Modelling basin-scale distribution of fish occurrence probability for assessment of flow and habitat conditions in rivers

Pengzhe Sui¹, Akito Iwasaki¹,², Oliver C. Saavedra V.¹,³ and Chihiro Yoshimura¹

¹Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Tokyo 152–8552, Japan
²Department of Mechanical and Environmental Informatics, Tokyo Institute of Technology, Tokyo 152–8552, Japan
³Energy Resources and Environmental Engineering Program, Egypt-Japan University of Science and Technology, Alexandria 21934, Egypt

Received 30 September 2012; accepted 24 June 2013; open for discussion until 1 September 2014

Editor Z.W. Kundzewicz; Guest editor M. Acreman

Citation Sui, P., Iwasaki, A., Saavedra, V.O.C., and Yoshimura, C., 2013. Modelling basin-scale distribution of fish occurrence probability for assessment of flow and habitat conditions in rivers. Hydrological Sciences Journal, 59 (3–4), 618–628.

Abstract Flow regimes play an important role in sustaining biodiversity in river ecosystems. However, the effects of flow regimes on riverine fish have not been clearly described. Therefore, we propose a new methodology to quantitatively link habitat conditions (such as flow indices and physical habitat conditions) to the occurrence probability (OP) of fish species. We developed a basin-scale fish distribution model by integrating the concept of habitat suitability assessment with a distributed hydrological model in order to estimate the OP of fish, with particular attention to flow regime. A generalized linear model was used to evaluate the relationship between the probabilities of fish occurrence and major environmental factors in river sections. A geomorphology-based hydrological model was adopted to simulate river discharge, which was used to calculate 10 flow indices. The occurrence probabilities of 50 fish species in the Sagami River in Japan were modelled. For the prediction accuracy, field survey results that included at least five observations of both the presence and the absence of each species were required to obtain relatively reliable prediction (accuracy > 60%). Using the developed model, important habitat conditions for each species were identified, which showed the importance of low-flow events for more than 10 species, including Hypomesus nipponensis and Rhinogobius fluviatilis. The model also confirmed the positive effects of natural flow and the negative effect of river-crossing structures, such as dams and weirs, on the OP of most species. The suggested approach enables us to evaluate and project the ecological consequences of water resource management policy. The results demonstrate the applicability of the fish distribution model to provide quantitative information on the flow required to maintain fish communities.

Key words flow regime; fish occurrence probability; distributed hydrological model; physical habitat; river fragmentation; Japan

Modélisation répartition échelle du bassin de la probabilité d’occurrence de poisson pour l’évaluation des flux et des conditions d’habitat dans les rivières

Résumé Régimes d’écoulement jouent un rôle important dans le maintien de la biodiversité dans les écosystèmes fluviaux. Cependant, les effets des régimes d’écoulement sur les poissons riverains n’ont pas été clairement décrit. Par conséquent, nous proposons une nouvelle méthode pour relier quantitativement les conditions de l’habitat (comme les indices d’écoulement et les conditions physiques de l’habitat) à la probabilité d’occurrence d’espèces de poissons. Nous avons développé un modèle de distribution des poissons échelle du bassin, en intégrant la notion d’évaluation de la convenance de l’habitat avec un modèle hydrologique distribué afin d’estimer la probabilité d’occurrence de poisson avec une attention particulière au régime d’écoulement. Un modèle linéaire généralisé a été utilisé pour évaluer la relation entre les probabilités d’occurrence des poissons et des principaux facteurs environnementaux dans les sections de la rivière. Un modèle hydrologique basé sur la géomorphologie a été adopté pour simuler le débit fluvial, qui a été utilisé pour calculer les 10 indices de fluidité. Les probabilités d’occurrence de 50 espèces de poissons dans la rivière de Sagami au Japon ont été modélisées. Pour la précision de la prédiction, les résultats de l’enquête de terrain qui comprenait au moins cinq
observations à la fois la présence et l’absence de chaque espèce ont été nécessaires pour obtenir la prévision relativement fiable (precision > 60%). En utilisant le modèle mis au point, des conditions importantes de l’habitat de chaque espèce ont été identifiées, ce qui montre l’importance des événements à faible débit pour plus de 10 espèces, y compris nipponensis Hypomesus et fluviatilis Rhinogobius. Le modèle a également confirmé les effets positifs de l’écoulement naturel et l’effet négatif des structures fluviales qui traversent, comme les barrages et déversoirs, sur la probabilité d’occurrence de la plupart des espèces. L’approche proposée permet d’évaluer et projeter les conséquences écologiques de la politique de gestion des ressources en eau. Les résultats démontrent l’applicabilité du modèle de distribution de poisson pour fournir des informations quantitatives sur le débit nécessaire pour maintenir les populations de poissons.

Mots clés régime d’écoulement ; probabilité d’occurrence du poisson ; modèle hydrologique distribué ; habitat physique ; fragmentation rivière ; Japon

1 INTRODUCTION

In recent years, there has been increasing concern about the global decline of biodiversity. In particular, species richness has declined faster in freshwater ecosystem than in terrestrial or marine systems over the past 30 years (Jenkins 2003). Such a long-term trend has been caused by multiple anthropogenic impacts, such as flow regime change, habitat fragmentation, channel alteration, water pollution and the presence of invasive species (Dudgeon et al. 2006), which have endangered 65% of the world’s river habitats and put thousands of aquatic wildlife species at risk (Vorosmarty et al. 2010).

An important approach to biodiversity conservation is to model the fish occurrence probability (OP) at the basin scale and link this to river habitat parameters, especially for the conservation of fish species that are rarely observed and may be endangered.

Flow regimes play an important role in sustaining the biodiversity in river ecosystems (Acreman and Dunbar 2004). Many flow indices of ecological importance have been proposed to describe flow regimes, on the basis of five important aspects, namely magnitude, frequency, duration, timing and rate of change (Richter et al. 1996, Olden and Poff 2003). All these flow indices have ecological roles for sustaining freshwater ecosystems (Richter et al. 1996, Poff et al. 1997, Acreman and Dunbar 2004). Xenopoulos et al. (2005) reported the linear relationship between the mean annual discharge of rivers and fish species richness. The ecological importance of high- and low-flow events for freshwater fish has also been investigated and emphasized (Iwasaki et al. 2012). Even for a single species, however, the effects of flow regimes on freshwater fish have not been clearly described, especially with regard to their population dynamics, because flow indices of ecological importance vary not only in spatial and temporal dimensions, but also among and within species. Furthermore, river discharge data are required for calculating flow indices. However, observed discharge data are usually available at only a few stations in a basin and thus cannot meet the requirements for the assessment of flow regimes as well as their ecological integrity along a river continuum. In this sense, the simulation of discharge in a river network by a distributed hydrological model (DHM) is an important and effective approach to provide essential data for the spatial description of flow regimes and their corresponding ecosystem assessment (Lester et al. 2011).

Flow regimes are also affected by river-crossing structures, such as dams and weirs. To date, flow regulation and channel fragmentation caused by dam construction have altered the flow regimes of over half of the large river systems in the world (Nilsson et al. 2005), resulting in habitat degradation and decline in aquatic species (Poff et al. 1997). The effects of dams on freshwater fish distribution have also been reported, particularly for diadromous fishes (Fukushima et al. 2007, Nislow et al. 2011). Besides flow regimes, water quality and physical habitat conditions, such as catchment area, river width, slope, altitude and riverbed material, are also key factors that affect the distribution and population of fishes.

In this study, a new methodology was proposed to quantitatively link habitat conditions (including flow indices, water quality and other physical habitat conditions) to the OP of fish species at the basin scale in a distributed manner. In order to determine the flow in the whole river basin, a geomorphology-based hydrological model (GBHM) (Yang et al. 2002) was used to simulate the discharge in Sagami River, Japan, and then flow indices were calculated. A basin-scale fish distribution model was developed using data collected on different habitat conditions and fish communities to estimate the OP of fish, with particular attention to flow regimes. Limitations to its application were discussed according to an analysis of key factors that affect the prediction accuracy of fish OP. By the developed model, environmental variables of habitat with ecological importance for each fish species were identified and discussed. Furthermore, as an example of its application, we also assessed the effects of flow regimes and river fragmentation by dams on the OP of fish species.
2 METHODS

2.1 Study area and hydrological simulation

The Sagami River basin is located in Kanagawa and Yamanashi Prefectures in Honshu Island, Japan (Fig. 1(a)). The basin area is about 1680 km², and its river length is 113 km. Sagami River originates from Fujii Mountain and flows into Sagami Bay of the Pacific Ocean. Both the main stream, Sagami River, and its biggest tributary, Nakatsu River, are dammed at several locations along the waterway (Fig. 1(b)). The target area is a part of the Sagami River basin, downstream from Shiroyama Dam and Miyagase Dam, which was divided into 10 sections (S1–S3, N1–N4, SN1–SN3) on the basis of the estuary, confluences of tributaries and river-crossing structures (dams, weirs), as shown in Fig. 1(b).

A GBHM was set up for the Sagami River basin to support the estimation of flow indices in the target river reach. Hourly discharge from 1998 to 2005 was simulated using precipitation, temperature, soil type, land-use etc., as input data (see Supplementary Material, Table S1). Furthermore, the operation of the Shiroyama and Miyagase Dam was included in the river routing module of the GBHM. The shuffled complex evolution method developed at the University of Arizona (SCE-UA) (Duan et al. 1992, 1994, Duan 2003) was used to calibrate model parameters using observed daily discharge. Five soil water parameters used in the GBHM were calibrated by the SCE-UA algorithm, including the saturated hydraulic conductivity of surface soil (Ksat, mm h⁻¹), soil anisotropy ratio (ani, −), maximum surface water detention (Sstmax, mm), surface Manning’s roughness coefficient (surf, −) and hydraulic conductivity of groundwater (kg, mm h⁻¹). The objective function was set up to minimize the difference between the observed and simulated discharge at the gauge station of Sagami–Oohashi (Fig. 1(b)) in 2001. After obtaining suitable parameter values by calibration, the GBHM was validated from 1998 to 2000.

2.2 Variables to represent habitat properties

A total of 30 explanatory variables were employed in this study, details and definitions of which are given in Table 1. The variables belong to three groups: flow indices (10), physical habitat properties (13) and water quality (7). Most of the physical habitat properties, such as CA, S, MS, DFS, AL, RW, SL and SI (see Table 1), were determined using geographic data in Google Earth and ArcGIS, while other habitat properties (IP, NUD, NLD, D50, UC) and water quality information were obtained from field survey results (Kanagawa Prefectural Government). Daily discharge from 1998 to 2005 obtained by hydrological simulation was averaged in each section, and then used to calculate flow indices, as described by Olden and Poff (2003). It should be noted that the number of high-flow peaks (NF) in each year was counted based on high flows four times larger than the grand mean discharge (the average discharge in all 8 years from 1998 to 2005). The ranges of habitat variables used in this study are shown in Table S2.
Table 1 Explanatory variables of habitat property for the fish distribution model.

| Category                  | Code | Definition                                      |
|---------------------------|------|-------------------------------------------------|
| Flow indices              | AMD  | Annual mean discharge in one year               |
|                           | MID  | Minimum discharge in one year                   |
|                           | RMM  | Ratio of maximum and minimum discharge in one year |
|                           | DV   | Variance of discharge in one year               |
|                           | SD   | Mean discharge during spring, from 1 March to 31 March |
|                           | MLF  | Magnitude of low flow (25th percentile discharge) |
|                           | MHF  | Magnitude of high flow (75th percentile discharge) |
|                           | DLF  | Mean duration of low flow (less than MLF)       |
|                           | DHF  | Mean duration of high flow (more than MHF)      |
|                           | NF   | Number of high flow peaks in one year           |
| Physical habitat property | CA   | Catchment area in each section                  |
|                           | S    | Mean longitudinal slope observed per 500 m from estuary |
|                           | MS   | Maximum longitudinal slope observed per 500 m    |
|                           | DFS  | Distance from estuary to midpoint of river section |
|                           | AL   | Average of altitude observed per 500 m from estuary |
|                           | RW   | Average of river width observed per 500 m from estuary |
|                           | SL   | Distance between midpoints of two contiguous river sections |
|                           | SI   | Sinuosity index, ratio of straight length to real length of section |
|                           | NUD  | Number of river-crossing structures in upstream |
|                           | NLD  | Number of river-crossing structures in downstream |
|                           | IP   | Isolation period, difference between observation year of fish survey and construction year of river-crossing structure |
|                           | D50  | Mean diameter of river bed material             |
|                           | UC   | Uniformity coefficient of river bed material    |
| Water quality             | COD  | Average chemical oxygen demand in each section  |
|                           | BOD  | Average biochemical oxygen demand in each section |
|                           | DO   | Average dissolved oxygen in each section        |
|                           | TN   | Average total nitrogen in each section          |
|                           | TP   | Average total phosphorus in each section        |
|                           | WT   | Water temperature                               |
|                           | PH   | Water pH                                        |

2.3 Fish data

For the OP of fish distribution, presence and absence (P/A) data were collected on fishes belonging to 24 families and 54 species (refer to Table S3 in the Supplementary Material, for a full list) in each river section. The “presence” of a fish in a section was defined as it being observed once or more in a year, while “absence” meant that it was not observed at all. Fish surveys were conducted by Kanagawa Prefectural Fisheries Technology Center in the autumn (from September to December) of each year. Survey sites were distributed all over the river with a total number that varied from 68 to 75 in the target river reach. Sampling was mainly done with fishing nets, supplemented by an electric shocker when fishing nets could not be used (Kanagawa Prefectural Fisheries Technology Center 1997–2006). Fish P/A data for 1998–2005 were used in this study.

2.4 Model structure and statistical analysis

A generalized linear model (GLM) was applied to describe the relationship between the spatial distribution of the OP of each fish species and habitat variables, as:

\[
\text{OP} = \frac{1}{1 + \exp[-(\alpha + \sum \beta_i X_i)]}
\]

where \(\alpha\) is an intercept, \(\beta_i\) is a regression coefficient and \(X_i\) is a habitat variable. Though other methods (e.g. generalized additive model, artificial neural network) may more adequately describe non-linear responses of many fish species to changes in environmental factors (Guegan et al. 1998, Ahmadi-Nedushan et al. 2006), a GLM was preferred due to its simplicity and the nature of the available data.

In order to avoid multivariate collinearity, principal components analysis (PCA) was first conducted on all environmental variables after their normalization using mean values and standard deviations Fig. S1 shows the PCA biplot of PC1 vs PC2. As shown in Table 2, PC1 was mainly correlated with physical habitat properties and water quality, while PC2 was correlated with flow indices; PC1 and PC2 contributed more than 60% to the total variance. The first 10 principal components (PC1–PC10) contributed more than 99% and were employed for the following principal components regression (PCR). After regression analysis with 10 PCs, the best model expressed by PCs was selected on the basis of the Akaike information criterion (AIC). The contributions of original variables to the OP of each fish species were obtained by (Legendre and Legendre 2012):
where \( \mathbf{b} \) is a vector of the contributions of the original variables to the regression equation, i.e. \( \beta_i \) in equation (1); \( \mathbf{c} \) is a vector of the regression coefficients obtained by PCR without including the intercept; and matrix \( \mathbf{U} \) is the eigenvectors of the matrix of the original variables (loadings of the PC, Table 2). An original environmental variable was considered important for the fish species if the absolute value of its regression coefficient (|\( \beta_i \)|) was more than half of the maximum absolute value in vector \( \mathbf{b} \). All statistical analyses were carried out using the free software R 2.14 (R Development Core Team 2011).

As an example of an application of the developed fish distribution model, a case study was carried out to assess the negative effects of flow regulation and river fragmentation by dams and weirs on the OP of fish species under three management scenarios. In the first scenario (D1), all dams and weirs matched the current situation in the Sagami River basin. In the second scenario (D2), the released flow from two dams (Miyagase and Shiroyama, located upstream of target area, Fig. 1(b)) was assumed to be exactly the same as their inflow. This means that the natural flow was maintained but there was fragmentation by weirs. In the third scenario (D3), it was supposed that there were no dams or weirs in the target area.

### 3 RESULTS AND DISCUSSION

#### 3.1 Hydrological simulation

The GBHM was validated by a comparison of the simulated river discharge against the observed discharge from 1998 to 2000 at Sagami–Oohashi Station, which is close to the estuary, as shown in Fig. 1(b). Both high and low flows could be

### Table 2 Results of principal components analysis (PCA).

| Contribution (%) | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Cumulative       | 40.4| 22.4| 12.9| 7.6 | 4.6 | 3.8 | 2.9 | 2.3 | 1.4 | 0.6  |

| Loadings of principal components | CA | MS | DFS | AL | RW | SL  | SI | NUD | SLD | D50 | UC  | PH  | COD | BOD | DO  | TN  | TP  | WT  |
|----------------------------------|----|----|-----|----|----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CA | -0.330 | 0.206 | -0.387 | 0.364 | 0.444 | -0.306 |
| MS | -0.264 | 0.239 | -0.205 | 0.364 | 0.444 | -0.306 |
| DFS | -0.248 | 0.239 | -0.205 | 0.364 | 0.444 | -0.306 |
| AL | -0.225 | 0.239 | -0.205 | 0.364 | 0.444 | -0.306 |
| RW | 0.237 | -0.417 | -0.212 | -0.228 | -0.273 | -0.217 |
| SL  | 0.232 | -0.204 | 0.256 | 0.268 | 0.217 | -0.205 |
| SI  | 0.232 | -0.204 | 0.256 | 0.268 | 0.217 | -0.205 |
| NUD | 0.232 | -0.204 | 0.256 | 0.268 | 0.217 | -0.205 |
| SLD | 0.232 | -0.204 | 0.256 | 0.268 | 0.217 | -0.205 |
| D50 | 0.232 | -0.204 | 0.256 | 0.268 | 0.217 | -0.205 |
| UC  | 0.266 | -0.426 | 0.250 | 0.417 | 0.384 | -0.241 |
| PH  | 0.257 | -0.448 | 0.250 | 0.417 | 0.384 | -0.241 |
| COD | 0.257 | -0.448 | 0.250 | 0.417 | 0.384 | -0.241 |
| BOD | 0.257 | -0.448 | 0.250 | 0.417 | 0.384 | -0.241 |
| DO  | 0.255 | -0.238 | 0.205 | 0.205 | 0.205 | 0.205 |
| TN  | 0.255 | -0.238 | 0.205 | 0.205 | 0.205 | 0.205 |
| TP  | 0.255 | -0.238 | 0.205 | 0.205 | 0.205 | 0.205 |
| WT  | 0.255 | -0.238 | 0.205 | 0.205 | 0.205 | 0.205 |
| AMB | 0.266 | -0.354 | 0.233 | -0.203 | -0.579 | 0.447 |
| MID | 0.266 | -0.354 | 0.233 | -0.203 | -0.579 | 0.447 |
| RMM | 0.219 | -0.488 | -0.250 | 0.288 | -0.207 | 0.288 |
| DV  | 0.219 | -0.488 | -0.250 | 0.288 | -0.207 | 0.288 |
| SD  | 0.311 | 0.271 | -0.203 | -0.203 | -0.203 | -0.203 |
| MFL | 0.311 | 0.271 | -0.203 | -0.203 | -0.203 | -0.203 |
| MFL | 0.238 | -0.235 | 0.202 | 0.202 | 0.202 | 0.202 |
| DLF | 0.238 | -0.235 | 0.202 | 0.202 | 0.202 | 0.202 |
| DLF | 0.238 | -0.235 | 0.202 | 0.202 | 0.202 | 0.202 |
| NF  | -0.212 | -0.379 | 0.264 | 0.246 | 0.264 | 0.246 |

Notes: See Table 1 for the definition and code of each variable. Loadings of principal components were omitted when its absolute value was smaller than 0.2.
simulated very well at this station with a root mean square error (RMSE) of 38.44 and a Nash-Sutcliffe efficiency (NSE) of 0.908 (see Supplementary Material, Fig. S2a). However, the simulation results from 2003 to 2005 at Isobe Station (upstream of Sagami–Oohashi station and close to Isobe Weir, as shown in Fig. 1(b)) were not as good as the downstream results (with RMSE of 223.5 and NSE of -12.24, Fig. S2b). The simulation period was different in the two figures, only because of the availability of observed discharge data.

3.2 Fish OP model

After regression analysis by GLM and selection by AIC, the best models for the relationship between fish OP and 10 PCs were obtained for 50 species (see Supplementary Material, Table S4). For the other four species, AIC selection results showed that the best model consisted of only the intercept (constant). Then the contribution of the original habitat variables to the OP of each fish species was calculated with equation (2). The environmental factors of the habitat with a high contribution varied among the fish species (Table S5). Taking the Amur goby CB (Rhinogobius sp. CB, species no. 42 in Table S3) as an example, as shown in Table 3(a), the regression coefficient of original variables showed that the OP of the Amur goby CB was negatively correlated with DHF (the biggest negative contribution, counting it as 100%) and RMM (98.7%), while it was positively correlated with water temperature (WT, counting its contribution as 100%), BOD (92.7%), pH (78.5%), TP (78.5%), MLF (75.5%), TN (74.9%), and SD (50.4%). The results demonstrated the importance of flow indices and water quality as well as the negative effects of segmentation. Two other typical species were the Japanese eel (Anguilla japonica, species no. 1 in Table S3) and common carp (Cyprinus carpio, species no. 20 in Table S3). The Japanese eel has a high risk for extinction and is marked as endangered (EN) in the Red Data Book published by the Japanese Ministry of the Environment (MOE), while the common carp is marked as VU vulnerable (VU) in the Red List of Threatened Species published by the International Union for Conservation of Nature (IUCN) (Freyhof and Kottelat 2008). As shown in Table 3(b), the maximum negative impact on the Japanese eel came from variance in the discharge within a single year (DV, counting as 100%), followed by MS (92.7%), NF (85.0%), NLD (72.5%), DO (63.4%), SI (61.5%), MLF (61.2%), and AMD (56.9%); while it was most positively correlated with DLF (100%) and IP (94.9%). The Japanese eel can be found in isolated lakes and migrates mainly by water movement (Kimura 2003), which could be a possible reason for its positive correlation to IP. As a catadromous species, NLD, MS and SI are related to its migration. However, it is interesting that the Japanese eel prefers a long duration of low flow, small mean discharge (MLF, AMD), and little variation in flow (DV, NF). It is well known that adult Japanese eels migrate thousands of kilometres from rivers in East Asia to their spawning area without feeding. However, they live in freshwater and estuaries for a number of years before their migration. The several years of their life-span in freshwater rivers may be more important for its conservation based on the results obtained in this study. A similar evaluation for common carp (Table 3(c)) showed that the most important factor was river width, since it favours large bodies of slow or standing water and soft, vegetative sediments.

Considering the conservation of all fish species in Sagami River, the important environmental factors for all species were identified. An original variable was counted as an important factor if its contribution to the fish OP was more than 50% of the contribution of the most important factor for each species (vector b in equation (2), and as shown in Table 3). Figure 2 summarizes the number of fish species affected by each habitat variable. The results confirm the negative isolating effect of river-crossing structures (dams and weirs) on the OP of fish species. However, the isolation period was found to be more important than the number of downstream or upstream dams since IP was negatively correlated with 14 fish species, while NLD and NUD were negatively correlated with two and one fish species, respectively. As the most important negative factor, the longer DLF corresponded to a lower OP for 22 fish species. Similarly, the MID was positively correlated with 10 species, including Hypomesus nipponensis and Rhinogobius fluviatilis. These two results confirmed the importance of low-flow events. Although DLF affected fish species negatively more than the duration of high flow (DFH) in the Sagami River basin, DFH was positively correlated with more fish species than DLF (with the highest number of correlated species, 32 out of the 50 fish species). It is interesting that so many fish species prefer a long DFH. DFH will increase with dam operation. Thus, dam operation may not always be harmful for fish species, and this needs to be further investigated. The distribution of freshwater fish was found to be determined not only by physical conditions of the habitat, but also by flow regimes.
Table 3 Contribution of environmental factors to occurrence probability of fish.

| No. | Positive Factor | Contribution | Percentage | No. | Negative Factor | Contribution | Percentage |
|-----|-----------------|--------------|------------|-----|-----------------|--------------|------------|
|     |                 |              |            |     |                 |              |            |
| (a) Rhinogobius sp. CB | | | | (b) Anguilla japonica | | | |
| 1   | WT              | 0.0550       | 100        | 1   | DHF             | -0.1213      | -100       |
| 2   | BOD             | 0.0509       | 92.7       | 2   | RMM             | -0.1197      | -98.7      |
| 3   | PH              | 0.0431       | 78.5       | 3   | AL              | -0.0586      | -48.3      |
| 4   | TP              | 0.0431       | 78.5       | 4   | SL              | -0.0520      | -42.8      |
| 5   | MLF             | 0.0415       | 75.5       | 5   | RW              | -0.0461      | -38.0      |
| 6   | TN              | 0.0411       | 74.9       | 6   | DLF             | -0.0401      | -33.1      |
| 7   | SD              | 0.0277       | 50.4       | 7   | D50             | -0.0382      | -31.5      |
| 8   | UC              | 0.0269       | 48.9       | 8   | DFS             | -0.0361      | -29.8      |
| 9   | MID             | 0.0233       | 42.5       | 9   | NLD             | -0.0293      | -24.2      |
| 10  | NUD             | 0.0203       | 36.9       | 10  | IP              | -0.0231      | -19.1      |
| 11  | S               | 0.0197       | 35.8       | 11  | CA              | -0.0200      | -16.5      |
| 12  | AMD             | 0.0188       | 34.1       | 12  | SI              | -0.0125      | -10.3      |
| 13  | COD             | 0.0145       | 26.3       | 13  | NF              | -0.0110      | -9.0       |
| 14  | MS              | 0.0128       | 23.3       | 14  | MHF             | -0.0099      | -8.2       |
| 15  | DV              | 0.0070       | 12.8       | 15  | DO              | -0.0025      | -2.1       |
|     |                 |              |            |     |                 |              |            |
| (c) Cyprinus carpio | | | | | | |
| 1   | RW              | 0.1340       | 100        | 1   | IP              | -0.2232      | -100       |
| 2   | NUD             | 0.1188       | 88.6       | 2   | S               | -0.1225      | -54.9      |
| 3   | CA              | 0.0767       | 57.2       | 3   | MHF             | -0.0838      | -37.5      |
| 4   | DLF             | 0.0710       | 52.9       | 4   | NF              | -0.0800      | -35.8      |
| 5   | MS              | 0.0267       | 19.9       | 5   | DFS             | -0.0404      | -18.1      |
| 6   | COD             | 0.0123       | 9.2        | 6   | TN              | -0.0373      | -16.7      |
| 7   | DHF             | 0.0071       | 5.3        | 7   | DV              | -0.0338      | -15.1      |
| 8   | SL              | 0.0068       | 5.1        | 8   | PH              | -0.0323      | -14.5      |
| 9   | RMM             | 0.0061       | 4.5        | 9   | AMD             | -0.0299      | -13.4      |
| 10  | AL              | 0.0045       | 3.3        | 10  | BOD             | -0.0274      | -12.3      |
| 11  | MID             | 0.0018       | 1.3        | 11  | WT              | -0.0243      | -10.9      |
| 12  | D50             | -0.0213      | -9.6       | 12  | D50             | -0.0213      | -9.6       |
| 13  | DO              | -0.0171      | -7.7       | 13  | DO              | -0.0171      | -7.7       |
| 14  | SI              | -0.0117      | -5.2       | 14  | SI              | -0.0117      | -5.2       |
| 15  | TP              | -0.0102      | -4.6       | 15  | TP              | -0.0102      | -4.6       |
| 16  | NLD             | -0.0048      | -2.2       | 16  | NLD             | -0.0048      | -2.2       |
| 17  | SD              | -0.0034      | -1.5       | 17  | SD              | -0.0034      | -1.5       |
| 18  | MLF             | -0.0017      | -0.7       | 18  | MLF             | -0.0017      | -0.7       |
| 19  | UC              | -0.0008      | -0.4       | 19  | UC              | -0.0008      | -0.4       |
3.3 Key factors affecting the prediction accuracy of the fish distribution model

The developed fish distribution model was validated by the observed fish occurrence data from 1998 to 2005. If the predicted OP was more than 50%, this species was assumed to be present in the corresponding river section, otherwise it was absent. Predicted presence/absence data were compared to the observations in each section from 1998 to 2005, and the prediction accuracy of the model was calculated (see Supplementary Material, Fig. S3). Eighteen fish species (accounting for 36% of all species) were predicted with an accuracy of more than 80%, including nine species with an accuracy of 100%. The worst prediction came for *Nipponocypris temminckii* and *Odontesthes bonariensis*, with an accuracy of 44.4%, which appeared only once in the 8 year observation in the target area of the Sagami River basin. The prediction accuracy was found to be highly correlated with the number of occurrences of each species, as plotted in Fig. 3. All frequently observed species (12 species observed more than nine times in the 8 year observation period) could be predicted with an accuracy of more than 72%. The prediction accuracy was less than 60% for all species observed only once. For rarely observed species (less than five times in the 8 year survey for 10 sections), although nine species (observed two, three, or five times) could be predicted accurately (100%), it was found that the intercepts in the GLMs, i.e. \( \alpha \) in equation (1), for these nine species ranged from \(-280\) to \(-75\), while the other intercepts ranged from \(-3.53\) to 0.944 (see Supplementary Material, Table S4). Negative and large absolute values contributed more to the predicted results than variation in the explanatory variables. From a statistical perspective, similarly, if fish species were observed too often (absent less than five times), the fish distribution model would not show reliable prediction accuracy. It was necessary that there be more than five observations of both the presence and the absence of a species in the 8 year survey for 10 sections in order to obtain relatively reliable prediction results (accuracy > 60%) in this study. It was reported that error rates were higher for less-frequent observations when predicting the occurrence of salmonid fish (Dunham *et al.* 2002). Furthermore, because of problems with detectability (species present but not observed) during the field survey, sampling variability was a significant portion of overall species occurrence for the species with low abundance levels (Bayley and Peterson 2001, Stauffer *et al.* 2002). Further studies are required to improve prediction accuracy of species occurrence in both statistical and biological aspects (Dunham *et al.* 2002). In the case of fewer occurrences of species, the careful examination of observation data set and incorporating the detectability into the species occurrence model are recommended for the application of the fish distribution model.

3.4 The effects of flow regulation and fragmentation on fish OP

In order to confirm the effects of flow regulation and fragmentation on fish occurrence, three scenarios, including D1 (with dam and weir), D2 (only weir,
no dam) and D3 (no dam, no weir), were employed in this study. River discharge under the three scenarios was simulated by GBHM, and then the corresponding OP of 50 fish species was calculated with the developed fish model. Fish species exhibited different ecological habits. Species that usually lived downstream or in estuaries would not appear in the upstream section. In order to avoid complications from upstream/downstream comparisons, section SN1 (refer to Fig. 1(b) for its location) was selected as a target section and fish occurrence in SN1 only was compared under different scenarios (Fig. 4). In scenario D2, 30 species had a higher OP while 20 species had a lower probability compared to D1. This demonstrates that more fish species favour “natural flow conditions” than regulated flow by dams. It is consistent with the hypothesis that natural flow or environmental flow is central to sustaining biodiversity and preserving aquatic ecosystems (Poff et al. 1997, Richter et al. 1997, Tharme 2003). Furthermore, in scenario D3, where no dams or weirs existed, i.e. there was no river fragmentation or flow regulation, the OP of 28 fish species would increase further compared to the increase in scenario D2, while only two fish species that had higher occurrence probabilities in D2 than D1 would decrease. As mentioned above, river fragmentation limits the migration of diadromous fish and decreases their OP (Fukushima et al. 2007, Han et al. 2008). Furthermore, the indirect effects of dams, e.g. changes of predator–prey interaction and nutrient distribution and concentration, are also important for fish distribution (Meffe 1984, Greathouse et al. 2006). The biological factors, such as predator, are not considered in the presented model and thus the careful application of the model and further developments are required.

4 CONCLUSIONS

A basin-scale fish distribution model was developed in this study by coupling with a DHM. In the Sagami River basin, the OP of 50 fish species was linked to habitat conditions, including flow indices of ecological importance, water quality and physical habitat conditions. The DHM was used to provide essential data for the assessment of flow indices along a river continuum. For the prediction accuracy of fish distribution model, field survey results with more than five observations of both the presence and the absence for each species were found to be required to obtain relatively reliable prediction results (accuracy > 60%). Using the developed model, important habitat conditions for each species were identified, and the results showed the importance of low-flow events for more than 10 species, such as Hypomesus...
nipponensis and Rhinogobius fluviatilis. As a demonstration of model application, the positive effects of natural flow conditions and the negative effects of river fragmentation by dams and weirs were confirmed. Overall, the developed model enables us to evaluate and project the ecological consequences of water resource management policy, including flood management and water supply, for different spatial sections through dam operation and water withdrawal. Therefore, this study demonstrated the applicability of the fish distribution model to provide quantitative information on flow conditions required to maintain fish communities.

Acknowledgements We thank Naoyuki Suguro and Yasushi Aizawa at Kanagawa Prefectural Fisheries Technology Center for providing fish data.

Funding This study was conducted as a part of the Core Research for Evolutional Science and Technology (CREST) project of the Japan Science and Technology Agency (JST) and was partly supported by a Grant-in-Aid for Scientific Research (KAKENHI).

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be accessed at: http://dx.doi.org/10.1080/02626667.2013.827791

REFERENCES

Acreman, M. and Dunbar, M.J., 2004. Defining environmental river flow requirements—a review. Hydrology and Earth System Sciences, 8 (5), 861–876.
Ahmadi-Nedushan, B., et al., 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications, 22 (5), 503–523.
Bayley, P.B. and Peterson, J.T., 2001. An approach to estimate probability of presence and richness of fish species. Transactions of the American Fisheries Society, 130 (4), 620–633.
Duan, Q., 2003. Global optimization for watershed model calibration. In: Q. Duan, et al., eds. Calibration of watershed models. Washington, DC: AGU, 89–104.
Duan, Q.Y., Sorooshian, S., and Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research, 28 (4), 1015–1031.
Duan, Q.Y., Sorooshian, S., and Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology, 158 (3–4), 265–284.
Dudgeon, D., et al., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews, 81 (2), 163–182.
Dunham, J.B., Rieman, B.E., and Peterson, J.T., 2002. Patch-based models to predict species occurrence: lessons from salmonid fishes in streams. In: J.M. Scott, et al., eds. Predicting species occurrences: issues of accuracy and scale. Covello, CA: Island Press, 327–334.
Freyhof, J. and Kottelat, M., 2008. Cyprinus carpio. In: IUCN 2013. IUCN Red List of Threatened Species [online]. Version 2013.2. Available from: www.iucnredlist.org [Accessed 7 Jan 2014].
Fukushima, M., et al., 2007. Modelling the effects of dams on freshwater fish distributions in Hokkaido, Japan. Freshwater Biology, 52 (8), 1511–1524.
Greathouse, E.A., et al., 2006. Indirect upstream effects of dams: consequences of migratory consumer extirpation in Puerto Rico. Ecological Applications, 16 (1), 339–352.
Guegan, J.F., Lek, S., and Oberdorff, T., 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. Nature, 391 (6665), 382–384.
Han, M., et al., 2008. How do dams affect freshwater fish distributions in Japan? Statistical analysis of native and nonnative species with various life histories. Ecological Research, 23 (4), 735–743.
Iwaski, Y., et al., 2012. Evaluating the relationship between basin-scale fish species richness and ecologically relevant flow characteristics in rivers worldwide. Freshwater Biology, 57 (10), 2173–2180.
Jenkins, M., 2003. Prospects for biodiversity. Science, 302 (5648), 1175–1177.
Kanagawa Prefectural Fisheries Technology Center, 1997–2006. Fish survey report in Sagami River. Kawasaki: Freshwater Fisheries Experiment Station in Kanagawa Prefectural Fisheries Technology Center (In Japanese).
Kanagawa Prefectural Government, Water quality of public water body and groundwater in Kanagawa Prefectural [online]. Kanagawa Prefectural Government. Available from: http://www.pref.kanagawa.jp/cnt/f41010/ [Accessed 21 September 2012] (In Japanese).
Kimura, Y., 2003. Freshwater Fishes in Japan. Tokyo: Gakken Education (In Japanese).
Legendre, P. and Legendre, L., 2012. Numerical ecology. 3rd ed. Oxford: Elsvier Science & Technology Books.
Lester, R.E., et al., 2011. Linking water-resource models to ecosystem-response models to guide water-resource planning —an example from the Murray—Darling Basin, Australia. Marine and Freshwater Research, 62 (3), 279–289.
Meffe, G.K., 1984. Effects of abiotic disturbance on coexistence of predator-prey fish species. Ecology, 65 (5), 1525–1534.
Nilsson, C., et al., 2005. Fragmentation and flow regulation of the world’s large river systems. Science, 308 (5720), 405–408.
Nislow, K.H., et al., 2011. Variation in local abundance and species richness of stream fishes in relation to dispersal barriers: implications for management and conservation. Freshwater Biology, 56 (10), 2135–2144.
Olden, J.D. and Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19 (2), 101–121.
Poff, N.L., et al., 1997. The natural flow regime. Bioscience, 47 (11), 769–784.
R Development Core Team, 2011. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
Richter, B.D., et al., 1996. A method for assessing hydrologic alteration within ecosystems. Conservation Biology, 10 (4), 1163–1174.
Richter, B.D., et al., 1997. How much water does a river need? Freshwater Biology, 37 (1), 231–249.
Stauffer, H.B., Ralph, C.J., and Miller, S.L., 2002. Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. In: J.M. Scott, et al., eds. Predicting species occurrences: issues of accuracy and scale. Covello, CA: Island Press, 357–365.
Tharme, R.E., 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. *River Research and Applications*, 19 (5–6), 397–441.

Vorosmarty, C.J., *et al*., 2010. Global threats to human water security and river biodiversity. *Nature*, 467 (7315), 555–561.

Xenopoulos, M.A., *et al*., 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology*, 11 (10), 1557–1564.

Yang, D., Herath, S., and Musiake, K., 2002. A hillslope-based hydrological model using catchment area and width functions. *Hydrological Sciences Journal*, 47 (1), 49–65.