Transferability of Monitoring Data from Neighboring Streams in a Physical Habitat Simulation

Byungwoong Choi, Sung-Uk Choi * and Hojeong Kang

School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea; E-Mails: bw628@yonsei.ac.kr (B.C.); hj_kang@yonsei.ac.kr (H.K.)

* Author to whom correspondence should be addressed; E-Mail: schoi@yonsei.ac.kr; Tel.: +82-2-2123-2797; Fax: +82-2-364-5300.

Academic Editor: Young-Seuk Park

Received: 28 May 2015 / Accepted: 18 August 2015 / Published: 20 August 2015

Abstract: Habitat simulation models heavily rely on monitoring data, which can have serious effects on the success of a physical habitat simulation. However, if data monitored in a study reach are not available or insufficient, then data from neighboring streams are commonly used. The problem is that the impact of using data from neighboring streams has rarely been studied before. Motivated by this, we report herein on an investigation of the transferability of data from neighboring streams in a physical habitat simulation. The study area is a 2.5 km long reach located downstream from a dam in the Dal River, Korea. Zacco platypus was selected as the target fish for the physical habitat simulation. Monitoring data for the Dal River and three neighboring streams were obtained. First, similarities in the data related to channel geometry and in the observed distribution of the target species were examined. Principal Component Analysis (PCA) was also carried out to see the characteristics of the habitat use of the target species. Habitat Suitability Curves (HSCs) were constructed using the Gene Expression Programming (GEP) model, and improved Generalized Habitat Suitability Curves (GHSCs) were proposed. The physical habitat simulations were then performed. The Composite Suitability Index (CSI) distributions were predicted, and the impact of using data from the neighboring streams was investigated. The results indicated that the use of data from a neighboring stream even in the same watershed can result in large errors in the prediction of CSI. The physical habitat simulation with the improved GHSCs was found to best predict the CSI.
1. Introduction

A physical habitat simulation is a numerical tool that quantifies physical habitat in terms of the flow depth, velocity, and substrate at a particular discharge for a given stream [1]. Thus, it is capable of predicting the impact of a change in flow on habitat availability for target species. Physical habitat simulations have been successfully used to estimate the environmental flows in rivers [2–5], in designing a river restoration [6–8], to evaluate river health [9–11], and to assess the impact of river works or river development [12–14].

For the success of a physical habitat simulation, acquiring relevant monitoring data is important. This is because most habitat simulation models are heavily dependent on the monitoring data [15–19]. However, a significant portion of previous physical habitat simulations have used monitoring data from neighboring streams due to the lack of sufficient data. The monitoring data include such physical habitat variables as flow depth, velocity, substrate, and related issues. Obtaining monitoring data is, in general, costly and time-consuming. For example, more than one hundred physical habitat simulations have been carried out for streams in Korea. However, only about 15% of these studies used data obtained for the actual stream being studied [20,21]. The situation is not much better in the US [18,22–25].

The use of monitoring data from neighboring streams involves an implicit hypothesis that the knowledge-based or data-based models constructed using data from a neighboring stream are applicable to the stream being studied. The similarity of HSCs between the study stream and neighboring streams has been studied by many researchers [18,23,25,26–32]. However, the impact of using such data from neighboring streams has rarely been investigated and a general and efficient solution to this problem has never been proposed. This motivated the present study.

The goal of this study was to assess the impact of using data from neighboring streams in a physical habitat simulation and to propose a generalized and efficient Habitat Suitability Index (HSI) model using data from neighboring streams. For this, a 2.5 km long reach in the Dal River, Korea was selected. This study reach is a gravel-bed stream located downstream from a dam. For the physical habitat simulation, Zacco platypus was selected for the target fish. Monitoring data from three neighboring streams were obtained for the physical habitat simulations of the study reach. Similarities of data for the channel geometry and for the observed distribution of the target fish against physical habitat variables were studied. Physical habitat simulations were carried out using the CCHE2D model and the HSI model for hydraulic and habitat simulations, respectively. For the HSI model, HSCs were constructed using the GEP model, and improved GHSCs were proposed using the suitable range concept. First, the impact of using data from neighboring streams in physical habitat simulations was examined quantitatively. Then, the improved GHSCs were used to predict the CSI distribution, and simulated results were compared.
2. Materials and Methods

2.1. Study Area

Figure 1 shows the study area in the Dal River, Korea and its neighboring streams. The Dal River is a mid-sized stream, a tributary of the Han River, and the basin area is 682.41 km$^2$. The study reach is 2.5 km in length and extends from the Sujeon Bridge to the Daesu Weir. The Goesan Dam, located 0.92 km upstream from the Sujeon Bridge, regulates the flow in the study reach. The Goesan dam discharges water irregularly only for hydropower generation. As a result, the role of the Daesu Weir is to maintain a constant flow during periods when the dam is not discharging water. For the study reach, the discharges for drought flow, low flow, normal flow, and averaged-wet flow are 1.82, 4.02, 7.23, and 17.13 m$^3$/s, respectively [33].

![Figure 1](image_url)  
**Figure 1.** The Dal River and its neighboring streams.

2.2. Monitoring Data

The three neighboring streams include the Hongcheon River, the Geum River, and the Chogang Stream, which are shown in Figure 1. For the Dal River and three neighboring streams, hydrologic, water quality, and fish monitoring data were collected for the period of 2007–2010 through government R&D projects [34,35]. To measure the water level and velocity, a radar water gauge and a price current meter were used, respectively, at the Sujeon Bridge. Dissolved oxygen and pH were measured using the handheld dissolved oxygen meter and pH meter, respectively. Turbidity was measured by turbidity...
meter (PT-200). Fish monitoring was carried out using cast nets and kick nets, revealing that dominant species in the study area is a minnow (*Zacco platypus*), followed by dark chubs (*Zacco temmincki*) and swiri (*Coreoleuciscus splendidus*). They account for 27%, 15% and 15%, respectively. In the present study, the adult minnow was selected as the target fish. Since the monitoring data includes the number of individuals, flow depth, velocity, and substrate, they are habitat use data of Category II based on the criterion by Bovee [15].

2.3. Habitat Suitability Curves

In the present study, the GEP was used to construct HSCs. GEP takes advantage of both the Genetic Algorithm (GA) and Genetic Programming (GP). The GEP uses linear chromosomes with a fixed length and nonlinear parse trees with varied sizes and shapes. The former is obtained from the GA and the latter from the GP. Therefore, in the GEP, individuals are encoded as chromosomes, which are then expressed as expression trees. The combination of these separate entities, chromosomes and expression trees, enables the GEP to perform with a high degree of efficiency compared to GA and GP.

For constructing HSCs, GeneXpro-Tools 5.0, a GEP software program developed by Ferreira [36,37] was used. Four basic arithmetic operators (+, −, ×, ÷) and some mathematical functions (power, abs, cos, tan, arctan, min, max) were used with a combination of all genetic operators, including mutation, transposition, and crossover.

2.4. Improved Generalized Habitat Suitability Curves

Maki-Petays *et al.* [25] introduced GHSCs for physical habitat simulations of juvenile salmon in four rivers in Finland. They constructed GHSCs for each habitat variable using the arithmetic means of the habitat suitability indices for the four neighboring rivers. To smooth the curves, Maki-Petays *et al.* [25] used the distance weighted least square method.

In the present study, a new method for constructing GHSCs by using the suitable range concept is proposed. The improved GHSCs use the arithmetic means of the HSCs of neighboring streams constructed by using data in a suitable range. The suitable range, the concept of which was proposed by Thomas and Bovee [26], is the range containing the central 95% of the occupied locations in the HSC.

The ranges of data for the Dal River and three neighboring streams are presented in Figure 2 where the total range, suitable range, and optimum range are denoted by dotted, black and red bold arrows, respectively. The optimum range is the interval containing the central 50% of the occupied locations in the HSC [26]. The suitable and optimum ranges of the Dal River data are shadowed by black and red, respectively, in the figure. It can be seen that the suitable and optimum ranges of the Chogang Stream data are the most similar to those of the Dal River data.
2.5. Hydraulic Simulation

CCHE2D, a numerical model for analyzing unsteady turbulent flows in an open-channel, was developed by the National Center for Computational Hydrosciences and Engineering at the University of Mississippi, US. The CCHE2D solves two-dimensional depth-averaged hydrodynamic equations using the efficient element method [38]. The continuity and longitudinal (x) and lateral (y) components of momentum equations are, respectively, given by:

$$\frac{\partial H}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = 0$$

(1)

$$\frac{\partial q_x}{\partial t} + \frac{\partial}{\partial x}(Uq_x) + \frac{\partial}{\partial y}(Vq_x) + \frac{g}{2}$$

$$= gH\left(S_{0x} - S_{h}\right) + \frac{1}{\rho}\left\{ \frac{\partial}{\partial x}(H\tau_{xx}) \right\} + \frac{1}{\rho}\left\{ \frac{\partial}{\partial y}(H\tau_{xy}) \right\}$$

(2)
\[
\frac{\partial q_x}{\partial t} + \frac{\partial}{\partial x}(Uq_x) + \frac{\partial}{\partial y}(Vq_y) + \frac{g}{2} \frac{\partial}{\partial x} H^2 = \begin{align*}
gH (S_{0x} - S_{x}) + \frac{1}{\rho} \left( \frac{\partial}{\partial x} (H\tau_{xy}) \right) + \frac{1}{\rho} \left( \frac{\partial}{\partial y} (H\tau_{xy}) \right) \end{align*}
\] (3)

where \( H \) is the flow depth, \( U \) and \( V \) are the depth-averaged velocities in the \( x \)- and \( y \)-directions, respectively, \( q_x \) and \( q_y \) are respective discharges per unit width (\( q_x = HU \), \( q_y = HV \)), \( g \) is the gravitational acceleration, \( \rho \) is the water density, \( S_{0x} \) and \( S_x \) are the river bed slope and friction slope in the \( i \)-direction, and \( \tau_{ij} \) is the horizontal turbulent stress tensor.

2.6. Method of Comparing Data and Results

For quantitative comparisons of geometric data of the streams, relative errors of each component defined below were computed.

\[
RE_i = \left[ \frac{\Phi_{i,NS} - \Phi_{i,Dal}}{\Phi_{i,Dal}} \right] \times 100 \% 
\] (4)

where \( \Phi_{i,Dal} \) denotes the \( i \)-th geometric component of the Dal River and \( \Phi_{i,NS} \) is the same component of the neighboring streams. In addition, to compare the observed and predicted CSIs, the following MAPE (Mean Average Percentage Error) is used:

\[
MAPE = \frac{1}{n} \sum \left[ \frac{CSI^o - CSI^p}{CSI^o} \right] \times 100 \% 
\] (5)

in which \( n \) is the number of data and \( CSI^o \) and \( CSI^p \) are the observed and predicted CSIs, respectively.

3. Results and Discussions

3.1. Data Variability between Target and Source Streams

The Hongcheon River belongs to the Han River basin, the same as the Dal River, whereas the other two belong to the Geum River basin. Detailed characteristics of channel geometry and the substrate of the Dal River and neighboring streams are given in Table 1. The shape factor, defined by the basin area divided by the length of the stream, in the table denotes the average width of the basin. It should be noted in the table that the Geum River is a large-sized sand-bed river whose average slope is mild compared to the other streams. The geometric components include basin area, stream length, average width, mean elevation, and mean slope. Average values of the relative errors of five geometric components in Table 1 are 51%, 357% and 25% for the Hongcheon River, the Geum River, and the Chogang Stream, respectively. This indicates that the Chogang Stream is the most similar to the Dal River in terms of channel geometry.
Table 1. Characteristics of the data used.

<table>
<thead>
<tr>
<th>Streams</th>
<th>Number of Datasets</th>
<th>basin Area (km²)</th>
<th>Length (km)</th>
<th>Average Width (km)</th>
<th>Shape Factor</th>
<th>Mean Elevation (EL. m)</th>
<th>Mean Slope</th>
<th>Mean Diameter of Substrate (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dal River</td>
<td>468</td>
<td>682.41</td>
<td>82.8</td>
<td>8.30</td>
<td>0.10</td>
<td>381.00</td>
<td>1/650</td>
<td>102.3–165.9</td>
</tr>
<tr>
<td>Hongcheon River</td>
<td>138</td>
<td>1566.2</td>
<td>140.2</td>
<td>11.17</td>
<td>0.08</td>
<td>423.00</td>
<td>1/550</td>
<td>35.1–148.0</td>
</tr>
<tr>
<td>Geum River</td>
<td>291</td>
<td>9912.15</td>
<td>397.8</td>
<td>24.92</td>
<td>0.063</td>
<td>217.03</td>
<td>1/2500</td>
<td>0.0097–1.70</td>
</tr>
<tr>
<td>Chogang Stream</td>
<td>689</td>
<td>665.2</td>
<td>66.3</td>
<td>10.03</td>
<td>0.15</td>
<td>289.00</td>
<td>1/770</td>
<td>46.8–111.8</td>
</tr>
</tbody>
</table>

3.2. Habitat Use Characteristics of Freshwater Minnow and Constructed HSCs

The distribution of the *Zacco platypus* in the Chogang Stream is the most similar to that in the Dal River (Figure 3). However, the target species in the Hongcheon River and Geum River appears to be distributed differently. In the figure, the classification scheme by Wentworth [39] was used for the substrate. Specifically, the ranges for the high population of *Zacco platypus* in the Dal River are a velocity of 0.4–0.65 m/s, a flow depth of 0.4–0.7 m, and a substrate of 4–6. However, in the Hongcheon River, the target species are densely populated in the ranges of a velocity of 0.6–1.0 m/s, a flow depth of 0.5–0.8 m, and a substrate of 2–4. In the Geum River, the ranges for a high population are a velocity of 0.1–0.6 m/s, a flow depth of 0.2–0.6 m, and a substrate of 2–5. The figure implies that monitoring data for the Chogang Stream can be used acceptably for a physical habitat simulation of the target fish in the study reach if data for the Dal River are not available.

Figure 3. Observed distribution of *Zacco platypus* against velocity, flow depth, and substrate for (a) Dal River; (b) Hongcheon River; (c) Geum River; (d) Chogang Stream.
In order to investigate the habitat use difference between the Dal River and the three neighboring streams, Principal Component Analysis (PCA) was carried out with seven variables, namely observed number of individuals, flow depth, velocity, substrate, pH, DO, and turbidity. The total number of data is 1586. A two-dimensional plot of PC2 versus PC1 is shown in Figure 4. It can be seen that the data can be grouped into four, which is the number of streams in this study. For PC1, the data for the Dal River lie in the range of −60–5. Ranges of data for PC1 are 10–60, 5–60, and −45–−10 for the Hongcheon River, the Geum River, and the Chogang Stream, respectively. For PC2, the data for the Dal River, the Hongcheon River, the Geum River, and the Chogang Stream range from −30–40, −60–0, 0–70, and 0–70, respectively. The results from PCA indicate that the pattern of the Chogang Stream data is the most similar to that of the Dal River data, followed by the Geum River data and the Hongcheon River data.

![Figure 4. Result of a principal component analysis.](image)

Before performing the physical habitat simulations, HSCs were constructed using the GEP model. For three physical habitat variables, HSCs were constructed with the three monitoring datasets from the neighboring streams and are compared with those constructed using the Dal River data. Figure 5 shows the resulting HSCs. The grey bars in the figure indicate that the number of observed individuals for each physical habitat variables. Choi and Choi [40] found that the HSCs by the GEP model are very similar to those by the method of Gosse [41] and the GEP model predicts HSCs better than the Adaptive Neuro Fuzzy Inference System (ANFIS) model. Furthermore, Choi and Choi [40] indicated that the GEP model is robust and non-subjective compared the method of Gosse [41].

For the flow depth and velocity, the HSCs for the Chogang Stream appear to be the most similar to those of the Dal River (Figure 5). However, the HSCs for the Hongcheon River are substantially different from those for the Dal River even though the two streams are part of the same watershed. The level of similarity can be expected from prior investigations in Table 1 and Figures 3 and 4.
Regarding the substrate, all HSCs differ seriously in Figure 5c. This is because the target fish, *Zacco platypus*, does not have a substrate preference [42,43]. That is, *Zacco platypus* lives in both sand-bed and gravel-bed streams. Thus, hereafter, the substrate will not be considered in the physical habitat simulations.

![Graphs showing habitat suitability curves for different factors](image_url)

**Figure 5.** (a) Flow depth; (b) velocity; (c) substrate. Habitat suitability curves.

The improved GHSCs are constructed and compared with the GHSCs constructed by the method proposed by Maki-Petays *et al.* [25] (Figure 6). It can be seen that both curves appear to be very similar except that the GHSCs have long tails in case where the values for the flow depth and velocity are large.

### 3.3. Transferability of HSCs and Validation of Improved GHSCs

In this section, CSI distributions of the Dal River are presented based on the physical habitat simulations. First, various HSCs in Figure 5, constructed using the study reach and the neighboring streams, were used for the HSI model. Then, GHSCs in Figure 6 were used to compute the CSI distributions. The former is to investigate the transferability of HSCs of the neighboring streams and the latter is to show the validity of the proposed GHSCs.
Using the HSCs in Figure 5, physical habitat simulations were carried out for *Zacco platypus*, and the resulting CSI distributions in the study reach are shown in Figure 7. The CSI was calculated using the multiplicative aggregation method. It can be seen that the CSI distribution computed using Chogang Stream data is the most similar to that of the Dal River. The CSI distribution computed using Geum River data is also similar, but the CSI distribution predicted with the use of Hongcheon River data is substantially different from the CSI distribution for the Dal River.

In order to evaluate the impact of using data from neighboring streams, the CSI of the Dal River versus the CSI predicted using data from neighboring streams are plotted in Figure 8. Only non-zero
values of CSI are plotted in the figure, where the 45 degree line indicates a perfect match. It can be seen that values of CSI predicted using Chogang Stream data provide the match best with those of the Dal River. However, the use of Hongcheon River data and Geum River data results in slight and serious over-predictions of CSI, respectively. The values for MAPE were also computed, and 67.5%, 39.2% and 25.1% were obtained for predictions using Hongcheon River data, Geum River Data, and Chogang Stream data, respectively. The results indicate that the use of data from a neighboring stream even in the same watershed results in larger errors in the prediction of the CSI. Prior investigations in Table 1 and Figures 3 and 4 can be useful for selecting appropriate datasets for a physical habitat simulation.

![Figure 8. Comparison of predicted CSIs.](image)

Figure 9 shows the CSI distributions for *Zacco platypus* in the Dal River. The GHSCs and improved GHSCs were used for CSI distributions in Figure 9a,b, respectively. It appears that the use of GHSCs substantially improves the CSI distribution compared to the CSI distributions in Figure 7a. Quantitatively, the uses of HSC constructed with the Chogang Stream data, GHSC, and improved GHSC result in MAPE values of 17.18%, 19.35% and 15.46%, respectively. This indicates that the use of the improved GHSC leads to a CSI distribution that is most similar to that of the Dal River.
4. Conclusions

This paper investigated the impact of using monitoring data from neighboring streams for a physical habitat simulation and proposed a new method for constructing GHSCs using monitoring data from neighboring streams. The present study showed that great attention should be paid to the use of data from neighboring streams when monitoring data are not available for the physical habitat simulation. That is, the data from a neighboring stream whose geometrical properties are similar to those of the study reach should be used. Even the data from a stream that shares the same watershed as the study reach can result in large errors in the prediction of the CSI.

In addition, for a general strategy, the present study proposed an improved GHSC, which can be used for the physical habitat simulation more confidentially with data from neighboring streams. The new method used the arithmetic means of the HSCs that were constructed with data only in a suitable range. The predicted CSI distribution was compared with that computed using the conventional HSI model, revealing that the prediction made using the improved GHSCs was better. However, the applicability of the proposed GHSCs can be investigated further by applying the methodology to various target streams that have their own monitoring data as well as that from neighboring streams.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (NRF-2014R1A2A1A11054236).
Author Contributions

All of the authors contributed extensively to the work. Byungwoong Choi and Sung-Uk Choi performed the calculation, result analysis, and wrote the research article. Hojeong Kang contributed to the article’s discussion and edition.

Conflicts of Interest

The authors declare no conflict of interest.

References


© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).