An information-theoretic approach for model selection in habitat preference evaluation of Japanese medaka (*Oryzias latipes*)

Shinji Fukuda [1)*], Kazuaki Hiramatsu [2)] and Shuji Okushima [3)]

Abstract

In keeping with the growing concern over sustainable development, there is an increasing need for clearer goals and applicable techniques to achieve ecological conservation and restoration. To accomplish nature-oriented planning and management, it is essential to understand the ecology and habitat requirements of target species. In this study, we applied the information-theoretic approaches of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to identify the best model and most significant factor which can explain the habitat selection of Japanese medaka (*Oryzias latipes*). A fuzzy preference intensity model (FPIM), i.e. a hybrid model of simplified fuzzy reasoning and a genetic algorithm, was introduced to evaluate the habitat preference of the fish. The present result suggests that the lateral cover ratio is the most significant factor governing the habitat preference of the fish, and the FPIM that considers the four factors of water depth, current velocity, lateral cover ratio, and percent vegetation coverage has the best prediction ability among the candidates.

Keywords: Information criterion, fish habitat, physical environment, habitat preference model, habitat assessment

Introduction

Today, sustainable development has now become a key issue all over the world. To achieve this, many research studies and practical activities have been carried out in many countries (e.g. Harby *et al*., 2004). Since the need for a balance between development and environmental conservation to achieve a better quality of life is widely recognized, clearer goals and applicable techniques are urgently required for the sustainable development and management. However, it is very difficult to maintain this balance because of the complicated trade-offs between ecological, social, and economical demands.

1) Institute of Tropical Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.
2) Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.
3) National Institute for Rural Engineering, Kannondai 2-1-6, Tsukuba-shi, Ibaraki 305-8609, Japan.
*Corresponding author
E-mail: shinji-fkd@agr.kyushu-u.ac.jp
TEL: +81-92-642-3075
FAX: +81-92-642-3077

Several studies and challenges for the conservation and restoration of river ecosystems have been conducted by using the physical habitat simulation system (PHABSIM: Bovee *et al*., 1998), habitat evaluation procedures (HEP: U. S. Fish and Wildlife Service, 1980a), etc. The objectives of these techniques are to numerically quantify the anthropogenic impacts on flora and fauna in target basins. In general, these impacts are evaluated as changes in habitat potential, often in terms of weighted usable area (WUA: Bovee *et al*., 1998) or habitat unit (HU: U. S. Fish and Wildlife Service, 1980a), in keeping with the physical environments such as water depth, velocity, and substrate. Since both WUA and HU are based on the habitat preference of a target species within a given environment, the evaluation techniques for habitat preferences are one of the essential components requiring continued study.

To date, many researchers have proposed and applied habitat prediction models to evaluate the habitat preference of a target species. These models are based on the habitat suitability index (HSI: U. S. Fish and Wildlife Service, 1980b; Lechowicz, 1982; Inoue and Nakano, 1999; Urabe and Nakano, 1999), linear and nonlinear regression models (Guey *et al*., 2000; Vadas and Orth, 2001; Guey *et al*., 2003), fuzzy rule-based models (Hiramatsu *et al*., 2003; Fukuda *et al*., 2005; Rüger *et al*., 2005; Fukuda *et al*., 2006b; Mouton *et al*., 2006), artificial neural network models (Brosse and Lek, 2000; Fukuda *et al*., 2006a), etc. In addition, these models are also used to understand the ecology and the habitat requirements of the target species.

Since the consideration of all the ecological requirements is practically impossible in planning and management, it is crucial to identify which factor should take priority during conservation and restoration. One of the methods is to consider the preference weight for each factor. Based on the study by Sekine *et al*. (1997), several weighting operations have been proposed. However, it appears to be difficult to apply these techniques to the studies based on field survey because habitat preference is affected by complicated interactions between multiple environmental factors. The other method is to employ an information-theoretic approach, e.g. Akaike information criterion (AIC) and Bayesian information criterion (BIC); this approach is now widely known and applied to various fields of research for the purpose of obtaining the 'best model' among candidates (Burnham and Anderson, 2002). This approach provides us with information regarding certain aspects of the significance of a particular factor by considering the balance between prediction errors and the number of parameters evaluated within the model employed.

The main aim of this paper is to demonstrate the applicability of the information-theoretic approach for the identification of the best model and the significant factors in predicting the spatial distribution of Japanese medaka (*Oryzias latipes*) dwelling in the agricultural canals of Japan. We further verify which factor significantly affects the habitat preferences of Japanese medaka.

Methods

Study site

The study was carried out in an agricultural canal located in Kurume City, Fukuoka, Japan. The spring-fed canal runs through paddy fields, is used for both irrigation and drainage purposes and flows into the Kose River, which is a tributary of the Chikugo River. The fish species dwelling in this canal are presented elsewhere (Fukuda *et al*., 2006b). We established a 50-m study reach (width, 1.6–2.0 m, gradient = 0.3%) in this canal. A concrete agricultural facility is located approximately 31 m from the downstream end. The riparian zone of the reach was not covered with any trees or bushes. The discharge of the reach varies widely between irrigation and non-irrigation periods; however, the discharge in each period generally remains stable. The water temperature remained stable (16.1–18.6 °C) during the survey.

Field survey

During the non-irrigation period, the spatial distributions of Japanese medaka and the physical habitat characteristics of water depth (henceforth referred to as depth), current velocity (velocity), lateral cover ratio (cover), and percent vegetation coverage (vegetation) in the study reach were surveyed on two sunny days, i.e. 14 October and 5 November 2004. The study was performed in this period because in the irrigation period, spatial uniformity in both fish distribution and physical environment occurs due to the backwaters that are caused by weir control. Hence, these spatial distributions were not affected by any agricultural activities or agricultural chemicals. Based on Yoshioka's study (Yoshioka, 1963), the survey is found to be conducted right after the end of the spawning season. We first observed the fish distribution (11:00–14:30) and then surveyed the physical habitat characteristics within the reach.

The spatial distribution of Japanese medaka was observed visually from the bank; the observer moved slowly and carefully to avoid any fish disturbance caused by the observer's activity. The number of fish was counted in units of five fish to take into consideration the patterns of school formation, i.e. fish in a small school (less than five fish) were not counted. Observations were repeated eight times, and the results were averaged to avoid observation variance.

Immediately after completing fish observation, the four physical habitat characteristics of depth, velocity, cover, and vegetation were surveyed to establish a relationship between the physical environment and habitat preferences of Japanese medaka. First, depth and velocity were measured to divide the study reach into small water bodies possessing similar characteristics with regard to these two physical

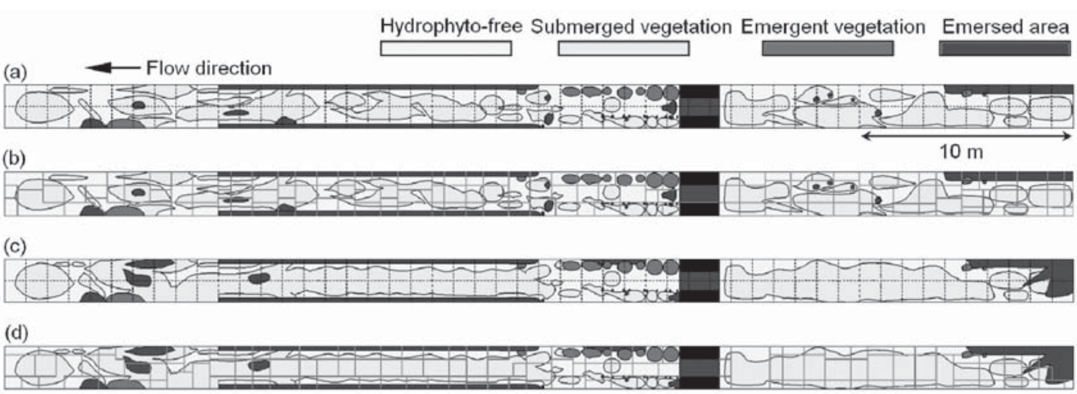

Fig. 1. Schematics of the overview of the study reach (**a** and **c**) and water bodies (**b** and **d**) in similar physical conditions of water depth and current velocity on 14 October (**a** and **b**) and 5 November (**c** and **d**) 2004. Solid lines denote the boundary between the water bodies. See text for the details of the study reach.

parameters (Fig. 1). Depth was measured with a stainless steel ruler and velocity, with a portable propeller current meter (KENEK, V-303) at three lateral points comprising a midpoint and two near-shore points at longitudinal intervals of 1 m. By using the measurements of depth and velocity, the reach was divided into water bodies (Fig. 1). Next, the other two factors of cover and vegetation were calculated from the schematic diagrams of the water bodies (Fig. 1). The lateral cover ratio—an index of the spatial structure of a water body - is defined as a function of the presence of lateral cover which comprises the water's edge, a dyke or anything that emerges from the water surface and surrounds the water body. The cover thus consists of four components (four lateral sides). The full, i.e. the maximum cover ratio is 100%. Each of the cover components is assigned a score of 25%. In the definition of the cover, objects attached to more than 90% of the boundary between water bodies, i.e. the solid lines in Fig. 1, were regarded as cover components, that is, we considered only instream and undersurface cover structures that may have had the same effects as the margin of the stream. Percent vegetation coverage is defined as the percentage of the area covered with aquatic vegetation in each water body. Both submerged and emergent vegetation were pooled because of their same roles, i.e. providing food and shelter from predators and fast-flowing currents.

In the following analyses, the fish distribution data that we used were the observed fish population density obtained for the *i*th water body $\rho_{o,i}$ (individuals per square metre), where i ($i = 1, 2, \ldots, n$) denotes the index of the water body and n, the total number of water bodies.

Habitat preference model

In this study, we utilized a fuzzy rule-based preference intensity model (FPIM) to quantify the habitat preference of Japanese medaka (Hiramatsu *et al.*, 2003; Fukuda *et al.*, 2005, 2006b). The FPIM is a hybrid model based on the two artificial intelligence techniques of fuzzy reasoning and simple genetic algorithms (GA). Fuzzy reasoning was introduced to consider the essential vagueness of fish behavior and habitat preference, the uncertainty in field measurement errors, and the dispersions of the physical environment within a water body. We applied simplified fuzzy reasoning in this study for its simplicity and high performance (Mizumoto, 1995). An overview of fuzzy if-then rules is shown in Fig. 2. The simple GA was employed to optimize the singletons in the consequence part of the fuzzy if-then rules. The following is a brief explanation of the optimization procedure. Firstly, the GA proposes an initial consequence part of the fuzzy rules. Secondly, the composite habitat preference is calculated by using a simple multiplication method expressed as

$$P_i = P_{d,i} \times P_{v,i} \times P_{c,i} \times P_{veg,i} \qquad [1]$$

where P_i denotes the habitat preference in the *i*th water body and the abbreviations d, v, c, and veg indicate depth, velocity, cover, and vegetation, respectively. Thirdly, the spatial distribution of Japanese medaka is predicted using the equation

$$\rho_{c,i} = \left(P_i \Big/ \sum_{i=1}^{n} P_i \right) \cdot \sum_{i=1}^{n} \rho_{o,i} \qquad [2]$$

where $\rho_{c,i}$ is the calculated fish population density in the *i*th water body and $\rho_{o,i}$, the observed density

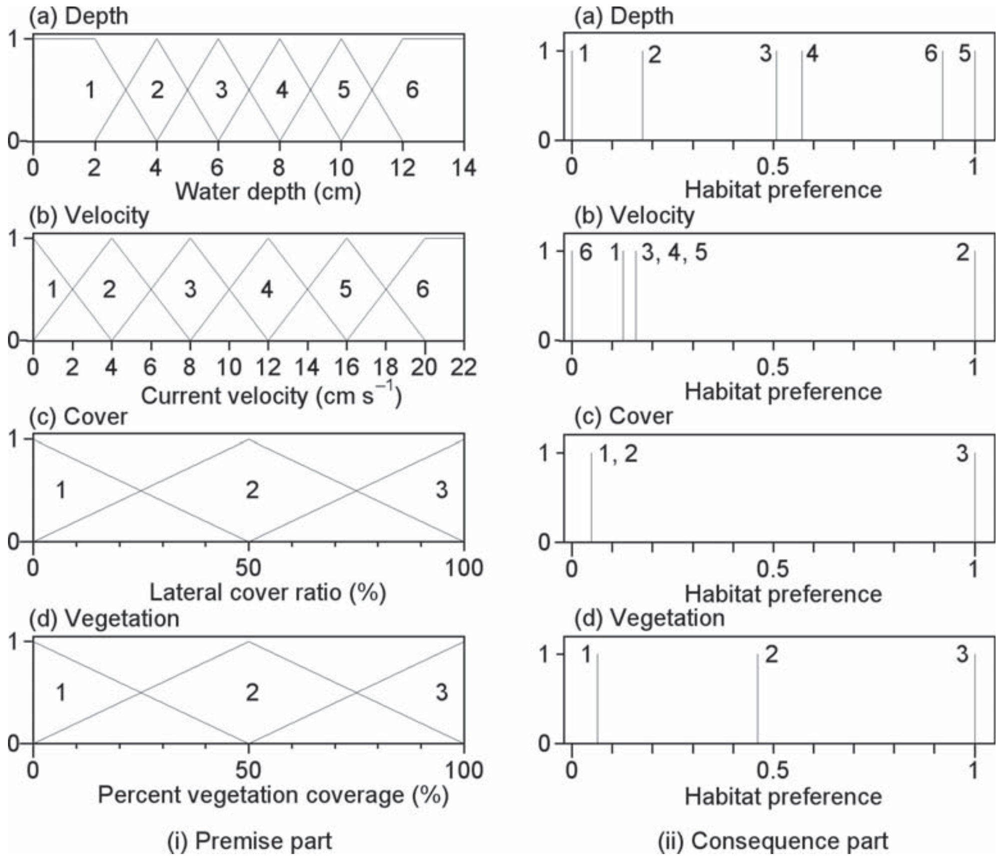

Fig. 2. Schematic showing the overview of the fuzzy if-then rules of the FPIM premise part (i) and consequence part (ii) of the four-factor composite model. The singleton 1 in the consequence part of depth is set to be zero to consider the ecology of Japanese medaka (Fukuda *et al*., 2006b).

(Fukuda *et al*., 2006b). Fourthly, the mean square error (MSE) between the predicted and observed fish population density is calculated. Then, the GA repeatedly modifies the singletons so as to minimize the MSE. Finally, the optimized FPIM is obtained. The details of the modelling procedure and calibration conditions have been described in Fukuda *et al*. (2005, 2006b).

Data analysis

To identify the most significant factor and the combination of factors for the prediction of the spatial distribution of Japanese medaka, we applied the information-theoretic approaches of AIC and BIC. The analyses were as follows. Firstly, we developed the FPIM for every combination of the four environmental factors of depth, velocity, cover, and vegetation (Table 1). Next, the MSE between the predicted and observed fish population density was calculated for all of the models. Finally, every model was evaluated by the AIC with a correction term for small sample sizes.

Table 1. Conditions and results of the $\mathrm{AIC_c}$ and BIC calculation. The number of parameters consists of the model parameters including one for σ^2.

Model index	Environmental factors	Number of parameters	MSE	$\mathrm{AIC_c}$	BIC	$\mathrm{AIC_c}$ difference	BIC difference	Rank		
								MSE	$\mathrm{AIC_c}$	BIC
1	D, V, C, Veg	18	10.68	1439.1	1501.1	0.0	0.0	1	1	1
2	D, V, C	15	15.32	1529.4	1581.4	90.3	80.3	3	3	5
3	D, V, Veg	15	16.44	1548.4	1600.4	109.3	99.3	5	6	6
4	D, C, Veg	12	15.78	1530.8	1572.7	91.6	71.6	4	4	3
5	V, C, Veg	13	13.70	1494.9	1540.2	55.8	39.1	2	2	2
6	D, V	12	20.49	1601.0	1642.9	161.9	141.8	9	10	11
7	D, C	9	20.21	1590.8	1622.4	151.7	121.3	8	8	9
8	D, Veg	9	25.51	1653.4	1685.1	214.3	184.0	13	14	14
9	V, C	10	16.82	1543.5	1578.6	104.4	77.5	6	5	4
10	V, Veg	10	20.62	1598.4	1633.5	159.2	132.4	10	9	10
11	C, Veg	7	19.78	1580.7	1605.4	141.6	104.3	7	7	7
12	D	6	28.20	1674.0	1695.2	234.9	194.1	15	15	15
13	V	7	23.23	1623.9	1648.6	184.8	147.6	12	12	12
14	C	4	22.42	1608.1	1622.4	169.0	121.3	11	11	8
15	Veg	4	26.14	1649.4	1663.6	210.3	162.5	14	13	13

$$\mathrm{AIC_c} = -2\log_e(\text{maximum likelihood}) + 2k\frac{n}{n-k-1} \qquad [3]$$

and

$$\mathrm{BIC} = -2\log_e(\text{maximum likelihood}) + k\log_e n \qquad [4]$$

where n denotes the sample size, i.e. the number of data used, and k, the total number of estimable parameters within the model (Burnham and Anderson, 2002; Johnson and Omland, 2004). Since all the FPIMs are optimized by the simple GA so as to minimize the MSE between the observed and predicted fish population density, the residuals of the prediction model (Eq. 2) are assumed to be independent and normally distributed with a constant variance σ^2. Hence, the model structure is expressed as

$$\rho_{o,i} = \rho_{c,i} + \varepsilon_i = \left(P_i \Big/ \sum_{i=1}^{n} P_i\right) \cdot \sum_{i=1}^{n} \rho_{o,i} + \varepsilon_i \qquad [5]$$

where $\rho_{o,i}$ is hypothesized to be a function of the environmental factors considered. The residuals of the models

$$\varepsilon_i = \rho_{c,i} - \rho_{o,i} \qquad [6]$$

have the probability distribution $f(\rho_{o,i}|\theta)$ given as

$$f(\rho_{o,i}|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot (\rho_{o,i} - \rho_{c,i})^2\right\} \tag{7}$$

where θ is a vector of k parameters, i.e. the number of model parameters (corresponding to the number of singletons in the consequence part of the fuzzy rules to be determined) including one for σ^2. The k values of the candidates are summarized in Table 1. The likelihood is simply the product of Eq. 7 over the n observations.

$$\begin{aligned} L(\rho_{o,i}|\theta) &= \prod_{i=1}^{n} f(\rho_{o,i}|\theta) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot (\rho_{o,i} - \rho_{c,i})^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (\rho_{o,i} - \rho_{c,i})^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} \varepsilon_i^2\right\} \end{aligned} \tag{8}$$

Taking the logarithm of Eq. 8, we then obtain the log likelihood

$$l(\rho_{o,i}|\theta) = -\frac{n}{2}\log_e 2\pi - \frac{n}{2}\log_e \sigma^2 - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} \varepsilon_i^2 \tag{9}$$

To achieve the maximum likelihood estimator $\hat{\sigma}^2$, we firstly differentiate Eq. 9 by σ^2 as follows.

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \cdot \sum_{i=1}^{n} \varepsilon_i^2 \tag{10}$$

we then derive the estimator by taking the condition that Eq. 10 is equal to zero, that is,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 \tag{11}$$

Substituting Eq. 11 into Eq. 9, the maximum log-likelihood (MLL) is thus obtained as

$$\mathrm{MLL} = -\frac{n}{2}\log_e 2\pi - \frac{n}{2}\log_e \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \frac{n}{2} \tag{12}$$

Consequently, Eqs. 3 and 4 are rewritten as follows.

$$\mathrm{AIC_c} = n\log_e 2\pi + n\log_e\left\{\frac{1}{n}\sum_{i=1}^{n}(\rho_{o,i} - \rho_{c,i})^2\right\} + n + 2k + \frac{2k(k+1)}{n-k-1} \tag{13}$$

$$\mathrm{BIC} = n\log_e 2\pi + n\log_e\left\{\frac{1}{n}\sum_{i=1}^{n}(\rho_{o,i} - \rho_{c,i})^2\right\} + n + k\log_e n \tag{14}$$

To compare the significance between the candidates and to rank them, we calculated the AIC_c and BIC differences Δ_m where m denotes the index of the candidate. Since the AIC_c and BIC values are relative values, these differences are important. The following rough rules of thumb are available for the comparison: models with $\Delta_m > 10$ have essentially no empirical support, models with $4 < \Delta_m < 7$ have considerably less support and models with $0 < \Delta_m < 2$ have substantial support (Burnham and Anderson, 2002).

Results

During the two surveys on 14 October and 5 November 2004, 269-point water bodies were observed within the study reach (Fig. 1). Of these, 139 were the result of the first survey and 130 were from the second survey. The total fish population density observed was 404 (individuals per square metre) in the former and 356 in the latter. Since each survey was conducted independently, we pooled the results, and the fish density was normalized for each survey as shown in Fig. 3. This figure would indicate that

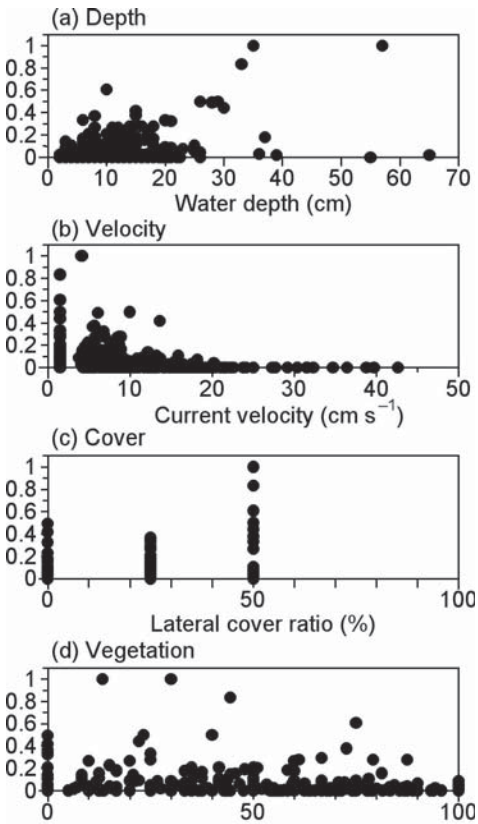

Fig. 3. Schematic diagrams showing the results of the field survey. The fish population density in each survey was normalized and the two results were pooled for the analysis.

Japanese medaka prefer deeper and slow-flowing water and areas with greater lateral cover and relatively lesser vegetation coverage; however, no area within the study reach exhibited more than 50% of lateral cover.

Using the above results, we developed FPIMs that considered every combination of the four environmental factors to minimize the MSE between the predicted and observed fish population density (Table 1). The model structure of the optimized four-factor composite FPIM is shown in Fig. 2 as an example. Figure 4 illustrates the habitat preference curves estimated by the four-factor composite (solid line) and single-factor (broken line) FPIMs. Although slight differences in the preference curves for depth and velocity were found between the two FPIMs, both curves for every factor were in fairly good agreement with each other. In addition, these curves demonstrated similar habitat preferences as mentioned earlier.

Since the habitat preference model was successfully achieved, we then predicted the spatial distribution of Japanese medaka using Eqs. 1 and 2. The results of the MSE calculation are summarized in Table 1. The significance of each factor and that of the combination of factors were analyzed by using the MSE together with AIC_c (Eq. 13) and BIC (Eq. 14) (Table 1) because the *n/k* ratio was low (*n/k* = 14.9

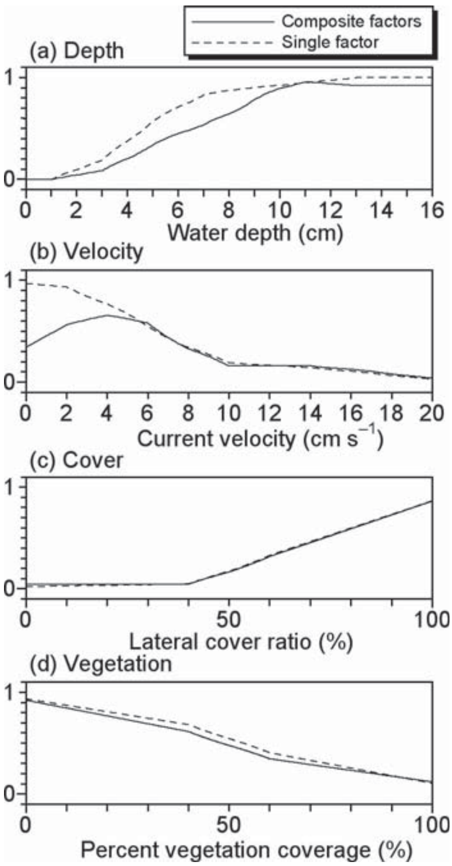

Fig. 4. Schematic diagrams showing the habitat preference estimated by the four-factor composite (solid line) and single-factor (broken line) FPIMs.

< 40). In order to compare the significance and to rank the models, the AIC_c and BIC differences were calculated as summarized in Table 2. From Table 2 and the rules of thumb, it is found that, for instance, the plausibility of model 7 is equal to that of model 14 from the viewpoint of BIC. The results in Tables 1 and 2 indicate that among the four environmental factors, cover is the most significant, followed by velocity, vegetation, and depth. In addition, the four-factor composite FPIM was found to be the best model among the candidates. The ranks based on AIC_c and BIC are different for each of the criteria, and the criteria exhibit different characteristics with regard to model selection; for example, BIC would select a model with a smaller number of parameters.

Discussion

We employed the FPIM to evaluate the habitat preferences of Japanese medaka. The applications of fuzzy rule-based models in ecological research have been increasing (Rüger *et al*., 2005; Adriaenssens *et al*., 2006; Mouton *et al*., 2006) because ecological data and expert knowledge contain uncertainty themselves (Bosserman and Ragade, 1982). The fuzzy rule-based approach has substantial features such as consideration of uncertainties, reflection of ecological and expert knowledge, etc. Knowledge-based modeling approaches, such as expert systems, in which fuzzy sets and rules are defined by experts such as biologists are often employed. Recently, some studies have introduced heuristic optimization techniques for the modification of the rules (Mouton *et al*., 2006). For comparison, the FPIM comprised

Table 2. AIC_c and BIC differences between models. The upper half matrix is the BIC difference; the lower half matrix, the AIC_c difference. See Table 1 for model index and conditions.

Model index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	–	80.3	99.3	71.6	39.1	141.8	121.3	184.0	77.5	132.4	104.3	194.1	147.6	121.3	162.5
2	90.3	–	19.0	8.8	41.2	61.5	41.0	103.7	2.8	52.0	24.0	113.8	67.2	40.9	82.2
3	109.3	19.0	–	27.7	60.2	42.5	22.0	84.7	21.8	33.1	5.0	94.8	48.3	22.0	63.2
4	91.6	1.4	17.6	–	32.5	70.2	49.8	112.4	6.0	60.8	32.7	122.6	76.0	49.7	91.0
5	55.8	34.5	53.5	35.9	–	102.7	82.2	144.9	38.4	93.3	65.2	155.0	108.5	82.2	123.4
6	161.9	71.6	52.6	70.2	106.1	–	20.5	42.2	64.3	9.4	37.5	52.3	5.8	20.5	20.7
7	151.7	61.4	42.4	60.0	95.9	10.2	–	62.6	43.8	11.0	17.0	72.8	26.2	0.1	41.2
8	214.3	124.0	105.0	122.7	158.5	52.4	62.6	–	106.4	51.6	79.7	10.2	36.4	62.7	21.4
9	104.4	14.1	4.8	12.8	48.6	57.5	47.3	109.9	–	54.8	26.8	116.6	70.0	43.7	85.0
10	159.2	69.0	50.0	67.6	103.5	2.6	7.6	55.0	54.8	–	28.1	61.8	15.2	11.1	30.2
11	141.6	51.3	32.3	49.9	85.8	20.3	10.1	72.7	37.1	17.7	–	89.8	43.2	17.0	58.2
12	234.9	144.6	125.6	143.2	179.1	73.0	83.2	20.6	130.4	75.6	93.3	–	46.6	72.9	31.6
13	184.8	94.5	75.5	93.2	129.0	22.9	33.1	29.5	80.4	25.6	43.2	50.1	–	26.3	15.0
14	169.0	78.7	59.8	77.4	113.2	7.2	17.4	45.3	64.6	9.8	27.5	65.8	15.8	–	41.3
15	210.3	120.0	101.0	118.6	154.5	48.4	58.6	4.0	105.9	51.0	68.7	24.6	25.5	41.3	–

the two artificial intelligence techniques of fuzzy reasoning and GA; this enabled us to evaluate habitat preference using field observation data (Hiramatsu *et al*., 2003; Fukuda *et al*., 2005, 2006b). The major difference between the FPIM and the models based on heuristic optimization is the parameter being optimized. In the former model, the premise part and if-then rules are predefined and then the GA optimizes the consequence part using field data; in the latter model, the rules are optimized, while the fuzzy sets, i.e. the premise and consequence parts, are unchanged. The applicability of the FPIM has been verified by Fukuda and Hiramatsu (2006), and the results indicate that the FPIM has a fairly good ability in predicting the spatial distribution of Japanese medaka.

The habitat preference curves evaluated by the four-factor composite and single-factor FPIMs were in good agreement with each other. This suggests the applicability of a univariate approach in habitat preference evaluation performed using field observation data. In a natural system, habitat selection by fish is affected by complicated interactions between multiple environmental factors that make it difficult to relate the physical environment to habitat preference. To overcome this problem, we employed a GA and estimated habitat preference with regard to the four environmental factors of depth, velocity, cover, and vegetation simultaneously (Fukuda *et al*., 2005, 2006b). For the same reason, we applied the fuzzy neural network to predict habitat selection (Fukuda, 2006a). Despite the difficulties and complexities involved, almost all the habitat preference models are generally based on the univariate approach and are widely applied in habitat evaluation (e.g. Guey *et al*., 2000; Nykänen and Huusko, 2004; Koizumi *et al*., 2005); however, few studies have verified whether the univariate approach can be applied. Therefore, further investigations are required for this verification.

The present FPIM represented the habitat preference of Japanese medaka as previously reported (Fukuda *et al*., 2006b). Japanese medaka avoided shallow and fast-flowing water and preferred greater lateral cover and less vegetation coverage. This avoidance of shallow water has been explained by Morishita and Morishita (1997): a fish requires a depth at least three times their body length to inhabit a stream. The report by Takemura *et al*. (2003) supports the observation that Japanese medaka avoid fast-flowing water: the individual swimming ability of adult medaka is limited to velocities ranging from 10 cm s^{-1} to 15 cm s^{-1}. The preference for a larger cover suggests the importance of a nearshore habitat that is now widely recognized as a necessary habitat condition for larval and juvenile stream-resident fish (Moore and Gregory, 1988; Wolter and Bischoff, 2001; Nykänen and Huusko, 2003). Since Japanese medaka generally grow up to approximately 2 cm in length, the results obtained for larval stream resident fish can be compared with those for adult medaka. The avoidance of larger vegetation coverage may be due to the feeding behavior of the medaka in the daytime. Rosenfeld and Boss (2001) reported that the position of the fish would be determined by the balance between the net gain of energy from feeding and the cost of being exposed to fast-flowing water. Since this study is based on daytime observations in one season, further studies concerning diurnal and seasonal changes in habitat preference are necessary for a deeper understanding.

Information criteria have been widely applied for the selection of the best model among available candidates; models are evaluated on the basis of the principle of parsimony by using the number of parameters within the model, the residuals, and the sample size (Burnham and Anderson, 2002). The prediction ability and the efficiency of models are generally difficult to analyse, but comparison of the differences between the models would be easy with the help of the criteria (see examples in Burnham and Anderson, 2002). Although the criteria have been originally developed for linear regression models, their application is now extended to non-linearly fitted models (Burnham and Anderson, 2002; Shimizu *et al*., 2002). Some studies have attempted to apply information criteria to non-linear systems, such as artificial neural networks, for quantifying the effective number of parameters (e.g. Murata *et al*., 1994); however, Anders and Korn (1999) have discussed this applicability and emphasized that care is needed due to the identification problems inherent in network models.

The result from the information-theoretic approach gave us three arguments. Firstly, cover was found to be the most significant factor in the current observation data from the viewpoint of model selection. This suggests that, among the single-factor models, the FPIM that considered lateral cover possessed superior explanatory abilities with regard to the habitat preference of Japanese medaka. The significance of cover also indicates the importance of lateral habitat, i.e. the water's edge (Moore and Gregory, 1988; Wolter and Bischoff, 2001; Nykänen and Huusko, 2003). Since the significance of the factor was evaluated using the balance between the prediction error and the number of parameters, the present result should be validated together with the other evaluation methods such as weighting operations. In contrast, previous studies have reported that velocity was found to be the most significant factor governing the habitat preference of Japanese medaka (Hiramatsu *et al*., 2003; Hiramatsu and Shikasho, 2004; Fukuda *et al*., 2005, 2006b). It is also necessary to consider other effects, such as seasonal changes, because the significance of environmental factors would differ according to the season, life stage, inter- and intra-specific competition, etc. Secondly, the FPIM that comprised the four factors of depth, velocity, cover, and vegetation possessed the best prediction ability among all the candidates. This result is supported by the rule of thumb, that is, every other model exhibited differences in the AIC_c and BIC that exceeded the value of ten. This may be because the four-factor composite FPIM simultaneously quantified the habitat preference for each factor and was thus able to reflect the effect of the complicated interactions between the environmental factors in its model structure. Thirdly, the AIC_c and BIC demonstrated different characteristics for model selection. BIC selected models with fewer model parameters, which corresponds to the previous study by Johnson and Omland (2004). Burnham and Anderson (2002) discussed the differences between the two criteria; one reason for the difference is the assumption in the derivation procedure. They suggested the use of AIC-type criteria in the biological sciences because it is difficult to keep the reality as sample size is increased by orders of magnitude in biological systems. This simple factor is a violation of the assumptions that form the basis of the BIC (Burnham and Anderson, 2002).

In conclusion, the information-theoretic approach appears to be very useful in the identification of the best model and the most significant factor that can explain the habitat selection of the target species. Together with ecological knowledge, this approach can provide us with additional information useful in decision making for the improved conservation and restoration of ecosystems.

Acknowledgements

The authors thank K. Abe and M. Mori at the Laboratory of Drainage and Water Environment, Kyushu University, for their intensive assistance in the field survey.

References

Adriaenssens, V., Goethals, P. L. M., De Pauw, N. (2006) Fuzzy knowledge-based models for prediction of Asellus and Gammarus in watercourses in Flanders (Belgium). Ecol. Model. 195: 3–10. DOI: 10.1016/j.ecolmodel.2005.11.043

Anders, U., Korn, O. (1999) Model selection in neural networks. Neural Networks **12**: 309–323.

Bosserman, R. W., Ragade, R. K. (1982) Ecosystem analysis using fuzzy set theory. Ecol. Model. **16**: 191–208.

Bovee, K. D., Lamb, B. L., Bartholow, J. M., Stalnaker, C. B., Taylor, J. Henriksen, J. (1998) Stream habitat analysis using the instream flow incremental methodology. U. S. Geological Survey, Biological Resources Division Information and Technology Report. USGS/BRD-1998-0004

Brosse, S., Lek, S. (2000) Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques. Freshwater Biol. **44**: 441–452.

Burnham, K. P., Anderson, D. R. (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer-Verlag, New York Berlin Heidelberg.

Fukuda, S., Hiramatsu, K. (2006) Model selection toward the quantification of fish habitat preference of Japanese medaka (*Oryzias latipes*). Proceedings of ICEM 2006 in Yamaguchi: 152–153, International Society for Ecological Modelling, August 2006, Yamaguchi, Japan.

Fukuda, S., Hiramatsu, K., Mori, M., Shikasho, S. (2005) Mathematical Characterization of fuzziness in fish habitat preference of Japanese medaka (*Oryzias latipes*) in agricultural canal. Trans. JSIDRE **239**: 43–49 (in Japanese with English abstract).

Fukuda, S., Hiramatsu, K., Mori, M. (2006a) Fuzzy neural network model for habitat prediction and HEP for habitat quality estimation focusing on Japanese medaka (*Oryzias latipes*) in agricultural canals. Paddy Water Environ. **4 (3)**: 119–124. DOI: 10.1007/s10333-006-0039-5

Fukuda, S., Hiramatsu, K., Mori, M., Shikasho, S. (2006b) Numerical quantification of the significance of aquatic vegetation affecting spatial distribution of Japanese medaka (*Oryzias latipes*) in an agricultural canal. Landscape Ecol. Eng. **2**: 65–80. DOI: 10.1007/s11355-006-0030-8

Guey, J. C., Boisclair, D., Rioux, D., Leclerc, M. (2000) Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*), Can. J. Fish. Aquat. Sci. **57**: 2065–2075.

Guey, J. C., Boisclair, D., Leclerc, M., Lapointe, M. (2003) Assessment of the transferability of biological habitat models for Atlantic salmon parr (*Salmo salar*), Can. J. Fish. Aquat. Sci. **60**: 1398–1408. DOI: 10.1139/F03-120.

Harby A., Baptist M., Dunbar M. J., Schmutz S. (eds.) (2004) COST Action 626 report: State-of-the-art in data sampling, modelling analysis and applications of river habitat modelling. COST Action 626 European Aquatic Modelling Network.

Hiramatsu, K., Fukuda, S., Shikasho, S. (2003) Mathematical modeling of preference intensity of Japanese medaka for instream water environment using fuzzy inference, Trans. JSIDRE *228*: 65–72 (in Japanese with English abstract).

Hiramatsu, K., Shikasho, S. (2004) GA-based model optimization for preference intensity of Japanese Medaka Fish (*Oryzias latipes*) to streamflow environments. Paddy Water Environ. **2**: 135–143. DOI: 10.1007/s10333-004-0052-5

Inoue, M., Nakano, S. (1999) Habitat structure along channel-unit sequences for juvenile salmon: a subunit-based analysis of in-stream landscapes. Freshwater Biol. **42**: 597–608.

Johnson, J. B., Omland, K. S. (2004) Model selection in ecology and evolution, Trends ecol. Evol. **19 (2)**: 101–108. DOI: 10.1016/j.tree.2003.10.013

Koizumi, N., Takemura, T., Okushima, S., Aiga, H., Yamamoto, S., Ebihara, S. (2005) Evaluation of suitable habitat for field Gudgeon in drainage canal based on HEP technique: A case study of Yatsu paddy field, the Shibata-Gawa river basin, Chiba prefecture, Proceedings of river engineering, JSCE **11**: 489–494 (in Japanese with English abstract).

Lechowicz, M. J. (1982) The sampling characteristics of electivity indices. Oecologia (Berl.) **52**: 22–30.

Mizumoto, M. (1995) Realization of PID controls by fuzzy control methods, Fuzzy Set Syst. **70**: 171–182.

Moore, K. M. S., Gregory, S. V. (1988) Response of Young-of-the-Year Cutthroat trout to Manipulation of Habitat Structure in a Small Stream. T. Am. Fish. Soc. **117**: 162–170.

Morishita, I., Morishita, Y. (1997) Kyosei no shizen-gaku, stream organisms in Japan: how they are affected by Japanese culture and how they express ecological health. Sankaido 38–39 (in Japanese with English description).

Mouton, A. S., Schneider, M., Goethals, P. L. M., De Pauw, N. (2006) Nature knows: heuristic search algorithms optimizing a fuzzy rule-based fish habitat suitability model for rivers. Proceedings of ICEM 2006 in Yamaguchi: 154–155, International Society for Ecological Modelling, August 2006, Yamaguchi, Japan.

Murata, N., Yoshizawa, S., Amari, S. (1994) Network Information Criterion— Determining the Number of Hidden Units for an Artificial Neural Network Model. IEEE Trans. Neural Networks **5 (6)**: 865–872.

Nykänen, M., Huusko, A. (2003) Size-related changes in habitat selection by larval grayling (*Thymallus thymallus* L.).

Ecol. Freshw. Fish **12**: 127–133.

Nykänen, M., Huusko, A. (2004) Transferability of habitat preference criteria for larval European grayling (*Thymallus thymallus*), Can. J. Fish. Aquat. Sci. **61**: 185–192. DOI: 10.1139/F03-156

Rosenfeld, J. S., Boss, S. (2001) Fitness consequences of habitat use for juvenile cutthroat trout: energetic costs and benefits in pools and riffles. Can. J. Fish. Aquat. Sci. **58 (3)**: 585–593. DOI: 10.1139/cjfas-58-3-585

Rüger, N., Schlüter, M., Matthies, M. (2005) A fuzzy habitat suitability index for *Populus euphratica* in the Northern Amudarya delta (Uzbekistan). Ecol. Model. **184**: 313–328. DOI:10.1016/j.ecolmodel.2004.10.010

Sekine, M., Imai, T., Ukita, M. (1997) A model of fish distribution in rivers according to their preference for environmental factors. Ecol. Model. **104**: 215–230.

Shimizu, Y., Tamura, T., Ono, M., Kasai, O., Nakajima, T. (2002) Application of Nonlinear Fitting and Selection of the Most Fitted Equation by AIC in Stability Test of Pharmaceutical Ingredients. Drug Dev. and Ind. Pharm. **28 (8)**: 931–937. DOI:10.1081DDC-120006425

Takemura, T., Koizumi, N., Okushima, S., Yamamoto, S., Kato, T. (2003) Experiments of Relationship between Physical Environment and Behavior of Medakafish Assuming Small-scale Channels. Tech. Rep. Natl. Inst. Rural. Eng. Japan **201**: 37–45 (in Japanese with English abstract).

Urabe, H., Nakano, S. (1999) Linking microhabitat availability and local density of rainbow trout in low-gradient Japanese streams. Ecol. Res. **14**: 341–349.

U. S. Fish and Wildlife Service (1980a) Habitat evaluation procedures (HEP): Ecological service manual 102. Washington, D. C., U. S. A.

U. S. Fish and Wildlife Service (1980b) Standards for the development of habitat suitability index models: Ecological service manual 103. Washington, D. C., U. S. A.

Vadas, R. L., Orth, D. J. (2001) Formulation of Habitat Suitability Models for Stream Fish Guilds: Do the Standard Methods Work? T. Am. Fish. Soc. **130**: 217–235.

Wolter, C., Bischoff, A. (2001) Seasonal Changes of Fish Diversity in the Main Channel of the Large Lowland River Oder. Regul. Rivers: Res. Mgmt. **17**: 595–608. DOI: 10.1002/rrr.645

Yoshioka, H. (1963) On the effects of environmental factors upon the reproduction of fishes: 2. Effects of short and long day-lengths on *Oryzias latipes* during spawning season. Bull. Fac. Fish. Hokkaido Univ. **14 (3)**: 137–151.