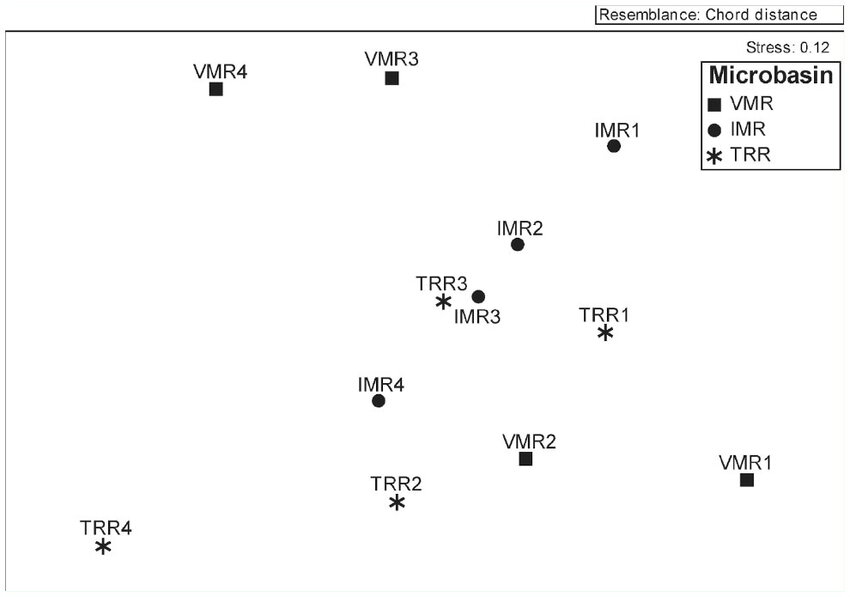
**Technical Report 1**

**Running a Non-Metric, Multidimensional Scaling (NMDS) to analyse community ecology on freshwater samples.**

**Introduction**

Ecosystems are dynamic and intricate, comprising of a multitude of species that interact within complex ecological networks. The increasing influence of environmental pressures has amplified the need to understand how communities respond and adapt over time. Biodiversity monitoring has become crucial in this endeavour, enabling the assessment of ecosystem health and resilience by linking shifts in species composition to underlying environmental parameters and spatial variability1. The accumulation of ecological data, alongside advances in analytical techniques allow these complex patterns to be understood, employing robust multivariate methods capable of managing the inherent variability and complexity within ecological systems, specifically in this report, freshwater.

Non-Metric Multidimensional Scaling (NMDS), implemented through the vegan package in R, provides a fundamental method for assessing community ecology2. NMDS is a powerful ordination technique that allows the visualisation of similarities and differences in the species assemblages between different sites without assuming a linear relationship. Unlike other biodiversity indices that condense community data into a single value (such as Shannon index or species richness), NMDS keeps the complexity of species relationships, providing a more exact understanding to ecological patterns.

**Figure 1:** An ideal example of sampling from different sites and different micro basins. An NMDS was carried out and the stress test is between 0.1-0.2 indicating a fair fit to carry on the analysis3

**Figure 1** illustrates a scenario with varying habitats across different sites. NMDS is particularly well-suited for such analyses due to its robustness in handling non-linear data and the common occurrence of zero counts in ecological community datasets. It functions by ranking pairwise dissimilarities between sample sites and projecting them into a low-dimensional space, where the distances between points visually represent the degree of community similarity. The combination of this clear visualisation and the ability to assess ordination reliability through a stress test highlights NMDS as a powerful and effective analytical technique.

**Methods & Data**

The data collection methods outlined in this section followed the procedures in **‘Braid Burn: River Habitat Survey.pdf’** by Andrew Innes4. Data analysis was conducted using a combination of the tutorial **‘Analysis of Ecological Communities in Vegan.pdf’** provided by Andrew Innes and additional R code help from Stack Overflow5.

**A. Data collection (Kick sampling)**

A map of a golf course

AI-generated content may be incorrect.To run the NMDS analysis, freshwater sampling was conducted using the kick sampling technique. Kick sampling is a widely used and well-established method in freshwater ecology, recognised for its reliability in providing representative samples compared to alternative techniques⁶. This method targeted benthic macroinvertebrate communities, which serve as key bioindicators of aquatic ecosystem health, and was carried out following the Scottish Environmental Protection Agency (SEPA) protocol⁷. Samples were collected along a 500-meter stretch of the river, with sampling points located both inside and outside designated habitats, as illustrated in **Figure 2**. At each site, kick sampling was performed over a standardised 3-minute period using a D-frame mesh net, ensuring consistency across all samples. The net was positioned downstream while the riverbed was disturbed by foot to dislodge invertebrates into the flow. Multiple habitats within each site, such as riffles and pools, were sampled to ensure a comprehensive representation of benthic communities. Invertebrates were then identified to the lowest practical taxonomic level and recorded as count data for each species, providing a robust dataset for subsequent NMDS analysis.

***Figure 2****:* *Map of the sampling area along a 500-meter stretch of the river, showing the 10 designated kick sampling sites. Sampling points were distributed to capture community composition both inside and outside habitat boundaries4*

**B. Data analysis (NMDS)**

**B.1 Data preparation**

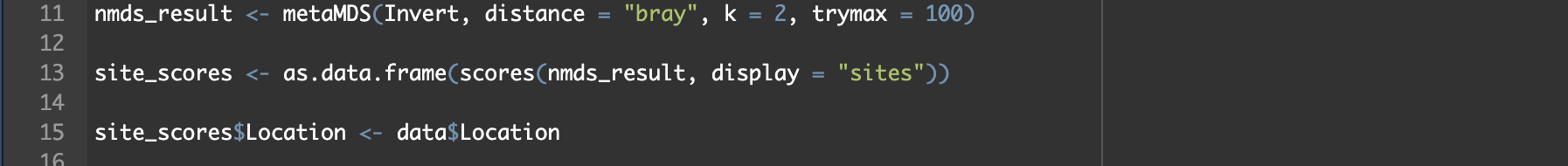
**A black background with white text

AI-generated content may be incorrect.**With the invertebrate data recorded as count data, the first step was to import it into R. The dataset was already formatted as a matrix, allowing for a direct import as a CSV file (**Figure 3)**. Since the metaMDS function only processes numerical data, the ‘Location’ column was temporarily removed prior to analysis but was reattached later for grouping. Additionally, any missing values (NAs) in the matrix were replaced with zeros for the same reason.



**Figure 3**: R code showing the data preparation stage of running an NMDS

**B.2 Dissimilarity measure:**

The NMDS analysis was performed using the metaMDS function from the vegan package (**Figure 4**). A two-dimensional solution (k = 2) was selected to simplify visual interpretation while maintaining the integrity of the data. The ‘Bray-Curtis’ dissimilarity index was used, as it is commonly applied to ecological count data and effectively accounts for species abundance. The function was run with 100 iterations, I made this choice to ensure stability within the analysis. After the NMDS is complete, the location column is reattached to the ordination results and stored ready for plotting which can be seen in results.



**Figure 4**: R code showing the NMDS analysis main part, as well as the rejoining and grouping of the location column.

**B.3 Stress test**

A stress test in NMDS analysis is further evaluation of how well this new ordination represents the original dissimilarities between samples8. The stress value quantifies this fit, with lower values indicating a more accurate representation. Generally, a stress value below 0.1 is considered excellent, between 0.1-0.2 is fair and values about 0.2 are a poor fit leading to misleading ordination9,10. A simple stress test is run prior to plotting any graphs to make sure the NMDS analysis is a good fit which can be seen in **Figure 5.**

**Figure 5**: A simple stress test ran in R returning with a value of below 0.1 indicating an 'excellent' fit

**Results & Analysis**

**NMDS ordination plot**

**Figure 6** presents the plot of the NMDS analysis, visualising the differences in community composition between sites located inside and outside the habitat. The degree of separation between inside (blue) and outside (red) samples are shown with some overlap. These overlapping ellipses, generated using the ggforce package, represent 95% confidence intervals, indicating that species assemblages inside and outside the habitat form distinct clusters while still sharing some community structure¹¹. The separation seen in the plot implies potential ecological differences between the two habitats and sites, which may be influenced by environmental variables such as substrate composition, habitat complexity, or something simple such as flow rate.

A graph with dots and numbers

AI-generated content may be incorrect.

**Figure 6**: NMDS analysis plot looking at the dissimilarity between inside and outside habitats. Each point represents a site, and the colour dictates the habitat.

**Stress test**

A graph of a number of objects

AI-generated content may be incorrect.The reliability of the ordination technique was assessed through a stress test, as described previously in Section **B.3**. The analysis produced a stress value of 0.07, indicating an excellent fit and a reliable NMDS ordination. The stress plot (**Figure 7**) shows a strong correlation between observed dissimilarities and ordination distances, with minimal scatter around the fit line. Additionally, the high R² values (non-metric fit = 0.995, linear fit = 0.974) confirm that NMDS effectively preserves the rank order of community dissimilarities, ensuring that the reduced-dimensional representation accurately reflects the relationships in the original dataset. If the R² values were lower, particularly below 0.9, it would instead suggest that the ordination is not adequately capturing the true dissimilarities between samples.

**Figure 7:** *Stress plot of the NMDS ordination, illustrating the relationship between observed dissimilarities and ordination distances*

**Discussion & Conclusions**

Following up from the results of the NMDS ordination, it successfully showed the differences in macroinvertebrate community composition between inside and outside habitat sites. The clustering within the plot suggest species assemblages are influenced by habitat conditions, with inside sites showing a tendency to group separately from outside sites. Although, some degree of overlap indicates shared species between inside and outside, which may be due to habitat connectivity or environmental gradients that influence community structure. The success of the NMDS analysis alongside the stress test results that not only indicate an excellent fit for the ordination but also demonstrated a strong correlation between observed dissimilarities and ordination distances with a high R2 score, justify the use of this technique as a suitable tool for assessing ecological communities in freshwater ecosystems.

While NMDS successfully visualised species distributions, its interpretation remains inherently descriptive, as it does not provide direct statistical evidence for differences between groups. To rigorously assess whether community composition varies significantly between inside and outside sites, a ‘Permanova’ test could be conducted5. This test would quantify the degree of dissimilarity between the two habitat types and determine if observed patterns are statistically significant rather than occurring by chance. NMDS does not inherently explain why species distributions differ. To address this limitation, environmental fitting analyses using envfit() could be employed accounting for variables such as flow rate, substrate type, or water quality, that had previously been mentioned in the report. Additionally, alternative ordination methods such as Canonical Correspondence Analysis (CCA), could be used if the goal is to directly correlate species distributions with environmental gradients12.

In conclusion, NMDS proved to be an effective multivariate tool, successfully distinguishing sites based on species assemblages and highlighting potential habitat-driven differences. While the clustering patterns observed suggest environmental influences, further statistical validation through analyses such as ‘Permanova’ or environmental fitting would strengthen these findings. Incorporating longitudinal data could provide insight into seasonal and interannual shifts, improving understanding of ecosystem resilience. Additionally, applying NMDS to larger datasets or integrating it with predictive modelling could enhance its relevance for biodiversity assessments. By moving beyond simplistic biodiversity indices, NMDS offers a more nuanced perspective on species distributions, reinforcing the importance of multivariate approaches in ecological research and ecosystem monitoring

**References**

1. Ramette, A., 2007. Multivariate analyses in microbial ecology. *FEMS microbiology ecology*, *62*(2), pp.142-160.
2. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2024). \_vegan: Community Ecology Package\_. R package version 2.6-8, <https://CRAN.R-project.org/package=vegan>.
3. Batalla Salvarrey, A.V., Kotzian, C.B., Spies, M.R. and Braun, B., 2014. The influence of natural and anthropic environmental variables on the structure and spatial distribution along longitudinal gradient of macroinvertebrate communities in southern Brazilian streams. *Journal of insect Science*, *14*(1), p.13.
4. Innes, A. ESM wk 4 ‘Freshwater Fieldtrip’ – ‘Braid Burn – River Habitat Survey.pdf’ 2025.
5. Innes, A ESM wk 4 ‘Freshwater Fieldtrip’ – ‘Analysis of Ecological Communities in Vegan.pdf’ 2025.
6. Letovsky, E., Myers, I.E., Canepa, A. and McCabe, D.J., 2012. Differences between kick sampling techniques and short-term Hester-Dendy sampling for stream macroinvertebrates. *Bios*, *83*(2), pp.47-55.
7. Scottish Environmental Protetion Agency (2021) Can be accessed here - <https://www.gov.scot/publications/monitoring-watercourses-in-relation-to-onshore-wind-farm-developments-generic-monitoring-programme/>
8. Dexter, E., Rollwagen‐Bollens, G. and Bollens, S.M., 2018. The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography: Methods*, *16*(7), pp.434-443.
9. Boyra, A., Sanchez-Jerez, P., Tuya, F., Espino, F. and Haroun, R., 2004. Attraction of wild coastal fishes to an Atlantic subtropical cage fish farms, Gran Canaria, Canary Islands. *Environmental Biology of Fishes*, *70*, pp.393-401.
10. Vizzini, S. and Mazzola, A., 2004. Stable isotope evidence for the environmental impact of a land-based fish farm in the western Mediterranean. *Marine Pollution Bulletin*, *49*(1-2), pp.61-70.
11. Pedersen T (2024). ggforce: Accelerating 'ggplot2'. R package version 0.5.0, <https://github.com/thomasp85/ggforce>
12. Ter Braak, C.J. and Verdonschot, P.F., 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic sciences*, *57*, pp.255-289.