

# 多分支的降水量概率预测模型研究

余霖 吕鑫 周思琪 刘璇

(河海大学计算机与信息学院 南京 211100)

**摘要** 降水量大小对水资源调度决策、防汛防旱预警等方面有着决定性作用。目前已有大量降水量预测模型被提出,但其由于缺乏对降水过程非线性性态的考虑,因此预测准确度不高。另外,单独的预测值难以对决策判断形成有效支持,使预测结果的应用性不好。针对上述问题,基于降水量的平稳性及周期性,构建了同比分支及环比分支预测模型,进而提出了一种多分支的降水量概率预测模型 MBPPFM。该模型采用十字交叉选择算法,精细化筛选同比、环比分支预测结果,提高了预测准确性,并能避免异常预测。同时,预测结果包括区间概率和结果置信度,能有效支持决策形成。

**关键词** 降水量预测, ARMA 模型, BP 网络, Markov 预测, 关联规则, 十字交叉模型

**中图分类号** TP274 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.10.040

## Research of Multi-branch Precipitation Probability Forecasting Model

YU Lin LV Xin ZHOU Si-qi LIU Xuan

(College of Computer and Information, Hohai University, Nanjing 211100, China)

**Abstract** The size of precipitation plays a decisive role in aspects of water dispatching decision, early warning of flood and drought control, etc. Currently, many precipitation forecasting models have already been put forward. However, due to lack of the nonlinear characteristic of precipitation process consideration, the forecasting accuracy is not high. In addition, it is difficult to use a single forecast value to effectively support the judgment, leading to the fact of lower applicability results. Aimed at the above-mentioned problems, forecasting models of year-on-year branch and month-on-month branch were constructed based on the stationarity and periodicity of precipitation, and then a multi-branch precipitation probability forecasting model (MBPPFM) was proposed. The cross selection algorithm was used in the model to well screen the forecasting results from year-on-year branch and month-on-month branch. Finally, the forecasting accuracy is improved and abnormal forecasting can be avoided. At the same time, probability and confidence values are included in the forecasting results to effectively support decision making.

**Keywords** Precipitation forecasting, ARMA model, BP network, Markov forecasting, Association rules, Cross model

## 1 引言

大气降水是水循环的重要组成部分。科学有效地预测降水量对于加强区域水资源的宏观管理、优化调度和开发利用具有重要的指导意义。同时,作为最主要的自然灾害之一,暴雨等引起的洪水灾害会严重威胁人类的生命和财产安全。提高降水量预测的准确度,减少暴雨洪水灾害带来的损失,是我们需要解决的重大课题。然而,由于气候条件的复杂多变及降水过程自身的非线性和多重复杂性等原因,加上蒸发量、积雪量等各种要素的影响,降水量表现为一种随机变量,有很强的随机性、不确定性和预测不准确<sup>[1]</sup>。基于此,很难通过分

析物理成因和气象因素来准确预测一个时段内的降水量。因而,在对历史降水量数据进行统计、分析和计算的基础上形成降水量预测方法是目前国内外大多数的降水量预测研究的主流思路<sup>[2]</sup>。

传统的预测方法大多数假设降水量系统是线性的,导致预测结果准确度低,因此诞生了新的预测方法,例如田俊武<sup>[3]</sup>、李希国<sup>[4]</sup>、王声锋<sup>[5]</sup>、Yang L L<sup>[6]</sup>等引入时间序列回归模型(如 AR, MA, ARMA 和 ARIMA 模型等)对降水量进行预测;何慧<sup>[7]</sup>、刘莉<sup>[8]</sup>、张继学<sup>[9]</sup>等引入神经网络模型对降水量序列进行预测等。虽然上述方法在不同程度上给出了降水量预测的方式,较好地解决了降水量预测某些方面的问题,

到稿日期:2016-11-21 返修日期:2017-01-02 本文受NSFC-广东联合基金重点项目(U1301252),水利部公益性行业科研专项重点项目(201501007),国家自然科学基金面上项目(61272543),国家科技支撑计划项目(HNKJ13-H17-04)资助。

余霖(1991-),女,硕士,主要研究方向为计算机应用技术、数据挖掘, E-mail: yl@hhu.edu.cn; 吕鑫(1983-),男,博士,讲师, CCF 会员,主要研究方向为密码学、网络信息安全、大数据隐私保护; 周思琪(1994-),女,硕士生,主要研究方向为云计算、数据分析; 刘璇(1989-),女,博士生,主要研究方向为数据挖掘。

但是在预测的准确性和应用性方面仍需进一步改进。主要原因有以下两点:1)采用回归模型<sup>[10]</sup>、人工神经网络等方法进行预测时简单的线性拟合方程不能表现降水量的丰富繁杂规律,存在线性缺陷,导致结果准确度比较低<sup>[11-12]</sup>;2)使用小波分析<sup>[13]</sup>、随机分析法<sup>[14]</sup>等方法进行降水量预测时,都仅仅针对降水量的单一特点进行了改进,对预测准确性和应用性的改善有限。降水量序列既包含众多的时序动态特性,又包括环流、气压、地形等环境气象因子的影响,是一种典型的非线性且多维的时间序列。因此将多种数据分析方法进行融合,对降水量数据进行非线性建模,能提高降水量预测的准确性及应用性。

本文首先根据降水量的平稳性与周期性特点,分别在同比和环比两个时间分支上选择传统的方法进行预测;然后,对两个分支的预测结果分别进行模糊区间化处理,同时得到区间结果出现的概率以及结果置信度,决策者可以根据降水量概率和置信度进行判断,以提高预测结果的应用性;最后,针对两分支的预测结果区间,提出多分支的降水量概率预测模型 MBPPFM。该模型定义了交叉选择算法,对模糊区间化处理形式、维度一致的预测区间进行精细化筛选,得到了更为准确的降水量预测结果。

## 2 降水量多分支预测及修正

本节首先分析了降水量数据的平稳性与周期性特点,然后分别对同比和环比分支使用传统方法进行降水量预测。对传统方法得出的降水量预测结果进行误差修正处理,每个月的降水量预测值被修正为一组由概率状态表示的月降水量预测区间值。

### 2.1 月降水量数据的平稳性与周期性

本文所有实验数据均来自国家气象信息中心资料服务室提供的中国气象科学数据服务网上数据集。截止到撰稿本文时,该数据集的年完整数据只更新到 2014 年。本文选取河海大学江宁校区附近观测点的数据作为实验数据,使用 MATLAB 软件将 1961 年 1 月—2014 年 12 月共 54×12 条降水量数据绘制成三维图(见图 1 和图 2),以展现降水量数据在年间的平稳性和年内的周期性。由图 1 可以看出,每年 1 月—12 月的降水量变化趋势是相似的:第一季度和第二季度降水量较少,并呈现上升趋势;第三季度降水量较丰富,尤其在 7 月前后达到峰值,之后降水开始呈现下降趋势,直到年末。这体现了降水情况的年内周期性。由图 2 可见,在不同年份的同一个月份,降水量值趋于平稳,体现了年间平稳性。

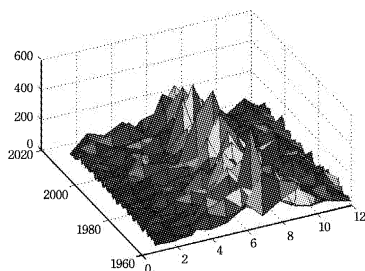


图 1 1961 年 1 月—2014 年 12 月降水量的折线图(从周期性角度)

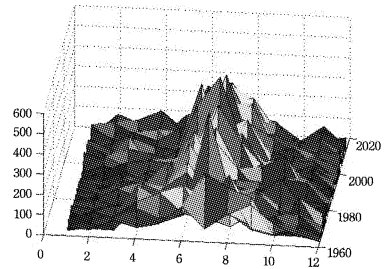


图 2 1961 年 1 月—2014 年 12 月降水量的折线图(从平稳性角度)

本文将分两个分支预测降水量,具体如下。

1)同比预测:将某年第  $n$  月与往年第  $n$  月进行对比。同比预测强调对不同年份同一月份的降水量进行比较。同比预测主要是为了消除季节变动的影响,用以说明本期降水量情况与往年同期降水量情况之间的关系。同比预测是基于降水量变化在年间所表现的平稳性。

2)环比预测:将某年第  $n$  月与同一年第  $n-1$  月进行比较。环比预测强调对同一年份相邻月份的降水量情况进行比较。环比预测主要考虑了季节变动的影响,用以说明本期降水量情况与上期降水量情况之间的关系。环比预测是基于降水量变化在年内所表现的周期性。

## 2.2 月降水量同比分支预测模型

### 2.2.1 平稳性识别

为预测 2014 年 8 月的降水量,首先绘制 1961—2013 年每年 8 月份的降水量数据时序图和自相关函数图,如图 3 和图 4 所示。由图无法判断降水量数据时间序列是否具有平稳性。

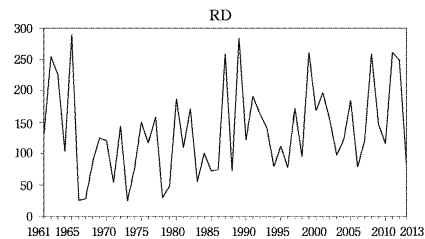


图 3 1961—2013 年每年 8 月份的降水量时序图

Sample: 1961 2013  
Included observations: 53

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.031	-0.031	0.0531	0.818
		2	0.138	0.137	1.1443	0.564
		3	0.082	0.092	1.5375	0.674
		4	-0.056	-0.072	1.7263	0.786
		5	-0.050	-0.082	1.8807	0.865
		6	-0.059	-0.052	2.0752	0.913
		7	0.140	0.174	3.3106	0.855
		8	-0.051	-0.018	3.4776	0.901
		9	-0.052	-0.110	3.6564	0.933
		10	0.206	0.184	6.5347	0.769
		11	-0.047	0.011	6.9901	0.824
		12	0.279	0.271	12.214	0.429
		13	-0.052	-0.091	12.414	0.494
		14	-0.187	-0.319	14.503	0.413
		15	0.033	0.050	14.589	0.482
		16	-0.228	-0.086	18.989	0.285
		17	0.067	0.108	19.069	0.325
		18	-0.066	-0.033	19.423	0.366
		19	0.084	-0.047	20.030	0.393

图 4 1961—2013 年每年 8 月份的降水量自相关函数图

在时序图和样本自相关函数图都无法判断时间序列是否具有平稳性时,进行 ADF 和 PP 单位根检验,检验结果如图 5 和图 6 所示。在给定显著性水平为 0.01 的情况下, $t$  统计量小于检验统计值。因此,ADF 检验和 PP 检验结果均证明该降水量序列是平稳的。

Augmented Dickey-Fuller Unit Root Test on RD		
Null Hypothesis: RD has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic - based on SIC, maxlag=10)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.257579	0.0000
Test critical values:	1% level	-3.582669
	5% level	-2.918778
	10% level	-2.597285

\*Mackinnon (1996) one-sided p-values.

图5 ADF 检验结果

Phillips-Perron Unit Root Test on RD		
Null Hypothesis: RD has a unit root		
Exogenous: Constant		
Bandwidth: 2 (Newey-West automatic) using Bartlett kernel		
	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-7.285293	0.0000
Test critical values:	1% level	-3.582669
	5% level	-2.918778
	10% level	-2.597285

\*Mackinnon (1996) one-sided p-values.

图6 PP 检验结果

### 2.2.2 平稳时间序列 ARMA 模型

根据降水量数据的自相关与偏相关函数,确定使用平稳时间序列 ARMA 模型。由自相关函数初步确定了预测 2014 年 8 月降水量的 ARMA 模型的参数:  $p=14, q=12$ 。通过  $t$  统计量检验了参数的估计值具有统计显著性,通过  $Q$  统计量和 LM 检验验证了残差序列的白噪声性。

### 2.2.3 误差修正

使用 ARMA 模型对 1961—2013 年每年 8 月的降水量数据进行预测,并与真实值进行比较,得到归一化相对误差折线图,如图 7 所示。由于使用的模型为 ARMA(12, 14),最高滞后阶数为 14,1961—1974 年的模型预测数据为空值,因此图 7 中只有后 39 组预测数据的误差分析结果。

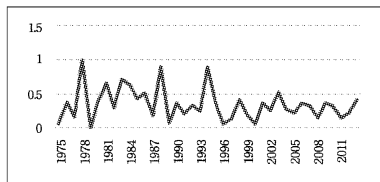


图7 ARMA 对 1975—2013 年每年 8 月降水量的预测归一化相对误差折线图

李世华等<sup>[15]</sup>引入 Markov 模型对降水状态和降雨量进行预测,证明了其具有一定可靠性。本文依此使用 Markov 模型预测出误差区间,模糊区间化 ARMA 预测结果。

由图 7 的数据分布规律可以将整个  $[0, 1]$  相对误差区间划分为 5 个区间:  $[0, 0.060]$ ,  $(0.060, 0.240]$ ,  $(0.240, 0.420]$ ,  $(0.420, 0.600]$  和  $(0.600, 1]$ 。根据上述分类可获得 1961—2013 年月降水量时间序列模型预测结果的 Markov 状态转移情况,并计算得到状态转移矩阵  $P$ :

$$P = \begin{bmatrix} 0 & 0.333 & 0.667 & 0 & 0 \\ 0.083 & 0.083 & 0.417 & 0.167 & 0.25 \\ 0.077 & 0.462 & 0.23 & 0.077 & 0.154 \\ 0 & 0.5 & 0.25 & 0.25 & 0 \\ 0.167 & 0.167 & 0.332 & 0.167 & 0.167 \end{bmatrix}$$

根据 Markov 链预测模型,可以得到 2014 年 8 月预测降水量误差状态向量为  $[0 \ 0.5 \ 0.25 \ 0.25 \ 0]$ 。利用状态向量对结果进行修正,得到 2014 年 8 月的预测降水量值,如表 1 所列。

表1 同比分支 Markov 修正结果分析表

预测月份	月降水量	ARMA 月降	同比分支月预测值/mm	
	实际值 /mm	水量预测值 /mm	最大概率预测区间	概率/%
2014 年 8 月	160.4	198.22	[123.8, 187.2)	50

由于降水量数据的非线性和复杂性,使用 ARMA 模型对降水量进行拟合,并给出预测的结果出现异常值的可能性。本文提出使用关联规则的方法对误差进行修正,通过样本中预测值和误差之间建立的关联关系,发现预测异常值,提出重新预测的建议。

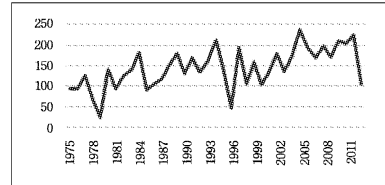


图8 ARMA 对 1975—2013 年每年 8 月降水量的预测值

根据图 8 可以看出,降水量预测值的分布具有随机性和均衡性,因此可以将降水量预测值等分,并进行关联规则初始项目区间划分。样本数据共 39 组,降水量 ARMA 预测值的最大值为 236.72mm,最小值为 24.82mm。根据预测值差值以及项目个数优化选择,确定将降水量 ARMA 预测值划分为 5 个区间,即对应 5 个项目:  $DI_1=[20, 64]$ ,  $DI_2=(64, 108]$ ,  $DI_3=(108, 152]$ ,  $DI_4=(152, 196]$  和  $DI_5=(196, 240]$ 。归一化误差状态区间划分方法仍采用 2.2.3 节中对误差区间使用的方法,即  $[0, 0.060]$ ,  $(0.060, 0.240]$ ,  $(0.240, 0.420]$ ,  $(0.420, 0.600]$  和  $(0.600, 1]$ 。将统一编号,将上述区间分别编号项目 ID 为  $EI_1-EI_5$ 。使用 Apriori 算法实现频繁项集和关联规则,分别设置 10% 和 5% 作为算法中一项集和二项集的支持度,得到生成各频繁二项集的关联规则及其对应的置信度:

- $DI_1 \rightarrow EI_1$  Confidence =  $2/2 = 100\%$
- $DI_2 \rightarrow EI_5$  Confidence =  $2/8 = 25\%$
- $DI_3 \rightarrow EI_2$  Confidence =  $2/11 = 18.2\%$
- $DI_3 \rightarrow EI_3$  Confidence =  $6/11 = 54.5\%$
- $DI_3 \rightarrow EI_4$  Confidence =  $2/11 = 18.2\%$
- $DI_4 \rightarrow EI_2$  Confidence =  $5/12 = 41.7\%$
- $DI_4 \rightarrow EI_3$  Confidence =  $5/12 = 41.7\%$
- $DI_4 \rightarrow EI_5$  Confidence =  $2/12 = 16.6\%$
- $DI_5 \rightarrow EI_2$  Confidence =  $4/6 = 66.7\%$

上述所有的关联关系构成降水量预测误差关联规则集  $S$ 。根据该关联规则集  $S$ ,可以对预测结果进行正向确认或反向拒绝,具体方法如下:

#### (1) 正向确认

将经过 Markov 修正的降水量预测值对应到降水量预测值区间,建立区间项与误差状态项之间的关联关系  $a$ ,如果  $a \in S$ ,即预测结果满足关联规则,则接受预测结果。

#### (2) 反向拒绝

如果  $a \notin S$ ,即预测结果不满足关联规则,则拒绝预测结果,给出重新选择 ARMA 模型参数和预测的建议。

对表 1 中 2014 年 8 月的降水量预测值进行关联规则修正。

预测的区间属于项 DI<sub>3</sub>,误差区间属于项 EI<sub>4</sub>. DI<sub>3</sub>→EI<sub>2</sub> 是关联规则集 S 中的关系,即预测结果得到正向确认,置信度为 18.2%。2014 年 8 月的降水量预测结果如表 2 所列。

表 2 2014 年 8 月降水量预测结果(同比分支)

预测月份	月降水量 实际值/mm	同比分支月预测值/mm		
		最大概率预测区间	概率/%	预测置信度/%
2014 年 8 月	160.4	[123.8,187.2)	50	18.2

由表 2 可以看出,月降水量实际值落在同比分支的预测区间内,说明预测结果有效;并且,同比分支预测给出了预测结果区间出现的概率以及预测置信度,决策者可以根据相关经验,在参考概率和预测置信度数值的情况下进行更加准确、有效的水资源分配以及暴雨灾害预警。

### 2.3 月降水量环比分支预测模型

#### 2.3.1 BP 神经网络月降水量预测

建立 BP 神经网络月降水量预测子模型的步骤主要分为网络层数设计、网络的输入和输出确定、输入输出数据归一化、网络参数的选择、网络函数的选择及网络的建立等。根据月降水量数据得到各个参数值,建立了如图 9 所示的 BP 神经网络月降水量预测模型。

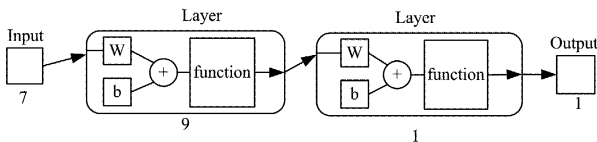


图 9 BP 神经网络月降水量预测模型

通过建立并实现 BP 神经网络月降水量预测子模型,得到了降水量在环比分支上的初步预测结果。

#### 2.3.2 误差修正

与 2.2.3 节类似,将整个 [0,1] 相对误差区间划分为 5 个区间:[0,0.21],[0.21,0.24],[0.24,0.27],[0.27,0.30]和 (0.30,1],得到 Markov 状态转移概率矩阵 P:

$$P = \begin{bmatrix} 0 & 0.571 & 0.286 & 0 & 0.143 \\ 0.111 & 0.389 & 0.333 & 0.167 & 0 \\ 0.2 & 0.2 & 0.4 & 0.067 & 0.133 \\ 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0.167 & 0.333 & 0 & 0.167 & 0.333 \end{bmatrix}$$

根据预测值差值及项目个数优化选择,确定将降水量 ARMA 预测值划分为 5 个区间,即对应 5 个项目:DI<sub>1</sub>=[30,80],DI<sub>2</sub>=(80,130],DI<sub>3</sub>=(130,180],DI<sub>4</sub>=(180,230]和 DI<sub>5</sub>=(230,280]。使用 Apriori 算法实现频繁项集和关联规则,分别设置 15%和 25%作为算法中一项集和二项集的支持度,得到生成各频繁二项集的关联规则及其对应的置信度:

- DI<sub>1</sub>→EI<sub>3</sub> Confidence=30.5%
- DI<sub>2</sub>→EI<sub>2</sub> Confidence=46.7%
- DI<sub>2</sub>→EI<sub>3</sub> Confidence=26.7%
- DI<sub>3</sub>→EI<sub>2</sub> Confidence=50.0%
- DI<sub>3</sub>→EI<sub>3</sub> Confidence=50.0%
- DI<sub>5</sub>→EI<sub>2</sub> Confidence=25.0%
- DI<sub>5</sub>→EI<sub>3</sub> Confidence=25.0%

由此得到 2014 年 8 月降水量的 BP-MA 模型预测结果,如表 3 所列。

表 3 2014 年 8 月降水量预测结果(环比分支)

预测月份	月降水量 实际值/mm	同比分支月预测值/mm		
		最大概率预测区间	概率/%	预测置信度/%
2014 年 8 月	160.4	[159.9,161.9)	40.0	50.0

### 3 多分支的降水量概率预测模型

对于特定的某年某月,上文将降水量数据先拆分成同比(时间轴横向)和环比(时间轴纵向)两个维度分支,选择各维度上具有预测优势的模型进行降水量预测,并且经过 Markov 和关联规则修正后,得到模糊区间化后的预测结果。在此基础上,提出多分支的降水量概率预测模型(Multi-Branch Precipitation and Probability Forecasting Model, MBPPFM),通过交叉交叉选择算法,实现纵、横向两个分支预测结果的合并,得到具有实际应用意义的降水量预测值及其概率和置信度。

在 MBPPFM 中,首先对搜集到的降水量数据进行预处理,选择某一地区的各年份全年月降水量值。将降水量值分别进行 ARMA 模型和 BP 模型降水量预测,再经过 Markov 和关联规则误差修正,对最后得到的值进行交叉交叉处理,得到 MBPPFM 的最终输出结果。具体流程如图 10 所示。

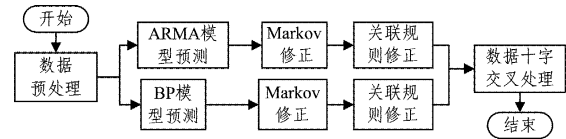


图 10 MBPPFM 流程图

#### (1)数据预处理

从中国气象科学数据服务网上获得中国地面降水月值 0.5°×0.5°格点数据集,使用 eclipse 进行 Java 编程,选择每个年月中对应观测点的值。输出的数据再经过 Excel 文本导入向导,获得符合要求的实验数据 Excel 表格。

#### (2)同比分支和环比分支预测

经过平稳性识别、模型识别、模型参数估计、模型检验和子模型预测得到 ARMA 模型降水量同比预测值;经过网络层数的设计、网络输入输出的确定、输入输出数据归一化、网络参数的选择、传递函数和学习函数的选择、BP 网络的建立和子模型的预测得到 BP 模型降水量环比预测值。

#### (3)Markov 误差修正

经过划分状态区间、计算状态转移矩阵和误差修正流程后,对子模型降水量预测值进行 Markov 误差修正,得到修正后的误差区间 a 和 b,以及预测出现该降水量区间的概率。

#### (4)关联规则误差修正

根据历史降水量预测误差值,划分降水量区间,建立降水量值与预测误差的关联规则,根据关联规则对降水量预测值进行正向或反向修正,从而得到降水量预测结果的置信度。

#### (5)数据交叉交叉处理

根据以上流程,可以获得 ARMA-MA 模型的降水量预测值 a、概率 ap、置信度 ac 和 BP-MA 模型的降水量预测值 b、概率 bp、置信度 bc。将这些值作为本流程的输入值,输出值即为 MBPPFM 的最终输出结果。具体流程如图 11 所示。

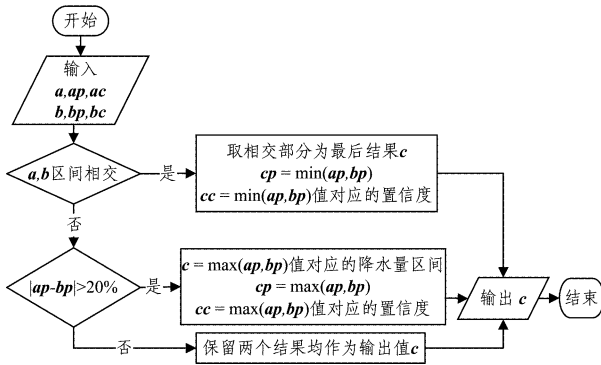


图 11 数据交叉交叉处理流程图

### 4 实验结果与分析

根据表 4 和交叉交叉处理算法,  $a, b$  区间相交, 选择相交区间为  $c = b = [159.9, 161.9]$  (区间预测概率  $cp = bp = 40.0\%$ , 预测置信度  $cc = bc = 50.0\%$ ). 由此可以得到 MBP-PPFM 对 2014 年 8 月的降水量预测结果为  $[159.9, 161.9]$  (单位为 mm), 区间预测概率为  $40.0\%$ , 预测置信度为  $50.0\%$ . 实际上 2014 年 8 月的降水量值为  $160.4\text{mm}$ , 预测区间包含了实际值, 表明预测结果准确. 此外, 与传统预测方法相比, 本模型预测方法所得的预测结果增加了预测概率值和置信度两个参数, 决策者可以根据这两个参数和经验判断该降水量预测结果实际出现的可能性, 作出水资源调度及洪水预警等相关的决策. 所提方法提高了预测结果的应用性.

表 4 2014 年 8 月降水量各分支预测结果

预测模型	实际值/ mm	传统方法 的预测值/mm	分支误差修正值		
			最大概率预测 区间/mm	概率/ %	置信度/ %
同比分支	160.4	198.22	[123.8, 187.2)	50.0	18.2
环比分支	160.4	160.09	[159.9, 161.9)	40.0	50.0

使用本文提出的 MBP-PPFM 模型对其他月份的降水量进行预测的效果均比较理想, 以预测 2014 年 7 月和 2013 年 8 月为例, 预测过程及结果数据如表 5 和表 6 所列.

表 5 2014 年 7 月降水量预测过程及结果

预测模型	实际值/ mm	传统方法的预测值/ mm	最大概率预测 区间/mm	概率/ %	置信度/ %
同比分支	305.7	245.7	[243.9, 556.7)	50.0	40.0
环比分支	305.7	279.4	[234.9, 308.3)	55.6	66.7
MBP-PPFM	305.7	245.7/279.4	[243.9, 308.3)	50.0	40.0

表 6 2013 年 8 月降水量预测过程及结果

预测模型	实际值/ mm	传统方法的预测值/ mm	最大概率预测 区间/mm	概率/ %	置信度/ %
同比分支	115.1	119.22	[87.4, 119.8)	42.9	63.6
环比分支	115.1	114.84	[114.7, 115.3)	50.0	38.5
MBP-PPFM	115.1	119.22/114.84	[114.7, 115.3)	42.9	63.6

本文实验在进行 Markov 和关联规则修正时按照数据特征情况进行状态划分, 该划分方法在考虑两端非频繁值的同时将频繁密集部分平均分为数个区间, 这符合较常用的统计学分类规则. 但因实验数据较少, 造成生成 Markov 转移矩阵和关联规则时频繁项与低频项的统计值差距不明显, 进而导致预测概率和置信度的百分比不具有直观优势. 但是, 对于 5 个状态转移百分比加和为 1 的情况, 1 个状态的概率

为  $40\%$ , 其他 4 个状态的平均概率只有  $15\%$ , 足以说明  $40\%$  是较高的统计概率, 如图 12 所示.

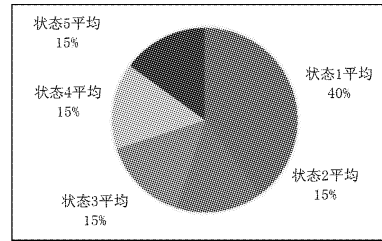


图 12 状态转移概率

**结束语** 本文通过研究基于数据挖掘技术的降水量预测方法, 提出降水量预测交叉交叉模型 MBP-PPFM, 对降水量预测的结果进行了改进, 提高了降水量预测结果的准确性和应用性. 本文在同比和环比预测分支上, 分别选用了平稳时间序列 ARMA 模型和 BP 神经网络模型, 在发展降水量预测模型的同时, 可以根据降水量的平稳性和周期性寻找其他新预测模型进行尝试, 找到分支上更加具有优势的模型.

### 参考文献

- [1] BI Z L, ZHANG Z Y, ZHU X G, et al. Precipitation Predicting Based on Improved RBF Neural Network and Markov Model [J]. Water Saving Irrigation, 2010(11): 1-3. (in Chinese) 闭祖良, 张展羽, 朱新国, 等. 基于 RBF 神经网络马尔可夫模型的降水量预测[J]. 节水灌溉, 2010(11): 1-3.
- [2] ZHANG A. Henan Zhaokou Irrigation Management System Based on WebGIS [D]. Zhengzhou: Zhengzhou University, 2012. (in Chinese) 张昂. 基于 WebGIS 的河南省赵口灌区管理信息系统[D]. 郑州: 郑州大学, 2012.
- [3] TIAN J W, SHANG S H, SUN Y L, et al. Stochastic characteristics of reference evapotranspiration and precipitation of Xiaohu Irrigation Areas, Shanxi Province [J]. Transactions of the CSAE, 2005, 21(10): 26-30. (in Chinese) 田俊武, 尚松浩, 孙艳丽, 等. 山西潇河灌区参考作物腾发量和降水的随机特性[J]. 农业工程学报, 2005, 21(10): 26-30.
- [4] LI X G, LIU X Z. Study on ARIMA stochastic model for precipitation in Yantai Region [J]. Journal of Water Resources and Water Engineering, 2006, 17(2): 505-510. (in Chinese) 李希国, 刘贤赵. 烟台地区降水量的 ARIMA 随机模型研究[J]. 水利科技与经济, 2006, 17(2): 505-510.
- [5] WANG S F, ZHANG Z Y, DUAN A W, et al. Time Sequence Characteristic Analysis of Precipitation in North Area of Henan Province [J]. China Rural Water and Hydropower, 2008(3): 13-16. (in Chinese) 王声锋, 张展羽, 段爱旺, 等. 豫北地区降水的时间序列特性分析[J]. 中国农村水利水电, 2008(3): 13-16.
- [6] YANG L L, LU W X. The Application of Time Series Analysis in Precipitation Forecast in Wuyuan County [C] // 2011 International Symposium on Water Resource and Environmental Protection (ISWREP). IEEE, 2011: 3063-3065.
- [7] HE H, JIN L, QIN Z N, et al. Downscaling Forecast of Monthly Precipitation over Guangxi Based on BP Neural Network Model [J]. Journal of Tropical Meteorology, 2007, 13(1): 97-100.
- [8] LIU L, YE W. Precipitation prediction of time series model based on BP artificial neural network [J]. Journal of Water Resources

- and Water Engineering, 2010, 21(5):156-159. (in Chinese)
- 刘莉,叶文. 基于 BP 神经网络时间序列模型的降水量预测[J]. 水资源与水工程学报, 2010, 21(5):156-159.
- [9] ZHANG J X, WANG P, ZHANG L, et al. Application of artificial neural network in short term precipitation forecast [J]. Technology Wind, 2016(17):123-124. (in Chinese)
- 张继学,王鹏,张琳,等. 神经网络在短期降水预测方面的应用研究[J]. 科技风, 2016(17):123-124.
- [10] CHI Z X, BAI H. The Study on Short-term Climatic Forecast Way in Southeast Guizhou [J]. Desert and Oasis Meteorology, 2005, 28(6):20-21. (in Chinese)
- 池再香,白慧. 黔东南地区短期气候预测方法研究[J]. 沙漠与绿洲气象, 2005, 28(6):20-21.
- [11] LU Z Y, YANG L, ZHAO Z C, et al. A Field Feature Extraction Method of Sand-dust Storm Ensemble Forecast System Based on ANN [J]. Computer Simulation, 2007, 24(6):341-344. (in Chinese)
- 路志英,杨乐,赵智超,等. 沙尘暴综合预报系统中场特征提取方法的研究[J]. 计算机仿真, 2007, 24(6):341-344.
- [12] HUANG J P. Research on PID Controller Based on BP Neural Network [J]. Computer Simulation, 2010, 27(7):167-170. (in Chinese)
- 黄剑平. 基于 BP 神经网络的 PID 控制研究[J]. 计算机仿真, 2010, 27(7):167-170.
- [13] CAO Y Q, HOU W P. Application Research of Non-linear Theory in Hydrology and its Prospect [J]. Water Power, 2005, 31(4):14-17. (in Chinese)
- 曹永强,侯文萍. 非线性理论在水文学中的应用研究及展望[J]. 水力发电, 2005, 31(4):14-17.
- [14] QING H. Exact Distribution Theory of Runs [J]. Chinese Journal of Applied Probability and Statistics, 1999, 15(2):199-212.
- [15] 李世华. 基于马尔可夫模型的澜沧县降雨状态及降水量预测应用 [C] // 云南省水利学会 2015 年度学术年会论文集. 昆明: 云南省水利学会, 2015:688-692.
- (上接第 221 页)
- zation [J]. Computational Intelligence Magazine IEEE, 2006, 1(4):28-39.
- [4] WOLPERT D H, MACREARY W G. No free lunch theorems for optimization [J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1):67-82.
- [5] YANG X, HOSSEIN GANDOMI A. Bat algorithm: a novel approach for global engineering optimization [J]. Engineering Computations, 2012, 29(5):464-483.
- [6] RASHEDI E, NEZAMABADI-POUR H, SARAZDI S. GSA: a gravitational search algorithm [J]. Information Sciences, 2009, 179(13):2232-2248.
- [7] OFTADEH R, MAHJOOB M J, SHARIATPANAH M. A novel meta-heuristic optimization algorithm inspired by group hunting of animals: Hunting search [J]. Computers & Mathematics with Applications, 2010, 60(7):2087-2098.
- [8] SIMON D. Biogeography-based optimization [J]. IEEE Transactions on Evolutionary Computation, 2008, 12(6):702-713.
- [9] ZHENG Y J. Water wave optimization: a new nature-inspired metaheuristic [J]. Computers & Operations Research, 2015, 55:1-11.
- [10] WU X B, LIAO J, WANG Z C. Water Wave Optimization for the Traveling Salesman Problem [M] // Intelligent Computing Theories and Methodologies. Springer International Publishing, 2015:137-146.
- [11] YANG F, HU C P, YAN X F. Particle swarm optimization algorithm of self-adaptive parameter based on ant system and its application [J]. Control Theory & Applications, 2010, 27(11):1479-1488. (in Chinese)
- 杨帆,胡春平,颜学峰. 基于蚁群系统的参数自适应粒子群算法及其应用 [J]. 控制理论与应用, 2010, 27(11):1479-1488.
- [12] YANG X, YUAN J, et al. A modified particle swarm optimizer with dynamic adaptation [J]. Applied Mathematics and Computation, 2007, 189(2):1205-1213.
- [13] PANIGRAHI B K, PANDI V R, DAS S. Adaptive particle swarm optimization approach for static and dynamic economic load dispatch [J]. Energy Conversion and Management, 2008, 49(6):1407-1415.
- [14] NICKABADI A, EBADZADEH M M, SAFABAKHSH R. A novel particle swarm optimization algorithm with adaptive inertia weight [J]. Applied Soft Computing, 2011, 11(4):3658-3670.
- [15] TANWEER M R, SURESH S, SUNDARARAJAN N. Self regulating particle swarm optimization algorithm [J]. Information Sciences, 2015, 294(10):182-202.
- [16] GAO Y, XIE S L. Particle swarm optimization algorithm based on Simulated annealing [J]. Computer Engineering and Applications, 2004, 40(1):47-50. (in Chinese)
- 高鹰,谢胜利. 基于模拟退火的粒子群优化算法 [J]. 计算机工程与应用, 2004, 40(1):47-50.
- [17] LIU A J, YANG Y, LI F, et al. Chaotic simulated annealing particle swarm optimization algorithm research and its application [J]. Journal of Zhejiang University (Engineering Science), 2013, 47(10):1722-1730. (in Chinese)
- 刘爱军,杨育,李斐,等. 混沌模拟退火粒子群优化算法研究及应用 [J]. 浙江大学学报(工学版), 2013, 47(10):1722-1730.
- [18] DAI M, TANG D, GIRET A, et al. Energy-efficient scheduling for a flexible flow shop using an improved genetic-simulated annealing algorithm [J]. Robotics and Computer-Integrated Manufacturing, 2013, 29(5):418-429.
- [19] ZHENG Y, ZHANG B. A simplified water wave optimization algorithm [C] // 2015 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2015:807-813.
- [20] DOWSLAND K A, THOMPSON J M. Simulated annealing [M]. Handbook of Natural Computing, Springer, 2012:1623-1655.
- [21] LIANG J J, QU B Y, SUGANTHAN P N, et al. Problem definitions and evaluation criteria for the CEC 2015 competition on learning-based real-parameter single objective optimization [R]. Computational Intelligence Laboratory, 2014.
- [22] ZHANG B, ZHENG Y J. Convergence Analysis of Water Wave Optimization Algorithm [J]. Computer Science, 2016, 43(4):41-44. (in Chinese)
- 张蓓,郑宇军. 水波优化算法收敛性分析 [J]. 计算机科学, 2016, 43(4):41-44.