A MATHEMATICAL FRAMEWORK FOR DESIGNING AND EVALUATING CONTROL STRATEGIES FOR WATER- & FOOD-BORNE PATHOGENS: A NOROVIRUS CASE STUDY

PAUL MCMENEMY

Doctor of Philosophy

Division of Computing Science and Mathematics

University of Stirling

February 2017

DECLARATION

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text and bibliography.

I also declare that this dissertation (or any significant part of my dissertation) is not substantially the same as any that I have submitted, or that is being concurrently submitted, for a degree or diploma or other qualification at the University of Stirling or similar institution.

I was admitted as a research student in September 2012 and a candidate for the degree of Doctor of Philosophy in June 2013. This dissertation is a record of the work carried out at the University of Stirling between 2012 and 2017, under the supervision of Professor Adam Kleczkowski, Professor Frans de Vries, Dr Nicholas G.H. Taylor and Dr Neil Auchterlonie.

I have read, and adhered to, the University's plaigarism policy, as detailed at: `http://www.plagiarism.stir.ac.uk/`

Stirling, February 2017

Paul McMenemy

ABSTRACT

Norovirus (NoV) is a significant cause of gastroenteritis globally, and the consumption of oysters is frequently linked to outbreaks. Depuration is the principle means employed to reduce levels of potentially harmful agents or toxins in shellfish. The aim of this thesis was to construct mathematical models which can describe the depuration dynamics of water-borne pathogens and specifically examine the dynamics of NoV during depuration for a population shellfish. Legislation is currently under consideration within the EU by the Directorate-General for Health and Consumers (DG SANCO) to limit the maximum level of NoV that consumers are exposed to via this route. Therefore it was important to the utility of the thesis that any models constructed should incorporate control measures which could be used to implement maximum NoV levels. Doing so allowed calculation of minimum depuration times that would be required to adhere to the control measures incorporated into the models.

In addition to modelling the impact on pathogens during the depuration, we wished to gain some insight into how the variability, and not just the mean levels, of water-borne pathogens can be as important with respect to the length of depuration required to minimise any food safety risks to the consumer. This proved difficult in the absence of any data sets that can be used to calculate variability measures, as little data is currently available to inform these values for NoV. However, our modelling techniques were able to calculate an upper limit on the variability of water-borne pathogens that can

be well approximated by lognormal distributions.

Finally we constructed a model which provided linkage between the depuration process and the accretion of pathogens by shellfish while still within farming waters. This model proposed that the pulses of untreated waste waters released by sewage treatment works due to high levels of rainfall would be transmitted into shellfish whilst filter-feeding. We carried out analysis of economic trade-offs with regards to the costs incurred by water treatment companies, shellfish farmers and the wider human community who could be impacted by an outbreak of NoV-induced gastroenteritis.

This research was supported by the University of Stirling via a PhD Impact Collaborative Studentship (Agreement Number DP227R), and the Centre for Environment, Fisheries and Aquaculture Science (CEFAS).

ACKNOWLEDGMENTS

I have so many people that I wish to thank, let alone acknowledge, for putting up with me during my PhD project. Firstly I wish to thank my primary supervisor, Professor Adam Kleczkowski of the University of Stirling, for his continual support during the highs and lows of the last four years. Adam not only provided his expert and extensive knowledge on the mathematical modelling of biological systems, but was also a constant spring of encouragement and motivation to me. There are not many certainties in life, but I am certain that without Adam as my supervisor I would have not have learned as much during this process, nor developed as person.

Secondly, I wish to thank my other supervisors for their academic support, friendship and encouragement I have received from them — Dr Nick Taylor of CEFAS in Weymouth and Professor Frans de Vries of Stirling. There are also many people within the Division of Computing Science & Mathematics at Stirling that I wish to thank: my office mate and friend and now Dr Adrian Worton, and fellow PhD students Brian Lee, Iona Paterson, Ken Reid and Jason Adair for their friendship, moral support and not even close to frequent enough supply of cupcakes and coffee. I also wish to thank my fellow PhD students and good friends Dr Vincent Marmara and Mrs Danika Marmara for their support, friendship and hospitality over the last few years. Without IT support we would all be doomed, therefore the patience and technical help given by Sam Nelson and Graham Cochrane has been vital in keeping the wheels of academia turning over the last four years. My heartfelt thanks are also given to the other staff members at Stirling who have made my time

there so memorable: Grace McArthur, Linda Bradley, Lynne Reilly, Gemma Gardiner, Rachel Norman, Donald Smith, Andy Hoyle, Penelope Jackson, Kate Howie, Anthony O'Hare, Jess Enright, Joszef Farkas... the list could go on for some time.

I would also like to thank the team at CEFAS in Weymouth for their input and expert guidance. David Lees, James Lowther, Carlos Campos, Anna Neish, Rachel Hartnell and Ioanna Katsiadaki were always accommodating towards myself when I visited with them at the Weymouth labs.

Finally, and most importantly, I wish to thank my family and friends for their support and encouragement over the time I have spent at the University of Stirling. They have been a constant and safe harbour of encouragement and strength during my studies. My parents Frank and Grace have been incredibly kind and generous with their love, support and presence, as they have always been throughout my life. My brothers, sisters, cousins, nieces, nephews, aunts and uncles may have seen much less of me over the last few years (and they may be happy for that to have occurred!), but that will hopefully change in the future. Tough luck you guys...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

cefas Centre for Environment, Fisheries and Aquaculture Science — UK government agency responsible for regulation and monitoring of marine and freshwater environments

nov Norovirus — a single-stranded RNA virus comprising the Norwalk or Norwalk-like virus, which can cause acute gastroenteritis in humans

stw Sewage Treatment Works — a processing station whose purpose is to remove all contaminants within the public water system

uk United Kingdom of Great Britain and Northern Ireland — a north-western European island nation

rrv real, randomly distributed variable

pdf A probability density function

cdf Cumulative distribution function of a distribution

pcr Polymerase chain reaction — enzymatic procedure which allows *in vitro* amplification of DNA

cpg copies per gram — quantification of pathogen loads in biological samples

mdt minimum depuration time — minimum depuration time required for a population of shellfish to conform to the control parameters Ψ and ϕ

eu European Union — an economic and political union between 28 European countries

wcv Worst case variability — an estimate of the measure of the spread of water borne pathogens based upon the depuration model which returns a maximum minimum (maximin) depuration time

1

AN OVERVIEW OF NOROVIRUS WITHIN THE SHELLFISH INDUSTRY

1.1 INTRODUCTION

"Never eat oysters unless there is an 'R' in the month..."

This quote is from an article in a book entitled "Strange Stories, Amazing Facts" published by Reader's Digest in 1975 [8]. The book was one of those that every house seemed to have in the late 1970's and early 1980's: your parents bought in the hope that it would be of interest to someone in the family. Our copy was of huge interest to me when I was young, and I read it from cover to cover many times. It contained articles on a wide range of subjects, from science to tales of the supernatural, from history to myths and legends.

The quote shown above is from a section entitled "Hard to Swallow — Beliefs about food that are completely unfounded", and goes on the say that before the widespread use of refrigeration, it may have been risky to consume oysters during the hotter summer months, which coincidently have no 'R' in their names. This phrase has stayed with me since I first read it, probably more for the fact that it had pointed out that the last four and first four months of the year each contained an 'R' in their names than anything to do with the consumption of oysters. The article does go on to say that there was no issue with eating oysters whatever the month is, at least this was the thinking at Reader's Digest in 1975. Evidence in the current literature shows that oysters

Viruses	Bacteria	Parasites
Norwalk-like viruses	*Vibrio parahaemolyticus*	Nematodes
Human calcioviruses	*Vibrio cholerae*	Trematodes
Sapporo viruses	*Escherichia coli*	Liver/Lung/Intestinal flukes
Hepatitis A,E virus	*Campylobacter jejuni*	Protozoa
	Campylobacter coli	Cestodes

Table 1.1: Table of infectious causes of seafood-related illness

may still contain higher pathogen levels during the summer months (without an 'R' in their name) [2, 9], but still can present some food safety issues to the consumer and the industry.

Food safety issues within the shellfish industry provide the impetus for the work in this thesis, with particular focus placed on the issues presented by the consumption of oysters, which are more often than not consumed raw. As the cooking process can render many pathogens in-viable that may be present in raw oysters, then these could be readily transferred to the human population once consumed. Many articles in the current literature identify consumption of oysters as being the most prevalent cause of food poisoning associated with bivalve shellfish, with many instances attributed to raw osyter consumption [10, 11, 12].

As shellfish are bivalve filter feeders, they can bioconcentrate any pathogens that are present in their immediate waters [13, 14, 15, 16, 17]. This exposes shellfish to a wide ranging list of potential pathogens from viruses, bacteria, biotoxins and parasites shown in Table 1.1 [18].

1.2 A BRIEF HISTORY OF NOROVIRUS

One pathogen of particular concern that is associated with oyster consumption is norovirus NoV, and is one of the dominant causes of global food-borne illness [18, 19, 20]. In 2011 in the United States alone, an estimated 58% of 9.4 million cases of food-borne illness were attributed to NoV [20]. A global increased prevalence of NoV has been reported [21], with children under 5 years old in developing countries deemed to be particularly vulnerable to the effects of acute gastroenteritis [19].

The virus was first identified in 1972 after an outbreak of gastroenteritis in an elementary school in Norwalk, Ohio and so became to called Norwalk virus. Its later genotype variations that were identified were classed as Norwalk-like viruses, and eventually the contraction to norovirus NoV became the most common label for these virus types [22].

The UK bivalve shellfish industry was worth an estimated £33 million in 2012 at farm gate prices, employing 750 people. This excludes any fleet catches of offshore shellfish species of crabs, scallops and nephrops, and refers to aquacultural farm bivalves only. Bivalve shellfish species that are predominantly farmed in the UK are mussels, oysters and clams (see Table 1.2).

Shellfish	tonnes	price per tonne (£) — UK average	value (£)
Mussels (*M. edulis*)	26,021.3	1.048.8	27,282,020
Pacific oyster (*C. gigas*)	1206.3	4,071.6	4,911,600
European oyster (*O. edulis*)	110.9	7,602.4	843,106
Northern quahog (hard clam)	8.6	3,090.2	26,576
Japanese carpet shell (Manila clam)	5.0	3,100.0	15,500
Great Atlantic scallop	7.0	14,300.0	100,100
Queen scallop	0.4	2,500.0	1,000
TOTAL	27,359.5		33,189,902

Table 1.2: 2012 Bivalve shellfish production in the UK, broken down by species. Also included are estimated farm gate prices, averaged across the constituent UK nations [3].

Shellfish farm locations are predominantly located in coastal waters, and so are leased to operators by the Crown Estate. Each farm location is given

a classification based on quantitative test results of coliforms present in the water [4, 5] and are restated here:

- **Class A** — site where local authority testing has recorded at least 70% < 230 *E. coli*/100g and 100% < 700 *E. coli*/100g. No depuration or relaying treatment is required.

- **Class B** — site where recorded tests must not exceed 4600 *E. coli* 100g in 10% of tests. Purification by either relaying, depuration or an approved cooking method is required.

- **Class C** — site where 100% of tests must not exceed 46000 *E. coli*/100g. Relaying in either class A or B sites for extended periods or approved cooking method is required.

- **Prohibited** — site where any tests recorded > 46000 *E. coli*/100g. No harvesting is permitted.

The testing frequency is, in most cases, once every month, but previous testing history can be factored into the classification. Some class B sites in England and Wales have been designated with a Long Term classification (B-LT) [23]. This status is conferred when at least 30 results are available for the previous 5 years, and the site must have had 90% or better compliance with < 4600 *E. coli*/100g over that period. This allows some sites to have a smoothed approach to their classification, as fluctuations of coliform counts can be accounted for by e.g. sewage treatment works failures or extraordinary rainfall events. In these events, further testing and investigations are carried out by Local Action Groups to ensure public health measures are maintained.

Within the UK in 2015/16, 425 distinct shellfish harvest sites were classified by government agencies and operated by shellfish enterprises. Of these sites, 154 were located in coastal waters with an A rating, 245 with a B rating and 26 sites with a C rating [24, 25, 26]. Of the class A sites, 60 had both A and

B classifications due to seasonal fluctuations in *E. coli* measurements, and so would need to depurate/relay any harvests collected during the time the site was stipulated as B classification. Additionally, there were 22 class B sites with seasonal B and C ratings, and would need to follow the correct procedure dependent upon which rating applied at any particular time.

	Classification	No. Sites	%	No. Transitory*	Adjusted %
England & Wales	A	2	0.96	0	0.96
	B	185	88.94	11	86.30
	C	21	10.10	0	12.74
	Totals	208	100.00	11	100.00
Scotland	A	147	83.05	60	66.10
	B	30	16.95	9	31.36
	C	0	0	0	2.54
	Total	177	100.00	69	100.00
Northern Ireland	A	5	12.50	0	12.50
	B	30	75.00	2	72.50
	C	5	12.50	0	15.00
	Totals	40	100.00	2	100.00
United Kingdom	A	154	36.24	60	29.18
	B	245	57.65	22	62.12
	C	26	6.12	0	8.81
	Totals	425	100.00	82	100.00

Table 1.3: Shellfish harvest site classifications - Scotland 2015/16. *Transitory sites are defined as sites which have a dual classification due to seasonal fluctuations in *E. coli* levels

The production and distribution of shellfish within the UK can be broken down into the following stages:

(A) **Harvest** — This refers to the actual cultivation of live shellfish from aquacultural fisheries, wild shellfish commercial collection, or wild collection by individuals for private consumption or 'black market' sale directly to retailers.

(B) **Depuration** — This stage is where shellfish will be cleansed by clean water, removing any bacteria and/or sand in the shellfish, as well as potentially reducing the level of other pathogens that may be present.

(C) **Wholesale** — Primary link between fisheries and retailers of shellfish e.g. Colchester Oyster Fishery Ltd, River Roach Oyster Company.

(D) **Imports** — Some importing of shellfish into the U.K. does exist, although levels have not been quantified.

(E) **Retail** — Retail incorporates supermarkets, fish markets, restaurants, fishmongers as entities that sell shellfish to the consumer.

(F) **Consumer** — The end consumer; a population comprised of individual entities that purchase and consume shellfish.

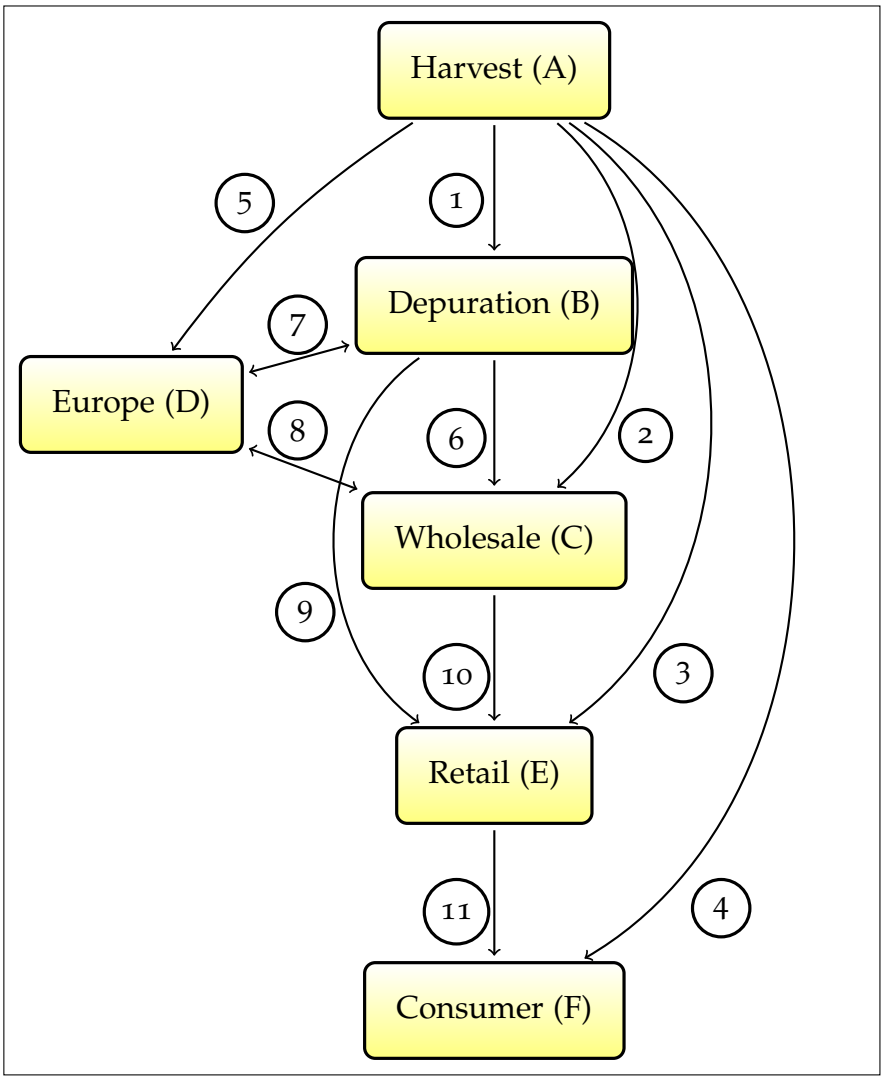


Figure 1.1: Flow diagram of UK shellfish industry's main components and their interactions

The trade connections between these stages are shown in Figure 1.1, with connections described as:

1) **Harvest** → **Depuration** Depending upon the water classification of the fishery where the shellfish are produced, shellfish may be depurated at the farm site, or removed to a specialist depuration plant off-site.

2) **Harvest** → **Wholesale** Farms with 'A' classification are not required by current legislation to depurate their harvests, therefore product may be sent directly to wholesale markets.

3) **Harvest** → **Retail** This link in the supply chain refers to both the 'black market' in shellfish production, mainly referring to private individuals who collect wild molluscs for their own sale at, for example, local fishmongers and restaurants; and Class 'A' product sold straight to retail establishments such as fishmongers, restaurants.

4) **Harvest** → **Consumer** This link refers to consumers who collect wild shellfish for their own consumption.

5) **Harvest** → **Imports** Shellfish harvests can be exported directly from source to markets outside the U.K. These levels will vary dependent on numerous factors such as home and foreign product demand; seasonality leading to variations in production numbers; and other externalities such as logistic costs.

6) **Depuration** → **Wholesale** Most of the UK farmed shellfish stock is from class B/C waters [24, 25, 26], therefore most of the UK shellfish harvest is either depurated at the harvest site, or transported live to depuration facilities.

7) **Imports** → **Depuration** Shellfish can be imported into the U.K. for depuration purposes, and then exported again or join the U.K. oyster supply chain at this point. This link is bi-directional, as oysters can be exported after depuration.

8) **Imports** → **Wholesale** Imports can join the U.K. supply chain at the wholesale stage if their provenance is from an (equivalent) class A water location in the country/region of harvest, or have already been depurated before import.

9) **Depuration** → **Retail** This link in the supply chain is mainly for direct supply to large retailers by individual fisheries. For example, supermarkets

such as Waitrose may have direct safe-sourcing agreements with fisheries and these arrangements would bypass the wholesale stage as a matter of economy.

10) **Wholesale** → **Retail** Establishments such as restaurants and fishmongers will source their supply of oysters most likely from a wholesaler. Wholesalers may still do business by open market (Billingsgate market in London being a good example of this), or sell to retailers via online business.

11) **Retail** → **Consumer** The product is consumed by a member of the general population via either purchase at a supermarket or consumption at a restaurant.

1.5 U.K. WATER QUALITY CLASSIFICATION

One pathway identified for NoV to pass into the human population is the consumption of bivalve shellfish [27, 28]. Shellfish filter-feed nutrients from their surrounding waters which, in addition to food, can concentrate contaminants and infectious agents often associated with faecal contamination into their digestive system [28, 16, 29, 30, 10]. The potential exists for transmission of such agents into the human population if the shellfish are consumed while they still contain such pathogens.

This is of special concern when shellfish are eaten raw, which is commonly the case for oysters such as the Pacific cupped oyster (*Crassostrea gigas*) and the American cupped oyster (*Crassostrea virginica*) [31]. To protect against the accumulation of pathogens in shellfish, farms should ideally be situated in waters with low pollution levels. However due to socio-geographic reasons this is not always possible, as many farms are located close to population centres

Harvest Area Classification	Classification Criteria	Treatment Required
A	80% $\leqslant$ 230 *E. coli*/100g & 100% $\leqslant$ 700 *E. coli*/100g	None
B	MPN tests carried out must not exceed 4600 *E. coli*/100g in more than 10% of test samples	Purification by either relaying or depuration, or cooking by approved method
C	MPN tests must not exceed 46000 *E. coli*/100g of test samples	Relaying for a long period, depuration or cooking by approved method
Prohibited	$>$ 46000 *E. coli*/100g of test samples	Harvesting not permitted

Table 1.4: Table of water classification criteria according to EU Regulations (EC) No. 854/2004 and No. 2015/2285. Class A waters are limited in availability in the U.K. and so in many instances further treatment of farmed shellfish is required [4, 5]. MPN — Most Probable Number

[32, 33, 34]. Recently, there have been increases in the volumes of most farmed fish species via aquaculture processes, however global mollusc production has remained relatively constant since 1990. Oyster production has exhibited a slight downward trend over the same time period [35], with NoV outbreaks linked to oyster consumption more frequently identified in recent times [36].

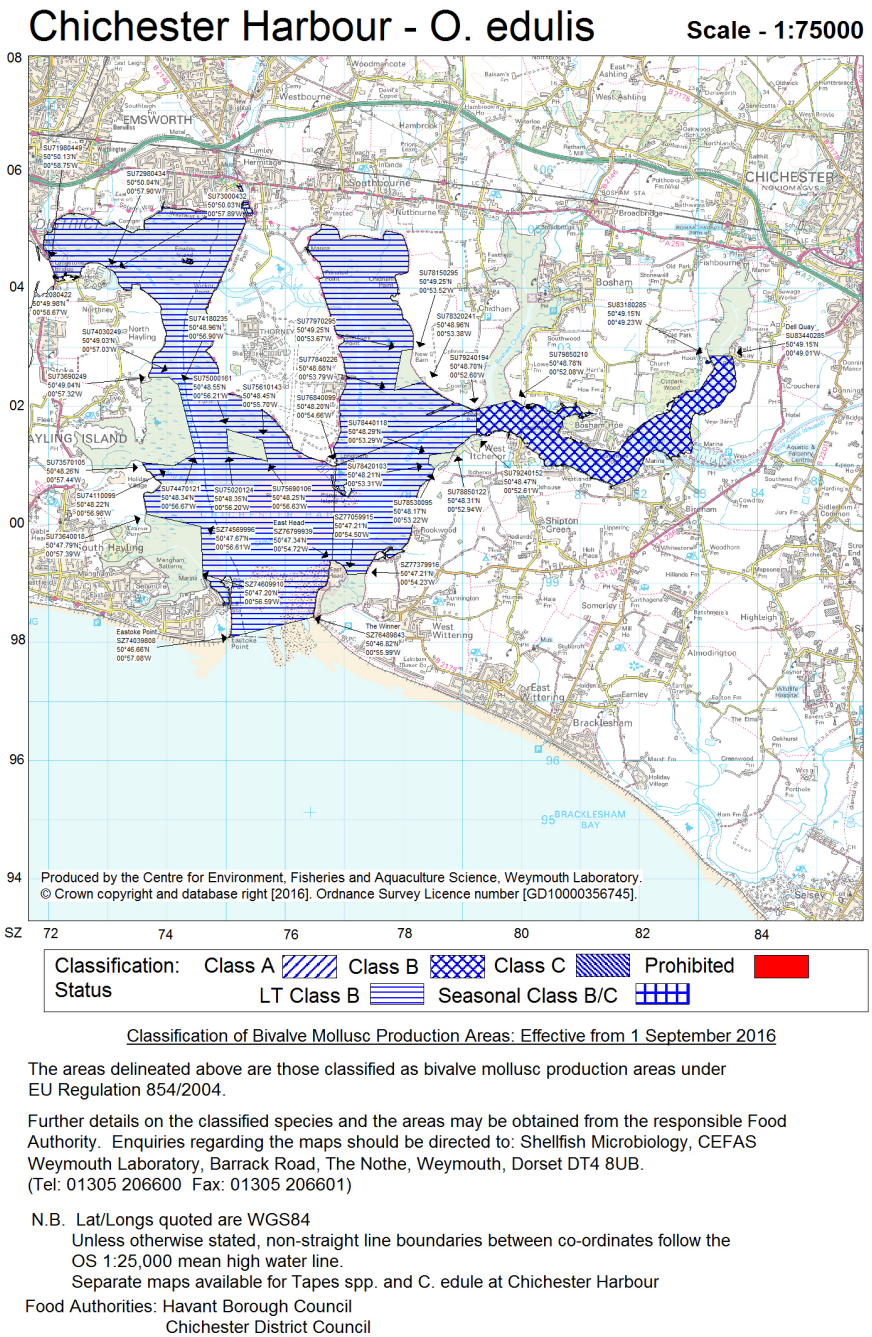


Figure 1.2: Map of Chichester harbour showing shellfish farm sites and water classifications [1]

In most industrialised countries, legislation has been put in place to minimise levels of faecal contamination found in shellfish, and thus reduce health risks to the consumer. European levels of faecal contaminants in shellfish are legislated for by EU Regulation (EC) No 854/2004, which states that "live

bivalve molluscs must come from: (a) a class A production area; (b) a relaying area; (c) a purification centre..." [4]. Shellfish harvest sites are classified as A, B, or C based on levels of the faecal indicator organism *Escherichia coli* detected in the shellfish [4]. A relaying area is a class A or B rated site where shellfish harvested from class B and C waters are respectively relocated for a time sufficient to reduce faecal contamination to acceptable levels based on *E. coli* counts. Class A waters can be limited in availability, and so in many instances an alternative to farming or relaying in these areas is required [37, 4].

The U.K. shellfish industry is dominated by two main species, mussels and the pacific oyster, which comprised over 95% of the market in 2012. More niche species comprised the remainder (Table 1.2). Geographically, the production values are reasonably equal amongst the constituent U.K. nations, with England holding 30% of the market, Wales with 27%, Scotland with 26%, and Northern Ireland with the remaining 17%. The value of the market is predominantly driven by fluctuating prices rather than production variations, with the rise from 2011 to 2012 driven by the imputed price per tonne [3].
The employment data for U.K. shellfish farms have been estimated, stating that in 2012 shellfish farming activity directly employed 705 people (Table 1.5). This excludes sites which are depuration plants only. These sites purchase shellfish at the farm gate from multiple Class B sites, and depurate in larger and more cost-effective facilities.

	No. of enterprises	Active farm sites	FT	PT	Male	Female	Total
England	68	—	166	92	250	8	258
Wales	10	—	31	3	32	2	34
Scotland	153	330	171	187	313	45	358
N. Ireland	17	—	—	—	—	—	55
U.K. Total	248	—	—	—	—	—	705

Table 1.5: U.K. Shellfish industry enterprise and employment information for 2012 [3].

This results in any retained water being discharged into the immediate waters [38]. This discharge from the oyster may increase any NoV present in the tank water, exposing the oyster batch being depurated to potential cross-contamination of NoV. However, the levels involved are not significant enough to account for detected NoV levels increasing during the initial hours of depuration carried out in experiments conducted in 2013 [39].

Consideration of the limitations of polymerase chain reaction (PCR) testing should be made, PCR being the current standard assay used to detect the presence of NoV within shellfish [15]. The main limitation is PCR can only quantify the levels of NoV within specific parts of the digestive system of oysters species. This is due to the presence of enzymes within oysters which inhibit the effectiveness of the test [40, 19, 30]. The digestive glands are the main repository for NoV within molluscs, with the rest of the pre-gland system accounting for an equal or smaller NoV load [41]. This specific accumulation of NoV in the digestive glands is attributed to the NoV's preferential molecular binding to the diverticulae of the animal's digestive system [42].

1.7 DEPURATION

From an industry standpoint, ensuring that shellfish reach the consumer pathogen-free (or at least virtually free) is important to the reputation of the producers, suppliers and the shellfish industry as a whole. From a societal position this is also of huge importance. A consumer may become ill after consumption of a shellfish which may contain pathogens, which can then lead to an epidemiological outbreak for the wider community. The consequences of an outbreak of gastroenteritis within the human population which have been linked to the consumption of shellfish (conclusively or otherwise) have been previously documented in both the literature and the general media [43, 10].

A loss of reputation to the retail establishment, supplier, and the producer always results from such incidents, and translates into a certain, short term economic cost to all the businesses involved, but can also translate into a long term loss of revenues and profits.

The implementation of correct processes also plays a vital part in maximising the effectiveness of depuration. The stacking configuration within the tank of the trays holding the animals is also important in ensuring the effectiveness of the depuration process is maximised [44, 45, 46]. At the end of a depuration cycle, a layer of excreted material from the shellfish will have been deposited on the bottom of the tank which will contain high levels of bacterial, viral and particulate contaminants that were excreted during depuration by the shellfish. It is important that this layer of detritus is not disturbed during the removal of the shellfish trays, as it would be re-suspended within the water of the depuration tank. Consequentially there would be a high risk that any animals still present in the tank would either re-ingest the contaminants, or at least

have them present on the exterior of their shells. Ensuring that the depuration tanks are thoroughly cleaned after each batch is also crucial in minimising the possibility of cross-contamination from one batch to the next. These criteria are maintained to ensure that the effectiveness of depuration is maximised, and thus the rate of contaminant and pathogen removal will be maximised.

1.7.1 *Criteria for Effective Depuration*

Depuration is a process which places harvested shellfish into tanks containing clean water, where they remain submerged for a period of time sufficient for the animals to excrete any contaminants that they may contain [47, 48, 44, 16, 31]. The efficacy of depuration relies on a number of criteria including but not restricted to:

(i) water temperature,

(ii) salinity,

(iii) oxygenation,

(iv) water flow rates,

(v) Water disinfection.

Water temperature and salinity should closely match the harvest location's conditions to minimise stress to the shellfish; the oxygen levels in the water should be sufficient for the density of shellfish held in the tank to allow normal metabolic activity to resume for all the animals within the depuration tank. In addition to its importance in terms of animal welfare, the water temperature is also an important factor in removing NoV from oysters. Recent studies by Neish (2013) and Doré (2010) indicate that increased water temperatures can

increase the rate of NoV excretion [7, 39].

The rate of pathogen removal for a population of shellfish due to depuration will clearly have a significant impact upon the actual pathogen loads within individual oysters, and thus upon the depuration time required to minimise any potential risk to consumers. Since depuration can incur significant costs to the shellfish industry [10, 44, 39], minimising any costs while at the same time minimising NoV levels in shellfish would be beneficial to both the industry and the consumer. This is due to the absence of any legislated depuration time for specifically viral contaminants (beyond the current 42 hours instigated for *E. coli*), and in tandem with the indicators within the current literature that NoV is particular food safety concern for shellfish which is consumed raw.

2

MATHEMATICAL MODELLING

2.1 INTRODUCTION

This chapter provides a short insight into the applications of mathematical modelling and how it is a relevant and appropriate tool to be used in understanding the dynamics of a water-borne pathogen, especially with regards to the depuration process. A summary of techniques and probability distributions which will form the foundations for the work carried out later in this thesis is also provided. Finally, we will discuss some of the previous modelling which has been carried out with regards to shellfish depuration and propose methods that will extend the current literature's findings.

There have been many attempts to sum up the field of applied mathematics in a single, succinct yet insightful phrase. One attempt which has resonated with me was written by G.G. Hall — “The goal of applied mathematics is to understand reality mathematically" [49]. Another by Edna Kramer is a more comparative and mathematical summation — “A mathematical model is an abstract idealization of various features of a real situation, in the same sense that pure Euclidean plane geometry is the abstract counterpart of the surveyor’s or the engineer’s conception of physical points..." [50].

Indeed, to be able to describe only a small part of reality using mathematical techniques requires a reasonable level of mathematical ability. It took the (not inconsiderable) skills of first Newton and then Einstein to derive mathematical descriptions of gravitational fields; Schrodinger, Dirac, Heisenberg and Feynman produced equations describing the nature of light. The obvious question therefore is why go to the considerable effort to produce mathematical descriptions that describe some aspect of the natural, or real, world?

Mathematical models can provide an alternative method to understand something about the real world. Often, when compared to obtaining experimental results, mathematical models can provide much more cost effective results and can sometimes provide additional understanding of the system in question beyond that which can be obtained from experimental measurements alone.

One possible example is in the field of prophylactic vaccination of children against any of a range of diseases such as measles. Constructing a math-

ematical model of the problem can give an insight into how many children should be inoculated, at what age to inoculate and spatial information on the places to concentrate your vaccination efforts. In the absence of modelling, only previous experience could be applied to consider a best approach to the problem. Only after several years would you know whether you had succeeded or not; if your efforts had failed then the results could be disastrous from a humanitarian viewpoint, as well as an ethical one.

Of course, you would only know how successful your modelling approach would have been after the same number of years. However, constructing a logical and reasonable mathematical facsimile of the real world problem would provide additional decision making tools to any legislators who would be required to make and implement programs that can have such an impact on the world at large.

2.2.1 *Systems of Differential Equations*

Producing such mathematical descriptions provides two vital insights into the system being described:

- a deeper understanding of the system;
- the ability to make predictions of the behaviour of the system.

The process of constructing a simulacrum of some facet of the real world is known as mathematical modelling.

A mathematical model will only ever be a simplification of the actual situation, no matter whether it is mechanical, biological or even financial in nature. However, any simplified model should describe the most important

aspects of the system being modelled. A well known mathematical description of a natural process is the predator-prey model, which describes the predation of one species by another. The model was first developed by Alfred Lotka and then independently by Vito Volterra to describe the fluctuations in the catch sizes of fish species in the North Atlantic. Thus, it is known as the Lotka-Volterra model, and is described by a system of differential equations:

$$\frac{dN}{dt} = N(a - bP) \tag{2.1}$$

$$\frac{dP}{dt} = P(cN - d)\,, \tag{2.2}$$

where $N(t)$ describes the population of the prey at time t, and $P(t)$ the population of the predator at time t. The parameter terms describe:

a — prey population growth rate;

b — effect of predation on the prey population;

c — prey's contribution to the growth of the predator population;

d — mortality rate of the predator

[51]. This system does not and cannot fully describe all of the environmental and biological factors that could also influence the population numbers of a particular predator-prey dynamic. However, the Lotka-Volterra model does provide a simplified description of the interactions between predator and prey that allows certain analyses of the model to be carried out. These have shown that, dependent upon the parameter values of a, b, c and d, the system can result in oscillatory behaviour of the population numbers for the predators and the prey. The system can also be used to determine equilibria points for the two population values, but again these are dependent upon the parameter values applied to the model.

As mathematical models are often constructed to describe how a particular system changes over a period of time, they can also provide predictions for the future. This is one of the most powerful benefits that can be obtained from modelling. The ability to predict future behaviours, results or outcomes using applied mathematical models is (when you stop and think about it) a fantastically powerful tool. Of course, this is dependent upon how representative the model is of the system it is based upon.

2.2.2 *Increasing the Realism, Increasing the Complexity*

The Lotka-Volterra model described by Equations 2.1–2.2 is an example of a system of differential equations, one of the primary tools of mathematical modelling. As previously stated, these equations are a simple representation of the interconnectivity between the populations of a predator-prey relationship. These can be adapted to take into account other aspects of the relationship, thus providing a more realistic portrayal of the system. The Lotka-Volterra system shown above is a basic example of its type. It can be adapted to include other aspects of a predator-prey system such as carrying capacity of the environment (K) with regards to the prey population:

$$\frac{dN}{dt} = aN(1 - \frac{N}{K}) - bNP$$
$$\frac{dP}{dt} = cNP - dP\,.$$

This model is also predicated on the unrealistic assumption that the predation level increases linearly as the number of prey increases. The model can be adapted to take into account a non-linear description of this interaction

between the predator and prey numbers by introducing a functional response that more accurately describes the number of prey caught per predator:

$$\frac{dN}{dt} = aN(1 - \frac{N}{K}) - P\ f(N,P) \tag{2.3}$$

$$\frac{dP}{dt} = cP\ f(N,P) - dP\,. \tag{2.4}$$

The form that the functional response $f(N,P)$ takes will depend upon any pre-existing data or knowledge derived from previous studies or models.

We have adapted Equations 2.1–2.2 to incorporate two additional considerations of a predator-prey system. Comparison of Equations 2.1–2.2 with 2.3–2.4 shows that it is apparent that the complexity of the model has increased. This increased complexity should provide more realistic results from an analysis of the model; however, any analysis will itself be more complex to carry out. Therefore, a balance is to be found when constructing a mathematical model of any system. The level of complexity of the model should not be such that any analysis that can be obtained is either too complex, or requires too much resource to carry out the analysis.

Any modelling work carried out should be representative of the system in question. Once construction of the model is completed, analysis and results should be obtained. The validity of the model is only confirmed when results conform to the current and future behaviours observed and recorded in the real world (see Figure 2.1) [52]. Continual refinement of a model based on new observations or insights will provide results which give greater understanding of the system as well as improved predictive capabilities.

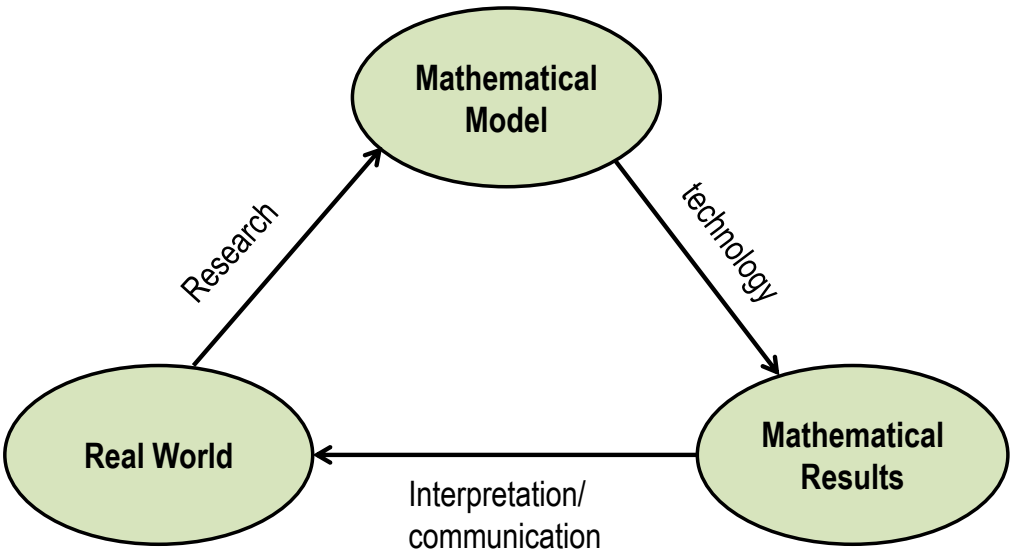


Figure 2.1: Mathematical modelling cycle

In this section we lay some foundations with regards to probability distributions that are used in the construction of the mathematical models further on in this thesis. We provide here some of the fundamental definitions of distribution types, looking at the normal and lognormal distributions in greater length than others discussed. This is carried out based on the current literature which states that water-borne pathogens are well-described by lognormal distributions, itself a close variant of the normal distribution [53, 44, 54, 55, 56].

2.3.1 *The Normal Distribution*

The normal distribution is often described as the most important probability distribution, as it and its related distribution family can describe behaviour of many interesting random variables [57, 58]. The normal distribution is known also as the Gaussian distribution due to the work carried out by Carl Friedrich Gauss in the early nineteenth century; it is also referred to as the "bell curve", due to its symmetric and convex unimodal shape.

2.3.1.1 *Probability Density Function*

The normal distribution is often described by the notation $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, where μ is the measure of location and σ^2 describes the variance of the distribution, with $X = x$ being a distribution which spans all real numbers, i.e. $x \in \mathbb{R}$. The probability density function of a normally distributed, real, random variable (RRV) is defined by

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}. \tag{2.5}$$

This describes the probability density function (PDF) of a normally distributed RRV, and provides a mathematical definition of $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

The total area under the curve of any continuous PDF must be equal to 1 and, furthermore, any probability value derived from a PDF must be in the range $[0, 1]$. These two conditions must be satisfied by a probability density function, and are restated in mathematical terms as (1): $0 \leqslant P(x) \leqslant 1$ and (2): $\int_{-\infty}^{\infty} P(x)dx = 1$.

The *standard*, normal distribution has arithmetic mean $\mu = 0$ and standard deviation $\sigma = 1$, and its PDF is often defined as $\varphi(x)$, where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-x^2}{2}\right\}. \tag{2.6}$$

This is extended for the *general*, normal distribution as

$$\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}. \tag{2.7}$$

2.3.1.2 *Cumulative Density Function*

A cumulative distribution function (CDF) describes the probability that some RRV will have a value less than or equal to some limit. In other terms, a CDF calculates the area under a specific PDF's curve up to some upper bound value:

$$P(x \leqslant a) = \int_{-\infty}^{a} P(x)dx\,, \tag{2.8}$$

where $P(x)$ is the PDF describing the distribution, and a is the limit which we require the RRV to be less than or equal to. The CDF of the standard, normal distribution (where $\mu = 0$, $\sigma = 1$) is defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-t^2}{2}\right\} dt\,. \tag{2.9}$$

Compare this with the definition of the error function which is defined as [58]

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\{-t^2\}\, dt\,. \tag{2.10}$$

We see that Equation 2.9 can be restated in terms of $\operatorname{erf}(x)$, and thus the CDF of the standard normal distribution ($\Phi(x)$) is

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right].$$

Thus the CDF of the general, normal distribution can therefore be stated as

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]. \tag{2.11}$$

Many random variables can be approximated by a normal distribution. The Central Limit Theorem states that, for independent and identically distributed random variables with a finite variability, then the sum of the of the distributions will be approximately normally distributed. Alternatively, if X_1, X_2, ..., X_n are identically distributed random variables, then $X_1 + X_2 + ... + X_n$ is approximately normally distributed, providing n is large [59].

2.3.2 *Lognormal Distribution & Its Properties*

Not all RRVs conform to being described by a normal distribution. Some biological systems will often combine the independent effects of processes which can be individually described by normality, however these processes interact multiplicatively, i.e. as products of each other, rather than by additive methods. Limpert *et al* provide an example of this, explaining "...for instance, exponential growth is combined with further symmetrical variation: with a mean concentration of, say, 10^6 bacteria, one cell division more - or less - will lead to 2×10^6 - or 5×10^5 - cells. Thus, the range will be asymmetrical; to be precise, multiplied or divided by 2 around the mean." [6]. This type of interaction will result in a distribution of cell counts across multiple cultures,

which can be described as a positively skewed distribution with a low mean in comparison to its variance, and which also has all variates greater than zero. The best known distribution of this type is the lognormal distribution.

This type of distribution describes some RRV X where $X = \ln(Y)$, with $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, and is defined as

$$P(x) = \frac{1}{\sqrt{2\pi}\ \sigma\ x} \exp\left\{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right\}. \tag{2.12}$$

The arithmetic Mean (m) and arithmetic variance (v) for log-normal distributions are defined as:

$$m = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \tag{2.13}$$

$$v = \left(\exp\left\{\sigma^2\right\} - 1\right) \exp\left\{2\mu + \sigma^2\right\} \tag{2.14}$$

Thus, the standard deviation of the lognormal distribution, which is equivalent to $\sqrt{v}$, can be stated as

$$\sqrt{v} = \sqrt{(\exp\{\sigma^2\} - 1)\exp\{2\mu + \sigma^2\}}$$

$$\Rightarrow \sqrt{v} = \sqrt{\exp\{\sigma^2\} - 1}\ \exp\left\{\mu + \frac{1}{2}\sigma^2\right\}. \tag{2.15}$$

The coefficient of variability (or coefficient of variation) is $\sqrt{v}/m$ therefore

$$cv = \left(\exp\{\sigma^2\} - 1\right)^{1/2}. \tag{2.16}$$

The properties of the normal and lognormal distributions are shown in Table 2.1 for comparison.

Property	**Normal Distribution**	**Lognormal Distribution**
Effects (Central Limit Theorem)	Additive	Multiplicative
Shape of distribution	Symmetrical	Positively skewed
Triangle shape	Isosceles	Scalene
Effects at each decision point x	$x \pm c$	$x\ ^{\times}/c$ [a]
Mean	$\bar{x}$, Arithmetic	$\bar{x}^*$, Geometric
Standard deviation	s, Additive	s^*, Multiplicative
Measure of dispersion	$cv = s/\bar{x}$ [b]	s^*
Confidence Interval:		
68.3%	$\bar{x} \pm s$	$\bar{x}^*\ ^{\times}/\ s$
95.5%	$\bar{x} \pm 2s$	$\bar{x}^*\ ^{\times}/\ s^2$
99.7%	$\bar{x} \pm 3s$	$\bar{x}^*\ ^{\times}/\ s^3$

Table 2.1: Comparison of characteristics of normal and lognormal distributions. (Table reproduced from Limpert *et al* [6, p.346]

[a] $^{\times}/$ = times/divide, corresponding to plus/minus for the established sign $\pm$

[b] cv = coefficient of variation

2.3.3 *Gamma and Binomial Distributions*

The two distributions defined next are less prominent in our subsequent modelling; however, providing definitions for them here simplifies the model construction carried out later in Chapter 6.

2.3.3.1 *Gamma Distribution*

A random variable $X = x$ has a gamma distribution if its PDF is defined as

$$P_G(x) = \frac{1}{\theta^k\,\Gamma(k)} x^{(k-1)} \exp\left\{\frac{-x}{\theta}\right\}, \tag{2.17}$$

where $x \in (0, \infty)$ [58, p. 166], and $\Gamma(k)$ is the gamma function. The gamma function is defined as

$$\Gamma(k) = \int_0^\infty z^{(k-1)} \exp\{-z\}\, dz \tag{2.18}$$

[60, p. 255].

The shape (k) and scale (θ) parameters of the Gamma PDF can be defined in terms of the arithmetic mean ($\bar{x}$) and the variance (σ^2), where

$$\bar{x} = k\theta \tag{2.19}$$

$$\sigma^2 = k\theta^2\,. \tag{2.20}$$

These allow definitions for the parameters k and θ in terms of the arithmetic mean and the variance:

$$k = \frac{\bar{x}^2}{\sigma^2} \tag{2.21}$$

$$\theta = \frac{\sigma^2}{\bar{x}}\,. \tag{2.22}$$

2.3.3.2 *Binomial Distribution*

The binomial distribution is a discrete probability distribution with parameters N, k and p. Formally, if a random variable $\mathcal{T}$ conforms to the binomial distribution, then the likelihood of obtaining k successes from N independent Bernoulli trials is described by the probability mass function[1] defined as

$$P_B(K = k) = \binom{N}{k} p^k\, q^{N-k} \qquad \text{[61].} \tag{2.23}$$

1 As this is a discrete distribution, it does not have a PDF

2.4 CURRENT MODELLING OF WATER-BORNE PATHOGENS IN DEPURATION

This section reviews the current literature that provides mathematical models of the dynamics of water-borne pathogens during the depuration process. There is a dearth of mathematical modelling with regards to the shellfish depuration process, and especially where viral contaminants are the subject of the work. However, two papers have been published in the last three years which have produced mathematical descriptions of viral contaminants within shellfish and the impact which depuration has on the virus levels present.

2.4.1 *Polo, Feal, Varela, Monteagudo and Romalde 2014 Model*

A 2014 paper in the Food Research International journal proposed a mathematical description of the “Depuration kinetics of murine norovirus in shellfish" based on experimental evidence the authors had obtained. The data pertained to the depuration induced reduction of murine norovirus (a “human norovirus surrogate") within Manila clams and Mediterranean mussels.

The authors proposed an equation which describes the viral load decay over time due to depuration as

$$c(t) = \alpha \exp\{-\beta t\} + \delta\,, \tag{2.24}$$

where c(t) describes the viral concentration at time t, relative to the initial viral concentration at $t = 0$. The parameter β describes the depuration decay rate per time period (in the case of this paper, per day is the unit of β), and δ is some residual viral load which is a constant level of contamination of the individual oyster.

The authors quantify the parameters based on experimental data, and conclude that murine norovirus is an effective analogous pathogen for human NoV. They also state that further studies are required to ascertain differences in depuration decay rates for combinations of different viruses and/or shellfish species [62].

2.4.2 *Polo, Feal and Romalde 2015 Model*

The paper "Mathematical model for viral depuration kinetics in shellfish: An useful tool to estimate the risk for the consumers" by similar authors expands upon their 2014 model. The paper constructs a mathematical model of the viral kinetics within an individual shellfish which incorporates some of the depuration criteria discussed in Section 1.7.1, most specifically including a term describing the filtration rate of the depuration process.

Again, the authors consider the total viral load within an individual animal as being split between a time-dependent compartment and a constant background viral level of pathogen that will be constant $\forall t,\ t \in \mathbb{R}^+$:

$$n(t) = n_d(t) + n_r\,, \tag{2.25}$$

where n(t) is the total viral load at time t. The constant background viral level is denoted by n_r and the viral load impacted by depuration $n_d(t)$. The authors go on to derive an expression of the total viral load at t as

$$n(t) = n_r + (n_0 - n_r)\exp\{-\lambda t\}, \tag{2.26}$$

where n_0 is the initial, total load and λ is the decay rate due to depuration. The difference of $n_0 - n_r$ describes the amount of pathogen present per animal that can be impacted by the depuration process. The rate of pathogen removal

due to depuration is defined as λ; the authors provide a breakdown of this rate as

$$\lambda = \frac{\alpha f}{v_d}, \tag{2.27}$$

where α is described as “the ratio of filtration rate" entering the digestive tissue, f is the filtration rate and v_d is the volume of the tracts of the digestive tissue [63]. Therefore the authors describe the depuration rate λ as being a consequence of only the flow rate of the depuration process being employed.

The papers summarised in Sections 2.4.1 and 2.4.2 present a logical and reasonable modelling approach to depuration dynamics. However, the models only describe how an individual shellfish's viral load is reduced during depuration. The models do not take into account the wide variability observed in viral loads present across a batch of shellfish which are being depurated at the same time.

This thesis is primarily focussed in the construction of mathematical models that describe how pathogen loads within populations of oysters are impacted by the depuration process. In Chapter 3, we build a mathematical model that incorporates both the decay of pathogen loads within an individual shellfish, and how this individual decay effects pathogen levels across a population of shellfish undergoing depuration simultaneously. We go on to incorporate pathogen control parameters into the model, parameters which would allow legislators to enforce control measures for pathogen levels across a shellfish population before they pass to either the wholesale or retail markets. This model is used to obtain estimates of the variability of pathogens across a population of shellfish, as well as provide guidance on the length of depuration required for a batch of shellfish based on previously obtained data from the harvest location of the shellfish in question.

Chapter 4 serves as a comparison between the application of using different probability distributions as the descriptors of the pathogen loads within the model constructed in Chapter 3, and how any differences observed impact resultant depuration times obtained from the different distribution types.

The work carried out in Chapter 5 extends the model from Chapter 3, incorporating a compartmentalised model of an individual oyster into the model. The Chapter goes on to examine how this compartmentalisation of pathogen loads within an individual mollusc can have consequences for, not only depuration times, but also for testing protocols currently employed.

Chapter 6 considers the question of how pathogens resulting from sewage treatment works discharges can drive up the levels of pathogens transmitted into shellfish farms in close proximity to the treatment plants. The model constructed here includes a mechanism to increase the total treatment capacity of a sewage treatment works, and how increasing and decreasing this capacity impacts the levels of pathogens in shellfish nearby. This Chapter also examines how this variable capacity can alter the length of depuration times required by industry stakeholders to ensure that clean shellfish are passed further into the supply chain, and analyses the trade-offs between capacity levels and minimum depuration times.

Finally, we propose further work on the back of this thesis, looking at open questions from the modelling carried out and what further research would be required to validate the mathematical models constructed.

3

MATHEMATICAL MODEL OF DEPURATION PROCESS

3.1 INTRODUCTION

This chapter explains in detail the construction of a mathematical model of depuration, and how it can be used as a pathogen mitigation tool within the shellfish industry. In particular this chapter considers the following questions:

(i) How are pathogen loads distributed across a shellfish population before depuration?

(ii) How do the pathogen loads decay due to the depuration process?

(iii) What is the probability that a pathogen load in a randomly sampled oyster from that population may exceed some legislated, threshold value?

(iv) What minimum length of time must a batch of oysters be depurated for to ensure they present a negligible food safety risk to consumers?

Answering these four questions allows us to construct a mathematical model which describes the dynamics of pathogen reduction due to depuration. The answers provided in this chapter are based upon robust evidence within the current literature, as well as mathematical processes that realistically describe the impact of depuration on pathogens within shellfish. The construction of the model is fully explained, with each step clearly explained. Parameterisation of the model is based upon available data and industry literature, and the results obtained from the model are presented. How these results pertain to the shellfish industry within the UK and further afield is discussed, and the

advantages that a population model of depuration brings to both the industry and regulators are also emphasised.

Bivalve shellfish use a filter feeding process to gain nutrients from their surrounding waters, and this feeding process can concentrate contaminants into the digestive system of the shellfish [10, 16, 28, 29, 30]. Therefore the potential exists for shellfish to contain bacterial, viral or particulate contaminants, any of which can transmit pathogens into the human population once eaten by a consumer. Thus, minimising any potential exposure to contaminants is of critical importance to both the industry and the consumer, as any illness which stems from the consumption of shellfish would result in, not only illness in the human population, but also damage to the reputation of the shellfish industry as a whole.

Ideally shellfish should be farmed in waters with very low levels of pollution, however this is not always possible due to economic and socio-geographic reasons, with many farms located close to population centres [32, 33, 34]. The issue of shellfish farming in waters which are not rated as Class A was highlighted in Chapter 1 and is not discussed further in this chapter. Modelling which describes how pathogens are transmitted into shellfish from pollution events is discussed later in Chapter 6.

The rating system of UK coastal waters is currently based upon the detection and quantification of the presence of the bacterium *E. coli* (see Section 1.5). Though *E. coli* is a reliable indicator for the presence of faecal pollution and is the industry metric used when classifying water quality, it is a poor indicator of viral and chemical contamination. Several current studies have shown that there is a poor correlation between levels of *E. coli* and other contaminants such NoV [29, 64, 65], and as such any viral pathogens present in shellfish

could reach the consumer in notable quantities due to this low correlation. This is of especial concern, as pathogens such as NoV are not currently controlled under any national or E.U. legislation [40]. As detailed in Section 1.5, to ensure any pathogenic transmission to the consumer is minimised, only class A sourced shellfish can pass straight into the wholesale and retail markets. Class B shellfish harvests are required by law to undergo a minimum of 42 hours of depuration before re-entering the supply chain. Section 1.7 discusses the technical details on the depuration process, and the criteria that must be considered when implementing efficient depuration practices.

The shellfish industry as a whole is and has been pragmatic to self-regulate for pathogens and other contaminants over and above the current coliform mitigation requirements. At present, NoV levels within shellfish are reduced only by methods put in place to mitigate other contaminants, despite posing a potential risk to consumer health [14, 18, 28]. Although several countries conduct monitoring for viral contamination, thus far no producer countries have implemented legislative standards [13, 17, 66]. Rather legislators control harvesting from sites which are adversely impacted by pollution events that mainly emanate from adjacent sewage treatment works. Harvest sites which do not attain the top classification of A require the pathogen controls to be implemented as discussed previously, one of which is depuration.

In the rest of this chapter, we employ mathematical techniques to construct a model of the depuration process by determining how NoV levels in shellfish populations change over time when the initial NoV load in each oyster, and the way in which the initial loads evolve over time are considered.

3.3.1 *Pre-Depuration Pathogen Distribution*

To construct our model of depuration we are first required to mathematically describe the distribution of pathogens across the batch or population of shellfish before they undergo the depuration process. This can be achieved by the analysis of a data set that records pathogen loads detected in a population of shellfish before they have been cleaned using depuration.

3.3.1.1 *Experiment*

A data set of NoV loads within a population of Pacific oysters (*Crassostrea gigas*) was provided to me by Dr James Lowther of CEFAS, who carried out the following steps to obtain the data.

Six hundred oysters were sourced from a class B site and, immediately upon receipt, ten oysters were randomly selected and tested for the presence of NoV using a standardised, Reverse Transcription Polymerase Chain Reaction (PCR) method [15]. The remaining oysters were divided between four commercial oyster sacks and split between two depuration/recirculation tanks in the experimental facility at the CEFAS Weymouth laboratory, each containing 500L of filtered seawater and set to a temperature of 18°C. The animals were left for four days with ultraviolet irradiation of the recirculated water to aid in the removal of any residual, undetectable virus and to allow restoration of filter feeding activity.

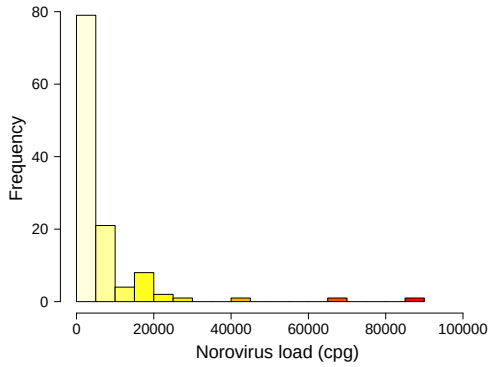
A bioaccumulation method [31] was then used in order to ensure that the oysters contained high levels of NoV at the outset of the subsequent study to

ensure detection and a high base line of load values. This involved adding 10ml of influent, municipal sewage collected from a large local water treatment works (population equivalent to 30,000), and was supplemented with 500µl of a suspension of G_I and G_{II} positive human faecal material (containing approximately 1.25 $\times 10^8$ viral genome copies for each genogroup), which was added to each of the tanks. Oysters were then left to bioaccumulate virus for 16 hours. After removal from the tanks, the oysters were removed from their holding sacks and thirty oysters from each were randomly selected for testing for each genotype.

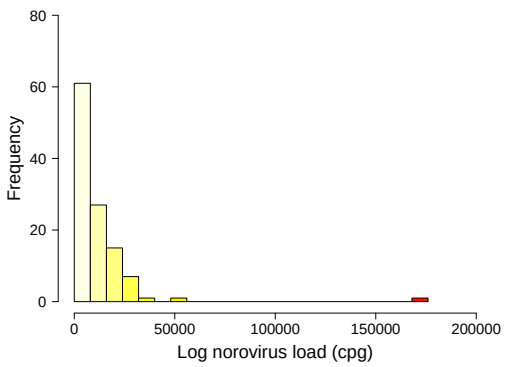
3.3.1.2 *Analysis*

The results from these samples were used as the foundation for determining the distribution type that reasonably describes NoV across an oyster population. Test results with less than 10 NoV cpg obtained as part of the depuration study described above were removed from the data sets, as these were below the limit of quantitation for NoV [2]; this affected less than 2% of data points for genotype I and none for genotype II.

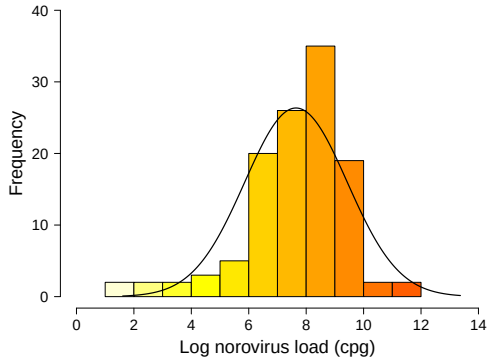
The NoV counts for both genotypes are plotted as histograms in Figure 3.1, firstly as unlogged data in Figures 3.1a and 3.1b , and as log-transformed in Figures 3.1c and 3.1d. The unlogged data obtained from the experiment are characterised by highly skewed distributions (genotype I : mean = 6323 cpg, standard deviation = 11520 cpg; genotype II : mean = 10921 cpg, standard deviation = 17577 cpg). However the log-data can be well described by a normal distribution (Kolmogorov-Smirnov test; genotype I: $p = 0.124$, $n = 116$; genotype II: $p = 0.104$, $n = 113$). Figures 3.1c and 3.1d both have fitted normal distribution curves added for comparison.



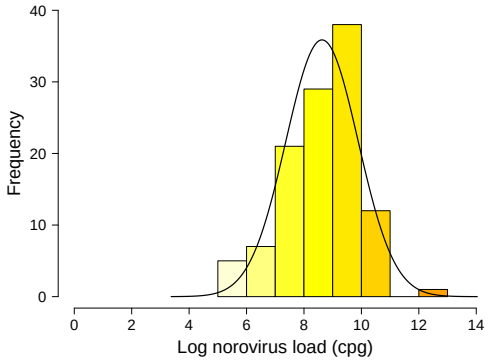
(a) Genotype I histogram



(b) Genotype II histogram



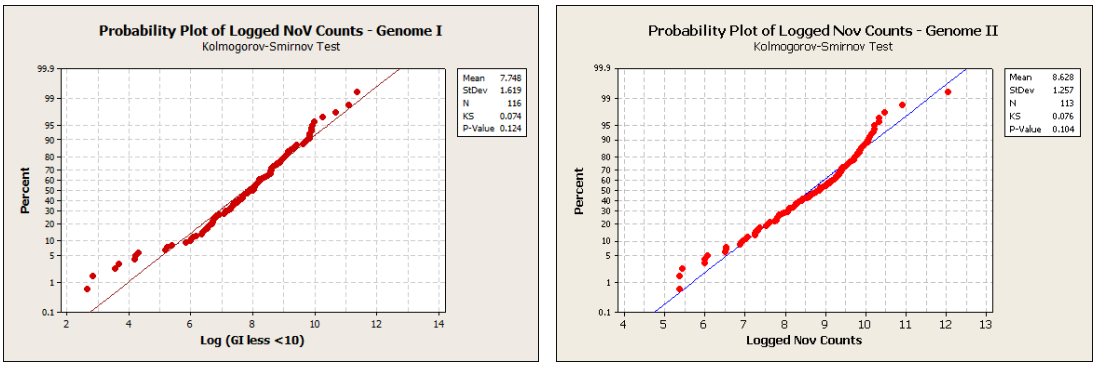
(c) Genotype I histogram — logged data



(d) Genotype II histogram — logged data

Figure 3.1: Histograms of bioaccumulation data supplied by Dr James Lowther of CEFAS. Figures (a) and (b) show the unlogged data for genotypes I and II respectively, while Figures (c) and (d) show the log norovirus loads cpg. Genotype I data has $n = 116$ and genotype II has $n = 113$

Figure 3.2 shows the goodness of fit plots for both genotypes using the Kolmogorov-Smirnov test methodology. These plots demonstrate that both genome types conform closely to a lognormal description, although they do indicate that there are some outliers at the lower and higher values, departing from the straight line which signifies full lognormality. These plots reinforce our assertion that NoV loads across a population of shellfish can be well described by a lognormal distribution, an assertion which is backed up by



(a) Genotype I

(b) Genotype II

Figure 3.2: Goodness of fit lognormal Kolmogorov-Smirnov Test plots

the literature where many water-borne pathogens have been previously been described as log-normal [44, 53, 54, 55, 56].

3.3.2 *Pre-Depuration Model*

We can now describe the distribution of NoV loads across a population of oysters by a lognormal distribution. We define the probability density distribution of NoV loads within a population of shellfish, $X_0 = x_0$, at depuration time $t = 0$ as $P(x_0)$ and so

$$P(x_0) = \frac{1}{\sqrt{2\pi}\,\sigma_0\,x_0}\exp\left\{\frac{-(\ln(x_0)-\mu_0)^2}{2\sigma_0^2}\right\}, \tag{3.1}$$

where μ_0 is the location parameter of the lognormal distribution, and σ_0 is the scale parameter [6, 58].

Echoing the definition from Equation 2.15, we also state here that the expected value of x_0 is $E(x_0) = \exp\{\mu_0 + \sigma_0^2/2\}$. This is equivalent to the arithmetic mean value of the distribution $P(x_0)$, therefore we also define

$$\bar{x}_0 = \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\}. \tag{3.2}$$

Again based on definitions in Section 2.3.2, we state the standard deviation of $P(x_0)$ as

$$SD(x_0) = \sqrt{\exp\{\sigma_0^2\} - 1}\ \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\}, \tag{3.3}$$

This formula can be rewritten as

$$SD(x_0) = \bar{x}_0\ \sqrt{\exp\{\sigma_0^2\} - 1} \tag{3.4}$$

and

$$SD(x_0) = CV(x_0)\ \bar{x}_0\,, \tag{3.5}$$

where $CV(x_0)$ is the coefficient of variation of the lognormal distribution [6].

3.3.3 *Maximum Likelihood Estimation of Parameters*

Current testing protocols provide a measure of the arithmetic mean of a sample, obtained from a ten oyster homogenate. However, if future technological advances allowed a robust and cost effective test procedure to be employed which could provide sufficiently large data sets of single oyster NoV loads from harvest locations, then lognormal distribution location and spread parameters could be obtained using maximum likelihood estimation methods.

Estimates of μ_0 and σ_0 parameters can be derived for X_i ($i \in \{1, 2, ..., n\}$), a set of n independent samples drawn from a lognormal distribution. For simplicity of calculation, an estimate of σ_0^2 will be derived. The parameter

estimates are obtained using the log-likelihood function of the lognormal distribution $P(X_i)$ which is defined as [67, 68]

$$\mathcal{L}(\mu_0, \sigma_0^2|X) = \ln\left(\prod_{i=1}^{n}[P(X_i)]\right)$$

$$\Rightarrow \mathcal{L}(\mu_0, \sigma_0^2|X) = \ln\left(\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_0^2}X_i}\exp\left[\frac{-\ln(X_i-\mu_0)^2}{2\sigma_0^2}\right]\right)$$

$$\Rightarrow \mathcal{L}(\mu_0, \sigma_0^2|X) = \ln\left(\frac{1}{\left(\sqrt{2\pi\sigma_0^2}\right)^n}\prod_{i=1}^{n}\frac{1}{X_i}\exp\left[\sum_{i=1}^{n}\frac{-\ln(X_i-\mu_0)^2}{2\sigma_0^2}\right]\right)$$

$$\Rightarrow \mathcal{L}(\mu_0, \sigma_0^2|X) = \ln\left((2\pi\sigma_0^2)^{-n/2}\prod_{i=1}^{n}\frac{1}{X_i}\exp\left[\sum_{i=1}^{n}\frac{-\ln(X_i-\mu_0)^2}{2\sigma_0^2}\right]\right). \quad (3.6)$$

Using the laws of logarithms, Equation 3.6 can be rearranged into

$$\mathcal{L}(\mu_0, \sigma_0^2|X) = -\frac{n}{2}\ln(2\pi\sigma_0^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}(\ln(X_i)-\mu_0)^2}{2\sigma_0^2},$$

and further simplified into

$$\mathcal{L}(\mu_0, \sigma_0^2|X) = -\frac{n}{2}\ln(2\pi\sigma_0^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}\ln(X_i)^2}{2\sigma_0^2} + \frac{\sum_{i=1}^{n}\ln(X_i)\mu_0}{\sigma_0^2} - \frac{n\mu_0^2}{2\sigma_0^2}. \quad (3.7)$$

We use first order derivatives of Equation 3.7 with respect to μ_0 and σ_0 to obtain estimates of these parameters. Starting with μ_0, we set $\partial\mathcal{L}/\partial\mu_0 = 0$ and solve for an estimate of μ_0:

$$\frac{\partial\mathcal{L}}{\partial\mu_0} = \frac{\sum_{i=1}^{n}\ln(X_i)}{\sigma_0^2} - \frac{2n\hat{\mu}_0}{2\hat{\sigma}_0^2} = 0$$

$$\Rightarrow \hat{\mu}_0 = \frac{\sum_{i=1}^{n}\ln(X_i)}{n}. \quad (3.8)$$

For an estimate of $\hat{\sigma}_0^2$, we follow a similar process, this time differentiating with respect to σ_0^2:

$$\frac{\partial\mathcal{L}}{\partial\sigma_0^2} = -\frac{n}{2\hat{\sigma}_0^2} + \frac{\sum_{i=1}^{n}(\ln(X_i)-\hat{\mu}_0)^2}{2\hat{\sigma}_0^4} = 0$$

$$\Rightarrow \hat{\sigma}_0^2 = \frac{\sum_{i=1}^{n}\left(\ln(X_i) - \frac{\sum_{i=1}^{n}\ln(X_i)}{n}\right)^2}{n}. \quad (3.9)$$

To show that the estimators of $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ maximize the log-likelihood function $\mathcal{L}(\mu_0, \sigma_0|X)$, we must consider the second order derivatives of the function. The Hessian matrix, $H(\mu_0, \sigma_0^2)$, of $\mathcal{L}(\mu_0, \sigma_0^2|X)$ can be shown to be

$$H(\mu_0, \sigma_0^2) = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \mu_0^2} & \frac{\partial^2 \mathcal{L}}{\partial \sigma_0^2 \partial \mu_0} \\ \frac{\partial^2 \mathcal{L}}{\partial \mu_0 \partial \sigma_0^2} & \frac{\partial^2 \mathcal{L}}{\partial (\sigma_0^2)^2} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\sigma_0^2} & 0 \\ 0 & \frac{-\sum_{i=1}^{n} (\ln(X_i) - \hat{\mu}_0)^2}{2(\sigma_0^2)^3} \end{pmatrix} \tag{3.10}$$

The determinant of $H(\mu_0, \sigma_0^2)$ is greater than zero and the top left entry of $H(\mu_0, \sigma_0^2)$ is always negative as n, $\sigma_0^2 > 0$. Therefore both the estimators $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ maximize the likelihood function $\mathcal{L}(\mu_0, \sigma_0^2|X)$ [67].

Thus, parameter values to accurately estimate the shape of a pre-depuration lognormal distribution can be obtained from single oyster data samples using Equations 3.8 and 3.9.

3.3.4 *Evolution Of The Depuration Distribution Over Time*

Let x_t be a NoV load in a single oyster at a time t. A probability distribution $P(x_t)$ can be derived for x_t, i.e. a PDF that describes the distribution of pathogens across a population of shellfish for any time $t \geqslant 0$ during the depuration process. To accomplish this we must first define how depuration changes each individual mollusc's pathogen load over time. We assume here that there is a constant, proportional removal of a mollusc's pathogen load for each defined time unit, and thus can be described by an exponential decay of pathogen loads in individual shellfish due to the depuration process [69, p. 418]:

$$x_t = x_0 \ \exp\{-bt\}, \tag{3.11}$$

where b quantifies the depuration decay rate per unit of time. Unlike the models summarised in Chapter 2, we do not assume a constant residual pathogen load that remains in the shellfish, no matter the length of depuration

undertaken [62, 63].

As x_0 is related to x_t by a continuous function, this enables the change of variables process to be applied to $P(x_0)$ to obtain a PDF for any time t during depuration [70, p. 66]. This change of variables process is shown here as

$$\begin{aligned} P(x_t)\ dx_t &= P(x_0)dx_0 \\ \Rightarrow P(x_t) &= P(x_0)\left(\frac{dx_0}{dx_t}\right) \\ \Rightarrow P(x_t) &= P(x_0)\left(\frac{dx_t}{dx_0}\right)^{-1}. \end{aligned} \tag{3.12}$$

Substituting in our definition of $P(x_0)$ from Equation 3.1 and subsequently the inverse of the derivative of Equation 3.11:

$$\begin{aligned} P(x_t) &= \frac{1}{\sqrt{2\pi}\sigma_0\ x_0}\exp\left\{\frac{-(\ln(x_0)-\mu_0)^2}{2\sigma_0^2}\right\}\left(\frac{dx_t}{dx_0}\right)^{-1} \\ \Rightarrow P(x_t) &= \frac{1}{\sqrt{2\pi}\sigma_0\ x_0}\exp\left\{\frac{-(\ln(x_0)-\mu_0)^2}{2\sigma_0^2}\right\}\exp\{bt\} \\ \Rightarrow P(x_t) &= \frac{\exp\{bt\}}{\sqrt{2\pi}\sigma_0\ x_0}\exp\left\{\frac{-(\ln(x_0)-\mu_0)^2}{2\sigma_0^2}\right\}. \end{aligned}$$

The change of variables is completed by replacing x_0 with $x_t\exp\{bt\}$ (from a rearrangement of Equation 3.11) to yield

$$\begin{aligned} P(x_t) &= \frac{\exp\{bt\}}{\sqrt{2\pi}\ \sigma_0\ x_t\exp\{bt\}}\exp\left\{\frac{-(\ln(x_t\exp\{bt\})-\mu_0)^2}{2\sigma_0^2}\right\} \\ \Rightarrow P(x_t) &= \frac{1}{\sqrt{2\pi}\ \sigma_0\ x_t}\exp\left\{\frac{-(\ln(x_t)+bt-\mu_0)^2}{2\sigma_0^2}\right\}. \end{aligned} \tag{3.13}$$

Thus Equation 3.13 describes the distribution of pathogen loads across a population of shellfish during depuration for any time $t \geqslant 0$. $P(x_t)$ also describes a lognormal distribution, with the same measure of spread, σ_0, as $P(x_0)$. However the measure of location is now described by $\mu_0 - bt$. Descriptions of μ_t and σ_t are discussed further in the next section.

3.3.5 *Measures of Location and Spread of* $P(x_t)$

A corollary on the measures of location and scale of the distribution describing depuration is shown here. Definitions of these measures will prove to be of significant use in interpreting results later.

3.3.5.1 *Measures of Location of* $P(x_t)$

We have already stated that the arithmetic mean of $P(x_0)$ is as follows (Equation 3.2)

$$\bar{x}_0 = \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\}. \tag{3.14}$$

This formula needs to be extended to describe the arithmetic mean for any time point during depuration, $\bar{x}_t$, and so we require definitions for both μ_t and σ_t.

To ascertain the value of the location parameter μ_t, we need to consider the equation of the depuration distribution $P(x_t)$ (Equation 3.13), where

$$P(x_t) = \frac{1}{\sqrt{2\pi}\,\sigma_0\,x_t}\exp\left\{\frac{-(\ln(x_t) + bt - \mu_0)^2}{2\sigma_0^2}\right\}, \tag{3.15}$$

and to compare it with the definition of the general lognormal distribution as defined by Equation 2.12:

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma\,x}\exp\left\{\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right\}. \tag{3.16}$$

This equation is the PDF of $P(x_0)$ less the subscripts of 0, thus we are comparing the pre-depuration and during depuration PDFs. The exponent in both equations performs the Z-score correction, with the numerator in the exponent being corrected by $-(bt - \mu_0)$. Thus the location parameter at any time $t \geqslant 0$ can be calculated from

$$\mu_t = \mu_0 - bt. \tag{3.17}$$

Following the same process for σ_t, it is apparent that the measure of spread is the same for both $P(X_0 = x_0)$ and $P(X_t = x_t)$. Thus, we can state that

$$\sigma_t = \sigma_0 \,. \tag{3.18}$$

Hence, substituting Equations 3.17 and 3.18 into Equation 3.14 will provide a definition of the arithmetic mean of $P(X_t = x_t)$:

$$\begin{aligned} \bar{x}_t &= \exp\left\{\mu_0 - bt + \frac{1}{2}\sigma_0^2\right\} \\ \Rightarrow \bar{x}_t &= \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\} \ \exp\{-bt\} \\ \Rightarrow \bar{x}_t &= \bar{x}_0 \ \exp\{-bt\} \,. \end{aligned} \tag{3.19}$$

Thus the value of $\bar{x}_t$ across the population decreases as depuration progresses, with a decay gradient of $\exp\{-bt\}$ after time t.

This result also follows from Equation 3.11 which defined how an individual NoV load decayed due to depuration:

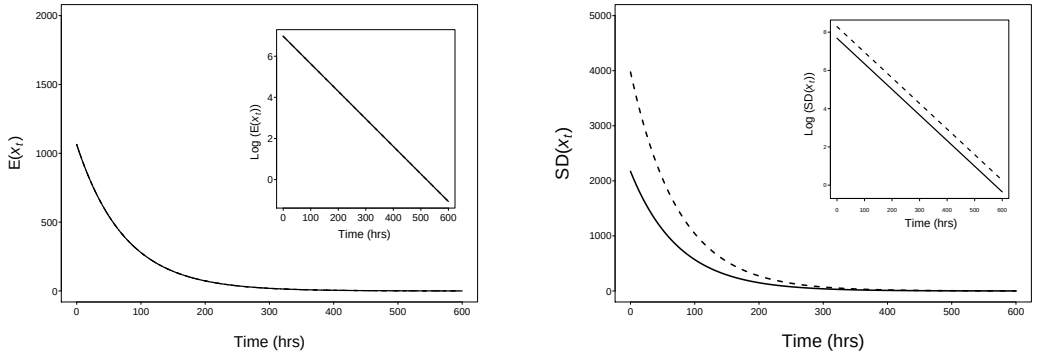
$$x_t = x_0 \ \exp\{-bt\} \,.$$

This describes a linear relationship between x_t and x_0, with the decay gradient fixed for the whole population at $\exp\{-bt\}$ after a time t. Therefore each individual NoV load will have decayed by $\exp\{-bt\}$, and so the arithmetic mean value $\bar{x}_0$ will also have decayed by $\exp\{-bt\}$.

3.3.5.2 *Measures of Spread of* $P(x_t)$

Equation 3.3 in Section 3.3.2 states that the standard deviation of $P(x_0)$ is

$$SD(x_0) = \sqrt{\exp\{\sigma_0^2\} - 1} \ \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\} \,.$$



(a) Depuration decay of expected value of x_t (b) depuration decay of std. deviation of x_t

Figure 3.3: Decay of expected value and standard deviation of x_t due to individual exponential decay. Inset plots show log vertical axes. Both plots obtained from $x_0 = 1064$ NoV cpg. Dotted line in Figure 3(b) corresponds to $\sigma_0 = 1.645$, while solid line has $\sigma_0 = 1.282$. All vertical axes in NoV cpg

Thus the variance of $P(x_0)$ is

$$\begin{aligned} V(x_0) &= [SD(x_0)]^2 \\ \Rightarrow V(x_0) &= \left(\exp\left\{\sigma_0^2\right\} - 1\right)\ \exp\left\{2\mu_0 + \sigma_0^2\right\} . \end{aligned}$$

We can obtain the variance and standard deviation of x_t by substitution of Equations 3.18 and 3.17:

$$\begin{aligned} V(x_t) &= \left(\exp\left\{\sigma_t^2\right\} - 1\right)\ \exp\left\{2(\mu_0 - bt) + \sigma_t^2\right\} \\ \Rightarrow V(x_t) &= \left(\exp\left\{\sigma_t^2\right\} - 1\right)\ \exp\left\{2\mu_0 - 2bt + \sigma_t^2\right\} \\ \Rightarrow V(x_t) &= \left(\exp\left\{\sigma_t^2\right\} - 1\right)\ \exp\left\{2\mu_0 + \sigma_t^2\right\} \exp\{-2bt\} \\ \Rightarrow SD(x_t) &= \underbrace{\sqrt{(\exp\{\sigma_t^2\} - 1)\ \exp\{2\mu_0 + \sigma_t^2\}}}_{SD(x_0)} \exp\{-bt\} , \end{aligned} \tag{3.20}$$

and so

$$SD(x_t) = SD(x_0)\ \exp\{-bt\} . \tag{3.21}$$

Thus, both the arithmetic mean ($\bar{x}_t$) and the standard deviation ($SD(x_t)$) decline to zero at the same rate as each individual oyster load.

Note that, as the coefficient of variation of the lognormal distribution is independent of the arithmetic mean (see Equation 2.16) and dependent only on σ_0, then the value of $CV(x_0)$ will hold constant for any change to the mean of the distribution for any $t \geqslant 0$.

3.3.6 *Tail Analysis*

Equation 3.13 describes the distribution of a water-borne pathogen across a population of shellfish at time t during the depuration process. The efficacy of the depuration process is described by the decay rate parameter b. We now introduce two control parameters:

Ψ : NoV load level — below which an individual shellfish is deemed to not constitute a risk to the consumer;

ϕ : NoV assurance level — acceptable proportion of the shellfish population which has a NoV level below that of Ψ.

Both of these parameters allow implementation of food safety controls within our model. Figure 3.4 demonstrates how these parameters split the distribution of into 2 parts. The diagram presents the general shape of a lognormal distribution, with the parameter Ψ shown on the horizontal axis.

The green area to the left of Ψ constitutes the proportion of the shellfish population with pathogen loads $x_t < \Psi$, and can be called as the body of the distribution. The red area corresponds to the probability of a randomly selected shellfish having a pathogen load $x_t > \Psi$, and is called the tail of the distribution. The tail of the distribution is of particular interest as it quantifies the risk of marketing 'non-compliant' oysters, i.e. oysters containing pathogen loads exceeding a certain value, Ψ. At time t, the probability in the tail of the distribution is given by

$$P(x_t > \Psi) = \int_{\Psi}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0\, x_t} \exp\left\{\frac{-(\ln(x_t) + bt - \mu_0)^2}{2\sigma_0^2}\right\}\, dx_t\,. \tag{3.22}$$

This describes a PDF, and so its value must be in the range $0 \leqslant P(x_t > \Psi) \leqslant 1$.

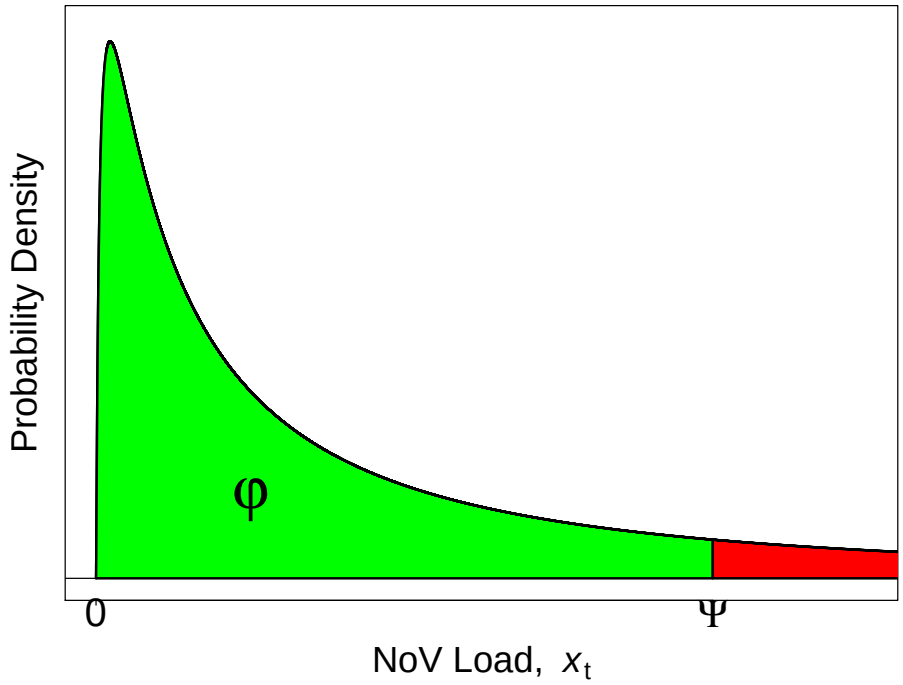


Figure 3.4: Probability distribution plot, $P(x_t)$. The red section corresponds to total probability that a randomly selected shellfish has a pathogen load x_t greater than the threshold limit Ψ. The green zone denotes the total probability that a random shellfish will have a pathogen load $x_t < \Psi$

Conversely, we describe the body of the distribution as the area under the curve from zero to Ψ (as shown as the green area in Figure 3.4) by

$$P(x_t < \Psi) = \int_0^{\Psi} \frac{1}{\sqrt{2\pi}\sigma_0\, x_t} \exp\left\{\frac{-(\ln(x_t) + bt - \mu_0)^2}{2\sigma_0^2}\right\} dx_t\,. \tag{3.23}$$

The value of $P(x_t < \Psi)$ will never be equal to 1 (at least mathematically) due to the inherent nature of the lognormal distribution, with the tail continuing towards ∞ along the horizontal axis. The consequence of this is that the red area of Figure 3.4 will never disappear completely, i.e. $P(x_t > \Psi) \neq 0$.[1]

Therefore we deploy a second control parameter, ϕ, the area under the curve of the PDF between 0 and Ψ. This is a proportion of the total shellfish

1 In reality, x_t does not extend to ∞, and so the "tail" in principle could disappear.

population with NoV loads less than Ψ, and so this threshold value will also be in the range $0 \leqslant \phi \leqslant 1$. We wish to maximise the number of shellfish which have pathogen NoV loads less than Ψ, therefore ideally $\phi \lesssim 1$ to ensure that almost all shellfish would have a pathogen load below Ψ.

Thus we can now state that the depuration model with pathogen controls incorporated is described by

$$P(x_t < \Psi) = \int_0^{\Psi} \frac{1}{x_t \sigma_0 \sqrt{2\pi}} \exp\left\{\frac{-(\ln(x_t) + bt - \mu_0)^2}{2\sigma_0^2}\right\} \, dx_t = \phi \,, \tag{3.24}$$

which is a definition of ϕ, given Ψ. These two parameters are independent of each other, and varying their values will allow legislators to tailor the level of pathogen control according to their own requirements.

3.3.7 *Minimum Depuration Time*

The aim of constructing a model of depuration is to provide the industry with a tool to minimise any food safety risks to the consumer, while maximising profitability for stakeholders within the industry itself. The inclusion of the control parameters as described in the previous section will provide a significant level of food security when they are conformed to by the depuration process. However, when does a depurator know they have achieved this?

Solving Equation 3.24 for t allows us to determine the minimum depuration time required to attain the assurance level ϕ with the maximum allowable pathogen load level Ψ being enforced. A solution can be obtained by applying the substitution

$$z = \frac{\ln(x_t) + bt - \mu_0}{\sqrt{2}\sigma_0} \Rightarrow dx_t = x_t\sqrt{2}\sigma_0 \; dz$$

and appropriately changing the limits of the integral:

$$\begin{aligned} \frac{1}{x_t\sigma_0\sqrt{2\pi}} \int_{-\infty}^{W} \exp\{-z^2\} \, x_t\sqrt{2}\sigma_0 \; dz = \phi \\ \Rightarrow \frac{1}{\sqrt{\pi}} \int_{-\infty}^{W} \exp\{-z^2\} \, dz = \phi \,, \end{aligned} \tag{3.25}$$

where:

$$W = \frac{\ln(\Psi) + bt - \mu_0}{\sqrt{2}\sigma_0} \,.$$

The left hand side of Equation 3.25 closely resembles the definition of the standard, normal distribution with $\bar{x} = \mu = 0$ and $\sigma = 1$ (cf. Equation 2.6):

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp\{-z^2\} \, dz = 1 \,,$$

and which is shown in Figure 3.5. As both are PDFs, then the total area under the curve of their distributions are equal to 1. As they are also both symmetric

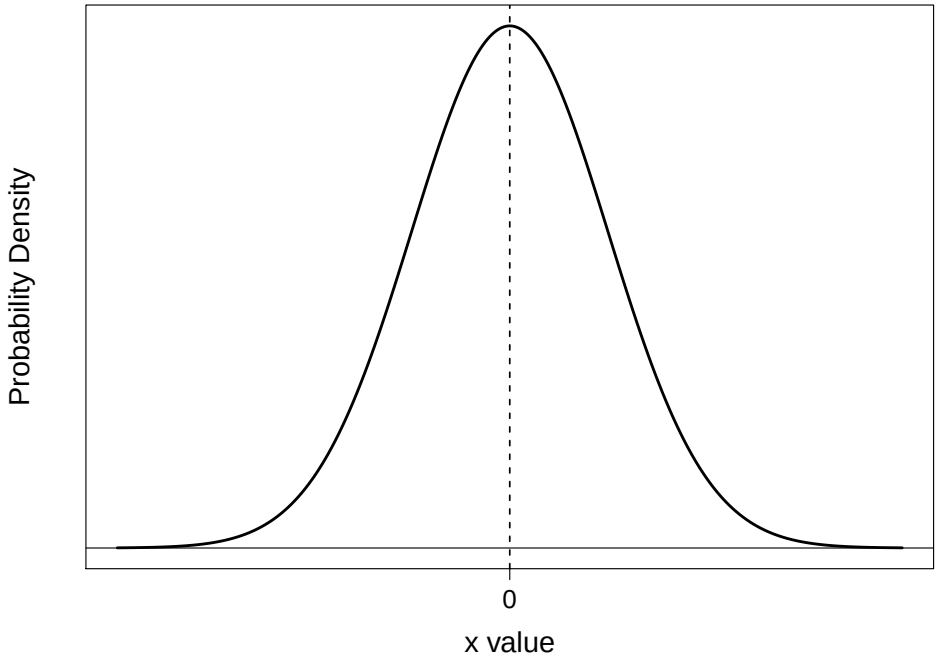


Figure 3.5: Probability density plot of standard, normal distribution with $\bar{x} = 0$

about zero, then the area under the standard normal distribution curve in the range $[-\infty, 0)$ is equal to $\frac{1}{2}$,

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{0} \exp\{-z^2\}\, dz = \frac{1}{2}\,.$$

Subtracting this from Equation (3.25) results in

$$\frac{1}{\sqrt{\pi}} \int_{0}^{W} \exp\{-z^2\}\, dz = \phi - \frac{1}{2}\,.$$

Multiplying both sides of this equation by 2 and applying the erf function which was previously defined by Equation 2.10 as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp\{-t^2\}\, dt \tag{3.26}$$

allows us to further derive:

$$erf(W) = 2\phi - 1\,. \tag{3.27}$$

Applying the inverse error function (erf^{-1}) to both sides of Equation 3.26 provides

$$W = erf^{-1}(2\phi - 1)$$
$$\Rightarrow \ln(\Psi) + bt - \mu_0 = \sqrt{2}\,\sigma_0\, erf^{-1}(2\phi - 1)$$
$$\Rightarrow bt = \sqrt{2}\,\sigma_0\, erf^{-1}(2\phi - 1) - \ln(\Psi) + \mu_0\,,$$

and finally solving for time t yields

$$T = t = b^{-1}\left[\sqrt{2}\,\sigma_0\, erf^{-1}(2\phi - 1) - \ln(\Psi) + \mu_0\right]. \tag{3.28}$$

Equation 3.28 calculates the MDT, $T = t$, required to ensure that the probability that a randomly chosen oyster has a NoV load which is less than some selected NoV threshold limit Ψ is equal to ϕ.

3.3.8 *Estimating Variability of the Distribution*

The MDT depends not only on ϕ and Ψ, but also on μ_0 and σ_0. The two control parameters ϕ and Ψ would either be determined by legislators, or self-imposed by the industry stakeholders. Obtaining realistic estimates for μ_0 and σ_0 is more problematic however, especially for NoV. Although we were able to obtain σ_0 for the data described in Section 3.3.1, this experiment involved so many measurements that it would be impractical and expensive to consider that it can be carried out on a regular basis with regards to NoV load monitoring for all harvests. The current standard assay for NoV detection is carried out on homogenates of 10 oysters [15], and so provides an arithmetic mean in terms of NoV cpg ($\bar{x}_0$) of the 10 oysters sampled.

We can convert the arithmetic mean of the lognormal distribution into the geometric mean (in this case μ_0) by applying the relationship between the

arithmetic mean and geometric mean of a lognormal distribution which is described by

$$\mu_0 = \ln(\bar{x}_0) - \frac{1}{2}\sigma_0^2, \tag{3.29}$$

[6]. However, to obtain a value of μ_0 we need to have a measure of the variability of the pathogen across the population, σ_0. Obtaining estimates of σ_0 from experimental data would require multiple tests to be performed, which is both time-consuming and costly [39]. However, modelling can be used to provide a 'worst case scenario' value of the variability of the pathogen's distribution. This, in turn, will provide a maximum estimate of the MDT.

To achieve this, Equations 3.28 and 3.29 are combined to provide the MDT in terms of the variability and the arithmetic mean of the distribution:

$$T(\sigma_0, \bar{x}_0) = b^{-1}\left[-\frac{1}{2}\sigma_0^2 + \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1)\,\sigma_0 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right], \tag{3.30}$$

where we explicitly note the dependence of T on μ_0 and σ_0.

Equation 3.30 is a concave quadratic function in terms of the variability scale parameter σ_0, confirmed by the negative value of its second derivative, where

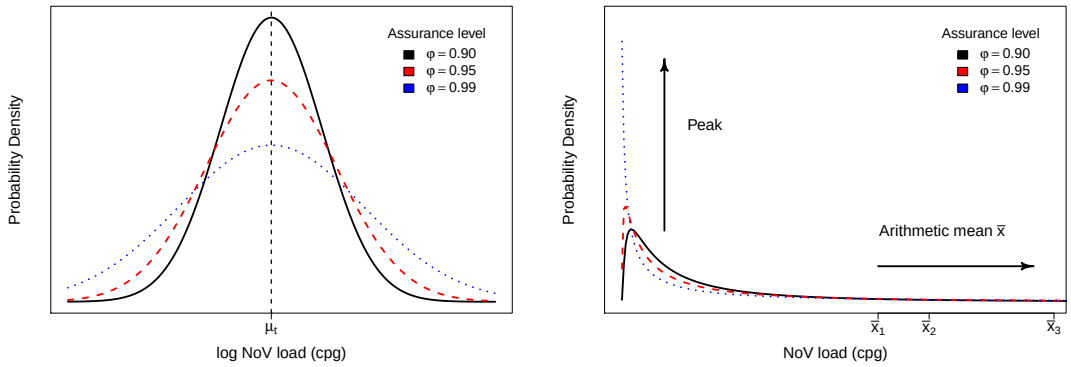
$$\frac{\partial T}{\partial \sigma_0} = b^{-1}\left(-\sigma_0 + \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1)\right)$$
$$\Rightarrow \frac{\partial^2 T}{\partial \sigma_0^2} = -b^{-1} < 0.$$

As $b \geqslant 0$, Equation 3.30 will have a maximum value for $T(\sigma_0, \bar{x}_0)$ when

$$\frac{\partial T}{\partial \sigma_0} = 0$$
$$\Rightarrow -\sigma_0 + \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1) = 0$$
$$\Rightarrow \sigma_0 = \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1) \tag{3.31}$$

Thus, when $\sigma_0 = \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1)$, the variability of the system is such that the MDT is maximised. This 'worst case scenario' variability is a consequence of

the flattening of the distribution curve as σ_0 increases, forcing the area under the distribution curve into higher NoV load values, while simultaneously displacing area into the lower value ranges.



(a) log-transformed (normal) PDFs

(b) Lognormal PDFs

Figure 3.6: Generic plots showing impact of increasing assurance level ϕ. (a) shows log-transformed normal distributions for $\phi = 0.90,\ 0.95,\ 0.99$, whereas (b) shows the equivalent unlogged distributions

Figure 3.6a displays three log-transformed normal PDFs, corrpesonding to three values of ϕ. This figure shows that increasing the value of ϕ (and consequentially increasing σ_0) flattens the shape of the log transformed PDF curve. This flattening increases the area under the curve at the extreme left and right of the plots, displacing area towards $-\infty$ and ∞.

Figure 3.6b represents the same three PDFs without any log-transformation. This plot shows that increasing ϕ displaces area under the curve towards zero, while at the same time displacing area under the curve towards ∞. As the lognormal distribution has a domain restriction of $x > 0$, any increase in area towards zero results in an increase in the PDF's kurtosis. However there is no upper limit on the value of the domain, and so displacing area towards ∞ results in an increase in the weight of the tail of the distribution.

We can rearrange Equation 3.30 by excluding the variability parameter σ_0, and substituting in Equation 3.31 to obtain T_{WCV}, the maximum (worst case scenario) minimum depuration time:

$$T_{WCV} = b^{-1}\left[-\frac{1}{2}\sigma_0^2 + \sqrt{2}\text{erf}^{-1}(2\phi - 1)\,\sigma_0 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right]$$

$$\Rightarrow T_{WCV} = b^{-1}\left[-\frac{1}{2}\left(\sqrt{2}\text{erf}^{-1}(2\phi - 1)\right)^2 + \left(\sqrt{2}\text{erf}^{-1}(2\phi - 1)\right)^2 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right]$$

$$\Rightarrow T_{WCV} = b^{-1}\left[-\left(\text{erf}^{-1}(2\phi - 1)\right)^2 + 2\left(\text{erf}^{-1}(2\phi - 1)\right)^2 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right]$$

Thus the minimum depuration time based on a (WCV) approach can be calculated from

$$T_{WCV} = b^{-1}\left[\left(\text{erf}^{-1}(2\phi - 1)\right)^2 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right] \tag{3.32}$$

This equation will return a maximum MDT, having applied a worst case variability to the calculation of the MDT.

Thus, to obtain the WCV estimate for the MDT, the only parameters we require are the arithmetic mean $\bar{x}_0$ and the control parameters ϕ and Ψ, as well as the depuration decay rate b. As the current testing protocols provide arithmetic mean values from homogenates of 10 shellfish, and the control parameters can be set exogenously by authority bodies, then MDT's can be calculated with little or minimal change to current test methodologies. Determining a value for the decay rate of the process would need be determined for each depuration facility, based upon further experimental data or an assessement of the criteria listed in Section 1.7.1.

3.4.1 *Pre-Depuration Data*

The data used to inform our selection of a lognormal distribution in Section 3.3.1 was obtained from an experiment where six hundred oysters had been artificially seeded with high levels of NoV. The mean values for the two genotypes from this bioaccumulated data are presumed to be much higher than that which would be found from standard shellfish harvest sites, and would be especially greater than expected from any class B site harvests. In the absence of at harvest/pre-depuration data, parameter values for $\bar{x}_0$ for our model must be sourced from the literature.

SITE CLASSIFICATION	Jul '09	Jan '10	Jul '10	Jan '11
B - $\bar{x}_0$	40	1062	38	1064
C - $\bar{x}_0$	< 100 [a]	13272	< 40 [b]	15369

Table 3.1: NoV load calculated means for class B and C sites at low and high temperature points throughout study duration, measured in cpg

[a] NoV loads recorded as < 40 cpg are designated as having value = 20 cpg (the midpoint between 0 and 40) [b] while < 100 cpg are quantified as 70 cpg (the midpoint between 40 and 100) [2]

Lowther *et al* analysed NoV loads in oysters (*C. gigas*) from 39 UK harvest sites, with samples collected each month over a two year period in 2011-12 [2]. The 39 sites were comprised of 6 class A, 31 class B and 2 class C sites from around mainland Britain. Class A sites exhibited low NoV loads, Class C sites must use relaying rather than depuration to reduce contaminants, so only the

class B data were analysed to obtain estimates of arithmetic mean values ($\bar{x}_0$) of NoV. Appendix C contains full details on analysis of the data set, with a summary of the values obtained from the data shown in Table 3.1.

The observations recorded in winter months exhibited higher NoV levels, with Class B Jan '10 average $\bar{x}_0 = 1062$ NoV cpg, and Jan '11 $\bar{x}_0 = 1064$ NoV cpg. As only class B harvests are legally allowed to proceed to depuration, we will use the winter 2011 month value as the parameter for our model. As no estimate of σ_0 is available from harvest data, we base our results on WCV, so Equations 3.31 and 3.28 are used throughout the results.

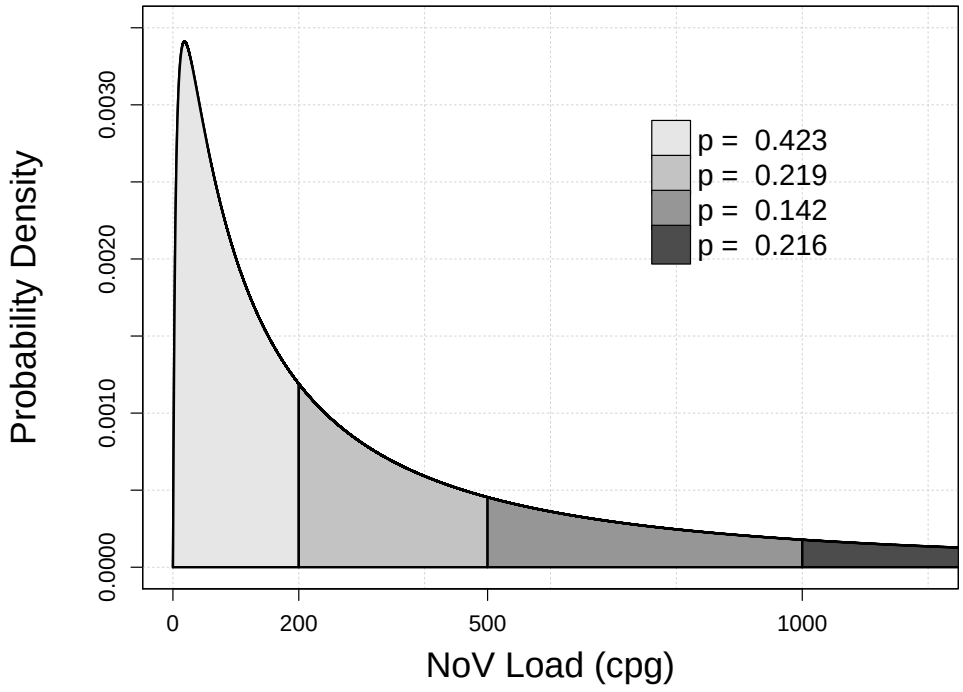


Figure 3.7: Plot of pre-Depuration probability distribution $P(x_0)$, with $\bar{x} = 1064$ NoV cpg (Class B, Jan '11) and $\sigma_0 = 1.645$. This variability corresponds to the worst case scenario (Equation 3.31) with assurance level set as $\phi = 95\%$

The pre-depuration distribution, $P(x_0)$, of Class B Jan '11 sampled oysters ($\bar{x}_0 = 1064$ NoV cpg) is shown in Figure 3.7. Assuming that $\phi = 0.95$ and using $\bar{x}_0 = 1064$ NoV cpg yields $\sigma_0 = 1.645$ under the WCV approach (Equation 3.31). The distribution was assessed for the probability that a randomly sample oyster's NoV load fell within a certain range of values, and are shown as probability values in Figure 3.7. The range values have been set at 200, 500, 1000 NoV cpg, and are arbitrary. However, in the next section, we parameterise $\Psi = 200$ NoV cpg from the literature, and so choose to split the PDF at this value, and others close to Ψ.

For these pre-depuration parameters, it was calculated that there was only a 42.3% likelihood that a randomly sampled oyster would have a NoV load below $\Psi = 200$ NoV cpg. Thus before being depurated 57.7% of a batch of Class B oysters with an average NoV load of $\bar{x}_0 = 1064$ cpg would have NoV loads above 200 cpg. Also note that there is a 21.6% probability that an oyster before depuration would have a NoV load greater than 1000 cpg.

3.4.2 *Model Parameterisation*

3.4.2.1 *Depuration Decay Rate Parameterisation*

	Pre-Dep.	During Dep.	Post-Dep.
NoV cpg	$\bar{x}_0 = 492$	$\bar{x}_{96} = 136$	$\bar{x}_{144} = 99$ (max.)
Total duration (hrs)	0	96	144
Decay rate, b	—	0.01339	0.01113

Table 3.2: Depuration decay rates derived from data in Doré *et al* (2010) [7, p. 2].

The literature can be used to derive further parameters for the model. In 2010, Doré *et al* [7] carried out a NoV survey of an Irish farm which had recognised they were selling oysters with greater than expected NoV levels. The farm voluntarily applied additional NoV mitigation methods to reduce any potential risk to their consumers. In the published report, the authors provide pre-depuration and during depuration data on NoV levels detected in oysters using PCR assay (Table 3.2).

Based on the exponential decay equation (cf. Equation 3.19)

$$\bar{x}_t = \bar{x}_0 \exp\{-bt\}\,, \tag{3.33}$$

the depuration decay parameter b can be established for 96 hours and 144 hours. For example, the depuration decay rate for 0–96 hours is derived using Equation 3.19:

$$\begin{aligned}
\bar{x}_t &= \bar{x}_0 \exp\{-bt\} \\
\Rightarrow \bar{x}_{96} &= \bar{x}_0 \exp\{-96\ b\} \\
\Rightarrow 136 &= 492 \exp\{-96\ b\} \\
\Rightarrow -96\ b &= \ln\left\{\frac{136}{492}\right\} \\
\Rightarrow b &= \ln\left\{\frac{136}{492}\right\} \div -96 \\
\Rightarrow b &= 0.01339 \quad \text{(to 5 d.p.)}\,.
\end{aligned}$$

This value is consistent with other values in the literature with regards to depuration decay rates.

Polo *et al* state several depuration decay rates of pathogens within shellfish due to depuration. They provide an average reduction rate of 0.56 day^{-1} of murine norovirus within Manila clams, equivalent to 0.0233 hours^{-1}; and a depuration decay rate equivalent to 0.0471 hours^{-1} for Mediterranean mussels

[62]. In their later paper, Polo *et al* provide equivalent decay rates of 0.0454 hours^{-1} for Hepatitis A virus in mussels, and 0.0179 in clams for the same virus [63]. Even though these are not exact comparisons to NoV within oysters, they provide a level of confirmation in our approach to the parameterisation of the decay rate shown above.

3.4.2.2 *Pathogen Load Limit Ψ Parameterisation*

The same report also states "Since 19 March 2010 more than 50,000 oysters have been placed on the market and no reports of illness have been received. NoV levels in these batches were less than 200 viral genome copies per g" [7]. This can be used to inform our selection of a value for the NoV load limit, and so we set our parameter $\Psi = 200$ cpg.

3.4.3 *NoV Dynamics During Depuration*

The dynamics of the Class B Jan '11 data are analysed using the model, applying the depuration decay rate derived from Doré (2010) ($b = 0.01339$ hours^{-1}) [7] for a reasonable depuration time period (0 — 100 hours). Thus the exponential decay of NoV loads in individual oysters induces the dynamics of the population during the process, and is characterised by $P(x_t)$ and shown in Fig. 3.8. The peak of the highly-skewed distribution moves towards lower values of x_t, creating a distribution that is increasingly skewed and with greater kurtosis.

Setting the NoV threshold limit at $\Psi = 200$ NoV cpg, the model calculates that, after $t = 50$ hours, 58.4% are below Ψ; after $t = 100$ hours, this increases to 73.2%.

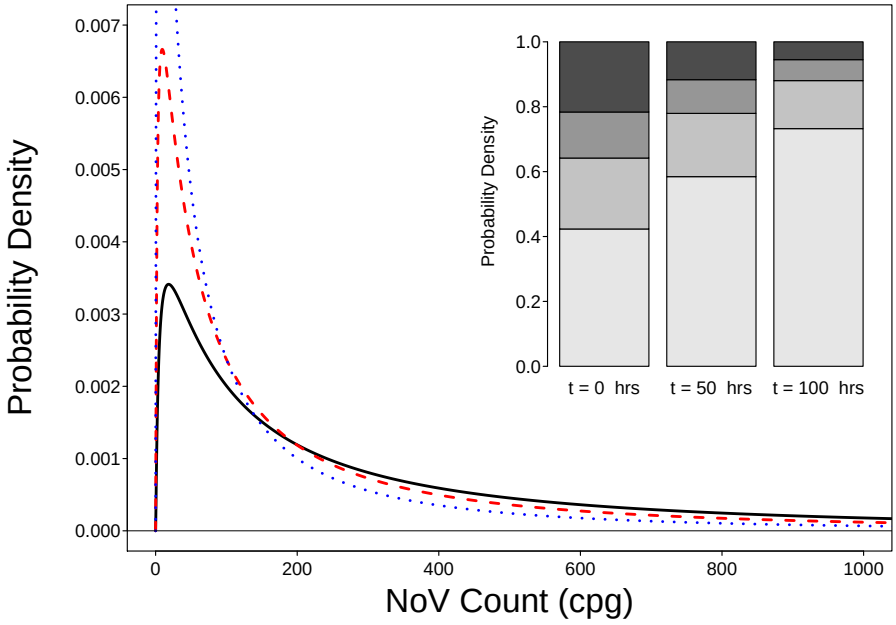


Figure 3.8: $P(x_t)$ dynamics during depuration, with decay rate $b = 0.01339$. Main plot shows probability distributions at $t = 0$ hrs (——), $t = 50$ hrs (– – –), $t = 100$ hrs (· · ·), using the same parameters and threshold values as Fig. 3.7. Inset bar plot shows the respective changes in section probabilities for each time point, with different shade bars representing values of x_t up to 200 (light grey), 500 (darker shade), 1000 (dark grey), and above 1000 (black)

3.4.4 *Minimum Depuration Time*

Equation 3.32 is used to calculate minimum depuration times for Class B 'Jan 11 data ($\bar{x}_0 = 1064$ NoV cpg). Applying $\phi = 90\%$, 95% and 99% assurance levels, MDTs were achieved after 186, 226, and 327 hours respectively. Note that the WCV approach was used here, and so these are maximum estimates of the MDTs. Figure 3.9 shows how MDTs change when the NoV load limit is in the range $0 < \Psi \leqslant 16000$ cpg. As expected MDTs are seen to be short when Ψ is very high. Increasing the load limit reassigns more of the tail of the distribution into the body, and so less depuration time would be required to

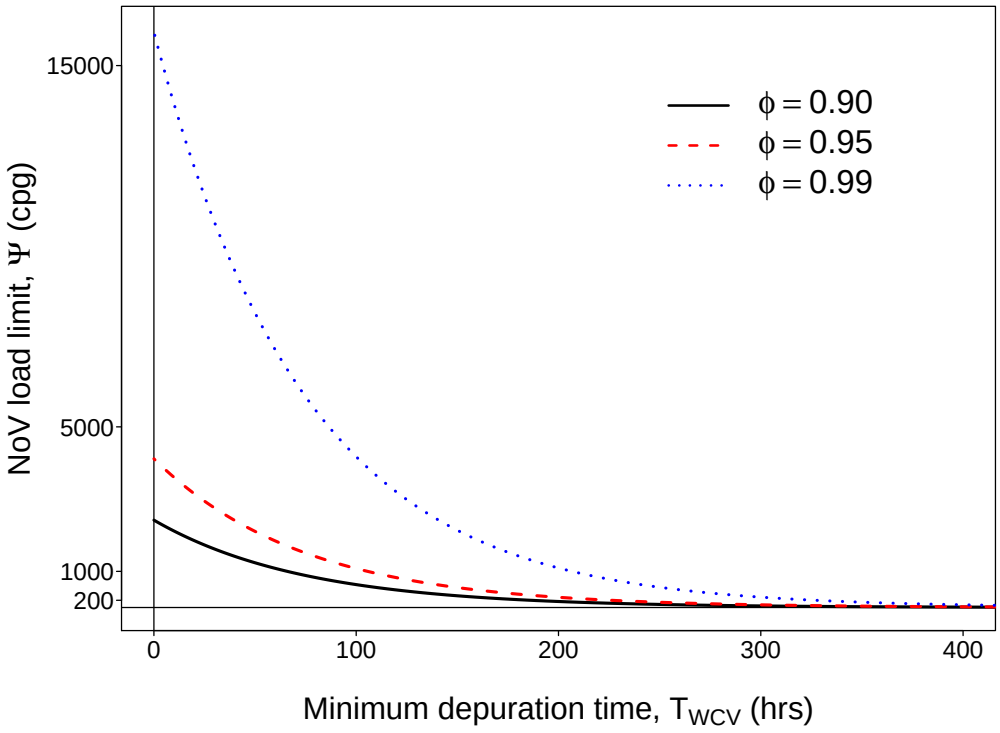


Figure 3.9: Plot of MDT versus NoV load level Ψ. MDTs with assurance level $\phi = 0.90$ plotted with (——) ; MDTs with assurance level $\phi = 0.95$ by (– – –); and $\phi = 0.99$ by (· · ·). Note that the response variable is on the horizontal axis and the independent parameter on the vertical axis, which is due to ease of display

achieve the desired assurance level ϕ. It can also be seen that as $\phi \to 1$, MDTs also increase. As the inverse error function has an asymptote at 1, increasing ϕ beyond 0.99 will result in greatly increased MDTs.

Figure 3.10 shows the same information except on a log-Ψ scale. The dynamics are shown to be log-linear, so confirming that the decay in Figure 3.9 is exponential. A line at $\Psi = 200$ cpg is shown, with the perpendicular lines highlighting the MDTs for each value of ϕ stated above.

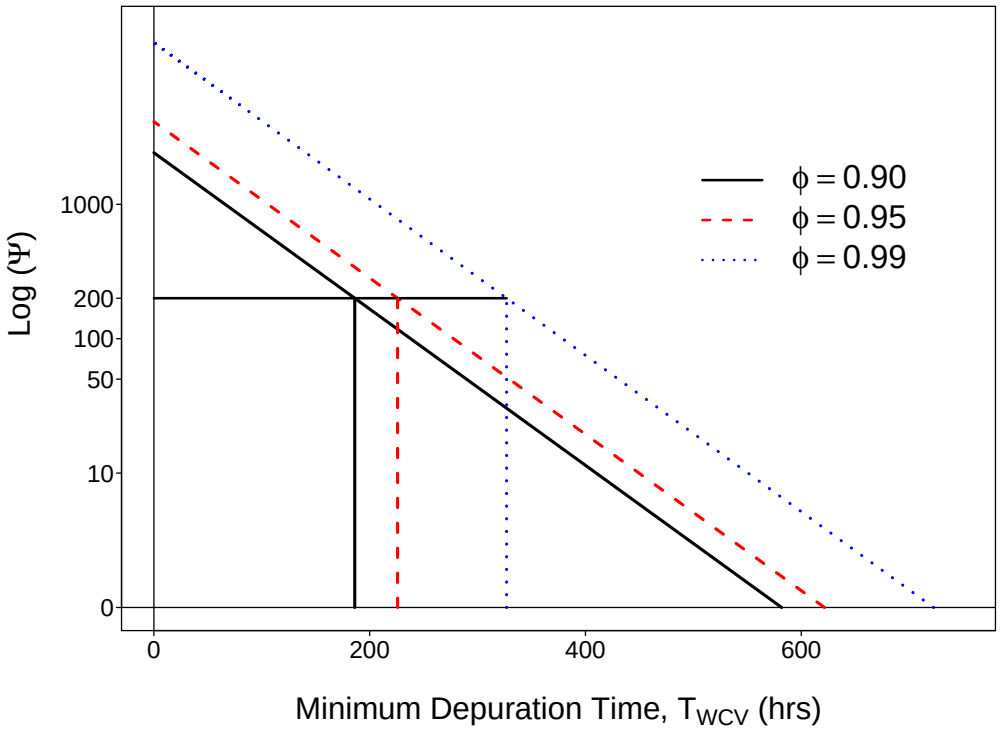


Figure 3.10: Plot of minimum depuration versus NoV load level Ψ on a log-Ψ scale

3.4.4.1 *Sensitivity Analysis: Depuration Decay rate*

The MDTs obtained from Equation 3.32 were then analysed by increasing the depuration decay rate. The depuration rate obtained from Doré (2010) [7] was used as a baseline, and b+10%, b+25%, b+50%, b+100% were used to calculate minimum depuration times for each increased value of b. The values obtained shown in Table 3.3 demonstrate that increasing the depuration rate has a significant impact upon the MDT. These results are a consequence of Equation 3.32, and show that the minimum depuration time is reduced by a factor of $\rho/(1+\rho)$ if depuration efficacy is altered by a proportion ρ. Increasing depuration efficacy by 10% provided a 9.09% decrease in depuration time; $b + 25\%$ decreased time required by 20.01%; $b + 50\%$ decreased time required by 33.30%; and doubling depuration efficacy halved the required depuration time. It is worth noting here that if depuration efficacy falls by (e.g.) 25%, then

minimum depuration times would be increased by 33.33%.

	$\phi = 90\%$	$\phi = 95\%$	$\phi = 99\%$
$b = 0.01339$	186	226	327
$b + 10\%$	169	205	297
$b + 25\%$	149	181	262
$b + 50\%$	124	151	218
$b + 100\%$	93	113	163
Test pass %	96	98	99

Table 3.3: Impact of changes in depuration efficiency on minimum depuration times (hrs) and simulated quantitative NoV tests of ten oyster homogenates, which had undergone depuration using each parameter set (ϕ, b, T_{WCV})

3.4.4.2 *Simulation of Current Testing Methods*

The model was run again using the Class B Jan '11 data ($\bar{x}_0 = 1064$ NoV cpg) and with a parameter set of (ϕ, b, T_{WCV}) as per values in Table 3.3. For example, a parameter set of ($\phi = 0.90$, $b + 10\%$, MDT=169) was used in conjunction with other parameters fixed at $\bar{x}_0 = 1064$ cpg and $\Psi = 200$ cpg.

Measures of spread and location for each parameter set's distribution after each specific MDT were obtained, and 10 random variates were selected from each distribution and averages calculated. This method synthesised the current NoV test protocols using homogenates of 10 oysters [15]. Each simulated test was deemed a 'pass' if it's average deviate value was less than $\Psi = 200$ NoV cpg. The average pass rate shown was calculated from 10 000 iterations of each parameter set. 10,000 iterations of each parameter set were run, and the proportion of test passes are shown at the bottom of Table 3.3. For each value of ϕ the

trade-off between minimum depuration time and b results in approximately the test same pass rate. This is due to the fact that the measure of location, μ_t, will be fixed by varying both b and the MDT. For each column in Table 3.3, the value of μ_t will be (approximately) constant when calculated from the same base value of μ_0, and the pairwise values of b and MDT (cf. Equation 3.17).

Using assurance levels of $\phi = 90\%$, 95% and 99%, pass rates of approximately 96%, 98% and 99% were achieved for ten oyster homogenates. This is in contrast to pass rates for single oyster simulated testing, which returned pass rates equal to the value of ϕ being applied. By our definition (cf. Equation 3.24), when $\phi = 0.90$ the probability of sampling an oyster with NoV loads greater than Ψ will be 0.90. Within a sample of ten oysters, only one will be expected to have an pathogen load greater than the designated value. However, if the other nine oysters have loads much less than Ψ, the mean load obtained from the homogenised sample will likely have a load value less than Ψ also. Testing of the ten oyster homogenates thus can result in any oysters with NoV loads greater than Ψ often being masked within the homogenate sample.

With this in mind, the option of using $\phi = 90\%$ is not recommended as the pass rate is much greater than that of the actual assurance level. A higher rate of either $\phi = 95\%$ or $\phi = 99\%$ would be more appropriate to minimise the probability that a shellfish with a pathogen load significantly greater than that of Ψ would be masked by the homogenisation of the PCR assay. Ideally $\phi = 99\%$ would provide a much greater level of food safety; however, this value results in greatly increased lengths of MDT required to conform to both control parameters Ψ and ϕ. With $\phi = 95\%$, a high level of food safety is ensured, as well as a reduced length of MDT in comparison to when $\phi = 99\%$.

This chapter has proposed a model of the depuration process, and its impact on the pathogen levels across a population or batch of shellfish. We have parameterised the model based on data pertaining to NoV within oysters. This combination of pathogen and mollusc will most often result in the transmission of pathogen into the human population via shellfish ingestion due to two factors: the prevalence of raw consumption of oysters compared to other molluscs; and the preferential binding of NoV that occurs within an oyster's digestive system, thus slowing down this pathogen's expulsion due to depuration.

However, the model can be applied for any mollusc/pathogen combination, assuming that the initial distribution is well approximated by lognormality. To accommodate different combinations of shellfish and pathogen, a reassessment of the value of the depuration decay rate b must be considered as well as changes to the values of the control parameters Ψ and ϕ. For example, evidence shows that *E. coli* is excreted much quicker by shellfish and so would require a larger value of b in comparison to NoV. Indeed, this is the case as the minimum 42 hour depuration time for Class B harvests is based on *E. coli*, and is deemed to be more than sufficient from a food safety standpoint to reduce any levels of *E. coli* to negligible levels [71]. In 2016 the Food Standards Agency relaxed the minimum 42 hour depuration time for depuration provided that the operators could justify the reduction [72]. The implications and justifications of this are not covered by the scope of this thesis.

It is apparent from the paucity of available data that further experimentation is required to better parameterise this model, as well as to test and validate

	Genotype I	Genotype II	Model Parameter
μ_0 (location)	7.645	8.628	5.617
σ_0 (scale)	1.787	1.257	1.645
$\bar{x}_0$ (arithmetic mean)	10318	12308	1064

Table 3.4: Genotypes I & II location and scale parameters for lognormal distribution calculated from bioaccumulation data in Section 3.3.1

the assumptions used in its construction. Of most note is the need to quantify the variability of a pathogen across a batch of oysters. Due to the lack of any pre-depuration or harvest field data that can be used to calculate variability, we have applied our concept of a “worst case variability". The value of the WCV in this model is an increasing, monotonic function calculated solely on the value of ϕ, which is the NoV assurance level either enforced on or adopted by the depurator (cf. Section 3.3.8).

Using the WCV approach, the model showed that significantly longer depuration times than the minimum 42 hours are required to achieve the levels set out by the control parameters as shown in Table 3.3. Depuration decay rates ranged from 96—327 hours, depending upon which assurance level was adopted and the depuration decay rate used. Previous results from experimentation in the literature have stated that 120 hours(5 days) of depuration would be required to remove quantifiable levels of human viruses from shellfish [16].

This is in line with the range of our results. As our results are based on a worst case approach to variability, then the MDT results shown are at the upper end of depuration times that may be required to provide adequate levels of food safety. A more accurate measure of σ_0 of NoV across harvested batches of shellfish would be expected to reduce the variability and so the depuration times.

The bioaccumulated data returned spread parameters of $\sigma_0 = 1.787$ for genotype I and $\sigma_0 = 1.257$ for genotype II (see Table 3.4), sandwiching the value of $\sigma_0 = 1.645$ obtained from our WCV approach. Note that the measure of location parameter used in our results is significantly less than that which can be obtained from the bioaccumulation data discussed in Section 3.3.1. As previously stated, the bioaccumulation data was intentionally seeded with high levels of NoV, and so would not be a useful basis to provide results pertaining to actual batches of harvested shellfish from coastal or farm waters.

This chapter goes further than previous work by extending the model of depuration across a population of shellfish rather than to describe an individual animal. This model also incorporates control parameters which would allow legislators to enforce legislative limits on the levels of NoV (and other pathogens). Thus the model provides a mathematical framework that could be used by industry and regulatory bodies to help determine the minimum depuration times required to reduce NoV levels to below a desired threshold.

The crux of this model has been the use of lognormality to describe the dispersal of a water-borne pathogen across a shellfish population. We have used a lognormal distribution as the basis for our model, informed by experimental data and the literature, and where most of the samples were close in

value to zero cpg. The tail of the distribution (as we defined it in Section 3.3.6) is the concern from a food safety perspective, and the weight of this tail is a consequence of the variability across the lognormal distribution of pathogens.

Comparison between the lognormal distribution and another distribution type is carried out in the next chapter.

4

LOGNORMAL VERSUS TRUNCATED NORMAL DISTRIBUTION

4.1 INTRODUCTION

Although the lognormal distribution is the most widely accepted type used to describe pathogen loads within shellfish populations, the paucity of available data does not preclude that other distributions may also be valid. Therefore this chapter considers another distribution type: the truncated normal distribution. We apply the same or similar model construction techniques (as used in Chapter 3) to a truncated normal distribution, and compare the results between it and the results obtained using the lognormal distribution in the previous Chapter.

The truncated normal distribution has been chosen as it meets the specific criteria to describe the variability of water-borne pathogens:

(1) all variates must be non-negative;

(2) distribution exhibits a strong, positive skewness;

(3) continuous domain with no upper bound.

The truncated normal distribution is a variant of the normal distribution, with the difference being that (as the name suggests) the domain of the truncated normal's PDF is restricted at either a lower or upper bound, or both. For our modelling needs, we truncate the normal distribution (fully described in

Section 2.3.1) at a lower bound of zero, thus fulfilling the three requirements listed above.

The truncated normal distribution can be configured as one of three different distribution types:

(1) lower truncated normal (single truncation);

(2) upper truncated normal (single truncation);

(3) upper and lower truncated normal (double truncation).

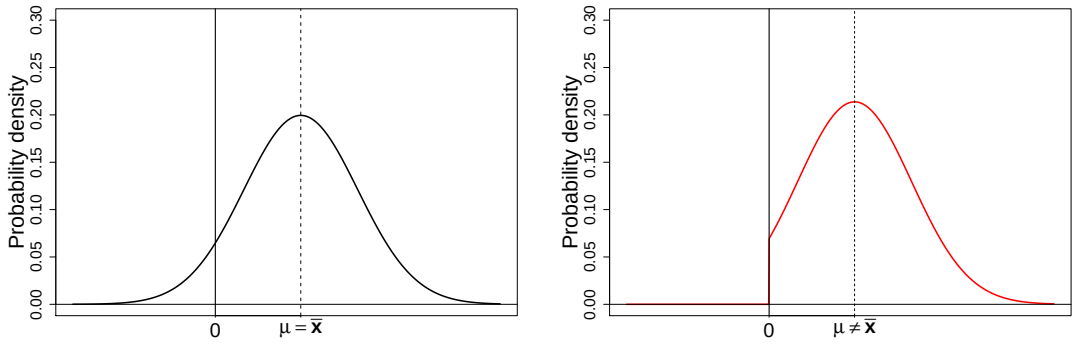
We are interested in the lower truncated normal type, and this is shown in Figure 4.1, providing a comparison of the shapes of a normal and a lower truncated normal distribution[1].

Figure 4.1a) shows a normal distribution with mean (and median and mode) located at $x = 3$, with standard deviation $\sigma = 2$. This distribution evidently does not meet all three of the requirements to describe water-borne pathogen variability, as part of its domain is in $-\infty < x < 0$. This will be the case no matter the value of the parameters μ and σ (> 0).

However we can be restrict the PDF of the normal distribution to a specific domain on $\mathbb{R}$ in one of the three ways mentioned above using either lower, upper or both types of domain truncation. Figures 4.1b shows how the shape of the density function changes when the distribution is truncated at a lower ($x = 0$) value [2].

1 These plots have been generated using the 'msm' package in "R" software (see Section A.1.2 in Appendix A for the code)

2 The selected lower bound could also have been set at 10, the limit of quantitation of NoV cpg for PCR



(a) Normal: $\bar{x} = \mu = 3$, $sd = \sigma = 2$ (b) Truncated lower at $x = 0$: $\mu = 3$, $sd = \sigma = 2$

Figure 4.1: Plots of normal and truncated normal probability density functions. Plot (a) shows a normal distribution with mean (and median and mode) located at $\bar{x} = \mu = 3$, and standard deviation $\sigma = 2$. Plot (b) shows a truncated normal distribution with the same values of $\mu = 3$ and $\sigma = 2$ with a lower truncation at $x = 0$

As the total area under the curve of the PDF must be equal to 1 to satisfy the conditions listed in Section 2.3.1.1, the area of Figure 4.1a in the range $[-\infty, 0)$ (i.e. the area that has been sliced off by the truncation) is proportionally redistributed across the curve in Figure 4.1b in $[0, \infty]$.

The peak of the distribution is maintained at the same location before and after any truncation, i.e. the mode of both the normal and truncated normal PDFs are located at $x = \mu$, the location parameter of both distribution types. However, where the mean of the normal distribution was also found at $x = \mu$, this will not be the case for the truncated distribution.

4.2.1 *Truncated Normal Distribution Definition*

Letting $P_T(x)$ denote a truncated normal distribution, Johnson & Kotz state that:

$$P_T(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \left[\frac{1}{\sigma\sqrt{2\pi}} \int_a^b \exp\left\{\frac{-(t-\mu)^2}{2\sigma^2}\right\} dt\right]^{-1}$$

$$\Rightarrow P_T(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} \left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]^{-1} \quad (4.1)$$

applicable when $a < x < b$, where a and b are the respective lower and upper truncation points of the distribution; otherwise $P_T(x) = 0$ [58, p. 81]. The function Φ (cf. Equation 2.3.1.2) is defined as

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \quad (4.2)$$

It can be shown that when $b = \infty$

$$\Phi\left(\frac{b-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right)\right] = 1, \quad (4.3)$$

and when $a = -\infty$

$$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)\right] = 0. \quad (4.4)$$

Thus, when $a = -\infty$, $b = \infty$ Equation 4.1 would be equivalent to the general normal distribution.

The distribution $P_T(x)$ has location parameter μ and spread parameter σ. However, unlike the normal distribution, $E(x) \neq \mu$ and $V(x) \neq \sigma^2$. Definitions for the mean and variance are derived in Section 4.3.2.1 below.

4.2.2 *Truncator Constant* — T

From Equation 4.1, we can define T as the truncator term of a truncated normal distribution, where

$$T = \left[\Phi\left(\frac{b-\mu}{\sigma} \right) - \Phi\left(\frac{a-\mu}{\sigma} \right) \right] . \tag{4.5}$$

This allows us to rewrite the PDF of the truncated normal distribution as

$$\Rightarrow P_T(x) = \begin{cases} \frac{1}{T\sigma\sqrt{2\pi}} \exp\left\{ \frac{-(x-\mu)^2}{2\sigma^2} \right\} & \text{when } \ a < x < b \\ 0 & \text{otherwise} . \end{cases} \tag{4.6}$$

The value of T is constant for a truncated distribution, and now can be seen as a component of the normalisation coefficient for a truncated normal distribution. Referring back to the discussion of Figure 4.1 above, we can now state that the specific role of T is to redistribute the area of the normal distribution (defined by $X \sim \mathcal{N}(\mu, \sigma^2)$) which is outside the domain of the truncation across the domain $a < x < b$.

Adopting a similar approach to the construction of our lognormal model as carried out in Section 3.3.2, we assume that the pre-depuration distribution of NoV loads within a population of shellfish can be reasonably described by a truncated normal distribution, with a lower truncation point equal to zero which restricts the domain of the PDF to non-negative values.

4.3.1 *Pre-Depuration Truncated Normal Model*

We assume the distribution of NoV across a population of shellfish is well described by a truncated normal population, with a lower truncation point at zero, so satisfying the third condition listed in Section 4.1. Thus we define the probability $X_0 = x_0$ of a shellfish bearing a NoV load before depuration as $P_T(X_0 = x_0)$.

First, we carry forward Equation 4.3 to adapt Equation 4.5, allowing us to define the truncator term T for $P_T(x_0)$ where $a = 0$ and $b = \infty$:

$$1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{0 - \mu_0}{\sqrt{2}\sigma_0}\right)\right] = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{-\mu_0}{\sqrt{2}\sigma_0}\right)\right] = T\,. \tag{4.7}$$

Combining Equations 4.6 and 4.7, the PDF of $P_T(x_0)$ is defined as

$$P_T(x_0) = \frac{1}{T\sigma_0\sqrt{2\pi}} \exp\left\{\frac{-(x_0 - \mu_0)^2}{2\sigma_0^2}\right\}, \tag{4.8}$$

where $x_0 \geqslant 0$.

4.3.2 *Evolution Of The Depuration Distribution Over Time*

As carried out in the previous chapter, we model the effect of depuration on NoV levels within a population of shellfish, describing the dynamics of pathogens during depuration by the truncated normal PDF $P_T(x_t)$, $\forall t \geqslant 0$.

We assume the same exponential decay across time t of individual shellfish NoV loads as according to Equation 3.11 in Section 3.3.4:

$$x_t = x_0 \ \exp\{-bt\}, \tag{4.9}$$

where b is the depuration decay rate for the specific depuration processes being employed. This allows us to develop the PDF for $P_T(x_t)$ using the same change of variables process as used in Section 3.3.4.

Again we apply the change of variables for Equation 4.8, shifting from variable x_0 to x_t as the variate of the distribution. This is achieved by simplification of the following equation (cf. Equation 3.12):

$$\begin{aligned} P_T(x_t) &= P_T(x_0)\frac{dx_0}{dx_t} \\ \Rightarrow P_T(x_t) &= \frac{\exp\{bt\}}{T\sigma_0\sqrt{2\pi}} \exp\left\{\frac{-(x_t e^{bt} - \mu_0)^2}{2\sigma_0^2}\right\} \end{aligned} \tag{4.10}$$

As exponential decay of NoV loads occurs across the whole population, then the location and scale parameters of the PDF will also decay exponentially. Thus

$$\mu_t = \mu_0 e^{-bt}, \tag{4.11}$$

$$\sigma_t = \sigma_0 e^{-bt}. \tag{4.12}$$

This allows Equation 4.10 to be simplified further:

$$P_T(x_t) = \frac{e^{bt}}{T\sigma_t e^{bt}\sqrt{2\pi}} \exp\left\{\frac{(x_t e^{bt} - \mu_t e^{bt})^2}{2\sigma_t^2 e^{2bt}}\right\}$$

$$\Rightarrow P_T(x_t) = \frac{1}{T\sigma_t\sqrt{2\pi}} \exp\left\{\frac{(x_t - \mu_t)^2}{2\sigma_t^2}\right\} \tag{4.13}$$

When $x_0 \geqslant 0$, so also will $x_t \geqslant 0$; therefore the resulting distribution for x_t will also be of a truncated normal form.

The truncator term T will be impacted by the exponential decay of the system. For $t \geqslant 0$, T can be rewritten as

$$T = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{-\mu_t e^{bt}}{\sqrt{2}\sigma_t e^{bt}}\right)\right]$$

$$T = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{-\mu_t}{\sqrt{2}\sigma_t}\right)\right], \tag{4.14}$$

as $\mu_t = \mu_0 e^{-bt}$ and $\sigma_t = \sigma_0 e^{-bt}$ (Equations 4.11 and 4.12). It is apparent that the value of the truncator T will remain constant, as the ratio of $-\mu_t/\sqrt{2}\sigma_t$ will remain constant $\forall t \geqslant 0$.

4.3.2.1 *Expected Value of Truncated Normal Distribution*

We have stated that the value of the location and scale parameters will decay in accordance with the depuration decay parameter b. Consideration of how the mean or expected value of the distribution behaves over time will also be useful in examining the behaviour of the distribution.

The expected value of x of a truncated normal distribution $P_T(x)$ is defined as

$$E(x) = \mu + \sigma\frac{Z\left(\frac{a-\mu}{\sigma}\right) - Z\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \tag{4.15}$$

where

$$Z(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\frac{-\alpha^2}{2}, \tag{4.16}$$

μ and σ are the location and shape parameters of $P_T(x)$, a and b are the lower and upper bounds of any truncation, and Φ is as defined earlier by Equation 4.2 [58, p. 81].

We have set our lower truncation when $a = 0$ and, as previously stated, when $b = \infty$

$$\Phi\left(\frac{b-\mu}{\sigma}\right) = 1\,.$$

We also note that for $b = \infty$,

$$Z\left(\frac{b-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \exp \frac{-(b-\mu)^2}{2\sigma^2} = 0\,. \tag{4.17}$$

Incorporating these into Equation 4.15, the expected value of the pre-depuration distribution $P_T(x_0)$ can be stated as

$$E_T(x_0) = \mu_0 + \sigma_0 \frac{Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} \tag{4.18}$$

$$\Rightarrow E_T(x_0) = \mu_0 + \sigma_0 \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_0)^2}{2\sigma_0^2}\right\}}{1 - \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{-\mu_0}{\sqrt{2}\sigma_0}\right)\right)}$$

$$\Rightarrow E_T(x_0) = \mu_0 + \sigma_0 \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_0)^2}{2\sigma_0^2}\right\}}{\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\mu_0}{\sqrt{2}\sigma_0}\right)\right)}\,. \tag{4.19}$$

Substituting in $\mu_0 = \mu_t e^{bt}$ and $\sigma_0 = \sigma_t e^{bt}$ derived from Equations 4.11 and 4.12 where appropriate and simplifying:

$$E_T(x_0) = \mu_t e^{bt} + \sigma_t e^{bt} \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_t e^{bt})^2}{2(\sigma_t e^{bt})^2}\right\}}{\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\mu_t e^{bt}}{\sqrt{2}\sigma_t e^{bt}}\right)\right)}$$

$$\Rightarrow E_T(x_0) = e^{bt}\left[\mu_t + \sigma_t \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_t)^2}{2(\sigma_t)^2}\right\}}{\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\mu_t}{\sqrt{2}\sigma_t}\right)\right)}\right]$$

$$\Rightarrow E_T(x_0)e^{-bt} = \mu_t + \sigma_t \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_t)^2}{2(\sigma_t)^2}\right\}}{\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\mu_t}{\sqrt{2}\sigma_t}\right)\right)} \tag{4.20}$$

The righthand side of Equation 4.20 now describes the expected value of $P_T(x_t)$, therefore

$$E_T(x_t) = E_T(x_0)e^{-bt}. \tag{4.21}$$

Thus the expected value of our truncated normal distribution will also decay in accordance with our depuration rate b.

4.3.2.2 *Variance of Truncated Normal Distribution*

Similar methods are applied to obtain the variance (and standard deviation) of $P_T(x_t)$. Johnson and Kotz provide a definition of the variance of a truncated normal distribution [58, p. 83]. Beginning with the pre-depuration distribution, the variance of $P_T(x_0)$ is defined as

$$V_T(x_0) = \sigma_0^2 \left[1 + \frac{\left(\frac{a-\mu_0}{\sigma_0}\right) Z\left(\frac{a-\mu_0}{\sigma_0}\right) - \left(\frac{b-\mu_0}{\sigma_0}\right) Z\left(\frac{b-\mu_0}{\sigma_0}\right)}{\Phi\left(\frac{b-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{a-\mu_0}{\sigma_0}\right)} - \left\{ \frac{Z\left(\frac{a-\mu_0}{\sigma_0}\right) - Z\left(\frac{b-\mu_0}{\sigma_0}\right)}{\Phi\left(\frac{b-\mu_0}{\sigma_0}\right) - \Phi\left(\frac{a-\mu_0}{\sigma_0}\right)} \right\}^2 \right]. \tag{4.22}$$

Substituting in Equations 4.3, 4.17 and $a = 0$ results in

$$V_T(x_0) = \sigma_0^2 \left[1 + \frac{\left(\frac{-\mu_0}{\sigma_0}\right) Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} - \left\{ \frac{Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} \right\}^2 \right]. \tag{4.23}$$

Previously we have stated that $\mu_0 = \mu_t e^{bt}$ and $\sigma_0 = \sigma_t e^{bt}$ (cf. Equations 4.11, 4.12). Thus we can state that

$$\frac{-\mu_0}{\sigma_0} = \frac{-\mu_t}{\sigma_t}, \tag{4.24}$$

and substitute this result into Equation 4.23 (along with $\sigma_t = \sigma_0 e^{-bt}$). This provides a definition of the variance of $P_T(x_t)$, where

$$\begin{aligned} V_T(x_t) &= \sigma_t^2 \left[1 + \frac{\left(\frac{-\mu_t}{\sigma_t}\right) Z\left(\frac{-\mu_t}{\sigma_t}\right)}{1 - \Phi\left(\frac{-\mu_t}{\sigma_t}\right)} - \left\{ \frac{Z\left(\frac{-\mu_t}{\sigma_t}\right)}{1 - \Phi\left(\frac{-\mu_t}{\sigma_t}\right)} \right\}^2 \right] \\ \Rightarrow V_T(x_t) &= \sigma_0^2 e^{-2bt} \left[1 + \frac{\left(\frac{-\mu_0}{\sigma_0}\right) Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} - \left\{ \frac{Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} \right\}^2 \right]. \end{aligned} \tag{4.25}$$

4.3.2.3 *Standard Deviation of Truncated Normal Distribution*

Generally speaking, standard deviation is often defined as $SD(x) = \sqrt{V(x)}$. Thus we can also state that, from Equation 4.23, the standard deviation of the pre-depuration distribution is

$$SD_T(x_0) = \sigma_0 \sqrt{\left[1 + \frac{\left(\frac{-\mu_0}{\sigma_0}\right) Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} - \left\{\frac{Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)}\right\}^2\right]}. \tag{4.26}$$

From Equation 4.25, we can also state that

$$SD_T(x_t) = \sigma_0 e^{-bt} \sqrt{\left[1 + \frac{\left(\frac{-\mu_0}{\sigma_0}\right) Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)} - \left\{\frac{Z\left(\frac{-\mu_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-\mu_0}{\sigma_0}\right)}\right\}^2\right]}$$

$$\Rightarrow SD_T(x_t) = SD_T(x_0) e^{-bt}, \tag{4.27}$$

Thus the standard deviation of the truncated normal model of depuration also decays according to the depuration rate parameter b.

4.4.1 *Inclusion of Control Parameters*

As with the lognormal model in Chapter 3, we must incorporate two additional parameters to our truncated normal model. Section 3.3.6 introduced and defined these as:

- Ψ — NoV load level — below which an individual shellfish is deemed to not constitute a risk to the consumer;
- ϕ — NoV assurance level — acceptable proportion of the shellfish population which has a NoV level below that of Ψ.

These two parameters implement the food safety requirements of legislators into our model.

Applying these two parameters to constrain a contaminant's level within a population of shellfish splits the population into two parts: the body of the population and the tail. Again following the processes used in Chapter 3, we define the body of the distribution as

$$P_T(x_t < \Psi) = \int_0^{\Psi} \frac{1}{T\sigma_t\sqrt{2\pi}} \exp\left\{\frac{(x_t - \mu_t)^2}{2\sigma_t^2}\right\} dx_t \leqslant \phi\,, \tag{4.28}$$

with the required tail defined by Equation 4.28's complement:

$$P_T(x_t > \Psi) = 1 - \int_0^{\Psi} \frac{1}{T\sigma_t\sqrt{2\pi}} \exp\left\{\frac{(x_t - \mu_t)^2}{2\sigma_t^2}\right\} dx_t \geqslant \phi\,. \tag{4.29}$$

This can be equivalently stated as

$$P_T(x_t > \Psi) = \int_{\Psi}^{\infty} \frac{1}{T\sigma_t\sqrt{2\pi}} \exp\left\{\frac{(x_t - \mu_t)^2}{2\sigma_t^2}\right\} dx_t \geqslant \phi\,.$$

4.4.2 *Minimum Depuration Time*

In Section 3.3.7, we were able to obtain an analytical solution to Equation 3.24 for t, allowing us to determine the minimum depuration time required to attain the assurance level ϕ. To obtain the minimum depuration time required to satisfy Equation 4.28 for the truncated normal model here, no analytical solution for t can be obtained. This is due to the multiple occurrences of time t within Equation 4.28, and the fact that they cannot be cancelled down to only one instance of t.

Therefore, numerical solutions must be employed to obtain results regarding minimum depuration times of the truncated normal model. Computer code which can be implemented in 'R' which obtains numerical solutions for MDT's from the truncated normal model is provided by Appendix A.1.4.

4.5 RESULTS

In this section we examine the dynamics of water-borne pathogens within a shellfish population using the truncated normal model developed above. The same parameters (where possible and appropriate) that were derived from the literature in Section 3.4 will be used to inform this model, thus allowing a more direct comparison of the dynamics of the lognormal and truncated normal distributions. Any different parameters that are required are obtained in the next section. We also differentiate between the truncated normal and lognormal PDFs by subscripts L and T. Thus $P_T(x_t)$ refers to the truncated normal distribution and $P_L(x_t)$ to the lognormal model's PDF.

4.5.1 *Parameterisation of Truncated Normal Model*

Only some of the parameters that were used in the lognormal model in Section 3.4 are appropriate for use with the truncated normal model. The values of the control parameters are set again at $\Psi = 200$ cpg and $\phi = 0.95$. These are not dependent upon the distribution type used as the basis for the model. We can also use the depuration decay rate of $b = 0.01339$ hr^{-1} can also be used from the literature [7], as well as $\bar{x}_0 = 1064$ cpg [2].

Thus, only values for the pre-depuration location (μ_0) and scale (σ_0) parameters of the truncated normal distribution need to be obtained. To carry out an "apples to apples" comparison between the lognormal model in Chapter 3 and the truncated model in this Chapter, we must use the same or similar parameters in both distribution types that inform their location and variability measures. However the values used in the lognormal model are not appropri-

ate here, as they are both defined specifically for the lognormal model and so we must find appropriate values for μ_0 and σ_0 by other methods.

4.5.1.1 *Scale Parameter* σ_0

We can parameterise σ_0 for the truncated normal model by assuming that the variability of both the lognormal and truncated normal models are the same and based upon our hypothesis of 'worst case variability' as described in Section 3.3.8. In that section we were able to estimate a worst possible variability parameter as stated in Equation 3.31:

$$\sigma_0 = \sqrt{2}\text{erf}^{-1}\left(2\phi - 1\right) \tag{4.30}$$

for the lognormal distribution. Assuming that $\phi = 0.95$, this results in a value of $\sigma_0 = 5.617$. This is the scale parameter for the pre-depuration *lognormal* distribution; therefore it is not appropriate for our truncated normal model which does not operate on a log scale.

However we can still use the worst case variability approach to calculate a value for σ_0. Re-defining the left hand side of Equation 4.30 as the lognormal scale parameter LN-$\sigma_0 = \sqrt{2}\text{erf}^{-1}\left(2\phi - 1\right)$, its result can be used (alongwith the value of $\bar{x}_0$) to obtain an unlogged value for the scale parameter. This is obtained from the equation

$$\begin{aligned} \sigma_0 &= \bar{x}_0\sqrt{\exp\left\{\text{LN-}\sigma_0^2 - 1\right\}} \\ \Rightarrow \sigma_0 &= \bar{x}_0\sqrt{\exp\left\{(\sqrt{2}\text{erf}^{-1}\left(2\phi - 1\right))^2 - 1\right\}} \end{aligned} \tag{4.31}$$

which itself is derived from the equation for the standard deviation of the lognormal distribution (Equation 3.3).

Thus the value of the scale parameter σ_0 can be calculated for the truncated normal distribution. Assuming a fixed value of $\bar{x}_0$ in our parameterisation, we

can obtain values for σ_0 based on the value of the NoV assurance level ϕ being applied. See Table 4.2 for values of σ_0 used in this section.

4.5.1.2 *Location Parameter* μ_0

Obtaining an appropriate value of μ_0 for our truncated normal model is more difficult than for our lognormal model. In Section 3.3.2 we stated the equation relating arithmetic mean to both the location and scale parameters for a lognormal distribution as

$$\bar{x}_0 = \exp\left\{\mu_0 + \frac{1}{2}\sigma_0^2\right\},$$

for the lognormal distribution. However this equation is not applicable for a truncated normal distribution.

Equation 4.19 above describes the pre-depuration relationship between arithmetic mean (expected value) and the location and scale parameters of our truncated normal distribution, and is reproduced below for ease of reference:

$$\bar{x}_0 = E_T(x_0) = \mu_0 + \sigma_0 \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\mu_0)^2}{2\sigma_0^2}\right\}}{\frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{-\mu_0}{\sqrt{2}\sigma_0}\right)\right)}. \tag{4.32}$$

We have values for both $\bar{x}_0$ and σ_0 but, due to the error function term on the denominator containing μ_0, we cannot obtain an analytical value solely for μ_0. Again we must turn to numerical approaches to obtain a value for the location parameter μ_0 that is appropriate for our model. The code executed in 'R' to achieve this is included in Section A.1.3 of Appendix A, and pairwise values for μ_0 and σ_0 are shown in Table 4.1 for differing values of the NoV assurance level ϕ, which drives the worst case variability.

It can be seen in Table 4.1 that, as the variability of the truncated normal distribution increases (due to increasing the value of ϕ), we must shift the value of μ_0 further down the number line to maintain an expected value of $E_T(x_0) = \bar{x}_0 = 1064$ cpg. If the value of μ_0 is outside the domain of the truncation(s) of the distribution, then the mode of the distribution will be

NoV assurance level, ϕ	Worst case variability, σ_0	Location parameter, μ_0
0.90	1467	-312
0.925	1819	-1315
0.95	2496	-3957
0.975	4405	-16209
0.985	5317	-24518

Table 4.1: Table of location and scale parameters applicable to truncated normal model, developed using worst case variability approach, to maintain an arithmetic mean value of $E_T(x_0) = \bar{x}_0 = 1064$ cpg. Coloured row indicates the parameters used as standard in the Results section below

located at the truncation value closest to μ_0. The value of μ_0 shown in Table 4.1 used in Figure 4.2a is negative and so outside the domain of the truncated normal distribution, therefore the mode is found at $x_0 = 0$ cpg. The negative value of μ_0 is a consequence of equation 4.32. Applying the value $\sigma_0 = 2496$ and $\bar{x}_0 = 1064$ cpg (Table 4.1) in Equation 4.32 provides a numercial solution value of $\mu_0 = -3957$ cpg. Therefore, if we imagine that a normal distribution with parameters $\mu_0 = -3957$ cpg and $\sigma_0 = 2496$ cpg had not been truncated at $x_0 = 0$ cpg, we would observe a symmetric, unimodal distribution centred at μ_0.

Parameter	b	$\bar{x}_0$	ϕ	μ_0	σ_0	Ψ
Value	0.01339	1064 [b]	0.95	-3957[c]	2496[c]	200

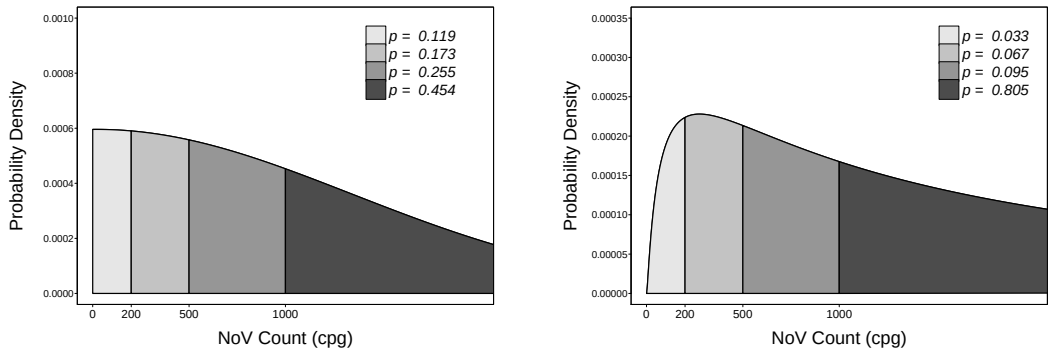
Table 4.2: Table of parameters applied to truncated normal model, derived from literature and numerical calculations.

[a] [7] [b] [2] [c] value based on $\phi = 0.95$

4.5.2 *Pre-Depuration: Truncated Normal versus Lognormal Model Comparison*

The shape and probability densities for specific NoV ranges of $P_T(x_0)$ (pre-depuration distribution using truncated normal), and $P_L(x_0)$ (pre-depuration distribution using lognormal) is shown in Figure 4.2a. The plot of $P_T(x_0)$ has been constructed using the pre-depuration parameters indicated in Table 4.2.

Figure 4.2b shows the pre-depuration distribution of NoV obtained from the lognormal model described in Chapter 3. Identical (where appropriate) parameters were used to allow a close comparison between the pre-depuration shapes.



(a) $P_T(x_0)$ pre-depuration PDF

(b) $P_L(x_0)$ pre-depuration PDF

Figure 4.2: Shape parameters are based on $\bar{x_0} = 1064$ NoV cpg (derived from Lowther data — Class B, Jan '11 [2]), $\phi = 0.95$

It is apparent that there is a marked difference between the shapes of the distributions. The lognormal version exhibits a notable peak away from zero, whereas the truncated normal version's maximum is at zero NoV cpg. Note that the vertical scales of the two plots differ, with the truncated normal plot exhibiting a greater peak than the lognormal plot.

More important to observe are the differences in the size areas highlighted by the different sections under each curve. The legend in each plot's top right corner detail the total probability density within each section, and are restated in Table 4.3 below for easy comparison.

NoV cpg range	Truncated Normal Model	Lognormal Model	Difference
$0 < x_0 \leqslant 200$	0.151	0.033	0.118
$200 < x_0 \leqslant 500$	0.192	0.067	0.125
$500 < x_0 \leqslant 1000$	0.240	0.095	0.145
$x_0 > 1000$	0.417	0.805	-0.388

Table 4.3: Comparison of probability densities within sections between truncated normal and lognormal pre-depuration distributions

The differences between the values for the models show that, even though we have employed the same worst case variability approach to inform the spread of both distributions, the lognormal model's pre-depuration shape has a much larger tail density than the truncated normal's shape. This is especially highlighted by the densities of both distributions when $x_0 > 1000$. The lognormal model returns a density 0.388 greater than the truncated normal in this range. In plainer language, this difference can be interpreted as there being 38.8% greater likelihood that a randomly sampled shellfish from the pre-depuration lognormal population will have a NoV count greater than 1000 cpg.

4.5.3 *During Depuration: Truncated Normal versus Lognormal Model Comparison*

Figure 4.3 shows the dynamics of the truncated normal model at four time snapshots during depuration — 0, 50, 100, 150 hours, and based on a depuration decay rate for the system of $b = 0.01339\,\mathrm{hr}^{-1}$. It can be seen that as the

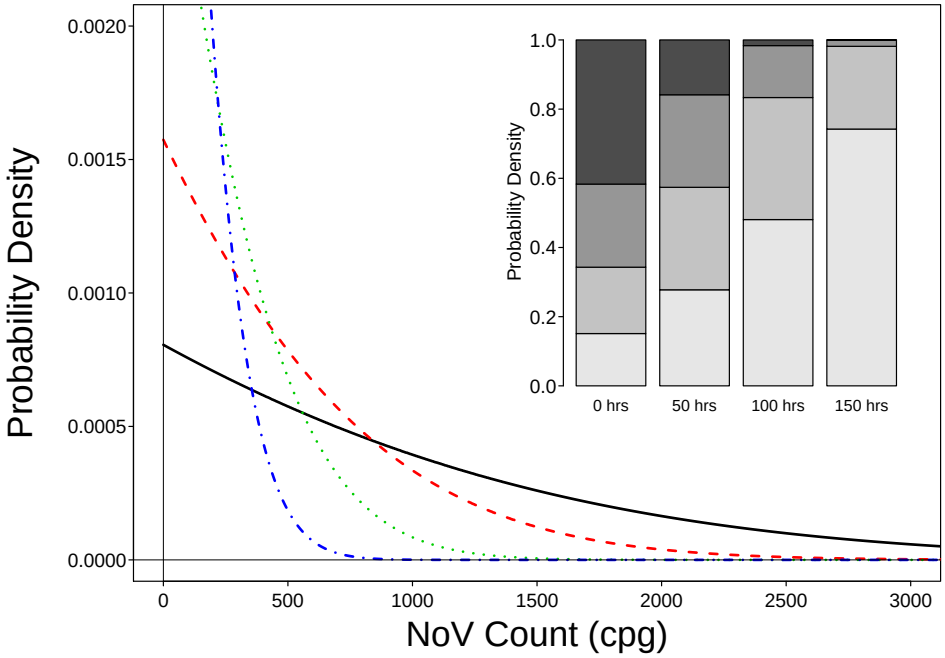


Figure 4.3: $P_T(x_t)$ dynamics during depuration, with decay rate $b = 0.01339$. Main plot shows probability distributions at $t = 0$ hrs (——), $t = 50$ hrs (– – –), $t = 100$ hrs (· · ·), $t = 150$ hours (– · –) using the same parameters and threshold values as Figure 4.2a. Inset bar plot shows the respective changes in section probabilities for each time point.

depuration process progresses, there is a marked shift of densities towards zero, resulting in an increase in the peak's height at zero.

The behaviour of the lognormal model during depuration using the same depuration rate is shown in Figure 4.4. In this plot we also observe a shift towards greater positive skewness and increased peak height close to zero, as we did with the truncated normal distribution model.

The main feature of these two plots are the barplots included as insets in both diagrams. The height of each individual bar within each time point represents the probability density of a particular NoV range, as per the ranges

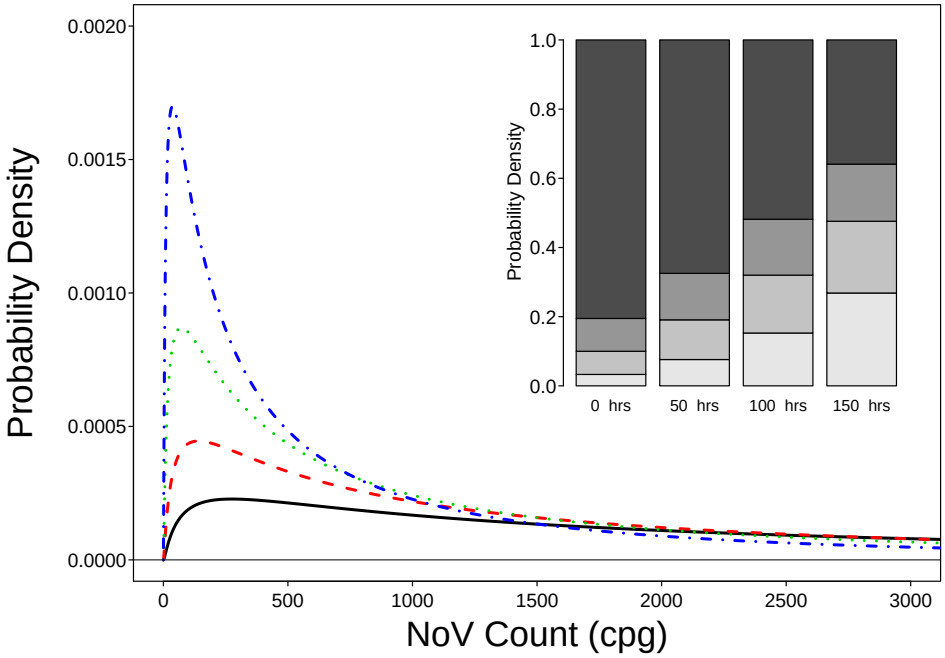


Figure 4.4: $P_L(x_t)$ dynamics during depuration, with decay rate $b = 0.01339$. Main plot shows probability distributions at $t = 0$ hrs (——), $t = 50$ hrs (– – –), $t = 100$ hrs (· · ·), $t = 150$ hours (– · –) using the same parameters and threshold values as Figure 4.2b. Inset bar plot shows the respective changes in section probabilities for each time point.

and grey colours used in Figures 4.2a and 4.2b. The densities for both barplots are enumerated for comparison in Table 4.4. We again define the body of a distribution as being the area under its curve which is less than the NoV assurance level Ψ, i.e. $P_{L/T}(x_t) < \Psi$. Conversely, the tail of the distribution is reiterated as $P_{L/T}(x_t) > \Psi$. We have set the value of the at $\Psi = 200$ NoV cpg, and the assurance level at $\phi = 0.95$ (Table 4.2) The values in Table 4.4 show that the lognormal model maintains a much larger tail during depuration, with only 26.8% of its distribution density below 200 NoV cpg even after 150 hours of depuration. The body of the distribution of the truncated normal model is measured as 0.742 compared to the 0.268 of the lognormal model. However, even after 150 hours, the density of the body is still less than the

Time, t	0 hrs			50 hrs			100 hrs			150 hrs		
Range	Trunc.	Log.	Diff.	Trunc.	Log.	Diff.	Trunc.	Log.	Diff.	Trunc.	Log.	Diff.
$0 < x_t \leqslant 200$	0.151	0.033	0.118	0.278	0.076	0.202	0.481	0.153	0.328	0.742	0.268	0.474
$200 < x_t \leqslant 500$	0.192	0.067	0.125	0.296	0.115	0.181	0.353	0.167	0.186	0.239	0.207	0.032
$500 < x_t \leqslant 1000$	0.240	0.095	0.145	0.268	0.134	0.134	0.150	0.162	-0.012	0.018	0.165	-0.147
$x_t > 1000$	0.417	0.805	-0.388	0.159	0.675	-0.516	0.016	0.518	-0.502	0.000	0.359	-0.359

Table 4.4: Comparison of probability densities within sections between truncated normal and lognormal models during the depuration process. Red values indicate when the lognormal model exhibits a greater probability density for a particular range and time, with green figures showing when the truncated normal model has a higher density value

NoV assurance level of 0.95.

These results do provide some insight into a comparison between the minimum depuration times of the two models, and this is explored in the next section.

4.5.4 *Minimum Depuration Time Comparison*

Summarising Section 4.4, we restate that the minimum depuration time of a particular population of shellfish with individual pathogen loads levels x_t is defined as the solution for time t of

$$P_{L/T}(x_t < \Psi) \leqslant \phi\,, \tag{4.33}$$

and applies to the distributions of both models., $P_L(x_t)$ and $P_T(x_t)$. The analytical solution as stated by Equation 3.32 has been used to calculate MDT's for the lognormal model, however only numerical methods for the truncated normal can be used (see Section 4.4.2). The crux of this chapter is the analysis and comparison of the minimum depuration times between the truncated normal

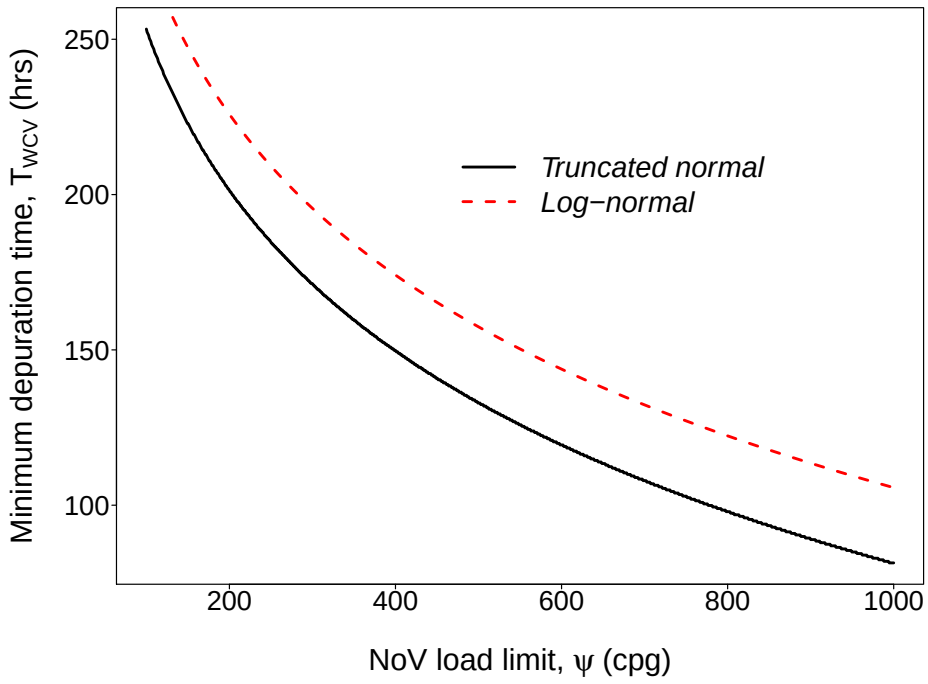


Figure 4.5: Plot of NoV load limit Ψ versus minimum depuration time. MDT's calculated using parameter values of $\phi = 0.95$, $\bar{x}_0 = 1064$ NoV cpg

and lognormal models constructed in this thesis. We have shown that the distribution shapes of the two models have significantly different tail values throughout the depuration process (Section 4.5.3), and understanding what impact this has on the minimum depuration time required by both models to achieve the NoV assurance level is critical.

Figure 4.5 plots MDT's for a range of values of the NoV load limit Ψ. For all values of Ψ, the MDT of the lognormal model is greater than that of the truncated normal model, with both plots following a monotonic, decreasing pattern. Figure 4.6 shows the same information with the notable exception that the horizontal axis has been transformed to a log-scale.

It is apparent that the MDT's for both plots follow the same negative, log-

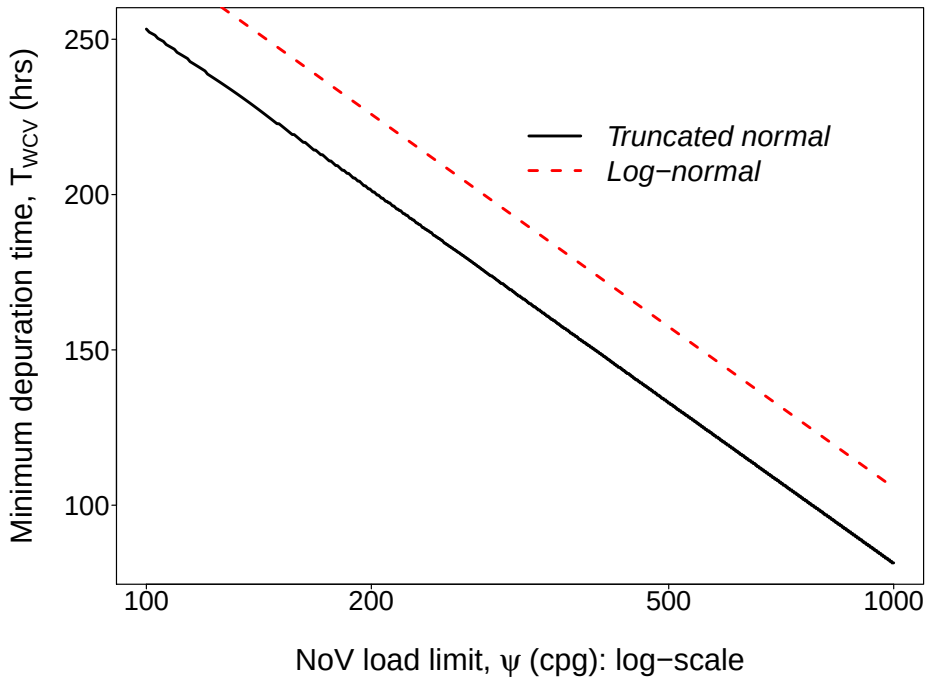


Figure 4.6: Plot of NoV load limit Ψ versus minimum depuration time (hrs). Note the log scale used on the horizontal axis. MDT's calculated using parameter values of $\phi = 0.95$, $\bar{x}_0 = 1064$ NoV cpg

scale gradient as one another. The equation which describes the MDT for the lognormal model (Equation 3.32 states

$$MDT_{WCV} = b^{-1}\left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right].$$

This can be rewritten as

$$\begin{aligned} MDT_{WCV} &= b^{-1}\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + b^{-1}\ln\left(\frac{\bar{x}_0}{\Psi}\right) \\ \Rightarrow MDT_{WCV} &= b^{-1}\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + b^{-1}\ln\left(\frac{\bar{x}_0}{\Psi}\right) \\ \Rightarrow MDT_{WCV} &= b^{-1}\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + b^{-1}\ln\left(\bar{x}_0\right) - b^{-1}\ln\left(\Psi\right). \end{aligned} \tag{4.34}$$

Thus MDT_{WCV} depends linearly on $\ln(\Psi)$, with a gradient equal to $-b^{-1}$ and a y-intercept of $b^{-1}\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + b^{-1}\ln\left(\bar{x}_0\right)$.

Figure 4.6 shows that all MDTs for the truncated normal model are less than that of the lognormal model. For a NoV load threshold level of $\Psi = 200$ NoV cpg, 200 hours depuration time is required by the truncated model, whereas this is much higher at 225 hours for the lognormal model.

4.6 DISCUSSION

In this chapter we examined the impact of applying a different statistical distribution other than lognormal as the basis for our depuration model. We selected a truncated normal distribution, based on the main requirements of non-negativity and positive skewness, to describe the dispersal of pathogens across a population of shellfish. The truncated normal distribution is a variant of the normal distribution, and we showed how a lower, singly-truncated normal distribution can be derived from the normal or Gaussian type. We then adopted the same methodology as was used in Chapter 3 in the construction of a mathematical depuration model based on a truncated normal distribution.

Parameters that had been used with the lognormal model were deployed within the truncated normal where appropriate, and others which could not be reused were derived by mathematical means. As previously discussed, there is a paucity of available data which could inform a reasonable approximation of the variability of water-borne pathogens across a shellfish population. Thus the 'worst case variability' approach which was developed in Section 3.3.8 demonstrates its worth, as it fills this gap in our parameterisation requirements, and so also allows a worst case variability MDT to be calculated for the truncated normal distribution.

Crucially, the pathogen control parameters Ψ and ϕ were again incorporated into our model, allowing us to calculate MDT's for the truncated normal model. This allowed us to compare the depuration times between the models, and we noted that the lognormal model requires greater depuration periods to achieve the desired NoV assurance levels.

This is a consequence of the differing shapes described by the two distributions, with the lognormal model exhibiting a significantly "heavier"tail in comparison with the truncated normal model. More succinctly, the lognormal model describes a shellfish population that will have a greater number of individual molluscs with a high pathogen load. This occurs even though the variability for both models is driven by the same parameter value.

This comparison between lognormal and truncated normal distributions in our modelling has highlighted an important aspect of pathogen control for the shellfish industry, especially where NoV is concerned. All evidence in the literature points towards a lognormal distribution as the best fit shape to describe pathogen loads across a population of shellfish, and consequentially there exists the probability of a small number of shellfish with a NoV load much greater than the rest of the batch being harvested. Our model enforces a pathogen assurance level ϕ where, when the value of ϕ is very close to 1, provides a greater level of food safety for the consumer; however, this is offset by the resulting extended depuration times required to achieve the increased assurance levels.

The consideration of increased food safety versus reasonable implementation of pathogen mitigation practices would be of primary importance if and when future legislation was introduced to monitor and control NoV. By using the lognormal distribution as the basis for our model, and not the truncated normal distribution, we additionally adopt a more cautious approach with regards to the MDTs required to conform to Ψ and ϕ as MDTs for the lognormal model are greater than those for the truncated normal model (see Figures 4.5 and 4.6).

5

MODEL 2 — COMPARTMENTAL OYSTER DEPURATION MODEL

5.1 INTRODUCTION

Chapter 3 introduced a mathematical model of NoV dynamics during depuration founded on two main premises:

1. The pathogen is lognormally distributed across a molluscan population;
2. Each mollusc's NoV load decayed exponentially at the same rate.

This model is also predicated on the assumption that each mollusc stores any NoV present in one compartment within its anatomy, namely its digestive glands. However, some recent studies have shown that molluscs compartmentalise NoV loads within biological tracts that are not currently tested for the presence of NoV. Evidence shows that molluscs internally sequester and transfer pathogen levels sequentially through their whole digestive system; passing through their gills, labial palps, mouth, oesophagus and intestines before reaching their stomach/digestive glands (Figure 5.1)[29, 30].

Special consideration must be made for oysters due to the fact that only their digestive glands are currently tested for NoV. As previously stated in Chapter 3, only the digestive glands of the oyster are currently tested for NoV using quantitative PCR assay, and thus only a proportion of the NoV burden present in an oyster is measured [41, 28], with the rest of the shellfish discarded. This includes the tract of the oyster's digestive system which precedes the digestive

glands [15, 73].

Based on this internal sequestration of NoV, we extend the lognormal depuration model from Chapter 3 and incorporate this compartmentalisation of water-borne pathogens within oysters into the model. This facilitates the development of a description of the dynamics of the compartmentalised pathogen loads across a population of oysters. The rest of the chapter examines the differences between the two modelling approaches, looking at the impact that the compartmentalised NoV approach would have upon NoV levels and minimum depuration times. Other implications for testing protocols are also considered and discussed.

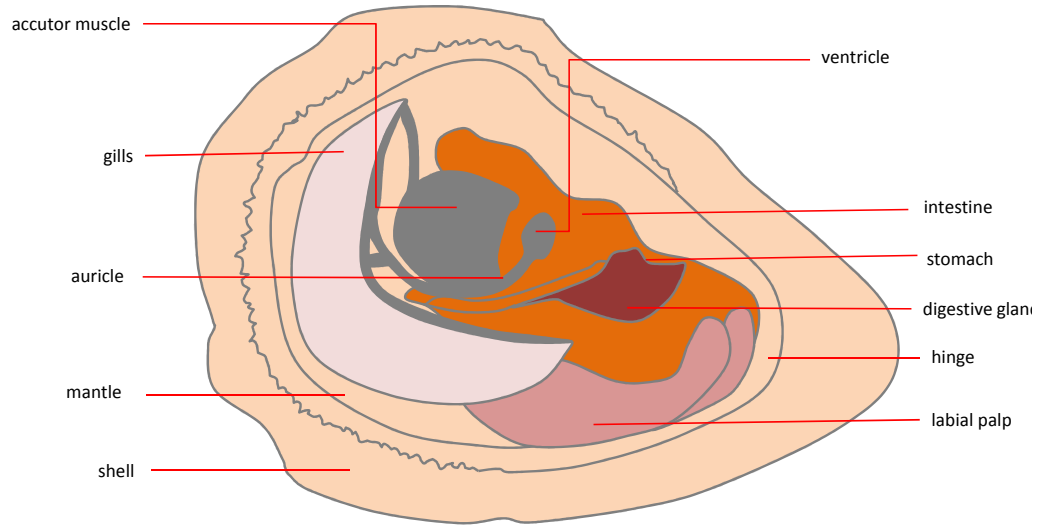


Figure 5.1: Digestive system of the oyster species

The most common assay employed for the detection of NoV within shellfish is polymerase chain reaction (PCR); however this only quantifies the NoV load within the digestive glands of the oyster [28]. The digestive glands only comprise one section of an oyster's overall digestive system, with the rest of the oyster unable to be tested for NoV and so are excluded in the PCR test. This includes the tract of the oyster's digestive system which precedes the digestive glands [15, 73]. As these sections of the animal comprise a significant proportion of the oyster's digestive system, the possibility exists that they may contain significant levels of NoV which cannot be measured by the current PCR assay.

As previously stated, evidence in the literature indicates that pathogens can be sequestered away from the digestive glands. In 1995, Dore and Lees carried out experiments analysing depuration effects on FRNA+ bacteriophage (a common NoV test surrogate) within oysters and mussels. They reported that

FRNA+ bacteriophage was still detected in approximately 60% of the digestive glands and 40% elsewhere within mussels. Wang carried out a similar study, analysing the sequestration of NoV in suminoe oysters (*Crassostrea ariakensis*) using immunohistochemical analysis [30]. They also reported that significant NoV levels were discovered out with the digestive glands of the animals. Therefore, it follows that a mathematical model of NoV dynamics during depuration should incorporate a presumption of pathogen loads sequestered out with the digestive gland, and consequentially outwith the capacity of PCR to quantify any pre-gland NoV burden.

5.3 EXPERIMENTAL DEPURATION DATA

A study was carried out in 2013 by Dr Anna Neish from the Weymouth laboratory of CEFAS, with the aim of determining whether variations in depuration water temperature and water treatment would have a significant impact upon the removal rate of NoV from shellfish. The data have been provided by Neish and presented here with permission [39]. The study compared the effectiveness of depuration water temperatures of 8°C versus 16°C, and ultraviolet radiation versus ozone as disinfectants of depuration tank water. The results obtained showed that depuration water temperature of 16°C was statistically more effective than water at 8°C, although still only showed a slight improvement in NoV levels after 330 hours. The study also noted that ozone did not significantly improve NoV mitigation compared with the use of ultraviolet irradiation. In this thesis, we concentrate on an aspect that was not analysed in great depth in [39], namely the depuration time dependence of NoV levels.

Figure 5.2 shows the data from one of the experiments carried out using ultraviolet radiation water treatment and tank water with a temperature of 16°C. Neish's data was obtained by carrying out PCR testing of four sets of 10 homogenates of *C. gigas*, at seven time points across a period of depuration. The first time point was a pre-depuration measure, with the second NoV sample taken after 42 hours — the minimum depuration time required by regulation for class B harvested shellfish. Beyond this, samples were taken at 90, 162, 210, 258 and 330 respectively and PCR quantitative analysis was carried out on all time samples.

Figure 5.2 shows that an increase between the geometric means of NoV levels between 0 and 42 hours was recorded by the study for the 16°C data.

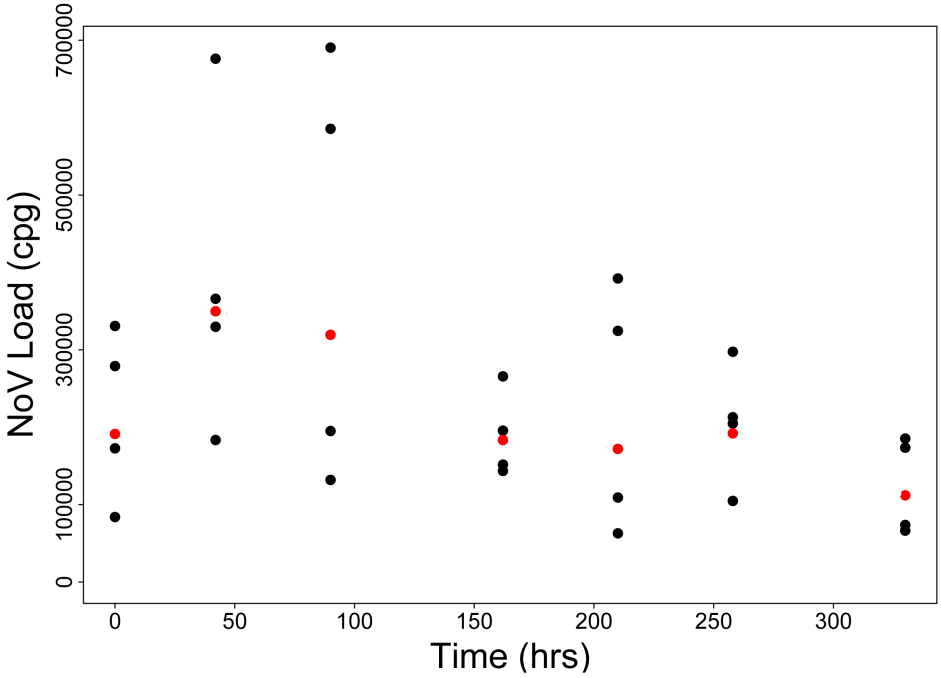


Figure 5.2: Plot of during depuration data set. Four homogenates , each comprised of ten oysters were tested for genogroup II and NoV loads at t = 0, 42, 90, 162, 210, 258, 330 hours are shown on the plot as black points. Red points are geometric means of each time point's data

This increase can be attributed to an internal transfer of NoV load from a compartment of the oyster that currently cannot be assayed, into the digestive glands of the oyster, the only section of the oyster which is currently tested by PCR.

5.4.1 *Internal NoV dynamics*

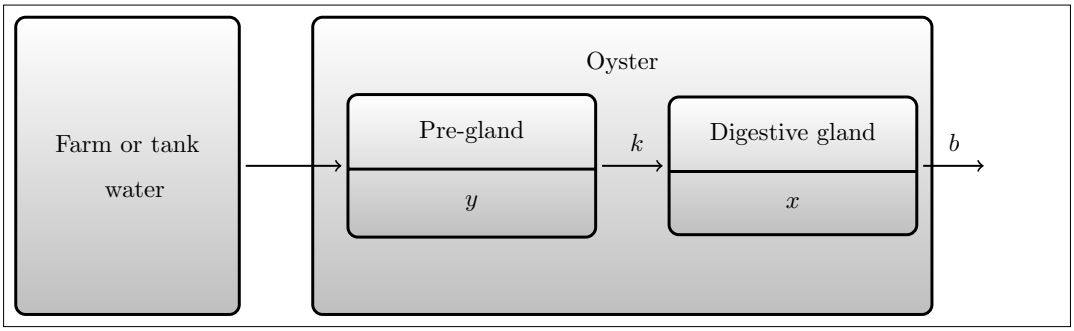


Figure 5.3: Representation of NoV transit through an oyster's digestive system during relaying/depuration

Once an oyster is placed into a depuration tank, any NoV present in the unmeasurable, pre-gland sections would recommence traversing through the animal's digestive system, finally passing into the digestive gland (see Figure 5.3). A compartmental model can be used to describe the flow of NoV through an oyster's digestive system. The total NoV count of the oyster at time t is defined as z_t, and is divided into two compartments, x_t and y_t, where

$$z_t = x_t + y_t \,. \tag{5.1}$$

The compartment labelled x_t in Figure 5.3 represents the NoV load within the section of the oyster's digestive system which can be currently measured using PCR assay, referred to as the **observable** NoV load. The NoV load in the remaining parts of the animal's digestive system which are not currently measured by PCR assay are denoted by y_t, also known as the **unobservable** NoV load per oyster. We assume that x_t and y_t, as continuous functions of time, satisfy a set of differential equations where

$$\frac{dx_t}{dt} = ky_t - bx_t \,, \tag{5.2}$$

$$\frac{dy_t}{dt} = -ky_t \,. \tag{5.3}$$

Here the parameter k quantifies the internal transfer rate of NoV from the unobservable compartment into the observable section. The parameter b describes the rate at which NoV is removed from the digestive gland by excretion during depuration.

It is assumed that at $t = 0$, the animal's total NoV load is split across these two compartments, with the observable load described by

$$x_0 = Az_0 \tag{5.4}$$

and the unobservable load at $t = 0$ defined as

$$y_0 = (1 - A)z_0 \,, \tag{5.5}$$

where A determines the proportion of an oyster's total, initial NoV load (z_0) present in the **observable** part of the digestive gland, with $0 < A \leqslant 1$.

5.4.2 *Obtaining Analytic Solutions for Compartmental Model*

The first order, homogeneous equations describing the compartmentalisation of NoV are simple enough to allow analytic solutions to be obtained for Equations 5.1, 5.2 and 5.3. Solutions are obtained below, beginning with Equation 5.3.

5.4.2.1 *Analytical Solution for Unobservable NoV Load*

Using the method of separation of variables, from

$$\frac{dy_t}{dt} = -ky_t$$

we obtain

$$y_t = B \exp\{-kt\}, \tag{5.6}$$

where B is an arbitrary constant. Setting $t = 0$ and using the initial condition of Equation 5.5), we obtain

$$y_0 = B$$
$$\Rightarrow B = (1 - A)z_0 \,.$$

Substituting this into Equation 5.6 gives an analytical solution for the unobservable compartment y_t:

$$y_t = (1 - A)z_0 \exp\{-kt\} \,. \tag{5.7}$$

Rearranging Equation 5.4 as $z_0 = x_0/A$ and substituting into Equation 5.7 restates y_t in terms of the observable NoV load:

$$y_t = \frac{(1 - A)}{A} x_0 \exp\{-kt\} \,. \tag{5.8}$$

5.4.2.2 *Analytical Solution for Observable NoV Load*

The analytical solution for the unobservable load (y_t) can be used to obtain an analytical solution for x_t. Substituting Equation 5.8 into Equation 5.2 gives

$$\frac{dx_t}{dt} = \frac{k(1 - A)}{A} x_0 \exp\{-kt\} - bx_t \,. \tag{5.9}$$

This is a linear, first-order, ordinary differential equation of the form

$$\frac{dy}{dx} + Py = Q \,.$$

This form of differential equation can be solved using the integrating factor method where an integrating factor I is deployed, with $I = \exp\left\{\int P \, dx\right\}$. The integrating factor for Equation 5.9 is $I = \exp\{bt\}$, therefore an analytical solution can obtained by firstly multiplying both sides by $\exp\{bt\}$:

$$\frac{dx_t}{dt} . \exp\{bt\} = \frac{k(1 - A)}{A} x_0 \exp\{-kt\} . \exp\{bt\}$$
$$\Rightarrow \frac{d}{dt} \left(x_t . \exp\{bt\}\right) = \frac{k(1 - A)}{A} x_0 \exp\{(b - k)t\} \,.$$

Integrating both sides with respect to t, we obtain

$$\begin{aligned} x_t \exp\{bt\} &= \frac{k(1-A)}{A} x_0 \int \exp\{(b-k)t\}\, dt \\ \Rightarrow x_t \exp\{bt\} &= \frac{k(1-A)}{(b-k)A} x_0 \exp\{(b-k)t\} + C \\ \Rightarrow x_t &= \frac{k(1-A)}{(b-k)A} x_0 \exp\{-kt\} + C\ \exp\{-bt\}, \end{aligned} \tag{5.10}$$

where C is the constant of integration of $\int \exp\{(b-k)t\}dt$. To obtain a solution for C, we let $t = 0$:

$$\begin{aligned} x_0 &= \frac{k(1-A)}{(b-k)A} x_0 + C \\ \Rightarrow C &= x_0 \left(1 - \frac{k(1-A)}{(b-k)A}\right). \end{aligned}$$

Thus Equation 5.10 can be rewritten as

$$\begin{aligned} x_t &= \frac{k(1-A)}{A(b-k)} x_0 \exp\{-kt\} + x_0 \left(1 - \frac{k(1-A)}{A(b-k)}\right) \exp\{-bt\} \\ \Rightarrow x_t &= x_0 \left[\frac{k\,(1-A)}{A(b-k)} \exp\{-kt\} + \left(1 - \frac{k(1-A)}{A(b-k)}\right) \exp\{-bt\}\right] \\ \Rightarrow x_t &= \frac{x_0}{A} \left[\frac{k\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\}\right] \\ \Rightarrow x_t &= x_0\, \Theta_t, \end{aligned} \tag{5.11}$$

where

$$\Theta_t = A^{-1} \left[\frac{k\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\}\right], \tag{5.12}$$

and $0 \leqslant A \leqslant 1$, $b \neq k$.

5.4.2.3 *Analytical Solution for total NoV load*

We can now obtain an analytical solution for the total NoV load in an oyster (z_t) by substituting in the solutions for x_t and y_t into Equation 5.1 which states that

$$
\begin{aligned}
& z_t = x_t + y_t \\
\Rightarrow\ & z_t = \frac{x_0}{A}\left[\frac{k\,(1-A)}{(b-k)}\exp\{-kt\} + \frac{(Ab-k)}{b-k}\exp\{-bt\}\right] + \frac{(1-A)}{A}x_0\exp\{-kt\} \\
\Rightarrow\ & z_t = \frac{x_0}{A}\left[\left(\frac{k\,(1-A)}{(b-k)} + (1-A)\right)\exp\{-kt\} + \left(\frac{Ab-k}{b-k}\right)\exp\{-bt\}\right] \\
\Rightarrow\ & z_t = \frac{x_0}{A}\left[\left(\frac{k-Ak+(1-A)(b-k)}{(b-k)}\right)\exp\{-kt\} + \left(\frac{Ab-k}{b-k}\right)\exp\{-bt\}\right] \\
\Rightarrow\ & z_t = \frac{x_0}{A}\left[\left(\frac{k-Ak+b-k-Ab+Ak}{(b-k)}\right)\exp\{-kt\} + \left(\frac{Ab-k}{b-k}\right)\exp\{-bt\}\right] \\
\Rightarrow\ & z_t = \frac{x_0}{A}\left[\left(\frac{b(1-A)}{b-k}\right)\exp\{-kt\} + \left(\frac{Ab-k}{b-k}\right)\exp\{-bt\}\right]. \qquad (5.13)
\end{aligned}
$$

Using the initial condition that $x_0 = Az_0$, we can restate Equation 5.13 as

$$
\begin{aligned}
& \Rightarrow z_t = z_0\left[\frac{b\,(1-A)}{(b-k)}\exp\{-kt\} + \frac{(Ab-k)}{(b-k)}\exp\{-bt\}\right] \\
& \Rightarrow z_t = z_0\Omega_t\,, \qquad (5.14)
\end{aligned}
$$

where

$$
\Omega_t = \left[\frac{b\,(1-A)}{(b-k)}\exp\{-kt\} + \frac{(Ab-k)}{(b-k)}\exp\{-bt\}\right]. \qquad (5.15)
$$

This solution can also be obtained from the principles of linearity:

$$
\begin{aligned}
z_t &= x_t + y_t \\
\Rightarrow \frac{dz_t}{dt} &= \frac{dx_t}{dt} + \frac{dy_t}{dt}\,,
\end{aligned}
$$

and this is expanded upon in Appendix B.1.1 and shown to hold true and equal Equations 5.14 and 5.15.

5.4.3 *NoV Transfer Rates Constraint*

Solutions to Equations 5.1, 5.2 and 5.3 have been obtained above using the initial conditions $x_0 = Az_0$ and $y_0 = (1-A)z_0$. These solutions are collated here:

$$x_t = x_0\,\Theta_t\,, \tag{5.16}$$

$$y_t = y_0 \exp\{-kt\}\,, \tag{5.17}$$

$$z_t = z_0\,\Omega_t\,, \tag{5.18}$$

where

$$\Theta_t = A^{-1}\left[\frac{k\,(1-A)}{(b-k)}\exp\{-kt\} + \frac{(Ab-k)}{(b-k)}\exp\{-bt\}\right]\,, \tag{5.19}$$

and

$$\Omega_t = \left[\frac{b\,(1-A)}{(b-k)}\exp\{-kt\} + \frac{(Ab-k)}{(b-k)}\exp\{-bt\}\right] \tag{5.20}$$

with $b \neq k$ and $0 < A \leqslant 1$.

We have previously stated that NoV selectively binds to the digestive gland of molluscs (x_t). If k, the internal transfer rate of NoV from y_t to x_t, was less than the excretion/depuration rate b, then the observable compartment x_t would only act as a conduit for the purging of viral loads from y_t. Thus there would be no accumulation of NoV in the observable compartment, which is contrary to literature findings [42, 74, 40]. Conversely, if $b < k$ then the outflow of NoV from the observable compartment x_t would be greater than the inflow from y_t. Therefore declaring that $b < k$ realistically describes the selective accumulation of NoV within the x_t compartment.

5.4.4 *Equilibrium of System*

We now consider the equilibrium of the system of Equations 6.15 and 6.16. Setting Equations 6.15 and 6.16 to equilibrium (i.e. that there is no change to either of x_t and x_t as time progresses) gives

$$\begin{aligned} \frac{dy_t}{dt} &= -ky_t = 0 \\ \frac{dx_t}{dt} &= ky_t - bx_t = 0\,. \end{aligned} \tag{5.21}$$

From Equation 5.21 we see that, at equilibrium, $ky_t = bx_t \Rightarrow \frac{y_t}{x_t} = \frac{b}{k}$. A is defined as the proportion of observable (x_t) NoV to the total ($z_t = x_t + y_t$) NoV load within a shellfish, therefore

$$\begin{aligned} A &= \frac{x_t}{x_t + y_t} = \frac{1}{1 + \frac{y_t}{x_t}} \\ \Rightarrow A &= \frac{1}{1 + \frac{b}{k}} \\ \Rightarrow A &= \frac{k}{k + b}\,. \end{aligned} \tag{5.22}$$

Therefore, at equilibrium, the magnitude of the observable NoV load A can be estimated using the internal transfer rate k and the depuration rate b.

5.4.5 *Probability Distributions of x_0 and z_0*

Equations 5.16 — 5.20 provide a description of the depuration dynamics of NoV within a single mollusc. As with the model of NoV dynamics described in Chapter 3, we must apply these equations across a whole population of shellfish to realistically understand and model the depuration process. For simplicity we assume that all variability in the system is associated with the total initial NoV loads, Z_0, and that A is fixed across the population. The

distribution of initial observable NoV loads is denoted by $X_0 = AZ_0$, and is given by

$$P(X_0 = x_0) = A^{-1}\, P(Z_0 = z_0)\,. \tag{5.23}$$

As in Chapter 3, we assume that the loads of NoV across a population of molluscs can be well described by a log normal distribution, with the PDF of observable NoV load described by:

$$P(x_0) = \frac{1}{x_0\sigma_0\sqrt{2\pi}} \exp\left\{\frac{-(\ln(x_0) - \mu_0)^2}{2\sigma_0^2}\right\}. \tag{5.24}$$

Thus, the probability distribution of total NoV loads for a population at pre-depuration can be derived from Equations 5.24 and 5.23:

$$\begin{aligned} P(z_0) &= A\; P(x_0) \\ \Rightarrow P(z_0) &= \frac{A}{x_0\sigma_0\sqrt{2\pi}} \exp\left\{\frac{-(\ln(x_0) - \mu_0)^2}{2\sigma_0^2}\right\}. \end{aligned}$$

Substituting in $x_0 = Az_0$ from Equation 5.4 and simplifying obtains

$$P(z_0) = \frac{1}{z_0\sigma_0\sqrt{2\pi}} \exp\left\{\frac{-(\ln(Az_0) - \mu_0)^2}{2\sigma_0^2}\right\}. \tag{5.25}$$

5.4.6 *Probability Distributions of x_t and z_t*

Equations 5.24 and 5.25 describe the pre-depuration distributions of the observable and total NoV loads respectively. The relationship between the pre-depuration and during depuration variables for the observable and total NoV loads in a single oyster are defined by Equations 5.16 and 5.18, where $x_t = x_0\,\Theta_t$ and $z_t = z_0\,\Omega_t$. As carried out in Section 3.3.4, these equations can be used to change the variables of Equations 5.24 and 5.25, so to provide PDF's for any time $t \geqslant 0$ during depuration.

Beginning with the observable NoV loads, we state the relationship between $P(x_0)$ and $P(x_t)$ as

$$\begin{aligned} P(x_0)dx_0 &= P(x_t)dx_t \\ \Rightarrow P(x_t) &= P(x_0)\frac{dx_0}{dx_t}. \end{aligned} \tag{5.26}$$

To obtain a term for dx_0/dx_t, we use Equation 5.16:

$$\begin{aligned} x_t &= x_0\,\Theta_t \\ \Rightarrow \frac{dx_t}{dx_0} &= \Theta_t \\ \Rightarrow \frac{dx_0}{dx_t} &= \frac{1}{\Theta_t}. \end{aligned}$$

Substituting this and Equation 5.24 into Equation 5.26 gives

$$P(x_t) = \frac{1}{\Theta_t x_0 \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln(x_0) - \mu_0\right)^2 / 2\sigma_0^2\right\}. \tag{5.27}$$

Rewriting Equation 5.16 as $x_0 = x_t/\Theta_t$, and substituting into Equation 5.27 then simplifying provides the description of the depuration distribution for all $t \geqslant 0$:

$$\begin{aligned} P(x_t) &= \frac{1}{\Theta_t x_0 \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln(x_0) - \mu_0\right)^2 / 2\sigma_0^2\right\} \\ \Rightarrow P(x_t) &= \frac{\Theta_t}{\Theta_t x_t \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln\left(\frac{x_t}{\Theta_t}\right) - \mu_0\right)^2 / 2\sigma_0^2\right\} \\ \Rightarrow P(x_t) &= \frac{1}{x_t \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln\left(\frac{x_t}{\Theta_t}\right) - \mu_0\right)^2 / 2\sigma_0^2\right\}. \end{aligned} \tag{5.28}$$

Thus Equation 5.28 (along with the definition of Θ_t in Equation 5.19) describes the distribution of the compartmentalised observable NoV loads before and during the depuration process.

The same process is used to obtain an equation for $P(z_t)$, the distribution of the total NoV load across a population of shellfish. The description of the relationship between $P(z_0)$ and $P(z_t)$ defined by

$$\begin{aligned} P(z_0)dz_0 &= dz_t P(z_t) \\ \Rightarrow P(z_t) &= \frac{dz_0}{dz_t} P(z_0). \end{aligned} \tag{5.29}$$

The derivative of Equation 5.18 is obtained:

$$\begin{aligned} z_t &= z_0\Omega_t \\ \Rightarrow \frac{dz_t}{dz_0} &= \Omega_t \\ \Rightarrow \frac{dz_0}{dz_t} &= \frac{1}{\Omega_t}, \end{aligned} \tag{5.30}$$

and substituted into Equation 5.29, along with $z_0 = z_t/\Omega_t$:

$$\begin{aligned} P(z_t) &= \frac{1}{\Omega_t z_0 \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln\left(Az_0\right) - \mu_0\right)^2 / 2\sigma_0^2\right\}) \\ \Rightarrow P(z_t) &= \frac{\Omega_t}{\Omega_t z_t \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln\left(\frac{Az_t}{\Omega_t}\right) - \mu_0\right)^2 / 2\sigma_0^2\right\} \\ \Rightarrow P(z_t) &= \frac{1}{z_t \sigma_0 \sqrt{2\pi}} \exp\left\{-\left(\ln\left(\frac{Az_t}{\Omega_t}\right) - \mu_0\right)^2 / 2\sigma_0^2\right\}. \end{aligned} \tag{5.31}$$

Thus Equation 5.31 describes the distribution of the total NoV load before and during the depuration process.

5.4.7 *Minimum Depuration Times*

In Chapter 3, two control parameters were applied to ensure the depuration process was successful from a food safety viewpoint:

i) Ψ — a NoV threshold value in copies per gram (cpg);

ii) ϕ — an acceptable proportion of oysters with NoV loads less than Ψ.

Figure 5.4 shows how the value of Ψ splits the distribution into 2 sections: the area shown in green represents the proportion of shellfish whose NoV load is below the value of Ψ. The red area indicates the proportion of the population who still have pathogen loads above Ψ, with this tail of the distribution continuing on to infinity. Therefore, we can quantify the value of ϕ with regards to the observable and total NoV loads during depuration in terms of

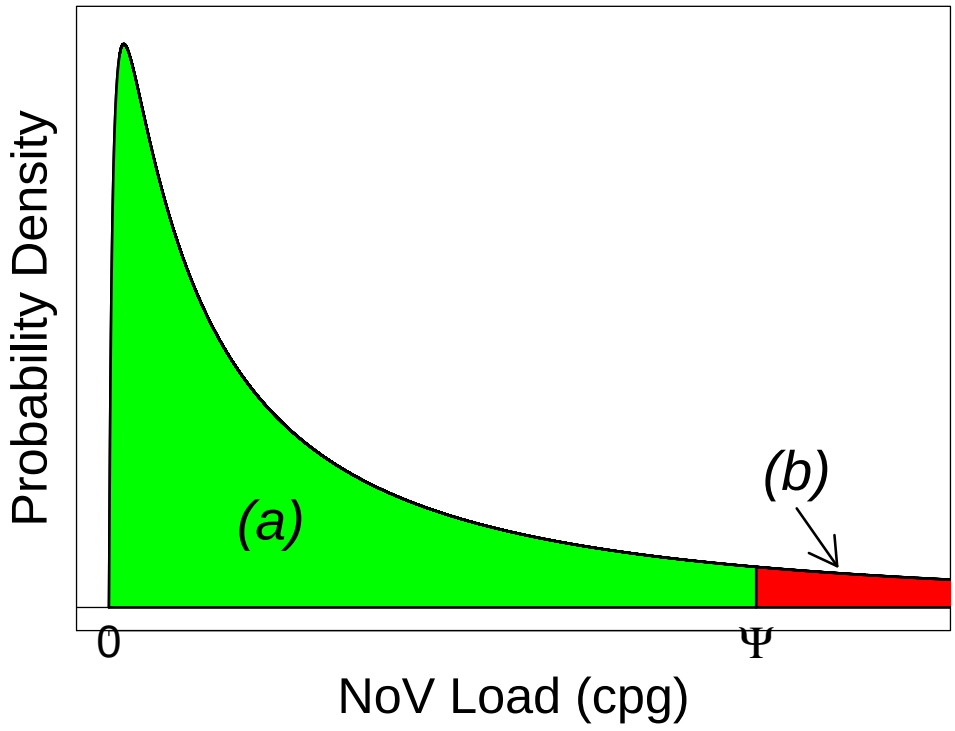


Figure 5.4: Generic plot of NoV distribution at some time t during depuration. Ψ is the NoV threshold limit, *(a)* represents the proportion of shellfish in a population whose NoV loads are within Ψ, and *(b)* indicates the proportion of shellfish whose pathogen loads are still above this threshold limit

Equations 5.28 and 5.31 by calculating the area under the distribution curve that equates to ϕ:

$$P(x_t < \Psi) = \int_0^{\Psi} \frac{1}{x_t \sigma_0 \sqrt{2\pi}} \exp\left\{ \frac{-\left(\ln\left(\frac{x_t}{\Theta}\right) - \mu_0\right)^2}{2\sigma_0^2} \right\} dx_t = \phi \qquad (5.32)$$

and

$$P(z_t < \Psi) = \int_0^{\Psi} \frac{1}{z_t \sigma_0 \sqrt{2\pi}} \exp\left\{ \frac{-\left(\ln\left(\frac{A z_t}{\Omega}\right) - \mu_0\right)^2}{2\sigma_0^2} \right\} dx_t = \phi\,. \qquad (5.33)$$

Equations 5.32 and 5.33 cannot be satisfied simultaneously. Each will equal ϕ after a different time t, dependent upon the value of A which determines the proportions of pathogens within the observable and unobservable compart-

ments.

The integrals of Equations 5.32 and 5.33 will be less than or equal to ϕ only after a specific depuration time t. This time was previously defined as the minimum depuration time required to conform to both of the constraints Ψ and ϕ, and in Chapter 3, an analytical solution to $P(x_t < \Psi) = \phi$ was obtained in terms of the minimum depuration time (T_{WCV}); however both Equations 5.32 and 5.33 are more complex equations with regards to instances of the time term t, and may not allow analytical solutions for t to be obtained. The variables Θ_t and Ω_t (Equations 5.19 and 5.20) are both in the generic form

$$\Theta_t, \Omega_t = [m \exp\{-kt\} + n \exp\{-bt\}] \,, \tag{5.34}$$

where $m, n \in \mathbb{R}$. This cannot be analytically solved for t, therefore analytical solutions for both $P(x_t < \Psi) = \phi$ and $P(z_t < \Psi) = \phi$ in terms of time t cannot be obtained. Calculating minimum depuration times for the observable and total NoV loads from Equations 5.32 and 5.33 would need to be carried out using numerical methods.

5.4.8 *Variability Estimation*

We have previously stated that obtaining parameter estimates for the variability of pathogens across a population of shellfish is costly and time-consuming [39]. To deal with this difficulty, in Section 3.3.8 we established a worst case approach to estimating the variability for the exponential model. This was based on deriving a solution for time t to Equation 3.24, with this equation restated here:

$$P(x_t < \Psi) = \int_0^{\Psi} \frac{1}{x_t \sigma_0 \sqrt{2\pi}} \exp\{-(\ln(x_t) + bt - \mu_0)^2 / 2\sigma_0^2\} \, dx_t = \phi \,. \tag{5.35}$$

Solving this for t provided a formula for calculating the minimum depuration time required to ensure that a proportion ϕ of the shellfish population being depurated had NoV loads less than a threshold level of Ψ cpg. The solution for the minimum depuration time from Equation 5.35 is

$$t(\mu_0, \sigma_0) = b^{-1}\left(\sqrt{2}\,\sigma_0\, erf^{-1}(2\phi - 1) - \ln(\Psi) + \mu_0\right). \tag{5.36}$$

Substituting in the equation which relates the location and shape parameters of the lognormal distribution to its mean

$$\mu_0 = \ln(\bar{x}_0) - \frac{1}{2}\sigma_0^2, \tag{5.37}$$

allowed us to restate t as a convex quadratic equation in terms of σ_0:

$$T(\sigma_0) = b^{-1}\left[-\frac{1}{2}\sigma_0^2 + \sqrt{2}erf^{-1}(2\phi - 1)\,\sigma_0 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right], \tag{5.38}$$

which is maximised when

$$\Rightarrow \sigma_0 = \sqrt{2}erf^{-1}(2\phi - 1). \tag{5.39}$$

However, we cannot assume that this description of worst case variability is applicable for the compartmental model derived in this chapter, so must look at a more general form of Equation 5.36 for any probability density function. This can be generally stated for the exponential model from Chapter 3 as

$$P(x_t < \Psi) = \int_0^{\Psi} P(x_t)\, dx_t = \phi.$$

This general approach can be used for both the observable ($P(x_t < \Psi)$) and total ($P(z_t < \Psi)$), carried forward from Equations 5.32 and 5.33.

These two equations can be generalised as

$$P(s_t < \Psi) = \int_0^{\Psi} \frac{1}{s_t \sigma_0 \sqrt{2\pi}} \exp\{-(\ln(s_t) + g(t) - \mu_0)^2 / 2\sigma_0^2\}\, ds_t = \phi, \tag{5.40}$$

where $g(t)$ is some function of time t, and s_t is either x_t or z_t. For the exponential model's $P(x_t < \Psi)$ described by Equation 5.35, it is apparent that $g(t) = bt$. For the observable and total loads described in this chapter,

$$g(t) = -\ln(\Theta_t) \tag{5.41}$$

$$g(t) = \ln(A) - \ln(\Omega_t) \tag{5.42}$$

respectively.

We now carry out the same steps as applied in Section 3.3.8 to obtain a solution for t, the minimum depuration time:

$$\int_0^{\Psi} \frac{1}{s_t \sigma_0 \sqrt{2\pi}} \exp\{-(\ln(s_t) + g(t) - \mu_0)^2 / 2\sigma_0^2\}\, ds_t = \phi$$

substituting in $u = \ln(s_t)$ and $ds_t = s_t\, du$ gives

$$\int_{-\infty}^{\ln(\Psi)} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\{-(u + g(t) - \mu_0)^2 / 2\sigma_0^2\}\, du = \phi\,.$$

A second substitution of $z = (u + g(t) - \mu_0)/\sqrt{2}\sigma_0$ and $du = \sqrt{2}\sigma_0\, dz$, plus application of the symmetry of the standard, normal distribution results in

$$\frac{1}{\pi} \int_{-\infty}^{D} \exp\{-z^2\}\, dz = \phi$$

$$\Rightarrow \frac{1}{\pi} \int_0^{D} \exp\{-z^2\}\, dz = \phi - \frac{1}{2}\,, \tag{5.43}$$

where

$$D = (\ln(\Psi) + g(t) - \mu_0)/\sqrt{2}\sigma_0\,.$$

Multiplying both sides of Equation 5.43 by 2 and applying the definition of the error function (cf. Equation 3.26) provides

$$\begin{aligned}
\frac{2}{\pi}\int_0^D \exp\{-z^2\}\,dz &= 2\phi - 1 \\
\Rightarrow \operatorname{erf}(D) &= 2\phi - 1 \\
\Rightarrow D &= \operatorname{erf}^{-1}(2\phi - 1) \\
\Rightarrow \ln(\Psi) + g(t) - \mu_0 &= \sqrt{2}\sigma_0 \operatorname{erf}^{-1}(2\phi - 1) \\
\Rightarrow g(t) &= \sqrt{2}\sigma_0 \operatorname{erf}^{-1}(2\phi - 1) - \ln(\Psi) + \mu_0\,. \qquad (5.44)
\end{aligned}$$

A final substitution of Equation 5.37 into Equation 5.44 yields

$$g(t) = -\frac{1}{2}\sigma_0^2 + \sqrt{2}\sigma_0 \operatorname{erf}^{-1}(2\phi - 1) - \ln(\Psi) + \ln(\bar{x}_0)\,.$$

This equation needs to be solved for t to obtain the minimum depuration time. For $g(t) = bt$ (as in Chapter 3) this can be solved numerically, however neither of Equations 5.41 or 5.42 allow an analytical solution for $g(t)$.

We are still able to find the worst case scenario variability for σ_0 as follows. The general equation for $g(t)$ is maximised (as it is in convex quadratic form with respect to σ_0) when $\partial g(t)/\partial\sigma_0 = 0$:

$$\begin{aligned}
\frac{\partial g(t)}{\partial \sigma_0} &= -\sigma_0 + \sqrt{2}\operatorname{erf}^{-1}(2\phi - 1) \\
\Rightarrow -\sigma_0 + \sqrt{2}\operatorname{erf}^{-1}(2\phi - 1) &= 0 \\
\Rightarrow \sigma_0 &= \sqrt{2}\operatorname{erf}^{-1}(2\phi - 1)\,. \qquad (5.45)
\end{aligned}$$

Therefore this equation describes the worst case variability, regardless of $g(t)$. Thus the description of the worst case variability for the exponential model in Chapter 3 immediately translates to the compartmental model for both x_t and z_t.

5.4.9 *Arithmetic Means of* $P(x_t)$ *and* $P(z_t)$

We can consider a different approach to calculating minimum depuration times required to meet the two constraints ϕ and Ψ. In Chapter 3 it was discussed that the current testing protocols for NoV are carried out on homogenates of 10 oysters, and so provide arithmetic means of testing samples NoV burdens. The equation to calculate minimum depuration time (Equation 3.32) applies the arithmetic mean ($\bar{x}_0$) to inform the initial NoV distributions measure of location. The variability described by Equation 3.31 provides the worst case scenario with regards to the measure of spread of the initial distribution, σ_0, and this is solely a function of the assurance level ϕ.
Using this same approach allows us to obtain initial values of μ_0 and σ_0 for our observable and total NoV distributions. We have already stated that only numerical methods will provide solutions for t from Equations 5.32 and 5.33; however if we consider how the arithmetic means of the total and observable distributions behave over time, then we will be able to obtain similarly robust estimates for minimum depuration times.

In Section 3.3.5 of Chapter 3, we stated that the arithmetic mean of a lognormal distribution is defined as

$$\bar{x} = \exp\left\{\mu + \frac{\sigma^2}{2}\right\} . \tag{5.46}$$

We compared the PDF of $P(x_t)$ (Equation 3.15) with that of the PDF of $P(x_0)$, and equated the components which performed the Z-score correction, to obtain values of μ_t and σ_t. Applying the same comparison between $P(x_0)$ and $P(x_t)$ we obtain

$$\mu_t = \mu_0 + \ln(\Theta_t)$$

$$\sigma_t = \sigma_0 .$$

We substitute these two equations into Equation 5.46 to obtain $\bar{x}_t$:

$$\bar{x}_t = \exp\left\{\mu_t + \frac{\sigma_t^2}{2}\right\}$$
$$\Rightarrow \bar{x}_t = \exp\left\{\mu_0 + \ln(\Theta_t) + \frac{\sigma_0^2}{2}\right\}$$
$$\Rightarrow \bar{x}_t = \exp\left\{\mu_0 + \frac{\sigma_0^2}{2}\right\}\ \Theta_t\,.$$

It follows that the arithmetic mean for any time t for the observable NoV load is described by

$$\bar{x}_t = \Theta_t \bar{x}_0\,. \tag{5.47}$$

Similarly, the arithmetic mean of the total NoV load $P(z_t)$ can be derived from its definitions of μ_t and σ_t

$$\mu_t = \mu_0 + \ln(\Omega_t)$$
$$\sigma_t = \sigma_0\,,$$

thus providing an equation for the mean of the total NoV load during depuration:

$$\bar{z}_t = \Omega_t \bar{z}_0\,. \tag{5.48}$$

5.4.10 *Salient Depuration Timescales*

The inclusion of an unobservable compartment of NoV within our model infers that the minimum depuration time for the total NoV load model must be greater than that derived in Chapter 3. Although x_t is observable, z_t $(> x_t)$ is what constitutes the risk to the consumer. Hence, depuration must be aimed at the total NoV load z_t and not only x_t. Therefore it is important to understand and be able to calculate this increase in depuration time by taking into account the initial, unobservable loads.

Examining the behaviour of the means of the observable $(\bar{x}_t)$ and total $(\bar{z}_t)$ NoV loads as shown in Figure 5.5, it is apparent that some time must elapse

in depuration before the total NoV load ($\bar{z}_t$) decays to equal the value of the initial, observable NoV load ($\bar{x}_0$). Note also in Figure 5.5 that $\bar{x}_t$ begins to closely approximate $\bar{z}_t$ after a significant time period in depuration; both of these times will depend on the levels of NoV that are observable in relation to the total NoV burden i.e. both times will be dependent upon the value of the parameter A.

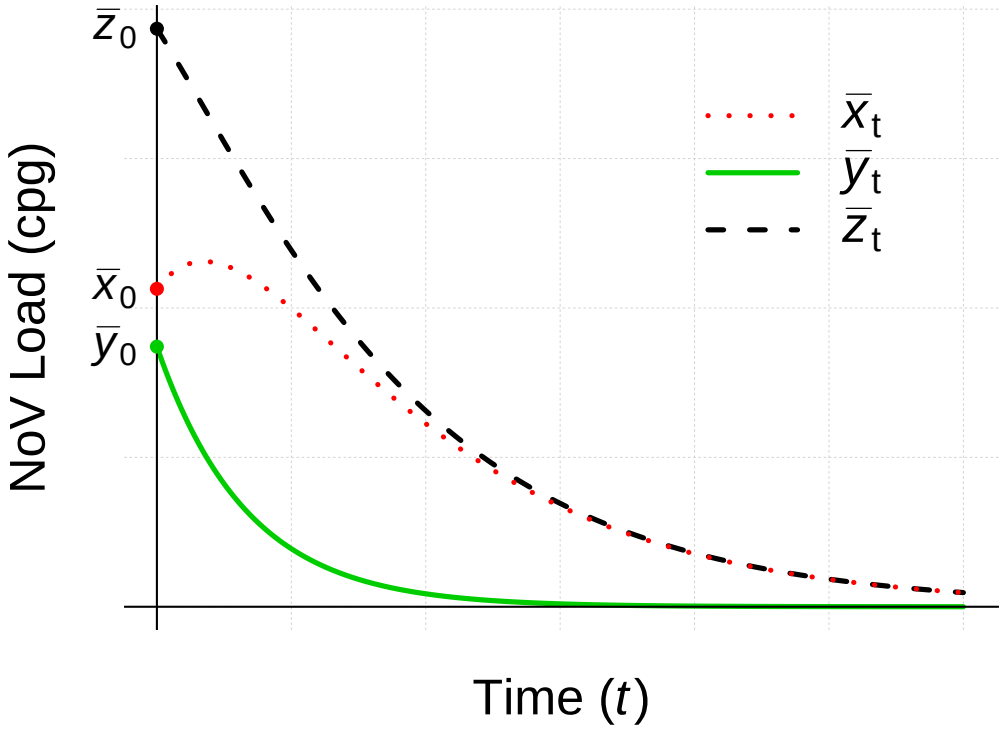


Figure 5.5: Generic plot of behaviour of arithmetic means of NoV loads within compartments $\bar{x}_t$, $\bar{y}_t$, $\bar{z}_t$

5.4.10.1 *Salient Depuration Times Summary*

In the above sections, we have compared the decay dynamics of the mean observable and mean total NoV loads along with the behaviour of the mean NoV loads from the exponential decay model described in Chapter 3. From

these comparisons, four salient depuration time points have been identified and mathematical descriptions have been obtained:

τ_1 — Mean total pathogen load has decayed to match initial, mean, observable load

τ_2 — Mean unobservable load has reduced to insignificant level

τ_3 — Mean observable load has exponentially decayed to equal mean, NoV threshold value

τ_4 — Mean total load has decayed to equal mean NoV threshold value

The more significant of these four times are τ_3 and τ_4, as these describe the minimum depuration times of the exponential model and compartmental models, and the difference between these times would have significant food safety implications, dependent on the value of A. If τ_4 is significantly greater than τ_3 then there exists an increased likelihood of allowing oysters with larger than expected NoV loads to the market.

5.4.10.2 τ_1 : *when* $\bar{z}_t$ *equals* $\bar{x}_0$

Figure 5.6 shows that at some time point during depuration (τ_1), the total NoV count ($\bar{z}_t$) will have decayed sufficiently to equal the initial, observable NoV count ($\bar{x}_0$), and this occurs when

$$\bar{z}_{\tau_1} = \bar{x}_0 \,. \tag{5.49}$$

Equations 5.47 and 5.48 state that $\bar{z}_t = \Omega_t \bar{z}_0$ and $A\bar{z}_0 = \bar{x}_0$. These are used to restate Equation 5.49:

$$\begin{aligned} \bar{z}_{\tau_1} &= \bar{x}_0 \\ \Rightarrow \Omega_{\tau_1} \bar{z}_0 &= \bar{x}_0 \\ \Rightarrow \Omega_{\tau_1} \bar{z}_0 &= A\bar{z}_0 \\ \Rightarrow \Omega_{\tau_1} &= A \,. \end{aligned} \tag{5.50}$$

As previously shown in Section 5.4.7, Ω_t does not afford an analytic solution, but can be solved numerically for τ_1 using Equation 5.50.

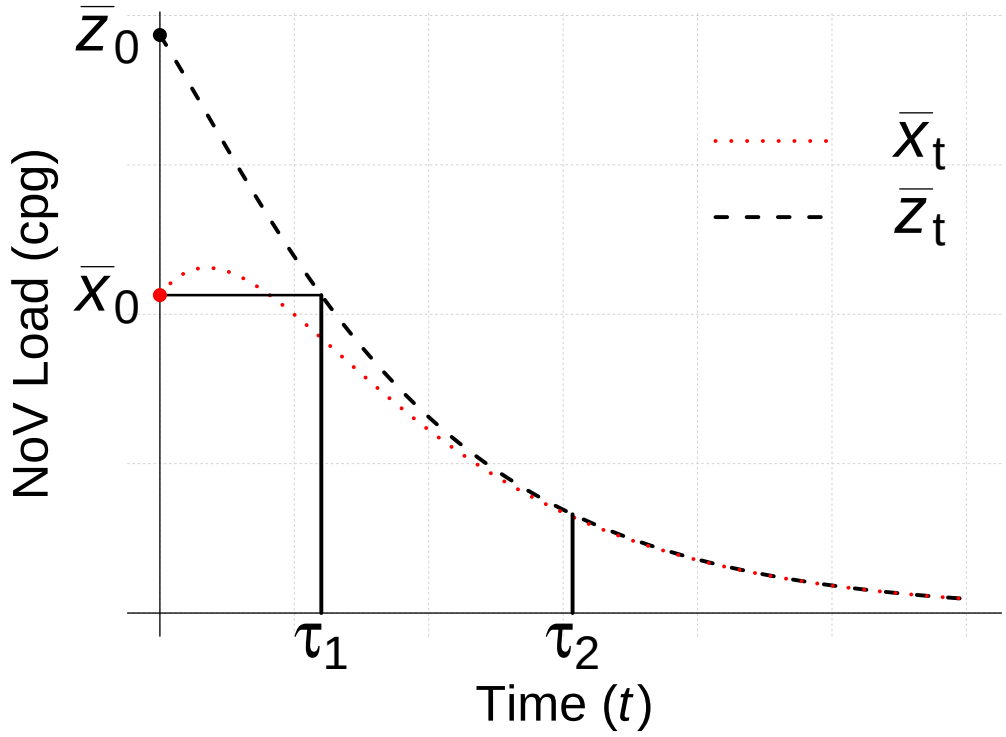


Figure 5.6: Generic plot showing instances of τ_1 and τ_2

5.4.10.3 τ_2 : *when* $\bar{x}_t$ *converges to* $\bar{z}_t$

When both $t \geqslant 0$ and $A \neq 1$, the observable NoV load and total NoV load will approach but never reach a convergence of values. However, we can identify a time during depuration at which they will have reached an approximate convergence (see Figure 5.6). This will occur at time τ_2, when a substantial enough proportion of the initial, unobservable NoV load (y_0) has migrated to or through the animal's NoV-measurable digestive gland (x_t) and has either been excreted by the animal, or is currently located there. Labelling this sufficient proportion as p (where $p \approx 1$) allows us to calculate when τ_2 occurs, i.e. when

$$\bar{y}_{\tau_2} = (1-p)\bar{y}_0 \,. \tag{5.51}$$

This can be simplified for τ_2 using $\bar{y}_{\tau_2} = \bar{y}_0 \exp\{-k\tau_2\}$:

$$\begin{aligned} \bar{y}_{\tau_2} &= (1-p)\bar{y}_0 \\ \Rightarrow \bar{y}_0 \exp\{-k\tau_2\} &= (1-p)\bar{y}_0 \\ \Rightarrow \exp\{-k\tau_2\} &= (1-p) \\ \Rightarrow -k\tau_2 &= \ln(1-p) \\ \tau_2 &= -\frac{1}{k}\ln\{1-p\}\,. \end{aligned} \tag{5.52}$$

When $0 \leqslant t < \tau_2$, the mean observable NoV load $\bar{x}_t$ is notably less than the mean total NoV load $\bar{z}_t$, with the difference decreasing as $t \to \tau_2$. When $t \geqslant \tau_2$ and $p \approx 1$, the NoV load in the unobservable compartment (y_t) becomes negligible and so $\bar{x}_t$ approximates $\bar{z}_t$. At this point, the decay rates for both the observable and total NoV loads equalise.

5.4.10.4 τ_3 : *Minimum Depuration Time of* $\bar{x}_t$ *From Exponential Model*

In Chapter 3, minimum depuration times for populations of shellfish were calculated using Equation 3.32, which is restated below:

$$T_{WCV} = b^{-1}\left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left(\frac{\bar{x}_0}{\Psi}\right)\right]\,. \tag{5.53}$$

Figure 5.7 shows the decay dynamics of $\bar{x}_t$ and $\bar{z}_t$, this time including the individual, exponential decay of the mean observable NoV load according to the model described in Chapter 3, where

$$\bar{x}_t = \bar{x}_0 \exp\{-bt\}\,. \tag{5.54}$$

Adopting this same approach allows us to calculate τ_3, the minimum depuration time of the observable load (i.e. ignoring the impact of any unobservable NoV within the shellfish) as

$$\tau_3 = b^{-1}\left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left\{\frac{\bar{x}_0}{\Psi}\right\}\right]\,, \tag{5.55}$$

where Ψ and ϕ are the NoV load limit and NoV assurance level applied to the depuration system respectively.

Therefore τ_3 is the minimum depuration time required to achieve the NoV threshold without accounting for any unobservable compartments (when $A = 1$), and as such

$$\tau_3 = T_{WCV},$$

with T_{WCV} described by Equation 3.32 in Chapter 3.

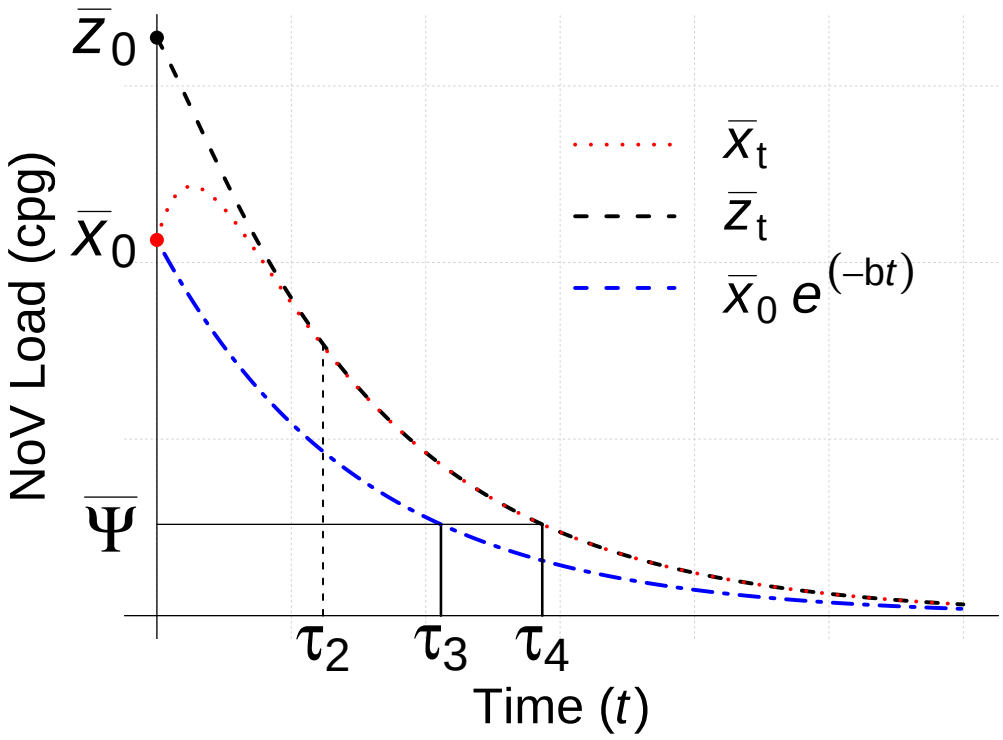


Figure 5.7: Generic plot showing occurrences of τ_2, τ_3 and τ_4

5.4.10.5 *Arithmetic Mean of Distribution Satisfying Ψ and ϕ Control Parameters*

As previously discussed, we cannot obtain analytical solutions for the MDTs of the distributions described by Equations 5.32 and 5.33 (cf. Section 5.4.7). Other than numerical solutions for time t of these two equations, we could potentially compare arithmetic means of these distributions with the mean value of a lognormal PDF that satisfies the control parameters Ψ and ϕ (and so would not require further depuration time). This can then be used as the

benchmark value that both $\bar{x}_t$ and $\bar{z}_t$ would be required to achieve the MDT.

In Section 5.4.9 we derived definitions for $\bar{x}_t$ and $\bar{z}_t$ (Equations 5.47 and 5.48), the arithmetic means of the PDFs that describe the distributions of the observable and total NoV loads during depuration. Therefore, we must obtain a definition for the arithmetic mean of a lognormal PDF that satisfies the control parameters Ψ and ϕ to allow us to use Equations 5.47 and 5.48 to determine MDTs.

Note that to allow this comparison, we must hold the variability constant across the PDFs being compared. Thus we must apply a consistent value of ϕ to allow calculation of τ_4, the minimum depuration time of the total NoV load.

To achieve this, we initially consider the PDF from Chapter 3 which was used to calculate τ_3 in the previous section. The location and scale parameters of that PDF are described by Equations 3.17 and 3.18 in Chapter 3, which state that

$$\sigma_t = \sigma_0 \tag{5.56}$$

and that

$$\mu_t = \mu_0 - bt\,. \tag{5.57}$$

Combining these with Equation 5.46 ($\bar{x} = \exp\{\mu + \sigma^2/2\}$ cf. Section 5.4.9) we calculated the appropriate mean NoV load of the distribution at time t that conforms to the constraints Ψ and ϕ.

This was carried out in Section 3.3.5, however we are interested in calculating the mean of the distribution at time τ_3 as shown in Figure 5.7. To do so, we

require the location and spread parameters of the distribution at time τ_3. Commencing from Equation 5.57, and restating this for τ_3 gives

$$\mu_{\tau_3} = \mu_0 - b\tau_3 \,.$$

Substituting in Equation 5.55 obtains

$$\begin{aligned}
\mu_{\tau_3} &= \mu_0 - \frac{b}{b}\left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left\{\frac{\bar{x}_0}{\Psi}\right\}\right] \\
\Rightarrow \mu_{\tau_3} &= \mu_0 - \left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left\{\frac{\bar{x}_0}{\Psi}\right\}\right] \\
\Rightarrow \mu_{\tau_3} &= \mu_0 - \left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 - \ln\left\{\frac{\bar{x}_0}{\Psi}\right\} \\
\Rightarrow \mu_{\tau_3} &= \mu_0 - \left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 - \ln\left(\bar{x}_0\right) + \ln\left(\Psi\right) \,. \qquad (5.58)
\end{aligned}$$

Equations 5.56 states that $\sigma_t = \sigma_0$. This holds true for the model in Chapter 3, as well as the observable and total load PDFs in this chapter. Equation 5.45 in Section 5.4.8 states that the worst case variability for both the observable and total NoV loads is described by

$$\sigma_t = \sigma_0 = \sqrt{2}\text{erf}^{-1}(2\phi - 1)\,. \qquad (5.59)$$

This can be permuted to aid simplification of Equation 5.58:

$$\begin{aligned}
\sigma_0^2 &= \left(\sqrt{2}\text{erf}^{-1}(2\phi - 1)\right)^2 \\
\Rightarrow \sigma_0^2 &= 2\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 \\
\Rightarrow \frac{1}{2}\sigma_0^2 &= \left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 \,. \qquad (5.60)
\end{aligned}$$

Substituting this into Equation 5.58 gives

$$\mu_{\tau_3} = \mu_0 - \frac{1}{2}\sigma_0^2 - \ln\left(\bar{x}_0\right) + \ln\left(\Psi\right) \,. \qquad (5.61)$$

This equation can be simplified further yet by considering Equation 5.46 which states that

$$\bar{x}_0 = \exp\left\{\mu_0 + \frac{\sigma_0^2}{2}\right\} \,.$$

Taking natural logs on both sides of this equation yields

$$\ln(\bar{x}_0) = \mu_0 + \frac{1}{2}\sigma_0^2, \tag{5.62}$$

which can be substituted into Equation 5.61:

$$\begin{aligned} \mu_{\tau_3} &= \mu_0 - \frac{1}{2}\sigma_0^2 - \mu_0 - \frac{1}{2}\sigma_0^2 + \ln(\Psi) \\ \Rightarrow \mu_{\tau_3} &= \ln(\Psi) - \sigma_0^2 . \end{aligned} \tag{5.63}$$

As it has previously been shown that exponential decay of individual NoV loads due to depuration does not impact the magnitude of a lognormal distribution's scale parameter σ (see Equation 5.56), we can state that the value of σ at time τ_3 is

$$\sigma_{\tau_3} = \sigma_0 . \tag{5.64}$$

Finally, we can substitute Equations 5.63 and 5.64 into the definition of the arithmetic mean of the lognormal distribution (Equation 5.46); doing so will provide an equation for the arithmetic mean of the lognormally distributed NoV burden across a shellfish population that has conformed to the constraints ϕ and Ψ:

$$\begin{aligned} \bar{x}_{\tau_3} &= \exp\left\{\mu_{\tau_3} + \frac{1}{2}\sigma_{\tau_3}^2\right\} \\ \Rightarrow \bar{x}_{\tau_3} &= \exp\left\{\ln(\Psi) - \sigma_0^2 + \frac{1}{2}\sigma_0^2\right\} \\ \Rightarrow \bar{x}_{\tau_3} &= \Psi \exp\left\{-\frac{1}{2}\sigma_0^2\right\} \end{aligned} \tag{5.65}$$

This equation describes the average NoV load required to conform to the constraints ϕ and Ψ, and is redesignated as

$$\overline{\Psi} = \Psi \exp\left\{-\frac{1}{2}\sigma_0^2\right\} \tag{5.66}$$

where $\sigma_0 = \sqrt{2}\mathrm{erf}^{-1}(2\phi - 1)$ from Equation 5.59.

This describes the arithmetic mean of the distribution of NoV that satisfies the control parameters Ψ and ϕ as per the model in Chapter 3. This term

is shown in Figure 5.7 as the target mean NoV load to be attained via the depuration process. Note that Figure 5.7 also shows that $\overline{\Psi}$ can be stated as

$$\overline{\Psi} = \bar{x}_0 \exp\{-b\tau_3\} \,. \tag{5.67}$$

5.4.10.6 τ_4 : *Minimum Depuration Time of* $\bar{z}_t$

Equation 5.48 states that the mean total pathogen load for any time t can be calculated from $\bar{z}_t = \Omega_t \bar{z}_0$, while Figure 5.7 shows the dynamics of the mean observable and mean total pathogen loads during depuration. From this plot and Equation 5.48, we can state that the minimum depuration time of the total NoV load for a population (here designated as τ_4) will occur when

$$\bar{z}_{\tau_4} = \overline{\Psi} \,. \tag{5.68}$$

Equations 5.48 and 5.66 can be substituted in to obtain

$$\begin{aligned} \Omega_{\tau_4} \bar{z}_0 &= \Psi \exp\left\{-\frac{1}{2}\sigma_0^2\right\} \\ \Rightarrow \Omega_{\tau_4} &= \frac{\Psi}{\bar{z}_0} \exp\left\{-\frac{1}{2}\sigma_0^2\right\} \,. \end{aligned} \tag{5.69}$$

Here Ω_{τ_4} is a function of the mean, total NoV load's minimum depuration time (τ_4). We have previously shown that analytical solutions for Ω_t are not attainable, therefore numerical methods must be used to obtain values for τ_4.

The value of τ_4 obtained from Equation 5.69 is the minimum depuration time of the shellfish population, taking into account the unobserved and observed NoV loads.

5.4.11 *Minimum Depuration Time — Compartmental vs Exponential Decay Model*

The compartmental model developed in this Chapter is based on the assumption that the total NoV load per shellfish is always at least equal to or greater

than the observable NoV load, i.e. $z_t \geqslant x_t$. These values are equal only when $A = 1$, i.e. when there is no unobservable load present. The proportion of the observable load present is described by the parameter A, and this value is fixed across the population of shellfish being depurated; this allows us to state that the mean of the total load will also at least always be equal to or greater than that of the mean of the observable load. It follows that the minimum depuration time of $\bar{z}_t$ will also at least always be equal to or greater than that for $\bar{x}_t$.

The magnitude of $\tau_4 - \tau_3$ corresponds to the additional depuration time required to conform to the constraints Ψ and ϕ, by taking into account the compartmentalisation of NoV within the observable and unobservable compartments, and is further analysed below.

The compartmental model constructed in this Chapter is posited on the hypothesis that there exists an unobservable tranche of NoV within each shellfish that is not captured by current testing methods. Therefore in this section we compare the behaviour of the compartmental model from this Chapter with that of the depuration model constructed in Chapter 3. We compare the initial, pre-depuration distribution shapes for both models and go on to examine how they behave during depuration.

We are also interested in the early pathogen dynamics during depuration, as any unobservable NoV load in each shellfish will have begun to traverse into the observable compartment according to the internal transfer rate k. The sensitivity of the parameters used in the compartmental model is also analysed. Finally, and most importantly, the value of the total NoV load's minimum depuration time, τ_4, is compared with that of the depuration model τ_3.

5.5.1 *Parameterisation of Compartmental Model*

Parameter	Literature value	Regression value [c]	Parameter value applied	Impacts time:
A	N/A	0.461	0.461	τ_1, τ_4
b	0.01339 [a]	0.00398	0.01339	τ_1, τ_3, τ_4
k	N/A	0.07453	0.07453	τ_1, τ_2, τ_4
$\bar{x}_0$	1064 [b]	191245	1064	τ_3,τ_4
ϕ	0.95	N/A	0.95	τ_3, τ_4
μ_0	5.617	N/A	5.617	τ_3, τ_4
σ_0	1.645	N/A	1.645	τ_3, τ_4
Ψ	200	N/A	200	τ_3, τ_4
p	N/A	N/A	0.99	τ_2

Table 5.1: Table of parameters and values derived from literature and non-linear least squares regression. The salient times impacted by changes in parameters values are also noted.

[a] [7] [b] [2] [c] All values obtained from Neish data using 'nls' function in R [39]

In Chapter 3, the depuration model was parameterised from the literature [7, 75] and available data (Appendix C). The compartmental model has been constructed on the foundation of the depuration model, with the additional assumption that each oyster has a proportion of pathogen that is not observed by current testing. The magnitude of this observable proportion (A) fixed across the population of oysters is currently unknown, due to the fact that the unobservable compartment is currently 'unobservable'.

An internal transfer rate from the unobservable to observable compartments has also been incorporated into the model, with this transfer rate previously designated as the parameter k. Again, there is no experimental data that has previously informed a value of this internal transfer rate. Therefore, other methods must be deployed to derive estimates of these two parameters.

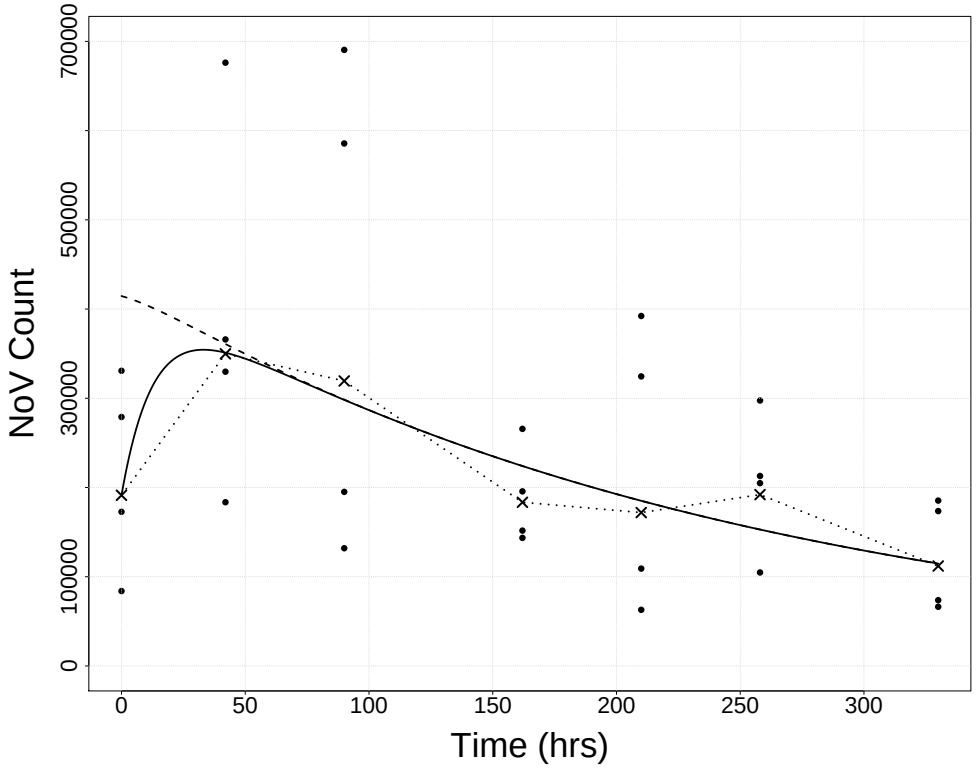


Figure 5.8: Plot of $\bar{x}_t$ highlighted by (——), and $\bar{z}_t$ highlighted by (– – –), fitted to $t = 0$ and $t = 320$ Neish data shown in Figure 5.2. Parameters used are $A = 0.461$, $\phi = 0.95$, $k = 0.07453$, $b = 0.01339$

Parameter value estimates for A and k have been derived from the data provided by Neish's experiments carried out in 2013 and which were previously shown in Figure 5.2 [39]. Parameter estimation was carried out using non-linear least squares regression on the Neish data using the software package R, and specifically the 'nls' command from the 'stats' software package for R. This code is provided in full in Appendix A.1.5, with parameters obtained

from the data shown in Table 5.1. These regression parameters are displayed alongside the parameters carried over from literature in Chapter 3. Where parameter values are obtained from both the literature and regression, the literature parameters have been used in the following results.

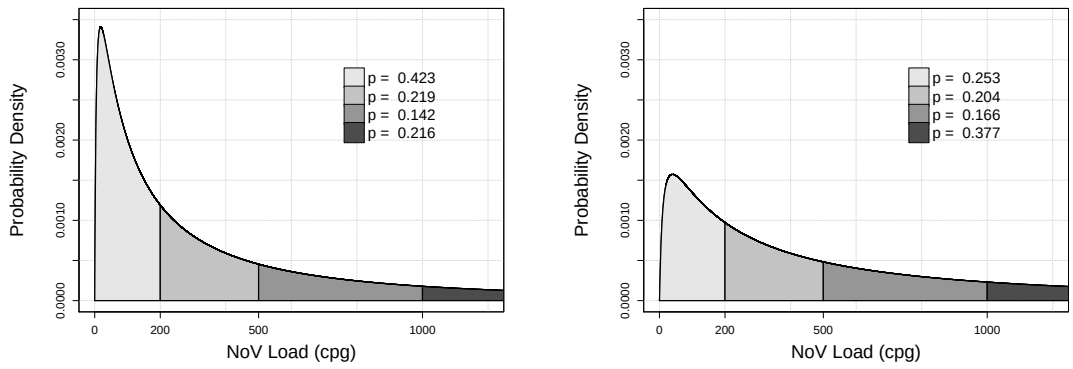
The regression derived value of $\bar{x}_0$ shown in Table 5.1 is significantly greater than that of the value of $\bar{x}_0$ deployed in Chapter 3. This is due to the regression value being derived from the artificially seeded oyster samples, which were exposed to artificially high levels of effluence before depuration to ensure that all samples would return NoV test result values above the limit of quantitation (10 NoV cpg).

Values of the parameter estimates $b = 0.01339$ and $k = 0.07553$ (cf. Table 6.2) allow us to also obtain a value for A using Equation 5.22 of $A = 0.848$.

Note that this value of A is greater than that obtained from the data supplied by Neish [39], where a value of $A = 0.461$ was obtained using regression techniques on time series of depuration data. The data used to obtain the regression value of A was noisy and based on the geometric mean value obtained from samples of only 4 measurements for each time point measured. Due to the sample size and the variability of the measurements at the timepoint $t = 0$ in Figure 5.2, the regression value of A could have been skewed somewhat. Larger samples sizes would be preferable if applying the same regression technique in future. However, we will deploy a value of $A = 0.461$ for our results simulations in the rest of this chapter.

5.5.2 *Comparison of Initial Distributions*

Equations 5.16 – 5.20 describe the depuration dynamics of water-borne pathogens within oysters, providing descriptions of both the observable and unobservable compartments across a population of oysters. If any unobservable NoV load does exist at the pre-depuration stage, (i.e. $A \neq 1$ when $t = 0$) then the pre-depuration distributions of the observable and total NoV loads would be anticipated to exhibit different shapes. This is due to the relationship between the observable and total load PDFs, $P(X_0 = x_0) = A^{-1}\ P(Z_0 = z_0)$ (cf. Equation 5.23).



(a) Pre-depuration observable distribution, $P(x_0)$ (b) Plot of initial total distribution, $P(z_0)$

Figure 5.9: Parameters used: $\mu_0 = 5.617$, $\sigma_0 = 1.645$ and $\bar{x} = 1064$ cpg, with the proportion of observable NoV $A = 0.461$. Distributions are shown segmented into four tranches: $0 < x_0 \leqslant 200$, $200 < x_0 \leqslant 500$, $500 < x_0 \leqslant 1000$ and $x_0 > 1000$

Figure 5.9 shows the pre-depuration distributions of the observable and total NoV loads respectively, parameterised with $\mu_0 = 5.617$, $\sigma_0 = 1.645$ and $\bar{x}_0 = 1064$ cpg, with the observable load in each single oyster fixed at the proportion $A = 0.461$. The p values shown in both plots represent the probab-

ility of a randomly selected oyster having a NoV load in that particular segment.

In Figure 5.9a, the distribution of the observable compartment, $P(x_0)$, exhibits a strong, positive skewness and a high peak is observed close to zero. These contribute to the probability $p = 0.423$ of the observable distribution's population that will have a NoV load less than 200 cpg. This can be interpreted as 42.3% of randomly selected oysters having a NoV load less than 200 cpg.

Figure 5.9b plots the total NoV load at pre-depuration ($P(z_0)$). We observe a notably smaller peak than that of $P(x_0)$. This is due to the addition of the unobservable compartment to the observable compartment, which results in a horizontal axis shift towards higher values for all variates from Figures 5.9a to 5.9b. As the mode is much smaller than that of the observable load distribution, then this reduction in the area between $0 < z_0 \leqslant 200$ will result in an increase in the area of the tail ($z_0 > 200$).

NoV range (cpg)	$P(x_0)$	$P(z_0)$	Difference
$0 < x_0, z_0 \leqslant 200$	42.3%	25.3%	-17.0%
$200 < x_0, z_0 \leqslant 500$	21.9%	20.4%	-1.5%
$500 < x_0, z_0 \leqslant 1000$	14.2%	16.6%	+2.4%
$x_0, z_0 > 1000$	21.6%	37.7%	+16.1%

Table 5.2: Probabilities of an oyster having a NoV load within a particular range of values. Parameters used are $\bar{x}_0 = 1064$ cpg, $\phi = 0.95$, $A = 0.461$

The proportions of the observable and total NoV distributions for each tranche are shown in Table 5.2, with differences between the models also shown. These differences in values show that, when the unobservable loads are included, there is a significant movement towards higher NoV load values. This

conforms with sensible expectations: the inclusion of additional, sequestered pathogen loads per mollusc should result in an across-the-board increase of the total pathogen loads, and this has been borne out by the values from the compartmental model.

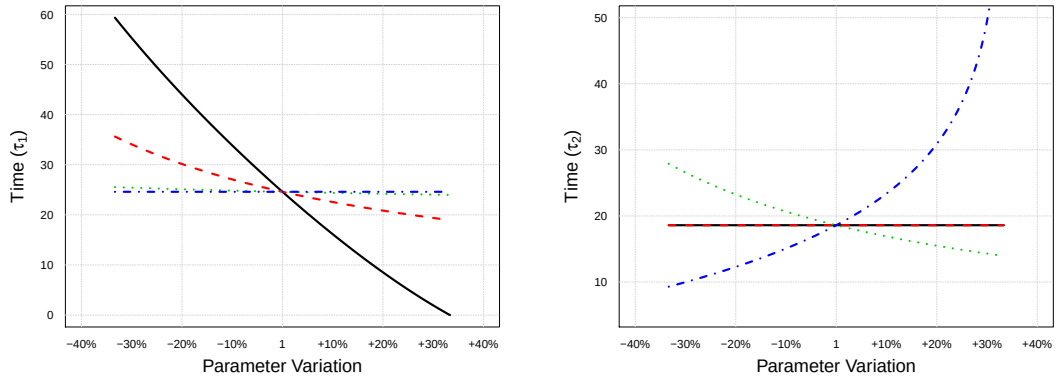
5.5.3 *Sensitivity of Salient Times to Parameter Variation*

The sensitivity of each of the salient times to variation of the four parameters A , b , k and p is shown in Figure 5.10. In each plot one of four variables is varied while holding all other parameters fixed, and the process is repeated for each of A , b , k and p. Each line pertains to the impact upon the particular salient time's value for each parameter. Each parameter has been varied across a range of $\pm 1/3$ of the starting value of each parameter being analysed.

For example, in Figure 5.10(a), the solid black line shows the impact upon the value of τ_1 of holding all other parameters constant at $b = 0.01339$, $k = 0.07453$, $p = 0.75$, $\bar{x}_0 = 1064$, $\Psi = 200$, $\phi = 0.95$ while varying the value of A across the range $A = 0.75 \pm A/3$. The other lines depicted correspond to varying the depuration rate b (– – –), the internal transfer rate k (· · ·), and the sufficient proportion parameter p (– · –).

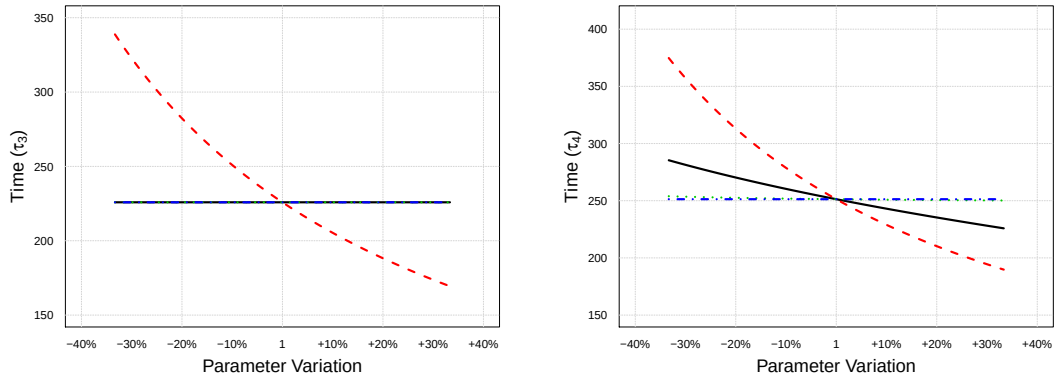
5.5.3.1 τ_1 *Sensitivity*

The value of time τ_1 corresponds to when the total NoV load has decayed during depuration to match the initial, observable pathogen load. Figure 5.10(a) highlights that any changes in A, the proportion of initial, observable NoV loads, has the biggest effect on the value of τ_1. As A increases, a marked decrease in the value of τ_1 is observed. This is readily anticipated as the right



(a) salient time τ_1 sensitivity analysis

(b) salient time τ_2 sensitivity analysis



(c) salient time τ_3 sensitivity analysis

(d) salient time τ_4 sensitivity analysis

Figure 5.10: Sensitivity analysis of parameters within compartmental model. Initial parameter values are: $A = 0.75$, $b = 0.01339$, $k = 0.07453$, $p = 0.75$, with other parameters fixed at $\bar{x}_0 = 1064$ cpg, $\Psi = 200$ cpg, $\phi = 0.95$. Variation of parameter A highlighted by (——); variation of b parameter by (– – –); k parameter by (· · ·); variation of p parameter (– · –). Times are shown in hours

hand side of Equation 5.50 ($\Omega_{\tau_1} = A$) is equal to A.

Consider when $\Omega_{\tau_1} = A$, the value of the mean, total NoV load ($\bar{z}_t$) will equal the mean, initial observable NoV load. In other terms,

$$\bar{z}_{\tau_1} = \Omega_{\tau_1} \bar{z}_0 = \bar{x}_0 \,.$$

The depuration decay rate b also has a notable effect on the value of τ_1, whereas the internal transfer rate k has a more limited impact.

The parameter p does not have any impact upon τ_1 as it is not a component of Equation 5.50.

5.5.3.2 τ_2 *Sensitivity*

The time τ_2 pinpoints when the values of the mean observable NoV load sufficiently approximates the mean total NoV load during depuration, i.e. the unobservable compartment y_t has been reduced to a sufficiently small amount that any remaining value of y_t can be reasonably ignored. The equation describing τ_2 (Equation 5.52) is solely dependent upon the parameters k and p, thus only changes to their values will affect any change to the value of τ_2.

When the parameter p is increased across the range $0.50 \leqslant p < 1$, τ_2 increases exponentially, approaching a vertical asymptote at $p = 1$. It is evident that, if $p = 1$, then $\tau_2 \to \infty$ as $\bar{x}_t$ converges towards, but never equals, $\bar{z}_t$ (refer to Section 5.4.10.3 for further details).

As k increases a relatively small decrease in the value of τ_2 is observed, demonstrating that the value of the internal transfer rate is of limited importance with respect to τ_2.

5.5.3.3 τ_3 *Sensitivity*

The salient time τ_3 is equivalent to the minimum depuration time T_{WCV} from Chapter 3, and so is not impacted by any parameters that are exclusively used by the compartmental model. Therefore τ_3 will only be altered by b, the only parameter examined here that was used in the depuration model in Chapter 3. Figure 5.10(c) confirms this, as only changes in the value of b (shown in red)

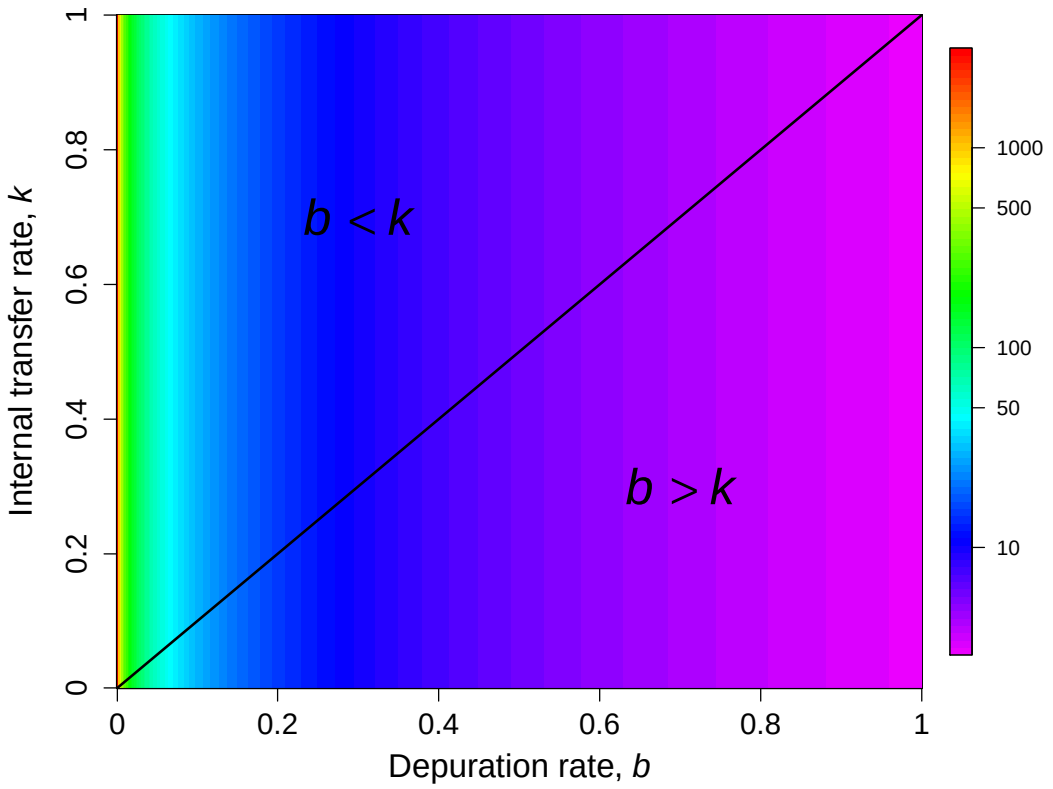


Figure 5.11: Heat plot of depuration rate b and internal transfer rate k versus τ_3, the minimum depuration time from exponential model. All other parameters fixed at $\bar{z}_0 = 1064$ cpg, $p = 0.99$, $A = 0.75$, $\phi = 0.95$, $\Psi = 200$ cpg. Note that the z-values of the plot are shown on a $\log_{10}$ scale (in hours)

results in changes in τ_3.

Clearly any increase in the depuration rate b will result in a decrease in minimum depuration time required, and this is shown in Figure 5.11. This figures shows that as $b \to 0$, $\tau_3 \to \infty$, and also further confirms that the internal transfer rate k has no impact upon the value of τ_3.

5.5.3.4 τ_4 *Sensitivity*

The minimum depuration time of the mean total NoV load is defined as τ_4, the numerical solution obtained from Equation 5.69. This equation utilises all four of the parameters (A , b , k and p) analysed in this section. In Figure

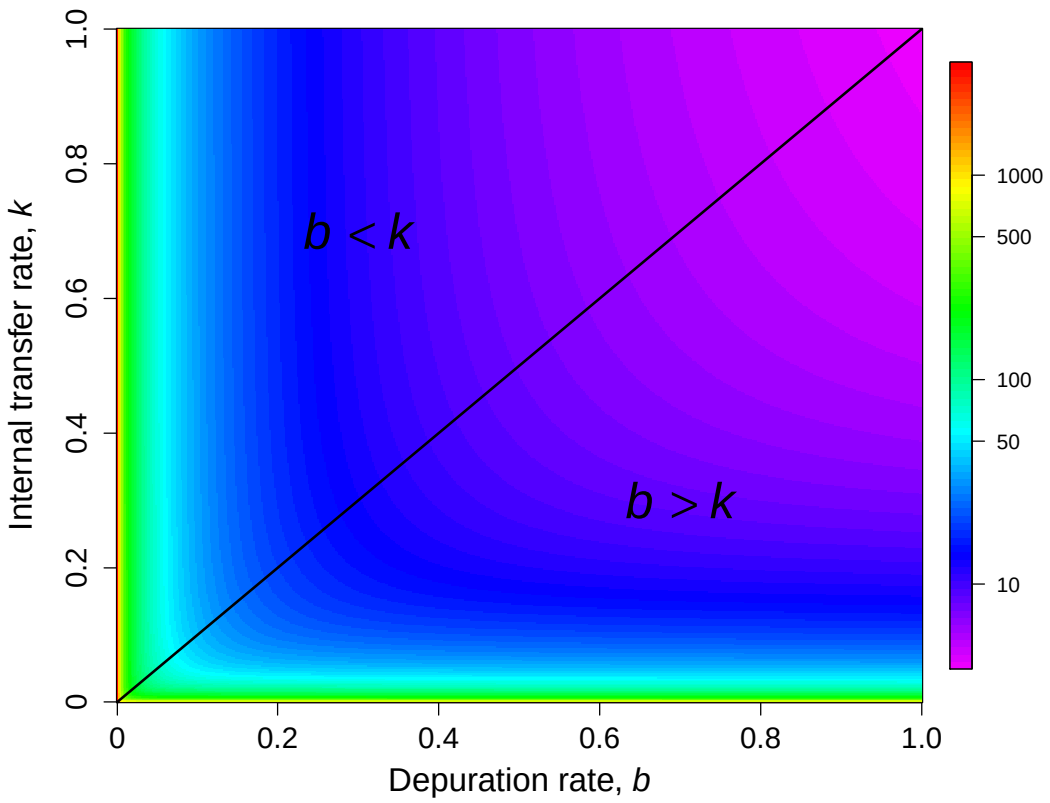


Figure 5.12: Heat plot of depuration rate b and internal transfer rate k versus total NoV load minimum depuration time τ_4. All other parameters fixed at $\bar{z}_0 = 1064$ cpg, $p = 0.99$, $A = 0.75$, $\phi = 0.95$, $\Psi = 200$ cpg. Note that the z-values of the plot are shown on a $\log_{10}$ scale (in hours)

5.10(d), each of A , b , k and p have been varied individually to test their impact upon the total NoV load's minimum depuration time. It is readily seen that increasing the depuration rate b induces a significant reduction in the value of τ_4, and so the total load's MDT is most sensitive to changes in the depuration decay rate b for the ranges stated.

The black line in Figure 5.10(d), shows that as A (the proportion of the total NoV load that is observable) increases towards 1, we see a steady decline in the value of τ_4. This is a consequence of the fact that more of the initial total NoV load is sequestered in the observable load for larger values of A, and not in the unobservable compartment. Thus there is a significantly smaller amount of

NoV which must pass through both y_t and x_t compartments. As the transition through each compartment takes time, the smaller the pathogen load which needs to pass through both compartments will result in a quicker MDT for the total load, τ_4.

As we increase the value of the parameters p and k, we see very little response in the value of τ_4, thus showing that any change in these parameter's values in the ranges examined has minimal impact on τ_4. As the sufficient proportion p increases it elicits a very slight increase in the value of τ_4. Any increase in the internal transfer rate k only results in a very small decrease in the value of τ_4.

The sensitivity analysis shown in Figure 5.10 only considers the sensitivity of the salient times τ_i across narrow ranges of values for each parameter. There we analysed the decay and transfer rates in the range $0.00893 < b < 0.01785$ and $0.0497 < k < 0.09937$ respectively. Consideration of how the decay parameter b impacts depuration times across a wider range of values is undertaken now, along with an examination of the internal transfer rate k across an expanded range.

Figure 5.12 shows the behaviour of τ_4 as both b and k operate across the larger range of $(0, 1]$. This plot shows that significantly increasing the values of both b and k would result in reduced depuration times. However, we have already discussed the feasibility of whether $b > k$ is biologically relevant (see Section 5.4.3), and so we need only consider the top left triangle of the figure, as the values above the diagonal $b = k$ black line correspond to when $b < k$. When either of the parameter's $b, k \to 0$, it is seen that $\tau_4 \to \infty$.

There is no mathematical reason to restrict $b,\ k \leqslant 1$. However Figure 5.12 shows that if both rates are greater than 1 then the MDT is unrealistically fast at less than 1 hour, much less than the current legislated MDT of 42 hours for Class B harvests.

5.5.3.5 $\tau_4 - \tau_3$ *Sensitivity*

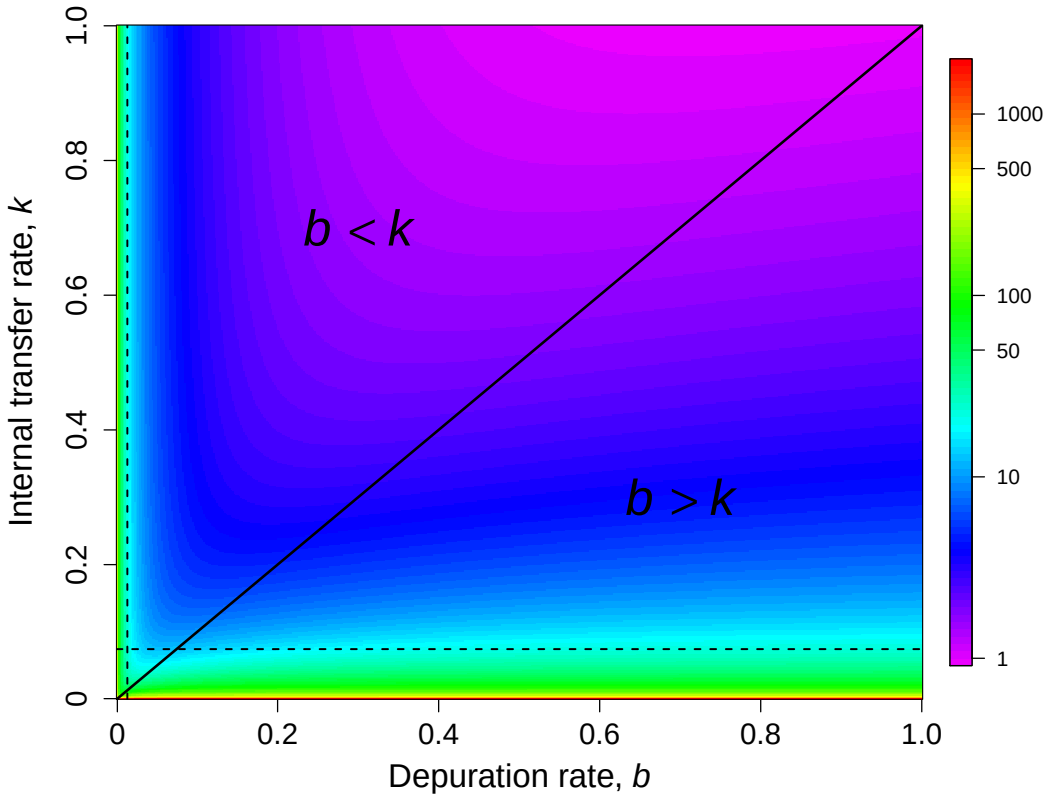


Figure 5.13: Heat plot of depuration rate b and internal transfer rate k versus $\tau_4 - \tau_3$, the additional depuration time required when including the unobservable NoV load. All other parameters fixed at $\bar{z}_0 = 1064$ cpg, $p = 0.99$, $A = 0.75$, $\phi = 0.95$, $\Psi = 200$ cpg. Note that the z-values of the plot are shown on a $\log_{10}$ scale (in hours)

Figure 5.13 shows the behaviour of $(\tau_4 - \tau_3)$, which corresponds to the additional depuration time required due to the inclusion of the unobservable compartment in our model, for $0 < b,\ k \leqslant 1$. It is apparent that $(\tau_4 - \tau_3) \to \infty$

when either $b \to 0$ or $k \to 0$, which is a consequence of the behaviour of τ_4 shown in Figure 5.12.

The minimum value of $\tau_4 - \tau_3$ observed at approximately $0.5 < b < 1$ where $k \approx 1$ is noteworthy here. So far we have not obtained a satisfying hypothesis as to why the relationship between b and k provides the contour 'funnel' observed at these values. This may be a consequence of the numerical methods used in calculating $\tau_4 - \tau_3$, but seems unlikely as the contour is consistently decreasing towards $(\approx 0.7, 1)$. Further analysis on this is required to understand this result.

Industry stakeholders have no control over the internal transfer parameter k; its value is a consequence of the biology of the shellfish. The value we have obtained by regression from the Neish data ($k = 0.07453$) is highlighted on Figure 5.13, along with the depuration decay rate we have obtained via the literature ($b = 0.01339$). Where these lines cross in Figure 5.13, it is shown that a minimum depuration time of approximately 100 hours would be required. Both the values of b and k would need to be significantly increased to reduce depuration times although, as a minimum of 42 hours in depuration is currently required by law, they would only need to be increased by a small factor.

5.5.4 *Analysis of* $(\tau_4 - \tau_3)$

In this section we especially note how the value of $(\tau_4 - \tau_3)$ is impacted by varying the four parameters A , b , k and p. The four salient times are restated for convenience below:

$$\Omega_{\tau_1} = A \tag{5.70}$$

$$\tau_2 = -\frac{1}{k}\ln(1-p) \tag{5.71}$$

$$\tau_3 = b^{-1}\left[\left(\text{erf}^{-1}\left(2\phi - 1\right)\right)^2 + \ln\left\{\frac{\bar{x}_0}{\Psi}\right\}\right] \tag{5.72}$$

$$\Omega_{\tau_4} = \frac{\Psi}{\bar{z}_0}\exp\left\{-\frac{1}{2}\sigma_0^2\right\} \tag{5.73}$$

The impact of varying the two parameters (ϕ and Ψ) on minimum depuration times was investigated in the previous Chapter 3.

5.5.4.1 *Varying parameter* A

$\boldsymbol{A}$	0.50	0.60	0.70	0.80	0.90	1.00
$\boldsymbol{\tau_1}$	**59**	**44**	**31**	**19**	**9**	**0**
τ_2	62	62	62	62	62	62
τ_3	226	226	226	226	226	226
$\boldsymbol{\tau_4}$	**285**	**270**	**257**	**246**	**235**	**226**
$\boldsymbol{\tau_4 - \tau_3}$	**59**	**44**	**31**	**20**	**9**	**0**

Table 5.3: Results of varying proportion of pre-Depuration total NoV load which is observable (A). Fixed parameters are: $b = 0.01339$, $k = 0.07453$, $\bar{x}_0 = 1064$, $\Psi = 200$, $\phi = 0.95$, $p = 0.99$. Impacted salient times (τ_i) and values (in hrs) are highlighted in bold

Table 5.3 shows the impact of varying A, the proportion of the intial observable NoV load. As expected, τ_1 is significantly impacted by variations in the

value of A. As τ_2 and τ_3 are not derived from A, no change in their values are observed when A is increased. However τ_4 is also altered by changes to the value of A. The reason for this is not apparent from the definition of τ_4 as it stands in Equation 5.73.

From Table 5.3 note that, when $A = 0.50$, the minimum depuration time of $\bar{z}_t$ (τ_4) is only 26% longer than that for τ_3, the minimum depuration time of the depuration model in Chapter 3. Figure 5.14 shows the impact on τ_4

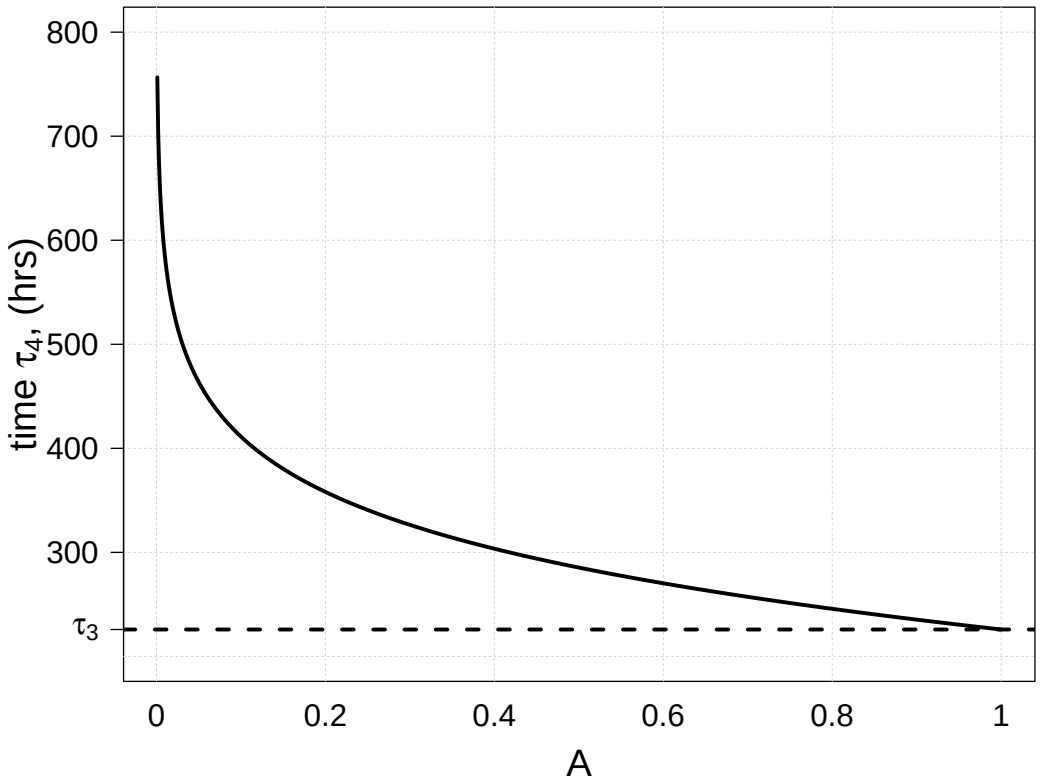


Figure 5.14: Plot of proportion of initial, observable NoV load A versus total NoV load minimum depuration time τ_4. All other parameters fixed at $\bar{x}_0 = 1064$ cpg, $p = 0.99$, $k = 0.07453$, $b = 0.01339$, $\phi = 0.95$, $\Psi = 200$ cpg

across the full range of possible values of A. When A is low we observe a large value for τ_4. As $A \to 1$, the resultant decrease in τ_4 tends towards the value of τ_3. When $A = 1$ (i.e. there is no initial, unobservable load) Figure 5.14 shows that the value of the minimum depuration time for the total NoV load equals

the minimum depuration time from the exponential model. In other words, $\tau_3 = \tau_4$ when $A = 1$.

5.5.4.2 *Varying parameter* p

$\boldsymbol{p}$	0.90	0.92	0.94	0.96	0.98	0.99
τ_1	59	59	59	59	59	59
$\boldsymbol{\tau_2}$	**31**	**34**	**38**	**43**	**52**	**62**
τ_3	226	226	226	226	226	226
τ_4	284	284	285	285	285	285
$\tau_4 - \tau_3$	58	58	59	59	59	59

Table 5.4: Results of varying sufficient proportion (p) expelled from unobservable compartment y_t. Other fixed parameters are: $b = 0.01339$, $k = 0.07453$, $\bar{x}_0 = 1064$, $\Psi = 200$, $\phi = 0.95$, $A = 0.5$. Impacted salient times (τ_i) and values (in hrs) are highlighted in bold

From Table 5.4, it can be seen that changes to the value of p affect the value of τ_2, which is apparent from Equation 5.71. Neither τ_4 or τ_3 are dependent upon p, and so $\tau_4 - \tau_3$ is not affected by any changes to its value.

5.5.4.3 *Varying parameter* b

Increasing the value of the depuration rate b results in an expected decrease in the significant times τ_1, τ_3 and τ_4. Chapter 3 showed that an increase in the depuration decay rate by some proportion ρ results in a $\rho/(1+\rho)$ proportional reduction in the minimum depuration time. This is borne out in the times for τ_3 in Table 5.5. This reduction is approximated by τ_4, but exhibits a smaller reduction in minimum depuration time. This is due to the inherent difference in the decay rates between τ_3 and τ_4, which have different decay rates, $\exp\{-bt\}$ and Ω_t respectively.

b	b	1.1b	1.25b	1.5b	1.75b	2b
τ_1	**59**	**55**	**49**	**42**	**37**	**33**
τ_2	62	62	62	62	62	62
τ_3	**226**	**205**	**181**	**151**	**129**	**113**
τ_4	**285**	**260**	**230**	**193**	**167**	**148**
$\tau_4 - \tau_3$	**59**	**55**	**49**	**42**	**38**	**35**

Table 5.5: Results when depuration/excretion rate (b) is varied. Other fixed parameters are: $p = 0.99$, $k = 0.07453$, $\bar{x}_0 = 1064$, $\Psi = 200$, $\phi = 0.95$, $A = 0.5$, $b = 0.01339$. Impacted salient times (τ_i) and values (in hrs) are highlighted in bold

5.5.4.4 *Varying parameter* k

k	0.5k	0.75k	k	1.25k	1.5k	2k
τ_1	**67**	**62**	**59**	**58**	**57**	**55**
τ_2	**124**	**82**	**62**	**49**	**41**	**31**
τ_3	226	226	226	226	226	226
τ_4	**295**	**288**	**285**	**284**	**282**	**281**
$\tau_4 - \tau_3$	**69**	**62**	**59**	**58**	**56**	**55**

Table 5.6: Results when internal transfer rate (k) is varied. Other parameters applied to model are: $k = 0.07453$, $b = 0.01339$, $\bar{x}_0 = 1064$ cpg, $\Psi = 200$, $\phi = 0.95$, $A = 0.5$, $p = 0.99$. Impacted salient times (τ_i) and values (in hrs) are highlighted in bold

The results of varying the value of k, the internal NoV transfer rate, show that this parameter has only a small impact on the salient times (Table 5.6). A four-fold increase in the value of k only resulted in a reduction of 14 hours in

the minimum depuration time for $\bar{z}_t$. This is significant due to the fact that this rate cannot currently be measured, and only by estimation using numerical methods has an approximation of k been obtained.

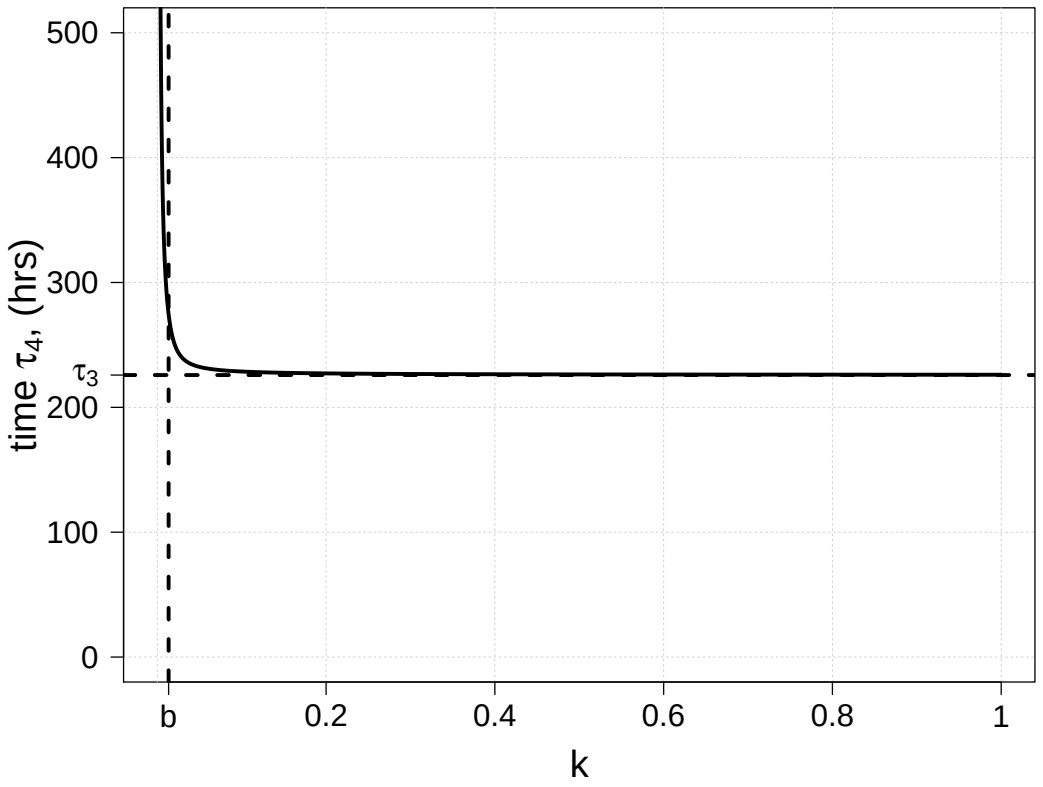


Figure 5.15: Plot of internal transfer rate k versus total NoV load minimum depuration time τ_4. All other parameters fixed at $\bar{x}_0 = 1064$ cpg, $p = 0.99$, $A = 0.75$, $b = 0.01339$, $\phi = 0.95$, $\Psi = 200$ cpg

So far we have analysed k in the range of 0.03765 ($0.5 \times k$) — 0.14906 ($2 \times k$). Further analysis of this parameter is carried out by looking at the full range of values for $k > 0$, as a more complete understanding of how k affects the value of τ_4 is required to be able to ascertain the full significance of this parameter. Figure 5.15 shows the behaviour of τ_4 for the range of $0 < k \leqslant 1$, and shows that the total NoV load minimum depuration time τ_4 approaches a vertical asymptote as $k \to 0$. In fact, as $k \to b$, the value of τ_4 begins to increase rapidly.

It was previously stated in Section 5.4.1 that $k \leqslant b$ does not make biological sense, something again borne out by this analysis.

This chapter has proposed an extended model describing the depuration dynamics of NoV within a population of shellfish, based upon the depuration model detailed in Chapter 3. The previous model has been extended by taking into account an assumption of internal sequestration of pathogens within individual shellfish, splitting the pathogen load into unobservable and observable compartments. This approach has been undertaken based on the evidence in the current literature that shows that pathogen loads are not centralised in the digestive gland but are found more distributed throughout the animal's whole biology. This compartmentalisation has implications for current detection regimes of NoV, and specifically for NoV detection within oyster species.

Parameterisation of the compartmental model has been based in part on values from the depuration model from Chapter 3. We have obtained reasonable estimates for k, A and b from regression techniques applied to the Neish data, and stated that the value of p should be $\lesssim 1$. However, the value of the internal transfer rate k stated in Table 5.1 must conform to $b < k$ for biological reasons. This is due to the fact that, if $b \geqslant k$, then the internal transfer rate would become the rate determining step of the depuration process. If this occurred then the digestive gland (compartment x_t in this model) would not be the primary repository of NoV in shellfish, which would be contrary to literature findings [29, 30].

For the sensitivity analysis carried out in Section 5.5.3, the impact of k upon any of the salient times τ_i is very limited, inducing only small variation in the values of τ_1, τ_2 and τ_4. This is significant to our methodology here as, in the absence of being able to obtain reliable estimates of k, we have had to

fall back upon indirect methods to derive a parameter value for k, and any inaccuracies involved in our estimate of k would only result in small changes to the model's results.

The salient result here is any observed increase in the length of minimum depuration time required due to an increase in the initial unobservable load proportion $1 - A$. This increase in the length of minimum depuration time due to compartmentalisation of the pathogen, represented by $(\tau_4 - \tau_3)$ in our model, is shown to be most responsive to changes in the value of A, the proportion of initial NoV loads which are observable to current testing practices. Figure 5.14 shows that at low values of A, minimum depuration times of τ_4 are much greater than that of τ_3. Only as $A \to 1$ do we observe the minimum depuration time of the total NoV load approach the minimum depuration times using the exponential model from Chapter 3.

Not only does this compartmental model show that consideration of internal pathogen sequestration must be made, but also that the timing of pathogen detection testing be undertaken with some caution. The salient time τ_1 indicates when the total NoV load has decayed during depuration to match the initial observable load. This time is most significantly impacted by the value of A, as shown by the black line in Figure 5.10(a). That plot shows that for the range $0.50 \leqslant A \leqslant 1$, τ_1 decreased steadily from 59 to 0 hours. Thus, if the value of A is low then any pathogen testing during depuration should not be carried out within the first hours of the process. In actuality, only post-depuration testing for the presence of NoV would be prudent, as at this point would any difference between the observable and total pathogen loads have approximated zero. Further studies are required to validate our hypothesis of pathogen compartmentalisation and especially with regards to NoV in oysters.

6

STOCHASTIC WATER-BORNE PATHOGEN MODEL

6.1 INTRODUCTION

This chapter presents a mathematical model describing how pathogens are transmitted into shellfish in a marine environment. We consider rainfall events that can overwhelm the processing capacity of sewage treatment works (STW), resulting in discharges of effluent into the immediate waters of the STW. Any shellfish farms that are in the proximity of the STW will experience an increase in human faecal contaminant levels in their waters, thus presenting an increased likelihood of any shellfish ingesting pathogens such as *E. coli*, NoV and other enteroviruses that may be present in the discharged effluence [13, 33, 76].

In this chapter, a mathematical model is constructed which can provide linkage between the discharge levels from an STW and the pathogen exposure levels at a shellfish farm located in close proximity. The model constructed to describe rainfall events and a STW's capacity to process them are based on available data and literature [77, 78, 79, 80]. We incorporate into this model a variable capacity for the STW to cope with significant rainfall, and examine how varying the overflow capacity impacts the levels of pathogens within harvested shellfish, as well as how this alters minimum depuration times (MDT's). We then adapt the compartmental model of depuration from Chapter 5 into a compartmental model that describes the dynamics of pathogens loads across a population of shellfish while still at a farm and before harvesting. We analyse

the results of the model, paying special consideration to the distribution of pathogens within shellfish harvests, and how varying the overflow capacity impacts upon the pathogen levels within shellfish populations.

Part of the motivation behind the construction of this model is to investigate the origin of the variability of water-borne pathogens, in an attempt to ascertain whether the random rainfall events which result in storm overflow discharges at STWs contribute in some way towards the observed positive skewness of the distributions of water-borne pathogens.

We initially examine the impact on pathogen levels in a population of oysters at the point of harvest when the treatment capacity of a sewage treatment works (STW) is varied. We subsequently analyse the trade-off between STW capacity levels and the minimum depuration time (MDT) required post-harvest time to conform to control parameter requirements.

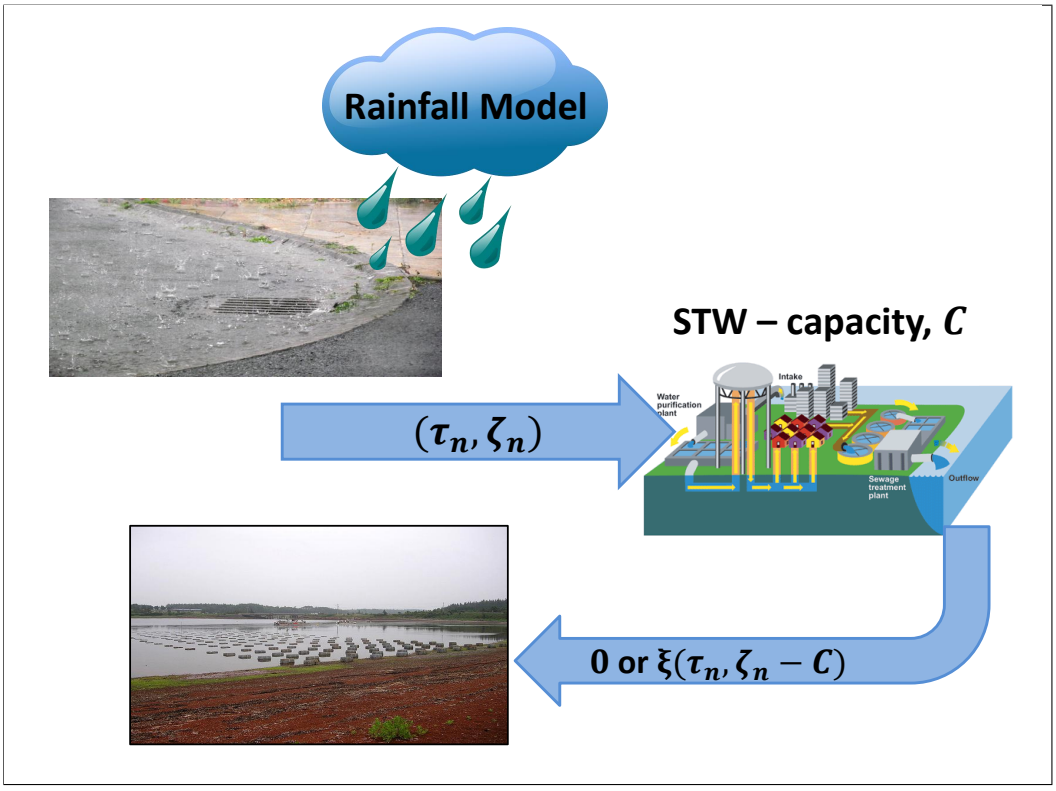
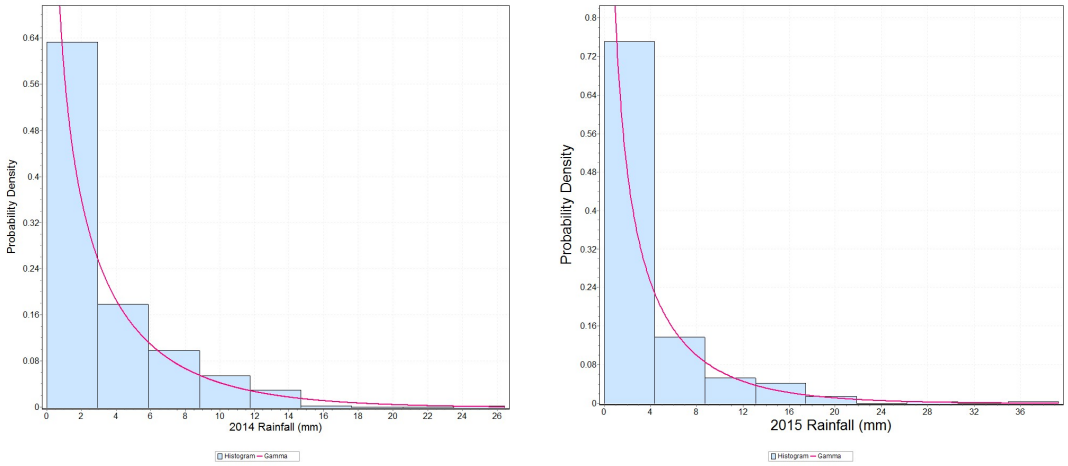


Figure 6.1: Rainfall to STW to shellfish farm diagram

To construct a mathematical model that describes how rainfall events can cause an increase in the levels of pathogens passing into the waters of shellfish farms, we need to first construct a model that accurately describes rainfall levels and its frequency.

6.2.0.1 *Gamma Distribution of Rainfall Levels*



(a) 2014 Rainfall

(b) 2015 Rainfall

Figure 6.2: PDF's of 2014 and 2015 daily rainfall levels in England and Wales. Gamma distribution curves fitted using parameters from Table D.1

Significant research has been carried out in describing rainfall levels around the world. In the UK and many other countries rainfall levels have been shown to be well described by a gamma distribution [78, 81, 80, 82, 83]. We have analysed rainfall data for England and Wales for the years 2006–2015, applying goodness of fit tests to the data for each year for a wide range of distribution types, to check the appropriateness of using a gamma distribution to inform rainfall levels for our model.

Figure 6.2 shows histograms with fitted gamma distributions for rainfall in England and Wales in 2014 and 2015. Rainfall has been plotted using a scale of per day, the timescale supplied by the Met Office [84]. Appendix D includes similar plots for the years 2006–2013, plus goodness of fit rankings for the gamma distribution when compared with a wide range of other PDF types. These plots, along with the current literature demonstrate that UK rainfall levels can be well described by the gamma distribution.

Section 2.3.3.1 in Chapter 2 provides equations defining the PDF of the gamma distribution. Based on Equation 2.17, we state that a random variable $X = x$ has a gamma distribution if its PDF is defined as

$$P_G(x) = \frac{1}{\theta^k \, \Gamma(k)} x^{(k-1)} \exp\left\{\frac{-x}{\theta}\right\}, \tag{6.1}$$

where $x \in (0, \infty)$ [58, p. 166], and $\Gamma(k)$ is the gamma function, defined as (cf. Equation 2.18)

$$\Gamma(k) = \int_0^\infty z^{(k-1)} \exp\{-z\} \, dz \,. \tag{6.2}$$

Equations 6.1 and 6.2 can be used to predict the level of rainfall per day in the UK, and more generally for any geographic location. This is accomplished by the use of appropriate parameter values of k and θ within Equations 6.1 and 6.2 (see Appendix D for UK parameter values for 2006–2015).

6.2.0.2 *Distribution of Rainfall Events*

The gamma distribution can describe the levels of rainfall per day, however we require another mechanism to inform the temporal frequency of rainfall i.e. how often does it rain on a particular day across a particular number of days?

Appendix D details data describing rainfall frequency within England and Wales. Looking at the data for the 30 year period of 1986–2015 (visualised in Figure 6.3), the median number of days with rainfall > 1 mm was 192, with

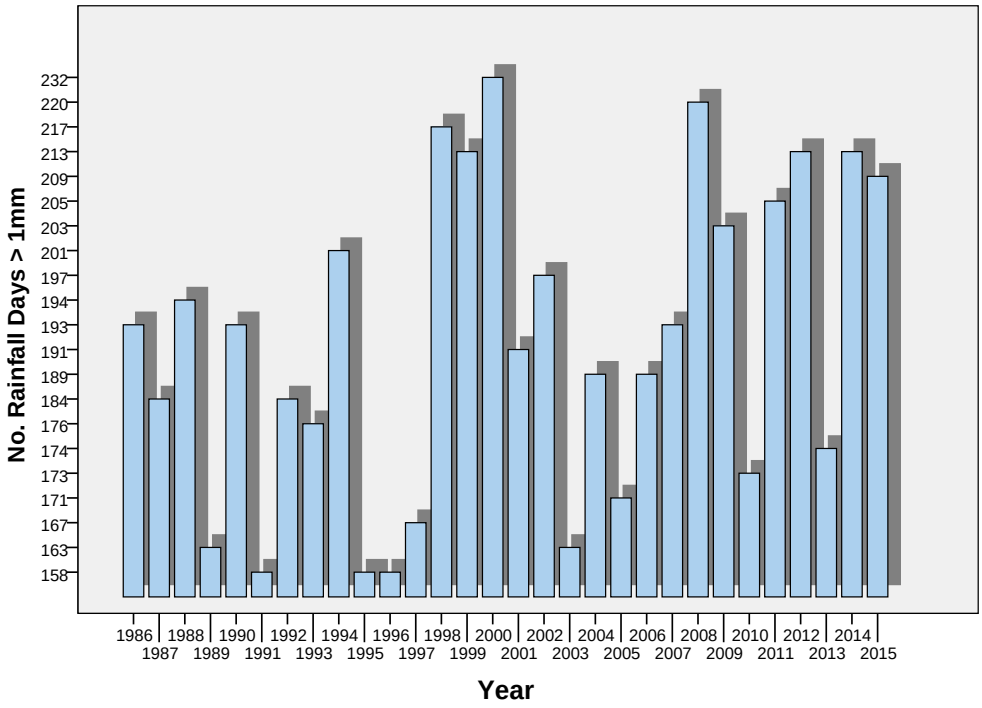


Figure 6.3: Number of days with rainfall > 1mm in England and Wales, 1986-2015

an average of 189.8 days per year. Using the median value of 192 (and taking each year as 365 days), we can estimate that each day in a particular year has probability of 192/365 = 0.5260 to have rainfall > 1 mm. This information can provide an estimate of the likelihood of rain for each day in the UK.

Whether a day has rainfall or not can be viewed as a success or failure outcome, and so can be described as a Bernoulli trial. For simplicity, we ignore any clustering of rainfall days e.g. due to seasonality of rainfall. As we have a fixed number of discrete Bernoulli trials to consider (365 in this case), we can use the binomial distribution as a means to find a random number of days with rainfall for our model. The negative binomial distribution was also considered but is not appropriate here as it is dependent on a fixed number of Bernoulli trial successes rather than a fixed number of trials.

As shown in Equation 2.23 in Chapter 2, the binomial distribution is a discrete distribution with parameters N, k and p:

$$P_B(\mathcal{T} = k) = \binom{N}{k} p^k \, q^{N-k} \qquad [61], \tag{6.3}$$

where N represents the number of days in a year, k represents the number of days that rainfall would be expected in a year, p defines the probability of any particular day being rainy, and $q = 1 - p$.

Selecting a random deviate from the probability mass function $P_B(\mathcal{T} = k)$ thus provides our model with a randomly selected number of rainfall days, $\mathcal{T}$, as defined by Equation 6.3.

6.2.0.3 *Seasonality of Rainfall*

For the purposes of our simulation, we randomly and uniformly select which days experience rainfall across each year. Thus any day of a year in the simulation is as likely as any other to experience rainfall. In actuality this does not reflect the seasonality of UK rainfall patterns which experience less rainfall frequency and intensity in the summer months. The model also does not take into account the dependence of rainfall per day upon the weather of the previous day, which often results in clustering of rainfall days in the UK.

6.2.0.4 *Rainfall Event Notation*

We now define the notation used to describe rainfall levels and occurrences within our model. A day with rainfall is denoted as τ_n, the nth day, which experiences a rainfall level of ζ_n, where $n = (1, 2, 3, \ldots, N)$ and N is the maximum number of days in the simulation. Thus each day with rainfall is described by a pairwise event

$$(\tau_n, \, \zeta_n) \, . \tag{6.4}$$

The number of days which experience rainfall is $\mathcal{T}$, obtained from the binomial distribution described above, with the distribution of rainfall days across N uniformly and randomly selected. The rainfall level per day (ζ_n) is drawn from the gamma distribution (Equation 6.1), independently of τ_n.

However, it can be readily seen that the pairwise rainfall event can be obtained by application of only the gamma distribution of rainfall levels. This can be selecting a rainfall level ζ_n for each nth day in our simulation, rather than for only each τ_n obtained from the binomial distribution described by Equation 6.3. Any days that have a rainfall level < 1mm could be set to have zero rainfall, however this would not be required due to model constructs described in the next Section (6.3).

Sewage treatment works (STWs) have only a finite capacity for dealing with the levels of waste and rain water they receive from public drainage and sewage systems [34, 48, 36]. Due to the effects of greater rainfall, an STW can experience pulsed levels of increased waste water which can overwhelm their treatment capacity. When this occurs, the capacity of a STWs can be breached and overflows of untreated sewage are released into their immediate environment[79, 44, 71]. When this occurs, there will be a spike in the levels of bacterial and viral contaminants discharged into proximal environmental waters. If this increase in contaminants are also expereinced by a nearby shellfish farm, then an increased risk of pathogens being ingested by the shellfish occurs, and thus can induce an increase in the food safety risk if the animals are subsequently harvested.

We incorporate the capacity of a STW into our model, thus taking into account the finite limitation of a STW for dealing with rainfall levels. Thus we define a STWs capacity as C, a measure of waste water volume beyond which the STW must discharge untreated liquids. We assume that, on any day without rainfall, a STW treats a constant level of sewage from public systems, and so we define C as a capacity which is over and above this baseline volume of sewage. Only when a rainfall event $(\tau_n,\ \zeta_n)$ occurs we need to consider any untreated effluence that may be discharged by the STW.

Therefore our model needs only to consider rainfall events where $\zeta_n > C$ for each rainfall day τ_n with volume C being fully processed by the STW. Thus

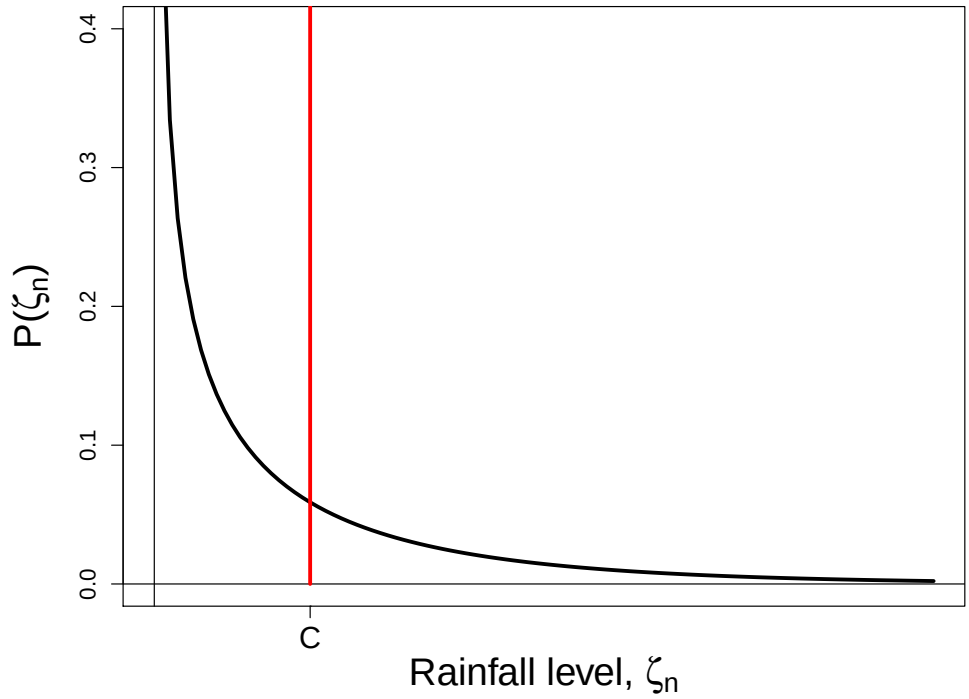


Figure 6.4: Generic rainfall levels conforming to a gamma distribution, with STW capacity C shown

the discharge of untreated waste water each day (with or without rainfall) is described by

$$\begin{cases} 0 & \text{if} \quad \zeta_n \leqslant C \\ \zeta_n - C & \text{if} \quad \zeta_n > C\,. \end{cases} \tag{6.5}$$

This allows us to state that each day with rainfall which results in an overflow discharge from a STW can be described by

$$(\tau_n,\ \zeta_n - C)\ . \tag{6.6}$$

As previously stated, our aim is to construct a model which describes the levels of pathogens which may pass into a shellfish farm's aquatic environment resulting from storm overflow processes from STWs. We have defined the events that will result in an overflow discharge as $(\tau_n,\ \zeta_n - C)$, where $n = (1, 2, 3, \ldots, N)$, and N is the maximum number of simulated days.

Discharge events which occur in rapid succession will result in a cumulative increase in these pathogens concentrations locally. Therefore, dependent upon the frequency and temporal spacing of discharge events, pathogen levels in any shellfish farms in close proximity to the STW will experience increased pathogen levels in their waters.

The number of rainfall events in one year is $\mathcal{T}$ where $0 \leqslant \mathcal{T} \leqslant N = 365$, which is dependent upon the parameters of the binomial distribution from which it has been randomly enumerated.

We must consider the natural reduction of pathogen levels within marine waters, resulting from dilution and/or dispersal of contaminants due to water currents and sedimentation. We designate the pathogen in water reduction rate as λ. This allows us to describe the concentration of pathogens within a shellfish farm's waters at time t by w_t, with the dynamics of w_t described by the differential equation:

$$\frac{dw_t}{dt} = -\lambda w_t + \sum_{n=1}^{N} (\zeta_n - C)\, \delta(t - \tau_n)\ , \tag{6.7}$$

where w_t describes the concentration of pathogen levels in the immediate water, and is a continuous function of time $t \in [0, \mathcal{T}]$. The term $\delta(t - \tau_n)$ describes a Dirac delta function,

$$\delta(n - \tau_n) = \begin{cases} 0 & \text{if} \quad n \neq \tau_n \\ \\ \infty & \text{if} \quad n = \tau_n \end{cases} . \tag{6.8}$$

Note that Equation 6.7 is comprised of both continuous and discrete components. The term $-\lambda w_t$ is the continuous component, describing the natural reduction in the current pathogen levels. The second term describes discrete influxes of pathogen levels, dependent upon the level of rainfall on a particular day.

6.4.1 *Parameter Units of Measurement*

We must ensure that the units of measurement used in Equation 6.7 are consistent with the system being described by the model. The term w_t describes pathogen concentrations in water, therefore the units of the summation term must also be in pathogen concentrations. Therefore we need to convert the measure of rainfall (in units of length) described by the summation term in Equation 6.7 into a pathogen concentration.

To do so we consider how the rainfall level passes pathogens to the STW. A level of rainfall τ_n will fall on the area of land served by the STW, with this catchment area is known as an agglomeration and indicated by L in our model. As we are assuming a linear relationship between the level of rainfall and the pathogen level passed to a STW, we can parameterise the input pathogen

concentration as ρ. Thus the total pathogen levels passed to the STW by rainfall level τ_n is described by $\rho\zeta_n$.

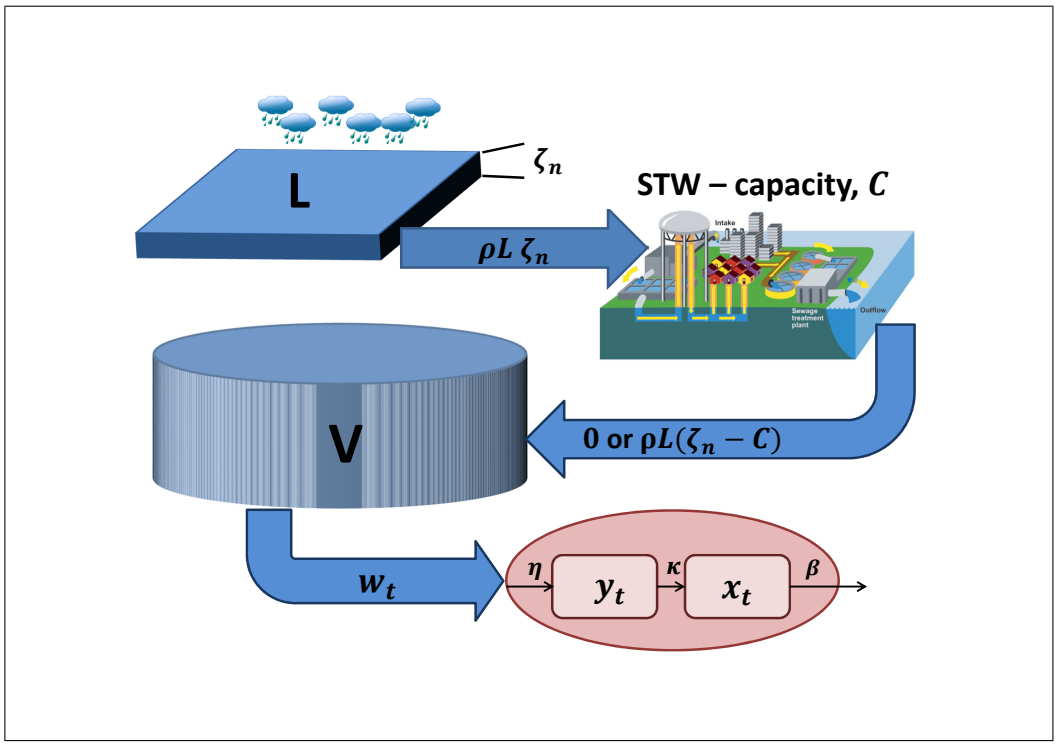


Figure 6.5: Diagram of pathogen transmission from agglomeration to STW to estuary shellfish farm per day

We now consider the pathogen levels passed into the estuary waters by a overflow discharge due to heavy rainfall. The STW will process up to its capacity C, and only allow an outflow of $\zeta_n - C$ of untreated water. Thus the STW can process pathogen levels of up to

$$\rho LC\,,$$

and so we can more accurately state the pathogen levels passed into the estuary waters are

$$\rho L(\zeta_n - C)\,.$$

We finally consider how this pathogen level is translated into a concentration in the estuary waters and thus in the shellfish farm waters. We assume that the volume V of the estuary water in which the STW and shellfish farm are located

is not increased or decreased due to the influx of overflow waters. This is a reasonable assumption as any additional water volume added to the estuary would not cause an increase in its nominal volume due to the equalisation of the water levels between the estuary and the open seawaters. Therefore we can state that any increase of pathogen concentration in the shellfish farm waters due to storm overflow discharges can be described by

$$\frac{\rho L}{V}(\zeta_n - C)\,. \tag{6.9}$$

We now include the term $\frac{\rho L}{V}$ in Equation 6.7 to fully describe the concentration levels within the estuary and shellfish farm waters:

$$\frac{dw_t}{dt} = -\lambda w_t + \sum_{n=1}^{N} \frac{\rho L}{V}\ (\zeta_n - C)\,\delta\,(t - \tau_n)\ , \tag{6.10}$$

Parameter	Description	Units
λ	pathogen water reduction rate	/day
ρ	Input pathogen concentration	copies/litre
L	Area of agglomeration	dm^2
V	Volume of estuary	litres
$\mathcal{T}$	Total no. of rain days/year	$\mathcal{T} \in \mathbb{Z}$
τ_n	Day n with rainfall event	$\tau_n \in \mathbb{Z}$
ζ_n	Rainfall level on day τ_n	dm/day
w_t	Farm water pathogen concentration	copies/litre

Table 6.1: Parameters and units used in stochastic environmental model

Table 6.1 states the units for each of the parameters that will be used within our model. Note that we have used equivalent volume measurements of 1 litre = 1 dm^3, and where 1 dm = 10 cm.

Equation 6.7 describes how the concentration of pathogens in a shellfish farm located close to a STW changes over time. The compartmental model of individual shellfish depuration dynamics described by differential Equations 5.1–5.2 in Chapter 5 is

$$\frac{dz_t}{dt} = \frac{dx_t}{dt} + \frac{dy_t}{dt}, \tag{6.11}$$

$$\frac{dy_t}{dt} = -ky_t \tag{6.12}$$

$$\frac{dx_t}{dt} = ky_t - bx_t. \tag{6.13}$$

These equations can be adapted to incorporate Equation 6.7, allowing a description of the internal dynamics of an individual, compartmentalised shellfish and while it is filter feeding in its farm environment.

In Equation 6.12, k is the internal pathogen transfer rate, whereas b is the depuration rate of each shellfish. For this model, we use the Greek letter equivalents to differentiate between the different stages of our modelling. Thus we set $\kappa = k$ and $\beta = b$ for the farm pathogen model, and both are incorporated:

$$\frac{dw_t}{dt} = -\lambda w_t + \sum_{n=1}^{N} \frac{\rho L}{V} \left(\zeta_n - C\right) \delta\left(t - \tau_n\right), \tag{6.14}$$

$$\frac{dy_t}{dt} = \eta w_t - \kappa y_t, \tag{6.15}$$

$$\frac{dx_t}{dt} = \kappa y_t - \beta x_t, \tag{6.16}$$

and where the total pathogen load per shellfish is described by

$$z_t = x_t + y_t. \tag{6.17}$$

The component ηw_t in Equation 6.15 defines the interface between the concentrations of pathogens in the marine environment and the ingestion/accretion

of pathogen concentrations by a shellfish whilst filter feeding.

This system of differential equations describes the farm-feeding dynamics of an individual shellfish, and is represented in Figure 6.6. Note that we assume full mixing of any pathogens transmitted into the environmental waters. We also assume that the volume of farm water is so large that the excretion of any pathogens by the shellfish (described by $-\beta x_t$) back into the environment has no effect on w_t.

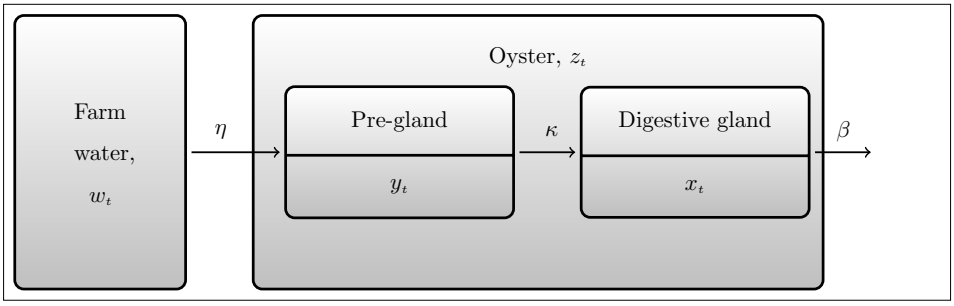


Figure 6.6: Representation of extended shellfish compartmental feeding model

Equation 6.14 is stochastic in nature, whereas Equations 6.15–6.17 are deterministic. However as Equation 6.14 instigates and drives the pathogen levels in the system, then this stochasticity will permeate through and into Equations 6.15–6.17.

This section provides parameter values from available data and literature, explaining parameter values which are directly used, as well as parameters which describe appropriate probability distributions used by the model.

6.6.1 *Rainfall Frequency and Intensity*

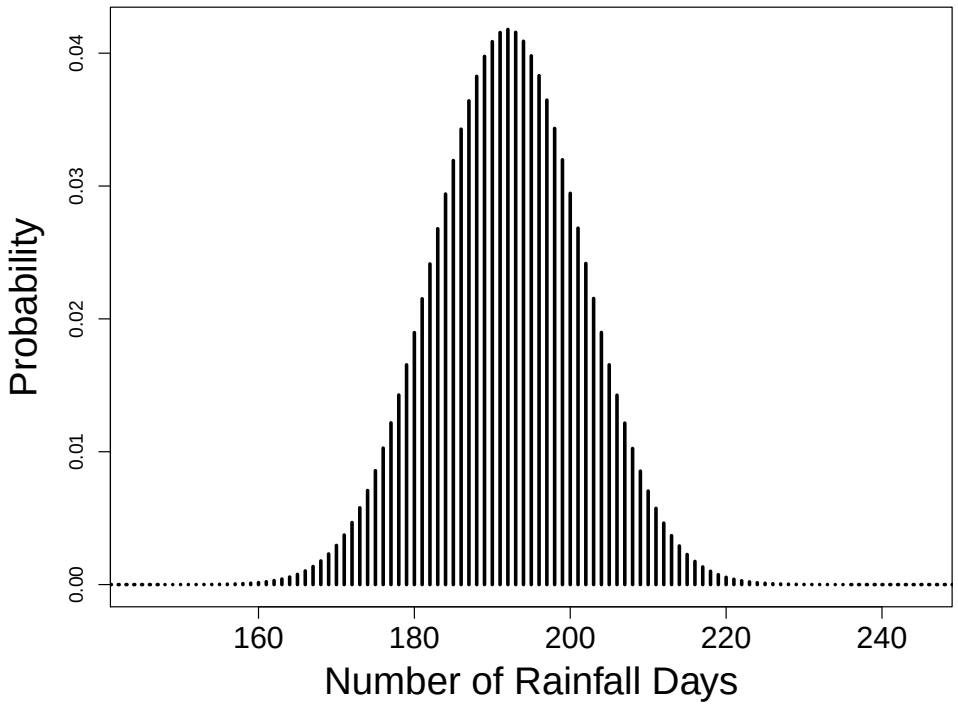


Figure 6.7: Probability mass function of binomially distributed number of rainfall days in England and Wales, based on the parameters $N = 365$, $k = 192$, and $p = 0.5260$

Equation 6.3 describes the binomial distribution, and allows us to select a random variable from this distribution using the parameter values of $N = 365$, $k = 192$, and $p = 0.5260$. These values have been obtained from the data detailed in Appendix D, where 192 is the median number of rainfall days

per year for the last 30 years. The probability mass function of this binomial distribution is shown in Figure 6.7. The random variable from this distribution yields $\mathcal{T}$, the number of days with rainfall per year for our stochastic model.

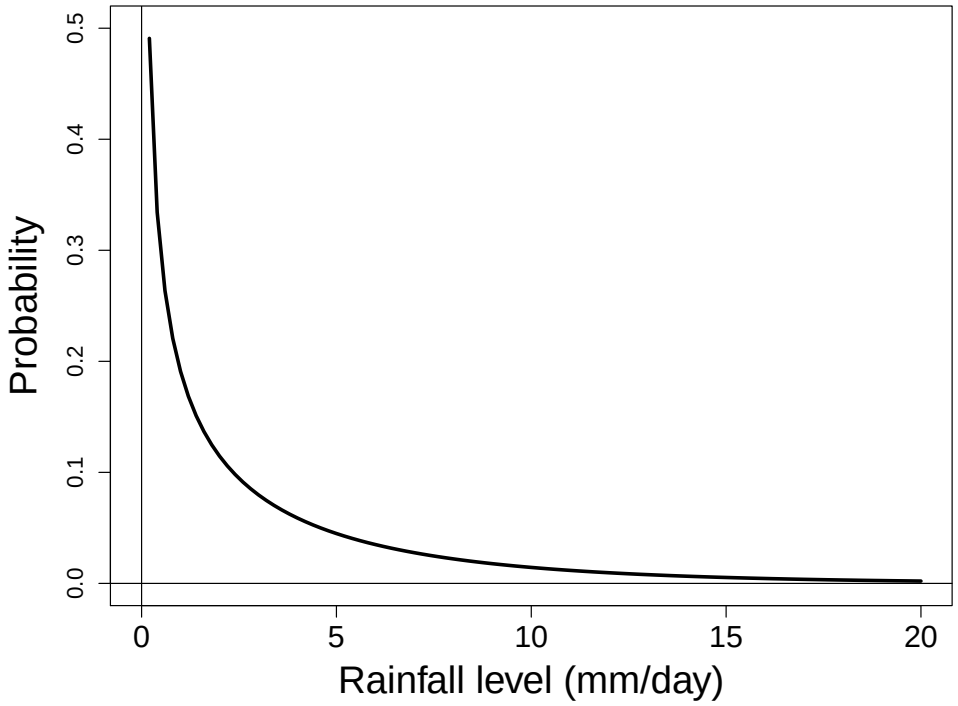


Figure 6.8: PDF of gamma distributed rainfall intensity derived from England and Wales, 2006–2015. Shape parameter $k = 0.49116$, scale parameter $\theta = 6.3675$ derived from Table D.1

Stochastic values of rainfall intensity levels ζ_n (for each rain day τ_n) are obtained from the gamma distribution shown in Figure 6.8, with shape parameter $k = 0.49116$ and scale parameter $\theta = 6.3675$ derived from Table D.1 in Appendix D. As previously stated, the model generates the number of rainfall days $\mathcal{T}$ from the binomial distribution shown in Figure 6.7. Therefore we require $\mathcal{T}$ random variables to be selected from the gamma distribution shown in Figure 6.8, one for each rainfall day τ_n.

Note that as we are operating in decimetres rather than millimetres for other length measurements, any value obtained from the above parameters applied to a gamma distribution must be divided by 100 to convert rainfall level into decimetres.

6.6.2 *Shellfish Pathogen Intake Rate*

A reasonable estimate of η, the parameter that describes the pathogen intake rate of an oyster, can be obtained from knowing the volume of water that the animal filter feeds during a particular time period.

Fox *et al* [85] reviewed previous experimental results designed to calculate the volume of water that molluscs filtrate per time period, stating that previous rates of filtration from a range of previous results were between 0.167 – 7.5 litres per hour. The authors carried out further experimentation with mussels species (a bivalve analog for oysters), arriving at a range of 2.2 – 2.6 litres per hour filtrated. We therefore reasonably assume a filtration volume rate of 2.5 litres/hour for our model, and equivalently is 60 litres/day.

6.6.3 *Environmental Water Pathogen Reduction Rate*

Bae and Schwab carried out a 2008 survey on the viral reduction rates of several enteric viruses within environmental waters [86]. This provides a NoV reduction rate of 0.03 $\log_{10}$/day, which equates to a NoV reduction rate of 0.06907 $\log_e$/day. Thus we set the water-borne pathogen reduction rate parameter at $\lambda = 0.06907$ $\log_e$/day.

6.6.4 *Area of Agglomeration*

The area of the agglomeration for a particular STW will be hugely variable; dependent upon many factors such as population density and geography [79]. Therefore any value we provide for L will be nominal and based only on best guess approach. The agglomeration area must be large enough to be financially viable for the local water company, but at the same time much less than the area of the UK. Based on literature, an average agglomeration with a population of 10000 – 20000 would have an area of 470 hectares [87]. We can round this value down to 400 hectares as not all of an agglomerations rainfall will be passed to a STW for processing. Converting this into decimetres^2, where 1 hectare = 1000000 dm^2, we obtain a value of $L = 400\,000\,000\ \mathrm{dm}^2$.

6.6.5 *Volume of Estuary*

Placing a value on the parameter V which enumerates the estuary volume is again an arbitrary exercise. The only reasonable constraint would be that the volume of the estuary must be much greater than any outflow from an adjacent STW and so we state that $V >> L\bar{\zeta}_n$, where $\bar{\zeta}_n$ is the average rainfall per day.

We can calculate an average rainfall level based on the parameters used in our gamma distribution, as well as Equation 2.20 which states the arithmetic mean of a gamma distribution as $\bar{x} = k\theta$. Using $k = 0.49116$ and $\theta = 6.3675$, we thus state that $\bar{\zeta}_n = 3.13\text{mm} = 0.0313\text{dm}$. and using the parameters decided upon previously we note that $V >> 400\,000\,000 \times 0.0313$ and so $V >> 12\,520\,000$ litres. Therefore we select a value of $V = 100\,000\,000$ litres,

but again emphasise that this parameter has been arbitrarily fixed.

Note that we assume absolute mixing of the waters in the estuary, resulting in a uniform pathogen concentration at any point in the estuary and the shellfish farm. We simulate spatial variation of concentrations within the shellfish population by running the simulation 30 times, once for each simulated shellfish sample. The stochastic nature of variables $\mathcal{T}$, τ_n and ζ_n will provide some randomness to the results.

6.6.6 *STW-Discharge Pathogen Concentrations*

We can obtain an estimate of the pathogen concentration in water in proximity to STWs from literature. In their 2005 paper, Lodder *et al* [76] stated that NoV concentration levels were in the range of 896–7499 PCR-detectable units of NoV per litre of treated sewage, and 5111–850000 PCR-detectable units of NoV per litre of untreated sewage. Flannery *et al* [88] reported a range of 29–323593 NoV copies per 100 ml for NoV genotype G_{II} after primary treatment, with a mean value of 2512 copies per litre. Equating these to the same volume measure of per litre, Flannery's range is equivalent 290–3235930 NoV copies per litre, with a mean of 25120 copies per litre.

These are broad ranges to consider when setting the input pathogen concentration parameter, ρ, for our model. For expediency, we set the pathogen concentration at $\rho = 2500$ copies per litre, based on the mean value reported by Flannery *et al* [88].

In essence we are applying a best guess approach to the parameterisation of ρ, L and V. Based on the breadth of the ranges found in the literature

that report water-borne pathogen concentrations of NoV, providing exact case study parameter values in this thesis would be difficult without extensive site surveys and fieldwork. However we are more interested in the general features of the system rather than the particular results. As stated in the introduction to this chapter, we are more concerned with how increasing the storm overflow capacity of a STW affects the pathogen loads in a nearby shellfish farm, as well as how minimum depuration times are impacted by any increase in pathogen levels.

Parameter label	Description	Parameter value used	Units
λ	pathogen water reduction rate	0.06907	/day
ρ	STW input pathogen concentration	2500	copies/litre
L	Area of agglomeration	400 000 000	dm^2
V	Volume of estuary water	100 000 000	litres
$\mathcal{T}$	Total no. of rain days/year	$\sim B(365, 0.5260)$	$\mathcal{T} \in \mathbb{Z}$
τ_n	Day n with rainfall event	$\sim U(1, 365)$	$\tau_n \in \mathbb{Z}$
ζ_n	Rainfall level on day τ_n	$[\sim \Gamma(0.49116, 6.3675)]$ /100	dm
κ	Pathogen internal transfer rate	1.78872	/day
β	Pathogen excretion rate	0.32136	/day
η	Feeding intake volume	60	litres/day

Table 6.2: Parameters used in simulation results from stochastic environmental model

6.7 RESULTS

This section presents the results obtained from the stochastic model describing the transmission of water-borne pathogens from STW storm discharges, flowing into the waters and shellfish of nearby farms. We analyse how this process impacts the pathogen levels in local waters as well as across populations of shellfish in nearby farms. We also examine the pathogen levels within the compartments of the animals, and analyse the shape of the distribution of pathogens within these compartments, comparing the shape with known distribution types.

6.7.1 *Environmental Model*

The environmental model described across Sections 6.2 – 6.4.1 was simulated using 'R' software, applying the parameters values discusseded in Section 6.6 and that are listed in Table 6.2. The simulation process is carried out in 'R' using the following steps:

1. time t is discretized to time steps of 1 day;
2. simulation is run for 1460 time steps, simulating 4 years;
3. w_t, y_t, x_t are calculated using Eulerian methods;
4. First 100 days data are removed to allow the first $(\tau_n, \zeta_n - C)$ events to take place, allowing a level of background pathogen concentration to be instigated.

The discretized equivalents of Equations 6.14 – 6.16 applied in the 'R' code are:

$$w_{n+1} = -\lambda w_n + \sum_{k=1}^{K} \xi \ (\zeta_k - C)\, \delta_{kn}\,,$$

$$y_{n+1} = \eta w_n - \kappa y_n\,,$$

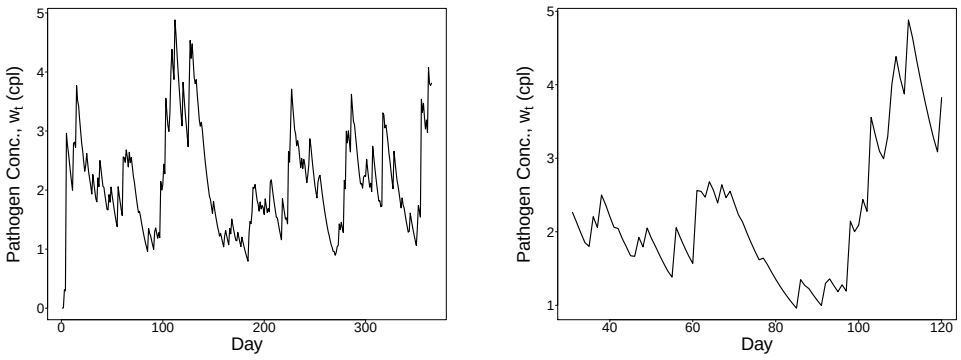
$$x_{n+1} = \kappa y_n - \beta x_n\,,$$

where δ_{kn} is a Kronecker delta,

$$\delta_{kn} = \begin{cases} 0 & \text{if } k \neq n \\ 1 & \text{if } k = n\,. \end{cases}$$

6.7.1.1 *Pathogen Concentration Dynamics*

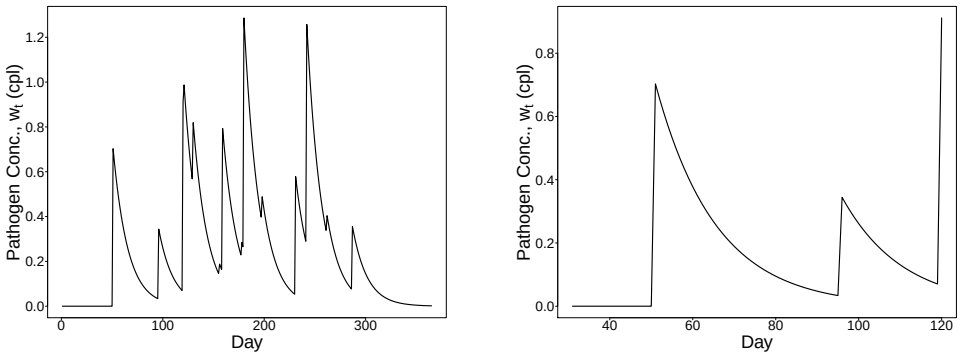
Figure 6.9 shows the pathogen concentrations within the environmental waters denoted by our model as w_t. Figure 6.9a shows the pathogen levels (in copies per litre) across days 0 – 365 simulated in our model, while 6.9b provides a closer look at the dynamics across 31 – 120 simulated days.



(a) Pathogen Conc. Days 1–365 (b) Pathogen Conc. Days 31–120

Figure 6.9: Plot of pathogen concentrations within estuary/shellfish farm water, w_t.
STW capacity = 0 dm of rainfall

Both plots exhibit multiple spikes, discrete increases in the pathogen concentrations which are proportional to any rainfall events, $(\tau_n, \zeta_n - C)$, which have rainfall levels greater than the STW capacity to process. These spikes are instantaneous, immediately followed by a decay of the pathogen level of rate λ.



(a) Pathogen Conc. Days 1–365 (b) Pathogen Conc. Days 31–120

Figure 6.10: Plot of pathogen concentrations, w_t, within estuary/shellfish farm water, when STW capacity = 0.1 dm of rainfall

Figure 6.9 shows pathogen levels when the local STW has capacity $C = 0dm$ of rainfall. Figure 6.10 shows pathogen level patterns when the STW has capacity $C = 0.1dm = 10mm$ of rainfall. It is apparent that both the number and magnitude of the spikes in concentration levels due to $(\tau_n, \zeta_n - C)$ events has decreased due to the increased treatment capacity. Note that the treatment capacity of $0.1dm$ is more than three times that of the average rainfall described by the parameters of the gamma distribution used to inform the rainfall levels ζ_n. Equation 2.20 in Chapter 2 defines the arithmetic mean of a gamma distribution as $\bar{x} = k\theta$. We have used values of $k = 0.49116$ for the shape parameter, and the scale parameter is $\theta = 6.3675$, resulting in $\bar{x} = 3.127mm = 0.0313dm$.

The spike-decay shape associated with each pairwise $(\tau_n, \zeta_n - C)$ event is similar to that reported in literature with regards to pathogen time series data.

Research carried out in 2008 by DEFRA measuring the levels of faecal indicator organisms at sites in England exhibited very similar spike-decay behaviour to that shown in Figures 6.9 – 6.10 [77].

6.7.1.2 *Pathogen Concentrations In Water*

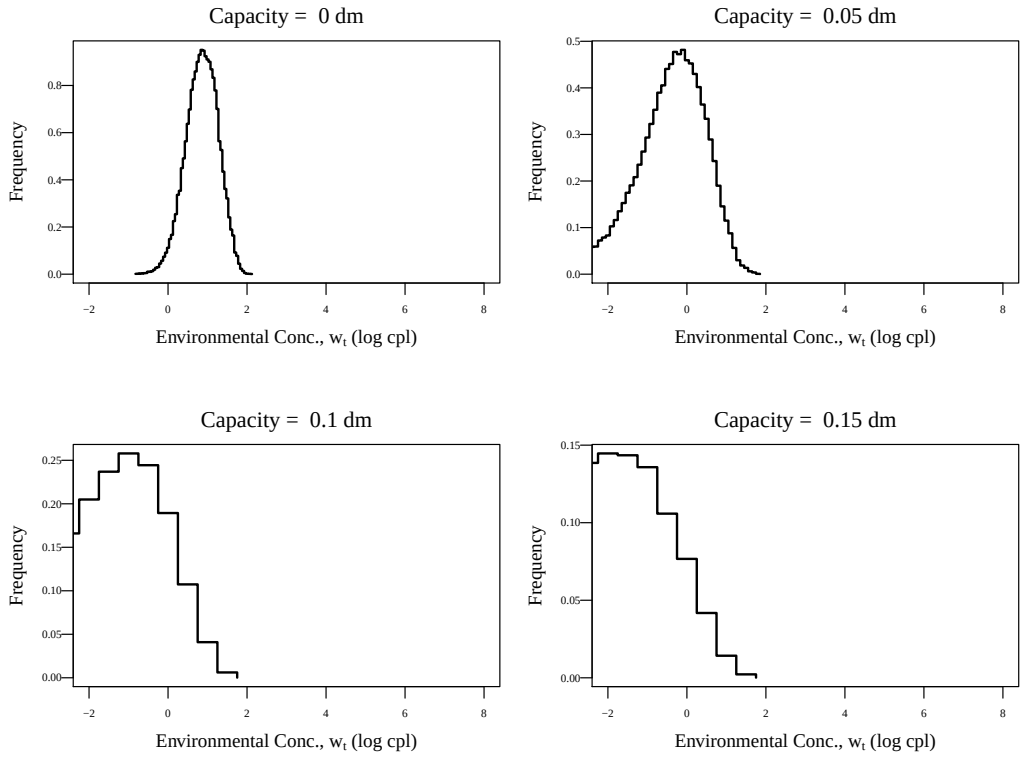


Figure 6.11: Histograms of pathogen concentrations within the environmental water w_t. The plots show $\log_e$-transformed values of w_t

Figure 6.11 shows the distribution shapes of pathogen levels, w_t, the environmental water component of our model. The histograms are plotted using log-transformed pathogen levels (log copies per litre). Each plot is a histogram of the pathogen levels each day, across days 100 – 1460. Previously Figure 6.10 showed that no pathogen level is present until the first $(\tau_n, \zeta_n - C)$ event occurs.

Each plot in Figure 6.11 shows that, as the capacity of the STW increases, the pathogen histogram experiences a shift towards lower concentration levels, as well as a decrease in event frequency. The shape also undergoes a change in symmetry, with the histogram for $C = 0dm$ appearing to be symmetric whereas, when the capacity increases, the shape becomes more negatively skewed.

6.7.1.3 *Pathogen Levels in Shellfish*

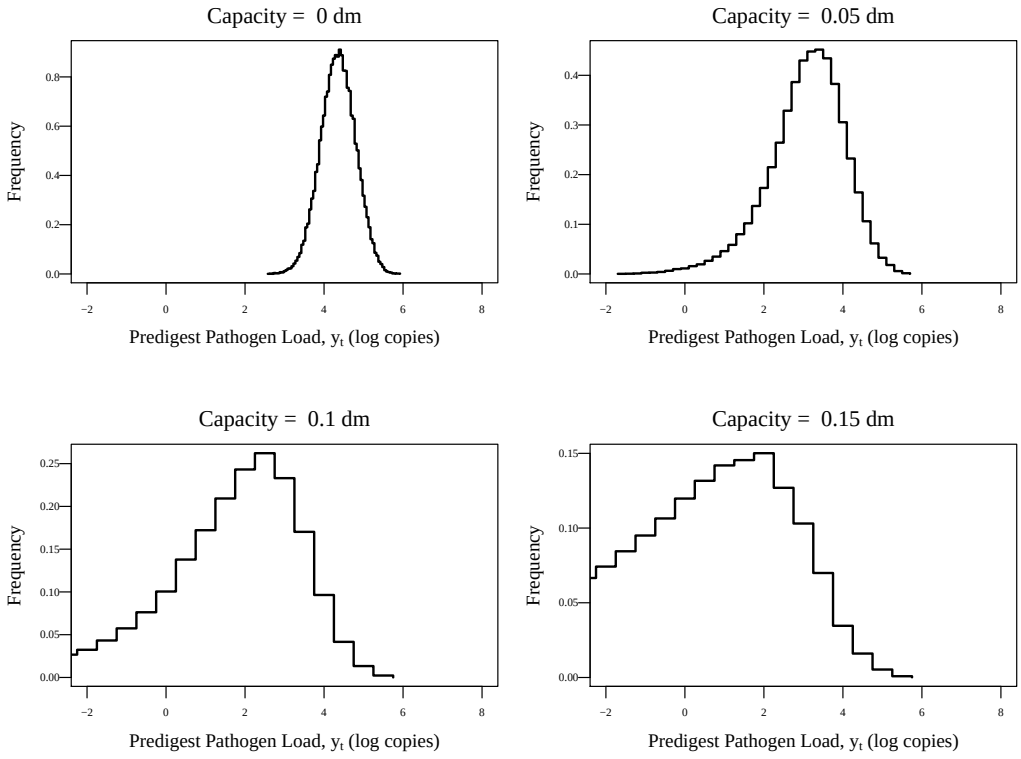


Figure 6.12: Histograms of pathogen concentrations within the unobservable compartment y_t. The plots show $\log_e$-transformed values of y_t

Figure 6.12 show the log-transformed histograms of simulated levels of pathogens in the unobservable compartments of shellfish y_t. Similar to the shape of the data in Figure 6.11, the histogram showing levels when the STW has capacity $C = 0dm$ is symmetric for log values. This suggests that y_t

could be well described by a lognormal distribution. Again, as the capacity is increased the symmetry of the data shifts to an increasingly negative skewness for the log transformed data.

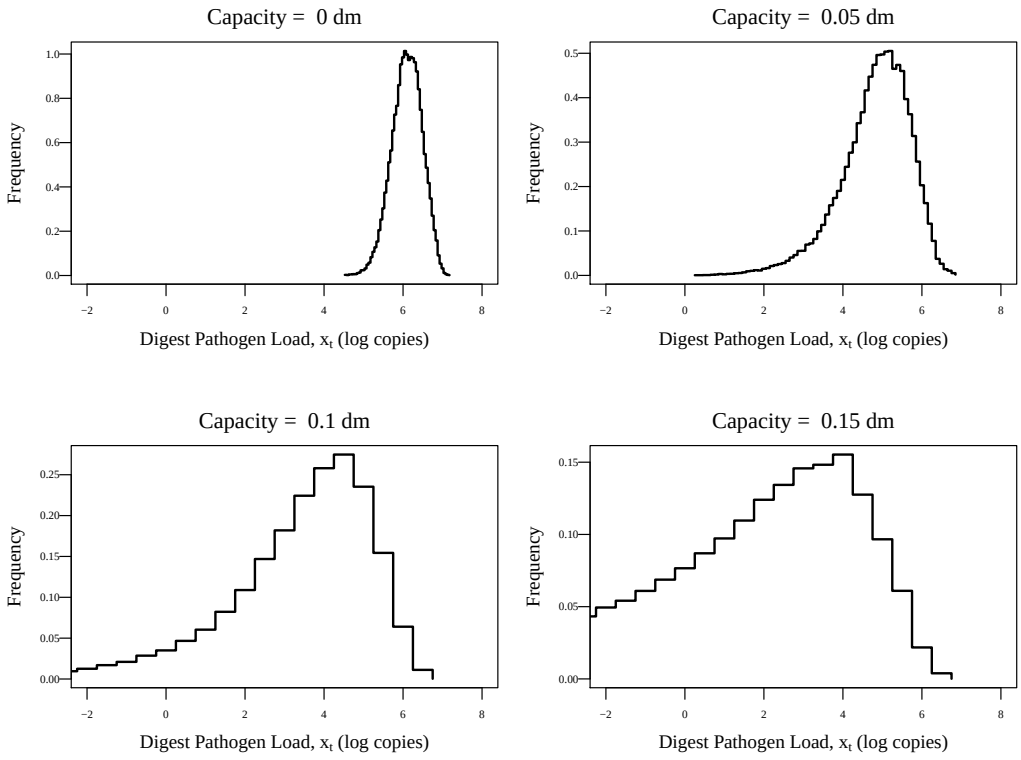


Figure 6.13: Histograms of pathogen concentrations within the observable compartment x_t. The plots show $\log_e$-transformed values of x_t

Figure 6.13 shows the observable compartment x_t under the same criteria as Figures 6.11 and 6.12. We again observe a symmetric shape when capacity $C = 0dm$, and similar changes to the shape of the data as the capacity increases towards $0.15dm$.

Note that mode of each of the peaks for the x_t data is located at higher pathogen levels when compared to the unobservable compartment y_t. This is a consequence of the relationship between the values of κ and β as used in this model. The rate κ describes the internal transfer of pathogens between the

unobservable (y_t) and observable (x_t) compartments of a shellfish, whereas β defines the pathogen excretion rate by the animal.

It is important to note that the shape of Figures 6.11(a), 6.12(a) and 6.13(a) all are unimodal, symmetric with no apparent outliers. As these histograms are plotted on a log horizontal axis, then they can be reasonably described as lognormal in shape. This corresponds with our use of lognormal distributions to describe the distribution of water-borne pathogens across a population of shellfish in Chapters 3 and 5.

A stochastic model of pathogen transmission from public waste water systems and into estuary/shellfish farm waters was constructed in Chapter 6. That model incorporated the system of differential equations that describes the dynamics of pathogen levels within individual shellfish (cf. Equations 5.2 – 5.3 in Chapter 5):

$$\frac{dy_t}{dt} = -ky_t \tag{6.18}$$

$$\frac{dx_t}{dt} = ky_t - bx_t\,, \tag{6.19}$$

with an adaptation to the equations to take into account the accretion of pathogens whilst filter feeding in contaminated waters (see Equations 6.14 – 6.17).

The results obtained from the model constructed in Chapter 6 can be used to inform the pathogen levels for a population of shellfish at the pre-depuration stage. Equations 6.18 – 6.19 shown above form the compartmental model of depuration from Chapter 5. Thus the models from both Chapters 5 and 6 provide the means to simulate a harvest and subsequent depuration of a batch of shellfish.

6.8.1 *Simulating Harvest of Shellfish*

There are two ways in which a simulated sample (or harvest) of oysters can be obtained from the environmental—compartmental model in Section 6.5. The first would be to generate multiple pairwise (y_t, x_t) values (say m pairwise values) using the model described by Equations 6.11—6.13, and select some time $t \in T$ in the model as the simulated harvest time, although t should be

sufficiently greater than zero to have allowed the burn-in of pathogen levels to have taken place. At the selected time point, m pairwise values of both y_t and x_t can then be randomly selected from the results of each simulation.

This sampling simulates the act of harvesting, and so can be used to enumerate the pre-depuration compartmentalised pathogen levels for a batch of m shellfish. These values can then used as the pre-depuration values ($t = 0$) in the depuration model detailed in Chapter 5, and so depuration of the m pairwise values (y_t, x_t) can be simulated according to Equations 6.12 and 6.13.

The second method would be to generate one pairwise set of y_t and x_t, across the period of n days, then select m pairwise (y_t, x_t) values randomly from the values as at each time point. This is possible due to the ergodic nature of the discrete time process used to create our values of y_t and x_t, which states that the statistical properties of an entire process can be obtained from a random sample of the process [89]. The ergodicity of the process is borne out by the calculation of the value of A described in Section 6.8.3 further on in this Chapter, which shows that this sampling method returned a value of A which conforms to Equation 5.22 derived in the previous Chapter.

Note that in the compartmental model in Chapter 5, we had to provide the parameter A, which split the unobservable and observable compartments at the pre-depuration stage by proportions $(1 - A)$ and A respectively. As we have been able to simulate pairwise sampling of both these compartments, we do not require this parameter in this connected model.

6.8.2 *Depuration of Simulated Harvest*

The parameters used to simulate depuration are shown in Table 6.3, and the 'R' code which implements the depuration process by the Euler method is shown in Appendix A. We have used the same values of k (the internal transfer rate) and b (the depuration rate) as applied in Chapter 5. Equivalent rates per day were also applied in the model describing the transit of pathogens from accretion to excretion in Chapter 6. Unlike the models described in Chapters 3 and 5, the NoV assurance level is not used to set the variability of the distribution of pathogen loads. The variability of the pathogen distribution is already inherent in the simulated samples randomly selected from the model in Chapter 6 at time t.In this instance, the values of ϕ are as described: they define the proportion of a shellfish batch which must be below some pre-defined pathogen load value. We have previously defined Ψ as this load limit value, and set $\Psi = 200$ copies as done previously.

Parameter label	Description	Parameter value used	Units
k	Pathogen depuration transfer rate	0.07453	/hour
b	Pathogen depuration rate	0.01339	/hour
ϕ	NoV assurance level	$\phi \in \{0.90,\ 0.95,\ 0.99\}$	$\phi \in \mathbb{R}^+$
Ψ	NoV threshold limit	200	copies

Table 6.3: Parameters used to obtain depuration results based values from stochastic environmental model

6.8.3 *Pre-Depuration Proportions*

Chapter 5 presented a model of unobservable and observable loads sequestered internally within each oyster. In that model, the most important parameter was

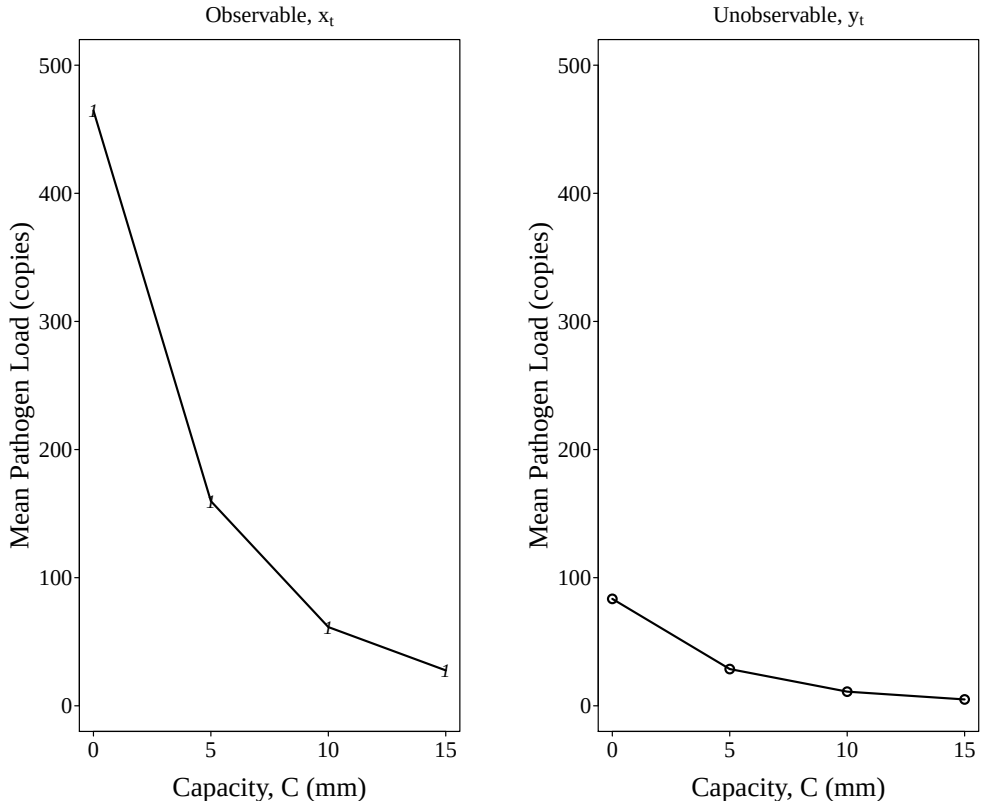


Figure 6.14: Plot of mean pre-depuration values of observable ($\bar{x}_t$) and unobservable ($\bar{y}_t$) loads for varying STW capacities

A, which determined the fixed level of observable NoV across the population of shellfish begin depurated.

We now analyse the mean values of the observable (x_t) and unobservable (y_t) values obtained using the stochastic model in Chapter 6, comparing the proportion of their values which comprise the total pathogen load (z_t). Figure 6.14 shows the mean pathogen loads of these compartments, and indicates that their values both follow exponential decay behaviour as the STW capacity is increased.

The proportion of pathogens obtained from calculating $\frac{x_t}{z_t}$ defined the value of parameter A. The values shown in Figure 6.14 are stated in Table 6.4, along

with the resulting value of A for each capacity C.

	Capacity, C			
	0 mm	5 mm	10 mm	15 mm
$\bar{y}_t$	83	29	11	5
$\bar{x}_t$	465	160	61	28
$\bar{z}_t = \bar{y}_t + \bar{x}_t$	548	189	72	33
A	0.849	0.847	0.847	0.848

Table 6.4: Parameters used to obtain depuration results based values from stochastic environmental model

The values of A are almost constant with $A \approx 0.848$ for the different values of capacity C. This aligns with the derivation of the value of A from the equilibrium of the system in Section 5.4.4, where the value of A was shown to be

$$A = \frac{k}{k+b}$$

in terms of the internal transfer and excretion/depuration rates during depuration, and equivalently as

$$A = \frac{\kappa}{\kappa+\beta} \tag{6.20}$$

in the environmental—compartmental model in Section 6.5. There we have used $\kappa = 1.78872$ and $\beta = 0.32136$ as the internal transfer and excretion/depuration rates, and so Equation 6.20 provides a value for $A = 0.8477$.

6.8.4 *Capacity v Pre-Depuration NoV Levels*

The processes to simulate shellfish harvest described in Section 6.8 were carried out, and results are now presented. Figure 6.15 provides values of the pre-

depuration observable load x_t at which the three NoV assurance levels have already been achieved. It can be seen that as the STW capacity C is increased, the value of x_t at which 90, 95, 99% of the simulated harvest sample is below decreases.

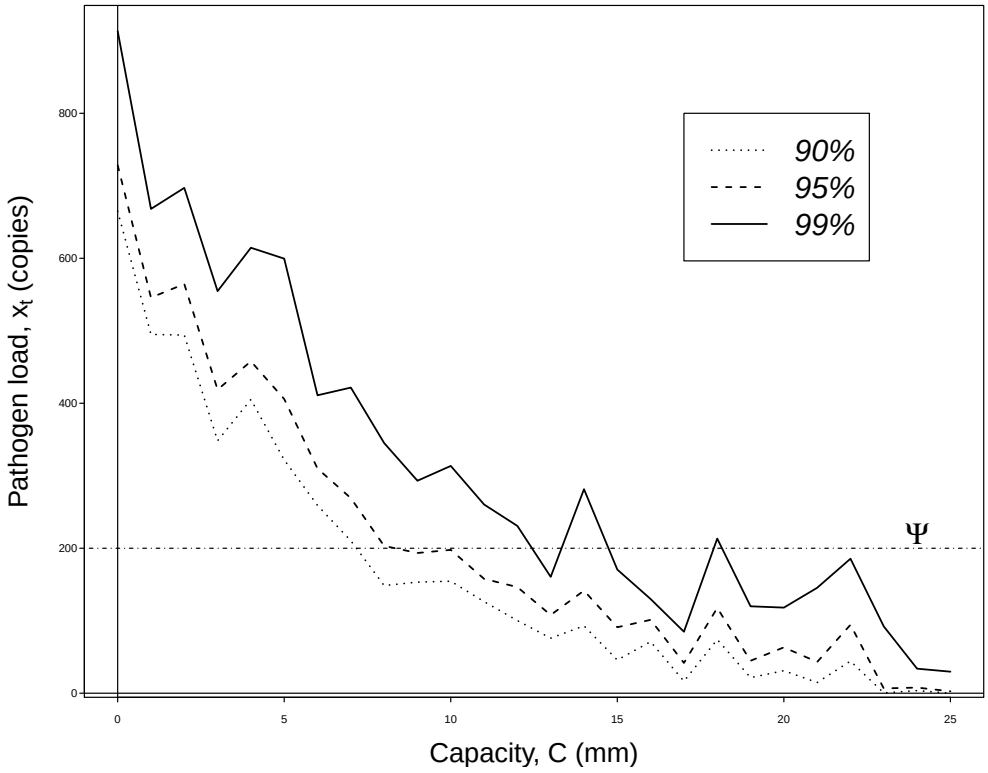


Figure 6.15: Plot of STW capacity versus pre-depuration observable NoV load x_t. Plot shows pathogen values of x_t at which the NoV assurance levels of $\phi = 0.90,\ 0.95,\ 0.99$ are already fulfilled

Results shown in Table 6.5 and Figure 6.15 are based upon simulated samples sizes of $n = 1000$. The code executed in 'R' to obtain these results is shown in Section A.1.7 of Appendix A.

	Capacity, C					
Assurance level	0 mm	5 mm	10 mm	15 mm	20 mm	25 mm
$\phi = 90\%$	665	322	155	46	31	0
$\phi = 95\%$	728	406	198	91	63	3
$\phi = 99\%$	913	599	314	171	118	30

Table 6.5: Observable pathogen load values at pre-depuration stage for increasing STW capacity C. Values shown are where ϕ % of the x_t samples have pathogen loads less than that value for each capacity C

6.8.5 *Capacity v Minimum Depuration Time*

The results shown in 6.5 and Figure 6.15 show that only after the STW capacity C has been increased to deal with rainfall levels of $\approx 10mm$ and above do we observe that the assurance levels of $\phi = 90\%, 95\%$ are already fulfilled for load limit value $\Psi = 200$ copies. Therefore no depuration would be required to meet these assurance levels. However some depuration would need to be undertaken to meet the higher assurance of $\phi = 99\%$.

Figure 6.16 shows the minimum depuration times (MDTs) required to achieve the assurance levels of 90%, 95% and 99%. In general, MDTs decrease with increasing capacity C. This shows that it is possible to reduce the MDTs by increasing C; this lowers the environmental concentrations of NoV and thus allows reduced (or no) requirement for depuration. As previously stated, when the STW capacity is at $C > 10mm$, only when we are required to achieve 99% of the sample with $x_t < \Psi = 200$ copies do we require some depuration to be instigated.

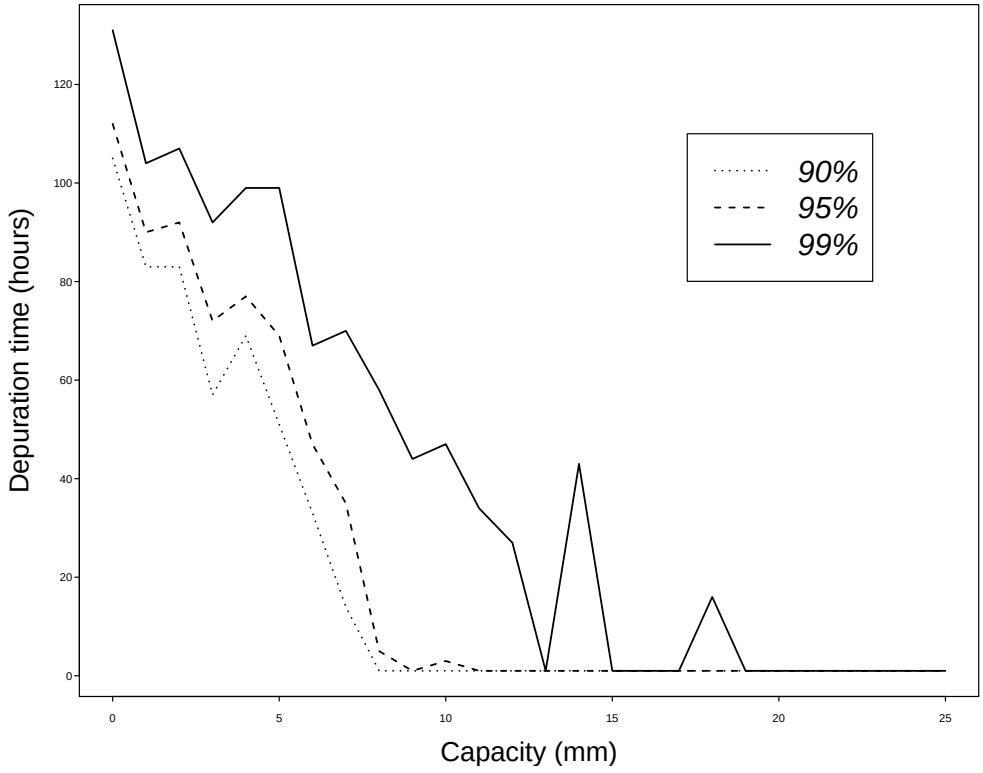


Figure 6.16: Plot of STW capacity versus MDT's. Plot reports MDT values for NoV assurance levels set at $\phi = 0.90,\ 0.95,\ 0.99$ to achieve load limit value $\Psi = 200$ copies

6.9 SUMMARY

The aim of this chapter was to construct a mathematical model that could provide some interconnectivity between the treatment capacity of a STW and the levels of pathogens within the waters of shellfish farms in close proximity. We achieved this by first analysing data on rainfall intensity and frequency, and constructing mathematical mechanisms that could simulate rainfall within our model. We considered how any significant rainfall events can overwhelm the waste water treatment capacity of a STW, and so incorporated the treatment capacity into our model. We then adapted the compartmental model of depuration constructed in Chapter 5 into a compartmental model describing the pathogens levels within the compartments of shellfish while they are filter-feeding in a pathogen rich environment.

The constructed model allowed us to consider how an increased capacity can have a positive impact on the levels of pathogens transmitted into waters and shellfish farms. The model was also used to produce results describing how an increased STW treatment capacity induces a reduction the pathogenic loads which become sequestered in shellfish due to their ability to bio-concentrate water-borne pathogens via their filter-feeding process. The model also showed that the shape of the distribution of pathogens that become sequestered in shellfish could be approximated as lognormal, and so could be a possible mechanism (or at least part form of a larger one) that drives the variability of water-borne pathogen levels across shellfish populations.

We have stated that the concentration of pathogens in the estuary is described by Equation 6.9, which states that the concentration passed into the estuary is

$$\frac{\rho L}{V}(\zeta_n - C)\,. \tag{6.21}$$

We should consider the pathogen concentration transmitted into the estuary, based on the parameters in Table 6.2, and no rainfall capacity in place at the STW. Values of $\rho = 2500$ copies per litre, $L = 400\,000\,000\,dm^2$, $V = 100\,000\,000$ litres, and an average rainfall level $\bar{\zeta}_n = 0.0313dm$ and assuming that $C = 0dm$ results in an estuary concentration of 313 copies per litre = 5.7446 $\log_e$ copies per litre.

This value is well above the range of the distribution shown in Figure 6.11. However this value is based on an average rainfall of $\bar{\zeta}_n = 0.0313dm$ occurring each day, and so does not take into account events when $\zeta_n < \bar{\zeta}$. More importantly the pathogen reduction rate λ will be the most significant factor in reducing the pathogen concentrations levels in w_t, above and beyond any increase in the value of the STW capacity C.

One aspect of this model of note has been the lognormal shape of pathogens in our results. Figures 6.11(a), 6.12(a) and 6.13(a) all closely approximate lognormal distributions and describe the distribution of pathogens within the environmental water (w_t),and unobservable (y_t) and observable (x_t) compartments of oysters. This provides some confidence in our application of lognormal distributions earlier in this thesis to describe pathogens within and across a population of shellfish at the harvest/pre-depuration stage.

7

DISCUSSION

Norovirus (NoV) is a significant cause of gastroenteritis globally [19, 20], and the consumption of oysters is frequently linked to outbreaks [14, 27]. Depuration is the principle means employed to reduce levels of potentially harmful agents or toxins in shellfish [90]. The aim of this thesis was to construct mathematical models which can describe the depuration dynamics of water-borne pathogens and specifically examines the dynamics of NoV during depuration. This was considered, not only for an individual shellfish, but across a population of shellfish. Legislation is currently under consideration within the EU by the Directorate-General for Health and Consumers (DG SANCO) to limit the maximum level of NoV that consumers are exposed to via this route. Therefore any models constructed should incorporate control measures which could be used to implement maximum NoV levels.

In addition to modelling the impact on pathogens during the depuration of shellfish, we wished to gain some insight into how the variability, and not just the mean levels, of water-borne pathogens can be as important with respect to the length of depuration required to minimise any food safety risks to the consumer. This proved difficult in the absence of any data sets that can be used to calculate variability measures, as little data is currently available to inform these values for NoV. However, our modelling techniques can provide an upper limit on the variability of water-borne pathogens that are well described by lognormal distributions.

The model described in Chapter 3 assumes an exponential decay of pathogen loads in a single mollusc due to depuration, and incorporates this decay into a lognormal distribution. This model is based on the assumptions that NoV is log-normally distributed throughout the oyster population [44, 53, 54, 55, 56] and that decay during depuration is exponential [62, 63, 69], both of which are consistent with available data and published records.

The model requires the input of four parameters: i) the initial average NoV load, ii) the depuration efficiency, iii) the desired assurance level and iv) the required NoV threshold limit. Based on these inputs the model provides an estimate of the minimum depuration time required to reduce NoV levels below a threshold which, in conjunction with the other parameters, can also be used to determine the probability of a batch of oysters testing below the detection threshold after depuration. A protocol for determining minimum depuration times using the model is as follows:

1. Measure $\bar{x}_0$ of oyster batch's harvest site;
2. Determine characteristic efficiency of overall depuration process, b^{-1};
3. Fix value of NoV load threshold, Ψ;
4. Select NoV assurance level, ϕ;
5. Apply these parameter values to the model to obtain recommended depuration period, T_{WCV}.

Steps 1–3 are anticipated to be carried out or fixed by regulators. The NoV assurance level parameter ϕ may not be fixed by legislation, however increasing values of this parameter in the model would provide increased confidence to

both depurators and consumers. This would ensure that oysters passing into the supply chain would have a diminished probability of containing significant NoV levels. The initial NoV load would be determined using the international standard for NoV detection [15] prior to depuration. This test provides an arithmetic mean of a NoV load across a population of ten oysters.

This in itself provides no information on variability within the population which is required in the calculation of depuration times. However, a worst case level of variability can be determined in the absence of this data. This worst case variability increases with the assurance level applied, as does the depuration time required. In the data sample analysed in Section 3.3.1.2 there was relatively good correlation between the worst case variability and the variability observed in the single oyster data when assurance levels of 90% and 95% were assumed. When real data becomes available for variability in NoV between oyster populations, this can replace our worst case variability approach, which should result in a reduction in the predicted depuration times.

The assurance level ϕ determines the desired (by the depurator or regulator) proportion of oysters in the population with NoV levels below the set threshold after depuration. This is important as, in addition to providing a confidence level associated with the safety of a batch depurated oysters, it is also directly linked to the probability that a sample of ten oysters will return a value below the threshold after depuration. Though increasing the assurance level also increases the required depuration time, it will also reduce the probability that a batch of oysters will fail testing, thus allowing stakeholders to evaluate the trade-offs between depuration times and an acceptable failure rate for themselves.

The results shown in Table 3.3 show that simulated tests of 10 homogenates of shellfish would provide differing pass rates compared to the value of the assurance level ϕ. With $\phi = 90\%$, a simulated test pass rate of 96% was observed from 10 000 test runs; when $\phi = 95\%$, a pass rate of 98% was observed; while $\phi = 99\%$ gave a pass rate of 99%. With this in mind, it would be prudent to use an assurance level of 95% or greater to ensure that post-depuration testing would reflect a pass rate close to the assurance level used to determine the MDT.

The final parameter is the NoV threshold value, which is likely to be set in future by an appropriate regulating body. The bigger the difference between the initial average NoV level and the threshold value, the longer the required period of depuration will be. Though this model uses a value of $\Psi = 200$ NoV cpg, this is purely for illustrative purposes and not a suggestion for such a limit, which would require a detailed understanding of the health risk posed by different levels of NoV.

Both of these values, Ψ and ϕ, may be controlled in the future by public health authorities [40], although changes to the status of the UK within the EU may require future UK legislators to adopt pathogen controls within shellfish beyond the current 42 hours required for Class B harvests.

7.1.1 *Model Limitations*

Little data is currently available regarding the depuration efficiency of different systems for the removal of NoV. For illustrative purposes this model used estimates obtained from the literature [7, 62, 63], but clearly there needs to be substantial research to determine accurate depuration efficiencies for NoV

before sensible predictions regarding depuration times can be made. However, regardless of the value used, it is possible to calculate the relationship between improvements in depuration efficiency and depuration time, which approximately halves as efficiency doubles.

The model has assumed an exponential decay of pathogen levels due to depuration within an individual shellfish. We have used a fixed value for the rate of decay (b) across the whole population during depuration, however this would not be realistic as pathogen levels would decay at different rates within the population. This could be due to differing metabolic rates, internal biological differences and position within the depuration tank are reasons that could vary the rate of b across the population.

This model does not incorporate any cross contamination between shellfish during depuration. Following on from the lognormal distribution of pathogen loads within shellfish, we can hypothesise that a small number of shellfish will carry a high pathogen load into the depuration tank. Once the shellfish have recommenced filter feeding, they will excrete pathogens into the depuration tank water, and some of this may be accreted by other shellfish in close proximity.

7.1.2 *Further Work on Exponential Decay Model*

Some refinements of the model from Chapter 3 could be undertaken in future. The inclusion of a variable decay rate instead of a fixed value of b would provide a more realistic depiction of depuration, as per the reasons stipulated previously. This could be implemented by selecting a value of b for each shellfish in the population from a probability distribution, e.g. each depuration rate

could be chosen from $\mathcal{N}(\mu = b, \sigma^2)$. Incorporating this selection method for b into the model would provide a more realistic depiction of the depuration process.

Where cross contamination during depuration is concerned, an experiment could be conducted to test for the significance of any transmission of pathogens between shellfish. Similar to the experiment carried out by Neish in 2013 [39], a batch of oysters could be exposed to high levels of NoV in artificially seeded tank waters. The shellfish would then ingest high levels of NoV as a consequence of their feeding activity.

Once this bioaccumulation has taken place, allowing the shellfish to bioconcentrate high pathogen loads, the animals could then be removed to a depuration tank. Amongst the bioaccumulated animals, shellfish which are known to have been grown and harvested in farms with track records of very low pathogen levels could be marked and placed within the depuration tanks along with the bioaccumulated specimens. Once depuration has been completed, the marked shellfish could then be tested for pathogen loads, and if they return higher than expected values, then cross-contamination within depuration would need to be taken into account within our modelling framework.

Chapter 5 built upon the depuration model, inheriting that models' features to include pathogen level controls as well as yield a worst case variability measure. The main difference between them was that the Chapter 5 model assumed a sequestration of NoV levels between two interconnected compartments within an oyster. This compartmentalised approach was based on evidence within the literature which states the pathogen loads are found not only in the digestive glands of a shellfish, but also in most other parts of the animal's biology [29, 30]. This is of special concern with regards to NoV in oyster species for two reasons: the standard assay for NoV is only viable for the digestive glands of oysters; oysters are traditionally consumed raw, unlike most other molluscs. The cooking process will render most or all viruses present in shellfish inviable, and so raw consumption can pose an increased food safety hazard to the public.

The inclusion of a fixed proportion (A) across a shellfish population of unobservable NoV load in our model resulted in extended depuration times being calculate. More importantly, the model demonstrated that mean observable NoV loads will experience an increase during the first hours of depuration, dependent upon the value of A. This initial increase in observed mean values was observed in data recorded by Neish in 2013 [39], and is also observed in results provided by Polo *et al* [63].

7.2.1 *Limitations of Compartmental Model*

Similar to the limitation of the depuration decay rate b, the proportion of observable NoV wold be variable rather than fixed across a population of shellfish. This would also be true for the internal transfer rate k. Including methods in our model to vary these parameter values would provide some greater resemblance to the reality of the depuration process.

7.2.2 *Further Work*

If our hypothesis of sequestered pathogens loads within oysters is correct, then this would have ramifications for the timing of testing practices. Due to the increase in the observable loads seen during the initial hours of depuration as predicted by our model, any food safety testing must be carried out at least after this initial increase has occurred (τ_2 in our model). Testing at the end point of depuration rather than before (or even during) the process would be prudent to minimise any potential food safety risk being passed to the consumer.

7.3 STW — SHELLFISH FARM PATHOGEN MODEL

The work carried out in Chapter 6 was not based on depuration, but rather had the principal aim of investigating the connection between overflow discharges from sewage treatment works (STW) due to high levels of rainfall, and the levels of faecal-transmitted pathogens that are found within untreated sewage waters. The model of pathogen concentrations passed into shellfish farm waters via a STW due to high rainfall levels was stochastic in nature, and was used to calculate observable and unobservable pathogen loads for a large sample of simulated shellfish. These values were then used to inform the pre-depuration pathogen loads for a large sample of simulated shellfish. These were then 'depurated' according to the equations used as the basis for the compartmental model in Chapter 5, applying Eulerian methods to do so.

7.3.1 *Limitations of Simulating Harvests*

The main limitation of this model is highlighted by the sampling of simulated shellfish. We produce sample batches of shellfish by randomly choosing time points of the simulation rather than selecting a batch at a specific time during the simulation. Therefore, our simulated harvests are temporal not spatial. Ergodic theory does allow this type of longitudinal sampling rather than from across a simulated batch of oysters; however, this does not reflect the actual harvesting of shellfish, who are removed en masse from farm waters at the same time.

Any variability in the pathogen loads they have accrued will be due to variances in their biology (which we could account for by the use of variable

rates of k, b and η), their location in the water column of the farm and distance form any source of pollution, as well as variances in the water currents and temperatures they individually experience over time.

7.3.2 *Further Work*

The inclusion of a spatial component in our modelling of a shellfish farm's waters would be an important next step in improving this model. We have used a temporal selection instead of a spatial one in our model and, as stated earlier, does not reflect the reality of shellfish harvesting. Further work on constructing a spatial model with components that describe variations in water column flows and temperatures, and an inclusion of turbulence in the pathogen levels due to tides and currents would provide the means to simulate a shellfish harvest within our model.

A further extension of this model would be to include variable values of the internal transfer (k) and depuration/excretion (b) rates per shellfish. we have previously stated that the lognormal distribution is a multiplicative type of distribution (cf. Section 2.3.2); incorporating a random selection of the values of k and b per oyster from an appropriate distribution of values would perhaps provide an additional explanation of the lognormal shape of pathogen distribution across shellfish populations. Further simulation and mathematical analysis is required to provide discover any validity to this hypothesis.

7.4 CONCLUSIONS

Depuration is one of the main tools through which shellfish industries aim to reduce NoV levels below possible future legislated thresholds. This may however be costly to the industry and therefore the consumer. This thesis arose from a desire to provide a mathematical framework to help industry and regulators understand the relationship between possible future NoV limits and required depuration times. In doing so the models described in this thesis can also provide tools with which to determine by how much depuration efficiencies may need to improve to reduce depuration times to levels deemed economically and logistically feasible by industry. It is recognised that NoV is harmful not just to shellfish consumers, but also to the shellfish industry's reputation; therefore, having the ability to determine the depuration times required to bring NoV and other water-borne pathogen loads to below threshold levels will be beneficial to all concerned.

However, recent developments with the vote on the EU Referendum in June 2016 will mean that any EU imposed regulations may not be enforced by the UK government. However UK legislators may deem it a necessity to enforce viral mitigation for shellfish before they reach the consumer, and these models can provide them and industry stakeholders with a management tool capable of providing a robust solution to the question of sufficient application of the depuration process.

BIBLIOGRAPHY

[1] Anonymous, "Map of Chichester harbour." [Online]. Available: https://www.cefas.co.uk/media/1442/chichester-o-edulis.jpg

[2] J. A. Lowther, N. E. Gustar, A. L. Powell, R. E. Hartnell, and D. N. Lees, "Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom." *Applied and environmental microbiology*, vol. 78, no. 16, pp. 5812–5817, aug 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22685151http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3406157{&}tool=pmcentrez{&}rendertype=abstract

[3] T. Ellis, R. Gardiner, M. Gubbins, A. Reese, and D. Smith, "Aquaculture statistics for the UK , with a focus on England and Wales 2012," Centre for Environment, Fisheries & Aquaculture Science, Weymouth, Tech. Rep., 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment{_}data/file/405469/Aquaculture{_}Statistics{_}UK{_}2012.pdf

[4] Anonymous, "Council of 29 April 2004 laying down specific rules for the organisation of official controls on products of animal origin intended for human consumption," *Official Journal of the European Union*, no. 854, pp. 83–127, 2004. [Online]. Available: http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Regulation+(EC)+No+854/2004+of+the+European+Parliament+and+of+the+Council+of+29+April+2004+laying+down+specific+rules+for+

the+organisation+of+official+controls+on+products+of+animal+origin+inte

[5] ——, "Commission Regulation (EU) 2015/2285 of 8 December 2015 amending Annex II to Regulation (EC) No 854/2004 of the European Parliament," 2015. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv{%}3AOJ.L{_}.2015.323.01.0002.01.ENG

[6] E. Limpert, W. a. Stahel, and M. Abbt, "Log-normal Distributions across the Sciences: Keys and Clues," *BioScience*, vol. 51, no. 5, p. 341, 2001. [Online]. Available: http://bioscience.oxfordjournals.org/cgi/doi/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

[7] B. Doré, S. Keaveney, J. Flannery, and P. Rajko-Nenow, "Management of health risks associated with oysters harvested from a norovirus contaminated area, Ireland, February-March 2010." *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, vol. 15, no. 19, p. pii=19567, may 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20483107

[8] Anonymous, *Strange Stories, Amazing Facts*. Reader's Digest Association Limited.

[9] S. R. Rippey, "Infectious diseases associated with molluscan shellfish consumption," *Clinical Microbiology Reviews*, vol. 7, no. 4, pp. 419–425, 1994.

[10] M. Fitzgerald, A; Syvret, M; Hamilton, A; Pyke, "Review and cost-benefit analysis for industry of reduced depuration times for the mussel Mytilus edulis," SARF - Scottish Aquaculture Research Forum, Tech. Rep., 2010. [Online]. Available: http://www.sarf.org.uk/cms-assets/documents/29383-342050.sarf066.pdf

[11] I. Potasman, A. Paz, and M. Odeh, "Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective." *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, vol. 35, no. 8, pp. 921–928, 2002.

[12] G. Simmons, C. Garbutt, J. Hewitt, and G. Greening, "A New Zealand outbreak of norovirus gastroenteritis linked to the consumption of imported raw Korean oysters," *New Zealand Medical Journal*, vol. 120, no. 1264, p. 8716, 2007.

[13] G. E. Greening and D. J. McCoubrey, "Enteric viruses and management of shellfish production in New Zealand," *Food and Environmental Virology*, vol. 2, no. 3, pp. 167–175, 2010.

[14] F. S. Le Guyader, F. Bon, D. DeMedici, S. Parnaudeau, A. Bertone, S. Crudeli, A. Doyle, M. Zidane, E. Suffredini, E. Kohli, F. Maddalo, M. Monini, A. Gallay, M. Pommepuy, P. Pothier, and F. M. Ruggeri, "Detection of multiple noroviruses associated with an international gastroenteritis outbreak linked to oyster consumption." *Journal of clinical microbiology*, vol. 44, no. 11, pp. 3878–82, nov 2006. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1698296{&}tool=pmcentrez{&}rendertype=abstract

[15] D. Lees, "International Standardisation of a Method for Detection of Human Pathogenic Viruses in Molluscan Shellfish," *Food and Environmental Virology*, vol. 2, no. 3, pp. 146–155, jun 2010. [Online]. Available: http://link.springer.com/10.1007/s12560-010-9042-5

[16] I. Muniain-Mujika, R. Girones, G. Tofiño-Quesada, M. Calvo, and F. Lucena, "Depuration dynamics of viruses in shellfish." *International journal of food microbiology*, vol. 77, no. 1-2, pp. 125–133, jul 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/12076030

[17] US FDA, "Guide for the Control of Molluscan Shellfish - Revised 2013," Interstate Shellfish Sanitation Conference; U.S. Food and Drug Administration, Tech. Rep., 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1755436514000590

[18] A. A. Butt, K. E. Aldridge, and C. V. Sanders, "Infections related to the ingestion of seafood Part I: Viral and bacterial infections." *The Lancet. Infectious Diseases*, vol. 4, no. 4, pp. 201–12, apr 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15050937

[19] M. M. Patel, M. A. Widdowson, R. I. Glass, K. Akazawa, J. Vinjé, and U. D. Parashar, "Systematic literature review of role of noroviruses in sporadic gastroenteritis," *Emerging Infectious Diseases*, vol. 14, no. 8, pp. 1224–1231, 2008.

[20] E. Scallan, R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M. A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin, "Foodborne illness acquired in the United States-Major pathogens," *Emerging Infectious Diseases*, vol. 17, no. 1, pp. 7–15, 2011.

[21] J. van Beek, K. Ambert-Balay, N. Botteldoorn, J. S. Eden, J. Fonager, J. Hewitt, N. Iritani, a. Kroneman, H. Vennema, J. Vinjé, P. a. White, M. Koopmans, and NoroNet, "Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012." *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin*, vol. 18, no. 1, pp. 8–9, 2013.

[22] A. Z. Kapikian, "The discovery of the 27-nm Norwalk virus: an historic perspective." *The Journal of infectious diseases*, vol. 181 Suppl, pp. S295–302, 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10804141

[23] Anonymous, "Long Term Classification of Designated Shellfish Harvesting Areas in England and Wales Guidance on Local Action Groups and Local Action Plans," Food Standards Agency, Tech. Rep. March, 2006.

[24] ——, "FSA Scotland Classification Report," 2016.

[25] ——, "Designated Bivalve Mollusc Production Areas in England and Wales," Tech. Rep. September, 2015. [Online]. Available: https://www.food.gov.uk/sites/default/files/shellfish-classification-list-2015-2016.pdf

[26] ——, "Current List of Classified Harvesting Areas in Northern Ireland 2014," p. 2015, 2014. [Online]. Available: http://multimedia.food.gov.uk/multimedia/pdfs/enforcement/shellfish-monitoring/shellfish-class-2014.pdf

[27] J. Schaeffer, J. C. Le Saux, M. Lora, R. L. Atmar, and F. S. Le Guyader, "Norovirus contamination on French marketed oysters," *International Journal of Food Microbiology*, vol. 166, no. 2, pp. 244–248, 2013.

[28] D. Lees, "Viruses and bivalve shellfish." *International journal of food microbiology*, vol. 59, no. 1-2, pp. 81–116, jul 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10946842

[29] W. J. Doré and D. N. Lees, "Behavior of Escherichia coli and male-specific bacteriophage in environmentally contaminated bivalve molluscs before and after depuration." *Applied and environmental microbiology*, vol. 61, no. 8, pp. 2830–4, aug 1995. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=167559{&}tool=pmcentrez{&}rendertype=abstract

[30] D. Wang, Q. Wu, X. Kou, L. Yao, and J. Zhang, "Distribution of norovirus in oyster tissues." *Journal of applied microbiology*,

vol. 105, no. 6, pp. 1966–72, dec 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19120643

[31] S. P. Nappier, T. K. Graczyk, and K. J. Schwab, "Bioaccumulation, retention, and depuration of enteric viruses by Crassostrea virginica and Crassostrea ariakensis oysters." *Applied and environmental microbiology*, vol. 74, no. 22, pp. 6825–31, nov 2008. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2583511{&}tool=pmcentrez{&}rendertype=abstract

[32] J. Hewitt, M. Leonard, G. E. Greening, and G. D. Lewis, "Influence of wastewater treatment process and the population size on human virus profiles in wastewater." *Water research*, vol. 45, no. 18, pp. 6267–76, nov 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21962483

[33] M. Iwai, S. Hasegawa, M. Obara, K. Nakamura, E. Horimoto, T. Takizawa, T. Kurata, S.-i. Sogen, and K. Shiraki, "Continuous presence of noroviruses and sapoviruses in raw sewage reflects infections among inhabitants of Toyama, Japan (2006 to 2008)." *Applied and environmental microbiology*, vol. 75, no. 5, pp. 1264–70, mar 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2648165{&}tool=pmcentrez{&}rendertype=abstract

[34] C. J. A. Campos and D. N. Lees, "Environmental Transmission of Human Noroviruses in Shellfish Waters." *Applied and environmental microbiology*, vol. 80, no. 12, pp. 3552–3561, apr 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24705321

[35] Anonymous, "The State of World Fisheries and Aquaculture 2012," Food And Agriculture Organization Of The United Nations, Tech. Rep., 2012. [Online]. Available: http://www.fao.org/docrep/016/i2727e/i2727e.pdf

[36] T. Westrell, V. Dusch, S. Ethelberg, J. Harris, M. Hjertqvist, N. J. da Silva, a. Koller, a. Lenglet, M. Lisby, and L. Vold, "Norovirus outbreaks linked to oyster consumption in the United Kingdom, Norway, France, Sweden and Denmark, 2010," *Eurosurveillance*, vol. 15, no. 12, pp. 7–10, 2010.

[37] Anonymous, "Guidance for inspection of shellfish purification systems for Local Food Authorities," Food Standards Agency (Scotland), Tech. Rep. July, 2009. [Online]. Available: http://multimedia.food.gov.uk/multimedia/pdfs/publication/shellfishpurificationsystemsscot.pdf

[38] M. N. L. Seaman, "Survival and aspects of metabolism in oysters, Crassostrea gigas, during and after prolonged air storage," *Aquaculture*, vol. 93, no. 4, pp. 389–395, 1991.

[39] A. Neish, "Investigative trials on the purification of oysters to identify ways of reducing norovirus," Tech. Rep., 2013. [Online]. Available: http://www.cefas.defra.gov.uk/media/607899/2013cefascontractreportc5224.pdf

[40] Anonymous, "Scientific Opinion on Norovirus (NoV) in oysters : methods , limits and control options," *EFSA Journal*, vol. 10, no. 1, pp. 1–39, 2012.

[41] F. S. Le Guyader, S. Parnaudeau, J. Schaeffer, A. Bosch, F. Loisy, M. Pommepuy, and R. L. Atmar, "Detection and quantification of noroviruses in shellfish." *Applied and environmental microbiology*, vol. 75, no. 3, pp. 618–24, feb 2009. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2632116{&}tool=pmcentrez{&}rendertype=abstract

[42] F. S. Le Guyader, F. Loisy, R. L. Atmar, A. M. Hutson, M. K. Estes, N. Ruvoën-Clouet, M. Pommepuy, and J. Le Pendu, "Norwalk virus-specific binding to oyster digestive tissues," *Emerging Infectious Diseases*, vol. 12, no. 6, pp. 931–936, 2006.

[43] Anonymous, "Foodborne illness at The Fat Duck restaurant," Health Protection Agency, Tech. Rep. February, 2009.

[44] R. Lee, A. Lovatelli, and L. Ababouch, "Bivalve depuration: fundamental and practical aspects," Food And Agriculture Organization Of The United Nations, Tech. Rep., 2008. [Online]. Available: http://library.wur.nl/WebQuery/clc/1886525

[45] R. Lee, "Cefas Protocol for Inspection and Approval of Purification (Depuration) Systems âĂŞ England and Wales." Centre for Environment, Fisheries & Aquaculture Science, Weymouth, Tech. Rep., 2010. [Online]. Available: http://www.cefas.defra.gov.uk/media/540968/purificationgenericprotocol{_}v8final.pdf

[46] Anonymous, "SFPA Guidance Document for Inspecting LBM Purification Centres," Tech. Rep. May, 2012. [Online]. Available: http://www.sfpa.ie/Portals/0/FoodSafety/Shellfish/LBMPurificationGuidanceDocument.pdf

[47] A. D. A. Corrêa, C. Rigotto, V. Moresco, C. R. Kleemann, A. L. Teixeira, C. R. Poli, C. M. O. Simões, and C. R. M. Barardi, "The depuration dynamics of oysters (Crassostrea gigas) artificially contaminated with hepatitis A virus and human adenovirus," *Memorias do Instituto Oswaldo Cruz*, vol. 107, no. February, pp. 11–17, 2012.

[48] K. L. Jackson and D. Ogburn, "Review of Depuration and its Role in Shellfish Quality Assurance SHELLFISH QUALITY ASSURANCE PROGRAM," Fisheries Research & Development Corporation, Tech. Rep. 13, 1999. [Online]. Available: http://www.dpi.nsw.gov.au/{_}{_}data/assets/pdf{_}file/0005/134636/Output-13.pdf

[49] G. Hall, *The Application of Mathematical Thinking*, 1st ed. Nottingham: University of Nottginham Press, 1963.

[50] E. Kramer, *The Nature and Growth of Modern Mathematics*, 1st ed. New York: Hawthorn Books, Inc, 1970.

[51] J. Murray, *Mathematical Biology*, 1st ed. New York: Springer-Verlag, 1993.

[52] R. Bronson and G. Costa, *Schaum's Outlines of Differential Equations*. McGraw-Hill Companies, Inc, 2006.

[53] C. N. Haas, J. B. Rose, C. Gerba, and S. Regli, "Risk assessment of virus in drinking water." *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 13, no. 5, pp. 545–52, oct 1993. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8259444

[54] R. T. Noble, S. B. Weisberg, M. K. Leecaster, C. D. McGee, K. Ritter, K. O. Walker, and P. M. Vainik, "Comparison of beach bacterial water quality indicator measurement methods," *Environmental Monitoring and Assessment*, vol. 81, no. 1-3, pp. 301–312, 2003.

[55] T. J. Wade, R. L. Calderon, E. Sams, M. Beach, K. P. Brenner, A. H. Williams, and A. P. Dufour, "Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness," *Environmental Health Perspectives*, vol. 114, no. 1, pp. 24–28, 2006.

[56] T. J. Wade, R. L. Calderon, K. P. Brenner, E. Sams, M. Beach, R. Haugland, L. Wymer, and A. P. Dufour, "High Sensitivity of Children to Swimming-Associated Gastrointestinal Illness," *Epidemiology*, vol. 19, no. 3, pp. 375–383, 2008. [Online]. Available: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage{&}an=00001648-200805000-00008

[57] R. Adams, *Calculus: A Complete Course*, 5th ed. Toronto: Addison Wesley Longman, 2003.

[58] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions, Volume 1*, 1st ed. Wiley-Interscience, 1970.

[59] N. A. Weiss, *A Course in Probability*, internatio ed. Pearson Education Inc., 2006.

[60] M. Abramowitz, I. a. Stegun, and D. Miller, "Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables (National Bureau of Standards Applied Mathematics Series No. 55)," *Journal of Applied Mechanics*, vol. 32, no. 1, p. 239, 1965.

[61] N. L. Johnson and S. Kotz, *Discrete Distributions*, v. 3 ed. Boston: Wiley, 1969.

[62] D. Polo, X. Feal, M. F. Varela, A. Monteagudo, and J. L. Romalde, "Depuration kinetics of murine norovirus in shellfish," *Food Research International*, vol. 64, pp. 182–187, 2014.

[63] D. Polo, X. Feal, and J. L. Romalde, "Mathematical model for viral depuration kinetics in shellfish: An useful tool to estimate the risk for the consumers," *Food Microbiology*, vol. 49, pp. 220–225, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.fm.2015.02.015

[64] W. J. Doré, K. Henshilwood, D. N. Lees, and W. J. Dore, "Evaluation of F-Specific RNA Bacteriophage as a Candidate Human Enteric Virus Indicator for Bivalve Molluscan Shellfish," *Applied and environmental microbiology*, vol. 66, no. 4, pp. 1280–1285, 2000.

[65] I. Muniain-Mujika, R. Girones, and F. Lucena, "Viral contamination of shellfish: evaluation of methods and analysis of bacteriophages and human viruses." *Journal of virological methods*, vol. 89, no. 1-2, pp. 109–118, sep 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10996644

[66] S. Cartwright, "Animal Products (Regulated Control Scheme âĂŤ Bivalve Molluscan Shellfish) Regulations Order in Council," 2008.

[Online]. Available: http://www.legislation.govt.nz/regulation/public/2006/0038/latest/DLM369647.html.

[67] B. F. Ginos, "Parameter Estimation for the Lognormal Distribution," 2009. [Online]. Available: http://scholarsarchive.byu.edu/etd/1928/

[68] R. Deutsch, *Estimation Theory*, 3rd ed. Englewood Cliffs: Prentice-Hall, Inc., 1965.

[69] S. R. Smith, N. L. Lang, K. H. M. Cheung, and K. Spanoudaki, "Factors controlling pathogen destruction during anaerobic digestion of biowastes," *Waste Management*, vol. 25, no. 4 SPEC. ISS., pp. 417–425, 2005.

[70] G. Grimmett and D. Welsh, *Probability An Introduction*. Oxford Science Publications, 1986.

[71] C. Mcleod, "A critical review of the current evidence for the use of indicator shellfish species for the purposes of biotoxin and chemical contaminants monitoring in the Scottish shellfish production areas," pp. 1–83, 2014. [Online]. Available: https://www.food.gov.uk/sites/default/files/865-1-1607{_}FS512006{_}VMcFarlane.pdf

[72] B. Kuster, "Proposed changes to the approval of purification establishments for live bivalve molluscs," Tech. Rep., 2016. [Online]. Available: https://www.food.gov.uk/sites/default/files/purificationletterinterestedparties{_}4.pdf

[73] F. Loisy, R. L. Atmar, P. Guillon, P. Le Cann, M. Pommepuy, and F. S. Le Guyader, "Real-time RT-PCR for norovirus screening in shellfish," *Journal of Virological Methods*, vol. 123, no. 1, pp. 1–7, 2005.

[74] Y. Ueki, M. Shoji, A. Suto, T. Tanabe, Y. Okimura, Y. Kikuchi, N. Saito, D. Sano, and T. Omura, "Persistence of caliciviruses in artificially contaminated oysters during depuration." *Applied and*

environmental microbiology, vol. 73, no. 17, pp. 5698–701, sep 2007. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2042073{&}tool=pmcentrez{&}rendertype=abstract

[75] J. A. Lowther, "Investigation into the prevalence, distribution and levels of norovirus titre in oyster harvesting areas in the UK," Cefas, FSA, Weymouth, Tech. Rep. August, 2011. [Online]. Available: http://www.cefas.defra.gov.uk/media/548908/p01009norovirussurveillancefinalreport.pdf

[76] W. J. Lodder and a. M. D. R. Husman, "Presence of Noroviruses and Other Enteric Viruses in Sewage and Surface Waters in The Netherlands Presence of Noroviruses and Other Enteric Viruses in Sewage and Surface Waters in The Netherlands," *Applied and Environmental Microbiology*, vol. 71, no. 3, pp. 1453–1461, 2005.

[77] Anonymous, "Input variables, output results and storm plots for the FIO in-stream model at Lower Moor Head Farm and Dedra Banks Farm," Tech. Rep., 2008. [Online]. Available: http://sciencesearch.defra.gov.uk/Default.aspx?Menu=Menu{&}Module=More{&}Location=None{&}ProjectID=15547{&}FromSearch=Y{&}Publisher=1{&}SearchText=pedal{&}SortString=ProjectCode{&}SortOrder=Asc{&}Paging=10{#}Description

[78] H.-K. Cho, K. P. Bowman, and G. R. North, "A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission," *Journal of Applied Meteorology*, vol. 43, no. 11, pp. 1586–1597, 2004.

[79] Defra, "UK Implementation of the EC Urban Waste Water Treatment Directive," *Water Services*, p. 20, 2002. [Online]. Available: http://www.defra.gov.uk/publications/files/pb6655-uk-sewage-treatment-020424.pdf

[80] N. Ison, A. Feyerherm, and L. Bark, "Wet Period Precipitation and the Gamma Distribution," *Journal of Applied Meteorology and Climatology*, vol. 10, no. 4, pp. 658–665, 1971. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450{%}281971{%}29010{%}3C0658{%}3AWPPATG{%}3E2.0.CO{%}3B2

[81] H. Dikko, I. David, and H. Bakari, "Modeling the Distribution of Rainfall Intensity using Quarterly Data," *IOSR Journal of Mathematics (IOSR-JM)*, vol. 9, no. 1, pp. 11–16, 2013. [Online]. Available: http://www.iosrjournals.org/iosr-jm/papers/Vol9-issue1/C0911116.pdf?id=7287

[82] L. W. Swift and H. Schreuder, "Fitting Daily Precipitation Amounts Using the SB Distribution," *Monthly Weather Review*, vol. 109, no. 12, pp. 2535–2540, 1981. [Online]. Available: http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1981)109{%}3C2535{%}3AFDPAUT{%}3E2.0.CO{%}3B2

[83] T. Volkova, A. Longobardi, and N. Krasnogorskaya, "Gamma distribution function as a tool for monthly precipitation generation in the Bashkortostan Republic , Russian Federation," in *Latest Trends in Energy, Environment and Development*, Salerno, 2014, pp. 88–95.

[84] Anonymous, "E&W Daily Rainfall Data from Met Office." [Online]. Available: http://www.metoffice.gov.uk/hadobs/hadukp/data/download.html

[85] D. L. Fox, H. U. Sverdrup, and J. P. Cunningham, "The Rate of Water Propulsion by the California Mussel," *Biological Bulletin*, vol. 72, no. 3, pp. 417–438, 1937. [Online]. Available: http://www.jstor.org/stable/1537700

[86] J. Bae and K. J. Schwab, "Evaluation of murine norovirus, feline calicivirus, poliovirus, and MS2 as surrogates for human norovirus in a model of viral

persistence in surface water and groundwater," *Applied and Environmental Microbiology*, vol. 74, no. 2, pp. 477–484, 2008.

[87] H. B. Wenban-Smith, "Economies of scale, distribution costs and density effects in urban water supply," pp. 1–279, 2009. [Online]. Available: http://etheses.lse.ac.uk/285/

[88] J. Flannery, S. Keaveney, P. Rajko-Nenow, V. O'Flaherty, and W. Doré, "Concentration of norovirus during wastewater treatment and its impact on oyster contamination," *Applied and Environmental Microbiology*, vol. 78, no. 9, pp. 3400–3406, 2012.

[89] C. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. Berlin: Springer-Verlag, 2004.

[90] G. Savini, C. Casaccia, N. B. Barile, M. Paoletti, and C. Pinoni, "Norovirus in bivalve molluscs: a study of the efficacy of the depuration system." *Veterinaria italiana*, vol. 45, no. 4, pp. 535–9, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20391417

[91] C. Jackson, "R 'msm' package." [Online]. Available: http://svitsrv25.epfl.ch/R-doc/library/msm/html/tnorm.html

[92] MathWave Technologies, "Easy Fit." [Online]. Available: http://www.mathwave.com/

Part I

APPENDICES

APPENDIX A

A.1 R CODE

This appendix provides the code executed in 'R' which has been referenced throughout the main body of this thesis.

A.1.1 *Pass/Fail Test Results Calculations*

This R code calculates the arithmetic mean values of 10000 test iterations of the depuration model in accordance with the results discussed in Section 3.4.4.

```
#Set value of NoV threshold value used to calaculate min. dep. time
psi_test=200

#define NoV xbar value to be used
xbar_test=1064

#Create inverse erf function
invErf = function(x) {
  qnorm((1+x)/2)/sqrt(2)
}

#define upper limit of NoV Load from literature
```

```
maxNoVLoad=10000


#define range of x(NoV cpg) to be used
xx=seq(0.0001,maxNoVLoad,length=4000000)


#phi values to test
phi_test_array=c(0.90,0.95,0.99)


#lambda values to test
lambda_test=0.01339
lambda_test_array=c(lambda_test,lambda_test*1.1,lambda_test*1.25,lambda_
    test*1.5,lambda_test*2)


#Number of iterations to run each test
test_number=10000


#build 3d array to store test run values
test_array = array(NaN,dim=c(test_number,length(phi_test_array),length(
    lambda_test_array)))


#create array to store pass(=1)/fail(=0) results
test_array_results=array(NaN,dim=c(test_number,length(phi_test_array),
    length(lambda_test_array)))


single_test_array=array(NaN,dim=c(test_number,length(phi_test_array),
    length(lambda_test_array)))


Fixed_Time=42      #Set value for Fixed Depuration Time for Pass/Fail
    rates
```

```
MDT_array <- matrix(c
    (186,226,327,169,205,297,149,181,262,124,151,218,93,113,163), nrow =
    5, ncol = 3)


################################Test Run Code for rlnorm function
    ############################################################


#Loop 1 - run through 5 differing values of lambda
for (k in 1:length(lambda_test_array)){
  #Loop 2 - run through 3 different phi values
  for (j in 1:length(phi_test_array)){

    #calculate initial sigma value using WCV equation
    sigma_initial=sqrt(2)*invErf(2*phi_test_array[j]-1)
    #calculate initial value of mu at t=0 to calculate min. dep. time
    mu_initial=log(xbar_test)-sigma_initial^2/2

    #Loop 3 - run number of test iterations for each variant of phi and
        lambda
    for (i in 1:test_number){

      #calculate MinDep_mu
      MinDep_mu=0
      #MinDep_mu= mu_initial-lambda_test_array[k]*Fixed_Time
      MinDep_mu= mu_initial-lambda_test_array[k]*MDT_array[k,j]

      #Store 10 random NoV values from minimum depuration distribution
          using inbuilt R function rlnorm
      ten_oyster_array=rep(NaN,10)
      ten_oyster_array=rlnorm(10,MinDep_mu,sigma_initial)
```

```
      #run test for single oyster samples
      single_oyster=rlnorm(1,MinDep_mu,sigma_initial)

      #Finally, store the average value of these 10 'random oysters' in
          the big array in the correct location
      test_array[i,j,k]=mean(ten_oyster_array)

      if (test_array[i,j,k]>psi_test){

        test_array_results[i,j,k]=0
      }
      else {
        test_array_results[i,j,k]=1
      }

      #same process for single oyster test
      if (single_oyster>psi_test){

        single_test_array[i,j,k]=0
      }
      else {
        single_test_array[i,j,k]=1
      }

    }

  }

}
```

```
mean(test_array_results[,1,1])     #to access results, use this code in
    console window and cycle up through k,j options
```

A.1.2 *Truncated Normal Distribution Generic Plots*

The code executed to produce the plots shown in Figure 4.1 is shown below. The 'msm' package was imported to provide the appropriate functions to produce the plots. [91].

```
# Generic Plots for normal and 3 truncated normal
    -------------------------

pdf("generic_normal_std.pdf", 9,7)

xx <- seq(-5,10, length=1000)

std_norm <- dnorm(xx, mean =3, sd =2)

plot(xx,std_norm, type="l", lwd=2.5, xaxt="n", ylab = "", xlab = "", ylim
     = c(0,0.30), cex.axis=1.6)

axis(1, at=c(0,3),labels=c(0, expression(mu == bar(bold(x)))),cex.axis
    =2.0,las=1, tck=-0.005)

mtext("Probability density",2,line=2.5,cex=2.2,las=0)

abline(v=0, lwd=1.8)
```

```
abline(v=3, lty=2, lwd =1.5)
abline(h=0)

dev.off()

pdf("generic_normal_std_lower.pdf", 9,7)

xx <- seq(-5,10, length=1000)

std_norm_lower <-  dtnorm(xx, mean =3, sd =2, lower =0, upper = Inf)

plot(xx,std_norm_lower, type="l", col = "red", lwd=2.5, xaxt="n", ylab =
    "", xlab = "", ylim = c(0,0.30), cex.axis=1.6)

axis(1, at=c(0,3),labels=c(0, expression(mu != bar(bold(x)))),cex.axis
    =2.0,las=1, tck=-0.005)

mtext("Probability density",2,line=2.5,cex=2.2,las=0)

abline(v=0, lwd=1.8)
abline(v=3, lty=2)
abline(h=0)

dev.off()

pdf("generic_normal_std_upper.pdf", 9,7)

xx <- seq(-5,10, length=1000)

std_norm_upper <-  dtnorm(xx, mean =3, sd =2, lower =-Inf, upper = 4)
```

```
plot(xx,std_norm_upper, type="l", col="blue", lwd=2.5, xaxt="n", ylab = "
    ", xlab = "", ylim = c(0,0.30), cex.axis=1.6)
#lines(xx,std_norm)

axis(1, at=c(0,3),labels=c(0, expression(mu != bar(bold(x)))),cex.axis
    =2.0,las=1, tck=-0.005)

mtext("Probability density",2,line=2.5,cex=2.2,las=0)

abline(v=0, lwd=1.8)
abline(v=3, lty=2)
abline(h=0)

dev.off()

pdf("generic_normal_std_combined.pdf", 9,7)

xx <- seq(-5,10, length=1000)

std_norm_upper <-  dtnorm(xx, mean =3, sd =2, lower =-Inf, upper = 4)

plot(xx,std_norm_upper, type="l", col="blue", lwd=2.5, xaxt="n", ylab = "
    ", xlab = "", ylim = c(0,0.30), cex.axis=1.6)
lines(xx,std_norm, lwd =2.5)
lines(xx,std_norm_lower, lwd = 2.5, col ="red")

axis(1, at=c(0,3),labels=c(0, expression(mu)),cex.axis=2.0,las=1, tck
    =-0.005)
```

```
mtext("Probability density",2,line=2.5,cex=1.8,las=0)


abline(v=0, lwd=1.8)

abline(v=3, lty=2)

abline(h=0)


dev.off()
```

A.1.3 *Calculation of σ_0 Parameter for Truncated Normal Model*

The code below was used to obtain a numerical solution for a value of $sigma_0$, based on the use of the literature-derived parameter $\bar{x}_0 = 1064$ cpg, and matching the value of the mode between the lognormal and truncated normal distributions. A numerical solution was required as no analytical solution for $sigma_0$ of Equation 4.32 can be obtained.

```
# Preamble
    ----------------------------------------------------------------


# Define inverse erf function
invErf <- function(x) {

  qnorm((1+x)/2)/sqrt(2)

}

# Define erf function to use in calculating C
erf <- function(x) {

  2*pnorm(x*sqrt(2))-1

}

# define value of mean NoV from James Lowther harvest data Jan '11
xbar0 <- 1064

# define values of assurance level phi as applied in lognormal model
```

```
phi_array <- c(0.90,0.95,0.99)


sigma_array <- rep(0, length(phi_array))
sigma_tnorm_array <- rep(0, length(phi_array))
mu_array <- rep(0, length(phi_array))
mode_array <- rep(0, length(phi_array))



# function to calculate sigma for tnormal model
# sigma_0 numerical calculation
    ----------------------------------------------


# sigma_0 is the scale parameter for the truncated distrbution
# we know that the mean value = xbar_0 = 1064, so we use the definition
    for the mean of the truncated normal distrbution to
# numerically derive a value for sigma_0


# first declare the definition of the mean for tnorm as a function in
    terms of sigma and equal to zero


mu_0 <- mode_array[2]


function_sigma <- function (sigma) {


  mu_0 + sigma*(((1/sqrt(2*pi))*exp(-((-mu_0) ^2)/(2 * sigma ^2)))/(0.5 *
       (1 - (erf((- mu_0) / (sqrt(2) * sigma)))))) - xbar0


}


# solve function_mu using uniroot
```

```
sigma_0 = uniroot(function_sigma,c(0,10000))$root


# check that mean = xbar_0


test_mean <- mu_0 + sigma_0*(((1/sqrt(2*pi))*exp(-((-mu_0) ^2)/(2 * sigma
    _0 ^2)))/(0.5 * (1 - (erf(- mu_0 / (sqrt(2) * sigma_0))))))
```

A.1.4 *Minimum Depuration Times for Truncated Normal Model*

The 'R' code reproduced below generates numerical solutions to determine the minimum depuration time required using the truncated normal model.

```
b = 0.01339 # Dore 2010 depuration decay rate

time <- seq(from = 0, to = 600, by = 0.5)

# Vary the value of Psi for these calculations

psi_array <- seq(100,1000,length=1000)

mu_array_psi <- array(data = 0, dim=c(length(psi_array),length(time)))
sigma_array_psi <- array(data = 0, dim=c(length(psi_array),length(time)))
K_array <- array(data = 0, dim=c(length(psi_array),length(time)))

# array to store values of areas under PDF between 0 and each Psi, for
    each time[j]
prob_Area_MDT <- array(data = 0, dim=c(length(psi_array),length(time)))

# array to store results
MDT_psi <- rep(0,length(psi_array))

# cycle through values of Psi
for (i in 1:length(psi_array)){

  # cycle through values of time
  for (j in 1:length(time)){
```

```
    # calculate values of mu and sigma for each time step
    mu_array_psi[i,j] <- mu_0*exp(-b*time[[j]])
    sigma_array_psi[i,j] <- sigma_0*exp(-b*time[[j]])
    K_array[i,j] <- (0.5*(1-erf(-mu_array_psi[i,j]/(sqrt(2)*sigma_array_
        psi[i,j]))))

    # function defining the truncated normal distribution at time[j]
    pxt_area=function(x)
    {
      return((1/(sqrt(2*pi)*sigma_array_psi[i,j]*K_array[i,j]))*exp(-(x-
          mu_array_psi[i,j])^2/(2*sigma_array_psi[i,j]^2)))
    }

    ### calculate definite integral value for each bin
    prob_Area_MDT[i,j]=integrate(pxt_area, lower = 0, upper = psi_array[i
        ])$value

  }

  MDT_psi[i] <- (which.min(abs(prob_Area_MDT[i,]-phi)))/(length(time)/max
      (time))

}
```

A.1.5 *Non-linear Least Squares Regression of Neish Data*

Below is shown the R code executed to obtain parameter estimates for the proportion of observable NoV load (A), as well as the internal transfer rate (k)

from the unobservable to observable compartments within the oysters tested by Dr Anna Neish in 2013 [39].

```
# Nonlinear least squares regression to obtain best fit parameters

# Original data from Anna Neish
matrix_AN <- matrix(c(0,83951,172746,330798,279012,
                      42,676135,329798,183490,365956,
                      90,195117,131936,690449,585570,
                      162,143542,195646,151702,265712,
                      210,109121,62891,392198,324491,
                      258,297505,204854,212936,104772,
                      330,73709,66332,173570,185330),
                    byrow=T,nrow=7)

# Obtain time data from matrix_AN
time_AN <- matrix_AN[,1]

# Count number of columns excluding time data in first row
count_AN <- matrix_AN[,-1]

# Remove time data column from Matrix_AN and log values
logdata_AN <- log(count_AN)

# Calculate geometric mean values for each time point
# of AN data for each time point's data
geomean_AN <- c(exp(mean(logdata_AN[1,])),exp(mean(logdata_AN[2,])),
                            exp(mean(logdata_AN[3,])),exp(mean(logdata_AN
                                [4,])),
                    exp(mean(logdata_AN[5,])),exp(mean(logdata_AN[6,])),
```

```
                    exp(mean(logdata_AN[7,])))

# Use nls to derive best fit values for parameters A, x_0, b and k from
    matrix_AN
derived_nls <- nls(geomean_AN ~ x_0*(exp(-b*time_AN)+(((1-A)*k)/(A*(b-k))
    )*
                                    (exp(-k*time_AN)-exp(-b*time_AN))),
                   start = list(A = 0.4, x_0 = 550000, b = 0.005, k =
                       0.03))

# Extract best fit data from summary of nls
coef_BF <- summary(derived_nls)\$coefficients[,1]

BF_A1 <- coef_BF[[1]]          #best fit of proportion of norovirus in
    unobservable compartment
BF_x0 <- coef_BF[[2]]          #Best fit value from nls for mean of t=0 AN
     data -
BF_b1 <- coef_BF[[3]]          #Best fit depuration decay rate in hours
BF_k1 <- coef_BF[[4]]          #Best fit interior transfer rate in hours
```

A.1.6 *Simulation of Environmental Stochastic Model*

Shown below is the code executed in 'R' to carry out the simulation of pathgoen transferral from waste waters through sewage treatment works overflows due to rainfall, and into environmental estuary waters where pathogen levels are bioconcentrated by simulations of shellfish with compartmentalised pathogen caches.

```
```

```
# This script sets the parameters values for all the results and
    modelling for Chapter 6 Thesis



# Environmental Parameters
    ------------------------------------------------


capacity <- seq(from = 0, to = 25,by = 1) #rainfall level processed by
    STW (in mm)

time <- 1460      # run the simulation for this number of days = 4 years
                  # 4 years used due to life cycle of oyster form spat to
                       adult oyster

batchSize <- 30   # number of oysters in simulation

lambda <- 0.06907          # natural decay/dispersal rate of pathogen in
    environmental waters

alpha <- 0.49116           # gamma distribution shape parameter of
    rainfall event level (note is in mm)
                           # - based on last 10 years E&W rainfall
theta <- 6.3675            # gamma distribution rate parameter of
    rainfall levels
                           # - based on last 10 years E&W rainfall



 yyy <- seq(from=0, to=100, by=0.1) # domain to produce test plot of
     gamma distributed rainfall
#
```

```
plot(dgamma(yyy, shape = alpha,scale = theta), type="n", xlim=c(0,0.2),
    ylim=c(0,0.5),cex.axis=1.2,
     xlab= "Rainfall level (dm/day)", ylab="Probability", cex.lab=2)
curve(dgamma(x, shape=alpha, scale=theta), add=TRUE, col=1, lwd=3)


numEvents <- 192*(time/365)   # number of (expected?) STW events in time
    period
#192 per year is figure based on median of last 30 years E&W rainfall for
     rainfall >= 1mm in a day
# Parameter is used in randomly generating number of rainfall events
    across all 'time'


Area <- 400000000   # agglomeration area in dm^2
rho <- 2500         # input into STW pathogen concentration copies/litre
Volume <- 100000000 # volume of estuary in litres = dm^3


# Compartmental Parameters
    ------------------------------------------------


b <- 0.01339*24          # excretion rate of each oyster per day,
    equivalent to depuration rate
transfer <- 0.07453*24   # Internal transfer rate of pathogen from
    compartment y to compartment x, per day


eta <- 60                # oyster intake rate litres/day


############### STW Section array declarations ###################


# array to store up to 'batchSize' number of time series of pathogen
    loads
```

```
pathogenLoad <- array(0,dim=c(time,batchSize,length(capacity)))
pathogenLoad0 <- array(0,dim=c(time,batchSize,length(capacity)))
# blank array to copy pathogenLoad data to and from
x1 <- array(0,dim=c(time,batchSize,length(capacity)))


###############Oyster Section array declarations ##################


# array to store unobservable pathogen loads
preDigest <- array(0,dim=c(time,batchSize,length(capacity)))
# array to store observable pathogen loads
Digest <- array(0,dim=c(time,batchSize,length(capacity)))



# Main Loop for environmental simulation
    ---------------------------------


#timer for how long the loop takes to run
ptm <- proc.time()

# This outer loop runs through the capacity values in array 'capacity'
for (c in 1:length(capacity)){

  # This loop runs for the number of oysters in your batch
  for (k in 1:batchSize){

    # n is the random number of time periods from which to instigate an
        event
    # rbinom(n,size,prob) gives n observations from 'size' trials with
        probability success = 'prob'
```

```
# binomial does not use clustered or seasonal rainfall pattern, see
    AK re implement seasonality
n = rbinom(1,time,p=numEvents/time)
n <- ifelse(n = 0,1,n)

# randomly selected points in time for rainfall events to occur, and
    puts in ascending order
beginEvent = sort(sample(1:(time-1),n))       #not seasonal or
    clustered at the moment######################

# sets value of level of rainfall event
x1[beginEvent,k,c] = rgamma(n,shape=alpha, scale=theta) #gamma
    distributed random level of rainfall set for each beginEvent

#set the level of each beginEvent's discharge to what would
    constitute the overflow: rainfall minus capacity
x1[beginEvent,k,c] <- ifelse(x1[beginEvent,k,c]<=capacity[c],0,x1[
    beginEvent,k,c]-capacity[c])

# divide by 100 to set rainfall minus capacity to decimetre units
    instead of millimetres
x1[beginEvent,k,c] <- x1[beginEvent,k,c]/10

#obtain vector of times of overflow events, as beginEvent currently
    holds ALL rainfall events
beginEvent <- beginEvent[x1[beginEvent,k,c] != 0]

# copy x1 array into pathogenLoad array
pathogenLoad[beginEvent,k,c] <- x1[beginEvent,k,c]
pathogenLoad0[beginEvent,k,c] <- x1[beginEvent,k,c]
```

```
    # this loop goes through the array of overflow events
    for (i in 1:length(beginEvent)) {
      afterEvent <- beginEvent[i]
      # this loop calculates the compound value of each time point for
          each overflow event
      # 'overlaying' the decayed value of each previous event on top of
          each value
      for (j in (afterEvent+1):time){
        pathogenLoad[j,k,c] <- pathogenLoad[j,k,c]+pathogenLoad0[
            afterEvent,k,c]*exp(-lambda*(j-afterEvent))
      }
    }
    #loop to pass environmental pathogenLoad through each oyster's
        digestive and predigestive systems
    for (m in 1:(time-1)){
      preDigest[m+1,k,c] <- preDigest[m,k,c] + eta*pathogenLoad[m,k,c] -
          transfer*preDigest[m,k,c]
      Digest[m+1,k,c] <- Digest[m,k,c] + transfer*preDigest[m,k,c] - b*
          Digest[m,k,c]
    }
  }
  print(c)
}
# Stop the clock
proc.time() - ptm

######Allow system to reach 'steady' state, so remove first 100 time
    units from results arrays###########
```

```
equilPathogenLoad <- array(0,dim=c(length(101:time),batchSize,length(
    capacity)))
equilPathogenLoad[,,] <- pathogenLoad[101:time,,]
equilPreDigest <- array(0,dim=c(length(101:time),batchSize,length(
    capacity)))
equilPreDigest[,,] <- preDigest[101:time,,]
equilDigest <- array(0,dim=c(length(101:time),batchSize,length(capacity))
    )
equilDigest[,,] <- Digest[101:time,,]
```

A.1.7 *Simulating Harvest and Depuration of Results From Stochastic Model*

This Section provides the code used to obtain and simulate the harvest and subsequent depuration of a population of shellfish with pathogen compartmentalised between y_t and x_t. The values have been obtained from simulations of the stochastic environmental model detailed in Chapter 6.

```
# use depTime from depuration_plots.R

# copies array of unobservable pathogen loads during depuration for 1000
    sampled oysters
y_t <- depPreDigest # dim = c(oyster,capacity,depTime)

# copies array of observable pathogen loads during depuration for 1000
    sampled oysters
x_t <- depDigest # dim = c(oyster,capacity,depTime)
```

```
#Plot impact of capacity on minimum dep time for assurance levels phi =
    0.90,0.95,0.99

psi <- 200 # NoV load threshold limit value

phi = c(0.9,0.95,0.99) #assurance levels

quantile.results <- array(0,dim=c(length(phi),length(capacity),length(
    depTime)))
ecdf.results <- array(0,dim=c(2,length(capacity),length(depTime)))
minDepTime.results <- array(0,c(length(capacity),length(phi)))

# obtain quantiles for both x_t and y_t for all capacities and for each
    value of phi

#loop through all capacity values
for (c in 1:length(capacity)){

  #loop through all depuration time values
  for (t in 1:length(depTime)){
    # define function to calculate quantiles of observable values (x_t)
    quantile.function <- quantile(x_t[,c,t],probs=phi)
    quantile.results[,c,t] <- quantile.function
    ecdf.function <- ecdf(x_t[,c,t])
    ecdf.results[1,c,t] <- ecdf.function(psi)
    ecdf.results[2,c,t] <- 1-ecdf.function(psi)
  }
  # calculate the minimum depuration time needed to achieve each
  #  of the pathogen load assurance levels for each capacity value
  for (p in 1:length(phi)){
```

```
    minDepTime.results[c,p] <- min(which(ecdf.results[1,c,]>phi[[p]]))
  }
}

# this plots the pathogens loads values at the 90%, 95%, 99% assurance
    levels for each capacity

pdf(file="capacity_vs_predepuration_loads.pdf",width=9,height=7)

plot(capacity,quantile.results[3,,1],type="l",xlab="Capacity",ylab="
    Pathogen load",lwd=1.5,
     main="Capacity vs pre-depuration pathogen loads\nat 90%, 95%, 99%
         assurance levels")
lines(capacity,quantile.results[2,,1],lty=2,lwd=1.5)
lines(capacity,quantile.results[1,,1],lty=3,lwd=1.5)
abline(h=0,v=0)
# add a line denoting the pathogen threshold level limit psi
abline(h=psi,lty=4)
legend(17,800,c("90%","95%","99%"),lty=c(3,2,1),lwd=1.5, cex=1.6)
text(24,220,bquote(Psi), cex = 1.6)

dev.off()

# this plots minimum depuration time required for each of the 3 assurance
     levels 0.90,0.95,0.99
pdf(file="capacity_vs_minimum_depuration_times.pdf",width=9,height=7)

plot(capacity,minDepTime.results[,3],type="l",xlab="Capacity",ylab="Time"
    ,lwd=1.5,
```

```
     main="Capacity vs minimum depuration time required\nto achieve 90%,
         95%, 99% assurance levels")
lines(capacity,minDepTime.results[,2],lty=2,lwd=1.5)
lines(capacity,minDepTime.results[,1],lty=3,lwd=1.5)
legend(17.25,110,c("90%","95%","99%"),lty=c(3,2,1),lwd=1.5, cex=1.6)


dev.off()
```

B

APPENDIX B

B.1 MATHEMATICAL DERIVATIONS

B.1.1 *Alternative derivation of analytical solution for total NoV load, z_t*

In Section 5.4.2.3, we obtained an analytical solution for z_t by summing the analytical solutions of x_t and y_t. This solution can be obtained by using a different method, and so can confirm that we obtained the correct definition for the total NoV load. Equation 5.1 states that

$$z_t = x_t + y_t \,,$$

and so taking derivatives of both sides of this equation using the principle of linearity yields

$$\frac{dz_t}{dt} = \frac{dx_t}{dt} + \frac{dy_t}{dt} \,. \tag{B.1}$$

Equations 5.2 and 5.3 state that

$$\begin{aligned} \frac{dx_t}{dt} &= ky_t - bx_t \\ \frac{dy_t}{dt} &= -ky_t \,, \end{aligned}$$

therefore

$$\frac{dz_t}{dt} = -bx_t \,. \tag{B.2}$$

Integrating both sides of this equation with respect to t gives

$$\begin{aligned} \int \frac{dz_t}{dt} dt &= \int -bx_t dt \\ \Rightarrow z_t &= -b \int x_t dt \,. \end{aligned}$$

Substituting in above the solution for x_t as stated by Equation 5.16 yields

$$
\begin{aligned}
z_t &= -b \int \frac{x_0}{A} \left[\frac{k\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\} \right] dt \\
\Rightarrow z_t &= \frac{-bx_0}{A} \left[\int \frac{k\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\} \right] dt \\
\Rightarrow z_t &= \frac{-bx_0}{A} \left[\frac{k\,(1-A)}{(b-k)} \int \exp\{-kt\}\,dt + \frac{(Ab-k)}{(b-k)} \int \exp\{-bt\}\,dt \right] \\
\Rightarrow z_t &= \frac{-bx_0}{A} \left[\frac{k\,(1-A)}{(b-k)} \left(-\frac{1}{k} \right) \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \left(-\frac{1}{b} \right) \exp\{-bt\} + C \right] \\
\Rightarrow z_t &= \frac{x_0}{A} \left[\frac{b\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\} - bC \right], \qquad \text{(B.3)}
\end{aligned}
$$

where C is the constant of integration. The initial condition of $x_0 = Az_0$ allows us to restate this as

$$
z_t = z_0 \left[\frac{b\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\} \right] + D
$$

where $D = -bz_0C$. Setting $t = 0$ elicits a term for D ans subsequently C:

$$
\begin{aligned}
z_0 &= z_0 \left[\frac{b\,(1-A)}{(b-k)} + \frac{(Ab-k)}{(b-k)} \right] + D \\
\Rightarrow z_0 &= z_0 \left[\frac{b - Ab + Ab - k}{b-k} \right] + D \\
\Rightarrow z_0 &= z_0 \left[\frac{b-k}{b-k} \right] + D \\
\Rightarrow z_0 &= z_0 + D \\
\Rightarrow D &= 0 \\
\Rightarrow -bz_0C &= 0\,.
\end{aligned}
$$

As both $b, z_0 \neq 0$, it follows that $C = 0$. This allows us to state that, cf. Equation B.3,

$$
z_t = z_0 \left[\frac{b\,(1-A)}{(b-k)} \exp\{-kt\} + \frac{(Ab-k)}{(b-k)} \exp\{-bt\} \right], \qquad \text{(B.4)}
$$

which is equal to the definition derived in Section 5.4.2.3.

B.1.2 Θ_t *and* Ω_t *when* $t = 0$

Equations 5.16 – 5.20 describe the dynamics of the respective observable and total NoV loads within a single oyster for all $t \geqslant 0$. Equations 5.16 and 5.18

state that $x_t = x_0\,\Theta_t$ and $z_t = z_0\,\Omega_t$. From these two equations, we would expect that $\Theta_0 = 1$ and $\Omega_0 = 1$ when $t = 0$. This can be shown to be true by setting $t = 0$ within, firstly, Equation 5.19 and simplifying:

$$\begin{aligned}
\Theta_0 &= A^{-1}\left[\frac{k\,(1-A)}{(b-k)}\exp\{-k.0\} + \frac{(Ab-k)}{(b-k)}\exp\{-b.0\}\right] \\
\Rightarrow \Theta_0 &= A^{-1}\left[\frac{k\,(1-A)}{(b-k)} + \frac{(Ab-k)}{(b-k)}\right] \\
\Rightarrow \Theta_0 &= \left[\frac{k\,(1-A) + (Ab-k)}{A\,(b-k)}\right] \\
\Rightarrow \Theta_0 &= \left[\frac{k - Ak + Ab - k}{A\,(b-k)}\right] \\
\Rightarrow \Theta_0 &= \left[\frac{A\,(b-k)}{A\,(b-k)}\right] \\
\Rightarrow \Theta_0 &= 1\,.
\end{aligned}$$

Again by setting $t = 0$, we also check that Equation 5.20 conforms to expectations:

$$\begin{aligned}
\Omega_0 &= \left[\frac{b\,(1-A)}{(b-k)}\exp\{-k.0\} + \frac{(Ab-k)}{(b-k)}\exp\{-b.0\}\right] \\
\Rightarrow \Omega_0 &= \left[\frac{b\,(1-A)}{(b-k)} + \frac{(Ab-k)}{(b-k)}\right] \\
\Rightarrow \Omega_0 &= \left[\frac{b - Ab + Ab - k}{b-k}\right] \\
\Rightarrow \Omega_0 &= \left[\frac{b-k}{b-k}\right] \\
\Rightarrow \Omega_0 &= 1\,.
\end{aligned}$$

Therefore, as required, $\Theta_t = 1$ and $\Omega_t = 1$ when $t = 0$. This demonstrates that the analytical solutions conform with our initial condition assumptions.

C

APPENDIX C

C.1 HARVEST DATA

To be able to derive a mean value of NoV in oyster samples from harvest locations, harvest data must of course be used. Dr James Lowther of CEFAS carried out an extensive study of NoV loads in oysters from 39 harvest sites over the course of 2 years. This study assayed 10-homogenate oyster samples per site for NoV using the standard quantitative PCR test, from samples collected each month over the two year period. The 39 sites were comprised of 6 class A, 31 class B and 2 class C sites from around mainland Britain [2]. Class A sites were reported as having almost negligible NoV copies/g, and as these sites are also not currently required by legislation to depurate, we have not examined their data.

The class B and C data can be used to derive estimates of arithmetic mean values ($\bar{x}_0$) of NoV for each site and at specific time frames (Tables 2,3). However we need to be able to calculate μ_0 for our exponential depuration model, which is the geometric mean of the un-logged data, not the arithmetic mean. Geometric mean is defined as the arithmetic mean of log-data values that follow a log-normal distribution. However, knowing values of both $\bar{x}_0$

and σ_0 we can use the definition of the arithmetic mean of a log-normal distribution as stated by Equation 3.2 and solve for μ_0:

$$\begin{aligned} \bar{x}_0 &= \exp\left(\mu_0 + \frac{1}{2}\sigma_0^2\right) \\ \Rightarrow \ln(\bar{x}_0) &= \mu_0 + \frac{1}{2}\sigma_0^2 \\ \Rightarrow \mu_0 &= \ln(\bar{x}_0) - \frac{1}{2}\sigma_0^2 \end{aligned} \tag{C.1}$$

Using $\sigma_0 = 1.645$ (cf. Section 3.4) and Equation C.1, we can calculate values of μ_0 for both harvest classifications at the high and low temperature time-points.

SITE NO.	Jul '09	Jan '10	Jul '10	Jan '11
11	$< 100^a$	13272	$< 40^b$	15369
14	$< 100^a$	6249	$< 40^b$	5042^c
$\bar{x}_0$	70	9761	20	10206
μ_0	2.896	7.833	1.643	7.878

Table C.1: NoV load cpg for class C sites at low and high temperature points throughout study duration, along with calculated means. [a] NoV loads recorded as < 40 cpg are designated as having value = 20 cpg (the midpoint between 0 and 40), [b] while < 100 cpg are quantified as 70 cpg (the midpoint between 40 and 100). [c] Nearest data to Jan '11 was Oct '10.

SITE NO.	Jul '09	Jan '10	Jul '10	Jan '11
1	< 40 [a]	609	< 100 [b]	774
2	< 40	2554	< 40	1990
3	< 40	< 40	< 40	< 40
4	< 40	1240	< 40	3274
5	< 100	10883	< 40	3657
6	< 40	< 40	< 40 [c]	< 100
7	< 40	348	< 40	3944
8	< 40	< 100	< 40	152
9	< 40	300 [d]	< 40	2739
12	< 100	2055 [e]	< 40	3114
13	< 100	1852	< 40	366
15	< 40	1123	< 40	< 100
16	< 40	300 [d]	< 40	1401
17	< 40	748	< 40	414
18	< 100	512	< 40	1312
19	< 40	< 100	< 40	465
20	< 40	1043	< 100	517
21	< 40 [f]	300 [d]	< 40	655
22	< 40	300 [d]	< 40	769
23	< 40	294	< 40	300 [d, g]
24	140	< 100	< 40	657 [h]
25	< 100	3638	357 [i]	235

SITE NO.	Jul '09	Jan '10	Jul '10	Jan '11
26	< 100	< 100	< 40	< 100
27	< 100	300 [d]	< 40	2756
28	< 100	< 100	< 40	532
29	< 40	< 100	< 100	< 100
30	< 100	300 [d]	< 40	824
32	< 100	300 [d, j]	< 40	< 100
33	< 40	1869	< 40	1175 [k]
35	< 40	< 40	< 40	< 40
37	< 40	1549	< 100	564
$\bar{x}_0$	40	1062	38	1064
μ_0	2.336	5.615	2.285	5.617

Table C.2: Genotype II NoV load cpg for class B sites at low and high temperature points through study duration, as well as calculated means. [a] NoV loads recorded as < 100 cpg are designated as having value = 70 cpg (midpoint between 40 and 100), [b] while < 40 cpg are quantified as 20 cpg (the midpoint between 0 and 40). [c] Nearest data to Jul '10 was Jun '10. [d] Midpoint of 100-500 cpg. [e] Nearest date to Jan '10 was Feb '10. [f] Nearest date to Jul '09 was Aug '09. [g] Nearest date to Jan '11 was Oct '10. [h] Nearest date to Jan '11 was Feb '11. [i] Nearest date to Jul '10 was Aug '10. [j] Nearest date to Jan '10 was Dec '09. [k] Nearest date to Jan '10 was Feb '10.

D

APPENDIX D

D.1 RAINFALL DATA

D.1.1 *England & Wales Rainfall Levels*

This appendix provides details of the analysis of rainfall data for England & Wales covering the last ten years. UK rainfall data is available online from the Met Office [84]. Daily rainfall data for England and Wales from 1931–2015 was obtained, with the last ten years of data analysed for a best fit distribution type. This was accomplished using a distribution fitting program called Easy Fit [92], which ranks a large number of distribution types against each year's rainfall data, and provides the parameter values used in the rankings per distribution.

The rankings are based upon goodness of fit tests carried out against a wide range of different distribution types. The software provided rankings based on goodness of fit results from Kolmogorov-Smirnov, Ryan-Joiner and Anderson-Darling tests. For our purposes, we used the Kolmogorov-Smirnov results and the gamma distribution ranked very highly, if not the best fit, for each year's rainfall examined. The other tests also returned very high rankings for the gamma distribution.

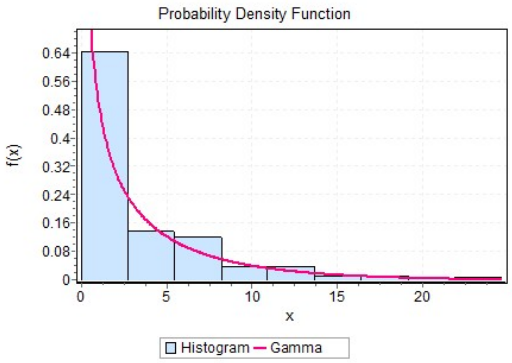
The histograms of each year's rainfall data for 2006–2015 with fitted Gamma distributions are shown in Figure D.0, whereas the ranking and parameters for the Gamma distribution for each year are shown in Table D.1.

Note that all values of the shape parameter $0 < k < 1$. This results in the the PDF being a monotonic decreasing function, with $P(x) \to 0$ as $x \to \infty$, and $P(x) \to \infty$ as $x \to 0$.

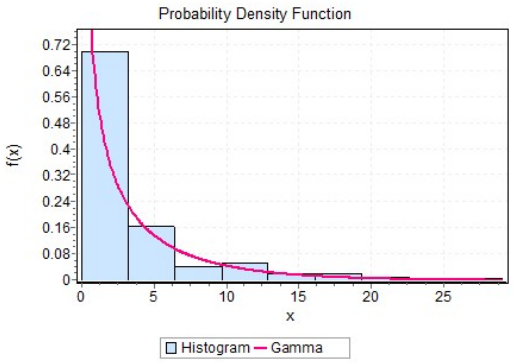
Other notable distribution types which ranked highly using this process were the beta, lognormal, Weibull and Pearson Type VI distributions.

Year	Rank	k	Θ	No. of rain days
2006	7/51	0.48668	6.7107	189
2007	9/51	0.45875	6.624	193
2008	1/51	0.49296	7.0049	220
2009	3/51	0.56039	5.6353	203
2010	2/52	0.42119	5.6354	173
2011	7/51	0.55637	5.3008	205
2012	2/51	0.50921	7.3544	213
2013	1/52	0.39687	6.6758	174
2014	6/51	0.55568	5.6596	213
2015	1/52	0.4735	7.0737	209
Mean	3.9	0.49116	6.3675	199.2

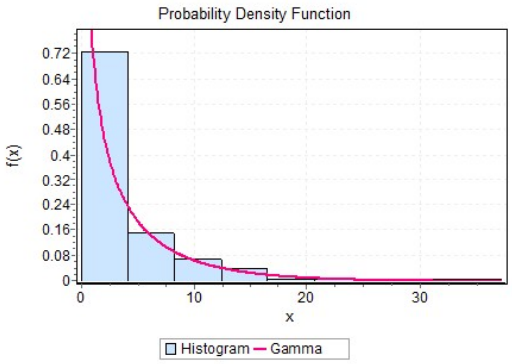
Table D.1: Rainfall 2006–2015 data — Gamma distribution's Kolmogorov-Smirnov goodness of fit ranking, with distribution parameters describing each year's rainfall. Final column reports the number of days with recorded rainfall > 1 mm



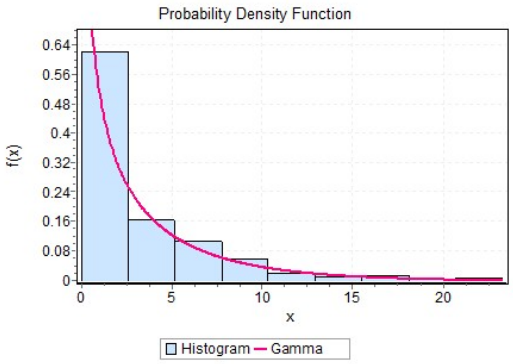
(a) 2006 Rainfall



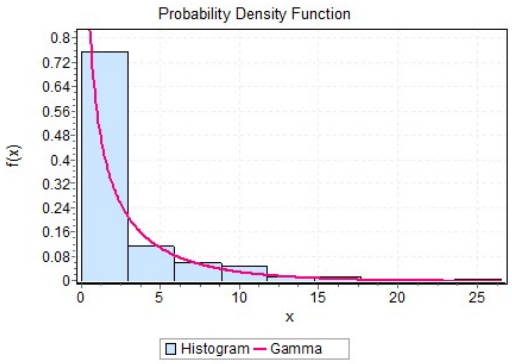
(b) 2007 Rainfall



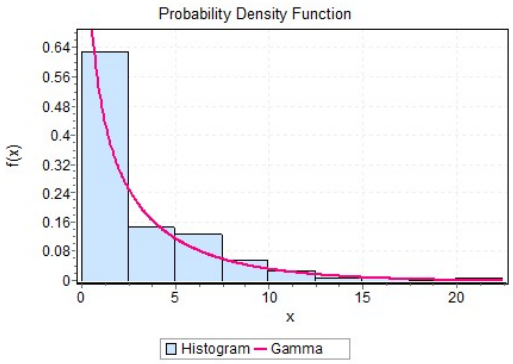
(c) 2008 Rainfall



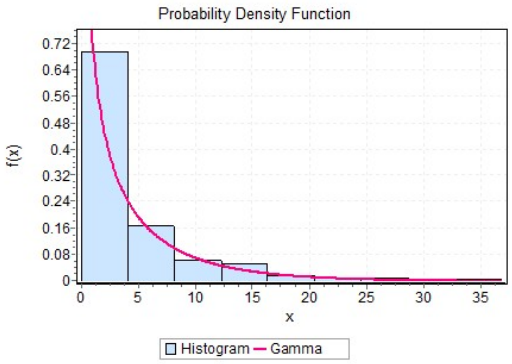
(d) 2009 Rainfall



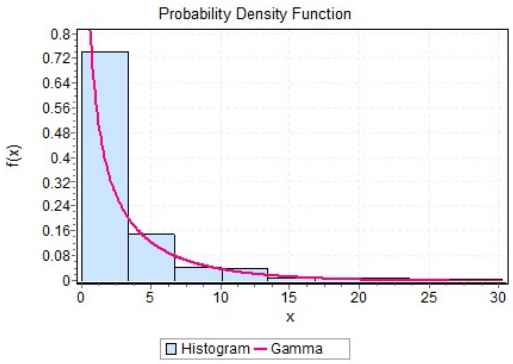
(e) 2010 Rainfall



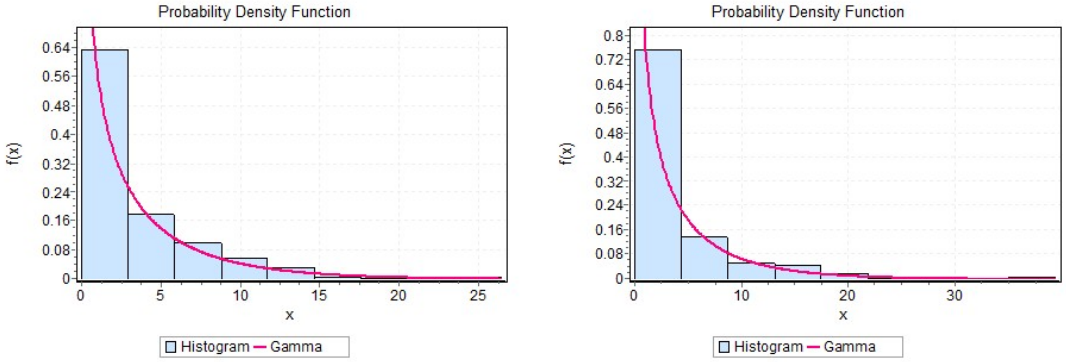
(f) 2011 Rainfall



(g) 2012 Rainfall



(h) 2013 Rainfall



(i) 2014 Rainfall

(j) 2015 Rainfall

Figure D.0: 2006–2015 England and Wales rainfall density histograms with fitted Gamma distribution curves using parameters from Table D.1

D.1.2 *Frequency of England and Wales Rainfall*

Years	Mean	Median	Q_1	Q_3	Min.	Max.
1986–2015	189.8	192	173.25	204.5	158	232
2006–2015	199.2	204	190	212	173	220

Table D.2: Rainfall 1986–2015 data — 30 years of England and Wales rainfall data for days with rainfall > 1 mm

Again analysing the rainfall data from the Met Office [84], we can ascertain the number of days that rainfall was recorded for each year, assessing the last 30 years of available data for 1986–2015. Table D.2 shows that between 2006–2015, England and Wales experienced days with rainfall > 1 mm within a range of 173–220 days per year. The average number of rain days is 199.2, with a standard deviation of 16.4 days. median number of rain days is 204.

For the last 30 years, the location of the number of rainfall days is much lower in comparison with only the 2006–2015 data, and boxplots of both data are shown in Figure D.1. Both timeframes exhibit a slight negative skew towards the upper values of the range. As the data is not symmetric, the median values are more appropriate measures of location than the mean.

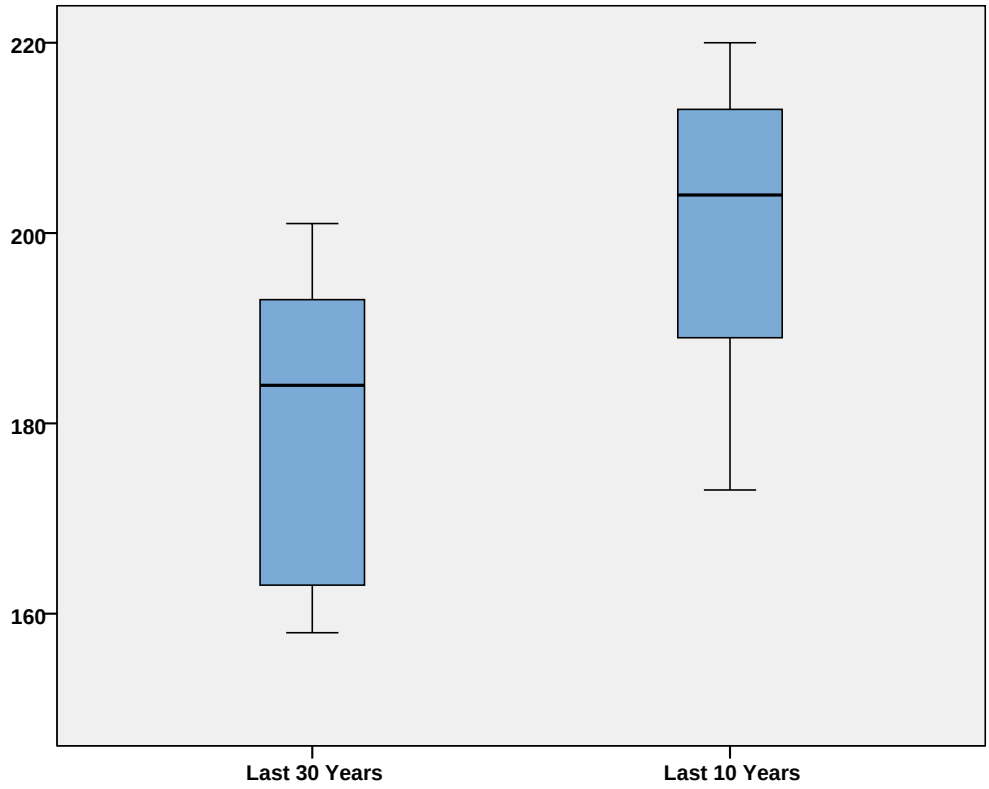


Figure D.1: Boxplots of 1986–2015 and 2006–2015 England and Wales rainfall, for days with rainfall > 1 mm