

**Technical Report****Zero-altered modeling of an aquatic parasite host with application to invasive species risk assessments**Mark K. Taylor<sup>1,\*</sup>, Colby Whelan<sup>2</sup>, Carl J. Schwarz<sup>3</sup>, Patrick C. Hanington<sup>4</sup> and Leland J. Jackson<sup>2</sup><sup>1</sup>Parks Canada Agency, Banff National Park, Banff, AB, Canada<sup>2</sup>University of Calgary, Department of Biological Sciences, Calgary, AB, Canada<sup>3</sup>Simon Fraser University, Department of Statistics and Actuarial Science, Vancouver, BC, Canada<sup>4</sup>University of Alberta, School of Public Health, Edmonton, AB, Canada

\*Corresponding author

E-mail: [mark.taylor@pc.gc.ca](mailto:mark.taylor@pc.gc.ca)

**Citation:** Taylor MK, Whelan C, Schwarz CJ, Hanington PC, Jackson LJ (2022) Zero-altered modeling of an aquatic parasite host with application to invasive species risk assessments. *Management of Biological Invasions* 13(4): 845–862, <https://doi.org/10.3391/mbi.2022.13.4.17>

**Received:** 8 March 2022**Accepted:** 12 July 2022**Published:** 31 October 2022**Thematic editor:** Matthew Barnes**Copyright:** © Taylor et al.

This is an open access article distributed under terms of the Creative Commons Attribution License ([Attribution 4.0 International - CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

**OPEN ACCESS****Abstract**

A fundamental consideration in aquatic invasive species risk management is the distribution of invasive taxa relative to the risk assessment area. However, sampling the distribution of aquatic invasive species (AIS) is costly and time consuming, especially when they are rare or clustered. While random sampling removes bias in most estimates, it could lead to many zeroes in the response variable if the target species is not evenly distributed. We surveyed the distribution of *Tubifex tubifex*, the obligate secondary host of *Myxobolus cerebralis*, an invasive parasite that causes whirling disease in salmonids. We used intensive grid and random sampling in a lake and two river systems but collected many samples with zero *T. tubifex*. Zero-altered models identified a single source of zero *T. tubifex* counts in each dataset and improved model validation compared to Poisson or negative binomial models. While *T. tubifex* counts were not predictable using the covariates we measured, the binomial processes identified zero odds of a *T. tubifex* occurrence at lake sites with water depths > 2 m and stream sites with channel slopes > 3%. These covariates could be used to stratify the landscape for future sampling, which would save survey time and resources and likely reduce uncertainty in parameter estimates. Model-based sample stratification could address some of the challenges with AIS surveys when species are at low abundance or are clustered.

**Key words:** *Tubifex*, distribution model, whirling disease, zero-inflation**Introduction**

Freshwater ecosystems are among the most endangered ecosystems globally and face increasing threat from aquatic invasive species (AIS; Dudgeon et al. 2006). Knowledge of the distribution and habitat suitability of AIS can be used to inform risk management and prevention strategies (Mandrak and Cudmore 2015; Rodríguez-Rey et al. 2019). Species distribution models have been widely used to guide management of AIS (e.g., Papes et al. 2016). These correlative techniques allow conservation practitioners to model the distribution of a species and map the spatial suitability of its habitat based on the statistical association between the species' occurrence (or abundance) and predictor variables (Elith and Leathwick 2009).

Traditionally, some form of random or systematic sampling has been employed to avoid sampling bias for distribution models (e.g., Yoccoz et al. 2001). However, AIS can have rare and/or patchy distributions because they are at non-equilibrium with their environment and simple random or systematic sampling can lead to few observations (Guisan et al. 2006; Hattab et al. 2017; Trochine et al. 2018). The challenge of modeling the distribution of AIS has been a topic of interest in the literature providing guidance on model bias, accuracy and validation (Rodríguez-Rey et al. 2019; Hattab et al. 2017); however, there is a lack of guidance regarding modeling zero AIS counts. Therefore, practical case studies that include analyses of AIS distributions with zero-inflated responses could support conservation practitioners with AIS risk management.

*Myxobolus cerebralis* (*Mc*; Hofer, 1903) is an invasive species to North America that has spread from its native habitat in Europe (Hoffman 1990). *Mc* is an obligate, two-host myxozoan parasite responsible for whirling disease (WD) in salmonids (Wolf and Markiw 1984). Fisheries managers perform risk assessments on whirling disease because it has impacted some salmonid populations (Nehring and Walker 1996; Vincent 1996). For an invasive parasite, the distribution of its host is predictive of the likelihood of invasion (Arneberg et al. 1998). Indeed, the presence of *Tubifex tubifex* (Müller, 1773), the intermediate host of *Mc*, is the most important risk factor for the development of WD in free-ranging salmonids (Hallett et al. 2009; Nehring et al. 2014). Therefore, an understanding of *T. tubifex* distributions would help guide waterbody-specific and regional prevention strategies.

*Tubifex tubifex* is an oligochaete that inhabits the substrate of temperate waterbodies (Anlauf and Moffitt 2008). While *T. tubifex* is a cosmopolitan species, its distribution is not ubiquitous across salmonid habitat in North America. At the microhabitat and habitat unit scale in rivers, *T. tubifex* have consistently been associated with the presence of fine sediment (Anlauf and Moffitt 2008) and leaf litter (Lazim and Learner 1987). These sediment properties are associated with low gradient river channels at the reach and segment scale (Anlauf and Moffitt 2008; Ayre et al. 2014). At broader scales, *T. tubifex* is found in habitat associated with agriculture (Anlauf and Moffitt 2010), nutrient enrichment (Kaeser and Sharpe 2006), water reservoirs (Zendt and Bergersen 2000; Nehring et al. 2013) and trout hatchery rearing ponds (Allen and Bergersen 2002; Bartholomew et al. 2007). A substantial proportion of the freshwater oligochaete population throughout the eastern slopes of the Rocky Mountains, in Alberta, Canada, is *T. tubifex* (Barry et al. 2021). Considerable regional differences in *T. tubifex* occurrences have been described within Alberta (Barry et al. 2021) and from other countries including the complete absence of *T. tubifex* in some other regions (Arsan et al. 2007). Therefore, there is a critical need for stream and lake surveys in western Canada to develop distribution models that predict the occurrence and abundance of *T. tubifex*. These models are needed to predict the risk of

whirling disease across broad landscapes that include species at risk such as *Oncorhynchus clarkii lewisi* (Richardson, 1836) and *Salvelinus confluentus* (Suckley, 1859) (Fisheries and Oceans Canada 2019, 2020).

The small body size, sessile lifestyle and sub-benthic habitat of *T. tubifex* might make them difficult to sample. Low-efficiency sampling methods coupled with low population densities can result in low detectability (Royle and Nichols 2003). Even within suitable habitat, a burrowing invertebrate like *T. tubifex* can have a patchy distribution because significant horizontal and vertical heterogeneities in microhabitat occur in lake substrata (Hutchinson 1957) that create an uneven distribution of benthic invertebrates. Thus, it is possible that a large number of samples might be devoid of the target organism (i.e., zeros), which leads to data that do not fit commonly used unimodal distributions (e.g., Poisson, quasi-Poisson, and negative-binomial).

Here, our purpose was to demonstrate that random and systematic allocation of sample sites can sometimes lead to excess zeros using case studies that are relevant to AIS risk management. Excessive zeroes can cause bias and uncertainty in model predictions. We aimed to also demonstrated how using zero altered models, which explicitly accommodate zero responses, can reduce uncertainty in model parameter estimates and provide covariates that can be used to stratify the landscape to make future sampling more efficient.

Our first objective was to model *T. tubifex* counts from samples collected in one lake and two contrasting fifth-order watershed stream systems. Our second objective was to determine environmental co-variates that explain observed *T. tubifex* patterns of abundance. While five *T. tubifex* lineages (I, III, IV, V, and VI) are known to occur in North America, only lineage III *T. tubifex* propagate high levels of *Mc* infection (Beauchamp et al. 2005; Nehring et al. 2014). Therefore, our third objective was to identify *T. tubifex* lineages and the proportion of *T. tubifex* infected with *Mc* in our study area. Finally, we briefly outlined a general procedure to model count data. Detailed reviews on modeling zero-inflated count data exist (e.g., Cunningham and Lindenmayer 2005; Martin et al. 2005; Blasco-Moreno et al. 2019); however, we provide a straightforward introduction to modeling excess zero counts of an AIS with emphasis on model selection and validation for conservation practitioners.

## General approach to count data modeling

Modelling count data is a multi-step process not unlike modeling Gaussian data using linear regression. The first step is data exploration. Outliers in the response and explanatory variables can be identified using histograms and Cleveland dotplots (Zuur et al. 2009). Pairplots are used to examine collinearity of explanatory variables because multi-collinearity causes problems in model fitting (e.g., suppression of marginal effects). Habitat covariates

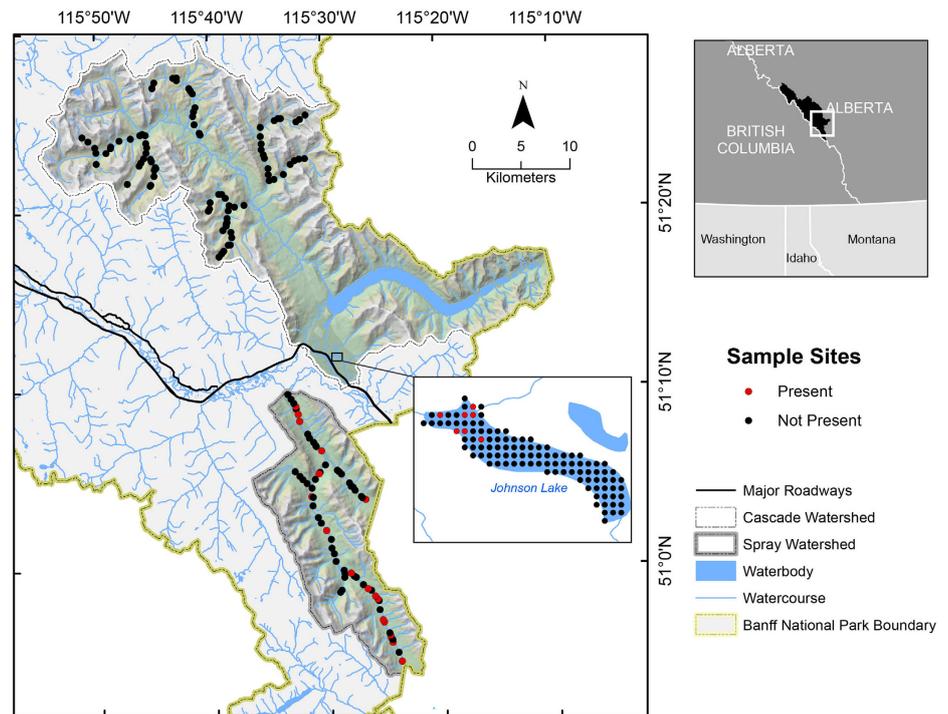
may need to be removed when they were correlated with one another; judgements on which variables to remove were based on the objectives of the study and past literature.

The standard model for count data is the generalized linear model (GLM; McCullagh and Nelder 1989) with a Poisson distribution. Akaike information criterion (AIC; Akaike 1974) can be used for Poisson model selection similar to linear regression (Zuur et al. 2009). Count data frequently exhibit two characteristics: (1) a variance significantly larger than the mean, known as *overdispersion* (Bliss and Fisher 1953); (2) an excess of zero values in comparison to standard distributions such as normal, Poisson, binomial and negative-binomial (herein: *zero-inflation*; Blasco-Moreno et al. 2019). These two phenomena are related to each other because zero-inflation also contributes to the data overdispersion (Blasco-Moreno et al. 2019). Ignoring overdispersion and zero inflation can cause biased parameter estimates and biased measures of uncertainty, resulting in misinterpretations of the species-environment relationship (Lambert 1992; MacKenzie et al. 2002). Given this, an important part of model selection is to examine evidence for overdispersion in models by looking at the dispersion parameter (Null deviance/degrees of freedom; Zuur et al. 2009).

In the absence of overdispersion, Poisson models may provide an adequate fit. Some amount of overdispersion can be modelled with a quasi-Poisson model or a negative binomial model (Cunningham and Lindenmayer 2005). However, quasi-Poisson and negative binomial models will not improve the dispersion parameter when overdispersion is caused by zero-inflation (Zuur et al. 2009). Model validation consisted of looking for trends in the deviance residuals versus fitted values. In addition, deviance residuals are plotted versus each covariate in the model and covariates dropped from the model (Zuur et al. 2009).

An alternate approach for data with an excess number of zeros are zero-inflation models (ZIP), zero-altered models (ZAP) or their corresponding negative-binomial versions (i.e. ZIPNB and ZAPNB). ZIP models are called mixture models because the zeros are modelled as coming from two different processes: the binomial-process from which only presence/absence is observed and a Poisson process in which counts and a proportion of the zero values, appropriate to the Poisson distribution, are observed (Zuur et al. 2009). ZAP models are called two-part models (or hurdle models), whereby the first part is a binomial process and the second part is a truncated Poisson count model (Cunningham and Lindenmayer 2005). ZAP models treat all zeros together and assume that zeros arise from a single process. Both ZIP and ZAP models produce two sets of parameter estimates – the probability that a species is present in a sample and then given it is present, the relative mean number of individuals per unit of sample.

The choice between ZIP or ZAP should be based on a priori knowledge of the cause of the zero-inflation (Zuur et al. 2009). ZIP models can be used



**Figure 1.** Map of the study area including the Spray watershed, Cascade watersheds and Johnson Lake (inset), Banff National Park, AB.

if more than one source of zeros exists regardless of the sources of zeros (e.g., false zeros). ZAP models are commonly used to model zero-inflated data when only one source of zeros exists (e.g., structural zeros; Lambert 1992; Greene and Young 1993) or when one cannot differentiate between structural and false zeros (Zuur et al. 2009). AICs can be used to compare nested models with different candidate covariates (Zuur et al. 2009). Model validation requires plotting Pearson residuals against the fitted values, against each explanatory variable in the model and against each explanatory variables that were removed from the model to ensure no patterns (Zuur et al 2009).

### Case Study 1 – Johnson Lake

Johnson Lake is a 16-hectare waterbody in montane forests of Banff National Park (BNP; UTM: 11U 605864E x 5672800N; Figure 1). BNP encompasses 6,641 square kilometres of mountainous landscape with numerous glaciers and icefields, dense coniferous forests and alpine landscapes. Johnson Lake was previously a wetland and was impounded in the 1930s to raise the water level 6 m to create a larger waterbody for recreational opportunities. The lake bottom is primarily fine sediment with patches of *Chara* sp., that forms thick mats. Johnson Lake supported a population of native *Catostomus commersonii* (Lacépède) and *Couesius plumbeus* (Agassiz, 1850) before rotenone treatment in 1967 (Ward 1974). Non-native *Salvelinus fontinalis* (Mitchill, 1814) and *Oncorhynchus mykiss* (Walbaum, 1792) were stocked immediately following the rotenone treatment (Ward 1974). Nonnative

*Salmo trutta* (Linnaeus, 1758) appeared in the lake more recently and are likely the result of illegal stocking. All of the nonnative trout were removed over the years 2017–2020 to remove the salmonid host and eradicate whirling disease.

A site was defined as a 1 m × 1 m cell spaced 40 m × 40 m apart from one another in a grid across Johnson Lake for a total of 107 possible cells. Samples were systematically sampled from all 107 sites while the lake surface was frozen between December 2016 and March 2017. A mini-Glew suction corer with an internal diameter of 3.8 cm (Glew 1991) was used to collect five replicate sediment cores within each site, for a total of 535 cores. Each replicate core was sampled adjacent to one another randomly in the 1 m × 1 m cell. Depth was measured with a graduated plumb line. The top 10 cm of each core was extruded and sliced longitudinally to create two equal subsamples. One half was searched for *T. tubifex* following a standardized protocol and the other was frozen for later analyses of sediment organic carbon and inorganic carbon composition via loss on ignition (LOI) at 550 °C and 990 °C, respectively (Heiri et al. 2000). All sediment from all samples could be washed through a 400 µm sieve which indicated almost all sediment would be size-classified as medium sand and smaller; *T. tubifex* don't discriminate between categories smaller than medium sand (Lazim and Learner 1987).

All samples were collected in the morning and searched for oligochaetes in the afternoon. By searching unpreserved samples, we were better able to identify potential *T. tubifex* by their movement. Samples were washed in a 400 µm sieve to remove fine sediment. Large debris was also removed. The remainder was transferred to a white tray and diluted with clean water. An illuminated magnifying glass was used to examine the sample and oligochaetes were removed and subsequently viewed under a light microscope. Any oligochaete that featured chaetal hairs or bifid chaetae (Brinkhurst 1986) was preserved in 70% ethanol and sent to the University of Alberta Molecular Biology Service Unit for genetic species and lineage confirmation. Every tenth sample was sorted a second time by a different researcher for quality assurance. DNA was analyzed following extraction from samples and amplification by qPCR following techniques described in Nehring et al. (2013). Lineages were analyzed following Beauchamp et al. (2005). *Tubifex tubifex* were also tested to determine if they were infected with *Mc* using a HSP70 adapted from Cavender et al. (2004).

We modeled the total number of *T. tubifex* from five sediment cores at each site following the general approach to count data modeling above. However, we used the offset command to standardize the response variable by the surface area of the five cores per site (0.00285 m<sup>2</sup>) thus modeling the density (total *T. tubifex*/m<sup>2</sup>). The surface area of individual sediment samples collected (0.00113 m<sup>2</sup>) was divided by two to account for the removal of one half of the core for sediment analysis (0.00057 m<sup>2</sup>). We consistently

**Table 1.** Model selection results for zero-altered model (ZAP) of *T. tubifex* densities in Johnson Lake, Banff National Park, Alberta, using Akaike information criterion (AIC), delta AIC ( $\Delta$ AIC) and AIC weights ( $AIC_w$ ).  $n = 107$ . Vertical lines between parameters separate Poisson process covariates from binomial process covariates. Candidate models contained the following covariates: water depth (depth) and organic carbon (OC).

Rank	Parameters	AIC	$\Delta$ AIC	$AIC_w$
1	1   depth	59.76	0	0.45
2	OC   depth	60.45	0.69	0.32
3	OC   depth + OC	62.17	2.41	0.13
4	depth + OC   depth + OC	62.63	2.87	0.11

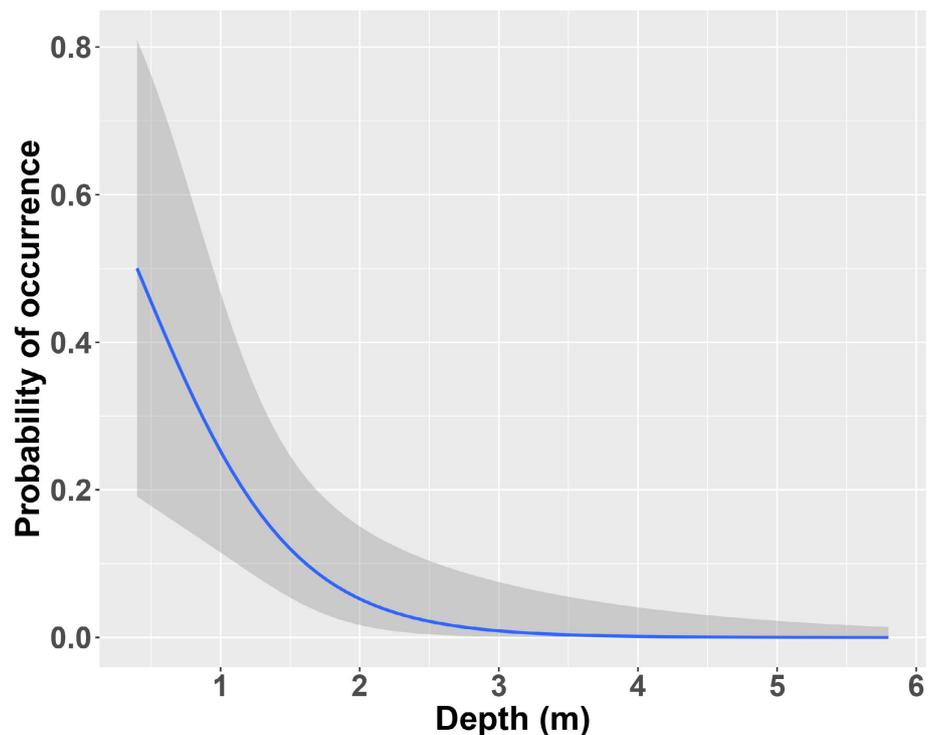
searched a 10 cm length of core at each site, so surface area rather than volume was used to be comparable to other published densities of *T. tubifex* (e.g., Zandt and Bergersen 2000). Candidate covariates for a *T. tubifex* count model included water depth (m), percent organic carbon (%; organic carbon) and percent inorganic carbon (%; inorganic carbon). No habitat covariate outliers in the data were found, but inorganic carbon was dropped from the analysis because it was inversely correlated with organic carbon. The analysis was completed using the open source program R (R Core Team 2021) and R package “pscl” (Jackman 2020).

Our intensive sampling revealed that there was a low density and patchy distribution of *T. tubifex* in Johnson Lake (Figure 1). Of the 107 sites sampled, *T. tubifex* were found only at seven sites (Supplementary material Table S1). There was a significant negative relationship between water depth and *T. tubifex* density in the Poisson model ( $P < 0.001$ ); however, the Poisson model was underdispersed (dispersion parameter = 0.64) which can occur when species counts are clustered (Zuur and Ieno 2012). Furthermore, the deviance residuals were cone-shaped when plotted vs. water depth suggesting poor model validation (Zuur et al. 2009). Negative binomial models are generally useful for overdispersion rather than underdispersion; therefore, we continued with ZAP models because of the knowledge of the large number of zeros.

We included water depth and organic carbon as covariates in both the truncated Poisson and binomial portions of the ZAP model (see Table 1). We did not report ZIP models because separating two sources of zeros brought unnecessary complexity to the modeling and did not affect our conclusions (e.g., Kuhn et al. 2016). Based on AICs, the final ZAP model had only water depth as a covariate in the binomial process and the intercept in the Poisson process (Table 1). For every 1 m increase in water depth at a given site, the results indicated a 17% decrease in the odds of finding a *T. tubifex* (Table 2; Figure 2). No *T. tubifex* were found in water deeper than 2 m. A ZAPNB model was not a better fit according to AICs. Water depth is likely just a proxy for sediment characteristics such as organic carbon and inorganic carbon content. We chose water depth as a covariate because it is easy to measure, and bathymetric maps are commonly available for water bodies. Model validation showed no patterns in Pearson residuals

**Table 2.** Parameter estimates and their standard errors for the zero-altered model (ZAP) of *T. tubifex* densities in Johnson Lake, Banff National Park, Alberta ( $n = 107$ ).

Process	Parameters	Estimates	Std. error	Z value	P value
Poisson	Intercept	6.213	0.377	16.48	< 0.0001
Binomial	Intercept	0.725	0.951	0.762	0.446
	Depth	-1.811	0.626	-2.89	0.003



**Figure 2.** The relationship between water depth and the probability of occurrence of *T. tubifex* in Johnson Lake, Banff National Park, Alberta. The shaded area is the 95% confidence interval.

which suggested that the ZAP model fit the data better than a standard unimodal Poisson model.

Of the sites where *T. tubifex* occurred in the sample, their predicted mean density was 493 ( $\pm$  S.E. = 1.46) individuals/m<sup>2</sup>. While the density of *T. tubifex* in lakes has not been reported frequently, the mean density of *T. tubifex* at occupied sites in Johnson Lake was much less than the density of *T. tubifex* from studies at the Windy Gap reservoir, in Colorado, U.S.A. Windy Gap Reservoir is known to have a high density population of *T. tubifex* ( $> 10^4$  individuals/m<sup>2</sup>) and has been implicated as a source of whirling disease for the Colorado River (Zendt and Bergersen 2000).

*Tubifex tubifex* in Johnson Lake were identified genetically as lineage III which is known to propagate the most triactinomyxon spores per myxospore ingested (Beauchamp et al. 2005) and is associated with the worst outbreaks of *M. cerebralis* among sympatric fish populations (Stevens et al. 2001). Homogenous lineage III *T. tubifex* populations are not common, yet have been found in Yellowstone National Park (Alexander et al. 2011) and the Deschutes River in Oregon (Zielinski et al. 2011). None of the *T. tubifex* samples tested positive for *Mc*. An individual *T. tubifex* that was positive

for *Mc* was identified during pilot sampling, but not from the main sampling effort. It is common for infection rates to be very low in all waterbodies (~ 1%; Zendt and Bergersen 2000) and the absence of an infected *T. tubifex* could be a result of the low number of *T. tubifex* captured. Despite the fact that none of the *T. tubifex* evaluated for *Mc* tested positive, all *Salvelinus fontinalis* sampled from Johnson Lake ( $n = 183$ ), tested in batches of 5 individuals, were positive for *Mc*. This supports the idea that a high prevalence of whirling disease can manifest from a low prevalence of the parasite.

## Case Study 2 – Spray and Cascade watersheds

The mainstem of the Spray watershed is regulated by a series of ponds, dykes and a dam associated with the Spray Reservoir. Since 1949 the infrastructure had significantly reduced stream flows in the Spray River mainstem and Goat Creek. The Spray River flows north and enters the Bow River as a fifth-order river after 35 kms downstream of the dam outlet. In contrast, the Cascade watershed is higher elevation and has a natural flow regime. The Spray and Cascade watersheds support native *Salvelinus confluentus* and non-native *Salvelinus fontinalis* populations. Native *Oncorhynchus clarkii lewisi* occur in the Cascade watershed but have been functionally extirpated from the Spray watershed below Canyon Dam (*unpublished data*).

Sampling reaches were distributed by random tessellation (Stevens and Olsen 2004) at a density of ~1 reach/km in both watersheds ( $n = 67$  in the Spray watershed;  $N = 86$  in the Cascade Watershed). Sampling was limited to first, second, and third-order reaches (Figure 1). The collection occurred in autumn of 2017 and 2018. Reaches within segments with a gradient greater than 15% were eliminated because they exceeded the maximum gradient where salmonids could reside. Furthermore, reaches in areas known to be above fish barriers were also eliminated as whirling disease cannot propagate to areas without both a fish and *T. tubifex* host present.

Three sites were sampled per reach at 0 m, 50 m, and 100 m from the pre-determined reach start location. Crews collected sediment at a transect, perpendicular to the stream bank, at each of the three sites per reach. The lateral position on the transect that contained the finest sediment at each site was sampled. An area of approximately 1 m<sup>2</sup> was sampled with a kick net (30 cm × 30 cm triangular opening, 400 μm mesh) for 3 minutes at each site. Crews kicked consistently to standardize effort. Samples were stored in plastic bags and preserved with 70% ethanol immediately after collection. Samples were searched in the lab as per the Johnson Lake case study.

At each site, an underwater camera (*Olympus TG-5, 4000×3000 pixels*) was used to capture images of the stream bottom, at 2 m intervals across the width of the site transect following the protocol in Turley et al. (2016).

Photos were imported into Photoshop (*Adobe Photoshop Ver. 19.0*) and overlaid with a grid of 100 squares. Squares where the majority of sediment was under 2 mm were counted using a standardized approach outlined in Turley et al. (2016) to get a measure of fine sediment (%) because previous studies have shown that *T. tubifex* abundance is associated with fine sediment (Anlauf and Moffitt 2008).

Slope (%) and contributing area (ha) were calculated from a hydroline network generated from 30 m digital elevation models (DEM). A segment was measured from stream node to stream node. Point values for each stream segment were extracted from the DEM raster to determine elevation drop over the segment. These were compared to the length of each segment using the ArcGIS 3D analyst tool to calculate slope. We excluded ephemeral streams by removing 1<sup>st</sup> order hydrolines that were tributaries of third-order and higher streams. Historically, these streams are dry in late summer. The contributing area was calculated for each reach with the ArcGIS Accumulate Values Downstream tool from the STARS toolset (Peterson and Ver Hoef 2014). Nodes were generated at each stream junction and the accumulated area was calculated at each node based on hydroline and DEM models. The most downstream node that was still above each sample point was taken as the value for watershed contributing area. The contributing area for the Spray watershed was set at a relative mark of zero for both nodes immediately below the Canyon Dam and the Goat Creek reservoir assuming that upstream surface runoff was captured by the reservoirs rather than the downstream lotic environment.

We modeled the total number of *T. tubifex* from three sites within each reach in the Spray watershed following the general approach to count data modeling (above). However, we used an offset command to standardize the response variable by the surface area of the three kick areas per reach ( $3 \times 1 \text{ m}^2$ ). Therefore, we modeled a density of *T. tubifex* (total *T. tubifex*/m<sup>2</sup>). Candidate covariates included the landscape-scale variables, contributing area and slope. We also considered a third variable, fine sediment that was an average of three measurements per reach (0 m, 50 m, and 100 m). The analysis was completed using R (R Core Team 2021) and the R package “pscl” (Jackman 2020).

Our sampling resulted in no *T. tubifex* from any of the reaches in the Cascade watershed (N = 86) and a patchy distribution and low density of *T. tubifex* in reaches of the Spray watershed (N = 67; Figure 1). Of the reaches sampled in the Spray watershed, *T. tubifex* were found only at 17 reaches (Table S2). Slope, contributing area and % fine sediment were all statistically significant in the Poisson model ( $P < 0.001$ ) describing *T. tubifex* densities in the Spray watershed. However, the model was overdispersed (dispersion parameter = 4.3). In addition to overdispersion, the relationship between deviance residuals and the best covariate was cone-shaped suggesting poor model fit (Zuur et al. 2009). As for the Johnson Lake samples, we had a high

**Table 3.** Model selection results for zero-altered model (ZAP) of *T. tubifex* densities in the mainstem of the Spray River and Goat Creek, Banff National Park, Alberta, using Akaike information criterion (AIC), delta AIC ( $\Delta$ AIC) and AIC weights (AIC<sub>w</sub>).  $n = 57$ . Vertical lines between parameters separate Poisson process covariates from binomial process covariates. Candidate models contained the following covariates: %slope (%; slope), contributing area (ha; cont\_area) and fine sediment (%; fines).

Rank	Parameters	AIC	$\Delta$ AIC	AIC <sub>w</sub>
1	1   slope	207.22	0.00	0.64
2	1   cont_area + slope	208.98	1.76	0.26
3	fines   cont_area + slope	210.95	3.72	0.10

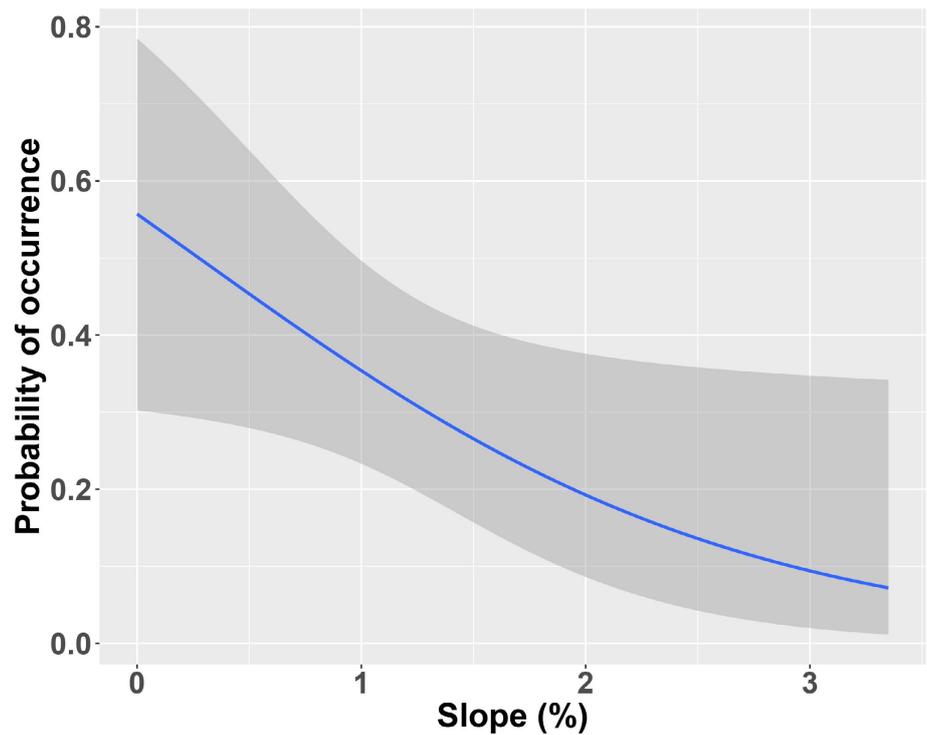
**Table 4.** Parameter estimates and their standard errors for the zero-altered model (ZAP) of *T. tubifex* densities in the Spray River and Goat Creek, Banff National Park, Alberta ( $n = 57$ ).

Process	Parameters	Estimates	Std. error	Z value	P value
Poisson	Intercept	0.311	0.120	2.605	0.009
	slope	-0.831	0.411	-2.020	0.043
Binomial	Intercept	0.229	0.544	0.422	0.673
	slope	-0.831	0.411	-2.020	0.043

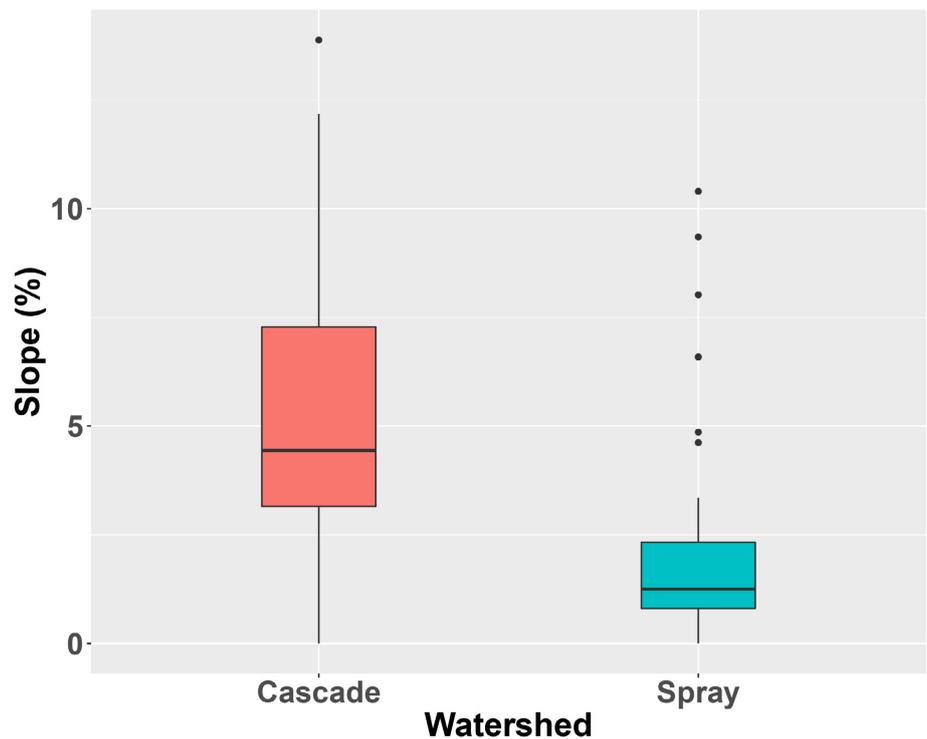
proportion of zeros in this dataset even though we excluded all of the reaches from the Cascade water due to zero responses. We had also excluded data from the tributaries of the Spray watershed to reduce the number of zeros (e.g. Arab et al. 2008). Therefore, only responses from reaches in the mainstem of the Spray River and Goat Creek were included in the modeling ( $n = 57$  reaches).

Like the Johnson Lake dataset, we proceeded with a ZAP model. We did not report ZIP models because separating two sources of zeros brought unnecessary complexity to the modeling and did not affect our conclusions (e.g., Kuhn et al. 2016). We considered contributing area and slope in the binomial process and % fine sediment in the count process (Table 3). AICs suggested removing % fine sediment and contributing area (Table 3). The final ZAP model had slope in the binomial process, but only the intercept in the truncated Poisson process (Table 4). The predicted odds of finding a *T. tubifex* decreased by a 44% for every 1% increase in slope (Table 4; Figure 3). Model validation produced no strong patterns in Pearson residuals. While AICs indicated that the ZIPNB was a better fit, the parameter estimates for the ZIPNB had very large standard errors, and therefore not reported.

Slope was the only predictor in the ZAP model with zero odds of *T. tubifex* occupying reaches with slopes  $> 3\%$ . Anlauf and Moffit (2008) also found that *T. tubifex* were associated with low reach slopes in the Pahsimeroi River drainage in Idaho, U.S. Reaches in the Cascade watershed had significantly higher slopes compared to the Spray watershed; nearly all of the sampled reaches in the Cascade watershed had a channel slope of  $> 3\%$  (Figure 4). Reaches in the tributaries of the Spray watershed also had slopes  $> 3\%$  and were also excluded from the model to reduce the number of zeros. In contrast, the sites with *T. tubifex* occurrences were located on the mainstem of the Spray River and Goat Creek which had slopes that were  $< 3\%$  (Figure 4). The presence of a hydropower facility that regulates river flow in the Spray watershed was a second notable difference between the Spray and Cascade



**Figure 3.** The relationship between slope (%) and the probability of occurrence of *T. tubifex* in the Spray watershed, Banff National Park, Alberta. The shaded area is the 95% confidence interval.



**Figure 4.** A boxplot comparison of slope (%) between the Spray and Cascade watersheds, Banff National Park, Alberta.

watersheds. *T. tubifex* have been associated with reservoirs upstream and downstream of hydropower infrastructure elsewhere (e.g., Zendt and Bergersen 2000; Dubey and Caldwell 2004) and may have been a contributing factor in this watershed comparison.

Of the occupied reaches, the predicted mean density of *T. tubifex* was 1.36 ( $\pm$  SE = 1.13) *T. tubifex*/m<sup>2</sup>. In contrast, densities of *T. tubifex* were lower than previously reported for other river systems. For example, our model predicted a mean of 1.36 *T. tubifex*/m<sup>2</sup> while tailwaters of the San Juan River, New Mexico, USA, were found to have  $> 10^3$  individuals/m<sup>2</sup> (Dubey and Caldwell 2004). However, it is difficult to compare studies who sampled different habitats. The density of *T. tubifex* was high in side channels of the Madison River, Montana, USA ( $10^2$  to  $10^3$  depending on water depth and organic material; Krueger et al. 2006). We did not specifically target side channels; however, side channels likely have lower flows and a higher proportion of fine sediment which may have supported higher densities of *T. tubifex*.

All of the *T. tubifex* sampled in the Spray watershed were lineage III and none of the *T. tubifex* captured tested positive for *Mc*. Despite this, 100% of *Salvelinus fontinalis* and *Salvelinus confluentus* ( $n = 234$ ) sampled from the same randomly allocated sites tested positive for *M. cerebralis* with two separate assays (pepsin digest and PCR) (*unpublished data* 2016). Comparatively, *Oncorhynchus clarkii lewisi*, *Salvelinus confluentus* and *Salvelinus fontinalis* collected from the Cascade watershed in 2016 ( $n = 235$ ) all tested negative for *M. cerebralis* using identical methods (*unpublished data*).

## Discussion and Conclusions

Traditionally, avoiding sampling bias is achieved using some form of random or systematic sampling (e.g., Yoccoz et al. 2001). Therefore, we started *T. tubifex* sampling using systematic (Johnson Lake) and random (Spray watershed) sampling approaches. However, *T. tubifex* occurred in only 7% and 25% of each of the modelled lake and stream sample units. We had more zeros than would be expected in a Poisson or negative binomial distribution (i.e. zero-inflation). These zeros made it difficult to meet assumptions of traditional count models; therefore, the estimated parameters and standard errors were biased (Zuur et al. 2009). Instead, we modelled *T. tubifex* densities from Johnson Lake and the Spray watershed using ZAP models. ZAP models are a two-part model that specifies one process for the zeros and one process for the counts. All of the zeros in these *T. tubifex* datasets were modelled together. Both ZAP models produced clear improvements over Poisson models.

We assumed that we had an excess of true zeros due to ecological processes, yet we acknowledge that there may have been false zeros due to sampling efficiency given the size of *T. tubifex* and the tools we used to sample sediments. ZIP models may provide insight into processes or mechanisms that generated the zeros (i.e., distinguish between true versus false zeroes; Martin et al. 2005); however, there is typically insufficient knowledge to distinguish the two types of zeros in practice (Cunningham

and Lindenmayer 2005). We chose ZAP rather than ZIP models because we were mostly interested in modeling all zeros, regardless of their source, to avoid sampling zeros in the future. This will inform larger-scale sampling that is being considered in order to support landscape-scale AIS risk assessments.

AIS can often have a rare and/or patchy distributions because they are at non-equilibrium with their environment and can occur in “hotspots” due to habitat suitability (Hattab et al. 2017; Trochine et al. 2018). Both ZAP and ZIP models and their negative binomial counterparts (ZAPNB and ZIPNB) are effective tools for conservation practitioners modelling the distribution of AIS. For example, hoop net sampling of nonnative burbot, *Lota lota* (Linnaeus, 1758) from randomly selected reaches of the Green River, Wyoming, U.S., resulted in zeros from > 50% of the sample reaches (Klein et al. 2015). Klein et al. (2015) used a ZAP model to determine that the probability of catching a burbot was related to boulder substrate and mean current velocity. Similarly, Erhardt and Tiffan (2016) had many zero catches of nonnative Siberian prawns, *Palaemon modestus* (Heller, 1862) from beam trawls of randomly selected transects in the Lower Snake River, Washington, USA. These authors used ZIPNB models and identified that the probability of catching a prawn increased in deeper water and close to dams. These ZIP and ZAP models were used to reduce bias in model parameter estimates, but the results can also be used to guide future sampling to reduce zero catches. Stratified random sampling has been used to improve efficiency and reduce uncertainty in distribution model estimates by focusing sampling on a portion of the landscape that is more likely to support the target organism (e.g., Guisan et al. 2006; Fois et al. 2018). The variables used to stratify the landscape could be determined from distribution models from previous studies. Indeed, model-based stratification helped identify new populations of target rare lichen and plant species in Switzerland and the U.S. Pacific Northwest (Edwards et al. 2005; Guisan et al. 2006). This approach resulted in a two- to fivefold increase in detection of rare lichens (Edwards et al. 2005).

When starting the study, we lacked sufficient knowledge of within-lake *T. tubifex* habitat preferences to know what habitat variables could be used to stratify sampling in Johnson Lake. Although *T. tubifex* have been associated with fine sediment in numerous studies, fine sediment occurred at all sites in Johnson Lake. In hindsight, lotic samples could have been stratified by slope in light of previous studies (see Anlauf and Moffitt 2008). Our results indicated that future sampling could benefit from stratified sampling by water depth in lentic environments and slope in lotic environments. Habitat stratification would allow conservation practitioners to focus resources and time in only the habitat where the species is most likely to occur (Edwards et al. 2005; Guisan et al. 2006).

The purpose of this manuscript was to provide guidance on modeling an AIS dataset with excess zeros. However, there are some additional statistical

nuances that were relevant to these data, related to the violation of independence. In both *T. tubifex* datasets, multiple samples were collected within the unit that was used in the analysis: five sediment cores were collected per site in Johnson Lake and three kick net samples were collected per reach in the Spray and Cascade Rivers. Site and reach were the units of analysis respectively and we aggregated the *T. tubifex* counts from samples within each unit of analysis which removed the most obvious violation of independence (spatial clustering of sampling sites). Another approach would have been to use zero-inflated mixed models (see Zuur and Ieno 2012), but these are harder to fit and interpret. Despite aggregating multiple samples from the same site or reach, spatial autocorrelation may still exist. The best practice is to plot residuals against spatial coordinates and look for patterns (Zuur et al. 2009). Additional correlation structures can be used to deal with temporal or spatial autocorrelation caused by clustering of sites or samples (Zuur et al. 2009; Zuur and Ieno 2018). However, we could not find examples of both alternative correlation structures and zero-inflation outside of textbooks and intended this manuscript to be a gentle introduction to zero-inflation modelling for conservation practitioners.

As AIS continue to threaten freshwater biodiversity, conservation practitioners need to optimize methods to assess the risk of invasions while reducing uncertainty in those estimates. We demonstrated that ZAP models, compared to models based on a standard unimodal distribution, can be an important tool to help resolve issues with zero-inflation. Mobilizing practical knowledge using case studies of the experiences of researchers and practitioners will be important to achieve on the ground actions to prevent AIS (Cvitanovic et al. 2016).

### Acknowledgements

We are grateful to anonymous reviewers who provided valuable feedback and guidance to make the manuscript as useful as possible. We also express gratitude to those that helped with the field work: Kayla Eykelboom, Dan Struthers, Chris Carli, Renee Lazar, Liam Horne, Paige Phaneuf, Hedin Nelson-Chorney, Kate Keenan, and Tim Town. We also thank the Alberta Conservation Association for support via the ACA Grants in Biodiversity Program.

### Funding declaration

This work was funded by Parks Canada Agency in the Banff Field Unit. Parks Canada's role in the study design; collection, analysis and interpretation of data, and writing is reflected in the authors. This project was also supported by the Alberta Conservation Association (ACA) via the ACA Grants in Biodiversity Program (ACAGiB #861). The ACA had no role in the study design, data collection and analysis, decision to publish or preparation of the article.

### Authors' contribution

MKT – research conceptualization, sample design and methodology, data collection, data analysis and interpretation, funding provision, writing original draft and review and editing; CW – sample design and methodology, data collection, data analysis and interpretation, writing original draft and review and editing; CJS – sample design and methodology, data analysis and interpretation, review and editing; PCH – sample design and methodology, data analysis and interpretation, review and editing; LJJ – research conceptualization, sample design and methodology, funding provision, writing original draft and review and editing.

## Ethics and permits

This study was permitted under Parks Canada Agency Research and Collection Permit #37046.

## References

- Akaike H (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723, <https://doi.org/10.1109/TAC.1974.1100705>
- Allen MB, Bergersen EP (2002) Factors Influencing the Distribution of *Myxobolus cerebralis*, the Causative Agent of Whirling Disease in the Cache La Poudre River, Colorado. *Diseases of Aquatic Organisms* 49: 51–60, <https://doi.org/10.3354/dao049051>
- Anlauf KJ, Moffitt CM (2008) Models of stream habitat characteristics associated with tubificid populations in an intermountain watershed. *Hydrobiologia* 603: 147–158, <https://doi.org/10.1007/s10750-007-9255-2>
- Anlauf KJ, Moffitt CM (2010) Modelling of landscape variables at multiple extents to predict fine sediments and suitable habitat for *Tubifex tubifex* in a stream system. *Freshwater Biology* 55: 794–805, <https://doi.org/10.1111/j.1365-2427.2009.02323.x>
- Arab A, Wildhaber M, Wikle C, Gentry C (2008) Zero-Inflated Modeling of Fish Catch per Unit Area Resulting from Multiple Gears: Application to Channel Catfish and Shovelnose Sturgeon in the Missouri River. *North American Journal of Fisheries Management* 28: 1044–1058, <https://doi.org/10.1577/M06-250.1>
- Arneberg P, Skorping A, Grenfell B, Read AF (1998) Host densities as determinants of abundance in parasite communities. *Proceedings of the Royal Society London, Series B* 265: 1283–1289, <https://doi.org/10.1098/rspb.1998.0431>
- Arsan EL, Hallett SL, Bartholomew JL (2007) *Tubifex tubifex* from Alaska and their susceptibility to *Myxobolus cerebralis*. *Journal of Parasitology* 93: 1332–1342, <https://doi.org/10.1645/GE-1117.1>
- Ayre KK, Caldwell CA, Stinson J, Landis WG (2014) Analysis of regional scale risk of Whirling disease in populations of Colorado and Rio Grande cutthroat trout using a bayesian belief network model. *Risk Analysis* 34: 1589–1605, <https://doi.org/10.1111/risa.12189>
- Barry DE, Veillard M, James CT, Brummelhuis L, Pila EA, Turnbull A, Oddy-van Oploo A, Han X, Hanington PC (2021) qPCR-based environmental monitoring of *Myxobolus cerebralis* and phylogenetic analysis of its tubificid hosts in Alberta, Canada. *Diseases of Aquatic Organisms* 145: 119–137, <https://doi.org/10.3354/dao03608>
- Bartholomew JL, Lorz HV, Atkinson SD, Hallett SL, Stevens DG, Holt RA, Lujan K, Amandi A (2007) Evaluation of a management strategy to control the spread of *Myxobolus cerebralis* in a lower Columbia River tributary. *North American Journal of Fisheries Management* 27: 542–550, <https://doi.org/10.1577/M06-151.1>
- Beauchamp KA, Kelley GO, Nehring RB, Hedrick RP (2005) The severity of whirling disease among wild trout corresponds to the differences in genetic composition of *Tubifex tubifex* populations in central Colorado. *Journal of Parasitology* 91: 53–60, <https://doi.org/10.1645/GE-327R>
- Blasco-Moreno A, Pérez-Casany M, Puig P, Morante M, Castells E (2019) What does a zero mean? Understanding false, random and structural zeros in ecology. *Methods in Ecology and Evolution* 10: 949–959, <https://doi.org/10.1111/2041-210X.13185>
- Bliss CI, Fisher RA (1953) Fitting the negative binomial distribution to biological data. *Biometrics* 9: 176–200, <https://doi.org/10.2307/3001850>
- Brinkhurst RO (1986) Guide to the freshwater aquatic microdrile oligochaetes of North America. Canadian Special Publication of Fisheries and Aquatic Sciences 84, 259 pp
- Cavender WP, Wood JS, Powell MS, Overturf K, Cain KD (2004) Real-time quantitative polymerase chain reaction (QPCR) to identify *Myxobolus cerebralis* in rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms* 60: 205–213, <https://doi.org/10.3354/dao060205>
- Cunningham RB, Lindenmayer DB (2005) Modeling count data of rare species: some statistical issues. *Ecology* 86: 1135–1142, <https://doi.org/10.1890/04-0589>
- Cvitanovic C, McDonald J, Hobday AJ (2016) From science to action: Principles for undertaking environmental research that enables knowledge exchange and evidence-based decision-making. *Journal of Environmental Management* 183: 864–874, <https://doi.org/10.1016/j.jenvman.2016.09.038>
- DuBey R, Caldwell C (2004) Distribution of *Tubifex tubifex* lineages and *Myxobolus cerebralis* infection in the tailwater of the San Juan River, New Mexico. *Journal of Aquatic Animal Health* 16: 179–185, <https://doi.org/10.1577/H04-018.1>
- Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Leveque C, Naiman, RJ, Prieur-Richard A-H, Soto D, Stiassny MLJ, Sullivan CA (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* 81: 163–182, <https://doi.org/10.1017/S1464793105006950>
- Edwards TC, Cutler DR, Zimmerman NE, Geiser L, Alegria J (2005) Model-based stratifications for enhancing the detection of rare ecological events. *Ecology* 86: 1081–1090, <https://doi.org/10.1890/04-0608>
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics* 40: 677–697, <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Erhardt JM, Tiffan KF (2016) Ecology of nonnative Siberian prawn (*Palaemon modestus*) in the lower Snake River, Washington, USA. *Aquatic Ecology* 50: 607–621, <https://doi.org/10.1007/s10452-016-9581-4>

- Fisheries and Oceans Canada (2019) Recovery Strategy and Action Plan for the Westslope Cutthroat Trout (*Oncorhynchus clarkii lewisi*) Alberta population (also known as Saskatchewan-Nelson River populations) in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa. vii + 61 pp + Part 2, <https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery/westslope-cutthroat-trout-2019-proposed.html>
- Fisheries and Oceans Canada (2020) Recovery Strategy for the Bull Trout (*Salvelinus confluentus*), Saskatchewan-Nelson Rivers populations, in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa. viii + 130 pp, <https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery-strategies/bull-trout-proposed-2020.html>
- Fois M, Cuena-Lombrana A, Fenu G, Bacchetta G (2018) Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling* 385: 124–132, <https://doi.org/10.1016/j.ecolmodel.2018.07.018>
- Green RH, Young RC (1993) Sampling to detect rare species. *Ecological Applications* 3: 351–366, <https://doi.org/10.2307/1941837>
- Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology* 20: 501–511, <https://doi.org/10.1111/j.1523-1739.2006.00354.x>
- Hallett SL, Lorz HV, Atkinson, SD, Rasmussen C, Xue L, Bartholomew JL (2009) Propagation of the myxozoan parasite *Myxobolus cerebralis* by different geographic and genetic populations of *Tubifex tubifex*: an Oregon perspective. *Invertebrate Pathology* 102: 57–68, <https://doi.org/10.1016/j.jip.2009.07.001>
- Hattab T, Garzón-López CX, Ewald M, Skowronek S, Aerts R, Horen H, Brasseur B, Gallet-Moron E, Spicher F, Decocq G, Feilhauer H, Honnay O, Kempeneers P, Schmidlein S, Somers B, Van De Kerchove R, Rocchini D, Lenoir J (2017) A unified framework to model the potential and realized distributions of invasive species within the invaded range. *Diversity and Distributions* 23: 806–819, <https://doi.org/10.1111/ddi.12566>
- Heiri O, Lotter A, Lemcke G (2000) Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *Journal of Paleolimnology* 25: 101–110, <https://doi.org/10.1023/A:1008119611481>
- Hoffman GL (1990) *Myxobolus cerebralis*, a worldwide cause of salmonid whirling disease. *Journal of Aquatic Animal Health* 2: 30–37, [https://doi.org/10.1577/1548-8667\(1990\)002<0030:MCAWCO>2.3.CO;2](https://doi.org/10.1577/1548-8667(1990)002<0030:MCAWCO>2.3.CO;2)
- Hutchinson G (1957) A Treatise on Limnology, Vol 1. Wiley, New York, USA, 1015 pp
- Jackman S (2020) pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory. United States Studies Centre, University of Sydney, Sydney, New South Wales, Australia. R package version 1.5.5, <https://github.com/atahk/pscl>
- Kaesler AJ, Sharpe WE (2006) Patterns of distribution and abundance of *Tubifex tubifex* and other aquatic oligochaetes in *Myxobolus cerebralis* enzootic areas in Pennsylvania. *Journal of Aquatic Animal Health* 18: 64–78, <https://doi.org/10.1577/H05-016.1>
- Klein ZB, Quist MC, Rhea DT, Senecal AC (2015) Habitat use of non-native burbot in a western river. *Hydrobiologia* 757: 61–71, <https://doi.org/10.1007/s10750-015-2176-6>
- Krueger RC, Kerans BL, Vincent ER, Rasmussen C (2006) Risk of *Myxobolus cerebralis* infection to rainbow trout in the Madison River, Montana, USA. *Ecological Applications* 16: 770–783, [https://doi.org/10.1890/1051-0761\(2006\)016\[0770:ROMCIT\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2006)016[0770:ROMCIT]2.0.CO;2)
- Kuhn JA, Frainger A, Knudsen R, Kristoffersen R, Amundsen P-A (2016) Effects of fish species composition on *Diphyllobothrium* spp. infections in brown trout - is three-spined stickleback a key species? *Journal of Fish Diseases* 39: 1313–1323, <https://doi.org/10.1111/jfd.12467>
- Lambert D (1992) Zero-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34: 1–14, <https://doi.org/10.2307/1269547>
- Lazim MN, Learner MA (1987) The influence of sediment composition and leaf litter on the distribution of tubificid worms (Oligochaeta). *Oecologia* 72: 131–136, <https://doi.org/10.1007/BF00385057>
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83: 2248–2255, [https://doi.org/10.1890/0012-9658\(2002\)083\[2248:ESORWD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2)
- Mandrak NE, Cudmore B (2015) Risk Assessment: Cornerstone of an Aquatic Invasive Species Program. *Aquatic Ecosystem Health and Management* 18: 312–320, <https://doi.org/10.1080/14634988.2015.1046357>
- Martin T, Wintle B, Rhodes J, Kuhnert P, Field S, Low-Choy S, Possingham H (2005) Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8: 1235–1246, <https://doi.org/10.1111/j.1461-0248.2005.00826.x>
- McCullagh P, Nelder J (1989) Generalised linear models (2<sup>nd</sup> ed.). Chapman and Hall, New York, NY, 510 pp, <https://doi.org/10.1007/978-1-4899-3242-6>
- Nehring RB, Walker PG (1996) Whirling disease in the wild: the new reality in the intermountain west. *Fisheries* 21: 28–30
- Nehring BR, Hancock B, Catanese M, Stinson MET, Winkelman D, Wood J, Epp J (2013) Reduced *Myxobolus cerebralis* actinospore production in a Colorado reservoir may be linked to changes in *Tubifex tubifex* population structure. *Journal of Aquatic Animal Health* 25: 205–220, <https://doi.org/10.1080/08997659.2013.788581>
- Nehring RB, Lukacs PM, Baxa DV, Stinson MET, Chiaramonte L, Wise SK, Poole B, Horton A (2014) Susceptibility to *Myxobolus cerebralis* among *Tubifex tubifex* populations from ten major drainage basins in Colorado where cutthroat trout are endemic. *Journal of Aquatic Animal Health* 26: 19–32, <https://doi.org/10.1080/08997659.2013.864719>

- Papes M, Havel JE, vander Zanden MJ (2016) Using maximum entropy to predict the potential distribution of an invasive freshwater snail. *Freshwater Biology* 61: 457–471, <https://doi.org/10.1111/fwb.12719>
- Peterson EE, Ver Hoef JM (2014) STARS: an ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *Journal of Statistical Software* 56: 1–17, <https://doi.org/10.18637/jss.v056.i02>
- Rodríguez-Rey M, Consuegra S, Börger L, Garcia de Leaniz C (2019) Improving species distribution modelling of freshwater invasive species for management applications. *PLoS ONE* 14: e0217896, <https://doi.org/10.1371/journal.pone.0217896>
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
- Royle JA, Nichols JD (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84: 777–790, [https://doi.org/10.1890/0012-9658\(2003\)084\[0777:EAFRPA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2)
- Stevens DL, Olsen AR (2004) Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99: 262–278, <https://doi.org/10.1198/016214504000000250>
- Stevens R, Kerans BL, Lemmon J, Rasmussens C (2001) The effects of *Myxobolus cerebralis* myxospore dose on triactinomyxon production and biology of *Tubifex tubifex* from two geographic regions. *Journal of Parasitology* 87: 315–321, [https://doi.org/10.1645/0022-3395\(2001\)087\[0315:TEOMCM\]2.0.CO;2](https://doi.org/10.1645/0022-3395(2001)087[0315:TEOMCM]2.0.CO;2)
- Trochine C, Brucet S, Jeppesen E, Winfield IJ, Argillier C, Arranz I, Ferreira T, Hesthagen T, Holmgren K, Jeppesen E, Kelly F, Krause T, Rask M, Volta P, Winfield IJ, Mehner T (2018) Non-native fish occurrence and biomass in 1943 western Palearctic lakes and reservoirs and their abiotic and biotic correlates. *Ecosystems* 21: 395–409, <https://doi.org/10.1007/s10021-017-0156-6>
- Turley MD, Bilotta GS, Arbocute G, Chadd RP, Extence CA, Brazier RE (2016) Quantifying submerged deposited fine sediments in rivers and streams using digital image analysis. *River Research and Applications* 33: 1585–1595, <https://doi.org/10.1002/rra.3073>
- Vincent ER (1996) Whirling disease and wild trout: the Montana experience. *Fisheries* 21: 32–34
- Ward JC (1974) The Fishes and their Distribution in the Mountain National Parks. Report prepared by the Canadian Wildlife Service. Calgary, AB. 41 pp + App. <http://friendsofkootenay.ca/fishes-and-their-distribution-mountain-national-parks-canada>
- Wolf K, Markiw ME (1984) Biology contravenes taxonomy in the Myxozoa: new discoveries show alternation of invertebrate and vertebrate hosts. *Science* 225: 1449–1452, <https://doi.org/10.1126/science.225.4669.1449>
- Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16: 446–453, [https://doi.org/10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)
- Zendt JS, Bergersen EP (2000) Distribution and abundance of the aquatic oligochaete host *Tubifex tubifex* for the salmonid Whirling Disease parasite *Myxobolus cerebralis* in the upper Colorado River basin. *North American Journal of Fisheries Management* 20: 502–512, [https://doi.org/10.1577/1548-8675\(2000\)020<0502:DAAOTA>2.3.CO;2](https://doi.org/10.1577/1548-8675(2000)020<0502:DAAOTA>2.3.CO;2)
- Zielinski CM, Lorz HV, Hallett SL, Xue L, Bartholomew JL (2011) Comparative susceptibility of Deschutes River, Oregon, *Tubifex tubifex* populations to *Myxobolus cerebralis*. *Journal of Aquatic Animal Health* 23: 1–8, <https://doi.org/10.1080/08997659.2011.545710>
- Zuur AF, Ieno EN (2012) Zero Inflated Models and Generalized Linear Mixed Models with R. Newburgh, UK: Highland Statistics Ltd., 324 pp
- Zuur AF, Ieno EN (2018) Beginner's Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA. Newburgh, UK: Highland Statistics Ltd., 362 pp
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed Effects Models and Extensions in Ecology with R. Springer, New York, 574 pp, <https://doi.org/10.1007/978-0-387-87458-6>

### Supplementary material

The following supplementary material is available for this article:

**Table S1.** Table of *T. tubifex* counts sample locations and model covariates for Johnson Lake, Banff National Park, Alberta, Canada.

**Table S2.** Table of *T. tubifex* counts sample locations and model covariates for the Spray and Cascade watersheds, Banff National Park, Alberta, Canada.

This material is available as part of online article from:

[http://www.reabic.net/journals/mbi/2022/Supplements/MBI\\_2022\\_Taylor\\_etal\\_SupplementaryMaterial.xlsx](http://www.reabic.net/journals/mbi/2022/Supplements/MBI_2022_Taylor_etal_SupplementaryMaterial.xlsx)