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 major goal in ecology. Field surveys using environmental DNA (eDNA) are widely used for detecting invasive or rare species and hotspot of biodiversity because the surveys of eDNA are easy to detect occurrence of target species, non-invasiveness, and high cost effectiveness rather than previous direct sampling method (Rees et al. 2014; Thomsen & Willerslev 2015) However, uncertainty of the occurrence of eDNA is a issue in eDNA studies because various factors, such as biological and environmental features, intricately affect the shedding (i.e., production of eDNA from organisms) and degradation (i.e., decay of eDNA from a system) of eDNA (Fig. 1). Therefore, it is necessary to consider the spatial uncertainty included in eDNA to infer the spatial distribution of species.

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. There are many commercially important species for fisheries that are called ”Edomae” in Tokyo Bay because these species have been used for Sushi since Edo Era (about 400 years ago). Catch weight 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 weight 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 data and includes zero data, while the weaknesses are that the data is less abundant due to spatial restriction of the survey and includes spatial uncertainty in the occurrence of eDNA as description in above.

In this paper, to infer spatial distribution of species from eDNA, we first make a spatial distribution model wichi considers spatial uncertainty of eDNA 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.

2 Materials and Methods

2.1 A general model to estimate species distribution from eDNA

Integrated spatial distribution model that account for explicitly spatial autocorrelation in occurrence were built by Pacifici et al. (2017), which shows three approaches to predict the spatial distribution of species: the joint likelihood (shared), correlation, and covariate methods. The joint likelihood method uses multiple data types to simultaneously estimate a shared set of parameters with constraining that the likelihoods of shared set of parameters to be equal across. The correlation method connects multiple data types indirectly through a shared covariance matrix that captures similar patterns present in each data sources. The covariate method incorporates information from a added dataset via a fixed effect.

Although each methods estimate the spatial distribution of species using multiple data sets, we need to select method depending on the data features for analysis because there are strengths and weaknesses (Pacifici et al. 2017; Miller et al. 2018). The joint likelihood method may be problematic when the second data is of poorly quality compared to correlation and covariate methods because each data can directly inform the latent occurrence state (probabilities?) and the weight given to estimate the parameters is naturally determined by their relative size and quality. Thus, it is not the best method when our second data is low quality while it is the best method when our second data is high quality (vise versa). The correlation method is added robustness to the joint likelihood because the second data indirectly inform the occurrence state. Thus, it is the best method when our second data is low quality while it is inferior to the joint likelihood method when both data are deemed reliable. The covariate method does not make full use of the information in the second data because the second data as a constructed covariate in the mean occurrence state. In addition, this method can reduce the computational cost because there are fewer parameters to estimate and the number of data locations can be reduced. Thus, it is the best method when the second data is low quality and/or there is computational limitation while it may not the best method when the information of the second data is needed.

When predicting the spatial distribution of species from eDNA using integrated species distribution model, the information that a species exists is needed as second data to consider spatial uncertainties of eDNA due to complex factors (Fig. 1). Hence, the second data is preferred to high quality as possible.

しかし，eDNAは直接的なモニタリングに比べて簡易的であるためより広い範囲で取得されている可能性が高く，eDNAのデータと同様の空間範囲で調査データのように質の高いデータを取得することは難しいかもしれない．その一方で，eDNAの空間的な不確実性を考慮するためには，種がいた証拠である2番目のデータの情報をeDNAのデータにしっかりと伝える必要がある．これらを考えると，integrated spetial distribution modelを用いたeDNAからの空間分布の推定には，以下のようなcorrelation methodが適切である:

$$\begin{aligned} p_e(s_i) &= \alpha_e + \textstyle\sum_k f_{e,k}(x_{e,k}(s_i)) + w\theta(s_i) + u_e(s_i) \\ p_a(s_i) &= \alpha_a + \textstyle\sum_k f_{a,k}(x_{a,k}(s_i)) + \theta(s_i) + u_a(s_i) \end{aligned} \tag{1}$$

where α and $x_k(s_i)$ are the intercept and the covariates at sites i for occurrence probabilities at sites i of the added data (p_a) and eDNA data (p_e), respectively. $u(s_i)$ is spatial error that is specific for each data following multivariate normal distributions $\mathrm{MVN}(0, \mathbf{R})$, where the variance–covariance matrix $\mathbf{R}$ is a Matérn correlation function. θ which is shared between two equations is the common spatial pattern between the two data, which cannot explain by each terms of the equations. That is, θ can be interpreted as ”true” spatial distribution of species.

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

2.2.1 eDNA data

Field surveys

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).

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 using eDNA and catch data by considering with spatial uncertainties of eDNA, we fitted the model (equation 1) to the presence/absence of eDNA and of catch collected in Tokyo Bay as follows:

$$\text{logit }\ p_e(s_i) = \alpha_e + \sum_k f_k(x_k(s_i)) + w\theta(s_i) + u_e(s_i)$$

$$\text{logit }\ p_c(s_i) = \alpha_c + \beta_i + \theta(s_i) + u_c(s_i)$$

where α is the intercept, and $x_k(s_i)$ is the covariates at sites i for occurrence probabilities of eDNA at sites i. In the study, seawater temperature, salinity, pH, and DO were used as covariates which effect on the occurrence of eDNA (i.e., $k = 4$). $u(s_i)$ is spatial error that is specific for each data following multivariate normal distributions $\mathrm{MVN}(0, \mathbf{R})$, where the variance–covariance matrix $\mathbf{R}$ is a Matérn correlation function. θ which is shared between eDNA and catch is the common spatial pattern between the two data, which cannot explain by each terms of the equations. That is, θ can be interpreted as the spatial distribution of species we want to know. 共変量の非線形性について書く Parameters in this model was estimated by Integrated Nested Laplace Approximation using using the R-INLA package (Lindgren, 2012) in R 3.6.1 (R Development Core Team, 2019).

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.

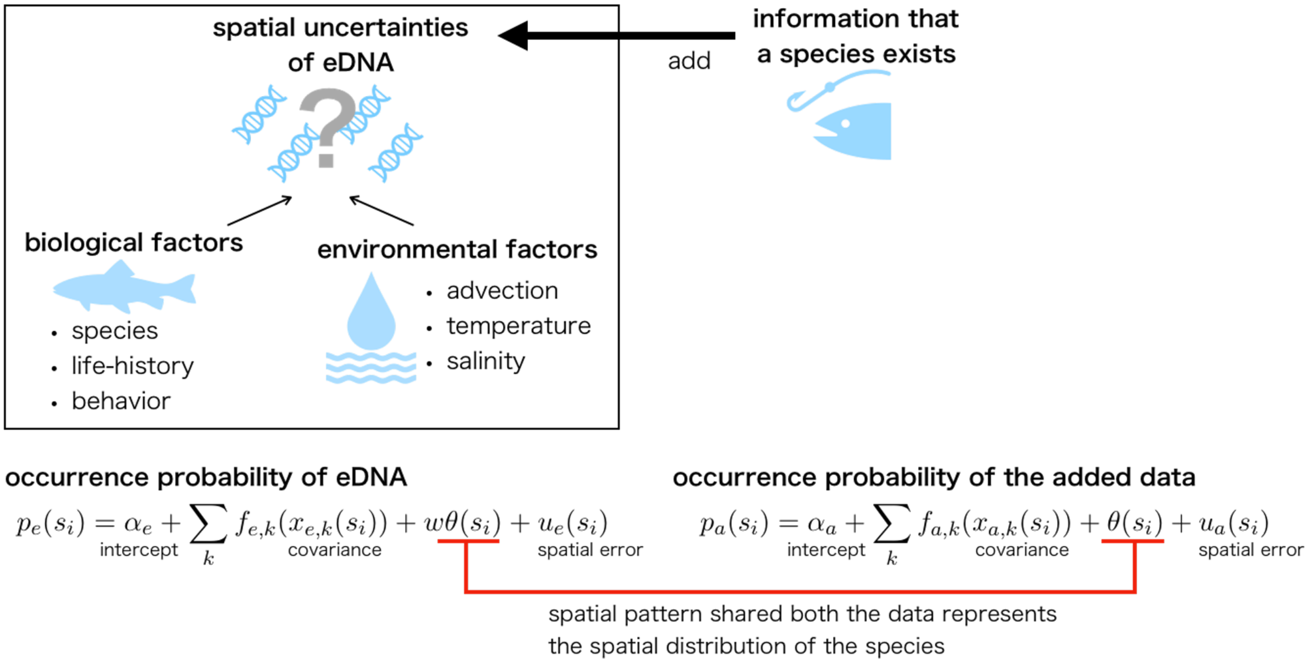
spatial uncertainties of eDNA
add
information that a species exists
biological factors
species
life-history
behavior
environmental factors
advection
temperature
salinity
occurrence probability of eDNA
$p_e(s_i) = \alpha_e + \sum_k f_{e,k}(x_{e,k}(s_i)) + w\theta(s_i) + u_e(s_i)$
intercept
covariance
spatial error
occurrence probability of the added data
$p_a(s_i) = \alpha_a + \sum_k f_{a,k}(x_{a,k}(s_i)) + \theta(s_i) + u_a(s_i)$
intercept
covariance
spatial error
spatial pattern shared both the data represents the spatial distribution of the species

Fig. 1: Conceptual diagram of this study.