Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating GIS and Machine Learning to Prioritise Environmental Monitoring of Water Quality and Antimicrobial Resistance

Brian O'Sullivan

brian.b.osullivan@universityofgalway.ie

Ollscoil na Gaillimhe – University of Galway

Aaron Golden

Ollscoil na Gaillimhe – University of Galway

Alexandre Barretto de Menezes

Ollscoil na Gaillimhe – University of Galway

Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11540020/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrating GIS and Machine Learning to Prioritise Environmental Monitoring of Water Quality and Antimicrobial Resistance

Brian O'Sullivan[1*], Aaron Golden[2,3,4†] and Alexandre Barretto de Menezes[1,4*†]

[1*]School of Biological and Chemical Sciences, Microbiology, University of Galway, University Road, Galway, H91 TK33, Co. Galway, Ireland.
[2]Physics, School of Natural Sciences, University of Galway, University Road, Galway, H91 TK33, Co. Galway, Ireland.
[3]Department of Physics and Electronics, Rhodes University, Drostdy Road, Makhanda (Grahamstown), 6139, Eastern Cape, South Africa.
[4]The Ryan Institute, University of Galway, University Road, Galway, H91 TK33, Co. Galway, Ireland.

*Corresponding author(s). E-mail(s):
brian.b.osullivan@universityofgalway.ie;
alexandre.demenezes@universityofgalway.ie;
Contributing authors: aaron.golden@universityofgalway.ie;
†These authors contributed equally to this work.

Abstract

Emerging pathogens and antimicrobial resistance (AMR) pose growing One Health challenges, yet limited environmental surveillance hampers identification of priority monitoring locations. Here, we integrate geographic information systems (GIS), machine learning, publicly available bathing water monitoring data (2020–2025) and other environmental datasets to prioritise environmental AMR monitoring in Ireland. Self-organising maps classified sites by faecal contamination, while GIS-based source attribution and source-specific AMR weightings were combined to estimate the spatial distribution of relative environmental AMR risk across coastal waters. Four priority areas were identified. Coastal analysis highlighted wastewater discharges as a major environmental pressure, with aquaculture and recreation representing potential exposure pathways. Inland

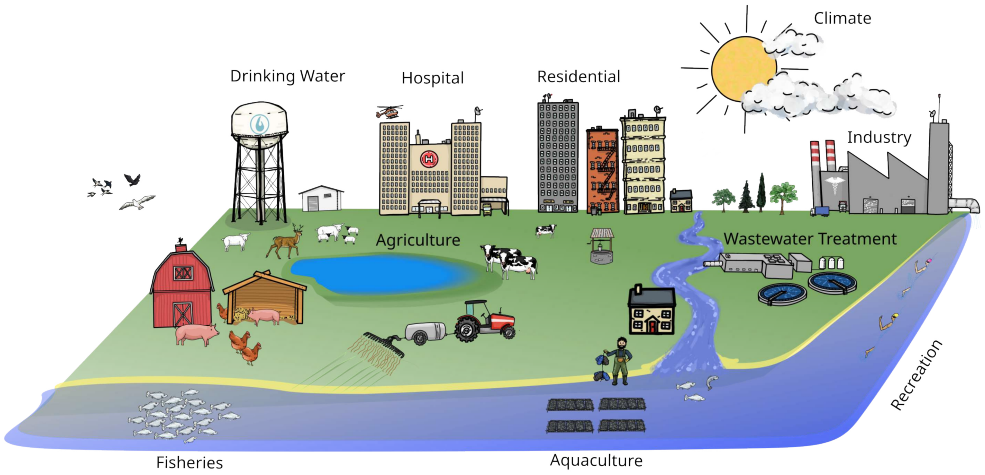


Fig. 1 Key drivers of emerging pathogens and antimicrobial resistance (AMR) within the One Health framework.

classification identified two sites with the highest faecal contamination nationally. Catchment analysis identified intensive poultry production as a prominent pressure at one site and multiple potential pressures at the other. This scalable framework demonstrates how existing environmental datasets can inform monitoring priorities, resource allocation and evidence-informed environmental policy.

Keywords: Antimicrobial resistance,AMR,Environmental surveillance,Faecal contamination,Geographic information systems,GIS,Machine learning,Water quality,Ireland

1 Introduction

Aquatic systems, including freshwater, marine, and groundwater environments, are essential to human and environmental health. They provide water for drinking, sanitation, and irrigation; support food production through fisheries and aquaculture; and sustain recreational and economic activities. At the same time, they receive effluents from domestic, industrial, and agricultural sources, making them potential reservoirs and dissemination pathways for pathogens and pollutants (Figure 1). The geographical extent of aquatic networks, coupled with the costs of advanced molecular analyses, presents significant challenges for environmental agencies seeking to implement effective national-scale surveillance programmes, detect pathogens of concern, and meet evolving regulatory requirements.

Under regulatory frameworks such as the EU Bathing Water Directive and US Environmental Protection Agency criteria, routine monitoring of the faecal indicator bacteria *Escherichia coli* and intestinal enterococci generates extensive spatial and

temporal datasets. Beyond assessing compliance with bathing water quality standards, these data could support the prioritisation of environmental monitoring locations when integrated with source attribution and geospatial information on environmental pressures. Such integration can help identify potential contamination hotspots, guide the placement of additional monitoring, and improve the allocation of surveillance resources. This approach may be particularly valuable for environmental antimicrobial resistance (AMR), where direct monitoring remains limited and the spatial distribution of potential contamination sources is poorly characterised.

Although AMR has not yet been systematically integrated into national water policy frameworks [1], its environmental surveillance is receiving increasing regulatory attention. In Europe, forthcoming requirements are expected to introduce surveillance of AMR in wastewater from 2030 [2], alongside proposals to incorporate AMR-related indicators into EU regulatory frameworks [3, 4]. National-scale environmental surveillance initiatives are also emerging in the United States. However, limited funding and the scale of aquatic networks make comprehensive monitoring of rivers, streams, and coastal waters difficult to achieve [5]. Surveillance and mitigation efforts must therefore be prioritised towards locations where potential risks warrant further investigation. Yet robust, scalable approaches for identifying such locations at national scale using publicly available data remain limited.

Geographic information systems (GIS) provide established tools for spatial analysis in environmental and public health. Their integration with environmental monitoring and machine learning has supported disease-cluster detection and more targeted interventions [6–9]. GIS-based approaches have also been widely applied to assess risks associated with climate change and air pollution [10]. However, applications to environmental AMR remain comparatively limited, with GIS often used to visualise data from relatively small, region-specific sampling campaigns [11, 12]. Few studies have integrated broader geospatial datasets to infer potential AMR sources and transmission pathways [13], and these investigations have generally been geographically restricted. To our knowledge, national-scale geotemporal microbiological monitoring data have not yet been used to support spatial assessments of relative environmental AMR risk.

Here, we present an integrated GIS and machine-learning framework that uses faecal contamination levels to inform the prioritisation of environmental AMR surveillance in Irish coastal waters and selected inland bathing sites. Six years of routine bathing water monitoring data were combined with national geospatial datasets on livestock distribution and wastewater discharges to classify faecal contamination levels and investigate potential contamination sources. For coastal waters, source-attributed faecal contamination estimates were combined with source-specific AMR weightings to generate a spatial assessment of relative environmental AMR risk. GIS analyses of four priority locations were then used to characterise potential environmental pressures and exposure pathways, identifying opportunities to spatially prioritise follow-up monitoring. By integrating existing microbiological monitoring with publicly available environmental data, the framework demonstrates how routine water-quality datasets can inform surveillance prioritisation. Its potential transferability to countries with comparable monitoring programmes and environmental datasets could

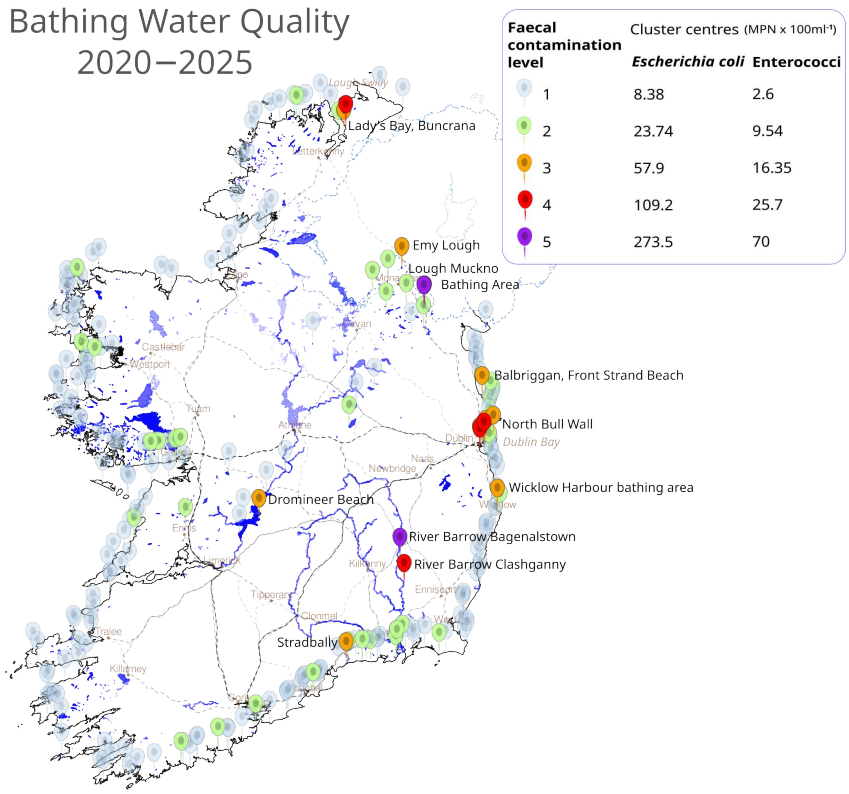


Fig. 2 Hierarchical clustering of self-organising map (SOM)-derived classifications based on median *Escherichia coli* and intestinal enterococci concentrations from routine bathing water monitoring by local authorities (2020–2025). Five levels of faecal contamination were identified and mapped across the Republic of Ireland using the `sf` package in R.

support the spatial prioritisation of AMR monitoring, strategic resource allocation, and evidence-informed environmental policy.

2 Results

2.0.1 SOM-based classification of faecal contamination levels from routine public monitoring

The within-cluster sum of squares (WCSS) elbow criterion identified five clusters as the optimal solution (Supplementary Figure 1A). Hierarchical clustering of the self-organising map (SOM) codebook vectors was then used to classify monitoring sites into five faecal contamination levels, ranging from Level 1 (lowest) to Level 5 (highest) (Supplementary Figure 1B; Figure 2). More than 75% of monitoring sites were assigned to Level 1, which was designated the national baseline for faecal contamination. Seasonal median *Escherichia coli* concentrations, calculated across all available

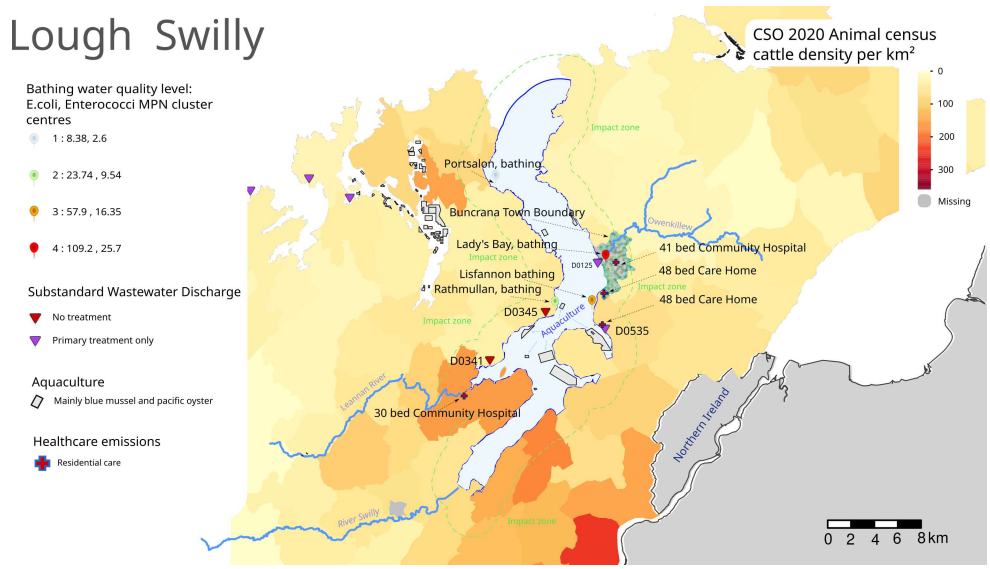


Fig. 3 Lough Swilly, one of two coastal segments with the highest estimated relative environmental AMR risk. Dominant anthropogenic pressures include primary-treated wastewater and untreated sewage discharges. Potential exposure pathways include shellfish aquaculture, recreational bathing, and water sports. Selected healthcare facilities within the wastewater catchments are shown.

monitoring years, are presented in Supplementary Figure 2 for four sites with high faecal contamination: North Bull Wall (Dublin), Lady's Bay (Buncrana), Lough Muckno (Monaghan), and River Barrow (Bagenalstown, Carlow).

2.1 Prioritising coastal sites for environmental AMR monitoring

The SOM-based classification identified 22 coastal segments containing monitoring sites with faecal contamination above the national baseline, while 50 segments remained at baseline levels (Supplementary Figure 3). Source-attributed faecal contamination levels were estimated for cattle, domestic wastewater treatment systems (DWWTS), and wastewater treatment plants (WWTPs), then combined with source-specific AMR weightings to generate a national map of relative environmental AMR risk (Supplementary Figure 4). Lough Swilly in the north and Dublin Bay in the east had the highest estimated relative AMR risk and were selected for further investigation of potential AMR sources and exposure pathways.

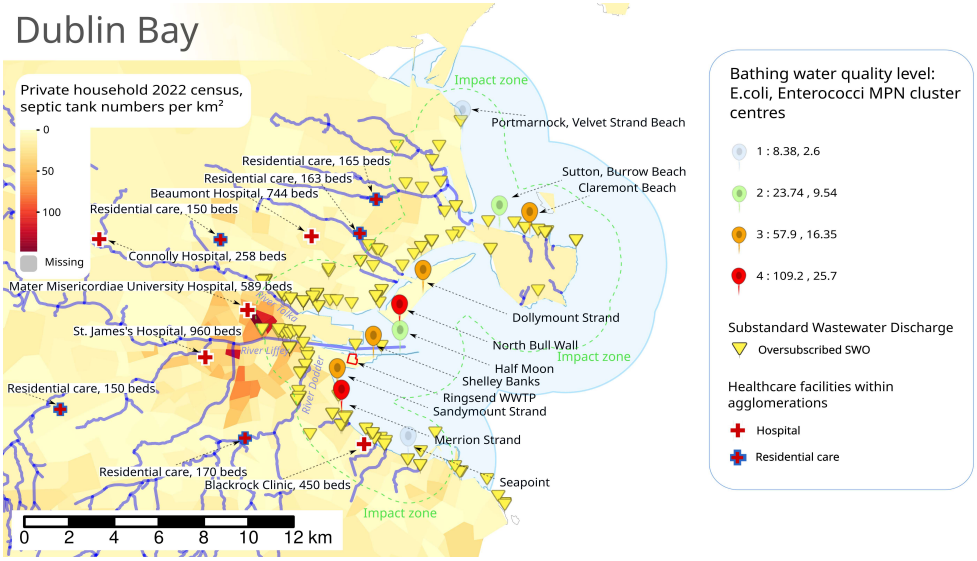


Fig. 4 Dublin Bay, a shallow, tidally dominated inlet. Bathing water data for *Escherichia coli* and intestinal enterococci, combined with source attribution, identified overflow discharges from the Ringsend WWTP as the dominant anthropogenic pressure. Dublin Bay ranked among the two coastal segments with the highest estimated relative environmental AMR risk. Popular bathing waters and selected major healthcare facilities are shown.

GIS analysis of Lough Swilly (Figure 3) identified wastewater discharges as the dominant anthropogenic pressure. Primary-treated effluent from the Buncrana and Fahan WWTPs (combined population served: approximately 10,200) represented the principal pressure along the eastern shoreline, whereas untreated sewage discharges from facilities serving Ramelton and Rathmullan (combined population served: approximately 2,600) predominated along the western shoreline. The WWTP agglomerations also contained healthcare facilities serving an estimated population of approximately 250 [14, 15]. Several non-compliant discharge events have been reported at the Buncrana facility. Septic tank density within the impact zone was approximately 47% higher than the national average [16], whereas cattle density was below the national mean [17]. Shellfish aquaculture, particularly mussel and Pacific oyster cultivation, and recreational bathing were identified as potential exposure pathways for AMR.

GIS analysis of Dublin Bay (Figure 4) identified overflow discharges from the Ringsend WWTP as the dominant anthropogenic pressure within the catchment [18, 19]. The Ringsend agglomeration serves healthcare facilities with a combined population of more than 10,000. Localised areas of elevated septic tank density were also identified [16].

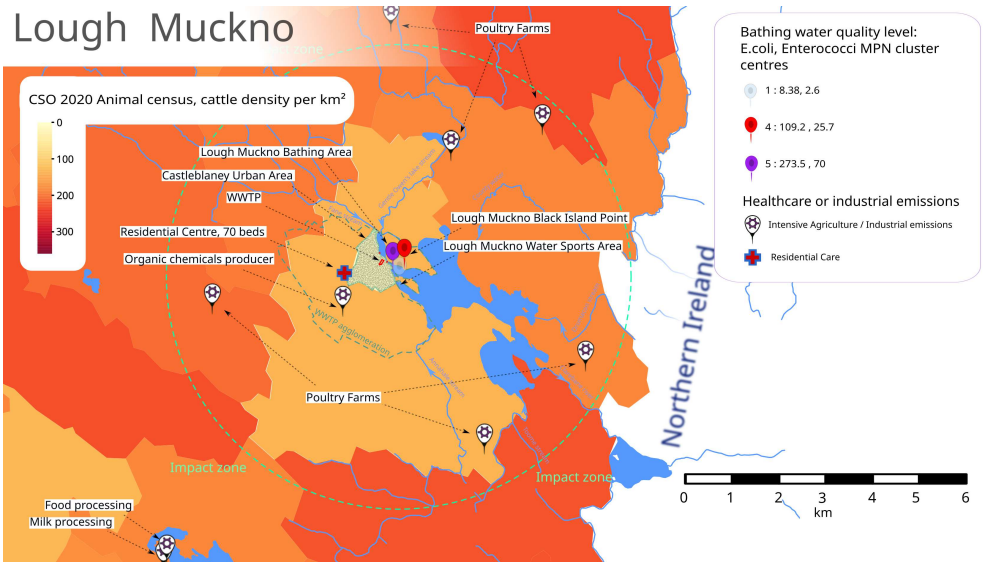


Fig. 5 Lough Muckno, a large lake in north-east Ireland draining to the Fane (Clarebane) River catchment. Anthropogenic pressures within a 5 km radius of the bathing sites indicate potential influence from intensive poultry production. The lake supports multiple recreational uses, including bathing, angling, and boating.

2.2 Prioritising inland sites for environmental AMR monitoring

Inland bathing sites had a significantly greater proportion of monitoring sites exceeding the national faecal contamination baseline than coastal sites (Fisher's exact test, $p < 0.008$). However, limited spatial coverage precluded interpolation and detailed source attribution across inland waters. Inland case studies were therefore selected based on SOM-derived contamination levels, and potential catchment-scale drivers were assessed.

Lough Muckno (Figure 5) was selected as the first inland case study. It lies within a cluster in north County Monaghan (Figure 2) containing more than half of all inland sites exceeding the national baseline. Within a 5 km impact zone centred on the bathing sites, intensive poultry production was the dominant catchment pressure, with an estimated 380,000 birds housed in the area [18]. Cattle and septic tank densities were approximately 60% and 42% above the national averages, respectively [16, 17]. A nearby WWTP providing tertiary treatment with phosphorus removal also discharges treated effluent and storm overflows to the lake (Figure 5) [18, 19].

The second inland case study comprised two adjacent monitoring sites along the River Barrow in south-eastern Ireland, both with relatively high faecal contamination levels. The impact zone spanning the upstream and downstream sites contained several potential AMR drivers, although no single dominant pressure was evident. The surrounding catchment is predominantly agricultural, with cattle and septic tank densities approximately 27% and 14% above national averages, respectively [16, 17]. It also

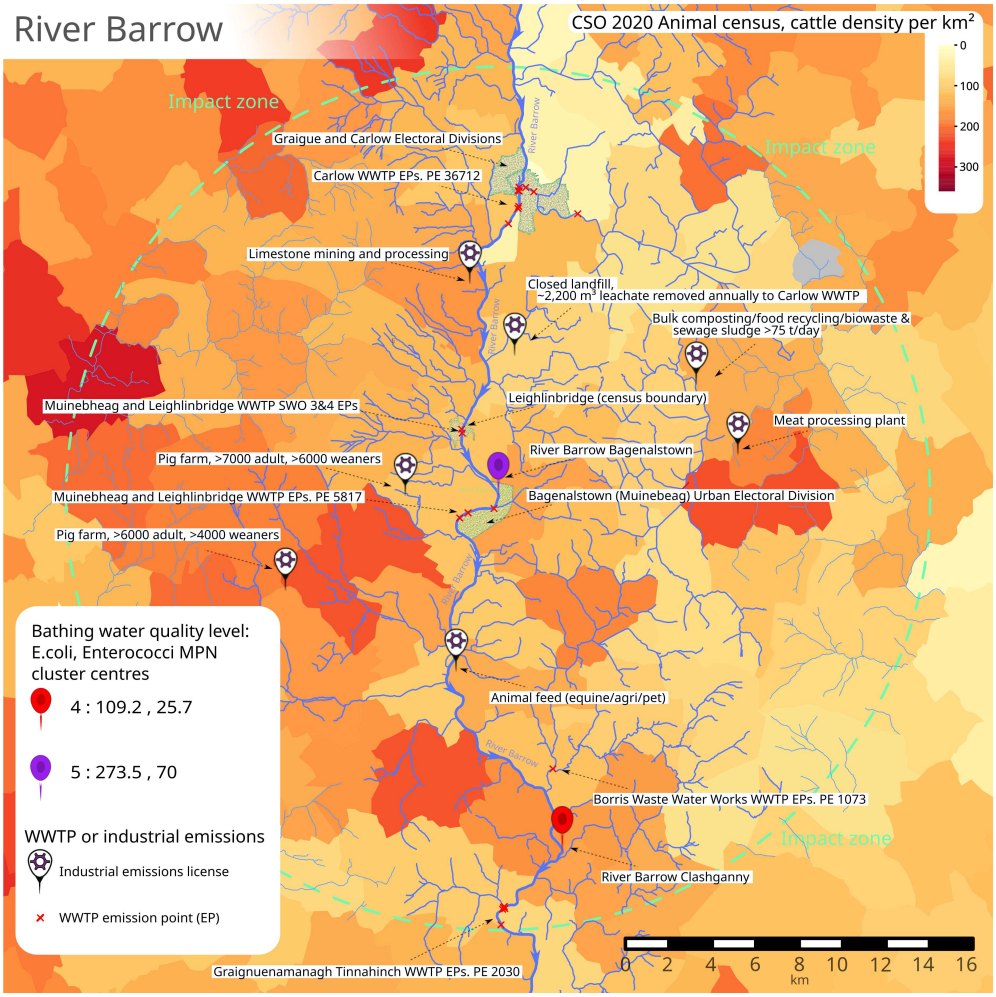


Fig. 6 The River Barrow flows through Bagenalstown and Clashganny, draining a broad catchment of tributaries, streams, and groundwater inputs from the surrounding countryside. The impact zone for this study extended from the Bagenalstown monitoring site downstream beyond the Clashganny monitoring site. In addition to designated bathing areas, this river section is widely used for angling.

contains two intensive pig farms housing more than 13,000 adult pigs and 10,000 weaners [18], as well as several licensed industrial emission facilities, including composting, meat-processing, and agri-food enterprises (Figure 6) [18]. All WWTPs discharging to this section of the River Barrow and its tributaries provide tertiary treatment, although non-compliant discharge events have been reported at some facilities [18, 19].

3 Discussion

This study demonstrates how routine water-quality monitoring and publicly available geospatial datasets can be integrated to inform the spatial prioritisation of microbiological monitoring. SOM-based classification identified four bathing sites with the highest faecal contamination levels nationally over the six-year study period, while GIS analysis characterised potential contamination sources and environmental pressures within their surrounding catchments. Together, these approaches provide a basis for prioritising locations for further investigation and identifying where additional monitoring could help resolve uncertainty about contamination sources. Although demonstrated here in the context of environmental AMR, the framework has broader potential to support water-quality surveillance and environmental management.

At Lough Swilly, long-term monitoring indicates increasing *Escherichia coli* and enterococci concentrations in recent years. At Lady's Bay, seasonal median *Escherichia coli* concentrations increased significantly between 2013 and 2025 (slope = 11.8 most probable number (MPN) per year, $p < 8 \times 10^{-5}$; Supplementary Figure 2B). By 2025, concentrations ranked among the highest recorded nationally at coastal bathing sites, with bathing warnings issued on 33 days during the season [20]. Wastewater discharges were identified as the principal anthropogenic pressure, alongside septic tank densities approximately 47% above the national average. Lough Swilly also supports recreational bathing, water sports, and intensive mussel and oyster aquaculture. Shellfish are of particular concern because they can bioaccumulate pathogens including antimicrobial-resistant bacteria, irrespective of bathing water restrictions [21–23].

In Dublin Bay, the Ringsend WWTP was identified as the dominant anthropogenic pressure. The plant treats approximately 40% of Ireland's public wastewater [24] and, before its recent upgrade, treated wastewater volumes equivalent to those generated by 2.6 million people, exceeding its design capacity of 2.1 million. This overloading may have contributed to the relatively poor bathing water quality observed during the study period. The large healthcare-associated population served by the plant represents an additional potential AMR pressure. Although upgrades completed in 2026 increased treatment capacity and reduced effluent nutrient loads, other pressures remain, including ageing sewerage infrastructure and combined stormwater overflows [25, 26]. These systems are relevant to environmental AMR because overflow events can release resistant microorganisms into receiving waters, particularly during heavy rainfall. Seasonal median *Escherichia coli* concentrations at North Bull Wall illustrate this vulnerability: the 2023 peak exceeded the previous maximum by more than threefold (Supplementary Figure 2A) and coincided with the second-wettest summer on record at Ringsend, based on rainfall data extending to 1941 [27]. With projected increases in extreme rainfall events, continued investment in sewer network upgrades will be important for reducing environmental contamination and associated AMR risks.

Inland bathing waters present greater challenges for identifying potential contamination sources. Compared with coastal sites, a significantly greater proportion of inland sites exceeded the national faecal contamination baseline, although this difference may partly reflect monitoring-site selection bias. Limited spatial coverage also restricts source attribution, while publicly available microbiological data are lacking for many inland waterbodies. Potential AMR pressures include agricultural slurry,

poultry litter, and sewage sludge applied to land [28]. However, limited information on the timing and location of these activities, together with limited access to DWWTS inspection results, constrains interpretation of monitoring data and assessment of environmental AMR risk.

At Lough Muckno, one of the lowest-performing inland bathing waters nationally, intensive poultry production was identified as the dominant pressure, with an estimated 380,000 birds housed within the catchment. Environmental compliance records indicate that wastewater is generally managed on-site, while poultry litter is transported off-site to locations not publicly identified. The catchment also supports extensive cattle grazing and dispersed rural housing, with cattle and septic tank densities above national averages. A nearby WWTP providing tertiary treatment with phosphorus removal discharges treated effluent and storm overflows to the lake. Seasonal variability in *Escherichia coli* concentrations at the monitoring sites suggests that contamination inputs fluctuate over time. Potential explanations include agricultural runoff associated with livestock and poultry management, alongside episodic wastewater releases and storm-driven inputs. Although ammonia and phosphorus noncompliances have been reported at the Castleblayney WWTP, their contribution to elevated faecal contamination at nearby bathing sites remains uncertain.

The River Barrow case study illustrates the challenges of prioritising environmental AMR monitoring locations across extensive, heterogeneous catchments. The section between Bagenalstown and Clashganny receives inputs from tributaries and groundwater and is surrounded by agricultural activities, licensed industrial facilities and WWTPs, all of which represent potential contamination sources. However, routine monitoring at only two sites limits robust source attribution and assessment of the relative contributions of individual pressures. Grab samples of *Escherichia coli* and intestinal enterococci, typically collected weekly or fortnightly during the bathing season, provide snapshots of microbiological conditions but may miss short-lived contamination events. These limitations complicate the identification and attribution of episodic inputs, particularly where multiple potential sources coexist.

Geospatial assessment of potential environmental pressures can help address these challenges by identifying priority locations for follow-up monitoring. Mapping potential contamination sources and their connectivity to receiving waters can guide additional sampling at selected tributary inputs or downstream of wastewater discharges. Depending on suspected sources and temporal patterns, monitoring could involve more frequent microbiological sampling, passive samplers to integrate contamination signals over time, or event-triggered autosamplers to capture short-lived discharge events. Microbial source tracking methods [29] could help distinguish human and animal sources of faecal contamination at sites with elevated faecal indicator levels and refine relative risk estimates at priority locations. Together, these approaches could strengthen source attribution and guide the allocation of limited monitoring resources, including microbiological and molecular methods for AMR quantification, to locations where they are likely to provide the greatest benefit.

Several limitations should be considered when interpreting the findings. Pathogen loadings, AMR concentrations and runoff rates were estimated using simplified, literature-derived values that did not account for site-specific conditions. Coastal

segments were treated as discrete units, without accounting for microbial transport between segments by coastal currents. Data availability also constrained the representation of potential contamination sources. Intensive agricultural operations below the Irish Environmental Protection Agency (EPA) Industrial Emissions licence threshold were excluded because publicly available data on their size and location were unavailable. Smaller WWTPs serving populations of fewer than 500 people were excluded from coastal source attribution because population data were unavailable through the EPA GeoPortal, although these facilities were considered in the detailed case studies. The timing and spatial distribution of slurry and poultry litter spreading, wastewater treatment sludge application and DWWTS inspection records were also unavailable. In addition, potential transboundary influences from Northern Ireland on AMR risk in Republic of Ireland water bodies were not assessed because acquiring and harmonising comparable datasets was beyond the scope of this study. Finally, the source-specific AMR weighting factors and resulting environmental AMR risk classifications had not been microbiologically validated at the time of publication. The resulting risk estimates should therefore be interpreted as relative indicators for spatial prioritisation rather than direct measurements of environmental AMR.

Despite these limitations, this study demonstrates how machine-learning classification, routine microbiological monitoring and publicly available geospatial datasets can be integrated to prioritise locations for environmental AMR monitoring. Beyond AMR, the framework illustrates how existing water-quality data can help identify potential contamination sources, guide follow-up sampling and support the strategic allocation of monitoring resources. Repeating the analysis with updated datasets and comparing results with historical assessments could help identify changes associated with evolving environmental pressures and wastewater infrastructure upgrades without requiring additional sampling for the initial assessment. The framework's adaptability to regions with comparable monitoring programmes and environmental datasets (Table 1) offers potential to support environmental management and evidence-informed policymaking.

4 Methods

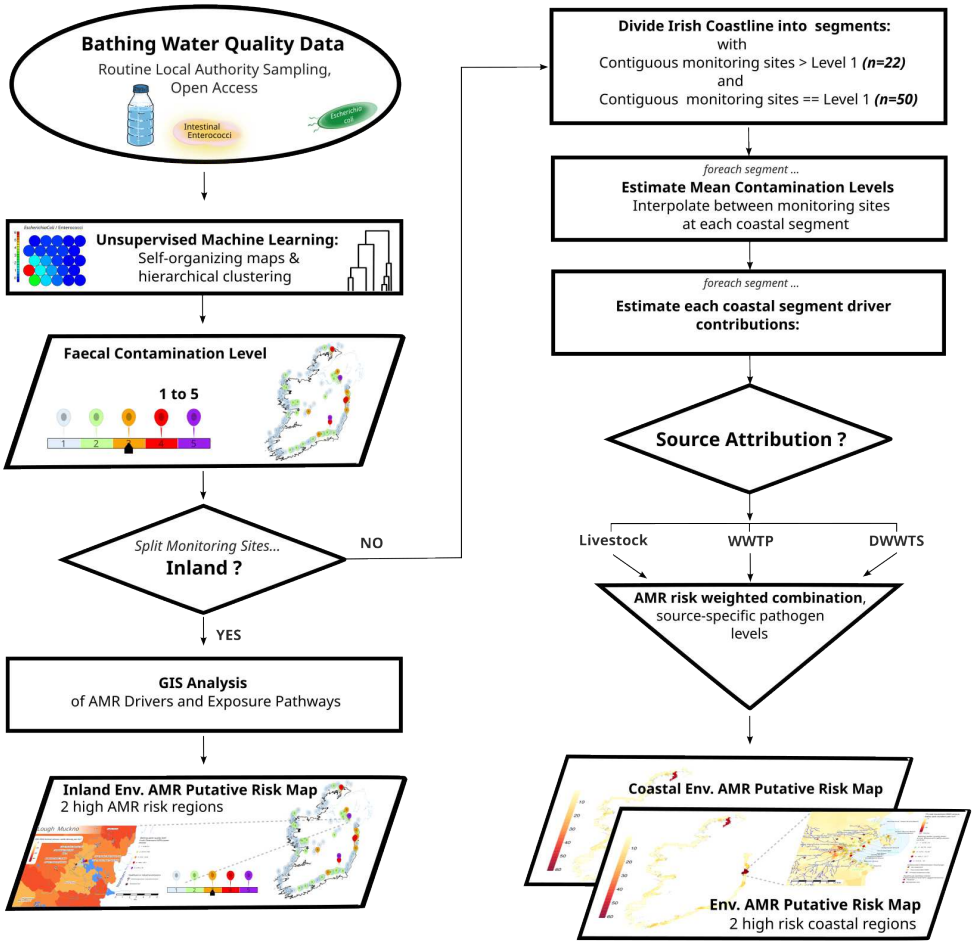


Fig. 7 Data-processing workflow used to identify inland monitoring sites and coastal segments with elevated faecal contamination levels and assess potential environmental AMR risk. GIS and unsupervised machine learning analyses of open datasets, including routine bathing water monitoring data for *Escherichia coli* and intestinal enterococci, were used to classify faecal contamination and infer potential anthropogenic sources. Selected case-study locations were analysed to identify contamination drivers and transmission pathways.

4.1 Data Sources

Dataset	Provider	Update frequency	Data year(s)	Licence
Livestock [17]	CSO	Every 10 years	2020	CC BY 4.0[a]
Domestic wastewater [16]	CSO	Every 5 years	2022	CC BY 4.0[a]
Urban wastewater plants [19]	EPA	Single snapshot[c]	2025	CC BY 4.0[a]
Urban wastewater emissions [30]	EPA	Single snapshot[c]	2025	CC BY 4.0[a]
Industrial emissions licences [31]	EPA	Single snapshot[c]	2025	CC BY 4.0[a]
Aquaculture [32]	DAFM	Single snapshot[c]	2025	CC BY 4.0[a]
Healthcare facilities [14, 15]	HIQA and OSM	Single snapshot[c]	2024 (HIQA); 2025 (OSM)	Open data
Bathing water quality [20]	EPA	1–3 times per month (1 June–15 September)[b]	2020–2025	CC BY 4.0[a]
Coastline of the Republic of Ireland [33]	Tailte Éireann	Single snapshot[c]	2025	CC BY 4.0
Coastline and land boundary of Northern Ireland [34]	Ordnance Survey of Northern Ireland	Single snapshot[c]	2025	OGL v3.0

Table 1 Publicly available datasets used in this study. Data providers: Central Statistics Office (CSO), Irish Environmental Protection Agency (EPA), Department of Agriculture, Food and the Marine (DAFM), Health Information and Quality Authority (HIQA), and OpenStreetMap (OSM). [a]CC BY 4.0: Creative Commons Attribution 4.0 International licence. [b]Sampling frequency for faecal contamination indicators, *Escherichia coli* and intestinal enterococci. [c]Single snapshot: a single data extract representing the dataset at the time of retrieval.

All datasets used in this study were obtained under open-data licences; no proprietary data were used (Table 1). Detailed data download and preprocessing methods are available on Zenodo (DOI: 10.5281/zenodo.23497114).

4.2 SOM-based classification of faecal contamination levels from routine public monitoring

Self-organising maps (SOMs) are unsupervised artificial neural networks that project high-dimensional data onto a lower-dimensional, typically two-dimensional, grid while preserving relationships among observations, making them well suited to clustering and visualising environmental datasets [35, 36].

Median MPN (most probable number, an estimate of the concentration of viable microorganisms in a sample) levels of the faecal indicators *Escherichia coli* and intestinal enterococci were calculated for each monitoring site using observations collected during the 2020–2025 bathing seasons. These data were then compiled into a georeferenced training dataset. The two indicator variables were standardised by centring and scaling using R's `scale()` function before classification using a two-stage hybrid clustering approach [36–38]. A SOM was first fitted to the standardised data using the `kohonen` package (v3.0.12), followed by hierarchical clustering of the SOM codebook vectors. The optimal number of clusters was determined using the within-cluster sum of squares (WCSS) elbow criterion. Final cluster assignments were generated using the `cutree` function from the base R `stats` package and mapped using `sf` (v1.0.21). The largest cluster, representing the lowest faecal contamination level, was designated the national faecal contamination baseline. Full implementation details are provided in the source code archived on Zenodo (DOI: 10.5281/zenodo.23497114).

Type	Source	Daily Discharge Volume (L day^{-1})	*Escherichia coli* (MPN 100 mL^{-1})	Enterococci (MPN 100 mL^{-1})	Runoff factor[a]	AMR weighting[b]
WWTP[c] [39]	Person	200	6.2×10^6	0.6×10^6	1.00	0.9
DWWTS[d] [40]	tank	800	2.1×10^6	0.2×10^6	0.05	0.85
Cattle [41]	head	47.14	1×10^8	1×10^7	0.05	0.5
Intensive Agriculture[e]	facility	–	–	–	–	–

Table 2 Summary of faecal contamination levels, volumes, runoff factors, and AMR weighting factors by source. [a]Runoff factor: estimates the fraction of the faecal indicator load in a discharge that reaches the receiving water body in a viable state. [b]Source-specific weighting factors representing the contribution of each faecal contamination source to estimated environmental AMR risk. [c]Values reported for wastewater treatment plants (WWTPs) represent untreated sewage and were adjusted during GIS analysis to account for treatment level. Primary-treated effluent was estimated as 0.7 times the equivalent raw sewage discharge [42]. Modern secondary treatment, which removes up to 99% of faecal indicator bacteria [43], and tertiary treatment were not included in the source attribution analysis. Overflow discharges from oversubscribed WWTPs were estimated as 0.08 times the equivalent raw sewage discharge from populations exceeding design capacity [44]. [d]Domestic wastewater treatment systems (DWWTS). [e]Only one intensive agriculture facility was identified within 4 km of a coastal segment, which exhibited baseline faecal contamination levels. Intensive agriculture was therefore excluded from further coastal source attribution. For inland bathing sites, site-specific factors including livestock numbers and type, and potential runoff from slurry, litter, and wastewater, were considered.

4.3 GIS analysis of coastal waters

Coastline classification and estimation of mean faecal contamination levels

A coastal subset of monitoring sites was extracted, and the Irish coastline was divided into discrete shoreline segments based on monitoring coverage and contamination status. Unmonitored shoreline sections exceeding 120 km (approximately 2% of the total coastline) were excluded from concentration estimation due to insufficient monitoring data. The remaining coastline was classified into two segment types: (i) contiguous monitoring sites above the national faecal contamination baseline, bounded by either a no-coverage section or a baseline site; and (ii) contiguous monitoring sites at baseline levels.

Mean faecal contamination levels were estimated for each segment by linearly interpolating concentrations between adjacent monitoring sites and integrating the resulting concentration profile using the trapezoidal rule. The integral was divided by segment length to obtain the distance-weighted mean (Figure 7).

Source attribution of faecal contamination levels

For each coastal segment, concentrations of *Escherichia coli* and intestinal enterococci were estimated using publicly available geospatial datasets (Table 1) describing livestock densities and wastewater inputs within a 2 km buffer around each segment, defined as the impact zone. These zones were used to characterise potential environmental pressures on the associated coastal segments. The pressures were scaled using literature-derived pathogen concentrations in domestic, municipal, and agricultural emissions (Table 2).

The resulting estimates were used to quantify the source-specific contributions from extensive agriculture, DWWTS, and WWTP discharges to faecal contamination recorded in public monitoring data (Figure 7).

Source-specific AMR weighting and coastal AMR risk assessment

Source-specific AMR weighting factors were applied to source-attributed faecal contamination levels to generate a spatial representation of potential environmental AMR hotspots along the Irish coastline (Figure 7). No published studies were identified that directly compare the environmental AMR risks associated with WWTP effluent, DWWTS effluent, and cattle-derived faecal contamination. Although livestock faeces may contain higher overall abundances of resistance genes [45], human faecal contamination was estimated to pose a greater immediate clinical risk because it is more likely to contain resistance to critically important antimicrobials and AMR strains adapted to human infection [46]. The resulting weighting factors (Table 2) should therefore be regarded as initial estimates rather than validated risk coefficients. The framework allows these weighting factors to be readily recalibrated to prioritise alternative risk contexts, such as AMR risks to livestock, without modifying the underlying model structure.

For simplicity, identical source weights were applied to *Escherichia coli* and intestinal enterococci. Although antimicrobial resistance gene (ARG) abundances in

DWWTS influent have been reported to be comparable to those in WWTP influent [47], the environmental AMR risk associated with DWWTS effluent was estimated to be lower because DWWTS are less likely to receive healthcare-associated influent. This assumption was partially offset by evidence that 59% of DWWTS inspected in Ireland during 2025 failed inspection [48], increasing the potential for inadequately treated wastewater to enter the environment. Accordingly, the AMR weighting factor assigned to DWWTS effluent was set slightly lower than that applied to WWTP effluent.

Coastal regions prioritised by estimated relative environmental AMR risk

The two coastal segments with the highest estimated relative environmental AMR risk were selected for further GIS-based analysis of potential environmental pressures within their respective impact zones (Figure 7). Aquaculture sites and recreational bathing locations were also assessed as potential exposure pathways. To characterise temporal patterns in faecal contamination, seasonal median concentrations across all available monitoring years were plotted for the monitoring site with the highest faecal contamination level within each segment.

4.4 GIS analysis of inland waters

SOM-based site classifications were spatially mapped using `sf` (v1.0.21) in R, with analyses restricted to inland monitoring sites. Routine bathing-season monitoring in Ireland is concentrated along the coastline (n = 216), with relatively few river and lake sites (n = 23). Due to this limited spatial coverage, faecal contamination levels could not be estimated along unmonitored sections of rivers or lakes. Inland sites were therefore treated as discrete point locations, with no interpolation between sites (Figure 7).

Inland bathing sites prioritised by faecal contamination levels

Two inland monitoring sites with the highest faecal contamination levels, based on SOM-derived site classifications, were selected for further analysis (Figure 7). Anthropogenic pressures within an impact zone surrounding each water body were quantified to assess their potential contribution to environmental AMR risk. Each impact zone represented the surrounding area within which environmental pressures could influence AMR risk at the associated monitoring site. Impact zone radii were estimated according to water body type and the distribution of potential AMR drivers within the wider catchment. To characterise temporal patterns in faecal contamination, seasonal median concentrations across all available monitoring years were plotted for the two inland monitoring sites with the highest faecal contamination levels.

5 Supplementary information

Supplementary information referenced in this article is available in the accompanying file, supplementary.pdf.

6 Data availability

A snapshot of the data used in the analysis has been archived on Zenodo (https://doi.org/10.5281/zenodo.23497114). The archive corresponds to the data used in this publication. All datasets used in this study were obtained under open-data licences; no proprietary data were used.

7 Code availability

The R scripts required to download and preprocess the data and reproduce the analyses are archived on Zenodo (DOI: 10.5281/zenodo.23497114). The latest project documentation and source code are available at https://github.com/BrianOSullivanGit/PRISMA_2023PSSRC925.

8 Acknowledgements

This research was supported by the Irish Department of Agriculture, Food and the Marine (DAFM) under Grant No. 2023PSSRC925.

The authors thank Caoimhe O'Sullivan for creating the artwork used in Figure 1.

9 Author contributions

B.O'S. conceived the study, curated the data, performed the formal analysis, and developed the methodology and software. B.O'S. wrote the original draft of the manuscript. A.D.M. contributed to conceptualisation, supervised the project, managed project administration, and acquired funding. A.G. contributed to the formal analysis, supervised the work, and acquired funding. A.D.M. and A.G. reviewed and edited the manuscript. All authors approved the final manuscript.

10 Competing interests

The authors declare no competing financial or non-financial interests.

11 Declaration of generative AI use

The authors used ChatGPT (OpenAI) to assist with language editing and improving phrasing. All content was reviewed and approved by the authors, who take full responsibility for the final manuscript.

References

[1] Grenni, P.: Antimicrobial resistance in rivers: A review of the genes detected and new challenges. Environmental Toxicology and Chemistry **41**(3), 687–714 (2022) https://doi.org/10.1002/etc.5289

[2] Hock, L., Luiken, R., Valério, E., Vargha, M., Vierheilig, J., Börjesson, S., Pitkänen, T., Schmitt, H.: Integrating amr surveillance into wastewater monitoring systems in 2025: a position on the implementation of article 17 of the urban wastewater treatment directive (uwwtd). Eurosurveillance **31**(3) (2026) https://doi.org/10.2807/1560-7917.es.2026.31.3.2500289

[3] European Commission: Proposal for a Directive amending Directive 2000/60/EC, Directive 2006/118/EC and Directive 2008/105/EC as regards surface water and groundwater pollutants. COM(2022) 540 final (2022)

[4] Bakos, V., Lóránt, B., Murray, A.K., Feil, E.J., Gaze, W.H., Plósz, B.G.: Antimicrobial risk assessment–aggregating aquatic chemical and resistome emissions. Water Research **271**, 122929 (2025) https://doi.org/10.1016/j.watres.2024.122929

[5] EU-WISH Consortium: EU-WISH T5.1 Mapping Survey. https://www.eu-wish.eu/results. Accessed 27 April 2026 (2024)

[6] José Becerra, M., Rocha, M.: Applications of Geotechnologies in the Field of Public Health. IntechOpen, London, UK (2024). https://doi.org/10.5772/intechopen.1003867 . http://dx.doi.org/10.5772/intechopen.1003867

[7] Nazia, N., Law, J., Butt, Z.A.: Modelling the spatiotemporal spread of covid-19 outbreaks and prioritization of the risk areas in toronto, canada. Health & Place **80**, 102988 (2023) https://doi.org/10.1016/j.healthplace.2023.102988

[8] Dahu, B.M., Alaboud, K., Nowbuth, A.A., Puckett, H.M., Scott, G.J., Sheets, L.R.: The role of remote sensing and geospatial analysis for understanding covid-19 population severity: A systematic review. International Journal of Environmental Research and Public Health **20**(5), 4298 (2023) https://doi.org/10.3390/ijerph20054298

[9] Lee-Cruz, L., Lenormand, M., Cappelle, J., Caron, A., De Nys, H., Peeters, M., Bourgarel, M., Roger, F., Tran, A.: Mapping of ebola virus spillover: Suitability and seasonal variability at the landscape scale. PLOS Neglected Tropical Diseases **15**(8), 0009683 (2021) https://doi.org/10.1371/journal.pntd.0009683

[10] Dheekwal, A., Singh, K., Sharma, A.: Gis-based modeling of traffic-related air pollution and public health risks: a systematic review with an emphasis on developing urban contexts. Computational Urban Science **5**(1) (2025) https://doi.org/10.1007/s43762-025-00225-6

[11] Asaduzzaman, M., Rousham, E., Unicomb, L., Islam, M.R., Amin, M.B., Rahman, M., Hossain, M.I., Mahmud, Z.H., Szegner, M., Wood, P., Islam, M.A.: Spatiotemporal distribution of antimicrobial resistant organisms in different water environments in urban and rural settings of bangladesh. Science of The Total Environment **831**, 154890 (2022) https://doi.org/10.1016/j.scitotenv.2022.

154890

[12] Knight, M.E., Farkas, K., Kiss, A., Jones, D.L.: National-scale insights into amr transmission along the wastewater-environment continuum. Water Research **282**, 123603 (2025) https://doi.org/10.1016/j.watres.2025.123603

[13] Chique, C., Cullinan, J., Hooban, B., Morris, D.: Mapping and analysing potential sources and transmission routes of antimicrobial resistant organisms in the environment using geographic information systems—an exploratory study. Antibiotics **8**(1), 16 (2019) https://doi.org/10.3390/antibiotics8010016

[14] Health Information and Quality Authority: Find a Centre. Accessed 19 May 2026 (2026). https://www.hiqa.ie/find-a-centre

[15] OpenStreetMap contributors: OpenStreetMap. Accessed 15 May 2026 (2026). https://www.openstreetmap.org

[16] Central Statistics Office: Census 2022: Permanent Private Households (SAP2022T6T7ED). data.gov.ie dataset. Accessed: 2026-04-29. Data retrieved via PXStat GeoMap API: https://ws.cso.ie/public/api.static/PxStat.Data.GeoMap_API.Read/c5b950f2f3ab85cc657c4c0082b9fd05 (2022). https://data.gov.ie/dataset/sap2022t6t7ed-permanent-private-households

[17] Central Statistics Office: Census of Agriculture 2020: Farms with Livestock (AVA42). data.gov.ie dataset. Accessed: 2026-04-29. Data retrieved via PXStat API: https://ws.cso.ie/public/api.restful/PxStat.Data.Cube_API.ReadDataset/AVA42/CSV/1.0/en (2023). https://data.gov.ie/dataset/ava42-farms-with-livestock/resource/b0874c44-5c8b-42a4-9920-c00678cf6920

[18] Environmental Protection Agency: Licence and Enforcement Access Portal. LEAP Online Environmental Data Portal. Accessed 1 May 2026 (2026). https://leap.epa.ie/

[19] Environmental Protection Agency: SIIF Urban Waste Water Emission Points. EPA GeoServer, Web Feature Service (WFS). Accessed: 2026-04-29. Type name EPA:SIIF_UWWEmissionPoints. Metadata and reuse conditions: https://gis.epa.ie/geonetwork/srv/api/records/8034f005-acf6-45b4-9c55-33c23fb74c78 (2025). https://gis.epa.ie/geoserver/EPA/ows

[20] Environmental Protection Agency: Bathing Water Open Data API. EPA Open Data Portal. Accessed: 2026-04-29 Also accessible through https://www.beaches.ie (2025). https://data.epa.ie/api-list/bathing-water-open-data

[21] Desdouits, M., Reynaud, Y., Philippe, C., Guyader, F.S.L.: A comprehensive review for the surveillance of human pathogenic microorganisms

in shellfish. Microorganisms **11**(9), 2218 (2023) https://doi.org/10.3390/microorganisms11092218

[22] Mohammed, R., Nader, S.M., Hamza, D.A., Sabry, M.A.: Public health concern of antimicrobial resistance and virulence determinants in e. coli isolates from oysters in egypt. Scientific Reports **14**(1) (2024) https://doi.org/10.1038/s41598-024-77519-y

[23] Rees, E.E., Davidson, J., Fairbrother, J.M., St. Hilaire, S., Saab, M., McClure, J.T.: Occurrence and antimicrobial resistance of *Escherichia coli* in oysters and mussels from atlantic canada. Foodborne Pathogens and Disease **12**(2), 164–169 (2015) https://doi.org/10.1089/fpd.2014.1840

[24] Uisce Éireann: Ringsend Wastewater Treatment Plant Upgrade Project. Accessed 1 May 2026 (2026). https://www.water.ie/projects/local-projects/ringsend-wastewater-treatment-plant-upgrade-project

[25] Uisce Éireann: Chief Executive Officer's Statement. Accessed 11 May 2026 (2024). https://www.water.ie/about/model-publication-scheme/financial-information/strategic-report/chief-executive-officers-statement.html

[26] Uisce Éireann: Grand canal storm water outfall extension project: Environmental impact assessment report. volume 2: Main report. Technical report, Uisce Éireann, Dublin, Ireland (2023). Accessed 11 May 2026. https://www.pleanala.ie/publicaccess/EIAR-NIS/313738/GCSWOE%20Planning%20Application/Environmental%20Impact%20Assessment%20Report/Volume%202%20EIA%20Main%20Report.pdf

[27] Met Éireann: Historical Climate Data (Rainfall). Accessed 1 May 2026 (2026). https://www.met.ie/climate/available-data/historical-data

[28] European Commission, Eurostat: Data Browser: TEN00030. Eurostat Database. Accessed 12 May 2026 (2026). https://ec.europa.eu/eurostat/databrowser/product/page/ten00030

[29] Harwood, V.J., Staley, C., Badgley, B.D., Borges, K., Korajkic, A.: Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS Microbiology Reviews **38**(1), 1–40 (2014) https://doi.org/10.1111/1574-6976.12031

[30] Environmental Protection Agency: LEMA Facilities – Urban Waste Water. EPA GeoServer, Web Feature Service (WFS). Accessed: 2026-04-29. Type name EPA:LEMA_Facilities_UWW. Metadata and reuse conditions: https://gis.epa.ie/geonetwork/srv/api/records/8034f005-acf6-45b4-9c55-33c23fb74c78 (2025). https://gis.epa.ie/geoserver/EPA/ows

[31] Environmental Protection Agency: LEMA Facilities (Industrial Emissions Licence). EPA GeoServer, Web Feature Service (WFS). Accessed: 2026-04-29. Type name EPA:LEMA_Facilities_IEL. Metadata and reuse conditions: https://gis.epa.ie/geonetwork/srv/api/records/7905844c-a43d-4dd4-b262-c95c7aa0e9c7 (2025). https://gis.epa.ie/geoserver/EPA/ows

[32] Marine Institute: Aquaculture Sites. Ireland's Marine Atlas, ArcGIS REST service. Accessed: 2026-04-29. Metadata and reuse conditions: https://www.marine.ie/site-area/online-policies/re-use-public-sector-information (2025). https://atlas.marine.ie/arcgis/rest/services/05_Aquaculture_Sites

[33] Tailte Éireann: Coast – National 250K Map of Ireland. data.gov.ie dataset. Open geospatial dataset, accessed 11 October 2026 (2025). https://data.gov.ie/en_GB/dataset/coast-national-250k-map-of-ireland

[34] Ordnance Survey of Northern Ireland: OSNI Open Data - Largescale Boundaries - NI Outline. Open Data NI. Open geospatial dataset, accessed 11 October 2026 (2025). https://opendatani.gov.uk

[35] Astel, A., Tsakovski, S., Barbieri, P., Simeonov, V.: Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Research **41**(19), 4566–4578 (2007) https://doi.org/10.1016/j.watres.2007.06.030

[36] Licen, S., Astel, A., Tsakovski, S.: Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review. Science of The Total Environment **878**, 163084 (2023) https://doi.org/10.1016/j.scitotenv.2023.163084

[37] Vesanto, J., Alhoniemi, E.: Clustering of the self-organizing map. IEEE Transactions on Neural Networks **11**(3), 586–600 (2000) https://doi.org/10.1109/72.846731

[38] Tripathi, K.: The novel hierarchical clustering approach using self-organizing map with optimum dimension selection. Health Care Science **3**(2), 88–100 (2024) https://doi.org/10.1002/hcs2.90

[39] Galvin, S., Boyle, F., Hickey, P., Vellinga, A., Morris, D., Cormican, M.: Enumeration and characterization of antimicrobial-resistant *Escherichia coli* bacteria in effluent from municipal, hospital, and secondary treatment facility sources. Applied and Environmental Microbiology **76**(14), 4772–4779 (2010) https://doi.org/10.1128/aem.02898-09

[40] Environmental Protection Agency: Code of practice for domestic waste water treatment systems (population equivalent <= 10). Technical report, Environmental Protection Agency (2021). Accessed 13 May 2026.

https://www.epa.ie/publications/compliance--enforcement/waste-water/2021_CodeofPractice_Web.pdf

[41] Vinten, A.J.A., Douglas, J.T., Lewis, D.R., Aitken, M.N., Fenlon, D.R.: Relative risk of surface water pollution by *E. coli* derived from faeces of grazing animals compared to slurry application. Soil Use and Management **20**(1), 13–22 (2004) https://doi.org/10.1079/sum2004214

[42] Raboni, M., Gavasci, R., Torretta, V.: Assessment of the fate of escherichia coli in different stages of wastewater treatment plants. Water, Air, & Soil Pollution **227**(12) (2016) https://doi.org/10.1007/s11270-016-3157-8

[43] Kay, D., Crowther, J., Stapleton, C.M., Wyer, M.D., Fewtrell, L., Edwards, A., Francis, C.A., McDonald, A.T., Watkins, J., Wilkinson, J.: Faecal indicator organism concentrations in sewage and treated effluents. Water Research **42**(1-2), 442–454 (2008) https://doi.org/10.1016/j.watres.2007.07.036

[44] O'Sullivan, K.: Over one-fifth of Dublin Bay water users become ill, pressure group claims. *The Irish Times*. Accessed 15 May 2026 (2021). https://www.irishtimes.com/news/environment/over-one-fifth-of-dublin-bay-water-users-become-ill-pressure-group-claims-1.4536227

[45] Li, B., Jiang, L., Johnson, T., Wang, G., Sun, W., Wei, G., Jiao, S., Gu, J., Tiedje, J., Qian, X.: Global health risks lurking in livestock resistome. Science Advances **11**(26) (2025) https://doi.org/10.1126/sciadv.adt8073

[46] Agga, G.E., Arthur, T.M., Durso, L.M., Harhay, D.M., Schmidt, J.W.: Antimicrobial-resistant bacterial populations and antimicrobial resistance genes obtained from environments impacted by livestock and municipal waste. PLOS ONE **10**(7), 0132586 (2015) https://doi.org/10.1371/journal.pone.0132586

[47] Price, S., Krometis, L.-A., Magee, J., Darling, A., Pruden, A.: Extending wastewater-based surveillance to serve rural communities: Comparison of antimicrobial resistance indicator genes in septage and wastewater influent. Science of The Total Environment **1000**, 180456 (2025) https://doi.org/10.1016/j.scitotenv.2025.180456

[48] Environmental Protection Agency: Domestic waste water treatment system inspections 2025. Annual inspection report, Environmental Protection Agency, Wexford, Ireland (2026). https://www.epa.ie/publications/compliance--enforcement/waste-water/septic-tanks/domestic-waste-water-treatment-system-inspections-2025.php Accessed 2026-07-12

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf