Title: Inetgrated spatial model estimates the fish distribution using environmental DNA and catch data

Yuki Kanamori[1*], Hiroshi Okamura[2], Shota Nishijima[2], Yuki Hongo[2], Yasuyuki Uto[3], Hisatoku Mita[4], Mitsuhiyo Ishii[4], Kiyoharu Akimoto[5], and Akane Kusano[6]

[1] Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 25-259 Shimomekurakubo, Samemachi, Hachinohe, Aomori 031-0841, Japan
[2] Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan
[3]
[4]
[5]
[6]

[*] Corresponding author
Email: kana.yuki@fra.affrc.go.jp

Abstract

Keywords

1 Introduction

Understanding of spatial distribution of species and underling its mechanism is a essential issue in ecology. Field surveys using environmental DNA (eDNA) are widely used for detecting invasive or rare species and hotspot of biodiversity (面倒なのでレビュー論文を引用) because the surveys of eDNA are easy to detect presence/absence of target species, non-invasiveness, and high cost effectiveness rather than previous direct sampling method (Rees et al. 2014; Thomsen & Willerslev 2015 However, the presence/absence of eDNA includes many types of uncertainties due to relating to environmental factors such as temperature and advection (). For example, in aquatic habitats, it is not sure whether target species are in a location or not when eDNA of target species is detected because eDNA are transported passively. Therefore, the consideration to the influence of environmental factors on eDNA is necessary for estimation of species distribution when we use eDNA methods.

One step towards overcoming these uncertainties is a understanding of the ”ecology of eDNA”: (Barnes & Turner 2016). Previous studies

Integrated species distribution models (IDMs) are now common spatial model to predict spatial pattern of species (Issac et al. 2020). The model use the different type of data with strengths and weaknesses, such as scientific survey data which is restricted spatially and quantitatively and opportunistic citizen data which is widely collected and abundant, and combine in a single model (Isaac et al. 2020; Miller et al. 2019).

The models combine the different type of data with strengths and weaknesses in a single model (). For example, scientific survey data are high quality but less abundant due to restriction of spatially costly while opportunistic data such as citizen data are widely collected and abundant but may be low quality due to not using consistent field methods. Combining both types of data can capitalize on the strengths of each data and perform better prediction than models when we use single data (Pacifici et al. 2017; Miller et al. 2019).

Tokyo Bay is a large enclosed coastal sea in Japan. In Tokyo Bay, there are many commercially important species for fisheries that are called "Edomae" because these species have been used for Sushi since Edo Era (about 400 years ago). Catch of some Edomae have been decreased because of habitat modification due to urbanization (e.g., landfill of tidal flats and water pollution). Catch statistics (total catch in each species, efforts, and geographic location of fishing) have been collected for stock assessment since 1990 by prefectures around Tokyo Bay. The strengths of this data are the direct evidence that a focal species occupies a location of fishing and abundant because of widely collected in Tokyo Bay. On the other hand, weakness of this data is like a opportunistic data because the data is likely to be biased towards areas to high density of focal species due to commercially fishes, consequently less zero data. In addition to this catch statistics, scientific survey of eDNA has been conducted monthly since 2018 for biodiversity monitoring because biodiversity also may decreased due to human-induced environmental changes in Tokyo Bay (Hongo et al., submitted). The strengths are that the data is systematically collected by scientific survey data and includes zero data, while the weaknesses are that the data is less abundant due to spatial restriction of the survey and includes uncertainties in presence/absence as description in above.

In this paper, to predict spatial distribution of species from eDNA, we first make a model which considers uncertainties of eDNA caused by environmental factors without additional laboratory experiments and numerical hydrodynamic models, by using an integrated spatial distribution model (eDNA-IDM). We then apply the model to both eDNA data and catch statistics for four Edomae fish in Tokyo Bay, Japan. The predicted spatial distribution of four fish form our model reduced

2 Materials and Methods

2.1 A general model to estimate species distribution from eDNA

To estimate the spatial distribution from eDNA considering with spatial biases due to degradation from environmental factors (e.g., temperature and advection),

$$\text{logit}(p_1(s_i)) = \alpha_1 + \beta(s_i) + \theta(s_i) + u_1(s_i)$$

$$\text{logit}(p_2(s_i)) = \alpha_2 + \sum_k f_k(x_k(s_i)) + w\theta(s_i) + u_2(s_i)$$

2.2 An application to a eDNA and catch data in Tokyo Bay

2.2.1 eDNA data

2.2.1.1 Field survey

Field surveys were conducted by prefectural experimental station in Chiba, following the consistent sampling design at 14 sites in Tokyo Bay from April to December in 2018 (Fig. 1). In each sites, seawater and environmental data were simultaneously collected. For eDNA analysis, two litter of bottom seawater was collected using a Niskin water sampler, and then it was separated for two 1L samples for replicate. Each samples filtered glass fiber membrane GF/F (0.7 μm pore size; Cytiva, Sheffield, UK) onboard and then the filters were frozen on a block of dry ice. These frozen filters were stored at $-30°$ in the laboratory until eDNA extraction. To lower the levels of cross-contamination, equipments for eDNA sampling were changed new one or washed in each sites. During sampling the bottom seawater, seawater temperature, salinity, pH, and dissolved oxygen (DO) at the same depth of seawater sampling for eDNA were measured by CTD (メーカー).

2.2.1.1 Laboratory experiments

In laboratory, eDNA extraction, eDNA amplification, and eDNA sequence were conducted. Total eDNA was extracted from the frozen filters using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following Yamamoto et al. 2019. Mitochondorial 12S rRNA gene was amplified using MiFish universal primers referring to Miya et al. 2015 with slight modification. The details was shown in Hongo et al. (受理されてないようだったら書くしかない). eDNA sequence were ....

2.2.2 Catch statistics

A part of catch statistics of small-scale bottom trawl fisheries recorded by several representative boats of Chiba Prefecture were provided by Chiba Prefecture. This data included date, geographic location, efforts (number of tows), gear, and catch weight (kg) in each fish. Almost of all gear was beam trawl although dredge net also used. The species which also detected by eDNA was *Conger myriaster* (マアナゴ), *Kareius bicoloratus* (イシガレイ), *Lateolabrax japonicus* (スズキ), and *Konosirus punctatus* (コノシロ). Thus, we estimated the spatial distribution of these four species using the eDNA-IDM. マコガレイ,

カマス類，クロダイ，イシモチ類も解析できる？？

2.2.3 Estimation of spatial distribution

To estimate the spatial distribution of four focal species from eDNA and catch data by considering uncertainties caused by environmental factors, we fitted the model (equation 1) to the presence/absence data of eDNA and of catch data collected in Tokyo Bay as follows:

equation examples

$$\begin{aligned} \text{logit } p_i &= \beta_p(t_i) + \omega_p(s_i) + \varepsilon_p(s_i, t_i) + \eta_p(v_i) + \lambda_p Q(i) \\ \log d_i &= \beta_d(t_i) + \omega_d(s_i) + \varepsilon_d(s_i, t_i) + \eta_d(v_i) + \lambda_d Q(i) \end{aligned} \tag{1}$$

where $\beta(t_i)$ is the intercept for year t, and $\omega(s_i)$ and $\varepsilon(s_i, t_i)$ are the spatial and spatio–temporal random effects for year t and location s, respectively. $\eta(v_i)$ is the overdispersion random effect of factor v_i, which is the interaction of year and month. λ is the effect of the chatchability covariate $Q(i)$:

$$Q(i) = \log(d_{chub}(s_i) + 0.1).$$

That is, this term considers the effect of species misidentification between chub mackerel and spotted mackerel; as mentioned earlier, we suspected overestimation of egg density of spotted mackerel because the difference in egg diameter has become ambiguous according to increase in egg density of chub mackerel and the distributions of egg diameters between species have overlapped (Yukami et al., 2019). The constant 0.1 was added because $\log 0$ (i.e., no chub mackerel eggs) is undefined, and the same result was obtained when using 1 in place of 0.1.

Acknowledgments

This research was financially supported by Grant-in-Aid for Fisheries Agency of Japan.

Authorship

YK conceived of the research idea. YH, YU, HM, MI, KA, and AK conducted field sampling. YH performed the laboratory experiments. YK, HO, and SN designed statistical analyses. YK wrote programs and performed the analyses. YK wrote the manuscript with input from all co-authors' comments.

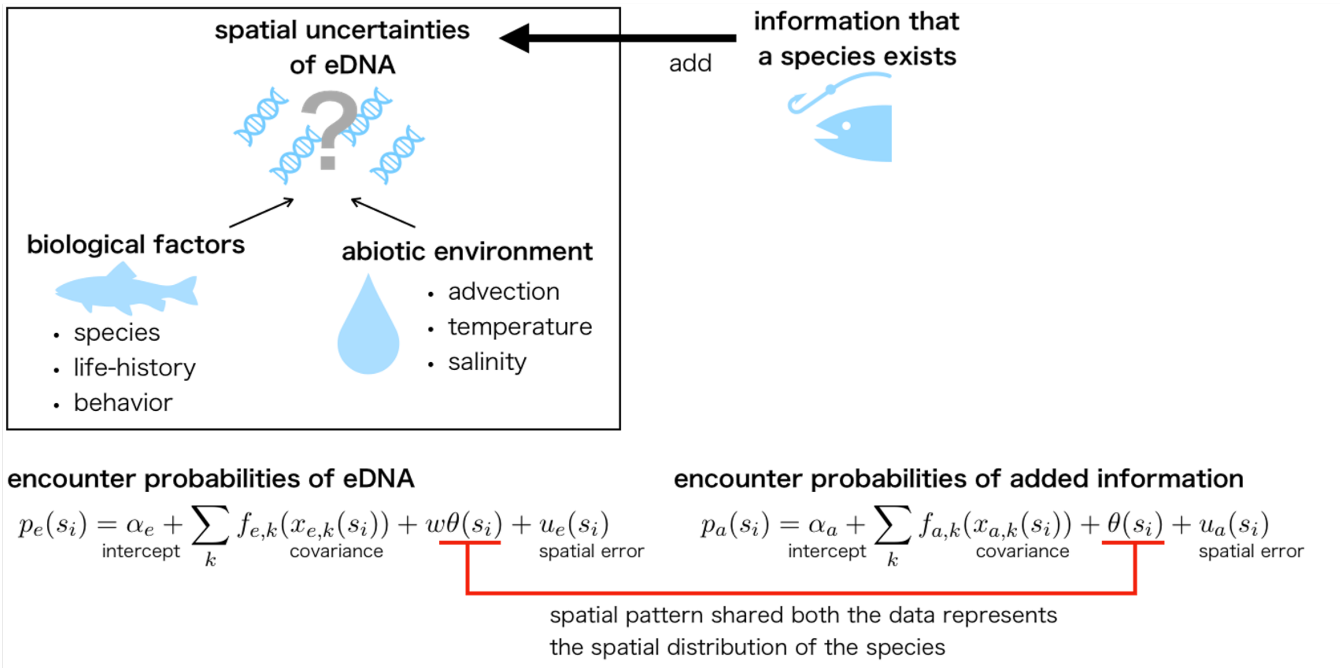

Fig. 1: Conceptual diagram of this study.