Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modeling

Running head: Estimating population abundance using eDNA

Keiichi Fukaya,[1,2*] Hiroaki Murakami,[3] Seokjin Yoon,[4] Kenji Minami,[5] Yutaka Osada,[6,7]
Satoshi Yamamoto,[8] Reiji Masuda,[3] Akihide Kasai,[4] Kazushi Miyashita,[9]
Toshifumi Minamoto,[10] and Michio Kondoh[6**]

[1] *National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan*

[2] *The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan*

[3] *Maizuru Fisheries Research Station, Field Science Education and Research Center, Kyoto University, Maizuru, Kyoto 625-0086, Japan*

[4] *Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Hokkaido 041-8611, Japan*

[5] *Estuary Research Center, Shimane University, Matsue, Shimane 690-8504, Japan*

[6] *Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan*

[7] *National Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan*

[8] *Laboratory of Animal Ecology, Department of Zoology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*

[9] *Field Science Center for Northern Biosphere, Hokkaido University, Hakodate, Hokkaido 040-0051, Japan*

[10] *Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo 657-8501, Japan*

*fukaya.keiichi@nies.go.jp

**michio.kondo.b8@tohoku.ac.jp

Abstract

Molecular analysis of DNA left in the environment, known as environmental DNA (eDNA), has proven to be a powerful and cost-effective approach to infer occurrence of species. Nonetheless, relating measurements of eDNA concentration to population abundance remains difficult because detailed knowledge on the processes that govern spatial and temporal distribution of eDNA should be integrated to reconstruct the underlying distribution and abundance of a target species. In this study, we propose a general framework of abundance estimation for aquatic systems on the basis of spatially replicated measurements of eDNA. The proposed method explicitly accounts for production, transport, and degradation of eDNA by utilizing numerical hydrodynamic models that can simulate the distribution of eDNA concentrations within an aquatic area. It turns out that,

under certain assumptions, population abundance can be estimated via a Bayesian inference of a generalized linear model. Application to a Japanese jack mackerel (*Trachurus japonicus*) population in Maizuru Bay revealed that the proposed method gives an estimate of population abundance comparable to that of a quantitative echo sounder method. Furthermore, the method successfully identified a source of exogenous input of eDNA (a fish market), which may render a quantitative application of eDNA difficult to interpret unless its effect is taken into account. These findings indicate the ability of eDNA to reliably reflect population abundance of aquatic macroorganisms; when the "ecology of eDNA" is adequately accounted for, population abundance can be quantified on the basis of measurements of eDNA concentration.

Introduction

Knowledge of the distribution and abundance of species is crucial for ecology and related applied fields such as wildlife management and fisheries. In particular, quantitative assessments are often required to effectively monitor and manage ecosystems because trends in environmental stressors such as climate change, habitat modification, and pollution can result in shifts in the distribution and the level of population abundance of species. Nevertheless, quantification of natural population of species can be challenging, if not impossible, at least due to the extensive effort required for field survey and the low detection probability of species or individuals (Yoccoz et al. 2001).

The detection and quantification of environmental DNA (eDNA) is an emerging methodology for ecological studies and could enhance the ability of investigators to infer occurrence and abundance of species. This approach has been applied, especially but not exclusively, to aquatic species such as fish and amphibians and has been identified as a powerful and yet cost-effective tool for species detection (Bohmann et al. 2014, Rees et al. 2014, Thomsen and Willerslev 2015, Goldberg et al. 2016, Deiner et al. 2017, Hansen et al. 2018).

Challenges remain, however, in quantitative applications of eDNA. Since earlier studies revealed positive correlations between species abundance and eDNA concentration (Takahara et al. 2012, Thomsen et al. 2012, Goldberg et al. 2013, Pilliod et al. 2013, Eichmiller et al. 2014), it has been expected that local population abundance may be inferred by measuring the concentration of eDNA at a given locality. Indeed, an analytical framework proposed for eDNA-based abundance estimation assumes a probability distribution that represents the quantitative relation between eDNA concentration and the underlying population size (Chambert et al. 2018). A recent empirical study showed that the abundance of anadromous fish in a river can be quantified based on frequent measurements of eDNA concentration, when streamflow is taken into consideration (Levi et al. 2019). Nonetheless, such a definite relation may not always be present, possibly depending on e.g., the shedding rate, transport, and exogenous input of eDNA (Pilliod et al. 2013, Eichmiller et al. 2014, Lacoursière-Roussel et al. 2016, Yamamoto et al. 2016, Jo et al. 2017), especially in natural environments as indicated by

3

a meta-analysis (Yates et al. 2019).

The fundamental factors that underlie such context dependency are the "ecology of eDNA": the distribution of eDNA in space and time stems from processes governing the origin, state, transport, and fate of eDNA particles (Barnes and Turner 2016). Thus, in applications of the eDNA methodology, detailed information about such processes may be critical. Without relevant knowledge of these processes, for example, the spatial and temporal scales of information provided by eDNA remain largely uncertain (Thomsen and Willerslev 2015, Goldberg et al. 2016, Hansen et al. 2018). Therefore, here, our purpose was to develop a general approach to eDNA-based abundance estimation that can fully account for the ecology of eDNA, i.e., the rate of production and degradation of eDNA as well as the transport of eDNA within a flow field in an aquatic area of interest (Figure 1). Although quantitative models of eDNA in which these processes are explicitly accounted for have been proposed for linear habitats such as rivers (Sansom and Sassoubre 2017, Carraro et al. 2018), no such model is currently available for general aquatic systems, including the marine environment.

In this study, we make use of a *tracer model*, namely, a numerical hydrodynamic model that can simulate the distribution of eDNA concentrations within an aquatic area (Shulman et al. 2003). Under certain assumptions, the behavior of the model can also be regarded mathematically as a linear function of an input vector representing the distribution of population abundance levels (densities) within the area. We show that the estimation of population abundance can then be achieved via a Bayesian inference of a generalized linear model (Figure 1). We applied this approach to a population of the Japanese jack mackerel (*Trachurus japonicus*, a commercially important fish species) in Maizuru Bay, Japan (Figure 2). On the basis of the eDNA concentration measurements and a tracer model configured for Maizuru Bay, we obtained an estimate of fish population abundance in the bay. This estimate was then verified via a parallel estimate of abundance obtained by a quantitative echo sounder method. The results suggest that the proposed approach can reliably quantify fish population abundance in the bay.

Materials and Methods

A general framework for abundance estimation

The tracer model as a linear function

Here, we define a *tracer model* as a numerical hydrodynamic model that simulates generation, transport, and decay of particles (i.e., eDNA) on the basis of a flow field determined by given physical conditions within an aquatic area of interest. In this study, we assume a tracer model for a three-dimensional discrete space in which the entire aquatic area of interest is discretized into grid cells of known volume. A tracer model can in principle simulate the ecology of eDNA and thus derives a spatial distribution of eDNA within the aquatic area, given that per capita and unit time shedding rates of eDNA, degradation rates of eDNA, and density (or equivalently, abundance) of organisms in each grid cell are specified, in addition to the flow field (Figure 1). The main idea that underlies the framework we propose is that we can regard a tracer model as a function that takes a vector of cell level density of organisms as an input and outputs eDNA concentration in each grid cell at a point in time; thus, the inference of abundance is an inverse problem: finding an input vector of a tracer model (i.e., density of organisms in each grid cell) that best explains measurements of eDNA concentration that are collected at a point in time and are replicated spatially within the aquatic area of interest.

Nevertheless, such a problem is difficult to solve under the general conditions where both the environment and abundance vary in a complex manner. We therefore make several key assumptions that simplify the problem (Figure 1). First, we assume that during two time points t and s ($< t$), key environmental variables for hydrodynamic processes are known from some observations and/or model prediction so that the flow field can be determined and plugged in to the tracer model. Here, t refers to the point in time at which eDNA concentration is observed at multiple locations within the aquatic area, and s denotes some point in time sufficiently far away from t such that eDNA concentration at t is virtually independent from that at s. Operationally, s and t define the time domain of the tracer model. Second, we assume that the rates of production and degradation of eDNA are known in each

¹¹² grid cell during the period between s and t. They may either be regarded as constant across

¹¹³ space and time or assumed to vary depending on known environmental variables, such as water

¹¹⁴ temperature, salinity, and pH, so that the rates of generation and disappearance of eDNA can

¹¹⁵ be determined completely in the tracer model. In addition, we assume that these rates are

¹¹⁶ independent of the eDNA concentration, and thus both production and degradation of eDNA

¹¹⁷ are linear processes. Third, we suppose that in each grid cell, all eDNA particles arise

¹¹⁸ exclusively from individuals of the target species that are identical in their eDNA-shedding

¹¹⁹ rate. Finally, we assume that abundance is stationary in each grid cell throughout the period

¹²⁰ between s and t (i.e., the demographic closure assumption; Williams et al. 2002).

¹²¹ Under these assumptions, a tracer model can be regarded as a linear function. We

¹²² denote density of organisms in cell i ($i = 1, \ldots, M$) by x_i and define $\boldsymbol{x} = (x_1, x_2, \ldots, x_M)$. Let

¹²³ us denote the water volume of each cell by $\boldsymbol{v} = (v_1, v_2, \ldots, v_M)$ so that abundance in cell i and

¹²⁴ in the whole aquatic area is expressed as $v_i x_i$ and $\boldsymbol{v}^\top \boldsymbol{x}$, respectively (here, $\boldsymbol{a}^\top$ means the

¹²⁵ transpose of vector $\boldsymbol{a}$). The tracer model predicts eDNA concentration in each grid cell at time

¹²⁶ point t that results from the generation, advection, diffusion, and degradation of eDNA

¹²⁷ occurring between s and t within a given flow field, which we denote (without an explicit index

¹²⁸ of t) by $\boldsymbol{c} = (c_1, c_2, \ldots, c_M)$. If a_{ij} is defined as the (per unit density) contribution of cell j to

¹²⁹ eDNA concentration in cell i at time t, then eDNA concentration can be expressed as

¹³⁰ $c_i = a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{iM} x_M$. If we designate $\boldsymbol{A} = (a_{ij})_{M \times M}$, then this equation can be

¹³¹ written in a matrix form as $\boldsymbol{c} = \boldsymbol{A}\boldsymbol{x}$. Thus, although a tracer model indeed represents

¹³² temporal evolution of eDNA concentration within the period between s and t according to

¹³³ some differential equations (presented below), its behavior can be described simply — under

¹³⁴ the assumptions noted above — by matrix $\boldsymbol{A}$, which maps the vector of density $\boldsymbol{x}$ onto the

¹³⁵ vector of eDNA concentration $\boldsymbol{c}$. For $i = 1, \ldots, M$, the ith column of $\boldsymbol{A}$ can be obtained

¹³⁶ numerically as a result of execution of the tracer model between time points s and t with a

¹³⁷ vector of density in which cell i has a unit density and all other cells have null density.

Fitting the tracer model to eDNA concentration data

We assume that eDNA concentration was measured in N samples collected within the aquatic area of interest at a point in time (or, in practice, within a sufficiently short period). Let us denote the observed eDNA concentration in sample n by y_n $(n = 1, \ldots, N)$ and express it with vector $\boldsymbol{y} = (y_1, \ldots, y_N)$. In the following text, we suppose that all eDNA measurements are positive (i.e., $y_n > 0$). Note, however, that negative samples could also be included in the analysis given that the detection process of eDNA is modeled jointly (Carraro et al. 2018). We define $i(n)$ as an index variable that means the index of the cell in which sample n was obtained. If we let $\boldsymbol{B} = (a_{i(n)j})_{N \times M}$, the prediction of the tracer model for the data vector, as a function of density vector $\boldsymbol{x}$ is then expressed as $\boldsymbol{Bx}$.

Because the tracer model yields a linear predictor for $\boldsymbol{y}$, we can apply the (generalized) linear modeling framework (McCullagh and Nelder 1989) to estimate density vector $\boldsymbol{x}$; in particular, we can regard $\boldsymbol{B}$ and $\boldsymbol{x}$ as a design matrix and a vector of coefficients of a linear regression model, respectively (note that because $\boldsymbol{x}$ represents density, the searches for estimates should be within the space of parameters such that $x_i \geq 0$ for all i). Considering that eDNA concentration data often represent a lognormal error structure (e.g., Takahara et al. 2012, Thomsen et al. 2012, Eichmiller et al. 2014, Wilcox et al. 2016, Jo et al. 2017), we can consider the following model:

$$\log \boldsymbol{y} \sim \mathcal{N} \left(\log \boldsymbol{Bx}, \sigma^2 \boldsymbol{I}_N \right), \tag{1}$$

where $\mathcal{N} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$; σ^2 is a residual variance, and $\boldsymbol{I}_m$ is a $m \times m$ identity matrix. Note that this is a generalized linear model with an exponential link function (McCullagh and Nelder 1989). The maximum likelihood method for this model yields an estimate $\hat{\boldsymbol{x}}$ that minimizes the residual square error $|\log \boldsymbol{y} - \log \boldsymbol{Bx}|_2^2$.

The standard maximum likelihood approach is, however, not applicable to this model when $M > N$ because the maximum likelihood estimate of $\boldsymbol{x}$ is not uniquely identified in this setting. This may be a typical situation at a reasonable level of spatial discretization for the

7

tracer model and sampling effort of eDNA. Additional assumption (or regularization) on the regression coefficients is necessary to make an inference based on such a singular model. We can apply a Bayesian approach in which a common probability distribution is specified on $\boldsymbol{x}$ as a prior distribution. This approach effectively specifies $\boldsymbol{x}$ as random effects, allowing to "borrow strength" to estimate each element of $\boldsymbol{x}$ (Kéry and Schaub 2012). Alternatively, when some covariates, assumed to covary with density, are available for each cell, an additional model component for density can be introduced to effectively reduce the number of unknown parameters (Carraro et al. 2018). Specifically, density of the target species can be modeled, for example, as $\log \boldsymbol{x} = \boldsymbol{Z}\boldsymbol{\beta}$, where $\boldsymbol{Z}$ is a matrix of covariates, and $\boldsymbol{\beta}$ is a vector of coefficients (including an intercept). Note, however, that the addition of a nonlinear component for density takes the model beyond the standard generalized linear modeling framework.

An application to a marine fish population

Measurement of jack mackerel eDNA in Maizuru Bay

The study was conducted in Maizuru Bay (Kyoto prefecture, Japan; 35°29′N, 135°23′E) to estimate abundance of the jack mackerel (*T. japonicus*) via concentration of eDNA. The bay has a surface area of $\sim$22.87 km^2 with a maximum water depth of approximately 30 m, and connects with Wakasa Bay through a narrow bay mouth in its north (Figure 2).

According to long-term underwater visual surveys, the jack mackerel is numerically the most dominant fish species in shallow (< 10 m in depth) coastal waters in this area (Masuda 2008); their body size ranges from 10 to 45 mm in standard length offshore and 40–120 mm standard length in the shallow rocky reef habitat (Masuda et al. 2008). The study was conducted during the peak season of jack mackerel recruitment from the offshore pelagic zone to a coastal shallow reef habitat, where the jack mackerel population in the bay is dominated by new recruits. In the following analysis, we therefore assumed that the population is represented by individuals of size $\sim$3 cm (body length) and $\sim$1 g (body weight; see Appendix S1, Supporting information).

We conducted the water sampling on 21 and 22 June 2016 from a research vessel at 100 stations located approximately on $\sim$400 m grids in Maizuru Bay (Figure 2). At each sampling

station, we captured a 1-L sea water sample at three depths: the surface, middle, and bottom. The middle and bottom depths were defined as 5 m from the surface, which was just below the pycnocline, and 1 m above the sea floor, respectively. We filtered water samples on the same day of the field survey, which were then subjected to eDNA extraction and subsequent quantitative PCR (qPCR). qPCR was carried out in triplicate for each sample. Details of the measurement of eDNA concentration are fully described in Appendix S2, Supporting information.

Development of the eDNA tracer model

To obtain the flow field in Maizuru Bay, we configured the Princeton Ocean Model (POM) with a scaled vertical coordinate (i.e., the sigma coordinate system, in which the number of layers of the water column is the same for every grid irrespective of the sea depth; Mellor 2002) for the bay. The model represented Maizuru Bay by 20,484 grid cells. Specifically, the bay was discretized by 2,276 horizontal lattice grids at a resolution of 100 m, and the grids had nine non uniform vertical layers, with finer resolution near the surface; the boundary of each layer in the sigma coordinate was set as $\sigma = 0.000, -0.041, -0.088, -0.150, -0.245, -0.374,$ $-0.510, -0.646, -0.796,$ and -1.000. The configuration of the model was achieved by means of the bottom topography of the bay, data and model estimates of surface meteorological conditions, estimated river discharges, and the model results of Wakasa Bay as the open boundary conditions (Yoon and Kasai 2017); additional details are described in Appendix S3. The model simulated flow fields within the bay from 1 June 2016, under the initial conditions interpolated from the model results of Yoon and Kasai (2017), to the final day for the water sampling (i.e., 22 June 2016). The time steps of the simulation were set to 0.1 s for the external mode to update the surface elevation and vertically averaged velocities, whereas they were 3 s for the internal mode to update the horizontal velocities, potential temperature, salinity, and turbulence quantities (Mellor 2002).

The concentration of eDNA of jack mackerels was then simulated based on the flow fields produced by the POM. The evolution of eDNA concentration, denoted by c, is represented as

9

$$\frac{\partial c}{\partial t} + u\nabla(c) = \nabla^*(d\nabla(c)) - \lambda c + \beta x, \tag{2}$$

where u is a three-dimensional velocity vector, $\nabla(c)$ is the gradient of c, $\nabla^*(\cdot)$ is the divergence operator, d is a vector of diffusion coefficients, λ is a degradation rate of eDNA, β is a per-capita shedding rate of jack mackerel DNA, and x is the density of jack mackerels in the cell. The eDNA degradation rate was assumed to be constant and was adopted from an estimate obtained in tank experiments where the same species-specific primer set was employed ($\lambda = 0.044$ h^{-1}; Jo et al. 2017). The eDNA shedding rate of the jack mackerel was assumed to be constant; it was derived mathematically and found to be $\beta = 9.88 \times 10^4$ copies per individual per hour, according to the results of tank experiments conducted by Maruyama et al. (2014) and Jo et al. (2017). Details of this derivation are provided in Appendix S4, Supporting information. The settling and resuspension of eDNA particles were ignored in Eq. 2 given that little is known about the rate of these processes.

For each of the 20,484 columns of the $\boldsymbol{A}$ matrix, the tracer model was executed from 1 June 2016 to 22 June 2016 to generate the values of the column elements. The elements in $\boldsymbol{A}$ were specified by the daily averages on 22 June and then used in the subsequent analyses.

Estimation of jack mackerel abundance based on the eDNA tracer model

We fitted the model with lognormal error distribution (Eq. 1) to the eDNA concentration data collected in Maizuru Bay. During the model fitting, we omitted negative samples in which the number of remaining observations was $N = 729$. For vector of density $\boldsymbol{x}$, we specified an independent lognormal prior with unknown prior mean μ and standard deviation τ:

$$\log \boldsymbol{x} \sim \mathcal{N}\left(\log\mu\mathbf{1}_M, \tau^2\boldsymbol{I}_M\right), \tag{3}$$

where $\mathbf{1}_M$ represents a vector of all ones with length M. Because N was significantly smaller than M, we were pessimistic about estimating the spatial variation in cell level density with reasonable precision. Our main goal of the inference was therefore to quantify the bay-level abundance $\boldsymbol{v}^\top\boldsymbol{x}$ along with its uncertainty.

10

With uniform positive priors on μ, τ, and σ, we fitted the model via a fully Bayesian approach. Posterior samples were obtained by the Markov chain Monte Carlo (MCMC) method implemented in Stan (version 2.18.1; Carpenter et al. 2017) in which three independent chains of 10,000 iterations were generated after 1,000 warm-up iterations. Each chain was thinned at intervals of 10 to save the posterior sample.

Convergence of the posterior was checked for each parameter with the $\hat{R}$ statistic. Posterior convergence was achieved at a recommended degree ($\hat{R} < 1.1$; Gelman et al. 2013) in almost all parameters except $\log x$ in four cells. We decided, however, that the results are solid because the posterior of the bay-level abundance — the target of the inference — fully converged. The goodness-of-fit assessment of the model, measured by the χ^2-discrepancy statistic (Conn et al. 2018), gave no clear indication of a lack of model fit (Bayesian p-value: 0.404).

We note that the independence assumption of the fitted model (Eq. 1) ignores the correlation between triplicates within each water sample, resulting in a potential underestimation of the uncertainty in the estimates of jack mackerel abundance. Although such a within-sample correlation could be accommodated in the model by explicitly accounting for the covariance structure of data, we here adopt the independence assumption for the purpose of illustrating the proposed framework.

Estimation of jack mackerel abundance from quantitative echo sounder data

An independent estimate of jack mackerel abundance was obtained based on a calibrated quantitative echo sounder by a standard acoustic survey method (Simmonds and MacLennan 2005). The acoustic survey was conducted during the survey cruise for the water sampling (described above). We used the KSE300 echo sounder (Sonic Co. Ltd., Tokyo, Japan) with two transducers (T-182, 120 kHz, and T-178, 38 kHz; beam type, split-beam; beam width, 8.5°; pulse duration, 0.3 ms; ping rate, 0.2 s), which were mounted off the side of the research vessel at a depth of 1 m. The acoustic devices were operated during the entire survey cruise to record all acoustic reflections, except when the research vessel stopped at each sampling station where the recording was stopped to avoid reflection from the sampling gear and cables.

11

271 The research vessel ran at ∼4 knots, on average, between the sampling stations. The echo

272 intensity data were denoised and cleaned in Echoview ver. 9.0 (Echoview Software Pty. Ltd.,

273 Tasmania, Australia). We omitted signals between the sea bottom and 0.5 m above it to

274 exclude the acoustic reflection from the sea floor. Additionally, we eliminated signals from sea

275 nettles (*Chrysaora pacifica*) by filtering reflections of −75 dB.

276   From the obtained acoustic data, the reflections of jack mackerel were extracted by the

277 volume back scattering strength difference (ΔS_V) method (Miyashita et al. 2004, Simmonds

278 and MacLennan 2005). ΔS_V was defined as the difference in the volume backscattering

279 strength (S_V) between the two frequencies as follows:

$$\Delta S_V = S_{V\,120\text{ kHz}} - S_{V\,38\text{ kHz}}. \tag{4}$$

280 According to field validation in Maizuru Bay combining acoustic surveys and visual

281 confirmation of jack mackerel schools by snorkelling, we assumed the range of ΔS_V of jack

282 mackerel between −6.4 and 5.2 dB. This criterion discriminates the jack mackerel from larval

283 Japanese anchovy (*Engraulis japonicus*), the subdominant species in the bay (Masuda 2008),

284 which reflects the high frequency echo strongly as compared to low frequency (Ito et al. 2011)

285 and was used to determine S_V of the jack mackerel in 1 m^3 water cubes. S_V values were

286 extracted for every 10 m segment of the survey line for every 1 m of depth using Echoview ver.

287 9.0.

288   Density of jack mackerel in a 1 m^3 water cube, denoted by D, was obtained as

$$D = \frac{10^{\frac{S_{V\,120\text{ kHz}}}{10}}}{10^{\frac{TS}{10}}} \tag{5}$$

289 where TS is the target strength of an individual jack mackerel. By assuming that jack

290 mackerel population in the bay was dominated by individuals of the size 3 cm, we chose

291 $TS = -59.6$ dB (Nakamura et al. 2013, Yamamoto et al. 2016). The fish density for each 10 m

292 segment and 1 m of depth on the echo sounder track lines was then matched with the grid

293 specification of the tracer model.

294   To obtain an estimate (and its associated uncertainty) of the bay-level abundance of

jack mackerel that can be compared to that obtained with the eDNA-based approach, a linear mixed model was fitted to the fish density data via a fully Bayesian approach. Details of the abundance estimation for quantitative echo sounder data are described in detail in Appendix S5, Supporting information.

Results

The proposed method yielded an estimate of fish abundance that was more than two times higher than the quantitative echo sounder estimate; although the point estimate of the eDNA method was of the same order of magnitude as the quantitative echo sounder estimate, there was no overlap in the two 95% highest posterior density intervals (HPDIs) (Table 1). In the eDNA method, however, we could identify a coordinate of grids in which density of jack mackerels was estimated to be unrealistically high; fish abundance in the nine vertical cells in this location was estimated at as much as tens of millions of individuals (posterior median and 95% HPDI: 1.35×10^7 [0.00 to 1.77×10^7] individuals; Figure 2b). It is located next to a wholesale fish market (Figure 2a), which has been suspected as a significant source of exogenous jack mackerel eDNA in Maizuru Bay (Yamamoto et al. 2016, Jo et al. 2017). We therefore regarded the extreme estimates in these cells as resulting from a massive eDNA input from the market and excluded them from the inference of the bay-level fish abundance. This correction for the eDNA method reduced the estimate of fish abundance in the bay. As a result, the abundance estimates of the two methods become more comparable; the estimate corrected for the eDNA method was 1.42 times higher than the estimate via the quantitative echo sounder method, with the two 95% HPDIs overlapping (Table 1).

The eDNA concentration spatial distribution predicted by the fitted model corresponded well to the observed values (posterior median and 95% HPDI of the correlation coefficient: 0.652 [0.638 to 0.664]; Figures 3 and 4). Nevertheless, there was a high level of uncertainty in the estimate of fish density in each grid cell (Figure 2b), as expected given the small number of samples relative to the number of grid cells. As a result, neither a discernible spatial pattern of fish density nor a clear correlation between the estimates of the two methods was evident at the cell scale (Figure 5 and Supplementary Table S1).

13

To examine the robustness of the inference, we conducted an additional analysis in which fish density was estimated in coarser resolutions while using the same tracer model. Specifically, we fitted three model variants wherein coarser grid blocks are defined by aggregating horizontally neighboring cells at intervals of 200, 300, or 400 m, respectively, for each of which fish density was estimated (Supplementary Figure S1). The three model variants yielded an estimate of the bay-level fish abundance that are close to the original estimate (Supplementary Table S1). Moreover, they indicated an extremely high fish density in the grid blocks that are next to the fish market (Supplementary Figure S1). The estimates of the bay-level fish abundance, corrected for the fish market, were also similar to the original estimate (Supplementary Table S1). A better correlation was observed in coarser grid blocks between fish density estimated with the eDNA method and that estimated with the quantitative echo sounder method (Supplementary Table S1).

Discussion

The eDNA methods are rapidly developing technologies that have a great potential to facilitate the understanding and management of aquatic species, although their quantitative applications are still the critical step. This study presents a novel approach to abundance estimation based on quantitative eDNA measurements into which a numerical tracer model is incorporated to explicitly account for the details of the ecology of eDNA (Figure 1). Briefly, it requires the following steps: (1) Develop a hydrodynamic model for an aquatic area of interest. (2) Collect spatially replicated samples at a point of time from the aquatic area to measure eDNA concentration. (3) Configure the tracer model for the field measurement of eDNA; this will require knowledge of the rate of eDNA shedding and degradation. Then, run the tracer model repeatedly to obtain a design matrix of a generalized linear model that relates the underlying density to observed eDNA concentration. (4) Fit the generalized linear model to eDNA concentration data statistically, to estimate density of organisms for each discretized grid cell. This approach may be flexibly applied to a wide array of aquatic systems in which hydrodynamics and rates of eDNA shedding and degradation are modeled, thereby broadening the scope of the general idea implemented recently in a one-dimensional lotic system with a

single eDNA source (Sansom and Sassoubre 2017, Levi et al. 2019) and in a river network system with multiple eDNA sources (Carraro et al. 2018).

The application of the proposed approach to the Japanese jack mackerel population in Maizuru Bay indicated that abundance of species can be reliably estimated by means of eDNA in a coastal ecosystem, where oceanographic processes drive transportation of eDNA (Andruszkiewicz et al. 2019). Furthermore, the results revealed that the method can distinguish major exogenous sources of eDNA, which have been recognized as a nuisance factor in eDNA applications especially for species subject to fishery (Yamamoto et al. 2016, Jo et al. 2017). These results suggest that when the processes of eDNA shedding, transport, and degradation were properly accounted for, an absolute estimation of abundance of aquatic macroorganisms can be practically achieved based on quantitative measurement of eDNA.

The proposed framework, however, has several limitations in its current form. For example, it requires several key assumptions, such as the stationarity (i.e., demographic closure) of the population and homogeneity of individuals in terms of their rate of eDNA shedding (Figure 1). This implies that the inferences can be biased by fluctuations in the spatial distribution of abundance, as well as heterogeneity in the shedding rate of eDNA and/or other factors that deviate from the modeling assumptions, if such factors are present. In addition, the number of eDNA samples may typically be smaller than the number of grid cells in the tracer model, thus requiring additional models describing among-cell variation in population density to make statistical inference (see *Materials and Methods*). Although our results indicated that the method can be applied even with these limitations, further methodological development would be warranted.

Ambiguity in the modeled processes could add further bias and uncertainty in the inferences based on a tracer model. The tracer model depends on the specified values for the rates of eDNA shedding and degradation, which have been quantified experimentally (e.g., Maruyama et al. 2014, Sassoubre et al. 2016, Jo et al. 2017, Nukazawa et al. 2018, Jo et al. 2019); yet our knowledge is still limited to predict them accurately, and even less is known about the difference between these rates in field and laboratory environments. Despite being ignored in our application, gravitational settling (and possibly, resuspension) of eDNA

particles may even have a significant effect on the spread of eDNA in a coastal ecosystem (Andruszkiewicz et al. 2019). However, little is known about the rates of these processes. Because the research on aquatic eDNA of macroorganisms is still in its infancy since its discovery (Ficetola et al. 2008), more work is needed to elucidate the processes that determine a distribution of eDNA in the field; a better understanding of the ecology of eDNA will improve the accuracy of quantitative eDNA approaches.

In our application, the local variation in fish density was not estimated reasonably; this is probably because the number of grid cells was too large relative to the number of eDNA samples. Given the logistical constraints in collecting eDNA samples, this may occur frequently in practical applications. A solution for better inference of local fish density may be to aggregate neighboring grid cells to yield coarser spatial units on which specific density is estimated. This approach may not only reduce the number of unknown quantities in the model, but it also may increase the likelihood that the assumption of population closure is met. In fact, we observed in our additional analysis an improved consistency in the estimated fish density obtained based on eDNA and the acoustic survey (Supplementary Table S1). Although it is computationally more challenging, specifying a prior distribution that explicitly accounts for spatial autocorrelation in fish density could be useful for obtaining smoothed estimates. When some relevant covariates are available for each spatial unit, spatial variation in density may be inferred more explicitly through an additional model component for local density (see *Materials and Methods*). As a possible alternative approach, eDNA measurements could be combined with classical protocols for abundance estimation to improve the inference on spatial variation in density (Chambert et al. 2018). However, it will require a further generalization of the model to formally accommodate different types of observations.

It has been argued that in an application of the eDNA method, careful consideration of details of the ecology of eDNA is critical (Bohmann et al. 2014, Rees et al. 2014, Thomsen and Willerslev 2015, Barnes and Turner 2016, Goldberg et al. 2016, Deiner et al. 2017, Hansen et al. 2018). We implemented this idea in a quantitative eDNA method, leading to integration of eDNA concentration measurements and hydrodynamic modeling for abundance estimation. The relatively less explored field of quantitative eDNA applications lies in the multispecies

16

context, which involves eDNA metabarcoding rather than the targeted quantitative PCR (qPCR) method (Deiner et al. 2017). The proposed approach would be applicable to the emerging quantitative metabarcoding technique (e.g., Ushio et al. 2018) in the same way, which may enable researchers to analyze many aquatic species at one time. Exploring between-species differences in the rate of eDNA shedding and degradation may therefore be worthwhile. In addition to remarkable efficiency in species detection, we expect that eDNA methodologies can enhance the ability of investigators to gain quantitative insights into aquatic ecosystems.

Acknowledgements

References

Andruszkiewicz, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., and Boehm, A. B. (2019). Modeling environmental DNA transport in the coastal ocean using Lagrangian particle tracking. *Frontiers in Marine Science*, 6:477.

Barnes, M. A. and Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17:1–17.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Douglas, W. Y., and de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29:358–367.

17

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

Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E., and Rinaldo, A. (2018). Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*, 115:11724–11729.

Chambert, T., Pilliod, D. S., Goldberg, C. S., Doi, H., and Takahara, T. (2018). An analytical framework for estimating aquatic species density from environmental DNA. *Ecology and Evolution*, 8:3468–3477.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88:526–542.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., and Bernatchez, L. (2017). Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Molecular Ecology*, 26:5872–5895.

Eichmiller, J. J., Bajer, P. G., and Sorensen, P. W. (2014). The relationship between the distribution of common carp and their environmental DNA in a small lake. *PloS ONE*, 9:e112611.

Ficetola, G. F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4:423–425.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J., and Waits, L. P. (2013). Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32:792–800.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., and Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7:1299–1307.

Hansen, B. K., Bekkevold, D., Clausen, L. W., and Nielsen, E. E. (2018). The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries*, 19:751–768.

Ito, Y., Yasuma, H., Masuda, R., Minami, K., Matsukura, R., Morioka, S., and Miyashita, K. (2011). Swimming angle and target strength of larval Japanese anchovy (*Engraulis japonicus*). *Fisheries Science*, 77:161–167.

Jo, T., Murakami, H., Masuda, R., Sakata, M. K., Yamamoto, S., and Minamoto, T. (2017). Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*, 17:e25–e33.

Jo, T., Murakami, H., Yamamoto, S., Masuda, R., and Minamoto, T. (2019). Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. *Ecology and Evolution*, 9:1135–1146.

Kéry, M. and Schaub, M. (2012). *Bayesian Population Analysis using WinBUGS: A Hierarchical Perspective*. Academic Press.

Lacoursière-Roussel, A., Rosabal, M., and Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources*, 16:1401–1414.

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

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., and Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PLoS ONE*, 9:e114639.

Masuda, R. (2008). Seasonal and interannual variation of subtidal fish assemblages in Wakasa Bay with reference to the warming trend in the Sea of Japan. *Environmental Biology of Fishes*, 82:387–399.

Masuda, R., Yamashita, Y., and Matsuyama, M. (2008). Jack mackerel *Trachurus japonicus* juveniles use jellyfish for predator avoidance and as a prey collector. *Fisheries Science*, 74:276–284.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC, 2nd edition.

Mellor, G. L. (2002). *Users Guide for A Three-Dimensional, Primitive Equation, Numerical Ocean Model*. (Available at http://www.ccpo.odu.edu/POMWEB).

Miyashita, K., Tetsumura, K., Honda, S., Oshima, T., Kawabe, R., and Sasaki, K. (2004). Diel changes in vertical distribution patterns of zooplankton and walleye pollock (*Theragra chalcogramma*) off the Pacific coast of eastern Hokkaido, Japan, estimated by the volume back scattering strength (S_V) difference method. *Fisheries Oceanography*, 13:99–110.

Nakamura, T., Hamano, A., Abe, K., Yasuma, H., and Miyashita, K. (2013). Acoustic scattering properties of juvenile jack mackerel *Trachurus japonicus* based on a scattering model and *ex situ* target strength measurements. *Nippon Suisan Gakkaishi*, 79:345–354. (in Japanese with English abstract)

Nukazawa, K., Hamasuna, Y., and Suzuki, Y. (2018). Simulating the advection and degradation of the environmental DNA of common carp along a river. *Environmental Science and Technology*, 52:10562–10570.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., and 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*, 70:1123–1130.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., and Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51:1450–1459.

Sansom, B. J. and Sassoubre, L. M. (2017). Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environmental Science and Technology*, 51:14244–14253.

Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., and Boehm, A. B. (2016). Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environmental Science and Technology*, 50:10456–10464.

Shulman, I., Haddock, S. H., McGillicuddy Jr, D. J., Paduan, J. D., and Bissett, W. P. (2003). Numerical modeling of bioluminescence distributions in the coastal ocean. *Journal of Atmospheric and Oceanic Technology*, 20:1060–1068.

Simmonds, J. and MacLennan, D. N. (2005). *Fisheries Acoustics: Theory and Practice*. John Wiley and Sons, 2nd edition.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., and Kawabata, Z. (2012). Estimation of fish biomass using environmental DNA. *PloS ONE*, 7:e35868.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., and Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21:2565–2573.

Thomsen, P. F. and Willerslev, E. (2015). Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183:4–18.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T., and Kondoh, M. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding and Metagenomics*, 2:e23297.

Wilcox, T. M., McKelvey, K. S., Young, M. K., Sepulveda, A. J., Shepard, B. B., Jane, S. F., Whiteley, A. R., Lowe, W. H., and Schwartz, M. K. (2016). Understanding environmental DNA detection probabilities: a case study using a stream-dwelling char *Salvelinus fontinalis*. *Biological Conservation*, 194:209–216.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press.

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T., Hongo, M., Nishida, J., Okugawa, Y., Fujiwara, A., Fukuda, M., Hidaka, S., Suzuki, K. W., Miya, M., Araki, H., Yamanaka, H., Maruyama, A., Miyashita, K., Masuda, R., Minamoto, T., and Kondoh, M. (2016). Environmental DNA

548 as a ò snapshotŐ of fish distribution: a case study of Japanese jack mackerel in Maizuru
549 Bay, Sea of Japan. *PLoS ONE*, 11:e0149786.

550 Yates, M. C., Fraser, D. J., and Derry, A. M. (2019). Meta-analysis supports further
551 refinement of eDNA for monitoring aquatic species-specific abundance in nature.
552 *Environmental DNA*, 1:5–13.

553 Yoccoz, N. G., Nichols, J. D., and Boulinier, T. (2001). Monitoring of biological diversity in
554 space and time. *Trends in Ecology and Evolution*, 16:446–453.

555 Yoon, S. and Kasai, A. (2017). Relative contributions of external forcing factors to circulation
556 and hydrographic properties in a micro-tidal bay. *Estuarine, Coastal and Shelf Science*,
557 198:225–235.

Data accessibility

Authors' contributions

M.K., R.M., A.K., K. Miyashita, and T.M. conceived and designed the eDNA survey. H.M.
and S. Yamamoto conducted the molecular experiments. K. Minami and K. Miyashita
analyzed the echo sounder data. S. Yoon and A.K. developed the tracer model. K.F. and Y.O.
designed the methodology and conducted data analyses. K.F., H.M., S. Yoon, K. Minami, and
A.K. led the writing of the manuscript. All the co-authors discussed the results and
contributed critically to the manuscript.

Tables and figures

Method	Abundance estimates	95% Bayesian credible interval
eDNA tracer model	3.31×10^7	$(2.32 \times 10^7, 6.32 \times 10^7)$
(fish market cells omitted)	2.23×10^7	$(0.77 \times 10^7, 5.29 \times 10^7)$
Quantitative echo sounder	1.57×10^7	$(1.51 \times 10^7, 1.64 \times 10^7)$

Table 1. Estimates of Japanese jack mackerel abundance in Maizuru Bay. The second row of the eDNA method gives the abundance estimate that excluded the nine vertical cells next to the wholesale fish market (indicated in Figure 2a), which were identified as extraordinary eDNA sources. The point abundance estimates and credible intervals are presented as posterior medians and highest posterior density intervals, respectively. In both estimation methods, estimates are obtained under the assumption that the size of jack mackerel individuals was 3 cm in body length and 1 g in body weight (see Materials and Methods).
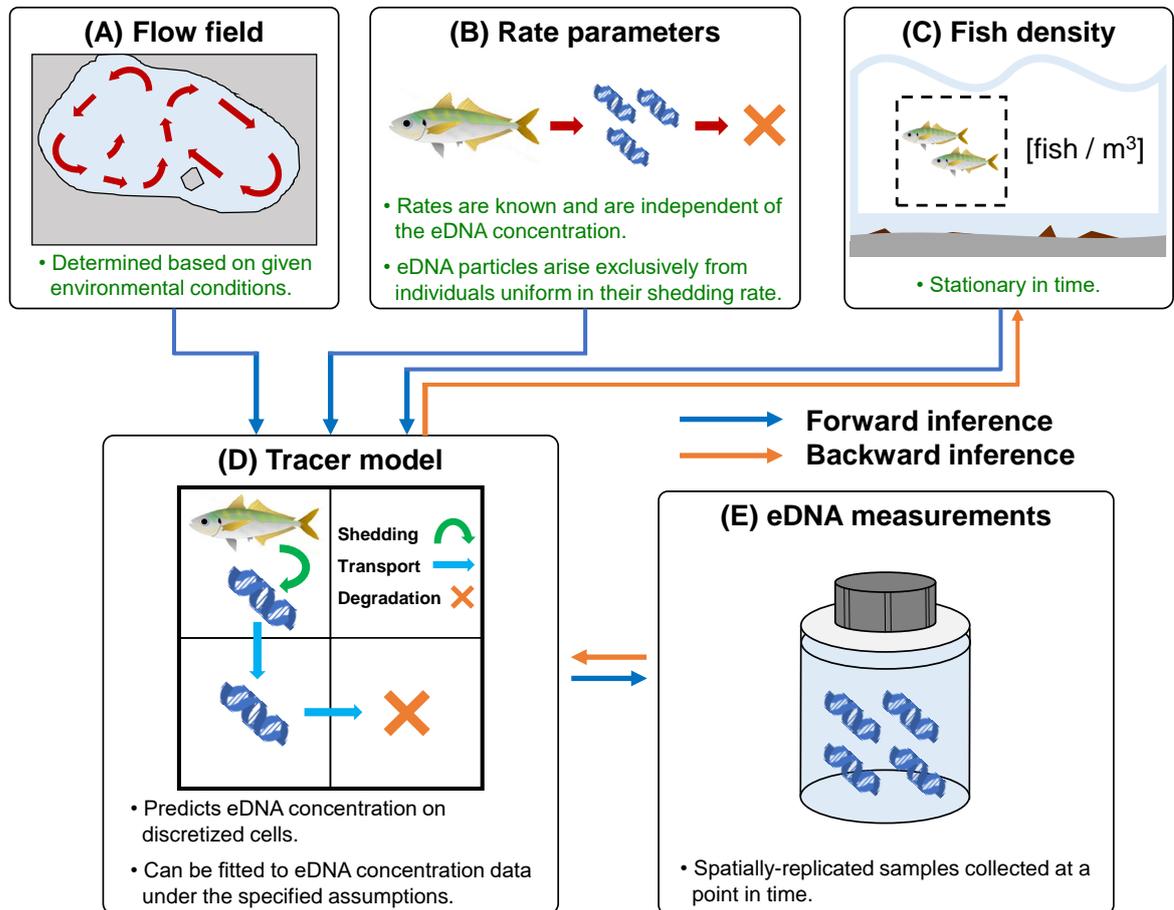
Fig. 1. Conceptual representation of the modeling framework. A tracer model (D) predicts the spatial distribution of eDNA concentration (E) within an aquatic area of interest under the specified flow field (A), rates of eDNA shedding and degradation (B), and spatial distribution of fish density (C). This is a forward inference of the model (blue arrows) represented by the governing equation (Eq. 2). Population density can be estimated by fitting the tracer model statistically to eDNA concentration measurements collected at a point of time within the modeled aquatic domain. This is a backward inference of the model (orange arrows) that can be achieved under the specified flow field and rate parameters in addition to some key assumptions described by green letters.
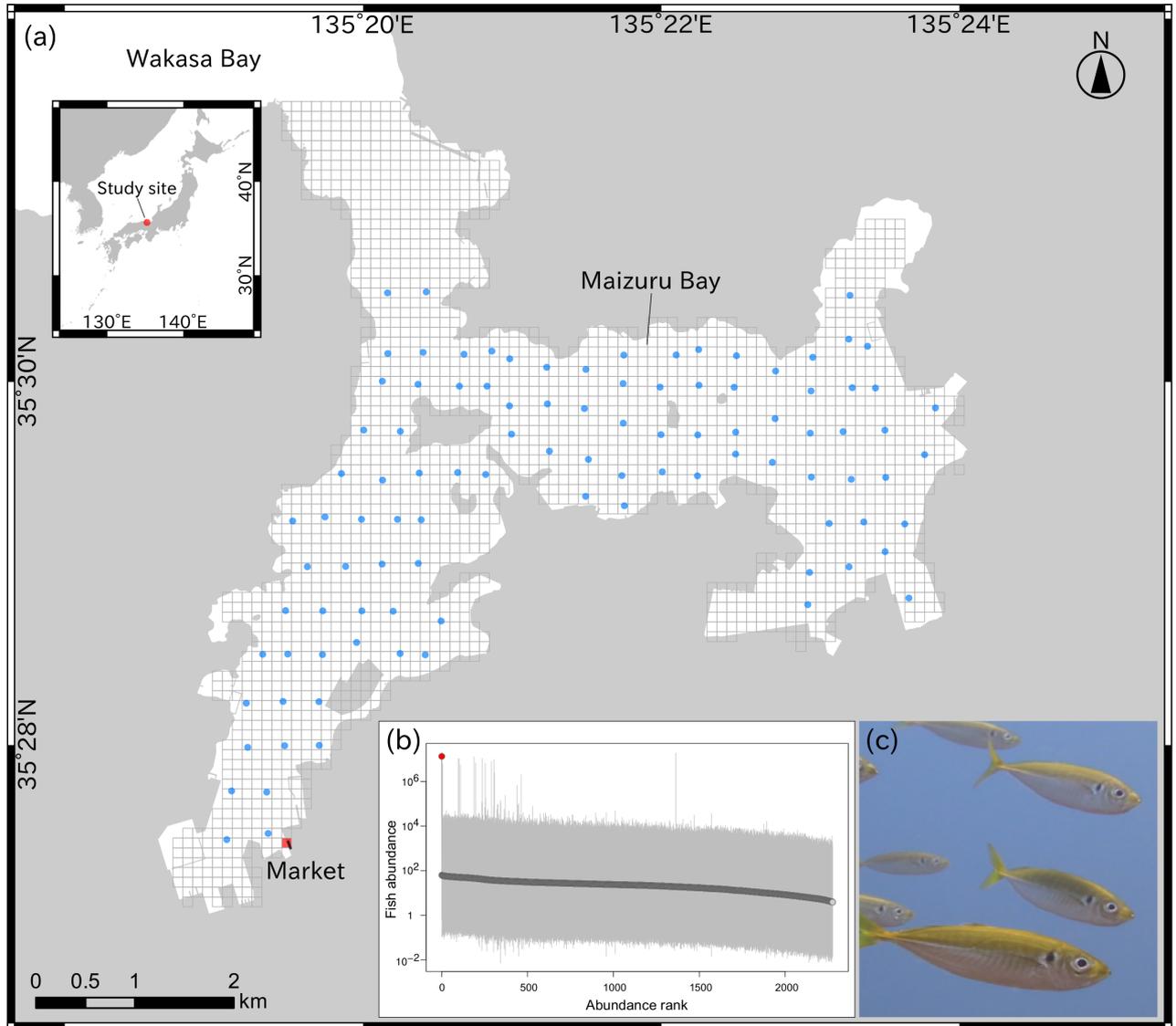
Fig. 2. Maizuru bay, the study site. (a) The 2,276 horizontal lattice grids for the eDNA tracer model (grey boxes) and the 100 water-sampling stations (blue circles). The grid in which estimates of jack mackerel density were extremely high is highlighted in red. The building of the fish market, overlapping with the red lattice grid, is depicted by a filled black box. (b) Fish abundance estimates (circles: posterior medians; bars: 95% highest posterior density intervals) in the 2,276 horizontal lattice grids. Abundance estimates in nine vertical cells were pooled for each grid. The lattice grid next to the market is highlighted in red. (c) The Japanese jack mackerel (*T. japonicus*) in Maizuru bay (photo credit: R. Masuda).
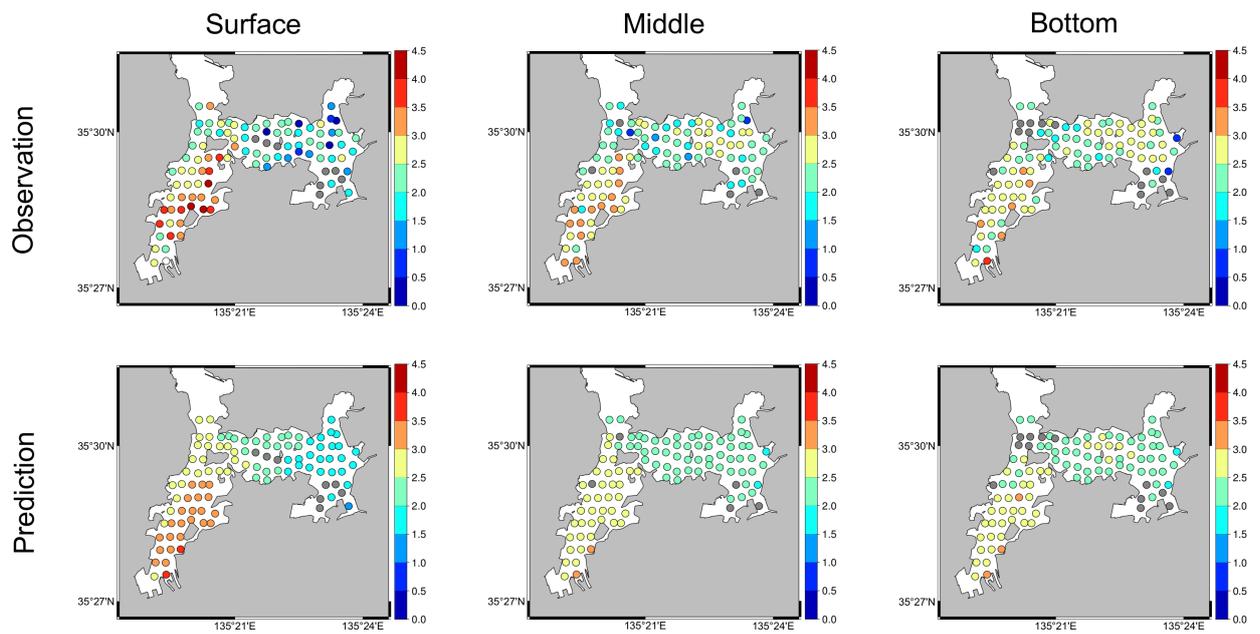
Fig. 3. Observed (upper panels) and predicted (lower panels) spatial distribution of eDNA concentration in 100 sampling locations. Left, center, and right panels show eDNA concentration at the surface, middle, and bottom layer, respectively. The color of each point represents the logarithm of eDNA concentration (copies/L) to base 10. The average concentration of positive samples was used for observed values. Posterior medians were used for predicted values. For visualization purposes, data collected from the surface layer of a station next to a wholesale fish market were omitted. Gray points indicate stations with only negative samples.
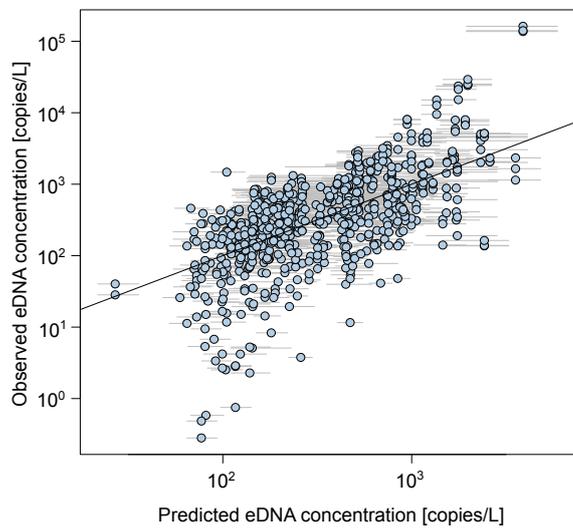
Fig. 4. Posterior predicted value and observed value of eDNA concentration in all samples. For predicted values, the posterior medians and 95% highest posterior density intervals are represented by circles and horizontal bars, respectively. The crossed line is the identity line.

(a) eDNA tracer model

(b) Quantitative echo sounder
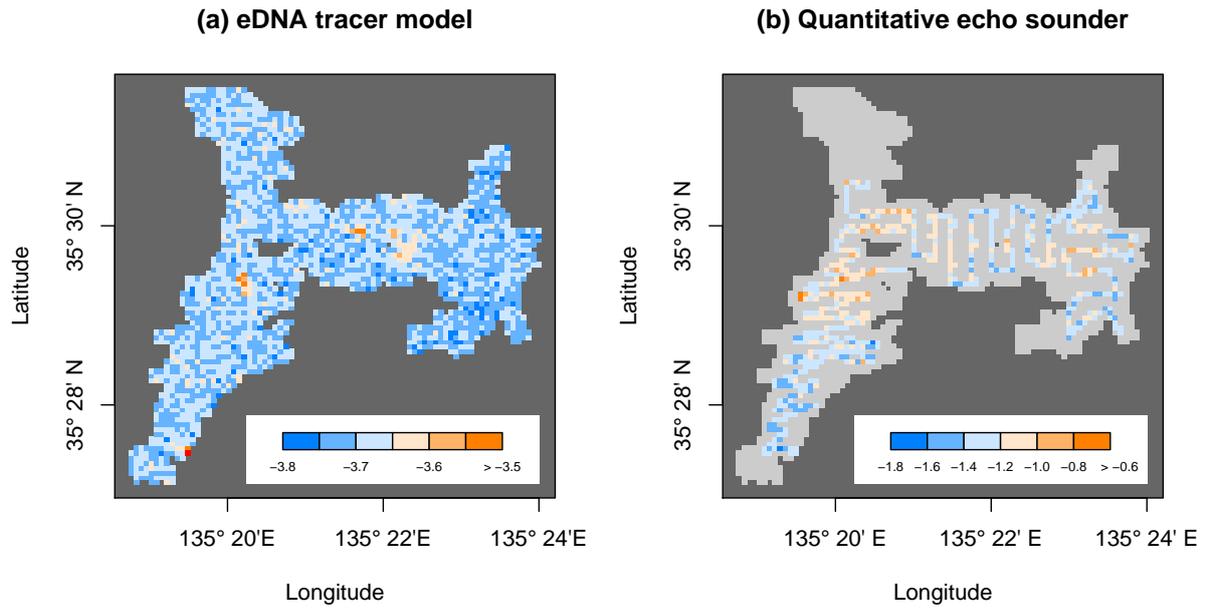
Fig. 5. Spatial distribution of density of Japanese jack mackerel estimated using (a) the proposed method and (b) the quantitative echo sounder method. Each panel shows the estimated density (posterior median) in the horizontal lattice grids, in which the colors represent the logarithm of fish density (individuals per 1 m^3) to base 10. Note the difference in the color scales. The uppermost color categories include a small number of outliers. In panel (a), the grid cell next to the fish market with an extremely high estimated density is highlighted in red. In panel (b), the horizontal lattice grids without quantitative echo sounder data are shown in gray.