THE UNIVERSITY
of EDINBURGH

Movement, mobility and disease modelling in three epidemic contexts

Anne-Sophie Ruget

Submitted for the degree of Doctor of Philosophy

Roslin Institute
University of Edinburgh

Friday 10th June, 2022

Declaration

This thesis is submitted to the University of Edinburgh in accordance with the requirements for the degree of Doctor of Philosophy in the faculty of Medicine and Veterinary Medicine. The work presented is the work of the author except where stated otherwise by reference and/or acknowledgement. Any work presented, which has been conducted by (or in collaboration with) others is explicitly acknowledged. No part of this work has been submitted for award or degree at any other university.

Abstract

Infectious diseases have played a considerable role in shaping human history. Although their global burden has significantly decreased through the past centuries, they are still among the main causes of human death worldwide. In livestock, infectious diseases can cause substantial production losses but also have detrimental impacts upon human health, and animal health and welfare. Changes in practices and development of treatments and vaccines have helped to dramatically mitigate the impact of infectious diseases, but infectious diseases remain an ongoing challenge, either because they are difficult to control (Tuberculosis, Malaria, HIV, FMD) or because they are emerging or re-emerging pathogens. Human mobility and livestock movements play a crucial role in epidemic spread as they allow for long-range transmission and can act as bridge between otherwise disconnected populations. Repeated importations of cases in disease-free areas make the eradication or control of a disease exceedingly difficult. The patterns of potentially infectious contacts, as recorded in mobility and movement data, can be described as a network. Understanding infection transmission on networks can provide useful insights in disease risk. Mathematical models have played an increasingly important role in helping to control epidemics in animal (FMD, Avian Influenza, Swine Fever) and in human (Measle, Malaria, SARS, Ebola) populations. Modelling tools are now a central feature in the decision-making process for policy makers, as illustrated by the ongoing COVID-19 pandemic and its management. The aim of this work is to show how disease models in combination with movement or mobility data can be useful in different epidemic contexts, namely in peacetime, at the start of an outbreak and once the pathogen is circulating.

Part One investigates how livestock movement data and network analysis can be used in peacetime to improve our understanding of disease risk and to propose tools for control. In this part, I consider a fast-spreading disease affecting cattle and sheep. First, I use multi-species movement networks to understand how the combination of cattle and sheep movement affects the potential for disease spread on the combined network. I compare results of single-species vs multi-species and static vs dynamic network analyses to show the importance of interspecies links and temporal network dynamics. My results show that depending on the season, up to 70% of the premises which are likely to drive the epidemic in the multi-species network differ from the ones in both the cattle

and the sheep networks. This indicates that their risk is derived from interaction between the two farming systems. Secondly, I propose the use of a dynamic network measure based on contact chains calculated in a network weighted with transmission probabilities to assess the importance of premises in an outbreak. Comparing results with disease simulation model outputs, I demonstrate that the measure proposed allows us to identify around 30% of the key farms in a simulated epidemic, ignoring markets. Whereas static network measures identify less than 10% of these farms.

Part Two explores how mobility data within disease models can be used during an epidemic: before the pathogen is introduced (importation phase) and once the pathogen is present (circulation phase). In this part, I use the COVID-19 pandemic and its spread in the Scottish Hebrides, an archipelago off the west coast of Scotland. First, human mobility data and a metapopulation model are used to estimate the risk of introduction in each of the Islands, according to season and potential for control. I show that in some islands the introduction risk is high even in the low season, when activity and movements from the mainland are expected to be reduced. This will be of particular concern if COVID-19 becomes a seasonal respiratory infection affecting temperate areas in winter concomitantly with other seasonal infections such as flu. In the high season, although in most cases movement control will not significantly delay a potential introduction, for some islands a 70% reduction of movements in peak summer tourist season has the potential for delaying the introduction risk for over 6 weeks, i.e. beyond the high risk summer holiday period. Secondly, data from an outbreak localised in Barra Island (Western Hebrides) are used to illustrate how adjusting model parameters to disease data can provide insight in transmission dynamic and control measure efficacy. Using Approximate Bayesian Inference, I estimate the most likely date of introduction, the basic reproduction number at the start of the outbreak and I quantify the impact of voluntary vs policy-induced measures. I find that transmission started to slow down two days after the first cases were reported and a week before restrictions were imposed by the authorities. Thus my analysis is most consistent with the outbreak being mostly contained by a combination of contact tracing and self-imposed measures, whilst the lockdown, which was later imposed, had only a negligible effect on the transmission dynamic.

Lay summary

Infectious diseases have played a considerable role in shaping human history. Although their global burden has significantly decreased through the past centuries, they are still among the main causes of human death worldwide. In livestock, infectious diseases can cause substantial production losses but also have detrimental impacts upon human health, and animal health and welfare. Changes in practices and development of treatments and vaccines have helped to dramatically mitigate the impact of infectious diseases, but infectious diseases remain an ongoing challenge, either because they are difficult to control (Tuberculosis, Malaria, HIV, FMD) or because they are emerging or re-emerging pathogens. Human mobility and livestock movement play a crucial role in epidemic spread as they allow for long-range transmission and can act as bridge between otherwise disconnected populations. Repeated importations of cases in disease-free areas make the eradication of a disease exceedingly difficult. Mathematical models have had an increasingly important role in helping to control epidemics in animal (FMD, Avian Influenza, Swine Fever) and in human (Measle, Malaria, SARS, Ebola), as they help make projections by giving the ability to study the 'what if' scenarios. Modelling tools are now a central piece in the decision process for policy makers, as illustrated by the ongoing COVID-19 pandemic and its management. The aim of this work is to show how disease models in combination with movement or mobility data can be useful in different epidemic contexts, namely in peacetime, at the start of an outbreak and during an ongoing outbreak. In the first part, I use livestock movement data and network analysis to improve our understanding of disease risk at a farm level when facing the next outbreak. I highlight the importance of integrating multiple species in the study if relevant for the disease in question and propose a tool to identify farms for control in the early stage of an outbreak. In the second part, I show how human mobility data and mathematical models can be used to inform the risk of introduction and potential for control. I estimate the risk of introduction of COVID-19 to each of the Hebridean islands situated off the west coast of mainland Scotland due to individual movements, and explore control strategies to mitigate this risk. In addition, I illustrate how modelling tools can help us measure the efficacy of control measures during an outbreak. I study the epidemic on Barra Island and find that a combination of contact tracing and self-imposed measures by the pop-

ulation were as effective in controlling the epidemic as the lockdown imposed by local authorities.

For my teachers

Acknowledgements

I would like to express my sincere gratitude to everybody who has helped me through this challenging and enjoyable journey. My deep thanks to my supervisors Rowland Kao, Chris Banks and Jessica Enright for their invaluable support during the whole time of this PhD. Rowland, for your constant support, time and guidance over the years, despite my repeated change of PhD plans :). Thanks for giving me the freedom to learn and grow and for encouraging me in my career decisions. Thanks to Chris for your understanding and support. I would like to thank you as well for your (always quick) responsiveness and help with any queries but especially computer and code related questions :)! To Jessica Enright, for your effectiveness and clarity in our meetings, for your understanding and ability to propose solutions to my problems, I've really enjoyed working with you! Further thanks also to my examiners, Chiara Poletto and Andrea Wilson, for taking the time and effort to review this work and for generating interesting discussion around it.

I would like to thank all members and visitors of my research group Tom, Daniel, Tijani, Jason, Gaël, but especially Gianluigi and Ewan for their patience, help, guidance and advice. Benefiting from your postdoc experience through formal and informal chats was really valuable. I was very lucky and grateful to be part of EPIC Scotland. I would like to thank all their members for making me feel welcome and allowing me to be in such a stimulating and interesting environment at the research-government interface. Thanks in particular to Theo, for your patience and being always available to help and answer SimInf-related or statistical questions.

I am very grateful for the financial aid provided by the University of Edinburgh and EPIC, without which this experience would not have been possible.

Thank you to past and present colleagues from the Roslin Institute. Being in such a diverse multifaceted and intellectually stimulating community has broadened my horizons. I wish to thank members from the Royal (Dick) School of Veterinary Studies, especially those who helped me to get a vet teaching experience and finally supported me in my decision to go back to vet practice alongside my research. Although it was a challenging shift to make, it has been a fulfilling choice so far, becoming now a career choice! Thank you to the Bishopton Vet and Raft Team, especially to Jonathan, for being so flexible and understanding, encouraging me to get back into practice whilst finishing my PhD!

To the friends I've made on the PhD journey within and outside of the research community, as the "PhD crises", which inevitably come along, couldn't be overcome without such a great support and loving bubble: Jon, Marina, Lucy, Lou-Eve, Lina, Darek, Mark... This also includes the amazing people I've had the pleasure of living with on Glengyle Terrace, Effie, Iain, Amanda and Dave, thanks for making my experience in Edinburgh so unforgettable! Thanks to Mark, I am very grateful for your constant support throughout this PhD. Thanks for showing me that it is possible to love your PhD, for being sometimes more excited than me about my work, for keeping me motivated, for helping me to learn Python and finally for providing this amazing thesis template :)! To my older friends from France, in particular Camille C., Camille F., Sophie P., Marion C., Géraldine, Charlotte W... for your support from the distance which has been a central element to achieve this.

Lastly, thank you to my family. First, to my dad, for your everlasting support, your trust, for always showing interest in what I do, and for doing everything you can to help me get where I want! Thanks to my mum, for keeping me focused, helping me do my best and, of course, for all the delicious food and cakes you bake to comfort me every time I come back to France. A special thank you to my sister, Elena, for your love, for your numerous visits, for being great at making me confident in my choices and for helping me 'dedramatise' whatever the situation is. This has been very precious all along my PhD. Thanks to my wider family, uncles, aunts and especially cousins, as being able to share my PhD experiences with people who have been through this, was always very helpful! Thanks to Cyrille, for making me feel like an amazing person and for being so proud of me. Your patience, your happiness and your energy have made the last stretch of this PhD so much more enjoyable. Also, thanks for knowing!

Friday 10th June, 2022

Contents

List of Figures

List of Tables

Abbreviation

BVD: Bovine Viral Diarrhoea
CC: contact chain
DZ: Data Zone
FMD: Foot and Mouth Disease
$GM - CC$: geometric mean contact chain
$GM - CC_w$: geometric mean of weighted contact chain
$GM - Deg$: Geometric mean degree
GSCC: Giant strongly connected component
GWCC: Giant weakly connected component
ICC: Ingoing contact chain
ICC_w: weighted ingoing contact chain
NPIs: Non-Pharmaceutical Interventions
OA: Output Area
OCC: Outgoing contact chain
OCC_w: weighted outgoing contact chain
SIR: Susceptible-Infectious-Recovered
SIMD: Scottish Index for Multiple Deprivation
SPL: Shortest Path Length
UK: United Kingdom

Publications

This thesis contains excerpts from the following papers:

Ruget, AS., Rossi, G., Pepler, P.T., Beaunée, G., Banks, CJ., Enright, J., & Kao, RR. Multi-species temporal network of livestock movements for disease spread. Appl Netw Sci 6, 15 (2021).
https://doi.org/10.1007/s41109-021-00354-x

Ruget, AS. , Banks, CJ., Enright, J., & Kao, RR. (2021). Risk of COVID-19 Introduction into the Scottish Hebrides and Strategies for Control. *Under consideration*

Ruget, AS., Beaunée, G., Enright, J., Banks, CJ., & Kao, RR. Role of voluntary measures for COVID-19 control in isolated territories: the case of Barra, Scotland. *In preparation*

1

Introduction

Infectious diseases have played a significant role in shaping human history. The Black Death during the 14th century is a famous example as the epidemic caused the death of at least one third of the European population. The pandemic had many social and economic consequences. For example, due to labour shortages serfs were able to demand higher wages and better working conditions. Subsequent social effects were felt immediately, and the outbreak is therefore thought to have changed the course of European history (Benedictow and Benedictow, 2004; Cohn, 2007). Similarly when Smallpox was introduced to the American continent by Europeans in the 16th century, the virus has had a devastating effect. It killed 90% of the native Americans, whose immune system had not been previously exposed to such infections. This same virus was also later used as possibly the first biological warfare during the American Revolutionary War in the 18th century (Patterson and Runge, 2002). One more example was when Cholera became prominent in the 19th century, it reached Great Britain and was responsible for a massive epidemic in London in 1854. At that time, the disease was thought to be caused by 'bad air', but when British physician John Snow mapped cholera cases in the area of London, he identified the source of the disease as the water from a public well pump (Snow, 1855). His studies of the pattern of the disease persuaded the local council to disable the well pump by removing its handle. John Snow is considered one of the fathers of modern epidemiology. Infectious diseases have had a similar devastating impact on animals. Rinderpest, also called cattle plague, is a disease of livestock which has had a considerable impact throughout history. It is a contagious viral disease of ruminants and pigs that could cause over 90% of morbidity and mortality. The disease is thought to have occurred as far back as the medieval period and was responsible for three major panzootics in Europe in the 18th century. At the time the epizootics were contained through measures such as travel bans, closure of markets and slaughter accompanied by compensation policies. It has been one of the most economically devastating diseases for domestic animals (Hünniger, 2020; Broad, 1983).

Until the beginning of the last century, infectious diseases were the major cause of human mortality, causing about fifty percent of all human deaths (CDC, 1999). More recently, advances in scientific and medical knowledge of infectious disease have helped to dramatically decrease this burden to less than a few percent in the industrialized world (Brachman, 2003). The spectacular decrease mainly relied on the implementation of public health measures and the development of antimicrobial treatments and vaccines. Two notable examples of successes in infectious disease control are the eradication of both smallpox virus and rinderpest virus through vaccination campaigns among humans and cattle, respectively (Henderson, 2011; Roeder, 2011). However, infectious diseases still constitute a considerable threat to human health. According to the Global Burden of Disease study, six infectious diseases were among the top ten causes of disability-adjusted life-years (DALYs) in children younger than 10 years in 2019 (Vos et al., 2020). In addition to known infectious diseases, new infectious diseases are continuously emerging. The majority of recurrent and emerging infectious disease threats are those which originate from animals and infect humans (Taylor et al., 2001). Initially, changes of practices over time, such as hunting of wild animals, domestication and human settlement in habitats high in wildlife biodiversity, had significantly driven the emergence of infectious diseases. More recently, fragmentation of wildlife habitat, intensification of agriculture and urbanisation have created more opportunities for pathogens to evolve, emerge and spread (Alirol et al., 2011; Jorgensen et al., 2004; Cohen, 2000; Reperant and Osterhaus, 2017; Reperant et al., 2012). Finally, globalisation has now increased the pandemic risk (Zimmermann et al., 2020). The emergence of COVID-19 in late 2019 has reemphasised the permanent threat that an emerging virus can represent and the need for continuously improving preparedness and response to such events (Velavan and Meyer, 2020).

Epidemiology is the study of the distribution and determinants of health-related states and events in specified populations (Dicker et al., 2006). The field of epidemiology is typically divided into two main branches: descriptive and analytic epidemiology. In descriptive epidemiology the objective is to provide information on disease patterns by considering various characteristics of person, place and time. The goal is to describe the health situation and to draw hypotheses based on these observations about the causes of these patterns and about the causes that increase risk of disease. Analytic epidemiology allows one to test these hypotheses to identify causative factors. By comparing groups of exposed and non-exposed individuals, it is possible to quantify associations between exposures and health outcomes. When sufficient evidence is available, identified factors associated with disease can help design appropriate prevention and control measures.

To support the evaluation of disease management activities, models in which disease transmission processes are represented explicitly, often referred to as mechanistic or process models, are widely used (Mancy et al., 2017). One of the values of process models lies in the ability to study the 'what if' scenarios in dynamical systems and provide *a priori* insight into consequences of disease spread and the impact of control strategies. Mathematical models give us the ability to develop intuitions and make predictions about behaviours beyond our observations. The use of mathematics as a tool to study the spread of infectious diseases has been increasing in the past decades and is now dominant in epidemic modelling.

1.1 Mathematical models for infectious diseases

A model is a simplification of reality, which can be used to help understand, explain, or predict a complex phenomenon. The first mathematical model in epidemiology was built in 1760 by Daniel Bernoulli to show that the inoculation of smallpox may be advantageous (Bacaër, 2011). In 1906 Hamer suggested that the spread of measles must depend on the number of susceptible and infectious individuals (Hamer, 1906). Ronald Ross received a Nobel Prize for his work on the dynamics of malaria between mosquitoes and humans, when he showed, by using differential equations, that reducing the mosquito population below a critical level would be sufficient to eliminate the disease (Ross, 1911). All these studies led to the foundation of the approach to epidemiology based on compartmental models.

1.1.1 Compartmental models

In 1927 Kermack and McKendrick extended the models of Ross and created a system dynamic model of infectious disease transmission: the compartmental model (Kermack and McKendrick, 1927). A compartmental model divides the population into compartments representing disease states. The SIR model is an example of a compartmental model where three states are included (see Fig. 1.1):

- -S stands for 'Susceptible', susceptible individuals are not infected, and can get infected;
- -I for 'Infectious', Infectious individuals are infected, and can transmit the disease;
- -R for 'removed', removed individuals are either immune or dead, and as such they do not contribute to the disease spread anymore.

The SIR model is relevant for diseases conferring a lifelong immunity, with a short incubation period, and which spreads quickly. Endemic diseases that propagate on a time-scale similar to the turnover in the population, or epidemic diseases where the turnover in the population is high relative to the generation time of the disease, are not represented well by this model (Vynnycky and White, 2010). A population, of size N is divided between the three states such that S + I + R = N. If the turnover in the population is low compared to the outbreak duration, N is assumed constant, otherwise not. If the size of the population changes over time, it is important to ask the question whether as the population grows, the number of contacts between individuals remains unchanged, in which case the transmission is said to be frequency-dependent; or the growth in population induces increased crowding and a higher number of contacts per time unit, so that the risk of infection increases. In this instance, transmission is said to be density-dependent.

Individuals are characterised by their state (S, I or R), they are otherwise identical. The transition between states in a compartmental model is governed by differential equations. Ordinary differential equations (ODEs) for a frequency dependent deterministic SIR model are given by:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\lambda(t)S(t)}{N} \\ \frac{dI(t)}{dt} &= \frac{\lambda(t)S(t)}{N} - \gamma I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t) \end{aligned} \tag{1.1}$$

Figure 1.1: Schematic representation of a compartmental model including three diseases states: 'S' for susceptible, 'I' for infectious, 'R' for removed, and the transition parameters between states $\lambda = \beta I$ and γ.

where, S(t), I(t), and R(t) are the numbers of susceptible, infectious, and removed individuals at time t. The differential equations define the rate of change in the number of individuals in each compartment. For example, the number of those susceptible who are newly infected per unit time is given by the product of the force of infection $\lambda(t)$ and the number of those susceptible at time t divided by the population size, as we consider a frequency dependent model. The force of infection, i.e. the rate at which susceptible individuals become infected per unit time, is a key parameter in epidemiology. If individuals are assumed to

contact each other randomly regardless of their age or other characteristics such as space (also called random or homogeneous mixing), $\lambda(t)$ can be replaced by:

$$\lambda(t) = \beta I(t) \tag{1.2}$$

β is the *per capita* rate at which two specific individuals come into *effective contact* per unit time. An *effective contact* is defined, by Abbey, as a contact which would be sufficient to lead to infection, were it to occur between a susceptible and an infectious individual (Abbey, 1952). They become removed (e.g. resistant to infection, as with measles) at a rate γ per unit time.

The temporal evolution of S(t), I(t), and R(t) can be calculated and plotted as a result of the model (see Fig. 1.2).

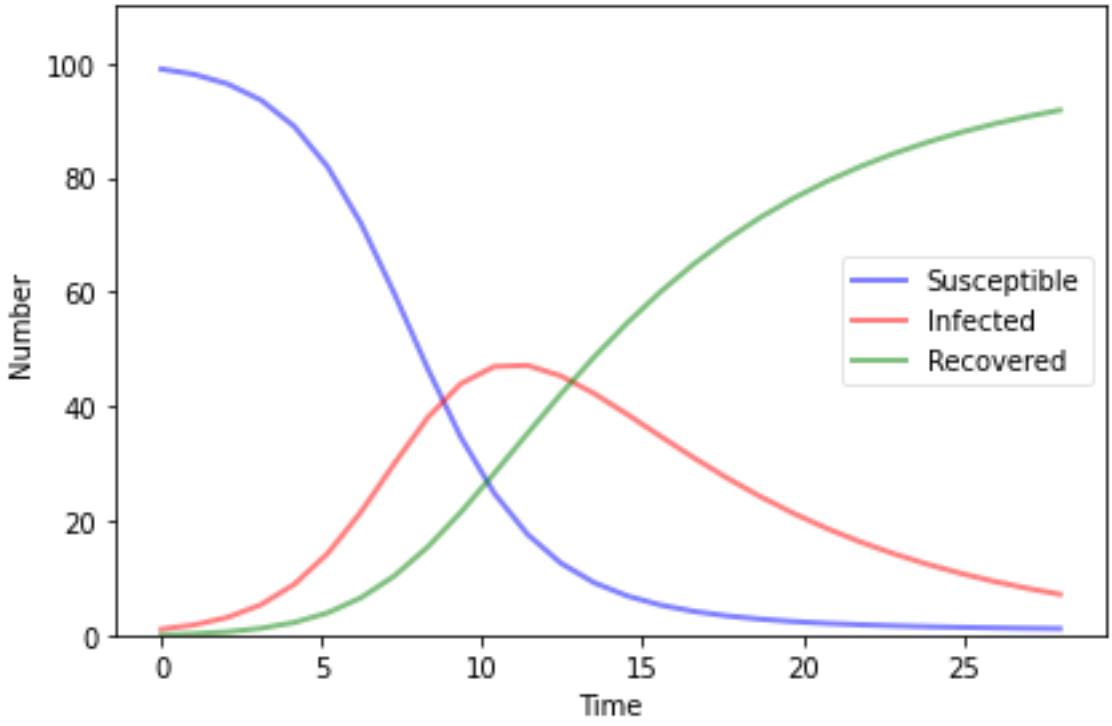

Figure 1.2: Number of susceptible (blue), infected (red) and recovered (green) individuals over time from a simple SIR model simulation. Model was initiated with 1 infected and 99 susceptible individuals in closed population and parameterised with $\beta = 0.75$, $\gamma = 0.15$.

The basic reproduction number R_0 is the average number of secondary cases resulting from one infectious individual following their introduction into a totally susceptible population. By definition the value of R_0 provides an information on whether the number of infected individuals will increase or decrease at the start

of the epidemic, i.e. whether $\frac{dI(t)}{dt} > 0$ at $t = 0$. We can deduct the formula for R_0 by rewriting the second equation of Eq. (3.1) combined with Eq. (1.2) as follows:

$$\frac{dI(t)}{dt} = \left(\frac{\beta S(t)}{N\gamma} - 1\right)\gamma I(t) \tag{1.3}$$

Since R_0 is calculated in a totally susceptible population, $S(t)$ and N simplify leaving the condition $\frac{dI(t)}{dt} > 0$ equivalent to $\frac{\beta}{\gamma} > 1$. In this case, the basic reproduction number is defined as:

$$R_0 = \frac{\beta}{\gamma} \tag{1.4}$$

The condition $R_0 > 1$ is considered an important threshold condition that must be satisfied for the epidemic to spread after the introduction of one infected individual in a totally susceptible population. If R_0 is less than one, the epidemic is expected to die out. The calculation of R_0 allows the testing of control strategies, for example by calculating the proportion of the population that would need to be vaccinated (Keeling, 1997), or the number of animals that might need to be culled to contain an epidemic (Ferguson et al., 2001a).

For a simple SIR model, R_0 can be easily calculated from the ODE. But for more complex compartmental models, especially those with more infected states, for example two pre-infectious states E_1 and E_2 from which individuals progress to infectiousness at different rates, R_0 is harder to calculate. Diekmann et al. (1990) proposed an alternative method to determine R_0 for an ODE compartmental model by using the next generation matrix (NGM).

We consider an SEIR model with the following states: S susceptible; E_1 latently infected of category 1; E_2 latently infected of category 2; I infectious; and R removed. We assume that susceptibles enter the E_1 and E_2 states following exposure to infection in a fixed ratio q and $1 - q$ respectively. ν_1 and ν_2 are the rates of leaving the respective pre-infectious states E_1 and E_2.

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\frac{\beta I(t) S(t)}{N} \\
\frac{dE_1(t)}{dt} &= q\frac{\beta I(t) S(t)}{N} - \nu_1 E_1(t) \\
\frac{dE_2(t)}{dt} &= (1-q)\frac{\beta I(t) S(t)}{N} - \nu_2 E_2(t) \\
\frac{dI(t)}{dt} &= \nu_1 E_1(t) + \nu_2 E_2(t) - \gamma I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t)
\end{aligned} \tag{1.5}$$

The matrix T corresponds to transmissions and the matrix Σ to transitions. The two matrices for our model are defined as follows:

$$T = \begin{pmatrix} 0 & 0 & q\beta \\ 0 & 0 & (1-q)\beta \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} -\nu_1 & 0 & 0 \\ 0 & -\nu_2 & 0 \\ \nu_2 & \nu_1 & -\gamma \end{pmatrix}$$

Diekmann et al. (1990) show that the dominant eigenvalue of the NGM defined as $-T\Sigma^{-1}$ is equal to R_0. This method can be used for complex models with multiple infected states, but also for models with asymmetric transmission between multiple hosts or heterogeneous contact structure (Diekmann et al., 2010).

The models described so far are deterministic and aimed to describe what would happen on average in a given situation. In deterministic models, outcomes are entirely predictable based on the initial conditions, and the parameter values. In contrast, the output of stochastic models depends on parameter values and the randomness of the underlying process. Deterministic models only need to be run once, because they have the same output every time; whereas stochastic models need to be run multiple times, to obtain a distribution of outputs, and give a good insight to the range of behaviours in the system. Because epidemiological systems are never fully known, it is often preferable to include stochasticity in the model to account for uncertainties. For example, although the average outcome of introducing one infected individual into a large population might be meaningful, it isn't necessarily the case in small populations. The model might predict that after one infectious person enters the population, < 1 new infections will subsequently occur each day, whereas it is unrealistic to have fractions of new infections. Finally, randomness might greatly affect the outbreak size, making the average outbreak size less informative.

The Reed-Frost model is a common way to include stochasticity in an epidemiological model and relies on a binomial stochastic process. In the deterministic formulation of the model, the risk of infection $\lambda(t)$ was assumed to be proportional to the number of infectious individuals in the population (cf Eq. (1.2)). This approximation is poor if the population is small, because of the probability that one susceptible individual will come into in contact with more than one

infectious increases and only one of those contacts would lead to infection. The Reed-Frost formula defines $\lambda(t)$ as follows instead:

$$\lambda(t) = 1 - (1 - p)^{I(t)} \tag{1.6}$$

where p is the probability of an effective contact between two specific individuals in each time step, $(1 - p)^{I(t)}$ being the probability that a susceptible individual does not become infected by at least one other individual at time t. The stochastic models used in this work use a similar method to implement randomness in the disease transmission process.

Since they were first developed, these types of models have helped to inform on disease dynamics and control programmes, quantifying the impact of measures applied to proportions of the population ("how many people should be vaccinated?") (Ramsay et al., 1994; Greenhalgh, 1990; Lipsitch et al., 2003). However, the assumptions made are strong. Some simplifications may limit the capability of mathematical models to represent the spread of diseases in detail. Some of these assumptions are listed below:

- The homogeneous and fully-mixed assumption means that individuals are fully connected and all make the same number of contacts per unit time. This is often unrealistic as for human populations for example, individuals usually contact only a small proportion of a large population. The heterogeneity of contact between individuals can have some important epidemiological implications.

- This assumption is limited in the variety of change in human behaviour due to infection or disease awareness which are not generalised across the population.

- Disease parameters such as infectious period, and probability of transmission are represented with mean values that fail to represent the heterogeneous nature of disease spread.

- A small set of parameters is considered to capture all the factors associated with the epidemic spread process when the reality is more complex.

To account for more details the population can be divided in subgroups within which homogeneous mixing is assumed but where subgroups have a varying rate of contacts. The subgroups can for example correspond to age groups as there is evidence that contact patterns may vary greatly according to age (Wallinga et al., 2006; Mossong et al., 2008). Taking into account the heterogeneous mixing potentially allows one to model control strategies as contact patterns strongly

determine the effect of interventions against infections (Brisson et al., 2000). The population can also be divided according to other characteristics such as professions or infectivity (Mkhatshwa and Mummert, 2010; Kwok et al., 2007). Individuals or events that generate disproportionately large numbers of secondary cases than average are termed *superspreaders*, and have been identified in many infectious diseases, including SARS (Stein, 2011).

In the past decades, the increasing availability of computational power and data have become useful in developing insights into the importance of population structure.

1.1.2 Networks

The heterogeneity of contact structures can be detailed using networks. The use of networks is especially meaningful if each individual is in contact with a small proportion of the general population. The earliest known description that we could call a network formulation was written by Euler in 1736, where he described the Seven Bridges of Koenigsberg problem using nodes, and links (Euler, 1741). The resolution of this problem laid the foundation of graph theory (Bondy et al., 1976).

1.1.2.1 *Movement network and network analysis*

A network encodes a group of entities, commonly called nodes, and the interactions between them, called links (see Fig. 1.3). For the study of disease spread, the nodes can represent various entities, such as a person, household, city, herd or farm. The links between nodes are disease relevant contacts.

A study of sexual contacts and the occurrence of AIDS is one of the first to evidence the importance of integrating social relationships into the study of disease transmission (Klovdahl, 1985). In human diseases, networks are relevant for sexually transmitted infection, as the number of contacts between individuals is small and heterogeneous (May and Anderson, 1987). However, studies rarely include explicit, exhaustive network data for human diseases, as those are often not available. Frequently networks are generated from partial data such as mobile phone, demographic or census data, using for example gravity models (Krings et al., 2009; Xia et al., 2004; Gu and Pang, 2008). In contrast, livestock movement data are now commonly available, since many countries have developed livestock movement databases after the bovine spongiform encephalopathy (BSE) investigations in Europe in the 90s (Dubé et al., 2009). Livestock movements have now been tracked and recorded on a daily basis, for several decades. These movements are particularly informative since, as opposed to human populations where people move freely and potentially have many epidemiologic-

ally relevant contacts each day, livestock generally remain on a farm until moved for trade purpose or agricultural events. Consequently livestock trade plays a crucial role in infectious disease transmission between farms (Fèvre et al., 2006; Green et al., 2006a) and movement data therefore allows us to generate disease relevant contact networks.

Many network Analysis concepts originate from the field of social science (Wasserman et al., 1994a) and can be used to characterise the structure of the network or the relative importance of nodes, which has proven informative for our understanding of disease risk. For example, livestock movement networks are often characterised by *small-world* properties and heavy right skew degree distribution (Kao R.R et al., 2006; Kiss et al., 2006a). The right skewed distribution of contacts means that there is a wide heterogeneity in the number of connections between individuals. A few nodes have many contacts, and most of the nodes have only few connections. A small-world network is defined as a highly clustered network, with a few long-range connections allowing for a short average path length (Watts and Strogatz, 1998). The clustering coefficient is the probability that two nodes are connected given that they have a common neighbour. These features can be exploited to target surveillance and control and drastically reduce the epidemic risk (Kao R.R et al., 2006; Kiss et al., 2006b).

Centrality measures are used to quantify the importance of nodes or links with regards to disease risk. Useful measures include:

- Degree centrality: number of links a node has (Newman, 2018)
- Betweenness centrality: considering all pair of nodes in the network, number of times a node or a link is traversed by the shortest-path between those nodes (Freeman, 1978),
- Eigenvector centrality: measure of influence in the network that takes into account second-order connections (i.e., connections of connections) (Silk et al., 2017),
- PageRank centrality: variant of eigenvector centrality, measure of a node's importance while giving consideration to the importance of its neighbors in a directed network (Newman, 2010)

The 2001 Foot and Mouth Disease epidemic in the UK illustrated how network analysis results can be relevant to disease control and therefore greatly enhanced the use of movement networks (Kao, 2002; Kao et al., 2006; Green et al., 2006b). In this occasion, network analysis concepts helped understand the dynamic of the initial spread. For example, a few long-distance movements acted as bridge, seeding the virus in new otherwise low risk communities (Shirley and Rushton, 2005; Kao, 2002; Gibbens et al., 2001). These movements are characterised by

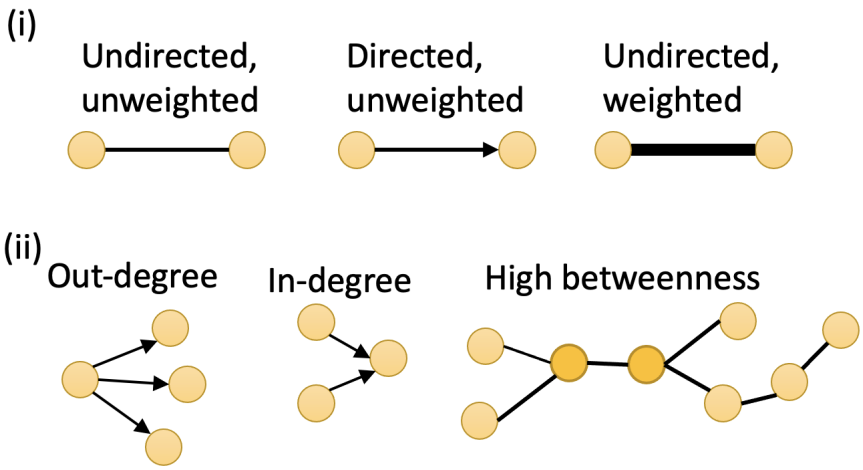

Figure 1.3: Schematic representation of network entities. Yellow dots represent "nodes" which are joined by "links", which can be undirected, directed, weighted or unweighted (i). The bottom part (ii) illustrates centrality measures: "out-degree" and "in-degree" in a directed network. The schema on the bottom left hand-side represents a network where darker orange nodes have a high betweenness.

a high betweenness centrality. Some markets (Longtown auction market, the largest in GB), characterised by a high degree centrality, were identified as high-risk premises in the network (Kao, 2002), as they played a dominant role in spreading disease. Markets are now known as high risk premises (Christley et al., 2005; Robinson and Christley, 2007a). Finally, PageRank has been identified as a useful measure to identify influential spreaders of information in real-world social networks or webpages link structure (Pei et al., 2014). It is a powerful metric which can also help predict the size of the epidemic (Bucur and Holme, 2019).

Most analytical results are available for static networks, where links between nodes are considered permanent. However, movement data is inherently dynamic. In some cases, these contacts might repeat over time, this is true for commuting patterns or social relationships, but livestock movement pattern for instance exhibit important variations which are not necessarily consistent over time (Bajardi et al., 2011a). This temporal variability has been shown to be a key determinant of the vulnerability to infection emergence and propagation (Bajardi et al., 2011a; Dutta et al., 2014).

1.1.2.2 *Network models*

The study of the structure and topology of networks can provide insights in disease risk, but networks can also be used to replace the hypothesized fully-mixed population in disease spreading models. Network models of disease work globally in the same way as the fully-mixed compartmental models. Here, the

transmission rate is defined as the probability per unit time that infection will be transmitted between two individuals connected by a link. As in compartmental models, individuals are characterized by their health states; numerical simulations of epidemics in complex networks use individual-based models to represent contact patterns between individuals and infection probability on links (Duan et al., 2015).

In addition, centrality measures are inherently limited in how they inform on outbreak dynamics due to strong fluctuations of movement patterns in time and the importance of outbreak initial conditions. Bajardi et al. (2012) show in their study how to use a network model instead of centrality measures to identify important links or nodes. They use network disease simulations to identify robust spreading paths that are stable across different initial conditions. Their results also provide guidance on the targeting of premises with high probability of early infection which can serve as sentinels and, if infected, would provide critical information on the outbreak origin. These results are useful to help design surveillance programmes or early response in an epidemic.

1.1.3 Meta-population

The metapopulation concept arises from theoretical ecology and conservation biology, where it is assumed that the distribution of species can be described as a system of sub-populations, each of which may be subject to extinction and recolonisation by dispersing individuals (Hanski and Gilpin, 1991). The persistence of a population depends on various factors such as population dynamics, population density, availability of resources, etc. The metapopulation is of the 'colonising' agent, e.g. for the spread of a disease the metapopulation is infected individuals and the resource is the proportion of susceptible (Grenfell and Harwood, 1997). Thus, this framework can be useful to model disease spread, as the global transmission dynamic can be sometimes best described by a combination of subpopulations, in which each of the disease dynamics is modelled independently, accounting for some interactions between subpopulations through 'migration' process (see Fig. 1.4).

Metapopulation models for infectious disease allow one to account for relevant details at various scales between populations and within populations. For Mycobacterium avium subsp. paratuberculosis (MAP) the herd size, age structure and farm management, e.g. renewal rate, are factors that largely impact the pathogen spread within and between populations. Beaunée et al. (2015); Brooks-Pollock et al. (2014) integrated this information in their modelling study to better understand pathogen spread (MAP and bovine Tuberculosis respectively) from local to regional or national scales. Bovine Viral Diarrhoea (BVD) is another example where integrating the age and sex structure of cattle herds into compart-

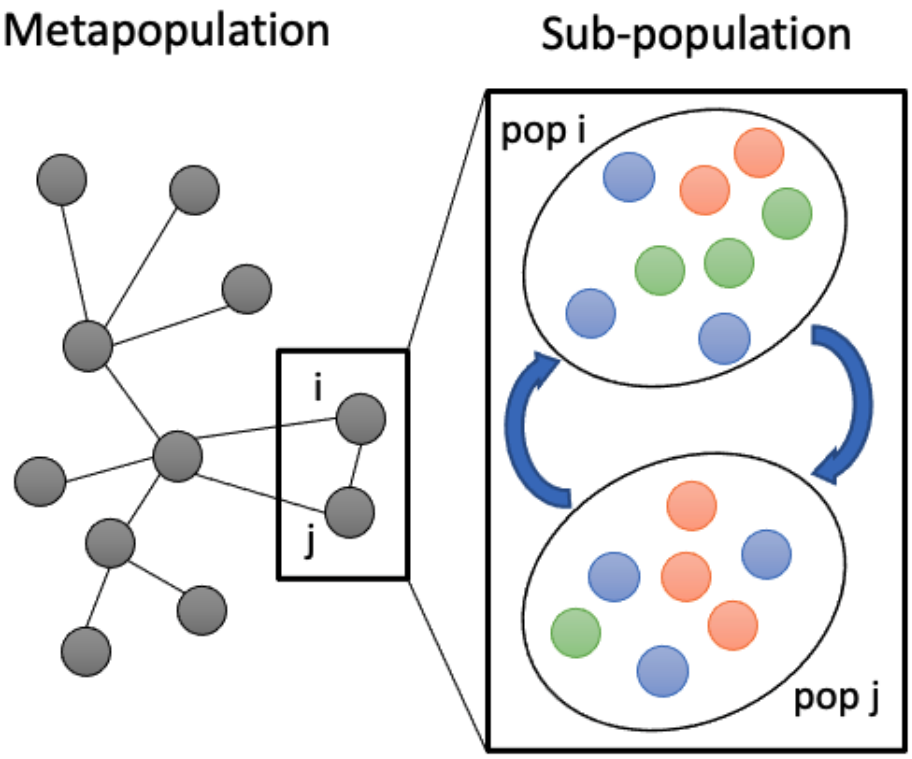

Figure 1.4: Schematic representation of a metapopulation model. The system is composed of a heterogeneous network of subpopulations or patches, connected by migration processes. Each patch contains a population of individuals who are characterized with respect to their disease state (e.g. susceptible, infected, removed), and identified with a different color in the picture. Individuals can move from a subpopulation to another on the network of connections among subpopulations.

mental frameworks is relevant due to the heterogeneity of transmission. With BVD, cows becoming infected during gestation may give birth to Persistently Infected (PI) calves, which remain highly infective throughout their life. Iotti et al. (2019) used a multiscale approach where the metapopulation model incorporates some relevant characteristics of the complexity of the infection dynamics (e.g. horizontal and vertical transmission, PIs...), alongside the heterogeneities and temporal variations of trade movements. The detailed model allowed them to demonstrate differences in roles for BVD spread between farm productive contexts, concluding that a control strategy targeting dairy farms would lead to significantly higher prevalence reduction, than targeting other production compartments. In the model developed by Keeling and Gilligan (2000) for bubonic plague, the spatial heterogeneity and the dynamic of infection within populations of humans, or rats and fleas are critical to understanding the probability for disease persistence in the rodent population and the chances for major human epidemics. Their results show that relatively small rodent metapopulations allow the disease to persist globally for many years, potentially explaining why

historically the plague persisted despite long disease-free periods, and how the disease re-occurred in cities with tight quarantine control.

Metapopulation models allow for a more detailed representation of the reality but adding complexity doesn't come without a cost. Most analytical results and calculation of useful parameters are better defined for simpler models. For example, the basic reproduction number, often used as a threshold to design control strategies, doesn't have a straightforward equivalent in metapopulation models. Colizza and Vespignani (2007), however, define an invasion threshold for heterogeneous metapopulation networks. They provide an explicit expression of the threshold that sets a critical value of the diffusion/mobility rate below, which the epidemic is not able to spread to a macroscopic fraction of subpopulations. This constitutes a useful result to understand the impact of travel restrictions in epidemic containment.

1.2 Mobility data and disease models in different epidemic contexts

1.2.1 Peacetime and preparedness for the next epidemic

The availability of epidemic data is an important constraint to the improvement of epidemic modelling and forecasting (De Angelis et al., 2015). As opposed to weather forecasting, where data becomes available every day, epidemic data is rare and context specific. Epidemics are intermittent as they refer to an increase, often sudden, in the number of cases of a disease above what is normally expected in that population in that area (CDC, 2012). In addition, two successive epidemics will often not involve the same pathogen. The disease relevant characteristics of the system are likely to change over time e.g. demography, contact pattern, behaviours, etc. All of these factors make epidemic modelling challenging.

Given the paucity of data, there is a need for improving our understanding and preparedness in peacetime, i.e. in between epidemics. For example, Ferguson et al. (2006); Halloran et al. (2008) developed a model to assess the impact of strategies for mitigating an influenza pandemic. When COVID-19 emerged and started to spread in late 2019, they were able to rapidly adapt this framework to give insight into the necessary restrictions to achieve mitigation or suppression (Ferguson et al., 2020).

Since the outcome of an epidemic depends on a combination between the intrinsic characteristics of the pathogen and the characteristics of the population including pattern of disease relevant contacts, studying contact structure on its own can help understand potential risks, speculate on outbreak outcomes or identify central nodes to target for strategies. The repeated outbreaks of avian influenza viruses in domestic birds (Chatziprodromidou et al., 2018) constitutes

an example where network analysis can help assess infection risk associated with movement patterns in peacetime. For example, the network analysis of the poultry industry in Great Britain by Dent et al. (2008) highlighted that premises sending birds to multiple slaughterhouses, or housing multiple species, could act as a bridge between otherwise separate sectors of the industry, resulting in the potential for large epidemics. Other studies of live bird markets (LBM) networks show that the removal of a small fraction of nodes (5%) is sufficient to dramatically reduce the network's connectedness (Moyen et al., 2018). The node removal can be achieved by the implementation of thorough, daily disinfection of the market environment as well as of traders' vehicles and equipment in only a small number of hubs, preventing disease spread (Fournié et al., 2013). These targeted interventions would be an effective alternative to a complete ban which is difficult to enforce. These network analysis findings, generated in peacetime, are of particular relevance for policy development.

A network can also encompass different type of contacts or different types of nodes. Silk et al. (2018) use a multilayer network to quantify transmission potential in multi-host systems (European badgers *Meles meles* and domestic cattle) and incorporating environmental transmission. Their network takes into account direct and indirect contacts between and within the two species. They find that indirect contacts via the environment at badger latrines on pasture are likely to be important for transmission within badger populations and between badgers and cattle. They also identify that badgers could play an important role for within-and between-species transmission. All these findings have implications for disease management interventions in this system, but can also provide general insights into transmission in other multi-host disease systems. There is a need for exploring the role of combining multiple species in network analysis since although most infectious diseases can affect several host species (Taylor et al., 2001), network analysis studies have generally focused on single species contact networks.

Network analysis studies performed in peacetime can focus on proposing new network metrics that could become useful for epidemic containment. Although there exists a number of static network metrics to assess node centrality in the network, available metrics for dynamic networks remain limited. Notable examples include: Valdano et al. (2015) define the node's loyalty as a local measure of its tendency to maintain contacts with the same elements over time, and uncover important non-trivial correlations with the node's epidemic risk. Deriving from the concept of reachability proposed by Holme (2005), Dubé et al. (2008) and Konschake et al. (2013) applied the metric to a livestock movement network under the name of infection chain or contact chain respectively. Vidondo and Voelkl (2018) show using a similar metric that their dynamic measures give bet-

ter risk estimates than the static counterpart and can help improve surveillance schemes.

1.2.2 During an epidemic: importation phase

When an infectious disease is emerging or circulating elsewhere, estimating the importation risk can help optimise surveillance and concentrate resources where needed. Gaining insights into the potential effectiveness of control measures to contain an outbreak can support decisions.

For example, when the new Influenza A(H1N1) virus started to spread in 2009, Balcan et al. (2009) proposed a Monte Carlo likelihood analysis based on human mobility to inform time of the epidemic peak in the Northern hemisphere as well as the number of hospitalisations and attack rate, i.e. the proportion of the population who will become infected. These results allow the improvement of preparedness and epidemic response. Similarly, Poletto et al. (2014a) assessed the impact of travel restrictions on the international spread of the 2014 West African Ebola epidemic and found that an 80% reduction of passenger traffic flow would delay international spread by only a few weeks. Such measures can, in addition, generate heavy logistical constraints on the management of the epidemic in affected countries causing shortages of food, energy, essential resources, and difficulties in importing medical personnel and materials (World Health Organization, 2014; African Union, 2014; FAO, 2014). Interestingly, for the 2002-2003 SARS epidemic or the 2009 H1N1 empirically observed travel reductions were due to a change of behaviours, self-imposed by the population, as they occurred without being enforced by the government. In 2014 for the Ebola epidemic, travel restrictions were imposed by national authorities or travel companies as a top-down decision.

As it has been shown by a number of studies (Hollingsworth et al., 2006; Cooper et al., 2006; Colizza et al., 2007a), travel reduction only leads to a delay in disease introduction. The mathematical reason is that a border control translates into a linear decrease of the number of infected travelling which aims at counterbalancing an exponential increase leading only to a shift of the epidemic curve (e.g. number of infected individuals over time) to the right (Tomba and Wallinga, 2008). There exists however a threshold effect, where one might achieve containment, but the level of reduction would be high (over 99% (Colizza et al., 2007a)) for an influenza pandemic for example. The threshold value doesn't depend only on mobility (i.e. "how many people are travelling"), but also on topology (i.e. "How this network is organised?") (Colizza and Vespignani, 2007; Bajardi et al., 2011b). If the airline travel network was organised as a grid, containment would be easily achieved but this network is characterised by a high reachability, thanks to hubs, allowing the connection of any two places with a minimal num-

ber of changes (Colizza et al., 2007b). As a consequence, it was assumed that feasible movement reductions could not prevent the further spread of a pandemic and were not recommended in pandemic preparedness plans, such as the Pandemic Influenza Preparedness Framework issued by the WHO for example.

But when COVID-19 started to spread, widespread travel restrictions were implemented, with a complete ban occurring sometimes, i.e. no movement of individuals allowed. Several studies looked into the impact of this travel ban on case exportation from China to new countries and observed that the containment strategies adopted in China were effective at first, until the circulation of the virus in other countries allowed them to behave as seeds (Pinotti et al., 2020; Chinazzi et al., 2020; Clifford et al., 2020). Total travel bans which were previously unthinkable because of their huge economic impact became the first line measure in combination with lockdown (Kraemer et al., 2020; Flaxman et al., 2020; Jia et al., 2020), as slowing the spread became a priority to avoid saturation of the healthcare system.

1.2.3 During an epidemic: transmission phase

Once the new pathogen is seeded somewhere, the phase of local transmission can start and mathematical models can again be used to gain insight, understanding and inform decision. The time pressure is a big component as data and models are usually not ready and not fast enough. A typical problem faced early in an epidemic is the lack of data. The first studies in an epidemic are usually based on knowledge from previous epidemics and tend to make a range of assumptions (Ferguson et al., 2020). This will provide useful insights but potentially with large uncertainties. But as data become more available over time, the models and their assumptions can be refined to be more relevant to the context.

As the epidemic unfolds the number of reported cases over time will help construct the epidemic curve (Christensen et al., 1953; Philip et al., 1959). The use of model framework helps gain insights in the size, the rate of increase and the timing of the epidemic. Efforts are concentrated early on in estimating the basic reproduction number, which measures the transmission potential of a disease in a susceptible population, as it is the most widely used estimator of how severe an epidemic outbreak can be.

The basic reproduction number can be inferred from the epidemic curve using the growth rate of the epidemic Λ. For example, Lipsitch et al. (2003) estimated the growth rate of the SARS epidemic using the cumulative number of cases over time. Depending on the assumptions about the average durations and the distributions of both the pre-infectious and infectious periods the basic reproduction number can be expressed differently. If the pre-infectious period is very short

compared to the infectious period (D) or individuals are infectious immediately, the equation for the basic reproduction number is:

$$R_0 = 1 + \Lambda D \tag{1.7}$$

This assumes that the infectious period follows the exponential distribution. If the pre-infectious (D') and infectious periods follow the exponential distribution, the basic reproduction number can be deducted from the following equation:

$$R_0 = (1 + \Lambda D')(1 + \Lambda D) \tag{1.8}$$

When facing a new pathogen, the lengths of the infectious and pre-infectious periods are largely unknown. The serial interval is defined as the interval between successive infections in a chain of transmission, also called generation time. If pathogen specific parameters such as infectious and pre-infectious periods are still unknown an alternative equation using the serial interval and the ratio between the infectious period and the serial interval (Vynnycky and White, 2010) can be used.

Secondly control strategies for disease containment models can be designed and optimised using mathematical model results. The herd immunity threshold, which is the proportion of the population which needs to be immune for the infection incidence to be stable, is useful to provide a target for immunisation programmes (Vynnycky and White, 2010). It can be calculated from the ODEs of the model. When restrictions are implemented, one of the challenges is to know how these restrictions translate in change of transmission dynamics, especially when the restrictions are unprecedented. For example, in the COVID-19 pandemic, control strategies have relied on long range movement restrictions (international and domestic mobility) and contact mitigation (social distancing, bans on large gatherings, nonessential business and school and university closures, and possibly physical isolation). But without data it is difficult to know to what extent restrictions are applied by the population and how this translates into parameter change. Mobile phone data, for instance, has been useful to study how mobility has changed throughout the pandemic (Oliver et al., 2020; Pullano et al., 2020). The changes in physical contacts became available thanks to contact surveys performed during the COVID-19 pandemic in England for example (Gimma et al., 2021). Once one has a better idea of what consequences restrictions could have on parameter values, one can explore the effect of those restrictions by tweaking parameters in a model.

Instead of running a model with chosen parameter values to analyse the model output, one can use disease data, such as number of cases over time, to

adjust model parameters by fitting model output to disease data. Disease spreading is a complex process, so it is inevitable that the observation of this process leads to build complex models. Usually, model parameters were estimated using likelihood-based inference methods, but as models become more complex and data limited, the likelihood function often becomes intractable or not known (Beaumont et al., 2002). Approximate Bayesian Computations (ABC) methods are powerful as they allow to circumvent the use of the likelihood function and are based on an inherently simple mechanism: simulating data under the model of interest and comparing the output to the observed dataset (Sisson et al., 2018). In addition, as in Bayesian inference, prior beliefs about the model parameters, expressed through the prior distribution, are updated in light of new evidence, helping to constrain models to where they are sensible. In this way one can learn and deduce from data, the fitted parameters giving information on the disease dynamic. After COVID-19 was introduced in Scotland, a rapid spread followed until lockdown was implemented on March 23rd 2020. A study by Banks et al. (2020) aimed at characterising the change of transmission dynamics after the lockdown was implemented. This helped quantify the impact of such restrictions which is an essential part of evaluating the necessity for future implementation of similar measures.

This method essentially allows one to quantify the effectiveness of control strategies by estimating the impact of restrictions on the transmission dynamic.

1.3 Contributions and thesis outline

Mathematical models, movement and mobility data have proven useful for infectious disease management at any stage of an epidemic. Given the themes described above, the goal of this work is to further explore how movement and mobility can be used in different epidemic contexts to improve our understanding of disease risk. I address this goal by focusing on two main tasks:

1. Explore how the use of rich movement data in peacetime can help improve preparedness to face the next outbreak (Part I). In this first part, I use livestock movement data to provide insights in disease risk for the spread of a fast-spreading disease similar to FMD and to propose a tool to facilitate rapid decision early in an epidemic.

2. Investigate how the use of mobility and disease data during an epidemic can help inform strategies to mitigate introduction risk or assess the efficiency of control measures *a posteriori* (Part II). In this second part, I use human mobility data and the spread of COVID-19 in the Scottish Hebrides.

In Part I, I consider two approaches applicable in epidemic peacetime. The first approach (Chapter 2) aims at improving preparedness by exploring how

combining movement data from multiple species in a network analysis impacts disease risk. Although livestock movement networks have been widely studied in the UK and elsewhere, multiple species have rarely been considered jointly in the analysis. Most studies have generally focused on single species contact networks, because aggregation of movement data from different species is often difficult. In Chapter 2, I use defined static and temporal network metrics to investigate the importance of multi-species links in a cattle and sheep movement network. The aim is to investigate how the combination of cattle and sheep movements affects the topology of the static and dynamic networks and the potential impact for policy decisions.

The second approach (Chapter 3) focuses on developing a network tool which can be useful in the early stage of an epidemic when only little is known about the pathogen. Using the same cattle and sheep movement data as in Chapter 2, I propose the use of a measure based on contact chains calculated in a network weighted with transmission probabilities to assess the importance of premises in an outbreak. The objective is to introduce a new measure relevant for multispecies systems, which is easy to compute and effective in identifying the most central farms in the network. With the aim of validating the suggested measure, I compare the performance of the measure with the results of disease simulation models with asymmetric probabilities of transmission between species. This work, performed in peacetime, helps improve preparedness for the next epidemic and early response.

In Part II, I consider two real-time examples of mathematical modelling studies performed during an epidemic to estimate the risk of introduction (Chapter 4) and to inform the local transmission phase (Chapter 5).

In Chapter 4, I estimate the risk of introduction of COVID-19 to each of the Hebridean islands situated off the west coast of mainland Scotland due to individual movements, and explore control strategies to mitigate this risk. I use a combination of real human mobility data and census data to generate seasonally varying patterns of human movements amongst the Hebrides and from elsewhere. I consider three distinct periods: each of summer and winter 2019, illustrating a year prior to the pandemic, and summer 2020 illustrating a *pandemic summer*. Movements during these periods serve as input to simulate COVID-19 transmission from the mainland to the archipelago in a stochastic meta-population model allowing an exploration of the impact of seasonal variations on the risk of introduction and the effectiveness of non-pharmaceutical interventions.

In Chapter 5, I use the case of the COVID-19 epidemic on Barra, Scottish Hebrides, to fit parameters to disease data to quantify the change of transmission dynamics throughout the epidemic. A change of dynamic can be attrib-

uted to a change in number of contacts per individual and a change in mobility which would impact the potential further spread of the virus. I consider three successive phases with different transmission dynamics to model the outbreak, i.e. initial spread, transmission reduction by voluntary measures and lockdown. A stochastic compartmental model for the spread of COVID-19 is adjusted to positive test data to estimate key disease parameters in these three phases. As individual movements are not explicitly detailed in this study, the change in parameter values between phases encompass both the change in contacts and in mobility. I use the fitted parameters to calculate the basic reproduction number of the epidemic and to assess the relative impact of voluntary and policy-induced measures.

Part I

In peacetime: before the next outbreak

2

Using movement data and network analysis to gain insight: Multi-species dimensions of the cattle and sheep static and dynamic network

The methods and results of this chapter have contributed to the following publication:

***Ruget, AS.**, Rossi, G., Pepler, P.T., Beaunée, G., Banks, CJ., Enright, J., & Kao, RR. Multi-species temporal network of livestock movements for disease spread. Appl Netw Sci 6, 15 (2021).*

Abstract

In this chapter, I explore how the analysis of a multi-species movement network can help improve our understanding of disease risk. More specifically the objective here is to investigate the potential change in disease risk when considering multiple species in a movement network. Although both cattle and sheep networks have been previously studied, cattle-sheep multi-species networks have not generally been studied in-depth. The central question of this chapter is how the combination of cattle and sheep movements affects the results of the network analysis for disease spread on the combined network.

This analysis considers static and temporal representations of networks based on recorded animal movements. Network-based node importance measures of two single-species networks are computed and the top-ranked premises are compared with the ones in the multi-species network. In addition, temporal contact

chains are used to estimate the maximal size of an epidemic according to season considering single-species or multi-species networks.

I find that up to 70% of the risky premises (i.e. top 100 ranked premises according to network metrics) in the multi-species temporal network are not identified as such in the cattle nor the sheep temporal networks. The premises which are likely to drive the epidemic in this multi-species network differ from the ones in both the cattle and the sheep networks, indicating that their risk is derived from interaction between the two farming systems. Although sheep movements are highly seasonal, the estimated size of an epidemic is significantly larger in the multi-species network than in the cattle network, independently of the period of the year.

2.1 Background

Infectious diseases in livestock are of great concern as they pose an economic burden, compromise animal health and welfare, and threaten human health by contributing to the emergence of new zoonotic diseases. Mathematical models of infectious disease spread are useful tools to help us understand the drivers of an outbreak, and inform policy decisions. In a world where pandemics are becoming more likely (Morse, 2001; Jones et al., 2008; Madhav et al., 2017), the usefulness of modelling techniques is well recognised (Colizza et al., 2007a; Dye and Gay, 2003; Lessler et al., 2014). Presently, models of infectious disease spread at a population scale are typically based on two phenomena: (i) the infection dynamic, which depends on the characteristics of the disease itself (transmission rate, infectious period, etc.), and (ii) the contact patterns allowing for disease transmission, depending on the transmission routes of the disease.

Here, our interest is in the transmission of infectious livestock diseases, where the movements of live animals between farms are known to be one of the main transmission routes (Fèvre et al., 2006). Better knowledge and understanding of contact patterns is a key element for building realistic and useful models. However, detailed models can be computationally costly, and require a substantial amount of data in order to be fitted properly.

When outbreaks occur, policy makers need rapid and robust information to define their strategy and support decisions at the early stages of the epidemic, when data are still limited. A better understanding of the structure of the livestock movement network and its characteristics is therefore useful, both to understand their role in the spread of endemic diseases such as bovine Tuberculosis (Boehm et al., 2009; Palisson et al., 2016; Brooks-Pollock et al., 2014; Green et al., 2008), or BVD (Tinsley et al., 2012)), and to inform policies to control a newly introduced disease in an early stage. The 2001 FMD epidemic provided considerable incentive to study and use livestock movements for network analysis (Ortiz-Pelaez et al., 2006; Christley et al., 2005; Kao et al., 2006; Robinson and Christley, 2007b; Robinson et al., 2007; Kiss et al., 2006a; Vernon and Keeling, 2009; Volkova et al., 2010). Analysis of contact networks has proven useful to help identify key actors in terms of disease spread.

Although most infectious diseases can affect several host species (Taylor et al., 2001), network analysis studies have generally focused on single species contact networks. Notable exceptions include Boehm et al. (2009), Nöremark et al. (2011), Kao et al. (2006), and Mohr et al. (2018). Practically, aggregation of movement data from different species is often difficult, because (i) data are recorded separately, often stored in different databases, and possibly managed by different administrative authorities; and (ii) the databases might have different formats or

contain different levels of information, and therefore need to be homogenised before use.

The cattle and sheep farming systems are strongly linked in Scotland, because on approximately half of the cattle farms, sheep are also raised. This allows ample opportunity for transmission of diseases between the two species, with FMD and bluetongue virus (BTV) being notable examples of diseases affecting both species. As a consequence, mixed-species farms can link groups of farms that would not be in contact in the network if the species were considered separately. From the network point of view, this might also have consequences for metrics describing the general structure, as well as the ranking of importance of nodes. It is therefore crucial to explore the multi-species network in order to highlight and quantify potential consequences for disease risk.

Livestock movement network analyses have been performed mostly on static networks, where most analytic results are available (Newman, 2018). A static network assumes that the change in the set of contacts are negligible over the course of the epidemic (Enright and Kao, 2018). In reality, livestock movements have an inherent temporal component, that are highly relevant to transmission, as they occur on a daily basis and constitute discrete events. As well as being intermittent, movements of livestock are not necessarily consistent over time (Bajardi et al., 2011a). A number of studies have shown that dynamic network analyses of livestock movements outperform those from static network analyses, when the aim is an in-depth understanding of disease spreading processes (Lentz et al., 2016; Vidondo and Voelkl, 2018; Rossi et al., 2017), or predictions of epidemic risks (Valdano et al., 2015). The study of the dynamics of the cattle-sheep network in Scotland is of interest, because as well as the general dynamics of livestock networks two farming systems are considered which have distinct seasonalities and varying trading behaviours, and the interaction between these systems.

The aim of this chapter is to understand how the sheep and cattle movement networks interact, and the implication for understanding disease spread. After describing the general characteristics of the networks, I analyse the static movement networks, and compare the results of the single- and multi-species networks. Secondly I analyse and compare the cattle, sheep, and multi-species dynamic networks. The dynamic network analysis exhibited important differences between the single-species and the multi-species networks, providing evidence that the premises driving epidemics would not be the same in the single-species and the multi-species networks. These results would have important consequences for disease control.

2.2 Material and Methods

2.2.1 Data

Cattle movement data were obtained from the Cattle Tracing System (CTS), Animal Plant and Health Agency (APHA), and sheep movement data were retrieved from ScotEID, the livestock traceability system for Scotland managed by the Scottish Agricultural Organization Society (SOAS) on behalf of the Scottish Government. I consider movements within Scotland only: between premises, which can be farms, markets or shows. Our interest is in the control of an outbreak after introduction, and therefore movements to or from outside of Scotland are ignored. Births, deaths, and movements to slaughterhouses are also ignored, because the length of the period considered in the study (i.e. four weeks) is short compared to the turnover in the population. Some characteristics of the data are summarised in Table 2.1.

Table 2.1: General characteristics of the setting in figures.

Network	Nodes	Movements 2016	Animals moved	Distinct movements
Cattle [mixed]	10,731 [6,039]	89,963	591,933	33,271
Sheep [mixed]	12,078 [6,039]	67,453	2,406,062	26,514
Multi-species	16,758	-	-	53,051

The number of nodes describe the total number of farms raising cattle or sheep reporting at least one animal movement during 2016. The number of movements (batches), the total number of animal moved per species, as well as the headcount of cattle and sheep in Scotland are shown. The number of distinct movements correspond to the number of unique pair of origin and destination in each network.

Overall the sheep population is larger accounting for 6.83 million heads, while the cattle were 1.76 million. There were slightly more sheep farms than cattle farms; of these, 6,039 farms raised cattle and sheep on the same premises (i.e. 50% of the sheep farms, and 56% of the cattle farms). In addition to the farms, the data include 26 auction markets.

2.2.2 Network construction

I construct networks by considering each premises as a node, and animal movements between two premises as a directed link. If one movement of an animal between two premises occurred during the period considered, I assign a permanent link between these two premises in the static network. When relevant, the links are weighted depending on the number of animals moved and the probability of an animal being infected:

$$1 - (1 - \mu)^n \tag{2.1}$$

where μ is the probability of an animal being infected, and n the number of animals moved. The probabilities depend on the type of movement and the species. I used the parameter values estimated by Kao et al. (2006) in the 2001 FMD epidemic in GB:

$\mu_1 = 0.02$ for a sheep movement between two farms;

$\mu_2 = 1$ for a cattle movement between two farms;

$\mu_3 = 0.004$ for a sheep movement from a market;

$\mu_4 = 0.02$ for a cattle movement from a market.

These weights are relevant for an infectious disease similar to FMD, where the infectiousness of sheep is lower than that of cattle (Geering, 1967; Gibson and Donaldson, 1986; Sørensen et al., 2000; Ferguson et al., 2001b).

In the dynamic network each link is annotated with a time variable equal to the date of the animal movement (i.e. I assume these movements occur on a single day).

2.2.3 Static network analysis

In the static network analysis, the links are weighted as defined by Eq. (2.1). I consider the static networks in successive 4-week periods. This allows us to highlight (i) short-term changes in the network structure, which would be relevant for the control of a fast-spreading disease, and (ii) temporal variation according to the season. Livestock movements are generally seasonal, depending on the species and type of production. In Scotland the cattle network typically shows two peaks; the largest is observed in spring, and the second largest in autumn (Robinson and Christley, 2006), whereas the sheep network has one main period of high trading activity around September (Kiss et al., 2006a). These peaks can be seen in Figure 2.1, which shows the number of cattle and sheep moved in

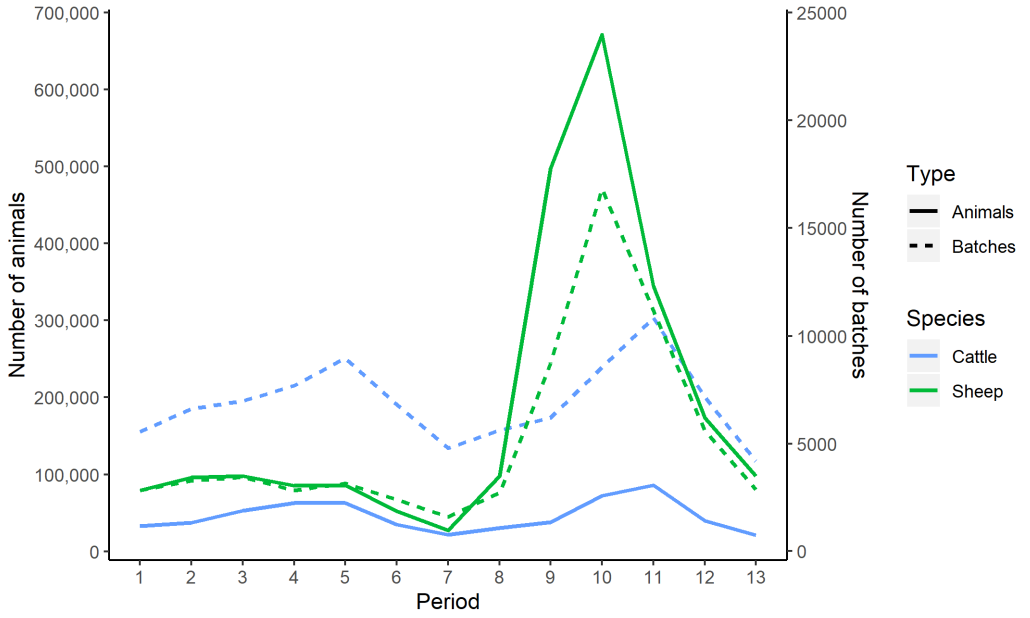

Figure 2.1: Variation in volume of cattle and sheep movements. The graph shows the number of animals (axis on right), and the number of batches (axis on left) moved per species per 4-week period of the year.

each 4-week period of the year. The static network is constructed considering weights as defined by Eq. (2.1).

The Fig. 2.2 shows a graphical representation of the movement network during a period dominated by cattle movements (5^{th}) or sheep movements (10^{th}).

I examine the overall characteristics of each network by calculating the average path length, clustering coefficient, edge density, component structure (number of components and sizes of the giant strongly and weakly connected components (GSCC and GWCC respectively) and diameter (definitions in Table 2.2). These measures are calculated for the single-species networks and the multi-species network, for each 4-week period of the year 2016.

I then calculate node centrality measures for all premises of the network, using the geometric mean degree, betweenness and PageRank (definitions in Table 2.2). In our case, *degree centrality* corresponds to the number of trading partners a farmer has. Because our network is directed, I differentiate in-degree (denoted $\mathrm{degree}_{\mathrm{in}}$), i.e. number of premises a farmer buys animals from, and out-degree (denoted $\mathrm{degree}_{\mathrm{out}}$), i.e. number of premises a farmer sells animals to. The geometric mean of the degree $\sqrt{\mathrm{degree}_{\mathrm{in}} \times \mathrm{degree}_{\mathrm{out}}}$ (denoted $\mathrm{GM-Deg}$), accounts for the risk of introducing the disease as well as spreading it further. *Betweenness centrality* is the frequency with which a premises is in the shortest path between pairs of premises in the network. Identifying high-betweenness premises is useful from a disease control point of view because these premises represent bridges, which can accelerate the epidemic by spreading diseases to previously unexposed communities of farms. *PageRank centrality* is based on an algorithm used

(a) 5th period (b) 10th period

Figure 2.2: Static network during the 5th and 10th periods of the year; Cattle movements are represented in blue, sheep movements in green. Nodes are represented by a pie chart where the light grey represents the proportion of outgoing movements, and the dark grey the proportion of ingoing movements. *Built with Transmissio https://www.gaelbn.com/transmissio*

by Google to rank web pages in their search engine (Page et al., 1999). PageRank centrality can capture useful information relevant to diffusion processes, such as epidemics, in networks (Bucur and Holme, 2019; Kandhway and Kuri, 2017). Data manipulation and analysis are conducted in R (R Core Team, 2019); the 'igraph' package (Csardi and Nepusz, 2006) is used for the network analysis.

I use these measures to rank the premises in each 4-week period for the single-species and multi-species networks respectively. The premises which shows the highest value (i.e. ranked first) is removed, and the measure is computed again. I focus on the top 100 premises in each network, and refer to these as the *risky premises*. These premises could be targeted for control strategies in the first stages of an epidemic.

I compare the set of risky premises from the multi-species network, with the set of risky premises in the cattle or sheep network by looking at the intersection. The size of the intersection in the set of risky premises between single-species and multi-species networks serves as a measure of how wrong one would be if considering only one species or the other, instead of the combination of both in

Table 2.2: Network analysis terminology.

Measure	**Definition**
Average Path Length	Average length of the shortest path between all pairs of nodes of the network (Watts and Strogatz, 1998)
Betweenness	Frequency with which a node is in the shortest path between pairs of nodes (Freeman, 1978)
Clustering Coefficient	Number of triplets of nodes all connected to each other (closed triplets) over the total number of triplets in the network (Watts and Strogatz, 1998)
Component	Subset of nodes of the network for which a path exists between any pair of nodes (Newman, 2010)
Giant Weakly Connected Component (GWCC)	Largest component of a directed network, when the directionality of edges is ignored (Newman, 2010)
Giant Strongly Connected Component (GSCC)	Largest subset of nodes for which a directed path exists between all pairs of them
Degree	Number of links a node has (Freeman, 1978)
Diameter	Length of the shortest path between the two most distant nodes of the network (Wasserman et al., 1994a)
Edge density	Proportion of links between nodes that actually exists in the network, calculated as the number of links, divided by the possible number of links (Wasserman et al., 1994a)
PageRank	A variant of Eigenvector Centrality, primarily used for directed networks: measure of a node's importance while giving consideration to the importance of its neighbors in a directed network (Newman, 2010)

the context of an outbreak where both species would be involved in the epidemiology.

2.2.4 Dynamic network analysis

Livestock movements for trade are occasional and not necessarily recurrent over time. Animal movements occur and are recorded on a daily basis, giving the network a temporal dimension. Thus, it is a system where network dynamics are both likely to be important and are well recorded. In the dynamic network, links are considered as an origin, a destination, and a date of occurrence. Two nodes are in contact if there exists a temporally logical path between them (see Fig. 2.3).

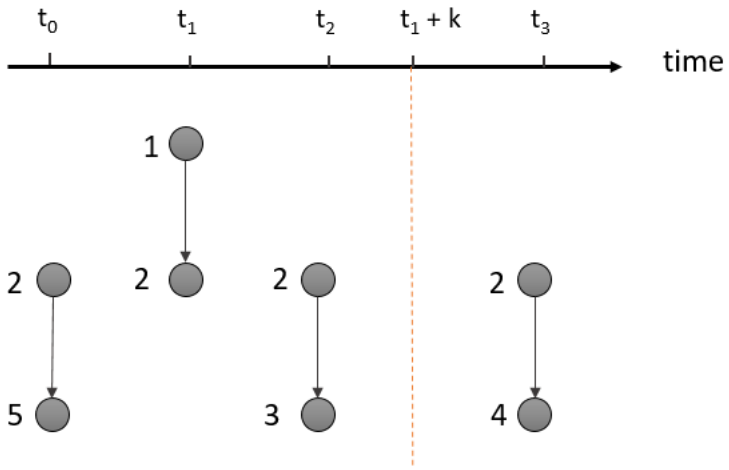

Figure 2.3: In this example there is a temporal path between node 1 and 3, if I consider an infectious period k, there is no path between 1 and 4. There is no temporal path between 1 and 5, since the movement from 2 to 5 occurs prior to the movement from 1 to 2; if $t_3 - t_0 < 28$ days, node 1 and 2 would be connected to 3, 4 and 5 in the static network.

In order to assess the importance of premises in the dynamic network, I calculate temporal Outgoing Contact Chains (OCC) and Ingoing Contact Chains (ICC), which are derived from the reachability, as described by Holme (2005). Contact chains (CC) were used in the context of diseases in livestock systems by Dubé et al. (2008) under the name of infection chain. Here, I use the method previously described by Konschake et al. (2013), where the OCC is defined as the number of premises that can be temporally reached from a primary infected node, considering an infectious period of k days. The ICC is the number of nodes from which a particular node can be temporally reached, accounting for the considered infectious period. I consider an infectious period of seven days, consistent with a fast-spreading FMD-like disease. In other words, the OCC of a premises corresponds to the largest possible epidemic size if the outbreak started in this premises; and the ICC of a premises is proportional to its probability of

being infected if an epidemic starts somewhere in the network. I use a method based on a Breadth-First-Search algorithm to calculate the contact chains for limited periods of four weeks. Starting from a designated node, I traverse the network by exploring all the neighbor nodes at the present depth prior to moving on to the nodes at the next depth level (see Fig. 2.4). I choose to compute the measure for a period of four weeks because: (i) I am interested in the early stage of the epidemic before the outbreak is detected and a movement ban applied; (ii) this makes these results comparable with the results of the static network analysis which have been performed for the same periods.

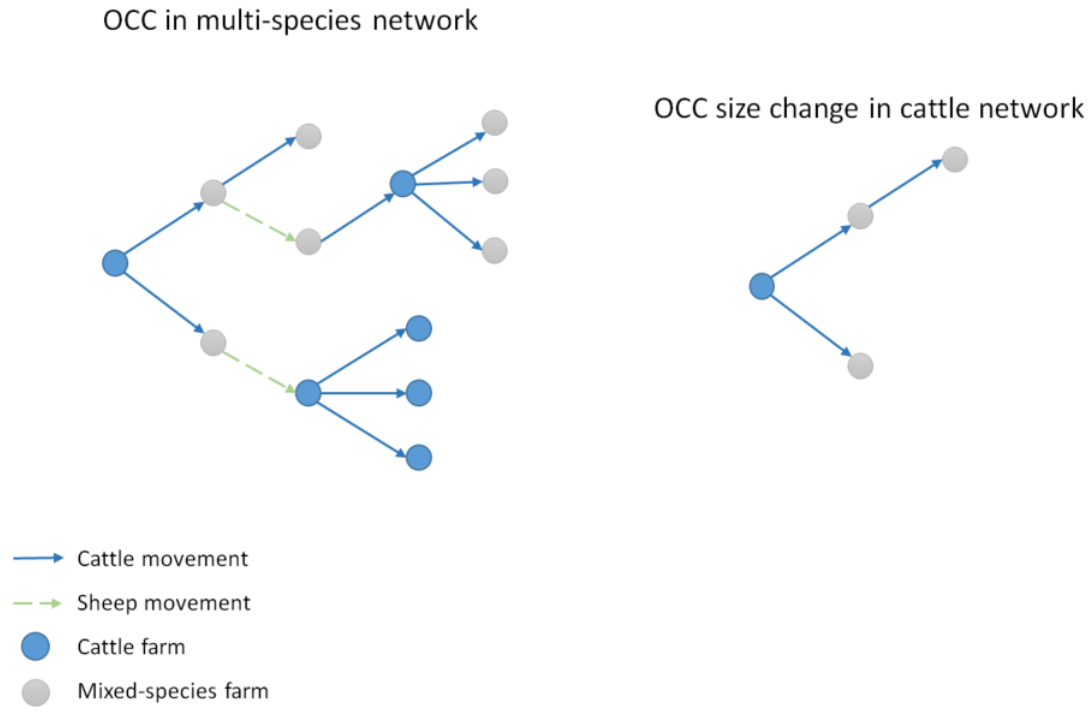

Figure 2.4: Schematic representation of Outgoing Contact Chains in the multi-species and cattle network.

For the sake of simplification, I consider unweighted links in this part. It also avoids making assumptions about the characteristics of the disease. This corresponds to the worst case scenario where the probability of transmission is certain given a link between premises. I compare the sets of risky premises according to the geometric mean of their contact chains ($GM - CC$), defined as $\sqrt{ICC \times OCC}$, in the different networks, i.e. comparing the top hundred risky premises in the single-species network and the multi-species network. This measure has been proven useful to assess the infection potential for fast spreading disease (Rossi et al., 2017). I also look at the changes in the set of risky premises according to geometric mean degree and geometric mean contact chain sizes for the same network, to understand the difference between considering a static or dynamic network.

In order to highlight potential shifts in estimated risk between the multi-species and the cattle systems, I look at the difference in maximal epidemic size between these two systems, by quantifying the change in the OCC of cattle premises taking into account the movements of both species or cattle movements only (see schematic representation in Figure 3.1). The maximal size of an epidemic is a critical parameter, often used in epidemiological studies to quantify the potential impact of an outbreak. Because I compute the OCC for a limited period of 28 days, the OCC is the potential size of the epidemic after 28 days of uncontrolled spread. I calculate for all cattle premises the factor by which their OCC is multiplied in the multi-species network; I create a *multiplication factor* which is:

$$\frac{OCC_M}{OCC_C} \tag{2.2}$$

where OCC_M and OCC_C are the OCC in the multi-species and cattle networks respectively.

2.3 Results

The number of sheep movements was consistently higher than the number of cattle movements (Figure 2.1). The highest volume of trading activity in the Scottish network occurs in late summer to early autumn, when the sheep movement volumes peak. Overall, most of the recorded movements went through markets, accounting for 75% of the trading operations for cattle, and 93% for sheep.

2.3.1 Static network analysis

The sheep network is more dense than the cattle and the multi-species networks, whereas the cattle network is generally more clustered (Table 2.3). As expected, average path length is longest in the cattle network, and shortest in the multi-species network. These shorter paths between pairs of premises enable faster spread of diseases. The sheep network is generally more connected, with typically only a single weakly connected component, while the cattle network is sparser, counting on average 187.1 components during any 4-week snapshot. The multi-species network is less fragmented than the cattle-movement network; this indicates that sheep movements connect different components of mixed or cattle premises in the multi-species network, which were disconnected in the cattle network.

Table 2.3: Static network measures for the 4-week animal movement networks in the year 2016 (mean [min, max]).

	Multi-Species	Cattle	Sheep
Edge Density ($\times 10^{-3}$)	2.3 [1.7, 3.1]	3.2 [2.5, 3.7]	4.9 [2.2, 9.0]
Clustering ($\times 10^{-3}$)	0.88 [0.6, 1.7]	1.5 [0.9, 2.6]	0.13 [0.01,0.55]
Mean Path Length	4.2 [4.1, 4.7]	4.5 [4.2, 5.4]	4.32 [2.8, 4.9]
Diameter	12 [10, 14]	13 [11, 16]	9.8 [6, 13]
No. of Components	138.6 [58, 218]	187.1 [133, 275]	1.8 [1, 4]

Average values are calculated over the thirteen 4-week periods of the year 2016.

The variation in sizes of the components follows the same pattern as the seasonality of the movements (Figure 2.5), i.e. the sizes of the strongly and weakly connected components show two peaks in the cattle network during the 5^{th} and

11^{th} 4-week periods of the year, whereas the sizes of both components sharply increase around the 10^{th} period of the year in the sheep network. Moreover, the membership to components across periods is mostly consistent, with over 60% of the premises constituting the GWCC of the cattle networks being the same in periods 5 and 11, when cattle movements peak. In the multi-species network, over 70% of the premises in the GWCC are remaining the same between these periods. In addition, the size of the GSCC in the multi-species network is always larger than the sum of the GSCC in the cattle and the sheep networks. This highlights how interconnected the two farming systems are. This is important, because the size of the GSCC corresponds to a lower bound on the maximum number of nodes that a newly introduced infectious agent might reach.

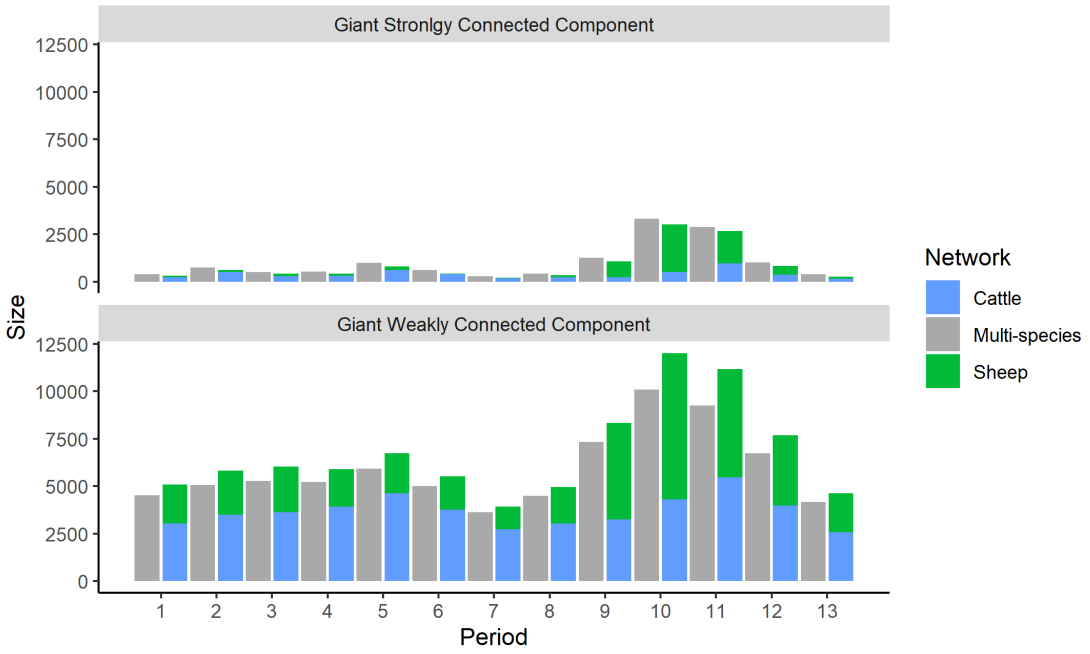

Figure 2.5: This histogram shows the size of the GSCC, and GWCC in the multi-species (in grey) and single-species networks (cattle in blue, and sheep in green) along the year 2016.

Figure 2.6 shows on the one hand that the three measures used are correlated, consistently identifying mostly the same premises as risky; but on the other hand that, for all periods of the year, the most risky premises in the multi-species network are more similar to the ones in the cattle network. The graph also shows, during the 9^{th} and 10^{th} periods, an increase in the number of identical risky premises between the sheep and the multi-species systems, whilst a decrease in this number is observed between the multi-species and cattle network. The majority (more than 96%, for all 4-week periods) of risky premises in the multi-species network are also considered most risky in the cattle network, for all three network measures considered. However, only around 20% of the risky premises in the multi-species network are also identified as most risky in the sheep network, in any 4-week period.

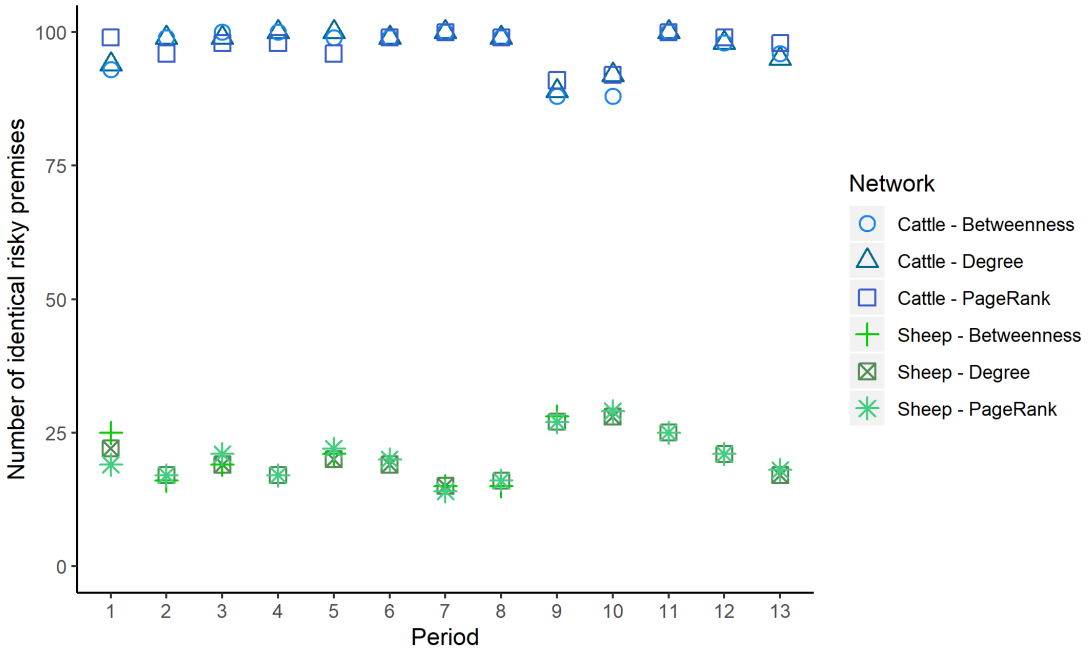

Figure 2.6: This graph details the variation in the number of identical risky premises between the multi-species and each of the single-species networks (cattle in blue, and sheep in green) along the year for the different static network measures.

2.3.2 *Dynamic network analysis*

In the dynamic network analysis, the risky premises are the top 100 premises with the largest $GM - CC$. The set of identified risky premises in the multi-species network are substantially different from the sets in the cattle or the sheep movement networks (Table 2.4). On average only 47.2% of the most risky premises in the multi-species network are most risky in the cattle networks as well, and 32.4% in the sheep networks.

Although for most of the year, the set of risky premises in the multi-species network is more similar to the one in the cattle network (Figure 2.7), during the 9^{th} and 10^{th} periods of the year, the trend reverses with the risky premises in the multi-species network becoming more similar to the ones in the sheep network. In addition, over all periods on average 29% of the risky premises in the multi-species network are not identified as risky in any of the single-species networks. This suggests that some premises with the largest ICC and OCC in the multi-species network exhibit large contact chains only through combination of cattle and sheep movements. The proportion of risky premises of this kind can be as high as 72% during the 11^{th} period.

These results differ considerably from those of the static analysis. I compare the sets of risky premises according to $GM - Deg$ in the static networks, to the risky premises according to $GM - CC$ in the dynamic networks (grey cells in Table 2.4). The percentage similarity between the risky premises according to

Table 2.4: Table showing average results in dynamic network analysis and comparing with static network analysis results

	$GM-CC$	$GM-CC/GM-Deg$
Multi-species	100	16.1 ($\pm$ 2.4)
Cattle	47.2 ($\pm$ 16.1)	16.8 ($\pm$ 3.3)
Sheep	32.4 ($\pm$ 12.3)	17.9 ($\pm$ 9.2)

Number of identical *risky premises* identified in both single- and multi-species networks (white cells), and by dynamic and static measures (grey cells). Values are averages over periods of the year, with standard deviations in brackets. $GM-CC$: geometric mean Contact Chain; $GM-Deg$: geometric mean degree.

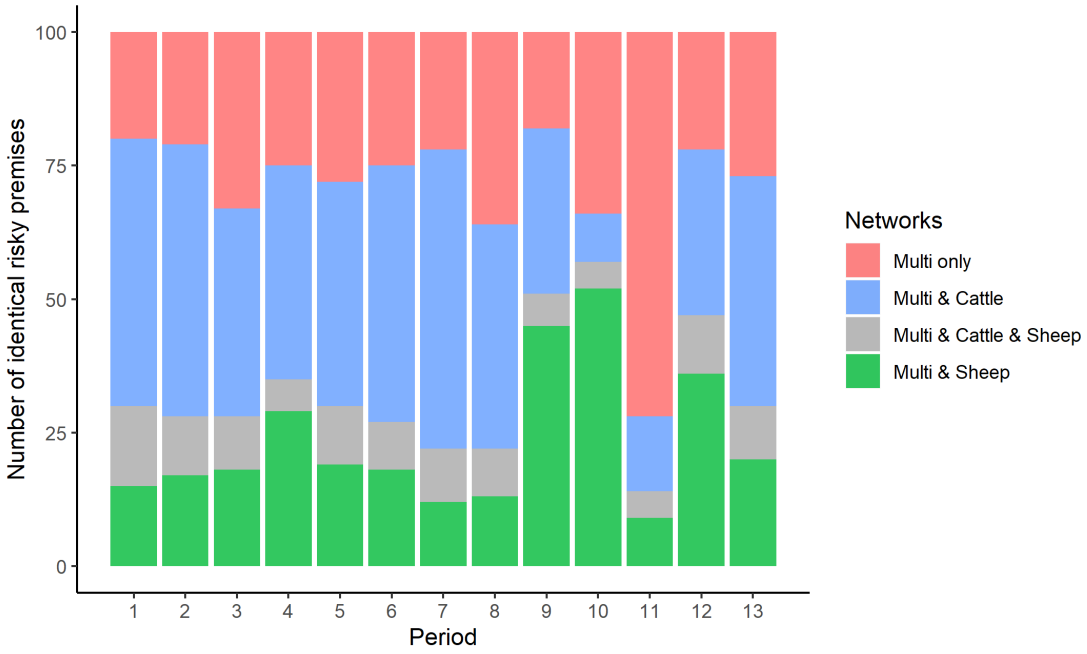

Figure 2.7: Comparison of the single-species and multi-species dynamic network analysis results; Number of identical risky premises between networks after ranking premises according to geometric mean contact chains. In grey is shown the number of premises which are risky in the multi-species, the sheep, and the cattle networks, in blue and in green is shown the number of premises risky in the multi-species and the cattle network but not in the sheep network, and the multi-species and the sheep, but not in the cattle network respectively; the red area represents the premises which are risky in the multi-species network only.

these two measures is low, with on average only around 17% of the premises being the same.

Overall, 99.5% of the cattle premises considered have a larger OCC when including sheep movements, and therefore a multiplication factor greater than 1. Very few cattle premises see their OCC unchanged when sheep movements are considered, even during the period of low activity in the sheep network (Figure 2.8). Sheep movements contribute to the construction of significantly larger OCC in most cases: over all periods of the year, half of the premises (54%) see their OCC multiplied by at least 2. As expected, large increases in OCC is more consistent during the period of high activity in the sheep network, i.e. 10^{th} period of the year. During this period, OCC of cattle premises are multiplied by an average of 8.9.

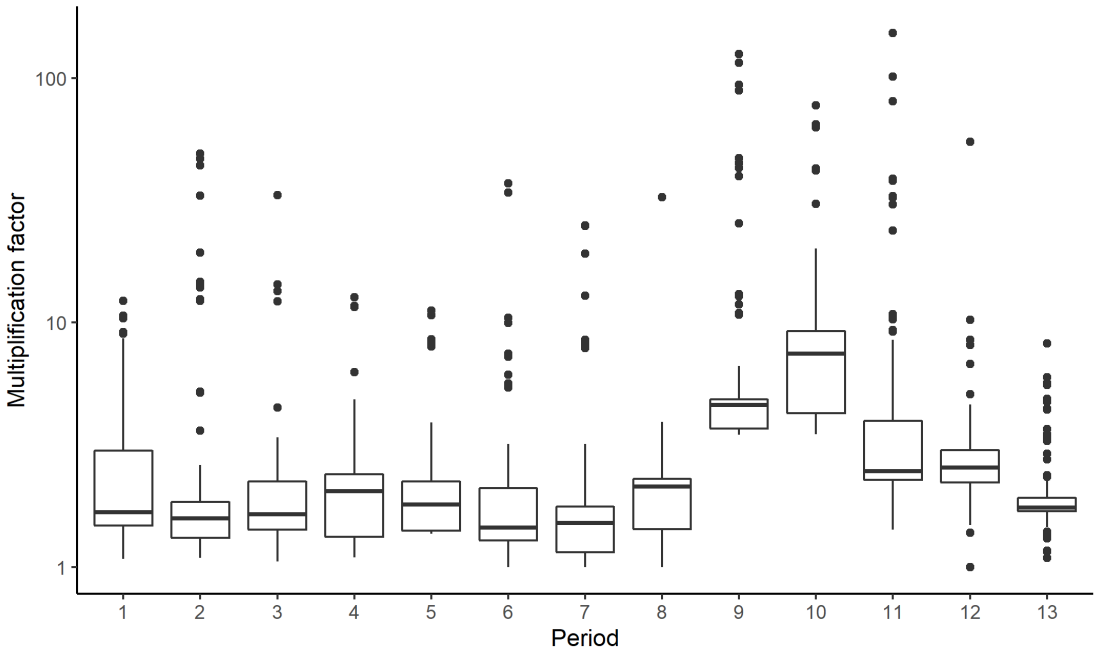

Figure 2.8: Multiplication factor distribution (y-axis log-scale) considering only cattle premises having an OCC of more than 10 premises.

Almost all cattle premises see their OCC increased in the multi-species network (Figure 2.9), but the range of OCC values as well as the range of increase are variable. The OCC values are similar in the 4^{th} and 10^{th} periods are similar for the cattle network, but differ considerably for the multi-species network, exhibiting sharper increases of the OCC during the 10^{th} period. During the 10^{th} period, on average 2513 additional premises can be reached in the multi-species network, which would not be reached through cattle movements only.

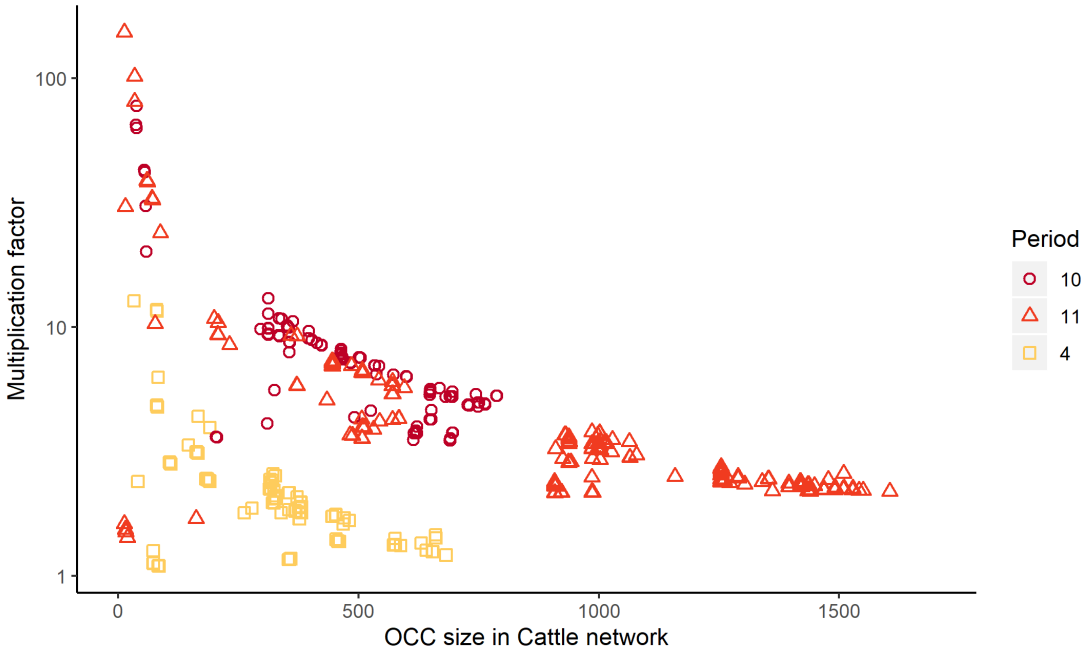

Figure 2.9: Multiplication factor of cattle premises in the multi-species network for three periods of the year; The graph shows the multiplication factor in log scale according to the OCC in the cattle network. Only premises with an OCC of at least ten premises are shown.

2.4 Discussion and conclusions

The objective of this chapter was to explore the Scottish cattle-sheep multi-species network characteristics to determine if this network was substantially different than the single-species ones. The temporal multi-species network exhibited significant differences in its structure, compared to the temporal cattle network. The results showed that more than a half of the risky premises in the multi-species network were not identified as risky in the cattle network. If the cattle network was used to identify risky premises in a context of a disease involving sheep and cattle, these premises would be missed. More importantly, a number of risky premises are identified as such in the multi-species network only: 72% in the 11^{th} period of the year, indicating that their risk is derived from interaction between the two farming systems. These differences indicate that, not only are the risks associated with multi-species epidemics higher, the premises likely to be driving those risks are also different. These differences are not captured by a static network representation of the system, and underlines the importance of temporality in livestock movement networks. These results confirm previous findings in livestock movement analysis and comparison between static and dynamic representation (Lentz et al., 2016; Vidondo and Voelkl, 2018).

In the temporal network, most OCC of cattle premises were significantly larger in the multi-species network than in the cattle network, for all periods of the year. By constructing longer contact chains, interaction between the farming systems

increases risk of larger epidemics throughout the year, and not only during the period of intensive trading in the sheep farming system, as one may expect.

These results ascertain the importance of combining species networks in Scotland, as well as considering layers of temporal livestock movements in detail for the study of disease risk. However, to be able to use such results in decision framework, there is a need for improving our understanding of how network metrics compare to more complex disease simulation models.

3

Proposing a tool to improve preparedness:
A dynamic network metric to identify farm for control measures

The methods and results of this chapter have contributed to the following publication:

***Ruget, AS.**, Rossi, G., Pepler, P.T., Beaunée, G., Banks, CJ., Enright, J., & Kao, RR. Multi-species temporal network of livestock movements for disease spread. Appl Netw Sci 6, 15 (2021).*

Abstract

In Chapter 2, I showed that the analysis of the multi-species network presents substantial differences compared to the analysis of each of single-species networks, which could have important consequences for disease risk.

Here, I further investigate the usefulness of different network measures by comparing their performance with the results of disease simulation models with asymmetric probabilities of transmission between species. In addition to the measures described in the previous chapter (degree, betweenness, PageRank, and contact chain), I propose the use of a measure based on contact chains calculated in a network weighted with transmission probabilities to assess the importance of premises in an outbreak.

I demonstrate that the measure based on contact chains allows us to identify around 30% of the key farms in a simulated epidemic, ignoring markets, whilst static network measures identify less than 10% of these farms.

3.1 Background

As mentioned previously in Chapter 1 and Chapter 2, livestock trade plays a crucial role in the spread of infectious diseases, since infectious animals can transmit a disease over long distance between premises. When an outbreak of a highly contagious disease, such as FMD, occurs, massive trade restrictions are rapidly implemented (Haydon et al., 2004). However, before the first disease case is detected, the disease can spread unrestrictedly via trade. In addition, trade restrictions are not necessarily implemented for endemic diseases such as BVD (Tinsley et al., 2012), in which case network analysis results can be informative to design control strategies.

Early-on there is an important trade-off between complexity of the method used and time pressure. It is therefore useful to have tools readily available to target farms for control strategies after the first cases are detected. Network metrics are useful in this case, since movement data should be immediately available and trade will have driven the spread until the movement ban is put in place. Network tools can help gain insight rapidly whilst more complex models are developed and adjusted. The specific pathogen characteristics and its disease parameters are, in addition, not precisely known in the early phase of the epidemic, making the use of more sophisticated epidemiological model challenging.

There is, however, a need for improving our understanding of how network metrics compare to models that explicitly include dynamics of transmission within nodes. Here, I consider static as well as dynamic network metrics, including the metrics presented in Chapter 2. I also propose a new metric based on contact chains taking into account transmission probabilities. I compare the 100 top ranked farms according to the defined network metrics with the results of a disease simulation model explicitly incorporating the temporal dynamics of the network. I demonstrate that a measure based on contact chains allows us to identify around 30% of the key farms in a simulated epidemic, ignoring markets, whilst static network measures identify less than 10% of these farms. In addition, my results show that even a mildly informed choice of transmission probability per link, without prior knowledge on the disease parameters, gives a better prediction of risk than an unweighted network. Thus, this suggests that once the outbreak pathogen is identified, parameters mildly informed from the literature would be sufficient to provide useful results.

3.2 Material and Methods

3.2.1 Networks and data

The data used and the networks constructed are the same as in Section 2.2.1 and Section 2.2.2. I consider static and dynamic network metrics as described in Section 2.2.3 and Section 2.2.4, in addition to a new metric based on contact chain in a weighted network. The method to calculate this dynamic network metric is detailed in the following section Section 3.2.2.

3.2.2 Weighted In- and Out-going Contact Chain

In Section 2.2.4, I describe the method used to compute In- and Out-going contact chains using a Breadth-First-Search algorithm in an unweighted network. But assuming that all movements are equally important—regardless of the species type, the number of animals, or the characteristics of the premises—neglects important and potentially useful information affecting the spread of a disease. I therefore calculate weighted Outgoing Contact Chains (OCC_w) where the weights are equal to Formula 2.1 and correspond to the probability of transmission given that the node is infected. I consider a network defined as a set of nodes V, and the set of edges $E\ j \xrightarrow{t,w_j} i$ where $i, j \in V$, t is a time, w_j is a weight. I denote the probability of being infected for a node i at time t, $p_I(i,t)$, the complementary probability of not being infected $p_{NI}(i,t)$. The probability of disease transmission for a movement from j to i at time t is consequently equal to $p_I(j,t-1) \times w_j$.

I adapt the algorithm used in the previous section, using a similar method to the one proposed by Enright and Kao (2016). In the initial conditions, all nodes are susceptible, except one root node u. At each discrete time step, I identify all edges $E\ j \xrightarrow{t,w_j} i$, where j has a non null probability of being infected. The probability of not being infected for the nodes i is updated by multiplying it by the probability for the edge $j \xrightarrow{t,w_j} i$ to not transmit infection, which is $1 - p_I(j,t-1) \times w_j$. I keep track of the probability of not having been infected so far, to consider cases of multiple potential infections. I present this algorithm as pseudo code in Algorithm 1.

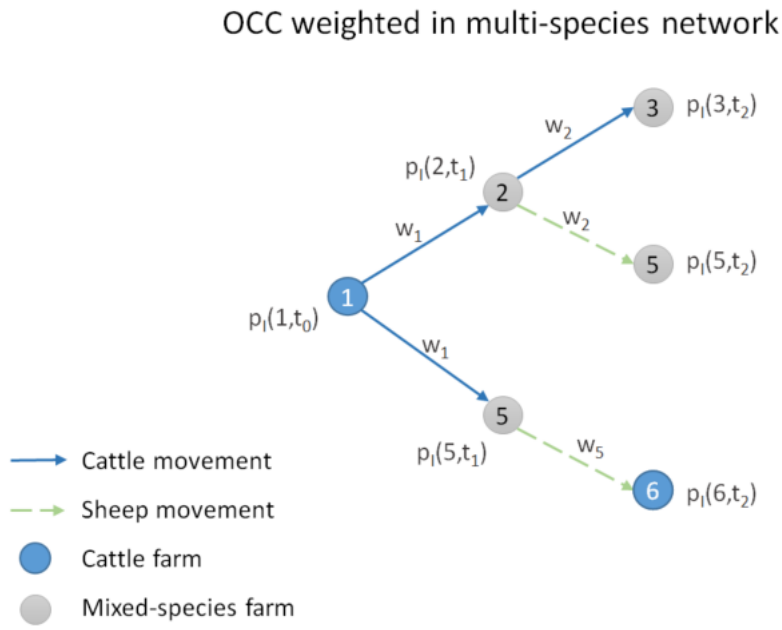

Figure 3.1: Schematic representation of Outgoing Contact Chains in the weighted multi-species network.

Algorithm 1: Algorithm to calculate OCC_w

Result: Average size of the epidemic after 28 days: OCC_w

For all $i \in V$, $p_I(i, t_0) = 0$, except a root node u for which $p_I(u, t_0) = 1$;

for $t \in [t_0 + 1, t_0 + 27]$ **do**

  for $i \in V$ **do**

    for $j \in V$ *such that* $(j \xrightarrow{t,w} i) \in E$ **do**

      $p_{NI}(i, t) = p_{NI}(i, t-1) \times [\prod_j (1 - p_I(j, t-1) \times w_j)]$;

    end

  end

end

$OCC_w = \sum_i (1 - p_{NI}(i, t_0 + 27))$

return OCC_w

Likewise, I calculate the weighted ICC (ICC_w) for all premises in the multi-species network and the different periods of the year. I rank premises according to the geometric mean of weighted contact chains ($GM - CC_w$ defined as $\sqrt{ICC_w \times OCC_w}$), expecting this ranking to be relevant to the prioritisation of control strategies.

3.2.3 Disease simulations

To investigate agreement between the network analysis results and a more realistic situation, I stochastically simulate transmission of a fast-spreading disease in both cattle and sheep. The simulation is based on a Susceptible-Infected-Recovered (SIR) metapopulation model (see Section 3.5, Fig. 3.2), compatible with an immunising infection. The time step is one day, to take into account the daily recorded animal movements, and disease transmission is frequency-dependent. I consider an infection with asymmetric transmission risk, where the rate of effective contacts β has the highest value between cattle, and the lowest from sheep to cattle (Table 3.1). The contact rate between sheep, and from cattle to sheep have intermediate values. The parameter values were chosen arbitrarily within the range of plausibility for a fast-spreading disease like FMD (Keeling, 2005). Parameter values are given in Table 3.1. The schematic representation of the model is detailed below.

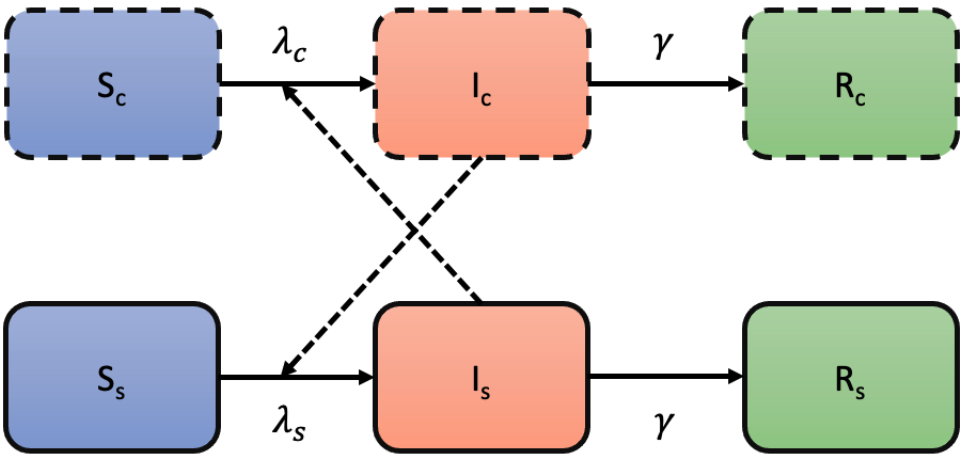

Figure 3.2: Schematic representation of the model with S_c and S_s the susceptible sheep and cattle, I_c and I_s the infected cattle and sheep and R_c and R_c the recovered cattle and sheep; $\lambda_s(t) = \beta_{ss}I_s(t) + \beta_{cs}I_c(t)$ and $\lambda_c(t) = \beta_{cc}I_c(t) + \beta_{sc}I_s(t)$

The equations governing the transition between states of the multi-species model are defined as follows:

$$\begin{aligned}
\frac{dS_s(t)}{dt} &= -\frac{(\beta_{SS}I_s(t) + \beta_{CS}I_c(t))S_s(t)}{N} \\
\frac{dI_s(t)}{dt} &= \frac{(\beta_{SS}I_s(t) + \beta_{CS}I_c(t))S_s(t)}{N} - \gamma I_s(t) \\
\frac{dR_s(t)}{dt} &= \gamma I_s(t) \\
\frac{dS_c(t)}{dt} &= -\frac{(\beta_{CC}I_c(t) + \beta_{SC}I_s(t))S_c(t)}{N} \\
\frac{dI_c(t)}{dt} &= \frac{(\beta_{CC}I_c(t) + \beta_{SC}I_s(t))S_c(t)}{N} - \gamma I_c(t) \\
\frac{dR_c(t)}{dt} &= \gamma I_c(t)
\end{aligned} \tag{3.1}$$

where

$S_s(t)$ and $S_c(t)$ are the number of susceptible sheep and cattle at time t,

$I_s(t)$ and $I_c(t)$ are the number of infected sheep and cattle at time t,

$R_s(t)$ and $R_c(t)$ are the number of recovered sheep and cattle at time t,

β_{SS}, β_{SC}, β_{CS} and β_{CC} are the respective effective contact rates as defined in Table 3.1,

$\frac{1}{\gamma}$ is the length of the infectious period.

Table 3.1: Daily rates for the parameters in the simulation model.

Parameter	Value	Definition
β_{CC}	0.2	effective contact rate between cattle
β_{SC}	0.19	effective contact rate between susceptible sheep and infectious cattle
β_{SS}	0.17	effective contact rate between sheep
β_{CS}	0.15	effective contact rate between susceptible cattle and infectious sheep
γ	0.14	recovery rate

For simplicity, I simulate epidemics starting only in premises having an OCC size greater than 100 premises, that is, premises that can potentially lead to an

epidemic of 100 premises or more. The simulations are run for a limited period of four weeks, starting at the first day of each 4-week period of the year. I use the SimInf package (Widgren et al., 2019) in R to perform 100 simulations per seed, and record the size of the epidemic after four weeks, as well as the number of times a premises is involved in the outbreak over all simulations for each period.

I define an indicator of the *epidemic risk* for each premises and each period as, $ER = N_E \times N_I$, where N_E is the average size of the epidemic at four weeks, and N_I is the number of times the premises is infected during the epidemic and is proportional to the probability of getting infected.

To evaluate the performance of network measures in identifying the most important farms, I compare the 100 premises with the highest ER according to the simulations with the 100 most risky premises according to the different measures ($GM-Deg$, $betweenness$, $PageRank$, $GM-CC$, and $GM-CC_w$). For this comparison, I consider only farms in the ranking, because markets and shows are already known to be high risk and would be targeted first for control measures.

3.3 Results

The static network measures identify at most 11% of the top 100 farms involved in the simulation, and their performance is very poor in most time periods (Figure 3.3). Dynamic network measures offer a clear improvement. The static measures identify on average only 7% of the 100 farms potentially most important in the epidemic, whereas the dynamic measures identify on average over 30% of the main actors of an outbreak (after markets and shows).

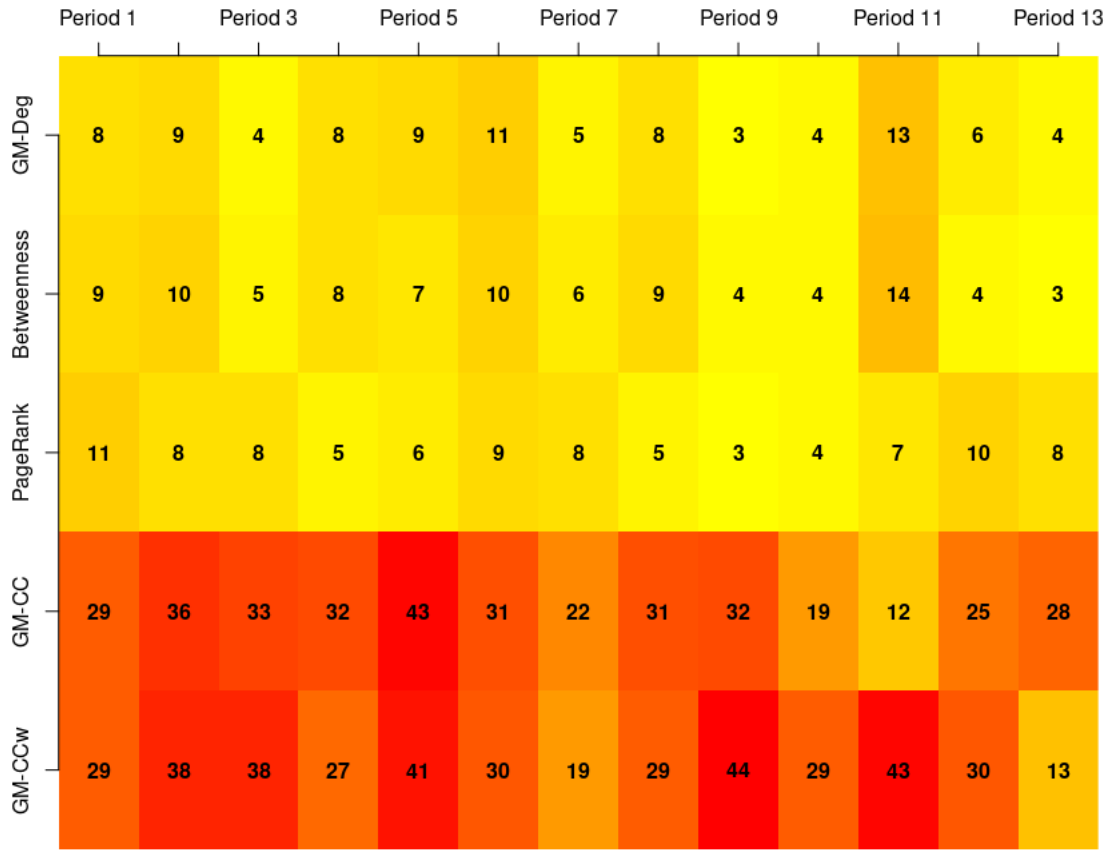

Figure 3.3: Matrix comparing the network analyses and simulations results per 4-week period; Percentage of most risky farms according to the simulations, correctly identified as such by the different network measures. $GM-Deg$: geometric mean degree; $GM-CC$: geometric mean unweighted contact chain; $GM-CC_w$: geometric mean weighted contact chain.

3.4 Discussion and conclusions

In this chapter, my aim was to assess the performance of network metrics to use as a tool to inform outbreak response in the early stage of an epidemic. I found that when disease spread was simulated in the multi-species system, the temporal measures performed better at identifying the most important farms than the static network measures. The inability of static descriptors to reliably predict outbreak risk is expected (Vidondo and Voelkl, 2018), when the pattern of contacts changes over timescales that are short compared to disease generation times. The measures based on contact chains take into account the more important aspects of the movement networks, such as temporal paths, which are relevant in the occurrence of an epidemic (Holme, 2005; Lentz et al., 2016). Usefulness of a measure based on ingoing and outgoing contact chains for assessing disease risk was already confirmed by other studies (Frössling et al., 2012; Vidondo and Voelkl, 2018). I used similar metrics, but took into account transmission probabilities. The weighted contact chains performed better than the simpler contact chains. Here I used weightings in line with the 2001 FMD epidemic in the UK, which substantially improved the predictive power of the metric. This shows that even a mildly informed choice of transmission probability per link, without prior knowledge on the disease parameters, gives a better prediction of risk than an unweighted network.

My work reinforces the importance of incorporating differences in transmission probability where they exist within a system (e.g. differences between host-species in a multi-species system (Dobson, 2004)). Significant variation can lead to critically different disease dynamics (Lloyd-Smith et al., 2005) which must be captured for predictive modelling. In the weighted network, the weights incorporate variation due to both the volume and species traded, which allows the network metric proposed to explicitly include this information. Similar methods based on contact chains have included more explicit simulation rather than metric calculation: e.g. Knific et al. (2020) reports work which filters a weighted temporal network and then simulates scenarios with differing transmission probabilities, thus somewhat decoupling the disease model from the underlying network. This metric differs in that it incorporates transmission probabilities directly into network and metric, and is thus most useful when consider a particular pathogen with known transmission probabilities that vary by category of edge.

While I demonstrated the importance of using a multi-species network to understand transmission of an FMD-like disease, additional work remains to be done. Cattle and sheep are not the only species vulnerable to FMD: in a future FMD epidemic, network layers including other species (e.g. pigs) could be included, as well as non-trade layers that could incorporate transmission risk due to shared equipment or human movements.

Because characteristics of the disease are included in the weighted temporal network construction, a number of adaptations would be needed to apply my work to other diseases. In particular, the time of aggregation has important consequences for the network's interaction with the pathogen (Bajardi et al., 2011a). The time scale in the approach should therefore be adapted to correspond to both the infectious period of a disease and the time scale of network dynamics (Kao et al., 2007).

While these methods have been implemented with reasonably efficient code, the focus has been on assessing the usefulness of the approaches as opposed to producing code optimised for speed and memory requirement. Computational performance could likely be improved, and further work may be required to deploy these approaches on very large or very dense networks.

Weighted contact chains can be a powerful tool to inform decisions in the early stages of an epidemic because it only relies on animal movement data that are immediately available. As well as being easy and fast to compute, it is deterministic, which means the metric can be calculated in a single computational run. In addition, the method proposed showed that weighting the network with reasonable transmission probabilities helps to improve the prediction of risk, which could aid decision making in the early stages of an epidemic when disease parameters are still unknown.

3.5 Appendix

As a member of EPIC (the Centre of Expertise on Animal Disease Outbreaks), my aim is to deliver scientific evidence to support the Scottish Government in the prevention of and preparation for important animal disease outbreaks. In this context, the findings of Chapter 2 and Chapter 3 lead to a research brief which will be communicated with the Animal Health and Welfare Division of the Scottish Government. See draft next page.

RESEARCH BRIEF: Links between Scottish cattle and sheep premises affect disease risk

Author: Anne-Sophie Ruget, Gianluigi Rossi, P. Theo Pepler, Gael Beaunée, Christopher J. Banks, Jessica Enright and Rowland R. Kao

Reviewed by: May Fujiwara, Amy Jennings, Lisa Boden

Contact details (institution and email address): Roslin Institute, a.ruget@sms.ed.ac.uk

Date: 03/09/2021

1. KEY MESSAGE

- Using a cattle-sheep movement temporal network (i.e. one that takes into account changes in the network over time) to target premises for control strategies, we found that the set of most important premises for disease spread in the combined cattle-sheep network differs substantially from the sets identified in the separate networks. Contact links between the sheep and cattle farming systems should not be ignored when targeting premises for outbreak control.
- For the early stage of an outbreak affecting multiple species, we propose a measure based on animal movement contact chains to rapidly identify key farms for disease surveillance and control.

2. MAJOR FINDINGS

- We found that more than a half of the risky premises (i.e. top 100 ranked premises according to network metrics) in the multi-species network were not identified as risky in the cattle network. Up to 70% of the risky premises were identified as such in the multi-species network only, indicating that their risk is derived from interaction between the two farming systems. These risky premises would be missed if the cattle network was used to identify risky premises in a context of a disease where both sheep and cattle play substantial roles.
- We demonstrated that a measure based on “contact chains” (sets of premises that are epidemiologically linked through animal movements) in the temporal network where links are weighted by a probability of transmission, allows us to identify around 30% of the risky farms in a simulated epidemic affecting cattle and sheep, ignoring markets. This is true even when specific disease parameters are only approximately known. In contrast, measures for a static network (where connections between the farms are fixed) identify less than 10% of these farms.

3. OBJECTIVES

- To show that connections between the cattle and sheep movement networks in Scotland changes disease risk and therefore the importance of considering them jointly,
- To propose a tool to rapidly identify farms to target for control measures in the early-stage of an outbreak.

4. POLICY IMPLICATIONS

- Weighted contact chains (that is, sets of premises that are epidemiologically linked through animal movements) can be a powerful tool to inform decisions in the early stage of an epidemic. This measure shows the potential for an epidemic to spread, even when no disease model is available.

- In an infectious disease outbreak affecting multiple species, disease surveillance and control strategies based on single species contact networks may be misleading. Links between the different species contact networks should be considered when identifying the riskiest premises for disease spread.

5. IMPORTANT ASSUMPTIONS AND LIMITATIONS

We considered directed cattle and sheep movement networks in successive 4-week windows, using observed animal movement records from 2016.

We calculated "contact chains" (sets of premises that are epidemiologically linked through animal movements) for various single species, multiple species, static and dynamic representations of the animal movement network.

Assumption 1: Assumed the infectious disease is spread only by animal-to-animal contact and animal movements between premises. Results may differ for diseases spread via indirect routes (e.g. to neighbouring premises, or via environmental contamination or wildlife reservoir).

Assumption 2: Assumed probability of infection broadly compatible with FMD spread, but deliberately not tuned to the specific 2001 FMD epidemic values, in order to test the value of imperfect knowledge.

Assumption 3: Used a 4-week period for calculating contact chains (sets of linked premises). This assumes a maximum time of 4 weeks for undetected spread of the disease before movement controls are introduced. Results may differ for slower moving, less transmissible diseases, potentially requiring a longer time window than 4 weeks.

6. FIGURE OR TABLE (optional; restrict to one if essential)

7. LINKS TO EXISTING PUBLICATIONS OR REPORTS

- Ruget, A. S., Rossi, G., Pepler, P. T., Beaunée, G., Banks, C. J., Enright, J., & Kao, R. R. (2021). Multi-species temporal network of livestock movements for disease spread. *Applied Network Science*, *6*(1), 1-20.

8. POLICY COMMENTS/RESPONSE (Has this scientific research had a beneficial impact on policy. Please give examples)

Part II

During the outbreak: importation and transmission phases

4

Importation phase, estimation of introduction risk:
Risk of COVID-19 introduction into the Scottish Hebrides and strategies for control

Abstract

In this chapter, I propose to estimate the risk of introduction of COVID-19 to each of the Hebridean Islands situated off the west coast of mainland Scotland according to movement levels. This is of interest because the natural boundaries of islands can allow for better estimation or control of movements between populations.

I use a combination of real human mobility data and census data to generate seasonally varying patterns of human movements amongst the Hebrides and from elsewhere. I consider three distinct periods: each of summer and winter 2019, illustrating a year prior to the pandemic, and summer 2020 illustrating a *pandemic summer*. Movements during these periods serve as input to simulate COVID-19 transmission from the mainland to the archipelago in a stochastic meta-population model allowing us to explore the impact of seasonal variations on the risk of introduction and the effectiveness of non-pharmaceutical interventions.

Despite strong seasonality in movement patterns, partly driven by tourism, I show that in some islands the introduction risk is high even in the low season, when activity and movements from the mainland are expected to be low. This will be of particular concern if COVID-19 becomes a seasonal respiratory infection affecting temperate areas in winter concomitantly with other seasonal

infections such as flu. In the high season, although in most cases movement control will not significantly delay a potential introduction, for some islands a 70% reduction of movements in peak summer tourist season has the potential for delaying the introduction risk for over 6 weeks, i.e. beyond the high risk summer holiday period.

4.1 Background

Numerous viruses causing respiratory diseases that have been responsible for large epidemics in the past were identified as a pandemic threat (MERS, SARS, Influenza... Reperant and Osterhaus (2017)). In December 2019, the first cases of pneumonia due to the new virus SARS-Cov-2, now called COVID-19, were identified. The virus rapidly spread to Europe including Scotland, where the first death was recorded on March 16th 2020 (National Records of Scotland, 2020a). Following the introduction and spread of the virus, the first lockdown was declared on March 23rd 2020. Since then the virus has continuously spread within the UK, at various levels. The first lockdown proved effective at reducing the circulation to a very low level (National Records of Scotland, 2020b). The rise in cases following the lifting of the restrictions in July 2020 was mostly due to importation (Lycett et al., 2021), showing that within the course of the epidemic, re-introduction of viruses between communities may play an important role and controlling these re-introductions may be key to reducing the spread and therefore the impact.

The geographical context of islands and their natural boundaries might have an impact on how the disease spreads compared to the mainland. In addition, in the context of a respiratory disease spread by movement of people, the natural boundaries of islands make measurement of the flow of people entering the islands, or circulating between the islands easier. It also potentially allows for a better control of flow between populations. With regards to COVID-19, New Zealand illustrated an example of how to take advantage of the geographic isolation to locally eliminate a pathogen (Cousins, 2020).

Modelling approaches can help us better understand the importation risk taking into account movement pattern and disease process. Such methods have been used in various contexts such as the Ebola epidemic in 2014 (Poletto et al., 2014b), and for influenza epidemics (Mateus et al., 2014), and have shown that movement restrictions can only delay the introduction but not prevent it, unless movements are suppressed. Whilst facing the spread of a new virus, every week allows for a better understanding of the disease dynamic, for potentially better treatments and interventions, or even a larger vaccine coverage of the population. In such a situation it might be useful to delay the introduction, even for a short period of time, as the breadth of knowledge and means are rapidly growing.

In Scotland, most of the cases due to COVID-19 have occurred in the most densely populated part of the country. The Scottish isles have been less affected by COVID-19, although multiple introductions of the virus have occurred. Movements between Scotland mainland and the isles experience high seasonal variations due to tourism and seasonal workers. My objective was to assess the risk of introducing a respiratory disease such as COVID-19 in the Hebrides, an ar-

chipelago off the west coast of mainland Scotland according to flow of people. I estimate the short-term risk of introduction in summer vs. winter prior to the pandemic situation using data from 2019, and highlighted the change of risk between seasons. I then consider the summer 2020 illustrating a peak season period within the pandemic and estimate the risk of introduction as well as the impact of movement restrictions on this risk. Finally I explore control scenarios based on movements restrictions and Non-Pharmaceutical Interventions (NPIs) inducing a mitigation of contacts between individuals and therefore disease transmission.

4.2 Materials and Methods

4.2.1 Data

To be able to reproduce the pattern of movement relevant to disease spread from the mainland to the islands, I combine publicly available data with data provided by transport companies. The Census Flow Data is an official statistics and corresponds to the number of people moving between declared home locations and workplace locations. This data is available online for academic institutions (UK data service, 2011). The Civil Aviation Authority (CAA) publishes monthly reports providing number of passengers between airport (Civil Aviation Authority, 2020). Skye is the single island in the study having a road link to the mainland. Transport Scotland shared with us statistics on direct vehicle movements into the Islands system via Skye bridge traffic, providing a number of cars in each direction per hour between January 2019 and December 2020. Finally the ferry company (Caledonian MacBrayne, or "Calmac") operating between 19 islands of the Hebrides provided number of passengers per ferry on scheduled routes between January 2019 and October 2020.

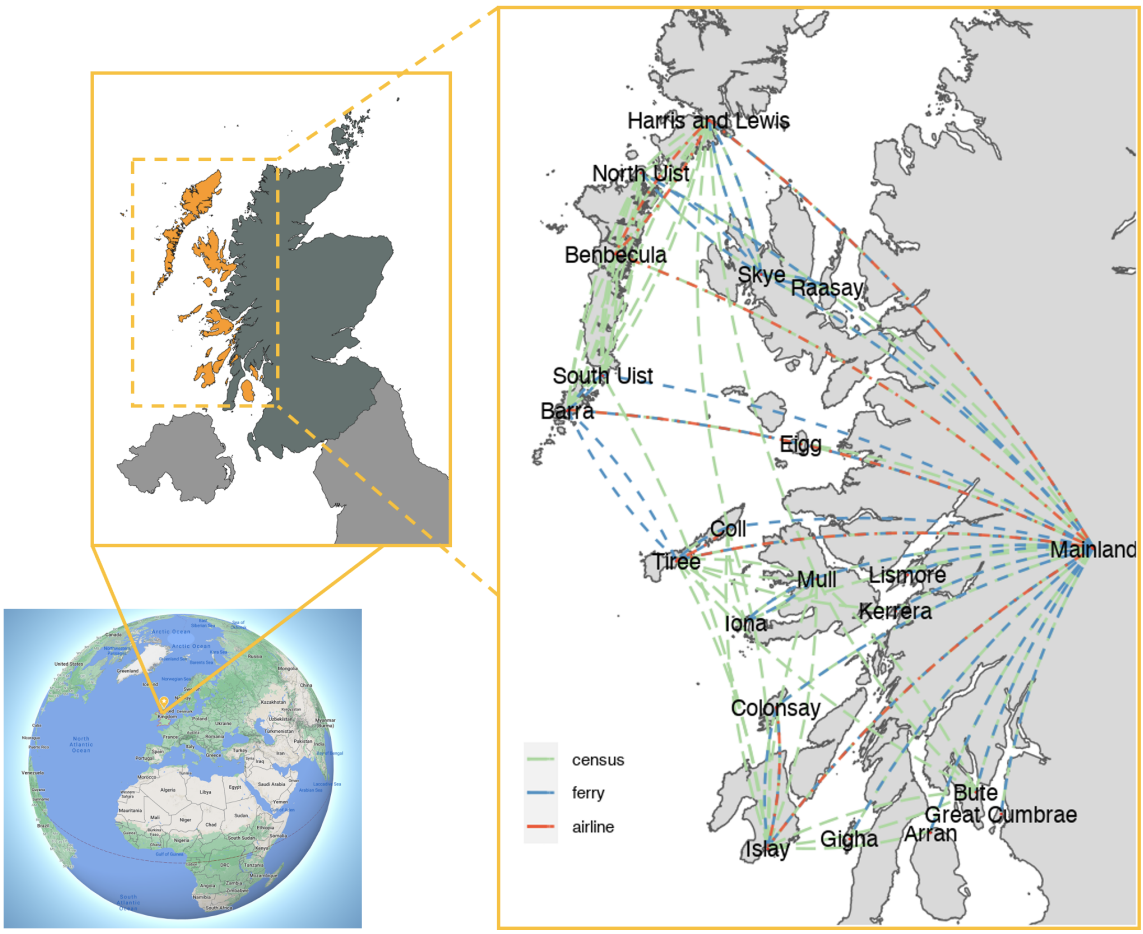

Figure 4.1: Maps showing the location of the Hebrides on earth (bottom left) and in Scotland (top left) and map of the islands displaying the superimposed networks (right hand-side).

I use these data to construct a combined network at the output area (OA) level. The output area is an administrative unit containing approximately 50 inhabitants. Fig. 4.1 shows the location of Scotland and the Hebrides, as well as an aggregation of the three networks at the island level superimposed with a map.

Since I consider a combination of data collected in real time and data from the 2011 census, I expect some movements to appear in more than one of the datasets. To avoid overestimating movement volumes, I calculate the edge weights, considering only the largest volume between the census data or the ferry and bridge data. I assume that there is no overlap between airline data and census data, as individuals are unlikely to commute regularly by air.

Large seasonal variation in volumes were clear in 2019, as well as large changes in volume due to restrictions in 2020 (see Fig. 4.2). When the combined network was constructed, three sets of weights for the edges compatible with three periods were considered: summer 2019, winter 2019, and summer 2020. These three periods reflect different situations: summer and winter 2019 illustrate the seasonal variation in a typical year, whereas summer 2020 represents a holiday period during the pandemic with no movement restrictions in Scotland (ISSN International Centre, 2020).

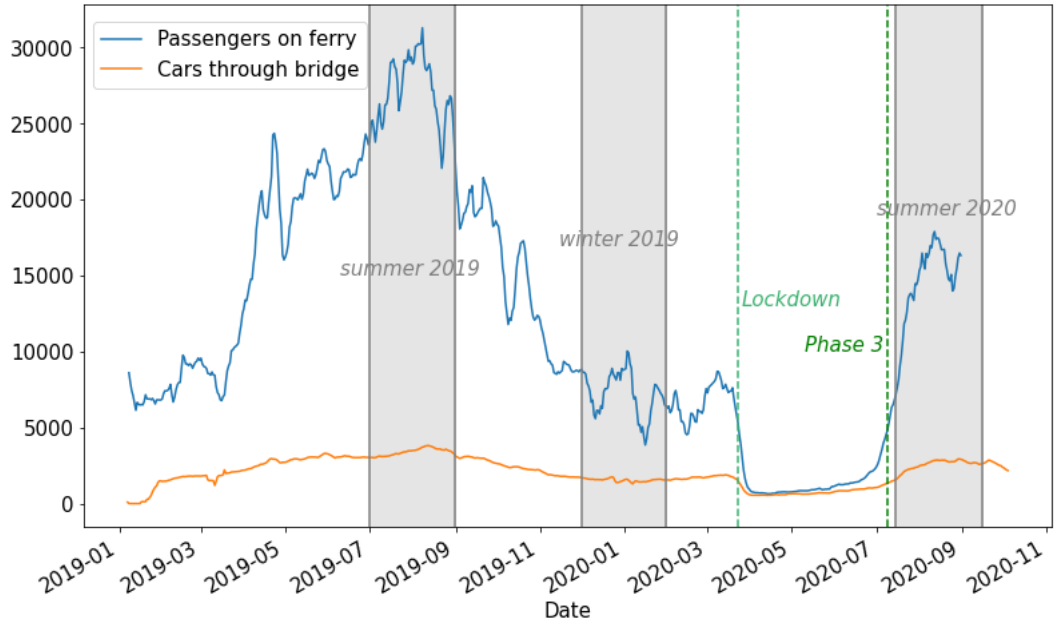

Figure 4.2: Variation of passenger and car volume over time. The green dashed lines show the date at which the lockdown was instigated and the date of phase 3 of lifting restrictions, when movements were allowed

While there may be some overlap, each dataset contains differing information: the ferry data contains a number of passengers from port to port per date and time, the airline data contains a monthly volume between pairs of airports, the Skye bridge data provides a number of cars per hour of the day from January

2019 to December 2020. Whilst this data represents measurement of real movements recorded during the period of interest, the Census Flow Data is an official statistics from 2011 provided at the OA scale. The CFD provides the network structure at the right scale for the model, whilst the other datasets are useful to inform the movement volumes in the network. To homogenise the data and construct between-OA movements of individuals, the data is processed as described below:

- **Ferry**: port to port movement volumes from the ferry data are used to adjust the between OA movements from CFD. As volumes were generally higher in the ferry data than the census, excess was randomly distributed between existing edges in the CFD, or uniformly distributed between OAs on each island if no connection existed in the CFD.
- **Skye bridge**: I assume that there is one passenger per vehicle in winter. I further assume that the increase in passenger flow over the bridge between seasons is proportional to the increase of passengers on ferry routes (on average $1.9\times$ the volume). This is consistent with Transport Scotland statistics where the average occupancy of car in 2018 was found to be 1.5 (Transport Scotland, 2011); whilst the assumed winter and summer passenger volumes is averaging 1.45 passengers per vehicle. The increased volume was then distributed between existing OA-OA routes in the census from the mainland to Skye with a probability proportional to the commute volume.
- **Airline**: no overlap with other sources of data was considered, movements from airport to airport were distributed between OAs using the same approach as for the ferry data.

4.2.2 Model

I use a stochastic network meta-population model, where each node represents an output area. Within each node the population is split into three age groups ([0-17), [17-70), and 70+) with populations derived from census data (Scotland's Census, 2011). I model the disease state of each node's population with a stochastic compartmental model including a latent (E) state, asymptomatic (A_2), presymptomatic (A) and symptomatic (I) infectious states, and subsequent hospitalised(H), recovered (R) or dead (D) states (schematic representation in Fig. 5.3). The parameter values and the model description are detailed in appendix.

To drive within-node infectious dynamics, I derive contact rates between individuals from age-structured mixing matrices based on survey data. For contacts before the pandemic, I use POLYMOD matrices (Mossong et al., 2017),

gathered before the pandemic, and for contacts during the pandemic, I use CoMix matrices (Jarvis et al., 2020a) gathered during the pandemic. The CoMix matrix used gave a mean of 6 contacts per individual per day overall, consistent with other estimates of the highest mean number of contacts in the UK during the pandemic (Jarvis et al., 2020b).

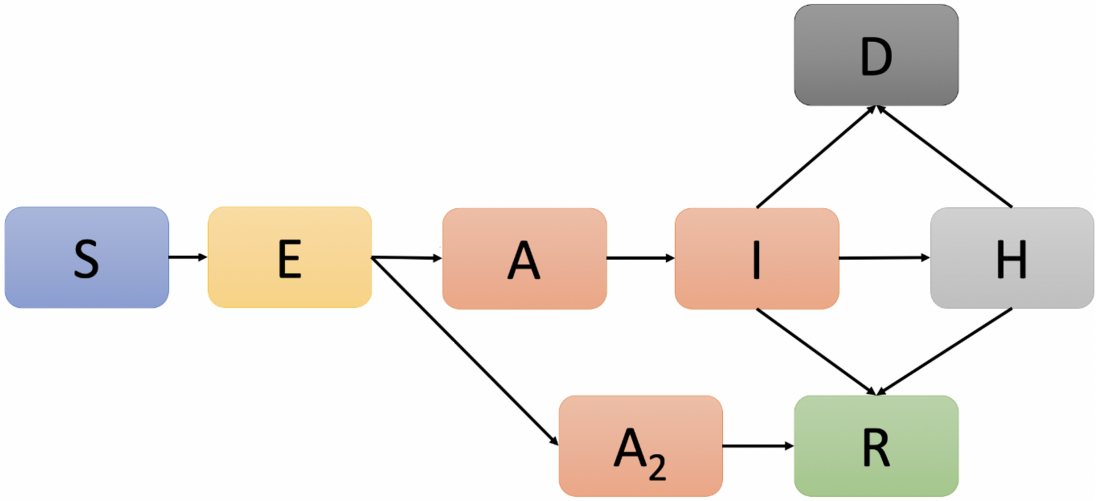

Figure 4.3: Schematic representation of the compartmental model. S - susceptible, E - exposed (but not infectious), A - presymptomatic infectious before symptom onset, I - symptomatic infectious, A_2 - asymptomatic infectious, H - hospitalised, R - recovered, D - dead.

4.2.3 *Variation in the risk of introduction*

I define the *risk of introduction* before time t as the probability that at least one individual had entered one of the disease states (E, A, A_2, I) before or at time t. This probability is calculated as the proportion of simulations when the introduction occurred over the total number of simulations. As in Fig. 4.2, volumes of movements exhibit high variations depending on circumstances. The variation in risk of introduction between a typical summer and winter using the data from 2019 is therefore first explored. These periods being prior to the COVID-19 pandemic, the average number of contacts between individuals is defined by the POLYMOD age mixing matrix. A fixed prevalence is assumed on the mainland of 1%, which is approximately the prevalence estimated in Scotland by the ONS survey at the end of 2020 (ONS, 2020). I run 200 simulations considering the winter edge weights and the summer edge weights respectively. As the focus is on short-term forecast here, the risk of introduction of the virus is calculated per island at 30 days after the start of the simulation, in summer and in winter.

I investigate statistical correlation of the introduction risk with network metrics, such as in-flow, closeness, betweenness, and length of the shortest path to mainland. The in-flow is the sum of the weights of incoming links (Wasserman

et al., 1994b). The closeness is the average of the shortest path length from the node to every other node in the network (Freeman, 1978). The betweenness is the frequency with which a node is in the shortest path between pairs of nodes (Freeman, 1978). The shortest path between an island and the mainland in the weighted network is the path which minimises the sum of the inverse of the weights (here, the weights represent closeness, consequently the inverse of the weights are interpreted as a distance). The shortest path length is the sum of the inverse of the link weights which form the shortest path. I also consider other indicators such as population size, and the health and access domains of the Scottish Multiple Deprivation Index (SIMD, Scottish Government (2020a)). SIMD is a relative measure of deprivation across datazones (DZs, which are areas containing approximately 500-1000 residents) in Scotland. SIMD looks at the extent to which an area is deprived across seven domains including health and access. The higher the score, the more deprived the area. The access domain is representative of the connectedness of the area taking into account drive or public transport travel time to facilities such as school, GP... The health domain measures the healthiness of the population. It has been shown to have a potential influence on COVID-19 mortality (Banks et al., 2020).

4.2.4 Mitigation of the introduction risk

4.2.4.1 *Relative importance of movement types*

I use data from the summer 2020 to illustrate the movement pattern of a *pandemic summer*, where no travel restrictions were in place at the national level (ISSN International Centre, 2020). Despite the absence of restrictions the volumes of passengers between the islands was twice as low as the previous year, according to the ferry data. This can be explained by the implementation of quarantine for international travel in Scotland, which can discourage international movements, in addition to a change in behaviour (Brinkman et al., 2020).

The risk of introduction before time t is defined as the probability that at least one individual has entered one of the disease states (E, A, A_2, I) before time t or at time t. To assess the importance of connections with neighbouring islands, I compare the conditional probability that an island k is infected given that one of its neighbours in the network is infected, with the probability the island k is infected.

The following notations are used to define the events of interest:

- event $I_{k,t}$: "introduction of COVID-19 on island k before or at time t"
- event $N_{k,t}$: "introduction of COVID-19 on one of the neighbours of island k before or at time t"

The probability of introduction of COVID-19 on island k before time t, $P(I_{k,t})$, is calculated as the ratio between the number of simulations with introduction in k before time t and the number of simulations.

The probability of introduction of COVID-19 on island k before time t given that COVID-19 has been introduced on at least one of the neighbouring islands $P(I_{k,t}|N_{k,t})$ is calculated as the number of simulations with introduction in k and any neighbour of k before time t divided by the number of simulations with introduction in any neighbour of k before time t, $\frac{P(I_{k,t} \cap N_{k,t})}{P(N_{k,t})}$.

I run simulations considering the summer-2020 level of movements and a mean of 6 contacts per person and day, and calculate the values of these two probabilities over time. To highlight any dependence between the probability for an island to get infected, and the probability for one of its neighbours to get infected, I compare $P(I_k)$ and $P(I_k|N_k)$. Finally, I consider situations in which restrictions would induce a reduction of 50% of the volume of movements from the mainland only, or a 50% reduction of movements from the mainland and between the islands. In the latter, a larger volume of movements is therefore removed. I compare the effect of these two scenarios on the risk of introduction per island over time.

4.2.4.2 *Delay in introduction according to restriction level*

To assess the effect of movement restrictions on the risk of introduction, I compare the spread of COVID-19 obtained from simulations of the model with and without movement reductions. I focus on short-term projections and calculate the probability of case importation per island predicted after 30 days of simulation in the baseline scenario without movement restrictions. I then compute the time delay needed to reach the same value of introduction probability per island in simulations with movement restrictions.

I run two scenarios considering a decrease of the movement volume from the mainland of 50% or 70%. I test correlation between the number of movements removed and the delay observed in virus introduction per island.

4.2.5 Summary of the COVID-19 cases on the islands

I compare the results to PCR test results data from PHS (PHS, 2020), which provides per test performed the date of the test and the home DZ of the person tested. I test the correlation between the number of months per island having at least one positive test and the probability of introduction calculated at 30 days in the "summer 2020" simulations.

4.2.6 Control scenarios exploration

I explore scenarios of control after introduction, with the aim at comparing the effect of controlling movements between OAs, as compared to controlling the number of contacts within OA, as well as the combination of these two measures.

Table 4.1: Table summarising the measures implemented for control in each scenario.

Scenario	Movement	Mean contacts
Scenario 1a	No movement reduction	3
Scenario 1b	Movements from the mainland reduced by 50%	3
Scenario 2	Movements from the mainland reduced by 50%	-
Scenario 3	All movements reduced by 50%	2.6

The baseline is summer 2020, when the average number of contacts per individual per day was 6. In the different scenarios, the simulation starts with the same disease parameters and movement network as the baseline. Control measures are implemented at 20 days. I consider, in scenario 1a, a decrease of the number of contacts from 6 to 3, which corresponds to the decrease observed in England when all public places inducing mixing are closed except for schools (Jarvis et al., 2021). In scenario 1b, I consider movement restrictions, applied to movement from the mainland only alongside contact mitigation. In scenario 2, I explore the effect of controlling the movements from the mainland alone by reducing them by 50%. Finally the scenario 3 simulates a lockdown-like situation, decreasing average number of contacts down to 2.6, similar to observations in March 2020 (Jarvis et al., 2020a, 2021), and all movements reduced by 50%.

To compare the impact of these various measures, I report the distribution of the number of individuals across simulations who has been infected in the whole of the Hebrides.

4.3 Results

4.3.1 Comparison of seasonal risks

I first assess correlations between introduction risk and network metrics in summer and winter, as well as with population size and two domains of the SIMD (Cf Table 4.2).

In-flow as weighted in-degree and shortest path length to mainland as minimum geodesic distance to the mainland were the measures showing the strongest correlation to risk. All other variables were significantly correlated apart from the *betweenness* and *closeness*. The variable *access* showed a negative correlation with the risk of introduction (-0.67 in winter, p-value 0.002, -0.65 in summer p-value 0.003), which is expected, since a higher access score means a longer driving/public transport time to facilities such as GP, schools or post office, i.e. poorer access. The variable *health* showed a positive correlation with the introduction risk in winter (0.54 p-value 0.022), indicating that areas with poorer health could be exposed to higher risk of introduction in winter.

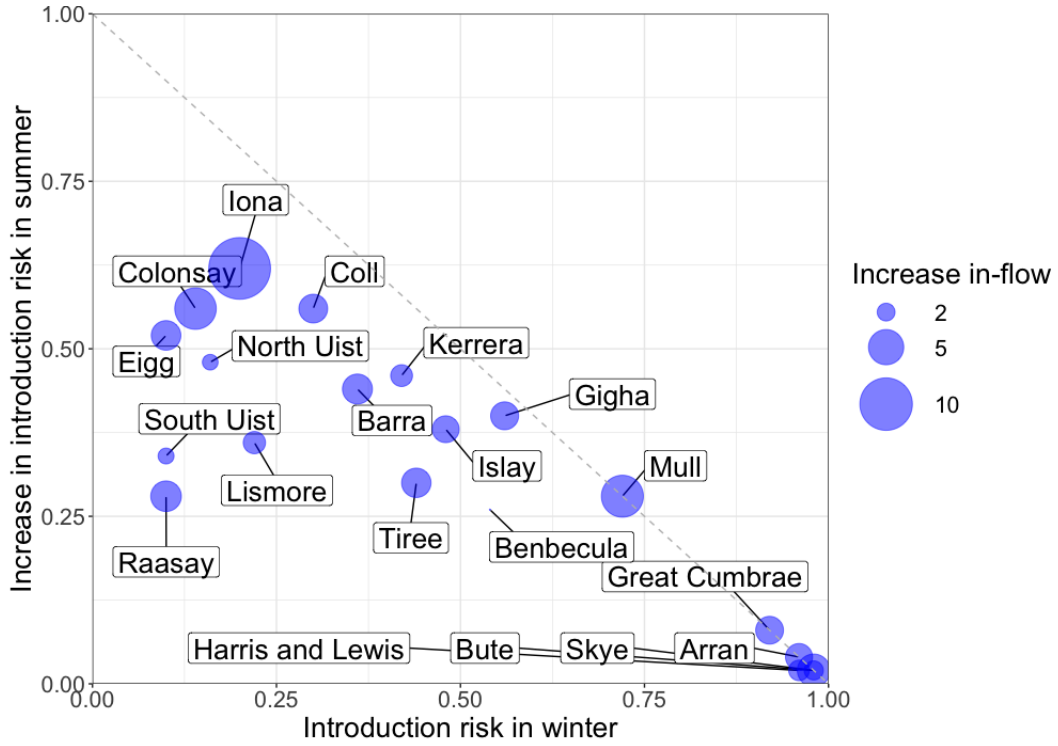

Figure 4.4: Increase in introduction risk in summer by introduction risk in winter. The area of the dot is proportional to the multiplying factor value between summer inflow and winter inflow.

Fig. 4.4 shows the increase in introduction risk in summer compared to winter plotted by introduction risk in winter. The area of the dot is proportional to the ratio of passenger volume between summer and winter. Skye, Harris and Lewis, Arran, Bute and Great Cumbrae showed a risk of introduction before 30 days close to one in summer as well as in winter. Other islands showed a more

Table 4.2: Spearman correlation between introduction risk and network metrics, population size, health and access domains from SIMD. (SPL: Shortest Path Length).

Variable	Winter		Summer	
	Correlation	p-value	Correlation	p-value
In-Flow	0.78	$< .001$	0.73	$< .001$
SPL to mainland	-0.75	$< .001$	-0.78	$< .001$
Population size	0.66	$< .001$	0.52	0.02
Access	-0.67	0.002	-0.65	0.003
Health	0.54	0.022	0.41	0.09

dramatic change in the risk between winter and summer. The most substantial change occurred in Iona, where the risk of introduction before 30 days is 0.2 in winter and 0.82 in summer.

4.3.2 Conditional probability and consequences for control

4.3.2.1 *Relative importance of movements*

Fig. 4.5 shows the probability for each island to introduce the virus over time (in red), and the conditional probability for these islands to introduce the virus given that at least one of their neighbour has introduced the virus (in blue). The limited differences between the two probabilities suggest that the event introduction of COVID-19 on island k before time t is independent to the event introduction of COVID-19 on one of the neighbours of island k before time t.

Secondly the impact on the introduction risk of controlling certain movements is evaluated: movements from the mainland and between islands concomitantly are reduced by 50% (light green, scenario a), or movements from the mainland only are reduced by 50% (dark green, scenario b) Fig. 4.5. Scenario a leads to a larger number of movements being suppressed compared to scenario b. Despite a larger reduction of movements in scenario a, the effect on introduction risk is very similar between the two scenarios.

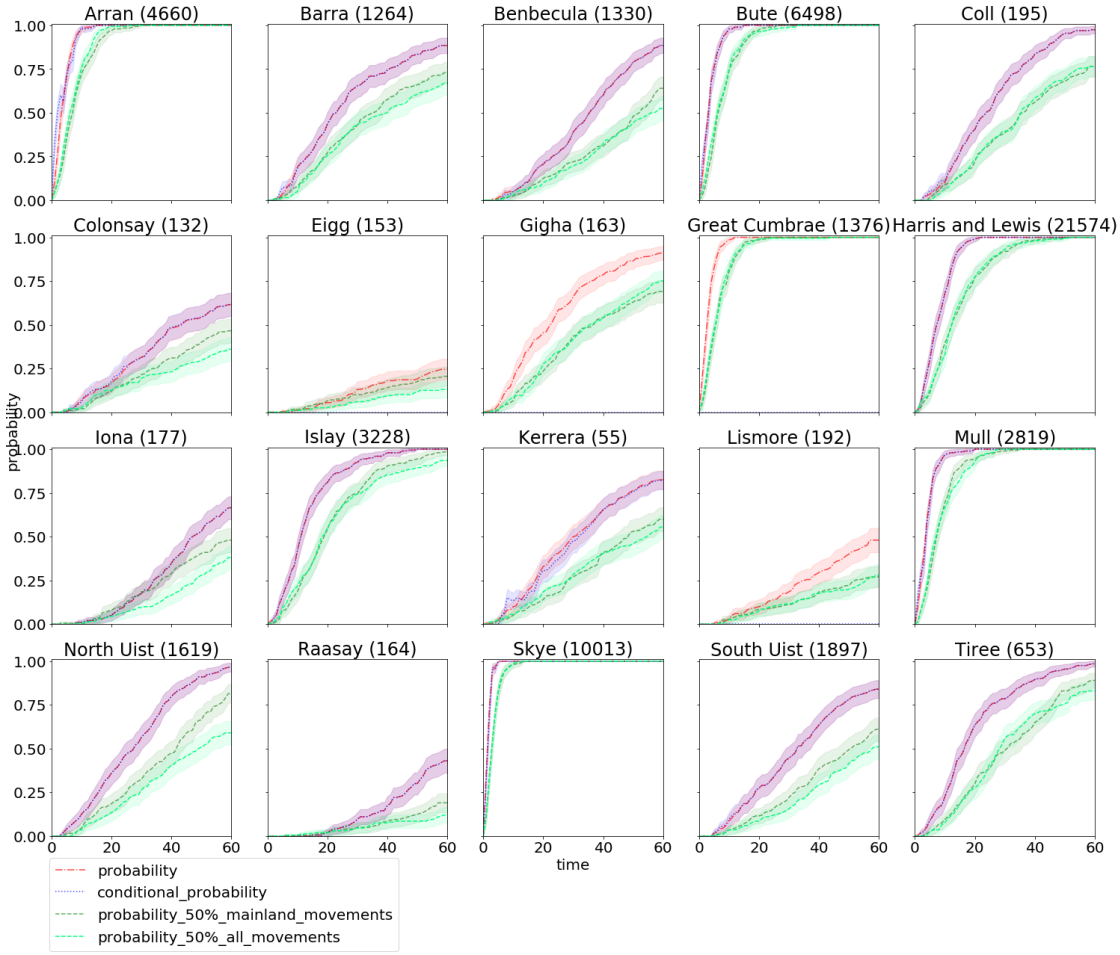

Figure 4.5: Comparison of the probability of introduction in summer (red) and the conditional probability of introduction (blue), with the probability of introduction if all movements from mainland and between islands are reduced by 50% (light green), or if only movements from the mainland are reduced by 50% (dark green). Note that the curve appears purple if the blue and red curves are superimposed.

4.3.2.2 *Delay observed following restrictions*

I consider a reduction by 50% and 70% of the movements from the mainland, where the summer 2020 constitutes the baseline of the scenario.

A cluster analysis using the k-means method allows us to distinguish two groups of islands (see Fig. 4.6), the clustering explaining 77.2% and 77.1% of the group's differences for the 50% and 70% reduction scenarios respectively. When movements are reduced by 50%, there is a strong linear correlation between the reduction in the number of passengers and resulting delay in days (Pearson correlation 0.78, p-value $< .001$) in the larger cluster of islands, as shown by the regression line. In the smaller cluster, although the decrease in number of passengers is higher there is nearly no delay in the introduction. This suggests that above a certain value, changes in movement volume have little impact, as the risk of introduction remains high.

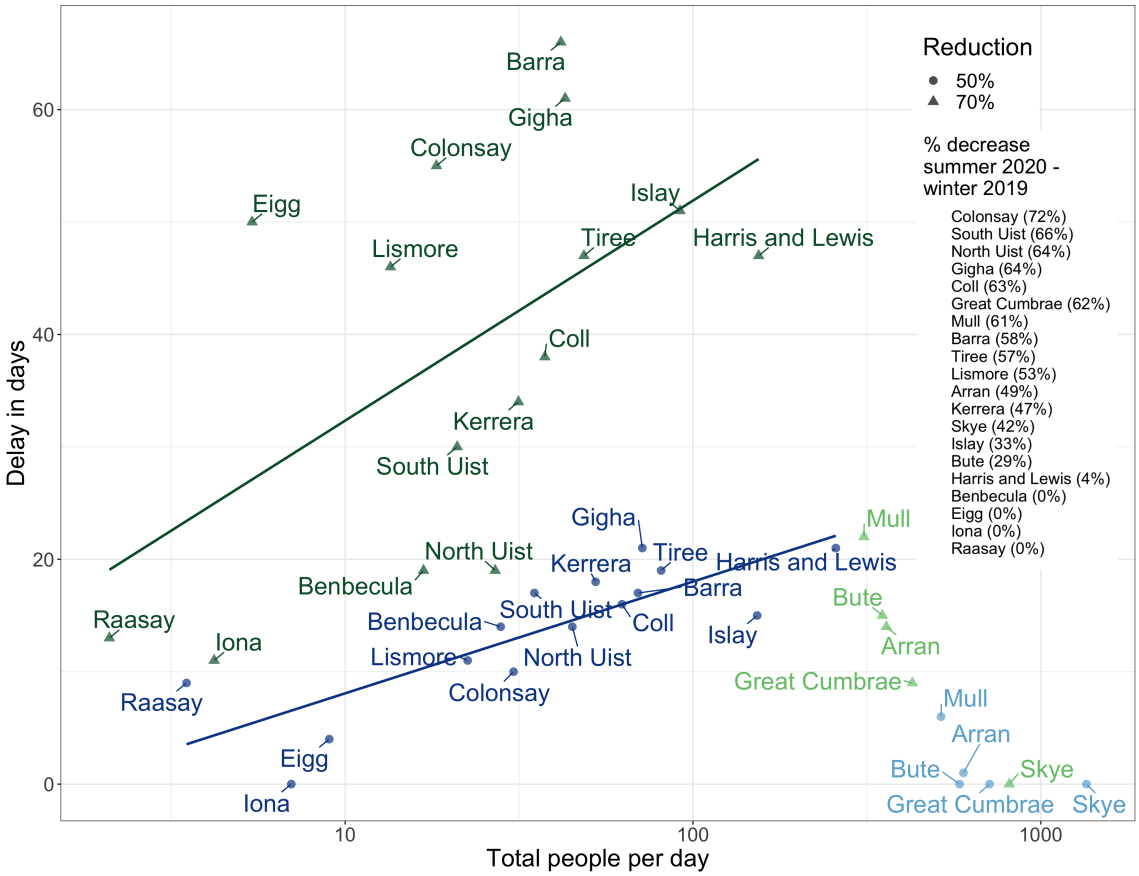

Figure 4.6: Delay in days with movements reduced by 50% (round blue dots) and 70% (triangular green dots) according to baseline probability of introduction before 30 days. The brighter and darker colours show the clusters for each reduction scenario. On the right hand-side the observed decrease in movement between the summer 2020 and the winter 2019 is shown, illustrating what could possibly be achieved by controlling summer movements.

The advantage of restricting movements varied between islands. For a number of islands like Barra, Gigha, Colonsay, there might be a real advantage of doing so, since I see a simulated delay of approximately two months. This means that the restriction would postpone the introduction beyond the length of an average summer holiday period, which also corresponds to a high risk period due to increased tourism activities. Beyond this period, the risk would spontaneously decrease with the seasonal decrease in movements.

4.3.3 Summary of the COVID-19 cases on the islands

COVID-19 has been introduced at several occasions on the islands since the start of the pandemic. The color matrix Fig. 4.7 shows the number of positive tests per island per month between March 2020 and February 2021. I assume that positive tests on different months potentially corresponded to independent introductions. I show that there is a positive correlation between the number of

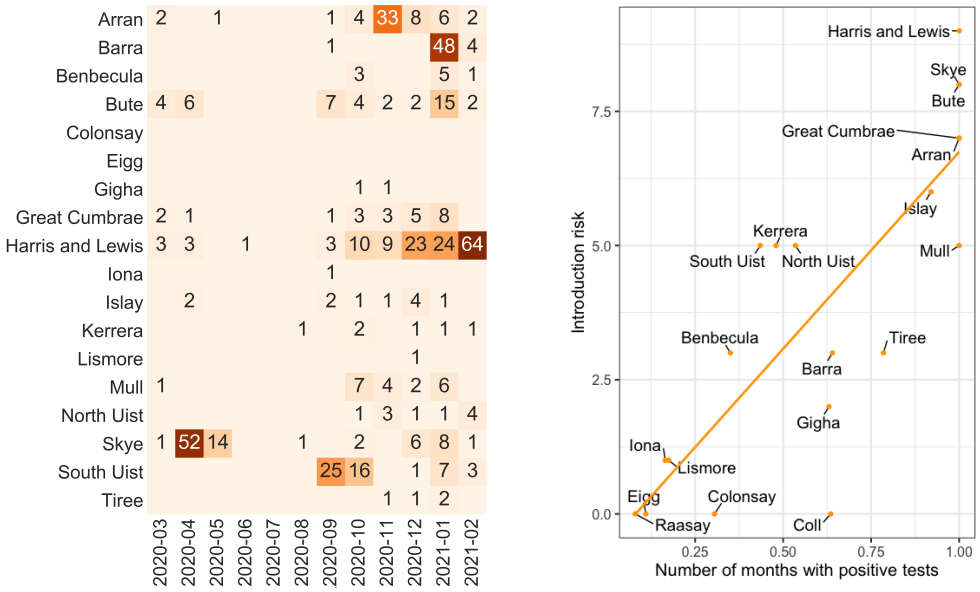

Figure 4.7: Heat-map showing the number of positive tests per month on each island, and correlation between number of months with positive cases and introduction risk.

introductions on each island and the baseline probability of introduction before 30 days calculated for summer 2020 (Pearson correlation 0.81, p-value $< .001$).

4.3.4 *Exploring control scenarios*

Fig. 4.8 highlights the difference in the distribution of the number of individuals who have been infected during the simulation according to the scenario considered.

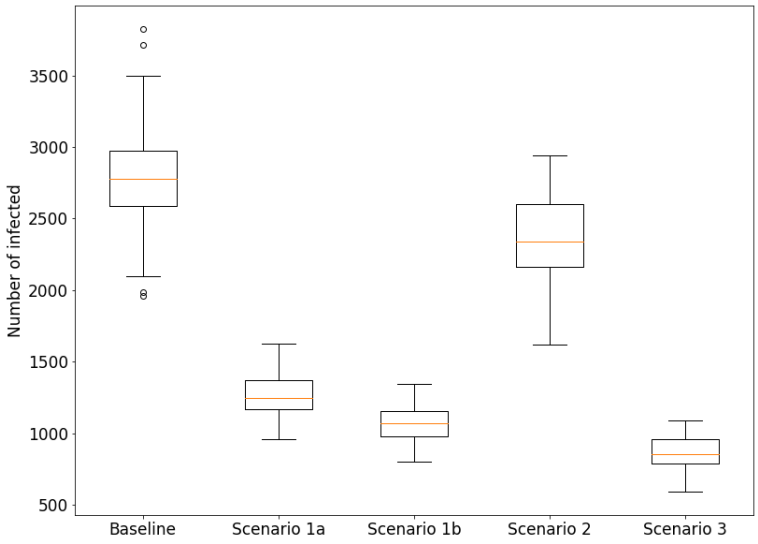

Figure 4.8: Comparison of the total number of individuals infected at the end of the simulation in the Hebrides according to scenarios.

Scenarios involving a reduction of the number of contact only (scenario 1a) are more effective than scenarios involving control on movements only (scenario 1b). This is to be expected as movements play an important role in spreading a disease to new places, but for COVID-19 once the pathogen has been imported, the circulation among smaller communities will be sufficient to sustain the epidemic provided that the number of recovered and immune individuals remain low.

Scenario 1b which is a combination of scenario 1a and 2, results in fewer infected individuals at the end of the simulation, but the decrease observed is not as great as the decrease between the two individual scenarios. This indicates that there is not a synergy when combining the two types of measure but some overlap in the disease transmission effect induced by each of those.

4.4 Discussion and conclusions

Movements play a crucial role in the spread of respiratory disease between territories. My study provides important information on the seasonal variation of risk and effect of control measures that can help decision making and preparedness to such events.

The analysis on the seasonal variation of risk showed that a number of islands (Skye, Bute, Arran, Great Cumbrae, Harris and Lewis) would have a high exposure to COVID-19 introduction from the mainland in summers and winters should the volumes of traffic be similar to *normal*, and the prevalence around 1% on the mainland. For these islands, the risk remains high even in winter when fewer people are moving from the mainland and between islands. Yet COVID-19 might become a seasonal respiratory infection (Audi et al., 2020) in temperate areas. In this case annual epidemics of such respiratory disease would affect the human population in the winter season, when environmental parameters and changes in human behavior become more favourable for the spread (Moriyama et al., 2020). This would be especially of concern since other respiratory infections circulate at the same time, such as flu (Nelson and Holmes, 2007), which consequently increases the bed occupancy during winter months (Bouscambert et al., 2015). The potential impact on hospitals could be more severe in isolated areas like the Hebrides. In the UK, in September 2021, over 43 million individuals had received two doses of COVID-19 (UK Government, 2021), but even with a good vaccination coverage the risk of the emergence of a new variant escaping the vaccine will remain (Domingo and Perales, 2021; Koyama et al., 2020). If and when there is evidence of an escape mutant circulating, additional measures in areas exposed to a higher risk should be considered. Furthermore, the correlation between health index and introduction risk in winter highlights that those islands the most at risk in winter are also the ones that are the most deprived on the health level supporting the need for additional measures.

The results also showed that movements from the mainland are more likely to play a role in the dissemination to new areas compared to movements from neighbouring islands (see Fig. 4.5). Consequently, in order to mitigate the risk of COVID-19 importation to the islands, restriction measures could primarily focus on reducing the passenger volume from the mainland but otherwise allowing continued traffic between islands with no restrictions.

In addition when movements from the mainland are reduced by 50%, some islands see the introduction of the virus delayed and the length of the delay is proportional to the decrease in number of passengers visiting the island. However, for a few other islands (Arran, Bute, Great Cumbrae, Mull and Skye) barely any delay results from the movement reduction. It is important to note that since the chosen prevalence on the mainland is 1%, if more than a thousand individu-

als are moving from the mainland to one of the island every day, it means that the probability to infect an individual on the island would be close to one every day. Therefore the islands that are highly connected to the mainland and experiencing a higher risk would not see any delay in the virus importation with a decrease of 50% of the mainland movements, since the volume of movements remains high enough to induce a high risk.

Finally my study suggests that the importation of the virus to the islands could be delayed by up to two months if movements from the mainland were reduced by 70% compared to the volume observed in summer 2020. This delay of two months should not be neglected in a period of massive vaccination. According to their vaccine delivery plan, the UK aims at vaccinating at least 2 million people per week from the end of January 2021 (UK Government, 2021). This means that any week of delay not only corresponds to time saved, but also to a higher vaccination coverage in the population. In addition, as the level of movements considered in the baseline scenario corresponds to a summer holiday period, the results show that for some islands a movement mitigation is likely to postpone the introduction of COVID-19 beyond the end of the holiday period, after which the risk will decrease with the lower seasonal winter volume of travel. For those islands which could benefit from such a gain, it might be interesting to consider controlling movement during holiday periods to maintain the risk at a low level.

The parameters used in the model were taken from other COVID-19 studies in the UK or elsewhere. These parameters, especially the number of contacts per person and per day was set as the same for all areas, whereas in reality some areas will be more rural, less populated than others. These differences might be especially important in the case of islands which are generally less populated. Using the same parameters as the ones estimated in urban areas might introduce a bias. In addition the prevalence on the mainland was arbitrarily fixed and does not take into account the change in circulation level of COVID-19 over time. However, the calculated probability of introduction before 30 days showed a strong correlation with the number of months with positive tests on each of Hebrides, which reinforces the confidence in the reliability of the method and the results.

Movement restrictions during peak periods might not be viable economically for the islands. According to the National Plan for Scotland's Islands published in 2019 by the Scottish government, sustainable economic development can be achieved through economic drivers such as marine activities, agriculture and crofting, fishing, tourism and the food and drink industry (Scottish Government, 2019). Most of these activities depend on free movement of people, essentially seasonal workers and tourists. Further investigations into the economic impact of these restrictions would have to be conducted to balance disease and economic

risks for the population. These impacts must be taken into account in policy decisions to insure the prosperity of the island communities.

4.5 Appendix

The parameters used in the simulations are detailed in Table 4.3.

To describe the infection process I used the following notations:

- $p = 0.05$ the infection probability of a contact between an infectious person and a susceptible person
- $A_{u,k}$ is age class k at node u.
- $X_t(A_{u,k})$ where X is a compartment in the compartmental model is the number of individuals in compartment X in age class A_k in node u at time t
- $N_t(A_{u,k})$ is the total number of individuals in age class A_k in node u at time t
- $\mathcal{I}_t(A_{u,k})$ is the number of individuals in *any* infectious compartment in age class A_k in node u at time t
- $p_{t,\mathcal{I}}(A_{u,k}) = \frac{\mathcal{I}_t(A_{u,k})}{N(A_{u,k})}$ is the proportion of age class A_k in node u at time t that is infectious, and as $p_{t,S}(A_{u,k}) = \frac{S_t(A_{u,k})}{N(A_{u,k})}$ the proportion of age class A_k in node u at time t that is susceptible
- C is a matrix describing contact between age classes, where $C_{i,j}$ is the expected number of contacts that an individual in age class A_i has with an individual in age class A_j.

The stochasticity is implemented in the calculation of the number of new infected. The number of infectious contacts at node u is calculated by first generating for each infected person in $A_{u,j}$ how many contacts this person had with people in $A_{u,i}$. I randomly sampled from a Poisson distribution, $\mathcal{P}(\gamma_{t-1} C_{i,j})$. Then for each infectious person, and its random number of contacts $contacts$, I randomly sampled from a hypergeometric distribution (the hypergeometric distribution is used instead of binomial because I assumed the contacts are with unique people, hence the random sampling must be done without replacement), with parameters $Hypergeometric(N_t(A_{u,i}), S_t(A_{u,i}), contacts)$, i.e. I sample $contact$ times in the total population of node u, with a "success" being a susceptible person. These potentially infectious contacts for all people are then summed.

The number of between node infectious contacts at time t is calculated by first randomly sampling the number of commutes originating from infectious people in u, using a Binomial distribution,

$$\text{commutes from infectious}_t \sim Bin(p_{t-1,\mathcal{I}}(A_u), x_{t-1} * w(u,v))$$

Table 4.3: Transition rates for the three age groups ('y' for young [0,17), 'w' for working age group [17,70), and 'o' for people older than 70). Parameters were taken from More et al. (2020); Byrne et al. (2020); He et al. (2020).

Age group	From	To	Parameter	Value
[0,17)	E	A_2	$(1-q)\nu_y$	0.15
[0,17)	E	A	$q\nu_y$	0.041
[0,17)	A	I	ζ_y	0.4
[0,17)	A_2	R	ρ_y	0.14
[0,17)	I	R	$p_{ry}\gamma$	0.14
[0,17)	I	H	$p_{hy}\gamma$	5.7×10^{-5}
[0,17)	I	D	$p_{dy}\gamma$	0.0
[0,17)	H	R	ρ_{hy}	0.85
[0,17)	H	D	μ_{hy}	0.0
[17,70)	E	A_2	$(1-q)\nu_w$	0.11
[17,70)	E	A	$q\nu_w$	0.082
[17,70)	A	I	ζ_w	0.4
[17,70)	A_2	R	ρ_w	0.14
[17,70)	I	R	$p_{rw}\gamma$	0.13
[17,70)	I	H	$p_{hw}\gamma$	7.1×10^{-3}
[17,70)	I	D	$p_{dw}\gamma$	2.0×10^{-4}
[17,70)	H	R	ρ_{hw}	0.81
[17,70)	H	D	μ_{hw}	0.04
70+	E	A_2	$(1-q)\nu_o$	0.060
70+	E	A	$q\nu_o$	0.13
70+	A	I	ζ_o	0.4
70+	A_2	R	ρ_o	0.14
70+	I	R	$p_{rw}\gamma$	0.11
70+	I	H	$p_{hw}\gamma$	0.026
70+	I	D	$p_{dw}\gamma$	2.1×10^{-3}
70+	H	R	ρ_{ho}	0.69
70+	H	D	μ_{ho}	0.17

Then given a number of such commutes, I selected those that targeted susceptible people in v.

$$\text{commutes from infectious to susceptible}_t \sim \text{Bin}(p_{t-1,S}(A_v), \text{commutes from infectious}_t)$$

After adding the number of within-node and between-node potentially infectious contacts, I sampled the number of new infections from a Binomial distribution:

$$\text{infectious contacts}_t \sim \text{Bin}(\text{contacts}_t, p)$$

The code for the model framework is publicly available[1].

[1] https://github.com/ScottishCOVIDResponse/simple_network_sim/

5

Transmission phase, Bayesian Inference to understand an ongoing epidemic: Role of individual vs community-wide measures to control Barra COVID-19 Outbreak

Abstract

In this chapter, I focus on the study of a recent epidemic. I consider the outbreak that occurred on the Isle of Barra at the beginning of 2021 and see how its analysis can help us assess the effectiveness of control measures. The use of the disease data in combination with a simulation model allows for the estimation of key disease parameters of the model using Approximate Bayesian Inference. The estimation of key parameter values at different time points throughout the outbreak can help quantify the effectiveness that control measures have had on the transmission dynamic.

I use a compartmental disease model similar to the one presented in Chapter 4 and data relative to number of cases over time. I consider three successive phases with different transmission dynamics to model the outbreak, i.e. initial spread, transmission reduction without population wide restrictions and lock-down. Parameters of the stochastic compartmental model are adjusted to positive test data to estimate key disease parameters in these three phases. I estimate the most likely date of introduction, the basic reproduction number at the start of the outbreak and I quantify the impact of measures in the different outbreak phases.

I find that the virus has most likely been introduced on New Year's Day, the transmission started to slow down two days after the first cases were reported and a week before further restrictions were imposed by the authorities. My results are consistent with the containment of the outbreak by contact tracing combined with self-imposed measures. The lockdown, which was later implemented, had likely no substantial effect on the transmission dynamic.

5.1 Background

Since the start of the COVID-19 pandemic, many countries have been placed on lockdown at different time points to slow the spread of COVID-19 (ACAPS, 2020). A lockdown is a combination of non-pharmaceutical interventions (NPIs) i.e. any methods used to reduce the spread of an infectious disease, apart from vaccination or medication. Non-pharmaceutical methods are crucial in curtailing infectious disease spread, as has been shown previously for influenza pandemics (Markel et al., 2006; Aledort et al., 2007). For COVID-19 and before vaccination was implemented, NPIs, including testing and isolation, represented the primary mitigation strategy as the benefit of individual treatment of patients is constrained by limitations on effectiveness and availability of resources (Song et al., 2020; Ranney et al., 2020). Physical distancing measures have therefore been central in government strategic plans for COVID-19 control (ACAPS, 2020) and their scale and severity were unprecedented. Measures implemented could include: closure of events, schools, restaurants, bars, gyms, and other leisure or hospitality-related businesses. In addition movements of individuals were restricted, and 'stay at home' orders could be enforced in the stricter form of the lockdown.

These decisions proved very effective in containing the spread of the virus (Alfano and Ercolano, 2020), since human contacts drive the transmission. But such measures are not harmless as they have a large impact on the economy and the population's mental health (Clemens et al., 2020; Buheji et al., 2020). The management of healthcare resource was one driver for control strategies (Jombart et al., 2020). More specifically, the availability of critical care beds was modelled as it was one of the main limiting factor for the management of COVID-19 infected patients (McCabe et al., 2021). For instance in March 2020, Ferguson et al. (2020) predicted the number of critical care beds occupied according to various scenarios and suggested that population-wide social distancing would have the potential to mitigate the spread sufficiently.

Later on, the Scottish COVID-19 response transitioned from national lockdown to localised interventions at Local Authority level. A 5-tier system was adopted in November 2020, where five levels of restrictions were defined and each of the 32 Local Authorities could be in any of the five levels. Five critical indicators were used to inform the allocation of levels (Scottish Government, 2020b), namely: (i) the number of cases per 100,000 people over the past seven days, (ii) the percentage of tests that are positive over the past seven days, (iii) forecasts of the number of cases per 100,000 consisting of the weekly number of cases in two weeks' time, (iv) current and projected future use of local hospital beds, compared with capacity, (v) current and projected future use of intensive care beds, compared with capacity. But these indicators might be challenging to

estimate or to interpret as Local Authorities might have different characteristics in terms of demography, distribution of rural/urban areas or healthcare facilities. There might also be more uncertainties in indicator predictions when dealing with small numbers, as it is the case for the Western Isles where 70 hospital beds are available on average, as opposed to several thousands in more populated areas (Public Health Scotland, 2020). In addition, using indicators based on a 7 day period induces a time lag between health event and decision.

Although epidemiological models are now recognised as a central tool to inform decisions (Alahmadi et al., 2020), using model forecasting for policies is often challenging, partly because the predictions highly depend on the interaction between interventions and their impact on parameters. These parameters are often unknown since they are context-specific, i.e. they depend on the pathogen of interest (contagiousness) but also the social context (demography, contact patterns). One can use parameter values estimated in other experimental or modelling studies, but this will not come without introducing bias. For the spread of a respiratory disease like COVID-19, contact and movement patterns are the main drivers of transmission. Contacts between individuals before and during the pandemic have been estimated by surveying a subset of the population in the UK and elsewhere (Mossong et al., 2017; Jarvis et al., 2020b). However these estimates have been made in a context which is not representative of the one of a Scottish Island, as the population density, commuting and contact patterns might be different. Barra is a small island of the Outer Hebrides (see Fig. 5.1) counting 1,174 inhabitants and approximately 20 inhabitants per square meter (compared to 3,555/km^2 in Glasgow for example). Barra is a remote rural area according to Scottish Government Urban Rural Classification (Scottish Government, 2021a).

If disease data are available, several methods exist to estimate parameters from data. This process, called model fitting, is usually based on likelihood methods (Csilléry et al., 2010). But in infectious disease modelling the likelihood function is often computationally intensive or intractable because of model complexity and data scarcity and incompleteness due to the nature of partially observed epidemics. Approximate Bayesian Computation is a method based on Bayesian Inference where the likelihood is approximated (Beaumont et al., 2002). The likelihood calculation is bypassed by comparing summary statistics of the simulated data to summary statistics of the observed data. As in Bayesian inference, prior beliefs about the model parameters, expressed through the prior distribution, are updated in light of new evidence obtained by simulations. Adjusting disease transmission parameters to data can help gain insight in the timeline of the transmission dynamics and therefore the effect of decisions.

Here, I analyse an outbreak which occurred on the Isle of Barra, situated in the Western Isles, off the West coast of mainland Scotland (see Fig. 5.1. The first two cases were reported on January 11th 2021. In the next days the NHS chief executive advised inhabitants of Barra to "limit contact with other households and leave home only if necessary" in his daily video update addressed to the Western Isles population. On January 15th, over 110 inhabitants were self-isolating, nearly 10% of the local population. Increasing numbers of new cases were being identified until January 19th, when Barra moved to "level 4" restrictions at midnight. On that day a total of 45 cases had been detected and linked to the epidemic. Three more cases were reported in the next few days, bringing the number of cases to 48 on February 2nd, followed by a whole week with no cases (NHS Western Isles, 2021). The role of the lockdown in controlling this epidemic is questionable as it occurred when the epidemic peak had already been reached. Our aim is to estimate the role of voluntary vs. policy-induced measures in containing the outbreak on Barra. The epidemic is first described with indicators, such as number of positive tests, positivity over time, growth rate. Using an ABC method, I propose to fit disease parameters to the available data of the epidemic to provide likely estimates of those in this context. Finally, I investigate the parameter values and their meaning in terms of transmission dynamics along the epidemic. I found that voluntary measures more likely played a central role in containing the outbreak compared to policy-induced restrictions.

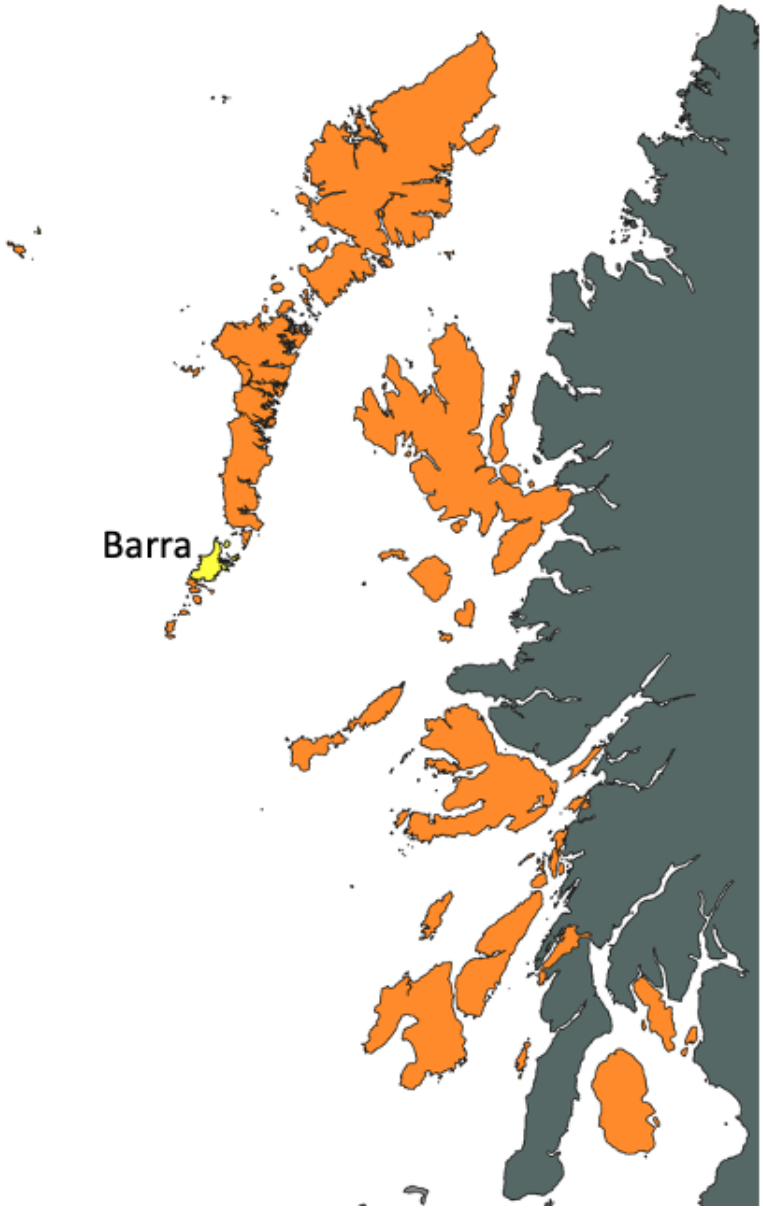

Figure 5.1: Map showing the location of the Isle of Barra (bright yellow), within the Hebrides (orange).

5.2 Methods

5.2.1 Data

I use data on COVID-19 held by Public Health Scotland (PHS) and made available via the PHS Electronic Data Research and Innovation Service (eDRIS, PHS (2020)). The data contains the number of positive and negative tests per day per data zone, with age group and gender for each observation. A timeline of the epidemic with restrictions imposed was retrieved from the NHS Western Isles COVID-19 website (NHS Western Isles, 2021). To populate the model I use age demographics data obtained from the 2011 census and consider three different age groups.

I consider cases occurring between January 10th and February 2nd as part of the epidemic of interest. Between these dates 167 tests were performed in Barra and 48 individuals tested positive. The overall positivity was relatively high, 29%, compared to the positivity in Scotland during the same period (24%). The percentage of individuals tested amongst the Barra population was 13% (whilst 4% of the Scottish population got tested during the same time frame) and the percentage of people testing positive 4% of the 1,264 inhabitants (according to the 2011 census, Scotland's Census (2011)). Fig. 5.2 shows the histogram of the distribution of tests and cases per age group. The number of positive tests per age group did not show significant variation.

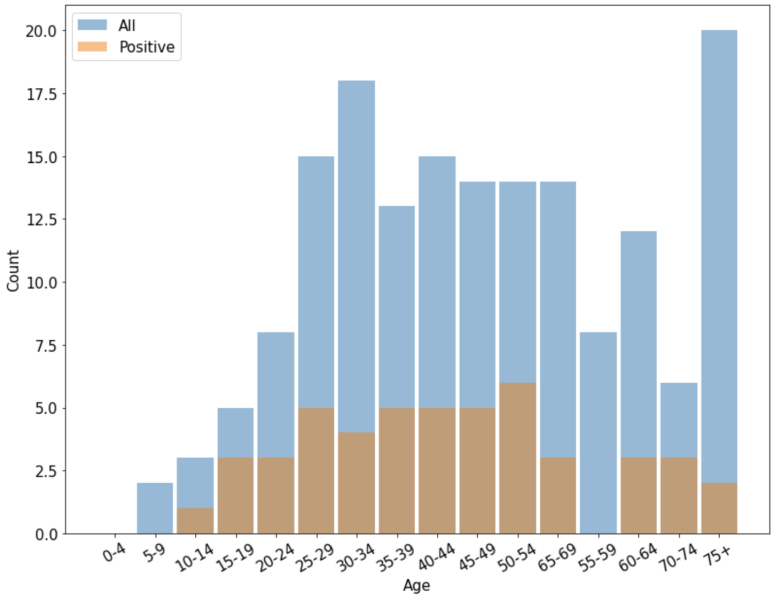

Figure 5.2: Histogram showing the number of individuals tested and positive per 5-year age group.

5.2.2 *Outbreak description*

The outbreak timeline is detailed in numbers and restrictions according to test results from PHS data and NHS reports (NHS Western Isles, 2021). I use indicators, such as number of positive tests, positivity over time, overall and per age group to describe the PHS data on tests performed during this outbreak.

Since most epidemics grow exponentially during the initial phase of an epidemic (Ma, 2020), one can infer the growth rate from the relationship between daily cumulative cases with time in log-linear scale. The initial growth rate of the epidemic is estimated graphically by linear regression using the log of the cumulative number of cases.

5.2.3 Parameter fitting

5.2.3.1 Model

I use a stochastic compartmental model representing the island population. The population is split into three age groups ([0-17), [17-70), and 70+) with populations derived from census data (Scotland's Census, 2011). The disease states of the model include a latent (E) state, asymptomatic (A2), presymptomatic (A) and symptomatic (I) infectious states, and subsequent hospitalised (H), recovered (R) or dead (D) states (schematic representation in Fig. 5.3). To drive infectious dynamics, contact rates between individuals were derived from age-structured mixing matrices based on POLYMOD survey data (Mossong et al., 2017).

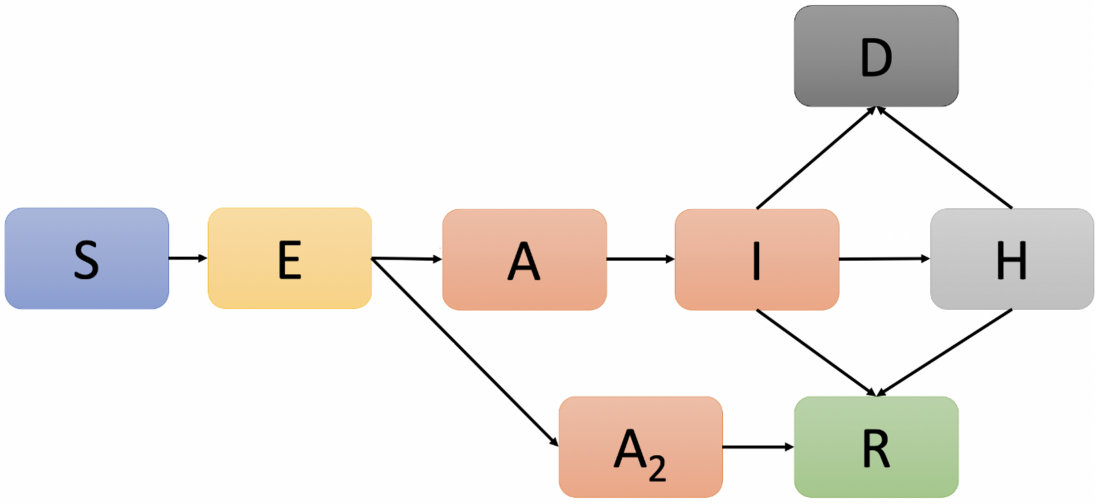

Figure 5.3: Schematic representation of the compartmental model.

The description of the model and the transition rates are available in the appendix of Chapter 4.

Three phases are considered to describe the epidemic (see Fig. 5.4):

- Phase 0: exponential growth of the epidemic,
- Phase 1: change in transmission dynamic after first cases are detected,
- Phase 2: change in transmission dynamic due to a lockdown declared on Jan 18th.

5.2.3.2 Principle and algorithm

An Approximate Bayesian Computation method is used to estimate disease parameters of the model from data. The general principle and the specific ABC method used here are detailed. The general idea as in Bayesian Inference is to update the probability for a hypothesis as more evidence or information becomes available. If the data is denoted D and the mathematical model parameters to fit

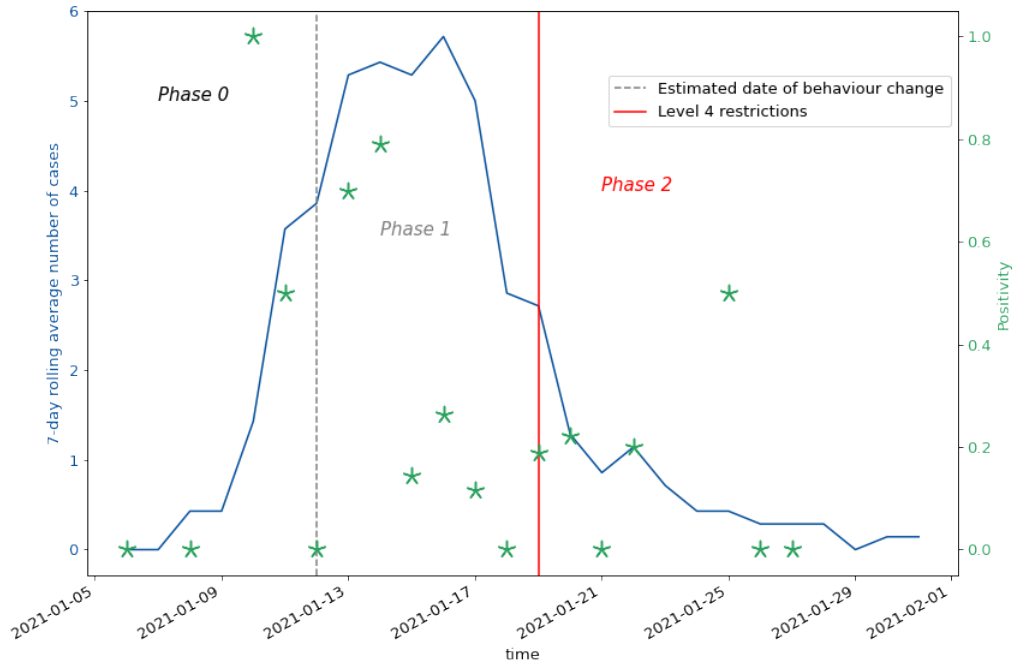

Figure 5.4: 7-day rolling average number of cases and positivity over time. Vertical lines show the transitions between epidemic phases defined in text.

θ, using the Bayes rule I can determine the posterior distribution of the parameters given the data $P(\theta|D)$:

$$\begin{aligned} P(\theta|D) &= \frac{P(D|\theta)P(\theta)}{P(D)} \\ &\propto P(D|\theta)P(\theta) \end{aligned} \tag{5.1}$$

where $P(\theta)$ is our prior belief (called prior distribution), and $P(D|\theta)$ is the likelihood function, the probability density function for the data given the parameters. An ABC algorithm allows for approximating the posterior probability $P(\theta|D)$ without using a likelihood function. In order to successively accept or reject a set of parameter values, so called particles, the algorithm measures whether the distance $d(s, s*)$ between the data $s = S(D)$ and the model simulated data $s* = S(D*)$ is less than or equal to some threshold ϵ. Multiple algorithms are available which mainly differ in the way the parameters are sampled. The way the parameter space is sampled influences the efficiency of the algorithm. ABC-SMC (Sequential Monte Carlo) tends to perform better than other algorithms (Toni et al., 2009).

With the ABC-SMC, a sequence of distributions is constructed by gradually decreasing ϵ. It first samples a finite number of parameter sets from a prior distribution $\pi(\theta)$ and obtains each intermediate distribution (generation) as a weighted sample from the previous distribution perturbed through a kernel $K(\theta|\theta^*)$. Particles sampled from the previous distribution are denoted by a single

asterisk, and after perturbation these particles are denoted by a double asterisk. The algorithm can be summarised as follows (Toni et al., 2009; Toni and Stumpf, 2009):

1. Set a number of generations G and a number of particles N
2. Set a tolerance schedule $\epsilon_1 > \epsilon_2 > ..., > \epsilon_G$ and a generation indicator g=1
3. Set particle indicator i=1
4. If g=1, sample parameter θ^{**} from the prior,

   Else sample parameter θ^* from the previous population $\theta^{(i)}_{g-1}$ with weights $w^{(i)}_{g-1}$ and perturb the particle to obtain $\theta^{**} \sim K_g(\theta \mid \theta^*)$, where K_g is a perturbation kernel.

   If $\pi(\theta^{**}) = 0$, return to step 4.
5. Simulate model outputs D^* with the sampled parameters. Compare the simulated data set D^* to the experimental data D using the distance function, $d(D, D^*)$, and tolerance ϵ_g;

   If $d(D, D^*) \leqslant \epsilon_g$, accept θ^*.
6. Set $\theta^{(i)}_g = \theta^{**}$ and calculate the weight for particle $\theta^{(i)}_g$,

$$w^{(i)}_g = \begin{cases} 1, & \text{if } g = 0, \\ \frac{\pi(\theta^{(i)}_g)}{\sum_{j=1}^{N} w^{(j)}_{g-1} K_g(\theta^{(j)}_{g-1}, \theta^{(i)}_g)}, & \text{if } g > 0. \end{cases} \tag{5.2}$$

   If $i < N$, set i=i+1, go to step 4.
7. Normalise the weights.

   If $g < G$, set g=g+1, go to 3.

The ABC-SMC is performed using the 'pyabc' package (Klinger et al., 2018) in python (Van Rossum and Drake Jr, 1995). The number of particles per generation is set to 500 and the number of generations to 25. I found that running five simulations and calculating D^* as the mean of those runs improved the efficiency of the algorithm avoiding to select parameter sets on runs that were non representative of the average behaviour of the model.

5.2.3.3 *Summary statistics and distance function*

The observed data D used to fit the model are the 7-day average number of positive cases per day over the period of the epidemic. To compare the output of the model to this data I choose as summary statistics D^* the mean number of

new infectious (number of individuals entering states A_2 or A) averaged over 7 days and multiplied by the proportion of cases reported ($\mathtt{reported\_proportion}$). The distance between observations and simulations is calculated as the sum of squared errors (Cf Eq. (5.3)).

$$d(D^*, D) = \sum_t (D^*(t) - D(t))^2 \tag{5.3}$$

5.2.3.4 *Parameters and priors*

The key parameters related to the infectious process are fitted, whereas other parameters, such as transition to R, H, D were taken from the literature. As little was known about the fitted parameters, the prior distributions are assumed to be uniform. They are therefore informative in a sense that they define a feasible range of values, but they are predominantly non-informative as they do not specify any further preference for particular values. This way the inference will mostly be informed by the information contained in the data. The start date, i.e. day at which an individual of the island entered state E, was fitted. The prior distribution was uniform, between December 20^{th} and January 4^{th}.

The length of the pre-infectious and infectious periods were fitted as they are important determinants for both the basic reproduction number and the epidemic dynamics (Sadun, 2020). The prior (see Table 5.1) of the latent period was between 1 and 5 days, whereas the infectious period could persist from 2 to 10 days (More et al., 2020; Byrne et al., 2020; He et al., 2020). In our model, the latency period is equal to the time spent in compartment E. The length of the infectiousness is equivalent to the time spent by an individual in A_2 if asymptomatic, or in A and I if symptomatic. $\mathtt{latent\_period}$ and $\mathtt{infectious\_period}$ are defined as follows:

$$\begin{aligned} \mathtt{latent\_period} &= \frac{1}{\nu} \\ \mathtt{infectious\_period} &= \frac{1}{\mu} + \frac{1}{\gamma} = \frac{1}{\rho} \end{aligned} \tag{5.4}$$

Other transmission parameters characterising the transition from S to E were adjusted allowing for variations between the three phases Fig. 5.4.

$$\begin{aligned} &\textbf{Phase 0}\ \beta_{ya}^{(0)} = p \times \beta_{ya} \times C_{ya} \\ &\textbf{Phase 1}\ \beta_{ya}^{(1)} = cm1 \times p \times \beta_{ya} \times C_{ya} \\ &\textbf{Phase 2}\ \beta_{ya}^{(2)} = cm2 \times p \times \beta_{ya} \times C_{ya} \end{aligned} \tag{5.5}$$

Where the contact rate in each of the phase i $\beta_{ya}^{(i)}$ is the effective contact rate between susceptible individuals in age group y and infectious individuals in age group a and is defined as the product between the probability of transmission given contact p, the contact rate β_{ya}, the contact matrix C_{ya} and the contact multiplier in phase (1) and (2), $cm1$, $cm2$.

Two changes in the dynamic are considered because the main change of restrictions (level 4 measures) was declared on Jan 18th taking effect on the 19th, which occurs after the epidemic peak according to the observed number of cases Fig. 5.4. An earlier change in dynamic is therefore implemented and the date of this change is fitted ($date\_change$).

It is generally assumed that cases in official statistics are likely an underestimate of the total (Battegay et al., 2020; Fauci et al., 2020; Colman et al., 2021). For example, in the particular case of COVID-19, numbers of infected asymptomatic individuals are likely to be missed. In Scotland, the Scottish Government encourages people to get a PCR test only if they have symptoms compatible with a COVID-19 infection (Scottish Government, 2021b). In addition, in an ongoing outbreak the testing effort is likely to be scaled up after the detection of the first cases, impacting the number of reported cases (Russell et al., 2020). This is illustrated by the variation of the positivity over time (see Fig. 5.4), which shows high values at the start of the epidemic, followed by a decrease as the total number of tests increases more than the number of positive tests. Although the proportion of reported cases is likely to vary over time during an epidemic, for the sake of simplicity, a constant parameter $reported\_prop$ is defined and is adjusted in the ABC.

The list of fitted parameters, their definition and prior are summarised in Table 5.1.

5.2.4 Outbreak analysis and impact of the lockdown

Two methods are used to estimate the basic reproduction number. Using the epidemic growth rate and the length of the pre-infectious and infectious period, R_0 is calculated using the equation:

$$R_0 = (1 + GR \times Le)(1 + GR \times Li) \tag{5.6}$$

Secondly a method based on the Next Generation Matrix (NGM) is used to estimate R_0. The NGM has been proposed by Diekmann et al. (2010) where the NGM is defined based on a *transmission* T and a *transition* Σ matrix and is given by the equation:

$$NGM = -T\Sigma^{-1} \tag{5.7}$$

Table 5.1: Fitted parameters.

Parameter	**Definition**	**Prior**
start_date	Date at which an individual transitioned in state E	U(20 Dec, 4 Jan)
p	Probability of transmission given contact	U(0,1)
latent_period	Latency period, time spent in compartment E	U(1, 5)
infectious_period	Length infectiousness, time spent in state A_2 or A and I	U(3,10)
reported_prop	Proportion of existing cases reported through testing	U(0.2, 0.8)
date_change	Date of start of phase 1	U(11 Jan, 18 Jan)
contact_multiplier_1 (cm1)	Constant by which the contact matrix is multiplied during phase 1	U(0, 1)
contact_multiplier_2 (cm2)	Constant by which the contact matrix is multiplied during phase 2	U(0, 1)

R_0 is obtained by computing the dominant eigenvalue of this matrix.

5.3 Results

5.3.1 Outbreak description

Fig. 5.5 shows the 7-day average cumulative cases over time in linear and log scale. On the log scale graph, the regression line is shown. The slope of this fitted regression is the growth rate of the epidemic $GR = 0.67 day^{-1}$.

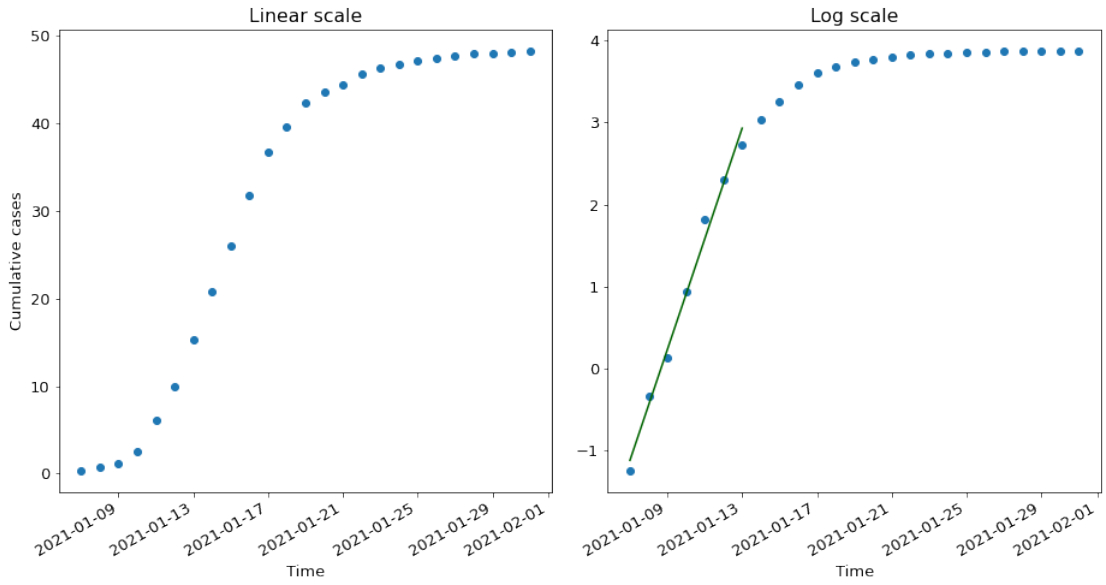

Figure 5.5: Graph showing the 7-day average cumulative number of cases over time in linear (left hand-side), and log scale (right hand-side).

5.3.2 Parameter fitting

Fig. 5.6 shows the posterior distributions of the parameters in all generations of the ABC-SMC. The mode (representing the most likely value) and 95% confidence intervals are reported in table Table 5.2.

According to our parameter estimates, the virus is likely to have been introduced around the New Year. The latency period and the infectious period were on average 2.1 and 3.7 days long respectively. The transmission probability was estimated at 0.18 during the initial phase, phase 0. The first contact multiplier took an average value of 0.023 reflecting a decrease in the transmission rate during phase 1, followed by a similar decrease (0.030) after the level 4 restrictions were declared (phase 2). I estimate that the first phase of transmission mitigation represented by phase 1 started on January 13^{th}. Finally, 41% of the overall proportion of cases were estimated to be detected through testing.

In Fig. 5.7, the matrix of correlations between parameters is shown. The parameters `cm1`, `date_change` and `p` are correlated with `latent_period`. In addition

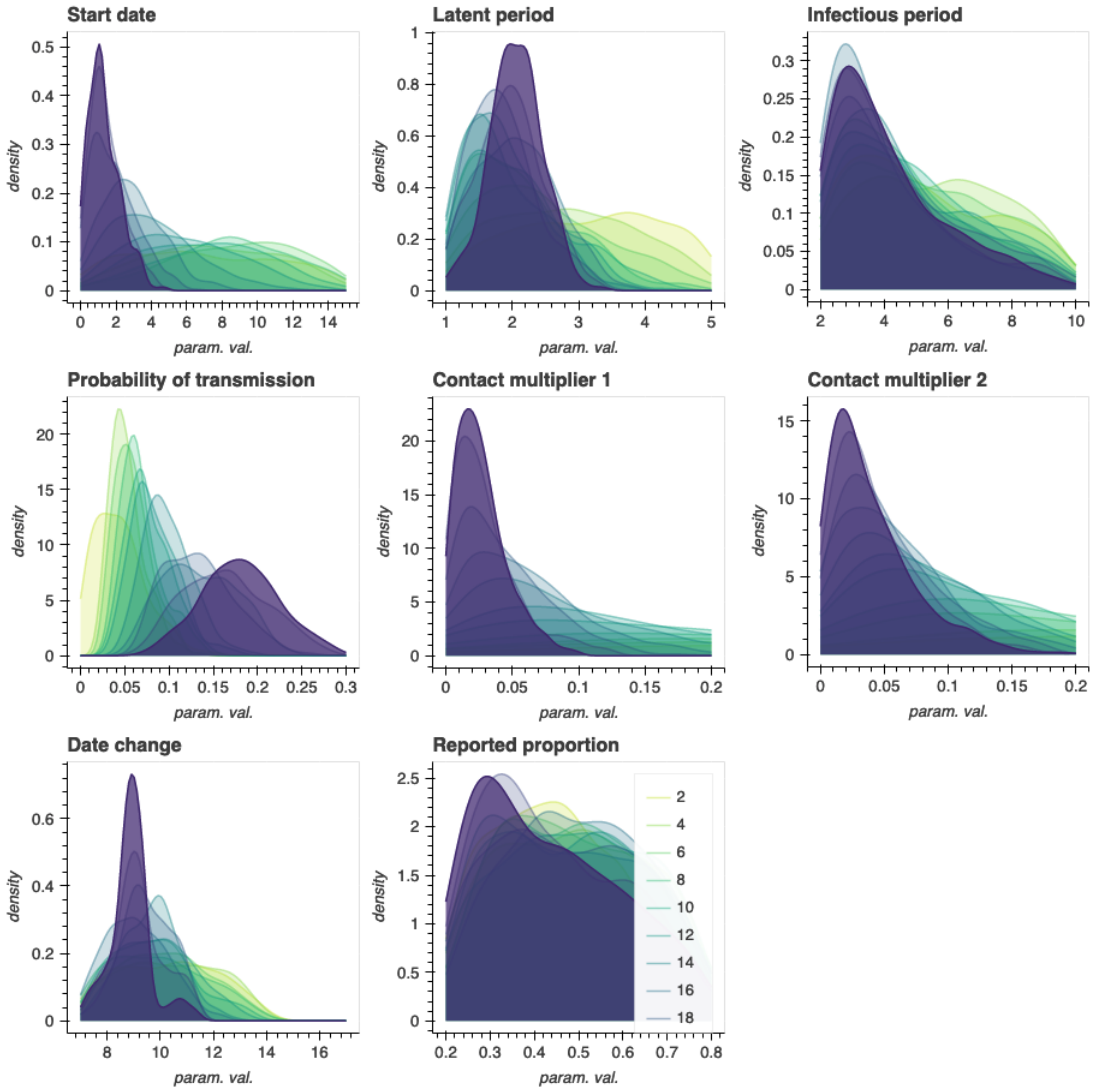

Figure 5.6: Posterior distributions of the fitted parameters in all generations of the ABC algorithm.

Table 5.2: Parameters fitted value ranges (median, 95% variation).

Parameter	Mode	95% quantiles
start_date	1 Jan	[30 Dec, 2 Jan]
latent_period	2.1	[1.4, 2.7]
infectious_period	3.7	[2.2, 7.8]
p	0.18	[0.11, 0.26]
contact_multiplier_1	0.023	[0.003, 0.064]
contact_multiplier_2	0.030	[0.003, 0.111]
date_change	13 Jan	[12 Jan, 14 Jan]
reported_prop	0.41	[0.22, 0.72]

p is correlated with date_change, and inf_period with cm2. All other parameters show low correlations.

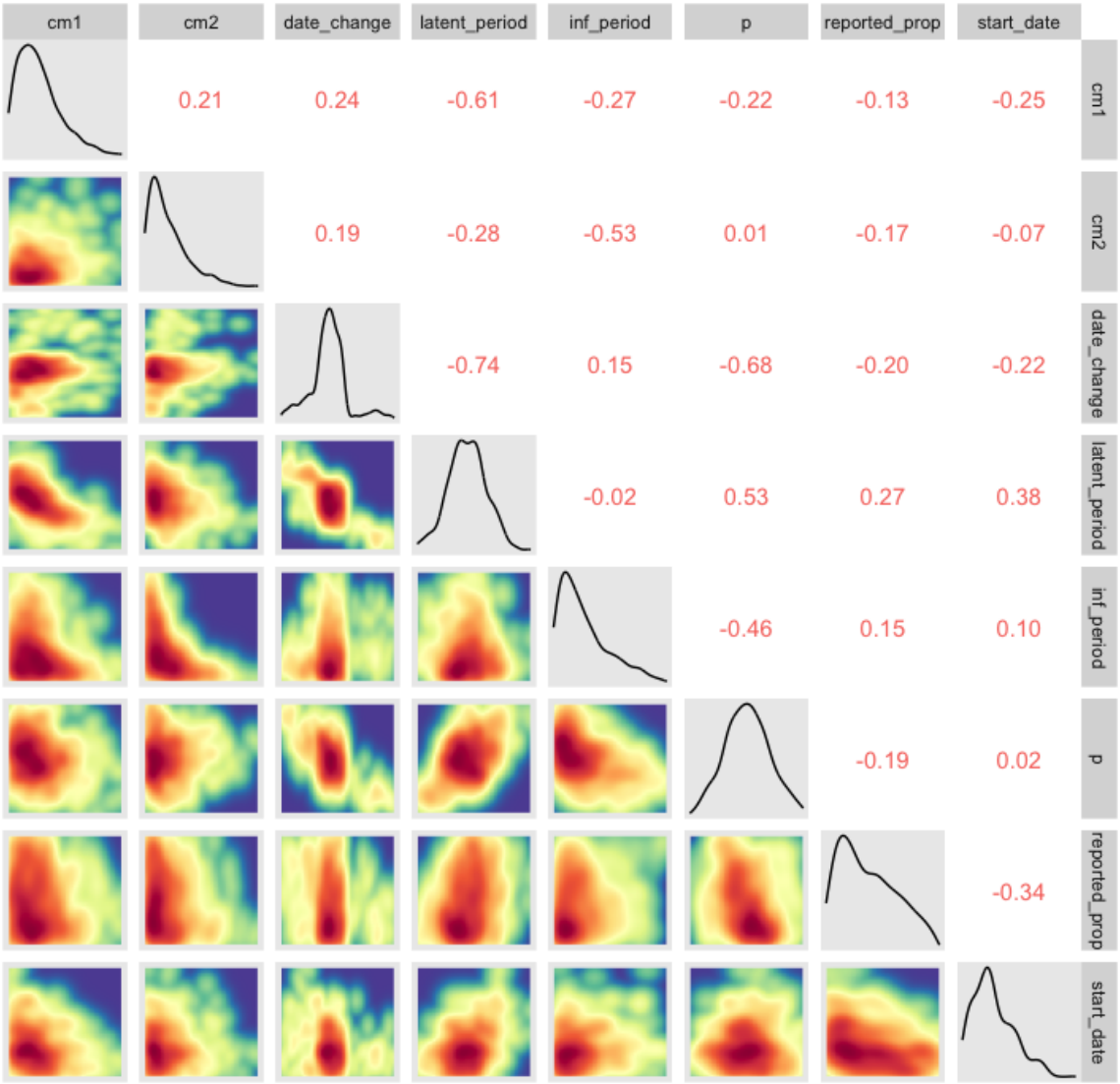

Figure 5.7: Graph showing the correlations between parameters with p the Probability of Transmission, cm1=contact_mulitplier_1, cm2=contact_mulitplier_2

The fit of the model to the data is shown in Fig. 5.8.

5.3.3 *Outbreak analysis and impact of the lockdown*

Using the NGM method, the basic reproduction number had a median value of 6.4 and a 95% confidence interval between 4.6 and 10.6.

In addition, the transmission rate observed an average 97.7% decrease between phase 0 and phase 1. During the phase 2, the dynamic observes a comparable decrease with a 97% decrease in transmission compared to the initial spread.

When the model was run considering only the decrease in transmission due to voluntary measures, no significant difference was observed in number of cases compared to the scenario with lockdown (cf Fig. 5.9). In a scenario where none

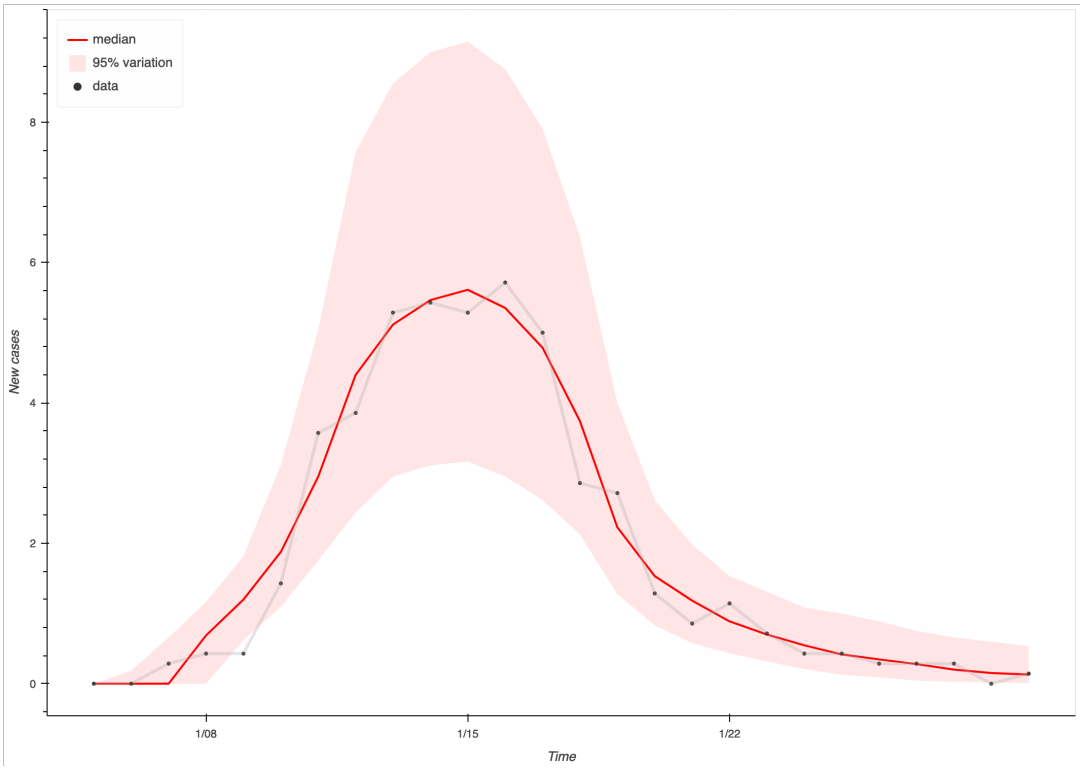

Figure 5.8: Graph showing the result of simulations using the posteriors.

of the latter restrictions happened the epidemic peak is reached 10 days later compared to scenarios with restrictions. The decrease in number of cases is then only driven by the depletion in susceptible in the population. The outbreak size would have been around 499 cases, i.e. a five fold increase compared to what was observed.

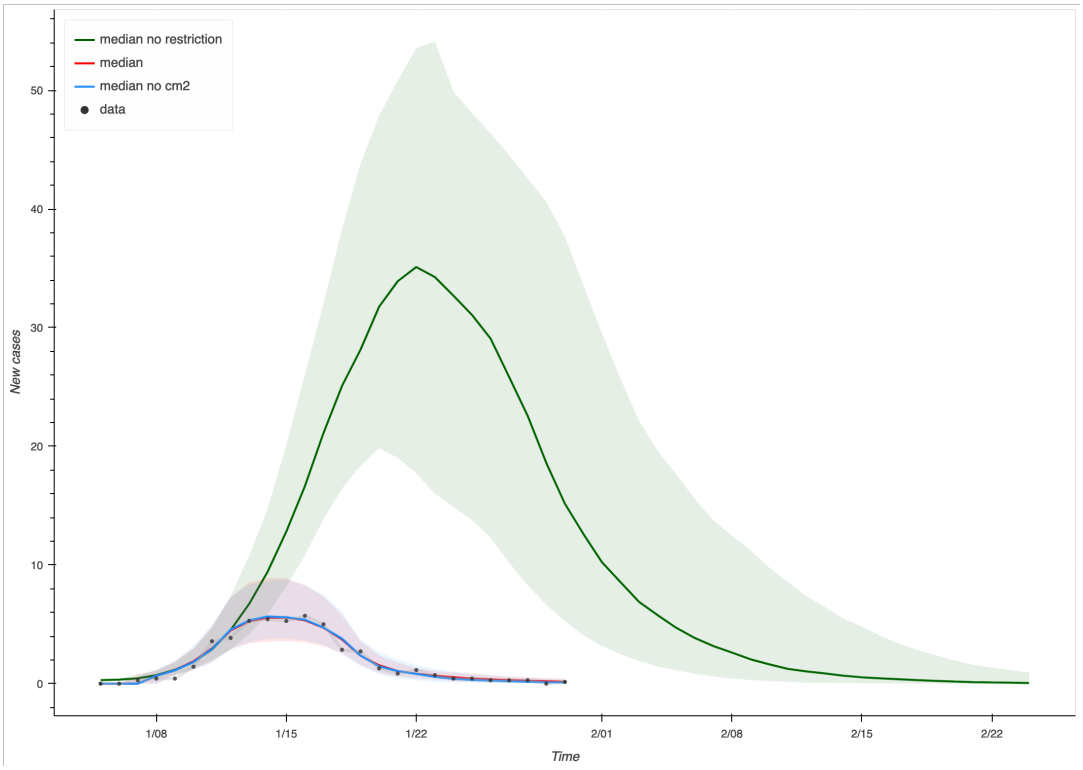

Figure 5.9: Graph comparing the result of simulations using the fitted parameter posteriors considering all three epidemic phases (red), if no lockdown had been implemented (blue) and if the two restriction stages hadn't happened (green).

5.4 Discussion and conclusions

I undertook a parameter estimation study to quantify the role of control measures. The results of the study provide an estimation of the basic reproduction number of the epidemic and the effect individual-based vs community-wide measures on the spread of COVID-19 in the Isle of Barra in January 2021.

I found that the virus has most likely been introduced in Barra on New Year's day.

The basic reproduction number estimated from the NGM had a median value of 6.4 [4.6, 10.6]. This estimate lies within the higher range of values that have been estimated for R_0 in various settings (Salom et al., 2021; Ahammed et al., 2021). It is important to note that R_0 is not an intrinsic characteristic of a given pathogen, but rather describes the transmissibility of that pathogen within a specific population and context. Our estimate might differ from others because the epidemic started during a festive period in the winter season, combining environmental and social factors which are advantageous for the spread of COVID-19 (Mecenas et al., 2020; Mallapaty, 2020). In terms of environmental conditions, laboratory experiments have revealed that lower temperatures and reduced UV levels increase viral persistence on surfaces and aerosols (Ratnesar-Shumate et al., 2020; Dabisch et al., 2021). Infectious virus also degrades faster on surfaces in warmer and more humid environments (Riddell et al., 2020). This

in addition to people spending more time indoors in poorly ventilated places can explain the rapid spread observed in winter in temperate areas. The high value of R_0 is in agreement with an outbreak starting in winter and during the festive period when number of contacts between individuals is unusually high. It is also in line with the sharp increase in number of cases observed at the start of the outbreak in the data.

The results confirm that a change in the disease dynamic had occurred prior to the implementation of the lockdown, and is likely to have happened around January 13^{th}, two days after the first cases were detected. The single control strategy implemented by the government in that case was contact tracing. Although contact tracing has proven effective especially when the number of cases to trace remains limited, there exists a delay between the start of the process and the effect due to logistics, e.g. time to get test results, to communicate with close-contacts, etc. Since the NHS chief adviser had recommended to stay at home and limit contact with other household, the population of the Island might have rapidly complied to the recommendations, as it has been observed in other settings (Kamerlin and Kasson, 2020; Yan et al., 2021). A change of behaviour can have an instantaneous effect by mitigating the number of contacts between individuals. These results also suggest that this early change in behaviour was very effective in containing the epidemic inducing a 97% reduction of the transmission. Although this reduction is substantial, it is important to remember that the contact mitigation depends on the contact rate in phase 0 which was high since the epidemic has started during the festive period. Consequently, the change in contact rate between phase 0 and phase 1 possibly entails several phenomena: return to normal life, contact tracing, change of behaviour with raise of awareness. According to the parameter estimation, the lockdown had likely no significant effect, since the mitigation process was already well under way. Since the model used was a compartmental model accounting for a unique population, the reduction of the transmission rate encompasses both, the potential mitigation in mobility and number of contacts.

I estimated that approximately 41% [22%, 72%] of the cases had been reported during the outbreak, meaning that the total number of cases could lie between 68 and 223.

Fig. 5.7 shows that there is some correlation between parameters `date_change`, `p`, `contact_multiplier_2`, `reported_prop` and `Le`, and between . This is not unexpected since these parameters all have a strong influence on the growth phase of the epidemic (phases 0 and 1 Cf Fig. 5.4). The correlation shows that there is likely an overlap in the parameters' role but these parameters have been included because they help represent the epidemic phases in a coherent and intelligible way.

For an island, controlling movement might be easier than in other settings thanks to their natural geographic boundaries. Movement restrictions can be implemented to delay introduction (Chapter 4) or to prevent further spread. But during this outbreak, seven cases had to be airlifted to the hospital in Stornoway or the mainland. A subsequent outbreak was declared in the hospital in Stornoway starting January 27^{th}. This reinforces the need for preventing introduction in isolated territories, as facilities for patient care are limited and transportation of patient might lead to the spread of the disease to new places.

During an outbreak, mathematical and simulation models are one element of the evidence - they can be used to synthesize data coming from different sources and project forward to give some insight into what might happen next. Epidemiological models can help produce estimates of the effect of various interventions in reducing disease burden, but one should remember that these models account for only a fraction of the factors potentially influencing the spread. Often using mathematical models to inform policy decisions is challenging, because, for the results to be timely, the data and modelling framework need to be readily available.

Appendix

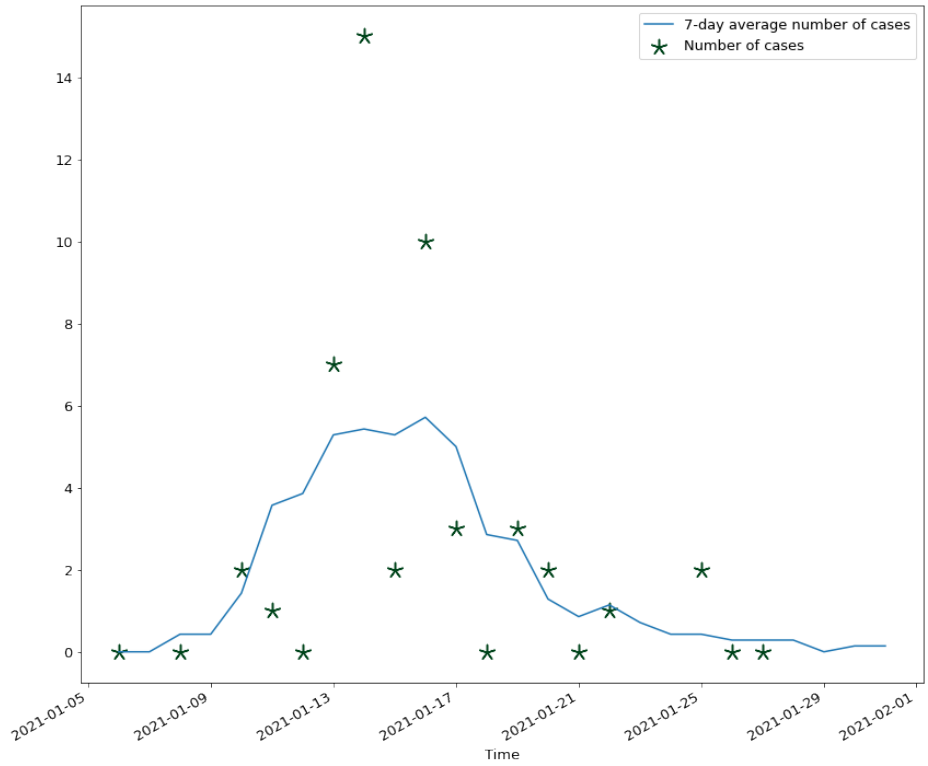

Figure 5.10: Number of cases per day and 7-day average.

Model equations for infected states for the age group 'a':

$$\begin{aligned}
\frac{dE_a(t)}{dt} &= \Lambda_a S_a(t) - \nu E_a(t) \\
\frac{dA_a(t)}{dt} &= q\nu E_a(t) - \mu A_a(t) \\
\frac{dI_a(t)}{dt} &= \mu A_a(t) - \gamma I_a(t) \\
\frac{dA_{2a}(t)}{dt} &= (1-q)\nu E_a(t) - \rho A_{2a}(t)
\end{aligned} \tag{5.8}$$

Where $S_a(t)$, $E_a(t)$, $A_a(t)$, $I_a(t)$ and $A_2(t)$ are the number of individuals in the respective state of age group a and at time t

a = age group (y: young, w: working, e: elderly)

Λ_a = force of infection age group a

$\frac{1}{\nu}$ = pre-infectious period

$\frac{1}{\mu}$ = length of presymptomatic phase before symptom onset

$\frac{1}{\gamma}$ = length of symptomatic phase

$\frac{1}{\rho}$ = length of asymptomatic phase for asymptomatic cases

The force of infection Λ_y for the 'young' age group during the phase (i) (0: initial growth, 1: transmission mitigation phase, 2: lockdown) is given by:

$$\Lambda_y^{(i)} = \frac{\sum_a \beta_{ya}^{(i)}(A_a(t) + I_a(t) + A_{2a}(t))}{N} \tag{5.9}$$

where $\beta_{ay}^{(i)}$ is the effective contact rate at which a susceptible person in age group a comes into effective contact with an infectious person from age group y and is defined in Eq. (5.5).

6

Closure

In the 20th century, following the hygiene and sanitation improvements, wide-scale manufacture and use of antibiotics, other antimicrobial medicines and vaccines, it seemed that the battle against infections was being won for the human population (CDC, 1999). Since then, however, and in addition to increasing antimicrobial resistance among bacterial pathogens, there has been an increase in the emergence of zoonotic diseases, sometimes causing fatal outbreaks of epidemic proportions (Cunningham et al., 2017).

The COVID-19 pandemic highlighted the unpreparedness of healthcare and governments to face such a pandemic (Alexandru, 2020). The pandemic risk remains high because of the increased global travel, urbanisation, changes in land use, greater exploitation of the natural environment and climate change (Madhav et al., 2017; Johnson et al., 2020; Jones et al., 2008). There is therefore an increasing need for better tools to help us manage such events.

This thesis provides results that are original and have potential for direct applications in the relevant context. The results of Part I contribute to the continuous efforts to improve our understanding and preparedness for the next outbreak. For instance, in Chapter 2, I show that combining movement networks from different livestock species substantially impact the farm's disease risk. The different farming systems need to be considered jointly to correctly identify farms to target for control strategies, as the premises which are likely to drive the epidemic in the multi-species network differ from the ones in both of the single-species networks. These two chapters provide insightful information on the multi-species dynamic movement network, which can be crucial in the face of an outbreak. In Chapter 3, I propose a tool to help rapidly identify risky premises in the face of an outbreak, when still little is known about the pathogen involved. This measure could be rapidly deployed and be useful in the early stage of an outbreak. These results have been communicated with the Scottish Government through a Policy Brief and can therefore contribute to the design of future control strategies.

Part II of this thesis described methods that can be used during an outbreak. Chapter 4 shows how weakly connected rural areas such as the Hebrides Islands can behave differently in terms of restrictions. They have options for movement control, which can in some cases help sufficiently mitigate the risk of introduction if this is their priority. These findings constitute evidence that policy makers can use when making decision. In Chapter 5, the results constitute a quantified analysis of the impact of the change in restrictions throughout the epidemic. The framework could be easily deployed in the occurrence of an outbreak in any of the Hebrides Islands and could therefore help inform decision.

In addition, these results are useful in the sense that they also illustrate the value that such studies can have in very various settings, i.e. in livestock diseases, human diseases, in peacetime, during an outbreak. The variety of work here also illustrates the opposition between using simplified approach which can provide a rapid answer and more complex model which are more accurate but slower to produce results.

6.1 Future work

There are a number of exciting possibilities for future work for each of the chapters.

Preparedness relies on the availability of data and the existence of tools which can be rapidly deployed when needed. This means that as more work is done beforehand, chances are higher that an appropriate tool will already exist. The network analysis in Chapter 2 helps increase our knowledge of the interaction between the cattle and sheep networks as well as the consequences for disease risk. This work could be broadened by considering additional species, other settings or other layers of potentially relevant contacts (such as movements of vehicles between farms, movements of individuals, spatial layer). Similarly, the metric developed in Chapter 3 could be adapted to other contexts involving other species. Furthermore, making more precise the frame and limits within which one can be confident that the results will be robust and therefore useful is valuable. Performing large scale sensitivity analyses for a broad range of parameter values for network weights and disease models would allow one to explore further correlation between metric results and disease context, helping to clarify areas of validity. In addition, while the method has been implemented with reasonably efficient code, the focus has been on assessing the usefulness of the approaches as opposed to producing code optimised for speed and memory requirement. Computational performance could likely be improved, and further work may be required to deploy these approaches on very large or very dense

networks. Ultimately the development of a user-friendly package could help the targeted audience to make use of the tool.

In the case of COVID-19, before any vaccine was developed the control of the spread of the virus relied on NPIs such as movement restrictions and social-distancing. It is important to note that although NPIs have proven effective to control the spread of the virus, these measures have a direct impact on the local and global economy (Brodeur et al., 2020). The COVID-19 pandemic has caused great economic loss at a global scale, likened to the economic impact of World War Two (Bank, 2020). Integrating economic considerations into epidemiological studies would consequently be relevant and potentially more useful to help decision makers as policy decisions can't be based on health impact only. Scotland developed a framework for COVID-19 decision-making based on *four harms*: direct impact of COVID-19 on people's health, on health and social care, on broader way of living and society, and on the economy (Scottish Government, 2020c). In the Scottish Isles, reducing movements is a plausible option as the natural boundaries allow for better control and could be effective to mitigate introduction risk in some cases (Chapter 4). But in addition to the specificity of the disease context, as a remote rural territory, the economic landscape of the islands differs from the national picture. The economic impact of movement restrictions in such settings is potentially greater than in urban areas. Rural and isolated territories partly depend on freedom of movements for tourism or movement of seasonal workers, so hindering their connectivity could compromise their economic sustainability (Currie and Falconer, 2014). Following-up with the results from Chapter 4 showing that controlling movements can be effective in preventing introduction, integrating economic aspects in terms of DALYs, impact of potential labour shortage or loss in GDP would be particularly interesting.

ABC is a powerful tool to make inferences with complex models, enabling parameter fitting when methods based on likelihood are intractable (Beaumont et al., 2002). Chapter 5 shows an illustration of this method applied to an outbreak localised in Barra. For the sake of simplification, I considered the whole island as one node within which contacts were driven by an age-mixing matrix. A natural next step could be to detail the model in smaller units, such as the datazones as used in Chapter 4, and adjust an additional parameter driving the transmission between datazones. This could inform us on the role of local transmission vs transmission driven by longer movements. In addition, the reported proportion of cases was fitted as a constant but could be fitted as a varying parameter over time as it is likely to vary according to testing efforts and change of behaviour. Since the epidemic has then spread from Barra to other isles of the Outer Hebrides, it would be interesting to integrate all of these islands in the model to understand the spread within and between islands. Finally, adapt-

ing the model for potential future outbreak, i.e. considering vaccination coverage and the more recent COVID-19 variants in the model would be useful (UK Health Security Agency, 2021).

6.2 Animal vs Human disease modelling

During this thesis, I have been lucky to have the opportunity to work on animal and human disease modelling in Scotland. These varied experiences gave me the opportunity to experience the differences between the two contexts from a researcher's point of view. I found it interesting *a posteriori* to relate the differences and similarities that struck me through my disease modelling work.

6.2.1 Data on contact patterns

Contact patterns play a central role in the model used for disease spread. Identifying and obtaining the data on contact structure between the modeled units (e.g., individuals or farms) can be challenging. In the UK, movement data of farm animals are now precisely recorded over time and centralised in a database by the government (British Cattle Movement Service, 2021). When these data are available, they are extremely useful for setting up a disease model at the country scale. Using the CTS and ScotEID data helped me grasp the richness that these data constitute thanks to their volume and level of details. Being able to track the movement of the 6 million cattle in the UK from birth to death over time and space is highly valuable. For animal diseases, other contact related parameters such as effective contact rate have sometimes been estimated in experimental studies which can provide good estimates and are useful for models (Alexandersen et al., 2003; De La Garza, 2010).

Because humans can normally move freely, there is more heterogeneity in their contact patterns compared to livestock populations that are mostly restricted to their farms. Social contacts survey, such as CoMix have been conducted in the past years making better data available for human contacts. But as shown by the results of this survey, the data are varying over time and need constantly updating (Jarvis et al., 2020a). Mobile phone data have been increasingly used in recent years to inform on human movements. But tracking human mobility using data from mobile phones raises ethical questions as these data can be disclosive and offer unprecedented power to surveil and control unwanted population movement (Taylor, 2016). These data usually contain a subset of the population and are not necessarily easy to generalise. They can nonetheless be useful to observe trend in changes, when new restrictions are put in place (Banks et al., 2020). In Chapter 4, I used movement data provided by transport companies and census flow data. Data from the transport companies offer a good level of precision but

are not exhaustive as people may use alternative ways to travel between islands for example. The census flow data, which were last collected in 2011, contains workplace and home location but disregards any other type of movement (students to university, children to school, leisure, etc.). This still provides a good idea of the general pattern of movements in the country, but the precision is nowhere near that available for livestock.

6.2.2 Disease characteristics and disease parameters

To study animal disease, experiments involving experimental inoculation can be used to measure relevant key disease parameters critical for disease models (Cox and Barnett, 2009; Guinat et al., 2016, 2014). When working on FMD-like disease there is a huge depth of studies available looking at disease models and parameters in various settings or contexts (Orsel et al., 2009; Garner and Beckett, 2005; Sanson et al., 2006; Bradhurst et al., 2021). In human outbreak, parameters have to be deduced from data available from the field as experimental inoculation can rarely be performed since it raises ethical questions (Roestenberg et al., 2018). Early in the COVID-19 outbreak parameters from previous outbreak or from similar pathogens have been used instead (Ferguson et al., 2020). In my modelling study for COVID-19 in Chapter 4 and Chapter 5, assumptions had to be made as some important information remains unknown. For example, the proportion of asymptomatic is difficult to assess from real world data. The spread of COVID-19 in closed populations like on the Diamond Princess cruise ship offered a great opportunity to estimate some parameters although there remained large uncertainty as the population was not representative of the general population (Mizumoto et al., 2020). As the COVID-19 outbreak progressed the number of studies estimating parameters grew, but using parameters from other studies can be difficult as the model framework or the context might be too different.

6.2.3 Data for modelling

For regulated animal diseases, centralised database may exist recording the animal or herd status over time. In the UK, this is the case for bTB or BVD (SAM, ScotEID databases) for example. In this case, there already exists a stream to collect, record, and make such data available for modelling studies (Tinsley et al., 2012; Brooks-Pollock et al., 2014). This is extremely valuable as, as illustrated by the COVID-19 outbreak creating these streams during an outbreak can be challenging, time-consuming and issues may be encountered potentially inducing data loss at crucial moments (Fetzer and Graeber, 2020).

In the COVID-19 pandemic, models were initially fitted to hospitalised data as tests were not widely performed. However, the definition of cases to consider between hospitalised due to COVID-19 or hospitalised with COVID-19 was sometimes unclear. Once hospitalisation became less relevant and tests widely available for the population, data on the number of positive tests could be used to fit models as seen in Chapter 5. Nevertheless, as opposed to animal settings where the design of sampling collection can be controlled, there are often large uncertainties in how and when individual will get tested, making the estimation of total number of infected difficult (Colman et al., 2021). The proportion of detected cases is likely inconsistent over time as the reason for tests can be influenced by local circumstances, changes in policies or changes in behaviour as illustrated in Chapter 5. The data used for the outbreak on Barra contained a 'reason for test' variable exists in the test result data for COVID-19 in Scotland, but the field contained mostly missing values and couldn't be exploited.

6.2.4 Control measures and management

Although the general objective is comparable (i.e. controlling the outbreak), the motivations and measures available can differ. Livestock diseases are prioritised according to their impact on animal and public health, food safety, food security, biodiversity and socioeconomic impact (OIE, 2014; F Gary, 2014). Objectives might differ depending on the disease and the country. For an FMD outbreak in the UK for example, the objective will be to eradicate the disease as quickly as possible and regain disease-free status to minimise overall cost (DEFRA, 2011). The disease-free status being necessary for international trade in Europe, losing it is hugely disruptive for the country economy (James and Rushton, 2002). The eradication is achieved by movement ban, depopulation of the affected herds followed by cleaning and disinfection, surveillance of neighboring herds, and tracing of contacts (DEFRA, 2011). Such an approach is controversial because of the ethical question around large scale animal slaughter and its impact on farmers' mental health, when other control means are available (Woods, 2013). Vaccination may be an option when a vaccine is available, but is often disregarded as it would induce a delay in retrieving disease free status (Backer et al., 2012).

In human diseases, the approach is different because the outcome on each individual's life matters and the range of restrictions available is limited by stricter ethical considerations. Interestingly during the COVID-19 pandemic, measures implemented had an unprecedented severity and scale (Zhang et al., 2020). Measures that were previously unthinkable were rapidly used to control the outbreak, i.e. movement ban, closure of schools, stay-at-home order, etc. These unpreced-

ented measures have however a long term impact on education, mental health and other health problems (Yunus et al., 2020).

6.2.5 Preparedness

Multiple animal outbreaks or outbreak threats have arisen in the past in the UK. FMD, Avian Influenza, BTV, ASF are a few examples which could explain why the veterinary sector is somehow prepared and organised to face an outbreak. Frameworks for models are available for these diseases and relevant data are recorded and can be made available in case needed (Bessell et al., 2016; Mohr et al., 2018; Nickbakhsh et al., 2011). To improve disease management and preparedness, EPIC, the Centre of Expertise for Animal Outbreaks in Scotland has been involved in disease exercises, held at UK or Scottish levels, which are designed to test the response of government and other stakeholders to outbreaks of exotic notifiable diseases. EPIC scientists have participated in disease exercises on Foot and Mouth Disease, Avian Influenza and Classical Swine Fever for example (EPIC, 2019). The COVID-19 pandemic highlighted the general lack of preparedness of many countries to manage a widespread epidemic in human populations (Fisher and Wilder-Smith, 2020). A number of European countries had seen their hospital bed capacity decrease in the past decades (Kroneman and Siegers, 2004), whereas a sufficient capacity is one of the criteria for hospital preparedness for epidemics according to WHO (WHO, 2014). Although an existing model of influenza virus spread could be used in the UK to advise the authorities (Ferguson et al., 2020), it wasn't necessarily the case in other countries in which case scientists had to build these models within a very short time at the expense of rigor and validation.

Interestingly, despite all these differences, from a research point of view working on livestock or human diseases did not feel any different as the intellectual challenge remained very similar.

6.3 Complexity vs timeliness

In applied research there is a will for making research results useful and for translating them into concrete actions (Hedrick et al., 1993). In the context of disease control at a population scale, the best way to make results valuable is to translate them into policies (Moyses Szklo, 2013). This is called *translational epidemiology* and it is the effective transfer of new knowledge from epidemiological studies into the planning of population-wide and individual-level disease control programs and policies. In the aftermath of the United Kingdom BSE outbreak, the Scientific Advisory Group for Emergencies was created in the UK.

Nevertheless, there are a variety of reasons why the link between much of research and policy isn't straightforward. This might be because often research is not designed to be relevant to policy or if it is so designed, it might fail to have an impact because of problems associated with presentation, manner of communication or because policy-makers do not see research findings as central to their decision-making (Diane Stone, 2001). This might be the case in epidemiology when most studies focus on one or a few disease related metrics to measure impact, e.g. number of cases, number of hospitalised or number of deaths, whereas the consequences of the disease as well as the restrictions have wider impacts on health, social, economic or political aspects. Research tends to reduce the complexity of real life problems to sub-questions that are studied separately, whilst policy makers need an integrated problem solution, ready-to-use in policy (Jansen et al., 2010). There is however an important distinction between policy making which is the making or formulation of a particular plan or course of action by the government or an organisation; as opposed to decision making which refers to the act or process of selecting a particular plan or course of action from a set of alternatives. Decision making is shorter term compared to policy making. In a situation where circumstances are changing fast such as the start of an outbreak, if model outputs are available, they can easily be used to inform decisions. Yet, even in an ideal case where the needed result or knowledge is available, practical constraints on rational decision-making will always exist. This concept was defined by Herbert Simon who developed a model of the policy process premised on the notions of *bounded rationality* and *satisficing*. Bounded rationality is the idea that rationality is limited when individuals make decisions. Limitations include the difficulty of the problem requiring a decision, the cognitive capability of the mind, and the time available to make the decision. Along the same lines, the term *satisficing* comes from the merging between the words satisfying and sufficing, translating the fact that individuals would rather accept a good enough option than optimal. Decision-makers, accepting the limits of their situation, choose compromise policies that satisfy (rather than maximise) organisational goals, and which are acceptable in the face of competing demands (Simon et al., 1984).

Timeliness and trust in the results are also major challenges for the translation of research results into decisions or policies. This appears as one of the most difficult hurdle to tackle in the occurrence of an outbreak: policy makers need nearly instantaneous and trustworthy results to inform their decisions on urgent matter, whilst the production of research results is constrained by the limited knowledge of the disease context at that point in time. Models to forecast potential future events or to predict events that could occur under certain circumstances or interventions are potentially most relevant to policies. This use

of models is challenging as it has to be done with what is available at the time, to get the best answer possible, which will inevitably be affected by large uncertainties. Communicating this uncertainty properly is crucial because if models are used to inform decisions, incorrect predictions can be harmful, giving model use an important ethical dimension. For emergent epidemics, the lack of data in addition to the necessity of implementing control measures early to successfully control the epidemic makes the use of process models particularly challenging. To get ahead of the curve extensive work is enormously beneficial prior to the start of the epidemic. The work presented in this thesis contributes to this long-term effort. But even though huge efforts can be made to collect data and create models beforehand, those are rarely rewarded as diseases rarely conform to our expectations.

Despite all the challenges, huge progress has been made in translational epidemiology (Khoury et al., 2010). This evolution is still ongoing as there are now more and more examples of initiatives that aim to develop stronger relationships between decision-makers and researchers and this is again enhanced by the current COVID-19 pandemic (van Schalkwyk and McKee, 2021). EPIC was created following the 2001 FMD epidemic to strengthen relationships between researchers in infectious disease epidemiology from various institutions in Scotland and The Animal Health and Welfare Division of the Scottish Government. This is a good example of an initiative designed to help build stronger and long-lasting relationships between academics and the Scottish Government. Such initiatives allow us to build person-to-person relationships improving the chances that communication will be effective when needed.

Process modelling remains the best tool to help gain insight in possible future events when dealing with non-linear, spatially heterogeneous and complex phenomena. It is a powerful approach to learn from data beyond what can be observed and to make predictions for the future when there is no data to guide us. Although there are always large uncertainties around predictions, it helps quantify uncertainty around potential scenarios. Past the initial stage when models need to be built and adjusted with sparse data, models output can then be integrated in the decision process. This is illustrated in the UK with $SPI-M$, which was initially developed to provide modelling summary for pandemic influenza. $SPI-M$ aims at giving expert advice to the Department of Health and Social Care and wider UK government on scientific matters relating to the UK's response to an influenza pandemic (or other emerging human infectious disease threats) (UK Government). Since the start of the pandemic, researchers of $SPI-M$ have been modelling scenarios for COVID-19 and communicating results to the government to support decision making (SAGE, 2021). Inevitably, policy decision-making must integrate different sources and types of inform-

ation across more than just epidemiology, even in a pandemic. The ability of process models to give some insight into future directions of the disease itself therefore is an invaluable underpinning of all those decisions, even if alone, it is rarely the only determinant of decisions.

6.4 Conclusion

In conclusion, among the realistic ingredients to be considered in the computational modeling of infectious diseases, mobility and movement represent a crucial piece. The increasing collection and availability of mobility and movement data have clearly helped to push forward the modelling techniques in the past decades. This helped to move from simple models assuming homogeneous mixing to network and metapopulation models where contacts are detailed. Disease modelling in the livestock sector has contributed to enhance the research on infectious disease spread on networks, thanks to the availability of very detailed animal movement networks as illustrated in Chapter 2 and Chapter 3. Chapter 4 showed how multiple sources of human mobility data can be combined to represent global movement patterns relevant to disease spread. In addition, greater source of mobility data have become available through the digital revolution. Since the start of the COVID-19 pandemic mobility data and contacts from survey have been central. The COVID-19 pandemic also enhanced the access to data for researchers through the initial global effort. Another crucial advance in modelling techniques is the development of tools allowing to make more from sparse data. Bayesian approaches such as model fitting and inference for infectious disease dynamics give opportunities to fit ever more complex mechanistic models to data. These powerful methods are valuable to improve our understanding of complex phenomena. The estimation of key parameter values helps quantify the importance of certain transmission routes Brooks-Pollock et al. (2014); O'Hare et al. (2014) or the effectiveness of control measures as seen in Chapter 5. In a world where more disease challenges are expected in the future, process models integrating more detailed movement or mobility data for infectious disease spread will play an increasingly important role in infectious disease management and should be better integrated into decision and policy processes.

References

Simon John More, Conor G McAloon, John M Griffin, Miriam Casey, Ann Barber, Elizabeth Lane, Andrew W Byrne, Áine B Collins, David McEvoy, Kevin Hunt, et al. Covid-19 epidemiological parameters summary document. Technical report, Department of Health, 2020. (Cited on pages xx, 79, and 91.)

Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, et al. Inferred duration of infectious period of sars-cov-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic covid-19 cases. *BMJ open*, 10(8):e039856, 2020. (Cited on pages xx, 79, and 91.)

Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020. (Cited on pages xx, 79, and 91.)

Ole Jørgen Benedictow and Ole L Benedictow. *The Black Death, 1346-1353: the complete history*. Boydell & Brewer, 2004. (Cited on page 1.)

Samuel Cohn. After the black death: labour legislation and attitudes towards labour in late-medieval western europe. *The Economic History Review*, 60(3): 457–485, 2007. (Cited on page 1.)

Kristine B Patterson and Thomas Runge. Smallpox and the native american. *The American journal of the medical sciences*, 323(4):216–222, 2002. (Cited on page 1.)

John Snow. *On the mode of communication of cholera*. John Churchill, 1855. (Cited on page 1.)

Dominik Hünniger. The "normative forces" of difference: Ecology, economy and society during cattle plagues in the eighteenth century. *Journal for the History of Environment and Society*, 5:91–100, 2020. (Cited on page 1.)

John Broad. Cattle plague in eighteenth-century england. *The Agricultural history review*, 31(2):104–115, 1983. (Cited on page 1.)

CDC. Achievements in public health, 1900-1999: Control of infectious diseases, 1999. URL https://www.cdc.gov/mmwr/preview/mmwrhtml/mm4829a1.htm. (Cited on pages 2 and 103.)

Philip S Brachman. Infectious diseases—past, present, and future, 2003. (Cited on page 2.)

Donald A Henderson. The eradication of smallpox–an overview of the past, present, and future. *Vaccine*, 29:D7–D9, 2011. (Cited on page 2.)

Peter L Roeder. Rinderpest: the end of cattle plague. *Preventive veterinary medicine*, 102(2):98–106, 2011. (Cited on page 2.)

Theo Vos, Stephen S Lim, Cristiana Abbafati, Kaja M Abbas, Mohammad Abbasi, Mitra Abbasifard, Mohsen Abbasi-Kangevari, Hedayat Abbastabar, Foad Abd-Allah, Ahmed Abdelalim, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258):1204–1222, 2020. (Cited on page 2.)

Louise H Taylor, Sophia M Latham, and Mark EJ Woolhouse. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1411):983–989, 2001. (Cited on pages 2, 15, and 26.)

Emilie Alirol, Laurent Getaz, Beat Stoll, François Chappuis, and Louis Loutan. Urbanisation and infectious diseases in a globalised world. *The Lancet infectious diseases*, 11(2):131–141, 2011. (Cited on page 2.)

Tove H Jorgensen, Henriette N Buttenschön, August G Wang, Thomas D Als, Anders D Børglum, and Henrik Ewald. The origin of the isolated population of the faroe islands investigated using y chromosomal markers. *Human Genetics*, 115(1):19–28, 2004. (Cited on page 2.)

Mitchell L Cohen. Changing patterns of infectious disease. *Nature*, 406(6797): 762–767, 2000. (Cited on page 2.)

Leslie A Reperant and Albert DME Osterhaus. Aids, avian flu, sars, mers, ebola, zika... what next? *Vaccine*, 35(35):4470–4474, 2017. (Cited on pages 2 and 60.)

Leslie A Reperant, Giuseppe Cornaglia, and Albert DME Osterhaus. The importance of understanding the human–animal interface. *One Health: The Human-Animal-Environment Interfaces in Emerging Infectious Diseases*, pages 49–81, 2012. (Cited on page 2.)

Klaus F Zimmermann, Gokhan Karabulut, Mehmet Huseyin Bilgin, and Asli Cansin Doker. Inter-country distancing, globalisation and the coronavirus pandemic. *The World Economy*, 43(6):1484–1498, 2020. (Cited on page 2.)

Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3):278, 2020. (Cited on page 2.)

Richard C Dicker, Fatima Coronado, Denise Koo, and R Gibson Parrish. Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics. 2006. (Cited on page 2.)

Rebecca Mancy, Patrick M Brock, and Rowland R Kao. An integrated framework for process-driven model construction in disease ecology and animal health. *Frontiers in veterinary science*, 4:155, 2017. (Cited on page 3.)

Nicolas Bacaër. Daniel bernoulli, d'alembert and the inoculation of smallpox (1760). In *A short history of mathematical population dynamics*, pages 21–30. Springer, 2011. (Cited on page 3.)

William Heaton Hamer. *The milroy lectures on epidemic diseases in england: The evidence of variability and of persistency of type; delivered before the royal college of physicians of london, march 1st, 6th, and 8th, 1906*. Bedford Press, 1906. (Cited on page 3.)

Ronald Ross. *The prevention of malaria*. John Murray, 1911. (Cited on page 3.)

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927. (Cited on page 3.)

Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. OUP oxford, 2010. (Cited on pages 4 and 18.)

Helen Abbey. An examination of the reed-frost theory of epidemics. *Human biology*, 24(3):201, 1952. (Cited on page 5.)

Matthew J Keeling. Modelling the persistence of measles. *Trends in microbiology*, 5(12):513–518, 1997. (Cited on page 6.)

Neil M Ferguson, Christl A Donnelly, and Roy M Anderson. Transmission intensity and impact of control policies on the foot and mouth epidemic in great britain. *Nature*, 413(6855):542–548, 2001a. (Cited on page 6.)

Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990. (Cited on pages 6 and 7.)

Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47):873–885, 2010. (Cited on pages 7 and 92.)

M Ramsay, N Gay, Elizabeth Miller, M Rush, J White, P Morgan-Capner, and D Brown. The epidemiology of measles in england and wales: rationale for the 1994 national vaccination campaign. *Communicable disease report. CDR review*, 4(12):R141–6, 1994. (Cited on page 8.)

David Greenhalgh. Vaccination campaigns for common childhood diseases. *Mathematical biosciences*, 100(2):201–240, 1990. (Cited on page 8.)

Marc Lipsitch, Ted Cohen, Ben Cooper, James M Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H Samore, et al. Transmission dynamics and control of severe acute respiratory syndrome. *science*, 300(5627):1966–1970, 2003. (Cited on pages 8 and 17.)

Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10):936–944, 2006. (Cited on page 8.)

Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008. (Cited on page 8.)

M Brisson, WJ Edmunds, NJ Gay, B Law, and G De Serres. Modelling the impact of immunization on the epidemiology of varicella zoster virus. *Epidemiology & Infection*, 125(3):651–669, 2000. (Cited on page 9.)

Thembinkosi Mkhatshwa and Anna Mummert. Modeling super-spreading events for infectious diseases: case study sars. *arXiv preprint arXiv:1007.0908*, 2010. (Cited on page 9.)

Kin On Kwok, Gabriel M Leung, Wai Yee Lam, and Steven Riley. Using models to identify routes of nosocomial infection: a large hospital outbreak of sars in hong kong. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610): 611–618, 2007. (Cited on page 9.)

Richard A Stein. Super-spreaders in infectious diseases. *International Journal of Infectious Diseases*, 15(8):e510–e513, 2011. (Cited on page 9.)

Leonhard Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, pages 128–140, 1741. (Cited on page 9.)

John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Macmillan London, 1976. (Cited on page 9.)

Alden S Klovdahl. Social networks and the spread of infectious diseases: the aids example. *Social science & medicine*, 21(11):1203–1216, 1985. (Cited on page 9.)

Robert M May and Roy M Anderson. Commentary transmission dynamics of hiv infection. *Nature*, 326(137):10–1038, 1987. (Cited on page 9.)

Gautier Krings, Francesco Calabrese, Carlo Ratti, and Vincent D Blondel. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):L07003, 2009. (Cited on page 9.)

Yingcun Xia, Ottar N Bjørnstad, and Bryan T Grenfell. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281, 2004. (Cited on page 9.)

Chao-lin Gu and Hai-feng Pang. Study on spatial relations of chinese urban system: Gravity model approach. *Geographical Research*, 27(1):1–12, 2008. (Cited on page 9.)

C. Dubé, C. Ribble, D. Kelton, and B. McNab. A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transboundary and emerging diseases*, 56(3):73–85, 2009. (Cited on page 9.)

Eric M. Fèvre, Barend M. de C. Bronsvoort, Katie A. Hamilton, and Sarah Cleaveland. Animal movements and the spread of infectious diseases. *Trends in Microbiology*, 14(3):125–131, March 2006. ISSN 0966-842X. doi: 10.1016/j.tim.2006.01.004. URL http://www.sciencedirect.com/science/article/pii/S0966842X06000175. (Cited on pages 10 and 26.)

D. M. Green, I. Z. Kiss, and R. R. Kao. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1602):2729–2735, 2006a. (Cited on page 10.)

Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994a. (Cited on pages 10 and 32.)

Kao R.R, Danon L, Green D.M, and Kiss I.Z. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597):1999–2007, August 2006. doi: 10.1098/rspb.2006.3505. URL https://royalsocietypublishing.org/doi/full/10.1098/rspb.2006.3505. (Cited on page 10.)

Istvan Z Kiss, Darren M Green, and Rowland R Kao. *Journal of the Royal Society Interface*, 3(10):669–677, 2006a. (Cited on pages 10, 26, and 29.)

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440, 1998. (Cited on pages 10 and 32.)

I. Z. Kiss, D. M. Green, and R. R. Kao. Infectious disease control using contact tracing in random and scale-free networks. *Journal of the Royal Society Interface*, 3(6):55–62, February 2006b. ISSN 1742-5689. doi: 10.1098/rsif.2005.0079. WOS:000235712600005. (Cited on page 10.)

Mark Newman. *Networks*. Oxford university press, 2018. (Cited on pages 10 and 27.)

Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978. (Cited on pages 10, 32, and 66.)

Matthew J Silk, Darren P Croft, Richard J Delahay, David J Hodgson, Mike Boots, Nicola Weber, and Robbie A McDonald. Using social network measures in wildlife disease ecology, epidemiology, and management. *BioScience*, 67(3): 245–257, 2017. (Cited on page 10.)

Mark Newman. Networks. Mar 2010. doi: 10.1093/acprof:oso/9780199206650.001.0001. URL http://dx.doi.org/10.1093/ACPROF:OSO/9780199206650.001.0001. (Cited on pages 10 and 32.)

Rowland R Kao. The role of mathematical modelling in the control of the 2001 fmd epidemic in the uk. *TRENDS in Microbiology*, 10(6):279–286, 2002. (Cited on pages 10 and 11.)

Rowland R Kao, Leon Danon, Darren M Green, and Istvan Z Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in great britain. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597): 1999–2007, 2006. (Cited on pages 10, 26, and 29.)

DM Green, IZ Kiss, and RR Kao. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2729–2735, 2006b. (Cited on page 10.)

MDF Shirley and SP Rushton. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. *Epidemiology & Infection*, 133(6):1023–1032, 2005. (Cited on page 10.)

JC Gibbens, JW Wilesmith, CE Sharpe, LM Mansley, E Michalopoulou, JBM Ryan, and M Hudson. Descriptive epidemiology of the 2001 foot-and-mouth disease

epidemic in great britain: the first five months. *Veterinary Record*, 149(24):729–743, 2001. (Cited on page 10.)

R. M. Christley, S. E. Robinson, R. Lysons, and N. P. French. Network analysis of cattle movement in Great Britain. *Proc. Soc. Vet. Epidemiol. Prev. Med*, pages 234–243, 2005. (Cited on pages 11 and 26.)

SE Robinson and RM Christley. Exploring the role of auction markets in cattle movements within great britain. *Preventive veterinary medicine*, 81(1-3):21–37, 2007a. (Cited on page 11.)

Sen Pei, Lev Muchnik, José S Andrade Jr, Zhiming Zheng, and Hernán A Makse. Searching for superspreaders of information in real-world social media. *Scientific reports*, 4(1):1–12, 2014. (Cited on page 11.)

Doina Bucur and Petter Holme. Beyond ranking nodes: Predicting epidemic outbreak sizes by network centralities. *arXiv preprint arXiv:1909.10021*, 2019. (Cited on pages 11 and 31.)

Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical patterns of cattle trade movements. *PloS one*, 6(5):e19869, 2011a. (Cited on pages 11, 27, and 52.)

Bhagat Lal Dutta, Pauline Ezanno, and Elisabeta Vergu. Characteristics of the spatio-temporal network of cattle movements in france over a 5-year period. *Preventive veterinary medicine*, 117(1):79–94, 2014. (Cited on page 11.)

Wei Duan, Zongchen Fan, Peng Zhang, Gang Guo, and Xiaogang Qiu. Mathematical and computational approaches to epidemic modeling: a comprehensive review. *Frontiers of Computer Science*, 9(5):806–826, 2015. (Cited on page 12.)

Paolo Bajardi, Alain Barrat, Lara Savini, and Vittoria Colizza. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of the Royal Society Interface*, 9(76):2814–2825, 2012. (Cited on page 12.)

Ilkka Hanski and Michael Gilpin. Metapopulation dynamics: brief history and conceptual domain. *Biological journal of the Linnean Society*, 42(1-2):3–16, 1991. (Cited on page 12.)

Bryan Grenfell and John Harwood. (meta) population dynamics of infectious diseases. *Trends in ecology & evolution*, 12(10):395–399, 1997. (Cited on page 12.)

Gaël Beaunée, Elisabeta Vergu, and Pauline Ezanno. Modelling of paratuberculosis spread between dairy cattle farms at a regional scale. *Veterinary research*, 46(1):1–13, 2015. (Cited on page 12.)

Ellen Brooks-Pollock, Gareth O Roberts, and Matt J Keeling. A dynamic model of bovine tuberculosis spread and control in great britain. *Nature*, 511(7508): 228–231, 2014. (Cited on pages 12, 26, 107, and 112.)

Bryan Iotti, Eugenio Valdano, Lara Savini, Luca Candeloro, Armando Giovannini, Sergio Rosati, Vittoria Colizza, and Mario Giacobini. Farm productive contexts and the dynamics of bovine viral diarrhea (bvd) transmission. *Preventive veterinary medicine*, 165:23–33, 2019. (Cited on page 13.)

Matt J Keeling and Chris A Gilligan. Metapopulation dynamics of bubonic plague. *Nature*, 407(6806):903–906, 2000. (Cited on page 13.)

Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical review letters*, 99(14):148701, 2007. (Cited on pages 14 and 16.)

Daniela De Angelis, Anne M Presanis, Paul J Birrell, Gianpaolo Scalia Tomba, and Thomas House. Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, 10:83–87, 2015. (Cited on page 14.)

CDC. Principles of epidemiology in public health practice, third edition; an introduction to applied epidemiology and biostatistics, 2012. URL https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section11.html. (Cited on page 14.)

Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006. (Cited on page 14.)

M Elizabeth Halloran, Neil M Ferguson, Stephen Eubank, Ira M Longini, Derek AT Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy C Germann, et al. Modeling targeted layered containment of an influenza pandemic in the united states. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644, 2008. (Cited on page 14.)

Neil Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College London*, 10(77482):491–497, 2020. (Cited on pages 14, 17, 83, 107, and 109.)

Ioanna P Chatziprodromidou, Malamatenia Arvanitidou, Javier Guitian, Thomas Apostolou, George Vantarakis, and Apostolos Vantarakis. Global avian influenza outbreaks 2010–2016: a systematic review of their distribution,

avian species and virus subtype. *Systematic reviews*, 7(1):1–12, 2018. (Cited on page 14.)

Jennifer E Dent, Rowland R Kao, Istvan Z Kiss, Kieran Hyder, and Mark Arnold. Contact structures in the poultry industry in great britain: exploring transmission routes for a potential avian influenza virus epidemic. *BMC Veterinary Research*, 4(1):1–14, 2008. (Cited on page 15.)

N Moyen, G Ahmed, S Gupta, T Tenzin, R Khan, T Khan, N Debnath, M Yamage, DU Pfeiffer, and G Fournie. A large-scale study of a poultry trading network in bangladesh: implications for control and surveillance of avian influenza viruses. *BMC veterinary research*, 14(1):1–12, 2018. (Cited on page 15.)

Guillaume Fournié, Javier Guitian, Stéphanie Desvaux, Vu Chi Cuong, Dirk Udo Pfeiffer, Punam Mangtani, Azra C Ghani, et al. Interventions for avian influenza a (h5n1) risk management in live bird market networks. *Proceedings of the National Academy of Sciences*, 110(22):9177–9182, 2013. (Cited on page 15.)

Matthew J Silk, Julian A Drewe, Richard J Delahay, Nicola Weber, Lucy C Steward, Jared Wilson-Aggarwal, Mike Boots, David J Hodgson, Darren P Croft, and Robbie A McDonald. Quantifying direct and indirect contacts for the potential transmission of infection between species using a multilayer contact network. *Behaviour*, 155(7-9):731–757, 2018. (Cited on page 15.)

Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza. Predicting epidemic risk from past temporal contact data. *PLoS Comput Biol*, 11(3):e1004152, 2015. (Cited on pages 15 and 27.)

Petter Holme. Network reachability of real-world contact sequences. *Physical Review E*, 71(4):046119, 2005. (Cited on pages 15, 33, and 51.)

C Dubé, C Ribble, D Kelton, and B McNab. Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in ontario, canada. *Transboundary and emerging diseases*, 55(9-10):382–392, 2008. (Cited on pages 15 and 33.)

Mario Konschake, Hartmut H. K. Lentz, Franz J. Conraths, Philipp Hövel, and Thomas Selhorst. On the Robustness of In- and Out-Components in a Temporal Network. *PLOS ONE*, 8(2):e55223, February 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0055223. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0055223. (Cited on pages 15 and 33.)

Beatriz Vidondo and Bernhard Voelkl. Dynamic network measures reveal the impact of cattle markets and alpine summering on the risk of epidemic outbreaks in the swiss cattle population. *BMC veterinary research*, 14(1):88, 2018. (Cited on pages 15, 27, 41, and 51.)

Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):1–12, 2009. (Cited on page 16.)

Chiara Poletto, MF Gomes, A Pastore y Piontti, Luca Rossi, Livio Bioglio, Dennis L Chao, IM Longini Jr, M Elizabeth Halloran, Vittoria Colizza, and Alessandro Vespignani. Assessing the impact of travel restrictions on international spread of the 2014 west african ebola epidemic. *Eurosurveillance*, 19(42):20936, 2014a. (Cited on page 16.)

World Health Organization. Un senior leaders outline needs for global ebola response, 2014. URL https://www.who.int/news/item/03-09-2014-un-senior-leaders-outline-needs-for-global-ebola-response. (Cited on page 16.)

African Union. African union's executive council urges lifting of travel restrictions related to ebola outbreak, 2014. URL https://au.int/en/pressreleases/20140908-1. (Cited on page 16.)

FAO. Ebola threatens food security in west africa: Fao, 2014. URL https://www.reuters.com/article/us-health-ebola-food-idUSKBN0GX0HB20140902. (Cited on page 16.)

T Déirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature medicine*, 12(5):497–499, 2006. (Cited on page 16.)

Ben S Cooper, Richard J Pitman, W John Edmunds, and Nigel J Gay. Delaying the international spread of pandemic influenza. *PLoS medicine*, 3(6):e212, 2006. (Cited on page 16.)

Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1), 2007a. (Cited on pages 16 and 26.)

Gianpaolo Scalia Tomba and Jacco Wallinga. A simple explanation for the low impact of border control as a countermeasure to the spread of an infectious disease. *Mathematical biosciences*, 214(1-2):70–72, 2008. (Cited on page 16.)

Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic. *PloS one*, 6(1):e16591, 2011b. (Cited on page 16.)

Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007b. (Cited on page 17.)

Francesco Pinotti, Laura Di Domenico, Ernesto Ortega, Marco Mancastroppa, Giulia Pullano, Eugenio Valdano, Pierre-Yves Boëlle, Chiara Poletto, and Vittoria Colizza. Tracing and analysis of 288 early sars-cov-2 infections outside china: A modeling study. *PLoS medicine*, 17(7):e1003193, 2020. (Cited on page 17.)

Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400, 2020. (Cited on page 17.)

Samuel Clifford, Carl AB Pearson, Petra Klepac, Kevin Van Zandvoort, Billy J Quilty, CMMID COVID-19 working group, Rosalind M Eggo, and Stefan Flasche. Effectiveness of interventions targeting air travellers for delaying local outbreaks of sars-cov-2. *Journal of travel medicine*, 27(5):taaa068, 2020. (Cited on page 17.)

Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497, 2020. (Cited on page 17.)

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020. (Cited on page 17.)

Jayson S Jia, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis. Population flow drives spatio-temporal distribution of covid-19 in china. *Nature*, 582(7812):389–394, 2020. (Cited on page 17.)

PE Christensen, HENNING Schmidt, HO Bang, Vera Andersen, BJARNE Jordal, OSKAR Jensen, et al. An epidemic of measles in southern greenland, 1951. measles in virgin soil. ii. the epid3mic proper. *Acta Medica Scandinavica*, 144(6): 430–49, 1953. (Cited on page 17.)

BN Philip, Karl R Reinhard, and DB Lackman. Observations on a mumps epidemic in a "virgin" population. *American Journal of Epidemiology*, 69(2):91–111, 1959. (Cited on page 17.)

Nuria Oliver, Bruno Lepri, Harald Sterly, Renaud Lambiotte, Sébastien Deletaille, Marco De Nadai, Emmanuel Letouzé, Albert Ali Salah, Richard Benjamins, Ciro Cattuto, et al. Mobile phone data for informing public health actions across the covid-19 pandemic life cycle, 2020. (Cited on page 18.)

Giulia Pullano, Eugenio Valdano, Nicola Scarpa, Stefania Rubrichi, and Vittoria Colizza. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the covid-19 epidemic in france under lockdown: a population-based study. *The Lancet Digital Health*, 2(12):e638–e649, 2020. (Cited on page 18.)

Amy Gimma, James D Munday, Kerry LM Wong, Pietro Coletti, Kevin van Zandvoort, Kiesha Prem, Petra Klepac, G James Rubin, Sebastian Funk, W John Edmunds, et al. Comix: Changes in social contacts as measured by the contact survey during the covid-19 pandemic in england between march 2020 and march 2021. *medRxiv*, 2021. (Cited on page 18.)

Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002. (Cited on pages 19, 84, and 105.)

Scott A Sisson, Yanan Fan, and Mark Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018. (Cited on page 19.)

Christopher J Banks, Ewan Colman, Thomas Doherty, Oliver Tearne, Mark E Arnold, Katherine Elizabeth Atkins, Daniel Balaz, Gaël Beaunée, Paul Bessell, Jessica Enright, et al. Disentangling the roles of human mobility and deprivation on the transmission dynamics of covid-19 using a spatially explicit simulation model. *medRxiv*, 2020. (Cited on pages 19, 66, and 106.)

Stephen S Morse. Factors in the emergence of infectious diseases. In *Plagues and politics*, pages 8–26. Springer, 2001. (Cited on page 26.)

Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, 2008. (Cited on pages 26 and 103.)

Nita Madhav, Ben Oppenheim, Mark Gallivan, Prime Mulembakani, Edward Rubin, and Nathan Wolfe. Pandemics: risks, impacts, and mitigation. In *Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition*. The International Bank for Reconstruction and Development/The World Bank, 2017. (Cited on pages 26 and 103.)

Chris Dye and Nigel Gay. Modeling the sars epidemic. *Science*, 300(5627):1884–1885, 2003. (Cited on page 26.)

Justin Lessler, Isabel Rodriguez-Barraquer, Derek AT Cummings, Tini Garske, Maria Van Kerkhove, Harriet Mills, Shaun Truelove, Rafat Hakeem, Ali Albarrak, Neil M Ferguson, et al. Estimating potential incidence of mers-cov associated with hajj pilgrims to saudi arabia, 2014. *PLoS currents*, 6, 2014. (Cited on page 26.)

Monika Boehm, Michael R Hutchings, and Piran CL White. Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine tb among badgers and cattle. *PLoS One*, 4(4), 2009. (Cited on page 26.)

Aurore Palisson, Aurelie Courcoul, and Benoit Durand. Role of cattle movements in bovine tuberculosis spread in france between 2005 and 2014. *PLoS One*, 11 (3), 2016. (Cited on page 26.)

Darren M Green, Istvan Z Kiss, Andrew P Mitchell, and Rowland R Kao. Estimates for local and movement-based transmission of bovine tuberculosis in british cattle. *Proceedings of the Royal Society B: Biological Sciences*, 275(1638): 1001–1005, 2008. (Cited on page 26.)

Mark Tinsley, Fraser I Lewis, and Franz Brülisauer. Network modeling of bvd transmission. *Veterinary research*, 43(1):11, 2012. (Cited on pages 26, 44, and 107.)

A Ortiz-Pelaez, DU Pfeiffer, RJ Soares-Magalhaes, and FJ Guitian. 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(1-2):40–55, 2006. (Cited on page 26.)

SE Robinson and RM Christley. Exploring the role of auction markets in cattle movements within great britain. *Preventive veterinary medicine*, 81(1-3):21–37, 2007b. (Cited on page 26.)

SE Robinson, MG Everett, and RM Christley. Recent network evolution increases the potential for large epidemics in the british cattle population. *Journal of the Royal Society Interface*, 4(15):669–674, 2007. (Cited on page 26.)

Matthew C Vernon and Matt J Keeling. Representing the uk's cattle herd as static and dynamic networks. *Proceedings of the Royal Society B: Biological Sciences*, 276 (1656):469–476, 2009. (Cited on page 26.)

Victoriya V Volkova, Richard Howey, Nicholas J Savill, and Mark EJ Woolhouse. Sheep movement networks and the transmission of infectious diseases. *PloS one*, 5(6), 2010. (Cited on page 26.)

Maria Nöremark, Nina Håkansson, Susanna Sternberg Lewerin, Ann Lindberg, and Annie Jonsson. Network analysis of cattle and pig movements in sweden: measures relevant for disease control and risk based surveillance. *Preventive veterinary medicine*, 99(2-4):78–90, 2011. (Cited on page 26.)

Sibylle Mohr, Michael Deason, Mikhail Churakov, Thomas Doherty, and Rowland R. Kao. Manipulation of contact network structure and the impact on foot-and-mouth disease transmission. *Preventive veterinary medicine*, 157:8–18, 2018. (Cited on pages 26 and 109.)

Jessica Enright and Rowland Raymond Kao. Epidemics on dynamic networks. *Epidemics*, 24:88–97, 2018. (Cited on page 27.)

Hartmut HK Lentz, Andreas Koher, Philipp Hövel, Jörn Gethmann, Carola Sauter-Louis, Thomas Selhorst, and Franz J Conraths. Disease spread through animal movements: a static and temporal network analysis of pig trade in germany. *PloS one*, 11(5), 2016. (Cited on pages 27, 41, and 51.)

Gianluigi Rossi, Giulio A De Leo, Stefano Pongolini, Silvano Natalini, Luca Zarenghi, Matteo Ricchi, and Luca Bolzoni. The potential role of direct and indirect contacts on infection spread in dairy farm networks. *PLoS computational biology*, 13(1), 2017. (Cited on pages 27 and 34.)

WA Geering. Foot-and-mouth disease in sheep. *Australian Veterinary Journal*, 43 (11):485–489, 1967. (Cited on page 29.)

CF Gibson and AI Donaldson. Exposure of sheep to natural aerosols of foot-and-mouth disease virus. *Research in veterinary science*, 41(1):45–49, 1986. (Cited on page 29.)

Jens Havskov Sørensen, DKJ Mackay, CØ Jensen, and Alex I Donaldson. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. *Epidemiology & Infection*, 124(3):577–590, 2000. (Cited on page 29.)

Neil M Ferguson, Christl A Donnelly, and Roy M Anderson. The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions. *Science*, 292(5519):1155–1160, 2001b. (Cited on page 29.)

Susan E Robinson and Rob M Christley. Identifying temporal variation in reported births, deaths and movements of cattle in britain. *BMC Veterinary Research*, 2(1):11, 2006. (Cited on page 29.)

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999. (Cited on page 31.)

K. Kandhway and J. Kuri. Using node centrality and optimal control to maximize information diffusion in social networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7):1099–1110, 2017. (Cited on page 31.)

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL http://www.R-project.org/. (Cited on page 31.)

Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL http://igraph.sf.net. (Cited on page 31.)

Daniel T Haydon, Rowland R Kao, and R Paul Kitching. The uk foot-and-mouth disease outbreak—the aftermath. *Nature Reviews Microbiology*, 2(8):675–681, 2004. (Cited on page 44.)

Jessica Enright and Rowland R Kao. A fast algorithm for calculating an expected outbreak size on dynamic contagion networks. *Epidemics*, 16:56–62, 2016. (Cited on page 45.)

Matt J Keeling. Models of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences*, 272(1569):1195–1202, 2005. (Cited on page 47.)

Stefan Widgren, Pavol Bauer, Robin Eriksson, and Stefan Engblom. SimInf: An R package for data-driven stochastic disease spread simulations. *Journal of Statistical Software*, 91(12):1–42, 2019. doi: 10.18637/jss.v091.i12. (Cited on page 49.)

Jenny Frössling, Anna Ohlson, Camilla Björkman, Nina Håkansson, and Maria Nöremark. Application of network analysis parameters in risk-based surveillance–examples based on cattle trade data and bovine infections in sweden. *Preventive veterinary medicine*, 105(3):202–208, 2012. (Cited on page 51.)

Andrew Dobson. Population dynamics of pathogens with multiple host species. *the american naturalist*, 164(S5):S64–S78, 2004. (Cited on page 51.)

James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005. (Cited on page 51.)

Tanja Knific, Matjaž Ocepek, Andrej Kirbiš, and Hartmut H. K. Lentz. Implications of cattle trade for the spread and control of infectious diseases in slovenia. *Frontiers in Veterinary Science*, 6:454, 2020. ISSN 2297-1769. doi: 10.3389/fvets.2019.00454. URL https://www.frontiersin.org/article/10.3389/fvets.2019.00454. (Cited on page 51.)

Rowland R Kao, Darren M Green, Jethro Johnson, and Istvan Z Kiss. 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, 2007. (Cited on page 52.)

National Records of Scotland. Deaths involving coronavirus (covid-19)in scotland week 14 (30 march to 5 april 2020), 2020a. URL https://www.nrscotland.gov.uk/files/statistics/covid19/covid-deaths-report-week-14.pdf. (Cited on page 60.)

National Records of Scotland. Deaths involving coronavirus (covid-19)in scotland week 29 (13 to 19 july 2020), 2020b. URL https://www.nrscotland.gov.uk/files/statistics/covid19/covid-deaths-report-week-29.pdf. (Cited on page 60.)

Samantha J Lycett, Joseph Hughes, Martin P McHugh, Ana da Silva Filipe, Rebecca Dewar, Lu Lu, Thomas Doherty, Amy Shepherd, Rhys Inward, Gianluigi Rossi, et al. Epidemic waves of covid-19 in scotland: a genomic perspective on the impact of the introduction and relaxation of lockdown on sars-cov-2. *medRxiv*, 2021. (Cited on page 60.)

Sophie Cousins. New zealand eliminates covid-19. *The Lancet*, 395(10235):1474, 2020. (Cited on page 60.)

Chiara Poletto, MF Gomes, A Pastore y Piontti, Luca Rossi, Livio Bioglio, Dennis L Chao, IM Longini Jr, M Elizabeth Halloran, Vittoria Colizza, and Alessandro Vespignani. Assessing the impact of travel restrictions on international spread of the 2014 west african ebola epidemic. *Eurosurveillance*, 19(42):20936, 2014b. (Cited on page 60.)

Ana LP Mateus, Harmony E Otete, Charles R Beck, Gayle P Dolan, and Jonathan S Nguyen-Van-Tam. Effectiveness of travel restrictions in the rapid containment of human influenza: a systematic review. *Bulletin of the World Health Organization*, 92:868–880D, 2014. (Cited on page 60.)

UK data service. Census flow data, 2011. URL https://census.ukdataservice.ac.uk/get-data/flow-data. (Cited on page 62.)

Civil Aviation Authority. Uk airport data, 2020. URL https://www.caa.co.uk/Data-and-analysis/UK-aviation-market/Airports/Datasets/UK-airport-data/. (Cited on page 62.)

ISSN International Centre. Coronavirus (covid-19) phase 3: Scotland's route map update - 9 july 2020, 2020. URL https://www.gov.scot/publications/coronavirus-covid-19-framework-decision-making-scotlands-route-map-through-out-cri pages/6/. (Cited on pages 63 and 66.)

Transport Scotland. Car occupancy, percentage of car stages 1 by car occupancy, 2008-2018, 2011. URL https://www.transport.gov.scot/publication/transport-and-travel-in-scotland-results-from-the-scottish-household-survey-1/table-td9-car-occupancy-percentage-of-car-stages-1-by-car-occupancy-2008-2018-2-3. (Cited on page 64.)

Scotland's Census. Output area, 2011. URL https://www.scotlandscensus.gov.uk/variables-classification/output-area-2011. (Cited on pages 64, 87, and 88.)

Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Polymod social contact data. 2017. (Cited on pages 64, 84, and 88.)

Christopher I Jarvis, Kevin Van Zandvoort, Amy Gimma, Kiesha Prem, Petra Klepac, G James Rubin, and W John Edmunds. Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. *BMC medicine*, 18:1–10, 2020a. (Cited on pages 65, 68, and 106.)

Christopher Jarvis, Amy Gimma, Kerry Wong, Kevin Van Zandvoort, et al. The effect of social distancing on the reproduction number and number ofcontacts in the uk from a social contact surveyreport for survey week 31, 2020b. URL https://cmmid.github.io/topics/covid19/reports/comix/CoMixWeeklyReport31.pdf. (Cited on pages 65 and 84.)

ONS. Coronavirus (covid-19) infection survey, uk, 2020. URL https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/18december2020. (Cited on page 65.)

Stanley Wasserman, Katherine Faust, et al. Social network analysis: Methods and applications. 1994b. (Cited on page 65.)

Scottish Government. Scottish index of multiple deprivation 2020, 2020a. URL https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/. (Cited on page 66.)

Jeffrey C Brinkman, Kyle Mangum, et al. *The Geography of Travel Behavior in the Early Phase of the COVID-19 Pandemic*. Research Department, Federal Reserve Bank of Philadelphia, 2020. (Cited on page 66.)

PHS. Electronic data research and innovation service (edris), 2020. URL https://www.isdscotland.org/Products-and-Services/EDRIS/. (Cited on pages 67 and 86.)

Christopher I Jarvis, Amy Gimma, Kevin van Zandvoort, Kerry LM Wong, and W John Edmunds. The impact of local and national restrictions in response to covid-19 on social contacts in england: a longitudinal natural experiment. *BMC medicine*, 19(1):1–12, 2021. (Cited on page 68.)

Amani Audi, Malak AlIbrahim, Malak Kaddoura, Ghina Hijazi, Hadi M Yassine, and Hassan Zaraket. Seasonality of respiratory viral infections: Will covid-19 follow suit? *Frontiers in Public Health*, 8:576, 2020. (Cited on page 75.)

Miyu Moriyama, Walter J Hugentobler, and Akiko Iwasaki. Seasonality of respiratory viral infections. *Annual review of virology*, 7:83–101, 2020. (Cited on page 75.)

Martha I Nelson and Edward C Holmes. The evolution of epidemic influenza. *Nature reviews genetics*, 8(3):196–205, 2007. (Cited on page 75.)

M Bouscambert, M Valette, and B Lina. Rapid bedside tests for diagnosis, management, and prevention of nosocomial influenza. *Journal of Hospital Infection*, 89(4):314–318, 2015. (Cited on page 75.)

UK Government. Uk covid-19 vaccines delivery plan, 2021. URL https://www.gov.uk/government/publications/uk-covid-19-vaccines-delivery-plan/uk-covid-19-vaccines-delivery-plan. (Cited on pages 75 and 76.)

Esteban Domingo and Celia Perales. The time for covid-19 vaccination. *Journal of Virology*, 95(8), 2021. (Cited on page 75.)

Takahiko Koyama, Dilhan Weeraratne, Jane L Snowdon, and Laxmi Parida. Emergence of drift variants that may affect covid-19 vaccine development and antibody treatment. *Pathogens*, 9(5):324, 2020. (Cited on page 75.)

Scottish Government. The national plan for scotland's islands, 2019. URL https://www.gov.scot/publications/national-plan-scotlands-islands/pages/6/. (Cited on page 76.)

ACAPS. Covid-19: Government measures created: 19/03/2020, 2020. URL https://www.acaps.org/sites/acaps/files/products/files/20200319_acaps_covid19_government_measures_report_0.pdf. (Cited on page 83.)

Howard Markel, Alexandra M Stern, J Alexander Navarro, Joseph R Michalsen, Arnold S Monto, and Cleto DiGiovanni Jr. Nonpharmaceutical influenza mitigation strategies, us communities, 1918–1920 pandemic. *Emerging infectious diseases*, 12(12):1961, 2006. (Cited on page 83.)

Julia E Aledort, Nicole Lurie, Jeffrey Wasserman, and Samuel A Bozzette. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC public health*, 7(1):1–9, 2007. (Cited on page 83.)

Yang Song, Min Zhang, Ling Yin, Kunkun Wang, Yiyi Zhou, Mi Zhou, and Yun Lu. Covid-19 treatment: close to a cure?–a rapid review of pharmacotherapies for the novel coronavirus. *International journal of antimicrobial agents*, page 106080, 2020. (Cited on page 83.)

Megan L Ranney, Valerie Griffeth, and Ashish K Jha. Critical supply shortages—the need for ventilators and personal protective equipment during the covid-19 pandemic. *New England Journal of Medicine*, 382(18):e41, 2020. (Cited on page 83.)

Vincenzo Alfano and Salvatore Ercolano. The efficacy of lockdown against covid-19: a cross-country panel analysis. *Applied health economics and health policy*, 18: 509–517, 2020. (Cited on page 83.)

Vera Clemens, Peter Deschamps, Jörg M Fegert, Dimitris Anagnostopoulos, Sue Bailey, Maeve Doyle, Stephan Eliez, Anna Sofie Hansen, Johannes Hebebrand, Manon Hillegers, et al. Potential effects of "social" distancing measures and school lockdown on child and adolescent mental health, 2020. (Cited on page 83.)

Mohamed Buheji, Katiane da Costa Cunha, Godfred Beka, Bartola Mavric, YL De Souza, S Souza da Costa Silva, Mohammed Hanafi, and T Chetia Yein. The extent of covid-19 pandemic socio-economic impact on global poverty. a global integrative multidisciplinary review. *American Journal of Economics*, 10 (4):213–224, 2020. (Cited on page 83.)

T Jombart, ES Nightingale, M Jit, O de Waroux, G Knight, S Flasche, R Eggo, AJ Kucharski, C Pearson, S Procter, et al. Forecasting critical care bed requirements for covid-19 patients in england. *CMMID Repository.[(accessed on 23 April 2020)]*, 2020. (Cited on page 83.)

Ruth McCabe, Mara D Kont, Nora Schmit, Charles Whittaker, Alessandra Løchen, Marc Baguelin, Edward Knock, Lilith K Whittles, John Lees, Nicholas F Brazeau, et al. Modelling intensive care unit capacity under different epidemiological scenarios of the covid-19 pandemic in three western european countries. *International journal of epidemiology*, 2021. (Cited on page 83.)

Scottish Government. Coronavirus (covid-19): allocation of levels to local authorities, 2020b. URL https://www.gov.scot/publications/coronavirus-covid-19-allocation-of-levels-to-local-authorities-17-november-2020/. (Cited on page 83.)

Public Health Scotland. Available beds by specialty nhs board of treatment, 2020. URL https://www.isdscotland.org/Health-Topics/Hospital-Care/Beds/. (Cited on page 84.)

Amani Alahmadi, Sarah Belet, Andrew Black, Deborah Cromer, Jennifer A Flegg, Thomas House, Pavithra Jayasundara, Jonathan M Keith, James M McCaw, Robert Moss, et al. Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. *Epidemics*, 32:100393, 2020. (Cited on page 84.)

Scottish Government. Rural scotland key facts 2021, 2021a. URL https://www.gov.scot/publications/rural-scotland-key-facts-2021/. (Cited on page 84.)

Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7):410–418, 2010. (Cited on page 84.)

NHS Western Isles. Nhs western isles – covid-19 video updates, 2021. URL https://www.coronavirus.wi.nhs.scot/?page_id=1506. (Cited on pages 85, 86, and 87.)

Junling Ma. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5:129–141, 2020. (Cited on page 87.)

Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009. (Cited on pages 89 and 90.)

Tina Toni and Michael PH Stumpf. Tutorial on abc rejection and abc smc for parameter estimation and model selection. *arXiv preprint arXiv:0910.4472*, 2009. (Cited on page 90.)

Emmanuel Klinger, Dennis Rickert, and Jan Hasenauer. pyabc: distributed, likelihood-free inference. *Bioinformatics*, 34(20):3591–3593, 2018. (Cited on page 90.)

Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995. (Cited on page 90.)

Lorenzo Sadun. Effects of latency on estimates of the covid-19 replication number. *Bulletin of Mathematical Biology*, 82(9):1–14, 2020. (Cited on page 91.)

Manuel Battegay, Richard Kuehl, Sarah Tschudin-Sutter, Hans H Hirsch, Andreas F Widmer, and Richard A Neher. 2019-novel coronavirus (2019-ncov): estimating the case fatality rate–a word of caution. *Swiss medical weekly*, 150 (0506), 2020. (Cited on page 92.)

Anthony S Fauci, H Clifford Lane, and Robert R Redfield. Covid-19—navigating the uncharted, 2020. (Cited on page 92.)

Ewan Colman, Jessica Enright, Gavrila Amadea Puspitarani, and Rowland Raymond Kao. Estimating the proportion of sars-cov-2 infections reported through diagnostic testing. *medRxiv*, 2021. (Cited on pages 92 and 108.)

Scottish Government. Coronavirus (covid-19): getting tested in scotland, 2021b. URL https://www.gov.scot/publications/coronavirus-covid-19-getting-tested/. (Cited on page 92.)

Timothy W Russell, Nick Golding, Joel Hellewell, Sam Abbott, Lawrence Wright, Carl AB Pearson, Kevin van Zandvoort, Christopher I Jarvis, Hamish Gibbs, Yang Liu, et al. Reconstructing the early global dynamics of under-ascertained covid-19 cases and infections. *BMC medicine*, 18(1):1–9, 2020. (Cited on page 92.)

Igor Salom, Andjela Rodic, Ognjen Milicevic, Dusan Zigic, Magdalena Djordjevic, and Marko Djordjevic. Effects of demographic and weather parameters on covid-19 basic reproduction number. *Frontiers in Ecology and Evolution*, 8: 524, 2021. (Cited on page 98.)

Tanvir Ahammed, Aniqua Anjum, Mohammad Meshbahur Rahman, Najmul Haider, Richard Kock, and Md Jamal Uddin. Estimation of novel coronavirus (covid-19) reproduction number and case fatality rate: A systematic review and meta-analysis. *Health science reports*, 4(2):e274, 2021. (Cited on page 98.)

Paulo Mecenas, Renata Travassos da Rosa Moreira Bastos, Antonio Carlos Rosário Vallinoto, and David Normando. Effects of temperature and humidity on the spread of covid-19: A systematic review. *PLoS one*, 15(9): e0238339, 2020. (Cited on page 98.)

Smriti Mallapaty. Why covid outbreaks look set to worsen this winter. *Nature*, 586(7831):653–654, 2020. (Cited on page 98.)

Shanna Ratnesar-Shumate, Gregory Williams, Brian Green, Melissa Krause, Brian Holland, Stewart Wood, Jordan Bohannon, Jeremy Boydston, Denise

Freeburger, Idris Hooper, et al. Simulated sunlight rapidly inactivates sars-cov-2 on surfaces. *The Journal of infectious diseases*, 222(2):214–222, 2020. (Cited on page 98.)

Paul Dabisch, Michael Schuit, Artemas Herzog, Katie Beck, Stewart Wood, Melissa Krause, David Miller, Wade Weaver, Denise Freeburger, Idris Hooper, et al. The influence of temperature, humidity, and simulated sunlight on the infectivity of sars-cov-2 in aerosols. *Aerosol Science and Technology*, 55(2):142–153, 2021. (Cited on page 98.)

Shane Riddell, Sarah Goldie, Andrew Hill, Debbie Eagles, and Trevor W Drew. The effect of temperature on persistence of sars-cov-2 on common surfaces. *Virology journal*, 17(1):1–7, 2020. (Cited on page 98.)

Shina CL Kamerlin and Peter M Kasson. Managing coronavirus disease 2019 spread with voluntary public health measures: Sweden as a case study for pandemic control. *Clinical Infectious Diseases*, 71(12):3174–3181, 2020. (Cited on page 99.)

Youpei Yan, Amyn A Malik, Jude Bayham, Eli P Fenichel, Chandra Couzens, and Saad B Omer. Measuring voluntary and policy-induced social distancing behavior during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(16), 2021. (Cited on page 99.)

Andrew A Cunningham, Peter Daszak, and James LN Wood. One health, emerging infectious diseases and wildlife: two decades of progress? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1725):20160167, 2017. (Cited on page 103.)

Udrescu Marius Alexandru. European solidarity in the context of covid-19 pandemic. 2020. (Cited on page 103.)

Christine K Johnson, Peta L Hitchens, Pranav S Pandit, Julie Rushmore, Tierra Smiley Evans, Cristin CW Young, and Megan M Doyle. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proceedings of the Royal Society B*, 287(1924):20192736, 2020. (Cited on page 103.)

Abel Brodeur, David M Gray, Anik Islam, and Suraiya Bhuiyan. A literature review of the economics of covid-19. 2020. (Cited on page 105.)

World Bank. Covid-19 to plunge global economy into worst recession since world war ii, 2020. (Cited on page 105.)

Scottish Government. Coronavirus (covid-19): framework for decision making - assessing the four harms, 2020c. URL https://www.gov.scot/publications/

covid-19-framework-decision-making-assessing-four-harms-crisis/pages/2/. (Cited on page 105.)

Christine Currie and Peter Falconer. Maintaining sustainable island destinations in scotland: The role of the transport–tourism relationship. *Journal of Destination Marketing & Management*, 3(3):162–172, 2014. (Cited on page 105.)

UK Health Security Agency. Covid-19 variants: genomically confirmed case numbers, 2021. URL https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers. (Cited on page 106.)

British Cattle Movement Service. British cattle movement service, 2021. URL https://www.gov.uk/government/organisations/british-cattle-movement-service/about. (Cited on page 106.)

Soren Alexandersen, Melvyn Quan, Ciara Murphy, Jeannette Knight, and Zhidong Zhang. Studies of quantitative parameters of virus excretion and transmission in pigs and cattle experimentally infected with foot-and-mouth disease virus. *Journal of comparative pathology*, 129(4):268–282, 2003. (Cited on page 106.)

Guadalupe Ray De La Garza. *Effective Contact of Cattle and Feral Swine Facilitating Potential Foot-and-mouth Disease Virus Transmission in Southern Texas, USA Rangeland*. PhD thesis, Texas A & M University, 2010. (Cited on page 106.)

Linnet Taylor. No place to hide? the ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D: Society and Space*, 34(2): 319–336, 2016. (Cited on page 106.)

Sarah J Cox and Paul V Barnett. Experimental evaluation of foot-and-mouth disease vaccines for emergency use in ruminants and pigs: a review. *Veterinary research*, 40(3):1, 2009. (Cited on page 107.)

Claire Guinat, Simon Gubbins, Timothée Vergne, Jose L Gonzales, Linda Dixon, and Dirk U Pfeiffer. Experimental pig-to-pig transmission dynamics for african swine fever virus, georgia 2007/1 strain. *Epidemiology & Infection*, 144 (1):25–34, 2016. (Cited on page 107.)

Claire Guinat, Ana Luisa Reis, Christopher L Netherton, Lynnette Goatley, Dirk U Pfeiffer, and Linda Dixon. Dynamics of african swine fever virus shedding and excretion in domestic pigs infected by intramuscular inoculation and contact transmission. *Veterinary research*, 45(1):1–9, 2014. (Cited on page 107.)

Karin Orsel, Annemarie Bouma, Aldo Dekker, JA Stegeman, and MCM De Jong. Foot and mouth disease virus transmission during the incubation period of the

disease in piglets, lambs, calves, and dairy cows. *Preventive veterinary medicine*, 88(2):158–163, 2009. (Cited on page 107.)

M Graeme Garner and SD Beckett. Modelling the spread of foot-and-mouth disease in australia. *Australian veterinary journal*, 83(12):758–766, 2005. (Cited on page 107.)

RL Sanson, MA Stevenson, GF Mackereth, and N Moles-Benfell. The development of an interspread plus parameter set to simulate the spread of fmd in new zealand. In *International Symposia on Veterinary Epidemiology and Economics (ISVEE) proceedings*, page 682, 2006. (Cited on page 107.)

Richard Bradhurst, Graeme Garner, Mark Hovari, Maria de la Puente, Koen Mintiens, Shankar Yadav, Tiziano Federici, Ian Kopacka, Simon Stockreiter, Ivanka Kuzmanova, et al. Development of a transboundary model of livestock disease in europe. *bioRxiv*, 2021. (Cited on page 107.)

Meta Roestenberg, Marie-Astrid Hoogerwerf, Daniela M Ferreira, Benjamin Mordmüller, and Maria Yazdanbakhsh. Experimental infection of human volunteers. *The Lancet Infectious Diseases*, 18(10):e312–e322, 2018. (Cited on page 107.)

Kenji Mizumoto, Katsushi Kagaya, Alexander Zarebski, and Gerardo Chowell. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance*, 25(10):2000180, 2020. (Cited on page 107.)

Thiemo Fetzer and Thomas Graeber. Does contact tracing work? quasi-experimental evidence from an excel error in england. 2020. (Cited on page 107.)

OIE. Guidelines for animal disease control, 2014. URL https://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/A_Guidelines_for_Animal_Disease_Control_final.pdf. (Cited on page 108.)

F Gary. Criteria and factors for rational prioritisation of animal diseases that should be covered by public health policies, 2014. URL https://web.oie.int/delegateweb/OIEdoc/sg82/gary/files/assets/basic-html/page1.html. (Cited on page 108.)

DEFRA. Foot and mouth disease control strategy for great britain, 2011. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69456/fmd-control-strategy111128.pdf. (Cited on page 108.)

Andrew D James and Jonathan Rushton. The economics of foot and mouth disease. *Revue scientifique et technique-office international des epizooties*, 21(3):637–641, 2002. (Cited on page 108.)

Abigail Woods. *A manufactured plague: the history of foot-and-mouth disease in Britain*. Routledge, 2013. (Cited on page 108.)

JA Backer, B Engel, A Dekker, and HJW Van Roermund. Vaccination against foot-and-mouth disease ii: regaining fmd-free status. *Preventive Veterinary Medicine*, 107(1-2):41–50, 2012. (Cited on page 108.)

Stephen X Zhang, Yifei Wang, Andreas Rauch, and Feng Wei. Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in china one month into the covid-19 outbreak. *Psychiatry research*, 288: 112958, 2020. (Cited on page 108.)

Wan Mohd Azam Wan Mohd Yunus, Siti Khadijah Zainal Badri, Siti Aisyah Panatik, and Firdaus Mukhtar. The unprecedented movement control order (lockdown) and factors associated with the negative emotional symptoms, happiness, and work-life balance of malaysian university students during the coronavirus disease (covid-19) pandemic. *Frontiers in psychiatry*, 11, 2020. (Cited on page 109.)

Paul R Bessell, Kate R Searle, Harriet K Auty, Ian G Handel, Bethan V Purse, and B Mark de C Bronsvoort. Assessing the potential for bluetongue virus 8 to spread and vaccination strategies in scotland. *Scientific reports*, 6(1):1–13, 2016. (Cited on page 109.)

Sema Nickbakhsh, Louise Matthews, Paul R Bessell, Stuart WJ Reid, and Rowland R Kao. Generating social network data using partially described networks: an example informing avian influenza control in the british poultry industry. *BMC Veterinary Research*, 7(1):1–16, 2011. (Cited on page 109.)

EPIC. Disease exercises, 2019. URL https://www.epicscotland.org/our-research/contingency-planning/disease-exercises/. (Cited on page 109.)

Dale Fisher and Annelies Wilder-Smith. The global community needs to swiftly ramp up the response to contain covid-19. *Lancet (London, England)*, 395(10230): 1109, 2020. (Cited on page 109.)

Madelon Kroneman and Jacques J Siegers. The effect of hospital bed reduction on the use of beds: a comparative study of 10 european countries. *Social science & medicine*, 59(8):1731–1740, 2004. (Cited on page 109.)

WHO. Hospital preparedness for epidemics, 2014. URL https://apps.who.int/iris/bitstream/handle/10665/151281/9789241548939_eng.pdf. (Cited on page 109.)

Terry Elizabeth Hedrick, Leonard Bickman, and Debra J Rog. *Applied research design: A practical guide*. Sage Publications, 1993. (Cited on page 109.)

Moyses Szklo. Why epidemiologists should get involved with policy, 2013. URL https://www.hsph.harvard.edu/news/features/why-epidemiologists-should-get-involved-with-policy/. (Cited on page 109.)

Diane Stone. Bridging research and policies, 2001. URL https://warwick.ac.uk/fac/soc/pais/research/csgr/research/keytopic/other/bridging.pdf. (Cited on page 110.)

Maria WJ Jansen, Hans AM Van Oers, Gerjo Kok, and Nanne K De Vries. Public health: disconnections between policy, practice and research. *Health Research Policy and Systems*, 8(1):1–13, 2010. (Cited on page 110.)

Herbert A Simon et al. Models of bounded rationality, volume 1: economic analysis and public policy. *MIT Press Books*, 1, 1984. (Cited on page 110.)

Muin J Khoury, Marta Gwinn, and John PA Ioannidis. The emergence of translational epidemiology: from scientific discovery to population health impact. *American journal of epidemiology*, 172(5):517–524, 2010. (Cited on page 111.)

May CI van Schalkwyk and Martin McKee. Research into policy: lessons from the covid-19 pandemic. *European Journal of Public Health*, 31(Supplement_4): iv3–iv8, 2021. (Cited on page 111.)

UK Government. Scientific pandemic influenza group on modelling (spi-m). URL https://www.gov.uk/government/groups/scientific-pandemic-influenza-subgroup-on-modelling. (Cited on page 111.)

SAGE. Spi-m-o: Summary of modelling for scenarios for covid-19 autumn and winter 2021 to 2022, 13 october 2021, 2021. URL https://www.gov.uk/government/publications/spi-m-o-summary-of-modelling-for-scenarios-for-covid-19-autumn-and-winter-2021-to- (Cited on page 111.)

Anthony O'Hare, RJ Orton, Paul R Bessell, and Rowland R Kao. Estimating epidemiological parameters for bovine tuberculosis in british cattle using a bayesian partial-likelihood approach. *Proceedings of the Royal Society B: Biological Sciences*, 281(1783):20140248, 2014. (Cited on page 112.)