

DOI: 10.5846/stxb201212171813

冯永玖,陈新军,杨晓明,高峰.基于遗传算法的渔情预报 HSI 建模与智能优化.生态学报,2014,34(15):4333-4346.

Feng Yongjiu^{1,2,3}, Chen X J, Yang X M, Gao F. HSI modeling and intelligent optimization for fishing ground forecasts using a genetic algorithm. Acta Ecologica Sinica, 2014, 34(15): 4333-4346.

基于遗传算法的渔情预报 HSI 建模与智能优化

冯永玖^{1,2,3,4}, 陈新军^{1,2,3,4,*}, 杨晓明^{1,2,3}, 高峰^{1,2,3}

(1. 上海海洋大学海洋科学学院, 上海 201306; 2. 大洋渔业资源可持续开发省部共建教育部重点实验室, 上海 201306

3. 国家远洋渔业工程技术研究中心, 上海 201306; 4. 远洋渔业协同创新中心, 上海 201306)

摘要: 鱼类栖息地适宜性指数模型(HSI)基于鱼类分布与海洋环境之间存在的非线性关系而构建。然而,海洋环境因子之间存在着传统方法无法消除的相关性,导致获取的 HSI 参数较难准确表达环境因子与渔场之间的复杂关系。基于遗传算法(GA),自动消除海洋环境因子之间的相关性,构建了一种通用的鱼类 HSI 建模与智能优化框架(GeneHSI)。GeneHSI 框架的核心是 HSI 建模空间向遗传算法空间的映射以及 GA 适应度函数的构建。该函数构建的思想是 HSI 预测的渔场概率与商业捕捞获取的渔场概率之间的累计误差值达到最小化。GeneHSI 由待解问题构建、GA 初始化和 GA 优化策略 3 部分组成。利用随机生成的标准化海洋环境数据与渔场概率数据,验证了 GeneHSI 模型框架的有效性。研究表明, GeneHSI 能够有效优化 HSI 的建模并能自动获取 HSI 参数。不同限制条件下,遗传算法获取的 HSI 具有较大的差异,其中一般优化策略下获取的 HSI 参数最差;不等式、等式和上下界条件下, GeneHSI 优化过程显著地更加合理,因此获取的 HSI 参数也更准确。此外,100、1000、5000 和 10000 样本量下的优化建模表明, GeneHSI 具有处理海量样本数据的能力。

关键词: 渔情预报; 栖息地适宜性指数; 遗传算法; 智能优化; 模拟数据

HSI modeling and intelligent optimization for fishing ground forecasts using a genetic algorithm

FENG Yongjiu^{1,2,3,4}, CHEN Xinjun^{1,2,3,4,*}, YANG Xiaoming^{1,2,3}, GAO Feng^{1,2,3}

1 College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China

2 The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources (Shanghai Ocean University), Ministry of Education, Shanghai 201306, China

3 National Distant-water Fisheries Engineering Research Center (Shanghai Ocean University), Shanghai 201306, China

4 Collaborative Innovation Center for Distant-water Fisheries, Shanghai 201306, China

Abstract: The construction of a habitat suitability index (HSI) model is a crucial problem in fishing ground forecasts. In general, the HSI model is established by estimating the relationship between marine environmental factors and fishing ground probabilities. However, the environmental factors observed by remote sensing technology and commercial fishing investigations are usually highly correlated, and conventional methods such as the continued product model, minimum model, maximum model, arithmetic mean model, and geometric mean model cannot eliminate the harmful effects caused by the correlation of fishing data. As a result, it is difficult for them to capture the complex relations between environmental factors and fishing ground probabilities. Based on the widely used intelligent optimization method of genetic algorithms (GAs), this paper presents a general framework called GeneHSI for HSI modeling and intelligent optimization. Most importantly, the GeneHSI framework can remove the harmful effects of correlation, allowing the automatic retrieval and

基金项目: 国家自然科学基金(41276156); 国家发改委产业化专项(2159999); 上海市自然科学基金面上项目(13ZR1419300); 教育部高等学校博士学科点专项科研基金新教师类项目(20123104120002); 上海市一流学科水产学(A类)共同资助

收稿日期: 2012-12-17; 网络出版日期: 2014-03-03

* 通讯作者 Corresponding author. E-mail: xjchen@shou.edu.cn;

optimization of the HSI parameters. The core of GeneHSI modeling is the construction of a fitness function. This function was built by projecting the logistic regression-based HSI space to that of a GA, and is used to guide the optimization process of GeneHSI. Specifically, the fundamental concept of the projection is to minimize accumulative errors between the computed ground probabilities and the observed probabilities converted from commercial fishing data. The proposed GeneHSI framework is composed of three elements. These are the construction of the problem to be solved, the initialization of the GA, and the optimization strategy of the GA. The validation and effectiveness of the GeneHSI framework have been demonstrated using simulation data, that is, randomly generated normalized marine environmental factors and fishing ground probabilities range from 0 to 1. Research shows that the GeneHSI framework is effective and efficient in retrieving and optimizing HSI parameters for fishing ground forecasts. Because of the stochastic characteristics of GAs, however, there is a high requirement for modelers and scientists to better control the implementation of the GeneHSI framework. The HSI parameters retrieved by the GeneHSI framework vary under different constraints. Such constraints used in GAs commonly include linear inequalities and linear equality constraints on the underlying relations between marine environmental factors and fishing ground probabilities, as well as constraints on the bounds of HSI parameters. Compared with the results under optimization strategies using these constraints, the results under a general optimization strategy are inferior in that the GeneHSI framework cannot obtain a good match between the best-fitness and mean-fitness curves. In theory, the fitness value is the accumulative error of the GeneHSI model; hence, a smaller value indicates a better result. However, a good convergence process does not necessarily lead to a minimum fitness value amongst fitness functions under different constraints. In this paper, therefore, an evaluation of the convergence process, instead of a minimum fitness value, is considered the fundamental standard for the assessment of a good set of HSI parameters. In addition, experience and professional knowledge are required for an exact assessment of the HSI parameters. Overall, the above constraints, especially those on the parameter bounds, greatly help the optimization of the GeneHSI framework to retrieve better HSI parameters. In addition, the implementation of the GeneHSI framework with 100, 1 000, 5 000, and 10 000 samples demonstrates its strong capability for processing the mass data of fishing grounds. It is expected that the GeneHSI framework can enrich the modeling methods and theories of fishing grounds, and hence guide the application of intelligent optimization methods in fishing ground HSI modeling.

Key Words: fishing ground forecasts; habitat suitability index; genetic algorithm; intelligent optimization; simulation data

栖息地适宜性指数(HSI)模型最早由美国科学家于20世纪80年代提出,用来模拟生物体对周围栖息环境要素的反应^[1]。目前已广泛应用于物种分布与管理等领域^[2],并逐渐在海洋渔场的分析与预测中得到了应用^[3-5],如印度洋大眼金枪鱼^[6-8]、大西洋及太平洋的鲑鱼^[9-11]、以及鲈鱼、秋刀鱼等^[12]。从研究方法来看,主要有权重求和法、几何平均法^[6-7]、分位数回归方法^[13]和主成分分析法^[14]等。HSI的研究不仅是为了从理论上认知渔场分布与海洋环境要素的关系,也是为了向捕捞生产和渔业资源管理者提供信息参考。

海洋环境要素与中心渔场之间存在动态交互关系并构成一个复杂的系统,环境要素之间通常存在一定的相关性。经典数理统计方法在构建HSI模型

时,由于无法消除环境要素固有的多重相关性,分析预报的精度受到极大的限制。因此,如何建立高精度的渔情预报模型,降低预报的不确定性,便成为了鱼类HSI建模需要解决的关键科学问题。

从最小化中心渔场预测值与实测值之间的差异的角度出发,利用人工智能方法进行自动优化,则可望提高渔情预报的精度和降低预报的不确定性。作为一种著名的组合优化算法,遗传算法(GA)常用于求解复杂函数的极值^[15]。遗传算法源自对达尔文生物进化论的模拟计算,是一种通过模拟自然进化过程搜索最优解的方法,目前已成为一种成熟的进化智能算法^[16]。近20年来,遗传算法在众多领域得到了广泛应用,例如在复杂地理系统的建模中遗传算法被用于建立非线性元胞自动机模型^[17],与之

类似的组合优化算法粒子群优化算法和模拟退火算法也得到了较好的应用^[18-19]。鉴于此,本研究基于遗传算法构建一种能够获取和优化中心渔场 HSI 参数的智能建模框架 GeneHSI,并利用模拟数据进行该框架的执行测试。同时,验证 GeneHSI 模型框架在一般优化策略、不等式、等式和上下界限条件下对渔业模拟数据的建模优化,并测试了 GeneHSI 模型框架对于不同样本量数据的优化解算能力。该基于遗传算法的 GeneHSI 框架将有助于实现渔情预报 HSI 的优化建模,并可望引导智能方法在渔业 HSI 建模中的应用。

1 GeneHSI 模型构建

1.1 遗传算法简介

遗传算法(GA)是一种基于生物自然选择与遗传机理的随机搜索算法。与传统搜索算法不同,GA 从一组随机产生的“种群”(初始可行解)开始搜索过程^[15]。待解问题的每一个参数即是一个“基因”,一组基因构成了问题的一个可行解,解是种群中的个体,称为“染色体”。在遗传算法中,染色体被编码为一串符号(二进制字符串),这些字符串在在后续迭代中根据遗传规则不断进化^[16]。遗传算法在适应度函数的引导下进行迭代搜索,通过适应度值来判断染色体的优劣并进行优胜劣汰;在未达到算法终止条件时,通过选择、交叉和变异 3 个算子产生下一代染色体。在形成新一代过程中,根据适应度值的大小选择部分后代,淘汰部分较差后代同时保持种群大小为常数,并继续进行随机搜索和迭代,直到算法收敛从而得到最优解(最优染色体)^[15-16]。

1.2 HSI 与遗传算法空间的映射关系

以综合栖息地基本模型(如连乘模型、算术平均模型、几何平均模型、logistic 回归模型等)为基础,以实际作业换算的 HSI 为基准^[20],构建 GA 算法的适应度函数^[15]。利用遗传算法自动寻找该适应度函数的最小值,其意义为 GA 获取的参数能够使模型预测值与实际作业换算的 HSI 值之间的差异达到最小化^[16]。对应于最小适应度函数值的染色体,就是一组通过 GA 优化所得的栖息地指数模型的参数,而对应的“基因”则为各海洋渔业环境要素的权重。由这样一组染色体构建的 HSI 模型,即为 GA 优化的渔场预报模型。

1.3 GeneHSI 模型中 GA 的适应度函数与优化策略

(1)染色体编码与种群初始化 染色体即是一组参数的可行解,采用浮点数编码方法,定义染色体的长度为有效解的变量长度,定义染色体编码为 $a = \{a_0, a_1, \dots, a_j, \dots, a_m\}$, a_j 是 HSI 模型参数,即 GA 算法中的基因, m 表示基因个数。

(2)适应度函数 适应度函数是渔场预报建模向遗传算法映射的核心,只有构建适应度函数才能够理解遗传算法的本质,也才能够进行遗传算法的渔情预报 HSI 建模。基于我们前期对元胞自动机、模拟退火算法和粒子群算法的大量研究^[16-19],构建了渔情预报 HSI 建模的适应度函数。在遗传算法中,利用适应度函数 $F(a)$ 评价染色体(可行解)的优劣,规定 $F(a)$ 的值越小则对应的染色体越优:

$$F(a) = \sqrt{\sum_{i=1}^n (\text{HSI}_c - \text{HSI}_a)^2} / n \quad (1)$$

式中, HSI_c 表示通过海洋环境因素计算所得的渔场概率,即 $\text{HSI}_c = 1 / [1 + \exp(-(a_0 + a_1 x_1 + \dots + a_m x_m))]$, HSI_a 表示通过商业捕捞数据计算而得的渔场概率, n 表示样本点的数量。其中, HSI_c 的表达式是一种 logistic 回归方程式,虽然目前为止这种表达形式在渔业栖息地研究中并未发现,但是 logistic 回归在野生动物的栖息地建模中广泛使用,一些经典论文的引用频次更是达到数百次之多,这表明 logistic 回归在动物栖息地建模中理论上是成立的^[21-22]。

国际上利用 logistic 回归进行野生动物栖息地建模的经典案例很多,包括格雷厄姆山红松鼠、高温蚱蜢和布什蟋蟀、高海拔稀有物种以及鹿和喜马拉雅斑羚等^[22-25]。海洋鱼类是典型的水生野生动物,其栖息地模式与文献中野生动物非常接近,从生态学上可以判定 logistic 回归符合海洋鱼类。在一项关于鱼类种群判别的研究中,认为 logistic 回归比判别分析的效果要好,这直接证明了 logistic 回归可以用于渔业资源与渔场学领域^[26]。在此将 logistic 回归方程作为核心构建适应度函数,但是在 GeneHSI 建模中并不需要真正执行 logistic 回归,只需要借用这种在理论上成立的方程式,通过遗传算法自动获取 HSI 参数从而完成 HSI 的建模。

(3)种群的选择、交叉与变异 确定适当的选择算子、交叉算子、变异算子以及算法终止条件,作为 GA 算法的优化策略^[15-16]。

(4)最优染色体的获取 最后获取的最优染色

体即为需要确定的权重参数^[15-16],用于建立基于遗传算法的综合 HSI 模型框架 GeneHSI。

基于遗传算法的基本理论以及从渔场预报向遗传算法的映射关系,构建用于渔情预报 HSI 参数获取与优化的建模框架 GeneHSI(图 1)。GeneHSI 模

型框架通过 Matlab“Optimization Tool”中的 GA 工具进行构建。由于该 GA 工具主要针对连续函数,而渔业 HSI 的适应度函数却是一种遍历所有样本点的离散函数,因此需要将该离散函数编写成为 Matlab 代码供 GA 工具调用和执行^[16]。

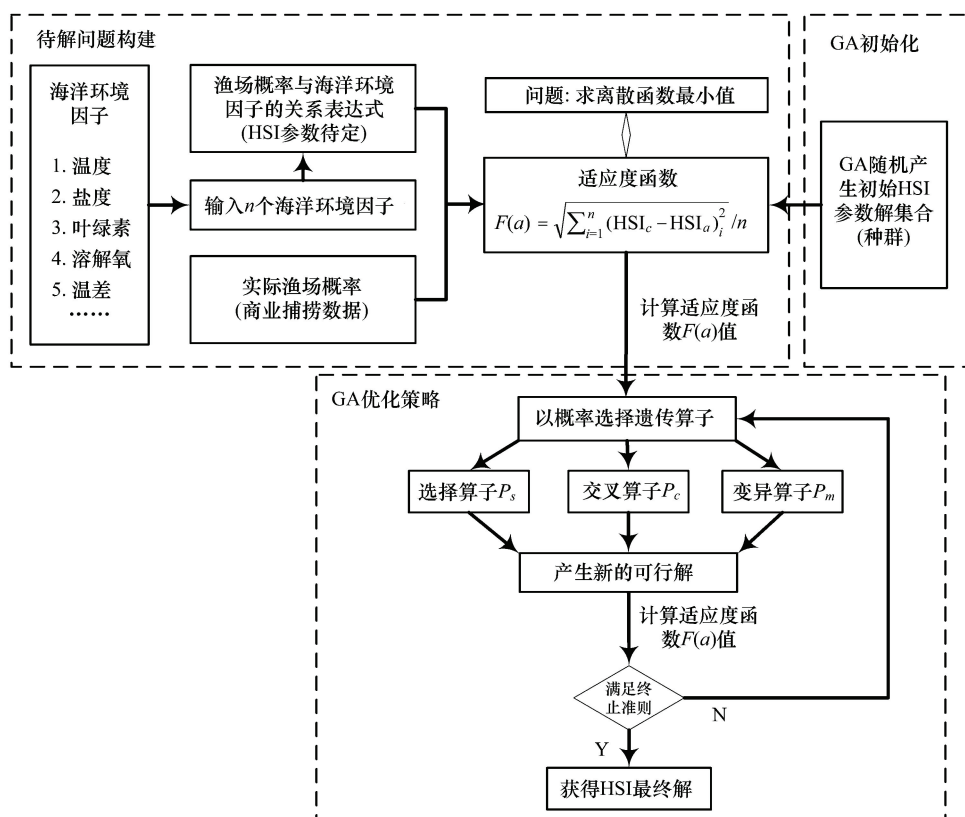


图 1 遗传算法用于渔情预报 HSI 参数获取与优化的建模框架

Fig.1 A general framework for retrieving and optimizing fishing forecast HSI parameters based on generic algorithm

图 1 表明 GeneHSI 由 3 个部分组成:(1)待解问题构建,即 HSI 模型向遗传算法空间的映射,通过海洋环境因子、商业捕捞数据、以及基于 logistic 回归的渔场概率方程构建遗传算法的适应度函数;(2)GA 初始化,即确定 GA 产生初始 HSI 参数的方案和 GA 初始种群数量等;(3)GA 优化策略,即确定适应度函数优化过程中的迭代方法,包括可行解判断方法、新解产生的选择、交叉和变异算子以及算法终止规则等。其中,选择概率 P_s 一般通过比例选择和保存策略等方法自动确定,交叉概率 P_c 一般选取 0.4—0.99 之间的值,而变异概率 P_m 一般选取 0.0001—0.1 之间的值^[15]。模型终止准则包括以下几个方面^[15-16]: 1) GeneHSI 最大计算时间(与计算机硬件关系较大),超出设定的时间则模型终止;2)最大迭代次数,达到该迭代次数则模型终止;3)适应度函数极限值,

当达到极限值则模型终止;4)两次计算所得的适应度函数值的差值(计算精度),当该差值小于某阈值则模型终止。

2 GeneHSI 框架的模拟数据应用与分析

2.1 渔业模拟数据

为验证该 GeneHSI 模型框架的正确性和有效性,更进一步探测其使用和控制方法,使用模拟数据进行测试。在模拟数据中,假设与渔场概率相关的海洋环境因素有温度、盐度、叶绿素、溶解氧、温差以及海面高度距平均值等。由于各种海洋环境因素的值域范围不一致,为了能够在同一个方程式中进行换算,将其进行归一化处理,所有环境因子的值域范围归一化到 $[0, 1]$ 之间。在实际操作中,利用 EXCEL 的随机数发生器 RAND() 产生模拟数据,

RAND()随机产生的数值均在[0—1]之间(表1)。10000等5种样本数据,用于检验 GeneHSI 模型的有效性以及模型对样本量的响应。

根据研究要求,分别产生了 50、100、1000、5000 和

表 1 归一化海洋环境因子与渔场概率的模拟数据

Table 1 Normalized simulated data for fishing ground probability and corresponding marine environmental factors							
数据项 No. of samples	海表温度 Sea surface temperature (SST)	盐度 Sea salinity (SS)	叶绿素 Chlorophyll-a (CHA)	溶解氧 Dissolved oxygen (DO)	温差 Temperature difference (TD)	海面高度 距平均值 Sea surface height anomaly (SSHA)	渔场概率 Fishing ground probability (FGP)
1	0.3003	0.2628	0.8899	0.2849	0.4754	0.8627	0.1919
2	0.6112	0.7253	0.5983	0.4180	0.9300	0.5017	0.7640
3	0.4074	0.8635	0.1323	0.5995	0.4017	0.7001	0.7629
4	0.2157	0.2779	0.2720	0.5612	0.2196	0.7543	0.6644
5	0.1506	0.4958	0.4862	0.3367	0.7677	0.9388	0.9665
6	0.7711	0.4506	0.4589	0.7334	0.1176	0.0071	0.9236
7	0.0386	0.6371	0.9703	0.0621	0.6235	0.6795	0.0886
8	0.1304	0.9272	0.6833	0.6298	0.6442	0.8488	0.8426
9	0.5302	0.1758	0.9313	0.2705	0.7852	0.8672	0.4537
10	0.5512	0.6660	0.3100	0.8085	0.3702	0.2512	0.0952
...	...	...	...	...	...	...	...

2.2 HSI 参数获取

在 GeneHSI 模型执行之前,首先必须确定模型的优化策略,即确定遗传算子与终止准则。本文使用轮盘赌法自动确定选择概率 P_s ,选取交叉概率 P_c 为 0.6,而变异概率 P_m 为 0.1。由于模型计算时间与适应度极值的影响较小,本研究设置模型终止准则为最大迭代次数 500 次和最优适应度函数差值(计算精度)为 $1e^{-6}$,达到其中任何一个条件则模型终止。在 GeneHSI 模型实际执行中,均由于计算精度达到 $1e^{-6}$ 而终止模型运行。根据 GeneHSI 模型以及模型优化策略,首先利用 50 个模拟数据样本点建立了适应度函数,从而获取 HSI 的参数。遗传算法在获取 HSI 参数的过程中,受到适应度函数的引导并进行优化,随着优化的推进适应度函数值不断降低,

最终达到最小值(图 2)。该最小值即为在最终获取的 HSI 参数下,预测渔场概率与真实渔场概率之间差异的最小值。

遗传算法对 HSI 参数进行计算与优化的适应度收敛过程。最佳适应度值收敛到 0.250918,平均适应度值收敛到 0.27452,根据适应度函数的构造理论,适应度值越小表明 HSI 参数导致的预测预报误差越小。最佳适应度值和平均适应度值均随迭代而减小,最佳适应度值在第 16 次迭代时收敛,但平均适应度值上下波动,未呈现收敛趋势(图 2)。直到迭代结束,最佳、最差(Worst)与平均适应度值之间的差异仍较大,表明 GeneHSI 在没有限制参数的条件下收敛性较差,结果并不理想(图 2)。因此,获取的 HSI 参数与真实情况存在一定的差异(表 2)。

表 2 一般优化策略下遗传算法获取的 HSI 参数

Table 2 HSI parameters retrieved using genetic algorithm under a general optimization strategy						
常数项 Constant	海表温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
-1.1925	-0.8240	0.2412	1.8855	0.9745	1.0139	-0.8059

2.3 限制条件及样本量对优化结果的影响

2.3.1 不等式限制条件

由于一般优化策略下获取的 HSI 参数并不理

想,因此需对 HSI 的参数范围进行限制,使 GeneHSI 在有限的范围内进行搜索,确保结果更加准确。遗传算法中,不等式、等式和参数范围限制条件较常

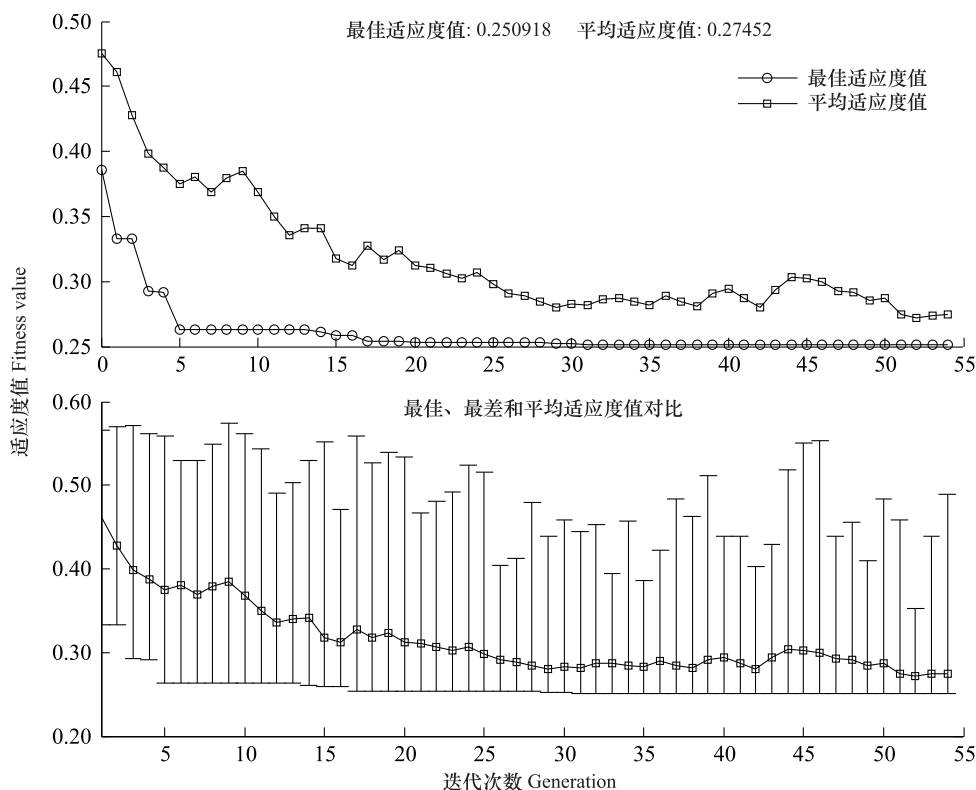


图2 一般优化策略下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.2 Fitness track and scores of the corresponding fitness of GeneHSI model under a general optimization strategy

用,其中不等式和等式限制条件指的是多种海洋环境因子(温度、盐度、叶绿素、溶解氧、温差以及海面高度距平均值)符合的某种关系,而参数范围限制条件指的是 HSI 参数上下界范围。一般地,不等式条件可以表达为:

$$A \times x \leq b \quad (2)$$

式中, A 和 b 是待定的限制条件,若 A 为矩阵,则 b 为向量;若 A 为向量,则 b 为数值; x 是海洋环境因子向量。在 GeneHSI 优化执行中, x 是既定的样本数据, A 和 b 的设置需经过理论探讨和经验分析加以确定,可以源于传统方法的计算结果、也可以源于对研究

对象和研究区域的认知。

为了探测不等式限制条件对 GeneHSI 结果的影响,设 $A=[1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$ 、 $b=5$ 进行 GeneHSI 模型执行,优化过程和 HSI 参数结果分别如图 3 和表 3。

图 3 表明,在不等式条件下最佳适应度值收敛到 0.256759,平均适应度值收敛到 0.264893。在迭代 30 次之后,平均与最佳适应度的收敛曲线较为接近,呈现较为明显的收敛状态(图 3)。当适应度函数曲线开始收敛,最佳与平均适应度值开始接近,而最差适与平均适应度值之间的差异仍较大(图 3)。

表3 不等式限制条件下遗传算法获取的 HSI 参数

Table 3 HSI parameters retrieved using genetic algorithm under linear inequalities

常数项 Constant	海面温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
-0.7064	-1.3827	-0.1444	1.0329	1.2298	0.7182	-0.0905

2.3.2 等式限制条件

与不等式限制条件类似,等式限制条件可以表达为:

$$Aeq \times x = beq \quad (3)$$

式中, Aeq 和 beq 是待定的限制条件,若 Aeq 为矩阵,则 beq 为向量;若 Aeq 为向量,则 beq 为数值; x 是海洋环境因子向量。与不等式条件类似, x 是既定的样本数据, Aeq 和 beq 通过理论和经验分析加以确定。

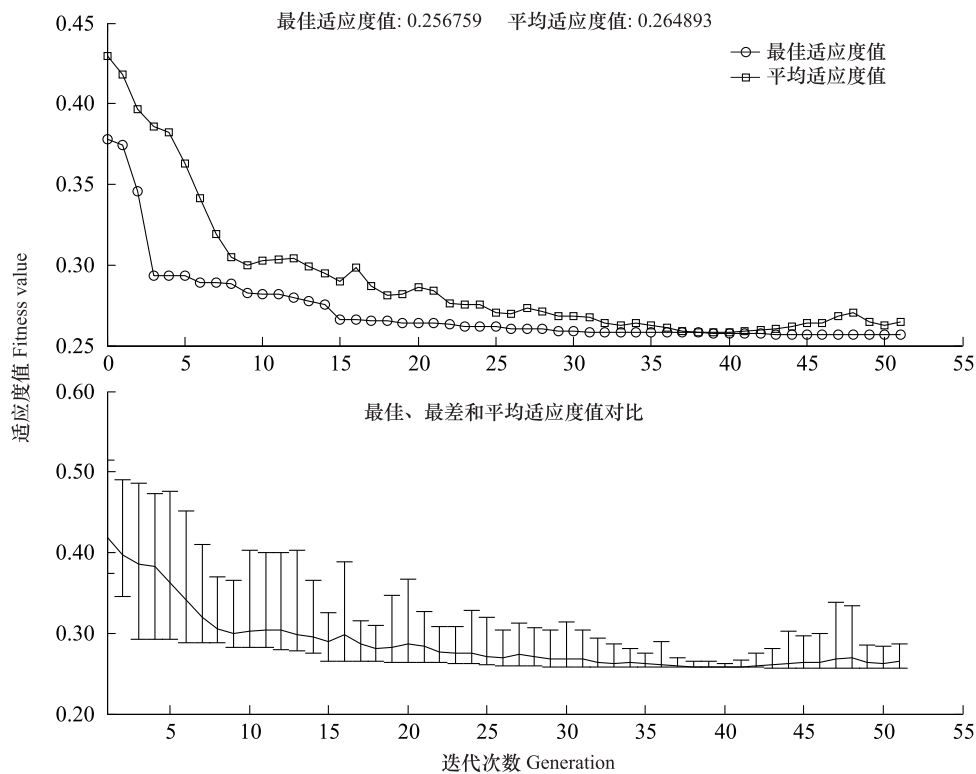


图3 不等式限制条件下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.3 Fitness track and scores of the corresponding fitness of GeneHSI model under linear inequalities

为探测等式限制条件对 GeneHSI 结果的影响,设置 $A_{eq} = [0 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$ 、 $b_{eq} = 3.5$ 进行优化。根据上述等式限制条件进行 GeneHSI 模型执行,其优化过程和 HSI 参数结果分别如图 4 和表 4。

等式限制条件下,最佳适应度值收敛到 0.284878,平均适应度值收敛到 0.284905。在迭代 25 次之后,平均与最佳适应度值完全重合,说明适应度函数曲线完全收敛,且效果较好。随着适应度函数开始收敛,最佳、最差和平均适应度值完全一致,表明 GeneHSI 优化效果非常理想(图 4)。

2.3.3 上下界限制

除了不等式和等式限制条件,在 GeneHSI 算法中也常用到上下界限限制条件。研究表明,传统 HSI 建模中获取的 HSI 参数值域均为 0 到 1 之间^[10-13],因此在 GeneHS 优化控制中可利用这个关键的信息对适应度函数进行引导。为此,设置 HSI

参数的上界为 $[1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$,而下界为 $[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$,优化过程如图 5。

从图 5 可知,上下界限限制条件下 GeneHSI 在迭代到 34 次时开始收敛,最佳适应度值收敛到 0.252479,平均适应度值收敛到 0.253145,收敛后最佳与平均适应度曲线重合较好。随着优化进程最差和最佳适应度值之间的范围不断缩小,开始收敛到最优适应度值,但是优化中最差适应度值仍有局部反弹迹象(图 5)。总体评估来看, GeneHSI 优化的效果相对较好,最终获取的 HSI 参数如表 5。

2.3.4 综合限制条件

为达到 GeneHSI 模型的最佳优化效果,通常同时使用不等式、等式和上下界 3 种限制条件,形成综合限制条件。图 6 显示了 3 种限制条件同时使用的适应度函数优化过程,具体限制条件与前述设置保持一致。

表 4 等式限制条件下遗传算法获取的 HSI 参数

Table 4 HSI parameters retrieved using genetic algorithm under linear equalities

常数项 Constant	海表温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
-1.3584	-0.4125	0.9702	0.1394	0.7752	1.7820	0.1041

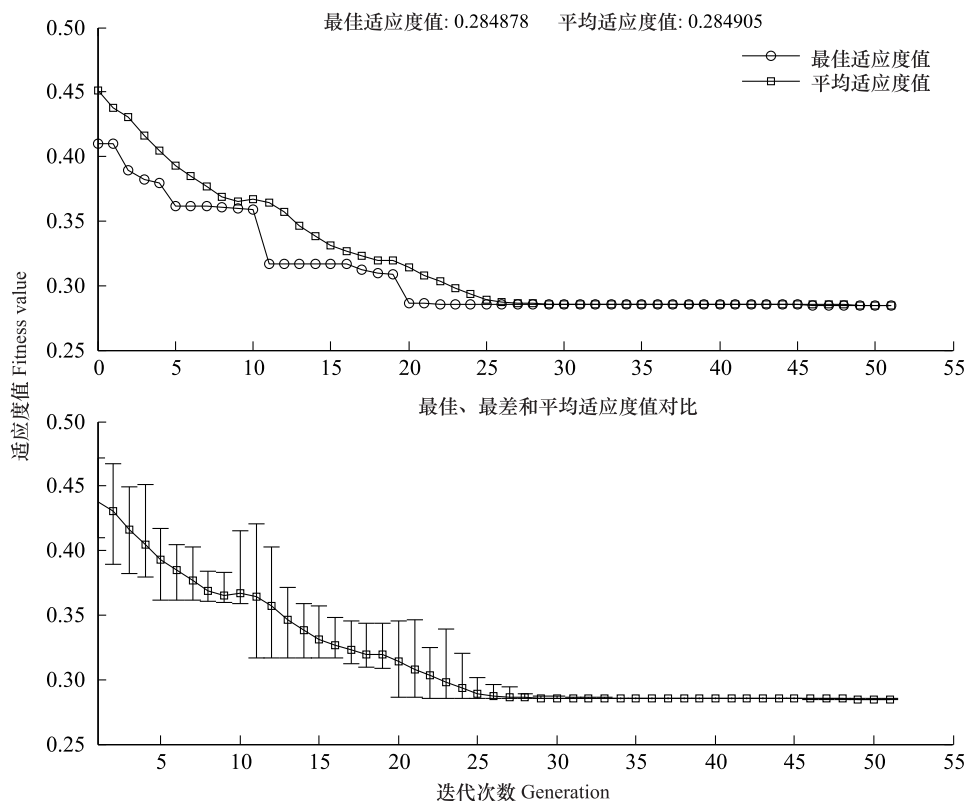


图 4 等式限制条件下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.4 Fitness track and scores of the corresponding fitness of GeneHSI model under linear equalities

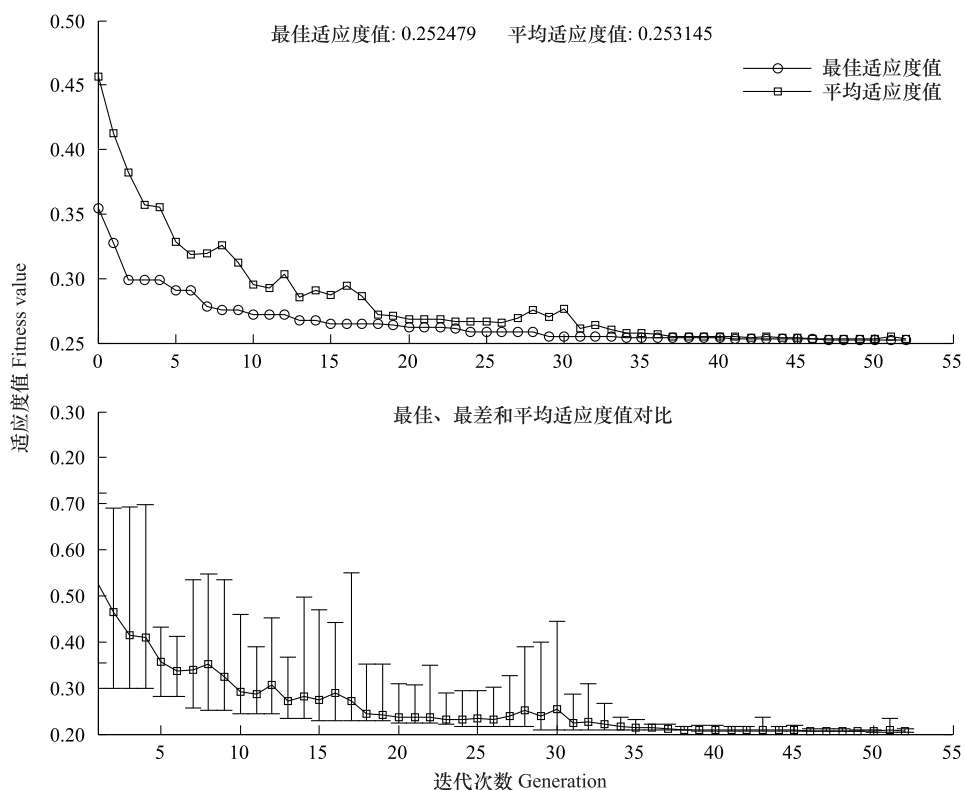


图 5 上下界限条件下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.5 Fitness track and scores of the corresponding fitness of GeneHSI model under parameters' bounds

表 5 上下界限条件下遗传算法获取的 HSI 参数

Table 5 HSI parameters retrieved using genetic algorithm under parameters' bounds						
常数项 Constant	海表温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
0.7356	0.8166	0.2455	0.9838	0.7269	0.9995	0.5973

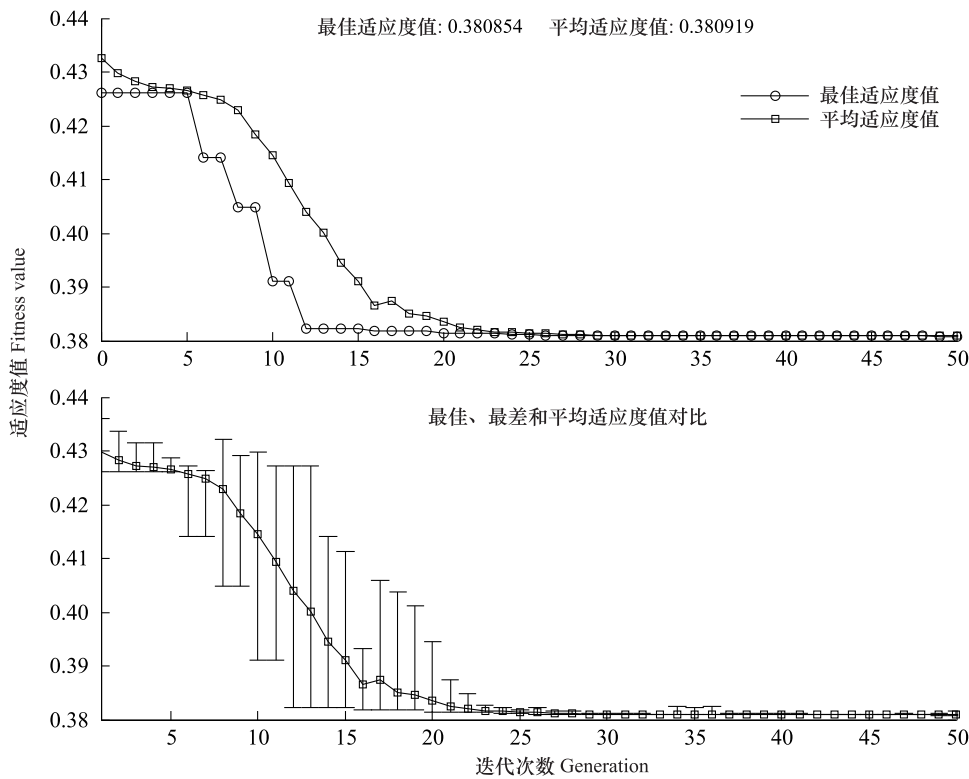


图 6 综合限制条件下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.6 Fitness track and scores of the corresponding fitness of GeneHSI model under comprehensive constraints

表 6 综合限制条件下遗传算法获取的 HSI 参数

Table 6 HSI parameters retrieved using genetic algorithm under comprehensive constraints						
常数项 Constant	海表温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
0.5077	0.7074	0.2709	0.5500	0.5668	0.6124	0.1986

综合限制条件下 GeneHSI 模型在迭代到 27 次时开始收敛,最佳适应度值收敛到 0.380855,平均适应度值收敛到 0.380919,收敛曲线表明优化效果非常理想。随着优化进程最差和最佳适应度值之间的范围收缩较快且非常彻底,收敛后曲线未发现任何波动(图 6),说明综合限制条件下 GeneHSI 取得了较佳的优化效果,获取的 HSI 参数如表 6。

2.4 样本量对优化结果的影响

除了各种限制条件,样本量对 GeneHSI 建模同样具有重要的影响。在渔场预报分析中,样本是通

过商业捕捞而产生的,因此对于研究区的特定鱼种,样本量并不是研究者主观决定的。但是对于一种智能建模框架而言,应该具有处理大样本的能力。因此,本文进一步分析 GeneHSI 模型对于各种样本量的处理能力,并分析优化过程与优化结果对于不同样本量的响应。为此,产生了 100、1000、5000 和 10000 四种样本量,样本产生的方法如 3.1 节所述。选择在综合限制条件下进行 GeneHSI 模型执行,具体限制参数和前述设置一致,最终模型优化过程如图 7。

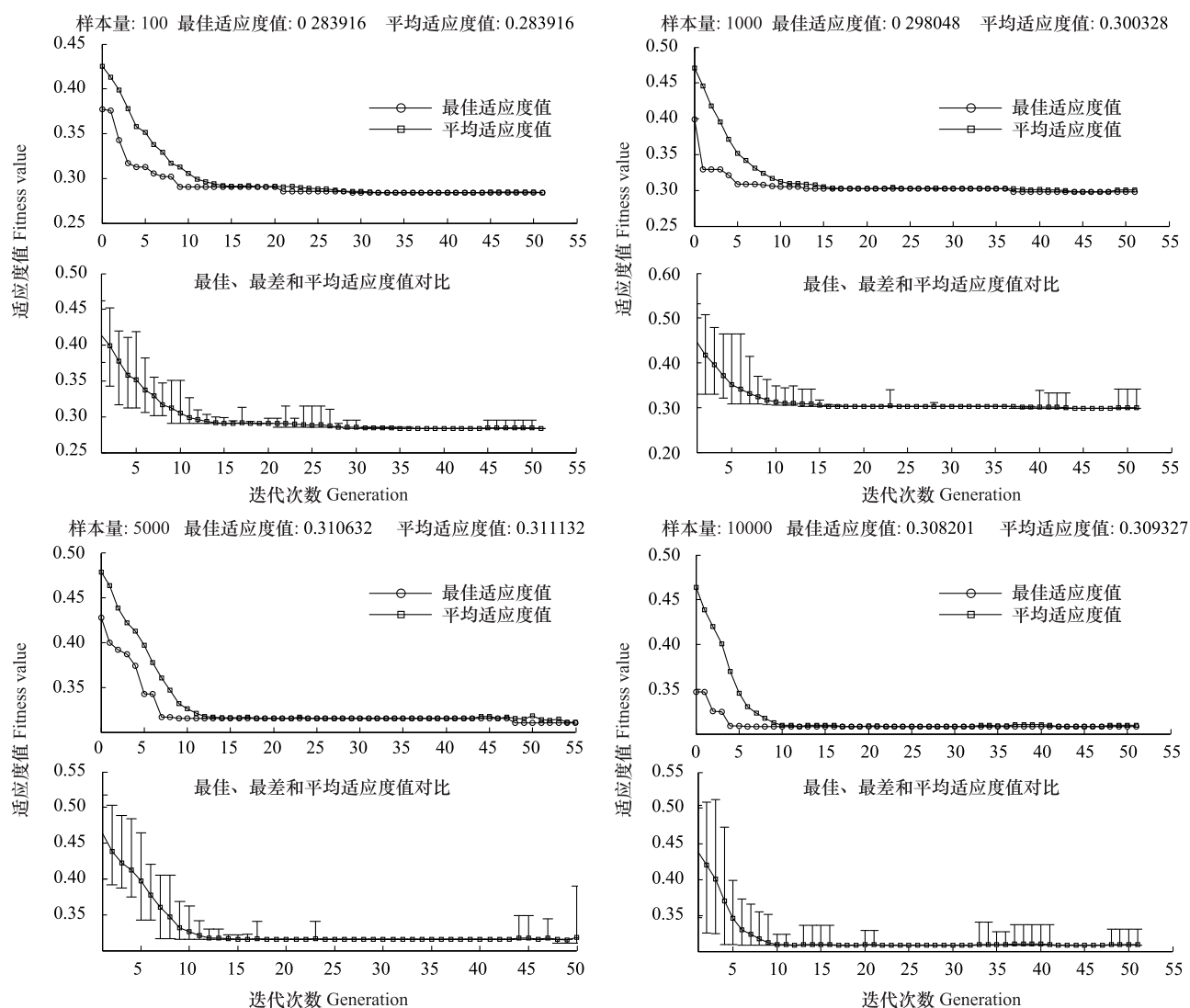


图 7 不同样本量下利用遗传算法获取与优化 HSI 参数的收敛过程

Fig.7 Fitness tracks and scores of the corresponding fitness of GeneHSI model under different samples

图 7 表明各种样本量在 GeneHSI 模型中均得到了优化处理,其中 100 样本量收敛于第 21 次迭代,最佳适应度值 0.283916,平均适应度值 0.283916;1000 样本量收敛于第 16 次迭代,最佳适应度值 0.288048,平均适应度值 0.300328;5000 样本量收敛于第 13 次迭代,最佳适应度值 0.310632,平均适应度值 0.311132;10000 样本量收敛于第 11 次迭代,最佳适应度值 0.308201,平均适应度值 0.309327。模型优化过程与收敛曲线显示,10000 样本量收敛最快,收敛效果最好;1000 和 5000 样本量收敛效果次之;而 100 样本量收敛的效果较晚,完全收敛出现在第 30 次迭代。4 种样本量在收敛之后,最差与平均适应度值之间存在局部的波动,其中大样本量 10000 的波动最为严重,而小样本量 100 的波动最小。优

化结果表明, GeneHSI 模型具有处理各种样本量的能力,同时对于样本量的响应也不尽一致,各种样本量下 GeneHSI 获取的参数如表 7 所示。

3 讨论

3.1 GeneHSI 应用效果与模拟数据测试情况

遗传算法用于渔场渔情预报,其核心思想在于可以使计算渔场概率值与真实渔场概率值的累计误差达到最小化。研究表明,遗传算法本质上是一种具有智能特性的随机优化算法^[15],能够行之有效地效优化鱼类 HSI 的建模并获取 HSI 参数。由于遗传算法的随机特性^[15],在对它进行控制并引导算法有效执行的过程中,对建模者有较高的技术要求(如对 Matlab 工具的熟练程度和 Matlab 代码编写的能力),

表 7 不同样本量下遗传算法获取的 HSI 参数

Table 7 HSI parameters retrieved using genetic algorithm under different samples

样本数 Number of samples	常数项 Constant	海表温度 SST	盐度 SS	叶绿素 CHA	溶解氧 DO	温差 TD	海面高度距平均值 SSHA
100	0.0413	0.0703	0.0134	0.0268	0.1669	0.0854	0.0010
1000	0.1316	0.0180	0.0273	0.0097	0.0631	0.1302	0.0512
5000	0.0605	0.1808	0.2788	0.0830	0.1148	0.0926	0.0229
10000	0.0999	0.1191	0.0142	0.0024	0.4149	0.0141	0.0524

需要研究者对研究区域与研究对象有较高的熟悉程度^[16-17]。本研究所采用的数据是随机产生的模拟数据,模拟数据的作用和关键在于:(1)模拟数据比之真实数据,其数据中存在的规律性要差,因此如果 GeneHSI 能在该模拟数据上得到很好的应用,那么证明 GeneHSI 模型具有很强的普适性,当应用于规律性更强的渔业真实数据时可望得到更好的结果;(2)遗传需要较强的人工干预与良好的参数控制经验^[16-17],模拟数据有利于从多角度探讨 GeneHSI 算法的控制与使用方法。这些使用控制方法为真实渔业数据的建模应用奠定了技术基础。

3.2 几种限制条件下 GeneHSI 优化效果的比较

与一般优化策略下的 GeneHSI 比较,不等式条件下最佳适应度值要大、而平均适应度值要低,表明整体优化结果稍好。但是不等式条件下的 GeneHSI 优化效果并未达到最佳状态,表明在不等式条件下虽然效果优于一般优化策略,因此仍需对算法进行深入的优化控制。与一般优化策略和不等式优化策略比较,等式限制条件下 GeneHSI 参数的获取效果更好,优化曲线收敛更加彻底。

与一般优化策略和不等式优化策略比较,上下界限条件下 GeneHSI 模型取得了更好的收敛效果;但比之等式条件,上下界条件下优化曲线的收敛情况却较差,表现为收敛较晚且最佳适应度曲线与平均适应度曲线重合性较差,但是却获得了更小的最佳和平均适应度值。综合限制条件与几种单项优化限制条件比较, GeneHSI 模型取得了更好的收敛效果,适应度函数的收敛更早、且最佳与平均适应度值重合得非常一致。

总体来看,在不同的限制条件下,遗传算法获取的 HSI 具有较大的差异,其中一般优化策略下获取的 HSI 参数最差,加以限制的不等式、等式和上下界条件下其优化过程显著地更加合理,因此获取的 HSI 参数也更准确。理论上,适应度函数值是渔场概率

计算值与真实值的累计误差^[15-17],但是函数优化收敛过程较好时,最终对应的适应度值并不是所有限制条件下可能得到的最低值。因此,评价 HSI 参数结果是否合理和准确,并不是以适应度值的高低为唯一标准的^[18-19];应当同时考虑到适应度函数的优化收敛过程,并结合对研究区渔场的专业知识进行判断。综合来看,对 GeneHSI 的优化过程加以限制,尤其是上下界等限制条件的使用将使 HSI 参数优化结果更加合理^[8-10]。

3.3 GeneHSI 处理样本数据的效率

针对于渔场预报 HSI 建模,增加样本量并不会显著增加遗传算法需要迭代的次数,也不会增加遗传算法的实际解算时间。这与地理复杂系统中的建模有所不同,后者会显著增加遗传算法的迭代次数和解算时间^[18-19]。这可能由两个原因造成的:本文使用的海洋环境因子相对较少,或渔业 HSI 建模与经典地理系统建模比较数据复杂程度稍低,因此遗传算法解算起来效率更高。

在对 GeneHSI 有效控制的前提下,大样本量反而有可能使遗传算法更早地收敛于合理的适应度值^[15]。研究表明,大样本量得到的适应度值一般会大于小样本量得到的适应度值^[16-17]。但是这种情况并不是绝对的,当用于建模的数据规律性较强时,大样本量反而能够得到更小的适应度值^[15]。在真实渔情预报中,受到商业捕捞数据的限制,一般很少能够获取到大样本量数据^[6-11],这进一步表明了 GeneHSI 能够高效地对渔场预报进行优化处理。

3.4 GeneHSI 与传统 HSI 模型的比较

传统 HSI 模型在渔业科学中应用广泛,主要用于栖息地质量评估、资源量估算、分析选择渔场和渔情预报等,但各自有不同的特点。针对不同目的和研究区,各种模型的适用性和效果则不同。与传统 HSI 模型比较, GeneHSI 模型存在诸多不同之处(表 8)。

表 8 GeneHSI 模型与传统 HSI 模型的比较

Table 8 A comparison between the GeneHSI model and conventional HSI models

模型 Models	单变量 Single variable	多变量 Multi-variable	变量相关性 Correlation of variables	复杂性-智能化 Complexity and intellectualization	参数意义明确 Explicit meaning of parameters	模型特点 Model features
连乘法	√	×	不考虑	低-低	×	计算结果保守,对零值敏感
最小值法	√	×	不考虑	低-低	×	估计结果保守,决定于最小 SI (suitability index) 因子
最大值法	√	×	不考虑	低-低	×	估计结果乐观,决定于最大 SI 因子
算术平均法	√	×	不考虑	低-低	×	估计结果折中,不受 SI 极值的影响,未考虑单因素 SI 的权重
几何平均法	√	×	不考虑	低-低	×	估计结果折中,效果较算术平均法差,参数越少则效果越优,受 SI 零值影响较大
分位数回归法	√	×	不考虑	中-低	×	能提供多种不同分位数的 SI 估计结果,阐释因变量的整体分配,结合前述几种模型计算最终 HSI 值,估计结果的保守或乐观程度决定于 HSI 计算模型的选取
遗传算法	√	√	自动消除	高-高	√	同时考虑多种海洋环境因素、对因素的权重进行自动赋权与智能优化,估计结果的保守或乐观程度决定于 GA 限制参数

关于连乘法、最小/最大值法、算术/几何平均法和分位数回归法的“模型特点”参见文献^[27]

研究表明,经典的相关和回归分析方法并不适合于生态学领域的相关因素关系研究^[27]。因此,渔业 HSI 建模一般地先采用单因子构建 SI 函数,再通过几何平均法^[6-7]、分位数回归方法^[13]和主成分分析法^[14]等计算 HSI 值。与表 8 中的连乘法等 6 种模型不同,GeneHSI 模型通过适应度函数直接搜索、优化并获取海洋环境因素的权重,进而直接计算得到 HSI 值。传统渔业 HSI 的应用研究实际上将各种不同类型的因素进行等值赋权,虽然这些方法都是经过渔业专家知识而获得的,但显然带有主观判断的成分^[27]。在等值赋权中,海洋环境因子的相关性通常被忽略,而这种相关性将对 HSI 的建模具有负面的影响。GeneHSI 模型同时考虑海表温度、盐度、叶绿素、溶解氧、温差和海面高度距平均值等不同类型的海洋环境因素,在遗传优化中自动消除这些因素相关性的负面影响。这表明,GeneHSI 模型是一个典型的多变量 HSI 模型,且 HSI 参数的含义就是海洋环境因素的权重,而传统 HSI 模型是对 SI 曲线的一种拟合,其参数不具备物理意义。

比较而言,传统 HSI 模型的复杂性较低,因此其建模难度和智能化程度较低^[6-7, 27];而 GeneHSI 模型复杂性和建模难度均较高,但其具有高度的智能化程度。GeneHSI 模型估计结果的保守或乐观程度决

定于 GA 限制参数,在应用方面较适合于栖息地质量评估、渔场分析和渔情预报等。

4 结论

基于广泛应用于各领域的遗传算法,本文提出了一种渔情预报 HSI 参数获取与智能优化的方法框架 GeneHSI,并对该模型框架的使用和参数控制进行了分析与讨论。本研究的目的在于提供一种基于遗传算法的渔业 HSI 建模框架和行之有效的智能建模方法与思路,因此侧重于模型建立的整体框架、模型建立的思路以及模型执行过程中的人为控制。

基于遗传算法的 HSI 建模有助于发展现有的渔情预报的理论和方法。但在以后的研究中,应将该 GeneHSI 模型框架应用到实际的渔情预报分析中,同时通过实际的应用与反馈来优化 GeneHSI 模型。

References:

[1] U. S. Fish and Wildlife Service. Habitat Evaluation Procedures (HEP). Washington; U. S. Fish and Wildlife Service, 1980.

[2] U. S. Fish and Wildlife Service. Standards for the development of habitat suitability index models [R]. U. S. Fish and Wildlife Service, 1981; 1-81.

[3] Store R, Jokimäki J. A GIS-based multi-scale approach to habitat

- suitability modeling. *Ecological Modelling*, 2003, 169(1): 1-15.
- [4] Jin L R, Sun K P, He H S, Zhou Y F. Research advances in habitat suitability index model. *Chinese Journal of Ecology*, 2008, 27(5): 841-846.
- [5] Zielinski W J, Dunk J R, Yaeger J S, LaPlante D W. Developing and testing a landscape-scale habitat suitability model for fisher (*Martes pennanti*) in forests of interior northern California. *Forest Ecology and Management*, 2010, 260(9): 1579-1591.
- [6] Fan W, Cui X S, Shen X Q. Progress in fishing-ground analysis and fishing condition forecasting. *Journal of Fisheries of China*, 2005, 29(5): 706-710.
- [7] Chen X J, Feng B, Xu L X. A comparative study on habitat suitability index of bigeye tuna, *Thunnus obesus* in the Indian Ocean. *Journal of Fishery Sciences of China*, 2008, 15(2): 269-278.
- [8] Fan J T, Chen X J, Qian W G, Liu B L. Distribution of fishing ground of *Thunnus Alalunnus* and its relationship with sea surface temperature in the waters around Vanuatu. *Transactions of Oceanology and Limnology*, 2011, (1): 71-78.
- [9] Song L M, Zhou Y G. Developing an integrated habitat index for bigeye tuna (*Thunnus obesus*) in the Indian Ocean based on longline fisheries data. *Fisheries Research*, 2010, 105(2): 63-74.
- [10] Chen F, Chen X J, Liu B L, Qian W G, Tian S Q. Relationship between fishing ground of *Ommastrephes bartramii* and vertical temperature structure in the northwestern Pacific Ocean. *Journal of Shanghai Ocean University*, 2010, 19(4): 495-504.
- [11] Chen X, Tian S, Chen Y, Liu B. A modeling approach to identify optimal habitat and suitable fishing grounds for neon flying squid (*Ommastrephes bartramii*) in the Northwest Pacific Ocean. *Fishery Bulletin*, 2010, 108(1): 1-14.
- [12] Tian S Q, Chen X J, Chen Y, Xu L X, Dai X J. Evaluating habitat suitability indices derived from CPUE and fishing effort data for *Ommastrephes bartramii* in the northwestern Pacific Ocean. *Fisheries Research*, 2009, 95(2/3): 181-188.
- [13] Li G, Chen X J, Guan W J. Stock assessment and risk analysis of management strategies for *Scomber japonicus* in the East China Sea and Yellow Sea using a Bayesian approach. *Journal of Fisheries of China*, 2010, 34(5): 740-750.
- [14] Feng B, Tian S, Chen X J. The habitat suitability index of *Illex argentinus* by using quartile regression method in the Southwest Atlantic. *Transactions of Oceanology and Limnology*, 2010, (1): 15-22.
- [15] Hu Z M, Chen X J, Zhou Y Q, Qian W G, Liu B L. Forecasting fishing ground of *Dosidicus gigas* based on habitat suitability index off Peru. *Acta Oceanologica Sinica*, 2010, 32(5): 67-75.
- [16] Xuan G N, Cheng R W. Genetic Algorithm and the Engineering Optimization. Yu X J, Zhou G G, translate. Beijing: Press of Tsinghua University, 2004.
- [17] Feng Y J, Liu Y, Han Z. Land use simulation and landscape assessment by using genetic algorithm based on cellular automata under different sampling schemes. *Chinese Journal of Applied Ecology*, 2011, 22(4): 957-963.
- [18] Feng Y, Liu Y. An optimised cellular automata model based on adaptive genetic algorithm for urban growth simulation // Yeh A, Shi J W, Leung Y, Zhou C, eds. *Advances in Spatial Data Handling and GIS, Lecture Notes in Geoinformation and Cartography*. Berlin, Heidelberg: Springer-Verlag, 2012: 27-38.
- [19] Feng Y J, Liu Y. A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing. *International Journal of Geographical Information Science*, 2013, 27(3): 449-466.
- [20] Feng Y J, Liu Y, Tong X H, Liu M L, Deng S S. Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape and Urban Planning*, 2011, 102(3): 188-196.
- [21] Tian S Q, Chen X J. Impacts of different calculating methods for nominal CPUE on CPUE standardization. *Journal of Shanghai Ocean University*, 2010, 19(2): 240-245.
- [22] Keating K A, Cherry S. Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management*, 2004, 68(4): 774-789.
- [23] Pereira J M C, Itami R M. GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing*, 1991, 57(11): 1475-1486.
- [24] Hein S, Voss J, Poethke H J, Boris S. Habitat suitability models for the conservation of thermophilic grasshoppers and bush crickets-simple or complex? *Journal of Insect Conservation*, 2007, 11(3): 221-240.
- [25] Singh N J, Yoccoz N G, Bhatnagar Y V, Fox J L. Using habitat suitability models to sample rare species in high-altitude ecosystems: a case study with Tibetan argali. *Biodiversity and Conservation*, 2009, 18(11): 2893-2908.
- [26] Singh A, Kushwaha S P S. Refining logistic regression models for wildlife habitat suitability modeling: A case study with muntjak and goral in the Central Himalayas, India. *Ecological Modelling*, 2011, 222(8): 1354-1366.
- [27] Prager M H, Fabrizio M C. Comparison of logistic regression and discriminant analyses for stock identification of anadromous fish, with application to striped bass (*Morone saxatilis*) and American shad (*Alosa sapidissima*). *Canadian Journal of Fisheries and Aquatic Sciences*, 1990, 47(8): 1570-1577.
- [28] Gong C, Chen X J, Gao F, Guan W J, Lei L. Review on habitat suitability index in fishery science. *Journal of Shanghai Ocean University*, 2011, 20(2): 260-269.

参考文献:

- [3] 金龙如, 孙克萍, 贺红士, 周宇飞. 生境适宜度指数模型研究

- 进展. 生态学杂志, 2008, 27(5): 841-846.
- [5] 樊伟, 崔雪森, 沈新强. 渔场渔情分析预报的研究及其进展. 水产学报, 2005, 29(5): 706-710.
- [6] 陈新军, 冯波, 许柳雄. 印度洋大眼金枪鱼栖息地指数研究及其比较. 中国水产科学, 2008, 15(2): 269-278.
- [7] 范江涛, 陈新军, 钱卫国, 刘必林. 瓦努阿图周边海域长鳍金枪鱼渔场分布及其与表温关系. 海洋湖沼通报, 2011, (1): 71-78.
- [9] 陈峰, 陈新军, 刘必林, 钱卫国, 田思泉. 西北太平洋柔鱼渔场与水温垂直结构关系. 上海海洋大学学报, 2010, 19(4): 495-504.
- [12] 李纲, 陈新军, 官文江. 基于贝叶斯方法的东、黄海鲈资源评估及管理策略风险分析. 水产学报, 2010, 34(5): 740-750.
- [13] 冯波, 田思泉, 陈新军. 基于分位数回归的西南太平洋阿根廷滑柔鱼栖息地模型研究. 海洋湖沼通报, 2010, (1): 15-22.
- [14] 胡振明, 陈新军, 周应祺, 钱卫国, 刘必林. 利用栖息地适宜指数分析秘鲁外海茎柔鱼渔场分布. 海洋学报, 2010, 32(5): 67-75.
- [15] 玄光男, 程润伟. 遗传算法与工程优化. 于歆杰, 周根贵, 译. 北京: 清华大学出版社, 2004.
- [16] 冯永玖, 刘艳, 韩震. 不同样本方案下遗传元胞自动机的土地利用模拟及景观评价. 应用生态学报, 2011, 22(4): 957-963.
- [20] 田思泉, 陈新军. 不同名义 CPUE 算法对 CPUE 标准化的影响. 上海海洋大学学报, 2010, 19(2): 240-245.
- [27] 龚彩霞, 陈新军, 高峰, 官文江, 雷林. 栖息地适宜性指数在渔业科学中的应用进展. 上海海洋大学学报, 2011, 20(2): 260-269.