

鳊塘水体生态因子的计算机模式识别法应用

赖子尼 吴淑勤 石存斌

(中国水产科学研究院珠江水产研究所, 广州 510380)

余煜棉 董奋强

(广东工业大学, 广州 510090)

钟邓文 劳启宁 吴柳平 卫东

(南海市水产养殖场, 528211)

摘 要 应用计算机模式识别技术研究鳊塘水体中溶氧量与九项生态因子间的相互关系。研究表明, 不同塘或不同时期其生态因子差异较大。在高维空间中, 高、低溶氧量两类样品点聚集在不同的区域, 周界清晰、分类十分成功, 表明池塘水体溶氧量与水中多项生态因子间存在显著规律性。用模式识别优化技术求得维持稳定的高溶氧状态的 9 项生态指标值, 优化生态因子的溶氧预报值大于 8.4mg/L。水体溶氧量与其它九项因子成多元的二次函数关系, 其模型复相关系数 $R=0.97$ 。对测量点溶氧量的回代准确率为 93.9%。

关键词 鳊, 池塘, 生态因子, 计算机模式识别

Application of the ecosystem factors in pond waters for *Siniperca chuatsi* by computer pattern recognition

Lai Zini, Wu Shuqing, Shi Cunbin

(Pearl River Fisheries Research Institute, CAFS, Guangzhou 510380)

Yu Yumian, Dong Fenqiang

(Guangdong Industry University, Guangzhou 510090)

Zhong Dengwen, Lao Qining, Wu Liuping, Xuan Weidong

(Nanhai City Fresh Water Culture Farm, 285211)

ABSTRACT The computer pattern recognition (CPR) technology is applied to the study on the relationships among dissolved oxygen (DO) concentration and nine ecosystem factors in the *Siniperca chuatsi* ponds. The factors differ greatly in distinct ponds or periods. In high-dimensional space, it is clear and successful that two kinds of samples, high and low DO, gather themselves in different regions. It indicates markedly regular relationships between DO and ecosystem factors. The nine superiority ecosystem targets for maintain high DO are sought by CPR favorable technology. In water, DO and nine ecosystem factors follow polybasic quadratic function relation, $R=0.97$. Using the mathematical model prediction DO, precision is 93.9%, predicting the DO is over 8.4 mg/L for favorable ecosystem factor.

KEYWORDS *Siniperca chuatsi*, pond, ecosystem factor, computer pattern recognition

“九五”农业部渔业重点科研资助项目(鳊鱼暴发性传染病防治研究), 95-A-96-03-01 号和南海市科技局资助项目(鳊鱼养殖水生生态平衡调节与疾病控制), 97-01 号。

第一作者简介: 赖子尼, 女, 1964 年 2 月生, Tel: 020-81510887

收稿日期: 1999-05-06

健康养殖必须保证池塘生态系统良好的平衡与稳定,而掌握池塘水体中各生态因子的关系是调控生态系统达到并维持在理想状态的基础。池塘生态系统是一个复杂的、多因素的、非典型分布的动态系统,用传统的单因素统计法往往难以有效地解决池塘的养殖研究^[1]。因此需要运用一种研究多因素问题的方法^[2-5]。灰色理论是其中的方法之一,多用于产量评估方面。而本文则首次用计算机模式识别法(Computer Pattern Recognition, CPR)研究鳊塘各种生态因子之间的关系^[6,7],在鳊塘的水体中选择了溶氧量(DO)为目标函数,以水温、透明度、pH、亚硝酸盐氮、铵氮、磷酸盐磷、化学耗氧量(COD)、叶绿素 a、异氧细菌数等九项因素为自变量,研究了高溶氧和低溶氧两类不同样本的聚类规律,确定体系的数学模型为多元二次非线性函数,模型的复相关系数达 0.97,模型对溶氧量的回代准确率达 93.9%。文中用 CPR 逆运算方法求得维持池塘高溶氧状态时九个生态因子的优化组合推荐值(对应溶氧量的预报值都不小于 8.4)。本文将为研究复杂的池塘生态系统提供了有效的方法。

1 调查方法

1998年,在广东南海市西樵、九江、新会荷塘建立三个实验点,选择 10 口鳊鱼生长正常的池塘为取样塘,水深约 1.2~ 1.5m,5~ 7d 投喂一批饵料鱼(鳊),通常情况下不注换新水(半封闭精养)。在养殖前、中、后期每口塘各取样一次,共取样 28 次。取样时间为上午 11:00~ 11:30,用有机玻璃采水器采集离水面约 30cm 处水样。九个生态因子的测定按常规分析方法^[8-10]。

2 计算机模式识别法

计算机模式识别是近年广泛应用在多因素问题的决策和预报上的多元统计技术。它利用 m 个自变量构成 m 维的模式空间,每一个样本点的 m 个变量值决定了该样本点在 m 维模式空间的位置。CPR 技术利用“物以类聚”的原则,把性质相近的样本点聚集在某一区域,而性质相异的另一类样本点聚集在另一区域,通过对这些样本点的空间区域的划分及其属性出发,研究多因素体系的数据特征的一种非函数方法。人类自身是无法识别三维以上的空间样本点的分布,因此 CPR 技术借助计算机降维技术,将 m 维空间的样本点分布状态在尽可能不失真的情况降至人类能判别的二维、三维空间上,不是通过推求变量间的严格的数学关系,而是利用一定的数学统计手段研究样本点在低维(3 维以下)空间的分布、聚类特点,进而建立识别样本聚类规律的判别模型,对待测样本的性质进行预报的图形识别新技术。本文利用 CPR 技术研究水温等九个生态因子总体对水体 DO 影响的规律,求出体系对应的数学模型和生态因子的优化值。

3 结果与讨论

3.1 目标因子的选取

测量了水体中水温等十项生态因子(其中 DO 为目标函数),并计算了各因子间的相关系数矩阵。结果表明,COD 与水温、叶绿素 a 为显著正相关,与透明度、亚硝酸盐氮为显著负相关;叶绿素 a 与水温、COD 成非常显著正相关,与透明度、亚硝酸盐氮接近为显著负相关;DO 与 pH 值呈非常显著正相关,与叶绿素 a 接近显著相关。其它因子间相关没有明显统计学意义(矩阵从略)。如果仅从与其它因子关系的密切程度来确定目标函数似乎应选取 COD 或叶绿素 a。但考虑氧是一切生物赖以生存的基础,DO 与养殖鱼类的生长速度和发病多寡有密切关系,且溶氧充足也是池塘形成良好微生态循环系统的重要标志,而 COD 或叶绿素 a 对养殖鱼类的影响则是间接的,所以本文还是选择 DO(x_{10})为目标函数(Y)研究塘水中生态因子间关系。

3.2 鳊塘生态因子的测量值

表 1 是十口塘十项生态因子 28 次测量的情况。表中各参数统计结果表明只有水温和 pH 值的变

化较小,全距幅度不大,各测量值的相对偏差小于8%,而余下的八个因子各次测量值变化较大,全距幅度大,测量值的相对偏差一般在百分之几十,亚硝酸盐氮甚至达153%。这就表明,不同的池塘或不同的养殖时期,其养殖环境不同。这就给养殖者提出了如何根据不同的水质情况,采用不同手段调控各项生态因子,让其处于良好状态的课题。

表1 鳊塘生态因子的测量结果

Tab. 1 The measurement values of ecosystem factor in the *Simperca chuatsi* ponds

变 量	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀
全距	27~ 34.3	16~ 45	7.3~ 8.6	0.001~ 0.675	0.272~ 5.22	0.001~ 0.148	7.0~ 19.1	52.7~ 422.7	0.1~ 6.6	3.3~ 9.6
平均值	30.7	24.8	7.96	0.0997	1.39	0.057	11.3	167.2	2.31	5.59
标准差	2.4	5.9	0.30	0.153	1.04	0.049	2.7	92.2	1.78	1.64
相对标准差 (%)	7.8	24	3.8	153	75	86	24	55	77	34

注: x₁, 水温(°C); x₂, 透明度(cm); x₃, pH值; x₄, 亚硝酸盐氮(mg/L); x₅, 铵氮(mg/L); x₆, 磷酸盐磷(mg/L); x₇, COD(mg/L); x₈, 叶绿素 a(μg/L); x₉, 异氧细菌数(10⁵个/L菌); x₁₀, 溶氧量(mg/L)。

3.3 模式识别技术对生态因子间关系的研究

表1指出不同池塘、不同时期的水环境差异较大,我们用CPR技术进行研究,将28次测量点按溶氧量高低分为两类。为使不同类样本数目相近,选择溶氧值不低于5.2为1类(15个),低于5.2为2类(13个)。通过研究两类样本点的聚集分布是否存在规律,推断水体溶氧量是否受其它九个因子所制约,并进而求出维持高溶氧值的水体环境的九个生态因子的优化组合推荐值。

3.3.1 不同溶氧量样本的聚类研究

图1是以pH值、化学耗氧量和叶绿素为自变量(3因素),图2则再增加异养细菌总数(4因素)为自变量,以溶氧为目标的二维分类图,图3、图4分别以九个因子为自变量,以溶氧为目标的二维、三维分类图。从图中看出,在所选择的因素下,两类样本大体上聚集在不同的区域,其周界基本上是清晰的,只是随着自变量个数的增加,其分类效果就更佳。这就给我们提供一个重要信息:虽然不同塘的水体环境不同,但溶氧量的高低都遵循一定的规律,均可以用一定的生态因子值来调控。这就为不同塘鳊鱼养殖寻找优化的生态因子值带来可能性。

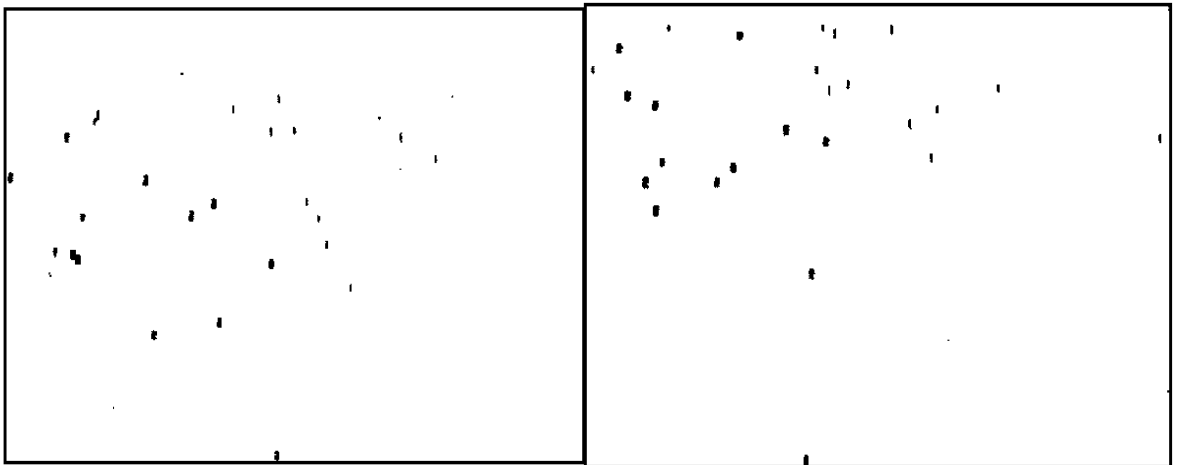


图1 pH、COD、叶绿素 a 样本分类图

Fig. 1 Three factors, pH, COD, chlorophyll a of samples of type distribute
1. 高溶氧 2. 低溶氧

图2 pH、COD、叶绿素 a、异养细菌数样本分类图

Fig. 2 Four factors, pH, COD, chlorophyll a, bacterium of samples of type distribute
1. 高溶氧 2. 低溶氧

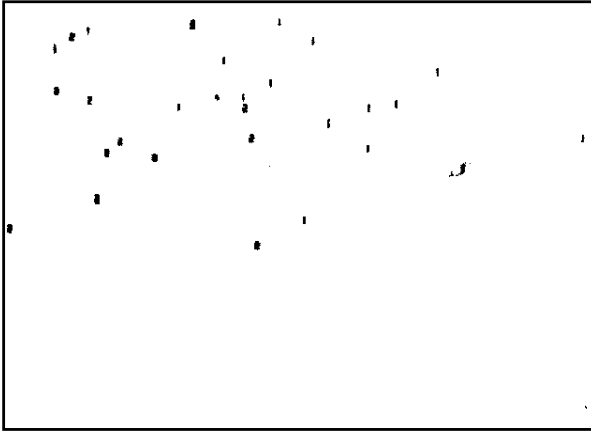


图 3 九因素样本二维聚类图

Fig. 3 Nine factors of samples of two-dimensional space type distribute

1. 高溶氧 2. 低溶氧

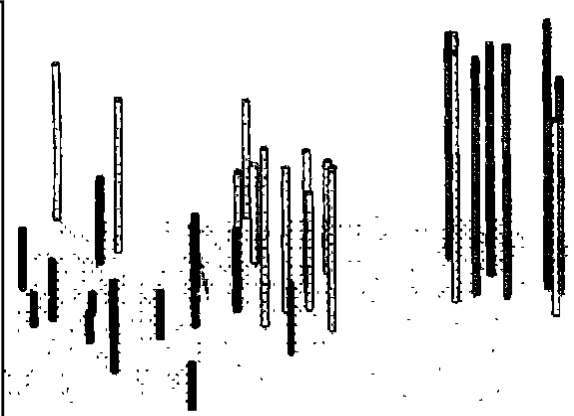


图 4 九因素样本聚类三维图

Fig. 4 Nine factors of samples of three-dimensional space type distribute

黑柱: 低溶氧; 白色柱: 高溶氧; 灰色柱: 优化样本

3.3.2 生态因子优化值的推求

图 3 可知, 低溶氧量的样本多在第三象限, 高溶氧量样本点多在第一象限。溶氧量的增长方向必然是由左往右, 从第三象限往第一象限变化。本文在高溶氧样本点附近, 沿溶氧增加方向, 设置若干优化点, 图 4 中的灰色柱, 用本文编制的程序施行逆方向的迭代计算, 推求对应点的九因素的优化值, 并利用本文编制的预报程序或下节求得数学模型, 推求其溶氧量, 均达 8.4 以上(表 2)。从而显示出利用计算机优化技术指导池塘水生态因子调控, 实现水质科学管理, 健康养殖的可能性。

表 2 不同模式点的生态因子寻优值

Tab. 2 The ecosystem factor advantage targets of different type

变 量	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀
模式点 1	30	30	8.5	0.10	0.3	0.08	8	200	0.60	8.8
模式点 2	31	28	8.6	0.1	0.2	0.11	11	197	2.5	9.7
模式点 3	30	29	8.4	0.1	0.3	0.10	10	210	1.7	8.9
模式点 4	33	30	8.4	0.1	0.6	0.1	12	242	2.3	9.3
模式点 5	29	28	8.5	0.1	0.05	0.16	9	182	3.5	8.4
模式点 6	30	27	8.5	0.15	1.1	0.06	10	220	1.5	8.6

注: 变量与表 1 同。

4 数学模型的建立

综合运用了多元线性、非线性、多项式拟合, 多元线性、非线性逐步回归技术推求各种条件下的数学模型。表 3 是部分回归结果。从表中可以看出, 我们选择与溶氧量较为密切的 1~ 3 个生态因子 (x₃, x₇, x₈) 或九个因子为自变量时, 线性回归均有显著的统计学意义, 当然九个因子的回归相对偏差最小, 这就表明溶氧量与 9 个生态因子应都有关系。但引入二次项非线性回归的结果显然比线性的好, 因此溶氧量的模型为多元的二次模型。逐步回归结果(表 3) 14 个因子时的模型最佳, 其复相关系数达 0.97, 回归相对偏差仅为 6.2%。

其线性模型为

$$Y = -28.0352 + 0.13x_1 + 0.0332x_2 + 3.7413x_3 + 2.19x_4 + 0.1874x_5 + 0.687x_6 - 0.402x_7 + 0.0147x_8 + 0.2684x_9;$$

非线性模型为

$$Y = -282.52 + 10.465x_1 - 0.1615x_1^2 + 0.0676x_2 + 27.37x_3 - 1.586x_3^2 + 25.78x_4 - 38.17x_4^2 - 2.72x_5 + 0.56x_5^2 - 90.54x_6^2 - 0.0173x_7^2 + 0.0105x_8 + 1.343x_9 - 0.134x_9^2。$$

表3 不同因子数建模效果

Tab.3 The mathematical model effect of different factor number

因子数	引用因子	相关系数(R)	回归相对偏差(%)	剩余标准差(σ)	说明
1	x_3	0.60	19.5	1.33	线性拟合
2	x_3, x_8	0.66	18.4	1.28	线性拟合
3	x_3, x_7, x_8	0.78	16.0	1.09	线性拟合
9	$x_1 \sim x_9$	0.87	12.3	0.994	线性拟合
14	$x_1, x_1^2 \sim x_9, x_9^2$ (缺 x_2^2, x_6, x_7, x_8^2)	0.97	6.2	0.622	非线性拟合
16	$x_1, x_1^2 \sim x_9, x_9^2$ (缺 x_7, x_8^2)	0.97	6.2	0.635	非线性拟合
18	$x_1, x_1^2 \sim x_9, x_9^2$	0.97	6.2	0.691	非线性拟合

5 结语

CPR 技术研究水生态因子复杂的相关问题, 从模型的建立到优化值的推求取得较满意的结果。本实验取样时间接近中午, pH、DO 都较高, 由 pH、DO 昼夜变化曲线还可以推求其不同时刻的数值^[11]。虽然本文只以鳊塘水样建立数学模型, 但其建模方法同样可以用于其它养殖池塘水体的讨论、预报。CPR 技术特有的强大的多元统计分析能力, 是研究自然界相互作用、相互依靠、相互竞争的生态平衡的新的有效技术, 鱼类病害与环境相关的预报、预测等方面的应用还有待进一步研究。

参 考 文 献

- 1 何志辉. 吉林镇赉养鱼场鱼池生态系统的分析. 水产学报, 1993, 17(1): 25~ 35
- 2 樊兰英. 生态环境与汾河水库输沙量的灰色关联分析和预测. 生态学杂志, 1990, 9(1): 28~ 30
- 3 谢 骏, 黄樟翰, 肖学铮等. 高产池塘混养鱼类关系的研究. 水产学报, 1998, 22(2): 106~ 110
- 4 Péron P, Mullon C, Pichon G. Experimental interactive software for choosing and fitting surplus production models including environmental variables. Rome: FAO Computerized Information Series Fisheries, 1993, 5: 1~ 76
- 5 Friedman J H, Stuetzle W. Projection pursuit regression. J Amer Statist Assoc, 1981, 76: 817~ 823
- 6 李金宗. 模式识别导论. 北京: 高等教育出版社. 1994, 1~ 133
- 7 王碧泉, 陈祖荫. 模式识别理论、方法和应用. 北京: 地质出版社. 1989, 1~ 68
- 8 湛江水产专科学校. 淡水养殖水化学. 北京: 农业出版社. 1979, 88~ 89
- 9 章宗涉, 黄祥飞. 淡水浮游生物研究方法. 北京: 科学出版社. 1995, 345~ 347
- 10 戴玉蓉, 卢敬让. 养殖水体浮游植物叶绿素 a 的简易测定. 齐鲁渔业, 1997, (2): 35~ 36
- 11 徐 宁, 李德尚. 养殖池塘溶氧平衡与日最低值预报的研究概况. 中国水产科学, 1998, 5(1): 84~ 87