetworks were found to be more sensitive to node removal than the full network: e.g. all subnetworks almost break-down with top 2% of the nodes removed based on the betweenness, whereas for the full network it takes about 5% of the nodes. However, since each of the subnetworks communicates with the others through intermediate nodes like markets and assembly centres, breaking down its structure does not necessarily mean it is also true for the global network.

On the field, interventions are generally implemented at a regional scale, in relation with a local organization of animal health management for many animal diseases, but also because it is more convenient to geographically group interventions. The fact that cattle movements largely occur in the neighbourhood of the source holding also points out towards a regional organization. Moreover, regional networks could have specificities which should be taken into account when implementing interventions, to enhance control effectiveness. Indeed, when comparing a mainly dairy region, as Brittany, with a mainly beef region, as Limousin, we found that not only the number of holdings and the number of movements differ, but also the reciprocity, the backbone stability, and the reachability of nodes do, all the three being higher in the latter. This corroborates the analysis of monthly variations of network indicators for dairy and beef subnetworks. Besides, the intrinsic temporal variation of animal trade network is an important element to be considered when evaluating outbreak burden and interventions. Namely, the RR, calculated as the average fraction of nodes reachable through time-respecting paths starting from any node, in a given time window, allows a refined assessment of epidemic load, compared to the GSCC (Holme and Saramäki, 2012). Indeed, since the GSCC does not make the distinction between temporally compatible and inconsistent paths, it overestimates the epidemic size. For the two regions studied, the annual GSCC includes about 26% of nodes in Brittany and 34% in Limousin, whereas the RR, when calculated over all the active nodes, is about 19% and 14%, respectively. Similar considerations, related to ingoing and outgoing infection chains, can be found in (Nöremark et al., 2011).

Even if links are not stable over the years, the network topology does not significantly change from 2005 to 2009, at both yearly and monthly scales, indicating that data on a single year is sufficient to describe it (but not the setting up of links between nodes). The period studied covers an animal health crisis related to the spread of bluetongue among cattle herds, started in 2006 and experiencing its geographical peak in terms of geographical spread and incidence in 2007, which induced regulations of animal movements for specific periods (Mintiens et al., 2008; Carpenter et al., 2009; Pioz et al., 2011; Vernon and Keeling, 2012). However, at the national level, the only effects on the network topology that could have been highlighted in our study, possibly in relation with bluetongue spread, concern the more severe decrease in 2007 in the number of active nodes and a noticeable increase during the same year in the shortest path length. The other network indicators are not noticeably impacted, whatever the spatial and temporal granularities considered, at least when focusing on within-country movements, excluding imports and exports. Therefore, the topology of the full network can be seen as quite robust to perturbations, due, for instance, to a spontaneous reorganization of the network. But this is also possibly due to the rapid progression of the partition into zones of trade related to the BTV status, towards very large zones, corresponding to the whole country. When considering regional specificities, the reachability ratio, accounting for the network ability to propagate pathogens through temporally compatible paths, was more variable among years and lower in Limousin, a beef region, than in Brittany, a diary region, especially in years 2008 and 2009. This could represent a signature of the bluetongue spread on the realised between-farm cattle movements, expected to be higher in Limousin which was affected more and earlier (www.fcoinfo.fr, 2011). Although further analyses are needed to investigate this hypothesis, this already points out the importance of the spatial scale in identifying consequences of external perturbations of such networks.

As the precise information on movements of cattle and pigs ensuring their traceability is available in most of the European countries, it is important to extend the analysis of these databases in both space and time. Indeed, this would allow assessing the spread of diseases at the level of more global markets, beyond the national level. The existence of common structural patterns between trade networks for different countries could contribute to a better control based on synergistic interventions. In parallel, a deeper understanding of livestock movements, accounting for economic mechanisms, can also contribute to improve prevention and control of pathogen spread.

Acknowledgements

We acknowledge financial support provided by the French Research Agency, program Investments for the future, project ANR-10-BINF-07 (MIHMES), by the European fund for the regional development FEDER, and by INRA, Cemagref and Basse-Normandie, Bretagne, Pays-de-la-Loire and Poitou-Charentes Regional Councils under SANCRE project, in the framework of ”For and About Regional Development” programs (France). We thank the French Ministry of Agriculture (DGAl) for providing the data used in this work.

References

Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97.

Bajardi, P., Barrat, A., Natale, F., Savini, L., Colizza, V., 2011. Dynamical patterns of cattle trade movements. PLoS One 6, e19869.

Bajardi, P., Barrat, A., Savini, L., Colizza, V., 2012. Optimizing surveillance for livestock disease spreading through animal movements. J R Soc Interface 9, 2814–2825.

Barrat, A., Barthlemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc Natl Acad Sci U S A 101, 3747–3752.

Barthélemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A., 2005. Characterization and modeling of weighted networks. Physica A: Statistical Mechanics and its Applications 346, 34–43.

Bigras-Poulin, M., Thompson, R.A., Chriel, M., Mortensen, S., Greiner, M., 2006. Network analysis of danish cattle industry trade patterns as an evaluation of risk potential for disease spread. Prev Vet Med 76, 11–39.

Buhnerkempe, M.G., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Webb, C.T., 2013. A national-scale picture of u.s. cattle movements obtained from interstate certificate of veterinary inspection data. Prev Vet Med 112, 318–329.

Büttner, K., Krieter, J., Traulsen, A., Traulsen, I., 2013. Efficient interruption of infection chains by targeted removal of central holdings in an animal trade network. PLoS ONE 8, e74292.

Carpenter, S., Wilson, A., Mellor, P.S., 2009. Culicoides and the emergence of bluetongue virus in northern europe. Trends Microbiol 17, 172–178.

Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM review 51, 661–703.

Courcoul, A., Ezanno, P., 2010. Modelling the spread of bovine viral diarrhoea virus (bvdv) in a managed metapopulation of cattle herds. Veterinary microbiology 142, 119–128.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal Complex Systems, 1695. URL: `http://igraph.org`.

Danon, L., Ford, A., House, T., Jewell, C., Keeling, M., Roberts, G., Ross, J., Vernon, M., 2011. Networks and the epidemiology of infectious disease. Interdisciplinary perspectives on infectious diseases 2011, 284909:28.

Dee, S., Otake, S., Oliveira, S., Deen, J., 2009. Evidence of long distance airborne transport of porcine reproductive and respiratory syndrome virus and mycoplasma hyopneumoniae. Veterinary Research 40, 39.

Dorjee, S., Revie, C., Poljak, Z., McNab, W., Sanchez, J., 2013. Network analysis of swine shipments in ontario, canada, to support disease spread modelling and risk-based disease management. Preventive Veterinary Medicine 112, 118 – 127.

Dubé, C., Ribble, C., Kelton, D., 2010. An analysis of the movement of dairy cattle through 2 large livestock markets in the province of ontario, canada. Can Vet J 51, 1254–1260.

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. Transbound Emerg Dis 56, 73–85.

Ezanno, P., Fourichon, C., Beaudeau, F., Seegers, H., 2006. Between-herd movements of cattle as a tool for evaluating the risk of introducing infected animals. Animal Research 55, 189–208.

Hagberg, A.A., Schult, D.A., Swart, P.J., 2008. Exploring network structure, dynamics, and function using NetworkX, in: Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA USA. pp. 11–15.

Holme, P., Saramäki, J., 2012. Temporal networks. Physics Reports 519, 97 – 125. Temporal Networks.

Iyer, S., Killingback, T., Sundaram, B., Wang, Z., 2013. Attack robustness and centrality of complex networks. PLoS One 8, e59613.

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. Proc Biol Sci 273, 1999–2007.

Karsai, M., Perra, N., Vespignani, A., 2014. Time varying networks and the weakness of strong ties. Scientific reports 4, 4001.

Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. J R Soc Interface 2, 295–307.

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. J R Soc Interface 3, 669–677.

Konschake, M., Lentz, H.H., Conraths, F.J., Hövel, P., Selhorst, T., 2013. On the robustness of in-and out-components in a temporal network. PloS one 8, e55223.

Krings, G., Karsai, M., Bernhardsson, S., Blondel, V.D., Saramäki, J., 2012. Effects of time window size and placement on the structure of an aggregated communication network. EPJ Data Science 1, 1–16.

Marcé, C., Ezanno, P., Seegers, H., Pfeiffer, D., Fourichon, C., 2011. Predicting fadeout versus persistence of paratuberculosis in a dairy cattle herd for management and control purposes: a modelling study. Veterinary Research 42, 36.

Martínez-López, B., Perez, A., Sánchez-Vizcaíno, J., 2009. Social network analysis. review of general concepts and use in preventive veterinary medicine. Transbound Emerg Dis 56, 109–120.

Mintiens, K., Mroc, E., Faes, C., Abrahantes, J.C., Hendrickx, G., Staubach, C., Gerbier, G., Elbers, A.R.W., Aerts, M., Clercq, K.D., 2008. Impact of human interventions on the spread of bluetongue virus serotype 8 during the 2006 epidemic in north-western europe. Prev Vet Med 87, 145–161.

Mweu, M.M., Fourni, G., Halasa, T., Toft, N., Nielsen, S.S., 2013. Temporal characterisation of the network of danish cattle movements and its implication for disease control: 2000-2009. Prev Vet Med 110, 379–87.

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, 341–350.

Newman, M., 2010. Networks: An Introduction. Oxford University Press, Inc., New York, NY, USA.

Newman, M.E., 2002. Assortative mixing in networks. Physical review letters 89, 208701.

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, 78–90.

Nöremark, M., Widgren, S., 2014. Epicontacttrace: an r-package for contact tracing during livestock disease outbreaks and for risk-based surveillance. BMC Vet Res 10, 71.

Perra, N., Gonalves, B., Pastor-Satorras, R., Vespignani, A., 2012. Activity driven modeling of time varying networks. Sci Rep 2, 469.

Pioz, M., Guis, H., Calavas, D., Durand, B., Abrial, D., Ducrot, C., 2011. Estimating front-wave velocity of infectious diseases: a simple, efficient method applied to bluetongue. Vet Res 42, 60.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://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, 110–120.

Rautureau, S., Dufour, B., Durand, B., 2012. Structural vulnerability of the french swine industry trade network to the spread of infectious diseases. Animal 6, 1152–1162.

Squartini, T., Picciolo, F., Ruzzenenti, F., Garlaschelli, D., 2013. Reciprocity of weighted networks. Sci Rep 3, 2729.

Vernon, M.C., 2011. Demographics of cattle movements in the united kingdom. BMC Vet Res 7, 31.

Vernon, M.C., Keeling, M.J., 2009. Representing the uk's cattle herd as static and dynamic networks. Proc Biol Sci 276, 469–476.

Vernon, M.C., Keeling, M.J., 2012. Impact of regulatory perturbations to disease spread through cattle movements in great britain. Prev Vet Med 105, 110–117.

www.fcoinfo.fr, 2011. Fcoinfo.fr. URL: www.fcoinfo.fr.

Appendix II: Summary of the thesis in French

Résumé long de la thèse en français

Propagation spatio-temporelle d’un agent pathogène dans une métapopulation bovine: application au virus de la diarrhée virale bovine (BVDV)

Introduction : contexte et objectifs

Les baisses de productivité et l’augmentation des pertes en élevage ont un fort impact économiques tant sur les filières qu’au niveau individuel des éleveurs. Une des causes majeures de pertes en élevage est la circulation d’agents pathogènes transmissibles. Les maladies animales, et notamment celles sans impact pour la santé de l’homme, sont généralement rarement considérées, hormis lorsqu’elles engendrent de très larges épizooties, aux conséquences immédiatement visibles, comme ça a été le cas pour la fièvre aphteuse par exemple. La gestion des maladies enzootiques est quant à elle laissée à la libre décision de l’éleveur ou des organisations d’éleveurs. Cependant, ces maladies peuvent persister à de faibles à moyennes prévalences pendant de très longues périodes, impactant finalement de manière notable la productivité des élevages [Bennett, 2003, Nieuwhof and Bishop, 2005]. De plus, leur maîtrise à l’échelle d’un élevage est généralement contrainte par la situation dans les élevages en contact (source d’animaux ou au voisinage) [Carslake et al., 2011].

Le complexe maladie des muqueuses / diarrhée virale bovine (BVD) est l’une de ces maladies enzootiques, touchant principalement les bovins, et actuellement présentes dans une large proportion des élevages de par le monde. Cette maladie virale engendre des baisses de performance reproductive et de production laitière, des retards de croissance et des mortalités des veaux, et favorise l’occurrence d’autres maladies en élevage [Lindberg, 2003]. L’impact économique de cette maladie sur l’industrie laitière a été quantifié à $10-40 millions par million de vêlages, avec une incidence annuelle de 20-40%, les variations étant attribuées à l’hétérogénéité des structures d’élevages et des modalités de conduite des troupeaux [Houe, 2003]. La propagation et la persistance de ce virus impliquent plusieurs échelles : une transmission entre individus dans un groupe d’animaux, entre groupes dans un troupeau structuré, entre troupeaux localisés dans un même territoire, et entre troupeaux géographiquement distants (métapopulation) [Ezanno et al., 2012]. Les contacts entre populations d’une métapopulation sont de plusieurs types : migration ou dispersion (les individus quittent définitivement leur population d’origine), visites (reposant sur des allers-retours entre populations), voisinage et contacts indirects (sans mouvement explicite des individus des populations mais avec des contacts à la frontière ou via le mouvement de supports inertes ou de vecteurs) [Keeling and Rohani, 2008]. Dans un contexte d’élevage, certains éleveurs achètent/vendent des animaux en lien avec leur activité de production ou pour remplacer des animaux reproducteurs, de manière à maintenir la taille du troupeau et sa productivité. Ils exposent ainsi leur élevage à l’introduction d’agents pathogènes et – s’ils vendent des animaux – les élevages de destination de même [Alvarez et al., 2011, Tinsley et al., 2012, Gates and Woolhouse, 2014]. Des contacts de voisinage peuvent aussi avoir lieu lors du pâturage (contact à la barrière, pâture commune, etc.) ou de l’échange de matériel par exemple. Une meilleure compréhension de la transmission du BVDV entre troupeaux en lien avec les caractéristiques des troupeaux en contact permettrait de proposer des mesures de maîtrise de l’infection pertinentes, à l’échelle du troupeau ou de la région, et adaptées aux spécificités territoriales (type de troupeau, densité animale, mouvements).

Une approche par modélisation apparaît comme pertinente pour représenter un tel système biologique complexe reposant sur des processus démographiques, épidémiologiques et économiques en interaction dans le temps et à différentes échelles spatiales, de l'animal à la filière de production primaire ou la région. Plusieurs modèles ont été développés pour représenter la propagation du BVDV dans un troupeau bovin laitier ou allaitant [voir pour revue Viet et al., 2007 ; Ezanno et al., 2007, Damman et al., en révision]. Plus récemment, des auteurs se sont intéressés à la propagation du BVDV à une plus large échelle [Courcoul and Ezanno, 2010, Ersbøll et al., 2010,Tinsley et al., 2012]. Cependant, ces travaux n'intègrent pas simultanément l'hétérogénéité des troupeaux (taille, structure, prévalence de l'infection) dans l'espace et dans le temps et des modalités réalistes de contacts entre troupeaux (voisinage et mouvements d'animaux), qui peuvent elles-mêmes être dynamiques dans le temps et hétérogènes en espace.

Une modélisation de la dynamique de propagation spatio-temporelle du BVDV dans une métapopulation de troupeaux bovins permettrait non seulement de mieux comprendre la persistance et les facteurs de propagation de ce virus, mais aussi d'évaluer des stratégies de maîtrise ciblées dans le temps et dans l'espace, voire sur une catégorie donnée de troupeaux ou d'animaux. L'objectif de ma thèse est de proposer un tel modèle, reposant sur la prise en compte explicite des mouvements de bovins entre troupeaux tels qu'enregistrés dans la base de données nationale d'identification des bovins (BDNI), ainsi que sur une représentation la plus réaliste possible des relations de voisinage entre troupeaux selon leur localisation géographique. Le statut épidémiologique des troupeaux sera représenté de manière dynamique par un modèle intra-troupeau adapté pour prendre en compte mouvements et voisinage, ainsi que l'hétérogénéité de taille des troupeaux.

Une première étape a concerné l'étude exhaustive du réseau de contact entre troupeaux basé sur les mouvements de bovins en France de 2005 à 2009. Ce réseau orienté et pondéré a été analysé non seulement d'un point de vue statique en agrégeant les informations sur des fenêtres de temps et/ou d'espace plus ou moins grandes, mais également d'un point de vue dynamique. Une deuxième étape a concerné la traduction d'un modèle existant de la propagation du BVDV dans un troupeau bovin laitier dans un langage de programmation plus performant (C++) de manière à le coupler au réseau des mouvements. Avant de passer à une échelle régionale, ce modèle a été étendu pour pouvoir tenir compte d'une hétérogénéité de taille des troupeaux et l'impact de cette taille sur la propagation locale du BVDV a été analysé. A l'échelle régionale, j'ai non seulement considéré les mouvements d'animaux entre troupeaux en me basant sur les mouvements observés, mais j'ai également tenu compte des relations de voisinage en considérant qu'un troupeau était au voisinage de tous les troupeaux localisés à une distance comprise dans un certain rayon, de manière à évaluer l'impact de l'intensité des relation de voisinage sur la propagation semi-locale du BVDV. Le modèle complet a été analysé pour mieux comprendre les déterminants de la propagation du BVDV à une échelle régionale et identifier des leviers d'actions possibles. La structure générale de la thèse est illustrée dans la figure 1. L'analyse du réseau des mouvements a été valorisée par une publication acceptée dans *Preventive Veterinary Medicine* (Annexe 1). L'élaboration et l'analyse du modèle épidémiologique régional de propagation du BVDV est en cours de valorisation pour *Veterinary Research.*

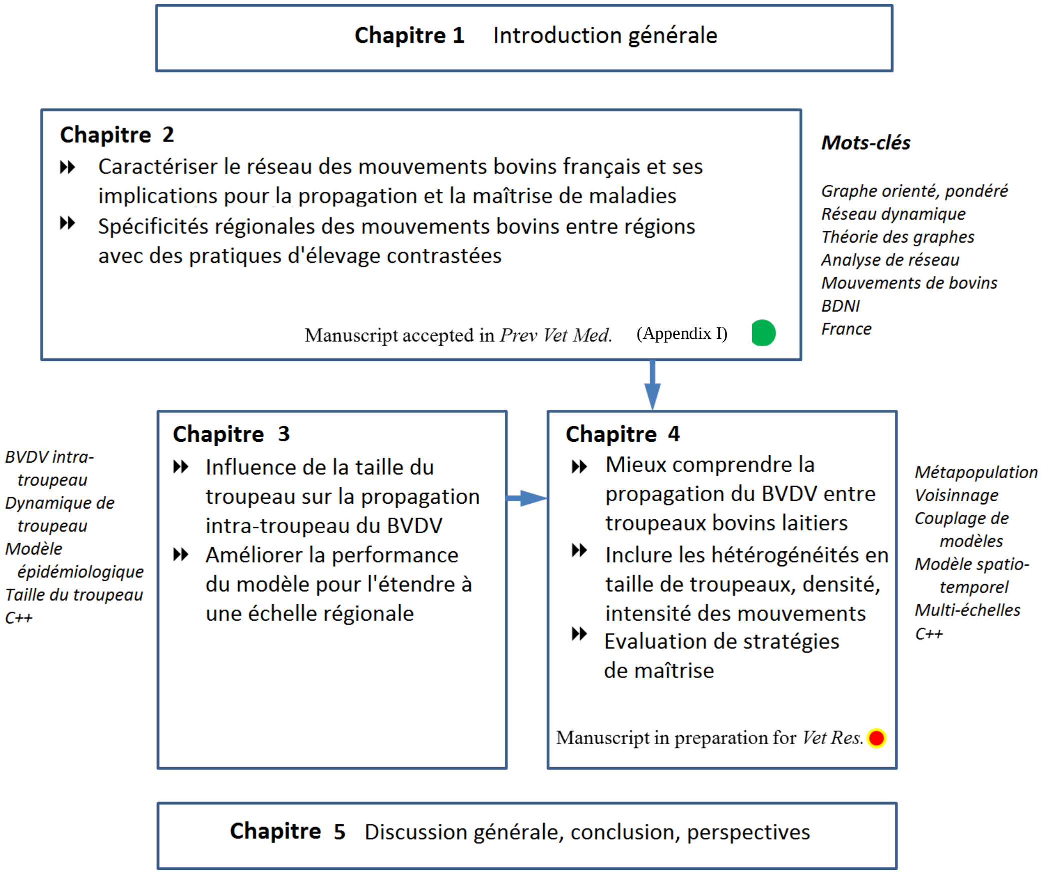


Fig. 1 : Structure générale de la thèse.

Mouvements de bovins entre troupeaux et risque de propagation d'agents pathogènes

Dans le Chapitre 2, j'ai étudié le réseau des mouvements de bovins en France (BDNI) de 2005 à 2009 en utilisant la théorie des graphes et les méthodes d'analyse de réseaux sociaux [Dubé et al., 2009]. J'ai réalisé un bref état de l'art tant sur les méthodes que sur les analyses parues dans la littérature portant sur des mouvements d'animaux d'élevage [parmi lesquels Kao et al., 2006, Kiss et al., 2006, Vernon and Keeling, 2009, Rautureau et al., 2011, Nöremark et al., 2011, Bajardi et al., 2011, Mweu et al., 2013]. J'ai décrit les données disponibles (BDNI), que j'ai ensuite analysées de deux manières : tout d'abord, les mouvements ont été agrégés dans le temps pour considérer un réseau statique dont les liens sont supposés durer l'intégralité de l'intervalle de temps considéré ; ensuite, la séquence temporelle des mouvements a été prise en compte pour considérer un réseau dynamique. L'ensemble des descripteurs de réseaux (tels que le degré de connectivité des noeuds – valeur moyenne et distribution, longueur moyenne des chemins les plus courts entre paires de noeuds, composante connexe maximale – GSCC, etc) pertinents pour les systèmes d'élevage bovins et l'épidémiologie animale ont été évalués, et ce à différentes échelles de temps (de la semaine à l'année) et de résolution spatiale (en considérant comme unité l'exploitation d'élevage ou la commune, d'une échelle régionale à nationale). Cette analyse poursuivait trois objectifs : (i) explorer les variations temporelles des principaux indicateurs topologiques du réseau, (ii) évaluer, à travers des mesures approchant la taille d'une épidémie, la vulnérabilité des réseaux à la propagation d'agents pathogènes en tenant compte des propriétés dynamiques du réseau, et (iii) mettre en évidence d'éventuelles spécificités dans la structure des sous-réseaux laitiers et allaitants ou correspondant à une région plutôt laitière et une région plutôt allaitante.

Les propriétés du réseau se sont avérées qualitativement similaires entre les différentes échelles de temps et d'espace considérées dans notre étude. Environ 40% des troupeaux et 80% des communes étaient connectés indirectement (i.e. appartenaient à la GSCC). L'intervalle de temps considéré ne semble pas avoir d'impact sur la distribution normalisée des descripteurs du réseau. Une fenêtre de 8 à 16 semaines devrait suffire pour une estimation robuste des principales caractéristiques du réseau considéré d'un point de vue statique. Cependant, une fenêtre d'observation plus longue permet de mieux considérer les queues de distribution. La faible superposition entre réseaux sur des intervalles de temps consécutifs mais de même taille révèle la nature profondément dynamique de ce réseau. Ainsi, si 65% des troupeaux impliqués dans des mouvements sont communs entre les cinq années observées, seulement 3% des liens sont conservés sur la période étudiée. Du point de vue des spécificités liées aux types d'élevages et à la race des animaux (laitière versus allaitante), nous avons remarqué que le nombre de bêtes échangées par noeud du réseau, la réciprocité et la cohésion (exprimée à travers la GSCC) étaient uniformément supérieurs pour les sous-réseaux laitiers (Fig. 2).

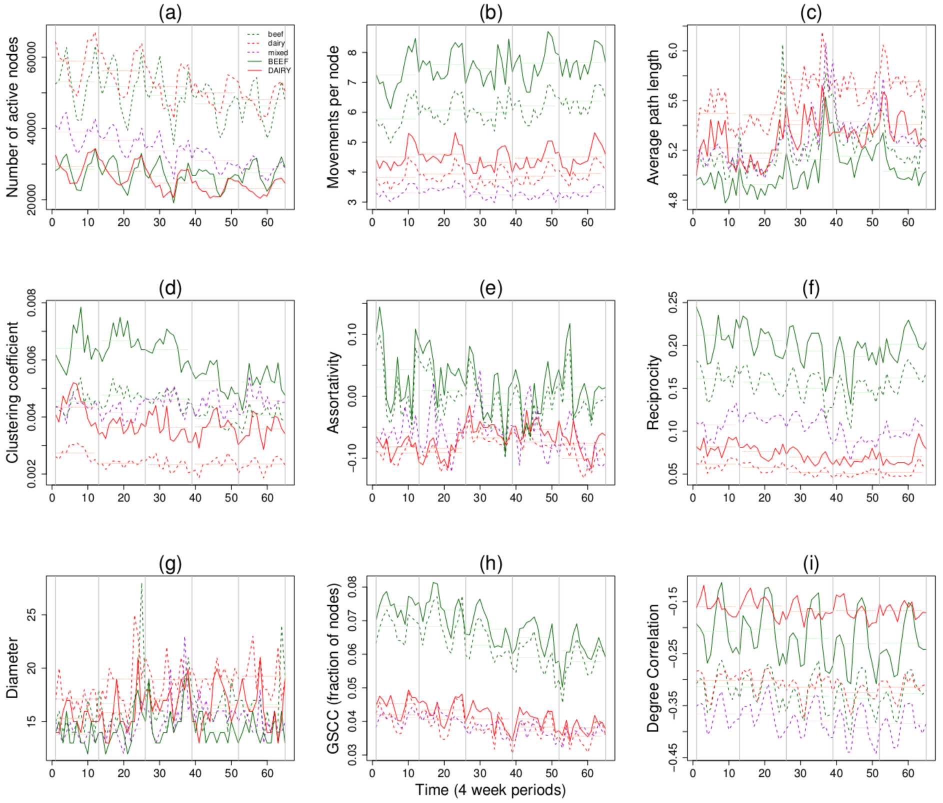


Fig. 2 : Variabilité temporelle des descripteurs des sous-réseaux (labélisés sur l'axe y de chaque sous-graphe) pour les sous-réseaux laitiers (en rouge), allaitants (en vert) et mixtes (en violet) (définis selon le type d'élevage ou la race des animaux échangés). Les lignes pointillées horizontales représentent les moyennes annuelles (même code couleur que ci-dessus) et les barres verticales indiquent des intervalles de 364 jours. Les données sont issues de la BDNI, à l'échelle de la France, pour la période 2005-2009 et chaque point de chaque courbe correspond à la valeur d'un indicateur donné pour un sous-réseau agrégé temporellement sur une fenêtre de 4 semaines.

Inversement, nous avons montré que pour une région à dominance laitière, l'estimateur d'une épidémie basé sur un réseau dynamique est plus important que pour une région à dominance allaitante. Une analyse de percolation a été réalisée pour différents scénarios d'action ciblée sur des nœuds spécifiques du réseau, permettant d'estimer la réponse du réseau global à des potentielles stratégies de maîtrise reposant sur le contrôle des mouvements d'animaux. Cette analyse montre que limiter les mouvements provenant de 1 à 5% des élevages ayant la plus forte centralité dans le réseau permettrait de réduire la taille maximale de l'épidémie de 80%. En revanche, une action non ciblée, contrôlant les nœuds de manière aléatoire n'empêche aucunement la propagation d'agents pathogènes (Fig. 3). L'implication de ces résultats pour la propagation de maladies infectieuses animales (et notamment la BVD) sur le réseau a été discutée. En particulier, nous avons noté que les spécificités locales et la dynamique temporelle des échanges commerciaux d'animaux doivent être prises en compte pour évaluer des stratégies de maîtrise de la propagation d'agents pathogènes sur le réseau.

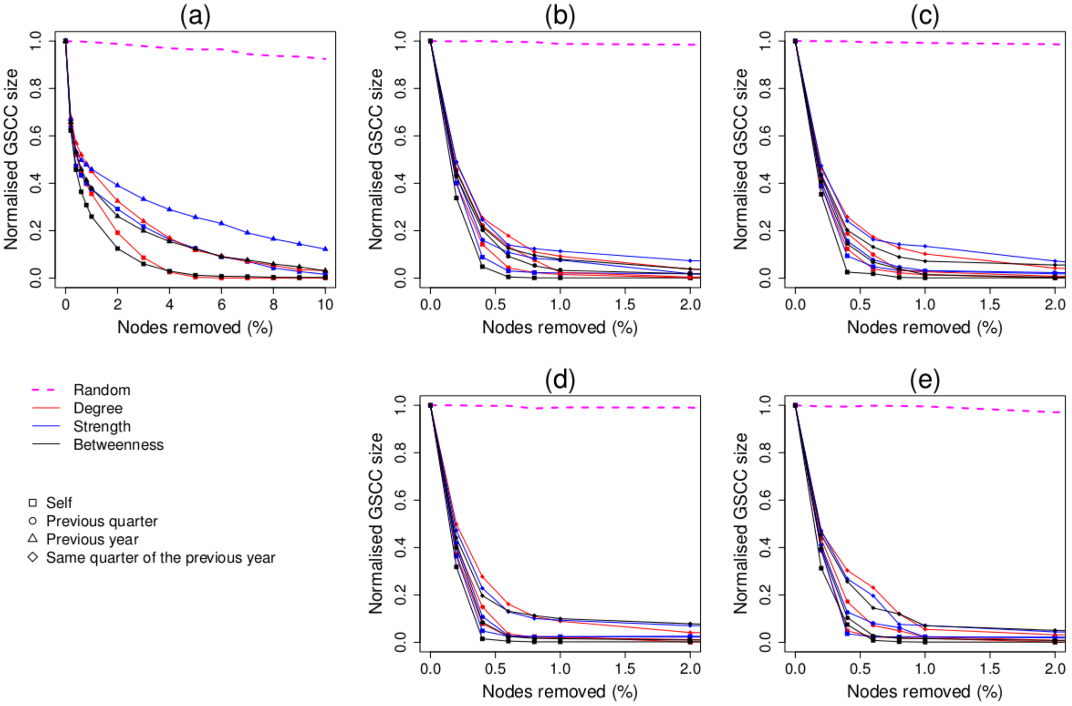


Fig. 3 : Effet des différentes stratégies d'isolement des nœuds sur la GSCC : ciblage des noeuds aléatoirement (rose ligne interrompue) ou par ordre décroissant des degrés (rouge), de l'intensité des échanges (bleu) et de la centralité betweeness (noir). L'information choisie pour classer les noeuds est soit celle du même réseau (signe carré), du réseau de l'année précédente (triangle), du trimestre précédent (cercle), ou du trimestre analogue de l'année précédente (losange). Données issues de la BDNI, à l'échelle de la France, 2009. Réseau annuel (a), réseaux trimestriels (b)-(e).

La propagation du BVDV dans une région d'élevage bovin laitier

La prévalence de l'infection par le BVDV dans un troupeau infecté étant très variable au cours du temps et entre troupeaux, j'ai choisi de modéliser la dynamique régionale de l'infection en considérant le couplage explicite de plusieurs dynamiques à l'échelle du troupeau. Le modèle tient donc compte : (1) de l'hétérogénéité de taille des troupeaux, (2) de leur dynamique propre d'infection lorsqu'ils sont infectés en lien avec leur dynamique démographique, un modèle existant pertinent [Ezanno et al., 2007] ayant été optimisé et adapté, (3) des mouvements qui les concernent tels qu'observés dans la BDNI et dont le risque de transmettre / acquérir l'infection dépend de la prévalence au sein de la classe d'âge concernée par le mouvement dans le troupeau source, et (4) de possibles relations de voisinage en lien avec les densités locales de troupeaux au sein de chaque commune. Ce modèle a été appliqué au département du Finistère en Bretagne, qui constitue un très important bassin de production laitière dans l'Ouest de la France.

Dans le chapitre 3, je décris le modèle utilisé et les principaux ajustements réalisés, notamment pour tenir compte de la diversité de taille des troupeaux et pour améliorer la performance informatique du modèle qui a été reprogrammé en C++. Les sorties de ce modèle mis à jour sont très comparables à celle du modèle initial. Le modèle a ensuite été utilisé pour évaluer l'effet de la taille du troupeau sur la propagation locale du BVDV. Ainsi, l'introduction unique d'un animal infecté dans un troupeau naïf donne lieu comme attendu à de plus importantes tailles d'épidémies dans des grands troupeaux en valeur absolue. Cependant, lorsque l'on normalise par la taille, les petits troupeaux sont plus impactés.

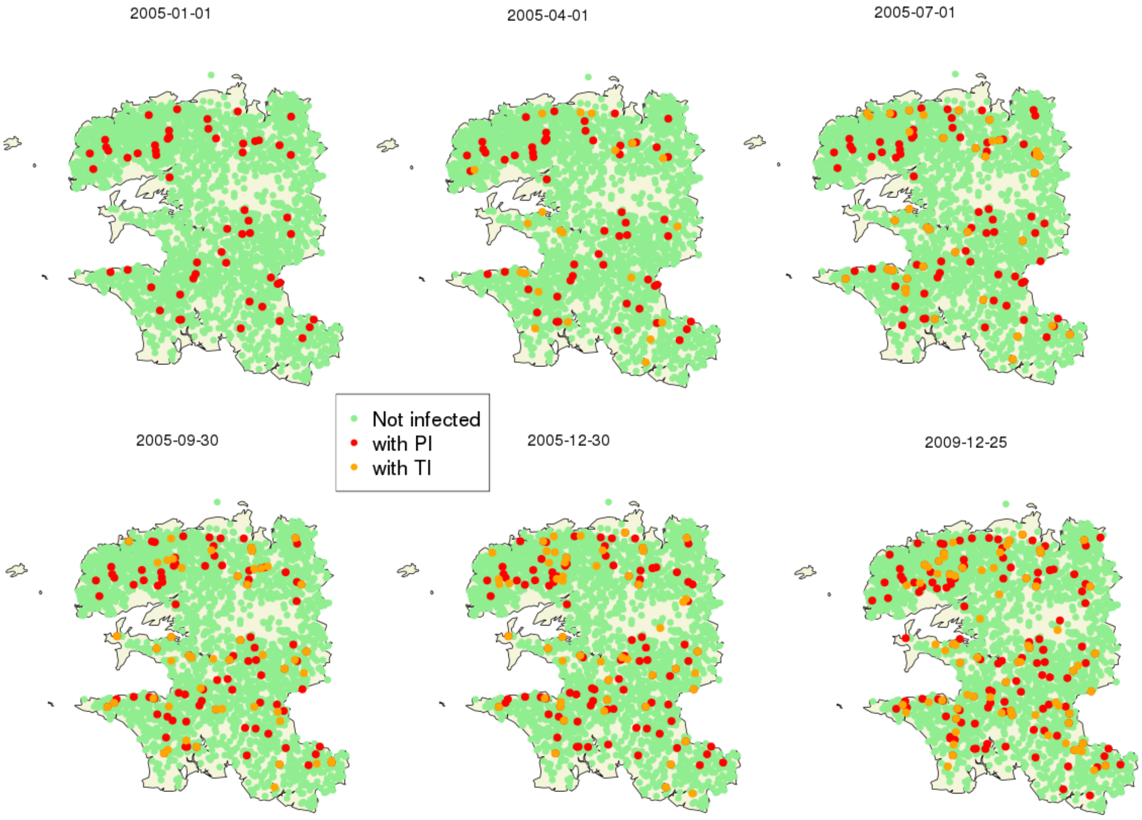


Fig. 4 : Simulation (une trajectoire stochastique) de la propagation spatiale de l'infection par le BVDV dans le département Finistère (en rouge présence d'IPI – animaux infectés à vie, en orange présence de TI – animaux infectés transitoires, en vert exploitations non atteintes par le BVDV). Les différentes cartes correspondent (de gauche à droite) au temps initial d'introduction du BVDV dans la métapopulation (le 01/01/2005 dans cette simulation) et 3 mois, 6 mois, 9 mois, 1 an et 5 ans après le début de la propagation de l'infection.

Dans le chapitre 4, après une présentation des modèles existants de la propagation du BVDV à une échelle régionale, un nouveau modèle inter-troupeau est justifié, présenté, analysé et discuté. Ce modèle stochastique à temps discret représente simultanément la dynamique d'infection locale des troupeaux infectés, les mouvements des animaux et les relations de voisinage. Les hétérogénéités de taille et de localisation des troupeaux sont prises en compte. Nos résultats préliminaires montrent que quel que soit le rayon de voisinage représenté, le virus persiste beaucoup plus longtemps dans la métapopulation que dans un troupeau isolé. La proportion moyenne de troupeaux infectés dans une région où le BVDV est présent est de 6 à 11%, en cohérence avec les observations de terrain qui indiquent une prévalence autour de 10% en région laitière avant mise en œuvre des stratégies de maîtrise. Nous avons confirmé que les mouvements jouaient un rôle prépondérant dans la propagation régionale du BVDV comparativement aux relations de voisinage, pour une intensité du voisinage telle que celle observée en Bretagne. Ainsi, pour un rayon de voisinage de 5 km (en moyenne, 41 voisins), le voisinage contribue à 1 à 3% des troupeaux nouvellement infectés. La robustesse de ce résultat a été explorée en considérant différents types d'animaux initialement infectés et différentes valeurs pour l'intensité de transmission et le rayon de voisinage. La propagation spatiale a été également étudiée et nous avons remarqué que, qualitativement, il n'y avait pas d'effet de regroupement des fermes infectées (Fig. 4). En effet, cela serait corroboré par le fait (observé dans l'analyse du réseau basé sur la BDNI) qu'à l'échelle départementale les échanges se font de manière assez homogène (i.e. ils ne sont pas restreints aux plus proches voisins).

Conclusions et perspectives

D'une part, cette thèse a fourni une analyse fine et exhaustive des échanges marchands entre exploitations de bovins en France sur plusieurs années et contribué à une meilleure compréhension de la structure de contact sous-jacente à la propagation de nombreux agents pathogènes pouvant guider le choix de stratégies de contrôle optimales. D'autre part, ce travail a fourni le premier modèle de simulation à large échelle, tout en considérant la dynamique de propagation intra-troupeau, de la propagation dans une métapopulation bovine de la BVD, maladie enzootique à large impact économique, et a permis de mieux comprendre les rôles relatifs des mouvements d'animaux et des contacts de voisinage dans la transmission de cette maladie. Ce modèle sera utilisé très prochainement pour l'évaluation de stratégies de maîtrise à l'échelle d'une région laitière (la Bretagne), en raffinant la prise en compte des relations de voisinage et en considérant aussi les mouvements d'animaux en provenance de l'extérieur de la métapopulation.

L'UNAM
l'université
nantes
angers
le mans

PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

Oniris
École Nationale Vétérinaire, Agroalimentaire et de l'Alimentation Nantes Atlantique
Santé et alimentation au cœur de la vie

Thèse de Doctorat

Bhagat Lal DUTTA

Modélisation spatio-temporelle de la propagation d'un agent pathogène dans une métapopulation bovine : application au virus de la diarrhée virale bovine (BVDV)

Spatio-temporal modelling of pathogen spread in a cattle metapopulation: application to the Bovine Viral Diarrhoea Virus (BVDV)

Résumé

L'infection au virus de la diarrhée virale bovine (BVDV) pose de sérieux problèmes socio-économiques. L'objectif de cette thèse a été de comprendre, par une approche de modélisation, la propagation du BVDV entre fermes à une échelle régionale, via des mouvements d'animaux et des relations de voisinage, ouvrant ainsi la voie à l'évaluation des stratégies de contrôle. Dans une première partie, le réseau des mouvements des bovins en France (2005-2009) a été analysé pour évaluer son évolution temporelle influençant sa capacité sous-jacente de propager des pathogènes et l'importance des spécificités régionales ou liées à la race des animaux. Les propriétés topologiques de ce réseau semblent temporellement relativement stables. La vulnérabilité relative des sous-réseaux laitiers et allaitants dépend de leur caractère agrégée ou dynamique. Dans une seconde partie, un modèle de métapopulation, stochastique et multi-échelles a été développé et efficacement implémenté pour la propagation du BVDV entre fermes laitières, en couplant des dynamiques de population et de transmission locales et une transmission du virus entre fermes par les échanges d'animaux et le voisinage. Les données des mouvements d'animaux et la géolocalisation ont été utilisées pour simuler la propagation du BVDV dans le Finistère, département à dominante laitière. Les simulations sur 10 ans ont montré que les mouvements des bovins jouaient le rôle prépondérant dans la transmission du BVDV à une échelle régionale. En accord avec les observations de terrain, il a été trouvé que 6 à 11% des fermes étaient infectées. Cette approche de modélisation sera utilisée pour tester les programmes de contrôle in silico.

Mots clés
BVDV, modèle de métapopulation, modélisation spatio-temporelle, couplage de modèles, analyse de réseaux, réseaux dynamiques, voisinage, simulations intensives

Abstract

Bovine Viral Diarrhoea Virus (BVDV) infection poses serious socio-economic concerns to cattle farms. The objective of this thesis was to understand, by a modelling approach, the propagation of BVDV between farms at a regional scale, interacting via animal movements and neighbouring relationships, thus paving the way for the evaluation of control strategies. In the first part, the network of cattle movements in France (2005-2009) was analysed to evaluate its temporal evolution influencing the underlying capacity of spreading pathogens via animal exchanges between farms and to explore the importance of its regional and breed-related specificities. Topological properties of this network were found to be quite stable over time. The relative vulnerability of beef and dairy sub-networks depended on if aggregated or dynamical views of these networks were considered. In the second part, a multi-level stochastic metapopulation model of BVDV spread in dairy herds was developed and efficiently implemented, coupling local herd population and pathogen spreading dynamics and possible inter-herd transmission of the pathogen through animal exchanges and neighbouring relationships. Animal movement data between farms and geo-location based neighbourhoods were used to simulate BVDV transmission in Finistère, a dominantly dairy department of Western France. The simulations spanning over 10 years showed that cattle movements played the main role in the transmission of BVDV at a regional scale. In agreement with field observations, it was also found that, on average, 6 to 11% of farms were infected. This modelling approach will be further used to test the control programs in-silico.

Key Words
BVDV, metapopulation model, spatio-temporal modelling, model coupling, network analysis, dynamic networks, neighbourhood, intensive simulations

L'Université Nantes Angers Le Mans