Developing an Assurance Scheme for Shellfish and Human Health

Final Report

Produced by

Bangor University, Centre for Applied Marine Sciences

&

UK Centre for Ecology & Hydrology

Suggested citation : Malham, S.K., Cooper, D., Robins, P., Dickson, N., Farkas, K., Jones, L., Mannion, F., Sorby, R., Winterbourn, B., Thorpe, J., Allender S. and Le Vay, L., 2023. Developing an Assurance Scheme for Shellfish and Human Health. Seafish Report No: SR776. ISBN No 978-1-911073-82-6

Acknowledgements

We would like to thank Tim Marshall of Porthilly Shellfish for his continuous support throughout the project. Without the fortnightly, and sometimes more frequent, shellfish samples collected over two years this work would not have been possible.

The Shellfish Stakeholder Working Group provided the initial context that informed the scope of the project and gave valuable feedback to interim reporting.

Seafish secured European Maritime and Fisheries Fund (EMFF) funding that provided the majority of the financial support for the DASSHH project, facilitated discussions around defining the concept of an Assurance Scheme for shellfish and provided invaluable support throughout the project.

The Food Standards Agency supported this project from early concept and provided funding.

The Environment Agency provided funding and access to environmental data used in the statistical modelling.

South West Water provided data related to their operations.

We also acknowledge support of MSci students Jake Warden and Matt Bridgman and placement student John Jones, all at the School of Ocean Sciences, Bangor University.

Executive Summary

Water quality, in terms of the bacteria and viruses present, affects the incidence of microbial contamination in shellfish. If shellfish is eaten raw or is only lightly cooked, some of these microbes can cause gastro-enteric illness in humans. The Official Control framework for shellfish hygiene uses the bacterium *Escherichia coli* (*E. coli)* in shellfish flesh as an indicator of faecal contamination. *E. coli* levels in shellfish are typically measured by the Most Probable Number (MPN) test (ISO 16649-3, 2016), with results used to broadly classify production areas and monitoring contamination levels.

The DASSHH study (2019-2021) was commissioned by Seafish, on behalf of the Shellfish Stakeholder Working Group. The study was commissioned in response to industry concerns about the MPN test method and the lack of flexibility in the current system, where monitoring results used for long-term area classification may, in some cases, inform short term management decisions on temporary closures and downgrades.

Environmental conditions (e.g. high rainfall), can lead to sewage spills and increased catchment run-off, both resulting in high levels of *E. coli* in shellfish in affected water bodies. The DASSHH project set out to demonstrate the potential for an adaptive approach to management, using environmental indicators to predict suitable/unsuitable conditions for shellfish harvesting. The intent is that such a regime could provide elevated levels of regulatory assurance alongside greater operator flexibility. The DASSHH study focused on one case study site, the Camel estuary, to allow detailed investigation of factors influencing *E. coli* in shellfish (oysters and mussels), with results informing wider application of the approach.

Variability in the MPN assay is acknowledged and was considered as a potential confounding factor influencing the reliability of predictive models being developed in the DASSHH project. Hence the pour plate (ISO 16649-2, 2001) method, also approved for use in official shellfish monitoring, was also used for time series sampling of shellfish from the Camel. The pour plate method consistently yielded less variable *E. coli* results for repeat measures of single samples than obtained by MPN, particularly for the upper range of *E. coli* concentrations. MPN results were also statistically higher than pour plate, whether considered at the level of the inherent measurement variability of individual MPN results or at the level of variability in a series of single results as used in the practical application of the official control regulations. These findings suggest that the variability in data collected by the MPN method may limit the precision of predictive models based on historical MPN data, for *E. coli* in shellfish. This also suggests that the MPN method has greater potential to generate outlier results that may influence application of monitoring results

An overarching finding from the Camel study is that a real-time predictive system for *E. coli* levels in shellfish is conceptually feasible. The best predictive model developed for the Camel was based on relatively simple environmental data (rainfall radar, river flow, temperature/season) that is readily available. The MPN and pour plate shellfish data that were collected by the DASSHH project supported improved fitting of predictive models,

compared to models based on historical MPN data. This may be due to the increased sampling frequency, capturing a greater range of environmental conditions, and the pooling of samples from each monitoring point. The explanatory power of the models based on pour plate *E. coli* data were in some cases improved over the course of the DASSHH project, with some strong predictive relationships demonstrated for individual beds. The findings indicate that bed-specific models may be more appropriate than a single whole-site model. The most reliable models correctly predicting when individual *E. coli* results in shellfish fell below the classification thresholds <230 and <700 *E. coli*/100g, with 90% and 88% reliability respectively (based on the probability of correct predictions compared to measured results). This rose to 98% reliability predicting individual results relative to the <4,600 *E. coli*/100g boundary. This demonstrates that there is good potential to develop a model-driven management system, but sufficient accuracy was only achieved where *E. coli* data supplementary to the Official Control sampling was included, and accuracy was greater when using pour plate *E. coli* data. However, it is acknowledged that the predictive modelling is based on relatively small data sets over a 12-month period and there is scope for further improvement of the models, as may be required for application in an assurance scheme.

The Camel study was unable to develop satisfactory predictive hindcast models for *E. coli* in shellfish based solely on historical MPN *E. coli* results from the Official Control sampling. These data were found to be highly variable and poorly related to explanatory variables considered. Hence the explanatory power of the environmental data were often limited, and strongly influenced by small numbers of extreme values. A few extremely high MPN values were difficult to characterise statistically, and some were not associated with preceding rainfall or any other explanatory variable. One potential reason for the differences in model performance between Official Control MPN *E. coli* data and those collected for the DASSHH project is that the latter were collected more frequently (two-weekly vs monthly) and systematically on the same day every two weeks. The improved performance of models based on pour plate *E. coli* data is unsurprising given the lower inherent variability in this method (as demonstrated in the DASSHH results). However, these results are based on relatively short data sets and further modelling over longer time series is required to confirm these findings and potentially improve the models.

The contribution of Combined Sewage Overflow (CSO) spills to *E. coli* levels in shellfish was not clearly demonstrated in predictive models for the Camel estuary, and these were not included in the final models developed. This does not mean that human sewage is not a significant contributor to *E. coli* levels. Hydrodynamic modelling clearly showed dispersal of CSO discharge over shellfish beds, with some variation in duration and extent, depending on time of year and location of the outfalls relative to net tidal flow in the estuary. Viral indicators also confirmed the influence of human sewage on shellfish. Available data for CSO operation were limited; only timing/duration of discharges were available, without any measure of volume or concentration and, for some locations, the operation time series data were apparently incomplete. As CSO operation is largely influenced by weather conditions it can

also be difficult to disentangle from rainfall as a driver of other catchment sources, and rainfall provided a key predictor of *E. coli* counts in shellfish without inclusion of CSO operation in the predictive models.

Measurement of *E. coli* in shellfish does not distinguish human and animal sources of contamination. Viral monitoring in shellfish from the Camel estuary was conducted over a two-year period, to (i) assess the potential for health risks associated with the consumption of shellfish as a result of accumulation of virus in shellfish and (ii) to investigate whether the bacterial contamination affecting the area originated from animal or human sources by tracking selected indicator viruses. Strong correlation between specific viral indicators suggests that human and animal waste inputs to the Camel estuary also correlate with each other and may be responding to common environmental drivers (e.g. rainfall), making it difficult to identify periods of contamination from either source separately. Norovirus was present seasonally and sporadically in a low proportion of samples but with no clear correlation with *E. coli* numbers. There was no clear correlation between environmental predictors and pathogenic viral contamination of shellfish. This may be due to the longer retention of viruses in shellfish tissue, so that measured levels are less responsive to environmental variation, and also the seasonal nature of occurrence of viral pathogens. Thus, while a predictive modelling system can use *E. coli* as an indicator of shellfish contamination, it may need to be supplemented by monitoring of norovirus to inform harvesting decisions.

Alongside the development of predictive models for *E. coli* in shellfish, an assessment of depuration times for mussels and oysters from a range of initial microbial loadings was conducted to inform recommendations for depuration times under specific conditions for mussel and oysters. These results, together with information from other published studies, demonstrate the potential for adoption of flexibility in depuration times (12 – 72 hours), taking into account predicted levels of *E. coli* in shellfish.

The successful development of relatively simple predictive models based on readily available environmental data suggests that transferring this approach to other catchments is feasible and could be linked to ongoing environmental monitoring programmes, with establishment of real-time data links opening up potential development of a forecast system. The need for incorporation of supplementary *E. coli* time series data into predictive models suggests that independent *E. coli* data collection, ideally using the pour plate method, should be considered in identifying suitable sites for predictive model development, with ongoing monitoring to allow refinement of models over time.

Key findings

- An overarching finding from the Camel study is that a real-time predictive system for *E. coli* levels in shellfish is conceptually feasible. The relatively simple model developed for the Camel is based on environmental data (rainfall radar, river flow, temperature/season) that is readily available.

- The study demonstrated the ability to predict *E. coli* levels in shellfish, with the most reliable models, based on pour plate data, correctly assigning individual predicted *E. coli* levels in shellfish relative to classification thresholds <230 and <700 *E. coli*/100g, with 90% and 88% reliability. This rose to 98% reliability for the <4,600 *E. coli*/100g boundary. However, these models are based on relatively short time series (over 12 months) of data that could usefully be extended to improve the reliability, as may be required for application in an assurance scheme.

- The best predictive models were based on high frequency *E. coli* samples taken during the study, and when using data from the pour plate method. Supplementary MPN *E. coli* data collected during the study were more consistent between shellfish beds than the Official Control results, and improved model performance. Satisfactory predictive models could not be developed based on historical Official Control MPN *E. coli* data, which were highly variable and less strongly related to the explanatory variables considered.

- The pour plate method yielded less variable *E. coli* results than obtained by MPN (for repeat measures of single samples), particularly for the upper range of *E. coli* concentrations (within the recommended application limits for the pour plate method). The MPN method also generated statistically significantly higher *E. coli* results compared to pour plate data (paired t-test). This suggests that the MPN method has potential to generate outlier results that may influence application of monitoring results.

- The significance of CSO spills in contributing to *E. coli* levels in shellfish was not clearly demonstrated in development of predictive models for the Camel estuary, due to limitations in available spill data. However, sewage contamination of shellfish was confirmed by hydrodynamic dispersal models and presence of human indicator and pathogenic viruses in shellfish.

- As the area surrounding the Camel is predominantly rural, agricultural run-off is clearly a contributor to *E. coli* in shellfish in the Camel estuary. There was a strong correlation between human and livestock indicator viruses, suggesting that CSO operation and farmland run-off respond similarly to catchment-scale environmental drivers.

- Norovirus was present seasonally and sporadically in a low proportion of samples, with no clear correlation with *E. coli* numbers. There was also no clear correlation between environmental predictors and viral contamination of shellfish, preventing predictive modelling of viral contamination in shellfish.

- Results of depuration experiments confirmed the potential for adoption of flexibility in depuration times in response to predicted levels of *E. coli* in shellfish, though with the

caveat that human health risk may arise from specific pathogens such as norovirus that are not readily depurated.

Recommendations

- The DASSHH study provides some useful guidance on approaches that could be taken in development of an assurance scheme based on predictive modelling of *E. coli* in shellfish. Successful development of relatively simple predictive models based on readily available environmental data suggests that transferring this approach to other catchments may be feasible, linking to ongoing environmental monitoring (rain radar, river flow). Establishment of real-time data links would open up the potential development of a predictive system.

- The need for incorporation of supplementary *E. coli* time series data into predictive models is a challenging outcome from the Camel study. Transfer to other catchments may require supplementary collection of two-weekly *E. coli* data for at least 12 months. The successful outcome of the modelling is not necessarily certain, even with this investment.

- Although not for this study, the observed statistically significant differences in pour plate vs MPN *E. coli* results have wider implications, e.g. potential for occurrence of unexplained high results, some of which may potentially influence shellfish area classifications occasionally. The comparison of results from the two methods for the Camel did not determine any effect on classification of this site. The present study provides an initial basis for assessing potential effects of a change in methods on other sites.

- Modelling studies in other catchments should always include CSO operation as a predictive factor, as the location (i.e. proximity to shellfish production sites) and relative scale and frequency of CSO discharges may make them a more readily quantifiable source of shellfish contamination. This can be investigated by hydrodynamic modelling or a less onerous statistical approach to screening the relative contribution of environmental factors (including CSOs). Ideally this would involve CSO data on volumes and concentrations of release, in addition to timings.

- The relative importance of human and animal source pollution may be expected to vary between catchments, and locations within catchments. The application of viral source identification of the human and agricultural/wildlife faecal contamination can be a valuable component of interpretation and application of predictive models developed for other locations, as the *E. coli* indictor does not separate these sources and could result in overestimation of risk (for example where high *E. coli* results reflect increases in agricultural rather than sewage inputs).

- The lack of significant correlation between norovirus and *E. coli* levels in shellfish highlights the well-recognised weakness in the use of this generic indicator to quantify the most prevalent human health risks of greatest concern. However, the lack of methods for measuring infective norovirus in shellfish samples remains a limitation to use of norovirus testing in official control monitoring. An assurance scheme based entirely on environmental factors will be unable to predict human health risk from norovirus reliably. This limitation may be overcome by incorporation of new approaches to monitoring prevalence of norovirus in the human population, for example viral surveillance of wastewater, that could feed into risk models underpinning management of shellfish production areas.

- The depuration results from the present study are likely to be broadly applicable, in conjunction with other published studies on rates of clearance of *E. coli.* Appropriate depuration times could be determined for predicted periods of lower or higher contamination, ranging from 12 to 72 hours, with guidelines universally applied across any catchments. However, consideration of norovirus prevalence may also need to be taken into account (Figure 1).

- The findings from the Camel study can inform the conceptual design of a predictive tool to inform shellfish harvesting decisions that would be intrinsic to an industry-operated assurance scheme. Figure 1 illustrates how such a tool could operate, based on the environmental indicators and predictive model outcomes from the DASSHH study.

- The most appropriate technical system for implementation of the predictive tool remains to be determined, but at its most accessible it is envisaged that harvesting recommendations might be delivered via an interface such as a mobile phone app and/or desktop version. The system would require bespoke algorithms for each shellfish production area, based on the predictive model developed for each catchment. A feed of environmental data inputs would be needed, requiring engagement with the appropriate organisations (e.g. Met Office, Environment Agency, water companies). Once established, the technical solution should be transferable between production areas, ideally with a centralised platform enabling effective technical support to multiple locations.

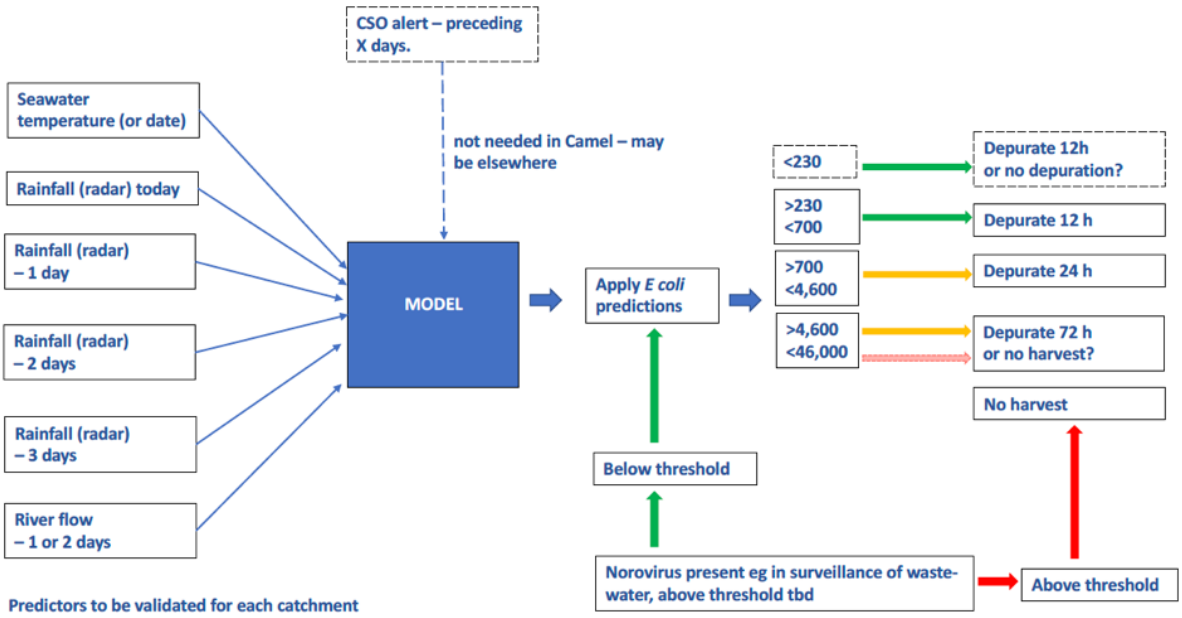


Figure 1 Generalised schematic for operation of a predictive decision tool in management of shellfish harvesting, based on findings of the DASSHH study of the Camel estuary. The predicted *E. coli* ranges are indicative and may vary from catchment to catchment.

Notes: (i) thresholds for depuration and harvesting decisions are only illustrative (ii) for B class areas, such as the Camel, shellfish depuration (or other approved processing) will always be required under current regulations if the area continues to be classed as B.

Contents

1 Introduction

1.1 The concept of an Assurance Scheme for Shellfish and Human Health

Water quality, in terms of the bacteria and viruses present, affects the incidence of microbial contamination in shellfish. If shellfish is eaten raw or is only lightly cooked, some of these microbes can cause gastro-enteric illness in humans. The current system for regulation of shellfish hygiene broadly classifies production area by microbial contamination levels, using *Escherichia coli (E. coli)* in shellfish flesh as an indicator of faecal contamination. In the UK, as elsewhere globally, *E. coli* levels in shellfish are typically measured by the Most Probable Number (MPN) test.

The Official Control framework for shellfish hygiene is set out in UK-retained EU legislation and regulations. EC Regulation 217/625 requires the routine monitoring of the microbiological quality of live bivalve molluscs in shellfish production areas, to classify areas into risk categories and to monitor contamination levels against regulatory thresholds that may require management action. This regulation requires the monitoring outlined in the repealed EC regulation 854/2004 to be retained. The results of classification monitoring determine what level of post-harvest treatment is required and when *E. coli* levels in monitoring samples exceed a classification threshold, an action state is triggered and the area may be temporarily downgraded or closed, unless investigation shows that the high result is due to an unusual event that has been resolved and the site returned to classification threshold requirements.

The detailed implementation of classification and monitoring programmes in line with EC Regulation 2019/627 is the responsibility of the relevant Competent Authority which in England, Wales and Northern Ireland is the Food Standards Agency (FSA). The FSA is required to undertake sanitary surveys of shellfish farms and wild beds, and their associated hydrological catchments and coastal waters, in order to establish the appropriate representative monitoring points (RMPs) for the subsequent monitoring programme. These RMPs can be decided in collaboration with the producer.

Production areas are classified as A, B or C, with A as the least and C as the most contaminated. Harvesting can occur from Class B and C waters but further treatment is required (depuration, heat treatment or relaying) before the shellfish can be offered for human consumption.

Routine official monitoring sampling is generally monthly, with required frequency for a particular site determined in the initial sanitary survey. Class A waters require a minimum of 10 samples per year compared to 8 samples for Class B or C waters (https://www.food.gov.uk/sites/default/files/media/document/Classification%20protocol%20-%20September%202022.pdf).

Production areas are classified as A, B or C, with A as the least and C as the most contaminated. Harvesting can occur from Class B and C waters but further treatment is

required (depuration, heat treatment or relaying) before the shellfish can be offered for human consumption.

The current use of MPN testing in the monitoring of shellfish production areas, to inform management decisions such as closure or downgrade, is a relatively imperfect way to determine health risk. This is primarily because:

a. It is a retrospective test that takes several days from sampling to results being issued; hence product can already be placed on the market by the time a high reading is recorded.

b. Producers are concerned about perceived high variability in MPN results and the risk of false positive results leading to closures and longer-term impacts on site classification.

c. The testing regime is not responsive enough. Combining the MPN test with the requirement for two consecutive tests to re-open a downgraded shellfish bed means that downgrades and closures may last for considerable periods, with significant business impact, even if risk levels have declined in the meantime.

d. The use of monthly samples to monitor shellfish areas only provides a coarse time series to assess variation in environmental quality and any associated periods of human health risk. Despite this, monitoring results used for long-term area classification may, in some cases, inform short term management decisions on temporary closures and downgrades.

The DASSHH project was initiated in response to the issues highlighted above, to explore opportunities for a finer scale and responsive management regime that complements the statutory regulations. The intent is that such a regime could provide elevated levels of regulatory assurance alongside greater operator flexibility, while continuing to sufficiently manage human health risks. As such the focus is to enable greater responsiveness within the regulatory system by equipping shellfish operators to proactively manage production areas, with an evidence-based system that avoids harvesting shellfish in periods where there is an elevated risk of microbial contamination.

Human sewage pollution may occur at a range of levels in UK coastal waters, higher levels can occur if the sewage treatment system is overloaded during environmental conditions such as high rainfall, spills of untreated sewage and increased catchment run-off from farmland of faecal material originating from livestock. The DASSHH project set out to demonstrate the potential for an adaptive approach to management of shellfish harvesting, using environmental indicators to predict periods of suitable/unsuitable conditions for shellfish harvesting.

The scope for achieving the full benefit from adoption of such adaptive management of shellfish production areas will depend on the extent to which risk management principles may

in future be aligned or integrated with the Official Control framework governing shellfish production. Although the exact mechanism to achieve this integration is still to be determined, it could take the form of an industry Assurance Scheme supported by domestic policy and regulatory guidance. However, it is acknowledged that the specifics of any solution could vary.

The initial scope of a proposed Assurance Scheme may be to use a series of environmental indicators relevant to the location of the shellfish farm (or catchment), to determine the optimal time to harvest product. This will ensure that harvesting does not occur during instances where there is a high likelihood of shellfish contamination because of water quality issues. The scheme will enable producers to make informed decisions about harvesting schedules to avoid high risk periods. It could also inform depuration, for example enabling producers to increase depuration during periods of higher risk. Conversely, in periods of low risk, depuration periods could be shortened without affecting product safety, so long as there is also information on prevalence of viral contamination (eg norovirus). Any requirement for the adaptation of depuration duration recognises that a significant amount of the UK's bivalve production is exported to EU and other international markets, hence depuration requirements may also need to be aligned with the food safety requirements of the importing country. An Assurance Scheme will need to be able to reflect that the required intervention may not be directly taken by the shellfish operator.

The assumption is that an Assurance Scheme would complement the Official Controls rather than seek to replace them. Shellfish waters would continue to be sampled and classifications awarded. A key benefit could be that if official sampling results record that there is a high likelihood of contamination the shellfish operator would already have been aware of this and adjusted their operations with positive proactive management, thus reducing potential health risk from shellfish contamination which in turn should limit the need for regulatory intervention.

Specific benefits of an Assurance Scheme could include:

- The data could inform the optimal timeframe for official sampling, or how results are interpreted or used, so as to avoid periods of isolated high results unnecessarily compromising classifications (because there will be no harvesting).

- Shellfish producers participating in the scheme may experience less frequent official sampling on the basis that they are taking positive action to reduce risks and comply with legislative requirements, verified through their participation in the Assurance Scheme.

- The regulatory burden in dealing with high *E. coli* results (downgrade/closure, investigation) could be avoided or reduced. If the official sampling returns a high result the default could be that no further action is required because the producer will be able

to demonstrate that they have monitored agreed environmental indicators and can evidence that they have either not harvested or that appropriate depuration will have taken place before the product reaches the consumer, either by the operator directly or by supply chain participants.

- Similarly, there could be a simpler, faster and more responsive process re-opening of sites that are subject to downgrade or temporary closure, informed by evidence from environmental indicators rather than current practice of re-testing. The timing of re-testing, if required following a high Official Control result, could also be informed by the predictive model.

- During low-risk periods, the operator may be able to decrease depuration times which would reduce costs and enable businesses to be more responsive to market needs.

- Greater public health assurance, and reputational benefits, as all activity (harvesting and sampling) will occur during lower risk periods within the required monthly sampling period. This would be an improvement over the current system, where businesses could inadvertently harvest during high-risk periods, due to the relatively sparse time series of Official Control sampling.

An Assurance Scheme would also need to align with current and emerging regulatory requirements and to comply with the existing legislative framework. It could provide an opportunity for development and adoption of a risk-based, proportionate, robust and resilient system for management of food safety by which:

- participating individual shellfish operators could demonstrate that they are producing food that is safe and are helping consumers make informed choices about the food they buy by branding product as coming from a shellfish farm operating under the scheme.

- a tailored and proportionate approach to regulating businesses is available to individual shellfish farms (or a consortium of operators within a single catchment), supported by and enabling a more collaborative approach between regulators and industry.

- regulators are provided with additional information that they can consider when making decisions to manage risk.

- any restrictions placed on businesses (such as closure or additional costs of production due to depuration) are appropriate and reflect the likelihood of risk such that the costs of regulation are no more than they need to be.

1.2 Research background

Estuarine environments provide a wealth of economic, social and natural benefits that include food, employment, recreation and habitation (Costanza *et al,* 1997; Barbier *et al,* 2011; Tuholske *et al,* 2021). Over 50% of the world's population lives within 100km of the coast and anthropogenic activities cause substantial impacts on the health of estuarine and ocean ecosystems (Stewart *et al,* 2008). Of particular significance is the introduction of human microbial pathogens from point and diffuse sources (Droppo *et al,* 2009; Malham et al, 2014; Tryland *et al*, 2014) and impact on the aquatic environment.

Bivalve shellfish are filter-feeding organisms and the sustainability of shellfish aquaculture is highly dependent on maintaining clean and healthy coastal waters. Microbial water quality and its relationship with pathogen load in shellfish is of particular importance with regards to protecting public health particularly as shellfish are known to bio-accumulate microbial contaminants (Potasman *et al,* 2002; Teplitski *et al,* 2009) and pose a significant public health risk when entering the human food chain (Bellou *et al,* 2013)

Faecal matter from humans and farmed livestock is the main source of pathogenic microorganisms (bacteria, protozoa and viruses) entering the estuary and coastal zone (Malham *et al,* 2014) Other sources, including wild animals such as deer, birds and rodents, play a lesser role but can be locally important (Santo Domingo and Edge*,* 2010) The pathways and timings by which these pathogenic bacteria enter the aquatic environment vary. The main pathway for human sources are point sources, such as Combined Sewer Overflows (CSOs) which discharge untreated waste into receiving waters when the sewerage infrastructure is overloaded by high rainfall (Droppo *et al,* 2009), sewage system malfunctions, septic tanks, as well discharges from boats. Pathways for animal wastes are mainly diffuse, driven by surface runoff from fields and hard-standings around farm buildings, and vary with farming practices (Kay *et al,* 2008). A third and under-appreciated pathway is from sediment and suspended sediment including 'flocs' within rivers and estuaries, which represent a significant reservoir of potential pathogens (Malham *et al,* 2014; Perkins *et al,* 2014; Bradshaw *et al,* 2016). Resuspension of sediments (Wilkinson *et al,* 2006) can substantially increase potential pathogen levels in the water column and identification of the sources and pathways is critical to accurately modelling their fate in the environment (Oliver *et al,* 2016). The most commonly-used indicator to assess contamination of shellfish with potential pathogens is the faecal indicator organism (FIO), *Escherichia coli,* which is derived from both human and non-human sources (Harwood *et al,* 2014; Oliver *et al,* 2016)

As the current Official Control framework is based on the use of *E. coli* as an indicator of faecal contamination, the DASSHH project has largely focused on developing an understanding of the relationships between environmental conditions and *E. coli* levels in shellfish. Regulation for the classification of shellfish areas in the UK utilises the ISO accredited (ISO 2016) Most

Probable Number (MPN) method for measuring *E. coli* in shellfish flesh. This method uses dilution tubes and a probability calculation to give the concentration of viable organisms in a given sample, based on the number of tubes that return a positive result (Walker *et al,* 2018). Alternative approved methods for measurement of *E. coli* in shellfish include the pour plate and impedance methods, both which have been validated and characterised against the MPN reference method (IFREMER 2014; Walker *et al,* 2018; Pol-Hofstad and Jacobs-Reitsma 2021).

Estimation of bacterial abundance in food and environmental samples is inherently variable and in the case of shellfish, *E. coli* variability is compounded across a range of sources and factors with fine scale spatial variation reported across individual shellfish beds (Beliaeff and Cochard 1995; Kay *et al,* 2008; Clements *et al,* 2015). The MPN method itself is also inherently variable, and development of statistical approaches to account for this measurement uncertainty has been integral to the evolution of the MPN method over the >100 years of its application (Jarvis *et al,* 2010). However, shellfish industry concerns about the reliability of Official Control monitoring based on the MPN method has been central to motivation for the development of new approaches to management of shellfish hygiene, including the establishment of the Shellfish Stakeholder Working Group and the investigation of the potential for adaptive management undertaken in the DASSHH project.

While the use of *E. coli* as an indicator is long-established, it is based on the assumption that these bacteria associate with human waste and hence would indicate any wastewater-derived microbial contamination in the aquatic environment. However, enteric viruses can be found at high concentrations in the faeces of infected individuals for prolonged times following infection (Aoki *et al,* 2010). Infectious viruses may be discharged into the aquatic environment where they may be accumulated by filter feeders, including oysters and mussels, with longer measurable persistence than *E. coli.* (Adriaenssens *et al*, 2021; Farkas *et al*, 2018; Lowther *et al,* 2012). Enteric viruses have also been shown to be more persistent in the environment than indicator bacteria, suggesting that *E. coli* may not be representative for viral contamination (Baggi *et al*, 2001; Espinosa *et al*, 2009). Furthermore, *E. coli* may associate with animal faecal matter as well as with human waste, and hence the presence of such bacteria may not indicate any human-derived health risks (although some animal-derived pathogens can occur) (Devane *et al*, 2020; Nguyen *et al,* 2018). To overcome these limitations and have a better understanding on viral contamination, comprehensive viral monitoring campaigns are necessary, both for specific pathogens of concern such as norovirus and for indicator viruses which can help distinguish animal and human sources in environmental samples.

The principle of an assurance scheme is to use real-time information on environmental conditions to assess the risk of microbial contamination of shellfish. Management intervention based on this risk assessment then informs both closure and reopening of shellfish beds. This approach has been implemented globally for bathing water quality (Frick *et al,* 2008; Stidson *et al,* 2011; Viegas *et al,* 2012; Shively *et al,* 2016; Avila *et al,* 2018). The more complex schemes involve real-time and short-range forecasting data sent via telemetry

to a web-server which predicts likely contaminant concentrations in target areas (Shively *et al,* 2016). In New Zealand and Australia, similar approaches use real time environmental monitoring data (e.g. rainfall, river flow, salinity) to trigger automatic alerts that inform management decisions for shellfish production areas. The majority of schemes use regression-type models to predict bathing water quality and a limited number for shellfish water quality, e.g. in the USA (Frick *et al,* 2008) and the UK (Vinten *et al,* 2004; Stidson *et al,* 2011; Zimmer-Faust *et al,* 2018).

A number of countries monitor shellfish waters (USA, Canada, New Zealand) rather than shellfish flesh (Europe and UK) to classify shellfish harvesting areas (Seafish 2021). In many locations in the USA and New Zealand, *E. coli* levels in water are considered to be a sufficiently robust proxy for *E. coli* in shellfish flesh. A range of statistical models for shellfish waters are being developed to predict spatial and temporal pollution in estuarine waters, with some approaches explaining exceedances in shellfish waters with 100 and 97% predictive power (Zimmer- Faust *et al,* 2018). In contrast, the use of predictive models for shellfish waters to predict *E. coli* in shellfish flesh, is limited (Bougeard *et al,* 2011; Campos *et al,* 2011; Schmidt *et al,* 2018). Prediction for shellfish is more challenging, given the additional uncertainty in terms of accumulation and depuration in the shellfish and requires adequate understanding of water movement within estuaries and coastal areas. Also, *E. coli* concentrations in shellfish are highly temporally and spatially variable both within and across estuaries (Malham *et al,* 2017). However, some usefully transferrable information can be drawn from the understanding developed in bathing water schemes and emerging shellfish water quality models.

Several regression type models (Zimmer- Faust *et al,* 2018) are based on factors such as catchment area, diffuse and point sources of pollution and the number of sewage treatment works (STWs) and combined sewer overflows (CSOs) (Malham *et al*, 2014; Crowther *et al*, 2002) as well as physicochemical factors including suspended particulate matter, nutrients, rainfall, tidal movements, seasonal variations, temperature, UV and salinity (Hassard *et al*, 2017; Malham *et al,* 2017). Further, catchment size, topography and soil characteristics including soil moisture at the time of rainfall may also be included (Campos *et al,* 2013). A number of studies have shown that survival rates and persistence of bacteria in the coastal zone is species and strain dependent (Campos *et al,* 2011; Hassard *et al*, 2017) and can differ between point and diffuse sources (Perkins *et al*, 2014). Bacterial survival is influenced by temperature, pH, turbidity, sunlight/UV and salinity (Dean and Mitchell, 2022). However, the interactions between these factors on rates of accumulation and depuration of bacteria in shellfish are poorly understood and difficult to include in models.

Understanding the various environmental drivers and how they interact and impact *E. coli* in shellfish could allow for statistically derived predictive models of faecal pollution to be developed. Predictive models could be used to inform a risk-based active management system providing alternative means of shellfisheries regulation for safe products whilst reducing the potential economic impact of shellfish area closures. Statistical models and

probabilities could be a basis for risk-based management of the shellfish beds by relating existing measurements of possible drivers to measured *E. coli* concentrations in shellfish. These "black-box" models are simple to apply and have been shown to be useful for management. An example of a statistical model for mainly Class A waters was undertaken for two bays in Cornwall utilising a General Linear Model with up to 99% accuracy through incorporation of historical *E. coli* data, rainfall, river flow and, for one bay, solar radiation (Schmidt *et al*, 2018). The high level of accuracy in this case was in part due to the relative simplicity of the outcomes that needed to be predicted, with few *E. coli* results above 230/100g occurring in these areas. Earlier work in the Dart estuary indicated rainfall and river flow as the main drivers of microbial quality of shellfish utilising general statistics (Campos *et al,* 2011).

Statistical models can be further augmented by mechanistic models which use hydrodynamic models to investigate dispersal and persistence of particles (potential pathogens) within river-estuary-coast systems and indicate locations and periods of heightened risk under different scenarios such as combined sewer overflow discharges (Robins *et al,* 2019). Due to the complexity of the system, estuarine and coastal impact modelling requires an integrated approach where catchment, and ocean models are appropriately coupled to predict dispersal plumes or retention zones through the catchment-to-coast continuum. Hydrodynamic models have been applied to predicting dispersal and concentration of *E. coli* in coastal embayments (Dabrowski *et al,* 2014), and in estuarine settings (Garcia Garcia *et al,* 2021). Existing catchment and hydrodynamic models are rarely integrated, and when coupled, models are often not optimised to minimise uncertainties (Robins *et al,* 2018). For instance, the small catchments and estuaries which dominate much of the western UK coast require hourly data to realistically model response to rainfall events, in contrast with larger catchments and estuaries which typically integrate weather conditions over longer time periods of days. A number of hydrodynamic models have been used to predict bathing water quality, such as MOHID, TELEMAC, DELF3D (Viegas *et al,* 2012; Robins *et al,* 2014) and have included regular physical forcing factors such as tides (Kashefipour *et al,* 2005). Hydrological models such as SWAT and INCA-Pathogens typically run at daily resolution, and only a few models such as the coupled CASCADE-TELEMAC model (Robins *et al,* 2018), have so far been optimised for such fine temporal and spatial scale applications.

1.3 Aims and scope of the DASSHH project

The DASSHH study was focused on one case study site, the Camel estuary, to allow detailed investigation of a range of factors that may influence *E. coli* in shellfish. This included collection of new time series environmental data to explore potential refinement of predictive models. The results were intended to inform development of a common approach to development of predictive tools and provide guidance to support wider development of a shellfish Assurance Scheme. To achieve this, the project set out to answer several inter-linked research questions:

- Can models of water movement driven by river flow and tides help our understanding of potential sources of microbial contamination affect shellfish? Coupled hydrodynamic modelling of the Camel estuary, including both tidal and river flow forcing, investigated dispersal of discharges from intermittent wastewater discharges from combined sewage overflows (CSO) This enabled determination of fine scale spatial and temporal patterns of microbial loading over shellfish beds following contamination events, as well as the effect of seasonal variations in river flow.

- Does the bacterial contamination affecting shellfish beds, measured as *E. coli,* originate from animal or human sources? Selected common enteric viruses that are specific to either humans and livestock were monitored in shellfish over a two-year period to investigate seasonal patterns of prevalence.

- Do patterns of risk, as assessed by measurement or modelling of *E. coli* levels in shellfish, coincide with trends in occurrence of specific enteric pathogens of concern for public health? Over a two-year period, shellfish samples were monitoring for levels of norovirus and sapovirus, and correlations with *E. coli* data from the same samples and environmental predictors were investigated.

- Are there differences in variability in *E. coli* data using two laboratory methods approved for use in Official Control monitoring of shellfish influence the precision of predictive models for *E. coli?* Over a one-year period, *E. coli* levels in shellfish collected from seven representative monitoring points, were measured using both the MPN and the pour plate methods, allowing comparison of data obtained using the methods and the performance of predictive models developed using the two data sets.

- Can reliable predictive models for *E. coli* be developed using "off the shelf" existing environmental data and Official Control monitoring records? Relationships between historical *E. coli* data and selected environmental data sets were explored, and the accuracy of predictive models based on these data sets investigated.

- Can predictive models be improved with addition of finer-scale time series for *E. coli* and more detailed environmental data*?* Two-weekly sampling of shellfish provided a time series of *E. coli* data, using both MPN and pour plate methods, as well as high resolution environmental data at each of the shellfish beds. Exploration of relationships between *E. coli* and environmental factors investigated the potential for improvement in the ability to correctly predict levels of shellfish contamination.

1.4 References

Adriaenssens EM, Farkas K, McDonald JE, Jones DL, Allison HE, McCarthy AJ (2021) Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity. Water Research 203: 117568.

Aoki Y, Suto A, Mizuta K, Ahiko T, Osaka K, Matsuzaki Y (2010) Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients. Journal of Hospital Infections 75: 42–46.

Avila R, Horn B, Moriarty E, Hodson R, Moltchanova E (2018) Evaluating statistical model performance in water quality prediction Journal of Environmental Management 206: 910-919.

Baggi F, Demarta A, Peduzzi R (2001) Persistence of viral pathogens and bacteriophages during sewage treatment: Lack of correlation with indicator bacteria. Research Microbiology 152: 743–751.

Barbier EB, Hacker SD, Kennedy C, Koch EW, Stier AC, Silliman BR (2011) The value of estuarine and coastal ecosystem services. Ecological Monographs 81: 169-193.

Beliaeff B, Cochard ML (1995) Applying geostatistics to identification of spatial patterns of fecal contamination in a mussel farming area (Havre de la Vanlee, France). Water Research 29: 1541-1548.

Bellou M, Kokkinos P, Vantarakis A. (2013) Shellfish-borne viral outbreaks: a systematic review. Food and environmental virology. 5: 13-23.

Bougeard M, Le Saux J-C, Perenne N, Baffaut C, Robin M, Pommepuy M (2011) Modeling of *Escherichia coli* fluxes on a catchment and the impact on coastal water and shellfish quality Journal of the American Water Resources Association 47: 350-366.

Bradshaw JK, Snyder BJ, Oladeinde A, Spidle D, Berrang ME, Meinersmann RJ, Oakley B, Sidle RC, Sullivan K, Molina M. (2016) Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. Water Research 15: 498-509.

Campos CJA, Kershaw S, Lee RJ, Morgan OC (2011) Rainfall and river flows are predictors for β-glucuronidase positive *Escherichia coli* accumulation in mussels and Pacific oysters from the Dart Estuary (England). Journal of Water and Health 9: 368-381.

Clements K, Quilliam RS, Jones DL, Wilson J, Malham SK (2015) Spatial and temporal heterogeneity of bacteria across an intertidal shellfish bed: Implications for regulatory monitoring of faecal indicator organisms. Science of the Total Environment 506: 1-9.

Costanza R, d'Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'Neill RV, Paruelo, J, Raskin RG, Sutton P, van den Belt M (1997) The value of the world's ecosystem services and natural capital. Nature 387: 253 – 260.

Dabrowski T, Dore WJ, Lyons K, Nolan GD (2014) Numerical modelling of blue mussel (Mytilus edulis) bacterial contamination Journal of Sea Research 89: 52-63.

Dean K, Mitchell J (2022) Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters. Water Research 211:118051

Droppo IG, Liss SN, Williams D, Nelson TC, Jaskot C, Trapp B (2009) Dynamic existence of waterborne pathogens within river sediment compartments. Implications for water quality regulatory affairs. Environmental Science and Technology 43: 1737–1743.

Devane ML, Moriarty E, Weaver L, Cookson A, Gilpin B (2020) Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research 185: 116204.

Espinosa AC, Arias CF, Sánchez-Colón S, Mazari-Hiriart M (2009) Comparative study of enteric viruses, coliphages and indicator bacteria for evaluating water quality in a tropical high-altitude system. Environmental Health 8: 49.

EU (2017). Community Guide to the Principles of Good Practice for the Microbiological Classification and Monitoring of Bivalve Mollusc Production and Relaying Areas with regard to Regulation 854/2004. https://ec.europa.eu/food/document/download/ee5d7924-08bc-4eb7-94ef-3edf657f51f4_en (accessed 19.05.2022).

Farkas K, Cooper DM, McDonald JE, Malham SK, de Rougemont A, Jones DL (2018) Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Science of the Total Environment 634: 1174–1183.

Frick WE, Ge Z, Zeep RJ (2008) Nowcasting and forecasting concentrations of biological contaminants at beaches: a feasibility and case study. Environmental Science & Technology 42: 4818-4824.

FSA (2016). Active Management Systems in Shellfish. A report by the Food Standards Agency.

Garcia-Garcia LM, Campos CJA, Kershaw S, Younger A, Bacon J (2021) Scenarios of intermittent *E. coli* contamination from sewer overflows to shellfish growing waters: the Dart Estuary case study. Marine Pollution Bulletin 167: 112332.

Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A (2014) Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiological Review 38: 1– 40.

Hassard F, Andrews A, Jones DL, Parsons L, Jones V, Cox B, Daldorph P, Brett H, McDonald JE, Malham SK (2017) Physicochemical factors influence the abundance and culturability of human enteric pathogens and faecal indicator organisms in estuarine, water and sediment. Frontiers in Microbiology 8: 1996.

IFREMER (2014) Enumeration of *Escherichia coli* in live bivalve molluscan shellfish by the direct impedance technique using the BacTrac 4300 series analyser. https://cefaswebsitedev.cefastest.co.uk/media/taulrzth/c-users-ab19-documents-coe-

seafood-safety-generic-protocol-enumeration-of-e-coli-in-bivalve-molluscs-using-impedance.pdf Accessed 19.05.2022.

ISO (2016) ISO 16649-3:2015 Microbiology of the food chain. Horizontal method for the enumeration of beta-glucuronidase-positive *Escherichia coli* detection and most probable number technique using 5-bromo-4-chloro-3-indolyl-β-D-glucuronide. Corrected 2016-12, confirmed 2020. 10 pp. https://www.iso.org/standard/56824.html. Accessed 19.05.2022.

Kashefipour SM, Lin BB, Falconer RA (2005) Neural networks for predicting seawater bacterial levels. Water Management 158: 111–118.

Kay D, Kershaw S, Lee R, Wyer MD, Watkins J, Francis C (2008) Results of field investigations into the impact of intermittent sewage discharges on the microbiological quality of wild mussels (*M. edulis)* in a tidal estuary. Water Research 42: 3033-3046.

Jarvis B, Wilrich C, Wilrich P-T (2010) Reconsideration of the derivation of Most Probable Numbers, their standard deviations, confidence bounds and rarity values. Journal of Applied Microbiology 109: 1660–1667.

Lowther JA, Gustar NE, Powell AL, Hartnell RE, Lees DN, (2012) Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom. Appl. Environmental Microbiology 78: 5812–5817.

Malham SK, Rajko-Nenow P, Howlett E, Tuson KE, Perkins TL, Pallett DW, Wang H, Jago CF, Jones DL, McDonald JE (2014) The interaction of human microbial pathogens, particulate material and nutrients in estuarine environments and their impacts on recreational and shellfish waters. Environmental Science. Processes & Impacts 16: 2145-2155.

Malham SK, Taft H, Cooper D, Ladd C, Seymour S, Robins PE, Jones DL, McDonald JE, Le Vay L, Jones L (2017) Review of current evidence to inform selection of environmental predictors for active management systems in classified shellfish harvesting areas. FSA Project FS103001. Report prepared by Bangor University and the NERC Centre for Ecology and Hydrology.

Nguyen KH, Senay C, Young S, Nayak B, Lobos A, Conrad J, Harwood VJ (2018) Determination of wild animal sources of fecal indicator bacteria by microbial source tracking (MST) influences regulatory decisions. Water Research 144: 424–434.

Oliver DM, Porter KDH, Pachepsky YA, Muirhead RW, Reaney SM, Coffey R, Kay D, Milledge DG, Hong E, Anthony SG, Page R, Bloodworth JW, Mellander P-E, Carbonneau PE, McGrane SJ, Quilliam RS (2016) Predicting microbial water quality with models: Over-arching questions for managing risk in agricultural catchments. Science of the Total Environment 544: 39-47.

Seafish (2021) Review of the application of the Official Control Regulations for shellfish production as they relate to microbial contamination. 103 pp. https://www.seafish.org/document/?id=621ffa11-4943-4698-bdd4-bcf3784bdda8.

Perkins TL, Clements K, Baas JH, Jago CF, Jones DL, Malham SK, McDonald JE. Sediment composition influences spatial variation in the abundance of human pathogen indicator bacteria within an estuarine environment. PloS One 14: 9 e112951.

Potasman I, Paz A, Odeh M (2002) Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective. Clinical Infectious Diseases 35: 921-928.

Pol-Hofstad IE, and Jacobs-Reitsma WF, (2021) Validation of the TBX pour plate method (ISO 16649-2) for the enumeration of Escherichia coli in Live Bivalve Molluscs: renewal study for alignment with EN ISO 16140-2: 2016. RIVM report 2021-0127. DOI 10.21945/RIVM-2021-0127.

Robins PE, Lewis MJ, Cooper D, Coulthard T, Skinner C, Freer J (2018) Improving estuary models by reducing uncertainties derived from river flows. Estuarine, Coastal and Shelf Science 20: 63-73.

Robins PE, Lewis MJ, Simpson JH, Howlett ER, Malham SK (2014) Future variability of solute transport in a macrotidal estuary. Estuarine, Coastal and Shelf Science 151: 88-99.

Robins PE, Farkas K, Cooper D, Malham SK, and Jones DL (2019) Viral dispersal in the coastal zone: A method to quantify water quality risk. Environment international 126: 430-442.

Santo Domingo JW, Edge TA (2010) World Health Organization (WHO). Safe Management of Shellfish and Harvest Waters Edited by G. Rees, K. Pond, D. Kay, J. Bartram and J. Santo Domingo. ISBN: 9781843392255. Published by IWA Publishing, London, UK. pp 51-90.

Schmidt W, Evers-King HL, Campos CJA, Jones DB, Miller PI, Davidson K, Shutler JD (2018) A generic approach for the development of short-term predictions of *Escherichia coli* and biotoxin in shellfish. Aquaculture Environment Interactions 10: 173-185.

Shively DA, Nevers MB, Breitenbach C, Phanikumar MS, Przybyla-Kelly K, Spoljaric AM, Whitman RL (2016) Prototypic automated continuous recreational water quality monitoring of nine Chicago beaches. Journal of Environmental Management 166: 285-293.

Stewart JR, Gast RJ, Fujioka RS, Solo-Gabriele HM, Meschke JS, Amaral-Zettler LA, del Castillo E, Polz MF, Collier TK, Strom MS, Sinigalliano CD, Moeller PDR, Holland AF (2008) The coastal environment and human health: microbial indicators, pathogens, sentinels and reservoirs. Environment and Health 7: S3.

Stidson RT, Gray CA, McPhail CD (2011) Development and use of modelling techniques for real-time bathing water quality predictions. Water and Environment Journal 26: 7-18.

Teplitski M, Wright AC, Lorca G. (2009) Biological approaches for controlling shellfish-associated pathogens. Current opinion in biotechnology 20: 185 - 90.

Tryland I, Myrmel M, Østensvik Ø, Wennverg AC, Robertson LJ (2014) Impact of rainfall on the hygienic quality of blue mussels and water in urban areas in the Inner Oslofjord, Norway. Marine Pollution Bulletin 85: 42-49

Tuholske C, Halpern BS, Blasco G Villasenor JC, Frazier M, Caylor K (2021) Mapping gloval inputs and impacts from human sewage in coastal ecosystems PLoS ONE 16: e0258898.

Viegas C, Neves R, Fernandes R, Mateus M (2012) Modelling tools to support an early alert system for bathing water quality. Environmental Engineering & Management Journal 11: 907-918.

Vinten AJ, Lewis DR, McGechan M, Duncan A, Aitken M, Hill C, Crawford C (2004) Predicting the effect of livestock inputs of *E. coli* on microbiological compliance of bathing waters. Water Research 38: 3215-3224.

Walker DI, Younger A, Stockley L, Baker-Austin C (2018) *Escherichia coli* testing and enumeration in live bivalve shellfish—Present methods and future directions. Food Microbiology 73: 29–38.

Wilksinson J, Kay D, Wyer M, Jenkins A (2006) Processes driving the episodic flux of faecal indicator organisms in streams impacting shellfish harvesting waters Water Research 40: 153-161.

Zimmer-Faust AG, Brown CA, Manderson A (2018) Statistical models of fecal coliform levels in Pacific Northwest estuaries for improved shellfish harvest area closure decision making Marine Pollution Bulletin 137: 360-369.

2 Site Description

2.1 The Camel Estuary

The Camel Estuary is a ria estuarine system in Cornwall, England, that supports a diverse range of habitats historically influenced by the geomorphologic evolution of the estuary (Brew and Gibbard, 2009). As a ria, the estuary is a deep valley, that has been drowned as a results of post-glacial sea level rise. Today the estuary extends 15km upstream from Strepper Point and Trebetherick Point to the tidal limit at Polbrock with a total intertidal area of around 6km^2 (Buck, 1993). A large part of this area is shallow and sandy, with 92% of the area being intertidal flats.

The Camel estuary is a predominantly rural area and historically, supported 4 species of bivalve molluscs for commercial purpose within the mid-estuary, from Rock and Ball Hill to Tregunna (CEFAS, 2015). These species include cockles (*Cerastoderma edule*), blue mussels (*Mytilus edulis*). Pacific oysters (*Magallana* (=*Crassostrea) gigas*) and the peppery furrow shell clam (*Scrobicularia plana*). Currently, blue mussels and Pacific oysters are the only species that are active for commercial harvesting. Within the context of shellfish hygiene, 3 blue mussel beds and 4 Pacific oyster beds are regularly monitored for classification purposes (Table 2.1). All sites have been considered long-term class B since 2014, though some have held the status beyond that.

Table 2.1 Shellfish harvesting areas in the Camel estuary with the species and classification status

Bed Name	Species	Classification	Year Classification Established
Porthilly Rock	Mussels	Long-Term B	2014
Gentle Jane	Mussels	Long-Term B	2010
Ball Hill	Mussels	Long-Term B	2014
Porthilly Rock	Pacific Oysters	Long-Term B	2005
Gentle Jane	Pacific Oysters	Long-Term B	2005
Longlands	Pacific Oysters	Long-Term B	2005
Ball Hill	Pacific Oysters	Long-Term B	2010

The shellfish beds in the estuary are subjected to microbial contamination by both point and diffuse source pollutions, (i.e. from both sewage discharges and agricultural runoff). Since 2009 upgrades have been made on four sewage treatment works (CEFAS, 2015). There are 17

continuous discharges and 58 intermittent discharges that are water company owned as well as 44 private discharges within the estuary. A number of these intermittent discharges are located near the shellfish beds, discharging into areas around Rock and Padstow (see Chapter 3).

2.2 Available Environmental Data

Available environmental data were collected from several sources for the project with a description given in section 6.

2.3 Field sampling

Shellfish and water samples were collected from Camel between 2019 and 2021 with interruption due to the COVID pandemic. Fortnightly sampling of the Camel estuary commenced on the 2nd June 2019 and continued until the 17th March 2020 (Table 2 and Table 3) when the research facilities were temporarily closed due to COVID-19. Two weekly sampling resumed in August 2020 with the last samples collected in August 2021. The samples of Pacific oyster (*Magallana* (=*Crassostrea) gigas*) and blue mussel (*Mytilus edulis*) samples were collected from seven sites, Gentle Jane, Ball Hill, Porthilly Rock and Longland (oysters only), comprising the 4 oyster and 3 mussel representative monitoring points (RMPs), which are the sampling locations for official control sampling, as determined in the sanitary survey for this site (Figure 2.1). Shellfish and water samples were collected from the RMPs for each species on each bed, plus shellfish samples were taken from two other adjacent stations within the beds where the RMPs were located (Figure 2.1). For the investigation of variability of *E. coli* test methods (see Section 4), the three shellfish samples were analysed separately. For time series monitoring of patterns of *E. coli* in shellfish, the three samples were pooled (Section 5). Supplementary samples have also been taken in addition to the fortnightly samples, to investigate events such as rain events or to provide a more detailed investigation of assay variability in the enumeration of *E. coli* in shellfish. These supplementary samples were collected on and ad-hoc basis when rainfall events were forecasted. These events were not predefined. Such samples were collected on 5 occasions and have been limited to sampling the Gentle Jane RMP for both mussels and oysters. A salinity and temperature logger was also deployed, attached to an oyster trestle on the western side of the Gentle Jane oyster bed.

Shellfish samples were analysed for faecal indicator bacteria enumeration via MPN and Pour plate methods with water samples analysed for *E. coli* and nutrient analysis (Table 2.2). Shellfish samples were also analysed for viruses (see Section 4 for detail).

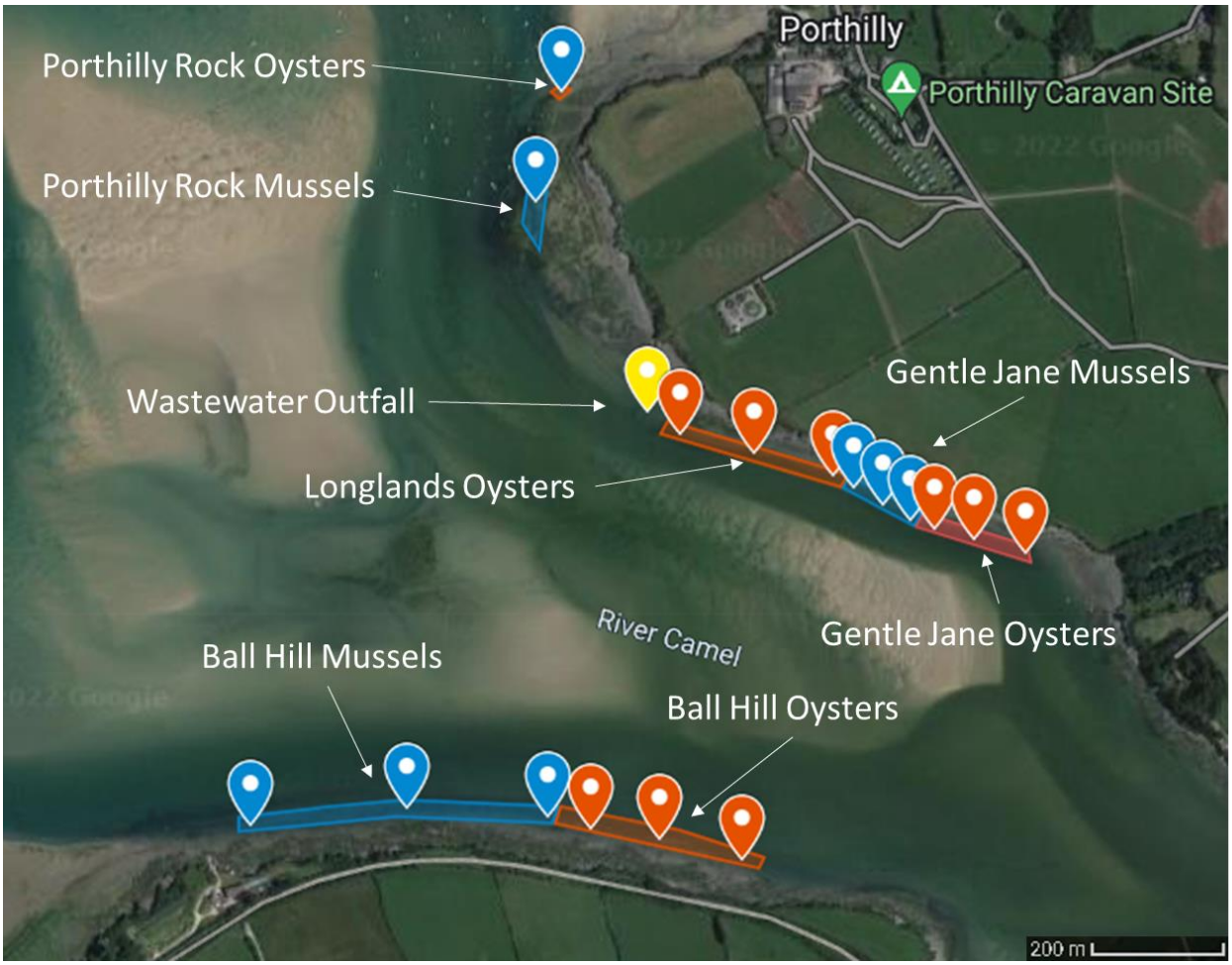


Figure 2.1 Satellite view of the Camel estuary shellfish area and the shellfish beds that were sampled routinely. Samples were collected from three areas within each bed and are marked as points on the map with the exception the Porthilly Rock beds, where a single sample was collected. Shellfish samples were pooled at the point of processing in the laboratory. The mussel (blue) and oyster (orange) shellfish beds are indicated on the map as boxes as well as the location of the wastewater outflow (yellow).

Table 2.2 Summary of all the field sampling undertaken for the DASSHH project. Samples included mussels, oysters, water and sediment taken every two weeks and assayed for bacteria, viruses, nutrients and turbidity.

Sample type	**Assay**	**Notes**
Oysters	MPN	*E. coli*/100g shellfish flesh
Mussels		
Oysters	Pour plate	*E. coli*/100g shellfish flesh
Mussels		
Oysters	qPCR	Norovirus, Adenoviruses Viruses/genome copies (gc/g shellfish digestive tissue)
Mussels		
Water from near Sewage Pipe		Viruses (gc/200µl)
Sediment		Viruses (gc/200µl), *E. coli* and coliforms (cfu/200µl)
Water	Membrane Filtration	*E. coli* and coliforms (cfu/100 ml)
Water	Nutrient Analysis	Total oxidised nitrogen, Nitrite Silicate, Phosphate, Ammonia
Water	Turbidity	NTU
Sediment	Spread plate	*E. coli* cfu/0.2g sediment

2.4 Methods

Methods for the MPN and pour plate techniques are described in section 5. Methods for viral analysis are described in Section 4. Nutrient concentrations were determined using a SEAL Analytical AA3 HR following the methods and procedures provided by SEAL Analytical Ltd. Total oxidised nitrogen, nitrite and silicate methods are based on standard colourimetric methodology (Grasshoff et al, 1983) adapted for segmented flow analysis. Phosphate was determined using the method of Murphy and Riley (1962) and optimised according to Drummond and Maher (1995). Ammonia was measured fluorometrically following the method of Kerouel and Aminot (1997). All runs were monitored for accuracy and precision using Nutrient Standard Solutions from OSIL, Ocean Scientific International Ltd. From August 2021 onwards, analytical accuracy for silicate, phosphate, nitrate, and nitrite was tracked with

KANSO CO LTD reference material nutrient seawater solution (RMNS) lot CH. Nutrient standards were run approximately once a week during analysis.

Sediments samples were processed for *E. coli* and other coliforms using Harlequin CCA Agar. Five grams of sediment were transferred into a 15ml centrifuge tubes before 5ml of 0.1% peptone water was added. Samples were then briefly vortexed to mix the sample with the 0.1% peptone water before they were placed on a shaker running at 275 RPM for 15 minutes. Harlequin CCA Agar plates were then inoculated with 200µl of the 1:1 dilution of the sample before being placed in an incubator at 37 °C for 24 hours. CCA agar allows for the enumeration of both E. coli and coliforms based on the colouration of the coliforms.

Water samples were processed for *E. coli* and other coliform enumeration according to ISO 9308-1 2014/AMD 1:2016 (ISO, 2016), a culture-based membrane filtration method. Briefly, water samples were filtered using a vacuum pump onto a 47mm, 0.45µm pore size cellulose nitrate filter. Two volumes of the sample were filtered to reduce the number of samples that returned uncountable colony recoveries. The turbidity of the water samples was analysed using an Oakton T-100 Turbidity meter.

2.5 References

Brew DS, Gibberd BB (2009) Geomorphological change in the Camel estuary, Cornwall, UK. Geoscience in South-West England 12: 95-100.

Buck AL (1993) An inventory of UK estuaries. Volume 2. South-west Britain. Joint Nature Conservation Committee, Peterborough. 178pp.

CEFAS (2015) Review of the Camel 2009 Sanitary Survey. CEFAS report on behalf of the Food Standards Agency, to demonstrate compliance with the requirements for classification of bivalve mollusc production areas in England and Wales under EC Regulation No. 854/2004.

Drummond L, Maher W (1995) Re-examination of the optimum conditions for the analysis of phosphate. Analytical Chimica Acta 302: 69-74.

Grasshoff K, Ehrhardt M, Kremling K (1983) Methods of Seawater Analysis. Verlag Chemie, New York. 419pp.

ISO (2016) ISO 9308-1:2014/AMD2017 Water quality. Enumeration of *Escherichia coli* and coliform bacteria Membrane filtration method for waters with low bacterial background flora. https://www.iso.org/standard/70094.html. Accessed 19.05.2022.

Kerouel R, Aminot A (1997) Fluorometric determination of ammonia in sea and estuarine waters by direct segmented flow analysis. Marine Chemistry 57: 265-275

Murphy J, Riley JP (1962) A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta 27:31-36.

3 Hydrodynamic Modelling

3.1 Summary

Intermittent discharges from CSOs into the Camel estuary are known to be a possible source of contamination to the mussel and oyster beds within the estuary. A depth-averaged hydrodynamic model was used to simulate spatiotemporal dispersal of potential microbial contamination through the river-estuary-coast over a two-year period. Available CSO operation data did not include flow volume or concentration of bacteria, so simulations were of relative bacterial dispersal patterns from all CSOs. Most discharge events occurred during winter, following high rainfall/river flow events, with discharges being rare but important during the summer months. Some CSO discharges were not associated with river flow events (daily river discharge was found to explain ~58% of the variance of CSO spills) and it is unclear what triggered them, although wastewater system failure has been known to trigger CSOs. The highest seasonal-mean bacterial tracer concentrations were found in the autumn and winter months, being much lower during the spring and especially the summer. In spring and summer, short-lived hotspots of maximum concentrations did occur when river discharge was low and dilution therefore was minimal. From these results, specific CSOs that pose a relatively high risk to shellfisheries could be identified, and further work is required to understand the potential microbial dispersal following specific environmental events (including extreme events not captured within our two-year period) from these CSOs in isolation.

3.2 Introduction

Secondary wastewater treatment processes, such as activated sludge and UV disinfection, are typically relatively inefficient at removing microbes including bacteria and viruses (Kitajima *et al*, 2014) and hence *E. coli* (FIO) and pathogenic microbes can enter the aquatic environment causing a potential risk to food sources and human health. Intermittent discharges from wastewater combined sewer overflows (CSOs) along with storm overflows (SOs) are known to be a widespread problem effecting rivers and estuaries globally (e.g. Hassard *et al*, 2017; Hata *et al*, 2014). However, few studies have attempted to quantify and map the contamination risks associated with multiple interacting CSO discharges (but see García-García *et al,* 2021). A typical estuary may be impacted by several CSOs discharging at different locations and operating at irregular time intervals and for periods of seconds to hours. These CSO discharges interact with river flows, tidal dynamics, residual coastal currents, and complex bathymetry. The resulting spatiotemporal patterns of contamination of estuary waters can vary over spatial scales of metres, and temporal scales of minutes but also seasonally/inter-annually (Bashawri *et al,* 2020; Robins *et al*, 2019), being contingent on estuary size and geomorphic type (Robins *et al,* 2018). Estuaries can have residence times for contaminated water of days or months; additionally, estuarine circulation patterns can concentrate contaminated water at certain hotspots (e.g., Brown *et al,* 1991; Robins *et al*,

2012). It is difficult to unpick these patterns by collecting and analysing field samples due to practical restrictions of sampling data that would be needed to resolve the system complexity. Hence, numerical modelling can be used to overcome some of the limitations associated with field data (e.g., Robins *et al,* 2019).

Intermittent discharges from CSOs into the Camel estuary are known to be a possible source of contamination to the mussel and oyster beds within the estuary (CEFAS Sanitary Survey, Camel, 2015). Yet a detailed understanding of the fate of contaminated water that enters the estuary is unclear. The aim, therefore, was to study the potential impact of intermittent CSO discharges on water quality around shellfish sites in the Camel estuary. A depth-averaged hydrodynamic model was used that could simulate the spatiotemporal dispersal of potential microbial contamination through the river-estuary-coast continuum. The hydrodynamic model simulated a two-year period, forced with tidal boundary conditions together with historical river flow data and microbial loading from nine permitted CSO sites that feed into the adjoining rivers of the Camel estuary. The residence concentration of microbial loading in the Camel estuary following contamination events from the CSOs, and the sensitivity of microbial concentrations to hydrodynamic conditions and microbial decay rate, were examined.

3.3 Methods

3.3.1 Model Setup

A hydrodynamic ocean model (Telemac Modelling System V7.2; www.opentelemac.org) was used to simulate the dispersal of tracers within the Camel estuary and surrounding coast. Telemac computes the depth-averaged shallow water Saint-Venant equations of momentum and continuity, derived from the Navier-Stokes equations (Hervouet, 2007), on an unstructured triangular mesh. The mesh is at a very high density in the estuary and around shellfish beds: ~10 m and increases out to sea up to ~800 m (Figure 3.1). The mesh is mapped on to observational bathymetry data that was obtained from two sources: (1) the UK Government's ADMIRALTY Marine Data Portal (www.admiralty.co.uk/ukho) at 200 m spatial resolution (EDINA, 2008); (2) LIDAR data in intertidal regions at 10 m resolution (freely available from the UK Environment Agency and Natural Resources Wales), surveyed in 2018.

The Telemac model is well suited to vertically mixed coastal and estuarine applications, such as the Camel, since the mesh is optimised to adequately resolve near-coast dynamic features that are important for pollutant dispersal. The N-TF advection scheme (Distributive scheme N adapted for tidal flats) was used which has the capability to deal with dry-zones within the model domain whilst preserving monotonicity and maintaining tracer mass conservation (Hervouet, et al, 2015). The classical k-ε turbulence model has been adapted into vertically averaged form to include additional dispersion terms (Rastogi and Rodi, 1978); a constant internal friction coefficient of 3×10^{-2} m was implemented in Nikuradse's law of bottom friction (Hervouet, 2007). Turbulent viscosity was set to a constant with the overall viscosity (molecular + turbulent) coefficient equal to 10^6.

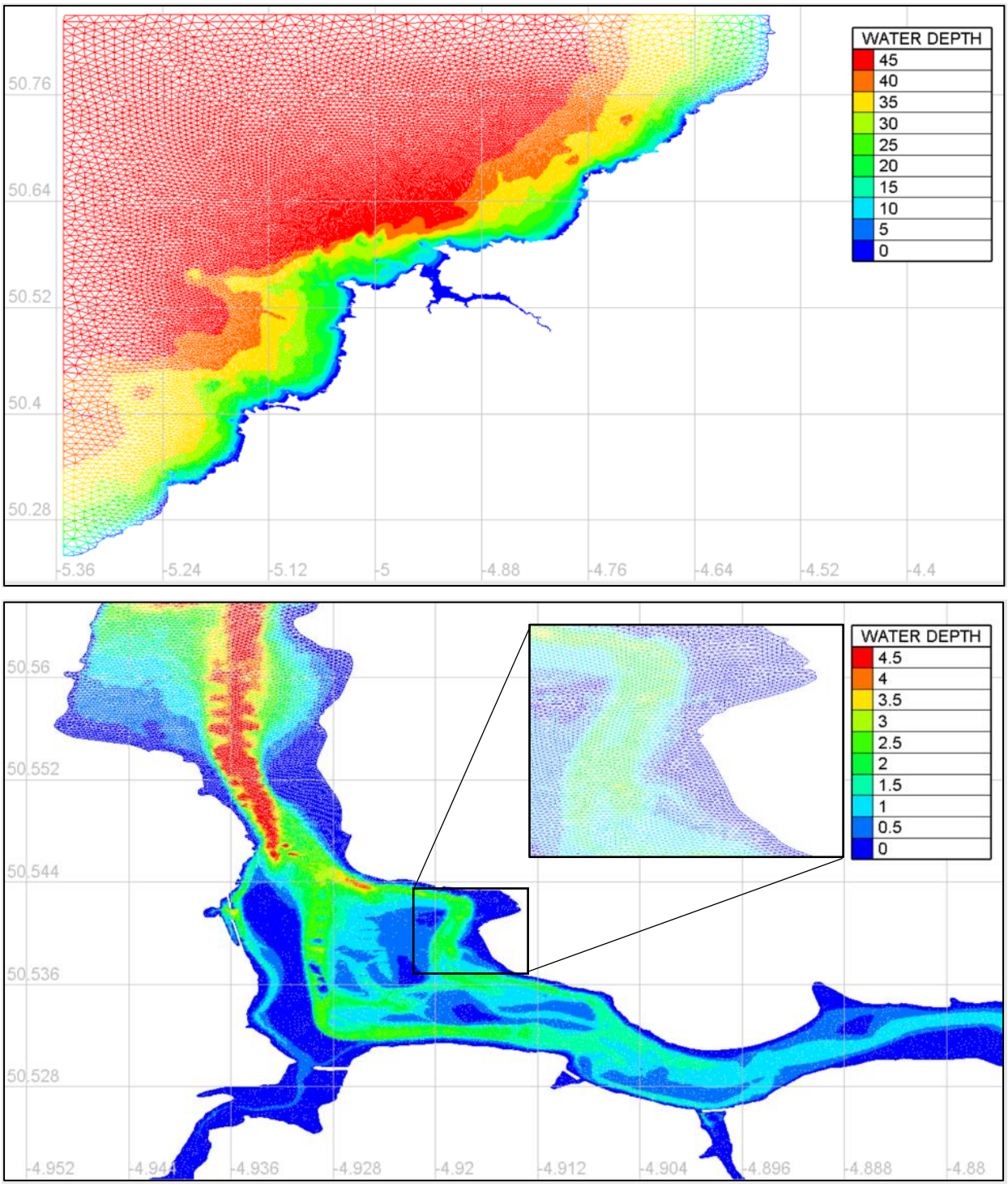


Figure 3.1 Top panel shows the extent of model domain and mesh; colours indicate the mean water depth (m) across the modelled period (grid spacing of ~7.5 km). Bottom panel shows the estuary at a finer scale (grid spacing of ~500 m), note the higher resolution mesh in the estuary /river channel, The Porthilly Rock area is zoomed in to give clearer indication of the mesh resolution used around the shellfish beds.

The hydrodynamic model was initially run (spun-up) for one month (December 2018) to create a steady-state horizontal salinity balance and to prime the estuary with tracers from CSO discharges that occurred prior to the modelling period. The salinity and tracer distributions from the spin-up simulation was used as initial conditions for subsequent two-year simulation. Comprehensive validation procedures have previously been conducted for hydrodynamics and salinity intrusion (see Robins et al, 2014) for the same modelling approach for the Conwy estuary which gives a reasonable degree of confidence that the model results for the Camel should also validate well given the similar characteristics of the estuaries.

Tidal forcing comprised the 13 primary harmonic constituents (M_2, S_2, N_2, K_2, K_1, O_1, P_1, Q_1, M_4, MS_4, MN_4, M_f and M_m), derived from the Topex/Poseidon TPXO global tidal database on a structured grid of 0.25° resolution (Egbert et al, 1994). Both surface elevation change and the deduced horizontal velocities were used at the boundaries. Hourly river discharge data were used to simulate river flow inputs from the River Camel and its four largest tributaries within the model domain, the rivers: Allen, Polmorfa, Amble and Petherick. Discharge data used for the Camel was obtained from the National River Flow Archive (NRFA) at the Denby Bodmin Dunmere gauge – 49007 (https://nrfa.ceh.ac.uk/). Hourly discharge time-series for the four (un-gauged) tributaries was approximated from the River Camel timeseries by using a scale factor for each tributary that was derived from relative catchment areas.

3.3.2 CSO Discharge

CSO discharges were simulated based on binary spill data (flow/no-flow condition) from nine permitted intermittent discharge sites around the Camel estuary (Table 3.1, Figure 3.2) provided by Southwest Water (SWW). Start and end time data from each CSO gauge were used to create a continuous two-year time-series for each site at minute resolution; no-spill periods and spill periods (defined per minute) were assigned logical values of zero and one respectively (Figure 3.3). Therefore, during the two-year simulation, tracer concentrations were periodically discharged (independently) from each of the nine CSO gauges according to the periods of spill/no-spill data. The 'Sum of Spills' index plotted was calculated by summing the number of CSOs spilling at each time point. No data were available for CSOs quantifying the volume of discharge when spilling occurred nor were data available quantifying the concentration of microbial contamination in discharged sewerage. It was assumed that when spilling each CSO discharged 1 litre/s of effluent (tracer) of unit concentration. This allows us to directly compare the hydrodynamic controls on spatio-temporal patterns in tracer concentrations. All downstream tracer concentrations in model outputs are therefore expressed as a percentage of CSO effluent concentration.

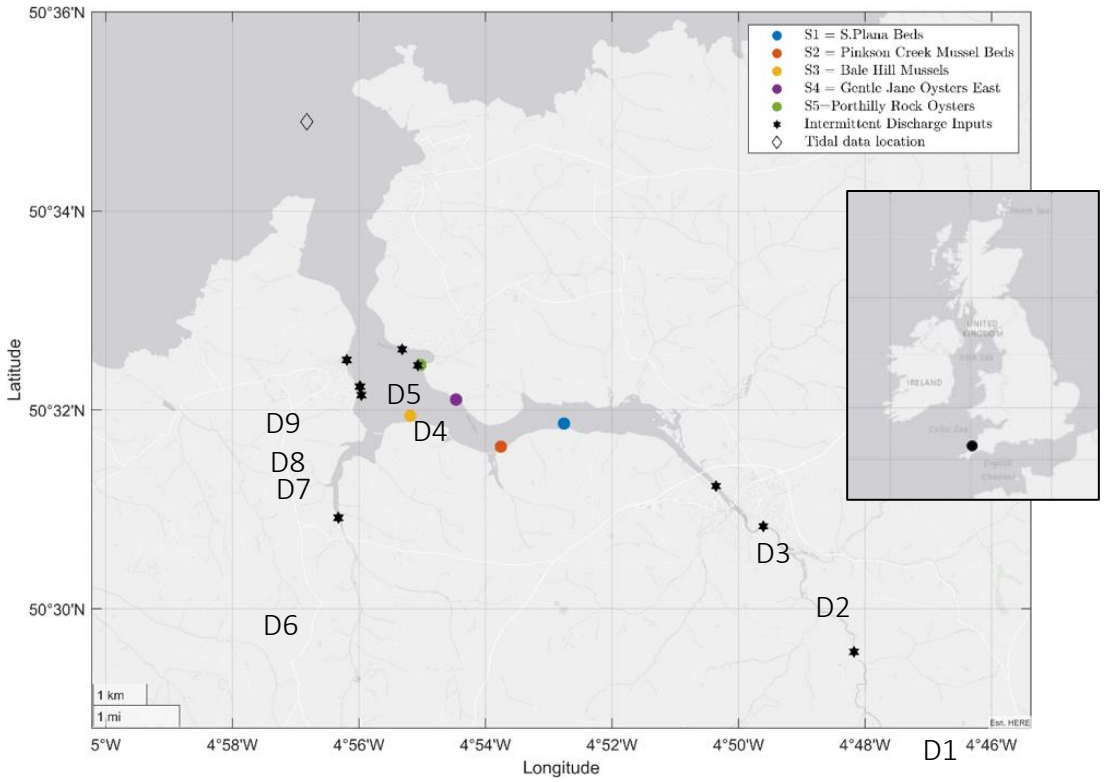


Figure 3.2 Map showing shellfish beds and intermittent discharge input locations. For details of discharge input locations see Table 3.1. Note: the location of Nanstallon pumping station emergency overflow (PSEO) (D6) shown is the tracer input location in the model (at the models upstream limit), Nanstallon PSEO is in fact located approximately 4.7 km further upstream (National Grid Reference SX0350067300).

Table 3.1 Intermittent discharges included in study. Time spilling is the percentage of the two-year study period when the spill gauge indicated discharge.

	Name (as given in CEFAS (2015)	**National Grid Reference**	**Time spilling (%)**
D1	Nanstallon Photovoltaic Solar Electrochemical Oxidation (PSEO) site	SX0350067300	13.215
D2	Egloshayle Pumping Station	SW9970972074	2.828
D3	Wadebridge Pumping Station	SW9885072720	2.923
D4	Porthilly Cove Pumping Station	SW9373075460	0.088
D5	Rock Pumping Station	SW9307075600	0.005
D6	Little Petherick Sewerage Treatment Works	SW9182072580	1.494
D7	Moyles Rd Combined Sewerage Overflow (CSO)	SW9225074780	0.049
D8	Padstow Foreshore Pumping Station	SW9224074920	0.211
D9	Padstow Harbour Pumping Station	SW9201075450	0.826

3.3.3 Simulations

The model was run over a two-year period from 1st Jan 2018 through to 31st Dec 2019 using a 1 s time-step and 12-minute resolution outputs were recorded. To investigate the sensitivity of the results to microbial decay rates, simulations were run using a first order kinetic decay model. Two different rates of exponential decay were used; T_{90} = 24 hrs and T_{90} = 672hrs (28 days), where T_{90} is defined as the time taken for 90% of the tracer to decay (die-off), or alternatively as the time taken for tracer concentration to decrease by one log unit. The decay rate of bacteria and viruses depends on their species as well as the physicochemical environmental conditions, i.e., temperature, salinity, turbidity UV-exposure (Murphy, 2017). Additionally, it is probable that decay rates for both bacteria and viruses would vary during the year; likely higher levels of decay during summer when UV and seawater temperature is higher. However, the model was not parameterised for variable decay rates during the year. Therefore, contrasting decay rates at opposite ends of the spectrum associated with bacteria and viruses likely to be present in the Camel were selected so that the sensitivity of the model results to a broad range of expected decay rates could be assessed. Selected rates were used from research demonstrating that increasing salt (calcium, magnesium, potassium and sodium) concentrations in the freshwater range can increase the survival rates of *E. coli* (DeVilbiss et al 2021).

3.3.4 Post Simulation Analysis

Times series of tracer concentrations were extracted at Sites 1-5 (S1-S5) as indicated in Figure 3.2 that are associated with (currently non-operational) peppery furrow clam (*Scrobicularia plana*) beds (S1), and also with operational mussel (*Mytilus* spp.) beds and Pacific oyster (*Magalla* (= *Crassostrea) gigas*) trestles (S2-S5). Seasonal risk maps were produced by calculating the mean concentration of the tracers at each node within the model domain in the Camel estuary for each meteorological season simulated (Spring; 1st March - 31st May, Summer; 1st Jun - 31st Aug, Autumn; 1st Sep - 30th Nov, Winter; 1st Dec - 28th Feb). Seasonal data for the two years were aggregated).

3.4 Results

3.4.1 Hydrodynamics and CSO Spillage

CSO events can occur following heavy/intense rainfall which can place pressure on the capacity of wastewater systems, whereby untreated sewerage bypasses the wastewater system and discharges directly into the river network which would likely be in spate because of the heavy rain. CSOs can also operate as a result of wastewater system breakdown/failure, and under this scenario the CSO and river flows may not necessarily be high. Hydrodynamic modelling of the Camel estuary over a two-year period that represents realistic flow/tide regimes and associated CSO spillages has been performed and the results are discussed

below. This section includes analysis and plots which have been undertaken to investigate whether river discharge could be a useful predictor of CSO events.

The sum of CSO spills for the River Camel for the study period is presented in Figure 3.3 and summarised by meteorological season in Table 3.2. The winters and autumns of the study period were characterised with relatively high base flow (i.e., highest mean discharge) of 10-12 m^3/s and frequent high magnitude flow events of 40-52 m^3/s. Overall base-flow declined to near-zero over the course of the spring, although occasional extreme flow events (>10 m^3/s) did occur. Base-flow was near zero during the summers and large flow events were infrequent and less extreme than at other times of year. The highest river discharge was recorded on 15th March 2018, at 52 m^3/s - this being a >99th percentile discharge value over the 30-year period 1984-2013 (maximum recorded discharge during this period was 150 m^3/s) (Lyddon *et al*, 2022).

Table 3.2 Summary of river Camel discharge data by meteorological season

	Spring	***Summer***	***Autumn***	***Winter***	***Overall***
Mean Discharge (m^3s^{-1})	5.993	1.478	8.061	12.076	6.9025
Max Discharge (m^3s^{-1})	52.37	9.2	45.57	40.76	52.37
Min Discharge (m^3s^{-1})	1.41	0.74	0.83	3.85	0.74
Max Daily Mean Discharge (m^3s^{-1})	32.04	6.58	35.46	29.84	35.46
Min Daily Mean Discharge (m^3s^{-1})	1.46	0.76	0.86	3.99	0.76

Time series of spill events for each of the CSOs are plotted in Figure 3.3 along with an aggregated time series 'sum of spills' which is number of CSO gauges spilling at any given time. Figure 3.3 shows that there were occasional CSO discharges that did not appear to be associated with any river flow events. When comparing the 'sum of spills' and river discharge time series, it was evident that, overall, the number of CSO events occurring in the catchment at a given time are associated with the discharge but the daily mean discharge alone was found to only explain 58.2% of variance in the 'sum of spills' (Figure 3.4). When the mean daily discharge exceeded a threshold of 22.6 m^3s^{-1} spillage occurred from at least one CSO but on some occasions spillage occurred at up to four CSOs when river discharge was below this threshold. The fact that CSO events can occur due to wastewater system breakdown/failure may be one reason why river discharge does not appear to be a strong predictor.

To explore the relationship between river discharge and spillage further at each CSO, two metrics of river discharge: maximum daily discharge and within day increase in discharge (i.e. daily max discharge minus the minimum discharge earlier that day), were plotted against the proportion of the day the CSO discharged for, and a linear regression was conducted between these variables for each CSO (Figure 3.5). For CSO D4-D9, spillages, there was no clear relationship with the maximum daily river discharge, and the within day increase in river

discharge was also found to be a poor predictor of CSO spillage. At sites D1-D3, D6, D8 and D9 spillage occurred on multiple days when the river discharge was declining.

Figures 3.6 and 3.7, when comparing the *T_{90} = 24 hrs* and *T_{90} = 28 days* plots, demonstrated that the tracer concentrations in the estuary are sensitive to tracer decay. The $T_{90} = 28\ days$$\mathrm{T}_{90} = 24\ \mathrm{hrs}$ tracer quickly dropped to negligible levels in between times of CSO spill events (Figure 3.7). Model results suggested that depending on the river discharge behaviour and tidal state, advection times for the tracer to travel from the upstream model limit (tidal limit) to the estuary mouth (13.6 km downstream) ranged from between 13.5 hrs up to 4 days. Thus, tracers with *T_{90} = 24 hrs* decay discharged from Nanstallon (D1), and to a lesser extent Egloshayle (D2) and Wadebridge (D3), reduced significantly in concentration$T_{90} = 28\ days$ did not significantly reduce tracer concentrations, as the majority of tracer discharge reached the estuary mouth within 4 days even during low river discharge conditions. It is worth noting that Figures 3.6 and 3.7 show seasonal-mean tracer concentrations throughout the estuary (spatially), whereas in reality the tidal flats dry out during each low tide – with several interesting potential physiochemical processes (e.g., pooling of water, heating) that are not captured by the modelling. Because the model runs for a long period (two years), these effects are assumed to be minimal.

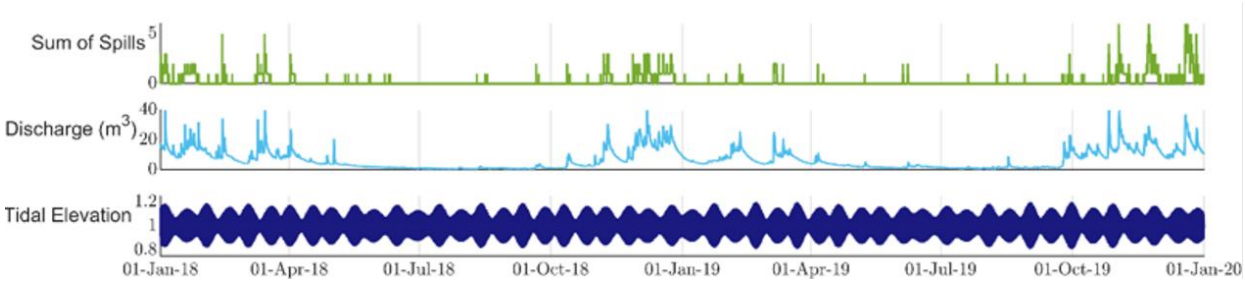


Figure 3.3 Time series for the study showing the sum of spills for the CSO gauges spilling at any given time (top), the Camel river discharge (middle) and the tidal variation (depth/mean depth for the study period at the estuary mouth) relative to the mean tide level (bottom).

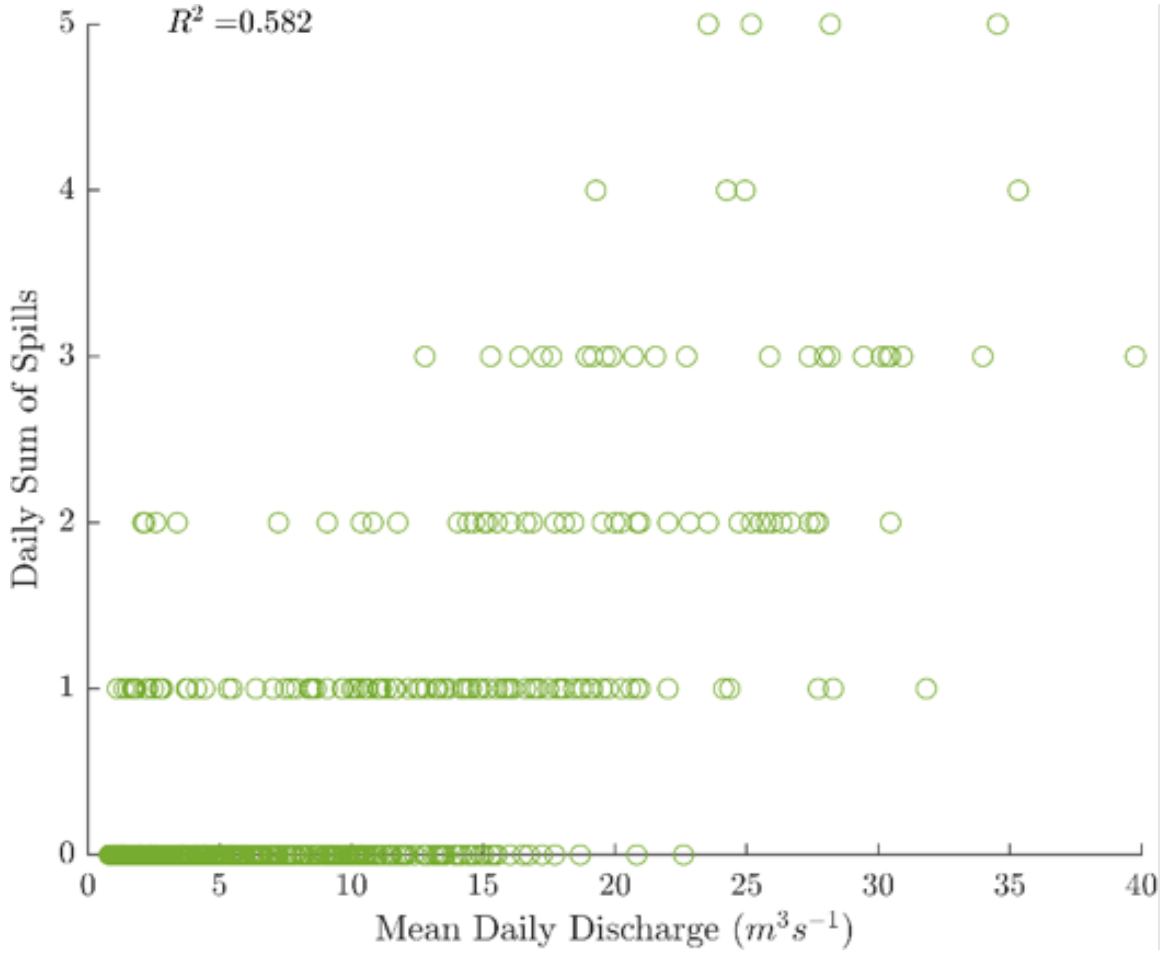


Figure 3.4 Scatter plot of the sum of spills (number of CSO spilling) on a given day against the mean daily river discharge. Note a maximum of five out of nine CSOs spilled on any one day.

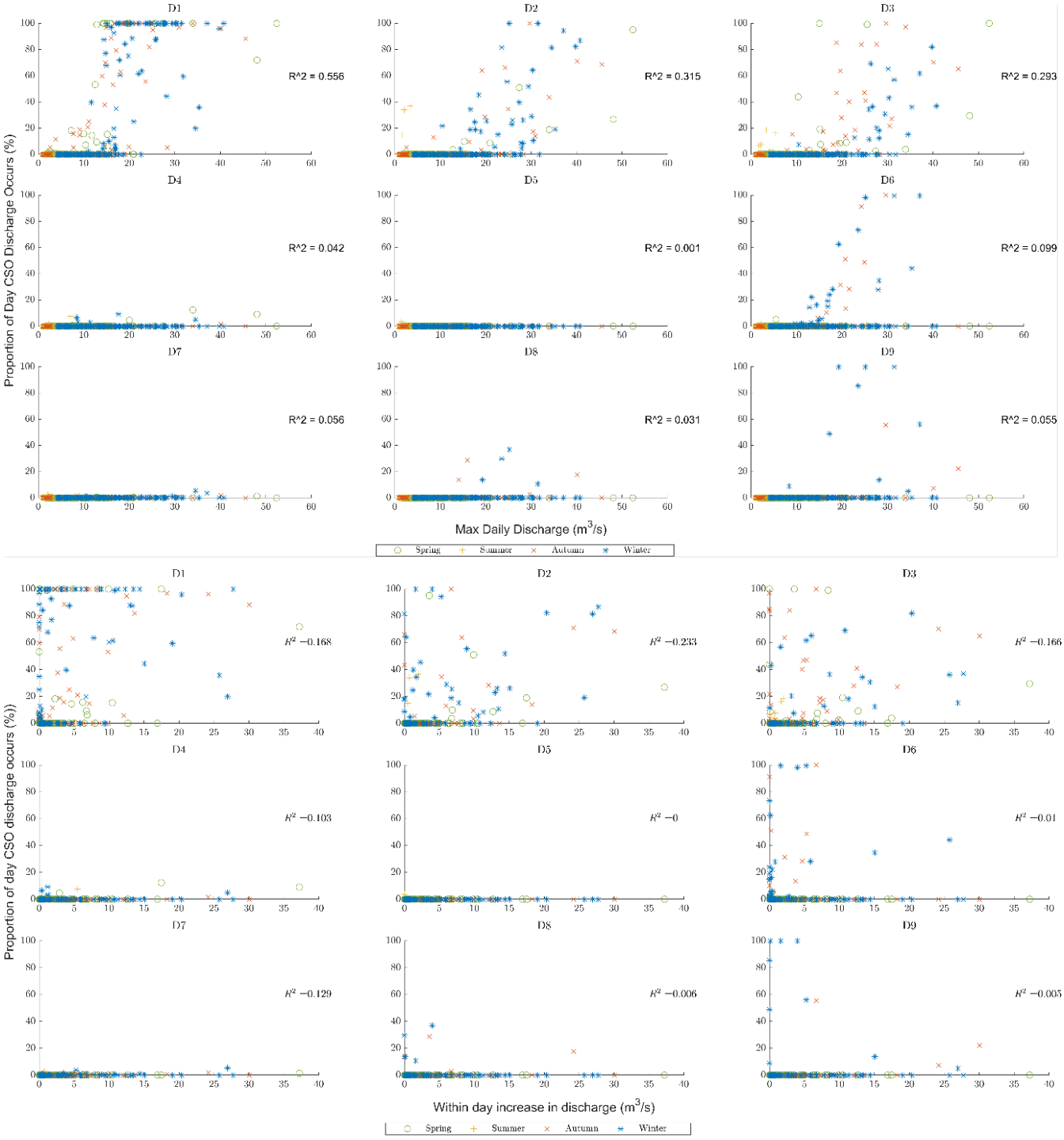


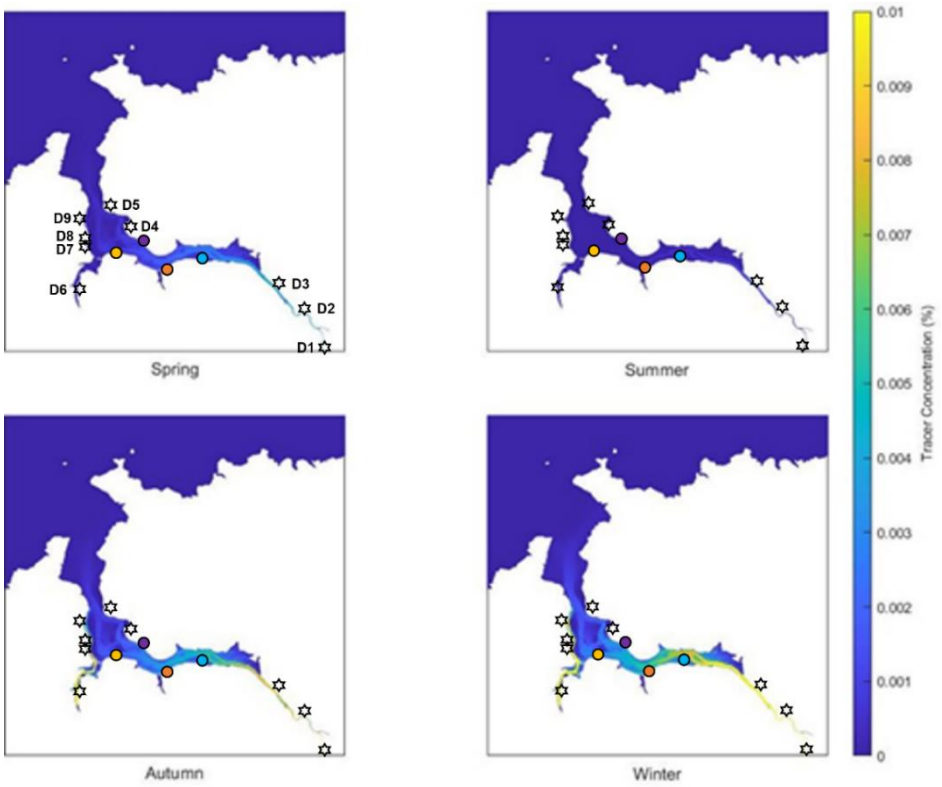
Figure 3.5 Scatter plots of the proportion of days that each CSO discharge on any given day against (top) the maximum river discharge on that day (Q_{dmax}) and (bottom) the within day increase in river discharge (Q_{dinc}), i.e., the overall increase in river discharge that occurred on each day from the daily min flow to the daily maximum flow where the daily minimum flow was always taken as the minimum flow proceeding the maximum flow; where the maximum river flow occurred at the beginning of the day the within day increase was recorded as zero. Circles are coloured by season. D1-D9 are locations of intermittent discharges given in Table 3.1

Results of maximum tracer concentrations for each season are presented in Figure 3.8. Somewhat in contrast with the mean tracer concentrations (Figure 3.6), maximum tracer concentrations were not greatest during autumn and winter. Indeed, when discharge was low during summer, maximum tracer concentrations in the main river channel as far downstream as the Gentle Jane Oyster beds (S4) were highest. In addition, during spring and summer, hotspots of maximum concentrations occurred around the Porthilly Rock Oysters site (S5) due to CSO spillage from Porthilly Cove Pumping Station (D4) when river discharge was low and dilution therefore was minimal. The time series presented in Figure 3.7 corroborates that tracer concentrations can reach high levels during summer at the shellfish sites, and indeed on occasion exceeded the levels present at other times of year. However, Figure 3.7 also highlights that high concentration events during summer tend to be short lived, as CSO discharges tend to not be sustained (see Figure 3.3). However, it is also probable that microbial decay may be higher during the summer than at other times, due to increased UV and temperature. So, whilst simulated maximum concentrations may have been higher during the summer, it is also possible that decay/inactivation may be higher.

With respect to maximum concentrations shown in Figure 3.8, the modelling results suggest that the shellfish sites in the Camel are well located (in the mid-estuary downstream of D3 upstream of D4) to reduce the risk of exposure from high concentration events. For example, its apparent from the spring-, autumn- and winter-averaged 24-hr tracer decay simulations (Figure 3.8a) that maximal concentrations were generally higher (~0.05%) upstream of the shellfish sites, associated with spillages from D1-D3, but concentrations were also up to 0.05% higher in the lower estuary (particularly the western channel) as a result of CSO spillages from sites D4-D9. This pattern was also seen to some extent in the autumn- and winter-averaged simulations with 28-day tracer decay in Figure 3.8b. Although the summer-averaged simulations (both 24-hr and 28-day tracer decay) showed high maximal concentrations reaching the eastern shellfish sites that are most upstream or proximal to the CSOs (D1-D3), and hence these shellfish sites are most at risk from short but intense contamination events.

The modelling results presented enable spatial and temporal comparisons to be made across the Camel Estuary of tracer concentrations arising from tracers with different decay rates under the assumption that all CSO's discharge the same volume and concertation of effluent. For this reason, the results have been presented qualitatively rather than quantitatively. It is not possible to infer directly from the results the absolute levels of contamination in the estuary arising from CSO spillage. To model this would require information on concentration and volume of effluent discharged from each CSO, which is not recorded for the CSOs discharging in the Camel.

***a*) $T_{90} = 24\ hrs$**



***b*) $T_{90} = 28\ days$**

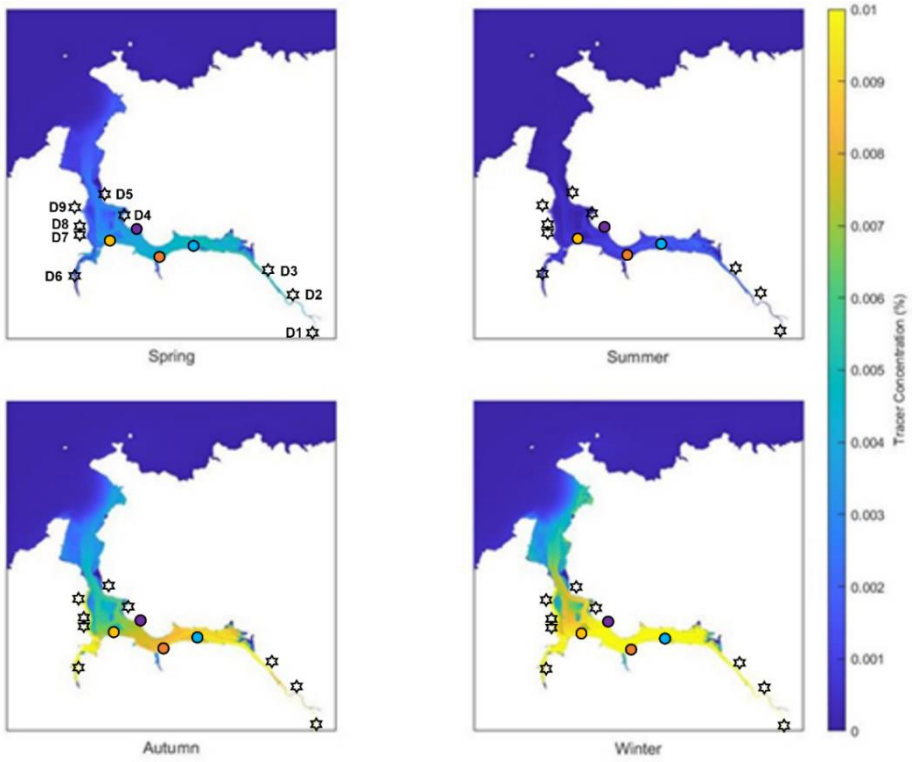
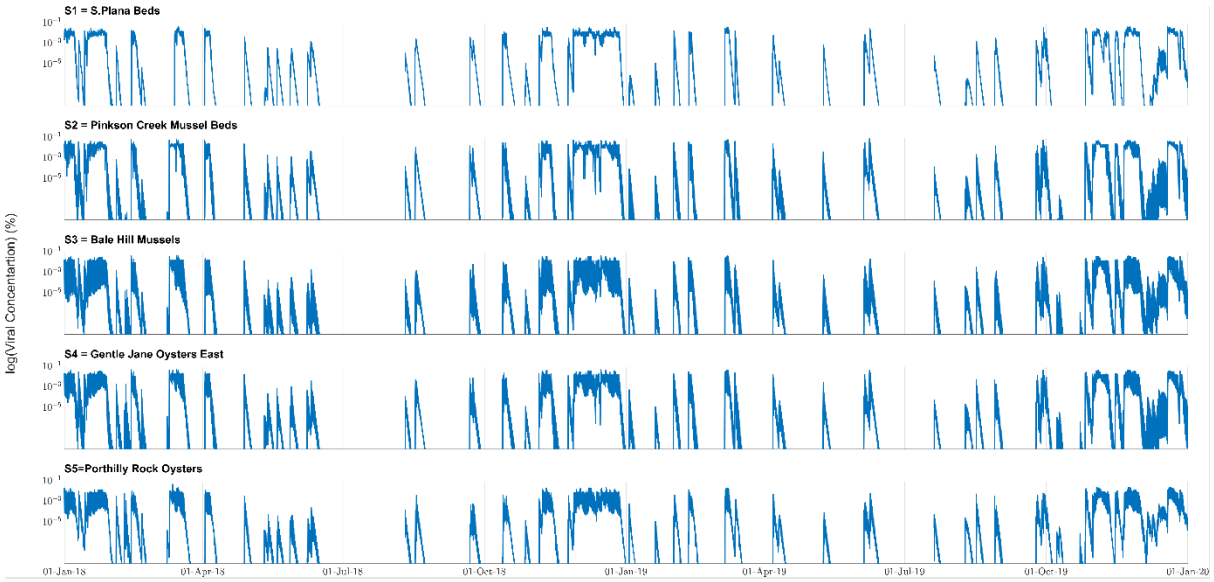


Figure 3.6 Seasonal-mean tracer concentration in the Camel river/estuary for each season simulated using decay constants; a) $T_{90} = 24\ hours$ and b) $T_{90} = 28\ days$.

***a*) $T_{90} = 24\ hrs$**



***b*) $T_{90} = 28\ days$**

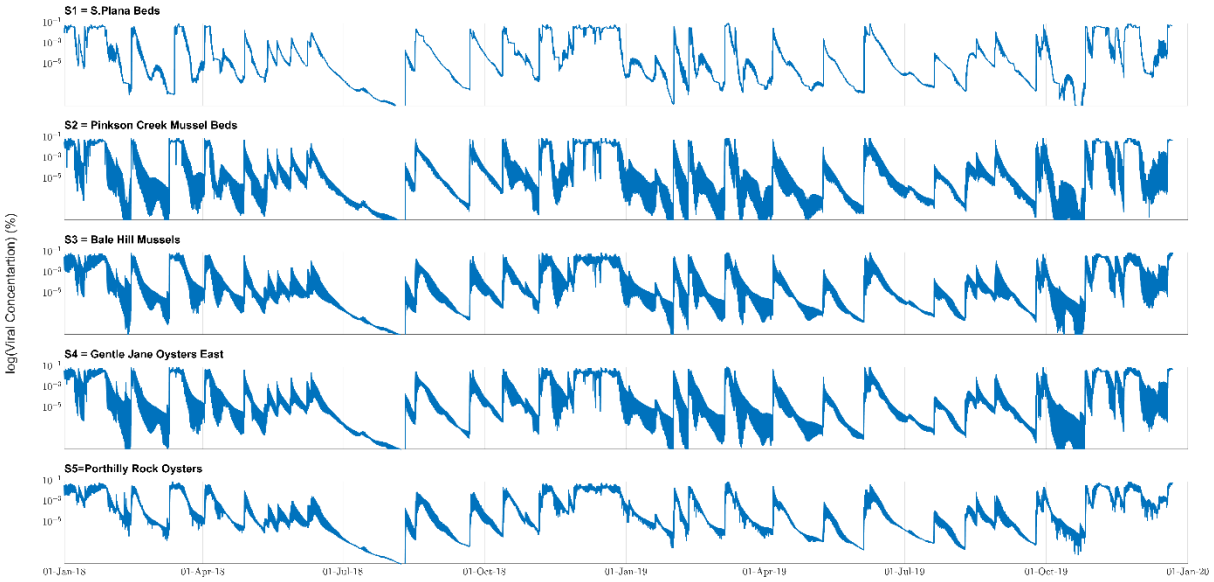
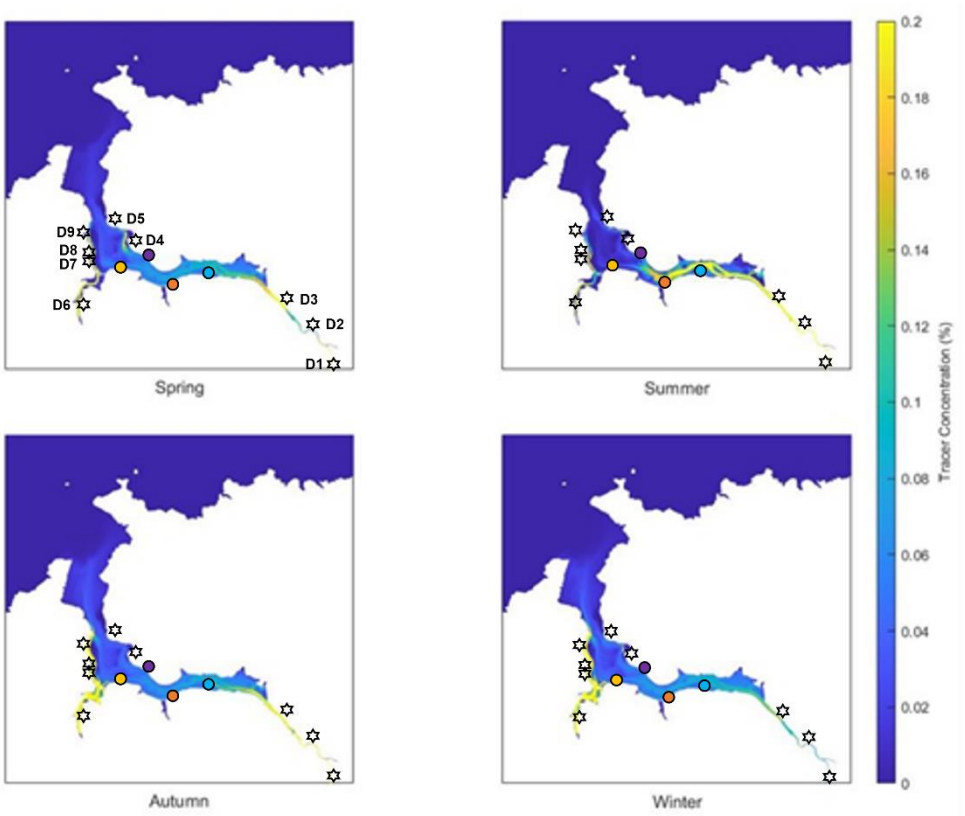


Figure 3.7 Time series plots for tracer concentration at shellfish sites simulated using decay constants; a) $T_{90} = 24\ hours$ and $b)\ T_{90} = 28\ days$.

***a*) $T_{90} = 24\ hrs$**



***b*) $T_{90} = 28\ days$**

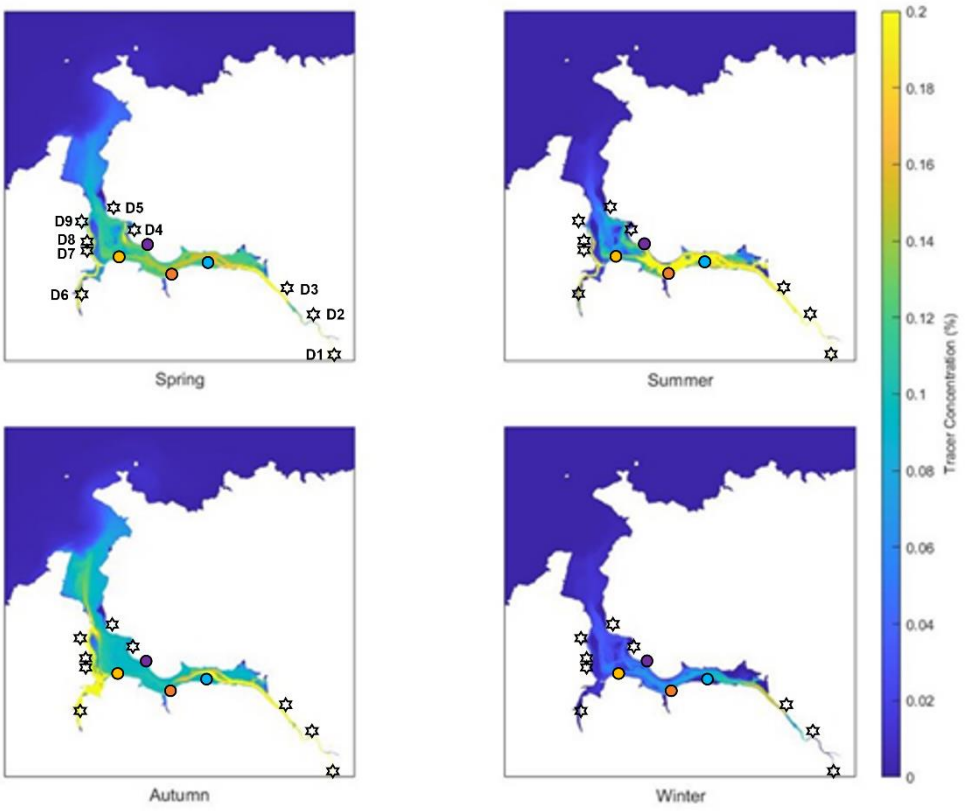


Figure 3.8 Seasonal-maximum tracer concentration in the Camel river/estuary for each season simulated using decay constants; a) $T_{90} = 24\ hours$ and b) $T_{90} = 28\ days$.

3.5 Discussion

Hydrodynamic modelling has been applied to the Camel estuary to simulate potential effluent dispersal from nine intermittent CSO discharges entering both the upper and lower estuary (upstream and downstream of the shellfisheries sites). The CSO discharge events occurred during a period of two years (2018-2019), with most discharges occurring during winter, following high rainfall/river flow events. Discharges were being rare but important during the summer months. Some CSO discharges were not associated with any river flow events and it is unclear what triggered them, although it is known that CSOs can operate as a result of wastewater system breakdown/failure, and under this scenario the CSO and river flows may not necessarily be high. The intricate coastline and bathymetric features (channels and sand flats) were well-resolved in the model, as was the tidal propagation and river behaviour (hourly flow data). Accurately resolving these boundary conditions has been shown to be crucial for modelling water quality in shallow, macro-tidal estuaries such as the Camel that are common throughout the UK (Regnier *et al*, 1998; Robins *et al*, 2018).

Over several tidal cycles (of the order of weeks) the salinity at any point along an estuary is often assumed to be constant implying that the freshwater flow, which acts to freshen the estuary and carry bacteria and viruses seawards, is balanced by diffusive mechanisms that carry salt landwards (Smith, 1997). This principle of saline intrusion/estuarine recovery acts to promote the retention of fluvial-sourced viruses within estuaries. It should be noted, however, that in many systems a position of steady-state salt balance/virus retention is rarely achieved due to high fluvial discharge/viral loading events in combination with strong and variable tidal dynamics. Therefore, the longer that bacteria and viruses are prevalent within the system (long decay rate), the more likely it is that they will not be retained within the estuary (Robins *et al,* 2019). The Camel is a bar built and well-mixed estuary that is typical in terms of the processes above.

The hydrodynamic simulations of tracer dispersal performed in this study provide a robust, albeit relative, framework within which to base management decisions. The main result is that the tracer simulations suggested that the shellfish sites in the Camel were well located (in the mid-estuary) to reduce the risk of prolonged exposure to CSO discharge, compared with other locations both upstream or downstream. This result held true for the simulations with 28-day decay but was less apparent for the simulations with 24-hr decay, particularly during spring-summer for the eastern/upstream shellfish sites that appeared to encounter high tracer concentrations. Even so, the locations of the beds are so that they flank the main water flow (particularly S1 that is closest to the upstream CSOs D1-D3), resulting in simulated tracer concentrations that were generally lower (indicated by bluer colour) than in the main channel. Further, the estuary area where the tracer concentrations are thought to be highest, in the upper estuary river channels, were clearly identified. Furthermore, the highest seasonal-mean tracer concentrations were found in the autumn and winter months and seasonal-mean concentrations were much lower during the spring and especially the summer. Whereas during spring and summer, hotspots of maximum concentrations occurred

due to CSO spillage from Porthilly Cove Pumping Station (D4) when river discharge was low and dilution therefore was minimal. High concentration events during summer tended to be short lived. Nevertheless, viral indicators of human source sewage pollution were observed in 50-60% of shellfish samples, confirming that contamination from CSO spillages does indeed occur across all the shellfish beds in the Camel (see Section 4 for details).

In the absence of realistic CSO discharge data inputted into the simulations, it was assumed that each discharge event contained 1 l s^{-1} of effluent (tracer) of unit concentration. This allowed assessment of the relative tracer dispersal patterns from all CSOs and events, which is an important initial step in the risk assessment for the estuary. For future studies, the simulations could be built upon by applying realistic discharge rates from the CSO events. This would allow a range of CSO discharge behaviours and bacteria and viral dilution rates to be distinguished, and would ultimately improve the risk assessment.

The simulations used tracer variables with a wide range of decay rates that replicate minimum/maximum potential bacterial and viral decay, and this parameter led to profound differences in the risk maps produced. Simulations with a longer decay rate led to more seaward dispersal of the tracers, in all cases and seasons. Concentrations were negligible outside the estuary, giving us some confidence that in most cases the estuary mouth can be thought of as a seaward limit of viral dispersal. Therefore, more knowledge of bacteria and virus decay rates under conditions prevailing in the Camel estuary would improve predictions of the potential influence of health risk within the Camel and other estuaries. A realistic two-year period was simulated from a holistic perspective where multiple discharges from nine CSOs were traced simultaneously. The results indicate that specific CSOs pose a relatively high risk to the current shellfish sites; for example, we have highlighted the eastern beds (S1: S. Planna and S2: Pinkson Creek) as being at a relatively higher risk than the western beds, as a result of CSO spillages from D1-D3 sites that discharge into the estuary from the upstream rivers. Further work is required to understand the potential microbial dispersal following specific environmental events (including extreme events not captured within the two-year period) from these CSOs in isolation. Further hydrodynamic modelling could include 3D processes of the water column and possibly sediment transport. The latter has been shown to be responsible for net import and deposition within the Camel estuary. There may also be tidal eddy circulation patterns that increase the resident times for faecal contaminants in areas around Rock and Porthilly (https://www.padstow-harbour.co.uk/Ecospan-Modelling-of-sand-movement-in-the-Camel-estuary-near-Padstow.pdf. Another improvement would be to investigate how the other environmental variables affect bacteria and virus survival and the ranges encountered during transportation through the river-estuary continuum.

3.6 References

Bashawri YM, Robins P, Cooper DM, McDonald JE, Jones DL, and Williams AP (2020) Impact of sediment concentration on the survival of wastewater-derived blaCTX-M-15-producing *E.*

coli, and the implications for dispersal into estuarine waters. International Journal of Environmental Research and Public Health 17: 7608.

Brown J, Turrell WR and Simpson JH (1991) Aerial surveys of axial convergent fronts in UK estuaries and the implications for pollution. Marine Pollution Bulletin 22: 397-400.

DeVilbiss SE, Steele MK, Krometis LAH and Badgley BD (2021) Freshwater salinization increases survival of Escherichia coli and risk of bacterial impairment. Water Research 191: 116812.

CEFAS (2015) *Review of the Camel 2009 Sanitary Survey*. Available at: https://www.cefas.co.uk/publications-data/food-safety/sanitary-surveys/england-and-wales/.

EDINA (2008) 'EDINA Marine Digimap Service, Hydrospatial bathymetry, scale 1: 45,000, tiles: NW25000040-NW25000080, NW25200040-NW25200080, NW25400040-NW25400080. Updated April 2008. UK: SeaZone Solutions Ltd.'

Egbert GD, Bennett AF, and Foreman MG (1994) TOPEX/POSEIDON tides estimated using a global inverse model. Journal of Geophysical Research: Oceans, 99(C12): 24821-24852.

Garcia-Garcia LM, Campos CJA, Kershaw S, Younger A, Bacon J (2021) Scenarios of intermittent *E. coli* contamination from sewer overflows to shellfish growing waters: the Dart Estuary case study Marine Pollution Bulletin 167: 112332.

Hassard F, Sharp, JH, Taft, H., Le Vay L, Harris JP, McDonald JE, Tuson K, Jones DL, Malham SK (2017) Critical Review on the Public Health Impact of Norovirus Contamination in Shellfish and the Environment: A UK Perspective. Food and Environmental Virology. 9: 123-141.

Hata A, Katayama H, Kojima K, Sano S, Kasuga I, Kitajima M, Furumai H (2014) Effects of rainfall events on the occurrence and detection efficiency of viruses in river water impacted by combined sewer overflows Science of the Total Environment 468–469: doi: 10.1016/j.scitotenv.2013.08.093.

Hervouet J (2007) *Hydrodynamics of free surface flows*. 1st ed. John Wiley and Sons, Press.

Hervouet J-M, Pavan S, and Ata R (2015) 'Distributive advection schemes and dry zones , new solutions', in *XXIIth TELEMAC-MASCARET Technical User Conference*.

Kitajima M, Ahmed W, Bibby K, Carducci A, Gerba CP, Hamilton KA, Haramoto E, and Rose JB (2020) SARS-CoV-2 in wastewater: State of the knowledge and research needs. Science of The Total Environment 739: 139076.

Lyddon C, Lewis M, Couthard T, Robins P (2022) Historic spatial patterns of compound flood events in UK estuaries. Estuaries and Coasts (In review).

Murphy HM (2017) 'Part Four: Management of risk from excreta and wastewater - septic systems', in Rose, J. B. and Jiménez-Cisneros, B. (eds) *Water and Sanitation for the 21st Century: Health and Microbiological Aspects of Excreta and Wastewater Management (Global*

Water Pathogen Project). Michigan State University, E. Lansing, MI, UNESCO. Doi.org/10.14321/waterpathogens.59.

Rastogi AK, and Rodi W (1978) 'Predictions of heat and mass transfer in open channels', Journal of the Hydraulics division 104: 397–420.

Regnier P, Mouchet A, Wollast R, and Ronday F (1998) A discussion of methods for estimating residual fluxes in strong tidal estuaries. Continental Shelf Research 18: 1543-1571.

Robins PE, Neill SP, and Giménez L (2012) A numerical study of marine larval dispersal in the presence of an axial convergent front. Estuarine, Coastal and Shelf Science 100: 172-185.

Robins PE, Lewis MJ, Simpson JH, Howlett ER, and Malham SK (2014) Future variability of solute transport in a macrotidal estuary. Estuarine, Coastal and Shelf Science 151: 88-99.

Robins PE, Lewis MJ, Freer J, Cooper DM, Skinner CJ, and Coulthard TJ (2018) Improving estuary models by reducing uncertainties associated with river flows Estuarine, Coastal and Shelf Science, 207: 63-73.

Robins PE, Farkas K, Cooper D, Malham SK, and Jones DL (2019) Viral dispersal in the coastal zone: A method to quantify water quality risk. Environment international 126: 430-442.

Smith, R. and Scott, C.F., 1997. Mixing in the tidal environment. *Journal of Hydraulic Engineering*, *123*(4), pp.332-340.

4 Viral monitoring of faecal source indicators and human pathogens

4.1 Summary

Viral monitoring in wastewater effluent and shellfish at the Camel estuary was conducted over a two-year period (2019-2021), to (i) assess the potential health risks associated with the consumption of mussels and oysters harvested in the study area and (ii) to investigate whether the bacterial contamination affecting the area is originated from animal or human sources. Noroviruses were detected sporadically in shellfish samples throughout the year, with peak titres in shellfish during the summer and/or when COVID-19 restrictions were lifted. This indicates that the movement and mixing of people may have resulted in norovirus illnesses and subsequently the virus reached the aquatic environment and accumulated in shellfish. Mastadenovirus F (exclusive to humans) and atadenoviruses (commonly infecting sheep, cattle, deer and goats) were common in shellfish samples, with strong correlation between the titres of these indicator viruses. This suggests that human and animal waste input to the Camel estuary also correlate with each other and may be responding to common environmental drivers. Of the human pathogenic viruses, sapovirus and hepatitis E virus concentrations correlated well with mastadenovirus titres, while norovirus showed no such pattern. This may be due to the differences in the accumulation and survival of the viruses in shellfish. There was no meaningful correlation between virus concentration data and *E. coli* counts, confirming that *E. coli* may not be a good indicator for prolonged viral contamination in shellfish.

4.2 Introduction

Noroviruses and other enteric viruses are becoming an emerging risk with increasing number of outbreaks globally. These pathogens are transmitted via the faecal-oral route and the infection causes severe gastroenteritis. The symptoms usually only last for a few days, although, in some cases the illness may be more severe and even life-threatening (Katayama and Vinje, 2017). Globally, there are 685 million cases with approximately 200,000 deaths (CDC, 2016; Katayama and Vinje, 2017), with a total of US$4.2 billion in direct health system costs and US$60.3 billion in societal costs every year (Bartsch *et al*, 2016).

The main route of enteric virus transmission is direct (person-to-person) contact, however, the number of cases associated with consumption of contaminated food and water is on the rise (Bellou *et al*, 2013; Radin, 2014; Williams and O'Brien, 2019). Enteric viruses can be found at high concentrations in the faeces of infected individuals for prolonged times following infection (Aoki *et al*, 2010). These viruses are extremely resistant to most wastewater treatment procedures (Kitajima *et al*, 2014; Qiu *et al*, 2015; Sidhu *et al*, 2017), and hence infectious viruses may be discharged into the aquatic environment (Farkas *et al*, 2018; Iaconelli *et al*, 2017) where they may be uptaken by filter feeders, including oysters and mussels (Adriaenssens *et al*, 2021, 2018; Farkas *et al*, 2018; Landry *et al,* 1983; Lowther *et al,*

2012). As these shellfish are often consumed raw or lightly cooked, such contamination can lead to foodborne outbreaks and subsequent shellfish bed closures.

The current classification system for water quality and shellfish hygiene in the UK is solely based on faecal indicator bacteria (*E. coli*) detection. That is based on the assumption that these bacteria associate with human waste and hence would indicate any wastewater-derived microbial contamination in the aquatic environment. However, enteric viruses have been shown to be more persistent in the environment than indicator bacteria, suggesting that *E. coli* may not always be representative for viral contamination (Baggi *et al*, 2001; Espinosa *et al*, 2009; Lin and Ganesh, 2013). Furthermore, *E. coli* is also being representative of non-human faecal matter too (i.e. wildlife and domesticated animals) as well as with human waste, and hence the presence of such bacteria may not indicate any human health risks (Devane *et al*, 2020; Nguyen *et al*, 2018). In order to overcome these limitations and have a better understanding on viral contamination, comprehensive viral monitoring campaigns are necessary.

In this study, viral monitoring was conducted in wastewater effluent and shellfish at the Camel estuary over a two-year period. The aims of the viral monitoring were to (i) assess the potential health risks associated with the consumption of mussels and oysters harvested in the study area and (ii) to investigate whether the bacterial contamination affecting the area is originated from animal or human sources. To assess health risks, the abundance of common and emerging human enteric viruses, namely, norovirus GI and GII, sapovirus and hepatitis A/E viruses, which are often associated with water- and foodborne gastroenteritis was investigated. In order to assess if the microbial contamination originated from humans or animals, we used adenoviruses, namely Mastadenovirus groups C and F, as source indicators. Adenovirus infections are common and most cases the infection remains asymptomatic enabling the rapid spread of the viruses within communities. As infections are common, viruses are found in wastewater at high concentrations and they can be used as indicators for human waste contamination in the aquatic environment (Farkas *et al*, 2020). Two groups of animal-associated adenoviruses were selected, ovine adenoviruses and atadenoviruses, commonly infecting sheep, cattle, goats and deer, as a proxy for animal waste-associated pollution.

4.3 Methods

During the study period of 2nd June 2019 to 8th August 2021, 167 oyster and 126 mussel samples were collected in parallel with the samples taken for bacterial assays (Table 4.1). A 100 ml aliquot of water was also collected in the estuary near a sewage pipe fortnightly from 2nd June 2019 to 11th April 2021 (Figure 2.1). No samples were collected during the national lockdown due to the COVID-19 pandemic between 2 March – 15 August 2020.

Table 4.1 The type and number of samples collected and tested at each sampling site in the Camel estuary. RNA viruses include norovirus GI/GII, hepatitis A/E viruses and sapoviruses. *n=24 for norovirus GII

Site	Species	Total samples	Tested for RNA viruses	Tested for adenoviruses
Ball Hill	Mussels	41	41	38
Ball Hill	Oysters	42	42	39
Gentle Jane	Mussels	45	45	40
Gentle Jane	Oysters	45	45	40
Longlands	Oysters	42	42	39
Porthilly Rock	Mussels	38	38	36
Porthilly Rock	Oysters	40	40	37
	Wastewater	33	33*	32

The mussel and oyster samples, 15 and 10 animals/site/sampling occasion, respectively, were processed according to the ISO 15216-2:2019 standard. In brief, the digestive tissue was extracted from the animals and a 2g aliquot of the homogenised tissue was spiked with murine norovirus (MNV) to provide an extraction efficiency control and treated with proteinase K prior to RNA/DNA extraction using the Nuclisens® extraction system (BioMerieux, France). A 30ml aliquot of the wastewater samples was spiked with MNV as an extraction control, and then ultrafiltered using 100kDa Amicon™ Ultra Centrifugal Filter Units (Merck, USA) as recommended by the manufacturer. The final volume was adjusted to 1ml using phosphate-buffered saline (PBS), pH 7.4. The RNA/DNA of the wastewater concentrates were also extracted using the Nuclisens extraction system.

Quantitative reverse transcription PCR (qRT-PCR) was used for the quantification of viral RNA in the shellfish and sewage extracts, as described elsewhere (Farkas *et al*, 2017; Kitajima *et al*, 2010). The limit of detection was 200 gc/g for all enteric RNA viruses. The limit of quantification was 3,600 gc/g for norovirus GI, 2,600 gc/g for norovirus GII, 1,800 gc/g for sapovirus, 2,400 gc/g for hepatitis E virus, as determined previously (Farkas et al, 2017). A novel quadruplex qPCR assay for the quantification of adenoviral DNA using existing primer and probe sequences (Wolf *et al*, 2010). These assays are well-established and highly specific to the target strains and genotype, hence false positive results are unlikely to occur. For quantification, dilution series of plasmid DNA incorporating the target sequencies were used. Molecular grade water was also used as a negative control in all qPCR assays and those were negative suggesting no cross-contamination.

All 326 samples were tested for norovirus GI, hepatitis A/E viruses and sapovirus, 317 samples were tested for norovirus GII and 301 samples were tested for adenoviruses (Table 4.1). The results are interpreted as genome copies (gc)/g shellfish digestive tissue. Two-tailed Spearman correlation analysis was performed using IBM SPSS Statistics v27.

4.4 Results

4.4.1 Quality control

Selected samples were spiked with known quantities of Murine Norovirus (MNV) as a process control to estimate the viral recovery rates. As the structure of MNV is similar to the structure of most enteric viruses, if the MNV is successfully recovered, it is expected that the target viruses would also be recovered during sample process and detection. The recommended recoveries of process controls according to the ISO 15216-2:2019 standard is >1%.

The mean MNV recovery was 62% in oyster samples (n=74), 50% in mussel (n=56) and 65% in wastewater samples (n=11). All samples had >1% recoveries and 94% of the samples having >10% recoveries. Therefore, we concluded that the concentration, extraction and detection methodologies are appropriate for the target viruses. However, not all samples were spiked with process control virus due to limited virus availability during COVID lockdowns, we are confident that the recoveries of the spiked samples are representative for the unspiked samples, as the quality of the shellfish was consistent during the whole sampling period. The frequent detection of human and animal viruses in the samples also suggest that viruses were successfully recovered even in unspiked samples.

4.4.2 Surveillance of human pathogenic viruses – assessment of health risks

Norovirus GI and GII were detected in shellfish and in wastewater, although, usually at very low concentrations. In general, norovirus GII was more frequently detected than norovirus GI at all sites in all sample types (Table 4.2). At two out of three sites, the norovirus detection rates were higher in oyster than in the corresponding mussel samples taken at each site, suggesting that the virus uptake may differ between the two species. Overall, the concentrations of noroviruses were very low with peak norovirus titres observed during known norovirus outbreaks in late August 2019 and between November 2019 and February 2020 (Public Health England, 2021a). Norovirus GI was detected after the national lockdown in August-September 2021 and in December 2020 – March 2021, when COVID-19 restrictions were temporarily lifted and tourism in the area increased, although, no gastroenteritis outbreaks were noted. Noroviruses were detected in the autumn and winter period of 2020-21, although sporadically. The highest concentrations and detection rates were observed in January-February 2021, when most norovirus cases are reported. Some samples were

positive for norovirus in March-April and in June, however, the concentrations were extremely low.

Hepatitis A virus was not detected in any of the samples. This corresponds with the Public Health England reports suggesting there were no cases of hepatitis A in the area during the sampling period. Sapovirus and hepatitis E virus were detected in shellfish and wastewater samples between August 2020 and August 2021 (Table 4.2), however, at very low concentrations. Even though enteric viruses were detected in the shellfish samples, it is hard to draw any conclusions on the temporal changes in viral abundance as in most cases the viral concentrations were close to the limit of detection, below the limit of quantification of the assay (Figure 4.1).

Table 4.2 The detection rates (i.e. the ratio of positive samples; D) and mean concentrations, expressed as genome copies (gc)/g shellfish digestive tissue or gc/ml wastewater, of animal and human-associated viruses in shellfish and sewage in the Camel estuary.

	Human norovirus GI		**Human norovirus GII**		**Human sapovirus**		**Hepatitis E virus**		**Human adenovirus F**		**Human adenovirus C**		**Ovine adenovirus (sheep, cattle)**		**Atadenovirus (sheep, cattle, deer, goats)**	
Sample	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**	**D**	**gc/g**
Ball Hill mussels	13%	128	18%	803	15%	152	8%	136	61%	104,000	37%	5,250	3%	484	55%	344,000
Ball Hill oysters	15%	132	22%	877	10%	192	7%	855	59%	122,000	33%	10,100	3%	430	64%	291,000
Gentle Jane mussels	14%	63	27%	2800	11%	179	2%	197	63%	152,000	33%	8,340	3%	4,757	60%	411,000
Gentle Jane oysters	11%	100	23%	547	9%	89	0%	-	65%	191,000	33%	7,340	10%	5,000	68%	598,000
Longlands oysters	22%	722	24%	904	5%	458	0%	-	67%	204,000	31%	15,900	10%	1,350	62%	733,000
Porthilly Rock mussels	8%	130	18%	2,540	13%	200	3%	39	50%	165,000	36%	5,390	8%	3,720	53%	488,000
Porthilly Rock oysters	13%	55	23%	7,770	15%	183	10%	204	51%	241,000	32%	7,380	8%	18,000	65%	642,000
Wastewater	3%	200	13%	130	6%	182	3%	857	41%	2,480	9%	366	13%	8,610	28%	6,600

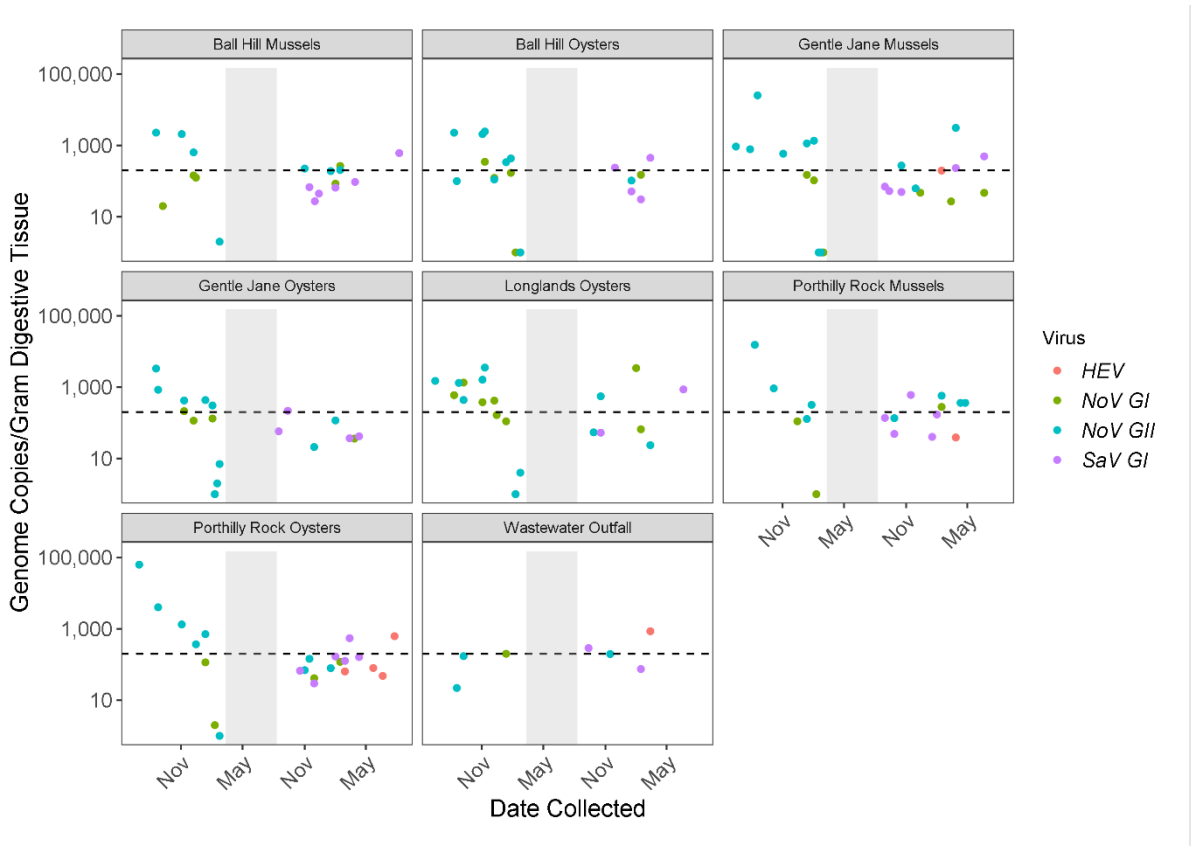


Figure 4.1 The concentration of norovirus GI and GII, sapovirus and hepatitis E virus in shellfish. The grey area indicates the time period when no samples were taken due to national COVID-19 lockdown. The colour of the datapoints indicate the virus that was tested. Dotted line indicates the limit of detection (200 gc/g). The limit of quantification were: (norovirus GI: 3,600 gc/g, norovirus GII: 2,600 gc/g, sapovirus: 1,800 gc/g, hepatitis E virus: 2,400 gc/g, as determined in Farkas *et al*, 2017.

4.4.3 Surveillance of indicator viruses – contamination source tracking

Human and animal-associated adenoviruses were detected in all sample types and locations, suggesting that the study area is vulnerable to both human and animal-related microbial contamination. The most frequently detected group was the atadenovirus (infecting cattle and sheep), and human mastadenovirus F, followed by the human mastadenovirus C, with the ovine adenovirus (infecting cattle, sheep, goats and deer) being the least prevalent (Table 4.2 and Figures 4.2-4.3). The prevalence of mastadenovirus F and atadenovirus was similar among sites and shellfish types (Figure 4.4), except for the mussels collected at Porthilly Rocks where the viruses were detected less frequently. The viral concentration results show that the atadenoviruses and mastadenovirus F and C were more abundant in Aug 2020 – Aug 2021 than during the period of Aug 2019 – Jan 2020, whereas the concentration of ovine adenoviruses was constant (Figures 4.2-4.3). That may be due to changes in weather, human or animal populations, however, no data were collected to investigate. Longer, multi-year monitoring would be required to investigate the temporal changes in adenovirus titres.

The enteric virus detection rates were also lower in the mussels at this site compared to other locations and oysters, suggesting that the mussels collected here were less exposed to viral contamination. The association between the detection rates of mastadenovirus F and noroviruses suggest that adenoviruses can indicate the magnitude of viral contamination in shellfish.

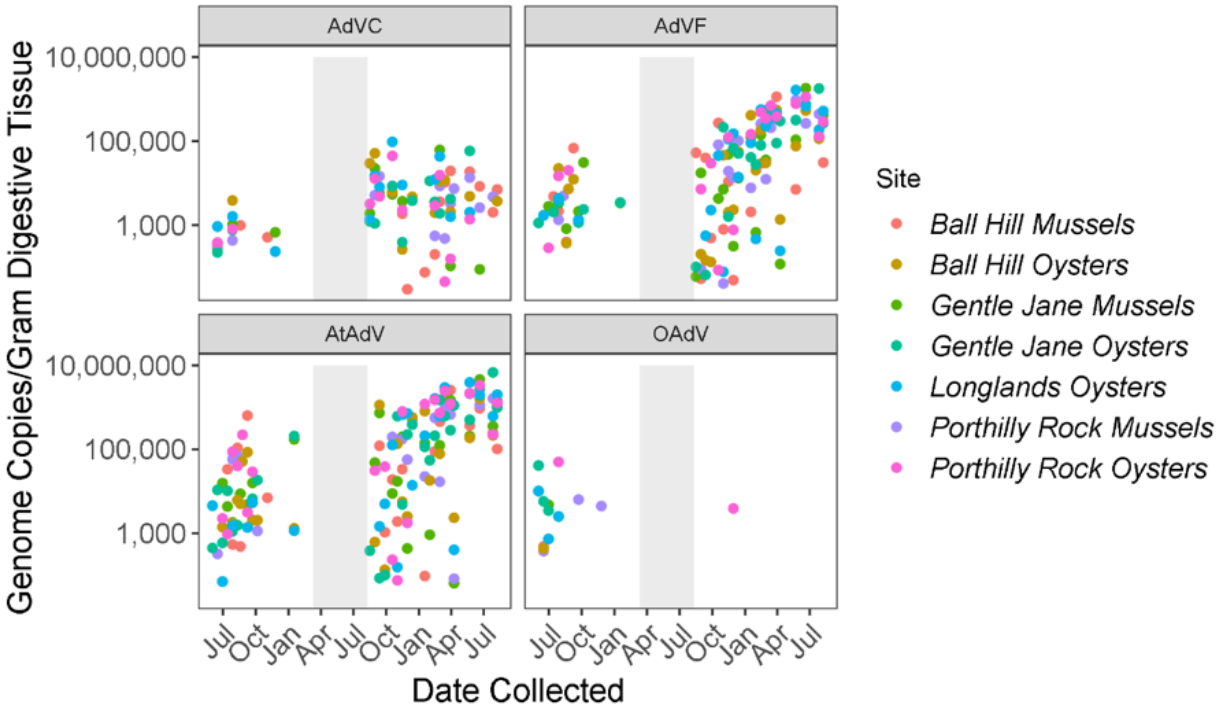


Figure 4.2 The concentration of human mastadenovirus C (AdVC), human mastadenovirus F (AdVF), ovine adenovirus (OAdV) infecting cattle and sheep, and atadenoviruses (AtAdV) infecting sheep, cattle, goats and deer. The colour of the datapoints indicate the species of shellfish and site tested. The grey area indicates the time period when no samples were taken due to national COVID-19 lockdown.

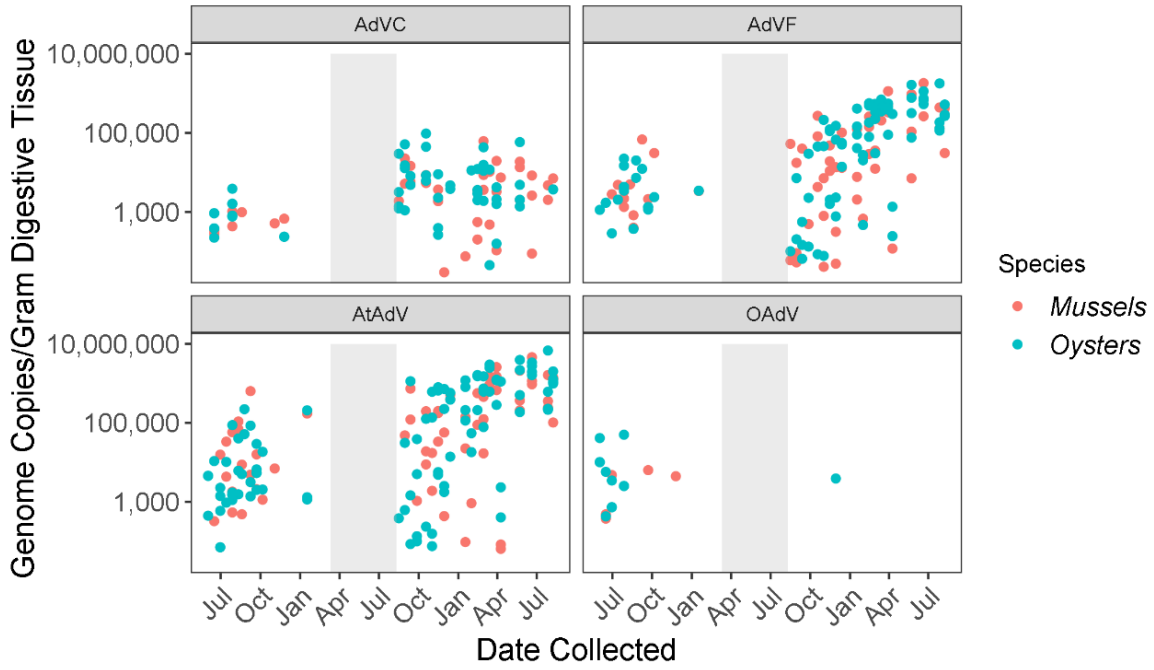


Figure 4.3 The concentration of human mastadenovirus C (AdVC), human mastadenovirus F (AdVF), ovine adenovirus (OAdV) infecting cattle and sheep, and atadenoviruses (AtAdV) infecting sheep, cattle, goats and deer. The grey area indicates the time period when no samples were taken due to national COVID-19 lockdown. The colour of the datapoints indicate the species of shellfish tested.

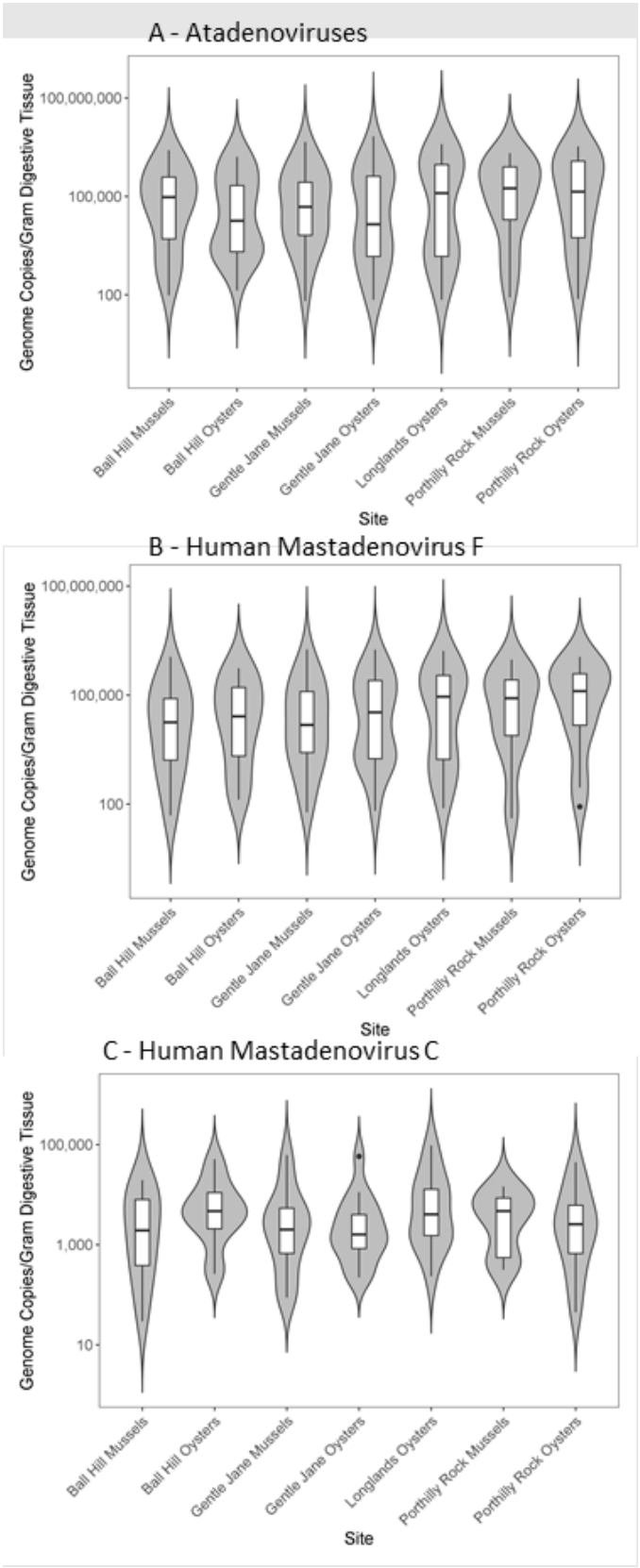


Figure 4.4 Violin plot displaying the distribution of the observed concentrations of (A) atadenoviruses infecting sheep, cattle, goats and deer; (B) human mastadenovirus F and (C) human mastadenovirus C at each shellfish site over the sampling period.

All four groups of adenoviruses were also detected in wastewater, suggesting that the water was contaminated with animal waste at the sampling point, approximately 20 metres downstream the pipe outlet. The contamination may be originated directly from animals grazing in the area or from the estuary during high tides. The most frequently detected adenoviruses, the human mastadenovirus F and the animal-associated atadenoviruses showed good correlation (Figure 4.5). Interestingly, the concentrations of viruses in wastewater were much lower that the corresponding concentrations of viruses in shellfish. This may be due to the accumulation of viruses in shellfish. Furthermore, the wastewater collected entered the environment after treatment and have been mixed with estuarine water and rainwater, whereas the shellfish beds may be vulnerable to untreated wastewater discharge via CSOs.

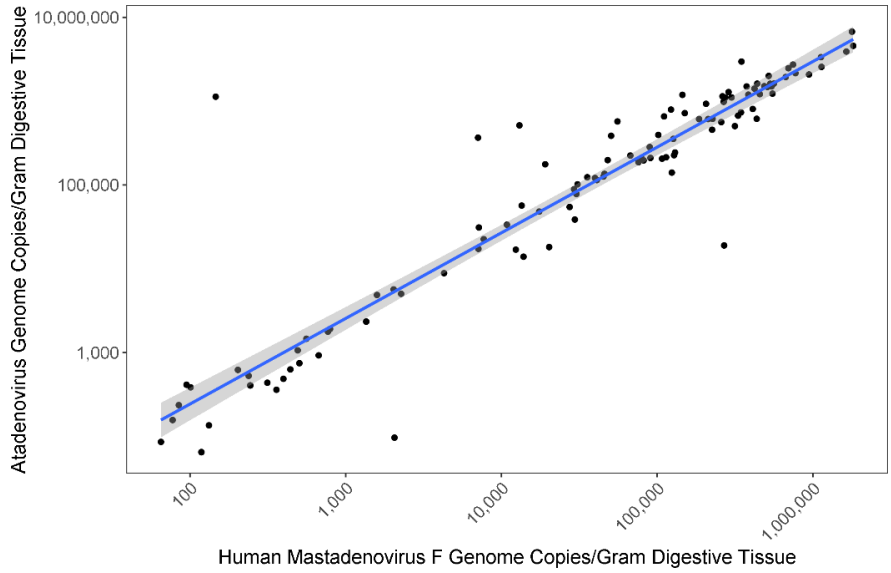


Figure 4.5 Scatterplot and regression analysis of concentrations of human mastadenovirus F and atadenoviruses infecting sheep, cattle, goats and deer observed in the digestive tissue of the mussels and oysters from the seven shellfish production areas sampled. $F_{1, 113} = 856.8$, $p < 0.001$.

4.4.4 Correlation between pathogenic viruses and indicators

Comparing norovirus positivity in samples with low (<230 MPN) and high (>230 MPN) *E. coli* concentrations measured with MPN and pour plating techniques, no correlation between *E. coli* and norovirus concentrations was found (Tables 4.3, 4.4 and Figure 4.6). Further correlation analysis suggested weak negative correlation between *E. coli* and ovine adenovirus and human sapovirus.

Weak correlation was observed for norovirus GI and GII detections, suggesting that these genotypes co-exist in the population. Hepatitis E virus and sapovirus concentrations correlated with human mastadenovirus F, C and atadenovirus titres, whereas weak negative correlation was observed with ovine adenovirus. Strong correlation was noted for human mastadenovirus F and atadenovirus (Figure 4.5).

Table 4.3 Relationship between *E. coli* load and the incidence of noroviruses in shellfish

	MPN ≤ 230	**Pour Plate ≤ 230**	**MPN > 230**	**Pour Plate > 230**
Number (%) of norovirus positive samples	26 (26.5%)	25 (30.9%)	58 (26.9%)	13 (18.3%)
Total paired samples	98	81	216	71

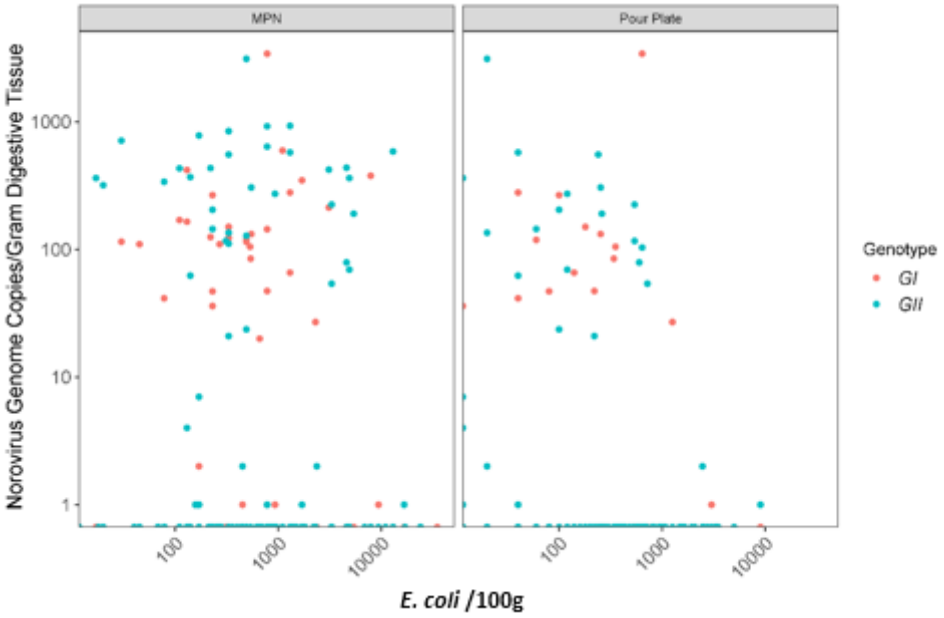


Figure 4.6 Scatterplot showing the relationship between *E. coli* and Norovirus counts using MPN and pour plate method for *E. coli* quantification and RT-qPCR for norovirus quantification.

Table 4.4 Spearman correlation coefficients (rs) observed amongst *E. coli* and viral concentrations in shellfish. ** Correlation is significant at the p value < 0.01 level (2-tailed). * Correlation is significant at the p value < 0.05 level (2-tailed). n: number of samples included in the analysis.

		E. coli MPN	*E. coli* Pour plate	Noro-virus GI	Noro-virus GII	Sapo-virus	Hepa-titis E virus	Mast-adeno-virus F	Mast-adeno-virus C	Ovine adeno-virus	Ataden-ovirus
E. coli MPN	r^2	1	**0.521****	-0.087	-0.046	-0.074	-0.006	-0.057	-0.031	**-0.154***	-0.007
	p	.	0	0.168	0.482	0.24	0.925	0.41	0.639	0.022	0.926
	n	388	248	252	236	253	253	208	227	221	192
E. coli Pour plate	r^2	**0.521****	1	-0.003	-0.153	**-0.192***	-0.049	0.05	0.098	.	0.05
	p	0	.	0.976	0.097	0.036	0.592	0.604	0.31	.	0.606
	n	248	256	120	119	120	120	109	109	109	109
Norovirus GI	r^2	-0.087	-0.003	1	**0.183****	-0.01	-0.066	0.087	-0.072	0.011	-0.077
	p	0.168	0.976	.	0.003	0.859	0.258	0.172	0.244	0.864	0.245
	n	252	120	294	270	294	294	246	266	258	229
Norovirus GII	r^2	-0.046	-0.153	**0.183****	1	-0.017	-0.083	0.009	-0.038	-0.011	0.018
	p	0.482	0.097	0.003	.	0.778	0.171	0.895	0.559	0.862	0.797
	n	236	119	270	271	271	271	226	243	235	211
Sapovirus	r^2	-0.074	**-0.192***	-0.01	-0.017	1	0.084	**0.220****	**0.178****	-0.06	**0.161***
	p	0.24	0.036	0.859	0.778	.	0.149	0.001	0.003	0.34	0.015
	n	253	120	294	271	295	295	247	267	259	229
Hepatitis E virus	r_s	-0.006	-0.049	-0.066	-0.083	0.084	1	**0.201****	**0.125***	-0.034	**0.140***
	p	0.925	0.592	0.258	0.171	0.149	.	0.002	0.042	0.588	0.035
	n	253	120	294	271	295	295	247	267	259	229
Mastadeno-virus F	r^2	-0.057	0.05	0.087	0.009	**0.220****	**0.201****	1	**0.364****	**-0.144***	**0.842****
	p	0.41	0.604	0.172	0.895	0.001	0.002	.	0	0.026	0
	n	208	109	246	226	247	247	247	247	239	224
Mastadeno-virus C	r^2	-0.031	0.098	-0.072	-0.038	**0.178****	**0.125***	**0.364****	1	0.043	0.**376****
	p	0.639	0.31	0.244	0.559	0.003	0.042	0	.	0.497	0
	n	227	109	266	243	267	267	247	267	257	229
Ovine adeno-virus	r^2	**-0.154***	.	0.011	-0.011	-0.06	-0.034	**-0.144***	0.043	1	-0.089
	p	0.022	.	0.864	0.862	0.34	0.588	0.026	0.497	.	0.185
	n	221	109	258	235	259	259	239	257	259	224
Atadeno-virus	r^2	-0.007	0.05	-0.077	0.018	**0.161***	**0.140***	**0.842****	**0.376****	-0.089	1
	p	0.926	0.606	0.245	0.797	0.015	0.035	0	0	0.185	.
	n	192	109	229	211	229	229	224	229	224	229

4.5 Discussion

4.5.1 Human enteric viruses in mussels and oysters

Over the two year- period of this study, noroviruses were detected sporadically in shellfish samples. Norovirus GI and GII genotype detections correlated well with each other, suggesting that outbreaks of the two genotypes may coincide. The lower detection rates of norovirus GI compared to norovirus GII verifies that norovirus GI is less prevalent than norovirus GII in the UK (Farkas *et al*, 2018; Lowther *et al*, 2018). Noroviruses were slightly more prevalent in oysters than mussels as observed previously (Kittigul *et al*, 2016) suggesting the bioaccumulation capacity may be different between species (Tian *et al*, 2007). However, the differences may be attributed to slightly different exposure. Although, most norovirus infections have been noted during the cold months (Katayama and Vinje, 2017), our results suggests that the infections may occur any time during the year. Specifically, peak norovirus titres in shellfish from the Camel estuary were observed during the summer and/or when COVID-19 restrictions were lifted. This indicates that the movement and mixing of people may have resulted in norovirus illnesses and subsequently the virus reached the aquatic environment and accumulated in shellfish.

Hepatitis A virus was not found in either mussel or oyster digestive tissue samples collected during the study period. That correlates with the low numbers of hepatitis A clinical cases in the area (Public Health England, 2021b). Hepatitis E virus and sapovirus were detected sporadically in shellfish samples. These viruses are only diagnosed at hospital settings, therefore, most infections are undetected and there is limited information available on case numbers. We found that sapovirus is detected along with or shortly after norovirus peaks in shellfish samples, as reported previously (Farkas *et al*, 2018). The presence of hepatitis A/E viruses and sapovirus in the aquatic environment in the UK is extremely rare and thus the health risks associated with environmental transmission are low. To our knowledge, this is the first time hepatitis E virus was detected in the aquatic environment in England and the second time it was noted in the UK (Crossan *et al*, 2012; O'Hara *et al*, 2019). These viruses are considered emerging pathogens, and their abundance in the UK may change in the future, in which case the study of shellfish and water sources would be a valuable tool for outbreak surveillance and mitigation.

4.5.2 The usefulness of bacterial and viral indicators in wastewater-derived viral pollution

In order to explore the magnitude of animal vs human waste distribution in shellfish, the samples were tested for mastadenovirus F and C (exclusive to humans), and two animal-associated adenoviruses, ovine adenovirus and atadenovirus (commonly infecting sheep, cattle, deer and goats). All these viruses are common in the host species and usually result in asymptomatic infections. As a result, they are commonly found in the hosts' faeces and subsequently in wastewater at high titres. Human mastadenoviruses have been commonly used as a proxy for wastewater contamination in the aquatic environment (Farkas *et al,* 2020), however, limited

information is available on the prevalence of animal associated adenoviruses (Wolf *et al*, 2010). This study found that mastadenovirus F and atadenoviruses were common in shellfish samples, whereas the other two viral species were less abundant and hence may not be suitable targets for source tracking. A significant correlation between the titres of the mastadenovirus F and atadenovirus was also found, suggesting that the human and animal waste input also correlate with each other. However, we have not collected data to investigate to reasons for this phenomenon, it is possible that the viruses are responding to common environmental drivers, such as rainfall events causing CSO spills and agricultural run-off.

While the sapovirus and hepatitis E virus concentrations correlated well with mastadenovirus titres, norovirus showed no such pattern. This may be due to the differences in the accumulation and survival of the viruses in shellfish. Previous research has found substantial differences in the time required for the depuration of different enteric viruses (noroviruses and hepatitis A virus) in oyster (Nappier *et al,* 2008; Polo *et al,* 2014). More data would be necessary to further investigate the bioaccumulation of pathogenic and indicator viruses.

When virus concentration data were compared with *E. coli* counts, no meaningful correlation was noted. This finding also suggests that *E. coli* may not be a good indicator for viral contamination in shellfish. A possible reason for that may be that bacteria inactivate more rapidly than noroviruses in shellfish and hence unable to bioaccumulate. However, the depuration of viruses in shellfish was not the scope of this study, such differences have been noted and discussed in previous research (Bazzardi *et al,* 2014; Burkhardt and Calci, 2000; Chung *et al,* 1998; Love *et al,* 2010). It is also important to note that PCR-based detection was used for the quantification of viruses in this study which does not indicate the infective nature of the target viruses. To date there are no reliable methods for the in vitro culturing of human noroviruses and sapoviruses and hence the persistency of these pathogens remains unknown.

However, most data on viral and bacterial contamination were collected during the COVID-19 pandemic with lockdowns and other restrictions in place which reduced the transmission of not only SARS-CoV-2 but other viruses as well, including noroviruses. The data collected prior to the pandemic were not sufficient for correlation analyses and hence more data from post-pandemic timeframes would be beneficial to further investigate the correlations between pathogenic viruses and potential faecal indicators.

4.6 References

Adriaenssens EM, Farkas K, Harrison C, Jones DL, Allison HE, McCarthy AJ (2018). Viromic analysis of wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses. mSystems. https://doi.org/10.1128/mSystems.00025-18

Adriaenssens EM, Farka K, McDonald JL, Jones DL, Allison HE, McCarthy AJ (2021) Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity. Water Research 203: 117568.

Aoki Y, Suto A, Mizuta K, Ahiko T, Osaka K, Matsuzaki Y (2010) Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients. Journal of Hospital Infectections 75: 42–46.

Baggi F, Demarta A, Peduzzi R (2001) Persistence of viral pathogens and bacteriophages during sewage treatment: Lack of correlation with indicator bacteria. Research Microbiology 152: 743–751.

Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY (2016) Global economic burden of norovirus gastroenteritis. PLoS One 11: 1–16

Bazzardi R, Fattaccio MC, Salza S, Canu A, Marongiu E, Pisanu M (2014) Preliminary Study on Norovirus, Hepatitis A Virus, Escherichia coli and their Potential Seasonality in Shellfish from Different Growing and Harvesting Areas in Sardinia Region. Italian Journal of Food Saftey 3: 1601.

Bellou M, Kokkinos P, Vantarakis A (2013) Shellfish-borne viral outbreaks: A systematic review. Food Environmental Virology 5: 13–23

Burkhardt W, Calci KR (2000) Selective Accumulation May Account for Shellfish-Associated Viral Illness. Applied Environmental Microbiology 66: 1375–1378

CDC, 2016. Norovirus worldwide [WWW Document]. URL https://www.cdc.gov/norovirus/worldwide.html (accessed 5.29.19).

Chung H, Jaykus L-A, Lovelace G, Sobsey MD (1998) Bacteriophages and bacteria as indicators of enteric viruses in oysters and their harvest waters. Water Science and Technolology 38: 37–44.

Crossan, C., Baker, P.J., Craft, J., Takeuchi, Y., Dalton, H.R. and Scobie, L., 2012. Hepatitis E virus genotype 3 in shellfish, United Kingdom. *Emerging infectious diseases*, *18*(12), p.2085.

Devane ML, Moriarty E, Weaver L, Cookson A, Gilpin B (2020) Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research 185: 116204.

Espinosa AC, Arias CF, Sánchez-Colón S, Mazari-Hiriart M (2009) Comparative study of enteric viruses, coliphages and indicator bacteria for evaluating water quality in a tropical high-altitude system. Environment and Health 8: 49.

Farkas K, Cooper DM, McDonald JE, Malham SK, de Rougemont A, de Rougemont A, Jones DL (2018) Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. Science of the Total Environment 634: 1174–1183

Farkas K, Malham SK, Peters DE, de Rougemont A, McDonald JE, Jones DL (2017) Evaluation of two triplex one-step qRT-PCR assays for the quantification of human enteric viruses in environmental samples. Food Environmental Virololgy 9: 343–349.

Farkas K, Walker DI, Adriaenssens EM, McDonald JE, Hillary LS, Malham SK, Jones DL (2020) Viral indicators for tracking domestic wastewater contamination in the aquatic environment. Water Research 181: 115926.

Iaconelli M, Muscillo M, Della Libera S, Fratini M, Meucci L, De Ceglia M, Giacosa D, La Rosa G (2017) One-year surveillance of human enteric viruses in raw and treated wastewaters, downstream river waters, and drinking waters. Food Environmental Virology 9: 79–88

ISO 15216-2:2019 Microbiology of the food chain – Horizontal method for determination of hepatitis A virus and norovirus using real-time RT-PCR – Part 2: Method for detection

Katayama H, Vinje J (2017) Norovirus and other Caliciviruses, in: Meschke, J.S., Gironés, R. (Eds.), Global Water Pathogen Project. Michigan State University, E. Lansing, MI, UNESCO.

Kitajima M, Iker BC, Pepper IL, Gerba CP (2014) Relative abundance and treatment reduction of viruses during wastewater treatment processes—Identification of potential viral indicators. Science of the Total Environment 488: 290–296.

Kitajima M, Oka T, Takagi H, Tohya Y, Katayama H, Takeda N, Katayama K (2010) Development and application of a broadly reactive real-time reverse transcription-PCR assay for detection of murine noroviruses. Journal of Virological Methods 169: 269–273

Kittigul L, Thamjaroen A, Chiawchan S, Chavalitshewinkoon-Petmitr P, Pombubpa K, Diraphat P (2016) Prevalence and molecular genotyping of noroviruses in market oysters, mussels, and cockles in Bangkok, Thailand. Food and Environmental Virology 8: 133–140.

Landry EF, Vaughn JM, Vicale TJ, Mann R (1983) Accumulation of sediment-associated viruses in shellfish. Applied Environmental Microbiology 45: 238–247.

Lin J, Ganesh A (2013) Water quality indicators: Bacteria, coliphages, enteric viruses. International Journal of Enviromental Health Research 23: 484-506

Love DC, Lovelace GL, Sobsey MD (2010) Removal of *Escherichia coli, Enterococcus fecalis,* coliphage MS2, poliovirus, and hepatitis A virus from oysters (Crassostrea virginica) and hard shell clams (Mercinaria mercinaria) by depuration. International Journal of Food Microbiology 143: 211–217.

Lowther JA, Gustar NE, Powell AL, Hartnell RE, Lees DN (2012) Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom. Applied Environmental Microbiology 78: 5812–5817.

Lowther JA, Gustar NE, Powell AL, O'Brien S, Lees DN (2018) A One-Year Survey of Norovirus in UK Oysters Collected at the Point of Sale. Food Environmental Virology 10: 278–287.

Nappier SP, Graczyk TK, Schwab, KJ (2008) Bioaccumulation, retention, and depuration of enteric viruses by Crassostrea virginica and Crassostrea ariakensis oysters. Applied Environmental Microbiology 74: 6825–6831.

Nguyen KH, Senay C, Young S, Nayak B, Lobos A, Conrad J, Harwood VJ (2018) Determination of wild animal sources of fecal indicator bacteria by microbial source tracking (MST) influences regulatory decisions. Water Research 144: 424–434.

O'Hara Z, Crossan C, Craft J, Scobie L (2018) First report of the presence of hepatitis E virus in Scottish-harvested shellfish purchased at retail level. Food and environmental virology. 10(2): 217-221.

Polo D, Álvarez C, Díez J, Darriba S, Longa Á, Romalde JL (2014) Viral elimination during commercial depuration of shellfish. Food Control 43: 206–212.

Public Health England, 2021a. PHE Weekly National Norovirus and Rotavirus Report.

Public Health England, 2021b. Laboratory reports of hepatitis A infections in England and Wales, 2019.

Qiu Y, Lee BE, Neumann N, Ashbolt N, Craik S, Maal-Bared R, Pang XL, Maal-Bared R, Pang XL (2015) Assessment of human virus removal during municipal wastewater treatment in Edmonton, Canadaian Journal of Applied Microbiology 119: 1729–1739.

Radin D (2014) New trends in food-and waterborne viral outbreaks. Archives Biololgocial Society 66: 1–9.

Sidhu JPS, Sena K, Hodgers L, Palmer A, Toze S (2017) Comparative enteric viruses and coliphage removal during wastewater treatment processes in a sub-tropical environment. Science of the Total Environment 616: 669–677.

Tian P, Engelbrektson AL, Jiang X, Zhong W, Mandrell RE (2007) Norovirus recognizes histo-blood group antigens on gastrointestinal cells of clams, mussels, and oysters: a possible mechanism of bioaccumulation. Journal of Food Protection 70: 2140–2147.

Williams and O'Brien, 2019. Assessing the contribution made by the food chain to the burden of UK-acquired norovirus infection. Norovirus Attribution Study, FS101040.

Wolf S, Hewitt J, Greening GE (2010) Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. Applied Environmental Microbiology 76: https://doi.org/10.1128/AEM.02249-09.

5 Comparison of MPN and Pour Plate methods for measurement of *E. coli* in shellfish samples.

5.1 Summary

The Most Probable Number (MPN) method (ISO 2016a), as the reference method, is internationally the most commonly used method for measurement of *E. coli* in official shellfish hygiene monitoring and is used for all Official Control sample analysis in the UK. Variability in the MPN assay is acknowledged and was considered as a potential confounding factor influencing the reliability of predictive models being developed in the DASSHH project. Hence the pour plate method (ISO 2001), also approved for use in official shellfish monitoring where validated, was also used for time series sampling of shellfish from the Camel, to allow comparison of (a) variability in results, focused on within sample measurement uncertainty and (b) differences in *E. coli* concentrations for single samples measured using both methods. The pour plate method consistently yielded less variable *E. coli* results (for repeat measures of single samples) than obtained by MPN, particularly for the upper range of *E. coli* concentrations. MPN results were also statistically higher than pour plate, whether considered at the level of the inherent measurement variability of individual MPN results or at the level of variability in a series of single results as used in the practical application of the official control regulations. These findings suggest that the MPN method has greater potential to generate outlier results that may hamper predictive modelling of *E. coli* in shellfish and could potentially influence the application of official monitoring results. The effect of the variability of the Official Control MPN results on predictive modelling of *E. coli* is explored in Section 6 of this report. The pour plate method is approved for use in Official Control sampling in the UK (within a range of 200 – 18,000 *E.coli*/100g), and the lower variability of *E.* coli results using this method could support consideration of its application in some production areas. This would be particularly relevant for sites where a) very high *E. coli* results are not commonly experienced and b) management decisions (e.g. downgrades and closures) may be influenced by a few high results. The relationships between MPN and pour plate results produced in the present report, may be applicable to evaluation of historical MPN results for other shellfish production areas, to assess the likelihood that the choice of method might influence classification outcomes.

5.2 Introduction

Where shellfish are grown in areas impacted by anthropogenic pollutants, they can accumulate and concentrate a range of contaminants include bacteria and viruses. These pollutants often include pathogenic bacteria and viruses associated with faecal matter that is introduced into the water column and originate from point and diffuse pollution such as from wastewater treatment facilities and agricultural runoff (Iwamoto *et al*, 2010). As bivalve shellfish are often eaten raw or only lightly cooked, they can become foodborne vectors for these pathogens with a potential risk of human illness (Pouillot *et al*, 2021).

To protect public health, regular monitoring of faecal indicator bacteria is undertaken with production areas classified according to the results of the monitoring. In the EU this framework is

referred to as the Shellfish Control Regulations and aims to ensure that the harvesting and selling of bivalve shellfish can only occur at sites that are deemed safe for consumption according to a standardised classification system (European Commission 2004; European Commission 2017;; European Commission 2019/627). Monitoring of the shellfish beds in England and Wales is usually by local authority representatives and shellfish samples are tested for levels of *E. coli*. Results are transferred from CEFAS to the Food Standards Agency (FSA) with advice on the appropriate classification of Class A, B or C depending on the levels of *E. coli* in shellfish flesh (Table 5.1). High levels of bacterial contamination can cause the closure of shellfish harvesting areas for extended periods of time to ensure the protection of public health.

Regulation for the classification of shellfish areas in the UK utilises the ISO accredited Most Probable Number or MPN method for measuring *E. coli* in shellfish flesh (ISO 2016a). This method uses dilution tubes and a probability calculation to give the concentration of viable organisms in a sample based on the number of tubes that return a positive result (West and Coleman, 1986). This method is specified in Codex Code of Practice/Standard EU legislation (CEFAS 2019; Walker *et al*, 2018). Three significant advantages of the dilution tube MPN method are the simplicity of the laboratory process itself relative to molecular methods and the volume of data that has been collected using the MPN method, and the wide range of *E. coli* concentrations that can be measured (Walker *et al,* 2018). However, the process is laborious and takes a minimum of 45 hours after sample arrival in the laboratory to yield results. Variability in the MPN assay is acknowledged and expected in any laboratory analysis with guidance available and estimates of measurement of uncertainty in food microbiology (ISO 2016; Lee and Murray, 2010; Walker *et al*, 2018).

Other methods are available that can be used to enumerate faecal indicator bacteria in shellfish. The main methods include the pour plate/TBX testing, impedance testing and plate spreading (spread plate) methods. Of these, the impedance (Dupont *et al*, 2004; Lee and Murray, 2010; IFREMER, 2014; Walker *et al,* 2018) and pour plate methods (EURL 2014; ISO 2001) have been validated and characterised against the MPN reference method (Walker *et al*, 2018; Pol-Hofstad and Jacobs-Reitsma 2021). The pour plate (ISO 2001), method is a culture-based technique that has been used for shellfish hygiene monitoring and relies on the counting of *E. coli* colonies on TBX or brilliance agar (Clements *et al*, 2013; Walker *et al*, 2018). Pour plate is a relatively simple and cost-efficient assay, providing results in 24 hours from start of sample testing in the laboratory. The RIKILT Institute of Food Safety, Wageningen University and the Netherlands National Institute for Public Health and Environment RIVM, have conducted several comparative and validation studies assessing the pour plate method against the reference MPN method for *E. coli* enumeration in shellfish samples (Mooijman *et al,* 2007; Jacobs-Reistma *et al,* 2010; Pol-Hofstad and Jacobs-Reitsma 2021). The most recent study concluded that "TBX pour plate' (ISO 2001) shows comparable performance to the reference method 'MPN method' (ISO 2016) for the enumeration of *Escherichia coli* in Live Bivalve Molluscs" (Pol-Hofstad and Jacobs-Reitsma 2021).

The pour plate has the advantages of relatively low-cost and faster turn-around times but is only approved for use (validated against the MPN method) in shellfish for a range between 200 - 18,000 *E. coli*/100 g, reflecting the higher variability observed at lower values and the difficulty in reading dense colonies on plates from highly contaminated samples (EURL 2014). This does not eliminate

its potential use in monitoring of shellfish production areas across all classifications, as a result returned as < 200 *E. coli*/100 g would still indicate conformity with the A/B classification threshold (< 230 *E. coli*/100 g). Use in areas that may return results above 18,000 *E. coli*/100 g would require some modification of the method. The most recent ISO comparison does not specify range limits, and variation in *e.g*. dilution may achieve results outside the prescribed range for shellfish of 200-18,000 *E. coli*/100 g. The impedance method for *E. coli* enumeration is also considered equivalent to the standard reference MPN method by the EURL (IFREMER, 2014; Walker *et al,* 2018). Whilst its use is not widespread in the EU, it is used in some French and Italian laboratories. The impedance process is rapid (5-10 hours from test commencing), simple to prepare and allows a high sample throughput. The equipment and laboratory set-up are expensive, thereafter however the cost per sample is much lower than other approved methods. The impedance method was not assessed as part of this study. Other methods that are not ISO accredited have been trialled against the MPN method, including a 3M Petrifilm technique, conductance using a Malthus 2000 instrument, Merck Chromocult agar with differing results (Ogden *et al,* 1998; Dupont *et al,* 1996; Walker *et al,* 2018). Ogden *et al* (1998) suggested that Merck Chromocult agar was in broad agreement with the MPN method, with advantages over the MPN on cost and time taken for analysis, however Dupont *et al* (1995) suggested conductance using a Malthus instrument gave better results. More recently a qPCR-MPN method was developed which could be used as a complementary method to the official MPN reference method (Walker *et al,* 2020). These methods have not been validated as comparable to the reference MPN method and as such were not tested in this study.

In the context of the DASSHH project, the overarching aim was to assess potential for prediction of *E. coli* levels in shellfish using environmental factors, based on models derived from measured *E. coli* data both from the official control sample records and from additional field sampling conducted during the study. The uncertainty in the MPN method was considered as a potential confounding factor that may influence the reliability of the predictive models being developed. Hence the pour plate (ISO 2001) method was also used for time series sampling of shellfish. This also allowed a comparison between the two methods.

This section reports results of comparison of two aspects of potential differences in *E. coli* values obtained by the two methods. These were (a) variability in results, focused on within sample measurement uncertainty, and (b) differences in *E. coli* concentrations for single samples measured using both methods. The implications of observed differences are considered in the context of potential to influence classification outcomes, as well as potential for improvement of predictive modelling of *E. coli* in shellfish using environmental predictors. In Section 6 of this report, data from the two methods are used to assess the effect on precision of predictive modelling of *E. coli* based on environmental variables.

Table 5.1. Outline of the classification system in place for UK bivalve production based on the health standards set out in Annex III of European Community Regulation 853/2004 and Articles 53, 54 and 55 of European Commission Regulation 2019/627.

Classification	Samples Required	*E. coli* Level Limits	Post-harvesting Treatment Options
A	Generally monthly (minimum of 10 samples per annum)	80 % of samples must be ≤ 230 *E. coli*/100 g. No result > 700 *E. coli*/100 g.	1. Shellfish can be harvested directly for human consumption.
B	Generally monthly (minimum of 8 samples per annum)	90 % of samples must be ≤ 4,600*E. coli*/100 g. No result > 46,000 *E. coli*/100 g.	1. Purification in an approved establishment. 2. Relaying for at least one month in a Class A relaying area. 3. An EC approved heat treatment process
C	Generally monthly (minimum of 8 samples per annum)	All samples ≤ 46,000 *E. coli*/100 g.	1. Relaying for at least 2 months in an approved Class B relaying area followed by treatment in an approved purification centre. 2. Relaying for at least 2 months in an approved class B relaying area. 3. An EC approved heat treatment process.
Prohibited		Results > 46,000 *E. coli*/100 g.	Shellfish from areas with consistently prohibited level results must not be subject to production or harvested.

5.3 Methods

5.3.1 Site/sample processing

Routine fortnightly sampling of bivalve shellfish beds was conducted in the Camel Estuary, from August 2020 to August 2021. This included samples from three blue mussel and four pacific oyster production areas (Figure 5.1). Bivalve shellfish samples were collected by Porthilly Shellfish, stored in food grade plastic bags in chilled conditions and transported to the laboratory at School of Ocean Sciences at Bangor University, maintaining official control shellfish hygiene monitoring requirements for sampling and transport to the laboratory (FSA, 2017). A minimum of 15 mussels or 12 oysters were processed for the enumeration of *E. coli*. Upon receipt, the temperature of the samples was recorded to ensure that they were between 2 and 10 °C. Samples were processed within 48 hours from collection on the shore.

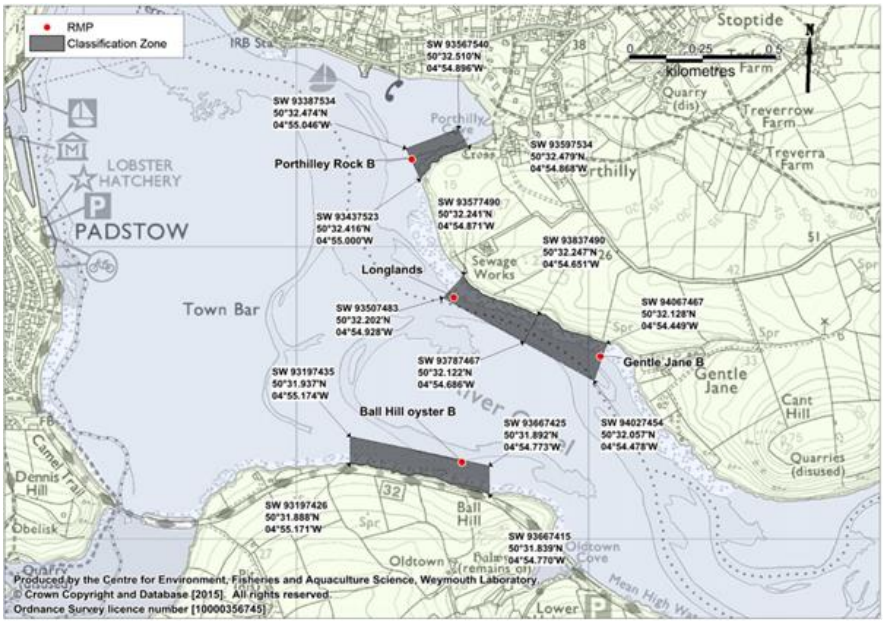


Figure 5.1 Location of recommended monitoring points (RMPs) within the shellfish beds are indicated by the red dots on the map. Shellfish beds were sampled from August 2020 – August 2022. Taken from the 2015 Camel Sanitary Survey (CEFAS, 2015)

All shellfish were processed according to the EURL ISO accredited reference method most probably number (MPN) method (CEFAS 2019; ISO 2016a) with some minor modifications to the dilution steps to allow for the pour plate ISO accredited method (EURL, 2014; ISO 2001) for *E. coli* enumeration in shellfish to be run in parallel (Figure 5.2). Briefly, the reference MPN method dilutes flesh and liquor by adding 2 ml of 0.1 % bacteriological peptone water (0.1 % PW) per gram of shellfish flesh, however samples were initially diluted with 1 ml of 0.1 % PW per gram of shellfish flesh before homogenisation as directed by the standard method for the pour plate (EURL, 2014). Once aliquots of the sample had been taken for use in the pour plate assay, the master 10^{-1} dilution was made by adding 80 ml of 0.1 PW to 20 ml of the 1:1 diluted sample.

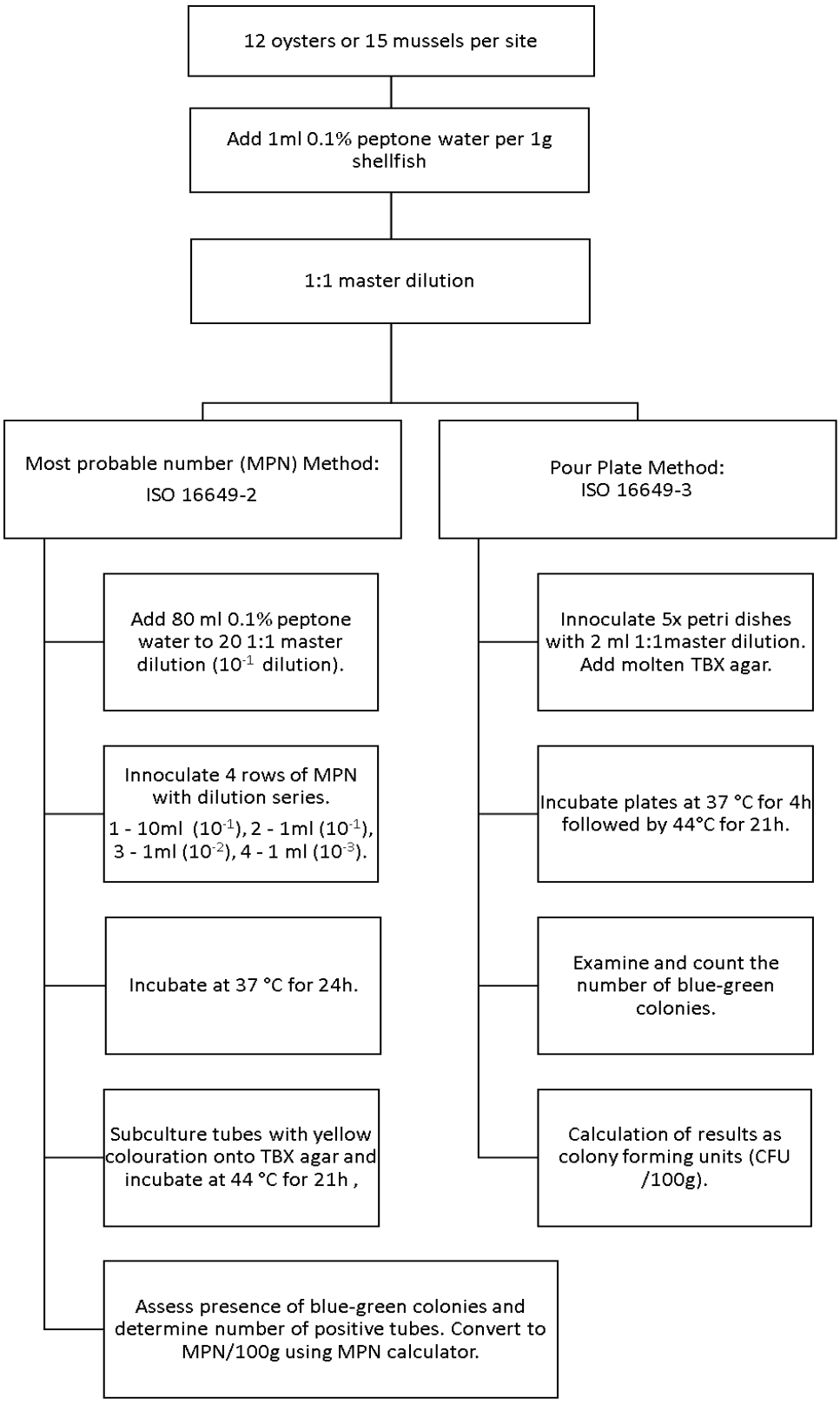


Figure 5.2 Flow chart summarising the laboratory process for the assessment of the concentration of *E. coli* in a single shellfish sample according to the ISO methods for both the pour plate and MPN methods. A 1:1 dilution of shellfish to 0.1 % peptone water was prepared using all shellfish flesh and liquor from the 12 oysters or 15 mussels and used for the pour plate method before further dilution to 10^{-1} for use in the MPN method.

Shellfish samples were processed via the four-dilution MPN format. For this, the master 10^{-1} dilution was serially diluted a further two times to create a 10^{-2} and 10^{-3} dilution for the MPN series. Each dilution series was inoculated into 5 tubes containing minerals modified glutamate broth before incubation at 37 ± 1 °C for 24 ± 2 hours. After incubation, positive MPN tubes were determined by examining whether they displayed yellow colouration that would indicate acid production. Confirmation of the presence of *E. coli* in the positive tubes was performed by subculturing onto tryptone bile X-glucuronide (TBX) agar using a sterile 1 µl loop before incubation at 44 ± 1 °C for 21 ± 3 hours. The subcultures were then assessed for the presence of blue/green colonies, confirming the presence of *E. coli*. The number of positive tubes at each dilution was then used to calculate the MPN of *E. coli* per 100 g of the sample using the calculation tool created by Jarvis et al (2010), which also generates 95% confidence interval for each sample.

5.3.1.1 Pour Plate Method

The pour plate technique involved inoculating 5 separate sterile Petri dishes with 2 ml of the 1:1 dilution of the shellfish sample before adding molten TBX agar held at a maximum temperature of 47 °C to prevent solidification of the agar before adding to the sample. The agar was removed from the water bath and decanted twice before being added to the sample at a temperature of 44 °C. The dishes were lightly mixed to combine the sample and the agar before incubation at 37 ± 1 °C for 4 hours as a resuscitation step. After this step they were transferred to an incubator at 44 °C for a further 21 ± 3 hours. The number of blue/green colonies were recorded across all 5 plates and an estimate of *E. coli* cells per 1 g shellfish sample was calculated by multiplying first by the dilution factor. This can be expressed by the equation:

$$N = (\sum c / V) \times tv$$

Where ∑c is the sum of the blue/green colonies across all 5 plates, V is the total volume in ml of the inoculum across the 5 dishes (10 ml) and tv is the total volume of the sample added to each plate (2 ml). This gives the number of colony-forming units (cfu) per 1 g of shellfish flesh. This number is then multiplied by 100 to give the cfu/100 g of shellfish flesh.

5.3.1.2 Estimation of measurement uncertainty confidence intervals for the pour plate method

Measurement uncertainty confidence intervals were calculated for each pour plate result, based on the individual plate count data for each sample, using the same statistical assumptions about the distribution of *E. coli* in the shellfish samples as are used in the MPN statistical calculations. Shellfish flesh is assumed to be infected with *E. coli* having concentration μ cfu/g. The cfus are assumed to be independently Poisson distributed. That is, a count μ of the number of cfus in a 1 g sample is Poisson distributed with mean μ.

In our laboratory analysis, counts μ are made of cfus in each of five 1 g samples, so that an estimate of μ is:

$$\hat{\mu} = \frac{\sum_{i=1}^{5} x_i}{5}; E(\hat{\mu}) = \mu \qquad (0.1)$$

Using the assumption of a Poisson distribution for each μ, the variance of each is also μ (since for the Poisson distribution the mean and variance are both μ). The variance of μ is then:

$$\mathrm{var}(\hat{\mu}) = \frac{\sum_{i=1}^{5} \mathrm{var}(x_i)}{25} = \frac{\mu}{5} \qquad (0.2)$$

We can approximate

$$\hat{\mathrm{var}}(\hat{\mu}) = \frac{\hat{\mu}}{5} \qquad (0.3)$$

From this we can compute an approximate confidence interval for μ.
To avoid a zero lower confidence limit we use the method of Jarvis *et al*, (2010).

$$\hat{\sigma}^2_{\ln\hat{\mu}} = \hat{\mathrm{var}}(\ln\hat{\mu}) = \hat{\mathrm{var}}(\hat{\mu})/\hat{\mu}^2 \qquad (0.4)$$

Hence, the interval:

$$\left[\ln\hat{\mu} - 2\hat{\sigma}_{\ln\hat{\mu}}, \ln\hat{\mu} + 2\hat{\sigma}_{\ln\hat{\mu}}\right] \qquad (0.5)$$

is an approximate 95% confidence interval for $\ln\mu$ and:

$$\left[\hat{\mu}\exp\left(-2\hat{\sigma}_{\ln\hat{\mu}}\right), \hat{\mu}\exp\left(+2\hat{\sigma}_{\ln\hat{\mu}}\right)\right] \qquad (0.6)$$

is an approximate 95 % confidence interval for the concentration μ.

Multiplication of $\hat{\mu}$ and the confidence limits by 100 will then give an estimate of cfu/100 g and its 95 % confidence limits.

5.3.1.3 *Comparison of variability of replicate and repeat measured* E. coli *results from MPN and pour plate methods*

The variability of both methods was compared using samples from the Gentle Jane mussel and oyster production bed. Firstly, for 3 sampling dates, on 02/02/2020, 16/02/2020 and 01/03/2020, triplicate samples each of 15 mussels or oysters were separately homogenised before being split into 3 aliquots. Each aliquot was tested for their *E. coli* concentrations by both the MPN and the pour plate methods. Hence for each date, three replicate samples were assayed as triplicate subsamples, allowing comparison of both variability between repeat samples taken on the same date and of measurement uncertainty within single samples. For this experiment, the standard deviations were calculated at the sample and subsample level to provide an indication of the measurement uncertainty and the variability between samples. Secondly, the comparison of measurement uncertainty was extended, comparing triplicate subsamples of single samples only (without the triplicate repeats). For all data where triplicate subsamples were analysed, variability across the observed range of *E. coli* levels was investigated by linear regression of the standard deviation of the triplicate results against the mean value for each sample.

In this section, which investigated variability of the two methods, the full range of *E. coli* results were included. This allowed comparison of variability between the two methods, which is not possible if low values are either excluded or assigned a fixed value (eg 100 or 200 *E.* coli/100 g). However, it should be noted that, for the pour plate method, values <200 *E. coli*/100 g are not approved for use in application of the official control regulations for shellfish hygiene.

5.3.1.4 *Statistical comparison of* E. coli *results, as practically applied, from MPN and pour plate methods*

All data were compared for *E. coli* from both MPN and pour plate methods for samples collected from August 2020 - August 2021, using similar approach to some aspects of previous cross-validation of the two methods (Mooijman *et al,* 2007; Pol-Hofstad and Jacobs-Reitsma, 2021). With pour plate values <200 *E.* coli/100 g set to = 100, the data were subjected to an orthogonal regression and following ISO (2016b) a visual interpretation of bias and extreme results was undertaken (Mooijman *et al,* 2007; Pol-Hofstad and Jacobs-Reitsma, 2021). The inter-comparability of the two methods was assessed by inspection of the 95 % confidence intervals for the intercept and slope of the fitted regression line relative to 0 and 1, respectively. A Bland-Altman plot of the data was also produced, with the differences in values for paired samples ($\log_{10}$ MPN *E. coli*/100 g – $\log_{10}$ Pour Plate *E. coli*/ 100 g) plotted against the $\log_{10}$ pour plate *E. coli*/100 g for *E. coli* for each sample.

To assess significant differences between the *E. coli* values from the MPN and the pour plate methods, the paired values reported by each method over the year-long comparisons between the MPN and the pour plate were subjected to both parametric paired t-test and a non-parametric Wilcoxon paired sample test. This analysis was performed for each site and for the combined samples across all production areas. For this analysis we explored the potential influence of low pour plate results (<200 *E. coli*/100 g) on the statistical comparison between the two methods. Low results were treated in three ways: a) included b) excluded or c) set to a midrange value of 100 *E.*

coli/100 g. This was intended to allow exploration of the differences between the methods, with consideration of how inclusion/exclusion of such low results might affect the comparison.

Further investigation of the statistical significance of differences between *E. coli* results from the two methods included consideration of the 95 % confidence interval data reported for individual MPN and pour plate samples. Simple binomial tests were applied to investigate the probabilities of the occurrence of the observed pour plate results relative to a null hypothesis of equality with the MPN results (0.5 probability MPN).

5.3.1.5 Comparison of MPN and pour plate measurements of E. coli *in experimentally spiked samples*

A laboratory-based experiment was conducted to compare the MPN method and pour plate methods at known (experimentally spiked) *E. coli* concentrations. Mussels (n = 30) were collected from the Menai Strait before being placed in a small-scale depuration unit for 7 days to ensure that any *E. coli* present in the mussels had been removed. Upon removal from the depuration unit the mussels were shucked, and the flesh and liquor were decanted into a sterile beaker. One ml of 0.1 % peptone water was then added per gram of shellfish before homogenisation using a blender (EURL, 2014). The shellfish homogenate was then split equally into six batches, one for a negative control. The remaining five batches had varying amounts of *E. coli* K12 culture added. Each batch was then split equally into 3 sub-batches.

E. coli K12 (LZB 035), supplied by Blades Biological (Kent, UK), was cultured overnight in Luria-Bertani Miller's medium (LB) (Miller, 1972). The optical density of the culture at 600 nm was measured using a spectrophotometer to estimate the concentration of *E. coli* cells. The culture was then serially diluted in 0.1 % peptone water to reach concentrations that were appropriate for the spiking of the shellfish homogenate. Diluted K12 culture was added to the beakers at the targeted concentrations of *E. coli* per 100 g of shellfish flesh at 50, 150, 300, 1000, 2000 and 5000 *E. coli* per 100 g, with each concentration run in triplicate. To verify the concentration of *E. coli* in the culture, the culture was serially diluted and filtered through a 0.45 µm cellulose nitrate membrane filter and placed on harlequin agar before incubation for 24 hours at 37 °C.

Once the spike was added to each beaker, the homogenates were blended again before they were processed for subsequent enumeration using the MPN and pour plate methods. One set of control triplicates was spiked with 0.1 % peptone water only to ensure that the depuration had successfully reduced the *E. coli* in the mussels to an undetectable level.

5.3.1.6 Comparison of the potential differences in shellfish area classification based on MPN and pour plate results

All shellfish (mussels and oysters) from the routine two-weekly sampling between August 2020 to August 2021 were homogenised first, the sample split to enable analysis using both MPN and pour plate methods and then assayed for *E. coli* in parallel (as described in 5.3.2.3., above). The proportion of observed results from the two methods which fell above or below the various

classification thresholds (230, 700, 4,600 *E. coli*/100 g – see Table 5.1), and the proportion of samples where the results from the two methods fell into different classification thresholds were compared. To extend this to more generalised consideration of classification outcomes across other shellfish sites, a Monte Carlo simulation was used to generate probabilities of the two methods generating values which differed across classification threshold levels.

5.4 Results

5.4.1 Comparison of variability of both replicate and repeat measure *E. coli* results from MPN and pour plate methods

Figures 5.3 and 5.4 show the *E. coli* concentrations in oysters and mussels from the Gentle Jane shellfish bed, measured using the MPN and pour plate methods, with triplicate samples collected in each week and repeat measurement of triplicate sub-samples for each sample. Generally, the variability of sub-samples was greater for MPN than for pour plate results (Tables 5.2 & 5.3). In some cases, results from repeat MPN measurements of sub-samples from a single sample ranged by an order of magnitude (e.g. <2000 - 30,000 or 300 – 3000 *E. coli*/100g). In contrast, pour plate values were more consistent with much smaller ranges across the three replicate measures, particularly at higher *E. coli* levels in the range tests. This difference between the methods was less marked at lower *E. coli* levels where some pour plate samples showed similar ranges of values to those for MPN.

In some sampling weeks, for both oysters and mussels, the variability between samples for MPN values was less than the variability within samples (standard deviation of the sample means vs standard deviation of the replicates). For pour plate results, the variability between sets of samples for a given day was generally higher than the internal variability of individual samples, suggesting greater potential for reliably identifying differences between samples.

The standard deviations for replicate measurements of a single MPN sample had a wider range (0.1 – 0.6 $\log_{10}$ *E. coli*/100 g) than the range of estimated standard deviations generated by the statistical calculations associated with each individual test (0.20 – 0.26 $\log_{10}$ *E. coli*/100 g).

For the longer-term comparative samples, the means and range of MPN and pour plate triplicate samples from the Gentle Jane oysters and mussels production sites are shown in Figures 5.5 and 5.6. For this data set, the differences in variability of the two methods increased across the range investigated, for both mussels and oysters. Figure 5.7 shows the standard deviations of the log transformed *E. coli* counts from the triplicated data plotted against the mean value for each sample. The observed variability was consistently higher across the range of mean values for the MPN method, with a slight increase in precision with increasing *E. coli* levels. In contrast, the precision of pour plate results increased markedly with increasing *E. coli* levels, which is not unexpected as the lower end of the tested range was below the recommended limit of detection for the method when applied in shellfish (200 *E. coli*/100 g, EURL 2014). Note that values below this threshold are included to allow investigation of differences in variability, but should be considered estimates rather than true counts.

Table 5.2 Standard deviations observed in log $_{10}$ transformed *E. coli* results within samples and between samples in mussels taken from the Gentle Jane shellfish production area in the Camel Estuary.

		Standard Deviation	
Method	Sampling Week	Within Samples	Between Samples
MPN	1	0.205	0.087
MPN	2	0.182	0.126
MPN	3	0.282	0.142
Pour Plate	1	0.074	0.038
Pour Plate	2	0.018	0.156
Pour Plate	3	0.039	0.116

Table 5.3. Standard deviations observed in log $_{10}$ transformed *E. coli* results within samples and between samples in oysters taken from the Gentle Jane shellfish production area in the Camel Estuary.

		Standard Deviation	
Method	Sampling Week	Within Samples	Between Samples
MPN	1	0.210	0.092
MPN	2	0.176	0.147
MPN	3	0.327	0.153
Pour Plate	1	0.106	0.114
Pour Plate	2	0.047	0.135
Pour Plate	3	0.058	0.067

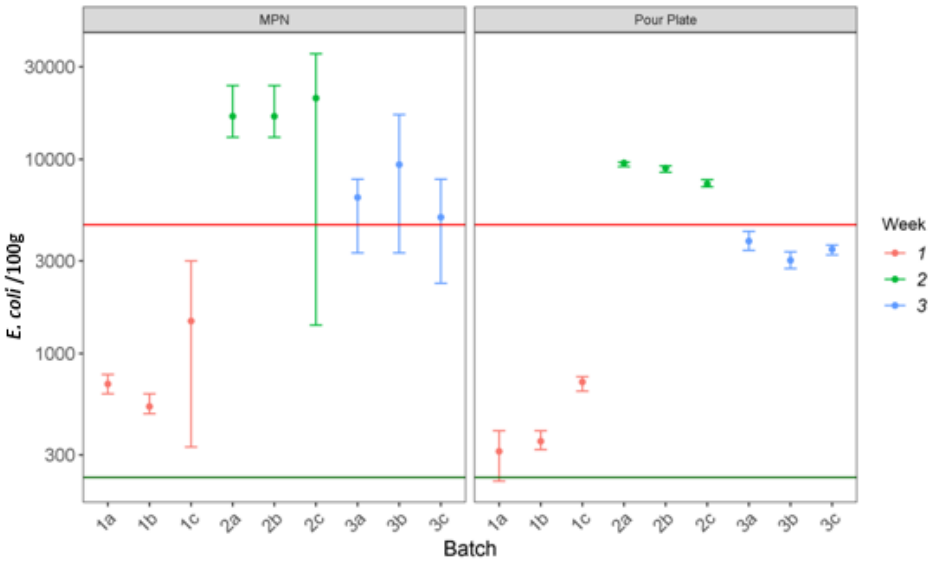


Figure 5.3 The mean *E. coli* concentrations reported by the pour plate and MPN methods. The vertical lines shows the range of values (± the highest and lowest) observed for each batch in mussels taken at the Gentle Jane RMP. Each week, three samples were collected and homogenised before being separated into 3 aliquots. Each aliquot was processed for enumeration of *E. coli* using both the MPN and pour plate methods. Green horizontal line indicates the classification boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification boundary at 4600 *E. coli*/100 g. Note that pour plate values <200 *E. coli*/100 g should be considered estimates rather than true counts.

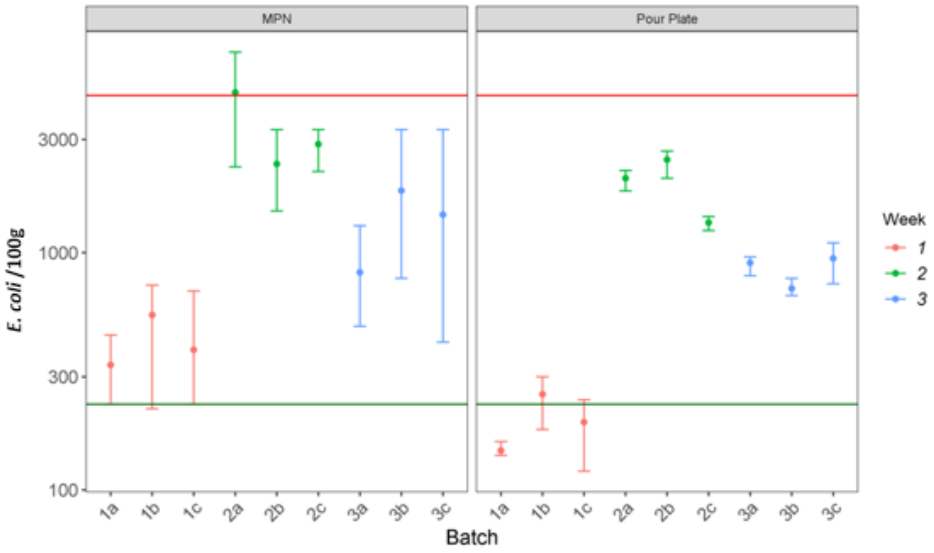


Figure 5.4. The mean *E. coli* concentrations reported by the pour plate and MPN methods. The vertical lines shows the range of values (± the highest and lowest) observed for each batch in oysters taken at the Gentle Jane RMP. Each week, three samples were collected and homogenised before being separated into 3 aliquots. Each aliquot was processed for enumeration of *E. coli* using both the MPN and pour plate methods. Green horizontal line indicates the classification boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification boundary at 4600 *E. coli*/100 g. Note that pour plate values <200 *E. coli*/100 g should be considered estimates rather than true counts.

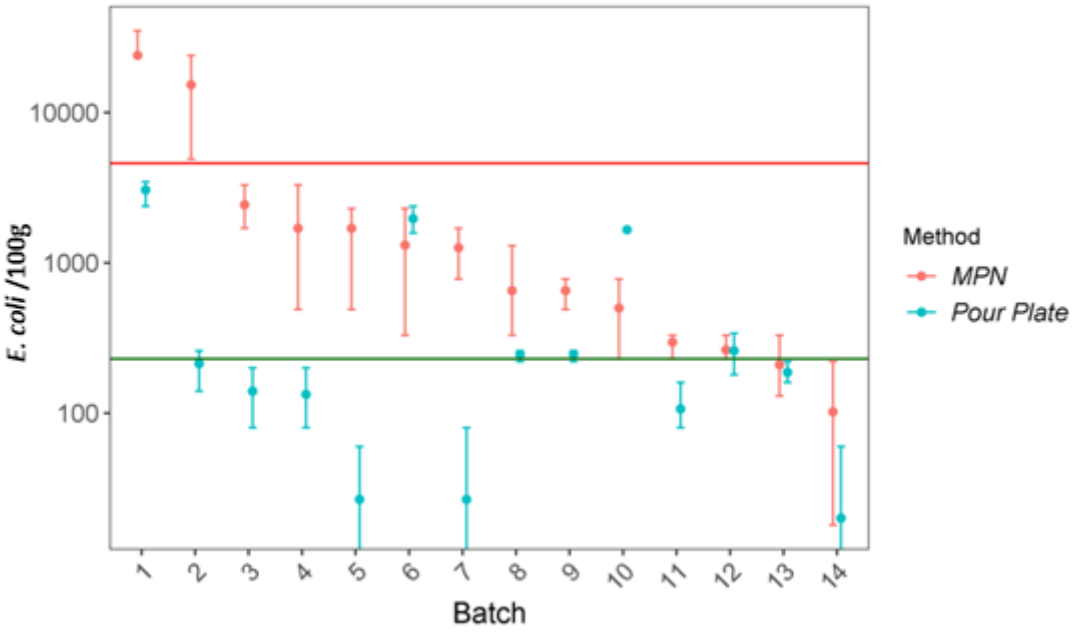


Figure 5.5 Comparison of *E. coli* counts in oysters at the Gentle Jane oyster site, from MPN and pour plate methods. Each batch represents paired data from a single sample, split and measured in triplicate by both methods. The vertical lines shows the range of values (± the highest and lowest). Batches have been placed in order of the highest mean MPN result to lowest mean MPN results observed. Green horizontal line indicates the classification boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification boundary at 4600 *E. coli*/100g. Note that pour plate values <200 *E. coli*/100 g should be considered estimates rather than true counts.

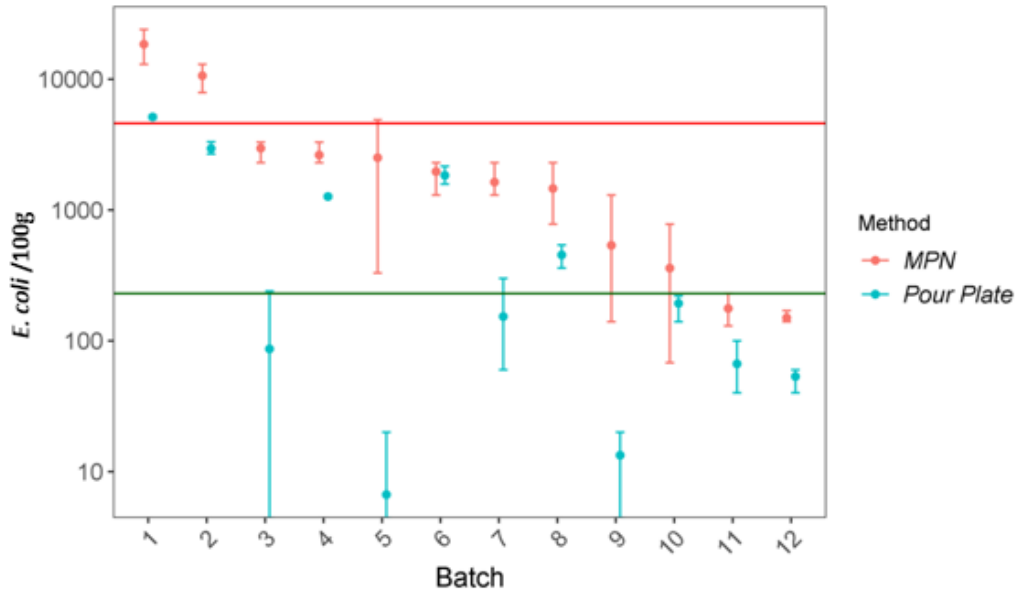


Figure 5.6 Comparison of *E. coli* counts in mussels at the Gentle Jane site, from MPN and pour plate methods. Each batch represents paired data from a single sample, split and measured in triplicate by both methods. Values are means and the vertical line shows ± highest and lowest values. Batches have been placed in order of the highest to lowest mean MPN results observed. Green horizontal line indicates the classification boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification boundary at 4600 *E. coli* /100 g. Note that pour plate values <200 *E. coli*/100 g should be considered estimates rather than true counts.

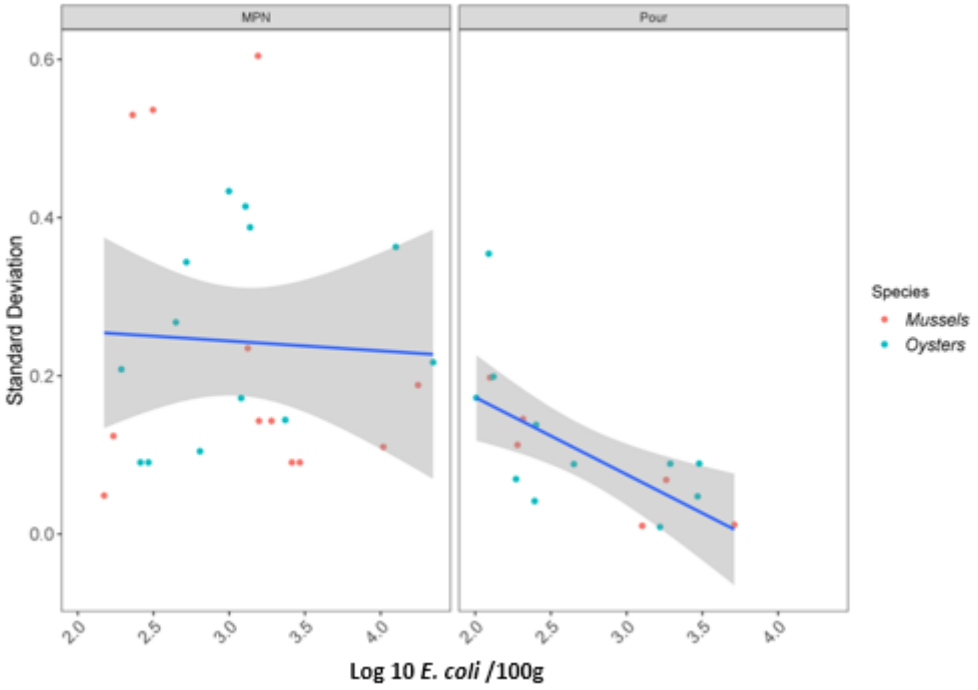


Figure 5.7 Standard deviation of triplicate measurements for shellfish samples from the Gentle Jane bed, plotted against mean *E. coli*/100 g for each sample. Data are log $_{10}$ transformed. Note that pour plate values <2.3 log $_{10}$ *E. coli*/100 g should be considered estimates rather than true counts.

5.4.2 Comparison of *E. coli* results, as practically applied, from MPN and pour plate methods

Figure 5.8 shows a summary of *E. coli* results from the MPN and pour plate methods from across all the shellfish beds in the Camel, sampled over a 12-month period from August 2020 - August 2021. As pour plate results of < 200 *E. coli*/100 g are not validated for use in shellfish official control, all pour plate results of < 200 *E. coli*/100 g are included = 100 *E. coli*/100 g, to replicate the treatment of *E. coli* values below the limit of quantification by the MPN method. Visual inspection of these plots shows a tendency for higher MPN than pour plate results across all beds and for both mussels and oysters. The seasonal trend in results from the two methods can be seen to follow similar patterns, though with some periods/beds where low pour plate results are not reflected in similar drops in MPN values.

The regression plot (Figure 5.9) of the same data set for log transformed MPN and pour plate results of both mussels and oysters from all the shellfish beds in the Camel, indicates that the MPN and pour plate methods are not comparable; whilst the intercept does not differ significantly from zero, the slope of the line differs significantly from 1 as the confidence interval for the slope do not include 1 (Table 5.4).

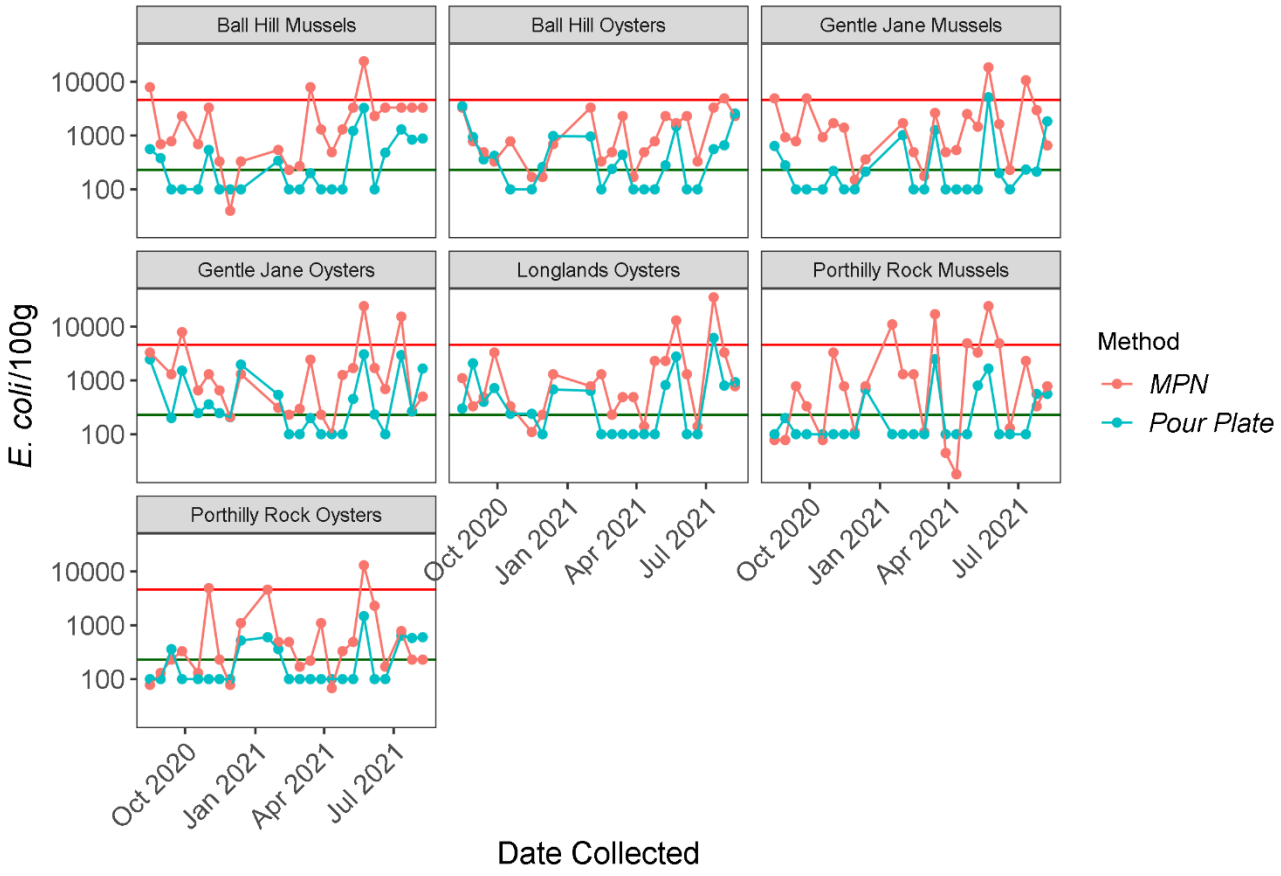


Figure 5.8. *E. coli* concentrations determined by the MPN and pour plate methods in shellfish samples across all seven shellfish sites in the Camel shellfish production area from August 2020-August 2021. Each sample was processed and split for enumeration of *E. coli* by the pour plate and MPN. Green horizontal line indicates the classification boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification boundary at 4600 *E. coli*/100 g. In these plots, all pour plate values are included, with results <200 *E. coli*/100 g included as = 100 *E. coli*/100 g.

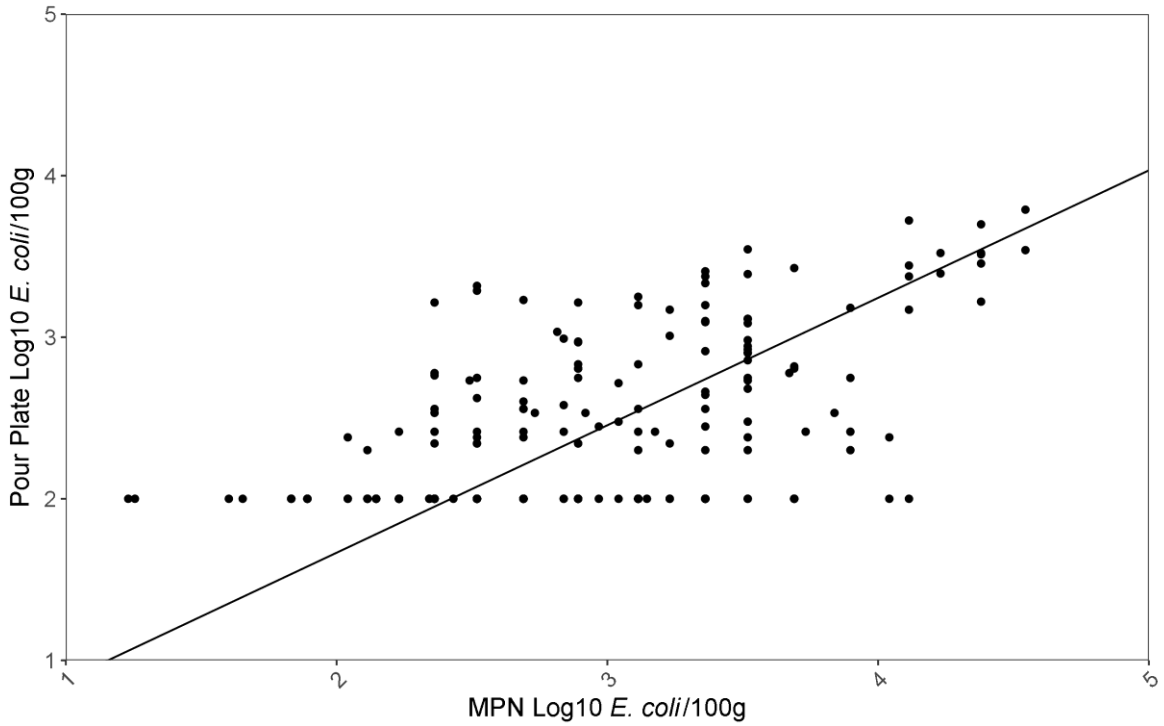


Figure 5.9 Regression of log transformed *E. coli* values from MPN and pour plate methods for all shellfish samples from the Camel shellfish production area from August 2020 - August 2021. Values of < 200 *E. coli* /100 g for the pour plate method included as = 100 *E.* coli/100g

Table 5.4 Results of regression of log_{10} transformed *E. coli* values from MPN and pour plate methods for all shellfish samples sites in the Camel shellfish production area from August 2020 - August 2021.

Predictor	**Coef**	**SE Coef**	**Z**	**P**	**95% CI**
Constant	0.088	0.245	0.358	0.720	(-0.0393, 0.568)
Log_{10} Pour Plate	0.789	0.082	9.587	<0.001	(0.628, 0.95)

Figure 5.10 shows the Bland-Altman plot for differences between pairs of *E. coli* results for each method plotted against the values for the mean value for both methods. Inspection of the spread of data shows relatively few values (5) that exceed the 95% confidence intervals. However, there is an increase in the (negative) differences with increasing counts and the confidence interval for the fitted regression line only cross the x-axis at the lower end of pour plate values. This indicates that the MPN method generated higher results over most of the range of pour plate values investigated.

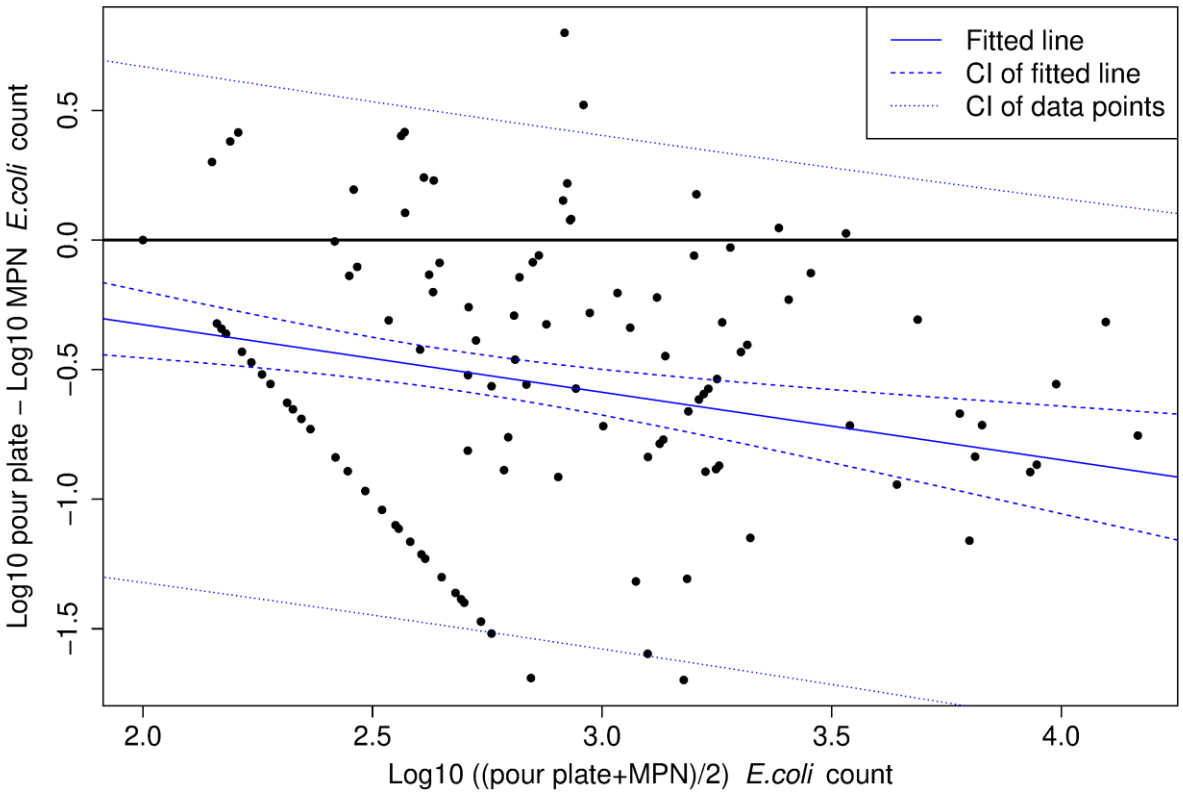


Figure 5.10 Bland-Altman plot of differences between paired log $_{10}$ *E. coli* results between the two methods against the mean log $_{10}$ *E. coli* results for the two methods. The dotted lines show the 95 % confidence range of the data for differences between the methods, the dashed lines show the 95 % confidence range for the fitted regression line. Data pairs where the pour plate result was < 200 cfu/100 g were set to 100 cfu/100g. These points appear as a diagonal line of dots across the figure.

These interpretations prompted further statistical investigation of the results from the two methods. To test differences between data sets with paired results measured by two methods from single samples, an alternative and more appropriate statistical analysis might be a paired t-test for samples with unequal variance (Welch) for normally distributed datasets, or its non-parametric equivalent (Wilcoxon's signed rank test). Because the pour plate counts of fewer than 200 *E. coli*/100 g are not validated for use in official control sampling, we consider three approaches to comparing these lower pour plate values with paired MPN values. We omit all samples for which both MPN and pour point count are zero. Following this initial sift, we first omit all pairs having a pour plate count lower than 200 *E. coli*/100 g. This discards all information from these pairs, although values lower than 200 are likely to reflect low contamination in the analysed sample. The results are shown in Table 5.5a. Taking all beds and species together ("All") shows a clear rejection of the hypothesis that the methods give counts with the same statistical distribution. The 95 % confidence interval shown indicates higher MPN counts than pour plate counts. When the data are subdivided by bed and species, the results are less clear cut. For Porthilly, in particular, there is no statistical difference identified between the two methods. However, for this site, there are very few

pour plate counts above 200, so that poor discrimination between the two methods might be expected. In a second analysis (Table 5.5b) we analyse all pairs both MPN and pour plate counts non-zero. This therefore includes the uncertain non-zero pour plate counts below 200 *E. coli*/100 g. This gives generally highly significant differences between the two methods, even for individual bed and species combinations. Finally, in Table 5.5c, we show the analysis with all pour plate counts below 200 *E. coli*/100 g set to 100. This gives intermediate levels of significance, between the results of the preceding two techniques, but overall highly significant differences. It is common practice to set values less than the limit of detection (LoD) to half the LoD value, when analysing environmental data. This retains the information that these values are less than the LoD, while acknowledging that the measured values are not expected to be accurate. Clearly there will be cases where the value of half the LoD strongly misrepresents the true mean value of the sub-LoD values and, where this is important, some other statistical approach may be needed. However, all three methods of dealing with the pour plate values below 200 *E. coli*/100 g show overall highly significant differences between the MPN and pour plate counts, with the MPN counts being higher.

Statistical comparison between the pour plate and MPN *E. coli* results from the Camel sampling programme can be further extended to include the measurement uncertainty derived with each sample for both method assays. Within the MPN estimation procedure, on the assumption that the number of micro-organisms in the sample tubes follows a Poisson distribution, probabilistic arguments allow a confidence interval to be computed for each MPN estimate (e.g. Jarvis *et al,* 2010). Confidence intervals were also calculated for each pour plate results, as described in the Methods section. While not in general use in practice, these confidence intervals have some relevance in the comparison of MPN and pour plate counts. Figure 5.1.1 plots sorted MPN values, associated pour plate counts and the computed 95 % confidence limits for both. This shows that at the lower end of the range (values < 230 *E. col*/100 g, = 2.3 $\log_{10}$ *E. coli*/100 g), while pour plate results are often lower, there is generally some overlap of confidence intervals for the two methods. However, above 230 *E. col*/100 g (which is the range for which the pour plate methods is approved), overlap of confidence intervals is less frequent indicating that results in this range are contributing to statistical differences between the methods. Furthermore, at the higher values the confidence intervals for the pour plate results tended to be narrower than for MPN. This is even clearer in Figure 5.1.2., which focuses on those samples having an MPN count greater than 4600 *E. coli*/100 g, the class B/C boundary. This further illustrates the discrepancy between the two methods for these apparently more contaminated samples. Only two pour plate values fall within the MPN confidence limits, and only 3 pour plate values are above the 4600 *E. coli*/100 g threshold.

Table 5.5 a,b,c. Paired statistical comparisons of log $_{10}$ (MPN) and log $_{10}$ (PP) *E. coli* concentrations in all shellfish production areas in the Camel estuary in samples taken from August 2020 to August 2021. P = values < 0.05 indicate statistically significant differences, * indicates differences that are highly statistically significant.

Table 5.5a. Paired non-zero data, excluding pour plate results < 200 *E. coli*/100 g

Site	**t-test 95% CI**	**df**	**t-test p-value**	**Wilcoxon statistic**	**Wilcoxon p-value**
All	0.29-0.49	106	<0.0001	4925	<0.0001*
Ball Hill Mussels	0.48-1.03	11	<0.0001	78	<0.0005*
Ball Hill Oysters	0.07-0.53	14	0.015	93	0.064
Gentle Jane Mussels	0.33-0.84	17	<0.001	168	<0.0001*
Gentle Jane Oysters	0.10-0.49	31	0.004	409	0.007
Longlands Oysters	0.04-0.60	14	0.028	99	0.026
Porthilly Rock Mussels	-0.12-0.98	5	0.10	18	0.15
Porthilly Rock Oysters	-0.26-0.52	8	0.46	25	0.82

Table 5.5b. Paired non-zero data, with full range of results included

Site	**t-test 95% CI**	**df**	**t-test p-value**	**Wilcoxon statistic**	**Wilcoxon p-value**
All	0.53-0.70	195	<0.0001	17828	<0.0001*
Ball Hill Mussels	0.67-1.19	22	<0.0001	253	<0.0001*
Ball Hill Oysters	0.18-0.57	20	<0.001	197	0.003
Gentle Jane Mussels	0.63-1.03	40	<0.0001	849	<0.0001*
Gentle Jane Oysters	0.26-0.59	45	<0.0001	918	<0.0001*
Longlands Oysters	0.29-0.73	23	<0.0001	274	<0.001*
Porthilly Rock Mussels	0.41-1.04	19	<0.001	195	<0.001*
Porthilly Rock Oysters	0.28-0.86	20	<0.001	207	<0.001*

Table 5.5c. Paired non-zero data with pour plate results < 200 *E. coli*/100 g set to 100 *E. coli*/100 g

Site	**t-test 95% CI**	**df**	**t-test p-value**	**Wilcoxon statistic**	**Wilcoxon p-value**
All	0.44-0.60	195	<0.0001	17742	<0.0001*
Ball Hill Mussels	0.53-0.92	22	<0.0001	272	<0.0001*
Ball Hill Oysters	0.10-0.69	20	<0.001	204	0.002
Gentle Jane Mussels	0.50-0.86	40	<0.0001	738	<0.0001*
Gentle Jane Oysters	0.26-0.57	45	<0.0001	929	<0.0001*
Longlands Oysters	0.24-0.66	23	0.0002	271	<0.001*
Porthilly Rock Mussels	0.29-0.87	19	<0.001	186	0.003
Porthilly Rock Oysters	0.10-0.62	20	0.009	186	0.015

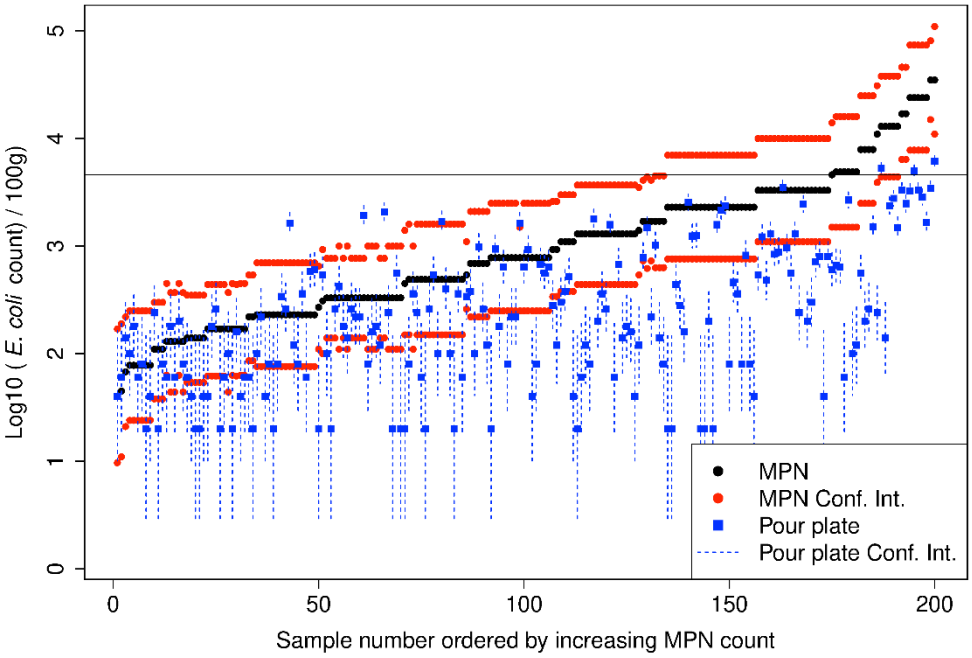


Figure 5.1.1 Comparison of *E. coli* concentrations determined by MPN and pour plate counts, including individual sample confidence limits, for shellfish samples collected across all seven shellfish sites in the Camel shellfish production area.

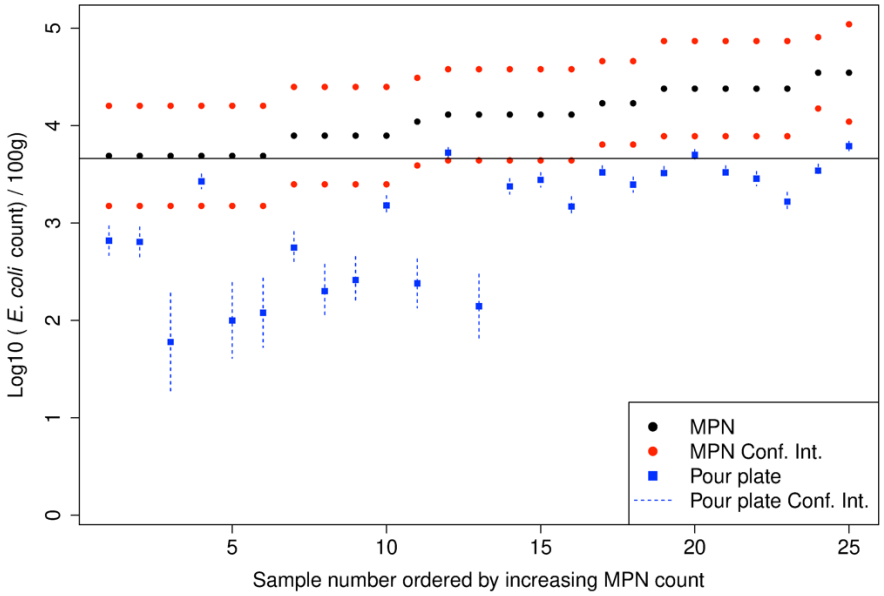


Figure 5.1.2 Comparison of *E. coli* concentrations determined by MPN and pour plate counts, including individual sample confidence limits for shellfish samples collected across all seven shellfish sites in the Camel shellfish production area, where MPN *E. coli* counts were greater than 4600/100 g

In addition to Welch's t-test and Wilcoxon's signed rank test, simple, more intuitive but less powerful binomial tests can formally be carried out on these data. Assuming MPN and pour plate values have the same mean value, the probability that the MPN count for a particular sample is

greater than the pour plate count is 0.5. This event occurs 192 times out of 225 in our samples. Using a binomial test, the null hypothesis that the probability is 0.5 is rejected with p-value < 0.0001. Also, we would expect equal probabilities of pour plate counts being above or below the respective MPN confidence limits. The counts for our samples are 3 and 140, with a p-value of < 0.0001 from the equivalent binomial test.

5.4.3 Comparison of MPN and pour plate measurements of *E. coli* in experimentally spiked samples

E. coli were measured by the MPN and pour plate methods for depurated shellfish samples that had been spiked with *E. coli* K12 at levels intended to be representative of the range of classification values (A-C) observed in the Camel time series samples. All samples processed by the MPN gave results that were higher than the intended spiked level of *E. coli* whereas the same samples processed by the pour plate returned results lower than the intended spiked level. (Figure 5.1.3).

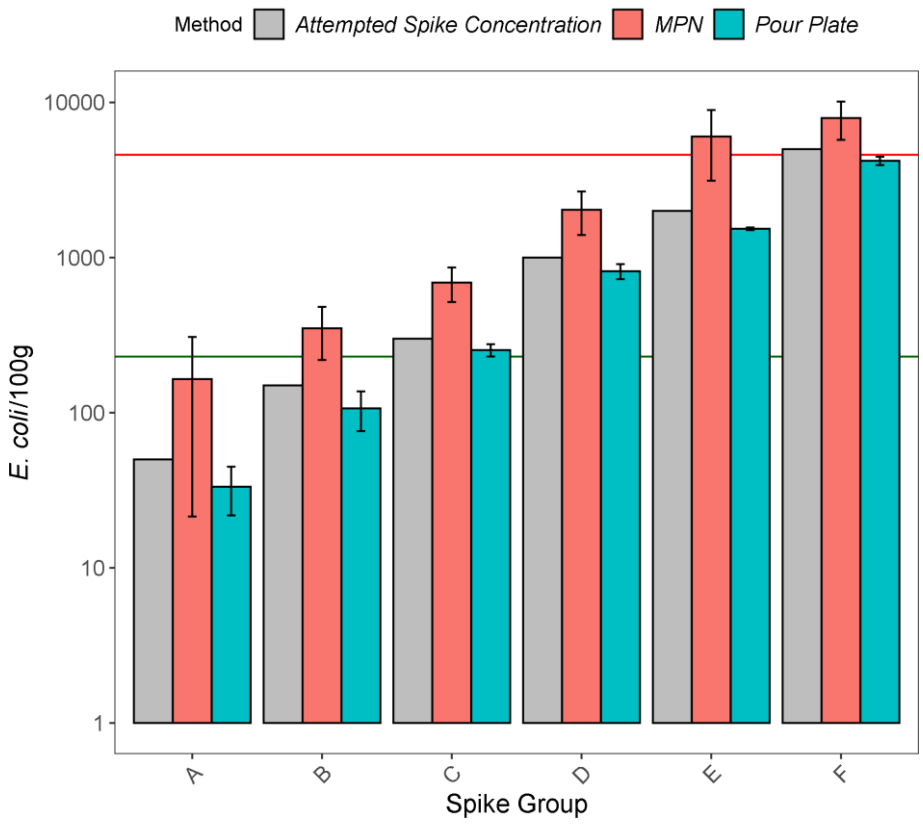


Figure 5.1.3 Comparisons of *E. coli* concentrations reported by the MPN and pour plate to the target *E. coli* concentration in a homogenised shellfish sample. At each spike level 3 batches of mussel homogenate were spiked with a known concentration of *E. coli* k12 and then processed for enumeration via both the MPN and the Pour Plate method. Green horizontal line indicates the classification class A/B boundary at 230 *E. coli*/100 g. Red horizontal line indicates the classification class B/C boundary at 4,600 *E. coli*/100 g. The intended spike concentrations of *E. coli*/100 g were as follows: A - 50, B – 150, C – 300, D – 1000, E – 2000, F – 5000. Note for the pour plate method, values < 200 *E. coli*/100 g should be considered estimates rather than true counts.

5.4.4 Comparison of the potential differences in shellfish area classification based on MPN and pour plate results

Inspection of results from the 12-month sampling of shellfish from the Camel beds shows that the frequency of results falling above/below the various classification boundaries differed between the two methods (Figure 5.1.4). The pour plate method gave a higher proportion of results < 230 *E. coli*/100 g, and a much lower proportion of results > 4600 *E. coli*/100 g results compared to the MPN method (1 result > 4600 *E. coli/*100 g vs 20).

The observed differences in *E. coli* results obtained by the MPN and pour plate methods indicates that a switch to the pour plate method for analysis of Official Control Regulations (OCR) samples has the potential to influence the overall classification of a shellfish production area. Since 2010, all mussel and oyster classification areas in the Camel have received a B or long-term B classifications. The *E. coli* concentrations reported by both the pour plate and the MPN from August 2020 to August 2021 conformed to the legislative requirements of a class B shellfish harvesting area. Across all beds during this period, no samples by either method returned values close to the maximum allowable *E. coli* concentration of > 46,000 *E. coli*/100 g, which no samples are permitted to exceed in a class B area, and any result of that magnitude would likely be reviewed as potentially anomalous (Figure 5.8). Classifications are determined by review of a three-year *E. coli* data set, for which in the present study we have only a 12-month subsample (though with 22 samples during that period). However, for three of the shellfish beds in the Camel, Ball Hill (mussels), Gentle Jane (oysters) and Longlands (oysters), 3 out of 22 samples measured by MPN exceeded 4600 *E. coli*/100 g. If this frequency of occurrence were extended over a three-year review period, it could potentially result in a C classification being determined. However, this would also depend on the outcome of the standard procedures for investigation of any individual high result, which is beyond the scope of the present study. In contrast, for the pour plate data collected for the Camel, no results exceeded 4,600 *E. coli*/100 g for any of the beds, suggesting that a C classification on that basis would be highly unlikely. At the lower end of the range, using the results of the pour plate over the course of the year would be very unlikely to lead to a change in classification of any bed from B to A. Whilst a greater proportion of the *E. coli* concentrations reported by the pour plate were within the class A boundaries, class B results were still regularly observed in all shellfish beds (Figure 5.8).

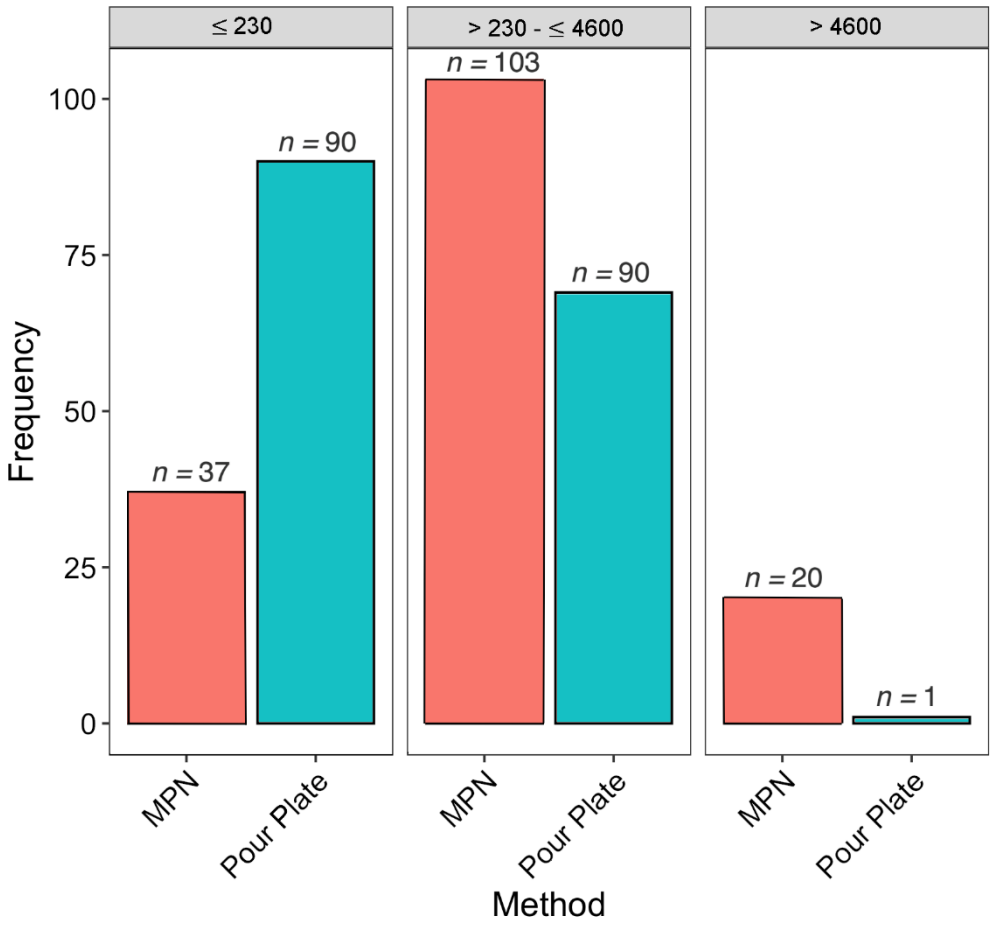


Figure 5.1.4 Comparison of values for MPN and pour plate *E. coli* results relative to classification thresholds, for all shellfish samples from the Camel production area from August 2020 - August 2021. Thresholds: < 230, 230 - 4,600 and > 4,600.

Considering where individual results fall relative to the classification thresholds (230, 230-4,600 and > 4,600 *E. coli*/100 g) can identify the proportion of results which could potentially contribute to a difference in classification outcomes for a production area (Table 5.6). Over half of all paired sample values fell within the same classification range. MPN results fell in a higher classification range than the corresponding pour plate in 41.1 % of samples. Conversely, pour plate results fell within a higher classification range in 3.7 % of samples.

Table 5.6. Frequency of agreement/disagreement relative to the classification thresholds between the MPN and pour plate methods for all shellfish samples from the Camel production area from August 2020 - August 2021. Frequencies of outcomes are given as a raw value and the relative frequency is given as a percentage. Thresholds: (230 - 4,600 and > 4,600 *E. coli*/100 g).

Outcome	Frequency (Percentage)
Results differ across a classification boundary (MPN higher)	74 (41.1 %)
Both tests fall within same classification level	100 (55.6 %)
Results differ across a classification boundary (pour plate higher)	6 (3.7 %)

The broader implications of the potential differences arising from applying the two methods to site classification was investigated by a more generalised probability model, using a Monte Carlo simulation (Figure 5.1.5) from the joint distribution of the $\log_{10}$ MPN and $\log_{10}$ pour plate *E. coli* counts. In this analysis, measured pour plate results < 200 *E.* coli/100 g were assigned a value of 100 *E.* coli/100 g. Assuming a linear relationship between these two variables, a bivariate normal distribution can be fitted, random paired values generated from that distribution, and the proportion of values falling in sectors of the distribution counted, particularly exceedances of classification thresholds.

Table 5.7 shows the probability of counts and proportion of observed results across three threshold values (230, 700, 4,600 *E. coli*/100 g). The second and third columns of probabilities refer to instances where the two methods lead to a different classification. Columns 5 and 6 show the greater probability of threshold exceedance using MPN counts. For the higher thresholds the simulated probabilities of a paired count falling within each sector are very close to measured proportions. The correspondence is a little poorer for the 230 threshold.

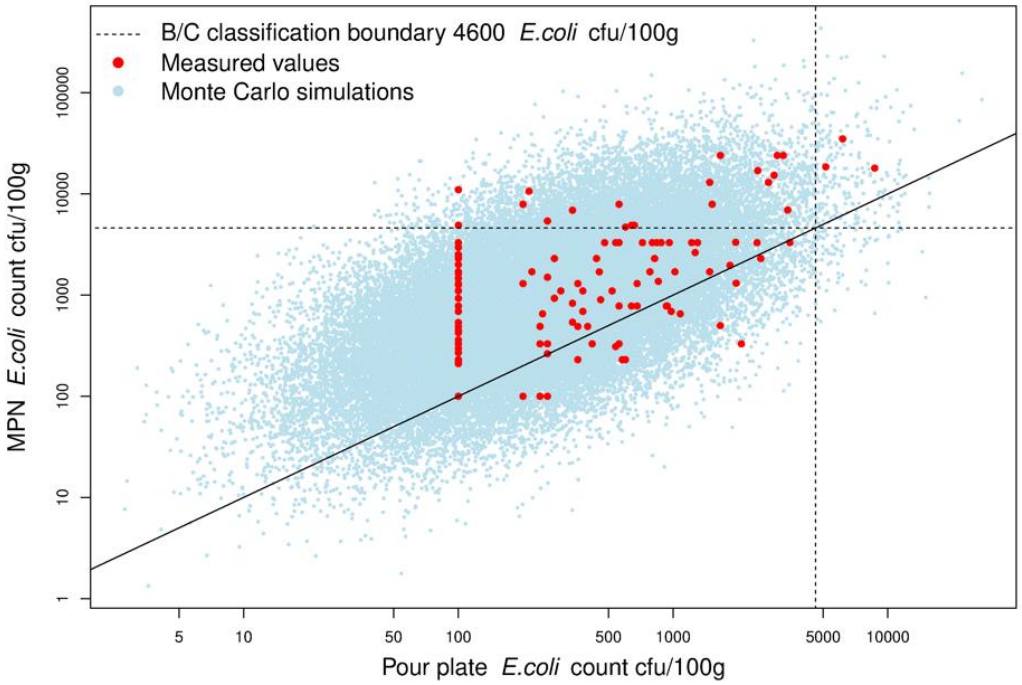


Figure 5.1.5 Monte Carlo simulated values (blue cloud points) for *E. coli* counts for the MPN and pour plate methods, as predicted from the relationship between measured values for mussels and oysters from the Camel estuary between August 2020 - August 2021. Note: In this modelling, measured pour plate results < 200 *E. coli* /100 g were assigned a value of 100 *E. coli* /100 g.

Table 5.7 Simulated probabilities (from the Monte Carlo simulation) and observed proportions of paired *E. coli* results by MPN and pour plate (PP) methods falling above or below classification thresholds 230, 700, 4600 *E. coli/100 g.*

Threshold	**Probability/proportion of outcomes of paired test results for the two methods relative to thresholds**					
	1 Both below	2 MPN below Pour plate above	3 MPN above Pour plate below	4 Both above	5 MPN above	6 Pour plate above
230 simulated	0.151	0.035	0.317	0.500	0.81	0.54
230 observed	0.205	0.027	0.346	0.422	0.77	0.45
700 simulated	0.421	0.024	0.386	0.169	0.55	0.19
700 observed	0.438	0.022	0.357	0.183	0.54	0.21
4600 simulated	0.872	0.002	0.121	0.005	0.122	0.006
4600 observed	0.865	0	0.119	0.016	0.135	0.016

5.5 Discussion

5.5.1 Method variability

Estimation of bacterial abundance in food and environmental samples is inherently variable. In the case of shellfish, *E. coli* variability is compounded across a range of sources and factors. Spatial variation across individual shellfish beds has been reported, which in part may be due to proximity to sources of contamination such as sewage discharges and dispersal plumes (Beliaeff and Cochard 1995; Kay *et al,* 2008). Hence, positioning of Representative Monitoring Points for OCR sampling typically takes into account location relative to likely contamination sources and areas of potentially highest contamination within a production area. However, some studies have also reported localised "hotspots" within a single shellfish bed that are not readily attributed to known sources of contamination and may in part be due to fine scale differences in tidal elevation and spatial differences in rates of bacterial uptake and elimination (Clements *et al,* 2015). Short-term temporal and longer-term seasonal variation in *E. coli* levels in shellfish may reflect a range of environmental variables, such as rainfall, river flow, temperature, water turbidity and sunlight, and it is variability at this level that is the focus of the DASSHH project (see Introduction to this report for review). This level of variation in *E. coli* levels in shellfish, responding to seasonal and environmental factors, is considered to be greater than variation between repeat samples within an area or measurement uncertainty for individual samples (Lee and Silk 2013; Walker *et al,* 2018). As with any laboratory analysis, an additional level of variability may be introduced by human operator or process differences in sample handing, storage, processing, assay performance and record keeping. The guidance given in the official protocols is designed to minimise these sources of variability through standardisation of processes and methods, and quality assurance monitoring via within-lab and inter-lab comparability testing. However, modelling by Gronewold and Wolpert (2008) for water samples indicates that differences in intra-sample variability between MPN and cfu methods was likely to arise from the intrinsic uncertainty introduced by the statistical method of calculation of MPN results, rather than human or process error.

The first part of this study focused on variability in laboratory measurement of *E. coli,* both within and between methods. In practical terms, this was largely based on individual samples split and analysed with two methods, i.e. estimating measurement uncertainty associated with each method. Post sampling variability in the laboratory may be derived from uneven distribution of bacteria in homogenised samples (whichever test method is applied) and potential for differences in growth of cells inoculated into test cultures, particularly relevant to the MPN method which uses multiple culture tubes for each sample. Development of statistical approaches to account for this intrinsic measurement uncertainty has been integral to the evolution of the MPN method over the > 100 years since its first application (McCrady, 1915). Hence, the MPN assay generates estimates of *E. coli* abundance derived from statistical calculations, with 95% confidence intervals and a rarity ranking score that helps identify erroneous results (Jarvis *et al*, 2010). This limitation to the precision of the MPN method is acknowledged, with the expanded uncertainty estimated at 0.66 ($\log_{10}$-transformed data) for shellfish samples (Lee and Silk, 2013; Walker *et al*, 2018), and in practical terms by the provision of recommended methods for laboratory determination of measurement uncertainty by laboratories (EURL/CEFAS 2021). However, in application of the official control

regulations, the classification of a shellfish production area is informed by results from a series of MPN results, for which the individual measurement uncertainty of each result is not directly considered.

5.5.2 Comparability of *E. coli* concentrations in shellfish samples

Comparison of two methods for measurement of *E. coli* abundance needs to take account the method uncertainty in both methods. The level at which variability can be considered may differ between methods and may also depend on the purpose of the comparison being made. For example, the pour plate method as applied in OCR shellfish testing does not generate an estimate of measurement error for each sample, as the colony counts from the replicate plates are summed in calculating the test result (EURL 2014). Assessment of measurement uncertainty is recommended in the official pour plate protocols, in terms of use of control tests and within-lab and inter-lab comparability testing (EURL 2014). In the present study, we have developed a method for calculation of confidence intervals for the pour plate method based in the replicate plate counts used in the test. The MPN method does provide an estimate of measurement error for each sample. For the MPN tests conducted, the average calculated standard deviation for each measurement was 0.24 $\log_{10}$ *E. coli*/100 g, which is consistent with the reported typical standard deviation for the method (0.24 $\log_{10}$ *E. coli*/100 g, Lee and Silk 2013; Walker *et al,* 2018). This compared reasonably with the range of standard deviations calculated for the pour plate data from the Camel samples (0.20 – 0.26 $\log_{10}$ *E. coli*/100 g), though the range of standard deviations was higher for MPN (0.1-0.6 $\log_{10}$ *E. coli*/100 g). The standard deviation of triplicate measurements from a single sample ranged from < 0.04 - 0.6 $\log_{10}$ *E. coli*/100 g for MPN and from 0.009 -0.35 $\log_{10}$ *E. coli*/100 g for pour plate, with an average standard deviation of 0.24 $\log_{10}$ *E. coli*/100 g for MPN and 0.10 $\log_{10}$ *E. coli*/100 g for pour plate. This wide range of observed standard deviations for MPN is unsurprising given the small sample number (= 3 replicates) (Walker *et al,* 2018) and guidance for experimental quantification of measurement uncertainty for MPN recommends a higher number (20) of replicate sub-samples (CEFAS 2021). However, the present data do allow some comparison of variability of the two methods, with the pour plate method yielding less variable results, particularly for the higher end of the range of *E. coli* concentrations measured. The high variability of replicate measures of single MPN samples, relative to the variability between mean values samples taken from the same bed on the same day, indicates the potential lack of discrimination in comparing repeat samples using this method and also the potential for random occurrence of high or low results that may influence interpretation or application of monitoring data. Over the range of values investigated, the difference in variability between the two methods seems to be of potentially greatest significance around the < 4,600 threshold, where high results could result in downgrade or closure of class B shellfish beds and the pour plate method provides more consistent/less variable results than MPN.

Previous comparisons of MPN and pour plate measurement of *E. coli* in shellfish, which followed ISO procedure for validation of new methods, compared data at the level of test results equivalent to those used in official control monitoring i.e. without considering the internal quantification of measurement uncertainty for individual samples that is available for MPN results (e.g. Pol-Hofstad

and Jacobs-Reitsma 2021). This pragmatic approach addresses the question of comparability of pour plate method for use in official shellfish testing, in generating results that are consistent with those derived from MPN. As in the present study, Pol-Hofstad and Jacobs-Reitsma (2021) statistical analyses investigated the relationship between results generated by the two methods, using sub-samples of homogenate from each shellfish sample. The comparability of the two methods (trueness) was assessed by the goodness of fit of the regression line for results from the two methods plotted against each other, and determination of the proportion of values for difference between results that exceeded the calculated 95 % confidence interval for differences between the methods (Bland-Altman difference plot). In the present study, we applied similar approaches to *E. coli* data from samples of oysters and mussels collected over a 12-month period from the Camel shellfish bed, and to some extent results are consistent between the two studies. However, the addition of the regression plot of differences between the methods in the Bland-Altman plot indicates that there were differences.

Statistically significant differences were found between *E. coli* values when applying MPN and pour plate methods to the Camel environmental shellfish samples using paired t-tests. The differences between the methods were also confirmed at the level of comparison of the individual pour plate results with the confidence intervals generated for each paired MPN result. To investigate this further, data from the Pol-Hofstad and Jacobs-Reitsma (2021) RIVM validation study of TBX pour plate (ISO 16649-2) against the MPN reference method were examined in a similar way. This analysis shows that, in their study, the MPN method also generated statistically higher *E. coli* values than the pour plate method, with the difference in mean results between the two methods similar to that observed for the Camel in the present study.

There are relatively few examples of similar direct comparisons of spike-recovery of *E. coli* in shellfish by MPN and plate culture methods, especially for the two ISO methods compared here. Data for mussel samples spiked with *E. coli*, from Pol-Hofstad and Jacobs-Reitsma (2021, Annex 4), show similar results to those observed in the present study; measured MPN values were generally substantially higher than the inoculated concentration, while pour plate results were closer to the inoculated values (though also higher in many cases). In other studies, the measured concentrations of *E. coli* may be expected to be slightly lower or equivalent to the introduced spiked concentrations. For example, Garcia *et al* (1995) compared spike recovery in soft shell clams (*Mya arenaria*) over a range of concentrations from 100 - 10,000 cfu/100 g, for which they found that 5-tube MPN yielded slightly but significantly lower results compared to ETPC/mFC rosilic acid agar plating method, and below the intended 100 – 10,000 cfu/100g spike levels.

Overall, the present results and comparison with previous studies indicate that determining comparability or differences between the two methods of measuring *E. coli* may depend on the question being asked, the statistical approach taken and the range of values being considered. It is clear that statistically significant differences may be observed between the MPN and pour plate methods, whether considered at the level of the inherent measurement variability of individual MPN results or at the level of variability in a series of single results as used in the practical application of the official control regulations. These statistical differences are in contrast to the apparent

conformity between the two methods of the assessments of equivalence applied in ISO-standard cross-validation studies.

The present results do not explain the potential underlying causes of the differences between the two methods, but do indicate that regardless of laboratory procedures there are substantial differences in their inherent statistical properties. The inclusion of the 24-hour resuscitation step in MPN could potentially result in greater recovery of viable but non-culturable bacteria (VNBC), and the pour plate method has a shorter resuscitation step of 4 hours at 37 °C. Some studies have also shown that MPN can record some bacteria other than *E. coli.* Research into species accuracy of the MPN method has also shown that other bacteria, mainly within the Enterobacteriaceae family, will give false positives with 10% of presumptive *E. coli* not confirmed as *E. coli,* leading to overestimation of counts (Grevskott *et al,* 2016). These authors suggest if specific enumeration is required then alternative chromogenic medium and/or biochemical verification or DNA based methods could be used to verify results.

5.5.3 Implications for shellfish area classification

Present results demonstrate the potential for differences in the *E. coli* monitoring results depending on the test methodology utilised, through differences in method variability and tendency for lower but more consistent values with the pour plate method. Whilst overestimation of *E. coli* can lead to the closure and incorrect classification of shellfish beds, underestimation could lead to increased public health risk. Gronewold and Wolpert (2008) considered how differences in variability in MPN and cfu-based methods of measuring faecal coliforms in environmental water samples might result in differences in management decisions (e.g. closure of shellfish waters, which in the USA is based on bacterial concentrations in water rather than shellfish flesh). Thus, intra-sample variability or measurement uncertainty may yield a range of results from a single sample, or a proportion of results in an overall data set, that crosses a management decision threshold.

The difference in measured values between the pour plate and MPN results can also be considerable. The OCR are applied on the basis of natural values (i.e. not log transformed), and hence the observed differences between the methods can lead to results that can be 2 -3 times higher when using MPN instead of pour plate. The regression relationships between MPN and pour plate results and the probability table for potential different classification threshold outcomes for a single sample may be applicable to evaluation of historical MPN results for a range of shellfish production areas elsewhere. Over the 12-month study period, no difference in potential classification of the Camel shellfish beds could be determined, based on *E. coli* results from the two methods. However, the relationships between the results obtained using the two methods, as reported above, could help provide an initial assessment of the likelihood that application of the pour plate could influence classification outcomes at other locations. Gronewold and Wolpert (2008) also highlighted the potential difficulties posed in transitioning from one method to another, especially where results may be used for monitoring long term trends (Noble *et al,* 2003). This could also have implications for any change of methods in monitoring of *E. coli* UK shellfisheries, where 3-year datasets are used as the basis for classification of production areas, while longer-term data may be used as evidence in monitoring trends in environmental quality.

5.5.4 Implications for predictive modelling of *E. coli* in shellfish from environmental factors

The purpose of the present study was not to investigate the potential for a change to alternative methods for use in OCR *E coli* monitoring, but more to explore the potential for using different measurements of *E. coli* in development for predictive models based on environmental factors. Also, initial exploratory environmental modelling work (see Section 7 of this report), identified very high MPN results (> 180,000 *E.coli*/100 g) in the official records that could not be readily attributed to any of environmental factors being considered. These unexplained high results contributed disproportionately to the uncertainty in initial model fits, and it was suspected that the overall variability of the MPN data might be contributing to the weak correlations observed between *E. coli* results from adjacent shellfish beds and the weak relationships between observed *E. coli* levels and environmental predictors. This was prompted by review of the MPN method, for which numerous reports discuss inherent variability of the method. The pour plate results recorded in the present study indicate potential for improvement of modelling of *E. coli* from selected environmental predictors and provide a data set for which this is explored in subsequent sections of this report.

5.5.4.1 Acknowledgements

Statistical analysis and plotting were carried out using the R statistical package and the ggplot2 package (Wickham, 2016; R Core Team, 2021)

5.6 References

Beliaeff B, Cochard ML (1995) Applying geostatistics to identification of spatial patterns of fecal contamination in a mussel farming area (Havre de la Vanlee, France). Water Research 29: 1541-1548.

CEFAS (2019) Enumeration of *Escherichia coli* in bivalve molluscan shellfish by the most probable number (MPN) technique. Generic protocol (based on ISO 16649-3). Issue date 17.09.2019.

CEFAS (2021) Determining measurement uncertainty for the enumeration of *E. coli* in bivalve molluscs by ISO 16649-3. https://www.cefas.co.uk/media/kr5gdmpi/determining-uncertainty-of-measurement-for-the-enumeration-of-e-coli-in-bivalve-molluscs-by-iso-16649-3-v2021.pdf

Clements K, Giménez L, Jones DL, Wilson J, Malham SK (2013) Epizoic Barnacles Act as Pathogen Reservoirs on Shellfish Beds. Journal of Shellfish Research 32: 533538.

Clements K, Quilliam R S, Jones DL, Wilson J, Malham SK (2015) Spatial and temporal heterogeneity of bacteria across an intertidal shellfish bed: Implications for regulatory monitoring of faecal indicator organisms**.** Science of the Total Environment 506: 1-9.

Dupont, J., Menard, D., Herve, C., Chevalier, F., Beliaeff, B., Minier, B. (1996). Rapid estimation of Escherichia coli in live marine bivalve shellfish using automated conductance measurement. Journal of Applied Bacteriology, 80(1), 81-90.

Dupont J, Dumont R, Menanteau C, Pommepuy M (2004) Calibration of the impedance method for rapid quantitative estimation of *Escherichia coli* in live marine bivalve molluscs Journal of Applied Microbiology 96: 894-902.

EURL (2014) EURL protocol - Enumeration of Escherichia coli in bivalve molluscan shellfish by the colony-count technique. Issue 3, 15.05.14.

European Commission (2004) Regulation of the European Parliament and of the Council of 29 April 2004 laying down specific hygiene rules for on the hygiene of foodstuffs, 853/2004/EC. Official Journal of the European Union L 139 30.04.2004.

European Commission (2017) Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products, amending Regulations (EC) No 999/2001, (EC) No 396/2005, (EC) No 1069/2009, (EC) No 1107/2009, (EU) No 1151/2012, (EU) No 652/2014, (EU) 2016/429 and (EU) 2016/2031 of the European Parliament and of the Council, Council Regulations (EC) No 1/2005 and (EC) No 1099/2009 and Council Directives 98/58/EC, 1999/74/EC, 2007/43/EC, 2008/119/EC and 2008/120/EC, and repealing Regulations (EC) No 854/2004 and (EC) No 882/2004 of the European Parliament and of the Council, Council Directives 89/608/EEC, 89/662/EEC, 90/425/EEC, 91/496/EEC, 96/23/EC, 96/93/EC and 97/78/EC and Council Decision 92/438/EEC (Official Controls Regulation). Official Journal of the European Unio L 95, 7.4.2017, pp. 1-142.

European Commission (2019). Commission implementing regulation (EU) 2019/627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and amending Commission Regulation (EC) No 2074/2005 as regards official controls. Official Journal of the European Union L 131, pp.51-100.

Food Standards Agency (2017) Protocol for sampling and transport of shellfish for the purpose of Official Control Monitoring of classified shellfish production areas under Regulation (EC) No. 854/2004

Garcia GR, Haymond RE, Sprague DM, Singleton ER, Peeler JT, Lancette GA and Sofos JN (1995) Comparison of a rapid plate count and MPN methods for enumeration of fecal coliforms and *Escherichia coli* in soft-shell clams. Journal of food protection 58: 1197-1200.

Grevskott DH, Svanevik CS, Wester AL, Lunestad, BT (2016) The species accuracy of the Most Probable Number (MPN) European Union reference method for enumeration of *Escherichia coli* in marine bivalves. Journal of Microbiological Methods 131: 73-77.

Gronewold AD, Wolpert RL (2008) Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. Water Research 42: 3327 – 3334.

IFREMER, 2014. Enumeration of *Escherichia coli* in live bivalve molluscan shellfish by the direct impedance technique using the BacTrac 4300 series analyser.

https://cefaswebsitedevtest.co.uk/media/taulrzth/c-users-ab19-documents-coe-seafood-safety-generic-protocol-using-impedence.pdf Accessed 19.05.2022.

ISO (2001) 16649-2. Microbiology of food and animal feeding stuffs—Horizontal method for the enumeration of beta-glucuronidase-positive *Escherichia coli*—Part 2: Colony-count technique at 44 degrees C using 5-bromo-4-chloro-3-indolyl beta-D-glucuronide. 8 pp. https://www.iso.org/standard/29824.html. Accessed 19.05.2022.

ISO (2016a) ISO 16649-3:2015 Microbiology of the food chain. Horizontal method for the enumeration of beta-glucuronidase-positive *Escherichia coli* detection and most probable number technique using 5-bromo-4-chloro-3-indolyl-β-D-glucuronide. Corrected 2016-12, confirmed 2020. 10 pp. https://www.iso.org/standard/56824.html. Accessed 19.05.2022.

ISO (2016b) ISO 16140-2:2016 Microbiology of the food chain — Method validation — Part 2: Protocol for the validation of alternative (proprietary) methods against a reference method (Accessed 01/12/2022)

Gronewold AD, Wolpert RL (2008) Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. Water Research 42: 3327 – 3334.

Iwamoto, M, Ayers T, Mahon BE, and Swerdlow DL (2010) Epidemiology of seafood-associated infections in the United States. Clinical microbiology reviews 23: 399-411.

Jacobs-Reistma W F, van Overbeek W, Franz E, and Pol-Hofstad IE (2010) Expert lab report on the MicroVal ISO 16140:2003 validation of the TBX pour plate method (ISO 16649-2) for enumeration of *Escherichia coli* in bivalve molluscs. Rikilt Report 507: 94pp.

Jarvis B, Wilrich C, Wilrich P-T (2010) Reconsideration of the derivation of Most Probable Numbers, their standard deviations, confidence bounds and rarity values. Journal of Applied Microbiology 109: 1660–1667.

Kay D, Kershaw S, Lee R, Wyer MD, Watkins J, Francis C (2008) Results of field investigations into the impact of intermittent sewage discharges on the microbiological quality of wild mussels (*M. edulis*) in a tidal estuary. Water Research 42: 3033-3046.

Lee R, Murray L (2010) Components of microbiological monitoring programmes. World Health Organization (WHO). Safe Management of Shellfish and Harvest Waters. Edited by Rees, G., Pond, K., Kay, D., Bartram, J., Santo Domingo, J. pp 91-109.

Lee RJ, Silk R (2013) Sources of variation of *Escherichia coli* concentrations in bivalve molluscs. Journal of Water Health, England 11: 78 - 83.

McCrady, M.H., 1915. The numerical interpretation of fermentation-tube results. The Journal of Infectious Diseases, pp.183-212.

Mooijman KA, Poelman M, Stegeman H, Warmerdam C, Teunis PFM, de Roda Husman, AM (2007) Validation and comparison of methods for enumeration of faecal coliforms and *Escherichia coli* in bivalve molluscs. RIVM rapport 330310001.

Noble RT, Weisberg SB, Leecaster MK, McGee CD, Ritter KJ, Walker KO, Vainik PM (2003) Comparison of beach bacterial water quality indicator measurement methods. Environmental Monitoring and Assessment 81: 301–312.

Ogden ID, Brown GC, Gallacher S, Garthwaite PH, Gennari M, Gonzalez MP, Jørgensen LB, Lunestad BT, MacRae M, Nunes MC, Petersen AC (1998). An interlaboratory study to find an alternative to the MPN technique for enumerating *Escherichia coli* in shellfish. International Journal of Food Microbiology, 40: 57-64.

Pol-Hofstad IE, Jacobs-Reitsma WF (2021). Validation of the TBX pour plate method (ISO 16649-2) for the enumeration of *Escherichia coli* in live bivalve molluscs: renewal study for alignment with EN ISO 16140-2: 2016. RIVM Report 2021-0127. 62pp.

Pouillot R, Smith M, Van Doren JM, Catford A, Holtzman J, Calci KR, Edwards R, Goblick G, Roberts C, Stobo J, White J (2021) Risk Assessment of Norovirus Illness from Consumption of Raw Oysters in the United States and in Canada. Risk Analysis. DOI 10.111/risa.3755

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Miller J (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY. pp.328-330.

Walker DI, Younger A, Stockley L, Baker-Austin C (2018). *Escherichia coli* testing and enumeration in live bivalve shellfish—Present methods and future directions. Food Microbiology. 73: 29–38.

Walker DI, Fok BCT, Ford CL (2020) A qPCR-MPN method for rapid quantification of *Escherichia coli* in bivalve molluscan shellfish. Journal of Microbiological Methods 178: 106067.

West PA, Coleman MR (1986) A tentative national reference procedure for isolation and enumeration of *Escherichia coli* from bivalve molluscan shellfish by most probable number method. Journal of Applied Microbiology 61: 505-516.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer.

6 Statistical Modelling for Shellfish Assurance Scheme

6.1 Summary

The DASSHH project set out to fit the best statistical model to historical MPN and project MPN. During the course of the project, an investigation of the merits of pour plate *E. coli* counts was included. The best and relatively simple model developed for the Camel is based on environmental data (rainfall radar, river flow, temperature/season) that is readily available. The limited available CSO spill data were included as initial explanatory variables but not found to have explanatory power beyond that provided by the remaining variables The MPN and pour plate *E. coli*/100 g shellfish flesh data that was collected by the DASSHH project gave better results than the historical MPN data. For the selected pour plate-based model, explanatory power of environmental variables and *E. coli* in shellfish was in some cases improved over previous studies when considered at the level of individual shellfish beds Figure 2.1). These findings suggest that, for the Camel, bed-specific and species-specific models may be more appropriate than a single whole-site model, with some strong predictive relationships demonstrated for individual beds.

The most reliable models correctly assigning predicted *E. coli* levels in shellfish to below Class A classification thresholds (<230 and <700 *E. coli*/100g), with 90% and 88% reliability. This rose to 98% reliability for the C class boundary (<4,600 *E. coli*/100g). These results suggest that there is potential to develop a model-driven management system, but with sufficient accuracy demonstrated only where *E. coli* data supplementary to the Official Control sampling is applied, especially the use of pour plate *E. coli* data. For each of the Class A, Class B and Class Classification thresholds, the model based on pour plate data were substantially better at predicting a pass (i.e. below the threshold, avoiding false negatives) than the MPN model, while the prediction accuracies for fails, above the threshold, were similar for both models. However, these results are based on relatively short data sets and further modelling over longer time series is required to confirm these findings and potentially improve the models.

The Camel study was unable to develop satisfactory predictive models based solely on historical MPN *E. coli* results from the Official Control sampling. These data were found to be highly variable and loosely related to explanatory variables considered. Hence the explanatory power of the environmental data were often limited, and strongly influenced by small numbers of extreme values. Small numbers of extremely high MPN values (>180,000 *E.coli*/100g)are difficult to characterise statistically and some are not associated with preceding rainfall or any other explanatory variable. One potential reason for the differences in model performance between Official Control MPN data and those collected for the DASSHH project is that the latter were collected more frequently (two-weekly vs monthly) and more systematically on the same day every two weeks whatever the weather, whereas statutory sampling usually occurs once a month and sampling date may vary depending on the weather, potentially introducing bias. The pooling of shellfish sampled from three points across each bed may also have reduced variability, compared to Official Control samples that are collected at a single monitoring point. The improved performance of models based on pour plate *E. coli* data is unsurprising given the lower inherent variability in this method.

The role of CSO spills in contributing to *E. coli* levels in shellfish was not clearly demonstrated in development of predictive models for the Camel estuary. However, this does not mean that human sewage sources are not significant contributors to *E. coli* levels. The data for CSO operation that were available for this modelling exercise were limited in two ways. First, the "on-off" nature of the data meant that only timing and duration of discharges could be included in statistical models, without any measure of volume or concentration. Secondly, for some of the wastewater source locations, the operation time series data were apparently incomplete. As CSO operation is largely influenced by weather conditions it can also be difficult to disentangle from rainfall as a driver of other catchment sources (eg agricultural), and rainfall is itself a key predictor of *E. coli* counts in shellfish.

Predictive relationships when using other response variables such as enteric viruses (noroviruses and Adenoviruses) were weak. The poor model fit with viruses could be due to the sporadic nature of their occurrence and their longer persistence (infective or not infective) in shellfish.

6.2 Introduction

Microbial contamination of shellfish is assessed from the faecal indicator bacteria *E. coli*/100g count in shellfish flesh. Key sources of *E. coli* and other faecal indicator organisms (FIOs) within catchments impacting upon shellfisheries include sewage discharges and agricultural activities, with further potential inputs from boating activity in coastal environments and tributary watercourses, and wild bird and mammal populations (Malham *et al,* 2014). The transport of faecal pollutants from source to shellfish beds is commonly triggered by rainfall events, with speed of transport and magnitude of impact on shellfish beds further influenced by a combination of catchment characteristics (hydrography, topography, geology, land use types and distributions, and how fast a catchment responds to rain events i.e. 'flashiness') (Malham *et al,* 2017). Other factors with the potential to influence *E. coli* concentrations in shellfish via their role in mediating persistence of *E. coli* once in water include solar radiation, temperature, salinity, pH, sediments and flocs (Campos *et al,* 2013; Malham *et al,* 2017).

In theory, statistical models based on these potential environmental influences can be used to provide an assessment of risk. Such models are now used increasingly to predict bathing water quality (Zimmer-Faust *et al,* 2018), but the challenges associated with predicting shellfish contamination are more complex. This is due to the additional pathways and factors governing accumulation and depuration within shellfish *in situ* (reviewed by Campos *et al,* 2013). As an example, work in in the Dart estuary indicated rainfall and river flow as the main drivers of microbial quality of shellfish (Campos *et al,* 2011). Such approaches have been used to develop a statistical model for shellfish mainly Class A waters in two bays in Cornwall utilising a General Linear Model through incorporation of historical *E. coli* data, rainfall, river flow and, for one bay, solar radiation (Schmidt *et al,* 2018).

A shellfish Assurance Scheme underpinned by robust modelling approaches would offer an alternative means of shellfisheries regulation based on the use of environmental parameters to predict the timing and location of elevated levels of faecal pollution. Such predictions could inform

decision thresholds for preventative closure of shellfish beds showing a high probability of contamination by faecal pollution. Beds would re-open and harvesting resume once the elevated levels had returned to 'background' levels. This form of adaptive management of the shellfishery could be used to reduce the duration of periods of closure.

The aim of the DASSHH project was to develop a statistical model for predicting *E. coli* concentrations in shellfish which is sufficiently robust to form the basis of a decision support system. The model would be based on *E. coli*/100g shellfish flesh data and environmental variables such as weather and river flow and would aim to calculate the probability of shellfish contamination as a basis for risk-based management of the shellfish beds.

6.3 Statistical analysis

Statistical modelling was undertaken for *E. coli* counts in shellfish in the Camel estuary, Cornwall, and a range of potentially predictive environmental variables. The analysis was in three main stages:

- Developing a relationship between historical statutory RMP shellfish monitoring data (1993-2019) based on the MPN method (CEFAS data) and rainfall and river flow. These data were collected monthly and the available series length differs between sites.
- Extending the analysis to include statutory shellfish monitoring MPN data and supplementary MPN data collected for the project via fortnightly oyster and mussel sampling in the Camel from May 2019 – March 2020, as well as additional environmental data over the same period.
- Revisiting the analysis using high frequency (2 weekly) supplementary shellfish *E. coli* data, collected from the Camel during the project using both MPN and pour plate methods, with updated environmental data.

6.4 Stage 1: Statistical modelling of historical statutory RMP monitoring data (1993-2019)

6.4.1 Shellfish sampling locations

Routine *E. coli* monitoring data for shellfish beds in the Camel estuary for oysters and mussels respectively were collated from CEFAS data for the Camel shellfish harvesting areas (early 1990s to December 2019). The site names and grid references are indicated in Tables 6.4.1 and 6.4.2 and their locations in the estuary mapped in Figures 6.4.1 and 6.4.2.

Table 6.4.1. Pacific oyster bed and their locations in the Camel Estuary.

Location (RMP_name)	Abbr.	Easting	Northing
Porthilly Rock (C. g) (B035L) *	PRO	193420	75330
Porthilly Rock B (Cg) (B35AC)		193400	75300
Longlands (C. g) (B035I) *	LLO	193540	74830
Gentle Jane_P. Farm (C. g) (B035A) *	GJO	193920	74680
Gentle Jane B (Cg) (B35AD)		194040	74630
Ball Hill Oyster (OYG) (B035Q) *	BHO	193420	74290
Pinkson Creek (C gigas) (B035R)		194610	73580

Table 6.4.2. Mussel beds and their locations in the Camel estuary.

Location (RMP_name)	Abbr.	Easting	Northing
Pinkson Creek (M) (B035M) *		194590	73600
Pinkson Creek B (M) (B35AF)		194520	73690
Gentle Jane_P. Cove (M) (B035B) *	GJM	193900	74680
Ball Hill West (B035U) *	BHM	193420	74280
Porthilly Cove (B035X) *	PRM	193420	75330
Porthilly Rock B (M) (B35AE)		193400	75300

* used for subsequent statistical analysis.

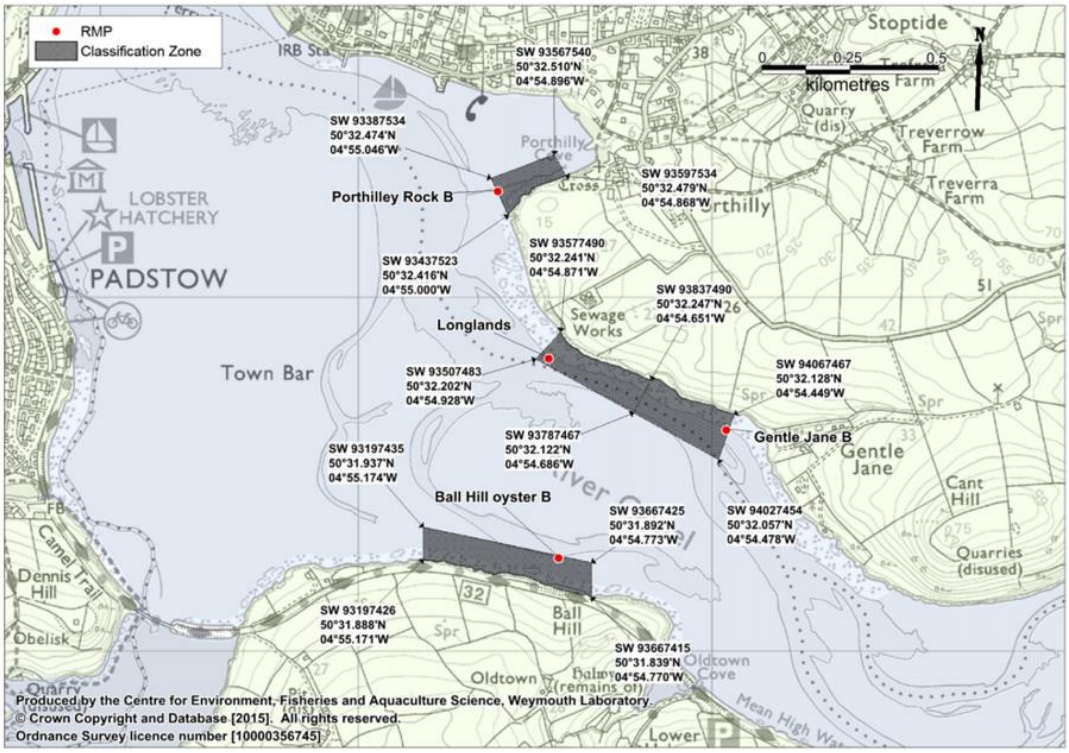


Figure 6.4.1. Zoning and monitoring arrangement for Pacific oysters (*Crassostrea gigas*) in the Camel Estuary. Red dots are the RMPS for the shellfish areas indicated by grey hatching. Taken from the 2015 Camel Sanitary Survey (CEFAS, 2015)

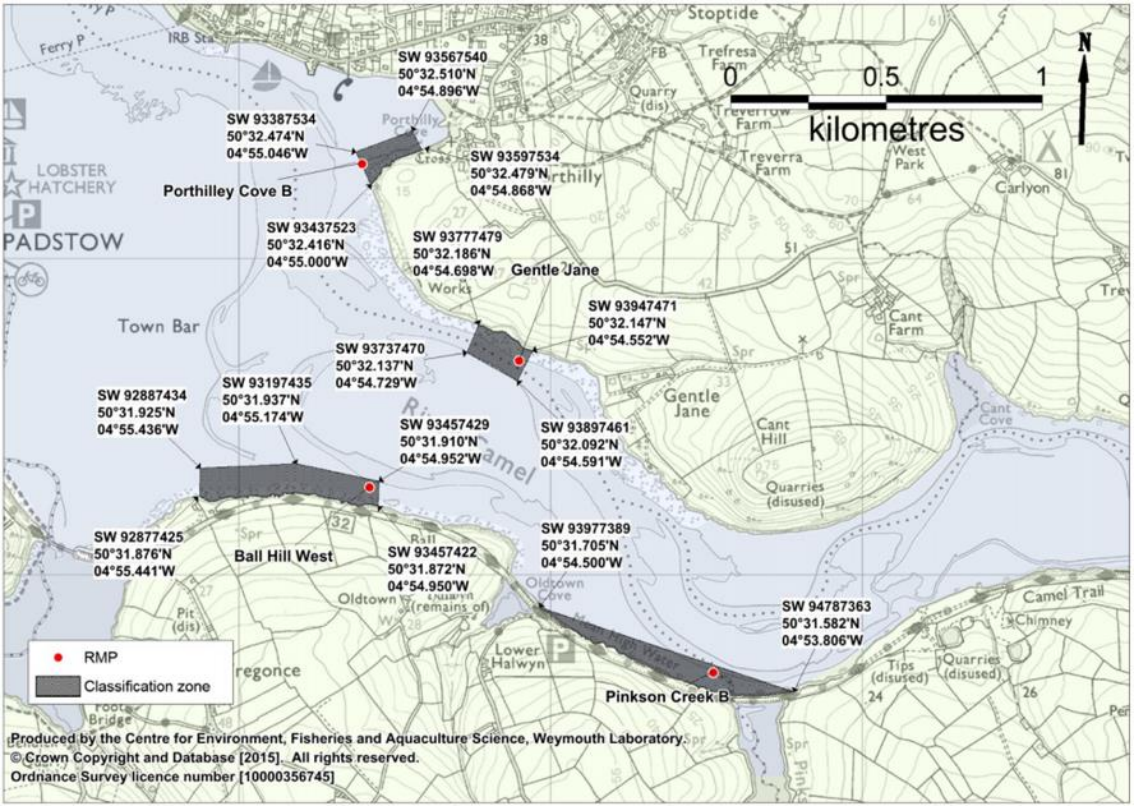


Figure 6.4.2. Zoning and monitoring arrangement for Mussels (*Mytilus* spp.) in the Camel Estuary. Red dots are the RMPS for the shellfish areas indicated by grey hatching. Taken from the 2015 Camel Sanitary Survey (CEFAS, 2015)

6.4.2 MPN count response variables

The Official Control monitoring data is measured using the standard (MPN) method for *E. coli* enumeration in shellfish. The method gives a statistically derived estimate of viable *E. coli* concentration based on the presence/absence of *E. coli* in replicate ten-fold serial dilutions, reporting the concentrations as the most probable number (MPN) per 100g of sampled material. The sampling record at the sites listed in Tables 6.4.1 and 6.4.2 starts in 1991, with varying frequency, but generally monthly or bimonthly. Figures 6.4.3 and 6.4.4 show time series of counts over the period of record.

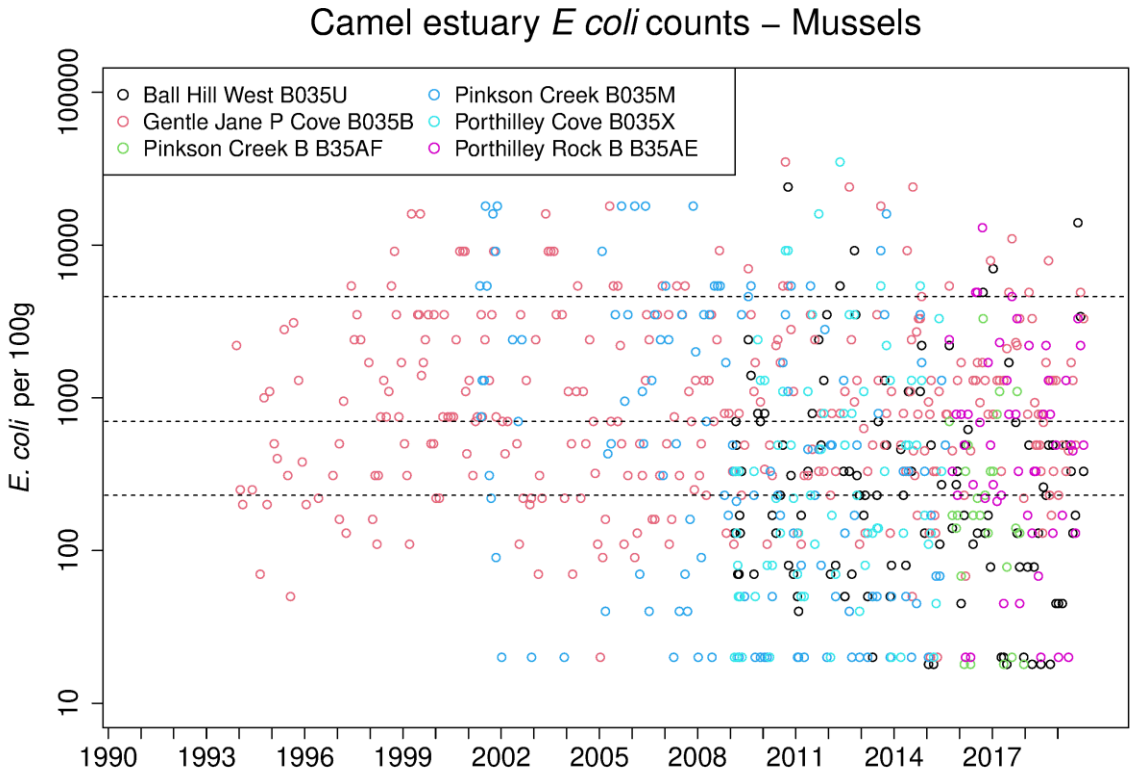


Figure 6.4.3. Time series plots (1993-2019) of Official Control data for *E. coli*/100g in mussels at the shellfish growing area recommended monitoring points in the Camel Estuary. Dashed horizontal lines indicate 230, 700, 4600 MPN threshold. Note the log scale on the y-axis.

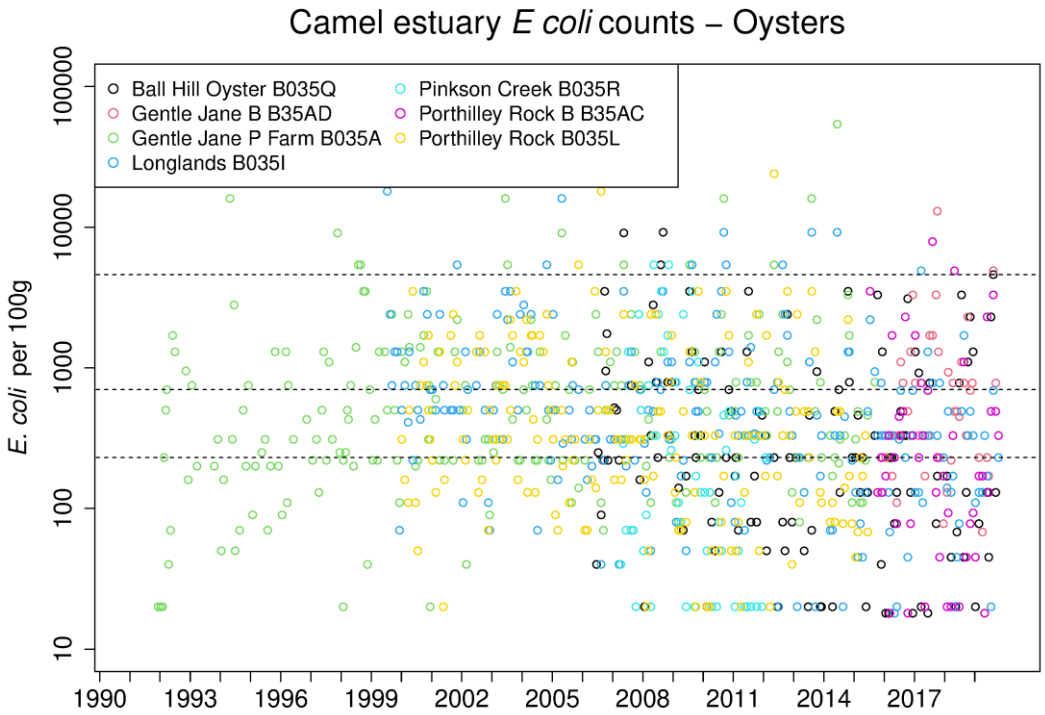


Figure 6.4.4. Time series plots of Official Control data for *E. coli*/100g in oysters at the shellfish growing area recommended monitoring points in the Camel Estuary. Dashed horizontal lines indicate 230, 700, 4600 MPN thresholds. Note the log scale on the y-axis.

6.4.3 Explanatory variables

Sources of the *E. coli* which enter shellfish are not directly measured, but evidence suggests they are related to environmental factors dominated by weather conditions, notably antecedent rainfall (Tryland *et. Al.*, 2014). Some measures of these environmental variables are available, and relationships between them and *E. coli* counts in shellfish may be investigated.

Environmental data as possible explanatory variables were collated from several sources (Table 6.4.3). The usefulness of all of these was not fully investigated during Stage 1, but they were identified as potentially useful, and sourced. Detailed discussion of these data sources is reserved for Stage 2, where they are investigated as potential explanatory variables in modelling. Variables used in Stage 1 were Met Station rainfall data and river flow.

The Environment Agency (EA) provided 15 minute flow measurements for the river Camel at the Denby (NGR E201748 N068159) and Bodmin (Dunmere) (NGR E204456 N067451) river gauging sites. The Denby site ran until January 2019 when it was replaced by Bodmin (Dunmere), with both sites operational for 2017-2019. A calibration equation (with units m^3s^{-1}) was established to simulate flows at Denby from January 2019 onwards, to allow a degree of consistency in the flow record:

$$\text{Log}_{10}\,(\text{Denby}) = 0.24 + 0.9\log_{10}\,(\text{Bodmin})$$

Hourly precipitation data at Cardinham (NGR E211197, N70053) meteorological stations were accessed from the CEDA database for January 2009 to December 2018.

Table 6.4.3 Summary of available environmental data collected in relation to the Camel catchment

Variable	From	To	Source	Comments
Combined sewer overflows	01/2004	12/2019	CEFAS, SWW	Location of CSOs (CEFAS Camel sanitary survey, SWW)
River/estuary water quality data	01/1991	12/2015	EA	Sparse data and short record. Includes Harmonised Monitoring Scheme at Polbrock Bridge
River flow data	01/1991	03/2020	EA	15 min flow measurements Denby (1964 – January 2019) and Camelford (2006 – December 2018)
Met Station data	01/1991	03/2020	BADC	Hourly rainfall at Cardinham and St Mawgan data from CEDA database Jan 2009-Dec 2018
Radar rainfall data	01/2014	03/2020	BADC/ NIMROD	5Km grid and 5 min time interval
Tidal state	01/1991	03/2020	www.ntslf.org	Tidal gauge and storm surge data Newlyn and Ilfracombe
Land cover data for Camel			CEH	Land cover map (LCM2007)
Solar radiation, wind speed and direction	01/1991	03/2020	BADC	Cardinham weather station data.

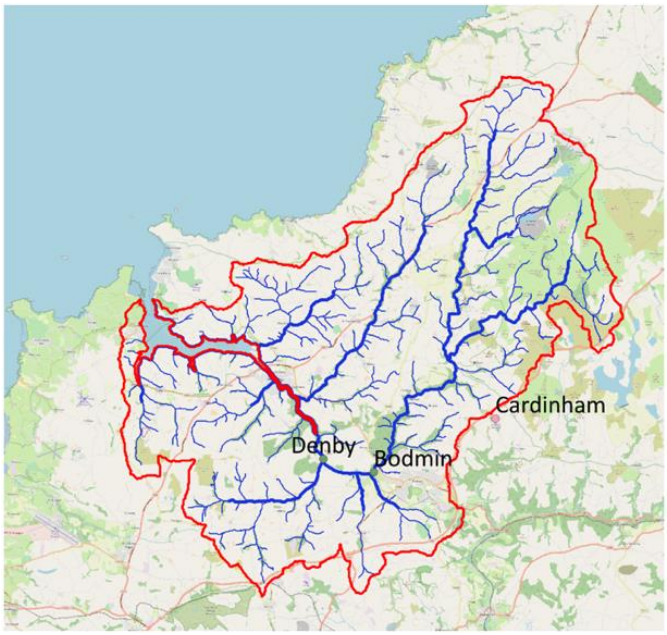


Figure 6.4.5. Location of Met station and river flow gauging sites

6.4.4 Statistical analysis and model prediction

In Stage 1 shellfish *E. coli* counts were compared between sites and a regression relationship fitted between counts and season, river flow in the Camel, and precipitation data from Cardinham, including lagged values of these hydrological variables.

For a between site comparison, those dates on which an MPN count was available at every site were selected. This provided a set of eight values on each of 70 different dates. Counts of *E. coli* have a skewed distribution, with a few samplings having particularly high values. Analysis was undertaken using logged counts, assuming raw data are approximately log-normally distributed.

The collated raw data from 8 shellfish sites for the statutory *E. coli* MPN analysis are presented in Figure 6.4.6. A paired comparison of logged *E. coli* at the 8 sites/100g demonstrated a weak relationship between logged counts at the eight sites (Figure 6.4.7)

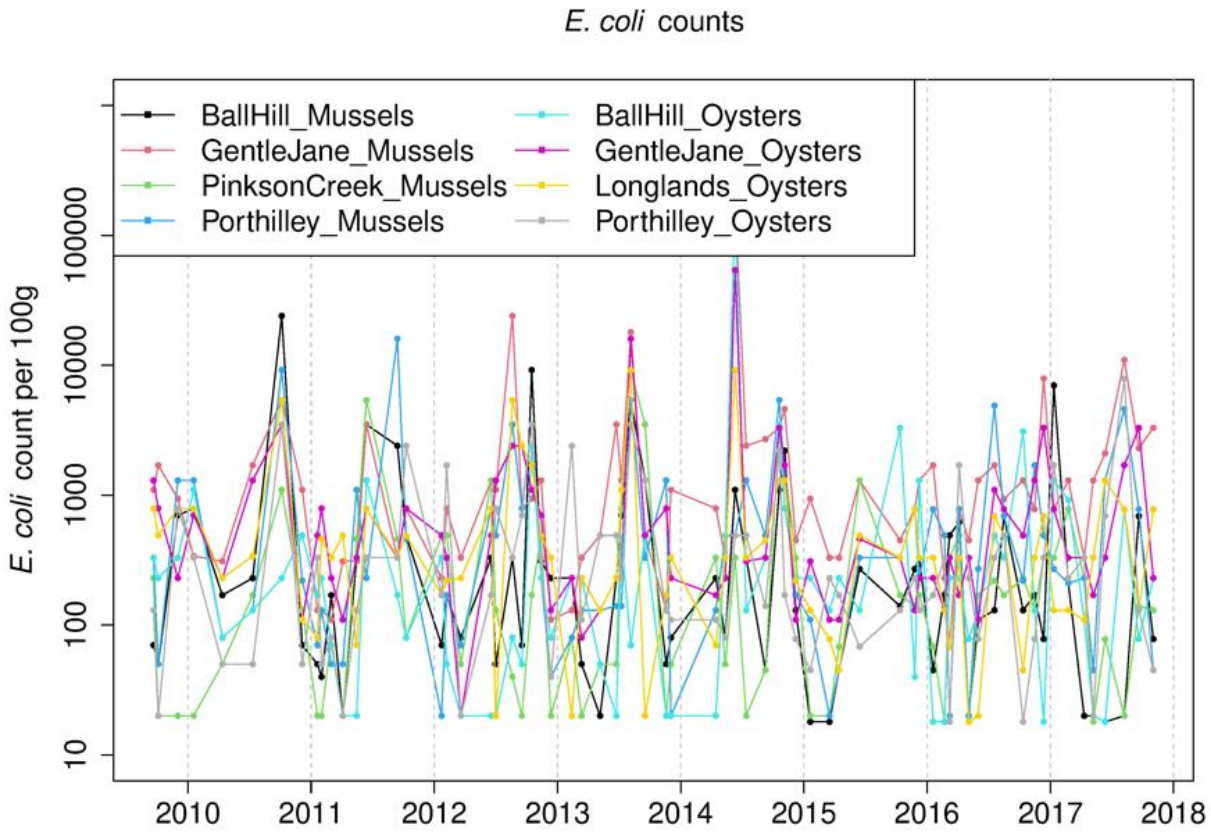


Figure 6.4.6. Raw *E. coli*/100g Official Control data from mussel and oyster recommended monitoring points in the Camel Estuary that have been used in statistical analysis. Each point represents a single sampling occasion on which every site was sampled

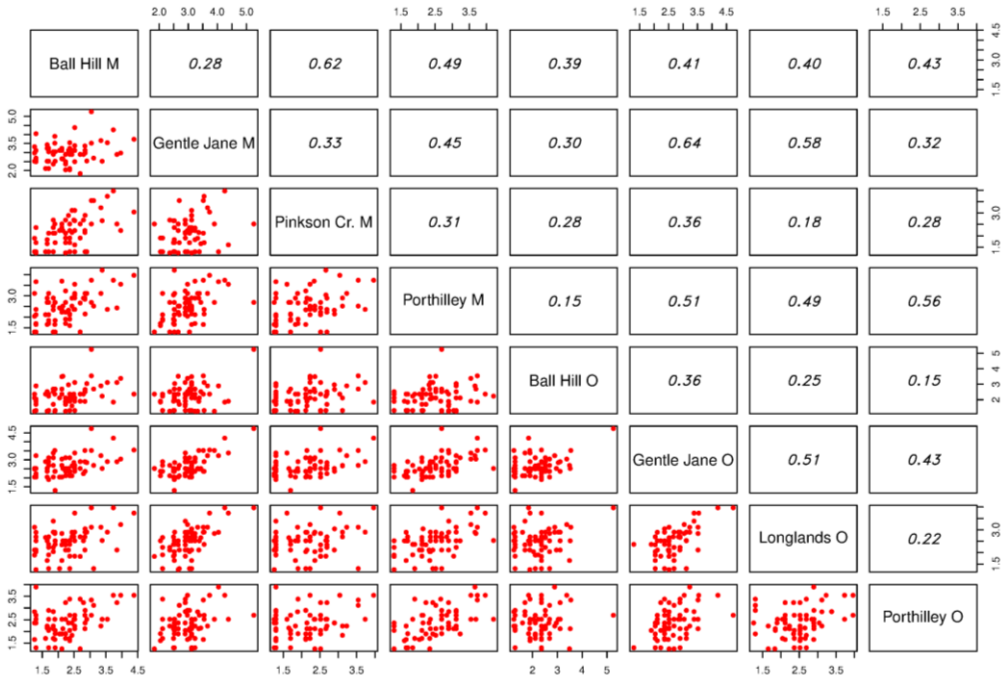


Figure 6.4.7. Paired comparison of logged Official Control *E. coli* counts/100g between selected shellfish Representative Monitoring Points in the Camel Estuary. The upper right triangle shows correlations.

For between-site comparison, a linear model was fitted to the logged *E. coli* counts with date and site (separately for oysters and mussels) as the sole explanatory variables. This provided a measure of the common response of all sites to real but unspecified environmental drivers. While there is good reason to expect there to be differences in the response of individual beds, the fitting of a common response gives an indication of how much common variability there is between beds, as a basis for comparison with model results for individual beds. In this model the logged *E. coli* count was assumed to be the sum of a common response on the day in question and a site-specific response. This model has an R^2 of 0.53, so that just over half of the total variability is explained. The residual standard error is 0.52, so the simulations have error bars of +/- an order of magnitude. The daily variability common to all sites was found to be approximately three times the variability between sites. The analysis demonstrated that the daily effect varied rapidly which is likely to be a response to short term changes in explanatory variables such as the weather or conditions within the estuary. The common response gives the maximum variability which could be accounted for by a statistical model using explanatory variables which are spatially invariant between sites. The additional unaccounted variability is associated with individual sites and can only be accounted for using a model with explanatory variable parameters that vary spatially and with time.

A common response was fitted to the raw data as a daily time series representing the mean *E. coli* count in shellfish (mussels and oysters) across all seven beds. A stepwise regression of this common response was then carried out using hydrological explanatory variables using the stepAIC routine in R, with backward and forward selection. Seasons (summer, winter), river flow and rainfall on the day of sampling and the summed values over the previous three and seven days were considered as possible explanatory variables. Lagged effects were included since an understanding of shellfish physiology suggests that the shellfish response will be cumulative and acts as a reflection of a limited past history of environmental conditions. There may also be a lag in the transfer of E coli through the catchment to the estuary.

Stepwise regression shows that the selected model accounts for 61% of the daily variability in the common response. The retained variables are season, and flow and rainfall on the day of sampling and the 7-day accumulated values. Of the retained variables, season and 7-day summed rainfall account for 53% of the daily variability. Figure 6.4.8 shows the common response and the fitted values. Applying stepwise regression to the raw data rather than the common response, with site also included as an explanatory factor, explaining 37% of the variability. Site, season and 7-day lagged rainfall alone accounted for 35%. The residual standard error is 0.88, implying a 95% confidence interval for count simulations extending to almost two orders of magnitude.

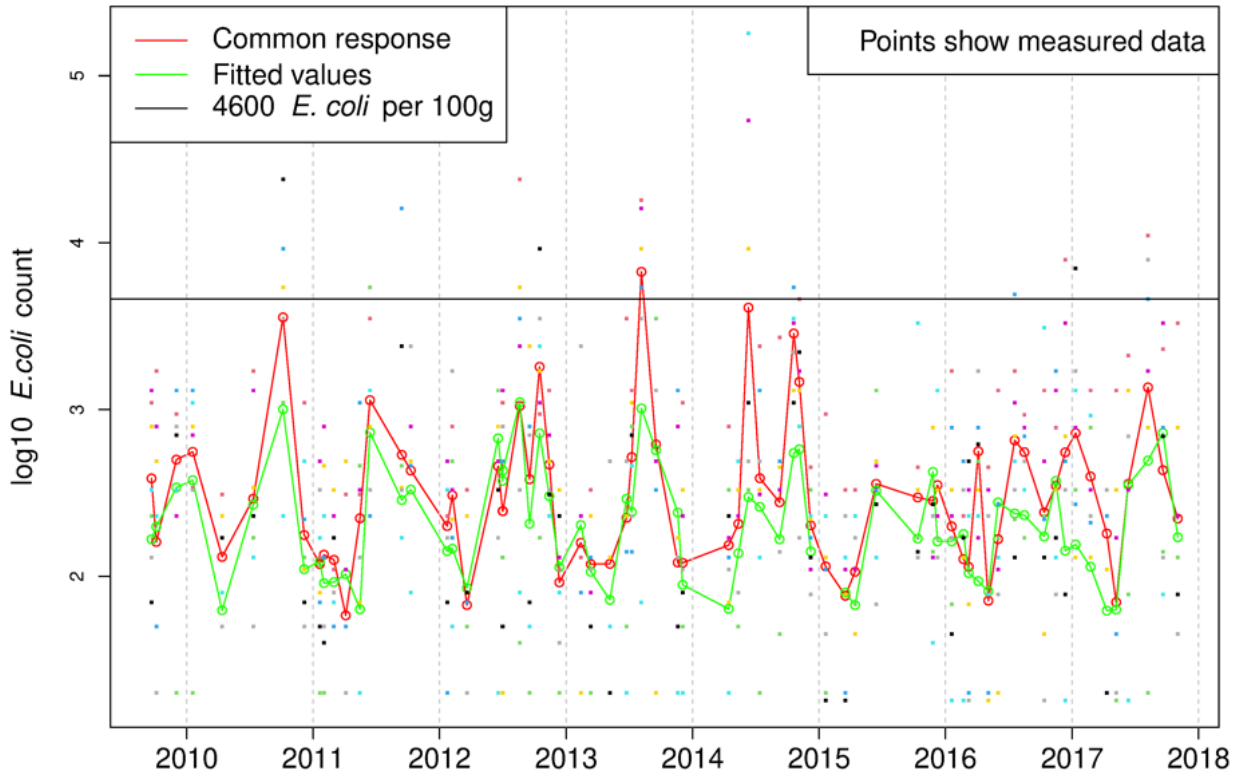


Figure 6.4.8 shows the common response obtained by calculating the daily mean of all sites, and the fitted model results after stepwise regression on the selected explanatory variables.

Further investigation was undertaken into what variability might be systematic using a principal components analysis on the residuals from the common response model. A biplot of these components (Figure 6.4.9) demonstrated that nearby sites in general show similar deviances from the simple linear model and suggested that the response of both oysters and mussels at the same location is generally similar. This analysis provided some measure of the replicability of counts by location, suggesting local effects of at least the order of the distance between the main shellfishery areas. This indicated how far additional model structure might account for the variability in *E. coli* counts between sites.

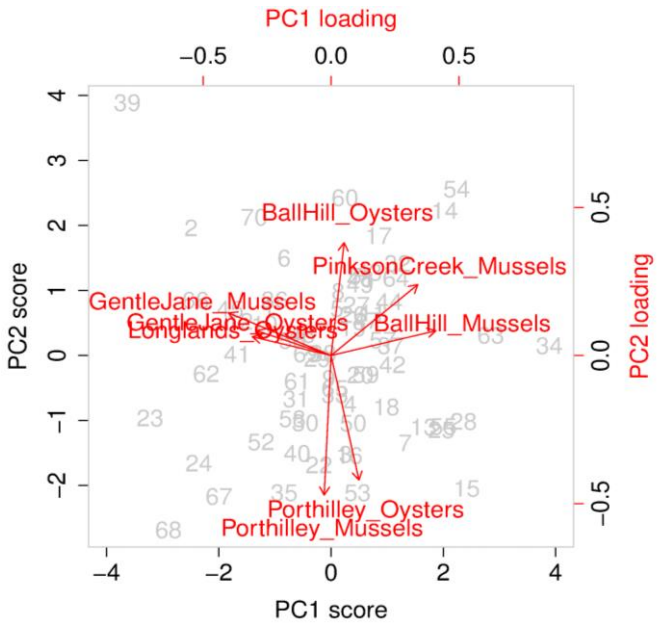


Figure 6.4.9. Principal components analysis of residuals for the common response model for *E. coli* in shellfish in the Camel Estuary. Greyed figures index sample number. Arrows indicate direction of increasing value of the variable.

The high residual variance found using a common model indicated the need for a spatially variable component in order to improve the model. This can be achieved by fitting a model with separate parameters at each site used with a single explanatory variable, or possibly using explanatory variables that are spatially variable rather than single-valued.

Stepwise regression was used to estimate models with separate parameters for each site, including the explanatory variables already identified. The results of the stepwise analysis are shown in Table 6.4.4, where lagged flow and rainfall refer to 3-day and 7-day accumulations respectively. Overall, the data indicated that the strongest association was with lagged rainfall and that the oyster beds showed no association with flow.

Table 6.4.4 Key explanatory variables for Official Control *E. coli* data for shellfish in the Camel, from stepwise regression on individual beds. R^2 refers to the model fit. A single tick implies significance at the 5% level, two ticks at 1%. A cross indicates not significant at the 5% level.

Site	Season	Rain	Lagged rain	Flow	Lagged flow	R^2
Ball Hill Mussels	×	✓✓	✓✓	✓	✓	0.56
Gentle Jane Mussels	✓	×	✓	×	✓	0.42
Pinkson Creek Mussels	×	✓	✓✓	✓	✓	0.45
Porthilly Mussels	✓	×	✓✓	×	×	0.46
Ball Hill Oysters	✓	×	×	×	×	0.09
Gentle Jane Oysters	✓✓	×	✓✓	×	×	0.39
Longlands Oysters	✓	✓	✓✓	×	×	0.39
Porthilly Oysters	×	×	✓✓	×	×	0.26

6.4.5 Model prediction

Once any systematic variability has been accounted for using the model identified by stepwise regression, empirical probability distributions of the unaccounted variability can be added to this component to simulate the distribution of logged counts of which a measured value is a sample. This generated probabilities of exceedance of thresholds for any particular day and shellfish bed, given the measure values of the explanatory variables. The performance of the model can be ascertained by plotting the probabilities of threshold exceedance against measured counts. The relationship in the theoretical situation where the model is prefect is shown in Figure 6.4.10. Every measured exceedance is predicted as an exceedance with probability 1 and every non-exceedance with probability zero.

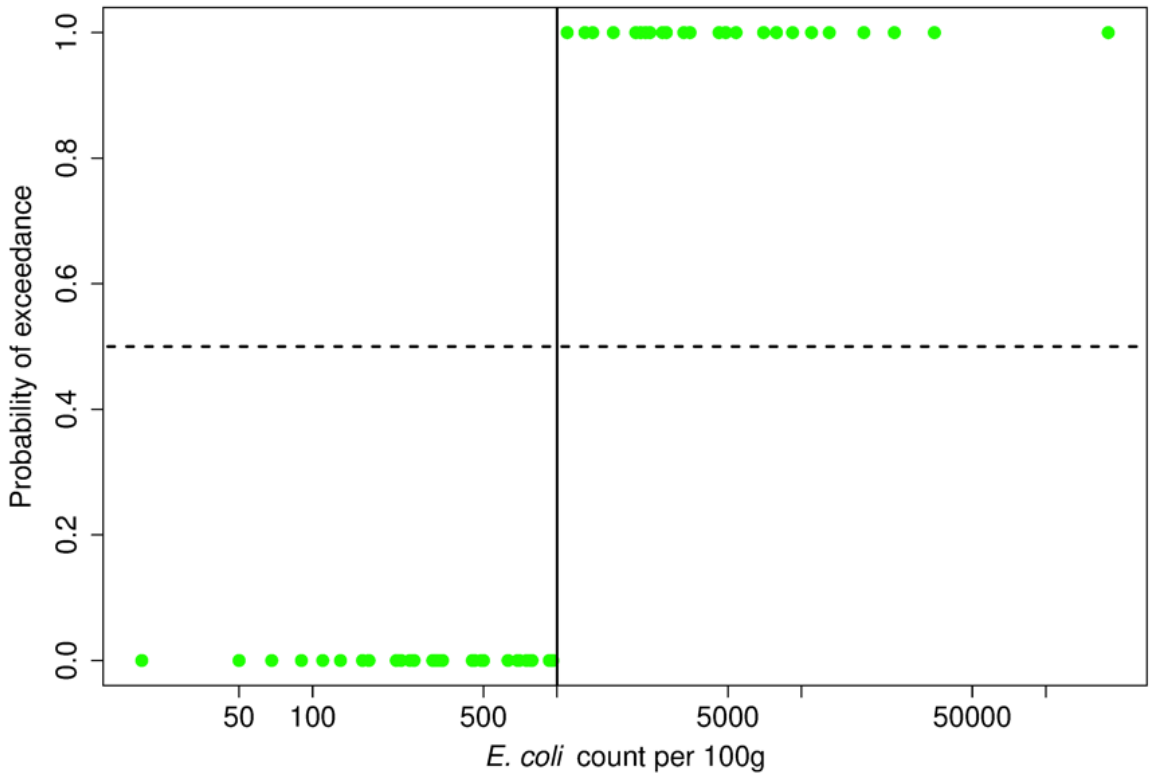


Figure 6.4.10. Example of a perfect prediction of exceedance. All true passes have a probability of 0 of failing and all true fails have a probability of 1 of failing

The stepwise model fitting process is unweighted, and so is likely to simulate average behaviour better than behaviour at extremes. Taking as an example the data and model fit for Gentle Jane mussels, many counts were above a threshold of 1,000 per 100g. The model was moderately successful in simulating exceedances for this threshold (Figure 6.4.11). In general, higher probabilities of exceedance are associated with higher measured counts.

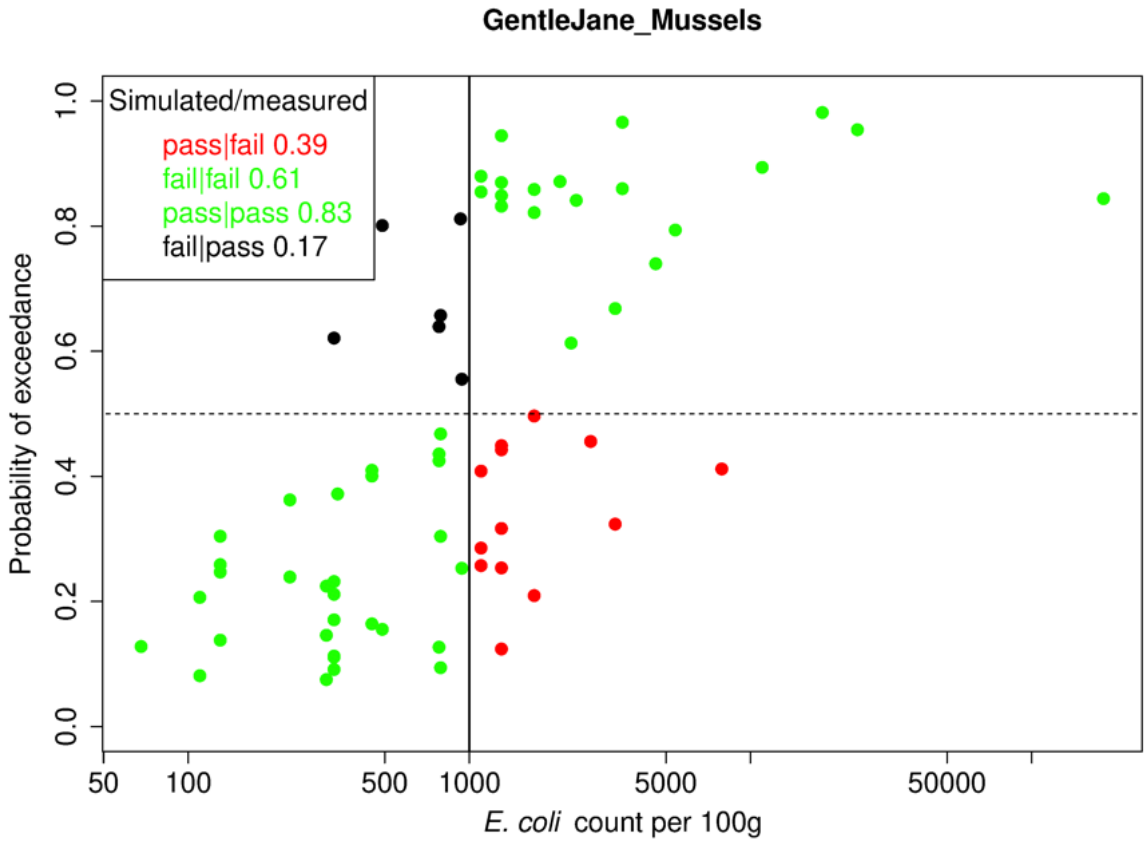


Figure 6.4.11. Model predictive performance for *E. coli* counts in Gentle Jane mussels. Threshold is 1,000 *E. coli*/100g of tissue. Red dots indicate that a pass was simulated by the model, but the measured value was a fail (pass/fail). Green indicates either fail/fail or pass/pass (i.e. a correct simulation). Black indicates a simulated fail, but a measured pass (fail/pass).

In contrast, for Ball Hill oysters (Figure 6.4.12), there was little relationship between probabilities and exceedances for a 1,000 count threshold. The model completely failed to identify the measured exceedances as being likely to be above the threshold. The pattern of behaviour of the probabilities is because they are based solely on season, and this and the poor fit are consistent with low R^2 value found in stepwise model fitting. The model for Ball Hill oysters was particularly poor and required further investigation. For the threshold count of 4,600 *E. coli* per 100g for Gentle Jane mussels (Figure 6.4.13), while there remains some correlation between probabilities and counts, decisions based on a threshold of 0.5 will seldom be reliable. Exceedances of 4,600 are infrequent, and insufficient to give a clear idea of the true performance of the model for this shellfish bed and threshold.

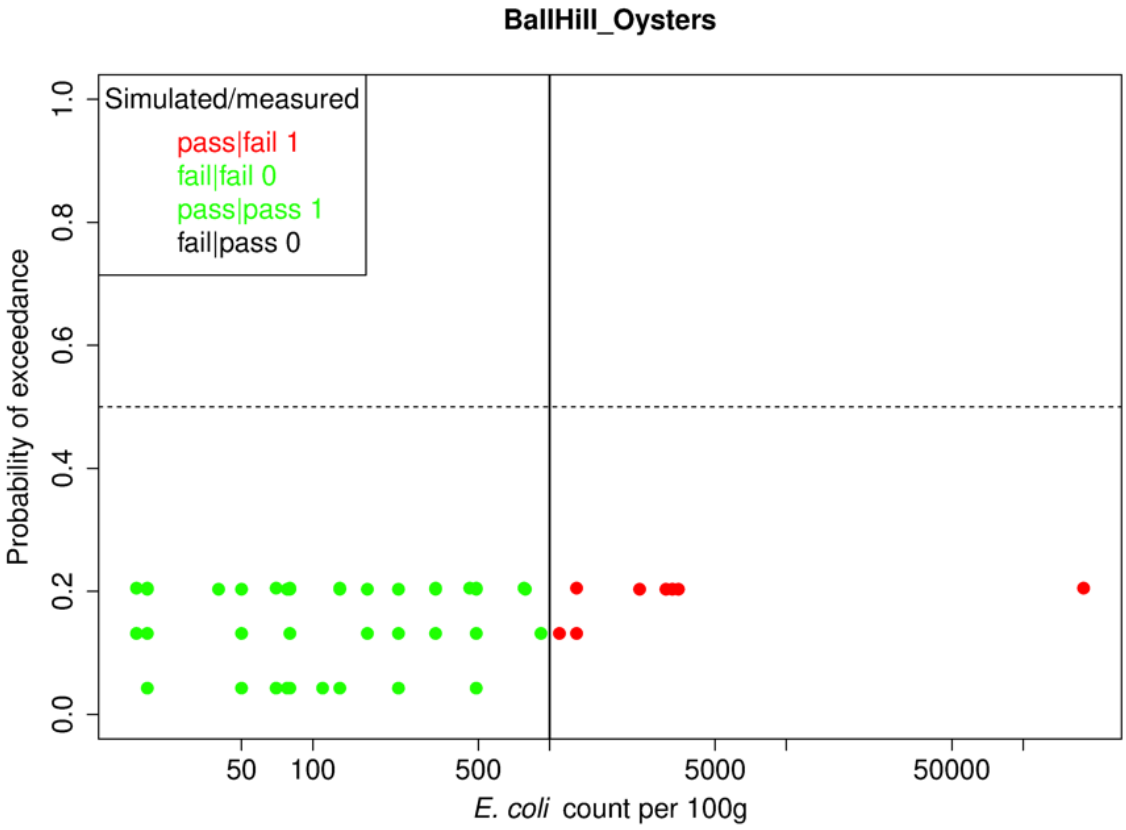


Figure 6.4.12. Model predictive performance for *E. coli* counts in Ball Hill oysters. Threshold is 1,000 *E. coli*/100 g oyster tissue.

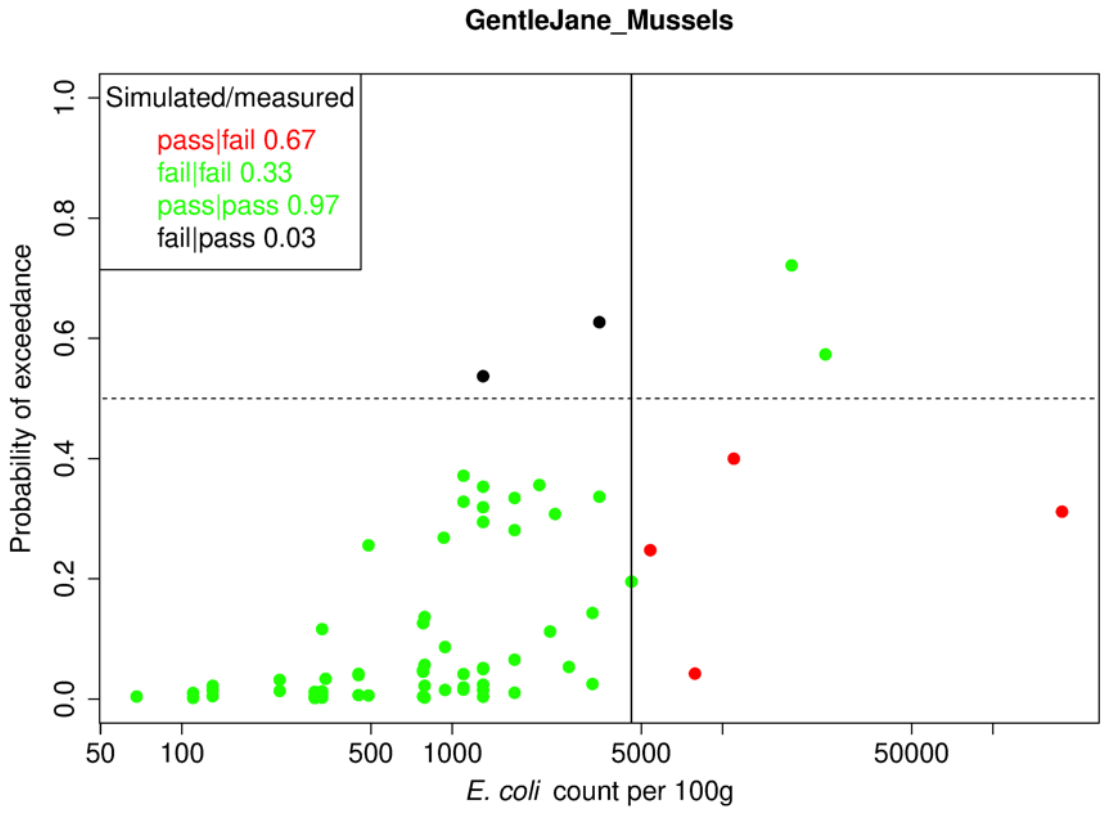


Figure 6.4.13. Model predictive performance for *E. coli* counts in Gentle Jane mussels. Threshold is 4,600 *E. coli/* 100g mussel tissue.

Where there is some relationship between probability of exceedance and measured exceedance, decision-making may be based on a lower probability threshold. This will reduce the probability of failing to predict an exceedance but at the cost of increasing the frequency of false positives. In this way, the probabilities calculated, and the decision threshold, can be built into a decision strategy that takes account of the relative cost of false positives or false negatives.

6.4.6 Summary Statistical modelling of statutory RMP monitoring data (1991 – 2019)

Historical MPN data were found to be highly variable and only loosely related to the hydrological variables selected as explanatory, even after the exclusion of apparently anomalous extreme high values (>180,000 *E.coli*/100g). Stepwise regression and principal component analysis demonstrated both related and unrelated patterns of concentrations between sites. Correlations between the *E. coli* counts of the various shellfish beds were inconsistent, and in some cases very weak e.g. strong relationship between Gentle Jane oysters and mussels, but no relationship between Ball Hill mussels and oysters. Data analysis for MPN relationships with environmental variables indicated the strongest association was with cumulative rainfall for the shellfish sites, whilst the oyster beds showed no association with river flow. Shellfish sites also differed in their response seasonally. The modelling approach does generate probabilities of exceedance of threshold counts given a set of environmental conditions. However, predictions based on models developed in Stage 1 are not sufficiently robust for application in shellfish management, due to the large differences in performance between shellfish beds and species.

6.5 Stage 2. Statistical modelling of statutory RMP monitoring data (1991-2019) and field collected MPN data (2019-2020)

6.5.1 Shellfish statutory MPN RMP and field collected MPN Data

Data for modelling collated for the second stage included those statutory RMP data analysed in Stage 1, together with DASSHH project data collected mainly fortnightly from May 2019 to March 2020. These datasets were combined prior to statistical analysis. Scatter plots over time of shellfish *E. coli/*100 g MPN counts at the locations listed in Tables 6.5.1 and 6.5.2. being considered in Stage 2 (both CEFAS and Bangor University MPN) are shown in Figures 6.5.1 and 6.5.1.

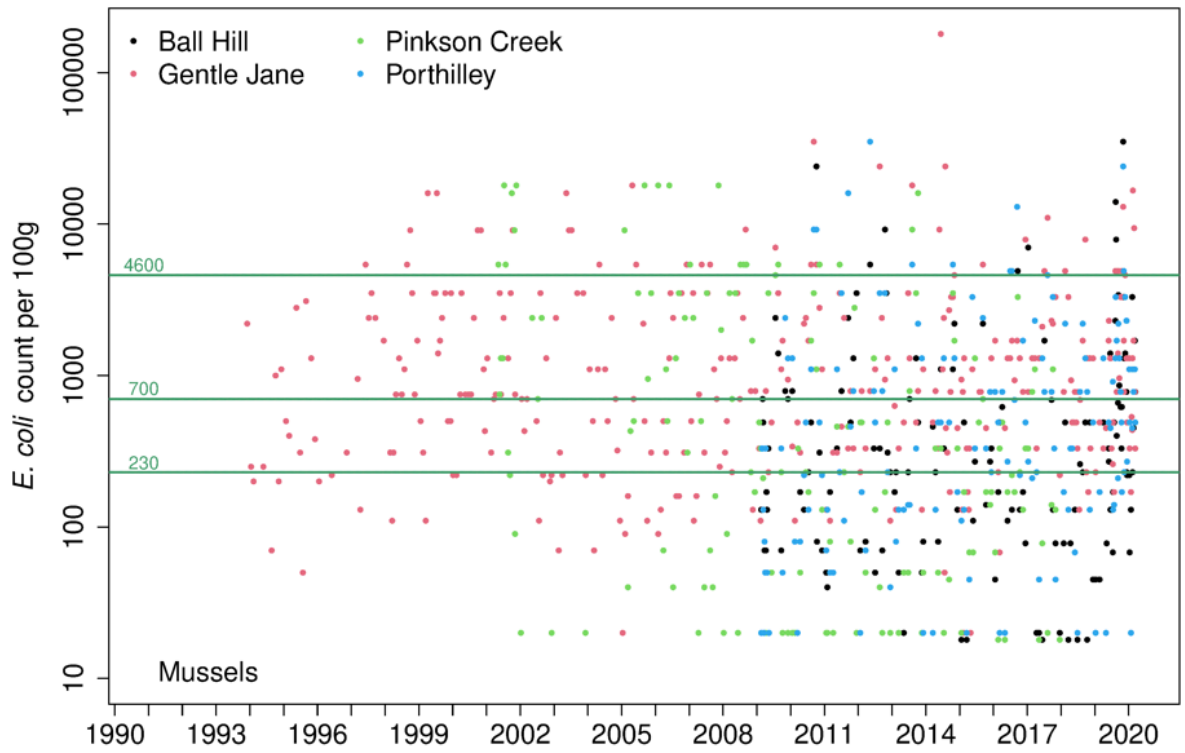


Figure 6.5.1. Time series plots of MPN *E. coli*/100g in mussels in the Camel Estuary, combining Official Control data and supplementary data collected during the DASSHH project. Green lines indicate threshold counts used for shellfish bed classification of 230, 700 and 4,600 *E. coli*/100 g shellfish flesh.

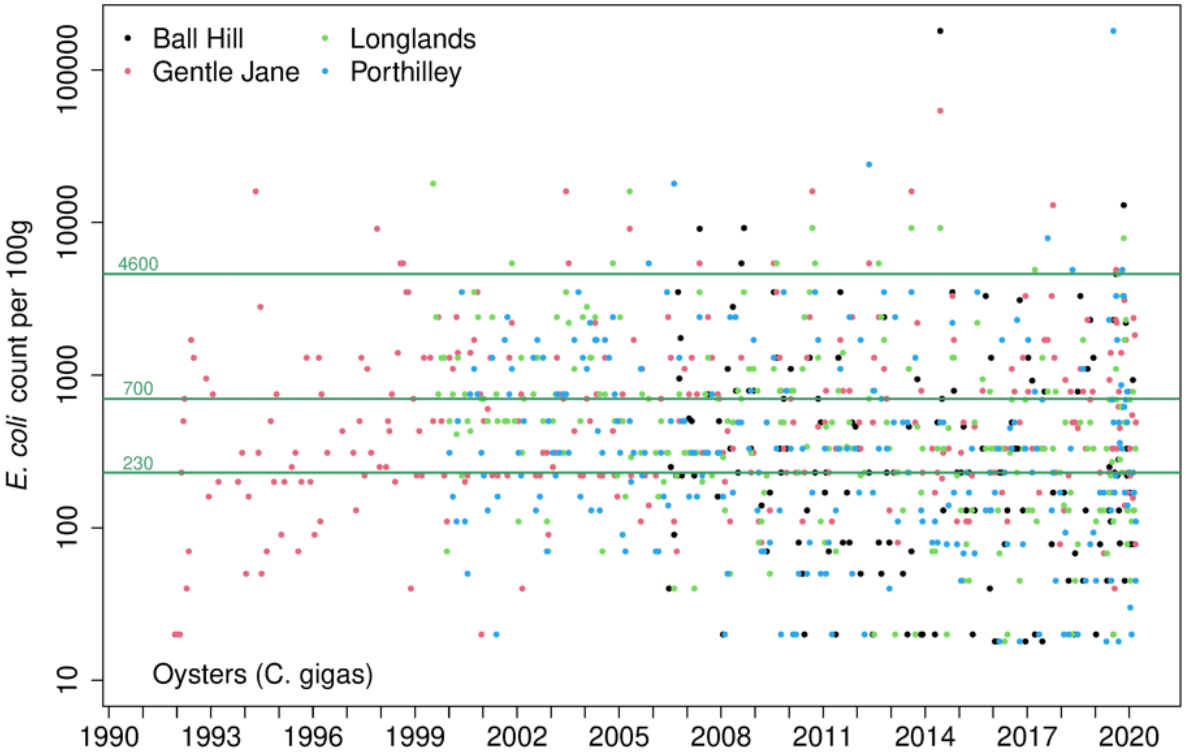


Figure 6.5.2. Time series plots of MPN *E. coli*/100g in oysters in the Camel Estuary, combining Official Control data and supplementary data collected during the DASSHH project. Green lines indicate threshold counts used for shellfish bed classification of 230, 700 and 4,600 *E. coli*/100 g shellfish flesh.

Boxplots of all MPN *E. coli* /100g counts for all data are shown in Figure 6.5.3. The data are presented on a log scale and show high value outliers. Table 6.5.1 shows proportions of counts within classification threshold bounds. A high proportion of values were above the absolute upper limit of 700 *E. coli*/100g required for Class A beds, with a small proportion of samples having counts above the B/C grade classification threshold of 4,600/100g. Counts above the class C threshold of 46,000 *E. coli*/100 g were recorded on two dates, 11 June 2014 at three sites and 14 July 2019 at Porthilly Oysters. On the first of these dates, counts were not associated with any measured environmental conditions liable to generate such high values. These measurements have therefore been excluded from the statistical analysis. On the second date, the solitary value (>180,000) has also been excluded from our analysis. This does not mean that the values are necessarily incorrect or unimportant, but they are not compatible with the statistical pattern of the remaining data. They cannot be treated in a statistical analysis but need separate consideration.

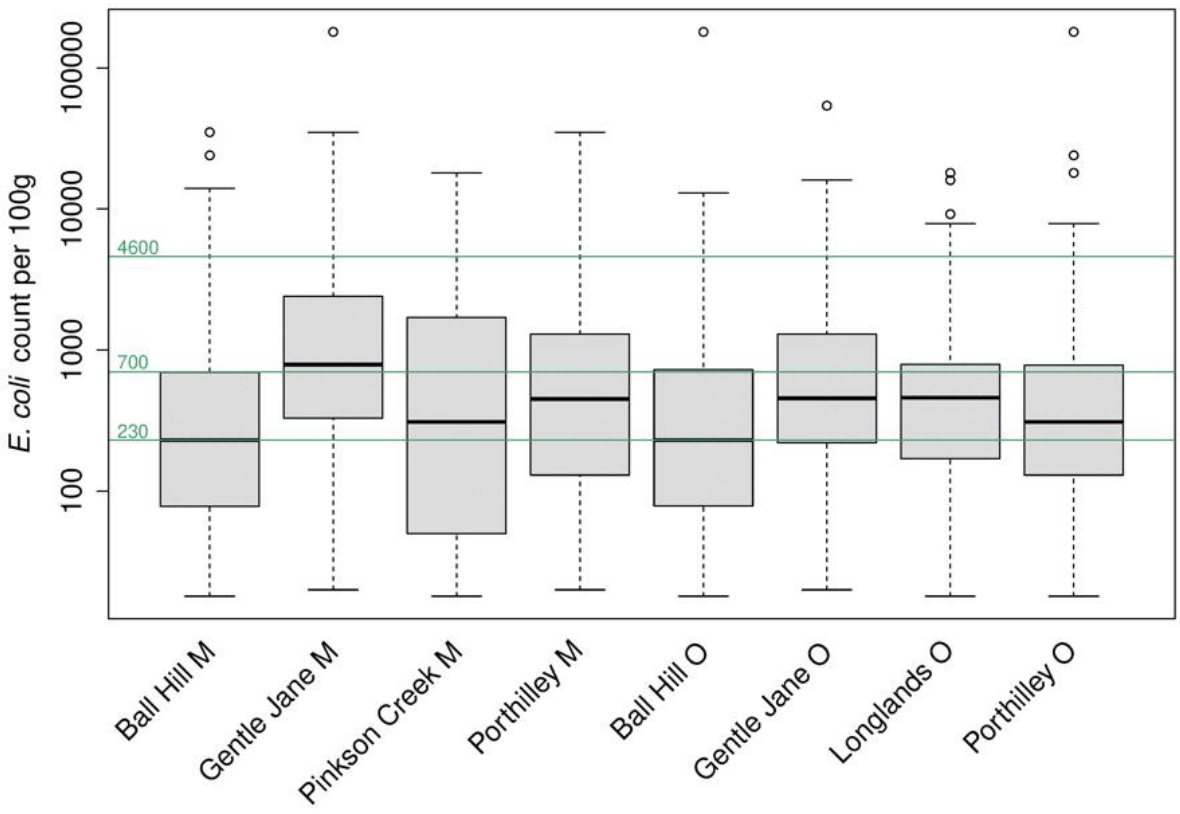


Figure 6.5.3. Boxplots of MPN *E. coli*/100g in oysters and mussels in the Camel Estuary, combining Official Control data and supplementary data collected during the DASSHH project. The heavy bar shows the median and each box covers the interquartile range.

Table 6.5.1. Proportion of MPN *E. coli*/100g samples in relation to shellfish bed classification, for both oysters and mussels in the Camel Estuary, combining Official Control data and supplementary data collected during the DASSHH project.

Site	Propn<230	Propn>230,<700	Propn>700	Propn>4,600
Ball Hill_Mussels	0.53	0.24	0.23	0.06
Gentle Jane_Mussels	0.18	0.24	0.58	0.12
Pinkson Creek_Mussels	0.49	0.16	0.35	0.13
Porthilly_Mussels	0.4	0.21	0.39	0.07
Ball Hill_Oysters	0.57	0.18	0.25	0.03
Gentle Jane_Oysters	0.35	0.28	0.37	0.05
Longlands_Oysters	0.33	0.35	0.33	0.05
Porthilly_Oysters	0.46	0.26	0.29	0.03

6.5.2 Explanatory variables

The explanatory environmental data tabulated in Stage 1 was extended to cover the period of the project and, additionally, sea surface temperature was included (Table 6.5.2).

Table 6.5.2 Summary of the explanatory variables investigated (updated)

Variable	From	To	Source	Comments
Combined sewer overflows	01/2004	12/2019	SWW	Variable record length
River/estuary water quality data	01/1991	12/2015	EA	Sparse data and short record. Not used in this analysis
River flow data	01/1991	03/2020	EA	
Met. Station data	01/1991	03/2020	BADC	
Radar rainfall data	01/2014	03/2020	BADC/NIMROD	
Tidal state	01/1991	03/2020	www.ntslf.org	Ilfracombe
Sea surface temperature	01/2000	03/2020		Extrapolated to 2019, 2020

6.5.2.1 Combined sewer overflows

Combined sewer overflows (CSOs) and sewage treatment works (STWs) may discharge untreated sewage, diluted by rainfall runoff, during extreme wet periods. The names and grid references of CSOs and STWs having data along the Camel estuary and near-catchment, provided by South West Water, are given in Table 6.5.3.

Table 6.5.3. Location of Camel estuary CSOs.

Location	**Easting**	**Northing**
Egloshayle Pumping Station	199709	72074
Little Petherick STW	191820	72580
Moyle Road CSO	192250	74780
Nanstallon PSEO	203500	67300
Padstow Foreshore Pumping Station	192240	74920
Padstow Harbour Pumping Station	192010	75450
Porthilly CSO	193379	75240
Porthilly Sewage Treatment Works	193510	74870
Rock Pumping Station	193070	75600
Sarah's View Pumping Station	192110	74430
Wadebridge PS	198850	72720

Figure 6.5.4 shows the locations of the CSOs and STWs discharging to the Camel estuary and Nanstallon, the main STW for Bodmin, and Figure 6.5.5 the location of those close to the shellfish beds for comparison.

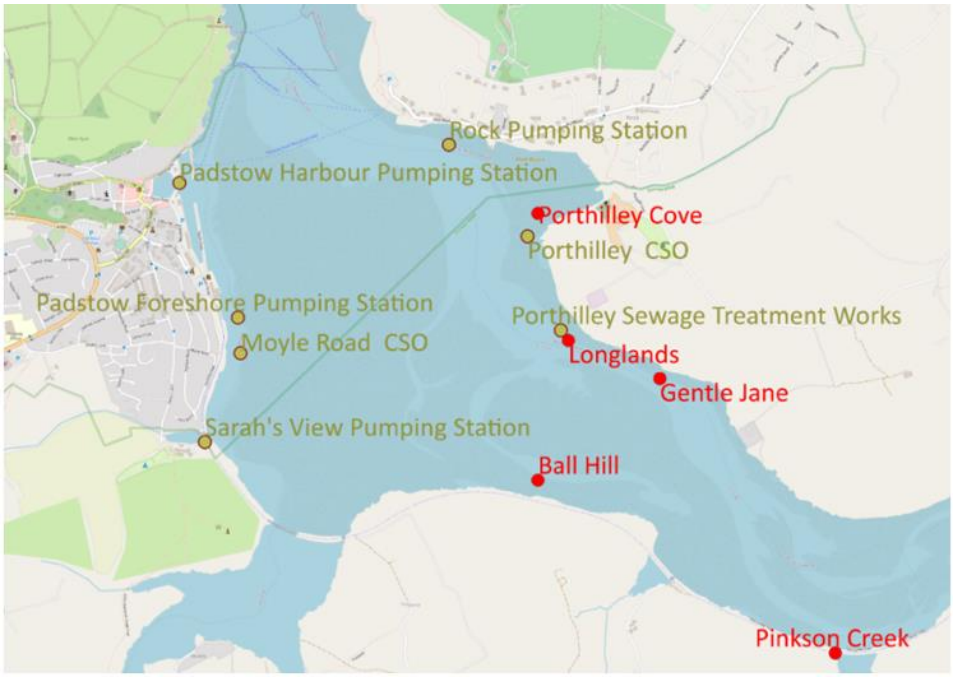


Figure 6.5.5. Location of combined sewage overflows and pumping stations (in green) close to shellfish beds (in red) in the Camel estuary.

Operational durations derived from the CSO start and stop times were aggregated to provide daily time series. However, it was uncertain whether the CSO data provide a full or partial record of operation. The sporadic nature of some records suggests either the data are incomplete for some years, or the values are a reflection of changes in operation (or malfunction) affecting operation. For these sporadic records the operational duration does not appear to be related to rainfall. For fuller records the durations generally show a stable pattern related to rainfall amount.

The CSOs closest to the estuary shellfish beds are Porthilly STW and WWPS, Rock WWPS and St Miniver STW. The records from these 4 sites are treated as independent, although it may be that the spillages from each are related, for example through redirecting discharge between sites. The record of spills from Porthilly STW is restricted to 2012 and early 2013. During this period the major spillages were associated with high winter flows, and other shorter spillages in summer and autumn of 2012 were associated with rainfall events in a proportionate way. There is no indication of any major spillage during this period whose influence on shellfish contamination would be distinguishable from an effect associated with high rainfall. The fact that there is no later record for Porthilly STW suggests this site was either no longer operational or was not being monitored after 2013. The discharge location of St Minver STW is uncertain, but is in the vicinity of Porthilly with records for 2010 and 2011. The generally short duration spills were associated with rainfall events over the winter period. Rock WWPS shows very occasional spillages throughout the period which are not always associated with rainfall events. It is not clear why these spillages occurred, but they were not obviously associated with any increase in shellfish contamination. Porthilly WWPS has a record of spillages which is broadly consistent throughout. A period of long duration spillages in July 2007 appears to be the result of some change in operation but does not appear associated with a recorded increase in shellfish contamination. Otherwise, recorded spillages are only loosely associated with rainfall.

Four CSO sites are located on the Padstow shore of the estuary. Amongst this group Padstow Foreshore pumping station was pre-eminent, with very frequent spillages. Moyle Road CSO is close by, and the records of the two sites suggest they are related, with diversion of spillages between the two. The records from these two sites are clearly rainfall-related. Of the remaining two sites, Padstow Harbour WWPS gave spillages in conditions of very high rainfall in the winters of 2012/13 and 2013/14, but apparently not during equally wet periods at other times. Sarah's View pumping station showed some sporadic spillages in 2013/14, not obviously related to rainfall.

Data has been provided for four further sites, none of which are close to shellfish beds. Little Petherick STW, at the head of a minor western tributary of the estuary, shows sporadic spills for most of the period but appeared to have become more active during the winter of 2019/20. The remaining sites are located on the main river near or upstream of the tidal limit and would appear to be less likely to influence shellfish contamination. All were responsive to rainfall, at times when they are operating. Nanstallon STW and Egloshayle WWPS do not show consistent behaviour throughout the period and may have an incomplete record or have undergone change during the period.

Those CSO records that were inconsistent over the period are not suitable for statistical modelling. While they may have had some influence on shellfish contamination, this cannot be statistically characterised with the limited data available. In the light of this, the best and longest records for Foreshore/Moyle Road and Wadebridge CSOs only are used in modelling.

The duration of operation of CSOs does not provide a measure of the volume of water being released, nor the bacteriological quality of the effluent. The availability of these data would in all likelihood greatly improve the modelling.

6.5.2.2 River water quality data

The *E. coli* that contaminate shellfish are derived from the water that flows over them, largely derived from locations in the catchment, both diffuse (agricultural) and point (municipal sewage) sources. Measurements of *E. coli* in river and estuary water may therefore provide an indication of likely shellfish contamination. Where these are associated with other water quality variables, those may also be potential shellfish contamination indicators. Nevertheless, these data are unlikely to be available in real time for operational decision-making.

The UKCEH Water Information Management System (WIMS) database of EA water quality data comprising concentrations of measured determinands at 177 sites within the catchment and 123 sites within and around the tidal part of the estuary. The catchment sites include EA Harmonised Monitoring Scheme (HMS) site at Polbrock Bridge. Water quality measurements have been made for a variety of purposes, including routine monitoring of rivers, monitoring sewage discharges and trade and agricultural effluent. The most recent data available from the WIMS database are for 2015, with varying coverage between locations in terms of both sampling rate and determinands measured. The key HMS site at Polbrock Bridge has data from 1974 for conductivity, 5-day BOD (with and without suppression of nitrification), total and dissolved organic carbon, total oxidised nitrogen, nitrate, un-ionised ammonium, suspended solids, soluble reactive phosphorus (orthophosphate) and total inorganic phosphorus. Other sites have a subset of these determinands, with some sites having other determinands where a particular contaminant is being investigated. Estuary measurements typically exclude any nitrogen species data.

In practice the sporadic nature of these data over much of the historical period being considered makes them unsuitable for use in statistical modelling. The availability of intensively sampled water quality measurements would undoubtedly be of value in supporting model construction.

6.5.2.3 River flow data

The EA river flow record at Bodmin (Dunmere) was extended to cover the period May 2019 to March 2020, and converted to an estimated flow at Denby using the calibration equation given earlier.

6.5.2.4 Met station data

Daily precipitation data from Cardinham were extended to cover the project period. Solar radiation, wind speed and wind direction data were extracted for the whole period from the Cardinham

historic data record. These measurements will only be indicative of values when extrapolated to the estuary, and for both of these variables it is estuary values which will be important rather than values at the Met site. This is in contrast to precipitation data whose influence is largely an aggregate over the catchment through the generation of river runoff to the estuary.

6.5.2.5 Radar rainfall data

Meteorological variables measured at Cardinham, at the eastern edge of the catchment may not be representative of meteorological conditions having greatest influence on the Camel estuary. Radar rainfall data from the Met Office Nimrod database held by the BADC was extracted. Rainfall estimates for the Camel catchment are provided by coverage from the Predannack radar record, with the data being available in what is considered a reliable form from 2014 onwards. The data have been aggregated from 5-minute rainfall estimates on a 5x5 km grid to hourly and daily values, giving six groups of grid squares to retain as possible explanatory variables (Figure 6.5.6).

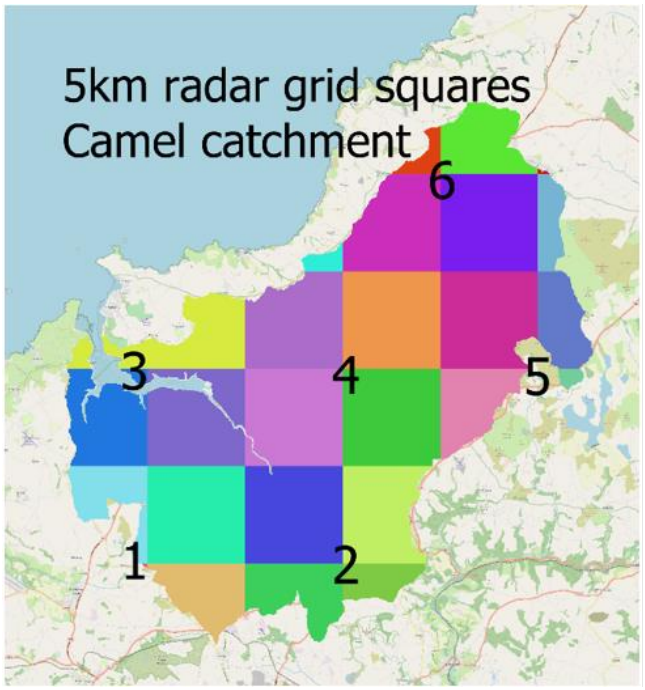


Figure 6.5.6. Aggregates of four 5x5 km rainfall radar grid squares. Each number 1 to 6 shows the centre of an aggregate of four squares.

6.5.2.6 Tidal conditions, sea surface temperature and salinity

Hourly tidal gauge and storm surge data were obtained from the National Tidal and Sea Level Facility database for the two tide gauge stations nearest to the Camel: Newlyn and Ilfracombe. The daily range of the Ilfracombe data were computed to provide an indication of neap and spring tides.

Daily sea surface temperatures (SST) at Padstow were extracted from the Copernicus Marine Service database of modelled values for the Atlantic-European North West Shelf. The data are provided on a 5km grid, and while the true temperature in the estuary will be influenced by river inputs and

shallower water depth, the SSTs are expected to provide a good estimate of the seasonal variability in estuary temperatures, which peak in late summer and are not as variable day-to-day as air temperatures. There is reason to expect temperature-dependent seasonal variability in biological activity, although temperature may not itself be a seasonal cause of measured differences. Using sea surface temperature gives a means of accounting for seasonal variability in response.

While salinity measurements were taken at the time of sampling, in this tidal estuary values vary considerably through the day. Overall salinity in the sense of the strength of the marine signal is a location variable which could be assigned a notional value. In our analysis any location difference in overall salinity is subsumed within the separate analyses for each site.

6.5.2.7 Synthetic explanatory data – a non-linear risk factor

The simplest exploratory regression models assume a linear relationship between the response and explanatory variables. Evidence from error patterns following model fitting may suggest quadratic or non-linear terms should also be included. New putative, possibly non-linear explanatory variables based on subject matter understanding were constructed. This, combined with exploratory data analysis, suggested that shellfish contamination in the Camel estuary is associated with heavy rainfall following a dry period. This is consistent with a "flush" effect, which has been well-studied in ephemeral rivers. A risk factor was constructed which incorporates a need for preceding dry conditions and heavy rainfall to account for an interaction effect between rainfall and dryness in an intuitive way. The risk factor is calculated as a combination of a risk associated with low flow (risk1) and heavy rainfall (risk2) to give an overall risk (t) as shown in the equation below, where the index t denotes time in days, Flow is daily mean flow in m^3s^{-1} and Rain is daily rainfall in mm:

$$risk1(t) = 0.98 \times risk1(t-1) + 1/\left(0.1 + \exp\left(Flow(t)\right)\right)$$
$$risk2(t) = 0.95 \times risk1(t-1) + Rain(t) + 50. \times \sqrt{\max(0, Rain(t) - 25)}$$
$$risk(t) = risk1(t) \times risk2(t)/100 + risk2(t)/100$$

Both risk1 and risk2 tend to decline at higher flows with no rain. A sequence of low flows tends to increase risk1, while rainfall increases risk2, with a high penalty for rainfall above 25mm in a day. There is clearly some arbitrariness in this particular formulation of a risk factor, but it encapsulates prior understanding of conditions under which contamination is likely to occur (Campos *et al* 2013). Beyond simple statistical modelling, the use of non-linear factors based on process understanding is likely to be a way forward in developing risk-based management strategies. The parameters included in such a synthetic variable could in principle be estimated from measured data. This risk variable is added to the base selection.

6.5.2.8 Other explanatory data not included

Land management and specific management operations are a possible stimulus to the release of *E. coli* to surface waters and, hence, shellfish contaminant (Campos *et al*, 2013). Land cover in itself is generally mapped as fixed, and the influence of particular land covers in particular locations cannot usefully be included in an estuary response model. Management operations (e.g. slurry spreading) can also pose a risk. These need to be located in time and space to be used as explanatory variables, but such data were available for the project.

6.5.3 Statistical modelling

Annual sequences of response and a selection of explanatory variables for 2004 to 2020 were plotted in Appendix A.6.1. MPN counts are recorded on a log scale on the left hand vertical axis. Other variables are plotted on the linear scale on the right hand axis. Tidal range is in m and sea surface temperature in ^{o}C. The statistical relationship between the 8 related sequences of counts (beds) was investigated before attempting to construct a model driven by explanatory environmental variables. Paired correlations between $\log_{10}$ values are shown in Figure 6.5.7. These plots are based on all data from 1991 to 2020 where a paired sample has been taken on the same day. The number of points varies according to the extent of the data record for different sites. In most cases there is a weak to moderate correlation between counts. The strongest correlations are between Gentle Jane mussels and oysters (0.54), Porthilly mussels and oysters (0.57) and Ball Hill and Pinkson Creek mussels (0.62) suggesting a common contamination source at these three pairs of sites. Longlands oyster counts also have a high correlation with the two Gentle Jane counts. These three sites are at a similar location on the north shore of the estuary. Positive correlations tend to indicate a common response to environmental drivers. The lack of high correlation between sites is indicative of either contamination, which has more of a random character, or is due to site by contamination source interactions.

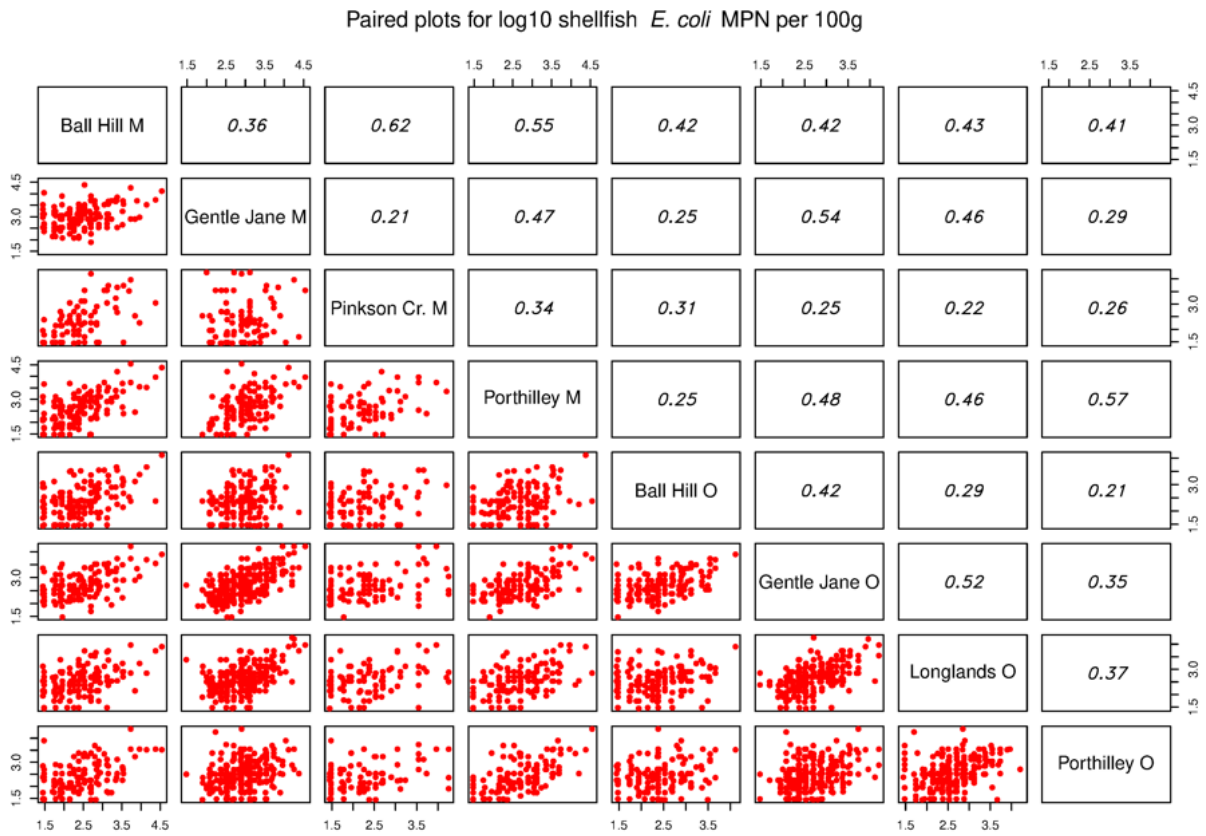


Figure 6.5.7. Paired plots and Spearman correlations between MPN *E. coli* counts in mussels and oysters from shellfish beds in the Camel Estuary, combining Official Control data and supplementary data collected during the DASSHH project from 1991-2020.

6.5.3.1 Selection of explanatory variables

Shellfish contamination is cumulative and is likely to be a response to catchment sources of *E. coli* reaching the estuary over a period of days. For past history, which might have influenced contamination, in addition to sampling day values, totals (rainfall) or means (flow, radiation) for the previous days 2-3, 4-7, 7-14 and 14-28 were included. Only present-day sea temperature was included, since this is only slowly varying. Tidal height and range are considered only on the day of sampling. To include a seasonal effect related to day length in the model, the year was divided into four seasons, from January 1. The data are not considered at a finer scale than daily, since information on timing within a day is not available for *E. coli* counts.

A large number of highly correlated investigatory environmental variables are available for use as possibly being associated with shellfish contamination, particularly with the inclusion of lagged variables. This gives potential for overfitting, and the undue influence of outliers in determining apparently significant drivers. Some reduction in the number of explanatory variables was therefore required before proceeding with modelling.

Radar rainfall data are available for six aggregates of four 5x5 km squares in the catchment. These will be highly correlated, and, with the inclusion of lagged variables, would add significantly to the number of potential variables. Therefore, only square 3 of the group was considered (indicated in

Figure 6.5.6), as this cluster is closest to the estuary. With elimination of some of the CSOs and radar rainfall squares, and inclusion of site, seasonal and trend effects, sea surface temperature, the risk factor and lagged values for the remaining hydrological variables and daily radiation comprised the full selection of variables for possible inclusion models listed in Table 6.5.4.

6.5.3.2 Base models

Model estimation used stepwise regression starting from a base model. Variables included in each base model, Models 1 to 5 of Table 6.2.4, comprise a subgroup from the full selection of variables shown in the first column.

Model 1 included a complete suite of variables, with the exception of the radar data. It used Cardinham rainfall with lagged values as its rainfall explanatory variable. Model 1 was run with data starting from 2004, before which data is patchy and may not be representative of present day conditions.

Model 2 replaced Cardinham rainfall data by Region 3 radar data. Since radar data run only from 2014, when comparing the relative usefulness of radar and Cardinham precipitation data we repeated a model 1 run with only data starting from 2014.

Model 3 replaced the daily total radar rainfalls with daily maxima measured over an hour. For lagged variables it used the maxima over the whole of the lagged period in question.

Models 4 and 5 use only rainfall totals for the previous days 2-3 and 4-7, either Cardinham (Model 4) or radar (Model 5) values, together with sea surface temperature and the risk factor.

Table 6.5.4. Base models for prediction of *E. coli* in mussels and oysters from shellfish beds in the Camel estuary, showing variables selected. Models were based on MPN *E. coli* counts in mussels and oysters, combining Official Control data and supplementary data collected during the DASSHH project.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Date (for trend)	√	√	√		
Season	√	√	√		
Site (all site fit only)	√	√	√	√	√
Flow	√	√	√		
Flow.Lag.3	√	√	√		
Flow.Lag.7	√	√	√		
Flow.Lag.14	√	√	√		
Flow.Lag.28	√	√	√		
Precip_Card	√				
Precip_Card.Lag.3	√			√	
Precip_Card.Lag.7	√			√	
Precip_Card.Lag.14	√				
Precip_Card.Lag.28	√				
Tidal Height	√	√	√		
Tidal Range	√	√	√		
Foreshore.WWPS	√	√	√		
Foreshore.WWPS.Lag.3	√	√	√		
Foreshore.WWPS.Lag.7	√	√	√		
Foreshore.WWPS.Lag.14	√	√	√		
Wadebridge.WWPS	√	√	√		
Wadebridge.WWPS.Lag.3	√	√	√		
Wadebridge.WWPS.Lag.7	√	√	√		
Wadebridge.WWPS.Lag.14	√	√	√		
Hourly_Mean_Radiation_KJm.2	√	√	√		
Radiation.Lag.3	√	√	√		
Radiation.Lag.7	√	√	√		
Sea_surface_Temp_deg_C	√	√	√	√	√
risk	√	√	√	√	√
Precip_region3.Max			√		

Precip_region3.Max.Lag.3			√		
Precip_region3.Max.Lag.7			√		
Precip_region3.Max.Lag.14			√		
Precip_region3.Max.Lag.28			√		
Precip_region3		√			
Precip_region3.Lag.3		√			√
Precip_region3.Lag.7		√			√
Precip_region3.Lag.14		√			
Precip_region3.Lag.28		√			

6.5.3.3 Model fitting

Modelling is by stepwise regression using the StepAIC routine of the computer package R. To avoid overfitting, initially, StepAIC was applied to 20 random subsets of 70 per cent of the total data, using the Akaike criterion (k=2 in StepAIC). These 20 runs may not all produce the same reduced model, depending on the subset of data selected. However, for a good model, it is expected that there will be some consistency between the reduced models. Having run the 20 stepwise regressions, those variables which have been selected at least 7 times were retained. Variables which appear less frequently than this were considered spurious. Following reduced variable selection, a final stepwise regression using the stricter Bayes criterion (k=log(n) in StepAIC is carried out. This gave the final reduced model (FRM).

Each FRM has an associated adjusted r^2 value and residual standard error. The former gives an indication of the quality of fit of the model, and the latter can be used, along with the model fitted values, to derive exceedance probabilities for thresholds.

6.5.3.4 Regression analysis results

Table 6.5.5 shows adjusted r^2 values for the 5 FRMs, once stepwise regression has been completed. The first Model 1 column shows the adjusted r^2 value using data from 1991, and the second, data from 2014, to enable comparison with Model 2. In general, use of data from 1991 seems to give a better fit than the more restricted dataset from 2014. A comparison between Model 1 (Cardinham) and Model 2 (radar) suggested Model 2 is slightly better. Model 3, which used rainfall maxima, was never an improvement on Model 2, and was rejected. Model 5, which used radar rainfall, appeared to be a slight improvement on Model 4. Note that Model 5 r^2 values are slightly lower than Model 2.

Table 6.5.5. Adjusted r^2 values for the five final reduced models for prediction of *E. coli* in mussels and oysters from shellfish beds in the Camel estuary. Models were based on MPN *E. coli* counts in mussels and oysters, combining Official Control data and supplementary data collected during the DASSHH project.

Site	Model1 (all years)	Model1	Model2	Model3	Model4	Model5
All sites	0.33	0.36	0.37	0.34	0.34	0.35
Ball Hill Mussels	0.45	0.38	0.37	0.24	0.34	0.33
Gentle Jane Mussels	0.27	0.22	0.26	0.17	0.25	0.21
Pinkson Creek Mussels	0.28	0.22	0.39	0.22	0.18	0.19
Porthilly Mussels	0.49	0.33	0.32	0.32	0.31	0.31
Ball Hill Oysters	0.28	0.23	0.23	0.19	0.17	0.19
Gentle Jane Oysters	0.29	0.37	0.40	0.37	0.32	0.35
Longlands Oysters	0.29	0.41	0.40	0.38	0.38	0.39
Porthilly Oysters	0.19	0.11	0.14	0.08	0.09	0.12

Three FRMs were selected: Models 1 (all years), 2 and 5. Appendix A.6.2 show regression coefficients for these models. Model 1(all years) was the most comprehensive model, and might be expected to give high r^2 values. The risk with this model is of overfitting. Model 2 appears better than Model 1, for those years where the two are comparable, so is presented in detail. Model 5 is the better of the two highly conservative models.

6.5.3.4.1 Model 1 (all years)

For Model 1 using concatenated data for all sites, taking *E. coli* counts in Ball Hill mussels as a reference, counts in Gentle Jane mussels and oysters and Porthilly mussels are significantly higher ($p<0.005$). Amongst the variables rejected by the stepwise regression is a seasonal effect when expressed as a factor with four levels corresponding to three-monthly periods commencing annually in January. Sea surface temperature is preferred as an explanatory variable representing seasonality. Cumulative daily rainfall between 15 and 28 days previously is included in the model, but between 8 and 14 days is not. This is likely to be a spurious artefact of model-fitting to the particular configuration of data. An alternative explanation could be that tidal effects were responsible for a delayed response. While possible, this is considered unlikely.

Analysis for individual beds showed that sea surface temperature and at least one lagged rainfall variable to be significant ($p<0.001$) for all site by species combinations apart from Pinkson Creek mussels, for which tidal height and the risk factor are selected. Rather than providing definitive

evidence of a tidal height influence, this suggested possible further investigation was required, as this may be an artefact. Other sites showed a variety of additional significant variables, notably risk (4 sites) and Wadebridge operation 2 to 3 days previously (Longlands oysters). Gentle Jane oysters also has tidal height. Ball Hill oysters includes river flow and two radiation variables. In this case river flow seems to substitute for rainfall as the key hydrological variable. Porthilly mussels includes Foreshore WWPS operation the previous 2 days and Wadebridge WWPS operation on the sampling day. The Wadebridge parameter is negative, which if correct would suggest that the operation of this WWPS was reducing contamination. Assuming this is unlikely, the result suggests some inadequacy in the model representation for this data set. Gentle Jane mussels includes an upward trend and a tidal range parameter.

6.5.3.4.2 Model 2

Model 2 investigated the use of radar rainfall data over the estuary, rather than using the gauge data from Cardinham Met station. The selected model for all sites included rainfall 2 to 3 days previously, together with the risk variable (accounting for preceding dry conditions and heavy rainfall) and sea surface temperature. It also included two lagged flow variables covering different time periods, one of which is negative. It is likely that these two variables are explaining some of the hydrological variability not accounted for by the radar data. The analyses for individual sites tended to show lagged radar data replacing the lagged Met office rainfall of Model 1. Sea surface temperature and the risk variable are identified as having predictive power in many of the selected models. Selections for Pinkson Creek mussels show two negative parameters and a high constant. Both increasing cumulative flow and increasing hourly radiation appear to reduce *E. coli* concentrations. This makes no sense in terms of our understanding of estuarine processes, and can be rejected as a statistical artefact. Equally, Ball Hill mussels showed a negative parameter associated with the operation of Wadebridge WWPS 4 to 7 days previously. It is unlikely that this represents a genuine effect.

6.5.3.4.3 Model 5

Model 5 allowed only Region 3 radar rainfall 2 and 3 days previously, sea surface temperature, and risk as potential explanatory variables. These were the commonest variables present in the selections made in Model 2, and make intuitive sense. Stepwise regression using all sites retains all three variables. All three are also retained in the single site analyses, with the exception of Pinkson Creek mussels (risk only) and Porthilly oysters (risk excluded). Model 5 therefore gives the most consistent pattern of results across all sites. Although some other models give a higher r^2 value for particular sites, the overall pattern of fit is not substantially different. Figure 6.5.8 shows the model fitted results from Model 5 against measured MPN values.

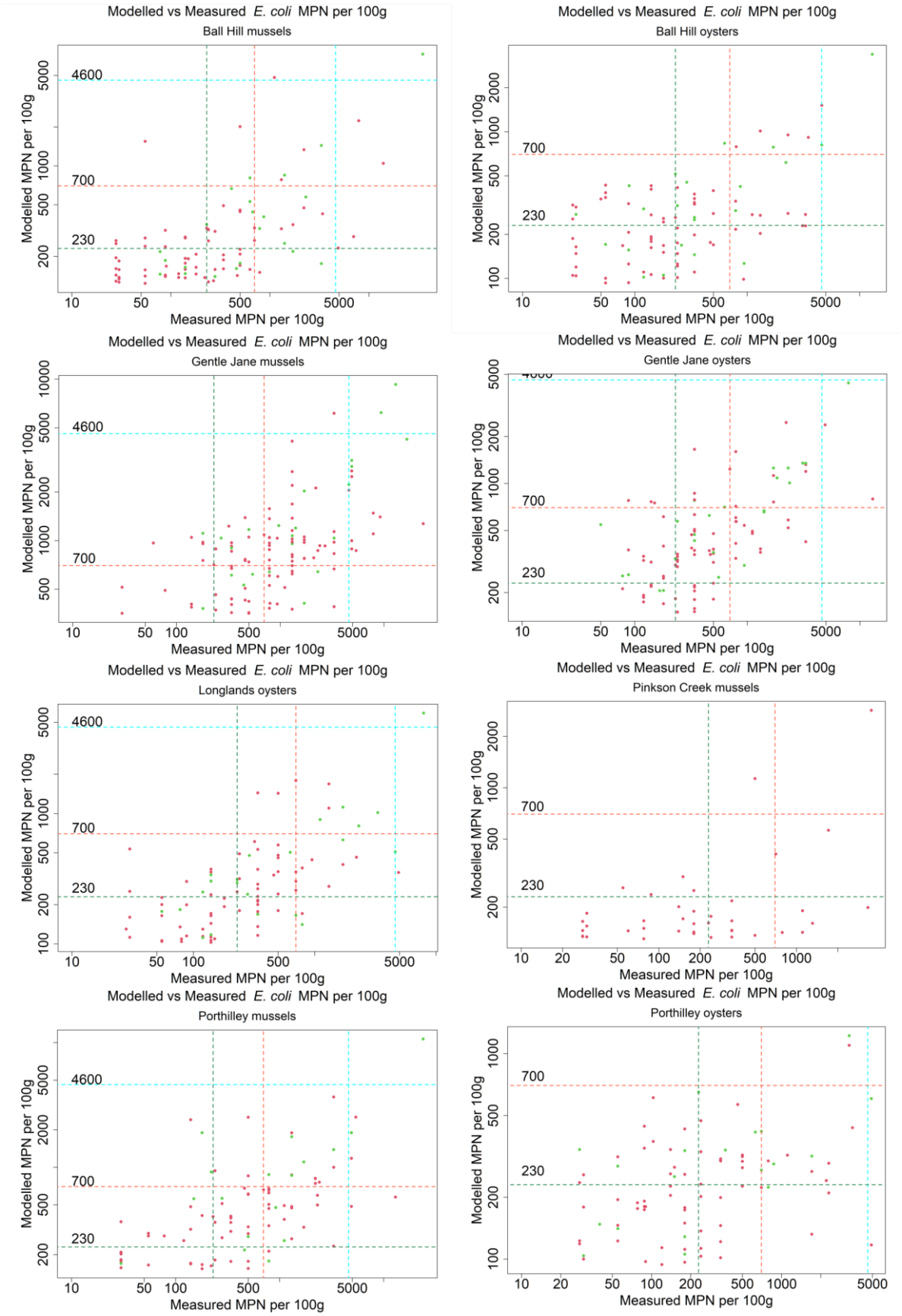
The poor fit for Pinkson Creek is apparent in Figure 6.5.8. This is partly because of a shorter data record and possibly location, as it is remote from the other sites, with the exception of Ball Hill. Porthilly oysters also had a notably poor fit. In most cases a few high unexplained MPN values had

a strong influence on the model fitting, which relies on these values being at least the right order of magnitude (see section 5 on accuracy).

Because of the small number of explanatory variables in the model, the coefficients can be concisely presented (Table 6.5.6). This gives an indication of the between-site variability in the fit of the model.

Table 6.5.6. Summary table of coefficients of Model 5 for prediction of *E. coli* in mussels and oysters from shellfish beds in the Camel estuary. Based on MPN *E. coli* counts, combining Official Control data and supplementary data collected during the DASSHH project. NS = Not significant. Abbreviated column titles refer respectively to the model 5 variables sea temperature, the synthetic risk factor and precipitation. RSE is the residual standard error and Adj. r^2 the adjusted r^2 from the model fit.

Site	**Sea temp.**		**Risk**		**Precip.**		**RSE**	**Adj. r^2**
	Estimate	p-val	Estimate	p-val	Estimate	p-val		
All sites	0.051	<0.001	0.028	<0.001	0.036	<0.001	0.48	0.35
Ball Hill mussels	0.053	0.03	0.047	0.003	0.043	0.005	0.51	0.33
Gentle Jane mussels	0.056	<0.001	NS		0.047	<0.001	0.45	0.21
Pinkson Creek	NS		0.051	0.003	NS		0.52	0.19
Porthilly mussels	0.072	0.002	0.033	0.035	0.041	0.013	0.51	0.31
Ball Hill oysters	0.059	0.02	NS		0.040	0.010	0.55	0.19
Gentle Jane oysters	0.051	0.003	0.025	0.021	0.029	0.003	0.38	0.35
Longlands oysters	0.046	0.011	0.037	0.003	0.039	0.010	0.42	0.39
Porthilly oysters	0.055	0.011	NS		0.034	0.008	0.55	0.12



Figures 6.5.8. Fitted against measured MPN values using Model 5. CEFAS data are shown in red and DASSHH supplementary project data in green. Dotted lines are the standard classification thresholds for shellfish beds.

6.5.3.5 Probability of threshold exceedance

Each model generated a point estimate of log(MPN) for a given value of the explanatory variables, together with a residual standard error. Assuming a Normal distribution of the residuals, the probability of threshold exceedance can be estimated for any given estimate of log(MPN). This is illustrated in Figure 6.5.9 for an example threshold MPN count of 230 and a modelled count of 130. The example standard error (s.e) of 0.5 is typical of the residual standard error from fitting Model 5.

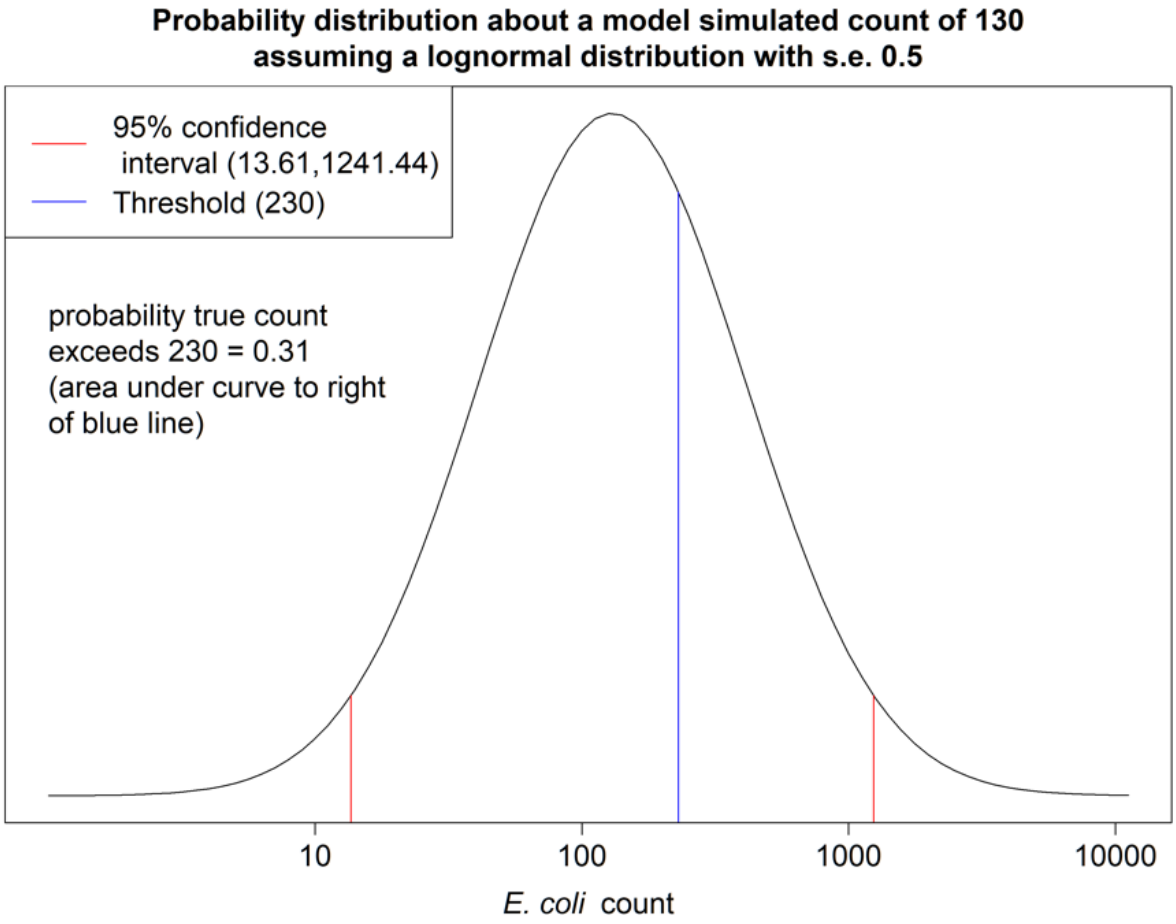
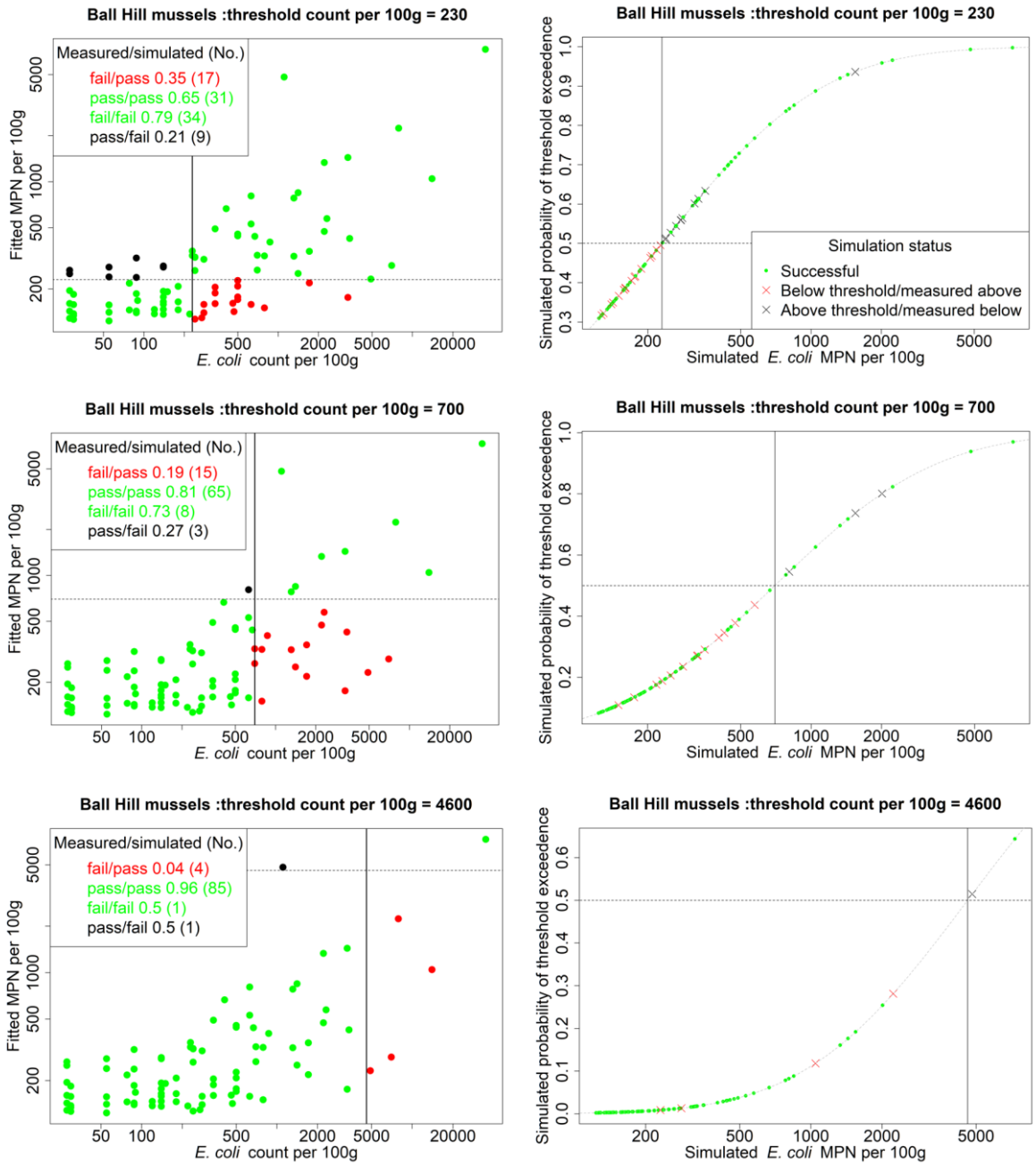


Figure 6.5.9. Probability plot for threshold exceedance

For model 5 applied to Ball Hill mussels data, critical 230, 700 and 4,600 threshold exceedance estimates are shown in Figure 6.5.10. The plots to the left all show the same measured and model-fitted data, but in relation to a different threshold. The plots to the right show, for any modelled MPN, the probability, given the selected model, that the threshold will be exceeded. These plots also show, as red and black crosses, the cases where the model predicted wrongly. Red crosses indicate where simulated values are below the threshold and measured are above, and black the converse. The closer a simulation is to the threshold value, the more likely the true MPN will be

misclassified. The further a model is from a threshold (i.e. far from borderline) the greater the confidence of a correct classification.



Figures 6.5.10. Threshold exceedance probabilities for *E. coli* in shellfish, using Model 5 fitted to Ball Hill mussels data. Based on MPN *E. coli* counts combining Official Control data and supplementary data collected during the DASSHH project.

Regardless of the model used, probability plots of the type shown in Figures 6.5.10 form a basis for decision-making in the face of uncertainty. They provide the probability of occurrence of an event given a set of environmental conditions. A correct decision will not always be made, but the likelihood of failure is known, and this can be attached to a measure of utility to estimate, under a particular decision rule. The principle of deriving probabilities of exceedance illustrated in Figure

6.5.101 can clearly be extended to derive probabilities of true counts lying within certain ranges, rather than above a threshold. By selecting a number of different ranges, a more refined decision rule could be constructed. However, the example for predictions shown here represents the best performing model for mussels at Ball Hill; predictions for other beds and species were not as reliable.

6.5.4 Summary Stage 2 – Statistical modelling of statutory RMP monitoring data (1991-2019) and field collected MPN data (2019-2020)

There is clear evidence that, in general, MPN counts are related to rainfall in the preceding days, and that there is a seasonal effect that can be accounted for in models by sea surface temperature, although this does not imply a causal relationship. The historic MPN data are characterised by small numbers of extremely high values which were difficult to characterise statistically. Some of these are not associated with preceding rainfall or any other obvious explanatory variable. Either the shellfish samples analysed were not as contaminated as suggested by the MPN count, or there were sporadic events generating high *E. coli* contaminant loads unrelated to rainfall.

While CSOs area highly probable source of *E. coli* during wet periods, the available data are not sufficient to identify or quantify a specific influence of CSO spillages on shellfish contamination, in part because of their correlation with rainfall and inherent limitations of the data.

The constructed risk factor based on subject understanding (heavy rainfall with dry preceding conditions) seems to improve prediction and was incorporated as a significant factor in the majority of models. Although there were differences in the factors that were significant in the individual site models, a few variables were commonly selected, and these form the structure of the simpler Model 5. Where models incorporated more factors, in some cases the parameter values estimated are not intuitively meaningful, for example the negative influences of CSOs.

The modelling approach generates probabilities of exceedance of threshold counts given a set of environmental conditions. This is the requirement of a decision support system for shellfish bed closure or resumption of harvesting. However, the performance of the models based on combining Official Control MPN *E. coli* data and supplementary data collected during the DASSHH were not considered sufficiently robust for use in management of the shellfish beds.

6.6 Stage 3: Statistical modelling of field collected MPN and pour plate data (2019 -2021)

The data analysed under Stage 3 excluded historical Official Control *E. coli* data, retaining only measurements made fortnightly within the DASSH project. These included not only MPN counts of *E. coli*, but also measurements of other estuary quality data, and pour plate counts of *E. coli*. The analysis explores the relationship among these variables and other explanatory variables.

6.6.1 Shellfish and field measurements: Response variables

The bacterial and water quality data analysed comprise laboratory measurements made on samples of shellfish, sediment and water collected at shellfish beds in the Camel estuary at approximately two-weekly intervals between May 8 2019 and August 8 2021. There was a break between 15 March

2020 and 16 August 2020 due to Covid restrictions. Table 6.6.1 lists the variables measured and the period of measurement. Sampling and analytical methodology are as described in Section 2 of this report.

Data are from shellfish beds Ball Hill (Pacific oysters (O), Mussels (M)), Gentle Jane (O,M), Longlands (O), Porthilly Rock (O,M). Shellfish samples were collected at east E, west W and mid M locations of each bed and bulked to a single sample for each bed on each sampling date, with subsequent analysis as described earlier in this report (Section 5). Water and sediment samples were also collected at E, W and M locations, but not bulked for analysis. Water from location M was analysed for the inorganic water quality variables (TON, NO_2, SiO_4, PO_4 and NH_4). Shellfish were also analysed for the presence of a number of viruses of human health interest and potential viral indicators specific to human and animal sources (see Section 4). Water samples were also collected at Porthilly sewage discharge point and analysed for turbidity, *E. coli* and non-*E. coli* coliforms, viruses and inorganic water quality variables.

Table 6.6.1. Response variables analysed (2019-2021)

Variable	**Units**	**Period of record (excluding Covid restricted)**
E. coli Most Probable Number (MPN)	cfu per 100g shellfish	8/5/2019-8/8/2021
E. coli pour plate (pour plate)	cfu per 100g shellfish	2/2/2020-8/8/2021
E. coli in sediment	cfu per 0.2g sediment	2/6/2019-8/8/2021
Non-*E. coli* coliforms in sediment	cfu per 0.2g sediment	30/6/2019-8/8/2021
E. coli in estuary water	cfu per 100ml water	16/6/2019-8/8/2021
Non-*E. coli* coliforms in estuary water	cfu per 100ml water	16/6/2019-8/8/2021
Viruses NoVGI, NoVGII, SaVGI, HAV, HEV, AdVF, AdVC, OAdV, AtAdV	Genome copies/g	2/6/2019-8/8/2021
Total oxidised nitrogen (TON), nitrite (NO2-), silicate(SiO4----), phosphate (PO4---) and ammonium (NH4+)	µmol L^{-1}	2/2/2020-8/8/2021
Turbidity	NTU	16/6/2019-8/8/2021

6.6.2 Explanatory variables

The explanatory variables used are those used in Stage 2. They are summarised in Table 6.6.2, with any changes noted. Environment Agency Water Quality data and tidal state were excluded.

Table 6.6.2. Summary of the explanatory variables investigated (Updated)

Variable	From	To	Source	Comments
Combined sewer overflows	12/01/2020	08/08/2021	SWW	Variable record length
River flow data	08/05/2019	08/08/2021	EA	
Met Station data	08/05/2019	08/08/2021	BADC	
Radar rainfall data	01/01/2019	08/08/2021	BADC/NIMROD	
Sea surface temperature	08/05/2019	08/08/2021		Extrapolated from 2018 values

6.6.2.1 River flow data

The EA provided 15 minute flow measurements for the river Camel at Bodmin (Dunmere). The Denby gauging site had been fully decommissioned prior to the start of the analysis period, so the data used are the untransformed Bodmin (Dunmere) data. The data have been aggregated to daily values, and lagged daily values also computed (e.g. rainfall two days previously). Missing flow data during autumn 2020 have been replaced by computed values using an autoregressive moving average (2,2) model. Over the fitting period this gave an r-squared value of 0.93.

6.6.2.2 Precipitation data

Because of a break in the rainfall record from Cardinham Met Station during spring 2021, radar data have been used as explanatory for the modelling. Flow and precipitation data are plotted against time for the Stage 3 sampling period in Figure 6.6.1.

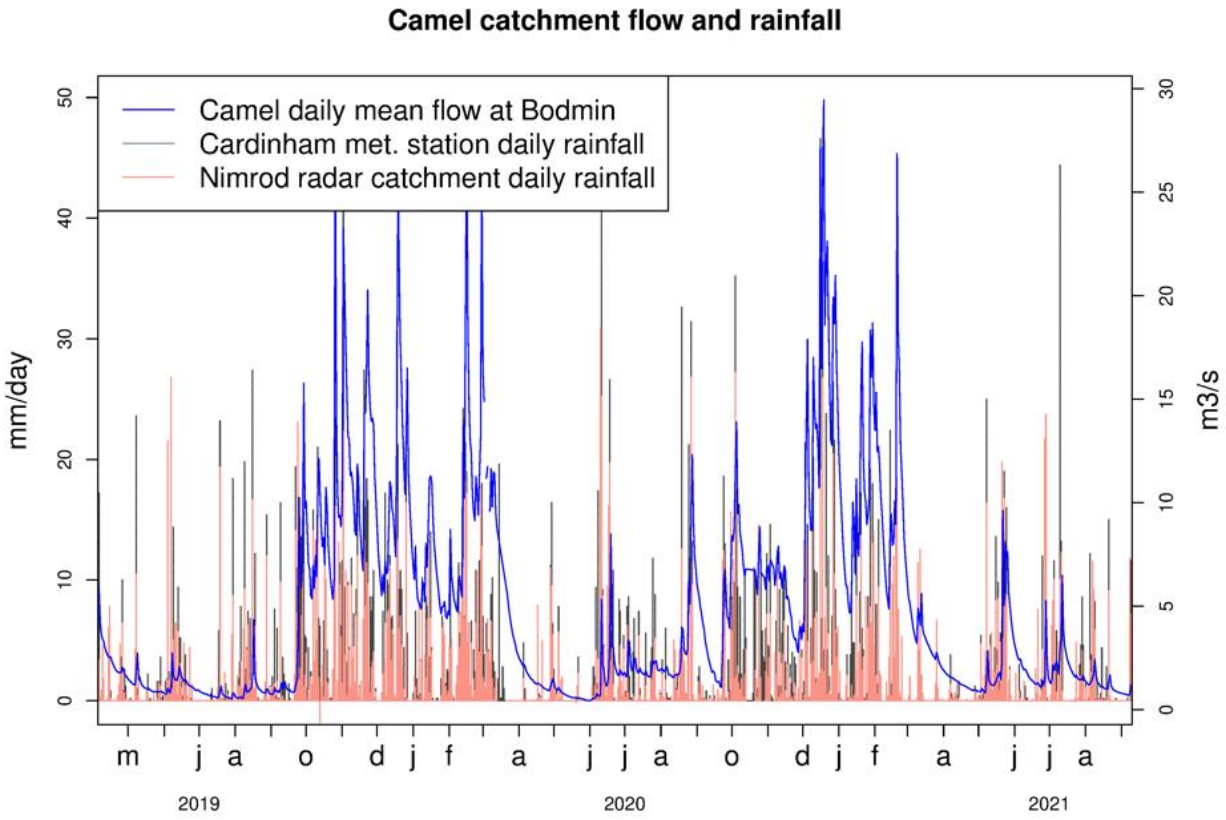


Fig 6.6.1. Camel catchment river flow (Bodmin), daily rainfall (Cardinham Met station) and Nimrod catchment daily rainfall over the study period 2019-2021.

Figures in Appendix A 6.3 show time series of response and inorganic variables against a background of rainfall and flow data.

6.6.2.3 Combined sewer overflows

South West Water provided a sequence of durations of operation and spill numbers for the CSOs and STWs listed in Table 6.2.3 for the period January 2020 to August 2021.The available data are limited since volumes of water released are not recorded. The duration of operation may be recorded as several days, with a small number of spills recorded within each period. It may not be clear on what days during operation a spill has occurred. More detailed information on spill volumes would be valuable.

6.6.2.4 Sea surface temperature

Estimates of daily sea surface temperatures (SST) at Padstow were extracted from the Copernicus Marine Service database of modelled values for the Atlantic-European Northwest Shelf. The data used are extrapolated from values for 2018, since variability on particular days between years is

very small compared to the annual cycle which is likely to have the major effect on biological processes in the estuary.

6.6.3 Statistical modelling

6.6.3.1 Exploratory data analysis

Figures 6.6.2 and 6.6.3 show the between-bed and species relationships of the key variables, log_{10} MPN and pour plate counts. MPN values cover the complete period from May 2019, while pour plate values run from the start of their record in January 2020 with the main data collected between August 2020 and August 2021. The upper right panel gives correlation coefficients for the paired variables. These are generally lower between Porthilly beds and the remainder.

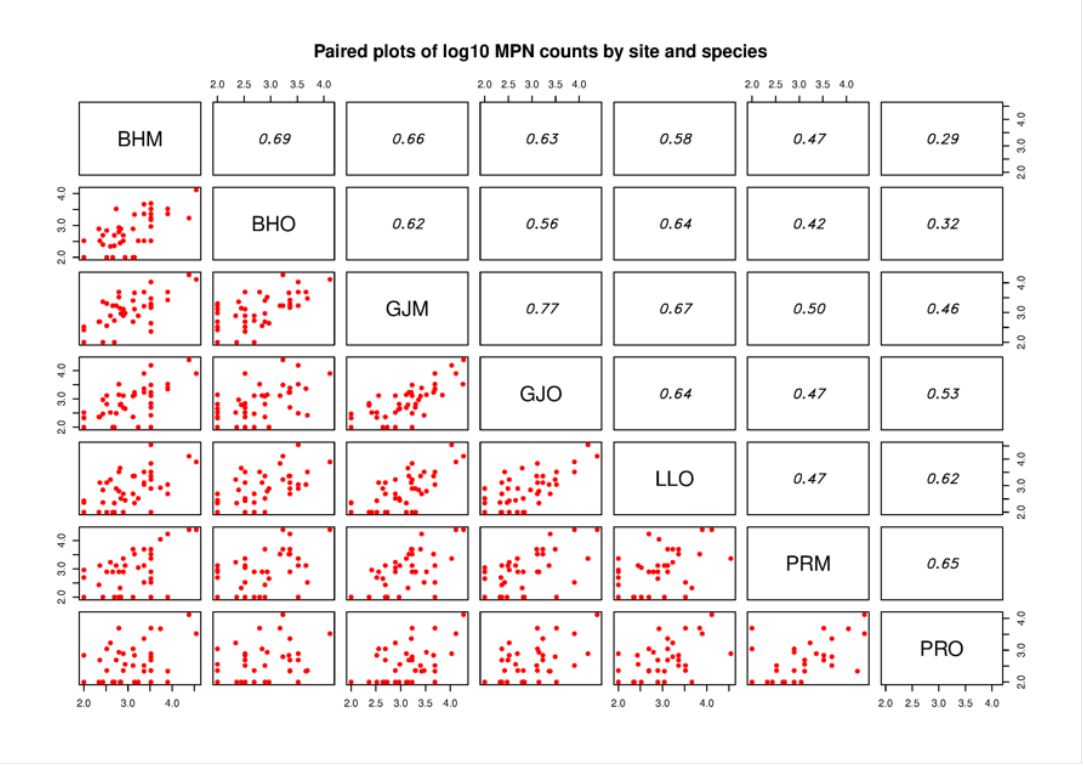


Figure 6.6.2. Paired plots and correlation coefficients for log_{10} MPN *E. coli* counts in mussels and oysters between shellfish beds in the Camel estuary between August 2020-August 2021. Sites are Ball Hill Mussels (BHM), Ball Hill Oysters (BHO), Gentle Jane Mussels (GJM), Gentle Jane Oysters (GJO), Longlands Oysters (LLO), Porthilly Mussels (PRM), Porthilly Oysters (PRO)

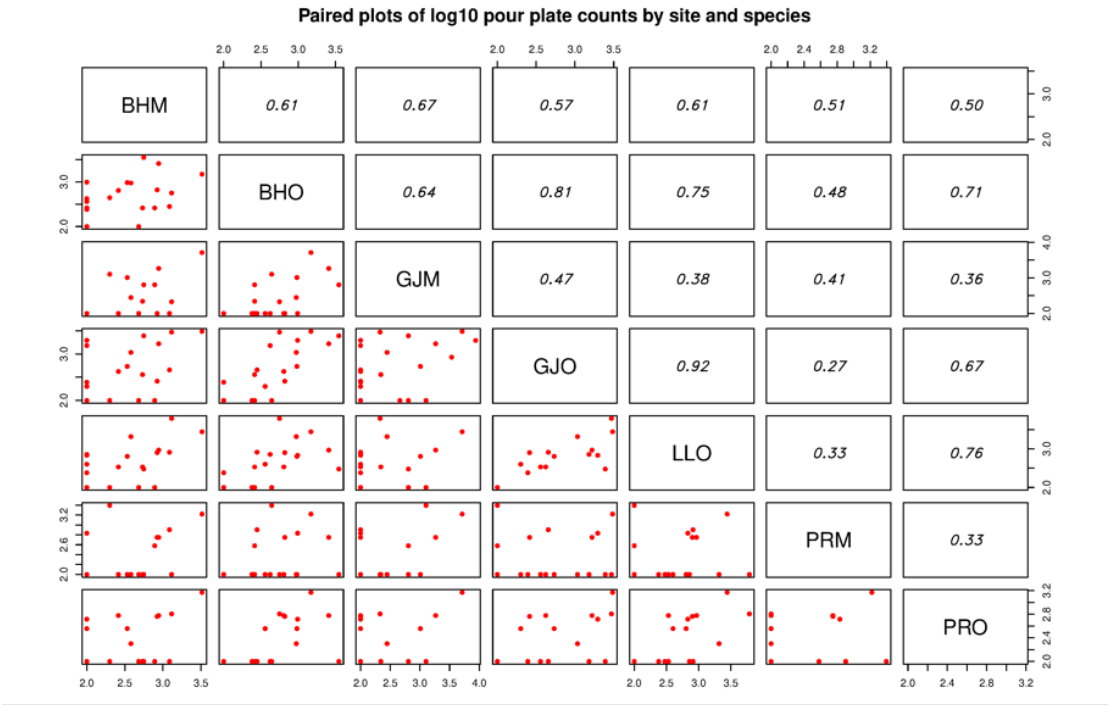


Figure 6.6.3. Paired plots and correlation coefficients for $\log_{10}$ pour plate *E. coli* counts in mussels and oysters between shellfish beds in the Camel estuary between August 2020-August 2021. Sites are Ball Hill Mussels (BHM), Ball Hill Oysters (BHO), Gentle Jane Mussels (GJM), Gentle Jane Oysters (GJO), Longlands Oysters (LLO), Porthilly Mussels (PRM), Porthilly Oysters (PRO)

Figure 6.6.4 shows the relationship between MPN and corresponding pour plate *E. coli* counts. A mean value has been taken for the replicate values quoted for Gentle Jane samples, usually a mean of 3 values. The bias towards higher MPN than pour plate values is apparent, notably the large number of high MPN counts when the pour plate count is near the lower detection limit, and the large number of exceedances of the B/C classification boundary by MPN counts compared to pour plate counts. Figure 6.6.5 shows a time series of the difference between the paired $\log_{10}$(MPN) and $\log_{10}$(pour plate) values, including the Gentle Jane replicate values individually, emphasising the apparent difference between the two methods. This discrepancy is discussed further in Section 5 of this report.

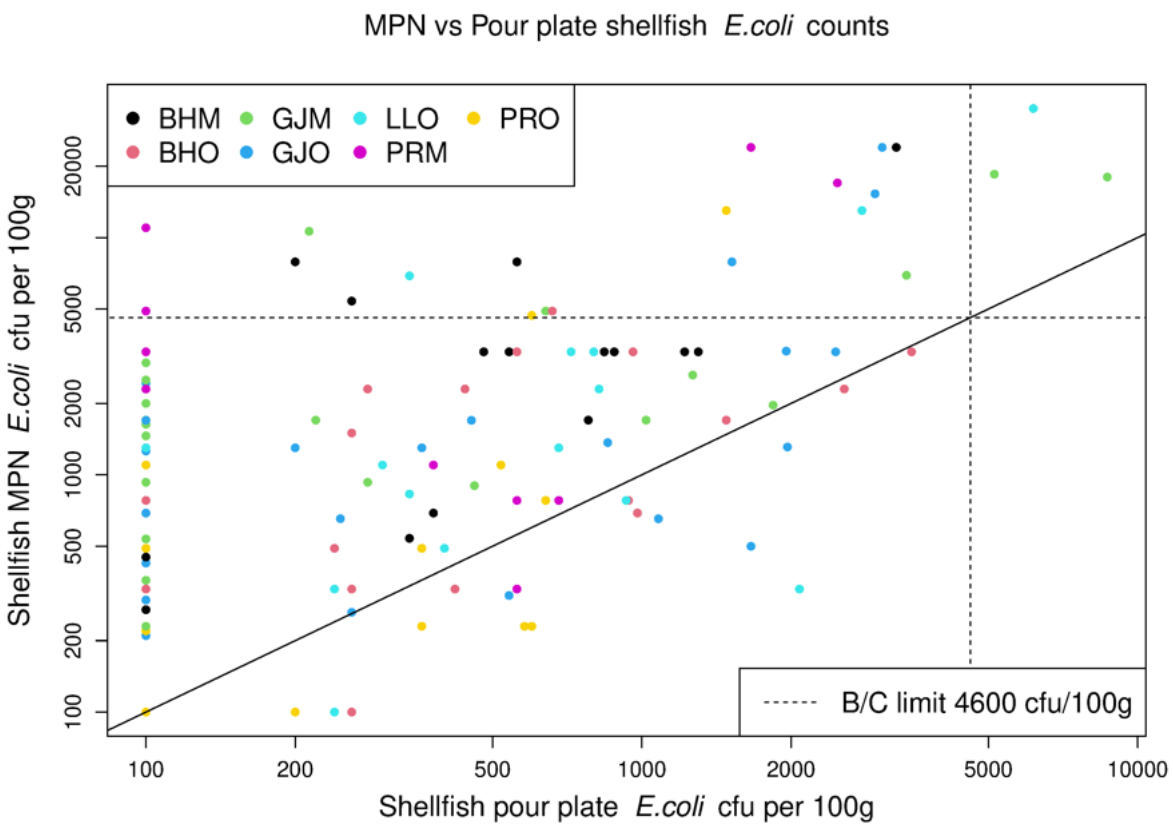


Figure 6.6.4. Scatter plot of MPN and pour plate *E. coli* counts for mussels and oysters sampled from the shellfish beds in the Camel estuary from August 2020-August 2021.

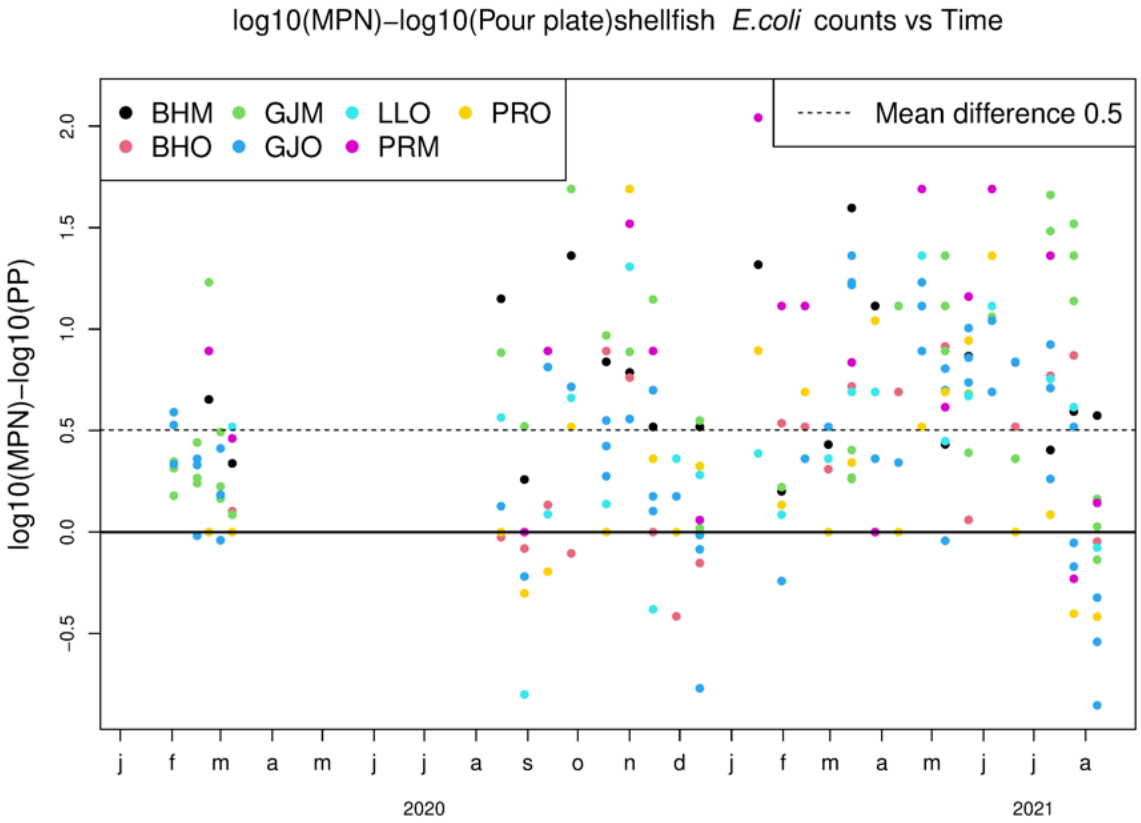


Figure 6.6.5. Time series of differences between $\log_{10}$ transformed MPN and pour plate *E. coli* counts for mussels and oysters sampled from the shellfish beds in the Camel estuary from August 2020-August 2021.

Figure 6.6.6 shows relationships between $\log_{10}$ values of all bacteriological counts. In the lower left scatter plots, different colours and symbols indicate different shellfish beds and species. There is little differentiation between them. The top right panel shows associated correlations between the variables. The relationship between shellfish counts and both sediment and water counts is poor, although counts of *E. coli* and non-*E. coli* coliforms in sediment are well correlated, as are within-water counts.

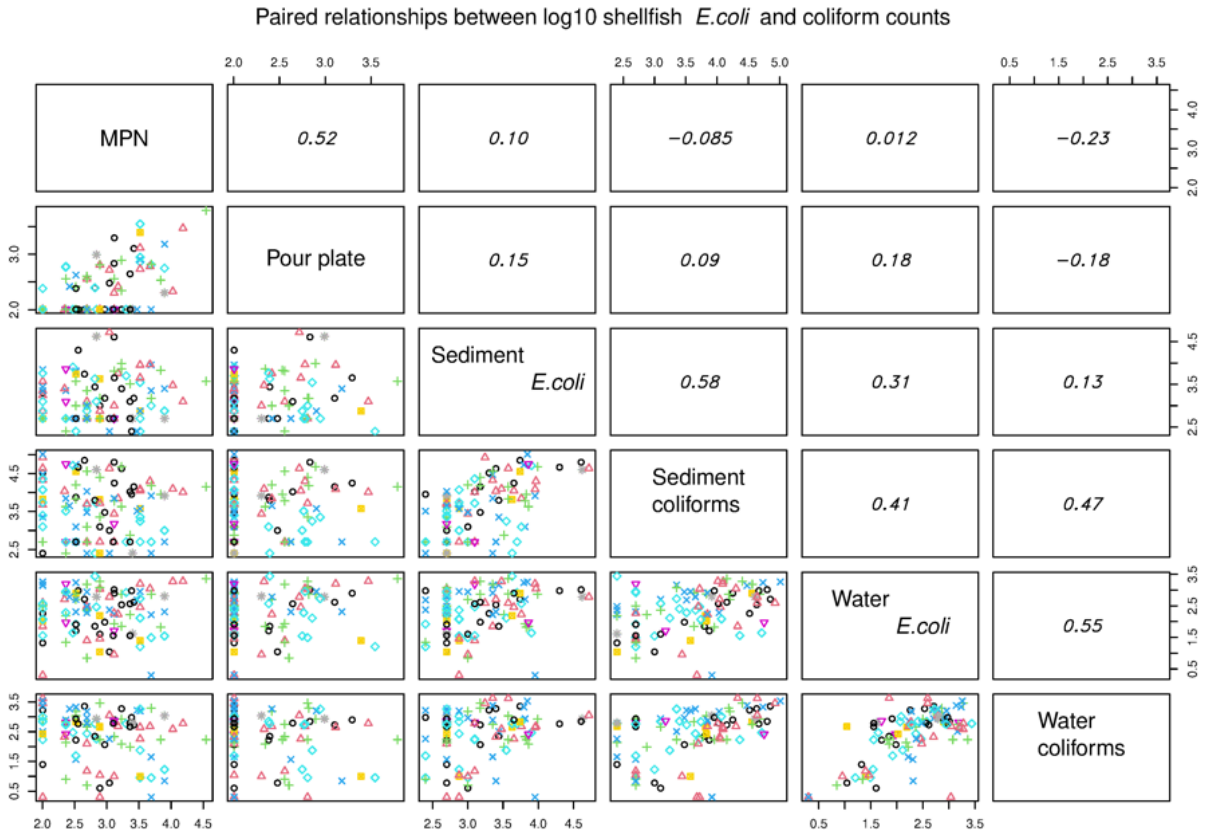


Figure 6.6.6. Paired plots and correlation coefficients between *E. coli* and coliform counts in shellfish and water over the shellfish beds in the Camel estuary from August 20220-August 2021.

Figures in Appendix A6.3 give similar plots showing relationships between $\log_{10}$ *E. coli* and coliform counts and $\log_{10}$ turbidity and inorganic nutrients. For shellfish *E. coli* counts there are no strong correlations, but both sediment and water *E. coli* and coliform counts are positively correlated with total organic nitrogen and silicate concentrations. These concentrations are thought to be governed by the salinity of the water at the time of sampling, with lower values when salinity is greater, so dependent on the state of the tide at sampling time.

Figures in Appendix A6.3 also show equivalent relationships between $\log_{10}$ *E. coli* and coliform counts and virus counts. Again, there is little structure in these data. There was a strong relationship between Human adenovirus F and Atadenovirus is apparent (cor. 0.78), and to some extent Human adenovirus C (cor. 0.37, 0.32). These virus results are discussed in detail in Section 4 of this report.

6.6.3.2 Selection of explanatory variables for modelling

A simple model was fitted to provide an indication of the limits of regression modelling using daily explanatory variables. In this model it was assumed the response of a variable in a linear regression model comprised the sum of a daily effect and a site effect. Beyond indicating variability sources, this model had little predictive power since the day effect is itself a function of other explanatory variables.

The stage 2 analysis identified lagged rainfall and a seasonal component represented by estimated sea temperature as key explanatory variables for MPN *E. coli* counts. As a base regression model these variables were considered. In addition, daily river flow, mean daily river flow on the previous two days, sea temperature and, separately, radar rainfall on days {0,-1,-2,-3 to -6) counting day 0 as the current day were also considered. Site and species were treated as factors. There is some evidence that turbidity influences *E. coli* and coliform counts in estuary water. This was, therefore, included as a potential explanatory variable.

While there was limited data on CSO inputs, these generally flow at times of heavy rain, and in a statistical analysis using some approximation to the presence of CSO inputs, the two effects were confounded. Including rainfall in the model rather than CSO operation used more reliable data and allowed for the possibility that sources other than CSOs may also be contributing to shellfish contamination, while not ruling out CSOs.

6.6.3.3 Modelling results

The r-squared values given by regressing $\log_{10}$ values of response variables on day, shellfish bed and species are shown in Table 6.6.3 as “max r-squared”. This is the fit which could be obtained with perfect prediction of the day effect, and not assuming any interaction between sites and days. Improvement on these r-squared values could be achieved, by using a differently parametrised model for each shellfish bed and species. The only response variable showing a difference between species is MPN, with oyster *E. coli*/100g counts being lower than mussel *E. coli*/100g counts ($p<0.05$). Several variables show a significant difference in sites, including all *E. coli* and coliform measurements. In all cases these variables show counts at Porthilly to be lower than elsewhere ($p<0.05$).

Following an initial regression model fit using $\log_{10}$ values of response variables and the selected possible explanatory variables, stepwise regression (R function StepAIC) using the Bayes Information Criterion (BIC) was carried out. This leads to more parsimonious models than the Akaike Information Criterion (AIC). For the more limited modelling carried out here, cross-validation was not included.

Stepwise regression yielded models whose remaining components ($p<0.05$) are summarised in Table 6.6.3. The acceptance of a variable as explanatory does not indicate the magnitude or direction of its effect on the response variable. In most cases the effect is positive. Where a current and past value of flow or rainfall are retained, the past value may have an apparent negative effect. This can be interpreted as reducing the effect of the current value, so, for example, the higher the past rainfall has been, the lower will be the effect of the current rainfall, while still being dominant. While this may be an artefact, it could be a genuine effect, for example due to flushing.

Note that most variables show that previous rainfall has some explanatory power, even for viruses. The model gives a higher r-squared value for pour plate results than for MPN counts, which is consistent with greater underlying variability in the MPN data. Nevertheless, both variables are driven by sea temperature and rainfall on the previous two days. The flow variable accepted differs but is a relatively small contribution to the model judged by the proportion of variance explained. Note that the MPN data show a difference between oyster and mussel *E. coli*/100g counts, the mussel counts being significantly higher than the oyster. No response variable showed a difference between shellfish beds in the final stepwise-reduced model, which was therefore omitted from Table 6.6.3.

Table 6.6.3. Summary of regression analysis of environmental predictor variables against a range of response variables covering *E. coli* concentrations in shellfish, water and sediment, and virus concentrations (all beds combined) for the Stage 3 modelling of 2019-2021 data. See start of this section for description of R2 and max R2.

Variable	**Turbidity**	**Species**	**Temp**	**Rain day no.**	**Flow day no.**	**R-squared**	**Max R-squared se**
Shellfish MPN	-	✓	✓	-1,-2	-1/-2	0.36	0.63
Shellfish pour plate	-	-	✓	-1,-2	1	0.50	0.65
Sediment *E. coli*	-	-	✓	-3 to -6	0	0.45	0.59
Sediment coliforms	-	-	-	0,-2	-1/-2	0.46	0.82
Water *E. coli*	✓	-	✓	-1,-2	0,-1/-2,	0.56	0.79
Water coliforms	✓	-	✓	-	0	0.28	0.80
Norovirus GI	-	-	✓	-	0,-1/-2	0.04	0.39
Norovirus GII	-	-	-	-1,-2,-3 to -6	-1	0.11	0.36
Sapovirus GI	-	-	-	-	-	-	0.25
Hepatitis E virus	-	-	✓	-2,-3 to -6	-	0.05	0.17
Human adenovirus F	-	-	✓	-1,-2	0	0.18	0.64
Human adenovirus C	-	-	-	-2	-	0.04	0.69
Ovine adenovirus	-	-	-	-	-	-	0.34
Atadenovirus	-	-	-	-2,-3 to -6	-	0.12	0.57

The presence of turbidity in the final model for estuary water beyond the effect of rainfall alone may be associated with increased turbidity due to the mobilisation of sediment by tidal or wind-driven marine action. The model fits for viruses are particularly poor, suggesting random outbreaks. The relatively good fit of water *E. coli* is not surprising, since this has a more direct link to sources than shellfish *E. coli*. The remainder of the analysis focuses on the contrast between MPN and pour plate counts. Figures 6.6.7 and 6.6.8 show the fitted against measured values following stepwise regression for these two variables.

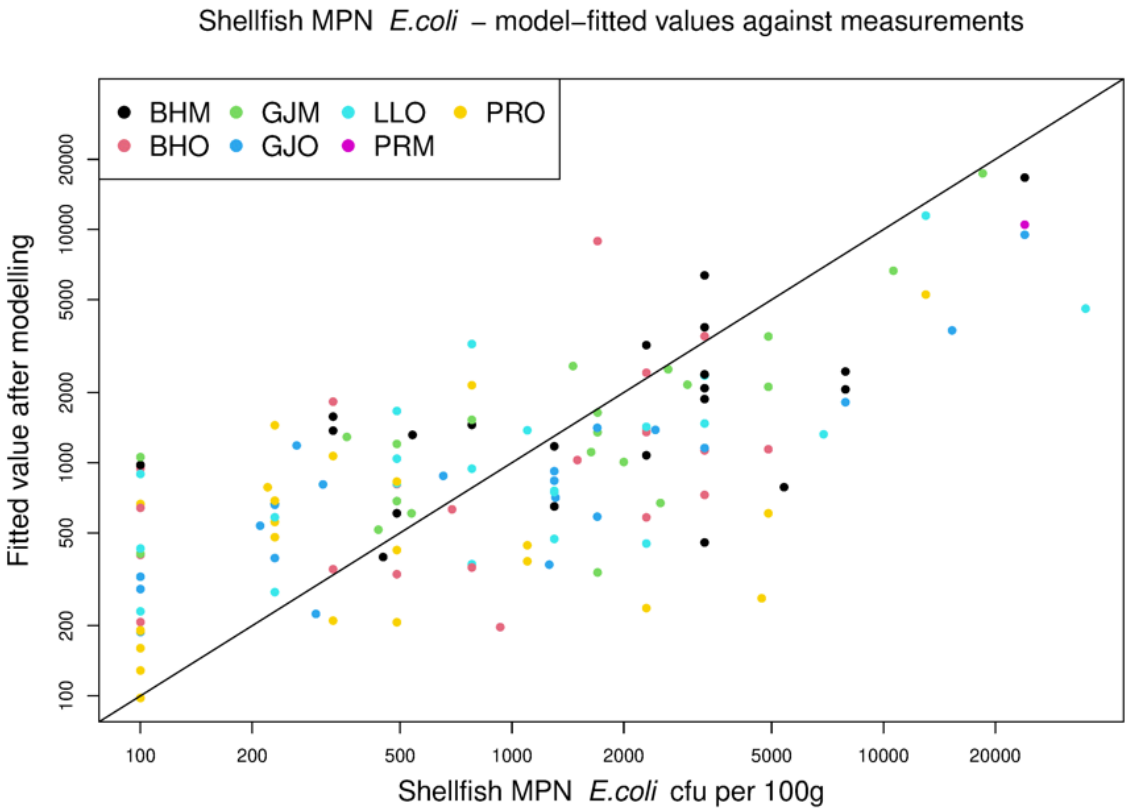


Figure 6.6.7. Plot of fitted vs measured MPN *E. coli* in mussels and oysters from the Camel shellfish beds (2020-2021)

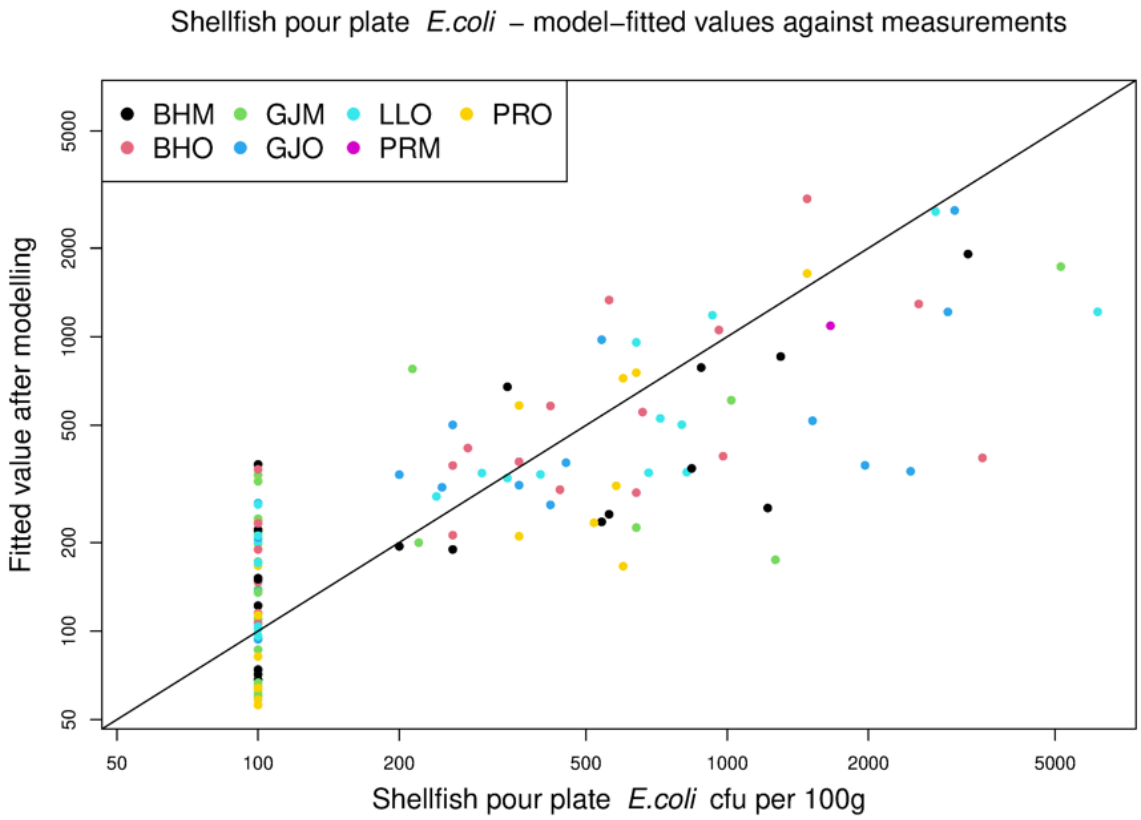


Figure 6.6.8. Plot of fitted vs measured pour plate *E. coli* counts in mussels and oysters from the Camel shellfish beds (2020-2021)

Tables 6.6.4 and 6.6.5 show in more detail the results of model fitting for individual shellfish beds and species for $\log_{10}$ MPN and pour plate counts. The fitting follows the same stepwise procedure as for the full data set (from Jan 1 2020) in each case. The variable r-squared values provide an indication of the relative predictability of counts in the different beds. However, for these analyses the numbers of data points are small, and many of the individual values are set to 100 as indicative of values below the reliable detection limit of 200 cfu/100g. This tends to inflate the r-squared values, as individual points have greater influence on the fitted relationship. The tables are nevertheless useful in clarifying the key variables identified as significant for all shellfish beds; sea temperature, as a surrogate for seasonality, and rainfall two days previously. Where flow and mean flow over the past two days (Flow_-1/-2) are present with positive and negative signs respectively, this indicates the rate of increase in flow, which may be rather different from the absolute value of flow and could be interpreted as a relatively quick response to rainfall. Although the model fits, judged by r-squared values, are uniformly better for pour plate counts than for MPN, the fitted models include broadly similar parameters.

Table 6.6.4. Model parameter estimates for log_{10} MPN counts by bed and species for mussels and oysters on the shellfish beds in the Camel estuary.

Variable	Sea Temp	Rain_0	Rain_-1	Rain_-2	Rain_-3	Flow	Flow_ -1/-2	R-squared
All sites	0.033	-	0.038	0.062	-	-	-.024	0.36
Ball Hill Mussels	0.077	-	-	0.097	-0.029	-	-	0.49
Ball Hill Oysters	0.058	-	0.048	-	-	-	-	0.31
Gentle Jane Mussels	0.071	-	-	0.078	-	0.079	-0.080	0.55
Gentle Jane Oysters	0.087	-	-	0.093	-	0.065	-0.090	0.65
Longlands Oysters	0.083	-	-	0.078	-	0.180	-0.173	0.39
Porthilly Mussels	-	-	0.055	0.062	-	-	-	0.29
Porthilly Oysters	-	-	-	0.061	-	-	-	0.23

Table 6.6.5. Model parameter estimates for log_{10} pour plate counts by bed and species for mussels and oysters on the shellfish beds in the Camel estuary.

Variable	Sea Temp	Rain_0	Rain_-1	Rain_-2	Rain_-3	Flow	Flow_ -1/-2	R-squared
All sites	0.056	-	0.033	0.046	-	0.017	-	0.50
Ball Hill Mussels	0.048	-	0.059	0.042	-	-0.028	-	0.73
Ball Hill Oysters	0.105	-	-	0.054	-0.021	-	0.067	0.56
Gentle Jane Mussels	-	-	-	0.079	-	0.080	-0.070	0.63
Gentle Jane Oysters	0.101	0.037	-	0.034		-	0.031	0.74

Longlands Oysters	0.083	0.022	0.035	0.032	0.010	-	-	0.81
Porthilly Mussels	-	-	-	0.075	-0.016	-	-	0.46
Porthilly Oysters	0.089	-	-	0.062	-0.039	-	0.083	0.69

6.6.4 Decision support

A key question for decision making is, given a model-simulated $\log_{10}$ *E. coli* count, what is the probability that the "true" concentration is above a chosen threshold? By way of example, for simulations in the hundreds a low probability that the "true" count will be above a threshold of 4,600 is expected if the model has some useful predictive power. Equally, with a simulation of 10,000 cfu/100g a high probability that the true count was above 4,600 is expected. To formalise this, it is necessary to estimate the probability distribution of true concentrations about the simulated concentration. This probability distribution is computed assuming a normal distribution of the "true" $\log_{10}$ count with mean value the $\log_{10}$ simulated count and standard deviation the residual standard error from the model fit. It is then straightforward to compute the probability that a value from this normal distribution will exceed any chosen value.

Figures 6.6.9 and 6.6.10 show the probability of "true" exceedance of three thresholds given a range of model-simulated $\log_{10}$ counts. The figures provide estimates of these probabilities for modelled values of the MPN and pour plate counts from Jan 1 2020. For demonstration purposes, the full models are used in each case, not the models reduced by stepwise elimination. Points are plotted as red if the measured count corresponding to the plotted simulated count is greater than the chosen threshold, and green if it is less than the chosen threshold. For an ideal model, all points to the right of the threshold line would be red, and all to the left would be green.

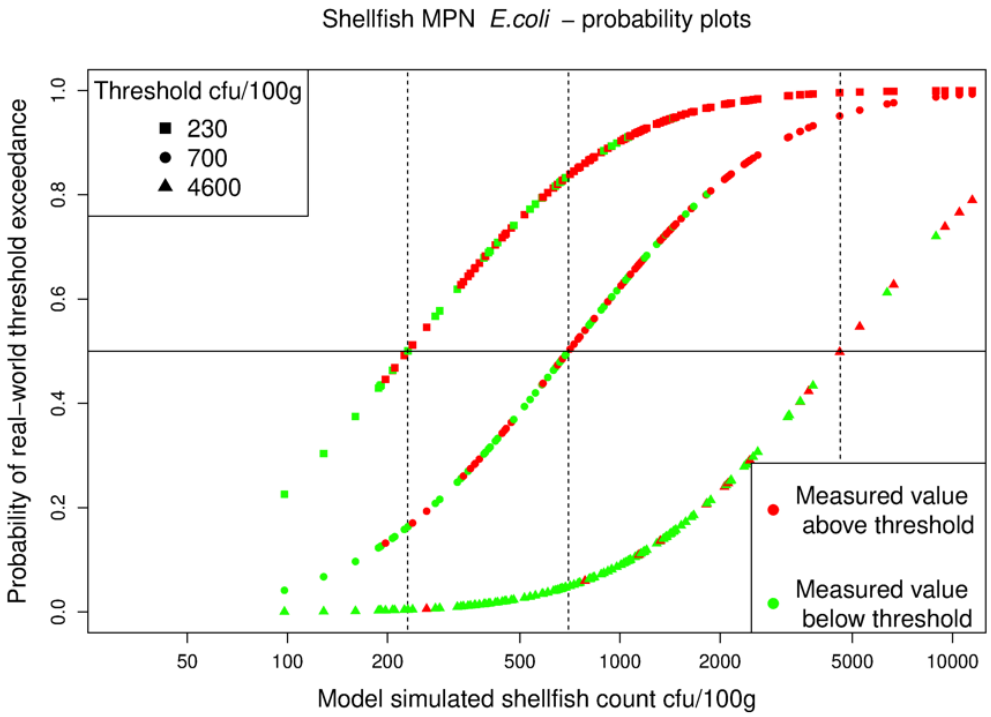


Figure 6.6.9. Shellfish MPN *E. coli* probability plots, showing the probability of simulated values correctly predicting actual observed values. Dotted vertical lines indicate *E. coli* classification thresholds of 230, 700 and 4,600 *E. coli* per 100g shellfish flesh.

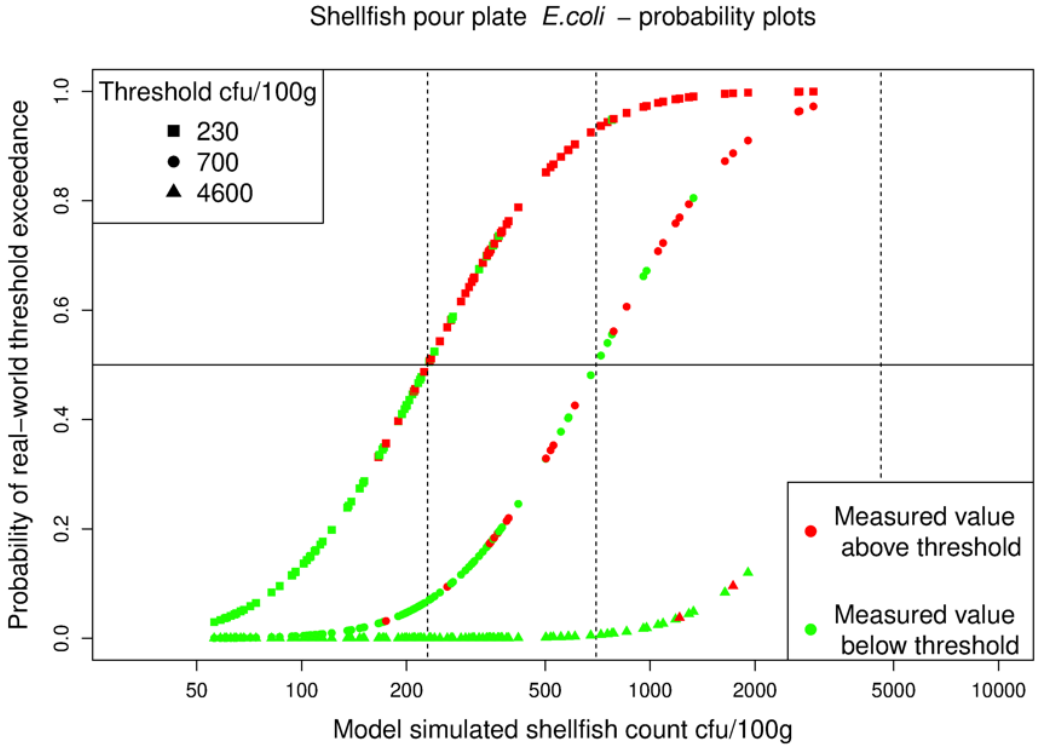


Figure 6.6.10. Shellfish pour plate *E. coli* probability plots, showing the probability of simulated values correctly predicting actual observed values. Dotted vertical lines indicate *E. coli* classification thresholds of 230, 700 and 4600 *E. coli* per 100g shellfish flesh.

Plots such as this can be used to derive error estimates for particular decision rules. If the real world pour plate is greater than 230 if the simulated value is greater than 230, the probability that the decision is correct using the present example for pour plate counts is 0.78. In terms of Figures 6.6.9 and 2.3.10 this is the proportion of red points to the right of 230 along the curve representing the behaviour of the data with respect to the 230 threshold.

To generalise this, Table 6.6.6 shows the probabilities of making correct decisions for MPN and pour plate counts for the example data and model. The correct assignment of counts below each threshold is apparently better for pour plate counts, while MPN and pour plate are comparable for correct assignment of values above the threshold. There is no value for pour plate values above 4,600 because no simulation above that value was made.

Table 6.6.6. Probabilities of correct decisions

Variable	**Threshold 230**		**Threshold 700**		**Threshold 4600**	
	Correct <	Correct >	Correct <	Correct >	Correct <	Correct >
MPN	0.71	0.78	0.69	0.71	0.91	0.78
Pour plate	0.9	0.78	0.88	0.70	0.98	-

6.6.5 Stage 3 Summary

Overall, the supplementary collected data for both MPN and pour plate performed better than the statutory data collected in Stage 1 of the report and the existing statutory data supplemented with field MPN data from the Camel collected by the project for Stage 2. This was attributable to the higher frequency of the data and to an extent the fortuitous timing of rainfall events in relation to sampling. However, for the purpose of developing a predictive model of shellfish *E. coli* contents the data record from 2019 to 2021 is very limited, taking account of the natural variability in the response and the opportunistic nature of the explanatory variables collected. Even with data collection as frequently as every two weeks, sampling is not focussed on periods immediately after likely adverse events, such as heavy rain. Intensive sampling following such events would be likely to improve modelling, but comes with significant logistical challenges in the absence of automated sampling.

As described in Stage 1 and 2 , *E. coli,* coliform and virus counts for the field collected data are highly variable in space and time. This variability has not been fully captured in the sampling program, and it should be noted that the sampling period included COVID restrictions during which human population behaviour was not representative (e.g. altered tourism patterns during the sampling period). However, the use of the field collected MPN *E. coli* data improved the explanatory power of the models compared to the historical official data used in Stage 1, with more consistent results between beds.

The model produced gives an indication of the limits of regression modelling and has limited predictive power but does indicate sources of variability within the data. The Max R-squared (Table

6.6.3) indicates the maximum variability which could be accounted for by a statistical model using explanatory variables selected. Explanatory variables included in the model were sea temperature, radar rainfall two days previously and flow (including where flow and mean flow over the past two days together may indicate increase in flow rate – i.e. a response to rainfall). Rainfall was included in the model rather than CSO operation as rainfall has more reliable data and allows for the inclusion of other potential *E. coli* sources. The model was fitted for the individual shellfish beds and species using $\log_{10}$ *E. coli* and analysed separately for MPN and for pour plate. Although explanatory power of the models was higher for both MPN and pour plate compared to those produced in Stages 1 and 2, the model for MPN was still weak for some beds. The explanatory power of the model using the pour plate date was better than for the MPN *E. coli*/100g shellfish flesh data. The time series of the paired MPN and pour plate *E. coli* data study series of data is too short to expect to generate a reliable predictive model for management, though the high frequency of sampling has provided a data set of 22 time points, across 7 data sets for species and shellfish beds. The results are promising, particularly where pour plate data were applied, and the models could be further refined by extended sampling.

The performance of risk prediction using probability plots and generalised probabilities of correct decisions indicated that the correct assignment of counts to shellfish classification threshold was greatly improved in this Stage 3 modelling. This is encouraging and indicates the feasibility of this approach, especially with the potential for model refinement with longer data sets.

6.7 General discussion Stages 1 - 3

6.7.1 Model development

The aim was to fit the best, relatively simple, statistical model to historical MPN and project MPN and pour plate *E. coli* counts. The best model developed for the Camel is based on environmental data (rainfall radar, river flow, temperature/season) that is readily available. The MPN and pour plate *E. coli*/100 g shellfish flesh data that was collected by the DASSHH project gave better results than the historical MPN data. For the selected pour plate-based model, explanatory power of environmental variables and *E. coli* in shellfish was in some cases greatly improved over previous studies (e.g. Campos *et al,* 2011), when considered at the level of individual beds (R^2 ranging from 0.46 – 0.81). However, this was based on relatively short data sets and further modelling over longer time series is required to confirm these findings and potentially improve the models.

The Camel study was unable to develop satisfactory predictive models based solely on historical MPN *E. coli* results from the Official Control sampling. These data were found to be highly variable and loosely related to explanatory variables considered. Hence the explanatory power of the environmental data were often limited, and strongly influenced by small numbers of extreme values. Small numbers of extremely high and anomalous MPN values are difficult to characterise statistically and some are not associated with preceding rainfall or any other explanatory variable. One potential reason for the differences in model performance between Official Control MPN data and those collected for the DASSHH project is that the latter were collected more frequently (two-weekly vs monthly) and more systematically on the same day every two weeks whatever the

weather, whereas statutory sampling usually occurs once a month and sampling date may vary depending on the weather, potentially introducing bias. The pooling of shellfish sampled from three points across each bed may also have reduced variability, compared to Official Control samples that are collected at a single monitoring point. The improved performance of models based on pour plate data is unsurprising given the lower inherent variability in this method (see Section 5).

Several studies have used regression analysis to relate water quality to explanatory variables using multivariate linear regression models (review by de Brauwere *et al,* 2014). That review concluded that explanatory variables often need to be site specific, that model fit and parsimony should be considered in the selection of explanatory variables, and that data should be collected over several years (de Brauwere *et al,* 2014).

The explanatory variables included in published regression based models vary across studies (de Brauwere *et al,* 2014; Zimmer-Faust *et al,* 2018; Schmidt *et al,* 2018). Models have included factors such as catchment area, diffuse and point sources of pollution and the number of sewage treatment works (STWs) and combined sewer overflows (CSOs) (Malham *et al,* 2014; Crowther *et al,* 2002) as well as physicochemical factors including suspended particulate matter, nutrients, rainfall, tidal movements, seasonal variations, temperature, UV, salinity (Hassard *et al,* 2017a), catchment topography and soil characteristics including soil moisture at the time of rainfall (Campos *et al,* 2013). Work on Pacific oysters in the Dart Estuary demonstrated significant relationships between *E. coli* and lagged rainfall also using historical MPN data (Campos *et al,* 2011). The best model predictors in our study were 7 day cumulative rainfall, river flow and sea surface temperature. The model fit with sea temperature suggests a mechanistic link to *E. coli* persistence in the environment, as well as possible seasonal controls on sources, while rainfall and flow variables capture the conditions under which flushing of *E. coli* occurs into water courses, from both catchment and point sources (see Section 3 of this report). Cumulative rainfall has been included in other regression based models related to shellfish hygiene (e.g. Campos *et al,* 2011; Campos *et al,* 2013; Schmidt *et al,* 2018) reinforcing the importance of antecedent conditions as a component of risk. The inclusion of radar rainfall in the models gave a better integrator of catchment conditions compared to the meteorological data suggesting that point-based meteorological data may not adequately represent the influence of local storms or varying intensity rainfall across the catchment.

The role of CSO spills in contributing to *E. coli* levels in shellfish was not clearly demonstrated in development of predictive models for the Camel estuary. However, this does not mean that human sewage sources are not significant contributors to *E. coli* levels. The data for CSO operation that were available for this modelling exercise were limited in two ways. First, the “on-off” nature of the data meant that only timing and duration of discharges could be included in both the hydrodynamic and statistical models, without any measure of volume or concentration. Secondly, at least for some of the wastewater source locations, the operation time series data were apparently incomplete. As CSO operation is largely influenced by weather conditions it can also be difficult to disentangle from rainfall as a driver of other catchment sources. Elsewhere, statistical models have identified CSOs as potentially important predictors of *E. coli in shellfish* (Conwy Estuary, North Wales, Malham *et al,* 2017). Wastewater treatment works can potentially contribute a significant proportion of total bacterial load into the aquatic environment (Wither *et al,* 2005; Campos *et al,* 2013) with combined

sewer overflows releasing significant volumes of highly contaminated water and affecting shellfish waters (Campos *et al,* 2013; Kay *et al,* 2008).

Relationships for *E. coli* in water at specific locations over each shellfish bed were not investigated in this report. The overall predictive relationship for *E. coli* in water was stronger than the overall result for *E. coli* in shellfish flesh (across all data for all beds), with a relatively good fit for water *E. coli* (R^2 of 0.56). This value is slightly lower than those achieved for other published models to predict shellfish water quality (R^2 0.61 (Gonzalez *et al,* 2012), but demonstrates potential for development in shellfish catchment models.Such approaches are already used by some Third Countries exporting bivalves to the EU and are considered equivalent to the European regulatory approach (Seafish 2021). Gonzalez *et al,* (2012) produced a model where rainfall, dissolved oxygen and salinity were significant. In the present study the exact time of water sample collection was not known so tidal state and associated variables such as salinity could not be used.

Predictive relationships when using other response variables such as enteric viruses (noroviruses and adenoviruses) were weak. The poor model fit with viruses could be due to the sporadic nature of their occurrence (Farkas *et al,* 2018), and their longer persistence (whether infective or not infective) in shellfish (Hassard *et al,* 2017a). Therefore, different approaches may need to be developed for viruses. In addition, there were poor correlations between norovirus and *E. coli*. However, there was a high positive correlation was observed between human mastadenoviruses F and animal atadenoviruses. Presence of the two adenoviruses suggests that both human and animal waste may be contributing to contamination in shellfish at the same time (Wolf *et al,* 2010). Monitoring for potential pathogenic viruses and bacteria at wastewater treatment plants, as currently occurs for SARS-CoV-2 (Hillary *et al,* 2021) may provide an early warning system for occurrence of these organisms in shellfish. These findings add to the information on shellfish contamination and further investigation would aid in understanding the potential use for source apportionment (see Section 4 of this report).

All analyses showed significant differences between shellfish beds and species, and other studies have also shown that *E. coli* concentrations in shellfish are highly variable temporally and spatially both within and across estuaries (Hassard *et al,* 2017b). For the Camel, bed-specific and species-specific models are more appropriate than a single model, with some very strong predictive relationships demonstrated for individual beds.

6.7.2 Application in a predictive tool

The literature on statistical models for water quality is more comprehensive than that for shellfish flesh, because it is a simpler problem to solve. Models for water quality have been developed to predict spatial and temporal pollution in estuarine waters (Zimmer- Faust *et al,* 2018). Those authors tested 5 models, multiple linear regression, Tobit regression, Firth's binary logistic regression (BLR), Classification trees and mixed effects regression, with Classification tree and Firths BLR approach showing the most promise. In early work in Stage 1, regression tree approaches were trialled, but proved unsuitable. A number of countries monitor shellfish waters rather than shellfish flesh (e.g. USA, Canada and New Zealand) to classify shellfish harvesting areas (Seafish 2021) . However, the use of predictive models for shellfish waters as a proxy for *E. coli* in shellfish flesh, is

still relatively limited (Bougeard *et al,* 2011; Campos *et al,* 2011; Campos *et al,* 2013; Schmidt *et al,* 2018).

Differences in *E. coli* concentrations in water and in shellfish flesh are due to a number of factors. Survival rates and persistence of bacterial in the coastal zone is species and strain dependent (Campos *et al,* 2011; Hassard *et al,* 2017b) and can differ between point and diffuse sources (Perkins *et al,* 2016). Bacterial survival is influenced by temperature, pH, turbidity, sunlight/UV and salinity (Campos *et al,* 2013). However, the interactions between these factors on rates of accumulation and depuration of bacteria in shellfish are poorly understood and the effects are difficult to capture in models. This may partly contribute to the unexplained variation which currently limits the accuracy of the approach presented here.

In the UK, models to predict whether *E. coli* concentrations in shellfish flesh would exceed thresholds have been trialled in Cornwall in two bays Schmidt *et al,* (2018) used generalised linear models (GLMs) in mainly A classified shellfish areas and suggested a prediction accuracy of 99-100% that concentrations would not exceed Class B thresholds, based on sample numbers of 107 and 13 respectively in the two bays (Schmidt *et al,* 2018). That study used explanatory variables of lagged rainfall, river flow, sea surface temperature and, for one bay, solar radiation, applied to historical MPN data. However, the stated prediction accuracies for exceedance of Class B thresholds are misleading since the models were derived and calibrated on data that was almost entirely within Class A.

Creating a workable model needs to balance two factors. First, risk to the public, where it is desirable that the model correctly predicts threshold exceedances in order to minimise risks to human health. In principle this can be achieved even with a poorly fitting model by taking a conservative approach to the selection of probability of exceedance. By setting a lower value it is possible to implement a model which correctly predicts all exceedances, but comes at the expense of a very high number of false negatives, i.e. it predicts exceedance when in reality the concentrations may be below the threshold. Such a conservative approach will, with a high proportion of false negatives can have a real and direct economic impact on shellfish producers. For this reason, a workable model needs to balance both false positives and false negatives and requires a higher prediction accuracy than bathing water predictive models. In the Camel study, the most reliable models correctly assigning predicted *E. coli* levels in shellfish to below Class A classification thresholds (<230 and <700 *E. coli*/100g) with 90% and 88% reliability. This rose to 98% reliability for the C class boundary (<4,600 *E. coli*/100g). These results suggest that there is potential to develop a model-driven management system, but with sufficient accuracy demonstrated only where *E. coli* data supplementary to the Official Control sampling is applied, especially the use of pour plate *E. coli* data. For each of the Class A, Class B and Class C thresholds, the model based on pour plate data were substantially better at predicting a pass (i.e. below the threshold, avoiding false negatives) than the MPN model, while the prediction accuracies for fails, (I.e. above the threshold), were similar for both models. However, these results are based on relatively short data sets and further modelling over longer time series is required to confirm these findings and potentially improve the models.

6.8 References

de Brauwere A, Ouattara NK, Servais P (2014) Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review Critical Reviews Environmental Science and Technology 44: 2380-2453.

Bougeard M, Le Saux J-C, Perenne N, Baffaut C, Robin M, Pommepuy M (2011) Modeling of *Escherichia coli* fluxes on a catchment and the impact on coastal water and shellfish quality Journal of the American Water Resources Association 47: 350-366.

Campos CJA, Kershaw S, Lee RJ, Morgan OC (2011) Rainfall and river flows are predictors for β-glucuronidase positive *Escherichia coli* accumulation in mussels and Pacific oysters from the Dart Estuary (England). Journal of Water and Health 09.2: 368-381.

Campos CJA, Kershaw SR, Lee RJ (2013) Environmental Influences on Faecal Indicator Organisms in Coastal Waters and Their Accumulation in Bivalve Shellfish. Estuaries and Coasts 36: 834-853.

Crowther J, Kay D, Wyer MD (2002) Faecal-indicator concentrations in waters draining lowland pastoral catchment in the UK: relationships with land use and farming practices Water Research 36: 1725-1734.

Farkas K, Cooper DM, McDonald JE, Malham SK, de Rougemont A, Jones DJ (2018) Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters Science of the Total Environment 634: 1174-1183.

Gonzalez RA, Conn KE, Crosswell JR, Noble RT (2012) Application of empirical predictive modelling using conventional and alternative fecal indicator bacteria in eastern Norther Carolina waters Water Research 46: 5871-5882.

Hassard F, Sharp JH, Taft H, LeVay L, Harris JP, McDonald, JE, Tuson K, Wilson J, Jones, DJ, Malham SK (2017) Critical Review on the Public Health Impact of Norovirus Contamination in Shellfish and the Environment: A UK Perspective Food and Environmental Virology 9: 123-141.

Hassard F, Andrews A, Jones DL, Parsons L, Jones V, Cox B, Daldorph P, Brett H, McDonald JE, Malham SK (2017b) Physicochemical Factors influence the Abundance and Culturability of Human Enteric Pathogens and Faecal Indicator Organisms in Estuarine Water and Sediment Frontiers in Microbiology 8: 1996.

Hillary LS, Farkas K, Maher KH, Lucaci A, Thorpe J, Distaso MA, Gaze WH, Paterson S, Burke T, Connor TR, Malham, SK, McDonald, J, Jones, D (2021) Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK Water research 200: 117214.

Kay D, Crowther J, Stapleton CM, Wyer MD, Fewtrell L, Edwards A, Francis CA, McDonald AT, Watkins J, and Wilkinson J (2008) Faecal indicator organism concentrations in sewage and treated effluents Water Research 42, 442–454.

Malham SK, Rajko-Nenow P, Howlett E, Tuson KE, Perkins TL, Pallett DW, Wang H, Jago CF, Jones DL, McDonald JE. 2014. The interaction of human microbial pathogens, particulate material and

nutrients in estuarine environments and their impacts on recreational and shellfish waters Environmental Science. Processes & Impacts. 16: 2145-2155.

Malham SK Taft H, Cooper D, Ladd C, Seymour S, Robins PE, Jones DL, McDonald JE, Le Vay L, Jones L. 2017. Review of current evidence to inform selection of environmental predictors for active management systems in classified shellfish harvesting areas. FSA Project FS103001. Report prepared by Bangor University and the NERC Centre for Ecology and Hydrology.

Perkins TL, Perrow K, Rajko-Nenow P, Jago CF, Malham, SK, Jones DL, McDonald JE (2016) Decay rates of faecal indicator bacteria from sewage and ovine faeces in brackish and freshwater microcosms with contrasting suspended particulate matter concentrations Science of The Total Environment 572: 1645-1652.

Schmidt W, Evers-King HL, Campos CJA, Jones DB, Miller PI, Davidson K, Shutler JD (2018) A generic approach for the development of short-term predictions of Escherichia coli and biotoxin in shellfish Aquaculture Environment Interactions 10: 173-185.

Seafish (2021) Review of the application of the Official Control Regulations for shellfish production as they relate to microbial contamination. 103 pp. https://www.seafish.org/document/?id=621ffa11-4943-4698-bdd4-bcf3784bdda8.

Wither A, Greaves J, Dunhill I, Wyer M, Stapleton C, Kay D, Humphrey N, Watkins J, Francis C, McDonald A, and Crowther J (2005) Estimation of diffuse and point source microbial pollution in the Ribble catchment discharging to bathing waters in the north west of England. Water Science and Technology 51: 191–198.

Wolf S, Hewitt J, Greening GE (2010) Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. Applied Environmental Microbiology 76: https://doi.org/10.1128/AEM.02249-09.

Zimmer-Faust AG, Brown CA, Manderson A (2018) Statistical models of fecal coliform levels in Pacific Northwest estuaries for improved shellfish harvest area closure decision making Marine Pollution Bulletin 137: 360-369.

7 Potential for variation of duration of post-harvest depuration for mussels and oyster in response to predicted environmental risk levels.

7.1 Summary

Alongside the development of predictive models for *E. coli* in shellfish, an assessment of depuration times for mussels and oysters from a range of initial microbial loadings was conducted to inform recommendations for depuration times under specific conditions. A series of experiments investigated the rate of clearance of *E. coli* from mussels and oysters that either been experimentally inoculated with *E. coil* in the laboratory or that were collected from Class B shellfish beds. Results of depuration experiments confirmed expected patterns, with the time taken to clear *E. coli* down to Class A levels (<230 *E. coli*/100g*)* varying with the initial level of contamination. Where inoculation achieved very high loadings of *E. coli* (>170,000/100g) depuration for up to 72h was not sufficient to reduce counts to <230 *E. coli/100g,* which is consistent with shellfish from areas with a Prohibited classification (>46,000 *E. coli*/100g, as measured by MPN) not being fit for consumption even after extended depuration. Where spiked *E. coli* levels were equivalent to occasional high values above 4,600 *E. coli*/100g that may occasionally be experienced in a B classified production area, shellfish were depurated to class A levels<230 *E. coli*/100g within 48 hours, which is broadly consistent with current industry practice. Where environmentally contaminated shellfish exhibited initially *E. coli* lower contamination (300 – 3,300 *E. coli*/100g) that was fully within the B-classification boundaries, depuration to very low levels could be achieved in 6-12 hours, justifying potential use of short depuration times in low-risk periods. These results, together with information other published studies, demonstrate the potential for adoption of flexibility in depuration times in response to predicted levels of *E. coli* in shellfish.

7.2 Introduction

Bivalve molluscs filter water, removing organic and inorganic particles. These particles can contain a wide variety of contaminants including bacteria and viruses which may be pathogenic to humans and may bioaccumulate in the bivalves. Faecal indicator bacteria are utilised as an indicator of microbial contamination which may present a health risk when bivalves are consumed raw or only lightly cooked (Malham *et al,* 2014)

In order to protect public health regulations in the UK and elsewhere require bivalve shellfish from Class B areas to be depurated, relayed or cooked (European Commission 2017; European Commission 2019). Depuration involves the immersion of bivalves in recirculating seawater in order to remove any potential human pathogens. The efficiency of depuration depends on several variables such as temperature, salinity, the species of bivalves and their initial loadings of the faecal indicator bacteria, *E. coli* (Love *et al*, 2010) Generally, depuration is undertaken for 42 hours.

One of the potential advantages of an assurance scheme enabling adaptive management of shellfish harvesting, is that producers could make informed decisions about harvesting schedules to avoid high risk periods. It could also inform depuration, for example enabling producers to increase

depuration during periods of higher risk. Conversely, in periods of low risk, depuration periods could be shortened without affecting product safety.

Alongside the development of predictive models for *E. coli* in shellfish, an assessment of depuration times for mussels and oysters from a range of initial microbial loadings was conducted to inform recommendations for depuration times under specific conditions.

7.3 Methods

Depuration experiments were undertaken using mussels and oysters that had either been inoculated in the laboratory with *E. coli* and then depurated or were collected from Class B shellfish beds (Conwy and Camel) and transported to the depuration unit at the School of Ocean Sciences in Menai Bridge.

7.3.1 Laboratory inoculated samples

Mussels and oysters were inoculated with *E. coli* K12 (LZB 035), supplied by Blades Biological (Kent, UK), which was cultured overnight in Luria-Bertani Miller's medium (LB) (Miller, 1972). The bacterial concentration was determined using a spectrophotometer OD_{600}. The culture was then serially diluted to the appropriate concentrations, centrifuged and resuspended in ¼ strength Ringers solution (Oxoid Ltd.) and added to 1L of algal feed before inoculation of the mussels or oysters held in a depuration tank (400-litre unit held at 10 ^{o}C with a flow rate of 20 litres min^{-1}, Seafish, 2018). Agar plates were performed to confirm the holding tanks were free of *E. coli* pre inoculation and to assess post-inoculation and post-accumulation *E. coli* concentrations. Samples of the inoculate, the water in the holding tanks pre- and post- inoculation, and water in the tanks post-accumulation were filtered and spread onto Harlequin agar plates and incubated at 37 °C. Two experiments were undertaken, with the first inoculated oysters at a calculated loading of 30,000 *E. coli/*100g and the second mussels at 15,000 *E. coli/*100g. The shellfish were left to accumulate *E. coli* for 24 hours in the depuration tank with the flow through pump on and the UV off, before the first set of samples were taken (Hour 0). Samples were taken in triplicate, with three batches of 12 oysters removed from the tank from various areas within the tank to ensure that they were representative. Following the accumulation period, an hour zero sample was taken for processing before the shellfish were removed and placed in a holding tank. The depuration tank was emptied, scrubbed and filled again with the UV lamps turned on for 1 hour to sterilise the water, before the mussels were replaced. Further samples of mussels were randomly removed at 6, 12, 24 and 48 hours. All mussels were quantified for their *E. coli* loadings by both the approved MPN and the pour plate methods as previously described (see Section 5 of this report).

7.3.2 Environmentally contaminated samples

Environmentally contaminated mussels and oysters were collected from the Conwy and Camel Estuaries and placed in the depuration tanks in Menai Bridge. As for the inoculation experiments, the shellfish were depurated in a 400-litre depuration unit. Shellfish were randomly selected from the depuration unit at various sampling times (see below). For experiments based on mussels

collected from the Conwy estuary, samples were removed from the depuration tank in triplicate, with a minimum of 15 mussels comprising each sample. For experiments based on oysters and mussels collected from the Camel estuary, shellfish could only be processed as single samples for each timeframe.

7.3.2.1 Conwy

Mussels were collected from the Conwy estuary near a combined sewer overflow (CSO). Animals were collected within 24 hours of a high rainfall weather event likely to trigger CSO overflow. Following collection, mussels were separated into those that were placed into the depuration unit or those that were processed immediately for *E. coli* enumeration using the MPN and Pour plate methods. Three replicate samples of 15 mussels per sample were removed from the depuration unit at 6, 24 and 48 hours and immediately shucked and blended. For each replicate, the mussels were processed independently using the MPN and the pour plate methods.

7.3.2.2 Camel

Mussels and oysters were sent from the Camel (Gentle Jane lay) as part of the regular bi-weekly sampling for the DASSHH project. Unlike the Conwy experiments, samples from the Camel were not specifically taken following rainfall events, unless this had opportunistically occurred before sampling. Upon receipt of the shellfish, some were processed immediately for *E. coli* enumeration with extra samples placed in the depuration unit and subsequently sampled at 6, 12 and 24 hours for mussels and oysters.

This study was restricted compared to initial scope, mainly due to the work being affected by the COVID -19 pandemic, with restrictions on hours of work as well as social distancing regulations reducing the number of people allowed in laboratories.

7.4 Results

7.4.1 Laboratory inoculated samples

Oysters inoculated at a target concentration of 30,000 *E. coli*/100g contained initial mean *E. coli* concentrations of 173,333 and 17,273 *E. coli*/100g, measured by MPN and pour plate methods, respectively. These levels had reduced to below the level of 4600 *E. coli*/100g shellfish flesh within 48 hours (Figure 7.1). However, reduction to <230 *E. coli*/100g was not achieved even after 72h depuration. A Wilcoxon rank test demonstrated no significant difference overall between MPN and Pour plate values although the initial starting values are significantly different (paired t-test, p = 0.0014).

Mussels inoculated at 15,000 *E. coli*/100g of shellfish contained initial mean *E. coli* concentrations of 7933 and 5047 *E. coli*/100g, measured by MPN and pour plate methods, respectively. These levels, measured by both methods, had reduced to below 230 *E. coli*/100g shellfish flesh within 48

hours (Figure 7.2). The starting values as measured by MPN and Pour plate methods were not significantly different (paired t-test, p = 0.1819).

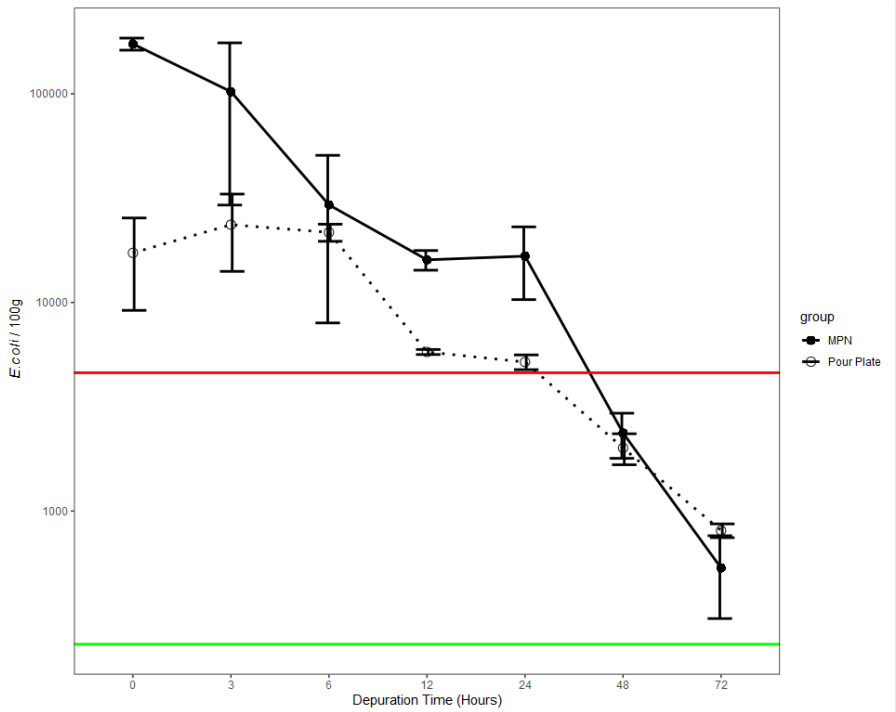


Figure 7.1 *E. coli* concentrations in oysters measured by MPN and pour plate methods following initial inoculation at a target level of 30,000 *E. coli/*100g, sampled at 0, 6, 12,24 and 48h. Values are means ± standard deviation, n= 3 for each point. The green and red horizontal line indicates the classification thresholds at 230 and 4,600 *E. coli*/100g, respectively.

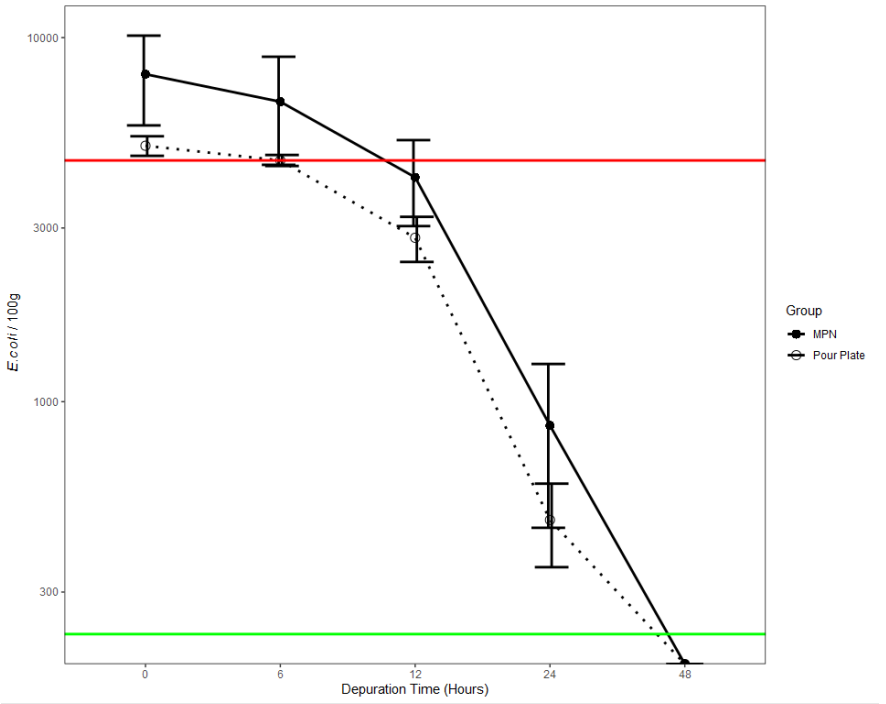


Figure 7.2 *E. coli* concentrations in mussels measured by MPN and pour plate methods following initial inoculation at a target level of 15,000 *E. coli/*100g, sampled at 0, 6, 12,24 and 48h. Values are means ± standard deviation, n= 3 for each point. The green and red horizontal line indicates the classification thresholds at 230 and 4,600 *E. coli*/100g, respectively.

7.4.2 Environmental samples

7.4.2.1 Conwy

The mussels collected from Conwy were within the classification boundaries 230 – 4600 *E. coli*/100g before depuration, apart from one sample (measured by MPN) at 0 hours which was above > 4600 *E. coli*/100g threshold. Mean starting values for Experiment 1 were 3300 and 2853 *E. coli*/100g, measured by MPN and pour plate methods, respectively. Mean starting values for Experiment 2 were 2900 and 1060 *E. coli*/100g, measured by MPN and pour plate methods, respectively. One sample, measured by MPN, was recorded above the *E. coli*/100g boundary after 6 hours. Analysis from both MPN and Pour plate data showed that in both experiments, *E. coli* concentrations had dropped to below < 230 *E. coli* /100g after 12 hours of depuration, (Figure 7.3). Overall, there was no significant difference between the MPN and Pour plate data for this experiment.

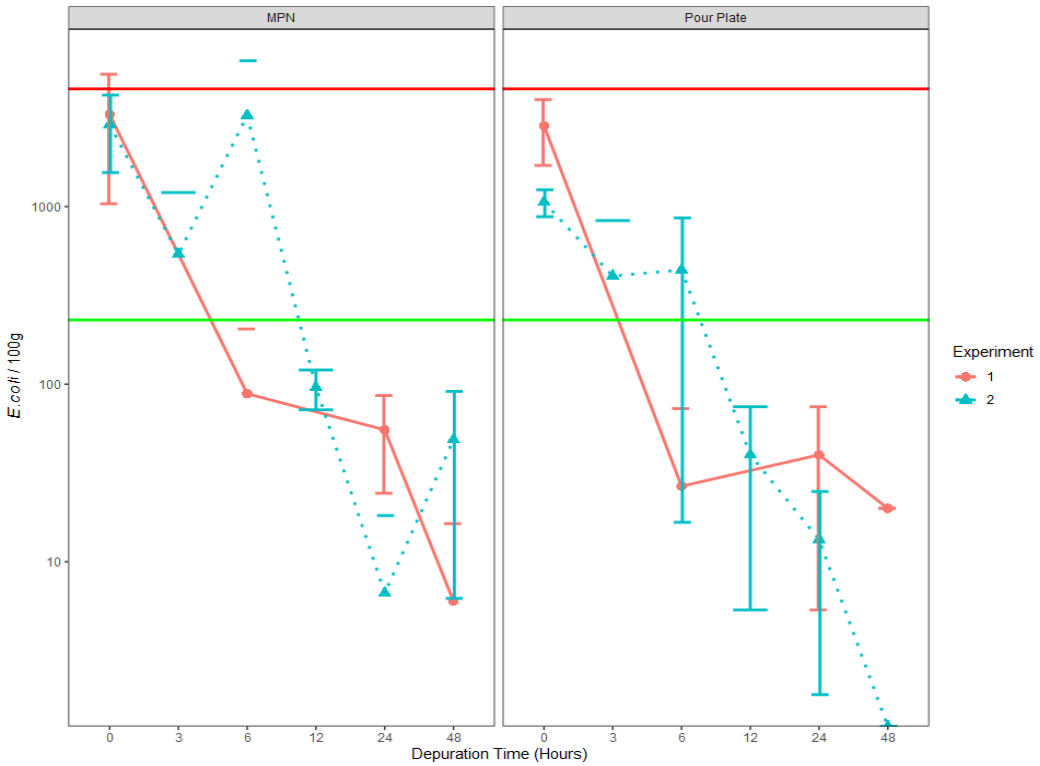


Figure 7.3. *E. coli* concentrations in environmentally contaminated mussels collected from the Conwy estuary, measured by MPN and pour plate methods, in two experiments following depuration over 48h. Values are means ± standard deviation, n= 3 for each point. The green and red horizontal line indicates the classification thresholds at 230 and thresholds at 230 and 4,600 *E. coli*/100g, respectively.

7.4.2.2 Camel

The mussels and oysters collected from Camel were mostly between 230 – 4600 *E. coli*/100g, as measured by the MPN method, with one sample of mussels starting at <230 *E. coli*/100g. Using the pour plate method, the *E. coli* concentrations were lower in all cases and were initially very low (not

detectable) in the first experiment in both oysters and mussels (Experiment A, Figure 7.4). In the second experiment, mussels were measured to be within the 230 – 4600 *E. coli*/100g classification boundaries before depuration, while the oysters started at <230 *E. coli*/100g. In both experiments, for both species, the MPN method showed *E. coli* levels were reduced to <230 *E. coli*/100g after 12h depuration, with this level reached after 6h of depuration in three experiments. Results were less clear using the pour plate method, wherein the mussels assessed in experiment B returned an *E. coli* concentration of > 230 *E. coli*/100g after 12h, despite an initial rapid drop to below this threshold after 6h which was also observed in the oysters during the same experiment (Figure 7.4).

Table 7.1 gives a summary of the depuration results showing initial *E. coli* concentration and the time taken for each of those *E. coli* loads to depurate to ≤ 230 *E. coli*/100g.

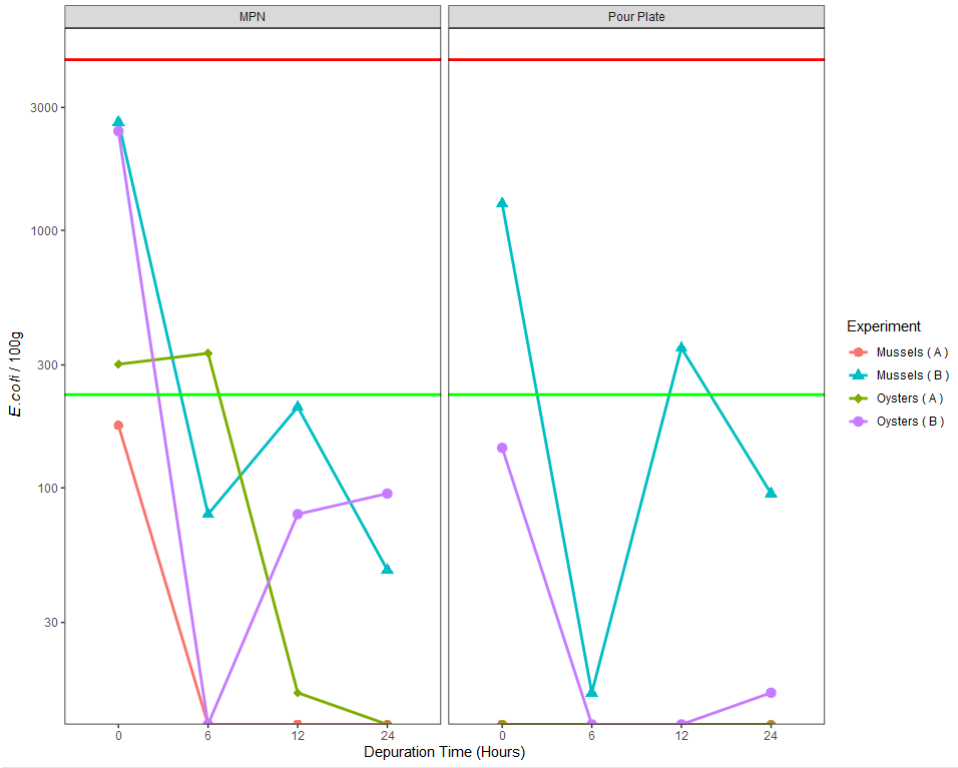


Figure 7.4 *E. coli* concentrations in environmentally contaminated oysters and mussels collected from the Camel estuary, measured by MPN and pour plate methods, in two experiments following depuration over 48h. Values are from single samples of 12 oysters or 15 mussels. The green and red horizontal line indicates the classification thresholds at 230 and 4,600 *E. coli*/100g, respectively.

Table 7.1 Summary of depuration results showing initial *E. coli* concentrations, the time until *E. coli* concentrations ≤ 230 *E. coli*/100g were confirmed and the values at the relevant time point. Values for *E. coli* are mean counts/100g for 3 replicates, for both the MPN and pour plate methods. nd = not detectable, no colonies observed. n/a = not applicable

	Initial *E. coli* / 100g		Hours to achieve <230 *E. coli*/100g		*E. coli*/100g concentration first confirmed <230/100g	
Experiment	MPN	Pour Plate	MPN	Pour Plate	MPN	Pour Plate
Lab Spike 1 - Oysters	173,333	17273	>72	>72	n/a	n/a
Lab Spike 2 - Mussels	7,933	5046	48	48	<18	nd
Conwy 1- Mussels	3,300	2853	6	6	88	27
Conwy 2- Mussels	2,900	1060	12	12	96	40
Camel Mussels A	175	Nd	6	0	nd	nd
Camel Mussels B	2,635	1270	6	24	79	100
Camel Oysters A	302	Nd	12	0	<18	nd
Camel Oysters B	2,429	143	6	nd	95	16

7.5 Discussion

Generally, the results show that depuration from high *E. coli* levels of 173,000 (MPN) or 7,900 (pour plate) *E. coli*/100g, from spiking undertaken in laboratory setting had not reached <230 *E. coli*/100g within 72 hours. These laboratory results are consistent with shellfish from areas with a Prohibited classification (>46,000 *E. coli*/100g, as measured by MPN) not being fit for consumption even after extended depuration (Table 7.1). Where spiked *E. coli* levels were lower but still above the 4,600 *E. E. coli/100g* classification threshold (*circ*a 8,000 and 5000 *E. coli*/100g, for MPN and pour plate methods, respectively), these are equivalent to occasional high values that may be experienced in a B classified production area, and depuration to <230 *E. coli*/100g in 48 hours was broadly consistent with current industry practice. Where environmentally contaminated shellfish exhibited initially *E. coli* lower contamination (300 - 3,300 *E. coli*/100g) that was fully within 4,600 - 46,000 *E. coli*/100g boundaries, depuration to very low levels could be achieved in 6-12 hours, justifying potential use of short depuration times in low-risk periods.

Higher levels of *E. coli* take longer to depurate compared to bivalves with lower levels of contamination (Oliveira *et al,* 2011). Additionally, the physiological state of the bivalves, as well as a range of factors including the size and age of the bivalves, alteration in environmental parameters and stress will also affect depuration (Richards, 1988). Consistent with the previous work (see section 5) differences in results obtained using the different methods for determining the concentration of *E. coli* per 100 grams of shellfish flesh were detected.

Several laboratory-based studies where bivalves have been inoculated with *E. coli* suggest that artificially contaminated bivalves may depurate more rapidly than environmentally contaminated

bivalves, possibly due to the use of pure cultures (Jones *et al,* 1991; Oliveira *et al,* 2011). Sharp *et al* (2021) demonstrated the removal of around 90% of spiked non-pathogenic *E. coli* from mussels within 42 hours from very high initial levels ($5x10^6$ cfu/100ml). Laboratory inoculations have also demonstrated that than bivalves with higher *E. coli* concentrations will depurate faster than those with lower *E. coli* contaminations, with the initial concentration of bacteria a significant factor in determining effective depuration rates (McGhee *et al,* 2008). Trials with the mussel *Mytilus galloprovincialis* from Class B and C indicated *E. coli* values of <230/100 g using MPN within 48h, which is broadly in line with the present study. However, mussels depurated from Class B reached Class A within 24h, compared to the much more rapid depuration seen in the present study (6-12h). It is possible that this difference in depuration rates observed in *M. galloprovincialis* and *M. edulis* in this study is related to the difference in species.

Relating these observations to the range of *E. coli* concentrations observed in the shellfish in the Camel study, a wide range of values were observed from August 2020 – August 2021. These ranged from <230 *E. coli*/100g up to occasional high *E. coli* concentrations of 35,000/100g, as measured by MPN (when pour plate values reached 3460 *E. coli*/100g for the same sample). The majority of *E.coli* MPN results for the Camel were within the shellfish classification thresholds 4,600 - 46,000 *E. coli*/100g (n=103), with pour plate values showing similar numbers both <230 and between 4,600 - 46,000 *E. coli*/100g (N=90 for both).

A caveat here is the limited range of experiments conducted for this section of the DASSHH report, due to Covid-19 restrictions. Hence, establishment of protocols for depuration times could usefully be informed by further experimental studies. There is a gap in the present data, for *E. coli* contamination levels between 8,000 and 43,000 *E.coli*/100g that could useful be explored. This reflects the fact that high values in this range may occur even if a shellfish area is B classified, as observed for the Camel. Present results demonstrate that during the highest risk periods observed in the Camel estuary, depuration of 48h or even 72 h may not be sufficient. This would be at times when *E. coli* contamination was significantly above the 4,600 *E.coli*/100g boundary. For the majority of the time, when B-class conditions are expected to prevail, depuration could be reduced to 12h, or even potentially eliminated during periods when A-class conditions are predicted. Further studies would be required to verify this reduction.

However, as the Section 6 of this report clearly shows, there was no clear correlation between environmental predictors and viral contamination in shellfish sampled from the Camel. Norovirus was present seasonally and sporadically in a low number of samples but with no overall correlation with *E. coli* numbers. Whilst *E. coli* is used as an indicator of human pathogens such as norovirus, it is cleared relatively rapidly from shellfish during depuration in comparison to the clearance of measurable (though not necessarily infective) norovirus, which takes substantially longer. Hence, any decisions relating to depuration times would also need to be informed by monitoring of norovirus prevalence (Jones *et al,* 1991; Lees 2000; Croci *et al,* 2002; Love *et al,* 2010; Hassard *et al,* 2017; Gyawali *et al,* 2019; Sharp *et al,* 2021).

7.6 References

Croci L, Suffredini E, Cozzi L, Toti L (2002) Effects of depuration of molluscs experimentally contaminated with *Escherichia coli, Vibrio cholerae* o1 and *Vibrio parahaemolyticus*. Journal of Applied Microbiology 92: 460-465.

European Commission (2017). Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant. Official Journal of European Communities L, 95: 1-142.

European Commission (2019). Commission implementing regulation (EU) 2019/627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and amending Commission Regulation (EC) No 2074/2005 as regards official controls. Official Journal of European Communities L, 131: 51-100.

Gyawali P, Fletcher GC, McCoubrey, D-J, Hewitt J (2019) Norovirus in shellfish: An overview of post-harvest treatments and their challenges. Food Control 99: 171-179.

Hassard F, Sharp, JH, Taft, H., Le Vay L, Harris JP, McDonald JE, Tuson K, Jones DL, Malham SK (2017) Critical Review on the Public Health Impact of Norovirus Contamination in Shellfish and the Environment: A UK Perspective. Food and Environmental Virology. 9: 123-141.

Jones SH, Howell TL, O'Neill KR (1991) Differential elimination of indicator bacteria and pathogenic *Vibrio* spp. from eastern oysters (*Crassostrea virginica* Gmelin, 1971) in a commercial controlled purification facility in Maine. Journal of Shellfish Research 10: 105-112.

Lees D (2000) Viruses and bivalve shellfish International Journal of Food Microbiology 59: 81-116.

Love DC, Lovelace GL, Sobsey MD (2010) Removal of *Escherichia coli, Enterococcus fecalis*, coliphage MS2, polivirus, and hepatitis A virus from oysters (Crassostrea virginica) and hardshell clams (*Mercenaria mercenaria*) by depuration. International Journal of Food Microbiology 143: 211-217.

Malham SK, Rajko-Nenow P, Howlett E, Tuson KE, Perkins TL, Pallett DW, Wang H, Jago CF, Jones DL, McDonald JE. 2014. The interaction of human microbial pathogens, particulate material and nutrients in estuarine environments and their impacts on recreational and shellfish waters. Environmental Science. Processes & Impacts. 16: 2145-2155.

McGhee TJ, Morris JA, Noble RT, Fowler PK (2008) Comparative Microbial Dynamics in *Crassostrea virginica* (Gmelin 1791) and *Crassostrea ariakensis* (Fujita 1913). Journal of Shellfish Research 27: 559-565.

Oliveira J, Cunha A, Castilho R, Romalde JL, Pereira MJ (2011) Microbial contamination and purification of bivalve shellfish: Crucial aspects in monitoring and future perspectives – A mini-review. Food Control 22: 805-816.

Richards GP (1988) Microbial purification of shellfish: a review of depuration and relaying. Journal of Food Protection 51: 218-251.

Miller J (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, pp.328-330.

Seafish (2018) Standard Design Purification Systems. Operating Manual for the Small Scale Shallow Tank Purification System. Seafish Report No: 721. ISBN No.: 978-1-911073-27-7.

Sharp JH, Clements K, Diggens M, McDonald JE, Malham SK, Jones DL (2021) *E. coli* is a poor end-product criterion for assessing the general microbial risk posed from consuming norovirus contaminated shellfish Frontiers in Microbiology 12: 150.

8 Summary of results

- An overarching finding from the Camel study is that a real-time predictive system for *E. coli* levels in shellfish is conceptually feasible.

- The relatively simple model developed for the Camel is based on environmental data (rainfall radar, river flow, temperature/season) that was readily available.

- The study demonstrated the ability to predict *E. coli* levels in shellfish, with the most reliable models, based on pour plate data, correctly assigning predicted *E. coli* levels in shellfish to below Class A classification thresholds (<230 and <700 *E. coli*/100g) with 90% and 88% reliability. This rose to 98% reliability for the C class boundary (<4,600 *E. coli*/100g). For the selected pour plate-based model, explanatory power of environmental variables and *E. coli* in shellfish was in some cases improved over previous studies when considered at the level of individual beds. These findings suggest that, for the Camel, bed-specific and species-specific models may be more appropriate than a single whole-site model, with some strong predictive relationships demonstrated for individual beds. However, it is acknowledged that the predictive modelling is based on relatively small data sets over a 12-month period and these is scope for improvement of the models, as may be required for application in an assurance scheme.

- The Camel study was unable to develop satisfactory predictive models based solely on historical MPN *E. coli* results from the Official Control sampling. These data were found to be highly variable and loosely related to explanatory variables considered. Hence the explanatory power of the environmental data were often limited, and strongly influenced by small numbers of extreme values. Small numbers of extremely high MPN values are difficult to characterise statistically and some are not associated with preceding rainfall or any other explanatory variable.

- The pour plate method consistently yielded less variable *E. coli* results than those from MPN, particularly for the higher end of the range of *E. coli* concentrations measured. This variability suggests that the MPN method has greater potential to generate outlier high results that may influence interpretation or application of monitoring data. This is exacerbated by the observation, in the Camel study, of higher results from the MPN method compared to pour plate data.

- More effective predictive models were based on the improved correlation between environmental predictors and additional *E. coli* MPN data gathered during the course of the study and, more so, *E. coli* data as measured by the pour plate method. The improvement in model fits using the MPN data collected by industry specifically for the DASSHH project seems to reflect the higher frequency and greater consistency of the supplementary *E. coli* data collected, though with limitation of relatively small data sets over only on annual cycle. The

pooling of shellfish sampled from three points across each bed may also have reduced variability, compared to Official Control samples that are collected at a single monitoring point. Improvement of the predictive modelling using the pour plate results is less surprising, given the lower intrinsic variability in *E. coli* measurement using this method.

- The hydrodynamic modelling clearly shows the potential for quantifiable CSO and wastewater treatment discharge dispersal over shellfish beds, though with some variation in duration and extent, depending time of year and location of the outfalls relative to net tidal flow in the estuary. Some outfalls close to the shellfish beds but slightly downstream may have less overall influence on shellfish beds than might be expected, relative to discharges further upstream.

- The role of CSO spills in contributing to *E. coli* levels in shellfish was not clearly demonstrated in development of predictive models for the Camel estuary. However, this does not mean that human sewage sources are not significant contributors to *E. coli* levels. The data for CSO operation that were available for this modelling exercise were limited in two ways. First, the "on-off" nature of the data meant that only timing and duration of discharges could be included in both the hydrodynamic and statistical models, without any measure of volume or concentration. Secondly, at least for some of the wastewater source locations, the operation time series data were apparently incomplete. A further confounding effect may be close correlation between rainfall and CSO operation. This probably explains why addition of CSO activation data adding little to the precision of predictive models, when rainfall is already a strong predictor of *E. coli* in shellfish.

- The presence of human source indicator viruses and enteric pathogens, at times at relatively high levels, clearly confirms human sewage contamination of shellfish within the Camel estuary. While the indicator viruses identified in the present study do not allow quantification of relative contribution from human and animal sources, agricultural run-off is clearly also a contributor to *E. coli* in shellfish in the Camel estuary. This contribution may be significant as the Camel catchment has a high proportion of improved grassland, reflecting the scale of livestock farming.

- Selected adenoviruses show good potential as indicators of animal and human sources of microbial contamination i.e. high frequency of occurrence and high abundance in shellfish. There was a very high degree of correlation between human and livestock indicator viruses, suggesting that sources of these (CSO operation and farmland run-off) are responding similarly to catchment-scale environmental drivers. This may confound modelling of risk based on *E. coli*, which does not distinguish human and animal sources.

- Any management measures would also need to take account of the less predictable occurrence of norovirus in shellfish. Norovirus was present seasonally and sporadically in a low proportion of samples but with no clear correlation with *E coli* numbers. There was also no clear correlation between environmental predictors and pathogenic viral contamination of shellfish. This is likely to be due to two key limitations: a) the seasonal and sporadic prevalence of enteric pathogen

viruses in human populations and b) the longer retention of measurable viruses in shellfish flesh. Both of these effects would be expected to limit the potential for predictive modelling of viral contamination in shellfish, in contrast to bacterial indicators such as *E. coli* that are consistently present in faecal sources and which are cleared from shellfish relatively quickly once input from sources of contamination declines.

- Predictive relationships when using other response variables such as enteric viruses (noroviruses and adenoviruses) were weak. The poor model fit with viruses could be due to the sporadic nature of their occurrence and their longer persistence (infective or not infective) in shellfish.

- Results of depuration experiments confirmed expected patterns, with the time taken to clear *E. coli* down to Class A levels (<230 *E. coli*/100g*)* varying with the initial level of contamination. Where *E. coli* levels were initially low (230-4600 *E. coli*/100g) clearance could be achieved in relatively short periods of 6-12h. Where contamination was high (well above 4600 *E. coli*/100g) 72h depuration could be required, and for very high contamination, effective depuration could not be achieved after 72h. This demonstrates the potential for adoption of flexibility in depuration times in response to predicted levels of *E. coli* in shellfish, while taking into account the prevalence and retention times of enteric viruses such as norovirus.

8.1 Conclusions

If a national-scale assurance scheme is to be developed, site-specific findings for the Camel clearly need to be extended to an understanding of general applicability to a range of shellfish sites. It is not expected that the relationships between environmental predictors and *E. coli* in shellfish will be the same across different catchments. However, the case study results provide some useful guidance and lessons on approaches that could be taken in assessing other sites for development of an assurance scheme based on predictive modelling of *E. coli* in shellfish. In summary these are:

- Successful development of relatively simple predictive models based on readily available environmental data suggests that transferring this approach to other catchments is unlikely to require the costly and time-consuming environmental data collection undertaken for the Camel, e.g. high frequency measurement of water turbidity and nutrients. Instead, linking to ongoing environmental monitoring programmes (rain radar, river flow) should be sufficient. Establishment of real-time data links would open up the potential development of a predictive system.

- The need for incorporation of supplementary *E. coli* time series data into predictive models is a challenging outcome. Official *E. coli* data records were too variable to allow development of practical predictive models and transfer to other catchments may a require supplementary collection of two-weekly *E. coli* data, ideally by the pour plate method, for at least 12 months

and possibly longer. The successful outcome of the modelling is not necessarily certain, even with this investment.

- Evaluation of other catchments suitability for the predictive modelling approach could still initially be based on investigation of relationships between "off the shelf" Official Control and environmental data records. Previous modelling across a range of catchments and shellfish production areas in England and Wales has indicated that the strength of relationships between *E. coli* and environmental predictors does vary between sites and catchments. Hence, at least some most likely candidate sites for a first phase of extension of results might be broadly identifiable on this basis.

- However, a key lesson from the DASSHH project is that evaluation of catchments for suitability for predictive modelling may under-estimate the potential for application of predictive models, if based only on the readily available Official Control data. As seen for the Camel, weak and inconsistent relationships between environmental predictors and Official Control *E. coli* results do not necessarily preclude development of effective models to underpin an assurance scheme. The potential for supplementary time series *E. coli* data, ideally the using pour plate method, to improve the reliability of predictive models should also be considered in identifying suitable sites for predictive model development.

- Even though CSO operation was not included in the predictive models developed for the Camel, modelling studies in other catchments and shellfish productions areas should always include CSO operation as a predictive factor, as the location and relative scale of CSO discharges may make them more readily quantifiable source of shellfish contamination. In a fully systematic approach, as applied in the Camel, this might be investigated by hydrodynamic source modelling studies. However, initial predictive model development for other sites might take a less onerous statistical approach to screening the relative contribution of environmental factors and sources (including CSOs), with iterative refinement identifying those factors which best explain *E. coli* levels in shellfish. Future improvement in CSO monitoring can potentially lead to incorporation of real-time activation data into predictive models. However, the value of such data will be greatest if flow rates of CSO spills are also recorded.

- The relative importance of human and animal source pollution may be expected to vary between catchments, and locations within catchments. The application of viral source identification of the human and agricultural/wildlife faecal contamination would be a valuable component of interpretation and application of predictive models developed for other locations, as the *E. coli* indictor does not separate these sources and could result in overestimation of risk (for example where high *E. coli* results reflect increases in agricultural rather than sewage inputs).

- The current system of shellfish area classification was designed to assess the overall level of sewage contamination, rather than to measure specific health risks. In the Camel study, the lack of any significant correlation between norovirus and *E. coli* levels in shellfish highlights the well-

recognised weakness in the use of this generic indicator to quantify specific human health risks. Furthermore, the lack of correlation between environmental predictors and viruses in shellfish, both from human and animal sources, shows that an assurance scheme based entirely on environmental factors will be equally unable to predict human health risk from norovirus reliably. In the future, this limitation may be overcome by development of new approaches to measurement of risk from norovirus. First, laboratory methods that can reliably assess levels of infective norovirus could be applied to routine monitoring of sampled shellfish and results used to inform management/harvesting decisions. Secondly, development of enhanced high-throughput methods for viral extraction from wastewater has enabled the establishment of the national wastewater-based Covid-19 surveillance programmes. This approach to rapid and high frequency screening of wastewater has the potential to be extended to other viruses of public health concern, including norovirus. If implemented in key wastewater treatment facilities within a catchment, this has the potential to provide close to real time information on norovirus prevalence that could feed into risk models underpinning adaptive management of shellfish production areas.

- The depuration results from the present study are likely to be broadly applicable, in conjunction with other published studies on rates of clearance of *E. coli.* Appropriate depuration times could be determined for predicted periods of lower or higher contamination, ranging from 12 to 72 hours. These guidelines could be universally applied across any catchments participating in an assurance scheme, but should include consideration of prevalence of norovirus in the human population and/or measured levels in shellfish.

- During the Camel study, the pour plate method was found to consistently yield less variable *E. coli* results than those from MPN, particularly for the higher end of the range of *E. coli* concentrations measured. This variability suggests that the MPN method has greater potential to generate outlier high results that may influence interpretation or application of monitoring data. This is exacerbated by the observation of statistically higher results for MPN compared to pour plate data. The pour plate method is approved for use in Official Control sampling, and the lower variability of *E. coli* results using this method support the case for its adoption. This would be particularly relevant for sites where management decisions (e.g. downgrades and closures) may be influenced by a few high results. For the Camel, changes in site classification by adoption of pour plate for Official Control samples is not considered likely. However, there may be instances elsewhere where this could be the case. The observed relationships between *E. coli* results obtained using the two methods for the Camel sampling could be more generally applied to other shellfish areas to investigate the potential effect on classifications elsewhere.

- The Camel study did not extend to comparison of other methods of testing *E. coli* in shellfish, specifically the impedance method which is also approved for use in Official Control monitoring. The rapid turnaround of results from the impedance method may be complementary to use of predictive modelling of *E. coli* levels in shellfish, where the quick response time could allow confirmatory results to support management decisions. The variability of *E. coli* results from the

impedance method, relative to the other methods, could also usefully be considered before collection of supplementary data sets to inform development of predictive models for use in an assurance scheme.

8.2 Conceptual design of a predictive tool to inform shellfish harvesting

The findings from the Camel case study can inform the conceptual design of a predictive tool to inform shellfish harvesting decisions that would be intrinsic to an industry-operated assurance scheme. Figure 8.1 shows an illustrative example of how such a tool could operate, based on the environmental indicators and predictive model outcomes from the Camel case study. The environmental inputs are likely to vary between catchments, depending on the models developed for each site, and the input of CSO operation may add to resolution of the predictive model in some locations. The output actions broadly reflect the range of *E. coli* results observed and predicted in the Camel. Clearly these may vary (or be redundant) in shellfish production sites where a different range of results is typical. For example, a site which largely operates across the A/B classification boundary might have a predictive tool that omits the potential closure due to results in the 4,600-46,000 *E. coli/100g* range.

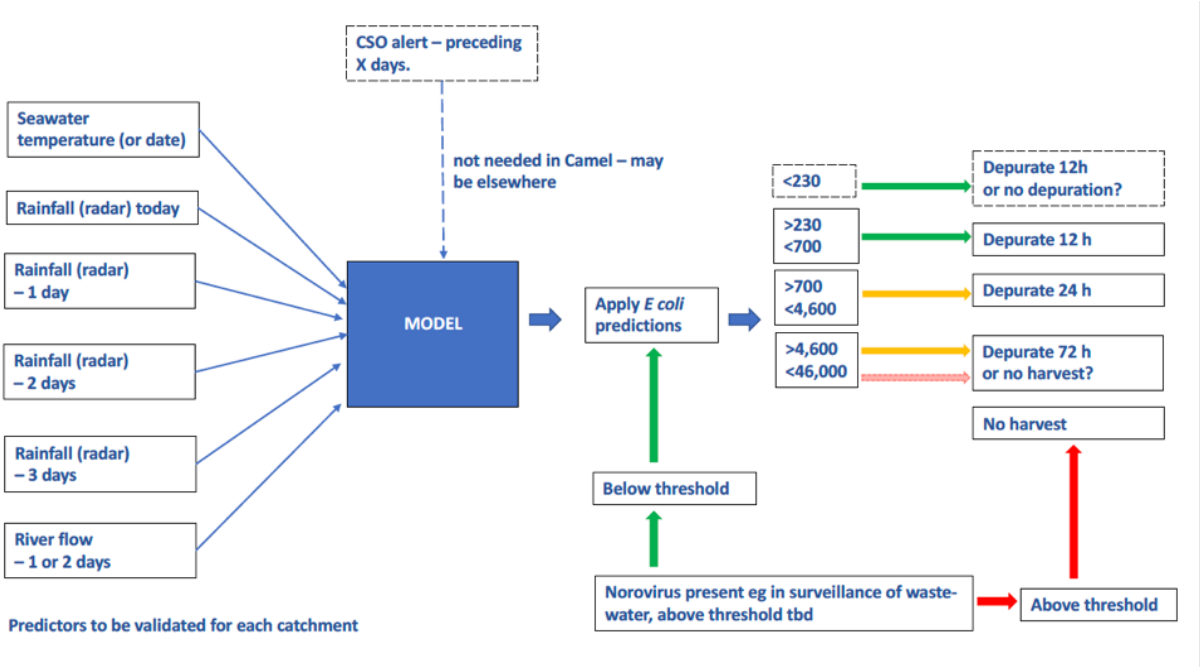


Figure 8.1 Generalised schematic for operation of a predictive decision tool in management of shellfish harvesting. The predicted *E. coli* ranges are indicative and many vary from catchment to catchment.

Notes: (i) thresholds for depuration and harvesting decisions are only illustrative (ii) for B class areas, such as the Camel, shellfish depuration will always be required under current regulations.

Notes: (i) thresholds for depuration and harvesting decisions are only illustrative (ii) for B class areas, such as the Camel, shellfish depuration will always be required under current regulations.

The Camel study has demonstrated the potential reliability of a predictive tool, which likely could be improved with ongoing data collection and algorithm refinement once the system is in operation. The exact level of reliability that would be sufficient for a model to be acceptable for application in an Assurance Scheme still needs to be determined and beyond the scope of the present study. It could be anticipated that reliability in correctly predicting high and low risk periods, will depend both on the precision of the underlying predictive model and on how broadly risk categories are defined.

The potential for inclusion of monitoring of norovirus prevalence is accounted for in the outline tool design. In the longer-term, this may become available in close to real-time (e.g. if wastewater surveillance is in in place). However, in the meantime, shellfish business operators may need to use current methods for monitoring of norovirus in shellfish to make over-riding harvest/closure decisions.

The most appropriate technical system for implementation of the predictive tool remains to be determined, but at its most accessible it is envisaged that harvesting recommendations might be delivered via an interface such as a mobile phone app and/or desktop version. The operator of the Assurance Scheme will be required to develop the technical solution to data acquisition and embedding of the predictive model algorithms in an appropriate hardware and software system with a user interface. The system would require bespoke algorithms for each shellfish production area, based on the predictive model developed for each catchment. The computational step could be set up locally on appropriate devices, or centrally managed by whichever organisation is contracted to deliver the assurance scheme, with results accessed remotely by users. In either case, a feed of environmental data inputs would be required to allow the computation of model outcomes. This will require engagement with agencies that gather environmental data (e.g. Met Office, Environment Agency, water companies). Once established the technical solution should be transferable between production areas, ideally with a centralised platform enabling effective technical support to multiple locations. The operator of the scheme will also need to put in place sufficiently robust internal audit and monitoring procedures to demonstrate continued compliance with the FSA's criteria for approving such schemes and UK Accreditation Service requirements or equivalent.

To participate in the Assurance Scheme, shellfish farms need to be evaluated (1) to understand the local environmental conditions that could influence fluctuations in water quality and levels of microbial contamination, (2) to confirm the specific environmental indicators that may provide an indication of water quality, and (3) to determine the data sources that will allow monitoring against these indicators. This is likely to be assessed on a case-by-case- basis for each production area; the Camel case study provides guidance on the likely data needed to establish a reliable site-specific predictive model to inform harvesting decisions, as well as appropriate sampling regimes and methods for initial model development and ongoing delivery of the assurance scheme.

Evaluation should consider both technical suitability and cost-benefits for shellfish operators and regulators. In some locations, sufficiently robust relationships between environmental predictors

and *E. coli* in shellfish may not be established. In others, the current classification system may be considered suitable and cost-effective by both producers and regulators.

Following agreement to proceed, a programme of supplementary time series sampling for *E. coli* data to inform predictive model development is likely to be needed, ideally using the pour plate method. Based on the experience from the Camel case study, it is anticipated that this may require a minimum 12 months of two weekly sampling. On completion of this initial data collection, initial modelling for the catchment can be undertaken. It is anticipated that model refinement will be ongoing, as new data is collected over time during participation in the scheme (as part of the Official Control sampling and/or more frequent supplementary sampling).