Supplementary Information

Integrating GIS and Machine Learning to Target Environmental Surveillance 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.

1 Determination of the Optimal Number of Clusters

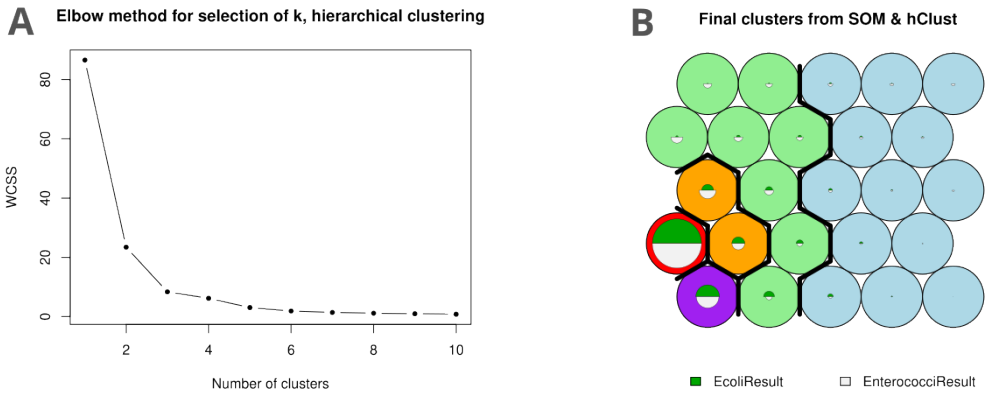


Fig. 1 **A:** The optimal number of clusters *(k)* was determined using the Elbow Method based on the Within-Cluster Sum of Squares (WCSS). WCSS was calculated across a range of cluster numbers and plotted against k. The optimal solution was identified as the elbow point, beyond which additional clusters yielded only marginal reductions in WCSS. **B:** Hierarchical clustering of the SOM codebook vectors using the base R stats::cutree function with k=5

2 Seasonal Escherichia coli measurements at the four lowest-quality bathing water monitoring sites nationally, 2020–2025

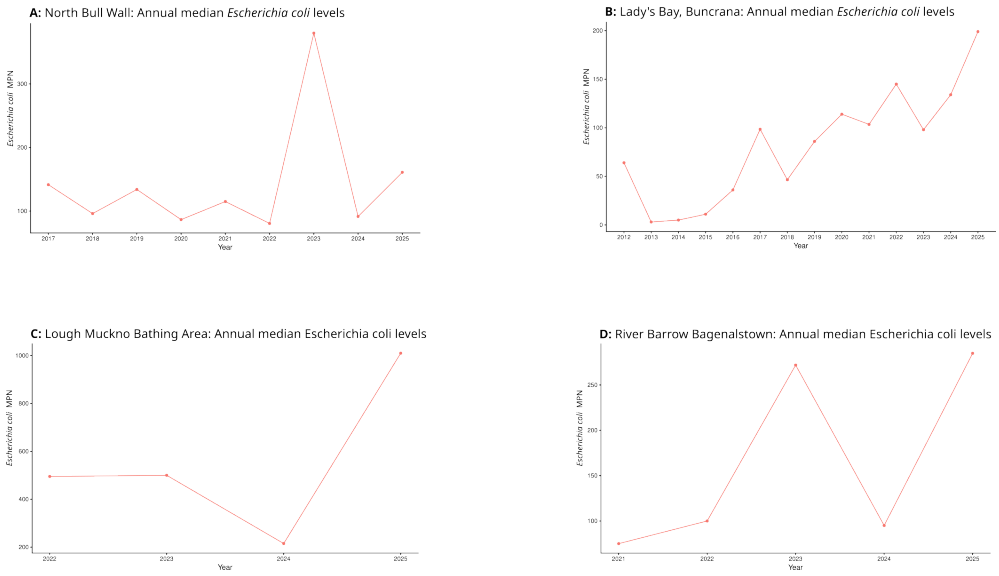


Fig. 2 Seasonal median Escherichia coli concentrations for all available years at four monitoring sites, including two coastal (A, B) and two inland (C, D) bathing sites with the lowest bathing water quality nationally.

3 Coastal segments identified based on analysis of routine bathing water microbiological monitoring data

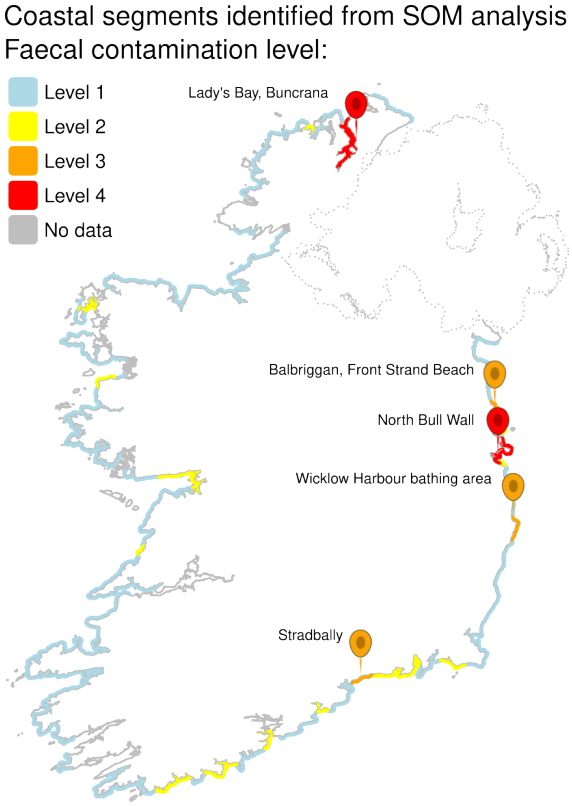


Fig. 3 Coastal segments are colour-coded according to the highest SOM-based faecal contamination level recorded at any monitoring station within each segment. Monitoring stations classified as Level 3 or higher are also shown to provide additional spatial context.

4 Coastal map of estimated relative environmental AMR risk

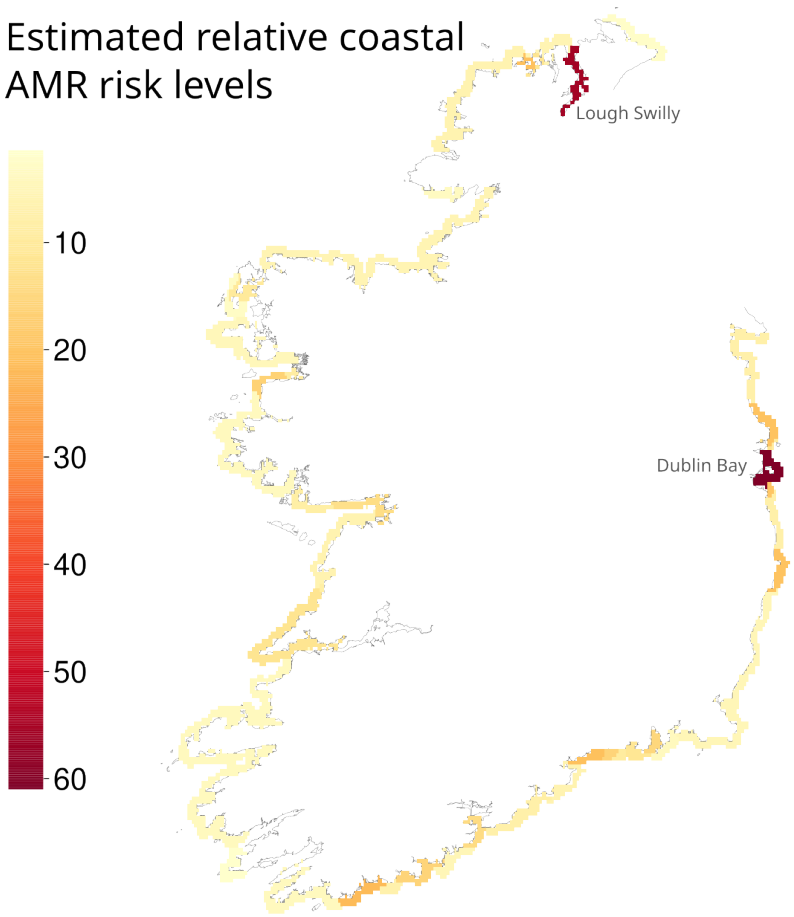


Fig. 4 Coastal map showing estimated relative environmental AMR risk.