

指数平滑法 - 马尔科夫模型在巢湖 水质预测中的应用

荣洁, 王腊春

(南京大学 地理与海洋科学学院, 南京, 210046)

摘要: 水质预测是水环境管理的重要内容,对工农业生产具有较好的指导意义。以巢湖为例,针对水质变化的随机性,应用指数平滑法 - 马尔科夫预测法,对合肥湖滨与巢湖裕溪口两大断面 2001 - 2010 年的 COD_{Mn} 、TP、TN 数据进行指数平滑处理,预测 2011 - 2013 年水质,发现未来几年需要注意 TN 以及东半湖 COD_{Mn} 的变化,这对宏观上把握水质变化趋势以及污染因子的动态有一定的指导意义,同时也发现指数平滑与马尔科夫预测的结果是一致的,将两者结合,可较好的运用于水质预测中。

关键词: 指数平滑法; 马尔科夫预测; 水质预测; 巢湖

中图分类号: X824 文献标识码: A 文章编号: 1672-643X(2013)04-0098-05

Application of the exponential smoothing law - Markov model in prediction of water quality of Chaohu lake

RONG Jie, WANG Lachun

(School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210046, China)

Abstract: Prediction of water quality is an important method of water environment management and has good guidance significance to industrial and agricultural production. Taking Chaohu Lake for example, aiming at the randomness of the changes in water quality, applying exponential smoothing method - Markov prediction, processing the data of COD_{Mn} , TP, TN of the two cross-section of the Hefei Lakeside and Chaohu Yuxikou during 2001 - 2010 by exponential smoothing, the paper forecasted the water quality from 2011 to 2013, and found out that the changes of TN as well as COD_{Mn} of the eastern half of Lake need to be paid attention to in the next few years. Which has certain guiding significance to grasp water quality change and pollution factor trend. Also found out that the forecast results of exponential smoothing method keep consistent with that of the Markov's prediction, the combination of the two can be better applied to the prediction of water quality.

Key words: exponential smoothing; Markov prediction; water quality; Chaohu lake

1 概述

在水环境日益严峻的情势下,水质变化趋势预测成为维护和管理当前水质状况的重要依据^[1],通过预测可以了解当地水域环境质量演变趋势,及时发现水质恶化原因,从而采取恰当的治理措施,减少水污染事件的发生率和社会财富的损失,为水污染控制及水资源长远规划与管理提供科学依据。

水质预测的方法有很多,归纳起来主要有:基于水流连续方程和扩散理论的物理模型预测法^[2],基于灰色系统理论的灰色预测法^[3],神经网络预测

法^[4],基于模糊理论的预测方法^[5],以及数理统计预测法^[6-7]。河流水质受多方面因素的影响,包括人为因素的和自然因素,所以水质状况好坏受多方面的影响,是一种随机过程。在外部因素随机的作用下,系统将从一个状态转移到另一个状态,并随时间的变化而变化,具有一定的随机性。现有的基于数学表达式的水质预测模型很难将这些因素都考虑进去,而水质预测的确定模型很难考虑所有影响因素,另外有些模型需要考虑到的参数很多,并且很难确定,运用这类模型对于不确定性就很难描述。这些都使得中、长期时间尺度的水质预测结果与实际

收稿日期:2013-04-08; 修回日期:2013-05-02

基金项目:国家水体污染控制与治理科技重大专项资助(2009ZX-07210-010)

作者简介:荣洁(1989-),女,江苏常熟人,硕士,主要从事水资源水环境方面的研究。

通讯作者:王腊春(1963-),男,江苏金坛人,博士,教授,博士生导师,主要从事水资源水环境方面的研究。

差异非常大,难以满足工作需要。

本文试图应用随机过程的有关理论,结合指数平滑法将马尔科夫模型运用到水质预测中。考虑水质序列的随机性和连续性,首先对水质序列进行时间序列分析,采用指数平滑法处理,作为一个新的序列,通过对水质数据的指数平滑法处理,降低序列的随机性。将水质按照地表水环境质量标准划分成5类,建立马尔科夫链(MC)预测模型,并把水质从一种类别转变为另一种类别的过程看作是无后效性的随机过程,对河流综合水质的预报作初步探讨,找出不同级别的水质间相互转化的概率,从宏观上把握水质的变化趋势,以便为水质预报的科学化提供依据。

2 指数平滑-马尔科夫模型

2.1 指数平滑法

指数平滑法的是特殊的加权移动平均法,加强了近期观测值对预测值的作用。由近至远的权数之间按等比级数递减,等比级数首项是平滑系数 α ,公比为 $(1-\alpha)$ 。平滑系数取值的大小决定了权数变化的快慢, α 的选择,直接影响过去数据对预测值的作用,如果时间序列较平稳,数据波动较小, α 取值则较小,取值范围在0.1~0.3之间,数据如果波动较大则取值在0.3~0.5之间,数据如果波动很大并且趋势比较明显,则在0.6~0.8之间^[8]。一次指数平滑法的计算公式为:

$$\begin{aligned} \hat{y}_{t+1} &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \\ &\quad \cdots + \alpha(1-\alpha)^{n-1} y_{t-n+1} \\ &= \sum_{j=0}^{n-1} \alpha(1-\alpha)^j y_{t-j} \\ &= \alpha y_t + (1-\alpha)\hat{y}_t \end{aligned} \quad (1)$$

式中: \hat{y}_{t-1} 、 \hat{y}_t 为预测值; y 为实际值。

2.2 马尔科夫预测

马尔科夫过程是研究某一事件的状态及状态之间转移规律理论的随机过程。其通过对 t_0 时刻事件不同状态的初始概率及状态间概率转移关系来研究 $t_0 + \Delta t$ 时刻状态的变化趋势^[9]。马尔科夫过程是具有无后效性的随机过程。即在事件的发展过程中,系统的第 n 次结果状态只与第 $n-1$ 次有关,与以前所处的状态无关。马尔科夫预测步骤:

(1) 计算初始概率:假定事件共有 N 个状态,在所有状态中 E_i 出现次数为 M_i , $N = \sum_{i=1}^n M_i (i \in R)$,则 E_i 出现频率为 $F_i = M_i/N$,样本容量足够大时,可用样本的分布情况 F_i 近似地描述理论分布 P_i , $P_i \approx$

$F_i = M_i/N$,初始概率分布总和为1。 $\sum_{i=1}^n P_i = 1$ 。

计算状态的一重转移概率 P_{ij} : $P_{ij} = (E_i \rightarrow E_j) = P(E_j/E_i) \approx F(E_j/E_i)$, $F(E_j/E_i)$ 为由 E_i 状态转向 E_j 状态的频率,若已知 E_i 状态转向 E_j 状态的个数为 M_{ij} ,则 $P(E_j/E_i) \approx F(E_j/E_i) = M_{ij}/M_i$ 状态为 i 的个数, $M_i = \sum_{j=1}^n M_{ij}$ 。

(2) 用一重转移概率预测:假定预测对象目前为 E_k 状态,则 P_{kj} 描绘目前 E_k 状态转向 E_j 状态的可能性,若存在 P_{kj}^* 突出大于其他各个 P_{kj} ,则未来处在 E_j^* 的可能性最大;若 P_{kj} 诸值相差不大,根据初始概率,计算二重转移概率,用二重转移概率预测。

(3) 用二重转移概率预测:二重转移概率 $P_{ijk} = (E_i \rightarrow E_j \rightarrow E_k) = M_{ijk}/M_{ij}$, $i_{jk} \in R$ 。 P_{ijk} 预测对象由状态 $E_i \rightarrow E_j \rightarrow E_k$ 的转移概率; M_{ij} 由 E_i 状态向 E_j 状态转移的次数; M_{ijk} 由 E_i 状态向 E_j 状态再转移到 E_k 状态的次数。

(4) 转移概率矩阵:对于马尔科夫链,用 P_{ij} 表示系统由状态 E_j 经过一次转移到达状态 E_j 的转移

概率。由转移概率构成矩阵 P ,即

$$\begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \vdots \\ P_{i1} & \cdots & P_{in} \end{bmatrix}$$

n 为状态空间的维数, P_{ij} 满足 $0 \leq P_{ij} \leq 1$, $\sum_{i=1}^n P_i = 1 (i, j = 1, 2, \dots, n)$ 。若 P 为标准概率矩阵,则存在非零向量 $\pi = [\pi_1, \pi_2, \dots, \pi_n]$, π_j 满足 $0 \leq \pi_j \leq 1$, $\sum_{j=1}^n \pi_j = 1$, $\pi(n+1) = \pi(n)P$ 。

3 实例分析

3.1 巢湖概况

巢湖位于安徽省中部,地处长江与淮河两大河流之间,属长江下游左岸水系,为我国五大淡水湖之一。其东西长54.5 km,南北宽21 km,水域面积约769.5 km²,为安徽境内最大的湖泊,沿岸为合肥市、巢湖市、庐江县所包围,以忠庙-姥山-齐头嘴为界,将巢湖划分为东、西两大湖区。近年来由于沿岸工业和生活污水的排放,水质日益恶化,富营养化程度严重。本文根据巢湖流域重点监测断面,合肥湖滨断面、巢湖裕溪口断面2001-2010年的水质监测数据,选取COD_{Mn},TP,TN为主要污染因子,运用指数平滑法和马尔科夫模型预测东西半湖湖水水质2011年和2012年的状态,并将两种方法的预测结

果对比。有助于了解巢湖水质状况与主要影响因素,为巢湖水质的改善与治理提供借鉴意义。

3.2 指数平滑法在巢湖水质中的应用

本次研究取巢湖合肥湖滨与巢湖裕溪口两大监测断面水质,即西半湖湖体和东半湖湖体水质,2001-2010年水质数据,数据来源于中国环境年鉴。平滑系数 α 取0.4,序列首项2001年数据取前三年平均值,后几年采用一次指数平滑法处理,经一次平滑处理后新序列如表1所示。2010年西半湖 COD_{Mn} 实测值是5.40,根据平滑法预测2010年 COD_{Mn} 数值为5.83,相对误差为8%,误差较小,其他因子类似,满足处理与预测需要,同时预测2011年水质结果示于表1,实测值与平滑处理值对比见图1,水质对应的状态见表2。

表1 巢湖水质实测值与平滑值 mg/L

年份	西半湖			东半湖		
	COD_{Mn}	TP	TN	COD_{Mn}	TP	TN
2001	5.85	0.28	3.45	4.34	0.117	1.75
2002	6.00	0.27	3.26	4.26	0.122	1.93
2003	5.82	0.25	3.24	4.47	0.120	1.78
2004	5.81	0.30	3.61	4.31	0.113	1.67
2005	5.77	0.31	3.67	4.18	0.117	1.48
2006	5.78	0.32	3.32	3.95	0.114	1.36
2007	6.27	0.27	2.61	4.28	0.113	1.48
2008	6.32	0.25	2.12	4.23	0.108	1.46
2009	6.11	0.24	2.27	4.45	0.097	1.46
2010	5.83	0.19	2.20	4.54	0.088	1.32
2011	5.66	0.17	2.24	4.40	0.081	1.19

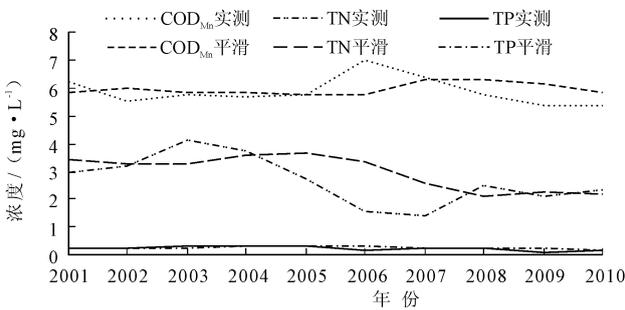


图1 实测与平滑数值对比

表2 指数平滑法预测2011年水质所处状态

污染因子	西半湖			东半湖		
	COD_{Mn}	TP	TN	COD_{Mn}	TP	TN
水质类别	III	III	劣V	III	II	IV

3.3 马尔科夫预测法在巢湖水质中的应用

以指数平滑法处理后的新序列,根据地表水环境质量标准(GB 3838-2002)将水质分为相应的等

级状态(表3)。以2001-2010年的数据为基础,预测2011年与2012年的状态,并与指数平滑法预测的具体数值对应的状态作对比。根据前10个时段的分类状态及转移次数,得到表4~8。

利用表4~8,得出转移举证 P 。以西半湖的 COD_{Mn} 为例, $P_{33} = 5/6, P_{34} = 1/6, P_{43} = 1/3, P_{44} = 2/3$,其余为0,得到西半湖 COD_{Mn} 的转移矩阵 P_{W_COD} ,以此类推得到其他因子转移矩阵,由于西半湖TN状态均为劣V,目前未存在状态转移,所以不计。2010年西半湖 COD_{Mn} 处于III状态,初始状态概率向量 $\pi(0) = [0, 0, 1, 0, 0]$,所以2011年状态 $\pi(1) = \pi(0) * P_{W_COD} = [0, 0, 5/6, 1/6, 0]$,2012年状态 $\pi(2) = \pi(1) * P_{W_COD} = [0, 0, 3/4, 1/4, 0]$, $\pi(3) = \pi(2) * P_{W_COD} = [0, 0, 17/24, 7/24, 0]$ 。以此类推,获得其他因子转移向量,见表9。

表3 巢湖水质主要污染因子所处状态

年份	西半湖			东半湖		
	COD_{Mn}	TP	TN	COD_{Mn}	TP	TN
2001	III	IV	劣V	III	III	V
2002	III	IV	劣V	III	III	V
2003	III	IV	劣V	III	III	V
2004	III	IV	劣V	III	III	V
2005	III	V	劣V	III	III	IV
2006	III	V	劣V	II	III	IV
2007	IV	IV	劣V	III	III	IV
2008	IV	IV	劣V	III	III	IV
2009	IV	IV	劣V	III	II	IV
2010	III	III	劣V	III	II	IV

表4 西半湖 COD_{Mn} 等级状态及转移次数统计表

起始状态	西半湖 COD_{Mn} 转移等级状态					合计
	I	II	III	IV	V	
I	0	0	0	0	0	0
II	0	0	0	0	0	0
III	0	0	5	1	0	6
IV	0	0	1	2	0	3
V	0	0	0	0	0	0

表5 西半湖TP等级状态及转移次数统计表

起始状态	西半湖TP转移等级状态					合计
	I	II	III	IV	V	
I	0	0	0	0	0	0
II	0	0	0	0	0	0
III	0	0	0	0	0	0
IV	0	0	1	5	1	7
V	0	0	0	1	1	2

表 6 东半湖 COD_{Mn} 等级状态及转移次数统计表

起始 状态	东半湖 COD _{Mn} 转移等级状态					合计
	I	II	III	IV	V	
I	0	0	0	0	0	0
II	0	0	1	0	0	1
III	0	1	7	0	0	8
IV	0	0	0	0	0	0
V	0	0	0	0	0	0

表 7 东半湖 TP 等级状态及转移次数统计表

起始 状态	东半湖 TP 转移等级状态					合计
	I	II	III	IV	V	
I	0	0	0	0	0	0
II	0	1	0	0	0	1
III	0	1	7	0	0	8
IV	0	0	0	0	0	0
V	0	0	0	0	0	0

表 8 东半湖 TN 等级状态及转移次数统计表

起始 状态	东半湖 TN 转移等级状态					合计
	I	II	III	IV	V	
I	0	0	0	0	0	0
II	0	0	0	0	0	0
III	0	0	0	0	0	0
IV	0	0	0	5	0	5
V	0	0	0	1	3	4

$$Pw_COD = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5/6 & 1/6 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Pe_COD = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1/8 & 7/8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Pw_TP = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/7 & 5/7 & 1/7 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

$$Pe_TP = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1/8 & 7/8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Pw_TN = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 \end{bmatrix}$$

表 9 变化状态向量

	预测	I	II	III	IV	V	Σ
COD _{Mn} 西	2011	0	0	5/6	1/6	0	1
	2012	0	0	3/4	1/4	0	1
	2013	0	0	17/24	7/24	0	1
COD _{Mn} 东	2011	0	1/8	7/8	0	0	1
	2012	0	1/9	8/9	0	0	1
	2013	0	1/9	8/9	0	0	1
TP 东	2011	0	1	0	0	0	1
	2012	0	1	0	0	0	1
	2013	0	1	0	0	0	1
TN 东	2011	0	0	0	1	0	1
	2012	0	0	0	1	0	0
	2013	0	0	0	1	0	0

由表 9 可以看出,如没有重大事故发生,西半湖的 COD_{Mn} 在 2013 年的时候有很大可能是处于 III 类水;由于近 10 年 TP 状态均处于 IV、V 类,未有处于 III 类状态,概率矩阵 Pw_TP 中,其他状态转移到 III 类状态的转移等级状态无,所以 TP 目前不好预测;TN 近 10 年均处于劣 V 类状态,所以未存在状态转移,没有特殊事件发生,TN 状态不会变。东半湖的 COD_{Mn} 在 2013 年的时候处于 III 类水的概率最大,TP 处于 II 类水,TN 在处于 IV 类水。上述状态和表 2 中运用指数平滑预测的具体数值所对应的状态一致。

近几年和未来几年,巢湖西半湖水质 TN 为劣 V 类,浓度一直很高,是主要污染因子,东半湖的水质要优于西半湖,TN 为主要污染因子,未来几年如不采取治理措施 COD_{Mn} 浓度将会升高。通过预测主要污染因子所处的区间,不仅可以宏观上把握水质的变化趋势,而且有助于重视超标因子或不容乐观的因子,从而采取相应的措施。

4 结论与展望

4.1 结论

根据 2001 - 2010 年时段的连续数据,通过指数平滑法,预测得到 2011 年具体的数值,将数值按照地表水质量标准划分为对应的等级状态。将马尔科夫模型运用到水质预测中,将平滑处理后的水质划分为若干个状态,统计水质状态间转换的概率,通过

概率矩阵,由初始的水质状态求得未来时段水质处于某种级别的概率。将两种方法得出的状态对比,发现水质预测结果一致。

由于近十年巢湖西半湖 TN 水质均处于劣 V,没有特殊事件发生的话仍将处于劣 V 状态,东半湖 TN 也不容乐观,未来几年需要注意巢湖 TN 的变化趋势,采取相应的措施降低总氮的浓度;东半湖 COD_{Mn} 目前水质状态还比较乐观,但也在 III 类水状态徘徊,所以也需要关注东半湖 COD_{Mn} 的变化,防止其成为主要污染因子。

马尔科夫模型不能准确地预测未来时间段中污染因子具体的数值,但能得出一个预测的范围,从而反映指标的变化趋势。如果将状态等级分类进一步细化,预测结果将趋向于具体的数值。但是如果收集的数据较少时,得出的概率转移矩阵随机性比较强,则不能较好的体现出变化趋势。本文结合了指数的平滑法,先对历年数据进行平滑处理,一来降低数据的随机性,二来可将马尔科夫预测法与指数平滑法预测的结果进行综合对比分析,这能够更好的预测水质变化情况。

4.2 展望

由于数据收集的困难,本文仅收集到巢湖东西半湖湖体 2010 年之前的水质数据,这不太适用于长时间尺度上的精确预测,但对于宏观上把握水质的变化趋势以及污染因子的动态有一定的意义,同时通过 2 种方法的结合,体现出指数平滑法 - 马尔科夫预测是一种较好的水质预测模型,尤其是在水质监测数据较多的情况下。在今后的工作中,继续收集近两年的水质数据,预测将来的水质,为水质预报与治理提供借鉴意义。

参考文献:

[1] 邱林,黄鑫,李洪良. 基于模糊权马尔可夫模型的综合水质预测[J]. 人民长江,2007,38(1):75-77.

[2] 张永祥,蔡德所,唐迎春. 马尔可夫链式灰色模型在南宁邕江水质预测中的应用[J]. 广西水利水电,2007(3):1-3.

[3] 孙志霞,孙英兰. GM(1,1)模型研究及其在水质预测中的应用[J]. 海洋通报,2009,28(4):116-120.

[4] 李明昌,梁书秀,孙昭晨,等. 海域组合单元水质模型参数反演方法研究[J]. 应用基础与工程科学学报,2010,

18(3):408-418.

[5] 刘章军,叶燎原. 模糊概率模型及其在地下水水质评价中的应用[J]. 应用基础与工程科学学报,2007,15(3):286-293.

[6] 徐坤,牟春友. 运用方差分析法分析白城市月亮湖水库水质[J]. 中国环境监测,2009,25(2):67-68.

[7] 岳遥,李天宏. 基于模糊集理论的马尔可夫模型在水质定量预测中的应用[J]. 应用基础与工程科学学报,2011,19(2):231-242.

[8] 王长江. 指数平滑法中平滑系数的选择研究[J]. 中北大学学报(自然科学版),2006,27(6):558-561.

[9] 刘造保,徐卫亚,张开普,等. 基于改进指数平滑法的岩体边坡变形预测[J]. 河海大学学报(自然科学版),2009,37(3):313-316.

[10] 张盼月,钟政林,曾光明,等. 马尔科夫过程在河流综合水质预报中的应用[J]. 环境工程,1997,15(2):41-44.

[11] 张树冬,李伟光,南军,等. 加权马尔可夫链预测源水水质的应用研究[J]. 中国给水排水,2009,25(13):111-113.

[12] 秦松. 基于加权的滑动平均 - 马尔科夫预测模型及其应用[J]. 水资源与水工程学报,2013,24(1):185-188.

[13] He B, Che C, Ga D H. Evaluation of water quality based on Markov process[J]. Environmental Engineering, 2003, 21(2):60-62.

[14] 牛军宜,冯平. 基于 Markov 状态切换的水质时序自回归预测模型[J]. 吉林大学学报(地球科学版),2010,40(3):657-664.

[15] Yun Yi, Zou Zhihong, Feng Wei, et al. Quantificational analysis on progress of river water quality in China original research article[J]. Journal of Environmental Sciences, 2009,21(6):770-773.

[16] 陈启明,陈华友. 基于 Markov 链的最优化模型在长江水质预测中的应用[J]. 安徽工业大学学报(自然科学版),2011,28(1):78-81.

[17] 程万里,李亦芳,郝伏勤,等. 黄河三门峡段基于马尔可夫模型的水质预测[J]. 电力环境保护,2008,24(3):8-11.

[18] 李娟,张维江,马轶. 滑动平均 - 马尔可夫模型在降水预测中的应用[J]. 水土保持研究,2005,12(6):196-205.

[19] 冯为,邹志红. 河流水质的动态马尔柯夫评价[J]. 环境工程学报,2007,1(8):132-135.

[20] 余秋梅,印士勇,周良伟. 巢湖水环境质量现状分析[J]. 人民长江,2001,32(7):29-30.