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 invesive 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-invesiveness, 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 uncertinties 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 uncertinties is a understanding of the ”ecology of eDNA”: (Barnes & Turner 2016). Previous studies

Integrated species distribution models 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 useful species for fisheries that are colled "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, Japan. 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 (Hongo et al., submitted). The strengths are that the data is systematically collected because of scientific survey data and includes zero data, while the weaknesses are that the data is less abundant and includes uncertinties 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 uncertinties of eDNA caused by environmental factors without additional laboratory experiments and numerical hydrodynamic models, by using an integrated spatial

distribution model. 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 Data sets

2.1.1 Survey and data

The egg density data with $30'$ latitude $\times$ $30'$ longitude horizontal square resolution in the areas from 122°E to 150°E and 24°N to 43°N was used. The egg density data set was derived from monthly egg surveys off the Pacific coast of Japan from January to June, 2005–2019 (Takasuka et al., 2008a, 2019). The aim of the surveys was to monitor the egg abundance of major small pelagic fish species, including chub mackerel and spotted mackerel, so that the spatial area and survey month of the data largely covered the major spawning grounds and spawning season. While some sampling locations were fixed, others varied for various reasons (e.g., environmental conditions). Accordingly, the survey design changed slightly each year (Kanamori et al., 2019). Although the sampling efforts were approximately consistent year-round, the efforts tended to be more intensive during early spring; effort was highest in February and decreased gradually thereafter (Takasuka et al., 2008b).

The egg surveys were conducted by 18 prefectural experimental stations or fisheries research institutes and two national research institutes of the Japan Fisheries Research and

Education Agency, following the consistent sampling designs, as a part of the stock assessment project. In the surveys, plankton nets were towed vertically from a depth of 150 m to the surface (if the depth was ¡150 m, nets were lowered to just above the bottom). This range of depths covers the vertical distributions of eggs of small pelagic fish. During the period from 2005 to 2019, the surveys used a plankton net with a mouth ring diameter of 0.45 m and a mesh size of 0.335 (partially 0.330 mm in 2015) (Takasuka et al., 2017). The samples were fixed with 5% formalin immediately after collection. In the laboratory, the samples were identified and sorted into eggs and larvae of different small pelagic species, based on the morphological characteristics (e.g., egg shape and size, number of oil globules, segmented yolk, perivitelline space ranging, yolk diameter, oil globule diameter). For the mackerel eggs, the egg diameters were measured to the nearest 0.025 mm by a micrometer for a maximum number of 100 individuals per sample (station or tow). Eggs with diameters $>$1.1 mm were identified as spotted mackerel, whereas those with diameters leq1.0 mm were identified as chub mackerel, according to Nishida et al. (2001). For any sample of $>$100 individuals, the proportion of the two species among 100 randomly selected individuals was assumed to be the same for the whole sample. Additionally, the number of eggs per unit area in the water column (number m^{-2}) for each sampling tow was calculated by flow-meter revolutions, flow-meter revolutions per meter tow in the calibration, wire length (m), opening mouth area of the net (m^{-2}), and wire angle. Then, the arithmetic average of the number of eggs was obtained with 30′ latitude $\times$ 30′ longitude horizontal square resolution. The mean proportion of the total number of eggs of spotted mackerel against the total number of eggs of *Scomber* was less than 20 % from 2005 to 2019. Therefore, the effect of the misidentification error that we considered was from chub

mackerel on spotted mackerel (i.e., we assumed that the effect of the misidentification error from spotted mackerel on chub mackerel was small.) More detailed descriptions of the surveys and data set are provided in previous studies of the reproductive biology of small pelagic fish species (e.g., Takasuka et al. 2008a,b, 2017, 2019).

2.2 Data analyses

2.2.1 Indices of egg density

In this study, we used the three indices of egg density of spotted mackerel; nominal, chub–, and chub+. The nominal index was the arithmetic mean of egg density for each year. The chub– index was the estimated egg density by considering sampling effects (i.e., spatio–temporal changes in survey design). The chub+ index was the estimated egg density by considering sampling effects and the effect of egg density of chub mackerel on the catchability of egg density of spotted mackerel. The process for estimating chub– and the chub+ is described in the following section.

2.2.2 Estimation of the indices of egg density

To estimate the chub– and the chub+ indices of egg density by considering sampling effects (i.e., spatio–temporal changes in survey design) as well as the effect of the egg density of chub mackerel on the catchability of egg density of spotted mackerel, we used the multivariate vector autoregressive spatio-temporal (VAST) model (Thorson and Barnett, 2017), which accounts for spatio-temporal changes in survey design, survey effort, and observation rates and can accurately estimate relative local densities at high resolution by standardizing sampling designs (Thorson and Barnett, 2017; Thorson, 2019). The model

includes two potential components because it is designed to support delta-models: (i) the encounter probability p_i for each sample i and (ii) the expected egg density d_i for each sample i when spawning occurs (i.e., egg density is not zero). The encounter probability p_i and the expected egg density d_i are, respectively, approximated using a logit-linked linear predictor and a log-linked linear predictor as follows (Thorson and Barnett, 2017):

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

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.