Statistical Analysis of Bathing Water Quality in Scotland

Ruth Alison Haggarty

A Dissertation Submitted to the

University of Glasgow

for the degree of

Master of Science

Department of Statistics

September 2008

Abstract

It is of interest, both environmentally and economically, for the water quality at beaches to be maintained and improved wherever possible. In 2006 a new European Community Directive was introduced which set compliance standards in terms of percentile values of different microbial indicators and, provided the public has been informed of the water quality via electronic message signs, permits samples to be discounted from compliance calculations. Consequently, the initial research question posed concerned the definition of a single sample limit (SSL) which could be used to determine the quality of a single sample of bathing water, whether or not it could be discounted and whether or not this could be set generically. The focus of the work later changed to become the definition of discounting limits that could be used to identify the samples which should be removed from the dataset on which compliance with the 2006 Directive is based.

Chapter 1 provides an introduction to the general context of the problem, a description of the data and gives details of how compliance is assessed. In Chapter 2 exploratory analysis of the data revealed extensive variation in each of the microbiological indicators considered, both across the bathing water sites and within the same site across different bathing seasons. The distribution of each of two microbial indicators, faecal streptoccoci (FS) and faecal coliforms (FC) was considered and other features of the data including multiple outliers and

evidence of bimodality were also apparent at some locations. All of this indicated that the definition of a generic single sample limit would not be achievable. The assumption of log-normality, on which the calculation of percentiles used to assess compliance with the Directive is based, was also investigated. Chapter 3 then used the level of compliance achieved in 2007 (using the data from bathing seasons in 2004 - 2007) as a measure of outcome in order to assess the effectiveness of several different candidate definitions of a single sample limit, including two site independent values and one formulaic approach.

Following on from the issues discovered in the initial exploratory analysis of the data and after discussion with SEPA a generic SSL did not seem feasible and hence the stated objective of the work was modified to identifying a discounting limit which could be used to identify samples that could potentially be removed from compliance calculations. Therefore from Chapter 4 onwards only discounting limits are considered and the idea of using extreme value models became the basis of the remaining chapters. Chapter 4 considers the use of extreme value theory, in particular block maxima, k-th largest order statistic and threshold models to identify suitable return levels which could be applied as discounting limits across all sites. The differences between the return level limits obtained from each of the models, their impact on the levels of compliance classification when all counts exceeding the limits were removed and the inclusion of a relevant covariate within the block maxima model were also considered here.

Chapter 5 focused on site specific threshold models, in particular, for the locations where electronic message signs are currently in place. The quantity of data removed at each site and the robustness of the discounting limits found using these models was also examined here. Finally, Chapter 6 provides a summary of the findings and discusses limitations of the study and possible future directions.

Acknowledgements

First and foremost I would like to thank my supervisors Prof. Marian Scott and Dr Claire Ferguson for their invaluable guidance and support throughout the past year. I would also like to thank the Scottish Environment Protection Agency for providing the data for this project and in particular, thanks are due to Ruth Stidson, Chinny Iroegbu and Colin Gray for their time, knowledge and advice. In addition, I gratefully acknowledge the funding from the Engineering and Physical Sciences Research Council that allowed me to undertake this work.

Thanks are due to all of my family and friends (old and new!) who have supported me along the way; either through being there for me or for providing distractions when they were needed, it has been very much appreciated!

Finally, to my mum and dad, thank you for your love, patience, encouragement and for never allowing me to doubt myself.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Background

Monitoring levels of water pollution has been a key focus of research and legislation in recent years. Following from this there has been increased investment and interest in the improvement of Scottish beaches and it is consequently of great importance, both environmentally and economically, to maintain a high standard of water quality at the sites throughout Scotland which are currently designated as bathing waters. In 2006, a revised European Community Bathing Water Directive (European Parliament, 2006) was introduced which set compliance standards for bathing waters in terms of safe limits for microbial, physical and chemical indicator quantities. These standards are required to be met and reported on annually by member states; in Scotland the Scottish Environment Protection Agency (SEPA) is the regulatory agency responsible for monitoring water quality and reporting back to the European Community.

In addition to the 2006 Bathing Water Directive, there have been several other pieces of European Community legislation which have been brought into force to assess and set targets for water quality criteria. For example, the Nitrates Directive (European Parliament, 1991) was introduced in 1991 with an aim to both identify polluted water environments and reduce the levels of nitrate pollution from agricultural sources. Similarly to the Bathing Water Directive, the Nitrates Directive sets specific limits for pollutants which cannot be exceeded and requires regular monitoring to be carried out. Furthermore, the 2000 European Community Water Framework Directive (WFD) (European Parliament, 2000) establishes a legal framework for the assessment and improvement of lakes and other surface waters across Europe. However, in contrast to the Bathing Water Directive which sets specific standards that should be achieved by 2015, whilst the WFD requires that 'good status' should be achieved in all lakes by a set date (2016), it does not specify the exact definition of 'good status' in terms of any particular indicators.

SEPA defines a bathing water to be "fresh or sea water where bathing is either explicitly authorised and is traditionally practised by a large number of bathers, or is not prohibited" (SEPA, 2007). The 61 locations across Scotland which are currently designated as bathing waters are shown in Figure 1.1.

As well as setting compliance guidelines for the protection of public health and the environment, the Bathing Water Directive outlines requirements with regards to frequency of sampling, methods of analysis, interpretation of the results obtained and the circumstances in which samples can be discounted from compliance calculations. Within this Directive limits are based on overall percentiles of 4 years of data and hence there is no specific limit set for assessing the quality of a single sample of water.

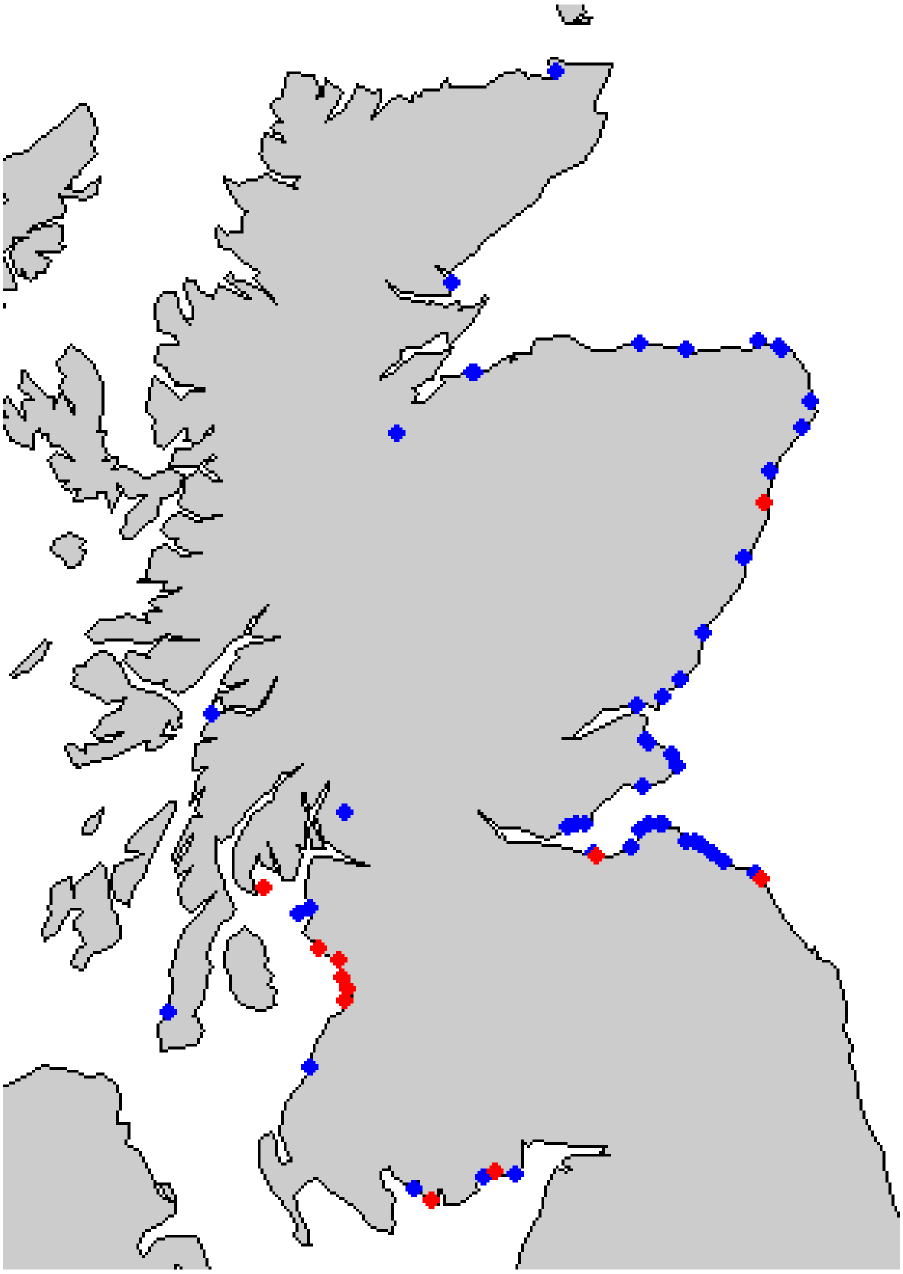

Figure 1.1. Location of bathing water sites in Scotland (2007)

Map of Scotland showing the location of sites classed as bathing waters by SEPA in 2007. Sites where electronic signs are in place are shown in red.

The initial aim of this research project was to determine the 'best' generic Single Sample Limit (SSL). A SSL can be defined as a single numerical value that serves two purposes; it can be used to determine whether or not advice against bathing should be issued to beach users and can also be used to identify which samples should be removed from the compliance dataset as part of the discounting procedure which is outlined in the new Directive. Further details of the discounting procedure are provided in Section 1.3. However after discussion with SEPA, at an early stage in the project, the focus of the work changed to be concerned only with the definition of discounting limits. The reasons for this are outlined fully in Chapters 2 and 3. A discounting limit identifies samples which can potentially be removed from the dataset on which compliance is based. The distinction between a SSL and a discounting limit is that while an SSL both protects human health (via predictions of poor water quality) and identifies the samples which can be removed, a discounting limit is only used for the latter of these two purposes.

1.2 Assessing Bathing Water Quality

In order to assess the quality of bathing water, samples are taken from each of the sites and are then analysed in labs to quantify the presence of different microbiological indicator bacteria including, but not exclusively, faecal streptococci (FS) and faecal coliforms (FC). Each of these indicators are expressed in terms of a count of colony forming units per 100ml (cfu/100ml). FS is referred to as Intestinal Enterococci in the EU and WHO documentation although both names are taken to represent the same group of bacteria. Based on the 90th and 95th percentiles of the samples over the current bathing season along with the previous

3 bathing seasons, the Directive sets limits which can be used to classify each of the bathing water sites into one of 4 categories; ‘Excellent’, ‘Good’, ‘Sufficient’, or ‘Poor’ (see Table 1.1).

Similarly to many biological and environmental determinants, FS and FC are thought to be $\log_{10}$ normally distributed at each of the sites and so this has been taken into account when calculating the percentiles. In accordance with the Directive, the 90th and 95th percentiles of the samples for each site are calculated using the procedure outlined below.

1. Convert each of the counts using a $\log_{10}$ transformation
2. Take the arithmetic mean of the $\log_{10}$ value of all samples ($\bar{x}$)
3. Find the standard deviation of the $\log_{10}$ value of all samples (s)
4. The 90th percentile is given by antilog($\bar{x}$ + 1.282 s)
5. The 95th percentile is given by antilog($\bar{x}$ + 1.65 s)

In fact, SEPA use more accurate normal quantile values to calculate the percentiles than those provided in the Directive; 1.281552 for the 90th percentile, and 1.644854 for the 95th. These more precise values are used in all subsequent percentile calculations within this thesis.

A site is classed as ‘Excellent’ or ‘Good’ if the 95th percentile falls below the respective stated guideline figures provided by the Directive, if it fails the requirement for both of these categories; the 90th percentile is compared to the ‘Sufficient’ value. If subsequently the 90th percentile exceeds this limit, then the site is classed as ‘Poor’. By the end of 2015, EU member states must aim to classify all bathing waters as ‘Good’ or ‘excellent’ while ensuring that all bathing waters at least meet the standards for the ‘Sufficient’ classification.

Table 1.1 below shows the numerical values for each of the four compliance categories given in the directive.

Parameter	**Excellent**	**Good**	**Sufficient**
FS	100(*)	200(*)	185(**)
FC	250(*)	500(*)	500(**)

(*) Based on 95th percentiles
(**) Based on 90th percentiles
Bathing Water is classed as Poor if it fails the conditions for Sufficient

Table 1.1. EC limits (cfu/100ml) for compliance categories (European Parliament, 2006)

There are also World Health Organization (WHO) guideline values for these indicators at bathing water sites and it was thought that these should be taken into account when considering different approaches to the definition of a SSL. The 2003 WHO Guidelines for Safe Recreational Water Environments (WHO, 2006) identify different microbial levels and the increased estimated risk of illness per exposure. Table 1.2 shows the guideline values for the microbial quality of recreational waters. The range of 95th percentile values for FS which *"represents a substantial elevation in the probability of all adverse health outcomes for which dose-response data are available"* is 201 - 500 cfu/100ml. Within this range it is believed that the risk of gastrointestinal illness from one exposure is between 1 and 5%, and for acute febrile respiratory illness is between 0.3 and 1.9%.

1.3 Discounting of Samples

The Directive encourages public awareness of the current water quality wherever possible. SEPA uses a statistical model which incorporates rainfall as a

Category	Estimated Risk Per Exposure
< 40 **A**	< 1 % GI illness risk < 0.3% AFRI risk
41 - 200 **B**	< 1 - 5% GI illness risk < 0.3 - 1.9% AFRI risk
201 - 500 **C**	5 - 10% GI illness risk < 1.9 - 3.9% AFRI risk
> 500 **D**	< 10% GI illness risk < 3.9% AFRI risk
GI- Gastrointestinal AFRI - Acute Febrile Respiratory Illness	

Table 1.2. Table of WHO guideline values (WHO, 2006)

variable to calculate daily predictions of the water quality at each site and at 11 of these sites (shown on Figure 1.1 in red) the predictions are updated and reported to beach users through the use of electronic variable message signs to indicate the standard of water quality. Under the new legislation, provided the public has been informed of short term water pollution episodes, up to 15% of the samples taken per year at each site can be discounted and not included in the figures used to gauge compliance. Therefore, it is only at the 11 signed sites that discounting of samples is currently acceptable. In practice, it is hoped that discounting will reduce the 4 year percentiles and could potentially improve compliance.

From each sample of bathing water taken there is a count obtained for FS and FC; however the discounting procedure is carried out separately for each of these two indicators. For example, for any given site and year, the water sample which contained the largest number of FS (cfu/100ml) may not have contained the largest number of FC (cfu/100ml). In such cases only the value for FS is removed from the corresponding percentile calculation and the value of FC is retained. If a site is classified differently according to each of the two indicators, then the overall compliance classification of the site is the poorer of the two categories.

Although the discounting limits considered may be varied from site to site, it is assumed that the limit used to identify whether or not the bathing water quality is poor will remain constant across all sites. This means that if a sample at one site indicates advice against bathing should be issued, the same count would be equally as unsafe at all other sites. Furthermore, for any given site the limit used to protect human health cannot be greater than the discounting limit in order to ensure that the samples which are removed from compliance calculations have already come from locations where poor water quality has been predicted.

From a single sample a count, x, is obtained for each indicator variable.

- If x exceeds the limit used to protect human health then the water quality is predicted as poor and the sample can *potentially* be discounted from the compliance dataset
- If x also exceeds the discounting limit then the sample is removed from the compliance dataset

For each variable, a maximum of 15% of samples can be discounted in each bathing season.

1.4 Faecal Pollution and Associated Health Risks

Since both FS and FC occur naturally in the gut of humans, and other warm blooded animals, their presence is an indication of sewage in the water which has not been adequately treated. However, there are a wide variety of potential sources of FS and FC at bathing water sites. Diffuse pollution does not have a single identifiable origin; it consists of pollution resulting from several different sources and land use activities. Each of these sources is indirect and although may only individually contribute a small amount of waste, together they can be collectively significant. One such source is agricultural run-off; for example rainfall washes manure used as fertilizer or livestock waste from surrounding fields either directly into the water itself or into connecting streams. In addition to waste from grazing livestock such as sheep and cattle, excrement from animals such as dogs, rodents, and birds such as swans, geese and seagulls can also increase the levels of bacteria in the water . Unspecified urban diffuse sources such as runoff from roads, faulty domestic septic tanks and un-sewered wastewater from boats are also all potential sources of faecal contamination (Environment Agency, 2007; Georgiou and Langford, 2006; WHO, 2006).

It is known that by ingestion or infection through wounds or mucous membranes, the bacteria in sewage can cause illness (Cabelli *et al.*, 1979; Rees, 1993). However, not all of the microbiological organisms in recreational bathing waters can cause disease; their presence indicates only faecal contamination (Efstratiou, 2001). FS and FC are not themselves thought to be the causative agents of disease. Large quantities of these indicator bacteria at a bathing water site signify pathenogenic micro-organisms may be present and therefore advice against bathing should be issued. The most common risks to human health from these disease causing bacteria are infections of the ears, eyes or skin, infections of the upper respiratory tract and most frequently, enteric illnesses such as gastroenteritis. Viruses which can lead to poliomyelitis, hepatitis A and meningitis have also been discovered in sewage polluted waters although there have been no reported cases of these which are thought to have been contracted through exposure to bathing waters (Rees, 1993; Georgiou and Langford, 2006).

1.5 Monitoring Bathing Waters

1.5.1 Sampling and Factors Affecting Water Quality

Under the directive a bathing water profile has to be created for each site which consists of a physical and geographical description of the bathing water, and the location of the monitoring point. This is the point where the samples used to gauge compliance will be taken from and should ideally be from within an area where either most bathers are expected, or where it is known there is an increased quantity of pollutants. Samples of at least 250ml in volume must be taken 30 centimeters below the surface of the water in an area which has a minimum depth of one metre. The profile must also contain the potential sources and rate of growth of pollutants and the expected rates and origins of short term pollution episodes.

There are several factors which will affect the short-term quality of the water, the most dominant being meteorological variables such as temperature, the number of sunshine hours and the quantity of rainfall. Unusual weather conditions can adversely affect water quality if the sample is taken immediately after; if this occurs, then an Abnormal Weather Waiver can be issued. This means the sample will be disregarded from the compliance calculations although, unlike the samples which can be discounted from signed sites, a replacement sample is taken within 7 days after the end of the abnormal weather event.

Guidelines for the frequency of sampling are also outlined within the directive. Samples are taken throughout the bathing season, which in Scotland, is approximately a 4 month period between the beginning of June and mid September. During this period no fewer than four samples should be analysed and one additional sample has to be taken before the beginning of the bathing season each year. The water quality at each of the 61 sites has to be monitored continuously throughout the bathing season, with samples being taken at intervals which last no longer than one month. SEPA takes around 20 samples from each of the 61 sites, although there tends to be fewer samples taken from sites which have been consistently classified as 'excellent' in previous years. The samples are taken at semi-random dates throughout the season, with at least one being taken at the weekend. However, there are restrictions on the times when samples can be taken due to the geographical locations of some of the sites and the fact that it is recommended samples are analysed on the same day they are taken.

1.5.2 Limits of Detection

Problems can occur in the analysis of some samples due to the equipment used to take measurements of bacteria. Scientific equipment often has saturation levels either above, or below which the exact quantity cannot be confirmed. Since the way in which limits of detection are dealt with could potentially alter the values of the percentiles there is some question as to how to treat these values which are effectively right or left censored observations. Both FS and FC have upper and lower limits of detection. For those samples which are identified as being greater than the upper limit of detection, SEPA includes the upper limit of detection in the percentile calculations, whilst for the lower limit, half the stated value is used since it is believed that the true count of the sample will lie somewhere between zero and this value.

It would invalidate any conclusions reached to simply ignore values which are affected by limits of detection and so they must be included in the analysis in some way. Eastoe *et al.* (2006) investigated different ways of handling censored observations in an environmental context concerning air pollutants and indicated it was necessary to include the values of censored observations rather than ignoring them altogether or by replacing them with a nominal fixed constant. Throughout this thesis observations which have been marked as being at the limits of detection have been dealt with in the same was as SEPA and have been included in any calculations or analysis.

1.6 Overview of Thesis

The difficulties associated with identifying a generic SSL are highlighted in Chapters 2 and 3 with the main focus changing to the identification of discounting limits in the remainder of the thesis. Specifically, in Chapter 2 the distribution of the indicators FS and FC is explored and the assumption that the data comes from a $\log_{10}$ population at each site is investigated. Following from this, Chapter 3 uses the level of compliance achieved as a measure of outcome in order to assess the effects on compliance of several different approaches to the definition a single sample limit. A sensitivity study also considers how compliance changes when the percentile used to assess the 'Sufficient' classification changes.

As mentioned previously, from Chapter 4 onwards the thesis focuses solely on the definition of discounting limits and extreme value analysis of the data is used to obtain different discounting limits; both site independent and site specific. In Chapter 4 block maxima, k-th largest order statistic and threshold models are fitted to data from all sites. This is done separately for each of the microbial indicators. Subsequently the return level limits obtained from these models have been applied to the data as discounting limits. The impact of including salinity as a covariate within the block maxima model is also considered here. Chapter 5 develops the use of threshold models further and applies models at individual site level in order to obtain suitable return level based discounting limits for each of the bathing water sites where electronic signage is currently in place. Finally, Chapter 6 ends with a summary of the work presented and discussion of limitations and future directions.

Chapter 2

Exploratory Analysis

The exploratory analysis has focussed on graphical tools to explore the distribution of the results at the individual sites and differences, if any, at the same site over the 4 year period from 2004 to 2007. While there were initially only 10 signed sites at which electronic signs were in place to inform the public of the water quality, a sign was later introduced at one additional site (Eyemouth) during the production of this thesis. Therefore, although the preliminary investigations only mention 10 signed sites, all subsequent analysis has been carried out on 11. The names and the number of samples collected in each bathing season between 2004 and 2007 at each of these signed sites is shown in Table 2.1.

	No. of Samples				
Site	2004	2005	2006	2007	Total
Portobello Central	20	20	20	19	79
Eyemouth	*20*	*20*	*20*	*19*	*79*
Sandyhills	19	20	20	19	78
Saltcoats/Ardrossan	20	20	20	19	79
Irvine	20	20	20	19	79
Troon	20	20	20	19	79
Prestwick	20	22	20	19	81
Ayr	20	20	20	19	79
Brighhouse	20	21	20	19	80
Ettrick Bay	20	20	20	19	79
Aberdeen	20	20	20	19	79

Table 2.1. Names of signed sites and number of samples taken (2004 - 2007)
Eyemouth is included in analysis of signed sites from Chapter 3 onwards

2.1 Descriptive Statistics

Currently samples can only be discounted from compliance calculations at sites where electronic signage is in place, therefore, these sites were initially explored in greater detail. Table 2.2 contains the summary statistics for all of the signed sites over the four bathing seasons between 2004 and 2007, while individual boxplots for each of the sites and the corresponding five number summaries are shown in Figures 2.1 and 2.2 and Tables 2.3 and 2.4.

For both microbial indicators, the summary statistics and boxplots reveal a high degree of variability in the data. The across year variation is clear, for example values of FS at the signed sites range from 1 to 1500 in 2004 and from 1 to 520 in 2005. In an attempt to reduce this variability, the data were standardized and the boxplots redrawn, however this added very little to the earlier impressions of the data and there continued to be extensive variation across both the sites,

FS	2004	2005	2006	2007
Min	1	1	1	4
1st Qu	8.5	14.75	10	10
Median	24	30	40	40
Mean	166.1	79	143.9	120.4
3rd Qu	88.25	100.5	92.5	90
Max	1500	520	1160	710
FC	2004	2005	2006	2007
Min	2	5	5	14
1st Qu	18.5	10	67.5	110
Median	85	55	160	140
Mean	546.1	184.2	457.1	833.8
3rd Qu	415	127.5	387.5	260
Max	6700	1500	4000	5900

Table 2.2. Summary statistics for the signed bathing water sites in each bathing season (2004 - 2007)

and to a greater extent, within the same site across different years. Notably, 2007 appears to have been a particularly poor year with regard to FC, while 2004 seems worse for both variables. The extensive across site heterogeneity is also apparent. For example, in 2004 the maximum observed value for FC at the Ettrick Bay site was 5900 cfu/100ml which is more than 35 times greater than the maximum value at the Portobello Central site in the same year.

2.2 Distribution of Data with respect to European Guidelines

Reference lines relating to the European Bathing Water Directive were included on the boxplots (Figures 2.1, 2.2) in order to assess the distribution of

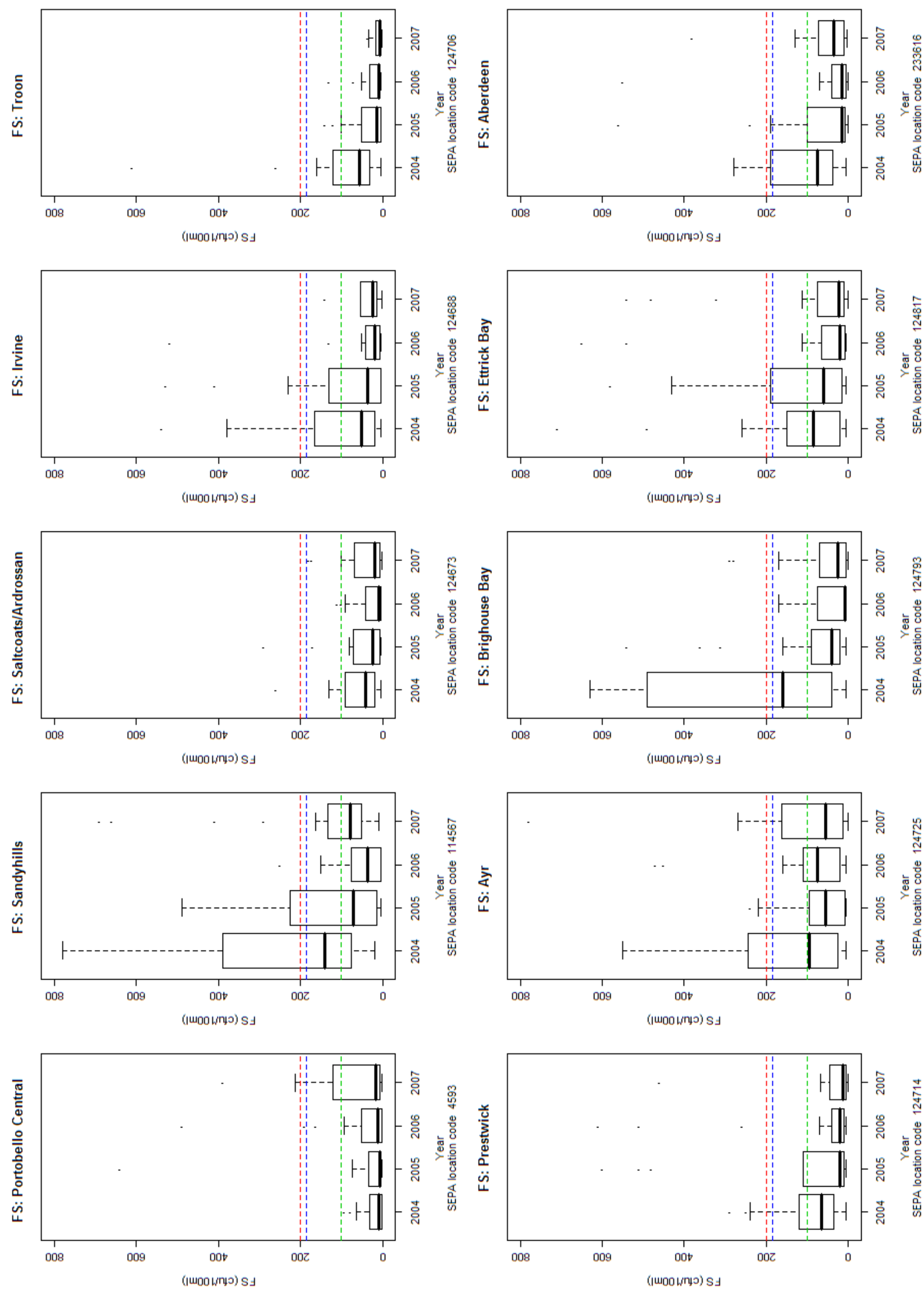

Figure 2.1. Boxplots of FS at Signed Sites

Boxplots of FS for each of the 10 signed sites, separately for year. The green line represents the Bathing Water Directive threshold for ‘Excellent’ water quality (based on 95th -percentile evaluation), the red line represents ‘Good’ (based on a 90-th percentile evaluation), and the blue line represents ‘Sufficient’ (based on 90th-percentile evaluation).

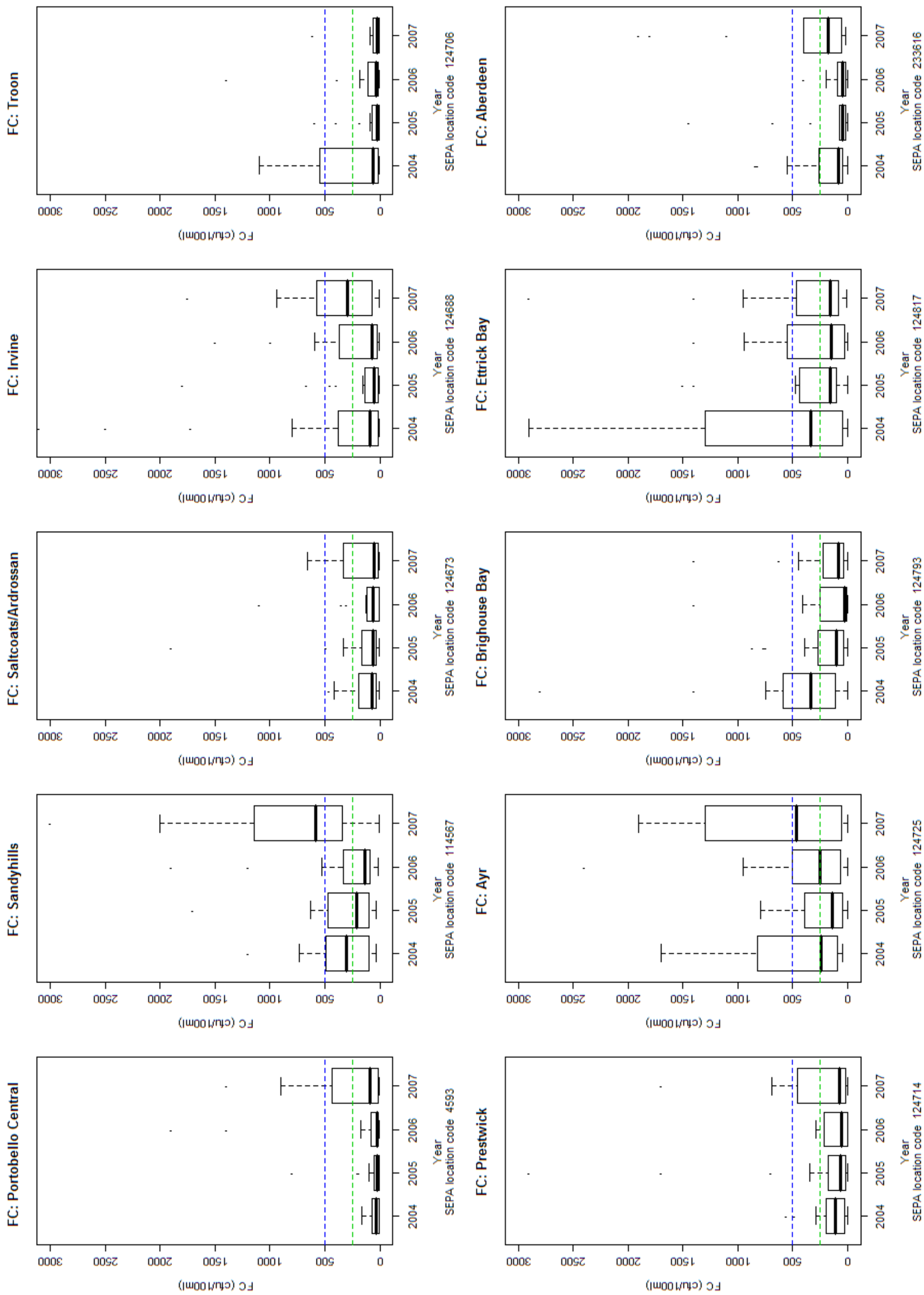

Figure 2.2. Boxplots of FC at Signed Sites

Boxplots of FC for each of the 10 signed sites, separately for year. The blue line represents the Bathing Water Directive threshold for both ‘Good’ (based on 95th-percentile evaluation) and ‘Sufficient’ (based on 90th-percentile evaluation). Similarly, the green line indicates the threshold for ‘Excellent’ water quality.

	FS. Signed Sites (2004 - 2007)					
YEAR	Site	Min	1st Qu	Median	3rd Qu	Max
2004	Portobello Central	1	1	11	40	390
	Sandyhills	20	70	120	350	780
	Saltcoats/Ardrossan	1	10	40	80	260
	Irvine	1	20	40	140	520
	Troon	2	8	30	100	1600
	Prestwick	2	13	55	100	1160
	Ayr	5	32	95	210	1500
	Brighouse	2	12	80	290	2500
	Ettrick Bay	1	10	58	140	930
	Aberdeen	4	20	51.5	162	5400
2005	Portobello Central	1	4	7	34	640
	Sandyhills	5	15	70	205	490
	Saltcoats/Ardrossan	5	7.5	25	70	290
	Irvine	5	5	35	130	530
	Troon	5	5	15	50	140
	Prestwick	5	15	20	110	1000
	Ayr	5	7.5	55	95	240
	Brighouse	5	15	30	75	540
	Ettrick Bay	5	15	60	190	1000
	Aberdeen	1	6.5	15	99	560
2006	Portobello Central	1	2	12	50	490
	Sandyhills	5	5	30	75	250
	Saltcoats/Ardrossan	5	5	10	40	110
	Irvine	5	7.5	20	40	520
	Troon	5	7.5	10	30	130
	Prestwick	5	7.5	20	30	610
	Ayr	5	20	75	110	470
	Brighouse	5	5	7.5	75	310
	Ettrick Bay	5	7.5	20	65	650
	Aberdeen	1	4	14.5	41	1500
2007	Portobello Central	1	8	26	83	1250
	Sandyhills	38	52	72	102	690
	Saltcoats/Ardrossan	1	5	20	34	1200
	Irvine	1	12	30	140	1500
	Troon	1	2	4	5	34
	Prestwick	1	4	6	50	1440
	Ayr	1	4	10	16	1500
	Brighouse	1	26	52	118	170
	Ettrick Bay	2	16	24	112	540
	Aberdeen	2	8	39	99	970

Table 2.3. Signed site summary statistics; FS (2004 - 2007)

	FC. Signed Sites (2004 - 2007)					
YEAR	Site	Min	1st Qu	Median	3rd Qu	Max
2004	Portobello Central	1	6	34	74	168
	Sandyhills	30	100	300	495	1200
	Saltcoats/Ardrossan	5	30	75	195	460
	Irvine	5	10	90	375	3100
	Troon	5	10	65	545	1100
	Prestwick	5	25	110	195	5900
	Ayr	50	90	240	820	1700
	Brighouse	5	110	340	585	2800
	Ettrick Bay	5	50	340	1300	5900
	Aberdeen	2	47	81	262	3500
2005	Portobello Central	1	7	24	54	800
	Sandyhills	30	100	215	475	1700
	Saltcoats/Ardrossan	5	30	65	160	1900
	Irvine	5	10	50	135	1800
	Troon	5	7.5	25	70	590
	Prestwick	5	20	65	180	2900
	Ayr	5	45	145	395	790
	Brighouse	5	40	100	270	870
	Ettrick Bay	5	100	155	435	4000
	Aberdeen	2	22	50	74	1450
2006	Portobello Central	1	11	23	84	1900
	Sandyhills	10	90	135	330	1900
	Saltcoats/Ardrossan	5	7.5	60	115	1100
	Irvine	5	20	75	370	5500
	Troon	5	10	35	110	1400
	Prestwick	5	5	60	215	5300
	Ayr	5	70	250	505	2400
	Brighouse	5	10	30	250	1400
	Ettrick Bay	5	25	150	550	7600
	Aberdeen	2	15	47	98	6700
2007	Portobello Central	1	19	86	438	7500
	Sandyhills	6	345	580	1145	3500
	Saltcoats/Ardrossan	2	14	48	330	15000
	Irvine	2	72	290	575	13000
	Troon	1	5	28	58	610
	Prestwick	1	22	78	460	11200
	Ayr	1	52	470	1300	15000
	Brighouse	4	39	80	225	3200
	Ettrick Bay	12	86	160	470	3200
	Aberdeen	18	55	174	400	1900

Table 2.4. Signed site summary statistics; FC (2004 - 2007)

the data in relation to the classification limits set. For FC, the same numerical value is used as the guideline limit for 'Good' and 'Sufficient' compliance, however while good is judged on the 95th percentile of the samples, 'Sufficient' is based on the 90th percentile. It is for the same reason that for FS, the limit for good compliance appears to be lower than the limit for the sufficient category.

Again there appears to be a high degree of variation amongst the sites. For example, for FS in 2004, while 3 of the signed sites could be classified as 'Good', there are several other sites where approximately 50% of the samples taken exceed the limit for this classification category.

2.3 Bimodality

It is evident from the boxplots and summary statistics that the distribution of the data is often highly skewed and at many sites there are several clear statistical outliers. One of the simplest rules for determining a single sample limit would be identification of individual outliers since their removal through discounting could effectively result in a reduced percentile value and therefore may improve the levels of compliance. However, dot-plots which were drawn in order to provide a clearer graphical display of the distribution demonstrate a potential problem with this approach. There is evidence of bimodality at some of the sites which can cause the 95th percentile to appear greater than the maximum recorded value. For example, at the Sandyhills site, if we consider only the samples taken in 2004 the maximum observed value for FS is 780, however due to the presence of several large counts which elevate the standard deviation, this value is less than the annual 95th percentile which has been calculated to be 969.

Dotplots for the Sandyhills site are shown in Figure 2.3. The red and blue lines respectively represent the **four year** 90th and 95th percentiles for this site. For this particular example, whilst removal of outliers in 2005 may improve compliance for 2004, and to some extent for 2007, there appears to be some evidence of bimodality, particularly when considering the 90th percentile.

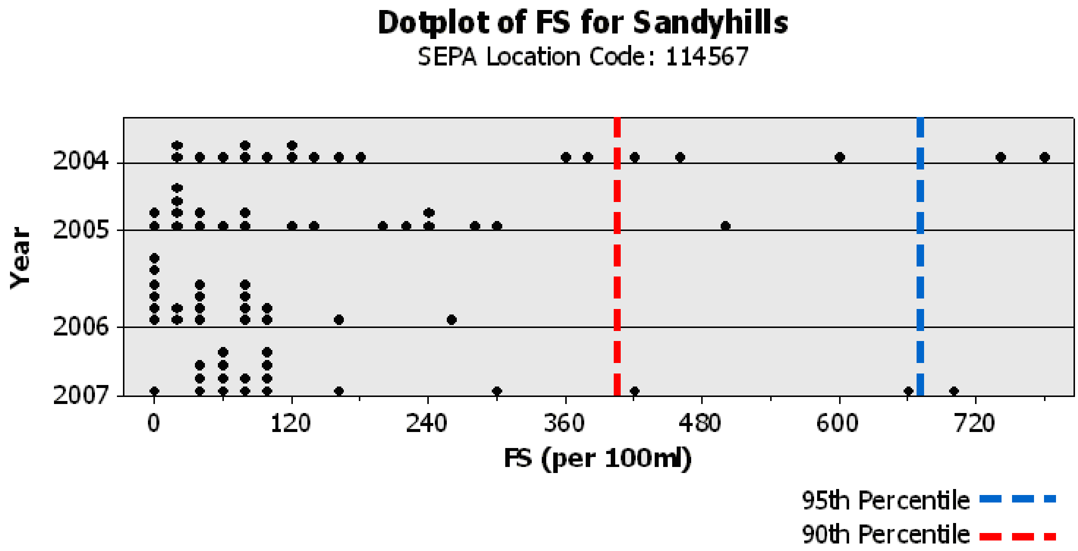

Figure 2.3. Dotplots of FS at Sandyhills site 2004 - 2007

In light of this preliminary analysis, due to the extreme variability in the data across both the different sites and the different years, and the distributional problems of bimodality, there does not appear to be a straightforward way to set a generic single sample limit which can be applied to all sites. It may therefore be more appropriate to explore a formula based definition of the SSL; a single rule could then be evaluated in terms of the data from each specific site.

2.4 Assessment of Distributional Assumptions

As mentioned previously, similarly to many environmental quantities, it is believed that each of the microbiological indicators, FS and FC, are $\log_{10}$-normally distributed at each bathing water site and so the Directive's methods of calculating percentiles are based on this assumption. It was clear from preliminary analysis that the data were positively skewed and although the $\log_{10}$-normal distribution is commonly used to model contaminant concentrations of this type, there are several possible alternative explanations for the high degree of asymmetry in the data. Biased sampling, the presence of one or more outliers and bimodality could all influence the calculation of the mean of the distribution at each site and may therefore have an effect on the subsequent compliance classification. There are several assumptions that are required to accurately calculate the 90th and 95th percentiles which are used to gauge compliance; including that;

- the $\log_{10}$ data points are normally distributed,
- there are very few censored observations and,
- the data have come from a single statistical population.

However, there are problems with the validity of these assumptions at some of the sites, particularly where there appears to be some indication of bimodality within the observed data.

Figure 2.4 illustrates some of the distributional problems which have been encountered. A histogram of the raw data for FS at the Elie (Harbour) and Earlsferry site appears to be highly skewed. However, there is some suggestion of either multiple outliers, or a smaller second distribution centered around 70 cfu/100ml. The curvature in the Q-Q plot obtained for this site (Figure 2.4) is another clear indication that the raw data is not normally distributed.

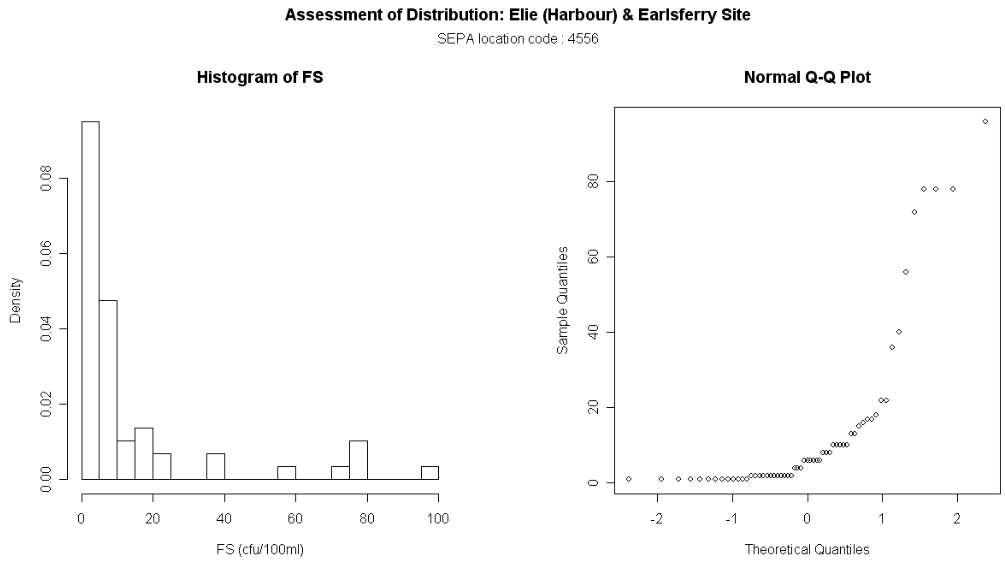

Figure 2.4. Histogram and Normal Q-Q plot of FS (cfu/100ml) at Elie Harbour site (4556)

Following from this a histogram and normal Q-Q plot were subsequently obtained for the $\log_{10}$ transformed data at the same site (Figure 2.5); if the data had originated from a $\log_{10}$-normal distribution then the transformed data should follow a normal distribution. As can be seen, the $\log_{10}$ data continues to be positively skewed and although there has been a marked improvement in the linearity of the residuals in the Q-Q plot, normality is still questionable. In addition to the Q-Q plot, a Shapiro-Wilks test of normality was performed. The p-value returned for this particular site was 0.0005 which provided further evidence that the $\log_{10}$ transformed sample did not come from a normally distributed population.

Despite the problems at selected sites there are several locations where the assumption of log normality seems reasonable. An example of one such site is

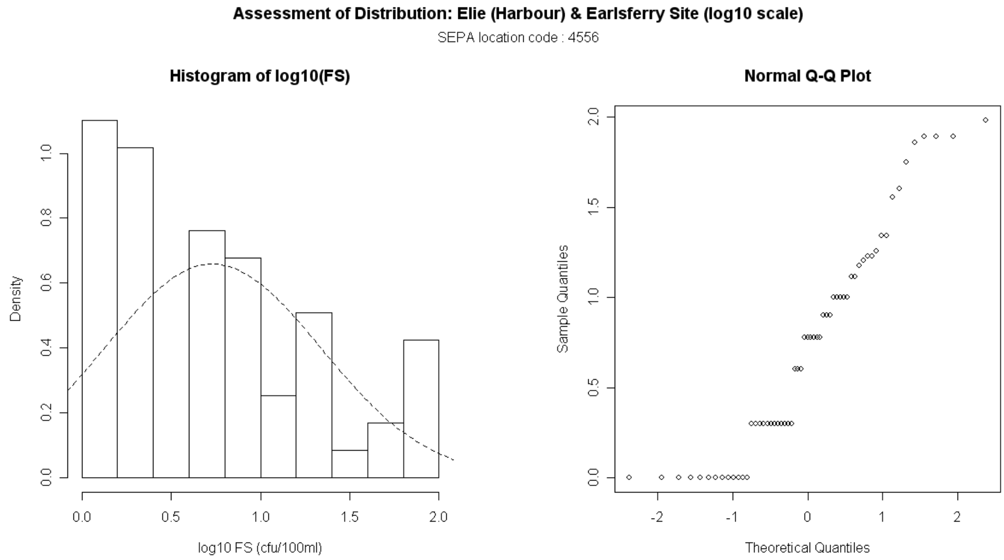

Figure 2.5. Histogram and Normal Q-Q plot of $\log_{10}$ FS (cfu/100ml) at Elie Harbour site (4556)

shown in Figure 2.6 which consists of the histogram and Normal Q-Q plot for the raw FS observations at the Cruden Bay site. Again, we can see that the sample is highly skewed with some evidence of outliers, however, in view of the plots of the $\log_{10}$ transformed data at this site (Figure 2.7), a normal distribution does appear to be reasonable. The Q-Q plot seems linear and although there is some divergence towards the tails this could possibly be where there are fewer observations, or alternatively where there are observations which have been affected by limits of detection. At the Cruden Bay site, a Shapiro-Wilks test of normality gave a p-value of approximately 0.56 indicating there is insufficient evidence to reject the null hypothesis that the sample has come from a normally distributed population.

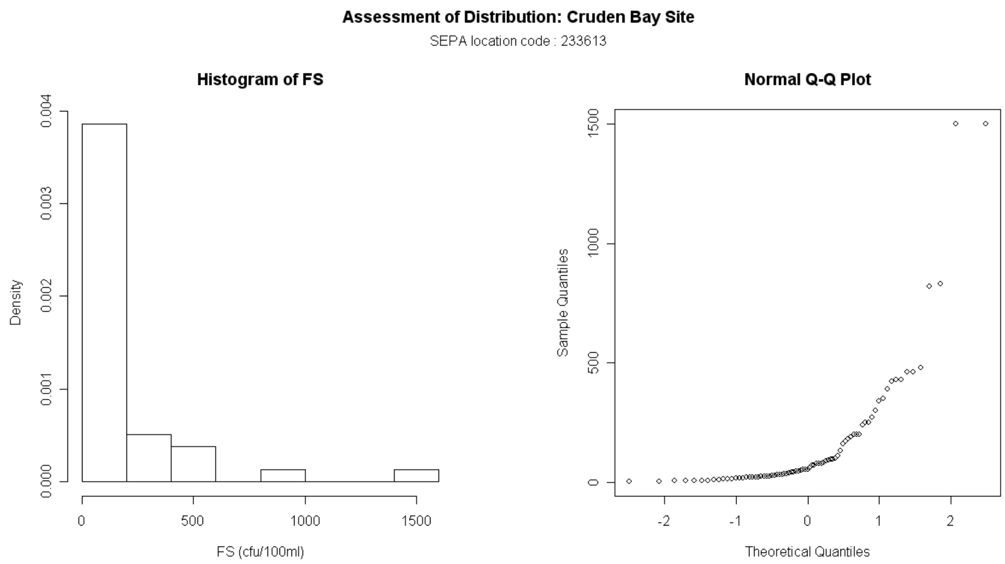

Figure 2.6. Histogram and Normal QQ Plot of FS (cfu/100ml) at Cruden Bay site (233613)

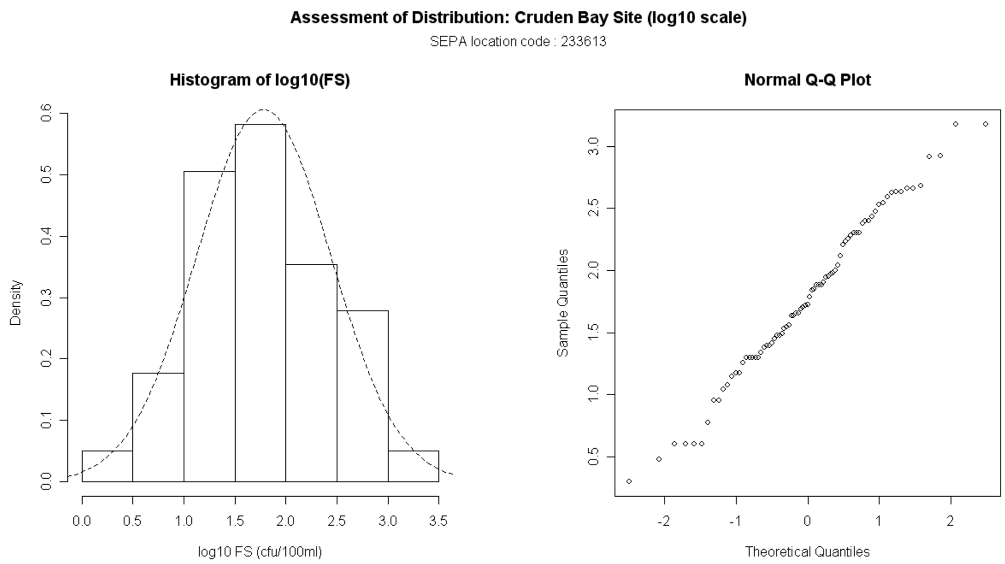

Figure 2.7. Histogram and Normal QQ Plot of $\log_{10}$FS (cfu/100ml) at Cruden Bay site (233613)

This variation between the sites is consistent with our initial impression of the distributions obtained from box plots of the data. While it is clear that at the majority of sites the data have come from a $\log_{10}$-normal distribution there are also several sites where problems arise. As noted previously, at some locations asymmetry in the distribution can appear deceptively large due to either multi-modality or the presence of multiple outliers. Although this can potentially result in the 90th percentile of the data, on which compliance is assessed, being misleadingly large, in practical terms, failure in the assumption of log-normality may not have a huge impact on the percentile calculations.

When a Shapiro Wilks distributional test was carried out at each site on the $\log_{10}$ counts only around 19% of sites appeared to have normally distributed data in terms of FS and around 44% in terms of FC. However, since these hypothesis tests have been carried out on a moderately large sample size of approximately 80 observations they may be particularly sensitive to inconsistencies in the data such as outliers. This has already been established as one of the notable features of this dataset. In order to further examine the true distribution of the data, the theoretical 90th percentile, calculated according to the directive, was compared to the empirical 90th percentile to see if these two values were similar. The theoretical percentile was obtained from the cumulative distribution function (CDF) formed using the mean and standard deviation estimated from the data but based on the distributional form while the empirical cumulative distribution function (ECDF) is formed using only the observed data. It was found that when the percentiles based on the distributional assumption of log-normality were compared to those obtained from the ECDF at each site, there was often little difference between the two. At the sites where there was a divergence between the theoretical percentile and that based on the ECDF, both values tended to be comparatively low in relation to the overall range of values. This implies that in practical terms, an apparent violation of the assumption of log-normality is not a problem.

Figure 2.8 shows plots of the empirical cumulative distribution function of $\log_{10}$ FS at the sites where the distributional assumptions were previously discussed (Elie Harbour and Cruden Bay). The 90th percentiles of the empirical functions have been shown as well as the normal curve which represents the corresponding theoretical distribution of data at each site.

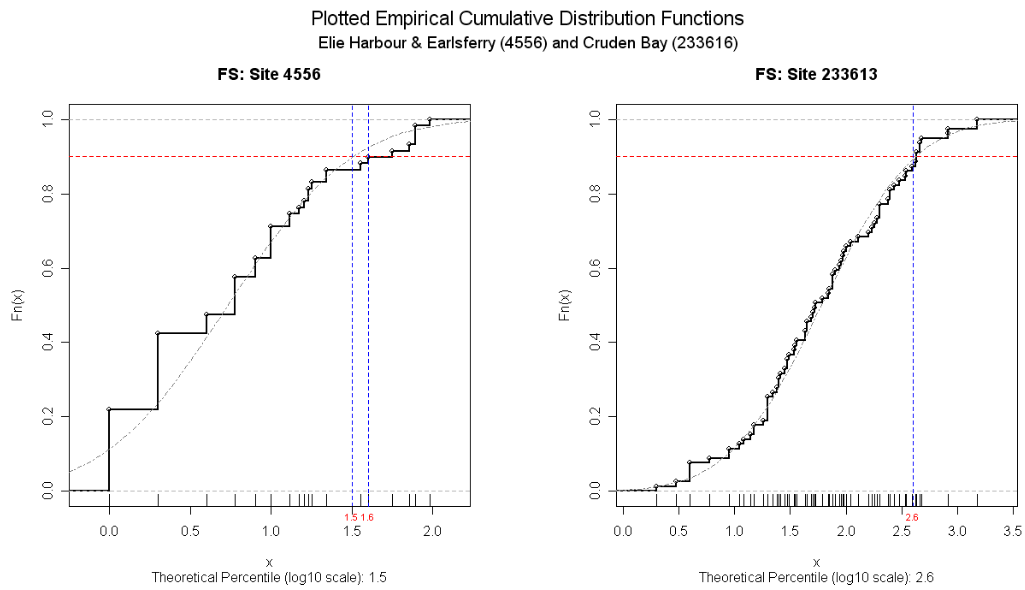

Figure 2.8. ECDF of FS (cfu/100ml) at Elie (Harbour) & Earlsferry and Cruden Bay sites

As can be seen, despite the initial concerns about bimodality and the consequent validity of the distributional assumption at the Elie (Harbour) and Earlsferry site there is very little difference between the empirical and theoretical distribution curves. The small difference between the percentile values, even when back transformed onto the original scale, is nothing which would affect the compliance classification of the site. In terms of values which can be interpreted in terms of the directive the distributional based 90th percentile for this site is approximately 32 cfu/100ml while the data driven empirical 90th percentile is 40 cfu/100ml; both of which are far lower than even the strictest EC guideline standard for 'Excellent' compliance. This is typical of the situation at most of the sites. If there is any difference between the 90th percentiles obtained from the plotted empirical distribution functions and those obtained from the calculations based on the assumption of $\log_{10}$-normality it is unlikely to affect the compliance classification of the site.

Classification Based on Empirical Percentiles

The 2007 compliance classification of each site was first calculated according to the directive guidelines based on the theoretical parametric percentiles, and then again based on the empirical percentiles obtained from the data. When the number of sites which met the mandatory standard of sufficient under each of these calculations were compared, based on the empirical percentiles there was only one additional site which fell into the sufficient category. Out of the 61 sites there were 49 where the classifications agreed exactly and 12 at which there was some difference. At 10 sites there was a difference of one classification class and at 2 locations there was a difference of 2 classes. At the sites where there was a discrepancy between the classifications, the parametric method did not consistently classify the sites as poorer than the empirical method or vice versa.

Table 2.5 shows the 2007 classification of each of the signed sites based on the parametric percentiles as well as the classification based on empirical percentiles.

As can be seen there is only one signed site where there is a difference in the classification based on parametric and empirical percentiles. Of the 61 bathing

2004 - 2007 Compliance Classification		
Site	**Parametric**	**Empirical**
Portobello Central	Sufficient	Sufficient
Eyemouth	Poor	Poor
Sandyhills	Poor	Poor
Saltcoats/Ardrossan	Poor	Good
Irvine	Poor	Poor
Troon	Sufficient	Sufficient
Prestwick	Poor	Poor
Ayr	Poor	Poor
Brighhouse	Poor	Poor
Ettrick Bay	Poor	Poor
Aberdeen	Poor	Poor
% Sufficient (or better)	18.2	27.3

Table 2.5. 2007 EC compliance classification (signed sites) based on parametric and empirical percentile values

water sites considered, Saltcoats/Ardrossan is the only one which meets the minimum required standard based on the empirical percentile but fails to do so when the theoretical percentile is used. At first this difference of two classes ('Poor'/'Good') may cause concern however on closer inspection, both the parametric and the empirical cases the values obtained are very close to the 'Sufficient' category boundaries. Prior to any discounting, it is only the FC indicator that does not meet the 'Sufficient' standard and the value of the parametric 90th percentile for FC is 513 cfu/100ml, which is only slightly larger than the sufficient category limit of 500 cfu/100ml.

2.5 Summary

From the preliminary analysis the most apparent feature of the data has been the extensive variation. There is also evidence of bimodality at some locations which brings into question the assumption that the observations have come from a single, $\log_{10}$ normal distribution at each site. The validity of the percentiles based on this parametric assumption was assessed and comparisons with data-driven percentiles have shown that, at the sites where the parametric assumption does not appear to hold, the distribution curves for the theoretical and the data driven distributions are in general fairly close and in most cases the classification is the same for both methods.

In Chapter 3 the feasibility of finding a generic single sample limit is considered by exploring the implications on the level of compliance achieved when different SSL definitions are applied.

Chapter 3

The Effects of Discounting on Compliance

In order to assess the effect of different single sample limits, the number of sites which met the Directive's 'Sufficient' standard was used as a measure of outcome. As mentioned previously, the SSL is concerned both with identifying which samples should be removed from the compliance dataset and with protecting human health and so site independent approaches to discounting were considered first. Discounting is currently only permitted at sites where there are electronic message signs in place since it is only at these locations that the public can be informed of up to date bathing water conditions. While the impacts of the different SSLs considered in this chapter were initially investigated only at the signed sites, the impact of each of the rules on all 61 sites has also been included.

In Section 3.1 simple methods of removing the maximum values at each site are investigated as well as a sensitivity study of changes in the level of compliance achieved when the percentile value used to gauge 'Sufficient' classification was altered. Section 3.2 weighs up the merits and drawbacks of two different site independent approaches to the SSL; one pragmatic definition and one that is based on World Health Organization guideline standards. Following from this, in Section 3.3 a site specific formulaic definition of an SSL which involves adjusting the geometric mean of the samples at each site in order to obtain individual limits is explored.

3.1 Removing Annual Maxima

The most straightforward method of discounting is to simply remove the maximum value (or top two annual site maxima) from each site in each bathing season and then observe the change in compliance achieved when the percentiles are recalculated. Table 3.1 shows the European Directive compliance classification in 2007 for each of the 11 locations where electronic signage is currently in place. The classification of each site is also shown when the site specific annual maximum value is removed and when the largest two annual values are removed.

As can be seen, without any discounting, only 2 of the 11 signed sites meet the mandatory EC standard of ‘Sufficient’. While removal of the annual maxima from percentile calculations resulted in only one additional site meeting the minimum required standard, the two sites which were previously classed as ‘Sufficient’ moved into the ‘Good’ category. Also removing the second highest annual value from each site increased the percentage of ‘Sufficient’ sites again; however over half still failed to reach the required standard. The removal of the top two annual values at each site is valid in terms of removing 15% of the data.

Electronic signage was introduced at locations for reasons which were neither entirely random, nor on the basis of prior knowledge about water quality.

Overall Compliance			
Site	Remove 0	Remove 1	Remove 2
Portobello Central	Sufficient	Good	Good
Eyemouth	Poor	Poor	Poor
Sandyhills	Poor	Poor	Poor
Saltcoats/Ardrossan	Poor	Sufficient	Good
Irvine	Poor	Poor	Poor
Troon	Sufficient	Good	Good
Prestwick	Poor	Poor	Sufficient
Ayr	Poor	Poor	Poor
Brighouse	Poor	Poor	Poor
Ettrick Bay	Poor	Poor	Poor
Aberdeen	Poor	Poor	Sufficient
% Sufficient (or better)	18.2	27.3	45.5

Table 3.1. 2007 EC compliance classification (signed sites) after removing 0, 1 and 2 samples

However, it is known that at a number of the signed locations the water quality had previously been poor and consequently there was an increased need to keep the public informed of bathing water conditions at these sites. In view of this it seems worthwhile to look at the effects of discounting on compliance of all 61 sites. Furthermore, it is envisaged that the number of places where discounting can be applied will increase as electronic signage is planned for additional sites within the next year. Following from this, Table 3.2 shows the percentages of all 61 bathing water sites which were sufficient or better in 2007 and, as before, shows the results when the annual site maxima or maximum two annual counts are removed.

Before removing any observations, a much greater proportion of sites - around 67% - met the minimum standards in 2007 compared to just 18.2% of the signed sites. Removal of each site's largest two annual values from the 90th percentile

Overall Compliance All Sites	
	%Sufficient (or better)
Remove 0	67.2
Remove 1	72.1
Remove 2	77

Table 3.2. Percentage of all sites compliant in 2007)

calculations increased this level of compliance by a further 10% although there continues to be around a quarter of sites which are classed as 'Poor'. The revised Directive states that all sites must be at least sufficient by the end of the bathing season in 2015.

The limited improvement in the number of compliant sites using this method of discounting could be anticipated due to the evidence of bimodality and multiple outliers at some sites as highlighted in the preliminary analysis of the data. These features of the data mean that removal of one or two large observations will do little in terms of reducing the percentile values in certain cases.

3.1.1 Sensitivity Study

In addition to investigating the effects of removing the top two annual values, the way in which the proportion of sites that were (at least) sufficient changed when the percentiles used to assess compliance were altered was also investigated. While it is unlikely that the percentile used within the Directive will change it was thought this would reveal if there was a distinct value at which the majority of sites met the minimum requirement. Table 3.3 shows the percentage of all sites which were classed as either 'Sufficient', 'Good' or 'Excellent' at a range of different percentile values in terms of FS and FC separately as well as the overall classification of the site. As before, the results are also shown for when the largest value from each year was removed from each site, and when the greatest two values were removed.

Percentile	70	75	80	85	*90*	95
FS						
Remove 0	98.4	98.4	90.2	78.7	*72.1*	60.7
Remove 1	98.4	98.4	96.7	90.2	*73.8*	70.5
Remove 2	100	98.4	98.4	95.1	*83.6*	73.8
FC						
Remove 0	100	95.1	88.5	77	*68.9*	57.4
Remove 1	100	98.4	93.4	85.2	*78.7*	67.2
Remove 2	100	100	96.7	93.4	*80.3*	70.5
Overall						
Remove 0	98.4	93.4	85.2	72.1	*67.2*	54.1
Remove 1	98.4	96.7	93.4	83.6	*72.1*	65.6
Remove 2	100	98.4	95.1	91.8	*77*	70.5

Table 3.3. Percentage of compliant sites assessed using different percentile values in 2007

2006 Directive uses 90th percentile to gauge 'Sufficient' compliance

Currently the 90th percentile is used to assess whether or not a site is sufficient although as discussed, only around 67% of sites successfully met this standard in 2007. As can be seen from the above table, if the 80th percentile value is considered there is a notable improvement in the proportion of sites which are sufficient or better and even with no discounting 85% of all sites were compliant with the Directive. In addition, removal of the two largest annual counts further elevated the level of compliance to approximately 95%.

The impact on the percentile calculations, and hence compliance of two potential numerical definitions of the SSL were next considered.

3.2 Site Independent Approaches

3.2.1 Pragmatic Approach

The first definition considered was based on an early pragmatic proposal of using the mandatory standard for FC of 2000 cfu/100ml from the previous 1976 Bathing Water Directive. It was suggested that, because in 30 years there has been no evidence of unacceptable risks to human health using this value as the minimum required standard then 2000 cfu/100ml would be an appropriate discounting limit for FC. In practical terms this discounting limit is the new Directive limit for the 'Sufficient' category multiplied by a factor of 4. Currently 185 cfu/100ml is the sufficient limit for FS and so, under the same rationale, an appropriate SSL for FS would be 4 x 185 = 740/100ml. Since this method is only a pragmatic approach it was thought it would be more convenient to use 750 cfu/100ml as the SSL for FS.

Tables 3.4 and 3.5 contain the levels of compliance achieved both for the signed sites and for all sites when **all** sample values which exceeded these SSLs were removed from the dataset and compliance reassessed. For the signed sites, when this rule was applied only 27.3% met the conditions required to be classified as sufficient, meaning only one additional site met the required standard compared to when no SSL had been applied. Even when looking at all 61 sites these SSLs only marginally improved compliance, increasing the number of sites which were sufficient or better from 67.2% to 70.5%. This was lower than the 77% compliance achieved when discounting the two largest annual values from each location.

Although this non technical definition is straightforward to implement across different sites, there is no scientific justification for using this definition of the SSL and, as shown, it provides limited benefit in terms of improving compliance. In view of this the second approach considered took into account the WHO Guideline Standards in order to define a more effective SSL.

Pragmatic Approach	
Site	**FC SSL = 2000, FS SSL=750**
Portobello Central	Sufficient
Eyemouth	Poor
Sandyhills	Poor
Saltcoats/Ardrossan	Sufficient
Irvine	Poor
Troon	Sufficient
Prestwick	Poor
Ayr	Poor
Brighhouse	Poor
Ettrick Bay	Poor
Aberdeen	Poor
% Sufficient (or better)	27.3

Table 3.4. 2007 Compliance classification with pragmatic SSLs applied (signed sites)

Pragmatic Approach (All Sites)	
	%Sufficient (or better)
FS	75.4
FC	72.1
Overall	70.5

Table 3.5. 2007 Percentage of compliant sites with pragmatic SSLs applied (all sites)

3.2.2 WHO Based Approach

Since Class D of the WHO guidelines (see Table 1.2) represents a level at which "*there may be a significant risk of high levels of minor illness transmission*" (WHO, 2006) it could be reasoned that consequently, the levels comprising class C do not represent this risk. The upper limit for class C is 500 cfu/100ml for FS and so the compliance of the bathing water sites was again assessed with this value used as the single sample limit for this indicator. Within the Directive the 'Sufficient' limit for FS was obtained via the equivalent limit for FC, using a ratio of the two indicators and so from this we can derive the SSL for FC to be (500/185) x 500 = 1350 cfu/100ml.

Similarly to the previous approaches considered, Tables 3.6 and 3.7 show the compliance classifications of the signed sites and the percentages of all sites which were classified as (at least) sufficient when all samples above these SSLs were removed from the dataset. At first sight this rule appears to be slightly more effective than the pragmatic approach for the signed sites; 45.5% of signed sites and 75.4% of all sites were compliant with the 'Sufficient' standard when the WHO based definitions were applied. However, this can be expected since these limits are numerically lower than those used in the pragmatic definition.

If the SSL is lower, there are potentially a greater number of samples that exceed it. This results in a greater number of samples being discounted and so the percentiles used to assess compliance will be calculated using a smaller number of lower counts. Whilst assessing the effects of different SSL definitions on compliance we need to be aware that the 2006 Directive enables us to discount at most 15% of the samples taken within each bathing season. For most sites the number

WHO Based Approach	
Site	**FC SSL = 1350, FS SSL = 500**
Portobello Central	Good
Eyemouth	Poor
Sandyhills	Poor
Saltcoats/Ardrossan	Sufficient
Irvine	Poor
Troon	Sufficient
Prestwick	Sufficient
Ayr	Poor
Brighhouse	Poor
Ettrick Bay	Poor
Aberdeen	Sufficient
% Sufficient (or better)	45.5

Table 3.6. 2007 Compliance classification with WHO based SSLs applied (signed sites)

WHO Based Approach (All Sites)	
	%Sufficient (or better)
FS	82
FC	78.7
Overall	75.4

Table 3.7. 2007 Percentage of compliant sites with WHO based SSLs applied (all sites)

of samples taken during each bathing season is approximately 20 and so the maximum number of samples which can be removed from each site in any given year is 3. For both the pragmatic and the WHO approach considered above, when this 15% restriction was introduced to discounting of samples above the SSLs, there was only one additional site (Nairn Central) where the percentiles were reduced enough to change the classification of the site from 'Poor' to 'Sufficient'.

It is clear that the large variation amongst the sites means that numerical site independent limits such as those considered above will have only limited success. For this reason, a formulaic definition of the SSL and its implications on compliance when applied at individual site level was investigated.

3.3 Adjusting The Geometric Mean

Another proposed method for obtaining a suitable generic SSL involved adjusting the geometric mean to return a specific 90th percentile value and using the subsequent 95th percentile as the single sample limit. There were a number of different values suggested for the standard deviation to be used for calculation of such a limit, including one of 0.8103 which was derived from a large WHO dataset comprising both fresh and marine waters across several different countries. However, it was thought to be more appropriate to use only SEPA data to estimate the standard deviation here. Although the WHO based estimate was considered by SEPA to be relatively large, the average standard deviations of the 61 bathing water sites in Scotland were only slightly lower than this at 0.6754 and 0.7371, for FS and FC respectively.

Rather than using this method to obtain a generic SSL, the procedure was carried out at a site specific level using the estimated standard deviation of the data at each site and setting the 90th percentile to equal 185 for FS, and 500 for FC (the limits for the sufficient categories of each indicator). From this the required adjusted mean was derived and the resulting 95th percentile calculated was applied as a discounting limit. All counts which exceeded this limit were removed and compliance was then re-assessed based on the 90th and 95th percentiles of the reduced data set.

The procedure outlined below was carried out separately for each site.

1. Set the 90th percentile, given by $10^{(\bar{x}+1.281552s)}$ to equal L (L = 185 for FS, L = 500 for FC)

2. Find the adjusted geometric mean, $\bar{x}_{adj}$, at each site from

$$\bar{x}_{adj} = \log_{10}(L) - 1.281552s$$

3. Using this adjusted mean and s as before we can find a site specific discounting limit from

$$D = (\bar{x}_{adj} + 1.644854s)$$

4. Compliance is then assessed using the 90th and 95th percentiles of the data at each site when all counts above D are removed

$$90th\text{percentile}: \quad 10^{(\bar{x}_{<D}+1.281552s_{<D})}$$

$$95th\text{percentile}: \quad 10^{(\bar{x}_{<D}+1.644854s_{<D})}$$

$\bar{x}_{<D}$: mean after all counts exceeding D are removed

$s_{<D}$: standard deviation after all counts exceeding D are removed

This procedure was used only at the 20 sites which were classed as 'Poor' in 2007 since if the site met the Directive's minimum required standard ('Sufficient') prior to any discounting, then the 90th percentile would already be smaller than the sufficient category limit (denoted by L above). Table 3.8 shows the percentage of these 20 sites which were compliant when each of the 92nd, 95th, 97th and 99th percentiles were used as discounting limits and Table 3.9 shows the corresponding percentage of all sites which met the Directive's minimum required standards.

Adjusted Geomean (Poor Sites)				
	92%ile	95%ile	97%ile	99%ile
FS	100	70	40	25
FC	100	100	95	40
Overall % Compliance	100	70	40	25
% Sites with > 15% removed	100	100	90	55

Table 3.8. Percentage of compliant sites in 2007 using adjusted geometric mean approach ('Poor' sites only)

Adjusted Geomean (All Sites)				
	92%ile	95%ile	97%ile	99%ile
FS	100	90.2	80.3	75.4
FC	100	100	98.4	80.3
Overall % Compliance	100	90.2	80.3	72.1

Table 3.9. Percentage of compliant sites in 2007 using adjusted geometric mean approach (all sites)

As can be seen, using the 92nd percentile, and to a lesser extent, the 95th percentile does improve the number of sites which meet the minimum requirement. However, there remains some concern as to the number of counts which were removed. While Table 3.8 shows the percentage of compliant sites when the number of counts removed from the percentile calculations has *not* been restricted, the last row in this table contains the percentage of sites where in at least one of the 4 bathing seasons, for at least one of the two indicators, more than 15% of the data exceeded the relevant SSL. Even for the 99th percentile limits which resulted in only a quarter of the sites meeting the minimum required standard, there were over half of the sites at which the quantity of data removed was greater than that currently permitted under the Directive.

In addition to this, while using the 97th or 99th percentile as a discounting limit does increase the level of compliance amongst the sites compared to when there was no limit applied, similarly to the two site independent approaches considered, this method does not perform as well as simply removing the two highest annual counts. It is also important to note that for nearly all of the sites, in terms of both indicators, the 97th and 99th percentile limits obtained were lower than the WHO guideline limits. If this is taken into account and values which exceed the relevant percentile limit, but still fall below the WHO guideline standards are retained, then using the 97th and 99th percentiles achieves the same results as the WHO based rule explored earlier.

3.4 Summary

While removal of the annual maxima or maximum two annual values from each site does increase the percentage of sites which are compliant this approach to discounting has limited success due to the occurrence of bimodality at some locations. Similarly, both the pragmatic and the WHO based SSL definitions considered above marginally increase the level of compliance amongst the sites however it is clear that the extreme variation between the sites inhibits the effectiveness of these SSLs. This indicated that a site specific approach may be more suitable, however an SSL should be site independent in order to protect human health. Although the geometric mean approach limits were site dependent, they did not achieve results which were markedly better than any of the other limits that have been explored. After application of each of the 92nd, 95th and 97th percentiles limits obtained by adjusting the geometric mean there was an increase in the minimum required standard, however at the majority of sites the percentage of counts which were removed from the data in order to achieve this level of compliance was far greater than that currently allowed under the directive. This was also a problem encountered with the generic limits.

The analysis in Chapters 2 and 3 has highlighted the extensive variation in the data and distributional problems which mean that consequently the impact of the proposed single sample limits on the level of compliance achieved has been limited. In light of this, and after discussion with SEPA, the stated objective of the work changed and the aim became to find a definition of a discounting limit which could be used to identify samples to be removed from compliance calculations. From this point onwards only discounting limits will be considered. While the procedures used for assessing the impact of a discounting limit are the same as that for a single sample limit, the sole purpose of a discounting limit is to identify the samples that could potentially be removed from compliance calculations as opposed to a generic single sample limit, which would be used to assess the quality of a single sample of bathing water as well as determining which samples should be removed. However, it was thought that a discounting limit would continue with the aim of protecting human health since it could only be applied at sites where the public is informed of the water quality.

There were several concerns about the early proposals for single sample limits including the lack of scientific justification for the pragmatic limits and the fact that, with the exception of the geometric mean method, none of the approaches to discounting considered above were based on the observed sample values. Consequently the remainder of the thesis uses statistical models which were fitted using the observed data to define appropriate discounting limits.

Chapter 4

Extreme Value Models

In terms of discounting to increase the number of compliant bathing water sites we are interested in the largest observations of FS and FC as opposed to the mean counts. Given that up to 15% of the data can be removed in each year and there are approximately 20 samples collected at each location during each bathing season, the observations of most interest are the two or three maximum annual values at each site. Similarly to Gaines and Denny (1993) who considered modelling ecological extremes such as sea surface temperatures, it was felt that while modelling mean counts would be irrelevant, a model which focussed on extreme observations would be both appropriate and useful. Extreme value models are used in a wide variety of disciplines including finance, meteorology, ecology and engineering. Specific ecological and environmental applications of these models are given in Coles and Tawn (1996); Robinson and Tawn (1997); Coles and Pericchi (2003); Smith (2004) and Katz *et al.* (2005). It was hoped that fitting such a model would enable a value which is expected to be exceeded a particular number of times during a set period to be found using the observed data. This value could subsequently be used as a discounting limit. Three such methods have been considered; modelling block maxima and K-th largest order statistics using generalized extreme value models and using threshold models which model observations that lie above a specified threshold.

The main references for the theoretical results presented in this chapter have been Coles (2001) and Davison and Smith (1990). The notation used throughout has been taken from Coles (2001).

4.1 Modelling Block Maxima

The first model considered was the the Generalised Extreme Value (GEV) distribution which is used to model block maxima of the form,

$$M_n = \max\{X_1, X_2, ..., X_n\}$$

where $X_1, X_2, ..., X_n$ is a sequence of independent random variables with a common distribution function F. Although the distribution of M_n can be obtained for all values of n as $\{F(z)\}^n$ this is of no practical use since the true function F is unknown. It is possible that the distribution of the data can be estimated, however substituting this estimate into $\{F(z)\}^n$ could prove problematic since small differences between this estimated function and the true distribution function F could potentially lead to very large differences in F^n.

If it is accepted that F^n is unknown, the limiting distribution of F^n as $n \to \infty$ can be used. A suitable re-parametrization is first required since for any $z < z_\alpha$, where z_α is the smallest value z such that $F(z) = 1$,

$$\begin{aligned} F(z) &< F(z_\alpha) = 1 \\ \Rightarrow \{F(z)\}^n &< \{F(z_\alpha)\}^n = 1 \\ \Rightarrow F(z)^n &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

The consequence of this is that the function F^n will always degenerate to a point mass on the upper end point of the distribution, z_α, and hence no meaningful limiting distribution can be found. To overcome this the block maxima are first rescaled in order to stabilize the scale and location parameters as n increases. The "Extremal Types Theorem" (Coles, 2001) states that if there exist sequences of constants $a_n > 0$ and b_n such that

$$Pr\left\{\frac{(M_n - b_n)}{a_n} \leq z\right\} \rightarrow G(z) \quad \text{as } n \rightarrow \infty$$

where G is a non-degenerate distribution function, then G is a member of the GEV family.

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{4.1}$$

defined on $\{z : 1+\xi(z-\mu)/\sigma > 0\}$ where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

Approximating the limiting distribution of the re-scaled block maxima $M_n^* = (M_n - b_n)/a_n$ by a GEV family distribution is equivalent to approximating the distribution of M_n by a different member of the GEV family since for a large enough value of n

$$Pr\left\{\frac{(M_n - b_n)}{a_n} \leq z\right\} \approx G(z) \qquad \text{and therefore equivalently}$$

$$
\begin{aligned}
Pr\{M_n \leq z\} &\approx G\left\{\frac{(z-b_n)}{a_n}\right\} \\
&= G^*(z)
\end{aligned}
$$

(where G^* is another member of the GEV family)

Depending on the value of ξ the shape of the GEV distribution takes one of three possible types. If $\xi = 0$ then the distribution is a light-tailed Gumbel distribution, if $\xi > 0$ it is a heavy tailed Fréchet distribution and if $\xi < 0$ it is a bounded Weibull distribution.

4.1.1 Return Levels

In ecological and meteorological contexts where GEV models are commonly used to model block maxima it is often of interest to look at quantiles of the fitted distribution rather than estimated parameter values. The return level, z_p, which is associated with a return period $1/p$ is the $(1-p)$-th quantile of the GEV distribution. For example, if considering annual maxima, $p = 0.1$ would correspond to a 10 year return level. This is the value which we would expect to be exceeded once every 10 years, or equivalently the value which we would expect to be exceeded in any given year with probability 0.1. Return levels can be found by inverting the distribution function of the GEV (Eqn 4.1) to obtain

$$
z_p = \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}] & : \quad \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\} & : \quad \xi = 0 \end{cases}
$$

In addition to actual values of specific return levels, a return level plot which is obtained by plotting maximum likelihood estimates of z_p against $-\log(1-p)$ is often useful for interpreting a GEV model and assessing model fit.

4.1.2 Diagnostics for Assessing Model Fit

One standard way to assess the fit of a model is to use a probability plot which compares the empirical distribution function to the fitted distribution function. If the model is a good fit to the data then these functions should be approximately equal and when plotting one against the other the points should lie in a straight line along the unit diagonal. Similarly, a quantile plot, in which the ordered observed block maxima are plotted against corresponding empirical estimates can also be used to assess model fit. Any evidence of non-linearity in either the probability or quantile plots suggests the fitted model is not adequate.

As noted previously, return level plots can also be used for model checking by comparing the fitted return level curve and empirical estimates of the return level function and ensuring there are no substantial discrepancies between them. In addition, a histogram of the observed data which shows the probability density curve of the fitted model can also be used. However care has to be taken in the interpretation of this since the level of agreement between the two can differ greatly according to the choice of interval width used for the histogram. Histograms are also of limited use as diagnostic plots when the number of observations used to estimate the model parameters is small.

4.1.3 Fitting a GEV model to Bathing Water Data

In order to fit a GEV model to the bathing water data there is some question as to how to define the blocks for the GEV model. In many contexts where extreme value analysis is used data is available as daily, weekly, or monthly counts taken at a single geographical location over a long period of time. The bathing water data is unusual in that it has been collected over a relatively short period of time at several different locations. However, since it is assumed that the observed counts of FS and FC taken during each year at each site are independent and have a common distribution function it seems reasonable to treat observations taken at a single site during each year as the blocks. This means that there are 224 blocks of approximate size 20 (around 20 observations are taken at each site throughout each bathing season) corresponding to the 61 sites and 4 years. Each M_n is therefore an annual site maximum.

Using the R package *ismev*, GEV models were fitted separately to the $\log_{10}$ transformed FS and FC block maxima. The following maximum likelihood estimates for each of the three parameters, and diagnostic plots (Figures 4.1, 4.2) for the fitted models were obtained;

$$
\begin{aligned}
(\hat{\mu}_{FS}, \hat{\sigma}_{FS}, \hat{\xi}_{FS}) &= (2.17, 0.61, -0.28) \\
(\hat{\mu}_{FC}, \hat{\sigma}_{FC}, \hat{\xi}_{FC}) &= (2.61, 0.71, -0.38)
\end{aligned}
$$

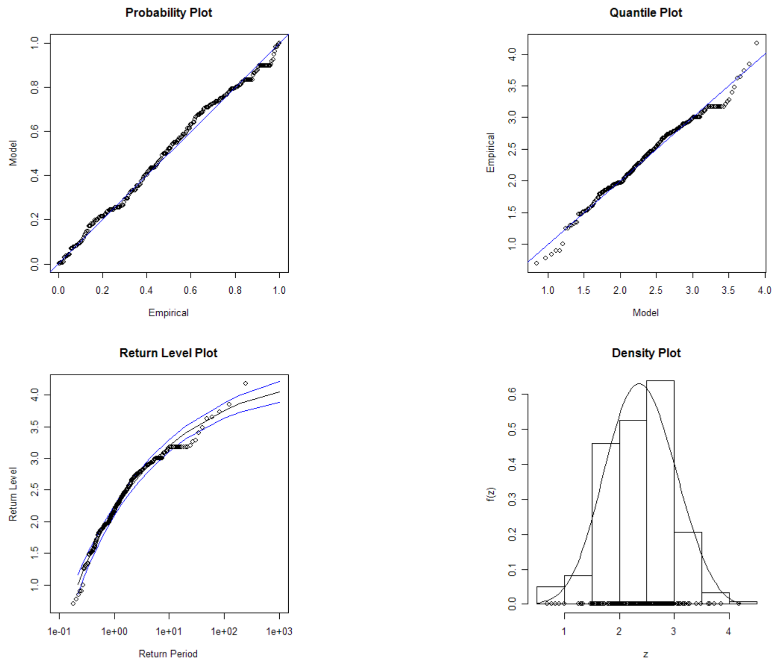

Figure 4.1. Diagnostic plots for fitted GEV model (FS)

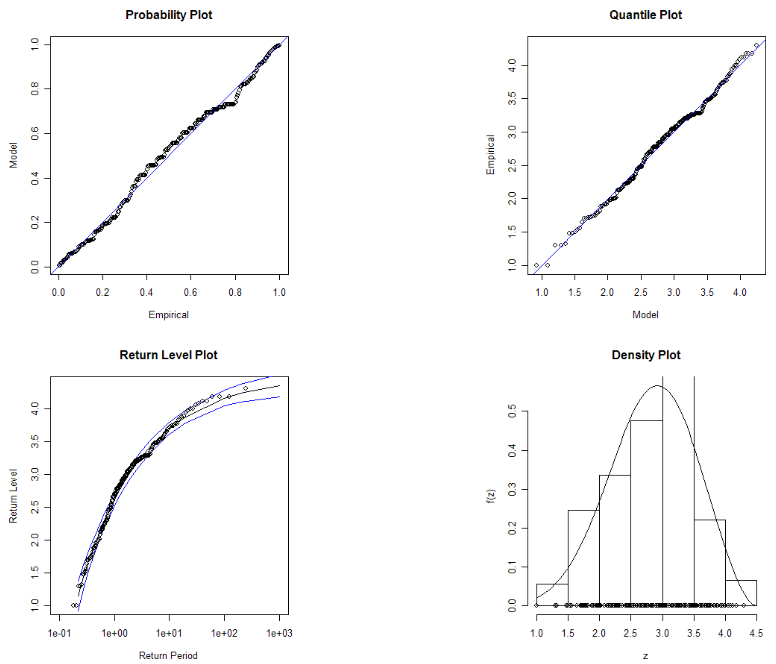

Figure 4.2. Diagnostic plots for fitted GEV model (FC)

For both the indicators, the shape parameters of the fitted models (ξ_{FS} and ξ_{FC}) were found to be significant based on both 95% confidence intervals and on generalized likelihood ratio tests which compared the GEV models to fitted Gumbel models where $\xi = 0$. In view of the diagnostic plots (Figures 4.1 and 4.2), although not ideal, for both indicators all four plots for each model indicate that the model is adequate. There is some evidence of non-linearity towards the tails of the probability and quantile plots but nothing which causes too much concern. Furthermore, in the return level plots, although there is some discrepancy between the fitted and empirical return levels, in general there is substantial agreement between the two. This is particularly true for FC where all empirical return level estimates lie within the plotted 95% confidence bands for the fitted model.

The $\log_{10}$ transformed data were used within these models in order to stabilize the variation in the data. It was also thought that since the Directive calculates the percentiles based on the $\log_{10}$ data and then transforms values back onto the original scale in order for the compliance classification of each site to be determined that a discounting limit should also be found using the transformed data and converted back.

Plots of the profile likelihood (Figure 4.3) were also obtained for the 4 year return levels (on the $\log_{10}$) of each indicator. From these plots maximum likelihood estimates and 95% confidence intervals were found after back transforming to be 649 cfu/100ml (CI: [531, 798]) for FS and 2080 cfu/100ml (CI: [1679, 2570]) for FC. In theory these values should be exceeded at each site in any given year with probability 0.25.

While the definition of return levels for each of the extreme value models that will be considered is based on the assumption that observations are collected over a 12 month period, the situation for the bathing water data is different. For these

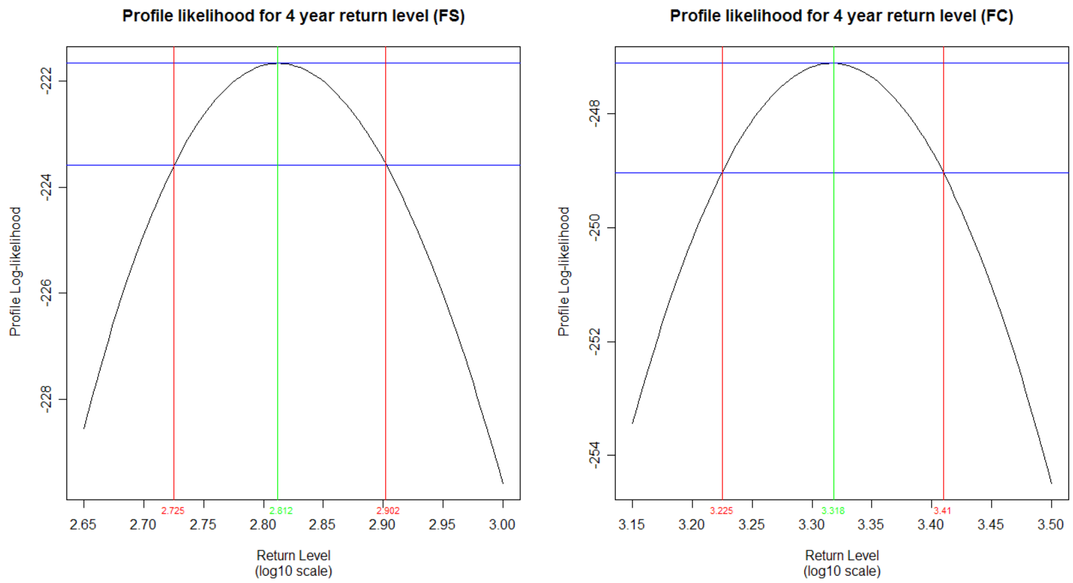

Figure 4.3. Block Maxima model profile likelihood plots for 4 year return levels of FS and FC

data samples are only taken during the bathing season, which in Scotland is approximately a 4 month period that runs from June to September. Consequently, the return periods that have been obtained from each of the models fitted have to be interpreted on a different time scale in order to fit within the context of the problem. For example a 2 year return level is the level which is expected to be exceeded once in every two year period, in this context this is equivalent to a 2 bathing season return level which is the level that we expect to be exceeded once every two bathing seasons.

Using Return Levels as Discounting Limits

Initially the 4 year return level for each indicator was applied as a discounting limit and compliance of each site was subsequently re-assessed by re-calculating the percentiles after all counts which exceeded these values had been removed from the dataset. It should be noted that the number of counts removed was not restricted to 15% at each site in each year and all counts that were greater than the limits were discounted. Table 4.1 shows the results which were obtained for all 61 sites.

All Sites	
	%Sufficient (or better)
FS	77.0
FC	72.1
Overall	70.5

Table 4.1. 2007 Percentage of compliant sites using 4 yr GEV return level discounting limits

As can be seen, using these limits achieved little in terms of improving compliance. There were only two additional sites which meet the minimum required standard compared to when there is no discounting (see Table 3.2). In fact, using these limits achieved the same level of compliance as the pragmatic rule (see Table 3.5) and was worse than both the WHO rule (see Table 3.7) and simply removing the largest/largest 2 annual maximum values (see Table 3.2). An explanation as to why compliance has not really changed from when there is no discounting can be found in the relatively large size of the 4 year return levels; for both indicators they were greater than the corresponding WHO limits, and for FC the value obtained was also greater than the suggested pragmatic limit (see Table 4.2).

This failure to improve compliance could be anticipated due to both the large differences between the annual maxima at each site and because, by their definition, 4 year return values are only expected to be exceeded at each site once every four years. For the data available, this means at each site we would assume only one observation should lie above this value and whilst at the majority of sites these levels were never reached within the four year period for which data are available, there were several others at which, with respect to FS in particular, the levels were exceeded on several occasions. This is again a clear indication of the extensive variation between the sites. Due to the large size of these return levels, in contrast to previous rules which have removed too many counts, using these values as discounting limits has removed too few and consequently there has been very little impact on the percentiles used to assess compliance.

Assuming at each site there are 20 or more observations taken throughout the bathing season, we would ideally like to obtain a return level which is exceeded three times per year however this would mean specifying a return period of 4 months. This return period cannot be obtained using the GEV model since choosing $p = 1/3$ corresponds to a probability of exceedance in any given year (bathing season) which is greater than 1. As an alternative to the 4 year return levels for each indicator, 2 year and 16 month return levels were also considered since both of these shorter time periods correspond to a greater probability of excess in any given year. The return level estimates obtained from the GEV models for each indicator as well as the pragmatic and WHO limits previously considered are shown in Table 4.2.

	16 month GEV RL	**2yr GEV RL**	**4yr GEV RL**	**Pragmatic**	**WHO**
FS	95	250	649	750	500
FC	234	719	2080	2000	1350

Table 4.2. Table of discounting limits

As can be seen, the 16 month discounting limits are very low and, for both indicators, are less than the Directive's limit for the 'Excellent' classification category. Again, this is believed to be the result of extensive differences between the maximum values at different sites. The consequence of these values being so low is that while using the 16 month level values as discounting limits achieves 100% compliance, in order to do this over 50% of the observations have to be removed from some sites. Application of the 2 year return levels as discounting limits similarly results in a substantial increase in the overall level of compliance from 67.2% with no discounting, to 91.8%. However, at some sites up to 30% of counts exceeded this value for the return level which is only expected to be exceeded twice in the period for which data is available.

Table 4.3 contains the percentage of all 61 bathing water sites which were at least 'Sufficient' after applying of each of the 16 month, 2 year and 4 year return levels obtained from the fitted block maxima models as discounting limits. The table also contains the percentage of sites where there was at least one bathing season and at least one indicator where more than 15% of the data were discounted.

Return Period	**Overall % Compliance**	**% Sites with > 15% removed**
16 months	100	73.8
2 years	91.8	42.6
4 years	70.5	14.8

Table 4.3. Percentage compliance achieved using GEV model return levels and corresponding percentage of sites where more than 15% of data are removed

One of the main limitations of the block maxima model is that using only one value from each site in each year wastes a lot of valuable information, particularly at sites where it is not only the annual maximum value which could be defined as extreme, but also the second and third largest etc. For this reason, k-th largest order statistic models and threshold models were also considered.

4.2 K-th Largest Order Statistic Models

The second model is a generalization of the block maxima model that focusses on the behaviour of the k-th largest order statistic within each block. This model would seem to be a more appropriate choice in the context of the bathing waters problem since the quantity of interest (assuming there are exactly 20 observations taken at each site during each year) is the level above which the three largest observations in each block lie.

If $X_1, X_2, ..., X_n$ are independent random variables with a common distribution function $F(z)$ then the variable of interest is

$$M_n^{(k)} = \text{k-th largest of}\{X_1, X_2, ..., X_n\}$$

Again the limiting distribution of this variable for a small fixed value of k as $n \to \infty$ is of interest. Again, as with the block maxima model, a suitable re-normalization of $M_n^{(k)}$ is first required in order to prevent degeneration of this distribution to a point mass at the smallest value z such that $F(z) = 1$. If there exist suitable sequences of constants such that a GEV distribution can be found for the block maxima, M_n (Eqn 4.1) then after application of the same re-scaling to $M_n^{(k)}$,

$$Pr\left\{\frac{(M_n^{(k)} - b_n)}{a_n} \leq z\right\} \to G_k(z) \quad \text{as } n \to \infty$$

where $G_k(z)$ is of the form.

$$\begin{aligned} G_k(z) &= \exp\{-\tau(z)\} \sum_{s=0}^{k-1} \frac{\tau(z)^s}{s!} \\ \text{with} \qquad \tau(z) &= \left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} \end{aligned}$$

defined on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$.

$G_k(z)$ is also within the GEV distribution family and uses the same location, shape and scale parameters as the corresponding GEV distribution for block maxima. However, additional observed extreme data is incorporated into this model and hence the return level estimates produced should be more accurate. As with the block maxima model, return level estimates can be obtained by inverting the distribution function, $G_k(z)$, and the same diagnostic plots can be used to assess model fit.

4.2.1 Fitting a K-th Largest Order Statistic Model

If blocks are defined in the same way as for the block maxima model, then the most relevant order statistic to look at is the third largest order statistic of each block i.e. the third largest annual value at each site. For both indicators a 3rd largest order statistic model was fitted in R to the $\log_{10}$ transformed data. In view of the diagnostic plots for the fitted models (Figures 4.4, 4.5) it was apparent that for FS and FC there is a clear lack of agreement between the empirical and fitted model estimates and therefore any return levels which were obtained using this approach would be invalid.

One possible explanation for this clear lack of fit was again the extensive variation amongst the sites. For example, while in 2004 the maximum FS count at the Dornoch site was 10 cfu/100ml, the third largest value at the Stonehaven site in the same year was over 1400 cfu/100ml. For this reason, a second model was then fitted to only the site/year blocks where the maximum value was greater than or equal to the Directive's limit for sufficient; the rationale behind this being that values below this limit could not really be considered as extreme. Despite this restriction, there was no improvement in the fit of the model and although, in theory, this type of model uses the available data more efficiently than the block maxima model, there continued to be some concern that the blocking structure of the GEV model did not fully take into account the heterogeneity between the sites.

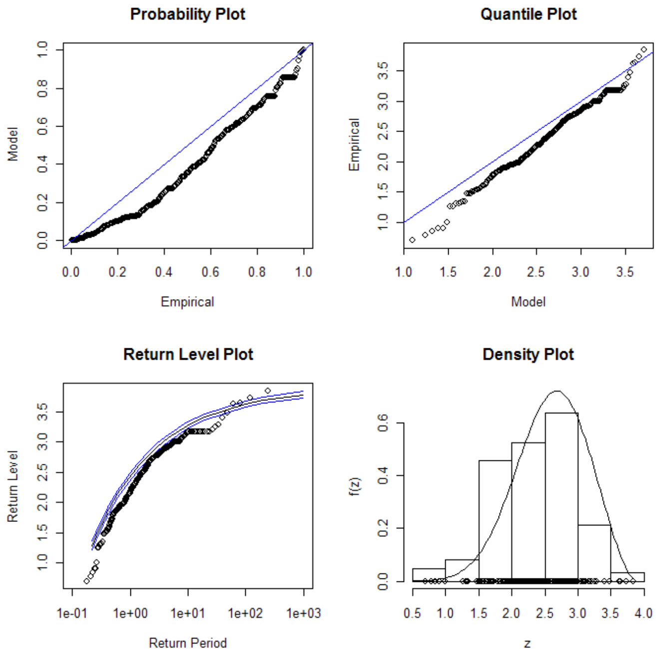

Figure 4.4. Diagnostic plots for K-th largest order statistic model: $\log_{10}(FS)$

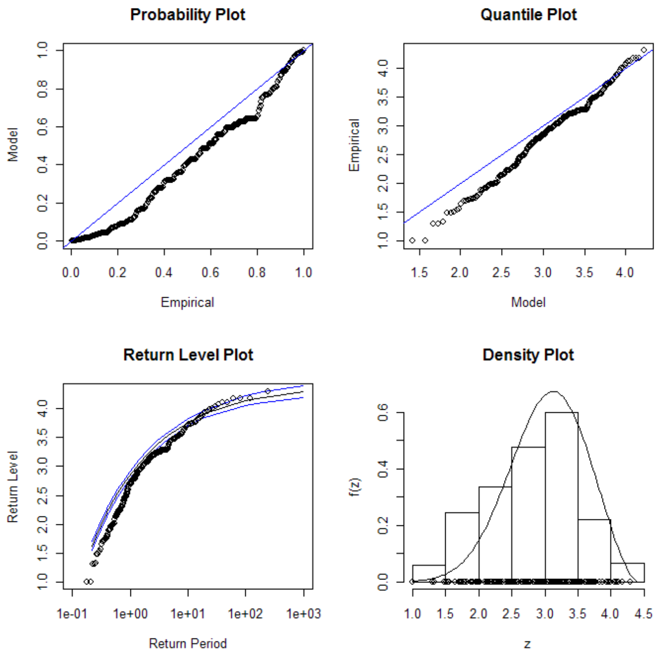

Figure 4.5. Diagnostic plots for K-th largest order statistic model: $\log_{10}(FC)$

4.3 Threshold Models

As has been shown, the first two extreme value models considered did not allow for the fact that some sites have many more large counts than others and consequently did not make use of a substantial amount of the data available. It was thought that threshold models overcome this problem since they are not based on only a single statistic from each block, but instead model all observations which exceed some pre-specified threshold.

If $X_1, X_2, ..., X_n$ is a sequence of independent random variables with a common distribution function, F, then all observations which exceed some large threshold, u can be defined as extreme events. The probability of threshold excess by y units for any arbitrary choice of X in the sequence X_i can then be defined as

$$Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)} \quad \text{where} \quad y > 0$$

While this cannot be used in practice since the true distribution, F, is unknown, if the limiting distribution of the block maxima can be approximated by a GEV distribution (Eqn 4.1) then the distribution of all threshold excesses, $(X - u)$ conditional on $X > u$, can be approximated by a Generalized Pareto Distribution (GPD). This has a distribution function of the form

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$

and is defined on $\{y : y > 0, (1 + \xi y/\tilde{\sigma}) > 0\}$. The shape parameter used within this distribution function is the same as that of the corresponding block maxima model and the scale parameter, $\tilde{\sigma}$, is also a function of the corresponding GEV parameters given by

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

In order to assess the suitability of a fitted model the same diagnostic plots that are available for the GEV models can be used; quantile plots, probability plots, return level plots and a histogram of the observed data including the probability density curve of the fitted model.

4.3.1 Return Levels for the Threshold Model

As with other extreme value models it is more informative to look at estimates of return levels rather than parameters when using this type of model. However because threshold models use all data exceeding some pre-specified threshold, the construction of N year return levels is different to that of the GEV family models which focus on one annual statistic from each block.

If there exists a suitable generalized Pareto distribution for observations which exceed some suitable threshold u with parameters ξ and $\tilde{\sigma}$ as before, then for $x > u$

$$Pr\{X > x | X > u\} = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} \tag{4.2}$$

If $\zeta_u = Pr(X > u)$ i.e. the probability that an observation exceeds the threshold u, then (Eqn 4.2) can also be written as

$$Pr\{X > x\} = \zeta_u\left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} \tag{4.3}$$

The m-observation return level x_m which is defined as the level expected to be exceeded on average by one observation in every m can then found as the solution of

$$\zeta_u \left[1 + \xi \left(\frac{x_m - u}{\xi} \right) \right]^{-1/\xi} = \frac{1}{m} \tag{4.4}$$

Rearranging equation 4.4

$$x_m = \begin{cases} u + \frac{\sigma}{\xi}[(m\zeta_u)^{\xi} - 1] & : \quad \xi \neq 0 \\ u + \sigma \log(m\zeta_u) & : \quad \xi = 0 \end{cases}$$

In practise, if k is the number of excesses and n is the total number of observations then k is thought to follow a $Bin(n, \zeta_u)$ distribution and so the maximum likelihood estimate of ζ_u is then $\hat{\zeta_u} = \frac{k}{n}$.

Since the return levels are defined as the level expected to be exceeded by m observations, for threshold models, it is possible to specify return periods of shorter than one year by substituting m in equation 4.3.1 as a fraction of the number of observations per year. So, for example, if there are 20 observations per year then a 6 month return level is the level expected to be exceeded once in every 10 observations. The N year return level, which is expected to be exceeded once during each N year period, can also be obtained by using $m = N \times n_y$, where n_y is the number of observations per year.

Return level plots which use the annual return levels can also be obtained and can be used both to graphically display fitted return level estimates and to assess the fit of the model.

4.3.2 Threshold Choice

One practical problem when fitting a threshold model is the question as to what choice of threshold is appropriate. While choosing a threshold which is too large will result in too few excesses on which to estimate the model, too small a threshold will violate the asymptotic properties on which the model is based. There is currently no automatic procedure to find a suitable threshold value although some exploratory diagnostics are available. The first is a mean residual life plot which plots the threshold, u, against the sample mean excess of the observations which exceed u. This was first suggested as an appropriate method of choosing a threshold by Davison and Smith (1990) and was also used by Coles (2001).

If the GPD model is suitable then we would expect a linear relationship between the threshold and the mean level of excess so within this plot we are looking for linearity after confidence bands are taken into account. Suitable thresholds will then correspond to the values within this area of linearity. Furthermore, after the model has been fitted, a sensitivity analysis can be used to assess the effect of choosing different threshold values on the estimated model parameters. Maximum likelihood estimates with confidence intervals for both the modified shape and scale parameters can be plotted against a range of different thresholds and can then be used to select appropriate threshold values by checking for which values there appears to be stability in the estimates.

4.3.3 Fitting a Threshold Model to the Bathing Waters Data

The directive permits that a maximum of 15% of counts to be discounted each year, therefore an initial step in determining an appropriate threshold was to consider the empirical 85th percentile for each indicator, which by definition is the value exceeded by 15% of the observed data. For FS and FC these were calculated to be 93 cfu/100ml and 260 cfu/100ml respectively, both of which are relatively small in terms of the limits of the Directive's classification categories. Mean residual life plots (Figure 4.6) for FS and FC indicated that the range of suitable threshold values was between around 90 cfu/100ml and 1000 cfu/100ml for FS, and between 160 cfu/100ml and 1100 cfu/100ml for FC.

Although the interpretation and subsequent choice of threshold using these plots is somewhat subjective, Coles (2001) suggests that the threshold chosen should be as low as possible subject to satisfying the relevant diagnostic plots. It was thought that observations which only just fell above the 85th percentile or the lower end of the scale of possible threshold values indicated by the mean residual life plots, could not truly be classed as extreme and for this reason, the threshold was initially set to be the limit of the Directive's 'Sufficient' category. It was believed that this value was both at a level above which observations could be considered large while still generating a sufficient number of excesses on which to estimate the model.

The range of potential threshold values indicated by the mean residual life plots was also used to assess the sensitivity of the stability of the estimated model parameters as the threshold value was altered. Figure 4.7 shows the modified

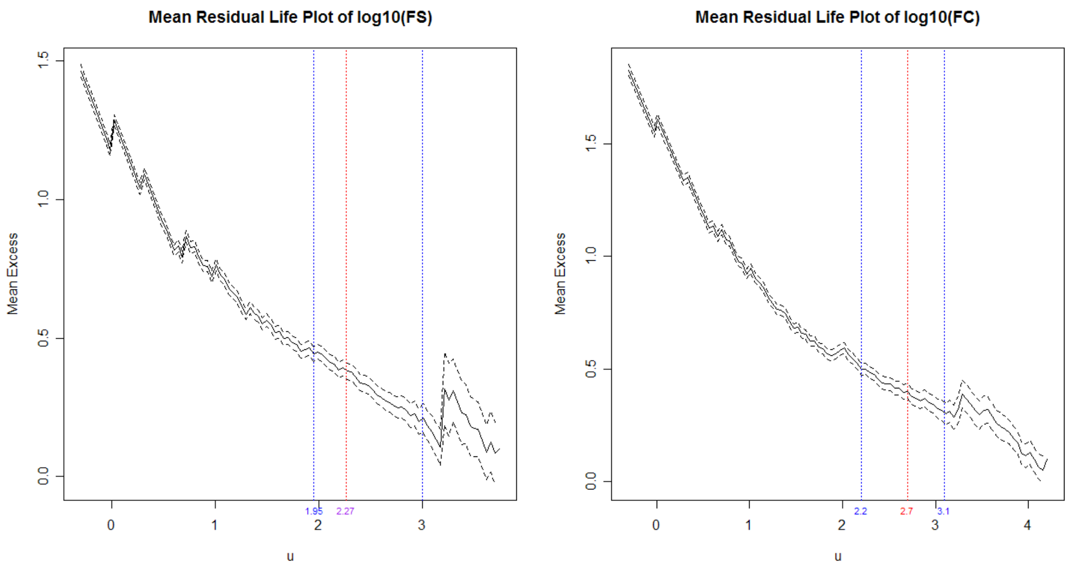

Figure 4.6. Mean residual life plots of $\log_{10}$ transformed data

The blue lines indicate an area of linearity corresponding to a range of suitable threshold values. The red lines indicate the Directive's 'Sufficient' category limits.

scale and shape parameter estimates for fitted models using the different thresholds. Although the interpretation of this is again fairly subjective, the choice of threshold for each of the indicators (shown on Figure 4.7 by the red lines) does not appear to be inappropriate.

Using the R package *ismev*, threshold models were fitted separately for each of the indicator variables using the sufficient category limits as thresholds. For FS this meant the threshold was 185 cfu/100ml which generated 382 excesses (8.6% of the total data) and for FC, the threshold was 500 cfu/100ml, resulting in 385 excesses (8.7% of the total data). Diagnostic plots (Figures 4.8 and 4.9) suggest that for both indicators the models were a reasonable fit for the data although there was some concern regarding divergence between the fitted and observed points towards the upper tail in the FS model.

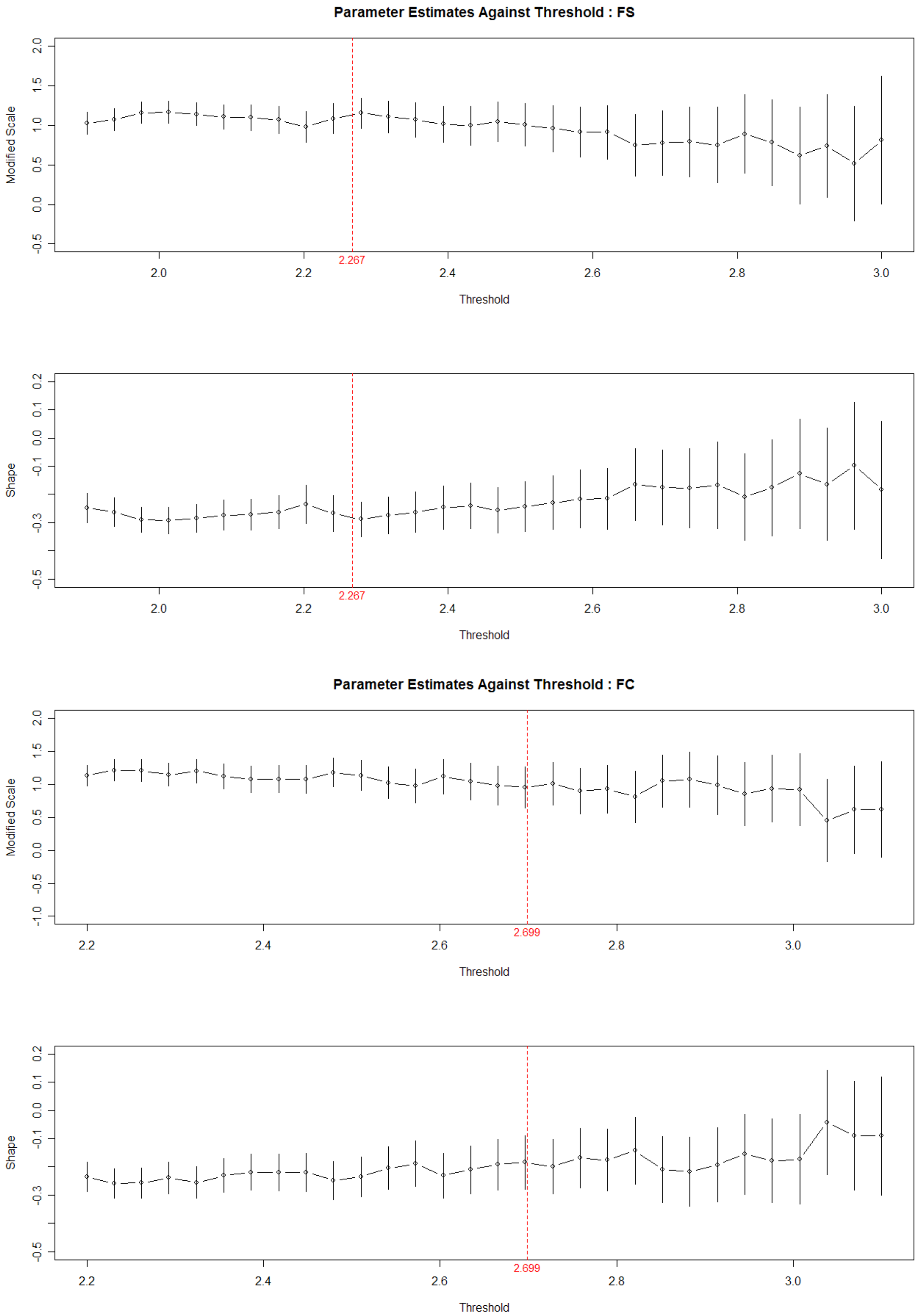

Figure 4.7. Plots of parameter estimates against threshold

The red lines indicate the Directive's 'Sufficient' category limits.

In order to estimate the GEV distribution whose parameters correspond to the GPD distribution being fitted and to calculate return levels, it is necessary to specify the number of observations in each year. In the models for the bathing water data this was set at 20 and although at the majority of sites there are 20 counts taken in each bathing season, the number of data points at each location ranges from 5 to 40. Subsequently, further models were fitted which were restricted to only include data from site/year combinations where there were exactly 20 counts and then to only including sites/years where there was between 18 and 22 counts. It was found that this restriction achieved very little in terms of improving the fit of the model.

Although the amount of data used was reduced in the restricted models, for all three models the proportion of threshold exceedances that is used in the calculation of return levels were very similar. As with the block maxima model the 16 month, 2 year and 4 year return levels were obtained and for each of the three models fitted for FS and FC the levels obtained are shown in Table 4.4 along with 4 month return levels which were also obtained for each of the threshold models.

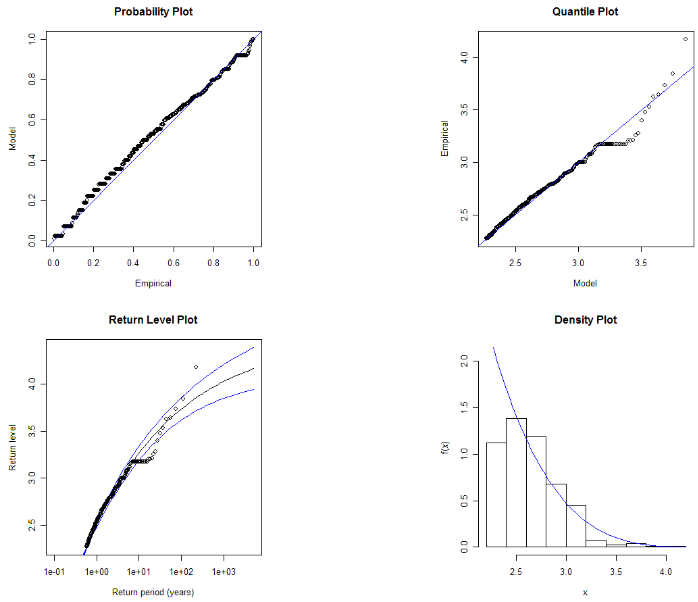

Figure 4.8. Diagnostic plots for fitted threshold model ($\log_{10}$(FS))

threshold = $\log_{10}(185)$

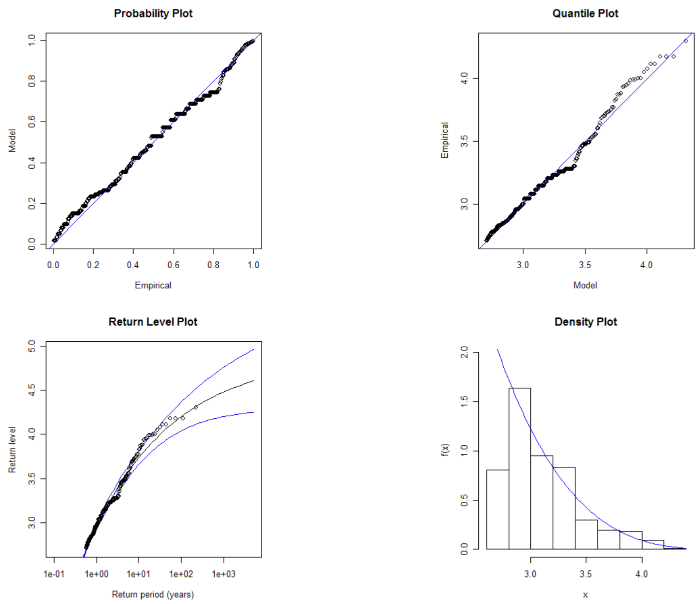

Figure 4.9. Diagnostic plots for fitted threshold model ($\log_{10}$(FC))

threshold = $\log_{10}(500)$

		Return Level (cfu/ 100ml)			
		4 month	16 month	2 Year	4 Year
All Obs. Model	FS	94	426	601	995
	FC	257	1191	1715	2977
18 - 22 Obs. Model	FS	99	445	625	1031
	FC	273	1241	1781	3077
20 Obs. Model	FS	94	402	567	968
	FC	239	1064	1516	2586

Table 4.4. Return levels for threshold models (cfu /100ml)

As can be seen there is a substantial difference between the 4 year, 2 year and 16 month return levels found using threshold models and those obtained from the block maxima model (Table 4.2). In fact the 16 month limits for the GEV model are very similar to the 4 month return levels obtained from the threshold model. One possible explanation for this is that the GEV model includes the maximum value from all sites, regardless of whether the maximum is very low in relative terms. In contrast, the threshold model only includes observations which have been defined as extreme so sites which already meet the minimum required standard of compliance are unlikely to contribute to the data that are used to estimate the model.

For all three of the threshold models which were fitted, application of each of the return levels as discounting limits achieved almost the same levels of compliance. Therefore, for simplicity, it was decided to use the limits obtained using the models which were estimated using observations from all sites. The results obtained when each of these discounting limits were applied are shown in Table 4.5. The third column in this table shows the percentage of sites where in at least one bathing season, more than 15% of the observations for either FS or FC exceeded the relevant discounting limits. Table 4.6 contains the classification of each signed site after application of the 4 month return level limits and the percentile values at each location before and after discounting. The percentage of counts removed at each site in each year, as well as the average percentage removed across the 4 years are shown for each site and each indicator.

Return Period	**Overall % of Compliance**	**% Sites with > 15% removed**
4 months	100	72.1
16 months	78.7	27.9
2 years	73.8	18.0
4 years	70.5	6.6

Table 4.5. Percentage compliance achieved in 2007 using threshold model return levels as discounting limits

Return levels are obtained from threshold models based on all observations.

All of threshold models return level limits result in some increase in the level of compliance compared to when there is no discounting (see Table 3.2). Although the 4 month levels achieved 100% compliance, far too much of the data were removed when using these limits. For example, it is expected that 3 counts (15%) of the data in each year exceed the 4 month return level but it can be seen from Table 4.6 that there are several sites at which more than 50% of the data lies above this value. At the Sandyhills site in 2007, 78% of the FC values are removed from the percentile calculation after application of the 4 month discounting limit. It was thought that further developing the idea of threshold models by application at individual site level would improve both the accuracy of the return level limits obtained and hence the consistency of the percentage of counts removed in each year.

2007 Overall Compliance **All Obs. Threshold Model** Signed Sites								
Site	**No Discounting** (90 %ile)	**4 Month RL** (90 %ile)	**% Removed**					
				2004	2005	2006	2007	Avg
Portobello Central	Sufficient 134(FS), 380(FC)	Good 53(FS), 145(FC)	FS	0	5	20	26	12.8
			FC	0	5	10	26.3	10.3
Eyemouth	Poor 443(FS), 1374(FC)	Good 58(FS), 200(FC)	FS	20	30	10	52	28
			FC	15	20	25	47	26.8
Sandyhills	Poor 403(FS), 1256(FC)	Good 101(FS), 240(FC)	FS	63	45	10	47	41.3
			FC	57	50	35	78	55
Saltcoats/ Ardrossan	Poor 139(FS), 513(FC)	Good 67(FS), 177(FC)	FS	25	10	10	21	16.5
			FC	15	15	15	31	19
Irvine	Poor 298(FS), 1298(FC)	Excellent 61(FS), 140(FC)	FS	40	30	15	21	26.5
			FC	35	35	30	57	36
Troon	Sufficient 108(FS), 307(FC)	Excellent 44(FS), 111(FC)	FS	40	15	5	0	15
			FC	30	10	10	5	13.8
Prestwick	Poor 267(FS), 847(FC)	Good 64(FS), 191(FC)	FS	35	27	15	15	23
			FC	20	18	20	32	23
Ayr	Poor 403(FS), 1725(FC)	Good 82(FS), 247(FC)	FS	50	25	40	36	37.8
			FC	50	45	50	57	51
Brighouse Bay	Poor 335(FS), 868(FC)	Good 64(FS), 215(FC)	FS	55	23	20	21	29.8
			FC	55	28	20	26	32.3
Ettrick Bay	Poor 333(FS), 1725(FC)	Good 133(FS), 831(FC)	FS	40	40	15	21	29.8
			FC	50	40	35	42	41.8
Aberdeen	Poor 349(FS), 868(FC)	Good 74(FS), 154(FC)	FS	45	25	10	21	25.3
			FC	25	15	10	42	23

Table 4.6. Compliance classification of signed sites in 2007 using threshold model 4 month return levels as discounting limits.

Return levels are obtained from threshold models based on all observations.

4.4 Independence

In order to ensure that parameter estimates are accurate one of the assumptions of all of the extreme value models considered is that the observations are independent. For the GEV model it is clear that this assumption holds since only one observation from each site during each year is included in the model. Although for the threshold model the assumption of independent observations is not automatic since all exceedances of the threshold are considered, regardless of the date when the sample was taken, there are several reasons why this would seem reasonable. It could be argued that if the threshold is set at a level high enough such that observations which lie above it can truly be considered as extreme events, then exceedances will occur at times points which are far apart and are subsequently independent of one another. In spite of this, in some situations where extreme value analysis is used it is known that clustering of several threshold threshold excesses in small intervals of time can occur (Gaines and Denny, 1993; Katz *et al.*, 2005). This does not appear to be a problem for the bathing waters data. Often clustering of extreme events happens when samples are taken close to each other in time, such as hourly or daily, however bathing water samples are taken from each site at a rate of (approximately) one per week and therefore there should be no possibility of one sample having any influence on the other. For example, it is very unlikely that if a sample was taken on a day where there was heavy rainfall that this would then affect the next sample which is taken a week later. Abnormal weather waivers for removal of samples affected by extremely poor weather are also likely to prevent this.

It is important to look for dependence at individual site level since looking at the data from all sites together could be misleading. Combining all sites will result in several observations which although may have been collected on the same day, have come from different sites which are potentially hundreds of miles apart and so would be independent, but may appear very close together. For each site, a plot of the absolute difference between consecutive observations in each bathing season against the corresponding time difference between these observations was produced in order to see if there was any evidence of dependence within the samples. A positive linear relationship in this plot would indicate that the samples which are taken close together in time are also close together in magnitude.

Figures 4.10 and 4.11 show an example of these plots, each for a different indicator at two different sites. Since there is no clear evidence of a linear relationship, both plots suggest that dependence amongst the observations does not appear to be a problem at these sites. This was typical of the situation at each of the sites and so there was no indication of consistent dependence in the samples for either FS or FC which would cause concern.

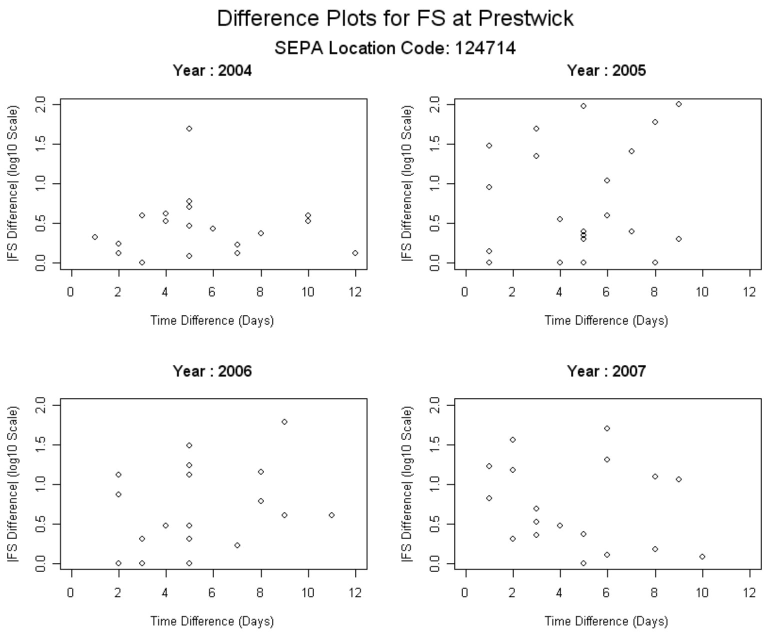

Figure 4.10. Plots of absolute difference in consecutive FS observations against corresponding time difference between samples at Prestwick Site

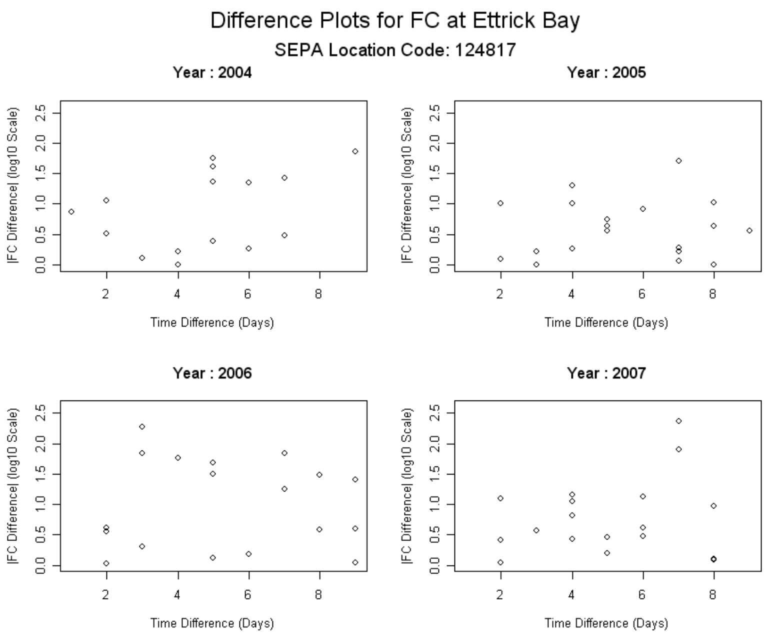

Figure 4.11. Plots of absolute difference in consecutive FC observations against corresponding time difference between samples at Ettrick Bay Site

4.5 Including a Covariate

All of the extreme value models that have been considered so far assume a distribution which is constant through time and until now the possibility that the behaviour and occurrence of extreme observations may be dependent on a covariate has not been explored. There are several reasons why this assumption of stationarity does not hold in the situations where extreme value analysis is most suited; for example, changes in weather conditions can often affect the distribution of observations from season to season, or, as with the bathing water data, on a much shorter term basis. It is already well established that the level of each microbiological indicator present varies according to meteorological conditions, this is the reason why abnormal weather waivers are sometimes issued for samples which are then completely removed from the dataset. Increased rainfall is likely to increase the level of diffuse pollution in the water. For this reason it was thought that incorporating extra information into the fitted extreme value models, through a measure of salinity, may produce more accurate models and therefore more appropriate return level estimates.

Salinity is a measure of the quantity of dissolved salt in the water and therefore can give an indication of the level of rain water present in each sample. Since low salinity values are likely to signify that samples have been diluted by rainfall it could be expected that extreme observations will be associated with samples that have a smaller salt content. The effect of salinity on the microbiological indicators within the context of the bathing water data was next considered.

4.5.1 The effect of salinity on extreme observations

Salinity data is available for 9 of the 11 signed sites in the form of one value, measured in parts per thousand, for each bathing water sample taken. Due to the limited quantity of data available at individual site level exploratory plots were initially produced for data from all 9 sites in order to examine the relationship between salinity and each microbiological indicator.

As well as working with the $\log_{10}$ transformed FS and FC values, the salinity variable was also transformed before any relationships were explored. As can be seen from Figure 4.12, on the original scale the distribution of salinity was highly negatively skewed. In order to overcome this and increase the symmetry in the distribution, values were first subtracted from a value of 38 to invert the direction of the skewness and subsequently the log transform was applied. A similar transformation for salinity was used by Satpute (2005). It is important to note that this transformation will reverse the direction of the expected relationship between salinity and the microbial indicators. Smaller values of salinity on the original scale will correspond to larger values for the transformed variable.

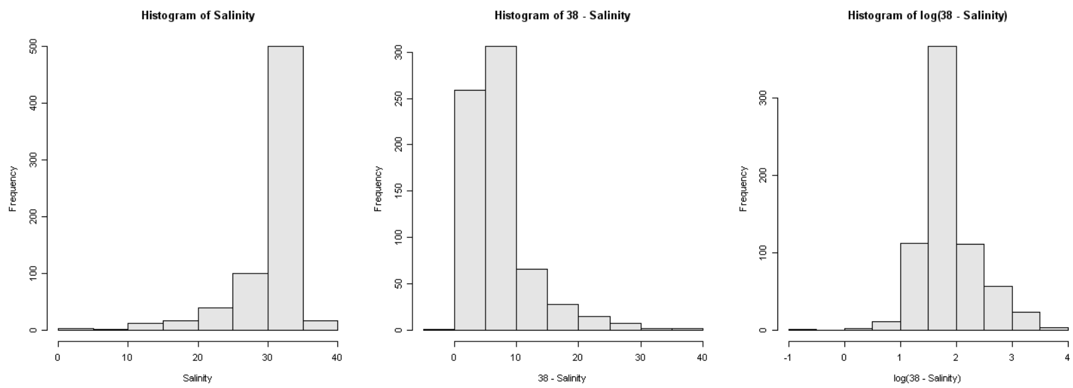

Figure 4.12. Histograms of salinity and transformed salinity at signed bathing water sites

Scatterplots (Figure 4.13) were first produced of block maxima (annual site maxima) against the corresponding annual average salinity value for each of FS and FC. Reference lines relating to the sufficient category limits of each indicator are shown in red.

Figure 4.13. Plots of annual site maxima against average annual salinity

Reference lines indicate the Directive's 'Sufficient' category limits.
FS: $\log_{10}(185)$, FC: $\log_{10}(500)$

For both FS and FC there is no strong evidence of any relationship between the two variables. It could be argued that for both indicators there is some indication of a weak positive relationship between transformed salinity and annual site maxima however due to the large variation in the block maxima this is not immediately obvious. To further examine the effects of salinity on occurrence of extreme observations plots were produced for each indicator of transformed salinity values against whether or not the observation exceeded the threshold determined by the limit of the Directive's sufficient category (Figure 4.14). This effectively treats FS and FC as binary rather than continuous variables in order to see if there is any relationship between presence of threshold excess and salinity.

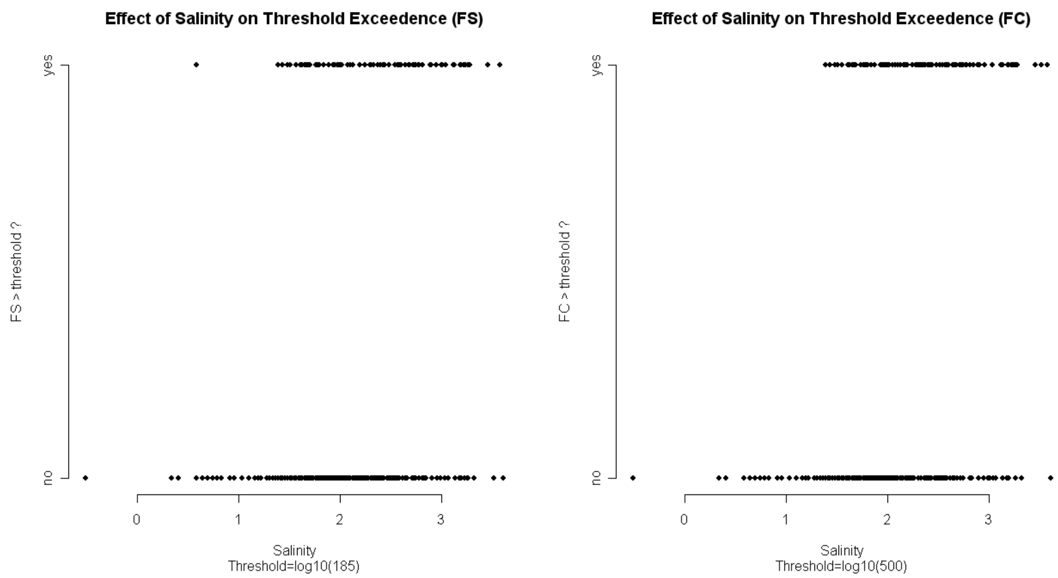

Figure 4.14. Plots of occurrence of threshold excess against transformed salinity

Again for both indicators there is very little evidence of a relationship between occurrence of threshold excess and salinity. Although threshold excesses are in general associated with greater values of the transformed salinity variable as could be expected, there is no distinction between the salinity of samples which were extreme events (threshold excesses) and those which were not. The data were next examined to determine whether there was any suggestion of a link between salinity and the magnitude of threshold excess'. Figure 4.15 shows scatterplots of the size of threshold excess against salinity. Similarly to the scatterplots of block maxima (Figure 4.13) there is almost no evidence of a relationship between salinity and the size of threshold excess. As before, for FC in particular, there is some indication of a very weak positive correlation although this is difficult to identify due to the large variation in the data.

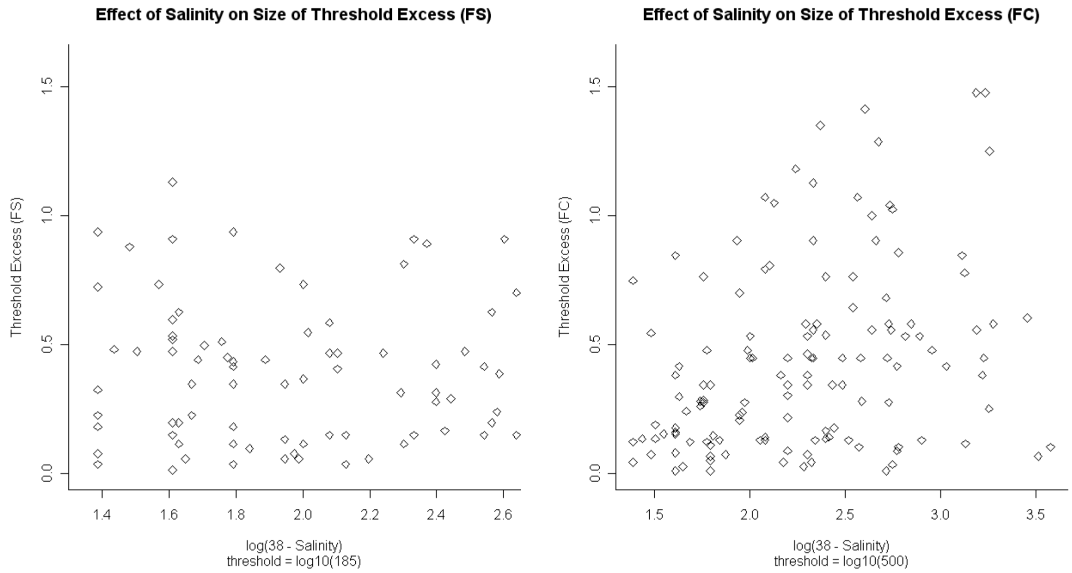

Figure 4.15. Plots of magnitude of threshold excess against transformed salinity

4.5.2 Including Salinity in Block Maxima Models

Although threshold models have been shown to be most appropriate for the bathing waters data as a result of the extensive variability between the sites and years, there is a limited quantity of salinity data available on which to estimate this type of model and so salinity was incorporated into a block maxima model which used information from all of the 9 sites. This suggestion seems reasonable as the previous block maxima models provided an acceptable fit for the data. The problem was with the return level estimates obtained from these models which did not accurately reflect the levels observed at some of the bathing water sites. Assessing the significance of salinity within block maxima models will therefore provide some indication of whether or not the additional information gained by inclusion of a relevant covariate would be worthwhile in terms of explaining more of the variation in the data.

If blocks are defined in the same way as for the previous GEV models considered (as observations taken from a single bathing season at each site) and it is assumed that the block maxima are from a generalized extreme value distribution with parameters μ, σ, and ξ then the following model can be used for X_t, the annual maximum value in year t,

$$\begin{aligned} X_t &\sim GEV(\mu(t), \sigma, \xi) \\ \text{where} \quad \mu(t) &= \beta_0 + (\beta_1 \times \mathrm{Sl(t)}) \end{aligned}$$

and where Sl(t) denotes a salinity statistic corresponding to year t.

Block maxima models were subsequently fitted to the annual site maxima from the signed sites, resulting in 36 observations on which to estimate the model parameters (9 sites x 4 years). For each of FS and FC, models were fitted which had a constant location parameter as well as two models which each incorporated a different salinity statistic; the mean salinity value in each block and the minimum salinity value in each block (which corresponds to the maximum value on the transformed scale).

For each of these three models, a summary of the parameter estimates and maximised log likelihood are shown (for each indicator) in Table 4.7.

It is apparent that for both FS and FC, the estimated coefficients of the two salinity statistics (mean and annual min) are not significant. The deviance statistics for the models which include average salinity in each year are 2(20.1-19.7) = 2(17.2-16.8) = 0.8 which is much smaller than the appropriate $\chi^2_{(1)}$ statistic of 3.84 and so indicates that the simpler model of the two is sufficient. Similarly for the annual minimum salinity models, the deviance statistics are not significant (0.4 and 1.2 for FS and FC respectively). In fact, for this model the

	Model	Log-likelihood	$\hat{\beta}$ (s.e. $\hat{\beta}$)	$\hat{\sigma}$ (s.e. $\hat{\sigma}$)	$\hat{\xi}$ (s.e. $\hat{\xi}$)
FS	$\beta = \beta_0$	-20.1	2.61 (0.08)	0.44 (0.05)	-0.33 (0.08)
	$\beta = \beta_0 + \beta_1 Sl(t)$ Sl(t) = annual mean salinity	-19.7	(2.16, 0.23) (0.51, 0.26)	0.43 (0.05)	-0.29 (0.09)
	$\beta = \beta_0 + \beta_1 Sl(t)$ Sl(t) = annual min salinity	-19.9	(2.78, -0.13) (0.39, 0.28)	0.45 (0.06)	-0.33 (0.08)
FC	$\beta = \beta_0$	- 17.2	3.27 (0.07)	0.38 (0.05)	-0.25 (0.14)
	$\beta = \beta_0 + \beta_1 Sl(t)$ Sl(t) = annual mean salinity	-16.8	(2.93, 0.17) (0.39, 0.20)	0.37 (0.05)	-0.24 (0.16)
	$\beta = \beta_0 + \beta_1 Sl(t)$ Sl(t) = annual min salinity	-16.6	(3.57, -0.21) (0.33, 0.24)	0.39 (0.05)	-0.28 (0.14)

Table 4.7. Maximised log-likelihoods and parameter estimates of block maxima models

(Models are estimated using observations from 9 of the 11 signed sites where salinity data are available)

coefficients of salinity are negative which suggests that the relationship between extreme observations and salinity is the opposite to what is expected. Although there was some problem due to missing salinity data, models were also fitted where the salinity statistic used was the value corresponding to the same sample as the maximum indicator count however as before, the model with the constant location parameter proved to be adequate.

In summary, it is clear that inclusion of the salinity data within the model would not result in more accurate return level based discounting limits. These findings are in agreement with the initial impressions of the data formed from the exploratory analysis, which did not provide evidence of any relationship between the salinity and extreme observations.

4.6 Summary

Extreme value analysis was used in order to obtain a discounting limit for each indicator since it enabled models to be fitted to the observations which are of most relevance; those which lie in the upper tails of the distribution. In terms of identifying suitable discounting limits that could be used to identify which samples should be removed from compliance calculations the largest samples from each bathing site are of the most interest. Using models which were fitted to the observed data, return levels could be extrapolated for different return periods and subsequently the suitability of each of these values as discounting limits could be assessed. There are however several problems which are encountered when using extreme value analysis, one of these is that by definition extreme values occur infrequently which can result in a limited quantity of data on which to estimate a model. While in many situations another question is what block size is appropriate for maxima and k-th largest models, for this project the choice was limited given the format of the data available. Similarly, for threshold models, there is often some difficulty in the identification of suitable threshold values which will generate enough data to use to fit the model while still identifying values that can truly be considered as extreme. However, a reasonable suggestion of using the 'Sufficient' category limits as the thresholds came from the context of the problem and proved to be suitable after looking at the appropriate diagnostic plots.

In terms of using extreme models to find an appropriate discounting limit the main difficulties were the specification of a suitable return level and the continuing problem of extensive variation between the sites. The 4 year return level estimates were exceeded too infrequently to make any notable difference to the compliance classification of sites when applied as discounting limits. While the block maxima model was a reasonable fit for the data, the way in which the model was constructed meant that the same amount of weight was given both to sites which did not contain any extreme values and to those where observations were consistently high. This resulted in 16 month and 2 year return level estimates which were misleadingly low and although these substantially improved the overall level of compliance, they also removed far more counts than that currently permitted by the 2006 Directive. In contrast, the construction of the threshold model overcame the problem of including information from sites where there were no large observations and enabled 4 month return levels to be estimated.

Inclusion of a covariate was also considered. The quantity of salinity data available meant that it was most appropriate to include this information within the block maxima model however it was clear from the block maxima models considered that incorporating salinity was of no additional benefit in terms of explaining the heterogeneity amongst the observations.

While using extreme value analysis to obtain a generic discounting limit for all sites did improve the level of compliance, particulary with the threshold model, and did not perform any worse than simply removing the top two annual values, the success of many of the limits obtained was limited due to the extensive variation across the sites which resulted in too many values being discounted. Following from this, Chapter 5 looks at extending the threshold model approach to a site specific level in order to achieve more accurate estimates. In addition, the quantity of data removed from each site using these individual site limits is also investigated.

Chapter 5

Site Specific Threshold Models

Each of the threshold models previously considered incorporated all observed data which exceeded a suitable threshold regardless of the location at which the count was taken. However, in order to fully account for the heterogeneity across the sites it was thought that any discounting limits obtained should be site specific and for this reason threshold models were next fitted individually at each of the 11 bathing water sites where there are currently electronic signs in place. Moreover, it was thought that site specific models may be more appropriate since extreme value analysis is ideally suited to situations where there are regular observations taken from a single location over a period of time.

5.1 Fitting Site Specific Models to the Bathing Waters Data

The limited quantity of data available at each site (approximately 80 observations in total over the 4 year period) causes problems in the selection of an appropriate threshold. As before, the value chosen at each site has to be large enough so that both the asymptotic properties on which the model is based are satisfied and that observations which exceed the threshold can be considered extreme. On the other hand, any threshold used has to generate an adequate number of excesses on which to estimate the model parameters. Since the choice of threshold is somewhat subjective it was decided to set the thresholds for each microbiological indicator, at each site, so that the rate of excess was 15% (resulting in around 12 threshold exceedances on which to estimate a model). Each of these thresholds was then checked using mean residual life plots to ensure it was within the range of reasonable values. The use of these diagnostic plots as an appropriate method for identifying a suitable threshold was indicated by Davison and Smith (1990). For the overall model which was previously fitted, the threshold values used were equal to the 'Sufficient' limit for each of the indicators. One of the reasons for this choice was that observations which fell above these limits could be considered extreme events.

Table 5.1 shows the thresholds for each indicator which were identified using a 15% rate of excess at each of the signed sites.

For both indicators there are considerable differences between the thresholds at each site, which again is a reflection of the heterogeneity across the bathing waters. It is hoped that the different thresholds applied for each of the site specific models take account of this variation. Further evidence that the 15% rate of excess thresholds were suitable came from the fact, with few exceptions, it was primarily at sites where the Directive's minimum standard had already been achieved, and therefore where the observations are in general relatively small, that the thresholds obtained were notably smaller than the 'Sufficient' category limits. Although the Saltcoats/ Ardrossan site was classed as 'Poor' the

Site	FS (cfu/ 100ml)	FC (cfu/ 100ml)
Portobello Central	**78**	**168**
Eyemouth	350	700
Sandyhills	280	650
Saltcoats/Ardrossan	90	330
Irvine	220	670
Troon	**70**	**180**
Prestwick	240	440
Ayr	220	950
Brighhouse	280	510
Ettrick Bay	260	1200
Aberdeen	170	341

Table 5.1. Site specific thresholds for signed sites (based on a 15% rate of excess)
Sites which met the 2006 Directive's minimum required standard prior to any discounting are shown in bold.

thresholds obtained were smaller than the 'Sufficient' category limits, the reasons for this can be seen by looking at the 90th percentiles on which the classification is based; 193 cfu/100ml for FS and 513 cfu/100ml for FC. The overall 'Poor' classification of the site is based on the FC indicator only and even then, the percentile value for this indicator is very close to the 'Sufficient' limit of 500 cfu/100ml.

Model Diagnostics

When models were fitted using thresholds based on a 15% exceedance rate, diagnostic plots obtained indicated that the model was a reasonable fit to the data at around half of the sites. An example of one such site (Aberdeen) is shown in Figure 5.1. Although at some sites there is some discrepancy between the fitted and the observed models in the probability and quantile plots, this could be anticipated given the small number of data points on which the model is based.

At several of the sites, all entries in the parameter covariance matrix for the fitted models were extremely small. This was evident in the diagnostic plots since for these sites the return level plots obtained showed no difference between the confidence bands and the fitted return level curve. One possible explanation for this is that because the models have been fitted using very few observations on the $\log_{10}$ scale, the variation between the threshold excesses at some sites is very small in comparison to the variance of all observations at the same site. Figure 5.2 shows the diagnostic plots obtained for one of the signed sites (Prestwick) where this appears to be a problem.

It was initially thought that small variation in threshold excesses could be a result of the number of counts which were at the upper limits of detection and so were effectively right censored observations, however very few observations at the signed sites were affected by this. Site specific models which were fitted to the raw data produced improved diagnostic plots due to the increased variation between the threshold excesses, however this resulted in return level estimates which were thought to be less scientifically justifiable. It could be argued that because the percentiles used to assess compliance are calculated on the $\log_{10}$ data and back transformed to obtain values which are meaningful in terms of the Directive, any return level estimates which are going to be used as discounting limits should also be found by using models based on $\log_{10}$ data and then converted back.

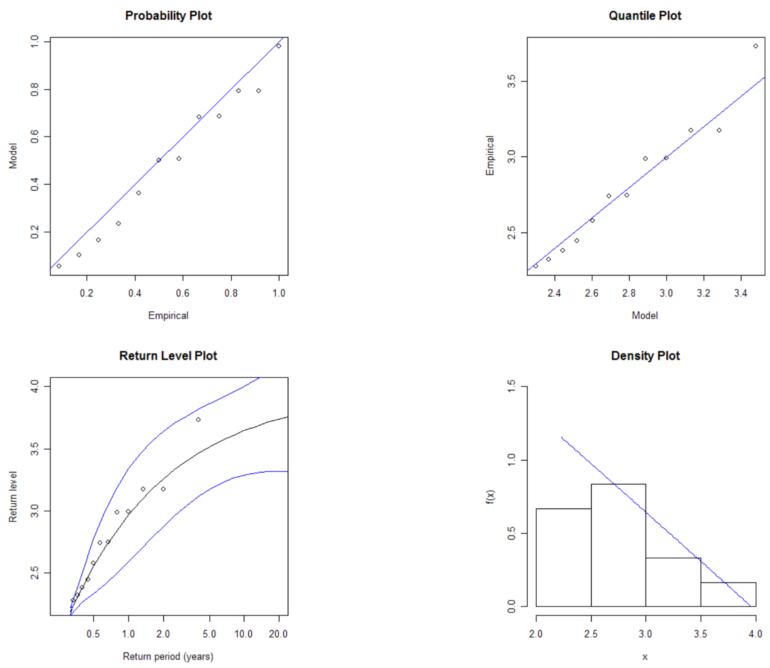

Figure 5.1. Diagnostic plots for Threshold Model at Aberdeen Site (FS)

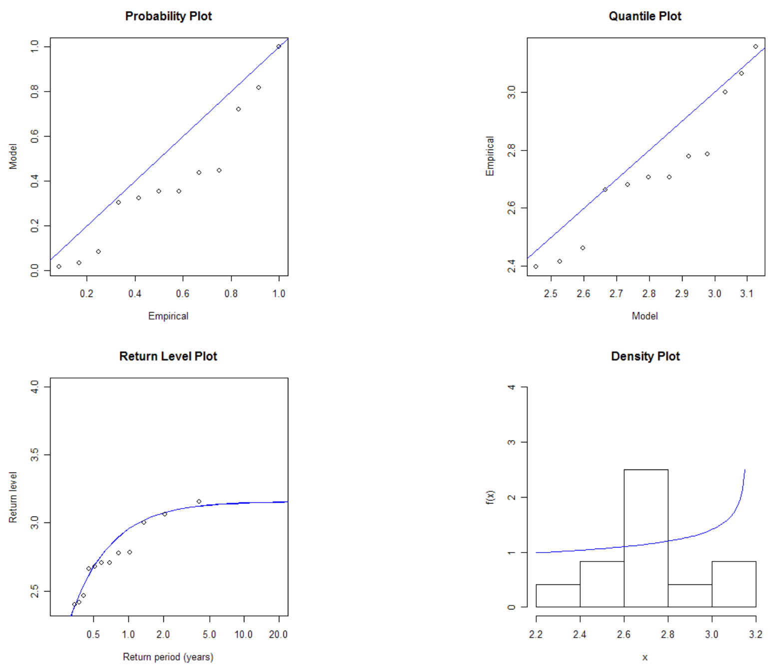

Figure 5.2. Diagnostic plots for Threshold Model at Prestwick Site (FS)

In addition, for models estimated using the $\log_{10}$ data, if the threshold is set at a very low level, so that the exceedance rate is above 40%, although there is a marked improvement in diagnostic plots due to the increased number of data points on which the model is estimated (and therefore the increased variation between the threshold excesses) again the return levels obtained are likely to be less accurate since the model has not been estimated using data which could truly be considered extreme. Subsequently it was decided that the models using the transformed data with thresholds based on a 15% exceedance rate were most appropriate. For each of these the return levels obtained seemed reasonable in the context of the data.

5.1.1 Applying Discounting Limits

As with the threshold models previously fitted, 4 month, 16 month, 2 year and 4 year return levels were obtained from each of these site threshold models and applied as discounting limits separately. While each of the 4 year, 2 year and 16 month return level discounting limits resulted in only one additional signed site meeting the minimum required standard, the 4 month limits increased the number of compliant sites from 2 (18.2%) to 8 (72.7%). Table 5.2 shows the compliance classification of each of the signed sites when the 4 month limits were applied as well as the percentile values, the percentage of counts removed in each year and the average percentage of counts removed across all four years.

As can be seen, at the three sites which continued to be classed as 'Poor' after application of the 4 month discounting limits there was a substantial reduction in percentiles on which compliance is based. Furthermore, at two of these three sites (Ayr and Ettrick Bay) the failure to meet the Directive's 'Sufficient' standard was only in the FC indicator.

Using these limits the problem of removing too many counts remained and at all of the signed sites there was at least one year where more than 15% of the data exceeded the discounting limit and was therefore removed from the percentile calculation used to gauge compliance. However it should be noted that at all sites where there is a classification change the overall percentage of counts removed across the four bathing seasons is very close to 15. In addition there were several sites where discounting all return level exceedances resulted in the classification improving beyond the minimum standard required, with several sites moving from ‘Poor’ to ‘Good’ classification. At the site where the largest quantity of data is removed (Troon, where 50% of FS is removed in 2004 and 18.8% on average) the ‘Sufficient’ standard had already been met prior to any discounting.

If 15% of counts were allowed to be removed across all 4 years then a discounting limit may be more effective in terms of improving the level of compliance. Subsequently, from the counts which are already identified as ‘Poor’ according to the limit used to protect human health, the discounting limit could also then be used to identify which of these samples, if any, should be discounted at the end of each year. This would be particularly useful in light of the year to year variation in both indicators which can again be seen from the differing percentages of counts removed in each year across each site. For example, at all sites 2006 appeared to be better than 2004 and 2007 in terms of the number of large counts observed.

Table 5.3 contains the levels of compliance achieved when the 4 month return level based limits were applied with a restriction of 3 samples at most being removed from each of the sites during each bathing season. Application of these

2007 Overall Compliance Signed Sites								
Site	**No Discounting** (90 %ile)	**4 Month RL** (90 %ile)	**% Removed**					
				2004	2005	2006	2007	Avg
Portobello	Sufficient	Excellent	FS	5	5	20	31.6	15.4
Central	134(FS), 380(FC)	48(FS), 138(FC)	FC	0	10	10	26.3	11.6
Eyemouth	Poor	Sufficient	FS	5	25	5	26.3	15.3
	443(FS), 1374(FC)	155(FS), 349(FC)	FC	10	15	10	26.3	15.3
Sandyhills	Poor	Poor	FS	36.8	5	0	15.8	14.4
	403(FS), 1256(FC)	225(FS), 721(FC)	FC	10.5	5	10	31.6	14.3
Saltcoats/	Poor	Good	FS	25	10	10	21	16.5
Ardrossan	139(FS), 513(FC)	67(FS), 211(FC)	FC	15	10	10	26.3	15.3
Irvine	Poor	Sufficient	FS	15	15	10	15.8	13.9
	298(FS), 1298(FC)	113(FS), 413(FC)	FC	20	5	15	21	15.3
Troon	Sufficient	Excellent	FS	50	15	10	0	18.8
	108(FS), 307(FC)	44(FS), 104(FC)	FC	35	10	10	5.3	15
Prestwick	Poor	Good	FS	15	18.2	15	15.8	16
	267(FS), 847(FC)	87(FS), 244(FC)	FC	15	13.6	5	31.6	16.3
Ayr	Poor	Poor	FS	25	5	10	21	15.3
	403(FS), 1725(FC)	182(FS), 831(FC)	FC	25	0	5	31.6	15.4
Brighouse	Poor	Sufficient	FS	40	14.3	0	10.5	16.2
Bay	335(FS), 868(FC)	135(FS), 398(FC)	FC	30	14.3	5	15.8	16.3
Ettrick Bay	Poor	Poor	FS	15	20	10	15.8	15.2
	333(FS), 1725(FC)	133(FS), 831(FC)	FC	20	10	5	10.5	11.4
Aberdeen	Poor	Good	FS	25	20	10	10.5	15.1
	349(FS), 868(FC)	117(FS), 398(FC)	FC	20	10	5	21	14

Table 5.2. Compliance classification in 2007 using 4 month site specific return levels as discounting limits

2007 Overall Compliance - with 15% restriction Signed Sites								
Site	**No Discounting** (90 %ile)	**4 Month RL** (90 %ile)	**% Removed**					
				2004	2005	2006	2007	Avg
Portobello	Sufficient	Good	FS	10	10	15	15.8	12.7
Central	134(FS), 380(FC)	69(FS), 167(FC)	FC	0	15	10	15.8	15.2
Eyemouth	Poor	Poor	FS	5	15	10	15.8	11.5
	443(FS), 1374(FC)	239(FS), 375(FC)	FC	15	15	15	15.8	15.2
Sandyhills	Poor	Poor	FS	15.8	5	0	15.8	9.2
	403(FS), 1256(FC)	305(FS), 813(FC)	FC	15.8	10	10	15.8	12.9
Saltcoats/	Poor	Good	FS	15	10	10	15.8	12.7
Ardrossan	139(FS), 513(FC)	86(FS), 224(FC)	FC	15	15	10	15.8	14.0
Irvine	Poor	Sufficient	FS	15	15	10	15.8	14.0
	298(FS), 1298(FC)	155(FS), 449(FC)	FC	15	10	15	15.8	14.0
Troon	Sufficient	Good	FS	15	15	10	0	10
	108(FS), 307(FC)	67(FS), 139(FC)	FC	15	15	15	5.5	12.6
Prestwick	Poor	Sufficient	FS	15	13.6	15	15.8	14.9
	267(FS), 847(FC)	136(FS), 296(FC)	FC	15	13.6	5	15.8	12.4
Ayr	Poor	Poor	FS	15	10	10	15.8	12.7
	403(FS), 1725(FC)	242(FS), 1019(FC)	FC	15	0	10	15.8	10.2
Brighouse	Poor	Poor	FS	15	14.3	0	5.3	8.7
Bay	335(FS), 868(FC)	226(FS), 464(FC)	FC	15	14.3	5	15.8	12.6
Ettrick Bay	Poor	Poor	FS	15	15	10	15.8	14.0
	333(FS), 1725(FC)	181(FS), 924(FC)	FC	15	15	10	15.8	14.0
Aberdeen	Poor	Poor	FS	15	15	10	10.5	12.6
	349(FS), 868(FC)	159(FS), 281(FC)	FC	15	10	5	15.8	11.5

Table 5.3. Compliance classification in 2007 using 4 month site specific return levels as discounting limits - a maximum of 3 threshold excesses were removed in each bathing season.

Percentage removed may be greater than 15% where there are fewer than 20 samples collected in each bathing season.

limits with this restriction imposed results in three additional sites meeting the minimum required standard in comparison to when there is no discounting. In addition, the two sites which were classed as 'Sufficient' prior to any discounting are classed as 'Good' when these limits were applied. The percentage of compliant sites in terms of FS only increased from 27.3% to 63.6% while for FS the increase was slightly greater, moving from 18.2% to 72.7% after discounting. Overall the number of compliant sites increased from 2 (18.2%) to 6 (54.5%). Although the increase in the level of compliance achieved is not as great as when all threshold excesses are discounted, it is reassuring to see that even with the restriction imposed there continues to be a considerable improvement in terms of the number of sites which meet the minimum required standard.

5.2 Assessing the Robustness of Discounting Limits

Due to the year to year variation in the data caused by changing weather conditions it is of interest to see if discounting limits obtained from these models will achieve a similar level of compliance for different periods of 4 bathing seasons. Data from the 2003 bathing season was used in order to do this. Table 5.4 contains summary statistics for the 11 signed sites in 2003 (summary statistics for the bathing seasons between 2004 and 2007 are shown in Table 2.2). Similarly to each of the bathing seasons that have previously been used to assess compliance, it is clear there is a high degree of variability at this site for both microbial indicators.

In order to assess how sensitive the procedure of using threshold models to

2003 Summary Statistics		
	FS	**FC**
Min	1	1
1st Qu	6	20
Median	28	70
3rd Qu	72	250
Max	1000	10000

Table 5.4. Summary statistics for signed sites in 2003

obtain discounting limits was to different data the above analysis was repeated on the 2003 to 2006 bathing water data. Threshold models were again fitted at individual site level for each indicator on the $\log_{10}$ transformed data with thresholds based on a 15% rate of excess. Return levels for 4 year, 2 year, 16 months and 4 months were obtained from these models and subsequently used as discounting limits.

As at the end of 2007, there were 2 out of the 11 signed sites using the 2003 - 2006 period which met the sufficient standard with no discounting. In addition, similarly to the models fitted using the 2004 - 2007 data, application of each of the 4 year, 2 year and 16 month return level discounting limits resulted in only one additional site meeting the Directive's minimum required standard. The 2003 - 2006 4 month return levels also achieved the same level of compliance as the 2004 - 2007 limits, improving the number of sites which were classed as sufficient or better from 2 (18.2%) to 8 (72.7%).

It is apparent from the results shown in Table 5.5 that at all of the sites the percentage of counts removed in each year continued to lie above the 15% that is currently permitted for some of the years. Although on average slightly more data exceeds the discounting limits in the 2003 - 2006 data than exceeds the

2006 Overall Compliance Signed Sites								
Site	**No Discounting** (90 %ile)	**4 Month RL** (90 %ile)	**% Removed**					
				2003	2004	2005	2006	Avg
Portobello Central	Sufficient 118(FS), 334(FC)	Excellent 33(FS), 97(FC)	FS	10	20	20	20	17.5
			FC	15	5	15	20	13.8
Eyemouth	Poor 232(FS), 654(FC)	Good 71(FS), 165(FC)	FS	10	15	30	10	16.3
			FC	10	15	20	20	16.3
Sandyhills	Poor 425(FS), 1033(FC)	Poor 239(FS), 573(FC)	FS	15	36	5	0	14
			FC	30	10	10	10	15
Saltcoats/ Ardrossan	Poor 149(FS), 505(FC)	Good 74(FS), 231(FC)	FS	20	25	10	10	16.3
			FC	30	15	15	10	17.5
Irvine	Poor 226(FS), 762(FC)	Good 91(FS), 221(FC)	FS	5	30	25	10	17.5
			FC	10	20	10	25	16.3
Troon	Sufficient 121(FS), 283(FC)	Excellent 51(FS), 82(FC)	FS	5	40	15	5	16.3
			FC	5	35	15	15	17.5
Prestwick	Poor 287(FS), 546(FC)	Good 93(FS), 173(FC)	FS	20	15	18	15	17
			FC	20	20	18	10	17
Ayr	Poor 363(FS), 1178(FC)	Poor 174(FS), 635(FC)	FS	20	25	10	10	16.3
			FC	20	25	5	15	16.3
Brighouse Bay	Poor 327(FS), 1072(FC)	Sufficient 115(FS), 446(FC)	FS	10	40	14	0	16
			FC	20	30	14	5	17.3
Ettrick Bay	Poor 337(FS), 1956(FC)	Poor 134(FS), 843(FC)	FS	10	20	25	10	16.3
			FC	25	20	10	5	12.5
Aberdeen	Poor 354(FS), 440(FC)	Good 117(FS), 163(FC)	FS	11	25	15	10	15.3
			FC	11	25	15	10	15.3

Table 5.5. Compliance classification in 2006 using 4 month site specific return levels as discounting limits

2006 Overall Compliance - with 15% restriction Signed Sites									
Site	**No Discounting** (90 %ile)	**4 Month RL** (90 %ile)	**% Removed**	2003	2004	2005	2006	Avg
Portobello Central	Sufficient 118(FS), 334(FC)	Good 81(FS), 211(FC)	FS	10	15	15	15	13.8
			FC	15	5	15	15	12.5
Eyemouth	Poor 232(FS), 654(FC)	Sufficient 117(FS), 192(FC)	FS	10	15	15	10	12.5
			FC	10	15	15	15	13.8
Sandyhills	Poor 425(FS), 1033(FC)	Poor 318(FS), 658(FC)	FS	15	15.8	5	0	9.0
			FC	15	10.5	10	10	11.4
Saltcoats/ Ardrossan	Poor 149(FS), 505(FC)	Good 93(FS), 246(FC)	FS	15	15	10	10	12.5
			FC	15	15	15	10	13.75
Irvine	Poor 226(FS), 762(FC)	Sufficient 129(FS), 277(FC)	FS	5	15	15	10	11.25
			FC	10	15	10	15	12.5
Troon	Sufficient 121(FS), 283(FC)	Good 78(FS), 126(FC)	FS	5	15	15	5	10
			FC	5	15	15	15	12.5
Prestwick	Poor 287(FS), 546(FC)	Sufficient 160(FS), 204(FC)	FS	15	15	13.6	15	14.7
			FC	15	15	13.6	10	13.4
Ayr	Poor 363(FS), 1178(FC)	Poor 238(FS), 725(FC)	FS	15	15	10	10	12.5
			FC	15	15	5	15	12.5
Brighouse Bay	Poor 327(FS), 1072(FC)	Poor 199(FS), 528(FC)	FS	10	15	14.3	0	9.8
			FC	15	15	14.3	5	12.3
Ettrick Bay	Poor 337(FS), 1956(FC)	Poor 189(FS), 903(FC)	FS	10	15	15	10	12.5
			FC	15	15	10	5	11.3
Aberdeen	Poor 354(FS), 440(FC)	Sufficient 161(FS), 184(FC)	FS	11.8	15	15	10	13.0
			FC	5.9	15	15	10	11.5

Table 5.6. Compliance classification in 2006 using 4 month site specific return levels as discounting limits - a maximum of 3 threshold excesses were removed in each bathing season.

Percentage removed may be greater than 15% where there are fewer than 20 samples collected in each bathing season.

corresponding 2004 - 2007 limit, there are again several sites where there is an improvement in the classification of more than one category change; for example Portobello moves from ‘Sufficient’ to ‘Excellent’ compliance.

Table 5.6 contains the results when the limits were applied with the restriction of removing at most 3 samples from each site during each bathing season. As before, even with this restriction there continues to be an increase in the number of sites which meet the Directive’s minimum required standard compared to when there is no discounting. The percentage of sites which are at least ‘Sufficient’ in terms of FS only increases from 27.3% to 63.7% and for FC only the increase is from 18.2% to 72.7% after discounting. Overall the number of compliant sites increased from 2 (18.2%) to 6 (54.5%). These are the same percentages which were achieved during the 2004 - 2007 data.

While it could be anticipated that the 2003 - 2006 models would perform just as well in terms of producing appropriate return level based discounting limits since 3 out of the 4 years data is the same as used in the 2004 - 2007 models, the 2007 bathing waters report produced by SEPA states that 2007 was a particularly poor year due to extensive coastal rainfall throughout the season. Consequently, the fact that the 2003 - 2006 limits obtained achieved the same levels of compliance and that very similar quantities of counts were removed at each site indicates that the method of using site specific threshold models is appropriate for determining discounting limits.

5.3 Summary

After considering the diagnostic plots there was some initial concern as to the suitability of the threshold models. However, it became clear this was due to the limited quantity of data on which the models were estimated.

As can be seen, application of the 4 month return levels as discounting limits at individual site level both increased the overall level of compliance amongst the sites and substantially reduced the percentiles even at the sites where there is no change in classification. While there continues to be a problem with the percentage of counts which are removed during each year the average percentage across the 4 year period is very close to 15% at most of the sites considered and when the number of counts removed at each site in each bathing season was limited to a maximum of 3 there continued to be an increase in the level of compliance achieved with 4 additional sites meeting the Directive's minimum required standard. In addition to this, when the effectiveness of discounting limits found using this procedure was tested on a different 4 year dataset very similar results were obtained in terms of the percentage of the counts which were removed. This is in spite of the year to year variation which indicates that the threshold models fitted are reflecting the variation in the observed counts.

Chapter 6

Discussions and Conclusions

6.1 Summary

The initial aim of this thesis was to obtain a Single Sample Limit for the microbial indicators FS and FC that could be used to assess the quality of water at Scottish beaches and to determine whether or not this could be set generically. However, after exploration of the data and using compliance as a measure of outcome to assess different approaches to a SSL, it quickly became clear that a single numerical value could not be found which could be used both to protect human health and which would identify the samples that should be discounted from the compliance dataset. This was due to the extensive variation both across the sites and within the same site from year to year. Subsequently the aim of the work became finding a discounting limit that could be used to determine the samples that could be removed from the percentile calculations on which compliance is based. Extreme value analysis was the statistical methodology that was used in order to do this.

Exploratory analysis and assessing the effects of discounting

Chapters 2 and 3 considered the distribution of the data and looked at the level of compliance achieved after application of several different candidate definitions of a SSL. The two site independent approaches which were considered - one pragmatic approach and one based on WHO guideline values - did improve the level of compliance amongst the sites, however they did not significantly improve on removing the two annual maximum values from each site. In addition, despite being applied at individual site level, the method of adjusting the geometric mean in order to obtain a single sample limit was difficult to apply since it could only be used at sites which were originally classed as 'Poor'. This method, although site specific, did not improve on the level of compliance achieved using generic limits without removing more than 15% of the data. Variation in the data, distributional properties and results from naive SSL approaches indicated that identification of a generic SSL would not be achievable. Therefore, from this point forward it is only discounting limits which are considered.

Extreme Value Models

Chapters 4 and 5 used return level limits obtained from fitted extreme value models as discounting limits and assessed their impact on the levels of compliance achieved. Initially each of the models was fitted to data from all of the locations since one inherent problem of extreme value analysis is that by definition, extreme observations do not occur regularly and so the quantity of data available to estimate model parameters is often limited.

While providing a reasonable fit to the data, due to its structure, the block maxima model failed to take account of the extensive variation in the data and consequently the return level estimates obtained were unsuitable for use as discounting limits. In addition, an appropriate return period could not be specified for this model since it included only one annual statistic from each location. The fit of the model did however indicate that the approach of using extreme value analysis was suitable for this data. For this reason, salinity was later included within this model in order to assess the significance of an additional relevant covariant in explaining some of the variation in the data. Theoretically, the K-th largest order statistic model which was next investigated included more information relating to the extreme observations, although when fitted it did not appear to be suitable for the bathing water data. The reason for this lack of fit was again a result of both the heterogeneity in the data, and the fact that in GEV family models equal weight is given to observations originating from all sites.

Threshold models were thought to overcome the main problem of the GEV models by including only information which could be regarded as extreme. There was some question as to what threshold was appropriate, however the limits for the 'Sufficient' category within the directive presented a reasonable option which seemed suitable after looking at relevant diagnostic plots. The threshold model also enabled a 4 month return level to be specified, relating to the idea of potentially removing 3 values in each bathing season. In spite of providing a reasonable fit to the data, for the overall model, none of the return level based discounting limits achieved a notable improvement in the level of compliance while not removing too many counts from the percentile calculations. This resulted in the extension of the threshold model approach to individual site level. The decision of what threshold to use at each site was determined by the limited quantity of available data and models were fitted separately for each indicator. The 4 month return level limits obtained were found to be the most effective to use as discounting limits and although application of these removed more than 15% of the data in each year the average quantities discounted over the 4 year periods were close to this value. The level of compliance amongst the signed sites increased from 18.2% to 72.7% after application of these limits compared to 45.5% which was achieved by simply removing the top two values from each year and each site. When the number of counts in each year was limited to a maximum of three threshold excesses in each year at each site there continued to be an increase in the level of compliance achieved with 6 (54.5%) of the 11 signed sites meeting at least the Directive's minimum required standard.

Variation in the Data

Throughout all of the analysis presented the main feature of the data has been the extensive variation in the observations. This variation restricted the effectiveness of any of the limits which have been obtained and often resulted in too much of the data being removed at some sites while very few counts are removed at others. It is clear that the Directive's restriction of discounting at most 15% of the data at each site during each year means that the best results will be achieved by removal of the maximum 2 or 3 annual values, however, the site specific threshold model based limits which were obtained improved the consistency of the percentage of counts which were removed across the sites. Application of these limits resulted in approximately the same quantity of data being removed from each site over the 4 year period which indicated that the return level limits obtained were suitable for each particular location. If slightly more than 15% is removed this is not necessarily a problem as the values used could identify the samples that could potentially be discounted at the end of the 4 year period. While there continued to be a large degree of year to year variation due to the unavoidable variability of British weather conditions, the results obtained for the 2003 - 2006 data set were very similar to those obtained for the 2004 - 2007 data. This is further evidence that this approach to finding discounting limits was suitable.

6.2 Limitations of the Study

Distributional Assumptions

Although the 2006 Directive sets standards in terms of 4 year percentile values it is important to note that since there is a marked distinction between population characteristics and sample statistics it may be misleading to use a sample based estimator of the population characteristic of interest. The mean and standard deviation of the sample used to obtain the upper percentile values may be significantly different to the corresponding population values. Bartram and Rees (2001) provides a general outline for the design and implementation of water quality monitoring programmes and considers problems such as this. While the distinction between sample and population estimates is acknowledged, it is also stated that accepting this sampling error can be justified and that the risk of misclassification, particularly at sites which are very close to the limits of compliance, cannot be avoided in practice. One idea proposed by Bartram and Rees (2001) to deal with this is to allow for sampling errors within any classification limit by defining a confidence region around the limit however guidance for this is not provided in the 2006 Directive. This may be an area for future development.

There is also some question as to the suitability of using percentiles based on the distributional assumption of log-normality to determine the compliance. The suitability of the log-normal is discussed both in Singh *et al.* (2007), with reference to its use in general environmental contexts, and in Chalwa and Hunter (2005) which is specifically related to its use as a basis for classification of bathing waters. Chalwa and Hunter (2005) considered using parametric percentile values to assess the classification of Irish bathing water sites before the 2006 Bathing Water Directive was published and concluded that using this method to gauge compliance was statistically unreliable due to the failure in the assumption of log-normality at many beaches.

This distributional assumption was also a concern for the Scottish bathing water data and consequently was investigated in Chapter 2. The Bathing Water Directive classification system is based on the assumption of log-normality and consequently this method was used throughout the thesis to examine the impact of different discounting limits on compliance since if any of these limits were to be applied by SEPA, it would be in accordance with the Directive. However, this assumption did not seem unreasonable as, although there were some discrepancies between the percentiles obtained using the two methods, it has been shown that overall there was general agreement between the parametric and empirical based classifications of each site with only a few exceptions at sites which lie on the borderline of two classification categories.

Sampling

While SEPA currently take far more samples from sites than is required by the Directive, a further increase in the number of samples collected would mean both that there is more choice in which samples should be discounted and that a greater number of samples could be discounted from the percentile calculations. In addition, a greater number of observations would increase the quantity of data available to estimate site specific threshold models and therefore improve the accuracy of any return level discounting limits which are obtained. Infrequency of data collection was identified as a general problem within environmental monitoring in a report by the Environmental Research Funders Forum (Slater *et al.*, 2006), whose aim is to improve the effectiveness of environmental research funding in the UK.

Although increasing the quantity of data available would be useful it is likely that this would be financially inefficient and any improvements would far be overshadowed by the cost of collection and analysis of additional samples. In addition, as discussed previously there is also the possibility that samples taken on days which are close together could result in dependence between the observations and so would invalidate one of the assumptions required in order to fit accurate extreme value models. Despite the infeasibility of increasing the number of samples there is a strong argument for improving the consistency of samples taken at each of the bathing water locations. The number of samples collected at each site during each bathing season is kept approximately constant however due to practical limitations such as site access or adverse weather conditions which prevent sampling, there are often different numbers of samples taken at each site during each year. More regular collection of samples, both in terms of date of collection and frequency, would enable fairer comparisons to be made between sites and moreover would improve the accuracy of the threshold models since the number of samples per year could be specified exactly. This is something which could be taken into account when compiling monitoring calendars, which are required to be established before the start of each bathing season under the revised Directive.

6.3 Conclusion

Due to the extensive variation in the data, using extreme value analysis has provided an approach which can be used to obtain site dependent discounting limits that are both appropriate given the context of the data and effective in reducing the percentiles used to assess compliance. While 100% compliance is not achieved, and there continues to be a problem in terms of the quantity of data which is removed in each year, the limits obtained from the site specific threshold models provide an indication as to which of the samples that have already been indicated as poor, should be discounted from the compliance calculations. It is hoped that with some judgement, these models can be used to obtain discounting limits that are both scientifically justifiable and will be effective in improving the level of compliance at the bathing water sites.

Bibliography

Bartram, J. and Rees, G. (2001), *Monitoring Bathing Waters. A Practical Guide to the Design and Implementation of Assessments and Monitoring Programmes*, first edn, London: E and FN Spon.

Brenniman, G. R., Rosenberg, S. H. and Northrop, R. L. (1981), 'Microbial Sampling Variables and Recreational Water Quality Standards', *American Journal of Public Health* **71**(3), 283–289.

Cabelli, V. J., Dufour, A. P., Levin, M. A., McCabe, L. J. and Haberman, P. W. (1979), 'Relationship of Microbial Indicators to Health Effects At Marine Bathing Beaches', *American Journal of Public Health* **69**(7), 690–696.

Cabelli, V. J., Dufour, A. P., McCabe, L. J. and Levin, M. A. (1982), 'Swimming-associated Gastroenteritis and Water Quality', *American Journal of Epidemiology* **115**(4), 606–616.

Chalwa, R. and Hunter, P. R. (2005), 'Classification of bathing water quality based on the parametric calculation of percentiles is unsound', *Water Research* **39**, 4552–4558.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, second edn, London: Springer-Verlag.

Coles, S. and Pericchi, L. (2003), 'Anticipating Catastrophes through Extreme Value Modelling', *Journal of the Royal Statistical Society, Series C* **52**(4), 405–416.

Coles, S. and Tawn, J. A. (1996), 'Modelling Extremes of the Areal Rainfall Process', *Journal of the Royal Statistical Society, Series B* **58**(2), 329–347.

Davison, A. C. and Smith, R. L. (1990), 'Models for Exceedances over High Thresholds', *Journal of the Royal Statistical Society, Series C* **52**(3), 393–442.

Eastoe, E. F., Halsall, C. J., Heffernan, J. E. and Hung, H. (2006), 'A statistical comparison of survival and replacement analyses for the use of censored data in a contaminant air database: A case study from the Canadian Artic', *Atmospheric Environment* **40**, 6528–6540.

Efstratiou, M. A. (2001), 'Managing Coastal Bathing Water Quality: The Contribution of Microbiology and Epidemiology', *Marine Pollution Bulletin* **42**(6), 425–432.

Environment Agency (2007), The Unseen Threat to Water Quality. Diffuse water pollution in England and Wales report., Technical report, Environment Agency.

European Parliament (1976), 'Directive 76/160/EEC. of 8 december 1975 concerning the quality of bathing water', *Official Journal of the European Communities* **31**, 1–7.

European Parliament (1991), 'Council directive 1991/676/EEC, concerning the protection of waters against pollution caused by nitrates from agricultural sources', *Official Journal of the European Communities* **L375**, 0001–0008.

European Parliament (2000), ‘Directive 2000/60/EC. of the European Parliament, establishing a framework for community action in the field of water policy’, *Official Journal of the European Communities* **327**, 1–72.

European Parliament (2006), ‘Directive 2006/7/EC. of the European Parliament, concerning the management of bathing water quality and repealing directive 76/160/EEC’, *Official Journal of the European Communities* **64**, 37–51.

Figueras, M. J., Polo, F., Inza, I. and Guarro, J. (1997), ‘Past, Present and Future Perspectives of the EU Bathing Water Directive’, *Marine Pollution Bulletin* **34**(3), 148–156.

Gaines, S. D. and Denny, M. W. (1993), ‘The Largest, Smallest, Highest, Lowest, and Shortest: Extremes in Ecology’, *Ecology* **74**(6), 1677–1692.

Georgiou, S. and Langford, I. H. (2006), Coastal Bathing Water Quality and Human Health Risks: A review of legislation, policy and epidemiology, with an assessment of current water quality, proposed standards, and disease burden in England and Wales, Technical report, Centre for Social and Economic Research on the Global Environment (CSERGE).

Katz, R. W., Brush, G. S. and Parlange, M. B. (2005), ‘Statistics of Extremes: Modelling Ecological Disturbances’, *Ecological Society of America* **86**(5), 1124–1134.

Kay, D., Bartram, J., Pruss, A., Ashbolt, N., Wyer, M. D., Fleisher, J. M., Fewtrell, L., Rogers, A. and Rees, G. (2004), ‘Derivation of numerical values for the World Health Organization guidelines for recreational waters.’, *Water Research* **38**, 1296–1304.

Rees, G. (1993), 'Health Implications of Sewage in Coastal Waters - the British Case', *Marine Pollution Bulletin* **26**(1), 14–19.

Robinson, M. E. and Tawn, J. A. (1997), 'Statistics for Extreme Sea Currents', *Journal of the Royal Statistical Society, Series C* **46**(2), 183 – 205.

Satpute, N. (2005), Statistical Modelling of Dissolved Oxygen in the Northern Sea Lochs of the Clyde Estuary, Master's thesis, University of Glasgow.

Scottish Executive Environment Group (2006), Better bathing waters: meeting the challenges of the revised Bathing Water Directive, Technical report, Scottish Executive.

SEPA (2007), Scottish Bathing Waters Report 2007, Technical report, Scottish Environmental Protection Agency.

Singh, A., Singh, A. and Engelhardt, M. (2007), The Lognormal Distribution in Environmental Applications, Technical report, United States Environmental Protection Agency. Technology Support Center Issue Paper.

Slater, J. A., Mole, A. C. and Waring, R. (2006), Strategic Analysis of UK Environmental Monitoring Activity for ERFF, Technical report, ADAS on behalf of Environmental Research Funders' Forum.

Smith, R. L. (2004), Extreme Values in Finance, Telecommunications, and the Environment, *in* B. Finkenstadt and H. Rootzén, eds, 'Statistics of Extremes with Applications in Environment, Insurance, and Finance', Chapman and Hall/CRC, pp. 1–78.

WHO (2006), Guidelines for Safe Recreational Water Environments, Technical report, World Health Organisation. Volume 1; Coastal and Fresh Waters.