Application of genetic programming to project climate change impacts on the population of Formosan Landlocked Salmon

Ching-Pin Tung*, Tsung-Yu Lee, Yi-Chen E. Yang, Yun-Ju Chen

Bioenvironmental Systems Engineering, National Taiwan University, No. 1 Roosevelt Road, Section 4, Taipei 106, Taiwan

ABSTRACT

This work presents a novel methodology, genetic programming (GP), for developing environmental response functions for Formosan Landlocked Salmon (*Oncorhynchus masou formosanus*); these functions are then applied to evaluate the impacts of climate changes. Average daily temperature and maximal flows between two sampling periods were adopted as principal factors for categorizing environmental conditions. The GP successfully identified the response functions for various environmental categories. The response functions were further applied to assess the impact of climate change. Fourteen future possible climate scenarios were derived based on the equilibrium and transition experiments by GCMs. Impact assessment results indicated that climate change may significantly influence populations of Formosan Landlocked Salmon due to more frequent higher temperatures. Adaptation strategies are required to mitigate the impact of global climate change as current conservation measures for Formosan Landlocked Salmon habitat only reduce local human-induced effects. In the situation of complicated relationships between fish population and environmental conditions, GP provides a useful tool to obtain some information from the limited data.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Formosan Landlocked Salmon (*Oncorhynchus masou formosanus*) is a very important species as it is the salmon living at the lowest latitude. Although widely distributed in the upper Dajia Creek in early 1900s, it is currently only found in the uppermost reaches of Dajia Creek, Chichiawan Creek and Gaoshan Creek tributaries as in Fig. 1. Human-induced land use changes, hydraulic structures, poor water quality, increased water temperature and flooding are principal factors reducing habitat and endangering Formosan Landlocked Salmon. Conversion of land uses from forest to agriculture has been forbidden in the upper Dajia Creek, and some hydraulic structures have been removed and new construction prohibited. The factor endangering Formosan Landlocked Salmon most will be climate change. Climate change may influence streamflows, water temperature and habitats.

Increased water temperature, the primary effect of climate change, affects growth and survival rates of fish and other aquatic creatures, as well as migration patterns, breeding, and competitive ability of fish. Moreover, transmission of pollutants and resulting chemical reactions are also influenced by water temperature. These landlocked salmon only live in a water temperature range of 9–17 °C and, during spawning, the temperature must be below 12 °C (Tzeng, 1999). Generally, water temperature increases as it flows downstream. According to recent surveys (1985–1997), the 12 °C isotherm has moved 1.56 km upstream in the upper section of Dajia creek (Yang, 1997). This reduction in suitable spawning stream habitats can seriously influence the distribution of these landlocked salmon. Additionally, high flow rates also impact Formosan salmon. Sediment and increased flow generated by storm events degraded habitat and flushed fish downstream into inhospitable reaches. However, increased streamflow can also benefit habitats by reducing any increases in water temperature.

Numerous researchers, who have investigated the relationship between fish populations and environmental conditions, have identified decreasing populations of salmon and trout. Bradford and Irvine (2000) evaluated the impact of land use change, climate change and over-fishing on the population density of salmon. Shaw and Richardson (2001) assessed the direct and indirect effects of sediment pulse duration on rainbow trout survival. Cattaneo et al. (2002) concluded that flooding is the predominant factor influencing salmon and trout populations. Developed by Chen et al. (2000), the fuzzy logic model can assess the functional relationship between the number of spawners and sea surface temperature, reconstruct historical fish recruitment time series and predict the future fish populations.

* Corresponding author. Tel.: +886 2 2362 0327; fax: +886 2 2363 5854.
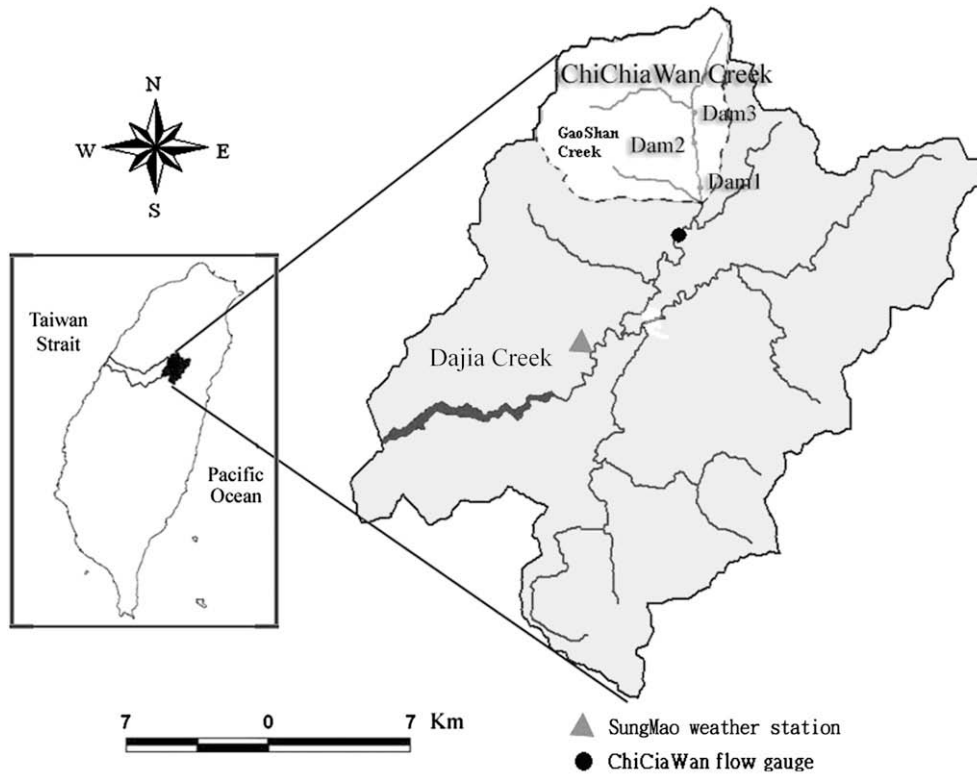 *E-mail address:* cptung@ntu.edu.tw (C.-P. Tung).

Fig. 1. The habitat of Formosan Landlocked Salmon – the upstream of the Dajia River.

According to IPCC 2007, ambient CO_2 concentrations will increase mean global air temperature by 1.1–6.4 °C by 2100 (IPCC WG I, 2007). Moreover, recent local research concluded that climate change may result in increased stream temperatures (Tung and Lee, 2006) and flows (Tung and Lee, 2001) in Formosan Landlocked Salmon's habitat. Thus, identifying response functions for assessing the impacts of climate change on Formosan Landlocked Salmon is necessary.

Differing environmental conditions impact ecosystems differently; thus, a predefined response function will likely limit predictability. In this study, environmental conditions are classified into several categories and a response function is then optimized by genetic programming (GP) for each category. GP is utilized here because it can find functions of various forms, not merely those involving a pre-determined polynomial with unknown coefficients. GP enhances the predictability particularly for this unknown situation. The study's goal is to identify environmental impact functions and their parameters for Formosan Landlocked Salmon. Section 2 describes the method employed to optimize response functions and Section 3 presents an analysis of critical environmental factors. Section 4 then addresses the method of evaluating climate change impacts on specific environmental factors. Simulation results and conclusions are presented in Sections 5 and 6, respectively.

2. Development of response functions

Different environmental conditions can have varying impacts of varying magnitudes on an ecosystem; thus, different functions may be required to predict the effects of various environmental conditions. GP generating functions helps us to know the response of a fish population to various environmental conditions. Furthermore, the functions can be applied to predict possible fish populations in the future. Environmental conditions in this study are categorized prior to optimizing response functions by GP. A brief description of the selection of environmental factors, data classification, and GP follows.

2.1. Selection of environmental factors

The most important environmental factors representing different environmental conditions are selected based on previous studies and available data. Then the relationships between these factors and the Formosan Landlocked Salmon population are analyzed. The final environmental factors adopted are utilized as predictors in the environmental response functions. Section 3 presents a detailed analysis of environmental factors.

2.2. Data classification

Environmental conditions are categorized based on environmental factor values. Each factor contains two groups: high and low values. A threshold (k_x) is chosen first for each environmental factor (X). If $X_i > k_x$, it is assigned to the high group, otherwise it is assigned to the low group. There are 2^n categories and associated functions if there are n environmental factors. For example, of two environmental factors, X_1 and X_2, there are $2^2 = 4$ categories and functions.

$$
\begin{aligned}
&\text{Category I} &&\text{if } x_{1i} > k_{x1} \text{ and } x_{2i} > k_{x2}, \text{then } Y_{I,i} = F_I(x_{1i}, x_{2i})\\
&\text{Category II} &&\text{if } x_{1i} > k_{x1} \text{ and } x_{2i} \le k_{x2}, \text{then } Y_{II,i} = F_{II}(x_{1i}, x_{2i})\\
&\text{Category III} &&\text{if } x_{1i} \le k_{x1} \text{ and } x_{2i} > k_{x2}, \text{then } Y_{III,i} = F_{III}(x_{1i}, x_{2i})\\
&\text{Category IV} &&\text{if } x_{1i} \le k_{x1} \text{ and } x_{2i} \le k_{x2}, \text{then } Y_{IV,i} = F_{IV}(x_{1i}, x_{2i})
\end{aligned}
$$

$$(1)$$

where x_{1i} and x_{2i} are the ith sampled data of environmental factors X_1 and X_2, and $Y_{j,i}$ is predicted fish population for category j by function F_j ($j = $ I,…,IV). The GP can be applied to identify functions

for each category with the objective function of minimizing root mean square error as Eq. (2).

$$\text{Min } Z = \left[\frac{1}{n_j}\sum_{i=1}^{n_j}\left(Y_{j,i} - y_{j,i}^0\right)^2\right]^{1/2} \tag{2}$$

where $y_{j,i}^0$ is the observed fish population and n_j is the number of observation data in category j. Thresholds k_x, which are determined by data analyses, can also be decision variables in an optimization process, although this study does not employ such optimization process.

2.3. Genetic programming

Genetic programming is an efficient method for constructing a relationship function comprising multiple independent variables (predictors). The GP optimizes both the coefficients and constants in a function and the function type itself. A possible function is determined by given mathematical operators, such as $+$, $-$, $\times$, sin, exp, etc. The GP encodes a function as a tree with nodes and branches, and then optimizes functions based on natural principles. The GP procedure is similar to that of a genetic algorithm, which generates solutions as a parent population, and then improves solutions by Selection, Crossover, and Mutation processes. The general procedure and primary components of GP are briefly described as follows (Koza, 1992).

Step 1 Generate initial parent population.
Step 2 Evaluate fitness of all alternatives.
Step 3 Select two parent alternatives for reproduction according to their fitness. Those with higher fitness are assigned greater probabilities to mate.
Step 4 Crossover to reproduce offspring and determine whether mutation occurs.
Step 5 Repeat steps 3–4 until the pre-determined population size is attained.
Step 6 Use the offspring population as a new generation and return to step 2 unless the stop criterion is met.

2.3.1. Encode and generate initial parent population

Different trees in GP represent different functions. For example, the functions of $y = x_1 + x_2$ and $y = (0.5 - x_1) + \exp(x_2)$ can be expressed as Fig. 2(a) and (b), respectively. The end node of each branch can only be filled with one constant or variable, whereas other nodes can only be filled with mathematical operators. These operators can be $+$, $-$, $\times$, $\div$, and any other operator, such as sin, cos,
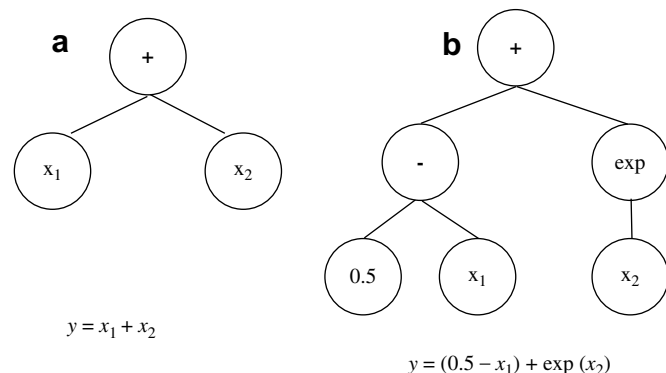

Fig. 2. The tree structure of the genetic programming method.

log, exponential, etc. Different operators have different child branches. For example, operators $+$, $-$, $\times$, $\div$ have two branches each, whereas sin and exponential have only one branch each. A function is only limited by given mathematical operators.

The initial parent population needs to be randomly generated. To generate a parent population, two sets must be defined, of which one set (S_1) contains variables and constants and the second set (S_2) contains designated mathematical operators. Fig. 3 presents the flowchart used to generate a tree (function). First, the maximum number of layers should be defined, allowing a node to continue branching downward until it reaches the predefined lowest layer or is filled with a constant or variable. A tree is generated when no more nodes are available for branching.

2.3.2. Selection

Selection can lead a search process toward a defined goal in the solution space. The principle of roulette wheel selection is used in this study. The solution that has superior fitness has an increased likelihood of being selected as a parent to generate new offspring (or a new solution). Since the objective function is to minimize root mean square errors as Eq. (2), the smaller values of the objective function, the better. Selection probabilities are assigned to solutions based on their objective function values using Eq. (3).

$$P_i = \frac{1/z_i}{\sum_{i=1}^{M} 1/z_i} \tag{3}$$

where P_i is the probability to select ith solution and z_i is the value of objective function of ith solution, and M is the population size. Uniformly distributed random numbers (ε) in the range between 0 and 1 are generated to select solutions among parent populations based on the criteria given in Eq. (4). One selects the kth solution when accumulated probability from the first to kth solutions is larger than the generated random number. For example, consider four solutions, and their probabilities $P_1 = 0.1$, $P_2 = 0.3$, $P_3 = 0.4$, and $P_4 = 0.2$, respectively. If we generate a random number $\varepsilon = 0.6$, then the third solution is selected due to $P_1 + P_2 = 0.5 < \varepsilon = 0.6 < P_1 + P_2 + P_3 = 0.8$.

$$\varepsilon \leq \min_k \sum_{i=1}^{k} P_i \tag{4}$$

2.3.3. Crossover

The crossover process produces new solutions by recombining parent chromosomes. One exchange node is first randomly determined for each selected parent tree. Then, the selected nodes and branches below the nodes are swapped. Fig. 4 shows with the right node (dot in tree (a)) and the left node (dot in tree (b)) selected as exchange nodes; crossover results in new solutions for trees (c) and (d).

2.3.4. Mutation

Mutation, a random change at a branch or node in an individual tree, diversifies the search process which avoids being restricted at a local optimum. Trees can be mutated in two ways. First, the content of one node is altered. A randomly chosen node is replaced. If the node was originally filled with an operator, only a new operator selected from set S_2 can replace the original operator. But the new operator must also have the same attributes (i.e. the number of child branches). For example, "$+$" requiring 2 branches can't be replaced with "exp" requiring only one branch. The second method changes one node and its following branches, providing more diversification than the first method. In applying this method, one node is first randomly selected and its branches deleted. A
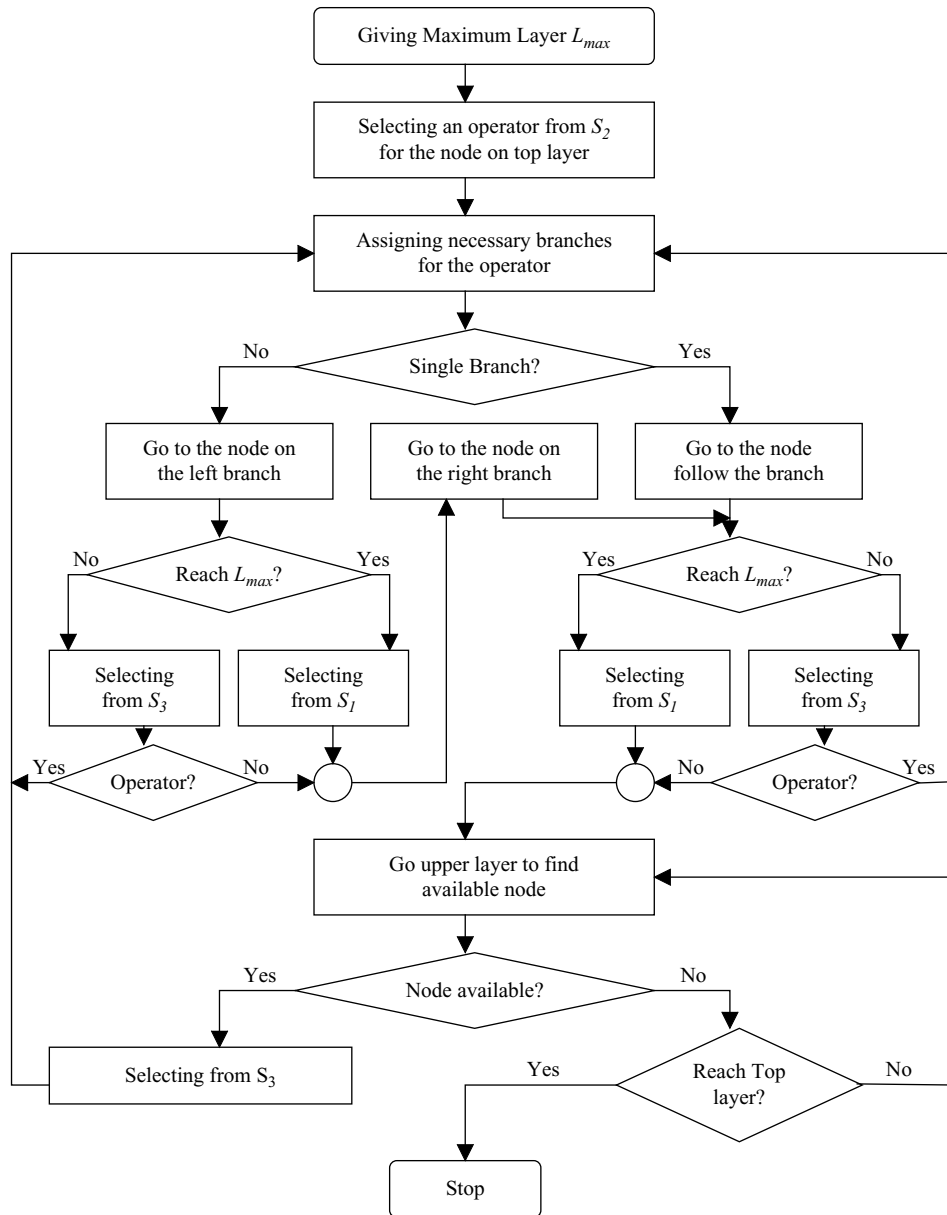
Fig. 3. The flowchart to generate a tree for genetic programming.

constant, a variable or an operator can be selected to replace the item in the node, and then new branches are generated with the same procedure used to generate a new tree described in Section 2. This study employs the first mutation method.

3. Analysis of environmental factors

Considerable researchers have concluded that the most important factors affecting fish populations include stream temperature, flow, sediment and nutrient levels (Yang, 1997; Shaw and Richardson, 2001; Cattaneo et al., 2002). The nutrient concentration meets water quality standards in the study area and is not considered to be a factor limiting fish population stability or growth. Furthermore, based on monitoring data from the study area, three factors were selected for further analysis: water temperature, streamflow and suspended sediment. The following analysis is applied to determine final indicators and their thresholds for classification.

3.1. Water temperature

Formosan Landlocked Salmon are very sensitive to water temperature. Suitable temperatures are in the range of 9–17 °C and must be below 12 °C during spawning (Tzeng, 1999). As a relationship exists between water and air temperatures (Mohseni and Stefan, 1999), air temperature is selected as an indicator to project climate change impact. The salmon population is measured twice in one year, once in summer and once in winter. Cumulative heat input to streams from winter in the previous year to summer influenced the population counted in summer, and the cumulative heat from summer to winter influenced the population measured in winter. Thus, average daily temperature between two sampling times can represent cumulative heat input to the stream in this study.

Fig. 5 presents the population of Formosan Landlocked Salmon plotted against average daily air temperature. Two data groups were easily identified, and the threshold between 17 °C and 18 °C was chosen. A linear decreasing trend was observed when
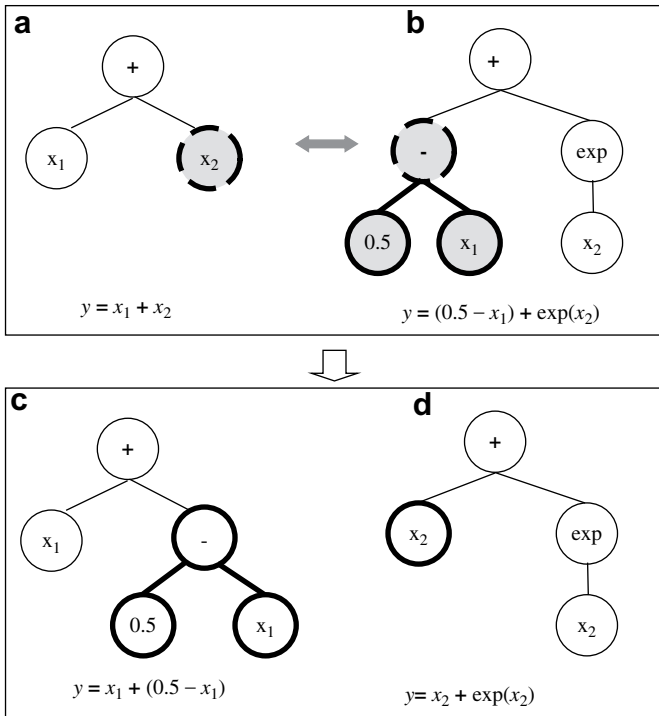
Fig. 4. An example of crossover in GP, where nodes with dot circle tree (a) and (b) are chosen as exchanging nodes.

temperature increased. Variation in the low temperature group was greater than in the high temperature group, suggesting that the indicator of average daily temperature is less dominant in the lower temperature group than the high temperature group. Based on these findings, average daily temperature between two sampling times was selected as an indicator, and air temperature of 17.5 °C was adopted as the threshold for separating temperature data into low and high temperature groups.

3.2. Streamflow

Fig. 6 shows the maximal flow observed between two sampling times versus the salmon population. Two groups were identified with a flow threshold between 40 cm and 60 cm. According to data analysis, a maximal daily streamflow threshold of 50 cm was adopted to separate flow data into low and high groups. The high flow group had fewer fish than the low flow group (Fig. 6). However, fish population changes in both groups due to streamflow
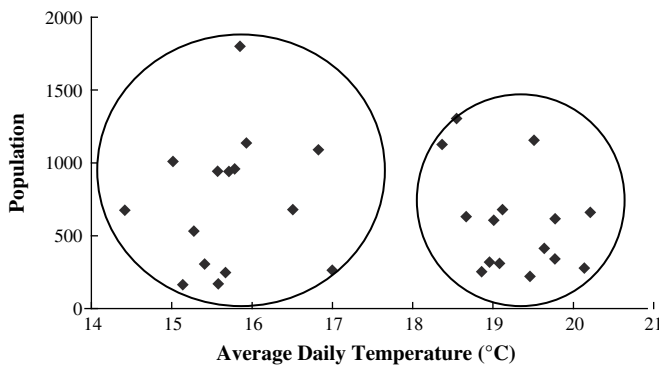

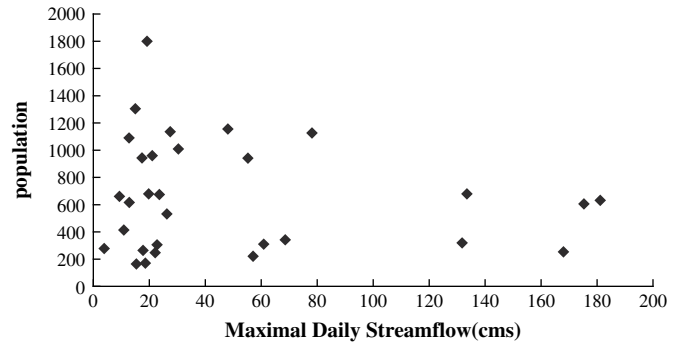**Fig. 5.** The plot of fish population versus average daily air temperature.

Fig. 6. The plot of fish population versus maximal daily streamflow.

increase were not significant – this change is further examined in Section 4.

3.3. Suspended sediment

Suspended sediment also affects stream habitat of these salmon. However, suspended sediment is not selected as an indicator in response function development, as suspended sediment and maximal streamflow are closely related. Fig. 7a presents the log linear relationship between suspended sediment and daily streamflow. Fig. 7b shows the linear relationship between average daily streamflow (total daily flow) and maximal daily streamflow. Thus, total suspended sediment and maximal daily streamflow are interdependent. Since maximal daily streamflow has been selected, suspended sediment is not included in response functions.

4. Climate change impact assessment

This section describes the procedure employed to assess the impact of climate change. First, future climate change scenarios derived from GCM simulations must be input into a weather generation model to produce weather data. Second, average daily
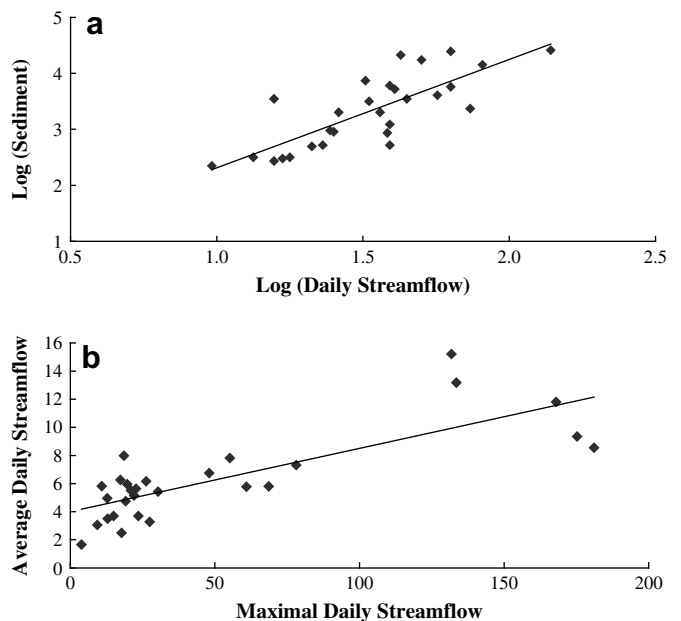

Fig. 7. (a) The plot of log sediment versus log daily streamflow; (b) the plot of maximal daily streamflow versus average daily streamflow.

air temperature and maximal streamflow in different climate conditions require evaluation. Then, the resulting data is input into the identified response functions to assess the impact of climate change on the population of Formosan Landlocked Salmon.

4.1. Climate change scenarios

4.1.1. Future climate scenarios

Climate scenarios are derived from various GCMs and experiments, including equilibrium experiments and transition experiments based on SRES (the Special Report on Emissions Scenarios) (Nakicenovic et al., 2000). Change in future temperature in the study area was assumed to be the same as the difference between the temperatures simulated by GCMs for future and current conditions at the nearest grid point. Thus, future climate scenarios can be estimated as

$$\mu'_{mT} = \mu_{mT} + (\mu_{mT,Future} - \mu_{mT,Current}) \qquad (5)$$

where μ_{mT} and μ_{mT}' are current and future mean monthly temperatures (°C), respectively, and $\mu_{mT,current}$ and $\mu_{mT,Future}$ are simulated mean monthly temperatures (°C) under current and future climate conditions, respectively. Change in precipitation is assumed as the ratio of the precipitation under future conditions to that under current conditions:

$$\mu'_{mP} = \mu_{mP}(\mu_{mP,Future}/\mu_{mP,Current}) \qquad (6)$$

where μ_{mP} and μ_{mP}' are current and future mean monthly precipitation values (cm), respectively, and $\mu_{mP,Current}$ and $\mu_{mP,Future}$ are simulated mean monthly precipitation values (cm) under current and future climate conditions, respectively.

The GCM predictions based on equilibrium experiments (1995 version) were downloaded from the US Country Studies Program on the NCAR ftp site (ftp://ncardata.ucar.edu/pub). Output from the CCCM (Canadian Center for Climate Modeling) and GISS (Goddard Institute for Space Studies) models was adopted. Table 1 lists the change of mean monthly temperature and the ratio of precipitation between $1 \times CO_2$ and $2 \times CO_2$ conditions.

The transitional experiments by CGCM2 and HADCM3 models based on A2 and B2 scenarios of SRES were applied to establish other future climate scenarios in this study. Three future scenarios are considered: short-term (2010–2039), mid-term (2040–2069), and long-term (2070–2099). Future climate scenarios are also determined by Eqs. (5) and (6), in which $\mu_{mT,Current}$ and $\mu_{mP,Current}$ are averaged from the GCM base simulation for 1961–1990 and $\mu_{mT,Future}$ and $\mu_{mP,Future}$ are average values of different periods.

Table 1
Changes of mean monthly temperature (ΔT, °C) and ratios of mean monthly precipitation (R_p, cm/cm) – equilibrium experiments.

Month	CCCM		GISS	
	ΔT	R_p	ΔT	R_p
1	2.71	0.81	2.74	1.13
2	3.66	1.05	2.59	0.68
3	4.73	0.67	3.31	1.03
4	4.16	1.11	3.75	1.33
5	4.21	1.13	3.58	0.91
6	2.52	1.39	4.69	1.51
7	2.09	1.00	4.52	1.30
8	1.75	1.19	4.18	1.23
9	2.62	1.40	3.52	1.30
10	2.45	1.01	2.92	0.98
11	2.39	0.86	3.72	1.15
12	3.42	0.66	2.42	0.94
Average	3.02	1.02	3.50	1.12

Tables 2 and 3 present the climate change estimates based on A2 and B2 scenarios, respectively.

4.1.2. Weather generation

A weather generation model generated daily air temperature and precipitation values for various climate scenarios. The weather generation model used in Tung and Haith (1995) and Tung and Haith (1998) was applied in the study. A sequence of daily precipitation and air temperature over 100 years was generated for each current or future climate scenario.

4.2. Impact assessment

The populations of Formosan Landlocked Salmon under different climate scenarios were predicted by input average daily temperatures and maximal flows. Two environmental factors under different climate scenarios were evaluated first. The streamflow component of the Generalized Watershed Loading Functions (GWLF) (Haith and Shoemaker, 1987) simulated streamflows. The plausibility of predictions by the streamflow model has been verified by Tung and Lee (2001) for the ChiChaWan Creek. The model requires daily temperature and precipitation as inputs and was modified to produce maximal flows. Generated weather data for both current and future climates input into the model determined maximal streamflows. Average daily temperatures were simply calculated from current and future synthetic weather data.

4.3. Data and model setup

The population of Formosan Landlocked Salmon has been surveyed twice a year since 1987. The fish population between Dam1 and Dam3, which includes major habitats for Formosan Landlocked Salmon, was used in the GP model. Daily air temperatures and streamflows in 1987–2001 were utilized to identify response functions, for which data were obtained from the SungMao weather station and Chichiawan gauge station, respectively. Data were classified into three categories: High Temperature/High Flow (HH); High Temperature/Low Flow (HL); and, Low Temperature/Low Flow (LL). No data exist for the Low Temperature and High Flow (LH) category.

The followings are the necessary settings for GP. Set S_1 contains two variables $\{X_1, X_2\}$ representing average daily air temperature and maximal streamflow, whereas set S_2 has operators $\{+, -, \times, \div\}$. Six maximal layers and a population (parents and offspring) of 1000 are adopted. The stop criterion for GP is 1000 generations. Air temperature and flow in each category are standardized to eliminate unit effects before applying in GP.

5. Results and discussions

The response functions identified are addressed first. These functions are then applied to assess the impact of climate changes on Formosan Landlocked Salmon.

5.1. Calibration of response functions

The thresholds of 17.5 °C for temperature and 50 cm for streamflow were applied to categorize data. Then, GP was applied to identify response functions for different categories. To verify identified response functions, observed data in each category were randomly selected as a data set for validation. Response functions based on the remaining data were then used to construct response functions shown as Eq. (7).

Table 2
Changes in mean monthly weather parameters predicted by the CGCM2 and HCCPR general circulation models for SRES-A2 scenario.

Month	CGCM2						HADCM3					
	ΔT			R_p			ΔT			R_p		
	S	M	L	S	M	L	S	M	L	S	M	L
1	0.94	1.62	2.76	1.22	1.12	1.02	0.84	1.82	3.11	1.13	0.85	0.95
2	1.81	2.57	3.27	1.35	1.01	0.89	0.46	1.55	2.40	0.95	0.68	1.00
3	1.25	2.64	3.78	1.3	1.03	0.88	0.52	1.44	2.45	0.95	1.07	0.92
4	0.6	1.82	4.64	1.13	0.8	0.76	0.61	1.52	2.49	1.13	1.31	1.28
5	−0.46	2.78	4.54	1.1	0.82	0.54	0.29	1.35	2.20	1.08	1.30	1.38
6	1.09	3.62	5.62	1.03	0.79	0.7	0.63	1.35	2.33	1.30	1.21	1.01
7	1.18	2.26	4.52	0.92	0.93	0.7	0.53	1.32	2.28	1.14	1.10	1.37
8	0.52	2.05	3.45	1.19	1.03	1.23	0.40	1.37	2.23	1.21	1.27	1.39
9	0.45	1.9	3.12	1.18	0.97	1.29	0.60	1.46	3.04	1.27	0.98	1.25
10	0.43	1.81	2.68	0.92	0.99	1.09	0.91	1.50	2.99	1.07	1.53	1.47
11	0.47	2.1	2.53	1.1	1.24	0.8	0.75	1.47	2.59	0.85	0.80	0.91
12	1.41	2.43	2.81	0.96	0.88	0.69	0.90	1.75	3.10	1.08	1.00	0.88
Average	0.81	2.30	3.64	1.12	0.97	0.88	0.62	1.49	2.60	1.10	1.09	1.15

For HH category: $\quad Pop = \dfrac{F}{3T(2F+T) - \dfrac{T+F}{T^2}}$

For HL category: $\quad Pop = F(2T - F) - \dfrac{\dfrac{T}{2} - \dfrac{1}{T}}{2T^2 + 2T^3}$ (7)

For LL category: $\quad Pop = 3F + T - 9TF$

$Pop = \dfrac{pop}{1800}; \quad F = \dfrac{f}{181.12}; \quad T = \dfrac{t - 14.42}{20.21 - 14.42}$

where pop, f, and t represent observed fish population, maximal streamflow, and average air temperature in measured units. Pop, F and T represent standardized fish population, maximal streamflow, and average air temperature, respectively. 1800 and 181.1 cms is the historical observed maximal fish population and streamflow; 20.21 °C and 14.42 °C are the observed maximal and minimal average air temperature during the study period. Standardization makes the magnitude of three variables of the same order to prevent the functions from being dominated by some variables. In fact, GP has the ability to adjust each variable to the most suitable magnitude in the function (through the combination of operator nodes and constant child branches). However, pre-standardization makes GP more efficient in searching the optimal solution. After the response functions were constructed, validation data sets were then used to test the forecast ability of the functions. The results are as follows.
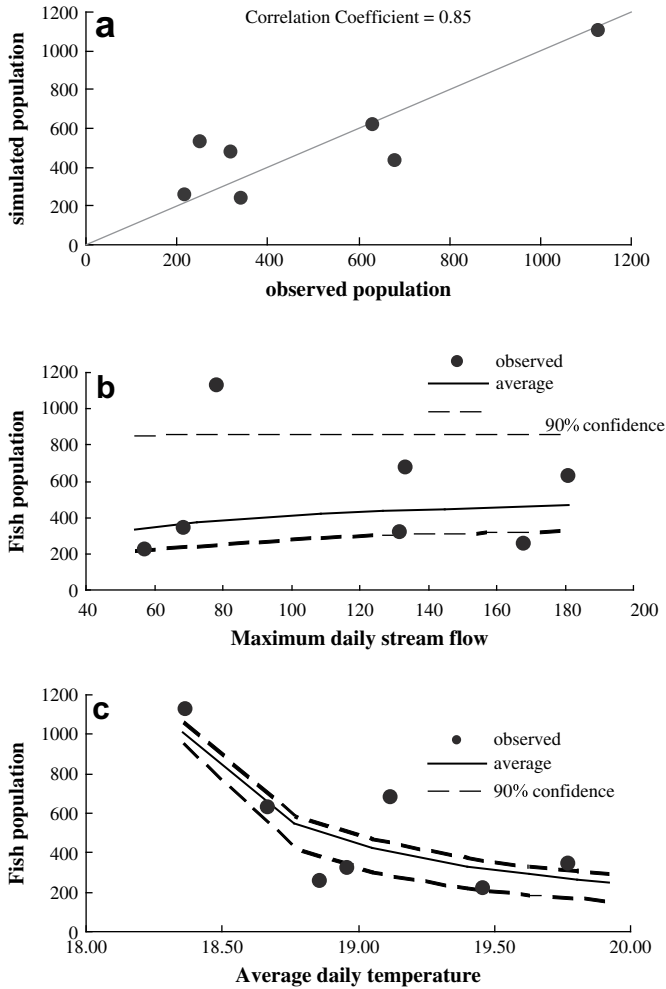
5.1.1. High temperature and high streamflow (HH)

Fig. 8a shows the scatter plot of simulated versus observed fish populations. The root mean square error and the correlation coefficient for this category were 158 and 0.85, respectively. Fig. 8b and c presents the plotted sensitivity of the response function, showing the changes of fish population when air temperature (Fig. 8b) or maximal flow (Fig. 8c) is fixed to average value and within 90% confidence interval. The salmon population increased slightly as flow increased under fixed air temperature, indicating that increasing flows to reduce stream temperature is favorable when air temperature is high and fixed (Fig. 8b). The salmon population decreased as air temperature increased under fixed flow (Fig. 8c).

As flooding can threaten fish populations, a fish population may decrease when streamflow increases. However, this relationship was not significant in this study. To confirm this analytical finding, the relationship between observed fish population and the maximal streamflow in the reach – a heavy storm caused by a typhoon – was analyzed (Table 4). No significant relationship between the number of surviving fish and maximal flow can be found for the study reach. Fish population decreased under some scenarios and increased under others, such as the AMBER and BILIS typhoons. The most upstream fish population may be flushed into the study reach to increase mainstream population, indicating that future studies need to extend the study area upstream.

Table 3
Changes in mean monthly weather parameters predicted by the CCCM and HCCPR general circulation models for SRES-B2 scenario.

Month	CGCM2						HADCM3					
	ΔT			R_p			ΔT			R_p		
	S	M	L	S	M	L	S	M	L	S	M	L
1	1.36	1.71	1.65	1.14	1.00	0.83	0.75	1.48	1.93	0.85	0.94	0.77
2	1.48	1.81	2.54	1.45	0.91	1.01	0.56	1.01	1.66	1.09	0.86	0.91
3	0.71	2.44	3.26	1.18	0.97	0.95	0.48	0.99	1.49	0.99	1.08	0.85
4	0.62	2.28	3.09	1.15	0.98	0.85	0.51	1.12	1.86	0.99	1.13	1.39
5	1.69	1.76	2.75	0.83	0.95	0.77	0.62	1.31	1.72	0.87	1.11	1.60
6	2.60	1.72	3.18	0.76	1.18	0.95	0.88	1.40	1.78	1.03	1.15	1.11
7	1.59	1.47	2.60	0.96	0.97	0.97	0.71	1.31	1.75	1.30	1.20	1.04
8	0.72	1.16	1.84	1.12	1.07	1.26	0.50	1.24	1.67	1.36	1.14	1.45
9	0.73	1.28	1.80	1.07	1.04	1.19	0.79	1.45	2.22	1.21	1.27	1.17
10	1.11	1.37	1.58	1.02	1.08	1.39	0.60	1.08	2.03	0.70	1.13	1.23
11	0.55	1.72	1.80	0.82	0.94	1.00	0.61	1.30	1.96	0.78	1.20	1.09
12	1.60	1.30	2.63	0.94	0.56	0.76	0.7	1.73	2.39	1.00	0.98	1.15
Average	1.23	1.67	2.39	1.04	0.97	0.99	0.64	1.29	1.87	1.01	1.10	1.15

Fig. 8. (a) The scatter plot of simulated versus observed fish population; (b) the relationship between fish population and maximal daily streamflow; (c) the relationship between fish population and average daily air temperature – the category of High temperature and High streamflow.

5.1.2. High temperature and low flow (HL)

Fig. 9a shows the plotted relationship between the simulated population versus observed population. The root mean square error (RMSE) and correlation coefficient were 142 and 0.93, respectively. Fig. 9b and c shows sensitivity analysis results. The population will increase when streamflow increases and air temperature is fixed, indicating that increased flow reduces stream temperature and is favorable to Formosan Landlocked Salmon under high air temperatures (Fig. 9b). Comparing Fig. 9b with Fig. 8b, the population increase rates for HL category were higher than those of the HH category, indicating that increasing streamflow benefits salmon more when flow is low. The identified response function for the HL
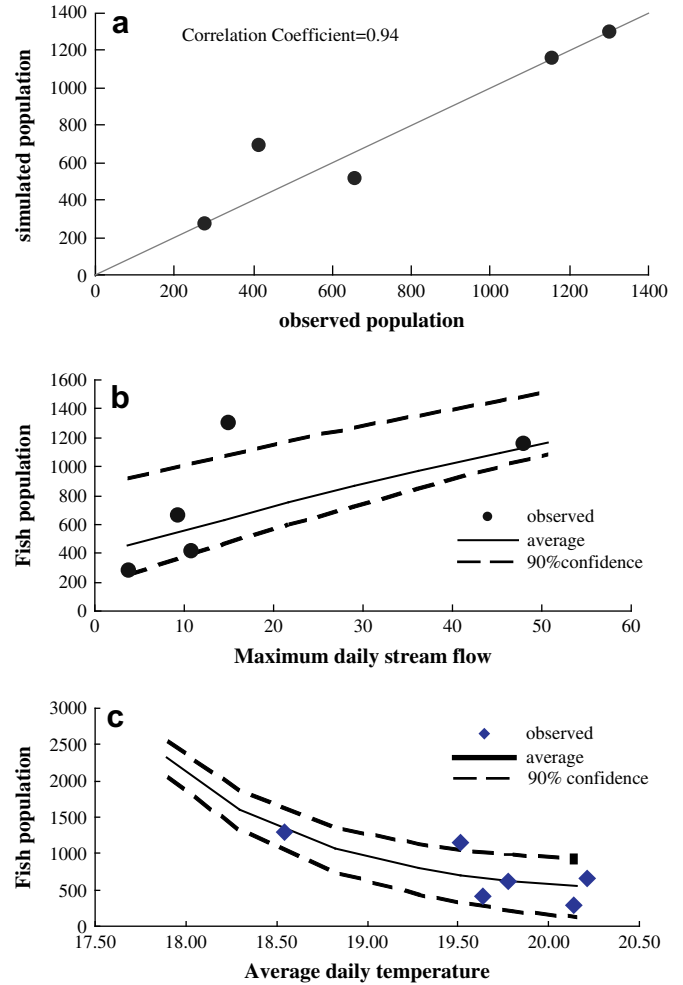
Fig. 9. (a) The scatter plot of simulated versus observed fish population; (b) the relationship between fish population and highest daily streamflow; (c) the relationship between fish population and average daily air temperature – the category of High air temperature and Low streamflow.

category successfully shows that the Formosan Landlocked Salmon does not favor higher temperature when streamflow is low and fixed (Fig. 9c).

5.1.3. Low temperature and low flow (LL)

Fig. 10 shows the relationship between simulated and observed salmon populations. The RMSE and correlation coefficient were 519 and 0.20, respectively. The identified function tends to predict the

Table 4
The relationship between fish population and the highest discharge due to typhoon.

Typhoon's name	Date	Q_{peak} (cms)	Population before typhoon	Population after typhoon
SARAH	1989/09/12	175.3	1136	606
DOT	1990/09/08	133.5	941	679
DOUG	1994/08/08	131.8	263	319
HERB	1996/08/31	181.1	675	631
AMBER	1997/08/29	78.1	1009	1127
BILIS	2000/08/23	68.6	305	341
XANGSANE	2000/11/01	15.4	341	164
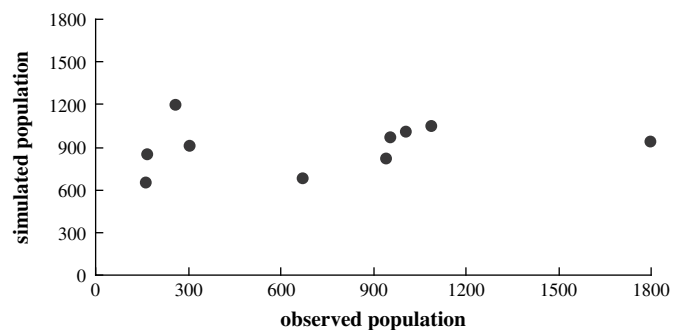
Fig. 10. The scatter plot of simulated versus observed fish population – the category of Low air temperature and Low streamflow.

Table 5
The results of verifications of response functions for different categories.

Category	Air temperature	Streamflow	RMSE	Correlation coefficient
Category 1 – HH	High	High	158	0.85
Category 2 – HL	High	Low	142	0.93
Category 3 – LH	Low	High	N/A	N/A
Category 4 – LL	Low	Low	519	0.20

mean salmon population (Fig. 10). The identified function did not respond to changes of the two environmental indicators, indicating that the two environmental indicators are not factors limiting the Formosan Landlocked Salmon population in this category. The factors influencing fish populations require further research.

5.2. Validation of response functions

Table 5 presents a summary of response function results. The functions had good predictability for categories HH and HL, but only forecast the mean value for category of LL. Thus, response functions for HH and HL categories were further validated. Fig. 11 presents these validation results. The solid and hollow dots in Fig. 11 represent the remaining and removed observed data for HH, respectively, and the solid and hollow triangles represent the remaining and removed observed data for HL, respectively. Due to limited available data, only three records for fish populations were used for validation. The response functions successfully predicted populations for HH and HL (Fig. 11). The response functions provided reasonable predictions for the categories of HH and HL and, thus were further applied to climate change impact assessment.

Comparing the performance of GP to the commonly used linear regression method, Table 6 shows the respective RMSE and correlation coefficient of the calibration and validation. Estimated values from both HH and HL categories are combined for evaluation. The LL category is ignored because both linear regression and GP estimate a mean value only and hence are undiscriminating. It is apparent that the GP method has the overall superiority. In the validation set, the correlation coefficient estimated by linear regression method is even less than zero. This means that the linear regression method can only delineate the overall trend of the data sets used and is not adequate for emulating the true data-generating process. Compared to the linear regression method, GP has the better ability to simulate how the fish population reflect to two of the variables that are indicative of existing environmental conditions.

6. Climate change impacts on Formosan Landlocked Salmon

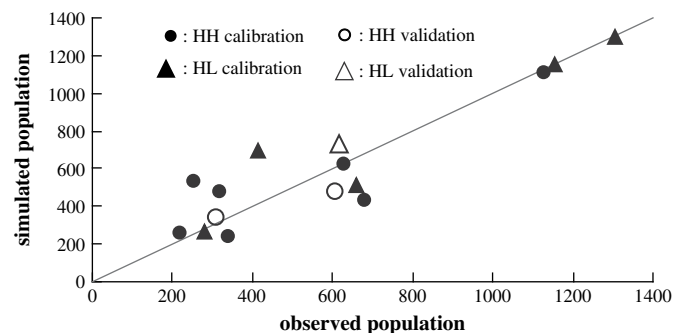Tables 7 and 8 show the impacts on Formosan Landlocked Salmon under High Temperature/High Flow and High Temperature/

Fig. 11. The validation of the response functions for the categories of HH and HL.

Table 6
The performance of linear and GP methods. The root mean square error (RMSE) and correlation coefficient (r) of the corresponding data sets (Calibration/Validation/ Calibration & Validation) in the combined High temperature and High flow (HH) and High temperature and Low flow (HL) category.

	Linear		GP	
	RMSE	r	RMSE	r
Calibration	179	0.87	151	0.91
Validation	251	−0.52	100	0.79
Calibration & Validation	196	0.82	142	0.91

Low Flow for different climate conditions. Fig. 12 presents a plot of the 25th, 50th, and 75th percentiles of the fish population for the two categories under different climate conditions. As the response function of Low Temperature/Low Flow did not successfully provide predictability, it was not further applied in the climate change impact evaluation.

There was a decrease for all the 25th, 50th, and 75th percentile fish populations in HH (Fig. 12a and Table 7). According to hypothesis test results (Table 8), all changes were significant at the 5% level. The results based on two transition experiments (CGCM2 and HADCM3) showed similar decreasing patterns, and the projected impacts on fish population based on two equilibrium experiments (CCCM and GISS) were similar to those based on the long-term predictions of the two transition experiments. Comparing the results caused by the A2 and B2 climate change scenarios, the fish population decrease under the A2 scenario was higher than that under B2, suggesting that the strategy of reducing emissions of greenhouse gases could mitigate climate change impacts on Formosan Landlocked Salmon.

In the HL category, the 50th percentile of the population may decrease until the mid-term period, and then begins to increase (Fig. 12b). This finding results from streamflow continually increasing under HL, mitigating the negative effect of high temperature in long-term climate conditions. The mean salmon population may be reduced in all future climate conditions (Table 8). However, changes in some long-term scenarios, such as in the GISS equilibrium experiment and both CGCM2 and HADCM3 A2 experiments, were not significant.

Climate change impact had differing effects on population variations under HH and HL. The coefficients of variation in HH may become smaller, whereas those in HL may change little or even

Table 7
The percentage of change of fish population under different climate scenarios– the "HH" and "HL" categories.

			HH			HL		
			P_{25}	P_{50}	P_{75}	P_{25}	P_{50}	P_{75}
Short-term	A2	CGCM2	−16.7%	−19.7%	−23.3%	−10.9%	−1.4%	−10.0%
		HADCM3	−12.6%	−15.9%	−21.1%	−4.8%	−4.0%	−9.7%
	B2	CGCM2	−30.0%	−34.9%	−33.8%	−21.9%	−17.6%	−18.2%
		HADCM3	−9.9%	−11.9%	−19.8%	−3.8%	1.3%	−5.3%
Mid-term	A2	CGCM2	−50.4%	−53.1%	−53.7%	−41.8%	−28.9%	−26.4%
		HADCM3	−33.5%	−35.9%	−42.8%	−15.0%	−4.9%	−9.0%
	B2	CGCM2	−34.9%	−39.5%	−41.5%	−23.7%	−19.9%	−20.1%
		HADCM3	−31.3%	−33.5%	−36.4%	−11.8%	−10.4%	−12.8%
Long-term	A2	CGCM2	−61.5%	−61.8%	−64.6%	−33.7%	−12.0%	−10.1%
		HADCM3	−52.9%	−50.3%	−53.4%	1.5%	0.5%	−10.2%
	B2	CGCM2	−42.4%	−45.8%	−47.1%	−16.0%	−12.4%	−7.9%
		HADCM3	−39.4%	−42.5%	−46.9%	−29.0%	1.4%	−10.4%
Equilibrium		CCCM	−48.2%	−47.2%	−52.0%	−15.6%	−12.6%	−17.8%
		GISS	−63.1%	−62.8%	−66.9%	−2.5%	1.6%	−5.6%

P_{25}, P_{50}, and P_{75} represent the 25th, 50th, and 75th percentiles of fish population, respectively.

Table 8
The statistics of simulated fish population under different climate scenarios – the "HH" and "HL" categories.

			HH			HL		
			μ	σ	σ/μ	μ	σ	σ/μ
Current			284	223	0.78	865	283	0.33
Short-term	A2	CGCM2	203[a]	57	0.28	768[a]	227	0.30
		HADCM3	210[a]	58	0.28	791[b]	188	0.24
	B2	CGCM2	169[a]	42	0.25	674[a]	214	0.32
		HADCM3	216[a]	56	0.26	801	205	0.26
Mid-term	A2	CGCM2	118[a]	28	0.24	595[a]	259	0.44
		HADCM3	155[a]	36	0.23	736[a]	235	0.32
	B2	CGCM2	151[a]	36	0.24	667[a]	232	0.35
		HADCM3	162[a]	39	0.24	734[a]	199	0.27
Long-term	A2	CGCM2	101[a]	72	0.71	760	427	0.56
		HADCM3	120[a]	32	0.26	810	248	0.31
	B2	CGCM2	135[a]	31	0.23	716[a]	261	0.36
		HADCM3	141[a]	32	0.23	733[a]	250	0.34
Equilibrium		CCCM	136[a]	91	0.67	712[a]	234	0.33
		GISS	89[a]	22	0.25	793	306	0.37

[a] Significant change with 95% reliable interval.
[b] Significant change with 90% reliable interval.

increase under future climate conditions. The reason for these projections may be that streamflow has a less mitigating influence on the negative effect of temperature under HH than HL. Higher temperature tends to reduce the variations in salmon populations. The category HH has high flows currently and, thus, an increase in streamflow due to future climate change will have less influence, making temperature a dominant factor affecting variation in salmon populations. The category HL had low flows and, thus, increased streamflow can mitigate the negative effect of temperature. Thus, higher variation of streamflow under HL may result in high variation of fish populations.

Fig. 12. The fish populations under different climate scenarios (a) "High temperature and High streamflow" category; (b) "High temperature and Low streamflow". P_{25}, P_{50}, and P_{75} represent the fish population of 25th, 50th, and 75th percentile, respectively.

7. Conclusions

This study proposes a procedure for identifying environmental response functions for Formosan Landlocked Salmon under different environmental conditions, to which GP is applied to optimize both the type and coefficients of response functions. Identified response functions were further applied to assess the impacts of climate change on Formosan Landlocked Salmon. Two principal environmental factors, average daily temperature and maximal flow, were used in categorizing data for the study area.
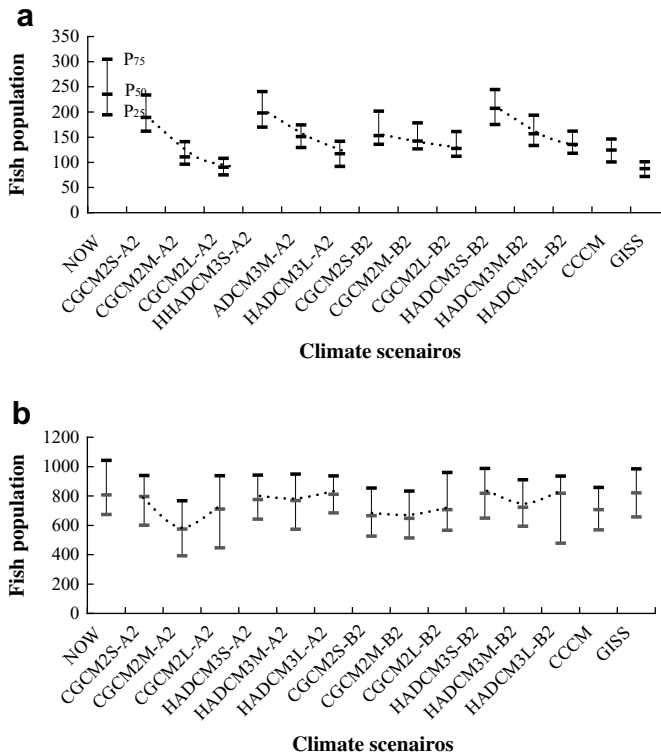
The results indicated that GP successfully identifies environmental response functions for categories of High Temperature and High Flow (HH) and High Temperature and Low Flow (HL), and provides a mean population for Low Temperature and Low Flow (LL). It is likely that other factors influence fish populations for Low Temperature and Low Flow and these require further study. Due to limited observational data, only three records were used for validation. More frequent observations are suggested. If more data are available, response functions can be more robustly identified. Furthermore, the proposed analysis procedure and methodology should work well for other aquatic species whose survival is closely related to environmental conditions.

Taiwan's conservation agencies have made considerable effort to improve Formosan Landlocked Salmon habitat, including dismantling weirs and stopping both farming and land use conversion in sensitive habitat; however, such measures may prove inadequate in the face of climate change. Impact assessment results demonstrated that the Formosan Landlocked Salmon population, under conditions of High Temperature and High Flow and High Temperature and Low Flow, may be significantly reduced by climate change. Tung and Lee (2001) concluded that the frequency of environmental conditions of HH and HL categories will likely increase. Thus, global climate change may seriously endanger Formosan Landlocked Salmon. Current conservation measures can halt local human-induced threats, but not the impact from global climate change. Identifying successful adaptive strategies to mitigate the future climate change impact is urgent. Climate change impact assessments should be integrated into the process of developing or modifying conservation strategies.

GP is useful, particularly for dealing with unknown relationships between dependent variables and predictors, because GP is able to emulate almost any functional form. Even if a data set has unknown linear or nonlinear relationships, GP is able to identify an "optimal" function. However, if the relationship between dependent variables and predictors is well known, such as linear/polynomial relationship, regression analysis is suggested because it may not be worth using GP. Therefore, we would suggest using GP when the relationship among variables is not clear. A very large data set increases computational time, but it actually comprises more-complete response relationships among variables, which help GP to identify functions that better emulate the true data-generating process. Collecting more data is encouraged and it is suggested that data are classified into several categories according to the level of predictors, which would increase the computational efficiency of GP.

References

Bradford, M.J., Irvine, J.R., 2000. Land use, fishing, climate change, and the decline of Thompson River, British Columbia, Coho salmon. Canadian Journal of Fisheries and Aquatic Sciences 57, 13–16.

Cattaneo, F., Lamouroux, N., Breil, P., Capra, H., 2002. The influence of hydrological and biotic process on brown trout (Salmo trutta) population dynamics. Canadian Journal of Fisheries and Aquatic Sciences 59, 12–22.

Chen, D.G., Hargreaves, N.B., Ware, D.M., Liu, Y., 2000. A fuzzy logic model with genetic algorithm for analyzing fish stock–recruitment relationships. Canadian Journal of Fisheries and Aquatic Sciences 57, 1878–1887.

Haith, D.A., Shoemaker, L.L., 1987. Generalized watershed loading functions for streamflow nutrients. Water Resource Bulletin 23 (3), 471–478.

IPCC Working Group I, 2007. In: Solomon, S. (Ed.), Climate Change 2007 – The Physical Science Basis. Cambridge University Press, UK.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press., Mass.

Mohseni, O., Stefan, H.G., 1999. Stream temperature/air temperature relationship: a physical interpretation. Journal of Hydrology 218, 128–141.

Shaw, E.A., Richardson, J.S., 2001. Direct and indirect effects of sediment pulse duration on stream invertebrate assemblages and rainbow trout growth and survival. Canadian Journal of Fisheries and Aquatic Sciences 58, 2213–2221.

Tung, C.P., Haith, D.A., 1995. Global warming effects on New York streamflows. ASCE Journal of Water Resources Planning and Management 121 (2), 216–225.

Tung, C.P., Haith, D.A., 1998. Climate change, irrigation and crop response. Journal of the American Water Resources Association 34 (5), 1071–1085.

Tung, C.P., Lee, T.Y., 2001. Climate change impact assessment of the Chichiawan Creek streamflow. Chinese Journal of Agricultural Engineering 47 (1), 65–74 (in Chinese).

Tung, C.P., Lee, T.Y., Yang, Y.C., 2006. Modelling climate-change impacts on stream temperature of Formosan landlocked salmon habitat. Hydrological Processes 20, 1629–1649.

Tzeng, C.S., 1999. Studies on Population Ecology of the Formosan Landlocked salmon Oncorhynchus masou formosanus (II). Construction and Planning Agency, Ministry of Interior, SHEI–PA National Park Management Office, Taiwan (in Chinese).

Yang, C.H., 1997. The effect of the water temperature on Taiwan landlocked salmon (Oncorhynchus masou formosanus) in Chichiawan stream Basin. M.S. thesis, Department of Life Science, National Tsing Hua University, Taiwan (in Chinese).

Nakicenovic, N., et al., 2000. Special Report on Emissions Scenarios: a Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, U.K., 599 pp.