

黄 炜. 蓝藻水华 Probit 短期预测模型[J]. 江苏农业科学, 2014, 42(1): 337–342.

# 蓝藻水华 Probit 短期预测模型

黄 炜

(上海大学管理学院, 上海 200444)

**摘要:**蓝藻水华反演图被用于判断某一水域在某一时刻是否暴发蓝藻水华,进而直接以蓝藻水华发生与否的二元变量为被预测变量,以水质、水文、气象 3 类监测变量为预测变量构建蓝藻水华暴发 Probit 短期预测模型。以太湖大贡山水域作为案例进行该预测模型的实证研究。结果表明,该预测模型的评价指标值较好;平均相对误差为 13.5%,接近或小于 2 个对照模型;该模型在空间精度和时间精度方面具有显著优势;隔天预测模型的准确性最高,预测周期加长时预测准确性降低;将所有可用监测变量都纳入预测模型时的预测准确度高于仅采用可用监测变量的若干子集时的准确度。

**关键词:**蓝藻水华;预测;Probit 模型

**中图分类号:** X524      **文献标志码:** A      **文章编号:** 1002–1302(2014)01–0337–06

蓝藻是一种原始而古老的藻类原核生物,气温较高尤其在夏季时容易在富营养化的湖泊、水库、河流中大量繁殖,在适宜的气象条件与水文条件时上浮、漂移、聚集,有时在水面聚集成一层蓝绿色而有腥臭味的浮沫,被称为蓝藻水华。蓝藻暴发和蓝藻水华暴发有所区别,蓝藻暴发可以只是水体中蓝藻大量增殖或聚集,即蓝藻丰度增大,但不一定覆盖水面;而蓝藻水华暴发则是指蓝藻覆盖了某一(些)水域的水面,其危害更大,如可使水体含氧量降低,造成鱼虾大量死亡。近年来,太湖蓝藻水华的持续时间有所增加,3—12 月都有发生<sup>[1]</sup>,而且发生频率和空间范围都大大增加<sup>[2–3]</sup>;近年来滇池

每年都会发生不同范围、不同程度的蓝藻水华<sup>[4]</sup>;波罗的海每年出现的夏末蓝藻水华也引起了公众关注<sup>[5]</sup>,蓝藻水华问题已成为国内外主要环境问题之一。对蓝藻水华的预测和预防十分困难,已成为环境保护领域的一个关键问题。本研究以太湖为例,以大贡山水域的蓝藻水华为对象,研究蓝藻水华的预测,以期为其他水域和水体的蓝藻水华预测提供参考。

## 1 蓝藻水华暴发预测的研究现状

已有研究表明,蓝藻水华暴发受多种因素影响。(1)蓝藻的生理特点和生长特性。蓝藻聚集是在水底越冬的蓝藻在适当条件时转换到浮性生活期的过程<sup>[6]</sup>。微囊藻借助其悬浮机制置身于水体中有利其自身生长的位置<sup>[7]</sup>。为了得到适宜的光照,蓝藻以形成水华的方式作为其适应环境的生态对策的一部分<sup>[8]</sup>。蓝藻生长与蓝藻水华形成由休眠、复苏、生长、上浮聚集等 4 个阶段组成<sup>[9]</sup>。(2)其他水生生物。浮游植物群落的规模是由(水生)捕食动物和营养盐共同控制

收稿日期:2013–06–14

基金项目:教育部“新世纪优秀人才支持计划”(编号:NCET–10–0938)。

作者简介:黄 炜(1972—),男,上海人,博士,研究方向为水环境管理。E-mail:hw3@shu.edu.cn。

禽养殖业污染风险势必会随之不断加大。因此,河南省应采取积极有效的措施来防治畜禽养殖业的污染。

## 参考文献:

- [1]国家统计局. 中国畜牧业年鉴:2000—2011[M]. 北京:中国农业出版社,2001—2012.
- [2]2011 年中国环境状况公报[EB/OL]. (2012–06–06)[2013–10–22]. <http://jcs.mep.gov.cn/hjzl/zkgb/2011zkqb>.
- [3]全国畜禽养殖污染防治“十二五”规划[[EB/OL]. (2013–01–04)[2013–10–22]. [http://www.zhb.gov.cn/zhxx/hjyw/201301/t20130104\\_244581.htm](http://www.zhb.gov.cn/zhxx/hjyw/201301/t20130104_244581.htm).
- [4]彭 里. 畜禽养殖环境污染及治理研究进展[J]. 中国生态农业学报,2006,14(2):19–22.
- [5]吴飞龙,林代炎,叶美锋. 福建省畜禽养殖业废弃物污染风险评估[J]. 中国农学通报,2009,25(24):445–449.
- [6]苏 杨. 我国集约化畜禽养殖场污染问题研究[J]. 中国生态农业学报,2006,14(2):15–18.
- [7]朱兆良,孙 波,杨林章,等. 我国农业面源污染的控制政策和措

施[J]. 科技导报,2005,23(4):47–51.

- [8]郑纯辉,赵 杰. 河南省集约化畜禽养殖业环境污染状况及管理对策[J]. 郑州牧业工程高等专科学校学报,2006,26(2):14–16,20.
- [9]河南省统计局. 河南省统计年鉴:2011[M]. 北京:中国统计出版社,2011.
- [10]李建国. 畜牧学概论[M]. 北京:中国农业出版社,2002.
- [11]GB 18596—2001 畜禽养殖业污染物排放标准[S]. 北京:中国环境出版社,2002.
- [12]农业部全国土壤肥料总站肥料处. 肥料检测实用手册[M]. 北京:中国农业出版社,1990:410–411.
- [13]国家环境保护总局自然生态保护司. 全国规模化畜禽养殖业污染情况调查及防治对策[M]. 北京:中国环境科学出版社,2002:77–78.
- [14]王方浩,马文奇,窦争霞,等. 中国畜禽粪便产生量估算及环境效应[J]. 中国环境科学,2006,26(5):614–617.
- [15]《中国环境年鉴》编辑委员会. 中国环境年鉴:2003[M]. 北京:中国环境出版社,2003.

的<sup>[10]</sup>。微囊藻细胞形成较大的群体后就可以有效抵御原生动物、浮游动物的捕食<sup>[11]</sup>。蓝藻能形成群体胶鞘,这降低了其被浮游动物摄食的可能;蓝藻能分泌他感物质,这对其他物种的生长有抑制作用;其分泌的藻毒素对其他物种也有抑制作用<sup>[9]</sup>。(3)水质情况。在相同条件时硝酸盐浓度与藻蓝素浓度呈正相关<sup>[12]</sup>。美国环境保护署指出,在湖泊与水库中总磷、总氮浓度分别超过 10、150  $\mu\text{g/L}$  时即有可能发生蓝藻水华<sup>[13]</sup>。Bulgakov 等认为,氮磷浓度比降至 5~10 常导致蓝藻成为浮游植物群落的主导者<sup>[14]</sup>。在澳大利亚鱼腥藻(属蓝藻门)水华在浑浊度低于 50 NTU 时发生<sup>[15]</sup>。(4)气象情况。徐恒省等研究表明,风速、风向和蓝藻在水体中的分布有内在联系<sup>[16]</sup>。武胜利等认为,较大的风速和降水量可以抑制蓝藻水华<sup>[17]</sup>。蓝藻在风速低于 3 m/s 时能够在水下 30 cm 内占据优势<sup>[18]</sup>。光照条件与蓝藻水华形成有密切关系<sup>[19]</sup>。温度在 25°C 以上时大多数蓝藻达到最高生长速率<sup>[20]</sup>。(5)水文情况。在具备充分营养盐与适合水温条件时,蓝藻水华的发生在很大程度上受水体稳定性的影响<sup>[21-22]</sup>。淀山湖水华暴发的频率和规模远低于其他同类型湖泊,一个重要原因是其平均换水周期较短,仅为 30 d 左右<sup>[18]</sup>。风浪和湖流的运动使湖区内的蓝藻聚集于湖岸形成水华<sup>[9]</sup>。水体发生垂直混合时微囊藻的悬浮机制优势丧失<sup>[23]</sup>。

目前国内外有害藻华的预警和预报技术主要是针对藻华的发生和运动进行监测和预测,预测方法主要包括依据指标经验临界值预警、确定性生态数学模型(或基于过程的模型)、不确定的数据驱动模型(包括人工神经网络方法和多元回归方法等)<sup>[1]</sup>。

依据指标经验临界值进行判断隐含着一种错误逻辑,即蓝藻水华的某些相关因素取其自身阈值以上或以下的值是蓝藻水华发生的条件。然而目前为止还没有这样的实证结果,即蓝藻水华的发生是以其某些相关因素的阈值为分水岭。有关影响因素对蓝藻水华的影响趋势并不一定是严格递增或递减的,例如并非氮浓度越高对蓝藻的竞争优势越有利,Blomqvist 等指出, $\text{NO}_3^- - \text{N}$  浓度在小于 200  $\mu\text{g/L}$  时才对蓝藻成为优势种类有利<sup>[24]</sup>。在氮、磷限制条件下微囊藻比其他藻类更具竞争优势,因此在氮、磷浓度较低的许多水体中也可发生蓝藻水华<sup>[9]</sup>。再如很多研究表明,蓝藻水华易在高温时暴发,然而 2007 年 12 月 8 日、2008 年 1 月 3 日太湖 2 次暴发蓝藻水华,而当时日均温分别为 10.6、1.7 °C;此外,2006 年 11 月 2 日太湖吴县东山气象站的日照时数为 0,但当天蓝藻仍然大面积暴发,说明充足的日照并非蓝藻水华暴发的必要条件<sup>[17]</sup>。由此可见,设定相关因素阈值进行预警的方法是不可取的。

许秋瑾等提出了一种太湖藻类生长的模型,考虑藻类的生长率和死亡率,估算藻类浓度<sup>[25]</sup>。确定性数学模型需要大量的训练数据,为了开发基于过程的模型,研究者还须要充分理解所有基础的物理过程和物理化学过程,以便用数学方法将其表达出来。由于自然系统的巨大复杂性,这种方法常常是行不通的<sup>[26]</sup>。开发基于过程的模型难度较大,这也降低了此类模型可推广性。另外,蓝藻水华暴发的原因非常复杂,湖泊内不同水域的环境因素与蓝藻水华暴发之间的响应关系往往并不相同,年度差别也相当明显,因而水华发生具有很大的不确定性。这些因素导致消耗大量时间和成本后仍不能得到

理想的生态数学模型<sup>[1]</sup>。上述情况也是导致此类模型的可推广性不强的原因——不能保证将已有模型用于不同水域时的可靠性,所以对不同水域须另行开发专用的基于过程的预测模型。还有,由于基于过程模型的自变量和因变量值往往属于同期,因此必须用自变量的未来值预测蓝藻水华。如果使用一个预测模型时须首先预测或估算诸多自变量甚至全部自变量在未来时期的值,双重预测显然会降低蓝藻水华预报结果的准确性。

相对确定性生态数学模型而言,开发不确定的数据驱动模型则快捷、简便、成本较低。人工神经网络方法具有较强的适应能力和学习能力。刘静玲等建立了基于人工神经网络的北京“六海”水华预警模型,以叶绿素 a 浓度为被预测变量<sup>[27]</sup>。Maier 等使用人工神经网络预测河流中的蓝藻水华<sup>[26]</sup>。但人工神经网络方法需要大量的训练数据,数据收集难度很大;而且该方法是一种完全的黑箱方法,研究者很难解释神经网络结构及其对输出变量的影响,也难以解释输入变量对输出变量的影响。陈翔等基于 18 个旬度数据记录构建了影响洪泽湖蓝藻水华暴发的气象因子类多因子交叉相关综合法预报方程,进行下一旬中期预报<sup>[24]</sup>。但该方法只能以旬为时间单位进行预报,不能精确到蓝藻水华日暴发情况。该方法适用于类似洪泽湖这样的蓝藻水华暴发频率不高的湖泊,但不适合太湖、滇池这种水华暴发频率很高的湖泊。另外,该方法将连续型的自变量离散化,导致信息量损失,降低了预测模型精度。离散化的做法还可能导致预测模型的逻辑错误或判断错误,导致水华误报或漏报。一些实际投入使用的预测系统则将定量方法和定性方法相结合,例如中国科学院南京地理与湖泊研究所的预报系统<sup>[2]</sup>和美国的墨西哥湾赤潮预报系统<sup>[28]</sup>。

## 2 蓝藻水华 Probit 短期预测模型

为了解决上述问题,笔者构建了以蓝藻水华是否发生为被预测变量,以水质、水文、气象 3 类监测变量为连续型预测变量的非线性预测模型。该模型属于数据驱动模型,且克服了传统多元(线性)回归分析方法和人工神经网络方法的不足。

蓝藻细菌的生长速率为 0.3~1.4 倍/d,比其他很多藻类慢得多<sup>[20]</sup>。蓝藻的生长发育时间为 7~12 d,在单体水中蓝藻大量增殖形成堆积,才会形成水华;但在非单体水中则不一定如此,蓝藻水华可由蓝藻输移形成<sup>[29]</sup>。蓝藻达到一定数量后,低流速、高温、弱风、弱降水、强光照、低气压可能导致水华发生<sup>[18]</sup>。在多数情况中,这种突然出现的水华是已存在的蓝藻群落在水中的重新分布(即上浮、聚集于水体表层)引起的;而蓝藻群落的重新分布则是由适宜的气象与水文条件引起的,例如对微弱湍流条件的反应。蓝藻水华并不全是由藻类的原位生长所致,不是短时间内蓝藻数量爆炸性增长的结果<sup>[30]</sup>。因此对蓝藻水华进行短期(如 1 周内)预测时,可以将重点放在考察水质、水文、气象 3 类环境因素方面。在蓝藻水华易暴发的季节,尤其是在研究水域的蓝藻密度较高时,蓝藻在自身生长和种群竞争方面已经占有优势,或者说这 2 方面因素已不是蓝藻水华暴发的限制因素,因此仅以这 3 类指标作为蓝藻水华短期预测模型的潜在自变量(预测变量)不会明显降低预测精度。

目前国内外的蓝藻水华预测模型主要预测叶绿素 *a* 浓度、藻类生物量、蓝藻密度/丰度或藻蓝素浓度等指标,或者用这些指标作为蓝藻水华暴发的替代指标,并未直接预测蓝藻水华是否暴发<sup>[25-27,31-32]</sup>。叶绿素 *a* 浓度或蓝藻密度等指标只能从微观角度表示蓝藻水华强度,并不能确切地度量或反映蓝藻水华发生与否;再加上蓝藻水华暴发条件的多样性和复杂性,叶绿素 *a* 浓度或蓝藻密度等指标与其预测变量之间的响应关系并不能代表蓝藻水华暴发与这些预测变量之间的响应关系。例如有研究表明,当湖体中叶绿素 *a* 含量超过 10  $\mu\text{g/L}$  或蓝藻密度达到 2 万细胞/mL 时可被称为藻类水华<sup>[1]</sup>;然而孔繁翔等的观测报告显示,2007 年 8 月上旬太湖水体中叶绿素 *a* 浓度大于水华形成阈值,但由于台风过境,几乎没有观测到蓝藻水华<sup>[2]</sup>。在水华易发生季节,没有人干扰时,只要环境合适富营养化湖泊中的叶绿素一般都保持高浓度状态,这期间叶绿素并没有迅速消失或增加<sup>[9]</sup>。另外尽管对蓝藻水华的定义并不统一,但在该领域内有一个较通行的观点,即蓝藻颗粒覆盖水面是蓝藻水华暴发的重要特征<sup>[33-36]</sup>。但水体中叶绿素 *a* 浓度或藻类密度等指标较高时蓝藻颗粒并不一定形成对水面的覆盖;另一方面,由于蓝藻颗粒上浮、覆盖水面,此时水体中叶绿素 *a* 浓度或藻类密度反而降低<sup>[9]</sup>。因此蓝藻颗粒对水面的覆盖与水体中叶绿素 *a* 浓度或藻类密度之间并无确定的关系,或者说关于叶绿素 *a* 浓度或藻类密度并没有一个科学、准确、确定的阈值可被用来判断蓝藻水华暴发与否。综上,不宜用叶绿素 *a* 浓度或藻类密度等指标作为蓝藻水华暴发与否的指标。

本研究中蓝藻水华反演图(图 1)被用来判断某个水域某个时刻是否暴发蓝藻水华,并以此指标作为被预测变量。目前国内外尚无研究者基于此判断方法和指标进行蓝藻水华预测。蓝藻水华反演图是基于卫星遥感影像数据制作的,而卫星遥感影像在反映蓝藻水华暴发情况方面具有较多优势,例如其从宏观角度反映蓝藻水华的暴发情况,因而更确切;它是由特定的卫星遥感传感器成像,因而反映情况较客观;它基于覆盖水面的蓝藻水华的光谱特征而成像,符合蓝藻水华通行的概念。

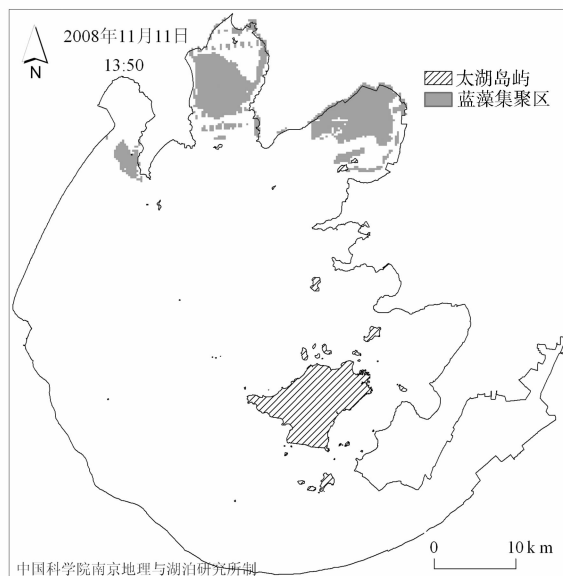


图1 某时刻太湖蓝藻水华反演图示例

在该蓝藻水华预测模型中,被预测变量是蓝藻水华暴发与否,是二元变量,只取“发生”或“未发生”2 个值;而预测变量大多为连续型变量。对于这一特殊的数据模式,一般的研究方法或模型均不适用,而离散因变量模型类型中的二元选择模型较合适。二元选择模型被用来研究在给定条件时作一种选择而舍弃另一种选择的概率。Probit 模型就是一种二元选择模型<sup>[37]</sup>,本研究中蓝藻水华预测模型采用 Probit 模型的形式。用“发生”和“未发生”这 2 个状态描述蓝藻水华,该研究方法包含如下概念:这 2 个状态分别代表一个指标或指标组合的某个取值范围。例如蓝藻在水面的集聚面积大到某个程度(可以是定性指标)以上、藻蓝素浓度高到某个程度以上时定义为蓝藻水华“发生”。因此根据研究背景 and 需要,可基于相关指标值确定蓝藻水华“发生”或“未发生”。例如笔者基于蓝藻水华反演图的宏观特性判断水华“发生”或“未发生”,即这 2 个状态的定义可因不同研究需要而变化,因此该研究方法具有广泛的适用性。

在 Probit 模型中因变量  $y$  是关于自变量向量  $x$  的、服从正态分布的非线性函数。以蓝藻水华发生或不发生的状态作为因变量  $y$ ,以若干潜在相关环境因素作为自变量向量  $x$ ,构建二元选择水华预测模型。总体模拟模型为:

$$y = 1 - F(-\beta^T x) + \mu \quad (1)$$

式中: $F(\cdot)$  为累积分布函数,对于 Probit 模型而言, $F(\cdot)$  为  $\Phi(\cdot)$ ,即标准正态分布函数; $\beta$  为系数向量; $T$  为向量转置运算符号; $\mu$  是均值为 0 的随机误差项。用  $\Phi(\cdot)$  替换  $F(\cdot)$ ,可得到模型(1)的具体形式:

$$y = \varphi(\beta^T x) + \mu = \int_{-\infty}^{\beta^T x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du + \mu \quad (2)$$

式中: $y = 1$  表示蓝藻水华发生, $y = 0$  表示蓝藻水华不发生。上述模型为一个通用模型,可用于湖泊、水库蓝藻水华的预测。

### 3 案例研究:太湖大贡山水域蓝藻水华暴发预测

调查结果显示,太湖流域藻类以蓝藻门为主,蓝藻门数量最高时约占藻类总量的 91.6%<sup>[38]</sup>。2007 年 5 月底蓝藻大规模暴发,蓝藻水华面积达全太湖面积的 1/3。统计结果表明:2007、2008 年蓝藻水华呈加剧态势<sup>[39]</sup>。2011 年太湖藻类水华暴发频次较往年有一定幅度下降,但小范围水华暴发频次居高不下。梅梁湾、贡湖湾、西部沿岸一带蓝藻水华形势没有根本好转。以下以大贡山水域的蓝藻水华为对象,研究蓝藻水华的预测。

#### 3.1 数据来源和处理

本研究中太湖湖区数据为太湖大贡山水域 2008 年 7—12 月和 2009 年 8—11 月的每日水质、水文、气象 3 类监测数据(共 17 个监测变量、86 条观测记录)和匹配的太湖蓝藻水华反演图,由湖泊-流域科学数据共享平台提供。数据采集时段为每天 08:47—15:44。17 个监测变量为:风速( $WS$ )、风向( $WD$ )、瞬时降雨( $IRF$ )、气温( $AT$ )、水温( $WT$ )、太阳辐射( $I$ )、气压( $AP$ )、水深( $WDEP$ )、电导率( $COND$ )、 $pH$  值( $PH$ )、溶解氧浓度( $DO$ )、电池电压( $V$ )、湿度( $H$ )、蓝藻密度( $CD$ )、叶绿素 *a* 浓度( $CHLA$ )、氨浓度( $NH_3$ )、盐度( $SA$ ),括号中为相应的变量名。这些监测变量属于本研究文献研究提到的后

3 类环境相关因素,因此可以作为潜在预测变量被纳入本预测模型。上述数据均采集自被预测水域即大贡山水域,而不是采用附近水域或地区的数据或更大范围的数据,然后作近似替代。本数据集的空间精度可为预测模型达到较高的预测准确度和较高的空间精度奠定基础。

将监测数据与水华反演图的信息即 0、1 值匹配、组合,并用 Excel 软件对监测数据作归一化处理。由于该模型中被预测变量与预测变量不属于同一期,所以须基于原始数据构造非同期数据集。隔天预测模型数据集的构造方法是:将对应的第 2 天水华暴发与否的值替换每日相应值,使模型能预测第 2 天是否暴发水华,在每日气温值后添加相应第 2 天的气温值(NEXTAT),由于天气预报能提供气温预报值,所以添加此项;将第 2 天没有水华暴发与否数据的日期的记录删除,原因是缺乏可验证预报值的因变量值。这样就形成了所需数据集。用类似方法为相差 2、3 d 进行的预测构造数据集。

3.2 模型拟合、预测检验与准确性分析

将第 1 个数据集(隔天预测数据集,共 18 个自变量、43 条记录,构造隔天预测模型的数据集时删除了部分无效数据记录)分成 2 部分。第 1 部分有 30 条记录,用于拟合预测模型;第 2 部分有 13 条记录,用于检验模型预测的准确性。根据“3.1”中的变量内容,用统计分析软件 EViews 6.0 构建相应的数据结构,并将上述第 1 部分的 30 条记录复制到该数据结构中。然后选择 EViews 中的 Probit 建模工具构建模型,并基于上述数据进行拟合,即估计式(2)中的各系数值。预测模型的拟合形式为:

$$\hat{y} = \Phi(-18.22 - 6.98IRF + 12.23V - 54.67WS - 76.38WD - 61.25AP - 24.49I - 124.33AT + 111.96NEXTAT - 28.66H - 14.45CD + 113.33PH - 57.14CHLA + 18.22DO + 231.24WDEP - 128.57NH_3 - 92.23WT - 81.51COND + 142.25SA) \quad (3)$$

式中: $\Phi(\cdot)$ 为标准正态分布函数。

该 Probit 模型的评价指标如表 1 所示。该模型整体的显著性较高。LR 统计量检验除了常数项以外所有系数都是 0 的假设,测试模型整体的显著性。LR 统计量渐进服从 $\chi^2(k)$

分布, $k$ 为自由度。LR 统计量值在 0.5% 水平上大于 $\chi^2$ 分布的临界值,通过整体显著性检验; $Prob(LR) < 0.5\%$ ,这也表示该模型在 0.5% 的显著性水平上通过整体显著性检验。Andrews 统计量值及其概率表明预测模型通过拟合优度检验,拟合优度较好。

表 1 对第 1 个数据集拟合得到的指标

模型指标	值
LR statistic	41.588 8
Prob(LR statistic)	0.001 3
Andrews Statistic	7.470 2
Prob. Chi - Sq(10)	0.680 4

再将数据集第 2 部分的 13 条记录复制到前述数据结构中,将其置于前述 30 条记录之后。用这 13 条数据记录进行检验。图 2 为预测结果,其中左边部分为蓝藻水华暴发概率的预测值,右边部分为预测结果评价,折线 YF 表示预测值,Y 表示实际值,Y 折线未显示于图中。Y 的情况是:样本 31~41 的值为 1,样本 42、43 的值为 0。结果显示:预测平均绝对误差为 20.4%,预测平均相对误差为 13.5%。平均相对误差(图 2 中平均绝对百分比误差)为预测值与真实值的偏差占真实值的比例。本预测模型的平均相对误差与以往研究结果相近,陈翔等建立的模型在 2008—2009 年试报中误差率为 12.5%<sup>[29]</sup>;孔繁翔等 2008 年的预报误差率为:梅梁湾 22%、贡山湾 16%、大太湖 12%,2009 年的预报误差率为:梅梁湾 40%、贡山湾 30%、大太湖 5%<sup>[2]</sup>。

偏差比例度量了预测值均值与数据序列实际值均值的偏离程度,表示系统误差,本模型的偏差比例较小,为 2.3%;方差比例度量了预测值方差与实际序列方差的偏离程度,图 2 中相应值为 0.4%,也较小。协方差比例衡量了剩余的非系统误差的大小。偏差比例、方差比例、协方差比例之和为 1。如果前两者较小,而协方差比例较大,则预测结果比较理想<sup>[40]</sup>。以上指标值显示本模型的预测结果较好。本案例研究还表明:将所有可用监测变量都纳入预测模型时的预测准确度高于仅采用可用监测变量诸多子集时的准确度,这印证了蓝藻水华暴发受多种因素影响的一般观点。

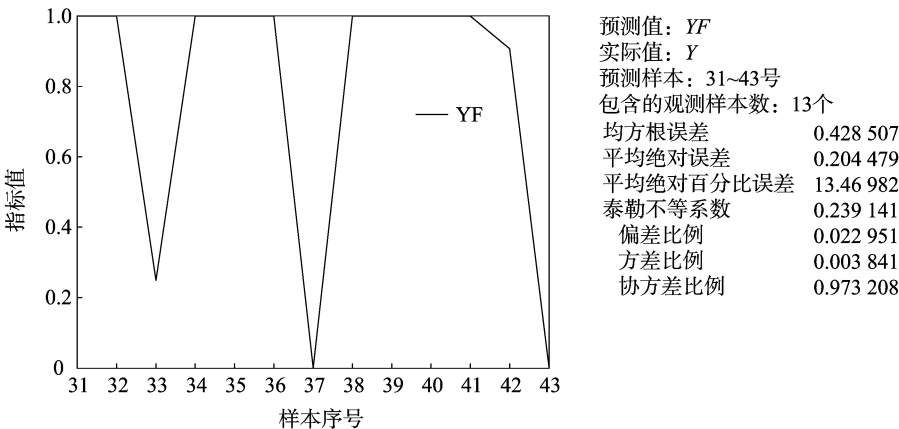


图2 隔天预测模型预测的结果

预测蓝藻水华时更须要知道蓝藻水华发生的确切位置,因为目前蓝藻水华在夏季、秋季容易暴发,但是只有湖滨岸带、重要景观湖区、饮用水源地的蓝藻水华才有可能给人们的

生活与社会稳定带来威胁<sup>[2]</sup>。本模型针对的是某一个水域,可以预测蓝藻水华暴发确切位置;而陈翔等的研究针对的是整个洪泽湖,从而有利于其预测准确率处于较高水平,但这

种准确率是以较低的空间精度为代价的<sup>[29]</sup>。因而本研究的空

间精度更高,或者说空间粒度更细。孔繁翔等的研究中部分预测的空间粒度也非常粗,其地点预测误差率在 16%~53%<sup>[2]</sup>。

本预测模型的时间精度为 1 d。而孔繁翔等的研究中蓝藻水华预报系统的预报时间精度是 3 d<sup>[2]</sup>,其时间精度没有本模型的高;陈翔等的模型时间精度更差,为旬<sup>[29]</sup>。这有利于这些模型的预测准确率处于较高水平,同样这种准确率是以较低的时间精度为代价的。也就是说本预测模型达到与前者相近的预测准确度时,本预测模型的实用性更强。因此整体而言,本预测模型的准确度高于对照模型。

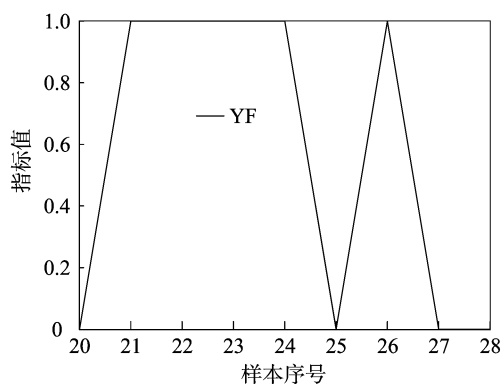


图3 预测周期为相差2 d的模型预测结果

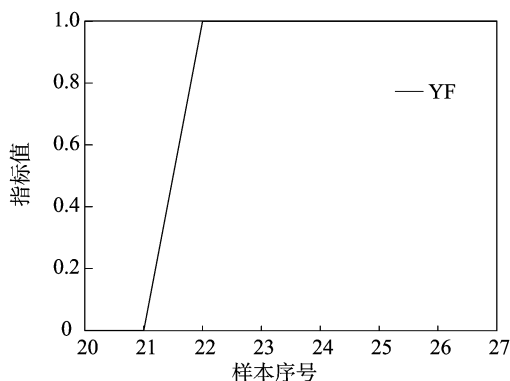


图4 预测周期为相差3 d的模型预测结果

第 2 个、第 3 个模型的预测效果变差的潜在原因除了拟合预测模型的可用样本数量明显减少外,还有预测周期加长导致不确定性因素增多的不利影响。

#### 4 结语

笔者根据蓝藻水华反演图判断太湖大贡山水域在某一时刻是否暴发蓝藻水华,用蓝藻水华暴发与否作为预测模型的被预测变量,用水质、水文、气象 3 类监测变量作为预测变量,基于太湖大贡山水域的监测数据,使用二元选择模型构建了蓝藻水华短期预测模型。该预测模型的平均相对误差为 13.5%,该模型的误差接近或小于对照模型误差,预测结果较为理想;并且该模型在空间精度和时间精度方面具有显著优势。整体而言,该模型的准确度比对照模型更高。本研究中被预测对象(蓝藻水华发生与否的二元变量)比诸多文献中的被预测对象更直接、有效。该预测模型能为有关部门提供蓝藻水华预测,有利于短期应急措施实施。本研究还涉及不

#### 3.3 其他预测周期与预测准确度分析

相差 2 d 的预测数据集共有 28 条观测记录,被分成 2 部分。第 1 部分 19 条记录,用于拟合预测模型;第 2 部分 9 条记录,用于检验模型预测精度。如图 3 所示,预测周期为相差 2 d 的模型预测平均绝对误差和平均相对误差均比隔天预测模型的大,其他评价指标值也偏大(差),预测精度有所降低。

相差 3 d 的预测数据集共有 27 条观测记录,被分成 2 部分。第 1 部分 19 条记录,用于拟合预测模型;第 2 部分 8 条记录,用于检验模型预测精度。如图 4 所示,预测周期为相差 3 d 的模型预测均方根误差、平均绝对误差、Theil 不等系数、偏差比例、方差比例都偏大,预测效果不好。

预测值: YF  
实际值: Y  
预测样本: 20~28号  
包含的观测样本数: 9个  
均方根误差 0.577 350  
平均绝对误差 0.333 349  
平均绝对百分比误差 22.222 22  
泰勒不等系数 0.369 657  
偏差比例 0.037 027  
方差比例 0.001 948  
协方差比例 0.961 025

预测值: YF  
实际值: Y  
预测样本: 20~27号  
包含的观测样本数: 8个  
均方根误差 0.707 107  
平均绝对误差 0.500 000  
平均绝对百分比误差 12.500 00  
泰勒不等系数 0.449 490  
偏差比例 0.125 000  
方差比例 0.008 975  
协方差比例 0.866 025

同预测周期对预测准确性的影响,结果显示隔天预测模型的准确性最高,预测周期加长时准确性降低。

本研究方法具有实用性和普适性,因为在诸多湖泊、水库、河流都设有自动监测设备或人工监测点,而卫星遥感服务又具有广泛覆盖性,只要通过合适方法获得这些数据后就能构建此类预测模型,并进行实际预测工作。该预测模型属于数据驱动型模型,适合处理蓝藻水华暴发的复杂性和不确定性,因而易于被应用到不同湖泊、水库,这适应了为不同水域的蓝藻水华暴发构建专用预测模型的要求,但却不必花费很多精力和成本另行开发模型(如采用确定性模型时)。由于各种原因,本研究中的数据未能涵盖太湖大贡山水域 2008、2009 年的所有有效数据记录,但即使这样,本预测模型亦具有较好的预测效果。可以预计,如果能获得更全面的监测变量和更完整的数据记录,以及更多年份的数据记录和匹配的蓝藻水华反演图,将之用于本预测模型的拟合,则可以获得与实际情况更吻合的模型和更高的预测准确率。

本研究中第 1 个预测模型具有更高准确性的原因是：(1) 对被预测变量的科学设计以及表示被预测变量数据的正确选择；(2) 大量、精确数据的支撑；(3) 选用了合适的建模工具——二元选择模型。

#### 参考文献：

- [1] 王 崇, 孔海南, 王欣泽, 等. 有害藻华预警预测技术研究进展[J]. 应用生态学报, 2009, 20(11): 2813–2819.
- [2] 孔繁翔, 马荣华, 高俊峰, 等. 太湖蓝藻水华的预防、预测和预警的理论与实践[J]. 湖泊科学, 2009, 21(3): 314–328.
- [3] 商兆堂, 任 健, 秦铭荣, 等. 气候变化与太湖蓝藻暴发的关系[J]. 生态学杂志, 2010, 29(1): 55–61.
- [4] 谢国清, 李 蒙, 鲁韦坤, 等. 滇池蓝藻水华光谱特征、遥感识别及暴发气象条件[J]. 湖泊科学, 2010, 22(3): 327–336.
- [5] Lips I, Lips U. Abiotic factors influencing cyanobacterial bloom development in the Gulf of Finland (Baltic Sea) [J]. Hydrobiologia, 2008, 614(1): 133–140.
- [6] Tan X, Kong F X, Cao H S, et al. Recruitment of bloom-forming cyanobacteria and its driving factors[J]. African Journal of Biotechnology, 2008, 7(25): 4726–4731.
- [7] Wallace B B, Hamilton D P. Simulation of vertical position of buoyancy regulating *Microcystis aeruginosa* in a shallow eutrophic lake[J]. Aquatic Sciences, 2000, 62(4): 320–333.
- [8] Paerl H W, Ustach J F. Blue-green algal scums: an explanation for their occurrence during freshwater blooms[J]. Limnology and Oceanography, 1982, 27(2): 212–217.
- [9] 孔繁翔, 高 光. 大型浅水富营养化湖泊中蓝藻水华形成机理的思考[J]. 生态学报, 2005, 25(3): 589–595.
- [10] Vanni M J, Temte J. Seasonal patterns of grazing and nutrient limitation of phytoplankton in a eutrophic lake[J]. Limnology and Oceanography, 1990, 35(3): 697–709.
- [11] Lynch M, Shapiro J. Predation, enrichment, and phytoplankton community structure[J]. Limnology and Oceanography, 1981, 26(1): 86–102.
- [12] 胡鸿钧. 水华蓝藻生物学[M]. 北京: 科学出版社, 2011: 146.
- [13] Gibson G, Carlson R, Simpson J, et al. Nutrient criteria, technical guidance manual: lakes and reservoirs[R]. Washington DC: United States Environmental Protection Agency, 2000.
- [14] Bulgakov N G, Levich A P. The nitrogen:Phosphorus ratio as a factor regulating phytoplankton community structure[J]. Archive für Hydrobiologia, 1999, 146(1): 3–22.
- [15] Sullivan C, Saunders J, Welsh D. Phytoplankton of the river murray [R]. Canberra: Murray – Darling Basin Commission (Australia), 1988.
- [16] 徐恒省, 翁建中, 李继影, 等. 太湖蓝藻水华预警监测与风速风向的关系研究[J]. 环境监控与预警, 2009, 1(2): 5–7.
- [17] 武胜利, 刘 诚, 孙 军, 等. 卫星遥感太湖蓝藻水华分布及其气象影响要素分析[J]. 气象, 2009, 35(1): 18–23.
- [18] 王铭伟, 徐启新, 车 越, 等. 淀山湖蓝藻水华暴发的气象水文因素探讨[J]. 华东师范大学学报: 自然科学版, 2011(1): 21–31.
- [19] Spencer C N, King D L. Regulation of blue-green algal buoyancy and bloom formation by light, inorganic Nitrogen, CO<sub>2</sub>, and trophic level interactions[J]. Hydrobiologia, 1987, 144(2): 183–191.
- [20] Mur L R, Skulberg O M, Utkilen H. Cyanobacteria in the environment[M], 1999.
- [21] 赵孟绪, 韩博平. 汤溪水库蓝藻水华发生的影响因子分析[J]. 生态学报, 2005, 25(7): 1554–1561.
- [22] Steinberg C E W, Hartmann H M. Plankton bloom-forming cyanobacteria and the eutrophication of lakes and rivers[J]. Freshwater Biology, 1988, 20(2): 279–287.
- [23] Visser P M, Ibelings B W, Veer B, et al. Artificial mixing prevents nuisance blooms of the cyanobacterium *Microcystis* in Lake Nieuwe Meer, the Netherlands [J]. Freshwater Biology, 1996, 36(2): 435–450.
- [24] Blomqvist P, Pettersson A, Hyenstrand P. Ammonium-nitrogen: a key regulatory factor causing dominance of non-nitrogen-fixing cyanobacteria in aquatic system[J]. Archive für Hydrobiologia, 1994, 132(2): 141–164.
- [25] 许秋瑾, 秦伯强, 陈伟民, 等. 太湖藻类生长模型研究[J]. 湖泊科学, 2001, 13(2): 149–157.
- [26] Maier H R, Dandy G C, Burch M D. Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia[J]. Ecological Modeling, 1998, 105(2/3): 257–272.
- [27] 刘静玲, 曾维华, 曾 勇, 等. 海河流域城市水系优化调度[M]. 北京: 科学出版社, 2008.
- [28] NOAA. Algal Bloom Forecasting System[OL]. [2013-10-31]. <http://tidesandcurrents.noaa.gov/hab/>.
- [29] 陈 翔, 禹继华, 刘 杰, 等. 洪泽湖蓝藻暴发的气象条件分析[J]. 安徽农业科学, 2010, 38(15): 8141–8142, 8177.
- [30] Reynolds C S. The ecology of the planktonic blue-green algae in the North Shropshire meres[J]. Field Studies, 1971, 3: 409–432.
- [31] Wei B, Sugiura N, Maekawa T. Use of artificial neural network in the prediction of algal blooms[J]. Water Research, 2001, 35(8): 2022–2028.
- [32] Rehnagel F. ANNA – Artificial neural network model for predicting species abundance and succession of blue-green algae[J]. Hydrobiologia, 2004, 349(1/2/3): 47–57.
- [33] Oliver R L, Ganf G G. Freshwater blooms. In: Whitton[C]. Netherlands: Kluwer Academic Publishers, 2000: 149–194.
- [34] 周立国, 冯学智, 王春红, 等. 太湖蓝藻水华的 MODIS 卫星监测[J]. 湖泊科学, 2008, 20(2): 203–207.
- [35] Reynolds C S, Irish A E, Elliott J A. The ecological basis for simulating phytoplankton responses to environmental change (PROTECH) [J]. Ecological Modelling, 2001, 140(3): 271–291.
- [36] Australian and New Zealand Environment and Conservation Council. Australian water quality guidelines for fresh and Marine waters, National water quality management strategy[R]. Canberra: Australian and New Zealand Environment and Conservation Council, 1992.
- [37] 张大维, 刘 博, 刘 琪. EViews 数据统计与分析教程[M]. 北京: 清华大学出版社, 2010: 160–161.
- [38] 顾苏莉, 陈 方, 孙将陵. 太湖蓝藻监测及暴发情况分析[J]. 水资源保护, 2011, 27(3): 28–32.
- [39] 李旭文. Landsat-7 SLC-OFF ETM 遥感数据下载及在太湖蓝藻水华监测中的应用[J]. 环境监测管理与技术, 2009, 21(3): 54–57.
- [40] 汪昌云, 戴稳胜, 张成思. 基于 EViews 的金融计量学[M]. 北京: 中国人民大学出版社, 2011: 59.