Appendix 1: Sampling considerations and effectiveness of eDNA in representing regional variation in the fish fauna.

1. Sampling location

Other than *Pseudorasbora parva* all fish species recorded were detected in samples collected from the shore (Fig. 1). Moreover, species on average had a higher occupancy and higher chance of detection in shore samples when offshore samples were also collected from the same lake (Fig. 2).

In general, it is reasonable to conclude that shore-based sampling will provide an adequate representation of the fish composition of a lake. Only samples collected from the shore were used in further analysis as all sites had samples taken from the shore, whereas only a subset had samples from both offshore and shore.

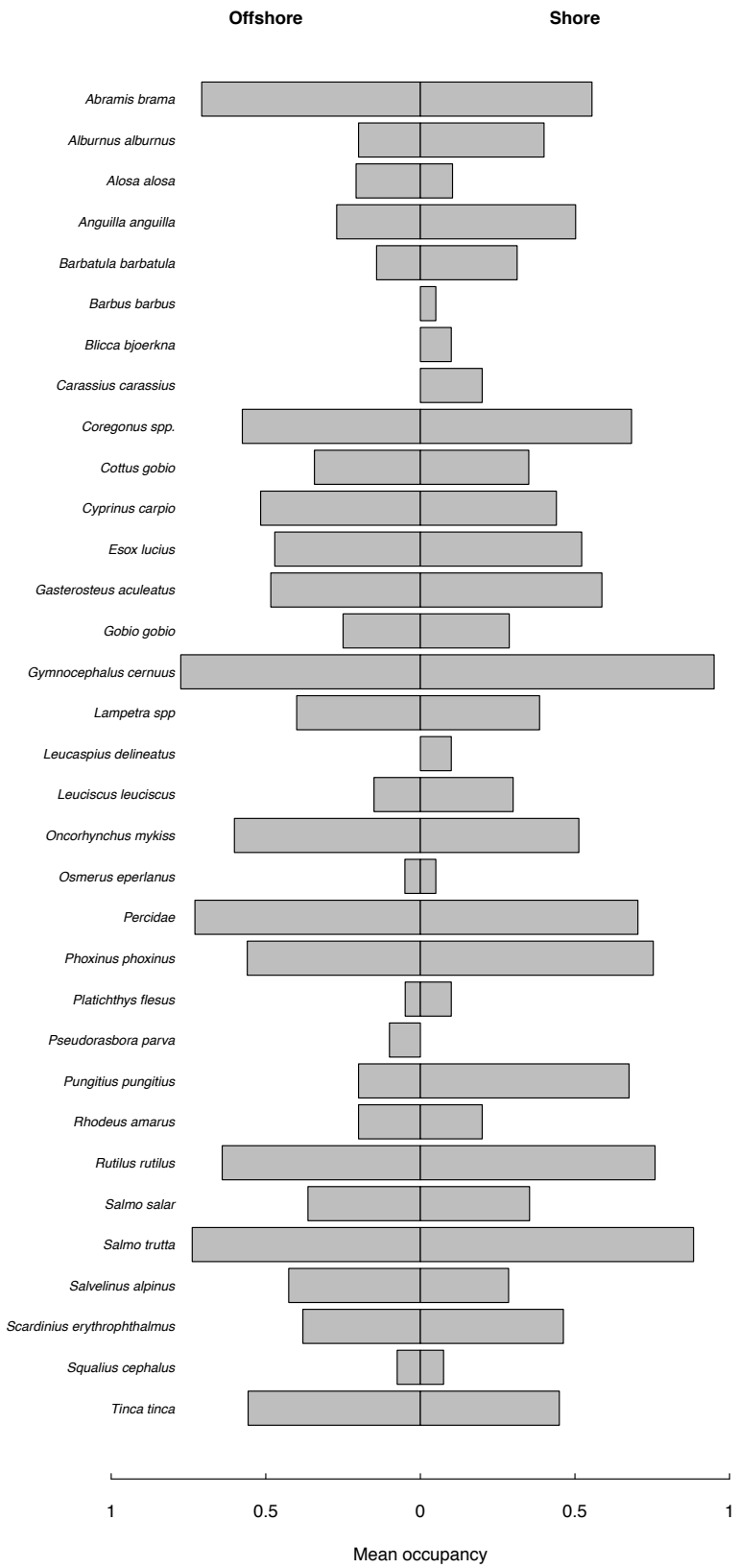

Fig. 1. The mean of species occupancy, averaged across those sites that had eDNA samples taken from both the shore and offshore (n=20).

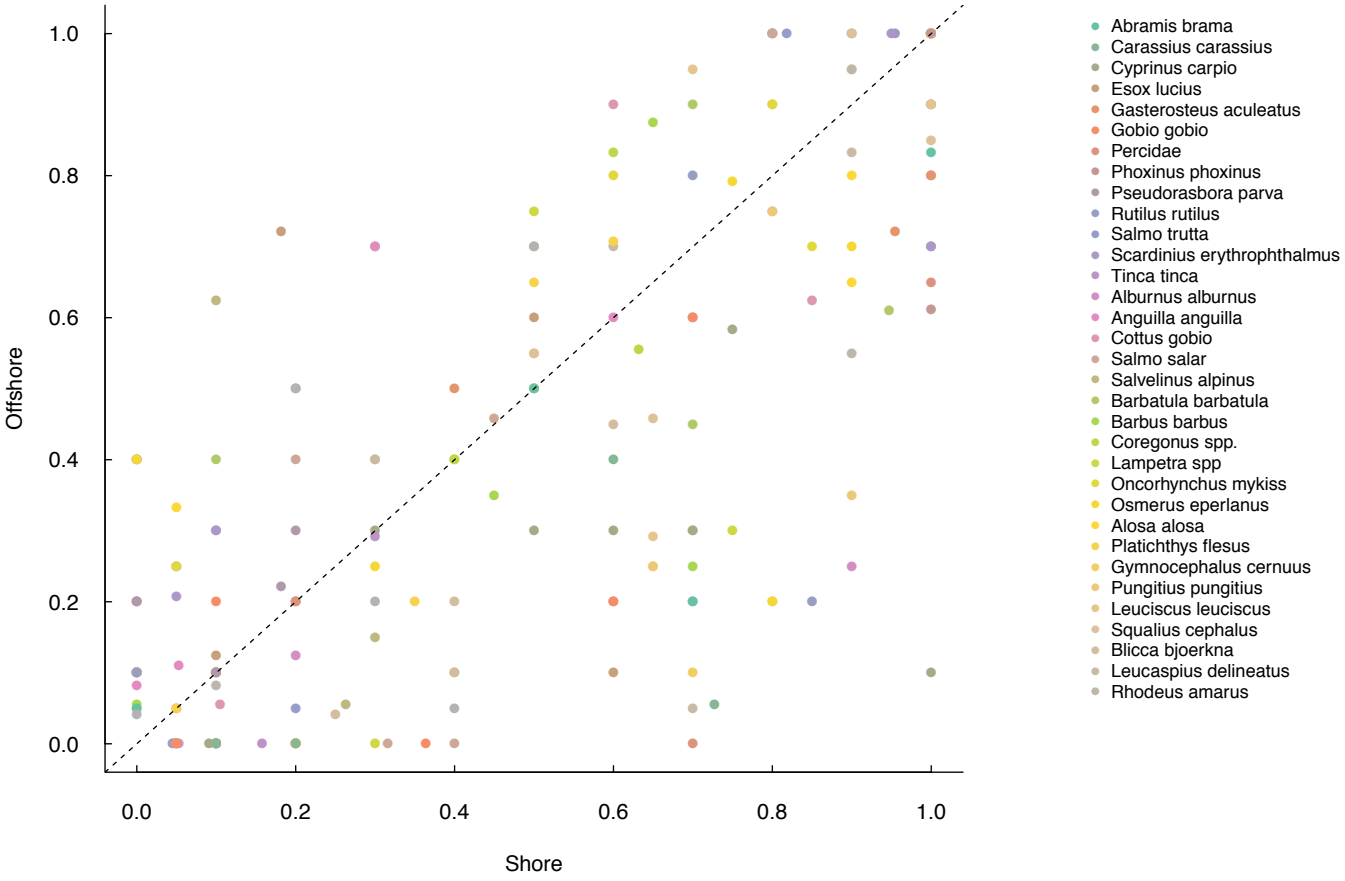

Fig. 2. Comparison of the species level occupancy across all sites with both offshore and shore samples for global dataset. The dashed line represents a 1:1 relationship.

2. Regional differentiation of the fish fauna

From Fig. 3 there is a clear distinction in the species composition found in samples from the Meres, South-East England and three Welsh sites (Kenfig Pool, Llan Bwch-llyn Lake & Llangorse Lake), for both abundance-weighted and presence only data. South West English lakes (Dozmary Pool, Little Sea, Slapton Ley and The Loe) displayed some overlap with North English lakes. Other Welsh lakes and the majority of those in North England clearly overlap in composition with Scottish lochs. The analysis demonstrates the effectiveness of the eDNA data in revealing major regional differences in the composition of the fish fauna.

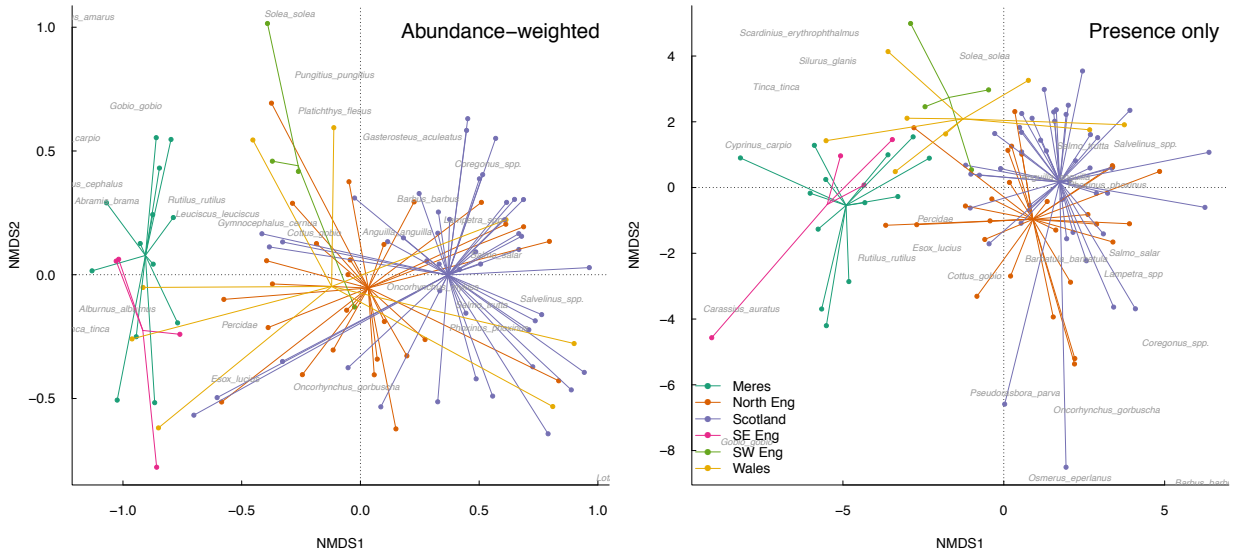

Fig. 3. Non-metric multidimensional scaling (NMDS) based on site occupancy. All regions and sites. Abundance is represented by each fishes occupancy (proportion of samples in which detected per site).

Due to the major compositional differences with the Meres, South East England, South West England and 3 Welsh lakes these were excluded from the analysis below. With the above-mentioned sites removed (n=24), there was a large overlap in between lakes from each region (Fig. 4), particularly when the abundance is considered (measured as the percentage of samples occupied). These patterns are indicative of similar fish species compositions across these regions, though some geographical variation in composition remains apparent.

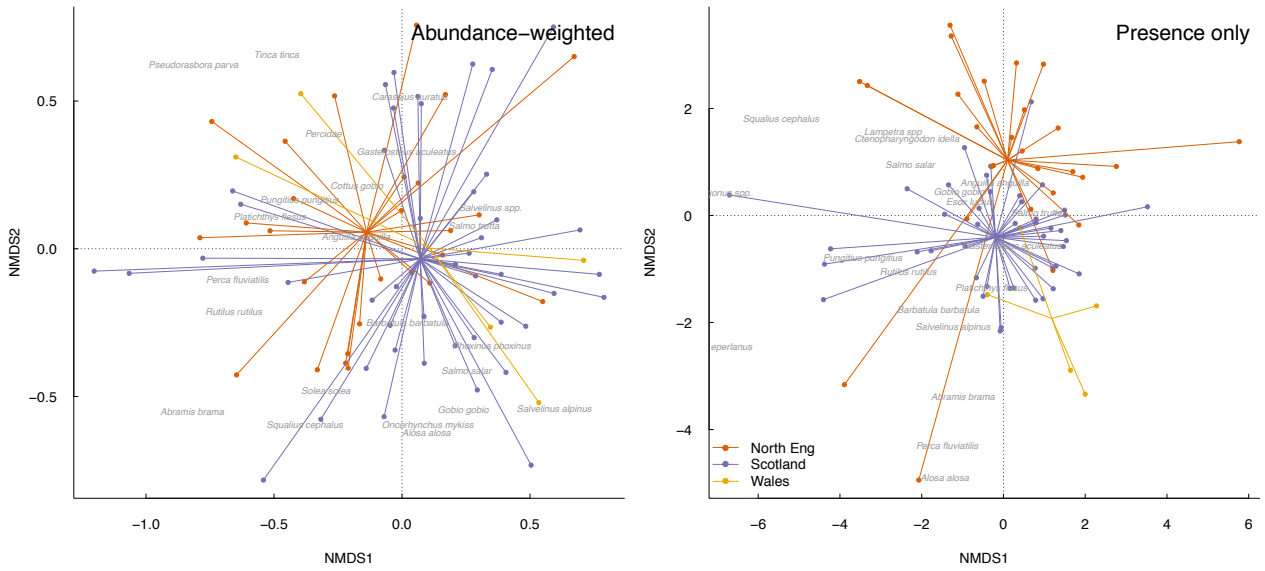

Fig. 4. Non-metric multidimensional scaling (NMDS) based on site occupancy. Selected northern region lakes (77 sites).

Fig. 5 displays a principal components analysis of the eDNA fish occurrence data per site, with and without potential false positives (based on expert judgement), to observe any natural groupings in fish communities regardless of region or environmental variables. There are general and predictable associations of fish taxa in both plots. On the lower right and upper left respectively, sites are characterised by coarse fish species characteristic of well vegetated lakes or littoral habitat e.g. roach, perch and pike. Whereas on the left and right side of each plot the pelagic/deeper-water salmonids dominate this grouping e.g. brown trout, salmon and arctic charr. Interestingly, minnow also contribute strongly to this latter grouping, suggesting their persistent presence in large deep-water lakes. The effect of false positives was therefore negligible in terms of affecting compositional groupings and explaining any additional variance.

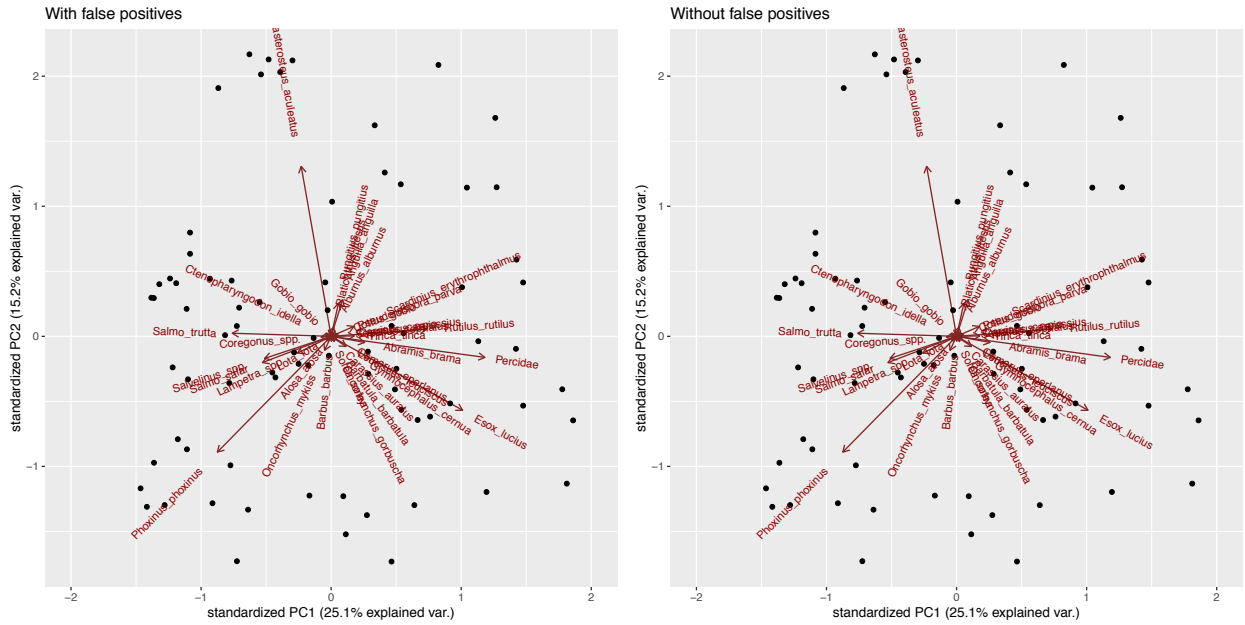

Fig. 5. Unconstrained ordination of the biological data using principal components analysis (PCA) (site no. = 79).

3. Adequacy of sampling

Based on the accumulation of species per lake, we can conclude that sampling within each region has been sufficient to detect the majority of species likely to be present in that region (Fig. 6), with the estimated sampling coverage consistently high (Table 1). Per lake, the majority of sites had samples taken at 20 locations. Judging by the trajectory of each sites' species accumulation curve the majority of species were detected within the first 10 samples (Fig. 7). Conducting species accumulation estimates for sites that had <12 samples (n = 9), we found that the estimated sample coverage was 91%, therefore lower than sites with 20 samples, but still capturing the majority of species present. Sites in which the extrapolated species richness was still rising strongly after 20 samples are very rare.

Table 1. Summary of observed and estimated species richness and sampling efficiency per region, excluding lakes in central and southern England. Includes potential false positives.

Group	No. of lakes	Mean richness per lake	Observed no. of species
All	77	8	35
North England	27	9	31
Scotland	45	7	29
Wales	5	7	15

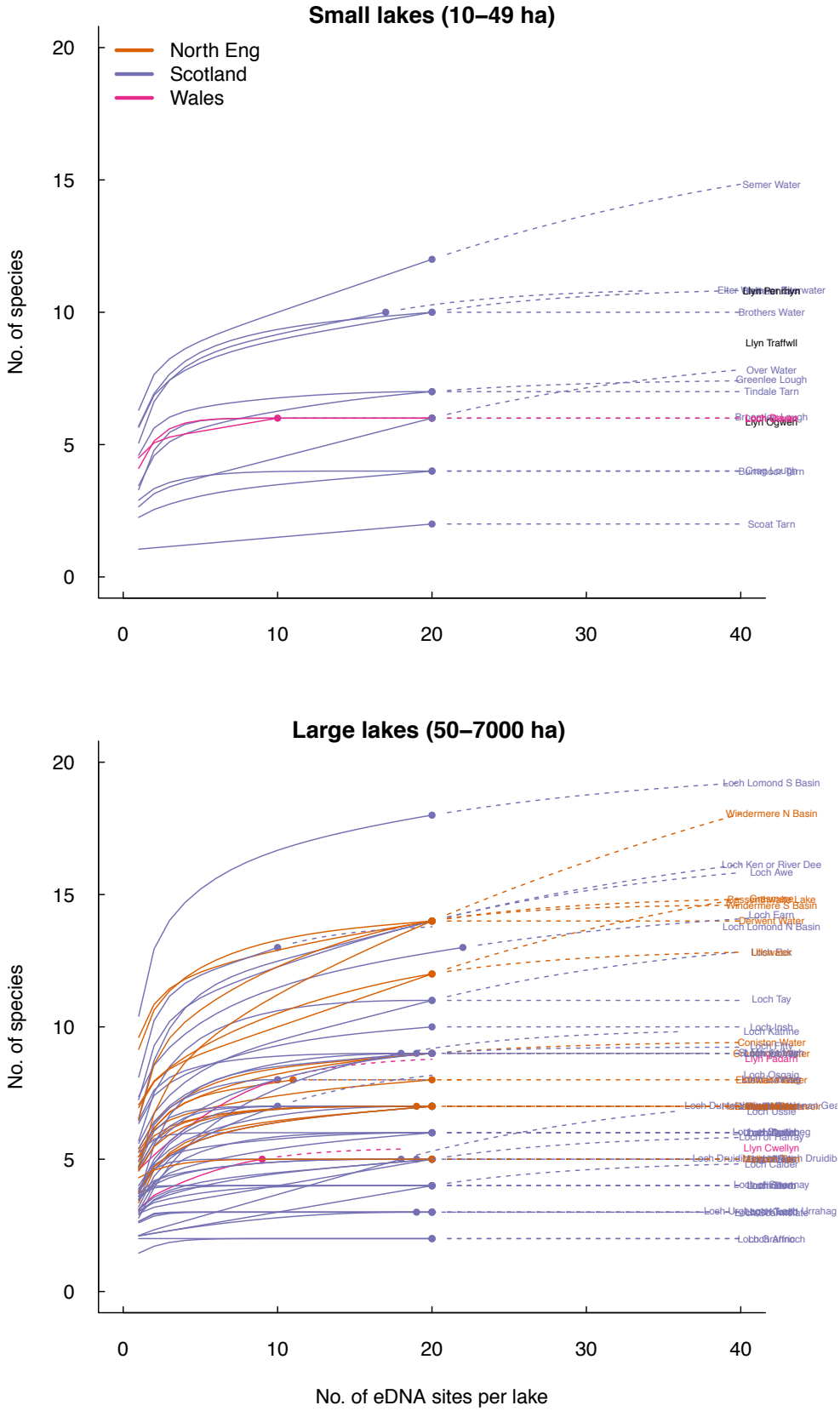

Fig. 6. Species accumulation per lake size category, based on species incidences per sample.