


RESEARCH ARTICLE

Open Access



Oceanographic indicators for the occurrence of anchovy eggs inferred from generalized additive models

Jin Yeong Kim^{1*} , Jae Bong Lee² and Young-Sang Suh²

Abstract

Three generalized additive models were applied to the distribution of anchovy eggs and oceanographic factors to determine the occurrence of anchovy spawning grounds in Korean waters and to identify the indicators of their occurrence using survey data from the spring and summer of 1985, 1995, and 2002. Binomial and Gaussian types of generalized additive models (GAM) and quantile generalized additive models (QGAM) revealed that egg density was influenced mostly by ocean temperature and salinity in spring, and the vertical structure of temperature, salinity, dissolved oxygen, and zooplankton biomass during summer in the upper quantiles of egg density. The GAM and QGAM model deviance explained 18.5–63.2% of the egg distribution in summer in the East and West Sea. For the principle component analysis-based GAMs, the variance explained by the final regression model was 27.3–67.0%, higher than the regular models and QGAMs for egg density in the East and West Sea. By analyzing the distribution of anchovy eggs off the Korean coast, our results revealed the optimal temperature and salinity conditions, in addition to high production and high vertical mixing, as the key indicators of the major spawning grounds of anchovies.

Keywords: Anchovy eggs, Indicators, GAMs, QGAMs, PCA-based GAMs

Background

Generalized additive models (GAMs) are non-parametric regression techniques that are not restricted by linear relationships, thus providing a flexible method for analysis when the relationship between variables is complex. Generalized additive models are powerful tools for analysis and are useful for investigating the effects of environmental factors on the distribution of organisms (Denis et al. 2002; Murase et al. 2009; Hoeffle et al. 2014). Based on theoretical considerations, the species response to an environmental factor is expected to be bell-shaped, i.e., “optima” in terms of environments, and recent studies have applied non-linear quantile regression models to consider the impact of environmental factors (Husson et al. 2020;

Waldock et al. 2019). Recently, principal component analysis-based (PCA-based) GAMs have been used in ecological research (Li et al. 2017; Liu et al. 2019; Alvarez-Fernandez et al. 2012). These provide a tool for reducing the number of variables by grouping variables that influence mutual factors, depending on the relevance of the variables.

Anchovy (*Engraulis japonicus*), a short-lived pelagic species, is one of the most important fisheries in Korean waters. Anchovies spawn from early spring to autumn, and their highest densities are found in the waters of Korea between April and August (Lim and Ok 1977; Kim and Lo 2001). Hydroacoustic and trawl surveys conducted during early spring have shown that anchovy wintering grounds are located in the coastal areas of the South Sea of Korea (Choi et al. 2001). Anchovy eggs and larvae are widely distributed in Korean waters but are concentrated along the frontal areas between the coastal

* Correspondence: jinykim3@gmail.com

¹Institute for Mathematical Convergence, Gyeongpook National University, #80 Daehakro, Bukgu, Daegu 41566, Republic of Korea
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

and offshore waters (Lim and Ok 1977; Kim 1983). To date, most ichthyoplankton research has focused on the South Sea of Korea, which constitutes the most important anchovy fishing ground and produces particularly high yields of juveniles and adults (Kim 1983; Kim and Lo 2001; Kim et al. 2005). However, little is known regarding anchovy spawning grounds and the oceanographic factors that influence their occurrence in the East and West Seas. Furthermore, the impacts of global warming have driven the expansion of fishery grounds offshore, which has caused an increase in the commercial catch during the winter season (Park et al. 2004; Kim et al. 2007). Therefore, from ecological and economical point of view, it is necessary to study all possible spawning grounds in Korean waters.

Short-lived pelagic fish vary widely in their patterns of distribution and migration with changes in oceanographic conditions that can drive quasi-decadal variations in population sizes (Kim and Lo 2001; Kim et al. 2007; Alheit and Bakun 2010). In Korean waters, spatial variability and long-term changes in the climate and oceanographic conditions, combined with the effects of anthropogenic activities, have resulted in the warming of regional seawater. This, in turn, has altered the marine ecosystem and influenced the dominant fish species in commercial fishery catches (Kim et al. 2007).

In this study, we quantified the occurrence of anchovy eggs using the egg density data collected during the spawning seasons and identified the oceanographic indicators that influence anchovy spawning grounds. We used regular GAMs, quantile generalized additive models (QGAMs), and PCA-based GAMs to compare the multivariate and quantile non-linear regressions of the absence-presence or the density of eggs depending on the oceanographic factors.

Materials and methods

Sampling and observation procedures

To determine the seasonal factors influencing anchovy spawning grounds, we used ichthyoplankton data collected from a zooplankton net (45 cm mouth diameter, 0.333 mm mesh size) during the oceanic survey regularly conducted by National Institute of Fisheries Science (NIFS). A total of 492 samples were collected along 19 lines in the East, West, and South Seas during April, June, and August of 1985 and 2002, and in June and August of 1995 (Fig. 1). Oceanographic data (seawater temperature, salinity, and dissolved oxygen (DO); at depths of 10 m and 50 m), coinciding with the same time period as the NIFS surveys, were retrieved from the Korea Oceanographic Data Center (KODC). As anchovy eggs are distributed mostly at the 10–50 m water layer (Kim and Choi 1988; Boyra et al. 2003), we used the

differences in temperature, salinity, and DO at 10 m and 50 m water layers as the levels of indication of vertical mixing in the ocean.

Egg density was calculated as the number of eggs per m^2 of sea surface area. Zooplankton biomass (mg/m^3) was estimated from the wet weight of each sample after removing the eggs. We used the egg distribution pattern to indicate the extent of the anchovy spawning ground, which was explained using a confocal ellipse of the distribution area and the density (D) at each station, and was calculated as follows:

$$(\bar{X}, \bar{Y}) = \left(\frac{\sum_i D_i X_i}{\sum_i D_i}, \frac{\sum_i D_i Y_i}{\sum_i D_i} \right)$$

where X and Y are the longitude and latitude of the station (i), and the \bar{X} and \bar{Y} bars are the confocal of the ellipses. The ellipses can be described by two principle axes of the fishing grounds: the major axis (b_1) and the minor axis (b_2), which are at right angles to each other. The slopes of b_1 and b_2 were calculated from the covariance of the longitude (X) and latitude (Y) of the station, weighted by the density (D), as estimated from the survey (Sokal and Rohlf 1995; Kim et al. 2005).

Statistical methods

Based on the oceanographic data system by the KODC, the total survey area was divided into three distinct areas: the East Sea (lines 102–106, 208–209), the South Sea (lines 203–207), and the West Sea (lines 307–312) of Korea. Box plots and descriptive statistics (mean and variance) were used to summarize the distribution of egg density and the oceanographic factors for the total monitoring period (total egg density) by year, month, and area. We applied statistical methods with the ordinated seven variables: temperature, salinity, and DO at the 10 m depth layer, the differences between the 10 m and 50 m depth layers for temperature, salinity, and DO and zooplankton abundance. To determine differences in egg density, we analyzed egg density over the years and survey lines during April, and among years, months, and regions during June and August using a multi-factor analysis of variance (MANOVA). Egg density data collected during June and August were combined for subsequent analysis as the MANOVA showed no significant differences between the densities in these 2 months.

We applied regular GAMs for the total and spring seasons (April) and for the data in each area during the summer season (June and August) to explain the occurrence of anchovy spawning grounds based on oceanographic factors and compared the results of the models. Since egg density is characterized by high variability due to drift and high embryonic mortality, we used two spatial strata and quantile GAMs to determine the

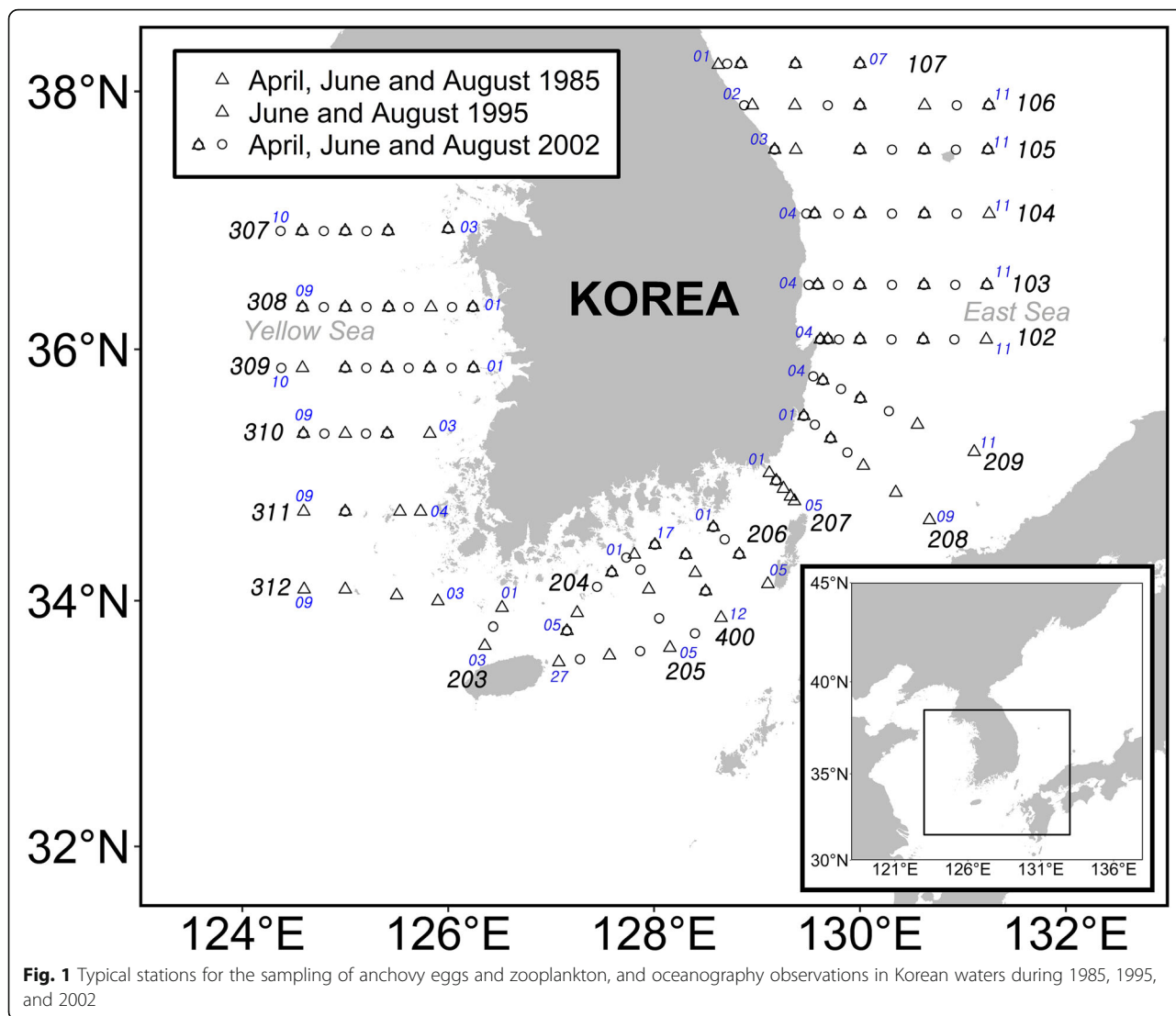


Fig. 1 Typical stations for the sampling of anchovy eggs and zooplankton, and oceanography observations in Korean waters during 1985, 1995, and 2002

indicators of the spawning grounds based on the absence/presence of eggs, the log-transformed egg density, and high density (above 70–90% quantiles) for the total data. For the first stratum (with egg presence as a response variable), GAMs were constructed using a binomial error structure and a logit-link function with an unbiased risk estimator (UBRE). For the second stratum, GAMs with log-transformed egg density as a response variable and oceanic factors as explanatory variables were built assuming a Gaussian error structure and an identity-link function with the generalized cross validation (GCV). Quantile generalized additive models were used to estimate either the conditional median or other quantiles of the response variables based on Elf with restricted maximum likelihood (REML) criterion. The upper boundary describes how abundance is limited by this factor, while the variation below the upper boundary reflects the limiting effect on the abundance of

environmental attributes other than the factor of interest (Vaz et al. 2008). We assumed that a high density of eggs indicates the center of spawning and the eggs spread out gradually as they developed. To confirm the effect of the oceanic factors on the egg density, confidence intervals around the predicted 0.10th and 0.90th quantiles in the quantile GAM were used to model the τ th quantile (where $\tau \in (0,1)$) of the response (y), and conditional on a p -dimensional vector of covariates (x). We fitted variables as a smooth term at the 0.10th, 0.50th, and 0.90th quantile of residual-abundance for total data and the 0.50th, 0.70th, 0.80th, and 0.90th quantile of residual-abundance for spring because of the seasonal changes in the ratio of absence stations: absence was far more frequent in spring and could overwhelm the shape of abundance distributions.

We applied the PCA and GAMs with the seven ordinated variables: temperature, salinity, and DO at the 10

m depth layer, and the difference between the 10 m and 50 m depth layers for temperature, salinity, and DO in addition to the zooplankton abundances based on the correlations. The PCA can be used to eliminate redundant information by removing correlated features in the predictor variables, and then using that to create new independent variables (Li et al. 2017; Liu et al. 2019). In our study, we used the PCA to simplify our models and eliminate multi-collinearity among oceanographic variables. The PCA-based GAMs were used to reduce variables, and smoothing functions were used on the predicted principal components (PCs). Independent variables were selected based on the environmental conditions and PCs were selected from the results of the PCA.

In this study, the general form of the GAM is explained by the following equation:

$$E[y] = a + s(V_i) + \dots + s(V_n) + \varepsilon$$

where $E[y]$ is the expected value of the response (dependent) variable y , a is the intercept of the parametric term which represents the mean of the response variable, s is the smoothing function based on the thin plate regression spline, V_i represents the independent variables, and ε is the error term. We restricted the number of degrees of freedom ($df = 4$, $k = 3$) to minimize model over-fitting.

The initial models of the GAM contained all covariates and were optimized with backward selection based on a reduction in Akaike's information criterion (AIC). We selected the best models based on the gain in deviance explained (% Dev) relative to the basic model, whilst minimizing the AIC, GCV, UBRE, and REML criteria. To avoid overfitting and to reduce collinearity, all covariates were examined using Pearson's rank correlations to identify variables with low correlation values (Pearson correlation $r < 0.6$) in the modeling process. Multi-

collinearity among the predictor variables was assessed by calculating the variance inflation factor (VIF) using a cutoff value of 5 with the function (Liu et al. 2019; Hoefle et al. 2014). The GAMs and QGAMs were fitted using the "mgcv" (Wood 2019) and "qgam" (Fasiolo et al. 2020) in R-3.4.2.

Results

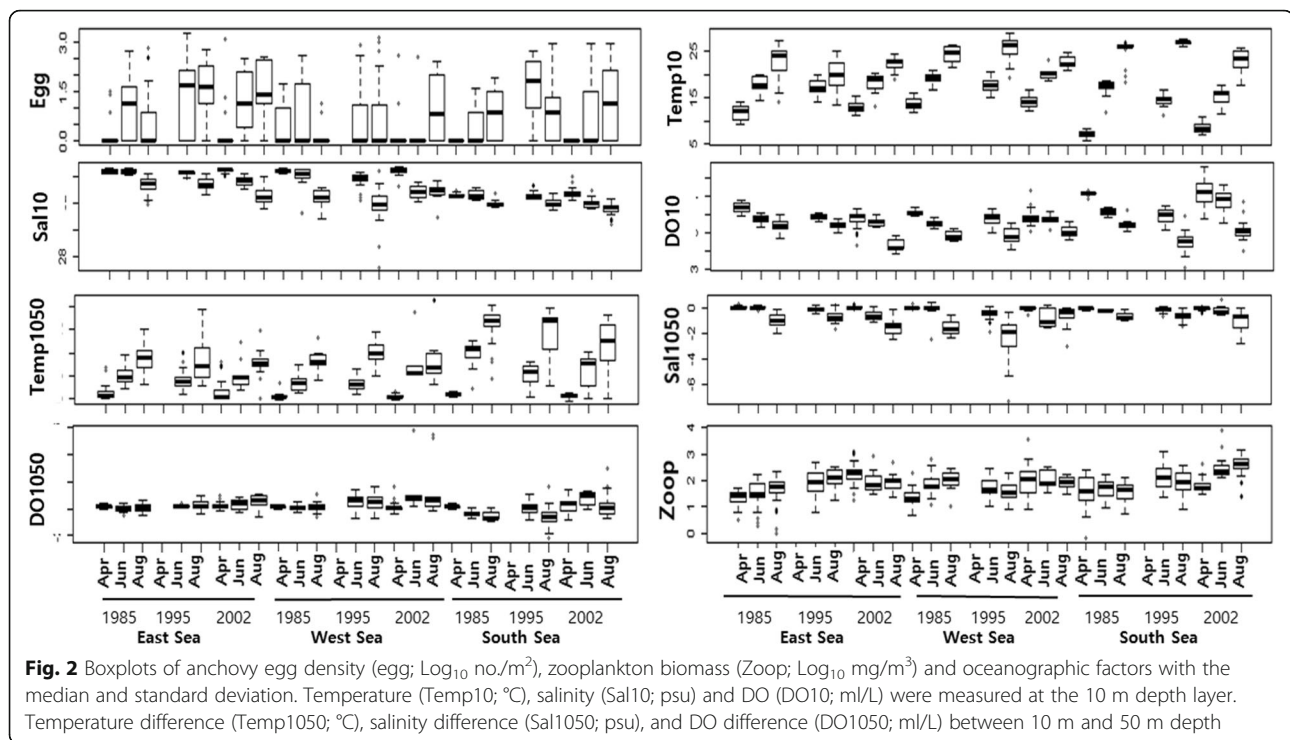
Distribution of egg density and oceanographic factors

Descriptive statistics were used to summarize the distribution of egg density and oceanographic factors (Table 1 and Fig. 2). Egg density and oceanographic factors varied widely and were highly skewed with a high level of kurtosis. Egg density was lowest in April and highest in June and August, with particularly high densities observed in the East and South Seas during 1995, and in the West Sea during August 2002. In all areas, the oceanographic parameters (measured at depths of 10 m) varied seasonally, with trends of increasing temperature and decreasing salinity and DO from April to August. Concurrently, the temperature difference between the 10 m and 50 m depth layers decreased, while the salinity difference increased from April to August. The DO difference between 10 m and 50 m varied widely during June and August. Zooplankton density increased from April to June and was particularly high during 1995 and 2002 in the East Sea and during 2002 in the South Sea (Fig. 2).

The results from the PCA revealed an optimal subset of explanatory variables, which represented a cumulative 77–88% of the total variance across three components (Fig. 3a). Temperature, salinity and DO at 10 m depth and the value of the difference between 10 and 50 m were mostly related to the first component (PC1) and represented 37–53% of the variance; DO and the DO difference between 10 m and 50 m were mostly related to the second component (PC2) and represented 16–31% of the variance; and zooplankton density was related

Table 1 Descriptive statistics summarizing the total anchovy egg density and oceanographic variables in Korean waters during April, June, and August of 1985 and 2002; and June and August of 1995

	No.	Mean	SD	Median	Min	Max	Skewness	Kurtosis	SE
Egg presence ratio	492	0.42	0.49	0.00	0.00	1.00	0.34	- 1.89	0.02
Egg density (Ind./m ²)	492	69.33	202.47	0.00	0.00	1893	4.95	30.75	9.13
10 m depth									
Temperature (°C)	492	18.10	5.60	17.62	5.70	28.88	- 0.03	- 0.84	0.25
Salinity (psu)	492	33.26	1.13	33.38	27.23	34.76	- 0.77	1.08	0.05
DO(ml/L)	492	5.67	0.78	5.67	3.12	8.62	0.29	0.88	0.03
50 m depth									
Temperature (°C)	492	12.24	3.97	12.81	2.45	24.16	- 0.06	- 0.55	0.18
Salinity (psu)	492	33.78	0.81	34.20	31.46	34.80	- 0.80	- 0.84	0.04
DO (ml/L)	492	5.51	0.83	5.58	3.08	8.20	- 0.27	0.78	0.04
Zooplankton biomass (mg/m ³)	492	162.62	443.53	71.50	0.66	8238.40	13.57	229.07	20.00



to the third component (PC3) and represented 14–20% of the variance, with correlation values varying among the area and the year (Fig. 3a, b–d).

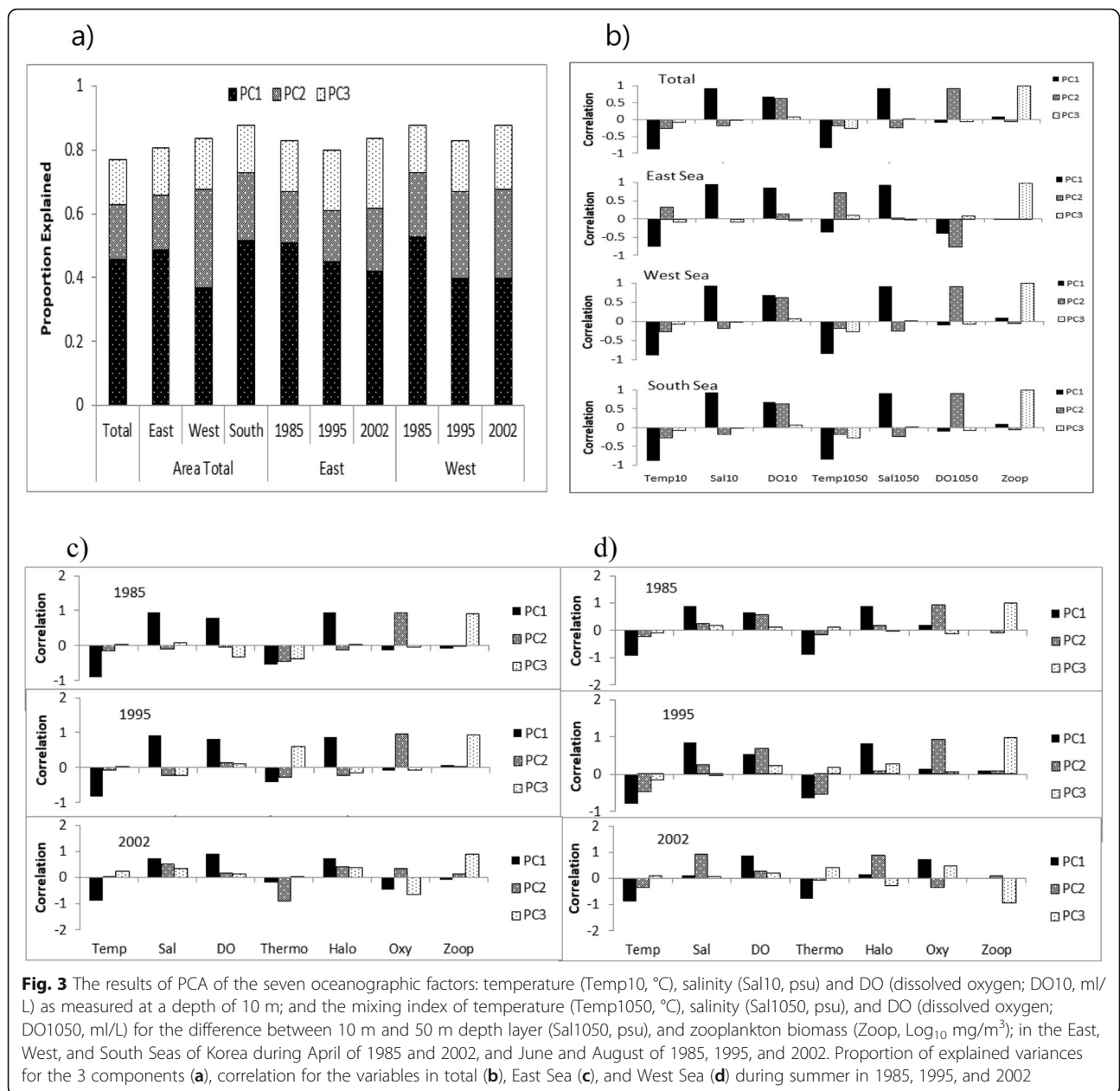
Based on measurements from 1985, 1995, and 2002, the monthly changes in the anchovy spawning grounds were characterized by seasonal variations in the density of anchovy eggs (Fig. 4). In April, the egg distribution was limited to the South Sea and the southern East Sea, whereas by June and August, eggs were widely distributed in the mid-East and West Seas. In April 1985, anchovy eggs were present at low densities (< 50 ind./m²) in the South Sea and in the southern East Sea. However, by April 2002, the egg distribution changed from the southern East Sea and southern coast of the mid-East Sea ($P = 0.041$, with a significant difference in densities at the survey line between April 1985 and 2002). In June 1985 and 1995, the eggs were distributed over wide areas of Korean waters. In August, the eggs were concentrated in the mid-East Sea and occurred along coastal areas of the mid and southern West Sea and the East Sea (> 500 ind./m²). In June 2002, the eggs were found at high densities in the waters of the mid-West and southern East Seas. In August 2002, the distribution of eggs was concentrated offshore of the West Sea (> 1000 ind./m²) and in coastal areas of the southern East Sea. Results of the MANOVA analysis of egg density during summer (June and August) showed significant differences by year ($P = 0.002$) and area ($P = 0.017$).

The distribution of anchovy eggs explained using confocal ellipses with axes and slopes are shown in Fig. 5.

The joint confidence regions were compared with the egg densities in each month between 1985, 1995, and 2002. Based on Fig. 5, it is apparent that their spawning ground was distributed in the East Sea in 2002 compared to 1985 and 1995, while their spawning ground changed northward in the West Sea in 1995 in comparison to 1985 and 2002.

Comparison of the results by GAMs, QGAMs, and PCA-based GAMs

The results of the regular (binomial and Gaussian) GAMs and QGAMs (Elf) revealed that various environmental factors at the 10 m depth layer and the values of difference between the 10 and 50 m depth layer had significant effects on egg density (Tables 2, 3, and 4). The shapes of the functional forms for the covariates are illustrated in Figs. 6, 7, and 8, where the solid lines represent the main effect of the independent variables and the dotted lines represent 95 % confidence intervals. The observed waves indicated that egg densities had non-linear responses to the covariates. Total egg density was non-linearly related to temperature, salinity, and DO at a depth of 10 m, the salinity difference between 10 m and 50 m, and log-transformed zooplankton density in the Gaussian error structure (Table 2, Fig. 6a). There was no significant effect of zooplankton density on egg presence based on results from the GAM with a binomial distribution. In addition, egg density was not significantly affected by DO (Table 2). Although the probability of egg presence was nearly maintained at a temperature of 10



m (Fig. 6b, 0.10th, and 0.50th quantiles), egg density probability peaked at 18–20 °C (Fig. 6a, b, 0.90th quantile) and especially increased in the presence of high zooplankton density in the upper 0.90th quantiles of egg density (Fig. 6b). Environmental variables explained 12.4% (binomial), 13.0% (Gaussian), and 17.7% (Elf) of the density of eggs, respectively (Table 2).

The shape of the functional forms during spring (April) indicated that egg density had a significant positive relationship with temperature (Fig. 6c), and a dome-type relationship with salinity in the upper 0.80th and 0.90th quantiles of egg density, which peaked with

salinity over 33.5 psu (Fig. 6d). This explained 16.8–48.0% of the deviance in 1985 and 2002 (Table 2).

Egg density during the summer (June and August) was mostly found to be non-linearly related to parameters in the Gaussian error structure, with steep increasing or decreasing trends in the upper 0.50th–0.90th quantiles (Fig. 7). Egg density in the East Sea was significantly related to the halocline index in conversely dome-type, and with zooplankton density in a positive linear shape during 1985 (Fig. 7a); with DO in a dome-type and oxycline index in an inversely linear shape in 1995 (Fig. 7b); with the DO difference between the 10 and 50 m layer linearly during 2002 (Fig. 7c), with 29.9–41.1% and

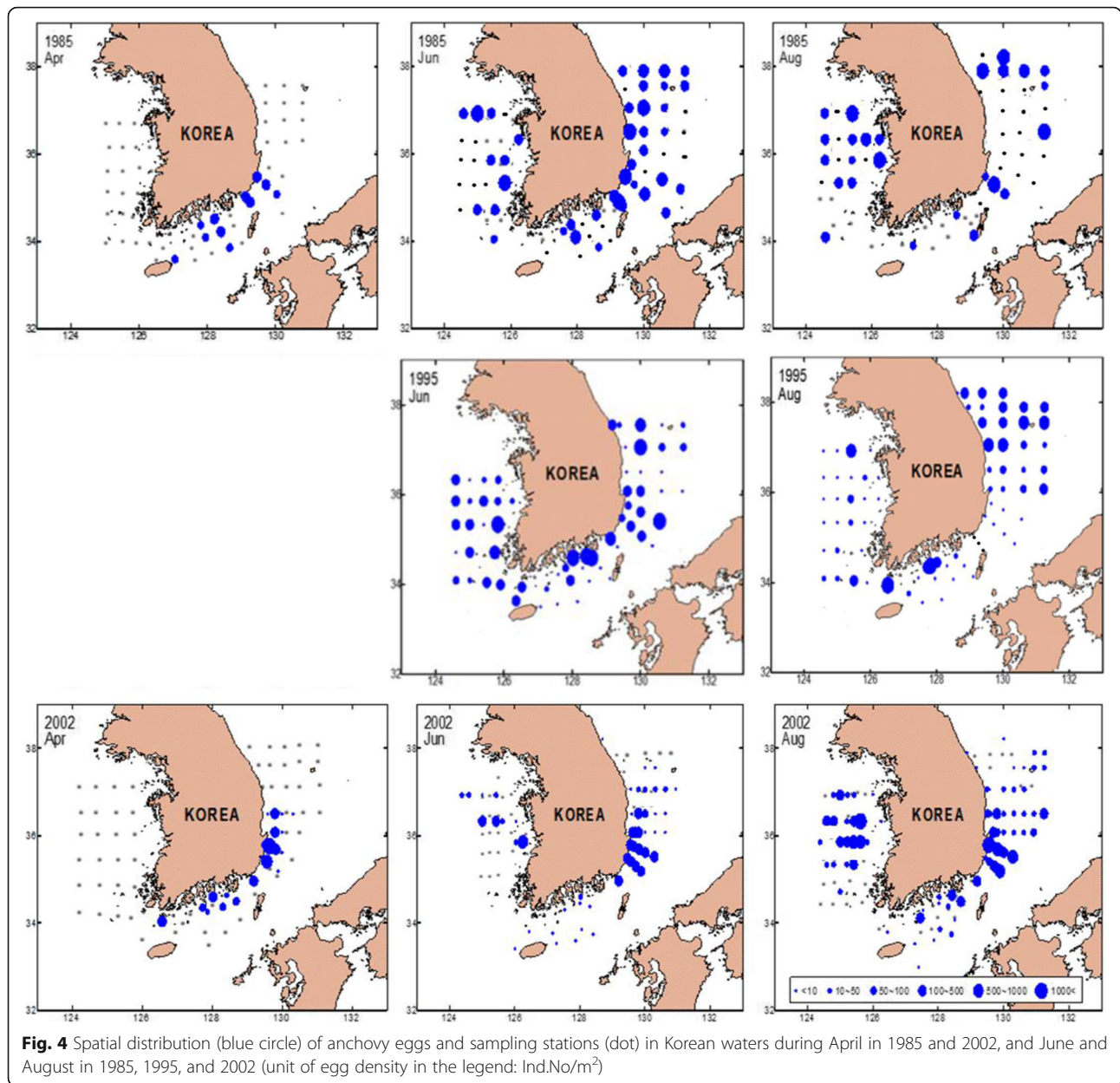


Fig. 4 Spatial distribution (blue circle) of anchovy eggs and sampling stations (dot) in Korean waters during April in 1985 and 2002, and June and August in 1985, 1995, and 2002 (unit of egg density in the legend: Ind.No/m²)

39.1–63.2% of the deviance explained in the GAM and QGAM, respectively (Table 3). Egg density in the West Sea had a negative relationship with salinity and a positive relationship with the DO difference between the 10 and 50 m layer in 1985 (Fig. 7d), and exhibited a negative relationship with the temperature difference between the 10 and 50 m layer and a positive relationship with zooplankton density in 1995 (Fig. 7e). It was also related to dome-type relationship with the salinity difference between the 10 and 50 m layer and a converse dome-shape with the DO difference between the 10 and 50 m layer in 2002 (Fig. 7f), with 18.5–41.6% and 38.6–56.3% of the deviance explained by the GAM and the QGAM, respectively (Table 3). Egg density in the South

Sea had a significant negative relationship with the temperature difference between the 10 m and 50 m layer in 1985, and a dome-shaped relationship with the salinity difference between the 10 m and 50 m layer in 1995 (Fig. 7h), with 14.3–24.1% and 14.3–38.3% of the deviance explained in the GAM and the QGAM, respectively (Table 3).

In the PCA-based GAM, the variance explained by the final regression model was 33.2–67.0% for egg density in the East Sea and 27.3–64.3% for density in the West Sea, which were higher than the values of the results obtained by the regular GAM and QGAM. Egg densities in the East Sea and West Sea exhibited similar effects of non-linear decrease for PC1 and increasing effects for

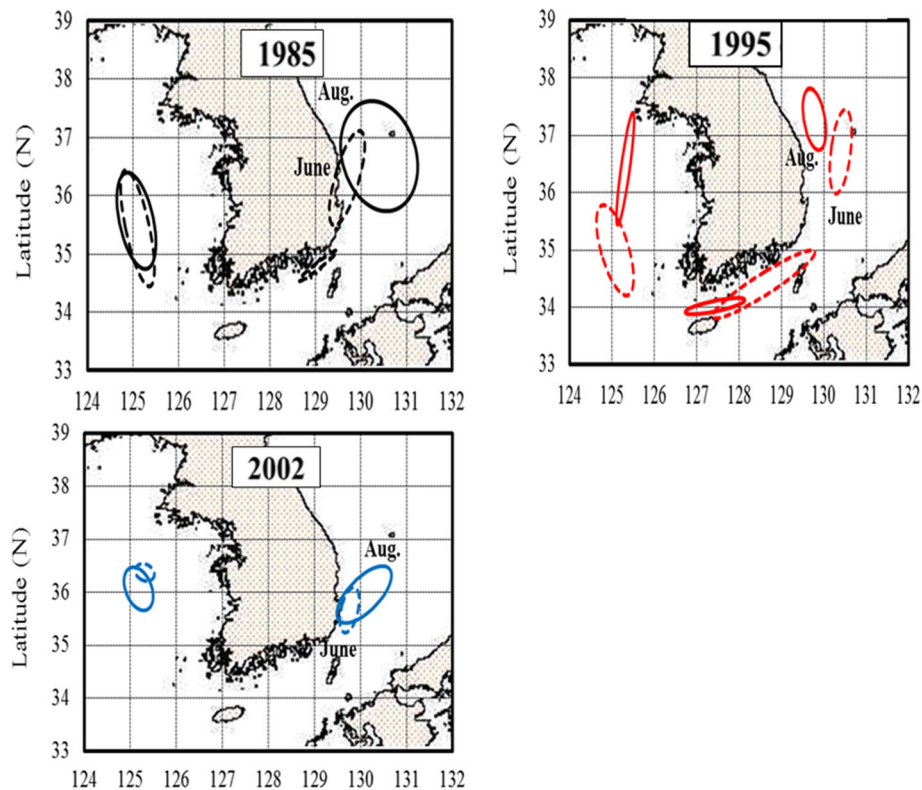


Fig. 5 Distribution of anchovy eggs explained using confocal ellipses of distribution areas, weighted by the density at each station, in Korean waters during June (dashed-line) and August (solid line) in 1985, 1995, and 2002

PC3, but had different effects for PC2, with a conversely dome-shape in the East Sea and a positive relationship in the West Sea. The yearly egg density had a significant relationship with PC1, PC2, and PC3 in the East Sea, but not in the West Sea (Table 4, Fig. 8).

Discussion

Changes in anchovy spawning grounds

Many studies have been conducted to investigate the factors affecting the distribution of anchovy eggs in Korean coastal waters; these have shown that anchovy eggs are distributed over a wide range, and their distribution shows a high skewness and kurtosis (Lim and Ok 1977; Kim 1983; Table 1 and Fig. 2). Fisheries research often uses catch statistics as an indirect survey method. Since the anchovy fishery is operated in the coastal areas of Korea, there is little information on the distribution of anchovy resources offshore. Anchovy eggs distributed on the coast of Korea are mostly 1–3-days old in spring and summer (Kim and Lo 2001). Therefore, this study aimed to determine the factors affecting the variation of anchovy spawning grounds by applying the latest analytical methods to egg distribution data collected in the past.

In this study, although egg densities differed across the region in spring, year and location, the ellipses of egg distribution in June and August mostly overlapped in adjacent areas in 1985 and 2002, while a change in spawning grounds occurred in 1995 (Fig. 5). Warm currents and nutrient upwellings strongly influence the meso-scale latitudinal heterogeneity of plankton, through physical and biological processes in Korean waters, and play a key ecological role in supporting substantial feeding environments (Jeong et al. 2009; Kim et al. 2010; Suh et al. 1999; Kim et al. 2011). Therefore, yearly or regional differences in egg density seem to be dependent on changes in oceanographic factors and productivity as well as the effect of population characteristics such as interspecies relations and density-dependent effects. Front and nutrient upwelling strongly influenced the mesoscale latitudinal heterogeneity of plankton through physical and biological processes in the southwestern East/Japan Sea (Kang et al. 2004) when water column stability was higher, favoring food concentration and egg and larval retention in the spawning areas (Basilone et al. 2006). We found that anchovy egg distribution was concentrated around areas of upwelling or vertically mixing areas in the central East Sea, reflecting the spawning grounds of adult anchovy to areas with higher nutrient concentrations.

Table 2 Results of the GAMs for the distribution of anchovy eggs in Korean waters: regular GAM (binomial with logit for the egg presence/absence and Gaussian with identity for the egg density) and quantile GAM (elf with identity at modelling the 0.10th, 0.50th, and 0.90th quantile) in total and 0.50th and 0.80th quantile (1985), 0.50th, 0.70th, and 0.90th quantile (2002) in April for the egg density (\log_{10})

	Total			1985			2002		
Family	Binomial	Gaussian	Elf	Binomial	Gaussian	Elf	Binomial	Gaussian	Elf
Link Function	Logit	Identity	Identity	Logit	Identity	Identity	Logit	Identity	Identity
R ² (ad)	0.132	0.12	0.134	0.213	0.144	0.100	0.235	0.235	0.194
DE (%)	12.4	13.0	17.7	30.5	16.8	20.0	40.5	26.3	48.0
GCV	0.211	0.814	675.630	-0.419	0.163	49.606	-0.582	0.199	82.153
Covariates(p value)									
10m depth									
Temperature	0.000***	0.000***	0.000***	0.016*	0.000***	0.000***			
Salinity	0.029*	0.085'					0.006**	0.000***	0.000***
DO		0.000***							
Difference between 10m and 50m									
Temperature									
Salinity	0.002**	0.099'	0.000***						
DO									
Feeding Index									
Zooplankton		0.021*	0.039*		0.048*				

***0.001, **0.01, *0.05, '0.1

In the eastern waters of Korea, juvenile anchovy formed fishing grounds near Kangwon Province between July and December (Park et al. 1996). This distribution was apparently related to the presence of cold water and a tidal front (Chen 2009). Phytoplankton communities and anchovy eggs in the eastern Yellow Sea are strongly affected by the formation of tidal fronts, thermal stratification, and coastal upwellings during summer (Hao et al. 2003; Kang et al. 2004). Therefore, the juvenile anchovy population in the coastal areas of the West and East Seas of Korea reflect spawning grounds and larval recruitment driven by eddies or tidal currents that are prevalent from July to November. This suggests the importance of spawning grounds in the East and West Sea during the summer season.

Oceanographic indicators of anchovy egg distribution

As anchovy schools migrate from the South Sea to the East Sea and West Sea during the spring and summer in response to oceanographic factors and feeding habitat, it is important to determine how annual fluctuations in the marine environment affect fish occurrence in the sea. By analyzing the anchovy egg data collected during 1985, 1995, and 2002, and their associated environmental variables with confocal ellipses, regular GAMs, QGAMs, and PCA-based GAMs, we aimed to address this knowledge gap.

The binomial models provided useful estimates for the probability of capturing anchovies in terms of spatial habitat evaluation (Stoner et al. 2001). In our study, because of the limitations of the sampling and observation methods (such as the high variability in egg density due to drift and high embryonic mortality, despite anchovy eggs hatching in 1–3 days in summer), we adopted a more conservative approach for the analysis, including the use of the egg presence binomial and egg density data. As shown in the results, egg density was found to be non-linearly related to temperature and salinity at a depth of 10 m, the salinity difference between 10 m and 50 m, and log-transformed zooplankton density in the Gaussian error structure. Egg presence was not significantly associated with zooplankton density, but had strong significant associations with salinity at 10 m depth and the salinity difference between 10 and 50 m depth in the binomial error structure (Table 2).

The analysis using a regular GAM may reveal the environmental parameters that control the timing of the migration of anchovies. However, regular GAMs may perform poorly when applied to unevenly distributed animals (Murase et al. 2009). The QGAM aims to estimate either the conditional median or other quantiles of the response variable, and is expected to be bell-shaped by nonlinear quantile regression models (Husson et al. 2020; Waldock et al. 2019). In this study, GAMs and QGAMs revealed that egg density was most influenced by oceanographic

Table 3 Yearly and area based results in the East, West, and South Sea during June and August in 1985, 1995, and 2002 for the egg density (\log_{10}) by regular GAMs (Gaussian with identity) of the egg density. Location was composed of longitude and latitude for the station

	East Sea			West Sea			South Sea	
	1985	1995	2002	1985	1995	2002	1985	1995
R^2 (adj)	0.250	0.370	0.444	0.133	0.322	0.456	0.506	0.111
DE (%)	29.9	41.1	51.1	18.5	35.7	53.0	60.2	14.3
GCV scores	0.658	0.781	0.453	0.507	0.689	0.770	0.478	0.949
Covariates (<i>p</i> value)								
10 m depth	Temperature							
	Salinity				0.019*			
	DO		0.000***					
Difference between 10m and 50m	Temperature				0.000***		0.017**	
	Salinity	0.003**				0.011*		
	DO		0.000***	0.045*	0.065'		0.071'	0.034*
Feeding index	Zooplankton	0.048*				0.042*		

***0.001, **0.01, *0.05, '0.1

factors with a dome shape in the upper 0.70–0.90th quantiles of egg density.

The results of the analysis with the regular GAM showed that the high egg density was positively related to water temperature, and had a dome shape relationship with salinity, which represents the characteristics of coastal water because the spawning adults school in the coastal water in spring. During summer, egg density had a dome shape relationship with temperature, and was inversely correlated with DO and positively correlated with salinity and zooplankton density (Table 2, Table 3, Fig. 6, and Fig. 7). The energy allocated to multiple spawning events in small pelagic fish is derived primarily from feeding, and spawning is related to both dietary intake and the nutritional status of the fish (Somarakis et al. 2004). This indicates that the spawning grounds occurred within the range of water temperatures in the productive areas of the East Sea. In the West Sea, egg density was inversely correlated to the thermocline,

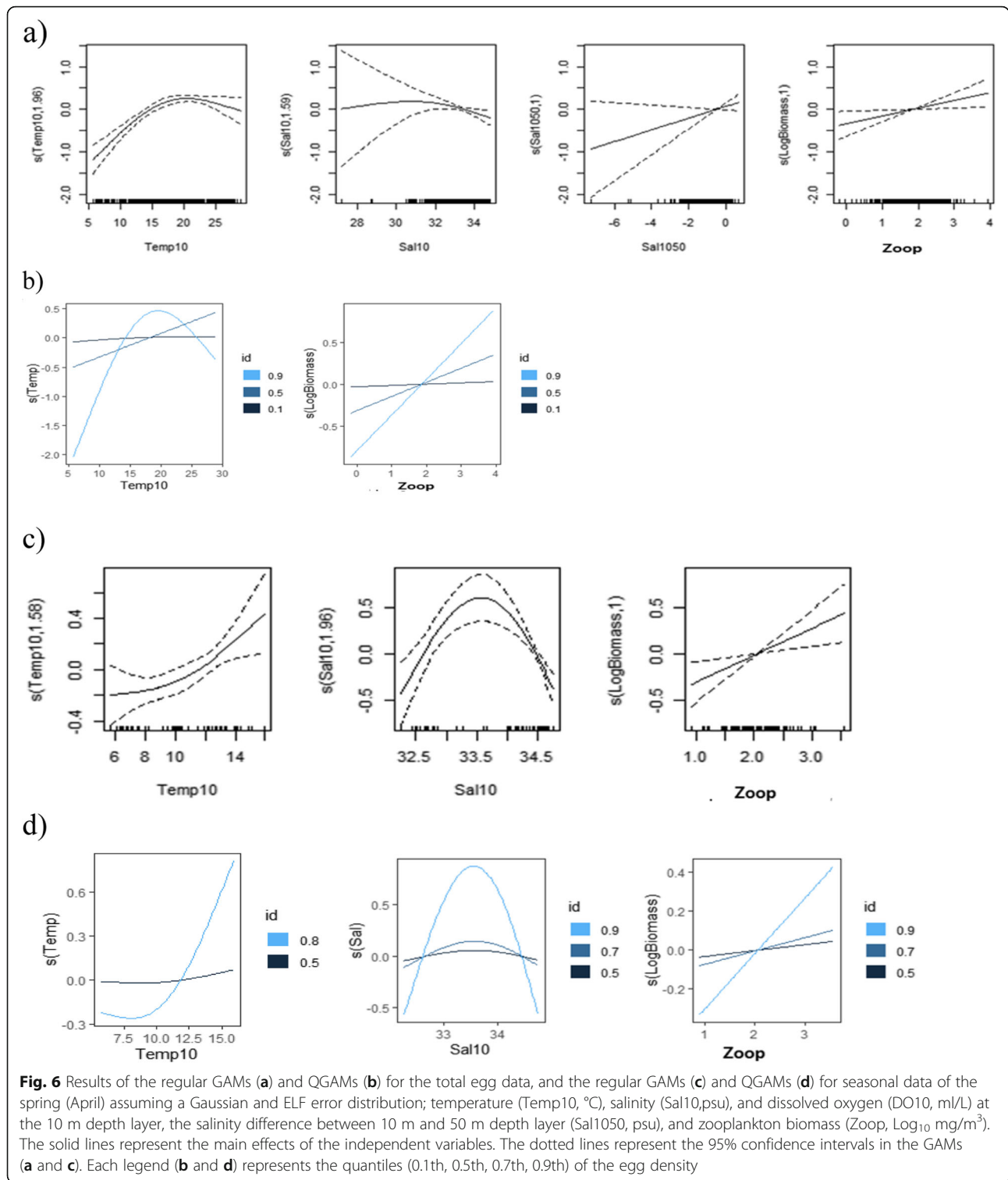
exhibited a dome shaped relationship with salinity and was associated with high densities of zooplankton, indicating that schooling occurs in the productive frontal areas between the coastal and offshore waters as the food becomes available for spawning. These observations were in line with those reported by Kim (1983).

The PCA-based GAM reduces multi-collinearity of the explanatory variables and improves the performance of regular GAMs (Alvarez-Fernandez et al. 2012; Liu et al. 2019). The PCA utilizes a mathematical method that transforms the original set of correlated variables into a new set of independent uncorrelated variables, which gives the linear combination of the original set of data (Li et al. 2017). The new variables are ordered in such a way that the first variable explains most of the variance in the data. Zhao et al. (2014) compared the performance of regular GAMs and PCA-based GAMs for modeling the relationship between diversity indices and environmental variables. They suggested that the PCA-based GAMs provide

Table 4 Results of the PCA-based GAM assuming a Gaussian error distribution, for the egg data from the summer (June and August) in the East Sea and West Sea. Composition of PC1, PC2, and PC3 are represented from Fig. 3

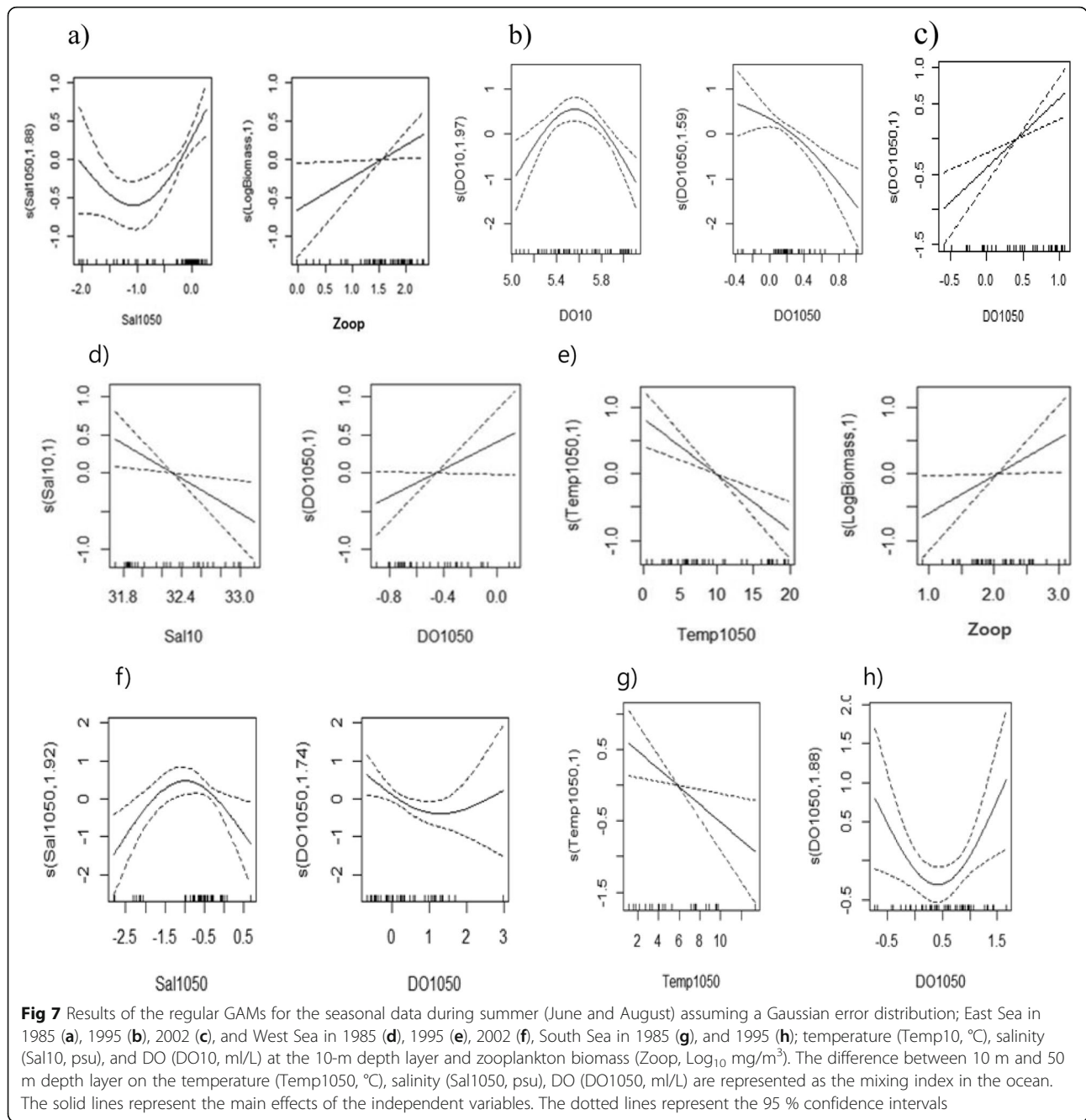
	East Sea				West Sea			
	Total	1985	1995	2002	Total	1985	1995	2002
R^2 (adj)	0.172	0.224	0.337	0.502	0.190	0.471	0.212	0.36
DE (%)	22.1	33.2	42.7	67.0	26.2	64.3	27.3	50.1
GCV	0.891	0.739	0.890	0.541	0.857	0.431	0.823	1.005
PC1	0.003**	0.042*	0.069'	0.009*	0.004**	0.104	0.423	0.089'
PC2	0.013*	0.011*	0.035*	0.029*	0.077	0.171	0.244	0.012*
PC3	0.002**	0.022*	0.008**	0.108	0.011*	0.009**	0.982	0.376

***0.001, **0.01, *0.05, '0.1



a better approach and a higher prediction accuracy. The variance explained by the PCA-based GAM was higher than the regular GAM and QGAM for egg density in the East and West Sea (Table 4). Therefore, PCA-based

GAMs are efficient tools for identifying the oceanographic factors influencing anchovy spawning grounds in Korea. In this study, the overall physical environment was mainly defined by PC1, the chemical environment by PC2, and



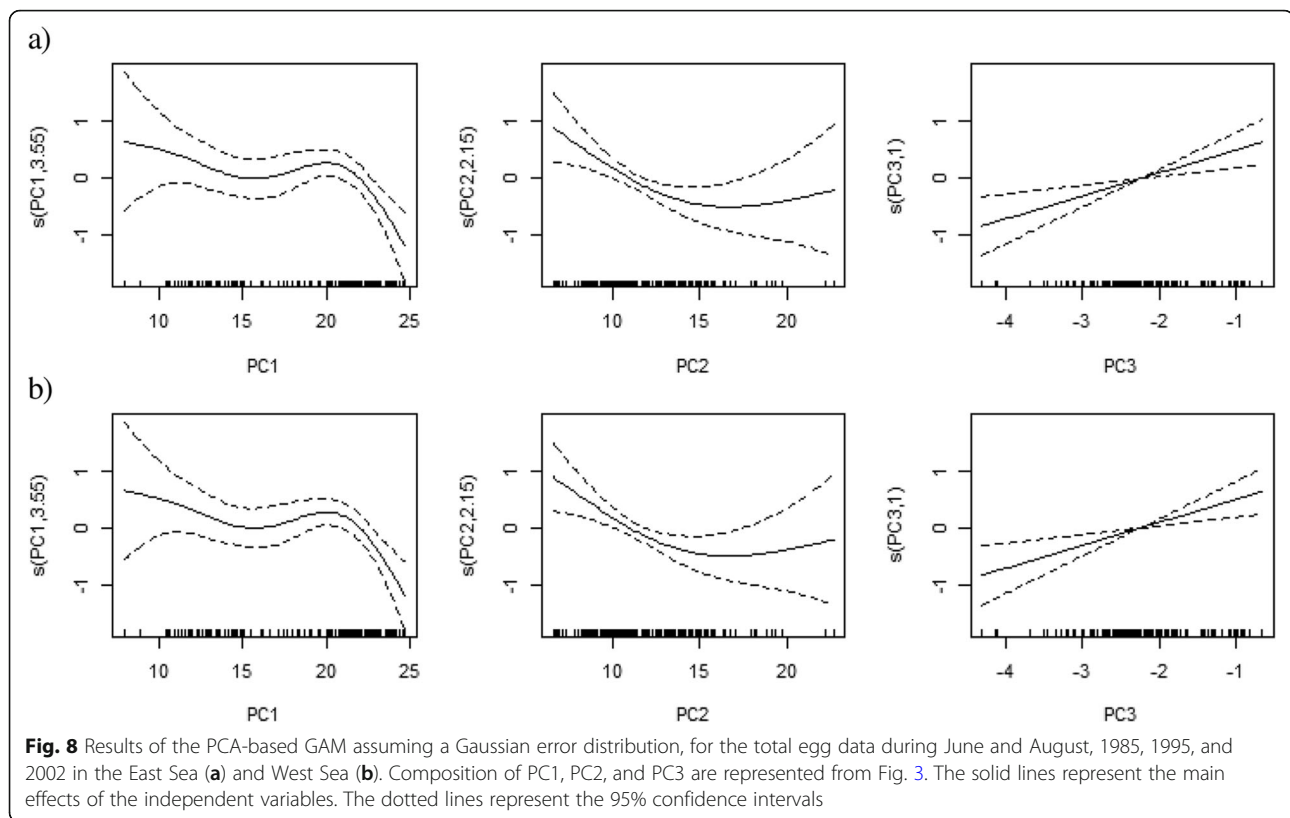
biological environment by PC3, with zooplankton biomass as an obvious indicator of spawning grounds in the East Sea and the West Sea (Table 4, Fig. 3 and Fig. 8). It appears to be a useful method in cases where the variables are more diverse and complex.

Although our results demonstrate that understanding the distribution of eggs for only three years can contribute to the identification of migration patterns of spawning anchovy populations, further improved results may be achieved through using long-term data in the future. Such

insights may improve fishery management because they are independent of fishery data and highlight important anchovy habitats in the South, East, and Western Seas of Korea, as well as their relationships to oceanographic conditions.

Conclusions

We analyzed the distribution of anchovy eggs in the spring and summer of 1985, 1995, and 2002, and identified the marine environmental factors influencing the



occurrence of anchovy spawning grounds. Anchovy eggs were widely distributed in the East and West seas during summer, and the spawning areas were affected by the marine environment. Aspects of the physical, chemical, and biological marine environment, as determined by regular GAMs, QGAMs and PCA-based GAMs were significantly and non-linearly related to the distribution of anchovy eggs. By analyzing the distribution of anchovy eggs off the Korean coast, our results revealed the optimal temperature and salinity conditions, in addition to highly productive and vertical mixing, as the key indicators of the location of major spawning grounds. Our study outcomes have demonstrated that the three kinds of GAMs are useful tools for identifying the oceanographic factors that influence anchovy spawning habitats in Korea. Changes in the climate and marine environment will continue to alter anchovy spawning grounds. This retrospective study can be used for predicting future spawning field variations and for the continuous management of fisheries resources.

Abbreviations

GAMs: Generalized additive models; QGAMs: Quantile generalized additive models; PCA: Principle component analysis; PC: Principle component; DO: Dissolved oxygen; NIFS: National Institute of Fisheries Science; KODC: Korea Oceanography Data Center; MANOVA: Multi-factor analysis of variance; UBRE: Un-biased risk estimator; GCV: Generalized cross-validation;

REML: Restricted maximum likelihood; AIC: Akaike's Information Criterion; DE: Deviance explained; VIF: Variance inflation factor

Acknowledgements

The authors thank all the people who helped publish this paper. This study was supported by a grant from the National Research Foundation of Korea (2017 R1D1A1B03031200). We also thank NIFS for its provision of data used in this paper.

Authors' contributions

All authors designed the direction of this study. JYK implemented the statistics and wrote the manuscript. JBL and YSS conducted some parts of the analysis and revised the manuscript. The authors read and approved the final manuscript.

Availability of data and materials

Data used in this paper were collected by a plankton survey in the Korean waters, operated by NIFS. Availability of the data would be determined by the institution.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute for Mathematical Convergence, Gyeongpook National University, #80 Daehakro, Bukgu, Daegu 41566, Republic of Korea. ²National Institute of Fisheries Science, #216 Gijanghaean-ro, Gijang-gun, Busan 46083, Republic of Korea.

Received: 31 March 2020 Accepted: 26 May 2020

Published online: 28 July 2020

References

- Alheit J, Bakun A. Population synchronies within and between ocean basins: Apparent teleconnections and implications as to physical–biological linkage mechanisms. *J Mar Syst.* 2010;79:267–85.
- Alvarez-Fernandez S, Lindeboom H, Meesters E. Temporal changes in plankton of the North Sea: community shifts and environmental drivers. *Mar Ecol Prog Ser.* 2012;462:21–38.
- Basilone G, Guisande C, Patti B, Mazzola S, Cuttitta A, Bonanno A, Vergara A, Maneiro I. Effect of habitat conditions on reproduction of the European anchovy (*Engraulis encrasicolus*) in the Strait of Sicily. *Fish Oceanogr.* 2006;15:271–80.
- Boyra G, Rueda L, Coombs S, Sundby S, Adlandsvik B, Santos M, Uriarte A. Modelling the vertical distribution of eggs of anchovy (*Engraulis encrasicolus*) and sardine (*Sardina pilchardus*). *Fish Oceanogr.* 2003;12:381–95.
- Chen C. Chemical and physical fronts in the Bohai, Yellow and East China seas. *J Mar Syst.* 2009;78:394–410.
- Choi S, Kim J, Kim S, Choi Y, Choi K. Biomass estimation of anchovy (*Engraulis japonicus*) by acoustic and trawl survey during spring seasons in the southern Korean waters. *J Kor Soc Fish Res.* 2001;4:20–9.
- Denis V, Lejeune J, Robin J. Spatio-temporal analysis of commercial trawler data using General Additive models: patterns of Loliginid squid abundance in the north-east Atlantic. *ICES J Mar Sci.* 2002;59:633–48.
- Fasiolo M, Wood S, Goude W, Nedellec R. Version 1.3.2. Package 'qgam' R-3.4.2; 2020.
- Hao W, Jian S, Ruijing W, Lei W, Yi'an L. Tidal front and the convergence of anchovy (*Engraulis japonicus*) eggs in the Yellow Sea. *Fish Oceanogr.* 2003;12:434–42.
- Hoeffle H, Solemdal P, Korsbrette K, Johannessen M, Bakkeplass K, Kjesbu 1 O. Variability of northeast Arctic cod (*Gadus morhua*) distribution on the main spawning grounds in relation to biophysical factors. *ICES J Mar Sci.* 2014;71:1317–31.
- Husson B, Certain G, Filin A, Planque B. Suitable habitats of fish species in the Barents Sea. 2020; doi:<https://doi.org/10.1101/2020.01.20.912816>.
- Jeong H, Cho K, Kwoun C, Kim S. Fluctuation of tidal front and expansion of cold water region in the Southwestern Sea of Korea. *J Korean Soc Mar Environ Saf.* 2009;15:289–96.
- Kang J, Kim W, Chang K, Noh J. Distribution of plankton related to the mesoscale physical structure within the surface mixed layer in the southwestern East Sea, Korea. *J Plankton Res.* 2004;26:1515–28.
- Kim J. Distribution of anchovy eggs and larvae off the western and southern coasts of Korea. *Bull Korean Fish Soc.* 1983;21:139–44.
- Kim J, Choi Y. Vertical distribution of anchovy, *Engraulis japonica* eggs and larvae. *Bull Korean Fish Soc.* 1988;21:139–44.
- Kim J, Lo N. Temporal variation of seasonality of egg production and the spawning biomass of Pacific anchovy, *Engraulis japonicus*, in the southern waters of Korea in 1983–1994. *Fish Oceanogr.* 2001;10:297–310.
- Kim J, Kang Y, Oh H, Suh Y, Hwang J. Spatial distribution of early life stages of anchovy (*Engraulis japonicus*) and hairtail (*Trichiurus lepturus*) and their relationship with oceanographic features of the East China Sea during the 1997–1998 El Niño Event. *Estuar Coast Shelf Sci.* 2005;63:13–21.
- Kim S, Zhang C, Kim J, Oh J, Kang S, Lee J. Climate variability and its effects on major fisheries in Korea. *OSJ.* 2007;42:179–92.
- Kim S, Go W, Kim S, Jeong H, Yamada K. Characteristics of ocean environment before and after coastal upwelling in the southeastern part of Korean peninsula using an in-situ and multi-satellite data. *J Kor Soc Mar Environ & Saf.* 2010;16:345–52.
- Kim D, Yang E, Kim K, Shin C, Park J, Yoo S, Hyun J. Impact of an anticyclonic eddy on the summer nutrient and chlorophyll a distributions in the Ulleung Basin, East Sea (Japan Sea). *ICES J Mar Sci.* 2011;69:23–9.
- Li S, Zhai L, Zou B, Sang H, Fang X. A Generalized Additive Model Combining Principal Component Analysis for PM2.5 Concentration Estimation. *ISPRS Int J Geo-Inf.* 2017;6:248.
- Lim J, Ok I. Studies on the occurrence and distribution of eggs and larvae anchovy in the Korean waters. *Bul Fish Res Dev Agency.* 1977;16:73–86.
- Liu X, Wang J, Zhang Y, Yu H, Xia B, Zhang C, Ren Y, Xue Y. Comparison between two GAMs in quantifying the spatial distribution of *Hexagrammos otakii* in Haizhou Bay, China. *Fish Res.* 2019;218:209–17.
- Murase H, Nagashima H, Yonezaki S, Matsukura R, Kitakado T. Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. *ICES J Mar Sci.* 2009;66:1417–24.
- Park J, Choi S, Kim J, Lee J. Distribution of anchovy, *Engraulis japonica* (Houttuyn), in the coastal waters of Kangwon Province in Korea. *Bull Korean Soc Fish Tech.* 1996;32:223–34.
- Park J, Im Y, Cha H, Suh Y. The relationship between oceanographic and fishing conditions for anchovy, *Engraulis japonica*, in the southern sea of Korea. *J Korean Soc Fish Res.* 2004;6:46–53.
- Sokal R, Rohlf F. *Biometry: The Principles and Practice of Statistics in Biological Research.* 3rd ed. New York: W. H. Freeman and Co.; 1995.
- Somarakis S, Palomera I, Garcia A, Quintanilla L, Koutsikopoulos C, Uriarte A, Motos L. Daily egg production of anchovy in European waters. *ICES J Mar Sci.* 2004;61:944–58.
- Stoner A, Manderson J, Pessutti J. Spatially explicit analysis of estuarine habitat for juvenile winter flounder: combining generalized additive models and geographic information systems. *Mar Ecol Prog Ser.* 2001;213:253–71.
- Suh Y, Jang L, Kim J. Study of a recurring anticyclonic eddy off Wonsan coast in northern Korea using satellite tracking drifter, satellite ocean color and sea surface temperature imagery. *J Kor Soc Rem Sens.* 1999;16:211–20.
- Vaz S, Martin C, Eastwood P, Ernande B, Carpentier A, Meaden G, Coppin F. Modelling species distributions using regression quantiles. *J Appl Ecol.* 2008;45:204–17.
- Waldock C, Stuart-Smith R, Edgar G, Bird T, Bates A. The shape of abundance distributions across temperature gradients in reef fishes. *Ecol Lett.* 2019;22:685–96.
- Wood S. Version 1.8-31, 2019 'mgcv' R-3.4.2.
- Zhao J, Cao J, Tian S, Chen Y, Zhang S, Wang Z, Zhou X. A comparison between two GAM models in quantifying relationships of environmental variables with fish richness and diversity indices. *Aquat Ecol.* 2014;48:297–312.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

