长江水质的评价和预测

摘　要

本文首先运用主成分分析法对长江流域主要城市水质检测报告进行分析，选取主成分，并把主成分得分按方差贡献率加权求和，得出每个地区的污染综合评价指数，进而可以计算每个月长江流域的污染综合评价指数。

通过一维河流的稳态水质模型，确定干流上污染物的变化，计算出各地区主要污染物的排放质量,确定高锰酸盐指数（CODMn）的主要污染源在湖南岳阳、湖北宜昌、江苏南京、江西九江等地区；氨氮（NH3-N）的主要污染源在湖南岳阳、江西九江、湖北宜昌等地区。

然后利用 GM(1,1)模型与 BP 神经网络模型联合完成对未来十年不同水质的河长比例的预测，考虑到数据少，预测期长。如果使用神经网络模型进行预测效果很差，考虑GM(1,1)模型在很少的数据下可得到较高的预测精度，因此首先使用 GM(1,1)模型对未来十年的排污量进行预测，结果如下：单位 (亿吨)

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
排污量	289.9	306.3	323.6	341.9	361.2	381.7	403.2	426.0	450.1	475.5

再根据排污量预测值，利用 BP 神经网络对未来十年的不同水质的河长比例进行了预测。

为了得到排污量与各类水质的河长比例，本文再次利用 BP 神经网络的高精度逼近能力对排污量与六类水质的河长比例的关系进行拟合。从而可以得到每年控制污染所应当处理的废水量：单位（亿吨）

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
废水处理量	58.2	123.6	133.3	174.3	163.0	189.9	245.4	272.1	300.5	300.7

关键词：主成分分析　GM(1,1)模型　　BP 神经网络

问题重述

长江是我国第一、世界第三大河流，长江水质的污染程度日趋严重，已引起了相关政府部门和专家们的高度重视。为了保护长江水资源，必须对长江水质进行评价和预测进而采取措施来治理水质的污染，根据题意，本文要解决的问题有：

1. 对长江近两年多的水质情况做出定量的综合评价，并分析各地区水质的污染状况。
2. 研究、分析长江干流近一年多主要污染物高锰酸盐指数和氨氮的主要污染源的位置。
3. 在不采取有效治理措施的情况下，根据过去 10 年的主要统计数据，对长江未来水质污染的发展趋势做出预测分析。
4. 根据(3)预测分析，确定每年处理的污水量使长江干流的Ⅳ类和Ⅴ类水的比例控制在20%以内，且没有劣Ⅴ类水。
5. 对解决长江水质污染问题提出切实可行的建议和意见。

问题假设

1. 假设干流的自然净化能力是均匀的；
2. 假设两个观测站之间河段的平均流速是等于两个观测站流速的平均值；
3. 假设废水的处理对各类污染程度的河流的影响是均匀的。

符号说明

X_1	溶解氧的浓度（DO）
X_2	高锰酸盐指数（CODMn）
X_3	氨氮浓度（NH3－N）
X_4	PH 值
c	污染物的浓度
V	水流的流量
k	污染物的降解系数
v	水流的流速
x	污染物流过的距离
M_n	第 n 个观测站（地区）水流所含污染物的质量
m_n	第 n 个观测站（地区）排放污染物的质量
R_i	第 i 类污染程度的河流总长度比例
W(t)	第 t 年排污量

问题分析

江水的质量是由多个指标来进行测量评估的，为了使得建立的模型能够客观、准确地对长江水质做出全面的评价，要求：

第一、能够消除指标之间可能存在的相关性，以避免数据的重叠冗余。

第二、必须可以确定不同的指标对水质影响的权重。

有很多传统的系统评估方法比如加权评估法、专家评估法、综合评分法以及层次分析法都不免受到主观因素不同程度的影响。而本文使用的基于主成分分析所构造的评估机制则可以避免主观因素对评估的影响，使得评估结果客观的反映系统状况。

主成分分析方法是一种将多维因子纳入同一系统进行定量化研究、理论成熟的多元统计分析方法。通过分析变量之间的相关性，使得所反映信息重叠的变量
被某一主成分替代，减少了变量数目，从而降低了系统评价的复杂性。再以方差贡献率作为每个主成分的权重，由每个主成分的得分加权即可完成对水质的综合评价。

为了确定主要污染物高锰酸盐指数（CODMn）和氨氮（NH3-N）的主要污染源,我们需要知道各个地区主要污染物的排放质量。而本地区污染物的排放质量可以通过当前观测站的污染物质量与上游对本地区影响部分质量的差值来确定。通过污染物的降解公式分析出上游对本地区影响部分质量变化关系，进而得出本地区污染物排放的质量关系式。根据长江干流近一年多的基本数据计算出各地区污染物的平均排放速度，进而确定主要污染源。

长江水质被分为六个级别，代表了不同程度的污染，不同水质河长的比例可以表征一定时期内的水污染状况。所以说预测长江未来十年的水污染趋势，就是要预测未来不同水质的河长的比例。对每年的排污量与不同水质河长的比例做一个相关性分析：

	第 I 类	第 II 类	第 III 类	第 IV 类	第 V 类	劣 V 类
相关系数	-0.8058	0.3164	-0.3371	0.3183	0.6624	0.9570

可见排污量与不同水质河长的比例有很高的相关性，与劣 V 类的相关系数更是达到了 0.9570 的水平，因此在作对不同水质河长的比例之前，必须先对未来的排污量有比较精确的预测。

由于附件中数据样本少,需要预测的时间长，直接应用神经网络很难取得理想的效果，因此本文采用 GM(1,1)模型与神经网络模型联合预测长江未来十年的水污染趋势，尝试着首先较精确预测出部分重要的数据，为建立神经网络预测未来不同水质的河长的比例提供更多的数据，从而完成对不同水质河长的比例的预测。GM(1,1)模型就可以用来较好的预测出未来的排污量。

基于主成分分析的水质综合评价机制

根据主成分分析的方法，分析长江沿线 17 个观测站（地区）近两年多主要的四种水质指标的检测数据。步骤如下：

Step1：为了消除不同变量的量纲的影响，首先需要对变量进行标准化，设检测数据中水质样本共有 n 个，指标共有 p 个，分别设为 $X_1, X_2, X_3 \cdots X_p$，令 x_{ij} (i=1…n; j=1…p) 为第 i 样本的第 j 个指标的值。做变换

$$Y_j = \frac{X_j - E(X_j)}{\sqrt{Var(X_j)}} \quad (\text{j}=1,2,3\cdots\text{p})$$

得到标准化的数据矩阵 $y_{ij} = \frac{x_{ij} - \overline{x_j}}{s_j}$ ，其中 $\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$， $s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2$。

Step2：在标准化数据矩阵 $Y = (y_{ij})_{n\times p}$ 的基础上计算 p 个原始指标相关系数矩阵 $R = (r_{ij})_{p\times p}$，其中：

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})(x_{kj}-\overline{x_j})}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})^2}\sqrt{\sum_{k=1}^{n}(x_{kj}-\overline{x_j})^2}} \qquad (i=1\cdots n;\ j=1\cdots p)$$

Step3：求相关系数矩阵 R 的特征值并排序 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_p$，再求出 R 的特征值的相应的正则化单位特征向量 $l_i=(l_{1i},l_{2i},\cdots l_{pi})$，则第 i 个主成分表示为各个指标 X_k 的线性组合 $Z_i=\sum_{k=1}^{p}l_{ki}X_k$。

Step4：确定主成分数目。在确定主成分数目前，需要先给出一个控制值 α，令 $\sum_{i=1}^{q}\lambda_i/\sum_{i=1}^{p}\lambda_i\geq 1-\alpha$，则对应满足条件的 q 的最小值即为保留的主成分的个数 m，本文中取 $\alpha=5\%$。

Step5：计算综合得分。首先计算得到第 i 个样本中第 k 个主成分的得分为 $F_{ik}=\sum_{j=1}^{p}l_{jk}X_j$，再以 m 个主成分的方差贡献率为权重，求得第 i 个样本的综合得分。

$f_i=\sum_{k=1}^{m}F_{ik}\times\lambda_k$ （i=1…n）

Step6：根据每个样本的综合得分进行排序。

附件中共有 28 个月的数据，这里仅随机选择 2005 年 4 月的数据来说明利用主成分分析进行水质综合评价的过程(同理可进行其他月份的数据分析)。

调用 MATLAB 统计工具箱 princomp 函数,格式为:

[pc,score,latent,tsquare]=princomp(ingredients)

其中 ingredients 指标准化后的样本指标矩阵，pc 是指各主成分关于指标的线性组合的系数矩阵,score 为各主成分得分，latent 是方差矩阵的特征值,tsquare 为 Hotelling T^2 统计量。

统计结果如下：

四种指标的相关系数矩阵为：

$$\begin{vmatrix} 1.0000 & -0.2294 & -0.1350 & -0.2758 \\ -0.2294 & 1.0000 & 0.5324 & 0.5383 \\ -0.1350 & 0.5324 & 1.0000 & 0.4074 \\ -0.2758 & 0.5383 & 0.4074 & 1.0000 \end{vmatrix}$$

各个主成分的贡献率：

表 1 主成分的贡献率表

	特征值	贡献率	累积贡献率
第一主成分	4.3121	64.84%	64.84%
第二主成分	1.1739	17.65%	82.49%
第三主成分	0.9431	14.18%	96.67%
第四主成分	0.2214	3.3%	100%

由表可看出，前三个主成分的累积贡献率已达到 96.67%,取控制参数α=0.06(因为28个月中前三个成分贡献率最低为94%),因此取前三个主成分对各地区水质进行综合评价。

根据R的特征值的相应的正则化单位特征向量，前三个主成分关于指标的线性组合为

$$y_1 = -0.0660x_1 + 0.6715x_2 + 0.5181x_3 + 0.5257x_4$$
$$y_2 = 0.0715x_1 - 0.0070x_2 + 0.7193x_3 - 0.6910x_4$$
$$y_3 = -0.0076x_1 - 0.7399x_2 + 0.4626x_3 + 0.4883x_4$$

根据线性表达式中的系数及符号，可对各主成分的实际意义作如下解释：第一主成分为除x_1之外的三项指标的综合；第二主成分与x_3成正相关，与x_4成负相关；第三主成分为除x_1之外的三项指标的综合。以各个主成分的方差贡献率为权重可得到水质的最终综合评价：

表2　综合评价表

	第一主成分	第二主成分	第三主成分	综合得分	综合排名
四川攀枝花	-1.4696	-0.68454	-0.91026	-1.2028	14
重庆朱沱	-1.7625	-0.94467	0.032861	-1.3049	15
湖北宜昌南津关	-0.33083	-0.31954	-1.2501	-0.44818	10
湖南岳阳城陵矶	2.0022	1.8388	-0.30098	1.5801	3
江西九江河西水厂	-1.4067	-0.02552	0.53935	-0.84008	13
安徽安庆皖河口	0.17222	0.14241	-1.0413	-0.01085	7
江苏南京林山	-1.2741	0.10172	0.43579	-0.74634	12
四川乐山岷江大桥	3.7118	0.6009	-1.3863	2.3162	2
四川宜宾凉姜沟	-2.2066	-1.0968	-0.0973	-1.6381	17
四川泸州沱江二桥	0.33637	0.27026	-1.1778	0.09879	6
湖北丹江口胡家岭	-2.0179	-0.26008	0.33508	-1.3068	16
湖南长沙新港	-0.46208	-0.72673	-0.06801	-0.43753	9
湖南岳阳岳阳楼	0.44676	1.8036	2.1496	0.91285	4
湖北武汉宗关	0.9091	1.4033	-0.13198	0.81843	5
江西南昌滁槎	5.447	-2.5132	1.3095	3.2738	1
江西九江蛤蟆石	-0.90294	0.36553	0.90899	-0.39204	8
江苏扬州三江营	-1.1922	0.044725	0.6528	-0.67255	11

上表给出了2005年4月17个地区的污染程度的综合评价，综合得分越高就说明污染越严重，排名越靠前，污染越严重。表中显示2005年4月江西，湖南等地的污染比较严重，且干流污染程度要小于支流的污染程度。

分别取28个月17个地区的综合得分总和作为当月长江流域的水质污染指数，其随时间变化如图1所示：

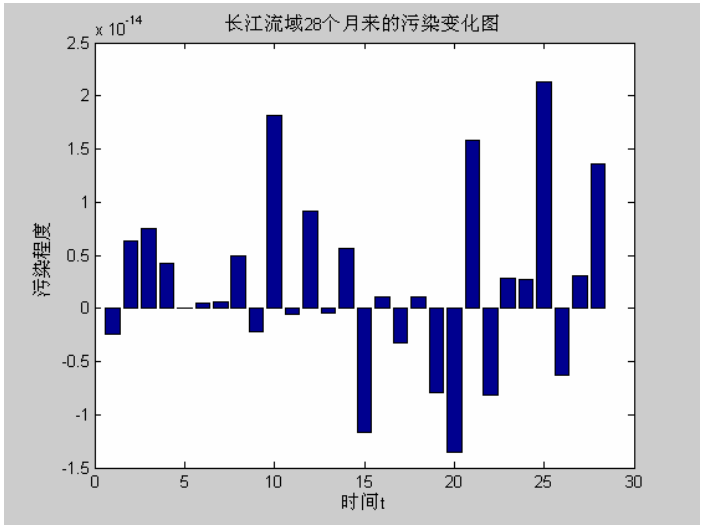

图 1 长江流域 28 个月的污染变化图

上图表明长江流域的水质变化存在一定的周期,污染状况两年来成波动恶化趋势,且在 2004 年 8 月~2005 年一月水的污染状况曾经有所缓解。从上图中还可以看出两年来长江流域的水质状况有三次集中的恶化表现，并且一次比一次更加严重，说明长江流域的水质正在面临不断加剧的污染。

分别取 17 个地区在 28 个月里综合得分的平均值作为该地两年来的污染指数。各地区污染情况如图 2 所示：

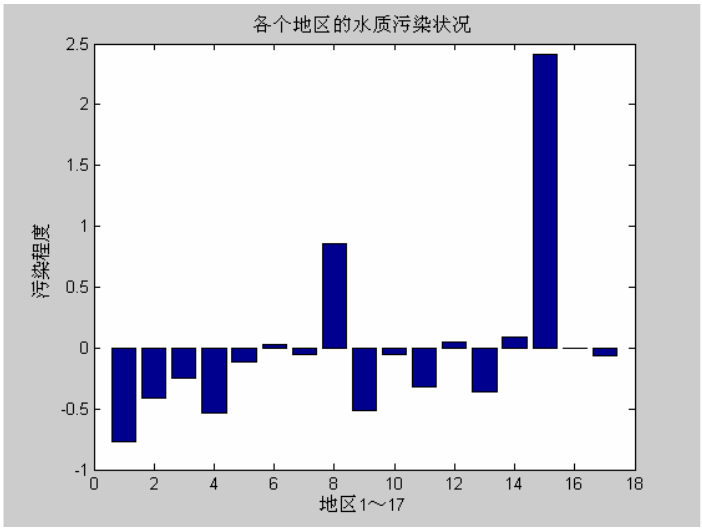

图 2 不同地区水质污染状况

综合排名依次为（按照表 2 的地区由上往下）：

17 14 11 16 10 5 7 2 15 8 12 4 13 3 1 6 9

仍然说明江西地区,湖南地区等地的污染十分严重,支流的污染程度普遍大于干流。

主要污染物的污染源分析

通过分析长江干流近一年多的基本数据来计算出各地区主要污染物的排放质量,即可确定主要污染物高锰酸盐指数（CODMn）和氨氮（NH3-N）的主要污染源。

观测站（地区）的水质污染的来源，主要来自本地区的排污部分和上游的污水。那

么，水流通过该观测站（第 n 个）时每秒所含污染物的质量为：

$$M_n = M^*_{n-1} + m_n$$

其中 m_n 为本地区的排污部分，M^*_{n-1} 为来自上游观测站部分。

为了确定 M^*_{n-1}，即来自上游的污水在流动过程中质量的变化情况,我们对各情况进行以下分析：

河流对污染物有净化能力，那么，水流从上游到下游所含的污染物会发生一定的变化，当流量不变时，由一维河流的稳态水质模型[5]，在下游 x 处污染物的浓度：

$$c(x) = c \cdot e^{-k \cdot t(x)} = c \cdot e^{-k \cdot x / v}$$

其中 c 为上游起点的污染物浓度，k 为污染物的降解系数，v 为该河段的平均流速。

由质量与浓度、体积的关系式 $m = c \cdot V$，可得下游 x 处污染物的质量：

$$m(x) = c(x) \cdot V = c \cdot V \cdot e^{-k \cdot x / v} = m \cdot e^{-k \cdot x / v}$$

考虑到流量变化的情况，我们对流量变化的水流进行分析。设在某一河段，水流的增量为 ΔV，水流的污染物原浓度为 c，原有水量为 V。

那么水流的浓度变化为：

$$c^* = \frac{c \cdot V}{V + \Delta V}$$

那么在下游 x 处污染物的浓度：

$$c^*(x) = c^* \cdot e^{-k \cdot t(x)} = \frac{c \cdot V}{V + \Delta V} \cdot e^{-k \cdot x / v}$$

那么污染物的质量变化为：

$$m^*(x) = c^*(x) \cdot (V + \Delta V) = c \cdot V \cdot e^{-k \cdot x / v} = m(x)$$

可以看出下游 x 处污染物的质量只与位置有关。

所以上游的观测站污染物到达下游相邻的观测站后的质量变为：

$$M^*_{n-1} = M_{n-1} \cdot e^{-k \cdot x / v}$$

其中 x 为两个观测站的距离，v 为该河段的流速。

同时，观测站污染物的质量也可以通过该观测站的水流量与污染物的浓度来确定：

$$M_n = c_n \cdot V_n$$

其中 c_n 为观测站的污染物浓度，V_n 为观测站的水流量。

简化上面公式，可得本地区的污染总量的计算公式：

$$m_n = M_n - M^*_{n-1} = M_n - M_{n-1} \cdot e^{-k \cdot x / v_n} = c_n \cdot V_n - c_{n-1} \cdot V_{n-1} \cdot e^{-k \cdot x / v_n} \quad (n=1,\cdots,7)$$

其中 v_n 为两个观测站的平均速度。

由相关参考资料[6]，我们确定长江对 CODMn 的降解系数为 0.28（单位：1/天），对 NH3-N 的降解系数为 0.43（单位：1/天），根据干流上 7 个观测站近一年多的基本数据解出各个月污染物排放质量的数据，从而得出这一年多时间内，各地区排放到长江的 CODMn 的平均含量，如表 3：

表 3　CODMn 的平均排放量表　　单位：　kg/s

	四川攀枝花	重庆朱沱	湖北宜昌	湖南岳阳	江西九江	安徽安庆	江苏南京
平均含量	8.986	33.553	39.434	51.824	37.878	21.482	38.124

各地区排放到长江的 NH3-N 的平均含量，如表 4：

表 4　NH3-N 的平均排放量表　　单位：　kg/s

	四川攀枝花	重庆朱沱	湖北宜昌	湖南岳阳	江西九江	安徽安庆	江苏南京
平均含量	0.4816	2.9507	3.8099	5.5019	4.0810	2.1791	1.3943

观察表 3 中的数据可知，高锰酸盐指数（CODMn）的主要污染源在湖南岳阳、湖北宜昌、江苏南京、江西九江等地区。

观察表 4 中的数据可知，氨氮（NH3-N）的主要污染源在湖南岳阳、江西九江、湖北宜昌等地区。

预测长江未来十年的水污染趋势

长江的水质问题是一个复杂的非线性系统,但是由于数据样本少,需要预测的时间长，直接应用神经网络很难取得理想的效果，因此本文采用 GM(1,1)模型与神经网络模型联合预测长江未来十年的水污染趋势，尝试着首先较精确预测出部分重要的数据，为建立神经网络预测未来不同水质的河长的比例提供更多的数据，GM(1,1)模型就可以用来较好的预测出未来的排污量。

1．模型 1　GM(1,1)模型：

考虑到污水排放量的变化规律是一个不确定的系统，且本题给出污水排放量数据样本比较少，还要求做出长达十年的预测，因此采用灰色预测方法来预测未来的污水排放量

灰色预测方法就是根据系统的普遍发展规律，建立一般的灰色微分方程，然后通过对数据序列的拟合，求得微分方程的系数，从而获得灰色预测模型方程。

利用灰色预测理论建立GM(1,1)模型,记1995年为第一年,第k年的排污量为 $\hat{X}^{(0)}(k)$，其中 $(k=1,2,\cdots n)$，对10个历史数据进行模拟并对未来的排污量进行预测，利用该数据列建立预测模型的步骤如下：

Step1:作一阶累加，形成生成数据序列

$X^{(1)}(k)=\sum_{m=1}^{k}X^{(0)}(m)$，$(k=1,2,\cdots 10)$

则相应的灰微分方程：

$\frac{dX^{(1)}}{dt}+aX^{(1)}=u$，此方程即为GM(1,1)的数值模型，式中a,u为待定系数，其中a为发展灰数，u为内生控制灰数。

Step2:求参数a和u

对微分方程进行离散化得关于a和u的超定方程组：

$X^{(1)}(k+1)=(1-a)X^{(1)}(k)-u$，$(k=2,3,\cdots 10)$

利用最小二乘法求超定方程得：

$[a,u]^T=(B^TB)^{-1}B^TY_n$，

$$\text{其中 } B=\begin{bmatrix} -\frac{1}{2}[X^{(1)}(1)+X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2)+X^{(1)}(3)] & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ -\frac{1}{2}[X^{(1)}(9)+X^{(1)}(10)] & 1 \end{bmatrix} Y_n=\begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \cdot \\ \cdot \\ \cdot \\ X^{(0)}(10) \end{bmatrix}$$

Step3:建立生成数据序列模型

将上面求得的参数代入上述的灰微分方程，求解微分方程得到GM(1,1)的灰色预测模型为：

$$\hat{X}^{(1)}(k+1)=\left[X^{(0)}-\frac{u}{a}\right]e^{-ak}+\frac{u}{a}\quad,\quad (k=1,2,3\cdots n)$$

Step4:建立原始数据序列模型，即由累减生成原始数据序列 $X^{(0)}$ 的模拟序列值：

$$\hat{X}^{(0)}(1)=X^{(0)}(1)$$

$$\hat{X}^{(0)}(k)=\hat{X}^{(1)}(k)-\hat{X}^{(1)}(k-1)$$

$$=(1-e^{a})\left(X^{(0)}(1)-\frac{u}{a}\right)e^{-a(k-1)}\qquad (k=2,3\cdots n)$$

这里 $\hat{X}^{(0)}(k)\ \ (k=1,2,\cdots 10)$ 是原始排污量数据序列 $\hat{X}^{(0)}(k)\ \ (k=1,2,\cdots 10)$ 的拟合值，$\hat{X}^{(0)}(k)\ \ (k>10)$ 是原始排污量数据序列的预测值。

根据上述的方法用MATLAB软件求得参数a=-0.0550,u=162.4114再把参数代回微分方程得到排污量的灰色预测模型为：

$$\hat{X}^{(1)}(k+1)=3119.5e^{0.0550k}-2952.9\quad (k=1,2,\cdots n)$$

对10个原始排污量数据的模拟模型为：

$$\hat{X}^{(0)}(1)=X^{(0)}(1)=174\text{（亿吨）}$$

$$\hat{X}^{(0)}(k)=166.9396e^{-0.0550(k-1)}\qquad (k=2,3,\cdots 10)$$

根据得出来的模型可得下面的两图：

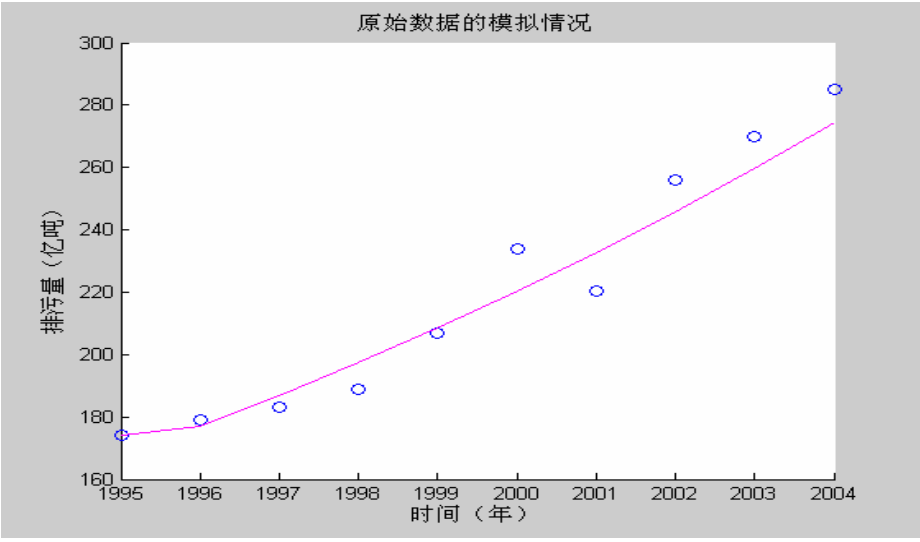

图3　原始数据与模拟曲线对照图

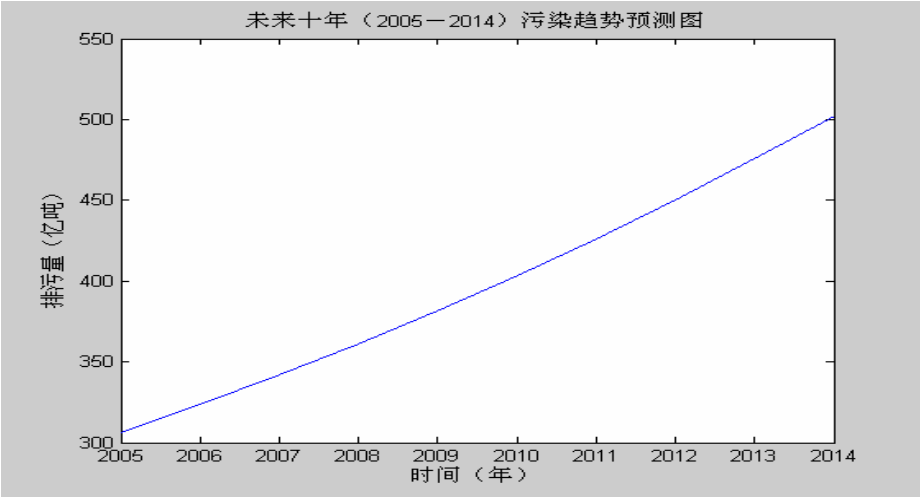

图4　未来十年排污量变化图

2．模型的检验

根据上面的检验步骤计算得：

表5　误差检验表

平均相对误差 $\overline{\Delta}$	关联度 r	均方差比值C	小误差概率P
0.0313	0.9815	0.2202	1

表6　常用的精度等级表

等级	平均相对误差 $\overline{\Delta}$	关联度 r	均方差比值C	小误差概率P
一级	0.01	0.90	0.35	0.95
二级	0.05	0.80	0.5	0.80
三级	0.10	0.70	0.65	0.70
四级	0.20	0.60	0.80	0.60

把误差检验表跟常用的精度等级表对比可知，模型的等级接近一级，也即是说，该模型的拟合精度很高，可用来预测。

3．模型 2　BP 神经网络预测模型

附件中根据污染程度不同把水质状况分为六类，可以分别针对各类水质状况的河流长度比例在未来十年的变化进行预测。得到未来六类不同水质河长比例的变化，从而可以全面显示未来十年污染趋势的变化

针对第 i 类污染程度的河流长度比例进行分析，首先选择输入数据，不同水质河长的比例必然同长江流域内的排污量有关，而未来十年的排污量已经由灰色模型预测得到。另外根据对附件中数据的分析，长江的污染程度表现出某种周期性的波动，可以预测不同水质河长的比例应有关于时间上的规律，因此输入数据中可以用待预测当年前三年的数据来显示这种波动。从而建立了四个输入变量一个输出变量的三层神经网络，隐层选择目前并没有可靠的成熟理论.可根据数据的复杂度尝试不同的隐层节点数目。本文中选择的网络拓扑结构图，如图 5

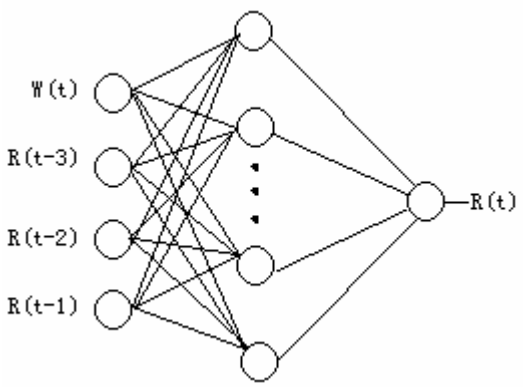

图 5　BP 神经网络预测模型拓扑结构图

4．网络训练与预测

进行网络训练前需要构建初始化网络，输入层网络激发函数选取对数 S 型传递函数 ，输出层选择线性函数，调用 MATLAB 神经网络工具箱函数：

net=newff(minmax(p),[10,1],{'logsig','purelin'},'trainlm')

即可建立一个 4——10——1 网络结构的网络。

由附件中的数据可以得到样本，利用 Levenberg_Marquardt 的 BP 算法训练函数，预测的算法流程如下：

Step1：对初始数据进行标准化。

Step2：利用原始数据对网络进行训练。

Step3：对未来第 t 年第 i 类污染程度的河流长度比例进行预测。

Step4：利用第 t 年预测得到的数据作为样本再对网络进行训练。

Step5：然后令 t=t+1，回到 Step2，直到 t=10。

依次对水文年全流域六类污染程度的河流长度比例重复进行十次预测，取平均值得到二十年六类污染程度的河流长度比例：

表 7　预测的六类污染程度的河流长度比例

	Ⅰ类	Ⅱ类	Ⅱ类	Ⅳ类	Ⅴ类	劣Ⅴ类
1995	24.7	35.7	30	2.9	6.7	0
1996	25.6	29.5	44.1	0	0.8	0
1997	14.6	27.6	44.5	13.3	0	0
1998	10.3	20.1	69.6	0	0	0
1999	0	56.4	30.8	5.5	7.3	0
2000	9.5	35.9	29.1	25.4	0	0
2001	2.3	30.1	35.3	18.7	7.8	5.8
2002	3.1	35.4	30.3	17.4	5.1	8.7
2003	8	17.8	68	1.5	4.6	0
2004	1.1	25.8	40.6	15.7	7.8	9
2005	6.4198	33.988	46.976	17.357	6.0823	17.521
2006	3.5728	35.978	43.328	10.99	6.1538	4.9205
2007	4.1559	33.508	48.264	10.997	6.6304	6.0133
2008	4.7921	33.658	45.881	11.487	6.8356	12.455
2009	2.9995	31.843	47.03	10.07	5.6425	8.6524
2010	4.6285	31.841	45.635	8.8012	7.0424	6.061

2011	4.0217	31.152	48.441	11.388	6.3627	12.09
2012	3.6442	31.843	44.942	10.414	6.7455	7.1099
2013	4.9101	35.932	44.834	8.5241	6.0965	8.7023
2014	3.4146	29.804	46.328	10.506	6.6459	13.684

5.求解问题 4

设 ： $W(t)=f(R_1(t),R_2(t),R_3(t),R_4(t),R_5(t),R_6(t))$，三层神经网络可以以任意精度逼近一个非线性函数，因此考虑建立 BP 神经网络来完成六类污染程度的河流长度比例同排污量的拟合。

设计按下列结构对网络进行训练：

表 8　网络结构

网络基本结构	输入激发函数	输出激发函数	训练函数	精度
6—13—1	对数 S 型函数	线性函数	Levenberg_Marquardt 算法	0.05

以模型 1 预测得到未来十年的排污量，通过模型 2 预测水文年干流未来十年的六类污染的河流长度比例，共有 20 组训练样本。网络经过训练后，即可按下面算法对问题 4 进行求解：

首先对未来十年的各类污染程度的河流长度比例数据进行调整:

如果 IV 类和 V 类污染程度的河流长度比例 $R_4(t)+R_5(t)$ 大于 20%

则令 $R_4(t)=R_4(t)-\dfrac{R_4(t)+R_5(t)-20\%}{2}$

$R_5(t)=R_5(t)-\dfrac{R_4(t)+R_5(t)-20\%}{2}$ ，使得 $R_4(t)+R_5(t)=20\%$

如果 $R_6(t)>0$, 则令 $R_6(t)=0$;

对调整后的数据进行模拟:

将调整后的数据输入网络进行模拟，得到新的输出值记为：$W_1(t)$

则 $\Delta W=W(t)-W_1(t)$ 即为第 t 年需要处理的污水量。由他们的相关系数可以证明，排污量同第 IV 类, 第 V 类, 劣 V 类均呈正相关，故 $\Delta W>0$

最终结果见表 9：

表 9 2005-2014 年需处理的污水量

年份	2005	2006	2007	2008	2009
排污量	289.9396	306.3231	323.6323	341.9197	361.2404
应处理废水量	58.206	123．578	133.331	174.314	163.025
年份	2010	2011	2012	2013	2014
排污量	381.6528	403.2187	426.0031	450.0751	475.5072
应处理废水量	189.860	245.429	272.145	300.524	300.669

7．灰色预测模型的评价及进一步改进

灰色预测模型具有如下主要特点：（1）少数据性（2）良好的时效性（3）较强的系统性和关联性，它将研究对象作为一个发展变化的系统，可对事物发展态势进行量化比较分析，其动态过程能反映系统已知信息和未知信息互相影响、互相制约的系统特征，并能揭示系统内涵的本质联系。（4）建模精度较高，可保持原系统的特征，能较好地

反映系统的实际状况。根据不同的预报等级和容许误差值，选用不同的模型，既可做长期趋势预报分析，也可做中、短期预测。

BP神经网络的特点有：(1)良好的逼近能力 (2)误差可以控制

灰色预测方法求解需要的计算量小，在少量样本的情况下可达到较高的精度；而神经网络计算精度高，且误差可控；这样二者优势互补，使预测结果更加符合实际。

建立 GM(1,1)模型后，可得到对原始数据数列的预测值，这些预测值与原始数据有一定的偏差，为了尽可能降低预测偏差，考虑将预测值与实际值之间的偏差关系综合到神经网络模型中：将灰色模型的预测值作为神经网络的输入样本，世纪之作为神经网络的输出样本，采取一定的网络结构，然后对神经网络训练，就可得到一系列对应于相应节点的权值和阈值。网络的结构固定之后，将 GM(1,1)下一个时刻或几个时刻的预测值作为训练好的神经网络的输入，得到的输出即为对应下一时刻或下几个时刻最终的预测值。这种结合算法即能尽可能的减少预测值与原始数据之间的偏差。

解决长江水质污染问题的建议和意见

近年来，长江的水污染日益严重，解决长江水质污染的问题已是刻不容缓的事了。解决水质问题关键在于治本：首先必须确定主要污染物；然后，分析主要污染物的主要污染源；最后，对主要的污染地区进行防治。

通过上面对长江主要污染物高锰酸盐指数（CODMn）和氨氮（NH3-N）主要污染源的分析，我们得出 7 个观测地区的平均排放量见表 1，表 2。由此可以知道各地区的主要污染物的平均排放量，从而确定主要污染物的主要污染源：高锰酸盐指数（CODMn）的主要污染源在湖南岳阳、湖北宜昌、江苏南京、江西九江等地区；氨氮（NH3-N）的主要污染源在湖南岳阳、江西九江、湖北宜昌等地区。

我们要着力于对这些主要污染源所在地区的防治：

1. 提高主要污染地区在主要污染物方面的排放指标。如，湖南岳阳在 CODMn 和 NH3-N 的污染排放指标。
2. 对于该地区内排放主要污染物较多的工厂、企业，要进行严格的监督，保证能及时地处理异常问题。
3. 提高该地区对主要污染物的净化能力。比如，买进较先进的净化仪器。
4. 加大对废水的处理量，可以使重污染河长的比例减小。

参考资料：

[1] 高惠璇，应用多元统计分析，北京，北京大学出版社，265~290 页，2005

[2] 王沫然，Matlab 与科学计算，北京，电子工业出版社，2003[

[3] 李涛，贺勇军等，Matlab 工具箱应用指南，北京，电子工业出版社，75~80 页,2000

[4] 耿继进，灰色预测理论若干问题研究，武汉测绘科技大学学报，第 19 卷第 1 期，第 57 页，1994

[5] 傅国伟，河流水质数学模型及其模拟计算，北京，中国环境科学出版社，第 88 页 1987

[6] 高锰酸盐指数和氨氮的降解系数
http://www.caep.org.cn/bbs/showthread.asp?threadid=342