

淡水养殖池塘水华发生及池水自净化研究

摘要

水产养殖是我国农业的重要组成部分，也是当前农村经济的主要增长点之一。本文针对池塘养殖中常见的水华灾害进行分析，探讨了养殖水体水质的影响因素，以及可持续发展生态养殖模式。

问题一中本文针对四个池塘的八个采样点的不同数据，采用了 *Pearson* 相关系数来度量水体、底泥与间隙水中常见主要理化因子之间的关系。发现池水与间隙水、池水与底泥之间理化因子无直接关联，而间隙水与底泥中的理化因子总体上具有较强线性相关性。探究其原因，底泥释放的物质直接作用于间隙水，而对体积庞大的池水影响甚微，池水中理化因子含量主要受到水生生物吸收与排放、人为调控等因素的影响。

对于问题二，水体质量可以直接从水体的理化因子指标上体现。本文结合我国地表水环境质量标准，选取了 6 个主要理化指标作为评价因子对池塘水质进行评价。考虑到评价因子的高维性，本文使用了可以探究高维数据特征的投影寻踪模型来进行评价系统的构建，利用基于实数编码的加速遗传算法(RAGA)对模型进行了求解，并将国家标准中不同等级的水体理化指标代入模型，得到在该模型下的水质评级标准。根据该模型可知各池水质处于动态变化，总体而言 2 号和 3 号池的水质较好，1 号池次之，4 号池最差。造成这种现象的主要原因在于养殖池中养殖密度的差异，过高的养殖密度会使水体溶氧降低，饵料、动物排放的废物浓度升高而引起水质下降。

关于问题三，由于藻类植物的生长与池水直接相关，故本文仅选择池水的主要理化指标来建立支持向量回归机模型，以刻画池水中理化因子与总藻类密度的关系，得到两者的回归函数参数（见表 7）。而后本文结合傅里叶逼近预测了未来 5 周 1 号池和 3 号池将会发生水华，其余两池水质较为乐观。接着我们使用网格搜索的方法逐一确定了水华发生时主要理化因子的波动范围（见表 9）。

对于问题四，本文对附件 6 的数据进行分析，剔除了部分异常值后，分别建立了两种鱼类体重与身长的指数回归方程，来描述身长与体重的关系。对于鱼类净化效果的研究，本文以鲢鱼为例，在考虑水循环的净化效果的基础上建立了池水种群的差分阻滞增长模型，求解该模型得到投入 166.4kg 鲢鱼时，在 20 周的演化后总藻类密度稳定在 $285.49 \times 10^6/L$ ，该方案可以较好地净化池中藻类。接着，本文对该模型的参数进行了敏感性分析，结果表明引入水循环，加之投入 166.4kg 鲢鱼的生态模式对一号池中藻类的净化效果是较为稳定可靠的。

对于问题五，本文以 1 号池为例创建生态养殖池，通过查阅相关资料并结合之前的结论，分别投放 300kg 鱼、 500kg 虾、 500kg 蟹，建立起一套鱼、虾、蟹、藻生态养殖模式，以总藻类密度作为指标，利用元胞自动机进行模拟。通过严重水华、一般水华、无水华三种情况的模拟结果可知，该生态系统能够起到很好的净化作用，但对于无水华的情况可能需要投放一定量人工饲料来保证鱼虾蟹类的正常生长。

最后，本文对模型进行了中肯的评价，并对模型的应用进行了推广。

关键词：水质评价、投影寻踪、支持向量回归机、差分阻滞增长模型、元胞自动机

淡水养殖池塘水华发生及池水自净化研究

目录

| | |
|-----------------------------------|----|
| 摘要 | 1 |
| 1 问题的提出 | 3 |
| 2 问题的分析 | 3 |
| 3 模型假设 | 4 |
| 4 符号说明 | 4 |
| 5 模型建立 | 4 |
| 5.1 问题一模型建立与求解 | 4 |
| 5.1.1 相关性分析 | 4 |
| 5.1.2 问题 1 结果的原因探究 | 5 |
| 5.2 问题二的模型建立与求解 | 6 |
| 5.2.1 水质指标选取 | 6 |
| 5.2.2 数据预处理 | 6 |
| 5.2.3 水质评价与分类投影寻踪模型的建立 | 7 |
| 5.2.4 基于实数编码的加速遗传算法求解投影寻踪模型 | 8 |
| 5.2.5 评价体系的构建 | 10 |
| 5.2.6 问题二结果分析 | 12 |
| 5.3 问题三的模型建立与求解 | 12 |
| 5.3.1 支持向量回归机的选择和建立 | 12 |
| 5.3.2 支持向量回归机模型的求解 | 13 |
| 5.3.3 各池塘水华情况的预测 | 14 |
| 5.3.4 支持向量回归机模型拟合精度检验 | 14 |
| 5.3.5 水华发生的理化因子范围确定 | 15 |
| 5.4 问题四的模型建立与求解 | 15 |
| 5.4.1 鱼类身长与体重相关模型 | 16 |
| 5.4.2 差分阻滞增长模型 | 17 |
| 5.4.3 差分阻滞增长模型的敏感性分析: | 18 |
| 5.5 问题五的模型建立与求解 | 19 |
| 5.5.1 生态养殖模式分析 | 19 |
| 5.5.2 基于元胞自动机的生态养殖模型 | 19 |
| 5.5.3 生态养殖池模拟结果 | 21 |
| 6 模型优缺点 | 23 |
| 7 模型推广 | 24 |
| 8 参考文献 | 24 |

1 问题的提出

水华（英文 Algal Blooms 或 Algae Blooms）是由于生活和工农业生产中，大量含有氮、磷、钾元素的废水排入淡水水体，从而引发的藻类大量繁殖的自然现象，是水体富营养化的特征^[1]。

近年来，淡水生态系统水体污染和富营养化程度加剧，作为淡水养殖主要方式之一的池塘养殖毫无疑问也面临着水华的威胁。水华会破坏养殖生态系统，导致养殖对象不同程度死亡，造成巨大经济损失。因此，针对水华现象的研究迫在眉睫，我们拟通过研究淡水养殖池塘主要理化因子、主要浮游生物数据及鱼虾生成等数据分析水华发生的原因，并在水质评价的基础上控制并预测水华的发生，从而最大程度上提高养殖产量，减小环境污染等。并期望通过对水华发生的研究，加强大家环保意识。

2 问题的分析

2.1 问题一的分析

要求分析水体、底泥与间隙水中常见主要理化因子之间的关系，并分析原因。即探寻鱼塘的上述三种结构两两组合下，各理化指标数值之间的相关程度，并结合水环境相关的知识给出原因。

2.2 问题二的分析

对四个池塘水体质量进行评价及分类，本文认为应选取合理的指标，并结合地表水环境质量国家标准建立分类和评价模型，以保证模型的准确性。根据国家标准划分 5 类标准水质等级，并用其衡量各池塘水质。最后根据四个池塘的水质情况，分析所养殖的动物（虾或鱼）对水体的影响。

2.3 问题三的分析

建立主要理化因子和常见浮游生物致害密度之间关系的模型，由于与浮游植物直接相关的为池水，本文选择池水的理化指标进行考虑。探究二者关系时，浮游生物量是因变量，理化因子是自变量，考虑沿用问题二中评价水质的指标因子，又由于与浮游植物直接相关的为池水，本文选择池水的理化指标进行考虑。已知 1 号池发生了轻微水华，故而可以据此确定水华发生的浮游生物临界值，通过建立模型预测出各池在一定滞后期内浮游生物的变化情况与临界值相比较，即可预测出各池未来水华发生的情况。对于水华发生时主要理化因子的范围，考虑使用网格搜索方法，通过控制变量来逐一确定各理化因子的变化范围与水华发生的关系。

2.4 问题四的分析

对于鱼类生长与体重相关模型，应先根据附件 6 的鱼类体重体长数据，绘制散点图并处理异常数据，再由数据分布决定使用何种模型进行拟合。

水华现象的产生与藻类的密度具有直接关系，构造一个与 1 号池相同大小的净化池，通过水循环，并放养鲢鱼或鳙鱼来净化藻类，结合生物的繁殖变化，本文建立考虑差分阻滞增长模型来模拟池塘中两种生物的变化过程。

2.5 问题五的分析

为了构建生态养殖模式，我们需要使水中生物构成一条和谐的生物链，实现池水的自净化，应基于前述问题考虑到多种生物之间具有复杂的捕食和共生关系，且纳入不同区域的水环境之间的相互作用。分析发现常见的方程不容易简单描述，故考虑采用元胞自动机进行模拟，以总藻类密度作为指标，分析各生物生长情况以及对水质的控制情况。

3 模型假设

- 假设各养殖池塘的水温、光照条件一致；
- 由于附件中所给的理化因子数据量较少，对样本的统计检验难以逼近实际情况，考虑到正态性是自然界中普遍存在特性，故假设池中理化因子值均服从正态分布。
- 为简化研究，假设池塘中常见的理化因子指标可以有效地反映浮游植物的生长情况，可作为影响池中浮游植物生长的主要因子。
- 假设本题附件中所提供的数据真实可靠，在实验过程中对藻类的生长繁殖无人影响。

4 符号说明

| 符号 | 含义 |
|------------|-----------|
| r | 相关系数 |
| x_{new} | 同趋化后的数据 |
| z_i | 投影值 |
| a | 投影方向 |
| σ_z | 投影值的标准差 |
| ρ_z | 投影值的局部密度 |
| $Q(a)$ | 投影指标函数 |
| s_i | 适应度函数 |
| $weight_1$ | 相邻元胞的影响权重 |
| $weight_2$ | 元胞自身的影响权重 |

5 模型建立

5.1 问题一模型建立与求解

5.1.1 相关性分析

为探寻池水、间隙水和底泥两两之间理化因子的相关关系，本文考虑使用皮尔森相关系数（Pearson correlation coefficient），选取总磷、磷酸盐磷、总氮、硝态氮、亚硝态氮、铵态氮这六种理化因子，通过定量分析得到相对可靠的结果。

5.1.1.1 Pearson 相关系数模型

Pearson 相关系数用来描述两个数据集的线性相关程度，其具体计算公式如下：

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

上式中 r 代表相关系数, n 样本数量, X_i 和 Y_i 分别表示两组样本数据集的第 i 个属性值。 r 的变化区间为 $[-1,1]$, 负值代表 X, Y 负相关, 正值代表 X, Y 正相关。其绝对值越接近于 1, 则代表相关性越强。为定量描述, 本文认为 r 的绝对值大于 0.8 时为高度相关, 小于 0.3 为低度相关, 介于二者之间时为中度相关。

本文拟探寻池水、间隙水、底泥两两之间相同理化因子的相关系数, 用 SAS 软件计算 (程序请见附录 1) 得到结果如下表:

表 1 池水、间隙水、底泥理化因子相关系数

| r | 总磷 | 磷酸盐磷 | 总氮 | 硝态氮 | 亚硝态氮 | 铵态氮 |
|--------|--------|----------------|----------------|----------------|--------|----------------|
| 池水与间隙水 | 0.060 | 0.173 | 0.165 | 0.043 | 0.138 | 0.054 |
| 池水与底泥 | 0.102 | -0.001 | 0.191* | 0.080 | -0.067 | -0.198* |
| 间隙水与底泥 | -0.094 | 0.425** | 0.615** | 0.468** | -0.034 | 0.609** |

注: **为 0.01 水平上显著,*为 0.05 水平上显著

根据上表, 下面逐个分析水体的三个部分 (池水、间隙水和底泥) 两两之间相同理化因子的相关程度。

池水与间隙水: 在六个常见主要理化因子的相关系数值均偏小, 尽管呈现正相关, 但都低于 0.3, 且范围仅介于 $[0,0.2]$ 之间, 表明池水与间隙水的理化因子之间呈现低度相关。故而可认为, 池水与间隙水中理化因子含量并无直接关联, 但可能存在一定关系。

池水与底泥: 相关系数呈现的情况与池水和间隙水之间比较时相似, 相关系数值介于 $[0,0.2]$ 之间, 说明池水和底泥的相关性不强, 本文认为池水与间隙水中理化因子含量也没有直接关系, 但可能存在一定关系。

间隙水与底泥: 与前面两个情况不同, 间隙水与底泥的磷酸盐磷和硝态氮含量相关系数超过了 0.4, 总氮和铵态氮含量相关系数达到 0.6, 可见两种介质的上述 4 种理化因子具有较强关联; 而我们也注意到, 只有总磷和亚硝态氮含量间的相关性较低。所以总体上来看, 间隙水与底泥在常见理化因子之间相关性较大, 但总磷和亚硝态氮例外。

5.1.2 问题 1 结果的原因探究

通过上述研究我们得到了如下结果: 间隙水或底泥与池水之间理化因子**无直接关联**, 而间隙水与底泥中的理化因子总体上具有**较强**线性相关性, 但总磷和亚硝态氮例外。对于这一结果本文进行了进一步探究。

(1) 池水中理化因子含量主要受到水生生物吸收与排放、人为调控等因素的影响, 相比起这些活跃因素, 底泥的物质释放作用甚微, 因此导致池水的理化因子与底

泥、间隙水没有直接联系。

(2) 间隙水与底泥之间的磷酸盐磷、总氮、硝态氮、铵态氮相关性较强，原因是这些是溶解态的物质，较易释放至间隙水中。例如，底泥对硝酸根离子的吸附作用很小，释放较充分，所以底泥与间隙水的硝酸根离子显著相关。

(3) 而关于总磷和亚硝态氮例外的原因：①底泥中总磷含量虽然很大，但磷元素的释放却少而慢，且磷的释放所受的干扰因素较多，因此大多数磷元素都沉积于底泥深处未进入间隙水。②间隙水的提取过程中涉及过滤步骤，所以最终间隙水中的总磷含量只包含了所有溶解态的磷，而原先含有的少量颗粒磷被滤除，致使间隙水中的总磷含量偏小，与底泥的相关性随之减小^[2]。

亚硝酸盐是硝化作用的中间产物，性质极不稳定的，在好氧条件下迅速转化为硝酸盐^[2]，所以在水体中的含量较少，因此致使底泥与间隙水中亚硝态氮含量相关性极低。

5.2 问题二的模型建立与求解

5.2.1 水质指标选取

本文认为评价水质要从污染程度、自净能力和生物承载能力等方面考虑来选取指标。由于总磷、总氮、化学需氧量 COD (Chemical Oxygen Demand) 是无机物和有机物污染参数；氨氮是水体中主要耗氧物，其含量高会对鱼类及其他水生生物有毒害^[3]；溶氧是指示水体自净能力的指标；PH 会直接或间接影响养殖对象的生存和繁殖，所以综上考虑，我们选取了总磷、总氮、COD、氨氮、溶氧、PH 这 6 个指标来进行水质评价与分类。

5.2.2 数据预处理

Step1: 三次样条插值补全缺失数据。

本文取每个水池中，A、B 两个采样点各理化因子实测值的均值作为各理化因子的计算值。总磷、总氮、氨氮 15 周的数据可以参考附件一获取。而附件 2 中 COD、溶氧、PH 的值均是间隔两周采样一次，数据量不足以用于建立合理的模型，因此我们考虑利用现有数据进行插值以补充数据。

插值方法选用三次样条插值，因为该方法可以很好地保持数据光滑性和连续性，减少信息量的损失。最终得到的 15 周的全部指标数据请见附录 2。

Step2: 数据同趋化。

所选指标中，一些指标属于高优指标，即数值越大越理想，如溶氧；一些指标是低优指标，即数值越小越理想，如总磷、总氮、氨氮、COD，另外还有 PH 这类值过低过高均不合理的指标。为处理上述性质不同的指标，正确反映不同作用力的综合结果，需要做同趋化处理。

假设每个指标含有 m 个数值， x_{new} 为一个指标同趋化后的数据， x_{max} 是该指标的最大值， x_{min} 为最小值。对于不同性质指标具有各自计算公式：

(1) 对于高优指标的同趋化采用如下公式：

$$x_{new} = \frac{x_{max} - x_i}{x_{max} - 0}, i = 1, 2, \dots, m$$

(2) 对于低优指标的同趋化采用如下公式：

$$x_{new} = \frac{x_i - x_{min}}{x_{max} - 0}, i = 1, 2, \dots, m$$

(3) 对于 PH 的同趋化, 由于我国地表水环境标准要求水体 PH 在[6,9]之间, 故采用其均值 7.5 作为同趋化标准, 具体如下公式:

$$x_{new} = 1 - \frac{|x_i - 7.5|}{|x - 7.5|_{max}}, i = 1, 2, \dots, m$$

5.2.3 水质评价与分类投影寻踪模型的建立

投影寻踪法 (PP) [4] 属于直接由样本数据驱动的探索性数据分析方法。它把高维数据 $x \in R^{m \times n}$ 通过某种组合投影到低维子空间上 $\{z(i)\}$ 进行聚类, 通过分析低维空间数据点的变化特征来研究高维数据结构特征。聚类就是按照一定的规则或者标准将对象聚为一个个集合, 所得到的分布是聚类前未确知的。它以类内具有较大密集度 (高内聚) 而类间具有相对大的散开度 (低耦合) 为目标寻找一维投影方向, 并根据每个样本相应的综合投影值对样本进行综合评价分析的方法。

我们运用投影寻踪法在评价样本集 $x \in R^{m \times n}$ 下进行讨论。其中 $x_{i,j}$ 为第 i 个样本第 j 个指标值。

首先构造投影指标函数。投影寻踪法就是把 n 维数据 $\{x_{i,j} | j = 1, 2 \dots n\}$ 综合成以 $a = (a_1, a_2 \dots a_n)$ 为投影方向的一维投影值 z_i :

$$z_i = \sum_{j=1}^n a_j \times x_{i,j}$$

然后根据 $\{z_i | i = 1, 2 \dots n\}$ 的一维散布图进行分类。上式中 a 为单位长度向量。在综合投影值时, 要求投影值 z_i 的散布特征应具有高内聚低耦合的特性。据此投影指标函数可构造为

$$Q(a) = \sigma_z \times \rho_z$$

式中, σ_z 为投影值 z_i 的标准差, ρ_z 为投影值 z_i 的局部密度。

样本序列的投影特征值标准差计算方式如下:

$$\sigma_z = \sqrt{\frac{\sum_{i=1}^n ((z_i - \bar{z}))^2}{n - 1}}$$

\bar{z} 为所有样本在投影方向上的投影特征值的均值, 根据概率论与数理统计的思想, 当 σ_z 越大, 样本集合的耦合度越低, 表现为类与类之间的散开程度越大。

设任意两个样本 k, l 之间的投影值之差为

$$r_{kl} = |z_k - z_l|$$

则投影值 z_l 的局部密度 ρ_z 可由以下公式计算得到:

$$\rho_z = \sum_{k=1}^m \sum_{l=1}^m (R - r_{kl}) \mu(R - r_{kl})$$

其中 R 为密度窗宽度参数。规定 $R > 0$, 且当 $k = l$ 时, 有 $(R - r_{kl}) \mu(R - r_{kl}) = R$, 同时 $r_{kl} = r_{lk}$, 于是可将上式整理如下:

$$\begin{aligned} \rho_z = mR + 2 \sum_{k=2}^m (R - r_{1k}) \mu(R - r_{1k}) + 2 \sum_{k=3}^m (R - r_{2k}) \mu(R - r_{2k}) + \dots \\ + 2 \sum_{k=n-1}^m (R - r_{m-2,m}) \mu(R - r_{m-2,m}) + 2(R - r_{1m}) \mu(R - r_{1m}) \end{aligned}$$

r_{kl} 愈小，则类内密度愈大；反之，则类内密度愈小。 $\mu(R - r_{kl})$ 为单位阶跃函数，当 $R > r_{kl}$ 时， $\mu(R - r_{kl}) = 1$ ，否则该项为0。密度窗宽度参数 R 的最佳取值为：

$$R = \frac{\max(r_{kl}) + n}{2}$$

对于投影指标函数 $Q(a)$ 。当各指标值的样本集给定时，投影指标函数 $Q(a)$ 只随投影方向 a 的变化而改变。可通过求解投影指标函数最大化问题，来估计最大可能展现高维数据某类特征结构的最佳投影方向。即该问题可视为如下优化问题：

目标函数：
$$\max_a Q(a) = \sigma_z \times \rho_z$$

约束条件：
$$\text{s.t. } \sum_{j=1}^n a_j^2 = 1$$

这是一个以 $\{a|j = 1, 2 \dots n\}$ 为优化变量的复杂非线性优化问题，运用 *Lingo* 等优化软件一般难以求出其解。所以我们考虑使用模拟生物优胜劣汰规则与群体内部染色体信息交换机制的基于实数编码的加速遗传算法（RAGA）来求解上述问题。

求解最佳投影方向后，本文利用求解所得的最佳投影方向 a^* 计算各个样本的投影指标值来进行分类评价：

$$z_i^* = \sum_{j=1}^n a_j^* \times x_{i,j}$$

5.2.4 基于实数编码的加速遗传算法求解投影寻踪模型

遗传算法是基于达尔文的进化论和孟德尔的遗传学说的理论基础发展得到的，是一种依据生物进化思想寻找全局最优解的概率优化算法，可以利用简单的编码技术和算法机制，模拟复杂的优化过程^[5]。

常用的标准遗传算法（SGA）基于二进制编码，但存在着一些问题如：易收敛于局部最优值、计算量大、解的精确度不高等。而基于实数编码的加速遗传算法

（RAGA）将标准遗传算法的二进制编码变为实数编码的形式，采用加速循环的方式求解使得求解的速度更为快捷，改善了标准遗传算法的部分问题，故而本文考虑使用基于实数编码的加速遗传算法求解所建立的投影寻踪模型^[6]。

RAGA 的求解步骤如下：

Step1: 编码。标准遗传算法采用二进制的编码方式，不但增加了算法执行过程中的计算量，且增加了产生额外最优解的可能性。RAGA 采用的实数编码利用了如下的线性变换：

$$a_j = \underline{a_j} + y_j (\overline{a_j} - \underline{a_j}) \quad (j = 1 \dots n)$$

其中 $[\underline{a_j}, \overline{a_j}]$ 为决策变量（投影方向分量） a_j 的定义域。可将该变换视为将投影方向 $a = (a_1, a_2 \dots a_n)$ 编码为 $y = (y_1, y_2 \dots y_n)$ 的染色体形式，且对任意的 j ，有 $y_1 \in [0, 1]$ 。

Step2: 初始化群体。设初始化群体个数为 M ，分别在定义域 $[\underline{a_j}, \overline{a_j}]$ 内随机生成 M 组初代投影方向 $\{dre_{j,i}^{(0)}\} (j = 1 \dots n, i = 1 \dots M)$ ，将其代入步骤一中进行实数编码后得到初

代编码序列 $\{y_{j,i}^{(0)}\}(j = 1 \cdots n, i = 1 \cdots M)$ ，将该序列依次代入目标函数 $Q(a)$ 中求取群体中各个个体的目标函数值 $\{f_i^{(0)}\}(i = 1 \cdots M)$ ，依据目标函数值的大小对个体进行升序排序，选择前 Num ($Num < M$)个个体作为优秀个体。

Step3: 适应度评价。由于目标函数值越小，表示该个体的适应度越高，再者考虑目标函数值为0的情况，故定义适应度函数为：

$$S_i = \frac{1}{(f_i^{(0)})^2 + 0.001}$$

Step4: 进行选择操作产生第1个子代群体 $\{y_{j,i}^{(0)}\}(j = 1 \cdots n, i = 1 \cdots M)$ 。第 i 个个体被选择的概率为 $E = \frac{S_i}{\sum_{i=1}^M S_i}$ ，定义 $\{p|p_i = \sum_{k=1}^i E_k\}$ ，则 p 将 $[0,1]$ 区间划分为 M 个子区间： $\{[0, p_1], (p_1, p_2] \cdots (p_M, 1]\}$ 。接着随机生成 $(M - Num)$ 个随机数 $Rand \in R^{M-Num}$ ，则个体选择方式为：

$$y_{j,k}^{(0)} = y_{j,i}^{(0)} \quad R^k \in (p_{i-1}, p_i)$$

对于剩余的 Num 五个个体，采用上一代中的优秀个体直接补充，即得第1个子代群体。

Step5: 进行杂交操作产生第2个子代群体 $\{y_{j,i}^{(2)}\}(j = 1 \cdots n, i = 1 \cdots M)$ 。根据步骤四中得到的选择概率 E_i 随机选择一对个体 $y_{j,k1}$ 与 $y_{j,k2}$ 进行杂交，因为编码方式的限制，故本文采用如下随机线性组合的方式进行杂交获得子代个体 $y_{j,i}^{(2)}$ ：

$$\begin{cases} y_{j,i}^{(2)} = r_1 \times y_{j,i}^{(0)} + (1 - r_1) \times y_{j,i}^{(0)} & r_3 < 0.5 \\ y_{j,i}^{(2)} = r_2 \times y_{j,i}^{(0)} + (1 - r_2) \times y_{j,i}^{(0)} & r_3 \geq 0.5 \end{cases}$$

其中 r_1 、 r_2 、 r_3 为在 $[0,1]$ 区间内的随机数。

Step6: 进行变异操作产生第3个子代群体 $\{y_{j,i}^{(3)}\}(j = 1 \cdots n, i = 1 \cdots M)$ 。采用 n 个随机数以 $pm_i = 1 - E_i$ 的概率来代替第 $y_{j,i}$ 个个体进行变异：

$$\begin{cases} y_{j,i}^{(3)} = r_4 & r_5 < pm_i \\ y_{j,i}^{(3)} = y_{j,i}^{(0)} & r_5 \geq pm_i \end{cases}$$

其中 r_4 、 r_5 为在 $[0,1]$ 区间内的随机数。

Step7: 演化迭代。将步骤四到步骤六中得到的个体依照适应度大小升序排序后，选取前 M 个个体构成新的种群，代入步骤三重新进行繁殖演化。

Step8: 加速循环。使用第一次与第二次演化迭代所产生的优秀个体对应的变量变化区间作为新的初始变化区间。算法重新进入步骤一。如此循环，直至算法达到预设的加速次数 N 时，结束迭代。在本问中，在如下参数值下进行求解。

表 2 参数设置表

| 变量名 | 变量取值 |
|-------|------|
| M | 400 |
| N | 7 |
| Num | 20 |

由于遗传算法是一个近似求解的算法，我们利用 *Matlab* 编程（程序见附录 3），设置运行 50 次，取目标函数的最大值作为投影寻踪法最终的最优值，其对应的解为最优解，即最佳投影方向。

运行得到最佳投影 α^* 与部分池塘的投影评价值（所有池塘投影评价值见附录 4）如下所示：

表 3 最佳投影方向及部分池塘的投影评价表

| 1#池 | 投影评价值 | 3#池 | 投影评价值 |
|--------|-------------|-----------------------------------------------------------------------------------|-------------|
| 1 | 1.172889074 | 1 | 1.23794772 |
| 2 | 1.490496041 | 2 | 1.454538707 |
| 3 | 1.608093057 | 3 | 1.434324726 |
| 4 | 1.498412196 | 4 | 1.120787322 |
| 5 | 1.466201903 | 5 | 1.292448181 |
| 6 | 1.474638961 | 6 | 1.020178215 |
| 7 | 1.593014056 | 7 | 1.485566789 |
| 8 | 1.247179561 | 8 | 1.499358493 |
| 9 | 0.889599071 | 9 | 1.27424924 |
| 10 | 1.315203435 | 10 | 1.222795578 |
| 11 | 1.285177612 | 11 | 1.439195359 |
| 12 | 1.27196821 | 12 | 1.376053346 |
| 13 | 1.105201678 | 13 | 1.280694159 |
| 14 | 1.0605161 | 14 | 1.273261307 |
| 15 | 1.140890083 | 15 | 1.45148397 |
| 最佳投影方向 | | (0.566210302, 0.557307031, 0.552752271, 0.052397675, 0.101274584, 0.224226752) | |

5.2.5 评价体系的构建

根据地表水环境质量国家标准 GB3838-2002，依据地表水水域环境功能和保护目标，按功能高低依次划分为 I~V 类水，所选指标在各类水的标准如下：

表 4 选取指标对应地表水环境质量标准

| | I 类 | II 类 | III类 | IV类 | V 类 |
|------------|------|------|------|-----|-----|
| PH | 6~9 | | | | |
| 溶氧 \geq | 7.5 | 6 | 5 | 3 | 2 |
| 总磷 \leq | 0.02 | 0.1 | 0.2 | 0.3 | 0.4 |
| 总氮 \leq | 0.2 | 0.5 | 1 | 1.5 | 2 |
| COD \leq | 15 | 15 | 20 | 30 | 40 |
| 氨氮 \leq | 0.15 | 0.5 | 1 | 1.5 | 2 |

我们将上述标准结合到所建立的投影寻踪模型，将上述每类水的 6 个理化指标范围化为 1 个综合范围，也就是类别之间的分界线，所得的 5 个类别范围如下：

表 5 类别范围表

| 水质分类 | I 类 | II 类 | III 类 | IV 类 | V 类 |
|-------|-------------|-------------|-------------|-------------|-------------|
| 指标值大小 | 1.713014173 | 1.623529207 | 1.492162265 | 1.379100947 | 1.260393328 |

同理，可利用所建模型将 4 个水池的 6 个理化指标数据，化为一个综合指标作为评价值，这样就可以将每个水池每周的综合指标与各类水体的划分标准对比，进而评价各池水体的质量。画出直观的分布图：

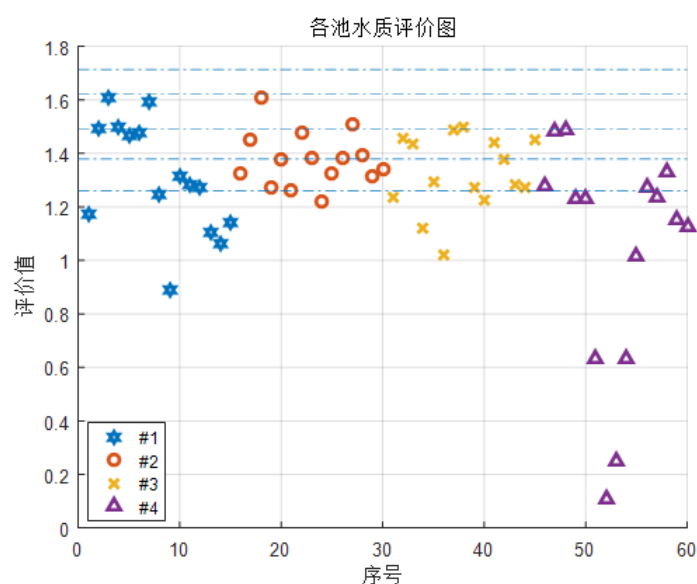


图 1 各池水质评价图

上图是 1~4 池每周的评价指标分布图，图中的五条直线是五类水的划分标准，我们看到水池水质的分类并非固定不变，而是动态变化的，计算各池水体属于各类别的周数占总周数（15 周）的比例如下：

表 6 各池水质类别比例

| | I 类 | II 类 | III 类 | IV 类 | V 类 | 劣 V 类 |
|-----|-----|------|-------|-------|-------|-------|
| 1#池 | 0 | 0 | 0.2 | 0.2 | 0.2 | 0.4 |
| 2#池 | 0 | 0 | 0.133 | 0.333 | 0.467 | 0.067 |
| 3#池 | 0 | 0 | 0.067 | 0.333 | 0.333 | 0.267 |
| 4#池 | 0 | 0 | 0 | 0.133 | 0.2 | 0.667 |

通过上表的定量计算，得到以下评价：四个水池在 15 周内水质均有波动，但均未达到过 I、II 类水标准，最高等级是 III 类水，且各池之间水质差异较为明显。1#池是四个池中达到 III 类水周数最多的，但其水质在 15 周内的两极分化比较明显，水质波动较大；2#池水质只有 1 周处于 V 类水以下，相比于其他水池其水质条件是 4 池中最好

的；3#池水质整体也比较良好，但仍有个别周劣于V类；4#池水质表现最差，没有达到过III类水，且有三分之二的时间段水质劣于V类。

5.2.6 问题二结果分析

相较而言 2#和 3#池的水质较好，1#池次之，4#池最差。根据题目附件 5 给出的各池放养对象和数量，1、2 池是虾池，3、4 池是鱼池，拥有不同养殖对象的池子之间相比，虾池的水质整体上优于鱼池。注意到虾池的投放量远小于鱼池，所以这一结论应与投放量有关。放养量多的池子，相应给予的饵料、动物排放的废物以及由此生长出来的水生植物都会较多，这些因素都将促进水质的下降。此外，鱼类的呼吸作用会消耗水体中的溶解氧（DO），溶氧下降后 COD 等污染物不易分解，造成水质变差。

再者我们由图 1 观察到 15 周内各池水质不断波动，我们推断这是由于换水、微生物净化以及水体自净能力造成的。

5.3 问题三的模型建立与求解

对于理化因子和常见浮游植物密度之间关系的研究，由于与浮游植物直接相关的为池水，本文选择池水的理化指标进行考虑，具体沿用水质评价的 6 个指标，即总磷、总氮、COD、氨氮、溶氧、PH。而池水中的浮游植物种类繁多，我们以藻类总量为研究对象进行拟合和分析。

再者考虑到数据呈现的波动性较大，难以采用常规的多元拟合等方法来探究两者间的关系，而支撑向量回归机（SVR）作为一种机器学习模型，可以通过训练学习来挖掘数据间深层的关系，故本文考虑采用支撑向量回归机来对理化因子和常见浮游植物密度进行训练拟合。

5.3.1 支持向量回归机的选择和建立

支持向量回归机是基于支持向量机（SVM）的回归拟合方法^[7]。Vapnik 等人在 SVM 分类的基础上引入了 ϵ 不敏感损失函数，从而得到了回归型支持向量机。SVM 应用于回归拟合分析时，旨在寻找一个最优分类面使得所有训练样本离该最优分类面的误差最小。为了防止过拟合现象的出现，本文使用每组采样点的前 12 个数据（涉及四个水池 A、B 两采样点，共 96 组）来作为训练样本，其他的数据来检验最优分界面的拟合度。

设投影到高维特征空间中建立的线性回归函数为：

$$f(x) = w\varphi(x) + b$$

其中， $\varphi(x)$ 为非线性映射函数，定义 ϵ 为线性不敏感损失函数，它可以忽略真实值在某个上下范围内的误差，它的解以函数的最小化为特征，确保对偶变量的稀疏性还可以确保全局最小解的存在和可靠泛化界的优化。通过规划的思想使得误差最小，然后引入 Lagrange 函数，并转化为对偶形式，如下：

$$\max_a \left[-\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) - \sum_{i=1}^l (a_i - a_i^*)\epsilon + \sum_{i=1}^l (a_i - a_i^*)y_i \right]$$

$$s. t. \begin{cases} \sum_{i=1}^l (a_i - a_i^*) = 0 \\ 0 \leq a_i \leq C \\ 0 \leq a_i^* \leq C \end{cases}$$

考虑到径向基函数(RBF 函数)不但可以将样本映射到一个高维空间, 而且它比多项式核函数等的参数更少, 可以减小模型的复杂程度, 故在训练中本文选择径向基核函数来作为核函数。式中 a_i 为引入的 Lagrange 系数, C 为惩罚因子, 它可以控制错分样本惩罚。求解可得最优解 a^* 和 a 向量, 进而可求出 w^* 和 b^* , 可得到最优的回归函数为:

$$f(x) = w^* \phi(x) + b^* = \sum_{i=1}^l (a_i - a_i^*) \phi(x_i) \phi(x) + b^* = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x) + b^*$$

若不为零, 则为对于的样本 x_i 的支持向量。

首先本文以 0.5 为步长, 利用交叉验证法来确定最优的 c 参数和 g 参数。通过 matlab 的支持向量机工具箱 libsvm 的 svmtrain 函数得出误差最小的相应参数。得到最佳参数之后, 使用 svmpredict 函数来求解出预测值, 定义测试集的均方误差 E 和决定系数 R^2 , 检验模型回归拟合的效果。为直观表述, 绘制 SVR 的大致结构图如下:

SVR结构图

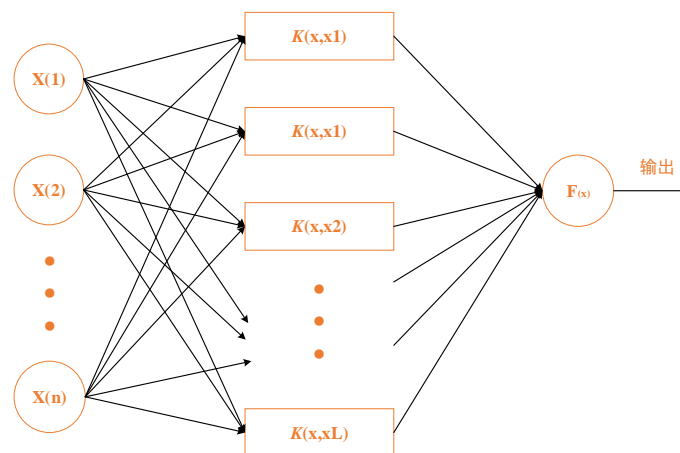


图 2 SVR 结构

5.3.2 支持向量回归机模型的求解

通过上述的分析, 执行程序 (见附录 5) 求解上述模型得支撑向量, 将其乘以对应的决策系数可得回归函数的系数 w 以及截距项 b :

表 7 回归系数与截距项

| 系数值 | |
|-----|--------------|
| b | 0.366269335 |
| w | -2.962752354 |
| | -1.143588937 |
| | -2.900956755 |
| | 0.899939428 |
| | -0.463251901 |
| | -0.371569136 |

5.3.3 各池塘水华情况的预测

为预测未来藻类的变化情况以判断水华的发生，首先应获得未来各池理化因子的数值。本文使用傅里叶曲线拟合计算出未来五周中4水池的6种理化因子预测值，并将之代入上述回归模型，计算得到未来五周藻类总量变化趋势如图：

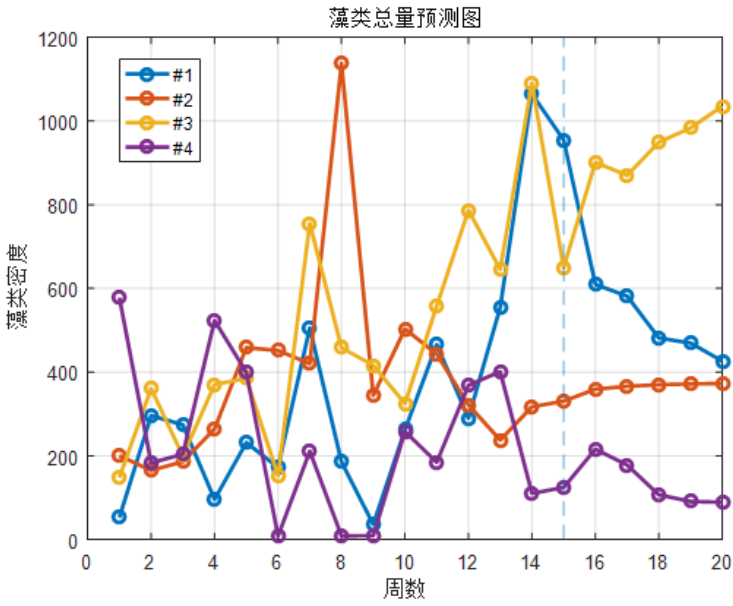


图 3 藻类总量预测图

已知1号池发生轻微水华，观察上图中1号池（蓝色线段）的藻类数量变化趋势，前期波动较小，但在12~14周藻类总量陡然上升，我们认为这一阶段发生了水华，认为13周总藻类密度 554.71×10^6 个/L为水华发生的**临界密度**，以此为界限可以判断未来四个水池水华发生情况。

本文预测的未来5周各池总藻类数量变化情况如下表：

表 8 各池未来五周总藻密度（个/L）预测表

| | 1#池 | 2#池 | 3#池 | 4#池 |
|-----|----------|----------|-----------|----------|
| 16周 | 609.5175 | 359.2779 | 901.3352 | 216.2291 |
| 17周 | 581.6579 | 366.2319 | 870.1059 | 178.1228 |
| 18周 | 481.6522 | 369.9249 | 949.4467 | 106.9568 |
| 19周 | 469.6187 | 371.8861 | 983.6214 | 91.0534 |
| 20周 | 426.4228 | 372.9276 | 1035.4775 | 89.2181 |

结合表8和图3我们发现，未来5周发生水华的是1号池和3号池，但1号池总藻密度在16、17周超过水华发生临界值后，有逐步下降趋势，说明污染态势得到控制，环境有一定改善；而3号池的总藻密度居高不下甚至渐有升高，表明随着时间的推移养殖环境恶化，如不采取有效措施将会造成损失。

相比而言，2号池总藻密度变化平缓，4号池逐步下降，两池的环境均达不到水华爆发的阈值，其未来5周整体环境较为乐观。

5.3.4 支持向量回归机模型拟合精度检验

为了检验标准 SVR 模型回归拟合的误差，以及与原样本的贴近程度，本文建立测试集的均方误差 E 和决定系数 R^2 来判别拟合结果的优劣。具体的计算公式如下：

$$E = \frac{1}{l} \sum_{i=1}^l (\hat{y}_i - y_i)^2$$

$$R^2 = \frac{(l \sum_{i=1}^l \hat{y}_i y_i - \sum_{i=1}^l \hat{y}_i \sum_{i=1}^l y_i)^2}{((l \sum_{i=1}^l \hat{y}_i^2 - (\sum_{i=1}^l \hat{y}_i)^2)(\sum_{i=1}^l y_i^2 - (\sum_{i=1}^l y_i)^2))}$$

其中 l 为测试集样本个数； $y_i (i = 1, 2 \cdots l)$ 为第 i 个样本的真实值； $\hat{y}_i (i = 1, 2 \cdots l)$ 为第 i 个样本的预测值。通过 MATLAB 求解预测值与真实值画出折线图并求出均方误差和决定系数，如图：

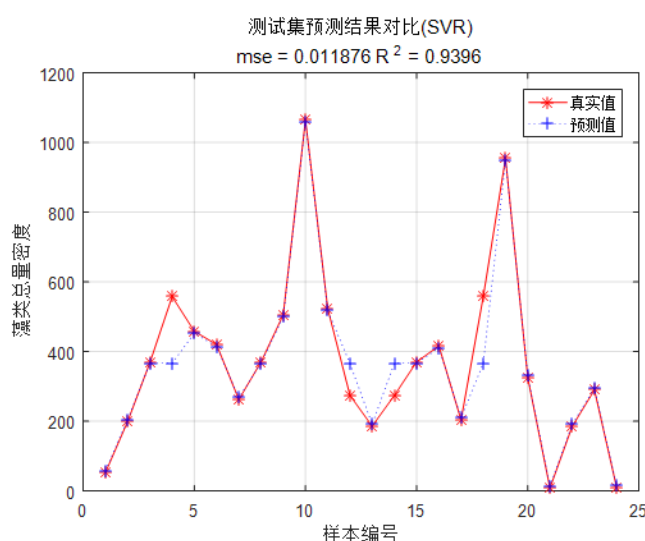


图 4 预测值与真实值对比

从上图可知， R^2 为 0.9396，均方差很小，预测值与真实值贴合较好，用该模型来拟合的预测值十分可信。

5.3.5 水华发生的理化因子范围确定

对于确定水华发生时各理化因子的范围，本文使用**网格搜索**的方法。

网格搜索会调控一个自变量因子按一定步长变化而控制其他因子不变，直到符合因变量的一定要求即停止。本问中我们以国家水质标准中 V 类水的理化因子界限为起始搜索点，高优指标（或称高优因子，数值越大越理想的指标）向低数值搜索，低优指标向高数值搜索，当计算得到的总藻类密度达到 554.71×10^6 个/L（发生水华的临界密度）为止。由此逐一计算各个因子得到高优指标的上界和低优指标的下界，最终搜索结果如表 9 所示：

表 9 主要理化因子波动范围

| 总氮 | 总磷 | 铵态氮 | 溶氧 | COD | PH |
|-------------|-------------|-------------|-------------|--------------|------|
| ≥ 1.71 | ≥ 6.82 | ≥ 1.34 | ≤ 9.36 | ≥ 20.62 | 9.33 |

5.4 问题四的模型建立与求解

对于鱼类生长与体重相关模型，本文根据附件 6 的鱼类体重体长数据，绘制出了散点图，发现存在部分异常值，剔除了这些异常数据后注意到两者的关系近于一个指数关系，故考虑使用指数模型进行回归拟合。

水华现象的产生与藻类的密度具有直接关系，构造一个与 1 号池相同大小的净化池，通过水循环，并放养鲢鱼或鳙鱼来净化藻类，考虑生物的繁殖变化，本文建立了差分阻滞增长模型来模拟池塘中两种生物的变化过程。

5.4.1 鱼类身长与体重相关模型

根据附件 6 中鱼类体重体长数据，本文以体长为自变量 x ，体重为因变量 y ，绘制了散点图如下：

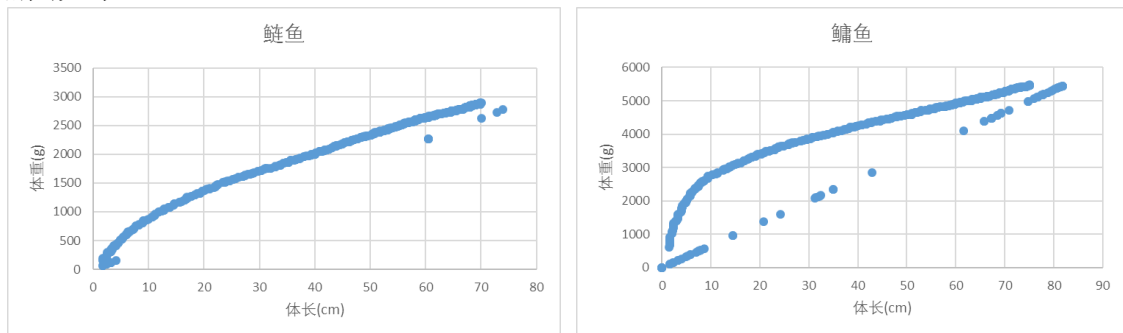


图 5 鲢鱼、鳙鱼体长体重散点图

根据散点图，可以观察到两种鱼的多数数据存在一个指数关系，同时也可以发现鲢鱼与鳙鱼的体重体长数据存在部分的异常值，本文采用 DBSCAN 算法对异常数据进行了剔除。

剔除异常数据后，本文对鲢鱼和鳙鱼的体重体长数据使用指数方程进行了回归拟合。得到结果如下：

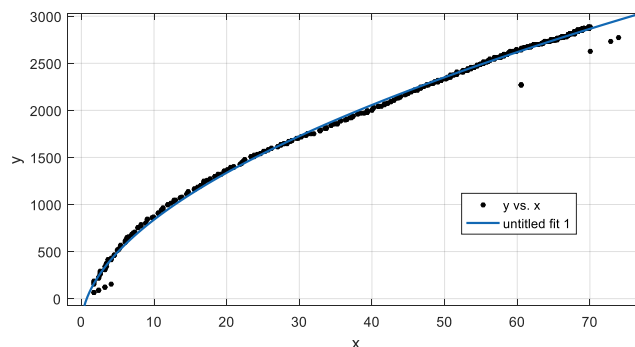


图 6 鲢鱼体重体长拟合效果图

鲢鱼体重体长相关方程：

$$y = 329.6x^{0.5311} - 280.9 \quad x \in [1.5, 71]$$

其中，鲢鱼体长为自变量 x ，体重为因变量 y ，使用 matlab 拟合工具箱得到上述拟合结果，调整 R^2 为 0.9967。

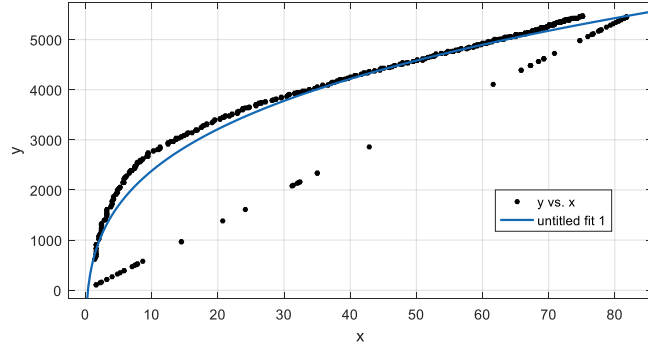


图 7 鳙鱼体重体长拟合效果图

鳙鱼体重体长相关方程：

$$y = 2152.6x^{0.2713} - 1638 \quad x \in [1.5, 71]$$

其中，鳙鱼体长为自变量 x ，体重为因变量 y ，使用 matlab 拟合工具箱得到上述拟合结果，调整 R^2 为 0.9224。

鲢鱼与鳙鱼体重体长的回归方程的调整 R^2 均大于 0.9，可见回归方程效果较好，但是注意到鳙鱼的拟合方程中，在方程回归曲线的下方仍有部分数据点存在，由于该部分数据比较分散，本文认为它部分数据不存在普遍性的统计规律，故不对它进行进一步的讨论分析。

5.4.2 差分阻滞增长模型

本文选择以投入鲢鱼来净化水质进行后续的研究。由于生物种群在自然条件下并非无限制的增长，本文考虑使用差分形式的阻滞增长模型来模拟水藻与鱼类种群的增长情况，

为讨论方便，本文视一周的时间为一个单位时间。假设第 k 时刻藻类总量密度为 A_k ，密度增长率为 γ_A ，其种群饱和密度为 N_A ，水循环对藻类密度的影响因子为 γ_W ，池中鲢鱼总质量为 F_k ，鲢鱼群体的质量增长率为 γ_F ，鲢鱼种群的饱和质量为 N_F ，鱼龄为 1、2、3 年的鲢鱼占比为 $\sigma_1, \sigma_2, \sigma_3$ 质量大小对藻类密度的影响因子分别为 $\gamma_1, \gamma_2, \gamma_3 \in [0, 1]$ ，由于在种群中，鱼龄为 4 年的鲢鱼所占比例仅为 0.87%，故不将其计入影响中。

在一个单位时间后，藻类密度的变化来自其自身的种群增长，水循环的抑制作用以及鲢鱼的净化作用。可建立差分方程模型如下：

$$\begin{cases} A_{k+1} - A_k = \gamma_A A_k \left(1 - \frac{A_k}{N_A}\right) - F_k \left(\sum_{i=1}^3 \sigma_i \gamma_i\right) - \gamma_W A_k \\ F_{k+1} - F_k = \gamma_F F_k \left(1 - \frac{F_k}{N_F}\right) \\ A_0 = 954.7 \quad k = 1, 2 \cdots n \end{cases}$$

对于上述差分阻滞增长模型，可以采用数值迭代求解。查阅相关文献^[8]，可知 γ_W 约为 0.13，观察到附件中，水藻总量密度的最高值为 $1064.82 \times 10^6/L$ ，故设一号池的种群饱和密度为 $1100 \times 10^6/L$ ，根据附件 5 的产量数据进行换算来作为鲢鱼群体的饱和质量为 $2058kg$ ，根据附件 7 可得鲢鱼种群中各鱼龄鱼类的占比为 $\sigma_1 = 5.47\%, \sigma_2 = 63.46\%, \sigma_3 = 30.19\%$ ，依据其增速关系可得 $\gamma_1 = 0.4622, \gamma_2 = 0.3109, \gamma_3 = 0.2269$ 。

对一号池进行 20 次数值迭代求解差分方程。对于上述模型而言，要求投入鲢鱼种

群后，总藻类的密度得到控制，若投入的鲢鱼种群质量越大，则藻类的控制越好，然而当藻类的数量减少时，鲢鱼的生长发育受到食物短缺因素的影响，也会逐渐降低，另一方面，由于鲢鱼的减少，少量的藻类会进行快速繁殖，总藻类的密度便又会上升，所以水藻总量密度与鲢鱼质量间存在一个动态的平衡关系。故假设总藻类密度较少时，对鲢鱼的抑制作用因子为 μ ，本文对差分方程中的鲢鱼质量差分进行如下修正：

$$F_{k+1} - F_k = \gamma_F F_k \left(1 - \frac{F_k}{N_F}\right) - \mu A_k$$

视密度为 $500 \times 10^6/L$ 为水藻丰足点，当密度值小于该点时，抑制因子 $\mu = 0.5$ ，反之则为 0。以0.1为步长搜索 $100kg$ 到 $1000kg$ 之间的鲢鱼投入量，取 20 期后最低总藻类含量的投入值即为最优的投入量（程序见附录 6）。得到结果如下：

表 10 投入量选择表

| 投入量 (kg) | 166. 2 | 166. 3 | 166. 4 | 166. 5 | 166. 6 |
|-----------------------------|---------|---------|---------|---------|---------|
| 20 期后总藻类 密度 ($10^6/L$) | 286. 19 | 285. 72 | 285. 49 | 285. 68 | 285. 95 |

可见投入 $166.4kg$ 鲢鱼时，总藻类密度为 $285.49 \times 10^6/L$ ，该方案可以很好地控制总藻类密度，达到抑制水华发生的目的。在该投放量下观察 20 期内总藻类密度变化的表格：

表 11 总藻类密度变化表

| 期数 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 总藻类密度值 | 954. 7 | 629. 2 | 949. 0 | 605. 8 | 908. 4 | 627. 7 | 847. 2 | 650. 6 | 765. 0 | 656. 9 |
| 期数 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 总藻类密度值 | 574. 7 | 524. 7 | 495. 4 | 459. 0 | 421. 4 | 381. 6 | 337. 9 | 286. 4 | 285. 6 | 285. 5 |

由上表可见，在前 10 期的模拟变化中，总藻类密度值在震荡波动，并有逐渐下降的趋势，这是由于鲢鱼生长繁殖以及藻类数量增加后单位时间内繁殖量加大等因素共同作用所引起的。在后 10 期的变化中，总藻类密度逐渐得到了控制，并且稳定在了 $285 \times 10^6/L$ 左右，投入 $166.4kg$ 鲢鱼可以在较短的时间内控制藻类植物的密度，有效地防止了水华现象的发生。

5. 4. 3 差分阻滞增长模型的敏感性分析：

对于差分阻滞增长模型中总水藻密度对鲢鱼的抑制因子 μ 进行敏感性分析。

取水藻丰足点分别为 200、300、400($\times 10^6/L$)， μ 以 0.1 为步长在区间[0,1)内进行取值，以 166.4kg 鲢鱼的投入量代入差分方程模型，得到 20 期后总藻类密度结果如下表所示：

表 12 灵敏度分析表

| μ | 丰足点总藻类密度 | | | | μ | 丰足点总藻类密度 | | | |
|-------|----------|-------|-------|-------|-------|----------|-------|-------|-------|
| | 200 | 300 | 400 | 500 | | 200 | 300 | 400 | 500 |
| 0.1 | 306.3 | 306.5 | 307.0 | 310.2 | 0.6 | 272.0 | 275.0 | 275.4 | 276.0 |
| 0.2 | 301.0 | 303.5 | 304.3 | 306.7 | 0.7 | 272.6 | 275.2 | 275.5 | 277.3 |
| 0.3 | 296.9 | 301.2 | 301.6 | 303.3 | 0.8 | 283.3 | 287.4 | 288.0 | 290.9 |
| 0.4 | 279.7 | 283.0 | 283.8 | 290.9 | 0.9 | 292.3 | 295.6 | 296.5 | 300.3 |
| 0.5 | 272.4 | 279.2 | 282.7 | 285.5 | / | | | | |

由上表可见，随着丰足点数值的下降，20 期后藻类总量普遍随之减小，这可能是由于丰足点减小，即当总藻类密度达到丰足点时，藻类总量的基数比较小，而鲢鱼的基数较大，使得反馈调节时，藻类在生态系统中处于不利的位置，而始终无法使得种群繁殖壮大。在实际中，若总藻类的密度过低，则对处于食物链上层的鲢鱼的反馈调节作用是较高的，故在上表中，抑制因子较低的部分不具有很大现实意义。

另一方面，就抑制因子的变化来看，无论抑制因子是过高或者过低都会使得总藻类密度增高。这可能是由于抑制因子过低时，对鲢鱼群体的反馈作用减弱，鲢鱼群体增长的同时导致食物短缺情况的出现，鲢鱼数量因此减少，从而使得藻类得以繁殖。抑制因子过高时，对鲢鱼群体的限制作用很大，从而使得鲢鱼群体数量在短时间内迅速减少，藻类密度因而提高。

总而言之，求解差分阻滞增长模型所得结果，即投入166.4kg的鲢鱼，在参数改变的情况下，20 期的表现较好，最终得到的总藻类密度在 $(285.5 \pm 30) \times 10^6/L$ 内波动，本文认为在 $285.5 \times 10^6/L$ 的基数下， $30 \times 10^6/L$ 的藻类数量波动对整个池塘的富营养化而言，其影响较小，故可认为使用水循环，加上投入166.4kg的鲢鱼的方案对池中的藻类具有较好的控制效果。

5.5 问题五的模型建立与求解

5.5.1 生态养殖模式分析

为了构建生态养殖模式，我们需要使水中生物构成一条和谐的生物链，实现池水的自净化，防止水华的发生，减少向江河湖海养殖废水的排放。根据相关资料^[9]可知，以虾、鱼、蟹、藻为基础的生态养殖池能够获得很好的效果，在维持水体正常营养的同时，显著提高了饲料利用率和经济效益。因此本文将以此为基础，在一号池的基础上构建一个生态养殖池，研究该方法对水体净化的作用。

考虑到多种生物之间具有复杂的捕食和共生关系，且不同区域的水环境之间也会有相互作用，常见的方程不容易简单描述，故考虑采用元胞自动机进行模拟，以总藻类密度作为指标，分析虾、鱼、蟹、藻生长情况。

5.5.2 基于元胞自动机的生态养殖模型

元胞自动机 (Cellular Automaton, CA) 是由Von Neumann提出的一种在时间、空间、状态上都离散的动力系统，利用相邻元胞之间的互相作用关系，能够以极简单的规则模拟复杂的行为，很适合用于生态养殖池的水质分析^[10]。

一个完整的元胞自动机系统通常包括元胞维数 (D)、邻居 (N)、状态 (S) 及演化规则 (f)，记为 $A = (D, N, S, f)$ 。其中，元胞是构成元胞自动机的最小单位，在本文中设置为 $S = (I, II, III, IV, V)$ 五种藻类密度状态，以总藻类密度作为衡量指标，具体划分如下， v 表示该元胞的总藻类密度值。

$$S_{ij} = \begin{cases} I & , 0 \leq v_{ij} < range_1 \\ II & , range_1 \leq v_{ij} < range_2 \\ III & , range_2 \leq v_{ij} < range_3, \\ IV & , range_3 \leq v_{ij} < range_4 \\ V & , range_4 \leq v_{ij} \end{cases} \quad range = (15, 50, 150, 500)$$

考虑到三维的元胞自动机太复杂，为了简单起见，本文采用维度为 $D = 2$ 。二维元胞自动机常见的邻居有Von Neumann型和Moore型两种，如图 8 所示，黑色为目标元

胞，灰色为其邻居。

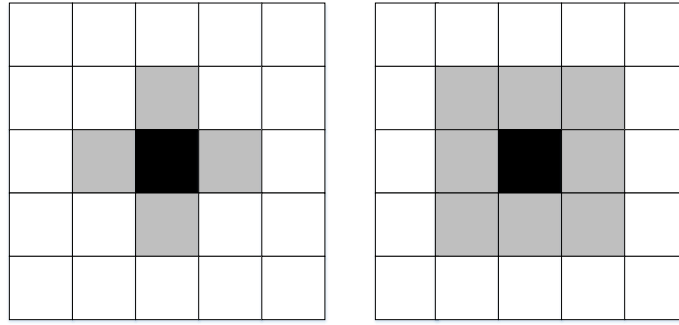


图 8 Von Neumann 型邻居（左）和 Moore 型邻居（右）

考虑到水体之间极强的流动性和分子扩散性，采用**Moore**型邻居会更合理。针对边界区域的处理，鉴于对象是水池，邻域间传播性很强，传统的定值型、周期型、反射型、绝热型四种边界可能不太合适，因此决定采用整个水池的藻类总值作为边界值。

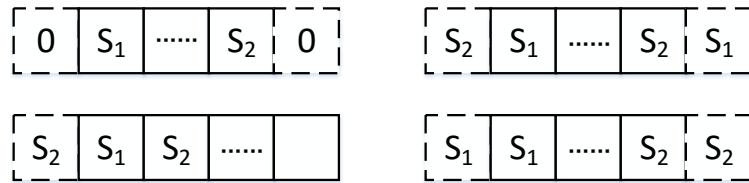


图 9 四种边界类型（左上：定值型，右上：周期型，左下：反射型，右下：绝热型）

由于所获得的原始数据是整个养殖池的藻类密度，为了构造二维元胞自动机，我们对数据进行随机交叉映射到二维上面，具体方法如下，其中 n 表示元胞自动机系统每一行的元胞个数。

$$\begin{cases} h = \text{randperm}(n) \\ v_{ij} = \frac{v'_i + v'_{h(j)}}{2} * (1 + \text{rand}) \end{cases}$$

关于元胞间演化规则，每个元胞下一阶段的状态取决于当前元胞及其邻居的状态。记相邻元胞影响的权重为 weight_1 ，自身影响的权重为 weight_2 ，鱼虾蟹类等影响系数为 λ ，鱼虾蟹类的数量为 ρ_1, ρ_2, ρ_3 ，不同状态下的消耗系数为 α_i ，成长系数为 β_i ，则演化规则的方程满足以下等式：

$$\begin{cases} z'_{i,j} = \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} z_{i,j} \times f(\text{weight}, x, y, i, j) \times \lambda \\ \lambda = 1 - \text{rand} \times (\rho_1, \rho_2, \rho_3) \times \alpha_i, \quad i = 1, 2, \dots, 5 \\ \rho_t = \rho_{t-1} \times \beta_i, \quad i = 1, 2, \dots, 5 \\ f(\text{weight}, x, y, i, j) = \begin{cases} \text{weight}_1, & (x, y) \neq (i, j) \\ \text{weight}_2, & (x, y) = (i, j) \end{cases} \\ 8 \times \text{weight}_1 + \text{weight}_2 = 1 \end{cases}$$

对于鱼虾蟹类之间的生长关系，由之前的模型和相关资料^[11]可知，在不多投放饲料的情况下，蟹类主要以鱼虾等动物尸体为食，所以其数目会受到鱼虾养殖量的约束，因此蟹类的数量受总藻类密度的间接影响，由此构造以下生长模式：

$$\begin{cases} \rho'_1 = \rho_1 \times \beta_1 \\ \rho'_2 = \rho_2 \times \beta_2 \\ \rho'_3 = \rho_3 \times \beta_3 \times (\beta_1 + \beta_2) \end{cases}$$

根据以上模型，构造元胞自动机，对生态养殖池的情况进行模拟(程序请见附录 7)。

5.5.3 生态养殖池模拟结果

基于前面问题的解决，设置虾、鱼、蟹的养殖量分别为 500 公斤、300 公斤、500 公斤，并分三种情况讨论。

情况一：对水华严重爆发的一号养殖池进行模拟，得到如下结果：

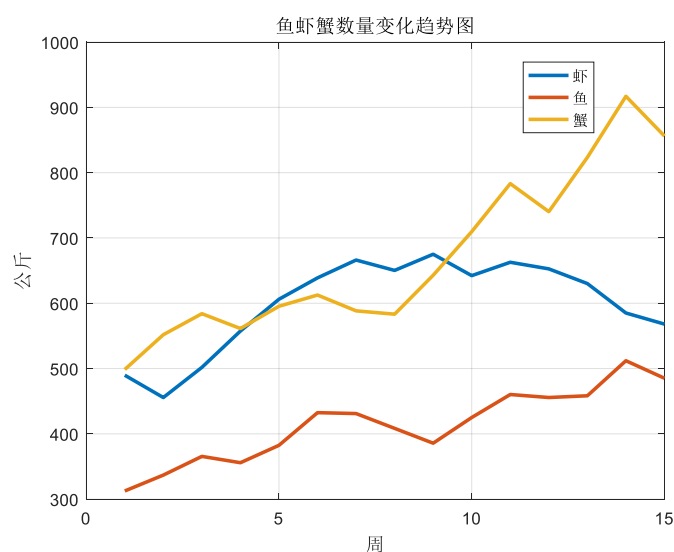
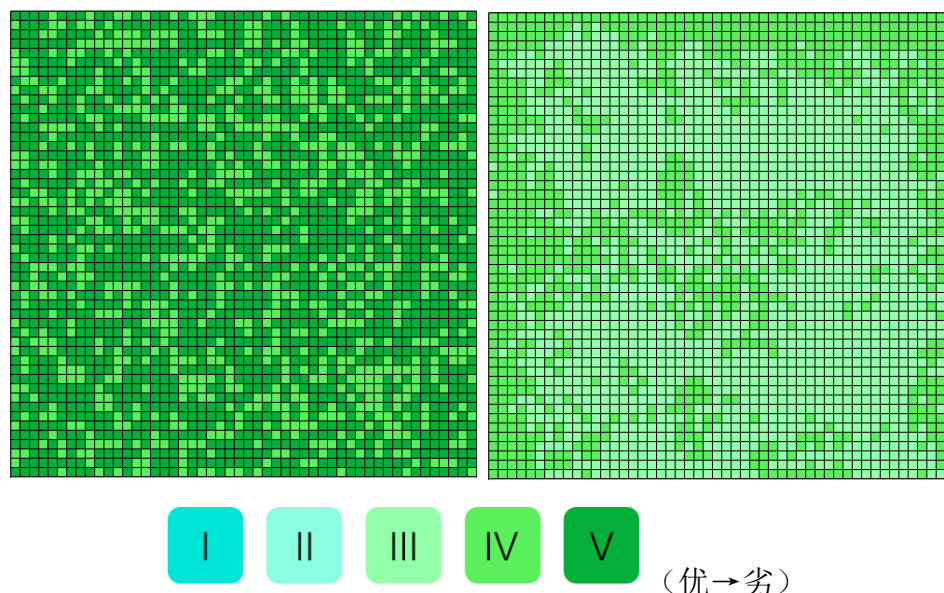


图 10 水华严重情况下模拟结果

根据以上结果可知，在对一号养殖池进行 15 周模拟的情况下，鱼、虾、蟹的数量都呈现出增长趋势，此时总藻类含量也在逐步下降，由最初的水华大爆发变为较轻程度的水华，说明该生态养殖模式在水华爆发时能够起到很好的抑制作用，并相应的获得不错的经济收益。另一方面，蟹的增长明显受到鱼虾增长的限制，具有一定滞后性，说明模型比较符合实际情况。

情况二：考虑水华轻微发生的时候，该生态养殖模式的发展情况。

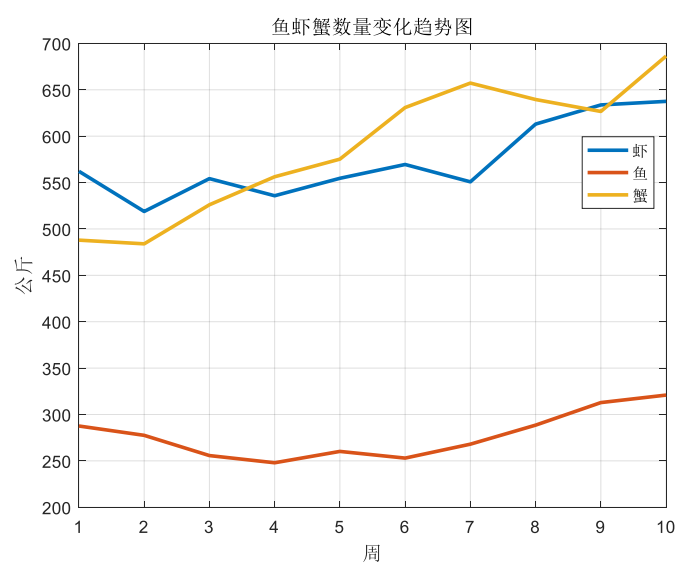
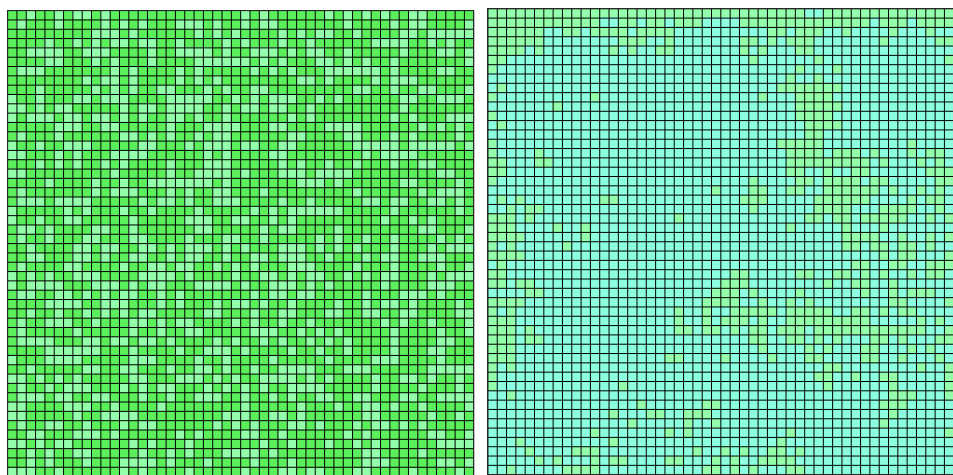
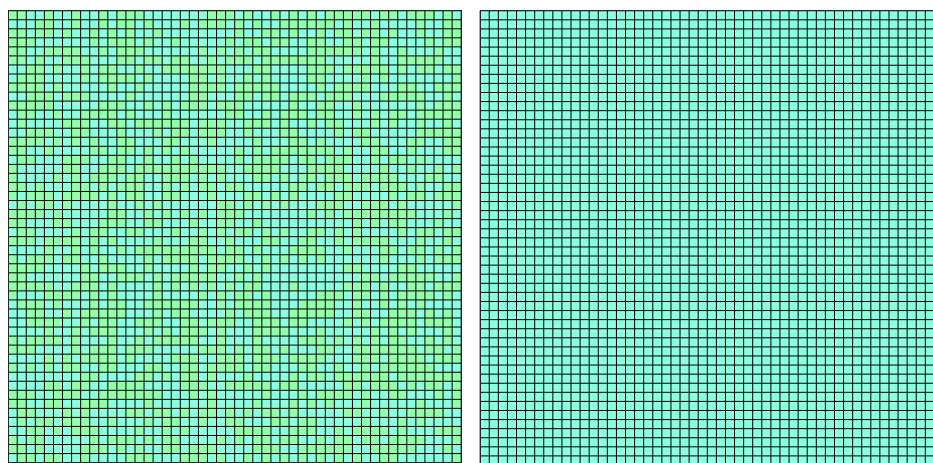


图 11 水华轻微情况下模拟结果

此时该生态养殖池依然具有不错的净化能力，大部分水质都有很大的改善，只是鱼、虾、蟹类的数量增长不如之前那么快。说明该系统此时的效果不错，能够有效改善水质。鱼虾的增长开始放缓，偶尔有波动。

情况三：对于无水华的情况进行模拟。



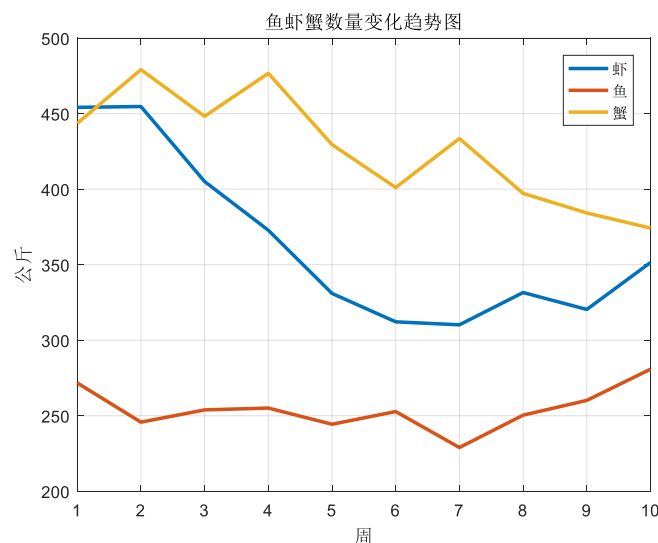


图 12 无水华情况下模拟结果

此时，经过 10 周的模拟，水质有了很大的改善，鱼、虾、蟹的数量有不同程度的下降，说明此时藻类的含量以及不足以支撑鱼、虾、蟹的生长，需要人工投放一定量的饲料。

综合以上模拟结果可知，该鱼、虾、蟹投入量能够有效应对水华大爆发及一般水华的情况，实现池水自净化的目标，减少向江河湖海养殖废水的排放，同时有一定的经济效益，不需要太多的人工干预。在无水华的情况下，可能需要人工投放一定饲料，或者采取适当地捕捞工作，减少鱼、虾、蟹的养殖量，以达到生态养殖池自给自足的目的。

6 模型优缺点

✧ 模型优点：

本文第一问中采用的皮尔森相关性分析，可以很好地测度两个变量间的相关性关系，具有坚实的数学理论基础。

在第二问中使用的投影寻踪模型选择将高维数据投影到一个低维空间，根据投影值来分析高维数据的分类结构特征，避免了高维数带来的计算复杂性，求解使用的基于实数编码的加速遗传算法在传统的遗传算法上进行了修改，减少了求解中计算量，使得求解的速度更为快捷。

第三问中建立的支持向量回归机模型，作为机器学习领域一种相对较为成熟的数据挖掘手段，具有坚实的理论基础，并且可以很好地表征数据间的深度联系。

而元胞自动机模型在数据的反映能力上形象直观，并且作为一个动态模型，它可以十分清楚地模拟事物的变化情况。

✧ 存在的问题：

本文的缺点在于，题中所提供的数据量较少，在支持向量回归机训练过程中，数据量对模型的精度存在一定的限制，若存在更多的数据，训练所得结果会更加贴近实际。

7 模型推广

本次建模中所使用投影寻踪模型可很好地处理高维数据，在数据挖掘领域具有广泛的实用性。基于实数编码的加速遗传算法具有很强的兼容性和扩展性，可以与其他很多算法相结合进行数据的分析处理，可以模拟复杂的优化问题，在函数优化、组合优化等领域中均已得到成功的应用。

而本文在预测时采用的支持向量机作为一个机器学习方法可以有效地解决分类、回归等数据挖掘问题，在天气评价，军事，产业结构、效能评估等领域均具有重要的借鉴意义。

处理第五问采用的元胞自动机模型，作为一种特殊的动力系统，可用作仿真任意复杂的计算过程，可以直观地表现事物的变化情况，在社会学、生物学、地理学中也具有广泛的应用。

8 参考文献

- [1]马健荣,邓建明,秦伯强,龙胜兴. 湖泊蓝藻水华发生机理研究进展[J]. 生态学报,2013,10:3020-3030.
- [2]刘峰. 养殖水体水华发生的原因探讨[D].东北农业大学,2007.
- [3]蔡继晗,沈奇宇,郑向勇,李凯,吕永林,陈强. 氨氮污染对水产养殖的危害及处理技术研究进展[J]. 浙江海洋学院学报(自然科学版),2010,02:167-172+195.
- [4] 赵小勇. 投影寻踪模型及其在水土资源中的应用[D].东北农业大学, 2006.
- [5] 金芬. 遗传算法在函数优化中的应用研究[D].苏州大学, 2008.
- [6] 金菊良,杨晓华,丁晶. 基于实数编码的加速遗传算法[J]. 四川大学学报(工程科学版), 2000,04:20-24.
- [7] 曾绍华.支持向量机回归机算法理论研究与应用[D].重庆大学,2006.
- [8]江成. 水体扰动对藻生长机制与 QCS 水库富营养化控制的影响研究[D].上海交通大学,2014.
- [9]申玉春,熊邦喜,叶富良,阮芳. 虾-鱼-贝-藻生态优化养殖及其水质生物调控技术研究[J]. 生态学杂志,2005,06:613-618.
- [10] Wolfram S. Theory and applications of cellular automata[M]. Singapore: World Scientific, 1986
- [11]施炜纲,刘凯,张敏莹,徐东坡. 春季禁渔期间长江下游鱼虾蟹类物种多样性变动(2001-2004年)[J]. 湖泊科学,2005,02:169-175.

附录：

附录 1：Pearson 相关系数（SAS9. 4）

```
data ex;
input x1-x3 @@;
cards;
<数据略>
;
proc corr data=ex pearson;
run;
```

附录 2：三次样条插值（Matlab2016a -interp1）

各水池插值数据表

| 1# | 溶氧 | COD | PH | 2# | 溶氧 | COD | PH |
|----|-------|--------|-------|----|-------|--------|-------|
| 1 | 5.12 | 21.9 | 9.31 | 1 | 5.21 | 18.21 | 8.75 |
| 2 | 4.16 | 20.95 | 9.225 | 2 | 5.43 | 18.855 | 8.62 |
| 3 | 3.2 | 20 | 9.14 | 3 | 5.65 | 19.5 | 8.49 |
| 4 | 4.96 | 23.4 | 9.14 | 4 | 5.705 | 20.545 | 8.57 |
| 5 | 6.72 | 26.8 | 9.14 | 5 | 5.76 | 21.59 | 8.65 |
| 6 | 5.04 | 27.265 | 9.215 | 6 | 6.08 | 20.215 | 8.715 |
| 7 | 3.36 | 27.73 | 9.29 | 7 | 6.4 | 18.84 | 8.78 |
| 8 | 2.88 | 25.565 | 9.255 | 8 | 5.43 | 19.36 | 8.735 |
| 9 | 2.4 | 23.4 | 9.22 | 9 | 4.46 | 19.88 | 8.69 |
| 10 | 3.27 | 23.075 | 9.275 | 10 | 5.89 | 21.795 | 8.68 |
| 11 | 4.14 | 22.75 | 9.33 | 11 | 7.32 | 23.71 | 8.67 |
| 12 | 5.285 | 24.055 | 9.245 | 12 | 6.645 | 24.53 | 8.645 |
| 13 | 6.43 | 25.36 | 9.16 | 13 | 5.97 | 25.35 | 8.62 |
| 14 | 5.515 | 25.695 | 9.21 | 14 | 5.895 | 23.93 | 8.695 |
| 15 | 4.6 | 26.03 | 9.26 | 15 | 5.82 | 22.51 | 8.77 |

| 3# | 溶氧 | COD | PH | 4# | 溶氧 | COD | PH |
|----|------|--------|-------|----|------|--------|------|
| 1 | 9.12 | 20.9 | 9.28 | 1 | 9.28 | 24.6 | 9.1 |
| 2 | 7.92 | 20.45 | 9.125 | 2 | 7.36 | 23.85 | 9.15 |
| 3 | 6.72 | 20 | 8.97 | 3 | 5.44 | 23.1 | 9.2 |
| 4 | 4.96 | 23.4 | 9.055 | 4 | 4.8 | 27.74 | 9.15 |
| 5 | 3.2 | 26.8 | 9.14 | 5 | 4.16 | 32.38 | 9.1 |
| 6 | 3.28 | 27.015 | 9.15 | 6 | 6.16 | 31.475 | 9.16 |
| 7 | 3.36 | 27.23 | 9.16 | 7 | 8.16 | 30.57 | 9.22 |
| 8 | 2.88 | 27.315 | 9.145 | 8 | 5.52 | 28.62 | 9.16 |
| 9 | 2.4 | 27.4 | 9.13 | 9 | 2.88 | 26.67 | 9.1 |
| 10 | 3.28 | 25.075 | 9.19 | 10 | 3.51 | 26.97 | 9.14 |
| 11 | 4.16 | 22.75 | 9.25 | 11 | 4.14 | 27.27 | 9.18 |

| | | | | | | | |
|----|-------|--------|-------|----|------|--------|-------|
| 12 | 5.295 | 23.555 | 9.205 | 12 | 4.83 | 24.975 | 9.155 |
| 13 | 6.43 | 24.36 | 9.16 | 13 | 5.52 | 22.68 | 9.13 |
| 14 | 5.515 | 25.195 | 9.07 | 14 | 4.29 | 24.525 | 9.09 |
| 15 | 4.6 | 26.03 | 8.98 | 15 | 3.06 | 26.37 | 9.05 |

附录 3: 投影寻踪 (Matlab2016a)

```
clear;
clc;
close all;
d=[];e=[];
load Variable
X=Variable;

%=====      数      据      的      同      趋      化      处      理
=====

normal = [0 0 0 0 0 0];
normal1 = [0.02 0.2 0.15 7.5 15 6];
normal2 = [0.1 0.5 0.5 6 15 6];
normal3 = [0.2 1 1 5 20 6];
normal4 = [0.3 1.5 1.5 3 30 6];
normal5 = [0.4 2 2 2 40 6];
X = [X;normal1;normal2;normal3;normal4;normal5];

[m,n]=size(X);
low=[1 2 3 4];high=[5];
%归一化
for i=1:m
    for j=1:size(high,2)
        x(i,high(j))=(X(i,high(j))-
normal(high(j)))/(max(X(:,high(j)))-normal(high(j)));
    end
end
for i=1:m
    for j=1:size(low,2)
        x(i,low(j))=(max(X(:,low(j)))-
X(i,low(j)))/(max(X(:,low(j)))-normal(low(j)));
    end
end
x(:,6) = abs(X(:,6) - 7.5);
x(:,6) = 1 - x(:,6) ./max(x(:,6));
for k=1:50
```

```

N=400;Pc=0.8;Pm=0.2;M=10;Ci=7;n=6;DaiNo=2;ads=1;
[a1,b1,ee,ff]=RAGA(x,N,n,Pc,Pm,M,DaiNo,Ci,ads);
d=[d,a1];e=[b1;e];
end
[a2 b2]=max(d);
e1=e(b2,:);
ff=e1*x';
figure(1);

title('样本评价图');

plot(1:15,ff(1,1:15),'k+');
hold on;
plot(16:30,ff(1,16:30),'r*');
plot(31:45,ff(1,31:45),'bo');
plot(46:60,ff(1,46:60),'gx');

legend('一号池样本','二号池样本','三号池样本','四号池样本');

xlabel('样本编号');

ylabel('评价值大小');

line([0,60],[ff(61),ff(61)]);
line([0,60],[ff(62),ff(62)]);
line([0,60],[ff(63),ff(63)]);
line([0,60],[ff(64),ff(64)]);
line([0,60],[ff(65),ff(65)]);
hold off;

function
[xStar,fStar,Lmin,Hmax]=RAGA(xx,N,n,Pc,Pm,M,K,Ci,ads)
if ads==0
    ad='ascend';
else
    ad='descend';
end

Low=zeros(1,n);High=ones(1,n);
for z=1:Ci
    for i=1:N
        while 1==1
            for p=1:n
                temp(p)=unifrnd(Low(p),High(p));
            end

```

```

        temp=sum(temp.^2);
        xStar=sqrt(temp.^2/temp);
        y=Feasibility(xStar);
        if y==1
            v(i,:)=xStar;
            break;
        end
    end
end

for s=1:K

    for i=1:N
        Favrite(i)=Target(xx,v(i,:));
    end

    [Favrite,i]=sort(Favrite,ad);
    v=v(i,:);
    %step2 end

    alpha=0.05;
    q(1)=0;
    for i=2:N+1
        q(i)=q(i-1)+alpha*(1-alpha)^(i-2);
    end
    for i=1:N
        r=unifrnd(0,q(N+1));
        for j=2:N+1
            if r>q(j-1) & r<=q(j)
                Mytemp1(i,:)=v(j-1,:);
            end
        end
    end

    while 1==1
        CrossNo=0;
        v1=Mytemp1;
        for i=1:N
            r1=unifrnd(0,1);
            if r1 < Pc
                CrossNo=CrossNo+1;
                Mytemp2(CrossNo,:)=v1(i,:);
                v1(i,:)=zeros(1,n);
            end
        end
    end
end

```

```

        end
    end
    if CrossNo~=0 & mod(CrossNo,2)==0
        break;
    elseif CrossNo==0|mod(CrossNo,2)==1
        Mytemp2=[];
    end
end
shengyuNo=0;
for i=1:N
    if v1(i,:)~=zeros(1,n)
        shengyuNo=shengyuNo+1;
        Mytemp3(shengyuNo,:)=v1(i,:);
    end
end

for i=1:CrossNo
    r2=ceil(unifrnd(0,1)*(CrossNo-i+1));
    Mytemp4(i,:)=Mytemp2(r2,:);
    Mytemp2(r2,:)=[];
end

for i=1:2:(CrossNo-1)
    while 1==1
        r3=unifrnd(0,1);
        v20(i,:)=r3*Mytemp4(i,:)+(1-r3)*Mytemp4(i+1,:);
        v20(i+1,:)=(1-
r3)*Mytemp4(i,:)+r3*Mytemp4(i+1,:);
        temp1=sum(v20(i,:).^2);
        temp2=sum(v20(i+1,:).^2);
        v2(i,:)=sqrt(v20(i,:).^2/temp1);
        v2(i+1,:)=sqrt(v20(i+1,:).^2/temp2);
        if Feasibility(v2(i,:))==1 &
Feasibility(v2(i+1,:))==1
            break;
        end
    end
end
end
%step5 end
v3=[Mytemp3;v2];
while 1==1
    MutationNo=0;
    v4=v3;

```

```

        for i=1:N
            r4=unifrnd(0,1);
            if r4<Pm
                MutationNo=MutationNo+1;
                vtemp5(MutationNo,:)=v4(i,:);
                v4(i,:)=zeros(1,n);
            end
        end
        if MutationNo~=0
            break;
        end
    end
end

Loss=0;
for i=1:N
    if v4(i,:)~=zeros(1,n)
        Loss=Loss+1;
        Mytemp6(Loss,:)=v4(i,:);
    end
end

DirectionV=unifrnd(-1,1,1,n);
for i=1:MutationNo
    tempNo=0;
    while 1==1
        tempNo=tempNo+1;

v5(i,:)=sqrt(((vtemp5(i,:)+M*DirectionV).^2)./sum((vtemp5(i,:)+M*DirectionV).^2));
        y=Feasibility(v5(i,:));
        if tempNo==200
            v5(i,:)=vtemp5(i,:);
            break;
        elseif y==1
            break;
        end
        M=unifrnd(0,M);
    end
end
end
%step6 end
vk=[v5;Mytemp6];
v=vk;
end

```

```

    for i=1:N;
        Favrite(i)=Target(xx,v(i,:));
    end

    [Favrite,i]=sort(Favrite,ad);
    v=v(i,:);
    vk=v;
    Nice=vk(1:20,:);
    t=1:n;
    Low(t)=min(Nice(:,t));
    High(t)=max(Nice(:,t));
    Lmin(z,:)=Low;
    Hmax(z,:)=High;
    if abs(Low-High)<=0.00001
        break;
    end
end
xStar=Favrite(1);
fStar=Nice(1,:);

```

```

function y=Feasibility(a)
b=sum(a.^2);
if abs(b-1)<=0.00001
    y=1;
else
    y=0;
end

```

```

function y=Target(x,a)
[m,n]=size(x);
for i=1:m
    temp1=0;
    for j=1:n
        temp1=temp1+a(j)*x(i,j);
    end
    z(i)=temp1;
end
Sz=std(z);
R=0.1*Sz;
temp2=0;
for i=1:m

```

```

for j=1:m
    r=abs(z(i)-z(j));
    t=R-r;
    if t>=0
        u=1;
    else
        u=0;
    end
    temp2=temp2+t*u;
end
end
Dz=temp2;
y=Sz*Dz;

```

附录4： 投影寻踪评价值（Matlab2016a）

| 各水池评价值 | | | |
|----------|----------|----------|----------|
| 一号水池 | 二号水池 | 三号水池 | 四号水池 |
| 1.172889 | 1.326263 | 1.237948 | 1.275469 |
| 1.490496 | 1.451523 | 1.454539 | 1.479945 |
| 1.608093 | 1.608198 | 1.434325 | 1.487064 |
| 1.498412 | 1.273098 | 1.120787 | 1.229696 |
| 1.466202 | 1.376882 | 1.292448 | 1.230765 |
| 1.474639 | 1.26229 | 1.020178 | 0.629912 |
| 1.593014 | 1.478782 | 1.485567 | 0.106438 |
| 1.24718 | 1.383694 | 1.499358 | 0.247815 |
| 0.889599 | 1.222392 | 1.274249 | 0.634788 |
| 1.315203 | 1.326347 | 1.222796 | 1.017751 |
| 1.285178 | 1.383302 | 1.439195 | 1.274099 |
| 1.271968 | 1.507355 | 1.376053 | 1.237568 |
| 1.105202 | 1.391723 | 1.280694 | 1.332014 |
| 1.060516 | 1.312745 | 1.273261 | 1.153444 |
| 1.14089 | 1.343123 | 1.451484 | 1.127406 |

附录5： SVR（Matlab2016a）

```

clear;
clc;
close all;
load Variable
attributes = Variable(:,1:6);
strength = Variable(:,7);
attributes=attributes';
strength=strength';
n = randperm(size(attributes,2));
p_train = attributes(:,n(1:96))';

```



```

t_train = strength(:,n(1:96))';
p_test = attributes(:,n(97:end))';
t_test = strength(:,n(97:end))';

[pn_train,inputps] = mapminmax(p_train');
pn_train = pn_train';
pn_test = mapminmax('apply',p_test',inputps);
pn_test = pn_test';

[tn_train,outputps] = mapminmax(t_train');
tn_train = tn_train';
tn_test = mapminmax('apply',t_test',outputps);
tn_test = tn_test';

[c,g] = meshgrid(-10:0.5:10,-10:0.5:10);
[m,n] = size(c);
cg = zeros(m,n);
eps = 10^(-9);
v = 5;
bestc = 0;
bestg = 0;
error = Inf;
for i = 1:m
    for j = 1:n
        cmd = ['-v ',num2str(v),' -t 2',' -c ',num2str(2^c(i,j)),' -g ',num2str(2^g(i,j)),' -s 3 -p 0.1'];
        cg(i,j) = svmtrain(tn_train,pn_train,cmd);
        if cg(i,j) < error
            error = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
        if abs(cg(i,j) - error) <= eps && bestc > 2^c(i,j)
            error = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
    end
end
cmd = ['-t 2',' -c ',num2str(bestc),' -g ',num2str(bestg),' -s 3 -p 0.01'];
model = svmtrain(tn_train,pn_train,cmd);

```

```

[Predict_1,error_1] = svmpredict(tn_train,pn_train,model);
[Predict_2,error_2] = svmpredict(tn_test,pn_test,model);

predict_1 = mapminmax('reverse',Predict_1,outputps);
predict_2 = mapminmax('reverse',Predict_2,outputps);

result_1 = [t_train predict_1];
result_2 = [t_test predict_2];

%% 绘图
figure(1)
plot(1:length(t_train),t_train,'r-*',1:length(t_train),predict_1,'b:+')
grid on
legend('真实值','预测值')
xlabel('样本编号')
ylabel('藻类总量密度')
string_1 = {'训练集预测结果对比';
            ['mse = ' num2str(error_1(2)) ' R^2 = ' num2str(error_1(3))]};
title(string_1)
figure(2)
plot(1:length(t_test),t_test,'r-*',1:length(t_test),predict_2,'b:+')
grid on
legend('真实值','预测值')
xlabel('样本编号')
ylabel('藻类总量密度')
string_2 = {'测试集预测结果对比(SVR)';
            ['mse = ' num2str(error_2(2)) ' R^2 = ' num2str(error_2(3))]};
title(string_2)

load testVariable
pme1_train = mapminmax('apply',pme1_train,inputps);
pme1_train=pme1_train';
pme2_train = mapminmax('apply',pme2_train,inputps);
pme2_train=pme2_train';
pme3_train = mapminmax('apply',pme3_train,inputps);
pme3_train=pme3_train';
pme4_train = mapminmax('apply',pme4_train,inputps);
pme4_train=pme4_train';
[Pme1,error_3] =
svmpredict(zeros(size(pme1_train,1),1),pme1_train,model);
[Pme2,error_4] =
svmpredict(zeros(size(pme2_train,1),1),pme2_train,model);

```

```

[Pme3,error_3] =
svmpredict(zeros(size(pme3_train,1),1),pme3_train,model);
[Pme4,error_3] =
svmpredict(zeros(size(pme4_train,1),1),pme4_train,model);
p1 = mapminmax('reverse',Pme1,outputps);
p2 = mapminmax('reverse',Pme2,outputps);
p3 = mapminmax('reverse',Pme3,outputps);
p4 = mapminmax('reverse',Pme4,outputps);

```

附录6： 差分方程 (Matlab2016a)

```

function [A, F] = chafen(A0, F0, K, gammaA, gammaF, NA, NF, Mu, Delta)
A = zeros(1,K+1);
F = A;
A(1) = A0;
F(1) = F0;
rho = 0.13;
gamma = [0.4622 0.3109 0.2269];
alpha = [0.055 0.635 0.301];

for i = 2 : K+1
    mu = 0;
    if(A(i-1)<Delta)
        mu = Mu;
    end;
    F(i) = F(i-1)*(1+gammaF*(1-F(i-1)/NF))-mu*A(i-1);
    delta = alpha(1)*gamma(1)+alpha(2)*gamma(2)+alpha(3)*gamma(3);
    A(i) = A(i-1)*(1-rho+gammaA*(1-A(i-1)/NA)) - F(i-1)*delta;
end;

```

附录7： 元胞自动机 (Matlab2016a)

```

clear
clc
close all;
%%%%%%%%%%%%
rho = 50;
fish = [500 300 500];
length = 50;
lambda = 0.8;
T = 10;
alpha = [0.00001, 0.000015, 0.00035, 0.00025 0.0002];
range = [15, 50, 150, 500];
%%%%%%%%%%%%

```

```

area = MyCreate(length, rho, lambda);
DrawArea(area, range);
n = length;
plot([(0:n)', (0:n)'] + 0.5, [0,n] + 0.5, 'k');
plot([0,n] + 0.5, [(0:n)', (0:n)'] + 0.5, 'k');
axis image;
set(gca, 'xtick', []);
set(gca, 'ytick', []);
figure;
data = zeros(T, 3);
for i = 1 : T
    [area, fish] = MyChange(area, fish, range, alpha);
    DrawArea(area, range);
    drawnow;
    data(i, :) = fish;
end
plot([(0:n)', (0:n)'] + 0.5, [0,n] + 0.5, 'k');
plot([0,n] + 0.5, [(0:n)', (0:n)'] + 0.5, 'k');
axis image;
set(gca, 'xtick', []);
set(gca, 'ytick', []);
figure;
plot(data, 'LineWidth', 2);
grid on;
legend('虾', '鱼', '蟹');
title('鱼虾蟹数量变化趋势图');
xlabel('周');
ylabel('公斤');

```

```

function DrawArea(area, range)
temp = area;
n = length(temp);
Area(:, :, 1) = temp;
Area(:, :, 2) = temp;
Area(:, :, 3) = temp;
for i=1:n
    for j=1:n
        if temp(i, j) < range(1)
            Area(i, j, :) = [0 230 214];
        elseif temp(i, j) <= range(2)
            Area(i, j, :) = [139 255 223];
        elseif temp(i, j) <= range(3)
            Area(i, j, :) = [147 255 170];
        end
    end
end

```

```

elseif temp(i, j) <= range(4)
    Area(i,j,:) = [91 240 91];
else
    Area(i,j,:) = [4 176 58];
end
end
end
Area = uint8(Area);
p = imagesc(Area);
hold on;
end

function area = MyCreate(length, rho, lambda)
    area = ones(length) * rho;
    area = area .* (1 + (rand(length) - 0.5) * lambda );
    fun = sort(round(unifrnd(1,length,1,4)/3));
    area(fun(1):fun(2), fun(3):fun(4)) = area(fun(1):fun(2),
fun(3):fun(4))...
        * unifrnd(0.8, 1.2, 1, 1);
    fun = sort(round(unifrnd(1,length,1,4)/2));
    area(fun(1):fun(2), fun(3):fun(4)) = area(fun(1):fun(2),
fun(3):fun(4))...
        * unifrnd(0.8, 1.2, 1, 1);
    fun = sort(round(unifrnd(1,length,1,4)/2));
    area(fun(1):fun(2), fun(3):fun(4)) = area(fun(1):fun(2),
fun(3):fun(4))...
        * unifrnd(0.8, 1.2, 1, 1);
    fun = sort(round(unifrnd(1,length,1,4)/2));
    area(fun(1):fun(2), fun(3):fun(4)) = area(fun(1):fun(2),
fun(3):fun(4))...
        * unifrnd(0.8, 1.2, 1, 1);
End

function total = MyRound(area, i, j)
    n = length(area);
    Area = ones(n + 2) * area(i, j);
    Area(2:n+1, 2:n+1) = area;
    i = i + 1;
    j = j + 1;

```

```

    total = 0.2*Area(i, j) + 0.1*Area(i+1, j) + 0.1*Area(i-1, j) + ...
           0.1*Area(i, j+1) + 0.1*Area(i, j-1) + 0.1*Area(i+1, j+1) + ...
           0.1*Area(i+1, j-1) + 0.1*Area(i-1, j+1) + 0.1*Area(i-1, j-1);
End

function [area, fish] = MyChange(area, fish, range, alpha)
    n = length(area);
    for i = 1 : n
        for j = 1 : n
            if MyRound(area, i, j) < range(1)
                area(i, j) = MyRound(area, i, j) * (1 - rand * mean(fish) *
alpha(1));
            elseif MyRound(area, i, j) < range(2)
                area(i, j) = MyRound(area, i, j) * (1 - rand * mean(fish) *
alpha(2));
            elseif MyRound(area, i, j) < range(3)
                area(i, j) = MyRound(area, i, j) * (1 - rand * mean(fish) *
alpha(3));
            elseif MyRound(area, i, j) < range(3)
                area(i, j) = MyRound(area, i, j) * (1 - rand * mean(fish) *
alpha(4));
            else
                area(i, j) = MyRound(area, i, j) * (1 - rand * mean(fish) *
alpha(5));
            end
        end
    end
    avg = mean(mean(area));
    if avg < range(1)
        fish = fish * (1 + (rand - 0.5) * 0.01);
    elseif avg < range(2)
        fish = fish .* [1.006 1.005 1.007];
    elseif avg < range(3)
        fish = fish .* [1.013 1.011 1.013];
    elseif avg < range(4)
        fish = fish .* [1.03 1.03 1.03];
    else
        fish = fish .* [1.05 1.04 1.05];
    end
    fish = fish + fish .* unifrnd(-0.12, 0.1, 1, 3);
end

```