Received: 11 October 2024 | Accepted: 24 January 2025

DOI: 10.1111/2041-210X.70000

Methods in Ecology and Evolution

APPLICATION

eDNAjoint: An R package for interpreting paired or semi-paired environmental DNA and traditional survey data in a Bayesian framework

Abigail G. Keller[1] | **Ryan P. Kelly**[2]

[1]Department of Environment Science, Policy, and Management, University of California, Berkeley, Berkeley, California, USA

[2]School of Marine and Environmental Affairs, University of Washington, Seattle, Washington, USA

Correspondence
Abigail G. Keller
Email: agkeller@berkeley.edu

Funding information
U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Grant/Award Number: DE-SC0024386

Handling Editor: Chloe Robinson

Abstract

1. Environmental DNA (eDNA) sampling is increasingly used in surveys of species distribution as a potentially sensitive and efficient monitoring method. Yet access to modelling tools designed specifically for interpreting this new data type lags behind its ubiquity. While occupancy modelling software has dominated the analytical landscape for eDNA data analysis of single species, this type of model may not always be the most appropriate. The rate of eDNA detection often corresponds to species density, rather than just occupancy, and researchers often have access to observations from non-genetic sampling methods at the same sites.
2. To provide users access to a modelling framework designed to maximize the use of all available data, we developed an R package, eDNAjoint. The package provides an easy-to-use interface for fitting a 'joint' model that integrates data from paired or semi-paired eDNA and traditional surveys in a Bayesian framework. The model can be used to estimate parameters like the probability of a false positive eDNA detection and mean catch rate at a site, and the package allows access to multiple model variations and Bayesian prior customization. Additional functionality can be used for model selection, summarising posteriors and comparing the relative sensitivities of the two survey methods.
3. We demonstrate the use of eDNAjoint by fitting a variation of the model with site-level covariates that scale the sensitivity of eDNA sampling relative to traditional sampling. The example workflow uses binary eDNA and seine count data for the endangered tidewater goby (*Eucyclogobius newberryi*) from a study by Schmelzle and Kinziger (2016). This use case includes a prior sensitivity analysis and an evaluation of the relationship between detection rates and environmental variables.
4. eDNAjoint has the potential to greatly increase the range of users who will be able to rigorously analyse eDNA and traditional survey data in a Bayesian

framework, understand if and how eDNA can improve monitoring practices, and gain confidence in the interpretability of eDNA data.

KEYWORDS
Bayesian methods, environmental DNA, false positive probability, software, species distribution, statistics

1 | INTRODUCTION

The availability of environmental DNA (eDNA) data has proliferated since its first application in assessing the species distribution of a vertebrate nearly 20 years ago (Ficetola et al., 2008). The potential for rapid and cost-effective surveys of common, endangered (Bonfil et al., 2021), rare (Pfleger et al., 2016) and invasive species (Larson et al., 2020) has introduced the extraction and identification of DNA from environmental samples as a routine monitoring technique. Additionally, the emergence of private industry with dedicated eDNA services (i.e. Smith-Root and Jonah Ventures) has increased the accessibility of eDNA sampling gear and laboratory analytical expertise.

Software tools designed specifically for interpreting eDNA data, however, have not kept pace with the broad availability of this new data source. Robust approaches linking environmental DNA to quantitative estimates of species abundance and distribution have grown in peer-reviewed literature (Guri et al., 2024; Levi et al., 2019; Shelton et al., 2022; Tillotson et al., 2018; Yates et al., 2021), yet these methods are confined to academic communities with the technical expertise necessary to develop bespoke models. Importantly, institutional inertia due to challenges in data interpretation has limited the adoption of this novel data source in management agencies' decision-making routines (Lee et al., 2024).

The application of occupancy models to eDNA detection–nondetection data has dominated the eDNA single-species analytical landscape (Burian et al., 2021; Erickson et al., 2017; Pilliod et al., 2013; Strickland & Roberts, 2019). Originally developed to account for imperfect detection in visual, auditory and capture surveys of animals (MacKenzie et al., 2002), this hierarchical model has proved to be a useful framework for understanding eDNA-based distribution data. Since uncertainty around false-positive detections presents a hurdle to adopting eDNA-based approaches (Jerde, 2021), the occupancy model has been extended to account for false-positive detections at multiple scales in the context of eDNA analysis (Guillera-Arroita et al., 2017; Lahoz-Monfort et al., 2015). Responding to the occupancy model's utility in interpreting eDNA data, recent software advances include the occuFP feature of unmarked that fits occupancy models when false positive detections occur (Fiske & Chandler, 2011), as well as eDNAoccupancy that fits Bayesian multiscale occupancy models (Dorazio & Erickson, 2018).

Occupancy models, however, may not always be the most appropriate analytical method for interpreting eDNA detection–nondetection data. Variation in species abundance can induce variation in detection probability with genetic methods, providing an opportunity to estimate species density, rather than occupancy, from repeated observations of the presence or absence of genetic material (Royle & Nichols, 2003). Additionally, observations from nongenetic datasets, hereafter referred to as 'traditional observations', often spatially and temporally overlap with some or all eDNA observations (Guillera-Arroita et al., 2017; Schmelzle & Kinziger, 2016). Occupancy modelling with eDNA data can also underestimate species occupancy and overestimate detection probability, yet this bias can be ameliorated by including traditional observations as a proxy for abundance (Randall et al., 2023). Existing software has not yet facilitated the integration of eDNA observations with measurements made by other means.

The R package eDNAjoint builds upon and makes accessible a model developed in Keller et al. (2022) that sets up an analytical framework for integrating eDNA and traditional data into management processes (Keller et al., 2022). By jointly modelling repeated observations from both survey methods, the two data streams can inform each other to quantify uncertainty in both methods, understand their relative detection sensitivities and estimate parameters including mean (expected) species catch rate and the probability of a false positive eDNA detection. As a software tool designed to facilitate the integration of multiple available data sources, eDNAjoint may bring eDNA analytical capabilities to a broader researcher and practitioner base.

2 | MODEL DESCRIPTION

The package eDNAjoint is intended for use with replicated, paired or semi-paired eDNA (binary, detection–nondetection) and traditional (count or continuous) observations at multiple sites across the landscape (Figure 1).

The package runs a Bayesian model that integrates these two data streams to jointly estimate parameters like the false positive probability of eDNA detection and expected catch rate at a site (i.e. expected number of individuals per traditional sample). Optional model variations allow inclusion of site-level covariates that scale the sensitivity of eDNA sampling relative to traditional sampling, as well as the estimation of gear scaling coefficients when multiple traditional gear types are used. Additional functions in the package facilitate the interpretation of model fits.

Below is a representation of the model used in eDNAjoint including all model variations. Note that inclusion of gear scaling

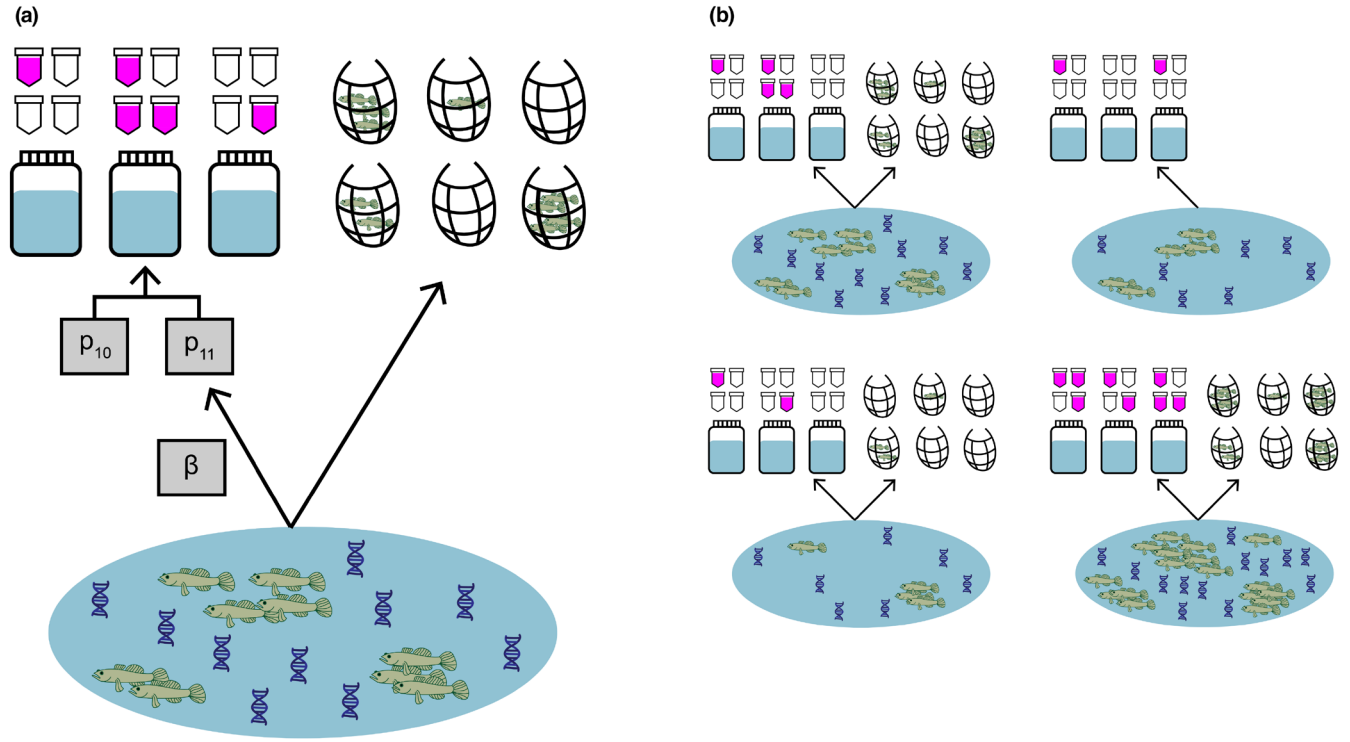

FIGURE 1 Conceptual diagram of joint model with (a) paired and (b) semi-paired data.

coefficients, q_k (Equation 2), and the regression with site-level covariates, α (Equation 4), are optional in implementation with eDNA-joint. A reduced version of the joint model without these variations is also described in Keller et al. (2022).

The model can accommodate both discrete count and continuous data from traditional surveys. Traditional observation, Y, of a species at site, i, in survey sample, j, of gear type, k, is drawn from either

1. a negative binomial distribution with mean (expected) species catch rate, $\mu_{i,k}$, and an overdispersion parameter, ϕ (Equation 1.1)
2. a Poisson distribution with mean (expected) species catch rate, $\mu_{i,k}$ (Equation 1.2).
3. a gamma distribution with shape parameter, α_{mu} and rate parameter, β_{mu}. The mean (expected) species catch rate, $\mu_{i,k}$ is equal to $\frac{\alpha_{mu}}{\beta_{mu}}$ (Equation 1.3),

$$Y_{i,j,k} \sim \text{NegativeBinomial}\left(\mu_{i,k}, \phi\right), \tag{1.1}$$

$$Y_{i,j,k} \sim \text{Poisson}\left(\mu_{i,k}\right), \tag{1.2}$$

$$Y_{i,j,k} \sim \text{Gamma}\left(\alpha_{mu,i,k}, \beta_{mu,i,k}\right). \tag{1.3}$$

Gear scaling coefficients, q_k, scale the catch rates of multiple gear types relative to gear type 1 (Equation 2),

$$\mu_{i,k} = q_k \times \mu_{i,1}. \tag{2}$$

The probability of a true positive eDNA detection, p_{11}, at site i, is a function of mean (expected) species catch rate, $\mu_{i,1}$ and scaling coefficient β_i (Equation 3),

$$p_{11,i} = \frac{\mu_{i,1}}{\mu_{i,1} + e^{\beta_i}}. \tag{3}$$

The scaling coefficient β_i relates the sensitivity of eDNA sampling to the mean (expected) species catch rate and is a function of site-level covariate coefficients, α_n and site-level covariate data, $A_{i,n}$ (Equation 4). Both q_k and β_i are dimensionless,

$$\beta_i = A_{i,n}^T \times \alpha_n. \tag{4}$$

The total probability of eDNA detection at site i, p_i, is the sum of the probability of a true positive eDNA detection at site i, $p_{11,i}$, and the probability of a false positive eDNA detection, p_{10} (Equation 5),

$$p_i = p_{11,i} + p_{10}. \tag{5}$$

The number of positive PCR eDNA detections, K, out of the number of trials, N, in eDNA water sample m at site i is drawn from a binomial distribution, with a probability of success on a single trial, p_i (Equation 6),

$$K_{i,m} \sim \text{Binomial}\left(N_{i,m}, p_i\right). \tag{6}$$

Non-uniform prior distributions are included in the model for parameters p_{10}, α_n and ϕ (if a negative binomial distribution is used to describe the traditional survey observations, Equation 1.2),

$$p_{10} \sim \text{Beta}(\alpha, \beta), \tag{7}$$

$$\phi \sim \text{Gamma}(\alpha, \beta), \tag{8}$$

$$\alpha_n \sim \text{Normal}(0, 10). \quad (9)$$

Hyperparameter values for p_{10} and ϕ prior distributions can be specified by the user in function arguments (see below).

3 | IMPLEMENTATION IN R

The models that can be run with eDNAjoint use Bayesian inference for parameter estimation. The models are specified in the probabilistic programming language Stan, which uses Hamiltonian Monte Carlo to obtain posterior simulations (Carpenter et al., 2017). The workflow of the package includes: (1) preparing data, (2) fitting the model and (3) interpreting model results. The following user guide includes installation instructions, a high-level overview of the functions, and example code: https://ednajoint.netlify.app/.

3.1 | Preparing the data

eDNAjoint is suitable for paired or semi-paired traditional and detection–nondetection eDNA survey data collected at multiple sites across the landscape for a single species. Environmental DNA and traditional data can both be collected at all sites (i.e. paired), or eDNA and traditional data can both be collected at some sites, while eDNA data is only collected at other sites (i.e. semi-paired). Both eDNA and traditional survey data should have a hierarchical structure:

- Sites (primary sample units) within a study area.
- eDNA and traditional samples (secondary sample units) collected from each site.
- eDNA subsamples (replicate observations) taken from each eDNA sample.

Users will input their data as a named list of matrices, where the matrix structure is designed to be similar to the structure of input data in the occupancy modelling package unmarked. All implementations of the model should contain the following named matrices in the list:

- pcr_n: matrix of dimensions *i* x *m*, representing the total number of positive PCR detections obtained for each site (row) and eDNA secondary sample (column).
- pcr_k: matrix of dimensions *i* x *m*, representing the total number of eDNA subsamples (replicate observations) collected at each site (row) in each eDNA secondary sample (column).
- count: matrix of dimensions *i* x *j*, representing the total number of animal individuals collected in each traditional secondary sample (column) at each site (row).

The following matrices can optionally be included in the data list:

- site_cov: matrix of dimensions *i* x *n*, representing values of site-level covariates (column) at each site (row) that scale the relationship between eDNA and traditional sampling sensitivity.
- count_type: matrix of dimensions *i* x *j*, representing the gear type of each traditional secondary sample (column) at each site (row).

3.2 | Fitting the model

The main functionality in eDNAjoint is the use of joint_model() that will fit the model to data. Along with the data as an input argument, the user can specify the desired model variation with additional inputs that describe the distribution used for the data-generating process of traditional samples (Equations 1.1–1.3), site-level covariates included in the model and the presence of multiple traditional gear types (Table 1). Additional arguments allow the user to specify parameters that control the MCMC sampling process.

The user can also specify the values of the hyperparameters used in the beta distribution for the p_{10} prior (Equations 5 and 7), as well as the gamma distribution for the ϕ prior used for overdispersed

TABLE 1 Model variation options and arguments used to implement variations in joint_model().

Model variation		Description	joint_model()**argument**
Distribution representing data-generating process of traditional survey data	Poisson	Count data, mean = variance	family = "poisson"
	Negative binomial	Count data, mean ≠ variance	family = "negbin"
	Gamma	Continuous data	family = "gamma"
Site-level covariates		Includes covariates that scale the relative detection sensitivities of eDNA and traditional sampling	cov = c("cov1", "cov2", "cov3")
Multiple traditional gear types		Includes gear scaling coefficients that scale the relative detection sensitivities of multiple traditional gear types	q = TRUE

count observations (Equations 1.2 and 2). The default specification for the p_{10} prior is beta($\alpha_{p_{10}} = 1$, $\beta_{p_{10}} = 20$) (mean: 0.048, var: 0.045), and the default specification for the ϕ prior is gamma(shape = 0.25, rate = 0.25) (mean: 1, var: 4). These default hyperparameter values for both prior distributions are relatively uninformative. For example, the p_{10} prior hyperparameter values represent an expectation that the probability of a false positive eDNA detection is likely less than 0.2: $P(p_{10} < 0.2 | \alpha_{p_{10}}, \beta_{p_{10}} = 0.99)$. More information about these prior distributions can be found in Appendix S1. The impact of these prior choices on inference can be evaluated through a prior sensitivity analysis, as detailed below, in Appendix S2, and in the user guide.

Using these inputs, joint_model() bundles the data into a format suitable for Stan and passes the data to the appropriate precompiled Stan model. Stan's MCMC sampler is then invoked via the rstan package (Stan Development Team, 2024), and the resulting model fit is an object of class stanFit.

3.3 | Interpreting model results

Additional functions in eDNAjoint facilitate interpretation of fitted models. joint_select() is a wrapper function of methods in the loo package that can be used for model selection using leave-one-out cross validation of a list of model fits (Vehtari et al., 2024). Stan model outputs will return information for all parameters and generated quantities in the model formulation, so joint_summarize() is a wrapper function of rstan's summary() that summarizes the posterior distributions of only parameters present in the model description for easier interpretation (Table 2).

As the species mean catch rate, μ, decreases, the probability of a true positive eDNA detection will ultimately decrease below the probability of a false positive eDNA detection. The function mu_critical() calculates the mean catch rate where the probability of a true positive and false positive eDNA detection are equal, μ_{critical}. This threshold effectively represents the species density at which eDNA detections become unreliable.

eDNAjoint allows users to compare the relative sensitivities of the two survey methods. The function detection_calculate() calculates the number of survey units necessary to detect species presence, and if site-level covariates are included in the model, detection_calculate() can be used to predict how covariate values scale the sensitivity of the two methods. The function detection_plot() creates visualisations of the sampling methods' relative ability to detect species presence.

All the models fit using eDNAjoint are of the stanfit class and can be analysed and manipulated with functions in the rstan package, in addition to the functions outlined above.

4 | ASSUMPTIONS AND LIMITATIONS

The models in eDNAjoint are formulated in a Bayesian framework, so users must have some knowledge of how to form a prior distribution and understand its influence on inference (Lemoine, 2019), as well as understand how to diagnose and mitigate MCMC problems such as lack of convergence (Kéry & Royle, 2020). To facilitate this, joint_summarize() in the eDNAjoint package is a wrapper function of rstan's summary() that provides convergence diagnostics such as $\hat{R}$ statistic and effective sample size (Stan Development Team, 2024; Vehtari et al., 2021), while filtering only parameters present in the model description and would be interpretable to the user (Table 2). Additionally, users can assess model convergence using diagnostics available in the shinystan package (Stan Development Team, 2017). More tips on MCMC troubleshooting and visualization can be found in the eDNAjoint user guide: https://ednajoint.netlify.app/.

Importantly, users will need to make decisions about eDNA and traditional observations considered 'paired'. Sample collection may not be identical in time and space, so users will need to determine a reasonable spatial and temporal window of

TABLE 2 Parameters included in the joint model, including symbols, names and descriptions.

Symbol	Name	Description
$\mu_{i,k}$	mu	Vector of mean (expected) catch rate at site, *i*. If multiple traditional gear types are used, mu is an array of mean (expected) catch rate at site, *i*, with gear type, *k*
p_{10}	p10	Probability of false positive eDNA detection
q_k	q	Vector of gear scaling coefficients for traditional gear type, *k*
α_n	alpha	Vector of regression coefficients for site-level covariates that scale the sensitivity of eDNA sampling relative to traditional sampling
β_i	beta	Parameter that scales the site-specific sensitivity of eDNA relative to traditional sampling. β is a vector of length, *i*, and a function of α_n. If site-level covariates are not used, β_i are all equal to α_1
ϕ	phi	Overdispersion parameter in negative binomial distribution, if used

observation overlap. The package is also likely not suitable for datasets where the species is present in high abundance across the landscape. The model uses eDNA binary detection–nondetection data, and there is a density threshold where detection saturates, therefore reducing heterogeneity in eDNA detection required to resolve eDNA's sensitivity. Inclusion of sites where the species is absent or below the true detection threshold by both survey methods may also be necessary for estimating the false positive probability of eDNA detection. Additionally, eDNAjoint could be used in conjunction with the package artemis, which provides a framework for defining qPCR detection–nondetection by rigorously accounting for qPCR censorship around Ct value thresholds (Espe et al., 2022).

Users of eDNAjoint are also limited to using only the available model variations and cannot further customize model structure. The models in the package simplify the data-generating process relative to other eDNA model formulations that have more complex representations of the eDNA data-generating processes (Guri et al., 2024) and/or explicitly represent eDNA spatial patterns (Shelton et al., 2022). Further extensions of this package could account for spatial or temporal autocorrelation among eDNA and traditional observations, make use of DNA copy number rather than binary detection–nondetection or distinguish the effects of false positives caused by PCR-related errors or by contamination in the field or in the laboratory (Guillera-Arroita et al., 2017).

5 | eDNAjoint USE CASE

We illustrate the functionality of the package with an example workflow using a case study with the endangered tidewater goby (*Eucyclogobius newberryi*). We exemplify how including site-level covariate data when fitting the model can be useful for understanding the detection probabilities of the two survey methods. Examples of other use cases, including the implementation of the joint model with multiple traditional gear types and semi-paired data, are provided in the user guide: https://ednajoint.netlify.app/.

The data used in this example come from a study by Schmelzle and Kinziger (2016), where eDNA samples were collected at 39 sites, along with paired traditional seine sampling for endangered tidewater gobies in California (Schmelzle & Kinziger, 2016). Environmental data were collected at each site, including salinity and average time to filter eDNA water samples.

We first fit three candidate models to the data with the function joint_model(): (1) a null model with no covariates, (2) a model with salinity as a site-level covariate and (3) a model with salinity and eDNA water sample filter time as site-level covariates (Appendix S2). All models used a Poisson distribution to describe the data-generating process for traditional seine data. We then compared the predictive accuracy of candidate models using leave-one-out cross-validation with the function joint_select(). Accuracy was measured using the expected log pointwise predictive density (ELPD) relative to the top-ranked model (ΔELPD; Vehtari et al., 2017). Model 3 had the highest predictive accuracy in the model set (model 2: ΔELPD = −31.9; model 1: ΔELPD = −34.7).

Recognizing the potential influence of prior distribution parameterization on posterior inference, we then conducted a prior sensitivity analysis for the probability of false positive eDNA detection, *p10* (Appendix S2). Using the covariate set associated with the highest predictive accuracy, we fit the model with multiple prior parameterizations, and we found that the posterior of *p10* was relatively insensitive to the prior distribution, suggesting that posterior inference is driven by the data rather than prior choice (Figure 2).

We then summarized the posterior distributions of the probability of a false positive eDNA detection, p10, and the regression coefficients for site-level covariates, alpha (Table 3). The sign of the coefficients indicates that the probability of true positive eDNA detection increases as salinity increases and decreases as the filter time of water samples (i.e. turbidity) increases. Using the posterior samples in the model fit object, we calculated the

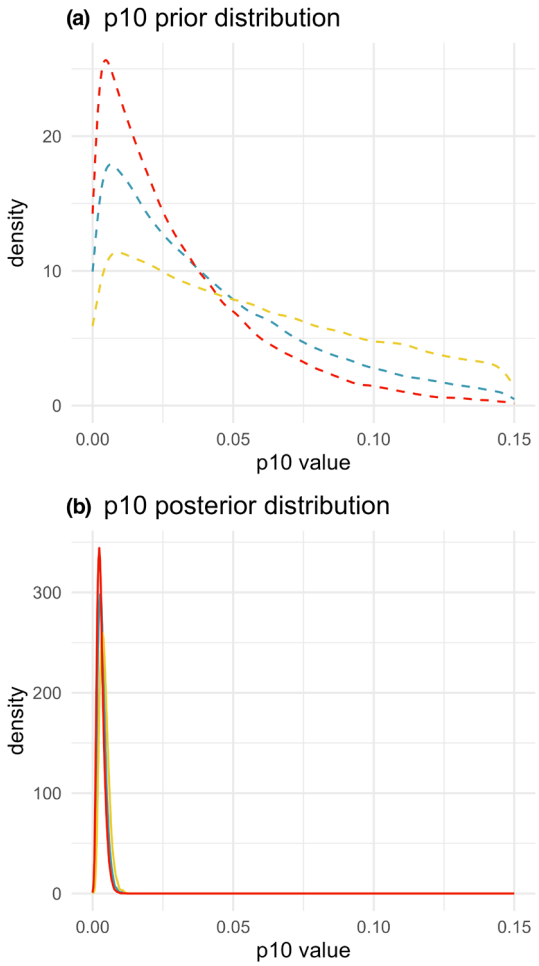

FIGURE 2 Results of prior sensitivity analysis for *p10*, probability of false positive eDNA detection of endangered tidewater gobies. (a) Prior distributions with varying shape parameters of beta distribution (red: (1, 30); blue: (1, 20); yellow: (1, 10)). (b) Posterior samples for *p10* with varying prior distributions (red: (1, 30); blue: (1, 20); yellow: (1, 10)). Code for sensitivity analysis can be found in Appendix S2.

TABLE 3 Summaries of posterior samples of parameters in the model fit with endangered tidewater goby data.

Parameter	Description	Mean	95% credibility interval
p_{10}	Probability of a false positive eDNA detection	0.003	0.001, 0.007
α_1	Intercept in site-level covariate regression that scale eDNA sample sensitivity relative to traditional sample sensitivity	0.543	0.346, 0.734
α_2	eDNA filter time coefficient for site-level covariate regression that scale eDNA sample sensitivity relative to traditional sample sensitivity	1.021	0.792, 1.249
α_3	Salinity coefficient for site-level covariate regression that scale eDNA sample sensitivity relative to traditional sample sensitivity	−0.351	−0.555, −0.144

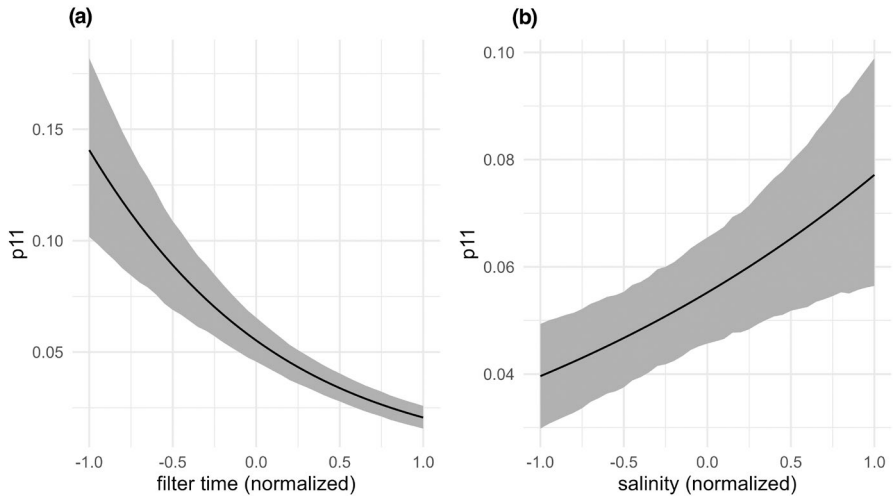

FIGURE 3 Probability of a true positive eDNA detection, *p11*, calculated at a mean (expected) catch rate of endangered tidewater gobies, μ, of 0.1. Calculations were made for (a) a range of site-level eDNA sample filter time values, with other site-level covariates held constant, and (b) a range of site-level salinity values, with other site-level covariates held constant. Covariate values are normalized and represent *z*-scores on the scale of the original covariate data. Solid line indicates the value of *p11* calculated at the mean value of posterior samples, and grey band indicates the 95% credibility interval. Code for calculations can be found in Appendix S3.

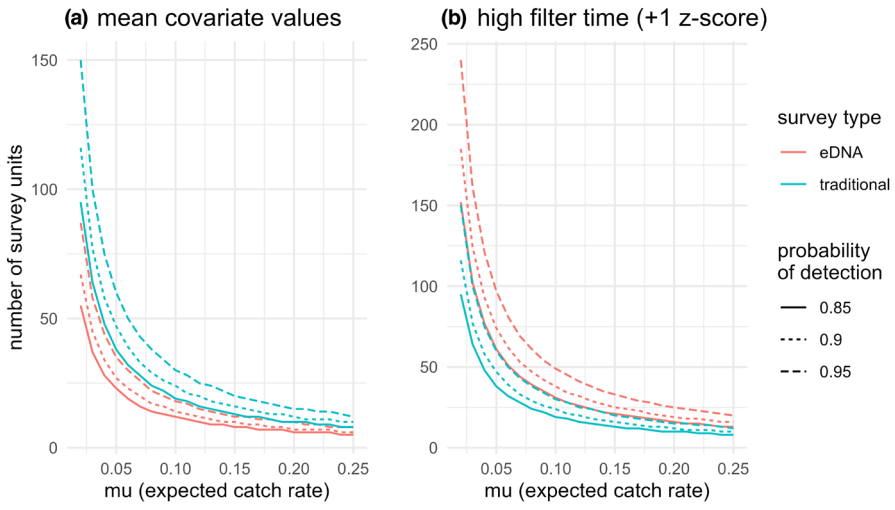

FIGURE 4 Number of traditional and eDNA survey units necessary to detect endangered tidewater goby presence with 85%, 90% and 95% probability at a range of mean (expected) catch rates, μ. Calculations are made with the eDNAjoint function detection_calculate() using (a) mean covariate values and (b) eDNA sample filter time one *z*-score above the mean, with all other covariate values held constant. Code for calculations can be found in Appendix S3.

probability of a true positive eDNA detection, *p11*, as a function of covariate values, while holding the mean (expected) catch rate, μ, constant (Figure 3).

Finally, to make the relative detection sensitivities of eDNA and seine sampling methods more concrete, we used detection_calculate() to determine the number of survey units of each method necessary to detect species presence with a defined probability (Figure 4). These calculations were made at the mean covariate values, as well as at sites where the time to filter eDNA water samples is one *z*-score above the mean. This analysis demonstrated that eDNA is on average more sensitive than seine samples (Figure 4a), but becomes less sensitive than seine samples at turbid sites where the time to filter water samples is high (Figure 4b).

AUTHOR CONTRIBUTIONS

Abigail G. Keller and Ryan P. Kelly conceived the ideas; Abigail G. Keller developed the package and wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

The authors would like to thank Carl Boettiger for R package development advice and Neha Acharya-Patel for valuable beta testing. This material is based upon work supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award Number DE-SC0024386.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70000.

DATA AVAILABILITY STATEMENT

The package eDNAjoint is available on the Comprehensive R Archive Network (CRAN; https://cran.r-project.org/web/packages/eDNAjoint/index.html). Raw data used as examples are available via: https://github.com/ropensci/eDNAjoint/tree/master/data for the tidewater goby data (Schmelzle & Kinziger, 2016) and European green crab data (Keller et al., 2022). All software code and associated data are available via https://doi.org/10.5281/zenodo.14750279 (Keller, 2025).

ORCID

Abigail G. Keller https://orcid.org/0000-0001-6068-0173
Ryan P. Kelly https://orcid.org/0000-0001-5037-2441

REFERENCES

Bonfil, R., Palacios-Barreto, P., Vargas, O. U. M., Ricaño-Soriano, M., & Díaz-Jaimes, P. (2021). Detection of critically endangered marine species with dwindling populations in the wild using eDNA gives hope for sawfishes. *Marine Biology*, *168*(5), 1–12. https://doi.org/10.1007/s00227-021-03862-7

Burian, A., Mauvisseau, Q., Bulling, M., Domisch, S., Qian, S., & Sweet, M. (2021). Improving the reliability of eDNA data interpretation. *Molecular Ecology Resources*, *21*(5), 1422–1433. https://doi.org/10.1111/1755-0998.13367

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32.

Dorazio, R. M., & Erickson, R. A. (2018). Ednaoccupancy: An r package for multiscale occupancy modelling of environmental DNA data. *Molecular Ecology Resources*, *18*(2), 368–380. https://doi.org/10.1111/1755-0998.12735

Erickson, R. A., Merkes, C. M., Jackson, C. A., Goforth, R. R., & Amberg, J. J. (2017). Seasonal trends in eDNA detection and occupancy of bigheaded carps. *Journal of Great Lakes Research*, *43*(4), 762–770. https://doi.org/10.1016/j.jglr.2017.06.003

Espe, M. B., Johnston, M., Blankenship, S. M., Dean, C. A., Bowen, M. D., Schultz, A., & Schumer, G. (2022). The artemis package for environmental dna analysis in r. *Environmental DNA*, *4*(3), 523–532.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fiske, I. J., & Chandler, R. B. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, *43*(10), 1–23. https://doi.org/10.18637/jss.v043.i10

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, *8*(9), 1081–1091. https://doi.org/10.1111/2041-210X.12743

Guri, G., Shelton, A. O., Kelly, R. P., Yoccoz, N., Johansen, T., Præbel, K., Hanebrekke, T., Ray, J. L., Fall, J., & Westgaard, J.-I. (2024). Predicting trawl catches using environmental DNA. *ICES Journal of Marine Science*, *81*, 1536–1548.

Jerde, C. L. (2021). Can we manage fisheries with the inherent uncertainty from eDNA? *Journal of Fish Biology*, *98*, 341–353. https://doi.org/10.1111/jfb.14218

Keller, A. G. (2025). Data and code for: 'eDNAjoint: An R package for interpreting paired or semi-paired environmental DNA and traditional survey data in a Bayesian framework' (version v0.3.1). *Zenodo*. https://doi.org/10.5281/zenodo.14750279

Keller, A. G., Grason, E. W., McDonald, P. S., Ramón-Laca, A., & Kelly, R. P. (2022). Tracking an invasion front with environmental DNA. *Ecological Applications*, *32*(4), 1–18. https://doi.org/10.1002/eap.2561

Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in r and bugs: Volume 2: Dynamic and advanced models*. Academic Press.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2015). Statistical approaches to account for false positive errors in environmental DNA samples. *Molecular Ecology Resources*, *16*, 673–685.

Larson, E. R., Graham, B. M., Achury, R., Coon, J. J., Daniels, M. K., Gambrell, D. K., Jonasen, K. L., King, G. D., LaRacuente, N., Perrin-Stowe, T. I., Reed, E. M., Rice, C. J., Ruzi, S. A., Thairu, M. W., Wilson, J. C., & Suarez, A. V. (2020). From eDNA to citizen science: Emerging tools for the early detection of invasive species. *Frontiers in Ecology and the Environment*, *18*(4), 194–202. https://doi.org/10.1002/fee.2162

Lee, K. N., Kelly, R. P., Demir-Hilton, E., Laschever, E., & Allan, E. A. (2024). Adoption of environmental DNA in public agency practice. *Environmental DNA*, *6*(1), 1–14. https://doi.org/10.1002/edn3.470

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in bayesian analyses. *Oikos*, *128*(7), 912–928.

Levi, T., Allen, J. M., Bell, D., Joyce, J., Russell, J. R., Tallmon, D. A., Vulstek, S. C., Yang, C., & Yu, D. W. (2019). Environmental DNA for the enumeration and management of Pacific salmon. *Molecular Ecology Resources*, *19*(3), 597–608. https://doi.org/10.1111/1755-0998.12987

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

Pfleger, M. O., Rider, S. J., Johnston, C. E., & Janosik, A. M. (2016). Saving the doomed: Using eDNA to aid in detection of rare sturgeon for conservation (Acipenseridae). *Global Ecology and Conservation*, *8*, 99–107. https://doi.org/10.1016/j.gecco.2016.08.008

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130. https://doi.org/10.1139/cjfas-2013-0047

Randall, L. A., Goldberg, C. S., & Moehenschlager, A. (2023). Environmental dna surveys can underestimate amphibian occupancy and overestimate detection probability: Implications for practice. *The Journal of Wildlife Management*, *87*(7), e22463.

Royle, A. J., & Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, *84*(3), 777–790.

Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, *16*(4), 895–908. https://doi.org/10.1111/1755-0998.12501

Shelton, A. O., Ramón-Laca, A., Wells, A., Clemons, J., Chu, D., Feist, B. E., Kelly, R. P., Parker-Stetter, S. L., Thomas, R., Nichols, K. M., & Park, L. (2022). Environmental DNA provides quantitative estimates of Pacific hake abundance and distribution in the open ocean. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1971), 20212613. https://doi.org/10.1098/rspb.2021.2613

Stan Development Team. (2017). *Shinystan: Interactive visual and numerical diagnostics and posterior analysis for bayesian models*. R package version 2.4.0. http://mc-stan.org/

Stan Development Team. (2024). *RStan: The R interface to Stan*. R package version 2.32.6. https://mc-stan.org/

Strickland, G. J., & Roberts, J. H. (2019). Utility of eDNA and occupancy models for monitoring an endangered fish across diverse riverine habitats. *Hydrobiologia*, *826*(1), 129–144. https://doi.org/10.1007/s10750-018-3723-8

Tillotson, M. D., Kelly, R. P., Duda, J. J., Hoy, M., Kralj, J., & Quinn, T. P. (2018). Concentrations of environmental DNA (EDNA) reflect spawning salmon abundance at fine spatial and temporal scales. *Biological Conservation*, *220*, 1–11.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2024). *Loo: Efficient leave-one-out cross-validation and waic for bayesian models*. R package version 2.8.0. https://mc-stan.org/loo/

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved r for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718.

Yates, M. C., Glaser, D., Post, J. R., Cristescu, M. E., Fraser, D. J., & Derry, A. M. (2021). The relationship between EDNA particle concentration and organism abundance in nature is strengthened by allometric scaling. *Molecular Ecology*, *30*(13), 3068–3082.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1: Specifying prior distributions.

Appendix S2: Example workflow, including fitting the joint model with site-level covariates, performing model selection, summarizing posterior distributions, and performing a prior sensitivity analysis.

Appendix S3: Covariate analysis.

How to cite this article: Keller, A. G., & Kelly, R. P. (2025). eDNAjoint: An R package for interpreting paired or semi-paired environmental DNA and traditional survey data in a Bayesian framework. *Methods in Ecology and Evolution*, *00*, 1–9. https://doi.org/10.1111/2041-210X.70000