Analysis of epidemiological models for disease control in single and multiple populations under resource constraints

Martin Vyska
Department of Plant Sciences
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College
May 2018
I would like to dedicate this thesis to my loving parents.
Declaration

This thesis is my own work, and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

I declare that this thesis is not substantially the same as any that I have submitted for a degree, diploma or other qualification at any other university. I further state that no part of my thesis has already been or is currently being submitted for any degree, diploma or other qualification.

Martin Vyska
May 2018
Acknowledgements

Above all, I would like to thank my supervisors, Chris Gilligan and Nik Cunniffe. Their guidance throughout my PhD helped me become a better researcher and our countless meetings throughout the years have been both enjoyable and invaluable.

I would also like to thank Ciara Dangerfield from the Epidemiology and Modelling group for her help during the preparation of this thesis. She made putting together such a substantial piece of work seem easy and I am very grateful for her time. Thanks to everybody else in the Epidemiology and Modelling group for helpful discussions, in particular Matt Castle, Ruairi Donnelly and Andrew Craig for their frequent and insightful comments on my work.

I would like to thank Trinity College, Cambridge and the Bakala Foundation for providing funding which made this PhD possible. And finally, I would like to thank my parents for their support throughout my studies - I never would have finished without you!
Abstract

Efficient management of epidemics is one of the primary motivations for computational modelling of disease dynamics. Examples range from reactive control measures, where the resources used to manage the epidemic in real time may be limited to prophylactic control measures such as deployment of genetically resistant plant varieties, which may lead to economic trade-offs. In such situations the question is how should resources for disease control be deployed to ensure the efficient management of the epidemic. Mathematical models are a powerful tool to investigate such questions since experiments are usually infeasible and the primary aim of this thesis is to study selected mathematical models of disease control to improve the current understanding of their behaviour.

We initially analyse the dynamical behaviour that arises from incorporating an economic constraint into two simple, but widely used epidemic models with reactive control. Despite the selection of simple models, the addition of constrained control leads to mathematically rich dynamics, including the coexistence of multiple stable equilibria and stable limit cycles arising from global bifurcations.

We use the analytical understanding obtained from the simple model to explore how to allocate a limited resource optimally between a number of separate populations that are exposed to an epidemic. Initially, we assume that the allocation is done at the beginning and cannot be changed later. We seek to answer the question of how the resource should be allocated efficiently to minimise the long-term number of infections. We show that the optimal allocation strategy can be approximated by a solution to a knapsack-type problem, that is the problem of how to select items of varying values and weights to maximise combined value without going over certain combined weight. The weights and values are given as functions of the population sizes, initial conditions, and the disease parameters. Later, we relax the assumptions to allow for reallocation and use the understanding of the dynamics gained from the simple models in the
beginning to devise a new continuous time reallocation strategy, which outperforms previously considered approaches.

In the final part of the thesis, we focus on plant disease and study a model of prophylactic control using a genetically resistant variety. We consider a trade-off where the genetic resistance carries with it a fitness penalty and therefore reduces yield. We identify the conditions on the parameters under which the resistant variety should be deployed and investigate how these change when the outbreak is uncertain. We show that deploying the resistant variety reduces the probability of an outbreak occurring and therefore can be optimal even when it would not be optimal to deploy it during the outbreak.
# Table of contents

1 Introduction ................................. 1
  1.1 History of mathematical modelling in epidemiology ......................... 1
  1.2 SIS and SIR(S) compartmental models .................................. 3
  1.3 Endemic equilibrium ........................................... 5
  1.4 Study and optimization of control interventions ........................... 6
  1.5 Aims and outline ........................................... 9
    1.5.1 Thesis outline ......................................... 11

2 Single population disease control with a resource constraint ............... 13
  2.1 Abstract ............................................... 13
  2.2 Introduction ........................................... 14
  2.3 Model description ........................................ 16
  2.4 The SIS type dynamics ..................................... 18
    2.4.1 Model analysis ..................................... 19
    2.4.2 Results ............................................ 22
  2.5 The SIRS type dynamics .................................... 24
    2.5.1 Model analysis ..................................... 24
    2.5.2 Results ............................................ 30
    2.5.3 Adding stochastic effects .............................. 35
  2.6 Discussion ............................................ 38
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.1</td>
<td>Author contributions</td>
<td>40</td>
</tr>
<tr>
<td>2.7</td>
<td>Appendix</td>
<td>41</td>
</tr>
<tr>
<td>2.7.1</td>
<td>A1. Dynamics of the standard SIS model</td>
<td>41</td>
</tr>
<tr>
<td>2.7.2</td>
<td>A2. Dynamics of the standard SIRS model</td>
<td>41</td>
</tr>
<tr>
<td>2.7.3</td>
<td>A3. Another way to model the effects of control</td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td>One-time allocation of a limited resource for disease control between multiple populations</td>
<td>43</td>
</tr>
<tr>
<td>3.1</td>
<td>Abstract</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Introduction</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>Single threat metapopulation with no coupling</td>
<td>46</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Model assumptions</td>
<td>46</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Model analysis</td>
<td>47</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Results</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>Multiple threats metapopulation</td>
<td>64</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Model assumptions</td>
<td>64</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Model analysis</td>
<td>64</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Results</td>
<td>67</td>
</tr>
<tr>
<td>3.5</td>
<td>Extension to the SIRS model structure</td>
<td>72</td>
</tr>
<tr>
<td>3.6</td>
<td>Discussion</td>
<td>75</td>
</tr>
<tr>
<td>3.7</td>
<td>Appendix</td>
<td>78</td>
</tr>
<tr>
<td>3.7.1</td>
<td>A1. Analysis of the knapsack approximation error when $n = 3$</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>Allocation of a limited resource for disease control between multiple populations with reallocation</td>
<td>81</td>
</tr>
<tr>
<td>4.1</td>
<td>Abstract</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Introduction</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Model with equilibrium resource reallocation</td>
<td>84</td>
</tr>
</tbody>
</table>
Table of contents

4.4 Model with continuous resource reallocation .......................... 89  
   4.4.1 Model description ............................................. 89  
   4.4.2 Analysis of the problem ........................................ 90  
   4.4.3 Strategy based on urgency ....................................... 94  
4.5 SIRS model with continuous resource reallocation ...................... 98  
   4.5.1 Model description ............................................. 98  
   4.5.2 Analysis of the problem ........................................ 99  
4.6 Discussion ............................................................. 102

5 Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective 105  
   5.1 Abstract .......................................................... 105  
   5.2 Introduction ...................................................... 106  
   5.3 Motivation from a simple model .................................... 108  
   5.4 Partial resistance and yield trade-off .............................. 111  
      5.4.1 Methods .................................................. 111  
      5.4.2 Results .................................................. 115  
   5.5 Discussion .......................................................... 121  
      5.5.1 Author contributions ........................................ 125  
   5.6 Appendix ........................................................... 126  
      5.6.1 A1. Sketch explaining the yield approximation ............... 126  
      5.6.2 A2. Calculation of the probability of a large scale outbreak 126  
      5.6.3 A3. Calculation of the yield function from the final size equations 127  
      5.6.4 A4. Extreme control results .................................. 128  
      5.6.5 A5. Numerical simulation: technical details ................. 129

6 Discussion ................................................................. 131  
   6.1 Summary of the results ............................................ 131
6.1.1 Main contributions ........................................ 134
6.2 Directions for future work ................................. 134

References .................................................................. 137
Chapter 1

Introduction

1.1 History of mathematical modelling in epidemiology

Infectious diseases have had significant impact on both human lives throughout history and on history itself (McNeill, 1976; Brauer, 2009). For example, Brauer (2009) hypothesises that the so-called Antonine plagues (possibly measles and smallpox) that invaded the Roman Empire in the 2nd century AD and caused high death tolls in the population as well as economic hardships may have facilitated the fall of the Roman Empire. Similarly, smallpox, followed by measles and diphtheria were important factors in the victory of Cortez over the Aztecs in 1519 and Pizarro over the Incas 1532. While the local populations were decimated by the disease, sometimes by as much as 90%, the European invaders had developed immunity to the pathogens. In Europe, the Black Death (bubonic plague) invaded in the 14th century and just between 1346 and 1350 it is estimated to have killed as many as one third of the population. It then kept recurring for more than 300 years and had a profound impact on the politics and economy of the medieval period.

Despite their impact, for most of our history, the causes of infectious diseases were poorly understood and typically ascribed to supernatural causes (Winslow, 1980). Among the first to suggest that epidemics in humans are caused by transmission was Hippocrates (Margotta, 1968). However, understanding progressed slowly and it took centuries before mathematics became useful in describing infectious disease outbreaks.
and assisting with their control. One of the first accounts comes from 1662, when John Graunt studied data on deaths due to disease in London and the surrounding countryside in the early 17th century (Graunt, 1662). Later, Daniel Bernoulli analysed smallpox mortality in Breslau in Poland and suggested that smallpox could be controlled using variolation (Bernoulli, 1760), that is inoculation of susceptible individuals with a small amount of infected material. This was an important insight and by the end of the 18th century, Edward Jenner successfully developed a vaccine against smallpox (Riedel, 2005) and thus started the practice of vaccination.

The understanding of transmission mechanisms advanced when John Snow mapped cholera cases during a London epidemic and realized that it is transmitted through water (Snow, 1855). John Snow also hypothesised that epidemics come to an end due to the depletion of susceptible hosts (Snow, 1853). His view was competing with that of William Farr who believed the pathogen loses potency as it infects further hosts, eventually leading to the end of the outbreak (Farr, 1866). In 1906 William Hamer tried to construct epidemic curves by considering a discrete time model with various assumptions on the transmission term. He noticed that the disease incidence, that is the number of new cases in each time period, was approximately proportional to the product of the number of susceptible and the number of infectious hosts (Hamer, 1906). This is the familiar law of mass action, first described in chemistry, in the study of chemical reactions (Guldberg and Waage, 1867). In the context of epidemiology, it of course assumes homogeneous mixing of hosts in the population, that is the probability of coming into contact is the same for every pair of hosts. It wasn’t until the work by Ross and Hudson (Ross and Hudson, 1917) and by Kermack and McKendrick (McKendrick, 1916) that researchers realised and explicitly argued that the incidence of new disease cases is a function of the contacts between infectious and susceptible individuals. This idea was at the heart of the theory of epidemics based on compartmental models later developed by Kermack and McKendrick (1927b, 1933).

A detailed account of the historical developments in the theory of epidemics can be found in Serfling (1952) while a historical account of the law of mass action can be found in Heesterbeek (2005).
Compartmental models are an important tool in studying epidemics. They are based on a simplifying assumption that every individual is in one of several pre-defined compartments with regards to its infection status (Brauer, 2008). One of the most simple and widely used models is to assume the population can be split into three compartments: susceptible individuals (S), who can be infected, infected and infectious individuals (I) and recovered or removed individuals (R) who can no longer transmit or contract the infection. These compartments can be further subdivided to include additional states of infection depending on the biological requirements of the pathogen being modelled. For instance, if a pathogen has a significant incubation period then it may be appropriate to include an exposed compartment (E), where the exposed individuals are infected but not yet infectious. This is appropriate for many pathogens, for example the causal agent of measles (Lessler et al., 2009; Grenfell et al., 2001). Or if an individual can be infectious but not showing symptoms, as is the case for a number of diseases, an infectious but asymptomatic compartment (A) can be included (Hsu and Hsieh, 2008). The model is then built in the following way: first the suitable compartments are selected and then the probabilistic rates of transition between these compartments are specified. The simplest and most common way to do this for susceptible, infected and recovered compartments is to assume a constant rate of transmission between infected-susceptible pairs and a constant rate of recovery of the infected individuals. This defines a stochastic process as individuals randomly transition between the compartments and effectively simulates an epidemic.

In this thesis, we will predominantly use deterministic compartmental models, which arise as a mean-field approximation to stochastic models (Barabási et al., 1999). Assuming that the infection and recovery rates are constant, that the individuals are homogeneously mixed in the population and that the population size is large (Keeling and Rohani, 2008), we arrive at a set of differential equations describing the trajectory of the system. Perhaps the most commonly used and the most famous is the SIR
model (Kermack and McKendrick, 1927b), given as

\[
\begin{align*}
\frac{dS}{dt} &= -\beta IS \\
\frac{dI}{dt} &= \beta IS - \mu I \\
\frac{dR}{dt} &= \mu I.
\end{align*}
\]

Here \( \beta \) represents the infection rate between infected and susceptible individuals and \( \mu \) is the rate at which infected individuals recover. The recovered individuals gain lifelong immunity and cannot be reinfected, an example of such a pathogen is provided by the measles virus (Grenfell et al., 2001). An SIR epidemic will spread through the population and slowly die out due to the depletion of susceptible individuals, unless new susceptible individuals appear in the population either due to births or immigration.

Not all diseases lead to permanent immunity after recovery, some only provide temporary immunity that wanes after a certain period of time, for example pertussis (Wendelboe et al., 2005) or malaria (Ngwa and Shu, 2000). The compartmental modelling framework allows for easy incorporation of this additional behaviour, by allowing transitions from the recovered compartment to the susceptible. The resulting model is referred to as SIRS model, and is governed by the following set of equations

\[
\begin{align*}
\frac{dS}{dt} &= \nu R - \beta IS \\
\frac{dI}{dt} &= \beta IS - \mu I \\
\frac{dR}{dt} &= \mu I - \nu R,
\end{align*}
\]

where \( \nu \) is the rate at which immunity wanes.

Some diseases don’t confer any immunity, and recovered individuals can be reinfected immediately. This is the case, for example, for gonorrhea (Lajmanovich and Yorke, 1976). Mathematically, this means we can drop the recovered class altogether and individuals who recover go straight into the susceptible compartment. The resulting model is the SIS model, governed by the equations

\[
\begin{align*}
\frac{dS}{dt} &= \mu I - \beta IS \\
\frac{dI}{dt} &= \beta IS - \mu I.
\end{align*}
\]
We use all of these three model structures in this thesis.

1.3 Endemic equilibrium

Out of the models mentioned above and in the absence of birth and death rates, only the SIS and SIRS allow for the existence of a stable endemic equilibrium. By this we mean a situation in which the prevalence is non-zero and the rates at which individuals leave the compartments exactly balance the rates at which they enter, and the compartment sizes therefore stay constant. An SIR epidemic can only be in an endemic equilibrium if further effects such as new births are added to the model (Keeling and Rohani, 2008). A simple way to describe epidemic dynamics is through the basic reproductive number (Heffernan et al., 2005), which gives the mean number of secondary infections caused by one infectious individual in a totally susceptible population and is denoted by $R_0$. This is typically used as a threshold parameter for whether the disease can invade the population (Kermack and McKendrick, 1927b; van den Driessche and Watmough, 2008), but in some simple models it can be used to describe the endemic equilibrium as well. The assumption of homogeneous mixing implies that the number of susceptible individuals infected by a single infectious host is proportional to the total number of susceptibles in the population. Therefore, when the number of susceptibles is reduced from $N$ to $N/R_0$, the average number of individuals infected by each infectious host will be exactly 1. This is exactly what we require from an endemic equilibrium and indeed, in both the SIS and SIRS models, the endemic proportion of susceptible individuals is equal to $1/R_0$ (Keeling and Rohani, 2008).

In more complicated epidemic models, for instance with explicitly modelled control measures, large numbers of compartments or allowance for spatial structure, calculating the endemic equilibrium is not as straightforward and there may in fact be multiple such endemic states, e.g. Kribs-Zaleta and Velasco-Hernandez (2000). The main subject of research in such cases is the stability and bifurcation properties of the fixed points. For example Hethcote (1985) studied two models with immunization and age-dependent infectivity and found conditions for the endemic equilibrium to be asymptotically stable. Others have studied the stability of endemic states in situations ranging from SIR, SIRS and SIS models with non-linear or non-constant transmission terms (Korobeinikov, 2006; van den Driessche and Watmough, 2000) to SIS models incorporating variable population size and time delay (Hethcote and...
van den Driessche, 1995) or the effects of media coverage on epidemic dynamics (Cui et al., 2008). Such external forcing can often affect the dynamics in a complicated way. For example, the endemic state can be periodic, in particular when periodic control such as pulse vaccination is implemented (Zhou and Liu, 2003) or when the there is environmental seasonality present (Aron and Schwartz, 1984; Schwartz, 1985, 1992). In this thesis we show that in the SIRS model with control, stable periodic endemic states can arise even in the absence of any external forcing (Vyska and Gilligan, 2016).

The concept of an endemic equilibrium is an important one in the field of epidemiology. In several of the optimization problems in this thesis, we seek to find control resource allocations which, if it is not possible to eradicate the disease, at least minimize the endemic number of infected individuals.

1.4 Study and optimization of control interventions

One of the primary motivations for mathematical modelling of disease dynamics is efficient management of epidemics. Infectious diseases are not only a significant burden on human health (Jones et al., 2008), but also on the economy, through loss of productivity or cost of care and in the case of livestock or plant pathogens. Plant diseases impose significant crop losses in agriculture, horticulture and forestry. Estimates of annual crop losses in agriculture range from 14% of crop yield worldwide (Agrios, 2005; Oerke et al., 1994; Oerke, 2006), extending to 20-40% when weeds and pests are included (Savary et al., 2012), with total elimination of crops in some severe epidemics (Madden, 2006). Therefore, insights into the dynamics of disease control provided by mathematical models can be of great value, especially as controlled experiments are typically infeasible. There has been much interest in the integration of epidemiological models of disease control with economic considerations (Klein et al., 2007; Geoffard and Philipson, 1996).

The models incorporating economics into epidemiology range from explicitly considering the cost of control to imposing a constraint on the available amount of the control resources. Often, researchers used optimal control theory (Seierstad and Sydsæter, 1986) to analyse optimal deployment of the control resource. This was done by Sethi and Staats (1978), who considered a simple compartmental Susceptible-Infected homogeneous model (Keeling and Rohani, 2008). They assumed both treatment of infected individuals and vaccination of susceptible individuals are available, at a fixed cost per
individual. They formulated an optimization problem by assigning an economic cost to new infections and identified conditions under which the treatment should be deployed and how much should be invested in it. Later, others built on their work and extended it by adding more realism. Goldman and Lightwood (2002) relaxed the assumption of a fixed cost of treatment and considered more complex forms of the function describing the cost of deploying treatment, including effects such as set-up cost or congestion effects at high levels of infection. Interestingly, they found that under the optimal treatment programme, less treatment resource is deployed at higher levels of infection. Later the optimal control analysis was extended to an Susceptible-Infected-Susceptible (SIS) model (Forster and Gilligan, 2007). The authors found that the optimal resource allocation rests heavily on the time-scale on which the control is applied. Importantly, they also investigated robustness of their approach to errors in estimating the disease parameters and to relaxing the homogeneous mixing assumption. Instead they imagined the hosts arranged on a lattice and infection spreading between nearby hosts.

Others looked at optimization problems related to types of control other than treatment of infected host. For example, in plant epidemiology, researchers looked at optimal allocation of control resource for removing (roguing) infected hosts between two types of host sharing a common pathogen (Ndeffo Mbah and Gilligan, 2010b), motivated by Sudden Oak Death (SOD) in California (Meentemeyer et al., 2011; Rizzo and Garbelotto, 2003). Using SOD as the model system, they considered the question of how to balance the allocation of a limited resource between the detection of infectious hosts and the removal of detected hosts (Ndeffo Mbah and Gilligan, 2010a).

Building on the analysis of the homogeneous models with control, there is a significant body of work integrating economic considerations with spatial structure of the host population. This is important not only because spatial structure has a significant effect on epidemiological dynamics, but also because it allows us to study problems involving allocation of resources between different regions. May and Anderson (1984) considered a subdivision of the population into smaller subpopulations, with less infection being transmitted between the subpopulations than within them. This is the so-called metapopulation model, a concept now well established in both ecology (Hanski and Ovaskainen, 2000) and epidemiology (Keeling and Gilligan, 2000b; Park et al., 2003). May and Anderson (1984) studied how to allocate vaccination doses between the subpopulations so as to achieve a set control goal with the least amount of doses possible. They noted the importance of considering spatial structure and found that
fewer vaccination doses are necessary than in the homogeneous model. Cairns (1989) extended the work by May and Anderson (1984) to arbitrary contact rates between the subpopulations and additionally explicitly modelled the rate at which hosts can be vaccinated. They showed that there is a unique optimal policy for allocating the vaccination doses between the subpopulations that minimizes the growth rate of the epidemic, and for the case of two subpopulations, they found an explicit formula for this optimum. The metapopulation approach to studying vaccination policies can be readily adapted so that the subpopulations describe different groups within the population, such as schoolchildren or elderly. Elveback et al. (1976) noted that influenza vaccination should be predominantly targeted at schoolchildren, because they are the age group most responsible for transmission. Dushoff et al. (2007) then investigated how to distribute a limited supply of influenza vaccine doses between a subpopulation highly effective at spreading the infection (such as schoolchildren) and a subpopulation particularly at risk from the infection (such as the elderly), in order to minimize the number of deaths resulting from the epidemic. Following from these earlier studies, Medlock and Galvani (2009) built a more general, realistic age-structure model (RAS) for an influenza epidemic and demonstrated that the recommendations for vaccine distribution from the U.S. Center for Disease Control and Prevention may need to be updated to include age structure of the population.

Another active area of research into the optimal allocation of control resource in metapopulations has been considering reactive control, in which treatment is applied to the infected individuals. Brandeau et al. (2003) studied a system of isolated populations with deployment of a control measure reducing the contact rates and thus transmission between the hosts, with the goal of minimizing the number of new infections or maximizing the quality-adjusted life years (QALYs) gained. They demonstrated that modelling the epidemiological dynamics is more appropriate than simple cost effectiveness analysis commonly used for healthcare resource allocation (Weinstein, 1995). Rowthorn et al. (2009) considered a situation in which two interconnected subpopulations of equal sizes share a common pool of a limited resource for control. They used SIS model dynamics and assumed that the resource can be continually reallocated between the populations at no extra cost. Using optimal control theory, they showed how the resource should be allocated between the two populations. Following from their work, Ndeffo Mbah and Gilligan (2011) extended the results into the Susceptible-Infectious-Recovered-Susceptible (SIRS) framework. Both Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) used deterministic epidemiological models and
arrived at the counter-intuitive result that, generally speaking, the control measures should preferentially target the sub-population with fewer infected hosts.

The approaches summarized above considered the problem from the perspective of a centralized authority with the power to decide resource allocation policy and ignored the behaviour of individuals in response to the epidemic and to control. This is often the case and in our work we follow the same line of reasoning. However, relaxing these assumptions leads to interesting areas of research. For example, behaviour in response to the epidemic becomes important when the individuals can travel freely between the regions. Gaythorpe and Adams (2016) studied optimal deployment of treatment facilities during outbreaks of waterborne infections. They considered a model without host behaviour and compared it with a model where infected individuals tend to gravitate towards the regions where such facilities are deployed. They found that such behaviour can have significant impact on the optimal deployment of the treatment facilities, because the infectious individuals who arrive to receive treatment in the region with a treatment facility increase the force of infection on the susceptible individuals living in the region.

Sometimes there is no central authority overseeing the allocation of the control resource. Instead each sub-population or individual can either attempt to “selfishly” optimize its own outcome or cooperate with the others, which necessitates the use of game theoretic methods. For example, a common problem with vaccination is so-called free riding (Ibuka et al., 2014), where some regions or countries rely on the vaccination efforts of their neighbours. This has been further addressed in Klepac et al. (2011) and Klepac et al. (2016), where the authors propose creating stable vaccination coalitions by incentivizing cooperation. They found such coalitions can achieve much greater vaccination coverage cheaper than by acting independently. This was also observed in the context of plant disease, where Epanchin-Niell and Wilen (2015) demonstrated how even a small amount of cooperation between independent landowners can bring significant social benefits when controlling an invasive biological agent.

1.5 Aims and outline

Generally speaking, mathematical models have two fundamental uses in epidemiology.
On one hand, they can be used to predict how an epidemic will develop. This includes the forecasting of future disease incidence, the final size of an epidemic (Arino et al., 2007) or the future endemic prevalence, in other words studying what will or might happen in reality, or in imagined scenarios given a particular control strategy. This is the most easily understandable application of mathematical theory. The models used for this purpose are often complex and attempt to balance the need for biological realism and inclusion of population-level heterogeneities with the need for feasibility of parametrisation from available data (Keeling and Rohani, 2008). The goal of such models is to inform policy making decisions regarding specific pathogens. There are many examples of such work. Stollenwerk and Jansen (2003) built a model to investigate the threshold number of clustered meningitis cases above which a local epidemic is developing and a quick action by the authorities is required. Attempting to answer the question of how to deal with a new outbreak of smallpox, Halloran et al. (2002) built a complex model comparing mass-vaccinations with local control measures. However, their model predicted advice conflicting with other models developed for that purpose, for example in (Meltzer et al., 2001), possibly because of the difficulty in accurately estimating the epidemiological parameters. Last but not least, models have been used to inform and guide vaccination influenza policies (Baguelin et al., 2013) In animal epidemiology, predictive models were useful during the 2001 foot-and-mouth epidemic in the UK, and predicted that local culling of livestock would dramatically reduce the number of cases (Keeling, 2005b).

On the other hand, models can be used to improve understanding of behaviour of epidemics and how the epidemic is affected by various features added to the model for greater realism in modelling the disease dynamics or control measures. For this purpose, simplistic models with a small number of features are often constructed in order to study those features in isolation. That way, the impact of the features on the epidemiological dynamics can be better understood. This can be used as a guiding light in building more complex, realistic models, where features’ contribution to the dynamics are often difficult to disentangle from each other. As an example of a development of understanding from simple models, we note the work by Keeling and Gilligan (2000b,a), who developed a simple compartmental model for bubonic plague and explained how the dynamics of the pathogen in the rat population led to the long seemingly disease-free periods in the 16th and 17th centuries (Keeling and Gilligan, 2000b,a). Another good example comes from the work by Rowthorn et al. (2009), who considered a simple compartmental model of two populations with a
limited control resource. This illustrates the point raised above: while the model in question is unlikely to describe any real disease outbreak to a high degree of accuracy, the insight it provides supplements intuition in designing control measures for more complex and realistic situations.

Throughout this thesis, we focus on mathematical models designed to improve understanding of the underlying dynamics and effectiveness of control, that is the models of the second kind as described above. Whenever we study a particular behaviour or dynamical property of a model, we try to construct the simplest example still exhibiting the property of interest, in order to isolate it and obtain intuitive understanding. Whenever possible, we aim to make analytical progress. This is useful, because it allows us to study the impact that the individual parameters have on the behaviour of the model or on the performance of a control strategy under consideration without sacrificing robustness or generality.

1.5.1 Thesis outline

In Chapter 2, we study the dynamics of two simple, but widely used, epidemic models, SIS and SIRS, with constrained reactive control. By constrained control we mean control which can be simultaneously applied to some, but not all the infected hosts in the population due to limited resource. We aim to build on existing models and investigate their interplay with constrained control in order to improve the analytic and intuitive understanding of the dynamics. While Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) investigated the optimal allocation of limited resource, they did not analyse the impact of such constrained control on the detailed dynamics of the disease model. However, this understanding turns out to be essential for extending their work to more general settings. We show that addition of constrained reactive control leads to complex dynamical behaviour. The models can have more than one stable endemic equilibrium and in the case of SIRS, there can also be stable periodic solutions in the form of a limit cycle.

In the two subsequent chapters, we use this understanding to investigate optimal allocation of control resource between multiple uncoupled populations of varying sizes in two different scenarios.

In Chapter 3, we assume that the resources have to be allocated between subpopulations at the beginning and cannot be reallocated later. This assumption would be
appropriate, for example, when reallocating the resources is very costly, or logistically infeasible. However, this assumption is also interesting mathematically as it leads to an optimization problem arising from the following trade-off. On one hand, allocating more of the control resource to a particular subpopulation is positive, in that it can eradicate the disease or at least significantly reduce the number of infected individuals in the endemic equilibrium. On the other hand, it has a negative consequence as the resources will be fixed to that subpopulation even after they are not needed any more and may be required in the other subpopulations.

In Chapter 4, we relax the assumption of fixed resource allocation. Following the work of Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) we assume that the control resource can be reallocated continuously in time. Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) investigated optimal allocation of control resource between two populations of equal sizes. We show that for some parameter values their results break when the assumption of equal sizes is relaxed. We consider a system of \( n \) subpopulations with varying sizes without coupling between the subpopulations and we find a new allocation strategy based on the understanding developed in Chapter 2.

In the final part of this thesis, Chapter 5, we investigate an economic trade-off in a prophylactic control of plant disease by means of the deployment of a resistant variety. We assume that while the resistant variety has lower susceptibility to the pathogen and lower infectiousness once infected, the resistance implies a fitness penalty which leads to reduced crop yield. A natural question that arises is how much of the resistant variety should be deployed in the population? The answer clearly depends on the dynamics of the disease and on the parameters governing the resistance and the yield penalty. We build a simple model to answer this question both when the epidemic outbreak in the absence of control is certain and when there is uncertainty as to whether an outbreak will happen in the first place.

In Chapter 6, we summarize the results and suggest avenues for potential future research.
Chapter 2

Single population disease control with a resource constraint

2.1 Abstract

Analysing schemes for control of infectious diseases is one of the main reasons for mathematical modelling in epidemiology. Often, control resources such as drugs or qualified medical personnel are insufficient to allow for treatment of all the infected individuals. In order to use the resources efficiently, we need to understand how the deployment of such constrained control impacts the dynamics of the epidemic.

Here, we analyse the dynamical behaviour of two simple, widely used models (SIS and SIRS) that integrate epidemiological dynamics with disease control and an economic constraint on the control resources. Despite their simplicity, the models exhibit mathematically rich dynamics, including multiple stable fixed points, hysteresis, that is the dependence of future behaviour on the past, and stable limit cycles arising from global bifurcations. We show that the endemic equilibrium reached by the system depends on the initial condition and that in some cases there is a threshold for the initial disease prevalence above which eradication is impossible. The understanding developed in this chapter turns out to be important for analysing problems of optimal allocation of control resources between multiple regions considered in the subsequent chapters. This chapter is based on published work (Vyska and Gilligan, 2016).
2.2 Introduction

A number of control measures that are deployed to slow an epidemic work by reducing the infectious period of infected individuals. This is a common way of modelling the effect of reactive control (Goldman and Lightwood, 2002; Forster and Gilligan, 2007) and it is readily included into the structure of compartmental models by increasing the rate of recovery of the infectious individuals. However, the control resources, such as antibiotics or qualified medical personnel, are often outstripped by demand (Cinti et al., 2009; Naicker et al., 2009). This translates to the modelling assumption that only a fixed number of infected individuals can be receiving treatment simultaneously. Such a control constraint has been studied before, in the context of optimal allocation of control resource (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011). The authors used optimal control theory to investigate which allocation strategies make the best use of the limited resource in an SIS and SIRS model settings, respectively. Rowthorn et al. (2009); Ndeffo Mbah and Gilligan (2011) did not study the detailed dynamics, such as epidemic curves or bifurcation diagrams of the models in question. However, an understanding of the dynamical properties turns out to be important for extending their work to more complicated scenarios. The main aim of this chapter is to develop this understanding through bifurcation analysis and long-term dynamics.

We select two simple, but widely used epidemic models with constrained control and we examine their deterministic dynamical behaviour. The two different model structures we consider are the SIS and SIRS compartmental models. Our goal is to improve understanding of the impact of an economic constraint on the control resource on the disease dynamics. The dynamical properties of these two simple models turn out to be interesting in their own right. We show that despite their simplicity, the models exhibit mathematically rich behaviour. Both models can give rise to multiple stable attractors and to catastrophic behaviour (Zeeman, 1977; Arnold, 1992) where a small change in the amount of control resources leads to a sudden, large change in the prevalence of disease. Furthermore, in the SIRS case the dynamics include stable limit cycles arising through global bifurcations. The presence of limit cycles in dynamical systems has long been of interest in mathematical biosciences, particularly in ecology (Rosenzweig and MacArthur, 1963; Kaung and Freedman, 1988; Hastings, 2001; Toupo and Strogatz, 2015) and epidemiology (Hethcote and Levin, 1989; Wang and Ruan, 2004; Jin et al., 2007). We demonstrate that the presence of limit cycles has important consequences for modelling the impacts of control. We also show that in some parts of
parameter space the SIRS model exhibits counter-intuitive behaviour in which lower initial disease prevalence leads to a higher prevalence at endemic equilibrium. This means that in some situations, depending on the initial conditions, it may be beneficial to delay the deployment of the treatment resources, in order to guide the system into a stable equilibrium with a lower endemic prevalence. This counter-intuitive result is a consequence of the interplay between the SIRS model dynamics and the constraint on the treatment resources. Whenever possible, we provide analytical conditions on the parameters of the model that give rise to the particular dynamics.

In the SIRS case, we also examine the sensitivity of the dynamical behaviour when stochasticity is introduced to the model to allow for inherent variability of the infection and recovery processes. We only present the stochastic effects in the SIRS model because of its complicated bifurcation behaviour, in particular limit cycles. These complications which give rise to interesting stochastic effects are not present in the SIS model. We do this by using the Gillespie (1976) construction to simulate every event in the system as an exponential random process with rates given by the deterministic model. Thus the stochastic effects we introduce are demographic in nature. We demonstrate that the existence of the limit cycles in the deterministic version of the model strongly impacts the behaviour of the stochastic version of the model. The stochastic fluctuations can cause transitions between different attractors of the system and in some cases can lead to extinction of the pathogen by perturbing the system onto a limit cycle which passes close to the line of zero prevalence in the phase space. Similar transitions between different attractors of the dynamical system have been previously studied in systems with seasonal forcing (Keeling et al., 2001).

Our work demonstrates that economical constraints on control in epidemiological models can lead to the existence of weakly stable attractors and complex bifurcation dynamics. While the analysis in this chapter is interesting in its own right, it also forms a basis for the resource allocation problems presented in the subsequent chapters. The work in this chapter is based on the paper by Vyska and Gilligan (2016).

In the two subsequent chapters, we apply the understanding developed in this chapter to a few selected optimal resource allocation problems.
2.3 Model description

A wide range of models are used for infectious disease dynamics. Of these, many are formulated as compartmental models (Kermack and McKendrick, 1927a; Anderson and May, 1991). The compartments represent groups of hosts that share an infection status, such as being infectious or susceptible. Considering all the hosts within one compartment as equivalent is a simplifying assumption that the transition rates between the compartments are constant, that is the underlying stochastic process is Markovian. In this chapter, we consider two types of compartmental models, an SIS type model without host demography and an SIRS type model with host demography. Here by demography we mean births and deaths independent of disease. We selected these models because they both lead to endemic behaviour and are relatively simple.

The structure of the SIS model is illustrated in Figure 2.1a. This describes a situation in which the time for which the hosts stay in the infected class after infection is exponentially distributed with mean $1/\mu$. After recovery, the hosts immediately move back to the susceptible class and can be reinfected. This model structure is appropriate for diseases that do not confer any immunity after recovery such as gonorrhea (Lajmanovich and Yorke, 1976; Anderson and May, 1991). The structure of the SIRS model is illustrated in Figure 2.1b. After recovery, hosts have temporary immunity and cannot be immediately reinfected. This immunity lasts for an exponentially distributed time period with mean $1/\nu$ after which the hosts rejoin the susceptible class. This model structure with temporary immunity is appropriate for diseases such as malaria (Aron, 1988; Filipe et al., 2007), tuberculosis (Castillo-Chavez and Feng, 1997) or syphilis (Grassly et al., 2004).

We assume the population size stays constant on the time scale of the epidemic and thus the per capita birth rate and death rate are the same and both equal to $\sigma$. The birth and death rate is assumed to be 0 in the SIS model. In fact, it is irrelevant, because it would only increase the recovery rate and thus it wouldn’t constitute a new, independent parameter.

Finally, in both models, the rate at which a susceptible host gets infected is $\beta I$, which assumes homogeneous mixing of the hosts. It can be understood as an aggregate of three terms, $n_C \times p_I \times I$ where $n_C$ is the number of contacts of an average host per unit time, $p_I$ is the probability of infection upon contact and $I$ is the proportion of infected individuals, that is the probability that the contact is with an infected
Fig. 2.1 The transition structure of a) SIS and b) SIRS compartmental models. All the rates are per capita and per unit of time. $\beta$ is the transmission rate and therefore $\beta I$ is the rate at which susceptible hosts get infected. $\mu$ is the rate of recovery and transition to the susceptible class (in the case of SIS) or the recovered class (in the case of SIRS). $\nu$ is the rate at which immunity is lost and hosts rejoin the susceptible class. Finally, $\sigma$ is both the birth and death rate, assumed to be equal.
individual. Note that the SIS model can be derived from the SIRS model by letting the rate at which immunity is lost, $\nu$, tend to infinity. We include a brief overview of the mathematical properties of the SIS model in Appendix A1 and of the SIRS model in Appendix A2.

The effects and effectiveness of control can be introduced in a number of ways. Here we consider a treatment that can be applied to infected individuals and increases their rate of recovery by a fixed amount $\eta$ (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011). To model the economic constraint, we assume that the control resources are constrained and no more than a proportion $\gamma$ of the hosts can be treated at any given time. The SIS model is then given by the equation for the proportions of susceptibles (S) and infecteds (I)

\[ \frac{dI}{dt} = \beta I (1 - I) - \mu I - \eta \min(I, \gamma), \]  
\[ S = 1 - I. \]

The SIRS model is described by a standard set of differential equations for the proportions of susceptibles (S), infecteds (I) and removed (R), given by

\[ \dot{I} = \beta IS - (\mu + \sigma)I - \eta \min(I, \gamma), \]  
\[ \dot{R} = \mu I + \eta \min(I, \gamma) - (\nu + \sigma)R, \]  
\[ S = 1 - I - R. \]

Here $\min(I, \gamma)$ refers to the smaller of $I$ and $\gamma$.

### 2.4 The SIS type dynamics

We start by analysing the comparatively simpler SIS model. We first give a full description of the bifurcation dynamics of the system and then discuss their implications for control.
2.4 The SIS type dynamics

2.4.1 Model analysis

The SIS model is characterized by equation (2.1). Here $\gamma$ is the proportion of hosts that can be treated at any given time, it therefore captures the limited treatment resources. Without loss of generality, we rescale the time so that one unit of time corresponds to the average length of one infectious period in the absence of treatment, that is $1/\mu$. This is equivalent to setting $\mu = 1$ and rescaling $\eta$ and $\beta$ appropriately, so that

$$\frac{dI}{dt} = \beta I (1 - I) - I - \eta \min(I, \gamma). \quad (2.6)$$

We wish to study the fixed points of (2.6). To that end, we have to consider the cases $I < \gamma$ and $I > \gamma$ separately. These correspond to the situations where we have enough resource to treat all the infected hosts ($I < \gamma$) and where we can only treat part of the infected hosts ($I > \gamma$).

We begin with $I < \gamma$. In this region, the fixed points are solutions to

$$0 = \beta I (1 - I) - (1 + \eta)I$$

$$= \beta I (C_T - I), \quad (2.7)$$

where $C_T$ is the endemic equilibrium proportion of infected hosts given full treatment. It is given by

$$C_T = 1 - \frac{1 + \eta}{\beta}. \quad (2.9)$$

Of course, if the treatment is strong enough to eradicate the disease completely, $C_T \leq 0$ and the interpretation of the endemic equilibrium is lost. In such a case, there is no realistic endemic equilibrium, but only a disease free equilibrium, $I = 0$. There are therefore two fixed points in this region, $A$ and $B$ with

$$I_A = 0 \quad (2.10)$$

$$I_B = C_T, \quad (2.11)$$

where clearly $B$ only has a meaningful interpretation when $C_T > 0$, otherwise it is unstable. As in the usual logistic growth equation, when $C_T > 0$, $A$ is unstable and $B$ is stable. When $C_T \leq 0$, $A$ is stable. However, $C_T > 0$ is not a sufficient condition for $B$ to exist, we also need $I_B < \gamma$, otherwise $B$ would be outside the relevant interval.
This translates to

\[ \gamma > C_T \quad (2.12) \]

and so \( B \) only exists when there is a sufficient amount of resource.

The other interval to consider is \( I > \gamma \). The fixed points are then solutions to the equation

\[ 0 = \beta I (1 - I) - I - \eta \gamma. \quad (2.13) \]

This is a quadratic equation and gives two fixed points, \( C \) and \( D \), given by

\[
I_C = \frac{1}{2} \left( C_0 - \sqrt{C_0^2 - \frac{4 \eta \gamma}{\beta}} \right) \quad (2.14) \\
I_D = \frac{1}{2} \left( C_0 + \sqrt{C_0^2 - \frac{4 \eta \gamma}{\beta}} \right) \quad (2.15)
\]

where \( C_0 \) is the endemic equilibrium proportion of infected hosts in the absence of treatment, that is \( C_0 = 1 - 1/\beta \). Before we address the question of when these fixed points exist, note that \( C \) is always unstable whereas \( D \) is always stable. This is because in the quadratic in (2.13), the coefficient in front of \( I^2 \) is negative. Therefore perturbing the system slightly from the root \( D \) in the positive (negative) direction results in a negative (positive) derivative and vice-versa for \( C \).

For these points to exist, we need three conditions to be satisfied, namely \( I_{C,D} < 1 \), \( I_{C,D} > \gamma \) and \( C_0^2 > 4\eta \gamma / \beta \).

First we check that \( I_D < 1 \). This follows immediately since

\[
\frac{1}{2} \left( C_0 + \sqrt{C_0^2 - \frac{4 \eta \gamma}{\beta}} \right) < 1 \\
\iff C_0 + \sqrt{C_0^2 - \frac{4 \eta \gamma}{\beta}} < 2 \\
\iff C_0^2 - \frac{4 \eta \gamma}{\beta} < 4 + C_0^2 - 4C_0 \\
\iff - \frac{\eta \gamma}{\beta} < 1 - C_0,
\]

which is always true.

The condition \( C_0^2 > 4\eta \gamma / \beta \) simply puts a restriction on \( \gamma \):

\[ \gamma < \frac{\beta C_0^2}{4 \eta} \equiv \gamma_{\text{max}}. \quad (2.20) \]
Next, we check when $I_D > \gamma$ is satisfied. This condition reads

$$2\gamma - C_0 < \sqrt{C_0^2 - \frac{4\eta \gamma}{\beta}}. \quad (2.21)$$

This is satisfied automatically whenever $\gamma < C_0/2$. If $\gamma \geq C_0/2$ we can square the inequality to obtain

$$\gamma - C_0 < -\frac{\eta}{\beta} \quad (2.22)$$

$$\iff \gamma < C_T. \quad (2.23)$$

In other words, for $I_D > \gamma$ we need

$$\gamma < \max(C_0/2, C_T). \quad (2.24)$$

Now, to ensure $C$ exists we need to check whether $I_C > \gamma$. We do not need to check that $I_C < 1$ because we already checked that $I_D < 1$ and by definition $I_C \leq I_D$. However, note that when both $B$ and $D$ coexist, $C$ must exist as in a one dimensional system we cannot have two stable fixed points without an unstable one in between them (Strogatz, 2014). Conversely, when $C$ exists, both $B$ and $D$ must exist (possibly with $B$ above replaced by $A$ at 0). This follows from basic topology, the solutions are restricted to the interval $(0, 1)$ and so the existence of an unstable fixed point implies the existence of a stable one both above and below it. This means that the existence of $C$ is equivalent to the simultaneous existence of $B$ and $D$. Putting together conditions (2.12), (2.20) and (2.24) gives that $C$ exists whenever $\gamma$ satisfies both

$$C_T < \gamma < C_0/2, \quad (2.25)$$

$$\gamma < \gamma_{\text{max}}. \quad (2.26)$$

It follows that $C$ can only exist when $C_0 > 2C_T$ and when $C_T < \gamma_{\text{max}}$. The first translates to

$$\eta > r \quad (2.27)$$
where we introduced a new parameter $r = \beta - 1$. The second translates to

\[
CT < \frac{\beta C_0^2}{4\eta} \quad \text{(2.28)}
\]

\[
\iff 4\eta(\beta - \eta - 1) < \beta^2 C_0^2 \quad \text{(2.29)}
\]

\[
\iff 4\eta^2 - 4\beta\eta C_0 + \beta^2 C_0^2 > 0 \quad \text{(2.30)}
\]

\[
\iff (2\eta - \beta C_0)^2 > 0. \quad \text{(2.31)}
\]

This is automatically satisfied and therefore we only need $\eta > r/2$. It is straightforward to check that when $\eta > r/2$ we have $C_0/2 > \gamma_{\text{max}}$ and therefore the full set of conditions equivalent to the existence of $C$ is

\[
\eta > r/2 \quad \text{and} \quad CT < \gamma < \gamma_{\text{max}}. \quad \text{(2.32)}
\]

### 2.4.2 Results

We can now summarize the fixed point behaviour of the system (2.1).

1. When $\eta < r/2$, $C$ cannot exist and therefore $B$ and $D$ cannot coexist. Therefore,

   - when $\gamma < CT$ only $D$ exists.
   - when $\gamma > CT$ only $B$ exists.

   This means there is a single non-zero endemic state which is given by $D$ when $\gamma < CT$ and by $B$ otherwise.

2. When $r/2 < \eta < r$, $C$ can exist. We then get that

   - when $\gamma < CT$ only $D$ exists.
   - when $CT < \gamma < \gamma_{\text{max}}$ then $B$, $C$ and $D$ coexist.
   - when $\gamma > \gamma_{\text{max}}$ only $B$ exists.

   This means that when $\gamma$ is outside of the interval $(CT, \gamma_{\text{max}})$ there is only one non-zero endemic state. However, when $\gamma \in (CT, \gamma_{\text{max}})$, there are two endemic states and which one is reached depends on the initial condition. The basins of attraction of these two equilibrium states are separated by the point $C$. 
Fig. 2.2 Illustration of the three different types of dynamical behaviour exhibited by the SIS model. Green lines correspond to stable fixed points, while the red dashed line corresponds to unstable. In (a) B and D never coexist and at $\gamma = C_T$, D continuously changes into B. Recall that $C_T$ is the endemic proportion of infected hosts given full treatment. In (b) $\gamma_{\text{max}} > C_T$ and the three points can coexist. Note that the amount of resources that need to be allocated to achieve the best possible long-term behaviour now depends on the initial condition. The three different initial conditions $I_1, I_2$ and $I_3$ are examples selected to illustrate this point. If the system starts with the initial prevalence $I_1$, we only need to allocate $C_T$. If the initial prevalence is $I_3$ we need to allocate $\gamma_{\text{max}}$ and in the case of $I_2$, the necessary $\gamma$ is given by the intersection of the fixed point $C$ with the line $I = I_2$. In (c) full treatment leads to eradication of the disease and so $B$ is replaced by $A$, the disease-free equilibrium. Note, however, that it is not globally stable. Again, the maximum allocation needed is dependent on the initial conditions. If the initial prevalence is $I_1$, the necessary $\gamma$ is the intersection of $C$ with $I = I_1$. If the system starts at $I_2$, we need to allocate $\gamma_{\text{max}}$.

3. When $\eta > r$, the behaviour is the same as in 2, only the treatment is strong enough to eradicate the disease and so $B$ is unrealistic and replaced by $A$ at 0.

Minimum amount of resource necessary

We might wish to know the minimum amount of resource necessary to be allocated to the population to achieve the best long term behaviour ($I \to C_T$ as $t \to \infty$). The above analysis shows that when $\eta > r/2$, this depends on the initial condition. Consider $\eta > r/2$. If the initial prevalence $I_0$ satisfies $I_0 < C_T$, then we only need to allocate $\gamma = C_T$. If the initial prevalence is greater than $I_C$ at the point where C and D annihilate, that is $I_0 > C_0/2$, then we need to allocate $\gamma = \gamma_{\text{max}}$. When the initial condition satisfies $C_T < I_0 < C_0/2$, the necessary $\gamma$ is given by the intersection of the
line \( I = I_0 \) with the curve \( I = I_C \), that is, it is the solution to the equation

\[
\frac{1}{2} \left( C_0 - \sqrt{C_0^2 - \frac{4\eta\gamma}{\beta}} \right) = I_0.
\] (2.33)

The solution is given by

\[
\gamma = \frac{\beta}{\eta} I_0 (C_0 - I_0).
\] (2.34)

In all three cases, when \( \gamma \) reaches the necessary value, the long term equilibrium state of the system undergoes a discontinuous transition from \( D \) to \( B \). This is even more pronounced when \( \eta > r \), since then \( B \) represents the disease-free equilibrium and so allocating the necessary \( \gamma \) leads to the eradication of the disease. The bifurcation diagrams of the system are plotted in Figure 2.2. Note that the system exhibits hysteresis, that is dependence of its state on its history (Krasnosel’skii and Pokrovskii, 2012; Noori, 2014). Consider Fig 2.2b and suppose the system is in state \( B \). If we decrease the amount of resource, as \( \gamma \) crosses \( C_T \), the system undergoes a quick transition into the state \( D \) with a much higher endemic disease prevalence. However, if we increase the resource above \( C_T \) again, the system does not transition back to \( B \), but stays in \( D \). We need to keep increasing the resource until \( \gamma \) exceeds \( \gamma_c \) for the system to transition to the state \( B \) again.

In the next chapter, we will use the understanding developed in this section to study the optimal allocation in a system of \( n \) non-interacting populations.

### 2.5 The SIRS type dynamics

In this section, we analyse the deterministic SIRS model and present the complex dynamical behaviour generated by the constrained treatment term. We then discuss its implications for control. We conclude the section with a brief discussion of the impact of the dynamics on the stochastic version of the model.

#### 2.5.1 Model analysis

To analyse the system (2.3-2.5), we calculate the fixed points and construct the bifurcation diagrams. We only consider the case when, in the absence of treatment, the pathogen can invade the population in the first place, that is the basic reproductive
number (Heffernan et al., 2005) satisfies $R_0 = \beta/(\mu + \sigma) > 1$. For the analysis it is useful also to define the “full treatment” basic reproductive number $R_T^0$ by

$$R_T^0 = \frac{\beta}{\mu + \eta + \sigma}. \quad (2.35)$$

The system of differential equations (2.3-2.5) can have at most four fixed points. There is always a fixed point at $(I, R) = (0, 0)$ which we denote as $A$. The point $A$ is unstable when $R_T^0 \geq 1$ and is stable otherwise. When $A$ is stable it means that the disease can be eradicated fully if the prevalence $I$ drops below a certain value. In the region $I < \gamma$ there can be another fixed point $B$ given by

$$I_B = \frac{(\nu + \sigma)(1 - 1/R_T^0)}{\eta + \nu + \mu + \sigma}, \quad R_B = \frac{(\mu + \eta)(1 - 1/R_T^0)}{\eta + \nu + \mu + \sigma}. \quad (2.36, 2.37)$$

This fixed point is stable whenever it exists (because it is the endemic equilibrium of the standard, simple SIRS model) and it exists whenever $R_T^0 > 1$ and

$$\gamma > \gamma_c \equiv \frac{(\nu + \sigma)(1 - 1/R_T^0)}{\eta + \mu + \nu + \sigma}. \quad (2.38)$$

This condition is simply $I_B < \gamma$. In the region $I > \gamma$ there can be two further fixed points, $C$ and $D$ (with $I_C < I_D$). These are found by solving

$$\beta I(1 - I - R) - (\mu + \sigma)I - \eta \gamma = 0 \quad (2.39)$$
$$\mu I + \eta \gamma - (\nu + \sigma)R = 0 \quad (2.40)$$

and are given by

$$I_C = \frac{\chi - \sqrt{\chi^2 - P}}{2\beta(\mu + \nu + \sigma)} \quad (2.41)$$
$$I_D = \frac{\chi + \sqrt{\chi^2 - P}}{2\beta(\mu + \nu + \sigma)} \quad (2.42)$$
$$R_C, D = \frac{\mu I_{C, D} + \gamma \eta}{\nu + \sigma} \quad (2.43)$$
where
\[
\chi = \beta(-\gamma \eta + \nu + \sigma) - (\mu + \sigma)(\nu + \sigma) \\
P = 4\beta \gamma (\nu + \sigma)(\mu + \nu + \sigma) > 0.
\]

To begin with, we must analyse the stability behaviour of these points. We begin with the point \(C\). When \(I > \gamma\), the Jacobian \(J\) of the system is given by
\[
J(I, R) = \begin{pmatrix}
\beta(1 - R - 2I) - (\mu + \sigma) & -\beta I \\
\mu & -(\nu + \sigma)
\end{pmatrix}.
\]

Denote the Jacobian at a point \(X\) by \(J_X\), thus for example \(J_C = J(I_C, R_C)\). We can then calculate the determinant of \(J_C\) as
\[
det J_C = -\beta(1 - R_C - 2I_C) - (\mu + \sigma)(\nu + \sigma) + \beta \mu I_C \\
= (\mu + \sigma)(\nu + \sigma) - \beta \left( \nu + \sigma - \gamma \eta - \frac{1}{\beta}(\chi - \sqrt{\chi^2 - P}) \right) \\
= -\sqrt{\chi^2 - P} < 0.
\]

Since the eigenvalues, \(\lambda\), of \(J_C\) are the solutions to the equation
\[
\lambda^2 - Tr(J_C)\lambda + det J_C = 0,
\]
the fact that \(det J_C < 0\) means their real parts will have opposite signs. This implies that the point \(C\) is always a saddle and thus unstable, whenever it exists. To investigate the stability properties of \(D\), note that as \(\gamma \to 0\), \(D\) is the endemic equilibrium of the standard SIRS model without treatment and therefore it must be stable (A1). The behaviour of \(D\) as \(\gamma\) increases then depends on the value of \(\eta\). There are five important regions on the \(\eta\) axis, I, II, III, IV resp. V, corresponding to \(\eta < \eta_1, \eta \in (\eta_1, \eta_2), \eta \in (\eta_2, \eta_3), \eta \in (\eta_3, \eta_4)\) resp. \(\eta > \eta_4\). The expressions for the boundaries \(\eta_i\) are given in the proof of the theorem below. The behaviour of the fixed points is plotted in the bifurcation diagrams in the Figure 2.3. Before going through the bifurcation diagrams, we prove the following theorem. The expressions for the critical values \(\eta_i\) can be found in its proof.

**Theorem 1.** The local fixed point behaviour shown in the Figure 2.3 is exhaustive, no other (local) bifurcation diagrams are possible.
Fig. 2.3 Bifurcation diagrams corresponding to the different values of $\eta$, for an SIRS model. Black lines mean that the fixed point is stable, red lines mean it is unstable. The numerical values used are: a $\eta = 0.3$, b $\eta = 0.65$, c $\eta = 1$, d $\eta = 1.5$ and e $\eta = 2.3$. The values of the remaining parameters are $\beta = 3$, $\mu = 1$, $\nu = 0.2$ and $\sigma = 0$. The vertical dashed lines are used to highlight the different regions on the $\gamma$-axis, $\gamma_c$ is the critical value at which the fixed point $B$ emerges, $\gamma'_c$ is defined in the proof of Theorem 1.
Single population disease control with a resource constraint

Proof. We have discussed stability of A and B in the text and we proved that C is always a saddle. To prove this theorem, we first consider what happens to D as $\gamma$ increases. $I_D$ must be a decreasing function of $\gamma$. What is its value when $\gamma = \gamma_c$?

Inserting the expression for $\gamma_c$ into the formula for $I_D$ gives that

$$I_D(\gamma_c) = \begin{cases} I_B & \text{if } \eta \leq \eta_2 \\ \frac{\eta(\nu + \sigma)}{\beta(\nu + \mu + \sigma)} > I_B & \text{if } \eta > \eta_2. \end{cases}$$

(2.51)

where

$$\eta_2 = -(\mu + \nu + \sigma) + \sqrt{(\mu + \nu + \sigma)(\beta + \nu)}. \quad (2.52)$$

Note this is the $\eta_2$ that separates regions II and III on the $\eta$ axis, in Figure 2.3 and therefore it is the value of $\eta$ at which the transition from D to B at $\gamma = \gamma_c$ becomes discontinuous. The argument presented here constitutes a derivation of its value.

At $\gamma = 0$, D is stable. It is easy to show that $\det J_D > 0$ and as $\gamma$ increases, $\text{Tr} J_D > 0$ when $\gamma \in (\gamma'_c, \gamma_0)$ where $\gamma'_c$ and $\gamma_0$ are the roots of the corresponding quadratic equation. This means that D can change its stability properties at most twice. For the purposes of the analysis here, only $\gamma'_c$ will be needed. It is given by

$$\gamma'_c = \frac{1}{2\beta\eta} \left( \mu^2 + 3\mu(\nu + \sigma) + 2(\nu + \sigma)(\beta + 2\nu + \sigma) \right)$$

$$- \frac{1}{2\beta\eta} \left( (\mu + 2\nu + 2\sigma)\sqrt{(\mu + \nu + \sigma)^2 + 4(\nu + \sigma)(\beta + \nu)} \right). \quad (2.53)$$

Now we need to check when D loses stability and whether it happens for $\gamma < \gamma_c$.

Solving the quadratic inequality $\gamma'_c > \gamma_c$, which, when satisfied, means that at $\gamma_c$, D is still stable, gives $\eta \notin (\eta_1, \eta_3)$ where $\eta_1$ and $\eta_3$ are the roots of the corresponding quadratic equation, given by

$$\eta_1 = \frac{1}{2} \left( -\mu - \nu - \sigma + \sqrt{(\mu + \nu + \sigma)^2 + 4(\nu + \sigma)(\beta + \nu)} \right) \quad (2.54)$$

$$\eta_3 = \frac{1}{2(\nu + \sigma)} \left( -(\mu + \nu + \sigma)(\mu + 3\nu + 3\sigma) + \sqrt{(\mu + \nu + \sigma)^2 + 4(\nu + \sigma)(\beta + \nu)} \right).$$

These are the critical values separating the regions I and II resp. III and IV on the $\eta$ axis, in Figure 2.3. This means that for $\eta < \eta_1$, D stays stable until it continuously transitions into B (see Figure 2.3a). Now we are ready to consider the other cases in turn:
• The case \( \eta \in (\eta_1, \eta_2) \). In this case, C cannot exist because D transitions into B continuously. We will show that it is impossible for D to lose stability and then become stable again before \( \gamma \) reaches \( \gamma_c \). To do this we consider the sign of \( \text{Tr} J_D \) at \( \gamma = \gamma_c \), when \( I_D(\gamma_c) = I_B \). This is equal to

\[
\text{Tr} J_D = \frac{\eta^2 - (\beta + \nu)(\nu + \sigma) + \eta(\mu + \nu + \sigma)}{\eta + \mu + \nu + \sigma}
\]  

Setting \( \text{Tr} J_D < 0 \) then gives a quadratic inequality which is satisfied if and only if \( \eta < \eta_1 \). Therefore when \( \eta \in (\eta_1, \eta_2) \) and obviously also when \( \eta \in (\eta_1, \eta_3) \) the only possible behaviour is such that D loses stability for some \( \gamma < \gamma_c \) and then continuously transitions into B. This justifies Figure 2.3b.

• The case \( \eta \in (\eta_2, \eta_4) \). We already know that in this case, \( I_D > I_B \) at \( \gamma = \gamma_c \). Now we will show that this means C always appears at \( \gamma = \gamma_c \). For C to exist, we must have \( I_C > \gamma \). This is once again a quadratic inequality and holds whenever \( \gamma \in (\gamma_c, \gamma_b) \) where the root \( \gamma_b \) is given by

\[
\gamma_b = \frac{(\beta - \mu - \sigma)(\nu + \sigma)}{\beta(\eta + 2(\mu + \nu + \sigma))}.
\]  

When does this interval exist? Setting \( \gamma_b > \gamma_c \) and solving for \( \eta \) reveals that the interval does exist when \( \eta > \eta_2 \). This proves that for \( \eta > \eta_2 \), C appears at \( \gamma = \gamma_c \). As \( \gamma \) increases further, there are two ways C can cease to exist. Either \( \gamma \) reaches \( \gamma_b \) or C and D collide, which happens when \( \chi^2 = P \). Solving this quadratic equation for \( \gamma \) reveals that this first happens at \( \gamma_a \) given by

\[
\gamma_a = \frac{\nu + \sigma}{\beta \eta} \left( \beta + \mu + 2\nu + \sigma - 2\sqrt{(\mu + \nu + \sigma)(\beta + \nu)} \right).
\]  

We will now show that \( \gamma_b > \gamma_a \) which means that C and D always collide and annihilate. We want to show that

\[
\frac{\nu + \sigma}{\beta \eta} \left( \beta + \mu + 2\nu + \sigma - 2\sqrt{(\mu + \nu + \sigma)(\beta + \nu)} \right) < \frac{(\beta - \mu - \sigma)(\nu + \sigma)}{\beta(\eta + 2(\mu + \nu + \sigma))}.
\]  

After solving for \( \eta \), this simplifies to \( \eta > (\nu + \mu + \sigma)(\beta + \nu - \eta_2)/\eta_2 \). We have divided by \( \eta_2 \) because it is always positive, as can be quickly checked using the assumption \( \beta > \mu + \sigma \). Simple algebra reveals that in fact \( (\nu + \mu + \sigma)(\beta + \nu - \eta_2)/\eta_2 = \eta_2 \) and therefore the above inequality reduces to \( \eta > \eta_2 \) which is trivially satisfied by...
assumption. To finish the justification of the bifurcation diagrams in Figure 2.3c and d, we need to check that when C and D collide, D is always unstable. To do this we need to show that $\gamma'_c < \gamma_a$ and a proof of this is the subject of Lemma 2.

- The final case, $\eta > \eta_4$. This region corresponds to $R_0^T \geq 1$ and therefore $\eta_4 = \beta - \mu - \sigma$. Since the disease can be eradicated, the point A is stable. The behaviour of D and C does not change compared to the case $\eta \in (\eta_2, \eta_4)$. They will be present as long as $\gamma_a > 0$ and for $\beta > \mu + \sigma$ this is always the case.

**Lemma 2.** $\gamma'_c < \gamma_a$.

*Proof.* Consider the quantity $D = 2\beta \eta (\gamma'_c - \gamma_a)$. Define new variables $x = \beta + \nu$ and $y = \sigma + \nu$ and without loss of generality, set $\mu = 1$ (this is just a rescaling of time). Then we can write

$$D(x, y) = 2y\sqrt{x(1 + y) + 1 + y - (1 + 2y)\sqrt{1 + 2y + y(4x + y)}}, \quad (2.59)$$

where $x > 1$ and $y > 0$. We will show that $D$ is always negative. After squaring $D < 0$ reads

$$4y^2x(1 + y) + (1 + y)^2 + 4(1 + y)y\sqrt{x(1 + y)} < (1 + 2y)^2(1 + 2y + y^2 + 4xy). \quad (2.60)$$

Separating the square root and squaring again gives

$$x(1 + y)^3 < [(1 + y)^3 + x(1 + 3y + 3y^2)]^2 \quad (2.61)$$

and rearranging the terms leads to

$$x^2(1 + 3y + 3y^2)^2 + (1 + y)^3(1 + 6y + 6y^2)x + (1 + y)^6 > 0, \quad (2.62)$$

which is evidently satisfied. This finishes the proof.

\[\square\]

**2.5.2 Results**

We have established that Figure 2.3 covers all the possible types of local dynamical behaviour exhibited by the system (2.3-2.5). In region I the bifurcation diagram is
2.5 The SIRS type dynamics

Fig. 2.4 A representative example of how the limit cycles emerge in the system and the point D loses stability after the unstable limit cycle shrinks around it as $\gamma$ increases. The values of $\gamma$ used are (from left to right) 0.0395, 0.0409 and 0.042. The values of the other parameters are $\eta = 0.65$, $\beta = 3$, $\mu = 1$, $\nu = 0.2$.

simple, with $D$ stable throughout and continuously transitioning into $B$ at the $\gamma = \gamma_c$ boundary (Fig. 2.3a). When $\eta$ increases into the region II, the fixed point $D$ loses stability at $\gamma = \gamma'_c$ before changing into $B$ (Fig. 2.3b). This has implications for the phase portraits, since when $D$ is unstable there is no stable fixed point in the system. Since the solutions are bounded, it follows from the Poincaré-Bendixson theorem (Teschl, 2012; Bendixson, 1901) that there must exist a stable limit cycle. In fact, for some unknown value or values of $\gamma < \gamma'_c$, a stable and an unstable limit cycle appear in the system through global bifurcation(s). The unstable limit cycle then shrinks around the fixed point $D$, causing it to lose stability. An example of this behaviour is presented in Fig. 2.4.

In Figure 2.5a we show an example of the phase portrait within region II just after the two limit cycles appear in the system. As $\eta$ increases (region III, Fig. 2.3c), the fixed points $B$ and $C$ appear through a saddle-node bifurcation. $D$ loses stability through the interaction with the unstable limit cycle as before and consequently there are limit cycles present. See Fig. 2.5b for an example of the phase portrait when all the fixed points are present in the system.

In region IV, $\gamma'_c > \gamma_c$ and so $D$ loses stability after $B$ appears in the system. Therefore, for values of $\gamma \in (\gamma_c, \gamma'_c)$ two stable endemic equilibria exist in the system. The corresponding bifurcation diagram is given in Figure 2.3d. The dynamical behaviour for values of $\eta$ in the region IV is complicated and here we give an example of a phase portrait showing both $B$ and $D$ stable (Fig. 2.5c). Note that the stable limit cycle in this case is large and comes close to the $I = 0$ axis. This has implications for the stochastic behaviour of the system, which are discussed in the next section, since there
Fig. 2.5  

**a** The two limit cycles, in an SIRS model, after they emerge through a global bifurcation around the stable fixed point $D$ in region II. Green corresponds to stable, red to unstable. Parameter values are $\eta = 0.65$, $\gamma = 0.04$.  

**b** All four fixed points coexisting with a stable limit cycle. The detail shows the basin of attraction of $B$ whose boundary is the stable manifold of the saddle point $C$. Note that the basin of attraction (grey) is very small and thus solutions are likely to end up on the large limit cycle. The parameter values are $\eta = 1$, $\gamma = 0.0315$.  

**c** The fixed points $B$ and $D$ are both stable. The behaviour is similar to that in (b) since the stable $D$ together with the unstable limit cycle around it act globally as an unstable fixed point. The parameter values are $\eta = 1.3$, $\gamma = 0.02$.  

**d** Coexistence of two stable fixed points without a limit cycle. Note that the trajectory starting at $Y$ leads to the disease-free state while that starting at $X$ does not, even though $I_Y > I_X$ and $R_Y < R_X$. Thus higher initial prevalence can lead to lower long term prevalence or even eradication. The parameter values are $\eta = 2.1$, $\gamma = 0.01$. In all the simulations, $\beta = 3$, $\mu = 1$, $\nu = 0.2$ and $\sigma = 0$. The unstable limit cycles were found numerically by reversing time. The vertical dashed red lines correspond to the prevalence below which all the infected individuals can be treated.
can be a significant probability of stochastic pathogen extinction on the limit cycle due to the very low minimal prevalence. This can happen even if the system initially starts at $D$, since stochastic fluctuations can perturb it outside of the unstable limit cycle.

Finally, region V corresponds to values of $\eta$ such that the fixed point $A$ becomes stable, that is the eradication of the pathogen becomes possible. This is equivalent to $R_0^T \leq 1$ and therefore $\eta = \beta - \mu - \sigma$. The bifurcation diagram is given in Figure 2.3e and a phase portrait showing both stable $D$ and the stable disease-free equilibrium $A$ coexisting in Figure 2.5d. Note that when two stable fixed points coexist in the system, the model predicts a counter-intuitive dependence of the endemic equilibrium on the initial conditions (Fig. 2.5d). Starting the system at X does not achieve eradication of the pathogen while starting it at Y does, even though at Y the prevalence is higher and the population resistance ($R$) is lower. The system also exhibits catastrophic behaviour. When $\gamma$ is increased just above $\gamma_c'$, the threshold for destabilizing $D$, the system undergoes a rapid transition to the disease-free equilibrium $A$.

**Implications for resource allocation**

Like in the SIS case, suppose we want to know the minimum amount of resources necessary in order to gain maximum benefit from the control, that is an amount such that the system ends up in either $B$ or $A$ and further allocation has no effect. Clearly, like in the case of the SIS model, this amount can depend on the initial conditions. It also depends on the region of $\eta$, that is the additional recovery rate due to treatment.

- In region I, the required amount $\gamma$ is simply the amount necessary for $D$ to transition into $B$.

$$\gamma = \gamma_c = \frac{(\nu + \sigma)(1 - 1/R_0^T)}{\eta + \mu + \nu + \sigma}. \quad (2.63)$$

- In region II, the situation is more complicated. When $D$ transitions into $B$, there is a stable limit cycle surrounding them. Since we cannot have a stable fixed point inside a stable limit cycle, an unstable limit cycle must be created as the unstable $D$ changes into the stable $B$. This means that allocating $\gamma_c$ like before is not enough, since the system would inevitably end up on the stable limit cycle, rather than on the fixed point $B$. Depending on the initial conditions, we need to allocate more, up to an upper limit $\gamma_{LC}$, where the two limit cycles annihilate. Since this is a global bifurcation, it is not possible to obtain an
analytical expression for $\gamma_{LC}$. It may be sufficient to allocate less than $\gamma_{LC}$, if the initial conditions are close enough to $B$ that they lie within the unstable limit cycle.

- In the regions III and IV, the situation is similar to the region II. When $B$ comes into existence, its basin of attraction can be very small (Fig 2.5) and in such a case it is very unlikely that the initial conditions will fall into it. The system will almost certainly end up on the stable limit cycle, unless we allocate enough resources that this limit cycle disappears. This again happens through a global bifurcation of limit cycles.

- In region V, $A$ is stable and the disease can be fully eradicated with treatment. The required amount of resources depends heavily on the initial conditions. The required amount $\gamma$ is such that the basin of attraction of the fixed point $D$ shrinks enough so that the initial conditions lie outside of it. In the worst case, we will have to allocate enough resources to destabilize $D$, which happens at

$$\gamma = \gamma'_c$$

defined in equation (2.53). Note that in particular, if the system is initialized in the treatment-free endemic state and the resources are added slowly, it will stay in $D$ until $D$ loses stability, so in this special case, the required amount is precisely $\gamma'_c$.

Unfortunately, the situation is further complicated by the behaviour exhibited in Figure 2.5d, which establishes that higher initial prevalence can lie outside of the basin of attraction of $D$ and thus lead to eradication of the disease, where a lower initial prevalence would only lead to $D$. This means that it is possible to get a better result by waiting before starting the treatment until the initial condition leaves the basin of attraction (under the desired treatment) of $D$. In particular, suppose that the initial conditions lie inside of the basin of attraction of $D$ and so we need $\gamma'_c$ to eradicate the disease. Suppose further that there exists $\gamma'_E < \gamma'_c$ such that when $\gamma'_E$ is allocated, the endemic equilibrium in the absence of treatment lies outside of the basin of attraction of $D$. This means we don’t need $\gamma'_c$ to eradicate the disease after all. We need to merely delay the treatment and let the system evolve towards the treatment-free endemic equilibrium. Once it is close enough to it, we switch on the treatment and allocate only $\gamma'_E$, which will cause the system to go to $A$ and thus achieve eradication. We
2.5 The SIRS type dynamics

Fig. 2.6 Left plot: Phase portrait for the SIRS model. In red is the path when the treatment is applied immediately. Since the initial conditions are inside the basin of attraction of the point B, the red path spirals towards B. In the green is the delayed treatment path, where we delayed the treatment by 5 infectious periods. The blue dot marks the spot where treatment is switched on, and also the position of the endemic equilibrium in the absence of treatment. Since the endemic equilibrium lies outside the basin of attraction of B, switching the treatment on later leads to eradication. Right plot: the same situation, but showing the proportion of infecteds plotted against time. The red corresponds to treating immediately, the green corresponds to delayed treatment. The dashed green line shows how the system would have evolved had we not treated at all. The parameter values used are $\beta = 3$, $\mu = 1$, $\eta = 2.1$, $\nu = 0.3$, $\sigma = 0$ and $\gamma = 0.02$. Note that this $\gamma$ corresponds to the $\gamma_E$ in the text.

illustrate this in the Figure 2.6. This is a highly counter-intuitive behaviour caused by the dynamical properties of the SIRS model.

2.5.3 Adding stochastic effects

We conclude this chapter with a brief discussion of the interaction of the dynamics described above with stochastic effects. It turns out that the dynamical behaviour discussed in the previous subsection strongly impacts the behaviour of the stochastic model. In this subsection we describe the method used to simulate the stochastic dynamics and then discuss their consequences for the behaviour of the model.

Simulation method

To simulate the full stochastic process, we use the standard Gillespie algorithm (Gillespie, 1976). We give a brief description of the algorithm here. Suppose there are $N$ individuals in the population. We first calculate the total rate $R_t$ of an event occurring, which is
the sum of the rate of an infection $R_i$, recovery $R_r$, loss of immunity $R_l$, a birth $R_b$ and a death $R_d$. These are given by

\[ R_i = \beta ISN \]  
\[ R_r = \mu IN + \eta \min(I, \gamma)N \]  
\[ R_l = \nu RN \]  
\[ R_b = \sigma N \]  
\[ R_d = \sigma N \]  
\[ R_t = N(\beta IS + \mu I + \eta \min(I, \gamma) + \nu R + 2\sigma), \]

using the initial values of $I$, $S$ and $R$. The algorithm then proceeds as follows:

1. Calculate the time until the next event $\tau$ by simulating an exponential random variable with mean $1/R_t$.

2. Determine which event occurs by noting that the probability of an event $x$ happening is equal to $R_x/R_t$.

3. Calculate the updated values of $S$, $I$ and $R$ after the time $\tau$ and save them.

4. Recalculate the event rates and the total rate and go to step 1.

The output of the algorithm is a series of times and the values of the state variables $S$, $I$ and $R$ at those times.

**Stochastic behaviour**

When the unstable limit cycle exists around the stable fixed point $D$, stochastic fluctuations can perturb the solution from $D$ over the limit cycle. The system then transitions to another stable state; either a stable limit cycle or another stable fixed point. Furthermore, in regions III and IV, the stable limit cycles have large amplitude and come close to the $I = 0$ axis. This means that once on the stable limit cycle, the pathogen might go extinct (Fig. 2.7a). The trajectories start at the fixed point $D$ and fluctuate around it. Eventually, they cross the unstable limit cycle and fall onto the stable limit cycle, which leads to large amplitude oscillations. Eventually, the trajectories lead to extinction as can be seen from the downward slope of the average (red curve). This is in stark contrast to the deterministic model (green curve).
Fig. 2.7 Stochastic realizations of the SIRS model in three different scenarios. The green curve shows the predicted deterministic behaviour, the red curve is the average of the stochastic realizations and the blue curve shows one of the stochastic realizations.

a The scenario from Fig. 2.5c. When the model is started at the stable fixed point $D$, the presence of the unstable limit cycle means that the stochastic fluctuations can perturb it outside and onto the large stable limit cycle. Once there, the pathogen is likely to go extinct due to the low minimum prevalence on the cycle.

b The scenario from Fig. 2.5d. The trajectories fluctuate around $D$, but rarely go extinct, as demonstrated by the small downward slope of the red curve.

c The same as b only with $\gamma$ increased from 0.01 to 0.011. The emergence of the limit cycle is enough to significantly increase the probability of extinction. In all the simulations, $\beta = 3$, $\mu = 1$, $\nu = 0.2$, $\sigma = 0$ and $N = 5000$. 

[Diagram showing different scenarios]
It shows that when even a simple economic constraint is added, the deterministic model diverges from the stochastic model by failing to capture the risk of extinction which can be appreciable not only when the disease prevalence is low but also in the endemic equilibrium where the disease prevalence is appreciable and where the risk of extinction would consequently be vanishing in the absence of the economic constraint. Figures 2.7b and 2.7c illustrate how this impacts control. In both, the system starts in the stable fixed point $D$. In Fig. 2.7b the eradication probability is low as demonstrated by the very small downward slope of the average (red curve). The effect of increasing the resources for control, $\gamma$, by a small amount (0.1% of the total population) is illustrated in Fig. 2.7c. A global bifurcation gives rise to an unstable limit cycle around the fixed point $D$ and consequently its basin of attraction shrinks. This significantly increases the probability of eradication, as can be seen from the steep drop in the average. This potential benefit of slightly increasing the control resources $\gamma$ would be completely hidden in the deterministic model.

### 2.6 Discussion

In this chapter, we studied the dynamics of two simple epidemic models with control treatment and an economic constraint on the treatment resource such that only a certain proportion $\gamma$ of the population can be treated at any given time. This corresponds to a limited amount of drug, insufficient infrastructure for administering the treatment or lack of specialised personnel. We assume that the effect of the treatment is to reduce the infectious period of the treated hosts, which is typical of antibiotic treatment. Another simple way of modelling the control could be by assuming it reduces the infectiousness of the treated hosts. This would correspond to antiviral or fungicide treatment, which would mainly reduce the transmission rate. In Appendix A3 we briefly discuss this way of modelling the control and show that the structure of the equations is the same. We therefore expect the same qualitative behaviour. We note that the control constraint could be modelled differently as well, for example the rate at which the treatment is applied could be limited (White et al., 2005).

The way of incorporating the effect of treatment in this work has been considered before, in different scenarios. Ndeffo Mbah and Gilligan (2011) were primarily concerned with optimal allocation of drugs across two subpopulations, following previous work by Rowthorn et al. (2009) who considered an SIS model. Others have previously considered
optimal allocation of constrained resources in a single population (Forster and Gilligan,
2007; Goldman and Lightwood, 2002; Sethi and Staats, 1978). Conventional analysis
centres around continuously adjusting the amount of resource available to optimize the
overall cost, using optimal control theory (Seierstad and Sydsaeter, 1986). However,
the detailed dynamics of either the SIS or the SIRS model with constrained control
resources have not been investigated before. Since optimal control theory becomes
mathematically intractably complex as more subpopulations are considered, such
understanding of the dynamics of the system with constrained control can be helpful
in guiding the intuition when studying more realistic problems of resource allocation.
At the same time, the non-linear dynamics of the constrained control system turn out
to be interesting in their own right, thanks to the insights they provide on the effect of
an economic constraint on the inherent dynamics of the epidemic system.

We showed that in both the SIS and the SIRS case, the system can have more than
one endemic equilibrium and exhibits hysteresis, that is dependence of the state of the
system on its history. This phenomenon is not uncommon in more complex models
in both ecology (Beisner et al., 2003) and epidemiology (Hdeeler, 1997; Gross et al.,
2006). Crucially, the final endemic equilibrium state which the solutions reach depends
non-trivially on the initial conditions. This type of behaviour can be very important
for the efficient control of the disease (White et al., 2005). When it occurs, there is a
threshold proportion of infected hosts, $I_0^c$, such that given a certain amount of resources,
we can achieve the lower-prevalence equilibrium (or even eradicate the disease) if and
only if $I < I_0^c$. This will be of critical importance later (specifically in Chapter 4)
when we consider optimal allocation of resources between several populations, because
neglecting to allocate resources to a population may cause the prevalence in that
population to climb over the threshold and in such situations the disease can never be
eradicated. Furthermore, in the SIRS case, it is possible for solutions with initially
fewer infected hosts to end up in a higher-prevalence equilibrium. This can lead to
counter-intuitive behaviour where delaying the deployment of treatment leads to a lower
long-term prevalence of infection. This is especially the case when the treatment-free
endemic equilibrium lies outside the basin of attraction of the higher-prevalence fixed
point.

In the SIRS case, the system exhibits global bifurcations, as the critical parameter
$\gamma$ (control resources) is changed, which gives rise to limit cycles. The existence of
the cycles has profound implications for the behaviour of the stochastic counterpart
Single population disease control with a resource constraint

of the deterministic model. Normally, stochastic solutions initiated at a stable fixed point fluctuate around it (on time-scales shorter than exponential in $N$, the number of individuals (Allen and Burgin, 2000)). However, when there is an unstable limit cycle surrounding the fixed point, the stochastic fluctuations can perturb the solutions across the cycle. The solutions then tend to a different stable attractor. This facilitates transitions between stable attractors which would be much less likely in a stochastic system without the unstable limit cycle and would not be possible at all in the deterministic system. This is of particular importance when one of the attractors in question is the disease-free equilibrium because the combined effect of the stochastic fluctuations and the deterministic dynamics might then facilitate disease eradication. This is a benefit of the control deployment which would not be revealed if the detailed dynamics were not considered. Furthermore, the stable limit cycle often comes very close to the $I = 0$ axis and thus may facilitate stochastic extinction of the pathogen even if the disease-free equilibrium of the deterministic system is not stable.

Most of the non-trivial dynamical behaviour occurs close to $\gamma = \gamma_c$, which for the parameter values considered in this paper corresponds to the ability of treating between 1% and 8% of the population simultaneously at any given time. These values are low but plausible in situations where the proportion of individuals that can be treated is limited by the shortage of infrastructure or personnel to administer the control. Furthermore, when designing an optimal control coverage (optimal value of $\gamma$), selecting $\gamma$ high above the critical threshold $\gamma_c$ leads to the number of the infected individuals in the endemic state being much smaller than the number that can be treated, that is, resources (if allocated with such a high $\gamma$) will be wasted. This means that the optimal value of $\gamma$ is likely to be close to $\gamma_c$ and thus in the region where the non-trivial dynamics are important.

In the next chapter, we apply the understanding of the dynamics developed here to investigate optimal resource allocation between different populations with a shared resource pool.

\subsection{Author contributions}

This chapter is based on published work (Vyska and Gilligan, 2016). M.V. conceived, designed and implemented the study. M.V. and C.A.G. contributed to the discussion of the work and writing of the paper.
2.7 Appendix

2.7.1 A1. Dynamics of the standard SIS model

A full description of the dynamical behaviour of the SIS model can be found in most standard textbooks, for example Keeling and Rohani (2008) or Anderson and May (1991). Here we list the most important properties for convenience. The SIS model at its most basic is described by the equations:

\[ \dot{I} = \beta IS - \mu I \]
\[ S = 1 - I. \]

(2.71)  
(2.72)

Here \( S, I \) are the proportions of the susceptible and infectious individuals respectively. \( \beta \) is the transmission rate of the infection and \( \mu \) is the recovery rate. When \( \beta \leq \mu \), the model only has one disease-free equilibrium at \((S, I) = (1, 0)\) which is globally stable. When \( \beta > \mu \), the model has two equilibria. One is the same disease-free equilibrium which is now unstable, the other is a stable endemic equilibrium given by

\[ I^* = 1 - \frac{\mu}{\beta}. \]

(2.73)

2.7.2 A2. Dynamics of the standard SIRS model

The SIRS model at its most basic is described by the equations:

\[ \dot{S} = \sigma + \nu R - \sigma S - \beta IS \]
\[ \dot{I} = \beta IS - (\mu + \sigma)I \]
\[ \dot{R} = \mu I - (\nu + \sigma)R. \]

(2.74)  
(2.75)  
(2.76)

Here \( S, I \) and \( R \) and the proportions of the susceptible, infectious and recovered individuals respectively. \( \beta \) is the transmission rate of the infection, \( \sigma \) is the birth rate and the death rate, which are assumed to be equal, \( \mu \) is the recovery rate and \( \nu \) is the rate with which recovered individuals lose immunity and rejoin the susceptible class.
The system has one or two equilibria. One is the disease-free equilibrium $I^* = 0$, $R^* = 0$ and $S^* = 1$. The other, whenever it exists, is an endemic equilibrium given by:

$$S^* = \frac{1}{R_0}$$

$$I^* = \frac{\nu + \sigma}{\nu + \sigma + \mu} \left(1 - \frac{1}{R_0}\right)$$

$$R^* = \frac{\mu}{\nu + \sigma + \mu} \left(1 - \frac{1}{R_0}\right),$$

where we have defined the so-called basic reproductive number $R_0 = \beta/(\mu + \sigma)$. The basic reproductive number measures on average how many new infections an infectious individual will cause before recovering. When $R_0 \leq 1$, an epidemic is impossible, the disease-free equilibrium is stable and the endemic equilibrium does not exist. When $R_0 > 1$, the disease-free equilibrium is unstable and the endemic equilibrium is a stable global attractor.

2.7.3 A3. Another way to model the effects of control

Suppose that instead of increasing the recovery rate of infected individuals, the treatment instead reduces their infectious rate. In the absence of treatment, the infectious rate is $\beta$ while under treatment, it is $\beta_T < \beta$. We will use the SIS model to illustrate the effects of this. When $\gamma \geq I$, every infected host can be treated and therefore the equation is simply

$$\frac{dI}{dt} = (1 - I)(\beta_TI - \mu I).$$

(2.80)

However, when $\gamma < I$, we get

$$\frac{dI}{dt} = (1 - I)[\beta_T\gamma + \beta(I - \gamma)] - \mu I$$

(2.81)

$$= (1 - I)\beta I - [\mu - \gamma(\beta - \beta_T)]I - \gamma(\beta - \beta_T).$$

(2.82)

These equations have a similar structure as in (2.1) where the control is modelled by its effect on the recovery rate. Therefore, while the details will be different between these two ways of modelling the control, we can expect the qualitative behaviour to be the same.
Chapter 3

One-time allocation of a limited resource for disease control between multiple populations

3.1 Abstract

Spatial structure of the host population is an important factor for both the dynamics of disease spread but also for the design of control interventions. A natural question that arises when the resources for control are limited and insufficient to treat all the infected individuals is: How should the resources be distributed in the population? This leads naturally to the study of the interplay between the constrained control and spatial structure.

In this chapter, we extend the analysis from the previous chapter to investigate how to allocate a limited control resource optimally between multiple isolated populations. The control resources are allocated between the populations initially and cannot be reallocated later. This is a simplifying assumption relevant in situations when moving the control resource is difficult or expensive. It is also a natural starting point, because reallocating resource will never be free and therefore it is important to know how well we can control the epidemic without reallocating at all. We show that the optimal strategy must allocate resources fully to some populations and not at all to others, and we construct an approximate, but simple and intuitive method for how this allocation should be executed in practice. We show that it is possible to allocate a value and a
cost to fully treating each of the populations, and that the resources should be allocated so as to maximize the overall value while keeping the cost below a certain limit.

3.2 Introduction

One of the most important factors that influence the behaviour of epidemics is the spatial structure of the host population. While simple models often assume that all the hosts within the population interact homogeneously, for better understanding and more accurate predictions of the spread of disease, the spatial distribution of the hosts in the landscape needs to be considered. This is a fact well established in both human epidemiology (Dye and Day, 2003; Grenfell et al., 2001) and plant epidemiology (Condeso and Meentemeyer, 2007). There are many ways spatial structure can be introduced into an epidemic model ranging from modelling the hosts as vertices of a lattice (Rhodes and Anderson, 1996), to modelling interactions between hosts using more general networks (Keeling, 2005a; Keeling and Eames, 2005). However, a very common middle ground between the homogeneously mixed models (easy to study, but simplistic) and the full network models (difficult to study) are the so-called structured metapopulation models (Grenfell and Harwood, 1997; Lin and Li, 2014). They arise from the assumption that the population can be divided into a number of subpopulations. The epidemic is then assumed to spread homogeneously within each subpopulation, with interaction between the subpopulations. By interaction we mean transfer of infection between the subpopulations, either through the movement of infected hosts in the case of human or animal epidemics or through the transfer of inoculum in the case of plant disease.

Clearly, since spatial structure is important for the dynamics of the disease, it is also important for control interventions. The need for incorporating spatial structure into the models of control of epidemics has been recognised across plant, animal and human epidemiology, for example in modelling the spread and control of Rhizomania in the United Kingdom (Stacey et al., 2004), in comparing various control strategies during the UK foot and mouth epidemic (Keeling et al., 2003) or in modelling pulse vaccination of measles (Agur et al., 1993).

Our aim is to use the understanding of the models with constrained control developed in Chapter 2 to investigate allocation of limited control resource in a metapopulation consisting of $n$ subpopulations of varying sizes. We model the effect of the treatment as
in Chapter 2 and assume that it can be applied to infected individuals to shorten their infectious period. In this chapter, we assume that the resource can be allocated between the subpopulations only once, and cannot be reallocated later. This assumption would be appropriate, for example, when reallocating the resources is very costly, or logistically infeasible. It is also an important base case, because reallocating resource will never be completely free and therefore it is important to know how successfully we can control the epidemic with only allocating the resource once.

The existence of the shared resource pool from which the resource is allocated between the subpopulation gives rise to an indirect interaction between them. When more resources are allocated to one of the subpopulations, the disease prevalence in that subpopulation will decrease but there will consequently be less resource to be allocated into the other subpopulations. The proportions of infected individuals in the different subpopulations are therefore anti-correlated. Our goal is to study this indirect interaction, and for this reason we ignore the direct coupling between the subpopulations. This is a simplifying assumption that allows us to make analytical progress and develop intuitive understanding of the optimal resource allocation. We note that optimal “one-time” allocation of a limited resource between multiple independent populations has been considered before with the effect of the control modelled as reducing the contact rate (Brandeau et al., 2003). In addition to the different way of modelling the control, the authors also considered different objective functions, namely the number of new infections and the quality adjusted life years gained. They focused on analysing the convexity or concavity of the objective function in various regions of the parameter space and while they provided some characterization of the optimal allocation strategies, for example whether they are bang-bang or not, they did not give any intuitive and simple-to-use guidelines for finding the optimal strategy. Developing such intuitive insight into the construction of an optimal resource allocation is our aim in this chapter.

While the assumption of no coupling between the populations is an idealization for most real pathogen systems, it is plausible in plant epidemiology where the different subpopulations can correspond to different host species each with its own pathogen. We refer to this situation as a multiple threat model and it is exemplified by the large number of different host species and pathogens in the UK Plant Health Risk Register (Baker et al., 2014). Mathematically, we model this by allowing for the disease
parameters, the cost and the effectiveness of the treatment and the value per host to vary between the subpopulations.

Because the resource is fixed after the initial allocation, what may initially be an efficient treatment allocation can later turn into an allocation where considerable part of the resource pool is wasted in populations where the disease prevalence has long dropped to very low values. Thus the resource allocation at the beginning must take into account the future dynamics of the disease and in particular the long term behaviour of the disease prevalence. For this reason we focus on the long term behaviour and consequently on the equilibrium dynamics.

Following this, we formulate the optimization problem by defining the objective function as the equilibrium number of infected hosts. This is also a simplifying assumption which allows us to make analytical progress and therefore fits with the broader theme of obtaining intuitive insight into the problem. We use both the SIS and SIRS model structure and derive a set of conditions that the optimal resource allocation must satisfy. We use these conditions to arrive at an approximation to the exact optimum that has a simple and intuitive interpretation. We show how this approximation can be implemented, and we compare it both to the exact optimum and to a number of simple and intuitive “benchmark” allocation strategies.

3.3 Single threat metapopulation with no coupling

In this section, we analyse the optimal allocation of a limited resource to a host metapopulation with no coupling. In the Model analysis subsection, we derive an approximate optimal allocation strategy. In the Results subsection we show this strategy is nearly optimal and compare it with several other candidate allocation strategies.

3.3.1 Model assumptions

In this chapter, we focus on the SIS model structure from the previous chapter. We consider \( n \) separate host populations, each with its own SIS epidemic, which share a common resource pool. The dynamics of the system in each population are given by equation (2.1) from Chapter 2. In this section, to derive the key ideas and to illustrate the methodology, we assume that the separate populations are separated
spatially and that the probability that an infected host in the population $i$ infects a susceptible host in the population $j$ can be neglected. We refer to this case as a single threat metapopulation, since it assumes one host type with one disease. Therefore the parameters do not vary between the different populations.

We assume that the control resource can be allocated between the populations only initially at time $t = 0$ and cannot be reallocated later. This would be relevant for example if reallocation is very expensive, but it also makes sense as a starting assumption because of its simplicity. We will relax this assumption in the next chapter. Suppose that the size of the $i^{th}$ population is $N_i$ and that we allocate a proportion $x_i \in [0, 1]$ of the resource to it. The objective function we seek to minimize by allocating the resource is given by

$$J(x_i) = \sum_i N_i I_i^\infty(x_i),$$

where $I_i^\infty$ denotes the long-term endemic (or disease-free) equilibrium reached by the system in the population $i$ and depends on $x_i$. Thus, we seek to minimize the long term number of infected hosts.

The objective function in (3.1) is natural and easy to understand. Mathematically, it is in fact a special case of a more general, commonly used (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011) objective function defined as the average over some time interval $(0, T)$ of the number of the infected individuals

$$J = \frac{1}{T} \sum_i N_i \int_0^T I_i(t) dt.$$  

The objective function we consider here can be interpreted as the average over a time interval $(0, T)$ when the time horizon $T$ becomes very large, in the limit $T \to \infty$. We expand on this point further in the discussion at the end of the chapter.

See table 3.1 for the description of the parameters and the variables used in this chapter.

### 3.3.2 Model analysis

Consider a system of $n$ non-interacting populations with sizes $N_i$ and initial conditions $I_0^{(i)}, i \in \{1, 2, \ldots, n\}$. Following the previous chapter, the dynamics in each of the
One-time allocation of a limited resource for disease control between multiple populations

<table>
<thead>
<tr>
<th>Parameter/variable (default value)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2)</td>
<td>Rate of infection.</td>
</tr>
<tr>
<td>$\mu$ (1)</td>
<td>Recovery rate of the infected hosts. Its value is set to 1 throughout this chapter, without loss of generality. One unit of time therefore corresponds to one infectious period.</td>
</tr>
<tr>
<td>$\eta$ (0.8)</td>
<td>Additional recovery rate provided by the treatment.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Proportion of the population which can be treated simultaneously.</td>
</tr>
<tr>
<td>$C_T$</td>
<td>Endemic disease prevalence given full treatment, $1 - \frac{1 + \eta}{\beta}$.</td>
</tr>
<tr>
<td>$C_0$</td>
<td>Endemic disease prevalence given no treatment, $1 - \frac{1}{\beta}$.</td>
</tr>
<tr>
<td>$r$</td>
<td>$\beta - \mu$.</td>
</tr>
<tr>
<td>$\gamma^{(i)}_C$</td>
<td>Proportion of the hosts in population $i$ that require simultaneous treatment necessary for saturation.</td>
</tr>
<tr>
<td>$N_i$</td>
<td>Size of the $i^{th}$ population.</td>
</tr>
<tr>
<td>$M$</td>
<td>The maximum available amount of resources; the maximum number of hosts that can be treated simultaneously.</td>
</tr>
<tr>
<td>$x_i$</td>
<td>Proportion of the resource that is allocated to the population $i$.</td>
</tr>
<tr>
<td>$m_0$</td>
<td>The index of the population that has the smallest size amongst the unsaturated populations.</td>
</tr>
<tr>
<td>$c_i$</td>
<td>The cost of treating one host in the population $i$, in the multiple threats model.</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>The value of one host in the population $i$, in the multiple threats model.</td>
</tr>
<tr>
<td>$n$</td>
<td>The number of subpopulations.</td>
</tr>
</tbody>
</table>

Table 3.1 Table showing the parameters and some of the variables used, together with their descriptions. Some of the derived parameters are also included, for convenience.

The populations is given by

$$\frac{dI}{dt} = \beta I (1 - I) - I - \eta \min(\gamma, I),$$

(3.3)

where $\beta$ is the rate of infection, $\eta$ is the additional recovery rate provided by the treatment and $\gamma$ is the proportion of the population that can be treated at any given time.

We refer to a population as saturated whenever the resources allocated to it cause the long term prevalence to be $C_T$ (the endemic prevalence given full treatment). Therefore, using the results from the previous chapter, the $\gamma^{(i)}_C$ necessary for saturation...
3.3 Single threat metapopulation with no coupling

is given by

\[ \gamma_c^{(i)} = C_T \quad \text{for } I_0^{(i)} < C_T \]  
(3.4)

\[ \gamma_c^{(i)} = \frac{\beta}{\eta} I_0^{(i)} (C_0 - I_0^{(i)}) \quad \text{for } C_T < I_0^{(i)} < C_0/2 \]  
(3.5)

\[ \gamma_c^{(i)} = \gamma_{\text{max}} \quad \text{for } I_0^{(i)} > C_0/2. \]  
(3.6)

From now on we refer to one unit of resource as the amount necessary for the treatment of one host. Suppose that the maximum number of hosts that can be treated at any given time is \( M \), or in other words, we only have \( M \) units of the resource. Consider an allocation strategy that allocates \( x_i \) resource into population \( i \), where \( x_i \in [0, 1] \) and \( \sum_i x_i = 1 \). Clearly the proportion of individuals that can be treated in the population \( i \) is \( \gamma_i = x_i M / N_i \).

The objective function to be minimized is the total number of infected individuals in the endemic equilibrium. The optimization problem can be stated as follows: we are looking for an allocation that puts a proportion \( \{x_i\} \) of the total resource into population \( i \), in order to minimize the objective function given by

\[ J = \sum_i N_i I_\infty^{(i)}(x_i) \]  
(3.7)

\[ = \sum_{i \in S} C_T N_i + \sum_{i \notin S} \frac{N_i}{2} \left( C_0 + \sqrt{C_0^2 - \frac{4\eta M x_i}{\beta N_i}} \right), \]  
(3.8)

where the set \( S \subset \{1, \ldots, n\} \) is the set of populations that are saturated under the allocation \( \{x_i\} \). To make progress, we first consider allocations under which no population is saturated. Let \( \{x_i\} \) be such an allocation. The objective function is then given by

\[ J = \sum_i \frac{N_i}{2} \left( C_0 + \sqrt{C_0^2 - \frac{4\eta M x_i}{\beta N_i}} \right). \]  
(3.9)

Setting \( x_n = 1 - x_1 - \ldots - x_{n-1} \) and differentiating with respect to \( x_i, i \in \{1, 2, \ldots, n-1\} \), we get

\[ \frac{\partial J}{\partial x_i} = -\frac{\eta M}{\beta} \left( \frac{1}{\sqrt{C_0^2 - \frac{4\eta M}{\beta N_i} x_i}} - \frac{1}{\sqrt{C_0^2 - \frac{4\eta M}{\beta N_n} (1 - x_1 - \ldots - x_{n-1})}} \right). \]  
(3.10)
Solving $\partial J/\partial x_i = 0$ gives

$$x_i = \frac{1 - x_1 - \ldots - x_{n-1}}{N_n} = \frac{x_n}{N_n}, \quad (3.11)$$

This system of equations has a unique solution $\{x^*_i\}$

$$x^*_i = \frac{N_i}{N_{tot}}, \quad (3.12)$$

where $N_{tot} = \sum_i N_i$. Thus there is a unique local extremum of $J$ when the resource allocation is proportional to the population size. We will prove that this point cannot be the solution to our optimization problem by demonstrating that it is not even a local minimum.

We denote the matrix of second derivatives $DJ$. For ease of notation, denote further $DJ^* = DJ|_{x=x^*}$. By differentiating $J$ twice, we obtain

$$DJ_{ii}^* = \frac{-2\eta^2 M^2}{\beta^2} \frac{1}{(C_0^2 - \frac{4\eta M}{\beta N_{tot}})^{3/2}} \left[ \frac{1}{N_i} + \frac{1}{N_n} \right], \quad (3.13)$$

$$DJ_{i\neq j}^* = \frac{-2\eta^2 M^2}{\beta^2} \frac{1}{(C_0^2 - \frac{4\eta M}{\beta N_{tot}})^{3/2}} \frac{1}{N_n}. \quad (3.14)$$

To simplify the notation, we define a new parameter $A$ as

$$A = \frac{2\eta^2 M^2}{\beta^2} \frac{1}{(C_0^2 - \frac{4\eta M}{\beta N_{tot}})^{3/2}} > 0. \quad (3.15)$$

The matrix $DJ^*$ is therefore given by

$$DJ^* = -\frac{A}{N_n} \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix} + \begin{pmatrix} -\frac{A}{N_1} & 0 & \ldots & 0 \\ 0 & -\frac{A}{N_2} & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \ldots & 0 & -\frac{A}{N_{n-1}} \end{pmatrix}. \quad (3.16)$$

To show that the fixed point is not a local minimum, we need to prove that the matrix $DJ^*$ has at least one negative eigenvalue (Arrowsmith and Place, 1992). The matrix $DJ^*$ is real, symmetric and all of its components are negative. Therefore, its trace is negative and since the sum of the eigenvalues is equal to the trace, there must be
at least one negative eigenvalue. This means that there exists a vector \( x \) such that \( x^T DJ^* x < 0 \) and so the fixed point \( x^* \) cannot be a local minimum. Since this is the only local extremal point of the system, the allocation strategy minimizing \( J \) must lie on the boundary, where some of the populations are saturated. This argument can be repeated iteratively. Suppose the optimum lies in the subspace of possible strategies where population \( m \) is saturated. The resources remaining after saturating \( m \) are \( M - N_m \gamma_c^{(m)} \). After we saturate \( m \), the problem is to minimize \( J \) in the remaining populations given the remaining resources. This is qualitatively the same as before, and so the solution must again lie on the boundary. Thus one of the remaining populations should be saturated. This argument can be repeated until we don’t have enough resources to keep saturating. Therefore, the optimal allocation is as follows: saturate some subset \( S \) of all the population such that no further saturating is possible and then allocate all the remaining resources into the remaining unsaturated populations. This raises two main questions.

1. How should the remaining resource be allocated?

2. How should \( S \) be chosen?

We start with the first question. This is the same as the problem of where to allocate resources if we cannot saturate any of the populations. That is, suppose we have a set of populations \( \{N_i\} \) and the amount of resources \( M \) such that

\[
M < N_i \gamma_c^{(i)} \quad \forall i.
\]  

(3.17)

This conditions merely expresses the fact that none of the populations can be saturated. We have the following result.

**Theorem 3.** When none of the populations can be saturated, the optimal allocation is \( \gamma_m = M/N_m \) for some \( m \) and \( \gamma_i = 0 \ \forall i \neq m \).

**Proof.** We prove this by contradiction. Suppose that in the optimal allocation, some subset \( X \) of the populations share the resources, that is more than one population have nonzero amount of resources allocated to it. Since there is no coupling, we can consider the subset \( X \) in isolation. Since none of the populations in \( X \) is saturated and none has zero resources allocated to it, as far as \( X \) is concerned, this allocation is an interior one. We proved above however, that there can be no interior local minimum.
One-time allocation of a limited resource for disease control between multiple populations of any system of any number of populations. Therefore we can move some of the resources to decrease the objective function and this means that the allocation cannot be optimal.

When all the resources are allocated to population $m$, the objective function becomes

$$J_m = C_0(N_{\text{tot}} - N_m) + \frac{N_m}{2} \left( C_0 + \sqrt{C_0^2 - \frac{4\eta M}{\beta N_m}} \right)$$

(3.18)

$$= C_0N_{\text{tot}} - \frac{1}{2}C_0N_m + \frac{1}{2}N_m\sqrt{C_0^2 - \frac{4\eta M}{\beta N_m}}.$$  

(3.19)

We will show that $J_m$ is strictly increasing as a function of $N_m$. Consider the first derivative of $J_m$ with respect to $N_m$. This is given by

$$J'_m = -\frac{C_0}{2} + \frac{1}{2} \sqrt{C_0^2 - \frac{4\eta M}{\beta N_m}} + \frac{1}{2} \frac{1}{\sqrt{C_0^2 - \frac{4\eta M}{\beta N_m}}} \frac{4\eta M}{\beta N_m}.$$ 

(3.20)

To simplify the notation, define a new variable $t = \frac{4\eta M}{\beta N_m}$. So $t \in (0, C_0^2)$. At the boundary of this interval, we can calculate the derivative directly. We get that $J'_m(t = 0) = 0$ and $J'_m(t = C_0^2) = \infty$. Furthermore solving $J'_m(t) = 0$ gives

$$0 = -C_0 + \sqrt{C_0^2 - t} + \frac{t}{\sqrt{C_0^2 - t}}$$

(3.21)

$$= -C_0\sqrt{C_0^2 - t} + C_0^2$$

(3.22)

$$= C_0(C_0 - \sqrt{C_0^2 - t}).$$

(3.23)

From this we can see that the only solution is $t = 0$ and so $t = 0$ is the only intersection with the $t$-axis. Therefore $J'_m > 0 \forall t \in (0, C_0^2)$. This means that to minimize the objective function, we should allocate the leftover resource to the remaining population with the smallest size, that is to the population with index $m_0$ such that

$$m_0 = \arg\min\{N_i | i \text{ not saturated}\}. $$

(3.24)

Now we can turn to the second question, that is, if we can saturate some of the populations, which should we pick? Suppose we saturate a set $S$ of the populations, so $S = \{i | i \text{ is saturated}\}$. Mathematically, we want to know which choice of $S$ minimizes the objective function. We refer to the set of the remaining unsaturated populations
as $R$, so $R = \{i|i \not\in S\}$. The resource left after saturating $S$, $M_R$, is given by

$$M_R = M - \sum_{i \in S} N_i \gamma_c^{(i)}. \tag{3.25}$$

The restriction on $S$ that no more populations can be saturated can be put in mathematical terms as

$$M_R < \gamma^{(i)}_c N_i, \quad \forall i \in R. \tag{3.26}$$

The objective function, as a function of $S$, is given by

$$J(S) = C_T \sum_{i \in S} N_i + C_0 \sum_{i \in R \setminus \{m_0\}} N_i + \frac{1}{2} N_{m_0} \left( C_0 + \sqrt{C_0^2 - \frac{4 \eta M_R}{\beta N_{m_0}}} \right), \tag{3.27}$$

where $m_0 = \text{argmin}\{N_i|i \in R\}$ is the index of the population receiving the resources left after saturation. The task is to minimize $J(S)$ by an appropriate choice of $S$. This is a difficult problem, mainly because it is hard to capture how $m_0$ depends on the choice of the set $S$.

We approach the optimization problem (3.27) by considering an approximation where we neglect the term $4 \eta M_R / \beta N_{m_0}$. This amounts to basically ignoring the resources that are left over after saturating the populations in $S$ when selecting the optimal $S$. The error in the objective function arising from this approximation is at most $\frac{1}{2} C_0 N_m$.

We can now make progress in finding the optimal choice of $S$. Let the approximate objective function be $J_a(S)$. Then

$$J_a(S) = C_T \sum_{i \in S} N_i + C_0 \sum_{i \in R} N_i. \tag{3.28}$$

The problem is to minimize $J_a(S)$ under the condition on $S$ that $\sum_{i \in S} N_i \gamma_c^{(i)} < M$. Let $N_S = \sum_{i \in S} N_i$. Then we can rewrite $J_a(S)$ as

$$J_a(S) = C_T N_S + C_0 (N_{\text{tot}} - N_S) \tag{3.29}$$

$$= C_0 N_{\text{tot}} - (C_0 - C_T) N_S. \tag{3.30}$$

To minimize $J_a$, we need to maximize $N_S$. In other words, we need to saturate a set $S$ such that the number of hosts in the saturated populations is as large as possible. We
One-time allocation of a limited resource for disease control between multiple populations can rephrase the problem as follows:

$$\max \sum_{i \in S} N_i \quad \text{subject to} \quad \sum_{i \in S} w_i < M, \quad (3.31)$$

where the weights are $w_i = N_i \gamma_i^{(i)}$. This is a variation on the knapsack problem well known in computer science (Martello and Toth, 1990; Skiena, 1999; Kellerer et al., 2004). The problem can be stated as follows: given a set of items with various values and weights, which items should be selected so that the overall value is maximised but the overall weight does not exceed a specified maximum allowed weight. Its uses are very wide ranging, for example finding the optimal way to cut raw materials (Kellerer et al., 2004), construction of investment portfolios (Kellerer et al., 2004) or the construction and scoring of tests (Feuerman and Weiss, 1973). In our case, we have to solve the so called 0-1 type knapsack problem, because each item is either included or not (each population either is saturated or not).

One of the standard approaches to computationally solving this type of the knapsack problem and the one we use here is the so-called Meet in the middle method (Horowitz and Sartaj, 1974). This algorithm is a variation on the brute-force approach searching through all the possible subsets $S$. The Meet in the middle algorithm consists of the following steps:

**Meet in the middle algorithm**

1. Split $\{1, 2, \ldots, n\}$ into two subsets of approximately equal size, $A$ and $B$.

2. Find the total weight and the total value ($\sum_i N_i$) of each subset of $A$ and each subset of $B$.

3. For each subset of $A$, find the subset of $B$ which maximizes the value with the combined weight less than the limit $M$. This can be done efficiently as follows. We can first sort the subsets of $B$ by weight. Then we remove all subsets of $B$ which have higher weight but smaller value than some other subset of $B$. That is, if for two subsets of $B$, $S_1$ and $S_2$ we have that $\text{weight}(S_1) \geq \text{weight}(S_2)$ but $\text{value}(S_1) \leq \text{value}(S_2)$, we remove $S_1$, because it will definitely not be in the optimal selection. After this procedure, the subsets of $B$ are sorted both in weight and value. To find the subset of $B$ which for some given subset of $A$ maximizes
the value while having combined weight less than $M$, we can just use binary search.

Both steps (2) and (3) require $O(n^{2n/2})$ operations and so the whole algorithm requires $O(n^{2n/2})$ operations. Computational time for a brute force approach to finding the optimal set $S$ exactly is $O(n2^n)$. Therefore, the *Meet in the middle* algorithm is significantly faster than the naive brute force search. There are other algorithms for solving this type of a knapsack problem, for example algorithms based on dynamic programming (Cormen et al., 2001) which can sometimes be faster, but for our purposes here the *Meet in the middle* algorithm is sufficient.

### 3.3.3 Results

The main advantage of this approximate approach to the solution over just finding the optimum numerically is that it has a simple and satisfying interpretation. Under the knapsack approximation,

we need to maximize the sum of the sizes of the saturated populations.

While when there are many different populations, the knapsack problem must be solved using a computer, thanks to its simple interpretation it is straightforward to solve when the number of populations is small. Another advantage of the knapsack approximation is that it does not require the detailed behaviour of the fixed point $I_D$ for implementation, whereas optimizing the full objective function from (3.27) does. In fact, the only information needed to implement the approximate optimal allocation is the sizes of the populations and the critical amounts of resources needed for saturation. These are given below in two important scenarios.

Firstly, when the disease has just invaded the population, the initial proportions of infected hosts are small and when $C_T > 0$, they are generally smaller than $C_T$. In such a case, the critical amounts of resources are simply equal to $N_iC_T$, so we only need to know the endemic equilibrium under full treatment. On the other hand, when the disease has been endemic in the population for a while, that is the initial conditions are endemic, the initial proportions of infected hosts are all equal to $C_0$. In such a
One-time allocation of a limited resource for disease control between multiple populations

situation, the critical proportions are given by (see equation 3.6)

\[ N_i \gamma_i^{(i)} = \frac{N_i C_0^2}{4 (C_0 - C_T)}. \]  

Therefore, we only need to know the endemic equilibrium under full treatment and under no treatment.

To demonstrate the effectiveness of the knapsack approximation, we introduce several other allocation strategies and compare their performance with the knapsack approximation with respect to the exact optimal allocation strategy. We obtain the exact optimal strategy by finding the optimal saturation set \( S \) that minimizes the objective function (3.27) numerically. Specifically, we use the brute force approach and scan through all the subsets \( S \subset \{1, 2, \ldots, n\} \) which can be saturated given the resource constraint and selecting the one that minimizes the objective function. This has a computational complexity proportional to \( n2^n \), however this is not a problem here since the largest \( n \) we consider is 11.

The other allocation strategies we consider are:

- **Proportional allocation.** The amount of resource allocated to the \( i^{th} \) population is proportional to its size \( N_i \), so

  \[ N_i \gamma_i = \frac{N_i}{N_{\text{tot}}}. \]

- **Equal allocation.** The same amount of resource is allocated to each of the populations, so \( N_i \gamma_i = 1/n \).

- **The “Greedy” Allocate to the largest strategy.** We look at which population we need to saturate to achieve the greatest decrease in the objective function. Since saturating population \( i \) amounts to decreasing the objective function by \( (C_0 - C_T)N_i \), this is the same as saturating the largest population. Therefore the allocation strategy is as follows: saturate the largest population. If there are resources left, saturate the second largest and continue in this manner until saturation cannot be achieved. At that point, put all the remaining resources into the largest population not yet saturated. For this reason we call the greedy allocation the “allocate to the largest” strategy.
3.3 Single threat metapopulation with no coupling

Fig. 3.1 The x-axis here represents the eleven populations. The green colour represents initial susceptible hosts and the red colour the infected ones. The columns correspond to the different populations and the height of the columns corresponds to the population size.

- **Allocate to the smallest strategy.** As the name suggests, it is the opposite of the above, in that we saturate the smallest population and then repeat until we cannot saturate any more, at which point we allocate the remaining resources to the smallest of the remaining populations.

Comparing the knapsack to the exact optimum is necessary in order to demonstrate that the approximation is a good one. The other test strategies above were selected for their simplicity and intuitive appeal. We compare them to the knapsack approximation to demonstrate that meaningful gains in the objective function can be achieved by correct resource allocation as opposed to one of the naive, simple strategies above.

The comparison with the exact optimum reveals that the knapsack approximation performs very well. The errors are on the order of tenth of a percent to a percent. We illustrate the behaviour on a concrete, generic example consisting of 11 populations. The sizes selected range from 100 to 400 and the initial proportions of infected hosts were selected at random, uniformly from the interval $(0, 1)$. The sizes together with the initial conditions are visualized in Figure 3.1. The green colour represents initial susceptible hosts and the red colour the infected ones. The columns correspond to the different populations and the height of the columns corresponds to the population size.
One-time allocation of a limited resource for disease control between multiple populations

Fig. 3.2 Comparison of several allocation strategies as the maximum available resources are varied (and therefore as different saturation sets $S$ become allowed). Note that the knapsack approximation is almost as good as the exact optimum. The allocate to the smallest strategy usually performs better than the allocate to the largest strategy (except when the resources are high). The proportional and equal allocation strategies are far worse than the rest. This shows that even if we saturate the wrong populations, saturating some populations is still better than spreading the resources equally or proportionally. The population sizes and the initial conditions are as in Fig. 3.1, the other parameter values are $\beta = 2 \text{ time}^{-1}$ and $\eta = 0.8 \text{ time}^{-1}$.

Note that as far as the model is concerned, the absolute magnitude of the population sizes does not matter, only their ratios ($N_i/N_j$).

In Figure 3.2 we plot the performance of the above allocation strategies as the amount of resource available ($M$) is varied. Note that the knapsack approximation is almost as good as the exact optimum. In fact the largest percentual error in the achieved objective function caused by using the approximation is about 1% (Figure 3.3). Both the allocate to the largest and allocate to the smallest strategies are noticeably suboptimal, with the allocate the smallest beating the allocate to the largest in most cases. The failure of the allocate to the largest strategy when compared with the optimal or knapsack allocations reflects the interaction between the populations through the common shared resource pool because greedily maximizing the profit from the first allocation does not take into account how that influences all the other possible allocations and their profits. The equal and proportional allocation strategies significantly under-perform when compared with the others. In Figure 3.3 we plot the percentage error of the
Fig. 3.3 The errors of the three allocation strategies. Note that the knapsack approximation is mostly optimal, with a very small error for some values of the maximum available resources. These errors occur when it is advantageous to achieve smaller sum of the sizes of fully saturated populations in exchange for more remaining resources allocated to the smallest of the remaining populations. The knapsack approximation misses these cases since it neglects the resources remaining after saturation. The parameter values are the same as before, that is $\beta = 2 \text{ time}^{-1}$ and $\eta = 0.8 \text{ time}^{-1}$. 
Fig. 3.4 When the treatment is weak (here $\eta = 0.3$), the differences between the knapsack approximation and the allocate to largest/smallest strategies is very small and they overlap in the plot. As before, $\beta = 2$.

knapsack approximation as well as the allocate to the largest and allocate to the smallest strategies. The error is calculated as

$$\text{error} = 100 \times \frac{J_{\text{strategy}} - J_{\text{optimal}}}{J_{\text{optimal}}}.$$

(3.34)

This plot once again demonstrates that the knapsack approximation is very good and much better than the more naive approaches.

The results for other populations and initial conditions are not qualitatively different. When $\eta < r/2$, that is when the treatment is weak, the differences between the knapsack approximation and the allocate to largest/smallest strategies become much smaller. This is illustrated in Figure 3.4. It can be seen that the results for the three leading strategies overlap. The reason is that when the treatment is weak, it matters less how precisely we allocate the resources.

**Knapsack approx. vs the exact optimum: detailed analysis**

We checked the performance of the knapsack approximation against the exact optimal strategy in a wide range of cases, and we summarize the results in the Table 3.2. In the table, we chose a range of values for $n$ (number of different populations) and for
Table 3.2 Table showing the percentual worst-case error of the knapsack approximation when compared with the exact optimum. The error was computed as follows: for each case, a random vector of the population sizes was generated with \( N_1 = 100 \) and the others chosen uniformly from \((100, 1000)\). Then, we ran through the possible resource limits \( M = 1, 2, 3, \ldots \) and selected the one that led to the largest error (therefore the worst-case error). This process was repeated 10000 times (1000 in the case \( n = 10 \)) and the summary statistics are shown in the table. The mean worst case error is quite low and in fact, for example in the case \( n = 3 \) and \( \eta = 1.2 \), the worst-case error was 0 in 95% of the cases. The other parameter value is \( \beta = 2 \).

For each combination of values, we generate the vector of population sizes \( N_i \) many times at random such that \( N_1 = 100 \) and for each \( i > 1 \), \( N_i \) is chosen uniformly from the interval \((100, 1000)\) and rounded up to the nearest integer. For each case we run through all the integer values of \( M \) (the resource limit) and calculate the worst-case relative percentual error of the knapsack approximation when compared with the exact optimum. The summary statistics in the table are then statistics of this worst-case error. So, for example, the mean corresponds to the mean over different landscapes of the worst-case error with respect to the resource limit. The maximum error in Table 3.2 is the global worst case, with respect to both the landscape and the resource limit. As can be seen in the Table 3.2, the mean is very low even though the maximum can get up to 10%. In fact, in most cases, the worst case error is 0, which is why the mean is so low. For example, when \( n = 3 \) and \( \eta = 1.2 \) the worst case error is 0 in roughly 95% of the landscape realizations. In Figure 3.5 we show the values of the worst case error in 10000 different cases of the random landscape realizations. The figure demonstrates that the cases with high error are rare.

In order to better understand why the errors occur, we also plotted 30000 realizations of the population sizes \( N_2, N_3 \) (for \( n = 3 \)) as points with colour density representing the magnitude of the error. The diagram can be found in Figure 3.6. In the figure, we can see that the cases with large error lie close to the lines \( N_3 = N_2 \pm 100 \). In
One-time allocation of a limited resource for disease control between multiple populations

Fig. 3.5 The distribution of the worst-case error of the knapsack approximation as the landscape (population sizes) is varied randomly. By worst-case we mean given a particular realization of the landscape sizes, we pick the resource limit which gives the worst error. The parameter values are $n = 3$, $\beta = 2$, $\eta = 1.2$ and the random population sizes are generated according to the procedure described in the text.

Fig. 3.6 The position of the red dots corresponds to the random realization of the population sizes $N_2$ and $N_3$ while the darker their colour, the greater the worst-case error. The plot reveals that the largest errors lie along the lines $N_3 = N_2 \pm N_1$. As before, $n = 3$, $\beta = 2$, $\eta = 1.2$
fact, since $N_1 = 100$, the cases with large error have $N_3 = N_1 + N_2 - \epsilon$, that is one of the sizes is just a little bit smaller than the sum of the other two. Furthermore, the largest error occurs when the resource limit $M$ is just below the amount necessary to saturate both the populations 1 and 3, but enough to saturate the populations 1 and 2. Intuitively, this makes sense since this maximizes the amount of resources left after saturation and the knapsack approximation effectively neglects the resource left after saturation when choosing which subset to saturate. Overall this demonstrates that obtaining a large error requires a carefully tuned situation. We give the supporting analysis in the Appendix A1. When the number of populations is greater than three, the criterion for the knapsack approximation to have a large error turns out to be the same. It happens precisely when there is a subset of the populations of size three $\{N_i, N_j, N_k\}$, in which $N_k = N_i + N_j - \epsilon$ and the resources are enough to saturate all the other populations so that we are effectively in the $n = 3$ case.

We note that unfortunately, it is not possible to derive a simplified heuristic for solving the knapsack approximation problem. It is tempting to consider a greedy strategy where we look at the ratios of the values $(N_i)$ to weights $(\gamma^c_i N_i)$, and saturate the populations from the highest ratio to the lowest, until the resources run out. This seems sensible, since the population with the highest ratio of the value to weight provides the largest value per one unit of resource invested. However, it is easy to show that such an approach can be almost arbitrarily wrong. As an illustration, consider the following example. Suppose the maximum allowable weight is $W = 50$ and the value/weight pairs are given by $(v_1, w_1) = (6, 1), (v_2, w_2) = (100, 20)$ and $(v_3, w_3) = (120, 30)$. The value to weight ratio is the largest for the first pair and once we include the first pair, the maximum value we can get without the combined weight going over $W$ is $v_1 + v_3 = 126$. On the other hand, the optimal solution is obviously combining the second and the third item, which yields total value $v_2 + v_3 = 220$, significantly larger than the value obtained by first picking the pair with the highest value to weight ratio.

While the knapsack approximation cannot be found using a simple heuristic, it is easier to find than the exact optimum and it has an intuitive interpretation. It is also more robust; in some cases the intuition behind it can be generalized to more complicated models, as we demonstrate in the next section.
3.4 Multiple threats metapopulation

In this section, we adapt the analysis from the previous section to the problem of optimally allocating resources between multiple threats which share a common resource pool. This is especially relevant for plant disease, where we can have several different types of crops each with its own pathogen.

3.4.1 Model assumptions

In this section the \( n \) different populations correspond to \( n \) different host species with \( n \) different diseases, or pathogen strains, each exclusive to its host species. In this scenario the assumption of no coupling between the populations becomes natural. The model assumptions are the same as in the single threat scenario, except now, in general, the epidemic in each population is described by a different set of parameters. Furthermore we allow for the hosts in different populations to have a different value, which we denote by \( \alpha_i \) and the treatment in each populations to cost a different amount of resource per host treated, \( c_i \). The objective function is then given by

\[
J(x_i) = \sum_i \alpha_i N_i I_i^\infty(x_i), \tag{3.35}
\]

while the resource constraint becomes

\[
\sum_i c_i N_i \gamma_i \leq M. \tag{3.36}
\]

Note that mathematically, the single threat metapopulation is a special case of the multiple threats metapopulation with all parameters identical across the different populations and \( \alpha_i = c_i = 1 \).

3.4.2 Model analysis

The model we will consider in this section is given by

\[
\frac{dI_i}{dt} = \beta_i (1 - I_i) I_i - I_i - \eta_i \min(\gamma_i, I_i), \tag{3.37}
\]
where the infection rate and the treatment rate are now dependent on the population. Likewise, the parameters $C_0$ and $C_T$ become $C_0^{(i)}$ and $C_T^{(i)}$. We now repeat the analysis from the previous section and show that the optimal solution can again be found approximately as a type of a knapsack problem with a new set of values and weights.

We begin by showing that as before, there is only one interior local extremum and it cannot be a minimum. Consider an allocation strategy $x_i$ which does not saturate any of the populations. The objective function is given by

$$J(x_1, \ldots, x_{n-1}) = \sum_i \frac{N_i \alpha_i}{2} \left( C_0^{(i)} + \sqrt{(C_0^{(i)})^2 - \frac{4 \eta_i M x_i}{\beta_i N_i c_i}} \right), \quad (3.38)$$

where $x_n = 1 - x_1 - \ldots - x_{n-1}$. To find the local extrema, we need to solve the set of equations $\partial J/\partial x_i = 0$. This yields that for each $i$, we have

$$\frac{\beta_i c_i}{\alpha_i \eta_i} \sqrt{(C_0^{(i)})^2 - \frac{4 \eta_i M}{\beta_i c_i N_i} x_i} = \text{const.} \equiv S. \quad (3.39)$$

To find the constant $S$ and hence the unique local extreme $x^*$, we use the condition $\sum_i x_i^* = 1$. This gives a solution as

$$x_i^* = a_i - \frac{\sum_k a_k - 1}{\sum_k b_k} b_i, \quad (3.40)$$

where

$$a_i = \frac{N_i (C_0^{(i)})^2 \beta_i c_i}{4 \eta_i M} \quad (3.41)$$

$$b_i = \frac{\eta_i N_i \alpha_i^2}{4 M \beta_i c_i}. \quad (3.42)$$

To show that this point cannot be a local minimum and therefore it cannot be the solution to the optimization problem, we need to find the $(n-1) \times (n-1)$ matrix of the second derivatives $DJ^*$. Differentiating $J$ twice yields

$$DJ_{ii}^* = -\frac{2 M^2}{S^{3/2}} \left( \frac{\beta_n c_n}{\eta_n N_n \alpha_n^2} + \frac{\beta_i c_i}{\eta_i N_i \alpha_i^2} \right) \quad (3.43)$$

$$DJ_{i \neq j}^* = -\frac{2 M^2}{S^{3/2}} \frac{\beta_n c_n}{\eta_n N_n \alpha_n^2}. \quad (3.44)$$
One-time allocation of a limited resource for disease control between multiple populations

This is again a real, symmetric matrix with negative entries and therefore it has a negative eigenvalue. This implies that the point $x^*$ cannot be a local minimum. We have thus shown that the optimal allocation must again lie on the boundary, that is we need to saturate populations until further saturation is not possible and then allocate all the remaining resources to a single population.

We start with the problem of where to allocate the remaining resources. Suppose we do not have enough resources to saturate any of the populations. If we choose to allocate them to the population $m$, the objective function will be

$$J(m) = \sum_i n_i \alpha_i C_0^{(i)} N_i - \frac{1}{2} n_m C_0^{(m)} N_m + \frac{1}{2} n_m N_m \sqrt{(C_0^{(m)})^2 - \frac{4\eta_m M}{\beta_m c_m N_m}}. \quad (3.45)$$

To minimize the objective function, we therefore need to maximize a function $f(m)$ given by

$$f(m) = \alpha_m C_0^{(m)} N_m \left(1 - \sqrt{1 - \frac{4\eta_m M}{\beta_m c_m N_m (C_0^{(m)})^2}}\right). \quad (3.46)$$

This is not as straightforward as before, where the answer was simply to choose $m$ which minimizes $N_m$. In this case, we pick the best $m$ numerically, simply by running through all $m \in \{1, 2, \ldots, n\}$ and selecting the one which maximizes $f(m)$.

Next we turn to the problem of what subset $S$ of the populations should be saturated. As before, we consider the knapsack approximation where we ignore the resources remaining after all the saturations. The objective function as a function of $S$ can then be written as

$$J(S) = \sum_{i \in S} \alpha_i C_0^{(i)} N_i + \sum_{i \notin S} \alpha_i C_0^{(i)} N_i$$

$$= \sum_{i} \alpha_i C_0^{(i)} N_i - \sum_{i \in S} \alpha_i (C_0^{(i)} - C_T^{(i)}) N_i. \quad (3.48)$$

Therefore the optimal $S$ is the solution to the knapsack problem

$$\text{maximize} \sum_{i \in S} \alpha_i (C_0^{(i)} - C_T^{(i)}) N_i \text{ given that } \sum_{i \in S} c_i \gamma_c^{(i)} N_i < M. \quad (3.49)$$
3.4 Multiple threats metapopulation

The solution is again a knapsack problem, but with a different set of values and weights, which can be read off equation (3.49) as

\[
\text{value}_i = \alpha_i \left( C_0^{(i)} - C_T^{(i)} \right) N_i \quad (3.50)
\]

\[
\text{weight}_i = c_i \gamma_c^{(i)} N_i. \quad (3.51)
\]

3.4.3 Results

As before, the knapsack approximation has an intuitive and simple interpretation. We need to

\textbf{maximize the host value saved by the treatment in the saturated populations.}

We have compared the same allocation strategies as in the previous section, on the same landscape. In addition, the infection rates, treatment rates and costs of treatment of one host are random numbers. Specifically, every infection rate $\beta_i$ was chosen uniformly from the interval $(1.5, 2.5)$, every treatment rate $\eta_i$ was chosen uniformly from the interval $(0.5, 1.1)$ and every cost of treating one host $c_i$ was chosen uniformly from the interval $(1, 1.5)$. The value of one host in this example is simply 1 for every population. The resulting plots for three different realisations of the random parameters can be found in Figure 3.7. We can see that the knapsack approximation does remarkably well. It only noticeably misses the optimal black line in Fig 3.7A for a small range of values of the resource limit and even then it is only by about 8%. In comparison both the allocate to the smallest and allocate to the largest strategies do not perform well and in fact, considerably worse than when there is no variability between the populations. For example, in both Fig 3.7 A and C the allocate to the largest strategy (the greedy strategy) does extremely badly, often even worse than the naive equal and proportional allocation strategies. To understand why, consider Figure 3.8. We plot the ordered population sizes and for each, their cost and value of saturation. These are given by

\[
\text{cost of saturation} = N_i \gamma_c^{(i)} c_i \quad (3.52)
\]

\[
\text{value of saturation} = N_i (C_0^{(i)} - C_T^{(i)}) \alpha_i. \quad (3.53)
\]

We can see that while in the Fig 3.8 B the largest populations have a relatively high value of saturation and relatively low cost of saturation, in Fig 3.8 A and C, the cost
One-time allocation of a limited resource for disease control between multiple populations

Fig. 3.7 Comparison of the performance of several allocation strategies for three different realisations of the random sets of parameters $\beta_i$, $\eta_i$ and $c_i$. 
3.4 Multiple threats metapopulation

Fig. 3.8 Visualisation of the different parameter values in the different populations. We show the ordered population sizes together with the cost of saturation and the value of saturation for each population. We can see that for the few largest populations, in B the cost is low and the value high, while in A and C, the cost is high and the value low. We would therefore expect to see the greedy allocation strategy perform better in B than in A and C.
One-time allocation of a limited resource for disease control between multiple populations

of saturation is higher and the value is lower. This is reflected in the bad performance of the allocate to the largest strategy in Fig 3.7 A and C and the comparatively better performance in Fig 3.7 B.

We also tested the knapsack approximation against the exact optimum across a wide range of different landscapes, for $n = 3$. The first population was kept the same with $\beta = 2$, $\eta = 1.2$, $c_1 = 1$, $N_1 = 100$ and the initial conditions endemic. The parameters for the remaining two populations were selected at random in the following manner. $\beta$ is uniform on $(2,3)$, $\eta$ is $\beta - 1 + \text{Unif}(0, 0.5)$, $c$ is uniform on $(1,1.5)$ and the size is uniform on $(100,1000)$. We considered 50,000 different realizations and for each we computed the worst-case error (the largest error as the resource limit is varied). The mean worst-case error was 0.28% and the standard deviation was 1.19%. As before, the distribution is highly skewed, since while the worst-case error was 0 in about 91% cases, the maximum was 16.7%. Correspondingly, the large values of the error were rare with about 1.2% of the cases having the worst-case error larger than 6%.

To investigate when these rare large errors occur, in Figure 3.9 we plotted two scatter plots of the values and weights in the two populations. The scatter plots reveal that when the worst-case error is large, the knapsack values in the populations are highly correlated and in fact, they lie along the lines where either two of the values are the same or one of the values is the sum of the other two. It is harder to interpret the scatter plot of the weights. While the weights $w_2$ and $w_3$ are not as well correlated (correlation coefficient of 0.61), we can see from the plot that they really split into two symmetric pieces. If we look at the data from each piece separately, we find a correlation coefficient of 0.91 and doing a linear regression reveals that very roughly, $w_3 \approx 0.5w_2$ or $w_3 \approx 2w_2$. However, we haven’t found a condition for the weights similarly general as the one for the values.

The relatively large values (greater than 6%) of the error only occur for a finely tuned combination of the values, weights and the resource limit and in vast majority of the cases, there is no error at all. This shows that even with the added complexity of parameters varying between the subpopulations, the knapsack approximation is still a good and intuitive solution to the allocation problem.
Fig. 3.9 Scatter plots of the different random realizations which gave the worst-case error higher than 6%. We can see that the values in the two populations $v_2$ and $v_3$ are almost perfectly correlated with each other and in fact lie along the three lines $v_3 = v_2$ and $v_3 = v_2 \pm v_1$. The colour coding just corresponds to the three different lines. The scatter plot of the weights is harder to interpret, the correlation coefficient of $w_2$ and $w_3$ is about 0.61.
3.5 Extension to the SIRS model structure

We conclude this chapter with a brief discussion of how our approach to solving the resource allocation problem in the SIS setting can be extended to the SIRS setting. The behaviour of the SIRS model differs from the SIS model in a few important aspects. As we have shown in the previous chapter, when the resources allocated are close to the saturation boundary, the stable attractor the system settles in is often a limit cycle. This is problematic for two reasons. Firstly, we cannot obtain an analytical expression for the proportion of infected individuals in the endemic state, which is naturally given by

$$\frac{1}{T} \int_0^T I(t) dt$$

(3.54)

where the integral is evaluated around the limit cycle and $T$ is the time it takes to traverse the cycle. Secondly, in most cases we cannot derive an expression for $\gamma_c$, the amount of resource required to saturate the population. Here by saturation we mean again the state in which adding more resources no longer has any effect on the objective function. Recall from Chapter 2 that in the SIRS setting this means the system has settled into $B$ (the endemic equilibrium in which all the infected hosts receive treatment) or $A$ (the disease free equilibrium), when it is stable. The reason is that while we can calculate the amount of resources where $B$ comes into existence and even where $D$ loses stability, usually when $D$ (the endemic equilibrium in which not every infected host can be treated, i.e. $I_D > \gamma$) loses stability the system does not transition into $B$ but instead onto a stable limit cycle. The amount of resource we need is therefore the amount to make this limit cycle disappear and since this happens through a global bifurcation, we cannot make progress. Even when there is no limit cycle involved, the basin of attraction of $D$ is given by a homoclinic orbit of the saddle point $C$, for which we have not found an analytical expression. To make matters worse, as we have shown in the previous chapter, sometimes it is optimal to delay the resource allocation and our SIS approach does not incorporate this at all.

Since the purpose of this investigation is to provide analytical and intuitive insight rather than run large scale numerical simulations, we consider a special case of the SIRS model where we can make progress. Suppose the treatment is strong enough to eradicate the disease entirely, given full coverage. Mathematically, this assumption translates to

$$\eta > \beta - \mu - \sigma.$$  

(3.55)
This is the region V from Chapter 2. In this region, there are no stable limit cycles and therefore we can obtain analytical expressions for the attractors in the system. One is the disease-free equilibrium (point $A$ at $(I, R) = (0, 0)$), the other is the point $D$, given by

\[
I_D = \frac{\chi + \sqrt{\chi^2 - P}}{2\beta(\mu + \nu + \sigma)} \quad (3.56)
\]

\[
R_D = \frac{\mu I_D + \gamma \eta}{\nu + \sigma} \quad (3.57)
\]

where

\[
\chi = \beta(-\gamma \eta + \nu + \sigma) - (\mu + \sigma)(\nu + \sigma) \quad (3.58)
\]

\[
P = 4\beta \gamma \eta(\nu + \sigma)(\mu + \nu + \sigma) > 0. \quad (3.59)
\]

This point is stable for $\gamma < \gamma_c$, expression for which is given in the previous chapter (eqn. 2.53). However, it is not possible to obtain an expression for the critical gamma $\gamma_c$ needed for saturation as a function of the initial conditions. This is because the saturation happens when the system transitions to the disease-free equilibrium which happens when the basin of attraction of $D$ shrinks enough so that the initial condition $(I_0, R_0)$ lies outside of it. Furthermore, if we consider delaying the allocation of resources (as discussed in the previous chapter, Fig. 2.5), then the condition for saturation is further relaxed; we need the basin of attraction to shrink enough so that either the initial condition $(I_0, R_0)$ or any future part of the trajectory of the system lies outside of the basin of attraction of $D$. In any case, we can at least provide an upper bound on $\gamma_c$ since clearly $\gamma_c < \gamma'_c$. We now show that the optimal resource allocation must again consist of saturating some subset of the populations and then allocating the remaining resource into one of the remaining populations.

Consider resource allocation strategy $\{x_i\}$ such that no population is saturated, that is all the populations are in the endemic equilibrium $D$. The objective function is given by

\[
J(x_i) = \sum_i N_i I_D^{(i)}(x_i) \quad (3.60)
\]

where

\[
I_D^{(i)} = \frac{\beta(\nu - \eta \gamma_i) - \nu + \sqrt{(\beta(\nu - \eta \gamma_i) - \nu)^2 - 4\beta \eta \gamma_i \nu(\nu + 1)}}{2\beta(1 + \nu)}. \quad (3.61)
\]
One-time allocation of a limited resource for disease control between multiple populations

Note we have set $\mu = 1$ and $\sigma = 0$ for simplicity. As before, $\gamma_i = Mx_i/N_i$. To prove that the optimal strategy must saturate populations until further saturation is not possible, we need to show that the above objective function does not have a local minimum. The elements of the Jacobian are

\[
D_{ii} = N_i \frac{\partial^2 I_D(i)}{\partial x_i^2} + N_n \frac{\partial^2 I_D(n)}{\partial x_n^2} \bigg|_{x_n=1-x_1...-x_{n-1}} \tag{3.62}
\]

\[
D_{i \neq j} = N_n \frac{\partial^2 I_D(n)}{\partial x_n^2} \bigg|_{x_n=1-x_1...-x_{n-1}}. \tag{3.63}
\]

As before, to show that there cannot be a local minimum, we need to show that the Jacobian always has at least one negative eigenvalue. This would certainly be true if $\frac{\partial^2 I_D}{\partial x^2} < 0$ and this is the subject of the following theorem.

**Theorem 4.** The endemic prevalence $I_D$ is a concave function of the resources allocated, that is $\frac{\partial^2 I_D}{\partial x^2} < 0$.

**Proof.** Define function $f$ by

\[
f(x) = \beta(\nu - \eta \gamma(x)) - \nu + \sqrt{(\beta(\nu - \eta \gamma(x)) - \nu)^2 - 4\beta \eta \gamma(x) \nu(\nu + 1)}. \tag{3.64}
\]

To prove the claim, we need to prove that $\frac{d^2 f}{dx^2} < 0$. To simplify notation, we introduce variable $z$ by $z = \beta \eta \gamma(x)$ and $r = \beta - 1$. Straightforward algebra then gives the expression for the second derivative of $f$ as

\[
\frac{d^2 f}{dx^2} = \frac{\sqrt{(rv - z)^2 - 4zv(\nu + 1)} - \frac{(rv - z + 2v(\nu + 1))^2}{\sqrt{(rv - z)^2 - 4zv(\nu + 1)}}}{(rv - z)^2 - 4zv(\nu + 1)}. \tag{3.65}
\]

Therefore, we have

\[
\frac{d^2 f}{dx^2} < 0 \tag{3.66}
\]

\[
\iff (rv - z)^2 - 4zv(\nu + 1) < (rv - z + 2v(\nu + 1))^2 \tag{3.67}
\]

\[
\iff 0 < \nu(\nu + 1) + rv, \tag{3.68}
\]

which is true. This proves the claim of the theorem. \qed

Therefore there cannot be any local minimum and the minimum of the objective function must lie on the boundary. This means that just like in the SIS case, the
optimal strategy must saturate some subset of the populations such that no further saturation is possible and then allocate the remaining resource into one of the remaining populations.

3.6 Discussion

In this chapter, acknowledging the importance of integrating disease control models with spatial structure (Dye and Day, 2003; Grenfell et al., 2001), we used the understanding of constrained control developed in chapter 2 to investigate the allocation of limited control resource between multiple separate populations. The shared resource pool introduces an effective interaction between the populations, because allocating resource to one of them implies there will be less of the resource left for the others. This has the effect of anti-correlating the levels of infection in the different populations. In order to better understand this effect, we made the simplifying assumption that the populations are otherwise uncoupled. This allows us to make analytical progress. It can also be thought of as a situation where different populations consist of different host species each with a different pathogen. This is a common problem in plant epidemiology as reflected by the large number of host-pathogen systems in the UK Plant Health Risk Register (Baker et al., 2014). We reflected this interpretation by allowing the disease and control parameters to vary between the different populations.

For the most part we focus on the SIS model because of its relative analytic simplicity but we include a brief extension of our results to a special case of the SIRS model. However, the full treatment of the SIRS model is problematic because the limit cycles make it difficult to study the long term behaviour analytically. We model the effects of control in the same way as in the previous chapter, and assume that the control resource is allocated between the populations initially and cannot be reallocated later. This means that what may be an efficient allocation initially may no longer be efficient later after the disease prevalence in some of the populations decreases. The allocation strategy must therefore take into account the long term behaviour of the epidemic.

Following this, we formulate an optimization problem by defining the objective function as the total number of infected hosts in the endemic equilibrium. There are other ways of defining the objective function, such as the total number of hosts infected in some time interval \((0, T)\) (Madden et al., 2007). In our analysis, we wanted to avoid
the arbitrariness of the time horizon $T$ and also to avoid introducing another parameter. In fact, defining the objective function as we do here is equivalent to considering a long-term horizon $T \rightarrow \infty$. Naturally, considering a short time horizon (Zaric and Brandeau, 2001) could potentially significantly change the results. Note that we do not explicitly consider any discounting of the cost of future infections (Forster and Gilligan, 2007). Generally, due to how our objective function is defined, an exponential discount factor $e^{-rt}$ would only multiply the objective function by a constant and thus not affect the analysis in any way.

In the single threat model, that is a metapopulation with the model parameters constant across the subpopulations, we find that the optimal allocation must saturate a certain subset $S$ of the populations until no further saturation is possible, and then allocate the remaining resources to the smallest of the remaining populations. The optimal set $S$ can be found numerically by searching through all the options. We derive an approximate solution to the problem which we call the knapsack approximation, due to the similarity to the knapsack problem from computer science (Martello and Toth, 1990; Kellerer et al., 2004). The knapsack approximation identifies a value of saturation for each population (given in the text) and the states that the optimal strategy maximizes the value given the resource constraint. The knapsack algorithm is very common across various fields, and has been used both in conservation biology (Joseph et al., 2008) and in epidemiology, for example in the economic evaluation of HIV policies (Kaplan, 1998).

The knapsack approximation implies that the set $S$ should be selected so that the sum of the sizes of the saturated populations is maximal. That is, under the approximate optimal allocation, we should maximize the number of hosts that reside in fully saturated populations. We give an algorithm to solve this problem and verify the accuracy of the knapsack approximation when compared with the exact solution. In the worst cases, the approximation underperforms the exact optimum by about 10%, and this only happens for very finely tuned combinations of the population sizes and the maximum available amount of resource.

To demonstrate the importance of the correct choice of the subset $S$, we also considered a number of naive, simplistic allocation strategies. Of these, the most intuitively appealing is the greedy allocation strategy, which identifies which population should be saturated to maximize the decrease in the objective function, saturates it and then repeats this process until there are no resources left. For an overview of such
greedy algorithms see chapter 16 in Cormen et al. (2001) and for an analysis of when such approaches fail see Bang-Jensen et al. (2004). This amounts to saturating the populations in the order of their size, from the largest to the smallest. This strategy significantly underperforms the knapsack approximation. The greedy allocation fails because allocating to the largest population does not take into account the large amount of resources that will not be available any more for allocation to the other populations. It therefore fails because it does not account for the interaction caused by the shared resource pool. Because it arises from the mathematical analysis of the SIS model, the knapsack approximation accounts for this interaction correctly. Note that both strategies significantly outperform other naive allocation strategies such as allocating the same amount of resources to each population, or allocating resources proportionally to the population size.

When the model parameters are allowed to vary between the populations, we show that the optimal allocation must again saturate a certain subset of populations and allocate the remaining resource to one population outside of this subset. We derive a different knapsack approximation which provides an (approximately) optimal solution to the problem. It turns out that we need to maximize the total value of the hosts saved by the treatment in the fully saturated populations.

The knapsack approximation represents a simple and intuitive allocation prescription that we set out to construct in this chapter. It specifies a value and a cost of saturation in each of the populations and reduces the resource allocation problem to a knapsack problem of how to maximize the total value while keeping the total cost below certain limit. Whereas the exact solution is more or less a black box, the knapsack approximation provides intuition that might be of use in more complex models.

Throughout this chapter, we assume that the treatment resource is allocated only once at the beginning and cannot be reallocated later. In the next chapter, we relax this assumption and analyse how the optimal allocation strategies are affected when the resource can be reallocated later, or even reallocated continuously as the system evolves.
One-time allocation of a limited resource for disease control between multiple populations

Fig. 3.10 Appendix 1. Figure showing the knapsack percentual error as a function of the resource limit $M$ for a landscape with population sizes $(100, 100, 198)$ and $\beta = 2$, $\eta = 1.2$. Note that these population sizes satisfy the condition outlined in the appendix for there to be a large error. The dashed red lines indicate the resource limits necessary for saturation of various subsets of the populations. From left to right, they are the limits for saturation of $(1)$ or $(2)$, $(3)$, $(1,2)$, $(1,3)$ or $(2,3)$. We can see that the error is largest when the resources are just below the limit necessary to saturate the set $(1,3)$ and in this appendix we aim to explain why that is.

3.7 Appendix

3.7.1 A1. Analysis of the knapsack approximation error when $n = 3$

Consider a situation like the one described in the text. We have three populations with sizes $N_1 \leq N_2 \leq N_3$ and furthermore $N_3 = N_1 + N_2 - \epsilon$ for $\epsilon$ a small, positive constant. In Figure 3.10 we plot the percentual error as a function of the resource limit $M$. Note that the error is mostly 0, and only becomes appreciable as we get close to the limit necessary to saturate populations $(1,3)$. The moment we go over this limit, it drops back to zero again, because both the knapsack approximation and
the exact optimum will simply saturate (1,3) as soon as they can. Here we aim to explain this behaviour. Suppose that the resource limit \( M \) is just below what is needed for saturating (1,3). Then the knapsack approximation saturates (1,2) while it turns out the exact optimum saturates (3). Since \( N_3 \approx N_1 + N_2 \), all we need to show is that given resources insufficient to saturate either, it is better to split \( N_3 \) into two pieces \( N_1 \leq N_2 \), neither of which can be saturated. Furthermore, to explain why does the largest error occur close to the resource limit needed for (1,3), we need to show that in fact when we split \( N_3 \), it is best to do it so that \( N_1 \) is almost small enough to be saturated. So, consider splitting \( N_3 \) into \( x \) and \( N_3 - x \). The knapsack objective function \( J_K \) and the exact one \( J_E \) are given by

\[
J_K = \frac{1}{2}N_3(C_0 + \sqrt{C_0^2 - a/N_3}) \tag{3.69}
\]

\[
J_E = (N_3 - x)C_0 + \frac{1}{2}x(C_0 + \sqrt{C_0^2 - a/x}) \tag{3.70}
\]

where \( a = 4\eta M_R/\beta \) and \( M_R \) are the resources remaining after saturation (in the knapsack case the saturation of (1,2), in the exact case the saturation of (3)). Clearly \( x \in (a/C_0^2, N_3) \), where \( x = N_3 \) corresponds to no splitting at all. To prove the above claims, we need to show that \( J_E \) is an increasing function of \( x \), since then it is the smallest (and consequently the error is the largest) when \( x \) is close to \( a/C_0^2 \) which is the size at which \( x \) can be saturated. It is easy to show that

\[
2\frac{dJ_E}{dx} = -C_0 + \sqrt{C_0^2 - a/x} + \frac{a}{2x\sqrt{C_0^2 - a/x}}. \tag{3.71}
\]

Then we can write

\[
\frac{dJ_E}{dx} > 0 \tag{3.72}
\]

\[
\iff C_0^2 - a/x + a/2x > C_0\sqrt{C_0^2 - a/x} \tag{3.73}
\]

\[
\iff (a/2x)^2 > 0, \tag{3.74}
\]

which finishes the proof.

This demonstrates, that when \( N_3 \approx N_1 + N_2 \) with \( N_3 \) slightly smaller (so that the knapsack chooses to saturate \( N_1 \) and \( N_2 \)), the exact optimum does better by saturating \( N_3 \). Furthermore, the difference is the largest when we can almost saturate both \( N_1 \) and \( N_3 \). This makes sense since this means the amount of resource remaining after
One-time allocation of a limited resource for disease control between multiple populations saturating is the largest it can be and since the knapsack approximation effectively neglects the resource remaining after saturation when determining the optimal set to saturate, it is not surprising that this is the point where it leads to the largest error.
Chapter 4

Allocation of a limited resource for disease control between multiple populations with reallocation

4.1 Abstract

In this chapter, we extend the analysis of optimal control resource allocation between multiple isolated populations. We relax the assumption of a one-time only allocation and allow for the resources to be reallocated. Clearly allowing for this extra flexibility means we can control the epidemic more efficiently. Our aim is to find simple and intuitive rules for how this reallocation should be done in order to maximize the benefits.

We start by demonstrating that significant gains can be made by reallocating the resources once the system reaches endemic equilibrium. Motivated by this, we turn to the opposite extreme and consider the question of how to construct an allocation strategy in continuous time, that is we allow for reallocation continually throughout the epidemic. We show how the dynamical understanding developed in chapter 2 can be used to approach this problem, and based on this we propose an intuitive allocation strategy that outperforms previously considered approaches.
4.2 Introduction

In the previous chapter, we assumed that after the initial allocation, the resources were fixed and could not be reallocated. This was a sensible starting assumption, because reallocating the resource will never be cost-free and it is therefore important to know how well we can do without reallocating at all. Naturally, being able to reallocate the resources will lead to better efficiency and therefore give better results. The problem of how much we can improve the control effort and how should the reallocation be done is the subject of this chapter.

Initially, we consider reallocation of resources whenever the system reaches an equilibrium, which gives rise to a mathematical problem interesting in its own right. We provide a solution and use it to demonstrate that allowing for reallocation can lead to substantial improvements in the management of the epidemic. This is a fairly intuitive result and one that has been seen before, albeit in a different model (Zaric and Brandeau, 2002).

In the main part of this chapter, we consider the opposite extreme to the one-time-only allocation, where the resource can be reallocated continuously in time. A similar problem was previously studied by Rowthorn et al. [2009], who considered the SIS model applied to two populations of equal size, with most of the transmission within the populations, but some transfer of infection between them as well. Mathematically, this is captured by the concept of metapopulation commonly used in both ecology (Hanski, 1999; Hanski and Ovaskainen, 2003) and epidemiology (Grenfell and Harwood, 1997; Lin and Li, 2014). Their assumption of the possibility of reallocating the resource continuously in time amounts to answering the following question: How should the resources be allocated between the populations at any given time in order to achieve the best overall outcome? To do this, Rowthorn et al. (2009) defined the objective function (to be minimized) as the total number of infected hosts up to a certain time horizon $T$, mathematically expressed using an integral. They further included a discounting factor (Downes and Goodman, 2003) $e^{-rt}$, taking into account that infections occurring now are more economically significant than infections occurring later (Forster and Gilligan, 2007).

To approach this problem, Rowthorn et al. (2009) used optimal control theory (Pontryagin et al., 1962; Seierstad and Sydsaeter, 1986) and constructed a numerical method for finding the optimal resource allocation. They found that, perhaps contrary to
intuition, the resource should preferentially be allocated to the less infected population, that is the one with fewer infected individuals. In fact, Rowthorn et al. (2009) went further and demonstrated that allocating the resources preferentially to the more infected population is the worst possible allocation strategy. At the end, they briefly discuss a system of $n$ populations of equal size and show that again, allocating preferentially to the population with the fewest infected hosts is better than allocating to the population with the most infected hosts. They offer an intuitive explanation for this phenomenon by noting that the population with the fewest infected hosts has the most susceptible hosts and should therefore be protected. In this chapter, we offer and discuss another explanation, based on our investigation of the system where eradication of the disease is possible, given full treatment coverage. Focusing on the populations with the lowest infection prevalence will eradicate the disease in the populations one by one, whereas equalizing the levels of infection effectively attempts to eradicate everywhere at once, which requires far greater amount of resources.

Ndeffo Mbah and Gilligan (2011) extended the study in Rowthorn et al. (2009) to the SIRS model. They found that the (approximately) best allocation strategy involves initially treating the population with more infected individuals and then switching to the population with fewer infected individuals. However, they note that this might be impractical due to the danger of missing the optimal switching time. They conclude that a strategy that represents good balance between optimality and practical considerations is to preferentially treat the population with more susceptible individuals.

We consider both the SIS and SIRS model and extend their results to multiple populations of varying sizes. Preliminary exploratory analysis showed that in the SIRS model, the strategy of preferentially treating the least infected population and the strategy of preferentially treating the most susceptible population sometimes significantly differ in their outcome and when that happens, usually it is because they reach different endemic equilibria. This behaviour was found independently of how strongly the populations were coupled, including in the absence of coupling, as we describe here.

To investigate this behaviour and to either find a criterion for deciding between these two strategies or an entirely new strategy, we make a number of simplifying assumptions. Like in Chapter 3, we ignore the coupling between the populations and focus on the indirect interaction provided by the shared resource pool. We also choose
Allocation of a limited resource for disease control between multiple populations with reallocation

the number of infected individuals in the endemic equilibrium as the objective function and thus we are effectively working with a long time horizon. These assumptions allow us to use the understanding of the dynamics developed in Chapter 2 and to construct a new allocation strategy which outperforms the previously considered approaches. This way we are able explain the large variability in outcome of the two strategies described above (focusing on the least infected or on the most susceptible).

4.3 Model with equilibrium resource reallocation

In this chapter, we consider what happens when we allow for resource reallocation. We are particularly interested in any insights our previous analysis can give into this new problem and also in what the value of reallocating resources is. In other words, how much can we improve the objective function if we are allowed to reallocate. In the previous chapter we established that the optimal value of the objective function is achieved when some of the populations are fully saturated. The problem, when we cannot reallocate the resources, is that in order to saturate a population we might need some amount of resources $N_i c_i \gamma^{(i)} c$, but after the population reaches the state $I_i = C_T^{(i)}$, we only need $N_i c_i C_T^{(i)}$ resources to maintain it. The surplus resources are not being used for anything at that point. Note that this means that when the initial conditions are such that $\gamma^{(i)} c = C_T^{(i)}$ for every $i$, the optimal allocation strategy is the same with or without reallocation. Otherwise, the idea is that when reallocation is possible, as long as initially the amount of resources we have is greater than or equal to $N_i c_i \gamma^{(i)} c$, we only need $N_i c_i C_T^{(i)}$ to saturate the population. This assumes that we can allocate the necessary resources to $i$, wait until it settles to its full treatment endemic equilibrium and then move the surplus resources elsewhere. In order to avoid having to worry about the transient dynamics of the other populations and how their $\gamma^{(j)} c$ might have changed in the meantime, we assume that the initial state of the system is the endemic equilibrium in the absence of treatment, $C_T^{(i)}$. Under this assumption, we have that

$$\gamma^{(i)} c = \gamma^{(i)}_{\text{max}}. \quad (4.1)$$

Note first that when saturation leads to complete disease eradication (that is $\eta > r$), the solution is trivial. Any population $i$ for which the total amount of available resources $M$ exceeds $N_i c_i \gamma^{(i)}_{\text{max}}$ can be completely cleared of the disease. No resources therefore need to be left in any of the populations long term and can be reallocated after the
eradication is achieved. Therefore, in all the populations for which \( N_i c_i \gamma \leq M \), the disease can be eradicated. The resources will then be allocated fully to one of the remaining populations. When \( \eta < r \), the situation is not as simple. The condition \( N_i c_i \gamma \leq M \) is now not sufficient for all those populations to be saturated, because after one of those populations is saturated, say population \( j \), the remaining resources are not \( M \) but \( M - N_j c_j C_T^{(j)} \) and this might not be enough to saturate the others anymore.

We take the following approach. At the end of the saturation procedure, all the saturated populations will be at the infection level \( C_T^{(i)} \). Therefore, we first solve the knapsack problem with weights given by \( w_i = N_i c_i C_T^{(i)} \) and total resources \( M \) on the subset of the populations on which saturation is theoretically possible, that is \( M \geq N_i c_i \gamma \) for each \( i \). This is the same as answering the question of what is the optimal final state of the system given the amount of resource \( M \). Suppose the solution to this problem is to saturate a subset \( S \) of the populations. If we can reach the saturation in this subset, we know this is as good as we can do (within the knapsack approximation framework). The next question is then as follows. We know that \( M \) is big enough for the system to be in this optimal final state, but is it also big enough to reach this state? Clearly the minimum requirement for being in the final state, that is \( M = \sum_{i \in S} N_i c_i C_T^{(i)} \) is not enough, because if that was the amount of resources available, when we reached the final population to be saturated, call its index \( z \) w.l.o.g., we would only have \( N_z c_z C_T^{(z)} \) resources available and we would need \( N_z c_z \gamma \). Suppose again that the last population saturated has index \( z \). Provided we can saturate all the previous ones, the amount of resources we need to ensure we will be able to saturate the final one is

\[
M_{\text{needed}} = \sum_{i \in S \setminus \{z\}} N_i c_i C_T^{(i)} + N_z c_z \gamma
\]

\[
= \sum_{i \in S} N_i c_i C_T^{(i)} + N_z c_z (\gamma - C_T^{(z)})
\]

\[
= f(z),
\]

where the final equality defines the function \( f \). Clearly if

\[
M < \min_z f(z)
\]
then we cannot reach a state in which \( S \) is saturated, because no matter which population we choose to saturate as the last one, we will not have enough resources once we reach it. This argument shows that the order in which we saturate the populations matters. We have the following result.

**Theorem 5.** Suppose that

\[
M > \max_z f(z).
\]

Then saturating \( S \) can be accomplished.

**Proof.** Suppose we re-index the populations so that they are ordered descending in the value of \( N_i c_i \gamma_i^{(i)} \). So \( N_1 c_1 \gamma_1^{(1)} \geq N_2 c_2 \gamma_2^{(2)} \geq \ldots \geq N_{|S|} c_{|S|} \gamma_{|S|}^{(|S|)} \). Suppose we want to saturate the populations in the order starting from \( |S| \) and going down to 1. Because \( M > f(1) \), we know we can finish the saturation procedure with population 1. The question is, could we have saturated population 2 before it? To saturate population 2 requires \( N_2 c_2 \gamma_2^{(2)} \) resources. Denote by \( M_i \) the amount of resources available to us when we need to saturate the population \( i \). Since we can finish with saturating 1, \( M_1 \geq N_1 c_1 \gamma_1^{(1)} \). We have

\[
M_2 = M_1 + N_2 c_2 \gamma_2^{(2)} \\
\geq N_1 c_1 \gamma_1^{(1)} + N_2 c_2 \gamma_2^{(2)} \\
\geq N_2 c_2 \gamma_2^{(2)},
\]

where the last inequality follows from the ordering of the re-indexing of the populations defined above. Therefore 2 can be saturated. The argument can be made iteratively by induction. Suppose we can saturate \( i \). Then

\[
M_{i+1} = M_i + N_{i+1} c_{i+1} \gamma_{i+1}^{(i+1)} \\
\geq N_i c_i \gamma_i^{(i)} + N_{i+1} c_{i+1} \gamma_{i+1}^{(i+1)} \\
\geq N_{i+1} c_{i+1} \gamma_{i+1}^{(i+1)},
\]

and so we can also saturate \( i + 1 \). Therefore we can saturate all of \( S \).

We have therefore established that when \( M < \min_z f(z) \), saturating \( S \) is impossible while when \( M > \max_z f(z) \), \( S \) can be saturated. What if \( M \) is in between these two values? Clearly there is some \( M_c \) such that \( M \geq M_c \) means saturating \( S \) is possible and \( M < M_c \) means it is impossible. To find \( M_c \), we first introduce some simplifying
notation. Define

\[ s_i = N_i c_i \gamma_{i \max} \]  \hspace{1cm} (4.13)
\[ t_i = N_i c_i C_T^{(i)} . \]  \hspace{1cm} (4.14)

Clearly \( s_i > t_i, \forall i \). Consider a permutation \( \sigma \) of the indexes \( 1, 2, \ldots, |S| \). The populations can be saturated in the order of the permutation \( \sigma \) if and only if

\[ M \geq \sum_{i=1}^{k-1} t_{\sigma(i)} + s_{\sigma(k)}, \ \forall k = 1, 2, \ldots, |S|. \]  \hspace{1cm} (4.15)

For each \( k \) this is simply the condition that we have enough resources to saturate \( \sigma(k) \) when we reach it, since \( \sum_{i=1}^{k-1} t_{\sigma(i)} \) is the amount of resources tied up in the populations saturated before \( \sigma(k) \) and \( s_{\sigma(k)} \) is the amount needed to then saturate \( \sigma(k) \) itself. This can be written as a single condition

\[ M \geq \max_k \left[ \sum_{i=1}^{k-1} t_{\sigma(i)} + s_{\sigma(k)} \right]. \]  \hspace{1cm} (4.16)

Clearly then, the expression for \( M_c \) is

\[ M_c = \min_{\sigma} \max_k \left[ \sum_{i=1}^{k-1} t_{\sigma(i)} + s_{\sigma(k)} \right]. \]  \hspace{1cm} (4.17)

The populations should then be saturated in the order of the permutation for which the minimum is achieved. We were not able to simplify equation (4.17) any further and the value \( M_c \) therefore needs to be identified numerically.

What if \( M < M_c \) and the optimal subset \( S \) therefore cannot be saturated? We found \( S \) as the solution to the knapsack problem with weights \( N_i c_i C_T^{(i)} \) that maximizes the value. If \( M < M_c \) and \( S \) is illegal, we need to go to the next best in the knapsack algorithm. So, let \( S_1 \) be subsets of the populations such that \( S_1 = S \) is the solution to the knapsack maximization problem, \( S_2 \) is the second best and so on. The optimal solution is saturating \( S_k \) for \( k \) the smallest such that \( M_c \) associated with the set \( S_k \) is smaller or equal to \( M \). To summarize the procedure, given amount of resource \( M \), we first look at the question of what are the best subsets of saturated populations we can support with the resource. We order them with \( S_1 \) being the best, \( S_2 \) begin the second best etc. At this point we however do not know whether these can actually be reached,
Fig. 4.1 The optimal value of the objective function as a function of the available resources when reallocation is possible (green) and when it is not (red). The three plots correspond to three different realisations of the random parameters $\beta_i$ chosen uniformly from $(1.8,2.2)$, $\eta_i$ chosen uniformly from $(0.5,1.1)$ and $c_i$ chosen uniformly from $(1,1.3)$, for each population $i$. We can see that the difference between the cases with and without reallocation can be very large.
we only know that if $S_1$ is reached, it is optimal. The above analysis then produces a critical value $M^*_i$ for each of these subsets $S_i$. If $M \geq M^*_i$, then $S_i$ can be reached from the initial state of the system. The optimal solution is then to saturate the subset $S_i$ for the smallest $i$ for which $M \geq M^*_i$.

We have plotted the optimal value of the objective function both with and without reallocation in Figure 4.1. Using the above results, we plot both cases for three different realisations of the random parameters $\beta_i$, $\eta_i$ and $c_i$. The underlying landscape has five populations with sizes 100, 200, 300, 150 and 220. We can see from the plots that the difference between the two can be quite dramatic, depending on the parameters of the model. This means that significant benefit can be gained from reallocating resources after the initial allocation.

### 4.4 Model with continuous resource reallocation

In this section, we consider the situation where the resources can be reallocated between the populations continuously in time. This would be the case for example when the resource reallocation is very cheap and the typical time scale needed for moving the resource between the populations is much shorter than the typical time scale of the epidemic (such as the recovery period). We begin the section with a brief description of the model. We then present the results for the previously investigated case of two populations of equal size (Rowthorn et al., 2009) and explain why the approach in their work might fail when there are more populations of potentially different sizes. We then present a strategy that incorporates the variable population sizes together with the underlying disease dynamics from Chapter 2 and improves on the previous results.

#### 4.4.1 Model description

The model is almost the same as the single threat metapopulation with no coupling from the previous chapter. The only difference is the allocation of the control resource. We denote the proportion of the total resource allocated to the population $i$ by $x_i$, so $\sum_i x_i = 1$. While in the previous chapter, these were constant, in this chapter they are
functions of time. The optimization problem can therefore be rephrased as

$$\min \sum_{i=1}^{n} N_i I_i^\infty,$$

(4.18)

given

$$\frac{dI_i}{dt} = \beta I_i(1 - I_i) - \mu - \eta \min(I_i, \frac{x_i(t)M}{N_i})$$

(4.19)

$$\sum_{i=1}^{n} x_i(t) = 1.$$  

(4.20)

This problem is fundamentally more difficult than when the allocations are fixed, because we cannot easily link the allocations $x_i(t)$ to the final state of the system. It is also challenging to solve computationally because we are now finding a minimum over a set of functions, rather than just numbers.

4.4.2 Analysis of the problem

In Rowthorn et al. (2009) the authors solved a similar problem for the case of two populations of equal size with coupling by considering a result from the optimal control theory, the Pontryagin Maximum Principle (Seierstad and Sydsaeter, 1986). The principle gives a set of equations and inequalities that the optimal solution must satisfy and these can be used for numerical exploration. Unfortunately, this approach is not suitable when the number of populations $n$ is greater than two, because the number of equations quickly becomes too large.

Instead of looking for an optimal allocation strategy, we build on the previous work (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011) by initially selecting three simple strategies and analyse their behaviour when applied to our model. They are

- **Treat the least infected population (TLI).** Under this strategy, at every time instant, the resource is fully allocated to the populations with the smallest number of infected hosts. Here by fully we mean that the resource thus allocated is sufficient to treat all the infected hosts in that population, if possible. If there are resources left, they are allocated to the population with the second smallest number of infected hosts and so on until the resource is depleted.
4.4 Model with continuous resource reallocation

Fig. 4.2 The objective function plotted as a function of the resources available. The TLI strategy makes a much better use of the limited resources than both TMI and TEO. Each drop in the TLI objective function corresponds to the disease eradicated in another population. In contrast, both TMI and TEO attempt to eradicate the disease in all the populations simultaneously and therefore they require much larger amount of resources to succeed. The parameter values used in this simulation are $\beta = 2$, $\eta = 1.2$ and the initial condition was assumed to be the endemic equilibrium in the absence of treatment.

- **Treat the most infected population (TMI).** This is the exact opposite of the above (TLI). Instead of the population with the smallest number of infected hosts, we allocate the resources to the population with the largest number of infected hosts.

- **Equal opportunity (TEO).** This strategy is selected for its simplicity and intuitive appeal. We simply allocate resources to each population proportionally to the number of infected hosts in that population. Under this strategy, each host has the same probability of receiving treatment, independently of the population it belongs to.
Both the TLI and TMI strategies have been considered in Rowthorn et al. (2009) and also in Ndeffo Mbah and Gilligan (2011), where the authors also considered the TEO strategy. Rowthorn et al. (2009) showed that TLI significantly outperforms TMI and that in fact TMI is the worst possible approach. The intuition behind this result is that TMI equalizes the disease prevalence between the populations and tries to control them all simultaneously. Unless the amount of available resource is high, this is likely to fail. On the other hand, TLI focuses on one population at a time and “picks them off” one by one.

This intuition translates fairly well into the case when the number of populations is greater than two and the populations are of various sizes. We illustrate this on an example landscape consisting of populations with sizes 100, 200, 150 and 300. In Fig 4.2 we plot the objective function as a function of the amount of the resource available. We can see that TLI makes a much better use of the available resource than both TMI and TEO. Furthermore, the “step” nature of the TLI curve comes from the fact that TLI eradicates the disease in the populations one by one. Each drop in the objective function corresponds to the resource limit becoming sufficient to eradicate the disease in the next population. In contrast, both the TMI and TEO attempt to equalize the disease prevalence across the populations and thus to eradicate the disease in all the populations simultaneously. This only becomes possible when the amount of available resource is large enough and this corresponds to the single large drop in the objective function of both TMI and TEO.

However, unlike in the case of equal population sizes, treating the population with fewer infected hosts might not always be a good strategy when the populations have different sizes. To understand why, we need to go back to the results from Chapter 2 regarding the dynamics of the SIS model with constrained control. In Chapter 2 we showed that given a resource constraint, there may exist a critical disease prevalence such that when the prevalence is below this critical value, the population can be saturated (and the disease eradicated) but when the prevalence is above this critical value, eradicating the disease is not possible. We also showed that this critical value is in fact equal to the prevalence $I_C$ in the unstable fixed point $C$, given by

$$I_C = \frac{1}{2} \left( C_0 - \sqrt{C_0^2 - \frac{4\eta M}{\beta N}} \right),$$  \hspace{1cm} (4.21)$$

where $C_0 = 1 - 1/\beta$ is the endemic equilibrium in the absence of treatment.
Suppose that we have two populations, one small and one large. At time $t = 0$ the situation is such that even if we allocated all the available resources into the large population, the disease prevalence would be just below the critical value $I_C$. Suppose further that the initial number of infected individuals in the small population is less than in the large one, simply due to the difference in sizes. Under these assumptions, TLI will initially allocate resources to the small population and quickly eradicate the disease there. Meanwhile the disease prevalence might cross the critical value $I_C$ in the large population so that we won’t be able to eradicate the disease there in the future. On the other hand, TMI would start treating in the large population and may in fact succeed in eradicating the disease in both. In other words, when one of the populations has a small window of opportunity for eradication, TLI might miss it and fail to eradicate the disease. We illustrate this on an example with two populations of sizes 100 and 1000. Suppose the initial disease prevalence is 50% in the smaller population and 6% in the larger population and that the total amount of available resource is enough to treat 50 hosts at any given time. The disease progress curves are plotted in Figure 4.3. We can see that as predicted for such a scenario, TLI fails to eradicate the disease by missing the small window of opportunity to eradicate the disease in the larger population.

Such cases where TMI outperforms TLI remain rare even when the population sizes are not equal. In Figure 4.4 we plot the difference between the objective functions of TMI and TLI and the objective functions of TEO and TMI for 3000 randomly generated landscapes consisting of five populations. The landscapes were generated such that $N_1 = 100$ and for all $i > 1$, $N_i$ was chosen uniformly from $(100, 1000)$. The initial conditions were also generated at random uniformly between 0 and the endemic equilibrium in the absence of treatment. Finally, the resource constraint was chosen at random uniformly so that on average, 15% of the initially infected hosts can be given treatment. The figure shows that TLI is better than TMI and thus that the behaviour where TMI outperforms TLI is rare. Furthermore, it demonstrates that TMI is better than TEO and so TLI is better than TEO.

However, while TLI certainly seems to be an excellent heuristic, the above analysis points at a problem with it which stems from the fact that it ignores the underlying dynamics of the model with constrained control. Suppose there is a “critical” population with a larger initial infection prevalence such that TLI will not allocate treatment to it, but such that delaying the treatment will lead to the prevalence crossing the critical
Fig. 4.3 Example illustrating a failure of the TLI strategy. The blue curve corresponds to the smaller population, the green to the larger one. The initial condition in the larger population was 6%, just under the critical value $I_C = 0.697$. The TLI strategy focused on the smaller population first and in doing so, it let the larger population irreversibly cross the critical prevalence. TMI on the other hand started treating in the larger population and managed to eradicate the disease in both the populations. The parameter values were $\beta = 2, \eta = 1.2$.

level $I_C$. In such a case, TLI will fail to eradicate the disease even though allocating the resources to the critical population would have succeeded. In the next section, we propose a modified strategy which takes these dynamics into account.

### 4.4.3 Strategy based on urgency

One way of including the dynamics of the underlying SIS model with resource constraint (Ch. 2) that is intuitive and easy to understand is a strategy we call “treat the most urgent” population preferentially, or TMU. It works as follows:

**Treat the most urgent population strategy**

- Partition the populations into three different classes.
  - Class A, most urgent. These are the populations where, given all the resource we have access to, we can eradicate the disease now, but not necessarily
Fig. 4.4 Scatterplots showing the differences between the TLI and TMI (left) and TMI and TEO (right) as a percentage of the total population size. The x-axis shows the proportion of infected hosts that can be treated at time $t = 0$, that is, it represents the amount of available resource. The population sizes, initial conditions and resource constraints were generated at random, the details can be found in the text. We can see that TLI outperforms TMI independently of the resource constraint and TMI outperforms TEO mostly when more resources are available. The parameter values are $\beta = 2$ and $\eta = 1.2$. 
later and certainly not once the disease prevalence reaches the treatment free endemic equilibrium. This means that $I_C < I_{equilibrium}$. 

- Class B. These are the populations where we can always eradicate the disease, even if the system starts in the treatment free endemic equilibrium. This means that $I_C > I_{equilibrium}$. 

- Class C, least urgent. These are the populations where the initial prevalence is so high that we cannot eradicate the disease at all. This means $I_{initial} > I_C$. Note that we assume $I_{equilibrium} > I_{initial}$. 

• Allocate resources to class A according to TLI. If resources are left, allocate resources to class B according to TLI. If resources are still left, allocate the remaining resource to class C.

The logic behind this allocation strategy is that preference should always be given to the populations where we can eradicate the disease now, but we might not be able to do so later, due to their size when compared with the resource constraint. Implementing the strategy computationally is easy, since at each time instant, we can quickly check which populations belong to which class just by comparing the current disease prevalence with the limit $I_C$. 

We implemented the above TMU strategy on the same set of 3000 landscapes as above and compared it with the TLI. The resulting plot can be found in Figure 4.5. The plot confirms that TMU is a better allocation strategy than TLI. While in most cases, both the strategies lead to the same result in about 16% of the cases, TMU significantly outperforms TLI. It achieves this by correctly identifying which populations should be treated most urgently. We give an example of this in Figure 4.6. Both the green and purple populations can be eradicated even from the treatment free endemic equilibrium, however red cannot. TMU correctly identifies this and preferentially allocates treatment to red. Later, it manages to eradicate the disease in all the populations. On the other hand, TLI initially allocates resources to the teal and blue populations and allows the red to cross the critical value beyond which it cannot be eradicated any more. Consequently, it becomes stuck in an endemic equilibrium where all the resources are going to red (but cannot achieve eradication) while purple and green are in their treatment-free equilibrium. The overall result is much worse than in the TMU case.
4.4 Model with continuous resource reallocation

Fig. 4.5 Plot showing the difference between the TLI and TMU as a percentage of the total population size. The x-axis shows the proportion of infected hosts that can be treated at time $t = 0$, that is, it represents the amount of available resource. The set of randomly generated landscapes is the same as in figure 4.4. Most of the points lie along the x-axis, meaning that both strategies lead to the same result. However, in about 16% cases, TMU significantly outperforms TLI because it manages to achieve eradication of the disease by correctly identifying the most urgent populations to allocate resource to. The parameter values are as before, $\beta = 2$ and $\eta = 1.2$.

Fig. 4.6 A figure showing an example of TMU outperforming TLI in the case marked with A in figure 4.5. The different colours correspond to the five populations. TMU correctly identifies that red is a population which should be treated preferentially. However, TLI allows the red population to cross the critical threshold and fails to eradicate the disease. The parameter values are $\beta = 2$ and $\eta = 1.2$. 
4.5 SIRS model with continuous resource reallocation

In this section, we extend the approach to optimizing treatment allocation developed in the previous section to the SIRS model dynamics. This is building on the work in Ndeffo Mbah and Gilligan (2011) where the authors investigated optimal allocation of limited resources between two populations of equal size in the SIRS setting. Their work built on the work in Rowthorn et al. (2009) and they identified several simple to implement resource allocation strategies which generally performed well in controlling the disease. They also considered more complex strategies which switch between these and numerically demonstrated their near-optimality. We aim to analyse how to generalize the results to the situation when there are $n$ populations of varying sizes. We focus on the case when the treatment is strong enough to eradicate the disease.

4.5.1 Model description

The model structure is the same as in the previous section, the only difference being that the dynamics are now SIRS, that is we have an extra recovered class. If we denote the proportion of the total resource allocated to the population $i$ by $x_i$, so $\sum_i x_i = 1$, the optimization problem can be phrased as

$$\min \sum_{i=1}^n N_i I_i^\infty,$$  

given

$$\frac{dI_i}{dt} = \beta I_i (1 - I_i - R_i) - I_i - \eta \min \left( I_i, \frac{x_i(t)M}{N_i} \right),$$  

$$\frac{dR_i}{dt} = I_i + \eta \min \left( I_i, \frac{x_i(t)M}{N_i} \right) - \nu R_i,$$  

$$\sum_{i=1}^n x_i(t) = 1,$$

where $\nu$ is the rate at which the immunity of the recovered hosts wanes.
4.5 SIRS model with continuous resource reallocation

Fig. 4.7 Plots showing the comparison of the TLI, TMS and TMI strategies as a percentage of the entire population. The middle and the right scatterplots show the comparison of TLI and TMS with TMI. We can see that both TMS and TLI outperform TMI, demonstrating once again that targeting the populations one by one is better than trying to eradicate the disease in all the populations at once. The leftmost plot shows the comparison of TLI and TMS. We can see that sometimes TLI is better and sometimes TMS is. The parameter values are $\beta = 2$ and $\eta = 1.2$ and $\nu = 0.5$.

4.5.2 Analysis of the problem

As before, due to the complexity of the problem, we initially focus on a few simple strategies. We consider the TLI and TMI from the previous section, and we add TMS, or “treat the most susceptible population” which allocates the treatment preferentially to the populations with the largest number of susceptible individuals (Ndeffo Mbah and Gilligan, 2011). Note that in the SIS model, TLI and TMS are the same strategy. Simulation shows (Fig. 4.7) that while both TLI and TMS outperform TMI as before, for the same reasons, there is no obvious way of distinguishing which one of them should be used. Sometimes TLI gives better results, sometimes TMS does. We generated 500 random sets of populations sizes, initial conditions and resource constraints and plotted the strategy comparison in figure 4.7. Similarly to the SIS model, the reason why TLI and TMS sometimes fail when compared with each other is that while they both attempt to eradicate the disease in the populations one by one, sometimes they miss the urgent populations, that is the populations where we can eradicate the disease initially, but not necessarily later.

We define a TMU strategy as before. Its implementation is slightly more complicated in the SIRS setting. While in the SIS setting, the criterion for whether we can eradicate or not was simply whether the disease prevalence is below or above a certain critical value, in the SIRS setting, no such simple criterion exists. In general, the disease can
Fig. 4.8 Plot showing the comparison of the TLI and TMS strategies with the TMU strategy. We can see that the TMU strategy outperforms both the TLI and TMS in the majority of the cases, with few rare exceptions, which we discuss in the text. The parameter values are $\beta = 2$, $\eta = 1.2$ and $\nu = 0.5$. The labels A and B refer to the cases highlighted in Figures 4.9 and 4.10.

be eradicated from a certain initial condition given a fixed amount of resources, if the initial condition lies outside of the basin of attraction of the fixed point $D$ (one of the stable endemic equilibria found in chapter 2). This basin of attraction is not simple, its boundary is sometimes formed by a homoclinic orbit, sometimes by a limit cycle and we cannot obtain an analytical expression for it. Therefore, whenever we need to decide which populations belong to which class A, B or C (defined as in the SIS case), we need to simulate them separately to determine whether the current values of $I$ and $R$ lie inside or outside of the basin of attraction. This makes the model run considerably slower than in the SIS setting.

We ran the TMU strategy on the same set of 500 landscape realizations as above and compared its performance with the TLI and TMS strategies. The results can be found in figure 4.8. The figure shows that TMU is again better than TLI and also than TMS. However, there are occasional cases where this is not the case. In one of the random realizations, TLI performed better than TMU (although only by about 1%). We plot that particular scenario in Figure 4.9. Both the strategies failed to eradicate the purple and the red populations, but TLI didn’t eradicate the teal while TMU didn’t eradicate the green. As it happened in that realization, the teal population had a smaller size than the green one and therefore TLI achieved a slightly better result.
The reason for this is that TMU itself uses a heuristic to decide on the order in which to treat the populations within each urgency class, in fact it uses the TLI heuristic. This is a good approach because when there are more populations in the most urgent class, it leads to a good chance of being able to eradicate the disease in all of them. However, occasionally, like in this case, it leads to a situation in which the size of the population in which the disease is not eradicated is larger than if another strategy was employed.

The reason why TMS outperformed TMU in three of the realizations is even more interesting. We selected the scenario in which the difference between the objective function of TMU and the objective of TMS was the largest for illustration and we plot the disease progress curves in Figure 4.10. In this scenario, all the populations actually belong to the class B, meaning we can eradicate the disease even if they start in their treatment free endemic equilibrium. The TMS strategy therefore manages to eradicate the disease in the populations one by one. How is it possible then that the TMU does not eradicate the disease in the teal population? This is actually an example of the counter-intuitive behaviour identified in Chapter 2, where delaying treatment can lead to eradication. If TMU did not treat the teal population at all, but instead allowed it to reach its treatment free endemic equilibrium and only after that started the treatment, it would eradicate the disease, because the treatment free equilibrium lies outside of the basin of attraction of the fixed point $D$. However, as the system evolves in the phase space towards this equilibrium, it passes through the
4.6 Discussion

In this chapter, we extended the model for optimizing resource allocation between isolated populations to include the possibility of reallocating the resources after the initial allocation. This is an important generalization for a couple of reasons. Firstly, reallocating resources can be costly and investigating the impact reallocation can have on the objective function is necessary in order to judge whether we should invest in the reallocation itself. Therefore, in the first section of this chapter, our aim was to answer the question of to what extent can allowing reallocation improve the objective function. A similar question has been investigated before (Zaric and Brandeau, 2002), albeit in a different model (*Susceptible-Infected* with births and deaths) and with a different objective function (quality-adjusted life years or the total number of new infections). Zaric and Brandeau (2002) found that reallocating resources can significantly improve the objective function.

We consider the multiple threat metapopulation model from the previous chapter and allow for reallocation whenever the system has reached an endemic equilibrium.
This avoids the need to introduce the time interval of the reallocation as an arbitrary parameter into the system. At the same time, it captures the essential feature of allowing the reallocation of resources, which is that after the prevalence of the disease decreases in a population as a result of the initial resource allocation, we can move the resources elsewhere. We show that depending on the exact combination of the parameter values in the different populations, reallocating resources can indeed lead to significant decreases in the objective function. The intuition behind this is very simple in the case where the treatment is strong enough to eradicate the disease (that is $R_0$ given full treatment is less or equal to 1). Consider the following idealised example. Given a number of populations, all in the treatment free endemic state, and resources enough to eradicate the disease in one of them only, without reallocation we will only succeed in eradicating the disease in a single population. If reallocation is allowed, we can move the resources from population to population, until the disease is eradicated everywhere.

When the treatment is not strong enough to eradicate the disease, we show that the order in which the populations are saturated is crucial for achieving the optimal outcome. We identify the condition on the amount of resource required to achieve the optimum and derive a formula for the order in which the populations should be saturated.

In the second part of the chapter we assume that the resources can be reallocated continuously in time. This means that the resources can be moved smoothly between the populations as required. This would be appropriate for example when the cost of reallocation is very low and when reallocation can be achieved very quickly, compared with the evolution of the epidemic. It is also interesting from a mathematical perspective, since it is in a sense the opposite extreme to the “no reallocation allowed” condition. A similar model of continuous resource reallocation in two equal-sized populations was considered before in Rowthorn et al. (2009); Ndeffo Mbah and Gilligan (2011). Here, our aim was to build on their work and extend it to the case of multiple populations of varying sizes. As in the previous chapter, we were interested in the indirect interaction provided by the shared resource pool and for this reason we assumed that the populations are otherwise uncoupled.

We investigated both the SIS (Rowthorn et al., 2009) and SIRS (Ndeffo Mbah and Gilligan, 2011) model dynamics. We showed that previous approaches to optimizing the resource allocation are insufficient when the population sizes are allowed to vary.
Instead, we identified a better allocation strategy, which is based on the understanding of the dynamical properties of the model with constrained control developed in Chapter 2. The strategy is based on preferentially allocating the resources to the populations based on their measure of urgency, which is defined as follows. A population is urgent if, given the initial conditions, we can eradicate the disease in the population now, but not necessarily later, if we wait. To compare the urgency-based strategy with the previously considered heuristics, such as allocating preferentially to the population with smallest disease prevalence (TLI) or to the population with the largest number of susceptible individuals (TMS), we ran simulations on a number of randomly selected landscapes. In both the SIS and SIRS models, the urgency-based strategy and the previously considered heuristics gave the same result in the majority of cases. However, in about 16% of the cases, the urgency based strategy was clearly superior, by as much as 40%. Additionally, in the SIRS case, in a few rare cases the urgency-based strategy underperformed the TLI heuristic by about 1%. This is because of the complexities of the SIRS model with constrained control discussed in chapter 2, for example the counter-intuitive phenomenon where sometimes having more control resources leads to a worse overall result.
Chapter 5

Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective

5.1 Abstract

In this chapter, we focus on a specific type of disease control, the deployment of crop varieties that are partially resistant to plant pathogens. While this is a common type of control, a trade-off may occur between the benefits of planting the resistant variety and a yield penalty, whereby the standard susceptible variety out-yields the resistant one in the absence of disease. This presents a dilemma: deploying the resistant variety is advisable only if the disease occurs and is sufficient for the resistant variety to out-yield the infected standard variety. Additionally, planting the resistant variety carries with it a further advantage in that the resistant variety reduces the probability of disease invading. Therefore, viewed from the perspective of a grower community, there is likely to be an optimal trade-off and thus an optimal cropping density for the resistant variety. We introduce a simple stochastic, epidemiological model to investigate the trade-off and the consequences for crop yield. Focusing on SIR epidemic dynamics, we use the final size equation to calculate the surviving host population in order to analyse the yield, an approach suitable for rapid epidemics in agricultural crops. We
identify a single compound parameter which we call the efficacy of resistance and which 
incorporates the changes in susceptibility, infectivity and durability of the resistant 
variety. We use the compound parameter to inform policy plots that can be used to 
identify the optimal strategy for given parameter values when an outbreak is certain. 
When the outbreak is uncertain, we show that for some parameter values planting the 
resistant variety is optimal even when it would not be during the outbreak. This is 
because the resistant variety reduces the probability of an outbreak occurring. This 
chapter is based on a published work (Vyska et al., 2016).

5.2 Introduction

Previous chapters focused on identification of heuristics for allocation of control resource 
between multiple separate populations. Here we switch back to a single population and 
consider a specific type of control of an SIR epidemic. We study the deployment of crop 
varieties that are genetically resistant to plant pathogens. The simplest case involves 
complete resistance to a pathogen that remains durable for long periods of time without 
being overcome by new virulent strains of the pathogen (Johnson, 1984; Mundt, 2014). 
Newly released resistant varieties should ideally also match or surpass the agronomic 
properties of the susceptible varieties that they replace. Historically many resistant 
varieties exhibited a qualitative resistance that is a form of complete resistance in which 
the pathogen is unable to infect the host (Jones and Dangl, 2006). There is increasing 
awareness, however, that qualitative resistance under single gene control in the host 
implies such strong pressures on the pathogen population to overcome resistance so 
that failure becomes almost inevitable (Crute et al., 1997; Brown, 2015). Accordingly, 
increasing attention is being paid to the release of partially resistant varieties that slow 
or otherwise reduce, but do not prevent, infection and multiplication of the pathogen 
on the host crop (Brown, 2015; Singh and Huerta-Espino, 1997). Such resistance is 
frequently under the control of few to many genes in the host crop (Burdon et al., 
2014; Ellis et al., 2014). Trade-offs may occur, however, between partial resistance 
and the agronomic properties of the host crop such that partial resistance may be 
associated with lower yield for reduced quality in the crop variety compared with the 
standard (Anon, 2015; Brown, 2002; Brown and Rant, 2013; Kjær et al., 1990). This 
presents individual growers with a dilemma: growing the resistant crop imposes a 
yield penalty that may be compensated for only if disease occurs and is sufficient for
the infected, partially resistant crop to out-yield an equivalently infected standard crop. The decision is further complicated: as more growers elect to plant the partially resistant crop the probability of a severe epidemic occurring decreases. Viewed from the perspective of a community of growers there is likely to be an optimal trade-off and hence a cropping density for the proportion of sites that are planted to the partially resistant variety. Moreover, even if an outbreak is certain, it is not necessary for all growers to have planted the partially resistant variety for the benefit to be realised by the community of growers as a whole (van den Bosch and Gilligan, 2003; Lo Iacono et al., 2013).

Previous epidemiological analyses on the introduction of partially resistant varieties in the landscape have focused on the effects of disease dynamics (van den Bosch and Gilligan, 2003; Lo Iacono et al., 2013; Papaïx et al., 2011; Fabre et al., 2012). Here we focus on the trade-off between disease and yield at the landscape scale. The primary aim of this chapter is, therefore, to provide insight into when it is worth deploying a partially resistant variety and in what proportions relative to a standard, higher yielding, susceptible variety. We do this using simple stochastic epidemiological models that characterise the spread of the pathogen, and hence disease, and the consequences for crop yield. We identify policy plots (Forster and Gilligan, 2007) that can be used to infer optimal strategies given some prior knowledge of the resistance-yield trade-offs. We make certain simplifying assumptions that we subsequently relax.

Our intention in this chapter is to address the generic problem of how to deploy partial resistance when there is a yield penalty and uncertainty. We intentionally propose a flexible modelling framework that encompasses a variety of epidemiological mechanisms that could be associated with partial resistance, with broad applicability to a range of host pathogen systems. Hence, we consider SIR epidemics, in which susceptible hosts (S) become infected (I), remain infectious for a period of time and then are removed (R). Removal may occur naturally by disease-induced death or by deliberate removal, for example by roguing of infected plants. SIR epidemics are typified by an increase in infected hosts followed by a decrease as the epidemic ‘burns itself out’ or is controlled. We assume that yield is a function of the amount of healthy, i.e. uninfected host, allowing for a yield penalty for healthy hosts of the partially-resistant compared with the susceptible variety. We initially assume that yield is accumulated over a long period of time relative to the period of crop growth and the time course of the epidemic. This assumption enables us to gain analytical insight to inform the
deployment of partially-resistant crops subject to yield penalties. We subsequently relax the assumptions to test the robustness of the conclusions. Our initial results apply to annual crops, in which epidemics happen fast, typified by potato late blight and rusts of small grain cereals. Here, $S$ and consequently $I$, are expressed as units of plant host tissue (Zadoks and Schein, 1979). Our results also apply to cassava virus diseases, in which yield is accumulated over a long period and roguing is practiced to remove infected hosts. In this case, $S$ and $I$ refer to whole plants. The model may also be applied, more generally, to perennial crops in which roguing of infected plants occurs, yield accumulates over long periods and there is continuing harvesting. We first consider the case where epidemics are inevitable and address two broad questions: under what circumstances is investing in the partially resistant variety likely to be profitable, and how much of the resistant variety should be deployed to maximise yield? We subsequently consider the robustness of the inferences about the deployment of the partially resistant variety when the occurrence of the epidemic is uncertain.

The rest of this chapter is structured as follows. We begin with a brief mathematical motivation for this work provided by the observation that in the SIR model, the final size of the epidemic will always be smaller if hosts are preventively culled prior to the epidemic. This can be viewed as an extreme case of the above mentioned resistance-yield trade-off with perfect “resistance” and no yield. Then, we turn to the more general case of disease resistance with a fitness penalty and therefore reduced yield. The work in this chapter is based on the paper (Vyska et al., 2016).

5.3 Motivation from a simple model

We begin with a brief discussion of a mathematical observation that motivates the main subject of this chapter, the resistance-yield trade-off. Suppose that the trade-off is perfect, that is the resistant hosts cannot be infected by the pathogen but do not contribute at all to the yield. This would be the case for example if instead of replacement by a resistant variety, the hosts were pre-emptively culled (thinned) (Humes et al., 1999; Vaast et al., 2005) or, in the case of forestry, replaced by a species that does not produce marketable timber.
We again use the familiar deterministic SIR model given by the set of equations

$$\frac{dS}{dt} = -\beta IS \tag{5.1}$$

$$\frac{dI}{dt} = \beta IS - \mu I \tag{5.2}$$

$$N = S + I + R. \tag{5.3}$$

As before, $\beta$ is the transmission rate and $\mu$ is the removal rate of the infected hosts. $S$, $I$ and $R$ stand for the numbers of the susceptible, infectious and recovered hosts and $N$ is the total number of hosts. This model is suitable for diseases which confer immunity upon recovery. Alternatively, in the case of plant pathogens, which are the main focus of this chapter, the $I \rightarrow R$ transition does not correspond to recovery but rather to the removal of the infectious hosts from the population. The model assumes density-dependent mixing between the individuals, which means the higher the density, the more likely the disease is to spread. Such an assumption corresponds well to plant disease where the infectious individuals produce inoculum that is then dispersed by wind or vectors. The more susceptible hosts that are in the vicinity, the more likely the inoculum is to come into contact with a susceptible host. This is made explicit in the model by the dependence of the basic reproductive number on the total population size,

$$R_0 = \frac{\beta N}{\mu}. \tag{5.4}$$

As the susceptible individuals become infected eventually their numbers decrease so that $S = \mu / \beta$, at which point each infectious host infects on average less than one new host during its infectious period. Therefore, the chain of transmissions is eventually broken and the epidemic ends with all the individuals either in the susceptible or removed class. The final size of the removed class $R(\infty)$ is referred to as the final size of the epidemic, since it corresponds to the total number of hosts that have been infected. For the simple SIR model, this is given by the final size equation (Diekmann et al., 2012),

$$N - R(\infty) = N \exp \left( -\frac{R_0 R(\infty)}{N} \right). \tag{5.5}$$

Consider the following theorem.

**Theorem 6.** The final size of the epidemic $R(\infty)$ as defined in (5.5) is always greater than $N(1 - 1/R_0)$ as long as $R_0 > 1$. Furthermore, within a deterministic framework,
Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective

\[ N(1 - 1/R_0) \text{ is precisely the number of hosts that would need to be removed in order to prevent the epidemic from taking off.} \]

**Proof.** Clearly, if we remove \( N(1 - 1/R_0) \) of the hosts, there will be only \( N/R_0 \) hosts left. The new basic reproductive number \( R'_0 \) will be given by \( R'_0 = \beta N/R_0 \mu = 1 \) which is the threshold for preventing the epidemic from taking off.

For the first part of the theorem, let \( x^* \) be the proportion of hosts lost due to the epidemic, that is \( x^* = R(\infty)/N \) and let \( x_T \) be the proportion of hosts lost due to thinning to bring the basic reproductive number down to 1 as above, that is \( x_T = 1 - 1/R_0 \). We need to show that \( x^* > x_T \).

If \( R_0 > 1 \), the curves \( x^* \) and \( x_T \) cannot cross. This is because when \( x^* = x_T \) we would have

\[ e^{R_0} = eR_0 \]

which is impossible. Therefore, on the interval \( R_0 \in (1, \infty) \) we have either \( x^* > x_T \) or \( x^* < x_T \). It is straightforward to check that at \( R_0 = e/(e-1) \), \( x^* > x_T \) and therefore \( x^* > x_T \) for all \( R_0 > 1 \).

This means that in the simple deterministic SIR model, thinning the host population and reducing the host density down to make the basic reproductive number 1 will always lead to smaller losses in terms of numbers of plants lost than letting the epidemic take place. This has a simple intuitive explanation. When \( R_0 \) is one, there is no epidemic outbreak. However, when \( R_0 > 1 \) once the epidemic starts, it won’t stop before it decreases the effective \( R_0 \) to one and at that point it will still continue for some time as the prevalence decreases to 0. Therefore the losses caused by the epidemic will always be greater than the losses due to culling the hosts to achieve \( R_0 = 1 \). The differences can be significant too. Consider for example \( R_0 = 2 \). Numerically solving the final size equation (5.5) shows that the losses due to the epidemic are 79.7%. However, culling 50% of the hosts prevents the epidemic altogether and therefore it is better by 29.7%.

In the next section, we introduce the full resistant-yield trade-off model where some of the hosts are partially resistant to the pathogen at the cost of reduced yield.
5.4 Partial resistance and yield trade-off

5.4.1 Methods

Epidemic model

We consider disease spreading through a metapopulation, comprising two types of host crop, a partially resistant and a fully susceptible “standard” variety. Plants of each variety can be in one of three classes: susceptible (\(S\), i.e. healthy), infected (\(I\)) or removed (\(R\), i.e. post-infectious). The principal parameters used in the models for disease spread and yield are summarised in Table 1, for ease of reference. We make the following assumptions:

- The epidemic follows SIR compartmental dynamics, with density-dependent mixing (Keeling and Rohani, 2008). The \(I \rightarrow R\) transition is realized either by disease-induced mortality, by roguing of the infected hosts, or by a combination of both.

- Hosts in the infected and removed classes do not contribute to the yield.

- The resistant hosts are less likely to become infected upon contact with the pathogen (by a factor \(\eta \in [0, 1]\)), produce less inoculum (by a factor \(\nu \in [0, 1]\)), have a different infectious period (by a factor \(\sigma \in (0, \infty)\)) and contribute less to the final yield (by a factor \(f \in [0, 1]\)). It is natural to consider both shorter and longer infectious periods for the resistant variety, because the resistant hosts might take longer to die or they might take longer to develop visible symptoms upon infection and thus avoid detection and subsequent removal (by, for example, roguing of symptomatic hosts in certain crops) for longer than the standard hosts.

We assume a fraction \(\rho = N_R/(N_S + N_R)\) of all the hosts to be resistant, where \(N_S\) and \(N_R\) are the numbers of standard and resistant hosts, respectively. The model for disease spread is given by (where subscripts S and R denote standard and resistant
Trade-off between disease resistance and crop yield: a landscape-scale mathematical modeling perspective

<table>
<thead>
<tr>
<th>Parameter, Value/Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$, [0,1]</td>
<td>Proportion of the resistant hosts in the population.</td>
</tr>
<tr>
<td>$\beta$, both 2 and 3 used</td>
<td>Transmission rate between the standard hosts.</td>
</tr>
<tr>
<td>$\mu$, 1</td>
<td>Infectious period of the standard hosts set to 1.</td>
</tr>
<tr>
<td>$\eta$, [0,1]</td>
<td>Susceptibility factor: reduction of susceptibility of the resistant hosts.</td>
</tr>
<tr>
<td>$\nu$, [0,1]</td>
<td>Infectivity factor: reduction of infectiousness of the resistant hosts.</td>
</tr>
<tr>
<td>$f$, [0,1]</td>
<td>Yield penalty: reduction of yield of the resistant hosts.</td>
</tr>
<tr>
<td>$\sigma$, (0, $\infty$)</td>
<td>Removal factor: change in the removal rate of the resistant hosts.</td>
</tr>
<tr>
<td>$R_0$, both 2 and 3 used</td>
<td>Basic reproductive number in a population of standard hosts.</td>
</tr>
<tr>
<td>$\xi$, [0,1]</td>
<td>Resistance efficacy, $\xi = 1 - \eta \sigma \nu$, see the Results section.</td>
</tr>
<tr>
<td>$\lambda$, 2</td>
<td>Probabilistic rate of import of the pathogen, see Section 2.3.</td>
</tr>
</tbody>
</table>

Table 5.1 List of all the parameters used.

$$\dot{S}_S = -\beta S_S[(1-\rho)I_S + \nu \rho I_R]$$ (5.7)
$$\dot{S}_R = -\eta \beta S_R[(1-\rho)I_S + \nu \rho I_R]$$ (5.8)
$$\dot{I}_S = \beta S_S[(1-\rho)I_S + \nu \rho I_R] - \mu I_S$$ (5.9)
$$\dot{I}_R = \eta \beta S_R[(1-\rho)I_S + \nu \rho I_R] - \frac{\mu}{\sigma} I_R$$ (5.10)

$$S_S + I_S + R_S = S_R + I_R + R_R = 1,$$ (5.11)

where $\beta$ is the transmission rate, $\mu \ (\mu/\sigma)$ is the removal rate of the standard (resistant) hosts respectively and all the state variables ($S_S, S_R$ etc.) are proportions of hosts in the respective class. It is convenient to introduce the basic reproductive number of the standard hosts in a monoculture, $R_0 = \beta/\mu$. This is defined in the standard way as the average number of secondary infections caused by a single infected host over the course of its infectious period.

**Yield without uncertainty**

When an outbreak is certain, we represent the yield per host and per unit time, accumulated by healthy plants over a fixed period $T$, by a straightforward adaptation...
of the integral over the susceptible hosts, a simple measure used in plant epidemiology (Madden et al., 2007; Cunniffe et al., 2015a)

\[ Y = \frac{1}{T} \int_0^T [(1 - \rho)S_S(t) + f\rho S_R(t)]dt. \]  

(5.12)

To simplify the analysis we consider the yield accumulated over a very long period of time relative to the duration of the longest epidemic \( T \to \infty \) in the above equation, so that it is effectively given by the proportions of susceptible hosts that survive the epidemic. We include an explanatory sketch in Appendix A1. Therefore the yield \( Y \) is given by

\[ Y = (1 - \rho)S_S^\infty + f\rho S_R^\infty. \]  

(5.13)

where \( S_i^\infty \) is the proportion of the susceptible hosts in the population \( i \in \{S, R\} \) after the outbreak has ended. Note that this assumption is made purely for mathematical convenience to make analytical progress, from which initial insights into the optimal strategy can be inferred. We subsequently relax this assumption. We introduce an arbitrary, finite-time horizon, \( T_E \) to Eqn 5.12. We show that the inferences derived for an infinite-time horizon hold when yield is accumulated over a finite period of time. The choice of \( T_E \), while arbitrary, is motivated by keeping the analysis generic and to avoid introducing another parameter to the model.

**Yield with uncertainty**

When there is uncertainty about whether or not an outbreak will occur, the yield function is given by the expected yield:

\[ Y(\rho) = p_E \times \text{Yield(outbreak)} + (1 - p_E) \times \text{Yield(no outbreak)} \]  

(5.14)

\[ = p_E[(1 - \rho)S_S^\infty + \rho f S_R^\infty] + (1 - p_E)[1 - \rho + f \rho], \]  

(5.15)

where \( p_E \) is the probability of an epidemic. To model this probability, we consider a situation in which the pathogen has a constant small rate of introduction into the host population over some period of time (for example because the climatic conditions are favourable during this period). The number of introductions then follows a Poisson distribution with mean \( \lambda_1 \). Each time the pathogen is introduced into the system, it infects an initial host with probability \( \lambda_2 (1 - \rho + \eta \rho) \) where \( \lambda_2 \) is the probability of a standard host getting infected upon contact with the pathogen. For convenience,
we define $\lambda = \lambda_1 \lambda_2$. Once the initial host has been infected, a large-scale outbreak will occur with probability $P_{\text{takeoff}}$ which can be calculated using standard arguments, see Appendix A2. Putting this together and using the thinning property of Poisson processes (Chiu et al., 2013) leads to the expression for the overall probability of an epidemic as

$$p_E = 1 - e^{-\lambda(1-\rho+\eta \rho)P_{\text{takeoff}}}.$$  \hspace{1cm} (5.16)

**Summary of the assumptions**

Here, for convenience and reference, we summarise the principal assumptions of the model and the subsequent analyses. We approach the problem generically, using biologically plausible parameter values and ranges to reflect classes of host-pathogen system rather than restricting the analysis to a single system. Instead, our analysis is designed to identify which parameters are important and what are their critical ranges. We introduce flexibility by allowing for the epidemiological mechanisms accounting for partial resistance to be expressed through changes in one or more of the following: the infectivity of infected hosts, the susceptibility of healthy hosts or the length of the infectious period. These effects can be tuned independently in our model through different parameters. In our model the unit of host can be either a whole plant or healthy tissue and we assume the infected hosts eventually die or are removed through roguing. For simplicity we assume that neither infected nor removed hosts contribute to the yield. This means there is no replanting or that it takes a long time for a replanted host to reach maturity, such as in the case of tree crops. The yield is modelled as an integral over the healthy hosts. Initially, in order to make the model analytically tractable we assume the yield is accumulated over a time-scale much longer than the epidemic duration so that it can be approximated by an infinite time horizon. In the second part of the analysis, we relax this assumption and consider yield accumulated only over the duration of the epidemic. The main purpose of this is to verify that our qualitative results are not simply an artefact of the infinite time horizon. Finally, when we allow for uncertainty in the occurrence of an epidemic, we assume that the pathogen has a constant rate of import into the host population over a certain time period preceding the potential epidemic.
5.4.2 Results

Model without uncertainty

Using the final size equations for an epidemic (Keeling and Rohani, 2008), it is possible to derive an analytic expression for the yield function (see the Appendix A3 for details),

\[ Y(x) = x - (x - f x^n) \frac{\ln(x) + R_0(1 - x)}{R_0[1 - x - \nu \sigma(1 - x^n)]}, \]  

(5.17)

where \( x \equiv S^S_\infty \) is the final size of the susceptible class of the non-resistant hosts and depends on \( \rho \). It is not possible to obtain an analytic expression for the yield as a function of \( \rho \) directly. However, the analytic solution (i.e. 5.17) shows three possible control scenarios to optimise yield. These are illustrated for three different yield penalties associated with the partially resistant variety in Fig. 5.1. There are two extreme scenarios: no control, i.e. grow only the standard variety when the yield penalty is high (Fig. 5.1A) and “full” control, i.e. grow sufficient resistant variety to bring the basic reproductive number below 1 (see below) when the yield penalty is low (Fig. 5.1C). As the yield penalty increases, so the cropping ratio of standard to resistant variety required to achieve an optimal yield increases leading to an intermediate control scenario (Fig. 5.1B). Note that the “full” control (Fig. 5.1C) does not necessarily mean \( \rho = 1 \) but corresponds to the density \( \rho \) being equal to

\[ \rho = \rho_{\text{full}} = \min(1, \rho_c), \]  

(5.18)

where \( \rho_c \) is a critical density such that the effective reproductive number of the system (1-4) \( R'_0 \) in the presence of the resistant hosts falls below 1. This is because once the basic reproductive number falls below one, the epidemic is prevented and further deployment of the resistant variety has no effect. The effective basic reproductive number \( R'_0 \) can be calculated using the next generation method (Heffernan et al., 2005): it is given by

\[ R'_0(\rho) = R_0(1 - \rho \xi), \]  

(5.19)

where we introduce the parameter \( \xi = 1 - \eta \sigma \nu \) which we will refer to as the resistance efficacy. From equation (5.17) we conclude that the optimal proportion \( \rho \) only depends on \( \nu \) and \( \sigma \) through their product \( \nu \sigma \). This allows us to make 2D policy plots showing where different types of control are optimal, for various values of yield penalty, \( f \) (see
Fig. 5.1 Three possible control scenarios, and optimal yields illustrated for three yield penalties associated with a partially resistant variety, \( f=0.1 \) (A), \( f=0.3 \) (B) and \( f=0.55 \) (C). The vertical red lines show the range of possible results, that is they show the final sizes corresponding to no control and to full control. The green dot shows the optimal final size in each case. Default parameter values: \( R_0 = 2 \), \( \eta = 0.79 \), \( \nu\sigma = 0.78 \). The values were selected in order to illustrate the three distinct types of behaviour.

Fig. 5.2 A Policy plot for \( f = 0.3 \) and \( R_0 = 3 \). The dark region corresponds to \( \xi < 0 \) that is controls that increase the basic reproductive number and support the spread of the pathogen, which we do not consider further. B shows how much yield per host is lost when we ignore the intermediate control, in the worst case scenario. Mathematically, the function plotted is \( \max_{\eta, \sigma\nu}(Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}})) \).
5.4 Partial resistance and yield trade-off

Fig. 5.2 for an example). Note that when $\xi = 1 - \eta \sigma \nu < 0$ (the dark region in Fig. 5.2A), the control supports the spread of the pathogen and we therefore do not consider this region any further. Examination of Fig. 5.2A and corresponding figures for a range of values of $f$ shows that the region of the parameter space where intermediate control is optimal is small. Accordingly we investigate the potential for loss in yield if the option of intermediate control is ignored, using $Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}})$ as a metric for yield loss in choosing no or 'full' control in place of intermediate control (cf. Fig. 5.1). The metric is a function of parameters $f$, $\eta$, $\nu$ and $\sigma$. In Figure 5.2B, we plot the worst yield loss as a function of $f$, that is $\max_{\eta,\sigma}(Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}}))$. The losses are small and decrease very quickly with $f$. Furthermore, in the $(\eta, \nu \sigma)$ plane they are only appreciable along the curve $Y_{\text{full}} = Y_{\text{none}}$ and close to $\eta = 1$, where the relative susceptibilities of resistant and standard varieties are identical. The explanation of this can be found in the Appendix A4. This means that for the most part, we can focus on the extreme controls since they are mostly optimal and when they are not, they provide a good approximation to the optimum.

**Extreme control optimization**

From equation (5.17) we can derive the conditions for deployment of a resistant variety under which “full” control is better than no control. There are two cases to consider. If the resistance is effective enough so that the outbreak can be prevented altogether, that is if $R'_0(\rho = 1) = (1 - \xi)R_0 < 1$, full control is better than no control when

$$\xi > \frac{(1 - 1/R_0)(1 - f)}{1 - s_0}$$  \hspace{0.5cm} (5.20)

where $s_0$ is the proportion of surviving hosts when no control is deployed. If the resistance is not effective enough to prevent the outbreak, that is if $(1 - \xi)R_0 > 1$, full control is better than no control when

$$\xi > 1 - \frac{f \ln(f/s_0)}{R_0} \frac{f}{f - s_0}.$$  \hspace{0.5cm} (5.21)

These conditions are derived in the Appendix A4. Note that these conditions only depend on the parameters $\eta$, $\sigma$ and $\nu$ via the resistance efficacy $\xi = 1 - \eta \sigma \nu$ and therefore there are effectively only three controlling parameters: $R_0$, $\xi$ and $f$. This allows us to plot a policy diagram, showing when the resistant variety is worth deploying.
Fig. 5.3 **A** Policy plot for $R_0 = 3$ to inform control scenarios when the resistance efficacy and yield penalty for the resistant variety are known. **B** Policy plot to show sensitivity of yield at optimal strategy to changing $f$ (yield penalty) and $\xi$ (resistance efficacy). The colour scale corresponds to the difference of these two rates, $\frac{\partial Y}{\partial f} - \frac{\partial Y}{\partial \xi}$. The additional black lines in 5.3B mark the boundaries of the blue and the red regions.

The plot for $R_0 = 3$ can be found in Fig. 5.3; other basic reproductive numbers are not qualitatively different. Figure 5.3A shows the region where full control is optimal and the boundary, which corresponds to the critical value of $\xi$. Figure 5.3B shows whether the optimal value of the yield, $Y$, increases more quickly as the yield penalty $f$ increases or when the resistance efficacy, $\xi$, increases, using the quantity $\frac{\partial Y}{\partial f} - \frac{\partial Y}{\partial \xi}$ as a metric. When the metric is positive, increasing $f$ is more important and when it is negative increasing $\xi$ is more important to increase yield. The effect of changing $\xi$ is most dramatic near the boundary $\xi = 1 - 1/R_0$ and therefore when $\xi$ is close to this boundary from below, it is much more profitable to increase it above the boundary than to attempt to increase the yield factor.

**Model with uncertainty**

The probability of invasion, $p_E$, derived from equation (5.16) decreases as the cropping ratio of the resistant hosts increases (Fig. 5.4). The decline in $p_E$ is steeper when the susceptibility factor ($\eta$, Table 1) of the resistant fraction of the host population becomes smaller, that is, when this fraction of hosts presents an increasing level of resistance. To investigate the impact on the yield of the deployment of the resistant variety in the
Fig. 5.4 Probability of an outbreak as a function of the proportion of the resistant hosts in the population. Parameters are $\lambda = 2$, $\xi = 0.2$ and $R_0 = 3$.

presence of uncertainty, we repeat our analysis from the previous section. Similarly to the case of a deterministic, certain outbreak, numerical analysis reveals that a good approximation is provided by only considering extreme controls, that is none or “full”. Note that because of the introduction of the probability $p_E$ there are now four independent controlling parameters, $\xi$, $R_0$, $f$ and $\eta$. In Figure 5.5 we show the policy plots corresponding to three different values of $\eta$ and $R_0 = 3$ for both the deterministic and the stochastic case. We can see that when the resistance significantly reduces the susceptibility to the pathogen, but its efficacy $\xi$ is low overall because $\sigma > 1$, it can happen that the control is optimal only in the presence of uncertainty. This means that while the control is not desirable during the outbreak, the benefits of the possibility of preventing the outbreak altogether are significant. We can formalize this by defining $\Delta Y = Y(\text{Optimal}) - Y(\text{No control})$, that is the yield gained by controlling correctly as opposed to not controlling at all. In Figure 5.6, we plot the difference between the yield gained by controlling with uncertainty and without uncertainty. We can see that without uncertainty the control provides greater yield gains for high resistance efficacy but, in agreement with Figure 5.5, when the resistance efficacy is low, the susceptibility factor $\eta$ is low and the yield parameter $f$ is high, the control provides significantly greater yield gains when the uncertainty is present. This demonstrates the importance of reducing the probability of an outbreak occurring even when during the outbreak the control is not desirable.
Fig. 5.5 $\lambda = 2$, $R_0 = 3$. Policy plots for three different values of $\eta$, with and without allowance for uncertainty in an outbreak. Note that when the resistance is relatively ineffective ($\xi$ low), but the resistant hosts are significantly less susceptible to the pathogen, full control becomes optimal when an outbreak is uncertain when it would not be optimal in the deterministic setting.

Fig. 5.6 $\lambda = 2$, $R_0 = 3$. The colour shows how much more yield is gained by the control when the uncertainty is present over when the outbreak is certain, mathematically $\Delta Y(\text{stochastic}) - \Delta Y(\text{deterministic})$. We can see that when the resistance efficacy is low, $f$ is high and the susceptibility is significantly reduced by the resistance, the control provides higher gains in yield when the uncertainty is present. This demonstrates the importance of reduction of the probability of an outbreak occurring even when during the outbreak the control is not desirable.
Yield model with finite time horizon

So far we have considered long-term yield. We have shown that contrary to intuition, the benefit of deploying a resistant variety can be greater when the disease outbreak is uncertain. We now verify this result when the assumption of the infinite time horizon is relaxed and the yield to be maximized is evaluated over a finite period of time. We selected the time of the duration of the epidemic $T_E$ as a representative of this finite time horizon. Therefore the yield is now given by

$$Y = \frac{1}{T_E} \int_{0}^{T_E} [(1 - \rho)S_S(t) + f\rho S_R(t)]\,dt. \quad (5.22)$$

Note that we assume no crop growth in this period. Whereas our initial model focuses on the final state of the epidemic, the beginning of the epidemic and how quickly the pathogen invades now become important factors. Note that since increasing the cropping ratio $\rho$ of the resistant variety leads to a decrease in $R_0$ of the system, the duration of the epidemic $T_E$ depends on $\rho$. It is no longer possible to use the final size of the epidemic to characterize the yield in this case and no analytic expression for the yield can be derived. To proceed, we ran numerical simulations of the model for randomly generated parameter values (see the Appendix A5 for the technical details). The results (Fig. 5.7) are in good agreement with the idealized “long-time” yield model (cf Fig. 5.6). As expected, the benefits of full control when an outbreak is uncertain correlate with lower values of $\eta$ and higher values of $\sigma$. The average value of $\eta$ in the simulations where the uncertainty leads to greater benefits of control was 0.48 while in those where the benefits of control were greater without uncertainty it was 0.63. For $\sigma$ the values were 4.47 and 1.65 respectively.

5.5 Discussion

In this chapter, we considered a specific type of control in a single population, namely the method of plant disease control where some of the hosts are replaced with a lower-yielding disease resistant variety. This study was motivated by the observation that in a simple SIR model, culling the host to prevent the outbreak (that is, decreasing the basic reproductive number down to 1) always leads to smaller losses than doing nothing. This can be considered as a special case of the above mentioned resistance-yield trade-off.
Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective

Fig. 5.7 $R_0 = 3$ and $\lambda = 2$. 2000 simulations with random parameter values out of which 1063 lead to the benefit of control being greater with the uncertainty present. The circles show the simulation points that lie in the appropriate interval for $\eta$. The colour shows the magnitude of the effect, that is how much greater is the benefit of the control with the uncertainty present. We can see that in agreement with the Figure 5.6, the effect is the strongest in the upper left corner of the $\xi$-$f$ plane and for low values of $\eta$.

where the resistance is perfect but implies no yield and is briefly discussed at the beginning of this chapter.

We analysed the impacts of a resistance-yield trade-off on disease dynamics and crop yield at landscape scales when attempting to control a disease by deploying a resistant crop variety in two different situations. In both cases, we assume that the resistant variety carries a yield penalty, i.e. it yields less than the standard variety in the absence of disease. First, we assumed that an epidemic is inevitable in order to analyse how the deployment of the resistant variety impacts the disease dynamics and how these in turn affect the final yield averaged over multiple fields in the landscape. Subsequently, we considered the situation where only the probability of an epidemic is known at the time of planting. We have not considered detailed models for demographic or environmental stochasticity (Keeling and Rohani, 2008). We merely allowed for uncertainty as to whether an outbreak will occur, coupled with deterministic dynamics. We selected this approach for two reasons. Firstly, it concentrates on the main component of variability that we wanted to investigate, that is whether or not an epidemic occurs. Secondly, it allows us to carry over most of the methodology from the analysis of the problem without uncertainty and it simplifies the numerical analysis when a different yield model with a finite time horizon is considered. We used a simple approach of integrating over the susceptible hosts or host tissue as appropriate (Madden et al., 2007;
Hall et al., 2007; Cunniffe et al., 2015a), which assumes that healthy tissue contributes to yield. Our intention throughout is to introduce a generic modelling framework. The framework is motivated by fast epidemics on agricultural crops in which the epidemic naturally burns itself out or by removal of infected hosts by roguing as may occur, for example, in cassava crops. The framework can, in principle, be extended to perennial crops. To simplify the mathematics and to minimize the number of free parameters, we initially assume that the yield is accumulated over a time period which is much longer than the time-scale of the epidemic such as in the case of potato late blight and some rust diseases of small grain cereals. This allows us to approximate the yield by the amount of susceptible host that survives the epidemic.

When an outbreak is certain, we showed that the susceptibility, infectivity and the removal factors $\eta$, $\nu$ and $\sigma$ introduced in the model can be combined into a single parameter $\xi = 1 - \eta \sigma \nu$ which we call the efficacy of the resistance. This aggregate parameter is a convenient means of integrating the components that characterise the differences between resistant and susceptible varieties. The approach is analogous to, yet different from, the approach advocated by Parlevliet (1979) directed at quantifying the components of resistance when comparing different varieties. Thus, Parlevliet (1979) first showed how to quantify epidemiological components of resistance such as infection frequency, latent period and spore production per unit time as well as the infectious period. From these analyses, Parlevliet (1979) was able to quantify and ascribe the components of resistance that accounted for differences between susceptible and resistant hosts. Savary et al. (1988) subsequently showed for peanut rust how to combine individual components into a product as a relative measure of resistance that reflected differences between epidemiological components, which is analogous to $\xi$. The difference lies in that $\xi$ is constructed so that the components relate directly to parameters that define rates in an epidemiological dynamical model. Thus $\nu$ is a measure of the reduction in the rate of transmission of infection; $\sigma$ is a measure of the change in the infectious period and $\eta$ is the measure of reduction in the transmission rate. We found the conditions that $\xi$ has to satisfy in order for deployment of the lower yielding, resistant variety to be optimal. Major gains can be achieved particularly when $\xi$ increases above $1 - 1/R_0$, which is also the condition for being able to prevent the outbreak altogether.

A further trade-off arises when uncertainty about an outbreak is added. Deployment of the resistant variety is wasteful in the absence of an epidemic. However, given
uncertainty about an epidemic outbreak, deploying a resistant variety renders the overall host population less susceptible to the pathogen and therefore decreases the probability that an outbreak will occur in the first place (Park et al., 2003). Fig. 5.3 and Fig. 5.5 provide the resolution to this trade-off.

Our results indicate that when the resistance is sufficient to prevent an epidemic altogether ($\xi$ large), the benefits of the control are greater when an outbreak is certain. However, when the resistance is not strong enough to prevent the outbreak ($\xi$ small) but offers significant reduction in susceptibility to the pathogen ($\eta$ is small), the benefits of control are overall greater when uncertainty is accounted for (cf Fig. 5.5 with Fig. 5.3). In such a case, reducing the invasion probability and thus possibly preventing the outbreak altogether outweighs the risk of wasting resources by deploying resistant cultivars. Note that this is independent of the assumption that the invasions follow a Poisson process. Rather it is a consequence of the fact that when the resistance is not very effective during the outbreak, it can still significantly reduce the probability of the pathogen invading. Biologically, this can happen when the resistant hosts have a longer infectious period (that is $\sigma > 1$). This is possible, for example, when the disease eventually kills the hosts, but takes longer to kill the resistant hosts. Alternatively, in the analogous case when the disease does not kill the hosts but rather they are removed from the population via roguing, the resistant hosts might take longer to show symptoms and therefore avoid detection.

We have made a number of important simplifying assumptions about the epidemiological model. These have allowed some insights to be gained about the trade-offs to be considered in deploying resistant varieties with lower yield potential compared with a susceptible variety. We have used a simple SIR epidemiological model (Eqns 5.7-5.11) that is parsimonious while allowing flexibility in attributing the effects of partial resistance to different epidemiological processes. Accordingly, there are just two parameters for the underlying epidemiological model ($\beta$, the transmission rate; $\mu$, a measure of the infectious period). To these we added three parameters (Table 5.1) to allow for resistance. One or more of these could be set to one and effectively eliminated as a separate parameter. We also introduced a similarly tunable parameter ($\rho$) for the ratio of resistant hosts in the population, while $\lambda$ is a measure of uncertainty about whether or not disease is likely to occur. The other parameters listed in (Table 5.1), ($R_0$) and ($\xi$), are compound parameters derived from the others.
We have also used a strongly simplifying assumption that crop yield can be assessed from the final level or the integral over time of susceptible hosts, which assumes that infected hosts do not contribute to yield. Our intention here, however, was to focus on principles relating to decision-making in relation to the deployment of resistance. For this we have preferred to keep the model for the epidemic and for yield simple. It is possible that infected hosts might well contribute to yield and this could readily be included in the model as could additional feedback loops for the effects of different levels of infection on growth dynamics of the host (Cunniffe et al., 2015a; Bailey and Gilligan, 2004; Bailey et al., 2006).

Our results have shown the importance of accounting for uncertainty to inform policy on behalf of an agricultural planner. We restricted our analyses to considering the probability of an outbreak occurring. Future analyses could address the robustness of the conclusions to short-term fluctuations associated with demographic and environmental stochasticity. Of more likely importance, however, are longer-term fluctuations that result in periodic epidemics, arising from repeated introductions of the pathogen from outside the system, and periods of long-term environmental suitability. The intricacy of this approach lies in the fact that while the environmental suitability forces outbreaks with a certain period, introducing the resistant variety would change the intrinsic time-scale of the epidemic (Lo Iacono et al., 2013). Combining these two effects could produce complex dynamics (Keeling and Rohani, 2008; Lo Iacono et al., 2013). Finally, it would be interesting to include the effects of heterogeneity in grower behaviour or a more formal treatment of risk-aversion (Gent et al., 2011; Gilligan et al., 2007). In this paper, we model the yield in the presence of uncertainty as the average of the yield when the outbreak epidemic does or does not occur. However, farmers and policy makers tend to prefer control strategies that minimize the probability of losses rather than those that maximize the probability of gains (McRoberts et al., 2011; Cunniffe et al., 2015b). Our model could be combined with game theoretic approaches to analyse these outcomes.

5.5.1 Author contributions

This chapter is based on published work (Vyska et al., 2016). M.V. conceived and designed the study. M.V. and N.J.C. implemented the study. M.V. and C.A.G. and N.J.C. contributed to the discussion of the work and writing of the paper.
5.6 Appendix

5.6.1 A1. Sketch explaining the yield approximation

![Diagram](image)

Fig. 5.8 The yield is modelled as the area under the curve $S(t)$. In the figure, the vertical red line marks the end of the epidemic. When the yield is accumulated over a time period much longer than the duration of the epidemic, we approximate it by the area of the green rectangle, that is by the length of the time period multiplied by the surviving population, $S(t = \infty)$.

5.6.2 A2. Calculation of the probability of a large scale outbreak

Let $q$ be the probability of no outbreak after the first infection has occurred and furthermore, let $q_R$ be this probability given the first infection was of a resistant host, and $q_S$ the same for standard host. Then by the standard argument (Diekmann et al., 2012), assuming large populations sizes, we have

\begin{align*}
q &= psq_S + p_Rq_R \quad \tag{5.23} \\
q_S &= \frac{\mu}{\mu + \beta(1 - \rho + \eta \rho)} + \frac{\beta(1 - \rho + \eta \rho)}{\mu + \beta(1 - \rho + \eta \rho)} q_S \quad \tag{5.24} \\
q_R &= \frac{\mu}{\mu + \beta \nu \sigma(1 - \rho + \eta \rho)} + \frac{\beta \nu \sigma(1 - \rho + \eta \rho)}{\mu + \beta \nu \sigma(1 - \rho + \eta \rho)} q_R \quad \tag{5.25}
\end{align*}
where \( p_S \) and \( p_R \) are the probabilities of a newly infected host being of the standard variety or the resistant variety and are given by

\[
\begin{align*}
p_S &= \frac{1 - \rho}{1 - \rho + \eta \rho} \quad (5.26) \\
p_R &= \frac{\eta \rho}{1 - \rho + \eta \rho}. \quad (5.27)
\end{align*}
\]

These are three equations for three unknowns \( q, q_S \) and \( q_R \) and solving them leads to a quadratic equation for \( q \). We have \( P_{\text{takeoff}} = 1 - q \) and this gives

\[
P_{\text{takeoff}} = \frac{-1}{\sigma \nu} + R_0 F + \sqrt{\left(\frac{1}{\sigma \nu} - 1\right)^2 + R_0^2 F^2 + 2R_0 \left(\frac{1}{\sigma \nu} - 1\right)((F - 2\eta \rho))}{2R_0 F} \quad (5.28)
\]

where, for simplicity, we have written \( F = 1 - \rho + \eta \rho \).

### 5.6.3 A3. Calculation of the yield function from the final size equations

It can be readily checked that the system (5.7-5.11) has two conserved quantities \( X_1 \) and \( X_2 \) given by

\[
X_i = -\ln S_i + R_0 \sum_j A_{ij} (I_j + S_j), \quad (5.29)
\]

where \( A \) is a \( 2 \times 2 \) matrix

\[
A = \begin{pmatrix}
(1 - \rho) & \nu \rho \sigma \\
\eta (1 - \rho) & \nu \eta \rho \sigma
\end{pmatrix}. \quad (5.30)
\]

This gives the final size equations (Diekmann et al., 2012)

\[
\begin{align*}
S_S^\infty &= e^{-R_0 (1 - \rho)(1 - S_S^\infty) - R_0 \alpha \nu \rho (1 - S_R^\infty)}, \quad (5.31) \\
S_R^\infty &= (S_S^\infty)^\eta. \quad (5.32)
\end{align*}
\]

Unsurprisingly, \( S_R^\infty > S_S^\infty \). These can be used to eliminate \( S_R^\infty \); however, the resulting equation for \( S_S^\infty \) cannot be solved analytically. It is possible to make progress however. First note that the map between \( S_S^\infty \) and \( \rho \) is one-to-one as \( S_S^\infty(\rho) \) is strictly increasing. It is useful to introduce some additional notation. When \( \rho = 0 \), the final size \( S_S^\infty \) will...
be denoted by $s_0$, where

$$s_0 = e^{-R_0 (1 - s_0)}.$$  \hfill (5.33)

When $\rho = 1$, the final size $S_S^\infty$ will be denoted by $s_1$, where

$$s_1 = e^{-R_0 \nu \sigma (1 - s_1^\infty)}.$$  \hfill (5.34)

Note that it is possible that $s_1 < s_0$, i.e. the proportion of the susceptible hosts in the standard population that survive the epidemic is lower with full control. This is not discussed in the main text because it has no impact on the general behaviour outlined in Figure 5.1. There are two cases to be considered. If $R_0 \nu \sigma \leq 1$, the outbreak is prevented for $\rho \geq \rho_c$ where

$$\rho_c = \frac{R_0 - 1}{R_0 (1 - \nu \sigma \eta)}.$$  \hfill (5.35)

In this case the allowed range of values of the final size is $s_0 \leq S_S^\infty \leq 1$, corresponding to $0 \leq \rho \leq \rho_c$. On the other hand, if $R_0 \nu \sigma > 1$, the outbreak cannot be prevented. The allowed range of values of the final size is then $s_0 \leq S_S^\infty \leq s_1$, corresponding to $0 \leq \rho \leq 1$. On these intervals the function $S_S^\infty(\rho)$ can be inverted to give

$$\rho(S_S^\infty) = \frac{\ln(S_S^\infty) + R_0 (1 - S_S^\infty)}{R_0 [1 - S_S^\infty - \nu \sigma (1 - (S_S^\infty)^\eta)]}.$$  \hfill (5.36)

and therefore the yield can be written as a function of $S_S^\infty$,

$$Y(S_S^\infty; R_0, f, \nu, \eta) = S_S^\infty - (S_S^\infty - f(S_S^\infty)^\eta) \frac{\ln(S_S^\infty) + R_0 (1 - S_S^\infty)}{R_0 [1 - S_S^\infty - \nu \sigma (1 - (S_S^\infty)^\eta)]}.$$  \hfill (5.37)

### 5.6.4 A4. Extreme control results

By scanning through possible values of the parameters, we have determined that the benefits of the intermediate control are not significantly higher than those of the extreme control. This is intuitive, since in Figure 5.2 we can see that the area where the intermediate control is optimal is small and so we would not expect the profit function to raise and drop too rapidly inside that area. Intuitively from Figure 5.2, the benefits of the intermediate control will approximately be the largest when $Y_N = Y_F$, where $Y_N$ is the yield when we do not control at all and $Y_F$ is the yield under full control. 


control. We have

\[ Y_N = s_0 \]
\[ Y_F = \begin{cases} 1 - \rho_{\text{full}} + f \rho_{\text{full}} & \text{if } R_0(1 - \xi) < 1 \\ f s_1^{\eta} & \text{if } R_0(1 - \xi) > 1 \end{cases} \]

(5.38) \hspace{1cm} (5.39)

where \( s_0 \) and \( s_1 \) are as above and \( \rho_{\text{full}} = \min(1, \rho_c) \). Note that raising equation (5.34) to the power \( \eta \) reveals that \( s_1^{\eta} \) is a function of \( \xi \) and \( R_0 \) only. Therefore \( Y_N = Y_F \) is represented by a curve in the \((\xi, f)\) plane. Numerical analysis reveals that along this curve, the benefit of intermediate control is the greatest when \( \eta = 1 \). This is visually intuitive from Figure 5.2. Also, comparing the function \( Y_F \) and \( Y_N \) immediately gives the conditions (5.20) and (5.21), they both follow from rearranging \( Y_F > Y_N \).

### 5.6.5 A5. Numerical simulation: technical details

We ran 2000 simulations where the parameters \( \nu, \eta, \xi \) and \( f \) were selected uniformly at random from the interval \((0.1, 1)\) and then discarded if they gave unreasonably high values of \( \sigma \) (above 20). To calculate the yield, the simulations were run with 10000 hosts, starting with 1 infected host (standard with probability \( 1 - \rho \) and resistant with probability \( \rho \)). Each simulation ended if either the outbreak has ended (the number of infected hosts dropped below 1) or if the duration of the epidemic reached the cut-off value of 1000 infectious periods. The cut-off value was introduced for the reasons of computational feasibility and since the duration of the outbreaks without any control was about 15 infectious periods, the cut-off does not have a significant effect on the analysis.
Chapter 6

Discussion

6.1 Summary of the results

In this thesis, we have studied the implications of a constraint on the amount of available resource for disease control on the dynamics of epidemics and on the allocation of limited resource between multiple separate populations. Our aim has been to build on existing models and improve intuition and understanding. For this reason, we have selected the problems studied in this thesis to expand on previous work, to be mathematically interesting and at least partially analytically tractable.

In Chapter 2, we consider a simple, non-spatial compartmental model with constrained control. Here by constrained control we mean control that can only be applied to a certain number of infected individuals at any given time, due to limited availability of resource for control. Such control constraint has been studied before, in the context of optimal control theory to optimize allocation of the control resource between two separate populations (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011). The authors however did not consider the detailed dynamics of the model with constrained control. We showed that an understanding of the dynamical properties of the model is crucial for extending the analyses of Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) to more general scenarios.

We showed that in both the SIS and the SIRS case, the system can have more than one endemic equilibrium and exhibits hysteresis, a phenomenon well known from more complicated models in both ecology (Beisner et al., 2003) and epidemiology (Hedeler, 1997; Gross et al., 2006) where the state of the system depends on its history. We
also demonstrated that which endemic equilibrium is reached depends on the initial conditions, a property important for control allocation. In the SIRS case, the system also exhibits global bifurcations of limit cycles, and the long term behaviour can therefore be oscillatory. We have shown how this impacts the behaviour of the stochastic model, by making transitions between different stable attractors more likely. Our work in Chapter 2 shows how even a simple model with constrained control can exhibit mathematically very rich dynamics that, in turn, can influence optimal control. It also forms the basis for the study of optimal allocation of control resource in the following chapters.

In Chapters 3 and 4, we use the understanding from Chapter 2 to investigate allocation of limited resource between several isolated populations of different sizes. In both chapters, our goal was to find a resource allocation strategy that would either eradicate the disease, or at least achieve the smallest possible endemic prevalence given a fixed amount of resource insufficient for treating all the infected individuals. We assumed that the transmission of infection happens only within these populations and not between, that is they are uncoupled. This simplifying assumption allowed us to avoid complex “black-box” simulations and instead we were able to devise heuristics and approximations characterizing the optimal allocation strategies. It also allowed us to focus on the indirect interaction between the populations arising from the fact that they share a common resource pool. This means that when more resources are allocated to one population, there are less remaining for the others and the levels of infection across the populations are therefore anti-correlated.

In Chapter 3, we considered a one-time resource allocation at the beginning, with no possibility of reallocating the resources later. This means that the allocation must take into account the long term dynamics of the epidemic. Following from Chapter 2, we defined a concept of saturating a population. This means that the amount of resource allocated to the population is such that allocating resources further no longer has any effect on the objective of minimizing the endemic disease prevalence. We demonstrated that the optimal allocation must saturate some populations and not allocate any resource to others. We further showed that it is possible to define a value of saturating a population such that a good approximation to the optimal allocation can be found by solving a knapsack problem in order to maximize this value given the available resource. This provides a simple and intuitive strategy of allocating the resource between the populations.
In Chapter 4, we relaxed the assumption of a one-time only resource allocation. We built on the work by Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) who assumed that the resource can be reallocated continuously in time and studied the problem of how it should be allocated between two populations of equal sizes. We have extended their results to $n$ populations of varying sizes and showed that when the populations are of unequal sizes, the previously considered approaches (Rowthorn et al., 2009; Ndeffo Mbah and Gilligan, 2011) might fail in eradicating the disease even when the amount of resource is sufficient for eradication, given correct allocation. Due to the complexities of considering $n$ populations we made the simplifying assumption that the populations are uncoupled. We showed how the understanding developed in Chapter 2 can be used to define a concept of urgency of a population, and we designed a new allocation strategy which outperforms the previously considered strategies.

Our work demonstrates that consideration of the detailed dynamics of the epidemiological model in question is necessary to efficiently allocate limited control resource between separate populations.

In Chapter 5, we shifted our attention to plant disease and investigated how to deploy a partially resistant variety efficiently assuming that it yields less than the susceptible one in the absence of disease. This gives rise to a trade-off between the yield lost due to the disease and the yield lost due to the deployment of the resistant variety. We built a model allowing for both reduction in the susceptibility and infectiousness of the resistant hosts and constructed policy plots showing how much of the resistant variety should be deployed in order to maximize the yield over a long time horizon. We also considered the possibility that the decision of whether or not to deploy the resistant variety has to be made before it is known whether an outbreak will occur or not. We assumed that the invasion of the pathogen is a stochastic event and we calculated how the probability of a major outbreak depends on the proportion of resistant hosts. We found the counter-intuitive result that it is possible for the deployment of the resistant variety to be optimal when there is uncertainty, even when it would not be optimal to deploy the resistant variety during the outbreak. This is a direct consequence of the reduction in the probability of an outbreak provided by deploying the resistant variety.
6.1.1 Main contributions

In this thesis we studied efficient allocation of resource for control of epidemics. We focused on obtaining intuitive insight into the problem, and for this reason, rather than using large scale simulations, we used analytical methods whenever possible. We achieved this by making a number of simplifying assumptions throughout with the aim to construct the simplest possible model exhibiting the behaviour of interest.

In Chapter 2 we showed that the addition of control with an economic constraint on the control resources to a simple SIS and SIRS model leads to mathematically very complex dynamics. With the help of bifurcation diagrams, we provided insight into the behaviour of the model and we argue that this insight is necessary for constructing an efficient control allocation scheme.

We continued the analysis in Chapters 3 and 4, where we extended the model to multiple uncoupled populations and considered the problem of allocating limited control resource between them. Initially we assumed a one-time fixed allocation of resources (Chapter 3) and later we allowed for the possibility of continually reallocating the resources (Chapter 4). In both cases, we found an elegant and intuitive heuristic based on the understanding of the dynamics developed in Chapter 2.

In Chapter 5 we switched focus back to a single population and we considered a specific type of control, namely the deployment of a resistant crop variety given the assumption that the resistance leads to a fitness trade-off and therefore reduced yield. We investigated the interplay of the effects the resistant variety has on the probability of an outbreak and on the dynamics of the epidemic during an outbreak. We found that sometimes it is optimal to deploy the resistant variety when we are not certain whether an outbreak will occur, even when it would not be optimal to deploy the resistant variety during an outbreak.

6.2 Directions for future work

In this concluding section, we would like to briefly touch upon some ways in which our work can be carried forward. In Chapter 2, and then in Chapters 3 and 4 which were based on the analysis in Chapter 2, we assumed that the effect of the treatment is to increase the recovery rate of the infected individuals. We briefly touch upon the topic of modelling the effect of the treatment by reducing the infectiousness of
the treated individuals, as it was done, for example, in Brandeau et al. (2003), but we do not analyse it in detail. For even greater realism, the two effects could be combined together. As mentioned in the appendix A3 in Chapter 2, while we expect the qualitative behaviour to be similar if not the same, it would be interesting to see if the model with different effects of treatment also leads to global bifurcations and stable limit cycles in the SIRS model.

In the metapopulation model in Chapters 3 and 4, the most obvious extension would be to add coupling between the subpopulations. The big differences between the performance of the strategies which allocate resources to the least infected populations and to the most susceptible populations are present with coupling. However our urgency-based strategy is based on the assumption that we can estimate in which populations the disease can be eradicated now, but not necessarily later. This would be significantly more complicated to determine in a coupled metapopulation. A first step in this direction could be to add initially very weak but increasingly stronger coupling and examine the performance of both the knapsack approximation and the urgency-based strategy.

Perhaps a more subtle, but more interesting generalization would be to include an asymptomatic class, that is individuals who are infectious but cannot be treated because they haven’t been detected. This is because both in the work by Rowthorn et al. (2009) and Ndeffo Mbah and Gilligan (2011) and in our work here, once the disease is eradicated or almost eradicated in one of the populations, keeping it that way requires arbitrarily small amount of resource. This is in part why treating the least infected population works so well, it eradicates the disease in the populations one by one, and it costs nothing to keep it that way, despite possible import of infection from the other populations. In practice, we would expect that it would cost some minimum amount to detect new infection cases in the populations where the disease has been eradicated. This could also be modelled by considering the cost of detection of new cases directly, which could perhaps scale with the size of the population.
References


References


Cinti, S. K., Barnosky, A. R., Gay, S. E., Goold, S. D., Lozon, M. M., Kim, K.,


Graunt, J. (1662). Natural and political observations made upon the bills of mortality.


ml-o powdery mildew resistance genes and agronomic traits in barley. *Euphytica*,

incentives and mathematical models of disease. *Environment and Development

and economic optima for control of immunizing infections. *PNAS*, 108(34):14366–
14370.


epidemiological models with non-linear transmission. *Bulletin of Mathematical
Biology*, 68(3):615–626.

Science and Business Media.


Lessler, J., Reich, N. G., Brookmeyer, R., Perl, T. M., Nelson, K. E., and Cummings, D.


Lo Iacono, G., van den Bosch, F., and Gilligan, C. A. (2013). Durable resistance to
crop pathogens: an epidemiological framework to predict risk under uncertainty.
*PLOS Computational Biology*, 9(1).

Madden, L. V. (2006). Botanical epidemiology: some key advances and its continuing

Madden, L. V., Hughes, G., and van den Bosch, F. (2007). *The Study of Plant Disease
Epidemics*. The American Phytopathological Society.


Implementations*. John Wiley & Sons.


Snow, J. (1853). Continuous molecular changes, more particularly in their relation to epidemic diseases.


