

黄河内蒙古段流凌预报可变模糊聚类循环迭代模型

冀鸿兰, 卞雪军, 徐 晶

(内蒙古农业大学水利与土木建筑工程学院, 内蒙古 呼和浩特 010018)

摘要: 为了模拟黄河内蒙古段冰情的发生、发展过程, 以可变模糊集理论为基础, 提出了可变模糊聚类循环迭代模型来预报黄河内蒙古段巴彦高勒站流凌日期。选定初始权重矩阵、初始模糊聚类矩阵、初始模糊聚类中心矩阵, 应用迭代公式求得最优权重矩阵、最优模糊聚类矩阵、最优模糊聚类中心矩阵, 建立符合相关性要求的回归方程, 从而实现预报识别检验。将模型预报结果与实测值对比分析, 预报结果合格率达到80%, 属于甲等预报方案, 表明基于物理成因分析和数学理论的可变模糊聚类循环迭代模型可用于黄河内蒙古段的流凌预报。

关键词: 冰凌; 可变模糊聚类循环迭代模型; 流凌预报; 黄河

中图分类号: P338⁺.4 **文献标志码:** A **文章编号:** 1006-7647(2013)04-0014-04

Variable fuzzy clustering loop iteration model for ice-run forecast in Inner Mongolia reach of Yellow River//JI Honglan, BIAN Xuejun, XU Jing (School of Water Conservancy & Civil Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: In order to simulate the initiation and development of ice regimes in the Inner Mongolia reach of Yellow River, a variable fuzzy clustering loop iteration model is proposed to forecast the ice-run date at Bayangaole station in the Inner Mongolia Reach of Yellow River based on the variable fuzzy set theory. The matrices of the initial weight, the initial fuzzy clustering and the initial fuzzy clustering center are chosen, and their optimal matrices are obtained by means of the loop iteration formulae. Regression equations to satisfy the correlation are established to achieve forecast examination. The forecast results are compared with the measured values, and the acceptability is 80%. The forecast schemes belong to the first-class ones. It indicates that the proposed variable fuzzy clustering loop iteration model based on physical factors and mathematical theories can be employed to predicate the ice-run date in the Inner Mongolia reach of Yellow River.

Key words: ice; variable fuzzy clustering loop iteration model; ice-run forecast; Yellow River

凌汛是河道里的冰凌对水流的阻力作用而引起的一种涨水现象。凌汛灾害是高纬度寒冷地区江河普遍存在的一种自然现象。黄河凌汛主要发生在石嘴山至头道拐的内蒙古河段和下游的山东河段。黄河内蒙古段几乎每年发生凌汛, 这是由于它特殊的地理位置、水文气象条件、河道特性决定的。首先, 内蒙古段地处黄河流域最北端, 海拔高程千米以上, 河道流向自南向北, 其上游甘肃、宁夏段气温高, 内蒙古段气温低, 故河道流凌封冻由内蒙古段溯源而上, 解冻开河则由甘肃、宁夏段逐渐向下发展; 其次, 内蒙古段处于内陆地带, 离海洋距离远, 暖湿气流难以到达, 常为蒙古高压控制, 呈典型的大陆性气候, 年降水量少, 夏季盛热短暂, 冬季严寒漫长, 气温在零度以下的时间持续四五个月, 最低气温可达-35℃, 河流

结冰期也长达四五个月, 大部分为稳定封冻河段; 再次, 内蒙古段呈几字形大弯曲, 其间浅滩弯道迭出, 坡度变缓, 逶迤曲折, 河中多夹心滩, 分流窄沟, 解冻开河时, 常在河道的曲湾或由宽变窄的狭窄段产生凌汛。黄河上游的内蒙古河段每年冬春季节都有不同程度的冰凌灾害发生, 较大范围的淹没损失平均两年一次。因此提供及时有效的冰情预报, 对防凌决策、防灾减灾及国民经济和社会的可持续发展都具有十分重要的意义。

冰情演化复杂多变, 每年差别很大, 处于动态变化中, 具有不确定性、非线性等特征^[1], 基于经验的传统预报方法已无法满足现实要求, 数学预报模型是未来高精度冰情预报发展的方向。目前中长期冰情预报数学模型主要有模糊模式识别神经网络预报

基金项目: 国家自然科学基金(51060001); 内蒙古自然科学基金(2010MS0618)
作者简介: 冀鸿兰(1970—), 女, 内蒙古呼和浩特人, 教授, 博士, 主要从事水文水资源与工程模糊集研究。E-mail: honglanji@sina.com
通信作者: 卞雪军(1987—), 男, 内蒙古赤峰人, 硕士研究生, 主要从事水文水资源与工程模糊集研究。E-mail: bianxuejun999@163.com

模型^[2]、基于遗传算法的模糊优选神经网络 BP 模型^[3]、投影寻踪回归模型^[4]、基于支持向量机 SVR 的黄河凌汛预报模型^[5]等人工智能模型以及偏最小二乘回归冰情预报模型^[6]、多元线性回归预报模型^[7]等线性回归模型。为了更好地反映流凌过程的渐变性及模糊性,本文以可变模糊集理论为基础^[8-10],提出可变模糊聚类循环迭代模型,并应用于黄河内蒙古段巴彦高勒站的流凌日期预报。

1 可变模糊聚类循环迭代模型

1.1 模型建立

设有待聚类的 n 个样本组成的集合 $\{x_1, x_2, \dots, x_n\}$, 用 m 个指标特征值向量 $(x_{1j}, x_{2j}, \dots, x_{mj})$ 对样本进行聚类, 则有指标特征值矩阵:

$$X = (x_{ij}) \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$$

式中: x_{ij} 为样本 j 指标 i 的特征值。

由于 m 个指标特征值物理量纲不同, 需对指标特征值进行归一化, 得到指标特征值归一化矩阵:

$$R = (r_{ij}) \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$$

设 n 个样本依据 m 个指标特征值按 c 个类别进行聚类, 则模糊聚类矩阵为

$$U = (u_{hj}) \quad (h = 1, 2, \dots, c; j = 1, 2, \dots, n)$$

式中: u_{hj} 为样本 j 隶属于类别 h 的相对隶属度, 满足条件:

$$\sum_{h=1}^c u_{hj} = 1 \quad 0 \leq u_{hj} \leq 1 \quad \sum_{j=1}^n u_{hj} > 0$$

设类别 h 的 m 个指标特征值归一化数为 h 类的聚类中心, 则 c 个类别的模糊聚类中心矩阵为

$$S = (s_{ih}) \quad (i = 1, 2, \dots, m; h = 1, 2, \dots, c)$$

式中: s_{ih} 为类别 h 指标 i 的聚类中心归一化数, $0 \leq s_{ih} \leq 1$ 。

考虑不同指标对聚类的影响不同, 引入指标权向量:

$$W = (w_1, w_2, \dots, w_m)$$

满足 $\sum_{i=1}^m w_i = 1, 0 \leq w_i \leq 1$ 。

初始指标权向量由样本与预报因子的相关系数的绝对值的归一化数确定:

$$W = \left(\frac{|\rho_1|}{\sum_{i=1}^m |\rho_i|}, \frac{|\rho_2|}{\sum_{i=1}^m |\rho_i|}, \dots, \frac{|\rho_m|}{\sum_{i=1}^m |\rho_i|} \right)$$

样本 j 与类别 h 之间的差异用广义欧氏权距离表示:

$$d_{hj} = \sqrt{\sum_{i=1}^m [w_i(r_{ij} - s_{ih})]^2}$$

样本 j 隶属于类别 h 的相对隶属度为 u_{hj} , 故有

加权广义欧氏权距离:

$$D_{hj} = u_{hj} d_{hj}$$

样本 j 与 c 个类别的差异综合权衡度量为

$$f_j(u_j, s, w) = \sum_{h=1}^c \left\{ u_{hj} \sqrt{\sum_{i=1}^m [w_i(r_{ij} - s_{ih})]^2} \right\}^2$$

则有

$$f(u, s, w) = (f_1(u_1, s, w), f_2(u_2, s, w), \dots, f_n(u_n, s, w))$$

建立目标函数

$$\min \{f(u, s, w) = (f_1(u_1, s, w), f_2(u_2, s, w), \dots, f_n(u_n, s, w))\}$$

由于样本集里各个样本之间公平竞争, 没有任何偏好关系, 故上述目标函数可用等权重的线性加权平均法集结为单目标优化问题:

$$\min \{F(U, s, w) = \sum_{j=1}^n f_j(u_j, s, w)\}$$

构造拉格朗日函数并求解, 得可变模糊聚类循环迭代模型:

$$w_i = \left\{ \frac{\sum_{h=1}^c [u_{hj}(r_{ij} - s_{ih})]^2}{\sum_{h=1}^c [u_{hj}(r_{kj} - s_{kh})]^2} \right\}^{-1} \quad (1)$$

$$u_{hj} = \frac{\sum_{i=1}^m [w_i(r_{ij} - s_{ih})]^2}{\sum_{i=1}^m [w_i(r_{ij} - s_{ik})]^2} \quad (2)$$

$$s_{ih} = \frac{\sum_{j=1}^n u_{hj}^2 w_i^2 r_{ij}}{\sum_{j=1}^n u_{hj}^2 w_i^2} \quad (3)$$

1.2 模型求解步骤

a. 给定聚类 c 及迭代计算精度 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 。

b. 设初始权重矩阵 $W^{(l)} = (w_i^{(l)})$ 、初始模糊聚类矩阵 $U^{(l)} = (u_{hj}^{(l)})$ 、初始模糊聚类中心矩阵 $S^{(l)} = (s_{ih}^{(l)})$, $l=0$ 。

c. 用公式(1)、(2)、(3)分别计算 $w_i^{(l+1)}$ 、 $u_{hj}^{(l+1)}$ 、 $s_{ih}^{(l+1)}$ 。

d. 如果满足 $\max |w_i^{(l+1)} - w_i^{(l)}| \leq \varepsilon_1$, $\max |u_{hj}^{(l+1)} - u_{hj}^{(l)}| \leq \varepsilon_2$, $\max |s_{ih}^{(l+1)} - s_{ih}^{(l)}| \leq \varepsilon_3$, 则迭代结束, $(w_i^{(l+1)})$ 、 $(u_{hj}^{(l+1)})$ 、 $(s_{ih}^{(l+1)})$ 为最优指标权重矩阵 W^* 、最优模糊聚类矩阵 U^* 、最优模糊聚类中心矩阵 S^* ; 否则令 $l+1=l$, 转入步骤 c 继续迭代。

e. 用公式(4)计算级别特征值 H 。

$$H = (1, 2, \dots, c) \begin{bmatrix} u_{11}^* & \dots & u_{1n}^* \\ \vdots & & \vdots \\ u_{c1}^* & \dots & u_{cn}^* \end{bmatrix} = (H_1, H_2, \dots, H_n) \quad (4)$$

检验样本特征值与级别特征值的相关性,要求相关系数 $\rho>0.9$ 。

f. 用回归方程(5)推求预报值。

$$Y-\bar{Y}=r\frac{\sigma_y}{\sigma_H}(H-\bar{H}) \tag{5}$$

式中: σ_y 、 σ_H 分别为 Y 与 H 的均方差。

2 模型应用

巴彦高勒站是黄河内蒙古段的入境站,研究资料系列选取 1970—1971 年度至 2007—2008 年度,确定预报因子 x_1 为气温转负日期至 12 月 1 日逐日累计气温和、 x_2 为气温转负日期至 12 月 1 日平均流量、 x_3 为 11 月 1 日至气温转负日的天数,预报对象 y 为 11 月 1 日至流凌日期的天数。根据可变模糊聚类理论所必要的资料数量,并考虑预报识别检验的需要,将研究系列中前 33 个年度资料用于确定模型参数,后 5 个年度资料用于预报识别检验。

计算各预报因子的相关系数: $\rho_1=0.56$, $\rho_2=0.16$, $\rho_3=0.72$ 。

将指标特征值矩阵 $X=(x_{ij})$ 进行归一化处理,得指标特征值归一化矩阵:

$$R=(r_{ij})$$

该矩阵为一个 3 行 33 列的矩阵,对应 3 个预报因子 33 个年度的归一化值,其第 1 行数据为 0.072, 0.681, 0.495, 0.887, 0.919, 0.643, 0.135, 0.91, 0.474, 0.319, 0.929, 0.861, 0.608, 0.729, 0.703, 0.688, 0.277, 0.565, 0.811, 0.527, 0.711, 0.824, 0.754, 0.000, 1.000, 0.73, 0.464, 0.592, 0.966, 0.68, 0.444, 0.763, 0.814;第 2 行数据为 0.370, 0.370, 0.370, 0.416, 0.515, 1.000, 0.039, 0.302, 0.584, 0.335, 0.231, 0.225, 0.519, 0.803, 0.287, 0.440, 0.076, 0.197, 0.197, 0.567, 0.458, 0.388, 0.431, 0.247, 0.536, 0.379, 0.146, 0.000, 0.336, 0.314, 0.331, 0.413, 0.255;第 3 行数据为 0.292, 0.583, 0.375, 0.708, 0.792, 0.458, 0.125, 0.833, 0.333, 0.167, 0.708, 1.000, 0.292, 0.375, 0.500, 0.625, 0.292, 0.833, 0.667, 0.083, 0.542, 0.000, 0.042, 0.375, 0.792, 0.500, 0.208, 0.375, 1.000, 0.750, 0.000, 0.750, 0.333。

令 $\varepsilon_1=\varepsilon_2=\varepsilon_3=10^{-4}$,初始指标权向量 $W=(0.39,0.11,0.50)$,应用可变模糊聚类循环迭代模型公式(1)~(3)确定最优指标权重矩阵 W^* 、最优模糊聚类矩阵 U^* 、最优模糊聚类中心矩阵 S^* :

$$W^*=(0.40,0.20,0.40)$$

$$U^*=(u_{hj}^*)$$

$$S^*=\begin{bmatrix}0.470&0.657&0.662\\0.242&0.384&0.387\\0.387&0.495&0.487\end{bmatrix}$$

U^* 也为一个 3 行 33 列的矩阵,其第 1 行数据为 0.310, 0.020, 0.311, 0.124, 0.144, 0.238, 0.163, 0.167, 0.255, 0.400, 0.185, 0.198, 0.135, 0.193, 0.088, 0.040, 0.653, 0.377, 0.193, 0.259, 0.037, 0.192, 0.210, 0.188, 0.165, 0.027, 0.771, 0.220, 0.188, 0.179, 0.543, 0.205, 0.263;第 2 行数据为 0.308, 0.253, 0.269, 0.434, 0.350, 0.170, 0.114, 0.250, 0.372, 0.189, 0.300, 0.343, 0.390, 0.396, 0.309, 0.297, 0.175, 0.198, 0.405, 0.367, 0.211, 0.396, 0.403, 0.203, 0.300, 0.458, 0.298, 0.180, 0.406, 0.308, 0.408, 0.292, 0.392;第 3 行数据为 0.248, 0.394, 0.210, 0.327, 0.348, 0.290, 0.190, 0.342, 0.300, 0.186, 0.408, 0.243, 0.365, 0.461, 0.451, 0.321, 0.172, 0.350, 0.400, 0.470, 0.502, 0.482, 0.422, 0.214, 0.421, 0.409, 0.114, 0.273, 0.470, 0.330, 0.358, 0.457, 0.423。检验样本特征值与级别特征值的相关性,其相关系数 $\rho=0.91$,满足要求,则回归方程为

$$Y=21.50H-9.73$$

图 1 为流凌历时拟合图,可见拟合效果较好,该回归方程可用于巴彦高勒站的流凌预报。

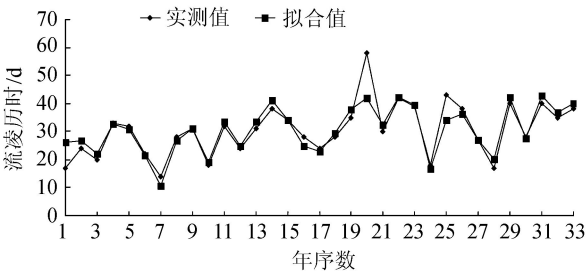


图 1 巴彦高勒站可变模糊聚类模型流凌历时拟合图

待预报识别检验对象的指标特征值归一化矩阵、相对隶属度矩阵、级别特征值向量分别为

$$R=\begin{bmatrix}0.734&0.779&0.887&0.841&0.904\\0.327&0.337&0.547&0.281&0.475\\0.000&0.500&0.542&0.792&0.792\end{bmatrix}$$

$$U^*=\begin{bmatrix}0.288&0.569&0.420&0.227&0.230\\0.136&0.110&0.267&0.665&0.662\\0.576&0.320&0.310&0.107&0.310\end{bmatrix}$$

$$H=(2.888,1.749,1.884,1.878,2.484)$$

代入回归方程,得后 5 个年度流凌历时的预报值:

$$Y=(39,26,31,31,44)$$

而后 5 个年度流凌历时的实测值为

$$Y_0=(36,26,34,37,43)$$

依据 GB/T 22482—2008《水文情报预报规范》并参照内蒙古水利厅防汛指挥办公室多年的防凌减灾实践,预见期为10 d,许可误差为3 d。分析表1的预报检验结果,合格率为80%,属于甲等预报方案。

表1 巴彦高勒站可变模糊聚类模型流凌日期预报值与实测值对比

年度	实测值	预报值	误差/d	合格否
2003—2004	12月6日	12月9日	3	合格
2004—2005	11月26日	11月26日	0	合格
2005—2006	12月4日	12月1日	3	合格
2006—2007	12月7日	12月1日	6	不合格
2007—2008	12月13日	12月14日	1	合格

3 结 语

本文将流凌历时作为一个相对模糊概念来研究,结合可变模糊聚类循环迭代模型的模糊聚类矩阵和模糊聚类中心矩阵,既考虑了各样本与最优聚类中心的差异,又考虑了指标权重和最优模糊矩阵的综合影响作用。利用1970—1971年度至2002—2003年度的实测资料,推测模糊聚类中心及不同指标的模糊聚类权重,根据模糊聚类中心与权重对2003—2004年度到2007—2008年度的流凌历时进行预报。预报结果准确率较高,对黄河内蒙古段的冰情预报有一定参考价值。

参考文献:

[1] 冯国华. 黄河内蒙古段冰凌特征分析及冰情信息模拟预报模型研究[D]. 呼和浩特:内蒙古农业大学,2009.
[2] 冀鸿兰,朝伦巴根,陈守煜. 基于模糊识别人工神经网络的冰凌预报模型[J]. 水力发电,2008,34(11):24-26. (JI Honglan,CHAOLUN Bagen,CHEN Shouyu. Fuzzy

recognition neural network model for ice forecasting[J]. Water Power,2008,34(11):24-26. (in Chinese))

[3] 陈守煜,冀鸿兰. 冰凌预报模糊优选神经网络BP方法[J]. 水利学报,2004(6):114-118. (CHEN Shouyu,JI Honglan. Fuzzy optimization neural network BP approach for ice forecasting[J]. Journal of Hydraulic Engineering, 2004(6):114-118. (in Chinese))
[4] 王志兴,李成振,陈刚. 冰情预报的投影寻踪回归模型[J]. 自然灾害学报,2009,18(5):174-177. (WANG Zhixing,LI Chengzhen,CHEN Gang. Projection pursuit regression model for ice situation forecast[J]. Journal of Natural Disasters,2009,18(5):174-177. (in Chinese))
[5] 李亚伟,陈守煜,韩小军. 基于支持向量机SVR的黄河凌汛预报方法[J]. 大连理工大学学报,2006,46(2):272-275. (LI Yawei,CHEN Shouyu,HAN Xiaojun. Yellow River ice flood prediction based on SVR[J]. Journal of Dalian University of Technology,2006,46(2):272-275. (in Chinese))
[6] 胡进宝. 黄河宁蒙段冰情中长期预报研究[D]. 南京:河海大学,2006.
[7] 张傲姐. 黄河内蒙古段冰情特点及预报模型研究[D]. 呼和浩特:内蒙古农业大学,2011.
[8] 陈守煜. 工程模糊集理论与应用[M]. 北京:国防工业出版社,1998.
[9] 陈守煜. 复杂水资源系统优化模糊识别理论与应用[M]. 长春:吉林大学出版社,2002.
[10] 陈守煜. 可变模糊集合理论与可变模型集[J]. 数学的实践与认识,2008,38(18):146-153. (CHEN Shouyu. Theory of variable fuzzy sets and variable model sets[J]. Mathematics in Practice and Theory,2008,38(18):146-153. (in Chinese))

(收稿日期:2012-12-19 编辑:熊水斌)



· 简讯 ·

第3届中国小水电论坛将于2013年下半年举办

由国际小水电联合会、中国水利学会水力发电专业委员会、中国水力发电工程学会小水电专业委员会主办,国际小水电中心、水利部农村电气化研究所承办的第3届中国小水电论坛将于2013年下半年举办。本届论坛将以“农村水电增效扩容改造”为主题,总结交流各地农村水电增效扩容改造试点的经验和做法,总结问题和不足,为后续改造工程的各参与方提供技术交流平台。论坛议题包括:①农村水电增效扩容改造建设管理;②农村水电增效扩容改造设计及施工安装技术;③农村水电增效扩容改造设备制造;④农村水电增效扩容改造新技术及应用;⑤农村水电增效扩容改造项目绩效评价;⑥农村水电增效扩容改造相关技术标准与工程应用;⑦农村水能资源规划与管理;⑧水电新农村电气化建设及小水电代燃料。

会议将出版论文集。详情可见 <http://www.hydropower.org.cn/info/shownews.asp?newsid=9569>。

(本刊编辑部供稿)