

# 水产养殖池塘综合研究

## 摘要:

本文基于 SPSS、Matlab 和 Lingo 软件, 综合利用相关性分析、主成分分析、因子分析、拟合、聚类分析等方法进行数据分析, 建立了 BP 神经网络、NAR 神经网络及线性规划等模型, 全面探讨了水产养殖池塘的 5 个问题。

问题 1 中, 将问题划分成两个维度, 横向探究各种理化因子自身间的相互影响转换关系。纵向探究各种理化因子在沉积物、间隙水、水体间的扩散转移。首先利用主成分分析法求出横向探究时的各理化因子的理化因子综合指标和纵向探究时各环境的池塘环境综合指标, 而后分别对两种综合指标做相关性分析, 判别各因子间的相关性关系。在此基础上, 进一步利用 matlab 拟合工具箱对各因子两两拟合, 深入探究其关系。最终发现, 总磷与总氮同步变化; 总磷与磷酸盐同步变化, 且总磷占据主导地位, 随着总磷的增多, 磷酸盐增多的速率逐渐增大, 其主要原因是水体中磷酸盐是磷的主要存在形式; 硝态氮与亚硝态氮具有“此消彼长”的关系, 其主要原因是因为在底泥和上层水体间不断发生硝化与反硝化作用; 底泥是各种理化因子的源, 底泥中各种含氮有机物不断分解, 含磷物质不断溶解是影响水体理化因子含量的主要原因, 从底泥到间隙水、从间隙水到池水存在理化因子的扩散转移。

问题 2 首先利用 SPSS 进行因子分析, 进一步确定综合评价函数求出水质综合评价值, 发现 1、2 池水质有差转好趋于稳定, 3 号池水质稳定, 4 号池水质由差转好并趋于稳定。对水质综合评价值进行系统聚类对水质进行分类, 判定 4 号池水质为优, 1 号和 3 号池水质为良, 2 号池水质为差。对附件 2 中各项指标进行逐一比对, 认为鱼池对水体的影响是减小水体轮虫密度, 并且以较大幅度扰动水体溶氧; 认为虾池对水体的主要影响是增大水体总碱度, 并使轮虫密度增大。

问题 3 中, 通过相关性分析得出池水中亚硝态氮、溶氧和 PH 是浮游植物密度总量的主要影响因子。以 3 个主要因子为输入, 以浮游植物总密度为输出, 建立单隐含层 BP 神经网络, 发现当隐含层节点数为 10 时预测误差最小, 因此以该神经网络为主要理化因子-浮游生物密度关系模型。接着以主要理化因子和浮游生物密度的相关关系为切入点, 利用 BP 神经网络, 以发生水华为条件, 在限定两个理化因子取实际情况中的最值的情况下, 求得另一个理化因子的范围。最终结果为发生水华时,  $\text{溶氧} \geq 4.9709 \text{mg} \cdot \text{L}^{-1}$ ,  $\text{亚硝态氮} \geq 0.0751 \text{mg} \cdot \text{L}^{-1}$ ,  $\text{PH} \leq 9.2482$ 。最后利用 NAR 神经网络, 建立水华预测模型, 预测 1 到 4 号池第 16 周蓝藻密度分别为 530.2644、300.9694、926.0669、 $291.4367 \times 10^6$  个/L。即 1、3 号池塘发生水华, 2、4 号池塘不发生水华。

问题 4, 通过 matlab 拟合工具箱, 对附录 6 中的数据进行拟合, 最终得到两种鱼类, 时间、体长、体重两两组合的共 6 个生长相关模型。以此为基础, 从鱼类增长与藻类消耗角度入手, 建立线性规划模型, 并利用 Lingo 求解。知在净化池中放养 4470 条鲢鱼和 636 条鳙鱼, 即鲢鱼的密度为 1.5 条/立方米, 鳙鱼的密度为 0.2 条/立方米是一种可行的净化池方案, 该方案对 1 号池具有较好的净化效果。

问题 5, 在总结前 4 问的探究成果基础上, 学习其他学者的相关见解, 最后提出一种主要包括菜地、养殖区、生态沟渠、排水渠四个部分的生态养殖模式。该模式下水体具备较好的自净化能力, 且能量可以被循环利用以提高利用率, 同时为不明原因的水质持续恶化提供了备用方案。总体上讲, 该生态养殖模式切实可行, 具备一定的科学性。

**关键词:** 水产养殖; 水质; 主成分分析; 神经网络; Matlab

# 目录

一 问题重述	3
1.1 题目背景	3
1.2 题目核心问题	3
二 问题假设	4
三 符号说明	4
四 第一问	4
1 问题分析	4
2 数据处理与分析	5
3 结果分析	11
五 第二问	11
1 问题分析	11
2 水质评价与分类	12
3 鱼虾影响探究	15
4 结果分析	16
六 第三问	16
1 问题分析	16
2 主要理化因子与浮游生物密度相关模型	17
3 求解理化因子范围	18
4 水华预测模型	20
七 第四问	24
1 问题分析	24
2 鱼类生长与体重模型	24
3 构造净化池	27
八 第五问	28
1 问题分析	28
2 理论依据	28
3 生态养殖模式	29
九 总结评价	31
参考文献	32
附录	33

# 一 问题重述

## 1.1 题目背景

目前我国水产养殖中，池塘养殖产量约占淡水养殖的 70%。近年来，随着淡水生态系统水体污染和富营养化进程的加剧，经常导致有害蓝藻、轮虫等常见的浮游生物高密度发生，很容易诱发大面积水华。水华造成严重的环境污染及水体污染，对养殖业是一个严重的打击。

水华的发生不仅直接影响了养殖对象的正常生长发育，严重时大量排泄废水造成淡水资源污染，还会破坏养殖生态系统的平衡，导致养殖对象的不同程度死亡，造成巨大经济损失。为此我们通过研究淡水养殖池塘相关主要理化因子，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，控制并预测水华的发生，从而提高养殖产量，减小环境污染等。通过对水华发生的了解，加强大家环保意识。

## 1.2 题目核心问题

- ①探究水体、底泥与间隙水三种环境中，包括总磷、总氮在内的 6 种理化因子间的关系，并给出解释。
- ②通过 7 周的理化性质指标，对四个养殖池水体质量进行评估，分析虾池与鱼池对水体的影响。
- ③建立基于主要理化因子，并以浮游生物的量为评判标准的水华预测模型，并给出水华发生时，主要理化因子的范围值。
- ④建立鱼类生长与体重的相关模型，构造净化池模型并对进化池做评价。
- ⑤构建一种水体自进化养殖模式。

## 二 问题假设

- ①附件中的数据真实可靠，可以作为分析依据
- ②理化因子间的转化关系在各环境下一致
- ③营养级间能量传递效率为 10%
- ④池塘处于较为理想状态，无特别扰动

## 三 符号说明

$\lambda_j$ : 进行主成分分析时的特征根

F: 因子分析后提取的主要因子

$\alpha$ : 累积方差贡献率

n: BP 神经网络输入层节点数

m: BP 神经网络输出层节点数

l: BP 神经网络隐含层节点数

a: 单个硅藻质量

b: 单个绿藻质量

c: 单个蓝藻质量

d: 单个裸藻质量

e: 单个隐藻质量

x: 净化池中鲢鱼数量

y: 净化池中鳙鱼数量

## 四 第一问

### 1 问题分析

氮在水中主要以离子形态存在，主要包括硝态氮、铵态氮和亚硝态氮，也有相当部分的氮以有机质的形式存储在水体有机物中。水中的各种含氮物质通过氨化、硝化、反消化和同化作用等进行相互转换。水体中的磷则以活性磷酸盐为主体，存在部分有机磷和无机磷颗粒。在沉积物中，氮主要以含氮有机物形式存在，并通过微生物作用转变为无机氮向水体扩散。而磷则无机与有机参半，其中能溶解并与水体交换的磷是各种评价的主要研究对象。<sup>[1]</sup>

在了解水体氮磷基本情况的前提下，探究池塘中常见 N、P 理化因子的关系，应当包含两个维度。一个维度是横向探究各种理化因子自身间的相互影响，相互转换关系。另一个维度是纵向各种理化因子在沉积物、间隙水、水体间的扩散转移。对于各维度的关系，可以在相对忽略另一维度关系影响的情况下，单独讨论。

附件一中包含的数据为“连续 15 周，对 1 到 4 号养殖池，分别在 A、B 两处采集底泥、间隙水、池水三种环境中的样本，并检测包括总氮与总磷在内的 6 种理化指标的含氮值”。因为理化因子间的相互关系具有稳定性，因此抽取 1 号养殖池数据分析，其结果可以代表整体。对于 A、B 两个采样点的数据取均值，作为被测理化因子的实际含量。

在横向探究理化因子间的自身影响时,采用主成分分析法求三种环境下各理化因子的综合指标,以消除理化因子纵向扩散转移关系对横向关系探究的影响。然后对 6 项理化因子综合指标做相关性分析,初步判别个理化因子间的相关关系。最后对,具有较强相关性的理化因子做进一步函数拟合,解读其相关关系。纵向探究不同环境中理化因子的扩散转移关系的思路与方法与横向探究一致。

## 2 数据处理与分析

### 2.1 综合指标提取

求综合指标时使用主成分分析法<sup>[2]</sup>,其基本步骤为“数据标准化”,“建立变量相关系数矩阵R”,“求R的特征根 $\lambda_1, \lambda_2, \lambda_3, \dots$ 及其对应的特征向量 $T_1^*, T_2^*, T_3^*, \dots$ ”,“由特征根及特征向量求主成分表达式与综合评价函数”,“带入标准化的原始数据,求综合评价值”。值得注意的是,从本质上讲,综合评价函数对原始指标的线性综合。从计算主成分到对之加权,经过两次线性运算后得到综合评价函数。求综合指标的 Matlab 代码在附录 1 中给出,这里仅给出计算结果。表 1 为横向探究中的理化因子综合指标,表 2 为纵向探究中的池塘环境综合指标。

表 1 理化因子综合指标

周数	总磷	磷酸盐磷	总氮	硝态氮	亚硝态氮	铵态氮
1	0.179	0.1601	0.1913	-0.0889	0.0724	-0.1744
2	0.1222	0.0566	0.1802	-0.1671	0.0825	-0.1445
3	0.1533	0.1032	0.1324	-0.1101	0.1266	-0.1904
4	0.1553	0.0871	0.1868	0.0071	0.1029	-0.2055
5	0.1617	0.1048	0.1828	-0.1187	0.0326	-0.1841
6	0.163	0.0945	0.1427	-0.2447	0.0622	-0.2622
7	0.1187	0.0924	0.0755	0.0062	0.0852	-0.0965
8	0.1617	0.1384	0.1457	-0.1542	0.3066	-0.1654
9	0.1839	0.2441	0.2173	-0.3453	0.8489	-0.2543
10	0.1621	0.1454	0.1619	-0.5243	0.3524	-0.1659
11	0.1749	0.1612	0.2281	-0.0898	0.0955	-0.1869
12	0.1775	0.1793	0.1827	-0.083	0.0446	-0.114
13	0.2079	0.2723	0.2446	-0.1135	0.0963	-0.2221
14	0.2153	0.3584	0.2596	-0.0164	0.0711	-0.1896
15	0.227	0.4416	0.2129	-0.0077	0.0752	-0.1764

表 2 池塘环境综合指标

周数	池水	间隙水	底泥	周数	池水	间隙水	底泥
1	0.2528	-0.2109	-0.3302	9	0.3534	-0.0819	-0.3638
2	0.0714	-0.2383	-0.3114	10	-0.002	-0.2551	-0.3609

3	0.0974	-0.2937	-0.3611	11	0.1125	-0.4581	-0.341
4	0.0968	-0.3002	-0.328	12	0.2319	-0.1433	-0.2748
5	-0.0471	-0.2742	-0.3503	13	0.2933	-0.235	-0.3996
6	-0.0483	-0.3176	-0.3335	14	0.4332	-0.4387	-0.4643
7	0.062	-0.1647	-0.3239	15	0.5473	-0.4504	-0.3924
8	0.139	-0.1213	-0.3601				

## 2.2 相关性分析

利用 matlab 中的相关性分析函数 corr() 分别对理化因子综合指标和池塘环境综合指标进行相关性分析, 结果分别如表 3 和表 4 所示。

表 3 理化因子综合指标相关性

	总磷	磷酸盐磷	总氮	硝态氮	亚硝态氮	铵态氮
总磷	1.0000	0.9122*	0.7734*	0.1193	0.0476	-0.3908
磷酸盐磷	0.9122*	1.0000	0.6547*	0.1824	0.1017	-0.1738
总氮	0.7734*	0.6547*	1.0000	0.0730	0.0784	-0.4127*
硝态氮	0.1193	0.1824	0.0730	1.0000	-0.6315*	0.3046
亚硝态氮	0.0476	0.1017	0.0784	-0.6315*	1.0000	-0.3747
铵态氮	-0.3908	-0.1738	-0.4127	0.3046	-0.3747	1.0000

可以发现总磷与磷酸盐磷存在很强的正相关关系, 总磷与总氮、总氮与硝酸盐磷间存在较强的正相关关系; 总氮与硝态氮、亚硝态氮无显著相关关系, 但与铵态氮存在较弱的负相关关系; 硝态氮与亚硝态氮存在较强的负相关关系。

表 4 池塘环境综合指标相关性

	池水	间隙水	底泥
池水	1.0000	-0.1979	-0.5357
间隙水	-0.1979	1.0000	0.4611
底泥	-0.5357	0.4611	1.0000

显然, 池水与底泥间存在一定的负相关关系, 底泥与间隙水间存在正相关关系, 池水与间隙水间的相关关系不明显, 可能是波动性地此消彼长关系, 需要进一步探究。

## 2.3 数据拟合

为进一步探究各因素之间的深层关系, 在分析线性相关性强弱的基础上, 利用 matlab 对数据进行进一步拟合。

### 2.3.1 横向数据拟合

由表 3 知总磷与磷酸盐磷, 硝态氮与亚硝态氮, 总磷与总氮, 总氮与铵态氮间显著相关且有必要进一步进行相关度更高的拟合, 探究其深层关系。拟合的中间步骤在附录 2 中给出, 正文中只给出具体分析用到的拟合图像和相关信息。

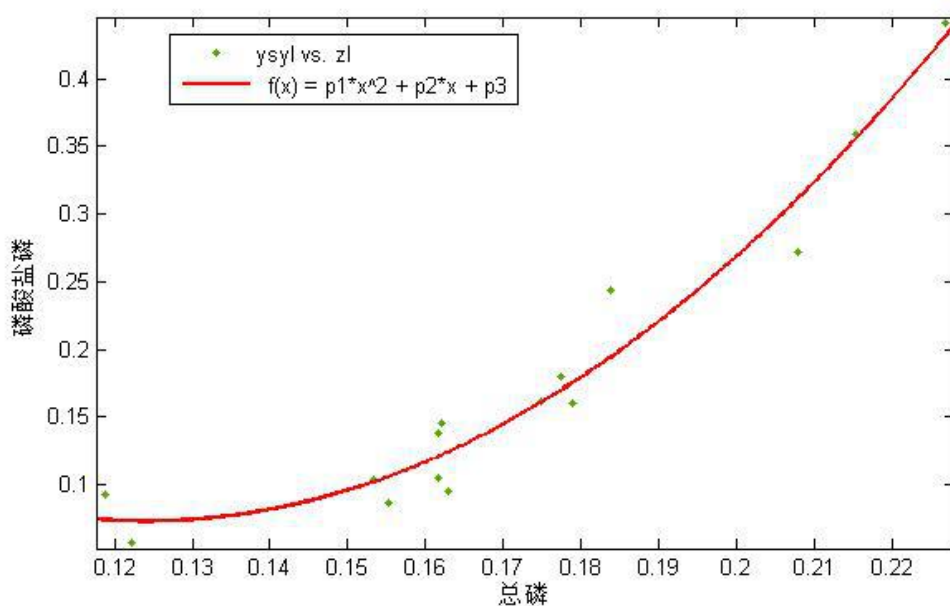


图 1 磷酸盐磷-总磷的二次多项式拟合

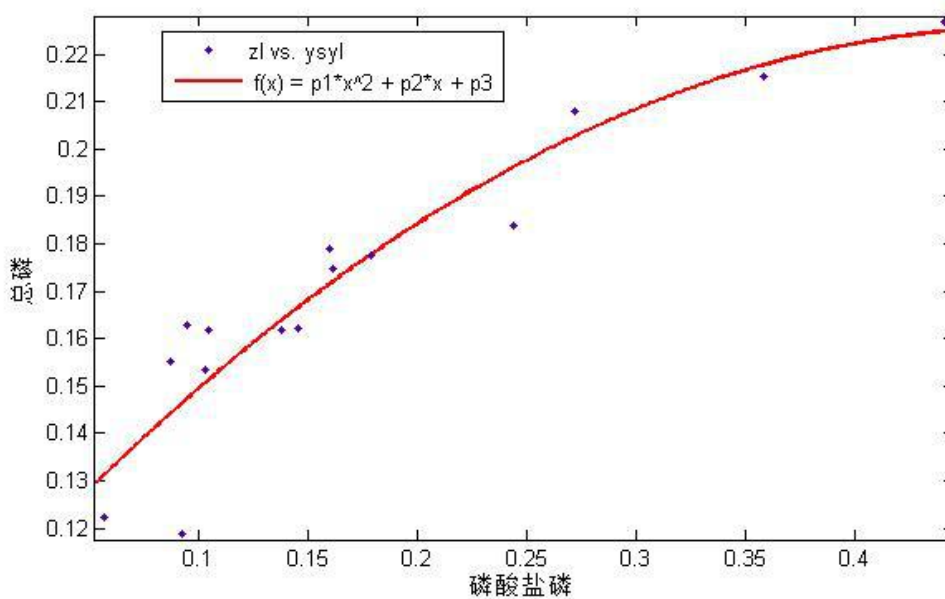


图 2 总磷-磷酸盐磷的二次多项式拟合

分别对磷酸盐磷、总磷做多项式拟合与指数拟合，发现二者总是呈正相关关系，但一次项拟合效果较差，二次及以上的多项式拟合及指数拟合效果较好，且无太大差别。因此可以得出的结论是，随着总磷浓度的增加，磷酸盐磷浓度的增加的速率逐渐增大，且无明显波动。调转  $x, y$  坐标数据，得到图 2，这种情况下拟合精度较图 1 中情况的更低。说明磷酸盐磷-总磷占主导影响的是总磷。

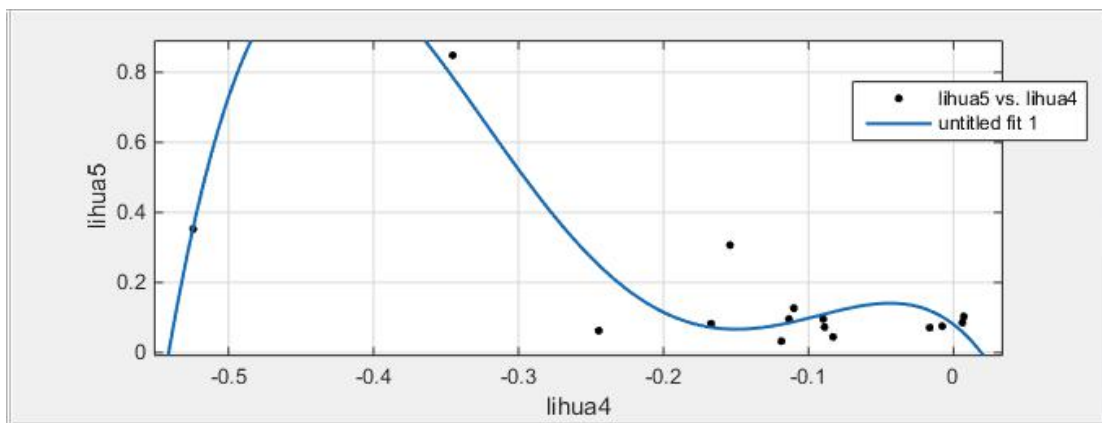


图 3 亚硝态氮-硝态氮的四次多项式拟合图

在使用四次项多项式进行拟合时， $R=0.8167$  具有较高的拟合度，从整体趋势上来看二者呈负相关关系。

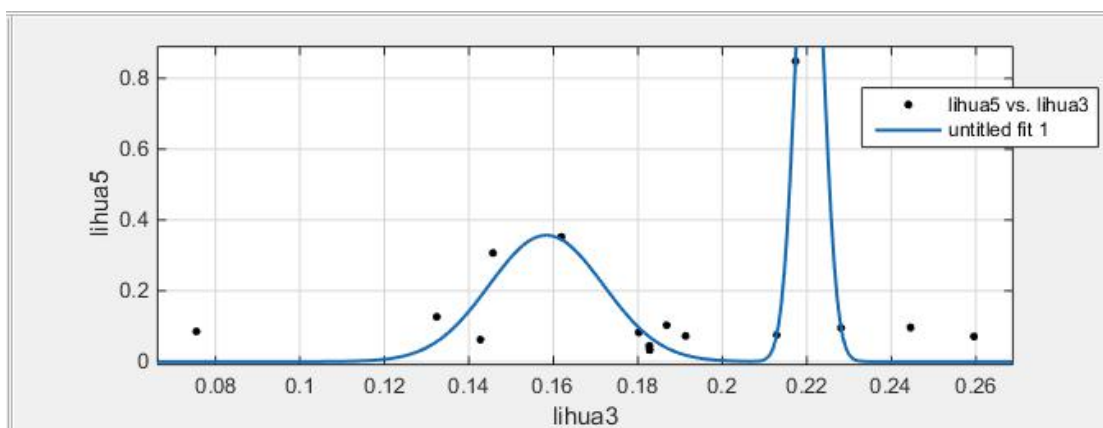


图 4 亚硝态氮-总氮的二次高斯拟合

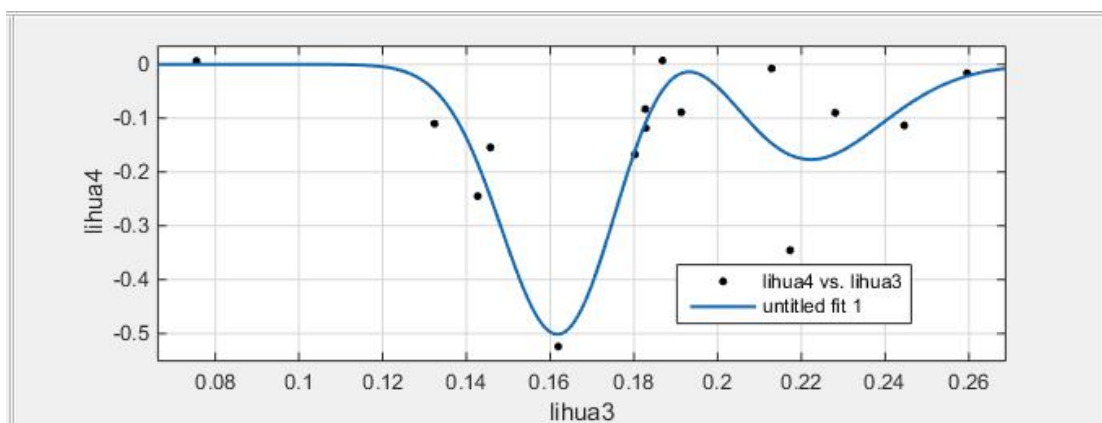


图 5 硝态氮-总氮的三次高斯拟合

亚硝态氮与硝态氮和总氮进行多项式拟合时无法达到较高拟合精度，因此总氮与这二者关系较小。利用高次高斯拟合，则能较好的拟合，说明二者与总氮的关系相对复杂，存在波动性。值得注意的是，在亚硝态氮-总氮的二次高斯拟合（图 4）中出现的两个波谷，和在硝态氮-总氮的三次高斯拟合（图 5）中出现的两个波峰具有位置一致性。



这佐证了硝态氮与亚硝态氮间相互转化“此消彼长”的关系。

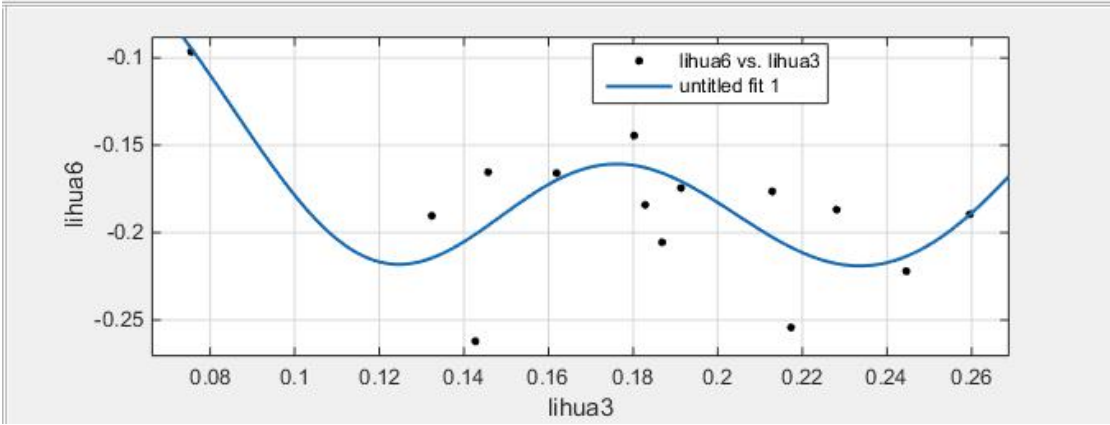


图 6 铵态氮-总氮的二次高斯拟合

可以发现，在总氮指标为 0.18 左右的地方，曲线出现了极大值。同时也存在两个极小值。而这两个极小值分别在亚硝态氮-总氮曲线和硝态氮指数-总氮指数曲线极值的两侧。

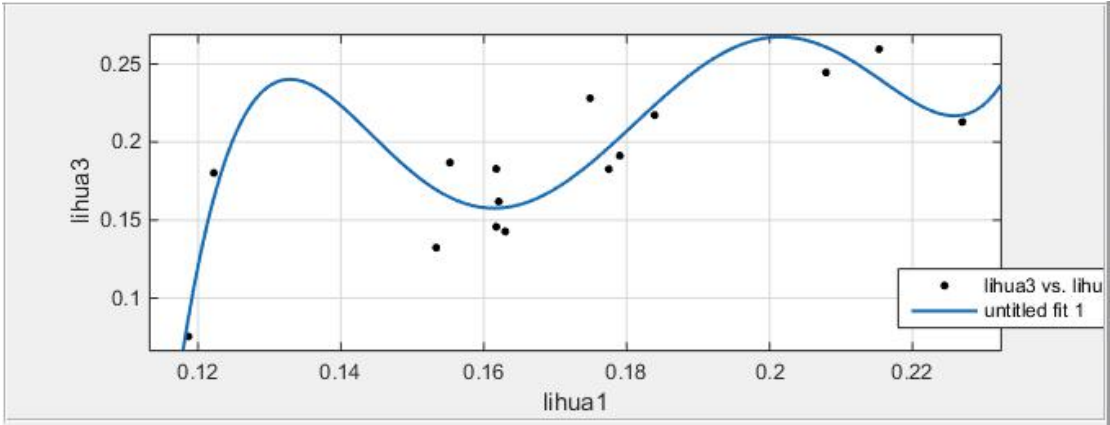


图 7 总氮-总磷的四次多项式拟合

显然，总磷与总氮间的呈明显正相关关系，存在一定的波动性，但波动性并不明显，以线性相关关系为主。

### 2.3.2 纵向数据拟合

以间隙水环境综合指标为横坐标，底泥环境综合指标为纵坐标进行多项式拟合和高斯拟合，发现三峰高斯函数拟合效果最佳。从总体上来看，间隙水环境综合指数与底泥环境综合指数呈现明显负相关，由此可以推测间隙水中的理化因子主要来自于底泥。曲线有两个明显的波峰波谷，可能是池水清淤和投放饲料等造成的。也可能是当间隙水和底泥理化因子含量较少时，底泥中的理化因子快速向间隙水中转移，当间隙水中的理化因子达到一定浓度后，容易发生沉降。

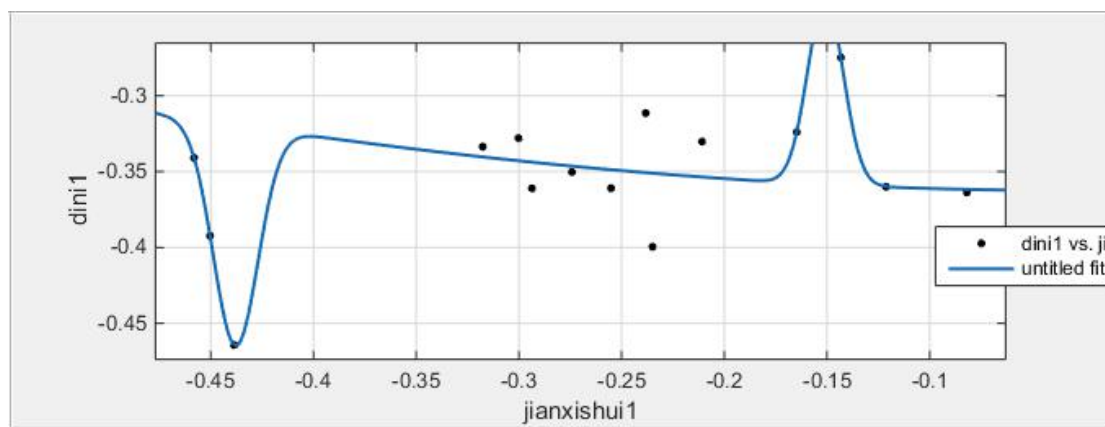


图 8 底泥-间隙水三峰高斯函数拟合

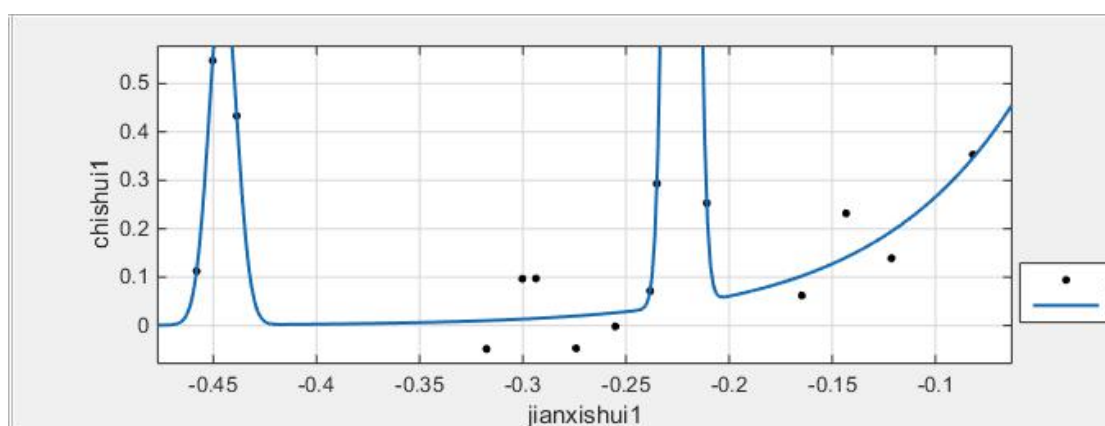


图 9 池水-间隙水三峰高斯函数拟合

以间隙水环境综合指标为横坐标,池水环境综合指标为纵坐标进行多项式拟合和高斯拟合,同样发现三峰高斯函数拟合效果最佳。从总体上看,二者呈明显的正相关关系,且随着池水综合指数增大速率逐渐变大。值得注意的是,图 9 中第一个波峰的位置和图 8 中波谷位置相同,推测该峰值是由于环境改变造成底泥理化物质迅速向水体转移。图 8 中的峰值稍延后于图 9 中的峰值,推测该峰值是由于污染水体流入造成,污染水体的流入造成池水氮磷含量骤增,而后发生沉降。

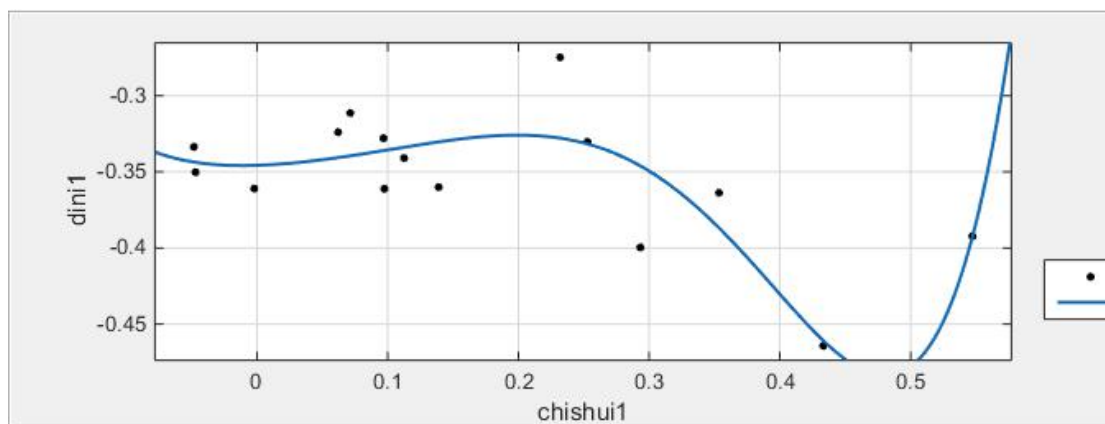


图 10 底泥-池水六次多项式函数拟合

底泥和池水的关系相对不明显,使用六次多项式拟合时,才能获得较高精度。从总体上看二者呈现负相关关系。因此,可以推断,水体中的各种理化因子主要来源于底泥,且理化

因子从底泥向间隙水扩散的速率高于理化因子从间隙水向池水扩散的速率。

### 3 结果分析

从理化因子的相互转化关系来看：<sup>[3]</sup>

- ①总氮与总磷同步变化，其主要原因是氮与磷是构成的生命的重要元素，在生物体及外界输入物质中同时存在。
- ②总磷与磷酸盐同步变化，且总磷占据主导地位，随着总磷的增多，磷酸盐增多的速率逐渐增大，其主要原因是水体中磷酸盐是磷的主要存在形式，水体中磷的增多主要得益于磷酸盐的溶解。
- ③总氮与铵态氮间有一定联系，呈有规律的波动，主要原因是将底泥中的有机氮转化成铵态氮的氨化作用。
- ④硝态氮与亚硝态氮具有“此消彼长”的关系，其主要原因是因为在底泥和上层水体间不断发生硝化与反硝化作用，这二者的相互转化对维持水体氮平衡有重要意义。

从理化因子在池塘不同环境中的转移关系来看：

- ①底泥是各种理化因子的源，理化因子从底泥向间隙水和池水扩散，且理化因子从底泥向间隙水扩散的速率高于理化因子从间隙水向池水扩散的速率。
- ②底泥中各种含氮有机物不断分解，含磷物质不断溶解是影响水体理化因子含量的主要原因，其分解或溶解速率与水体理化环境有密切关系。
- ③污染物进入会造成水体理化因子异常偏高，而后会发生沉降作用，使得水体中的含氮与含磷物质沉积于底泥中。

## 五 第二问

### 1 问题分析

问题二的核心是根据理化因子值评价水质<sup>[4-6]</sup>。水质评价作为生态环境评价的重要内容之一，其涉及到的因子数量众多，数据分析时，重要信息时常被因子间错综复杂的关系所掩盖。目前较好的水环境质量综合评价方法有指数法、模糊综合评价法、灰色数学法、神经网络法等它们均无法回避多因子问题。而因子分析法能够将多个因子化成少数几个因子，在分析问题，用前后方差较大的主要因子取代所有因子，一方面大大减少计算量，另一方面较好的解读出影响综合评价的主要成分。因此可以通过因子分析法，求主要因子综合分数作为池塘水质的评判依据。

附件 2 中 4 个池中采集的主要理化指标主要有叶绿素 A、叶绿素 B、叶绿素 C、轮虫、溶氧、COD、水温、PH 值、盐度、透明度、总碱度、CA2+、MG+和氯离子。但四个池子均有数据的指标，仅有轮虫、溶氧、COD、水温、PH 值、盐度、透明度、总碱度八项。为统一水质评价标准，使用这八个指标作为水质评价主成分分析的原始成分。附件 8 为我国水质标准，其中将水质划分成了五大类，但因其水质评价指标仅有溶氧和 COD 与附件 2 中的指

标相同。因此不能在主成分分析法中引入水质的国家标准[3]。水质评价最终以主成分综合分数为度量标准，仅能得到水质的相对好坏。得到池塘水质综合得分后，对 4 个池塘 7 周的得分进行系统聚类，可以得到池塘的水质类别划分。

附件 5 中明确指出，池塘 1 和 2 为虾池，池塘 3 和 4 为鱼池。单独分析池塘个理化指标的相对含量和变化趋势，可以得出虾池与鱼池对水体各项指标的影响。

## 2 水质评价与分类

以 SPSS 为工具进行因子分析，数据源为 4 个池塘 8 个周中包括轮虫、溶氧、COD 在内的 8 项指标，总计 32 条记录。利用 matlab 编程求得综合得分，代码见附录。

### 2.1 数据标准化

原始数据由 32 个样本和 8 个因子构成，构造原始变量矩阵  $X$ 。 $X_{ij}$  为指标含量， $i$  从 1 到 32，依次指示从池塘 1 到池塘 4，从 1 周到 15 周的数据样本。 $j$  从 1 到 8，依次指示轮虫、溶氧、COD、水温、PH 值、盐度、透明度、总碱度八项指标。

为统一数据数量级，并去除量纲，进行数据标准化处理。处理结果如表 5 所示。

表 5 理化指标标准化

No	轮虫	溶氧	COD	水温	PH 值	盐度	透明度	总碱度
1	1.29056	-.02325	-.66120	-1.41951	1.13336	-.19734	.61272	.57516
2	-.16886	.02635	-1.74752	-1.15820	-1.25603	1.55678	2.26191	.47277
3	-.81638	2.18126	-.95560	-1.72438	1.00536	-.19734	-1.21971	-.94143
4	-.81193	2.26944	.13367	-1.15820	.23734	-.72357	.24623	-1.03976
5	1.32589	-1.08141	-1.22055	-1.02755	.40801	.67972	-.12025	.73795
6	-.08165	.26885	-1.36775	-1.41951	-2.36540	.32890	2.44516	.54694
7	-.81695	.85855	-1.22055	-1.59372	-.31734	.15349	-.66998	-.91019
8	-.81197	.15311	-.30792	-1.68082	.66402	-1.42522	-.85323	-.91648
9	1.29829	.85855	.78135	-.54848	.40801	-.02193	.24623	.90872
10	.01770	.32947	-.75246	-.28717	-1.68271	2.08301	1.89542	.52159
11	-.81842	-1.08141	.78135	-.37427	.40801	-.19734	-.48674	-.93009
12	-.81256	-.55233	2.42409	-.11296	.23734	.32890	-1.40296	-.87276
13	1.61292	-.99323	1.05514	-.02586	1.04803	.32890	-.48674	.90674
14	.10271	.68219	-1.56205	-.02586	-1.12803	1.03054	1.16245	.73835
15	-.81836	-.99323	.90794	1.01938	.49335	.50431	.24623	-.94722
16	-.81343	1.65217	1.89123	1.54200	.74935	.15349	-.48674	-.86949
17	1.67363	-1.52232	-2.1960	1.01938	.74935	.32890	-.48674	.95551
18	.12810	-.38699	-1.25588	.88873	-1.51204	.85513	.97921	.61620
19	-.81896	-1.52232	.95799	-.11296	.36534	-.19734	-.12025	-.97891
20	-.81745	-1.25777	.74308	-.20007	.23734	2.08301	.42948	-.96515
21	1.72000	-.56335	-.41096	.84518	1.21870	-1.42522	-.85323	1.43354
22	.15018	1.18923	-.12834	.75807	-1.59738	-1.24981	.42948	1.09463
23	-.82017	-.55233	-.41096	.67097	.87736	-1.24981	-.48674	-.97044

24	-.81786	-.56335	.91972	1.01938	.57868	.85513	.06299	-.97510
25	1.81163	.69873	.35742	-.20007	.49335	-.72357	-1.03647	1.46305
26	.15570	.44521	.35447	.49676	-1.81072	1.03054	.61272	1.27085
27	-.81981	.69873	.06302	-.11296	.49335	-1.60063	-.66998	-1.03233
28	-.81867	.19720	-.43157	.06124	.36534	.15349	-.85323	-.91573
29	1.85026	-.30984	.55466	1.19359	.92003	-.72357	-.30350	1.55935
30	.18550	.36254	-.48162	1.01938	-1.17070	.15349	-.12025	1.39294
31	-.82006	-.30984	.55466	1.28069	-.27467	-1.42522	.79596	-.95653
32	-.81958	-1.15857	.65476	1.36780	.02400	-1.24981	-1.76944	-.97267

## 2.2 主要因子确定

求标准化数据的特征值  $\lambda_j$ ，并按公式  $\alpha = \sum_{j=1}^p \lambda_j / \sum_{j=1}^8 \lambda_j$  求解累积方差率，求解结果如下表所示。

表 6 主成分特征值及其方差贡献率及累积贡献率

主成分	特征值	方差贡献率(%)	累积方差贡献率 (%)
1	2.69	33.57	33.57
2	1.85	23.14	56.71
3	1.37	17.06	73.77
4	.90	11.30	85.06
5	.61	7.59	92.66
6	.31	3.89	96.54
7	.26	3.21	99.76
8	.02	.24	100.00

根据表 6，计算机自动确定第 1、2、3 主成分为主要因子，用于说明全体理化因子信息。由表 7 知，由表 3 知，因子 1 反映除轮虫外其他主要因子所共有的成分，因子 2 主要反映轮虫、溶氧和总碱度间的共有成分，因子 3 则反映全体因子的共有成分。

表 7 主要因子荷载

	轮虫	溶氧	COD	水温	PH 值	盐度	透明度	总碱度
因子 1	0.107	0.074	-0.265	-0.115	-0.286	0.211	0.313	0.217
因子 2	0.483	-0.208	0.068	0.238	0.133	-0.068	-0.091	0.414
因子 3	-0.196	-0.45	0.279	0.418	-0.201	0.383	0.199	-0.112

## 2.3 综合评价函数确定

由特征向量可以求得主要因子表达式如下

$$\begin{cases} F_1 = 0.288Z_1 + 0.199Z_2 - 0.712Z_3 - 0.308Z_4 - 0.769Z_5 + 0.568Z_6 + 0.841Z_7 + 0.583Z_8 \\ F_2 = 0.895Z_1 - 0.385Z_2 + 0.126Z_3 + 0.440Z_4 + 0.246Z_5 - 0.127Z_6 - 0.168Z_7 + 0.767Z_8 \\ F_3 = -0.267Z_1 - 0.614Z_2 + 0.381Z_3 + 0.571Z_4 - 0.274Z_5 + 0.522Z_6 + 0.271Z_7 - 0.152Z_8 \end{cases}$$

根据方差贡献率，又可求得综合评价函数如下

$$F = (\lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3) / \sum_{k=1}^3 \lambda_k = 0.455F_1 + 0.314F_2 + 0.231F_3$$

## 2.4 评价结果分析

由综合评价函数求得综合评价值如表 8 所示，值得注意的是，综合评价值越小则水质越好。

表 8 水质综合评价值

	第 1 周	第 3 周	第 5 周	第 7 周	第 9 周	第 11 周	第 13 周	第 15 周
1 号池	0.2285	-0.0867	-0.1165	-0.0592	6.21E-04	0.0571	0.0274	0.0167
2 号池	0.7661	0.1335	0.0407	0.0203	0.0713	0.075	0.0711	0.0444
3 号池	0.0063	0.0059	-0.0252	0.005	0.0405	0.0337	0.0211	0.03
4 号池	-0.2904	-0.1656	-0.1011	-0.0538	0.046	0.0212	-0.005	1.64E-07

利用 Excel 将水质综合评价值可视化，如下所示

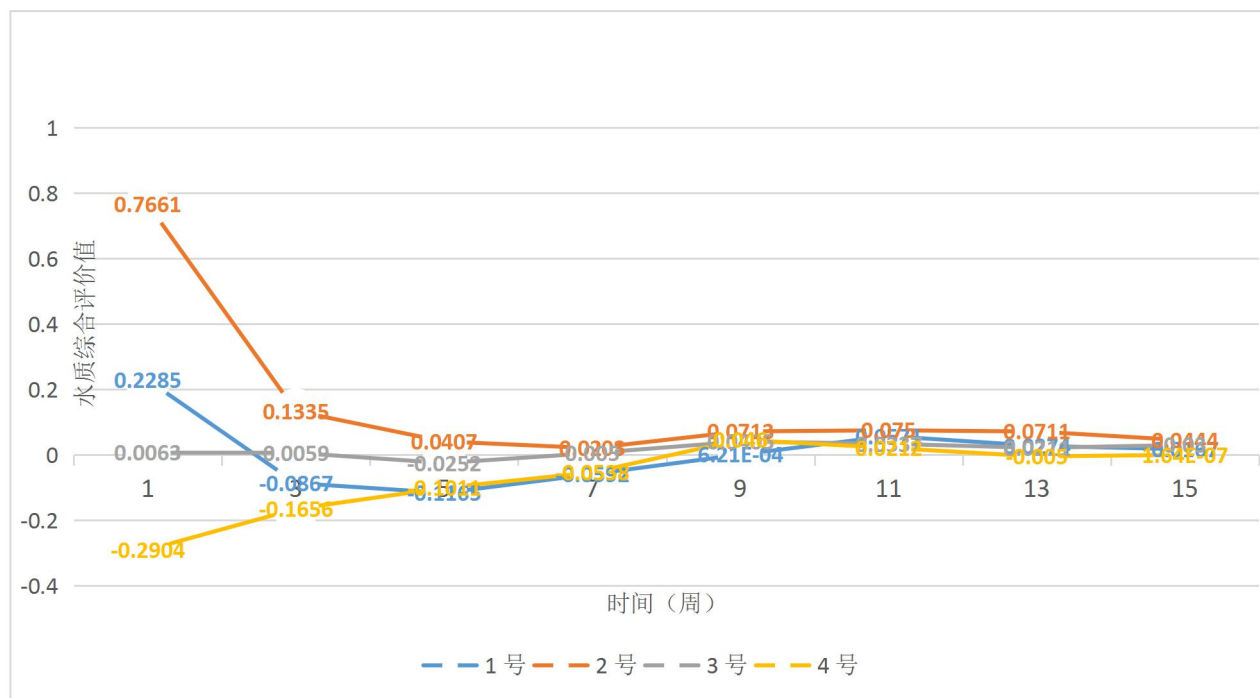


图 11 水质综合评价值

可以发现，1、2 号池水质一开始时相对较差，之后逐渐好转并趋于稳定。3 号池水质较

为稳定，4号池刚开始时水质较好，随着时间的推移，水质逐渐变差并趋于稳定。水质趋于稳定时，水质从优到劣的排序为4号池、3号池、1号池、2号池。由附件知，1、2号池为虾池，3、4号池为鱼池。因此可以推断，鱼池的水质从整体上讲优于虾池的水质。

### 2.5 水质评价聚类分析

对于各池塘的水质综合评分，利用 SPSS 进行系统聚类分析，得到图 12 所示的冰柱图，由冰柱图可以发现。若将池水水质分成 3 类，则 1 号和 3 号池为一类，2 号和 4 号池各为一类。若将池水水质分成 2 类，则 1 号池、3 号池、4 号池为一类，2 号池单独为一类。因此，可以认为，4 号池水质为优，1 号和 3 号池水质为良，2 号池水质为差。

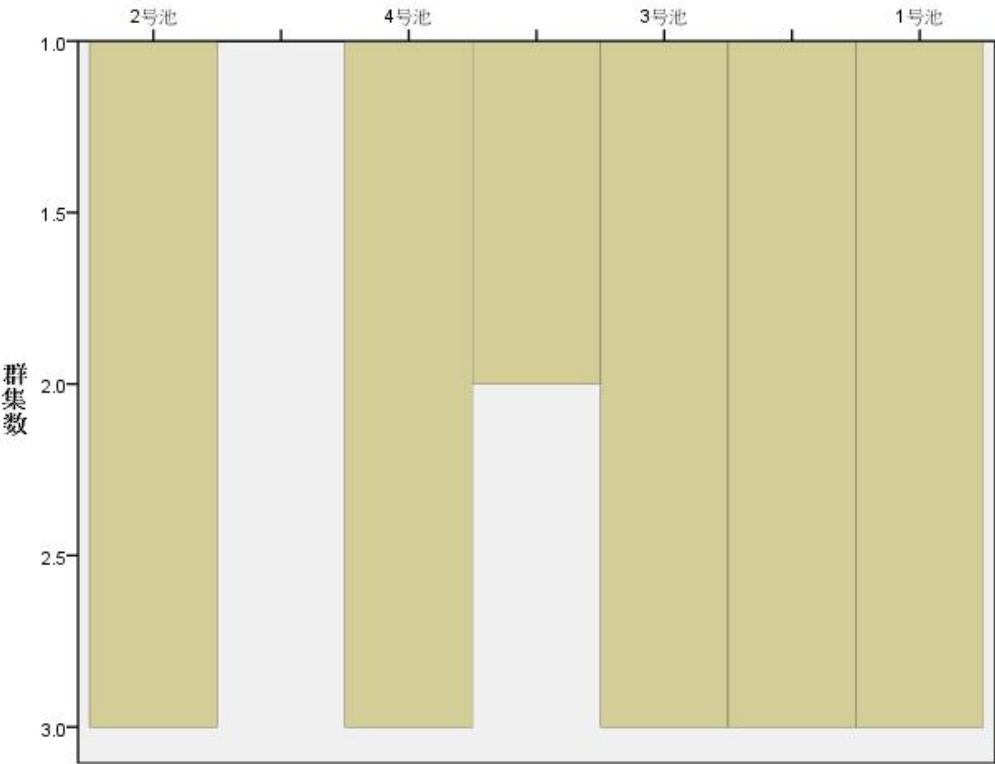


图 12 水质系统聚类冰柱图

### 3 鱼虾影响探究

附件 5 中清晰表明 1、2 号池为虾池而 3、4 号池为鱼池。对附件 2 中的各项指标进行逐一比对，来探究虾池鱼池对水体的影响。

- ①从 PH 值来看：四个池塘水体呈偏碱性，其值随时间呈微小波动，只有 2 号池碱性相对偏弱。
- ②从轮虫密度：1、2 号池明显高于 3、4 号池，且 1、2 号池轮虫密度逐渐增大，3、4 号池轮虫密度逐渐减小；
- ③从溶氧来看：1、2、3、4 号池基本相同，3、4 号池溶氧波动性略高于 1、2 号池。
- ④从 COD 来看：1、2、3、4 号池基本相同，前期有较大波动，后来趋于稳定；

⑤从总碱度来看：1、2号池明显高于3、4号池，且1、2号池总碱度呈明显增大趋势；3、4号池塘总碱度波动较小趋于稳定；

⑥从透明度来看：2号池明显高于其他三个池，且透明度呈明显下降趋势，1、3、4号池透明度波动较小。

⑦从水温和盐度上看，4个池塘基本趋于一致。

⑧对于叶绿素A含量：1、2号池塘没有数据，3、4号池塘相差较小；对于叶绿素B含量：1、2号池塘没有数据，4号明显高于3号池塘；对于叶绿素C含量：1、2号池塘没有数据，4号较高于3号池塘；对于钙离子、镁离子含量：1、2号池塘没有数据，3、4号池塘相差较小。对于氯离子含量：2号明显高于1号池塘，3、4号池塘没有数据；对于生物量：1号明显高于2号池塘，3、4号池塘没有数据。

## 4 结果分析

通过因子分析法获得水体质量的综合评价值，发现1、2号池水质一开始时相对较差，之后逐渐好转并趋于稳定。3号池水质较为稳定，4号池刚开始时水质较好，随着时间的推移，水质逐渐变差并趋于稳定。水质趋于稳定时，水质从优到劣的排序为4号池、3号池、1号池、2号池。进一步通过系统聚类，发现可以将池水划分为三类，其中4号池水质为优，1号和3号池水质为良，2号池水质为差。从总体上讲，鱼池的水质优于虾池。

通过对鱼池和虾池的理化因子含量即变化趋势的逐一比较，认为鱼池对水体的影响是减小水体轮虫密度，并且以较大幅度扰动水体溶氧，预计其主要原因是鱼池中的鱼在生长过程中能摄食大量浮游生物。认为虾池对水体的主要影响是增大水体总碱度，并使轮虫密度增大，预计其主要原因因为虾池中的虾主要在水体下层活动，会扰动各种物质间的相互转化和其在不同环境中的扩散转移，并且，虾子对浮游生物的摄食不明显，二者共生。

## 六 第三问

### 1 问题分析

第三问中实际上包含3个问题，第一个问题是建立一个模型，用于描述主要理化因子与浮游生物密度的关系。首先通过各理化因子与浮游生物密度间的相关性，选择出相关性强的主要理化因子。之后通过以主要理化因子为输入，以浮游生物密度为输出，构建合适的BP神经网络作为第一个问题的模型。

第二个问题是探究水华发生时，主要理化因子间的范围。水华发生指一定，池塘中浮游生物密度达到某值，由于题中明确指出池塘1发生轻微水华，因此以池塘1中浮游生物密度的平均值作为水华发生指标。将其他理化因子的值，取实际中最可能发生水华的值，限定浮游生物密度为水华发生指标，反解可得指定理化因子在发生水华时的范围。这里的反解，采用新构造的，以指定理化因子为输出的BP神经网络。

第三个问题是预测水华发生，为避免理化因子间错综复杂的关系，并充分利用现有数据。这一问使用NAR神经网络进行处理，不同于上述“主要理化因子与浮游生物密度相关模型



“的 BP 神经网络，NAR 神经网络不采用主要理化因子预测浮游植物含量，而采用时间序列进行直接预测，其输入为各池的依时间序列的藻含量，通过藻含量的时间变化来直接预测第 N+1 周的藻含量。进一步，可通过藻含量值与水华发生指标做对比，确定该预测浓度下是否会发生水华。

## 2 主要理化因子与浮游生物密度相关模型

### 2.1 数据选择与处理

附件 1 含有总磷、磷酸盐磷、总氮、硝态氮、亚硝态氮、铵态氮 6 种理化因子数据，附件 2 中 4 个池塘共有的理化因子为轮虫、溶氧、COD、水温、PH、盐度、透明度、总碱度 8 种。为统一模型适用性，以附件 1 和 2 中总共 16 种理化因子为基础，探究影响浮游生物密度的主要理化因子。以池塘 1、2 的数据为准，16 种理化因子与浮游生物密度间的相关性分析，分析结果如表 1 所示。

表 9 理化因子与浮游生物密度相关性

理化因子	R²	理化因子	R²	理化因子	R²
总磷	-0.0662	铵态氮	-0.0004	PH	-0.5169
磷酸盐磷	-0.1430	轮虫	-0.3860	盐度	-0.1098
总氮	0.0488	溶氧	0.5329	透明度	0.0328
硝态氮	0.2952	COD	0.0855	总碱度	0.2191
亚硝态氮	0.6307	水温	0.3174		

由表 9 知，亚硝态氮态氮、溶氧和 PH 是影响浮游生物密度的主要理化因子。因此，在之后的建模中，以这三种主要理化因子为研究对象，忽略其他次要理化因子。BP 神经网络的输入值即为，池水中亚硝态氮氮、溶氧和 PH，输出则为浮游植物密度总量。

### 2.2 模型建立

本文基于 Matlab 软件<sup>[7-9]</sup>，构建了一个三层前向神经网络（代码参见附录）。模型核心是确定数据训练方式和隐含层数量。

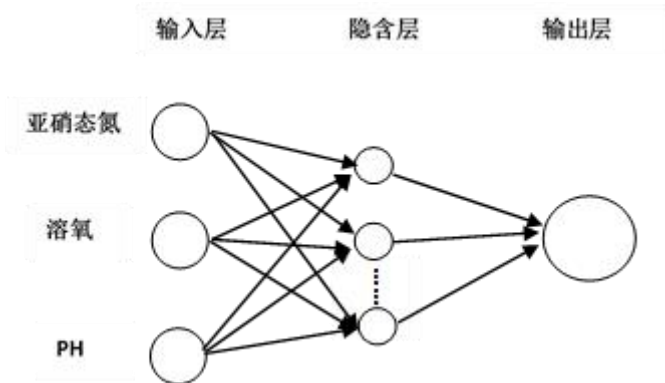


图 13 BP 神经网络拓扑结构

从附件中共可以提取出 4 个池塘，8 周总计 32 条包含亚硝态氮、溶氧、PH 和浮游植物密度总量的数据。以其中的 26 条数据作为 BP 神经网络训练数据，剩余的 6 条作为检验数据。

就训练数据量少的问题，将 26 条训练数据进行备份，利用训练数据原件和训练数据备份对神经网络进行重复训练，以提高拟合精度。

在 BP 神经网络构建过程中，输入层节点数为  $n=3$ ，输出节点数为  $m=1$ 。根据经验公式  $l \leq \sqrt{m+n+a}$ （其中  $a$  为从 0 到 10 的整数，该公式可以防止过拟合现象发生），可以确定隐含层节点数目有 2 到 12 共十种可能情况。建立这 10 种情况，并对这些神经网络的精度进行比较，结果如下。

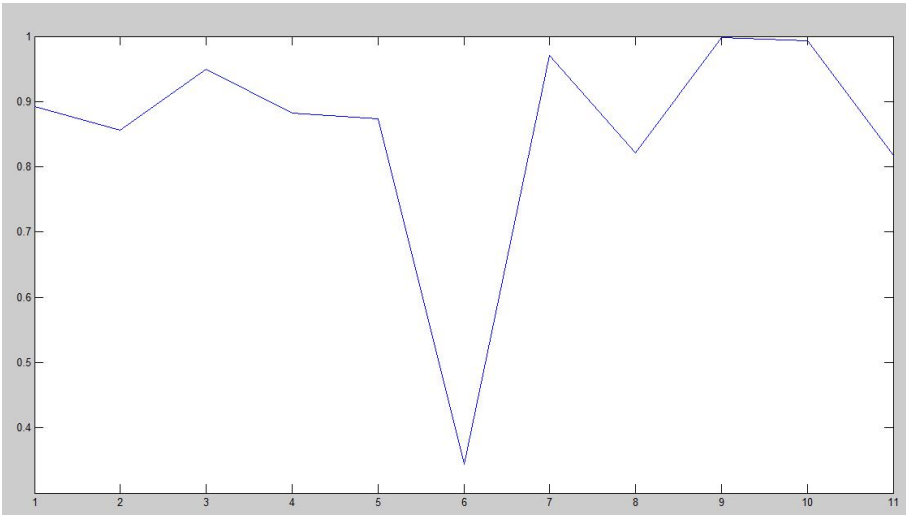


图 14 不同隐含层节点数下 BP 神经网络预测的  $R^2$

结果显示，当隐含层节点数为 10 时，神经网络拟合误差最小，因此将隐含层节点数为 10 的 BP 神经网络作为主要理化因子与浮游生物密度相关模型。

### 2.3 模型评价

上述模型为单隐含层 BP 神经网络。以第  $n$  周池水中亚硝态氮、溶氧和 PH 的含量为输入，以第  $n$  周浮游植物密度总量为输出。利用 26 条数据进行双重训练，并取隐含层节点数为 10。模型能够利用主要理化因子求得当前浮游生物密度，精度较高，适用于 4 个池塘。

## 3 求解理化因子范围

### 3.1 求解原理

由上述模型知，影响水华发生的主要理化因子为池水中亚硝态氮、溶氧和 PH 含量。其中，浮游生物量与亚硝态氮成正相关关系，与溶氧成正相关关系，与 PH 成负相关关系。水华发生意味着浮游生物密度达到某一值，该情况下亚硝态氮指数存在最小值，溶氧指数存在最小值，PH 值存在最大值。

为探究三个主要理化因子间的联系，对其三者做两两拟合。

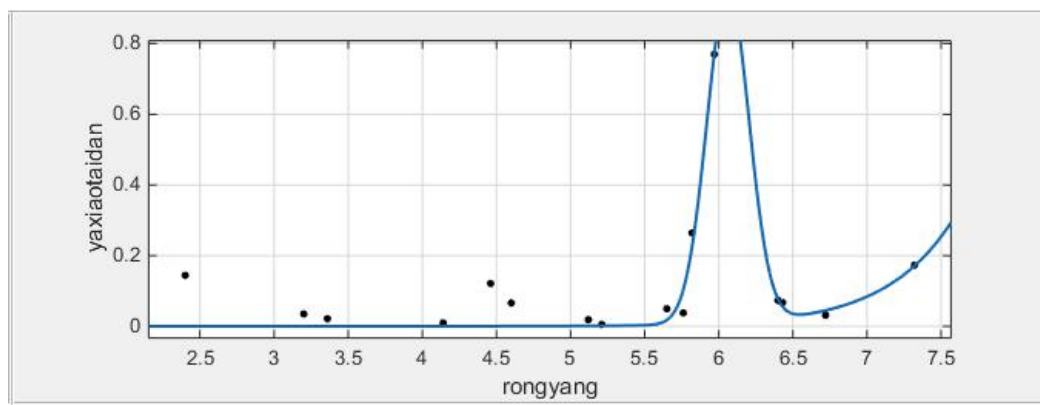


图 15 亚硝态氮与溶氧

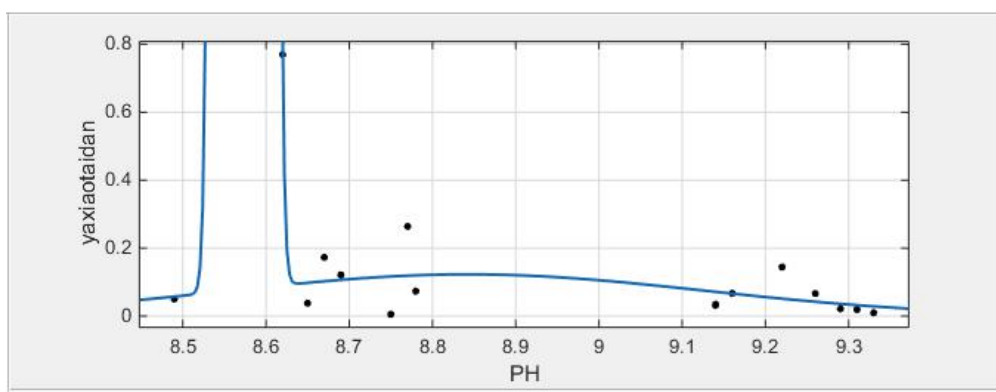


图 16 溶氧与 PH

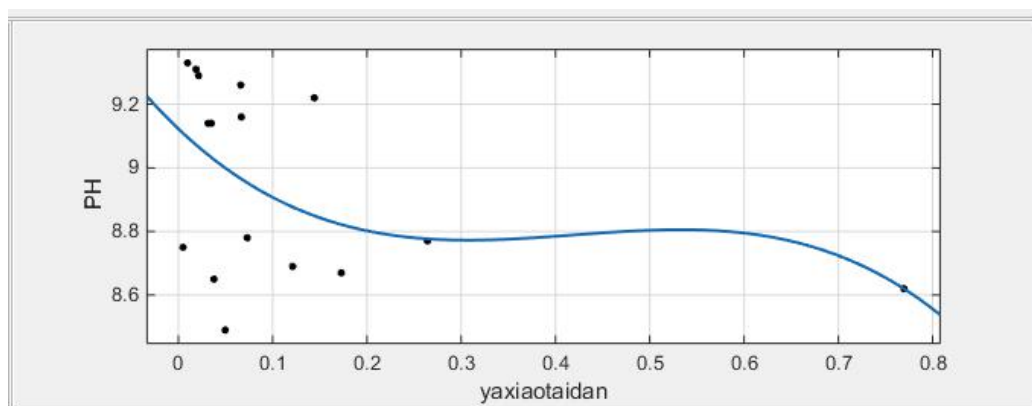


图 17 PH 与亚硝态氮

由上述拟合关系知，要使得溶氧取得最小值，则 PH 越大越好，亚硝态氮越小越好；要使亚硝态氮取得最小值，溶氧越小越好，PH 越大越好；要使得 PH 取最大值，则溶氧越小越好，亚硝态氮越小越好。

### 3.2 训练 BP 神经网络

BP 神经网络训练和验证方式与“主要理化因子与浮游生物密度相关模型”中的一致。分别训练 ryNET, yxtdNET, phNET 用于求解主要理化因子范围。

以浮游植物总密度、池水中亚硝态氮和 PH 为输入，以池水中的溶氧为输出，训练并验证 BP 神经网络 ryNET。

以浮游植物总密度、池水中溶氧和 PH 为输入，以池水中的亚硝态氮为输出，训练并验证 BP

神经网络 yxtdNET。  
以浮游植物总密度、池中亚硝态氮和溶氧为输入，以池水中的 PH 为输出，训练并验证 BP 神经网络 phNET。

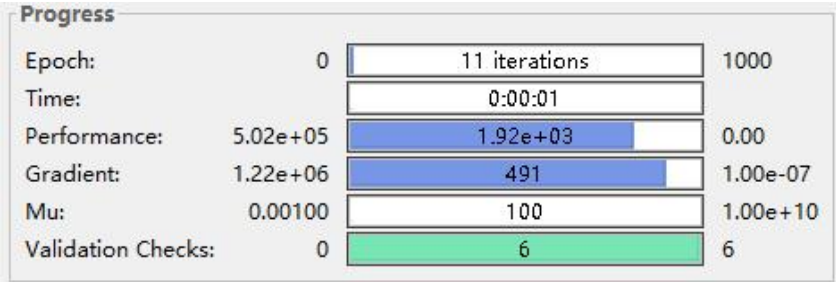


图 18 BP 神经网络训练过程

### 3.3 计算求解

由题知 1 号池中发生轻微水华，,因此以 1 号池中浮游植物总密度 359.0056 作为水华发生标准。

以水华发生时浮游植物总密度 359.0056，实际中 PH 的最大值 9.33，实际中亚硝态氮的最小值 0.005 为输入，利用 3.2 中名为 ryNET 的神经网络进行求解，得到发生水华时溶氧的最小值 4.9709。

以水华发生时浮游植物总密度 359.0056，实际中溶氧的最小值 2.4，实际中 PH 的最大值 9.33 为输入，利用 3.2 中名为 yxtdNET 的神经网络进行求解，得到发生水华时亚硝态氮的最小值 0.0751。

以水华发生时浮游植物总密度 359.0056，实际中亚硝态氮的最小值 0.005，实际中溶氧的最小值 2.4 为输入，利用 3.2 中名为 phNET 的神经网络进行求解，得到发生水华的 PH 的最大值为：9.2482。

因此发生水华主要理化因子的范围如下表所示：

表 10 水华发生时主要理化因子范围

主要理化因子	范围
溶氧	$\geq 4.9709mg \cdot L^{-1}$
亚硝态氮	$\geq 0.0751mg \cdot L^{-1}$
PH	$\leq 9.2482$

## 4 水华预测模型

### 4.1 NAR 神经网络原理

NAR 模型是 NARMA 模型的一种特殊形式，NARMA 模型可以定义为：

$$y_n = G(y_{n-1}, \dots, y_{n-k}, x_n, \dots, x_{n-l}) \quad (1)$$

①式中： $x$ —模型输入； $y$ —模型输出； $n$ —离散时间； $G$ —非线性函数。当实测数据为时间序列，模型输入是未知时，可以将式① 改写成式② 所示的 NAR 模型形式。

$$y_n = F(y_{n-1}, \dots, y_{n-k}) + k\varepsilon_n \quad (2)$$

②式中： $F$ —非线性函数； $k$ —常数； $\varepsilon_n$ —服从高斯分配的随机变量。

不同于上述“主要理化因子与浮游生物密度相关模型”的 BP 神经网络，NAR 神经网络<sup>[10]</sup>不采用主要理化因子预测浮游植物含量，而采用时间序列进行直接预测，其输入为各池的依时间序列的藻含量，通过藻含量的时间变化来直接预测第  $N+1$  周的藻含量。

## 4.2 模型建立

以池 1 数据为例，说明 NAR 神经网络建立方法，即“水华预测模型”。

选用贝叶斯正则化算法训练 NAR 神经网络，将网络权值均方与误差均方的加权作为性能函数，在保证网络均方误差尽可能小的前提下，有效地控制网络的复杂程度，从而获得较好的泛化性能提高网络推广能力。在 matlab 环境下（相关代码请参见附录），调用 `narnet` 函数构建网络，建模样本中分为 7 个训练样本、4 个验证样本、4 个测试样本，传递函数设定为 `tansig` 函数，训练函数设定为 `trainbr` 函数，权值自适应学习函数设定为 `learnbd` 函数，性能函数采用 MSE。网络经过 169 次训练，训练误差为  $5.3 \times 10^{-6}$ 。训练结束后建立了如图 5 所示的 NAR 神经网络蓝藻预测模型，预测结果如图 6 所示。

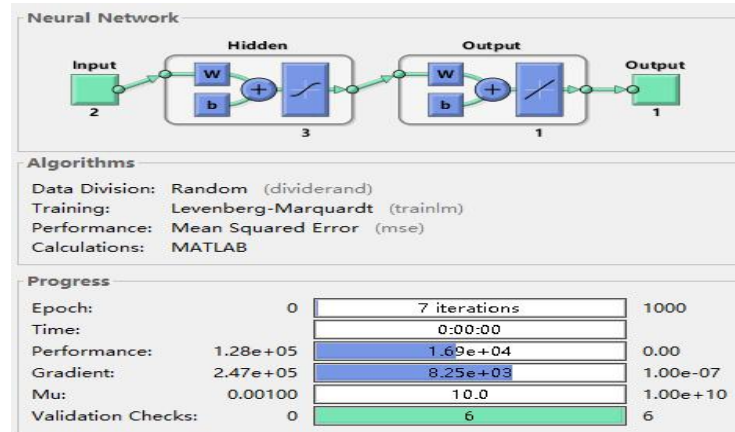


图 19 NAR 神经网络结构

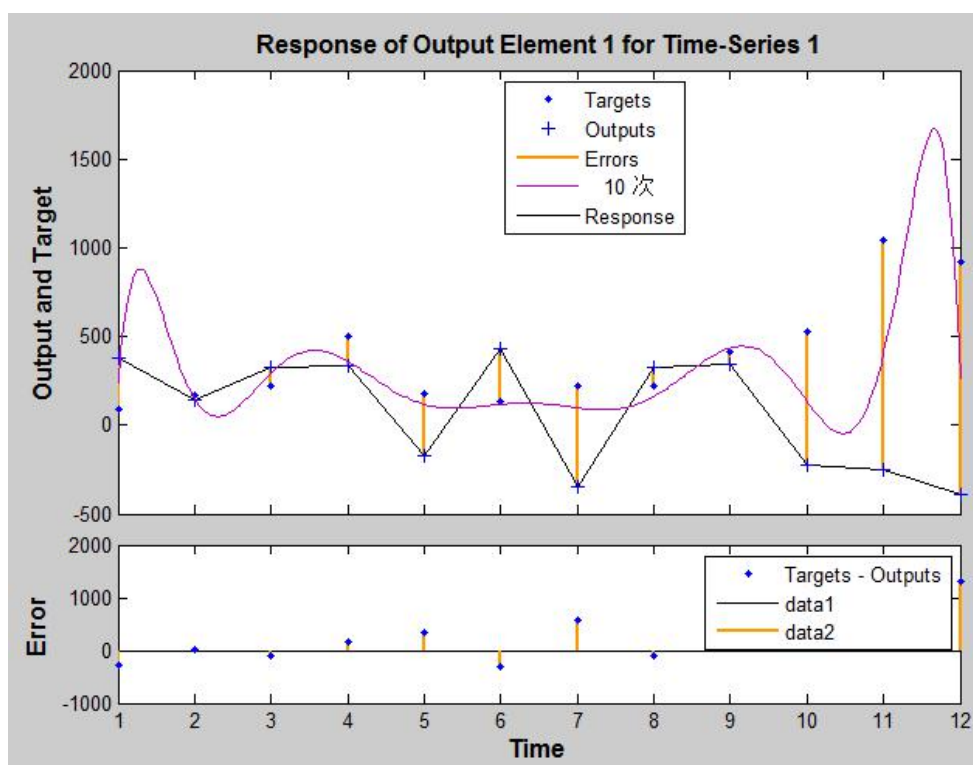


图 20 NAR 神经网络拟合效果

### 4.3 预测结果

利用 4 个池塘的前 15 周蓝藻含量构建 NAR 神经网络模型，预测第 16 周各池蓝藻含量，结果如表 11 所示。

表 11 NAR 神经网络预测结果

池塘编号	预测周数	预测蓝藻量 ( $10^6$ 个/L)
1	16	530.2644
2	16	300.9694
3	16	926.0669
4	16	291.4367

由题知，池塘 1 发生轻微水华，因此取池塘 1 从第一到第 15 周的平均蓝藻含量 345.714 作为水华发生判别标准。在此标准下，根据模型预测结果，认为 1 号和 3 号池塘在第 16 周发生水华，2 号和 4 号池塘在第 16 周不发生水华。

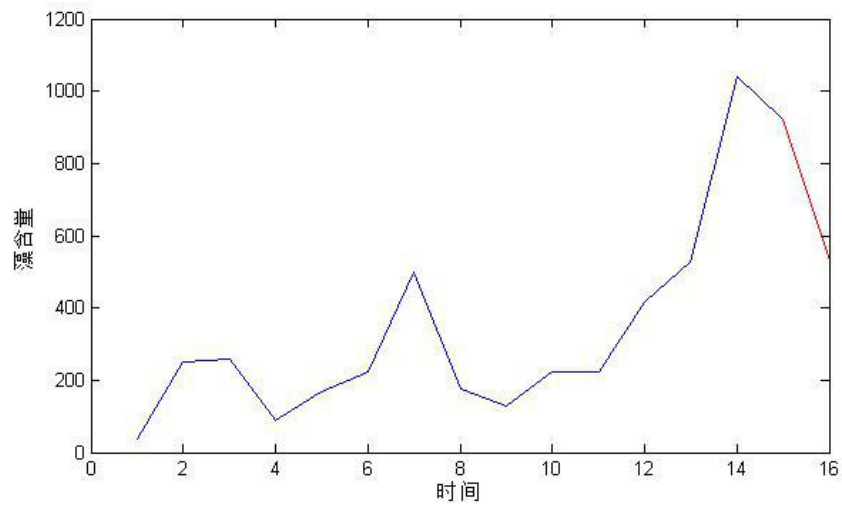


图 21 池塘 1 蓝藻含量及其预测

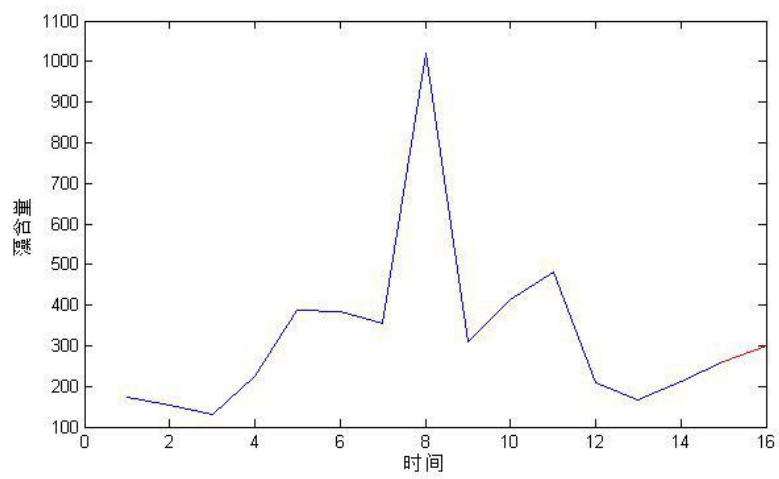


图 22 池塘 2 蓝藻含量及其预测

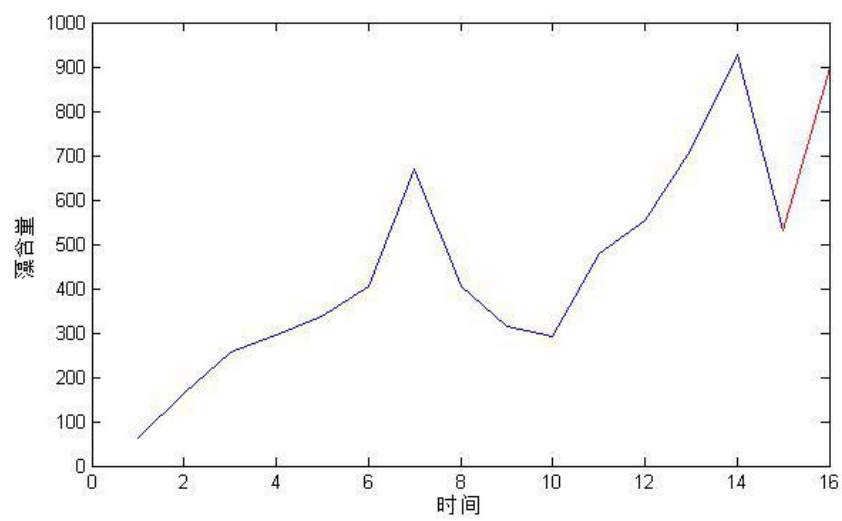


图 23 池塘 3 蓝藻含量及其预测



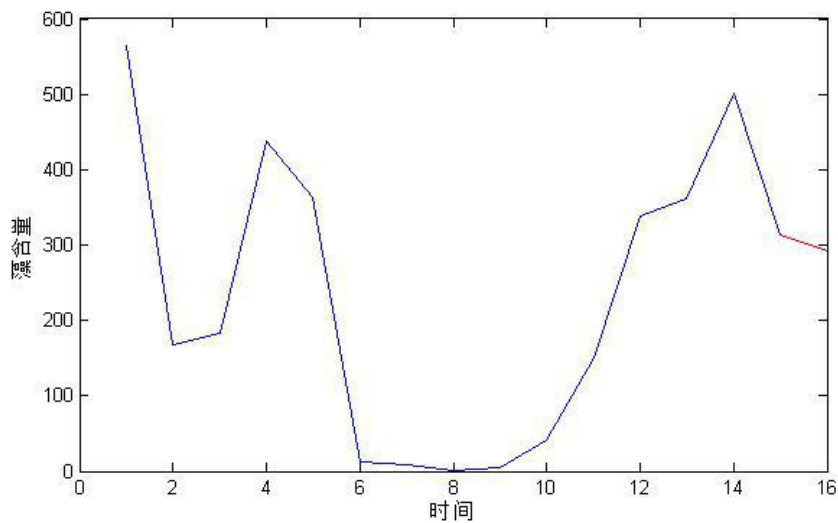


图 24 池塘 4 蓝藻含量及其预测

## 七 第四问

### 1 问题分析

附件 6 “鱼体重体长”和附件 7“鲢鱼鳙鱼相关数据”涉及到鱼类生长的相关信息。在鱼类学中，鱼类的生长与体重模型可以被描述为“不同年龄的鱼具有不同的平均体重”，一般有 5 种增长模型，但附件 7 中有关鲢鱼和鳙鱼相关数据量极少，不足建立模型。因此将“鱼类生长与体重相关模型”理解为描述鱼类生长时间、体重、体长三者间关系的模型，可以以“附件 6”为数据源，利用 matlab 拟合工具进行拟合，获取相应模型。

第 4 问的第 2 小问，希望构造一个与 1 号池一样大小的池塘，在其中放养鲢鱼或鳙鱼，通过鱼类的滤食作用净化 1 中藻类。净化意味着鱼类吃藻的速度大于藻类产生的速度，但是按照这种思路进行建模求解会有极多未知量，无法解算。因此将思路简化成，放养的鱼类在 15 周内能吃掉 1 号池中的藻类。假定鱼类摄食能量传递效率为 10%，且其食物构成与附件 7 中“某地区某湖水中各龄鲢鱼、鳙鱼饵料的生物组成”一致，则可以求定鱼类每增长 1g 需要消耗多少藻类。进一步代入藻类生物量与质量间的关系，即可求得鱼的总量。

### 2 鱼类生长与体重模型

#### 2.1 数据处理

附件 6 中时间数据以周为单位，522 周约为 10 年，而鱼的寿命普遍只有 3、4 年，因此



该数据的时间单位错误，应该更正为天。

对于附件 6 中鲢鱼、鳙鱼的体长、体重数据进行绘制，发现有较多数据正常值，为粗差，不能在拟合中使用，需要去粗。如图 1 所示，鳙鱼体长随时间变化数据中有明显的不合理的峰与谷出现，为错误数据。

利用 matlab 拟合工具箱中的排除工具，进行可视化的粗差排除，保留那些正常值作为进一步 matlab 拟合数据。

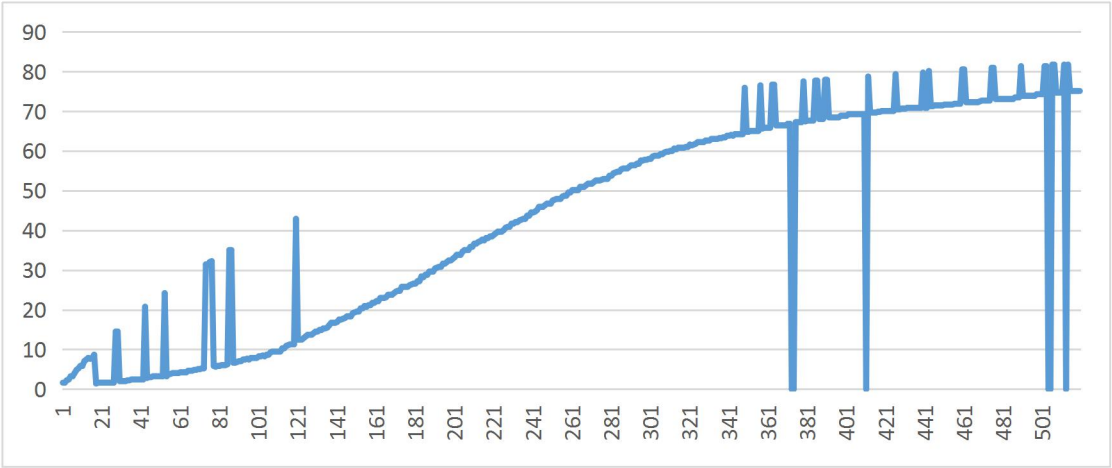


图 25 鳙鱼体长随时间变化

## 2.2 模型建立

使用 matlab 拟合工具箱对去除粗差后的数据进行拟合，其效果如图 2-7 所示，以其拟合方程作为鱼类生长与体重相关模型，即表 1。

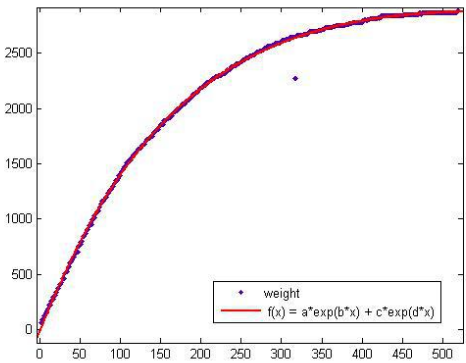


图 26 鲢鱼体重-时间拟合

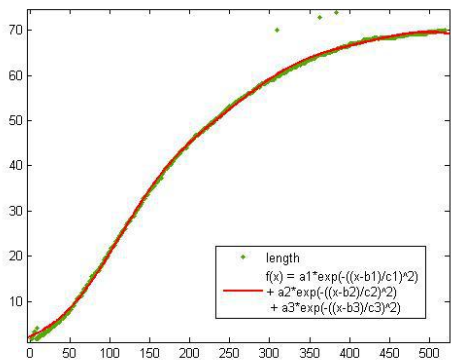


图 27 鲢鱼体长-时间拟合

鱼种类	关系	模型	R2
鲢鱼	体重-时间	$W_1(t) = 3687e^{-0.0003288t} - 3698e^{-0.005331t}$	0.9985
	体长-时间	$L_1(t) = 35.81 e^{-\frac{(t-289.8)^2}{147}} + 66.49 e^{-\frac{(t-533.5)^2}{220.8}} + 17.07 e^{-\frac{(t-154)^2}{97.22}}$	0.9988
	体重-体长	$W_1(L_1) = 329.6L_1^{0.5311} - 280.9$	0.9967
鳙鱼	体重-时间	$W_2(t) = 7572e^{-0.0004085t} - 7454e^{-0.00459t}$	0.9997
	体长-时间	$L_2(t) = 351.2 / (4.816 + 167 e^{-0.01648t})$	0.9985
	体重-体长	$W_2(L_2) = 3075 e^{0.00775L_2} - 2859e^{-0.1645L_2}$	0.9982

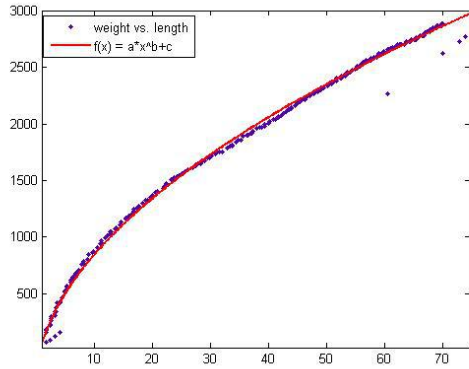


图 28 鲢鱼鱼体重-体长拟合

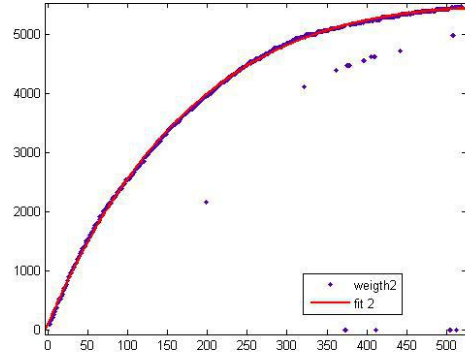


图 29 鳙鱼体重-时间拟合

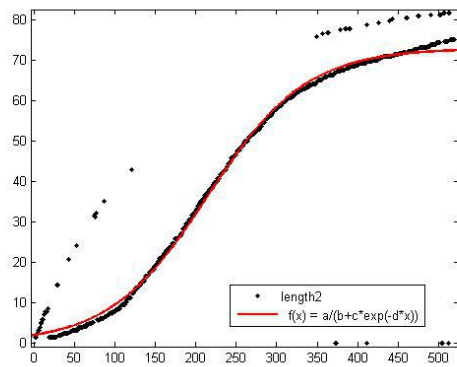


图 30 鳙鱼体长-时间拟合

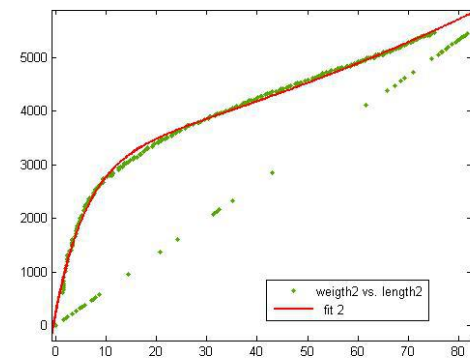


图 31 鳙鱼体重-体长拟合

表 12 鱼类生长与体重相关模型

### 3 构造净化池

#### 3.1 数据确定

在池水中主要含有蓝藻、绿藻、隐藻、硅藻、裸藻共五种藻类。鲢鱼 5 种藻类均有摄食，且这些藻类在其食物中所占百分比为，23.96%、3.9%、1.54%、38.13%、0.63%而鳙鱼则只摄食硅藻和绿藻，二者在其食物中占据 1.92%和 0.94%。

##### ①藻类含量：

第 15 周时，1 号池中有蓝藻： $3.3775 \times 10^{14}$  个；有绿藻： $1.1105 \times 10^{13}$  个；有隐藻： $5.5968 \times 10^{11}$  个；有硅藻： $6.36 \times 10^{11}$  个；有裸藻： $2.6076 \times 10^{11}$  个。

##### ②鱼类摄取藻量：

由鲢鱼的体重 - 时间模型求得，每条鲢鱼 15 周体重增长为： $w_1(105) - w_1(1) = 1449 - 7.4494 = 1441.55(g)$ ；由鳙鱼的体重-时间模型修的，每条鳙鱼 15 周体重增长为： $w_2(105) - w_2(1) = 2576.5 - 29.5925 = 2546.9(g)$ 。

假定能量传递效率为 10%，则可进一步求得每条鲢鱼 15 周的时间对藻类的消耗量：

消耗硅藻： $1441.55 \times 0.3813 \times 10 = 5496.63g$ ；

消耗绿藻： $1441.55 \times 0.039 \times 10 = 562.205g$ ；

消耗蓝藻： $1141.55 \times 23.96 \times 10 = 3453.954g$ ；

消耗裸藻： $1141.55 \times 0.0063 \times 10 = 90.818g$ ；

消耗隐藻： $1141.55 \times 0.0154 \times 10 = 221.999g$ 。

每条鳙鱼对藻类的消耗量：

消耗硅藻： $2546.9 \times 0.0192 \times 10 = 489.005g$ ；

消耗绿藻： $2546.9 \times 0.0094 \times 10 = 239.409g$ 。

#### 3.2 建立模型

要使得鱼类对藻类的消耗大于藻类含量，假设单个硅藻的重量是  $a$  克，单个绿藻的重量是  $b$  克，单个蓝藻的重量是  $c$  克，单个裸藻的重量是  $d$  克，单个隐藻的重量是  $e$  克，鲢鱼和鳙鱼的数量分别为  $x, y$ 。则可建立以下模型：

$$\begin{aligned} \min &= x + y \\ s.t. &\begin{cases} x \times \frac{5496.63}{a} + y \times \frac{489.005}{a} \geq 5.5968 \times 10^{11} \\ x \times \frac{562.205}{b} + y \times \frac{239.409}{b} \geq 1.1105 \times 10^{13} \\ x \times 3453.954 / c \geq 3.3775 \times 10^{14} \\ x \times 90.818 / d \geq 2.6076 \times 10^{11} \\ x \times 221.999 / e \geq 5.5968 \times 10^{11} \\ x, y \in N^+ \end{cases} \end{aligned}$$

### 3.3 模型求解与评价

令  $a=b=c=d=10^{-8}$ , 可以求得一个可行解为  $x=4470, y=636$ 。即应当在净化池中放养 4470 条鲢鱼和 636 条鳙鱼, 即鲢鱼的密度为 1.5 条/立方米, 鳙鱼的密度为 0.2 条/立方米。

由于藻类生物量和其质量间的转化关系无法获得, 因此这里采用预估值, 就模型求解结果而言, 这个预估值是合理的。由于模型没有考虑藻类的自增长和鱼类的细微差别, 因此模型不能全面客观地求解净化池中实际应当放养多少鱼, 仅能作为参考。

由模型建立条件知, 在该情况下, 净化池中的鱼能够较好的净化 1 号池中的藻类, 使藻类含量大幅下降。

## 八 第五问

### 1 问题分析

通过前 4 问的分析, 理清了“水体、底泥与间隙水中常见主要理化因子间的关系”、“虾池与鱼池对水质的影响”、“影响浮游生物量的主要理化因子”等问题, 并定量分析了鱼类对藻类的净化作用。第 5 问建立在前 4 问背景上, 其目的是建立一种养殖模式, 使其能完成水体自净化。前几问分析中对构建生态养殖模式有重要意义的是:

- ① 水中氮磷等营养物质主要来源于底泥;
- ② 增加水体溶氧, 保证氮磷等处于较低水平是提高水质的关键性因素;
- ③ 鱼类能够净化大量藻类, 防止水华发生。

针对这几点, 可以提出对应策略分别为“定时清除淤泥”、“引入植物等净化水体和增加溶氧”、“鱼虾混养净化藻类”。以这 3 个策略为指导, 并结合前人理论依据, 可以构建出合理的生态养殖模式。

### 2 理论依据

#### 2.1 水质恶化机理

当前我国池塘养殖产量与效率<sup>[11]</sup>的提高多依赖于提高鱼类放养密度和饲料投入。大量饲料投入池塘, 但由于饲料转化率较低, 因此相当比例的饲料以未被利用的残渣和鱼类排泄的形式存于水体中, 并逐步沉降并在底泥中积累。底泥中的有机质逐渐增多, 而它们的积累和分解会大量消耗水体溶氧。并且, 这些物质分解产生的氮、磷无机营养盐和有机物, 会引起水体富营养化和有机物污染, 造成水质恶化, 影响鱼类生长。水体富营养化则会造成浮游植物, 如蓝藻等的爆发性增长。最终水质环境的快速恶化造成鱼类死亡, 水质恶化进入恶性循环。

## 2.2 生物净化作用

生物净化作用<sup>[12]</sup>是指利用生物的生命代谢活动来降低存在于环境中有害物质的浓度或使其完全无害化,从而使受到污染的生态环境能够部分或完全恢复到原初状态的过程。它包括利用植物、动物和微生物吸收、降解、转化水体和底泥中的污染物,使污染物的浓度降低到可接受的水平,或将有毒有害的污染物转化为无害的物质,亦或将污染物稳定化,以减少其向周边环境的扩散。其主要方式有 3 种,第一是“配养生态互补型水生动物”,如鱼虾混养或主养吃鱼的池塘中搭配一定量的鲢鳙鱼;第二是“投放微生态制剂”,如在水体中投放芽孢杆菌、光合细菌、硝化和反硝化细菌、酵母菌、复合微生物制剂等;第三是“人为决定藻类优势种和种植水生高等植物”,在池塘中种植有益藻类,它不仅可以吸收有害藻类所需要的营养,而且,一旦它成为优势种,必然会限制其他无益藻类的生长。至于水生高等植物,他们不仅能够快速吸收水体和沉积物中的营养盐,而且对水体生态系统的物理、化学及生物学特性亦有重要影响,有助于稳定水体生态。

## 2.3 已有生态养殖模式

①多品种混养:利用鱼类不同的生活习性!在同一池塘内混养多种鱼类!从而充分利用池塘的各个水层和食物链的作用,有利于保持良好的水质,这是我国池塘养鱼的重要内容”。

②鱼菜共生系统:我国传统的稻田养鱼、莲田养鱼、苇塘养鱼等是鱼菜共生系统的雏形,将其原理应用于池塘并改进提高是池塘生态养殖的重要方向。

③桑基鱼塘:基面种桑、桑叶养蚕、蚕沙喂鱼、塘泥肥桑,通过桑叶、蚕沙、塘泥之间的物质循环和能量流动联成一个完整的农业生态系统—桑基鱼塘系统。

## 3 生态养殖模式

综合前述分析,结合前人经验,构建一种生态养殖模式如图 32 所示。

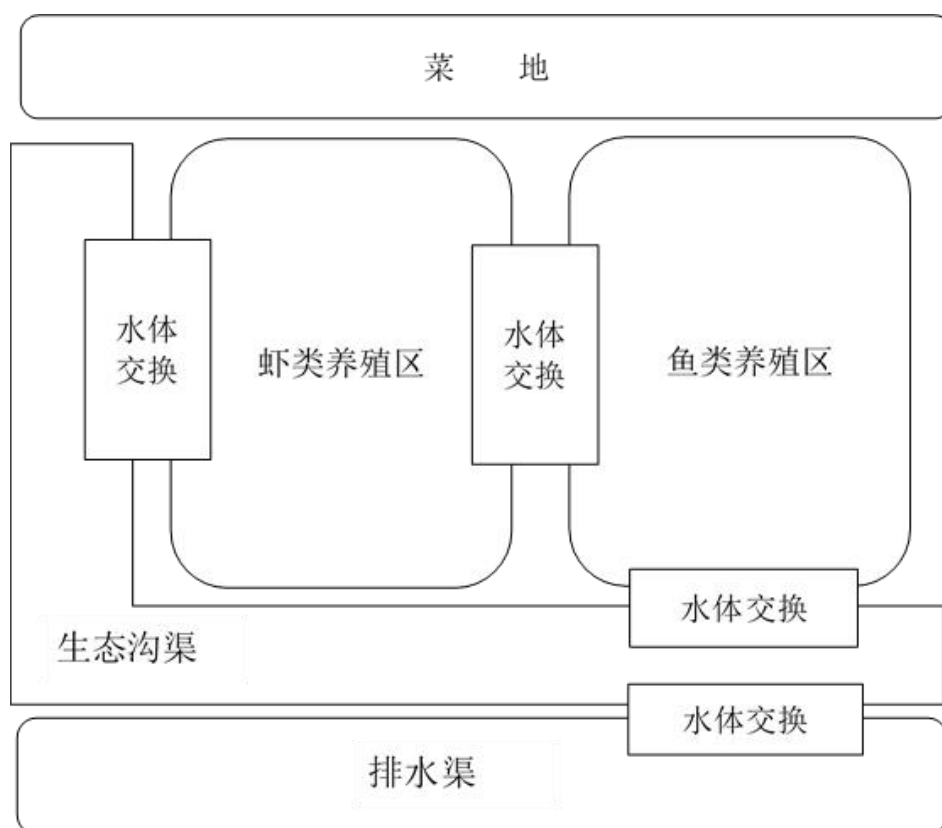


图 32 生态养殖模式示意图

该生态模式主要包括菜地、养殖区、生态沟渠、排水渠四个部分。其中养殖区为鱼虾分池喂养，中间进行水体交换，其核心目的是利用鱼类净化虾池中快速生长的藻类，防止水华爆发，并且分池养殖有利于管理和分别喂养。养殖区域由于饲料转换率较低，因此一段时间后底泥中会沉积较多营养物质，为防止营养物质的分解转化污染水池，进行定期的池底清淤工作，并把淤泥用作菜地肥料，实现营养物质再利用。菜地产出可以通过某种途径，再转化为养殖区域鱼虾的饲料，从而实现能量的高效利用。

生态沟渠的是生态养殖模式能进行水体自进化的核心，其作用是直接性的水体净化。在生态沟渠中种植水鳖、睡莲、伊乐藻、黑藻、金鱼藻、水芹、芦苇、茭尾、葛蒲等水生植物，这些水生植物的综合作用能有效去除水体中的 N 与 P，减少 COD 并增加水体溶氧。

同时，由于小球藻能够吸收 N、P 产生氧气，具有较好的水质净化效果，并能直接成为鱼类食物。因此在生态养殖模式的各个组件水体中接种适量小球藻，用于辅助净化水体。

当该生态模式出现失衡，水质迅速下降时，可以引入外部水源，将原有的水质较差的池水，通过排水渠排出。在排出前，综合使用理化方法处理，使水质达到排放标准，不至于污染环境。

## 九 总结评价

总体来说本文基于 SPSS、Matlab 和 Lingo 软件，综合利用相关性分析、主成分分析、因子分析、拟合、聚类分析等方法进行数据分析，建立了 BP 神经网络、NAR 神经网络及线性规划等模型，全面探讨了水产养殖池塘的 5 个问题。取得了较好的效果。

但是，由于写作的原因，中间步骤必然会有部分疏漏。就模型而言，存在普适性的欠缺，就数据分析而言，尚未做到尽善尽美。

## 参考文献

- [1] 陈永川,汤利. 沉积物-水体界面氮磷的迁移转化规律研究进展[J]. 云南农业大学学报, 2005(4)
- [2] 蔡金榜,李文奇,逢勇,等. 洋河水库水质主成分分析[J]. 中国环境监测, 2007(2): 62-66.
- [3] 俞盈,付广义,陈繁忠,等. 水体中三氮转化规律及影响因素研究[J]. 地球化学, 2008(6):
- [4] 冯健,崔广柏,张红举. 太湖水质指标主成分分析[J]. 水资源保护, 2004(3): 49-50, 63-70.
- [5] 张巍,王学军,江耀慈,等. 太湖水质指标相关性与富营养化特征分析[J]. 环境污染与防治, 2002(1): 50-53.
- [6] 钟晶晶,刘茂松,王玉,等. 太湖流域河流与湖泊间主要水质指标的空间关联特征[J]. 生态学杂志, 2014(8): 2176-2182.
- [7] 张克鑫,陆开宏,朱津永,等. 基于 BP 神经网络的藻类水华预测模型研究[J]. 中国环境监测, 2012(3): 53-57.
- [8] 殷高方,张玉钧,胡丽,等. BP 神经网络水华预测模型的敏感性分析[J]. 北京理工大学学报, 2012(12): 1288-1293.
- [9] 殷高方,张玉钧,胡丽,等. BP 神经网络在水华短期预测中的应用[J]. 北京理工大学学报, 2012(6): 655-660.
- [10] 汪翔,何吉祥,余磊,张静. 基于 NAR 神经网络对养殖水体亚硝酸盐预测模型的研究[J]. 渔业现代化,2015,04:30-34..
- [11] 刘兴国. 池塘养殖污染与生态工程化调控技术研究\_刘兴国[D]. [出版地不详]: 南京农业大学, 2011.
- [12] 张德隆,李刚,肖志国,等. 我国池塘生态养殖的回顾与展望\_张德隆[J]. 中国渔业经济, 2004(2): 44-45.



## 附录

### 求标准化矩阵

```
function std=cwstd(vector)
%FUNCTION 求向量的标准化矩阵
cwsun=sum(vector,1); %求向量的和
[a,b]=size(vector); %求向量的维数
std=ones(a,b);

for i=1:a
    for j=1:b
        std(i,j)=vector(i,j)/cwsun(j);
    end
end
```

### 求解相关系数矩阵、特征根特征向量

```
function result=cwfac(vector)
%RESULT 求向量的相关系数矩阵、特征根特征向量
fprintf('向量的维数:\n');
std=corrcoef(vector); %求向量的相关系数矩阵
fprintf('g\n',std);
disp(std);
fprintf('特征根和特征向量:\n');
[vec,val]=eig(std); %求特征根和特征向量
fprintf('vec=\ng\n',vec);
fprintf('val=\ng\n',val);
disp(vec);
disp(val);

newval=diag(val);
[y,i]=sort(newval); %求特征根和特征向量的排序
fprintf('排序后的特征根和特征向量:\n');
for z=1:length(y)
    newy(z)=y(length(y)+1-z);
end
%newy=排序后的特征根和特征向量
fprintf('g\n',newy);
disp(newy);
```

```

rate=y/sum(y);
fprintf('1+I×AÊ°\n');
newrate=newy/sum(newy);
disp(newrate);
sumrate=0;
newi=[];
for k=length(y):-1:1
    sumrate=sumrate+rate(k);
    newi(length(y)+1-k)=i(k);
    if sumrate>1.85
        break;
    end
end
%ÇİÂÂÛ¼E¹±İ×ÂÊ´óÓÚ85%µÄÏØ÷ÖµµÄÐò°Ã·ÃÊënewiÖÐ

fprintf('Ö÷³É·ÖÊý:%g\n\n',length(newi));
fprintf('Ö÷³É·ÖÏØ°É:\n');
for p=1:length(newi)
    for q=1:length(y)
        result(q,p)=sqrt(newval(newi(p)))*vec(q,newi(p));
    end
end
disp(result);

```

## 主成分分析打分

```
function score=cwscore(vector1,vector2)
%FUNCTION ¼ÆËäµÃ·Ö $
sco=vector1*vector2;
csum=sum(sco,2);
[~,i]=sort(-1*csum);
[~,j]=sort(i);
fprintf('¼ÆËäµÃ·Ö:\n');
score=[sco,csum,j];
%µÃ·Ö¼ØÓó°scoîª,÷÷³É·ÖµÃ·ÖË»csumîª×Û°îµÃ·ÖË»jîªÃÐò¼ª¹ü
disp(score);
```

## 神经元算法拟合浮游生物量

```
load('μÚÊŸÎÊÊŸ³⁴Ÿ.mat');
%ÔòÎªÊñ³⁴-ÔªÊä ···ÖĐĂĈÒ»ĂĐ×÷ÎªÒ»ÎöÊŸ³⁴Ÿ£-ÈùòÔò×öÒ»´Î×ªÖǺ
%input=input';
%output=output';
```

```

R=zeros(1,11);
input=[input;input];
output=[output;output];

for i=2:12
    [inputn,inputps]=mapminmax(input');
    [outputn,outputps]=mapminmax(output');
    net=newff(inputn,outputn,i);
    net=train(net,inputn,outputn);
    net=train(net,inputn,outputn);
    %Ô¤²â
    output_yucen=sim(net,inputn);

    output_yuce=mapminmax('reverse',output_yucen,outputps);

    R(i-1)=corr(output,output_yuce')^2;
    if R(i-1)>0.95
        break;
    end
end

disp(R);
plot(R);

```

## 神经元算法拟合亚硝态氮

```

load('µÜËÿîÊËÿŸ.mat');
%ÒòîªÉñ¼-ÔªËã·ÖÐÃ;ÒªÁÐ×÷îªÒ»îöÊÿŸŸ£-ÊùÒÒ×öÒ»´î×ªÖÃ
%inputÀiÃæÊÇ,;ÓîÉúîiÁÇ;¢ÈÜÑð;¢PH
input=[fuyoushengwuliang,rongyang,PH];
output=yaxiaotaidan;
%input=input';
%output=output';
R=zeros(1,11);
input=[input;input];
output=[output;output];

for i=2:12
    [inputn,inputps]=mapminmax(input');
    [outputn,outputps]=mapminmax(output');
    net=newff(inputn,outputn,i);
    net=train(net,inputn,outputn);

```

```

net=train(net,inputn,outputn);
%Ô¤²â
output_yucen=sim(net,inputn);

output_yuce=mapminmax('reverse',output_yucen,outputps);

R(i-1)=corr(output,output_yuce')^2;
if R(i-1)>0.95
    break;
end
end

disp(R);
plot(R);

```

## 神经元短发拟合溶氧

```

load('µÜËÿîÊËÿÝ.mat');
%ÒòîªÉñ¼-ÔªËã·ÖÐÃ;Ò»Ã×÷îªÒ»îðËÿÝ£-ËùÒÔ×öÒ»´îªÖÃ
%inputÀiÃæÊÇÑÇîðî-µª;ç.;óîÉúîiÃ;çPH
input=[yaxiaotaidan,fuyoushengwuliang,PH];
output=rongyang;
%input=input';
%output=output';
R=zeros(1,11);
input=[input;input];
output=[output;output];

for i=2:12
    [inputn,inputps]=mapminmax(input');
    [outputn,outputps]=mapminmax(output');
    net=newff(inputn,outputn,i);
    net=train(net,inputn,outputn);
    net=train(net,inputn,outputn);
    %Ô¤²â
    output_yucen=sim(net,inputn);

    output_yuce=mapminmax('reverse',output_yucen,outputps);

    R(i-1)=corr(output,output_yuce')^2;
    if R(i-1)>0.95
        break;
    end
end

```

```
end
```

```
disp(R);
```

```
plot(R);
```

## 神经元算法拟合 PH

```
load('púÈÿîÊËÿŷ.mat');
%ÒòîªÉñ¼-ÔªËã·ÖÐÃ;Ò»ÁÐ×÷îªÒ»ìòÈÿŷŷŷ-ÈùÒÔ×òÒ»´î×ªÖÃ
%inputÀiÃæÊÇÑÇîòî-µª;çÈÜÑð;ç.;óîÉúîiÁ;
input=[yaxiaotaidan,rongyang,fuyoushengwuliang];
output=PH;
%input=input';
%output=output';
R=zeros(1,11);
input=[input;input];
output=[output;output];

for i=2:12
    [inputn,inputps]=mapminmax(input');
    [outputn,outputps]=mapminmax(output');
    net=newff(inputn,outputn,i);
    net=train(net,inputn,outputn);
    net=train(net,inputn,outputn);
    %Ô¤²â
    output_yucen=sim(net,inputn);

    output_yuce=mapminmax('reverse',output_yucen,outputps);

    R(i-1)=corr(output,output_yuce')^2;
    if R(i-1)>0.95
        break;
    end
end

disp(R);
plot(R);
```

## 求亚硝态氮的最小值

```
load('ÑÇîòî-µªµÃÃª°î');
%input_test=[359.0056;7.32;8.49];
```

```

input_test=[359.0056;2.4;9.33];

inputn_test=mapminmax('apply',input_test,inputps);

outputn_test=sim(net,inputn_test);

output_test=mapminmax('reverse',outputn_test,outputps);

disp(output_test);

```

## 求溶氧的最小值

```

load('EÜÑöµÄÄ°İ');
%input_test=[0.7696;359.0056;8.49];
input_test=[0.0050;359.0056;9.33];

inputn_test=mapminmax('apply',input_test,inputps);

outputn_test=sim(net,inputn_test);

output_test=mapminmax('reverse',outputn_test,outputps);

disp(output_test);

```

## 求 PH 的最大值

```

load('PHµÄÄ°İ');

input_test=[0.0050;2.4;359.0056];

inputn_test=mapminmax('apply',input_test,inputps);

outputn_test=sim(net,inputn_test);

output_test=mapminmax('reverse',outputn_test,outputps);

disp(output_test);

```

## 磷酸盐磷-总磷

以总磷为横坐标，磷酸盐磷为纵坐标可得到一次多项式耦合时数据如下：

Linear model Poly1:

$$f(x) = p1*x + p2$$

Coefficients (with 95% confidence bounds):

$$p1 = 3.298 \quad (2.41, 4.187)$$

$$p2 = -0.3878 \quad (-0.5418, -0.2338)$$

Goodness of fit:

SSE: 0.02801

R-square: 0.8319

Adjusted R-square: 0.819

RMSE: 0.04642

二次多项式耦合时可得到数据如下：

Linear model Poly2:

$$f(x) = p1*x^2 + p2*x + p3$$

Coefficients (with 95% confidence bounds):

$$p1 = 33.9 \quad (21.14, 46.66)$$

$$p2 = -8.416 \quad (-12.85, -3.981)$$

$$p3 = 0.5953 \quad (0.2161, 0.9746)$$

Goodness of fit:

SSE: 0.007387

R-square: 0.9557

Adjusted R-square: 0.9483

RMSE: 0.02481

三次多项式耦合时可得到数据如下：

Linear model Poly3:

$$f(x) = p1*x^3 + p2*x^2 + p3*x + p4$$

Coefficients (with 95% confidence bounds):

$$p1 = -50.36 \quad (-645.4, 544.7)$$

$$p2 = 59.67 \quad (-245.1, 364.5)$$

$$p3 = -12.7 \quad (-63.5, 38.11)$$

$$p4 = 0.8259 \quad (-1.928, 3.58)$$

Goodness of fit:

SSE: 0.007364

R-square: 0.9558

Adjusted R-square: 0.9438

RMSE: 0.02587

四次多项式耦合时数据如下：

Linear model Poly4:

$$f(x) = p1*x^4 + p2*x^3 + p3*x^2 + p4*x + p5$$

Coefficients (with 95% confidence bounds):

p1 = 1.673e+04 (-6477, 3.993e+04)  
p2 = -1.171e+04 (-2.79e+04, 4476)  
p3 = 3062 (-1113, 7236)  
p4 = -350.3 (-821, 120.5)  
p5 = 14.79 (-4.757, 34.34)

Goodness of fit:

SSE: 0.005854

R-square: 0.9649

Adjusted R-square: 0.9508

RMSE: 0.02419

五次多项式拟合的精度如下:

Linear model Poly5:

$$f(x) = p1*x^5 + p2*x^4 + p3*x^3 + p4*x^2 + p5*x + p6$$

Coefficients (with 95% confidence bounds):

p1 = -7.494e+04 (-1.106e+06, 9.562e+05)  
p2 = 8.299e+04 (-8.291e+05, 9.951e+05)  
p3 = -3.489e+04 (-3.542e+05, 2.844e+05)  
p4 = 7065 (-4.82e+04, 6.233e+04)  
p5 = -691.7 (-5417, 4034)  
p6 = 26.28 (-133.2, 185.8)

Goodness of fit:

SSE: 0.005836

R-square: 0.965

Adjusted R-square: 0.9455

RMSE: 0.02546

一次指数函数的拟合精度如下:

General model Exp1:

$$f(x) = a*\exp(b*x)$$

Coefficients (with 95% confidence bounds):

a = 0.005792 (0.002632, 0.008952)  
b = 19.04 (16.36, 21.72)

Goodness of fit:

SSE: 0.008201

R-square: 0.9508

Adjusted R-square: 0.947

RMSE: 0.02512

二次指数函数的拟合精度如下:

General model Exp2:

$$f(x) = a*\exp(b*x) + c*\exp(d*x)$$

Coefficients (with 95% confidence bounds):



```

a = 1.429e+11 (-2.05e+13, 2.078e+13)
b = -245.1 (-1457, 967.1)
c = 0.005174 (0.002058, 0.00829)
d = 19.58 (16.64, 22.52)

```

Goodness of fit:

SSE: 0.006996

R-square: 0.958

Adjusted R-square: 0.9466

RMSE: 0.02522

经过上面的拟合，可以发现，磷酸盐磷与总磷之间的关系是正相关关系，但用一次多项式拟合效果并不理想。当将耦合函数的次数增大或者是改为二次指数函数时，发现拟合效果比较理想，并且一次多项式拟合和二次多项式拟合的效果差别比较大，但是二次多项式拟合和更高次的多项式、指数函数拟合的效果差别并不显著。由此我们推论，对于磷酸盐磷-总磷模型来说，要想达到更高的拟合精度，必须为模型提供斜率变化的机会（因为一次多项式的斜率不能发生变化，而高次多项式和指数函数的斜率是可以变化的。）。也就是说，随着总磷浓度的增加，磷酸盐磷浓度的增加的速率不是一定的。

并且从图像中我们可以看到斜率是有增大的趋势的。为了进一步讨论斜率变化的趋势，我们做了如下处理：

调换 x 坐标 y 坐标，即以总磷为纵坐标，以磷酸盐磷为横坐标，做一次。二次指数函数拟合：

一次指数函数拟合数据如下：

General model Exp1:

$$f(x) = a \cdot \exp(b \cdot x)$$

Coefficients (with 95% confidence bounds):

```

a = 0.1345 (0.1225, 0.1466)
b = 1.305 (0.9371, 1.674)

```

Goodness of fit:

SSE: 0.002524

R-square: 0.802

Adjusted R-square: 0.7867

RMSE: 0.01393

二次指数函数拟合数据如下：

General model Exp2:

$$f(x) = a \cdot \exp(b \cdot x) + c \cdot \exp(d \cdot x)$$

Coefficients (with 95% confidence bounds):

```

a = 0.1619 (0.09261, 0.2313)
b = 0.7839 (-0.3173, 1.885)
c = -0.1033 (-0.2308, 0.02424)
d = -14.47 (-47.36, 18.43)

```

Goodness of fit:

SSE: 0.001508

R-square: 0.8817

Adjusted R-square: 0.8494

RMSE: 0.01171

可以看出，二次指数函数拟合的精度明显大于一次指数函数拟合的精度。我们做了如下分析：因为函数总的趋势是递增的，所以无论一次指数函数的各项系数如何变化，其斜率趋势是递增的。而二次函数不存在这个限制条件。而得到的结果是，二次指数函数拟合的精度明显大于指数函数拟合的精度，所以我们可以得到结论：当以总磷为纵坐标，以磷酸盐磷为横坐标时，函数的斜率是递减的。

也就可以得到，当以总磷为横坐标，以磷酸盐磷为纵坐标时，函数的斜率是递增的。

结合上面的分析，可以得到：随着总磷浓度的增加，磷酸盐磷的浓度是随之增长的，而且增长的速率是越来越快的。

而通过对比两种坐标模式下的拟合数据，我们发现当以总磷为横坐标，以磷酸盐磷为纵坐标时，拟合的精度更高，由此我们得出以下结论：磷酸盐磷的浓度是和总磷的浓度相关的，是受总磷浓度影响的。

原因假设如下：在水中，磷元素是存在化学平衡的。当总磷的浓度增加时，化学平衡朝着其他形式的磷转化为磷酸盐磷的方向移动。因此，随着总磷浓度的增加，磷酸盐磷浓度增加的速率是递增的。

## 硝态氮-总氮

以硝态氮为纵坐标，以总氮为横坐标，进行拟合，得到只有三次高斯函数拟合的精度较为合理。但拟合得到的图像没有明显的趋势，数据与图像如下：

General model Gauss3:

$f(x) =$

$$a1 \cdot \exp(-((x-b1)/c1)^2) + a2 \cdot \exp(-((x-b2)/c2)^2) + a3 \cdot \exp(-((x-b3)/c3)^2)$$

Coefficients (with 95% confidence bounds):

a1 =	-0.5604	(-17.37, 16.25)
b1 =	0.163	(0.04912, 0.2768)
c1 =	0.01973	(-0.1127, 0.1522)
a2 =	-3.19	(-3.082e+05, 3.082e+05)
b2 =	0.2072	(-65.96, 66.37)
c2 =	0.0306	(-15.34, 15.41)
a3 =	3.097	(-3.082e+05, 3.082e+05)
b3 =	0.2057	(-77.44, 77.85)
c3 =	0.0309	(-13.72, 13.78)

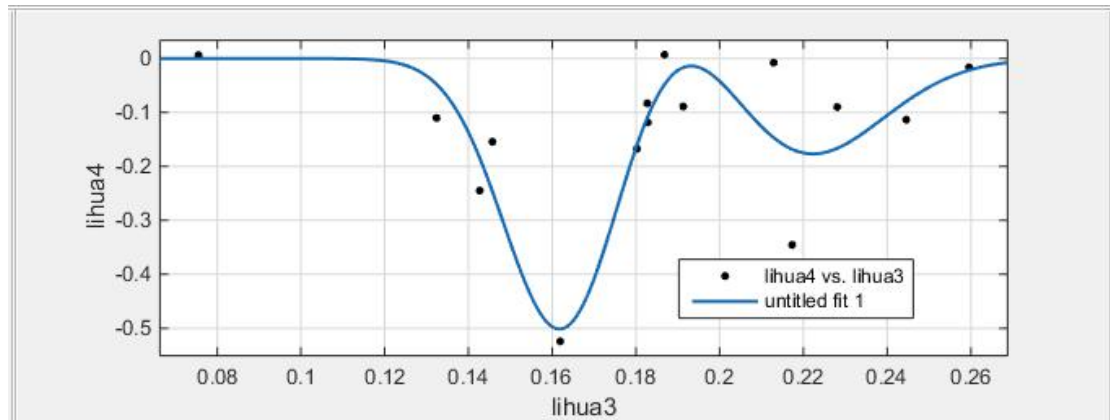
Goodness of fit:

SSE: 0.0833

R-square: 0.7106

Adjusted R-square: 0.3248

RMSE: 0.1178



而用一次多项式拟合得到的函数的斜率几乎为 0，这说明二者之间没有明显的趋势。通过高斯函数拟合，可以看出在总氮指标为 0.16 左右的位置和在总氮指标为 0.22 左右的位置，硝态氮-总氮指标曲线存在两个低谷。

## 亚硝态氮-总氮

以亚硝态氮指标为纵坐标，以总氮指标为横坐标，进行拟合发现：

当进行多项式拟合时，函数曲线没有表现出明显的趋势，拟合精度也非常低。这说明二者之间没有明显的相关性。

而以二次高斯函数进行拟合时，发现拟合精度远远大于其他形式的拟合，数据与图像如下：

General model Gauss2:

$$f(x) = a1 \cdot \exp(-((x-b1)/c1)^2) + a2 \cdot \exp(-((x-b2)/c2)^2)$$

Coefficients (with 95% confidence bounds):

a1 =	1.473	(0.05509, 2.891)
b1 =	0.2206	(0.2187, 0.2226)
c1 =	0.004503	(0.002796, 0.00621)
a2 =	0.3568	(0.1851, 0.5286)
b2 =	0.1583	(0.1525, 0.1642)
c2 =	0.01911	(0.01176, 0.02646)

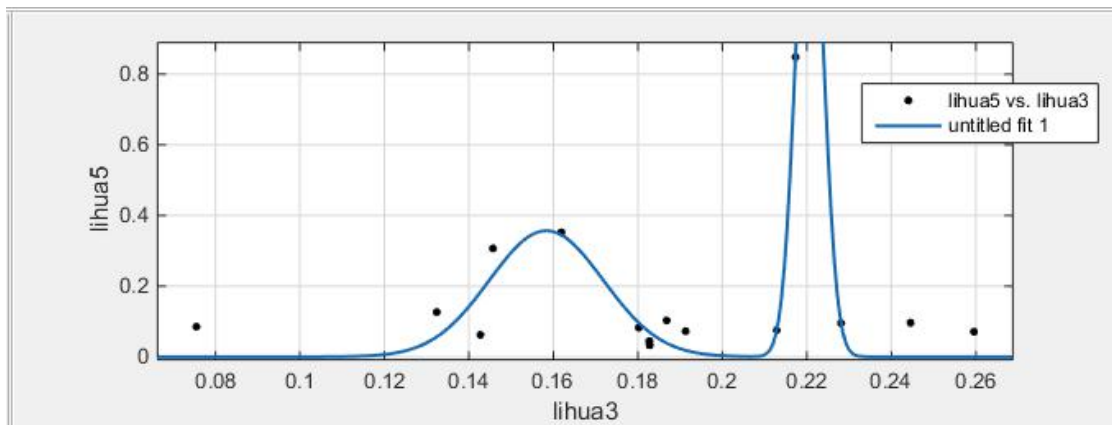
Goodness of fit:

SSE: 0.0561

R-square: 0.9094

Adjusted R-square: 0.859

RMSE: 0.07895



可以发现，在总氮指数为 0.16 和 0.22 的位置处，亚硝态氮指数-总氮指数曲线有两个明显的峰。结合硝态氮-总氮指数曲线的分析，发现，亚硝态氮-总氮曲线极大值所在的位置和硝态氮-总氮曲线极小值所在的位置是大致一样的。

经过分析，我们对此现象做出如下解释：

在池塘中，亚硝态氮和硝态氮之间的关系是“此消彼长”的关系。而两个极值的出现很可能是由于投放饲料和捕捞收货有关系的。

## 亚硝态氮-硝态氮

以亚硝态氮指标为纵坐标，以硝态氮指标为横坐标进行拟合发现：

以四次多项式拟合可以得到较高的精度值，数据和图像如下：

Linear model Poly4:

$$f(x) = p1*x^4 + p2*x^3 + p3*x^2 + p4*x + p5$$

Coefficients (with 95% confidence bounds):

p1 =	-279	(-501.9, -56.14)
p2 =	-232.7	(-446, -19.31)
p3 =	-50.31	(-110.7, 10.13)
p4 =	-3.16	(-8.335, 2.015)
p5 =	0.08043	(-0.04067, 0.2015)

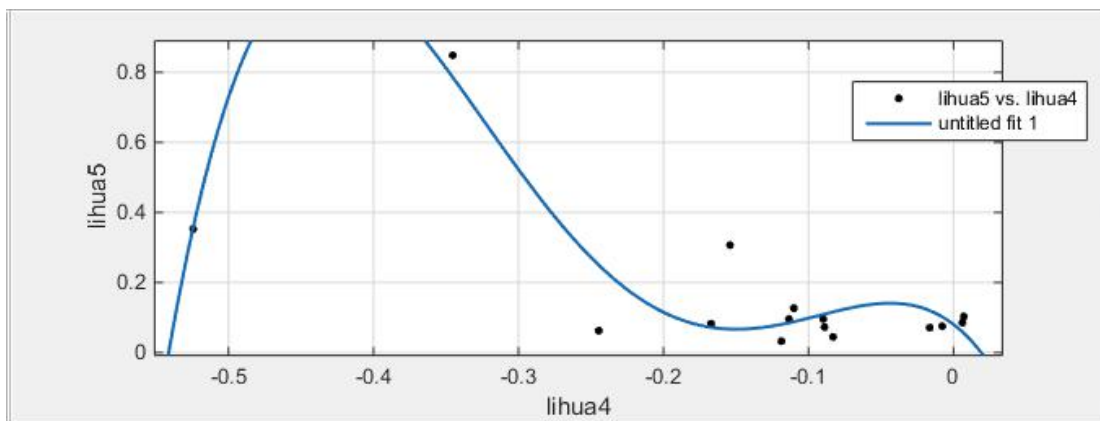
Goodness of fit:

SSE: 0.1135

R-square: 0.8167

Adjusted R-square: 0.7433

RMSE: 0.1065



可以发现，二者呈现出较为明显的负相关关系。原因可能是亚硝态氮与硝态氮之间的相互转化。与上面的分析结果是一致的。

## 铵态氮-总氮

以铵态氮为纵坐标，以总氮为横坐标进行拟合发现：

当用二次高斯函数进行拟合时，拟合精度相对较好。拟合数据和图像如下：

General model Gauss2:

$$f(x) = a1 \cdot \exp(-((x-b1)/c1)^2) + a2 \cdot \exp(-((x-b2)/c2)^2)$$

Coefficients (with 95% confidence bounds):

a1 =	-0.2056	(-0.2945, -0.1167)
b1 =	0.1203	(0.07277, 0.1678)
c1 =	0.0505	(-0.008762, 0.1098)
a2 =	-0.2178	(-0.2675, -0.168)
b2 =	0.2348	(0.2042, 0.2653)
c2 =	0.06659	(-0.03965, 0.1728)

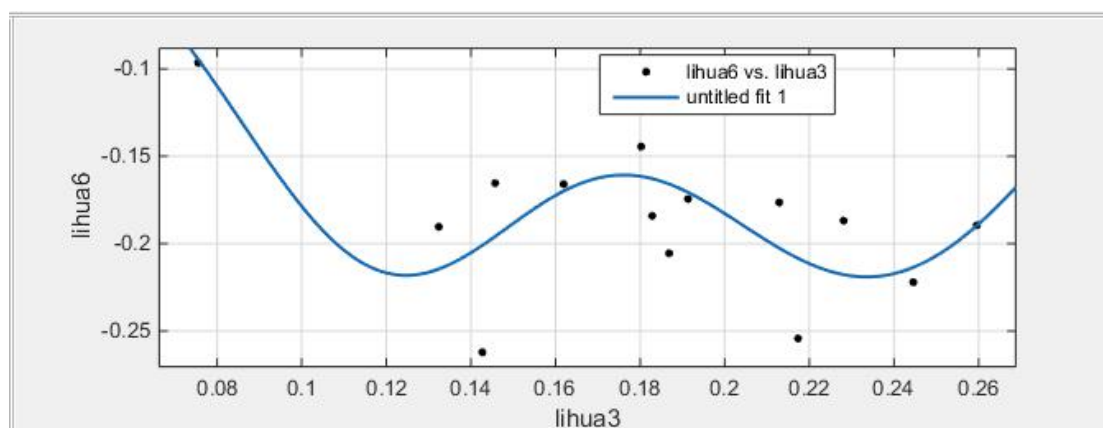
Goodness of fit:

SSE: 0.0138

R-square: 0.5059

Adjusted R-square: 0.2314

RMSE: 0.03915



可以发现，在总氮指标为 0.18 左右的地方，曲线出现了极大值。同时也存在两个极小值。而这两个极小值分别在亚硝态氮-总氮曲线和硝态氮指数-总氮指数曲线极值的两侧。

原因分析如下：（不知道……）

## 总氮-总磷

以总氮为纵坐标，以总磷为横坐标进行拟合：

五次多项式拟合数据和图像如下：

Linear model Poly5:

$$f(x) = p1 \cdot x^5 + p2 \cdot x^4 + p3 \cdot x^3 + p4 \cdot x^2 + p5 \cdot x + p6$$

Coefficients (with 95% confidence bounds):

p1 = 9.84e+05 (-6.706e+04, 2.035e+06)  
p2 = -8.877e+05 (-1.817e+06, 4.205e+04)  
p3 = 3.161e+05 (-9384, 6.416e+05)  
p4 = -5.551e+04 (-1.118e+05, 830.7)  
p5 = 4804 (-13.14, 9621)  
p6 = -163.7 (-326.3, -1.09)

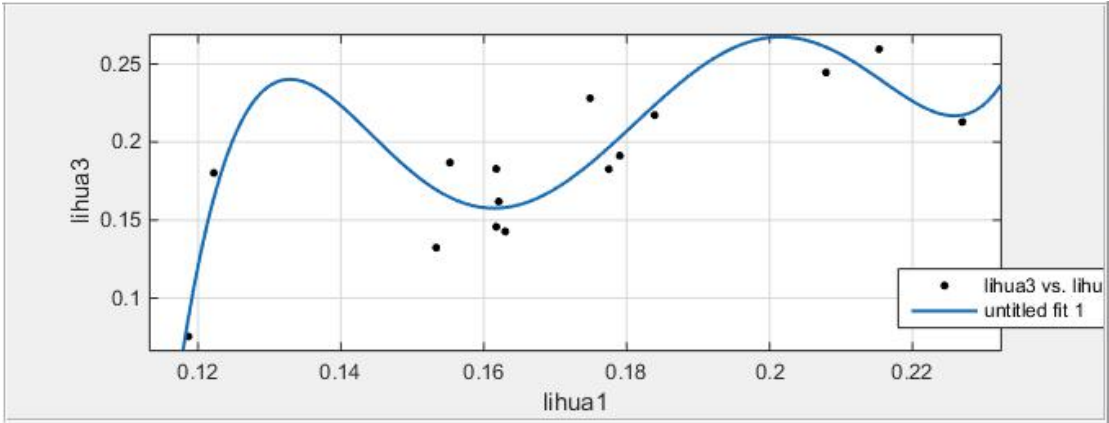
Goodness of fit:

SSE: 0.006064

R-square: 0.807

Adjusted R-square: 0.6997

RMSE: 0.02596



四次多项式拟合数据和图像如下：

Linear model Poly4:

$$f(x) = p1 \cdot x^4 + p2 \cdot x^3 + p3 \cdot x^2 + p4 \cdot x + p5$$

Coefficients (with 95% confidence bounds):

p1 = -1.76e+04 (-4.651e+04, 1.131e+04)  
p2 = 1.183e+04 (-8339, 3.2e+04)  
p3 = -2936 (-8137, 2265)  
p4 = 319.8 (-266.7, 906.3)  
p5 = -12.78 (-37.13, 11.58)

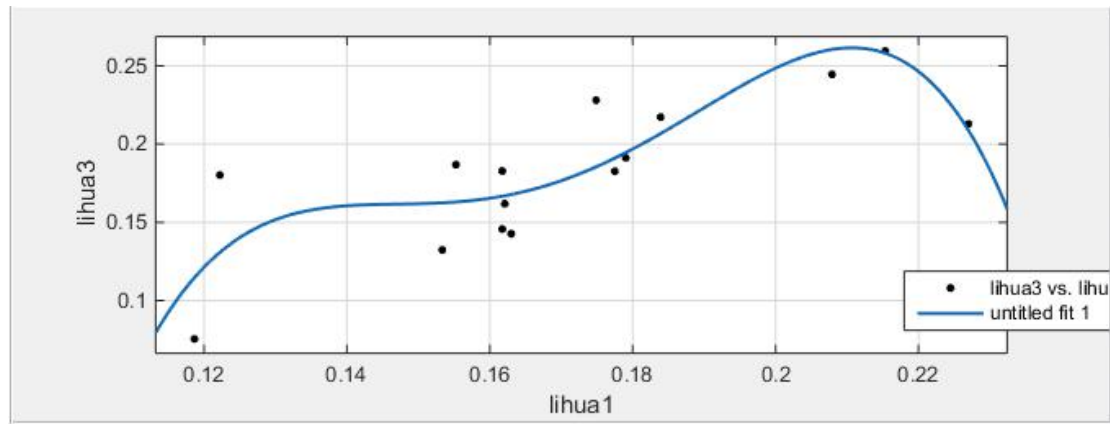
Goodness of fit:

SSE: 0.009087

R-square: 0.7108

Adjusted R-square: 0.5951

RMSE: 0.03014



可以看出，拟合精度较为满意，而且呈现出正相关的趋势。这说明总磷和总氮的变化是共同收到外界因素影响的，例如：温度、投放饲料、捕捞等。

以上只是一个维度的，即不同种类的理化因子；还有一个维度没有讨论，即不同位置的理化因子。这个明天再搞吧。

## 不同位置的理化因子之间的关系

(1) 首先，以间隙水为横坐标，以底泥为纵坐标进行多项式拟合，发现进行一次到五次多项式拟合的时候精度都较低。而进行更高次数的多项式函数拟合的时候，六次多项式拟合的精度较五次多项式拟合的精度有较大的上升，而更高次数的多项式拟合的精度较六次多项式拟合的精度没有较大的提高。因此，我们选定六次多项式拟合为较好的多项式拟合（对于此模型来说）。

六次多项式拟合的数据及图像如下：

Linear model Poly6:

$$f(x) = p1*x^6 + p2*x^5 + p3*x^4 + p4*x^3 + p5*x^2 + p6*x + p7$$

Coefficients (with 95% confidence bounds):

p1 =	2.235e+04	(-2396, 4.711e+04)
p2 =	3.42e+04	(-6250, 7.465e+04)
p3 =	2.061e+04	(-5590, 4.681e+04)
p4 =	6205	(-2336, 1.475e+04)
p5 =	972.3	(-493, 2438)

p6 = 74.1 (-50.07, 198.3)  
p7 = 1.782 (-2.238, 5.802)

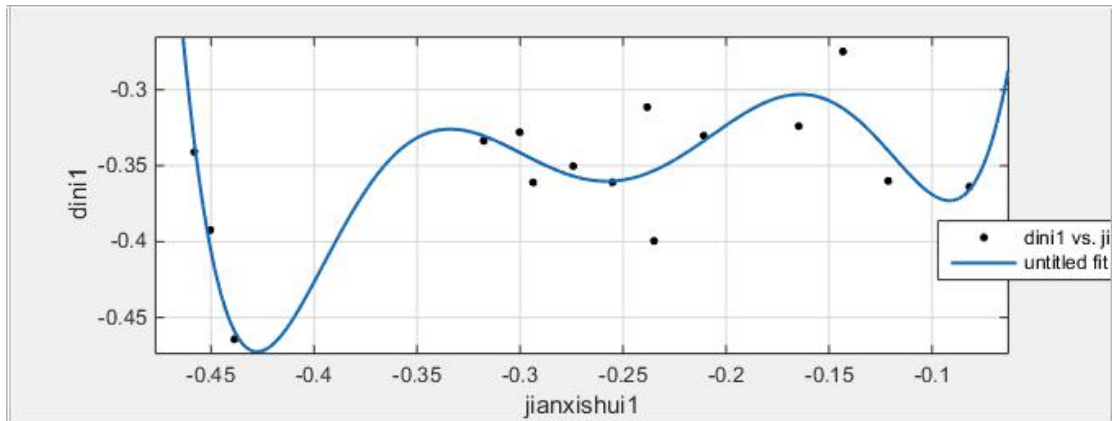
Goodness of fit:

SSE: 0.006974

R-square: 0.7393

Adjusted R-square: 0.5438

RMSE: 0.02953



可以发现，在此图像中，没有明显的趋势，但是却有明显的波峰波谷存在。为了进一步探讨波峰波谷的意义，我们进行了高斯拟合。经拟合发现，三峰高斯拟合的效果最为突出，而且拟合精度较多项式拟合也有提高。三峰高斯函数拟合的效果如下：

General model Gauss3:

$f(x) =$

$$a1 \cdot \exp(-((x-b1)/c1)^2) + a2 \cdot \exp(-((x-b2)/c2)^2) + a3 \cdot \exp(-((x-b3)/c3)^2)$$

Coefficients (with 95% confidence bounds):

a1 = -0.1454 (-0.4282, 0.1374)  
b1 = -0.4379 (-0.4965, -0.3794)  
c1 = 0.01535 (-0.06571, 0.09642)  
a2 = 0.1217 (-2.781, 3.025)  
b2 = -0.1508 (-0.2635, -0.0381)  
c2 = 0.01232 (-0.2044, 0.2291)  
a3 = -0.3625 (-0.4937, -0.2312)  
b3 = -0.03042 (-3.038, 2.977)  
c3 = 1.145 (-8.942, 11.23)

Goodness of fit:

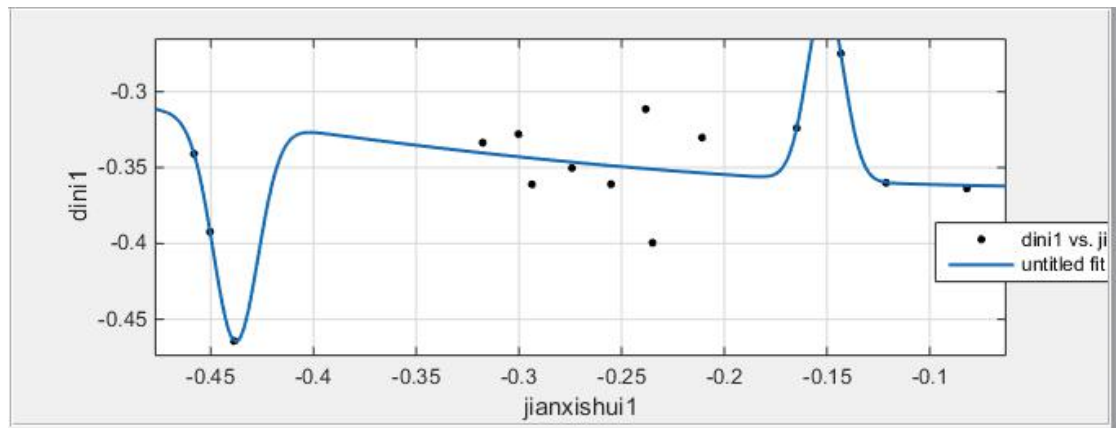
SSE: 0.005182

R-square: 0.8063

Adjusted R-square: 0.5481

RMSE: 0.02939





可以发现，在间隙水指数为-0.44 和-0.15 左右处函数曲线出现较为明显的波峰波谷。运用数据游标发现间隙水指数为-0.44 左右的位置大约对应着总氮指数 0.26，而间隙水指数为-0.15 左右的位置大约对应着总氮指数 0.18。

可以看出，间隙水指数和底泥指数整体来看呈现出较为明显的负相关关系。由此可以推测，间隙水中的理化因子多来自于底泥中。

## 底泥-池水

以底泥为纵坐标，以池水为横坐标利用多项式函数拟合发现：

当以六次多项式函数拟合的时候，拟合精度较高，且整体呈现出负相关关系。数据如下：

Linear model Poly6:

$$f(x) = p1*x^6 + p2*x^5 + p3*x^4 + p4*x^3 + p5*x^2 + p6*x + p7$$

Coefficients (with 95% confidence bounds):

p1 =	126.5	(-1786, 2039)
p2 =	-102.3	(-2772, 2567)
p3 =	23.89	(-1305, 1353)
p4 =	-5.654	(-288.6, 277.3)
p5 =	1.147	(-22.43, 24.73)
p6 =	0.0278	(-0.7667, 0.8222)
p7 =	-0.3457	(-0.4162, -0.2752)

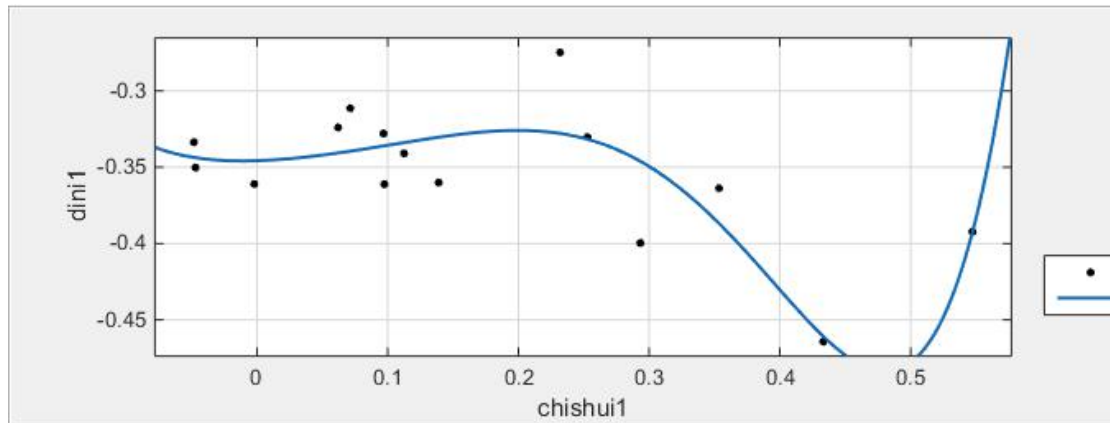
Goodness of fit:

SSE: 0.009288

R-square: 0.6529

Adjusted R-square: 0.3925

RMSE: 0.03407



## 池水-间隙水

以池水为纵坐标，以间隙水为横坐标进行拟合发现三峰高斯函数拟合的效果最为理想。数据与图像如下：

General model Gauss3:

$$f(x) = a_1 \exp(-((x-b_1)/c_1)^2) + a_2 \exp(-((x-b_2)/c_2)^2) + a_3 \exp(-((x-b_3)/c_3)^2)$$

Coefficients (with 95% confidence bounds):

a1 =	0.7259	(0.3561, 1.096)
b1 =	-0.4454	(-0.4474, -0.4434)
c1 =	0.009311	(0.004459, 0.01416)
a2 =	1.789e+13	(-7.05e+16, 7.054e+16)
b2 =	4.26	(-531.4, 539.9)
c2 =	0.7726	(-46.33, 47.87)
a3 =	5.401	(-42.23, 53.03)
b3 =	-0.2232	(-0.2244, -0.2219)
c3 =	0.006779	(-0.00275, 0.01631)

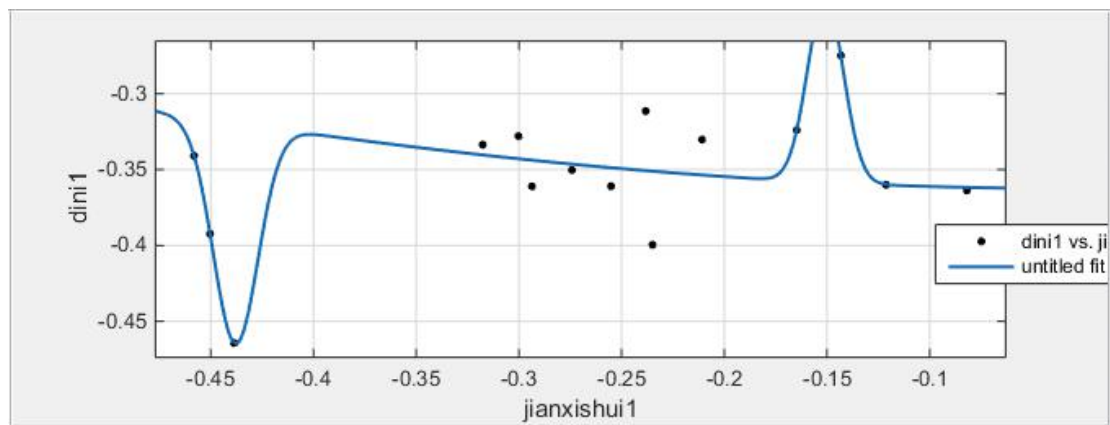
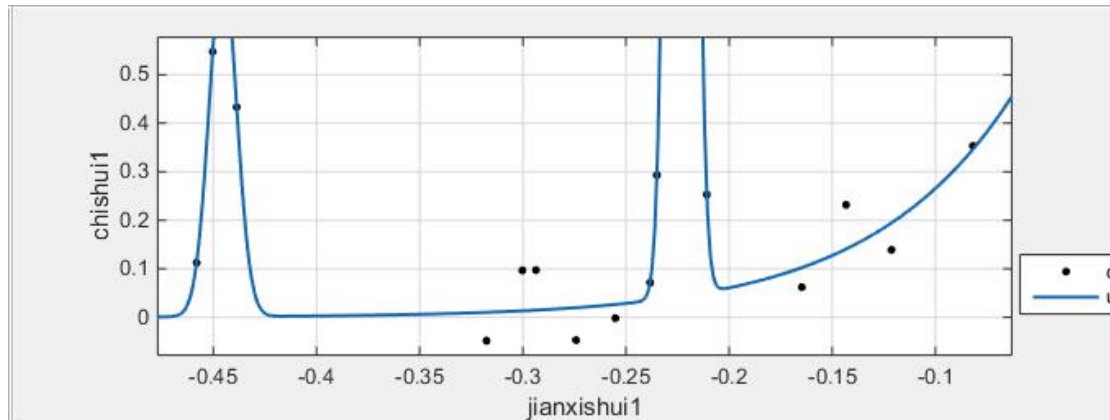
Goodness of fit:

SSE: 0.03558

R-square: 0.9176

Adjusted R-square: 0.8077

RMSE: 0.07701



发现在间隙水指数为-0.44 左右的位置和-0.23 左右的位置出现两个波峰。可能是捕捞和投放饲料导致的。

但整体呈现出正相关关系。

结合上面的分析结果：间隙水指数与底泥指数、池水指数与底泥指数呈现出良好的负相关关系，而间隙水指数与池水指数呈现出良好的正相关关系。由此我们可以推测，间隙水与池水中的理化因子来自于底泥。

根据附件 1 中的数据，我们发现，间隙水中的理化因子的浓度全部大于池水中的理化因子的浓度。由此我们推测，池水中的理化因子来自于间隙水。

至此，我们得出：池水中的理化因子来自于间隙水，间隙水中的理化因子来自于底泥。