**3 | RESULTS**

Our goal here was not to make predictions on spatial distribution of the parasitic infection but rather to investigate how prevalence of black spot disease in fish communities is perceived across three scale levels (landscape, lake, site). First, we built landscape-level infection prevalence accumulation curves, then we compared the frequency distribution of the lake’s prevalence for each sampling method, and lastly, we identified predictors of site-scale prevalence patterns.

***3.1. Landscape-scale***

We used a resampling approach to compare how different sampling methods change the mean prevalence estimate through an increasing sampling effort across the landscape (Figure 3).

The estimate of landscape prevalence differed between the sampling methods, varying between 21% and 36%. After 35 samples, the transect method generated the highest mean prevalence (36 %) followed by the methods combination (31 %), the minnow trap method (24 %) and the seine method (21 %). However, the minnow trap curve did not stabilize after the 35 random samples suggesting that the prevalence value obtained is higher than the actual estimate (landscape infection prevalence measured by the minnow trap is 19%, see Table Sx). However, relatively few random samples are necessary to estimate a landscape prevalence for the transect method (approximately 10 samples) while an accurate prevalence estimate only occurs after 30 samples for the seine method. Even if the method combination curve (in gray) stabilizes around 20 samples, it nevertheless displays some variation around the curve because of the variability among the different methods used. While we were able to generate trend curves, some data points remain outside the confidence interval suggesting that extreme values a frequent. This is especially the case for lower sampling effort and for the methods combination. Each method curves overestimated the mean prevalence of low N values (number of samples) suggesting that infection prevalence is spatially heterogenous in our landscape with presence of infection hotspots.

***3.2. Lake-scale***

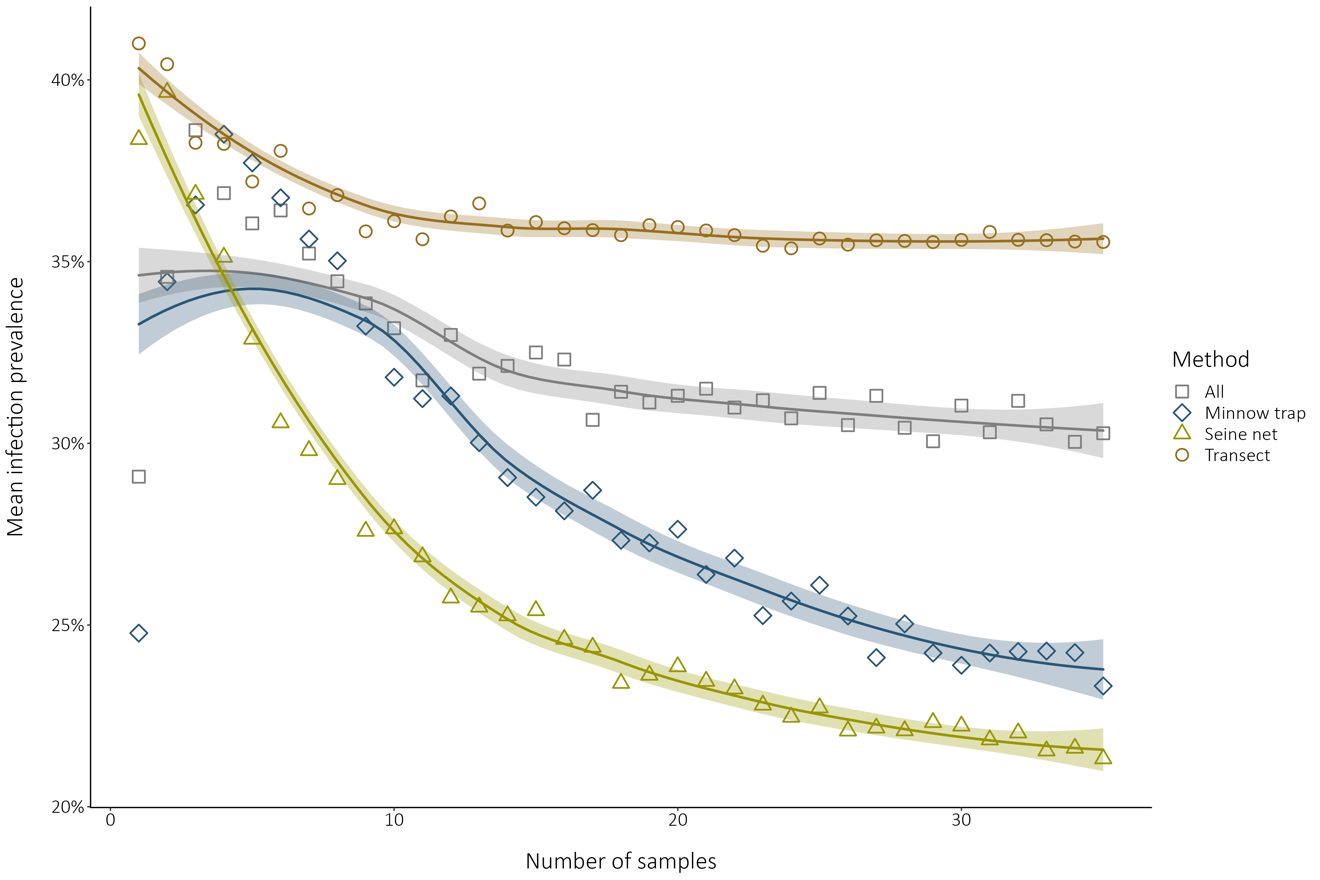
We compared the frequency distribution of the lake’s prevalence for each sampling method to see how the landscape prevalence is distributed among lakes (Figure 4). We then put the lake’s prevalence in their geographical context to inspect spatial patterns (Figure Sx).

Following the landscape results, the lake fish community prevalence estimates are not constant across the landscape. The lake prevalence frequency distributions for the combined methods and for the transect method show a bimodal distribution, with the landscape composed of many low-prevalence and high-prevalence lakes. However, neither method yielded prevalence estimates over 80%. The distribution patterns for the two fishing methods (seine net and minnow trap) are less clear. These methods show a right-skew distribution representing more low-prevalence lakes. Accordingly, seine and minnow traps captured very similar prevalence estimates at the landscape-scale (20.4% and 19.2% respectively). Heavily infected and less infected lakes do not appear to be clustered in space at the regional scale (Figure 2). Moreover, close and connected lakes do not appear to follow a spatial infection gradient, suggesting that geographic attributes (e.g., position in the landscape, connectivity, distance to nearest lake, belonging watershed) are not important drivers of the local infection prevalence (Figure 2). Both frequency distributions and maps support a difference between methods in terms of sampling fish communities and/or behavior.

***3.2. Site-scale***

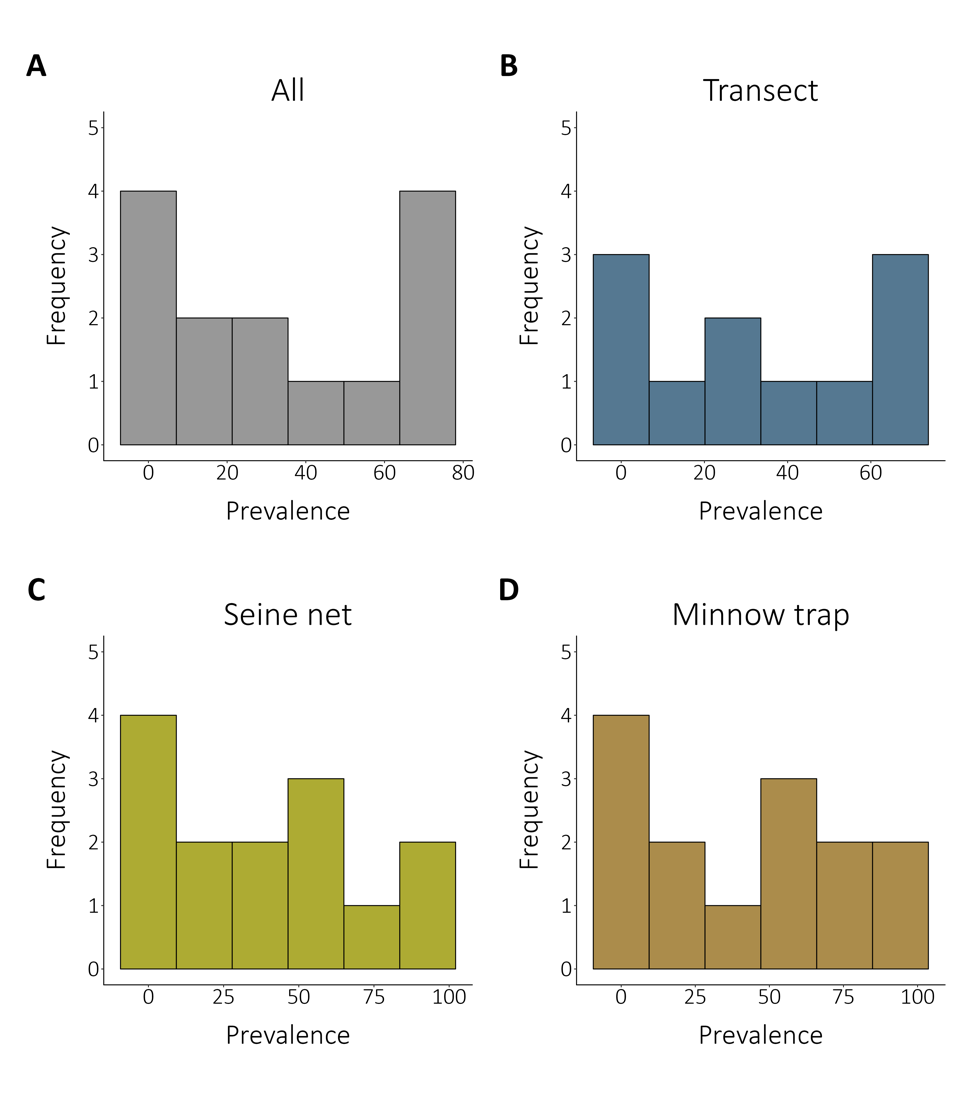
The relationships between the potential predictors and the site-scale prevalence (transect-level prevalence) were assessed with generalized additive mixed effects models.

The partial effects of environmental variables on prevalence are shown for all significant models in figure 5. The turbidity model has the best fit (D2 = 88,71%) (see Table S17) and is mostly non-linear. . The relation evidence a plateau for high-prevalence values, indicating that a prevalence saturation is reached for turbidity values above 2 NTU. That said, this model must be carefully interpreted as we sampled only a few high-turbidity sites. The relationship between TN:TP ratio and prevalence is unimodal but highly non-linear. For the macrophyte coverage, we found a decreasing relationship between the amount of macrophyte cover and the prevalence of infection, meaning that low macrophyte cover correlates with high prevalence of site-scale communities. Water temperature has an increasingly proportional relationship with infection prevalence. The prevalence estimate increased proportionally with pH, so that prevalence increases in more alkaline lakes. The same pattern is observed for the dissolved oxygen. However, we must take into consideration that the variation interval is very large for low-concentration oxygen values because of only three values below 7 mg/L were recorded. The conductivity and prevalence have a non-linear, unimodal relationship peaking around 80 (μS/cm). The relationship must also be carefully interpreted because of some gaps in the conductivity values (between 61.3 and 129.2 μS/cm). A parabolic curve is also observed in the Area:Perimeter model although, high ratio values are more uncommon increasing the variation interval. The relationship between the prevalence and the species diversity index is the only significant model related to the fish community structure per se and shows a decreasing trend. At the site-scale, fish communities are slightly to moderately diverse (Simpson’s diversity index between 0 and 0.64) indicating the dominance of some species. The models for TN, TP, TOC, lake area, maximum depth, mean depth, water residence time, drainage area, elevation, distance to the nearest lake, centrarchid abundance and species richness were not significant (see Table S17 for models’ summary values.

******

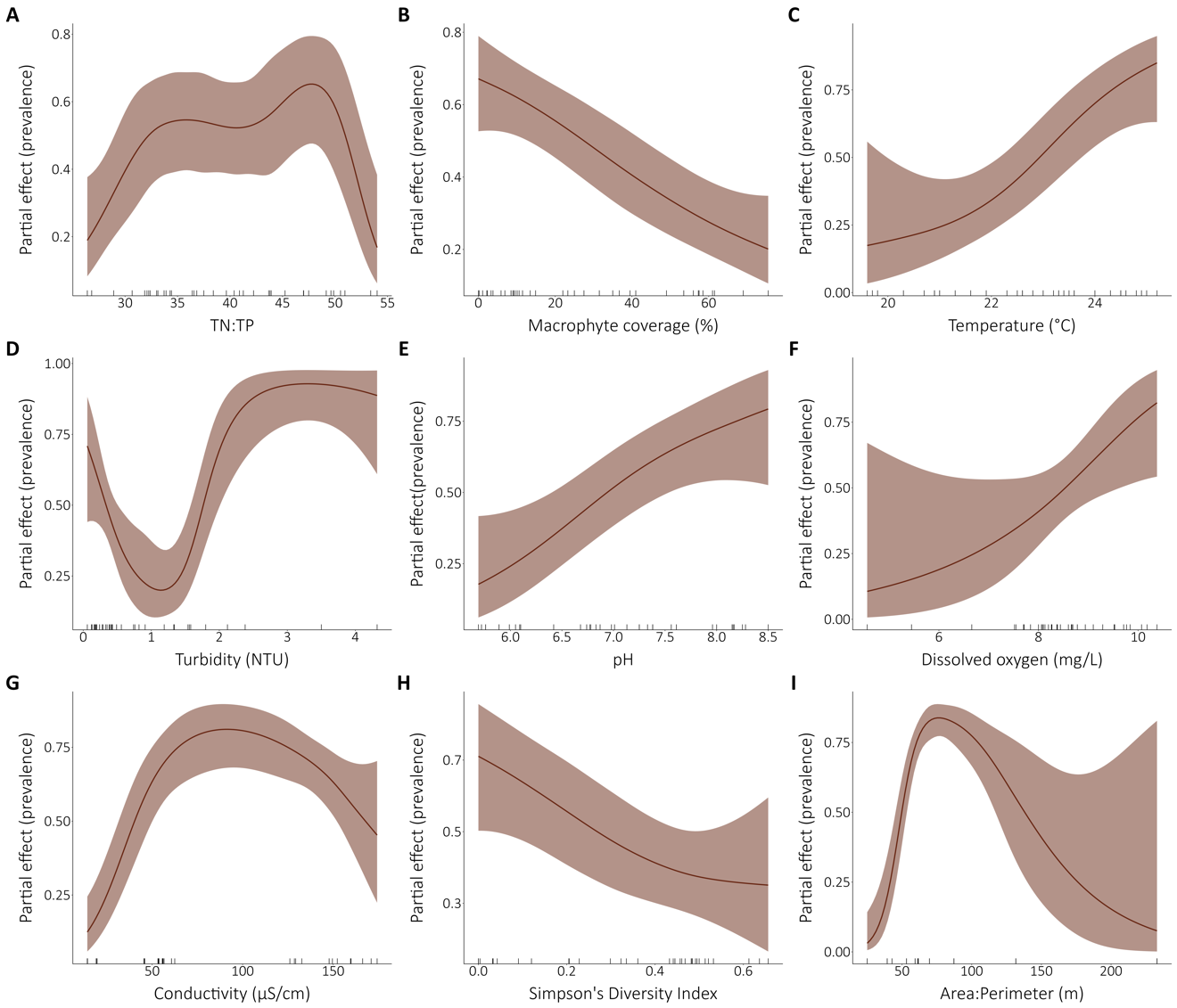
**FIGURE 3**

Simulated method comparison of landscape prevalence estimates sampling through an increasing random sampling effort. The data points indicate the mean prevalence for a given number of samples (N) randomly selected. The shadings indicate the 95% confidence interval on the loess smooth. The pool data used for the simulation with the minnow traps, the seine nets, the transects and all methods combined respectively contained 225, 75, 39 and 339 samples.



**FIGURE 4**

Comparison of the frequency distribution of the local community prevalence depending on the sampling method. All the frequency distributions considered 14 lakes, except for the transect method that survey 11 lakes. The figure shows frequency distribution among lakes for **(A)** all the sampling methods combined (weighted mean regional prevalence is 29.55 %), **(B)** the transect method only (weighted mean regional prevalence is 35.55 %), **(C)** the seine nets only (weighted mean regional prevalence is 20.44 %), **(D)** the minnow traps (weighted mean regional prevalence is 19.20 %).



**FIGURE 5**

Relationships between the fine-scale community infection prevalence and selected environmental drivers. All the models presented are univariate binomial generalized additive mixed models with the lake as a random effect on the intercept. The partial effects of the environmental variables on the prevalence are presented for the significant models. The ticks on the x-axes indicate a data point. **(A)** TN:TP ratio, **(B)** macrophyte coverage, **(C)** temperature, **(D)** turbidity, **(E)** pH, **(F)** dissolved oxygen, **(G)** conductivity and **(H)** Simpson’s Diversity Index are site-scale measurements while **(I)** area:perimeter ratio is a lake attribute. The perimeter model is not presented because it was highly non-linear and not interpretable although the model was significative (Figure S1).